541A Final Solutions[1]

1. Question 1

Let $X_1,\ldots,X_n$ be a random sample of size $n\geq 1$. Let $m\geq 1$. Recall that the **bootstrap sample** $Y_1,\ldots,Y_m$ is defined as follows. Given $X_1,\ldots,X_n$, let $Y_1,\ldots,Y_m$ be a random sample of size m from the values $\{X_1,\ldots,X_n\}$ with replacement.

Show that the covariance of Y_1 and Y_2 is equal to $\mathrm{Var}(X_1)/n$.

Conclude that $Y_1,\ldots,Y_m$ are not independent when $\mathrm{Var}(X_1)>0$.

(Hint: to compute $\mathbf{E}Y_1Y_2$, first condition on $X_1,\ldots,X_n$).

Solution. Using the conditional independence, we have

$$\begin{aligned}\mathbf{E}Y_1Y_2 &= \mathbf{E}\Big[\mathbf{E}(Y_1Y_2|X_1,\ldots,X_n)\Big] = \mathbf{E}\Big[\mathbf{E}(Y_1|X_1,\ldots,X_n)\cdot\mathbf{E}(Y_2|X_1,\ldots,X_n)\Big]\\ &= \mathbf{E}\Big[\Big(\mathbf{E}(Y_1|X_1,\ldots,X_n)\Big)^2\Big] = \mathbf{E}\overline{X}^2.\end{aligned}$$

Meanwhile

$$\mathbf{E}(\overline{Y}|X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^n\Big(\frac{1}{n}\sum_{j=1}^n X_j\Big) = \overline{X}. \qquad (\ddagger)$$

So, the covariance of Y_1 and Y_2 is

$$\mathbf{E}(Y_1-\mathbf{E}Y_1)(Y_2-\mathbf{E}Y_2) = \mathbf{E}Y_1Y_2 - (\mathbf{E}Y_1)(\mathbf{E}Y_2) = \mathbf{E}\overline{X}^2 - (\mathbf{E}\overline{X})^2 = \mathrm{Var}\overline{X} = \frac{\mathrm{Var}(X_1)}{n}.$$

So, if X_1 is nonconstant, this covariance is nonzero.

2. Question 2

Let $X_1,\ldots,X_n$ be a random sample from the Bernoulli distribution with unknown parameter $0<p<1$, so that, for all $1\leq i\leq n$,

$$\mathbf{P}(X_i=1)=p, \qquad \mathbf{P}(X_i=0)=1-p.$$

- Find a complete sufficient statistic for p. (As usual, justify your answer.)
- Find the UMVU for p^3. (You may assume $n\geq 3$.)
  (Hint: $X_1X_2X_3$ is an estimator for p^3.)

Solution. The joint distribution of $(X_1,\ldots,X_n)$ is

$$\begin{aligned}\mathbf{P}(X_1=x_1,\ldots,X_n=x_n) &= \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\\ &= e^{(\log p-\log(1-p))\sum_{i=1}^n x_i + n\log(1-p)}.\end{aligned}$$

From the Factorization Theorem, $\sum_{i=1}^n X_i$ is sufficient for p. Also, by some exercises from the homework, $\sum_{i=1}^n X_i$ is complete and sufficient for the exponential family (since the set $\{(\log p-\log(1-p))\colon p\in(0,1)\}$ contains an open interval in $\mathbf{R}$). Alternatively, note that $Z:=\sum_{i=1}^n X_i$ has the binomial distribution and Z is complete, as shown in Example 5.21 in the notes. In any case, $Z:=\sum_{i=1}^n X_i$ is complete and sufficient for θ. Also, $\frac{1}{n}\sum_{i=1}^n X_i$ is unbiased for p, so $\frac{1}{n}\sum_{i=1}^n X_i$ is UMVU for p.

[1]May 2, 2019,

Suppose we want to estimate p^3. We use the unbiased estimate $Y := X_1X_2X_3$ (noting that $\mathbf{E}_pY = \mathbf{E}_\theta X_1 \mathbf{E}_\theta X_2 \mathbf{E} X_3 = p^3$, by independence.) The UMVU is then $\mathbf{E}(Y|Z)$. Let $3 \leq z \leq n$ be an integer. Note that $Y = 1$ when $X_1 = X_2 = X_3 = 1$ and $Y = 0$ otherwise. So,

$$\begin{aligned}
\mathbf{E}_p(Y|Z = z) &= \mathbf{E}_p(1_{X_1=X_2=X_3=1}|Z = z) = \mathbf{P}_p(X_1 = X_2 = X_3 = 1|Z = z)\\
&= \mathbf{P}_p(X_1 = X_2 = X_3 = 1|\sum_{i=1}^{n} X_i = z) = \frac{\mathbf{P}_p(X_1 = X_2 = X_3 = 1, \sum_{i=1}^{n} X_i = z)}{\mathbf{P}_p(\sum_{i=1}^{n} X_i = z))}\\
&= \frac{\mathbf{P}_p(X_1 = X_2 = X_3 = 1, \sum_{i=4}^{n} X_i = z-3)}{\mathbf{P}_p(\sum_{i=1}^{n} X_i = z))} = \frac{p^3\binom{n-3}{z-3}p^{z-3}(1-\theta)^{n-z}}{\binom{n}{z}p^z(1-p)^{n-z}}\\
&= \frac{1}{n(n-1)(n-2)}\frac{(n-z)!z!}{(n-z)!(z-3)!} = \theta^2\frac{z(z-1)(z-2)}{n(n-1)(n-2)}.
\end{aligned}$$

Additionally, $\mathbf{E}_p(Y|Z = z) = 0 = p^3\frac{z(z-1)(z-2)}{n(n-1)(n-2)}$ for $z = 1, 2$ and for $z = 3$. So,

$$\mathbf{E}_p(Y|Z = z) = \frac{z(z-1)(z-2)}{n(n-1)(n-2)}, \qquad \forall\, 0 \leq z \leq n.$$

So, the UMVU is

$$\mathbf{E}_\theta(Y|Z) = \frac{Z(Z-1)(Z-2)}{n(n-1)(n-2)}.$$

3. Question 3

Let $X_1, \ldots, X_n$ be a random sample of size n, so that X_1 is a sample from the uniform distribution on the interval $[\theta - 1/2, \theta + 1/2]$, where $\theta \in \mathbf{R}$ is unknown. Let $g\colon \mathbf{R} \to \mathbf{R}$ be a nonconstant differentiable function of $\theta \in \mathbf{R}$. Show that no UMVU of $g(\theta)$ exists when $n = 1$.

Solution. We have $\int_{\theta-1/2}^{\theta+1/2} u(x)dx = 0$ for all $\theta \in \mathbf{R}$. Differentiating this condition and applying the Fundamental Theorem of Calculus, we get $u(\theta + 1/2) = u(\theta - 1/2)$ for a.e. $\theta \in \mathbf{R}$. (It is assumed that $\mathbf{E}_\theta |U| < \infty$ for all $\theta \in \mathbf{R}$ in order to be an unbiased estimator of 0.) For an example of an unbiased estimator of 0, consider e.g. $u(x) := \mathrm{sign}(\sin(2\pi x))$ for all $x \in \mathbf{R}$.

Assume that W is UMVU. By Theorem 6.18, an Alternate Characterization of UMVU, we must have $\mathbf{E}_\theta WU = 0$ for all $\theta \in \mathbf{R}$ when U is unbiased for 0. That is, $\mathbf{E}_\theta WU = 0$ for all $\theta \in \mathbf{R}$. By step one, we conclude that

$$w(x + 1)u(x + 1) = w(x)u(x), \qquad \text{for a.e. } x \in \mathbf{R}.$$

Using the u we produced there, we can divided both sides by u to conclude that

$$w(x + 1) = w(x), \qquad \text{for a.e. } x \in \mathbf{R}.$$

Now, since W is unbiased for $g(\theta)$, we have $g(\theta) = \int_{\theta-1/2}^{\theta+1/2} w(x)dx$. Differentiating both sides, we get

$$g'(\theta) = w(\theta + 1/2) - w(\theta - 1/2) = 0,$$

so that $g(\theta)$ is constant in θ.

4. Question 4

Let $f\colon \mathbf{R}^n \to \mathbf{R}$ be a convex function. Let $x \in \mathbf{R}^n$ be a local minimum of f. Show that x is in fact a global minimum of f.

Now suppose additionally that f is a C^1 function (all derivatives of f exist and are continuous), and $x \in \mathbf{R}^n$ satisfies $\nabla f(x) = 0$. Show that x is a global minimum of f.

*Solution.*Let $y \in \mathbf{R}^n$ with $y \neq x$. Let $t \in (0,1)$. Assume for the sake of contradiction that $f(y) < f(x)$. By convexity of f,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) < (t + (1-t))f(x) = f(x).$$

Letting $t \to 1^-$ shows that points near x have smaller f values than x, contradicting the local minimality of x. We conclude that $f(y) \geq f(x)$ for all $y \in \mathbf{R}^n$, so that x is a global minimum of f.

In the case that f is C^1, $\nabla f(x) = 0$ implies that x is a local minimum of f, so that the first assertion implies that x is a global minimum of f. To see this, note that the definition of convexity of f implies that the function f lies above the horizontal tangent plane of f at x.

5. Question 5

Give an example where the maximum likelihood estimator and the method of moments estimator are the same. That is, for a random sample of size $n \geq 1$, if Y_n is the MLE for the sample of size n, and if Z_n is the Method of Moments estimator for the sample of size n, then $Y_n = Z_n$. (As usual, justify your answer.)

Solution. We use the Example from Question 2. Let $X_1, \ldots, X_n$ be a random sample from the Bernoulli distribution with unknown parameter $0 < p < 1$, so that, for all $1 \leq i \leq n$,

$$\mathbf{P}(X_i = 1) = p, \qquad \mathbf{P}(X_i = 0) = 1 - p.$$

Suppose we want to estimate p. The joint distribution of $(X_1, \ldots, X_n)$ is

$$\begin{aligned}\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \\ &= e^{(\log p - \log(1-p))\sum_{i=1}^n x_i + n\log(1-p)}.\end{aligned}$$

Let $f(p) := (\log p - \log(1-p))\sum_{i=1}^n x_i + n\log(1-p)$. The MLE is the value of p maximizing f. We have

$$f'(p) = \left(\frac{1}{p} + \frac{1}{1-p}\right)\sum_{i=1}^{n} x_i - n\frac{1}{1-p}.$$

So, if $f'(p) = 0$, $\frac{1}{p(1-p)}\sum_{i=1}^n x_i = n\frac{1}{1-p}$, so that $p = \frac{1}{n}\sum_{i=1}^n x_i$. So, the MLE for p is

$$Y := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The Method of Moments Estimator for p is also Y since $\mathbf{E}X_1 = p$ and we are estimating p, so the MoM estimator is exactly the sample mean as well.

6. Question 6

Let $X := (X_1, \ldots, X_n)$ be a random sample of size n from a Gaussian distribution with unknown mean $\mu \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$. $\forall\, n \geq 1$, define

$$Y = Y_n = Y_n(X_1, \ldots, X_n) := \frac{1}{n}\sum_{j=1}^{n}\Big(X_j - \frac{1}{n}\sum_{i=1}^{n} X_i\Big)^2.$$

You are required to show that Y_n has asymptotically optimal variance as $n \to \infty$ (you are not allowed to use the theorem about the limiting distribution of the MLE; you are required to argue directly.) More specifically, the Cramér-Rao inequality shows that for any unbiased estimator $Z = Z_n(X_1, \ldots, X_n)$ of σ^2, there is a constant $c_n > 0$ such that

$$\mathrm{Var}_\sigma(Z) \geq c_n.$$

You are required to compute the lower bound c_n given by the Cramér-Rao Inequality, and you should also show that

$$\lim_{n\to\infty} \frac{\mathrm{Var}_\sigma(Y_n)}{c_n} = 1.$$

Solutoin. We will show that Y has asymptotically optimal variance without using the exponential family. If we fix $\mu \in \mathbf{R}$ and look at the information of the n-dimensional Gaussian X, we get by modifying an example from the notes, and using that the information of independent random variables is the sum of the informations,

$$\begin{aligned} I_X(\sigma) = nI_{X_1}(\sigma) &= n\mathrm{Var}_\sigma\Big(\frac{d}{d\sigma}\frac{-(X_1-\mu)^2}{2\sigma^2}\Big) = n\sigma^{-6}\mathrm{Var}_\sigma[(X_1-\mu)^2] \\ &= n\sigma^{-6}\mathbf{E}_\sigma((X_1-\mu)^4 - \sigma^4) = 2n\sigma^{-2}. \end{aligned}$$

By Proposition 4.7 from the notes, $\frac{1}{\sigma^2}\sum_{j=1}^{n}\Big(X_j - \frac{1}{n}\sum_{i=1}^{n} X_i\Big)^2$ is a chi-squared distributed random variable with $n-1$ degrees of freedom, so $\mathbf{E}_\sigma(Y) = \sigma^2(n-1)/n$. By the Cramér-Rao Inequality, with $g(\sigma) = \mathbf{E}_\sigma(Y) = \sigma^2(n-1)/n$, the variance of any unbiased estimator Z of $\sigma^2(n-1)/n$ satisfies

$$\mathrm{Var}_\sigma(Z) \geq \frac{|g'(\sigma)|^2}{I_X(\sigma)} = \frac{4\sigma^2(n-1)^2}{n^2 2n\sigma^{-2}} = \frac{2\sigma^4(n-1)^2}{n^3}.$$

Using Proposition 4.7 again, and that the fourth moment of a standard gaussian is 3,

$$\mathrm{Var}_\sigma(Y) = \mathrm{Var}_\sigma\Big[\frac{\sigma^2}{n}\frac{1}{\sigma^2}\sum_{j=1}^{n}\Big(X_j - \frac{1}{n}\sum_{i=1}^{n} X_i\Big)^2\Big] = \frac{\sigma^4}{n^2}2(n-1) = \frac{2\sigma^4(n-1)}{n^2}.$$

In summary,

$$\lim_{n\to\infty}\frac{\mathbf{E}Y}{\sigma^2} = 1, \qquad \lim_{n\to\infty}\frac{\mathrm{Var}_\sigma(Y)}{|g'(\sigma)|^2/I_X(\sigma)} = 1.$$

7. Question 7

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample of size n on the disc $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 \leq s^2\}$ where $s > 0$ is unknown. That is, the probability density function of (X, Y) is

$$f_s(x, y) := \frac{1}{\pi s^2}1_{[0,s]}(\sqrt{x^2+y^2}), \qquad \forall\, (x, y) \in \mathbf{R}^2$$

Find a complete sufficient statistic Z for s and find the density function of Z.

(Hint: first try the $n = 1$ case, then try the $n > 1$ case.)

Solution. Suppose $n = 1$. The Factorization Theorem immediately implies that $Z := \sqrt{X^2 + Y^2}$ is a sufficient statistic for θ. To find the distribution of Z, note that for any $0 < z < \theta$,

$$\mathbf{P}(Z < t) = \mathbf{P}(X^2+Y^2 < t^2) = \int_{\{(x,y)\in\mathbf{R}^2\colon\, x^2+y^2\leq t^2\}} f_\theta(x,y)dxdy = \frac{1}{\pi s^2}\int_{\theta=0}^{\theta=2\pi}\int_{r=0}^{r=t} rdrd\theta = \frac{t^2}{s^2}.$$

So, Z has density $(d/dt)t^2/s^2 = 2ts^{-2}1_{t\in[0,s]}$. And Z is complete since if $\mathbf{E}_s f(Z) = 0$ for all $s > 0$, we have $h(s) := 2s^{-2}\int_0^s tf(t)dt = 0$, we have

$$0 = h'(s) = 2s^{-2}sf(s) - 4s^{-3}\int_0^s tf(t)dt = 2f(s)/s - 2h(s)/s = 2f(s)/s.$$

Therefore, $f(s) = 0$ for all $s > 0$, i.e. $f(Z) = 0$, so that Z is complete.

In the case $n > 1$, note that $\prod_{i=1}^n 1_{[0,s]}(\sqrt{x_i^2 + y_i^2}) = 1[0,s](\max_{i=1,\ldots,n}\sqrt{x_i^2+y_i^2})$, so $Z := \max_{i=1,\ldots,n}\sqrt{X_i^2+Y_i^2}$ is a sufficient statistic for s by the Factorization Theorem. To find the distribution of Z, note that for any $0 < z < \theta$,

$$\mathbf{P}(Z < t) = \mathbf{P}(\sqrt{X^2+Y^2} < t)^n = \frac{t^{2n}}{s^{2n}}.$$

So, Z has density $(d/dt)t^{2n}/s^{2n} = 2nt^{2n-1}s^{-2n}1_{t\in[0,s]}$. And Z is complete since if $\mathbf{E}_s f(Z) = 0$ for all $s > 0$, we have $h(s) := 2ns^{-2n}\int_0^s t^{2n-1}f(t)dt = 0$, we have

$$0 = h'(s) = 2ns^{-2n}s^{2n-1}f(s) - 4n^2s^{-2n-1}\int_0^s t^{2n-1}f(t)dt = 2nf(s)/s - 2n^2h(s)/s = 2nf(s)/s.$$

Therefore, $f(s) = 0$ for all $s > 0$, i.e. $f(Z) = 0$, so that Z is complete.

8. Question 8

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample of size n on the disc $\{(x,y)\in\mathbf{R}^2\colon x^2+y^2 \leq s^2\}$ where $s > 0$ is unknown. That is, the probability density function of (X,Y) is

$$f_s(x,y) := \frac{1}{\pi s^2}1_{[0,s]}(\sqrt{x^2+y^2}), \qquad \forall\,(x,y)\in\mathbf{R}^2$$

Find the UMVU for s.

Solution. Suppose $n = 1$. Then Z has density $(d/dt)t^2/s^2 = 2ts^{-2}1_{t\in[0,s]}$, so that

$$\mathbf{E}_s Z = \int_0^s 2t^2 s^{-2}dt = (2/3)s.$$

So, $Y := (3/2)Z$ is an unbiased function of the complete sufficient statistic, so it is UMVU by Lehmann-Scheffé. (Lehmann-Scheffé says that $W := \mathbf{E}_s(Y|Z)$ is UMVU, and since Y is a multiple of Z, we have $W = Y$, since $\mathbf{E}_s(Z|Z) = Z$, a known property of conditional expectation.)

In the case $n > 1$, Z has density $(d/dt)t^{2n}/s^{2n} = 2nt^{2n-1}s^{-2n}1_{t\in[0,s]}$, so that

$$\mathbf{E}_s Z = \int_0^s 2nt^{2n}s^{-2n}dt = \frac{2n}{2n+1}s.$$

So, $\frac{2n+1}{2n}Z$ is an unbiased function of the complete sufficient statistic, so it is UMVU by Lehmann-Scheffé.