Lab 5 Handout

1. Mean Square Error

If the random variable Y has a Binomial(n, p) distribution (with n known), consider the two estimators of p:

$$\hat{p_1} = \frac{Y}{n} \qquad\qquad \hat{p_2} = \frac{Y+1}{n+2}$$

(a) Show that $\hat{p_1}$ is unbiased. *Hint*: Remember that, for a Binomial(n, p) random variable Y, we have $E(Y) = np$ and $Var(Y) = np(1-p)$.
(b) Derive the bias of $\hat{p_2}$.
(c) Derive MSE$(\hat{p_1})$ and MSE$(\hat{p_2})$. *Hint:* Recall that MSE$(\hat{\theta})$ = $[\text{Bias}(\hat{\theta})]^2 + Var(\hat{\theta})$
(d) For what values of p is MSE$(\hat{p_1})$ < MSE$(\hat{p_2})$?

(a) $E(\hat{p}_1) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = \frac{1}{n}np = \boxed{p}$ ✓ (unbiased)

(b) $\text{Bias}(\hat{p}_2) = E(\hat{p}_2) - p = E\left(\frac{Y+1}{n+2}\right) - p = \frac{E(Y)+1}{n+2} - p$

$= \frac{np+1}{n+2} - p = \frac{np+1-np-2p}{n+2} = \boxed{\frac{1-2p}{n+2}}$

(c) $\text{MSE}(\hat{p}_1) = [\text{Bias}(\hat{p}_1)]^2 + Var(\hat{p}_1) = 0 + Var\left(\frac{Y}{n}\right) =$

$= \frac{1}{n^2}Var(Y) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}.$

$\text{MSE}(\hat{p}_2) = [\text{Bias}(\hat{p}_2)]^2 + Var(\hat{p}_2) = \frac{(1-2p)^2}{(n+2)^2} + Var\left(\frac{Y+1}{n+2}\right)$

$= \frac{(1-2p)^2}{(n+2)^2} + \frac{1}{(n+2)^2}Var(Y+1)$

$= \frac{(1-2p)^2}{(n+2)^2} + \frac{1}{(n+2)^2}Var(Y) = \frac{(1-2p)^2}{(n+2)^2} + \frac{np(1-p)}{(n+2)^2}$

$= \boxed{\frac{np(1-p)+(1-2p)^2}{(n+2)^2}}$

(d) $\text{MSE}(\hat{p}_1) < \text{MSE}(\hat{p}_2)$ when

$\frac{p(1-p)}{n} < \frac{np(1-p)+(1-2p)^2}{(n+2)^2}$

$(n+2)^2 p(1-p) < n^2 p(1-p) + n(1-2p)^2$

$\cancel{n^2 p(1-p)} + [n+4]\,p(1-p) < \cancel{n^2 p(1-p)} + n(1-2p)^2$

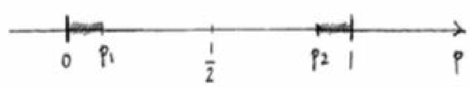
$(4n+4)p - (4n+4)p^2 < n - 4np + 4np^2$

$(4n+4n+4)p^2 - (4n+4n+4)p + n > 0$

$p^2 - p + \frac{n}{8n+4} > 0$

The roots of $p^2 - p + \frac{n}{8n+4} = 0$ are

$p = \frac{1 \pm \sqrt{1 - \frac{4}{8n+4}}}{2} = \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{8n}{8n+2}} = \begin{cases} p_1 = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{4n}{4n+1}} \\ p_2 = \frac{1}{2} + \frac{1}{2}\sqrt{\frac{4n}{4n+1}} \end{cases}$



So $\text{MSE}(\hat{p}_1) < \text{MSE}(\hat{p}_2)$ if and only if $\begin{cases} 0 \le p < p_1 \\ \text{or} \\ p_2 < p \le 1 \end{cases}$

(Note that, as $n \to \infty$, $p_1 \to 0$ and $p_2 \to 1$, so it is "never" the case that $\text{MSE}(\hat{p}_1) < \text{MSE}(\hat{p}_2)$)

2. Sufficiency

Suppose you observe $X_1, \ldots, X_n$ i.i.d Bernoulli(p) distribution. Recall that if $0 \leq p \leq 1, 0 \leq p(1-p) \leq 1/4$.
(a) What is the MLE of p?
(b) Is $\sum X_i$ sufficient for p? Show that X_1 is not sufficient and $X_1, \ldots, X_n$ is sufficient.
(c) Find the estimator with minimum variance among all unbiased estimators. Explain why your choice has this property.
(d) Suppose you want to estimate $\theta = p(1-p)$. Find an unbiased estimator $\hat{\theta}$ of θ that depends only on $\bar{X}_{n.}$ What is the CR bound for any unbiased estimator of p(1-p)?
Solution:
(a)

$$lik(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{(1-x_i)} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$
$$\log(lik(p)) = l(p) = \sum X_i \log(p) - \sum X_i \log(1-p) + n\log(1-p)$$

Taking the derivative and set it equal to 0, we get

$$0 = \frac{d}{dp} l(p) = \frac{\sum X_i}{p} + \frac{\sum X_i}{1-p} - \frac{n}{1-p}$$
$$0 = \sum X_i(1-p) + \sum X_i(p) - np$$
$$p = \frac{\sum X_i}{n} = \bar{X}$$

(b)
$P(X_1 = x_1, \ldots, X_n = x_n | p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{(1-x_i)} = p^{\sum x_i}(1-p)^{n-\sum x_i}$
where $g(T(X_1, \ldots, X_n), p) = p^{\sum X_i}(1-p)^{n-\sum X_i}$
and $h(X_1, \ldots, X_n) = 1$, T = $\sum X_i$. Therefore, by the factorization theorem, $T = \sum X_i$ is a sufficient statistic for p. Similarly, since $\sum X_i$ is a function of $X_1, \ldots, X_n$, the function g is also a function of the set of statistics $X_1, \ldots, X_n$, which is sufficient. The reason why X_1 is not sufficient is because MLE, $\bar{X}$, is not a function of X_1, but a function of all X_i (contrapositive of Corollary A on Page 309).

(c)

The smallest possible variance among all unbiased estimators is the Cramer Rao lower bound $\frac{1}{nI(\theta)}$.

$\hat{p}_{MLE}$ is the estimator with minimum variance among all unbiased estimators. The reason is because

① $\hat{p}_{MLE}$ is unbiased.

② $Var(\hat{p}_{MLE}) = \frac{1}{nI(p)}$. Therefore, $\hat{p}_{MLE}$ achieves the Cramer Rao lower bound.

$$I(p) = -E\left[\frac{\partial^2}{\partial p^2}\log f(X|p)\right]$$
$$= -E\left[\frac{\partial^2}{\partial p^2}\log p^X(1-p)^{1-X}\right]$$
$$= -E\left[\frac{\partial^2}{\partial p^2}\left[X\log p+(1-X)(\log(1-p))\right]\right]$$

$$\Rightarrow I(p) = -E\left[\frac{\partial^2}{\partial p^2}\left[X\log p+\log(1-p)-X\log(1-p)\right]\right]$$
$$= -E\left[\frac{\partial}{\partial p}\left[\frac{X}{p}+\frac{X-1}{1-p}\right]\right]$$
$$= -E\left[-\frac{X}{p^2}+\frac{X-1}{(1-p)^2}\right] = \frac{p}{p^2}+\frac{1-p}{(1-p)^2} = \frac{1}{p(1-p)}$$

$$Var(\hat{p}_{MLE}) = Var(\bar{X}_n)$$
$$Var(\hat{p}_{MLE}) = \frac{1}{n^2}\sum Var(X_i)$$
$$Var(\hat{p}_{MLE}) = \frac{p(1-p)}{n}$$

$$\therefore \frac{1}{nI(p)} = \frac{p(1-p)}{n}$$

This is Cramer Rao Lower Bound

(d)

$$\hat{\theta} = \hat{p}_{MLE}(1-\hat{p}_{MLE})$$
$$\hat{\theta} = \bar{X}_n(1-\bar{X}_n)$$
$$E(\hat{\theta}) = E(\bar{X}_n(1-\bar{X}_n))$$
$$= E[\bar{X}_n - \bar{X}_n^2]$$
$$= E(\bar{X}_n) - E(\bar{X}_n^2)$$
$$= E(\bar{X}_n) - \left[Var(\bar{X}_n) + [E(\bar{X}_n)]^2\right]$$
$$= \frac{1}{n}\sum E(X_i) - \left[\frac{1}{n^2}\sum Var(X_i) + \left[\frac{1}{n}\sum E(X_i)\right]^2\right]$$
$$= \frac{np}{n} - \left[\frac{np(1-p)}{n^2} + \left[\frac{np}{n}\right]^2\right]$$
$$= p - \left[\frac{p(1-p)}{n} + p^2\right]$$
$$= p - \frac{p(1-p)}{n} - p^2$$
$$= p - \frac{p-p^2}{n} - p^2$$
$$= \frac{np - p + p^2 - np^2}{n}$$
$$= \frac{p(n-1) + p^2(1-n)}{n}$$
$$= \frac{(n-1)(p-p^2)}{n}$$
$$= \frac{n-1}{n}(p(1-p))$$

So $\frac{n}{n-1}\hat{p}(1-\hat{p})$ is unbiased.

* To find the CR bound for unbiased estimators of $p(1-p)$, we apply Theorem E of Lecture 5 to get:

$$\text{CR\_bound}(p(1-p)) = \frac{[1-p-p]^2}{nI(p)} = \frac{p(1-p)(1-2p)^2}{n}.$$

3. FI and CR lower bound

Consider a SRS $Y_1, \ldots, Y_n$ with pdf $f(y) = \theta^{-y\theta}$ for $\theta > 0$ and $y > 0$.
(a) Derive the likelihood function. Where in your derivation are you using independence?
(b) Use the fatorization theorem to find a sufficient statistic for θ. Show your work.
(c) Find the MLE for θ.
(d) Find the MLE for $Var(Y) = 1/\theta^2$. State which property of MLEs you are using.
Solution:
(a) $lik(\theta) = \prod_{i=1}^{n} \theta e^{-y_i\theta} = \theta^n e^{-\theta \sum y_i}$. The likelihood function is the product of pdfs because of independence of data.
(b) $lik(\theta) = g(\theta, \sum y_i) * h(y)$ where $g(\theta, \sum y_i) = \theta^n e^{-\theta \sum y_i}$ and $h(y) = 1$. By factorization theorem, $\sum y_i$ is a sufficient statistic for θ.
(c)

$$\begin{aligned}
\log(lik(\theta)) &= n\log(\theta) - \theta \sum y_i \\
\frac{d}{d\theta}\log(lik(\theta)) &= \frac{n}{\theta} - \sum y_i = 0 \\
\therefore \hat{\theta}_{MLE} &= \frac{n}{\sum y_i} = \frac{1}{\bar{Y}}
\end{aligned}$$

(d) By equivariance property of MLEs, the MLE for $\frac{1}{\theta^2}$ is $\frac{1}{(\hat{\theta}_{MLE})^2} = \frac{1}{(1/\bar{Y})^2} = \bar{Y}^2$