Module 2: Parametric Estimation

<u>Goal:</u> Based on data $\underline{X}$, to estimate $g(F)$ (or $g(\theta)$)

$\mathcal{A}$: action space = set of all values that can be allowed as estimates of $g(F)$

$g(F)$ (or $g(\theta)$) is generally called estimand

Randomized Estimator: For each $\underline{x} \in \mathcal{X}$, a probability measure / function $(\delta(\cdot|\underline{x}))$ defined for subsets of $\mathcal{A}$.

Here, if for any given $\underline{x} \in \mathcal{X}$, $Y_{\underline{x}}$ is a r.v. corresponding to $\delta(\cdot|\underline{x})$, then

$$\delta(A|x) = Pr(Y_x \in A), \qquad A \subseteq \mathcal{A}.$$

After observing $\underline{X}$ (= $\underline{x}$, say) we do further randomization by generating a random observation $Y_{\underline{x}}$ and take $Y_{\underline{x}}$ as estimate of $g(F)$ (or $g(\theta)$).

Non-Randomized Estimator: A randomized estimator $\delta(\cdot|\underline{x})$ that for each $\underline{x} \in \mathcal{X}$ takes only values 0 or 1 (i.e., for each $\underline{x} \in \mathcal{X}$ $Y_{\underline{x}}$ is a degenerate r.v.) Here, after observing the data $\underline{X}$ (= $\underline{x}$, say) we estimate $g(F)$ (or $g(\theta)$) by degenerate r.v. $Y_{\underline{x}}$. If $Pr(Y_{\underline{x}} = a(\underline{x})) = 1$, then we may take estimate to be $a(\underline{x})$. Thus a non-randomized estimator can be defined to be a function $\delta : \mathcal{X} \to \mathcal{A}$ (no further randomization is required).

Example Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\theta, 1)$ r.v.s where $\theta \in \Theta = \mathbb{R}$ is unknown.

$\mathcal{X} = \mathbb{R}^n$

<u>Goal:</u> To estimate $g(\theta) = \theta$, $\theta \in \mathbb{R}$

Action space $\mathcal{A} = \Theta = \mathbb{R}$

<u>A randomized estimator</u>

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and let $\bar{x}$ be the observed value of $\bar{X}$. For given $\underline{x} \in \mathcal{X}$, let $Y_{\underline{x}} \sim N(\bar{x}, 1)$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Consider

$\delta(B|\underline{x}) = \Pr(Y_{\underline{x}} \in B)$, $B \subseteq \mathcal{A}$.

The randomized estimator, on observing $\underline{X} = \underline{x}$, estimates θ by generating a random observation $Y_{\underline{x}}$ from $N(\bar{x}, 1)$ and takes $Y_{\underline{x}}$ as the estimate of θ.

<u>A non-randomized estimator</u>

Let $Y^*_{\underline{x}}$ be a r.v. that is degenerate at $\bar{x}$.

$$\delta^*(B|\underline{x}) = \Pr(Y^*_{\underline{x}} \in B) = \begin{cases} 1 & \text{if } \bar{x} \in B \\ 0 & \text{otherwise} \end{cases}, \quad B \subseteq \mathcal{A}.$$

Here the non-randomized estimator estimates $g(\theta) = \theta$ by $\bar{x}$ (no further randomization is required after the data $\underline{X} = \underline{x}$ has been collected). We can write estimator δ^* as $\delta^*(\underline{X}) = \bar{X}$.

Loss function: $L: \mathcal{P} \times \mathcal{A} \to [0, \infty)$
or
$L: \Theta \times \mathcal{A} \to [0, \infty)$

Loss incurred in estimating $g(F)$ (or $g(\theta)$)

For $a \in \mathcal{A}$ and $F \in \mathcal{P}$ (or $\theta \in \mathcal{A}$)

$L(F, a)$ or $L(\theta, a)$: Loss incurred in estimating $g(\theta)$ by 'a' when θ is the true parametric value (state of nature).

For a randomized estimator $\delta(\cdot|\underline{x})$, define

$$L^*(F, \delta, \underline{x}) = \int_{\mathcal{A}} L(F, a)\, d\delta(a|\underline{x})$$

or

$$L^*(\underline{\theta}, \delta, \underline{x}) = \int_{\mathcal{A}} L(\underline{\theta}, a)\, d\delta(a|\underline{x}).$$

For a randomized estimator $\delta(\underline{x})$ we have

$$L^*(F, \delta, \underline{x}) = L(F, \delta(\underline{x}))$$

or

$$L^*(\underline{\theta}, \delta, \underline{x}) = L(\underline{\theta}, \delta(\underline{x}))$$

Generally we have $\mathcal{A} \subseteq \mathbb{R}^k$ and consider a loss function $L(\underline{\theta}, a)$ that for every $\underline{\theta} \in \Theta$ is convex in $a \in \mathcal{A}$. In such situations it suffices to consider only non-randomized estimators (Why!!, using Jensen's inequality)

Result:
- $\mathcal{A} \subseteq \mathbb{R}^k$
- For any $F \in \mathcal{P}$, $L(F, a)$ is a convex function of $a \in \mathcal{A}$
- δ is a given randomized estimator satisfying

$$\int_{\mathcal{A}} \|a\|\, d\delta(\underline{a}|\underline{x}) < \infty \quad \forall\ \underline{x} \in \mathcal{X}.$$

Define

$$\delta_1(\underline{x}) = \int_{\mathcal{A}} a\, d\delta(\underline{a}|\underline{x}), \quad \underline{x} \in \mathcal{X}.$$

Then

$$L^*(F, \delta, \underline{x}) \leq L(F, \delta_1(\underline{x})) \quad \forall\ \underline{x} \in \mathcal{X},$$

with strict inequality if L is strictly convex for any $\underline{x} \in \mathcal{X}$ and $F \in \mathcal{P}$.

Note that δ_1 is a non-randomized estimator and thus

$$L(F, \delta_1(\underline{x})) \equiv L^*(F, \delta_1, \underline{x}), \quad \underline{x} \in \mathcal{X}.$$

Rao Blackwell Theorem: Suppose that:

- $\mathcal{A} \subseteq \mathbb{R}^k$ is convex,
- for any $F \in \mathcal{P}$ (or $\theta \in \Theta$) $L(F, \underline{a})$ (or $L(\theta, \underline{a})$) is a convex function of $\underline{a} \in \mathcal{A}$;
- $T \equiv T(\underline{X})$ is a sufficient statistic for $F \in \mathcal{P}$
- $\delta_0 \in \mathbb{R}^k$ is a non-randomized estimator (decision rule) such that $E_F(\|\delta_0(\underline{X})\|) < \infty$.

Define

$$\delta_1(T) = E(\delta_0(\underline{X}) \mid T)$$
$$= (E(\delta_{01}(\underline{X}) \mid T), \ldots, E(\delta_{0k}(\underline{X}) \mid T)),$$

where $\delta_0 \equiv (\delta_{01}, \ldots, \delta_{0k})$.

Then $\quad R_{\delta_1}(F) \le R_{\delta_0}(F) \quad \forall F \in \mathcal{P} \quad \ldots\ldots (*)$

(or $R_{\delta_1}(\theta) \le R_{\delta_0}(\theta) \quad \forall \theta \in \Theta$).

If L is strictly convex in $\underline{a}$ and δ_0 is not sufficient then we have strict inequality for $F \in \mathcal{P}$ (or $\theta \in \Theta$) in (*).

Remark:

(i) $\left\{ \begin{array}{l} \cdot\ \mathcal{A} \subseteq \mathbb{R}^k \text{ is convex} \\ \cdot\ \forall F \in \mathcal{P}\ \ L(F, \underline{a}) \text{ is convex in } \underline{a} \in \mathcal{A} \end{array} \right\} \Rightarrow$ It is enough to consider only non-randomized decision rule based on minimal sufficient statistic

(ii) Concepts of admissibility and sufficiency eliminate some decision rules. However, even after elimination there may be many decision rules left and the best decision rule (or estimator) (in terms of having the smallest risk for every $F \in \mathcal{P}$) may not exist.