Lecture 12. Confidence intervals - - revisited

Igor Rychlik

Chalmers
Department of Mathematical Sciences

Probability, Statistics and Risk, MVE300 • Chalmers • May 2011

ML estimates for typical models:

Distribution	ML estimates	$(\sigma^2_{\mathcal{E}})^*$
$X \in \mathrm{Po}(\theta)$	$\theta^* = \bar{x}$	$\dfrac{\theta^*}{n}$
$K \in \mathrm{Bin}(n,p)$	$\theta^* = \dfrac{k}{n}$	$\dfrac{\theta^*(1-\theta^*)}{n}$
$X \in \mathrm{Exp}(\theta)$	$\theta^* = \bar{x}$	$\dfrac{(\theta^*)^2}{n}$
$X \in \mathrm{N}(m,\sigma^2)$	$\theta^* = (\bar{x}, s_n^2)$	$\left(\dfrac{s_n^2}{n}, \dfrac{2(s_n^2)^2}{n}\right)$

Example - times between earthquakes: Model exponential cdf $\mathrm{P}(T \le t) = 1 - \exp(-t/a)$; With $\theta = a$ table gives $a^* = 437.2$ days; Variance of estimation error

$$(\sigma^2_{\mathcal{E}})^* = \frac{(\theta^*)^2}{n} = \frac{437.2^2}{63} = 3034, \quad \mathrm{day}^2.$$

Hence the standard deviation is $\sqrt{3034} = 55.08$ days. "Common sense" uncertainty $437.2 \pm 2 \cdot 55.08$.

In general

- Choose a cdf $F(x;\theta)$ for data (θ unknown parameter to be selected).
- Compute likelihood function $L(\theta)$ (odds for θ). Find θ^* - the value of parameter maximizing the likelihood function (having maximal odds).
- $e = \theta - \theta^*$ - estimation error (unknown) and modeled as rv. $\mathcal{E}$.
  - If $\mathrm{E}[\mathcal{E}] = 0$ then estimation is unbiased.
  - If standard deviation of the error $\sigma_{\mathcal{E}} \to 0$ as $n \to \infty$ then estimation is consistent.
- For large n (number of observations) $\mathcal{E}$ is approximately normally distributed $N(0, \sigma_{\mathcal{E}}^2)$, $\sigma_{\mathcal{E}}$ is an estimated by $\sigma_{\mathcal{E}}^*$.

Error in expected time between earthquakes $\mathcal{E}$ is approx. $N(0, 3083)$.

Quantiles of $\mathcal{E}$:

The error distribution $F_{\mathcal{E}}(e)$ describes sizes of errors as well as the frequencies with which those occur. Often one uses quantiles e_α to describe the variability. Obviously we have that

$$P(e_{1-\alpha/2} \le \mathcal{E} \le e_{\alpha/2}) = 1 - \alpha, \quad \text{or} \quad P(\mathcal{E} \le e_\alpha) = 1 - \alpha.$$

If $\mathcal{E}$ is approximately $N\left(0, (\sigma^2_{\mathcal{E}})^*\right)$ then

$$e_{\alpha/2} \approx \lambda_{\alpha/2} \cdot \sigma^*_{\mathcal{E}}, \qquad e_{1-\alpha/2} \approx -\lambda_{\alpha/2} \cdot \sigma^*_{\mathcal{E}}.$$

Quantiles of the standard normal distribution.

α	0.10	0.05	0.025	0.01	0.005	0.001
λ_α	1.28	1.64	1.96	2.33	2.58	3.09

Approximative confidence interval:

Example - times between earthquakes: Let choose $\alpha = 0.05$, then

$e_{\alpha/2} \approx 1.96 \cdot \sqrt{3083} = 108.8,$ and $e_{1-\alpha/2} \approx -1.96 \cdot \sqrt{3083} = -108.8.$

Consequently $\mathrm{P}(-108.8 \leq \mathcal{E} \leq 108.8) \approx 0.95$. Since $\theta^* = 437.2$ we say that with approximate confidence $1 - \alpha = 0.95$

$$-108.8 \leq \theta - \theta^* \leq 108.8, \quad \text{or} \quad \theta \in [437.2 - 108.8,\ 437.2 + 108.8]$$

Confidence interval can be seen as an **interval estimate** of a parameter, i.e. instead of one value we give a set of possible values.

In general for any ML-estimator, the approximate $1 - \alpha^*$ confidence interval is

$$\theta^* - \lambda_{\alpha/2} \cdot \sigma^*_{\mathcal{E}} \leq \theta \leq \theta^* + \lambda_{\alpha/2} \cdot \sigma^*_{\mathcal{E}}.$$

Since $\theta^* \approx \mathsf{E}[\Theta^*]$ while $\sigma_{\mathcal{E}}^* \approx \sqrt{\mathsf{V}[\Theta *]}$ one can also give the following alternative formulation

$$\mathsf{E}[\Theta^*] - \lambda_{\alpha/2} \cdot \sqrt{\mathsf{V}[\Theta^*]} \leq \theta \leq \mathsf{E}[\Theta^*] + \lambda_{\alpha/2} \cdot \sqrt{\mathsf{V}[\Theta^*]}, \tag{1}$$

with approximative confidence $1 - \alpha$. Here Θ^* is approx. $N(\theta^*, (\sigma_{\mathcal{E}}^2)^*)$.

Example: Suppose we are interested in probability that distance between earthquakes is longer than 1500 days, viz. $p = \mathsf{P}(T > 1500)$. An possible estimate is

$$p^* = \exp(-1500/\theta^*) = \exp(-1500/437.2) = 0.0324.$$

Confidence interval: Let write $P^* = \exp(-1500/\Theta^*)$ and employ (**??**):

$$\mathsf{E}[P^*] - \lambda_{\alpha/2} \cdot \sqrt{\mathsf{V}[P^*]} \leq \theta \leq \mathsf{E}[P^*] + \lambda_{\alpha/2} \cdot \sqrt{\mathsf{V}[P^*]},$$

then use **Gauss' formulae** to evaluate $\mathsf{E}[P^*]$, $\mathsf{V}[P^*]$, see blackboard.
This approach is called Delta-method.

More complex example:

Suppose that we are measuring the concentration of radon in buildings. At some location 40 houses were selected at random out of 200. Then average yearly concentration X were measured. The requirement is that the yearly mean concentration should be below 200 Bq/m^3. By plotting the 40 measurement on normal probability paper we conclude that the measured values are $N(m, \sigma^2)$. The $m^* = \bar{x} = 120$ while $(\sigma^2)^* = s_n^2 = 400$. One decided to compute the quantile $x_{0.001}$,

$$x^*_{0.001} = 120 + 3.09 \cdot \sqrt{400} = 181.8 < 200.$$

Hence the number of houses that can have concentration above 181 is $160 \cdot 0.001 = 0.16$ which is small. Find confidence interval for $x_{0.001}$ instead of $x^*_{0.001}$! Let $X^*_{0.001}$ be the estimator then employ (**??**):

$$\mathsf{E}[X^*_{0.001}] - \lambda_{\alpha/2} \cdot \sqrt{\mathsf{V}[X^*_{0.001}]} \leq \theta \leq \mathsf{E}[X^*_{0.001}] + \lambda_{\alpha/2} \cdot \sqrt{\mathsf{V}[X^*_{0.001}]}.$$

Let M^* and Σ^* be the estimators of mean m and σ^2, resp., then

$$\mathsf{V}[X^*_{0.001}] = \mathsf{V}[M^* + 3.09 \cdot \sqrt{\Sigma^*}].$$

Use Gauss formulas and $\big(\mathsf{V}[M^*], \mathsf{V}[\Sigma^*]\big) \approx \big(\frac{s_n^2}{n}, \frac{2(s_n^2)^2}{n}\big)$, $\mathsf{Cov}[M^*, \Sigma^*] = 0$.

Gauss' Approximations.

Let X and Y be independent random variables with expectations m_X, m_Y, respectively. For a smooth function h the following approximations

$$\begin{aligned} \mathsf{E}[h(X,Y)] &\approx h(m_X, m_Y), \\ \mathsf{V}[h(X,Y)] &\approx \big[h_1(m_X,m_Y)\big]^2\,\mathsf{V}[X] + \big[h_2(m_X,m_Y)\big]^2\,\mathsf{V}[Y] \\ &\qquad +2h_1(m_X,m_Y)\,h_2(m_X,m_Y)\,\mathsf{Cov}[X,Y]. \end{aligned}$$

where

$$h_1(x,y) = \frac{\partial}{\partial x}h(x,y), \qquad h_2(x,y) = \frac{\partial}{\partial y}h(x,y).$$

In our case $X = M^*$, $Y = \Sigma^*$ and $h(x,y) = x + 3.09\sqrt{y}$ hence

$$h_1(x,y) = 1, \qquad h_2(x,y) = 3.09 \cdot /(2\sqrt{y}).$$

Even more complex example:

Recall the data of 22 lifetimes (there were ($n = 22$) units tested). For ball bearings lifetime X the *rating life*, L_{10} should be estimated.[1] Assume the Weibull model is valid for the distribution of the lifetime:

$$F_X(x) = 1 - \mathrm{e}^{-(x/a)^c}, \quad x \geq 0,$$

then $L_{10} = a \cdot \left(-\ln(1 - \frac{1}{10})\right)^{1/c}$. For our data set $a^* = 82.08$ and $c^* = 2.06$ and hence

$$L_{10}^* = a^* \cdot \left(-\ln(1 - \frac{1}{10})\right)^{1/c^*} = 27.53 \cdot 10^6$$

What about the confidence interval? Again employ (**??**):

$$\mathsf{E}[L_{10}^*] - \lambda_{\alpha/2} \cdot \sqrt{\mathsf{V}[L_{10}^*]} \leq \theta \leq \mathsf{E}[L_{10}^*] + \lambda_{\alpha/2} \cdot \sqrt{\mathsf{V}[L_{10}^*]},$$

and use **Gauss' formulae** to evaluate $\mathsf{E}[L_{10}^*]$, $\mathsf{V}[L_{10}^*]$.

[1] L_{10} satisfies $\mathsf{P}(X \leq L_{10}) = 1/10$.

- $X = A^*$ and $Y = C^*$, $m_X = a^* = 82.08$, $m_Y = c^* = 2.06$
- $h(x,y) = x \cdot \left(-\ln(1-\frac{1}{10})\right)^{1/y}$.
- $\mathrm{V}[A^*] \approx 1.087\frac{(a^*/c^*)^2}{n}$, $\mathrm{V}[C^*] \approx 0.608\frac{(c^*)^2}{n}$, $\mathrm{Cov}[A^*, B^*] \approx 0.255\frac{a^*}{n}$.
- $h_1(x,y) = \left(-\ln(1-\frac{1}{10})\right)^{1/y}$,
  $h_2(x,y) = -\frac{x}{y^2} \cdot \left(\ln(-\ln(1-\frac{1}{10}))\right) \cdot \left(-\ln(1-\frac{1}{10})\right)^{1/y}$.

Now we can compute the variance

$$\begin{aligned} \mathrm{V}[L_{10}^*] &= \mathrm{V}[h(A^*, C^*)] \approx h_1(a^*, c^*)^2 \mathrm{V}[A^*] + h_2(a^*, c^*)^2 \mathrm{V}[C^*] \\ &\qquad + 2\, h_1(a^*, c^*) \cdot h_2(a^*, c^*)\, \mathrm{Cov}[A^*, C^*] = 18.5. \end{aligned}$$

Since $\sqrt{\mathrm{V}[L_{10}^*]} \approx 4.3$, then with approximate confidence 95% L_{10} is in

$[\,27.53 - 1.96 \cdot 4.3,\ 27.53 + 1.96 \cdot 4.3\,] = [\,19.1,\ 36.0\,]$ milions of revolutions.

Connection to hypothesis testing:

If one wishes to test whether a parameter θ has a specific value

$$H_0 : \theta = \theta_0$$

One chooses size of error α, i.e. probability of rejecting a true hypothesis is α. Then the test can be performed by constructing an interval that with confidence $1 - \alpha$ contains the true value of the parameter.

If θ_0 is not contained in the interval than one rejects the hypothesis H_0 that $\theta = \theta_0$.

Suppose that a dealer claims that $L_{10} = 40$ millions of resolutions. Since our confidence interval $[\,19.1,\ 36.0\,]$ does not contain value 40 thus, with "about" 5% probability of making error, we reject the hypothesis that the quality of the ball bearings is $L_{10} = 40$ millions of resolutions.

Examples of exact confidence intervals:

Suppose we have n observations $\bar{\mathbf{x}} = \sum x_i/n$ then:

- $1-\alpha$ confidence interval for m in $\mathrm{N}(m, \sigma^2)$ (σ *unknown*)

$$\left[\bar{\mathbf{x}} - t_{\alpha/2}(n-1)\frac{s_{n-1}}{\sqrt{n}},\ \bar{\mathbf{x}} + t_{\alpha/2}(n-1)\frac{s_{n-1}}{\sqrt{n}}\right]$$

  where $s_{n-1}^2 = \sum (x_i - \bar{\mathbf{x}})^2/(n-1)$.

- $1-\alpha$ confidence interval for m in $\mathrm{Po}(m)$

$$\theta \in \left[\frac{\chi^2_{1-\alpha/2}(2n\,\bar{\mathbf{x}})}{2n},\quad \frac{\chi^2_{\alpha/2}(2n\,\bar{\mathbf{x}}+2)}{2n}\right].$$

- $1-\alpha$ confidence interval for a in $\mathrm{Exp}(a)$

$$\theta \in \left[\frac{2n\,\bar{\mathbf{x}}}{\chi^2_{\alpha/2}(2n)},\quad \frac{2n\,\bar{\mathbf{x}}}{\chi^2_{1-\alpha/2}(2n)}\right].$$

Credibility intervals:

- In the Bayessian approach the lack of knowledge of parameter value θ is described using the probability densities $f(\theta)$ (odds). Random variable Θ having the pdf $f(\theta)$ models our knowledge about θ.
- The initial knowledge is described using $f^{\text{prior}}(\theta)$ density and as the data are gathered it is updated

$$f^{\text{post}}(\theta) = c\, L(\theta) f^{\text{prior}}(\theta).$$

- The pdf $f^{\text{post}}(\theta)$ summarizes our knowledge about θ. However if one value of for the parameter is needed then

$$\theta^{\text{predictive}} = \mathsf{E}[\Theta] = \int \theta f^{\text{post}}(\theta)\, d\theta.$$

- If one wishes to describe the variability of θ by means of an interval then the so called credibility interval can be computed

$$[\,\theta^{\text{post}}_{1-\alpha/2},\ \theta^{\text{post}}_{\alpha/2}\,]$$