Estimators and hypothesis testing

Principles of data reduction

1. **Any statistic $T(\mathbf{X})$ defines a data reduction or data summary** for the random variable vector $\mathbf{X} = \{X_1, \ldots, X_n\}$.

2. **Sufficiency principle** A statistic $T(\mathbf{X})$ is sufficient statistic for θ if,

$$P\Big(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = T(\mathbf{x})\Big) = \frac{p(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)}$$

   is independent of θ, and where $p(\mathbf{x}|\theta)$ is the joint pdf (or pmf) of the sample $\mathbf{X}$, and $q(T(\mathbf{x})|\theta)$ is the pdf (or pmf) of $T(\mathbf{x})$.

3. **Factorization theorem** Let $f(\mathbf{x}|\theta)$ be the joint pdf (or pmf) of the sample $\mathbf{X}$. $T(\mathbf{X})$ is sufficient statistic for θ if there exists $g(t|\theta)$ and $h(\mathbf{x})$, such that for all sample points $\mathbf{x}$ and all parameter points,

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$$

   $g(T(\mathbf{x})|\theta)$ depends on the sample only through $T(\mathbf{x})$, while $h(\mathbf{x})$ does not depend on θ or $T(\mathbf{x})$.

Estimators

1. **Methods of moments**

2. **Maximum likelihood estimator** Let $L(\theta|X) = f(X|\theta)$ be the likelihood function of $X = \{X_1, X_2, \ldots, X_n\}$ then the ML solves

$$\hat{\theta} = \arg\max L(\theta|X) = \arg\max[\log L(\theta|X)]$$

3. **Invariance property of MLE** If $\hat{\theta}$ is a MLE, for any function $\tau(\cdot)$, the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

4. **Evaluating estimators**:

   (a) MSE: $E_\theta(\hat{\theta} - \theta)^2 = V_\theta(\hat{\theta}) + (E_\theta\hat{\theta} - \theta)^2$

   (b) Bias: $E_\theta(\hat{\theta} - \theta)$.

5. **Cramer-Rao inequality** If

$$\frac{d}{d\theta}E_\theta\hat{\theta} = \int \frac{\partial}{\partial\theta}\hat{\theta}f(\mathbf{X}|\theta)d\mathbf{X} \quad \text{and} \quad V_\theta(\hat{\theta}) < \infty$$

then

$$V_\theta(\hat{\theta}) \geq \frac{\left(\frac{d}{d\theta} E_\theta \hat{\theta}\right)^2}{E_\theta \left(\frac{\partial}{\partial\theta} \log f(\mathbf{X}|\theta)\right)^2}$$

If $\{X_i\}$ is iid then:

$$V_\theta(\hat{\theta}) \geq \frac{\left(\frac{d}{d\theta} E_\theta \hat{\theta}\right)^2}{n E_\theta \left(\frac{\partial}{\partial\theta} \log f(X|\theta)\right)^2}$$

6. **Information matrix equality:** For exponential family

$$E_\theta \left(\frac{\partial}{\partial\theta} \log f(X|\theta)\right)^2 = -E_\theta \left(\frac{\partial^2}{\partial\theta^2} \log f(X|\theta)\right)$$

Hypothesis testing

1. **Hypothesis test** is a rule that specifies:
   - For which sample values the decision is made to accept H_0 as true
   - For which sample values H_0 is rejected and H_1 is accepted as true
2. **Critical region or rejection region** the space for which H_0 is rejected R.
3. **Power function** (probability of type I error for $\theta \in \Theta_0$, one minus prob. of type II error $\theta \in \Theta_0^c$)

$$\beta(\theta) = P_\theta(\mathbf{X} \in R)$$

4. **Neyman-Pearson Lemma** Consider $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, where the pdf or pmf of θ_i is $f(\mathbf{x}|\theta_i)$, for $i = 0, 1$. Using a test rejection region R that satisfies,

$$\mathbf{x} \in R \quad if \quad \frac{f(\mathrm{x}|\theta_0)}{f(\mathrm{x}|\theta_1)} < k$$

$$\mathbf{x} \in R^c \quad if \quad \frac{f(\mathrm{x}|\theta_0)}{f(\mathrm{x}|\theta_1)} > k$$

for some $k \geq 0$, and

$$\alpha = P_{\theta_0}(\mathbf{X} \in R)$$

Then, any test that satisfies this conditions is a UMP α test (sufficiency condition).