Statistics

Luc Veldhuis

November 2016

Recap

Aim: Quantify the uncertainty in a point estimate.
Means: Confidence interval/region
Handy: Pivot
Often unavailable $\Rightarrow$ Approximate pivot.
Maximum Likelihood Estimator as approximate pivots. To this end define:

Definition

Score function $\theta \to l_{\dot\theta}(X_1) = \frac{d}{d\theta} log(p_\theta(X_1)$ and Fisher function $i_\theta = \mathbb{V}(l_{\dot\theta}(X_1)) = -\mathbb{E}(\frac{d^2}{d\theta^2} log(p_\theta(X_1)))$

Recap

Theorem

Let $X_1, \ldots, X_2 \sim p_\theta$ and $\hat{\theta}_n$ is the ML estimator on the basis of this sample. Then under some conditions:
$\sqrt{n}(\hat{\theta}_n - \theta) \to N(0, i_\theta^- 1)$ if $n \to \infty$.
Define the approximate pivot $\sqrt{n i_\theta}(\hat{\theta}_n - \theta) \sim N(0, 1)$
Now an approximate $(1-\alpha)$-confindence interval for θ is:
$\theta = \hat{\theta}_n \pm (n i_\theta)^{\frac{1}{2}} \xi_{1-\frac{\alpha}{2}}$

Note

i_θ is unknown and needs to be estimated. Two approaches:

- Plug in estimator: $\hat{i}_\theta = i_{\hat{\theta}}$ (substitute ML estimator of θ in i_θ)
- Observed information: $\hat{i}_\theta = -\frac{1}{n} \sum_{i=1}^{n} \frac{d^2}{d\theta^2} log(p_\theta(X_1)|_{\theta=\hat{\theta}_n}$

Wald interval

Definition

The Wald-interval is $\theta = \hat{\theta}_n \pm \frac{1}{\sqrt{n i_{\hat{\theta}}}} \xi_{1-\frac{alpha}{2}}$

Example

Derivation of the Wald interval for θ if $X_1, \ldots, X_n \sim Geom(\theta)$

ML-estimator: $L(\theta; X_1, \ldots, X_n) = \prod_{i=1}^{n} (1-\theta)^{x_i - 1}\theta$

$log(L) = \sum_{i=1}^{n} ((x_i - 1)\frac{1}{1-\theta} - \frac{1}{\theta}) = 0$

$\frac{d^2}{d\theta^2} log(L) = \frac{-1}{(\theta-1)^2} \sum_{i=1}^{n} (x_i - 1) - \frac{n}{\theta^2}$

Solve for θ: $(\sum x_i - n)\theta = n - n\theta \Rightarrow \bar{\theta_{ML}} = \frac{1}{\bar{x}}$

Fisher information: $i_\theta = \frac{1}{\theta^2(1-\theta)}$ using $\mathbb{E}(X_i) = \frac{1}{\theta}$

Fisher information

Definition

- $\hat{i_\theta}$ with plug-in: $i_\theta = \frac{1}{\theta^2(1-\theta)} = \frac{\bar{X}^3}{\bar{X}-1}$
- i_θ with observed information: $l_{\ddot{\theta}}(X_i) = \frac{X_i-1}{(1-\theta)^2} - \frac{1}{\theta^2}$

$$\hat{i_\theta} = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-1}{(1-\theta)^2} - \frac{1}{\theta^2}\right)\Big|_{\theta=\hat{\theta}_n}$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-1}{(1-\frac{1}{\hat{x}})^2} - \frac{1}{\frac{1}{\hat{x}}}\right) = \cdots =$$

$$\frac{\bar{X}^3}{\bar{X}-1}$$

Then, $(1-\alpha)$-confidence interval (approximate) is:

$\theta = \hat{\theta_{ml}} \pm \frac{\sqrt{\bar{X}-1}}{\sqrt{n\bar{X}}}\xi_{1-\frac{\alpha}{2}}$ In general, plug-in $\neq$ observed!

Confidence intervals

Theorem

Let for every $\theta_0 \in \Theta$ there be a test for $H_0 : \theta = \theta_0$ of significance level α. Then, the set of all values θ_0 for which H_0 is not rejected (on the basis of the data available) forms a $(1-\alpha)$-confidence interval for θ. Reverse: if G_x is an $(1-\alpha)$-confidence interval for θ, then the test that rejects $H_0 : \theta = \theta_0$ for $\theta_0 \notin G_x$ is of significance level α.

Proof

"$\Rightarrow$": Given a test with critical K_{θ_0} for $H_0 : \theta = \theta_0$, significance level α means $P_{\theta_0}(X \in K_{\theta_0}) \leq \alpha$ for all $\theta_0 \in \Theta$. Or $P_{\theta_0}(X \notin K_{\theta_0}) \geq 1-\alpha \ \forall \theta_0 \in \Theta$

Confidence intervals & tests

Definition

$G_x = \{\theta_0 : X \notin K_{\theta_0}\}$ Prove this is a confidence interval.
For G_x it holds:
$P_\theta(\theta \in G_x) = P_\theta(\theta \in \{\theta_0 : X \notin K_{\theta_0}\}) = P_\theta(X \notin K_\theta) \geq 1 - \alpha$ for all $\theta \in \Theta$.
This is indeed a $(1 - \alpha)$-confidence interval for θ.

Example

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, μ, σ^2 unknown. For $H_0 : \mu = \mu_0$ take the t-test with test statistic: $T = \sqrt{n}\frac{\bar{X} - \mu_0}{S_x}$.
If $|T| \leq t_{n-1;1-\frac{\alpha}{2}}$, then H_0 is not rejected. The $(1 - \alpha)$-confidence interval is therefore:
$\{\mu_0 : \sqrt{n}\frac{|\bar{X} - \mu_0|}{S_x} \leq t_{n-1;1-\frac{\alpha}{2}}\} = \{\mu_0 : t_{n-1;\frac{\alpha}{2}} \leq \sqrt{n}\frac{\bar{X} - \mu_0}{S_x} \leq t_{n-1;1-\frac{\alpha}{2}}\}$
$= [\bar{X} - \frac{S_x}{\sqrt{n}} t_{n-1;1-\frac{\alpha}{2}}, \bar{X} + \frac{S_x}{\sqrt{n}} t_{n-1;1-\frac{\alpha}{2}}]$ (as before).

Confidence intervals

For the $N(\mu, \sigma^2)$ distribution we now have 3 equivalent statements:

- $H_0 : \mu = \mu_0$ is not rejected at level α_0
- μ_0 lies in the $(1 - \alpha)$ confidence interval of μ
- p-values p_l, p_r are both larger then $\frac{\alpha}{2}$

In this the uncertainty α_0 is either:

- $P_{H_0}(|T| > t_{n-1;1-\frac{\alpha}{2}})$
- The probability that the true μ_0 does not lie within the confidence interval

This is due to the equivalent event:

$|T| \geq t_{n-1;1-\frac{\alpha}{2}} \Leftrightarrow \frac{\sqrt{n}}{S_x}|\bar{X} - \mu| > t_{n-1;1-\frac{\alpha}{2}}$
$\Leftrightarrow |\bar{X} - \mu| > \frac{S_x}{\sqrt{n}} t_{n-1;1-\frac{\alpha}{2}}$
$\Leftrightarrow \mu \notin \bar{X} \pm \frac{S_x}{\sqrt{n}} t_{n-1;1-\frac{\alpha}{2}}$

Likelihood-ratio regions

The principle from the theorem (Confidence interval $\Leftrightarrow$ test) can be applied to the general: LR-test. The likelihood-ration statistic is:

$2log(\lambda_n(X)) = 2log(\frac{p_{\hat{\theta}}(X)}{p_{\hat{\theta_0}}(X)})$ (relates to $H_0 : \theta = \theta_0$ vs $H_a : \theta \in \Theta$)

With critical region $[\chi^2_{k-k_0;1-\alpha}, \infty)$ with k & k_0 defined as in theorem.

Significance level α (approximately!). A $(1-\alpha)$-confidence interval (approximate) for θ (based on the LR-test for $H_0 : \theta = \theta_0$):

$\{\theta_0 : 2log(\lambda_n(X)) \leq \chi^2_{k-k_0;1-\alpha}\}$

$= \{\theta_0 : log(p_{\hat{\theta_{ML}}}(X)) - log(p_{\hat{\theta_0}}(X)) \leq \frac{1}{2}\chi^2_{k-k_0;1-\alpha}\}$

$= \{\theta_0 : log(p_{\hat{\theta_0}}(X)) - log(p_{\hat{\theta_{ML}}}(X)) \geq \frac{1}{2}\chi^2_{k-k_0;1-\alpha}\}$

Hence, observe that values of θ for which $log(p_\theta(X))$ is large are in the LR-confidence interval.

Likelihood-ration regions

Example

$X_1, \ldots, X_n \sim Poisson(\theta)$, $\theta > 0$ unknown. A $(1-\alpha)$-confidence interval for θ? Likelihood: $p_\theta(X) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta} \Rightarrow \theta\hat{_M}L = \bar{X}$

Now need log-likelihood:

$log(p_\theta(X)) = \sum_{i=1}^{n} x_i log(\theta) - \sum_{i=1}^{n} log(x_i!) - n\theta$ and a $(1-\alpha)$-confidence interval is then (approximately):

$\{\theta : \sum_{i=1}^{n} (log(\theta) - log(\bar{X}) - n(\theta - \bar{X}) \geq -\frac{1}{2}\chi^2_{1;1-\alpha}\}$

Optimality theory

Sufficient statistics

Often the observations $X_1, \ldots, X_n$ can be summarized by a lower dimensional statistic $V(X)$ without loss of relevant information on the parameter of interest.

Exmple

$X_1, \ldots, X_n \sim Bernoulli(p)$. Then $V(X) = \sum_{i=1}^{n} X_i$ sufficient for p.

In a study aimed to assess the fraction of students whose bike has been stolen, it suffices to know how many students in the sample experienced this. Not who.

Optimality theory

Definition

Let X be discrete and following p_θ. A statistic $V(X)$ is called <u>sufficient</u> if $P(X = x|V = v)$ does not depend on the unknown parameter θ for x and v.

Example

$V(X) = \sum_{i=1}^{n} X_i$

$P((X_1, \dots, X_n) = (x_1, \dots, x_n)|V = v) = \frac{P(X_1=x_1,\dots,X_n=x_n,V=v)}{P(V=v)}$

$= 0$ if $\sum x_i \neq v$

$= \frac{p^v(1-p)^{n-v}}{\binom{n}{v}p^v(1-p)^{n-v}} = \frac{1}{\binom{n}{v}}$ if $\sum x_i = v$.

This probability does not depend on p.

Hence, $\sum x_i = V$ is sufficient.

Optimality theory

Theorem

Let X be a discrete random variable. A statistic $V = v(x)$ is suficient iff there are function g and h such that $p_\theta = g_\theta(v(x))h(x)$ for all x and θ, with p_θ the probability distribution of x.

Example

$p_\theta(X_1, \ldots, X_n) = P(X_1 = x_1, \ldots, X_n = x_n) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$
$g_\theta(v(x))h(x) = 1$ met $v(x) = \sum x_i$