UMVUE

Def: $\hat{\tau}$ is called uniformly minimum variance unbiased for $\tau(\theta)$ if: $\begin{cases} \text{i) } E(\hat{\tau})=\theta \text{ for all } \theta \\ \text{ii) for } \forall \text{ other } \tilde{\tau} \text{ s.t. } E(\tilde{\tau})=\theta. \\ Var(\tilde{\tau}) \geq Var(\hat{\tau}) \text{ [Best Unbiased]} \end{cases}$

↳ Thm: If $\hat{\tau}$ is UMVU for $\tau(\theta)$, it's unique.

Thm: [Rao-Blackwell] ⇒ Suppose T is sufficient for θ and $\tilde{\tau}$ is an unbiased estimator for $\tau(\theta)$, then we can construct $\hat{\tau} = E(\tilde{\tau}|T(x))$ is an unbiased estimator for $\tau(\theta)$ and $Var(\hat{\tau}) \leq Var(\tilde{\tau})$

Thm: [Lehman-Scheffe] ⇒ Suppose T is a complete sufficient statistic for θ, if $\hat{\tau}(T)$ is unbiased for $\tau(\theta)$, then $\hat{\tau}$ is just the UMVUE.

Applying Lehman-Scheffe:

1. Find a complete sufficient statistic T.
2. If possible, identify the distribution of T
3. Determine some function $h(T)$ s.t. $E(h(T)) = \tau(\theta) \Rightarrow h(T)$ is the UMVUE for $\tau(\theta)$.

PS: Alternatively, find any simple unbiased estimator $\tilde{\tau}$ and compute $E(\tilde{\tau}|T)$.

Definition: Score function: $S(\theta|x) = \frac{\partial l(\theta)}{\partial \theta}$, $l(\theta) = \log L(\theta|x)$. Also, under regular condition, the Fisher Information: $I(\theta) = E(S(\theta|x)^2) = E(\frac{\partial^2}{\partial \theta^2} l(\theta))$

Theorem: Under regular condition:

1) $E(S(\theta|x)) = 0$ for all θ
2) $Cov(h(x), S(\theta|x)) = \frac{d}{d\theta} E(h(x))$
3). $I(\theta) = Var(S(\theta|x))$

⇒ Corollary: [If $x_i \overset{iid}{\sim} f(x|\theta)$ [Assuming "☆"]. then, $I(\theta) = n \cdot Var(\frac{d}{d\theta} \log f(x|\theta)) = n \cdot I_1(\theta)$. Also, $I_1(\theta) = -E(\frac{\partial^2}{\partial \theta^2} l(\theta))$]

Theorem: Cramer-Rao: Given "☆" and assuming $E(\hat{\tau}(\theta))$ is differentiable w.r.t. θ ⇒

$Var(\hat{\tau}(x)) \geq (\frac{\partial}{\partial\theta} E(\hat{\tau}(x)))^T I^{-1}(\theta) (\frac{\partial}{\partial\theta} E(\hat{\tau}(x))) \Rightarrow Var(\tilde{\tau}(x)) \geq [\frac{d}{d\theta} E(\hat{\tau}(x))]^2 / I(\theta)$ for 1-param case.

Special case: when $\hat{\tau}$ is unbiased for θ: $Var(\hat{\tau}(x)) \geq 1/I(\theta)$

Hypothesis Testings:

Definition: A hypothesis is called simple if the model is completely parameterized by the parameter space Θ and $H_0: \theta = \theta_0$ (i.e, $\Theta_0 = \{\theta_0\}$) is a single set. O.w. the hypothesis is called composite.

Definition: A hypothesis test is a rule that translates the data into a decision between H_0 and H_1: It specifies whether to: a: Reject H_0 in favor of H_1 b: Not reject H_0. Such rule is often of form:

$\begin{cases} T(x) \in A \Rightarrow H_0 \text{ is not rejected} \\ T(x) \notin A \Rightarrow H_0 \text{ is rejected} \end{cases}$ ⇒ [The set $A \in \mathbb{R}^n$ is called acceptance region. The set A^c is called rejection region.

Definition: $\psi(x) = I(T(x) \in A^c)$ is called decision rule. ⇒ $\begin{cases} \psi(x)=1 & : \text{reject } H_0 \\ \psi(x)=0 & : \text{do not reject } H_0 \end{cases}$

Type I error: Reject H_0 when H_0 is true.

Type II error: Fail to reject H_0 when H_1 is true.

Power of a test is the chance that H_0 is rejected $= P(T(x) \notin A) = P(\psi(x)=1) = E(\psi(x))$

H_0 is true: Power = chance of type I error

H_1 is true: Power = 1 - chance of type II error.

...finition: In a parametric family: $H_0: \theta \in \Theta_0 \quad H_1: \theta \in \Theta_1$

i) The power function is:

$\beta(\theta)$ = Power when θ is true value $= E(\varphi(x)|\theta) = P(T(x) \notin A | \theta)$

ii) The size of the test is: $\sup_{\theta \in \Theta_0} \beta(\theta)$ = max chance of type I error.

Also, test is said to have level α if its size is no more than α.

Choosing a sample size determination for a hypothesis testing.

1) Determine the test that will have size α for any n.

2) For a "meaningful" alternative $\theta_1 \in \Theta_1$, solve $1-\beta(\theta_1) = \gamma$. γ is the desired chance of type II error.

Definition: A hypothesis test (decision rule) with power function $\beta(\theta)$ is called unbiased if $\beta(\theta_1) \geq \beta(\theta_0)$ for all $\theta_0 \in \Theta_0$, $\theta_1 \in \Theta_1$. otherwise, it maybe more likely to reject in cases where H_0 is true than it's not.

Definition: A test $\varphi(x)$ is uniformly most powerful (UMP) in class C if:

$$\beta_\varphi \overset{def}{=} E(\varphi(x)|\theta) \geq \beta_{\varphi^*} \overset{def}{=} E(\varphi^*(x)|\theta)$$ for all $\varphi^*(x)$ in the class C and all the θ in Θ.

Theorem: [Neyman-Pearson Lemma] Consider testing simple hypothesis $H_0: \theta = \theta_0$ and $H_1: \theta = \theta_1$. Then the test given by reject H_0 iff $\frac{L(\theta_1|x)}{L(\theta_0|x)} \geq K$ ($K > 0$). is UMP among tests of its size or smaller. i.e. It is all level α tests where $\alpha = P\left(\frac{L(\theta_1|x)}{L(\theta_0|x)} \geq K\right)$. Furthermore, the test is essentially unique.

Corollary: The UMP test in N-P lemma is a function of the minimal sufficient statistic. If $T = t(x)$ is sufficient, then the test the same as reject H_0 is $\frac{f_T(t(x)|\theta_1)}{f_T(t(x)|\theta_0)} > K$

Theorem: Suppose $\theta \subset \mathbb{R}$, T is a 1-dim sufficient statistic. If f_T satisfies for any C, $\theta_1 > \theta_0$. there exists $K \geq 0$ such that $t > C \iff \frac{f_T(t|\theta_1)}{f_T(t|\theta_0)} > K$ ($\star$). Then:

i) $T > C$ is a UMP test for $\left\{\begin{array}{l} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{array}\right\} \to$ of size $\alpha = P(T > C | \theta_0)$ and the power function is nondecreasing in θ.

ii) $T < C$ is a UMP test for $\left\{\begin{array}{l} H_0: \theta \geq \theta_0 \\ H_1: \theta < \theta_0 \end{array}\right\} \to$ and the power is non-~~de~~increasing.

Definition: If ($\star$) holds, T is said to have monotone likelihood ratio. Because it's equivalant to saying $\frac{f_T(t|\theta_1)}{f_T(t|\theta_0)}$ is non-decreasing in t if $\theta_1 > \theta_0$.

Theorem: Suppose x_i is iid: $f(x|\theta) = h(x) \cdot c(\theta) \cdot \exp(w(\theta) \cdot t(x))$ and $w(\theta)$ is non-decreasing in θ then $T = \sum_{i=1}^n t(x_i)$ has monotone likelihood ratio.

Definition: Let $\hat{\theta}_0$ maximize the likelihood under the constraint $\theta \in \Theta_0$. The likelihood ratio statistic for testing $H_0: \theta \in \Theta_0$ vs $H_1: \theta \in \Theta_1$ is: $\lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta|x)}{\sup_{\theta \in \Theta} L(\theta|x)} = \frac{L(\hat{\theta}_0|x)}{L(\hat{\theta}_{MLE}|x)}$

And the level α test is to reject if $\lambda(x) \leq C_\alpha$, where C_α satisfies $\sup_{\theta \in \Theta_0} P(\lambda(x) \leq C_\alpha | \theta = \theta_0) \leq \alpha$. Note. small $\lambda(x)$ indicates $\theta \in \Theta_0$ is not likely.

Note: As in the case of N-P test, $\lambda(x)$ is a function of the (minimal) sufficient statistic $T(x)$, and the test may be expressed in terms of $T(x)$.

Theorem: Assume $X_i \overset{iid}{\sim} f(x|\theta) = h(x)\cdot c(\theta)\cdot \exp(w(\theta)\cdot t(x))$, $w(\theta)$ strictly $\uparrow$, let $T = \sum_{i=1}^{n} t(x_i)$

Define $\varphi(x) = \mathbb{I}[t(x)<t_1] + \mathbb{I}[t(x)>t_2]$, $t_1<t_2$. If $P(T(x)<t_1 \text{ or } T(x)>t_2 \mid \theta_0) = \alpha$ and $E_{\theta_0}(T(x)\varphi(x)) = \alpha\cdot E(T(x)\mid\theta_0)$, then φ is a UMPU size α test of $H_0: \theta=\theta_0$ vs $H_1: \theta\neq\theta_0$

Definition: Suppose T is test statistic for $H_0: \theta\in\Theta_0$ vs $H_1: \theta\in\Theta_1$, with rejection of the form $T>C$ and cdf F_T, then p-value (Significance level) of the test is $P = 1-F_\varphi(T)$.

Interpretation:

1). $P<\alpha$ iff $T>C_\alpha$, where $F_T(C_\alpha) = 1-\alpha$, so P is a test statistic with rejection region "$P<\alpha \Leftrightarrow T>C$"

2). P is a randome variable.

3). P is a simple measure of how much the evidence favors retaining H_0 [Small $P \Rightarrow$ Little evidence for keeping H_0]

Interval Zstimate

Definition: X: Sample Data, $\theta\in\mathbb{R}^p$, let $\tau=\tau(\theta)$ be real value.

i) An interval estimator for τ is a random 2-d statistic $[L(x), U(x)]$ with $L(x)\le U(x)$. The inference we make is "$\tau(\theta)$ is between the calculated values $L(x)$ and $U(x)$".

ii). A set estimator for θ is a random set $C(x)$. $C(\theta)$ depending only on data X.

Definition: Let $C=C(x)$ be a set estimator of θ.

i) The coverage probability of C is: $P(C(x) \text{ contains } \theta)$ depends on θ.

ii) The confidence level of C is $\inf_{\theta\in\Theta} P(C(x) \text{ contains } \theta)$

Several ways to identify confidence sets:

i) Use of "pivots" ii) Use of hypothesis tests iii) Use of asymptotic distributions

iv) Bayesian (credible) intervals. v) Likelihood (confidence) intervals.

Definition: A random ~~quantile~~ variable $Q(x,\theta)$ is pivotal quantity if:

i) The distribution of $Q(x,\theta)$ does not depend on θ

ii) There are sets A, $C(x)$ s.t. $Q(x,\theta)\in A \Leftrightarrow \theta\in C(x)$, $C(x)$ is a useful estimator of θ.

If $P(Q(x,\theta)\in A) = 1-\alpha$, then $C(x)$ is a $1-\alpha$ confidence set for θ.

Theorem: Suppose $X_i \overset{iid}{\sim} f(x|\mu,\sigma^2)$ where $f(\cdot|\mu,\sigma^2)$ is a location-scale family.

i) μ unknown, σ^2 known.

T is location invariant $\Rightarrow t-\mu$ is a pivot.

ii) μ known, σ^2 unknown.

S is scale invariant $\Rightarrow S/\sigma$ is a pivot.

iii). Both μ and σ^2 are unknown:

(T,S) is location scale invariant $\Rightarrow \left(\frac{T-\mu}{\sigma}, \frac{S}{\sigma}\right)$ is pivot.

Theorem: Suppose T is a 1-dim statistic with continuous distribution $F_T(t|\theta)$.

i). $Q(T,\theta) = F_T(T|\theta)$ is a pivot. [Probability Transform]

ii). If $F_T(t|\theta) = h(Q(t,\theta))$, $Q(t,\theta)$ is monotone in $t \Rightarrow Q(T,\theta)$ is Pivot.

Theorem: Let $A(\theta_0)$ be a level α acceptence region for a test of: $H_0: \theta = \theta_0$

Define $C(x) = \{\theta_0 : x \in A(\theta_0)\}$, then $C(x)$ is a confidence set with confidence $1-\alpha$.

Bayesian Approach: A $1-\alpha$ credible set $C(x)$ is a subset of Θ such that $\int_{C(x)} \pi(\theta|x)\, d\theta = 1-\alpha$