Homework 1 STAT 6202

Yang Liu
Instructor: Professor Tapan K. Nayak

January 26, 2019

1.

Let $X_1, \cdots, X_n$ be i.i.d. Bernoulli variables with success probability θ, when $n > 2$, and let $T = \sum_{i=1}^n X_i$. Derive the conditional distribution $X_1, \cdots, X_n$ given $T = t$.

Proof. Since $X_1, \cdots, X_n \overset{i.i.d.}{\sim} Bernoulli(\theta)$, and $T = \sum_{i=1}^n X_i \sim Binomial(n, \theta)$

$$P(X_1 = x_1, \cdots, X_n = x_n) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}$$

$$P\left(X_1 = x_1, \cdots, X_n = x_n, T = \sum_{i=1}^n X_i = t\right) = \theta^t (1-\theta)^{n-t}$$

$$P\left(X_1 = x_1, \cdots, X_n = x_n | \sum_{i=1}^n X_i = t\right) = \frac{\theta^t(1-\theta)^{n-t}}{\begin{pmatrix} n \\ t \end{pmatrix} \theta^t (1-\theta)^{n-t}} = \frac{1}{\begin{pmatrix} n \\ t \end{pmatrix}}$$

□

2.

Suppose X_1 and X_2 are iid $Poisson(\theta)$ random variables and let $T = X_1 + 2X_2$.

(a) Find the conditional distribution of (X_1, X_2) given $T = 7$.

(b) For $\theta = 1$ and $\theta = 2$, respectively, calculate all probabilities in the above conditional distribution and present the two conditional distributions numerically.

Proof. (a) Since $X_1, X_2 \overset{i.i.d.}{\sim} Poisson(\theta)$, then we have

$$\{X_1 + 2X_2 = 7\} = \{(X_1 = 1, X_2 = 3), (X_1 = 3, X_2 = 2), (X_1 = 5, X_2 = 1), (X_1 = 7, X_2 = 0)\}$$

and $(X_1 = 1, X_2 = 3), (X_1 = 3, X_2 = 2), (X_1 = 5, X_2 = 1), (X_1 = 7, X_2 = 0)$ are mutually exclusive, then

$$
\begin{aligned}
P\left(T=7\right) &= P\left(X_1=1, X_2=3\right) + P\left(X_1=3, X_2=2\right) \\
&\quad + P\left(X_1=5, X_2=1\right) + P(X_1=7, X_2=0) \\
&= \frac{\theta}{1}e^{-\theta}\cdot\frac{\theta^3}{3!}e^{-\theta} + \frac{\theta^3}{3!}e^{-\theta}\cdot\frac{\theta^2}{2!}e^{-\theta} + \frac{\theta^5}{5!}e^{-\theta}\cdot\frac{\theta^1}{1!}e^{-\theta} + \frac{\theta^7}{7!}e^{-\theta}\cdot e^{-\theta} \\
&= \frac{\theta^4 e^{-2\theta}}{6}\left(1+\frac{\theta}{2}+\frac{\theta^2}{20}+\frac{\theta^3}{840}\right)
\end{aligned}
$$

Then the conditional distribution of (X_1, X_2) given $T=7$ is

$$
\begin{aligned}
P\left(X_1=1, X_2=3 \middle| T=7\right) &= \frac{P\left(X_1=1, X_2=3\right)}{P\left(T=7\right)} \\
&= \frac{\frac{\theta^4 e^{-2\theta}}{6}}{\frac{\theta^4 e^{-2\theta}}{6}\left(1+\frac{\theta}{2}+\frac{\theta^2}{20}+\frac{\theta^3}{840}\right)} \\
&= \frac{840}{840+420\theta+42\theta^2+\theta^3} \\
P\left(X_1=3, X_2=2 \middle| T=7\right) &= \frac{P\left(X_1=3, X_2=2\right)}{P\left(T=7\right)} \\
&= \frac{\frac{\theta^4 e^{-2\theta}}{6}\cdot\frac{\theta}{2}}{\frac{\theta^4 e^{-2\theta}}{6}\left(1+\frac{\theta}{2}+\frac{\theta^2}{20}+\frac{\theta^3}{840}\right)} \\
&= \frac{420\theta}{840+420\theta+42\theta^2+\theta^3} \\
P\left(X_1=5, X_2=1 \middle| T=7\right) &= \frac{P\left(X_1=5, X_2=1\right)}{P\left(T=7\right)} \\
&= \frac{\frac{\theta^4 e^{-2\theta}}{6}\cdot\frac{\theta^2}{20}}{\frac{\theta^4 e^{-2\theta}}{6}\left(1+\frac{\theta}{2}+\frac{\theta^2}{20}+\frac{\theta^3}{840}\right)} \\
&= \frac{42\theta^2}{840+420\theta+42\theta^2+\theta^3} \\
P\left(X_1=7, X_2=0 \middle| T=7\right) &= \frac{P\left(X_1=7, X_2=0\right)}{P\left(T=7\right)} \\
&= \frac{\frac{\theta^4 e^{-2\theta}}{6}\cdot\frac{\theta^3}{840}}{\frac{\theta^4 e^{-2\theta}}{6}\left(1+\frac{\theta}{2}+\frac{\theta^2}{20}+\frac{\theta^3}{840}\right)} \\
&= \frac{\theta^3}{840+420\theta+42\theta^2+\theta^3}
\end{aligned}
$$

(b) The conditional distribution of $(X_1, X_2)|T=7$ is given in table 1.

Table 1: Conditional distribution of (X_1, X_2) given $T=7$

$P(X_1=x_1, X_2=x_2 \mid T=7)$	$(x_1=1, x_2=3)$	$(x_1=3, x_2=2)$	$(x_1=5, x_2=1)$	$(x_1=7, x_2=0)$
$\theta=1$	$\frac{840}{1303}$	$\frac{420}{1303}$	$\frac{42}{1303}$	$\frac{1}{1303}$
$\theta=2$	$\frac{840}{1856}$	$\frac{840}{1856}$	$\frac{169}{1856}$	$\frac{8}{1856}$

□

3.

Let $X_1, \cdots, X_n$ be i.i.d. random variables with mean μ and variance σ^2. Let $\bar{X}$ denote the sample mean and $V = \sum_{i=1}^n \left(X_i - \bar{X}\right)^2$.

(a) Derive the expected value of $\bar{X}$ and V.

(b) Further suppose that $X_1, \cdots, X_n$ are normally distributed. Let $A_{n\times n} = ((a_{ij}))$ be an orthogonal matrix whose
first row is $(\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}})$ and let $Y = AX$, where $Y = (Y_1, \cdots, Y_n)'$ and $X = (X_1, \cdots, X_n)$ are (column) vectors. (It is not necessary to know a_{ij} for $i = 2, \cdots, n$, $j = 1, \cdots, n$ for answering the following questions.)

(i) Find $\sum_{j=1}^n a_{ij}$ for $i = 1, \cdots, n$ and show that $\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n X_i^2$ (Use properties of orthogonal matrices.)

(ii) Express $\bar{X}$ and V in terms (or as functions) of $Y_1, \cdots, Y_n$.

(iii) Use (only) *transformation of variables* approach to find the joint distribution of $Y_1, \cdots, Y_n$. Are $Y_1, \cdots, Y_n$ independently distributed and what are their marginal distributions?

(iv) Prove that $\bar{X}$ and V are independent given their marginal distributions.

Proof. (a) Since $E[X_i] = \mu$ and $Var[X_i] = \sigma^2$ for $i = 1, \cdots, n$

$$E[\bar{X}] = E\left[\frac{\sum_{i=1}^m X_i}{n}\right] = \frac{\sum_{i=1}^n E[X_i]}{n} = \frac{n\mu}{n} = \mu$$

$$\begin{aligned} Var\left[\bar{X}\right] &= E\left[\left(\bar{X}-\mu\right)^2\right] = Var\left[\frac{\sum_{i=1}^n X_i}{n}\right] \\ &= \frac{\sum_{i=1}^n Var[X_i]}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

$$\begin{aligned} E\left[V\right] &= E\left[\sum_{i=1}^n \left(X_i - \bar{X}\right)^2\right] = E\left[\sum_{i=1}^n \left((X_i-\mu) - (\bar{X}-\mu)\right)^2\right] \\ &= E\left[\sum_{i=1}^n (X_i-\mu)^2\right] + nE\left[\left(\bar{X}-\mu\right)^2\right] - 2E\left[\sum_{i=1}^n (X_i-\mu)(\bar{X}-\mu)\right] \\ &= E\left[\sum_{i=1}^n (X_i-\mu)^2\right] - nE\left[\left(\bar{X}-\mu\right)^2\right] \\ &= nVar[X_i] - nVar[\bar{X}] \\ &= n\sigma^2 - n\cdot\frac{\sigma^2}{n} = (n-1)\sigma^2 \end{aligned}$$

Or since

$$\begin{aligned} E\left[\bar{X}^2\right] &= Var[\bar{X}] + E[\bar{X}^2] \\ &= \frac{\sigma^2}{n} + \mu^2 \end{aligned}$$

$$
\begin{aligned}
E[V] &= E\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] = E\left[\sum_{i=1}^{n}X_i^2 - 2\sum_{i=1}^{n}X_i\bar{X}^2 + n\bar{X}^2\right] \\
&= E\left[\sum_{i=1}^{n}X_i^2 - n\bar{X}^2\right] \\
&= n\cdot\left[\sigma^2+\mu^2\right] - n\cdot\left[\frac{\sigma^2}{n}+\mu^2\right] \\
&= (n-1)\sigma^2
\end{aligned}
$$

(b) (i) Due to the orthogonality of A, $A'A = AA' = I_{n\times n}$ ($I_{n\times n}$ is diagonal matrix of 1's. Let $A = (a_{1\cdot}, \cdots, a_{n\cdot})'$ where $a_{j\cdot}$ is the j^{th} row vector. Then we have for $i, j = 1, \cdots, n$

$$
a_{i\cdot}a_{i\cdot}' = 1 \; and \; a_{i\cdot}a_{j\cdot}' = 0
$$

$$
\begin{aligned}
a_{1\cdot}a_{1\cdot}' &= \sum_{k=1}^{n}\frac{1}{\sqrt{n}}\cdot\frac{1}{\sqrt{n}} = \frac{\sum_{j=1}^{n}a_{1j}}{\sqrt{n}} = 1 \\
a_{i\cdot}a_{1\cdot}' &= \sum_{j=1}^{n}a_{ij}\cdot\frac{1}{\sqrt{n}} = \frac{\sum_{j=1}^{n}a_{ij}}{\sqrt{n}} = 0
\end{aligned}
$$

Hence $\sum_{j=1}^{n}a_{ij} = \sqrt{n}$ for $j = 1$ and $\sum_{j=1}^{n}a_{ij} = 0$ for $j = 2, \cdots, n$.

$$
\begin{aligned}
\sum_{i=1}^{n}Y_i^2 &= Y'Y = X'A'AX = X'(A'A)X \\
&= X'X = \sum_{i=1}^{n}X_i^2
\end{aligned}
$$

(ii) Note that $Y_1 = \sum_{i=1}^{n}\frac{1}{\sqrt{n}}\cdot X_i = \frac{\sum_{i=1}^{n}X_i}{\sqrt{n}} = \sqrt{n}\cdot\bar{X}$

$$
\begin{aligned}
\sum_{i=2}^{n}Y_i^2 &= Y'Y - Y_1^2 = \sum_{i=1}^{n}X_i^2 - \left(\sqrt{n}\bar{X}\right)^2 \\
&= \sum_{i=1}^{n}X_i - n\bar{X}^2 \\
&= \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2
\end{aligned}
$$

Therefore $\bar{X} = \frac{Y_1}{\sqrt{n}}$ and $\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 = \sum_{i=2}^{n}Y_i^2$

(iii) since $X_1, \cdots, X_n \overset{i.i.d.}{\sim} N\left(\mu, \sigma^2\right)$

$$
\begin{aligned}
f_{X_1,\cdots,X_n}\left(x_1,\cdots,x_n\right) &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
&= (2\pi)^{-\frac{n}{2}}\,\sigma^{-n}e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} \\
&= (2\pi)^{-\frac{n}{2}}\,\sigma^{-n}e^{-\frac{(X-\mathbf{1}\mu)'(X-\mathbf{1}\mu)}{2\sigma^2}}
\end{aligned}
$$

where $\mathbf{1} = (1, \cdots, 1)'$. Let $A = (a_{\cdot 1}, \cdots, a_{\cdot n})$, $A' = (a_{\cdot 1}, \cdots, a_{\cdot n})'$ where $a_{\cdot j}$ is the j^{th} column vector, since $Y = AX$, $X = A'AX = A'Y$, $\frac{d}{dY}X = A$, $\left|\frac{d}{dY}X\right| = \det(A) = \det(A'A) = 1$, then we have

$$\begin{aligned}
f_{Y_1,\cdots,Y_n}(y_1, \cdots, y_n) &= f_{X_1,\cdots,X_n}(x_1, \cdots, x_n) \left| \left| \frac{d}{dY} X \right| \right|_{X=A'Y} \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{(X-\mu)'(X-\mu)}{2\sigma^2}} \Big|_{X=A'Y} \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{(A'Y-\mathbf{1}\mu)'(A'Y-\mathbf{1}\mu)}{2\sigma^2}} \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{(Y-A\mathbf{1}\mu)'AA'(Y-A\mathbf{1}\mu)}{2\sigma^2}} \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{(Y-A\mathbf{1}\mu)'(Y-A\mathbf{1}\mu)}{2\sigma^2}} \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{\sum_{i=1}^n (y_i - a_{i\cdot}\mathbf{1}\mu)^2}{2\sigma^2}} \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{\sum_{i=1}^n \left(y_i - \sum_{j=1}^n a_{ij}\mu\right)^2}{2\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_1-\sqrt{n}\mu)^2}{2\sigma^2}} \cdot \prod_{i=2}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y_i^2}{2\sigma^2}}
\end{aligned}$$

The last equation is due to (b) (i). Note that $E[Y] = AE[X] = A\mathbf{1}\mu$, $Var[Y] = A'Var[X]A = A'A\sigma^2 = I \cdot \sigma^2$, therefore $Y_1 \sim N(\sqrt{n}\mu, \sigma^2) \perp Y_2, \cdots, Y_n \overset{i.i.d.}{\sim} N(0, \sigma^2)$
Or

$$\begin{aligned}
\sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n X_i^2 - 2\mu \sum_{i=1}^n X_i + n\mu^2 \\
&= \sum_{i=1}^n Y_i^2 - 2\sqrt{n}Y_i + n\mu^2 \\
&= \sum_{i=2}^n Y_i^2 + (Y_1 - \sqrt{n}\mu)^2
\end{aligned}$$

The second equation is due to (b) (i). Hence

$$\begin{aligned}
f_{Y_1,\cdots,Y_n}(y_1, \cdots, y_n) &= f_{X_1,\cdots,X_n}(x_1, \cdots, x_n) \left| \left| \frac{d}{dY} X \right| \right|_{X=A'Y} \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_1-\sqrt{n}\mu)^2}{2\sigma^2}} \cdot \prod_{i=2}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y_i^2}{2\sigma^2}}
\end{aligned}$$

(iv) Since $Y_1 \sim N(\sqrt{n}\mu, \sigma^2) \perp Y_2, \cdots, Y_n \overset{i.i.d.}{\sim} N(0, \sigma^2)$, $\bar{X} = \frac{Y_1}{\sqrt{n}} \sim N\left(\mu, \frac{\sigma^2}{\sqrt{n}}\right) \perp Y_2, \cdots, Y_n$ and $\frac{Y_2^2}{\sigma^2}, \cdots, \frac{Y_n^2}{\sigma^2} \overset{i.i.d.}{\sim} \chi_1^2$, then $\sum_{i=2}^n \frac{Y_i^2}{\sigma^2} \sim \chi_{n-1}^2$. Therefore $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{\sqrt{n}}\right) \perp V = \sum_{i=2}^n Y_i^2 \sim \sigma^2 \cdot \chi_{n-1}^2$

□

4.

Consider a large population of individuals and let θ denote the (unknown) proportion of the population belonging to a sensitive group A (e.g. drug users).

Suppose, we randomly select n individuals from the population and ask each person to select a card from a deck and answer the question written on the card. Each card in the deck has one of the two questions: Q_1: Do you belong to A? and Q_2: Do you not belong to A? Also, 85% percent of the cards ask Q_1 and the remaining 15% ask Q_2.

Assume that each person answers Yes or No truthfully to the selected question. For $i = 1, \cdots, n$, let $X_i = 1$ if the i^{th} person answers 'Yes' otherwise $X_i = 0$. So, the data are the observed values of $X_1, \cdots, X_n$.

Give the joint distribution of $X_1, \cdots, X_n$ and the distribution of the total number of Yes responses.

Proof. We first consider to calculate the probability for the i^{th} person to answer 'Yes'

$$\begin{aligned} P(X_i = 1) = \quad & P(answer\ Q_1) \cdot P\left('Yes'\ as\ response \middle|\ answer\ Q_1\right) \\ & + P(answer\ Q_2) \cdot P\left('Yes'\ as\ response \middle|\ answer\ Q_2\right) \\ & = 0.85 \times \theta + 0.15 \times (1 - \theta) \\ & = 0.15 + 0.7\theta \end{aligned}$$

Then we have $X_1, \cdots, X_n \overset{i.i.d.}{\sim} Bernoulli\,(0.15 + 0.7\theta)$, therefore

$$\begin{aligned} P\left(X_1 = x_1, \cdots, X_n = x_n\right) &= \prod_{i=1}^{n} (0.15 + 0.7\theta)^{x_i} \left(1 - (0.15 + 0.7\theta)\right)^{1 - x_i} \\ &= (0.15 + 0.7\theta)^{\sum_{i=1}^{n} x_i} (0.85 - 0.7\theta)^{n - \sum_{i=1}^{n} x_i} \end{aligned}$$

Let $Y_n = \sum_{i=1}^{n} X_i$ be the total number of 'Yes' response, then $Y_n \sim Binomial\,(n, 0.15 + 0.7\theta)$

$$P\left(Y_n = y\right) = \left(\begin{array}{c} n \\ y \end{array}\right) (0.15 + 0.7\theta)^y (0.85 - 0.7\theta)^{n-y} \quad for\ y = 0, \cdots, n$$

□