STAT 4015 Q

Solution to the "Final"

Problem 1. European roulette wheel has 37 numbers on it, which we assume to be 1, 2, ..., 37. The roulette is rolled 4 times and the number obtained is recorded. Let X be the largest of the four numbers.

(1) Compute $\mathrm{P}(X = 20)$.
(2) Compute the $\mathbb{E}(X)$.

Solution. Let Y_i, $i = 1, \ldots, 4$, be the recorded numbers. We assume that they are independent and uniformly distributed over the 37 numbers. By the definition, $X = \max_{i=1}^4 Y_i$. Note that

$$\mathrm{P}(X \le x) = \mathrm{P}(Y_1 \le x, \ldots, Y_4 \le x) = \prod_{i=1}^{4} \mathrm{P}(Y_i \le x) = (x/37)^4 ,$$

for $x = 1, \ldots, 37$. It follows that

$$\mathrm{P}(X = 20) = \mathrm{P}(X \le 20) - \mathrm{P}(X \le 19) = (20^4 - 19^4)/37^4$$

and

$$\mathbb{E}(X) = \sum_{x=1}^{\infty} \mathrm{P}(X \ge x) = \sum_{x=1}^{37} \left[1 - (x-1)^4/37^4\right] .$$

Problem 2. The density of a random variable X is given in the form $f_X(x) = ce^{-3x^2+7x}$, for x real.

(1) Evaluate c.
(2) Find the value x with the property that $\mathrm{P}(X > x) = 0.1$

Solution. Notice that X has a normal distribution. Completing the squares we get that

$$-3x^2 + 7x = -3(x - 7/6)^2 + 49/12 = -\frac{(x - 7/6)^2}{2(1/6)} + 49/12 .$$

It follows that the mean of the random variable is $\mu = 7/6$ and the variance is $\sigma^2 = 1/6$. Completing the function into a density we get that

$$ce^{49/12}\sqrt{2\pi/6} = 1 \Longrightarrow c = e^{-49/12}/\sqrt{2\pi/6} .$$

Finally, $\mathrm{P}(X > x) = \mathrm{P}(Z > (x-\mu)/\sigma) = 0.1$. Hence, $(x-\mu)/\sigma = 1.28$ and $x = 1.28\sigma + \mu = 1.28/\sqrt{6} + 7/6$.

Problem 3. The joint density of X and Y is given in the form $f_{XY}(x, y) = ce^{y-x}y^2x^{2.3}$, for $0 \le y \le 1$ and $x > 0$.

(1) Evaluate c.
(2) Compute the marginal density of X. Are X and Y independent or not?
(3) Compute the probability $\mathrm{P}(X < Y)$.

Solution: X and Y are independent since

$$f_{XY}(x,y) = \Big\{c_1 x^{2.3} e^{-x} I_{\{x>0\}}\Big\} \times \Big\{c_2 e^y y^2 I_{\{0\le y\le 1\}}\Big\}\ ,$$

for some normalizing constants c_1 and c_2 such that $c_1 \cdot c_2 = c$. Notice that the marginal distribution of X is Gamma$(3.3, 1)$. Therefore, $c_1 = 1/\Gamma(3.3)$. In order to obtain the value of c_2 we should integrate the density of Y over the range:

$$1/c_2 = \int_0^1 y^2 e^y dy = \big[y^2 e^y\big]_0^1 - 2\int_0^1 y e^y dy = e - 2\big\{\big[ye^y\big]_0^1 - \textstyle\int_0^1 e^y dy\big\} = e - 2\ .$$

It can bee concluded that $c = 1/[(e-2)\Gamma(3.3)]$.

For the given probability one should integrate the joint density over the event:

$$\mathrm{P}(X < Y) = c\int_0^1 \Big[\int_x^1 y^2 e^y dy\Big] x^{2.3} e^{-x} dx\ .$$

This observation is sufficient for the exam. An alternative representation of the probability may be obtained by integration with respect to y, that yields

$$\begin{aligned}\int_x^1 y^2 e^y dy &= \big[y^2 e^y\big]_x^1 - 2\int_x^1 y e^y dy \\ &= e - x^2 e^x - 2\big\{\big[ye^y\big]_x^1 - \textstyle\int_x^1 e^y dy\big\} \\ &= e - x^2 e^x - 2\big\{e - xe^x - e + e^x\big\} \\ &= e - 2e^x + 2xe^x - x^2 e^x\ ,\end{aligned}$$

and then integrating with respect to x:

$$\mathrm{P}(X < Y) = \frac{e}{e-2}\mathrm{P}(X < 1) + \frac{1}{\Gamma(3.3)(e-2}\int_0^1 (-2x^{2.3} + 2x^{3.3} - x^{4.3})dx\ .$$

The CDF of X is computable via the incomplete Gamma function. The other integral can be integrated to produce an explicit result.

Problem 4. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed according to the $U(0,1)$ distribution.

(1) Identify the distribution of $Y_i = -\log X_i$ and compute its mean and its variance.
(2) Use the Central Limit Theorem in order to give and approximation to the probability $\mathrm{P}(\prod_{i=1}^n X_i \le x)$, for $0 < x < 1$.

Solution. Let $X \sim U(0,1)$. Consider the CDF of the transformed uniform for any positive x:

$$\mathrm{P}(-\log X \le x) = \mathrm{P}\big(X > e^{-x}\big) = 1 - e^{-x}\ .$$

It follows that $Y = -\log Y \sim \mathrm{Exp}(1)$, which has mean and variance both equal to one. Next,

$$\begin{aligned}\mathrm{P}\Big(\prod_{i=1}^n X_i \le x\Big) &= \mathrm{P}\Big(\sum_{i=1}^n(-\log X_i) \ge -\log x\Big) \\ &\approx \mathrm{P}(Z > (-\log x - n)/\sqrt{n}) = \Phi\big((\log x + n)/\sqrt{n}\big)\ .\end{aligned}$$

Problem 5. An urn contains 200 black and 200 white balls. They are removed from the urn in pairs. Let X be the number of pairs where both balls are black.

(1) Compute $\mathbb{E}(X)$.
(2) Compute (X).

Solution. Let X_i be the indicator of the i-th pair being composed of two black balls, for $i = 1, \dots, 200$. Notice that $X = \sum_{i=1}^{200} X_i$. By the linearity of the expectation

$$\mathbb{E}(X) = \sum_{i=1}^{200} \mathbb{E}(X_i) = 200 \cdot \mathrm{P}(X_1 = 1) = 200 \cdot \frac{200}{400} \cdot \frac{199}{399} \ .$$

The variance of X_i is

$$\mathbb{V}\mathrm{ar}(X_i) = \frac{1}{4} \cdot \frac{398}{399} \ ,$$

based on the computation of the variance of an Hypergeometric distribution. For the covariance between X_1 and X_2, notice that $\mathbb{E}(X_1 X_2) = \mathrm{P}(X_1 = 1, X_2 = 1)$. The latter is the probability that all 4 balls are black:

$$\mathrm{P}(X_1 = 1, X_2 = 1) = \frac{200}{400} \cdot \frac{199}{399} \frac{198}{398} \cdot \frac{197}{397} \ .$$

The covariance, for $i \neq j$, is thus

$$\mathbb{C}\mathrm{ov}(X_i, X_j) = \frac{200}{400} \cdot \frac{199}{399} \frac{198}{398} \cdot \frac{197}{397} - \left[\frac{200}{400} \frac{199}{399} \right]^2$$

Finally,

$$\mathbb{V}\mathrm{ar}(X) = 200 \cdot \mathbb{V}\mathrm{ar}(X_1) + 200 \cdot 199 \cdot \mathbb{C}\mathrm{ov}(X_1, X_2) \ .$$