Non-Random Parameter Estimation

Given an observation y drawn from $p_\theta(y)$, how do we estimate the parameter θ?
Here we assume **no** prior statistical knowledge of θ.

1. Terminology/Formulae:

   (a) Bias
   The bias of estimator $\hat{\theta}$ is $b(\hat{\theta}) = \mathbf{E}_\theta\left\{\hat{\theta}(y)\right\} - \theta$.
   An estimator is *unbiased* if $b(\hat{\theta}) = 0$.

   (b) Sufficient statistic
   A function $T(y)$ is a sufficient statistic for estimation of θ if $P_\theta(y|T(y))$ is not dependent on θ.

   (c) Minimal sufficient statistic
   A sufficient statistic for θ is minimal if it is a function of every other sufficient statistic for θ.

   (d) Jensen's Inequality
   If $f(x)$ is a convex function, then $\mathbf{E}\{f(X)\} \geq f(\mathbf{E}\{X\})$ with equality iff $X = E\{X\}$ w.p. 1.

   (e) Neyman-Fisher Factorization Theorem
   $T(y)$ is a sufficient statistic for θ iff there exist functions g_θ and $h(y)$ such that

   $$p_\theta(y) \;=\; g_\theta(T(y))h(y) \;\; \forall\, \theta \in \Lambda, y \in \Gamma$$

   (f) Rao-Blackwell Theorem
   Suppose that $\hat{g}(y)$ is any unbiased estimator of $g(\theta)$ and that $T(y)$ is a sufficient statistic for θ. Define

   $$\tilde{g}(T(y)) \;=\; \mathbf{E}_\theta\left\{\hat{g}(Y)|T(Y) = T(y)\right\},$$

   Then

   - $\tilde{g}(T(y))$ is an unbiased estimator of $g(\theta)$
   - $\mathbf{Var}_\theta\left\{\tilde{g}(Y)\right\} \leq \mathbf{Var}_\theta\left\{\hat{g}(Y)\right\}$ with equality iff $P_\theta\left[\tilde{g}(Y) = \hat{g}(Y)\right] = 1$.

   (g) Complete family
   The family $\{P_\theta; \theta \in \Lambda\}$ is said to be complete when $\mathbf{E}_\theta\left\{f(Y)\right\} = 0 \;\; \forall\, \theta \in \Lambda \;\Rightarrow P_\theta\left[f(Y) = 0\right] = 1$.
   Note that if $\{P_\theta; \theta \in \Lambda\}$ is complete, all sufficient statistics for θ are trivial, i.e., every sufficient statistic is related to the observation by a $1:1$ function.

   (h) Complete sufficient statistic
   Suppose $T(y)$ is sufficient for $\{P_\theta; \theta \in \Lambda\}$ and $T(Y) \sim Q_\theta$ when $Y \sim P_\theta$.
   If $\{Q_\theta; \theta \in \Lambda\}$ is a complete family, then $T(y)$ is a complete sufficient statistic.
   Any unbiased estimator that is a function of a complete sufficient statistic is unique and is the MVUE!
   Complete sufficient statistics are minimal.

   (i) Exponential Family
   A class of distributions is said to be an exponential family if there exist the real-valued functions $C, Q_1, \cdots Q_m$, $T_1, \cdots, T_m$, and h such that

   $$p_\theta(y) \;=\; C(\theta)h(y)\exp\left\{\sum_{l=1}^{m} Q_l(\theta)T_l(y)\right\} \quad \forall y \in \Gamma \subset \mathbb{R}^n,\; \forall \theta \in \Lambda \subset \mathbb{R}^m$$

   The T_l above are *complete sufficient statistics* when Λ contains an m-dimensional rectangle.

   (j) The Information Inequality
   Under suitable regularity conditions,

   $$\begin{aligned} \mathbf{Var}_\theta\left[\hat{\theta}(Y)\right] \;&\geq\; \frac{\left[\frac{\partial}{\partial\theta}\mathbf{E}_\theta\left\{\hat{\theta}(Y)\right\}\right]^2}{I_\theta} \\ \text{where } I_\theta \;&\equiv\; \mathbf{E}_\theta\left\{\left(\frac{\partial}{\partial\theta}\log p_\theta(Y)\right)^2\right\} \quad \text{Fisher's information} \\ &=\; -\mathbf{E}_\theta\left\{\frac{\partial^2}{\partial\theta^2}\log p_\theta(Y)\right\} \end{aligned}$$

   Information lower bound attained $\;\Leftrightarrow\;$ $p_\theta(y)$ a single-parameter exponential pdf.

(k) Cramér-Rao Lower Bound (CRLB)

If the estimator is unbiased, i.e., $\mathbf{E}_\theta\left\{\hat{\theta}(Y)\right\} = \theta$, then information inequality becomes

$$\mathbf{Var}_\theta\left[\hat{\theta}(Y)\right] \quad \geq \quad \frac{1}{I_\theta}$$

Estimators which attain the CRLB are called *efficient*.

2. Procedure for finding a MVUE:
   1. Find a complete sufficient statistic $T(y)$ for θ.
   2. Find $\hat{g}(y)$, any unbiased estimator for $g(\theta)$.
   3. Form $\tilde{g}(T(y)) = \mathbf{E}_\theta\left\{\hat{g}(Y)|T(Y) = T(y)\right\}$.

3. Notes:
   - Finding a complete sufficient statistic for a non-exponential family can be quite difficult.
   - Form your unbiased estimator by exploiting the statistics of your observations.
     This is part *ad hoc* procedure, part magic.
   - The information inequality tells you what is the **best** you can do given the observations.

     if $\hat{\theta}(y)$ achieves the CRLB then $\hat{\theta}(y) = \hat{\theta}_{\text{MVUE}}(y)$

     but $\hat{\theta}(y) = \hat{\theta}_{\text{MVUE}}(y) \not\Longrightarrow$ CRLB achieved