Recitation/Discussion 8

Midterm II Review

accompaniment to MATH 170E: Introduction to Probability and Statistics 1: Probability
at the University of California, Los Angeles

Allen Lin

In this recitation, we will review material such as

- exponential, Erlang, and normal distributions;
- conditional and joint distributions;
- covariance, conditional expectations, and conditional variance; and
- moment generating functions.

Exit Survey

Please fill out the exit survey.

This is so I can get feedback and know what to focus on next time!

Covariance

Last time, we talked about joint distributions of X and Y and discussed when X and Y are independent. Perhaps we want more *quantitative* ways of describing how X and Y are related. The covariance does exactly that: it measures the joint variability of two random variables by measuring the tendency in the linear relationship between the two variables.

Definition 1. The *covariance* of X and Y is

$$\operatorname{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Note that $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) > 0$ when X and Y are both above or below their means. Thus X and Y tend to move in the same direction if $\operatorname{Cov}(X, Y) > 0$.

By linearity of expectation, we can write

$$\begin{aligned}\operatorname{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - \mathbb{E}[Y]X - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

This is often esaier to compute.

From the definition of variance $\operatorname{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$, we see that the variance is a special case of the covariance when $Y = X$, so that $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$.

There is a normalized version of covariance known as the *correlation coefficient*. The normalizing factor here is $\sigma_X \sigma_Y$ (the product of the standard deviations), so that the $\operatorname{Cov}(X, Y)/\sigma_X \sigma_Y \in [-1, 1]$.

Definition 2. The *correlation coefficient* of X and Y, given $\sigma_X \neq 0$ and $\sigma_Y \neq 0$, is

$$\operatorname{Corr}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Definition 3. Let X and Y be random variables. Then

- X and Y are *positively correlated* iff $\operatorname{Cov}(X, Y) > 0$ iff $\operatorname{Corr}(X, Y) > 0$;
- X and Y are *negatively correlated* if $\operatorname{Cov}(X, Y) < 0$ iff $\operatorname{Corr}(X, Y) < 0$; and
- X and Y are *uncorrelated* if $\operatorname{Cov}(X, Y) = 0$ iff $\operatorname{Corr}(X, Y) = 0$.

In the same way that expected value is linear, covariance is bilinear.

Theorem 4. The covariance of two random variables is *bilinear* where

This is probably the most important property about covariance.

(a) $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$ (symmetric),

(b) $\operatorname{Cov}(aX, Y) = a \operatorname{Cov}(X, Y) = \operatorname{Cov}(X, aY)$ for any $a \in \mathbb{R}$,

(c) $\operatorname{Cov}(X_1 + X_2, Y) = \operatorname{Cov}(X_1, Y) + \operatorname{Cov}(X_2, Y)$, and

(d) $\operatorname{Cov}(X, Y_1 + Y_2) = \operatorname{Cov}(X_1, Y_1) + \operatorname{Cov}(X_1, Y_2)$.

In the special case of variance, we have the following.

Corollary 5. *For any $a \in \mathbb{R}$, $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$.*

The variance scales the factor into a square.

Exercise 1. Roll a pair of fair 6-sided die. Let X denote the smaller of the two rolls, and let Y denote the larger of the two rolls. Compute $\mathrm{Cov}(X, Y)$.

Exercise 2. Let $U = X + Y$ and $V = X + 2Y$. Calculate $\mathrm{Cov}(U, V)$.

You should remember this!

Recall that variance is not so quite linear in the sense that

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

is not *quite* true. But when is this true? We can use covariance and see that

$$\begin{aligned} \mathrm{Var}(X + Y) &= \mathrm{Cov}(X + Y, X + Y) \\ &= \mathrm{Cov}(X, X) + 2\,\mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, Y) \\ &= \mathrm{Var}(X) + 2\,\mathrm{Cov}(X, Y) + \mathrm{Var}(Y), \end{aligned}$$

so variance is *linear* when $\mathrm{Cov}(X, Y) = 0$, i.e., when X and Y are uncorrelated.

Theorem 6. If $\mathrm{Cov}(X, Y) = 0$, then $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Remember, *linear* things are nice! Computations are often simpler when things are linear.

Checking whether two variables are uncorrelated can be arduous. Luckily, there is an easier, more sufficient condition that we can check.

Theorem 7. If X and Y are independent, then X and Y are uncorrelated.

Sufficient, but not necessary.

Corollary 8. *If X and Y are independent, then* $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Note that the converse is not true: If X and Y are uncorrelated, that doesn't necessarily mean that X and Y are independent.

So remember: independence implies uncorrelated, but not conversely.

Exercise 3. Find two random variables X and Y such that X and Y are uncorrelated but not independent.

Exercise 4. If $X \sim \mathrm{Bin}(n, p)$, then $\mathrm{Var}[X] = np(1 - p)$.

Exercise 5. If $X \sim \mathrm{NB}(r, p)$, then $\mathrm{Var}(X) = r(1-p)/p^2$.

This is r times the variance of a geometric random variable.

Conditional Distributions

Now that we have joint distributions that describes the *joint* probability of two random variables, how can we describe conditional probability with two random variables? That is, if there is a prior on a random variable, how does that affect the posterior of the other random variable? We can define this in the natural way.

Theorem 9. If X and Y are two random variables, then the *conditional probability (mass or density) function of X given Y* is

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)}$$

Note that in the discrete case this definition matches exactly what we want.

and the *conditional probability (mass or density) function of Y given X* is

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

With this we have conditional expectation and conditional variance.

Definition 10. The *conditional expectation of X given Y* is

$$\mathbb{E}[X \mid Y = y] = \sum_x x f_{X|Y}(x \mid y)$$

in the discrete case and

$$\mathbb{E}[X \mid Y = y] = \int x f_{X|Y}(x \mid y)\, dx$$

in the continuous case. The *conditional expectation of Y given X* is defined analogously.

You should write this out to be sure.

Definition 11. The *conditional variance of X given Y* is

$$\mathrm{Var}(X \mid Y = y) = \mathbb{E}\left[X^2 \mid Y = y\right] - \mathbb{E}[X \mid Y = y]^2.$$

The *conditional variance of Y given X* is defined analogously.

There are two fundamental results involving conditional expectation and conditional variance.

Theorem 12 (Law of total expectation)**.** For two random variables X and Y, we have

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

The term $\mathbb{E}[X \mid Y]$ is a random variable depending on $Y = y$. You should prove this.

Theorem 13 (Law of total variance)**.** For two random variables X and Y,

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

Again, $\mathrm{Var}(X \mid Y)$ is a random variable depending on $Y = y$. You should prove this.

There is a law of total covariance. Can you guess what that is?

Moment Generating Functions

The moment generating function of a random variable X is a powerful way of encoding all the moments of X. Recall that the nth moment of X is $\mathbb{E}[X^n]$. To encode all the moments of X we need a function that has all the powers of X. The natural function that comes to mind from calculus is the exponential function

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

with its Taylor series expansion. In particular, we can write

$$\mathbb{E}\left[e^{tX}\right] = 1 + t\mathbb{E}[X] + \frac{t^2\mathbb{E}\left[X^2\right]}{2!} + \frac{t^3\mathbb{E}\left[X^3\right]}{3!} + \cdots$$

by applying linearity of expectation. Observe that $\mathbb{E}\left[e^{tX}\right]$ now has all the moments encoded in it (up to some multiple of t). Define

$$M_X(t) = \mathbb{E}\left[e^{tX}\right]$$

for the moment generating function of X. Now to obtain each moment separately, we can take iterated derivatives and then take the value when $t = 0$. To see this more clearly, observe that

$$M_X(0) = 1$$

is the constant term. If we want to obtain $\mathbb{E}[X]$, we take one derivative and get

$$M_X^{(1)}(t) = \mathbb{E}[X] + \frac{2t\mathbb{E}\left[X^2\right]}{2!} + \frac{3t^2\mathbb{E}\left[X^3\right]}{3!} + \cdots.$$

Now $\mathbb{E}[X]$ is the constant term, so

$$M_X^{(1)}(0) = \mathbb{E}[X].$$

What if we want $\mathbb{E}\left[X^2\right]$? We can take the derivative once more and get

$$M_X^{(2)}(t) = \frac{2\mathbb{E}\left[X^2\right]}{2!} + \frac{3 \cdot 2t\mathbb{E}\left[X^3\right]}{3!} + \cdots$$

so $2\mathbb{E}\left[X^2\right]/2! = \mathbb{E}\left[X^2\right]$ is the constant term. Thus

$$M_X^{(2)}(0) = \mathbb{E}\left[X^2\right].$$

In general, to obtain the nth momemnt we can take n derivatives. Specifically,

$$M_X^{(n)}(0) = \mathbb{E}[X^n]$$

so to calculate the 2026th moment, in principle we can take the 2026th derivative.

That's why the moment generating function is so powerful. Rather than calculating each moment by summing or integrating, we can just calculate the derivative of the mgf, which is often easier and faster.

Practice Problems

Exercise 6. Customers arrive at the local movie theatre for *The Odyssey* according to a Poisson process with rate $\lambda = 3$ customers per minute. Let T_n denote the time until the nth customer arrives.

(a) Find the probability that the third customer arrives within the first minute.

(b) Find the probability that the fourth customer arrives between the first and third minute.

(c) Calculate $\mathbb{E}[T_4]$ and $\text{Var}(T_4)$.

(d) Given that exactly two customers arrive within the first thirty seconds, what is the probability that the fourth customer arrives within the first minute?

Exercise 7. Two 6-sided dice numbered 1, 2, 3, 4, 5, and 6 are rolled independently. Both dice are weighted so that the probability of rolling some $i \in \{1, \ldots, 6\}$ on each die is $i/21$. Let X and Y denote the outcome of the bigger and smaller die roll, respectively.

(a) Find the joint probability mass function of X and Y.

(b) Are X and Y independent? Justify your answer.

(c) A player's prize is calculated by taking four times the larger outcome and deducting six times the smaller outcome. What are the expected winnings after 441 rolls?

Hint. Use the helpful formulas

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2},$$
$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6},$$
$$\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4},$$
$$\sum_{i=1}^{n} i^4 = \frac{n(n+1)(2n+1)\left(3n^2+3n-1\right)}{30}.$$

Exercise 8. The weight in grams of a randomly chocolate bar from a factory follows a normal distribution with mean 100 grams and standard deviation 10 grams.

(a) Calculate the probability that a randomly chocolate bar weighs more than 115 grams.

(b) Calculate the probability that a randomly chocolate bar weighs between 90 and 110 grams.

(c) Find a such that 95% of all chocolate bars weigh less than a grams.

(d) Given that a chocolate bar weighs more than 105 grams, calculate the probability that it weighs more than 115 grams.

(e) A chocolate bar is *defective* if its weight differs by more than 15 grams from its mean. Calculate the probability that a randomly chocolate bar is defective.

(f) Suppose 5 chocolate bars are selected independently. Find the probability that their average weight is more than 105 grams.

Hint. If U and V are independent random variables, then $\mathrm{Var}(c(U+V)) = c^2(\mathrm{Var}(U) + \mathrm{Var}(V))$.

Exercise 9. Let X and Y be independent random variables with

$$X \sim \text{Pois}(3) \quad \text{and} \quad Y \sim \text{Pois}(5).$$

(a) Find the moment generating functions $M_X(t)$ and $M_Y(t)$ of X and Y, respectively.

(b) Let $Z = X + Y$. Find the moment generating function $M_Z(t)$ of Z. Use this to identify the distribution of Z.

Hint. If U and V are independent random variables, then $\mathbb{E}[UV] = \mathbb{E}[U]\mathbb{E}[V]$.

(c) Calculate $\mathbb{E}[Z]$ and $\text{Var}(Z)$.

(d) Calculate $\mathbb{E}\left[Z^3\right]$.

Exercise 10. Suppose X has moment generating function

$$M_X(t) = \frac{1}{1-4t} \quad \text{for } t < \frac{1}{4}.$$

(a) Identify the distribution of X.

(b) Calculate $\mathrm{Var}(X)$.

(c) Prove that

$$\sum_{k=0}^{n} \frac{(-1/4)^k}{k!} \mathbb{E}\left[X^k\right] = \begin{cases} 1 & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd.} \end{cases}$$

Hint. Calculate the derivatives of M_X. Can you identify a general formula for the nth derivative of M_X?

(d) Let $Y = 170X$. Calculate $\sqrt[2026]{\mathbb{E}\left[Y^{2026}/2026!\right]}$.

Hint. Find $M_Y(t)$, the moment generating function of Y.

Exercise 11. As an avid water scroll collector, Katara is bartering for ancient water scrolls. The barterer has scrolls in box A and box B. Box A contains 4 water scrolls, 3 fire scrolls, and 2 earth scrolls. Box B contains 8 water scrolls and 2 air scrolls. The barterer asks Katara to first select a box and then draw a scroll five times. The barterer replaces the scrolls in each box so that the number of scrolls of each element type are fixed, i.e., if Katara picks a water scroll, then the barterer will replace the picked scroll with another water scroll. Assume that Katara favors box A with $2/3$ probability and favors box B with $1/3$ probability.

(a) What is the expected number of water scrolls that Katara will pick?

(b) What is the variance in the number of water scrolls that Katara will pick?

Exercise 12. Let X be a continuous random variable with pdf

$$f(x) = \begin{cases} x^2 e^{-x/2}/16 & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the moment generating function of X on the interval on which it exists, and use that to calculate $\mathbb{E}\left[X^2\right]$.

(b) Calculate $\mathbb{E}\left[X^{10}\right]$.

(c) Calculate $\text{Cov}\left(X^9, X\right)$.

(d) Let $Y = aX + b$. Find a and b such that $\mathbb{E}[Y] = 13$ and $\text{Var}(Y) = 1$.

Exercise 13. A store opens at 9 am, selling the long-awaited game *Silksong*. The arrival times (in minutes after opening) of Ari and William are uniformly distributed on $[0, 2]$. Let X be the arrival time of Ari, and let Y be the arrival time of William.

(a) Prove that X and Y are independent using joint and marginal distributions.

(b) What are the expected arrival times of both customers?

(c) Calculate $\text{Cov}(X, Y)$.

Exercise 14. A point is selected uniformly at random from the region

$$R := \left\{(x, y) \in \mathbb{R}^2 \;\middle|\; 0 \leq x \leq 2 \text{ and } x^2 \leq y \leq 4\right\}.$$

Let X and Y be the x- and y-coordinates of the point, respectively.

(a) Calculate the joint probability density function fo X and Y.

(b) Are X and Y independent? Justify your answer.

(c) Calculate $\mathrm{Cov}(X, Y)$.

(d) Let $U = 4X + 2Y$ and $V = 4X - 5Y$. Calculate $\mathrm{Cov}(U, V)$.

Exercise 15. A company manufactures two components, known as type I and type II systems, for an electronic device. Let X denote the number of defective type I systems, and let Y denote the number of defective type II systems. The joint probability mass function of X and Y is given by

$$f(x,y) = \begin{cases} c(x+y+1) & \text{if } x,y \in \{0,1,2\}, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Determine the value of c.

(b) Are X and Y independent?

(c) Suppose the inspector reports that the batch contains exactly one defective type I component. What is the conditioned expected number of defective type II components? What is the conditional variance of the number of defective type II components?

(d) A batch is rejected if the total number of defective components is at least three. Determine the probability that a randomly selected batch is rejected.