stats6257conf 625.6 - Confidence intervals

Gunnar Stefansson

October 28, 2021

Contents

1 Confidence intervals

1.1 Interval Estimation

1.1.1 Handout

Recall from previous chapters: If $X_1, ..., X_n \sim n(\mu, \sigma^2)$ are i.i.d. random variables then

$$\bar{X} \sim n(\mu, \sigma^2/n) \quad \text{and} \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim n(0, 1)$$

A method for obtaining a level α confidence interval is by the so called method of inversion:

$$P\left[-z_{1-\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right] = P\left[\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

So the random set

$$C(\mathbf{X}) = \left[\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

has probability $1 - \alpha$ of covering μ.

Once we have actual data, outcomes of the random variables, we can compute a realisation of the set:

$$C(\mathbf{x}) = \left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

and we call this a $100(1 - \alpha)\%$ confidence interval for μ.

Note carefully: The random set has probability $1 - \alpha$ of covering μ. Once we have data $\mathbf{x}$ we have a fixed set $C(\mathbf{x})$ and there is no probability any more. We simply **claim** $\mu \in C(\mathbf{x})$.

Recall that we tested $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ under the usual Gaussian assumptions, using $Z := \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ and rejected if the numerical outcome satisfies

$$|z| > z_{1-\alpha/2}.$$

Note that we use the natural definition of quantiles and subscript therefore denote the lower-tail probability, i.e.

$$\Phi(z_\alpha) = \alpha$$

so that

$$\Phi\left(z_{1-\alpha/2}\right) = 1 - \alpha/2.$$

<u>**Note 1:**</u> Suppose $X_1, ..., X_n \sim f_\theta, \theta \in \Theta$ and ϕ_{θ_0} is a test function for $H_0 : \theta = \theta_0$. Then we can think of the entire collection of such tests $\{\phi_{\theta_0}\}_{\theta_0 \in \Theta}$ or simply $\{\phi_\theta\}_{\theta \in \Theta}$.

Principle of generating confidence sets from test functions: Suppose $\{\phi_\theta\}_{\theta \in \Theta}$ is a collection of tests for the situation where $\mathbf{X} \sim f_\theta, \theta \in \Theta$ and define $C(\mathbf{X}) := \{\theta : \phi_\theta(\mathbf{x}) = 0\}$.

Theorem 1.1 If the tests $\{\phi_\theta\}_{\theta \in \Theta}$ are all level α tests, then the set $C(\mathbf{X})$ has coverage probability at least $1 - \alpha$.

Proof.

$$P_\theta(\theta \in C(\mathbf{X})) = P_\theta(\phi_\theta(\mathbf{X}) = 0) = 1 - P_\theta(\phi_\theta(\mathbf{X}) = 1) \geq 1 - \alpha$$

Consider the simplest cases where $C(\mathbf{X}) \subseteq \mathbb{R}$.

Definition 1 An interval $[L(\mathbf{x}), U(\mathbf{x})]$ is an interval estimate and $[L(\mathbf{X}), U(\mathbf{X})]$ is an interval estimator if $X_1, ..., X_n$ are random variables and $L, U: \mathbb{R}^n \to \mathbb{R}$ with $L \leq U$.
Note: $\mathbb{R}' = \mathbb{R} \cup \{-\infty, \infty\}$ is permitted.

Definition 2 Let $\mathbf{X} \sim f_\theta, \theta \in \Theta$. Then

(a) $\inf_{\theta\in\Theta} P_\theta(\theta \in C(\mathbf{X})) =:$ coverage probability $=: 1-\alpha$

(b) $100(1-\alpha)\%$ is the confidence of the set, i.e. $C(\mathbf{X})$ is a $100(1-\alpha)$ confidence set if α is as above.

1.2 Location, scale and location-scale families

1.2.1 Handout

Consider a location family with $f_\mu(x) = f(x-\mu)$.

Let $X \sim f_\mu$ and write $Q(X, \mu) = X - \mu$. Then we obtain

$$\begin{aligned} F_Q(t) :&= P_\mu[Q(X,\mu) \leq t] \\ &= P_\mu[X - \mu \leq t] = P_\mu[X \leq t + \mu] \\ &= \int_{-\infty}^{t+\mu} f_\mu(x)dx = \int_{-\infty}^{t+\mu} f(x-\mu)dx \end{aligned}$$

so the density is $\frac{d}{dt}F_Q(t) = f(t)$.

Notice that Q, or $Q(X,\mu)$, is a function of both the sample and the unknown parameters. This quantity is therefore neither an ordinary random variable not a parameter.

Notice also that this random quantity has a fixed distribution, which no longer depends on the parameter itself.

Random quantities with a distribution which is free of the parameter are called **pivotal quantities**.

In the same way we see that if $X \sim f_\sigma(x) = f(\frac{x}{\sigma})$ then

$$\frac{X}{\sigma} \sim f(x)$$

and if $X \sim f_{\mu,\sigma}(x) = f(\frac{x-\mu}{\sigma})$ then

$$\frac{X-\mu}{\sigma} \sim f(x)$$

.

Many statistics such as $T = \bar{X}$, $X_{(n)}$, $X_{(1)}$, $\widetilde{X} :=$ median$\{X_1, \ldots, X_n\}$ are linear, i.e.

$$T\left(\frac{X_1-\mu}{\sigma}, \ldots, \frac{X_n-\mu}{\sigma}\right) = \frac{T(\mathbf{X})-\mu}{\sigma}$$

or have scaling property: $R(\frac{\mathbf{X}}{\sigma}) = \frac{1}{\sigma}R(\mathbf{X})$.

Example 1 Let $X_1, \ldots, X_n \sim f_{\mu,\sigma}$ iid, where $f_{\mu,\sigma}(x) = f(\frac{x-\mu}{\sigma})$. Suppose f is a known density but μ, σ unknown. We know that $\frac{X_i - \mu}{\sigma} \sim f$ if $X_i \sim f_{\mu,\sigma}$ and therefore

1. $\frac{\overline{X}-\mu}{\sigma} = \frac{1}{n}\sum \frac{X_i-\mu}{\sigma}$,
2. $\frac{X_{(n)}-\mu}{\sigma}$,
3. $\frac{\widetilde{X}-\mu}{\sigma}$,
4. $\frac{X_{(1)}-\mu}{\sigma}$,
5. $\frac{1}{\sigma}S = \frac{1}{\sigma}\sqrt{\frac{X_i-\overline{X}}{n-1}}$,
6. $\frac{\overline{X}-\widetilde{X}}{S}$

are pivotal quantities.

Note that we want to use a probability statement of the form $P_\theta[a \leq Q(\mathbf{X}, \theta) \leq b] = 1-\alpha$, $\forall\theta$, and "pivot" this to generate an equivalent statement $P_\theta[\theta \in \zeta(\mathbf{X})] = 1-\alpha$, $\forall\theta$.

Example 2 In a location-scale family, one can e.g. use $\frac{S}{\sigma}$ to make inference on σ, even if is unknown; use $\frac{\overline{X}-\mu}{\sigma}$ for μ, if σ is known, $\frac{\overline{X}-\mu}{S}$ for μ even if σ is unknown, etc.

But $\frac{\overline{X}-\widetilde{X}}{S}$ does not involve the parameters and has a distribution free of the parameters, so it provides us information. It is an **ancillary** statistic and is useless here. Alternatives to S in this context include the range $X_{(n)} - X_{(1)}$ and MAD $= \text{median}\{|X_1 - \widetilde{X}|, |X_2 - \widetilde{X}|, \ldots, |X_n - \widetilde{X}|\} =$ median absolute deviation.

1.3 Seeking shorter confidence intervals

sometimes want to optimise the length of the CI
(add text...)
We now want to evaluate

$$(*) \int_a^b f_Y(t)dt = 1-\alpha$$

and find conditions which give a short confidence interval.

$$(*)B$$

$$\int_a^b \frac{t^{r-1}e^{-nt}}{\Gamma r(1/n)^r}dt = 1-\alpha$$

Could choose cutoffs $\alpha/2$, i.e.

$$\int_0^{\alpha/2} \frac{t^{r-1}e^{-nt}}{\Gamma(r)(1/n)^r}dt = \frac{\alpha}{2}$$

This is what is usually done. It is optimal for the normal distribution, but not for other, asymmetric distributions.

1.3.1 Examples

Example: Consider $X_1, ..., X_n \sim n(\mu, \sigma^2)$ and we want to find a confidence interval for σ^2. We know that (X, S^2) is sufficient for (μ, σ^2) and for σ^2 its natural to consider the pivotal quantity

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

so we can easily find a and b such that

$$\mathbf{P}(a \le \frac{(n-1)S^2}{\sigma^2} \le b) \ge 1 - \alpha$$

e.g. choose $a = \chi^2_{n-1,\alpha/2}$ and $b = \chi^2_{n-1,1-\alpha/2}$ to obtain the usual $100(1-\alpha)\%$ confidence interval.

$$\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}$$

which turns out to be not of the shortest length.

Next consider a gamma density:

We want to find a and b such that

$$P[a \le \frac{\bar{X}}{\beta} \le b] = 1 - \alpha$$

i.e.

$$P[\frac{\bar{X}}{b} \le \beta \le \frac{\bar{X}}{a}] = 1 - \alpha$$

We would generally prefer a short interval.

We know that

$$P[a \le \frac{\bar{X}}{\beta} \le b] = \int_a^b f_{\bar{X}/\beta}(t)dt$$

and $\frac{X_i}{\beta}$ has the density

$$\frac{x^{r-1}e^{-x}}{\Gamma(r)} \text{ i.e. } \Gamma(r, 1)$$

So if $Y_i := \frac{X_i}{\beta} \sim \Gamma(r, 1)$ then we need the density of $\bar{Y} = \frac{1}{n}\sum Y_i = \frac{\bar{X}}{\beta}$
The moment generating function for $\Gamma(r, \beta)$ is

$$E[e^{tX}] = \int_0^\infty e^{tx} \frac{x^{r-1}e^{\frac{-x}{\beta}}}{\Gamma(r)\beta^r} dx$$

$$= \int_0^\infty \frac{x^{r-1}}{\Gamma(r)\beta^r} e^{\frac{x}{\frac{1}{t-\frac{1}{\beta}}}} dx$$

$$\frac{\frac{1}{(t-1)/\beta}^r}{\beta^r} 1 = \frac{1}{\beta^r(\frac{1}{\beta} - t)^r} = \frac{1}{(1-\beta t)^r}$$

So

$$M_{Y_i}(t) = \frac{1}{(1-t)^r}$$

and

$$M_{\bar{Y}}(t) = E[e^{t\frac{1}{n}\sum Y_i}] = M_{Y_i}(\frac{t}{n})^n = \frac{1}{(1-\frac{t}{n})^{rn}}$$

$$M_{\sum Y_i}(t) = \frac{1}{(1-t)^{rn}}$$

We now want to evaluate $\int_a^b f_{\bar{Y}}(t)dt = 1 - \alpha$ and find conditions which give a short confidence interval.

1.4 A few more examples and background

1.4.1 Handout

20131101_092159.jpg :
Example 1
Let $X_1, ..., X_n \sim n(\mu, \sigma^2)$, i.i.d., σ^2 known. Define:

(a) $C(\mathbf{x}) := (-\infty,\ \bar{x} + z_{1-\alpha}\dfrac{\sigma}{\sqrt{n}}]$

(b) $C(\mathbf{x}) := [\bar{x} - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, \infty)$

(c) $C(\mathbf{x}) := [\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}]$

are all $100(1-\alpha)\%$ confidence intervals.

20131101_100326.jpg :
Note that in a location family with $f_\mu(x) = f(x-\mu)$ we know that if $X \sim f_\mu$, then if we write $Q(X, \mu) = X - \mu$ to obtain

$$\begin{aligned} F_Q(t) &:= P_\mu[Q(X,\mu) \leq t] \\ &= P_\mu[X - \mu \leq t] \\ &= P_\mu[X \leq t + \mu] \\ &= \int_{-\infty}^{t+\mu} f_\mu(x)dx \\ &= \int_{-\infty}^{t+\mu} f(x-\mu)dx \end{aligned}$$

so the density is $\frac{d}{dt}F_Q(t) = f(t)$ and in the same way we see that if $X \sim f_\sigma(x) = f(x/\sigma)$ then $\dfrac{X}{\sigma} \sim f(x)$ and if $X \sim f_{\mu,\sigma}(x) = f\left(\dfrac{x-\mu}{\sigma}\right)$ then $\dfrac{X-\mu}{\sigma} \sim f(x)$.

Many statistics, such as: $\bar{X}$, $X_{(n)}$, $X_{(1)}$, $\tilde{X} := \text{median}\{X_1, ..., X_n\}$ are linear, i.e.

$$T\left(\frac{X_1-\mu}{\sigma}, ..., \frac{X_n-\mu}{\sigma}\right) = \frac{T(\mathbf{X})-\mu}{\sigma}$$

or e.g. $R(\mathbf{X}) := S$.

2 Inverting test statistics

2.1 Some examples and preliminaries

2.1.1 Handout

Definition 3 Confidence sets-overview $\mathcal{C}(\underline{x})$ is a $100(1-\alpha)\%$ is Confidence set if $P_\theta[\theta \in \mathcal{C}(\underline{x})] \geq 1-\alpha \forall \theta \in \Theta$

Definition 4 Inverting tests:
If $\phi_\theta : \mathbb{R}^n \to \{0,1\}$ is such that $P_\theta[\phi_\theta(\underline{X}) = 1] = \alpha$ then $\mathcal{C} := \{\theta : \phi_\theta(\underline{x}) = 0\}$ is a $100(1-\alpha)\%$ C-set for θ

Example 3 (Inverting a normal test): Let $X_1, \ldots, X_n$ be i.i.d. $n(\mu, \sigma^2)$ and consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. For a fixed α level, a reasonable test has rejection region $\{\mathbf{x} : |\bar{x} - \mu_0| > z_{\alpha/2}\sigma/\sqrt{n}\}$. Note that H_0 is accepted for sample points with $|\bar{x} - \mu_0| \leq z_{\alpha/2}\sigma\sqrt{n}$ or, equivalently

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Since the test has size α, this means that $P(H_0 \text{ is rejected } | \mu = \mu_0) = \alpha$ or, stated in another way $P(H_0 \text{ is accepted } | \mu = \mu_0) = 1 - \alpha$. Combining this with the above characterization of the acceptance region, we can write

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \middle| \mu = \mu_0\right) = 1 - \alpha.$$

But this probability statement is true for *every* μ_0. Hence, the statement

$$P_\mu\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \middle| \mu = \mu_0\right) = 1 - \alpha.$$

is true. The interval $[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$, obtained by *inverting* the acceptance region of the level α test, is a $1-\alpha$ confidence interval.

3 Pivotal quantities

3.1 Definitions and examples of the use of pivotal quantities

3.1.1 Handout

Definition 5 Pivotal quantities:
$Q : \mathbb{R}^n \times \Theta \to \mathbb{R}$ is a pivotal quantity if

$$P_\theta[Q((\underline{}X), \theta) \in A]$$

is constant in θ.

If the set $A \subset \mathbb{R}$ is chosen such that

$$P_\theta[Q((\underline{}X), \theta) \in A] = 1 - \alpha$$

then

$$\{\theta : Q(\underline{x}, \theta) \in A\}$$

is a $100(1-\alpha)\%$ C-set for θ

Example 4 If $X_1, \cdots, X_n \sim F_\theta$ where $F_\theta(x) = F(x-\theta)$ and $\phi_\theta(\underline{x} = \phi(\underline{x} - \theta)$ where ϕ is level-α test of $H_0 : \theta = 0$ then

$$\mathcal{C}(\underline{x}) = \{\theta : \phi_\theta(\underline{x}) = 0\}$$

$$= \{\theta : \phi(\underline{x} - \theta) = 0\}$$
$$= \{\theta : \underline{x} - \theta \in \phi^{-1}(\{0\})\}$$
$$= \{\theta : \theta \in \underline{x} - \phi^{-1}(\{0\})\}$$
$$= \underline{x} - \underbrace{\phi^{-1}(\{0\})}_{\text{acceptance region for } \phi}$$

Example 5 Let $X_1, ..., X_n \sim U(0, \theta)$. By scaling by $1/\theta$, we get
$X_1/\theta, ..., X_n/\theta \sim U(0,1)$.
We know that the n-th order statistic $X_{(n)}$ is sufficient for θ. The distribution of T:= $\frac{X_{(n)}}{\theta}$ can be found by noting that

$$P[T < t] = \int_0^t 1dt = t$$

so for our sample, we have

$$P[X_1 < t, ..., X_n < t] \prod_{i=1}^{n} T = t^n$$

Differentiating w.r.t t, we get

$$T \sim nt^{n-1}$$

The distribution of T is independent of θ and thus $Q(X,\theta) = T$ is a pivotal quantity. A confidence interval for θ can be found by using

$$P(a < \frac{X_{(n)}}{\theta} < b) = \int_a^b nt^{n-1}dt = 1-\alpha$$

Example 6 If $X_1, \ldots, X_n \sim \textit{exponential}(\lambda)$ we can construct a 95% confidence interval for the parameter using a pivotal quantity and using the method of inverting the acceptance region of a test. We start by defining $T := \sum X_i$ and $Q(T,\lambda) := \frac{2T}{\lambda} \sim \chi^2_{2n}$.

Now write X^2_{2n} for a generic random variable with a χ^2_{2n} distribution, so that X^2_{2n} has the same distribution as $Q(T,\lambda)$.

Since X^2_{2n} for has a fixed and known distribution, we can choose constants a and b to satisfy

$$P(a \leq X^2_{2n} \leq b = 1-\alpha.$$

We then obtain

$$P_\lambda(a \leq \frac{2T}{\lambda} \leq b) = P_\lambda(a \leq Q(T,\lambda) \leq b) = P(a \leq X^2_{2n} \leq b) = 1-\alpha$$

Inverting the set $A(\lambda) = \{t : a \leq \frac{2t}{\lambda} \leq b)\}$ gives $C(t) = \{\lambda : \frac{2t}{b} \leq \lambda \leq \frac{2t}{a})\}$ which is a $1-\alpha$ confidence interval.

For example if $n = 10$, then consulting a table of χ^2 cutoffs shows that a 95% confidence interval is given by $\{\lambda : \frac{2T}{34.17} \leq \lambda \leq \frac{2T}{9.59})\}$