STA261

Ian Zhang

February 2023

1 Change of variable

1.1 Single variable

Let X be a $\mathbb{R}$-r.v. with PDF f_X and $\beta : \mathbb{R} \to \mathbb{R}$ be injective and strictly monotonic. Let $Z = \beta(X)$. Then

$$f_Z(z) = \begin{cases} f_X(\beta^{-1}(z))|(\beta^{-1}(z))'| & z \in \beta(\mathbb{R}) \\ 0 & \text{otherwise} \end{cases}$$

1.2 Multivariable

Let X be a $\mathbb{R}^d$-random vector with PDF $f_X : \mathbb{R}^d \to [0, \infty)$ and $\beta : \mathbb{R}^d \to \mathbb{R}^d$ be a function, and $Z = \beta(X)$ be a random vector. Let $(S_k)_{k \in N^0}$ partition $\mathbb{R}^d$ (ie: $\bigcup_k S_k = \mathbb{R}^d$ and for all $i, j \in \mathbb{N}^0$, $i \neq j \implies S_i \cap S_j = \emptyset$). Suppose $\int_{S_0} dV = 0$ and for all $k \in \mathbb{N}$, S_k is open and $\beta_k := \beta|_{S_k}$ is injective and $D\beta_k(a)$ (Jacobian of β_k at a) is invertible for all $a \in S_k$ (β_k is a diffeomorphism). Then

$$f_Z(z) = \begin{cases} f_X(\beta^{-1}(z))|\det D\beta^{-1}(z)| & z \in \beta(S_k), k \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$

2 Statistical inference introduction

Definition. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space.

- A random sample of size n is a set of r.v. $(X_1, \ldots, X_n)$
- An observed sample is a set of *constants* $(x_1, \ldots, x_n)$

Definition. A statistical model is a family of PDFs given by

$$\{f_\theta : \theta \in D\}$$

where D is the domain of θ.

Definition. $(X_i)_{i\in\mathbb{N}}$ are mutually independent if $\mathbb{P}(\bigcap_{i\in I}\{x_j \in A_n\}) = \prod_{i\in I}\mathbb{P}(x_i \in A_i)$ for any finite $I \subseteq \mathbb{N}$ and any $A_1, A_2, \ldots, \subseteq \mathbb{R}$.

Definition. If we write $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} f$, where f is a PDF, we mean $(X_1, \ldots, X_n)$ are mutually independent and X_i all have a PDF of f

3 Sums of random variables from a random sample

Definition. Let $(X_1, \ldots, X_n)$ be a random sample.

- $\overline{X} := \frac{X_1+\ldots+X_n}{n}$ is the sample mean
- $S^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ is the sample variance
- $S := \sqrt{S^2}$ is the sample standard deviation

$\overline{X}, S^2$ and S are all random variables and statistics.

Definition. Let $(x_1, \ldots, x_n)$ be an observed sample.

- $\overline{x} := \frac{x_1+\ldots+x_n}{n}$ is the observed sample mean
- $s^2 := \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2$ is the observed sample variance
- $s := \sqrt{s^2}$ is the observed sample standard deviation

Theorem. Let $x_1, \ldots, x_n$ be an numbers and $\overline{x} = \frac{x_1+\ldots+x_n}{n}$. Then

1. $\min_a \sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \overline{x})^2$
2. $(n-1)s^2 = \sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=1}^n x_i^2 - n\overline{x}^2$

Lemma. Suppose $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} f$. Let $g : \mathbb{R} \to \mathbb{R}$ be such that $\mathbb{E}|g(X_1)| < \infty$ and $\mathbb{E}|g(X_1)^2| < \infty$. Then

- $\mathbb{E}\left(\sum_{i=1}^n g(X_i)\right) = n\mathbb{E}(g(X_1))$
- $\text{Var}\left(\sum_{i=1}^n g(X_i)\right) = n\text{Var}(g(X_1))$

Theorem. Suppose $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} f$ with $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}|X_1^2| < \infty$. let $\mathbb{E}(X_1) = \mu$ and $\text{Var}(X_1) = \sigma^2$. Then

1. $\mathbb{E}(\overline{X}) = \mu$
2. $\text{Var}(\overline{X}) = \frac{\sigma^2}{n}$
3. $\mathbb{E}(S^2) = \sigma^2$

Generalized Chebyshev's Inequality: If Y is a $\mathbb{R}$-r.v., then

$$\mathbb{P}(|Y| \geq \varepsilon) \leq \frac{\mathbb{E}(|Y|^p)}{\varepsilon^p}, \forall \varepsilon > 0, p \in [1, \infty)$$

In particular, if $p = 2$, then we get the inequality from STA257.

Theorem (Distribution of Independent Sums)**.** If X, Y are independent continuous random variables with respective densities f_X, f_Y, then the density of $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z - w) dw$$

4 Sampling from the normal distribution

Definition. If a random variable Z has PDF $f_Z(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right)$ where $\mu \in \mathbb{R}$ and $\sigma > 0$, then we say $Z \sim N(\mu, \sigma^2)$.

Lemma. a) If $Z \sim N(\mu, \sigma^2)$, then $\mathbb{E}(Z) = \mu$ and $\text{Var}(Z) = \sigma^2$.

b) Let $Z_1 \sim N(\mu_1, \sigma_1^2)$ and $Z_2 \sim N(\mu_2, \sigma_2^2)$. If $Z_1 \perp\!\!\!\perp Z_2$, then $aZ_1 + bZ_2 \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

c) Let $Z_i \sim N(\mu_1, \sigma^2)$ and $(Z_1, \ldots, Z_n)$ be mutually independent. Then, $\sum_{i=1}^n a_i Z_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$

Definition. Let $Z \sim N(0, 1)$. Then, Z^2 has χ_1^2 distribution with 1-degree of freedom.
Let $(Z_1, \ldots, Z_n) \overset{\text{iid}}{\sim} N(0, 1)$. We say $\sum_{i=1}^n Z_i^2$ has χ_n^2 distribution.

The PDF of χ_n^2 is

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where

$$\Gamma(z) := \int_0^\infty r^{z-1} e^{-r} dr, z > 0 \qquad \Gamma(k) = (k-1)!$$

Notice that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, so the PDF of χ_1^2 is

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\frac{x}{2}} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Lemma. Let $(X_1, \ldots, X_n)$ be mutually independent. If $X_i \sim \chi_{P_i}^2$, then

$$\sum_{i=1}^n X_i \sim \chi_{P_1 + \ldots + P_n}^2$$

Theorem. Let $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. Then

1. $\overline{X} \perp\!\!\!\perp S^2$

2. $\overline{X} \sim N(\mu, \sigma^2/n)$

3. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Note that b) is useful for error analysis when estimating μ with a *known* σ^2.

$$\mathbb{P}(|\overline{X} - \mu| > \varepsilon) = \int_{-\infty}^{-\varepsilon} \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\frac{\sigma^2}{n}}\right) dx + \int_{\varepsilon}^{\infty} \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\frac{\sigma^2}{n}}\right) dx$$

c) is useful for finding the PDF of $\frac{S^2}{\sigma^2}$. If f is the PDF of χ^2_{n-1}, let $\beta(x) = \frac{1}{n-1}x$, thus $\beta' > 0$ and $\beta^{-1}(y) = (n-1)y \implies (\beta^{-1})' = n-1$. Then

$$f_{\frac{S^2}{\sigma^2}}(y) = f((n-1)x)(n-1) = \frac{n-1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}(n-1)^{\frac{n}{2}-1}x^{\frac{n}{2}-1}e^{-\frac{n-1}{2}x}$$

To perform actual estimations of μ and σ, we consider the random variable $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$. If μ is unknown but we know σ, then μ is the only unknown quantity and we can use this random variable to estimate μ (give confidence interval, etc.).
However, if we don't know σ, then we can use the random variable $S = \sqrt{S^2}$ to estimate σ and then perform error analysis (since $\mathbb{E}(S^2) = \sigma^2$). Thus, we have a random variable

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{1}$$

Definition. Let $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} N(0,1)$. The random variable presented in (1) has Student's t-distribution with $(n-1)$-degrees of freedom.
A random variable T has Student's t-distribution with p degrees of freedom if T has PDF

$$f_T(t) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\sqrt{p\pi}} \frac{1}{\left(1 + \frac{t^2}{p}\right)^{\frac{p+1}{2}}} \longrightarrow T \sim t_p$$

Definition. Let $(X_1, \ldots, X_n)$ be a random sample from a $N(\mu_X, \sigma_X^2)$ population and let $(Y_1, \ldots, Y_m)$ be a random sample from a $N(\mu_Y, \sigma_Y^2)$ population. The random variable

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$$

has Snedcor's F-distribution with $n-1$ and $m-1$ degrees of freedom.
A random variable F has F-distribution with p and q degrees of freedom if F has PDF

$$f_F(x) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{\frac{p}{2}} \frac{x^{\frac{p}{2}-1}}{\left(1 + \frac{p}{q}x\right)^{\frac{p+q}{2}}} \quad 0 < x < \infty$$

Let $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ be random samples as defined in the above definition.

We can compare relative variability (ie: $\frac{\sigma_X^2}{\sigma_Y^2}$) by observing $\frac{S_X^2}{S_Y^2}$. Notice the relation

$$\frac{S_X^2/S_Y^2}{\sigma_X^2\sigma_Y^2} = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$$

From an above Theorem, we know that $\frac{(n-1)S_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$ and $\frac{(m-1)S_Y^2}{\sigma_Y^2} \sim \chi^2_{m-1}$. Additionally, the quantity presented has $F_{n-1,m-1}$ distribution by definition. Observing the expected value, we have

$$\begin{aligned}
\mathbb{E}(F_{n-1,m-1}) &= \mathbb{E}\left(\frac{\chi^2_{n-1}/(n-1)}{\chi^2_{m-1}/(m-1)}\right) \\
&= \mathbb{E}\left(\frac{\chi^2_{n-1}}{n-1}\right)\mathbb{E}\left(\frac{m-1}{\chi^2_{m-1}}\right) && \text{independence} \\
&= \frac{n-1}{n-1}\frac{m-1}{m-3} \\
&= \frac{m-1}{m-3}
\end{aligned}$$

5 Convergence

Definition. A sequence of $\mathbb{R}$-r.v.s $(Y_n)_{n\in\mathbb{N}}$ converges in probability to a random variable Y if for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - Y| \geq \varepsilon) = 0 \iff \lim_{n\to\infty} P(|Y_n - Y| < \varepsilon) = 1$$

If so, we write $Y_n \xrightarrow{\mathbb{P}} Y$.

5.1 Converges in probability

Theorem (Weak Law of Large Numbers)**.** Let $(X_i)_{i\in\mathbb{N}}$ be an iid sequence of R-r.v.s with finite mean μ and finite variance $\sigma^2 < \infty$. Define $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. For every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| < \varepsilon) = 1$$

That is, $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$.

Proof. Fix $\varepsilon > 0$. By Chebyshev's Inequality, we have $\mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\mathbb{E}(|\overline{X}_n - \mu|^2)}{\varepsilon^2} = \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$. Taking $n \to \infty$, we have $\frac{\sigma^2}{n\varepsilon^2} \to 0$, thus $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$ by definition. ∎

Theorem. Suppose $(Y_n)_{n\in\mathbb{N}}$ converges in probability to Y. Let $h : \mathbb{R} \to \mathbb{R}$ be continuous. Then, the sequence of random variables $\{h(Y_n)\}_{n\in\mathbb{N}}$ converges in probability to $h(Y)$.

5.2 Almost surely convergence

Definition. A sequence of $\mathbb{R}$ random variables $(X_i)i \in \mathbb{N}$ converges almost surely to X if for all $\varepsilon > 0$,

$$\mathbb{P}(\lim_{n\to\infty} |X_n - X| < \varepsilon) = 1$$

If so, we say $X_n \xrightarrow{a.s.} X$.

Note that almost surely **implies** convergence in probability, but not vice versa.

Theorem. Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that $Y_n \xrightarrow{a.s.} Y$. Let $h : \mathbb{R} \to \mathbb{R}$ be continuous. Then, the sequence $(h(Y_n))_{n\in\mathbb{N}}$ converges almost surely to $h(Y)$.

Theorem (Strong Law of Large Numbers)**.** Let $(X_i)_{i\in\mathbb{N}}$ be iid $\mathbb{R}$-r.v.s and X a $\mathbb{R}$-r.v. Suppose $\mathbb{E}|X_i| < \infty, \mathbb{E}|X_i^2| < \infty$ and $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$. Define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then $\overline{X}_n \xrightarrow{a.s.} \mu$.

5.3 Convergence in distribution

Definition. A sequence of $\mathbb{R}$-r.v.s $(X_n)_{n\in\mathbb{N}}$ converges to a $\mathbb{R}$-r.v. X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all x at which F_X is continuous. If so, then $Y_n \xrightarrow{D} Y$.

Theorem. If $Y_n \xrightarrow{P} Y$, then $Y_n \xrightarrow{D} Y$.

A case where the converse is true is presented in the following theorem

Theorem. If $Y_n \xrightarrow{D} Y = c$, then $Y_n \xrightarrow{\mathbb{P}} Y$.

Theorem (Central Limit Theorem)**.** Let $(X_i)_{i\in\mathbb{N}}$ be iid with $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}|X_1^2| < \infty$. Let $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $Y_n = \frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu)$. Then $Y_n \xrightarrow{D} Z \sim N(0,1)$.

- Notice this theorem does not depend on the assumption that all X_i are $N(0,1)$ distributed

Theorem (Slutsky's)**.** Let $(Y_n)_{n\in\mathbb{N}}$ and $(Z_n)n \in \mathbb{N}$ be sequences of random variables with $Y_n \xrightarrow{D} Y$ and $Z_n \xrightarrow{\mathbb{P}} c$ for some $c \in \mathbb{R}$. Then we have

- $Y_n + Z_n \xrightarrow{D} Y + c$
- $Z_n Y_n \xrightarrow{D} cY$

Additionally, if $c > 0$, then we have $\frac{Y_n}{Z_n} \xrightarrow{D} \frac{Y}{c}$.

Suppose $(X_1, \dots, X_n) \overset{\text{iid}}{\sim} f$. Suppose $\mathbb{E}(X_i) = \mu < \infty$ and $\text{Var}(X_i) = \sigma^2 < \infty$.
Suppose we know σ but not μ. Then we can use $\overline{X}_n$ to estimate μ. By the CLT, $\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu)$ as a CDF close to that of Φ (the CDF of a $N(0,1)$ distribution)
IF we don't know σ either, we can use S_n to estimate σ and Slutsky's theorem. Let $H_n := \frac{\sqrt{n}}{S_n}(\overline{X}_n - \mu) = \frac{\sigma}{S_n}\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu)$. Notice that $\frac{\sigma}{S_n} \overset{\mathbb{P}}{\to} 1 > 0$ since $S_n \overset{\mathbb{P}}{\to} \sigma$. We also know from CLT that $\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) \overset{D}{\to} Z \sim N(0,1)$. Thus, by Slutsky's theorem, we know $H_n \overset{D}{\to} X$. This means we have

$$\mathbb{P}\left(\left|\frac{X_n - \mu}{S_n}\right| > \varepsilon\right) = \mathbb{P}(|\overline{X}_n - \mu| > S_n\varepsilon) \approx \Phi(-\sqrt{n}\varepsilon) + (1 - \Phi(\sqrt{n}\varepsilon))$$

6 Method of finding estimators

Definition. A point estimator is any function $W(X_1, \dots, X_n)$ of a sample $(X_1, \dots, X_n)$.

6.1 Method of moments

Consider $(X_1, \dots, X_n) \overset{\text{iid}}{\sim} f_{\theta_1, \dots, \theta_k}$ where $\theta_1, \dots, \theta_k$ are unknown. Let $(\eta_1, \dots, \eta_k)$ be dummy variables of $(\theta_1, \dots, \theta_k)$. For $l \in \mathbb{N}$, define

$$v_l(\eta_1, \dots, \eta_k) := \mathbb{E}_{\eta_1, \dots, \eta_k}(x^l) = \int_{\mathbb{R}} x^l f_{v_1, \dots, v_k}(x)dx$$

and

$$M_l := \frac{1}{n}\sum_{i=1}^n x_i^l$$

The method of moments estimator $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ of $(\theta_1, \dots, \theta_k)$ is obtained by solving for the $(\eta_1, \dots, \eta_k)$ that solves

$$\begin{cases} M_1 = v_1(\eta_1, \dots, \eta_k) \\ \quad\vdots \\ M_L = v_L(\eta_1, \dots, \eta_k) \end{cases}$$

6.2 Maximum likelihood estimator

Definition. Let $(X_1, \dots, X_n) \overset{\text{iid}}{\sim} f_\theta$ where $\theta \in D \subseteq \mathbb{R}^k$ with joint PDF $f_{\theta,(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$. The likelihood of θ given realized sample $(x_1, \dots, x_n)$ is

$$L(\theta | x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$$

Definition. Let f_θ be a PDF or PMF with $\theta \in D \subseteq \mathbb{R}^k$. For $(X_1, \dots, X_n) \in \mathbb{R}^n$, let $\hat{\theta}(X_1, \dots, X_n)$ attain $\max_{\theta \in D} L(\theta | x_1, \dots, x_n)$ where $(x_1, \dots, x_n)$ are fixed. The MLE of parameter θ based on random sample $(X_1, \dots, X_n)$ is $\hat{\theta}(X_1, \dots, X_n)$. IF D is not the largest possible set where f_θ is well-defined, call $\hat{\theta}(X_1, \dots, X_n)$ the restricted MLE.

- Alternatively, we can observe the behaviour of

$$\log(L(\theta|x_1, \dots, x_n)) = \sum_{i=1}^{k} \log(L(\theta|x_1, \dots, x_n))$$

What if we want to estimate $\tau(\theta_0)$ with MLE?

- We can define likelihood of $\eta = \tau(\theta)$
  - If τ is injective, then define $L^\tau(\eta|x_1, \dots, x_n) = L(\tau^{-1}(\eta)|x_1, \dots, x_n)$
  - If not, set $L^\tau(\eta|x_1, \dots, x_n) = \sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta|x_1, \dots, x_n)$

Let $\hat{\eta}(X_1, \dots, X_n)$ attain $\max_\eta L^\tau(\eta|x_1, \dots, x_n)$. We say $\hat{\eta}(X_1, \dots, X_n)$ is the MLE of $\tau(\theta)$.

Theorem. If $\hat{\theta}(X_1, \dots, X_n)$ attains $\max_{\theta \in D} L(\theta|x_1, \dots, x_n)$, then for any $\tau(\theta)$, $(X_1, \dots, X_n) \mapsto \tau(\hat{\theta}(X_1, \dots, X_n))$ also attains $\max_{\eta \in \tau(D)} L^\tau(\eta|x_1, \dots, x_n)$.

7 Best unbiased estimators

Definition. An estimator W^* is a best unbiased estimator of $\tau(\theta)$ where $\tau : D \to \mathbb{R}$ if W^* satisfies

- $\mathbb{E}_\theta(W^*) = \tau(\theta)$ for all $\theta \in D$
- For any other unbiased estimator of $\tau(\theta)$, W, we have $\text{Var}_\theta(W^*) \leq \text{Var}_\theta(W)$ for all $\theta \in D$

W^* is also called a uniform minimum variance unbiased estimator of $\tau(\theta)$.

Lemma (Hölder's Inequality)**.** Let $p, q \in (1, \infty)$ satisfy $p^{-1}+q^{-1} = 1$. Let X, Y be $\mathbb{R}$-r.v. and not constant. Then, $\mathbb{E}(|XY|) \leq \mathbb{E}(|X|^p)^{\frac{1}{p}}\mathbb{E}(|Y|^q)^{\frac{1}{q}}$.

Corollary (Cauchy-Schwarz)**.** Let X, Y be $\mathbb{R}$-r.v., not constant. Then,

$$\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(|X|)^2\mathbb{E}(|Y|^2)} \qquad \text{Cov}(X, Y)^2 \leq \text{Var}(X)\text{Var}(Y)$$

Equalities iff $\exists k \geq 0$ s.t. $|X| = k|Y|$ or $\exists k \in \mathbb{R}$ s.t. $X - \mathbb{E}(X) = k(Y - \mathbb{E}(Y))$ respectively.

Theorem (Cramer-Rao Inequality)**.** Let $\theta \in D \subseteq \mathbb{R}$ and $(X_1, \dots, X_n) \sim g_\theta$ that is C^∞ in θ and each X_i is $\mathcal{X}$-valued. Suppose $W = W(X_1, \dots, X_n)$ is any estimator satisfying

$$\frac{d}{d\theta} E_\theta(W) = \int_{\mathcal{X}^n} \frac{\partial}{\partial\theta}(W(x_1, \dots, x_n) \cdot g_\theta(x_1, \dots, x_n)) d(x_1, \dots, x_n)$$

and $\mathrm{Var}_\theta(X) < \infty$. Then,

$$\mathrm{Var}_\theta(W) \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta(W(X_1, \dots, X_n))\right)^2}{\mathbb{E}\left(\left(\frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right)^2\right)}$$

Proof. Notice that

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right) &= \int_{\mathcal{X}^n} \frac{\partial}{\partial\theta}\log(g_\theta(x_1, \dots, x_n))d(x_1, \dots, x_n) \\
&= \int_{\mathcal{X}^n} \frac{\partial}{\partial\theta} g_\theta(x_1, \dots, x_n) d(x_1, \dots, x_n) \\
&= \frac{\partial}{\partial\theta}(1) \\
&= 0
\end{aligned}$$

So, we have

$$\begin{aligned}
\mathrm{Cov}\left(W, \frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right) &= \mathbb{E}\left(W\frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right) \\
&- \quad \mathbb{E}(W)\mathbb{E}\left(\frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right) \\
&= \mathbb{E}\left(W\frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right) \\
&= \int_{\mathcal{X}^n} W(x_1, \dots, x_n)\frac{\partial}{\partial\theta}\log(g_\theta(x_1, \dots, x_n))g(x_1, \dots, x_n)d(x_1, \dots, x_n) \\
&= \int_{\mathcal{X}^n} W(x_1, \dots, x_n)\frac{\partial}{\partial\theta} g_\theta(x_1, \dots, x_n)d(x_1, \dots, x_n) \\
&= \frac{d}{d\theta}E_\theta(W) \qquad \text{by assumption} \\
&\leq \sqrt{\mathrm{Var}(W)\mathrm{Var}\left(\frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right)} \qquad \text{by Cauchy-Schwarz}
\end{aligned}$$

But notice

$$\begin{aligned}
\mathrm{Var}\left(\frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right) &= \mathbb{E}\left(\left(\frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right)^2\right) - \mathbb{E}\left(\frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right)^2 \\
&= \mathbb{E}\left(\left(\frac{\partial}{\partial\theta}\log(g_\theta(X_1, \dots, X_n))\right)^2\right)
\end{aligned}$$

which completes the proof. ■

Corollary (Cramer-Rao; iid case)**.** Assume the $(X_1, \dots, X_n)$ from the assumptions of

Cramer-Rao are iid. Then

$$\operatorname{Var}_\theta(W) \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta(W)\right)^2}{n\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log(g_\theta(X_1))\right)^2\right)}$$

Lemma. If f_θ satisfies $\frac{d}{d\theta}\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\log(f_\theta(X_1))\right) = \int_{\mathcal{X}} \frac{d}{d\theta}\left[\left(\frac{\partial}{\partial\theta}\log(f_\theta(x))\right) f_\theta(x)\right] dx$, then

$$\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log(f_\theta(X_1))\right)^2\right) = -\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log(f_\theta(X))\right)$$

Corollary (Attainment)**.** Let $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} f_\theta$ where $\theta \in D \subseteq \mathbb{R}$ and f_θ satisfies assumptions of Cramer-Rao. Let $L(\theta|x_1, \ldots, x_n) = \prod_{i=1}^n f_\theta(x_i)$ be the likelihood of θ. Let $W = W(X_1, \ldots, X_n)$ be an unbiased estimator of $\tau(\theta)$. Then, W attains equality in Cramer-Rao iff there exists $a : D \to [0, \infty)$ such that $a(\theta)(W(X_1, \ldots, X_n) - \tau(\theta)) = \frac{\partial}{\partial\theta}\log(L(\theta|x_1, \ldots, x_n))$.

8 Sufficiency and unbiasedness

Definition. A family of PDFs/PMFs $\{f_\theta : \theta \in D \subseteq \mathbb{R}^d\}$ of $\mathbb{R}$-r.v. is an exponential family if

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

for all $\theta \in D$ where $h(x) \geq 0$, $t_i(x)$ are real-valued functions, $c(\theta) \geq 0$ and $w_i(\theta)$ are real-valued functions.

Definition. Let $\{f_\theta : \theta \in D\}$ be an exponential family where $D \subseteq \mathbb{R}^d$ is non-empty and open. Let $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} f_\theta$. Call $T(x_1, \ldots, x_n) = \left(\sum_{j=1}^n t_1(x_j), \ldots, \sum_{j=1}^n t_k(x_j)\right)$ a complete sufficient statistic. $\tilde{T}$ is also a complete sufficient statistics if there exists an injective g such that $\tilde{T} = g(T)$.

Theorem. Let $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} f_\theta$ where $\theta \in D \subseteq \mathbb{R}^d$ and $T = T(X_1, \ldots, X_n)$ a complete sufficient statistic for a parameter θ and $\varphi(T)$ be any estimator based only on T. Then $\varphi(T)$ is the unique best unbiased estimator of $\mathbb{E}_\theta(\varphi(T))$.

Definition. Let X be a $\mathbb{R}$-r.v., Y be a $\mathbb{R}^d$-r.v., $\eta : \mathbb{R}^d \to \mathbb{R}$. The conditional expectation of $\eta(Y)$ given X is a real-valued function of X that satisfies

$$\mathbb{E}(h(X)\eta(Y)) = \mathbb{E}(h(X)\mathbb{E}(\eta(Y)|X))$$

for all bounded $h : \mathbb{R}^n \to \mathbb{R}$.

Properties:

- $\mathbb{E}(a\eta(Y) + b\gamma(Z)|X) = a\mathbb{E}(\eta(Y)|X) + b\mathbb{E}(\gamma(Z)|X)$ for all $a, b \in \mathbb{R}$
- $\mathbb{E}(\eta(X)|X) = \eta(X)$
- $\mathbb{E}(E(\eta(Y)|X)) = E(\eta(Y))$
- $\text{Var}(\eta(Y)) = \text{Var}(\mathbb{E}(\eta(Y)|X)) + \mathbb{E}(\text{Var}(\eta(Y)|X))$ where $\text{Var}(\eta(Y)|X) := \mathbb{E}((\eta(Y) - \mathbb{E}(\eta(Y)|X))^2|X) = \mathbb{E}(\eta(Y)|X) - \mathbb{E}(\eta(Y)|X)^2$

Theorem (Rao-Blackwell)**.** Let W be any unbiased estimator of $\tau(\theta)$ and T be a sufficient statistic of θ. Let $\varphi(T) = \mathbb{E}_\theta(W|T)$. Then, $\mathbb{E}_\theta(\varphi(T)) = \tau(\theta)$ and $\text{Var}_\theta(\varphi(T)) \leq \text{Var}_\theta(W)$ for all θ. In other words, $\varphi(T)$ is a better unbiased estimator than W.

A way to find the BUE:

- Given $(X_1, \dots, X_n) \overset{\text{iid}}{\sim} f_\theta$ where $\theta \in D \subseteq \mathbb{R}$ where f_θ is from an exponential family and $D \neq \emptyset$ is open, $T = T(X_1, \dots, X_n)$ is a complete sufficient statistic, and we want to estimate $\tau(\theta)$ where $\tau : D \to \mathbb{R}$
  - Find $H = H(X_1, \dots, X_n)$ such that $\mathbb{E}_\theta(H) = \tau(\theta)$ for all $\theta \in D$ so H is unbiased
  - Compute $\varphi(T) = \mathbb{E}_\theta(H|T)$, then $\varphi(T)$ is the BUE by a theorem above and Rao-Blackwell

Theorem. If W is the BUE of $\tau(\theta)$ then W is unique.

Theorem. Let $\tau : D \to \mathbb{R}$ and W be an unbiased estimator of $\tau(\theta)$. Then, W is the BUE iff $\text{Cov}_\theta(W, V) = 0$ for all θ and estimators V such that $\mathbb{E}_\theta(V) = 0$.

9 Hypothesis testing

Definition. A hypothesis is a statement about a population parameter

Definition. 2 complementary hypotheses are called null hypothesis and alternative hypothesis, denoted H_0 or H_1 respectively.

The general format for a hypothesis test is $H_0 : \theta \in D_0$, $H_1 : \theta \in D_0^c$ where D_0 is a subset of the parameter space, D.

Definition. A hypothesis testing procedure is a rate that specifies

1. For which sample values is H_0 accepted
2. For which sample values is H_1 accepted (H_0 is rejected)

The set of sample values for which H_0 is rejected is called the rejection region and denoted R.

9.1 Typical construction of a test

1. Specify domain D, null and alternative hypotheses
2. Specify test statistic $W = W(X_1, \ldots, X_n)$
3. If possible, set up R based on W such that $\mathbb{P}_\theta((X_1, \ldots, X_n) \in R)$ is small for $\theta \in D_0$ but large for $\theta \in D_0^c$
   - This is so that if H_0 is rejected, we're confident H_1 holds

9.2 Method of finding tests

Definition. The likelihood ratio test statistic for $H_0 : \theta \in D_0 \subseteq D$ vs $H_1 : \theta \in D_0^c$ is

$$\lambda(x_1, \ldots, x_n) = \frac{\sup_{\theta \in D_0} L(\theta|x_1, \ldots, x_n)}{\sup_{\theta \in D} L(\theta|x_1, \ldots, x_n)}$$

A likelihood ratio test (LRT) is any test with rejection region

$$R = \{(x_1, \ldots, x_n) : \lambda(x_1, \ldots, x_n) \leq c\}$$

for some $c \in [0, 1]$.

Definition. Let $(X_1, \ldots, X_n) \sim f_\theta$. Suppose that $\prod_{i=1}^n f_\theta(x_i) = g_\theta(T(x_1, \ldots, x_n))h(x_1, \ldots, x_n)$ for some T, h that don't depend on θ and g_θ not on $(x_1, \ldots, x_n)$. Then T is called a sufficient statistic.

Theorem (Theorem 8.2.4)**.** Under the setting of the definition of sufficient statistic, let $\lambda(x_1 \ldots, x_n)$ be a LRT. Then, $\exists \lambda^*$ on the range of T such that $\lambda^*(T(x_1, \ldots, x_n)) = \lambda(x_1, \ldots, x_n)$.

Proof. By definition,

$$\begin{aligned}\lambda(x_1, \ldots, x_n) &= \frac{\sup_{\theta \in D_0} L(\theta|x_1, \ldots, x_n)}{\sup_{\theta \in D} L(\theta|x_1, \ldots, x_n)} \\ &= \frac{\sup_{\theta \in D_0} g(T(x_1, \ldots, x_n))h(x_1, \ldots, x_n)}{\sup_{\theta \in D} g(T(x_1, \ldots, x_n))h(x_1, \ldots, x_n)} \\ &= \frac{\sup_{\theta \in D_0} g(T(x_1, \ldots, x_n))}{\sup_{\theta \in D} g(T(x_1, \ldots, x_n))}\end{aligned}$$

Setting $\lambda^*(x_1, \ldots, x_n) = \frac{\sup_{\theta \in D_0} g(T(x_1, \ldots, x_n))}{\sup_{\theta \in D} g(T(x_1, \ldots, x_n))}$, the rest of the claim follows. ■

9.3 Methods of evaluating tests

Definition.
- Type I Error: incorrectly reject H_0 but $\theta \in D_0$, probability of Type I Error is

$$\mathbb{P}_\theta((X_1, \ldots, X_n) \in R)$$

- Type II Error: incorrectly accept H_0 but $\theta \in D_0^c$, probability of Type II Error is

$$\mathbb{P}_\theta((X_1, \ldots, X_n) \notin R)$$

Definition. The power function of a hypothesis test with rejection region R is

$$\beta(\theta) = \mathbb{P}_\theta((X_1, \ldots, X_n) \in R)$$

Probability of Type I Error: $\beta(\theta), \theta \in D_0$
Probability of Type II Error: $1 - \beta(\theta), \theta \in D_0^c$

Definition. For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a size α test if $\sup_{\theta \in D_0} \beta(\theta) = \alpha$.

Definition. For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a level α test if $\sup_{\theta \in D_0} \beta(\theta) \leq \alpha$.

9.3.1 Picking c in the likelihood ratio test

- Pick R by choosing c such that $\beta(\theta) \leq \alpha$, α small

Suppose $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} N(\theta, 1)$, $\theta \in \mathbb{R}$, $H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$. As per the LRT, define the rejection region

$$R = \{(x_1, \ldots, x_n) : |\overline{x} - \theta_0| \geq \frac{c}{\sqrt{n}}\}$$

Then,

$$\begin{aligned}
\beta(\theta_0) &= \mathbb{P}_{\theta_0}\left(|\overline{X} - \theta_0| \geq \frac{c}{\sqrt{n}}\right) \\
&= \mathbb{P}_{\theta_0}\left(\overline{X} - \theta_0 \leq -\frac{c}{\sqrt{n}}\right) + \mathbb{P}_{\theta_0}\left(\overline{X} - \theta_0 \geq \frac{c}{\sqrt{n}}\right) \\
&= \mathbb{P}_{\theta_0}(\sqrt{n}(\overline{X} - \theta_0) \leq -c) + \mathbb{P}_{\theta_0}(\sqrt{n}(\overline{X} - \theta_0) \geq c) \\
&= 2(1 - \Phi(c))
\end{aligned}$$

since $\overline{X} \sim N(\theta, \frac{1}{n})$. Setting $\alpha = 2(1 - \Phi(c))$, we get a size α test. Equivalently, we set $c = z_{\frac{\alpha}{2}}$ [1].

Definition. A test with power function $\beta(\theta)$ is unbiased if $\beta(\theta') \leq \beta(\theta'')$ for any $\theta' \in D_0$ and $\theta'' \in D_0^c$.

When designing a test,

1. Want a level α with small α so Type I Error is unlikely so we can be confident when rejection occurs

[1] For $q \in [0, 1]$, define z_q to a number such that for a random variable $Z \sim N(0, 1)$, $\mathbb{P}(Z > z_q) = q$

2. Want Type II Error as small as possible so that it's easier to reject given H_1 is true

Definition. A test is the uniformly most powerful level α test if $\beta(\theta) \geq \beta_1(\theta)$ for all $\theta \in D_0^c$ and every β_1 that is a power function of another level α test.

Theorem (Neyman Pearson)**.** Let $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} f_\theta$, $\theta \in \{\theta_0, \theta_1\}$, $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$. Suppose there exists $k > 0$ such that

$$R \supseteq \left\{(x_1, \ldots, x_n) : \prod_{i=1}^n f_{\theta_1}(x_i) > k \prod_{i=1}^n f_{\theta_0}(x_i)\right\}$$

and

$$R^c \supseteq \left\{(x_1, \ldots, x_n) : \prod_{i=1}^n f_{\theta_1}(x_i) < k \prod_{i=1}^n f_{\theta_0}(x_i)\right\}$$

Additionally, suppose $\mathbb{P}_{\theta_0}((X_1, \ldots, X_n) \in R) = \alpha$ for some $\alpha \in (0, 1)$. Then, the test with R is the UMP size α test.

Proof. WLOG, suppose f_θ is a PDF. Clearly, the test of of size α, thus it's also a level α test. Let $\beta(\theta) = \mathbb{P}_\theta((X_1, \ldots, X_n) \in R)$ be the power function. Let R' be the rejection region of another level α test, and $\beta'(\theta) = \mathbb{P}_\theta((X_1, \ldots, X_n) \in R)$ be the corresponding power function. Note that

$$(\mathbb{1}_R(x_1, \ldots, x_n) - \mathbb{1}_{R'}(x_1, \ldots, x_n)) \left(\prod_{i=1}^n f_{\theta_1}(x_i) - k \prod_{i=1}^n f_{\theta_0}(x_i)\right) \geq$$

since

$$\mathbb{1}_R(x_1, \ldots, x_n) = 1 \Longleftarrow \prod_{i=1}^n f_{\theta_1}(x_i) - k \prod_{i=1}^n f_{\theta_0}(x_i) \geq 0$$
$$\mathbb{1}_R(x_1, \ldots, x_n) = 0 \Longleftarrow \prod_{i=1}^n f_{\theta_1}(x_i) - k \prod_{i=1}^n f_{\theta_0}(x_i) \leq 0$$

by assumption. So,

$$\begin{aligned}
0 &\leq \int_{\mathbb{X}^n} (\mathbb{1}_R(x_1, \ldots, x_n) - \mathbb{1}_{R'}(x_1, \ldots, x_n)) \left(\prod_{i=1}^n f_{\theta_1}(x_i) - k \prod_{i=1}^n f_{\theta_0}(x_i)\right) d(x_1, \ldots, x_n) \\
&= \beta(\theta_1) - \beta'(\theta_1) - k(\beta(\theta_0) - \beta'(\theta_0)) \\
&\leq \beta(\theta_1) - \beta'(\theta_1)
\end{aligned}$$

■

Corollary (8.3.13)**.** Suppose the assumption of Neyman-Pearson. Let $T(X_1, \ldots, X_n)$ be a sufficient statistic and g_θ be the PDF/PMF of T. Suppose $\exists k > 0$ such that

$$R \supseteq \{(x_1, \ldots, x_n) : g_{\theta_1}(T(x_1, \ldots, x_n)) > k g_{\theta_0}(T(x_1, \ldots, x_n))\}$$

and

$$R^c \supseteq \{(x_1, \ldots, x_n) : g_{\theta_1}(T(x_1, \ldots, x_n)) < k g_{\theta_0}(T(x_1, \ldots, x_n))\}$$

Suppose additionally $\alpha = \mathbb{P}_{\theta_0}((X_1, \ldots, X_n) \in R)$. Then, this test is a UMP level α test.

Proof. Recall from the definition of a sufficient statistic that

$$\prod_{i=1}^{n} f_{\theta_i}(x_i) = g_{\theta_i}(T(x_1, \ldots, x_n))h(x_1, \ldots, x_n)$$

Thus,

$$\prod_{i=1}^{n} f_{\theta_1}(x_i) > k \prod_{i=1}^{n} f_{\theta_0}(x_i) \iff g_{\theta_1}(x_1, \ldots, x_n) > k \prod_{i=1}^{n} g_{\theta_0}(T(x_1, \ldots, x_n))$$

and the same holds for $<$. By Neyman-Pearson, the rest follows. ■

Theorem (Karlin-Rubin)**.** Let $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} f_\theta$, $\theta \in \mathbb{R}$, $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$, T a sufficient statistic of θ. Suppose $\{f_\theta : \theta \in D\}$ has strict MLR. Then, for any $t_0 > 0$, the test with $R = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) > t_0\}$ is an UMP level α test.

Definition. Let $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} f_\theta$, $\theta \in D \subseteq \mathbb{R}$. Suppose T is a sufficient statistic. $\{f_\theta : \theta \in D\}$ has strict monotone likelihood ratio if for all $\theta' > 0$,

$$t \mapsto \frac{g_{\theta'}(t)}{g_\theta(t)}$$

is increasing on $\{t \in \mathbb{R} : g_\theta(t) > 0 \text{ and } g_{\theta'}(t) > 0\}$.

10 Interval estimation

Definition. An interval estimate of a real-valued parameter θ is any pair of functions $L(x_1, \ldots, x_n)$ and $H(x_1, \ldots, x_n)$ satisfying $L(x_1, \ldots, x_n) \leq U(x_1, \ldots, x_n)$ for all $(x_1, \ldots, x_n)$. The interval $[L(X_1, \ldots, X_n), H(X_1, \ldots, X_n)]$ is an interval estimator. If either $L \equiv -\infty$ or $H \equiv \infty$, then it's a one-sided estimator.

Definition. For an interval estimator $[L(X_1, \ldots, X_n), U(X_1, \ldots, X_n)]$ of a parameter θ, the coverage probability is $\mathbb{P}_\theta(L(X_1, \ldots, X_n) \leq \theta \leq U(X_1, \ldots, X_n))$.

Definition. The confidence coefficient is $\inf_{\theta \in D} \mathbb{P}_\theta(L(X_1, \ldots, X_n) \leq \theta \leq U(X_1, \ldots, X_n))$.

Definition. Call $[L(X_1, \ldots, X_n), U(X_1, \ldots, X_n)]$ a $(1-\alpha)$-confidence interval estimate if its confidence coefficient is $1-\alpha$. Call a $2^{\mathbb{R}}$-valued function $C(x_1, \ldots, x_n)$ a $(1-\alpha)$-confidence set estimate if $\inf_{\theta \in D} \mathbb{P}_\theta(C(X_1, \ldots, X_n) \ni \theta) \geq 1 - \alpha$.

10.1 Inverting test statistic

Theorem. For each $\theta_0 \in D$, let $A(\theta_0)$ be the acceptance region of a level α test of $H_0 : \theta = \theta_0$. For each $(x_1, \ldots, x_n)$, define

$$C(x_1, \ldots, x_n) = \{\theta_0 \in D : (x_1, \ldots, x_n) \in A(\theta_0)\}$$

Then, $C(x_1, \dots, x_n)$ is a $(1-\alpha)$-confidence set. Conversely, let $C(X_1, \dots, X_n)$ be a $(1-\alpha)$-confidence set. For any $\theta_0 \in D$, define $A(\theta_0) = \{(x_1, \dots, x_n) : \theta_0 \in C(X_1, \dots, X_n)\}$. Then, $A(\theta_0)$ is the acceptance region of a level α test of $H_0 : \theta = \theta_0$.

10.2 Pivotal quantities

Definition. A RV $Q(X_1, \dots, X_n; \theta)$ is a pivotal quantity if the distribution of $Q(X_1, \dots, X_n; \theta)$ is independent of θ.

With a pivot $Q(x_1, \dots, x_n; \theta)$, find $-\infty \leq a \leq b \leq \infty$ such that

$$\mathbb{P}_\theta(a \leq Q \leq b) = 1 - \alpha, \forall Q \in D$$

By the test inversion theorem, $C(x_1, \dots, x_n) = \{\theta_0 \in D : a \leq Q(x_1, \dots, x_n; \theta) \leq b\}$ is a $(1-\alpha)$-confidence set.

11 Asymptotic evaluations

11.1 Point estimators

Definition. A sequence of estimators $W_n = W_n(X_1, \dots, X_n)$ is a weakly consistent sequence of estimators of a parameter $\theta_i \in \mathbb{R}$ if for any $\varepsilon > 0$ and every $\theta_i \in D$, $\lim_{n\to\infty} \mathbb{P}_{\theta_i}(|W_n - \theta_i| < \varepsilon) = 1$.

Theorem (10.1.3)**.** Suppose $(W_n)_{n\in\mathbb{N}}$ satisfies $\lim_{n\to\infty} \mathrm{Var}_\theta(W_n) = 0$ and $\lim_{n\to\infty} \mathrm{Bias}_\theta(W_n) = 0$ for every $\theta \in d$. Then, $(W_n)_{n\in\mathbb{N}}$ is consistent.

Theorem. Let $(W_n)_{n\in\mathbb{N}}$ be consistent. Consider $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} a_n = 1$ and $\lim_{n\to\infty} b_n = 0$. Let $U_n = a_n W_n + b_n$. Then $(U_n)_{n\in\mathbb{N}}$ is also consistent.

Theorem. Let $(X_1, \dots, X_n) \overset{\text{iid}}{\sim} f_\theta$ and $L(\theta|x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$. Let $\hat{\theta}$ be the MLE of θ. Let $\tau(\theta)$ be $\mathbb{R}$-valued continuous function of θ. Under suitable regularity conditions, we have

$$\lim_{n\to\infty} \mathbb{P}_\theta(|\tau(\hat{\theta}(x_1, \dots, x_n)) - \tau(\theta)| \geq \varepsilon) = 0$$

That is, $(\tau(\hat{\theta}_n(X_1, \dots, X_n)))_{n\in\mathbb{N}}$ is consistent of $\tau(\theta)$.

Definition. $(W_n)_{n\in\mathbb{N}}$ is asymptotically efficient for $\tau(\theta)$ if $\sqrt{n}(W_n - \tau(\theta)) \xrightarrow{D} N(0, v(\theta))$ and

$$v(\theta) = \frac{\left(\frac{\partial}{\partial\theta}\tau(\theta)\right)^2}{\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f_\theta(x)\right)\right)^2}$$

which is the C-R lower bound.

Theorem. Let $(X_1, \ldots, X_n) \overset{\text{iid}}{\sim} f_\theta$, $\theta \in D \subseteq \mathbb{R}$. Let $\hat{\theta}$ be the MLE of θ and $\tau(\theta)$ be continuous. Under suitable regularity conditions,

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{D} N(0, v(\theta))$$

where $v(\theta)$ is the C-R lower bound.