

Solution to Tutorial Week 3

STAT3023: Statistical Inference

Semester 2, 2022

1. (a) We have $\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$. We have proved in the lecture that $X_i \sim \text{Bin}(m, p_i)$, so $E(X_i) = mp_i$ for any i , so we only need to compute $E(X_i X_j)$.

We have proved in the lecture that $X_i \sim \text{Bin}(m, p_i)$ for any $i = 1, \dots, n$, so we will use the argument that a binomial random variable is the sum of independent Bernoulli random variables. More specifically, we can write $X_i = \sum_{k=1}^m X_{ik}$ and $X_j = \sum_{l=1}^m X_{jl}$ where $X_{ik} = 1$ if the k th trial results in the i th outcome, and $X_{jl} = 1$ if the l th trials results in the j th outcome, and 0 otherwise.

Therefore,

$$\begin{aligned}\text{Cov}(X_i, X_j) &= \text{Cov}\left(\sum_{k=1}^m X_{ik}, \sum_{l=1}^m X_{jl}\right) \\ &= \sum_{k=1}^m \sum_{l=1}^m \text{Cov}(X_{ik}, X_{jl}).\end{aligned}$$

Note that if $j \neq l$, then $\text{Cov}(X_{ik}, X_{jl}) = 0$ since the k th and the j th trials are independent. Hence, we can reduce the above double sum to be

$$\text{Cov}(X_i, X_j) = \sum_{k=1}^m \text{Cov}(X_{ik}, X_{jk}) = \sum_{k=1}^m \{E(X_{ik} X_{jk}) - E(X_{ik})E(X_{jk})\}$$

For the k th trial, the product $X_{ik} X_{jk}$ is always equal to zero, because we can only observe only one outcome (i.e at most only one of X_{ik} and X_{jk} can be one), hence $E(X_{ik} X_{jk}) = 0$. Furthermore, since X_{ik} is Bernoulli with probability p_i , then $E(X_{ik}) = p_i$. As a result, we have

$$\text{Cov}(X_i, X_j) = \sum_{k=1}^m \text{Cov}(X_{ik}, X_{jk}) = \sum_{k=1}^m (-p_i p_j) = -mp_i p_j.$$

- (b) We have

$$\begin{aligned}P(X_1 = x_1, X_2 = x_2) &= \sum_{(x_3, \dots, x_n) \in \mathcal{A}} p(x_1, \dots, x_n) \\ &= \frac{m!}{x_1! x_2!} p_1^{x_1} p_2^{x_2} \sum_{(x_3, \dots, x_n) \in \mathcal{A}} \frac{1}{x_3! \dots x_n!} p_3^{x_3} \dots p_n^{x_n}\end{aligned}\tag{1}$$

$p_3 + \dots + p_n = 1 - p_1 - p_2$, and over the set \mathcal{A} , we have $x_3 + \dots + x_n = m - x_1 - x_2$. The multinomial theorem gives

$$(1 - p_1 - p_2)^{m - x_1 - x_2} = (p_3 + \dots + p_n)^{m - x_1 - x_2} = \sum_{(x_3, \dots, x_n) \in \mathcal{A}} \frac{(m - x_1 - x_2)!}{x_3! \dots x_n!} p_3^{x_3} \dots p_n^{x_n},$$

so we obtain

$$\sum_{(x_3, \dots, x_n) \in \mathcal{A}} \frac{1}{x_3! \dots x_n!} p_3^{x_3} \dots p_n^{x_n} = \frac{(1 - p_1 - p_2)^{m - x_1 - x_2}}{(m - x_1 - x_2)!}.$$

Substituting it into (1), we have

$$P(X_1 = x_1, X_2 = x_2) = \frac{m!}{x_1! x_2! (m - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{m - x_1 - x_2}.$$

We can recognize that this probability is the joint pmf for a multinomial distribution with m trials and the cell probabilities $\mathbf{p}_{1,2} = (p_1, p_2, 1 - p_1 - p_2)$.

(c) By definition of conditional probability, we have

$$P(X_3 = x_3, \dots, X_n = x_n | X_1 = x_1, X_2 = x_2) = \frac{P(X_1 = x_1, \dots, X_n = x_n)}{P(X_1 = x_1, X_2 = x_2)}.$$

The numerator is the joint pmf of \mathbf{X} at (x_1, \dots, x_n) , so it is equal to

$$P(X_1 = x_1, \dots, X_n = x_n) = p(x_1, \dots, x_n) = \frac{m!}{x_1! \dots x_n!} p_1^{x_1} \dots p_n^{x_n}.$$

The denominator is the marginal distribution of X_1 and X_2 we computed. Therefore,

$$\begin{aligned} & P(X_3 = x_3, \dots, X_n = x_n | X_1 = x_1, X_2 = x_2) \\ &= \frac{\frac{p_3^{x_3} \dots p_n^{x_n}}{x_3! \dots x_n!}}{\frac{(1 - p_1 - p_2)^{m - x_1 - x_2}}{(m - x_1 - x_2)!}} \\ &= \frac{(m - x_1 - x_2)!}{x_3! \dots x_n!} \left(\frac{p_3}{1 - p_1 - p_2} \right)^{x_3} \dots \left(\frac{p_n}{1 - p_1 - p_2} \right)^{x_n}. \end{aligned}$$

This is the joint pdf of a multinomial distribution with $m - x_1 - x_2$ trials and the cell probabilities

$$\mathbf{p}_{-1,-2} = \left(\frac{p_3}{1 - p_1 - p_2}, \dots, \frac{p_n}{1 - p_1 - p_2} \right)$$

2. Since $g(x) = F_X(x)$ is a monotone increasing function, so is its inverse g^{-1} . Hence, for any $0 < y < 1$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\{F_X(X) \leq y\} \\ &= P\{X \leq F_X^{-1}(y)\} \\ &\stackrel{(i)}{=} F_X(F_X^{-1}(y)) = y, \end{aligned}$$

where step (i) follows from the definition of F_X . Hence,

$$f_Y(y) = F_Y'(y) = 1,$$

so Y follows a continuous uniform distribution on $(0, 1)$.

3. Y is a discrete random variable and has realized values on the integer set. For any $y = 1, 2, \dots$, the probability mass function of Y is given by

$$\begin{aligned}
 P(Y = y) &= P(y-1 \leq X \leq y) = \int_{y-1}^y f_X(x) dx \\
 &= \int_{y-1}^y e^{-x} dx \\
 &= -e^{-x} \Big|_{x=y-1}^{x=y} \\
 &= e^{-(y-1)} - e^{-y} = e^{-(y-1)}(1 - e^{-1})
 \end{aligned}$$

This is the pmf of a Geometric distribution, with $p = 1 - e^{-1}$.

4. The transformation between (X, Y) and (Z, Y) is one-to-one, since we can recover (X, Y) from (Z, Y) . In fact, we have $X = Z/Y^2$. The Jacobian of the transformation is given by

$$\begin{bmatrix} \frac{\partial X}{\partial Z} & \frac{\partial X}{\partial Y} \\ \frac{\partial Y}{\partial X} & \frac{\partial Y}{\partial Y} \end{bmatrix} = \begin{bmatrix} 1/Y^2 & -Z/2Y^3 \\ 0 & 1 \end{bmatrix}$$

and the corresponding determinant is $|J| = 1/Y^2$. Hence, the joint density of (Y, Z) is given by

$$f_{Z,Y}(z, y) = f_{X,Y}(z, y)|J| = 12 \frac{z}{y^2} y(1-y) \frac{1}{y^2} = \frac{12z(1-y)}{y^3}.$$

Since the range of X and Y are between 0 and 1, this above joint density is only defined when $0 < y < 1$ and $0 \leq z/y^2 < 1$, i.e $0 < y < 1$ and $0 < z < y^2$. This range is critical in establishing the density of Z as

$$\begin{aligned}
 f_Z(z) &= \int_{\sqrt{z}}^1 f_{Z,Y}(z, y) dy = 12z \int_{\sqrt{z}}^1 \frac{1-y}{y^3} dy = 12z \int_{\sqrt{z}}^1 (y^{-3} - y^{-2}) dy \\
 &= 12z \left(\frac{y^{-2}}{-2} - \frac{y^{-1}}{-1} \right) \Big|_{y=\sqrt{z}}^{y=1} \\
 &= 12z \left(-\frac{1}{2} + 1 + \frac{1}{2z} - \frac{1}{\sqrt{z}} \right) \\
 &= 6z + 6 - 12\sqrt{z}, \quad 0 < z < 1.
 \end{aligned}$$

5. (a) Since U and V are independent, then

$$\begin{aligned}
 f_{U,V}(u, v) &= f_U u f_V(v) = \frac{1}{\Gamma\left(\frac{m}{2}\right) 2^{m/2}} u^{\frac{m}{2}-1} e^{-\frac{u}{2}} \frac{1}{\Gamma\left(\frac{n}{2}\right) \cdot 2^{n/2}} v^{\frac{n}{2}-1} e^{-\frac{v}{2}} \\
 &= \frac{1}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right) 2^{(m+n)/2}} u^{\frac{m}{2}-1} v^{\frac{n}{2}-1} e^{-\frac{u+v}{2}}.
 \end{aligned}$$

Consider the transformation from (U, V) to $X = \frac{U/m}{V/n}$ and $Y = V$. This bivariate transformation is one-to-one, since we can recover (U, V) from (X, Y) .

Specifically, we would have $U = (m/n)XY$ and $V = Y$. The corresponding Jacobian is given by

$$J = \begin{bmatrix} \frac{\partial U}{\partial X} & \frac{\partial U}{\partial Y} \\ \frac{\partial V}{\partial X} & \frac{\partial V}{\partial Y} \end{bmatrix} = \begin{bmatrix} \frac{m}{n}Y & \frac{m}{n}X \\ 0 & 1 \end{bmatrix}$$

so its determinant is $|J| = \frac{m}{n}Y$. Hence, the joint density of (X, Y) is given by

$$\begin{aligned} f_{X,Y}(x, y) &= f_{U,V}(x, y)|J| \\ &= \frac{1}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{(m+n)/2}} \left(\frac{m}{n}xy\right)^{\frac{m}{2}-1} y^{\frac{n}{2}-1} e^{-\frac{(m/n)xy+y}{2}} \frac{m}{n}y \\ &= \frac{1}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{(m+n)/2}} \left(\frac{m}{n}\right)^{m/2} x^{\frac{m}{2}-1} y^{\frac{m+n}{2}-1} e^{-\frac{(m/n)xy+y}{2}}. \end{aligned}$$

Finally, we can get the density of X by integrating Y out of the joint density.

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X,Y}(x, y)dy \\ &= \frac{1}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{(m+n)/2}} \left(\frac{m}{n}\right)^{m/2} x^{\frac{m}{2}-1} \int_0^\infty y^{\frac{m+n}{2}-1} \exp\left\{-y\left(\frac{mx/n+1}{2}\right)\right\} dy. \end{aligned}$$

The last integral is the unnormalized density of a Gamma distribution with rate $\alpha = (m+n)/2$ and scale $\beta = 2/(mx/n+1)$. Hence,

$$\int_0^\infty y^{\frac{m+n}{2}-1} \exp\left\{-y\left(\frac{mx/n+1}{2}\right)\right\} dy = \Gamma\left(\frac{m+n}{2}\right) \left(\frac{2}{mx/n+1}\right)^{(m+n)/2}.$$

Substituting this into the density of X and rearranging terms, we get

$$f_X(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{m/2} x^{\frac{m}{2}-1} \left(1 + \frac{mx}{n}\right)^{-(m+n)/2}.$$

- (b) While we can derive the density of T^2 , an easier way to show it is to use the definition of the t and the F -distribution. Let Z be the standard normal random variable and V be a chi-square distribution with n degrees of freedom and be independent of Z . Then, we can write

$$T = \frac{Z}{\sqrt{V/n}}.$$

Therefore,

$$T^2 = \frac{Z^2}{V/n} = \frac{U}{V/n} = \frac{U/1}{V/n}$$

where $U = Z^2 \sim \chi_1^2$. Since Z is independent of V , so is U . Comparing it with the definition of an F distribution, we can conclude $T^2 \sim F_{1,n}$.