STA 131A Introduction to Probability Theory

Practice Final Exam – Version A Solution

Instructor: Dogyoon Song

Problem 1. Warm-up

(a)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{2}{5} + \frac{1}{2} - \frac{1}{5} = \frac{7}{10}.$$

Also,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/5}{1/2} = \frac{2}{5}.$$

Since

$$P(A)P(B) = \frac{2}{5} \cdot \frac{1}{2} = \frac{1}{5} = P(A \cap B),$$

the events A and B are independent.

(b) Since $Y = (X-1)^2$,

$$\begin{cases} Y = 0 & \text{when } X = 1, \\ Y = 1 & \text{when } X = 0 \text{ or } X = 2. \end{cases}$$

Therefore,

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

Thus,

$$p_Y(y) = \begin{cases} 1/2, & y = 0, \\ 1/2, & y = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Hence

$$\mathbb{E}[Y] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

(c)

$$P(X \geq 1/2) = \int_{1/2}^{1} 2x\,dx = \left[x^2\right]_{1/2}^{1} = 1 - \frac{1}{4} = \frac{3}{4}.$$

Also,

$$\mathbb{E}[X] = \int_0^1 x \cdot 2x\,dx = 2\int_0^1 x^2\,dx = \frac{2}{3}.$$

(d) For president, vice president, and treasurer, order matters:

$$\frac{8!}{(8-3)!} = 8 \cdot 7 \cdot 6 = 336.$$

For a 3-person committee, order does not matter:

$$\binom{8}{3} = \frac{8!}{(8-3)!\,3!} = 56.$$

(e) Standardize:

$$P(X > 85) = P\left(Z > \frac{85-70}{10}\right) = P(Z > 1.5).$$

Thus,

$$P(X > 85) = 1 - \Phi(1.5).$$

Using $\Phi(1.5) \approx 0.9332$,

$$P(X > 85) \approx 1 - 0.9332 = 0.0668.$$

Problem 2. Conditioning, Bayes' rule, and expectation

(a) By the law of total probability,

$$\begin{aligned} P(E) &= P(E \mid H = A)P(H = A) + P(E \mid H = B)P(H = B) + P(E \mid H = C)P(H = C) \\ &= 0.8 \cdot \frac{1}{4} + 0.3 \cdot \frac{1}{2} + 0.1 \cdot \frac{1}{4} \\ &= 0.2 + 0.15 + 0.025 = 0.375 = \frac{3}{8}. \end{aligned}$$

(b) By Bayes' rule,

$$P(H = A \mid E) = \frac{P(E \mid H = A)P(H = A)}{P(E)} = \frac{0.8(1/4)}{3/8} = \frac{8}{15}.$$

Similarly,

$$\begin{aligned} P(H = B \mid E) &= \frac{0.3(1/2)}{3/8} = \frac{2}{5}, \\ P(H = C \mid E) &= \frac{0.1(1/4)}{3/8} = \frac{1}{15}. \end{aligned}$$

Thus,

$$P(H = A \mid E) = \frac{8}{15}, \qquad P(H = B \mid E) = \frac{2}{5}, \qquad P(H = C \mid E) = \frac{1}{15}.$$

(c) The unconditional expected payoff is

$$\mathbb{E}[R] = 12 \cdot \frac{1}{4} + 4 \cdot \frac{1}{2} + (-4) \cdot \frac{1}{4} = 3 + 2 - 1 = 4.$$

Using the posterior probabilities from part (b),

$$\begin{aligned} \mathbb{E}[R \mid E] &= 12P(H = A \mid E) + 4P(H = B \mid E) - 4P(H = C \mid E) \\ &= 12 \cdot \frac{8}{15} + 4 \cdot \frac{2}{5} - 4 \cdot \frac{1}{15} \\ &= \frac{96}{15} + \frac{24}{15} - \frac{4}{15} = \frac{116}{15} \approx 7.733. \end{aligned}$$

Problem 3. Joint PDF and derived distribution

(a) The support is the triangular region

$$0 \le y \le x \le 1.$$

Its area is 1/2, so normalization requires $c \cdot \frac{1}{2} = 1$, and hence $c = 2$.

Alternatively, we can formally write the argument above as follows:

$$\int\int_{\mathbb{R}^2} f_{X,Y}(x,y)\,dx\,dy = \int_0^1 \int_0^x c\,dy\,dx = \int_0^1 cx\,dx = \frac{cx^2}{2}\bigg|_0^1 = \frac{c}{2} = 1, \qquad \text{hence,} \qquad c = 2.$$

The random variables X and Y are not independent. One reason is that the joint support is triangular rather than rectangular: values with $0 < x < y < 1$ are impossible even though both marginal variables can take values in $[0,1]$.

(b) For $0 \le x \le 1$,

$$f_X(x) = \int_0^x 2\,dy = 2x.$$

Note that $f_X(x) = 0$ for x outside $[0,1]$. Thus,

$$f_X(x) = \begin{cases} 2x, & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

For $0 < x \le 1$ and $0 \le y \le x$,

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x}.$$

Therefore,

$$f_{Y|X}(y \mid x) = \begin{cases} 1/x, & 0 \le y \le x, \\ 0, & \text{otherwise.} \end{cases}$$

Hence $Y \mid X = x \sim \text{Uniform}(0, x)$, so

$$\mathbb{E}[Y \mid X = x] = \frac{x}{2}.$$

(c) Let

$$Z = X - Y.$$

For $0 \le z \le 1$, the event $Z = z$ corresponds to $x = y + z$. The allowed values satisfy

$$0 \le y \le y + z \le 1,$$

so

$$0 \le y \le 1 - z.$$

Thus,

$$f_Z(z) = \int_0^{1-z} 2\,dy = 2(1-z), \qquad 0 \le z \le 1.$$

Therefore,

$$f_Z(z) = \begin{cases} 2(1-z), & 0 \le z \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Problem 4. Dependence, total variance, and random sums

(a) The correlation coefficient is

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{-3}{\sqrt{9\cdot 4}} = -\frac{1}{2}.$$

Also,

$$\begin{aligned}\mathrm{Var}(2X+Y) &= 4\mathrm{Var}(X) + \mathrm{Var}(Y) + 4\mathrm{Cov}(X,Y)\\ &= 4(9) + 4 + 4(-3)\\ &= 36 + 4 - 12 = 28.\end{aligned}$$

(b) By total expectation,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Z]] = 0\cdot P(Z=0) + 4\cdot P(Z=1) = 4\cdot\frac{1}{4} = 1.$$

For variance, we use the law of total variance. First, observe that

$$\mathbb{E}[\mathrm{Var}(X \mid Z)] = 1\cdot\frac{3}{4} + 9\cdot\frac{1}{4} = \frac{3}{4} + \frac{9}{4} = 3.$$

Also,

$$\mathbb{E}[X \mid Z] = 4Z, \qquad \text{so} \qquad \mathrm{Var}(\mathbb{E}[X \mid Z]) = \mathrm{Var}(4Z) = 16\mathrm{Var}(Z).$$

Since $Z \sim \mathrm{Bernoulli}(1/4)$,

$$\mathrm{Var}(Z) = \frac{1}{4}\cdot\frac{3}{4} = \frac{3}{16}. \qquad \text{and thus,} \qquad \mathrm{Var}(\mathbb{E}[X \mid Z]) = 16\cdot\frac{3}{16} = 3.$$

Therefore, by the law of total variance:

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Z)] + \mathrm{Var}(\mathbb{E}[X \mid Z]) = 3 + 3 = 6.$$

(c) Here

$$\mathbb{E}[N] = 10\cdot\frac{1}{2} = 5, \qquad \mathrm{Var}(N) = 10\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{5}{2}.$$

Also, for $X_i \sim \mathrm{Bernoulli}(0.2)$,

$$\mathbb{E}[X_i] = 0.2, \qquad \mathrm{Var}(X_i) = 0.2(0.8) = 0.16.$$

The random-sum formulas give

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}[N]\mathbb{E}[X_i] = 5(0.2) = 1,\\ \mathrm{Var}(S) &= \mathbb{E}[N]\mathrm{Var}(X_i) + (\mathbb{E}[X_i])^2\mathrm{Var}(N) = 5(0.16) + (0.2)^2\left(\frac{5}{2}\right) = 0.8 + 0.1 = 0.9.\end{aligned}$$

Now use MGFs. Using the definition of MGF $M_X(t) = \mathbb{E}[e^{tX}]$, we have

$$M_N(t) = \left(\frac{1}{2} + \frac{1}{2}e^t\right)^{10}, \qquad M_X(t) = 0.8 + 0.2e^t.$$

Thus,

$$\begin{aligned}M_S(t) &= M_N(\log M_X(t))\\ &= \left(\frac{1}{2} + \frac{1}{2}M_X(t)\right)^{10}\\ &= \left(\frac{1}{2} + \frac{1}{2}(0.8 + 0.2e^t)\right)^{10}\\ &= (0.9 + 0.1e^t)^{10}.\end{aligned}$$

This is the MGF of a $\mathrm{Binomial}(10, 0.1)$ random variable. Therefore,

$$S \sim \mathrm{Binomial}(10, 0.1).$$

Problem 5. Deviation inequalities and the law of large numbers

(a) For $X \sim \text{Exponential}(1/2)$,

$$P(X \geq 6) = e^{-(1/2)6} = e^{-3} \approx 0.050.$$

Using Markov's inequality,

$$P(X \geq 6) \leq \frac{\mathbb{E}[X]}{6} = \frac{2}{6} = \frac{1}{3}.$$

Using Chebyshev's inequality, since $X \geq 6$ implies $|X - 2| \geq 4$,

$$P(X \geq 6) \leq P(|X - 2| \geq 4) \leq \frac{\text{Var}(X)}{4^2} = \frac{4}{16} = \frac{1}{4}.$$

(b) We have

$$\mathbb{E}[\bar{X}_n] = \mu, \qquad \text{Var}(\bar{X}_n) = \frac{9}{n}.$$

By Chebyshev,

$$P(|\bar{X}_n - \mu| \geq 0.2) \leq \frac{9/n}{(0.2)^2} = \frac{9}{0.04n} = \frac{225}{n}.$$

We want

$$\frac{225}{n} \leq 0.05.$$

Thus,

$$n \geq \frac{225}{0.05} = 4500.$$

So $n \geq 4500$ is sufficient.

(c) For Y_n, for any fixed $\epsilon > 0$, when $n^2 > \epsilon$,

$$P(|Y_n - 0| \geq \epsilon) = P(Y_n = n^2) = \frac{1}{n} \to 0.$$

Thus,

$$Y_n \xrightarrow{P} 0.$$

However,

$$\mathbb{E}[Y_n] = n^2 \cdot \frac{1}{n} = n,$$

so $\mathbb{E}[Y_n] \to \infty$, and $\mathbb{E}[Y_n]$ does not converge.

For Z_n, if $0 < \epsilon \leq 1$,

$$P(|Z_n - 0| \geq \epsilon) = P(Z_n = 1) = \frac{1}{2},$$

which does not go to 0. Thus Z_n does not converge in probability to 0. Its expectation is

$$\mathbb{E}[Z_n] = \frac{1}{2},$$

regardless of n, so $\mathbb{E}[Z_n] \to 1/2$.

Problem 6. Central limit theorem and normal approximation

(a) By the CLT,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}}$$

is approximately standard normal for large n. Therefore,

$$P(S_n \le c) \approx \Phi\left(\frac{c - n\mu}{\sigma\sqrt{n}}\right).$$

To have

$$P(S_n \le c) \approx 0.95,$$

we want

$$\frac{c - n\mu}{\sigma\sqrt{n}} \approx z_{0.95},$$

where $\Phi(z_{0.95}) = 0.95$. From the normal table, $z_{0.95} \approx 1.645$. Hence

$$c \approx n\mu + 1.645\,\sigma\sqrt{n}.$$

(b) For $X_i \sim \text{Exponential}(2)$,

$$\mu = \mathbb{E}[X_i] = \frac{1}{2}, \qquad \sigma^2 = \text{Var}(X_i) = \frac{1}{4}, \qquad \sigma = \frac{1}{2}.$$

For $n = 100$,

$$\text{SD}(\bar{X}_{100}) = \frac{\sigma}{\sqrt{100}} = \frac{1/2}{10} = 0.05.$$

Thus,

$$\begin{aligned} P(0.45 \le \bar{X}_{100} \le 0.55) &\approx P\left(\frac{0.45 - 0.5}{0.05} \le Z \le \frac{0.55 - 0.5}{0.05}\right) \\ &= P(-1 \le Z \le 1) \\ &= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1. \end{aligned}$$

Using $\Phi(1) \approx 0.8413$,

$$P(0.45 \le \bar{X}_{100} \le 0.55) \approx 2(0.8413) - 1 = 0.6826.$$

(c) For $S \sim \text{Binomial}(200, 0.5)$,

$$\mathbb{E}[S] = np = 100, \qquad \text{Var}(S) = np(1-p) = 200(0.5)(0.5) = 50.$$

Use

$$G \sim N(100, 50)$$

as the normal approximation. With continuity correction,

$$P(90 \le S \le 110) \approx P(89.5 \le G \le 110.5).$$

Therefore,

$$\begin{aligned} P(90 \le S \le 110) &\approx \Phi\left(\frac{110.5 - 100}{\sqrt{50}}\right) - \Phi\left(\frac{89.5 - 100}{\sqrt{50}}\right) \\ &= \Phi(1.48) - \Phi(-1.48) \\ &= 2\Phi(1.48) - 1. \end{aligned}$$

Using $\Phi(1.48) \approx 0.9306$,

$$P(90 \le S \le 110) \approx 2(0.9306) - 1 = 0.8612.$$

Using 1.49 instead gives approximately 0.8638, so a reasonable numerical answer is about

$$0.862.$$