Statistical Theory Notes

June 29, 2019

Contents

1 Point Estimation

Problem. *From the observed data, choose a plausible value for unknown θ, or $\psi(\theta)$ for some known ψ.*

1.1 Consistency

Definition 1.1. A sequence of estimators T_n based on a sample $X_1, \dots, X_n$ is said to be consistent of $\psi(\theta)$ if

$$T_n \xrightarrow{\mathbb{P}} \psi(\theta)$$

for each $\theta \in \Theta$.

T_n is called a_n-consistent if $a_n(T_n - \psi(\theta)) = o_p(1)$.

Proposition 1.2. *If $\mathbf{E}T_n \to \psi(\theta)$ and $\mathrm{Var}T_n \to \psi(\theta)$, then T_n is consistent for $\psi(\theta)$.*

1.2 Sufficient statistics and minimal sufficient statistics

Definition 1.3. Let $X_1, \dots, X_n \overset{iid}{\sim} F_\theta,\ \theta \in \Theta$. A statistic $T(X_1, \dots, X_n)$ is sufficient for θ if the distribution of $X|T = t$ does not depend on θ for any t.

Example 1. Let $X_i \overset{iid}{\sim} N(\theta, 1)$. Let $U_{n\times n}$ be an orthogonal matrix s.t. the first row is $u_1 = \frac{1}{\sqrt{n}}(1, \dots, 1)$. If $Y = UX$, then

$$Y_j \sim N(\sqrt{n}\theta u_j^T u_1, 1).$$

So $Y_1 = \sqrt{n}\bar{X}$ is sufficient; however, $Y_2, \dots, Y_n \overset{iid}{\sim} N(0, 1)$ contain no information about θ

Theorem 1.4. *Let $X_1, \dots, X_n \overset{iid}{\sim} f_\theta,\ \theta \in \Theta$. $T(X)$ is sufficient for θ if and only if there are non-negative functions h and g s.t.*

$$f_\theta(x_1, \dots, x_n) = h(x_1, \dots, x_n)g(T(X); \theta).$$

Remark.

- **Invariance.**

  If T is sufficient for θ, and f is one-to-one, then $f(T)$ is also sufficient.

Example 2. $X_1, \dots, X_n \overset{iid}{\sim} U(\theta_1, \theta_2),\ \theta_2 > \theta_1,\ \theta_j \in \mathbb{R}$.

$$\begin{aligned} f_\theta(x_1, \dots, x_n) &= \prod_i \frac{\mathbf{1}(\theta_1 < x_i < \theta_2)}{\theta_2 - \theta_1} \\ &= (\theta_2 - \theta_1)^{-n} \cdot \mathbf{1}(\theta_1 < x_{(1)})\mathbf{1}(x_{(n)} < \theta_2) \end{aligned}$$

$\implies \quad T(X) = (X_{(1)}, X_{(n)})$.

Example 3. $X_1, \dots, X_n \overset{iid}{\sim} U(-\theta, \theta),\ \theta > 0$. (so $(X_{(1)}, X_{(n)})$ is sufficient)

$$\begin{aligned} f_\theta(x_1, \dots, x_n) &= \prod_i \frac{\mathbf{1}(-\theta < x_i < \theta)}{2\theta} \\ &= (2\theta)^{-n} \cdot \mathbf{1}(\max(-x_{(1)}, x_{(1)}) < \theta) \end{aligned}$$

$\implies \quad T(X) = \max(-X_{(1)}, X_{(1)})$.

Definition 1.5. $T(X)$ is called minimal sufficient if

a) it is sufficient, and

b) If $S(X)$ is sufficient, $\exists w$ s.t. $T(X) = w \circ S(X)$

Theorem 1.6. *Let $A = \{(x,y) \mid \exists k(x,y) \neq 0 \text{ s.t. } f_\theta(x) = k(x,y) f_\theta(y)\ \forall \theta \in \Theta\}$, and T is sufficient. T is minimal sufficient if*

$$(x,y) \in A \implies T(x) = T(y).$$

Remark. Usually, we can follow the recipe below to show the minimal sufficiency of T:

1. Show T is sufficient.
2. Check $(x,y) \in A \implies T(x) = T(y)$;
3. or if $\{x : f_\theta(x) \geq 0\}$ doesn't depend on θ, check $f_\theta(x)/f_\theta(y)$ indep. of $\theta \implies T(x) = T(y)$

Example 4. $X_1, \ldots, X_n \overset{iid}{\sim} U(\theta, -\theta),\ \theta > 0$. Notice that $f_\theta(x) = \theta^{-n} \mathbf{1}(x_{(n)<\theta})$.
$\implies$ $T(X) = X_{(n)}$ is sufficient.
$\implies$ Taking $(x,y) \in A$, we have, for some $k(x,y) \neq 0$,

$$\theta^{-n} \mathbf{1}(x_{(n)<\theta}) = k(x,y) \theta^{-n} \mathbf{1}(y_{(n)<\theta}).$$

$\implies$ $T(x) = T(y)$. Thus, T is minimal sufficient.

Example 5. $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$. Obviously, $T = \sum X_i$ is sufficient.
If we assume

$$\frac{f_\theta(x)}{f_\theta(y)} = \exp\left(\frac{1}{2}[\sum y_i^2 - \sum x_i^2]\right) \exp\left(\mu[T(x) - T(y)]\right)$$

is indep. of μ, we must have $T(x) = T(y)$. By Theorem 1.6, T is minimal.

1.3 Complete statistics

Definition 1.7.

- Let $\mathcal{F} = \{f_\theta \mid \theta \in \Theta\}$ be a family of pmfs or pdfs. Then $\mathcal{F}$ is complete if

$$\mathbf{E}_\theta g(X) = 0\ \forall \theta \implies \mathbb{P}_\theta(g(X) = 0) = 1\ \forall \theta.$$

- A statistic T is called complete if the induced family of distributions for T is complete, i.e.

$$\mathbf{E}_\theta g(T(X)) = 0\ \forall \theta \implies \mathbb{P}_\theta(g(T(X)) = 0) = 1\ \forall \theta.$$

Example 6. $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bin}(1,p),\ 0 < p < 1$. Consider $T(X) = \sum_{i=1}^n X_i$. Then

$$\begin{aligned}
\mathbf{E}_p g(T) &= \sum_{t=0}^n \mathbb{P}(T = t) \cdot g(t) \\
&= \sum_{t=0}^n g(t) \binom{n}{t} p^t (1-p)^{n-t} \\
&= (1-p)^n \sum_{t=0}^n g(t) \binom{n}{t} (\frac{p}{1-p})^t
\end{aligned}$$

is a polynonial in $\frac{p}{1-p}$. Thus,

$$\mathbf{E}_p g(T) = 0\ \forall p \implies g(t)\binom{n}{t} = 0\ \forall t.$$

It means $g(t) = 0$ for $t \in \{0, \ldots, n\}$. T is a complete statistic.

Example 7 (not complete)**.** $X \sim \text{Bin}(n,p),\ p \in \{1/4, 3/4\}$, is not a complete family.
Construct g s.t. the definition of compelteness is not satisfied.

$$g(X) = (X - \frac{n}{4})(X - \frac{3n}{4}) - \frac{3n}{16}.$$

1.4 Ancillary statistics

Definition 1.8. A statistic A is called ancillary if its distribution doesn't depend on θ.

Remark. Usually, we have two ways to prove something is ancillary:

1. Compute its distribution directly.
2. Check if $\mathbb{P}_\theta(A(X) \in B)$ is a function of θ.

Example 8. $X_i \overset{iid}{\sim} N(\mu, \sigma_0^2)$. σ_0^2 known. We know that $S^2 \sim \frac{\sigma_0^2}{n-1}\chi^2_{n-1}$. It doesn't depend on θ. Moreover, by the Basu's theorem, $\bar{X}$ is independent of S^2.

Example 9. Let f be a pdf, and for $\theta \in \mathbb{R}$, set $f_\theta(x) = f(x-\theta)$ (location family).

If $X_i \overset{iid}{\sim} f_\theta$, $X_i - \bar{X}$ are all ancillary for θ. It is because $X_i - \bar{X}$ is location invariant. Let S be location invariant; that is

$$S(x) = S(x+c),$$

then we have

$$\mathbb{P}_\theta(S(\underline{X}) \in B) = \mathbb{P}_\theta(S(\underline{X} - \theta) \in B).$$

Notice that $\underline{X} - \theta$ doesn't depend on θ.

Example 10. Let f be a pdf, and for $\theta \in \mathbb{R}$, set $f_\theta(x) = \frac{1}{\theta} f(\frac{x}{\theta})$, $\theta > 0$ (location-scale family).

If $X_i \overset{iid}{\sim} \theta$, then $\frac{\bar{X}}{S}$ is ancillary for θ. It is because this statistic is location-scale invariant! So we don't need to compute its distribution.

We can summarize the two examples above as follow:

$f(X-\theta)$	Location Family	$X - \theta \sim f$	Location Invariant
$\frac{1}{\theta} f(\frac{X}{\theta})$	Scale Family	$\frac{X}{\theta} \sim f$	Scale Invariant
$\frac{1}{\sigma} f(\frac{X-\mu}{\sigma})$	Location-Scale Family	$\frac{X-\mu}{\sigma} \sim f$	Location-Scale Invariant

The following theorem sometimes could be used to prove independence.

Theorem 1.9 (Basu)**.** *If S is complete and sufficient, S is independent of any ancillary statistics.*

Proof. Let A be ancillary and $Y = \mathbf{E}_\theta(\mathbf{1}(A \le a) | S)$. To show that A is independent of S, it suffices to show

$$Y = \mathbf{E}_\theta(\mathbf{1}(A \le a)).$$

Clearly, $\mathbf{E}_\theta Y = \mathbb{P}(A \le a)$. So $\mathbf{E}_\theta(Y - \mathbb{P}(A \le a)) = 0$ holds for all θ.
By completeness, $Y = \mathbb{P}(A \le a)$ almost surely; that is A and S are independent. □

1.5 Unbaised estimation

Definition 1.10. Let $\mathcal{F}_\theta$ be a family of distributions, and φ be a function of θ.

- A statisticc T is unbiased for θ if
$$\mathbf{E}_\theta T = \varphi(\theta), \ \forall \theta \in \Theta.$$
- Any function φ is called estimable if there always exists an unbiased estimator.

Remark.

- Unbiased estimates may not exist.
- If T is unbiased for θ, $g(T)$ may not be so for $g(\theta)$.
- Usually, we take $\varphi = \mathbf{Id}_\Theta$.

1.6 Uniform minimal variance unbaised estimation (UMVUE)

Definition 1.11. Let $\mathcal{U}$ be the set of all unbaised estimators of $\varphi(\theta)$ that have finite variance. $T \in \mathcal{U}$ is called uniformly minimum variance unbiaed estimator (UMVUE) of θ if

$$\text{Var}_\theta T \leq \text{Var}_\theta S, \quad \forall S \in \mathcal{U},\ \forall \theta \in \Theta.$$

Remark. **Invariance**.

- If T_i is the UMVUE for ψ_i, then $\sum_{i=1}^n \lambda_i T_i$ is the UMVUE for $\sum_{i=1}^n \lambda_i \psi_i$.
- Let T_n be a sequence of UMVUEs. If $T_n \xrightarrow{L^2} T$, then T is also a UMVUE.

Theorem 1.12. *Let $\mathcal{U}_0 = \{v : \mathbf{E}_\theta(v) = 0$ and $\text{Var}_\theta(v) < \infty\}$. Then $T \in \mathcal{U}$ is the UMVUE of $\varphi(\theta)$ if and only if $\mathbf{E}(Tv) = 0$ for all θ and for all $v \in \mathcal{U}_0$.*

Theorem 1.13 (Rao-Blackwell)**.** *Let $\mathcal{F}_\theta$ be a paremetric family of distributions, and $h \in \mathcal{U}$ an unbiaed estimator of $\psi(\theta)$. If T is sufficient for θ, then $\mathbf{E}(h|T) \in \mathcal{U}$ and*

$$\text{Var}_\theta\Big(\mathbf{E}(h|T)\Big) \leq \text{Var}_\theta(h), \quad \forall \theta \in \Theta$$

with equality if and only if h is a function of T.

Theorem 1.14 (Lehmann-Scheffé)**.** *Suppose T is complete and sufficient. If there exists h s.t.*

$$\mathbf{E}_\theta(h) = \psi(\theta) \text{ and } \text{Var}_\theta(h) < \infty,$$

then $\mathbf{E}_\theta(h|T)$ is the UMVUE for ψ.

Remark.

- In Rao-Blackwell, we only require the sufficiency of T; however, in Lehmann-Scheffé, we require both of the completeness and sufficiency of T.
- By LS, we can follow this recipe to find the UMVUE:
  1. Find a complete sufficient statistic T and a unbiased estimate h.
  2. Compute $\mathbf{E}_\theta(h|T)$.

Example 11. $X_i \overset{iid}{\sim} \text{Poisson}(\lambda)$. Obviously, $\bar{X}$ is complete and sufficient for $\lambda \in (0, \infty)$.

- Since $X_i \in \mathcal{U}$, and $T = \bar{X}$ is complete and sufficient, by LS,

$$\mathbf{E}(X_i|\bar{X}) = \bar{X}$$

  is the UMVUE for λ. (Recall that $X_i | \sum_{j=1}^n X_j \sim \text{Bin}(n\bar{X}, \frac{1}{n})$.)
- Or we can directly choose $h = \bar{X}$. Notice that $\mathbf{E}_\lambda(\bar{X}) = \lambda$, so $\bar{X}$ is the UMVUE for λ.

Example 12. $X_i \overset{iid}{\sim} \text{Exp}(\lambda)$. Find the UMVUE of $\psi(\lambda) = \mathbb{P}_\lambda(X_1 \leq 1)$. A complete sufficient statistic is $T = \sum_{i=1}^n X_i$. And let

$$h(\underline{X}) = \mathbf{1}(X_j \leq 1)$$

be a unbiased estimator for $\psi(\lambda)$. Therefore, the UMVUE of $\psi(\lambda)$ is

$$\begin{aligned}
\mathbf{E}(h(X)|T) &= \mathbb{P}(X_j \leq 2 | \sum_{i=1}^n X_i = t) \\
&= \mathbb{P}(\frac{X_j}{\sum_{i=1}^n X_i} \leq \frac{2}{t} | \sum_{i=1}^n X_i = t) \\
&= \mathbb{P}(\frac{X_j}{\sum_{i=1}^n X_i} \leq \frac{2}{t}) \\
&= \mathbb{P}(Z \leq \frac{2}{t})
\end{aligned}$$

where $Z \sim \text{Beta}(1, n-1)$. Finally, we get the UMVUE of $\psi(\lambda)$:

$$\mathbf{E}(h(X)|T) = \begin{cases} 1 & T \leq 1; \\ 1 - (1 - \frac{1}{T})^{n-1} & T > 1. \end{cases}$$

Proposition 1.15. *If T is complete and sufficient, and $\mathbf{E}_\theta(T^2)$ is finite for all θ, then T is minimal sufficient.*

Proof. By LS, T is UMVUE for $\mathbf{E}_\theta(T)$. Let S be any sufficient statistic, and define

$$h(S) = \mathbf{E}_\theta(T|S).$$

Obviously, it is unbiased for $\mathbf{E}_\theta(T)$ and satisfies

$$\text{Var}_\theta(h(S)) \leq \text{Var}_\theta(T)$$

by Rao-Blackwell. However, as T is the UMVUE, by the uniqueness, $h(S) = T$ almost surely; i.e. T is a function of S. By the definiton, T is minimal sufficient. □

1.7 Lower bound for variance in unbiased estimation

Definition 1.16. Let $\mathcal{F}_\Theta$ be a parametric familiy of distributions for a RV X.

- The score function is defined as

$$\frac{\partial}{\partial\theta} \log f_\theta(x).$$

- The Fisher information is defined as the variance of the score function:

$$I(\theta) = \text{Var}_\theta\Big(\frac{\partial}{\partial\theta} \log f_\theta(x)\Big).$$

Remark. If $X_i \overset{iid}{\sim} f_\theta$, let $I_n(\theta)$ denote the FI for $\prod f_\theta(x)$.

Proposition 1.17 (Properties of Fisher information)**.** *Under regularity conditions, we have:*

- $I(\theta) = \mathbf{E}_\theta\Big((\frac{\partial}{\partial\theta} \log f_\theta(x))^2\Big) = -\mathbf{E}_\theta\Big(\frac{\partial^2}{\partial\theta^2} \log f_\theta(x)\Big)$;
- $I_n(\theta) = nI_1(\theta)$.

Theorem 1.18. *If $\Theta \subset \mathbb{R}$ is an open interval and*

(i) $s = \{x : f_\theta(x) > 0\}$ *is indep. of* θ

(ii) The score exists and is finite for all $x \in s,\ \theta \in \Theta$.

(iii) $\exists\ \mathbf{E}_\theta(h(x))$ *for all* θ *implies:*

$$\int h(X) \frac{\partial}{\partial\theta} f_\theta(x)\ \mathrm{d}x = \frac{\partial}{\partial\theta} \int h(x) f_\theta(x)\ \mathrm{d}x.$$

then if T is an unbiased estimator of $\varphi(\theta)$, and $0 < I(t) < \infty$,

$$\text{Var}_\theta(T) \geq \frac{[\varphi'(\theta)]^2}{I(\theta)}.$$

Remark.

- The lower bound is attained if and only if $T(\underline{X})$ and $\frac{\partial}{\partial\theta} \log f(\underline{X})$ are perfectly correlated, that is,

$$T(X) - \psi(\theta) = k(\theta) \frac{\partial}{\partial\theta} \log f(\underline{X})$$

  for some function $k(\theta)$.

- If $\theta \in \mathbb{R}^k$,

$$\mathrm{Var}_\theta(T(X)) \geq \psi'(\theta)^T I(\theta)^{-1}\psi(\theta).$$

- Suppose $\eta = \eta(\theta)$ is strictly monotonic, then

$$I(\eta) = \mathrm{Var}(\frac{\partial}{\partial \eta}\log f_\eta(X)) = \mathrm{Var}(\frac{\partial}{\partial \theta}\cdot\frac{\partial \theta}{\partial \eta}\cdot \log f_\theta(X)) = I(\theta)\cdot(\frac{\mathrm{d}\theta}{\mathrm{d}\eta})^2.$$

  and letting $\tilde{\psi}(\eta) = \psi(\theta)$,

$$\frac{[\frac{\mathrm{d}}{\mathrm{d}\theta}\psi(\theta)]^2}{I(\theta)} = \frac{[\frac{\mathrm{d}}{\mathrm{d}\eta}\frac{\mathrm{d}\eta}{\mathrm{d}\theta}\psi(\theta)]^2}{I(\eta)/(\frac{\mathrm{d}\theta}{\mathrm{d}\eta})^2} = \frac{[\frac{\mathrm{d}}{\mathrm{d}\eta}\tilde{\psi}(\eta)]^2}{I(\eta)}.$$

- Note: Scale families with bouded support and $U(0, \theta)$ don't satisfy the conditions.
- If a unbiased estimator attains the lower bound of variance, then it is UMVUE!

Example 13. $X_i \overset{iid}{\sim} \mathrm{Exp}(\lambda)$. Then

$$f_\lambda(x) = \frac{1}{\lambda^n}e^{-T(x)/\lambda}\mathbf{1}(X_{(1)} > 0).$$

- **Compute the Fisher information for λ**
  $\Longrightarrow$ $T(X) = \sum_{i=1}^n X_i$, $\frac{\partial}{\partial\lambda}\log f_\lambda(x) = T(X)/\lambda^2 - n = n\bar{X}/\lambda^2 - n$.
  $\Longrightarrow$ $I(\lambda) = \mathrm{Var}_\lambda(T(X)/\lambda^2) = \frac{1}{\lambda^4}n\lambda^2 = \frac{n}{\lambda^2}$.
- **Lower bound for variance of λ**
  $\Longrightarrow$ If $S(X)$ is unbiased for λ, $\mathrm{Var}S(X) \geq \frac{1}{I(\lambda)} = \frac{\lambda^2}{n} = \mathrm{Var}_\lambda(\bar{X})$.
- **Lower bound for variance of $\psi(\lambda) = \mathbb{P}_\lambda(X_1 \leq 1)$**
  For $\psi(\lambda) = \mathbb{P}_\lambda(X_1 \leq 1)$, $\psi'(\lambda) = -e^{-1/\lambda}/\lambda^2$
  $\Longrightarrow$ If $S(X)$ is unbiased for $\psi(\lambda)$, $\mathrm{Var}S(X) \geq \frac{[\psi'(\lambda)]^2}{I(\lambda)} = e^{-2/\lambda}/n\lambda^2$.

Theorem 1.19. *Assume $\theta \mapsto f_\theta$ is injective, and T is unbiased for $\psi(\theta)$, and $\mathbf{E}_\theta(T(X)) < \infty$. Let $\theta \in \Theta$ and*

$$S_\theta = \Big\{\varphi \in \Theta : \{x : f_\varphi(x) > 0\} \subset \{x : f_\theta(x) > 0\}\Big\}\backslash\Big\{\theta\Big\}.$$

Then

$$\mathrm{Var}_\theta(T(X)) \geq \sup_{\varphi\in S_\theta}\frac{[\psi(\varphi) - \psi(\theta)]^2}{\mathrm{Var}_\theta(\frac{f_\varphi(x)}{f_\theta(x)})}.$$

Example 14. $X \sim U(0, \theta)$. Then $S_\theta = (0, \theta)$. And $2X$ is the UMVUE for θ with the variance

$$\mathrm{Var}(2X) = 4\mathrm{Var}X = \frac{\theta^2}{3}.$$

Notice that $\frac{f_\varphi}{f_\theta} = (\frac{\theta}{\varphi})\cdot\mathbf{1}(0, \varphi)$ for $\varphi \in S_\theta = (0, \theta)$. Then

$$\begin{aligned}\sup_{0<\varphi<\theta}\frac{[\varphi - \theta]^2}{\mathrm{Var}_\theta[(\frac{\theta}{\varphi})\cdot\mathbf{1}(0, \varphi)]} &= \sup_{0<\varphi<\theta}\frac{(\varphi-\theta)^2}{\frac{\theta^2}{\varphi^2}\cdot\frac{\varphi}{\theta}\cdot(1-\frac{\varphi}{\theta})}\\ &= \sup_{0<\varphi<\theta}\frac{(\varphi-\theta)^2}{\frac{\theta}{\varphi}-1}\\ &= \frac{\theta^2}{4}\end{aligned}$$

Although $2X$ is the UMVUE, $\mathrm{Var}(2X) > \frac{\theta^2}{4}$.

1.8 Exponential family: Part I

Definition 1.20. Let $\{f_\theta\}$ be a family of PDFs with

$$f_\theta(x) = h(x)\exp\Big\{\sum_{j=1}^{k} Q_i(\theta)T_j(x) + D(\theta)\Big\}.$$

Theorem 1.21 (Sufficient and complete statistics)**.** *Let $\mathcal{F}_\theta = \{f_\theta : \theta \in \Theta\}$ be a k-parameter exponential family on $\mathbb{R}^n$, where $\Theta \subset \mathbb{R}^k$ is an interval and $k \le n$. Then*

a) T is sufficient.

b) If the range of $(Q_1, \dots, Q_k)$ contains an open set in $\mathbb{R}^k$, T is complete.

The theorem above gives a simple way to find sufficient statistics (see the example below); however, T may not be complete in general.

Example 15. $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

We re-write its pdf as the form of exponential family:

$$\begin{aligned} f_{\mu,\sigma^2}(x) &= (2\pi)^{-n/2}\exp\Big\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{n}{2}\log(\sigma^2)\Big\} \\ &= (2\pi)^{-n/2}\exp\Big\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n}x_i - \frac{n\mu^2}{2\sigma^2} - \frac{n}{2}\log(\sigma^2)\Big\} \end{aligned}$$

Thus, $T_1(X) = \sum_{i=1}^n X_i$, $T_2(X) = \sum_{i=1}^n X_i^2$, and (T_1, T_2) is sufficient.

Moreover, we are interested in its completeness. Notice that $Q_i(\mu,\sigma^2) = \frac{\mu}{\sigma^2}$ and $Q_2(\mu,\sigma^2) = -\frac{1}{2\sigma^2}$. The range of $Q = (Q_1, Q_2)$ is $\mathbb{R} \times \mathbb{R}^-$, and it contains an open set in $\mathbb{R}^2$. So T is complete.

Example 16. $X_i \overset{iid}{\sim} N(\theta, \theta^2)$, $\theta > 0$.

Obviously, (T_1, T_2) is still sufficient for θ, since

$$f_\theta(x) = (2\pi)^{-n/2}\exp\Big\{\frac{1}{\theta}T_1(x) - \frac{1}{2\theta^2}T_2(x) + D(\theta)\Big\}.$$

However, T is not complete.

Notice that $T_1 \sim N(n\theta, n\theta^2) \implies \mathbf{E}_\theta T_1^2(X) = n(n+1)\theta^2$. Similarly, $\mathbf{E}_\theta T_2(X) = 2n\theta^2$. So

$$\mathbf{E}_\theta\Big(2T_1^2(X) - (n+1)T_2(X)\Big) = 0, \forall\theta.$$

Thus, we can construct $g : (t_1, t_2) \mapsto 2t_1^2 - (n+1)t_2$ that is not identically 0 on $\mathbb{R} \times \mathbb{R}^+$.

1.9 Methods of moment

Definition 1.22. The method of moments estimator of $\theta = h(m_1, \dots, m_k)$ is

$$T_h = h(\hat{m}_1, \dots, \hat{m}_k)$$

where $\hat{m}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$.

Remark. Note: $m_n := \mathbf{E}X^n$. And $m_{n_1,\dots,n_k} := \mathbf{E}X_1^{n_1}\dots X_k^{n_k}$.

Example 17. $X_i \overset{iid}{\sim} \text{Bin}(m, p)$. $h(p) = \mathbb{P}_p(X_1 = 2) = \binom{m}{2}\frac{(mp)^2}{m^2}(1 - \frac{mp}{m})^{m-2}$.

The method of moments estimator is

$$T_h(X) = \binom{m}{2}\frac{(\bar{X}^2)^2}{m^2}(1 - \frac{\bar{X}}{m})^{m-2}.$$

Example 18. $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$. $h(\mu, \sigma^2) = \begin{pmatrix}\mu \\ \sigma^2\end{pmatrix} = \begin{pmatrix}\mu \\ \mathbf{E}(X^2) - \mu^2\end{pmatrix}$.

The method of moments estimator is

$$T_h(X) = \begin{pmatrix}\bar{X} \\ \frac{1}{n}\sum X_i^2 - \bar{X}^2\end{pmatrix} = \begin{pmatrix}\bar{X} \\ \frac{n-1}{n}S^2\end{pmatrix}.$$

2 Maximum likelihood

2.1 Maximum likelihood estimators (MLE)

Definition 2.1. Let $\mathcal{F}_\Theta$ be a family of pmfs/pdfs.

- The likelihood function is
$$L(\theta; x) = f_\theta(x), \quad \theta \in \Theta.$$
- The log-likelihood is
$$l(\theta; x) = \log L(\theta; x).$$

Remark. If $X_i \overset{iid}{\sim} f_\theta$, then $L(\theta; X) = \prod_{i=1}^n f_\theta(X_i)$ and $l(\theta; X) = \sum_{i=1}^n \log f_\theta(X_i)$.

Definition 2.2. If $X_i \overset{iid}{\sim} f_\theta$ and $X = x$ is observed.
$$\hat{\theta}(x) = \arg\max_{\theta\in\Theta} L(\theta; x),$$
if it exists, is called a maximum likelihood estimate of θ.

Remark. By the strict monotonity of log, we have
$$\hat{\theta}(x) = \arg\max_{\theta\in\Theta} l(\theta; x) = \arg\max_{\theta\in\Theta} \sum_{i=1}^n f_\theta(x_i).$$

Example 19. $X_i \overset{iid}{\sim} \text{Poisson}(\theta),\ \Theta = (0, \infty)$.
Compute its likelihood function:
$$L(\theta; x) = e^{-n\theta} \cdot \frac{e^{(\log\theta)\cdot\sum x_i}}{\prod x_i!}$$
$$l(\theta; x) = (\sum x_i)\log\theta - n\theta - \sum \log(x_i!)$$

Compute its partial derivatives:
$$\frac{\partial}{\partial\theta} = \frac{\sum x_i}{\theta} - n = 0 \implies \theta = \bar{x}$$
$$\frac{\partial^2}{\partial\theta^2} = -\frac{\sum x_i}{\theta^2} \le 0$$

Thus, $\hat{\theta}(x) = \bar{x}$ is the MLE except when $\bar{x} = 0$; because when $\bar{x} = 0$, $\theta = 0 \notin \Theta$.

Example 20. $X_i \overset{iid}{\sim} U(\theta_1, \theta_2)$.
Compute its likelihood function:
$$L(\theta; x) = \prod f_i(x_i) = \prod \left(\frac{1}{\theta_2 - \theta_1}\mathbf{1}(\theta_1 \le x_i \le \theta_2)\right)$$
$$= \begin{cases} 0 & \theta_1 \ge x_{(1)} \text{ or } \theta_2 < x_{(n)} \\ \frac{1}{(\theta_2-\theta_1)^n} & \text{o.w.} \end{cases}$$

Notice: when $\theta_1 \le x_{(1)}$ and $\theta_2 \ge x_{(n)}$,
$$(\theta_2 - \theta_1)\downarrow \implies L(\theta; x)\uparrow.$$

Therefore, $(\hat{\theta}_1, \hat{\theta}_2) = (x_{(1)}, x_{(n)})$ is the MLE.

Proposition 2.3. *Let T be sufficient for θ for a family of pdfs/pmfs. If an MLE exists, there is an MLE such that $\hat{\theta} = g(T)$.*

Proof. Compute its likelihood function:

$$\begin{aligned} L(\theta; x) &= f_\theta(x) \\ \text{(By Thm 1.4.)} \quad &= h(x) g_\theta(T(x)) \end{aligned}$$

Assume θ^* maximizes $L(\theta; x)$. It also maximizes $w_x(\theta) = g_\theta(T(x))$.

Define $S(x) = \{\theta^* \in \Theta : g_{\theta^*}(T(x)) = \max_\theta g_\theta(T(x))\}$. (Note: the maxima may not be unique.)

Notice that $T(x) = T(y) \implies S(x) = S(y)$, so we can choose $\hat{\theta}(x) \in S(x)$ such that it is a function of $T(x)$. □

2.2 Uniqueness and existence of MLEs

The following example shows: (1) MLE may not be unique. (2) MLE could be a function of T; however, some MLEs may not be a function of T.

Example 21. $X_i \overset{iid}{\sim} U(\theta - 1, \theta + 1)$.

Compute its likelihood function:

$$\begin{aligned} L(\theta; x) &= \frac{1}{2^n} \cdot \mathbf{1}(x_{(1)} \geq \theta - 1) \cdot \mathbf{1}(x_{(n)} \leq \theta + 1) \\ &= \frac{1}{2^n} \cdot \mathbf{1}(x_{(n)} - 1 \leq \theta \leq x_{(1)} + 1) \end{aligned}$$

$\implies$ any estimator $\hat{\theta}(x) \in [x_{(n)} - 1, x_{(1)} + 1]$ is an MLE. (not unique)

In particular,

$$\hat{\theta}(x) = \alpha(x_{(n)} - 1) + (1 - \alpha)(x_{(1)} + 1)$$

for $0 \leq \alpha \leq 1$ is an MLE that is a function of $T = (x_{(1)}, x_{(n)})$; however, so is

$$\sin^2(\bar{x})(x_{(n)} - 1) + \cos^2(\bar{x})(x_{(1)} + 1),$$

not a function of T.

Theorem 2.4.

- ***Existence***

  Suppose $l : \Theta \to \mathbb{R}$, Θ open in $\mathbb{R}^k$, is continuous. If $l(\theta; x) \to -\infty$ as $\theta \to \partial\Theta$, then

  $$\{\theta \in \Theta : l(\theta) = \max_{\theta \in \Theta} l(\theta)\} \neq \emptyset.$$

- ***Existence and uniqueness***

  Suppose $X \sim f_\theta$, $\theta \in \Theta \subset \mathbb{R}^k$ open set. If $l(\theta; x)$ is strictly convave, is continuous, and moreover, $l(\theta; x) \longrightarrow -\infty$ as $\theta \to \partial\Theta$, then the MLE exists and is unique.

2.3 Exponential family: Part II

Lemma 2.5. *Let $\mathcal{F}_\eta$ be a k-parameter exponential family in canonical parameter. The following statements are equivalent:*

a) The log-likelihood function $l(\eta; x)$ is strictly convave

b) $A(\eta)$ is strictly convex

c) $A''(\eta) = \mathrm{Var}(T) > 0$ (aka full rank).

Theorem 2.6. *Suppose $\mathcal{F}_\Theta$ is a k-parameter exponential family with*

$$f_\eta = h(x) \exp\left\{ \sum_{j=1}^k \eta_j T_j(x) - A(\eta) \right\}$$

such that Θ is open and $A''(\eta) > 0$. Let x be the deserved value and $t_0 = T(x) \in \mathbb{R}^k$.

a) If $\mathbb{P}_\eta(c^T T(x) > c^T t_0) > 0$ for all $c \neq 0$, $\eta \in \Theta$, then $\hat{\eta}$ exists, is unique, and satisfies

$$A'\big(\hat{\eta}(x)\big) = \mathbf{E}_{\hat{\eta}(x)}(T(x)) = t_0.$$

b) If $\exists c \neq 0$ such that $\mathbb{P}(c^T T(x) > c^T t_0) = 0$, there is no MLE.

Corollary 2.7. *Let C_T be the convex hull of the support of T. Then the MLE exists and is unique if and only if $t_0 \in C_T^\circ$.*

Corollary 2.8. *If T has a continuous distribution, the MLE exists and is unique.*

Corollary 2.9. *Let the exponential family be*

$$f_\theta(x) = h(x) \exp\Big\{\sum_{j=1}^{k} Q_j(\theta) T_j(x) - B(\theta)\Big\}.$$

If $\mathbf{E}_\theta T_j = T_j$ have a solution $\hat{\theta}(X) \in Q(\Theta)^\circ$, it is the unique MLE.

Example 22. $X \sim \mathrm{Bin}(n, \theta)$. Then $\hat{\theta} = \frac{X}{n}$ is the MLE unless $X = 0$.

Example 23. $X_i \overset{iid}{\sim} \mathrm{Gamma}(\alpha, \beta)$. The MLE exists and is unique.

2.4 Invariance

Theorem 2.10. *Let $\mathcal{F}_\theta$ be a family of pdfs/pmfs, $\theta \in \mathbb{R}^k$. If $\hat{\theta}$ is an MLE and $h : \mathbb{R}^k \to \mathbb{R}^p$ with $p \leq k$, then $h(\hat{\theta})$ is an MLE for $h(\theta)$.*

Example 24. $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ and $\sigma > 0$. Obviously, $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ are MLEs for μ and σ^2. We may be interested in the MLE of μ/σ.

Let $h : (x, y) \mapsto \frac{x}{y}$, then $h(\hat{\mu}, \hat{\sigma})$ is the MLE for $h(\mu, \sigma)$. Thus, the MLE for μ/σ is $\bar{X}/\hat{\sigma}$.

2.5 Asymptotic consistency and normality

Theorem 2.11 (Wald)**.** *Recall that $D(\theta_0, \theta) = \mathbf{E}_{\theta_0}(\log f_\theta(x))$. Suppose*

$$\sup_{\theta \in \Theta} \Big(\frac{1}{n}\sum_{i=1}^{n} \log f_\theta(x) - D(\theta_0, \theta)\Big) \xrightarrow[\theta_0]{\mathbb{P}} 0,$$

and for all $\epsilon > 0$,

$$\sup_{\theta : |\theta - \theta_0| \geq \epsilon} D(\theta_0, \theta) < D(\theta_0, \theta_0).$$

Then we have

$$\hat{\theta} \xrightarrow[\theta_0]{\mathbb{P}} \theta_0.$$

Remark. Generally, consistency of $\hat{\theta}$ can be found in other ways (e.g. continuous mapping theorem, WLLN).

The following theorem gives a sufficient conditions for a sequence of MLEs $\hat{\theta}_n$ based on a sample $X_1, \ldots, X_n \overset{iid}{\sim} f_\theta$ to be asymptotically normal. Let $\theta_0 \in \Theta$ be the true parameter.

Theorem 2.12. *If the following conditions hold*

(A1) The score function ψ is well-defined and $0 < I(\theta) < \infty$;

(A2) $\frac{\partial^2}{\partial \theta^2}\psi(x; \theta)$ is continuous;

(A3) For some ϵ, g such that $\mathbf{E}_{\theta_0} g(X) < \infty$,

$$\sup_{|\theta - \theta_0| \leq \epsilon} \Big|\frac{\partial^2}{\partial \theta^2}\psi(x; \theta)\Big| < g(x);$$

and $\hat{\theta}_n$ exists, is unique, and is consistent under H_0, then

$$\hat{\theta} = \theta_0 + \frac{1}{nI(\theta_0)} \sum_{i=1}^{n} \psi(X_i; \theta) + o_p(n^{-1/2}),$$

and

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow[\theta_0]{D} N(0, I^{-1}(\theta_0)).$$

Remark. For suitable h, we can also show AN of $h(\hat{\theta})$ using the delta-method.

Example 25. $X_i \overset{iid}{\sim} \text{Gamma}(\alpha, 1)$. The MLE $\hat{\alpha}$ is the solution to

$$\frac{n\Gamma'(\alpha)}{\Gamma(\alpha)} = \sum_{i=1}^{n} \log(X_i).$$

It can only be computed numerically. If we want to do inference for α, since

$$I(\alpha) = -\mathbf{E}_\alpha(\frac{\partial^2}{\partial\alpha^2} \log f_\alpha(x)) = \frac{\Gamma''(\alpha)\Gamma(\alpha) - \Gamma'(\alpha)^2}{\Gamma(\alpha)^2},$$

$$\sqrt{nI(\alpha)}(\hat{\alpha} - \alpha) \xrightarrow[\alpha]{D} N(0, 1).$$

Example 26. $X_i \overset{iid}{\sim} U(0, \theta)$. The conditons for AN do not hold. Its MLE is $\hat{\theta} = X_{(n)}$. So

$$n(\theta - \hat{\theta}) \xrightarrow{D} \text{Exp}(\theta).$$

3 Hypothesis Testing

3.1 Introduction to hypothesis testing

Definition 3.1. Let φ be a test, and $\beta_\varphi(\theta) = \mathbf{E}_\theta(\varphi(X))$.

- The size of a test φ is defined as

$$\sup_{\theta\in\Theta_0} \beta_\varphi(\theta) = \sup_{\theta\in\Theta_0} \mathbf{E}_\theta(\varphi(X)).$$

- Let φ be a test of size α. For any $\theta \in \Theta_1$, the power of φ for detecting θ is

$$\beta_\varphi(\theta) = \mathbf{E}_\theta(\varphi(X)) = \mathbb{P}_\theta(H_0 \text{ rejected}).$$

Remark. As a function of θ, β_φ is called the power function. If $\varphi(X) = \mathbf{1}(T(X) \in C)$, T is called a test statistic, and C is called the critical region.

The size is also called the Type I error; it represents the probability that H_0 is correct, but we reject it. The power is also called the Type II error; it represents the probability that H_0 is wrong, but we accept it.

Example 27. $X_i \overset{iid}{\sim} N(\mu, \sigma^s)$, $\mu \in \mu_0, \mu_1$ $(\mu_0 < \mu_1)$, and $\sigma^2 > 0$ known. $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1$.

Conider a rule $\varphi(\bar{X}) = \mathbf{1}(\bar{X} > k)$, for some k, corresponding to the critical region $c_k = \{X : \bar{X} > k\}$. Fix its size:

$$\beta_\varphi(\mu_0) = \mathbb{P}_{\mu_0}(\bar{X} > k) = 1 - \Phi(\frac{\sqrt{n}(k-\mu_0)}{\sigma}) = \alpha.;$$

so we take k s.t. $\frac{\sqrt{n}(k-\mu_0)}{\sigma} = \Phi^{-1}(1-\alpha) = z_{1-\alpha}$; i.e.

$$k = \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha},$$

leading the test

$$\varphi(\bar{X}) = \begin{cases} 1 & \bar{X} > \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \\ 0 & \text{o.w.} \end{cases}.$$

The power function is given by

$$\beta_\varphi(\mu_1) = \mathbb{P}_{\mu_1}(\bar{X} > \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}) = 1 - \Phi(\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} + z_{1-\alpha}).$$

Definition 3.2. Let Φ_α be all test functions of size $\leq \alpha$. Then $\varphi^* \in \Phi_\alpha$ is said to be most powerful against $\theta \in \Theta_1$, if

$$\beta_{\varphi^*}(\theta) \geq \beta_\varphi(\theta) \quad \forall \varphi \in \Phi_\alpha.$$

And φ^* is said to be uniformly most powerful if

$$\beta_{\varphi^*}(\theta) \geq \beta_\varphi(\theta) \quad \forall \varphi \in \Phi_\alpha,\ \theta \in \Theta_1.$$

3.2 Neyman-Pearson theory

Theorem 3.3 (Neyman-Pearson)**.** *Let $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$, be simple hypothese. Then*

a) any test of the form

$$\varphi(x) = \begin{cases} 1 & f_1(x) > kf_0(x) \\ \gamma(x) & f_1(x) = kf_0(x) \\ 0 & f_1(x) < kf_0(x) \end{cases} \tag{1}$$

for $k \geq 0$ and $0 \leq \gamma(x) \leq 1$ is most powerful for its size.

b) Given $\alpha \in (0,1)$, there exists a test of the form above with $\gamma(x) = \gamma$ a constant s.t. φ has size α.

Proof. This proof is important. Because it gives us a method to construct the most powerful test under the simple hypothese.

For part (a), let φ^* be a test which size is less than φ; that is,

$$\mathbf{E}_{\theta_0}\varphi^*(X) \le \mathbf{E}_{\theta_0}\varphi(X).$$

We hope prove $\mathbf{E}_{\theta_1}\varphi^*(X) \le \mathbf{E}_{\theta_1}\varphi(X)$. Notice that

$$\begin{aligned}\mathbf{E}_{\theta_1}\varphi(X) - \mathbf{E}_{\theta_1}\varphi^*(X) &\le \mathbf{E}_{\theta_1}\varphi(X) - \mathbf{E}_{\theta_1}\varphi^*(X) - k\big[\mathbf{E}_{\theta_0}\varphi(X) - \mathbf{E}_{\theta_0}\varphi^*(X)\big] \\ &= \int D(x)[f_1(x) - kf_0(x)]\ \mathrm{d}x\end{aligned}$$

where $D := \varphi - \varphi^*$. Let $A_0 = \{f_1 < kf_0\}$ and $A_1 = \{f_1 > kf_0\}$. In continuous case,

$$\begin{aligned}\int D(x)[f_1(x) - kf_0(x)]\ \mathrm{d}x &= \int_{A_0} D(x)[f_1(x) - kf_0(x)]\ \mathrm{d}x + \int_{A_1} D(x)[f_1(x) - kf_0(x)]\ \mathrm{d}x \\ &\ge 0\end{aligned}$$

by noticing that $D \le 0$ on A_0 and $D \ge 0$ on A_1.

Part (b). Let $\alpha \in (0,1]$. We want to find a test of the form (1) with size α where $\gamma(x)$ is a constant γ. Thus, we have the following equation:

$$\mathbf{E}_{\theta_0}\varphi(X) = \alpha;$$

that is,

$$\begin{aligned}\mathbb{P}_{\theta_0}\Big(f_1(X) > kf_0(X)\Big) + \gamma\mathbb{P}_{\theta_0}\Big(f_1(X) = kf_0(X)\Big) &= \alpha \\ \mathbb{P}_{\theta_0}\Big(f_1(X) \le kf_0(X)\Big) - \gamma\mathbb{P}_{\theta_0}\Big(f_1(X) = kf_0(X)\Big) &= 1 - \alpha.\end{aligned}$$

Let $\lambda = \frac{f_1}{f_0}$. G_0 be the CDF of λ under θ_0. So we have

$$G_0(k) - \gamma\mathbb{P}_{\theta_0}\Big(\lambda(X) = k\Big) = 1 - \alpha. \tag{2}$$

Define $k = G_0^{-1}(1-\alpha) = \inf\{\tilde{k} : G_0(\tilde{k}) > 1 - \alpha\}$.

- **Case (i).** If G_0 is continuous at k, let $\gamma = 0$.
- **Case (ii).** If G_0 is not continuous at k, let $\gamma = \frac{G_0(k) - (1-\alpha)}{\mathbb{P}_{\theta_0}(\lambda(X) = k)}$.

□

Proposition 3.4. *If T is sufficient for X, the NP test is a function of T.*

Example 28. $X \sim \text{Poisson}(\lambda)$, $H_0 : \lambda = \lambda_0 = 1$ vs $H_1 : \lambda = \lambda_1 = 2$.

- Compute the CDF of $\frac{f_1}{f_0}$:

  Since $\frac{f_1(x)}{f_0(x)} = \frac{e^{-\lambda_1}\frac{\lambda_1^x}{x!}}{e^{-\lambda_0}\frac{\lambda_0^x}{x!}} = e^{\lambda_0 - \lambda_1}(\frac{\lambda_1}{\lambda_0})^x = \frac{2^x}{e}$,

$$\mathbb{P}_{\lambda_0}(\frac{f_1(X)}{f_0(X)} \le k) = \mathbb{P}_{\lambda_0}(\frac{2^X}{e} \le k) = \mathbb{P}_{\lambda_0}(X \le \frac{\ln k + 1}{\ln 2}).$$

- Compute k and γ:

  The formula (2) becomes:

$$\mathbb{P}_{\lambda_0}(X \leq \frac{\ln k + 1}{\ln 2}) - \gamma \mathbb{P}_{\lambda_0}(\frac{2^X}{e} = k) = 1 - \alpha.$$

  If $\alpha = 0.05$, $F_{\lambda_0}^{-1}(1-\alpha) = 3$, so we set $k = \frac{8}{e}$,

$$\gamma = \frac{0.981 - 0.95}{0.061} = 0.5$$

  and thus the NP test is

$$\varphi(x) = \begin{cases} 1 & x > 3 \\ 0.5 & x = 3 \\ 0 & x < 3 \end{cases}.$$

The test statistic is X itself, while the p-value is $\mathbb{P}_\lambda(X > x_0)$, where x_0 is the observed value (since $\lambda_1 > \lambda_0$).

3.3 Monotone likelihood ratio (MLR) property

Definition 3.5. Let $\mathcal{F}_\Theta$ be a family of pdfs/pmfs, where $\Theta \subset \mathbb{R}$ is an interval. We say $\mathcal{F}_\Theta$ has the monotone likelihood ratio (MLR) property in $T(X)$ if, for $\theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$, $\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)}$ is an non-decreasing function of $T(X)$ on $\{x : f_{\theta_1}(x) \neq 0 \text{ or } f_{\theta_2}(x) \neq 0\}$.

Example 29. $X_i \overset{iid}{\sim} U(0,\theta),\ \theta > 0$. Let $\theta_1 < \theta_2$, so for $x \in \mathbb{R}^n$ such that $x_{(n)} < \theta_2$,

$$\begin{aligned} \frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} &= \frac{(\frac{1}{\theta_2})^n \mathbf{1}(x_{(n)<\theta_2})}{(\frac{1}{\theta_1})^n \mathbf{1}(x_{(n)} < \theta_1)} \\ &= \frac{\theta_1^n}{\theta_2^n} \cdot \frac{1}{\mathbf{1}(x_{(n)} < \theta_1)} \\ &= \begin{cases} \frac{\theta_1^n}{\theta_2^n} & \theta_{(1)} > x_{(n)} \\ \infty & \theta_{(1)} \leq x_{(n)} < \theta_2 \end{cases} \end{aligned}$$

$\implies$ it has the MLR in $T(X) = X_{(n)}$.

Example 30. $X_i \overset{iid}{\sim} N(0,\sigma^2),\ \sigma^2 > 0$. Let $\sigma_1^2 < \sigma_2^2$.

$$\frac{f_{\sigma_2}(x)}{f_{\sigma_1}(x)} = \frac{\sigma_1^n}{\sigma_2^n} + \frac{1}{2}(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}) \sum_{i=1}^n x_i^2;$$

so it has the MLR property in $T(X) = \sum_{i=1}^n X_i^2$.

Theorem 3.6.

- *If $X \sim f_\theta$, where $\{f_\theta : \theta \in \Theta\}$ has the MLR property in $T(X)$, then for $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$, any test of the form*

$$\varphi(x) = \begin{cases} 1 & T(x) > t_0 \\ \gamma & T(x) = t_0 \\ 0 & T(x) < t_0 \end{cases}$$

  has $\beta_\varphi(\theta)$ non-decreasing and is UMP for size $\alpha = \mathbf{E}_{\theta_0}(\varphi(X))$ if this is non-zero.

- *Also, for any $\alpha \in (0,1)$, $\exists t_0 \in \mathbb{R}$ and $\gamma \in (0,1)$ s.t. the above test is UMP of size α.*

Example 31. $X_i \overset{iid}{\sim} \text{Gamma}(\alpha, 1),\ \alpha > 0$. Find a UMP test for $H_0 : \alpha \geq \alpha_0$ vs $H_1 : \alpha < \alpha_0$. Note that

$$f(x) = \frac{1}{[\Gamma(\alpha)]^n \prod_{i=1}^n x_i} \exp\{\alpha \sum_{i=1} n \log x_i - \sum_{i=1}^n x_i\}$$

has the MLR property in $T(x) = \sum_{i=1}^n \log(x_i)$. Therefore, applying the theorem, any test of the form

$$\varphi(x) = \begin{cases} 1 & T(x) < t_0 \\ 0 & T(x) \geq t_0 \end{cases}$$

is UMP for its size $\alpha^* = \mathbf{E}_{\alpha_0}(\varphi(X))$.

For a fixed $\alpha^* \in (0,1)$, let F_0 be the CDF of $T(X)$ under α_0, and choose $t_0 = F^{-1}(\alpha^*)$, so that

$$\mathbf{E}_{\alpha_0}(\varphi(X)) = \mathbb{P}_{\alpha_0}(T(X) < t_0) = \alpha^*.$$

3.4 Unbiased tests

Definition 3.7.

- A test φ of $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ is said to be unbiased at size α if

$$\beta_\varphi(\theta) \leq \alpha \quad \forall \theta \in \Theta_0$$
$$\beta_\varphi(\theta) \geq \alpha \quad \forall \theta \in \Theta_1$$

- Let U_α be the class of all unbiased size α tests.
- If $\exists \varphi \in U_\alpha$ s.t. $\beta_\varphi(\theta) \geq \beta_{\varphi'}(\theta)\ \forall \varphi' \in U_\alpha,\ \forall \theta \in \Theta_1$, then φ is called a UMP unbiased test.

Definition 3.8.

- A test φ is said to be α-similar on $\Theta^* \subset \Theta$ if

$$\beta_\varphi(\theta) = \alpha \quad \forall \theta \in \Theta^*.$$

- Let $\Lambda = \bar{\Theta}_0 \cap \bar{\Theta}_1$.
- A test which is UMP over all tests that are α-similar on Λ is said to be a UMP α-similar test.

Remark. If $\beta_\varphi(\theta)$ is continuous in θ for all φ, then any unbiased size α test φ is α-similar on Λ.

It is easier to find a UMP α-similar test than to find a UMP unbiased test. The following theorem tells us tests that are UMP α-similar on the boundary are often UMP unbiased.

Theorem 3.9. *If β_φ is continuous in θ for all φ. And φ^* is UMP α-similar test on Λ with size α, then φ^* is a UMP unbiased test.*

3.5 Exponential family: Part III

Theorem 3.10. *The 1-parameter exponential family*

$$f_\theta(x) = h(x) \exp\{Q(\theta)T(x) - D(\theta)\}$$

has the MLR in T if Q is non-decrease.

Remark. Depending on the parametrization, Q may be non-increasing. Take $Q' = -Q$ and $T' = -T$.

Example 32. $X_i \overset{iid}{\sim} \text{Poisson}(\lambda),\ \lambda > 0$. The sufficient statistic is $T(X) = \sum_{i=1}^n X_i$, where $Q(\lambda) = \log(\lambda)$ is increasing.

Corollary 3.11. *Let $\mathcal{F}_\Theta$ be a 1-par exponential family. There exists a UMP test of*

$$H_0 : \theta \leq \theta_{00} \text{ or } \theta \geq \theta_{01} \text{ vs } H_1 : \theta_{00} < \theta < \theta_{01}$$

of the form

$$\varphi(x) = \begin{cases} 1 & t_{00} < T(x) < t_{01} \\ \gamma_j & T(x) = t_{0j} \\ 0 & T(x) < t_{00} \text{ or } T(x) > t_{01} \end{cases}$$

with t_{0j} determined by $\mathbf{E}_{\theta_{00}}(\varphi(X)) = \mathbf{E}_{\theta_{01}}(\varphi(X)) = \alpha$.

Remark. UMP tests for one-parameter exponential families don't exist for

- $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, or
- $H_0 : \theta_{00} \leq \theta \leq \theta_{01}$.

Theorem 3.12. *Let $\mathcal{F}_\Theta$ be a one-parameter exponential family, so that β_φ is continuous in θ for all φ. Consider testing*

a) $H_0 : \theta_1 \leq \theta \leq \theta_2$ vs $\theta < \theta_1$ or $\theta > \theta_2$

b) $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

Then

$$\varphi_a(x) = \begin{cases} 1 & T(x) < c_1 \text{ or } T(x) > c_2 \\ \gamma_i & T(x) = c_i \\ 0 & o.w. \end{cases}$$

where c_i, γ_i are chosen s.t. $\mathbf{E}_{\theta_1}\varphi_a(X) = \mathbf{E}_{\theta_2}\varphi_a(X) = \alpha$, is a UMP unbiased size α test, and

$$\varphi_b(x) = \begin{cases} 1 & T(x) < d_1 \text{ or } T(x) > d_2 \\ \gamma_i & T(x) = d_i \\ 0 & o.w. \end{cases}$$

where d_i, γ_i are chosen s.t. $\mathbf{E}_{\theta_0}\varphi_b(X) = \alpha$ and $\mathbf{E}_{\theta_0}(T(X)\varphi_b(X)) = \alpha\mathbf{E}_{\theta_0}(T(X))$, is a UMP unbiased size α test.

3.6 Generalized likelihood ratio tests (GLRT)

Definition 3.13. For testing $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, we could use the likelihood ratio

$$r(x) = \frac{\sup_{\theta \in \Theta_1} f_\theta(x)}{\sup_{\theta \in \Theta_0} f_\theta(x)}$$

and reject H_0 if $r(x)$ is large.

Definition 3.14. The generalized likelihood ratio is

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_0} f_\theta(x)}{\sup_{\theta \in \Theta} f_\theta(x)}$$

and a test that rejects H_0 if $\lambda(x) < c$ is a generalized likelihood ratio test (GLRT).

Remark. We choose c such that $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\lambda(x) > c) = \alpha$.

Proposition 3.15.

a) $r(x) > k \iff \lambda(x) < c$ for some $c = c(k)$.

b) If T is sufficient, then λ can be writen as the function of T.

Proposition 3.16.

a) The NP tests are GLRT's.

b) MLR one-sided tests are GLRT's.

Example 33. $X_i \overset{iid}{\sim} N(\mu, 1)$. $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$. Then

$$\varphi(x) = \begin{cases} 1 & |\bar{x}| > \sqrt{n} z_{1-\alpha} \\ 0 & \text{o.w.} \end{cases}$$

is UMPU. Now, compute the GLR,

$$\lambda(x) = \exp(-\frac{n}{2}\bar{x}^2) < c$$

$\iff |\bar{x}| > c'$, so an α-level GLRT is UMPU.

Example 34. $X_i \overset{iid}{\sim} f_{\theta,a}$, $f_{\theta,a} = \frac{1}{\theta} e^{-\frac{(x-a)}{\theta}} \mathbf{1}(x \geq a)$. $H_0 : \theta = 1$ vs $H_1 : \theta \neq 1$.
Compute the MLEs:

$$\hat{a} = X_{(1)}, \quad \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} (X_i - X_{(1)}).$$

Then the GLR is

$$\lambda(x) = \frac{\exp(-\sum_{i=1}^{n}(x_i - x_{(1)}))}{\frac{1}{\hat{\theta}^n} \sum_{i=1}^{n}(x_i - x_{(1)})} = \hat{\theta}^n \exp(-n(\hat{\theta} + 1));$$

and the GLRT rejects H_0 if and only if $\hat{\theta} < c_1$ or $\hat{\theta} > c_2$. Note that, under H_0, the distribution of $\hat{\theta}$ is independent of a.

Definition 3.17. A test function φ is said to have asymptotic size α if

$$\limsup_n \sup_{\theta \in \Theta_0} \beta_\varphi(\theta) \leq \alpha.$$

Theorem 3.18 (Wilk)**.** *Under the regularity conditions, if $H_0 : \theta = \theta_0$, $\hat{\theta}_n$ is the MLE for $\theta \in \Theta \subset \mathbb{R}^k$, and $X_i \overset{iid}{\sim} f_\theta$. Then*

$$-2 \log \lambda(x) \xrightarrow{w} \chi^2_k.$$

3.7 Other large sample tests

Definition 3.19. Begin again with

$$H_0 : \theta = \theta_0 \quad H_1 : \theta \neq \theta_0.$$

- **Rao score test**

  $$R_n = n\psi_n(\theta_0)^T I^{-1}(\theta_0)\psi_n(\theta_0)$$

  where $\psi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \psi(x_i; \theta)$.

- **Wald test**

  $$W_n = n(\hat{\theta}_n - \theta_0)^T I(\theta_0)(\hat{\theta}_n - \theta_0)$$

  where $\hat{\theta}_n$ is the general MLE.

Proposition 3.20. *a) $R_n \xrightarrow{w} \chi^2_k$ as $n \to \infty$.*

b) $W_n \xrightarrow[H_0]{w} \chi^2_k$ as $n \to \infty$.

c) $W_n = -2 \log \lambda(x) + o_p(1)$.

4 Decision Theory and Bayes Methods

4.1 Basic Setting: Bayes methods and decision theory

Definition 4.1. Let $X \sim f_\theta = f(\theta|x)$.

- A prior distribution π is a probability distribution of Θ.
- The posterior distribution for θ is
$$\pi(\theta|x) = \frac{f(\theta|x)\pi(\theta)}{f(x)}$$
or $\pi(\theta|x) \propto f(\theta|x)\pi(\theta)$.
- Let $\mathcal{F}_\Theta$ be a class of pdfs/pmfs. A family Π of prior distributions on Θ is a conjugate family for $\mathcal{F}_\Theta$ if
$$\pi(\theta|x) \in \Pi$$
for all x and for all $\pi \in \Pi$.

Example 35. $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$. σ^2 known. $\mu \sim N(\mu_0, \tau_0^2)$.
Compute the posterior distribution:

$$\begin{aligned}\pi(\theta|x) &\propto f(\theta|x)\pi(\theta)\\ &= \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\} \cdot \exp\{-\frac{1}{2\tau_0^2}(\mu-\mu_0)^2\}\\ &\propto \exp\{-\frac{1}{2\sigma^2}[n\mu^2 - n\bar{x}\mu] - \frac{1}{2\tau_0^2}[\mu^2 - 2\mu\mu_0]\}\\ &= \exp\{-\frac{1}{2}(\frac{n}{\sigma^2}+\frac{1}{\tau_0^2})\mu^2 + (\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\tau_0^2})\mu\}\\ &\propto \exp\{-\frac{1}{2\tau_1^2}(\mu-\mu_1)^2\}\end{aligned}$$

$\implies \mu|X = x \sim N(\mu_1, \tau_1^2)$.

Example 36. $X_i \overset{iid}{\sim} \text{Bin}(m, p)$. m known.
$f(x|p) = \binom{m}{x} \exp\{x \log(\frac{p}{1-p} + n\log(1-p))\}$.

Example 37. $X_i \overset{iid}{\sim} N(\theta, \sigma^2)$. σ^2 known. $\pi(\theta) \propto 1$. So

$$\begin{aligned}\pi(\theta|x) \propto f(\theta|x) &\propto \exp\{\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\theta)^2\}\\ &\propto \exp\{-\frac{n}{2\sigma^2}(\theta-\bar{x})^2\}\end{aligned}$$

$\implies \theta|X = x \sim N(\bar{x}, \sigma^2/n)$.

Definition 4.2.

- Model: $\mathcal{F}_\Theta$ a space of distributions.
- Action Space: $\mathcal{A}$ is the set of valid decisions one can make.
- Loss Function: $l : \Theta \times \mathcal{A} \to \mathbb{R}^+$ indicating the loss caused by taking action $a \in \mathcal{A}$ if $\theta \in \Theta$ is the true parameter value.
- Decision Rule: $\delta : \underline{X} \to \mathcal{A}$ a statistic.

Definition 4.3. Let $\mathcal{D}$ be the class of decision rules and l be a specified loss function. The risk function of $\delta \in \mathcal{D}$ is
$$R(\theta, \delta) = \mathbf{E}_\theta\big(l(\theta, \delta(X))\big).$$

4.2 Bayes rules

Definition 4.4.

- For a given prior π on Θ, the Bayes' risk of $\delta \in \mathcal{D}$ is

$$r(\pi, \delta) = \mathbf{E}_\pi\Big(R(\theta, \delta(X))\Big) = \mathbf{E}_\pi\Big(\mathbf{E}\big(l(\theta, \delta(X) \mid \theta)\big)\Big).$$

- A Bayes' rule δ^* satisfies

$$r(\pi, \delta^*) = \inf_{\delta \in \mathcal{D}} r(\pi, \delta)$$

  for some prior π.

- The posterior risk of decision a given $X = x$ and a prior π is

$$r_\pi(a|x) = \mathbf{E}\Big(l(\theta, a)\big|X = x\Big).$$

Example 38. Let $X \sim \text{Bin}(n, p)$. Find the min-max rule with the form $\alpha X + \beta$. Assume $p \sim \text{Beta}(\alpha, \beta)$, the Bayes rule is

$$\delta = \frac{X + \alpha}{n + \alpha + \beta}.$$

Then compute the risk $R(\delta, p)$:

$$R(\delta, p) = \mathbf{E}(\frac{X + \alpha}{n + \alpha + \beta} - p)^2 = \frac{1}{(n + \alpha + \beta)^2}\left[((\alpha + \beta)^2 - n)p^2 + (n - 2\alpha(\alpha + \beta))p + \alpha^2\right].$$

Let the risk be a constant (not rely on p) and solve α and β:

$$\alpha = \beta = \frac{\sqrt{n}}{2}.$$

Finally, we find

$$\delta^* = \frac{X + \frac{\sqrt{n}}{2}}{n + \sqrt{n}}.$$

Remark. Note we use the fact that every Bayes rule with constant risk is a min-max rule.

5 Confidence Estimation

5.1 Confident bounds and confident intervals

Definition 5.1. Begin with a family $\mathcal{F}_\Theta$, $\Theta \subset \mathbb{R}$.

- For $\alpha \in (0,1)$, $\underline{\theta}(X)$ is a lower confident bound (LCB) for θ of level $1-\alpha$ if

$$\inf_\theta \mathbb{P}_\theta(\underline{\theta}(X) \leq \theta) \geq 1-\alpha.$$

- For $\alpha \in (0,1)$, $\bar{\theta}(X)$ is a upper confident bound (UCB) for θ of level $1-\alpha$ if

$$\inf_\theta \mathbb{P}_\theta(\bar{\theta}(X) \geq \theta) \geq 1-\alpha.$$

- $(\underline{\theta}(X), \bar{\theta}(X)$ is a level $1-\alpha$ confident interval (CI) if

$$\inf_\theta \mathbb{P}_\theta(\underline{\theta}(x) \leq \theta \leq \bar{\theta}(x)) \geq 1-\alpha.$$

Remark. Confident bounds and intervals are not unique.

Example 39. $X \sim N(\theta, \sigma^2)$. σ known. (So $\frac{X-\theta}{\sigma} \sim N(0,1)$.)
We show: A LCB is $\underline{\theta}(X) = X - \sigma z_{1-\alpha}$. Since

$$\mathbb{P}_\theta(X - \sigma z_{1-\alpha} \leq \theta) = \mathbb{P}(\frac{X-\theta}{\sigma} \leq z_{1-\alpha}) = 1-\alpha.$$

Similarly, a UCB is $\bar{\theta}(X) = X + \sigma z_{1-\alpha}$. Since

$$\mathbb{P}_\theta(X + \sigma z_{1-\alpha} \geq \theta) = \mathbb{P}(\frac{X-\theta}{-\sigma} \leq z_{1-\alpha}) = 1-\alpha.$$

And a CI is $(X - \sigma z_{1-\frac{\alpha}{2}}, X + \sigma z_{1-\frac{\alpha}{2}})$.

5.2 Confident sets and uniformly most accuracy (UMA)

Definition 5.2.

- Suppose $\underline{\theta}^1, \underline{\theta}^2$ are level $1-\alpha$ lower confident bounds. We say $\underline{\theta}^1$ is more accurate than $\underline{\theta}^2$ if for any $\theta \in \Theta$ and $\tilde{\theta} < \theta$,

$$\mathbb{P}_\theta(\underline{\theta}^1(X) \leq \tilde{\theta}) \leq \mathbb{P}_\theta(\underline{\theta}^2(X) \leq \tilde{\theta}).$$

- Let $\underline{\theta}^*$ be a level $1-\alpha$ LCB. If for any other level $1-\alpha$ LCB $\underline{\theta}$, $\underline{\theta}^*$ is more accurate than $\underline{\theta}$, then $\underline{\theta}^*$ is uniformly most accurate (UMA).

Remark. We try to minimize the false converage rate $\mathbb{P}_\theta(\underline{\theta}(X) \leq \tilde{\theta})$. The related notions for UCB are similar.

Definition 5.3.

- A set-valued statistic $S : \underline{X} \to 2^\Theta$ is a level $1-\alpha$ confident set if

$$\inf_\theta \mathbb{P}_\theta(S(X) \ni \theta) \geq 1-\alpha.$$

- S^* is said to be uniformly most accurate if $\forall \theta \in \Theta$, $\tilde{\theta} \neq \theta$, and S another level $1-\alpha$ confident set

$$\mathbb{P}_\theta(S^*(X) \ni \tilde{\theta}) \leq \mathbb{P}_\theta(S(X) \ni \tilde{\theta}).$$

5.3 Duality between confident sets and hypothesis tests

In this subsection, we focus on the relationship between the confident sets and hypothesis tests. Usually, we can construct a level $1-\alpha$ confident set using a deterministic size α test; and conversely, if we have a level $1-\alpha$ confident set, we can define a deterministic size α test. The correspondence is described below

1. For each $\theta_0 \in \Theta$, assume there is a size α test for $H_0 : \theta = \theta_0$:

$$\varphi(x;\theta_0) = \begin{cases} 1 & x \notin A(\theta_0); \\ 0 & x \in A(\theta_0). \end{cases}$$

   Recall that if $\varphi(x;\theta_0) = 1$ means H_0 is rejected; that is $\theta \neq \theta_0$. Thus, if the observed data X is in $A(\theta_0)$, it means θ_0 is closed to the real parameter θ. We define

$$S(X) = \{\theta \in \Theta : X \in A(\theta)\}.$$

2. Let $S(X)$ be a level $1-\alpha$ confident set. For each $\theta_0 \in \Theta$, define a test for $H_0 : \theta = \theta_0$ by

$$\varphi(x;\theta_0) = \mathbf{1}(\theta_0 \notin S(x)).$$

More generally, we can construct a confident set using a randomized test. Letting $u \sim U(0,1)$ independent of X, set $\tilde{\varphi}_{\lambda_0}(x) = \mathbf{1}(\varphi_{\lambda_0}(x) \geq 1-u)$.

Proposition 5.4. *Let φ be a size α randomized test, and $\tilde{\varphi}$ defined above.*

a) $\tilde{\varphi}$ and φ have the same power functions.

b) $\tilde{\varphi}$ and φ have the same size.

Proof. We only consider the simplest case. Assume $\varphi = \begin{cases} 1 \\ \gamma \\ 0 \end{cases}$. Then we can compute the $\mathbf{E}_\theta(\tilde{\varphi})$:

$$\begin{aligned} \mathbf{E}_\theta(\tilde{\varphi}) &= \mathbb{P}(\varphi = 1)\mathbb{P}(1-\gamma > u > 0) + [\mathbb{P}(\varphi = 1) + \mathbb{P}(\varphi = \gamma)]\mathbb{P}(u \leq 1-\gamma) \\ &= \mathbb{P}(\varphi = 1) + \gamma\mathbb{P}(\varphi = \gamma) \\ &= \mathbf{E}_\theta(\varphi) \end{aligned}$$

Notice they are always same whenever $\theta \in \Theta_1$ or $\in \Theta_0$. □

Theorem 5.5. *Let $A : \Theta \to 2^{\underline{X}}$ and $S(X) = \{\theta \in \Theta : X \in A(\theta)\}$. Then $S(X)$ is a level $1-\alpha$ confident set if and only if $\mathbb{P}_\theta(X \notin A(\theta)) \leq \alpha,\ \forall \theta \in \Theta$.*

Example 40. $X_i \overset{iid}{\sim} \text{Poisson}(\lambda)$. $H_0 : \lambda = \lambda_0;\ H_1 : \lambda \neq \lambda_0$. Its UMPU test is of form

$$\varphi_{\lambda_0}(x) = \begin{cases} 1 & \bar{x} < c_1,\ \bar{x} > c_2 \\ \gamma_i & \bar{x} = c_j \\ 0 & \text{o.w.} \end{cases}$$

where c_j and γ_j are chosen to have size α. Now, we want to find a level $1-\alpha$ confident set for λ.

Letting $u \sim U(0,1)$ independent of X_i, set

$$\tilde{\varphi}_{\lambda_0} = \mathbf{1}(\varphi_{\lambda_0}(x) \geq 1-u);$$

notice that $\tilde{\varphi}$ is a size α deterministic test. Its acceptance region is:

$$A(\lambda_0) = \begin{cases} (c_1, c_2) & \min(\gamma_1,\gamma_2) > 1-u \\ [c_1, c_2) & \gamma_1 < 1-u \leq \gamma_2 \\ (c_1, c_2] & \gamma_2 < 1-u \leq \gamma_1 \\ [c_1, c_2] & \max(\gamma_1,\gamma_2) < 1-u \end{cases}$$

Theorem 5.6 (UMP $\implies$ UMA)**.** *Let $\underline{\theta}$ be a level $1-\alpha$ LCB for $\theta \in \mathbb{R}$ for which*

$$\varphi(x;\theta_0) = \begin{cases} 1 & \underline{\theta}(x) > \theta_0 \\ 0 & o.w. \end{cases}$$

is a UMP size α test for $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$, $\forall \theta_0 \in \Theta$. Then $\underline{\theta}$ is UMA.

5.4 Unbiased confident sets

Definition 5.7.

- A confident set $S(X)$ of level $1-\alpha$ is unbiased if

$$\mathbb{P}_\theta(S(X) \ni \theta) \geq 1-\alpha \quad \forall \theta$$
$$\mathbb{P}_\theta(S(X) \ni \tilde{\theta}) \leq 1-\alpha \quad \tilde{\theta} \neq \theta$$

- A level $1-\alpha$ confident set $S(X)$ is uniformly most accurate unbiased (UMAU) if it is unbiased and for any other unbiased level $1-\alpha$ confident set $S'(X)$

$$\mathbb{P}_\theta(S(X) \ni \tilde{\theta}) \leq \mathbb{P}_\theta(S'(X) \ni \tilde{\theta}), \quad \forall \theta \in \Theta,\ \tilde{\theta} \neq \theta.$$

Theorem 5.8 (UMPU $\implies$ UMPA)**.** *For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a size α UMPU test of $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Then $S(X) = \{\theta : X \in A(\theta)\}$ is UMAU level $1-\alpha$.*

5.5 Pivots

Definition 5.9. Let $X \sim f_\theta$. A RV $T(X,\theta)$ is called a pivot if its distribution is free of θ.

Theorem 5.10. *If a set C satisfies $\mathbb{P}(T(X,\theta) \in C) \geq 1-\alpha$, then*

$$S(X) = \{\theta \in \Theta : T(X,\theta) \in C\}$$

is a level $1-\alpha$ confident set.

Example 41. $X_i \overset{iid}{\sim} U(\theta, \theta+1)$. Note $X_{(n)} - \theta$ is a pivot. Let

$$\mathbb{P}(a \leq X_{(n)} - \theta \leq b) = 1-\alpha.$$

Then we get $(X_{(n)} - b, X_{(n)} - a)$.

5.6 Shortest length confident intervals

Example 42. $X_i \overset{iid}{\sim} U(\theta, \theta+1)$. Let $L = b - a$ such that $F(b) - F(a) = 1-\alpha$. First case $b \geq 1$ and $a \in (0,1)$. We solve $1 - a^n - 1 - \alpha$ and get $L = 1 - \alpha^{\frac{1}{n}}$.

Second case $b \in (0,1)$ and $a < 0$. We find $L = (1-\alpha)^{1/n}$.

Finally, we need to compare $1 - \alpha^{1/n}$ and $(1-\alpha)^{1/n}$.

5.7 Bayes credible intervals

Definition 5.11. A level $1-\alpha$ credible interval is a random set $S(X) \subset \Theta$ such that

$$\mathbb{P}(\theta \in S(X) \mid X = x) = 1-\alpha.$$

Example 43. $X_i \overset{iid}{\sim} \text{Bin}(1,p)$. $p \sim \text{Beta}(\alpha,\beta)$.

Compute its posterior: $p|X = x \sim \text{Beta}(\alpha + n\bar{X}, \beta + n - n\bar{X})$.

Compute $l(x)$ and $u(x)$ such that

$$\mathbb{P}(l(x) \leq p \leq u(x) \mid X = x) = 1-\alpha.$$

Then $(l(x), u(x))$ is a level $1-\alpha$ credible interval.

5.8 Large sample confident intervals

Example 44. $X_i \overset{iid}{\sim} \text{Bin}(1, p)$.

- **Option 1**

  Notice that
  $$\sqrt{n}(\hat{p} - p) \xrightarrow{w} N(0, p(1-p))$$
  where $\hat{p} = \bar{X}$.

  By Slusky's, $\sqrt{n}(\hat{p} - p)/\sqrt{\hat{p}(1-\hat{p})} \xrightarrow{w} N(0, 1)$.

  $\implies \hat{p} \pm \sqrt{\hat{p}(1-\hat{p})/n}\, z_{1-\alpha}$ is asymptotic level $1 - \alpha$.

- **Option 2**

  Let $g : x \mapsto 2 \arcsin(\sqrt{x})$. Then
  $$\sqrt{n}(g(\hat{p}) - g(p)) \xrightarrow{w} N(0, 1).$$

  $\implies g(\hat{p}) \pm \frac{1}{\sqrt{n}} z_{1-\alpha}$ is an asymptotic level $1 - \alpha$ CI for $g(p)$.

  $\implies S(X) = \{p : |g(p) - g(\hat{p})| \le \frac{1}{\sqrt{n}} z_{1-\alpha}\}$ is an asymptotic level $1 - \alpha$ CI for p.