Capítulo 6

Point estimation

6.1. Statistical Models. Statistical Parametric Models

Let $X_1, \ldots, X_n$ be a random sample of X. That is, $X_1, \ldots, X_n$ are independent identically distributed (i.i.d.) random variables with the same distribution as X. Let F be the distribution function of X. Let $\mathcal{F}_{\text{All}}$ be the set of all possible distribution functions.

Definition 6.1.1 (Statistical Model)**.** *A subset $\mathcal{F}$ of $\mathcal{F}_{\text{All}}$ is called a* statistical model. *We say that X follows the statistical model $\mathcal{F}$ when $X \sim F \in \mathcal{F}$.*

Example 6.1.1.

- $\mathcal{F}$, the set of all distribution functions corresponding to continuous probability distributions. In this case $\mathcal{F}$ can be identified with the set of density functions.

- $\mathcal{F}$, the set of distribution functions corresponding to continuous probability distributions that are symmetric around 0. In this case $\mathcal{F}$ can be identified with the set of density functions $f(x)$ such that $f(x) = f(-x)$ for all $x \in \mathbb{R}$.

- $\mathcal{F}$, the set of normal distributions with expected value μ and variance σ^2, with $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}$, $\sigma > 0$:

$$X \sim F \in \mathcal{F} \Leftrightarrow X \sim N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma \in \mathbb{R}, \sigma > 0.$$

□

Definition 6.1.2 (Statistical Parametric Model). *A subset of $\mathcal{F}_{\text{All}}$ is called a* statistical parametric model *if it can be expressed as*

$$\mathcal{F}_\Theta = \{F(x;\theta) : \theta \in \Theta \subseteq \mathbb{R}^k\},$$

where $F(x;\theta)$ is a known function of x and θ.

When $F(x;\theta)$ is the distribution function of a continuous random variable, it is equivalent to know the expression of the density function $f(x;\theta) = F'(x;\theta)$.

When $F(x;\theta)$ is the distribution function of a discrete random variable, it is equivalent to know the expression of the probability function $p(x;\theta) = \Pr_\theta(X = x)$.

Example 6.1.2.

- The statistical model $N(\mu, \sigma^2)$ is a parametric model with $k = 2$ parameters:
$$\theta = (\mu, \sigma) \in \Theta = \mathbb{R} \times \mathbb{R}^+.$$
- $\mathcal{F}$, the set of distribution functions corresponding to continuous probability distributions that are symmetric around 0, is not a parametric model.

□

Definition 6.1.3 (Parameters). *Assume that $X \sim F \in \mathcal{F}$. Let*

$$\begin{array}{rccl} \Phi : & \mathcal{F} & \to & \mathbb{R}^k \\ & F & \mapsto & \theta = \Phi(F) \end{array}$$

be a function that associates k real numbers $\theta = (\theta_1, \dots, \theta_k)$ to any random variable following the statistical model $\mathcal{F}$. We say that $\theta = (\theta_1, \dots, \theta_k)$ are k parameters of the distribution of X.

Example 6.1.3.

Consider the statistical model $\mathcal{F}$, the set continuous probability distributions symmetric around 0 with finite variance. Let X be a random variable with distribution in $\mathcal{F}$ and let $f(x)$ its density function. Then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{0} xf(x)dx + \int_{0}^{\infty} xf(x)dx = 0$$

because both integrals are finite (finite variace implies well defined expectation), equal in absolute value (by symmetry) and of opposite sign. The variance is a parameter of X:

$$\sigma^2 = \mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = 2\int_{0}^{\infty} x^2 f(x)dx < \infty.$$

$\square$

When X follows a parametric model, $X \sim \mathcal{F}_\Theta$, the function

$$\begin{array}{rccl} \Phi : & \mathcal{F}_\Theta & \to & \mathbb{R}^k \\ & F(\cdot;\theta) & \mapsto & \theta = \Phi(F(\cdot;\theta)) \end{array}$$

defines the k parameters $\theta_1, \ldots, \theta_k$ that identify the probability distribution of X. In this case, knowing these k parameters is equivalent to knowing the whole distribution of X.

Definition 6.1.4 (Statistics)**.** *Let $X_1, \ldots, X_n$ be a random sample of X. Assume that X takes values in $\mathcal{X} \subseteq \mathbb{R}$. Any function*

$$\begin{array}{rccl} T : & \mathcal{X}^n & \to & \mathbb{R} \\ & (x_1, \ldots, x_n) & \mapsto & T(x_1, \ldots, x_n) \end{array}$$

applied to $(X_1, \ldots, X_n)$ is a statistic*:*

$$T_n = T(X_1, \ldots, X_n).$$

It is possible to define k-dimensional statistics just considering the k-dimensional vector which components are k (one-dimensional) statistics.

Definition 6.1.5 (Estimators)**.** *Let $X_1, \ldots, X_n$ be a random sample of $X \sim F \in \mathcal{F}$. Let $\theta = \Phi(F) \in \Theta$ be a parameter of X. A statistic $T_n = T(X_1, \ldots, X_n)$ is an* estimator *of θ if T_n is used as an approximation of the value of θ. In this case we usually write*

$$\hat{\theta}_n = T_n = T(X_1, \ldots, X_n).$$

Definition 6.1.6 (Estimators based on the plug-in principle)**.** *Let $X_1, \dots, X_n$ be a random sample of $X \sim F \in \mathcal{F}$. Let $\theta = \Phi(F) \in \Theta$ be a parameter of X. Let*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$$

be the empirical distribution function. When $\Phi(F_n)$ can be computed, we say that

$$\hat{\theta}_n = \Phi(F_n)$$

is the plug-in estimator *of θ.*

Example 6.1.4.

Let $X_1, \dots, X_n$ be a random sample of $X \sim F$, with finite j-th *population moment* $\mu_j = \mathbb{E}(X^j)$. The plug-in estimator of μ_j is

$$m_j = \mathbb{E}(X_e^j)$$

where X_e is a random variable with distribution function F_n, that is, X_e is a discrete random variable that, for $i = 1, \dots, n$, takes the same value as X_i with probability $1/n$. So, when the observed sample is $x_1, \dots, x_n$,

$$m_j = \frac{1}{n} \sum_{i=1}^{n} x_i^j.$$

As a function of the random variables $X_1, \dots, X_n$ it is

$$m_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j.$$

m_j is called the j-th *sample moment.* □

6.2. Method of Moments

Definition 6.2.1 (Method of Moments). *Let X be a random variable, $X \sim F \in \mathcal{F}$, a statistical model for which the first k moments are assumed to be finite. Let $\theta = \Phi(F) \in \Theta \subseteq \mathbb{R}^k$ be a k-dimensional parameter of X, $\theta = (\theta_1, \dots, \theta_k)$. Let $\mu_1, \dots, \mu_k$ be first k* population moments *of X. Assume that there is a one-to-one relationship between $\theta_1, \dots, \theta_k$ and $\mu_1, \dots, \mu_k$:*

$$\left\{ \begin{array}{c} \mu_1 = g_1(\theta_1, \dots, \theta_k), \\ \vdots \\ \mu_k = g_k(\theta_1, \dots, \theta_k), \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{c} \theta_1 = h_1(\mu_1, \dots, \mu_k), \\ \vdots \\ \theta_k = h_k(\mu_1, \dots, \mu_k). \end{array} \right\}$$

The estimator of $(\theta_1, \dots, \theta_k)$ based on the method of moments *(or* the method of moments estimator of $(\theta_1, \dots, \theta_k)$*) is obtained when replacing the population moments by the sample moments in the last identities:*

$$\begin{aligned} \hat{\theta}_1 &= h_1(m_1, \dots, m_k), \\ &\vdots \\ \hat{\theta}_k &= h_k(m_1, \dots, m_k). \end{aligned}$$

Example 6.2.1.

Let X be a random sample with finite variance $\sigma^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \mu_2 - \mu_1^2$. Therefore the method of moments estimator od σ^2 is

$$\hat{\sigma}^2 = m_2 - m_1^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

□

Observe that working with (μ_1, μ_2) is equivalent to working with (μ_1, σ^2) because

$$\sigma^2 = \mu_2 - \mu_1^2 \Leftrightarrow \mu_2 = \sigma^2 + \mu_1^2.$$

Therefore, in the definition of method of moments estimators it is equivalent working with $(\mu_1, \mu_2, \mu_3, \dots, \mu_k)$ that working with $(\mu_1, \sigma^2, \mu_3, \dots, \mu_k)$. Once the expressions of $\theta_1, \dots, \theta_k$ as functions of $(\mu_1, \sigma^2, \mu_3, \dots, \mu_k)$ have been obtained, these values are replaced by $(m_1, \hat{\sigma}^2, m_3, \dots, m_k)$.

Example 6.2.2.

Consider $X \sim \gamma(\alpha, \beta)$, where $\alpha > 0$ is the `shape` parameter, and $\beta > 0$ is the `rate` parameter: X has density function

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},\ x > 0.$$

Then

$$\mu_1 = \mathbb{E}(X) = \frac{\alpha}{\beta},\ \sigma^2 = \mathrm{Var}(X) = \frac{\alpha}{\beta^2}.$$

We express α and β as a function of μ_1 and σ^2:

$$\beta = \frac{\mu_1}{\sigma^2},\ \alpha = \beta\mu_1 = \frac{\mu_1^2}{\sigma^2}.$$

Therefore the method of moments estimators of parameters α and β are

$$\hat{\beta} = \frac{m_1}{\hat{\sigma}^2},\ \alpha = \frac{m_1^2}{\hat{\sigma}^2}.$$

□

6.3. Maximum Likelihood Estimators

Definition 6.3.1 (Likelihood function)**.** *Let $X_1, \ldots, X_n$ be a random sample of X following a parametric model: $X \sim F(x; \theta)$, with $\theta \in \Theta \subseteq \mathbb{R}^k$. Assume that X takes values in $\mathcal{X} \subseteq \mathbb{R}$. The* likelihood function *is a function of the parameter values and the observed values of the sample,*

$$\begin{array}{rccl} L: & \Theta \times \mathcal{X}^n & \to & \mathbb{R} \\ & (\theta; x_1, \ldots, x_n) & \mapsto & L(\theta; x_1, \ldots, x_n) \end{array}$$

that for a discrete random variable X is computed as

$$L(\theta; x_1, \ldots, x_n) = \Pr(X_1 = x_1, \ldots, X_n = x_n; \theta) = \prod_{i=1}^n \Pr(X_i = x_i; \theta),$$

where $\Pr(X = x; \theta)$ is the probability function of X when it is distributed according the value θ of the parameter.

When X is continuous, the likelihood function is computed as

$$L(\theta; x_1, \ldots, x_n) = f_{(X_1, \ldots, X_n)}(x_1, \ldots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta),$$

where $f(\cdot; \theta)$ is the density function of X when it is distributed according to the value θ of the parameter.

The likelihood function tries to measure how likely is that the values $x_1, \ldots, x_n$ have been observed in the case that θ is the true value of the parameter. This idea leads to the following definition.

Definition 6.3.2 (Maximum Likelihood Estimate). *Let $X_1, \ldots, X_n$ be a random sample of X following the parametric model $F(x;\theta)$, with $\theta \in \Theta \subseteq \mathbb{R}^k$. Let $x_1, \ldots, x_n$ the observed values of $X_1, \ldots, X_n$, respectively. The maximum likelihood estimate of θ for this observed sample is*

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta; x_1, \ldots, x_n).$$

It is useful to define the *log-likelihood function*,

$$\ell(\theta; x_1, \ldots, x_n) = \log L(\theta; x_1, \ldots, x_n),$$

because the logarithm functions transform products into sums. For the discrete case,

$$\ell(\theta; x_1, \ldots, x_n) = \sum_{i=1}^{n} \log \Pr_{\theta}(X_i = x_i).$$

For the continuous case,

$$\ell(\theta; x_1, \ldots, x_n) = \sum_{i=1}^{n} \log f(x_i; \theta).$$

Observe that

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta; x_1, \ldots, x_n) = \arg \max_{\theta \in \Theta} \ell(\theta; x_1, \ldots, x_n)$$

because the logarithm is a monotone function. It is usually much easier to maximize the log-likelihood function than to maximize the likelihood function.

Both, the likelihood function and its logarithm, can be evaluated in the random sample $X_1, \ldots, X_n$ instead doing it in the observed values of the sample, $x_1, \ldots, X_n$. This way we obtain two random functions depending on the parameter values θ

$$L(\theta; X_1, \ldots, X_n) \text{ and } \ell(\theta; X_1, \ldots, X_n).$$

Maximizing any of these functions we obtain the *Maximum Likelihood Estimator*, as a random variable:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta; X_1, \ldots, X_n) = \arg \max_{\theta \in \Theta} \ell(\theta; X_1, \ldots, X_n).$$

Assuming, for instance, that X is continuous and that $X \sim f(x;\theta_0)$,

$$\frac{1}{n}\ell(\theta;X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^{n}\log f(X_i;\theta) \approx \mathbb{E}\left(\log f(X;\theta)\right) = \int_{-\infty}^{\infty}\log f(x;\theta)f(x;\theta_0)dx.$$

It can be proved that

$$\int_{-\infty}^{\infty}\log f(x;\theta)f(x;\theta_0)dx \leq \int_{-\infty}^{\infty}\log f(x;\theta_0)f(x;\theta_0)dx \text{ for all } \theta\in\Theta.$$

So

$$\theta_0 = \arg\max_{\theta\in\Theta}\mathbb{E}\left(\log f(X;\theta)\right).$$

Proposició 6.3.1. *Let f and g two density functions such that, for all $x\in\mathbb{R}$, $f(x)>0$ if and only if $g(x)>0$. Then*

$$\int_S \log(g(x))f(x)dx \leq \int_S \log(f(x))f(x)dx,$$

where S is the subset of $\mathbb{R}$ where f and g are strictly positive.
Proof:

$$\int_S \log(g(x))f(x)dx - \int_S \log(f(x))f(x)dx = \int_S \log\left(\frac{g(x)}{f(x)}\right)f(x)dx$$

$$\leq \int_S\left(\frac{g(x)}{f(x)}-1\right)f(x)dx = \int_S g(x)dx - \int_S f(x)dx = 1-1=0.$$

We have used that for all $u>0$ it happens that $\log(u)\leq u-1$. □

This is a theoretical justification for the definition of the Maximum Likelihood Estimator:

$$\hat{\theta} = \arg\max_{\theta\in\Theta}\sum_{i=1}^{n}\log f(X_i;\theta) = \arg\max_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\log f(X_i;\theta) \approx \arg\max_{\theta\in\Theta}\mathbb{E}\left(\log f(X;\theta)\right) = \theta_0.$$

Theorem 6.3.3 (Invariance principle)**.** *Let $\psi=\Psi(\theta)$, with Ψ a 1-1 function, be a* reparameterization *of the parametric model. If $\hat{\theta}$ is the maximum likelihood estimator of θ, then $\hat{\psi}=\Psi(\hat{\theta})$ is the maximum likelihood estimator of ψ.*

Example 6.3.1.
Consider $X \sim \text{Bin}(n, p)$. Let $\psi = \Psi(p) = \log(p/(1-p))$. Parameter ψ is known as *log-odds*, and it is an alternative way of expressing probabilities. The function Ψ, that is 1-1 from $(0,1)$ to $\mathbb{R}$, is known as *logistic* function. The inverse of function Ψ is $p = \Psi^{-1}(\psi) = e^{\psi}/(1+e^{\psi})$.

The maximum likelihood estimator of p is $\hat{p}_n = X/n$, the sampling proportion of successes. So the maximum likelihood estimator of ψ is

$$\hat{\psi}_n = \log\left(\frac{\hat{p}_n}{1-\hat{p}_n}\right).$$

□

6.4. Properties of an estimator

6.4.1. Sampling distribution. Standard error

Let $X_1, \ldots, X_n$ be a random sample of X, with distribution F in the statistical model $\mathcal{F}$. Let $\Phi : \mathcal{F} \to \mathbb{R}$ be the function defining the value of a one-dimensional parameter of interest for the population X from its probability distribution: $\theta = \Phi(F)$. We can use the notation $\theta_F = \Phi(F)$ to make explicit that the parameter value depends on the specific probability distribution of X.

Definition 6.4.1 (Sampling distribution and standard error)**.** *Let $T_n = T(X_1, \ldots, X_n)$ be a statistic. The probability distribution of T_n is known as its* sampling distribution*. The standard deviation of T_n is known as its* standard error*.*

6.4.2. Bias and Mean Square Error

Definition 6.4.2 (Bias of an estimator)**.** *Let $X_1, \ldots, X_n$ be a random sample of X, with distribution F. Let $\hat{\theta} = T(X_1, \ldots, X_n)$ an estimator of $\theta = \Phi(F)$. The* bias *of $\hat{\theta}$ as estimator of θ is defined as*

$$Bias_{\theta}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta.$$

Definition 6.4.3 (Unbiased estimator)**.** *Let $X_1, \ldots, X_n$ be a random sample of X, with distribution $F \in \mathcal{F}$. Let $\hat{\theta} = T(X_1, \ldots, X_n)$ an estimator of $\theta_F = \Phi(F)$. We say that $\hat{\theta}$ is an* unbiased estimator *of θ_F when*

$$\mathbb{E}_F(\hat{\theta}) = \theta_F, \ \textit{for all } F \in \mathcal{F},$$

where $\mathbb{E}_F(\hat{\theta})$ *indicates that the expectation is computed assuming that the true distribution function of* X *is* F*.*

In the particular case of a statistical parametric model, $X \sim F(x;\theta)$ with $\theta \in \Theta \subseteq \mathbb{R}^k$, $\widehat{\tau(\theta)}$ is an unbiased estimator of $\tau(\theta)$ when

$$\mathbb{E}_\theta(\widehat{\tau(\theta)}) = \tau(\theta) \text{ for all } \theta \in \Theta,$$

where $\mathbb{E}_\theta(\widehat{\tau(\theta)})$ indicates that the expectation is computed assuming that it was θ the value of the parameter that generated the data $X_1, \ldots, X_n$.

Definition 6.4.4 (Mean Square Error of an estimator)**.** *Let* $X_1, \ldots, X_n$ *be a random sample of* X*, with distribution* F*. Let* $\hat{\theta} = T(X_1, \ldots, X_n)$ *an estimator of* $\theta = \Phi(F)$*. The* mean square error *of* $\hat{\theta}$ *as estimator of* θ *is defined as*

$$MSE_\theta(\hat{\theta}) = \mathbb{E}\left((\hat{\theta} - \theta)^2\right).$$

Observe that

$$\begin{aligned}
\mathrm{MSE}_\theta(\hat{\theta}) =& \mathbb{E}\left((\hat{\theta} - \theta)^2\right) \\
=& \mathbb{E}\left((\{\mathbb{E}(\hat{\theta}) - \theta\} + \{\hat{\theta} - \mathbb{E}(\hat{\theta})\})^2\right) \\
=& (\mathbb{E}(\hat{\theta}) - \theta)^2 + \mathbb{E}\left((\hat{\theta} - \mathbb{E}(\hat{\theta}))^2\right) + 2(\mathbb{E}(\hat{\theta}) - \theta)\mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right) \\
=& (\mathbb{E}(\hat{\theta}) - \theta)^2 + \mathbb{E}\left((\hat{\theta} - \mathbb{E}(\hat{\theta}))^2\right) \\
=& \left(\mathrm{Bias}_\theta(\hat{\theta})\right)^2 + \mathrm{Var}(\hat{\theta}).
\end{aligned}$$

6.4.3. Consistency and asymptotic normality

Let $X_1, \ldots, X_n, \ldots$ be a sequence of independent identically distributed random variables with the same distribution as $X \sim F \in \mathcal{F}$. Let

$$\hat{\theta}_n = T_n(X_1, \ldots, X_n),\ n \geq 1,$$

be a sequence of estimators of the parameter $\theta_F = \Phi(F)$.

Definition 6.4.5 (Consistency (in probability))**.** *The sequence of estimators* $\hat{\theta}_n$ *is consistent (in probability) for* θ_F *if*

$$\hat{\theta}_n \underset{n\to\infty}{\longrightarrow} \theta_F \text{ in probability}$$

for every $F \in \mathcal{F}$. That is, for all $\varepsilon > 0$

$$\lim_n \Pr\left(|\hat{\theta}_n - \theta_F| > \varepsilon\right) = 0,$$

where the probabilities are computed assuming that F is the true distribution of X, and that happens for every $F \in \mathcal{F}$.

Similar definitions can be done for consistency almost surely and in quadratic mean, just replacing convergence in probability by convergence almost surely and in quadratic mean, respectively. Any of these two kinds of consistency imply consistency in probability. When talking about consistency, without more specific indication, it is usually understood that it refers to consistency in probability.

Proposition 6.4.6. *Let $\hat{\theta}_n$, $n \geq 1$, be a sequence of estimators of θ_F. If*

$$\lim_n \text{Bias}_{\theta_F}(\hat{\theta}) = 0 \text{ and } \lim_n \text{Var}(\hat{\theta}) = 0$$

for every $F \in \mathcal{F}$, then $\hat{\theta}_n$ is consistent in quadratic mean (and in probability) for θ_F.

Definition 6.4.7 (Asymptotic normality)**.** *Let $\hat{\theta}_n$, $n \geq 1$, be a consistent sequence of estimators of θ_F. Let $s.e.(\hat{\theta}_n)$ be the standard error of $\hat{\theta}_n$. We say that the sequence is* asymptotically normal *if*

$$\frac{\hat{\theta}_n - \theta_F}{s.e.(\hat{\theta}_n)} \underset{n\to\infty}{\longrightarrow} Z \sim N(0,1) \text{ in distribution}$$

for every $F \in \mathcal{F}$.

In most cases s.e.$(\hat{\theta}_n) \approx \sigma_{\hat{\theta}}(\theta_F)/\sqrt{n}$ when n goes to infinity, with $\sigma_{\hat{\theta}}(\theta_F)$ not depending on n, in the sense that

$$\lim_n \frac{\text{s.e.}(\hat{\theta}_n)}{\sigma_{\hat{\theta}}(\theta_F)/\sqrt{n}} = 1.$$

In this cases we can also write that

$$\sqrt{n}\left(\hat{\theta}_n - \theta_F\right) \underset{n\to\infty}{\longrightarrow} N(0, \sigma^2_{\hat{\theta}}(\theta_F)).$$

Definition 6.4.8 (Asymptotic relative efficiency)**.** *Let T_n and S_n two sequences of asymptotically normal estimators of a parameter θ, such that*

$$\sqrt{n}\,(T_n - \theta) \underset{n\to\infty}{\longrightarrow} N(0, \sigma^2_T(\theta)) \text{ in distribution, and}$$

$$\sqrt{n}\,(S_n - \theta) \underset{n\to\infty}{\longrightarrow} N(0, \sigma_S^2(\theta)) \textit{ in distribution.}$$

The asymptotic relative efficiency *of S_n with respect to T_n as*

$$\mathrm{ARE}(\theta, S_n, T_n) = \frac{1/\sigma_S^2(\theta)}{1/\sigma_T^2(\theta)} = \frac{\sigma_T^2(\theta)}{\sigma_S^2(\theta)}.$$

The value of asymptotic relative efficiency can be interpreted as the quotient of the sample sizes needed to obtain the same asymptotic precision (or the same asymptotic variance) by the two estimators in the estimation of θ. Let's see that this is so. Choosing sample size m for T and n for S, the asymptotic variances are, respectively, $\sigma_T^2(\theta)/m$ and $\sigma_S^2(\theta)/n$. If we force both of them to be equal, we have

$$\frac{\sigma_T^2(\theta)}{m} = \frac{\sigma_S^2(\theta)}{n} \Longleftrightarrow \frac{m}{n} = \frac{\sigma_T^2(\theta)}{\sigma_S^2(\theta)} = \mathrm{ARE}(\theta, S_n, T_n).$$

That is, if $\mathrm{ARE}(\theta, S_n, T_n) = 0{,}5$ then S is asymptotically less efficient that T: To have the same precision with the S estimator, we need a sample twice as large as if we were using T ($\mathrm{ARE} = 0{,}5 = m/n \Longrightarrow n = 2m$).

Asymptotic properties of the method of moments estimator

Proposition 6.4.9. *Let $\hat{\theta}_n = (\hat{\theta}_{1,n}, \ldots, \hat{\theta}_{k,n})$ be the method of moments estimator of the parameter $\theta = (\theta_1, \ldots, \theta_k)$, defined as*

$$\begin{aligned} \hat{\theta}_{1,n} &= h_1(m_{1,n}, \ldots, m_{k,n}), \\ &\vdots \\ \hat{\theta}_{k,n} &= h_k(m_{1,n}, \ldots, m_{k,n}). \end{aligned}$$

1. *If the functions $h_1, \ldots, h_k$ are continuous then $\hat{\theta}_n$ is consistent (almost surely) for θ.*
2. *Under mild conditions on the derivatives of the functions $h_1, \ldots, h_k$, $\hat{\theta}_n$ is asymptotically normal.*

Asymptotic properties of the maximum likelihood estimator

Proposition 6.4.10. *Let $\hat{\theta}_n = (\hat{\theta}_{1,n}, \ldots, \hat{\theta}_{k,n})$ be the maximum likelihood estimator of the parameter $\theta = (\theta_1, \ldots, \theta_k)$. Under* regularity conditions*,*

1. *$\hat{\theta}_n$ is consistent (almost surely) for θ.*
2. *$\hat{\theta}_n$ is asymptotically normal.*

3. *For the one-dimensional parameter case ($k = 1$), $\hat{\theta}$ is the asymptotically optimal estimator of θ, in the sense that for any other asymptotically normal estimator $\tilde{\theta}_n$ of θ, it happens that*

$$\mathrm{ARE}(\theta, \tilde{\theta}_n, \hat{\theta}_n) \leq 1$$

for every $\theta \in \Theta$.

Let us give some extra details about asymptotic normality of the maximum likelihood estimator for the one-dimensional parameter case ($k = 1$). Let

$$\ell(\theta; X) = \log f(X; \theta)$$

be the log-likelihood function for just one observation. Then the log-likelihood function for a sample os size n is

$$\ell(\theta; X_1, \ldots, X_n) = \sum_{i=1}^{n} \log f(X_i; \theta) = \sum_{i=1}^{n} \ell(\theta; X_i).$$

The *score function* is defined as

$$S(\theta; X_1, \ldots, X_n) = \frac{\partial \ell(\theta; X_1, \ldots, X_n)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta} = \sum_{i=1}^{n} S(\theta; X_i).$$

Under *regularity conditions* (support of X not depending on θ, second differentiability of $\ell(\theta; x)$ with respect to θ, interchangeability of integrals on x and derivatives with respect to θ) it can be proved that

$$\mathbb{E}_\theta(S(\theta; X_1, \ldots, X_n)) = 0, \text{ and } \mathrm{Var}(S(\theta; X_1, \ldots, X_n)) = n\mathbb{E}_\theta \left(-\frac{\partial^2 \ell(\theta; X)}{\partial \theta^2} \right).$$

The *Fisher's Information of the model* is defined as

$$I(\theta) = \mathrm{Var}(S(\theta; X)) = \mathbb{E}_\theta \left(-\frac{\partial^2 \ell(\theta; X)}{\partial \theta^2} \right),$$

and the *Fisher's Information for a sample* is

$$I_n(\theta) = \mathrm{Var}(S(\theta; X_1, \ldots, X_n)) = nI(\theta).$$

It can be proved that, under regularity conditions,

$$\sqrt{n}(\hat{\theta}_n - \theta) \underset{n \to \infty}{\longrightarrow} N(0, I(\theta)^{-1}) \text{ in distribution,}$$

or, equivalently, that

$$(nI(\theta))^{1/2}\,(\hat{\theta}_n - \theta) \underset{n\to\infty}{\longrightarrow} N(0,1) \text{ in distribution.}$$

The result remains true when replacing θ by its maximum likelihood estimator in $I(\theta)$,

$$\left(nI(\hat{\theta}_n)\right)^{1/2}(\hat{\theta}_n - \theta) \underset{n\to\infty}{\longrightarrow} N(0,1) \text{ in distribution,}$$

and even if the Fisher's Information for the sample, $nI(\theta)$, is replaced by the *observed Fisher's Information*, defined as

$$\hat{I}(\theta; X_1, \ldots, X_n) = -\left.\frac{\partial^2 \ell(\theta; X_1, \ldots, X_n)}{\partial \theta^2}\right|_{\theta=\hat{\theta}_n}.$$

When we are interested in estimating a one-to-one transformation $\tau(\theta)$ of the parameter, the Invariance Principle tells us that the maximum likelihood estimator of $\tau(\theta)$ is just $\tau(\hat{\theta}_n)$, where $\hat{\theta}_n$ is the maximum likelihood estimator of θ. It can be proved that, under the previous regularity conditions and assuming derivability of $\tau(\theta)$,

$$\sqrt{n}(\tau(\hat{\theta}_n) - \tau(\theta)) \underset{n\to\infty}{\longrightarrow} N(0, (\tau'(\theta))^2/I(\theta)) \text{ in distribution.}$$

6.5. Estimation of the mean and the variance

Let $X_1, \ldots, X_n$ be a random sample of the random variable X with $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2 < \infty$. Consider the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the sample variance

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Proposition 6.5.1. *$\bar{X}_n$ and S_n^2 are unbiased and consistent estimators of μ and σ^2, respectively. Additionally,*

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n},$$

that can be consistently estimated by

$$\widehat{\mathrm{Var}(\bar{X}_n)} = \frac{S_n^2}{n}.$$

Moreover, by the Central Limit Theorem,

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \underset{n\to\infty}{\longrightarrow} Z \sim N(0,1) \text{ in distribution.}$$

The result is still true when estimating σ *by* $S_n = \sqrt{S_n^2}$*:*

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} \underset{n\to\infty}{\longrightarrow} Z \sim N(0,1) \text{ in distribution,}$$

Teorema 6.5.1 (Fisher's Theorem)**.** *If* $X \sim N(\mu, \sigma^2)$*,*

(a) $\bar{X}_n$ *and* S_n^2 *are independent random variables.*

(b) $\bar{X}_n \sim N(\mu, \sigma^2/n)$ *or, equivalently,* $\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \sim N(0,1)$.

(c) $(n-1)S_n^2/\sigma^2 \sim \chi^2_{n-1}$.

Corol·lari 6.5.2. *Let* $X_1, \ldots, X_n$ *be i.i.d.r.v.* $N(\mu, \sigma^2)$ *and let* $\bar{X}$ *and* S^2 *be the sample mean and variance, respectively. Then*

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1).$$

6.6. Estimation of the population proportion

Let $Y_1, \ldots, Y_n$ be a random sample of the random variable $Y \sim \mathrm{Bernoulli}(p)$, $p \in [0,1]$. Let $X = \sum_{i=1}^n Y_i$, so $X \sim \mathrm{B}(n,p)$. Let

$$\hat{p}_n = \bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i = \frac{X}{n}$$

be the sampling proportion of successes in the n Bernouilli trials.

Proposition 6.6.1. $\hat{p}_n$ *is an unbiased and consistent estimator of* p*,*

$$Var(\hat{p}_n) = \frac{p(1-p)}{n},$$

that can be consistently estimated by

$$\widehat{\mathrm{Var}(\hat{p}_n)} = \frac{\hat{p}_n(1-\hat{p}_n)}{n}.$$

Moreover, by the Central Limit Theorem,

$$\frac{\sqrt{n}\,(\hat{p}_n - p)}{\sqrt{p(1-p)}} \underset{n\to\infty}{\longrightarrow} Z \sim N(0,1) \text{ in distribution.}$$

The result is still true when estimating the standard error of $\hat{p}_n$:

$$\frac{\sqrt{n}\,(\hat{p}_n - p)}{\sqrt{\hat{p}_n(1-\hat{p}_n)}} \underset{n\to\infty}{\longrightarrow} Z \sim N(0,1) \text{ in distribution.}$$

Apèndix A: Algunes distribucions usuals

A1. La família de distribucions Gamma

Definició 6.6.1 (Funció Gamma). *Si $\alpha > 0$ definim la* FUNCIÓ GAMMA *com*

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx.$$

Propietats de la funció Gamma

- $\Gamma(1) = 1$.
- $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$, es veu integrant per parts:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx = \left\{ \begin{array}{ll} u = x^{\alpha-1} & du = (\alpha-1)x^{\alpha-2} \\ dv = e^{-x}dx & v = -e^{-x} \end{array} \right\}$$

$$= \left(-e^{-x}x^{\alpha-1}\right)\Big|_0^\infty + \int_0^\infty (\alpha-1)x^{\alpha-2}e^{-x}\, dx = (0-0) + (\alpha-1)\Gamma(\alpha-1).$$

- $\Gamma(n) = (n-1)!$.
- $\Gamma(1/2) = \sqrt{\pi}$:

$$\Gamma(1/2) = \int_0^\infty x^{(1/2)-1} e^{-x}\, dx \overset{\left\{ \begin{array}{c} x = \frac{1}{2}y^2 \\ dx = y\, dy \end{array} \right\}}{=} \int_0^\infty \sqrt{2}\frac{1}{y} e^{-\frac{1}{2}y^2} y\, dy =$$

$$\sqrt{2}\sqrt{2\pi} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\, dy = \sqrt{2}\sqrt{2\pi}\frac{1}{2} = \sqrt{\pi}.$$

  Hem fet servir que $\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\, dy = 1$, perquè és la integral de la funció de densitat de la $N(0,1)$, funció simètrica respecte a 0.

Definició 6.6.2. *Una variable aleatòria X segueix una* LLEI GAMMA *de paràmetres $\alpha > 0$ i $\beta > 0$ si X té una funció de densitat definida per*

$$f(x|\alpha,\beta) = \left\{ \begin{array}{ll} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & x > 0, \\ 0 & \textit{altrament}, \end{array} \right.,$$

on α és un paràmetre de forma (`shape`*, en anglès) i β és una paràmetre d'escala inversa (*`rate`*, en anglès). Escriurem que $X \sim \gamma(\alpha,\beta)$.*

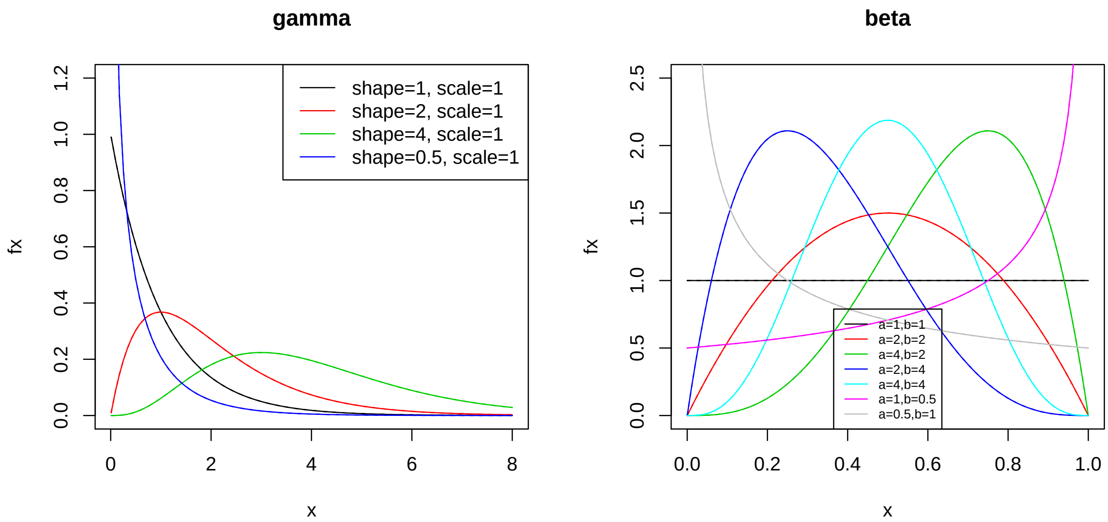

Figura 6.1: Funcions de densitat de diverses distribucions Gamma (esquerra) i Beta per a diferents combinacions de paràmetres.

La llei Gamma és una distribució útil per a modelar variables no negatives i, en particular, temps entre esdeveniments d'interès, com ara tems de supervivència de pacients diagnosticats d'una malaltia terminal, o temps de funcionament de màquines fins que s'espatllen.

En determinats llibres podeu trobar una altra parametrització d'aquesta distribució, que s'obté definint el paràmetre d'escala $b = 1/\beta$ (`scale`, en anglès):

$$f(x|\alpha, b) = \begin{cases} \frac{1}{\Gamma(\alpha) b^\alpha} x^{\alpha-1} e^{-\frac{x}{b}} & x > 0 \\ 0 & \text{altrament} \end{cases},$$

La Figura 6.1 (esquerra) mostra la funció de densitat de la Gamma per diferents combinacions dels paràmetres α i $b = 1/\beta$ (`shape` i `scale`, respectivament, al gràfic).

Exemple 6.6.1.
Suposem $Z \sim \mathrm{N}(0,1)$. Si definim $X = Z^2$ llavors

$$F_X(x) = \Pr(X \le x) = \Pr(-\sqrt{x} \le Z \le \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}).$$

$$f_X(x) = \varphi(\sqrt{x})\frac{1}{2\sqrt{x}} + \varphi(-\sqrt{x})\frac{1}{2\sqrt{x}} = \frac{1}{\sqrt{x}}\varphi(\sqrt{x}) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}.$$

Notem que prenent $\alpha = \beta = 1/2$ hem provat que la llei de X és $\gamma(1/2, 1/2)$ i que

$$\frac{\beta^\alpha}{\Gamma(\alpha)} = \frac{1}{\sqrt{2\pi}} \Rightarrow \Gamma(1/2) = \sqrt{\pi}.$$

La distribució $\gamma(1/2, 1/2)$ també s'anomena χ^2_1.

$\square$

Un canvi en les unitats de mesura implica un canvi en el valor del paràmetre β però no afecta la forma de la distribució.

Exemple 6.6.2.
Sigui $X \sim \gamma(\alpha, \beta)$ i $Y = aX$, $a > 0$. Llavors

$$f_Y(y) = f_X\left(\frac{y}{a}\right)\frac{1}{a} = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{y}{\alpha}\right)^{\alpha-1} e^{-\beta\frac{y}{a}}\frac{1}{a}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1}\frac{1}{a^\alpha}e^{-\beta\frac{y}{a}} = \left(\frac{\beta}{a}\right)^\alpha \frac{1}{\Gamma(\alpha)} y^{\alpha-1}e^{-\frac{\beta}{a}y}.$$

Per tant $Y \sim \gamma(\alpha, \frac{\beta}{a})$.

$\square$

Observació: La funció de distribució d'una Gamma no admet una expressió analítica:

$$F(x|\alpha,\beta) = \int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} u^{\alpha-1}e^{-\beta u}\, du.$$

Si definim la FUNCIÓ GAMMA INCOMPLETA (que està escrita en taules):

$$I(k, x) = \frac{1}{\Gamma(k)}\int_0^x u^{k-1}e^{-u}\, du,$$

es pot comprovar que $F(x|\alpha, \beta) = I(\alpha, \beta x)$.

Càlcul d'alguns moments

Calculem moments importants d'una variable aleatòria X que segueix una llei Gamma:

$$\mathbb{E}(X) = \int_0^\infty x\ f(x|\alpha,\beta)\, dx = \int_0^\infty x\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\, dx$$

$$= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\beta x}\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \int_0^\infty \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\beta x}\, dx$$

$$= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{1}{\beta} \int_0^\infty f(x|\alpha+1,\beta)\, dx = \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)} \frac{1}{\beta} = \frac{\alpha}{\beta}.$$

$$\mathbb{E}(X^2) = \frac{\alpha(\alpha+1)}{\beta^2}.$$

Amb els moments calculats podem trobar la variància fàcilment:

$$\mathrm{Var}(X) = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}.$$

Raonant com abans i aplicant el principi de inducció, és fàcil provar que

$$\mathbb{E}(X^k) = \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{\beta^k}$$

per a tot k natural.

Casos particulars de les distribucions Gamma

- Si $\alpha = 1$ i $\beta > 0$ tenim la llei exponencial.
- Si $\alpha = \frac{n}{2}$, amb $n \in \mathbb{N}$ i $\beta = 1/2$ tenim una χ^2_n. En particular, recordeu que si $Z \sim N(0,1)$ aleshores $Z^2 \sim \chi^2_1$. Fent servir la propietat que diu que la suma de Gammes independents és també Gamma (ho provarem més endavant) es té que si $Z_1,\ldots,Z_n$ són $N(0,1)$ independents, aleshores
$$\sum_{i=1}^n Z_i^2 \sim \chi^2_n.$$
- Si $X \sim \gamma(k,\beta)$ aleshores $2\beta Y \sim \gamma(\frac{2k}{2}, 1/2) = \gamma(k,1/2) \equiv \chi^2_{2k}$.

Exemple 6.6.3.

Càlcul del coeficient d'apuntament de la Normal

Sigui $X \sim N(\mu,\sigma^2)$ i sigui $Z = (X-\mu)/\sigma$. El coeficient d'apuntament de X es defineix com $\mathbb{E}(Z^4)$. Sigui $Y = Z^2$. Hem vist que $Y \sim \chi^2_1 \equiv \gamma(1/2,1/2)$. Per tant

$$\mathbb{E}(Y) = \frac{1/2}{1/2} = 1,\ V(Y) = \frac{1/2}{(1/2)^2} = 2.$$

Observeu que

$$\mathrm{CAp}(X) = \mathbb{E}(Z^4) = \mathbb{E}(Y^2) = V(Y) + \mathbb{E}(Y)^2 = 2 + 1 = 3.$$

$\square$

Proposició 6.6.3. *Si $X_1 \sim Gamma(\alpha_1, \beta)$ i $X_2 \sim Gamma(\alpha_2, \beta)$ són independents, aleshores $X_1 + X_2 \sim Gamma(\alpha_1 + \alpha_2, \beta)$.*

Demostració: Recordeu que si X i Y són variables aleatòries contínues i independents, aleshores la funció de densitat de la seva suma és la convolució de les seves funcions de densitat:

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx \\ &= \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy. \end{aligned}$$

Aplicant aquesta fórmula, la funció de densitat de $X_1 + X_2$ serà

$$\begin{aligned} f_{X_1+X_2}(y) = & \int_0^y \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} (y-x)^{\alpha_1-1} x^{\alpha_2-1} e^{-\beta(y-x)-\beta x} dx \\ & \text{(fem el canvi de variable } t = x/y,\ dt = (1/y)dx) \\ = & \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y^{\alpha_1+\alpha_2-1} e^{-\beta y} \int_0^1 (1-t)^{\alpha_1-1} t^{\alpha_2-1} dt \\ = & \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y^{\alpha_1+\alpha_2-1} e^{-\beta y} \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} \\ = & \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} y^{\alpha_1+\alpha_2-1} e^{-\beta y} \end{aligned}$$

i es conclou que $X_1 + X_2 \sim \mathrm{Gamma}(\alpha_1 + \alpha_2, \beta)$.

Hem fet servir la igualtat

$$\int_0^1 (1-t)^{\alpha_1-1} t^{\alpha_2-1} dt = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}.$$

que es prova a continuació:

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \int_0^{\infty} y^{\beta-1} e^{-y} dy = \int_0^{\infty} \int_0^{\infty} x^{\alpha-1} y^{\beta-1} e^{-(x+y)} dy dx$$

$$\left\{\begin{array}{c} s=x+y \\ y=s-x \\ ds=dy \end{array}\right\} \overset{}{=} \int_0^\infty \int_x^\infty x^{\alpha-1}(s-x)^{\beta-1}e^{-s}dsdx = \int_0^\infty \int_0^s x^{\alpha-1}(s-x)^{\beta-1}e^{-s}dxds$$

$$= \int_0^\infty s^{\alpha+\beta-2}e^{-s} \int_0^s \frac{x^{\alpha-1}}{s^{\alpha-1}}\frac{(s-x)^{\beta-1}}{s^{\beta-1}}dxds$$

$$= \int_0^\infty s^{\alpha+\beta-1}\frac{1}{s}e^{-s} \int_0^s \left(\frac{x}{s}\right)^{\alpha-1}\left(1-\frac{x}{s}\right)^{\beta-1} dxds$$

$$\left\{\begin{array}{c} t=x/s \\ dt=(1/s)dx \end{array}\right\} = \int_0^\infty s^{\alpha+\beta-1}e^{-s} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dtds$$

$$= \int_0^\infty s^{\alpha+\beta-1}e^{-s}ds \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt = \Gamma(\alpha+\beta)\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt.$$

□

A2. La família de distribucions Beta

A partir de la igualtat

$$\Gamma(\alpha)\Gamma(\beta) = \Gamma(\alpha+\beta)\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$$

es defineix la funció *beta*:

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Observeu també que

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}t^{\alpha-1}(1-t)^{\beta-1}dt = 1,$$

d'on es segueix que la funció

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}t^{\alpha-1}(1-t)^{\beta-1}I_{[0,1]}(t)$$

és una funció de densitat. Això permet fer la següent definició.

Definició 6.6.4. *Una variable aleatòria X direm que segueix una* LLEI BETA *amb paràmetres $\alpha > 0$ i $\beta > 0$ si*

$$f(x|\alpha,\beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & si\ 0 < x < 1 \\ 0 & si\ x \leq 0 \end{cases}.$$

Un cas particular d'aquesta distribució és $U[0,1]=B(1,1)$.

La Figura 6.1 (dreta) mostra la funció de densitat de la beta per diferents combinacions dels paràmetres a i b (a les funcions d'R corresponents a la distribució beta s'anomenen `shape1` i `shape1`).

Es pot provar que

$$\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta},\ \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Apèndix B: Distribucions relacionades amb la normal

La primera de les distribucions relacionades amb la normal que definirem és la χ^2 (llegim *txi quadrat*), que ja ha aparegut quan hem definit la distribució Gamma.

Definició 6.6.5. *Siguin $X \sim N(0,1)$ i sigui $Y = X^2$. Direm que Y segueix una distribució χ^2_1, que llegirem* TXI QUADRAT AMB UN GRAU DE LLIBERTAT.

Siguin $X_1, \ldots, X_n$ v.a.i.i.d. $N(0,1)$ i sigui $Y_n = \sum_{i=1}^n X_i^2$. Direm que Y_n segueix una distribució χ^2_n, que llegirem TXI QUADRAT AMB n GRAUS DE LLIBERTAT.

Hem vist abans (quan hem parlat de la distribució Gamma) que $Y_n \sim \chi^2_n \iff Y \sim \gamma(n/2, 1/2)$.

Definició 6.6.6. *Siguin $X \sim N(0,1)$ i $Y \sim \chi^2_r$ independents. Sigui*

$$T = \frac{X}{\sqrt{Y/r}}.$$

Es diu que T segueix una distribució t de Student amb r graus de llibertat i s'escriu $T \sim t(r)$ o $T \sim t_r$.

Proposició 6.6.7. *Si T és una t de Student amb r graus de llibertat la seva funció de densitat és*

$$f_T(t) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\,\Gamma(r/2)} \left(1 + \frac{t^2}{r}\right)^{-(r+1)/2},\ -\infty < t < \infty.$$

Proposició 6.6.8. *Sigui T una t de Student amb r graus de llibertat. Aleshores*

$$\mathbb{E}(T) = 0 \text{ si } r > 1$$

(per $r \in (0,1]$ la esperança no existeix) i

$$Var(T) = \frac{r}{r-2} \text{ si } r > 2$$

(per $r \in (1,2]$ la variància és infinita i per $r \in (0,1]$ no existeix).

Definició 6.6.9. *Siguin $X \sim \chi^2_r$ i $Y \sim \chi^2_s$ independents. Sigui*

$$F = \frac{X/r}{Y/s}.$$

Es diu que F segueix una distribució F amb r i s gruas de llibertat, i s'escriu $F \sim F(r,s)$ o $F \sim F_{r,s}$.

Proposició 6.6.10. *Sigui F amb distribució F amb r i s gruas de llibertat. La seva funció de densitat és*

$$f(x) = \frac{r\Gamma((r+s)/2)}{s\Gamma(r/2)\Gamma(s/2)} \frac{(rx/s)^{(r/2)-1}}{[1+(rx/s)]^{(r+s)/2}},\ x > 0.$$

Proposició 6.6.11. $F^{-1} \sim F(s,r)$ *i* $T^2 \sim F(1,r)$ *si* $T \sim t(r)$.

Proposició 6.6.12. *Siguin $X_1, \ldots, X_n$ v.a.i.i.d. $N(\mu_1, \sigma_1^2)$, i $Y_1, \ldots, Y_m$ v.a.i.i.d. $N(\mu_2, \sigma_2^2)$, dues mostres independents de dues distribucions normals. Siguin S_X^2 i S_Y^2 les variàncies mostrals calculades a partir d'aquestes mostres. Aleshores*

$$\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F(n-1, m-1).$$

Proposició 6.6.13. *Sigui F amb distribució F amb r i s graus de llibertat. Aleshores*

$$\mathbb{E}(F) = \frac{s}{s-2} \text{ si } s > 2$$

(per $s \in (0,2]$ la esperança és infinita) i

$$Var(F) = \frac{2s^2(r+s-2)}{r(s-2)^2(s-4)} \text{ si } s > 4$$

(per $s \in (2,4]$ la variància és infinita i per $s \in (0,2]$ no existeix).