
THERE ARE 4 MORE WEEKS UNTIL
SUMMER BREAK!
FALSE. THERE ARE 19 MORE DAYS TO
FILL CHILDREN'S MINDS WITH
KNOWLEDGE.
makeameme.org

CONGRATS ON FINISHING ALL QUIZZES!!! :-)

Today's Agenda

Last time:

- Multinomial Distribution

Today (Lec 31, 07/18):

- (Joint) expectations
- Linear combinations

Recall the expectation of a function of **one** discrete random variable:

$$\mathsf{E}[g(X)] = \sum_{\text{all } x} g(x)f(x).$$

We can easily extend this definition into the multivariate case.

Definition

Suppose X and Y are discrete random variables with joint probability function $f(x, y)$. Then for a function $g : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathsf{E}\left[g(X, Y)\right] = \sum_{(x,y)} g(x, y) f(x, y).$$

More generally, if $g : \mathbb{R}^n \to \mathbb{R}$, and $X_1, ..., X_n$ are discrete random variables with joint probability function $f(x_1, ..., x_n)$, then

$$\mathrm{E}\left[g(X_1, ..., X_n)\right] = \sum_{(x_1, ..., x_n)} g(x_1, ..., x_n) f(x_1, ..., x_n).$$

NOTE: We can define the expectation for multivariate continuous random variables as well, but that requires multiple integrals, and that's beyond the scope of this course. Wait until STAT 330.

Example

Suppose X and Y have joint probability function given by the following table:

	f(x,y)	x = 0	x = 1	x = 2
y	0	.2	.3	.1
	2	.25	.13	.02

Compute $\mathrm{E}[XY]$ and $\mathrm{E}[Y]$.

Linearity of the expected value - again!

Linearity of expectation carries through as well. Just like the one variable case, linearity applies to both discrete and continuous multivariate distributions.

a) $E[ag_1(X, Y) + bg_2(X, Y)] = a \cdot E[g_1(X, Y)] + b \cdot E[g_2(X, Y)]$.

b) $E[X + Y] = E[X] + E[Y]$

Try to prove property 1!

Example

Let $(X_1, X_2, X_3) \sim Mult(n, p_1, p_2, p_3)$. Show that

$$\mathrm{E}[X_1 X_2] = n(n-1)p_1 p_2.$$

So far, independence is the only concept that describes the relationship between two (or more) random variables.

However, what if X and Y are not independent? Can we say anything else about them?

Yes we can, and that's the next topic: covariance.

Definition

If X and Y are jointly distributed, then $Cov(X, Y)$ denotes the **covariance** between X and Y. It is defined by

$$Cov(X, Y) = \mathrm{E}\left[(X - E(X))(Y - E(Y))\right].$$

Shortcut formula:

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

Example

Suppose X and Y have joint probability function given by the following table:

			x	
	f(x,y)	0	1	2
y	0	0.2	0.3	0.1
	2	0.25	0.13	0.02

Compute $Cov(X, Y)$.

Theorem

If X and Y are independent, then $Cov(X, Y) = 0$.

The converse statement is FALSE, namely if $Cov(X, Y) = 0$ then X and Y are not necessarily independent. Counter example: Let $X \sim N(0, 1)$, and let $Y = X^2 - 1$.

Proof

Question

Suppose $X \sim Binomial(n, p)$ and $Y \sim Binomial(m, r)$. Which of the following might not be true:

A $E(XY) \geq 0$

B $E(X + Y) = np + mr$

C $E(XY) = E(X)E(Y)$

D $P(X + Y > n + m) = 0$

Definition

The **correlation** of X and Y is denoted $corr(X, Y)$, and is defined by

$$corr(X, Y) = \rho = \frac{Cov(X, Y)}{SD(X)SD(Y)}.$$

It follows from the Cauchy-Schwarz inequality that

$-1 \leq corr(X, Y) \leq 1$, and if $|corr(X, Y)| = 1$, $X = aY + b$.

Definition

We say that X and Y are uncorrelated if $Cov(X, Y) = 0$ (or $corr(X, Y) = 0$).

Remark:

- If X and Y are independent, then X and Y are uncorrelated.
- $Cov(X, X) = Var(X)$
- The correlation is unit-free.

Properties of correlation:

a) $\rho = corr(X, Y)$ has the same sign as $Cov(X, Y)$

b) $-1 \leq \rho \leq 1$

c) $|\rho| = 1 \implies X = aY + b$

d) X, Y independent $\implies corr(X, Y) = 0$

e) $corr(X, X) = cov(X, X)/SD(X)^2 = Var(X)/Var(X) = 1$

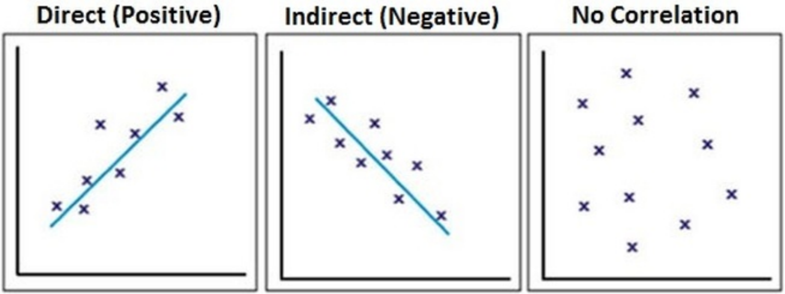
Properties 2 and 3 can be proved using Cauchy-Schwarz inequality.

ASIDE: While correlation is useful, it can be misleading as well.

This is more of a statistics issue than probability issue, but visualise the data first. Do not trust correlation blindly.

Take a look at the following link:
`https://www.autodeskresearch.com/publications/samestats`

Basic examples of correlation



The previous three plots are the “ideal” examples for correlation, because its value describes the pattern pretty accurately.

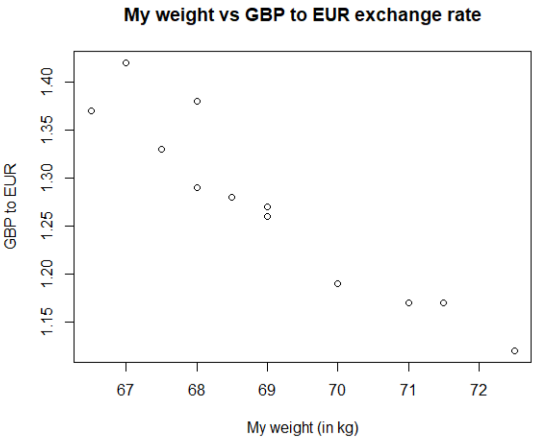
However, there are lots of cases where correlation itself can be very misleading.

This is more of a statistics issue than probability issue, but visualise the data first. Do not trust correlation blindly.

Take a look at the following link:
`https://www.autodeskresearch.com/publications/samestats`

“Correlation doesn’t imply causation”

Two variables being correlated does not always imply that one variable causes another to behave in certain ways.



Question

Suppose X, Y, and Z are jointly distributed random variables such that $corr(X, Y) = 1$, and $corr(Y, Z) \neq 0$. Which of the following is not necessarily true:

A X and Z are dependent

B $corr(X, Z) \neq 0$

C Y and Z are dependent

D $|corr(X, Z)| \neq |corr(Y, Z)|$

E $X = Y$

Definition

Suppose that $X_1, ..., X_n$ are jointly distributed RVs with joint probability function $f(x_1, ..., x_n)$. A **linear combination** of the RVs $X_1, ..., X_n$ is any random variable of the form

$$\sum_{i=1}^{n} a_i X_i$$

where $a_1, ..., a_n \in \mathbb{R}$. If $\mathbf{X} = (X_1, ..., X_n)^\top$, $\mathbf{a} = (a_1, ..., a_n)^\top$, then a linear combination is

$$\mathbf{X}^\top \mathbf{a}.$$

Example

Some "famous" linear combinations (ones you will see in STAT 231/241) are

a) The total

$$T = \sum_{i=1}^{n} X_i \quad a_i = 1, \;\; 1 \leq i \leq n$$

b) The sample mean

$$\bar{X} = \sum_{i=1}^{n} \frac{1}{n} X_i \quad a_i = \frac{1}{n}, \;\; 1 \leq i \leq n$$

Expected Value of a Linear Combination

Theorem

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

This follows directly from the linearity of expected value.

Example

Let $P_1, P_2, \ldots, P_7$ represent the number of cans of pop that Harold drinks each day (from day 1 to day 7). If each random variable P_i has mean $\mu = 6$, what is the expected number of cans consumed per day during those 7 days?

Question

Suppose $X \sim N(1,1)$ and $Y \sim U(0,1)$. Compute $E(2X - 4Y)$.

A $E(2X - 4Y) = 2$

B $E(2X - 4Y) = -4$

C $E(2X - 4Y) = 3$

D $E(2X - 4Y) = -1$

E $E(2X - 4Y) = 0$

Bilinearity of *Cov*

Theorem

Let X, Y, U, V be random variables, and $a, b, c, d \in \mathbb{R}$. Then,

$$\begin{aligned} &Cov(aX + bY, cU + dV) \\ &= acCov(X, U) + adCov(X, V) + bcCov(Y, U) + bdCov(Y, V) \end{aligned}$$

Proof: Exercise.

Remark: You can generalise this to a linear combination of arbitrary length, but the formula becomes messy.

Variance of a linear combination

The following result shows how the variance of a linear combination is "broken down" into pieces.

Theorem

Let X, Y be random variables, and $a, b \in \mathbb{R}$. Then,

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y).$$

In general,

$$Var(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i a_j Cov(X_i, X_j)$$

Example

Let X, Y be independent random variables, and $a, b \in \mathbb{R}$. What is $Var(aX + bY)$?

Example

Suppose $X \sim N(3, 4)$, and $Y \sim U(0, 1)$, and $Cov(X, Y) = -0.1$, compute $Var(2X - Y)$.