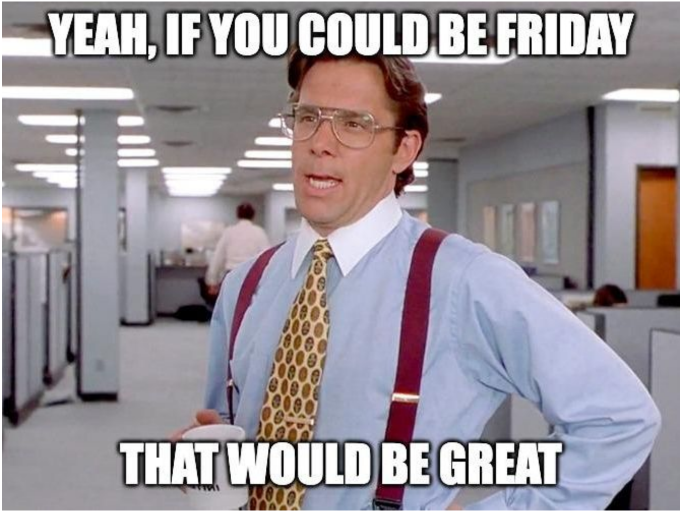
STAT 230
SECTION 2
LECTURE 28



Today's Agenda

Last time:

- The normal distribution

Today (Lec 28, 07/08):

- More examples on the normal distribution
- Multivariate Distributions (CHAPTER 9!!!)

Recap of normal distribution

- If $X \sim N(\mu, \sigma^2)$ (or $G(\mu, \sigma)$), then

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \quad x, \mu \in \mathbb{R},\ \sigma > 0$$

- Z follows standard normal distribution if $Z \sim N(0, 1)$. In this case, we have special notation for pdf and cdf:

$$f(z) = \phi(z), \quad F(z) = \Phi(z).$$

Properties of normal distribution

- $\mathrm{E}[X] = \mu$, and $Var(X) = \sigma^2$
- Normal distribution is symmetric about the mean μ:

$$P(X > \mu + t) = P(X < \mu - t).$$

- The standard normal distribution is symmetric about its mean $\mu = 0$, so that

$$P(Z \leq -x) = \Phi(-x) = 1 - \Phi(x), \quad x \in \mathbb{R}.$$

- If μ increases, the curve moves to the right, and if μ decreases, the curve moves to the left.
- Higher σ^2 means that the density curve is more "spread out".

- Standardising any normal distribution $X \sim N(\mu, \sigma^2)$:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

- ... and

$$P(X \leq x) = P(Z \leq (x - \mu)/\sigma) = \Phi((x - \mu)/\sigma)$$

  can be looked up in the Z table (use $\Phi(-x) = 1 - \Phi(x)$ for negative values of x)

- We call

$$z = \frac{x - \mu}{\sigma}$$

  the z-score of x. It's a unit free standardized measure giving the position of x on the standard normal curve.

Example

Let $X \sim N(4, 2)$. Compute the following:

- $P(X < 2.5)$
- $P(X > 1)$
- $P(-3 \leq X \leq 1)$
- $P(X \leq 4)$

Short solution

- $P(X < 2.5) = P(X \leq 2.5) = P\left(\frac{X-4}{\sqrt{2}} \leq \frac{2.5-4}{\sqrt{2}}\right) = \Phi(-1.0607) = 1 - \Phi(1.0607) = 1 - 0.85543$
- $P(X > 1) = P\left(\frac{X-4}{\sqrt{2}} > \frac{1-4}{\sqrt{2}}\right) = P(Z > -2.1213) = P(Z \leq 2.1213) = 0.98341$
- $P(-3 \leq X \leq 1) = P\left(\frac{-3-4}{\sqrt{2}} \leq Z \leq \frac{1-4}{\sqrt{2}}\right) = P(Z \leq -2.1213) - P(Z \leq -4.95) = \Phi(-2.1213) - \Phi(-4.95) = \Phi(4.95) - \Phi(2.1213) = 1 - 0.9830 = 0.017$
- $P(X \leq 4) = P\left(\frac{X-4}{\sqrt{2}} \leq 0\right) = \Phi(0) = 0.5$

The Z-table can also be used to obtain the percentile. Let's see how.

Example

Compute the

a) 75th percentile of the standard normal distribution

b) 58th percentile of the $N(5, 9)$ distribution

c) Let $Z \sim N(0, 1)$. Find c such that

$$P(-c \leq Z \leq c) = 0.95$$

An interesting empirical rule about normal distribution is the **68-95-99.7** rule, which states:

If $X \sim N(\mu, \sigma^2)$, then

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997.$$

Example

In men's tennis, the distribution of the first serve speed is normal with $\mu = 201.7$km/h and $\sigma = 9.904$km/h.

(Aside: Milos Raonic, a Canadian pro tennis player, is one of the fastest servers with the average first serve speed of 230km/h.)
Let X denote the first serve speed of a male tennis player.

a) Compute $P(X \geq 230)$.
b) Compute the probability that X is at least one standard deviation away from its mean.

Chapter 9: Multivariate distributions

"What? There are more than one ex's?"

Many, many problems can involve more than one random variable. For example, your course grade could just be represented by one random variable X, but that would ignore tons of information.

Assigning a random variable to each assessment would be more accurate:

- $X_1 =$ assignment grade,
- $X_2 =$ midterm grade,
- $X_3 =$ final exam grade.

Question

Think of a scenario where multiple random variables would be better than a single random variable.

A Don't use the course grade example. I did it already.

B This question has no "right" or "wrong" answer.

C Why are you still reading this? Start thinking!

D I'm serious! Start thinking!

As discussed previously, in many cases, we are interested in more than one random variable pertaining to an experiment.

Even more examples:

- Suppose we roll two fair six sided die,
  $X =$ outcome of the first die roll, and
  $Y =$ outcome of the second die roll.
- On a given day, let $X =$ temperature in Toronto at noon and
  $Y =$ temperature in Johannesburg at noon.
- Let $X_1, ..., X_{150}$ denote the heights of people in this room.

Repeated measurements of the same quantity, or measurements of many different quantities can be framed as observations of multiple random variables.

What is the goal of probability theory for more than one random variable?

- Computing probabilities relating to each random variable and their "joint" behaviour
  - What is the analog of the probability function/ probability density?
- Compute summary quantities describing the random variables/
  - How to calculate expected values and variances?
- Determine the distributional properties of transformations of the random variables, e.g. the sum/average.

Definition

Suppose that X and Y are *discrete* random variables defined on the same sample space (in general, when we consider two or more random variables it is assumed they are defined on the same sample space.)

The **joint probability function** of X and Y is

$$f(x, y) = P(\{X = x\} \cap \{Y = y\}) \quad x \in X(S), y \in Y(S).$$

A shorthand for this is

$$f(x, y) = P(X = x, Y = y).$$

For a collection of n discrete random variables, $X_1, ..., X_n$, the joint probability function is defined as

$$f(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n).$$

What does a 2-variable probability distribution look like?

Let $X \in \{1, 2, 3\}$ and $Y \in \{1, 2\}$, and suppose that every outcome of (X, Y) is equally likely.

Since there are $3 \times 2 = 6$ outcomes, we have $f(x, y) = 1/6$ for all (x, y) in the sample space.

So, if we tabulate this, we have

$f(x, y)$		x 1	2	3	
y	1	1/6	1/6	1/6	
	2	1/6	1/6	1/6	
					1

Table: A table of joint probability distribution of (X, Y).

Let's try one ourselves.

Example

Suppose two fair six sided die are rolled. Let X denote the outcome of the first die roll, and let Y denote the outcome of the second die roll. Compute the joint probability function of X and Y.

Properties of the joint probability function:

a) $f(x, y) \geq 0$

b) $\sum_{x,y} f(x, y) = 1.$

Point:joint probability function is still a probability function.

Recall that, when we consider multiple events, certain combinations are impossible.

Example: $A =$ "A student goes to Laurier", and $B =$ "A student majors in engineering". Then $A \cap B$ is impossible.

Similarly, certain values of (x, y) may be impossible, and in this case, we assign probability 0.

Example

Suppose a fair coin is tossed 3 times. Define the random variables $X =$ "number of Heads", and

$$Y = \begin{cases} 1 & \text{Head occurs on the first toss,} \\ 0 & \text{Tail occurs on the first toss.} \end{cases}$$

Find the joint probability function for (X, Y).

Suppose we have a multivariate random variable $(X_1, X_2, \ldots, X_n)$. We may be interested X_1. How can we "get rid of" the remaining random variables?

Question:

Let $f(x, y)$ denote the joint probability function of X and Y. Suppose that the probability functions of X and Y are $f_X(x)$ and $f_Y(y)$, respectively. Which of the following is false:

A $0 \leq f(x, y) \leq 1$

B $f(x, y) = f_X(x) f_Y(y)$

C $f(x, y) \leq f_X(x)$

D $f(x, y) \leq f_Y(y)$

Definition

Suppose that X and Y are *discrete* random variables with joint probability function $f(x, y)$.
The **marginal probability function** of X is

$$f_X(x) = P(X = x) = \sum_{y \in Y(S)} f(x, y).$$

Similarly, the marginal distribution of Y is

$$f_Y(y) = P(Y = y) = \sum_{x \in X(S)} f(x, y).$$

How does it look? Let's consider the very first joint distribution example:

			x		
$f(x,y)$		1	2	3	
y	1	1/6	1/6	1/6	
	2	1/6	1/6	1/6	
					1

If we want the marginal probability function of X, then we can sum the joint probability function values vertically, per value of x.

$f(x,y)$		x 1	2	3	
y	1	1/6	1/6	1/6	
	2	1/6	1/6	1/6	
$P(X=x)$		2/6	2/6	2/6	1

Table: Joint probability function with marginal probability of X filled out

Similarly, for the marginal probability function of Y, we can sum across the values of x, per value of y.

$f(x, y)$		x = 1	2	3	$P(Y = y)$
y	1	1/6	1/6	1/6	3/6
	2	1/6	1/6	1/6	3/6
$P(X = x)$		2/6	2/6	2/6	1

Table: Joint probability function with marginal probability of X and Y filled out

A common mistake is to think that there is a difference between the marginal distribution of X and the probability function of X. They are the same!

That being said, now it's important to distinguish probability functions by writing f_X, f_Y instead of just f.