

P5 - Multiple random variables

STAT 5870 (Engineering)
Iowa State University

August 28, 2024

Multiple discrete random variables

If X and Y are two discrete variables, their **joint probability mass function** is defined as

$$p_{X,Y}(x, y) = P(X = x \cap Y = y) = P(X = x, Y = y).$$

CPU example

A box contains 5 PowerPC G4 processors of different speeds:

#	speed
2	400 mHz
1	450 mHz
2	500 mHz

Randomly select two processors out of the box (without replacement) and let

- X be speed of the first selected processor and
- Y be speed of the second selected processor.

CPU example - outcomes

	Ω	1st processor (X)				
		400_1	400_2	450	500_1	500_2
2nd processor (Y)	400_1	-	x	x	x	x
	400_2	x	-	x	x	x
	450	x	x	-	x	x
	500_1	x	x	x	-	x
	500_2	x	x	x	x	-

Reasonable to believe each outcome is equally probable.

CPU example - joint pmf

Joint probability mass function for X and Y :

		1st processor (X)			
		mHz	400	450	500
2nd processor (Y)	400	2/20	2/20	4/20	
	450	2/20	0/20	2/20	
	500	4/20	2/20	2/20	

- What is $P(X = Y)$?
- What is $P(X > Y)$?

CPU example - probabilities

What is the probability that $X = Y$?

$$\begin{aligned}P(X = Y) &= p_{X,Y}(400, 400) + p_{X,Y}(450, 450) + p_{X,Y}(500, 500) \\&= 2/20 + 0/20 + 2/20 = 4/20 = 0.2\end{aligned}$$

What is the probability that $X > Y$?

$$\begin{aligned}P(X > Y) &= p_{X,Y}(450, 400) + p_{X,Y}(500, 400) + p_{X,Y}(500, 450) \\&= 2/20 + 4/20 + 2/20 = 8/20 = 0.4\end{aligned}$$

Expectation

The **expected value** of a function $h(x, y)$ is

$$E[h(X, Y)] = \sum_{x,y} h(x, y) p_{X,Y}(x, y).$$

CPU example - expected absolute speed difference

What is $E[|X - Y|]$?

Here, we have the situation $E[|X - Y|] = E[h(X, Y)]$, with $h(X, Y) = |X - Y|$. Thus, we have

$$\begin{aligned} E[|X - Y|] &= \sum_{x,y} |x - y| p_{X,Y}(x, y) = \\ &= |400 - 400| \cdot 0.1 + |400 - 450| \cdot 0.1 + |400 - 500| \cdot 0.2 \\ &\quad + |450 - 400| \cdot 0.1 + |450 - 450| \cdot 0.0 + |450 - 500| \cdot 0.1 \\ &\quad + |500 - 400| \cdot 0.2 + |500 - 450| \cdot 0.1 + |500 - 500| \cdot 0.1 \\ &= 0 + 5 + 20 + 5 + 0 + 5 + 20 + 5 + 0 = 60. \end{aligned}$$

Marginal distribution

For discrete random variables X and Y , the **marginal probability mass functions** are

$$\begin{aligned} p_X(x) &= \sum_y p_{X,Y}(x, y) & \text{and} \\ p_Y(y) &= \sum_x p_{X,Y}(x, y) \end{aligned}$$

Marginal distribution

Joint probability mass function for X and Y :

	mHz	1st processor (X)		
		400	450	500
2nd processor (Y)	400	2/20	2/20	4/20
	450	2/20	0/20	2/20
	500	4/20	2/20	2/20

Summing the rows within each column provides

x	400	450	500
$p_X(x)$	0.4	0.2	0.4

Summing the columns within each row provides

y	400	450	500
$p_Y(y)$	0.4	0.2	0.4

CPU example - independence

Are X and Y independent?

X and Y are **independent** if $p_{x,y}(x, y) = p_X(x)p_Y(y)$ for all x and y .

Since

$$p_{X,Y}(450, 450) = 0 \neq 0.2 \cdot 0.2 = p_X(450) \cdot p_Y(450)$$

they are **not** independent.

Covariance

The **covariance** between two random variables X and Y is

$$\text{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

where

$$\mu_X = E[X] \quad \text{and} \quad \mu_Y = E[Y].$$

If $Y = X$ in the above definition, then

$$\text{Cov}[X, X] = \text{Var}[X].$$

CPU example - covariance

Use marginal pmfs to compute:

$$E[X] = E[Y] = 450 \quad \text{and} \quad \text{Var}[X] = \text{Var}[Y] = 2000.$$

The covariance between X and Y is:

$$\begin{aligned} \text{Cov}[X, Y] &= \sum_{x,y} (x - E[X])(y - E[Y])p_{X,Y}(x, y) = \\ &= (400 - 450)(400 - 450) \cdot 0.1 \\ &\quad + (450 - 450)(400 - 450) \cdot 0.1 \\ &\quad + \dots \\ &\quad + (500 - 450)(500 - 450) \cdot 0.1 \\ &= 250 + 0 - 500 + 0 + 0 + 0 - 500 + 250 + 0 \\ &= -500. \end{aligned}$$

Correlation

The **correlation** between two variables X and Y is

$$\rho[X, Y] = \frac{Cov[X, Y]}{\sqrt{Var[X] \cdot Var[Y]}} = \frac{Cov[X, Y]}{SD[X] \cdot SD[Y]}.$$

Correlation properties

- ρ is between -1 and 1
- if $\rho = 1$ or -1 , Y is a linear function of X :
 - $\rho = 1 \implies Y = mX + b$ with $m > 0$,
 - $\rho = -1 \implies Y = mX + b$ with $m < 0$,
- ρ is a measure of linear association between X and Y
 - ρ near ± 1 indicates a strong linear relationship,
 - ρ near 0 indicates a lack of linear association.

CPU example - correlation

Recall

$$\text{Cov}[X, Y] = -500 \quad \text{and} \quad \text{Var}[X] = \text{Var}[Y] = 2000.$$

The correlation is

$$\rho[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \cdot \text{Var}[Y]}} = \frac{-500}{\sqrt{2000 \cdot 2000}} = -0.25,$$

and thus there is a weak negative (linear) association.

Continuous random variables

Suppose X and Y are two continuous random variables with **joint probability density function** $p_{X,Y}(x, y)$. Probabilities are calculated by integrating this function. For example,

$$P(a < X < b, c < Y < d) = \int_c^d \int_a^b p_{X,Y}(x, y) dx dy.$$

Then the **marginal probability density functions** are

$$\begin{aligned} p_X(x) &= \int p_{X,Y}(x, y) dy \\ p_Y(y) &= \int p_{X,Y}(x, y) dx. \end{aligned}$$

Continuous random variables

Two continuous random variables are **independent** if

$$p_{X,Y}(x, y) = p_X(x) p_Y(y).$$

The expected value is

$$E[h(X, Y)] = \int \int h(x, y) p_{X,Y}(x, y) dx dy.$$

Properties of variances and covariances

For any random variables X , Y , W and Z ,

$$\text{Var}[aX + bY + c] = a^2\text{Var}[X] + b^2\text{Var}[Y] + 2ab\text{Cov}[X, Y]$$

$$\begin{aligned}\text{Cov}[aX + bY, cZ + dW] &= ac\text{Cov}[X, Z] + ad\text{Cov}[X, W] \\ &\quad + bc\text{Cov}[Y, Z] + bd\text{Cov}[Y, W]\end{aligned}$$

$$\begin{aligned}\text{Cov}[X, Y] &= \text{Cov}[Y, X] \\ \rho[X, Y] &= \rho[Y, X]\end{aligned}$$

If X and Y are independent, then

$$\begin{aligned}\text{Cov}[X, Y] &= 0 \\ \text{Var}[aX + bY + c] &= a^2\text{Var}[X] + b^2\text{Var}[Y].\end{aligned}$$

Summary

- Multiple random variables
 - joint probability mass function
 - marginal probability mass function
 - joint probability density function
 - marginal probability density function
 - expected value
 - covariance
 - correlation