Chapter 1 & 2: Basics and Sufficiency *

1 Set-up for Statistics

- $(\Omega, \mathcal{F}, \mathcal{P}) \xrightarrow{X} (\mathcal{X}, \mathcal{B})$
- X is $\langle \mathcal{F}, \mathcal{B} \rangle$- measurable if $X^{-1}(B) \in \mathcal{F}$ for any $B \in \mathcal{B}$.
- Dominance: $\mu_1 \ll \mu_2$. Equivalently, $\mu_2(B) = 0 \implies \mu_1(B) = 0$
- σ-finite measure μ: $\exists A_1, \ldots \in \Omega, \Omega = \cup_{i=1}^{\infty} A_i, \mu(A_i) < \infty$.
- Radon-Nikodym Theorem: $P_\theta \ll \mu, \mu$ is σ-finite, then for each $\theta \in \Theta, \exists f_\theta : \mathcal{X} \to \mathbb{R}$ such that

$$P_\theta(B) = \int_B f_\theta d\mu, \forall B \in \mathcal{B}.$$

- $f_\theta(X)$ is called the likelihood function, with respect to measure μ. When X is completely discrete, μ is counting measure; when X is completely continuous, μ is Lebesgue measure; otherwise, μ can be a mixture of these two.

2 Sufficiency and Related Concepts

- Statistic $T(X) : (\mathcal{X}, \mathcal{B}) \to (\mathcal{T}, \mathcal{F})$ is a measurable map.
- σ-algebra generated by T is $\mathcal{B}_T := \sigma\langle T \rangle = \{T^{-11}(A) : A \in \mathcal{F}\} \subset \mathcal{B}$
- Sufficiency: $[X|T(X)] \perp \theta$. Equivalently, $T(x) = T(x') \implies f_\theta(x) = c(x, x')f_\theta(x')$. Equivalently, $\forall B \in \mathcal{B}, \exists \mathcal{B}_T$-measurable random variable $Y_B : \mathcal{X} \to \mathbb{R}$ such that

$$Y_B \equiv E_\theta(I_B|\mathcal{B}_T) = P_\theta(B|\mathcal{B}_T)$$

a.s. P_θ for all $\theta \in \Theta$, i.e. the conditional proabablity of X given $T(X)$ does not depend on θ.

- Factorization Theorem (Halmos-Savage): Suppose $\mathcal{P} \ll \mu, \mu$ is a σ-finite measure on $(\mathcal{X}, B)$. Then $T(X)$ is sufficient for $\mathcal{P} \iff \exists$ nonnegative $\mathcal{B}$-measurable function $h : \mathcal{X} \to \mathbb{R}$ and a $\mathcal{F}$- measurable function $g_\theta : \mathcal{T} \to \mathbb{R}$ such that

$$\frac{dP_\theta}{d\mu}(x) = f_\theta(x) = g_\theta(T(X))h(x)$$

a.s. μ for all $\theta \in \Theta$.

  - Lehmann's Theorem: (Lemma 1.2 / Page 37 of Shao): Let $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{T}, \mathcal{F})$ be measurable and let $\phi : (\mathcal{X}, \mathcal{B}) \to \left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right)$ be $\mathcal{B}_T \equiv \sigma\langle T \rangle$ -measurable. Then, there exists an $\mathcal{F}$ -measurable function $\psi : (\mathcal{T}, \mathcal{F}) \to \left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right)$ such that

$$\phi(x) = \psi(T(x))$$

  - Lemma 03 (Lemma 2.1/Page 104 of Shao): $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is dominated by a σ -finite measure μ if and only if $\mathcal{P}$ is dominated by a probability measure λ of the form

$$\lambda = \sum_{i=1}^{\infty} c_i P_{\theta_i}$$

for some countable subset $\{\theta_i\} \subset \Theta$ and a countable $\{c_i\}$ with $c_i \geq 0$ and $\sum_{i=1}^{\infty} c_i = 1$

*STA643: Advanced Theory of Statistical Inference. Instructed by Dr. Daniel Nordman. Arranged by Zhiling Gu

- Lemma 04: Suppose $\mathcal{P}=\{P_\theta : \theta \in \Theta\}$ is dominated by a σ -finite measure μ and $T:(\mathcal{X}, \mathcal{B}) \rightarrow(\mathcal{T}, \mathcal{F})$. Then, T is sufficient for $\mathcal{P}$ if and only if there exists a nonnegative $\mathcal{F}$ -measurable function $g_\theta:(\mathcal{T}, \mathcal{F}) \rightarrow(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\frac{d P_\theta}{d \lambda}(x)=g_\theta(T(x)) \quad \text { a.s. } \lambda$$

  using a fixed probability measure λ in the form of Lemma 03.

- Minimal Sufficiency: A sufficient statistic $T:(\mathcal{X}, \mathcal{B}) \rightarrow(\mathcal{T}, \mathcal{F})$ is minimal sufficient for $\mathcal{P}($ or $\theta)$ provided for every sufficient statistic $S:(\mathcal{X}, \mathcal{B}) \rightarrow(\mathcal{S}, \mathcal{G})$, there is a function $U: \mathcal{S} \rightarrow \mathcal{T}$ such that

$$T=U \circ S \quad \text { a.s. } \mathcal{P}$$

  (that is, the set $A=\{x \in \mathcal{X}: T(x) \neq U(S(x))\}$ satisfies $P_\theta(A)=0$ for any θ)

  - Theorem 06: Suppose $\mathcal{P}$ is dominated by a σ-finite measure μ and $T:(\mathcal{X}, \mathcal{B}) \rightarrow(\mathcal{T}, \mathcal{F})$ is sufficient. Suppose further that, if given versions of densities $\frac{d P_\theta}{d \lambda}=f_\theta$ and some $\mathcal{P}$ -null set N_0, it turns out that, for two datasets $x, y \in \mathcal{X} \backslash N_0$, the existence of a constant $k(x, y)>0$ such that

$$f_\theta(x)=f_\theta(y) k(x, y) \quad \forall \theta \in \Theta \quad(*)$$

    in turn implies that $T(x)=T(y)$. Then, T is minimal sufficient. (Null set $N_0 \in \mathcal{B}$ and $P_\theta\left(N_0\right)=0$ for all $P_\theta \in \mathcal{P}$.) Proof: only need to show T is a function of $S \Longleftrightarrow$ wherever $S(X)=S(Y), T(X)=T(Y)$.

  - Theorem 07: For finite dimension measure $\mathcal{P}=\left\{P_i\right\}_{i=1}^k$, $T(X)=\left(\frac{f_1(X)}{f_0(X)} \cdots \frac{f_k(X)}{f_0(X)}\right)$ is minimal sufficient for $\mathcal{P}$.

  - Theorem 08: $\mathcal{P} \ll \mathcal{P}_0\left(i . e . P(B)=0, \forall P \in \mathcal{P}_0 \Longrightarrow P(B)=0, \forall P \in \mathcal{P}\right), \mathcal{P}_0 \subset \mathcal{P}$, T sufficient for $\mathcal{P}, \mathcal{P}_0$, T minimal sufficient for $\mathcal{P}_0 \Longrightarrow T$ minimal sufficient for $\mathcal{P}$.

- Ancillary: A statistic $T:(\mathcal{X}, \mathcal{B}) \rightarrow(\mathcal{T}, \mathcal{F})$ is said to be ancillary for $\mathcal{P} \equiv\left\{P_\theta: \theta \in \Theta\right\}($ or $\theta)$ if the distribution of $T(X)$ does not depend on $\theta($ i.e. , is the same for all $\theta)$

- Pivot: A function of observation and probably parameters whose distribution does not depend on θ. e.g. $X_i-\mu, X_i \sim N(\mu, 1)$, note $X_i-\mu$ is not a statistic.

- 1st order Ancillary: A statistic $T:(\mathcal{X}, \mathcal{B}) \rightarrow(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is said to be 1 st order ancillary for $\mathcal{P} \equiv\left\{P_\theta: \theta \in \Theta\right\}$ (or θ) if $\mathrm{E}_\theta T(X)$ does not depend on θ (i.e., is the same for all θ)

- Completeness:

  - (Version A) A statistic $T:(\mathcal{X}, \mathcal{B}) \rightarrow(\mathcal{T}, \mathcal{F})$ (or $\mathcal{P}^T$) is complete for $\mathcal{P}$ (or θ) if
    $(i) \quad h:(\mathcal{T}, \mathcal{F}) \rightarrow(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is $\mathcal{F}$ -measurable
    $(ii) \quad \mathrm{E}_\theta h[T(X)]=0, \quad$ for all $\theta \in \Theta$ imply that $h \circ T=0$ a.s. P_θ, for all $\theta \in \Theta$
  - (Version B) A statistic $T:(\mathcal{X}, \mathcal{B}) \rightarrow(\mathcal{T}, \mathcal{F})$ (or $\mathcal{P}^T$) is complete for $\mathcal{P}$ (or θ) if
    $(i) \quad U:(\mathcal{X}, \mathcal{B}) \rightarrow(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is $\mathcal{B}_T \equiv\left\{T^{-1}(F): F \in \mathcal{F}\right\}$ measurable
    $(ii) \quad \mathrm{E}_\theta U(X)=0, \quad$ for all $\theta \in \Theta$ imply that $U=0$ a.s. P_θ, for all $\theta \in \Theta$
  - Boundedly complete: A statistic $T:(\mathcal{X}, \mathcal{B}) \rightarrow(\mathcal{T}, \mathcal{F})$ (or $\mathcal{P}^T$) is boundedly complete for $\mathcal{P}($ or $\theta)$ if
    $(i) \quad h:(\mathcal{T}, \mathcal{F}) \rightarrow(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is $\mathcal{F}$ -measurable and bounded
    $(ii) \quad \mathrm{E}_\theta h[T(X)]=0, \quad$ for all $\theta \in \Theta$ imply that $h \circ T=0$ a.s. P_θ, for all $\theta \in \Theta$

  Note: By Lehmann's Theorem (Lemma 02), versions A and B above are equivalent. The only function that makes the statistic have expectation zero is an almost surely zero function. h here does not depend on θ. Completeness is a property of the statistic and the model class for the data $\mathcal{P} \equiv\left\{P_\theta: \theta \in \Theta\right\}$. Essentially, if $\mathcal{P}$ is large, the requirement $E_\theta h(T(X))=0, \forall \theta \in \Theta$ becomes stringent, and h can only be satisfied if $h(T(X))=0$ a.s. $P_\theta, \forall \theta \in \Theta$.

  - Proposition 12: If T is complete for $\mathcal{P} \equiv\left\{P_\theta: \theta \in \Theta\right\}$ and $\tilde{\Theta} \subset \Theta$, then T need not be complete for $\tilde{\mathcal{P}} \equiv\left\{P_\theta: \theta \in \tilde{\Theta}\right\}$
  - Theorem 13: If $\Theta \subset \Theta^{\prime}$ and Θ dominates $\Theta^{\prime}$ (meaning that $P_\theta(B)=0$ for all $\theta \in \Theta$ implies $P_\theta(B)=0$ for all $\left.\theta \in \Theta^{\prime}\right)$, then T complete (or boundedly complete) for $\mathcal{P} \equiv \left\{P_\theta: \theta \in \Theta\right\}$ implies that T is complete (or boundedly complete) for $\mathcal{P}^{\prime} \equiv\left\{P_\theta: \theta \in \Theta^{\prime}\right\}$. Proof similar to Theorem 08 (use Factorization theorem).

- Theorem 14 (Bahadur's Theorem): Suppose that a statistic $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{T}, \mathcal{F})$ is sufficient and boundedly complete. Then,
  (i) if $(\mathcal{T}, \mathcal{F}) = \left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right)$ then T is minimal sufficient.
  (ii) if there is any minimal sufficient statistic, then T is minimal sufficient.
  * Example: (Exponential Families) Suppose that X has distribution P_θ where $\mathcal{P} =$

$$\{P_\theta : \theta \in \Theta\} \ll \mu(\sigma \text{ -finite }) \text{ and } \Theta \subset \mathbb{R}^k, \text{ and}$$
$$\frac{dP_\theta}{d\mu}(x) = c(\theta) \exp\left(\sum_{i=1}^k \theta_i T_i(x)\right), \quad \theta = (\theta_1, \ldots, \theta_k)$$

    Then, $T(X) = (T_1(X), \ldots, T_k(X))$ is sufficient for $\mathcal{P}$ by the factorization theorem. Theorem 2.74 (page 108) of Schervish says that T is complete if the parameter space Θ additionally contains an open set in $\mathbb{R}^k$.
- Theorem 15 (Basu's Theorem): If T is a boundedly complete and sufficient statistic and U is an ancillary statistic, then the variables T and U are independent under P_θ for any $\theta \in \Theta$