i

Formulas from Casella and Berger

Leibnitz rule:

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx = f(b(\theta),\theta)\cdot\frac{db}{d\theta}(\theta) - f(a(\theta),\theta)\cdot\frac{da}{d\theta}(\theta) + \int_{a(\theta)}^{b(\theta)}\frac{\partial f}{\partial\theta}(x,\theta)dx$$

Double expectation and double variance:

$$E[X] = E[E[X|Y]], \quad \mathrm{Var}[X] = E[\mathrm{Var}[X|Y]] + \mathrm{Var}[E[X|Y]]$$

Chebychev's inequality:

$$P(g(X)\geq r)\leq \frac{E[g(X)]}{r}$$

Cauchy-Schwarz inequality:

$$|E[XY]|\leq E[|XY|]\leq\sqrt{E\left[|X|^2\right]E\left[|Y|^2\right]}$$

Jensen's inequality:

If $g(x)$ is a convex function we have

$$E[g(X)]\geq g(E[X]).$$

Equality holds if and only if, for every line $a+bx$ that is a tangent to $g(x)$ at $x=E[X]$, $P(g(X)=a+bX)=1.$

Delta method:

Let Y_n be a sequence of random variables that satisfy $\sqrt{n}(Y_n-\theta)\longrightarrow N(0,\sigma^2)$ in distribution. For a given function $g(\theta)$ and a specific value of θ, suppose $g'(\theta)$ exists and is not 0. Then

$$\sqrt{n}(g(Y_n)-g(\theta))\longrightarrow N(0,\sigma^2(g'(\theta))^2)$$

in distribution.

Second-order delta method:

Let Y_n be a sequence of random variables that satisfy $\sqrt{n}(Y_n-\theta)\longrightarrow N(0,\sigma^2)$ in distribution. For a given function $g(\theta)$ and a specific value of θ, suppose $g'(\theta)=0$ and $g''(\theta)$ exists and is not 0. Then

$$\sqrt{n}(g(Y_n)-g(\theta))\longrightarrow \frac{\sigma^2 g''(\theta)}{2}\chi_1^2$$

in distribution.

Exponential family:

$$f(x|\theta)=h(x)c(\theta)\exp\left\{\sum_{i=1}^k w_i(\theta)t_i(x)\right\}$$

Expectation and variance formulas for an exponential family:

When X is a random variable from an exponential family, we have

$$E\left[\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right] = -\frac{\partial}{\partial \theta_j} \ln c(\theta),$$
$$\text{Var}\left[\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right] = -\frac{\partial^2}{\partial \theta_j^2} \ln c(\theta) - E\left[\sum_{i=1}^{k} \frac{\partial^2 w_i(\theta)}{\partial \theta_j^2} t_i(X)\right]$$

Complete statistic:

Let $f(t|\theta)$ be a family of pdfs/pmfs for a statistic $T(\mathbf{X})$. The family of probability distributions is called complete if $E_\theta[g(T)] = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ. Equivalently, $T(\mathbf{X})$ is called a complete statistic.

Basu's theorem:

If $T(\mathbf{X})$ is a complete and minimal sufficient statistic, then $T(\mathbf{X})$ is independent of every ancillary statistic.

Mean squared error:

The mean squared error (MSE) of an estimator W of a parameter θ is the function of θ defined by $E[(W-\theta)^2]$.

Cramer-Rao inequality:

$$\text{Var}_\theta[W(\mathbf{X})] \geq \frac{\left(\frac{d}{d\theta} E_\theta[W(\mathbf{X})]\right)^2}{E_\theta\left[\left(\frac{\partial}{\partial \theta} \ln f(\mathbf{X}|\theta)\right)^2\right]}$$

Rao-Blackwell:

Let W be any unbiased estimator of $\tau(\theta)$, and let T be a sufficient statistic for θ. Define

$$\phi(T) = E[W|T].$$

Then $E_\theta[\phi(T)] = \tau(\theta)$ and $\text{Var}_\theta[\phi(T)] \leq \text{Var}_\theta[W]$ for all θ; that is, $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$.

Risk function:

$$R(\theta, \delta) = E_\theta[L(\theta, \delta(\mathbf{X}))]$$

Bayes risk:

$$r(\delta) = E[R(\theta, \delta)]$$

Power function:

$$\beta(\theta) = P_\theta(\mathbf{X} \in R)$$

Likelihood ratio test:

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta|\mathbf{x})}{\sup_{\theta \in \Theta} L(\theta|\mathbf{x})}$$

Size α test:

$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$

Level α test:

$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$

Neyman-Pearson lemma:

Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1,$ where the pdf or pmf corresponding to θ_i is $f(\mathbf{x}|\theta_i), i = 0, 1,$ using a test with rejection region R that satisfy

$\mathbf{x} \in R$ if $f(\mathbf{x}|\theta_1) > k\, f(\mathbf{x}|\theta_0)$

and

$\mathbf{x} \in R^C$ if $f(\mathbf{x}|\theta_1) < k\, f(\mathbf{x}|\theta_0)$

for some $k \geq 0,$ and

$\alpha = P_{\theta_0}(\mathbf{X} \in R).$

Then any test that satisfy these three conditions is an UMP level α test.

P-value:

A p-value $p(\mathbf{X})$ is a test statistic satisfying $0 \leq p(\mathbf{x}) \leq 1$ for every sample point $\mathbf{x}.$ Small values of $p(\mathbf{X})$ give evidence that H_1 is true. A p-value is valid if, for every $\theta \in \Theta_o$ and every $0 \leq \alpha \leq 1,$

$P_\theta(p(\mathbf{X}) \leq \alpha) \leq \alpha.$

Valid p-value:

Let $W(\mathbf{X})$ be a test statistic such that large values of W give evidence that H_1 is true. For each sample point $\mathbf{x},$ define

$p(\mathbf{x}) = \sup_{\theta \in \Theta_0} P_\theta(W(\mathbf{X} \geq W(\mathbf{x})).$

Then $p(\mathbf{X})$ is a valid p-value.

Information number:

$\mathbf{I}(\theta) = E_\theta\left[\left(\frac{\partial}{\partial \theta} \ln f(\mathbf{X}|\theta)\right)^2\right] = -E_\theta\left[\frac{\partial^2}{\partial \theta^2} \ln f(\mathbf{X}|\theta)\right]$

Score test:

$S(\theta) = \frac{\partial}{\partial \theta} \ln f(\mathbf{X}|\theta)$