Answers to QUIZ 1

1. Let P be the proportion of voters who will vote "YES". Suppose the prior probability distribution of P is given by $\mathbf{Pr}(P < p) = p$ for $0 < p < 1$. You take a poll by choosing nine voters at random, the choice of each being independent of who else was chosen. It is found that six of the nine will vote "YES". Find the posterior probability that more than half of all voters in the whole population will vote "YES".

Answer: The prior probability density function of P is

$$f_P(p) = \frac{d}{dp}F_P(p) = \frac{d}{dp}\mathbf{Pr}(P \le p) = \frac{d}{dp}p = 1 \text{ if } 0 < p < 1$$

and $= 0$ if $p < 0$ or $p > 1$. In other words, P is uniformly distributed on the interval $[0, 1]$, or, in yet other words, $P \sim \text{Beta}(1, 1)$.

Let X be the number of voters in the sample of nine who will vote "YES". Then the likelihood function is

$$L(p) = \mathbf{Pr}(X = 6 \mid P = p) = \binom{9}{6}p^6(1-p)^3.$$

Multiplying the prior density by the likelihood gives us

$$[\text{constant}] \cdot 1 \cdot p^6(1-p)^3 = [\text{constant}] \cdot p^{7-1}(1-p)^{4-1},$$

so we have $P \mid [X = 6] \sim \text{Beta}(7, 4)$. In order to find $\mathbf{Pr}(P > 1/2 \mid X = 6)$, we need the value of the normalizing constant; we write the density as

$$f_{P|[X=6]}(p) = \frac{\Gamma(7+4)}{\Gamma(7)\Gamma(4)}p^{7-1}(1-p)^{4-1} = \frac{10!}{3!6!}p^6(1-p)^3 = 840p^6(1-p)^3$$

for $0 < p < 1$. Then

$$\begin{aligned}
\mathbf{Pr}(P > 1/2) &= \int_{1/2}^{1} 840p^6(1-p)^3\,dp = \int_{1/2}^{1} 840\left(p^6 - 3p^7 + 3p^8 - p^9\right)\,dp \\
&= 840\left[\frac{p^7}{7} - \frac{3p^8}{8} + \frac{p^9}{3} - \frac{p^{10}}{10}\right]_{p=1/2}^{p=1} = \left[120p^7 - 315p^8 + 280p^9 - 84p^{10}\right]_{p=1/2}^{p=1} \\
&= \underbrace{120 - 315 + 280 - 84}_{\uparrow} - \left(\frac{120}{2^7} - \frac{315}{2^8} + \frac{280}{2^9} - \frac{84}{2^{10}}\right)
\end{aligned}$$

If this $\neq 1$ then we can infer that an error has occurred!

$$= 1 - \frac{240 - 315 + 140 - 21}{2^8} = 1 - \frac{44}{2^8} = 1 - \frac{11}{2^6} = \frac{53}{64} = 0.828125.$$

2. Suppose a family of probability distributions of a random variable X is indexed by a parameter θ.

   (a) What does it mean to say that $T(X)$ is a sufficient statistic for θ?

   Answer: It means that the conditional probability distribution of X given $T(X)$ does not depend on θ; the conditional distribution remains the same as θ changes.

   (b) Suppose $T(X)$ is a sufficient statistic for θ. Explain why the value of the Rao-Blackwell estimator $\mathbf{E}(\delta(X) \mid T(X))$ does not depend on θ, even though the probability distribution of $\delta(X)$ must depend on θ in order that $\delta(X)$ make sense as an estimator of θ.

   Answer:

   The conditional distribution of X given $T(X)$ does not depend on θ.
   $\therefore$ The conditional distribution of $\delta(X)$ given $T(X)$ does not depend on θ.
   $\therefore$ The conditional expectation of $\delta(X)$ given $T(X)$ does not depend on θ.

3. Suppose $X_1, X_2 \sim$ i. i. d. Bernoulli(p), i.e., they are independent and identically distributed and

$$X_1 = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases}$$

   (a) Show that $X_1 - X_2$ is not a complete statistic.

   Answer: It is enough to find some function g such that $\mathbf{E}(g(X_1 - X_2))$ remains zero as p changes. But we have $\mathbf{E}(X_1 - X_2) = \mathbf{E}(X_1) - \mathbf{E}(X_2) = p - p = 0$, so we can take g to be the identity function.

   (b) Show that $X_1 + X_2$ is a sufficient statistic for p.

   Answer: One way to do this is by appealing directly to the definition of sufficiency, i.e., by finding $\mathbf{Pr}(X_1 = x_1 \;\&\; X_2 = x_2 \mid X_1 + X_2 = t)$ and observing that no "p" appears in the answer.

$$\mathbf{Pr}(X_1 = x_1 \;\&\; X_2 = x_2 \mid X_1 + X_2 = t) = \frac{p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2}}{\binom{2}{t} p^t (1-p)^{2-t}}$$

$$= \frac{p^{x_1+x_2}(1-p)^{2-(x_1+x_2)}}{\binom{2}{t} p^t (1-p)^{2-t}} = \frac{p^t (1-p)^{2-t}}{\binom{2}{t} p^t (1-p)^{2-t}} = \frac{1}{\binom{2}{t}},$$

   and no "p" appears here.

(c) You may use the fact that $X_1 + X_2$ is a complete statistic. Show that X_1X_2 is an unbiased estimator of p^2, and find the best unbiased estimator of p^2, i.e, the one with the smallest mean squared error among all unbiased estimators of p^2.

Answer: The Lehman-Scheffé theorem says that the conditional expectation of an unbiased estimator given a complete sufficient statistic is the unique best unbiased estimator. So we seek $\mathbf{E}(X_1X_2 \mid X_1 + X_2)$. Notice that X_1X_2 must be either 0 or 1, and is 1 if and only if both X_1 and X_2 are 1, and that happens if and only if $X_1 + X_2 = 2$. So

$$\mathbf{E}(X_1X_2 \mid X_1{+}X_2) = \mathbf{Pr}(X_1X_2 = 1 \mid X_1{+}X_2) = \begin{cases} 1 & \text{if } X_1 + X_2 = 2, \\ 0 & \text{if } X_1 + X_2 = \text{ either 0 or 1.} \end{cases}$$

Since we also have

$$X_1X_2 = \begin{cases} 1 & \text{if } X_1 + X_2 = 2, \\ 0 & \text{if } X_1 + X_2 = \text{ either 0 or 1,} \end{cases}$$

we can say that $\mathbf{E}(X_1X_2 \mid X_1{+}X_2) = X_1X_2$. In other words, X_1X_2 is already the best unbiased estimator of p^2, and is unchanged by the Rao-Blackwell process of improving an estimator.

(d) Find the maximum likelihood estimator of p^2.

Answer: By "invariance" of maximum-likelihood estimators, the maximum-likelihood estimator of p^2 is just the square of the maximum-likelihood estimator of p. The likelihood function is

$$L(p) = P(X_1 = x_1 \;\&\; X_2 = x_2) = p^{x_1+x_2}(1-p)^{2-x_1-x_2}.$$

Therefore

$$\ell(p) = \log L(p) = (x_1 + x_2)\log p + (2 - x_1 - x_2)\log(1-p).$$

$$\ell'(p) = \frac{x_1 + x_2}{p} - \frac{2 - x_1 - x_2}{1-p}$$

$$= \frac{2\left(\dfrac{x_1+x_2}{2}\right) - p}{p(1-p)} \quad \begin{cases} > 0 & \text{if } 0 < p < (x_1+x_2)/2, \\ = 0 & \text{if } p = (x_1+x_2)/2, \\ < 0 & \text{if } (x_1+x_2)/2 < p < 1. \end{cases}$$

Consequently $\widehat{p} = (X_1 + X_1)/2$, and so the maximum-likelihood estimator of p^2 is $(X_1 + X_2)^2/4$.

(e) Consider the two estimators that you found above: the best unbiased estimator of p^2 and the maximum likelihood estimator of p^2. Which has a smaller mean squared error when $p = 1/2$?

Answer: The best unbiased estimator is

$$\begin{cases} 0 & \text{if } X_1 + X_2 \in \{0, 1\}, \\ 1 & \text{if } X_1 + X_2 = 2, \end{cases}$$

and so it is

$$\begin{cases} 0 & \text{with probability } 1 - p^2, \\ 1 & \text{with probability } p^2. \end{cases}$$

Its mean squared error is

$$\overbrace{(0 - p^2)^2}^{\substack{\text{Squared error when the} \\ \text{estimator is 0} \\ \downarrow}} \cdot \underbrace{(1 - p^2)}_{\substack{\uparrow \\ \text{Probability that} \\ \text{the estimator is 0}}} + \overbrace{(1 - p^2)^2}^{\substack{\text{Squared error when the} \\ \text{estimator is 1} \\ \downarrow}} \cdot \underbrace{p^2}_{\substack{\uparrow \\ \text{Probability that} \\ \text{the estimator is 1}}} \overset{\substack{\text{when } p = 1/2 \\ \downarrow}}{=} \frac{3}{16} = 0.1875.$$

The maximum-likelihood estimator is

$$\begin{cases} 0 & \text{if } X_1 + X_2 = 0, \\ 1/4 & \text{if } X_1 + X_2 = 1, \\ 1 & \text{if } X_1 + X_2 = 2, \end{cases}$$

and so it is

$$\begin{cases} 0 & \text{with probability } (1 - p)^2, \\ 1/4 & \text{with probability } 2p(1 - p), \\ 1 & \text{with probability } p^2. \end{cases}$$

Its mean squared error is therefore

$$\overbrace{(0 - p^2)^2}^{\substack{\text{squared} \\ \text{error} \\ \downarrow}} \cdot \underbrace{(1 - p)^2}_{\substack{\uparrow \\ \text{probability}}} + \overbrace{(1/4 - p^2)^2}^{\substack{\text{squared} \\ \text{error} \\ \downarrow}} \cdot \underbrace{2p(1 - p)}_{\substack{\uparrow \\ \text{probability}}} + \overbrace{(1 - p^2)^2}^{\substack{\text{squared} \\ \text{error} \\ \downarrow}} \cdot \underbrace{p^2}_{\substack{\uparrow \\ \text{probability}}} \overset{\substack{\text{when } p = 1/2 \\ \downarrow}}{=} \frac{5}{32} = 0.15625$$

So the M.S.E. of the M.L.E. is slightly smaller than that of the best unbiased estimator when $p = 1/2$.

4. Among families with two children, let X be the score on a statistics test taken by the first child at age 21, and let Y be the income of the second child at age 40. Suppose the pair (X, Y) has a bivariate normal distribution, and $\mathbf{E}(X) = 65$, $\mathbf{SD}(X) = 10$, $\mathbf{E}(Y) = \$50,000$ per year, $\mathbf{SD}(Y) = \$10,000$ per year, and $\mathbf{corr}(X, Y) = 1/2$. [All of this is fiction.] Among families in which the first child scores 75 on the statistics test at age 21, in what proportion of cases does the second child have an income of at least $\$59,330$ at age 40?

   Answer: On page 315 of DeGroot & Schervish, we learn that

$$\begin{aligned}
\mathbf{E}(Y \mid X) &= \mathbf{E}(Y) + \mathbf{corr}(X, Y)\,\mathbf{SD}(Y)\left(\frac{X - \mathbf{E}(X)}{\mathbf{SD}(X)}\right) \\
&= 50,000 + (1/2)(10,000)\left(\frac{X - 65}{10}\right). \\
\text{So } \mathbf{E}(Y \mid X = 75) &= 55,000, \\
\text{and } \mathbf{var}(Y \mid X) &= (1 - \mathbf{corr}(X, Y)^2)\,\mathbf{SD}(Y) \\
&= (3/4) \cdot 10,000^2. \\
\text{So } \mathbf{SD}(Y \mid X = 75) &= \sqrt{\frac{3}{4}} \cdot 10,000 = \frac{10000\sqrt{3}}{2}.
\end{aligned}$$

   Since the conditional distribution of Y given that $X = 75$ is normal, we can say

$$\mathbf{Pr}(Y \geq 59,330 \mid X = 75) = 1 - \mathbf{Pr}(Y \leq 59,330 \mid X = 75)$$

$$= 1 - \mathbf{Pr}\left(\frac{Y - 55,000}{10000\sqrt{3}/2} \leq \frac{59,330 - 55,000}{10000\sqrt{3}/2} \,\middle|\, X = 75\right)$$

$$= 1 - \Phi\left(\frac{59,330 - 55,000}{10000\sqrt{3}/2}\right) \cong 1 - \Phi(0.500) \cong 1 - 0.6915 = 0.3085.$$

   So the event of interest occurs in about 30.85% of all cases.