STAT400. Sample questions for midterm 2.

1. *In this problem you may neglect the probability of twins.*

(a) A family has 5 children. Find the probability that they have 2 boys and 3 girls.

(b) A family decides to have children until they have 3 girls. Find the probability that they have 2 boys.

(c) A family decides to have children until they have 3 girls. Let C be the total number of children in the family. Compute EC and VC.

Solution. (a) The number of boys has binomial distribution with parameters $(5, 1/2)$. So the answer is $\binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \binom{5}{2}\left(\frac{1}{2}\right)^5$.

(b) The number of boys has negative binomial distribution with parameters $(3, 1/2)$. So the answer is $\binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \binom{4}{2}\left(\frac{1}{2}\right)^5$.

(c) $C = B+3$ where B is the number of boys. So using the formulas for negative binomial distribution we get $E(C) = E(B) + 3 = 3 + 3\frac{1/2}{1/2} = 6$, $V(C) = V(B) = 3\frac{1/2}{(1/2)^2} = 6$.

2. *(a) 40% of lightbulb produced by Shining Beauty company have suboptimal performance. 5 bulbs are chosen for test. Find the probability that 2 bulbs or less have suboptimal performance.*

(b) 40% of lightbulb produced by Shining Beauty company have suboptimal performance. 500 bulbs are chosen for test. Compute approximately the probability that 210 bulbs or less have suboptimal performance.

(c) 0.4% of lightbulb produced by Light company have suboptimal performance. 500 bulbs are chosen for test. Compute approximately the probability that 2 bulbs or less have suboptimal performance.

Solution. (a) The number of suboptimal has binomial distribution with parameters $(5, 0.4)$. So the answer is $\binom{5}{2}(0.4)^2(0.6)^3 + \binom{5}{1}(0.4)^1(0.6)^4 + (0.6)^5$.

(b) Let S be the number of suboptimal bulbs. Then $ES = 500 * 0.4 = 200$, $VS = 500 * 0.4 * 0.6 = 120$, so $\sigma_S = \sqrt{VS} \approx 11$. Thus using normal approximation to binomial we get $S \approx N(200, 120) \approx 200 + 11Z$ where Z is a standard normal. Hence

$$P(S \leq 210) = P(S \leq 210.5) = P(200{+}11 \leq 210.5) = P\left(Z \leq \frac{10.5}{11}\right) \approx P(Z \leq 0.95) \approx 0.83.$$

(c) Let S be the number of suboptimal bulbs. Then $ES = 500 * 0.004 = 2$. Let Y denote Poisson random variable with parameter 2. Using Poisson approximation to binomial we get

$$P(S \le 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) = e^{-2}\left(1 + 2 + \frac{2^2}{2}\right) \approx 0.68.$$

3. *Calls to a customer service center form Poisson process with intensity 4 calls per hour.*
(a) Find the probability that there are less than 3 calls between 9:00 and 10:00.
(b) John works from 9:00 am to 3:00 pm. His shift is divided into 6 one hour intervals. John calls an interval easy if there are less than 3 calls during that interval. Find the probability that during a particular day he has at least one easy interval.
(c) Find the probability that the first call during John watch arrives between 9:00 and 9:20 and the second between 9:20 and 9:40.

Solution. (a) The number of calls between 9:00 and 10:00 has Poisson distribution with parameter 4. Hence

$$P(N(9, 10) \le 2) = e^{-4}\left(1 + 4 + \frac{4^2}{2}\right) \approx 0.24.$$

(b) The probability that none of the intervals are easy is $0.76^6 \approx 0.2$. So the probability of having at least one easy interval. $1 - 0.2 = 0.8$.

(c) For this to occur there needs to be exactly one call between 9:00 and 9:20 (probability $\frac{4}{3}e^{-4/3}$ and more than one call between 9:20 and 9:40 (probability $1 - e^{-4/3}$) so the answer is $\frac{4}{3}e^{-4/3}(1 - e^{-4/3})$.

4. *Let X have cumultive density function*

$$F(x) = \begin{cases} 0, & \text{if } x \le 0 \\ \frac{x^2+x^4}{2} & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x \ge 0 \end{cases}.$$

(a) Compute density of X.
(b) Compute EX and VX.
(c) Let $Y = X^3$. Compute EY and VY.

Solution. (a) $f(x) = F'(x) = x + 2x^3$.

$$(b) \quad EX = \int_0^1 x(x + 2x^3)dx = \int_0^1 x^2 dx + 2\int_0^1 x^4 dx = \frac{11}{15}.$$

$$EX^2 = \int_0^1 x^2(x + 2x^3)dx = \int_0^1 x^3 dx + 2\int_0^1 x^5 dx = \frac{7}{12}.$$

$$V(X) = \frac{7}{12} - \left(\frac{11}{15}\right)^2 = \frac{41}{900}.$$

$$(c) \quad EY = \int_0^1 x^3(x+2x^3)dx = \int_0^1 x^4 dx + 2\int_0^1 x^6 dx = \frac{17}{35}.$$

$$EY^2 = \int_0^1 x^6(x+2x^3)dx = \int_0^1 x^7 dx + 2\int_0^1 x^9 dx = \frac{12}{35}.$$

$$V(X) = \frac{12}{35} - \left(\frac{17}{35}\right)^2 = \frac{131}{1225}.$$

5. *Let X have normal distribution with mean 2 and standard deviation 2.*
(a) Compute 25 and 75th percentiles.
(b) Compute $P(X > 5)$.
(c) Compute $E(X^2)$.

Solution. $X = 2 + 2Z$ where $Z \sim N(0,1)$. Thus $P(X \le x) = P(Z \le \frac{x-2}{2})$. So if $P(Z \le \frac{x_{0.75}-2}{2}) = 0.75$ then $\frac{x_{0.75}-2}{2} = 0.67$ and so $x_{0.75} = 2 + 2 * 067 = 3.34$. Likewise $x_{0.25} = 2 - 2 * 0.67 = 0.66$.

$$(b) \quad P(X > 5) = P(2 + 2Z > 5) = P(Z > 1.5) \approx 0.0668.$$

$$(c) \quad E(X^2) = V(X) + (EX)^2 = 4 + 4 = 8.$$

6. *Let the distribution of X and Y be given in the following table.*

$X\backslash Y$	1	2	3
0	.05	.10	15
1	.05	.05	.10
2	.20	.05	.25

(a) Compute the marginal distributions of X and Y.
(b) Compute $P(X = Y)$.
(c) Compute $Cov(X, Y)$.

Solution.

(a)

$X\backslash Y$	1	2	3	
0	.05	.10	15	0.30
1	.05	.05	.10	0.20
2	.20	.05	.25	0.50
	0.30	0.20	0.50	

(b) $P(X = Y = 1) + P(X = Y = 2) = 0.10$.

(c) $EX = 0 * 0.30 + 1 * 0.2 + 2 * 0.5 = 1.2$, $EY = 1 * 0.30 + 2 * 0.2 + 3 * 0.5 = 3.2$,

$E(XY) = 0*(0.05+0.10+0.15)+1*0.05+2*(0.20+0.05)+3*0.10+4*0.05+6*0.25 = 2.55$

$$Cov(X, Y) = 2.55 - 1.2 * 2.2 = -0.09.$$

7. *Let (X,Y) have uniform distribution on the trapezoid $0 \le y \le 1$, $0 \le x \le 1+y$.*
(a) Compute the marginal distributions of X and Y.
(b) Compute $V(X)$.
(c) Compute $Cov(X,Y)$.

Solution. Area(Trapezoid)$= \frac{1+2}{2} * 1 = \frac{3}{2}$ so the density of $(X,Y) = \frac{2}{3}$.

(a) Hence if $x \in [0,1]$ then $f_X(x) = 1 * \frac{2}{3} = \frac{2}{3}$ and if $x \in [1,2]$ then $f_X(x) = \dfrac{2}{3}(1-(1-x)) = \dfrac{2(2-x)}{3}$ since the slanted side of the trapezoid has form $x = 1+y$, that is $y = x-1$. $f_Y(y) = \dfrac{2}{3} * (1+y) = \dfrac{2(1+y)}{3}$.

$$(b) \quad EX = \int_0^1 \frac{2}{3} x dx + \int_1^2 \frac{2}{3}(2-x) x dx - = \frac{x^2}{3}\Big|_0^1 + \int_1^2 \frac{4x}{3} dx - \int_1^2 \frac{2x^2}{3} dx = \frac{1}{3} + \frac{2x^2}{3}\Big|_1^2 - \frac{2x^3}{9}\Big|_1^2 = \frac{7}{9}.$$

Likewise

$$EY = \int_0^1 \frac{2(1+y)}{3} y dy = \int_0^1 \frac{2y}{3} dy + \int_0^1 \frac{2y^2}{3} dy = \frac{1}{3} + \frac{2}{9} = \frac{5}{9}.$$

$$EX^2 == \int_0^1 \frac{2}{3} x^2 dx + \int_1^2 \frac{2}{3}(2-x) x^2 dx - = \frac{2x^2}{9}\Big|_0^1 + \int_1^2 \frac{4x^2}{3} dx - \int_1^2 \frac{2x^3}{3} dx = \frac{1}{3} + \frac{4x^3}{9}\Big|_1^2 - \frac{x^4}{6}\Big|_1^2 = \frac{5}{6}.$$

$$V(X) = \frac{5}{6} - \left(\frac{7}{9}\right)^2 = \frac{37}{162}.$$

$$(c) E(XY) = \frac{2}{3} \int_0^1 y \left(\int_0^{1+y} x dx \right) dy = \int_0^1 \frac{y(1+y)^2}{3} = \frac{1}{3} \int_0^1 (y + 2y^2 + y^3) dy = \frac{17}{36}.$$

Hence $Cov(X,Y) = \dfrac{17}{36} - \dfrac{7}{9} * \dfrac{5}{9} = \dfrac{13}{324}$.

8. *Let X be independent, X have uniform distribution on $[0,1]$ and Y have exponential distribution with parameters 1. Let $Z = X + Y$.*
(a) Compute the density of Z.
(b) Compute $Cov(X,Z)$.
(c) Compute $P(3Y > Z)$.

Solution. Recall that x has density equal to 1 on $[0,1]$ and equal to 0 elsewhere and y has density e^{-y} if $y \ge 0$ and 0 elsewhere.

(a) If $z \le 1$ then

$$p(z) = \int_0^z e^{-x} dx = 1 - e^{-z}.$$

If $z \ge 1$ then

$$p(z) = \int_{z-1}^z e^{-x} dx = e^{-z}(e-1).$$

(b) $Cov(X,Z) = Cov(X, X+Y) = Cov(X,X) + Cov(X,Y) = V(X) + 0 = \frac{1}{12}$.

$$(c) \quad P(3Y > X+Y) = P(2Y > X) = P(Y > X/2) = \int_0^1 e^{-x/2} dx = 2(1 - e^{-1/2}) \approx 0.78.$$

9. *Let $S = X_1 + X_2 + \dots X_{162}$ where X_j are independent identically distributed random variables. Suppose that X_j have density equal to $2x$ if $0 \le x \le 1$ and equal to 0 otherwise.*

(a) Compute ES.

(b) Compute VS.

(c) Compute approximately $P(S > 110)$.

Solution. (a) $EX_1 = \int_0^1 2x * x dx = \int_0^1 2x^2 dx = \dfrac{2}{3}$. Hence $ES = 162 * \dfrac{2}{3} = 108$.

(b) $E(X_1^2) = \int_0^1 2x * x^2 dx = \int_0^1 2x^3 = \dfrac{1}{2}$. Hence $VX_1 = \frac{1}{2} - \frac{2^2}{3^2} = \dfrac{1}{18}$. Thus $VS = 162 * \frac{1}{18} = 9$ and $\sigma_S = 3$.

(c) By the central Limit Theorem $S \approx N(108, 9)$, that is $S \approx 108 + 3Z$ where $Z \sim N(0,1)$. Hence

$$P(S > 110) \approx P(108 + 3Z > 110) = P(Z > \frac{2}{3}) \approx 0.75.$$

10. *60 scientists from 30 universities attend a conference. The conference includes a lunch in a cafetria which has 30 tables each suitable for 2 people. The people seated for lunch at random. Let $X_j = 1$ if the scientists from university j seat togather and $X_j = 0$ otherwise.*

(a) Compute $E(X_1)$ and $V(X_1)$.

(b) Compute $Cov(X_1, X_2)$.

(c) Let $X = X_1 + X_2 + \dots X_{30}$ be the number of tables occupied by the people from the same university. Compute EX and VX.

Solution. (a) X_1 has Bernoulli distribution and $P(X_1 = 1) = \dfrac{1}{59}$. Hence $E(X_1) = 1 = \dfrac{1}{59}$, $V(X_1) = \dfrac{1}{59} * \dfrac{58}{59} = \dfrac{58}{59}^2$.

(b) $E(X_1 X_2) = P(X_1 = 1, X_j = 2) = P(X_1 = 1) P(X_2 = 1 | X_1 = 1) = \dfrac{1}{59} * \dfrac{1}{57}$. $Cov(X_1, X_2) = \dfrac{1}{59}\left(\dfrac{1}{57} - \dfrac{1}{59}\right) = \dfrac{2}{57 * 59^2}$.

(c) $EX = E(X_1 + X_2 + \dots X_{30}) = \dfrac{30}{59}$.

$$V(X) = \sum_{j=1}^{30} V(X_j) + 2 \sum_{i<j} Cov(X_i, X_j) = 30 * \frac{58}{59^2} + 30 * 29 * \frac{2}{57 * 59^2} = \frac{30 * 58^2}{57 * 59^2}.$$