STOR 565 Homework: Probability and Statistics

1. Let $X > 0$ be a positive, continuous random variable with density f_X. Use the CDF method to find the density of $Y = X^{-1}$ in terms of f_X.

2. Recall that the variance of a random variable X is defined by $\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$. Carefully establish the following.

   a. If a, b are constants, then $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$

   b. $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ (expand the square in the definition)

   c. $\mathbb{E}X^2 \geq (EX)^2$.

3. In this problem we find an upper bound on the variance of a random variable with values in a finite interval. Let X be a random variable taking values in the finite interval $[0, c]$. You may assume that X is discrete, though this is not necessary for this problem.

   a. Show that $\mathbb{E}X \leq c$ and $\mathbb{E}X^2 \leq c\,\mathbb{E}X$.

   b. Recall that $\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$. Use the inequalities above to show that

   $$\mathrm{Var}(X) \;\leq\; c^2[u(1-u)] \quad \text{where} \quad u \;=\; \frac{\mathbb{E}X}{c} \in [0,1].$$

   c. Use this inequality and simple calculus to show that $\mathrm{Var}(X) \leq c^2/4$ if $X \in [0, c]$.

   d. Use this result to show that if X is a random variable taking values in an interval $[a, b]$ with $-\infty < a < b < \infty$ then $\mathrm{Var}(X) \leq (b-a)^2/4$

   e. It turns out that the general bound cannot be improved. To see this, show that the variance of the random variable $X \in [a, b]$ with $\mathbb{P}(X = a) = \mathbb{P}(X = b) = 1/2$ is equal to the bound you found above.

4. Show that if $f(x)$ is bounded and $X \sim \mathrm{Poiss}(\lambda)$ then $\mathbb{E}[\lambda f(X+1)] = \mathbb{E}[X f(X)]$. Here $\mathrm{Poiss}(\lambda)$ denotes the usual Poisson distribution with pmf $p(k) = e^{-\lambda}\lambda^k/k!$ for $k \geq 0$.

5. The probability that an individual has a certain rate disease is about 1 percent. If they have the disease, the chance that they test positive is 90 percent. If they do not have the disease, the chance that they nevertheless test positive is 9 percent. What is the probability

that someone who tests positive actually has the disease? (Use Bayes Formula.) What does this say about the test?

6. The empirical cumulative distribution function (CDF) of a sample $x = x_1, \ldots, x_m$ is defined by

$$F_x(t) \;=\; m^{-1} \sum_{i=1}^{m} \mathbb{I}(x_i \leq t)$$

The sum in the definition counts the number of data points that are less than or equal to t, so $F_x(t)$ is the fraction of data points that are less then or equal to t. Suppose that x has four points: -3, -1, -1, and 5.

a. Find the following values of the empirical CDF by using the formula above: $F_x(-4)$, $F_x(0)$, $F_x(-1)$, $F_x(6)$

b. Sketch the empirical CDF for this data set as a function of t.

c. For what values of t is $F_x(t) = 0$?

d. For what values of t is $F_x(t) = 1$?

7. Let X, X' be independent random variables with the same distribution. In this case we say that X' is an independent copy of X. Show that $\mathrm{Var}(X) = \frac{1}{2}\mathbb{E}(X - X')^2$

8. Let $x = x_1, \ldots, x_n$ be a univariate sample, and let $\tilde{x} = \tilde{x}_1, \ldots, \tilde{x}_n$ be the standardized version of x with $\tilde{x}_i = (x_i - m(x))/s(x)$. Show that $m(\tilde{x}) = 0$ and $s(\tilde{x}) = 1$.

9. Let $r(x, y)$ be the sample correlation of a bivariate data set $(x, y) = (x_1, y_1), \ldots, (x_n, y_n)$.

a. Let $ax + b$ denote the data set $ax_1 + b, \ldots, ax_n + b$ and define $cy + d$ similarly. Show that $r(ax + b, cy + d) = r(x, y)$ if $a, c > 0$.

b. Use the Cauchy-Schwarz inequality to show that $r(x, y)$ is always between -1 and $+1$.

10. Let P be a probability measure on a set $\mathcal{X}$. Recall that if A and B are subsets of $\mathcal{X}$ and $P(B) > 0$, then the conditional probability of A given B is defined by

$$P(A \,|\, B) \;=\; \frac{P(A \cap B)}{P(B)}$$

Show the following.

a. If A and B are disjoint then $P(A \cup B \,|\, C) = P(A \,|\, C) + P(B \,|\, C)$

b. $P(A^c \,|\, B) = 1 - P(A \,|\, B)$

c. If $A \subseteq B$ then $P(A \,|\, C) \leq P(B \,|\, C)$

11. Let $\mathcal{X}$ be a set and let A, B be subsets of $\mathcal{X}$. Recall that the indicator function of A is defined by

$$\mathbb{I}_A(x) \;=\; \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \in A^c \end{cases}$$

a. Show that $\mathbb{I}_{A^c} = 1 - \mathbb{I}_A$.

b. Show that $\mathbb{I}_A - \mathbb{I}_B = \mathbb{I}_{B^c} - \mathbb{I}_{A^c}$.

c. Show that $\mathbb{I}_{A \cap B} = \mathbb{I}_A \, \mathbb{I}_B$.

d. Let $u, v \in \{0, 1\}$. Show that $\mathbb{I}(u \neq v) = |\mathbb{I}(u = 1) - \mathbb{I}(v = 1)|$. Hint: Consider separately the cases $\mathbb{I}(u \neq v) = 0$ and $\mathbb{I}(u \neq v) = 1$.

12. Let (X, Y) be a discrete random pair with joint probability mass function $p(x, y)$. Recall from the lecture notes that we may define $\mathbb{E}(Y|X) = \varphi(X)$ where $\varphi(x) = \sum_y y \, p(y|x)$. Establish the following.

a. If $Y \geq 0$ then $\mathbb{E}(Y|X) \geq 0$

b. $\mathbb{E}(aY + b|X) = a \, \mathbb{E}(Y|X) + b$

c. $\mathbb{E}\{\, \mathbb{E}(Y|X) \,\} = \mathbb{E}Y$

13. Let X, Y be non-negative random variables with joint density function $f(x, y) = y^{-1} \, e^{-x/y} \, e^{-y}$ for $x, y \geq 0$.

a. Find the marginal density $f(y)$ of Y

b. Find the conditional density $f(x \,|\, y)$ of X given $Y = y$

c. Find $\mathbb{E}[X \,|\, Y = y]$

d. Find $\mathbb{E}[X \,|\, Y]$

*14. Let X be a discrete random variable taking values in a finite (or countably infinite) set $\mathcal{X}$, and having probability mass function $p(x) = \mathbb{P}(X = x)$. Let $h : \mathcal{X} \to [a, b]$ be any function.

a. Write down the sum for $\mathbb{E}h(X)$.

b. Show that $\mathbb{E}h(X) = a$ if $p(x) > 0$ only when $h(x) = a$.

c. Establish the reverse implication: if $\mathbb{E}h(X) = a$ then $p(x) > 0$ only when $h(x) = a$. Hint: Assume to the contrary that $p(x') > 0$ for some $x' \in \mathcal{X}$ with $h(x') \neq a$. As h takes values in $[a, b]$, we have $h(x') > a$. Use this to show $\mathbb{E}h(X) > a$.

d. Following the arguments above, show that $\mathbb{E}h(X) = b$ if and only if $p(x) > 0$ implies $h(x) = b$.

*15. Let $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ be a jointly distributed pair. Assume that $\mathcal{X}$ and $\mathcal{Y}$ are finite. Recall that X and Y are independent if $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y)$.

a. Show that if X and Y are independent then $\mathbb{P}(X = x \,|\, Y = y)$ does not depend on y.

b. Let $y \in \mathcal{Y}$ be fixed. Show that if $\mathbb{P}(Y = y \,|\, X = x)$ does not depend on x then it is equal to $\mathbb{P}(Y = y)$.

c. Suppose that for each $y \in \mathcal{Y}$ the conditional probability $\mathbb{P}(Y = y \,|\, X = x)$ does not depend on x. Show that X and Y are independent.

16. Let X have a $\mathcal{N}(\mu, \sigma^2)$ distribution. Show that $\mathbb{E}X = \mu$.

17. Let $Z \sim \mathcal{N}(0, 1)$. Use the CDF method to find the density of $X = aZ + b$.

18. Let $X \in \mathbb{R}^k$ be a random vector and $A \in \mathbb{R}^{r \times k}$. Use the definition of expected value, variance, and linear algebra to establish the following.

a. $\mathbb{E}(AX) = A\,\mathbb{E}X$

b. $\mathrm{Var}(X)$ is symmetric and non-negative definite

c. $\mathrm{Var}(X)_{ij} = \mathrm{Cov}(X_i, X_j)$

d. $\mathrm{Var}(AX) = A\,\mathrm{Var}(X)A^t$

19. Let $X \sim \mathcal{N}_d(\mu, \Sigma)$ and let $Y = \Sigma^{1/2} Z + \mu$ where $Z \sim \mathcal{N}_d(0, I)$.

(a) Show that $\mathbb{E}Y = \mathbb{E}X$ and that $\mathrm{Var}(Y) = \mathrm{Var}(X)$. Hint: Use the definition of $\mathcal{N}_d(\mu, \Sigma)$ and basic properties of multivariate means and variances.

(b) Fix $v \in \mathbb{R}^d$. Find the distributions of the random variable $v^t X$.

20. Let $X \sim \mathcal{N}_k(\mu, \Sigma)$ and let $Y = AX + b$ where $A \in \mathbb{R}^{l \times k}$ and $b \in \mathbb{R}^l$.

a. Find $\mathbb{E}Y$ and $\mathrm{Var}(Y)$.

b. Argue carefully that Y is multinormal and find its distribution.

c. Fix $v \in \mathbb{R}^l$. Using the results above, find the distribution of $U = \langle v, Y \rangle$.

21. Let $\mathcal{P} = \{f_\theta : \theta > 0\}$ be the family of exponential pdfs $f_\theta(x) = \theta e^{-\theta x}$ for $x \geq 0$. Suppose that we draw n samples independently from a fixed distribution $f_{\theta_0} \in \mathcal{P}$ and obtain data $x_1, \ldots, x_n \in \mathbb{R}$. The *likelihood function* for the family $\mathcal{P}$ is defined by $L(\theta) = \prod_{i=1}^n f_\theta(x_i)$. In words, $L(\theta)$ is just the joint density of the data $x_1, \ldots, x_n$ under f_θ, viewed as a function of the parameter θ, with the data held fixed. The *log-likelihood* is the log of the likelihood, $\ell(\theta) = \log L(\theta)$.

a. Note that $\ell(\theta) = \sum_{i=1}^n \log f_\theta(x_i)$. Find a simple expression for the log likelihood in this case.

b. The maximum likelihood estimate of the true parameter θ_0 is defined by $\hat{\theta}_n^{\mathrm{MLE}} = \mathrm{argmax}_{\theta > 0}\, \ell(\theta)$. Use calculus to find $\hat{\theta}_n^{\mathrm{MLE}}$ in terms of the data $x_1, \ldots, x_n$.

c. What is the relationship between $\hat{\theta}_n^{\mathrm{MLE}}$ and the average of the observations $x_1, \ldots, x_n$?

22. Let X be a standard normal random variable and let $Y = X^2$.

a. Use the cdf method to find the density of Y.

b. Are X and Y independent? Why or why not? An intuitive answer is fine.

c. Show that one of the events $\{Y \leq 1\}$ and $\{X \leq 1\}$ is contained in the other.

d. Show that $\mathbb{P}(X \leq 1, Y \leq 1) \neq \mathbb{P}(X \leq 1)\, \mathbb{P}(Y \leq 1)$. Thus X, Y are *not* independent.

e. What is $\mathrm{Cov}(X, Y)$? What do these results reveal about the relationship between covariance and independence?

23. Let $X \geq 0$ be a random variable with $\mathbb{E}X = 10$ and $\mathbb{E}X^2 = 120$.

a. Find an upper bound on $\mathbb{P}(X \geq 14)$ using Markov's inequality.

b. Let $0 < c < 10$. Find an upper bound on $\mathbb{P}(X \geq c)$ using Markov's inequality. Note that the bound is greater than one, and therefore uninformative. Argue informally that this is not a shortcoming of Markov's inequality, that is, $\mathbb{P}(X \geq c)$ may be equal to one.

c. Find an upper bound on $\mathbb{P}(X \geq 14)$ involving $\mathbb{E}X^2$.

d. Find an upper bound on $\mathbb{P}(X \geq 14)$ using Chebyshev's inequality. How does this bound compare to those above?

24. Let X be a random variable with $\mathrm{Var}(X) = 3$. Use Chebyshev's inequality to find upper bounds on $\mathbb{P}(|X - \mathbb{E}X| > 1)$ and $\mathbb{P}(|X - \mathbb{E}X| > 2)$. Comment on the potential usefulness of these bounds.

25. Recall that the moment generating function of a random variable X is defined by $M_X(s) = \mathbb{E}e^{sX}$ for all s such that the expectation is finite. Find the moment generating function (MGF) of the following distributions.

a. Poisson(λ)

b. $\mathcal{N}(0, 1)$

26. State and prove Markov's probability inequality.

27. Let X and Y be independent random variables with moment generating functions $M_X(s)$ and $M_Y(s)$, respectively. Show that $S = X + Y$ has moment generating function $M_S(s) = M_X(s)\, M_Y(s)$.

28. (Hoeffding's MGF Bound) Let X be a discrete random variable with probability mass function $p(\cdot)$. Assume that $a \leq X \leq b$ and that $\mathbb{E}X = 0$. Let $M_X(s) = \mathbb{E}e^{sX}$ be the moment generating function of X and define $\varphi(s) := \log M_X(s)$.

a. Show that

$$\varphi'(s) = \frac{\mathbb{E}[Xe^{sX}]}{\mathbb{E}e^{sX}} \quad \text{and} \quad \varphi''(s) = \frac{\mathbb{E}[X^2e^{sX}]}{\mathbb{E}e^{sX}} - (\varphi'(s))^2$$

b. Verify that $\varphi(0) = \varphi'(0) = 0$

Now fix $t > 0$ and let U be a new random variable having the "exponentially tilted" probability mass function

$$q(x) \;=\; \frac{p(x)e^{tx}}{\mathbb{E}e^{tX}}$$

c. Verify that $q(\cdot)$ is a probability mass function, that is, $q(x) \geq 0$ and $\sum_x q(x) = 1$.

d. Argue that $a \leq U \leq b$. This follows from the fact that U has the same possible values as X, only with different probabilities.

e. Show that $\mathbb{E}(U) = \varphi'(t)$ and that $\mathrm{Var}(U) = \varphi''(t)$.

f. Using the variance bound for bounded random variables, conclude from (c) and (d) that $\varphi''(t) \leq (b-a)^2/4$.

g. Use the second order Taylor series expansion of φ around $s = 0$ and what you've shown above to establish that $\varphi(s) \leq s^2(b-a)^2/8$ for $s > 0$.

h. Exponentiating the inequality in (g) gives Hoeffding's MGF bound.

29. Let $X_1, \ldots, X_n$ be iid Uniform$(-\theta, \theta)$ random variables.

a. Use Chebyshev's inequality to find a bound on $\mathbb{P}(\sum_{i=1}^n X_i \geq t)$ for $t \geq 0$.

b. Use Hoeffding's inequality to find a bound on $\mathbb{P}(\sum_{i=1}^n X_i \geq t)$ for $t \geq 0$.

30. Let U_1, U_2 be uncorrelated random variables with mean zero and variance one. Define $U = (U_1, U_2)^t$. Let $X = (X_1, X_2)^t$ be a random vector with components

$$X_1 \;=\; a\,U_1 + b\,U_2 \;\text{ and }\; X_2 \;=\; c\,U_1 + d\,U_2$$

a. Find $\mathbb{E}[U]$.

b. What is $\mathrm{Var}(U)$?

c. Find $\mathbb{E}X$.

d. Find the matrix $\mathrm{Var}(X)$ by directly calculating each entry using the definitions of X_1 and X_2.

e. Find $\mathbf{A}$ such that $X = \mathbf{A}U$.

f. Find $\mathrm{Var}(X)$ using the formula for $\mathrm{Var}(\mathbf{A}U)$.

g. In terms of a, b, c and d, when is $\mathbf{A}$ invertible?

31. Chi-squared distribution. A random variable X has a chi-squared distribution with $k \geq 1$ degrees of freedom, written $X \sim \chi^2_k$, if X has the same distribution as $Z_1^2 + \cdots + Z_k^2$ where $Z_1, \ldots, Z_k$ are iid $\sim \mathcal{N}(0,1)$.

a. Find $\mathbb{E}X$ and $\mathrm{Var}(X)$ when $X \sim \chi^2_k$. You may use the fact that $\mathbb{E}Z^4 = 3$ if $Z \sim \mathcal{N}(0,1)$.

b. If $X \sim \chi^2_k$ and $Y \sim \chi^2_l$ are independent, what is the distribution of $X + Y$?

32. Let $X_1, \ldots X_n$ be independent and identically distributed random variables. Calculate $\mathbb{E}[X_1 | X_1 + \ldots + X_n = x]$. (Hint: Consider $\mathbb{E}[S_n | X_1 + \ldots + X_n = x]$ where $S_n = X_1 + \ldots + X_n$ and use symmetry.)

33. Let $\phi(x)$ and $\Phi(x)$ be the density function and cumulative distribution function, respectively, of the standard normal distribution. Here we will find a useful upper bound on $1 - \Phi(x)$, which is the probability that a standard normal random variable exceeds x.

(a) Write down the formula for the density $\phi(t)$, and compute the derivative $\phi'(t)$.

(b) Justify the following sequence of equalities: For $x > 0$,

$$1 - \Phi(x) \;=\; \Phi(-x) \;=\; \int_{-\infty}^{-x} \phi(t)\, dt \;=\; \int_{-\infty}^{-x} \frac{1}{t} \cdot t\, \phi(t)\, dt.$$

(c) Integrate the last term above by parts to establish the useful inequality $1 - \Phi(x) \leq x^{-1}\, \phi(x)$ for $x > 0$.

34 (The weak law of large numbers). Let $U_1, U_2, \ldots, U$ be iid random variables with finite variance, and let $X = n^{-1} \sum_{i=1}^n U_i$ be the average of $U_1, \ldots, U_n$.

a. Find $\mathbb{E}X$ in terms of $\mathbb{E}U$.

b. Find $\mathrm{Var}(X)$ in terms of $\mathrm{Var}(U)$.

c. Use these calculations and Chebyshev's inequality to establish that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_i - \mathbb{E}U\right| \geq t\right) \leq \frac{\mathrm{Var}(U)}{nt^2}$$

d. What happens to the right side of the inequality above when t is fixed and n tends to infinity?

36. Let X, Y be random variables and let $a, b > 0$. Define events

$$A = \{|X| \geq a\} \quad B = \{|Y| \geq b\} \quad C = \{|X+Y| \geq a+b\}$$

a. Argue that if $a > |X|$ and $b > |Y|$ then $a + b > |X + Y|$.

b. Conclude that $A^c \cap B^c \subseteq C^c$.

c. Show using Boolean algebra that $C \subseteq A \cup B$.

d. Conclude using the properties of probability that $\mathbb{P}(C) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

e. Use the same type of reasoning to show that $\mathbb{P}(|XY| \geq a) \leq \mathbb{P}(|X| \geq a/b) + \mathbb{P}(|Y| \geq b)$.