MMAT5340 - Probability and Stochastic Analysis 2018-2019

Suggested Solutions of Mid-term

1. (25marks)

a. We need to choose a committee of six people: three French and three Germans, out of six French and seven Germans. How many ways are there to do this? Your answer should be in the form of a number, say 10 or 23.

b. Consider two independent events A and B. Find $P(B)$, if you know that

$$P(A) = 2P(B), \text{ and } P(A \setminus B) = 0.1.$$

c. Toss two fair dice. Let A = the sum on the dice is even, B = the first die is even. What is $P(A|B)$ and $P(B|A)$?

Solution:

a. The number is $\binom{6}{3} \cdot \binom{7}{3} = 20 \cdot 35 = 700$.

b. Since A and B are independent, we have

$$P(A) = P(A \cap B) + P(A \setminus B) = P(A)P(B) + 0.1.$$

Let $x = P(B)$. Then $P(A) = 2x$, and the equation becomes

$$2x = 2x^2 + 0.1$$
$$20x^2 - 20x + 1 = 0.$$

Solving, we have $P(B) = x = \dfrac{1}{10}(5 - 2\sqrt{5}) \approx 0.0528$ ($x = \dfrac{1}{10}(5 + 2\sqrt{5})$ is rejected since $P(A) = 2x \approx 1.7746 > 1$.)

c. Note that

$$P(A) = \frac{3 \times 3 + 3 \times 3}{6^2} = \frac{1}{2}, \quad P(B) = \frac{3}{6} = \frac{1}{2} \quad \text{and} \quad P(A \cap B) = \frac{3 \times 3}{6^2} = \frac{1}{4}.$$

Hence

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{1/2} = \frac{1}{2},$$

and

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{1/2} = \frac{1}{2}.$$

2. (20marks)

a. Toss a fair coin 3 times. Let X be the total number of heads. What is the probability space Ω, and what is the distribution of X?

b. An insurance company determines that N, the number of claims received in a week, is a random variable with

$$P(N = n) = \frac{1}{2^{n+1}}, n = 0, 1, 2, \ldots$$

The company also determines that the number of claims received in a given week is independent of the number of claims received in any other week. Determine the probability that exactly 7 claims will be received during a given two-week period.

c. Consider two random variables X and Y with joint distribution

$$P(X = -1, Y = 0) = P(X = 1, Y = 0) = P(X = 0, Y = 1) = P(X = 0, Y = -1) = \frac{1}{4}.$$

Are they independent? Find $\mathrm{Cov}(X, Y)$.

d. A client has losses $X \sim \mathrm{Geo}(0.6)$. The insurance policy has a deductible of 3: if the losses are 3 or less, the company does not pay anything, while if the losses are greater than 3, the company pays 80% of the difference. Find the expected value of the payment.

Solution:

a. The probability space Ω consists of all possible elementary outcomes. Thus

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

For $k = 0, 1, 2, 3$,

$$P(X = k) = \binom{3}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{3-k} = \frac{1}{8}\binom{3}{k}.$$

b.

$$\begin{aligned}
&P(\text{exactly 7 claims in a given two-week period}) \\
&= \sum_{n=0}^{7} P(n \text{ claims in 1st week and } 7-n \text{ claims in 2nd week}) \\
&= \sum_{n=0}^{7} P(n \text{ claims in 1st week and })P(7-n \text{ claims in 2nd week}) \\
&= \sum_{n=0}^{7} \frac{1}{2^{n+1}} \cdot \frac{1}{2^{(7-n)+1}} \\
&= \sum_{n=0}^{7} \frac{1}{2^9} \\
&= \frac{8}{2^9} = \frac{1}{2^6}.
\end{aligned}$$

c. Note that

$$P(X = 1) = P(X = 1, Y = 0) = \frac{1}{4},$$

and

$$P(Y=0) = P(X=-1, Y=0) + P(X=1, Y=0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

However

$$P(X=1, Y=0) = \frac{1}{4} \neq \frac{1}{4} \cdot \frac{1}{2} = P(X=1)P(Y=0).$$

Hence X and Y are not independent. Since

$$\begin{aligned}
\mathbf{E}(X) &= (-1)\frac{1}{4} + (1)\frac{1}{4} + (0)\left(\frac{1}{4} + \frac{1}{4}\right) = 0 \\
\mathbf{E}(Y) &= (-1)\frac{1}{4} + (1)\frac{1}{4} + (0)\left(\frac{1}{4} + \frac{1}{4}\right) = 0 \\
\mathbf{E}(XY) &= (-1)(0)\frac{1}{4} + (1)(0)\frac{1}{4} + (0)(1)\frac{1}{4} + (0)(-1)\frac{1}{4} = 0
\end{aligned}$$

we have

$$\mathrm{Cov}(X, Y) = \mathbf{E}(XY) - (\mathbf{E}X)(\mathbf{E}Y) = 0.$$

d. Since $X \sim \mathrm{Geo}(0.6)$, we have $\mathbf{E}(X) = \dfrac{1}{0.6}$ and

$$P(X=k) = (1-0.6)^{k-1}(0.6) = (0.4)^{k-1}(0.6), \quad k = 1, 2, 3, \ldots$$

Let Y be the payment. Then

$$Y = \begin{cases} (X-3) \times 0.8 & \text{if } X > 3 \\ 0 & \text{if } 1 \leq X \leq 3. \end{cases}$$

Now

$$\begin{aligned}
\mathbf{E}(Y) &= \sum_{k=4}^{\infty} 0.8(k-3)P(X=k) \\
&= 0.8\left(\sum_{k=1}^{\infty}(k-3)P(X=k)\right) - 0.8\left[(1-3)P(X=1) + (2-3)P(X=2) + (3-3)P(X=3)\right] \\
&= 0.8\mathbf{E}(X) - 0.8 \times 3 - 0.8\left[-2(0.6) - (0.4)(0.6)\right] \\
&= \frac{32}{375}.
\end{aligned}$$

3. (30marks)

a. Two random variables X and Y have joint density

$$\begin{cases} c(x+y), & 1 \leq x \leq 3, 0 \leq y \leq 3 \\ 0, & \text{otherwise}, \end{cases}$$

where c is some constant. Find c and find the probability $P(X \geq 2, Y \leq 2)$.

b. Consider (X, Y) with density

$$p(x, y) = cxy, \text{ for } 0 \leq x, y; x + y \leq 1,$$

where is c is some constant. Find c. Find $\mathbb{E}X$. Find $\mathrm{corr}(X, Y)$. Find $P(X \leq 0.5 | Y \geq 0.5)$.

c. There are $N = 10000$ car drivers. Each gets into an accident with probability 15%. In case of an accident, the losses are distributed uniformly on $[0, 10]$. Assume independence. Find the probability that the total losses exceed 8000. Find the value at risk at the confidence level 95%.

Solution:

a. Since

$$\begin{aligned}
1 &= \int p(x,y)dxdy \\
&= \int_1^3 \int_0^3 c(x+y)dydx \\
&= c\int_1^3 (xy + \frac{1}{2}y^2)\Big|_0^3 dx \\
&= c\int_1^3 (3x + \frac{9}{2})dx \\
&= c(\frac{3}{2}x^2 + \frac{9}{2}x)\Big|_1^3 \\
&= 21c,
\end{aligned}$$

we have $c = \dfrac{1}{21}$.

$$\begin{aligned}
P(X \geq 2, Y \leq 2) &= \int_2^3 \int_0^2 \frac{1}{21}(x+y)dydx \\
&= \frac{1}{21}\int_2^3 (xy + \frac{1}{2}y^2)\Big|_0^2 dx \\
&= \frac{1}{21}\int_2^3 (2x+2)dx \\
&= \frac{1}{21}(x^2 + 2x)\Big|_2^3 \\
&= \frac{1}{3}.
\end{aligned}$$

b. Since

$$\begin{aligned}
1 &= \int p(x,y)dxdy \\
&= c\int_0^1 \int_0^{1-y} xy \ dxdy \\
&= c\int_0^1 y(1-y)^2 \ dy \\
&= \frac{c}{24},
\end{aligned}$$

we have $c = 24$. Now

$$\begin{aligned}\mathbf{E}(X) &= \int xp(x,y)dxdy \\ &= 24\int_0^1\int_0^{1-x} x^2y\ dydx \\ &= 24\int_0^1 \frac{1}{2}x^2(1-x)^2dx \\ &= \frac{2}{5},\end{aligned}$$

$$\begin{aligned}\mathbf{E}(X^2) &= \int x^2p(x,y)dxdy \\ &= 24\int_0^1\int_0^{1-x} x^3y\ dydx \\ &= 24\int_0^1 \frac{1}{2}x^3(1-x)^2dx \\ &= \frac{1}{5},\end{aligned}$$

and similarly, $\mathbf{E}(Y) = \frac{2}{5}$ and $\mathbf{E}(Y^2) = \frac{1}{5}$. Moreover,

$$\begin{aligned}\mathbf{E}(XY) &= \int xyp(x,y)dxdy \\ &= 24\int_0^1\int_0^{1-x} x^2y^2\ dydx \\ &= 24\int_0^1 \frac{1}{3}x^2(1-x)^3dx \\ &= \frac{2}{15}.\end{aligned}$$

Thus

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{1}{5} - \left(\frac{2}{5}\right)^2 = \frac{1}{25},$$

$$\mathrm{Cov}(X,Y) = \mathbf{E}(XY) - (\mathbf{E}(X))(\mathbf{E}(Y)) = \frac{2}{15} - \left(\frac{2}{5}\right)^2 = -\frac{2}{75},$$

and therefore

$$\mathrm{corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} = \frac{-\frac{2}{75}}{\sqrt{\frac{1}{25}}\sqrt{\frac{1}{25}}} = -\frac{2}{3}.$$

Since $p(x,y) = 0$ when $x + y \geq 1$, we have $P(X \geq 0.5 \text{ and } Y \geq 0.5) = 0$, and hence

$$P(X \leq 0.5 \text{ and } Y \geq 0.5) = P(Y \leq 0.5) - P(X \geq 0.5 \text{ and } Y \geq 0.5) = P(Y \leq 0.5).$$

Therefore

$$P(X \leq 0.5 | Y \geq 0.5) = \frac{P(X \leq 0.5 \text{ and } Y \geq 0.5)}{P(Y \geq 0.5)} = 1.$$

c. Let X_i be the loss of driver i. Then $X_1, X_2, \ldots, X_N$ are i.i.d. with

$$\mu := \mathbf{E}(X_i) = (0.15)\frac{10+0}{2} = 0.75,$$

$$\mathbf{E}(X_i^2) = (0.15)\frac{10^2 + (10)(0) + 0^2}{3} = 5,$$

$$\sigma^2 := \mathrm{Var}(X_i) = \mathbf{E}(X_i^2) - \mathbf{E}(X_i)^2 = 5 - 0.75^2 = 4.4375.$$

Then the total loss is $S_N := X_1 + X_2 + \cdots + X_N$. By Normal approximation,

$$\begin{aligned} P(S_N > 8000) &= P\left(\frac{S_N - N\mu}{\sqrt{N}\sigma} > \frac{8000 - (10000)(0.75)}{100\sqrt{4.4375}}\right) \\ &= P\left(\frac{S_N - N\mu}{\sqrt{N}\sigma} > 2.3736\right) \\ &\approx 0.0088. \end{aligned}$$

The value at risk VaR_α at the confidence level $\alpha = 95\%$ is given by

$$\mathrm{VaR}_\alpha = \mu N + x_\alpha \sigma \sqrt{N} = (0.75)(10000) + (1.645)\sqrt{4.4375}\sqrt{10000} \approx 7846.53.$$

4. (15marks) Proposition A is on the ballot. A poll of $N = 200$ people finds that 112 of them support A. A week later, another poll of $M = 250$ people finds that 134 of them support A.

a. Can we reject the hypothesis that the support remains unchanged, with confidence level 95%?

b. Assume you have the following hypotheses for the share of support: $p = 50\%$, $p = 60\%$, $p = 40\%$, with prior probabilities 50%, 25%, 25%. What are the posterior probabilities after the first poll? After the first and second poll?

Solution:

a. Let H_0 be the hypothesis that the support remains unchanged, that is each vote is distributed as Bernoulli random variable with $p = \frac{112}{200} = 0.56$. This distribution has mean $\mu = p = 0.56$ and variance $\sigma^2 = 0.56 \cdot 0.44 = 0.2464$. At the confidence level 95%, we should check whether $\overline{x} = \frac{134}{250} = 0.536$ satisfies

$$-\frac{x_{97.5\%}\sigma}{\sqrt{N_2}} + \mu \le \overline{x} \le \frac{x_{97.5\%}\sigma}{\sqrt{N_2}} + \mu$$

If this is true, then we do not reject the hypothesis H_0. It turns out that

$$\frac{1.960 \cdot \sqrt{0.2464}}{\sqrt{250}} + 0.56 \approx 0.6215 > 0.536 = \overline{x},$$

and

$$-\frac{1.960 \cdot \sqrt{0.2464}}{\sqrt{250}} + 0.56 \approx 0.4985 < 0.536 = \overline{x}.$$

Therefore, we do not reject the hypothesis H_0.

b. Let H_1, H_2, H_3 be the hypotheses:

$$H_1 : p = 50\%; \quad H_2 : p = 60\%; \quad H_3 : p = 40\%,$$

with prior probabilities

$$P(H_1) = 50\%, \quad P(H_2) = 25\%, \quad P(H_3) = 25\%.$$

Let A_1, A_2 be the event that the first and second polls happened, respectively. Then

$$P(A_1 \mid H_1) = \binom{200}{112} 0.5^{112} 0.5^{88} \approx 0.0134,$$

$$P(A_1 \mid H_2) = \binom{200}{112} 0.6^{112} 0.4^{88} \approx 0.0293,$$

$$P(A_1 \mid H_3) = \binom{200}{112} 0.4^{112} 0.6^{88} \approx 0.0000.$$

By Bayes's formula,

$$\begin{aligned} P(H_1 \mid A_1) &= \frac{P(A_1 \mid H_1)P(H_1)}{P(A_1 \mid H_1)P(H_1) + P(A_1 \mid H_2)P(H_2) + P(A_1 \mid H_3)P(H_3)} \\ &= 0.4774, \end{aligned}$$

and similarly,

$$P(H_2 \mid A_1) = 0.5225, \quad P(H_3 \mid A_1) = 0.0000.$$

Now, after the second poll,

$$P(A_2 \mid H_1) = \binom{250}{134} 0.5^{134} 0.5^{116} \approx 0.0264,$$

$$P(A_2 \mid H_2) = \binom{250}{134} 0.6^{134} 0.4^{116} \approx 0.0062,$$

$$P(A_2 \mid H_3) = \binom{250}{134} 0.4^{134} 0.6^{116} \approx 0.0000.$$

By Bayes's formula,

$$\begin{aligned} P(H_1 \mid A_2) &= \frac{P(A_2 \mid H_1)P(H_1 \mid A_1)}{P(A_2 \mid H_1)P(H_1 \mid A_1) + P(A_1 \mid H_2)P(H_2 \mid A_1)) + P(A_1 \mid H_3)P(H_3 \mid A_1)} \\ &= 0.7963, \end{aligned}$$

and similarly,

$$P(H_2 \mid A_2) = 0.2037, \quad P(H_3 \mid A_2) = 0.0000.$$

5. (20marks)

a. For independent random variables

$$X \sim \mathcal{N}(1, 3), Y \sim \mathcal{N}(0, 2), Z \sim \mathcal{N}(4, 1),$$

consider the random variable

$$U := 2X - 4Y - Z + 5.$$

Find the expectation, variance, and the moment generating function for U.

b. Let $Y_1, \ldots, Y_{150}$ each be the number of tosses for a fair coin you need to get your first Heads. Assume all these random variables are independent. Estimate

$$P(|Y_1 + \ldots + Y_{150} - 300| \geq 50),$$

using Chebyshev's inequality.

c. Two random variables X and Y have joint density

$$p(x, y) = cx, 0 \leq x \leq 1, x \leq y \leq x + 1; p(x, y) = 0 \text{ otherwise},$$

where c is constant. Find c. Determine the conditional variance of Y given $X = x$.

Solution:

a. Since X, Y, Z are independent,

$$\begin{aligned}
\mathbf{E}(U) &= \mathbf{E}(2X - 4Y - Z + 5) \\
&= 2\mathbf{E}(X) - 4\mathbf{E}(Y) - \mathbf{E}(Z) + 5 \\
&= 2(1) - 4(0) - (4) + 5 \\
&= 3,
\end{aligned}$$

$$\begin{aligned}
\text{Var}(U) &= \text{Var}(2X - 4Y - Z + 5) \\
&= 2^2\text{Var}(X) + 4^2\text{Var}(Y) + \text{Var}(Z) \\
&= 4(3) + 16(2) + (1) \\
&= 45,
\end{aligned}$$

$$\begin{aligned}
F_U(t) &= \mathbf{E}\left(e^{t(2X-4Y-Z+5)}\right) \\
&= \mathbf{E}\left(e^{(2t)X}\right)\mathbf{E}\left(e^{(-4t)Y}\right)\mathbf{E}\left(e^{(-t)Z}\right)e^{5t} \\
&= \exp\left(2t(1) + \frac{(2t)^2(3)}{2}\right)\exp\left((-4t)(0) + \frac{(-4t)^2(2)}{2}\right)\exp\left(-t(4) + \frac{(-t)^2(1)}{2}\right)e^{5t} \\
&= \exp\left(3t + \frac{45t^2}{2}\right).
\end{aligned}$$

b. Since $Y_1, \ldots, Y_{150}$ are i.i.d with $Y_i \sim \text{Geo}(0.5)$, we have

$$\mathbf{E}(Y_i) = \frac{1}{0.5} = 2, \quad \text{Var}(Y_i) = \frac{1 - 0.5}{0.5^2} = 2,$$

and hence

$$\mu = \mathbf{E}(Y_1 + \cdots + Y_{150}) = 150 \cdot 2 = 300, \quad \sigma^2 = \text{Var}(Y_1 + \cdots + Y_{150}) = 150 \cdot 2 = 300.$$

Therefore,

$$P(|Y_1 + \ldots + Y_{150} - 300| \geq 50) \leq \frac{\sigma^2}{50^2} = \frac{300}{50^2} = \frac{3}{25}.$$

c.

$$\begin{aligned} 1 &= \int_0^1 \int_x^{x+1} cx\ dydx \\ &= \int_0^1 cxdx \\ &= \frac{c}{2}. \end{aligned}$$

Hence $c = 2$. Now

$$p_X(x) = \int p(x,y)dy = \int_x^{x+1} 2x\ dy = 2x, \quad \text{if } 0 \leq x \leq 1,$$

so that

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X}(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1,\ x \leq y \leq x+1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, for $0 \leq x \leq 1$, $(Y|X = x) \sim U[x, x+1]$, and hence

$$\text{Var}(Y|X = x) = \frac{1}{12}(x + 1 - x)^2 = \frac{1}{12}.$$

Alternative method: For $0 \leq x \leq 1$,

$$\begin{aligned} \mathbf{E}(Y|X = x) &= \int y\, p_{Y|X}(y|x)\ dy \\ &= \int_x^{x+1} y\ dy \\ &= x + \frac{1}{2}; \end{aligned}$$

$$\begin{aligned} \mathbf{E}(Y^2|X = x) &= \int y^2\, p_{Y|X}(y|x)\ dy \\ &= \int_x^{x+1} y^2\ dy \\ &= \frac{1}{3}(3x^2 + 3x + 1); \end{aligned}$$

so that

$$\begin{aligned} \text{Var}(Y|X = x) &= \mathbf{E}(Y^2|X = x) - [\mathbf{E}(Y|X = x)]^2 \\ &= \frac{1}{3}(3x^2 + 3x + 1) - (x + \frac{1}{2})^2; \\ &= \frac{1}{12}. \end{aligned}$$

For $x < 0$ or $x > 1$, $\text{Var}(Y|X = x) = 0$.