Probability and Statistics – Lancaster/BJTU 2017 Final Exam Paper A with Answer

1 Fill in the blanks (20 points)

1. Suppose $P(A) = 0.3$, $P(B) = 0.8$, and $P(A) = P(A|B)$. Then $P(B^c|A^c) = \underline{0.2}$.
2. Of the following six families of distributions: binomial, Poisson, negative binomial, uniform, normal, and Gamma, there are two families that a random variable having distribution in them may take negative values with positive probability. These two families of distributions are the <u>uniform</u> distributions and the <u>normal</u> distributions.
3. Suppose that there is a binomial distribution with parameter n and p, where n is large. 1) If np is moderate, then it can be approximated by a <u>Poisson</u> distribution. 2) If np is large, then according to the <u>central limit</u> theorem, it can be approximated by a <u>normal</u> distribution.
4. The lower and upper limits of correlation coefficients are <u>-1</u> and <u>1</u>, respectively.
5. Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2, then an unbiased and consistent estimator of μ is $\underline{\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}}$, and an unbiased and consistent estimator of σ^2 is $\underline{s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2}$.

2

(8 points) Suppose that a box contains r red balls and w white balls. Suppose also that balls are drawn from the box one at a time, at random, without replacement.
(a) What is the probability that all red balls will be obtained before any white balls are obtained?
(b) What is the probability that the first two balls drawn will be of the same color?

Solution: (a)

$$\frac{1}{\binom{r+w}{r}} = \frac{r!w!}{(r+w)!}$$

(b)

$$\frac{r(r-1) + w(w-1)}{(r+w)(r+w-1).}$$

3

(12 points) Two students, Adam and Brian, are going to take a test. They take the test independently of each other. The probability for Adam to pass the test is a, and the probability for Brian to pass the test is b. Let X be the number of people of these two students who pass the test. Prove that $\mathrm{Var}(X) = a - a^2 + b - b^2$.

Solution Method 1: Let A be the random variable such that $A = 1$ if Adam passes the test and $A = 0$ otherwise. Let B be the random variable such that $B = 1$ if Brian passes the test and

$B = 0$ otherwise. Then A and B are independent, A has the Bernoulli distribution with parameter a, and B has the Bernoulli distribution with parameter b. It is also obvious that

$$X = A + B.$$

It then follows that

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}(A) + \mathrm{Var}(B) \\ &= a(1-a) + b(1-b). \end{aligned}$$

Solution Method 2: Let A be the event that Adam passes the test, and let B be the event that Brian passes the test. Obviously X can only be 0,1,2.

$$P(X = 0) = P(A^c \cap B^c) = P(A^c)P(B^c) = (1-a) \times (1-b) = 1 - a - b + ab$$

$$P(X = 2) = P(A \cap B) = P(A)P(B) = ab$$

$$P(X = 1) = 1 - P(X = 0) - P(X = 2) = 1 - (1 - a - b + ab) - ab = a + b - 2ab.$$

Therefore

$$E(X) = (1 - a - b + ab) \times 0 + (a + b - 2ab) \times 1 + ab \times 2 = a + b$$

$$E(X^2) = (1 - a - b + ab) \times 0 + (a + b - 2ab) \times 1 + ab \times 4 = a + b + 2ab$$

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = a + b + 2ab - (a + b)^2 = a - a^2 + b - b^2.$$

4

(12 points) Suppose that random variables X and Y are **independent**. X has the uniform distribution on the interval $[0, 1]$, and Y has the exponential distribution with parameter β. Suppose also that $E(X) = E(Y)$.

(a) Determine the value of β.

(b) Find $P(X > Y)$.

Solution: (a) $E(X) = \frac{1}{2}$ and $E(Y) = \frac{1}{\beta}$, therefore $\beta = 2$.

(b) The joint p.d.f of X and Y is

$$f(x, y) = \begin{cases} 2e^{-2y} & \text{for } 0 \le x \le 1 \text{ and } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Thus

$$P(X > Y) = \int_0^1 dx \int_0^x 2e^{-2y} dy = \int_0^1 (1 - e^{-2x}) dx = \frac{1 + e^{-2}}{2}.$$

5

(8 points) Suppose that X and Y are independent random variables that both have the Poisson distribution and are such that $\mathrm{Var}(X) + \mathrm{Var}(Y) = 5$. Evaluate $P(X + Y < 2)$.

Solution: Let $X \sim \mathrm{Poisson}(\lambda_1)$ and $Y \sim \mathrm{Poisson}(\lambda_2)$. Since X and Y are independent, it follows from the additivity of Poisson distributions that $X + Y \sim \mathrm{Poisson}(\lambda_1 + \lambda_2)$. Since $\lambda_1 = \mathrm{Var}(X)$ and $\lambda_2 = \mathrm{Var}(Y)$, we have $\lambda_1 + \lambda_2 = 5$, hence

$$P(X + Y < 2) = (1 + 5)e^{-5} = 0.0404276819945128...$$

6

(14 points) A random sample $X_1, \ldots, X_n$ is to be taken from a distribution with mean μ and standard deviation σ^2 .

(a) Use the Chebyshev inequality to determine the smallest n such that the following relation will be satisfied:

$$P(|\overline{X}_n - \mu| < \frac{\sigma}{4}) > 0.99$$

(b) Suppose additionally that the random sample $X_1, \ldots, X_n$ has normal distribution. Determine the smallest n such that the above relation will be satisfied.

Solution: (a) $\overline{X}_n$ has mean μ and variance $\frac{\sigma^2}{n}$, therefore

$$P(|\overline{X}_n - \mu| \geq \frac{\sigma}{4}) \leq \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{16}} = \frac{16}{n},$$

and

$$P(|\overline{X}_n - \mu| < \frac{\sigma}{4}) > 1 - \frac{16}{n}.$$

To allow the relation to hold, we need to have $1 - \frac{16}{n} \geq 0.99$, i.e. $n \geq 1600$.

(b) Now $\overline{X}_n$ has the normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, hence

$$P(|\overline{X}_n - \mu| < \frac{\sigma}{4}) = P(\frac{|\overline{X}_n - \mu|}{\frac{\sigma}{\sqrt{n}}} < \frac{\frac{\sigma}{4}}{\frac{\sigma}{\sqrt{n}}}) = 2\Phi(\frac{\sqrt{n}}{4}) - 1.$$

To allow the relation to hold, we need to have $2\Phi(\frac{\sqrt{n}}{4}) - 1 > 0.99$, i.e. $\Phi(\frac{\sqrt{n}}{4}) > 0.995$, i.e. $\frac{\sqrt{n}}{4} > 2.58$, i.e. $n \geq 107$.

7

(14 points) Suppose that X and Y have the **bivariate normal distribution** with covariance matrix $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$. The real number a is such that

$$aX + Y \text{ and } aX - Y \text{ are independent, and}$$

$$\mathrm{Var}(aX + Y) = 1.$$

(a) Determine the value of a.

(b) Compute $\mathrm{Var}(aX - Y)$.

Solution: (a) Since X and Y are bivariate normal, we know that $aX + Y$ and $aX - Y$ are bivariate normal too. Therefore they are independent if and only if $\mathrm{Cov}(aX + Y, aX - Y) = 0$. Since

$$\mathrm{Cov}(aX + Y, aX - Y) = a^2\mathrm{Var}(X) - \mathrm{Var}(Y) = a^2 - 1,$$

it then follows that $a^2 = 1$. On the other hand,

$$1 = \mathrm{Var}(aX + Y) = a^2\mathrm{Var}(X) + \mathrm{Var}(Y) + 2a\mathrm{Cov}(X, Y) = a^2 + 1 + a,$$

which implies $a^2 + a = 0$. Therefore, $a = -1$.

(b)

$$\mathrm{Var}(aX - Y) = \mathrm{Var}(-X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y) = 1 + 1 + 1 = 3.$$

8

(12 points) Suppose that $X_1, \ldots, X_n$ form a random sample from a distribution for which the p.d.f. is as follows:

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1} & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

where $\theta > 1$ is the parameter to be estimated. Derive the M.L.E. of θ.

Solution: For observed values $x_1, \ldots, x_n$, the likelihood function is

$$L(\theta) = \theta^n (\prod_{i=1}^n x_i)^{\theta-1}.$$

$$\log L(\theta) = n\log(\theta) + (\theta - 1)\log(\prod_{i=1}^n x_i).$$

The maximizer of $\log L(\theta)$ is the solution of the following equation:

$$\tfrac{d\log L}{d\theta} = \tfrac{n}{\theta} + \log(\textstyle\prod_{i=1}^n x_i) = 0 \;\; ,$$

$$\theta = \frac{-n}{\log(\prod_{i=1}^n x_i}).$$

Thus the maximum likelihood estimator of θ is $\hat{\theta}_{MLE} = \frac{-n}{\log(\prod_{i=1}^n X_i)}$.