PSY 503: Foundations of Psychological Methods
Lecture 10: Joint Distributions of Random Variables

Robin Gomila

Princeton

September 30, 2020

Joint distributions

- We introduced the concept of random variable
- We saw how to use the probability mass function (PDF) and the cumulative distribution function (CDF) to compute the probability that a discrete random variable Y takes some given set of values $y_1, \ y_2, \ \cdots y_n$
- We introduced parametric distributions
- We learned how to summarize random variables
- So far, we have treated each rv **in isolation**. Today, we focus on the study of relationships between random variables.

Joint Probability Distribution: Relevance

- Suppose that you are the Secretary of Defense of your country. You have to decide on the total size of the army for the coming year
- Your main concern is the possibility that a military conflict starts either in Country A or in Country B
  - If a conflict starts in Country A, you will need to send 10,000 soldiers to Country A
  - If a conflict starts in Country B, you will need to send 10,000 soldiers to Country B
- One objective: have the smallest possible army subject to the constraint that, no matter what happens, there is no shortage of soldiers

Joint Probability Distribution: Relevance

- Let X_A be a Bernoulli rv equals to 1 if a conflict starts in Country A (and 0 otherwise). Let X_B be a Bernoulli rv equals to 1 if a conflict starts in Country B (and 0 otherwise).
- Assume

$$X_A \sim Bern(0.5)$$

$$X_B \sim Bern(0.5)$$

- Do you have enough information to make an efficient decision about how many many soldiers to enroll?

Joint Probability Distribution: Relevance

- No. As the Secretary of Defense, you only care about the probability that there is a conflict both in Country A and in Country B
- If the probability that there is a conflict both in Country A and in Country B is zero, you should decide to have an army of only 10,000 soldiers
- If this probability is positive, then you should plan to have an army of 20,000 soldiers
- In this setting, the separate PMFs of X_A and X_B do not provide enough information
  - No information on the degree of association between these two random variables
  - No way to compute $P(\{X_A = 1\} \cap \{P(X_B = 1\})$

Joint Probability Mass Function

Given a pair of discrete rvs X_A and X_B, the **joint probability mass function** describes the probability that X_A takes value x_A and X_B takes value x_B, for any possible value of (x_A, x_B)

$$\begin{aligned} p(x_A,\, x_B) &= P(\{X_A = x_A\} \cap \{P(X_B = x_B\}) \\ &= P(\{X_A = x_A\},\, \{P(X_B = x_B\}) \end{aligned} \tag{1}$$

Joint Probability Distribution: Relevance

In our example, one possible joint probability distribution of X_A and X_B is

	$X_B = 0$	$X_B = 1$
$X_A = 0$	0	0.5
$X_A = 1$	0.5	0

- In this configuration, you should enroll exactly 10,000 soldiers
- Probability that both countries suffer a military conflict simultaneously is 0
- Probability that none of them suffers a military conflict simultaneously is 0

Joint Probability Distribution: Relevance

- What else could have happened?

	$X_B = 0$	$X_B = 1$
$X_A = 0$	0.5	0
$X_A = 1$	0	0.5

	$X_B = 0$	$X_B = 1$
$X_A = 0$	0.25	0.25
$X_A = 1$	0.25	0.25

- In all 3 joint probability distributions above, it is true that $P(X_A = 1) = 0.5$ and $P(X_B = 1) = 0.5$. The only feature that changes across them is the dependence or association between the random variables X_A and X_B

Properties of Joint PMFs

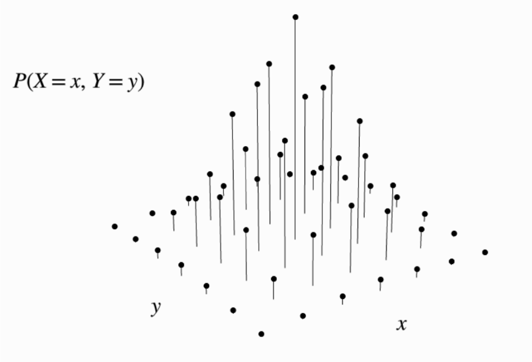
Just like univariate PMFs, joint PMFs are nonnegative and sum to 1. For two random variables X and Y:

$$\sum_x \sum_y P(X = x, Y = y) = 1$$

From joint PMF to marginal PMFs

- We can use the joint distribution of X and Y to get the univariate PMFs of X and Y
- To get the univariate PMF of X we simply sum over the possible values of Y
- This univariate PMF is called the **marginal distribution** of X, defined as
$$P(X = x) = \sum_y P(X = x, Y = y)$$
- Keep in mind that there is no way to derive the joint distribution of X and Y from the (separate) univariate distributions of X and Y

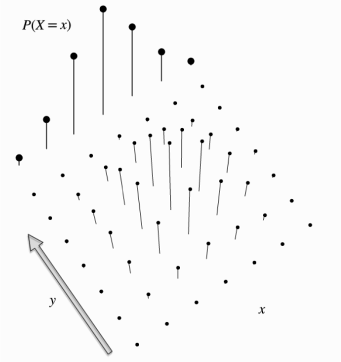
From joint PMF to marginal PMFs



- The total height of vertical bars is:

$$\sum_x \sum_y P(X = x, Y = y) = 1$$

From joint PMF to marginal PMFs



- For any value of x the probability $P(X = x)$ is the total height of the bars in the corresponding column of the joint PMF

$$P(X = x) = \sum_{y} P(X = x, Y = y)$$

Conditional PMF

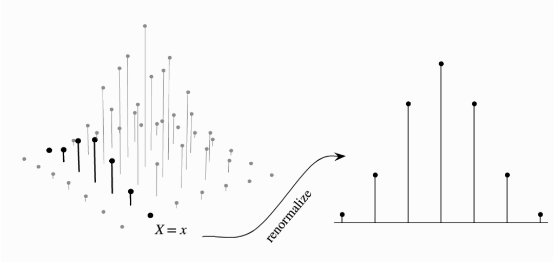
- The conditional PMF of Y given X is

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

- This should be understood as a function of Y for fixed x
- When it is most convenient, we can relate the conditional distribution of Y given $X = x$ to that of X given $Y = y$, using Baye's rule:

$$P(Y = y|X = x) = \frac{P(X = x|Y = y)P(Y = y)}{P(X = x)}$$

Conditional PMF



- We first take the joint PMF and focus on the vertical bars where X takes on the value x (i.e., the conditional PMF). All the other vertical bars are inconsistent with the knowledge that $X = x$ occurred
- Then we renormalize the conditional PMF by dividing by $P(X = x)$, which ensures that the conditional PMF will sum to 1

Independence of random variables

- Probability theory: two events A and B are independent if $P(A|B) = P(A)$
  - Remember $P(A \cap B) = P(A)P(B)$
- Analogously, two random variables Z and Y are independent if:
  - Z tells us nothing about Y. Y tells us nothing about Z.
- This is very important in experimental research!
  - Independence of treatment and background characteristics

Properties of independent random variables

$$P(Z = z, Y = y) = P(Z = z)P(Y = y)$$

$$\mathbb{E}[ZY] = \mathbb{E}[Z]\mathbb{E}[Y]$$

Quantifying the Dependence Between Random Variables

Relationship between PMFs

- Joint PMFs often contain too much information
- We often want to summarize joint distributions
  - Strength of their relationship, i.e., how dependent they are from each other

Covariance

Let X and Y be two random variables, the covariance between X and Y is defined as

$$\mathbb{C}(X, Y) = \mathbb{E}\Big[(X - E[X])(Y - E[Y])\Big]$$

Note that this is equivalent to:

$$\mathbb{C}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Properties of Covariance

$$\mathbb{C}[X, X] = \mathbb{V}[X]$$

$$\mathbb{C}[X, Y] = \mathbb{C}[Y, X]$$

$$\mathbb{C}[X + a, Y + b] = \mathbb{C}[X, Y]$$

$$\mathbb{C}[aX, bY] = ab\, \mathbb{C}[X, Y]$$

Limitations of covariance

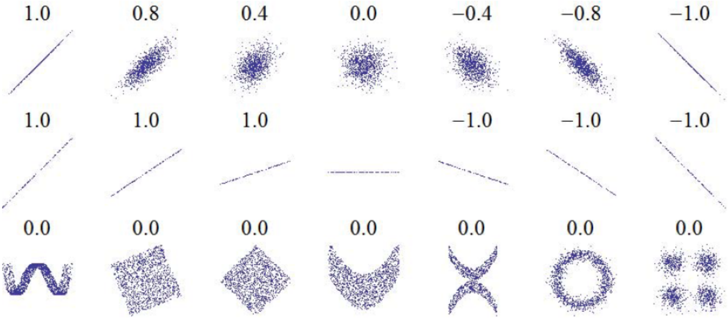
- Magnitude of $\mathbb{C}[X, Y]$ depends on the scales of X and Y
  - Difficult use covariance to interpret the strength of the relationship between two variables
- Use correlation

Correlation

$$\rho(X, Y) = \frac{\mathbb{C}[X, Y]}{\sigma_X \sigma_Y}$$

- Bounds: $-1 \leq \rho_{XY} \leq 1$

Correlation is linear



Correlation given independence

- Does knowing that X and Y are two independent random variables inform us about their covariance?
  - By definition, when variables are independent, one variable cannot be used to predict the other
  - Correlation is null
- Proof starts with definition of covariance:
  - $\mathbb{C}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
  - If independence: $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$
  - Therefore, $\mathbb{C}(X, Y) = 0$. And as a result, correlation is null.

Independence given (null) correlation

- Knowing that $\rho(X, Y) = 0$ does **not** imply that X and Y are independent
- See figure above

Properties of variance for two variables

$$\mathbb{V}[aX + bY] = a^2\,\mathbb{V}[X] + b^2\,\mathbb{V}[Y] + 2ab\,\mathbb{C}[X, Y]$$

Therefore, if X and Y are independent:

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$$

Properties of variance for two variables

$$\mathbb{V}[aX - bY] = a^2\,\mathbb{V}[X] + b^2\,\mathbb{V}[Y] - 2ab\,\mathbb{C}[X, Y]$$

Therefore, if X and Y are independent:

$$\mathbb{V}[X - Y] = \mathbb{V}[X] + \mathbb{V}[Y]$$