MSO 201a: Probability and Statistics

2019-20-II Semester

Assignment-V

Instructor: Neeraj Misra

1. For $\mu \in \mathbb{R}$ and $\lambda > 0$, let $X_{\mu,\lambda}$ be a random variable having the p.d.f.

$$f_{\mu,\sigma}(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x-\mu}{\lambda}}, & \text{if } x \geq \mu \\ 0, & \text{otherwise} \end{cases}.$$

(a) Find $C_r(\mu, \lambda) = E((X - \mu)^r), r \in \{1, 2, \ldots\}$ and $\mu'_r(\mu, \lambda) = E(X^r_{\mu,\lambda}), r \in \{1, 2\}$;

(b) For $p \in (0, 1)$, find the p-th quantile $\xi_p \equiv \xi_p(\mu, \lambda)$ of the distribution of $X_{\mu,\lambda}$ ($F_{\mu,\lambda}(\xi_p) = p$, where $F_{\mu,\lambda}$ is the distribution function of $X_{\mu,\lambda}$);

(c) Find the lower quartile $q_1(\mu, \lambda)$, the median $m(\mu, \lambda)$ and the upper quartile $q_3(\mu, \lambda)$ of the distribution of $X_{\mu,\lambda}$;

(d) Find the mode $m_0(\mu, \lambda)$ of the distribution of $X_{\mu,\sigma}$;

(e) Find the standard deviation $\sigma(\mu, \lambda)$, the mean deviation about median MD$(m(\mu, \lambda))$, the inter-quartile range IQR(μ, λ), the quartile deviation (or semi-inter-quartile range) QD(μ, λ), the coefficient of quartile deviation CQD(μ, λ) and the coefficient of variation CV(μ, λ) of the distribution of $X_{\mu,\lambda}$;

(f) Find the coefficient of skewness $\beta_1(\mu, \lambda)$ and the Yule coefficient of skewness $\beta_2(\mu, \lambda)$ of the distribution of $X_{\mu,\lambda}$;

(g) Find the excess kurtosis $\gamma_2(\mu, \lambda)$ of the distribution of $X_{\mu,\lambda}$;

(h) Based on values of measures of skewness and the kurtosis of the distribution of $X_{\mu,\lambda}$, comment on the shape of $f_{\mu,\sigma}$.

2. Let X be a random variable with p.d.f. $f_X(x)$ that is symmetric about μ ($\in \mathbb{R}$), i.e., $f_X(x + \mu) = f_X(\mu - x)$, $\forall\ x \in (-\infty, \infty)$.

(a) If q_1, m and q_3 are respectively the lower quartile, the median and the upper quartile of the distribution of X then show that $\mu = m = \frac{q_1+q_3}{2}$;

(b) If $E(X)$ is finite then show that $E(X) = \mu = m = \frac{q_1+q_3}{2}$.

3. A fair dice is rolled six times independently. Find the probability that on two occasions we get an upper face with 2 or 3 dots.

4. Let n (≥ 2) and $r \in \{1, 2, , n-1\}$ be fixed integers and let $p \in (0,1)$ be a fixed real number. Using probabilistic arguments show that

$$\sum_{j=r}^{n} \binom{n}{j} p^j (1-p)^{n-j} - \sum_{j=r}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = \binom{n-1}{r} p^r (1-p)^{n-r}.$$

5. Let $X \sim \text{Bin}(n, p)$, where n is a positive integer and $p \in (0,1)$. Find the mode of the distribution of X.

6. Eighteen balls are placed at random in seven boxes that are labeled $B_1, \ldots, B_7$. Find the probability that boxes with labels B_1, B_2 and B_3 all together contain six balls.

7. Let T be a discrete type random variable with support $S_T = \{0, 1, 2, \ldots\}$. Show that T has the lack of memory property if, and only if, $T \sim \text{Ge}(p)$, for some $p \in (0,1)$.

8. A person repeatedly rolls a fair dice independently until an upper face with two or three dots is observed twice. Find the probability that the person would require eights rolls to achieve this.

9. (a) Consider a sequence of independent Bernoulli trials with probability of success in each trial being $p \in (0,1)$. Let Z denote the number of trials required to get the r-th success, where r is a given positive integer. Let $X = Z - r$. Find the marginal probability distributions of X and Z. For $r = 1$, show that $P(Z > m+n) = P(Z > m)P(Z > n)$, $m, n \in \{0, 1, \ldots\}$ (or equivalently $P(Z > m + n|Z > m) = P(Z > n)$, $m, n \in \{0, 1, \ldots\}$; this property is also known as the lack of memory property).

   (b) Two teams (say Team A and Team B) play a series of games until one team wins 5 games. If the probability of Team A (Team B) winning any game is 0.7 (0.3), find the probability that the series will end in 8 games.

10. **(Relation between binomial and negative binomial probabilities)** Let n be a positive integer, $r \in \{1, 2, \ldots, n\}$ and let $p \in (0,1)$. Using probabilistic arguments and also otherwise show that

$$\sum_{k=r}^{n} \binom{n}{k} p^k (1-p)^{n-k} = p^r \sum_{k=0}^{n-r} \cdot \binom{r+k-1}{k} (1-p)^k,$$

    i.e., for $r \in \{1, 2, \ldots, n\}$, $P(\text{Bin}(n,p) \geq r) = P(\text{NB}(r,p) \leq n-r)$.

11. A mathematician carries at all times two match boxes, one in his left pocket and one in his right pocket. To begin with each match box contains n matches. Each

time the mathematician needs a match he is equally likely to take it from either pocket. Consider the moment when the mathematician for the first time discovers that one of the match boxes is empty. Find the probability that at that moment the other box contains exactly k matches, where $k \in \{0, 1, \dots, n\}$.

12. Consider a population comprising of N (≥ 2) units out of which a ($\in \{1, 2, \dots, N-1\}$) are labeled as S (success) and $N - a$ are labeled as F (failure). A sample of size n ($\in \{1, 2, \dots, N-1\}$) is drawn from this population, drawing one unit at a time and without replacing it back into the population (i.e., sampling is without replacement). Let A_i ($i = 1, 2, \dots, n$) denote the probability of obtaining success (S) in the i-th trial. Show that $P(A_i) = \frac{a}{N}$, $i = 1, \dots, n$ (i.e., probability of obtaining S remains the same in each trial). **Hint:** Use induction by first showing that $P(A_1) = P(A_2) = \frac{a}{N}$.

13. (a) (**Binomial approximation to hypergeometric distribution**) Show that a hypergeometric distributed can be approximated by a $\text{Bin}(n, p)$ distribution provided N is large ($N \to \infty$), a is large ($a \to \infty$) and $a/N \to p$, as $N \to \infty$;
(b) Out of 120 applicants for a job, only 80 are qualified for the job. Out of 120 applicants, five are selected at random for the interview. Find the probability that at least two selected applicants will be qualified for the job. Using Problem 13 (a) find an approximate value of this probability.

14. Urn U_i ($i = 1, 2$) contains N_i balls out of which r_i are red and $N_i - r_i$ are black. A sample of n ($1 \leq n \leq N_1$) balls is chosen at random (without replacement) from urn U_1 and all the balls in the selected sample are transferred to urn U_2. After the transfer two balls are drawn at random from the urn U_2. Find the probability that both the balls drawn from urn U_2 are red.

15. Let $X \sim \text{Po}(\lambda)$, where $\lambda > 0$. Find the mode of the distribution of X.

16. (a) **(Poisson approximation to negative binomial distribution)** Show that $\lim_{r\to\infty} \binom{r+k-1}{k} p_r^r (1 - p_r)^k = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0, 1, 2, \dots$, provided $\lim_{r\to\infty} p_r = 1$ and $\lim_{r\to\infty}(r(1 - p_r)) = \lambda > 0$.

    (b) Consider a person who plays a series of 2500 games independently. If the probability of person winning any game is 0.002, find the probability that the person will win at least two games.

17. (a) If $X \sim \text{Po}(\lambda)$, find $E((2 + X)^{-1})$;
(b) If $X \sim \text{Ge}(p)$ and r is a positive integer, find $E(\min(X, r))$.

18. A person has to open a lock whose key is lost among a set of N keys. Assume that out of these N keys only one can open the lock. To open the lock the person tries keys one by one by choosing, at each attempt, one of the keys at random from the unattempted keys. The unsuccessful keys are not considered for future attempts. Let Y denote the number of attempts the person will have to make to open the lock. Show that $Y \sim U(\{1, 2, \ldots, N\})$ and hence find the mean and the variance of the r.v. Y.

19. Let $a > 0$ be a real constant. A point X is chosen at random on the interval $(0, a)$ (i.e., $X \sim U(0, a)$).
(a) If Y denotes the area of equilateral triangle having sides of length X, find the mean and variance of Y;
(b) If the point X divides the interval $(0, a)$ into subintervals $I_1 = (0, X)$ and $I_2 = [X, a)$, find the probability that the larger of these two subintervals is at least the double the size of the smaller subinterval.

20. Using a random observation $U \sim U(0, 1)$, describe a method to generate a random observation X from the distribution having
(a) probability density function

$$f(x) = \frac{e^{-|x|}}{2}, \quad -\infty < x < \infty;$$

(b) probability mass function

$$g(x) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x}, & \text{if } x \in \{0, 1, \ldots, n\} \\ 0, & \text{otherwise} \end{cases},$$

where $n \in \mathbb{N}$ and $\theta \in (0, 1)$ are real constants.

21. Let $X \sim U(0, \theta)$, where θ is a positive integer. Let $Y = X - [X]$, where $[x]$ is the largest integer $\le x$. Show that $Y \sim U(0, 1)$.

22. Let $U \sim U(0, 1)$ and let X be the root of the equation $3t^2 - 2t^3 - U = 0$. Show that X has p.d.f. $f(x) = 6x(1 - x)$, if $0 \le x \le 1$, $= 0$, otherwise.

23. **(Relation between gamma and Poisson probabilities)** For $t > 0, \theta > 0$ and $n \in \{1, 2, \ldots\}$, using integration by parts, show that

$$\frac{1}{\theta^n \Gamma(n)} \int_t^\infty e^{-\frac{x}{\theta}} x^{n-1} dx = \sum_{j=0}^{n-1} \frac{e^{-\frac{t}{\theta}} (\frac{t}{\theta})^j}{j!},$$

i.e., $P(G(n, \theta) > t) = P(\mathrm{Po}(\frac{t}{\theta}) \le n - 1)$.

24. Let Y be a random variable of continuous type with $F_Y(0) = 0$, where $F_Y(\cdot)$ is the distribution function of Y. Show that Y has the lack of memory property if, and only if, $Y \sim \text{Exp}(\theta)$, for some $\theta > 0$.

25. **(Relation between binomial and beta probabilities)** Let n be a positive integer, $k \in \{0, 1, \ldots, n\}$ and let $p \in (0, 1)$. If $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Be}(k, n-k+1)$, show that $P(\{X \geq k\}) = P(\{Y \leq p\})$.

26. Let $\underline{X} = (X_1, X_2)'$ have the joint p.d.f.

$$f_{\underline{X}}(x_1, x_2) = \phi(x_1)\phi(x_2)[1+\alpha(2\Phi(x_1)-1)(2\Phi(x_2)-1)], \ -\infty < x_i < \infty, \ i = 1, 2, \ |\alpha| \leq 1.$$

(a) Verify that $f_{\underline{X}}(x_1, x_2)$ is a p.d.f.;

(b) Find the marginal p.d.f.s of X_1 and X_2;

(c) Is (X_1, X_2) jointly normal?

27. Let $\underline{X} = (X_1, X_2)' \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ and, for constants a_1, a_2, a_3 and a_4 ($a_i \neq 0$, $i = 1, 2, 3, 4$, $a_1 a_4 \neq a_2 a_3$), let $Y = a_1 X_1 + a_2 X_2$ and $Z = a_3 X_1 + a_4 X_2$.

(a) Find the joint p.d.f. of (Y, Z);

(b) Find the marginal p.d.f.s. of Y and Z.

28. (a) Let $(X, Y)' \sim N_2(5, 8, 16, 9, 0.6)$. Find $P(\{5 < Y < 11\}|\{X = 2\})$, $P(\{4 < X < 6\})$ and $P(\{7 < Y < 9\})$;

(b) Let $(X, Y)' \sim N_2(5, 10, 1, 25, \rho)$, where $\rho > 0$. If $P(\{4 < Y < 16\}|\{X = 5\}) = 0.954$, determine ρ.

29. Let X and Y be independent and identically distributed $N(0, \sigma^2)$ r.v.s.

(a) Find the joint p.d.f. of (U, V), where $U = aX + bY$ and $V = bX - aY$ ($a \neq 0, b \neq 0$);

(b) Show that U and V are independent;

(c) Show that $\frac{X+Y}{\sqrt{2}}$ and $\frac{X-Y}{\sqrt{2}}$ are independent $N(0, \sigma^2)$ random variables.

30. Let $\underline{X} = (X_1, X_2)' \sim N_2(0, 0, 1, 1, \rho)$.

(a) Find the m.g.f. of $Y = X_1 X_2$;

(b) Using (a), find $E(X_1^2 X_2^2)$;

(c) Using conditional expectation, find $E(X_1^2 X_2^2)$.

31. Let $\underline{X} = (X_1, X_2)'$ have the joint p.d.f.

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

Show that marginals of $f(\cdot,\cdot)$ are each $N(0,1)$ but $f(\cdot,\cdot)$ is not the p.d.f. of a bivariate normal distribution.

32. Let $f_r(\cdot,\cdot)$, $-1 < r < 1$, denote the pdf of $N_2(0,0,1,1,r)$ and, for a fixed $\rho \in (-1,1)$, let the random variable (X,Y) have the joint p.d.f.

$$g_\rho(x,y) = \frac{1}{2}[f_\rho(x,y) + f_{-\rho}(x,y)].$$

Show that X and Y are normally distributed but the distribution of (X,Y) is not bivariate normal.

33. Consider the random vector (X,Y) as defined in Problem 32.

    (a) Find $\mathrm{Corr}(X,Y)$;

    (b) Are X and Y independent?

34. (a) Suppose that $\underline{X} = (X_1, X_2, X_3, X_4) \sim \mathrm{Mult}(30, \theta_1, \theta_2, \theta_3, \theta_4)$. Find the conditional probability mass function of (X_1, X_2, X_3, X_4) given that $X_5 = 2$, where $X_5 = 30 - \sum_{i=1}^{4} X_i$.
    (b) Consider 7 independent casts of a pair of fair dice. Find the probability that once we get a sum of 12 and twice we get a sum of 8.

35. (a) Let $X \sim F_{n_1,n_2}$. Show that $Y = \frac{n_2}{n_2+n_1X} \sim \mathrm{Be}(\frac{n_1}{2}, \frac{n_2}{2})$;
    (b) Let Z_1 and Z_2 be i.i.d. $N(0,1)$ random variables. Show that $\frac{Z_1}{|Z_2|} \sim t_1$ and $\frac{Z_1}{Z_2} \sim t_1$ (Cauchy distribution);
    (c) Let X_1 and X_2 be i.i.d. $\mathrm{Exp}(\theta)$ random variables, where $\theta > 0$. Show that $Z = \frac{X_1}{X_2}$ has a F-distribution;
    (d) Let X_1, X_2 and X_3 be independent random variables with $X_i \sim \chi^2_{n_i}, i = 1,2,3$. Show that $\frac{X_1/n_1}{X_2/n_2}$ and $\frac{X_3/n_3}{(X_1+X_2)/(n_1+n_2)}$ are independent F-variables.

36. Let $X_1, \ldots, X_n$ be a random sample from $\mathrm{Exp}(\theta)$ distribution, where $\theta > 0$. Let $X_{1:n} \le \cdots \le X_{n:n}$ denote the order statistics of $X_1, \ldots, X_n$. Define $Z_1 = nX_{1:n},\ Z_i = (n-i+1)(X_{i:n} - X_{i-1:n}),\ i = 2, \ldots, n$. Show that $Z_1, \ldots, Z_n$ are independent and identically distributed $\mathrm{Exp}(\theta)$ random variables. Hence find the mean and variance of $X_{r:n},\ r = 1, \ldots, n$. Also, for $1 \le r < s \le n$, find $\mathrm{Cov}(X_{r:n}, X_{s:n})$.