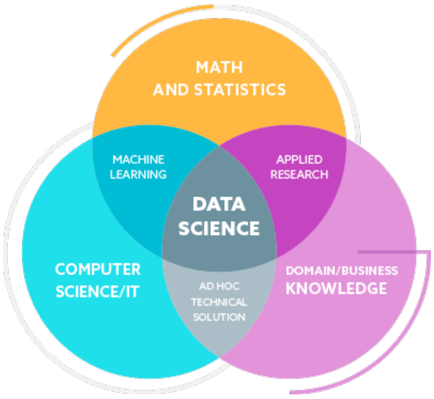
Computational Statistics Lecture Notes

Giovani Loiola da Silva - giovani.silva@tecnico.ulisboa.pt
Dep. Mathematics - Instituto Superior Técnico

7 November 2023

Contents

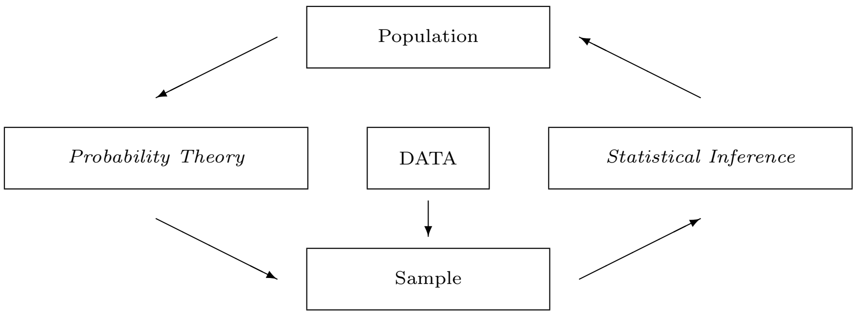


Circle of Data Science

1 Introduction

Brief description of the course object

- **Statistics**: Classification of experiments according to the exact (un)predictability of their results. Examples of *random experiments*:
  - Breakdown of structural defects in sheet metal used in industry;
  - Sex of the living being resulting from a fertilized egg.
- **Computational Statistics**: A class of statistical methods characterized by *computational intensity* and the supporting theory for such methods.



Schematic representation of a data set (sample) framed in Probability Theory and Statistical Inference

1.1 Review of classical methods of statistical inference*

Point estimation

Definition 1.1: Given a sample $(X_1 \dots, X_n)$ from a population X, a *statistic* T is a random variable (random vector) that is function of the sample, *i.e.*, $T = T(X_1, \dots, X_n)$.

Definition 1.2: A *parameter* is a measure used to describe a population characteristic, sometimes denoted by θ.

Definition 1.3: Let $(X_1 \dots, X_n)$ be a random sample (r.s.) of a population X indexed by the θ parameter. An *estimator* of θ is a statistic $T = T(X_1, \dots, X_n)$ used to estimate θ with its observed value $t = T(x_1, \dots, x_n)$ known as *estimate*.

Note 1.1: Estimation methods: maximum likelihood, moments, least squares, etc.

Interval estimation

Let $(X_1, \dots, X_n)$ be a random sample of a population X, indexed by a parameter θ. Sometimes it becomes more valuable to specify a range containing the true value of θ with a certain degree of confidence than just estimating θ as a point.

Definition 1.4: Let $(X_1, \dots, X_n)$ be a random sample of a population X indexed by a parameter $\theta \in \Theta$. If $T_i = T_i(X_1, \dots, X_n)$, $i = 1, 2$, are two statistics such that $T_1 < T_2$ and

$$P(T_1 < \theta < T_2) = \gamma.$$

where γ is a fixed value between 0 and 1, it is said that (T_1, T_2) is a *random confidence interval* (RCI) for θ with confidence γ.

Example 1.1: Let $(X_1, \ldots, X_n)$ be a r.s. from $X \sim N(\mu, \sigma^2 = 4)$. What is the RCI for μ with a confidence level of 95%?

$\bar{X}$ is known to be an estimator of μ and

$$\bar{X} \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad Z = \frac{\bar{X} - \mu}{\sqrt{4/n}} \sim N(0,1)$$

On the other hand, $P(-1.96 < Z < 1.96) = \gamma = 0.95$ and therefore

$$P(\bar{X} - 1.96\sqrt{4/n} < \mu < \bar{X} + 1.96\sqrt{4/n}) = 0.95,$$

indicating that the 95% random confidence interval for μ is expressed by (T_1, T_2), where

$$T_1 = \bar{X} - 1.96\sqrt{4/n} \quad \text{and} \quad T_2 = \bar{X} + 1.96\sqrt{4/n}.$$

Given a particular sample $(x_1, \ldots, x_n)$, the realization of the random confidence interval for μ with confidence degree γ is called confidence interval at $100\gamma\%$ to μ, given by

$$(t_1, t_2) = (t_1(x_1, \ldots, x_n), t_2(x_1, \ldots, x_n)) = \bar{x} - 1.96\sqrt{4/n}. \qquad \blacksquare$$

The probability γ is interpreted as the relative frequency of all intervals (t_1, t_2) containing θ obtained from an infinitely large sequence of repeated observations of $(X_1, \ldots, X_n)$ (frequentist perspective). Meantime,

$$\gamma \neq P(t_1 < \theta < t_2) = \begin{cases} 1, & \theta \in (t_1, t_2), \\ 0, & \text{o.w.} \end{cases}$$

Note 1.2: Confidence intervals are obtained here by the pivotal method.

Hypothesis testing

A parametric hypothesis test usually aims to compare different values for parameters of a given population X. For example, for the unknown parameter μ of $X \sim N(\mu, \sigma^2)$.

General procedure of a parametric hypothesis testing:

1. Hypotheses of interest:
   - Null hypothesis H_0 (*e.g.*, $\mu = \mu_0$; $\mu \geq \mu_0$ ou $\mu \leq \mu_0$).
   - Alternative hypothesis H_1 (*e.g.*, $\mu \neq \mu_0 \to$ bilateral test; $\mu < \mu_0$ or $\mu > \mu_0 \to$ unilateral tests).
2. Errors associated with the testing rule, whose corresponding probabilities are given by
   - $\alpha = P($ Type I error$) = P(\text{Reject } H_0 | H_0 \text{ true})$.
   - $\beta = P($ Type II error$) = P(\text{Accept } H_0 | H_0 \text{ false})$.
3. Critical region (CR):
   - Region leading to rejection of null hypothesis H_0 by testing rule. Constructed based on an appropriate statistic $T = T(X_1, \ldots, X_n)$ named testing statistic.

- The CR is constructed such that $P(T \in CR|H_0\,\text{true}) = \alpha$, with α (significance level) previously set to usual values 1%, 5% and 10%. This CR will be denoted by CR_α.

4. Hypothesis testing rule:

   - If $T \in CR_\alpha$, H_0 is rejected at the significance level of $100\alpha\%$. Otherwise, H_0 is not rejected at $100\alpha\%$.
   - As lower the significance level of the test as greater the precaution against the risk of incorrect rejection of H_0.

Determining β requires specifying each alternate value for the parameter under testing, since H_1 is usually composed (*e.g.*, $\beta(\mu) = P(T \notin CR|\mu \neq \mu_0)$).

The function $1-\beta(\mu)$ is known as test power for H_1 true. That is, for a given value of μ, the test power is the probability of rejection of H_0 when μ is the true value of the parameter.

$$P(\text{Reject}\,H_0|\mu) = \begin{cases} \alpha(\mu), & H_0 \text{ true} \\ 1-\beta(\mu), & H_1 \text{ true.} \end{cases}$$

Example 1.2: Let $X_1, \ldots, X_{16}$ be a r.s. of X (amount of coffee per packet) where the empirical mean and variance of one its realization are $480g$ and $800g^2$, respectively. Considering $X \sim N(\mu, \sigma^2)$, test if the machine is filling coffee packets with at least $500g$, at 5% significance level.

Hypothesis testing:

1. Hypotheses: $H_0 : \mu \geq 500$ versus $H_1 : \mu < 500$.
2. Test statistics: $T = \frac{\bar{X}-500}{S/\sqrt{n}} \overset{\mu=500}{\sim} t_{(15)}$, whose observed value is $t_0 = (480-500)/\sqrt{800/16} = -2.83$.
3. Unilateral critical region: Fixed $\alpha = 0.05$, $F^{-1}_{t_{(15)}}(0.95) = 1.753$ and $CR_{5\%} = (-\infty, -1.753)$, since decreasing values of T tend to reflect smaller values of μ.
4. Conclusion: Since $t_0 \in CR_{5\%}$, H_0 is rejected at a significance level of 5%, *i.e.*, there is evidence against the hypothesis of filling coffee packets with at least $500g$.

Note that the test decision varies with the choice of α, *i.e.*,

α	CR_α	test decision
0.05	$(-\infty, -1.753)$	reject H_0
0.01	$(-\infty, -2.602)$	reject H_0
0.006	$(-\infty, -2.857)$	not reject H_0

The smallest value of the significance level α leading to the rejection of H_0 is $P = P(T < -2.83|H_0) = 0.0063$. ■

P-value of the test

Definition 1.5: The *P-value* of a hypothesis testing is the probability under H_0 that the test statistic will take as much or more unfavorable values to H_0 than its observed value. Thus, H_0 will be rejected at all α significance levels such that $P < \alpha$ and accepted otherwise.

Pearson's chi-square goodness of fit test

Now it is interesting to know how one can test hypotheses about the distributional form of a given population, object of the so-called goodness of fit tests.

Construction of Pearson's chi-square test statistic:

1. Consider a random sample of n elements over which a variable X is observed, and their observations are classified into a real line partition, $B_1, \ldots, B_k$, so that O_i denotes the number of sample elements grouped into B_i, $i = 1, \ldots, k$, such that $\sum_{i=1}^{k} O_i = n$.

2. Let $p_i = P(X \in B_i)$ be the probability of obtaining an observation on i-th part of the partition, $i = 1, \ldots, k$, such that $\sum_{i=1}^{k} p_i = 1$.

3. The random vector $\mathbf{O} = (O_1, \ldots, O_k)$ has p.m.f. given by
$$f_{\mathbf{O}}(o_1, \ldots, o_k) = \frac{n!}{o_1! \ldots o_k!} \, p_1^{o_1} \, p_2^{o_2} \, \cdots \, p_k^{o_k},$$
known as Multinomial distribution $(n, \mathbf{p} = (p_1, \ldots, p_k))$, and it can be proved that $O_i \sim \text{Binomial}(n, p_i)$, $i = 1, \ldots, k$.

4. Hypotheses:
   - $H_0 : X \sim F_X(\cdot) \;\Rightarrow\; p_i = p_i^0, \; \forall \, i = 1, \ldots, k.$
   - $H_1 : X \nsim F_X(\cdot) \;\Rightarrow\; p_i \neq p_i^0, \; \exists \, i = 1, \ldots, k.$

5. Test statistics:
$$Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \underset{a}{\overset{H_0}{\sim}} \quad \chi^2_{(k-m-1)},$$
where $E_i = E(O_i | H_0) = n \, p_i^0$ and m is the total estimated parameters of $F_X(\cdot)$ under H_0. If $m > 0$, $\{p_i^0\}$ are still unknown implying that E_i are (appropriate) estimators of expected frequencies.

Regression model

There are random variables (Y) that can be explained by joint action of deterministic $(g(x)$ and random (ϵ) factors. Assuming an additive structure between them, the variable of interest is given by

$$Y = g(x) + \epsilon.$$

In this scenario, the dataset is made up of n pairs (y_i, x_i), $i = 1, \ldots, n$, with the x_i supposedly specified without error. Considering a random sample (Y_i, x_i), $i = 1, \ldots, n$, a statistical model for relating Y and x is the *simple linear regression model*

$$Y_i = \beta_0 + \beta_1 \, x_i + \epsilon_i,$$

where Y_i is the response variable of the i-th element of the sample, x_i is the corresponding value of the (fixed) explanatory variable, β_0 and β_1 are (unknown) parameters and ϵ_i is the random error of the element i of the sample.

Usual assumptions for random errors ϵ_i, $i = 1, \ldots, n$:

- $E(\epsilon_i) = 0$. This implies that for a given value of x,
$$E(Y|x) = \beta_0 + \beta_1 \, x,$$
so-called the equation or regression line of the model.

- $Var(\epsilon_i) = \sigma^2, \forall\, i$ constant variance.
- $\epsilon_1, \ldots, \epsilon_n$ are uncorrelated (or independent).
- ϵ_i follows a Normal distribution, $i = 1, \ldots, n$.

Interpretation of regression parameters:

- The intecept β_0 is the expected value of Y for a null value of the covariate x.
- The slope of the regression line β_1 is the variation of the expected value of Y for each unit increment by x.

1.2 Bayesian inference summary

Classical Statistics versus Bayesian Statistics

Data: realization of a r.v. X, $x \in \mathcal{X}$ / $x = (x_1, \ldots, x_n)$, $\mathcal{X} \subseteq I\!R^n$

Statistical model: Specification based on the nature of the phenomenon, pretreatment of analogous phenomena, experimental evidence, study objectives, parsimony requirements and interpretability.

$$\mathcal{F} = \{f(x|\theta),\, x \in \mathcal{X} : \theta \in \Theta\},$$

but unaware of the value of the θ index that produced the data; *e.g.*,

$$\Theta \in I\!R^k, f(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \quad \text{(sample model).} \tag{1.1}$$

Classic paradigm

Principle of repeated sampling: Evaluator of inferential procedures by analyzing their behavior in an indefinite number of hypothetical repetitions under essentially identical conditions of the random scheme originating the sample (assumption).

$\Longrightarrow$ uncertainty measurement based on the *frequentist concept* of probability.

Inferential way: (Totally or partially) observable random variables and their sample distributions associated with $\mathcal{F}$, based on which the properties of inferences are evaluated pre-experimentally.

Example 1.3: Clinical trial on a n patient sample to infer the probability θ of controlling their disease with a new drug.

Data: $(x_1, \ldots, x_n) \longleftarrow (X_1, \ldots, X_n) : X_i,\, i = 1, \ldots, n, \underset{iid}{\sim}$ Bernoulli(θ).

- (Frequentist) interpretation of θ: $\lim_{n\to\infty} \dfrac{1}{n} \sum_{i=1}^{n} X_i$.
- Test of $H_0 : \theta \geq 80\%$.
  Using the test statistic: $T = \frac{\bar{X} - 0.80}{\sqrt{0.80 \times 0.20/n}} \underset{\theta=0.80}{\overset{a}{\sim}} N(0, 1)$, with the maximum probability of incorrect rejection of H_0 of 5%, there is evidence against such conjecture if it turns out that

$$\bar{X} \leq 0.80 - 1.645\sqrt{0.8 \times 0.2/n}.$$

- Estimation by intervals of θ: $\frac{\bar{X} - \theta}{\sqrt{\bar{X}(1-\bar{X})/n}} \overset{a}{\sim} N(0, 1)$

  $\Rightarrow P\left[\left(\bar{X} \pm 1.96\sqrt{\bar{X}(1 - \bar{X})/n}\right) \text{ contain value } \theta \mid \theta\right] = 0.95.$ ■

Bayesian paradigm

Bayes' theorem (Theorem A.1): Fundamental inferential instrument.

Work elements:

1. Model sample **data** $\{f(x|\theta) : \theta \in \Theta\}$ in (1.1),

2. Prior information (previous or external to such a sample) about what is unknown quantified in (initial) **prior distribution**:

$$h(\theta) : \theta \in \Theta. \tag{1.2}$$

On the basis of the crucial argument, 'All that is unknown is uncertain and all uncertainty is likely to be probabilistically quantified!'

$\Rightarrow$ parameters of the sample models viewed as random on a typically subjective basis.

Subjectivist concept of probability — Degree of personal belief in the occurrence of the event (truth of the proposition), based on available evidence.

$\Longrightarrow$ Posterior distribution (final):

$$h(\theta|x) = \frac{h(\theta)f(x|\theta)}{p(x)} \equiv \frac{h(\theta)f(x|\theta)}{\int_\Theta f(x|\theta)h(\theta)d\theta} \quad \theta \in \Theta. \tag{1.3}$$

where $p(x), \forall x$, translates to the **marginal distribution** of observable data X.

$\Rightarrow$ *Precision of inferences*: post-experimental (final).

Example 1.4 (*vide* Example 1.3):

- Sample model: $f(x|\theta) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$.

- Prior distribution: $\theta \sim \text{Beta}(a, b)$, i.e., $h(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$

$$E(\theta) = a/(a+b) = 0.40, Var(\theta) = ab/((a+b)^2(a+b+1)) = 0.01$$
$$\Rightarrow \theta \sim \text{Beta}(9.2, 13.8).$$

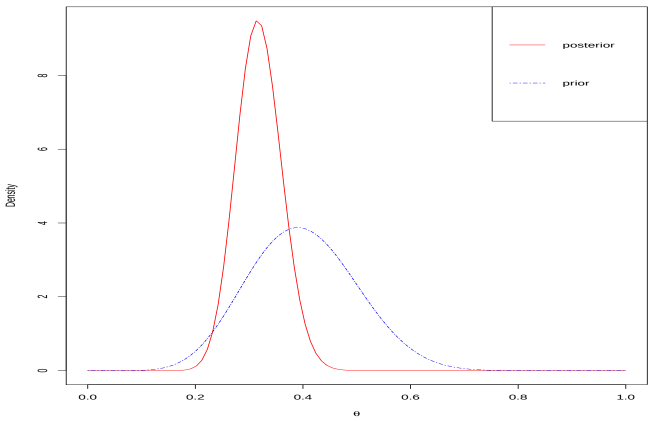
$$\Rightarrow h(\theta|x_1, \ldots, x_n) \propto \theta^{a+\sum_i x_i - 1}(1-\theta)^{b+n-\sum_i x_i - 1},\ \theta \in (0,1)$$
$$\therefore \quad \theta|x_1, \ldots, x_n \sim \text{Beta}(A, B),\ A = a + \sum_i x_i;\ B = b + n - \sum_i x_i.$$

If in the data, $n = 100$, $\sum_i x_i = 30$, then $A = 39.2$, $B = 83.8$ $\Rightarrow$

- $E(\theta|x) = 0.319$;

- $\sqrt{Var(\theta|x)} = 0.04$;

- $P\left[\theta \in (0.238, 0.401)|x\right] = 0.95$.



Prior Beta(9.2,13.8) and posterior Beta(39.2,83.8) distributions

■

Characteristics of Bayesian methodology

- The posterior distribution $h(\theta|x)$, $\theta \in \Theta$, is the complete description of current knowledge about θ, obtained from quantifying prior information (in $h(\theta)$) and sample information (in $f(x|\theta)$).
- The relevant part of $f(x|\theta)$ for inferential purposes is the factor involving θ. Taking it as the likelihood function $L(\theta|x)$, this is regarded as the vehicle of all sample information.

  $\Rightarrow$ Proportional likelihood functions lead to the same posterior distribution, which is the fulcrum of all inferences about the parameter.

  $\Rightarrow$ It satisfies principles (sufficiency, conditionality and likelihood) that all Statistical Inference is supposed to respect.

- The Bayesian knowledge update operation has a sequential nature: $X = (X_1, X_2)$, $\quad X_1 \perp\!\!\!\perp X_2 \mid \theta$

$$h(\theta|x) = \frac{h(\theta|x_1)f(x_2|\theta)}{\int_\Theta h(\theta|x_1)f(x_2|\theta)d\theta}, \quad p(x_1) > 0$$

  $\therefore \quad h(\theta|x)$ is the update of $h(\theta|x_1)$ as prior distribution by likelihood $f(x_2|\theta)$.

- Conceptual simplicity and uniformity in eliminating disturbing parameters, $\theta = (\gamma, \phi) \in \Gamma \times \Phi$, γ parameter of interest.

  $\Rightarrow \quad$ Calculation of $h(\gamma|x) = \int_\Phi h(\gamma, \phi|x)d\phi$.

Point estimation

This problem consists in determining a typical point of the posterior distribution. Possible choices:

- Posterior mode

$$\hat{\theta}: \quad h(\hat{\theta}|x) = \max_{\theta\in\Theta} h(\theta|x) = \max_{\theta\in\Theta} \left[h(\theta)f(x|\theta)\right].$$

- Posterior mean

$$\hat{\theta} = E\left[\theta|x\right]: \quad E\left[\theta_i|x\right] = \int_\Theta \theta_i h(\theta|x)d\theta, \; \forall\, \theta_i \text{ de } \theta.$$

- Posterior median vector

$$\hat{\theta} = (\hat{\theta}_i) : \quad \begin{cases} P\,[\,\theta_i \geq \hat{\theta}_i\ |x\,] \geq 1/2 \\ P\,[\,\theta_i \leq \hat{\theta}_i\ |x\,] \geq 1/2,\ \forall i. \end{cases}$$

Example 1.5 (*vide* Example 1.4): Bernoulli $\wedge$ Beta Bayesian model. Taking $A = a + \sum_i x_i = 39.2$ and $B = b + n - \sum_i x_i = 83.8$, $\theta|\{x_i\} \sim \text{Beta}(A, B)$

- Posterior mode: $\theta_{mo} = \dfrac{A-1}{A+B-2} = 0.316$
- Posterior mean: $\theta_{me} = \dfrac{A}{A+B} = 0.319$
- Posterior median: $\theta_{md} = F^{-1}_{\text{Beta}(A,B)}(1/2) = 0.318$. ■

Note 1.3: A complete justification for any given option (beyond its relevance to the problem) requires the incorporation in the analysis of additional information on the consequences (costs) of each one
$\Rightarrow$ Bayes estimators (scope of **Statistical Decision Theory**).

Region estimation

A summary of $h(\theta|x)$ more informative than any point estimate is obtained from a region of Θ that contains a substantial part of the posterior probabilistic mass — the parallel Bayesian of confidence region:

Definition 1.6: $R(x)$ is a *credible region* (CR) γ for θ, if

$$P\,[\theta \in R(x)|x] \;\equiv\; \int_{R(x)} h(\theta|x)d\theta \;\geq\; \gamma.$$

Observations:

- The entire credible region is numerically defined (*i.e.*, not random) and allows for a direct and unambiguous probabilistic interpretation — in contrast to the classical confidence region.
- Given the infinity of CR with the same credibility γ, it obviously matters to select the one that encompasses all the most credible posterior values of θ, *i.e.*, the one that satisfies the condition

$$h(\theta_1|x) \geq h(\theta_2|x), \quad \forall\, \theta_1 \in R(x),\ \theta_2 \notin R(x).$$

Definition 1.7 *HPD* criteria (*High Posterior Density*) or minimum volume): $R(x)$ is the credibility region γ with maximum posterior density (probability) if

$$R(x) = \{\theta : h(\theta|x) \geq c_\gamma\}\,,$$

with $c_\gamma > 0$ the largest constant such that $P\,[\theta \in R(x)|x] \geq \gamma$.

- HPD CRs are not invariant against nonlinear parametric transformations.
- Determining HPD CRs in practice often requires recourse to numerical methods, unless for $\theta \in I\!R$ $h(\theta|x)$ is a symmetric function. For posterior continuous distributions in $I\!R$, the numeric calculation of HPD CR $R(x|c) = \{\theta : h(\theta|x) \geq c\}$ demands:
  - a first subroutine that finds the solutions of the equations $h(\theta|x) = c$ to $c > 0$ variable defining $R(x|c)$;

- a second subroutine that evaluates the probabilities $P\left[\theta \in R(x|c)|x\right]$.

Once c is found such that $P\left[\theta \in R(x|c)|x\right] = \gamma$, the region $R(x|c)$ will be HPD with credibility γ.

Example 1.6 (*vide* Example 1.5): Bernoulli $\wedge$ Beta Bayesian model
$\Rightarrow \theta|x \sim \text{Beta}(39.2, 83.8)$

- 95% HPD CI: (0.238, 0.401)
  - determinable by software FIRST BAYES
  (`http://tonyohagan.co.uk/1b/`)

- 95% HPD CI: (0.240, 0.403)
  - determinable by software R
  (`http://www.r-project.org/`)[1].

■

Example 1.7: (Known standard deviation) Normal $\wedge$ 'Uniform' Bayesian model
$\{X_1, \ldots, X_n\}$ i.i.d. of $\{N(\mu, \sigma^2), \sigma^2 \text{ known}\}$ and $h(\mu) = k$
$\Rightarrow \quad \mu|x \sim N(\bar{x}, \sigma^2/n)$.
Point estimate of μ: $\hat{\mu} = \bar{x}$
$100\gamma\%$ HPD credible region:

$$R(x) = \left\{ \bar{x} \pm \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{1+\gamma}{2}\right) \right\}$$

Note 1.4:
$$\begin{array}{rcll} P\left[\mu \in R(x)|x\right] & = & \gamma & \text{(final precision measurement)} \\ P\left[\mu \in R(X)|\mu\right] & = & \gamma & \text{(initial precision measurement)} \\ P\left[\mu \in R(x)|\mu\right] & = & I_{R(x)}(\mu). & \end{array}$$

■

Hypothesis testing

The problem of testing

$$H_0 : \theta \in \Theta_0 \text{ contra } H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$$

is also conceptually simpler than in a classical context. Given the direct probabilistic interpretation of the confronting hypotheses, one has no more than to calculate the respective posterior probabilities and to opt for one of them on the basis of some criterion based on their relative magnitude.
$\Rightarrow$ Calculation of the *posterior odds* pro-H_0:

$$O(H_0, H_1|x) = \frac{P\left[H_0|x\right]}{P\left[H_1|x\right]}. \tag{1.4}$$

In order to measure the influence of the data x on changing the relative credibility of H_0 and H_1, we choose to counteract the posterior odds in favor of H_0 to their respective posterior odds via
Bayes Factor pro-H_0:

$$B(x) = \frac{P\left[H_0|x\right]/P\left[H_1|x\right]}{P\left[H_0\right]/P\left[H_1\right]} = \frac{\int_{\Theta_0} f(x|\theta)h_0(\theta)d\theta}{\int_{\Theta_1} f(x|\theta)h_1(\theta)d\theta}, \tag{1.5}$$

where $h_i(\theta)$ is the (p.d.f.) prior distribution conditioned on H_i, i.e., $h_i(\theta) = h(\theta)/\int_{\Theta_i} h(\theta)d\theta$, $i = 0, 1$.

[1] R function *qbeta e.g.* $qbeta(0.975, 39.2, 83.8) = 0.4033296$.

A situation where $B(x) >> 1$ or $B(x) << 1$ reflects a very strong trend in the data in favor of one hypothesis against another, in the sense that a hypothesis is much more or much less likely *a posteriori* than it was *a priori*.

In inferential practice, it is often used guiding rules on the interpretation of evidence contained in the data, such as these

$B(x)$	$2\ln B(x)$	Evidence
< 1	< 0	pro-H_1
$1-3$	$0-2$	weak pro-H_0
$3-20$	$2-6$	pro-H_0
$20-150$	$6-10$	strong pro-H_0
> 150	> 10	very strong pro-H_0

Example 1.8 (*vide* Example 1.6): Model Bernoulli $\wedge$ Beta with $\theta \sim \text{Beta}(9.2, 13.8)$ e $n = 100$, $\sum_i x_i = 30$ $\Rightarrow \theta|x \sim \text{Beta}(39.2, 83.8)$.

Problem: $\theta \geq 35\%$ *versus* $\theta < 35\%$

$$O(H_0, H_1) = \frac{0.679}{0.321} = 2.115; \quad O(H_0, H_1|x) = \frac{0.225}{0.775} = 0.29$$

$\Rightarrow B(x) = 0.137 \Leftrightarrow 1/B(x) = 7.3$

$\therefore$ The posterior odds for H_1 is more than 7 times the respective prior odds, implying that H_1 is even more likely (but only about 3 times) than H_0 as a consequence of H_1 is less likely than H_0 *a priori*. ■

Comments:

1. The form of Bayesian tests eliminates the need for a formal distinction between what is the null hypothesis and what is the alternative hypothesis and the asymmetrical nature of the classical test.

2. The *P-value* of unilateral tests may have Bayesian justification as a posterior probability of the null hypothesis and the two quantities may be similar (but not equal) or radically different.

3. Although $P[H_0|x]$ and the critical level (P-value) coincide, the conclusions of the Bayesian and classical test may be contrary.

4. Given the essential form of Bayesian tests, the problem of testing multiple hypotheses does not entail additional difficulties compared to the usual problem of confronting two hypotheses.

Prediction

Based on observations x of a random vector $X \sim f(x|\theta)$ (and eventually all accumulated knowledge about θ), it is intended to predict Y with sample distribution dependent on θ.

This problem, which may be controversial in the classical approach, also has, in Bayesian perspective, a conceptually (at least) simple solution: calculation of *posterior predictive distribution*

$$p(y|x) = \int_{\Theta} f(y|x, \theta) h(\theta|x) d\theta. \tag{1.6}$$

Once this is obtained, measurements that can be summarized can be determined, such as point predictions (mode, mean prediction, etc.) and regional predictions (prediction regions with the highest predictive density) of Y.

Example 1.9 (*vide* Example 1.7): (known standard deviation) Normal $\wedge$ 'Uniform' Bayesian model

$$f(x|\mu) \propto \left(2\pi\frac{\sigma^2}{n}\right)^{-1/2} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\} \;\wedge\; h(\mu) = k$$

$\Longrightarrow h(\mu|x) =$ f.d.p. de $N(\bar{x}, \sigma^2/n)$.

Mean prediction $\bar{Y}$ from m future i.i.d. observations from distribution $N(\mu, \sigma^2)$ and independent of X given μ, $\bar{Y}|x, \mu \sim N(\mu, \sigma^2/m)$:

$$\therefore \quad \bar{Y}|x \sim N(\bar{x}, \sigma^2\left((1/m) + (1/n)\right)).$$

$\Rightarrow$ Point prediction of $\bar{Y}$: $\bar{x}$

$\Rightarrow$ 95% HPD prediction interval for $\bar{Y}$: $\left(\bar{x} \pm 1.96\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}\right)$ ■

Representation of prior information

Eliciting a distribution that represents one's prior beliefs: This is a particularly difficult task and is surrounded by a series of contingencies.

Special situations:

- Scarce ('vague', 'diffuse') prior knowledge state

  $\Rightarrow$ *Non-informative distributions*

- Adoption of a suitable functional form and specification of hyperparameters (through their relationship to quantiles and/or moments *a priori*) according to elicited prior beliefs

  $\Rightarrow$ *Natural conjugate distributions*

Objectives of using non-informative prior distributions:

1. Description of situations where prior knowledge is little or nothing significant with respect to sample information;

2. Performance of a reference role, although prior strong beliefs as a way of:

   a to deduce posterior beliefs for those who start from a scarce knowledge, *i.e.*, when the sample provides most of the parameter information;

   b to allow comparison with the results of classical inference that 'only' uses the sample information (in whole or in part);

   c to ascertain the influence on the inferences of the prior distribution that describes the information that actually exists, when compared with those resulting from the use of the prior distribution of reference.

Jeffreys' rule

- θ is a location parameter ($\Theta \in I\!R$)

  Invariance under one-to-one translations: for every a, the intervals $(\theta_0, \theta_0 + a)$, $\forall\theta_0 \in I\!R$, must be the same probability.

$$\Rightarrow \quad h(\theta) = c, \quad \theta \in \Theta \quad \text{('continuous uniform')}$$

- θ is a scale parameter ($\Theta = I\!R^+$)

  Invariance under scale transformations: For each $b > 0$, the intervals $(\theta_0, b\theta_0)$, $\forall \theta_0 \in I\!R^+$, must have the same probability $\Rightarrow$ as $\ln\theta$ is the location parameter of the data logarithmic transformation

$$\begin{aligned} h^*(\ln\theta) = c \quad &\Rightarrow \quad h(\theta) \propto \theta^{-1},\ \theta > 0 \\ &\Rightarrow \quad h_*(\theta^a) \propto (\theta^a)^{-1},\ \theta > 0, \quad \forall a \in \mathbb{Z} \end{aligned}$$

- θ is a generic vectorial parameter

  Invariance that guarantees the identity of resulting inferences from the use of any biunivocal transformation - satisfied with the use of Fisher's information measure, $\mathcal{I}(\theta)$[2]

  $\Rightarrow \quad h(\theta) \propto [|\mathcal{I}(\theta)|]^{1/2}$

  Note 1.5: Since it is reasonable to admit prior independence (as can be in distinct type parameters), the prior distribution must verify this condition with the marginal distributions defined by applying the previous rule.

Example 1.10: Parameter $\theta = (\mu, \sigma^2)$ of the model $\{N(\mu, \sigma^2)\}$

$$\mathcal{I}(\mu, \sigma^2) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{pmatrix}$$

Multiparametric Jeffreys' rule $\quad\Rightarrow\quad h(\mu, \sigma^2) \propto 1/\sigma^3, \quad (\mu, \sigma^2) \in I\!R \times I\!R^+$
Uniparametric Jeffreys' rule + prior independence $\quad\Rightarrow\quad h(\mu, \sigma^2) \propto 1/\sigma^2, \quad (\mu, \sigma^2) \in I\!R \times I\!R^+$ $\blacksquare$

$\therefore$ Argument subject to criticism (often inappropriate nature; dependence on sample model) and counter-criticism.

Natural conjugate distributions

The success of the chosen distributional form in quantifying prior beliefs and triggering inferences is naturally associated with:

- family versatility;
- simplicity of the analytical derivation of the posterior distribution;
- ease of interpretation of the Bayesian operation in the combination of prior and sample information.

Definition 1.8: The family $\mathcal{H}$ is said to be *natural conjugate* of the $\mathcal{F} = \{f(x|\theta) : \theta \in \Theta\}$ if $h(\theta|x) \in \mathcal{H}$ whenever the corresponding $h(\theta) \in \mathcal{H}$.

Example 1.11 (*vide* Example 1.8): Bernoulli $\wedge$ Beta Bayesian model

$$\begin{aligned} f(x|\theta) &= \theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}, \quad 0 < \theta < 1 \\ &\equiv \text{ kernel of Beta}(\Sigma x_i + 1, n - \Sigma x_i + 1) \end{aligned}$$

If $h(\theta) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}I_{(0,1)}(\theta)$ (member pf the Beta family)
$\Rightarrow \quad \theta|x \sim Be(a + \Sigma x_i, b + n - \Sigma x_i), \quad a, b > 0$

That is, the Beta family is natural conjugate of a random sampling of the model $\{Ber(\theta)\}$ being:

[2]If $\log f(x|\theta)$ is twice differentiable with respect to θ, and under certain regularity conditions, the Fisher information may be written as $\mathcal{I}(\theta) = -E[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)|\theta]$.

1. The Beta family is quite versatile.
2. The update of $h(\theta)$ is made within the family.
3. The information in $h(\theta|x)$ translates into the sum of the successes and failures of the real sample with those of the fictional sample (a, b). ■

Example 1.12: Poisson $\wedge$ Gamma Bayesian model
$L(\theta|x) \propto \theta^{\sum_i x_i} e^{-n\theta} \equiv$ kernel of $Ga(\sum_i x_i + 1, n)$
$\therefore$ natural conjugate distribution: $\theta \sim Ga(a, b)$, $a, b > 0$
$\Rightarrow$ $\theta|x \sim Ga(a + \sum_i x_i, b + n)$
$\Rightarrow$ non-informative distribution: $\tau_0 = (0, 0)$

$$h(\theta) \propto \theta^{-1} I_{(0,+\infty)}(\theta) \quad \Leftrightarrow \quad h^*(\psi) = c, \quad \text{where } \psi = \ln\theta.$$

■

1.3 Use of stochastic simulation

'Computer-intensive statistics' is statistics that could only be done with modern computing resources (Ripley, 1987), typically either

- Statistical inference on small problems which needs a lot of computation to do at all, or to do well.
- Statistical inference on 'huge' problems.

Computers have revolutionized Statistics in namely two important issues:

- possibility of making inferences without the assumptions which standard techniques necessitate in order to obtain analytic solutions - normality, linearity, independence, etc.
- application of standard models to situations of big data complexity - missing, censoring, etc.

Systems with components partially or totally subject to random behavior are the basis of Statistics and their *simulation process*[3] is stochastic, *i.e.*, based on probability distributions (Gamerman & Lopes, 2006).

The starting point of stochastic simulation is the construction of a random number generator. Usually this mechanism generates numbers in the range $[0, M]$ for a given value of M. You can reduce that generation to a number by $[0, 1]$ by dividing it by M, being more appropriate to say *pseudorandom numbers* (see *e.g.* Ripley, 1987).

There are some methods of generating random quantities with discrete or continuous probability distribution based on a u random quantity generated from an *uniform* distribution in the range $[0, 1]$.

Inverse transform method

Theorem 1.1: If X is a continuous r.v. with strictly increasing distribution function F, then $F(X)$ is uniformly distributed on $(0, 1)$.

[3] The term simulation refers to the treatment of real problems through reproductions in investigator-controlled environments.

If $U \sim$ Uniforme$(0,1)$, then for all $x \in I\!R$

$$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

and therefore $F^{-1}(U)$ has the same distribution as X.

This *inverse transform method* can be applied for generating continuous and discrete r.v., being summarized as follows:

1. Derive and compute the inverse function $F_X^{-1}(u)$.
2. For each random quantity, generate a random u from Uniform$(0,1)$ and calculate $x = F_X^{-1}(u)$.

Example 1.13: To simulate from the exponential distribution with parameter λ, we use that

$$F(x) = 1 - \exp(-\lambda\, x),\ \ x > 0,$$

so

$$F^{-1}(u) = -\lambda^{-1} log(1-u),\ \ 0 < u < 1.$$

Since $U \sim$ Uniform$(0,1)$ implies that $1 - U \sim$ Uniform$(0,1)$, we have that

$$-\lambda^{-1} log(1-U_1), \ldots, -\lambda^{-1} log(1-U_n)$$

is a sequence of independent random variables from the exponential distribution with parameter $\lambda > 0$. ■

Example 1.14: To simulate from $X \sim$ Bernoulli(θ), we use that $F(0) = f(0) = 1 - \theta$, and $F(1) = f(0) + f(1) = 1$, and thus we divide the interval $(0,1)$ into $k = 2$ intervals $I_i = (F_{i-1}, F_i]$, where $F_0 = 0$, $F_i = F(i)$, $i = 1, \ldots, k$. Thus,

$$F^{-1}(u) = \begin{cases} 0, & u \leq 1 - \theta, \\ 1, & u > 1 - \theta \end{cases}$$

The generator should therefore provide the numerical value of the logical expression $u > 1 - \theta$. ■

In the generic case of a quantity X with values in $\{x_1, \ldots, x_k\}$ and respective probabilities $p_1, \ldots, p_k$, restricted to $\sum_{i=1}^k p_i = 1$, we can use the intervals above, where each interval corresponds to an unique value x and after observe the generating value of u, we verify the interval I_i in which is contained that value.

Transformation method

In addition to the inverse transformation method, other transformation methods can be applied to simulate random variables (Rizzo, 2019), using *e.g.* the following relations between probability distributions:

- If $Z \sim N(0,1)$, then $V = Z^2 \sim \chi^2_{(1)}$.
- If $Z \sim N(0,1)$ and $V \sim \chi^2_{(n)}$ are independent, then $T = \frac{Z}{\sqrt{V/n}}$ has t-Student distribution with n degrees of freedom.
- If $U \sim \chi^2_{(m)}$ and $V \sim \chi^2_{(n)}$ are independent, then $F = \frac{U/m}{V/n}$ has the F distribution with (m,n) degrees of freedom.
- If $U \sim$ Gamma(a,b) and $V \sim$ Gamma(c,b) are independent, then $X = \frac{U}{U+V}$ has the Beta(a,c) distribution.

Example 1.15: The following relation between Normal and Uniform distributions provides another standard Normal generator.

If $U, V \sim$ Uniform$(0,1)$ are independent, then

$$Z_1 = \sqrt{-2\log U}\,\cos(2\pi V) \qquad Z_2 = \sqrt{-2\log V}\,\sin(2\pi U)$$

are independent standard Normal variables.

These transformations determine an algorithm for generating two random $N(0,1)$ variables

1. Generate a random u from $U(0,1)$.
2. Generate a random v from $U(0,1)$.
3. calculate $z_1 = \sqrt{-2\log u}\cos(2\pi v)$ and $z_2 = \sqrt{-2\log v}\sin(2\pi u)$. ■

Mixture method

Definition 1.9: Let $X_1, \ldots, X_n$ be independent and identically distributed r.v. with $X_i \sim X$. The distribution function of $S = \sum_{i=1}^n X_i$ is called the *convolution* of X.

It is straightforward to simulate a convolution by directly generating $X_1, \ldots, X_n$ and computing the sum S.

Definition 1.10: A r.v. X is a discrete or continuous *mixture* if the distribution of X is, respectively:

- $F_X(x) = \sum_{i=1}^n p_i F_{X_i}(x)$ for some sequence of r.v. $X_1, X_2, \ldots$ and $p_i > 0$ (mixing weight) such that $\sum_{i=1}^n p_i = 1$.
- $F_X(x) = \int_{-\infty}^{\infty} F_{X|Y=y}(x)\, f_Y(y)\, dy$ for a conditional family $X|Y=y$ and weighting function f_y such that $\int_{-\infty}^{\infty} f_Y(y)\, dy = 1$.

Example 1.16: Suppose $X_1 \sim N(-1,1)$ and $X_2 \sim N(2,1)$ are independent r.v.. Let $S = X_1 + X_2$ denote a convolution. Define a Normal mixture X with $F_X(x) = 0.3\, F_{X_1}(x) + 0.7\, F_{X_2}(x)$.

To simulate the convolution:

1. Generate $x_1 \sim N(-1,1)$ and $x_2 \sim N(2,1)$.
2. Calculate $s = x_1 + x_2$.

To simulate the mixture:

1. Generate an integer $k \in \{1,2\}$, where $P(1) = 0.3$, $P(2) = 0.7$.
2. If $k=1$, calculate random x from $N(-1,1)$;
   If $k=2$, calculate random x from $N(2,1)$. ■

2 Classic Estimation Methods and Algorithms

There are several estimation methods for making inference especially under *Classic/Frequentist approach e.g.*

- Least square method in linear model.
- Maximum likelihood method.
- Moment method.

Sometimes these methods require algorithms taking into account several specific contexts *e.g.*

- Newton-Raphson algorithm.
- EM algorithm.
- Data augmented algorithm.

2.1 Least square method*

The *least squares method* has been employed for estimating the regression parameters of the general linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ is the response variable observation vector, $\mathbf{X}$ is a matrix $n \times p$ of model specification (design matrix), with i-th row $\mathbf{x}_i^T = (x_{i1}, \ldots, x_{ip})$ representing the observation of p covariates of the individual i, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the vector of regression parameters and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$ is the vector of random components.

Assume $r(\mathbf{X}) = p$, $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$ and $\boldsymbol{\epsilon} \sim N_p(\cdot, \cdot)$. That is so-called as *full rank (Normal) linear model.*

A method of estimation of the regression coefficients $\boldsymbol{\beta}$ is the method of least squares that is to minimize the sum of squares of random errors. That is, the value that minimizes the function

$$\begin{aligned} SS(\boldsymbol{\beta}) &= \textstyle\sum_{i=1}^n \epsilon_i^2 = \boldsymbol{\epsilon}^T\boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}, \end{aligned} \tag{2.2}$$

denoted by $\widehat{\boldsymbol{\beta}}$, is called the *least squares estimator* of $\boldsymbol{\beta}$.

The partial derivative of $SS(\boldsymbol{\beta})$, valued at $\{(y_i, \mathbf{x}_i)\}$ related to $\boldsymbol{\beta}$

$$\begin{aligned} &\tfrac{\partial SS(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\mathbf{X}^T\mathbf{Y} - (\mathbf{Y}^T\mathbf{X})^T + [(\mathbf{X}^T\mathbf{X}) + (\mathbf{X}^T\mathbf{X})^T]\boldsymbol{\beta} = -2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\ &\tfrac{\partial SS(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0} \quad \Rightarrow \quad \mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{Y} \quad \text{(normal equations)} \end{aligned}$$

As $r(\mathbf{X}) = r(\mathbf{X}^T\mathbf{X}) = p$, the matrix $\mathbf{X}^T\mathbf{X}$ is not unique and therefore the only solution of the normal equations is $\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$. Besides,

$$\frac{\partial^2 SS(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} = \frac{\partial}{\partial \boldsymbol{\beta}}\left(\frac{\partial SS(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\right) = \frac{\partial}{\partial \boldsymbol{\beta}}\left(\frac{\partial SS(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)^T = 2\mathbf{X}^T\mathbf{X} \ .$$

and the matrix $\mathbf{X}^T\mathbf{X}$ is positive definite[4] (or $2\mathbf{X}^T\mathbf{X}$) for any value of $\boldsymbol{\beta}$. Therefore, the *least squares estimator* of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{2.3}$$

- $E(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E(\mathbf{Y}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\,\boldsymbol{\beta} = \boldsymbol{\beta}$.
- $Var(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Var(\mathbf{Y})\,\mathbf{X}\,(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$.
- $\widehat{\boldsymbol{\beta}}$ is an *unbiased* estimator to $\boldsymbol{\beta}$.
- An unbiased estimator for σ^2 (*mean squares of error/residual*) is

$$MSE \equiv \frac{SSE}{n-p} = \frac{\mathbf{r}^T\mathbf{r}}{n-p} = \frac{(\mathbf{Y}-\hat{\mathbf{Y}})^T(\mathbf{Y}-\hat{\mathbf{Y}})}{n-p} = \frac{\mathbf{Y}^T(\mathbf{I}-\mathbf{H})\mathbf{Y}}{n-p}, \tag{2.4}$$

  where $\hat{\mathbf{Y}} \equiv \widehat{E}(\mathbf{Y}) = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{Y}$ and $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ (*'hat' matrix*). For further details, see *e.g.* Ross, 2014; Kutner *et al.* 2005.

Definition 2.1: It is said that $T = T(Y_1, \ldots, Y_n)$ is a *linear* estimator of θ, if T is a linear combination of $\mathbf{Y}$.

Theorem 2.1: For the general linear model (2.1), Gauss-Markov structure and full rank ($r(\mathbf{X})=p$), $\mathbf{c}^T\widehat{\boldsymbol{\beta}}$ is the *unbiased linear estimator* of $\mathbf{c}^T\boldsymbol{\beta}$ with minimum variance, where $\mathbf{c}$ is a $p \times 1$ vector of constants.

2.2 Maximum likelihood method*

Definition 2.2: Given a r.s. $(X_1 \ldots, X_n)$ from a population X with p.m.f. or p.d.f. $f_X(x|\theta)$ indexed by the (unknown) parameter θ, the *likelihood* function of θ related to the sample $(x_1, \ldots, x_n)$, denoted by $L(\theta|x_1, \ldots, x_n)$, is the function of θ which is numerically identical to the sample probability distribution evaluated at $(x_1, \ldots, x_n)$, *i.e.*,

$$L(\theta|x_1, \ldots, x_n) \equiv f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\theta) = \prod_{i=1}^n f_X(x_i|\theta). \tag{2.5}$$

The maximum likelihood method is to maximize the likelihood function to obtain the most likely said value of θ, called the maximum likelihood estimate of θ.[5]

Example 2.1: Let $(X_1 \ldots, X_n)$ be a r.s. of $X \sim$ Poisson(λ). What is the maximum likelihood estimator (MLE) of λ?

The likelihood function of λ, given $(x_1, \ldots, x_n)$, is

$$L(\lambda|x_1, \ldots, x_n) = \prod_{i=1}^n \frac{\mathrm{e}^{-\lambda}\lambda^{x_i}}{x_i!}.$$

As $L_\lambda \equiv \log L(\lambda|x_1, \ldots, x_n) = -n\,\lambda + \log\lambda\sum_{i=1}^n x_i - \log\prod_{i=1}^n x_i!$.

- $\frac{dL_\lambda}{d\lambda} = -n + \lambda^{-1}\sum_{i=1}^n x_i = 0 \quad \Rightarrow \ \lambda = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}$

[4]As $\mathbf{X}^T\mathbf{X}$ is a symmetric matrix, $\mathbf{z}^T\mathbf{X}^T\mathbf{X}\mathbf{z} = \sum_{i=1}^n(\mathbf{x}_i^T\mathbf{z})^2 \ \geq \ 0, \forall\,\mathbf{z} \in I\!R^p$, and $\mathbf{X}\mathbf{z} = \mathbf{0}$ only when $\mathbf{z} = \mathbf{0}$, then $\mathbf{z} = \mathbf{0}$ is the only system solution.

[5]In determining the maximum of $L(\theta|x_1, \ldots, x_n)$, it is often used that $L(\theta|x_1, \ldots, x_n)$ and $\log L(\theta|x_1, \ldots, x_n)$ have their maximum at the same value of θ.

- $\frac{d^2 L_\lambda}{d\lambda^2} = -\lambda^{-2} \sum_{i=1}^n x_i < 0, \forall \lambda.$

$\therefore$ $\bar{x}$ is the maximum likelihood estimate of λ and the MLE of λ is

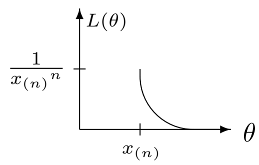
$$\hat{\lambda} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i. \qquad \blacksquare$$

Theorem 2.2: If $\hat{\theta}$ is the maximum likelihood estimator (MLE) of a parameter θ, then $g(\hat{\theta})$ is the MLE of $g(\theta)$ (*invariance property*).

Example 2.2: Let $(X_1 \ldots, X_n)$ be a r.s. of $X \sim$ Uniform$(0, \theta]$. What is the maximum likelihood estimator of $\log \theta$?

The likelihood function of θ, given $x_1, \ldots, x_n$, is

$$\begin{aligned} L(\theta | x_1, \ldots, x_n) &= \prod_{i=1}^{n} \tfrac{1}{\theta} \; I_{(0,\theta]}(x_i) \\ &= \tfrac{1}{\theta^n} \; I_{[x_{(n)}, \infty)}(\theta) \end{aligned}$$



$\Rightarrow$ $X_{(n)} = \max(X_1, \ldots, X_n)$ is the MLE of θ.

$\therefore$ From the MLE invariance property, $\log X_{(n)}$ is the MLE of $\log \theta$. $\blacksquare$

2.3 Newton-Raphson method

Standard techniques for optimization can be used to determine estimates that minimize various functions of the residuals *e.g.* (2.2) or maximize various likelihood functions in (2.5). The algorithms to solve *e.g.* the latter are usually iterative.

If the likelihood $L(\boldsymbol{\theta}|\mathbf{x})$ or $\log L(\boldsymbol{\theta}|\mathbf{x})$ is twice differentiable, one algorithm is Newton's method, in which the minimizing value of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$, is obtained as a limit of the iterates

$$\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)} - [H(\boldsymbol{\theta}^{(k-1)})]^{-1} \, s(\boldsymbol{\theta}^{(k-1)}), \quad k = 1, 2, \ldots, \tag{2.6}$$

where $H(\boldsymbol{\theta})$ and $s(\boldsymbol{\theta})$ denote the Hessian matrix and the gradient (score function) of $L(\boldsymbol{\theta}|\mathbf{x})$ or $\log L(\boldsymbol{\theta}|\mathbf{x})$, both evaluated at $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$, with $\mathbf{x} = (x_1, \ldots, x_n)^T$. (2.6) is also called the *Newton-Raphson method*.

For various computational considerations, instead of the exact Hessian H, a matrix $\tilde{H}$ approximating the Hessian is often used. In this case the technique is called a *quasi-Newton method*.

A common quasi-Newton method for optimizing $L_{\boldsymbol{\theta}} \equiv \log L(\boldsymbol{\theta}|\mathbf{x})$ is *Fisher scoring method*, in which the Hessian in Newton's method is replaced by its expected value. The expected value can be replaced by an estimate, such as the sample mean. The iterates then are

$$\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)} - [\widetilde{E}(\boldsymbol{\theta}^{(k-1)})]^{-1} \, s(\boldsymbol{\theta}^{(k-1)}), \tag{2.7}$$

where $\widetilde{E}(\boldsymbol{\theta})$ is an estimate or an approximation of $E(H(\boldsymbol{\theta})|\mathbf{X})$.[6]

Under suitable regularity conditions:

1. If θ is a scalar, the square of the first derivative of $L_{\boldsymbol{\theta}}$ is the negative of the second derivative or in general
$$s(\boldsymbol{\theta}) \, s(\boldsymbol{\theta})^T \equiv \left[\frac{\partial}{\partial \theta_i} L_{\boldsymbol{\theta}}\right] \left[\frac{\partial}{\partial \theta_i} L_{\boldsymbol{\theta}}\right]^T = -H(\boldsymbol{\theta}) \equiv -\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} L_{\boldsymbol{\theta}}\right].$$

[6]Note that (2.6)-(2.7) came from the tangent line equation of $y = f(x)$ at $x = x_n$, $y = f'(x_n)(x - x_n) + f(x_n)$ when $(x, y) = (x_{n+1}, 0)$, *i.e.* $x_{n+1} = x_n - f(x_n)/f'(x_n)$.

2. The MLE $\widehat{\boldsymbol{\theta}}$ is a consistent estimator and is asymptotically normal with mean $\boldsymbol{\theta}_\star$ and variance-covariance matrix
$$[E(-H(\boldsymbol{\theta}_\star)|\mathbf{X}]^{-1} = [\mathcal{I}(\boldsymbol{\theta}_\star)]^{-1},$$
which is the inverse of the Fisher information matrix. For further details, see *e.g.* Gentle (2002).

Example 2.3: If $\mathbf{X} = (X_1, X_2, X_3, X_4) \sim \text{Multinomial}(N, p_1, p_2, p_3, p_4)$ such that $p_1 = \frac{1}{2}+\frac{1}{4}\theta$, $p_2 = p_3 = \frac{1}{4}-\frac{1}{4}\theta$, $p_4 = \frac{1}{4}\theta$, where $0 \le \theta \le 1$ (Gentle, 2002), find the maximum likelihood estimate of θ using the Fisher scoring method for $N{=}197$, $x_1{=}125$, $x_2{=}18$, $x_3{=}20$, $x_4{=}34$ and $\theta^{(0)} = 0.5$.

Given an observation (x_1, x_2, x_3, x_4), one has to

$$\begin{aligned}
&L_\theta = x_1 \log(2+\theta) + (x_2+x_3)\log(1-\theta) + x_4 \log(\theta) + c\\
&s(\theta) = dL_\theta/d\theta = x_1/(2+\theta) - (x_2+x_3)/(1-\theta) + x_4/\theta\\
&\mathcal{I}(\theta) \equiv -E(d^2 L_\theta/d\theta^2|\mathbf{X}) = (N/4)(1/(2+\theta) + 2/(1-\theta) + 1/\theta)
\end{aligned}$$

Therefore, the implementation of Fisher's scoring method

$$\theta^{(k)} = \theta^{(k-1)} + [\mathcal{I}(\theta^{(k-1)})]^{-1}\, s(\theta^{(k-1)}), \quad k = 1, 2, \ldots$$

results in the maximum likelihood estimate of θ of 0.62682. ■

2.4 EM algorithm

Other optimization methods that are also applied directly to the likelihood or to the posterior density are here presented as data augmentation algorithms, including the *EM algorithm.*

All of these data augmentation algorithms share a common approach to problems: rather than performing a complicated maximization or simulation, one augments the observed data with latent data, *e.g.* missing data, that simplifies the calculation and subsequently performs a series of simple maximizations or simulations (see *e.g.* Tanner, 1996).

The EM method arises from an entirely different approach alternates between updating the parameter $\boldsymbol{\theta}^{(k)}$ by using the alternating steps involve an *expectation* and a *maximization.* The method was described by Dempster *et al.* (1977).

Let $X = (U, V)$ be a random vector that consists of two components, one observed U and one unobserved V, indexed by the parameter $\boldsymbol{\theta}$.

Denote $L_{\boldsymbol{\theta}} = \log L_c(\boldsymbol{\theta}|u, v)$ as the log-likelihood for the complete sample. The likelihood for the observed U is $L(\boldsymbol{\theta}|u) = \int L_c(\boldsymbol{\theta}|u, v)\, dv$.

The *EM algorithm* for maximizing $L(\boldsymbol{\theta}|u)$ has two steps that begin with a value $\boldsymbol{\theta}^{(0)}$. The steps are iterated until convergence.

- **E step:** for $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$ and $\boldsymbol{\theta}' = \boldsymbol{\theta}^{(k-1)}$, compute
$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}') = E_{V|u,\boldsymbol{\theta}'}(L_{\boldsymbol{\theta}}) = \int \log L_c(\boldsymbol{\theta}|u, v) p(v|\boldsymbol{\theta}', u)\, dv,$$
- **M step:** determine $\boldsymbol{\theta}^{(k)}$ so as to maximize $Q(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^{(k-1)})$, subject to any constraints on acceptable values of $\boldsymbol{\theta}$.

The sequence $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots$ converges to a local maximum of the observed-data likelihood $L(\boldsymbol{\theta}|u)$ under fairly general conditions.

Example 2.4: (Gentle, 2002) Consider an experiment with light bulbs such that their lifetime follow an exponential distribution with mean θ. In order to estimate θ, i) n bulbs were tested and their failure times were recorded as $u_1, \ldots, u_n$; ii) other m bulbs were also tested, but their failure times $v_1, \ldots, v_m$ were not recorded (*missing data*); only the number r of bulbs that had failed at time t was recorded.

For determining the kth E step, notice that

$$L_\theta \equiv \log L_c(\theta|u, v) = -n\,(\log\theta + \bar{u}/\theta) - m\,\log\theta - \textstyle\sum_{i=1}^{m} v_i/\theta,$$

and its expected value, $E_{V|u,\theta'}(L_\theta)$, is

$$Q(\theta, \theta') = -(n+m)\log\theta - (1/\theta)[n\bar{u} + (m-r)\,(t+\theta') + r\,(\theta' - t\,h(t,\theta'))],$$

where $h(t, \theta') = \exp(-t/\theta')/[1 - \exp(-t/\theta')]$ and $\theta' = \theta^{(k-1)}$.

The kth M step determine the maximum that occurs at

$$\theta^{(k)} = (\frac{1}{n+m})[n\bar{u} + (m-r)\,(t+\theta^{(k-1)}) + r\,(\theta^{(k-1)} - t\,h(t, \theta^{(k-1)}))]. \qquad \blacksquare$$

Comments:

1. Specific techniques were used for the computations in the two steps, it is not necessary for the EM method to use those same inner-loop algorithms (Gentle, 2002).

2. For the E step there are not so many choices. With some luck, the expectation can be computed in closed form. Otherwise, computing the expectation is a numerical quadrature problem.

3. For the maximization step there are more choices, *e.g.* Dempster *et al.* (1977) suggested requiring only an increase in the expected value; that is, take $\boldsymbol{\theta}^{(k)}$ so that $Q(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^{(k-1)}) \geq Q(\boldsymbol{\theta}^{(k-1)}, \boldsymbol{\theta}^{(k-2)})$. This is called a *generalized EM algorithm* (GEM).

4. Dempster *et al.* (1977) show that the EM algorithm converges at a linear rate, with the rate depending on the proportion of information about $\boldsymbol{\theta}$ in $L(\boldsymbol{\theta}|X)$ that is observed. This can imply quite slow convergence if a large portion of the data are missing.

2.5 Data augmentation algorithm

Unlike the EM algorithm, the *data argumentation algorithm* aims to obtain the full likelihood function or the posterior distribution rather than just its maximizer and the curvature at the maximizer. Because that, the data augmentation algorithm will provide a way of improving the inference in a small samples, whereas in large samples those two functions are consistent with the Normal approximation (Tanner, 1996).

1. In the data augmentation algorithm, data X are also augmented with latent data V, *i.e.* $X = (U, V)$ where U represents the observed data.

2. To obtain the posterior $h(\theta|u)$, one generates multiple values (imputations) of V from the predictive distribution $p(v|\theta, u)$ and then computes the average of $h(\theta|u, v)$ over the imputations.

The data augmentation algorithm is motivated by the following relations:

1. The posterior relation:
$$h(\theta|u) = \int_V h(\theta|u,v)\, p(v|u)\, dv,$$
2. The predictive relation:
$$p(v|u) = \int_\Theta p(v|\theta',u)\, h(\theta'|u)\, d\theta',$$
where $h(\theta|u,v)$ is the conditional distribution of θ given X *i.e.* the augmented posterior, and $p(v|\theta',u)$ is the conditional predictive distribution.

Substituting the predictive relation into the posterior relation, one has

$$\begin{aligned} g(\theta) \equiv h(\theta|u) &= \int_\Theta \left[\int_V h(\theta|u,v)\, p(v|\theta',u)\, dv\right] h(\theta'|u)\, d\theta' \\ &\equiv \int_\Theta K(\theta,\theta') g(\theta')\, d\theta'. \end{aligned} \tag{2.8}$$

In order to solve equation (2.8), one can use the method of successive substitution *i.e.* starting with $g_0(\theta)$, calculate successively

$$g_k(\theta) = Tg_{k-1}(\theta) \equiv \int_\Theta K(\theta,\theta')\, g(\theta')\, d\theta', \quad k = 1, 2, \ldots. \tag{2.9}$$

Tanner (1996) refers to using the Monte Carlo method to perform the integration in (2.8). In particular, applying the Monte Carlo method to the posterior identity yields an iterative scheme, giving rise to the *data augmentation algorithm.*

The data augmentation algorithm consists of iterating between the following two steps:

1. Imputation step: Generate a sample $v_1, \ldots, v_m$ from the current approximation to the predictive distribution $p(v|u)$.
   1.1 Generate θ' from $g_k(\theta)$.
   1.2 Generate v from $p(v|\theta',u)$, where θ' is the value generated in 1.1.
2. Posterior step: Update the current approximation to $h(\theta|u)$ to be the mixture of augmented posteriors of θ, given the augmented data from step 1, *i.e.*
$$g_k(\theta) = \frac{1}{m}\sum_{j=1}^m h(\theta|v^{(j)}, u).$$

When m is large, steps 1 and 2 will provide a close approximation to the iteration of (2.9). For further details, see Tanner (1996).

Example 2.5: (vide Example 2.3) Augment the observed data, *i.e.* $u = (u_1, u_2, u_3, u_4) = (125, 18, 20, 34)$, by splitting the first cell into two cells with probabilities $\frac{1}{2}$ and $\frac{\theta}{4}$. The augmented data are given by $v = (v_1, v_2, v_3, v_4, v_5)$ such that $v_1 + v_2 = u_1$, $v_3 = u_2$, $v_4 = u_3$, $v_5 = u_4$.

Notice that the observed posterior (under a flat prior) is proportional to

$$(2+\theta)^{u_1}(1-\theta)^{u_2+u_3}\theta^{u_4},$$

while the augmented posterior (under a flat prior) is proportional to

$$\theta^{v_2+v_5}(1-\theta)^{v_3+v_4}.$$

That is, augmented posterior $h(\theta|u,v)$ is the Beta(v_2+v_5+1, v_3+v_4+1) distribution. Besides conditional predictive distribution $p(v|\theta,u)$ is indeed the Binomial$(n=125, p=\theta/(2+\theta))$.

Thus, the data augmentation algorithm is given as:

1. Imputation step:

   1.1 Draw θ' from the current estimate of the posterior.

   1.2 Generate v_2 from Binomial$(n{=}125, p{=}\theta'/(2{+}\theta'))$.

   Repeat steps 1.1 and 1.2 m times to yield $v_2^{(1)}, \ldots, v_2^{(m)}$.

2. Posterior step: Set the posterior of θ equal to the mixture of Beta distributions, mixed over the m imputed values of v_2 *i.e.*

$$h(\theta|u) = \frac{1}{m}\sum_{j=1}^{m} Beta(v_2^{(j)}{+}v_5{+}1, v_3{+}v_4{+}1).$$

Steps 1.1, 1.2 and 2 are to be iterated until convergence of the algorithm is achieved. Notice that the Uniform(0,1) prior is assumed in this example. ■

Generalized linear models - I

Generalized Linear Models (GLM), introduced by Nelder and Wedderburn (1972), synthesize the normal linear model that has a linear regression structure and have in common that the response variable belongs to the exponential distribution family.

Particular cases of MLG are: i) Normal linear regression model; ii) Analysis of variance model; iii) Logistic regression model; iv) Log-linear models for contingency tables, etc.

Notation: The data $\{(y_i, \mathbf{x}_i),\ i = 1, \ldots, n\}$, are realizations of the response variable Y and the covariate vector $\mathbf{x}$ in n individuals, being the Y_i components of the random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ independent.

Definition 2.3: The r.v. Y is said to have distribution belonging to the *(dispersion) exponential family* if its p.d.f. or p.m.f. can be written in the form

$$f(y\,|\,\theta, \phi) = \exp\Big\{\frac{y\,\theta - b(\theta)}{a(\phi)} + c(y, \phi)\Big\}, \tag{2.10}$$

where θ and ϕ are scalar parameters, $a(\cdot)$, $b(\cdot)$ and $c(\cdot,\cdot)$ are known real functions. ■

In Definition 2.3, θ is the canonical form of the location parameter and ϕ is a (known) dispersion parameter. In this case the distribution (2.10) is part of the uniparametric exponential family. It is also assumed that the function $b(\cdot)$ is differentiable and that the distribution support does not depend on the parameters. Sometimes $a(\phi) = \phi/w$, where w is a known constant.

Example 2.6: Normal - If $Y \sim N(\mu, \sigma^2)$, p.d.f. of Y is given by

$$f(y|\mu, \sigma^2) = \exp\Big\{\frac{1}{\sigma^2}(y\mu - \frac{\mu^2}{2}) - \frac{1}{2}(\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\Big\}$$

for $y \in I\!R$. It has then been that this function is of type (2.10) with

$$\begin{aligned} &\theta = \mu,\ a(\phi) = \phi = \sigma^2, \\ &c(y, \phi) = -\tfrac{1}{2}(\tfrac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)), \\ &b(\theta) \ = \ \mu^2/2,\ b'(\theta) = \mu,\ b''(\theta) = V(\mu) = 1 \end{aligned}$$

It is known that $E(Y) = \mu$ e $Var(Y) = \sigma^2$. In this case, the variance function is $V(\mu) = 1$, the canonical parameter is the expected value μ and σ^2 is the scatter parameter. ■

Example 2.7: Binomial - If Y is such that $mY \sim Bi(m, \pi)$, your p.m.f. is

$$f(y|\pi) = \binom{m}{ym}\pi^{ym}(1-\pi)^{m-ym} = \exp\Big\{m\big(y\theta - \ln(1+e^{\theta})\big) + \ln\binom{m}{ym}\Big\}$$

with $y \in \{0, \frac{1}{m}, \frac{2}{m}, \ldots, 1\}$ and $\theta = \ln(\pi/(1-\pi))$, being of the form (2.10) with

$$\begin{aligned} &\theta = \ln(\tfrac{\pi}{1-\pi}),\ a(\phi) = \tfrac{\phi}{\omega},\ \phi = 1,\ \omega = m \\ &c(y,\phi) = \ln\tbinom{m}{ym},\ b(\theta) = \ln(1+e^{\theta}), \\ &b'(\theta) = \tfrac{e^{\theta}}{1+e^{\theta}} = \pi,\ b''(\theta) = V(\mu) = \tfrac{e^{\theta}}{(1+e^{\theta})^2} = \pi(1-\pi). \end{aligned}$$

Directly get $E(Y) = \pi$ and $Var(Y) = \pi(1-\pi)/m$.

The canonical parameter is the *logit* function, $\ln(\frac{\pi}{1-\pi})$. ■

Table 1: Some exponential family distributions.

Distribution	Normal	Binomial	Poisson	Gamma
Notation	$N(\mu,\sigma^2)$	$B(m,\pi)/m$	$P(\lambda)$	$Ga(\nu, \frac{\nu}{\mu})$
Support	$(-\infty,+\infty)$	$\{0, \frac{1}{m}, ..., 1\}$	$\{0, 1, ...\}$	$(0,\infty)$
θ	μ	$\ln(\frac{\pi}{1-\pi})$	$\ln\lambda$	$-\frac{1}{\mu}$
$a(\phi)$	σ^2	$\frac{1}{m}$	1	$\frac{1}{\nu}$
ϕ	σ^2	1	1	$\frac{1}{\nu}$
ω	1	m	1	1
$c(y,\phi)$	$-\frac{1}{2}(\frac{y^2}{\phi}+\ln(2\pi\phi))$	$\ln\binom{m}{my}$	$-\ln y!$	$\nu\ln\nu - \ln\Gamma(\nu)$ $+(\nu-1)\ln y$
$b(\theta)$	$\frac{\theta^2}{2}$	$\ln(1+e^{\theta})$	e^{θ}	$-\ln(-\theta)$
$b'(\theta) = E(Y)$	θ	$\pi = \frac{e^{\theta}}{1+e^{\theta}}$	$\lambda = e^{\theta}$	$\mu = -\frac{1}{\theta}$
$b''(\theta) = V(\mu)$	1	$\pi(1-\pi)$	λ	μ^2
$Var(Y)$	σ^2	$\pi(1-\pi)/m$	λ	μ^2/ν

MLGs are an extension of the normal linear model (2.1), $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, which is made in two directions:

- The distribution considered does not have to be Normal and can be any distribution of the *exponential family*;
- Maintaining the linearity structure, the *link function* (any differentiable function) $g(\mu) = \eta$ relates the expected value μ and the linear predictor $\eta = \mathbf{x}^T\boldsymbol{\beta}$.

The choice of *link function* depends on the response type and the intended analysis. When $\theta = \eta = \mathbf{x}^T\boldsymbol{\beta}$, the link function is *canonical*.

Table 2: Some link functions.

identity	reciprocal	inverse quadratic	logarithmic
μ	$1/\mu$	$1/\mu^2$	$\ln(\mu)$
logit	*probit*	square root	complementary log-log
$\ln(\frac{\mu}{1-\mu})$	$\Phi^{-1}(\mu)$	$\sqrt{\mu}$	$\ln[-\ln(1-\mu)]$

In GLM the estimation is made using the methods:

- Maximum likelihood for the parameter of interest $\boldsymbol{\beta}$;
- Moments for the nuisance parameter ϕ, when it exists.

The *likelihood function*, as a function of $\boldsymbol{\beta}$, is given by

$$L(\boldsymbol{\beta}) = \exp\left\{\frac{1}{\phi}\sum_{i=1}^{n}\omega_i(y_i\theta_i - b(\theta_i)) + \sum_{i=1}^{n}c(y_i,\phi,\omega_i)\right\} \quad (2.11)$$

and so the likelihood logarithm is

$$\ln L(\boldsymbol{\beta}) \equiv \ell(\boldsymbol{\beta}) = \textstyle\sum_{i=1}^{n}\ell_i(\boldsymbol{\beta}), \quad (2.12)$$

where $\ell_i(\boldsymbol{\beta}) = \frac{\omega_i}{\phi}(y_i\theta_i - b(\theta_i)) + c(y_i,\phi,\omega_i)$ is the contribution of the observation y_i in likelihood. Assuming certain conditions of regularity (Sen and Singer, 1993), the ML estimators of $\boldsymbol{\beta}$ are obtained as a solution to the likelihood equation system

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n}\frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \beta_j} = 0, \quad j = 1, \ldots, p.$$

To get the previous equations write

$$\frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \beta_j} = \frac{\partial \ell_i(\theta_i)}{\partial \theta_i}\frac{\partial \theta_i(\mu_i)}{\partial \mu_i}\frac{\partial \mu_i(\eta_i)}{\partial \eta_i}\frac{\partial \eta_i(\boldsymbol{\beta})}{\partial \beta_j}$$

being

$$\begin{array}{l} \frac{\partial \ell_i(\theta_i)}{\partial \theta_i} = \frac{\omega_i(y_i - b'(\theta_i))}{\phi} = \frac{\omega_i(y_i-\mu_i)}{\phi}, \\ \frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = \frac{\omega_i Var(Y_i)}{\phi} \quad \text{e} \quad \frac{\partial \eta_i(\boldsymbol{\beta})}{\partial \beta_j} = x_{ij}. \end{array}$$

Thus

$$\frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \beta_j} = \frac{\omega_i(y_i-\mu_i)}{\phi}\frac{\phi}{\omega_i Var(Y_i)}\frac{\partial \mu_i}{\partial \eta_i}x_{ij}$$

and the likelihood equations for $\boldsymbol{\beta}$ are

$$\sum_{i=1}^{n}\frac{(y_i-\mu_i)x_{ij}}{Var(Y_i)}\frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 1, \ldots, p. \quad (2.13)$$

The score function is the p-dimensional vector

$$\mathbf{s}(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\mathbf{s}_i(\boldsymbol{\beta}), \quad (2.14)$$

where $\mathbf{s}_i(\boldsymbol{\beta})$ is the $p \times 1$ vector of components $\frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \beta_j} = \frac{(y_i-\mu_i)x_{ij}}{Var(Y_i)}\frac{\partial \mu_i}{\partial \eta_i}$, $i=1,\ldots,n$, $j=1,\ldots,p$.
Fisher's information matrix, defined by $\mathcal{I}(\boldsymbol{\beta}) = E[-\frac{\partial \mathbf{s}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}]$, whose elements for regular families are

$$\begin{array}{rcl} -\sum_{i=1}^{n}E\left(\frac{\partial^2 \ell_i(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k}\right) & = & \sum_{i=1}^{n}E\left(\frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \beta_j}\frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \beta_k}\right) \\ & = & \sum_{i=1}^{n}E\left[\left(\frac{(Y_i-\mu_i)x_{ij}}{Var(Y_i)}\frac{\partial \mu_i}{\partial \eta_i}\right)\left(\frac{(Y_i-\mu_i)x_{ik}}{Var(Y_i)}\frac{\partial \mu_i}{\partial \eta_i}\right)\right] \\ & = & \sum_{i=1}^{n}\frac{x_{ij}x_{ik}}{Var(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \end{array}$$

In the matrix form one has

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) = \mathbf{X}^T\mathbf{W}\mathbf{X}\ , \quad (2.15)$$

where $\mathbf{W}$ is the diagonal matrix whose i element is

$$\varpi_i = \Big(\frac{\partial\mu_i}{\partial\eta_i}\Big)^2 \frac{1}{Var(Y_i)} = \Big(\frac{\partial\mu_i}{\partial\eta_i}\Big)^2 \frac{\omega_i}{\phi V(\mu_i)}. \tag{2.16}$$

For strictly concave functions $\ell(\boldsymbol{\beta})$ the maximum likelihood estimators are really unique, when they exist.

Problem: Equations (2.13) generally have no analytical solution and, therefore, their resolution implies numerical methods.

The iterative method for solving equations (2.13) is a *weighted least squares* method based on Fisher's method.

Let $\widehat{\boldsymbol{\beta}}^{(0)}$ be an initial estimate for *betab*. Fisher's *scoring* process proceeds with the successive calculation of $\widehat{\boldsymbol{\beta}}$ through the relationship:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \widehat{\boldsymbol{\beta}}^{(k)} + \Big[\boldsymbol{\mathcal{I}}(\widehat{\boldsymbol{\beta}}^{(k)})\Big]^{-1}\mathbf{s}(\widehat{\boldsymbol{\beta}}^{(k)}),$$

where $\mathcal{I}(\cdot)^{-1}$ is the inverse of the Fisher information matrix (2.15) and $S(\cdot)$ is the score function (2.14).

The previous expression can be written as

$$\Big[\boldsymbol{\mathcal{I}}(\widehat{\boldsymbol{\beta}}^{(k)})\Big]\widehat{\boldsymbol{\beta}}^{(k+1)} = \Big[\boldsymbol{\mathcal{I}}(\widehat{\boldsymbol{\beta}}^{(k)})\Big]\widehat{\boldsymbol{\beta}}^{(k)} + \mathbf{s}(\widehat{\boldsymbol{\beta}}^{(k)}).$$

The right side of this equation is a generic element vector

$$\sum_{j=1}^{p}\Big[\sum_{i=1}^{n}\tfrac{x_{ij}x_{il}}{Var(Y_i)}\Big(\tfrac{\partial\mu_i}{\partial\eta_i}\Big)^2\Big]\beta_j^{(k)} + \sum_{i=1}^{n}\tfrac{(y_i-\mu_i)x_{il}}{Var(Y_i)}\tfrac{\partial\mu_i}{\partial\eta_i}$$

and therefore in matrix form one has

$$\boldsymbol{\mathcal{I}}(\widehat{\boldsymbol{\beta}}^{(k)})\widehat{\boldsymbol{\beta}}^{(k+1)} = \mathbf{X}^T\mathbf{W}^{(k)}\mathbf{u}^{(k)},$$

where $\mathbf{u}^{(k)}$ is a vector with generic element

$$u_i^{(k)} = \eta_i^{(k)} + (y_i - \mu_i^{(k)})\frac{\partial\eta_i^{(k)}}{\partial\mu_i^{(k)}}, \tag{2.17}$$

with $\eta_i^{(k)} = \sum_{j=1}^{p} x_{ij}\beta_j^{(k)}$, and the matrix $\mathbf{W}^{(k)}$ represents the matrix $\mathbf{W}$ calculated at $\widehat{\boldsymbol{\mu}}^{(k)}$.

So, given (2.15), we get the final expression for the estimate of $\boldsymbol{\beta}$ na $(k+1)$-th iteration

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \Big(\mathbf{X}^T\mathbf{W}^{(k)}\mathbf{X}\Big)^{-1}\mathbf{X}^T\mathbf{W}^{(k)}\mathbf{u}^{(k)}. \tag{2.18}$$

Recall from the previous statement that the 'proposed algorithm operates through a sequence of *weighted least squares*' problems.

In fact, the equation (2.18) is identical to that which would be obtained for the weighted least squares estimators if the linear regression of the response $\mathbf{u}^{(k)}$ in $\mathbf{X}$ was performed at each step, where $\mathbf{W}^{(k)}$ is a weight matrix. For further details on MLG, see *e.g.* Amaral Turkman and Silva (2000).

In short: Calculating the maximum likelihood estimates for $\boldsymbol{\beta}$ is iteratively done in two steps:

i) Given $\widehat{\boldsymbol{\beta}}^{(k)}$ (with k starting at 0), $\mathbf{u}^{(k)}$ is calculated using (2.17) and $\mathbf{W}^{(k)}$ using (2.16).

ii) The new iterate $\widehat{\boldsymbol{\beta}}^{(k+1)}$ is calculated at (2.18).

Iterations stop when a suitable criterion is achieved, for example when

$$\|\widehat{\boldsymbol{\beta}}^{(k+1)} - \widehat{\boldsymbol{\beta}}^{(k)}\|/\|\widehat{\boldsymbol{\beta}}^{(k)}\| \;\leq\; \epsilon,$$

for any previously set $\epsilon > 0$ value.

In general *convergence* is reached after a few iterations. If the iterative process does not seem to converge, this could be due to a poor initial estimate or the non-existence of MLE within the allowable $\boldsymbol{\beta}$ region.

Method of moments*

In order to estimate an unknown parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) \in \Theta$ that characterizes the distribution $f_X(x|\boldsymbol{\theta})$ of the random variable X, suppose the first k moments of the true distribution (the 'population moments') can be expressed as functions of $\boldsymbol{\theta}$:

$$\mu_1 \equiv E(X) = g_1(\boldsymbol{\theta}),\;\; \mu_2 \equiv E(X^2) = g_2(\boldsymbol{\theta}),\;\; \cdots\;\; \mu_k \equiv E(X^k) = g_k(\boldsymbol{\theta}).$$

Consider the sample values of X, $x_1, \ldots, x_n$ to estimate μ_j by the j-th sample moment *i.e.* $\hat{\mu}_j = (1/n)\sum_{i=1}^{n} x_i^j$, $j = 1, \ldots, k$.

The method of moments estimator for $\theta_1, \ldots, \theta_k$, denoted by $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$ is defined as the solution (if there is one) to the equations:

$$\hat{\mu}_1 = g_1(\widehat{\theta}_1, \ldots, \widehat{\theta}_k),\;\; \hat{\mu}_2 = g_2(\widehat{\theta}_1, \ldots, \widehat{\theta}_k),\;\; \cdots\;\; \hat{\mu}_k = g_k(\widehat{\theta}_1, \ldots, \widehat{\theta}_k).$$

GLM selection*

In the face of many covariates, there is an interest in finding the 'best model', which should be a model that achieves a good balance between the three factors: good fit, parsimony and interpretation.

The most used models during the selection process are: (i) saturated (S) or complete ($\widehat{\mu}_i = y_i$); (ii) null ($E(Y_i) = \mu$); (iii) maximal (more parameters); (iv) minimal (less parameters); (v) current (M).

If we compare the model M with the model S through the likelihood ratio statistics, we obtain

$$D^*(\mathbf{y}; \widehat{\boldsymbol{\mu}}) = -2(\ell_M(\widehat{\boldsymbol{\beta}}_M) - \ell_S(\widehat{\boldsymbol{\beta}}_S)) = \frac{1}{\phi} D(\mathbf{y}; \widehat{\boldsymbol{\mu}}) = \frac{1}{\phi} \sum_{i=1}^{n} d_i, \tag{2.19}$$

where

$$d_i = 2\omega_i \Big\{ y_i(q(y_i) - q(\widehat{\mu}_i)) - b(q(y_i)) + b(q(\widehat{\mu}_i)) \Big\} \tag{2.20}$$

measures the difference in logarithms of the observed and fitted likelihoods for the observation i, $i = 1, \ldots, n$, denoting $q(y_i)$ and $q(\widehat{\mu}_i)$ as the estimates of canonical parameters under the models S and M, respectively.

The measure $D^*(\mathbf{y}; \widehat{\boldsymbol{\mu}})$ defined in (2.19) is called *reduced deviance*, while $D(\mathbf{y}; \widehat{\boldsymbol{\mu}})$ is called *deviance* for the current model, and this is only a function of the data.

Based on (2.20), it is possible to define the residual of the *deviance* corresponding to the i-th observation of the deviance above

$$R_i^D \;=\; \delta_i \sqrt{d_i}, \tag{2.21}$$

where $\delta_i = sinal(y_i - \widehat{\mu}_i)$.

Another important property of *deviance* is additivity for nested models. Suppose we have two intermediate models M_1 and M_2 with M_2 nested in M_1. If $D(\mathbf{y}; \widehat{\boldsymbol{\mu}}_j)$ is used to the deviance of the model M_j, $j = 1, 2$, then the likelihood ratio statistic to compare these two models is summarized in

$$-2(\ell_{M_2}(\widehat{\boldsymbol{\beta}}_2) - \ell_{M_1}(\widehat{\boldsymbol{\beta}}_1)) = [D(\mathbf{y}; \widehat{\boldsymbol{\mu}}_2) - D(\mathbf{y}; \widehat{\boldsymbol{\mu}}_1)]/\phi.$$

Without loss of generality, it is known that, under the hypothesis that the model M_1 is true, then

$$[D(\mathbf{y};\widehat{\boldsymbol{\mu}}_2) - D(\mathbf{y};\widehat{\boldsymbol{\mu}}_1)]/\phi \quad \overset{a}{\sim} \quad \chi^2_{p_1-p_2}, \tag{2.22}$$

where p_j, represents the dimension of the vector $\boldsymbol{\beta}$ for the model M_j, $j = 1, 2$. The comparison of nested models is based on (2.22).

3 Monte Carlo Methods

Using Monte Carlo methods is an appropriate alternative to numerical methods for solving integrals, particularly in multidimensional scenarios.

Monte Carlo methods are based on stochastic simulation, *i.e.* , on the reproduction of probability distribution values. (vide *e.g.* Gentle, 2004; Robert & Casella, 2004; Paulino *et al.* , 2018; Rizzo, 2019).

According to Robert and Casella (2004), two major classes of numerical problems that arise in statistical inference are:

- Optimization - generally associated with the likelihood approach;
- Integration - generally associated with the Bayesian approach.

Monte Carlo inference (or Monte Carlo integration) can be formulated as estimation of a definite integral

$$\mathcal{J} = \int_{\mathcal{X}} f(x)dx, \tag{3.1}$$

where where $\mathcal{X}$ is uni- or multidimensional, and f is a function that satisfies certain optimality conditions.

If the function f is decomposed so as to have a factor that is a probability density function, *i.e.* , $f(x) = g(x)\,h(x)$, then the integral $\mathcal{J}$ is the expectation of the function g of the random variable with probability density h, that is,

$$\mathcal{J} = E[g(X)] = \int_{\mathcal{X}} g(x)\,h(x)dx. \tag{3.2}$$

3.1 Simple Monte Carlo method

With a random sample $x_1, \ldots, x_n$ from the distribution with probability density h, the integral $\mathcal{J}$ is approximated by the empirical average

$$\hat{\mathcal{J}} \equiv \overline{g_n} = \frac{1}{n}\sum_{i=1}^{n} g(x_i) \tag{3.3}$$

which converges

$$\overline{g_n} \to E_h[g(X)]$$

by the Strong Law of Large Numbers.[7]

Note: The simple Monte Carlo (MC) estimator of $\int_0^1 g(x)dx$ is $\overline{g_n}$.

One uses this technique in many settings in Statistics (Gentle, 2004). There are three steps:

1. decompose the function of interest to include a probability density function as a factor;
2. identify an expected value;
3. use a sample (simulated or otherwise) to estimate the expected value.

Example 3.1: Let $g(x) = \sqrt{1-x^2}$ and let $h(x)$ be the density of the uniform distribution on $(0, 1)$. In order to calculate one-quarter of the area of the unit circle ($\pi/4 \approx 0.7854$), one can generate 5000 sample of $h(x)$ and gets $\overline{g_n} \approx 0.7851$. ■

[7] According to the law of large numbers, the average of the results obtained from a large number of trials should be close to the expected value (see Appendix).

For Bayesian inference, the problem is to approximate an integral of the form

$$\int g(\theta)h(\theta|x)d\theta = E[g(\theta)|x], \tag{3.4}$$

where θ and x can be vectors whose existence is admitted.

Many posterior quantities of interest are expressible by (3.4) for some kind of integrable $g(\theta)$ function, *e.g.*, posterior covariances of components of θ, where $g(\theta) = [\theta_i - E(\theta_i \mid x)]\,[\theta_j - E(\theta_j \mid x)]\,, \forall\, i, j$.

If you can simulate a random sample $(\theta_{(1)}, \ldots, \theta_{(n)})$ of posterior density $h(\theta \mid x)$, the simple Monte Carlo method approximates the integral (3.4) by the empirical average

$$\widehat{E}\,[g(\theta) \mid x] = \frac{1}{n}\sum_{i=1}^{n} g(\theta_{(i)}). \tag{3.5}$$

An estimate of the variance of the estimator (3.3) is

$$v_n = \frac{1}{n-1}\sum_{i=1}^{n}[g(x_i) - \overline{g_n}]^2. \tag{3.6}$$

This is because the elements of the set of random variables $\{g(X_i)\}$, on which we have observations $\{g(x_i)\}$, are (assumed to be) independent, and thus to have zero correlations.

For n large,

$$\frac{\overline{g_n} - E_h[g(X)]}{\sqrt{v_n}} \overset{a}{\sim} N(0,1). \tag{3.7}$$

Note: This can lead to the construction of a convergence test and of confidence bounds on the approximation of $E_h[g(X)]$.

In Bayesian scenario, the precision of the MC estimator (3.5) can also be measured by the Monte Carlo standard error (estimated) given by

$$\frac{1}{\sqrt{n(n-1)}}\left\{\sum_{i=1}^{n}\left[g(\theta_{(i)}) - \frac{1}{n}\sum_{i=1}^{n} g(\theta_{(i)})\right]^2\right\}^{1/2}, \tag{3.8}$$

when the quantity $E\{[g(\theta)]^2|x\}$ is finite.

In short, if you can simulate samples from the posterior distribution $h(\theta \mid x)$, applying the simple Monte Carlo method to solve for (3.4)-type integrals is then trivial, for instance:

1. the evaluation of posterior probabilities,
2. marginal posterior densities,
3. credible intervals,
4. quantities associated with the posterior predictive distribution.

Posterior probabilities

When $g(\theta)$ is the indicator function of some subset A of parametric space, the Monte Carlo approximation (3.5) represents the proportion of sample values included in A. For example, calculating the posterior probability of the smallest HPD interval containing a fixed value $\theta_0 \in \mathbb{R}$,

$$P(\theta_0) = P_{h(\theta|x)}\left(\{\theta : h(\theta \mid x) \geq h(\theta_0 \mid x)\}\right),$$

The Monte Carlo estimate can be expressed by

$$\hat{P}(\theta_0) = \frac{\#\left\{\theta_{(i)},\, 1 \le i \le n : L(\theta_{(i)} \mid x)h(\theta_{(i)}) \ge L(\theta_0 \mid x)h(\theta_0)\right\}}{n}. \tag{3.9}$$

Note that if the univariate density normalizing constant $h(\theta|x)$ is unknown, this does not preclude its determination.

Marginal posterior densities

If $\theta = (\theta_1, \ldots, \theta_k) \in I\!R^k$, $k > 1$ and the objective is to evaluate marginal posterior densities based on a random sample $\theta_{(i)} = (\theta_{(i)1}, \ldots, \theta_{(i)k})$, $1 \le i \le n$ of $h(\theta \mid x)$, several methods can be applied (see Paulino *et al.*, 2018).

For simplicity, assume $k = 2$ and let Θ be the support of the posterior density of $\theta = (\theta_1, \theta_2)$, $h(\theta_1, \theta_2 \mid x)$. Denote $\Theta_{-1}(\theta_1)$ the subset of Θ that represents the support of $h(\theta_1, \theta_2 \mid x)$ for θ_1 fixed, *i.e.*, $\Theta_{-1}(\theta_1) = \{\theta_2 : (\theta_1, \theta_2) \in \Theta\}$. In a coherent notation denote the support of conditional density $h(\theta_1 \mid \theta_2, x)$ by $\Theta_1(\theta_2) = \{\theta_1 : (\theta_1, \theta_2) \in \Theta\}$.

This reasoning will later be generalized to partitioning $\theta = \left(\theta^{(m)}, \theta^{(-m)}\right)$, with $\theta^{(m)} = (\theta_1, \ldots, \theta_m) \in I\!R^m$ for $m = 1, \ldots, k-1$ fixed, and $\theta^{(-m)} = (\theta_{m+1}, \ldots, \theta_k)$.

Fixed a value θ_{1*} of θ_1, one has (assuming the validity of Fubini's theorem)

$$\begin{aligned}
h(\theta_{1*} \mid x) &= \int_{\Theta_{-1}(\theta_{1*})} h(\theta_{1*} \mid \theta_2, x)\; h(\theta_2 \mid x)\, d\theta_2 \\
&= \int_{\Theta_{-1}(\theta_{1*})} h(\theta_{1*} \mid \theta_2, x) \int_{\Theta_1(\theta_2)} h(\theta_1, \theta_2 \mid x)\, d\theta_1 d\theta_2 \\
&= \int_{\Theta} h(\theta_{1*} \mid \theta_2, x)\; h(\theta \mid x)\, d\theta.
\end{aligned}$$

This expression implies that the marginal posterior density of the fixed vector $\theta^{(m)}$ can be approximate by Monte Carlo method applied to the random sample of $h(\theta \mid x)$, $\theta_{(i)} = (\theta_{(i)}^{(m)}, \theta_{(i)}^{(-m)})$ with $\theta_{(i)}^{(m)} = \left(\theta_{(i)1}, \ldots, \theta_{(i)m}\right)$ and $\theta_{(i)}^{(-m)} = \left(\theta_{(i)m+1}, \ldots, \theta_{(i)k}\right)$, $i = 1, \ldots, n$, by

$$\hat{h}\left(\theta_*^{(m)} \mid x\right) = \frac{1}{n}\sum_{i=1}^{n} h\left(\theta_*^{(m)} \mid \theta_{(i)}^{(-m)}, x\right). \tag{3.10}$$

Credible intervals

Consider now that $(\theta_{(i)},\ 1 \le i \le n)$ is a random sample of the univariate $h(\theta \mid x)$ posterior density, with distribution function $H(\theta \mid x)$, which is intended to be summarized by a γ credible interval.

The exact determination of this requires a complete knowledge of the posterior distribution, which is not always the case because of the normalizing constant (although this does not make it impossible to obtain a sample of it, as will be seen below).

The equal tail $100\gamma\%$ credible interval for θ is defined as $R_c(\gamma) = \left(\theta_{\frac{1-\gamma}{2}}, \theta_{\frac{1+\gamma}{2}}\right)$, whose extremes define the posterior probability quantiles $\frac{1-\gamma}{2}$ e $\frac{1+\gamma}{2}$, respectively, of θ, *i.e.*, $H\left(\theta_\beta \mid x\right) = \beta$.

A Monte Carlo approximation of $R_c(\gamma)$ is obtained by ordering the random sample and using the empirical quantiles. Specifically, now representing $(\theta_{[i]}, 1 \le i \le n)$ the ordered sample, the Monte Carlo estimate of $R_c(\gamma)$ is defined by

$$\hat{R}_c(\gamma) = \left(\theta_{\left\langle n\left(\frac{1-\gamma}{2}\right)\right\rangle}, \theta_{\left\langle n\left(\frac{1+\gamma}{2}\right)\right\rangle}\right), \tag{3.11}$$

where $\langle n\alpha \rangle$ denotes the integer part of $n\alpha$.

As noted earlier, the interval $R_c(\gamma)$ of $h(\theta \mid x)$ is not the best interval summary of a unimodal distribution when it is not symmetrical, so it is clearly deprecated in favor of the interval. HPD interval $R_0(\gamma) = \{\theta : h(\theta \mid x) \geq k_\gamma\}$, where k_γ is the largest constant for which the posterior probability of $R_0(\gamma)$ is at least γ. $R_0(\gamma)$ is harder to determine than $R_c(\gamma)$.

Predictive quantities

Since the posterior predicted density ordinances of Y are the expected value $p(y \mid x) = E_{\theta|x}\left[f(y \mid \theta, x)\right]$, easily get the respective Monte Carlo approximation

$$\hat{p}(y \mid x) = \frac{1}{n}\sum_{i=1}^{n} f(y \mid \theta_{(i)}, x) \tag{3.12}$$

based on i.i.d. simulated values of $h(\theta \mid x)$.

For Monte Carlo estimation of quantities associated with the predictive distribution $p(y \mid x)$, a random sample of this distribution must be obtained. This is possible through the so-called composition method (Tanner, 1996) if it is possible to sample from the sample distribution of y, obtaining then the sample $(y_{(1)}, \ldots, y_{(n)})$ from $p(y \mid x)$ as follows:

1. Take a realization $(\theta_{(1)}, \ldots, \theta_{(n)})$ from a random sample of the distribution $h(\theta \mid x)$;
2. For each i, remove $y_{(i)}$ of $f(y \mid \theta_{(i)}, x)$, $i = 1, \ldots, n$.

Based on this sample approximations of various summaries of the predictive distribution can easily be calculated. For example, estimates of the mean prediction and the HPD prediction interval for the future observation $y \in I\!R$ are derived from it in the same way as the posterior mean and the HPD credible interval for θ are estimated from the posterior sample distribution of θ, as indicated above.

Example 3.2: Consider the hierarchical Normal/Normal model defined by

$$X \mid \theta \sim N_p(\theta, \sigma_1^2 I_p), \quad \theta_i,\ i = 1, \ldots, p \mid \mu, \sigma_2^2 \underset{iid}{\sim} N(\mu, \sigma_2^2),$$

where $\theta = (\theta_i, i = 1, \ldots, p)$ and σ_1^2 is assumed to be known.

Paulino et al (2018) - Example 8.7 - describe:

- the conditional posterior distribution of θ give (μ, σ_2^2) (p-variate Normal),
- the conditional posterior distribution of μ given σ_2^2 (univariate Normal),
- the kernel, $\bar{h}_2(\sigma_2^2 \mid x)$, of the marginal posterior distribution of σ_2^2.

Since the posterior marginal distribution of θ has no explicit form, its estimation by the Monte Carlo method requires at least the simulation of the distribution $h_2(\sigma_2^2 \mid x)$, only partially known. Using as prior distribution of σ_2^2 the "Uniform distribution" on $I\!R_+$, the kernel $\bar{h}_2(\sigma_2^2 \mid x)$ can be evaluated in a grid of N values $\sigma^2_{2(l)}$, uniformly spaced, covering the effective range of σ_2^2 (which should be determined approximately by trial).

Normalization is achieved by summing all values and attaching to each one the weight $p_l = \bar{h}_2\left(\sigma^2_{2(l)}\right) / \sum_{l=1}^{N} \bar{h}_2\left(\sigma^2_{2(l)}\right)$.

The distribuition $h_2(\sigma_2^2 \mid x)$ is now represented by the discrete approximation $\left\{\sigma^2_{2(l)}, p_l\right\}$, which can be generated from simulated values $\sigma^2_{2(j)}$, $j = 1, \ldots, n$ (*vide, e.g.* , Ripley, 1987). Then, we easily simulate

$\mu_{(j)}$ of the distribution $h_2\left(\mu \mid \sigma^2_{2(j)}, x\right)$ and, if necessary, $\theta_{(j)} = \left(\theta_{(j)1}, \ldots, \theta_{(j)p}\right)$ of $h_1\left(\theta \mid \mu_{(j)}, \sigma^2_{2(j)}, x\right)$, $j = 1, \ldots, n$.

The posterior mean of θ is then estimated by the empirical average of $\theta_{(j)}$. If you want to evaluate the marginal posterior density of θ, you can use the conditional estimate.

$$\hat{h}(\theta \mid x) = \frac{1}{n}\sum_{j=1}^{n} h\left(\theta \mid \mu_{(j)}, \sigma^2_{2(j)}, x\right).$$

Wanting to predict a future observation y with sample distribution $Y \mid \theta_q \sim N(\theta_q, \sigma_1^2)$ for a given q, $q = 1, \ldots, p$, simulate $y_{(j)}$ from $f(y \mid \theta_{(j)q})$, $j = 1, \ldots, n$ - remember that σ_1 was assumed known -, and take the empirical average of the simulated values. Obtaining the Monte Carlo approximation of the y predictive distribution can avoid this additional simulation scheme by resorting to

$$\hat{p}(y \mid x) = \frac{1}{n}\sum_{j=1}^{n} f(y \mid \theta_{(j)q}). \qquad \blacksquare$$

Example 3.3: (Bayesian decision theory)

Bayes estimators are not always posterior expectations, but rather solutions of the minimization problem

$$\min_{\delta} \int_{\Theta} \mathcal{L}(\theta, \delta) f(x|\theta) h(\theta) d\theta. \tag{3.13}$$

- Quadratic loss: for $\mathcal{L}(\theta, \delta) = (\theta - \delta)^2$, the Bayes estimator is the posterior mean,
- Absolute loss: for $\mathcal{L}(\theta, \delta) = |\theta - \delta|$, the Bayes estimator is the posterior median,
- With no loss function: use the maximum posteriori estimator

$$\arg\max_{\theta} L(\theta|x)h(\theta). \qquad \blacksquare$$

Variance reduction

Monte Carlo estimators associated with the various representations have variable precision, with implications for the computational effort required to obtain reliable estimates.

Increasing the number of replicates n clearly reduces the variance of the Monte Carlo estimator. To reduce the standard error from 0.01 to 0.0001, one would need approximately 10000 times the number of replicates.

There are several approaches to *reducing the variance* in the sample mean estimator of $\mathcal{J} = E[g(X)]$: Antithetic variables, Control variates, Importance sampling, Stratified sampling, etc. (see *e.g.* Robert and Casella, 2004).

Definition 3.1: If $\hat{\mathcal{J}}_1$ and $\hat{\mathcal{J}}_2$ are MC estimators of the parameter $\mathcal{J} = E[g(X)]$, and $Var(\hat{\mathcal{J}}_2) < Var(\hat{\mathcal{J}}_1)$, then the percent reduction in variance achieved by using $\hat{\mathcal{J}}_2$ instead of $\hat{\mathcal{J}}_1$ is

$$100\left(\frac{Var(\hat{\mathcal{J}}_1) - Var(\hat{\mathcal{J}}_2)}{Var(\hat{\mathcal{J}}_1)}\right). \tag{3.14}$$

If σ^2 is the variance of the estimators $\hat{\mathcal{J}}_1$ and $\hat{\mathcal{J}}_2$ in Definition 3.1, the variance of the average of these estimators of $\mathcal{J}$ is given by

$$Var\left[(\hat{\mathcal{J}}_1 + \hat{\mathcal{J}}_2)/2\right] = \sigma^2/2 + Cov(\hat{\mathcal{J}}_1, \hat{\mathcal{J}}_2)/2.$$

Therefore, the variance of $Var[(\hat{\mathcal{J}}_1 + \hat{\mathcal{J}}_2)/2]$ is smaller if $\hat{\mathcal{J}}_1$ and $\hat{\mathcal{J}}_2$ are negatively correlated than when the variables are independent.

Antithetic variables

Let $X_1,\ldots,X_n$ be a random sample from the distribution of X, simulated via the inverse transform method. That is, we have generated $U_j \sim \text{Uniform}(0,1)$, and computed $X_j = F_X^{-1}(U_j)$, $j = 1,\ldots,n$.

Note that if U is uniformly distributed on $(0,1)$ then $1-U$ has the same distribution as U, but U and $1-U$ are negatively correlated.

Then, for $g(\cdot)$ a function of $X_1,\ldots,X_n$, $Y_1 = g(F_X^{-1}(U_1),\ldots,F_X^{-1}(U_n))$ and $Y_2 = g(F_X^{-1}(1-U_1),\ldots,F_X^{-1}(1-U_n))$ have the same distribution.

Proposition 3.1: If $g = g(X_1,\ldots,X_n)$ is monotone, then $Y_1 = g(F_X^{-1}(U_1),\ldots,F_X^{-1}(U_n))$ and $Y_2 = g(F_X^{-1}(1-U_1),\ldots,F_X^{-1}(1-U_n))$ are negatively correlated. Demonstration: See Corollary 6.1 in Rizzo (2019).

Example 3.4: (Antithetic variables) In order to estimate the standard normal cdf $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} dt$, we can find the approximate reduction in standard error by using antithetic variables.

As $\Phi(x) = \frac{1}{2} + \int_0^x \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} dt = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^1 x\, \mathrm{e}^{-(x\,u)^2/2} du$, if $x > 0$, the quantity of interest is $\mathcal{J} = E_U[x\, e^{-(xU)^2/2}] = \int_0^1 x\, \mathrm{e}^{-(x\,u)^2/2} du$, where $U \sim \text{Uniform}(0,1)$.

Being $g(u) = xe^{-(ux)^2/2}$ a monotone function, one can generate random numbers $u_1,\ldots,u_{n/2}$ from Uniform(0,1) and compute half of the replicates using $Y_j = g(u_j) = x\, e^{-(u_j x)^2/2}$, $j = 1,\ldots,\frac{n}{2}$ as before, but compute the remaining half of the replicates using $Y_j = x\, e^{-((1-u_j)x)^2/2}$, $j = 1,\ldots,\frac{n}{2}$.

The sample mean

$$\hat{\mathcal{J}} = \overline{g_n(x)} = \frac{1}{n} \sum_{j=1}^{n/2} \left(xe^{-(u_j x)^2/2} + xe^{-((1-u_j)x)^2/2} \right)$$

converges to $E(\hat{\mathcal{J}}) = \mathcal{J}$ as $n \to \infty$. ■

3.2 Monte Carlo method with importance sampling

For estimating an integral $\mathcal{J} = \int_a^b g(x)dx$, we can easily rewrite it as $\mathcal{J} = (b-a) \int_a^b g(x) \frac{1}{b-a} dx$, where an uniform weight function is applied over the interval (a,b).

The simple Monte Carlo method generates replicates of $X_1,\ldots,X_n$ uniformly distributed on (a,b) and estimates $\mathcal{J}$ by the sample mean

$$\hat{\mathcal{J}} = \frac{b-a}{n} \sum_{j=1}^{n} g(X_j).$$

However, one can consider other weight functions than Uniform. That is, other strategy for calculating an expected value $\mathcal{J}$ is to generate samples from the distribution with density $p(x)$ that is called *important function.*

If X is a random variable with density $p(x)$, such that $p(x) > 0$ on the set $\{x : g(x) > 0\}$, and Y the random variable $g(X)/p(X)$, then

$$\int g(x)dx = \int \frac{g(x)}{p(x)} p(x)dx = E(Y).$$

$E(Y)$ is estimated by simple Monte Carlo, *i.e.*, compute the average

$$\frac{1}{n} \sum_{j=1}^{n} Y_j = \frac{1}{n} \sum_{j=1}^{n} \frac{g(X_j)}{p(X_j)},$$

where the random variables $X_1,\ldots,X_n$ are generated from the distribution with density $p(x)$.

In an *importance sampling method*, the variance (standard error) of the estimator based on $Y = \frac{g(X)}{p(X)}$ is $\frac{Var(Y)}{n}$, and $Var(Y)$ should be small. The variance of Y is small if Y is nearly constant, so the density $p(x)$ should be 'close' to $g(x)$.

In order to get a more precise MC estimate for a given sample size, one can assume the simulated distribution is not uniform. In this case, a weighted average would be better than the unweighted sample mean to correct potential bias. This method is called *importance sampling* (see *e.g.* Robert and Casella, 2004).

Suppose that $h(x)$ is a density supported on a set A. If $p(x) > 0$ on A, then $\mathcal{J} = \int_A g(x)h(x)dx = \int_A g(x)\frac{h(x)}{p(x)}p(x)dx$. And if $p(x)$ is a density on A, then an estimator of $\mathcal{J} = E_p[g(X)h(X)/p(X)]$ is

$$\hat{\mathcal{J}} = \frac{1}{n}\sum_{j=1}^{n} g(X_j)\frac{h(X_j)}{p(X_j)},$$

where $X_1, \ldots, X_n$ is a random sample from density $p(x)$, which is the importance sampling function. There are many densities $p(x)$ that are convenient to simulate. Typically one should choose one such that $p(x) \approx |g(x)|h(x)$ on A.

For the *Bayesian scenario*, it is usually not possible to get an i.i.d. sample directly from the posterior distribution $h(\theta|x)$, so there is a need to find alternative strategies.

Let $p(\theta)$ be a density function whose support (say Θ_p) includes that of

$$h(\theta|x) = cf(x|\theta)h(\theta).$$

The posterior quantity of interest $\mathcal{J} = E(g(\theta))$ may be expressed in order of this distribution $p(\theta)$ as the expected value of the original function g adjusted by the multiplicative factor $h(\theta|x)/p(\theta)$, which is always finite by the condition imposed on the support of the proposed distribution for sampling, $p(\theta)$.

Therefore, the proposal of simulating from $p(\theta)$ instead of $h(\theta|x)$ leads to redefining the amount of interest through $(\Theta_p, g\,h/p, p)$.

On the other hand, this new representation of the amount of interest only requires that the posterior distribution be known less than the proportionality constant c, and this observation also applies to the instrumental distribution $p(\theta)$. Indeed,

$$\int g(\theta)h(\theta|x)d\theta = \frac{\int g(\theta)\frac{f(x|\theta)h(\theta)}{p(\theta)}p(\theta)d\theta}{\int \frac{f(x|\theta)h(\theta)}{p(\theta)}p(\theta)d\theta} \equiv \frac{\int g(\theta)w(\theta)p(\theta)d\theta}{\int w(\theta)p(\theta)d\theta}. \tag{3.15}$$

Since $(\theta_{(1)}, \ldots, \theta_{(n)})$ is then a sample of $p(\theta)$, you can apply the Monte Carlo method to approximate $\mathcal{J} = E\left[g(\theta) \mid x\right]$ per

$$\hat{\mathcal{J}} = \widehat{E}\left[g(\theta) \mid x\right] = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i g(\theta_{(i)}), \tag{3.16}$$

where $w_i = f(x \mid \theta_{(i)})h(\theta_{(i)})/p(\theta_{(i)})$.

If the support of $p(\theta)$ includes $h(\theta|x)$ and the integral $\int g(\theta)\, h(\theta|x)d\theta$ exists and is finite, Geweke (1989) shows, when $\theta_{(i)}$ are an i.i.d. sample from $p(\theta)$, which

$$\frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i g(\theta_{(i)}) \quad \rightarrow \quad \int g(\theta)\, h(\theta|x)d\theta \qquad a.s.,$$

with a Monte Carlo standard error estimated by

$$\frac{1}{\sum_{j=1}^{n} w_j}\left[\sum_{i=1}^{n}\left\{g(\theta_{(i)}) - \frac{1}{\sum_{j=1}^{n} w_j}\sum_{i=1}^{n} w_i g(\theta_{(i)})\right\}^2 w_i^2\right]^{1/2},$$

under the thinning variance of the Monte Carlo estimator, *i.e.* , the posterior expected value of the product of $[g(\theta)^2]$ by the importance ratio $h(\theta|x)/p(\theta)$ (which is the expected value according to p from the square of $g(\theta)h(\theta|x)/p(\theta)$).

Example 3.5: (vide Example 2.3) Based on a function of importance for obtaining the posterior distribution, one can calculate the posterior expected value and variance for θ. Considering the prior distribution Beta(a, b), the corresponding posterior density function will be

$$h(\theta|x) \propto (2+\theta)^{x_1}(1-\theta)^{x_2+x_3+b-1}\theta^{x_4+a-1}, \quad 0 \le \theta \le 1,$$

where for $L(\theta) \equiv \log h(\theta|x)$ one has

$$\begin{aligned}
L(\theta) &\propto x_1\log(2+\theta) + (x_2+x_3+b-1)\log(1-\theta) + (x_4+a-1)\log(\theta) \\
L'(\theta) &= \frac{x_1}{2+\theta} - \frac{x_2+x_3+b-1}{1-\theta} + \frac{x_4+a-1}{\theta} \\
-L''(\theta) &= \frac{x_1}{(2+\theta)^2} + \frac{x_2+x_3+b-1}{(1-\theta)^2} + \frac{x_4+a-1}{\theta^2}.
\end{aligned}$$

Although the Normal density function is widely used, as θ varies in the range $[0, 1]$, the Beta density function may also be considered suitable as a candidate for the importance function, $p(\theta)$.

Let $\hat{\theta}$ be the value of θ for which

- $L'(\theta) = 0$, and
- $\hat{\sigma}^2 = \{-L''(\hat{\theta})\}^{-1}$.

Consider these values as the first approximations, respectively, for the expected value and variance of the posterior distribution, based on which the characteristic parameters of the instrumental distribution $p(\theta)$[8].

Once this is completely specified, proceed as follows:

1. Simulate the sample $(\theta_{(1)}, \ldots, \theta_{(n)}) \overset{iid}{\sim} p(\theta)$;
2. Calculate the weights of importance $w_i = \frac{f(x|\theta_{(i)})h(\theta_{(i)})}{p(\theta_{(i)})}$;
3. Find the estimates $\frac{1}{\sum_{i=1}^n w_i}\sum_{i=1}^n w_i g(\theta_{(i)})$, taking:
   - $g(\theta) = \theta$ for the approximate calculation of the posterior distribution mean;
   - $g(\theta) = \theta^2$ for getting an approximation of the posterior distribution variance.

The proportionality constant of $h(\theta|x)$ can still be obtained via the Monte Carlo method at the expense of the importance weights w_i since

$$\int f(x|\theta)\, h(\theta)\, d\theta = E_{p(\theta)}\left[\frac{f(x|\theta)\, h(\theta)}{p(\theta)}\right] \approx \frac{1}{n}\sum_{i=1}^{n} w_i.$$

Figure 1 presents the plots for two samples ($N = 197, 20$) of the exact posterior density obtained using numerical integration, as well as the approximations by the Normal and Beta instrumental densities with parameters estimated by importance sampling. We used $n = 250$ and an Uniform prior distribution.

[8]The instrumental density $p(\theta)$ has been known in the literature as a function of importance, possibly by allowing the simulation process to be unlocked and to better cover the region of importance for the evaluation of the integral in question.

To plot the approximations to $h(\theta|x)$ and compare with the exact distribution, the procedure given above to approximate the proportionality constant of $h(\theta|x)$ for each of the importance functions used, $r = 100$ was repeated, and the average of the r values obtained was approximated.

Note that the importance sampling method provides good approximations for both considered importance functions.

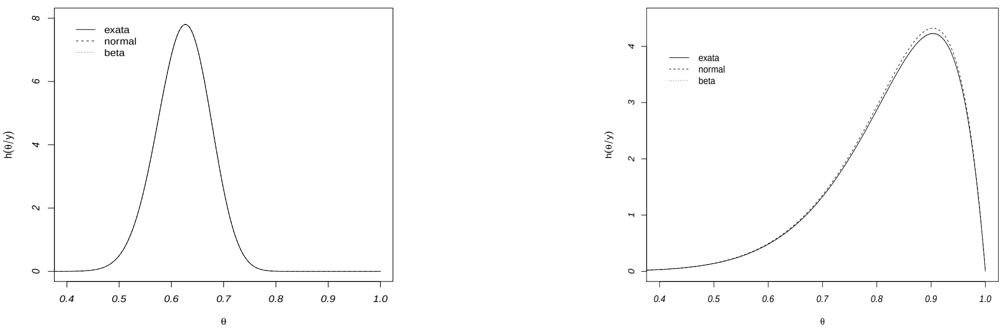


Figure 1: Exact and approximate posterior distributions of θ according to the importance sampling method: $N = 197$ (left), $N = 20$ (right).

Table 3.1 summarizes the information regarding the mean values and variances obtained using the method discussed here.

	$N = 197$		$N = 20$	
	$E(\theta\|x)$	$\mathrm{Var}(\theta\|x)$	$E(\theta\|x)$	$\mathrm{Var}(\theta\|x)$
Normal importance function	0.6233	0.002566	0.8377	0.009263
Beta importance function	0.6227	0.002584	0.8307	0.011741
Exact	0.6228	0.002595	0.8311	0.011651

It is observed that the method using the importance function Beta gives the closest values to the exact values, but the values obtained with both importance functions are similar. ■

3.3 Other stochastic simulation methods

3.3.1 Methods of rejection

Let $\pi(x) = c\pi^*(x)$ be a probability density function (p.d.f.), where c is the normalization constant. Suppose it is difficult to sample directly from π, but nonetheless there is a way to simulate from a p.d.f. $p_u(x)$ based on which you create a function that delimits superiorly π (known as *envelope*), *i.e.* , such that for any x in the π support you have $\pi(x) \leq Mp_u(x)$, where $M > 1$ is a specified constant.

The *basic rejection method* that returns a value x from a distribution $X \sim \pi(x)$ is explained in the following algorithm which is schematically illustrated in Figure 2.

Basic rejection algorithm

1. Generate y from p.d.f. p_u.

2. Generate u from an Uniform distribution in $(0, 1)$.

3. If $u \leq \dfrac{\pi(y)}{Mp_u(y)}$ take $x = y$; if not, it returns to 1.

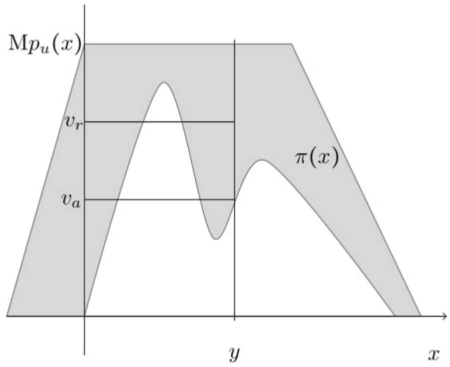


Figure 2: Schematic representation of the basic rejection algorithm where $v = uMp_u(y)$, $r \equiv$ rejected, and $a \equiv$ accepted.

In fact, taking X as the Y variable when it is accepted gives $\forall v$

$$P\left[Y \leq v, U \leq \frac{\pi(Y)}{Mp_u(Y)}\right] = \int_{-\infty}^{v} p_u(y) \int_0^{\frac{\pi(y)}{Mp_u(y)}} dudy = M^{-1} \int_{-\infty}^{v} \pi(y)dy,$$

whence

$$P(X \leq v) = P(Y \leq v | U \leq \tfrac{\pi(Y)}{Mp_u(Y)}) = \frac{P(Y \leq v, U \leq \frac{\pi(Y)}{Mp_u(Y)})}{\int P(U \leq \frac{\pi(Y)}{Mp_u(Y)} | y) p_u(y) dy} = \int_{-\infty}^{v} \pi(y)dy.$$

Example 3.6: A simple rejection method for sampling from posterior density $h_x(\theta)$ (just knowing its kernel $h_x^*(\theta)$ from the above result), when prior density is proper, is to take this as the envelope function and M equals to the maximum likelihood since

$$h_x^*(\theta) = h(\theta) f(x|\theta) \leq Mh(\theta).$$

Thus, generating $\theta_0 \sim h(\cdot)$ and $u_0 \sim Unif(0,1)$, one has $\theta_0 \sim h_x^*(\cdot)$ if $u_0 < \frac{f(x|\theta_0)}{M}$. ■

Applying this method requires finding an instrumental density p_u that fits well with π, having heavier tails than this, and a simple generator for it. In addition, the constant M must be chosen to be as small as possible for the algorithm to be efficient.

3.3.2 Adaptive rejection algorithm

Gilks and Wild (1992) suggest an automatic method of generating bounding functions for sampling of target-densities $\pi(x)$ (or its relevant factor $\pi^*(x)$) logarithmically concave, that is, whose logarithm is a concave function.

It is known that any concave function can be upperly and lowerly limited by envelopes formed by linear sections.

To construct them we consider points on the function graph and pass through and between these points, respectively, tangents and cords to the graph - see Figure 3.

Let $\pi(x) \propto \exp(L(x))$ then be a log-concave univariate probability density function supported $D \subset I\!R$ and $T_k = \{x_i, i = 1, \ldots, k\}$ a set of k ordered points, $x_1 \leq x_2 \leq \ldots \leq x_k$, for which $L(x)$ and $L'(x) = dL(x)/dx$, if π is continuous and differentiable in D.

Set the envelope function in T_k to $\pi(x)$ as $\exp[u_k(x)]$ where $u_k(x)$ is the linear piecewise *upper envelope* of $L(x)$

$$u_k(x) = L(x_j) + (x - x_j)L'(x_j),$$

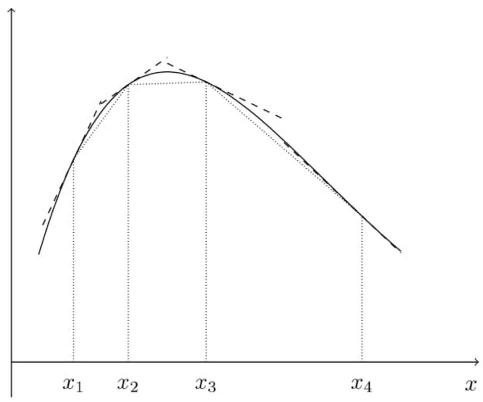


Figure 3: Upper and lower linear envelopes by sections to delimit the $L(x) = \ln \pi(x)$ function in the adaptive rejection method.

for $x \in [z_{j-1}, z_j]$ and $j = 1, \ldots, k-1$ with

$$z_j = \frac{L(x_{j+1}) - L(x_j) - x_{j+1}L'(x_{j+1}) + x_j L'(x_j)}{L'(x_j) - L'(x_{j+1})}$$

the intersection point of tangents to the curve $l(x)$ at x_j and x_{j+1}. The points z_0 and z_k are taken, respectively, as the lower limit of D (or $-\infty$ if D is not limited lower) and higher than D (or $+\infty$ if D is not limited upper).

Also set the lower framing function in T_k of $\pi(x)$ as $\exp[l_k(x)]$, where $l_k(x)$ is the linear piecewise *lower envelope* of $L(x)$

$$l_k(x) = \frac{(x_{j+1} - x)L(x_j) + (x - x_j)L(x_{j+1})}{x_{j+1} - x_j}$$

for $x \in [x_j, x_{j+1}]$ and $j = 1, \ldots, k-1$. For $x < x_1$ or $x > x_k$, $l_k(x) = -\infty$.

Since it is assumed that $L(x)$ is concave, we have $l_k(x) \leq L(x) \leq u_k(x)$ for all x in D.

Adaptive rejection algorithm (a sample of n points from $\pi(x)$)

1. x is obtained from normalized envelope $S_k(x) = \exp[u_k(x)] / \int_D \exp[u_k(y)]dy$.
2. U is obtained from an Uniform distribution in $(0, 1)$ and
   - if $u \leq \exp\{l_k(x) - u_k(x)\}$, x is accepted without making any calculation of the function $L(x)$ at this point; otherwise, $L(x)$ is calculated and the following rejection test is taken;
   - if $u \leq \exp\{L(x) - u_k(x)\}$, x is accepted; otherwise x is rejected;
   - resume the previous steps until you accept the generated candidate.
3. Once the previous cycle with acceptance of the candidate value is over, the upper and lower envelopes are updated by adding x to T_k, and increasing k by one unit.
4. Return to 1.
5. It ends when the number of points initially decided to sample is reached.

4 Markov Chain Monte Carlo Methods

Markov chain Monte Carlo *MCMC* Methods are used to generate a sample directly from e.g. a posterior distribution $h(\theta|x)$, according to the following procedure:

1. Construction of a Θ state space Markov chain that is simple to simulate and whose stationary distribution is $h(\theta|x)$;
2. Simulation of this chain over a long period using simulated chain values to draw inferences about posterior quantities (3.4) through the Monte Carlo integration method,

$$E(g(\theta)|x) \;\approx\; \frac{1}{m}\sum_{j=1}^{m} g(\theta^{(j)}),$$

   where $\theta^{(j)}$ is the j-th value for θ in a chain with m iterations.

Notions and basic results about Markov chains:

Definition 4.1: A *stochastic process* is any collection of random variables defined over the same probability space, $\{U^{(t)} \equiv U_t, t \in T\}$, where T is a subset of $I\!R$ that, for convenience, is understood as a class of time instants (known as index (or parameter) set).

When this class is the discrete set of positive integers $T = \{0, 1, 2, \ldots\}$, the discrete-time stochastic process is usually denoted by $\{U_n, n \geq 0\}$. This is the typical situation in the context of a stochastic simulation scheme.

The $\mathcal{U}$ set of variable values is named *state space* representing the support of a parameter vector. Note that for each $t \in T$, U_t is a random variable, whereas for each element of $w \in \mathcal{U}$, U_t is a function of t.

Definition 4.2: The process $\{U_n, n \geq 0\}$ satisfying the Markov conditional independence property is called *Markov chain* and can be defined by

$$P(U_{n+1}\in A|U_0=u_0, \ldots, U_n=u) = P(U_{n+1}\in A|U_n=u) \equiv P_n(u, A), \tag{4.1}$$

for the entire event A and $n \geq 0$, where the $P_n(u, A)$ symbol denotes the call *transition function* (in one step) when leave from instant n.

When the transition function is invariable with n, then denoted by $P(u, A)$, the Markov chain is called *homogeneous*.

Note that for a discrete Markov chain $\{U_t, t \in I\!N\}$, the transition probabilities in (4.1) *i.e.* $p(u, v) = P(U_{n+1} = v|U_n=u)$ satisfy: i) $p(u, v) \geq 0$, ii) $\sum_v p(u, v) = 1$, $\forall\, u, v \in \mathcal{U}$.

Example 4.1: Consider a particle moving independently left or right on the line with successive displacements from the current position governed by a probability function $f(u)$ with $u \in \mathbb{Z}$ and U_n representing its position at instant n, $n \in I\!N$. Initially, U_0 is distributed according to $\pi(0)$, being able to write

$$U_1 = U_0 + Z_1, \;\cdots\; U_n = U_{n-1} + Z_n = Z_1 + \cdots + Z_n,$$

where Z_i are independent random variables with probability function f. So $\{U_n, n \in I\!N\}$ is a Markov chain in $\mathbb{Z}$.

If $f(1) = p$, $f(-1) = q$ and $f(0) = r$ with $p+q+r = 1$, the chain transition probabilities are $p(u, v) = p$, for $v = u + 1$, $p(u, v) = q$, for $v = u - 1$, $p(u, v) = r$ for $v = u$, $p(u, v) = 0$ for $v \neq u - 1, u, u + 1$.

Note that to know where the chain is at the moment $t = n$, just know the distribution of $Z_1 + \cdots + Z_n$. This chain is known as *random walk*. ■

Studying the asymptotic behavior ($n \to \infty$) of chains is fundamental to MCMC methods and the following concept plays a crucial role in it.

Definition 4.3: It is said that a probability distribution $\pi(u), u \in \mathcal{U}$ is *stationary* if

$$\pi(v) = \sum_u \pi(u)\, p(u,v). \tag{4.2}$$

In particular, the initial distribution $P(U_0 = u) = \pi(u)$ is stationary iif the distribution of U_n is invariant with n, *i.e.* $P(U_n = u) = \pi(u), \forall\, n \geq 0$.

Convergence to the stationary distribution π depends on the chain having some stability properties known as *irreducibility* and *recurrence* and more broadly being *ergodic*, involving the strong law of large numbers (see Paulino *et al.* , 2018).

Example 4.2: Let $\{U_n : n \geq 0\}$ be a Markov chain in $\mathcal{U} = \{0,1\}$ with $\pi_0 = (\pi_0(0), \pi_0(1))$ and transition matrix $P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$.

Its stationary distribution $\pi = (\pi(0), \pi(1))$ is the system solution $\pi = \pi\, P$ that results in the equations

$$\pi(j) = \pi(0)\, p(0,j) + \pi(1)\, p(1,j), \quad j = 0,1.$$

Using the first equation *i.e.* $\pi(0) = (4/3)\pi(1)$ in the restriction $\pi(0) + \pi(1) = 1$, the solution is $\pi = (4/7, 3/7)$.

Note that the n-step transition matrix P_n, with the genetic element $p_n(u,v) = \sum_k p_{n-1}(u,k)p(k,v)$ can be given by

$$P_n = \frac{1}{0.7}\begin{pmatrix} 0.4 & 0.3 \\ 0.4 & 0.3 \end{pmatrix} + \frac{0.3^n}{0.7}\begin{pmatrix} 0.3 & -0.3 \\ -0.4 & 0.4 \end{pmatrix}.$$

So, $\lim_{n\to\infty} P_n = (1,1)^T \times \left(\frac{4}{7}, \frac{3}{7}\right)$ approaching the stationary distribution π. ∎

4.1 Metropolis-Hastings algorithm

The origin of the Metropolis-Hastings (M-H) algorithm dates back to the simulation method used in Metropolis *et al.* (1953), and was later generalized by Hastings (1970).

For convenience, U here represents the old k-parametric vector θ ($k \geq 2$), so $U^{(t)} \equiv U_t = \theta_{(t)}$, while stationary distribution continues to be denoted by $\pi(u), u \in \mathcal{U}$, and $\theta_{(t)j}$ denotes the j-th component of $\theta_{(t)}$.

The fundamental instrument of the M-H algorithm is a conditional distribution $q(v|u) \equiv q(u,v)$ to which the proposed simulated value generator role is reserved (defining a chain if this distribution is a corresponding function of transition). Therefore, a requirement of $q(\cdot|\cdot)$ (proposing distribution) is that it can be easily simulated.

The values $v^{(t)}$ generated successively from $q(\cdot|\cdot)$ are subject to a stochastic sieve based on $q(\cdot|\cdot)$ and $\pi(\cdot)$, which determines the acceptance or rejection of each, where the value replacing the $v^{(t)}$ rejected is the previous simulated value that was accepted.

Metropolis-Hastings (M-H) algorithm

1. Given $u^{(t)}$, $t = 0,1,2,\ldots$, generate a value of $V^{(t)} \sim q(v|u^{(t)})$.
2. Calculate the value of the ratio M-H $R(u^{(t)}, V^{(t)})$, where $R(u,v) = \frac{\pi(v)q(u|v)}{\pi(u)q(v|u)}$, and consider the probability

$$\alpha(u,v) = \min\left\{\frac{\pi(v)q(u|v)}{\pi(u)q(v|u)}, 1\right\}. \tag{4.3}$$

3. Take the next chain value as the realization of

$$U^{(t+1)} = \begin{cases} V^{(t)}, & \text{with probability } \alpha(u^{(t)}, V^{(t)}) \\ u^{(t)}, & \text{with probability } 1 - \alpha(u^{(t)}, V^{(t)}). \end{cases}$$

Each step of the chain thus consists of either replacing the current value u with the candidate v generated from the proposing distribution $q(u, \cdot)$ or, alternatively, retaining the chain at the current value. Acceptance decision (with probability $\alpha = \alpha(u, v)$) or not of the transition $u \to v$ is performed using the following procedure:

1. A value z is generated from a variable $Z \sim U(0, 1)$;
2. If $z \leq \alpha$, the transition to v is accepted; if $z > \alpha$, the candidate value is rejected, thus immobilizing the chain at u.

A set of pertinent observations about this algorithm are presented in Paulino et al. (2018, Notes 9.1-9.5). For example, the M-H algorithm has limited requirements on the π and q distributions to ensure the convergence of the chain to π.

Given the generic character of the M-H algorithm, some of its specialties are now illustrated.

Independence M-H algorithm

The designation of this algorithm means that the proposal distribution does not depend in any way on the iterations, *i.e.* $q(v|u) = q(v)$.

An example is given by a Normal distribution (usually multivariate) with mean vector v_0 and covariance matrix Σ_0 fixed regardless of the iteration. This implies that the probability of acceptance of each value generated by it is rewritten as

$$\alpha(u^{(t)}, v^{(t)}) = min\left\{ \frac{\pi(v^{(t)})\, q(u^{(t)})}{\pi(u^{(t)})\, q(v^{(t)})}, 1 \right\}, \; t \geq 0, \tag{4.4}$$

thus, it keeps varying with the accepted value of the previous iteration.

Example 4.3: An illustration of this algorithm framed in the simulation of a posterior, *i.e.* $\pi(\theta) = h(\theta|x) \propto L(\theta|x)h(\theta)$, where $\{U^{(t)} \equiv \theta^{(t)}\}$, is realized when considering $q(\theta) = h(\theta)$.

Note that in this case, the support of q covers that of π, although the two densities can be quite distinct.

The M-H ratio in this case is characterized by a likelihood ratio,

$$R(\theta^{(t)}, V^{(t)}) = \frac{L(V^{(t)}|x)}{L(\theta^{(t)}|x)}.$$

■

Note: The acceptance probability (4.3) shows that the efficiency of this algorithm is as greater as closer the instrumental distribution $q(\theta)$ is to the posterior distribution $\pi(\theta)$.

Random walk M-H algorithm

This algorithm starts from a Markov chain in the simulation of the instrumental distribution defined by $V^{(t)} = U^{(t)} + \varepsilon_t$, where ε_t represents a random error with a distribution q^* regardless of $U^{(t)}$.

It is thus a random walk associated with the transition density $q(v|u) = q^*(v - u)$. Usual choices for the symmetric proposal distribution q^* include Uniform distributions on a ball centered on the origin, Gaussian and t-Student.

Note: If the proposal distribution is symmetric, *i.e.* $q(v|u) = q(u|v)$, the MH ratios are simplified to $R(u,v) = \frac{\pi(v)}{\pi(u)}$, clearly showing that they do not need to know the normalizing constant of the target distribution (see Metropolis *et al.* , 1953).

Example 4.4: Assume that a random sample $(X_1, \ldots, X_n)$ from a two-component normal mixture is observed. The mixture distribution is denoted by $p\,N(\mu_1, \sigma_1^2) + (1-p)\,N(\mu_2, \sigma_2^2)$ and the density of the mixture is

$$\pi(x) = p\,f_1(x) + (1-p)\,f_2(x)$$

where f_1 and f_2 are the densities of the two Normal distributions, respectively.

If the densities f_1 and f_2 are completely specified, the issue is to make inference on the mixing parameter p given the observed sample.

In this sense, one can generate a chain using an independence M-H algorithm that has the posterior distribution of p as the target distribution.

The most obvious choice of the proposal distribution is the Beta(a,b) distribution. With no prior information on p, one might consider the Beta(1,1) proposal distribution (*i.e.* Uniform(0,1)).

The candidate value $v^{(t)}$ is accepted with probability

$$\alpha(u^{(t)}, v^{(t)}) = \min\left(\frac{\pi(v^{(t)})q(u^{(t)})}{\pi(u^{(t)})q(v^{(t)})}, 1\right),$$

where $q(v) \propto v^{a-1}\,(1-v)^{b-1}$ and the M-H ratio is

$$R(u^{(t)}, v^{(t)}) = \frac{(u^{(t)})^{a-1}\,(1-u^{(t)})^{b-1}\prod_{i=1}^{n} v^{(t)}\,f_1(x_i) + (1-v^{(t)})\,f_2(x_i)}{(v^{(t)})^{a-1}\,(1-v^{(t)})^{b-1}\prod_{i=1}^{n} u^{(t)}\,f_1(x_i) + (1-u^{(t)})\,f_2(x_i)}.$$

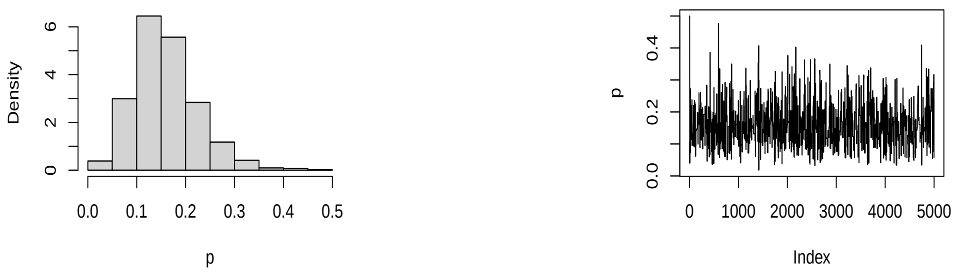


Figure 4: The histogram and trace plot of the generated sample ($m = 5000$) for the posterior of p based on simulate data ($n = 30$) from the normal mixture, $0.2\,N(0,1) + 0.8N(5,1)$, and the proposal distribution $Beta(1,1) = U(0,1)$.

■

4.2 Gibbs sampling

Among the various ways of constructing Markov chains in MCMC methods, we highlight the Gibbs sampling method introduced by Geman and Geman (1984) to simulate multivariate distributions in *image processing models.*

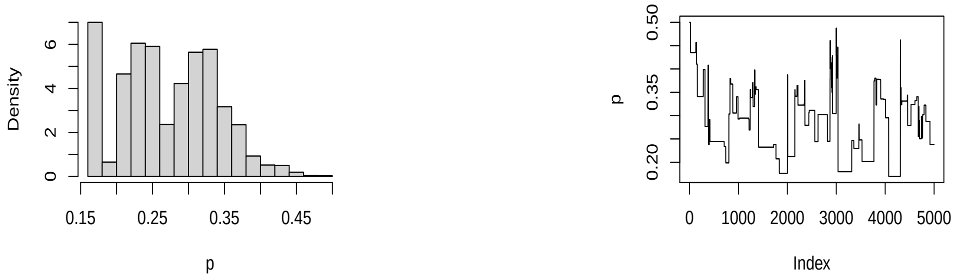


Figure 5: The histogram and trace plot of the generated sample ($m = 5000$) for the posterior of p based on simulate data ($n = 30$) from the normal mixture, $0.2\,N(0,1) + 0.8N(5,1)$, and the proposal distribution $Beta(5,2)$.

The Gibbs algorithm is based on the fact that a joint posterior distribution $h(\theta|x)$ with $\theta = (\theta_1, \ldots, \theta_k)^T$ can be in general conditions determined solely by k full conditional posterior distributions

$$h(\theta_j \,|\, \theta_{-j}, x), \quad j = 1, \ldots, k, \tag{4.5}$$

where θ_{-j} is the vetor θ without the component θ_j.

The *Gibbs sampling* algorithm has the following dynamic Markovian scheme for sampling conditional distributions (4.5):

1. A starting value set is chosen for θ,

$$\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_k^{(0)})^T;$$

2. θ components are generated from (4.5), considering the iterative procedure below for the l-th iteration

$$\begin{array}{lll}
\theta_1^{(l)} & \text{from} & h(\theta_1 \,|\, \theta_2^{(l-1)}, \ldots, \theta_k^{(l-1)}, x), \\
\theta_2^{(l)} & \text{from} & h(\theta_2 \,|\, \theta_1^{(l)}, \theta_3^{(l-1)}, \ldots, \theta_k^{(l-1)}, x), \\
\theta_3^{(l)} & \text{from} & h(\theta_3 \,|\, \theta_1^{(l)}, \theta_2^{(l)}, \theta_4^{(l-1)}, \ldots, \theta_k^{(l-1)}, x), \\
\vdots & & \vdots \\
\theta_{k-1}^{(l)} & \text{from} & h(\theta_{k-1} \,|\, \theta_1^{(l)}, \ldots, \theta_{k-2}^{(l)}, \theta_k^{(l-1)}, x), \\
\theta_k^{(l)} & \text{from} & h(\theta_k \,|\, \theta_1^{(l)}, \ldots, \theta_{k-1}^{(l)}, x)\ ;
\end{array}$$

3. The previous step is repeated s times until the generation of m *independent samples* of θ. Note that each succession element $\theta^{(1)}, \ldots, \theta^{(s)}, \ldots$ is an realization of a Markov chain with state space Θ and transition probabilities given by

$$p(\theta^{(s)}, \theta^{(s+1)}) \;=\; \prod_{j=1}^{k} h(\theta_j^{(s+1)} \,|\, \theta_{l>j}^{(s)}, \theta_{l<j}^{(s+1)}, x).$$

When $s \to \infty$ in the above procedure, $\theta^{(s)} = (\theta_1^{(s)}, \ldots, \theta_k^{(s)})^T$ tends to distribution for a random vector with p.d.f. $h(\theta|x)$ (Tanner, 1996). See also Casella and George (1992) and Gelfand and Smith (1990).

In particular, the j-th marginal posterior distribution can be obtained using its empirical distribution with the m *samples i.e.*

$$h(\theta_j|x) \;\approx\; \frac{1}{m}\sum_{l=1}^{m} h(\theta_j\,|\,\theta_{-j}^{(l)}, x), \tag{4.6}$$

where $h(\theta_j\,|\,\theta_{-j}^{(l)}, x)$ is the distribution (4.5) with the $\theta_{j'}$, $j' \neq j = 1, \ldots, k$, replaced by their respective values in the iteration l, $l = 1, \ldots, m$.

Note that the $s - m$ iterations, $m < s$, of the procedure in question are ignored in estimating the quantities of interest, as they are part of the chain's *burn-in* period, where there is believed to be a greater correlation between the $\theta^{(s)}$, $s = 1, 2, \ldots$.

Predictive distribution

MCMC methods are also used to predict future observation y from a model indexed by the θ parameter concerned. For example, predictive distribution

$$p(y|x) \;=\; \int f(y|\theta, x)h(\theta|x)d\theta, \tag{4.7}$$

where $f(y|\theta, x)$ is the distribution of y under this parametric model, can be estimated by

$$\widetilde{p}(y|x) \;=\; \frac{1}{m}\sum_{l=1}^{m} f(y|\theta^{(l)}, x),$$

sendo $\theta^{(l)}$, $l = 1, \ldots, m$, the values obtained for θ in the m samples referred to above.

In order to estimate $p(x_i|x_{(-i)})$, Gelfand (1996) suggests the use of the harmonic mean of $\{f(x_i|x_{(-i)}, \theta_{(j)}), j = 1, \ldots, m\}$. Considering that

$$p(x)h(\theta|x) = h(\theta)f(x|\theta) = h(\theta)f(x_i|x_{(-i)}, \theta)f(x_{(-i)}|\theta),$$

one has

$$\begin{aligned} p(x_i|x_{(-i)}) \;&=\; \frac{p(x)}{p(x_{(-i)})} = \frac{1}{\int \frac{f(x_{(-i)}|\theta)\;h(\theta)\;h(\theta|x)}{p(x)\;h(\theta|x)}\;d\theta} \\ &=\; \frac{1}{\int \frac{1}{f(x_i|x_{(-i)},\theta)}\;h(\theta|x)\;d\theta}, \end{aligned}$$

and therefore, if $\{\theta_{(j)}; j = 1, \ldots, m\}$ is a sample of $h(\theta|x)$, then

$$\hat{p}(x_i|x_{(-i)}) = \frac{1}{\frac{1}{m}\sum_{j=1}^{m}\frac{1}{f(x_i|x_{(-i)},\theta_{(j)})}}. \tag{4.8}$$

Example 4.5: Let (Y, X) be a random pair where Y conditional on $X = x$ follows a Poisson distribution with mean value $\lambda(x) = \delta^x$ and X has a Normal distribution with mean value μ and precision $\tau = 1/\sigma^2$. The likelihood function concerning data consisting of n i.i.d. observations of this random pair, $\mathcal{D} = \{(y_1, x_1), \ldots, (y_n, x_n)\}$, is

$$f(x, y|\theta) = \prod_{i=1}^{n}\frac{[\delta^{x_i}]^{y_i}}{y_i!}e^{-\delta^{x_i}}\left[\frac{\tau}{2\pi}\right]^{1/2}\exp\left\{-\frac{\tau}{2}(x_i - \mu)^2\right\}$$

where $\theta = (\delta, \mu, \tau)$ for $\delta, \tau > 0$ and $-\infty < \mu < +\infty$. If you consider a non-informative prior distribution $h(\delta, \mu, \tau) \propto (\delta\tau)^{-1}$, the posterior probability density function for θ is:

$$\begin{aligned} h(\theta|\mathcal{D}) \;&\propto\; \tau^{n/2-1}\delta^{\sum_i x_i y_i - 1}\exp\{-\textstyle\sum_i \delta^{x_i}\} \times \\ &\times\; \exp\left\{-\frac{\tau}{2}\left[\textstyle\sum_i (x_i - \bar{x})^2 + n(\mu - \bar{x})^2\right]\right\}, \end{aligned}$$

being the corresponding complete conditional distributions easily identified by

$$\begin{array}{rcl} h(\delta|\mathcal{D},\mu,\tau) & \propto & \delta^{\sum_i x_i y_i - 1} \exp\left\{-\sum_i \delta^{x_i}\right\}, \quad \delta > 0; \\ h(\mu|\mathcal{D},\delta,\tau) & = & N\left(\bar{x}, (\tau n)^{-1}\right), \quad -\infty < \mu < +\infty; \\ h(\tau|\mathcal{D},\delta,\mu) & = & Ga\left(\frac{n}{2}, \frac{A}{2}\right), \quad \tau > 0, \end{array} \tag{4.9}$$

where $A = \sum_i (x_i - \bar{x})^2 + n(\mu - \bar{x})^2$.

The distributions concerning μ and τ are well-known, sampling in the corresponding Gibbs steps by using known and efficient simulation algorithms. The situation regarding δ is no longer trivial requiring the use of other methods such as rejection (*e.g.* the adaptive rejection method, proposed by Wild & Gilks, 1993). ■

4.3 Convergence diagnostic techniques

There are several instruments and methods available for monitoring and diagnosing both types of convergence (chains or empirical means), some of which are automatically included in more specific software or more general Bayesian analysis.

The best known instrument for monitoring convergence for stationary distribution is the graphical representation for each scalar quantity of simulated chain values over successive iterations, connected by a continuous line.

Figure 6 depicts typical aspects of the start (left) and end (right) trace plots of the simulated value sequence.

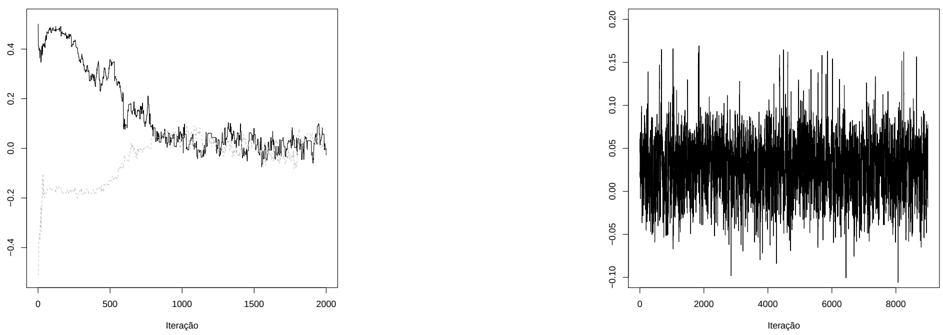


Figure 6: Trace plots for the same parameter of 2 chains over the first 1,000 iterations (left) and 1 chain over the last 9,000 iterations (right).

That informal *monitoring tool* to detect the occurrence of stability is the graphical overlap of posterior marginal density estimates as the number of iterations used in the estimation increases.

Another type of method is the use of nonparametric tests to determine the distributional chain stabilization, *e.g.* , the Kolmogorov-Smirnov test for comparison of two subsamples.

To monitor the convergence of the empirical averages of scalar quantities $g(\theta)$, given by $S_m = \sum_{i=1}^{m} g(\theta_{(i)})/m$, one possible technique is to construct the *plot of cumulative sums*, given by $D_l = \sum_{i=1}^{l} \left[g(\theta_{(i)}) - S_m\right]$, $l = 1, \ldots, m$. Rapidly exploiting chains of target support tend to present this chart with an uneven look and generally concentrated around 0.

An alternative technique is to construct the *plot of mean evolution* over iterations,

$$S_m = \left[(m-1)S_{m-1} + g(\theta_{(m)})\right]/m, \; m = 1, 2, \ldots, \tag{4.10}$$

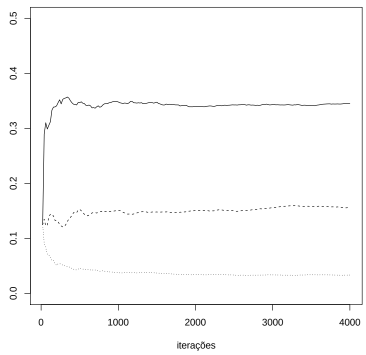


Figure 7: Evolution plot of the means of three of the probabilistic parameters of a model related to a 2^3 contingency table.

or other empirical quantities such as quartiles. These simple plots show clearly the stabilization period of the corresponding distributional characteristics, after an initial phase of instability of their estimates, of varying duration from case to case.

Figure 7 shows the evolutionary behavior of the empirical mean of three parameters of a model along their respective chains starting at the common value 0.124. The plot illustrates a rapid stabilization (in the 1000-1500 iteration range) of the three curves for their posterior mean.

Gelman and Rubin method

Gelman and Rubin (1992) suggest the use of variance components of chain multiple sequences, simulated from a variety of dispersed starting points by using the following steps:

- $m \geq 2$ sequences, each of length $2n$, are simulated from simulated starting points of an overdispersed distribution relative to the target distribution (equilibrium distribution).
- Discard the first n iterations of each sequence.
- Since $g(\cdot)$ is the scalar quantity of interest to be estimated (g is typically a function of the parameter θ), the components of variance W and B are calculated based on their simulated values, that is, the variance within each sequence and the variance between the sequences, respectively.
- The target-mean of g is estimated as a sample mean of all mn simulated values of g.
- The target-variance of $g(\theta)$ (V) is estimated as a weighted mean of W and B.
- Calculate the *scaling reduction factor* $\hat{R} = \sqrt{V/W}$. This ratio decreases to 1 when $n \to \infty$. Values of $\hat{R} \approx 1$ are indications that each of the m sequences of n simulated observations approaches the target distribution.

Geweke method

Being $\theta^{(t)}, t = 1, \ldots, N$ a sequence of values ??simulated by the MCMC procedure and $g(\theta)$ a function of θ to be estimated, the trajectory $g^{(t)} = g(\theta^{(t)})$, $t = 1, 2, \ldots$, defines a time series. Geweke's method (1992) is based on the application of usual techniques in series to verify the convergence of the simulated sequence.

Look at the series over a sufficiently long number N of iterations and calculate the average $g_a = \frac{1}{n_a}\sum g(\theta^{(t)})$ a trick on n_a from the first iterations, as well as the average $g_b = \frac{1}{n_b}\sum g(\theta^{(t)})$ a trick on n_b from the last iterations.

If the chain is stationary, then the average of the first part of the chain should be similar to the average of the second part of the chain. Assuming that n_a/N and n_b/N are fixed and $N \to \infty$ can be shown that

$$\frac{(g_a - g_b)}{\sqrt{(s_a^2/n_a) + (s_b^2/n_b)}} \to N(0,1),$$

where s_a^2 and s_b^2 are independent estimates of the asymptotic variances of g_a and g_b adjusted for autocorrelation. According to the result of this statistic, it can be verified whether or not there is an indication of convergence.

Raftery and Lewis method

Suppose you want to estimate a posterior quantile q of a parameter function, with a certain tolerance r and a probability s of being within these tolerance limits. The Raftery and Lewis (1992) method calculates the number of iterations N and the number of iterations M of the burn-in period required to satisfy the specified conditions.

The result of this diagnostic method has, in addition to N and M, N_{min} as the minimum number for a pilot sample, and $I = (M+N)/N_{min}$, called the dependency factor, interpreted as the proportional increment in the number of iterations attributable to the serial dependency. High values of this factor (> 5) may indicate influential initial values, high correlation between coefficients, or a poorly mixed chain in the support of the posterior distribution.

Heidelberg and Welch method

Heidelberg and Welch (1983) proposed a test statistic, based on the Cramer-von Mises statistical test, to test the null hypothesis that the simulated Markov chain comes from the stationary distribution, applied to each monitored variable as follows:

- Generates a N dimension chain and sets a level α (*e.g.* 0.05).
- For each monitored variable, the value of the test statistic is calculated using the N iterations. According to the result of the test, the decision is made as to whether or not to reject the null hypothesis.
- If the null hypothesis is rejected, the test statistic is recalculated by discarding 10% from the first iterations. This procedure is repeated if the null hypothesis is still rejected.
- If the null hypothesis is still rejected when the number of iterations used in the calculation of the test statistic reaches 50% of the initial N, then the iterative process must continue as the chain has not yet reached equilibrium. In this case, the software executing this method results in the test statistic and indicates that the chain has failed the stationarity test.
- Otherwise, the non-rejecting portion of the chain is used to estimate the mean (m) and the asymptotic standard error (s) of the mean, which is calculated using a time series method. If $1.96\, s < m\, epsilon$, with small ϵ (*e.g.* 0.1), then the chain passes the stationarity test. If $1.96\, s \geq m\,\epsilon$, this means there is a need to continue with the iterative process.

Software

The software R has a variety of packages that can be used to make Bayesian inference. It is advised to consult the webpage

`http://cran.r-project.org/web/views/Bayesian.html`

where one can find the DPpackage which contains functions for making nonparametric Bayesian inference, the package bayesSurv specific for making Bayesian inference in survival models, etc.

Software that implements stochastic simulation-based methods can be used by linking to **R**, namely OpenBUGS (Thomas et al., 2006), JAGS (Plummer, 2003), INLA (Rue et al., 2009), BayesX (Belitz et al., 2013) and Stan (Carpenter et al., 2017).

Chain convergence monitoring can be done using software CODA and BOA, both packages of **R**.

5 Statistical Models

A *statistical model* is a mathematical model that embodies a set of statistical assumptions concerning the generation of sample data (Cox, 2006). It is usually specified as a mathematical relationship between one or more random variables and other non-random variables.

Formally, a statistical model is usually thought of as a pair $(\Omega, \mathcal{P})$, where Ω is the set of possible observations, *i.e.* the sample space, and $\mathcal{P}$ is a set of probability distributions on Ω, being almost always parameterized: $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ (McCullagh, 2002). The intuition behind this is that $\mathcal{P}$ contains the "true" distribution, but in practice that may not be the case.

George Box: "Essentially, all models are wrong, but some are useful".

5.1 Model assessment

The model verification or evaluation (*model checking*) is the phase of analysis that aims to assess the adequacy of the model's fit to the data and to the substantive knowledge of the problem at hand.

Note that the verification of the model is not intended to answer the question whether the data were generated from it - a question that never quite admits an affirmative answer.

Model critical examination aims rather to quantify discrepancies with the data, to assess whether or not they are due to chance and to uncover ways to obtain from it models that are likely to be more promising (Paulino *et al.* , 2018).

5.1.1 Regression model selection

Building a good regression model can be summarized in the following steps (Kutner *et al.* , 2005):

1. Preparation and description of the data.
2. Reduce the number of covariates.
3. Evaluation and selection of the "best" model.
4. Validation and interpretation of the selected model.

Example 5.1: Consider the following regression model with response Y and explanatory x_1, x_2 and x_3 variables, to illustrate the steps for building a good model.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon.$$

1. In the model fit, one can conclude that the normality assumption is not satisfied (quantile-quantile or Q-Q plot).
2. The number of regression submodels increases substantially with the number of covariates, *e.g.* there are $2^3 = 8$ regression submodels.
3. The model evaluation can also be made via diagnostic techniques, *e.g.* residual analysis, with residual defined as $r \equiv y - \widehat{E}(Y)$.
4. To validate the regression model, one can use two strategies:

- Using another dataset to confirm the selected model and evaluate its predictive ability.
- Compare the obtained results with expected results in theory or approximate empirical results.

The "best" model should be a model that achieves a good balance between three factors: good fit, parsimony and interpretation. ■

Stepwise regression

The basic idea of the automatic procedure to reduce covariates in the regression analysis, known as *stepwise* regression, is to start from a given model and make a series of steps to exclude terms (covariates) of the model or to add terms candidates for inclusion.

In the stepwise regression, the model selection is made:

- Starting with the model with more parameters and considering alternative models by the exclusion of covariates (regressive elimination or *backward elimination*),
- Starting from the model with fewer parameters, *e.g.* , the model without covariates, and considering alternative models by the inclusion of covariates (progressive selection or *forward selection).*

The removal of the regression model assumptions may lead to unreliable results before the model fit (Kutner *et al.* , 2005).

The *residual analysis* is useful not only to a local quality assessment of a model fit with respect to the choice of distribution and of the linear regression equation terms, but also to help identify poorly fitted observations.

In linear regression model (2.1), one has $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ as least squares estimator of $\boldsymbol{\beta}$, whereby the vector of the residuals is naturally given by

$$\mathbf{r} \equiv \mathbf{y} - \hat{E}(\mathbf{Y}) = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{H}\mathbf{y}, \tag{5.1}$$

where $\mathbf{r} = (r_1, \ldots, r_n)^T$ and $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

Transformations

The use of transformations for the response variable (Y) or explanatory variables ($\mathbf{x}$) is often sufficient to ensure the assumptions of the regression model when applied to transformed data.

- Simple transformations of $\mathbf{x}$ may be useful to linearize non-linear regression relationship without affecting the distribution of Y.
- Transformations $\sqrt{Y}$ and $\log Y$ are recommended when the variance of the random error increases proportionally to x_i and x_i^2, respectively, $i = \ldots, n$.
- The simultaneous transformation of Y and $\mathbf{x}$ may be required.

Autocorrelation

If possible, one can use the plot of residuals *versus* time to investigate any correlation between the random errors over time (*autocorrelation* or *correlation in time series*). For example, in econometric time series studies, that could be due to the absence of one or more important covariates in the model.

If there is autocorrelation, you must fit the following regression model

$$Y_t = \mathbf{x}_t^T \boldsymbol{\beta} + \epsilon_t \quad \text{with} \quad \epsilon_t = \rho\epsilon_{t-1} + u_t, \tag{5.2}$$

where $u_t \sim N(0, \sigma^2)$ independent and $|\rho| < 1$, autocorrelation parameter, $t=1, \ldots, n$.

Note: The Durbin-Watson test is used to test the hypotheses $H_0 : \rho = 0$ *versus* $H_1 : \rho > 0$ (Kutner *et al.* , 2005).

Outliers

The designation *outliers* refers to anomalous observations to data *i.e.* , atypical, discordant or aberrant values.

The presence of outliers is easily detected by plots of residual *versus* a covariate x or the fitted values $\widehat{E}(Y)$.

What should we do when there are discordant values? The response requires a study of the situation, since eliminating these values may affect the analysis. Following two steps before the decision:

- To evaluate the effect of these values on the estimates and forecasts in the model.
- If delete them, check whether these values can be studied separately.

Multicollinearity

When two or more independent variables are highly correlated, *i.e.* , one or more covariates may be expressed in a linear form from the other ones, it is said that these variables are collinear and regression analysis shows *multicolinearity.*

Multicollinearity increases the standard errors of the regression coefficient estimators, causing some covariates are not statistically significant, when they should be significant.

- Example of uncorrelated covariates: $M_1 : \widehat{E}(Y) = 23 + 5.4x_1$, $M_2 : \widehat{E}(Y) = 27 + 9.3x_2$, $M_3 : \widehat{E}(Y) = 0.5 + 5.4x_1 + 9.3x_2$.
- Example of correlated covariates: $M_1 : \widehat{E}(Y) = 0.5 + 0.8x_1$, $M_2 : \widehat{E}(Y) = 1.2 + 2.9x_2$, $M_3 : \widehat{E}(Y) = 35 + 4.3x_1 - 2.3x_2 + 2x_3$.

5.1.2 Bayesian model adequacy

Critical examination of the model is usually based on the posterior predictive distribution of the model, $p(y|x)$ on the basis that its simulable y data should reflect (or not) the observed data expectantly x, in case of a good (bad) model fit.

For this purpose, variables $V(x, \theta)$ are used as the basis for measuring the discrepancy between observed and observable data according to the model.

Measures of discrepancy between observed (x) and observable according to the model (y) based on:

1. Variables $V(x, \theta)$, such as *e.g.* $\ln f(x|\theta)$ or $\sum_{i=1}^n \frac{[x_i - E(X_i|\theta)]^2}{Var(X_i|\theta)}$;
2. Simulated data $\{(y^{(j)}, \theta^{(j)}),\ j = 1, \ldots, m)\}$ from the joint distribution of (y, θ) conditional on x.

- Scatterplot of $\left\{\left(V(y^{(j)}, \theta^{(j)}), V(x, \theta^{(j)})\right),\ j = 1, \ldots, m\right\}$.

- Histogram of the values $\{V(y^{(j)},\theta^{(j)}) - V(x,\theta^{(j)}),\ j=1,\dots,m\}$.
- *Bayesian P-values* (posterior predictive)

$$P_B = P[V(Y,\theta) \geq V(x,\theta)|x] \simeq \frac{\#\{(y^{(j)},\theta^{(j)}) : V(y^{(j)},\theta^{(j)}) \geq V(x,\theta^{(j)})\}}{m}. \tag{5.3}$$

  P_B value too small or too large indicates poor fit of the model to the data in terms of what V translates.

- Bayesian residuals in cross validation.
  - Data decompounded into training sample x and validation sample $y = \{y_j\}$. *Standardized predictive residuals*:

$$d_j = \frac{y_j - E(Y_j|x)}{\sqrt{var(Y_j|x)}}, \quad j = 1,\dots,l. \tag{5.4}$$

    Criterion: The smaller $\sum_{j=1}^{l} |d_j|$, the more suitable the model.
  - Cross-validation with an outsider: For every $i = 1,\dots,n$ training sample $x_{(-i)} = (x_j, j \neq i)$; validation sample x_i. *Standardized elimination residuals*:

$$d'_i = \frac{x_i - E(Y_i|x_{(-i)})}{\sqrt{var(Y_i|x_{(-i)})}},\ i = 1,\dots,n, \tag{5.5}$$

    calculated from conditional predictive distributions $p(y_i|x_{(-i)})$.

$$\begin{aligned} p(y_i|x_{(-i)}) &= \int f(y_i|\theta, x_{(-i)})\ h(\theta|x_{(-i)})\ d\theta \qquad (5.6)\\ &\simeq \frac{1}{m}\sum_{j=1}^{m} f(y_i|\theta^{(j)}_{(-i)}, x_{(-i)}), \quad \{\theta^{(j)}_{(-i)}\} \leftarrow h(\theta|x_{(-i)})\\ &\simeq \frac{1}{\frac{1}{m}\sum_{j=1}^{m}\frac{1}{f(y_i|x_{(-i)},\theta^{(j)})}}, \qquad \{\theta^{(j)}\} \leftarrow h(\theta|x). \end{aligned}$$

- *Conditional predictive ordinates* (CPO): $\forall i, p(y_i|x_{(-i)})$ with $y_i = x_i$ in (5.6), estimated at (5.6) via MCMC methods.

  Criterion: The higher $\sum_{i=1}^{n} \ln CPO_i = \ln \prod_{i=1}^{n} p(x_i|x_{(-i)})$ the more suitable the model.
- *Pseudo-Bayes factor*: For comparison of the models M_1 and M_2,

$$PBF(M_1/M_2) = \prod_{i=1}^{n} \frac{p(x_i|x_{(-i)}; M_1)}{p(x_i|x_{(-i)}; M_2)}. \tag{5.7}$$

Example 5.2: From a study of the performance of car models measured by fuel consumption a data set was collected.

These data refer to efficiency values, E_f, measured in miles traveled per gallon of gasoline, weight in pounds (X_1), horsepower power (X_4^*) and number of gears gearbox at levels 3, 4 and 5 represented together by the indicator variables of category 4 (X_2) and 5 (X_3).

Following preliminary work, Normal regression models were considered in the transformed response variable $Y = 100/E_f$, expressed in gallons consumed per 100 miles, and $X_4 = X_4^*/X_1$ (power per unit of weight).

One of the models considered involves the explanatory variables X_1, (X_2, X_3) and $X_4 = X_4^*/X_1$ (power per unit weight) using the multiple regression function.

$$\begin{aligned} &M_1 : \mu \equiv E(Y) = \beta_0 + \textstyle\sum_{j=1}^{4} \beta_j X_j + \beta_5 X_2 X_4 + \beta_6 X_3 X_4, \\ &M_2 : \mu = \beta_0 + \textstyle\sum_{j=1}^{4} \beta_j X_j, \\ &M_3 : \mu = \beta_0 + \beta_1 X_1 + \beta_4 X_4. \end{aligned}$$

The regression model $Y_i, i = 1, \ldots, n = 29 \overset{ind}{\sim} N(\mu_i, \sigma^2)$ has been supplemented with non-informative prior distributions of the usual type, specifically $\beta_j \sim N(0, 10^4)$ and $1/\sigma^2 \sim Ga(10^{-3}, 10^{-3})$. Bayesian analysis of this linear model under a natural conjugated prior distribution or the usual non-informative distribution for (μ, σ^2) is in many ways analytically obtainable in exact terms (Paulino *et al.*, 2018).

For illustration of quantities described here under the Bayesian model above and some reductions of it, with μ parameterized in terms of regression coefficients, the Monte Carlo method simulation path was followed. θ denotes the parameter vector, consisting of the regression and variance coefficients of each model.

Figure 8 presents the scatterplots for the M_1 and M_2 models of the discrepancy measure translated by the standardized quadratic mean deviations,

$$V(x_i, \theta_k) = \sum_{i=1}^{n} \frac{[x_i - E(X_i|\theta_k)]^2}{Var(X_i|\theta_k)},$$

calculated from simulated values θ_k from the posterior distribution of θ (given the observed values of Y and $\{X_j\}$) for the actual data *versus* predicted data.

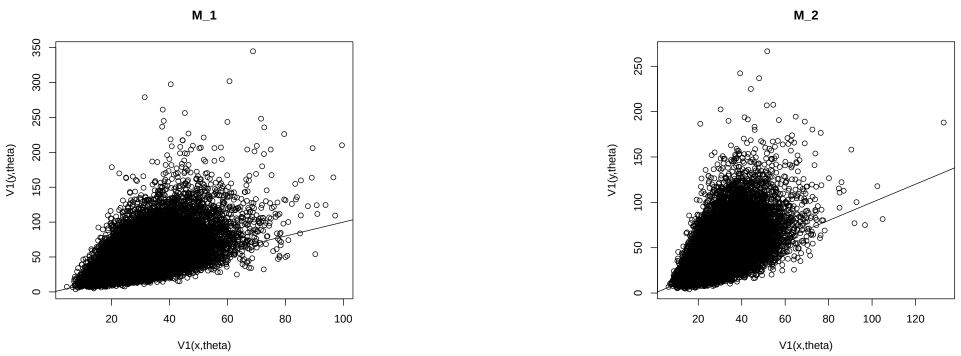


Figure 8: Scatterplots for the models M_1 and M_2.

The asymmetric configuration of the point cloud, with respect to the first quadrant bisector of the two scatterplots, indicates that in both predicted deviations tend to be larger than the corresponding deviations for the observed data, less pronounced for the model M_2.

In Table 3, the Bayesian P-values associated with the discrepancy between the function V evaluated in the observed data and the posterior distribution of $V(Y^*, \theta)$ for predictable data Y^*, point out that the reduced models behave better than the M_1 model and the sum of the CPO logarithms points essentially in the same direction (pro M_2).

Predictive performance measures

Idea: Reflect the extra-sample predictive accuracy with correction of the sample double-usage so that the lower the value, the better the model will perform.

Table 3: Diagnostic measures of the models in comparison.

Model	M_1	M_2	M_3
P_B	0.844	0.778	0.687
$\sum \ln CPO$	-27.577	-23.665	-23.817

■

- Akaike information criterion (AIC)

$$AIC = -2 \ln f(x|\hat{\theta}) + 2p, \tag{5.8}$$

  where $\hat{\theta}$ is MLE and p is the dimension of the parametric space Θ.

- Schwarz/Bayes information criterion (SIC/BIC)

$$BIC = -2 \ln f(x|\hat{\theta}) + p \ln n. \tag{5.9}$$

Schwarz/Bayes information criterion variants (SIC/BIC):

- Carlin-Louis (2000):

$$BIC_{CL} = -2E\left[\ln f(x|\theta)|x\right] + p \ln n. \tag{5.10}$$

- Raftery *et al.* (2007):

$$BIC_R = -2(\bar{l} + s_l^2) + \tilde{p} \ln n, \tag{5.11}$$

  where $\bar{l}$ and s_l^2 are the empirical mean and variance of the simulated values of $l = \ln f(x|\theta)$, $\{l^{(j)} = \ln f(x|\theta^{(j)})\}$; $\tilde{p}$ equals p if it is known or estimated at $2\, s_l^2$, otherwise.

- Deviation information criterion (DIC)

$$DIC = \overline{D(\theta)} + p_D \equiv E_{\theta|x}\left[D(\theta)\right] + \left[\overline{D(\theta)} - D(\bar{\theta})\right] \tag{5.12}$$

  with $D(\theta) = -2 \ln\left[f(x|\theta)/g(x)\right]$, $\overline{D(\theta)} = E_{\theta|x}\left[D(\theta)\right]$ and $\bar{\theta} = E(\theta|x)$, where $g(x)$ denotes some data-only function with purely standardizing effect. For example, $g(x) = f(x|\tilde{\theta})$ where $\tilde{\theta}$ is the estimate of θ in the saturated model.

  In most cases, the expected values in p_D (effective number of parameters) are calculated by Monte Carlo from a simulated sample of $h(\theta|x)$.

  An alternative proposal for the term of complexity of the model ensuring its positivity is $p_D^* = 2Var\left[\ln f(x|\theta)|x\right]$.

- Widely applicable information criterion [of Watanabe] (WAIC)

$$WAIC = -2 \sum_{i=1}^{n} \ln E_{\theta|x}\left[f(x_i|\theta)\right] + 2p_W, \tag{5.13}$$

  with two proposals for the "effective model dimension", p_W:

  ► Analog somehow to p_D used in DIC

$$\begin{aligned} p_{W_1} &= -2 \sum_{i=1}^{n} \{E_{\theta|x}\left[\ln f(x_i|\theta)\right] - \ln E_{\theta|x}\left[f(x_i|\theta)\right]\} \qquad (5.14) \\ &\simeq -2 \sum_{i=1}^{n} \left\{ \frac{1}{m} \sum_{j=1}^{m} \ln f(x_i|\theta^{(j)}) - \ln \left[\frac{1}{m} \sum_{j=1}^{m} f(x_i|\theta^{(j)}) \right] \right\}. \end{aligned}$$

▶ Somewhat similar to p_D^* used alternatively in the DIC

$$\begin{aligned} p_{W_2} &= \sum_{i=1}^{n} Var_{\theta|x}\left[\ln f(x_i|\theta)\right] \\ &\simeq \sum_{i=1}^{n}\left\{\frac{1}{m-1}\sum_{j=1}^{m}\left[l^{(j)}(x_i)-\bar{l}(x_i)\right]^2\right\}, \end{aligned} \tag{5.15}$$

where $l^{(j)}(x_i)=\ln f(x_i|\theta^{(j)})$ and $\bar{l}(x_i)=\frac{1}{m}\sum_{j=1}^{m} l^{(j)}(x_i)$.

Pair analysis via Bayes factor

$$B_{kl}(x)=\frac{p(x|M_k)}{p(x|M_l)}\equiv\frac{p(M_k|x)/p(M_l|x)}{p(M_k)/p(M_l)} \tag{5.16}$$

where for each model

$$p(x|M_r)=\int f_r(x|\theta_r)h_r(\theta_r)d\theta_r \text{ and } p(M_r|x)=\frac{p(M_r)p(x|M_r)}{p(x)},\ \forall r.$$

Calculation options without any distributional impropriety:

- *Simple Monte Carlo method* (usually inefficient)

  Simulation of $h_r(\theta_r)\longrightarrow(\theta_r^{(j)}, j=1,\ldots,m)$

  $$p(x|M_r)\simeq\frac{1}{m}\sum_{j=1}^{m} f_r(x|\theta_r^{(j)})$$

- *Newton-Raftery method* (usually unstable):

  Simulation of $h_r(\theta_r|x)\longrightarrow(\theta_r^{(j)}, j=1,\ldots,m)$

  $$p(x|M_r)=\left[\int\frac{h_r(\theta_r|x)}{f_r(x|\theta_r)}d\theta_r\right]^{-1}\simeq\left(\frac{1}{m}\sum_{j=1}^{m}\left[f_r(x|\theta_r^{(j)})\right]^{-1}\right)^{-1}$$

- *Gelfand-Dey method*:

  Let $g_r(\theta_r)$ be a good approximation of $h_r(\theta_r|x)$ (proper density). With the simulated values of the true posterior distribution,

  $$\begin{aligned} p(x|M_r) &= \left[\int\frac{g_r(\theta_r)}{f_r(x|\theta_r)h(\theta_r)}h_r(\theta_r|x)d\theta_r\right]^{-1} \\ &\simeq \left[\frac{1}{m}\sum_{j=1}^{m}\frac{g_r(\theta_r^{(j)})}{f_r(x|\theta_r^{(j)})h(\theta_r^{(j)})}\right]^{-1}. \end{aligned}$$

Example 5.3: Continuing with Example 5.2 for the purpose of comparative evaluation of 3 multiple regression models in terms of their predictive behavior, the following are the pseudo-Bayes factors (PBF) for comparing models two by two:

$$PBF(M_1/M_2)=0.809; PBF(M_1/M_3)=0.941; PBF(M_2/M_3)=1.164.$$

These values show well the earlier findings that M_2 is the best while M_1 is the worst of the 3 models, in terms of the criterion based on conditional predictive ordinates.

Table 4: Measures DIC, BIC and WAIC for models in comparison.

Model	DIC (p_D)	BIC (p)	WAIC (p_{W_2})
M_1	48.69 (8.27)	67.36 (8)	46.77 (5.38)
M_2	47.32 (6.19)	61.33 (6)	46.70 (4.78)
M_3	47.58 (4.12)	56.93 (4)	47.40 (3.48)

In terms of the information criteria shown in Table 4, it shows that the M_2 model is the best model in terms of the Bayesian measures DIC and WAIC, being beaten by M_3 in BIC, without much amazement as it is known that this measure benefits the simplest models.

Combining all the results obtained here and in the previous example, the applied criteria indicate that the best of the three models is the one in the middle of the complexity scale measured by the number of parameters. ■

5.2 Applications to various statistical problems

Statistics is used in a wide range of types of scientific and social research. Some fields of inquiry use applied statistics so extensively that they have specialized terminology. Namely:

- Astrostatistics (statistical evaluation of astronomical data).
- Actuarial science (assesses risk in the insurance and finance).
- Biostatistics, including medical statistics.
- Data mining (pattern recognition to discover knowledge from data).
- Demography (statistical study of populations).
- Econometrics (statistical analysis of economic data).
- Epidemiology (statistical analysis of disease).
- Geography and geographic information systems (spatial analysis).
- Reliability engineering (survival analysis).
- Sociology and social statistics.

5.2.1 Generalized linear models - II

Generalized linear models (GLM) can be divided into the type of the response variable (Y): i) continuous nature, ii) dichotomous nature or proportional form, iii) counting form.

Logistic model

Given n independent variables $Y_i \sim B(1, \pi_i)$, $i=1, \ldots, n$, *i.e.* $f(y_i|\pi_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$, $y_i = 0, 1$, with $\pi_i = P(Y_i = 1)$, suppose each individual i is associated with a vector $\mathbf{z}_i$ (covariates).

According to Table 5, $E(Y_i) = \pi_i$ and $\theta_i = \ln(\frac{\pi_i}{1-\pi_i})$, when making$\theta_i = \eta_i = \mathbf{z}_i^T\boldsymbol{\beta}$, concluding that the canonical link function is the logit function. Therefore, π_i is related to $\mathbf{z}_i$ through

$$\pi_i = \frac{\exp(\mathbf{z}_i^T\boldsymbol{\beta})}{1+\exp(\mathbf{z}_i^T\boldsymbol{\beta})}. \tag{5.17}$$

Table 5: Some generalized linear models.

random component	structural component		model
	link function	covariates	
Normal	identity	continuous	linear regression
Normal	identity	categorized	analysis of variance
Normal	identity	mixed	analysis of covariance
Binomial	*logit*	mixed	logistic regression
Poisson	logarithmic	mixed	log-linear

Table 6: Beetle mortality (Bliss, 1935).

i	x_i	n_i	y_i	i	x_i	n_i	y_i	i	x_i	n_i	y_i	i	x_i	n_i	y_i
1	1.6907	59	6	3	1.7552	62	18	5	1.8113	63	52	7	1.8610	62	61
2	1.7242	60	13	4	1.7842	56	28	6	1.8369	59	52	8	1.8839	60	60

Note that the function $F : I\!R \to [0,1]$, defined by $F(x) = \frac{\exp(x)}{1+\exp(x)}$, is a logistic distribution function. Thus Binomial GLM with logit canonical link function is known as *logistic model.*

Probit and complementary log-log models

If the relation between the probability π_i and the covariate vector $\mathbf{z}_i$ is

$$\pi_i = \Phi(\eta_i) = \Phi(\mathbf{z}_i^T \boldsymbol{\beta}), \tag{5.18}$$

where $\Phi(\cdot)$ is the $N(0,1)$ distribution function, one gets a probit link function $g(\mu_i) = \Phi^{-1}(\mu_i)$. The resulting GLM from the Binomial response model, with the probit link function leads to the *probit model.*

Another candidate for the inverse function of the link function is the Gumbel distribution function, $F(x)=1-\exp(-\exp(x))$, $x \in I\!R$. Considering then $h(\mathbf{z}_i^T\boldsymbol{\beta})=F(\mathbf{z}_i^T\boldsymbol{\beta})=\pi_i$, one gets the complementary log-log link function

$$\ln(-\ln(1-\pi_i)) = \mathbf{z}_i^T\boldsymbol{\beta} \tag{5.19}$$

for link function and its MLG is the *complementary log-log model.*

Example 5.4: Bliss (1935) studied the behavior of adult beetles on exposure to carbon disulfide gas (CS_2) for five hours observing 481 beetles divided into 8 groups.

Variables: n (number of beetles exposed), y (number of dead beetles) and x (dosage of $\log_{10} CS_2(mg/litro)$).

Objective: Estimate the dose-response curve for beetle mortality from different dosages.

Results: Estimated regression equations (DIC in (5.12)):

- Logistic: $\ln[\widehat{\pi}(x)/(1-\widehat{\pi}(x))] = -60.87 + 34.36\,x$.
- Probit: $\Phi^{-1}(\widehat{\pi}(x)) = -35.04 + 19.79\,x$.
- Complementary log-log: $\ln(-\ln(1-\widehat{\pi}(x))) = -39.73 + 22.13\,x$.

■

Table 7: Estimated proportions of dead beetles (Bliss, 1935).

dosage x	observed	logistic DIC=41.39	Probit DIC=40.31	clog-log DIC=33.60
1.6907	0.1017	0.0605	0.0585	0.0956
1.7242	0.2167	0.1658	0.1798	0.1884
1.7552	0.2903	0.3629	0.3794	0.3376
1.7842	0.5000	0.6054	0.6045	0.5416
1.8113	0.8254	0.7945	0.7878	0.7578
1.8369	0.8983	0.9021	0.9032	0.9165
1.8610	0.9839	0.9540	0.9614	0.9842
1.8839	1.0000	0.9781	0.9863	0.9986

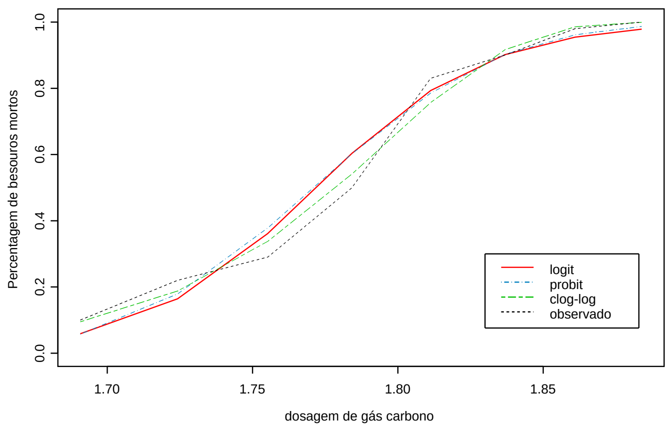


Figure 9: Plot of fitted proportions based on the three models.

Poisson model

To *count data e.g.* the number of telephone calls, the Poisson model plays a key role in its analysis.

Assuming the Y_i, $i=1,\ldots,n$, are independent variables, well modeled by a Poisson distribution of mean value $\mu_i = \exp(\mathbf{z}_i^T\boldsymbol{\beta})$ *i.e.*

$$f(y_i|\mathbf{x}_i) = \mathrm{e}^{-\mu_i}\mu_i^{y_i}/y_i! = \exp\{-e^{\mathbf{z}_i^T\boldsymbol{\beta}} + y_i\mathbf{z}_i^T\boldsymbol{\beta} - \ln y_i!\} \tag{5.20}$$

where $y_i = 0, 1, \ldots$, one gets a GLM with (*logarithmic* canonical link function, known as Poisson model or log-linear model.

Under certain conditions, the analysis of a *contingency table* under Poisson sampling is the same as the analysis under Multinomial or Multinomial product sampling.

There are three essential steps to modeling data through a GLM:

1. Model formulation:
   - Choice of distribution for the response variable.
   - Choice of covariates and formulation of design matrix.
   - Choice of link function.

2. Model fit. Estimation of model parameters: $\boldsymbol{\beta}$ regression coefficients and ϕ dispersion parameter. Goodness-of-fit tests.

3. Model selection and validation. A good model is one that can strike a balance among the three factors: adequacy, parsimony, and interpretation.

5.2.2 Reliability and survival analysis

- The study of survival or reliability analysis focuses on a set of units that are observed until the occurrence of some event of interest (*e.g.* death). Often the event does not occur for some of the units during the observation period (censoring).

- These units (individuals, electronic components, etc.) give rise to survival data that are essentially comprised of lifetime or *survival times* or unit failure times, *e.g.* the elapsed time from diagnosis of a particular disease in a patient to death due to that disease.

Example 5.5: Larynx cancer - In a Dutch hospital 90 male larynx cancer patients were diagnosed and treated during the period 1970 to 1978 (Kardaun, 1983).

Survival times observed in this study were the times elapsed between the first treatment of each patient and their death or the end of the study (03/01/1981).

For each patient, age at diagnosis, year of diagnosis and stage of disease were also observed. These stages are ordered from least severe (stage 1) to most severe (stage 4).

Table 8: Survival data of larynx cancer patients.

Stage	Lifetimes in years (* censored), Age, Year of diagnosis					
1	0.6,77,76	1.3,53,71	2.4,45,71	2.5*,57,78	3.2,58,74	3.2*,51,77
	3.3,76,74	3.3*,63,77	3.5,43,71	3.5,60,73	4.0,52,71	4.0,63,76
	4.3,86,74	4.5*,48,76	4.5*,68,76	5.3,81,72	5.5*,70,75	5.9*,58,75
	5.9*,47,75	6.0,75,73	6.1*,77,75	6.2*,64,75	6.4,77,72	6.5,67,70
	6.5*,79,74	6.7*,61,74	7.0*,66,74	7.4,68,71	7.4*,73,73	8.1*,56,73
	8.1*,73,73	9.6*,58,71	10.7*,68,70			
2	0.2,86,74	1.8,64,77	2.0,63,75	2.2*,71,78	2.6*,67,78	3.3*,51,77
	3.6,70,77	3.6*,72,77	4.0,81,71	4.3*,47,76	4.3*,64,76	5.0*,66,76
	6.2,74,72	7.0,62,73	7.5*,50,73	7.6*,53,73	9.3*,61,71	
3	0.3,49,72	0.3,71,76	0.5,57,74	0.7,79,77	0.8,82,74	1.0,49,76
	1.3,60,76	1.6,64,72	1.8,74,71	1.9,72,74	1.9,53,74	3.2,54,75
	3.5,81,74	3.7*,52,77	4.5*,66,76	4.8*,54,76	4.8*,63,76	5.0,59,73
	5.0,49,76	5.1*,69,76	6.3,70,72	6.4,65,72	6.5*,65,74	7.8,68,72
	8.0*,78,73	9.3*,69,71	10.1*,51,71			
4	0.1,65,72	0.3,71,76	0.4,76,77	0.8,65,76	0.8,78,77	1.0,41,77
	1.5,68,73	2.0,69,76	2.3,62,71	2.9*,74,78	3.6,71,75	3.8,84,74
	4.3*,48,76					

■

Basic concepts

One of the functions of interest in survival analysis for a unit population, whose survival time is represented by T, is the *survival function* which describes the distributional form of survival times through the probability of a unit. survive at least until the moment t,

$$S(t) \equiv P(T \geq t), \tag{5.21}$$

(a) Original scale

(b) Survival scale

Figure 10: Survival times (* censored unit).

where T is a nonnegative random variable (rv), where (5.21) is a non-increasing monotonous function of t in the time interval $\mathcal{J} = [0, \infty)$ such that $S(0) = 1$ and $S(\infty) \equiv \lim_{t\to\infty} S(t) = 0$.[9]

Survival models are usually formulated by the *hazard function* or *failure rate* or *force of mortality*), *i.e.* the occurrence rate of the event of interest at the moment $t \in \mathcal{J}$, defined by

$$\lambda(t) \;=\; \lim_{dt\to 0^+} \frac{P(t \le T < t + dt | T \ge t)}{dt}. \tag{5.22}$$

In Example 5.5, the function (5.22) represents the probability of an individual dying of larynx cancer in the infinitesimal range $[t, t + dt)$ since he lived so far just before t .

The *cumulative or integrated hazard function* (5.22) is given by $\Lambda(t) \;\equiv\; \int_0^t \lambda(u)du$ which is finite for some $t > 0$ and $\int_0^\infty \lambda(u)du = \infty$.

If the survival times are absolutely continuous, the hazard function completely determines the probability distributions of these (continuous) times by the following relationships:

$$\begin{aligned} \lambda(t) &= \tfrac{f(t)}{S(t)} = -\tfrac{d}{dt} \ln S(t) \\ S(t) &= \exp\left[-\int_0^t \lambda(u)du\right] = \exp[-\Lambda(t)], \end{aligned} \tag{5.23}$$

where $f(\cdot)$, $S(\cdot)$ and $\Lambda(\cdot)$ are respectively the probability density (pdf) function, the survival function, and the cumulative hazard function of generic survival time T.

Likelihood function construction

For homogeneous populations, $\mathcal{D} = \{(t_i, \gamma_i), i=1, \ldots, n\}$, while for *heterogeneous populations*

$$\mathcal{D} \;=\; \{(t_i, \gamma_i, \mathbf{z}_i), i=1, \ldots, n\},$$

where t_i is the observed value of $T_i = \min(X_i, C_i)$ and $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})^T$ is the covariate or function vector covariates associated with the i unit, for simplicity $\mathbf{z}_i$ with p covariates,$i=1, \ldots, n$.

In this scenario, survival times from a population with f.d.p. $f(\cdot)$ and $S(\cdot)$ survival function are continuous and subject to right censorship and C_i, $i=1, \ldots, n$ censorship times, are not random..

For independent random pairs from a homogeneous population with $f(\cdot)$ and $S(\cdot)$ dependent on a parameter vector $\boldsymbol{\theta}$ and non-informative censoring mechanism for $\boldsymbol{\theta}$, the *likelihood function* of $\boldsymbol{\theta}$ given the

[9]A consequence of this survival function definition is that its distribution function $F(t) \equiv 1 - S(t)$, $\forall\, t \in \mathcal{J}$, becomes continuous left, contrary to the usual definition of distribution functions.

set $\mathcal{D}$ is

$$L(\boldsymbol{\theta}|\mathcal{D}) \;=\; \prod_{i=1}^{n} f(t_i|\boldsymbol{\theta})^{\gamma_i} S(t_i|\boldsymbol{\theta})^{1-\gamma_i}. \tag{5.24}$$

Using the relations in (5.23), the likelihood function (5.24) can be expressed equally in terms of the hazard function $\lambda(t|\mathbf{z}, \boldsymbol{\theta})$, dependent on the parametric vector $\boldsymbol{\theta}$ and the covariate vector $\mathbf{z}_i$ of the i-th unit, *i.e.*

$$L(\boldsymbol{\theta}|\mathcal{D}) \;=\; \prod_{i=1}^{n} \lambda(t_i|\mathbf{z}_i, \boldsymbol{\theta})^{\gamma_i} \exp\left[-\int_0^{t_i} \lambda(u|\mathbf{z}_i, \boldsymbol{\theta}) du\right]. \tag{5.25}$$

Parametric survival models

Parametric survival models are constructed from families of specific probability distributions, *e.g.* :

- Exponential distribution;
- Gamma distribution;
- Log-normal distribution;
- Weibull distribution.

Choosing the survival distribution can always resort to the properties of the various univariate distributions that best fit the concrete study of survival (Lawless, 2003).

Weibull regression

In the face of heterogeneous populations, the likelihood functions (5.24) and (5.25) are redefined by introducing in their hazard functions a function of the covariates $\psi(\mathbf{z})$.

For example, the *Weibull regression model* is defined by the following conditional hazard function at $\mathbf{z}$ reparameterizing your scale parameter δ as $\delta\,\psi(\mathbf{z})$:

$$\lambda(t|\mathbf{z}) \;=\; \delta \alpha t^{\alpha-1} \psi(\mathbf{z}) \;, \quad t \geq 0, \tag{5.26}$$

where $\psi(\mathbf{z}) = \exp[\mathbf{z}^T \boldsymbol{\beta}]$ and $\boldsymbol{\beta}$ is the regression coefficient vector associated with the $\mathbf{z}$ covariate vector. (5.26) is known as accelerated lifetime model because of its scale parameter.

A feature of the model (5.26) is that the ratio of two-unit hazard functions with covariate vector $\mathbf{z}_1$ and $\mathbf{z}_2$,

$$\frac{\lambda(t|\mathbf{z}_1)}{\lambda(t|\mathbf{z}_2)} \;=\; \frac{\psi(\mathbf{z}_1)}{\psi(\mathbf{z}_2)} \tag{5.27}$$

does not depend on t, *i.e.* different units have proportional hazard functions (*proportional hazard models*).

The Weibull regression model is the only survival model that is both a location-scale (Lawless, 2003) and proportional hazards (5.27) model.

The inferential methods in these parametric models are obtained based on the likelihood function (5.25) using the respective hazard function (Lawless, 2003).

Example 5.6: Disease stage lifetimes of Example 5.5.
Results: Based on 5000 samples (*burn-in* = 5000), the following parameter estimates were obtained:

Table 9: Buyer frequencies by gender and model.

gender	model 1	model 2	model 3	total
male	160	140	40	340
female	40	60	60	160
total	200	200	100	500

	mean	sd	2.5%	50%	97.5%
alpha	1.16	0.1361	0.90	1.16	1.44
beta[1]	-2.76	0.3531	-3.42	-2.78	-2.06
beta[2]	-2.86	0.4391	-3.80	-2.86	-2.01
beta[3]	-2.08	0.3242	-2.76	-2.08	-1.41
beta[4]	-0.99	0.3370	-1.69	-0.98	-0.35
mediana[1]	8.13	2.0242	5.14	7.78	12.90
mediana[2]	9.24	3.9023	4.60	8.50	18.34
mediana[3]	4.47	0.9700	2.91	4.38	6.74
mediana[4]	1.77	0.5094	0.99	1.69	2.95
contra1	-0.10	0.4683	-1.03	-0.08	0.83
contra2	0.68	0.3623	-0.01	0.68	1.39
contra3	1.77	0.4184	0.93	1.77	2.57
deviance	288.20	3.1145	284.10	287.60	296.30

Simulated chains evaluation via Geweke, Heidelberg-Welch and Raftery-Lewis convergence diagnostic techniques:

```
# Fraction in 1st window = 0.1    # Fraction in 2nd window = 0.5
#   alpha     beta[1]            beta[3]
#   -0.5724   -0.1426            1.6625
-------------------------------------------------------------------
#             Stationarity test  start iteration  p-value
#   alpha     passed             1                0.256
#   beta[1]   passed             1                0.807
#   beta[3]   passed             1                0.301
#             Halfwidth test     Mean             Halfwidth
#   alpha     passed             1.16             0.00377
#   beta[1]   passed             -2.76            0.00979
#   beta[3]   passed             -2.08            0.00865
-------------------------------------------------------------------
# Quantile (q) = 0.025    # Accuracy (r) = +/- 0.005    # Probability (s) = 0.95
#             Burn-in   Total   Lower bound   Dependence
#             (M)       (N)     (Nmin)        factor (I)
#   alpha     2         3930    3746          1.050
#   beta[1]   2         3829    3746          1.020
#   beta[3]   2         3680    3746          0.982
```

■

5.2.3 Analysis of categorical data

Example 5.7: One car manufacturer suspects that the sale of its last three models is related to the gender of its buyers. Based on the following contingency table involving 500 buyers, test the hypothesis of independence between the type of car models and the gender of buyers.

Suppose that each of the n sampled elements of a population can be classified according to two characteristics X and Y, with r and s categories, respectively.

Let $p_{ij} = P(X=i, Y=j)$ be the (joint) probability that an element of the population belongs to the (i,j) category of (X,Y), $i=1,\dots,r$, $j=1,\dots,s$.

Consequently, the (marginal) probabilities of the two characteristics are given by

$$\begin{aligned} p_{i\bullet} &= P(X=i) = \textstyle\sum_{j=1}^{s} P(X=i, Y=j) \\ p_{\bullet j} &= P(Y=j) = \textstyle\sum_{i=1}^{r} P(X=i, Y=j). \end{aligned}$$

- Hypothesis of interest:

$$H_0 : p_{ij} = p_{i\,\bullet} \times p_{\bullet\,j}, \forall\, i, j.$$

■

Multinomial distribution

Let $n = (n_1, \ldots, n_c)$ be a random vector whose components take nonnegative integer values such that $1_c' n = N$ is fixed. If the probability function of any subset of $c-1$ components of n is

$$f(n|N, \theta) = N! \prod_{i=1}^{c} \theta_i^{n_i}/n_i!$$

where $\theta = (\theta_1, \ldots, \theta_c)'$ with $\theta_i > 0$ and $1_c'\theta = 1$, n is said to have a Multinomial distribution of parameters N and θ — write $n|N, \theta \sim M_{c-1}(N, \theta)$ highlighting the $c-1$ dimensionality of this distribution to counter notational abuse (practiced for convenience) to apply it to n instead of its $(c-1) \times 1$ subvector, *e.g.*, to $\bar{n} = (n_1, \ldots, n_{c-1})'$.

The moment generating function of the Multinomial distribution under analysis is

$$M_n(t) \equiv E(e^{t'n}|N, \theta) = \left(\sum_{j=1}^{c-1} \theta_j e^{t_j} + \theta_c \right)^N$$

where $t = (t_1, \ldots, t_{c-1}, 0)'$, which concludes in particular that the first two moments of the vector n are given by

$$E(n|N, \theta) = N\theta \text{ e } Var(n|N, \theta) = N(D_\theta - \theta\theta')$$

where $D_\theta = \text{diag}(\theta_1, \ldots, \theta_c)$. When the distribution in question is considered to be reported at $\bar{n}$, the respective mean vector and covariance matrix are

$$\mu = N\bar{\theta} \text{ e } \Sigma = N(D_{\bar{\theta}} - \bar{\theta}\bar{\theta}')$$

where $\bar{\theta} = (\theta_1, \ldots, \theta_{c-1})'$.

The conditional distribution of n given $M \equiv (M_1, \ldots, M_s) = Z'n$, where $M_k = z_k' n = \sum_{i \in C_k} n_i, \quad k = 1, \ldots, s$, is then the product of s Multinomial distributions for $n^{(k)} = (n_i, i \in C_k)'$

$$n^{(k)}|M, \theta, \; k = 1, \ldots, s \underset{ind.}{\sim} M_{d_k - 1}(M_k, \pi_k)$$

where $\pi_k = (\theta_i/\alpha_k, \; i \in C_k)'$. For example, in a two-dimensional contingency table $I \times J$ the conditional distribution of $n = (n_{ij})$ given $M = (n_{i\cdot}, \; i = 1, \ldots, I)'$ is the product of I Multinomials

$$(n_{ij}, \; j = 1, \ldots, J)'|n_{i\cdot}, \theta, \; i = 1, \ldots, I \underset{ind.}{\sim} M_{J-1}(n_{i\cdot}, \pi_i)$$

where $\pi_i = (\theta_{(i)j}, j = 1, \ldots, J)'$, $\theta_{(i)j} = \theta_{ij}/\theta_{i\cdot}$.

Dirichlet distribution

Θ is said to have a Dirichlet distribution of parameter $a = (a_1, \ldots, a_c) \in I\!R_+^c$, which is symbolically indicated by $\theta \sim D_{c-1}\ (a)$, if its S_{c-1} density function is expressed (again in oversized notation, for convenience) by

$$h(\theta|a) = [B(a)]^{-1} \prod_{i=1}^{c} \theta_i^{a_i - 1}$$

where

$$B(a) = \int\limits_{S_{c-1}} \prod_{i=1}^{c} \theta_i^{a_i-1} d\theta = \frac{\prod_{i=1}^{c} \Gamma(a_i)}{\Gamma(a.)}$$

with $a. = \sum_i a_i$, is the multivariate beta function (Dirichlet integral $(c-1)$-dimensional).

This distribution family is the multivariate extension of the Beta family that is reduced when $c = 2$. It is easy to show that θ moments can be obtained from

$$E\left[\prod_{i=1}^{c} \theta_i^{r_i} |a\right] = \frac{B(a+r)}{B(a)}$$

where $r = (r_1, \ldots, r_c)'$. In particular,

$$E(\theta_k|a) = a_k/a.$$

$$\text{Var}(\theta_k|a) = a_k(a. - a_k)/\{a_.^2(a. + 1)\}$$

$$\text{Cov}(\theta_k, \theta_l|a) = -a_k a_l/\{a_.^2(a. + 1)\}$$

for $k,\ l = 1, \ldots, c,\ l \neq k$.

Multinomial regression

Each combination of explanatory factors is supposed to give rise to a multinomial response with a logistic link, so that for lake i, length j, the observed vector of counts $X_{ij\cdot}$ has multinomial distribution *i.e.*

$$X_{ij\cdot} = (X_{ij1}, \ldots, X_{ij5}) \sim \text{Multinomial}(n_{ij}, p_{ij\cdot})$$

where $n_{ij} = \sum_k X_{ijk}$ and $p_{ij\cdot} = (p_{ij1}, \ldots, p_{ijk})$, $i = 1, \ldots, 4,\ j = 1, 2,\ k = 1, \ldots, 5$.To use the *logit* link function, consider repairing $p_{ijk} = \phi_{ijk}/\sum_k \phi_{ijk}$, with

$$\log \phi_{ijk} = \alpha_k + \beta_{ik} + \gamma_{jk}, \tag{5.28}$$

where $\alpha_1, \beta_{i1}, \beta_{1k}, \gamma_{j1}, \gamma_{1k} = 0$ for model identifiability, which is known as *multinomial logistic regression model.*

Example 5.8: Agresti (2019) analyzes a food choice data set of 221 alligators, where the response measure for each alligator is one of 5 categories. Possible explanatory factors are alligator length (2 categories) and lake (4 categories).

Results: Based on 5000 samples, after a burn-in period of 5000, the following parameter estimates were obtained.

■

Table 10: Choice of alligator primary foods (Agresti, 2019).

Lake	Length	fish	invertebrate	reptile	bird	other
Hancock	<= 2.3	23	4	2	2	8
	> 2.3	7	0	1	3	5
Oklawaha	<= 2.3	5	11	1	0	3
	> 2.3	13	8	6	1	0
Trafford	<= 2.3	5	11	2	1	5
	> 2.3	8	7	6	3	5
George	<= 2.3	16	19	1	2	3
	> 2.3	17	1	0	1	3

Table 11: Estimated proportions based on the model (5.22).

Lake	Length	fish	invertebrate	reptile	bird	other
Hancock	<= 2.3	0.5309	0.0830	0.0565	0.0706	0.2590
	> 2.3	0.5519	0.0198	0.0836	0.1435	0.2011
Oklawaha	<= 2.3	0.2580	0.6008	0.0794	0.0095	0.0523
	> 2.3	0.4649	0.2487	0.1894	0.0287	0.0684
Trafford	<= 2.3	0.1857	0.5179	0.0916	0.0367	0.1680
	> 2.3	0.3014	0.1942	0.1998	0.1047	0.2000
George	<= 2.3	0.4540	0.4128	0.0124	0.0292	0.0916
	> 2.3	0.6592	0.1381	0.0250	0.0787	0.0990

6 Resampling Methods

Resampling methods treat an observed sample as a finite population, and random samples are generated (resampled) from it to estimate population characteristics and make inferences about the sampled population.

Although subsampling, resampling, or otherwise rearranging a given dataset cannot increase the information content of the dataset, these procedures can sometimes be useful in extracting information.

- *Jackknife methods* are resampling methods for estimating bias, and standard error.
- *Bootstrap methods* are nonparametric Monte Carlo methods that estimate the distribution of a population by resampling.

6.1 Jackknife methods

Jackknife methods make use of systematic partitions of a dataset to estimate properties of an estimator computed from the full sample. Quenouille (1949, 1956) suggested the technique to estimate the bias of an estimator, while John Tukey used the term '*jackknife*' to refer to the method, and showed that the method is also useful in estimating the variance of an estimator (see *e.g.* Gentle, 2002; Rizzo, 2019).

The jackknife is like a 'leave-one-out' (LOO) type of cross-validation.

Let $\mathbf{x} = (x_1, \ldots, x_n)$ be an observed random sample, and define the i-th jackknife sample $\mathbf{x}_{(-i)}$ to be the subset of $\mathbf{x}$ that leaves out the i-th observation x_i. That is, $\mathbf{x}_{(-i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. If $T_n(\mathbf{x})$ is a statistics based on all sample, define the its i-th jackknife replicate as $T_{n-1}(\mathbf{x}_{(-i)})$, $i = 1, \ldots, n$.

Suppose the parameter $\theta = t(F)$ *i.e.* , θ is a function of the cumulative distribution function F and F_n denotes its empirical function based on a random sample from F. The 'plug-in' estimate of θ is $\hat{\theta} = t(F_n)$ that is 'smooth' in the sense that small changes in the data correspond to small changes in $\hat{\theta}$. For example, the sample mean is a plug-in estimate for the population mean, contrary the sample median to the population median.

For the estimation of θ, if $T = t(F_n(\mathbf{x}))$ is a 'smooth' (plug-in) statistic, then $T_{(-i)} = t(F_n(\mathbf{x}_{(-i)}))$, and the *jackknife estimate of bias* is defined as

$$\widehat{B_J} = (n-1)(\bar{T}_{(\cdot)} - T), \tag{6.1}$$

where $\bar{T}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} T_{(-i)}$ is the mean of the estimates from the leave-one-out samples, and T is the estimate computed from the original observed sample.

Example 6.1: Consider $\theta = Var(X) \equiv \sigma^2$ in order to illustrate the presence of the *factor* $n-1$ in (6.1). If $x_1, \ldots, x_n$ is a random sample from the distribution of X, the plug-in estimate of the variance of X is $T = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. The estimator T is biased for σ^2 with

$$B(T) = E(T - \sigma^2) = \cdots = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}.$$

Each jackknife replicate computes the estimate $T_{(-i)}$ on a sample size $n-1$, so that the bias in the jackknife replicate is $-\frac{\sigma^2}{n-1}$. Thus, for $i = 1, \ldots, n$ we have

$$E(T_{(-i)} - T) = B(T_{(-i)}) - B(T) = -\frac{\sigma^2}{n-1} - (-\frac{\sigma^2}{n}) = \frac{B(T)}{n-1}.$$

Thus, the jackknife estimate (6.1) with factor $(n-1)$ gives the correct estimate of bias in the plug-in estimator of variance (Rizzo, 2019). ■

A *jackknife estimate of standard error* (Tukey, 1958) is defined as

$$\widehat{SE}_J = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}(T_{(-i)} - \bar{T}_{(\cdot)})^2}, \tag{6.2}$$

for 'smooth' statistics T.

Example 6.2: Consider $\theta = E(X)$ *i.e.* the population mean in order to illustrate the presence of the factor $\frac{n-1}{n}$ in (6.2). A estimator of θ is $T = \bar{X}$ and the standard error of the mean of X is $\sqrt{Var(X)/n}$. So, under this factor, $\widehat{SE}_J$ is an 'unbiased' plug-in estimator of $\sqrt{Var(X)/n}$, since

$$E\big(\frac{n-1}{n}\sum_{i=1}^{n}(\bar{X}_{(-i)} - \bar{\bar{X}}_{(\cdot)})^2\big) = E\big(\frac{n-1}{n}\frac{S^2}{n-1}\big) = \frac{Var(X)}{n},$$

where $\bar{X}_{(-i)} = \frac{\sum_{j\neq i=1}^{n} X_j}{n-1}$, $\bar{\bar{X}}_{(\cdot)} = \frac{\sum_{i=1}^{n}\bar{X}_{(-i)}}{n}$ and $S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$. ■

The generalized jackknife

Schucany *et al.* (1971) suggested a method of systematically reducing the bias by combining higher order jackknifes (*generalized jackknife*). First consider two biased estimators of θ, T_1 and T_2. Let $w = \frac{B(T_1)}{B(T_2)}$, where $B(T_j)$ is the bias of estimation of T_j, $j = 1, 2$.

Now consider the weighted combination of the estimators *i.e.*

$$T_w = (T_1 - w\,T_2)/(1 - w), \tag{6.3}$$

which is an unbiased estimator of θ since

$$E(T_w) = \frac{1}{1-w}(\theta + B(T_1)) - \frac{w}{1-w}(\theta + B(T_2)) = \theta.$$

Notice if $w = \frac{n-1}{n}$, then the *jackknife estimator* T_J is unbiased, where

$$T_J = nT - (n-1)\bar{T}_{(\cdot)}. \tag{6.4}$$

- The jackknife estimator (6.4) can also be write as $T_J = \frac{i}{n}\sum_{i=1}^{n} T_i^{\star}$, where $T_i^{\star} = nT - (n-1)T_{(-i)}$ are called as *pseudo-values* of the jackknife.
- The jackknife can fail when the statistic T is not '*smooth*'. The median is an example of a statistic that is not smooth. Because in leaving out one observation at a time, the median of the reduced samples will only take on at most two different values, the jackknife procedure cannot lead to a good estimate of the variance.
- When the statistic is not smooth, the *delete-d jackknife* (leave d observations out on each replicate) can be applied (see Efron and Tibshirani, 1993). If $\sqrt{n}/d \to 0$ and $n - d \to \infty$, then the delete-d jackknife is consistent for the median. The computing time increases because there are a large number of jackknife replicates when n and d are large.

6.2 Bootstrap methods

The *bootstrap* was introduced by Efron (1979), with further developments in Efron (1981), and numerous other publications including the book of Efron and Tibshirani (1993).

Bootstrap methods are a class of nonparametric methods that estimate the distribution of a population by resampling. The term 'bootstrap' can refer to nonparametric bootstrap or parametric bootstrap (see Monte Carlo methods). In the former, the distribution is not specified.

A basic idea in *bootstrap resampling* is that, because the observed sample contains all the available information about the underlying population, the observed sample can be considered the 'population'. Hence, the distribution of any relevant test statistic can be simulated by use of random samples from the 'population' consisting of the original sample (Gentle, 2002).

Definition 6.1: A *bootstrap sample* $\mathbf{x}^\star = (x_1^\star, \dots, x_n^\star)$ is obtained by randomly sampling n times, with replacement, from the original observed random sample $\mathbf{x} = (x_1, \dots, x_n)$. Resampling generates a random sample $\mathbf{X}^\star = (X_1^\star, \dots, X_n^\star)$ by sampling with replacement from $\mathbf{x}$. The random variables $X_i^\star$ are i.i.d., uniformly distributed on the set $\{x_1, \dots, x_n\}$.

The empirical cumulative distribution function (*ecdf*) $F_n(x)$ is an estimator of the cumulative distribution function (*cdf*) $F(x)$. $F_n(x)$ is itself the cdf of a random variable; namely the random variable that is uniformly distributed on the set $\{x_1, \dots, x_n\}$. Hence the ecdf F_n is the cdf of $\mathbf{X}^\star$. Thus in bootstrap, there are two approximations:

1. The ecdf F_n is an approximation to the cdf F_X.
2. The ecdf $F_n^\star$ of the bootstrap replicates is an approximation to the ecdf F_n.

Example 6.3: Consider the observed sample $x = 2, 2, 1, 1, 5, 4, 4, 3, 1, 2$. Resampling from x, the cdf $F_{X^\star}$ of a randomly selected replicate is exactly the ecdf $F_n(x)$ *i.e.*

$$F_{X^\star} = F_n(x) = \begin{cases} 0, & x < 1, \\ 0.3 & 1 \leq x < 2, \\ 0.6 & 2 \leq x < 3, \\ 0.7 & 3 \leq x < 4, \\ 0.9 & 4 \leq x < 5, \\ 1 & x \geq 5. \end{cases}$$

If F_n is not close to F_X then the distribution of the replicates will not be close to F_X. Resampling from x a large number of replicates produces a good estimate of F_n but not a good estimate of F_X. Here x is a Poisson(2) sample and the bootstrap samples will never include 0. ■

To generate a *bootstrap random sample* by resampling $\mathbf{x}$, generate n random integers $\{i_1, \dots, i_n\}$ uniformly distributed on $\{1, \dots, n\}$ and select the bootstrap sample $\mathbf{x}^\star = (x_{i_1}, \dots, x_{i_n})$.

Suppose θ is the parameter of interest, and T is an estimator of θ. Then the bootstrap estimate of the distribution of T is obtained as follows (Rizzo, 2019).

1. For each bootstrap replicate, indexed $a = 1, \dots, A$:
   (a) Generate sample $\mathbf{x}^{\star(a)} = (x_1^\star, \dots, x_n^\star)$ by sampling with replacement from the observed sample $\mathbf{x} = (x_1, \dots, x_n)$.
   (b) Compute the a-th replicate $T^{(a)}$ from the a-th bootstrap sample.
2. The bootstrap estimate of $F_T(\cdot)$ is the empirical distribution of the replicates $T^{(1)}, \dots, T^{(A)}$.

The *bootstrap estimate of standard error* of an estimator T of the parameter θ is the sample standard deviation of the bootstrap replicates $T^{(1)}, \ldots, T^{(A)}$ *i.e.*

$$\widehat{SE}_B = \sqrt{\frac{1}{A-1}\sum_{a=1}^{A}(T^{(a)} - \bar{T}^{(\cdot)})^2}, \tag{6.5}$$

where $\bar{T}^{(\cdot)} = \frac{1}{A}\sum_{a=1}^{A} T^{(a)}$.

According to Efron and Tibshirani (1993), the number of replicates needed for good estimates of standard error is not large; $A = 50$ is usually large enough, and rarely is $A > 200$ necessary. (Much larger A will be needed for confidence interval estimation.)

The *bootstrap estimation of bias* uses the bootstrap replicates of T to estimate the sampling distribution of T. For the finite population $\mathbf{x} = (x_1, \ldots, x_n)$, the 'parameter' is $T(\mathbf{x})$ and there are A independent and identically distributed estimators $T^{(a)}$. The sample mean of the replicates $\{T^{(1)}, \ldots, T^{(A)}\}$ is unbiased for its expected value $E(T^{(\cdot)})$, so the bootstrap estimate of bias is

$$\widehat{B_B} = (\bar{T}^{(\cdot)} - T), \tag{6.6}$$

where $\bar{T}^{(\cdot)} = \frac{1}{A}\sum_{a=1}^{A} T^{(a)}$, and $T = T(\mathbf{x})$ is the estimate computed from the original observed sample.

Note that, in bootstrap, F_n is sampled in place of F_X, so we replace θ with T to estimate the bias. Positive bias indicates that T on average tends to overestimate θ.

Example 6.4: The law school dataset from Efron and Tibshirani (1993) contains LSAT (average score on law school admission test score) and GPA (undergraduate grade-point average) for 15 law schools.

LSAT	576	635	558	578	666	580	555	661	651	605	653	575	545	572	594
GPA	339	330	281	303	344	307	300	343	336	313	312	274	276	288	296

Correlation between LSAT and GPA scores is defined by

$$\theta = Cov(LSAT, GPA)/\sqrt{Var(LSAT)\,Var(GPA)}.$$

- θ is empirically estimated by 0.7763745.
- The bootstrap estimate of the standard error of the sample correlation ($A = 200$ replicates) is $\widehat{SE}_B = 0.1358393$.
- The histogram of the replicates of T is shown in Figure 11.

■

Jackknife-after-Bootstrap

For getting the variance of bootstrap estimates of standard error and bias, which are random variables, we can try the jackknife-after-bootstrap computing an estimate for each 'leave-one-out' sample. Let $B(i)$ denote number of bootstrap samples that do not contain x_i and $J(i)$ the corresponding indices. Then compute the jackknife replication leaving out the $B - B(i)$ samples that contain x_i (Efron and Tibshirani, 1993).

The jackknife-after-bootstrap estimate of standard error is computed by

$$\widehat{SE}_{JaB} = \widehat{SE}_J(\widehat{SE}_{B(1)}, \ldots, \widehat{SE}_{B(n)}), \tag{6.7}$$

where $\widehat{SE}_J$ is calculated by (6.2), $\widehat{SE}_{B(i)} = \sqrt{\frac{1}{B(i)}\sum_{j\in J(i)}[T^{(j)} - \bar{T}^{(J(i))}]^2}$ and $\bar{T}^{(J(i))} = \frac{1}{B(i)}\sum_{j\in J(i)} T^{(j)}$ that is the sample mean of the estimates from the leave-x_i-out jackknife samples.

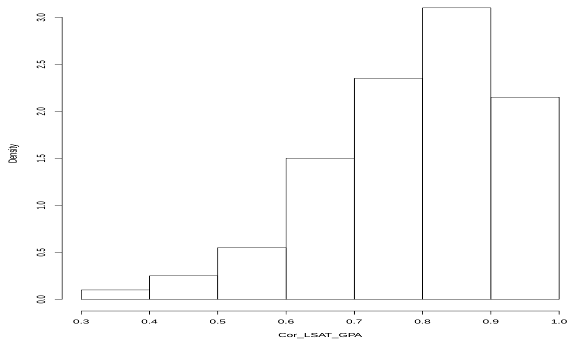


Figure 11: Bootstrap replicates for law school data in Example 6.4.

Bootstrap confidence intervals

A method of forming a confidence interval for a parameter θ is to find a pivotal quantity that involves θ and a statistic T, $f(T,\theta)$, and then to rearrange the terms in a probability statement of the form

$$P(q_{\frac{\alpha}{2}} \leq f(T,\theta) \leq q_{1-\frac{\alpha}{2}}) = 1-\alpha. \tag{6.8}$$

For the case $f(T,\theta) = T-\theta$ em (6.8), the $(1-\alpha)\%$ *basic bootstrap confidence interval* is defined by

$$P(T - q_{1-\frac{\alpha}{2}} \leq \theta \leq T - q_{\frac{\alpha}{2}}) = 1-\alpha. \tag{6.9}$$

where the quantiles $q_{\frac{\alpha}{2}}$ and $q_{1-\frac{\alpha}{2}}$ may be estimated from the ecdf of the bootstrap replicates $T^{(\cdot)}$. Gentle (2002) suggested their estimation from the quantiles of a Monte Carlo sample of $T^\star - T_0$, where $T^\star$ and T_0 are the values of T in the bootstrap sample and a given sample, respectively.

Methods of inference based on a Normal distribution often work well even when the underlying distribution is not Normal. A useful approximate confidence interval for a location parameter can often be constructed using the confidence interval for the mean of a Normal distribution.

A confidence interval for any parameter constructed in this pattern is called a *bootstrap-t interval* that has the form

$$(T - \hat{q}^t_{1-\frac{\alpha}{2}}\sqrt{\widehat{V}(T)}, T - \hat{q}^t_{\frac{\alpha}{2}}\sqrt{\widehat{V}(T)}). \tag{6.10}$$

where $\hat{q}^t_{(\cdot)}$ is the estimated quantile from the studentized statistic, $T = (T^\star - T_0)/\sqrt{\widehat{V}(T^\star)}$.

The variance $V(T)$ is often not available for many estimators T. But it could be estimated using a bootstrap sample *i.e.* the sample variance of $T^\star$ based on the m samples of size n taken from F_n.

A *bootstrap percentile confidence interval* uses the empirical distribution of the bootstrap replicates as the reference distribution. The quantiles of the empirical distribution are estimators of the quantiles of the sampling distribution of T.

From the ecdf of the bootstrap replicates, compute the $\frac{\alpha}{2}$ quantile $q^\star_{\frac{\alpha}{2}}$, and the $1-\frac{\alpha}{2}$ quantile $q^\star_{1-\frac{\alpha}{2}}$. In practice we generally use Monte Carlo and m bootstrap samples to estimate these quantities. The $(1-\alpha)100\%$ (equal-tailed) bootstrap percentile confidence interval is thus

$$(q^\star_{1-\frac{\alpha}{2}}, q^\star_{\frac{\alpha}{2}}). \tag{6.11}$$

where $q^\star_\alpha$ is the $[\alpha m]$-th order statistic of a m-size sample of T.

Efron and Tibshirani (1993) show that the percentile interval has some theoretical advantages over the standard normal interval and somewhat better coverage performance.

Example 6.5 (*vide* Example 6.3): Using R package *boot*, we can compute 95% bootstrap confidence interval estimates for the correlation statistic in the *law* data (see output below for $A = 2000$ replicates).

...
Intervals :

Level	Normal	Basic	Percentile
95%	(0.5182, 1.0448)	(0.5916, 1.0994)	(0.4534, 0.9611)

Calculations and Intervals on Original Scale

All three intervals cover the correlation $\rho = 0.76$ of the universe of all law schools (*law82* data). One reason for the difference in the percentile and normal confidence intervals could be that the sampling distribution of correlation statistic is not close to Normal (see Figure 11). ■

Appendix and Bibliography

Appendix A: Probability and Statistics Concepts

Notion of probability

1. *Laplace interpretation*: For a random experiment E with finite sample space $\Omega = \{1, \ldots, N\}$, assuming the N results are equally likely, the probability of any A event is the ratio of Ω results favorable to A.

2. *Frequentist interpretation*: The probability of an event A is the limit of the relative frequency of A occurring in a long succession of experiments conducted under the same conditions.

3. *Subjectivist interpretation*: The probability of an event A is understood as a personal measure (between 0 and 1) of the degree of belief about the occurrence of A.

Bayes' theorem

Definition A.1: The events $A_1, \ldots, A_n$ form a *partition* of the sample space Ω when i) $A_i \cap A_j = \emptyset$, $\forall i \neq j = 1, \ldots, n$; ii) $\cup_{i=1}^{n} A_i = \Omega$.

Theorem A.1: If the events $A_1, \ldots, A_n$ form a partition of the sample space Ω and B is any event of Ω with $P(B) > 0$, then $\forall\, i = 1, \ldots, n$, (*Bayes' theorem*)

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}.$$

Definition A.2: Two events A and B of the same sample space Ω are said to be *independent* if

$$P(A \cap B) = P(A) \times P(B).$$

Random variables

Definition A.3: A *random variable* (r.v.) X is a function that associates a real number with each possible result of a random experiment.

- The r.v.'s can assume a finite or infinite number (numerable/discrete or non-numerable/continuous) of possible values.

- The probabilistic model induced in $I\!R$ by X can be fully defined in various ways, *e.g.*, through the distribution function.

Definition A.4: Given a random variable X, the (cumulative) *distribution function* of X is given by

$$F_X(x) \equiv P(X \leq x), \forall\, x \in I\!R.$$

Discrete and continuous random variables

Definition A.5: X is said to be a *discrete* r.v., with possible values $x_1, x_2, \ldots$, if there is a function ($\mathbb{R} \to [0,1]$) $f_X(x) = P(X = x)$, denoting the probability of occurrence of $\{x\}$, known as the probability mass function (*p.m.f.*), and satisfying the conditions: i) $f_X(x_i) > 0$, $\forall\, i = 1, 2, \ldots$; ii) $\sum_{i \geq 1} f_X(x_i) = 1$.

Definition A.6: X is said to be a *continuous* r.v., if there is a function f_X, called the probability density function (*p.d.f.*) such that: i) $f_X(x) \geq 0$, $\forall\, x \in \mathbb{R}$; ii) $\int_{\mathbb{R}} f_X(x)dx = 1$; iii) The distribution function is continuous and given by

$$F_X(x) \equiv P(X \leq x) = \int_{-\infty}^{x} f_X(u)du.$$

Random variable functions

If X is a *discrete* r.v. with p.m.f. $f_X(x)$ and codomain $D = \{x_1, x_2, \ldots\}$, then $Y = g(X)$ is also a discrete r.v. with p.m.f.

$$f_Y(y) = P(Y = y) = P(X \in A_y) = \sum_{x_i \in A_y} f_X(x_i),\ y \in D^*$$

where $A_y = \{x \in D : g(x) = y\}$ and $D^* = g(D)$ is the codomain of Y.

If X is a *continuous* r.v., the continuity of $Y = g(X)$ depends on the type of the function $g(\cdot)$. For example, for X with f.d.p. $f_X(x) = 1$, if $0 < x < 1$, 0, otherwise, f.d.p. of $Y = \mathrm{e}^X$ é $f_Y(y) = \frac{1}{y}$, if $0 < y < \mathrm{e}$, 0, otherwise, considering that $\forall y > 0$,

$$F_Y(y) \equiv P(Y \leq y) = P(X \leq \log y) \equiv F_X(\log y) \to f_Y(y) = f_X(\log y)\frac{1}{y}.$$

Discrete and continuous random vectors

Definition A.7: Let $(X_1, \ldots, X_n) \in \mathbb{R}^n$ be a random vector, where X_i, $1 \leq i \leq n$ are discrete and/or continuous random variables. $(X_1, \ldots, X_n)$ is said to be a *discrete* or *continuous* random vector with distribution function $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$, when there is a nonnegative function $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ checking, resp.,

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \sum_{u_1 \leq x_1} \cdots \sum_{u_n \leq x_n} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n)$$
$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n)du_1 \ldots du_n$$

$$\therefore \quad \sum_{u_1 \leq \infty} \cdots \sum_{u_n \leq \infty} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n) = 1$$
$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n)du_1 \ldots du_n = 1.$$

Independent random variables

Definition A.8: $X_1, \ldots, X_n$ are v.a. *independent*, if the distribution function of $(X_1, \ldots, X_n)$ is given by

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \equiv P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} F_{X_i}(x_i),$$

where $F_{X_i}(x_i)$ is the marginal distribution function of X_i, $i=1,\ldots,n$.
or equivalently, if the joint p.m.f. (p.d.f.) of $X_1,\ldots,X_n$

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=\prod_{i=1}^{n} f_{X_i}(x_i),$$

where $f_{X_i}(x_i)$ is the marginal p.m.f. (p.d.f.) of X_i, $i=1,\ldots,n$.

Expected value and variance of a random variable

Definition A.9: Given a discrete (continuous) r.v. X with p.m.f. (p.d.f.) $f_X(x)$, the *expected value* (or mean value or mathematical expectation) of X, if any, is given by

$$E(X)=\begin{cases}\sum_{x_i} x_i\ f_X(x_i) & \text{(discrete)}\\ \int_{I\!R} x\ f_X(x)dx & \text{(continuous).}\end{cases}$$

Definition A.10: Given a discrete (continuous) r.v. X with p.m.f. (p.d.f.) $f_X(x)$, the *variance* of X is the central second order moment of X, *i.e.*,

$$Var(X)=E[(X-E(X))^2]=\begin{cases}\sum_{x_i}(x_i-E(X))^2\ f_X(x_i) & \text{(discrete)}\\ \int_{I\!R}(x-E(X))^2\ f_X(x)dx & \text{(continuous).}\end{cases}$$

Momentos simples e centrais

Definition A.11: Let X be a discrete (continuous) r.v. with p.m.f. (p.d.f.) $f_X(x)$ and k positive integer. The expected value of X^k, known as the simple order moment k of X, if any, is

$$E(X^k)=\begin{cases}\sum_{x_i} x_i^k\ f_X(x_i) & \text{(discreet case)}\\ \int_{I\!R} x^k\ f_X(x)dx & \text{(continuous case).}\end{cases}$$

Definition A.12: Let X be a discrete (continuous) r.v. with p.m.f. (p.d.f.) $f_X(x)$ and k positive integer. The expected value of $(X-E(X))^k$, known as the k-order central moment of X, if any, is

$$E((X-E(X))^k)=\begin{cases}\sum_{x_i}(x_i-E(X))^k\ f_X(x_i) & \text{(discreet case)}\\ \int_{I\!R}(x-E(X))^k\ f_X(x)dx & \text{(continuous case).}\end{cases}$$

Covariance and Correlation

Definition A.13: Given two r.v. X and Y, the *covariance* of X and Y is the expected value of the product of the average deviations of X and Y, *i.e.*,

$$Cov(X,Y)=E[(X-E(X))(Y-E(Y))]=E(XY)-E(X)E(Y).$$

Definition A.14: Given a random pair (X,Y), the (linear) *correlation coefficient* of X and Y is a dimensionless parameter given by

$$Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}.$$

Note A.1: i) $-1\leq Corr(X,Y)\leq 1$, ii) $Y=a\,X+b \Leftrightarrow Corr(X,Y)=\pm 1$.

Expected value and covariance matrix of the vector X

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector in $I\!R^n$, where $X_1, \ldots, X_n$ are r.v. with $E(X_i) = \mu_i$, $Var(X_i) = \sigma_i^2$, and $Cov(X_i, X_j) = \sigma_{ij}$, $j \neq i = 1, \ldots, n$. The expected value of $\mathbf{X}$ is understood to be

$$E(\mathbf{X}) = \boldsymbol{\mu} \equiv (\mu_1, \ldots, \mu_n)^T$$

while the covariance matrix of $\mathbf{X}$ is

$$Var(\mathbf{X}) = E((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T) \equiv \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}, \quad \sigma_{ij} = \sigma_{ji}$$

If $X_1, \ldots, X_n$ are independent r.v. and $Var(X_i) = \sigma^2$, $i = 1, \ldots, n$, then $Var(\mathbf{X}) = \sigma^2 \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix.

Conditional expected value and properties

Definition A.15: Given a discrete (continuous) random pair (X, Y) with conditional p.m.f. (p.d.f.) of X given $Y = y$, denoted by $f_{X|Y=y}(x)$, the *conditional expected value* of X given $Y = y$ is

$$E(X|Y = y) = \begin{cases} \sum_{x_i} x_i f_{X|Y=y}(x_i) & \text{(discrete)}, \\ \int_{I\!R} x f_{X|Y=y}(x)\, dx & \text{(continuous)}, \end{cases}$$

where *e.g.* the conditional p.d.f. $f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$, if $f_Y(y) > 0$.
Properties: If (X, Y) is a random pair with joint p.m.f. (p.d.f.) $f_{X,Y}(x, y)$ and marginal p.m.f. (p.d.f.) of Y $f_Y(y)$,

$$E(E(X|Y)) = E(X), \quad \text{in case that } E(X) < \infty.$$

Some discrete distributions

1. *Discrete uniform:* r.v. X with equally likely $x_1, \ldots, x_k$ *i.e.* its p.m.f. $f_X(x) = \frac{1}{k}$, if $x = x_1, \ldots, x_k$, 0, otherwise (o.w.).
2. *Binomial:* $X \sim Bi(n, p)$ iff $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$.
3. *Hypergeometric:* r.v. $X \sim Hpg(N, M, n)$ iff $f_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$, $\max\{0, n - N + M\} \leq x \leq \min\{n, M\}$.
4. *Geométrica:* $X \sim Geo(p)$ iff $f_X(x) = (1-p)^{x-1} p$, $x = 1, 2, \ldots$.
5. *Poisson:* $X \sim Poi(\lambda)$ iff $f_X(x) = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, \ldots$.

Some continuous distributions

1. *Continuous uniform:* $X \sim U(a, b)$ iff $f_X(x) = \frac{1}{b-a}$, $a < x < b$.
2. *Gamma:* $X \sim Ga(a, b)$ iff $f_X(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \mathrm{e}^{-b\,x}$, $x \geq 0$, where $\Gamma(a) = \int_0^\infty x^{a-1} \mathrm{e}^{-x} dx$ and $\Gamma(a) = (a-1)!$, $a \in I\!N$. Notice that if $a = 1, b = \lambda$, $X \sim$ Exponential(λ) and if $a = \frac{\nu}{2}, b = \frac{1}{2}$, $X \sim$ chi-squared(ν) $\equiv \chi^2_{(\nu)}$.

3. *Weibull:* $X \sim Wei(a,b)$ iff $f_X(x) = b\,a\,x^{a-1}\mathrm{e}^{-b\,x^a}$, $x \geq 0$.

4. *Beta:* $X \sim Be(a,b)$ iff $f_X(x) = \frac{1}{B(a,b)}\,x^{a-1}\,(1-x)^{b-1}$, $0 < x < 1$, where $B(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$.

5. *Normal (or Gaussian):* $X \sim N(\mu,\sigma^2)$ iff $f_X(x) = \frac{1}{\sqrt{2\,\pi}\,\sigma}\exp\left[-\frac{1}{2\,\sigma^2}(x-\mu)^2\right]$, $\infty < x < \infty$.

Convergence in distribution

Definition A.16: Let $X, X_1, X_2\ldots$ be r.v.'s with their distribution functions $F_X, F_{X_1}, F_{X_2}, \ldots$. The succession $\{X_n\}$ is said to converge on distribution to X ($X_n \xrightarrow{\mathcal{D}} X$), if

$$F_{X_n}(x) \to F_X(x), \text{quando } n \to \infty,$$

$\forall\, x$ continuity point of F_X. That is,

$$\forall x,\ \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

$\Leftrightarrow \forall\delta > 0, \exists n_1(\delta) : n > n_1(\delta) \Rightarrow |F_{X_n}(x) - F_X(x)| < \delta, \forall x$.

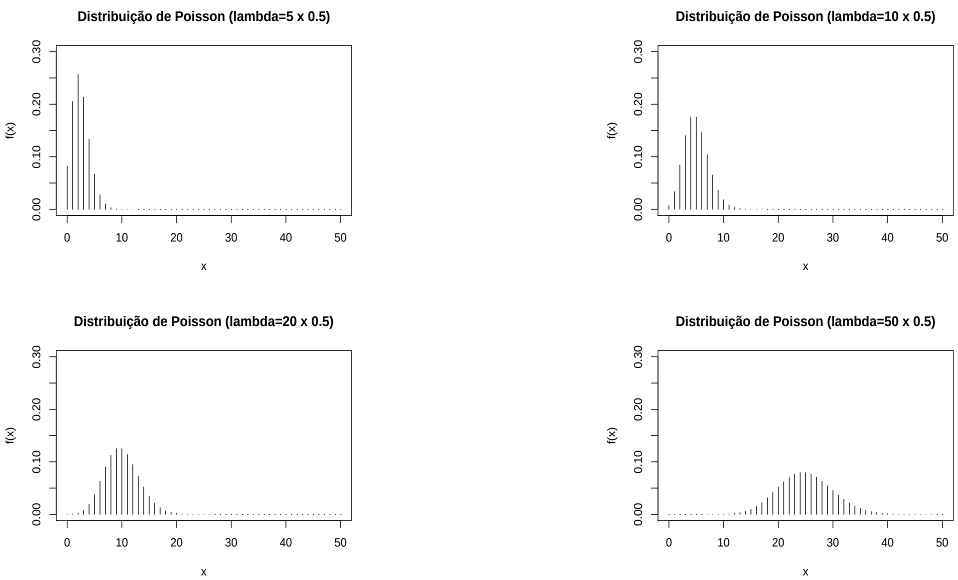
Central Limit Theorem

Theorem A.2 (C.L.T.): Let $X_1, X_2\ldots$ be a succession of independent and identically distributed (i.i.d) r.v. with expected value μ and variance σ^2, both finite. For $S_n = \sum_{i=1}^{n} X_i$, you have

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad \xrightarrow{\mathcal{D}} \quad N(0,1).$$

That is, for reasonably large n, $P\big(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\big) \approx \Phi(x)$, where $\Phi(\cdot)$ is the standard normal distribution function, *i.e.*, $N(0,1)$. So, $S_n \overset{a}{\sim} N(n\mu, n\sigma^2)$ for n large enough.

Application to Poisson distribution: $X_i \sim$ Poisson(λ), $i = 1,2,\ldots \Rightarrow S_n \sim$ Poisson$(n\lambda) \overset{a}{\sim} N(n\lambda, n\lambda)$ (next).



Sampling

Definition A.17: Given a population that is associated with a random variable X with a certain probability distribution, a n size *random sample* (r.s.) of that population is a sequence of n independent and identically distributed (i.i.d.) random variables $X_1 \dots, X_n$ with the same distribution of X.

Definition A.18: Given a random sample $(X_1 \dots, X_n)$ from a population X with p.m.f. (p.d.f.) $f_X(x)$, the *sample probability distribution* (joint p.m.f. or p.d.f.) is given by.

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} f_X(x_i).$$

Estimator Properties

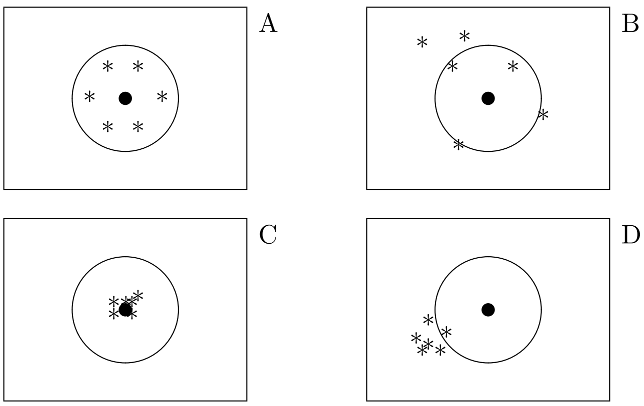
The basic properties of the estimators are related to notions of accuracy and precision similar to the characterization of experimental methods of measuring an unknown quantity in terms of the agreement of the repeated measurements obtained, in which one considers

$$\boxed{\textit{Accuracy} = \text{agreement of observations with target value.}}$$

$$\boxed{\textit{Precision} = \text{concordance of observations with each other.}}$$

Accuracy is associated with systematic errors, *e.g.*, measurement instrument deficiencies, while precision refers to the random errors that are responsible for small unpredictable variations in the measurements taken, whose causes are not completely known.

(Informal) illustration of targeting players ('estimators') with good accuracy (A,C) and good precision (C,D).



Definition A.19: Let $(X_1 \dots, X_n)$ be a r.s. of X with distribution indexed by the parameter θ. The estimator $T = T(X_1, \dots, X_n)$ is said to be an *unbiased* (centered) estimator of θ if $E(T) = \theta$.

Definition A.20: Let $T = T(X_1, \dots, X_n)$ be an estimator of the parameter θ. A measure of the variability of the estimator T is *mean square error* (MSE), given by

$$EQM(T) \equiv E((T-\theta)^2) = Var(T) + (E(T) - \theta)^2.$$

Definition A.21: Let $T = T(X_1, \dots, X_n)$ and $U = U(X_1, \dots, X_n)$ be two estimators of the parameter θ. T is said to be more *efficient* than U if

$$EQM(T) \leq EQM(U),\ \forall\, \theta$$

with strict inequality for some θ.

Other distributions

1. *Negative binomial:* $X \sim BiN(r,p)$ iff $f_X(x) = \binom{r+x-1}{r-1} p^r\,(1-p)^x$, $x = 0,1,2,\ldots$.
2. *t-Student:* $X \sim t_{(k)}$ iff $f_X(x) = \frac{1}{\sqrt{k\,\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-(\frac{k+1}{2})}$, $\infty < x < \infty$.
3. *Fisher-Snedecor:* $X \sim F(a,b)$ iff $f_X(x) = \frac{a^{a/2}\,b^{b/2}}{Beta(a/2,b/2)} x^{a/2-1} \mathrm{a\,x{+}b}^{-(a+b)/2}$, $x \geq 0$.
4. *Multinomial:* $\mathbf{X} = (X_1,\ldots,X_k) \sim M_k(n, \mathbf{p} = (p_1,\ldots,p_k))$ iff $f_{\mathbf{X}}(\mathbf{x}) = \dfrac{n!}{x_1!\cdots x_k!}\displaystyle\prod_{i=1}^{k} p_i{}^{x_i}$, $x_i = 0,1,\ldots,n$, with $\displaystyle\sum_{i=1}^{k} p_i = 1$, $\displaystyle\sum_{i=1}^{k} x_i = n$.
5. *Multivariate Normal:* $\mathbf{X} = (X_1,\ldots,X_k) \sim N_k(\boldsymbol{\mu}, \Delta)$ iff $f_{\mathbf{X}}(\mathbf{x}) = (2\,\pi)^{-\frac{k}{2}} |\Delta|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Delta^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]$, $\infty < x_i < \infty$.

The law of large numbers

According to the law, the average of the results obtained from a large number of trials should be close to the expected value.

Two different versions of the law of large numbers are:

1. *Weak law:* it states that the sample average converges in probability towards the expected value, that is, $\bar{X}_n \overset{P}{\to} \mu$ when $n \to \infty$ *i.e.*, for any positive number ϵ, $\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$.
2. *Strong law:* it states that the sample average converges almost surely to the expected value, that is, $\bar{X}_n \overset{a.s.}{\to} \mu$ when $n \to \infty$ *i.e.*, $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$.

Bibliography

Agresti A (2019). *An Introduction to Categorical Data Analysis*, 3rd edition. Wiley.

Amaral Turkman MA, Paulino CD, Müller P (2019). *Computational Bayesian Statistics: An Introduction*. Cambridge University Press, Cambridge.

Amaral Turkman MA, Silva GL (2000). *Modelos Lineares Generalizados: da teoria à prática*. Edições SPE, Lisboa.

Belitz C, Brezger A, Kneib T, Lang S, Umlauf N (2017). *BayesX: Software for Bayesian Inference in Structured Additive Regression Models*. Version 1.1.

Bliss CI (1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology*, 22, 134-167.

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1-32.

Casella G, George EI (1992). Explaining the Gibbs sampler. *The American Statistician*, 46, 167-174.

Cox DR (2006). *Principles of Statistical Inference*. Cambridge University Press.

Dempster AP, Laird NM, Rubin, DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39, 1-38.

Efron B (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7:1–26.

Efron B (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap, and other methods. *Biometrika*, 68, 589–599.

Efron B, Tibshirani RJ (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton, FL.

Gamerman D, Lopes HF (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd edition. Chapman & Hall/CRC, London.

Gelfand AE (1996). Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice*, WR Gilks, S Richardson and DJ Spiegelhalter (eds.), 145-161, Chapman and Hall, London.

Gelfand AE, Smith AFM (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, 85, 398-409.

Gelman A, Rubin DB (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7, 457-511.

Geman S, Geman D (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Gentle JE (2002). *Elements of Computational Statistics*. Springer, New York.

Geweke J (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57, 1317-1339.

Geweke J (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4*, JM Bernardo, JO Berger, AP Dawid and AFM Smith (eds.), 169-193, University Press, Oxford.

Gilks WR, Wild P (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337-348.

Härdle W, Okhrin O, Okhrin Y (2017). *Basic Elements of Computational Statistics*. Springer.

Hastie T, Tibshirani R, Friedman J (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, 2nd edition. Springer-Verlag.

Hastings WK (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97-109.

Heidelberger P, Welch PD (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31, 1109-1144.

Kardaun O (1983). Statistical survival analysis of male larynx-cancer patients - a case study. *Statistica Neerlandica*, 37, 103-125.

Kutner MH, Nachtsheim CJ, Neter J, Li W (2005). *Applied Linear Statistical Models*, 5th Edition. McGraw-Hill/Irwin, Boston.

Lawless JF (2003). *Statistical Models and Methods for Lifetime Data*, 2nd edition. Wiley, New York.

Marin J-M, Robert CP (2007). *Bayesian Core - A Practical Approach to Computational Bayesian Statistics*. Springer, New York.

McCullagh P (2002). What is a statistical model? *Annals of Statistics*, 30(5), 1225–1310.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21, 1087-1092.

Nelder JA, Wedderburn RWM (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, 135, 370-384.

Paulino CD, Amaral Turkman MA, Murteira B, Silva GL (2018). *Estatística Bayesiana*, 2ª edição. Fundação Calouste Gulbenkian, Lisboa.

Plummer M (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *3rd International Workshop on Distributed Statistical computing (DSC 2003)*, K Hornik, F Leisch and A Zeileis (eds.).

Quenouille MH (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Series B*, 11, 68-84.

Quenouille MH (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.

Raftery AE, Lewis S (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics 4*, JM Bernardo, JO Berger, AP Dawid and AFM Smith (eds.), 169-193, University Press, Oxford.

Ripley BD (1987). *Stochastic Simulation*. Wiley, New York.

Rizzo ML (2019). *Statistical Computing with R*, 2nd edition. Chapman and Hall/CRC.

Robert CP, Casella G (2004). *Monte Carlo Statistical Methods*, 2nd Edition. Springer, New York.

Ross, SM (2014). *Introduction to Probability and Statistics for Engineers and Scientists*, 5th edition, Academic Press.

Rue H, Martino S, Chopin N (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. Royal Statist. Soc. B*, 71(2), 319-392.

Schucany WR, Gray HL, Owen DB (1971), On bias reduction in estimation. *Journal of the American Statistical Association*, 66, 524–533.

Sen PK, Singer JM (1993). *Large Sample Methods in Statistics. An Introduction with Applications*. Chapman and Hall, New York.

Silva GL, Dean CB (2006). *Uma Introdução à Análise de Modelos Espaço-temporais para Taxas, Proporções e Processos de Multi-estados*. Edições ABE, Caxambú - Brasil.

Tanner MA (1996). *Tools for Statistical Inference*, 3rd Edition. Springer Verlag, New York.

Thomas A, O'Hara B, Ligges U, Sturtz S (2006). Making BUGS open. *R News*, 6(1), 12-17.

Tukey J (1958). Bias and confidence in not quite large samples (abstract). Annals of Mathematical Statistics, 29, 614.

Wickham H, Grolemund G (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media.

Exercises[10]

1. Introduction

REVIEW OF PROBABILITY AND STATISTICS (Ross, 2004/2009):

1.1.1 An experiment measuring the percent shrinkage on drying of 50 clay specimens produced the following data:

18.2	21.2	23.1	18.5	15.6	20.8	19.4	15.4	21.2	13.4
16.4	18.7	18.2	19.6	14.3	16.6	24.0	17.6	17.8	20.2
17.4	23.6	17.5	20.3	16.6	19.3	18.5	19.3	21.2	13.9
20.5	19.0	17.6	22.3	18.4	21.2	20.4	21.4	20.3	20.1
19.6	20.6	14.8	19.7	20.5	18.0	20.8	15.8	23.1	17.0

a) Draw a stem and leaf plot of these data.

b) Compute the sample mean, median, and mode.

c) Compute the sample variance.

d) Group the data into class intervals of size 1 percent starting with the value 13.0, and draw the resulting histogram.

e) For the grouped data acting as if each of the data points in an interval was actually located at the midpoint of that interval, compute the sample mean and sample variance and compare this with the results obtained in parts (b) and (c). Why do they differ?

1.1.2 Prostate cancer is the most common type of cancer found in males. As an indicator of whether a male has prostate cancer, doctors often perform a test that measures the level of the PSA protein (prostate specific antigen) that is produced only by the prostate gland. Although higher PSA levels are indicative of cancer, the test is notoriously unreliable. Indeed, the probability that a noncancerous man will have an elevated PSA level is approximately 0.135, with this probability increasing to approximately 0.268 if the man does have cancer. If, based on other factors, a physician is 70 percent certain that a male has prostate cancer, what is the conditional probability that he has the cancer given that

a) the test indicates an elevated PSA level;

b) the test does not indicate an elevated PSA level?

Repeat the preceding, this time assuming that the physician initially believes there is a 30 percent chance the man has prostate cancer.

1.1.3 The joint density function of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 2, & 0 < x < y,\ 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

a) Compute the density of X.

b) Compute the density of Y.

[10] The proposed exercises are in the following bibliography books: Gentle (2002), Ross (2004/2009), Robert & Casella (2004), Paulino et al. (2018), Rizzo (2019) and Amaral-Turkman *et al.* (2019), including solutions to some of these.

c) Are X and Y independent?

1.1.4 Independent trials, each of which is a success with probability p, are successively performed. Let X denote the first trial resulting in a success. That is, X will equal k if the first $k-1$ trials are all failures and the kth a success. X is called a geometric random variable. Compute

a) the probability (mass) function of X, $f_X(k) = P(X = k),\ k = 1, 2, \ldots$;

b) the expected value of X, $E(X)$.

Let Y denote the number of trials needed to obtain r successes. Y is called a negative binomial random variable. Compute

c) the probability (mass) function of Y, $f_Y(k) = P(Y = k),\ k = r, r+1, \ldots$.

1.1.5 If U is uniformly distributed on $(0, 1)$, show that $V = a + (b - a)\,U$ is uniform on (a, b).

1.1.6 A certain component is critical to the operation of an electrical system and must be replaced immediately upon failure. If the mean lifetime of this type of component is 100 hours and its standard deviation is 30 hours, how many of the components must be in stock so that the probability that the system is in continual operation for the next 2000 hours is at least 0.95?

1.1.7 The following are the burning times in seconds of floating smoke pots of two different types.

Type I					Type II				
481	506	527	661	501	526	511	556	542	491
572	561	501	487	524	537	582	605	558	578

Find a 99 percent confidence interval for the mean difference in burning times assuming normality with unknown but equal variances.

1.1.8 In an experiment 10 albino rats were used to study the effectiveness of carbon tetrachloride as a treatment for worms. Each rat received an injection of worm larvae. After 8 days, the rats were randomly divided into two groups of 5 each; each rat in the first group received a dose of 0.032 cc of carbon tetrachloride, whereas the dosage for each rat in the second group was 0.063 cc. Two days later the rats were killed, and the number of adult worms in each rat was determined. The numbers detected in the group receiving the 0.032 dosage were $421, 462, 400, 378, 413$, whereas they were $207, 17, 412, 74, 116$ for those receiving the 0.063 dosage. Do the data prove that the larger dosage is more effective than the smaller?

1.1.9 Among 100 vacuum tubes tested, 41 had lifetimes of less than 30 hours, 31 had lifetimes between 30 and 60 hours, 13 had lifetimes between 60 and 90 hours, and 15 had lifetimes of greater than 90 hours. Are these data consistent with the hypothesis that a vacuum tube's lifetime is exponentially distributed with a mean of 50 hours?

1.1.10 A sample of 300 cars having cellular phones and one of 400 cars without phones were tracked for 1 year. The following table gives the number of these cars involved in accidents over that year.

	Accident	No Accident
Cellular phone	22	278
No phone	26	374

Use the table above to test the hypothesis that having a cellular phone in your car and being involved in an accident are independent.

BAYESIAN STATISTICS (Paulino et al., 2018):

1.2.1 A commercial director, considering the proportion of individuals in a given population who would prefer the company's product, established, for ease of analysis, five cases to which he gives the following prior probabilities:

Proportion of consumers	Prior probabilities
0.10	0.25
0.15	0.35
0.20	0.20
0.25	0.15
0.30	0.05

After a market study, it was found that among 10 individuals consulted, there were 2 who said they were consumers of the product. Update the odds of the 5 cases considered.

1.2.2 An analytical system that aids blood type determination, according to the ABO group, consists in observing in each person a random variable X with density function.

$$f(x \mid \theta) = e^{-(x-\theta)} I_{(\theta,+\infty)}(x), \quad \theta > 0.$$

The classification in each blood type depends on the value of θ according to the following match:

$$\begin{array}{ccl} 0 < \theta < 1 & \Longrightarrow & \text{type O} \\ 1 \leq \theta < 2 & \Longrightarrow & \text{type A} \\ 2 \leq \theta < 3 & \Longrightarrow & \text{type B} \\ \theta \geq 3 & \Longrightarrow & \text{type AB} \end{array}$$

The distribution of the θ feature over the universe of people at any given time is as follows.

$$h(\theta) = e^{-\theta} I_{(0,+\infty)}(\theta).$$

a) Determine the prior probability of a randomly chosen person having each of the blood types.

b) Update the probabilities of (a) for a person whose analysis resulted in $X = 4$.

1.2.3 In describing the results of an investigation a statistician stated that the posterior distribution of the mean θ of a Normal distribution with variance 100 was still Normal with mean 52 and variance 10. He further added that the experimental information consisted of a 4-element sample averaging 55.

Fully specify the prior distribution of θ which is also known to be Normal.

1.2.4 The number of ships entering the river bar each day has a Poisson distribution of mean θ whose prior distribution is Exponential of mean 1. Knowing that in 5 days we observed the entries 3, 5, 4, 3 and 4:

a) Determine the posterior distribution of θ.

b) Obtain 90% and 95% credible intervals for θ.

c) Calculate the Bayes factor if the daily mean of ships entering the bar is greater than 3.8.

Hint: If $Z \sim \text{Gamma}(a, b)$, then $2b\, Z \sim \chi^2_{(2a)}$.

1.2.5 One researcher is interested in the difference between the means, $\theta = \mu_1 - \mu_2$, of two independent Normal populations with a common variance of 100. A sample of size 25 is taken from each population with their means $\bar{x}_1 = 80$ and $\bar{x}_2 = 60$. If *a priori* if you have $\theta \sim N(10, 50)$:

a) Determine the posterior distribution of θ.

b) Calculate the posterior and prior odds ratios in favor of the hypothesis $H_0 : \mu_1 \leq \mu_2$ against $H_1 : \mu_1 > \mu_2$.

1.2.6 Determine a $\psi = g(\theta)$ reparameterization whose Jeffreys' prior distribution, $h(\psi)$, is constant in the following scenarios:

a) $X|\theta \sim Poi(\theta)$

b) $X|\theta \sim Ga(\alpha, \theta^{-1}),\ \alpha = 1, 2.$

Hint: Notice that $h(\theta) = h(\psi(\theta))\, \psi'(\theta)$.

1.2.7 Suppose your interest in a given problem is focused on a probability density function expressed by

$$f(x|\theta) = \frac{2x}{\theta} \mathrm{e}^{-x^2/\theta}, \quad x > 0, \theta > 0.$$

Determine Jeffreys' prior distribution for the θ parameter of this model.

1.2.8 Let X be a random variable with density function

$$f(x \mid \theta_1, \theta_2) = (\theta_2 - \theta_1)^{-1} I_{(\theta_1, \theta_2)}(x),\ \theta_1 < 0 < 1 < \theta_2$$

where the parameter pair (θ_1, θ_2) is prior distributed according to the density function

$$h(\theta_1, \theta_2) = 2(\theta_2 - \theta_1)^{-3},\ \theta_1 < 0,\ \theta_2 > 1.$$

a) Determine the marginal density function of X to $0 < x < 1$ and the resulting posterior density of (θ_1, θ_2).

b) Show that if $x > 1$

$$h(\theta_1, \theta_2 \mid x) = 6x^2(\theta_2 - \theta_1)^{-4},\ \theta_1 < 0,\ \theta_2 > x,$$

and that if $x < 0$

$$h(\theta_1, \theta_2 \mid x) = 6(1 - x)^2(\theta_2 - \theta_1)^{-4},\ \theta_1 < x,\ \theta_2 > 1.$$

1.2.9 Let $X_1, \ldots, X_n$ be i.i.d. observations of the $NegativeBinomial(m, \theta)$ model.

a) Determine the natural conjugate family of this model.

b) Being *a priori* $\theta \sim Beta(a, b)$, show that

$$E[\theta \mid \{x_i\}] = \frac{a + mn}{a + b + mn + \sum_{i=1}^n x_i}$$

and compare with the posterior mode of θ.

c) Derive the predictive probability function from an independent future observation such that $Y \mid m, \theta \sim NegativeBinomial(m, \theta)$.

1.2.10 Let $\mathcal{F} = \left\{f(x \mid \theta) = \theta^{-1} I_{[0,\theta]}(x),\, \theta > 1\right\}$ where θ is distributed *a priori* according to density

$$h(\theta) = \theta^{-2} I_{(1,+\infty)}(\theta).$$

Made an observation, you get $X = x_1 > 1$.

a) Determine the $100\gamma\%$ HPD interval for θ and say that x_1 values are greater width than the respective prior credible interval.

b) Determine the predictive probability of a 2nd observation of X exceeding x_1.

c) Show that if $x_1 < 1$ the $100\gamma\%$ HPD interval has the form $\left(1, (1-\gamma)^{-\frac{1}{2}}\right)$.

1.2.11 Let $X_1, \ldots, X_n$ be the instants of passage through a given traffic light of n vehicles supposedly independent and uniformly distributed in the interval $(0, \theta)$, where $\theta > 0$ is the time interval where the signal is open. Consider for the prior distribution the Pareto density of hyperparameters $c > 0$ and $b > 0$ ($Pareto(c, b)$), i.e., $h(\theta) = b\, c^b\, \theta^{-(b+1)}\, I_{[c,\infty)}(\theta)$.

a) Show that the posterior distribution is the distribution $Pareto(C, B)$, where $C = \max(c, x_{(n)})$ and $B = b + n$, with $x_{(n)}$ representing the maximum sample.

b) Based on model in a), determine the minimum number of vehicles to pass so that the HPD interval for $\ln\theta$ of amplitude 5% has a degree of credibility *a posteriori* of at least 0.95 when the prior median of θ is $2c$.

1.2.12 Let X be the number of defects per meter of a clothing piece modeled on a Poisson(θ) distribution, with the expected number of defects per meter distributed *a priori* according to the expected value natural conjugate family member 1 and variance 1/2. The n meter inspection result of a randomly chosen part yielded for θ *a posteriori* an expected value of 1.5 and a variance of 0.125.

Deduct the number of meters inspected and the corresponding mean number of defects per meter.

* COMPLEMENTARY EXERCISES (Amaral-Turkman et al., 2019): 1.1, 1.2, 1.3, 1.4, 2.1, 2.2, 2.3, 2.4, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6.

STOCHASTIC SIMULATION (Gentle, 2002; Rizzo, 2019):

1.3.1 Prove that if X is a random variable with an absolutely continuous distribution function $F_X(x)$, $F_X(X)$ has a Uniform distribution in $(0, 1)$.

1.3.2 A random variable X is said to have a Logistic distribution with parameters μ and $\sigma > 0$ if its probability density function is

$$f_X(x|\mu, \sigma) = \frac{\mathrm{e}^{\frac{x-\mu}{\sigma}}}{\sigma\left(1 + \mathrm{e}^{\frac{x-\mu}{\sigma}}\right)^2}, \quad -\infty < x < \infty.$$

Show an inverse transform method to generate a random sample of size n for this distribution.

1.3.3 Computationally implement the method presented in 1.3.2 to obtain a random sample of X when $n = 1000$, $\mu = 0$ and $\sigma = 1$.

1.3.4 In a bridge hand, an ace is assigned a value of 4, a king 3, a queen 2, and a jack 1. All other cards from the 52 card deck are assigned a value of 0. Let X be the value of the cards with probability mass function

x	0	1	2	3	4
$f_X(x)$	36/52	4/52	4/52	4/52	4/52

a) Develop an algorithm to generate random samples from this distribution. b) Generate a sample random of size 1000 from the distribution of X. c) Compare the empirical with the theoretical probabilities.

1.3.5 Taking into account the previous question, show a method to generate the sum of the values of the cards in the hand, if a player has 13 cards.

1.3.6 Let X be a random variable with Logarithm(θ) distribution *i.e.* probability mass function $f_X(x|\theta) = \frac{-1}{\log(1-\theta)}\frac{\theta^x}{x}$, $x \geq 1$, $0 < \theta < 1$. If U, V are independent Uniform(0,1) random variables, then $X = \left\lfloor 1 + \frac{\log(V)}{\log(1-(1-\theta)^U)} \right\rfloor$ has the Logarithm(θ) distribution, where $\lfloor x \rfloor$ denotes the integer part of x. Based on this transformation, provide a simple and efficient generator for this distribution.

1.3.7 Show the steps of a sampling method in order to generate a Gamma(n, λ) random variable as the convolution of n independent and identically distributed Exponential(λ) random variables. For $n = 10$, $\lambda = 2$, obtain a 1000-size sample random based on this generating method and compare the empirical with the theoretical mean and variance.

1.3.8 The Negative Binomial distribution is a mixture of Poisson(λ) distributions, where λ has a gamma distribution. Specifically, if $X|\lambda \sim$ Poisson(λ) and $\lambda \sim$ Gamma(a, b), then X has the Negative Binomial distribution with parameters a and $\theta = b/(1+b)$. Considering $a = 4$ and $b = 3$, get two 1000-size random samples for the Negative Binomial distribution by using both this mixture distribution and its probability mass function. Compares the two samples via the corresponding histograms.

2. Classic estimation methods and algorithms

2.1 (Ross, 2004/2009) Exercises: 9.32, 9.37, 9.44, 9.48.

2.2 (Ross, 2004/2009) Exercises: 7.1, 7.2, 7.3, 7.5, 7.7.

2.3 (Gentle, 2002) Exercises: 1.7, 1.9, (Tanner, 1996) Example: pg.66, *i.e.*,

2.3.1 Consider the Multinomial($n, \mathbf{p}$) distribution for the random vector $\mathbf{X} = (X_1, X_2, X_3, X_4)$, given n and $\mathbf{p} = (p_1, p_2, p_3, p_4)$, with $\sum_{i=1}^4 x_i = n$ and $\sum_{i=1}^4 p_i = 1$. Assume that the probabilities are related by a single parameter θ: $p_1 = \frac{1}{2} + \frac{1}{4}\theta$, $p_2 = p_3 = \frac{1}{4} - \frac{1}{4}\theta$, $p_4 = \frac{1}{4}\theta$, where $0 < \theta < 1$. Let $n = 197$ and $\mathbf{x} = (125, 18, 20, 34)$. a) use the Newton-Raphson method do determine the maximum likelihood estimation for θ. b) Compare the previous method with the Fisher scoring method for estimating θ by writing two programs and starting with $\theta^{(0)} = 0.5$.

2.3.2 Assume a random sample $X_1, \ldots, X_n$ from a gamma distribution with parameters α and β. a) Use the Newton-Raphson method for determining the maximum likelihood estimate of α and β. Does it have a closed-form solution? b) Make a program to get a sample of size n related to a) and use it to compute the estimate of α and β based on an artificial sample of size 500 from a Gamma(5,2) distribution. c) Computes an approximation of the variance-covariance matrix using the inverse of the Fisher information matrix.

2.4 (Gentle, 2002) Exercise: 1.9, (Tanner, 1996) Example: pg.67, *i.e.*,

2.4.1 Consider the Multinomial($n, \mathbf{p}$) distribution for the random vector $\mathbf{U} = (U_1, U_2, U_3, U_4)$, given n and $\mathbf{p} = (p_1, p_2, p_3, p_4)$, with $\sum_{i=1}^4 u_i = n$ and $\sum_{i=1}^4 p_i = 1$. Assume that the probabilities are related by a single parameter θ: $p_1 = \frac{1}{2} + \frac{1}{4}\theta$, $p_2 = p_3 = \frac{1}{4} - \frac{1}{4}\theta$, $p_4 = \frac{1}{4}\theta$, where $0 < \theta < 1$. Let $n = 197$ and $\mathbf{u} = (125, 18, 20, 34)$. Augment the observed data by splitting the first cell into two cells with probabilities $\frac{1}{2}$ and $\frac{\theta}{4}$. The augmented data are $\mathbf{v} = (v_1, v_2, v_3, v_4, v_5)$ with $v_1 + v_2 = 125$, $v_3 = 18$, $v_4 = 20$, $v_5 = 34$. Under a flat prior, find the observed and augmented posterior distributions and write a program to determine the estimate of θ by EM algorithm, again starting with $\theta^{(0)} = 0.5$.

2.4.2 Consider an experiment involving ten motorettes that were tested at each of the four temperatures: 150^o, 170^o, 190^o and 220^o. The failure and censoring times in hours are given below, where a star indicates that a motorette was taken off the study without failing at the indicated time. For these data, a regression model was fitted $t_i = \beta_0 + \beta_1 x_i + \sigma\epsilon_i$, where $\epsilon_i \sim N(O, 1)$, $x_i = 1000/(\text{temperature} + 273.2)$ and $t_i = log_{10}$ (ith failure time). Reorder the data so that the first m observations are uncensored (i.e. a failure is observed at t_i) and the remaining n–m are censored (c_i denotes a censored event time). Notice that the conditional distribution $f(v_i|\beta_0, \beta_1, \sigma^2, c_i)$ is the conditional normal distribution, conditional on the unobserved failure time V_i is greater than c_i. Find the observed and augmented likelihood functions and write a program to determine the estimate of β_0, β_1, σ^2 by EM algorithm.

150^o	8064*	8064*	8064*	8064*	8064*	8064*	8064*	8064*	8064*	8064*
170^o	1764	2772	3444	3542	3780	4860	5196	5448*	5448*	5448*
190^o	408	408	1344	1344	1440	1680*	1680*	1680*	1680*	1680*
220^o	408	408	504	504	504	528*	528*	528*	528*	528*

Hint: $E(\epsilon_i|\epsilon_i > r_i) = \phi(r_i)/(1 - \Phi(r_i))$ and $E(\epsilon_i^2|\epsilon_i > r_i) = (r_i\,\phi(r_i) - (1 - \Phi(r_i)))/(1 - \Phi(r_i))$, where $r_i = \frac{c_i - \mu_i}{\sigma}$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are p.d.f. and c.d.f. of the $N(0, 1)$, respectively.

2.5 (Gentle, 2002) Exercises: 1.9, (Tanner, 1996) Example: pg.66, *i.e.*,

2.5.1 Consider the Multinomial($n, \mathbf{p}$) distribution for the random vector $\mathbf{U} = (U_1, U_2, U_3, U_4)$, given n and $\mathbf{p} = (p_1, p_2, p_3, p_4)$, with $\sum_{i=1}^4 u_i = n$ and $\sum_{i=1}^4 p_i = 1$. Assume that the probabilities are related by a single parameter θ: $p_1 = \frac{1}{2} + \frac{1}{4}\theta$, $p_2 = p_3 = \frac{1}{4} - \frac{1}{4}\theta$, $p_4 = \frac{1}{4}\theta$, where $0 < \theta < 1$. Let $n = 197$ and $\mathbf{u} = (125, 18, 20, 34)$. Augment the observed data by splitting the first cell into two cells with probabilities $\frac{1}{2}$ and $\frac{\theta}{4}$. The augmented data are $\mathbf{v} = (v_1, v_2, v_3, v_4, v_5)$ with $v_1 + v_2 = 125$, $v_3 = 18$, $v_4 = 20$, $v_5 = 34$. Under a flat prior, write a program to determine the estimate of θ by data augmentation algorithm, again starting with $\theta^{(0)} = 0.5$.

2.5.2 Based on the program elaborate in the previous exercise, determine the estimate of θ for $n = 20$ e $\mathbf{u} = (14, 0, 1, 5)$. Comment the results.

4. Monte Carlo Methods

4.1 (Paulino *et al.*, 2018) Exercise: 7.7, (Rizzo, 2019) Exercises: 5.1, 5.3, 5.4, 5.11, *i.e.*,

4.1.1 Compute a Monte Carlo estimate of $\int_0^{\frac{\pi}{2}} \cos x \, dx$ and compare your estimate with the exact value of the integral.

4.1.2 Compute a Monte Carlo estimate $\hat{\mathcal{J}}$ of $\mathcal{J} = \int_0^{0.5} \mathrm{e}^{-x} \, dx$ by sampling from Uniform(0,0.5), and estimate the variance of $\hat{\mathcal{J}}$. Find another Monte Carlo estimator $\tilde{\mathcal{J}}$ by sampling from the exponential distribution. Which of the variances (of $\hat{\mathcal{J}}$ and $\tilde{\mathcal{J}}$) is smaller? Why?

4.1.3 Write a code to compute a Monte Carlo estimate of the Beta(3,3) cumulative distribution function, $F(x)$. Use the code to estimate $F(x)$ for $x = 0.1, 0.2, \ldots, 0.9$. Compare the estimates with the values returned by the *pbeta* function in R.

4.1.4 If $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are unbiased estimators of θ, and $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are antithetic and identically distributed variables, show that $c* = 1/2$ is the optimal constant that minimizes the variance of $\widehat{\theta}_c = c\,\widehat{\theta}_1 + (1-c)\,\widehat{\theta}_2$.

4.1.5 Consider the model $X_i,\, i = 1, \ldots, n \overset{ind}{\sim} Bi(m, \theta_i)$ with $\{\theta_i\} \overset{iid}{\sim} Be(\alpha, \beta)$, both distributions being viewed as conditional on their (hyper)-parameters. Let us assume that the prior distribution hyperparameters of θ_i in turn have a prior distribution corresponding to the independent Uniform distributions for $\lambda = \alpha/(\alpha+\beta)$ (proper) and $\delta = (\alpha+\beta)^{-1/2}$ (improper).

a) Show that the joint posterior distribution of θ_i is a mixture of

$$\theta_i \mid \alpha, \beta,\, i = 1, \ldots, n \underset{ind}{\sim} Be(\alpha + x_i, \beta + m - x_i)$$

for the (α, β) posterior distribution defined less than the constant of proportionality by

$$h(\alpha, \beta \mid x) \propto (\alpha+\beta)^{-5/2} \prod_{i=1}^{n} \frac{Be(\alpha + x_i, \beta + m - x_i)}{Be(\alpha, \beta)}.$$

b) Indicate how you can get a Monte Carlo approximation of the joint distribution emph a posteriori and the usual Bayes estimate of θ_i.

4.2 (Paulino *et al.*, 2018) Exercises: 7.9, (Rizzo, 2019) Exercises: 5.14, *i.e.*,

4.2.1 Obtain a Monte Carlo estimate of $\int_1^\infty \frac{x^2}{\sqrt{2\,\pi}} \mathrm{e}^{-x^2/2} dx$ by importance sampling.

4.2.2 Suppose the joint posterior distribution of $\theta = \left(\theta^{(m)}, \theta^{(-m)}\right)$, for some $m = 1, \ldots, k-1$ fixed, is only known by its kernel, *i.e.* $\bar{h}\,(\theta \mid x) = L\,(\theta \mid x)\; h\,(\theta)$, but the conditional distribution $h\left(\theta^{(m)} \mid \theta^{(-m)}, x\right)$ is completely known, contrary to marginal distribution of $\theta^{(-m)}$ that is unknown. Being $p\left(\theta^{(-m)}\right)$ an appropriate density of importance for $h\left(\theta^{(-m)} \mid x\right)$ from which they are generated regardless of the values $\theta^{(-m)}_{(i)}$, $1 \le i \le n$, based on which the values $\theta^{(m)}_{(i)}$ are obtained per simulation of $h\left(\theta^{(m)} \mid \theta^{(-m)}_{(i)}, x\right)$:

a) Indicate what estimates the following quantity

$$\sum_{i=1}^{n} w_i h\left(\theta^{(m)} \mid \theta^{(-m)}_{(i)}\right) / \sum_{i=1}^{n} w_i$$

where

$$w_i = \frac{\bar{h}\left(\theta^{(m)}_{(i)}, \theta^{(-m)}_{(i)} \mid x\right)}{h\left(\theta^{(m)}_{(i)} \mid \theta^{(-m)}_{(i)}, x\right) p\left(\theta^{(-m)}_{(i)}\right)},\; i = 1, \ldots, n.$$

b) Considering the discrete distribution defined by the previously simulated values $\theta^{(-m)}_{(i)}$ coupled to the masses $p_i = w_i / \sum_{i=1}^{n} w_i$, which generates l (resampling) values denoted by $\theta^{(-m)}_{(j)*}$, show that the Monte Carlo estimate with importance sampling of $h\left(\theta^{(m)} \mid x\right)$ can be expressed by

$$\hat{h}\left(\theta^{(m)} \mid x\right) = \frac{1}{l} \sum_{j=1}^{l} h\left(\theta^{(m)} \mid \theta^{(-m)}_{(j)*}, x\right).$$

4.3 (Paulino *et al.*, 2018) Exercises: 7.13, 7.15, (Rizzo, 2019) Exercises: 3.7, *i.e.*,

4.3.1 Write a code to generate a random sample of size n from the Beta(a, b) distribution by the acceptance-rejection method. Generate a random sample of size 1000 from the Beta$(3, 2)$ distribution. Graph the histogram of the sample with the theoretical Beta$(3, 2)$ density superimposed.

4.3.2 In the rejection method frame, let be $Y \sim p(\cdot)$ and $W = UMp(Y)$ where $U \sim Unif([0, 1])$ independent of Y. Show that:

a) $(Y, W) \sim Unif(B)$ where $B = \{(y, w) : 0 \le w \le Mp(y)\}$.
[*Hint*: Determine the conditional distribution $W|Y = y$.]

b) The distribution of Y on acceptance is $X \sim \pi$, detailing all calculations.

c) This method can be viewed as a Uniform sampling in the two-dimensional region under the curve $Mp(y)$ followed by discarding points falling above $\pi(y)$.

4.3.3 Considering the probability density functions explained below, show that:

a) are log-concave as follows:
- $X \sim N(\mu, \sigma^2)$: $f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)}{\sigma^2}^2} I_{(-\infty,+\infty)}(x)$
- $X \sim \text{Beta}(\alpha, \beta)$: $f(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{(0,1)}(x)$
- $X \sim \text{Logistics}(\mu, \sigma)$: $f(x|\mu, \sigma) = \frac{\exp(-\frac{x-\mu}{\sigma})}{\sigma[1+\exp(-\frac{x-\mu}{\sigma})]^2} I_{(-\infty,+\infty)}(x)$

b) are not log-concave as follows:
- $X \sim \text{Cauchy}(\lambda, \delta)$: $f(x|\lambda, \delta) = \frac{\delta}{\pi\{\delta^2+(x-\lambda)^2\}} I_{(-\infty,+\infty)}(x)$
- $X \sim LN(\mu, \sigma^2)$: $f(x|\mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(\ln x-\mu)}{\sigma^2}^2} I_{(0,+\infty)}(x)$.

5. Markov chain Monte Carlo methods

5.1 (Paulino *et al.*, 2018) Exercises: 9.1, 9.2 (Rizzo, 2019) Exercises: 9.1, *i.e.*,

5.1.1 In a Markov chain $\{U_n, n \ge 0\}$ with discrete state space and transition function $p(\cdot, \cdot)$, show that:

a) The initial distribution π_0 is stationary if and only if $P(U_n = u) = \pi_0(u), \forall n \ge 1$.

b) If the n step transition function converges to the distribution π and π is stationary, then the limit-distribution of U_n coincides with π, regardless of the initial distribution.

c) Reversibility is equivalent to the so-called detailed equilibrium condition if the distribution π is stationary.

5.1.2 Construa the Metropolis-Hastings algorithm to generate a sample from a Rayleigh(σ) distribution, whose density is $f(x) = \frac{x}{\sigma^2} \mathrm{e}^{-x^2/(2\sigma^2)}$, $x \ge 0$, $\sigma > 0$. For the proposal distribution, try the chisquared distribution with degrees of freedom $U_t \equiv X_t$. Compare the performance of the Metropolis-Hastings sampler for Rayleigh($\sigma = 4$) and Rayleigh($\sigma = 2$) distributions. In particular, what differences are obvious from the plot the 10000-size sample *versus* the time index in both scenarios?

5.1.3 Considering a chain generated by the Metropolis-Hastings algorithm, show that:

a) The respective transition function is given by and deduce from it the expression of the corresponding density $p(u, v)$.

$$P(u, A) = \int_A q(u, v)\alpha(u, v)dv + r(u)\, I_A(u),$$

b) The detailed condition of equilibrium is satisfied by it.

5.2 (Paulino *et al.*, 2018) Exercises: 9.5, 9.6, 9.7, 9.8, 9.9, 9.13, (Rizzo, 2019) Exercises: 9.8, *i.e.*,

5.2.1 Consider the bivariate density $f(x, y) \propto \binom{n}{x} y^{x+a-1}(1-y)^{n-x+b-1}$, $x = 0, 1, \ldots, n$, $0 \le y \le 1$. For fixed a, b and n, construct the Gibbs algorithm to generate a chain with target joint density $f(x, y)$.

5.2.2 Consider the bivariate auto-exponential model

$$\pi(u_1, u_2) \propto \exp[-\alpha(u_1 + u_2) - \beta u_1 u_2], \;\; u_1, u_2 > 0, \alpha, \beta > 0.$$

a) Identify the full conditional distributions associated with the Gibbs algorithm for sampling $\pi(u_1, u_2)$.

b) Considering each Gibbs sampler cycle as a Metropolis-Hastings algorithm, show that the M-H ratio is given, for $u = (u_1, u_2)$ and $v = (v_1, v_2)$, for

$$R(u, v) = \frac{(\alpha + \beta v_2)(\alpha + \beta u_1)}{(\alpha + \beta u_2)(\alpha + \beta v_1)} \; \exp[\beta(u_2 v_1 - v_2 u_1)].$$

c) Show that the ratio in b) can be less than 1 for infinite values of (u, v).
[*Hint*: Consider $\alpha = \beta = 1$ (*e.g.*)and (u, v) such that $v_2/v_1 = u_2/u_1$.]

5.2.3 Assume the space defined by the Cartesian product $\{0, 1\} \times \{0, 1\}$ equipped with probability distribution $(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}) = (1/2, 1/4, 1/8, 1/8)$. Verify that the Gibbs procedure does not satisfy the detailed equilibrium condition, showing that $\pi_{00} q_{00,11} \neq \pi_{11} q_{11,00}$, where $q_{00,11}$ and $q_{11,00}$ are the entries in the Q matrix when begins sampling from the first component order and then the second component through the full conditional distributions.

5.2.4 Let (x_i, y_i), $i = 1, \ldots, n$ be data from a pair of variables for which the linear regression model is assumed, $\{Y_i\} \underset{ind}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$. It is intended to infer about $\beta = (\beta_0, \beta_1) \in I\!R^2$ and $\sigma^2 > 0$ from the usual non-informative prior distribution $h(\beta, \sigma^2) \propto \sigma^{-2}$ and using an MCMC method for sampling the posterior distribution.

a) Show that the Gibbs sampler only needs to resort to direct simulation methods, specifying univariate full conditional distributions.

b) Say if the use of an informative prior distribution of the type of $\beta|\sigma^2 \sim N_2(b_0, \sigma^2 V_0) \wedge \sigma^2 \sim IGa(c_0, d_0)$, with fixed hyperparameters, makes the referred simulation process unfeasible in a).

5.2.5 In a study for wind turbine deployment in a given area, the wind speed X (in m/s) was measured at some point over several occasions, obtaining the data $x = (x_i, i = 1, \ldots, n)$. The model commonly used to describe the variation of X is the Weibull model with scale and shape parameters denoted by δ and α, respectively, whose probability density function is expressed by

$$f(x|\delta, \alpha) = \delta \alpha x^{\alpha - 1} e^{-\delta x^\alpha} I_{(0,\infty)}(x), \;\; \delta, \alpha > 0.$$

Let us assume that *a priori* δ and α are independent with Gamma $Ga(a, b)$ and Log-Normal $LN(c, d)$ distributions of hyperparameters completely specified, with $a, b, d > 0$, and $c \in I\!R$.

Assuming the data is a realization of a random sample of this model, specify the full conditional densities and discuss how to sample from them in each cycle of the Gibbs algorithm.

5.2.6 Let $\mathcal{D} = \{(y_1, x_1), \ldots, (y_n, x_n)\}$ be a realization of a random sample from the Poisson-Normal model characterized by $Y|X \sim Poi(\eta \delta^X)$ and $X \sim N(\mu, \tau^{-1})$. Consider for $\theta = (\eta, \delta, \mu, \tau)$ the non-informative prior distribution, given by $h(\eta, \delta, \mu, \tau) \propto (\eta \delta \tau)^{-1}$.

a) Specify the posterior distribution of θ, making explicit the Gibbs sampler steps.

b) Entering the parameterization $\eta = e^{\beta_0}$, $\delta = e^{\beta_1}$, derives the full conditional distributions of $\phi = (\beta_0, \beta_1, \mu, \tau)$ and comment on how you can sample Gibbs.

5.2.7 Assume there are y_{ij} observations from the model

$$Y_{ij} = \mu_i + \epsilon_{ij},\ \ i = 1, \dots, k,\ \ j = 1, \dots, m,$$

where ϵ_{ij} are i.i.d. $N(0, \sigma_\epsilon^2)$, μ_i are i.i.d. $N(\mu, \sigma_\mu^2)$, and that ϵ_{ij} and μ_i are independent. Suppose further that the parameters $\sigma_\epsilon, \sigma_\mu$ and μ are independent with the following distributions: $\sigma_\epsilon^2 \sim GaI(a_1, b_1)$; $\sigma_\mu^2 \sim GaI(a_2, b_2)$; $\mu \sim N(\mu_0, \sigma_0^2)$. Describe the Gibbs sampler and get explicit forms for the following distributions:

a) μ given $\{y_{ij}\}$, $\{\mu_i\}$, σ_ϵ^2 and σ_μ^2;

b) μ_i given $\{y_{ij}\}$, μ, σ_ϵ^2 and σ_μ^2;

c) σ_ϵ^2 given $\{y_{ij}\}$, $\{\mu_i\}$, μ and σ_μ^2;

d) σ_μ^2 given $\{y_{ij}\}$, $\{\mu_i\}$, μ and σ_ϵ^2.

Solutions to exercises

1.1 Look for them in Ross (2004/2009).

1.2 Look for them in Amaral-Turkman *et al.* (2019). Alternatively,
1.2.1 $h(\theta|x=2) = 0.1868, 0.3724, 0.2329, 0.1629, 0.0450$ for $\theta = 0.10, 0.15, 0.20, 0.25, 0.30$, respectively.
1.2.2 *a)* $P(O) = 0.6321$; $P(A) = 0.2326$; $P(B) = 0.0855$; $P(AB) = 0.0498$. *b)* $P(O|x=4) = P(A|x=4) = P(B|x=4) = P(AB|x=4) = 0.25$.
1.2.3 $a = E(\theta) = 50$, $b^2 = Var(\theta) = 100/6$.
1.2.4 a) Gamma$(20, 6)$. b) 90% Central CI: (2.2091,4.6465), 90% HPD CI: (2.1129,4.5232), 95% Central CI: (2.0361,4.9451), 95% HPD CI: (1.9431,4.8199). c) 14.58.
1.2.5 a) $N(18.6207, 6.896)$. b) $O(H_1, H_0) = 11.71$, $O(H_1, H_0|x_1 - x_2) = 1.5 \times 10^{12}$.
1.2.6 a) $\psi = \sqrt{\theta}$. b) $\psi = \ln \frac{1-\sqrt{1-\theta}}{1+\sqrt{1-\theta}}$. c) $\psi = \ln \theta$.
1.2.7 $h(\theta) \propto [I(\theta)]^{1/2} = \frac{1}{\theta}$.
1.2.8 a) $h(\theta) \propto 1/\sqrt{\theta}$, $\theta > 0$.
1.2.9
1.2.10 a) Beta family. b) $E(\theta|\{x_i\}) > M_o(\theta|\{x_i\}) = \frac{a+mn-1}{a+b+mn+\sum_{i=1}^n x_i - 2}$ for $a+mn > 1$, $a+b+mn+\sum_{i=1}^n x_i > 1$. c) $p(y|\{x_i\}) = \frac{B(A+m, B+y)}{(m+y)B(m, y+1)B(A, B)}$, $A = a + mn$, $B = b + \sum_i x_i$.
1.2.11 a) 100γ% HPD CI: $\left[x_1, x_1\sqrt{(1-\gamma)^{-1}}\right]$. b) 1/3.
1.2.12 $n = 59$.
1.2.13 $n = 10$, $\bar{x} = 1.6$.

1.3 (Gentle, 2002), (Rizzo, 2019)

2.1 Look for them in Ross (2004/2009).

2.2 Look for them in Ross (2004/2009).

2.4 *2.4.2* $\widehat{\beta}_0 = -6.019$, $\widehat{\beta}_1 = 4.311$, $\widehat{\sigma}^2 = 0.067184$.

4.2 *4.2.2* a) The posterior marginal density of $\theta^{(m)}$, evaluated at each fixed point. b) Apply the SIR method.

4.3 *4.3.2* b) Calculate the conditional distribution function of Y given $U \le \pi(Y)/[Mp(Y)]$. c) Take into account a) and b).
4.3.3 Note any restrictions on parametric space.

5.2 *5.2.2* a) $\pi_i(u_i|u_j) \sim \text{Exp}(\alpha + \beta\, u_j)$, $i, j = 1, 2$, $i \ne j$. b) Note that the transition function from $u = (u_1, u_2)$ to $v = (v_1, v_2)$ in each cycle is $p(u, v) = \pi_1(v_1|u_2)\pi_2(v_2|v_1)$.
5.2.3 Note that the $q(u, v)$ entries of Q are such that

$$q(u,v) = \begin{cases} 1/10, & u = (0,0) \text{ e } v = (1,1), \\ 2/9, & u = (1,1) \text{ e } v = (0,0). \end{cases}$$

5.2.4 a) $\beta_0|\beta_1, \sigma^2, y \sim N(\widehat{\beta}_0 - \bar{x}(\beta_1 - \widehat{\beta}_1), \sigma^2/n)$, $\beta_1|\beta_0, \sigma^2, y \sim N(\widehat{\beta}_1 - \frac{n\,\bar{x}}{\sum_i x_i^2}(\beta_0 - \widehat{\beta}_0), \sigma^2/(\sum_i x_i^2))$, $\sigma^2|\beta, y \sim GaI(\frac{n}{2}, \frac{(n-2)\,s^2 + (\beta-\widehat{\beta})' X' X (\beta - \widehat{\beta})}{2})$. b) Negative response.
5.2.5 a) $\delta|\alpha, x \sim Ga(a+n, b + \sum_i x_i^\alpha)$, $h(\alpha|\delta, x) \propto \alpha^{n-1}(\prod_i x_i)^\alpha \exp\{-\frac{1}{2\,d}(\ln\alpha - c)^2 - \delta \sum_i x_i^\alpha\}$.
5.2.6 a) $\eta|\delta, \mu, \tau, D \sim Ga(\sum_i y_i, \sum_i \delta^{x_i})$, $D = \{(y_i, x_i)\}$; $h(\delta|\eta, \mu, \tau, D) \propto \delta^{\sum_i x_i\,y_i - 1} \exp(-\eta \sum_i \delta^{x_i})$; $\mu|\eta, \delta, \tau, D \sim N(\bar{x}, (n\,\tau)^{-1})$; $\tau|\eta, \delta, \mu, D \sim Ga(\frac{n}{2}, \frac{\sum_i (x_i-\bar{x})^2 + n(\mu-\bar{x})^2}{2})$.
5.2.7 a) $\mu|\{y_{ij}\}, \{\mu_i\}, \sigma_\epsilon^2, \sigma_\mu^2 \sim N(c, \frac{\sigma_0^2\,\sigma_\mu^2}{\sigma_\mu^2 + k\,\sigma_0^2})$, $c = \frac{\frac{k}{\sigma_\mu^2}\bar{\mu} + \frac{1}{\sigma_0^2}\mu}{\frac{k}{\sigma_\mu^2} + \frac{1}{\sigma_0^2}}$, $\bar{\mu} = \frac{1}{k}\sum_i \mu_i$. b) $\{\mu_i\}|\{y_{ij}\}, \mu, \sigma_\epsilon^2, \sigma_\mu^2 \underset{ind.}{\sim} N(A_i, \frac{\sigma_\epsilon^2\,\sigma_\mu^2}{\sigma_\epsilon^2 + m\,\sigma_\mu^2})$, $A_i = \frac{\frac{1}{\sigma_\mu^2}\mu + \frac{m}{\sigma_\epsilon^2}\bar{y}_{i\cdot}}{\frac{1}{\sigma_\mu^2} + \frac{m}{\sigma_\epsilon^2}}$. c) $\sigma_\epsilon^2|\{y_{ij}\}, \mu, \{\mu_i\}, \sigma_\mu^2 \sim GaI(a_1 + \frac{m\,k}{2}, b_1 + \frac{1}{2}\sum_{i,j}(y_{ij} - \mu_i)^2)$. d) $\sigma_\mu^2|\{y_{ij}\}, \mu, \{\mu_i\}, \sigma_\epsilon^2 \sim GaI(a_2 + \frac{k}{2}, b_2 + \frac{1}{2}\sum_i(\mu_i - \mu)^2)$.

Tables

Table T1: Binomial distribution function $F_X(x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$

n	$x \setminus p$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
1	0	0.9900	0.9800	0.9700	0.9600	0.9500	0.9400	0.9300	0.9200	0.9100	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0	0.9801	0.9604	0.9409	0.9216	0.9025	0.8836	0.8649	0.8464	0.8281	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9999	0.9996	0.9991	0.9984	0.9975	0.9964	0.9951	0.9936	0.9919	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.9703	0.9412	0.9127	0.8847	0.8574	0.8306	0.8044	0.7787	0.7536	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9997	0.9988	0.9974	0.9953	0.9928	0.9896	0.9860	0.9818	0.9772	0.9720	0.9393	0.8960	0.8438	0.7840	0.7183	0.6480	0.5748	0.5000
	2	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995	0.9993	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3				1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.9606	0.9224	0.8853	0.8493	0.8145	0.7807	0.7481	0.7164	0.6857	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9994	0.9977	0.9948	0.9909	0.9860	0.9801	0.9733	0.9656	0.9570	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	1.0000	1.0000	0.9999	0.9998	0.9995	0.9992	0.9987	0.9981	0.9973	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3			1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
	4									1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.9510	0.9039	0.8587	0.8154	0.7738	0.7339	0.6957	0.6591	0.6240	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.9990	0.9962	0.9915	0.9852	0.9774	0.9681	0.9575	0.9456	0.9326	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	1.0000	0.9999	0.9997	0.9994	0.9988	0.9980	0.9969	0.9955	0.9937	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3		1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4						1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
	5											1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.9415	0.8858	0.8330	0.7828	0.7351	0.6899	0.6470	0.6064	0.5679	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.9985	0.9943	0.9875	0.9784	0.9672	0.9541	0.9392	0.9227	0.9048	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	1.0000	0.9998	0.9995	0.9988	0.9978	0.9962	0.9942	0.9915	0.9882	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
	3		1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9995	0.9992	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6563
	4					1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5										1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
	6												1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	0.9321	0.8681	0.8080	0.7514	0.6983	0.6485	0.6017	0.5578	0.5168	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.9980	0.9921	0.9829	0.9706	0.9556	0.9382	0.9187	0.8974	0.8745	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
	2	1.0000	0.9997	0.9991	0.9980	0.9962	0.9937	0.9903	0.9860	0.9807	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
	3		1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9988	0.9982	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
	4				1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
	5								1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
	6											1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922
	7													1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	0	0.9227	0.8508	0.7837	0.7214	0.6634	0.6096	0.5596	0.5132	0.4703	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
	1	0.9973	0.9897	0.9777	0.9619	0.9428	0.9208	0.8965	0.8702	0.8423	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
	2	0.9999	0.9996	0.9987	0.9969	0.9942	0.9904	0.9853	0.9789	0.9711	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445
	3	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9987	0.9978	0.9966	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
	4			1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367
	5							1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555
	6											1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648
	7												1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
	8														1.0000	1.0000	1.0000	1.0000	1.0000
9	0	0.9135	0.8337	0.7602	0.6925	0.6302	0.5730	0.5204	0.4722	0.4279	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.9966	0.9869	0.9718	0.9522	0.9288	0.9022	0.8729	0.8417	0.8088	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195
	2	0.9999	0.9994	0.9980	0.9955	0.9916	0.9862	0.9791	0.9702	0.9595	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898
	3	1.0000	1.0000	0.9999	0.9997	0.9994	0.9987	0.9977	0.9963	0.9943	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539
	4			1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9995	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
	5						1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
	6										1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
	7												1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805
	8													1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
	9															1.0000	1.0000	1.0000	1.0000

n	$x \setminus p$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0	0.9044	0.8171	0.7374	0.6648	0.5987	0.5386	0.4840	0.4344	0.3894	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9957	0.9838	0.9655	0.9418	0.9139	0.8824	0.8483	0.8121	0.7746	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107
	2	0.9999	0.9991	0.9972	0.9938	0.9885	0.9812	0.9717	0.9599	0.9460	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	1.0000	1.0000	0.9999	0.9996	0.9990	0.9980	0.9964	0.9942	0.9912	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4			1.0000	1.0000	0.9999	0.9998	0.9997	0.9994	0.9990	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5					1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
	6									1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
	7											1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
	8												1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
	9														1.0000	1.0000	0.9999	0.9997	0.9990
	10																1.0000	1.0000	1.0000
11	0	0.8953	0.8007	0.7153	0.6382	0.5688	0.5063	0.4501	0.3996	0.3544	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
	1	0.9948	0.9805	0.9587	0.9308	0.8981	0.8618	0.8228	0.7819	0.7399	0.6974	0.4922	0.3221	0.1971	0.1130	0.0606	0.0302	0.0139	0.0059
	2	0.9998	0.9988	0.9963	0.9917	0.9848	0.9752	0.9630	0.9481	0.9305	0.9104	0.7788	0.6174	0.4552	0.3127	0.2001	0.1189	0.0652	0.0327
	3	1.0000	1.0000	0.9998	0.9993	0.9984	0.9970	0.9947	0.9915	0.9871	0.9815	0.9306	0.8389	0.7133	0.5696	0.4256	0.2963	0.1911	0.1133
	4			1.0000	1.0000	0.9999	0.9997	0.9995	0.9990	0.9983	0.9972	0.9841	0.9496	0.8854	0.7897	0.6683	0.5328	0.3971	0.2744
	5					1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9973	0.9883	0.9657	0.9218	0.8513	0.7535	0.6331	0.5000
	6								1.0000	1.0000	1.0000	0.9997	0.9980	0.9924	0.9784	0.9499	0.9006	0.8262	0.7256
	7											1.0000	0.9998	0.9988	0.9957	0.9878	0.9707	0.9390	0.8867
	8												1.0000	0.9999	0.9994	0.9980	0.9941	0.9852	0.9673
	9													1.0000	1.0000	0.9998	0.9993	0.9978	0.9941
	10															1.0000	1.0000	0.9998	0.9995
	11																	1.0000	1.0000
12	0	0.8864	0.7847	0.6938	0.6127	0.5404	0.4759	0.4186	0.3677	0.3225	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.9938	0.9769	0.9514	0.9191	0.8816	0.8405	0.7967	0.7513	0.7052	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
	2	0.9998	0.9985	0.9952	0.9893	0.9804	0.9684	0.9532	0.9348	0.9134	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
	3	1.0000	0.9999	0.9997	0.9990	0.9978	0.9957	0.9925	0.9880	0.9820	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
	4		1.0000	1.0000	0.9999	0.9998	0.9996	0.9991	0.9984	0.9973	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
	5				1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
	6							1.0000	1.0000	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
	7										1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
	8											1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
	9												1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10														1.0000	0.9999	0.9997	0.9989	0.9968
	11															1.0000	1.0000	0.9999	0.9998
	12																	1.0000	1.0000
13	0	0.8775	0.7690	0.6730	0.5882	0.5133	0.4474	0.3893	0.3383	0.2935	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001
	1	0.9928	0.9730	0.9436	0.9068	0.8646	0.8186	0.7702	0.7206	0.6707	0.6213	0.3983	0.2336	0.1267	0.0637	0.0296	0.0126	0.0049	0.0017
	2	0.9997	0.9980	0.9938	0.9865	0.9755	0.9608	0.9422	0.9201	0.8946	0.8661	0.6920	0.5017	0.3326	0.2025	0.1132	0.0579	0.0269	0.0112
	3	1.0000	0.9999	0.9995	0.9986	0.9969	0.9940	0.9897	0.9837	0.9758	0.9658	0.8820	0.7473	0.5843	0.4206	0.2783	0.1686	0.0929	0.0461
	4		1.0000	1.0000	0.9999	0.9997	0.9993	0.9987	0.9976	0.9959	0.9935	0.9658	0.9009	0.7940	0.6543	0.5005	0.3530	0.2279	0.1334
	5				1.0000	1.0000	0.9999	0.9999	0.9997	0.9995	0.9991	0.9925	0.9700	0.9198	0.8346	0.7159	0.5744	0.4268	0.2905
	6						1.0000	1.0000	1.0000	0.9999	0.9999	0.9987	0.9930	0.9757	0.9376	0.8705	0.7712	0.6437	0.5000
	7									1.0000	1.0000	0.9998	0.9988	0.9944	0.9818	0.9538	0.9023	0.8212	0.7095
	8											1.0000	0.9998	0.9990	0.9960	0.9874	0.9679	0.9302	0.8666
	9												1.0000	0.9999	0.9993	0.9975	0.9922	0.9797	0.9539
	10													1.0000	0.9999	0.9997	0.9987	0.9959	0.9888
	11														1.0000	1.0000	0.9999	0.9995	0.9983
	12																1.0000	1.0000	0.9999
	13																		1.0000
14	0	0.8687	0.7536	0.6528	0.5647	0.4877	0.4205	0.3620	0.3112	0.2670	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
	1	0.9916	0.9690	0.9355	0.8941	0.8470	0.7963	0.7436	0.6900	0.6368	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
	2	0.9997	0.9975	0.9923	0.9833	0.9699	0.9522	0.9302	0.9042	0.8745	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
	3	1.0000	0.9999	0.9994	0.9981	0.9958	0.9920	0.9864	0.9786	0.9685	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
	4		1.0000	1.0000	0.9998	0.9996	0.9990	0.9980	0.9965	0.9941	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
	5				1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120
	6						1.0000	1.0000	1.0000	0.9999	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953
	7									1.0000	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047
	8											1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880
	9												1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102
	10													1.0000	0.9998	0.9989	0.9961	0.9886	0.9713
	11														1.0000	0.9999	0.9994	0.9978	0.9935
	12															1.0000	0.9999	0.9997	0.9991
	13																1.0000	1.0000	0.9999
	14																		1.0000
15	0	0.8601	0.7386	0.6333	0.5421	0.4633	0.3953	0.3367	0.2863	0.2430	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.9904	0.9647	0.9270	0.8809	0.8290	0.7738	0.7168	0.6597	0.6035	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9996	0.9970	0.9906	0.9797	0.9638	0.9429	0.9171	0.8870	0.8531	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	1.0000	0.9998	0.9992	0.9976	0.9945	0.9896	0.9825	0.9727	0.9601	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
	4		1.0000	0.9999	0.9998	0.9994	0.9986	0.9972	0.9950	0.9918	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5			1.0000	1.0000	0.9999	0.9999	0.9997	0.9993	0.9987	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6					1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7								1.0000	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8											0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
	9											1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10												1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11													1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12														1.0000	0.9999	0.9997	0.9989	0.9963
	13															1.0000	1.0000	0.9999	0.9995
	14																	1.0000	1.0000

n	$x \setminus p$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
16	0	0.8515	0.7238	0.6143	0.5204	0.4401	0.3716	0.3131	0.2634	0.2211	0.1853	0.0743	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000
	1	0.9891	0.9601	0.9182	0.8673	0.8108	0.7511	0.6902	0.6299	0.5711	0.5147	0.2839	0.1407	0.0635	0.0261	0.0098	0.0033	0.0010	0.0003
	2	0.9995	0.9963	0.9887	0.9758	0.9571	0.9327	0.9031	0.8689	0.8306	0.7892	0.5614	0.3518	0.1971	0.0994	0.0451	0.0183	0.0066	0.0021
	3	1.0000	0.9998	0.9989	0.9968	0.9930	0.9868	0.9779	0.9658	0.9504	0.9316	0.7899	0.5981	0.4050	0.2459	0.1339	0.0651	0.0281	0.0106
	4		1.0000	0.9999	0.9997	0.9991	0.9981	0.9962	0.9932	0.9889	0.9830	0.9209	0.7982	0.6302	0.4499	0.2892	0.1666	0.0853	0.0384
	5			1.0000	1.0000	0.9999	0.9998	0.9995	0.9990	0.9981	0.9967	0.9765	0.9183	0.8103	0.6598	0.4900	0.3288	0.1976	0.1051
	6					1.0000	1.0000	0.9999	0.9999	0.9997	0.9995	0.9944	0.9733	0.9204	0.8247	0.6881	0.5272	0.3660	0.2272
	7							1.0000	1.0000	1.0000	0.9999	0.9989	0.9930	0.9729	0.9256	0.8406	0.7161	0.5629	0.4018
	8										1.0000	0.9998	0.9985	0.9925	0.9743	0.9329	0.8577	0.7441	0.5982
	9											1.0000	0.9998	0.9984	0.9929	0.9771	0.9417	0.8759	0.7728
	10												1.0000	0.9997	0.9984	0.9938	0.9809	0.9514	0.8949
	11													1.0000	0.9997	0.9987	0.9951	0.9851	0.9616
	12														1.0000	0.9998	0.9991	0.9965	0.9894
	13															1.0000	0.9999	0.9994	0.9979
	14																1.0000	0.9999	0.9997
	15																	1.0000	1.0000
17	0	0.8429	0.7093	0.5958	0.4996	0.4181	0.3493	0.2912	0.2423	0.2012	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
	1	0.9877	0.9554	0.9091	0.8535	0.7922	0.7283	0.6638	0.6005	0.5396	0.4818	0.2525	0.1182	0.0501	0.0193	0.0067	0.0021	0.0006	0.0001
	2	0.9994	0.9956	0.9866	0.9714	0.9497	0.9218	0.8882	0.8497	0.8073	0.7618	0.5198	0.3096	0.1637	0.0774	0.0327	0.0123	0.0041	0.0012
	3	1.0000	0.9997	0.9986	0.9960	0.9912	0.9836	0.9727	0.9581	0.9397	0.9174	0.7556	0.5489	0.3530	0.2019	0.1028	0.0464	0.0184	0.0064
	4		1.0000	0.9999	0.9996	0.9988	0.9974	0.9949	0.9911	0.9855	0.9779	0.9013	0.7582	0.5739	0.3887	0.2348	0.1260	0.0596	0.0245
	5			1.0000	1.0000	0.9999	0.9997	0.9993	0.9985	0.9973	0.9953	0.9681	0.8943	0.7653	0.5968	0.4197	0.2639	0.1471	0.0717
	6					1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9917	0.9623	0.8929	0.7752	0.6188	0.4478	0.2902	0.1662
	7							1.0000	1.0000	1.0000	0.9999	0.9983	0.9891	0.9598	0.8954	0.7872	0.6405	0.4743	0.3145
	8										1.0000	0.9997	0.9974	0.9876	0.9597	0.9006	0.8011	0.6626	0.5000
	9											1.0000	0.9995	0.9969	0.9873	0.9617	0.9081	0.8166	0.6855
	10												0.9999	0.9994	0.9968	0.9880	0.9652	0.9174	0.8338
	11												1.0000	0.9999	0.9993	0.9970	0.9894	0.9699	0.9283
	12													1.0000	0.9999	0.9994	0.9975	0.9914	0.9755
	13														1.0000	0.9999	0.9995	0.9981	0.9936
	14															1.0000	0.9999	0.9997	0.9988
	15																1.0000	1.0000	0.9999
	16																		1.0000
18	0	0.8345	0.6951	0.5780	0.4796	0.3972	0.3283	0.2708	0.2229	0.1831	0.1501	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000
	1	0.9862	0.9505	0.8997	0.8393	0.7735	0.7055	0.6378	0.5719	0.5091	0.4503	0.2241	0.0991	0.0395	0.0142	0.0046	0.0013	0.0003	0.0001
	2	0.9993	0.9948	0.9843	0.9667	0.9419	0.9102	0.8725	0.8298	0.7832	0.7338	0.4797	0.2713	0.1353	0.0600	0.0236	0.0082	0.0025	0.0007
	3	1.0000	0.9996	0.9982	0.9950	0.9891	0.9799	0.9667	0.9494	0.9277	0.9018	0.7202	0.5010	0.3057	0.1646	0.0783	0.0328	0.0120	0.0038
	4		1.0000	0.9998	0.9994	0.9985	0.9966	0.9933	0.9884	0.9814	0.9718	0.8794	0.7164	0.5187	0.3327	0.1886	0.0942	0.0411	0.0154
	5			1.0000	0.9999	0.9998	0.9995	0.9990	0.9979	0.9962	0.9936	0.9581	0.8671	0.7175	0.5344	0.3550	0.2088	0.1077	0.0481
	6				1.0000	1.0000	1.0000	0.9999	0.9997	0.9994	0.9988	0.9882	0.9487	0.8610	0.7217	0.5491	0.3743	0.2258	0.1189
	7							1.0000	1.0000	0.9999	0.9998	0.9973	0.9837	0.9431	0.8593	0.7283	0.5634	0.3915	0.2403
	8									1.0000	1.0000	0.9995	0.9957	0.9807	0.9404	0.8609	0.7368	0.5778	0.4073
	9											0.9999	0.9991	0.9946	0.9790	0.9403	0.8653	0.7473	0.5927
	10											1.0000	0.9998	0.9988	0.9939	0.9788	0.9424	0.8720	0.7597
	11												1.0000	0.9998	0.9986	0.9938	0.9797	0.9463	0.8811
	12													1.0000	0.9997	0.9986	0.9942	0.9817	0.9519
	13														1.0000	0.9997	0.9987	0.9951	0.9846
	14															1.0000	0.9998	0.9990	0.9962
	15																1.0000	0.9999	0.9993
	16																	1.0000	0.9999
	17																		1.0000
19	0	0.8262	0.6812	0.5606	0.4604	0.3774	0.3086	0.2519	0.2051	0.1666	0.1351	0.0456	0.0144	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000
	1	0.9847	0.9454	0.8900	0.8249	0.7547	0.6829	0.6121	0.5440	0.4798	0.4203	0.1985	0.0829	0.0310	0.0104	0.0031	0.0008	0.0002	0.0000
	2	0.9991	0.9939	0.9817	0.9616	0.9335	0.8979	0.8561	0.8092	0.7585	0.7054	0.4413	0.2369	0.1113	0.0462	0.0170	0.0055	0.0015	0.0004
	3	1.0000	0.9995	0.9978	0.9939	0.9868	0.9757	0.9602	0.9398	0.9147	0.8850	0.6841	0.4551	0.2631	0.1332	0.0591	0.0230	0.0077	0.0022
	4		1.0000	0.9998	0.9993	0.9980	0.9956	0.9915	0.9853	0.9765	0.9648	0.8556	0.6733	0.4654	0.2822	0.1500	0.0696	0.0280	0.0096
	5			1.0000	0.9999	0.9998	0.9994	0.9986	0.9971	0.9949	0.9914	0.9463	0.8369	0.6678	0.4739	0.2968	0.1629	0.0777	0.0318
	6				1.0000	1.0000	0.9999	0.9998	0.9996	0.9991	0.9983	0.9837	0.9324	0.8251	0.6655	0.4812	0.3081	0.1727	0.0835
	7						1.0000	1.0000	0.9999	0.9999	0.9997	0.9959	0.9767	0.9225	0.8180	0.6656	0.4878	0.3169	0.1796
	8								1.0000	1.0000	1.0000	0.9992	0.9933	0.9713	0.9161	0.8145	0.6675	0.4940	0.3238
	9											0.9999	0.9984	0.9911	0.9674	0.9125	0.8139	0.6710	0.5000
	10											1.0000	0.9997	0.9977	0.9895	0.9653	0.9115	0.8159	0.6762
	11												1.0000	0.9995	0.9972	0.9886	0.9648	0.9129	0.8204
	12													0.9999	0.9994	0.9969	0.9884	0.9658	0.9165
	13													1.0000	0.9999	0.9993	0.9969	0.9891	0.9682
	14														1.0000	0.9999	0.9994	0.9972	0.9904
	15															1.0000	0.9999	0.9995	0.9978
	16																1.0000	0.9999	0.9996
	17																	1.0000	1.0000
20	0	0.8179	0.6676	0.5438	0.4420	0.3585	0.2901	0.2342	0.1887	0.1516	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.9831	0.9401	0.8802	0.8103	0.7358	0.6605	0.5869	0.5169	0.4516	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
	2	0.9990	0.9929	0.9790	0.9561	0.9245	0.8850	0.8390	0.7879	0.7334	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002
	3	1.0000	0.9994	0.9973	0.9926	0.9841	0.9710	0.9529	0.9294	0.9007	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013
	4		1.0000	0.9997	0.9990	0.9974	0.9944	0.9893	0.9817	0.9710	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059
	5			1.0000	0.9999	0.9997	0.9991	0.9981	0.9962	0.9932	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207
	6				1.0000	1.0000	0.9999	0.9997	0.9994	0.9987	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
	7						1.0000	1.0000	0.9999	0.9998	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
	8								1.0000	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
	9										1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119
	10											1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881
	11												0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483
	12												1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
	13													1.0000	0.9997	0.9985	0.9935	0.9786	0.9423
	14														1.0000	0.9997	0.9984	0.9936	0.9793
	15															1.0000	0.9997	0.9985	0.9941
	16																1.0000	0.9997	0.9987
	17																	1.0000	0.9998
	18																		1.0000

Table T2: Poisson distribution function $F_X(x) = \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!}$

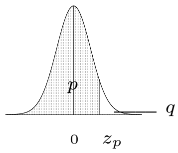
λ	x	0	1	2	3	4	5	6	7	8	9
0.01		0.9900	1.0000								
0.02		0.9802	0.9998	1.0000							
0.03		0.9704	0.9996	1.0000							
0.04		0.9608	0.9992	1.0000							
0.05		0.9512	0.9988	1.0000							
0.06		0.9418	0.9983	1.0000							
0.07		0.9324	0.9977	0.9999	1.0000						
0.08		0.9231	0.9970	0.9999	1.0000						
0.09		0.9139	0.9962	0.9999	1.0000						
0.10		0.9048	0.9953	0.9998	1.0000						
0.15		0.8607	0.9898	0.9995	1.0000						
0.20		0.8187	0.9825	0.9989	0.9999	1.0000					
0.25		0.7788	0.9735	0.9978	0.9999	1.0000					
0.30		0.7408	0.9631	0.9964	0.9997	1.0000					
0.35		0.7047	0.9513	0.9945	0.9995	1.0000					
0.40		0.6703	0.9384	0.9921	0.9992	0.9999	1.0000				
0.45		0.6376	0.9246	0.9891	0.9988	0.9999	1.0000				
0.50		0.6065	0.9098	0.9856	0.9982	0.9998	1.0000				
0.55		0.5769	0.8943	0.9815	0.9975	0.9997	1.0000				
0.60		0.5488	0.8781	0.9769	0.9966	0.9996	1.0000				
0.65		0.5220	0.8614	0.9717	0.9956	0.9994	0.9999	1.0000			
0.70		0.4966	0.8442	0.9659	0.9942	0.9992	0.9999	1.0000			
0.75		0.4724	0.8266	0.9595	0.9927	0.9989	0.9999	1.0000			
0.80		0.4493	0.8088	0.9526	0.9909	0.9986	0.9998	1.0000			
0.85		0.4274	0.7907	0.9451	0.9889	0.9982	0.9997	1.0000			
0.90		0.4066	0.7725	0.9371	0.9865	0.9977	0.9997	1.0000			
0.95		0.3867	0.7541	0.9287	0.9839	0.9971	0.9995	0.9999	1.0000		
1.00		0.3679	0.7358	0.9197	0.9810	0.9963	0.9994	0.9999	1.0000		
1.10		0.3329	0.6990	0.9004	0.9743	0.9946	0.9990	0.9999	1.0000		
1.20		0.3012	0.6626	0.8795	0.9662	0.9923	0.9985	0.9997	1.0000		
1.30		0.2725	0.6268	0.8571	0.9569	0.9893	0.9978	0.9996	0.9999	1.0000	
1.40		0.2466	0.5918	0.8335	0.9463	0.9857	0.9968	0.9994	0.9999	1.0000	
1.50		0.2231	0.5578	0.8088	0.9344	0.9814	0.9955	0.9991	0.9998	1.0000	
1.60		0.2019	0.5249	0.7834	0.9212	0.9763	0.9940	0.9987	0.9997	1.0000	
1.70		0.1827	0.4932	0.7572	0.9068	0.9704	0.9920	0.9981	0.9996	0.9999	1.0000
1.80		0.1653	0.4628	0.7306	0.8913	0.9636	0.9896	0.9974	0.9994	0.9999	1.0000
1.90		0.1496	0.4337	0.7037	0.8747	0.9559	0.9868	0.9966	0.9992	0.9998	1.0000
2.00	0	0.1353	0.4060	0.6767	0.8571	0.9473	0.9834	0.9955	0.9989	0.9998	1.0000
2.20	0	0.1108	0.3546	0.6227	0.8194	0.9275	0.9751	0.9925	0.9980	0.9995	0.9999
	10	1.0000									
2.40	0	0.0907	0.3084	0.5697	0.7787	0.9041	0.9643	0.9884	0.9967	0.9991	0.9998
	10	1.0000									
2.60	0	0.0743	0.2674	0.5184	0.7360	0.8774	0.9510	0.9828	0.9947	0.9985	0.9996
	10	0.9999	1.0000								
2.80	0	0.0608	0.2311	0.4695	0.6919	0.8477	0.9349	0.9756	0.9919	0.9976	0.9993
	10	0.9998	1.0000								
3.00	0	0.0498	0.1991	0.4232	0.6472	0.8153	0.9161	0.9665	0.9881	0.9962	0.9989
	10	0.9997	0.9999	1.0000							
3.20	0	0.0408	0.1712	0.3799	0.6025	0.7806	0.8946	0.9554	0.9832	0.9943	0.9982
	10	0.9995	0.9999	1.0000							
3.40	0	0.0334	0.1468	0.3397	0.5584	0.7442	0.8705	0.9421	0.9769	0.9917	0.9973
	10	0.9992	0.9998	0.9999	1.0000						
3.60	0	0.0273	0.1257	0.3027	0.5152	0.7064	0.8441	0.9267	0.9692	0.9883	0.9960
	10	0.9987	0.9996	0.9999	1.0000						
3.80	0	0.0224	0.1074	0.2689	0.4735	0.6678	0.8156	0.9091	0.9599	0.9840	0.9942
	10	0.9981	0.9994	0.9998	1.0000						
4.00	0	0.0183	0.0916	0.2381	0.4335	0.6288	0.7851	0.8893	0.9489	0.9786	0.9919
	10	0.9972	0.9991	0.9997	0.9999	1.0000					
4.20	0	0.0150	0.0780	0.2102	0.3954	0.5898	0.7531	0.8675	0.9361	0.9721	0.9889
	10	0.9959	0.9986	0.9996	0.9999	1.0000					
4.40	0	0.0123	0.0663	0.1851	0.3594	0.5512	0.7199	0.8436	0.9214	0.9642	0.9851
	10	0.9943	0.9980	0.9993	0.9998	0.9999	1.0000				
4.60	0	0.0101	0.0563	0.1626	0.3257	0.5132	0.6858	0.8180	0.9049	0.9549	0.9805
	10	0.9922	0.9971	0.9990	0.9997	0.9999	1.0000				
4.80	0	0.0082	0.0477	0.1425	0.2942	0.4763	0.6510	0.7908	0.8867	0.9442	0.9749
	10	0.9896	0.9960	0.9986	0.9995	0.9999	1.0000				
5.00	0	0.0067	0.0404	0.1247	0.2650	0.4405	0.6160	0.7622	0.8666	0.9319	0.9682
	10	0.9863	0.9945	0.9980	0.9993	0.9998	0.9999	1.0000			
5.20	0	0.0055	0.0342	0.1088	0.2381	0.4061	0.5809	0.7324	0.8449	0.9181	0.9603
	10	0.9823	0.9927	0.9972	0.9990	0.9997	0.9999	1.0000			
5.40	0	0.0045	0.0289	0.0948	0.2133	0.3733	0.5461	0.7017	0.8217	0.9027	0.9512
	10	0.9775	0.9904	0.9962	0.9986	0.9995	0.9998	0.9999	1.0000		
5.60	0	0.0037	0.0244	0.0824	0.1906	0.3422	0.5119	0.6703	0.7970	0.8857	0.9409
	10	0.9718	0.9875	0.9949	0.9980	0.9993	0.9998	0.9999	1.0000		
5.80	0	0.0030	0.0206	0.0715	0.1700	0.3127	0.4783	0.6384	0.7710	0.8672	0.9292
	10	0.9651	0.9841	0.9932	0.9973	0.9990	0.9996	0.9999	1.0000		

λ	x	0	1	2	3	4	5	6	7	8	9
6.00	0	0.0025	0.0174	0.0620	0.1512	0.2851	0.4457	0.6063	0.7440	0.8472	0.9161
	10	0.9574	0.9799	0.9912	0.9964	0.9986	0.9995	0.9998	0.9999	1.0000	
6.20	0	0.0020	0.0146	0.0536	0.1342	0.2592	0.4141	0.5742	0.7160	0.8259	0.9016
	10	0.9486	0.9750	0.9887	0.9952	0.9981	0.9993	0.9997	0.9999	1.0000	
6.40	0	0.0017	0.0123	0.0463	0.1189	0.2351	0.3837	0.5423	0.6873	0.8033	0.8858
	10	0.9386	0.9693	0.9857	0.9937	0.9974	0.9990	0.9996	0.9999	1.0000	
6.60	0	0.0014	0.0103	0.0400	0.1052	0.2127	0.3547	0.5108	0.6581	0.7796	0.8686
	10	0.9274	0.9627	0.9821	0.9920	0.9966	0.9986	0.9995	0.9998	0.9999	1.0000
6.80	0	0.0011	0.0087	0.0344	0.0928	0.1920	0.3270	0.4799	0.6285	0.7548	0.8502
	10	0.9151	0.9552	0.9779	0.9898	0.9956	0.9982	0.9993	0.9997	0.9999	1.0000
7.00	0	0.0009	0.0073	0.0296	0.0818	0.1730	0.3007	0.4497	0.5987	0.7291	0.8305
	10	0.9015	0.9467	0.9730	0.9872	0.9943	0.9976	0.9990	0.9996	0.9999	1.0000
7.20	0	0.0007	0.0061	0.0255	0.0719	0.1555	0.2759	0.4204	0.5689	0.7027	0.8096
	10	0.8867	0.9371	0.9673	0.9841	0.9927	0.9969	0.9987	0.9995	0.9998	0.9999
	20	1.0000									
7.40	0	0.0006	0.0051	0.0219	0.0632	0.1395	0.2526	0.3920	0.5393	0.6757	0.7877
	10	0.8707	0.9265	0.9609	0.9805	0.9908	0.9959	0.9983	0.9993	0.9997	0.9999
	20	1.0000									
7.60	0	0.0005	0.0043	0.0188	0.0554	0.1249	0.2307	0.3646	0.5100	0.6482	0.7649
	10	0.8535	0.9148	0.9536	0.9762	0.9886	0.9948	0.9978	0.9991	0.9996	0.9999
	20	1.0000									
7.80	0	0.0004	0.0036	0.0161	0.0485	0.1117	0.2103	0.3384	0.4812	0.6204	0.7411
	10	0.8352	0.9020	0.9454	0.9714	0.9859	0.9934	0.9971	0.9988	0.9995	0.9998
	20	0.9999	1.0000								
8.00	0	0.0003	0.0030	0.0138	0.0424	0.0996	0.1912	0.3134	0.4530	0.5925	0.7166
	10	0.8159	0.8881	0.9362	0.9658	0.9827	0.9918	0.9963	0.9984	0.9993	0.9997
	20	0.9999	1.0000								
8.20	0	0.0003	0.0025	0.0118	0.0370	0.0887	0.1736	0.2896	0.4254	0.5647	0.6915
	10	0.7955	0.8731	0.9261	0.9595	0.9791	0.9898	0.9953	0.9979	0.9991	0.9997
	20	0.9999	1.0000								
8.40	0	0.0002	0.0021	0.0100	0.0323	0.0789	0.1573	0.2670	0.3987	0.5369	0.6659
	10	0.7743	0.8571	0.9150	0.9524	0.9749	0.9875	0.9941	0.9973	0.9989	0.9995
	20	0.9998	0.9999	1.0000							
8.60	0	0.0002	0.0018	0.0086	0.0281	0.0701	0.1422	0.2457	0.3728	0.5094	0.6400
	10	0.7522	0.8400	0.9029	0.9445	0.9701	0.9848	0.9926	0.9966	0.9985	0.9994
	20	0.9998	0.9999	1.0000							
8.80	0	0.0002	0.0015	0.0073	0.0244	0.0621	0.1284	0.2256	0.3478	0.4823	0.6137
	10	0.7294	0.8220	0.8898	0.9358	0.9647	0.9816	0.9909	0.9957	0.9981	0.9992
	20	0.9997	0.9999	1.0000							
9.00	0	0.0001	0.0012	0.0062	0.0212	0.0550	0.1157	0.2068	0.3239	0.4557	0.5874
	10	0.7060	0.8030	0.8758	0.9261	0.9585	0.9780	0.9889	0.9947	0.9976	0.9989
	20	0.9996	0.9998	0.9999	1.0000						
9.20	0	0.0001	0.0010	0.0053	0.0184	0.0486	0.1041	0.1892	0.3010	0.4296	0.5611
	10	0.6820	0.7832	0.8607	0.9156	0.9517	0.9738	0.9865	0.9934	0.9969	0.9986
	20	0.9994	0.9998	0.9999	1.0000						
9.40	0	0.0001	0.0009	0.0045	0.0160	0.0429	0.0935	0.1727	0.2792	0.4042	0.5349
	10	0.6576	0.7626	0.8448	0.9042	0.9441	0.9691	0.9838	0.9919	0.9962	0.9983
	20	0.9992	0.9997	0.9999	1.0000						
9.60	0	0.0001	0.0007	0.0038	0.0138	0.0378	0.0838	0.1574	0.2584	0.3796	0.5089
	10	0.6329	0.7412	0.8279	0.8919	0.9357	0.9638	0.9806	0.9902	0.9952	0.9978
	20	0.9990	0.9996	0.9998	0.9999	1.0000					
9.80	0	0.0001	0.0006	0.0033	0.0120	0.0333	0.0750	0.1433	0.2388	0.3558	0.4832
	10	0.6080	0.7193	0.8101	0.8786	0.9265	0.9579	0.9770	0.9881	0.9941	0.9972
	20	0.9987	0.9995	0.9998	0.9999	1.0000					
10.00	0	0.0000	0.0005	0.0028	0.0103	0.0293	0.0671	0.1301	0.2202	0.3328	0.4579
	10	0.5830	0.6968	0.7916	0.8645	0.9165	0.9513	0.9730	0.9857	0.9928	0.9965
	20	0.9984	0.9993	0.9997	0.9999	1.0000					
10.50	0	0.0000	0.0003	0.0018	0.0071	0.0211	0.0504	0.1016	0.1785	0.2794	0.3971
	10	0.5207	0.6387	0.7420	0.8253	0.8879	0.9317	0.9604	0.9781	0.9885	0.9942
	20	0.9972	0.9987	0.9994	0.9998	0.9999	1.0000				
11.00	0	0.0000	0.0002	0.0012	0.0049	0.0151	0.0375	0.0786	0.1432	0.2320	0.3405
	10	0.4599	0.5793	0.6887	0.7813	0.8540	0.9074	0.9441	0.9678	0.9823	0.9907
	20	0.9953	0.9977	0.9990	0.9995	0.9998	0.9999	1.0000			
11.50	0	0.0000	0.0001	0.0008	0.0034	0.0107	0.0277	0.0603	0.1137	0.1906	0.2888
	10	0.4017	0.5198	0.6329	0.7330	0.8153	0.8783	0.9236	0.9542	0.9738	0.9857
	20	0.9925	0.9962	0.9982	0.9992	0.9996	0.9998	0.9999	1.0000		
12.00	0	0.0000	0.0001	0.0005	0.0023	0.0076	0.0203	0.0458	0.0895	0.1550	0.2424
	10	0.3472	0.4616	0.5760	0.6815	0.7720	0.8444	0.8987	0.9370	0.9626	0.9787
	20	0.9884	0.9939	0.9970	0.9985	0.9993	0.9997	0.9999	0.9999	1.0000	
12.50	0	0.0000	0.0001	0.0003	0.0016	0.0053	0.0148	0.0346	0.0698	0.1249	0.2014
	10	0.2971	0.4058	0.5190	0.6278	0.7250	0.8060	0.8693	0.9158	0.9481	0.9694
	20	0.9827	0.9906	0.9951	0.9975	0.9988	0.9994	0.9997	0.9999	1.0000	
13.00	0	0.0000	0.0000	0.0002	0.0011	0.0037	0.0107	0.0259	0.0540	0.0998	0.1658
	10	0.2517	0.3532	0.4631	0.5730	0.6751	0.7636	0.8355	0.8905	0.9302	0.9573
	20	0.9750	0.9859	0.9924	0.9960	0.9980	0.9990	0.9995	0.9998	0.9999	1.0000
13.50	0	0.0000	0.0000	0.0001	0.0007	0.0026	0.0077	0.0193	0.0415	0.0790	0.1353
	10	0.2112	0.3045	0.4093	0.5182	0.6233	0.7178	0.7975	0.8609	0.9084	0.9421
	20	0.9649	0.9796	0.9885	0.9938	0.9968	0.9984	0.9992	0.9996	0.9998	0.9999
	30	1.0000									

λ	x	0	1	2	3	4	5	6	7	8	9
14.00	0	0.0000	0.0000	0.0001	0.0005	0.0018	0.0055	0.0142	0.0316	0.0621	0.1094
	10	0.1757	0.2600	0.3585	0.4644	0.5704	0.6694	0.7559	0.8272	0.8826	0.9235
	20	0.9521	0.9712	0.9833	0.9907	0.9950	0.9974	0.9987	0.9994	0.9997	0.9999
	30	0.9999	1.0000								
14.50	0	0.0000	0.0000	0.0001	0.0003	0.0012	0.0039	0.0105	0.0239	0.0484	0.0878
	10	0.1449	0.2201	0.3111	0.4125	0.5176	0.6192	0.7112	0.7897	0.8530	0.9012
	20	0.9362	0.9604	0.9763	0.9863	0.9924	0.9959	0.9979	0.9989	0.9995	0.9998
	30	0.9999	1.0000								
15.00	0	0.0000	0.0000	0.0000	0.0002	0.0009	0.0028	0.0076	0.0180	0.0374	0.0699
	10	0.1185	0.1848	0.2676	0.3632	0.4657	0.5681	0.6641	0.7489	0.8195	0.8752
	20	0.9170	0.9469	0.9673	0.9805	0.9888	0.9938	0.9967	0.9983	0.9991	0.9996
	30	0.9998	0.9999	1.0000							
16.00	0	0.0000	0.0000	0.0000	0.0001	0.0004	0.0014	0.0040	0.0100	0.0220	0.0433
	10	0.0774	0.1270	0.1931	0.2745	0.3675	0.4667	0.5660	0.6593	0.7423	0.8122
	20	0.8682	0.9108	0.9418	0.9633	0.9777	0.9869	0.9925	0.9959	0.9978	0.9989
	30	0.9994	0.9997	0.9999	0.9999	1.0000					
17.00	0	0.0000	0.0000	0.0000	0.0000	0.0002	0.0007	0.0021	0.0054	0.0126	0.0261
	10	0.0491	0.0847	0.1350	0.2009	0.2808	0.3715	0.4677	0.5640	0.6550	0.7363
	20	0.8055	0.8615	0.9047	0.9367	0.9594	0.9748	0.9848	0.9912	0.9950	0.9973
	30	0.9986	0.9993	0.9996	0.9998	0.9999	1.0000				
18.00	0	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0.0029	0.0071	0.0154
	10	0.0304	0.0549	0.0917	0.1426	0.2081	0.2867	0.3751	0.4686	0.5622	0.6509
	20	0.7307	0.7991	0.8551	0.8989	0.9317	0.9554	0.9718	0.9827	0.9897	0.9941
	30	0.9967	0.9982	0.9990	0.9995	0.9998	0.9999	0.9999	1.0000		
19.00	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0005	0.0015	0.0039	0.0089
	10	0.0183	0.0347	0.0606	0.0984	0.1497	0.2148	0.2920	0.3784	0.4695	0.5606
	20	0.6472	0.7255	0.7931	0.8490	0.8933	0.9269	0.9514	0.9687	0.9805	0.9882
	30	0.9930	0.9960	0.9978	0.9988	0.9994	0.9997	0.9998	0.9999	1.0000	
20.00	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008	0.0021	0.0050
	10	0.0108	0.0214	0.0390	0.0661	0.1049	0.1565	0.2211	0.2970	0.3814	0.4703
	20	0.5591	0.6437	0.7206	0.7875	0.8432	0.8878	0.9221	0.9475	0.9657	0.9782
	30	0.9865	0.9919	0.9953	0.9973	0.9985	0.9992	0.9996	0.9998	0.9999	0.9999
	40	1.0000									
21.00	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0011	0.0028
	10	0.0063	0.0129	0.0245	0.0434	0.0716	0.1111	0.1629	0.2270	0.3017	0.3843
	20	0.4710	0.5577	0.6405	0.7160	0.7822	0.8377	0.8826	0.9175	0.9436	0.9626
	30	0.9758	0.9848	0.9907	0.9945	0.9968	0.9982	0.9990	0.9995	0.9997	0.9999
	40	0.9999	1.0000								
22.00	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0006	0.0015
	10	0.0035	0.0076	0.0151	0.0278	0.0477	0.0769	0.1170	0.1690	0.2325	0.3060
	20	0.3869	0.4716	0.5564	0.6374	0.7117	0.7771	0.8324	0.8775	0.9129	0.9398
	30	0.9595	0.9735	0.9831	0.9895	0.9936	0.9962	0.9978	0.9988	0.9993	0.9996
	40	0.9998	0.9999	1.0000							
23.00	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008
	10	0.0020	0.0044	0.0091	0.0174	0.0311	0.0520	0.0821	0.1228	0.1748	0.2377
	20	0.3101	0.3894	0.4723	0.5551	0.6346	0.7077	0.7723	0.8274	0.8726	0.9085
	30	0.9360	0.9564	0.9711	0.9813	0.9882	0.9927	0.9956	0.9974	0.9985	0.9992
	40	0.9996	0.9998	0.9999	0.9999	1.0000					
24.00	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0004
	10	0.0011	0.0025	0.0054	0.0107	0.0198	0.0344	0.0563	0.0871	0.1283	0.1803
	20	0.2426	0.3139	0.3917	0.4728	0.5540	0.6319	0.7038	0.7677	0.8225	0.8679
	30	0.9042	0.9322	0.9533	0.9686	0.9794	0.9868	0.9918	0.9950	0.9970	0.9983
	40	0.9990	0.9995	0.9997	0.9998	0.9999	1.0000				
25.00	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002
	10	0.0006	0.0014	0.0031	0.0065	0.0124	0.0223	0.0377	0.0605	0.0920	0.1336
	20	0.1855	0.2473	0.3175	0.3939	0.4734	0.5529	0.6294	0.7002	0.7634	0.8179
	30	0.8633	0.8999	0.9285	0.9502	0.9662	0.9775	0.9854	0.9908	0.9943	0.9966
	40	0.9980	0.9988	0.9993	0.9996	0.9998	0.9999	0.9999	1.0000		
30.00	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	10	0.0000	0.0001	0.0002	0.0004	0.0009	0.0019	0.0039	0.0073	0.0129	0.0219
	20	0.0353	0.0544	0.0806	0.1146	0.1572	0.2084	0.2673	0.3329	0.4031	0.4757
	30	0.5484	0.6186	0.6845	0.7444	0.7973	0.8426	0.8804	0.9110	0.9352	0.9537
	40	0.9677	0.9779	0.9852	0.9903	0.9937	0.9960	0.9975	0.9985	0.9991	0.9995
	50	0.9997	0.9998	0.9999	0.9999	1.0000					
35.00	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0006	0.0012	0.0023
	20	0.0043	0.0076	0.0128	0.0208	0.0324	0.0486	0.0705	0.0988	0.1343	0.1770
	30	0.2269	0.2833	0.3449	0.4102	0.4775	0.5448	0.6102	0.6721	0.7291	0.7802
	40	0.8249	0.8631	0.8950	0.9209	0.9415	0.9575	0.9697	0.9788	0.9854	0.9902
	50	0.9935	0.9957	0.9973	0.9983	0.9989	0.9993	0.9996	0.9998	0.9999	0.9999
	60	1.0000									
40.00	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002
	20	0.0004	0.0007	0.0014	0.0026	0.0045	0.0076	0.0123	0.0193	0.0294	0.0432
	30	0.0617	0.0855	0.1153	0.1514	0.1939	0.2424	0.2963	0.3547	0.4160	0.4790
	40	0.5419	0.6033	0.6618	0.7162	0.7657	0.8097	0.8479	0.8804	0.9075	0.9297
	50	0.9474	0.9613	0.9719	0.9800	0.9860	0.9903	0.9934	0.9956	0.9971	0.9981
	60	0.9988	0.9992	0.9995	0.9997	0.9998	0.9999	0.9999	1.0000		

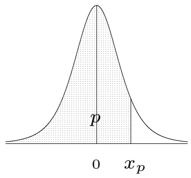
Table T3: Quantile of the Normal distribution function $Z \sim N(0,1)$

$$z_p = \Phi^{-1}(p) = \Phi^{-1}(1-q)$$



q	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.010	
0.00	∞	3.0902	2.8782	2.7478	2.6521	2.5758	2.5121	2.4573	2.4089	2.3656	2.3263	0.99
0.01	2.3263	2.2904	2.2571	2.2262	2.1973	2.1701	2.1444	2.1201	2.0969	2.0748	2.0537	0.98
0.02	2.0537	2.0335	2.0141	1.9954	1.9774	1.9600	1.9431	1.9268	1.9110	1.8957	1.8808	0.97
0.03	1.8808	1.8663	1.8522	1.8384	1.8250	1.8119	1.7991	1.7866	1.7744	1.7624	1.7507	0.96
0.04	1.7507	1.7392	1.7279	1.7169	1.7060	1.6954	1.6849	1.6747	1.6646	1.6546	1.6449	0.95
0.05	1.6449	1.6352	1.6258	1.6164	1.6072	1.5982	1.5893	1.5805	1.5718	1.5632	1.5548	0.94
0.06	1.5548	1.5464	1.5382	1.5301	1.5220	1.5141	1.5063	1.4985	1.4909	1.4833	1.4758	0.93
0.07	1.4758	1.4684	1.4611	1.4538	1.4466	1.4395	1.4325	1.4255	1.4187	1.4118	1.4051	0.92
0.08	1.4051	1.3984	1.3917	1.3852	1.3787	1.3722	1.3658	1.3595	1.3532	1.3469	1.3408	0.91
0.09	1.3408	1.3346	1.3285	1.3225	1.3165	1.3106	1.3047	1.2988	1.2930	1.2873	1.2816	0.90
0.10	1.2816	1.2759	1.2702	1.2646	1.2591	1.2536	1.2481	1.2426	1.2372	1.2319	1.2265	0.89
0.11	1.2265	1.2212	1.2160	1.2107	1.2055	1.2004	1.1952	1.1901	1.1850	1.1800	1.1750	0.88
0.12	1.1750	1.1700	1.1650	1.1601	1.1552	1.1503	1.1455	1.1407	1.1359	1.1311	1.1264	0.87
0.13	1.1264	1.1217	1.1170	1.1123	1.1077	1.1031	1.0985	1.0939	1.0893	1.0848	1.0803	0.86
0.14	1.0803	1.0758	1.0714	1.0669	1.0625	1.0581	1.0537	1.0494	1.0451	1.0407	1.0364	0.85
0.15	1.0364	1.0322	1.0279	1.0237	1.0194	1.0152	1.0110	1.0069	1.0027	0.9986	0.9945	0.84
0.16	0.9945	0.9904	0.9863	0.9822	0.9782	0.9741	0.9701	0.9661	0.9621	0.9581	0.9542	0.83
0.17	0.9542	0.9502	0.9463	0.9424	0.9385	0.9346	0.9307	0.9269	0.9230	0.9192	0.9154	0.82
0.18	0.9154	0.9116	0.9078	0.9040	0.9002	0.8965	0.8927	0.8890	0.8853	0.8816	0.8779	0.81
0.19	0.8779	0.8742	0.8706	0.8669	0.8632	0.8596	0.8560	0.8524	0.8488	0.8452	0.8416	0.80
0.20	0.8416	0.8381	0.8345	0.8310	0.8274	0.8239	0.8204	0.8169	0.8134	0.8099	0.8064	0.79
0.21	0.8064	0.8030	0.7995	0.7961	0.7926	0.7892	0.7858	0.7824	0.7790	0.7756	0.7722	0.78
0.22	0.7722	0.7688	0.7655	0.7621	0.7588	0.7554	0.7521	0.7488	0.7454	0.7421	0.7388	0.77
0.23	0.7388	0.7356	0.7323	0.7290	0.7257	0.7225	0.7192	0.7160	0.7128	0.7095	0.7063	0.76
0.24	0.7063	0.7031	0.6999	0.6967	0.6935	0.6903	0.6871	0.6840	0.6808	0.6776	0.6745	0.75
0.25	0.6745	0.6713	0.6682	0.6651	0.6620	0.6588	0.6557	0.6526	0.6495	0.6464	0.6433	0.74
0.26	0.6433	0.6403	0.6372	0.6341	0.6311	0.6280	0.6250	0.6219	0.6189	0.6158	0.6128	0.73
0.27	0.6128	0.6098	0.6068	0.6038	0.6008	0.5978	0.5948	0.5918	0.5888	0.5858	0.5828	0.72
0.28	0.5828	0.5799	0.5769	0.5740	0.5710	0.5681	0.5651	0.5622	0.5592	0.5563	0.5534	0.71
0.29	0.5534	0.5505	0.5476	0.5446	0.5417	0.5388	0.5359	0.5330	0.5302	0.5273	0.5244	0.70
0.30	0.5244	0.5215	0.5187	0.5158	0.5129	0.5101	0.5072	0.5044	0.5015	0.4987	0.4958	0.69
0.31	0.4958	0.4930	0.4902	0.4874	0.4845	0.4817	0.4789	0.4761	0.4733	0.4705	0.4677	0.68
0.32	0.4677	0.4649	0.4621	0.4593	0.4565	0.4538	0.4510	0.4482	0.4454	0.4427	0.4399	0.67
0.33	0.4399	0.4372	0.4344	0.4316	0.4289	0.4261	0.4234	0.4207	0.4179	0.4152	0.4125	0.66
0.34	0.4125	0.4097	0.4070	0.4043	0.4016	0.3989	0.3961	0.3934	0.3907	0.3880	0.3853	0.65
0.35	0.3853	0.3826	0.3799	0.3772	0.3745	0.3719	0.3692	0.3665	0.3638	0.3611	0.3585	0.64
0.36	0.3585	0.3558	0.3531	0.3505	0.3478	0.3451	0.3425	0.3398	0.3372	0.3345	0.3319	0.63
0.37	0.3319	0.3292	0.3266	0.3239	0.3213	0.3186	0.3160	0.3134	0.3107	0.3081	0.3055	0.62
0.38	0.3055	0.3029	0.3002	0.2976	0.2950	0.2924	0.2898	0.2871	0.2845	0.2819	0.2793	0.61
0.39	0.2793	0.2767	0.2741	0.2715	0.2689	0.2663	0.2637	0.2611	0.2585	0.2559	0.2533	0.60
0.40	0.2533	0.2508	0.2482	0.2456	0.2430	0.2404	0.2378	0.2353	0.2327	0.2301	0.2275	0.59
0.41	0.2275	0.2250	0.2224	0.2198	0.2173	0.2147	0.2121	0.2096	0.2070	0.2045	0.2019	0.58
0.42	0.2019	0.1993	0.1968	0.1942	0.1917	0.1891	0.1866	0.1840	0.1815	0.1789	0.1764	0.57
0.43	0.1764	0.1738	0.1713	0.1687	0.1662	0.1637	0.1611	0.1586	0.1560	0.1535	0.1510	0.56
0.44	0.1510	0.1484	0.1459	0.1434	0.1408	0.1383	0.1358	0.1332	0.1307	0.1282	0.1257	0.55
0.45	0.1257	0.1231	0.1206	0.1181	0.1156	0.1130	0.1105	0.1080	0.1055	0.1030	0.1004	0.54
0.46	0.1004	0.0979	0.0954	0.0929	0.0904	0.0878	0.0853	0.0828	0.0803	0.0778	0.0753	0.53
0.47	0.0753	0.0728	0.0702	0.0677	0.0652	0.0627	0.0602	0.0577	0.0552	0.0527	0.0502	0.52
0.48	0.0502	0.0476	0.0451	0.0426	0.0401	0.0376	0.0351	0.0326	0.0301	0.0276	0.0251	0.51
0.49	0.0251	0.0226	0.0201	0.0175	0.0150	0.0125	0.0100	0.0075	0.0050	0.0025	0.0000	0.50
	0.010	0.009	0.008	0.007	0.006	0.005	0.004	0.003	0.002	0.001	0.000	p

Table T4: Quantile of the Student's t-distribution function $X \sim t_{(n)} : x_p = F_X^{-1}(p)$



$n \setminus p$	0.6	0.7	0.75	0.8	0.85	0.9	0.925	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.325	0.727	1.000	1.376	1.963	3.078	4.165	6.314	12.706	31.821	63.656	318.289	636.578
2	0.289	0.617	0.816	1.061	1.386	1.886	2.282	2.920	4.303	6.965	9.925	22.328	31.600
3	0.277	0.584	0.765	0.978	1.250	1.638	1.924	2.353	3.182	4.541	5.841	10.214	12.924
4	0.271	0.569	0.741	0.941	1.190	1.533	1.778	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.559	0.727	0.920	1.156	1.476	1.699	2.015	2.571	3.365	4.032	5.894	6.869
6	0.265	0.553	0.718	0.906	1.134	1.440	1.650	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.549	0.711	0.896	1.119	1.415	1.617	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.546	0.706	0.889	1.108	1.397	1.592	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.543	0.703	0.883	1.100	1.383	1.574	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.542	0.700	0.879	1.093	1.372	1.559	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.540	0.697	0.876	1.088	1.363	1.548	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.539	0.695	0.873	1.083	1.356	1.538	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.538	0.694	0.870	1.079	1.350	1.530	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.537	0.692	0.868	1.076	1.345	1.523	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.536	0.691	0.866	1.074	1.341	1.517	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.535	0.690	0.865	1.071	1.337	1.512	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.534	0.689	0.863	1.069	1.333	1.508	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.534	0.688	0.862	1.067	1.330	1.504	1.734	2.101	2.552	2.878	3.610	3.922
19	0.257	0.533	0.688	0.861	1.066	1.328	1.500	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.533	0.687	0.860	1.064	1.325	1.497	1.725	2.086	2.528	2.845	3.552	3.850
21	0.257	0.532	0.686	0.859	1.063	1.323	1.494	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.532	0.686	0.858	1.061	1.321	1.492	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.532	0.685	0.858	1.060	1.319	1.489	1.714	2.069	2.500	2.807	3.485	3.768
24	0.256	0.531	0.685	0.857	1.059	1.318	1.487	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.531	0.684	0.856	1.058	1.316	1.485	1.708	2.060	2.485	2.787	3.450	3.725
26	0.256	0.531	0.684	0.856	1.058	1.315	1.483	1.706	2.056	2.479	2.779	3.435	3.707
27	0.256	0.531	0.684	0.855	1.057	1.314	1.482	1.703	2.052	2.473	2.771	3.421	3.689
28	0.256	0.530	0.683	0.855	1.056	1.313	1.480	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.530	0.683	0.854	1.055	1.311	1.479	1.699	2.045	2.462	2.756	3.396	3.660
30	0.256	0.530	0.683	0.854	1.055	1.310	1.477	1.697	2.042	2.457	2.750	3.385	3.646
40	0.255	0.529	0.681	0.851	1.050	1.303	1.468	1.684	2.021	2.423	2.704	3.307	3.551
45	0.255	0.528	0.680	0.850	1.049	1.301	1.465	1.679	2.014	2.412	2.690	3.281	3.520
50	0.255	0.528	0.679	0.849	1.047	1.299	1.462	1.676	2.009	2.403	2.678	3.261	3.496
60	0.254	0.527	0.679	0.848	1.045	1.296	1.458	1.671	2.000	2.390	2.660	3.232	3.460
70	0.254	0.527	0.678	0.847	1.044	1.294	1.456	1.667	1.994	2.381	2.648	3.211	3.435
80	0.254	0.526	0.678	0.846	1.043	1.292	1.453	1.664	1.990	2.374	2.639	3.195	3.416
90	0.254	0.526	0.677	0.846	1.042	1.291	1.452	1.662	1.987	2.368	2.632	3.183	3.402
100	0.254	0.526	0.677	0.845	1.042	1.290	1.451	1.660	1.984	2.364	2.626	3.174	3.390
120	0.254	0.526	0.677	0.845	1.041	1.289	1.449	1.658	1.980	2.358	2.617	3.160	3.373
150	0.254	0.526	0.676	0.844	1.040	1.287	1.447	1.655	1.976	2.351	2.609	3.145	3.357
∞	0.253	0.524	0.675	0.842	1.036	1.282	1.440	1.645	1.960	2.327	2.576	3.091	3.291

Table T5: Quantile of the chi-square distribution function $X \sim \chi^2_{(n)} : x_p = F_X^{-1}(p)$

$n \setminus p$	0.0005	0.001	0.005	0.01	0.025	0.05	0.075	0.10	0.15	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.85	0.90	0.925	0.950	0.975	0.990	0.995	0.999	0.9995
1	3.9E-07	1.6E-06	3.9E-05	0.0002	0.0010	0.0039	0.0089	0.0158	0.0358	0.0642	0.148	0.275	0.455	0.708	1.074	1.642	2.072	2.706	3.170	3.841	5.024	6.635	7.879	10.83	12.12
2	0.0010	0.0020	0.0100	0.0201	0.0506	0.103	0.156	0.211	0.325	0.446	0.713	1.022	1.386	1.833	2.408	3.219	3.794	4.605	5.181	5.991	7.378	9.210	10.60	13.82	15.20
3	0.0153	0.0243	0.0717	0.115	0.216	0.352	0.472	0.584	0.798	1.005	1.424	1.869	2.366	2.946	3.665	4.642	5.317	6.251	6.905	7.815	9.348	11.34	12.84	16.27	17.73
4	0.0639	0.0908	0.207	0.297	0.484	0.711	0.897	1.064	1.366	1.649	2.195	2.753	3.357	4.045	4.878	5.989	6.745	7.779	8.496	9.488	11.14	13.28	14.86	18.47	20.00
5	0.158	0.210	0.412	0.554	0.831	1.145	1.394	1.610	1.994	2.343	3.000	3.656	4.351	5.132	6.064	7.289	8.115	9.236	10.01	11.07	12.83	15.09	16.75	20.51	22.11
6	0.299	0.381	0.676	0.872	1.237	1.635	1.941	2.204	2.661	3.070	3.828	4.570	5.348	6.211	7.231	8.558	9.446	10.64	11.47	12.59	14.45	16.81	18.55	22.46	24.10
7	0.485	0.599	0.989	1.239	1.690	2.167	2.528	2.833	3.358	3.822	4.671	5.493	6.346	7.283	8.383	9.803	10.75	12.02	12.88	14.07	16.01	18.48	20.28	24.32	26.02
8	0.710	0.857	1.344	1.647	2.180	2.733	3.144	3.490	4.078	4.594	5.527	6.423	7.344	8.351	9.524	11.03	12.03	13.36	14.27	15.51	17.53	20.09	21.95	26.12	27.87
9	0.972	1.152	1.735	2.088	2.700	3.325	3.785	4.168	4.817	5.380	6.393	7.357	8.343	9.414	10.66	12.24	13.29	14.68	15.63	16.92	19.02	21.67	23.59	27.88	29.67
10	1.265	1.479	2.156	2.558	3.247	3.940	4.446	4.865	5.570	6.179	7.267	8.295	9.342	10.47	11.78	13.44	14.53	15.99	16.97	18.31	20.48	23.21	25.19	29.59	31.42
11	1.587	1.834	2.603	3.053	3.816	4.575	5.124	5.578	6.336	6.989	8.148	9.237	10.34	11.53	12.90	14.63	15.77	17.28	18.29	19.68	21.92	24.73	26.76	31.26	33.14
12	1.935	2.214	3.074	3.571	4.404	5.226	5.818	6.304	7.114	7.807	9.034	10.18	11.34	12.58	14.01	15.81	16.99	18.55	19.60	21.03	23.34	26.22	28.30	32.91	34.82
13	2.305	2.617	3.565	4.107	5.009	5.892	6.524	7.041	7.901	8.634	9.926	11.13	12.34	13.64	15.12	16.98	18.20	19.81	20.90	22.36	24.74	27.69	29.82	34.53	36.48
14	2.697	3.041	4.075	4.660	5.629	6.571	7.242	7.790	8.696	9.467	10.82	12.08	13.34	14.69	16.22	18.15	19.41	21.06	22.18	23.68	26.12	29.14	31.32	36.12	38.11
15	3.107	3.483	4.601	5.229	6.262	7.261	7.969	8.547	9.499	10.31	11.72	13.03	14.34	15.73	17.32	19.31	20.60	22.31	23.45	25.00	27.49	30.58	32.80	37.70	39.72
16	3.536	3.942	5.142	5.812	6.908	7.962	8.707	9.312	10.31	11.15	12.62	13.98	15.34	16.78	18.42	20.47	21.79	23.54	24.72	26.30	28.85	32.00	34.27	39.25	41.31
17	3.980	4.416	5.697	6.408	7.564	8.672	9.452	10.09	11.12	12.00	13.53	14.94	16.34	17.82	19.51	21.61	22.98	24.77	25.97	27.59	30.19	33.41	35.72	40.79	42.88
18	4.439	4.905	6.265	7.015	8.231	9.390	10.21	10.86	11.95	12.86	14.44	15.89	17.34	18.87	20.60	22.76	24.16	25.99	27.22	28.87	31.53	34.81	37.16	42.31	44.43
19	4.913	5.407	6.844	7.633	8.907	10.12	10.97	11.65	12.77	13.72	15.35	16.85	18.34	19.91	21.69	23.90	25.33	27.20	28.46	30.14	32.85	36.19	38.58	43.82	45.97
20	5.398	5.921	7.434	8.260	9.591	10.85	11.73	12.44	13.60	14.58	16.27	17.81	19.34	20.95	22.77	25.04	26.50	28.41	29.69	31.41	34.17	37.57	40.00	45.31	47.50
21	5.895	6.447	8.034	8.897	10.28	11.59	12.50	13.24	14.44	15.44	17.18	18.77	20.34	21.99	23.86	26.17	27.66	29.62	30.92	32.67	35.48	38.93	41.40	46.80	49.01
22	6.404	6.983	8.643	9.542	10.98	12.34	13.28	14.04	15.28	16.31	18.10	19.73	21.34	23.03	24.94	27.30	28.82	30.81	32.14	33.92	36.78	40.29	42.80	48.27	50.51
23	6.924	7.529	9.260	10.20	11.69	13.09	14.06	14.85	16.12	17.19	19.02	20.69	22.34	24.07	26.02	28.43	29.98	32.01	33.36	35.17	38.08	41.64	44.18	49.73	52.00
24	7.453	8.085	9.886	10.86	12.40	13.85	14.85	15.66	16.97	18.06	19.94	21.65	23.34	25.11	27.10	29.55	31.13	33.20	34.57	36.42	39.36	42.98	45.56	51.18	53.48
25	7.991	8.649	10.52	11.52	13.12	14.61	15.64	16.47	17.82	18.94	20.87	22.62	24.34	26.14	28.17	30.68	32.28	34.38	35.78	37.65	40.65	44.31	46.93	52.62	54.95
26	8.537	9.222	11.16	12.20	13.84	15.38	16.44	17.29	18.67	19.82	21.79	23.58	25.34	27.18	29.25	31.79	33.43	35.56	36.98	38.89	41.92	45.64	48.29	54.05	56.41
27	9.093	9.803	11.81	12.88	14.57	16.15	17.24	18.11	19.53	20.70	22.72	24.54	26.34	28.21	30.32	32.91	34.57	36.74	38.18	40.11	43.19	46.96	49.65	55.48	57.86
28	9.656	10.39	12.46	13.56	15.31	16.93	18.05	18.94	20.39	21.59	23.65	25.51	27.34	29.25	31.39	34.03	35.71	37.92	39.38	41.34	44.46	48.28	50.99	56.89	59.30
29	10.23	10.99	13.12	14.26	16.05	17.71	18.85	19.77	21.25	22.48	24.58	26.48	28.34	30.28	32.46	35.14	36.85	39.09	40.57	42.56	45.72	49.59	52.34	58.30	60.73
30	10.80	11.59	13.79	14.95	16.79	18.49	19.66	20.60	22.11	23.36	25.51	27.44	29.34	31.32	33.53	36.25	37.99	40.26	41.76	43.77	46.98	50.89	53.67	59.70	62.16
31	11.39	12.20	14.46	15.66	17.54	19.28	20.48	21.43	22.98	24.26	26.44	28.41	30.34	32.35	34.60	37.36	39.12	41.42	42.95	44.99	48.23	52.19	55.00	61.10	63.58
32	11.98	12.81	15.13	16.36	18.29	20.07	21.30	22.27	23.84	25.15	27.37	29.38	31.34	33.38	35.66	38.47	40.26	42.58	44.13	46.19	49.48	53.49	56.33	62.49	64.99
33	12.58	13.43	15.82	17.07	19.05	20.87	22.12	23.11	24.71	26.04	28.31	30.34	32.34	34.41	36.73	39.57	41.39	43.75	45.31	47.40	50.73	54.78	57.65	63.87	66.40
34	13.18	14.06	16.50	17.79	19.81	21.66	22.94	23.95	25.59	26.94	29.24	31.31	33.34	35.44	37.80	40.68	42.51	44.90	46.49	48.60	51.97	56.06	58.96	65.25	67.80
35	13.79	14.69	17.19	18.51	20.57	22.47	23.76	24.80	26.46	27.84	30.18	32.28	34.34	36.47	38.86	41.78	43.64	46.06	47.66	49.80	53.20	57.34	60.27	66.62	69.20
36	14.40	15.32	17.89	19.23	21.34	23.27	24.59	25.64	27.34	28.73	31.12	33.25	35.34	37.50	39.92	42.88	44.76	47.21	48.84	51.00	54.44	58.62	61.58	67.98	70.59
37	15.02	15.97	18.59	19.96	22.11	24.07	25.42	26.49	28.21	29.64	32.05	34.22	36.34	38.53	40.98	43.98	45.89	48.36	50.01	52.19	55.67	59.89	62.88	69.35	71.97
38	15.64	16.61	19.29	20.69	22.88	24.88	26.25	27.34	29.09	30.54	32.99	35.19	37.34	39.56	42.05	45.08	47.01	49.51	51.17	53.38	56.90	61.16	64.18	70.70	73.35
39	16.27	17.26	20.00	21.43	23.65	25.70	27.09	28.20	29.97	31.44	33.93	36.16	38.34	40.59	43.11	46.17	48.13	50.66	52.34	54.57	58.12	62.43	65.48	72.06	74.72
40	16.91	17.92	20.71	22.16	24.43	26.51	27.93	29.05	30.86	32.34	34.87	37.13	39.34	41.62	44.16	47.27	49.24	51.81	53.50	55.76	59.34	63.69	66.77	73.40	76.10
50	23.46	24.67	27.99	29.71	32.36	34.76	36.40	37.69	39.75	41.45	44.31	46.86	49.33	51.89	54.72	58.16	60.35	63.17	65.03	67.50	71.42	76.15	79.49	86.66	89.56
60	30.34	31.74	35.53	37.48	40.48	43.19	45.02	46.46	48.76	50.64	53.81	56.62	59.33	62.13	65.23	68.97	71.34	74.40	76.41	79.08	83.30	88.38	91.95	99.61	102.7
70	37.47	39.04	43.28	45.44	48.76	51.74	53.75	55.33	57.84	59.90	63.35	66.40	69.33	72.36	75.69	79.71	82.26	85.53	87.68	90.53	95.02	100.4	104.2	112.3	115.6
80	44.79	46.52	51.17	53.54	57.15	60.39	62.57	64.28	66.99	69.21	72.92	76.19	79.33	82.57	86.12	90.41	93.11	96.58	98.86	101.9	106.6	112.3	116.3	124.8	128.3
90	52.28	54.16	59.20	61.75	65.65	69.13	71.46	73.29	76.20	78.56	82.51	85.99	89.33	92.76	96.52	101.1	103.9	107.6	110.0	113.1	118.1	124.1	128.3	137.2	140.8
100	59.89	61.92	67.33	70.06	74.22	77.93	80.41	82.36	85.44	87.95	92.13	95.81	99.33	102.9	106.9	111.7	114.7	118.5	121.0	124.3	129.6	135.8	140.2	149.4	153.2
120	75.47	77.76	83.85	86.92	91.57	95.70	98.46	100.6	104.0	106.8	111.4	115.5	119.3	123.3	127.6	132.8	136.1	140.2	143.0	146.6	152.2	159.0	163.6	173.6	177.6
150	99.46	102.1	109.1	112.7	118.0	122.7	125.8	128.3	132.1	135.3	140.5	145.0	149.3	153.8	158.6	164.3	168.0	172.6	175.6	179.6	185.8	193.2	198.4	209.3	213.6
200	140.7	143.8	152.2	156.4	162.7	168.3	172.0	174.8	179.4	183.0	189.0	194.3	199.3	204.4	210.0	216.6	220.7	226.0	229.5	234.0	241.1	249.4	255.3	267.5	272.4

Table T6a: Quantile of the F-Snedcor distribution function

$$X \sim F_{(n,m)} : x_p = F_X^{-1}(p) \text{ com } p = 0.90$$

$m \setminus n$	1	2	3	4	5	6	7	8	9	10	20	40	120	∞
1	40	50	54	56	57	58	59	59	60	60	62	63	63	63
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.44	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.18	5.16	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.84	3.80	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.21	3.16	3.12	3.11
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.84	2.78	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.59	2.54	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.42	2.36	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.30	2.23	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.20	2.13	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.12	2.05	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.06	1.99	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.01	1.93	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	1.96	1.89	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	1.92	1.85	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.89	1.81	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.86	1.78	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.84	1.75	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.81	1.73	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.79	1.71	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.78	1.69	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.76	1.67	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.74	1.66	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.73	1.64	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.72	1.63	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.71	1.61	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.70	1.60	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.69	1.59	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.68	1.58	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.67	1.57	1.50	1.46
31	2.87	2.48	2.27	2.14	2.04	1.97	1.92	1.88	1.84	1.81	1.66	1.56	1.49	1.45
32	2.87	2.48	2.26	2.13	2.04	1.97	1.91	1.87	1.83	1.81	1.65	1.56	1.48	1.44
33	2.86	2.47	2.26	2.12	2.03	1.96	1.91	1.86	1.83	1.80	1.64	1.55	1.47	1.43
34	2.86	2.47	2.25	2.12	2.02	1.96	1.90	1.86	1.82	1.79	1.64	1.54	1.46	1.42
35	2.85	2.46	2.25	2.11	2.02	1.95	1.90	1.85	1.82	1.79	1.63	1.53	1.46	1.41
36	2.85	2.46	2.24	2.11	2.01	1.94	1.89	1.85	1.81	1.78	1.63	1.53	1.45	1.40
37	2.85	2.45	2.24	2.10	2.01	1.94	1.89	1.84	1.81	1.78	1.62	1.52	1.44	1.40
38	2.84	2.45	2.23	2.10	2.01	1.94	1.88	1.84	1.80	1.77	1.61	1.52	1.44	1.39
39	2.84	2.44	2.23	2.09	2.00	1.93	1.88	1.83	1.80	1.77	1.61	1.51	1.43	1.38
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.61	1.51	1.42	1.38
45	2.82	2.42	2.21	2.07	1.98	1.91	1.85	1.81	1.77	1.74	1.58	1.48	1.40	1.35
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73	1.57	1.46	1.38	1.33
55	2.80	2.40	2.19	2.05	1.95	1.88	1.83	1.78	1.75	1.72	1.55	1.45	1.36	1.31
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.54	1.44	1.35	1.29
70	2.78	2.38	2.16	2.03	1.93	1.86	1.80	1.76	1.72	1.69	1.53	1.42	1.32	1.27
80	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68	1.51	1.40	1.31	1.25
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67	1.50	1.39	1.29	1.23
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66	1.49	1.38	1.28	1.22
110	2.75	2.35	2.13	2.00	1.90	1.83	1.77	1.73	1.69	1.66	1.49	1.37	1.27	1.20
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.48	1.37	1.26	1.19
∞	2.71	2.30	2.08	1.95	1.85	1.77	1.72	1.67	1.63	1.60	1.42	1.30	1.17	1.03

Table T6b: Quantile of the F-Snedcor distribution function
$X \sim F_{(n,m)} : x_p = F_X^{-1}(p)$ com $p = 0.95$

$m \setminus n$	1	2	3	4	5	6	7	8	9	10	20	40	120	∞
1	161	199	216	225	230	234	237	239	241	242	248	251	253	254
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.45	19.47	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.66	8.59	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.80	5.72	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.56	4.46	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.87	3.77	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.44	3.34	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.15	3.04	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	2.94	2.83	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.77	2.66	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.65	2.53	2.45	2.41
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.54	2.43	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.46	2.34	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.39	2.27	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.33	2.20	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.28	2.15	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.23	2.10	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.19	2.06	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.16	2.03	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.12	1.99	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.10	1.96	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.07	1.94	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.05	1.91	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.03	1.89	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.01	1.87	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	1.99	1.85	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	1.97	1.84	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	1.96	1.82	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	1.94	1.81	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	1.93	1.79	1.68	1.62
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15	1.92	1.78	1.67	1.61
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	1.91	1.77	1.66	1.60
33	4.14	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18	2.13	1.90	1.76	1.64	1.58
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	1.89	1.75	1.63	1.57
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	1.88	1.74	1.62	1.56
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	1.87	1.73	1.61	1.55
37	4.11	3.25	2.86	2.63	2.47	2.36	2.27	2.20	2.14	2.10	1.86	1.72	1.60	1.54
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	1.85	1.71	1.59	1.53
39	4.09	3.24	2.85	2.61	2.46	2.34	2.26	2.19	2.13	2.08	1.85	1.70	1.58	1.52
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.84	1.69	1.58	1.51
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	1.81	1.66	1.54	1.47
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.78	1.63	1.51	1.44
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01	1.76	1.61	1.49	1.41
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.75	1.59	1.47	1.39
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.72	1.57	1.44	1.35
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.70	1.54	1.41	1.33
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.69	1.53	1.39	1.30
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.68	1.52	1.38	1.28
110	3.93	3.08	2.69	2.45	2.30	2.18	2.09	2.02	1.97	1.92	1.67	1.50	1.36	1.27
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.66	1.50	1.35	1.26
∞	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.57	1.40	1.22	1.03

Table T6c: Quantile of the F-Snedcor distribution function
$X \sim F_{(n,m)} : x_p = F_X^{-1}(p)$ com $p = 0.99$

$m \setminus n$	1	2	3	4	5	6	7	8	9	10	20	40	120	∞
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6209	6286	6340	6366
2	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.45	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	26.69	26.41	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.02	13.75	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.55	9.29	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.40	7.14	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.16	5.91	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.36	5.12	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.81	4.57	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.41	4.17	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.10	3.86	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	3.86	3.62	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.66	3.43	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.51	3.27	3.09	3.01
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.37	3.13	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.26	3.02	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.16	2.92	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.08	2.84	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.00	2.76	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	2.94	2.69	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	2.88	2.64	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	2.83	2.58	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	2.78	2.54	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	2.74	2.49	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.70	2.45	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.66	2.42	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.63	2.38	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.60	2.35	2.17	2.07
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.57	2.33	2.14	2.04
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.55	2.30	2.11	2.01
31	7.53	5.36	4.48	3.99	3.67	3.45	3.28	3.15	3.04	2.96	2.52	2.27	2.09	1.98
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93	2.50	2.25	2.06	1.96
33	7.47	5.31	4.44	3.95	3.63	3.41	3.24	3.11	3.00	2.91	2.48	2.23	2.04	1.93
34	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89	2.46	2.21	2.02	1.91
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.44	2.19	2.00	1.89
36	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86	2.43	2.18	1.98	1.87
37	7.37	5.23	4.36	3.87	3.56	3.33	3.17	3.04	2.93	2.84	2.41	2.16	1.96	1.86
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83	2.40	2.14	1.95	1.84
39	7.33	5.19	4.33	3.84	3.53	3.30	3.14	3.01	2.90	2.81	2.38	2.13	1.93	1.82
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.37	2.11	1.92	1.81
45	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74	2.31	2.05	1.85	1.74
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.27	2.01	1.80	1.68
55	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.23	1.97	1.76	1.64
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.20	1.94	1.73	1.60
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.15	1.89	1.67	1.54
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.12	1.85	1.63	1.50
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52	2.09	1.82	1.60	1.46
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.07	1.80	1.57	1.43
110	6.87	4.80	3.96	3.49	3.19	2.97	2.81	2.68	2.57	2.49	2.05	1.78	1.55	1.40
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.03	1.76	1.53	1.38
∞	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	1.88	1.59	1.33	1.05