Math 5010 § 1. **Final Exam** Name:______________

Treibergs $\sigma\tau$ May 1, 2009

---

1. *Suppose X is distributed according to Pascal's distribution with parameter $\mu > 0$, where*

$$f_X(x) = \begin{cases} \frac{\mu^x}{(1+\mu)^{x+1}}, & \text{for } x = 0, 1, 2, 3, \ldots; \\ 0, & \text{otherwise.} \end{cases}$$

Show that for $0 \le s, t$ we have $\mathbf{P}(X \ge s + t \mid X \ge s) = \mathbf{P}(X \ge t)$.

For $z \ge 0$, the probability

$$\mathbf{P}(X \ge z) = \sum_{x=z}^{\infty} f_X(x) = \sum_{x=z}^{\infty} \frac{\mu^x}{(1+\mu)^{x+1}} = \frac{\mu^z}{(1+\mu)^{z+1}} \sum_{x=0}^{\infty} \left(\frac{\mu}{1+\mu}\right)^x$$

$$= \frac{\left(\frac{\mu}{1+\mu}\right)^z}{(1+\mu)\left(1 - \frac{\mu}{1+\mu}\right)} = \left(\frac{\mu}{1+\mu}\right)^z.$$

Hence the conditional probability is

$$\mathbf{P}(X \ge s+t \mid X \ge s) = \frac{\mathbf{P}(X \ge s+t \text{ and } X \ge s)}{\mathbf{P}(X \ge s)} = \frac{\mathbf{P}(X \ge s+t)}{\mathbf{P}(X \ge s)}$$

$$= \frac{\left(\frac{\mu}{1+\mu}\right)^{s+t}}{\left(\frac{\mu}{1+\mu}\right)^{s}} = \left(\frac{\mu}{1+\mu}\right)^t = \mathbf{P}(X \ge t).$$

2. *Suppose that an urn has 5 red balls and 6 white balls. You choose four balls randomly from the urn without repalcement. What is the probability that you have chosen at least two white balls? What is the expected number of red balls chosen, given that you have chosen at least two white balls?*

Let X be the number of red balls chosen. Then there are $W = 4 - X$ white balls. The probability of choosing k white balls is, using the hypergeometric pmf, for $k = 0, 1, 2, 3, 4$,

$$\mathbf{P}(W = k) = \mathbf{P}(X = 4 - k) = \frac{\binom{5}{4-k}\binom{6}{k}}{\binom{11}{4}}.$$

Thus the probability of two or more white balls is

$$\mathbf{P}(W \ge 2) = \mathbf{P}(W = 2) + \mathbf{P}(W = 3) + \mathbf{P}(W = 4)$$

$$= \frac{\binom{5}{2}\binom{6}{2}}{\binom{11}{4}} + \frac{\binom{5}{1}\binom{6}{3}}{\binom{11}{4}} + \frac{\binom{5}{0}\binom{6}{4}}{\binom{11}{4}} = \frac{10 \cdot 15 + 5 \cdot 20 + 1 \cdot 15}{330} = \frac{265}{330}.$$

Because $W \ge 2$ if and only if $X = 4 - W \le 2$, the conditional pmf is

$$f_X(k \mid W \ge 2) = \frac{\mathbf{P}(X = k \text{ and } X \le 2)}{\mathbf{P}(W \ge 2)} = \frac{\mathbf{P}(X = k)}{\mathbf{P}(W \ge 2)} = \frac{\binom{5}{k}\binom{6}{4-k}}{\frac{265}{330}\binom{11}{4}} = \frac{1}{265}\binom{5}{k}\binom{6}{4-k}$$

for $k = 0, 1, 2$ and zero otherwise. Thus the conditional expectation

$$\mathbf{E}(X \mid W \ge 2) = 0 \cdot f_X(0 \mid W \ge 2) + 1 \cdot f_X(1 \mid W \ge 2) + 2 \cdot f_X(2 \mid W \ge 2)$$

$$= \frac{1}{265}\left\{1 \cdot \binom{5}{1}\binom{6}{3} + 2 \cdot \binom{5}{2}\binom{6}{2}\right\} = \frac{1 \cdot 5 \cdot 20 + 2 \cdot 10 \cdot 15}{265} = \frac{400}{265} \approx 1.509.$$

3. *Let X and Y be two independent normal random variables with zero means and variances σ^2 and τ^2, resp. Find* $\mathbf{E}\left(e^{-X^2}Y^2\right)$

By independence, $\mathbf{E}\left(e^{-X^2}Y^2\right) = \mathbf{E}\left(e^{-X^2}\right)\mathbf{E}\left(Y^2\right)$. Using the normal density for X,

$$\mathbf{E}\left(e^{-X^2}\right) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2}\, e^{-\frac{x^2}{2\sigma^2}}\,dx = \frac{c}{c\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2c^2}}\,dx = \frac{c}{\sigma}$$

since the total probability of an $N(0,c^2)$ variable is one and where

$$\frac{1}{2c^2} = 1 + \frac{1}{2\sigma^2} \qquad \text{so} \qquad c = \frac{\sigma}{\sqrt{2\sigma^2+1}}.$$

Also, since $\mathbf{E}(Y) = 0$ we have $\mathbf{E}(Y^2) = \mathbf{E}(Y^2) - \mathbf{E}(Y)^2 = \mathbf{Var}(T) = \tau^2$. Thus

$$\mathbf{E}\left(e^{-X^2}Y^2\right) = \frac{\tau^2}{\sqrt{2\sigma^2-1}}.$$

4. *Suppose you roll a standard die 1200 times.*
   - *(a) What is the exact probability that the number of sixes X satisfies $200 \le X \le 210$?*
   - *(b) Approximate the probability. State carefully why you can make the approximation. You can leave your answer in terms of the cumulative distribution function $\Phi(z)$ of the standard normal variable.*

Let X be the number of successes in $n = 1200$ independent trials with the probability of success (rolling a six) is $p = \frac{1}{6}$. Thus $X \sim \text{binomial}(n = 1200, p = \frac{1}{6})$. Thus the exact probability is

$$\mathbf{P}(200 \le X \le 210) = \sum_{x=200}^{210} f_X(x) = \sum_{x=200}^{210} \binom{1200}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{1200-x}.$$

By the de Moivre-Laplace Theorem, for fixed p, the standardized binomial variable converges to the standard normal $Z \sim N(0,1)$ in distribution as $n \to \infty$. Assume that n is sufficiently large that this approximation is valid. By the rule of thumb of Math 3070, $npq = 1200\cdot\frac{1}{6}\cdot\frac{5}{6} = \frac{1000}{6} \approx 166.7$ which exceeds ten and the approximation is acceptable. The conclusion also follows from the Central Limit Theorem. This time $X = S_n = X_1 + \cdots + X_2$ where the X_i's are IID Bernoulli variables with probability of success $p = \frac{1}{6}$. Thus all $\mathbf{E}(X_i) = p$, $\mathbf{Var}(X_i) = pq$ and the mgf is $M_{X_i}(t) = q + pe^t$ which is defined for all t, thus in a neighborhood ot $t = 0$. We have $\mathbf{E}(X) = np$ and $\mathbf{Var}(X) = npq$. Both theorems imply for all $z \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{S_n - np}{\sqrt{npq}} \le z\right) = \mathbf{P}(Z \le z).$$

Assuming that for $n = 1200$ both values are close,

$$\mathbf{P}(200 \le X \le 210) = \mathbf{P}\left(\frac{200-np}{\sqrt{npq}} \le \frac{X-np}{\sqrt{npq}} \le \frac{210-np}{\sqrt{npq}}\right)$$

$$\approx \mathbf{P}\left(0 \le Z \le \frac{210-200}{\sqrt{\frac{1000}{6}}}\right) = \Phi\left(\frac{10}{\sqrt{\frac{1000}{6}}}\right) - \Phi(0) = \Phi\left(0.775\right) - \frac{1}{2}$$

which equals approx. $0.7807 - 0.5000 = 0.2807$.

5. *Suppose that X and Y be continuous variables whose joint density is defined on an infinte wedge*

$$f(x,y) = \begin{cases} e^{-x}, & \text{if } 0 < y < x < \infty; \\ 0, & \text{otherwise.} \end{cases}$$

Find $\mathbf{P}(X \le 4 \text{ and } Y \le 5)$. Find the marginal densities $f_X(x)$ and $f_Y(y)$. Are X and Y independent? Why? Find $\mathbf{E}(X)$.

The region $\{(x,y) \in \mathbb{R}^2 : x \le 4 \text{ and } y \le 5\}$ cuts a triangle from the wedge, hence since $5 > 4$,

$$\mathbf{P}(X \le 4 \text{ and } Y \le 5) = \int_{x=0}^{4} \int_{y=0}^{x} e^{-x}\, dy\, dx \left(= \int_{x=0}^{4} x\, e^{-x}\, dx = \Big[-xe^{-x} - e^{-x} \Big]\Big|_0^4 = 1 - 5e^{-4}. \right)$$

The marginals are for $x, y \ge 0$,

$$f_X(x) = \int_{y=-\infty}^{\infty} f(x,y)\, dy = \int_{y=0}^{x} e^{-x}\, dy = xe^{-x};$$
$$f_Y(y) = \int_{x=-\infty}^{\infty} f(x,y)\, dx = \int_{x=y}^{\infty} e^{-x}\, dx = \Big[-e^{-x} \Big]\Big|_y^{\infty} = e^{-y}.$$

X and Y are not independent, because, *e.g.*, $0 = f(2,3) \ne f_X(2) f_Y(3) = 2e^{-5}$.

Finally, X is distributed as a Gamma variable with $\lambda = 1$ and $r = 2$. Thus $\mathbf{E}(X) = r\lambda^{-1} = 2$, as on p. 320. Or we integrate (by guessing the antiderivative)

$$\mathbf{E}(X) = \int_0^{\infty} x^2\, e^{-x}\, dx = \Big[-(x^2 + 2x + 2)\, e^{-x} \Big]\Big|_0^{\infty} = 2.$$

6. *A continuous random variable has the cumulative distribution*

$$F_X(x) = \begin{cases} 0, & \text{if } x \le 0; \\ e^x - 1, & \text{if } 0 < x < \ln 2; \\ 1, & \text{if } \ln 2 \le x. \end{cases}$$

(a) *Find the moment generating function $M_X(t)$. Where is it defined?*

(b) *Let $Y = e^X$. Find the density $f_Y(y)$.*

The density function is

$$f_X(x) = F_X'(x) = \begin{cases} e^x, & \text{if } 0 < x < \ln 2; \\ 0, & \text{otherwise.} \end{cases}$$

Thus the moment generating function is

$$M_X(t) = \mathbf{E}(e^{tX}) = \int_0^{\ln 2} e^{tx}\, e^x\, dx = \begin{cases} \ln 2, & \text{if } t = -1; \\ \dfrac{e^{(1+t)\ln 2} - 1}{1+t}, & \text{if } t \ne -1. \end{cases}$$

It is defined for all $t \in \mathbb{R}$, since the integral is taken over a finite interval and is finite for any $t \in \mathbb{R}$. The cumulative distribution function of $Y = e^X$ is for $e^0 = 1 < y < 2 = e^{\ln 2}$,

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(e^X \le y) = \mathbf{P}(X \le \ln y) = F_X(\ln y),$$

so that

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\ln y) = \frac{F_X'(\ln y)}{y} = \begin{cases} 1, & \text{if } 1 < y < 2; \\ 0, & \text{otherwise.} \end{cases}$$

7. *Let $X_1, X_2, X_3, \ldots, X_n, \ldots$ be a sequence of mutually independent random variables. Suppose that each X_n has the Laplace density with $\lambda = \frac{1}{3}$, i.e., with density function*

$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|} \text{ for } x \in \mathbb{R}.$$

How big does n have to be so that

$$\mathbf{P}\left(\frac{|X_1 + X_2 + \cdots + X_n|}{n} \geq 1\right) \leq \frac{1}{100} \text{ ?}$$

A random variable distributed according to the Laplace density has $\mathbf{E}(X_i) = 0$, since it's symmetric about zero. Its variance is $\mathbf{Var}(X_i) = 2\lambda^{-2} = 18$ (see, *e.g.*, p. 320). If $S_n = X_1 + X_2 + \cdots + X_n$ is a sum of independent variables then, by linearity and independence,

$$\mathbf{E}\left(\frac{1}{n}S_n\right) = 0, \qquad \mathbf{Var}\left(\frac{1}{n}S_n\right) = \frac{18}{n}.$$

Applying Chebychov's Inequality, for a rv Y and $a > 0$,

$$\mathbf{P}(|Y - \mathbf{E}(Y)| \geq a) \leq \frac{\mathbf{Var}(Y)}{a^2},$$

with $Y = \frac{1}{n}S_n$ and $a = 1$,

$$\mathbf{P}\left(\left|\frac{1}{n}S_n\right| \geq 1\right) \leq \frac{18}{n},$$

which is less than 0.01 if $n \geq 1800$.

8. *Ten married couples, (Mr. & Mrs. Able, Mr. & Mrs. Baker, Mr. & Mrs. Charles, &c.) with different last names play the following party game. All 20 people write their last name on a slip of paper and put it into a hat. Then everyone takes one of the slips randomly from the hat without replacement. Anyone who chooses their own name wins a prize. Let X be the number of people who choose their own last name. Find $\mathbf{E}(X)$ and $\mathbf{Var}(X)$.*

Number the people $i = 1, 2, 3, \ldots, 20$. Let the indicator $I_i = 1$ if the ith person draws their own last name and 0 otherwise. Observe that $X = I_1 + I_2 + \cdots + I_{20}$ is the number of people who draw their own names. Then for the ith person, the number of correct names divided by the number of slips is

$$\mathbf{E}(I_i) = \mathbf{P}(i\text{th person draws own name}) = \frac{2}{20} = \frac{1}{10}.$$

Thus

$$\mathbf{E}(X) = \mathbf{E}\left(\sum_{i=1}^{20} I_i\right) = \sum_{i=1}^{20}\mathbf{E}(I_i) = 20 \cdot \frac{1}{10} = 2.$$

To compute the variance, we need

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \mathbf{E}\left(\left[\sum_{i=1}^{20} I_i\right]^2\right) - 4 = \sum_{i=1}^{20}\sum_{i=1}^{20}\mathbf{E}(I_i I_j) - 4.$$

If $i = j$ then $I_i^2 = I_i$ so $\mathbf{E}(I_i^2) = \mathbf{E}(I_i) = 0.1$. if $i \neq j$ but i and j are spouses

$$\mathbf{E}(I_i I_j) = \mathbf{P}(I_i = 1 \text{ and } I_j = 1) = \mathbf{P}(I_j = 1)\,\mathbf{P}(I_i = 1 \mid I_j = 1) = \frac{1}{10} \cdot \frac{1}{19} = \frac{1}{190}$$

since the second person has only one remaining slip with their name on it. If $i \neq k$ and i and k are not spouses, then

$$\mathbf{E}(I_i I_k) = \mathbf{P}(I_i = 1 \text{ and } I_k = 1) = \mathbf{P}(I_k = 1)\,\mathbf{P}(I_i = 1 \mid I_k = 1) = \frac{1}{10} \cdot \frac{2}{19} = \frac{2}{190}$$

since there are two slips with k's name left in the hat. Partitioning the sum,

$$\begin{aligned}
\mathbf{E}(X^2) &= \sum_i \mathbf{E}(I_i^2) + \sum_{\substack{i \neq j,\ i \text{ and } j \\ \text{are spouses}}} \mathbf{E}(I_i I_j) + \sum_{\substack{i \neq k,\ i \text{ and } k \\ \text{are not spouses}}} \mathbf{E}(I_i I_k) \\
&= 20 \cdot \frac{1}{10} + 20 \cdot \frac{1}{190} + 360 \cdot \frac{2}{190} = \frac{380 + 20 + 720}{190} = \frac{112}{19}
\end{aligned}$$

Finally,

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \frac{112}{19} - 4 = \frac{36}{19}.$$