Probability Theory STAT 461/661
Final Exam May 13, 2024

Name _____

Your exam should consist of 10 problems and a Z-table.

You are allowed a double-sided handwritten note and a calculator.

Score 100 corresponds to 100%, above that is extra credit.

Problem	Points	S C O R E
1	10	
2	10	
3	10	
4	10	
5	10	
6	10	
7	10	
8	15	
9	15	
10	15	
Total:	115	

1. (10 points) A test for an infection gives us false positive result (that the patient has this infection if, in fact, the patient does not have this infection) with probability 10%. This test gives us a false negative result (that the patient does not have this infection if, in fact, the patient has this infection) with probability 5%. The overall rate of infection in the entire population is 2%. We conduct a test for a random person, and the result is positive. What is the probability that a person, in fact, has this infection?

   Solution: Let A be the event that the test is positive. Let $S =$ sick, $H =$ healthy. Then

   $$\mathbb{P}(S) = 0.02 \quad \mathbb{P}(H) = 0.98, \quad \mathbb{P}(A \mid S) = 0.95, \quad \mathbb{P}(A \mid H) = 0.1.$$

   By Bayes' formula,

   $$\mathbb{P}(S \mid A) = \frac{\mathbb{P}(A \mid S) \cdot \mathbb{P}(S)}{\mathbb{P}(A \mid S) \cdot \mathbb{P}(S) + \mathbb{P}(A \mid H) \cdot \mathbb{P}(H)} = \frac{0.02 \cdot 0.95}{0.02 \cdot 0.95 + 0.98 \cdot 0.1} = \boxed{16.2\%}$$

2. (10 points) The *Maxwell distribution* of a molecule speed V has density

$$f(v) = (2/\pi)^{1/2}\, v^2 \exp\left[-0.5v^2\right], \quad v > 0.$$

Find the density $g(w)$ of the energy $W = mV^2$ for (constant) mass m.

Solution: We can express $v = (w/m)^{0.5}$. The derivative $\mathrm{d}w/\mathrm{d}v = 2mv$. Thus

$$g(w) = \frac{f(v)}{2mv} = \frac{f\left((w/m)^{0.5}\right)}{2(wm)^{0.5}} = \left(\frac{2}{\pi}\right)^{1/2} \frac{w}{m} \frac{e^{-w/2m}}{2(wm)^{0.5}} = \boxed{\frac{1}{\sqrt{2\pi}} \frac{w^{1/2}}{m^{3/2}} e^{-w/2m}}$$

Note this is (scaled) χ^2 law with 3 degrees of freedom. This is not surprising, since $V^2 = Z_1^2 + Z_2^2 + Z_3^2$ where $Z_k \sim \mathcal{N}(0,1)$ are independent identically distributed random variables: three dimension velocity components.

3. (10 points) A drawer contains 4 white and 4 black socks. Draw two socks at random. Find the probability that these socks have the same color.

Solution: The number of ways to draw 2 socks out of 8 is $\binom{8}{2} = \frac{8 \cdot 7}{2} = 28$. The number of ways to draw 2 socks out of 4 white is $\binom{4}{2} = \frac{4 \cdot 3}{2} = 6$. The number of ways to draw 2 socks out of 4 black is also 6. Thus the number of ways to draw 2 matching socks is $6 + 6 = 12$. The probability of this is $12/28 = \boxed{3/7}$

4. (10 points) We model large stock returns as $X \sim \mathcal{N}(\mu, \sigma^2)$, for $\mu = 11.8$ and $\sigma = 19.3$. Then we invest in a portfolio with $\beta = 1.2$ and $\alpha = -1.4$, which has returns $Y = \alpha + \beta X$. Find $\mathbb{P}(Y > 0)$.

   Solution: We can represent $X = \mu + \sigma Z$ for standard normal Z. Therefore,

$$Y = \alpha + \beta X = \alpha + \beta\mu + \beta\sigma Z = 12.76 + 23.16Z.$$

   Thus, $Y > 0$ if and only if $Z > -\frac{12.76}{23.16} = -0.55$. This has probability

$$\mathbb{P}(Z > -0.55) = \mathbb{P}(Z < 0.55) = \boxed{70.88\%}$$

5. (10 points) Take two independent random variables: a Gamma Y with density ye^{-y} for $y > 0$, and a standard normal Z. Combine them to make another random variable $X = \sqrt{Y}Z$.

   (STAT 461) Find the standard deviation σ of X.

   (STAT 661) Find the kurtosis of X: $\sigma^{-4} \cdot \mathbb{E}\left[(X - \mathbb{E}X)^4\right]$.

   Solution: The Gamma distribution of Y is $\Gamma(2, 1)$, so $\mathbb{E}[Y] = \mathrm{Var}(Y) = 2$, and $\mathbb{E}[Y^2] = \mathrm{Var}(Y) + (\mathbb{E}[Y])^2 = 2 + 2^2 = 6$. By independence of Y and Z, the mean of X is $\mathbb{E}[X] = \mathbb{E}[\sqrt{Y}] \cdot \mathbb{E}[Z] = 0$.

   (STAT 461) Using independence of Y and Z, we get: the variance is $\sigma^2 = \mathrm{Var}(X) = \mathbb{E}[X^2] = \mathbb{E}[(\sqrt{Y}Z)^2] = \mathbb{E}[Y] \cdot \mathbb{E}[Z^2] = 2 \cdot 1 = 2$. Thus $\sigma = \boxed{\sqrt{2}}$

   (STAT 661) First, $\mathbb{E}[(X - \mathbb{E}X)^4] = \mathbb{E}[(\sqrt{Y}Z)^4] = \mathbb{E}[Y^2] \cdot \mathbb{E}[Z^4] = 6 \cdot 3 = 18$. Indeed, the MGF of Z is $M(t) = e^{t^2/2}$, and its fourth derivative is $M^{(4)}(t) = (t^4 + 6t^2 + 3)e^{t^2/2}$, which gives us $\mathbb{E}[Z^4] = M^{(4)}(0) = 3$. Thus the kurtosis is $18/(\sqrt{2}^4) = \boxed{4.5}$

6. (10 points) Let S_0, S be the stock prices at the beginning and end of year. Then $Q = \ln(S_1/S_0)$ are called *log returns*. Assume we model them as $Q \sim \mathcal{N}(0.1, 0.16)$, and $S_0 = 12$. Thus $S_1 = 12e^Q$.

   (a, 4 points) Find the mean of S_1.

   (b, 6 points) Find the variance of S_1.

   Solution: The random variable Q has MGF $M(t) = \mathbb{E}[e^{tQ}] = \exp(0.1t + 0.08t^2)$.

   (a) Therefore, $\mathbb{E}[S_1] = 12 \cdot \mathbb{E}[e^Q] = 12M(1) = 12\exp(0.18) = \boxed{14.4}$

   (b) Also, the second moment is

$$\mathbb{E}[S_1^2] = 12^2 \cdot \mathbb{E}[e^{2Q}] = 12^2 M(2) = 12^2 \cdot \exp(0.1 \cdot 2 + 0.08 \cdot 4) = 12^2 \exp(0.56)$$

   and the variance is computed as

$$\mathrm{Var}(S_1) = 12^2(M(2) - M^2(1)) = 12^2\left(\exp(0.56) - \exp(0.36)\right) = \boxed{45.7}$$

7. (10 points) Let us roll two dice 24 times. Possible outcomes for each double roll are 11, 12, ..., 16, 21, 22, ..., 26, ..., 61, ..., 66.

   (a, 5 points) Find the (exact) probability that there is at least one double-six.

   (b, 5 points) Approximate this probability using the Poisson distribution.

   Compute both probabilities up to three decimal points, for example 10.4%.

   Solution: (a) The probability of drawing 66 in one double roll is $p = 1/36$. Thus the number X of 66 is distributed as $\mathrm{Bin}(N, p)$ where $N = 24$. The probability of $X > 0$ is $1 - (1-p)^N = \boxed{49.1\%}$

   (b) This approximation is $\mathrm{Poi}(\lambda)$ where $\lambda = Np = 2/3$. Thus the probability of $X > 0$ is $1 - e^{-\lambda} = \boxed{48.7\%}$

8. (15 points) Two random variables X and Y have joint density

$$f(x,y) = c(x+y), \quad 1 \le x \le 3, \quad 1 \le y \le 3.$$

(a, 3 points) Are X and Y independent?

(b, 4 points) Find the constant c.

(c, 4 points) Find the mean of X.

(d, 4 points) Find the probability that $X < Y$.

Solution: (a) No, because the density is not product form.

(b) Total probability must be 1. Therefore,

$$c\int_1^3\int_1^3 (x+y)\,\mathrm{d}x\,\mathrm{d}y = \int_1^3\int_1^3 f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1.$$

Compute this double integral I:

$$\begin{aligned} I &= \int_1^3 x\,\mathrm{d}x \cdot \int_1^3 \mathrm{d}y + \int_1^3 \mathrm{d}x \cdot \int_1^3 y\,\mathrm{d}y \\ &= 2 \cdot \frac{3^2-1^2}{2} \cdot 2 = 16. \end{aligned}$$

Thus $cI = 1$ and $c = \boxed{1/16}$

(c) The mean of X is

$$\begin{aligned} &\int_1^3\int_1^3 x \cdot \frac{1}{16}(x+y)\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{16}\int_1^3 x^2\,\mathrm{d}x \cdot \int_1^3 \mathrm{d}y + \frac{1}{16}\int_1^3 x\,\mathrm{d}x \cdot \int_1^3 y\,\mathrm{d}y \\ &= \frac{1}{16}\,\frac{x^3}{3}\Big|_{x=1}^{x=3} \cdot 2 + \frac{1}{16} \cdot \frac{x^2}{2}\Big|_{x=1}^{x=3} \cdot \frac{y^2}{2}\Big|_{y=1}^{y=3} \\ &= \frac{1}{16}\frac{26}{3} \cdot 2 + \frac{1}{16} \cdot 4 \cdot 4 = \boxed{\frac{25}{12}} \end{aligned}$$

(d) By symmetry of density, (X,Y) and (Y,X) have the same distribution. Therefore, $\mathbb{P}(X<Y) = \mathbb{P}(X>Y)$. And $\mathbb{P}(X=Y) = 0$. Thus $\mathbb{P}(X<Y) = \boxed{0.5}$ Alternatively,

$$\mathbb{P}(X<Y) = \int_1^3\int_x^3 \frac{1}{16}(x+y)\,\mathrm{d}y\,\mathrm{d}x = \frac{1}{2}.$$

9. (15 points) An insurance company has 12000 clients with independent losses. For 10000 clients, losses are identically distributed with density

$$f(x) = 6x(1-x), \quad 0 \le x \le 1.$$

For the remaining 2000 clients, losses are identically distributed with mass function

$$p(0) = p(1) = 0.2, \quad p(0.5) = 0.6.$$

The company is required to charge each client the same premium s, so the total capital of this company is $12000s$. Find the s such that the total capital exceeds the sum of all losses with probability 90%.

Solution: The mean of the continuous loss is equal to

$$m_c = \int_0^1 xf(x)\,\mathrm{d}x = \int_0^1 6x^2(1-x)\,\mathrm{d}x = \left(2x^3 - 1.5x^4\right)\Big|_{x=0}^{x=1} = 0.5.$$

The second moment of the continuous loss is equal to

$$a_c = \int_0^1 x^2 f(x)\,\mathrm{d}x = \int_0^1 6x^3(1-x)\,\mathrm{d}x = \left(1.5x^4 - 1.2x^5\right)\Big|_{x=0}^{x=1} = 0.3.$$

Thus the variance is equal to $\sigma_c^2 = a_c - m_c^2 = 0.3 - 0.5^2 = 0.05$. Similarly, the mean of the discrete loss is equal to

$$m_d = 0 \cdot p(0) + 0.5 \cdot p(0.5) + 1 \cdot p(1) = 0 \cdot 0.2 + 0.5 \cdot 0.6 + 1 \cdot 0.2 = 0.5.$$

The variance of the discrete loss is equal to

$$\sigma_d^2 = (0-0.5)^2 \cdot p(0) + (0.5-0.5)^2 \cdot p(0.5) + (1-0.5)^2 \cdot p(1) = 0.25 \cdot 0.2 + 0.25 \cdot 0.2 = 0.1.$$

Thus the mean of the sum S of all 12000 losses is

$$m = 10000m_c + 2000m_d = 12000 \cdot 0.5 = 6000.$$

By independence, the variance of the sum S is

$$\sigma^2 = 10000\sigma_c^2 + 2000\sigma_d^2 = 10000 \cdot 0.05 + 2000 \cdot 0.1 = 700.$$

By the Central Limit Theorem,

$$\frac{S-m}{\sigma} \approx Z \sim \mathcal{N}(0,1).$$

With probability 90%, $Z < 1.282$ and $S < m + 1.282\sigma = 6000 + 1.282 \cdot \sqrt{700} = 6034$. Dividing this by 12000, we get $s = \boxed{0.503}$

10. (15 points) The water level in Truckee River each year is modeled as an exponential random variable with mean m. Each year, water levels are independent of previous years. Let M be the minimal water level in 10 years.

(a, 7 points) Find the density of M.

(b, 4 points) Find the median of M.

(c, 4 points) Find the mean of M.

Solution: (a) Denote by X_i the water level in year i, for $i = 1, \ldots, 10$. The density function of X_i is $f(x) = (1/m)e^{-x/m}$ for $x > 0$ (the mean is the inverse of the rate λ). Therefore, the tail function of this exponential random variable is $\mathbb{P}(X_i > x) = \int_x^\infty f(y)\,\mathrm{d}y = G(x) = e^{-x/m}$. Thus, the tail function of M is

$$\mathbb{P}(M > x) = \mathbb{P}(X_1 > x, \ldots, X_{10} > x).$$

By independence, we can apply the product rule:

$$G_M(x) := \mathbb{P}(M > x) = \mathbb{P}(X_1 > x) \cdot \ldots \cdot \mathbb{P}(X_{10} > x) = G^{10}(x) = \left[e^{-x/m}\right]^{10} = e^{-10x/m}.$$

Taking the derivative of this tail function, we get the density:

$$f_M(x) = -G'_M(x) = \boxed{\frac{10}{m}e^{-10x/m}}$$

(b) The median μ is found from $G_M(\mu) = 0.5$. Solving for this, we get: $e^{-10\mu/m} = 0.5$ so $\mu = \boxed{0.1 \ln 2 \cdot m}$

(c) This is an exponential distribution as well, with rate $10/m$ and mean $\boxed{m/10}$ We can also find it as

$$\mathbb{E}[M] = \int_0^\infty x \cdot (10/m)e^{-10x/m}\,\mathrm{d}x = \frac{m}{10}.$$