Probability and Statistics
Answer of Mock Exam 2020

Yiping Cheng

December 11, 2020

1. (20pt) Fill in the blanks:

(1) A and B are two events with $P(A) = 0.4$, $P(A \cup B) = 0.7$, then $P(B|\overline{A}) = \underline{\mathbf{0.5}}$.

(2) Three people are assigned randomly and independently into 4 rooms numbered A–D. Then the expected number of people in room A is $\underline{\mathbf{\frac{3}{4}}}$ and the probability that room A has exactly 2 people is $\underline{\mathbf{\frac{9}{64}}}$.

(3) The PDF of X which has the standard normal distribution is $\varphi(x) = \underline{\mathbf{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}}$.

(4) Suppose that random variables X and Y are independent, $X \sim Exp(1)$ and $Y \sim U[0,1]$. Then the joint PDF of (X, Y) is $f(x, y) = \underline{\begin{cases} e^{-x}, & \text{if } x > 0 \text{ and } 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}}$, and $\mathrm{Var}(X - Y) = \underline{\mathbf{\frac{13}{12}}}$.

(5) Of the following families of distributions: binomial, Poisson, geometric, uniform, normal, and exponential, the **Poisson** and **normal** distributions are completely additive. A distribution family is additive if X and Y are independent and belong to this distribution family, then $X + Y$ also belongs to this distribution family.

(6) For an estimator (actually a sequence of estimators) to be practically usable, if must be **consistent**. For any distribution, the unbiased and most efficient estimator of population expectation is sample mean, or $\mathbf{\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i}$.

2. (10pt) Suppose that random variables X and Y are independent, $X \sim N(0, 1)$ and $Y \sim U[0, 1]$. Let

$$Z = \begin{cases} X, & \text{if } Y < 0.4, \\ 2X - 1, & \text{otherwise.} \end{cases}$$

Find the CDF and PDF of Z. (You can simply write the CDF and PDF of the standard normal distribution by $\Phi(\cdot)$ and $\varphi(\cdot)$, respectively.)

Solution: Denote the CDF and PDF of Z by $F_Z(\cdot)$ and $f_Z(\cdot)$, respectively.

$$F_Z(x) = P(Z \le x)$$

$$= P(Y < 0.4)P(Z \le x | Y < 0.4) + P(Y \ge 0.4)P(Z \le x | Y \ge 0.4)$$

$$= 0.4\, P(Z \le x | Y < 0.4) \,+\, 0.6\, P(Z \le x | Y \ge 0.4)$$

$$= 0.4\,P(X \le x|Y < 0.4)\; +\; 0.6\,P(2X - 1 \le x|Y \ge 0.4)$$

$$= 0.4\,P(X \le x|Y < 0.4)\; +\; 0.6\,P(X \le \frac{x+1}{2}|Y \ge 0.4)$$

$$= 0.4\,P(X \le x)\; +\; 0.6\,P(X \le \frac{x+1}{2})$$

$$= 0.4\,\Phi(x)\; +\; 0.6\Phi(\frac{x+1}{2}).$$

And

$$f_Z(x) = F_Z'(x) = 0.4\,\varphi(x)\; +\; 0.3\varphi(\frac{x+1}{2}).$$

3. (12pt) The PDF of X is given by

$$f(x) = \begin{cases} 1 - |x|, & \text{if } -1 < x < 1, \\ 0, & \text{otherwise}. \end{cases}$$

Let $Y = \cos(\frac{\pi X}{2})$. Find the PDF of Y.

Solution (Method 1): Let $G(y)$, $g(y)$ denote the CDF and PDF of Y, respectively.
1) If $y \le 0$, then $G(y) = 0$.
2) If $y \ge 1$, then $G(y) = 1$.
3) If $0 < y < 1$,

$$G(y) = P(Y \le y) = P(\cos(\frac{\pi X}{2}) \le y)$$

$$= P(-\frac{\pi}{2} \le \frac{\pi X}{2} \le -\arccos y \text{ or } \arccos y \le \frac{\pi X}{2} \le \frac{\pi}{2})$$

$$= P(-\frac{\pi}{2} \le \frac{\pi X}{2} \le -\arccos y) + P(\arccos y \le \frac{\pi X}{2} \le \frac{\pi}{2})$$

$$= P(-1 \le X \le -\frac{2}{\pi}\arccos y) + P(\frac{2}{\pi}\arccos y \le X \le 1)$$

$$= F(-\frac{2}{\pi}\arccos y) - F(-1) + F(1) - F(\frac{2}{\pi}\arccos y).$$

Hence $g(y) = G'(y)$

$$= f(-\frac{2}{\pi}\arccos y)\frac{2}{\pi\sqrt{1-y^2}} + f(\frac{2}{\pi}\arccos y)\frac{2}{\pi\sqrt{1-y^2}}$$

$$= [f(-\frac{2}{\pi}\arccos y) + f(\frac{2}{\pi}\arccos y)]\frac{2}{\pi\sqrt{1-y^2}}$$

$$= 2(1 - \frac{2}{\pi}\arccos y)\frac{2}{\pi\sqrt{1-y^2}}$$

$$= (1 - \frac{2}{\pi}\arccos y)\frac{4}{\pi\sqrt{1-y^2}}.$$

To sum up,

$$g(y) = \begin{cases} (1 - \frac{2}{\pi}\arccos y)\frac{4}{\pi\sqrt{1-y^2}}, & \text{if } 0 < y < 1, \\ 0, & \text{otherwise}. \end{cases}$$

Solution (Method 2): Let $g(y)$ denote the PDF of Y. Let $U = |X|$, since cos is an even function, we have $Y = \cos(\frac{\pi U}{2})$. Let $h(u)$ be the PDF of U, then

$$h(u) = \begin{cases} 2(1-u), & \text{if } 0 \le u < 1, \\ 0, & \text{otherwise.} \end{cases}$$

The function $y = \cos(\frac{\pi u}{2})$ maps $(0,1)$ to $(0,1)$ one-to-one. Then for $0 < y < 1$, we have

$$\begin{aligned} g(y) &= h(\frac{2}{\pi}\arccos y)|\frac{d(\frac{2}{\pi}\arccos y)}{dy}| \\ &= 2(1 - \frac{2}{\pi}\arccos y)\frac{2}{\pi\sqrt{1-y^2}} \\ &= (1 - \frac{2}{\pi}\arccos y)\frac{4}{\pi\sqrt{1-y^2}}. \end{aligned}$$

To sum up,

$$g(y) = \begin{cases} (1 - \frac{2}{\pi}\arccos y)\frac{4}{\pi\sqrt{1-y^2}}, & \text{if } 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

4. (10pt) Suppose that random point (X, Y) is uniformly distributed in the disk $x^2 + y^2 < 9$.
(1) Find the conditional PDF $f_{Y|X}(y|x)$.
(2) Determine $P(Y > 0|X = 2)$.

Solution: (1) The joint PDF of X and Y is

$$f(x,y) = \begin{cases} \frac{1}{9\pi}, & \text{if } x^2+y^2 < 9, \\ 0, & \text{otherwise.} \end{cases}$$

Denote the PDF of X by $f_X(x)$.
If $x < -3$ or $x > 3$, then $f_X(x) = 0$.
If $-3 \le x \le 3$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\sqrt{9-x^2}}^{\sqrt{9-x^2}} \frac{1}{9\pi}dy = \frac{2}{9\pi}\sqrt{9-x^2}.$$

Hence for $-3 \le x \le 3$,

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f(x,y)}{f_X(x)} \\ &= \begin{cases} \frac{1}{2\sqrt{9-x^2}}, & \text{if } x^2+y^2 < 9, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

(2)

$$f_{Y|X}(y|2) = \begin{cases} \frac{1}{2\sqrt{5}}, & \text{if } -\sqrt{5} < y < \sqrt{5}, \\ 0, & \text{otherwise.} \end{cases}$$

$$P(Y > 0|X = 2) = \int_0^{\infty} f_{Y|X}(y|2)dy$$

$$= \int_0^{\sqrt{5}} \frac{1}{2\sqrt{5}} dy$$
$$= \frac{1}{2}.$$

5. (14pt) The joint PDF of X and Y is given by

$$f(x,y) = \begin{cases} x+y, & \textbf{if } 0 < x < 1 \textbf{ and } 0 < y < 1, \\ 0, & \textbf{otherwise}. \end{cases}$$

Find
(1) $E(X)$ and $E(Y)$;
(2) $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$;
(3) $\mathrm{Cov}(X,Y)$.

Solution: (1)

$$E(X) = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x,y) dy dx$$
$$= \int_0^1 x \int_0^1 (x+y) dy dx$$
$$= \int_0^1 x(xy + \frac{y^2}{2})|_0^1 dx$$
$$= \int_0^1 x(x+\frac{1}{2}) dx = \int_0^1 x^2 dx + \int_0^1 \frac{1}{2} x dx$$
$$= \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

By symmetry of X and Y in $f(x,y)$, we have $E(Y) = E(X) = \frac{7}{12}$.
(2)

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \int_{-\infty}^{\infty} f(x,y) dy dx$$
$$= \int_0^1 x^2 \int_0^1 (x+y) dy dx$$
$$= \int_0^1 x^2(xy + \frac{y^2}{2})|_0^1 dx$$
$$= \int_0^1 x^2(x+\frac{1}{2}) dx = \int_0^1 x^3 dx + \int_0^1 \frac{1}{2} x^2 dx$$
$$= \frac{1}{4} + \frac{1}{6} = \frac{5}{12}.$$

$$\mathrm{Var}(X) = E(X^2) - E^2(X) = \frac{5}{12} - (\frac{7}{12})^2 = \frac{11}{144}.$$

By symmetry of X and Y in $f(x,y)$, we have $\mathrm{Var}(Y) = \mathrm{Var}(X) = \frac{11}{144}$.
(3)

$$E(XY) = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} y f(x,y) dy dx$$

$$= \int_0^1 x \int_0^1 (xy + y^2) dy dx$$
$$= \int_0^1 x(\frac{xy^2}{2} + \frac{y^3}{3})|_0^1 dx$$
$$= \int_0^1 x(\frac{x}{2} + \frac{1}{3}) dx = \int_0^1 \frac{x^2}{2} dx + \int_0^1 \frac{1}{3} x dx$$
$$= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{1}{3} - \frac{7}{12} \cdot \frac{7}{12} = -\frac{1}{144}.$$

6. (10pt) Suppose that random variables X, Y and Z have $E(X) = E(Y) = 1$, $E(Z) = -1$, $\mathrm{Var}(X) = \mathrm{Var}(Y) = \mathrm{Var}(Z) = 1$, $\rho_{X,Y} = 0$, $\rho_{X,Z} = \frac{1}{2}$, $\rho_{Y,Z} = -\frac{1}{2}$. Let $W = X - Y + Z$. Find $E(W)$ and $\mathrm{Var}(W)$.

Solution:
$$E(W) = E(X) - E(Y) + E(Z) = 1 - 1 - 1 = -1.$$

We have
$$\mathrm{Cov}(X, Y) = \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}\rho_{X,Y} = 0,$$
$$\mathrm{Cov}(X, Z) = \sqrt{\mathrm{Var}(X)\mathrm{Var}(Z)}\rho_{X,Z} = \frac{1}{2},$$
$$\mathrm{Cov}(Y, Z) = \sqrt{\mathrm{Var}(Y)\mathrm{Var}(Z)}\rho_{Y,Z} = -\frac{1}{2}.$$

Therefore,

$$\mathrm{Var}(W) = \mathrm{Var}(X - Y + Z)$$
$$= \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z) - 2\mathrm{Cov}(X, Y) + 2\mathrm{Cov}(X, Z) - 2\mathrm{Cov}(Y, Z)$$
$$= 1 + 1 + 1 - 2 \cdot 0 + 2 \cdot \frac{1}{2} - 2 \cdot (-\frac{1}{2}) = 5.$$

7. (12pt) An insurance company runs a certain disease insurance policy, which has 10,000 policy holders. Each year, each policy holder pays the company a premium of $\$170$, and if he gets the disease in that year, he receives from the company a compensation of $\$20,000$. The probability for a person to get the disease in a year is 0.006. Using the central limit theorem, for this company approximately evaluate

(1) the probability that the annual profit of this policy is at least $\$200,000$;

(2) the probability that the annual profit of this policy is positive.

Solution: Let X be the number of policy holders who get the disease in the year. Then $X \sim B(10000, 0.006)$, and $E(X) = 60$ and $Var(X) = 59.64$.

Let Y be the annual profit of this policy in 10^6 dollars. Then $Y = 1.7 - 0.02X$.

(1) The annual profit of this policy being at least $\$200,000$ is equivalent to $Y \geq 0.2$, which is equivalent to $X \leq 75$. By the central limit theorem, the desired probability is

$$P(X \leq 75) = P(X < 75.5) = P(\frac{X - 60}{\sqrt{59.64}} < \frac{75.5 - 60}{\sqrt{59.64}})$$

$$\approx \Phi(\frac{75.5-60}{\sqrt{59.64}}) = \Phi(2.007072) = 0.977629.$$

(2) The annual profit of this policy being positive is equivalent to $Y > 0$, which is equivalent to $X < 85$, i.e. $X \leq 84$. By the central limit theorem, the desired probability is

$$P(X \leq 84) = P(X < 84.5) = P(\frac{X-60}{\sqrt{59.64}} < \frac{84.5-60}{\sqrt{59.64}})$$

$$\approx \Phi(\frac{84.5-60}{\sqrt{59.64}}) = \Phi(3.172468) = 0.999244.$$

Note: Solutions without using correction for continuity are also considered correct.

8. (12pt) Let $X_1, \cdots, X_n$ be a random sample from $B(100, p)$.
(1) Derive the maximum likelihood estimator of p.
(2) Find the bias of the estimator you just derived. Is it unbiased?

Solution: (1) For observed values $x_1, \ldots, x_n$, the likelihood function is

$$L(p) = \prod_{i=1}^{n} C_{100}^{x_i} p^{x_i}(1-p)^{100-x_i} = \prod_{i=1}^{n} C_{100}^{x_i} \cdot p^{\sum_{i=1}^n x_i}(1-p)^{100n-\sum_{i=1}^n x_i}.$$

$$\ln L(p) = \ln(\prod_{i=1}^{n} C_{100}^{x_i}) + \ln p \sum_{i=1}^{n} x_i + \ln(1-p)\,(100n - \sum_{i=1}^{n} x_i).$$

The maximizer of $\ln L(p)$ is the solution of the following equation:

$$0 = \frac{d \ln L}{dp} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}(100n - \sum_{i=1}^{n} x_i),$$

$$p = \frac{\sum_{i=1}^n x_i}{100n}.$$

Thus the maximum likelihood estimator of p is

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{100n}.$$

(2)

$$E(\hat{p}) = \frac{\sum_{i=1}^n E(X_i)}{100n} = \frac{\sum_{i=1}^n 100p}{100n} = \frac{100pn}{100n} = p.$$

Thus the bias is $E(\hat{p}) - p = 0$, and $\hat{p}$ is unbiased.