50.034 – Introduction to Probability and Statistics

January–May Term, 2019

Homework Set 5

Due by: Week 6 Cohort Class (7 Mar 2019 or 8 Mar 2019)

Reminder: There is Mini-quiz 2 in Week 6 during your cohort class.

Question 1. Suppose an investor comes to you, wishing to invest $85,000 completely in a portfolio consisting of some combination of Stock A or Stock B, where Stock A costs $100 per share, and Stock B costs $10 per share. Assume that the number of shares for each stock must be a whole number. (So if the portfolio consists of a shares of Stock A and b shares of Stock B, then $100a + 10b = 85000$ for some non-negative integers a and b.)

Let X and Y be random variables representing the rates of return (in dollars) per share of Stock A and Stock B respectively, for a period of one year. From your research on these two stocks, you believe that X and Y have variances 16 and 4 respectively, while their correlation is $\rho(X, Y) = -0.5$. What should the investment portfolio be to minimize the variance of the overall return after one year?

Solution. Suppose the portfolio consists of a shares of Stock A and b shares of Stock B. This implies that $100a + 10b = 85\,000$, hence $b = 8500 - 10a$. Our goal is to find the value of a that minimizes $\mathrm{var}(aX + bY) = \mathrm{var}(aX + (8500 - 10a)Y)$.

Note that $\mathrm{cov}(X, Y) = \rho(X, Y)\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)} = -4$, hence

$$\begin{aligned}
\mathrm{var}(aX + bY) &= a^2\mathrm{var}(X) + b^2\mathrm{var}(Y) + 2ab\,\mathrm{cov}(X, Y)\\
&= 16a^2 + 4(8500 - 10a)^2 - 8a(8500 - 10a)\\
&= 496a^2 - 784\,000a + 289\,000\,000\\
&= 496\big(a - \tfrac{23\,375}{31}\big)^2 + \tfrac{216\,750\,000}{31},
\end{aligned}$$

which is minimized at $a = \frac{23375}{31} \approx 754.032$. Note that $8500 - 10(754) = 960$.
Therefore, the portfolio should have 754 shares of Stock A and 960 shares of Stock B to minimize the variance of the overall return.

Question 2. A new electronics shop has opened. As part of its promotional advertising, the owner has organized an instant lucky draw via a digital “spin-the-wheel”, where every customer who enters the lucky draw wins a $2 cash voucher with probability 0.8, wins a $5 cash voucher with probability 0.15, wins a $10 cash voucher with probability 0.049, and wins a $1000 cash voucher with probability 0.001. Customers can only enter this lucky draw if they are the first 500 customers, and as long as nobody has won a $1000 cash voucher yet. Once someone wins a $1000 cash voucher, the lucky draw ends immediately after that. If none of the first 500 customers has won a $1000 cash voucher, then the lucky draw ends. What is the expected total amount (in dollars) of cash vouchers given out for the lucky draw?

(Hint: What if you let customers spin the wheel “for fun”, even after the lucky draw has ended?)

Solution. To simplify the scenario, assume that every customer can participate in spinning the wheel, whether or not the lucky draw is ongoing, but each customer can win a cash voucher (based on the outcome of the wheel spin) only if the lucky draw is still ongoing. This assumption does not change the total amount of cash vouchers given out for the entire lucky draw.

First, we define the following random variables:

- Let X be the total amount (in dollars) of cash vouchers given out for the lucky draw.
- Let Y be the total number of customers immediately before the first wheel spin for a \$1000 cash voucher. (This wheel spin may not necessarily occur during the lucky draw.)

Let y be a non-negative integer. If $y \leq 499$, then $Y = y$ implies that the $(y+1)$-th customer wins a \$1000 cash voucher, and the first y customers has a $\frac{0.8}{0.999}$ probability to win a \$2 cash voucher, $\frac{0.15}{0.999}$ probability to win a \$5 cash voucher, and a $\frac{0.049}{0.999}$ probability to win a \$10 cash voucher. Hence

$$\mathbf{E}[X|Y=y] = \big(\tfrac{0.8}{0.999}(2) + \tfrac{0.15}{0.999}(5) + \tfrac{0.049}{0.999}(10)\big)y + 1000 = \tfrac{2840}{999}y + 1000$$

in this case ($y \leq 499$).
If instead $y \geq 500$, then $Y = y$ implies that none of the first 500 customers has won a \$1000 cash voucher. Hence

$$\mathbf{E}[X|Y=y] = \big(\tfrac{0.8}{0.999}(2) + \tfrac{0.15}{0.999}(5) + \tfrac{0.049}{0.999}(10)\big)(500) = \tfrac{1420000}{999}$$

in this case ($y \geq 500$).

Note that Y is a geometric random variable with parameter 0.001, so $\Pr(Y=y) = 0.001(0.999)^y$. By the law of total probability for expectations, $\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]]$, thus

$$\begin{aligned}
\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]] &= \sum_{y=0}^{\infty} \mathbf{E}[X|Y=y]\Pr(Y=y) \\
&= \left(\sum_{y=0}^{499} \big(\tfrac{2840}{999}y + 1000\big)(0.001)(0.999)^y\right) + (0.999^{500})\big(\tfrac{1420000}{999}\big) \\
&= \left(\sum_{y=0}^{499}(2.84)(0.001)y(0.999)^{y-1}\right) + \left(\sum_{y=0}^{499}(0.999)^y\right) + (0.999^{500})\big(\tfrac{1420000}{999}\big).
\end{aligned}$$

Recall that the formula for a geometric progression with $n+1$ terms, with ratio r, and with first term 1, is $\sum_{y=0}^{n} r^y = \frac{(1-r^{n+1})}{1-r}$. Treating both sides of this formula as functions of r, we can take their derivatives with respect to r to get $\sum_{y=0}^{n} yr^{y-1} = \frac{nr^{n+1}-(n+1)r^n+1}{(1-r)^2}$. Consequently,

$$\sum_{y=0}^{499} y(0.999)^{y-1} = \frac{499(0.999)^{500} - 500(0.999)^{499} + 1}{(0.001)^2}, \quad \text{and} \quad \sum_{y=0}^{499}(0.999)^y = \frac{1-0.999^{500}}{0.001},$$

which implies that

$$\begin{aligned}
\mathbf{E}[X] &= (2.84)\frac{499(0.999)^{500} - 500(0.999)^{499} + 1}{(0.001)} + \frac{1-0.999^{500}}{0.001} + (0.999^{500})\big(\tfrac{1420000}{999}\big) \\
&\approx 1511.50
\end{aligned}$$

Therefore, the expected total amount of cash vouchers given out is approximately \$1511.50.

Question 3. Let X and Y be independent random variables, and suppose they have moment generating functions given by:

$$\begin{aligned}
\psi_X(t) &= e^{t^2+t}, \quad \text{for } -\infty < t < \infty; \\
\psi_Y(t) &= e^{t^2-t}, \quad \text{for } -\infty < t < \infty;
\end{aligned}$$

respectively.

(i) What is the moment generating function of $2X + 3Y$?
(ii) What is the mean of $2X + 3Y$?
(iii) What is the variance of $2X + 3Y$?

Solution. (i) By definition, the moment generating function of $2X + 3Y$ is the function $\psi(t) = \mathbf{E}[e^{t(2X+3Y)}] = \mathbf{E}[e^{(2t)X}e^{(3t)Y}]$. Since X and Y are independent, it follows that

$$\mathbf{E}[e^{(2t)X}e^{(3t)Y}] = \mathbf{E}[e^{(2t)X}]\mathbf{E}[e^{(3t)Y}] = \psi_X(2t)\psi_Y(3t).$$

(See Week 5's cohort class slides or see Theorem 4.4.4 of course textbook.) Therefore,

$$\psi(t) = \psi_X(2t)\psi_Y(3t) = e^{4t^2+2t}e^{9t^2-3t} = e^{13t^2-t}, \qquad \text{for } -\infty < t < \infty.$$

(ii) Using $\psi(t) = e^{13t^2-t}$ from the previous part, we compute the first derivative

$$\psi'(t) = (26t-1)e^{13t^2-t},$$

thus $\mathbf{E}[2X + 3Y] = \psi'(0) = -1$.

(iii) Next, we compute the second derivative:

$$\psi''(t) = \tfrac{d}{dt}(26t-1)e^{13t^2-t} = (676t^2 - 52t + 27)e^{13t^2-t}.$$

This implies that $\mathbf{E}[(2X + 3Y)^2] = \psi''(0) = 27$. Therefore, the variance of $2X + 3Y$ is

$$\mathrm{var}(2X + 3Y) = \mathbf{E}[(2X + 3Y)^2] - (\mathbf{E}[2X + 3Y])^2 = 27 - (-1)^2 = 26.$$

Question 4. An automated juice vending machine dispenses orange juice, whose amount per cup follows a normal distribution with a mean of 250 ml and a standard deviation of 27 ml.

(i) What should the minimum cup size be (in ml), so that the orange juice dispensed would overflow at most 1% of the time?
(ii) If the cup size is fixed at 300 ml, what is the probability that the orange juice dispensed for one cup would overflow?

Solution. (i) Let X be the amount (in ml) of orange juice dispensed per cup. We are given that X has mean $\mu = 250$ and standard deviation $\sigma = 27$. Define the random variable $Z = \frac{X-\mu}{\sigma} = \frac{X-250}{27}$. Then Z has the standard normal distribution. Let c be the cup size (in ml). We want to find the minimum value for c such that $\Pr(X > c) = 0.01$, or equivalently, $\Pr(X \leq c) = 0.99$. In other words, we want

$$\Pr(X \leq c) = \Pr\big(Z \leq \frac{c-\mu}{\sigma}\big) = \Pr\big(Z \leq \frac{c-250}{27}\big) = \Phi(\tfrac{c-250}{27}) = 0.99,$$

where $\Phi(z)$ here denotes the standard normal cumulative distribution function.

By checking the standard normal distribution table in the course textbook, the closest value we can find for z satisfying $\Phi(z) = 0.99$ is $z = 2.33$. Thus, $\frac{c-250}{27} \approx 2.33$, which implies $c \approx 312.91$. Therefore, the cup should have a minimum size of approximately 312.91 ml, so that the orange juice would overflow at most 1% of the time.

(ii) Using the same random variable Z and the same function $\Phi(z)$ as defined in the previous part, we check that

$$\Pr(X > 300) = \Pr(Z > \frac{300-250}{27}) = 1-\Pr(Z \leq \tfrac{50}{27}) \approx 1-\Phi(1.85) \approx 1-0.9678 = 0.0322,$$

therefore the probability that the orange juice dispensed for one 300 ml cup would overflow is approximately 3.22%.

Question 5. A new animal species has recently been discovered in the heart of the Amazon rainforest. Let X be the random variable representing the weight (in grams) of this animal, and let Y be the random variable representing the length (in cm) of this animal. Suppose that X and Y have the bivariate normal distribution with means $\mu_X = 230$, $\mu_Y = 6.3$, variances $\sigma_X^2 = 3600$, $\sigma_Y^2 = 0.49$, and correlation $\rho = 0.65$.

(i) Suppose that a randomly selected animal of this species has length 7.0 cm. What is the probability that it has weight at least 280 g?
(ii) Suppose that a second randomly selected animal of this species has weight 200 g. What is its expected length?

(Hint: Did you read Section 5.10 of the course textbook?)

Solution. (i) By Theorem 5.10.4 of the course textbook, the conditional distribution of X given that $Y = 7.0$ is a normal distribution with mean

$$\mu_{X|Y=7.0} = \mu_X + \rho\sigma_X\Big(\frac{7.0-\mu_Y}{\sigma_Y}\Big) = 269,$$

and variance

$$\sigma^2_{X|Y=7.0} = (1-\rho^2)\sigma_X^2 = 2079.$$

Define the random variable $Z_1 = \frac{X-269}{\sqrt{2079}} = \frac{X-269}{\sqrt{2079}}$, and note that Z_1 has the standard normal distribution. Thus,

$$\begin{aligned}\Pr(X \geq 280|Y = 7.0) &= \Pr\left(Z_1 \geq \frac{280-269}{\sqrt{2079}}\right) = 1 - \Pr\left(Z_1 \leq \frac{11}{\sqrt{2079}}\right)\\ &\approx 1 - \Phi(0.24) \approx 1 - 0.5948 = 0.4052,\end{aligned}$$

where $\Phi(z)$ denotes the standard normal cumulative distribution function. Therefore, the probability that it has weight at least 280 g is approximately 0.4052.

(ii) By Theorem 5.10.4 of the course textbook, the conditional distribution of Y given that $X = 200$ is a normal distribution with mean

$$\begin{aligned}\mu_{Y|X=200} &= \mu_Y + \rho\sigma_Y\Big(\frac{200-\mu_X}{\sigma_X}\Big) = 6.3 + (0.65)(0.7)\Big(\frac{200-230}{60}\Big)\\ &= 6.0725,\end{aligned}$$

therefore its expected length is 6.0725 cm.