Math 3215: Midterm 2

Directions: Answers in the form of $\Pr[Z < 2]$ or $\Pr[-1 \leq T_5 \leq 1]$ are fine. Otherwise answers should be numbers or numbers easily computable by a calculator (eg $e^{1.5}, 1 - 179/340$ etc..). Your answers should not contain $\sum_{i=1}^{n}$ or $\binom{100}{35}$.

1 (20 points)

Assume that $\Pr[-3 \leq T_{16} \leq 3] \sim .99$. We take a sample of 16 players and find the sample average height to be 77 inches and S^2 of the sample to be 9.

1. Find a 99% confidence interval for the mean of a randomly chosen NBA player's height.
2. Does this mean that approximately 99% of players will have a height within this confidence interval? Why or why not?

answer:

1. The confidence interval is:

$$
\begin{aligned}
&[77, -3S/\sqrt{n}, 77 + 3 \cdot S/\sqrt{n} \\
&= [77 - 9/4, 77 + 9/4]
\end{aligned}
$$

2. No. The confidence interval says something about where the *average* height of a player is located. It is very likely that as our sample size (n) grows, the confidence interval shrinks until eventually very few players' heights are within the interval, but the average height still is.

2 (16 points)

Let X_1 and X_2 be independent random variables both with mean 1 and variance 4. Let

- $Y = X_1 - X_2$
- $Z = 2X_2 - 3X_1$

Calculate:

1. $cov(X_1, Y)$
2. $cov(Y, Z)$

answer:

1. $\mathbb{E}Y = 0$, so $cov(X_1, Y) = \mathbb{E}[X_1 Y] - 0$.

$$\begin{aligned} cov(X_1, Y) = \mathbb{E}[X_1 Y] &= \mathbb{E}[X_1^2 - X_1 X_2] \\ &= \mathbb{E}[X_1^2] - \mathbb{E}[X_1 X_2] \\ &= (4+1) - 1 \cdot 1 = 4 \end{aligned}$$

   since X_1, X_2 are independent. $\mathbb{E}[X_1^2] = var(X_1) + (\mathbb{E}X_1)^2 = 5$.

2. again since $\mathbb{E}Y = 0$, $cov(Y, Z) = \mathbb{E}[YZ]$.

$$\begin{aligned} \mathbb{E}[YZ] &= \mathbb{E}[-3X_1^2 - 2X_2^2 + 5X_1X_2] \\ &= -5 \cdot 5 + 5\mathbb{E}X_1X_2 \\ &= -25 + 5 = -20 \end{aligned}$$

3 (16 points)

The chances of winning the mega-millions lottery are 1 in $170,000,000$ for each ticket. Say $340,000,000$ tickets were sold.

1. What is the expected number of winning tickets?
2. Estimate the probability that at least 2 winning tickets were sold.

answer:

1. Let X_i be the indicator r.v. that ticket i is a winner. Then $\mathbb{E}X_i = \frac{1}{170000000}$, and there are 340000000 tickets, so the expected number of winning tickets is 2.
2. The number of winning tickets is a binomial random variable (assuming tickets numbers are picked indepedently), with $np = 2$. This calls for Poisson approximation, not binomial approximation. $\Pr[X \geq 2] = 1 - \Pr[X = 0] - \Pr[X = 1] = 1 - e^{-2} - 2e^{-2} = 1 - 3e^{-2}$.

4 (20 points)

We want to test whether McDonald's or Burger King has bigger hamburgers. We sample 400 random burgers from Burger King and find an average of 4.1 oz, and 50 random burgers from McDonald's and find an average weight of 3.9 oz. Assume that a random burger from each restaurant has a normal distribution with unknown mean and variance 1. We want to determine whether μ_b or μ_m is greater, where μ_b is the true mean of a Burger King burger and μ_m is the true mean of a McDonald's burger. Call the sampled weights of the burgers $b_1, b_2, \ldots b_{400}$ and $m_1, m_2, \ldots m_{50}$.

1. Find a good estimator for $\mu_b - \mu_m$. Call this estimator Y.
2. (Bonus) Show that Y is the maximum likelihood estimator.
3. What is $\mathbb{E}Y$?
4. What $var(Y)$? Show the whole calculation.
5. Find a 95% confidence interval for $\mu_b - \mu_m$.
6. In terms of Z probabilities, calculate your confidence that $\mu_b \geq \mu_m$.

answer:

1. The natural estimator is $\frac{b_1+\cdots+b_{400}}{400} - \frac{m_1+\cdots+m_{50}}{50} = \overline{b} - \overline{m}$

2.

3. $\mathbb{E}Y = \mu_b - \mu_m$

4. We use the fact that our random samples are independent. $var(b_i/400^2) = \sigma^2/400^2 = 1/400^2$, and $var(m_i/50) = 1/50^2$), so $var(Y) = 400/400^2 + 50/50^2 = 1/400 + 1/50 = 9/400$.

5. Since $\Pr[-2 \leq Z \leq 2] = .95$, our 95% confidence interval for $\mu_b - \mu_m$ is

$$[.2 - 2 \cdot sd(Y), .2 + 2 \cdot sd(Y)]$$
$$= [.2 - 2 \cdot \frac{3}{20}, .2 + 2 \cdot \frac{3}{20}]$$
$$= [-.1, .5]$$

6. This confidence is the $\Pr[Z \cdot sd(Y) \geq -.2]$

$$= \Pr[Z \geq -.2/sd(Y)]$$
$$= \Pr[Z \geq -4/3]$$

5 (12 points)

X_1 and X_2 are independent Poisson random variables with means λ_1 and λ_2. Let $Z = X_1 + X_2$.

1. Calculate $\mathbb{E}Z$
2. Calculate $var(Z)$
3. Prove that Z has a Poisson distribution

answer:

1. $\lambda_1 + \lambda_2$
2. $\lambda_1 + \lambda_2$ by independence and the fact that the variance of a Poisson equals its mean.
3. To show Z has a Poisson distribution we calculate:

$$\Pr[Z = k] = \sum_{i=0}^{k} \Pr[X_1 = i] \Pr[X_2 = k - i]$$
$$= \sum_{i=0}^{k} \frac{e^{-\lambda_1}\lambda_1^i}{i!} \frac{e^{-\lambda_2}\lambda_2^{k-i}}{(k-i)!}$$
$$= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=1}^{k} \frac{k!}{i!(k-i)!} \lambda_1^i \lambda_2^{k-i}$$
$$= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} (\lambda_1 + \lambda_2)^k$$

which is a Poisson distribution with parameter $\lambda_1 + \lambda_2$.

6 (8 points)

Let X be uniform on $[0, 2]$ and Y uniform on $[1, 3]$.

- Find the maximum possible value of $cov(X, Y)$
- Find the minimum possible value of $cov(X, Y)$

answer: Covariance is highest when X and Y are as highly correlated as possible - that is, when X is big Y is big. This happens when $Y = X + 1$ (notice that Y is still uniform on $[1, 3]$. We can calculate the covariance in this case:

$$\begin{aligned} cov(X, Y) &= \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y \\ &= \int_0^2 \frac{1}{2} x(x+1)\, dx - 1 \cot 2 \\ &= \frac{x^3}{6} + \frac{x^2}{4}|_0^2 - 2 \\ &= \frac{8}{6} + 1 - 2 = \frac{1}{3} \end{aligned}$$

Similarly, covariance is lowest when X and Y are negatively correlated - when X is large Y is small. In this case, the extreme is to set $Y = 3 - X$. Again that keeps Y unform on $[1, 3]$. We can do a similar calculation and find $cov(X, Y) = -\frac{1}{3}$.

7 (8 points)

Let $X_1, X_2, \ldots$ be i.i.d. exponential random variables with mean 1. I.e. each has density function $f(x) = e^{-x}$, $x \geq 0$.

1. What is $\Pr[X_1 > 1]$?
2. What is $\Pr[X_1 \leq 1 \text{ and } X_1 + X_2 > 1]$?
3. What is $\Pr[X_1 + X_2 \leq 1 \text{ and } X_1 + X_2 + X_3 > 1]$?
4. What is $\Pr[X_1 + X_2 + \cdots + X_j \leq 1 \text{ and } X_1 + X_2 + \cdots + X_j + X_{j+1} > 1]$?

answer:
This is mainly a problem in integration but with a nice answer with probabilistic meaning.

1.

$$\begin{aligned} \Pr[X > 1] &= \int_1^\infty e^{-x}\, dx \\ &= -e^{-x}|_1^\infty = e^{-1} \end{aligned}$$

2.

$$\begin{aligned} \Pr[X_1 \leq 1 \text{ and } X_1 + X_2 > 1] &= \int_0^1 \int_{1-x_1}^\infty e^{-x_1 - x_2}\, dx_2 dx_1 \\ &= \int_0^1 e^{-x_1} \cdot (-e^{-x_2}|_{1-x_1}^\infty)\, dx_1 \\ &= \int_0^1 e^{-x_1} e^{x_1 - 1}\, dx_1 \\ &= e^{-1} \end{aligned}$$

3.

$$\begin{aligned}\Pr[X_1 + X_2 \leq 1 \text{ and } X_1 + X_2 + X_3 > 1] &= \int_0^1 e^{-x_1} \int_0^{1-x_1} e^{-x_2} \int_{1-x_1-x_2}^{\infty} e^{-x_3}\, dx_3 dx_2 dx_1 \\ &= \int_0^1 e^{-x_1} \int_0^{1-x_1} e^{-x_2} e^{x_2+x_1-1} dx_2 dx_1 \\ &= e^{-1} \int_0^1 \int_0^{1-x_1} dx_2 dx_1 \\ &= \frac{e^{-1}}{2}\end{aligned}$$

4.

$$\begin{aligned}&\Pr[X_1 + X_2 + \cdots + X_j \leq 1 \text{ and } X_1 + X_2 + \cdots + X_j + X_{j+1} > 1] \\ &= \int_0^1 \int_0^{1-x_1} \cdots \int_0^{1-x_1-\cdots-x_{j-1}} \int_{1-x_1-\cdots-x_j}^{\infty} e^{-x_1-\cdots-x_{j+1}}\, dx_{j+1} \cdots dx_1 \\ &= e^{-1} \int_0^1 \int_0^{1-x_1} \int_0^{1-x_1-x_2} \cdots \int_0^{1-x_1-\cdots-x_{j-1}} dx_j \cdots dx_1 \\ &= e^{-1} \int_0^1 \int_0^{1-x_1} \int_0^{1-x_1-x_2} \cdots \int_0^{1-x_1-\cdots-x_{j-2}} 1 - x_1 - \cdots - x_{j-1}\, dx_{j-1} \cdots dx_1 \\ &= e^{-1} \int_0^1 \int_0^{1-x_1} \int_0^{1-x_1-x_2} \cdots \int_0^{1-x_1-\cdots-x_{j-3}} \frac{(1 - x_1 - \cdots - x_{j-2})^2}{2}\, dx_{j-2} \cdots dx_1 \\ &= e^{-1} \int_0^1 \int_0^{1-x_1} \int_0^{1-x_1-x_2} \cdots \int_0^{1-x_1-\cdots-x_{j-3}} \frac{(1 - x_1 - \cdots - x_{j-2})^3}{2 \cdot 3}\, dx_{j-3} \cdots dx_1 \\ &\cdots \\ &= \frac{e^{-1}}{j!}\end{aligned}$$

Which is exactly $\Pr[Pois(1) = j]$.