Estimating two parameters.

We have seen many distributions with two (real-valued) parameters, e.g. Negative Binomial (discrete), Normal, gamma, Cauchy, beta and Uniform (continuous). In the following, we denote the two parameters θ and λ, and assume their (unknown) values are to be estimated based on a random independent sample of size n (denoted $X_1,\ X_2, ...X_n$) from the corresponding distribution. Similarly to estimating one parameter, we need to differentiate between the REGULAR case of estimation when the (now two-dimensional) *support* of the distribution is *not* affected by the values of θ and λ, and the exceptional non-regular case (such as Uniform distribution). In this section, we discuss the **regular case** only.

In analogy with one-parameter estimation, one can show (this time, we do not go over the actual proof) that the *variance-covariance* matrix (say $\mathbb{V}$) of the sampling distribution of any *unbiased* estimators of θ and λ (denoted $\hat{\theta}$ and $\hat{\lambda}$), i.e.

$$\begin{bmatrix} \operatorname{Var}\left(\hat{\theta}\right) & \operatorname{Cov}\left(\hat{\theta},\hat{\lambda}\right) \\ \operatorname{Cov}\left(\hat{\theta},\hat{\lambda}\right) & \operatorname{Var}\left(\hat{\lambda}\right) \end{bmatrix}$$

must be (in a sense explained shortly) bigger or equal to the following Rao-Cramer bound (we will still refer to it by the RCV acronym) for all potential values of θ and λ:

$$\text{RCV} = \frac{1}{n}\cdot\begin{bmatrix} -\mathbb{E}\left(\dfrac{\partial^2 \ln f(x|\theta,\lambda)}{\partial\theta^2}\right) & -\mathbb{E}\left(\dfrac{\partial^2 \ln f(x|\theta,\lambda)}{\partial\theta\partial\lambda}\right) \\ -\mathbb{E}\left(\dfrac{\partial^2 \ln f(x|\theta,\lambda)}{\partial\theta\partial\lambda}\right) & -\mathbb{E}\left(\dfrac{\partial^2 \ln f(x|\theta,\lambda)}{\partial\lambda^2}\right) \end{bmatrix}^{-1}$$

To be able to compute this RCV, one must remember that a 2 by 2 matrix is inverted using the following scheme

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{d}{ad-bc} & \dfrac{-b}{ad-bc} \\ \dfrac{-c}{ad-bc} & \dfrac{a}{ad-bc} \end{bmatrix}$$

Also: saying that $\mathbb{V} \geq$ RCV means that (i) $\operatorname{Var}\left(\hat{\theta}\right) \geq$ first main-diagonal element of RCV, (ii) $\operatorname{Var}\left(\hat{\theta}\right) \geq$ second main-diagonal element of RCV, and (iii) the determinant of $\mathbb{V}-$ RCV (the product of its main-diagonal elements minus the product of the two off-diagonal elements) cannot be negative; the individual bound on each variance is of particular interest to us. In addition to providing a lower bound for each variance, the same RCV is also the actual *asymptotic* variance-covariance matrix of the corresponding *maximum-likelihood* estimators (whose distribution is, approximately, *bivariate* Normal). Finding MLEs is done in the next section; now we go over a few examples of computing the RCV matrix.

Examples:

- Normal distribution ($\sigma > 0$)

$$
\begin{aligned}
\ln f &= -\frac{(X-\mu)^2}{2\sigma^2} - \ln\sigma - \frac{\ln(2\pi)}{2} \\
\begin{bmatrix} -\dfrac{\partial^2 \ln f}{\partial \mu^2} & -\dfrac{\partial^2 \ln f}{\partial \mu \partial \sigma} \\ -\dfrac{\partial^2 \ln f}{\partial \mu \partial \sigma} & -\dfrac{\partial^2 \ln f}{\partial \sigma^2} \end{bmatrix} &= \begin{bmatrix} \dfrac{1}{\sigma^2} & \dfrac{2(X-\mu)}{\sigma^3} \\ \dfrac{2(X-\mu)}{\sigma^3} & 3\dfrac{(X-\mu)^2}{\sigma^4} - \dfrac{1}{\sigma^2} \end{bmatrix} \\
\text{RCV} &= \frac{1}{n} \cdot \begin{bmatrix} \dfrac{1}{\sigma^2} & 0 \\ 0 & \dfrac{2}{\sigma^2} \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{\sigma^2}{2n} \end{bmatrix}
\end{aligned}
$$

  since $\mathbb{E}(X) = \mu$ and $\mathbb{E}\left((X-\mu)^2\right) = \sigma^2$. Note that inverting a diagonal matrix is easy - replace each diagonal element by its reciprocal; the two ML estimators will be asymptotically uncorrelated.

- gamma distribution (both parameters and X itself must be positive)

$$
\begin{aligned}
\ln f &= (\alpha - 1)\ln X - \frac{X}{\beta} - \ln\Gamma(\alpha) - \alpha\ \ln\beta \\
\begin{bmatrix} -\dfrac{\partial^2 \ln f}{\partial \alpha^2} & -\dfrac{\partial^2 \ln f}{\partial \alpha \partial \beta} \\ -\dfrac{\partial^2 \ln f}{\partial \alpha \partial \beta} & -\dfrac{\partial^2 \ln f}{\partial \beta^2} \end{bmatrix} &= \begin{bmatrix} \Psi'(\alpha) & \dfrac{1}{\beta} \\ \dfrac{1}{\beta} & \dfrac{2X}{\beta^3} - \dfrac{\alpha}{\beta^2} \end{bmatrix} \\
\text{RCV} &= \frac{1}{n} \cdot \begin{bmatrix} \Psi'(\alpha) & \dfrac{1}{\beta} \\ \dfrac{1}{\beta} & \dfrac{\alpha}{\beta^2} \end{bmatrix}^{-1} = \frac{\begin{bmatrix} \alpha & -\beta \\ -\beta & \beta^2\Psi'(\alpha) \end{bmatrix}}{n \cdot (\alpha\ \Psi'(\alpha) - 1)}
\end{aligned}
$$

  meaning that *each* element of the matrix is divided by the denominator. This implies that the asymptotic correlation coefficient of the two ML estimators equals $-\dfrac{1}{\sqrt{\alpha \cdot \Psi'(\alpha)}}$ (it approaches -1 as α increases; the two errors will be highly correlated). Note that $\Psi(\alpha)$ is a shorthand for $\dfrac{d\ln\Gamma(\alpha)}{d\alpha}$, and thus $\Psi(\alpha)' = \dfrac{d^2\ln\Gamma(\alpha)}{d\alpha^2}$, and recall that $\mathbb{E}(X) = \alpha\beta$.

- Cauchy ($\hat{\sigma} > 0$):

$$\begin{aligned}
\ln f &= \ln \tilde{\sigma} - \ln(\tilde{\sigma}^2 + (X - \tilde{\mu})^2) - \ln \pi \\
\begin{bmatrix} -\frac{\partial^2 \ln f}{\partial \tilde{\mu}^2} & -\frac{\partial^2 \ln f}{\partial \tilde{\mu}\partial\tilde{\sigma}} \\ -\frac{\partial^2 \ln f}{\partial \tilde{\mu}\partial\tilde{\sigma}} & -\frac{\partial^2 \ln f}{\partial \tilde{\sigma}^2} \end{bmatrix} &= \begin{bmatrix} \frac{2\left(\tilde{\sigma}^2 - (X-\tilde{\mu})^2\right)}{\left(\tilde{\sigma}^2 + (X-\tilde{\mu})^2\right)^2} & \frac{4(X-\tilde{\mu})\tilde{\sigma}}{\left(\tilde{\sigma}^2 + (X-\tilde{\mu})^2\right)^2} \\ \frac{4(X-\tilde{\mu})\tilde{\sigma}}{\left(\tilde{\sigma}^2 + (X-\tilde{\mu})^2\right)^2} & \frac{1}{\tilde{\sigma}^2} + \frac{2}{\tilde{\sigma}^2 + (X-\tilde{\mu})^2} - \frac{4\tilde{\sigma}^2}{\left(\tilde{\sigma}^2 + (X-\tilde{\mu})^2\right)^2} \end{bmatrix} \\
\text{RCV} &= \frac{1}{n} \cdot \begin{bmatrix} \frac{1}{2\tilde{\sigma}^2} & 0 \\ 0 & \frac{1}{2\tilde{\sigma}^2} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{2\tilde{\sigma}^2}{n} & 0 \\ 0 & \frac{2\tilde{\sigma}^2}{n} \end{bmatrix}
\end{aligned}$$

  where the expected values can be easily computed by Maple. The two ML estimators are asymptotically uncorrelated.

- Log-normal (a more 'exotic' distribution, $b > 0$ and $X > 0$):

$$\begin{aligned}
\ln f &= -\frac{(\ln X - a)^2}{2b^2} - \ln X - \ln b - \frac{\ln(2\pi)}{2} \\
\begin{bmatrix} -\frac{\partial^2 \ln f}{\partial a^2} & -\frac{\partial^2 \ln f}{\partial a\partial b} \\ -\frac{\partial^2 \ln f}{\partial a\partial b} & -\frac{\partial^2 \ln f}{\partial b^2} \end{bmatrix} &= \begin{bmatrix} \frac{1}{b^2} & \frac{2(\ln X - a)}{b^3} \\ \frac{2(\ln X - a)}{b^3} & \frac{3(\ln X - a)^2 - b^2}{b^4} \end{bmatrix} \\
\text{RCV} &= \frac{1}{n} \cdot \begin{bmatrix} \frac{1}{b^2} & 0 \\ 0 & \frac{2}{b^2} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{b^2}{n} & 0 \\ 0 & \frac{b^2}{2n} \end{bmatrix}
\end{aligned}$$

  since $\mathbb{E}\,(\ln X - a) = 0$ and $\mathbb{E}\left((\ln X - a)^2\right) = b^2$ (Maple).

- Laplace (ditto, $b > 0$):

$$\begin{aligned}
\ln f &= -\frac{|X - a|}{b} - \ln b - \ln 2 \\
\begin{bmatrix} -\frac{\partial^2 \ln f}{\partial a^2} & -\frac{\partial^2 \ln f}{\partial a\partial b} \\ -\frac{\partial^2 \ln f}{\partial a\partial b} & -\frac{\partial^2 \ln f}{\partial b^2} \end{bmatrix} &= \begin{bmatrix} \frac{\text{sign}'(x-a)}{b} & \frac{\text{sign}(x-a)}{b^2} \\ \frac{\text{sign}(x-a)}{b^2} & \frac{2\,|x-a|\,-b}{b^3} \end{bmatrix} \\
\text{RCV} &= \frac{1}{n} \cdot \begin{bmatrix} \frac{1}{b^2} & 0 \\ 0 & \frac{1}{b^2} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{b^2}{n} & 0 \\ 0 & \frac{b^2}{n} \end{bmatrix}
\end{aligned}$$

  The 'sign' function is equal to 1 or -1, depending on the sign of its argument (positive or negative, respectively). It is needed only in this example (students will never need to use it).

Finding MLEs by solving the corresponding two normal equations.

- Normal distribution. Solve

$$\begin{aligned}\frac{\sum_{i=1}^{n}(X_i-\mu)}{\sigma^2} &= 0\\ \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^3}-\frac{n}{\sigma} &= 0\end{aligned}$$

Solution:

$$\begin{aligned}\hat{\mu} &= \bar{X}\\ \hat{\sigma} &= \sqrt{\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n}}=\sqrt{\overline{X^2}-\bar{X}^2}\equiv\sqrt{S^2}\end{aligned}$$

Note that we are dividing by n, not by $n-1$; for large n, this is of minor difference.

- gamma distribution. Solve

$$\begin{aligned}\sum_{i=1}^{n}\ln X_i-n\ \Psi(\alpha)-n\ln\beta &= 0\\ \frac{\sum_{i=1}^{n}X_i}{\beta^2}-\frac{n\ \alpha}{\beta} &= 0\end{aligned}$$

The second equation implies that $\hat{\beta}=\frac{\bar{X}}{\hat{\alpha}}$. Substitute into the first equation and get

$$\ln\hat{\alpha}-\Psi(\hat{\alpha})=\ln\bar{X}-\overline{\ln X}$$

which needs to be solved numerically for $\hat{\alpha}$ (Maple).

- For Cauchy distribution, we have to solve

$$\begin{aligned}\sum_{i=1}^{n}\frac{X_i-\tilde{\mu}}{\tilde{\sigma}^2+(X_i-\tilde{\mu})^2} &= 0\\ 2\sum_{i=1}^{n}\frac{\tilde{\sigma}^2}{\tilde{\sigma}^2+(X_i-\tilde{\mu})^2} &= n\end{aligned}$$

This can be done only numerically by Maple's 'fsolve'.

- Log-normal: Solve

$$\begin{aligned}\frac{\sum_{i=1}^{n}(\ln X_i-a)}{b^2} &= 0\\ \frac{\sum_{i=1}^{n}(\ln X_i-a)^2-n\ b^2}{b^3} &= 0\end{aligned}$$

implying that

$$\begin{aligned}\hat{a} &= \overline{\ln X}\\ \hat{b} &= \sqrt{\overline{(\ln X)^2}-\overline{\ln X}^2}\end{aligned}$$

Note the similarity with Normal distribution (X is replaced by $\ln X$).

- Laplace distribution with the following pdf

$$f(x) = \frac{\exp\left(-\frac{|x-a|}{b}\right)}{2b}$$

(for any real x), where $b > 0$. The ln of the correponding likelihood function is given by

$$-\frac{\sum_{i=1}^{n}|X_i - a|}{b} - n\ln b - n\ln 2$$

It is obvious that to minimize the denominator (whatever the value of b is), one must take $\hat{a} = \tilde{X}$ (the sample median) - just visualize what happens when a has k observations to its left and $m > k$ observations to its right (by increasing a the sum of the absolute values of the differences correspondingly decreases, right?). Once we have solved for $\hat{a}$, maximizing the above expression with respect to b is done by making its b derivative equal to zero, thus:

$$\frac{\sum_{i=1}^{n}|X_i - \hat{a}|}{b^2} - \frac{n}{b} = 0$$

implying that

$$\hat{b} = \frac{\sum_{i=1}^{n}|X_i - \hat{a}|}{n} = \overline{\left|X - \tilde{X}\right|}$$

Again, we know how to find the (exact) distribution of $\tilde{X}$ (see the order-statistics chapter); the technique itself guarantees that both estimators are asymptotically unbiased, their RCV indicates that each of them has the asymptotic variance of $\frac{b^2}{n}$ and that they are asymptotically uncorrelated.

We can also get MLEs for any **non-regular** case by similarly maximizing the likelihood function (or its ln, whichever is more convenient); this time, it is *not* done by solving normal equations. Also, the (asymptotic) variance-covariance matrix of the two estimators must be computed separately (RCV does not exists).

Example:

- $\mathcal{U}(a,b)$. The corresponding likelihood function is

$$\frac{\prod_{i=1}^{n} G_{a,b}[X_i]}{(b-a)^n} = \frac{G_{a,b}[X_{(1)}]G_{X_{(1)},b}[X_{(n)}]}{(b-a)^n}$$

To maximize it, we have to make b as small as possible, but not smaller than $X_{(n)}$ (at which point the LF would drop down to 0); at the same time a must be made as large as possible (minimizing the denominator), but not bigger than $X_{(1)}$, for the same reason. This implies that the two ML estimators are $\hat{a} = X_{(1)}$ and $\hat{b} = X_{(n)}$; we know how to construct their (exact) joint pdf, based on what we learned in the previous chapter (on order statistics). Since their (exact) correlation coefficient is $\frac{1}{n}$, they are *asymptotically* uncorrelated; both of their (exact) variances are equal to $\frac{n\,(b-a)^2}{(n+1)^2(n+2)}$.

MM estimators

These are sometimes identical to MLEs, but more often inferior to them (having bigger variances, in non-regular cases dramatically so!); on occations, they may not even exist. On top of it, getting their variance-covariance matrix is now a lot more complicated. They are thus clearly an obsolete 'hangover' from the days before computers, but we will still explain how the technique works and give a few examples.

The idea is to solve $\mathbb{E}(X) = g(\theta, \lambda)$ and $\mathbb{E}\left(X^2\right) = h(\theta, \lambda)$ for θ and λ (these will be our MM estimators), and then replace $\mathbb{E}(X)$ and $\mathbb{E}\left(X^2\right)$ by the corresponding sample means $\bar{X}$ and $\overline{X^2}$ respectively; the resulting estimators will always be (regular case or not) asymptotically bivariate Normal.

Examples:

- $\mathcal{N}(\mu, \sigma)$. Since $\mathbb{E}(X) = \mu$ and $\mathbb{E}\left(X^2\right) = \sigma^2 + \mu$, we get $\hat{\mu} = \bar{X}$ and $\hat{\sigma} = \sqrt{\overline{X^2} - \bar{X}^2}$ (same as MLEs).

- gamma(α, β). Recall that

$$f(x) = \frac{x^{\alpha-1}\exp\left(-\dfrac{x}{\beta}\right)}{\Gamma(\alpha)\beta^{\alpha}} \qquad \text{when} \quad x > 0$$

and compute

$$\begin{aligned} \mathbb{E}(X) &= \alpha\,\beta \\ \mathbb{E}\left(X^2\right) &= \alpha\,\beta^2 + \alpha^2\beta^2 \\ \mathrm{Var}(X) &= \alpha\beta^2 \\ \mathrm{Var}\left(X^2\right) &= 2\alpha(1+\alpha)(3+2\alpha)\beta^4 \\ \mathrm{Cov}\left(X, X^2\right) &= 2\alpha(1+\alpha)\beta^3 \end{aligned}$$

Solve

$$\begin{aligned} \alpha\,\beta &= \bar{X} \\ \alpha\,\beta^2 + \alpha^2\beta^2 &= \overline{X^2} \end{aligned}$$

which yields

$$\begin{aligned} \hat{\alpha} &= \frac{\bar{X}^2}{\overline{X^2} - \bar{X}^2} \\ \hat{\beta} &= \frac{\overline{X^2} - \bar{X}^2}{\bar{X}} \end{aligned}$$

Taylor-expand each of them in terms of $\bar{X}$ (at $\alpha\beta$) and $\overline{X^2}$ (at $\alpha\beta^2+\alpha^2\beta^2$), which is easily done by Maple's 'mtaylor', getting

$$\begin{aligned}\hat{\alpha} &= \alpha + \frac{2(1+\alpha)}{\beta}\cdot\left(\bar{X}-\alpha\beta\right) - \frac{1}{\beta^2}\left(\overline{X^2}-\beta^2-\alpha^2\beta^2\right)+...\\ \hat{\beta} &= \beta - \frac{1+2\alpha}{\alpha}\cdot\left(\bar{X}-\alpha\beta\right) + \frac{1}{\alpha\beta}\left(\overline{X^2}-\beta^2-\alpha^2\beta^2\right)+...\end{aligned}$$

(each a linear combination of RVs whose variances and covariance are known). Here you have to recall that

$$\begin{aligned}g(x,y) &= g(x_0,y_0) + \left.\frac{\partial g(x,y)}{\partial x}\right|_{x=x_0,y=y_0}\cdot(x-x_0)\\ &\quad+\left.\frac{\partial g(x,y)}{\partial y}\right|_{x=x_0,y=y_0}\cdot(y-y_0)+...\end{aligned}$$

(Maple's 'mtaylor' will do it for you).
This implies that

$$\begin{aligned}&\mathrm{Var}(\hat{\alpha}) \simeq \left(\frac{2(1+\alpha)}{\beta}\right)^2\cdot\mathrm{Var}(\bar{X})+\left(-\frac{1}{\beta^2}\right)^2\cdot\mathrm{Var}(\overline{X^2})\\ &+2\cdot\frac{2(1+\alpha)}{\beta}\cdot\left(-\frac{1}{\beta^2}\right)\mathrm{Cov}(\bar{X},\overline{X^2}) = \frac{2\alpha(1+\alpha)}{n}\end{aligned}$$

Similarly

$$\mathrm{Var}(\hat{\beta}) \simeq \frac{(3+2\alpha)\beta^2}{\alpha\ n}$$

and

$$\begin{aligned}&\mathrm{Cov}(\hat{\alpha},\hat{\beta}) \simeq \frac{2(1+\alpha)}{\beta}\cdot\left(-\frac{1+2\alpha}{\alpha}\right)\cdot\mathrm{Var}(\bar{X})\\ &+\left(-\frac{1}{\beta^2}\right)\cdot\frac{1}{\alpha\beta}\cdot\mathrm{Var}(\overline{X^2})\\ &+\left[\frac{2(1+\alpha)}{\beta}\cdot\frac{1}{\alpha\beta}+\left(-\frac{1+2\alpha}{\alpha}\right)\cdot\left(-\frac{1}{\beta^2}\right)\right]\cdot\mathrm{Cov}(\bar{X},\overline{X^2})\\ &=\frac{2\alpha(1+\alpha)\beta^3}{n}\end{aligned}$$

This can be converted to

$$\rho_{\hat{\alpha},\hat{\beta}} \simeq \frac{\mathrm{Cov}(\hat{\alpha},\hat{\beta})}{\sqrt{\mathrm{Var}(\hat{\alpha})\cdot\mathrm{Var}(\hat{\beta})}} = -\sqrt{\frac{2+2\alpha}{3+2\alpha}}$$

(a rather large negative correlation of the two estimators, implying that a positive error of one estimator will most likely go with a negative error of the other - this will be reflected by the corresponding confidence ellipse).

- Cauchy distribution has indefinite mean and infinite variance - there are no MMEs.
- Log-Normal. Since $\mathbb{E}(X) = \exp\left(a + \frac{b^2}{2}\right)$ and $\mathbb{E}(X^2) = \exp(2a + 2b^2)$, solving for b yields

$$\hat{b} = \sqrt{\ln \frac{\overline{X^2}}{\bar{X}^2}}$$

implying that

$$\hat{a} = 2\ln \bar{X} - \frac{\ln \overline{X^2}}{2}$$

We will not explore their asymptotic properties.

- $\mathcal{U}(a, b)$. Solving $\bar{X} = \frac{a+b}{2}$ and $\overline{X^2} = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4}$, we get

$$\begin{aligned} \hat{a} &= \bar{X} - \sqrt{3S^2} \\ \hat{b} &= \bar{X} + \sqrt{3S^2} \end{aligned}$$

Expanding:

$$\begin{aligned} \hat{a} &= a + \frac{2a+4b}{b-a}\left(\bar{X} - \frac{a+b}{2}\right) + \frac{3}{b-a}\left(\overline{X^2} - \frac{a^2+ab+b^2}{2}\right) + \ldots \\ \hat{b} &= a + \frac{4a+2b}{b-a}\left(\bar{X} - \frac{a+b}{2}\right) + \frac{3}{b-a}\left(\overline{X^2} - \frac{a^2+ab+b^2}{2}\right) + \ldots \end{aligned}$$

which leads to $\mathrm{Var}(\hat{a}) = \frac{2(b-a)^2}{15n}$, $\mathrm{Var}\left(\hat{b}\right) = \frac{2(b-a)^2}{15n}$, and $\rho_{\hat{a},\hat{b}} = \frac{1}{4}$.

- Laplace. Since $\mathbb{E}(X) = 0$ (not a function of a and/or b), the MM technique does not work!

Joint sufficient statistics

The idea is similar to the unvariate case: simplify the joint pdf of X_1, $X_2, \ldots X_n$, delete factors free of θ and λ, delete factors free of the X_is, and if in what remains you find only two combinations (i.e. single-valued functions) of the X_is, there are the two sufficient statistics for estimating θ and λ. They contain all the information there is about the values of the two paramteters, meaning that any sensible estimators should be built out of these sufficient statistics (any other estimators can be improved by properly designed functions of sufficient statistics). The trouble is that sufficient statistics may not always exist; but when they do, they enable us to build the 'best' estimators (with the smallest variance-covariance matrix). This provides justification for calling the MLEs the 'best' even in the non-regular case (as they will always be functions of sufficient statistics). In the regular case, each sufficient statistic is invariably a sample sum or a product (see our examples); in the latter case, it is advisable to take its ln, thus converting it into the corresponding sum of logarithms (an equivalent way of presenting it, as it contains the same information).

Examples:

- Normal

$$\frac{\prod_{i=1}^{n} \exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)}{(2\pi)^{n/2}\sigma^n} = \frac{\exp\left(-\frac{\sum_{i=1}^{n} X_i^2 - 2\mu\sum_{i=1}^{n} X_i + n\mu^2}{2\sigma^2}\right)}{(2\pi)^{n/2}\sigma^n}$$

  implying that $\sum_{i=1}^{n} X_i^2$ and $\sum_{i=1}^{n} X_i$ are jointly sufficient for estimating μ and σ.

- gamma

$$\frac{\prod_{i=1}^{n} X_i^{\alpha-1} \exp\left(-\frac{X_i}{\beta}\right)}{\Gamma(\alpha)^n \beta^{n\ \alpha}} = \frac{\left(\prod_{i=1}^{n} X_i\right)^{\alpha-1} \exp\left(-\frac{\sum_{i=1}^{n} X_i}{\beta}\right)}{\Gamma(\alpha)^n \beta^{n\ \alpha}}$$

  making it obvious that $\sum_{i=1}^{n} X_i$ and $\prod_{i=1}^{n} X_i$ or, equivalently $\sum_{i=1}^{n} \ln X_i$ are jointly sufficient for estimating α and β.

- For Cauchy, such a separation becomes impossible (there are no sufficient statistics for this estimation).

- Log-normal

$$\frac{\prod_{i=1}^{n} \exp\left(-\frac{(\ln X_i - a)^2}{2b^2}\right)}{(2\pi)^{n/2} b^n \prod_{i=1}^{n} X_i} = \frac{\exp\left(-\frac{\sum_{i=1}^{n} (\ln X_i)^2 - 2a\sum_{i=1}^{n} \ln X_i + na^2}{2b^2}\right)}{(2\pi)^{n/2} b^n \prod_{i=1}^{n} X_i}$$

  implying that $\sum_{i=1}^{n} (\ln X_i)^2$ and $\sum_{i=1}^{n} \ln X_i$ are jointly sufficient for estimating a and b.

- Uniform (non-regular)

$$\frac{\prod_{i=1}^{n} G_{a,b}[X_i]}{(b-a)^n} = \frac{G_{a,b}[X_{(1)}]G_{X_{(1)},b}[X_{(n)}]}{(b-a)^n}$$

  Here $X_{(1)}$ and $X_{(n)}$ are jointly sufficient for estimating a and b.