Yang Wenxiao.

Homework 3

Homework Policy

- You must turn in your own work, do not copy answers from others. This is not helpful to you at all.
- You are welcome and encouraged to work together on homework. If you work with others, write the following on the top of your assignment: I worked with .... on this assignment.
- I am fully aware that it is possible to use online resources and merely copy the derived solutions. This does not help you to learn the material and is going to be detrimental for you in the exams.
- Show all supporting work.

Concepts:

- Multinomial Distribution
- Multivariate Normal Distribution
- Multivariate Change of Variables

Exercises

1. **Multinomial Distribution**(3 points)

   Consider the birthdays of 50 people. Assume people's birthdays are independent and the 365 days of the year are equally likely. Find the covariance (1 point) and correlation (2 points) of how many of the people were born on January 1 and how many were born on January 2.

2. **Multinomial and Binomial**(3 points)

   In humans (and many other organisms), genes come in pairs. Consider a gene of interest, which comes in two types (alleles): type a and type A. The genotype of a person for that gene is the types of the two genes in the pair: AA, Aa, or aa (aA is equivalent to Aa). According to the Hardy-Weinberg law, for a population in equilibrium the frequencies of AA, Aa, aa will be p^2, $2p(1-p)$ and $(1-p)^2$ respectively, for some p with $0 < p < 1$. Suppose that the Hardy-Weinberg law holds, and that n people are drawn randomly from the population, independently. Let X_1, X_2, X_3 be the number of people in the sample with genotypes AA, Aa, aa, respectively.

   (a) What is the joint PMF of (X_1, X_2, X_3)? (1 point)

   (b) What is the distribution of the number of people in the sample who have an A? (1 point)

(c) What is the distribution of how many of the $2n$ genes among the people are As? (1 point. This may require some thinking.)

3. **Bivariate Normal Distribution**(3 points)

Let (X, Y) be Bivariate Normal with $X \sim N(0, \sigma_1^2)$ and $Y \sim N(0, \sigma_2^2)$ marginally and with $Corr(X, Y) = \rho$.

(a) Show that $X, Y - cX$ also follows a Bivariate Normal distribution for any constant c. (1 point)

(b) Find a constant c such that $Y - cX$ is independent of X. (2 points)

Hint: Find c such that $Cov(X, Y - cX) = 0$. Now argue why that also ensures independence.

4. **Change of Variables**(3 points)

Let X, Y be i.i.d $N(0, 1)$ random variables and (R, θ) be the polar coordinates for the random point X, Y. So, $X = R\,Cos\theta$ and $Y = R\,Sin\theta$ with $R \geq 0$ and $\theta \in [0, 2\pi)$.

Hint: This is the inverse of the Box Muller example done in the lecture. Define g to be the map that takes (x, y) to (R, θ). Then g is invertible and continuously differentiable. The inverse map is $g^{-1}(R, \theta) = (R\,Cos\theta, R\,Sin\theta)$. Find jacobian of the map g^{-1}. Now use change of variables theorem for part (a).

(a) Find the joint pdf of R and θ. (1 point)

(b) Are R and θ independent? (1 point)

(c) Also find the marginal distributions of R^2 and θ, giving their names if they are distributions we have studied before. (1 point)

The maximum you can score in this homework is 12. This homework is worth 10 points in the overall scheme of things as will be every homework. If you score 11 or 12 you will earn 1 or 2 extra credit points respectively.

1. $X \sim Mult(50, \quad P \quad), \quad P_i = \frac{1}{365} \quad i = 1, 2, \cdots 365$

$X_1 \sim Bin(50, \frac{1}{365}) \qquad X_2 \sim Bin(50, \frac{1}{365})$

$$Cov(X_1, X_2) = \frac{1}{2}(Var(X_1 + X_2) - Var X_1 - Var X_2)$$

$$= \frac{1}{2}\left(50 \frac{2}{365} \frac{363}{365} - 50 \frac{1}{365} \frac{364}{365} \times 2\right)$$

$$= -\frac{2}{5329}$$

$$\rho_{X_1, X_2} = \frac{Cov(X_1, X_2)}{\sqrt{Var X_1 \, Var X_2}} = \frac{-50 \times \cancel{\frac{1}{365}} \times \cancel{\frac{1}{365}}}{\sqrt{\frac{2}{365} \times \cancel{\frac{363}{365}} \times \cancel{\frac{1}{365}} \times \cancel{\frac{364}{365}}}} = \frac{-25}{11\sqrt{546}}$$

2. (a) $(X_1, X_2, X_3) \sim Mult(n, (p^2, 2p(1-p), (1-p)^2))$

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1! \, x_2! \, x_3!} (p^2)^{x_1} (2p(1-p))^{x_2} ((1-p)^2)^{x_3}$$

$$= \frac{n!}{x_1! \, x_2! \, x_3!} 2^{x_2} p^{2x_1 + x_2} (1-p)^{x_2 + 2x_3}$$

(b) : $X_A = X_1 + X_2 \sim Bin(n, p^2 + 2p(1-p))$

$$P\{X_A = k\} = \binom{n}{k} (2p - p^2)^k (p^2 - 2p + 1)^{n-k}$$

(c)

$$P\{Y = k\} = P\{2X_1 + X_2 = k\}$$

(k is even:)

$$= \sum_{k_1=0}^{\frac{k}{2}} P\{X_1 = k_1,\ X_2 = k-2k_1,\ X_3 = n-k+k_1\}$$

$$= \sum_{k_1=0}^{\frac{k}{2}} \frac{n!}{k_1!\,(k-2k_1)!\,(n-k+k_1)!} 2^{k_1} p^{k+k_1} (1-p)^{2n-k}$$

$$= n!\,(1-p)^{2n-k} p^k \sum_{k_1=0}^{\frac{k}{2}} \frac{2^{k_1} p^{k_1}}{k_1!\,(k-2k_1)!\,(n-k+k_1)!}$$

(k is odd)

$$= \sum_{k_1=0}^{\frac{k-1}{2}} P\{X_1 = k_1,\ X_2 = k-2k_1,\ X_3 = n-k+k_1\}$$

$$= n!\,(1-p)^{2n-k} p^k \sum_{k_1=0}^{\frac{k-1}{2}} \frac{2^{k_1} p^{k_1}}{k_1!\,(k-2k_1)!\,(n-k+k_1)!}$$

3. (a) Linear combination of $X, Y-cX$, $\forall c \in \mathbb{R}$.

$aX + b(Y-cX) = (a-bc)X + bY$, $\forall a, b \in \mathbb{R}$

is also linear combination of X, Y

$(X, Y) \sim BVN \Rightarrow$ Every linear combination of X, Y has normal distribution $\Rightarrow$ Every linear combination of $X, Y-cX$ has normal distribution

$\Rightarrow (X, Y-cX) \sim BVN$. $\forall c \in \mathbb{R}$.

(b) $Cov(X, Y-cX) = Cov(X,Y) - c\,Var(X)$.

$= \rho\sigma_1\sigma_2 - c\,\sigma_1^2 = 0 \implies c = \frac{\sigma_2}{\sigma_1}\rho$

Since $(X, Y-\frac{\sigma_2}{\sigma_1}\rho X) \sim BVN$,

and $Cov(X - Y - \frac{\sigma_2}{\sigma_1}\rho X) = 0$, we can conclude

X, $Y-\frac{\sigma_2}{\sigma_1}\rho X$ are independent.

4. (a) $g(X,Y) = (R,\theta)$, $g^{-1}(R,\theta) = (R\cos\theta, R\sin\theta)$.

$$J_{(R,\theta)} = \begin{vmatrix} \frac{\partial x}{\partial R} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial R} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -R\sin\theta \\ \sin\theta & R\cos\theta \end{vmatrix} = R$$

$$P(R,\theta) = \frac{\partial^2 F(R,\theta)}{\partial R\partial\theta} = \frac{\partial^2 F(X(R,\theta), Y(R,\theta))}{dx \cdot dy} \cdot R$$

$$= f(x,y)\cdot R = \frac{R}{2\pi} e^{-\frac{x^2+y^2}{2}} = \frac{R}{2\pi} e^{-\frac{R^2}{2}}$$

(b) Yes $f_{R|\theta}(r|t_1) = f_{R|\theta}(r|t_2)$

$$= \frac{r}{2\pi} e^{-\frac{r^2}{2}}$$

(c) $$P_\theta(\theta) = \int_0^{+\infty} \frac{R}{2\pi} e^{-\frac{R^2}{2}} dR = \frac{1}{2\pi} \sim U(0, 2\pi).$$

$$p_R(R) = \int_0^{2\pi} \frac{R}{2\pi} e^{-\frac{R^2}{2}} d\theta = R e^{-\frac{R^2}{2}}$$