Stat 2911 Lecture Notes

Class 34, 2017

Uri Keich

© Uri Keich, The University of Sydney

Confidence Intervals: motivation, definition and derivation (Rice 7.3 is relevant)

Norbert Quirk Mathematical Essay Competition

Confidence Intervals

Estimation: given a sample $x_1, \ldots, x_n \sim F_\theta$ find θ.

For example, $X_i \sim N(\theta, 1) \Rightarrow \hat{\Theta} = \bar{X}$ (MLE, moments)

How close are we to θ?

$$\sum_{i=1}^{n} X_i \sim N(\quad , \quad) \quad \text{(why?)}$$

$$\Rightarrow \quad \hat{\Theta} = \bar{X} \sim N(\quad , \quad),$$

and in particular,

$$P_\theta(\hat{\Theta} = \theta) = \qquad .$$

One way to measure how well the estimator is doing is the $MSE(\hat{\Theta}) = E[(\hat{\Theta} - \theta]^2)$.

However, as we saw the MSE often depends on θ, and regardless, it's not clear how to interpret it beyond comparing estimators.

We cannot expect to know how well the estimator $\hat{\Theta}$ is doing on any given sample (as this would imply we know θ).

It would however be useful to give a bound on the probability that $\hat{\Theta}$ deviates significantly from θ for a randomly drawn sample.

Example. $P_\theta(|\hat{\Theta}-\theta| \geq \varepsilon) = P_\theta(|\hat{\Theta}-\theta|^2 \geq \varepsilon^2)$

$$\overset{?}{\leq} \frac{E_\theta(\hat{\Theta}-\theta)^2}{\varepsilon^2}$$

$$= \frac{MSE_\theta(\hat{\Theta})}{\varepsilon^2}.$$

Notice the tradeoff:
proximity to θ (ε) $\longleftrightarrow$ confidence (upper bound on the prob. of deviation)

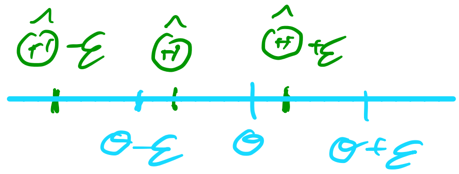
If our sample is again from the $N(\theta,1)$ dist. then

$$P_\theta(|\hat{\Theta}-\theta| \geq \varepsilon) \leq \frac{V_\theta(\hat{\Theta})}{\varepsilon^2}$$

$$= \frac{1}{n\varepsilon^2}.$$

Equivalently,

$$P_\theta(|\hat{\Theta}-\theta| < \varepsilon) \geq 1 - \frac{1}{n\varepsilon^2}.$$



$$\Rightarrow \quad P_\theta(\theta \in (\hat{\Theta}-\varepsilon, \hat{\Theta}+\varepsilon)) \geq 1 - \frac{1}{n\varepsilon^2}.$$

That is, the random interval $(\bar{X}-\varepsilon, \bar{X}+\varepsilon)$ contains the unknown θ with prob. $\geq 1 - \frac{1}{n\varepsilon^2}$.

We say that $(\bar{X}-\varepsilon, \bar{X}+\varepsilon)$ is a $100\times(1-1/n\varepsilon^2)\%$ confidence interval for θ.

© Uri Keich, The University of Sydney

Warning. For any specific sample $x_1, \ldots, x_n$ $\theta \in (\bar{x}-\varepsilon, \bar{x}+\varepsilon)$ is **not** an event. Rather, it is a deterministic statement that's either true or false.

So $P(\theta \in (\bar{X}-\varepsilon, \bar{X}+\varepsilon)) = \ldots$ is meaningless. Instead, we say $(\bar{x}-\varepsilon, \bar{x}+\varepsilon)$ is a confidence interval (CI) for θ.

How can we improve the CI in our example?

We can evaluate $P(\theta \in (\hat{\Theta}-\varepsilon, \hat{\Theta}+\varepsilon))$ more accurately;

Markov's Inequality is rather crude.

$$\hat{\Theta} = \bar{X} \sim N(\theta, 1/n) \Rightarrow \sqrt{n}(\bar{X}-\theta) = \frac{\bar{X}-\theta}{1/\sqrt{n}} \sim N(\ , \ )$$

Therefore,

$$P(\theta \in (\bar{X}-\varepsilon, \bar{X}+\varepsilon)) = P(-\varepsilon < \bar{X}-\theta < \varepsilon)$$

$$= P(-\sqrt{n}\cdot\varepsilon < \sqrt{n}(\bar{X}-\theta) < \sqrt{n}\varepsilon)$$

("=" not "$\geq$"!) $\quad \overset{?}{=} \Phi(\sqrt{n}\varepsilon) - \Phi(-\sqrt{n}\varepsilon)$,

where Φ is the $N(0,1)$ CDF.

For example, with $\varepsilon = 3$ we have

n	$\Phi(\sqrt{n}\cdot 3) - \Phi(-\sqrt{n}\cdot 3)$	$1 - 1/9n$
1	0.997	0.889
3	0.99978	0.944
10	$1 - 2\times 10^{-21}$	0.989

Our level of confidence increases with n since our CI is of a fixed length $(2\varepsilon = 6)$.

Better: trade some of that ridiculously high conf. for a tighter interval, or keep the conf. level fixed and adjust the interval's length.

<u>Def.</u> An interval valued function

$$I_\alpha^{(n)} : \mathbb{R}^n \longrightarrow \{(a,b) \in \mathbb{R}^2 : a < b\}$$

e.g. $I_\alpha^n(\underline{x}) = (\bar{x} - \alpha, \bar{x} + \alpha)$, where $\underline{x} = (x_1, \ldots, x_n)$

defines a $100(1-\alpha)\%$ confidence interval for θ if for $\forall \theta$ and $X_1, \ldots, X_n$ ind. F_θ-distributed RVs

$$P_\theta\left(\theta \in I_\alpha^{(n)}(X_1, \ldots, X_n)\right) \geq 1 - \alpha.$$

In this case, given a sample $x_1, \ldots, x_n$, $I_\alpha(x_1, \ldots, x_n)$ is a $100(1-\alpha)\%$ CI for θ.

It is typically not difficult to verify a given interval function defines a CI, but how do you find one to begin with?

© Uri Keich, The University of Sydney

Let $\hat{\Theta}$ be an estimator of θ and let $\hat{\theta} = \hat{\theta}(\underline{x})$ be the estimate on a particular sample $\underline{x} = (x_1, \dots, x_n)$.
For that estimate $\hat{\theta}$ define the function

$$h(\theta; \hat{\theta}) = P_\theta(\hat{\Theta} \geq \hat{\theta}). \qquad (\hat{\theta} \text{ is fixed})$$

For example, for the $N(\theta,1)$ problem

$$h(\theta; \hat{\theta} = \bar{x}) = P_\theta(\bar{X} \geq \bar{x}).$$

Suppose that for any value of $\hat{\theta}$, h is cont. and strictly $\uparrow$ in θ (so h^{-1} wrt θ exists for any $\hat{\theta}$).

In our example, $\hat{\Theta} = \bar{X} \sim N(\theta, 1/n)$ so

$$h(\theta; \bar{x}) = P_\theta(\bar{X} \geq \bar{x})$$
$$= P(\underbrace{\sqrt{n}(\bar{X} - \theta)}_{\text{dist. of} \downarrow \text{ is invariant of } \theta!} \geq \sqrt{n}(\bar{x} - \theta))$$
$$= 1 - \Phi(\sqrt{n}(\bar{x} - \theta)).$$

$$\Rightarrow \frac{d}{d\theta} h(\theta; \bar{x}) = \varphi(\sqrt{n}(\bar{x} - \theta)) \cdot \sqrt{n} > 0 \qquad \left(\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\right)$$
$$\Rightarrow h \uparrow$$

Define $\theta_L = \theta_L(\underline{x}) = h^{-1}(\alpha/2; \hat{\theta}(\underline{x}))$, i.e.

$$h(\theta_L; \hat{\theta}(\underline{x})) = P_{\theta_L}(\hat{\Theta} \geq \hat{\theta}(\underline{x})) = \alpha/2.$$

Ex. $\theta_L = \min\{\nu : P_\nu(\hat{\Theta} \geq \hat{\theta}(\underline{x})) \geq \alpha/2\}$

© Uri Keich, The University of Sydney

In our example with $\hat{\Theta} = \bar{X} \sim N(\theta, 1/n)$

$$h(\theta; \bar{x}) = 1 - \Phi(\sqrt{n}(\bar{x} - \theta))$$

$$\Rightarrow \alpha/2 = h(\theta_L; \bar{x}) = 1 - \Phi(\sqrt{n}(\bar{x} - \theta_L))$$

$$\Rightarrow \Phi(\sqrt{n}(\bar{x} - \theta_L)) = 1 - \alpha/2$$

$$\Rightarrow \sqrt{n}(\bar{x} - \theta_L) = \Phi^{-1}(1 - \alpha/2) =: z_{1-\alpha/2}$$

$$\Rightarrow \theta_L = \bar{x} - \frac{z_{1-\alpha/2}}{\sqrt{n}} .$$

Similarly, define

$$g(\theta; \hat{\theta}) = P_\theta(\hat{\Theta} \le \hat{\theta}) .$$ ($\hat{\theta}$ is fixed)

For example, for the $N(\theta, 1)$ problem

$$g(\theta; \hat{\theta} = \bar{x}) = P_\theta(\bar{X} \le \bar{x}) .$$

Suppose that for any value of $\hat{\theta}$, g is cont. and strictly $\downarrow$ in θ (so g^{-1} wrt θ exists for any $\hat{\theta}$).

In our example, $\hat{\Theta} = \bar{X} \sim N(\theta, 1/n)$ so

$$g(\theta; \bar{x}) = P(\underbrace{\sqrt{n}(\bar{X} - \theta)}_{\text{dist. of } \downarrow \text{ is invariant of } \theta!} \le \sqrt{n}(\bar{x} - \theta))$$

$$= \Phi(\sqrt{n}(\bar{x} - \theta)) ,$$

$$\Rightarrow \frac{d}{d\theta} g(\theta; \bar{x}) = -\varphi(\sqrt{n}(\bar{x}-\theta)) \cdot \sqrt{n} < 0 \quad \Rightarrow g \downarrow$$

Define $\theta_R = \theta_R(\underline{x}) = g^{-1}(\alpha/2; \hat{\theta}(\underline{x}))$, i.e.,

$$g(\theta_R; \hat{\theta}(\underline{x})) = P_{\theta_R}(\hat{\Theta} \le \hat{\theta}(\underline{x})) = \alpha/2.$$

Ex. $\theta_R = \max\{\nu : P_\nu(\hat{\Theta} \le \hat{\theta}(\underline{x})) \geq \alpha/2\}.$

In our example with $\hat{\Theta} = \bar{X} \sim N(\theta, 1/n)$

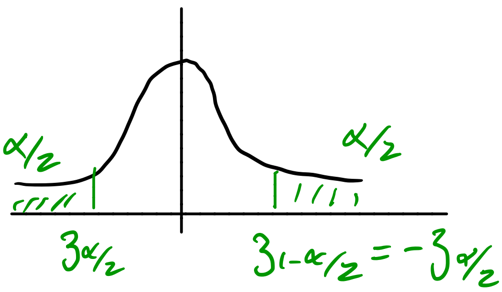
$$g(\theta; \bar{x}) = \Phi(\sqrt{n}(\bar{x}-\theta))$$

$$\Rightarrow \alpha/2 = g(\theta_R; \bar{x}) = \Phi(\sqrt{n}(\bar{x}-\theta_R))$$

$$\Rightarrow \sqrt{n}(\bar{x}-\theta_R) = \Phi^{-1}(\alpha/2) =: z_{\alpha/2}$$

$$\Rightarrow \theta_R = \bar{x} - \frac{z_{\alpha/2}}{\sqrt{n}}$$

$$\overset{\text{sym}}{=} \bar{x} + \frac{z_{1-\alpha/2}}{\sqrt{n}}$$



<u>Claim.</u> (θ_L, θ_R) is a $100(1-\alpha)\%$ CI for θ.

$\Rightarrow$ Given a sample $x_1, \ldots, x_n$ from the $N(\theta, 1)$ dist.

$\left(\bar{x} - \frac{z_{1-\alpha/2}}{\sqrt{n}}, \bar{x} + \frac{z_{1-\alpha/2}}{\sqrt{n}}\right)$ is a $100(1-\alpha)\%$ CI for θ.

For a fixed α, the length of this CI is $2\cdot\frac{z_{1-\alpha/2}}{\sqrt{n}}$, which shrinks like $1/\sqrt{n}$:

n	1	10	100	1000
$z_{1-\alpha/2}$	1.96	0.62	0.196	0.062

$(\alpha = 0.05)$

Conversely, for a fixed n, as we increase $1-\alpha \uparrow 1$ (so $\alpha \downarrow 0$), $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2) \uparrow \infty$.

Again: conf. vs. precision tradeoff.

© Uri Keich, The University of Sydney