Lecture 8: Miscellaneous Topics on Probability

- Joint/Bivariate distributions
- Marginal and conditional distributions
- Conditional expectations
- Covariance and correlation
- Functions of a random variable

Introduction

– We have already talked about random variables, their distributions, density (or mass) functions and expectations

– Today we will extend those concepts to two or more variables jointly. We will cover joint distributions for bivariate distributions as well as conditional densities, covariance and correlation, conditional expectation

– Lastly, we will introduce how to find the distribution of some functions of random variable

Joint/Bivariate Distributions — Discrete R.V.s

– Let X_1, X_2 be discrete random variables with support $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. The **joint PMF** of (X_1, X_2) is a function that maps $x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2$ to $P(X_1 = x_1, X_2 = x_2)$ (the comma means "and")

$$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$$

– If X_1 and X_2 are **independent**, the joint PMF will be the product of the PMFs of X_1 and X_2. Given a joint PMF, we can find **conditional PMF** as

$$f(x_1|x_2) = P(X_1 = x_1|X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)}$$

The **marginal PMF** of X_1 (or X_2) can be found by adding up in the joint PMF:

$$f(x_1) = \sum_{k \in \mathcal{X}_2} f(x_1, k), \quad f(x_2) = \sum_{k \in \mathcal{X}_1} f(x_2, k)$$

Example: Party and Policy

Suppose we are in an imaginary country with two parties, which we will represent as $D = 0$ and $D = 1$, and all individuals in the country are either in favor of a certain policy $(Q = 1)$ or not $(Q = 0)$. The joint PMF is given by the table

	$D = 0$	$D = 1$
$Q = 0$	0.1	0.4
$Q = 1$	0.3	0.2

For instance, the probability that somebody supports party 0 and doesn't support the policy is $P(D = 0, Q = 0) = 0.1$.
The conditional PMF of Q given that $D = 0$ is

$$P(Q = 0|D = 0) = \frac{P(Q = 0, D = 0)}{P(D = 0)} = \frac{0.1}{0.4} = \frac{1}{4}$$

$$P(Q = 1|D = 0) = \frac{P(Q = 1, D = 0)}{P(D = 0)} = \frac{0.3}{0.4} = \frac{3}{4}$$

Example: Party and Policy (Cont'd)

The marginal PMF of D is

$$P(D = 0) = P(D = 0, Q = 0) + P(D = 0, Q = 1) = 0.4$$
$$P(D = 1) = P(D = 1, Q = 0) + P(D = 1, Q = 1) = 0.6$$

Q and D are dependent (the probability of being in favor of the policy depends on the party one supports). Indeed,

$$P(Q = 1|D = 0) = 3/4$$
$$P(Q = 1|D = 1) = \frac{P(Q = 1, D = 1)}{P(D = 1)} = 0.2/0.6 = 1/3$$

Joint/Bivariate Distributions — Continuous R.V.s

Let X_1 and X_2 be continuous random variables with support $\mathcal{X}_1$ and $\mathcal{X}_2$. The **joint PDF** of (X_1, X_2) is a nonnegative function f such that

$$P((X_1, X_2) \in A) = \int_A f(x_1, x_2) d(x_1, x_2)$$

for $A \subset \mathcal{X}_1 \times \mathcal{X}_2$. The **conditional PDF** of X_1 given $X_2 = x_2$ is

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)},$$

and the **marginal PDF** of X_1 is

$$f_1(x_1) = \int_{\mathcal{X}_2} f(x_1, x_2) dx_2$$

Joint PDF and CDF

First,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$$

where $f(x,y)$ is the non-negative joint **probability density function (pdf)**. That is, in order to be a joint density function, $f(x,y)$ must integrate to 1. So that,

$$\mathbb{P}[a \le X \le b \text{ and } c \le Y \le d] = \int_{c}^{d}\int_{a}^{b} f(x,y)\,dxdy$$

Also, the joint **cumulative distribution function (cdf)** is:

$$F(x,y) = \mathbb{P}[X \le x \text{ and } Y \le y] = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)\,dxdy$$

Example: Share Price of Two Stocks

Example 1: You compare the share price of two stocks, A and B. The share values of stock A (denoted X) cannot exceed 6.00, and the share values of stock B (denoted Y) cannot exceed 9.00.

Random market forces cause the values to vary. Assume that on a given day, the share prices in X and Y have joint probability density function

$$f(x,y) = \begin{cases} \frac{1}{4374}xy^2 & \text{for } 0 \le x \le 6, 0 \le y \le 9 \\ 0 & \text{otherwise.} \end{cases}$$

Qualitatively, this joint density function puts less probability on extremely low values, which might be reasonable. It also suggests that the probability of large values for B increases faster than the probability of large values for A, which might also be plausible.

Example: Share Price of Two Stocks (Cont'd)

Is the function in our stock value example a joint pdf? Yes, since:

$$\begin{aligned}\int_0^9 \int_0^6 \frac{1}{4374} xy^2 \, dxdy &= \frac{1}{4374} \int_0^9 18y^2 \, dy \\ &= \frac{1}{4374}(18 * 243) = 1.\end{aligned}$$

The marginals give us the density for just the X random variable or just the Y random variable, ignoring the other. If we just wanted to know the marginal density function for the share value of an A stock, it would be

$$\begin{aligned}f_1(x) &= \int_{-\infty}^{\infty} f(x,y) \, dy && \text{for } -\infty < x < \infty \\ &= \frac{1}{4374} \int_0^9 xy^2 \, dy && \text{for } 0 \le x \le 6 \\ &= \frac{1}{18} x && \text{for } 0 \le x \le 6.\end{aligned}$$

Example: Share Price of Two Stocks (Cont'd)

Similarly, if I wanted to know just the probability density function for share price of stock B, a similar integration would give

$$\begin{aligned} f_2(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx && \text{for } -\infty < x < \infty \\ &= \frac{1}{4374}\int_0^6 xy^2\,dx && \text{for } 0 \le y \le 9 \\ &= \frac{1}{243}y^2 && \text{for } 0 \le y \le 9. \end{aligned}$$

Two random variables are **independent** if and only if

$$f(x,y) = f_1(x) * f_2(y).$$

Independence and Conditional Densities

In our example, are the the two share prices independent?

Yes, since $(1/243)y^2 * (1/18)x = (1/4374)xy^2$**.**

What is the conditional density for a share of stock A value if a share of B costs \$5.00?

$$g_1(x|y=5) = \frac{f(x,5)}{f_2(5)} = \frac{\left(\frac{5^2}{4374}x\right)}{\left(\frac{5^2}{243}\right)} = \frac{1}{18}x \qquad \text{for } 0 \le x \le 6.$$

Since the random variables are independent, knowing the value of B does not change our probability for the value of A.

Note: *As usual, we use X and Y to denote random variables, and x and y to denote values they may take. Above, we observed that the outcome for Y was \$5.00 and sought the corresponding density for X.*

Conditional Expectation

Let X and Y be random variables such that the mean of Y exists and is finite. The **conditional expectation** of Y given $X = x$ is denoted by $E(Y|x)$ and is defined to be the expectation of the conditional distribution of Y given $X = x$.

– For example, if Y has a continuous conditional distribution given $X = x$ with conditional pdf $g_2(y|x)$, then

$$E(Y|x) = \int_{-\infty}^{\infty} y g_2(y|x) dy$$

– Similarly, if Y has a discrete conditional distribution given $X = x$ with conditional pmf $g_2(y|x)$, then

$$E(Y|x) = \sum_{\text{All } Y} y g_2(y|x)$$

– **Conditional mean of Y given X is a random variable!**

Covariance

The expectation of a function $h(X, Y)$ is

$$\mathbb{E}[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\, dxdy.$$

For example, the expected value of the product of X and Y is

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\, dxdy.$$

A particularly useful function is $h(X,Y) = (X - \mu_X)(Y - \mu_Y)$. Its expectation is called the **covariance**. The covariance between two random variables is a measure of joint variability.

– **Properties:**

1. $\mathrm{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$
2. If X_1 and X_2 are independent, $\mathrm{Cov}(X_1, X_2) = 0$. Warning: the reverse is not true!
3. $\mathrm{Cov}(X_1, X_1) = V(X_1)$

Correlation

When there are more than two random variables, e.g., there are X, Y, and Z, then one can calculate the covariance matrix whose entries are all possible pairs of covariances.

The covariance is important because it allows us to calculate the **correlation** between X and Y:

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X * \sigma_Y}.$$

Recall that σ_X is the standard deviation of X, or the square-root of $V(X) = \mathbb{E}[X^2] - \mu_X^2$.

Correlation is a measure of strength of association/linear dependence between two random variables. The advantage of correlation is that it is always between -1 and 1, so it is easier to interpret than covariance (which depends on the unit of the variables)

Properties of Correlation

Some useful facts about the correlation and covariance:

- If X and Y are independent, then $\text{Corr}(X, Y) = 0$.
- The converse fails: if $\text{Corr}(X, Y) = 0$, then X and Y may be dependent.
- $-1 \leq \text{Corr}(X, Y) \leq 1$.
- $\text{Cov}(aX, b) = 0$.
- $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$.
- $\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$.
- The $\text{Corr}(X, Y) = \pm 1$ if and only if $Y = aX + b$ for some $a \neq 0$.

The book covers this material in Chapter 4, and treats both continuous and discrete distributions. As we have seen before, that simply means that summations replace integrals in the formulae shown here.

Example: Party and Policy (Cont'd)

Suppose we are in an imaginary country with two parties, which we will represent as $D = 0$ and $D = 1$, and all individuals in the country are either in favor of a certain policy $(Q = 1)$ or not $(Q = 0)$. The joint PMF is given by the table

	$D = 0$	$D = 1$
$Q = 0$	0.1	0.4
$Q = 1$	0.3	0.2

The covariance between Q and D is

$$\mathrm{Cov}(Q, D) = E(QD) - E(Q)E(D),$$

where

$$P(QD = 0) = P(Q = 0, D = 0) + P(Q = 0, D = 1) + P(Q = 1, D = 0) = 0.8$$
$$P(QD = 1) = 1 - P(QD = 0) = 0.2$$

So $E(QD) = 0.2$, and $P(Q = 0) = 0.5$, $P(Q = 1) = 0.5$, $P(D = 0) = 0.4$, $P(D = 1) = 0.6$, from which we can find $E(Q)E(D) = 0.5 \times 0.6 = 0.3$.

Example: Party and Policy (Cont'd)

Therefore,

$$\text{Cov}(Q, D) = E(QD) - E(Q)E(D) = 0.2 - 0.3 = -0.1,$$

and

$$V(Q) = E(Q^2) - E(Q)^2 = 0.5 - 0.5^2 = 0.25$$
$$V(D) = E(D^2) - E(D)^2 = 0.6 - 0.6^2 = 0.24$$

Finally,

$$\text{Corr}(Q, D) = \frac{\text{Cov}(Q, D)}{\sqrt{V(Q)V(D)}} \approx -0.408$$

Motivation

– Let X be a continuous random variable, and $Y = h(X)$, a function of X. We have already talked about how to find the expectation of Y from previous lectures. What if we want more than that?

– For simplicity, we will assume that $h(X)$ is a well-behaved function of X. That is, h is continuous, monotone and its inverse h^{-1} exists.

– Notice that Y must also be a random variable that has its own distribution since it is a transformation of a random variable. Can we derive that?

Illustration

Example 2: Let X be a continuous random variable with pdf $f_X(x) = 4x^3, 0 < x < 1$ and let $Y = X^2$. What is the pdf of Y?

– Sometimes it makes more sense (and can be relatively straightforward) to first find the cdf of Y and then derive the pdf using calculus and what we have learned so far.

– Back to our example,

– What the cdf of X is?

Illustration

– Then, $F_Y(y) = P(Y \leq y) = P(X \leq \sqrt[+]{y}) = F_X(\sqrt[+]{y}) = (\sqrt[+]{y})^4 = y^2$

Do you know why $X \nleq \sqrt[-]{y}$?

– Now that we know what the cdf of Y is, can you figure out what the pdf of Y is?

– What is the support for y?

Another Illustration

Example 3: Let $X_1, X_2, \cdots, X_n$ be a random independent and identically distributed (iid) sample from the distribution with pdf $f(x) = 3x^2$ on $0 \leq x \leq 1$

Now, Let $Y = max\{X_1, X_2, \cdots, X_n\}$. Can we find E($Y$)?

First, can we figure out what the cdf of X is?

(Think about this: what does being the maximum really mean? If the maximum of X random variables is less than y, what is the relationship between each of the X's and y?)

Also, what is the pdf of Y?

Another Illustration (Cont'd)

As in the previous example, first find the cdf $F_X(x)$ of X.

Thus, $F(y) = P(Y \leq y) = P(X_1 \leq y, X_2 \leq y, \cdots, X_n \leq y)$

$\Rightarrow P(Y \leq y) = P(X_1 \leq y \; and \; X_2 \leq y \; and \; \cdots, and \; X_n \leq y)$

$\Rightarrow P(Y \leq y) = P(X_1 \leq y) \times P(X_2 \leq y) \times \cdots \times P(X_n \leq y)$

Can you see why this is the case?

$\Rightarrow F_Y(y) = P(Y \leq y) = F_{X1}(y) \times F_{X2}(y) \times \cdots \times F_{Xn}(y)$

$\Rightarrow F_Y(y) = y^3 \times y^3 \times \cdots \times y^3 = y^{3n}$

Another Illustration (Cont'd)

$\Rightarrow f_Y(y) = (3n)y^{3n-1}$ on $0 \le y \le 1$

Do you know why that is the support of Y?

Now,

$$E(Y) = \int_0^1 y(3n)y^{3n-1}dy = 3n \int_0^1 y^{3n}dy$$

Therefore, $E(Y) = \dfrac{3n}{3n+1}$

How about $Z = min\{X_1, X_2, \cdots, X_n\}$, can you find $E(Z)$?

Shortcut to the cdf technique

If the function $h(X)$ is a continuous function with an inverse (which we already assumed), there is a neater and faster way to find its pdf.

Let $Y = h(X)$, then if the inverse exists, $X = h^{-1}(Y)$.

Also, the derivative of X with respect to Y is $\dfrac{dX}{dY}$

Thus, the pdf of Y, $f_Y(y) = f_X[h^{-1}(Y)] \times \left|\dfrac{dX}{dY}\right|$

Example 2 again

$f_X(x) = 4x^3$ on $0 \leq x \leq 1$ and $Y = X^2$.

Inverse of $Y = X^2$ is $X = h^{-1}(Y) = \sqrt[+]{y}$ and $\dfrac{dX}{dY} = \dfrac{1}{2\sqrt[+]{y}}$

Then, $f_X[h^{-1}(Y)] = f_X[\sqrt[+]{y}] = 4(\sqrt[+]{y})^3 = 4y(\sqrt[+]{y})$

Therefore, $f_Y(y) = \dfrac{1}{2\sqrt[+]{y}} \times 4y(\sqrt[+]{y}) = 2y$

Do we have the same answer as before?

Another Example

Example 3 (D.S. Chapter 3.8 Exercises Question 4 – to be done in class): Suppose that the pdf of a random variable X is $\frac{x}{2}$ for $0 \leq x \leq 2$. What is the cdf of $Y = 4 - X^3$? We'll try both methods!

Answer:

$$f_Y(y) = \frac{1}{6}(4-y)^{-\frac{1}{3}}, \quad -4 \leq y \leq 4$$

Recap

We talked about

- Joint, marginal and conditional distributions
- Covariance and correlation
- Derive the marginal distribution of functions of a continuous random variable using the CDF method
- Use the change of variable technique as a shortcut to the CDF method

Suggested reading:

- D.S. Sec. 3.4, 3.5, 3.6, 3.8, 4.6, 4.7