STAT 135 Solutions to Homework 2: 30 points

Spring 2015

Problem 1: 10 points

Suppose X_1, X_2 are independent samples from a Bernoulli distribution $B(p)$, and let $T = X_1 + X_2$.

1. Show that $T \sim Bin(2, p)$

Note that we can consider each X_i as the outcome (1 = success, 0 = failure) of an individual experiment. In this case, it is clear that $T = X_1 + X_2$ counts the number of successes of the two independent trials. T takes values in $\{0, 1, 2\}$, which is the definition of a Binomial random variable. Mathematically, we can show that

$$\begin{aligned}
P(T=0) &= P(X_1 + X_2 = 0) \\
&= P(X_1 = 0, X_2 = 0) \\
&= P(X_1 = 0)P(X_2 = 0) && \text{(by independence)} \\
&= (1-p)^2 \\
&= \binom{2}{0} p^0 (1-p)^2
\end{aligned}$$

and

$$\begin{aligned}
P(T=1) &= P(X_1 + X_2 = 1) \\
&= P(X_1 = 1, X_2 = 0) + P(X_1 = 0, X_2 = 1) \\
&= P(X_1 = 1)P(X_2 = 0) + P(X_1 = 0)P(X_2 = 1) && \text{(by independence)} \\
&= p(1-p) + (1-p)p \\
&= 2p(1-p) \\
&= \binom{2}{1} p^1 (1-p)^1
\end{aligned}$$

and

$$\begin{aligned}
P(T=2) &= P(X_1 + X_2 = 2) \\
&= P(X_1 = 1, X_2 = 2) \\
&= P(X_1 = 1)P(X_2 = 1) && \text{(by independence)} \\
&= p^2 \\
&= \binom{2}{2} p^2 (1-p)^0
\end{aligned}$$

So T clearly has the same probability mass function as a Binomial(2,p) random variable:

$$P(T=k) = \binom{2}{k} p^k (1-p)^{2-k}$$

This implies that $T \sim Bin(2, p)$.

2. If $\theta = \gamma_0 + \gamma_1 p + \gamma_2 p^2$, show that

$$\widehat{\theta} = \begin{cases} \gamma_0 & \textbf{if } T = 0, \\ \frac{1}{2}(\gamma_1 + 2\gamma_0) & \textbf{if } T = 1, \\ \gamma_0 + \gamma_1 + \gamma_2 & \textbf{if } T = 2 \end{cases}$$

is an unbiased estimator of θ.

Recall that an estimator $\hat{\theta}$ is unbiased if then $E[\hat{\theta}] = \theta$, that is, the expected value of the estimator is equal to the true value of the parameter. So we have

$$E[\hat{\theta}] = \sum_{k=0}^{2} \hat{\theta}_k P(\hat{\theta} = \hat{\theta}_k)$$

where $\hat{\theta}_k$ is the value of $\hat{\theta}$ when $T = k$. Thus

$$\begin{aligned} E[\hat{\theta}] &= \gamma_0 P(T=0) + \frac{1}{2}(\gamma_1 + 2\gamma_0) P(T=1) + (\gamma_0 + \gamma_1 + \gamma_2) P(T=2) \\ &= \gamma_0 \times (1-p)^2 + \frac{1}{2}(\gamma_1 + 2\gamma_0) \times 2p(1-p) + (\gamma_0 + \gamma_1 + \gamma_2) \times p^2 \\ &= \gamma_0(1 - 2p + p^2 + 2p - 2p^2 + p^2) + \gamma_1(p - p^2 + p^2) + \gamma_2 p^2 \\ &= \gamma_0 + \gamma_1 p + \gamma_2 p^2 \\ &= \theta \end{aligned}$$

Hence $\hat{\theta}$ is unbiased.

3. Show that when $\theta = 2^p$ there is no unbiased estimator of θ from T.
Hint: Note that $E\left(\hat{\theta}(T)\right) = (1-p)^2\hat{\theta}(0) + 2p(1-p)\hat{\theta}(1) + p^2\hat{\theta}(2)$**.**

Note that the expectation is equal to a second-order polynomial in p of order 2:

$$\begin{aligned} E\left(\hat{\theta}(T)\right) &= (1-p)^2\hat{\theta}(0) + 2p(1-p)\hat{\theta}(1) + p^2\hat{\theta}(2) \\ &= \hat{\theta}(0) + (2\hat{\theta}(1) - 2\hat{\theta}(0))p + (\hat{\theta}(0) - 2\hat{\theta}(1) + \hat{\theta}(2))p^2 \end{aligned}$$

which can never be equal to 2^p.

Problem 2: 10 points

Recall that $X \sim U([a,b))$ is uniformly distributed in an interval $[a,b)$ if its density function is supported in $[a,b)$ and uniform:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b), \\ 0 & \text{otherwise .} \end{cases}$$

Let $X_1, \ldots, X_n$ be iid uniform variables $\sim U([0,\theta))$ with $\theta > 0$. Our goal is to estimate θ.

1. Show that the likelihood function for θ is

$$l(\theta) = \begin{cases} \frac{1}{\theta^n} & \textbf{if } \theta \geq \max_i X_i, \\ 0 & \textbf{otherwise}. \end{cases}$$

Recall that the likelihood function

$$l(\theta) = \prod_{i=1}^{n} f_\theta(X_i) = \begin{cases} \prod_{i=1}^{n} \frac{1}{\theta - 0} & \text{if } X_i \leq \theta \text{ for all } i = 1, 2, ..., n \\ 0 & \text{otherwise} \end{cases}$$

However, notice that the condition that $X_i \leq \theta$ for all $i = 1, 2, ..., n$ is equivalent to the condition that $\theta \geq \max_i X_i$, so

$$l(\theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } \theta \geq \max_i X_i \\ 0 & \text{otherwise} \end{cases}$$

2. Show that $\hat{\theta}_{ML} = \max_i X_i$.

Note that the likelihood is certainly not maximized when $\theta < \max_i X_i$, since it attains its minimum value of 0 over this range. We do notice however, that $\frac{1}{\theta^n}$ tends to zero as $\theta \to \infty$ (in particular, $f(x) = \frac{1}{x^n}$ is a decreasing function). Thus, the likelihood will be at its maximum for the smallest value of θ for which the likelihood is non-zero. That is, when $\theta = \max_i X_i$. So

$$\hat{\theta}_{ML} = \max_i X_i$$

3. Is $\hat{\theta}_{ML}$ biased? (*Hint: If X is a non-negative random variable and $F(x)$ is its cumulative distribution, then* $E[X] = \int_0^\infty (1 - F(x))dx$).

To see whether $\hat{\theta}_{ML}$ is biased, we check to see whether or not $E[\hat{\theta}_{ML}] = \theta$. Using the hint provided, we have that

$$\begin{aligned} E[\hat{\theta}_{ML}] &= E[\max_i X_i] \\ &= \int_0^\infty 1 - F_{\max_i X_i}(x) dx \qquad (*) \end{aligned}$$

So we need to identify $F_{\max_i X_i}(x)dx$. Note that

$$\begin{aligned}
F_{\max_i X_i}(x) &= P(\max_i X_i \leq x) \\
&= P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x) \\
&= P(X_1 \leq x)P(X_2 \leq x)...P(X_n \leq x) && \text{(By independence)} \\
&= (P(X_1 \leq x))^n && \text{(By identical distirbutions)}
\end{aligned}$$

Moreover, since $X_1 \sim U(0, \theta]$, we have

$$P(X_1 \leq x) = \begin{cases} 0 & x < 0 \\ \frac{x}{\theta} & x \in [0, \theta) \\ 1 & x > \theta \end{cases}$$

from which it follows that

$$F_{\max_i X_i}(x) = \begin{cases} 0 & x < 0 \\ \frac{x^n}{\theta^n} & x \in [0, \theta) \\ 1 & x > \theta \end{cases}$$

Thus substituting into (*), we have that

$$\begin{aligned}
E[\hat{\theta}_{ML}] &= \int_0^\infty 1 - F_{\max_i X_i}(x)dx \\
&= \int_0^\theta 1 - \frac{x^n}{\theta^n}dx \\
&= x - \frac{x^{n+1}}{(n+1)\theta^n}\Big|_0^\theta \\
&= \theta - \frac{\theta}{n+1} \\
&= \frac{n}{n+1}\theta
\end{aligned}$$

4. Show that $\hat{\theta}_2 = \frac{n+1}{n} \max_i X_i$ is unbiased.

Recall from the previous question that

$$E(\max_i X_i) = \frac{n}{n+1}\theta$$

which, by the linearity of expectation, implies that

$$\begin{aligned}
E[\hat{\theta}_2] &= E\left[\frac{n+1}{n} \max_i X_i\right] \\
&= \frac{n+1}{n}\frac{n}{n+1}\theta \\
&= \theta
\end{aligned}$$

so $\hat{\theta}_2$ is unbiased.

Problem 3: 10 points

Let $X_1 \sim \mathcal{N}(\theta_1, 1)$ and $X_2 \sim \mathcal{N}(\theta_2, 1)$ be independent Normal random variables, where θ_1 and θ_2 are unknown.

1. $Y \sim \mathbf{Exp}(\lambda)$ follows the exponential distribution if

$$P(0 \leq Y \leq \alpha) = 1 - \exp\left(-\lambda\alpha\right) .$$

Show that $Y = (X_1 - \theta_1)^2 + (X_2 - \theta_2)^2 \sim \mathbf{Exp}(1/2)$.

Note that since $X_1 \sim N(\theta_1, 1)$ and $X_2 \sim N(\theta_2, 1)$, we must have that

$$Y_1 := X_1 - \theta_1 \sim N(0, 1)$$

and

$$Y_2 := X_2 - \theta_2 \sim N(0, 1)$$

So that

$$Y = Y_1^2 + Y_2^2$$

is the sum of two standard normal random variables.

Next, since the joint density of Y_1 and Y_2 is just the product of their densities by independence, we have that

$$\begin{aligned}
P(0 \leq Y \leq \alpha) &= \int_{y_1^2+y_2^2 \leq \alpha} \frac{e^{-\frac{y_1^2}{2}}}{\sqrt{2\pi}} \frac{e^{-\frac{y_2^2}{2}}}{\sqrt{2\pi}} dy_1 dy_2 \\
&= \int_{y_1^2+y_2^2 \leq \alpha} \frac{e^{-\frac{y_2^2+y_1^2}{2}}}{2\pi} dy_1 dy_2 \\
&= \int_0^{\sqrt{\alpha}} \int_0^{2\pi} r \frac{e^{-\frac{r^2}{2}}}{2\pi} d\theta dr \qquad \text{(Change of variables to Polar coordinates)} \\
&= \int_0^{\sqrt{\alpha}} r e^{-\frac{r^2}{2}} dr \\
&= -e^{-\frac{r^2}{2}} \Big|_0^{\sqrt{\alpha}} \\
&= 1 - e^{-\alpha/2}
\end{aligned}$$

which implies that $Y \sim Exp(1/2)$

2. Show that both the square S and the circle C in $\mathbb{R}^2$, given by

$$S = \{(\theta_1, \theta_2) \,:\, |\theta_1 - X_1| \leq q(1 - 0.0125); \, |\theta_2 - X_2| \leq q(1 - 0.0125); \} \ ,$$

$$C = \{(\theta_1, \theta_2) \,:\, (X_1 - \theta_1)^2 + (X_2 - \theta_2)^2 \leq 5.991\}$$

are 95% confidence regions for (θ_1, θ_2). (q is the quantile function for the Standard Normal distribution).

A confidence region 95% confidence region, A, for our parameter vector (θ_1, θ_2) satisfies that

$$P((\theta_1, \theta_2) \in A) = 0.95$$

Moreover, recall that $Y_i := X_i - \theta_i \sim N(0,1)$ for $i = 1, 2$. Thus

$$
\begin{aligned}
P((\theta_1, \theta_2) \in S) &= P(|Y_1| \leq q(1-0.0125), |Y_2| \leq q(1-0.0125)) \\
&= P(|Y_1| \leq q(1-0.0125))P(|Y_2| \leq q(1-0.0125)) && \text{(Since } Y_1 \text{ and } Y_2 \text{ are independent)} \\
&= (P(|Y_1| \leq q(1-0.0125)))^2 && \text{(Since } Y_1 \text{ and } Y_2 \text{ are identically distributed)} \\
&= (0.975)^2 \approx 0.95
\end{aligned}
$$

Next, recall that we know $(X_1 - \theta_1)^2 + (X_2 - \theta_2)^2 \sim Exp\left(\frac{1}{2}\right)$, so

$$
\begin{aligned}
P((\theta_1, \theta_2) \in C) &= P((X_1 - \theta_1)^2 + (X_2 - \theta_2)^2 \leq 5.991) \\
&= 1 - e^{\frac{-5.991}{2}} && \text{(By the CDF of the exponential distribution)} \\
&= 0.95
\end{aligned}
$$

3. Which region would you pick and why?

Although both confidence regions have 95% coverage probability, they do not have the same surface. The square S has surface $(2 * q(1-0.0125))^2 = (4.472)^2 = 20$, whereas the circle C has a surface of $\pi * 5.991 \approx 18.82$. Therefore, we should pick C, the confidence region with smallest surface, because it provides more information about (θ_1, θ_2).