

# Lecture 2: Probability

## STATS 101: Foundations of Statistics

Linh Tran

ThetaHat.AI@gmail.com

November 21, 2019

# Announcements

- ▶ 30 students submitted homework
- ▶ New course textbook: *DeGroot & Schervish*
- ▶ Next assignment will be posted tonight (due 12/4 @ 11:59pm)
- ▶ Re: prizes for top 3 students
  - ▶ Amazon gift cards
  - ▶ Extra points are awarded for:
    1. Class participation (e.g. asking/answering questions, etc)
    2. Catching & correcting errata/typos
    3. Answering questions / participating in discussions on Piazza
  - ▶ Blinded top 3 scores will be posted to course website
- ▶ No class next week (Happy Thanksgiving!)

All things linear algebra

- ▶ Sample space
- ▶ Probability function
- ▶ Probability space
- ▶ Random variables

**Warning:** I am assuming

- ▶ Fluency with algebra, calculus
- ▶ Familiarity with linear algebra
- ▶ Comfort with mathematical notation

**Warning:** I am assuming

- ▶ Fluency with algebra, calculus
- ▶ Familiarity with linear algebra
- ▶ Comfort with mathematical notation

**Warning:** This lecture pace is fast.

# Sample space

The set of all possible values is called the *sample space*  $S$ .

- It's the space where realizations can be produced.

# Sample space

The set of all possible values is called the *sample space*  $S$ .

- It's the space where realizations can be produced.

**Example:** Tossing a coin

$$S = \{Heads, Tails\} \quad (1)$$

# Sample space

The set of all possible values is called the *sample space*  $S$ .

- ▶ It's the space where realizations can be produced.

**Example:** Tossing a coin

$$S = \{Heads, Tails\} \quad (1)$$

More notation:

- ▶  $\emptyset$  is the *empty set*. Can be denoted as  $\emptyset = \{\}$ .
- ▶  $\cup_{i=1}^{\infty} B_i$  is the union of sets  $B_i$ . Formally,
  - ▶  $\cup_{i=1}^{\infty} B_i = \{s \in S : s \in B_i \forall i\}$
- ▶  $B \subseteq S$  means  $B$  is a *subset* of the sample space.
- ▶ *Heads*, without curly braces, is an *element* of set  $B$ .
- ▶  $B^C = S \setminus B$  is the complement of set  $B$



# Probability function

A *probability function* is a function  $P : \mathcal{B} \rightarrow [0, 1]$ , where

- ▶  $P(S) = 1$
- ▶  $P(\cup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i)$  when  $B_1, B_2, \dots$  are disjoint

# Probability function

A *probability function* is a function  $P : \mathcal{B} \rightarrow [0, 1]$ , where

- ▶  $P(S) = 1$

- ▶  $P(\cup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i)$  when  $B_1, B_2, \dots$  are disjoint

n.b. We can define the domain  $\mathcal{B}$  many ways, e.g.  $\mathcal{B} = 2^S$

# Probability function

A *probability function* is a function  $P : \mathcal{B} \rightarrow [0, 1]$ , where

- ▶  $P(S) = 1$
- ▶  $P(\cup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i)$  when  $B_1, B_2, \dots$  are disjoint

n.b. We can define the domain  $\mathcal{B}$  many ways, e.g.  $\mathcal{B} = 2^S$

**Example:** For flipping a coin, we have

$$\mathcal{B} = 2^S = \{\emptyset, \{Heads\}, \{Tails\}, \{Heads, Tails\}\} \quad (2)$$

This implies that

$$P(B) = \begin{cases} 1 & B = \{Heads, Tails\} \\ \frac{1}{2} & B = \{Heads\} \\ \frac{1}{2} & B = \{Tails\} \\ 0 & B = \emptyset \end{cases} \quad (3)$$

n.b. The power set is a 'set of sets'

**Problem:** Power sets don't work well for  $\mathbb{R}$ .

# Probability function domains

**Problem:** Power sets don't work well for  $\mathbb{R}$ .

**Solution:** Define the domain using  $\sigma$ -algebra:

- ▶  $\emptyset \in \mathcal{B}$
- ▶  $B \in \mathcal{B} \Rightarrow B^C \in \mathcal{B}$
- ▶  $B_1, B_2, \dots \in \mathcal{B} \Rightarrow \bigcup_{i=1}^{\infty} B_i \in \mathcal{B}$

# Probability function domains

**Problem:** Power sets don't work well for  $\mathbb{R}$ .

**Solution:** Define the domain using  $\sigma$ -algebra:

- ▶  $\emptyset \in \mathcal{B}$
- ▶  $B \in \mathcal{B} \Rightarrow B^C \in \mathcal{B}$
- ▶  $B_1, B_2, \dots \in \mathcal{B} \Rightarrow \bigcup_{i=1}^{\infty} B_i \in \mathcal{B}$

**Example:**

- ▶ The *discrete*  $\sigma$ -algebra:  
 $\mathcal{B} = 2^S = \{\emptyset, \{Heads\}, \{Tails\}, \{Heads, Tails\}\}$
- ▶ The *trivial*  $\sigma$ -algebra:  $\mathcal{B} = \emptyset \cup S = \{\emptyset, \{Heads, Tails\}\}$

n.b. For uncountable sets, we use the *Borel*  $\sigma$ -algebra.

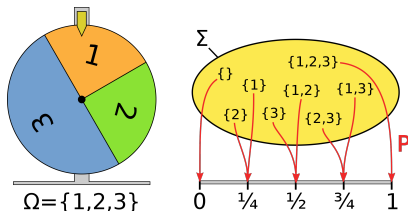
# Probability space

## Def:

A *probability space* is a triple  $(S, \mathcal{B}, P)$ .

- ▶  $S$  is the set of possible singleton events
- ▶  $\mathcal{B}$  is the set of questions to ask  $P$
- ▶  $P$  maps sets into probabilities

n.b. They represent the ingredients needed to talk about probabilities



Some properties of  $P(\cdot)$

- ▶  $P(B) = 1 - P(B^C)$
- ▶  $P(\emptyset) = 0$ , since  $P(\emptyset) = 1 - P(S)$
- ▶  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ , implying that
  - ▶  $P(A \cup B) \leq P(A) + P(B)$
  - ▶  $P(A \cap B) \geq P(A) + P(B) - 1$



# Conditional probability

For events  $A$  and  $B$  where  $P(B) > 0$ , the *conditional probability* of  $A$  given  $B$  (denoted  $P(A|B)$ ) is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (4)$$

**Example:** In an agricultural region with 1000 farms, we want to know if the farm has vineyards or cork trees.

		Cork Trees	
		Yes	No
Vineyard	Yes	200	50
	No	150	600

Table: Frequency counts

# Conditional probability

**Example:** In an agricultural region with 1000 farms, we want to know if the farm has vineyards or cork trees.

		Cork Trees	
		Yes	No
Vineyard	Yes	20%	5%
	No	15%	60%

Table: Joint probabilities

## Questions:

- ▶ What is the probability of seeing cork trees in a farm with vineyards?
- ▶ Among farms with cork trees or vineyards, what is the probability of having both?

# Conditional probability

Let's assume the following joint probabilities

		Cork Trees	
		Yes	No
Vineyard	Yes	25%	25%
	No	25%	25%

We have that  $P(A \cap B) = P(A) \cdot P(B)$ , meaning that they are *independent*

# Law of total probability

Let  $B_1, B_2, \dots, B_k \in \mathcal{B}$  and  $P(B_i) > 0 : i = 1, \dots, k$ . The *law of total probability* states that

$$P(A) = \sum_{i=1}^k P(B_i)P(A|B_i) \quad (5)$$

# Law of total probability

Let  $B_1, B_2, \dots, B_k \in \mathcal{B}$  and  $P(B_i) > 0 : i = 1, \dots, k$ . The *law of total probability* states that

$$P(A) = \sum_{i=1}^k P(B_i)P(A|B_i) \quad (5)$$

The *conditional law of total probability* states that

$$P(A|C) = \sum_{i=1}^k P(B_i|C)P(A|B_i, C) \quad (6)$$

# Bayes' Theorem

Let  $B_1, B_2, \dots, B_k \in \mathcal{B}$ ,  $P(B_i) > 0 : i = 1, \dots, k$ , and  $P(A) > 0$ .  
Then Bayes' Theorem states that for  $i = 1, \dots, k$

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^k P(B_j)P(A|B_j)} \quad (7)$$

n.b. Can be proven using the def of conditional probability

**Example:** You test positive for disease  $X$ , which has 90% sensitivity and a FPR of 10%. Past genetic screening has indicated that you have a 1 in 10,000 chance of having the disease. What is the probability of having disease  $X$ ?

**Example:** You test positive for disease  $X$ , which has 90% sensitivity and a FPR of 10%. Past genetic screening has indicated that you have a 1 in 10,000 chance of having the disease. What is the probability of having disease  $X$ ?

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} \quad (8)$$

$$= \frac{(0.9)(0.0001)}{(0.9)(0.0001) + (0.1)(0.9999)} = 0.0009 \quad (9)$$



**Example:** You test positive for disease  $X$ , which has 90% sensitivity and a FPR of 10%. Past genetic screening has indicated that you have a 1 in 10,000 chance of having the disease. What is the probability of having disease  $X$ ?

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} \quad (8)$$

$$= \frac{(0.9)(0.0001)}{(0.9)(0.0001) + (0.1)(0.9999)} = 0.0009 \quad (9)$$

Notes:

- ▶  $P(B_1)$  is often referred to as the *prior* probability
- ▶  $P(B_1|A)$  is often referred to as the *posterior* probability

# Random variables

A *random variable* is a (Borel measurable) function

$$X : S \rightarrow \mathbb{R}$$

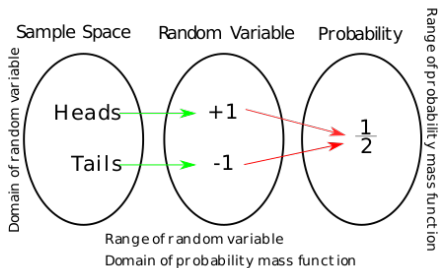
# Random variables

A *random variable* is a (Borel measurable) function

$$X : S \rightarrow \mathbb{R}$$

**Example:** For coin tossing, we have  $X : \{Heads, Tails\} \rightarrow \mathbb{R}$ , where

$$X(s) = \begin{cases} 1 & \text{if } s = Heads \\ 0 & \text{if } s = Tails \end{cases} \quad (10)$$



# Cumulative distribution function

The *cumulative distribution function* (cdf) of a random variable  $X$  is the function  $F_X : \mathbb{R} \rightarrow [0, 1]$ .

# Cumulative distribution function

The *cumulative distribution function* (cdf) of a random variable  $X$  is the function  $F_X : \mathbb{R} \rightarrow [0, 1]$ .

**Example:** For coin tossing, we have

$$X : \{Heads, Tails\} \rightarrow \mathbb{R},$$

we have

where

$$X(s) = \begin{cases} 1 & \text{if } s = Heads \\ 0 & \text{if } s = Tails \end{cases} \quad (11)$$

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \quad (12)$$

# Cumulative distribution function

The *cumulative distribution function* (cdf) of a random variable  $X$  is the function  $F_X : \mathbb{R} \rightarrow [0, 1]$ .

**Example:** For coin tossing, we have

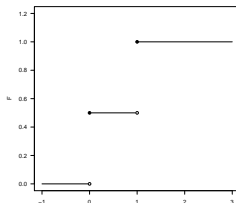
$$X : \{Heads, Tails\} \rightarrow \mathbb{R},$$

we have

where

$$X(s) = \begin{cases} 1 & \text{if } s = Heads \\ 0 & \text{if } s = Tails \end{cases} \quad (11)$$

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \quad (12)$$



# Cumulative distribution function

n.b. We have two ways of thinking about probabilities:

1. Probability functions
2. Cumulative distribution functions

**Question:** Which one should we use?

n.b. We have two ways of thinking about probabilities:

1. Probability functions
2. Cumulative distribution functions

**Question:** Which one should we use?

**The Correspondence Theorem:** Let  $P_X(\cdot)$  and  $P_Y(\cdot)$  be probability functions and  $F_X(\cdot)$  and  $F_Y(\cdot)$  be their associated cdfs. Then

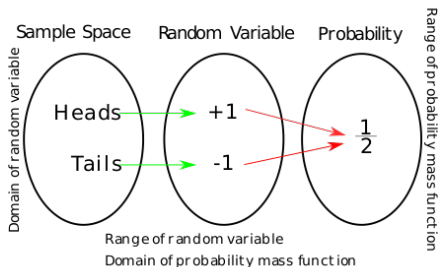
$$P_X(\cdot) = P_Y(\cdot) \iff F_X(\cdot) = F_Y(\cdot) \quad (13)$$



# Cumulative distribution function

Some properties for cdfs:

- ▶  $\lim_{x \Rightarrow -\infty} F(x) = 0$
- ▶  $\lim_{x \Rightarrow \infty} F(x) = 1$
- ▶  $F(\cdot)$  is non-decreasing
- ▶  $F(\cdot)$  is right-continuous



# Quantile function

Let  $X$  be a continuous rv and one-to-one over the the possible values of  $X$ . Then

$$F^{-1}(p) = \inf\{x \in \mathbb{R} : p \leq F(x)\} \quad (14)$$

Is the quantile function of  $X$ .

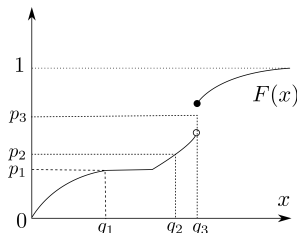
# Quantile function

Let  $X$  be a continuous rv and one-to-one over the the possible values of  $X$ . Then

$$F^{-1}(p) = \inf\{x \in \mathbb{R} : p \leq F(x)\} \quad (14)$$

Is the quantile function of  $X$ . Let  $X$  be a *discrete* rv and one-to-one over the the possible values of  $X$ . Then  $F^{-1}(p)$  states that we take the smallest value of  $x$ .

**Example:**



# Nature of random variables

A random variable  $X$  is

- ▶ *Discrete* if  $\exists f_X : \mathbb{R} \rightarrow [0, 1] \ni F_X(x) = \sum_{t \leq x} f_X(t), x \in \mathbb{R}$ 
  - ▶  $f_X$  is referred to as the probability mass function (pmf)
- ▶ *Continuous* if  $\exists f_X : \mathbb{R} \rightarrow \mathbb{R}_+ \ni F_X(x) = \int_{-\infty}^x f_X(t) dt, x \in \mathbb{R}$ 
  - ▶  $f_X$  is referred to as the probability density function (pdf).
  - ▶ n.b. We can have multiple pdf's consistent with the same cdf.
  - ▶ n.b. For any specific value of a continuous random variable, its probability is 0, i.e.  $P(\{x\}) = 0 \forall x \in \mathbb{R}$ .

# Nature of random variables

A random variable  $X$  is

- ▶ *Discrete* if  $\exists f_X : \mathbb{R} \rightarrow [0, 1] \ni F_X(x) = \sum_{t \leq x} f_X(t), x \in \mathbb{R}$ 
  - ▶  $f_X$  is referred to as the probability mass function (pmf)
- ▶ *Continuous* if  $\exists f_X : \mathbb{R} \rightarrow \mathbb{R}_+ \ni F_X(x) = \int_{-\infty}^x f_X(t)dt, x \in \mathbb{R}$ 
  - ▶  $f_X$  is referred to as the probability density function (pdf).
  - ▶ n.b. We can have multiple pdf's consistent with the same cdf.
  - ▶ n.b. For any specific value of a continuous random variable, its probability is 0, i.e.  $P(\{x\}) = 0 \forall x \in \mathbb{R}$ .

n.b. pmf's and pdf's sum to 1, i.e.

- ▶  $f : \mathbb{R} \rightarrow [0, 1]$  is the pmf of a discrete RV iff  $\sum_{x \in \mathbb{R}} f(x) = 1$
- ▶  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  is the pdf of a continuous RV iff  $\int_{-\infty}^{\infty} f(x)dx = 1$

**Example #1:** Coin tossing

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \quad (15)$$

Here,  $F_X$  is a step function with pmf

$$f_X(x) = \begin{cases} \frac{1}{2} & x \in \{0, 1\} \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

**Example #2:** Uniform distribution on  $(0,1)$

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \quad (17)$$

Here,  $F_X$  is a continuous function. Two consistent pdfs include

$$f_X(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

$$f_X(x) = \begin{cases} 1 & x \in (0, 1) \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

# Transformations of random variables

Suppose  $Y = g(X)$ , where  $g : \mathbb{R} \rightarrow \mathbb{R}$  and  $X$  is a *discrete* rv with cdf  $F_X$ .



# Transformations of random variables

Suppose  $Y = g(X)$ , where  $g : \mathbb{R} \rightarrow \mathbb{R}$  and  $X$  is a *discrete* rv with cdf  $F_X$ .

Since the function is applied to a rv,  $Y$  is also a random variable with probability function

$$f_Y(y) = P_Y(g(X) = y) = \sum_{x:g(x)=y} f_X(x) \quad (20)$$

# Transformations of random variables

Suppose  $Y = g(X)$ , where  $g : \mathbb{R} \rightarrow \mathbb{R}$  and  $X$  is a *discrete* rv with cdf  $F_X$ .

Since the function is applied to a rv,  $Y$  is also a random variable with probability function

$$f_Y(y) = P_Y(g(X) = y) = \sum_{x:g(x)=y} f_X(x) \quad (20)$$

## Example:

Let  $X$  be a uniform random variable on  $\{-n, -n+1, \dots, n-1, n\}$ . Then  $Y = |X|$  has mass function

$$f_Y(y) = \begin{cases} \frac{1}{2n+1} & \text{if } x = 0 \\ \frac{2}{2n+1} & \text{if } x \neq 0 \end{cases} \quad (21)$$

# Transformations of random variables

Suppose  $Y = g(X)$ , where  $g : \mathbb{R} \rightarrow \mathbb{R}$  and rv  $X$  with cdf  $F_X$ .

# Transformations of random variables

Suppose  $Y = g(X)$ , where  $g : \mathbb{R} \rightarrow \mathbb{R}$  and rv  $X$  with cdf  $F_X$ .

Then  $Y$  is also a random variable with cdf

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{\{x : g(x) \leq y\}} f_X(x) dx \quad (22)$$

We can get the probability function by taking the derivative

$$f_Y(y) = \frac{\partial}{\partial y} F_Y(y) \quad (23)$$

# Transformations of random variables

Suppose  $Y = g(X)$ , where  $g : \mathbb{R} \rightarrow \mathbb{R}$  and rv  $X$  with cdf  $F_X$ .

Then  $Y$  is also a random variable with cdf

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{x : g(x) \leq y} f_X(x) dx \quad (22)$$

We can get the probability function by taking the derivative

$$f_Y(y) = \frac{\partial}{\partial y} F_Y(y) \quad (23)$$

## Example:

Let  $X$  be a uniform rv on  $[-1, 1]$ . Then  $Y = X^2$  has cdf

$$\begin{aligned} F_Y(y) &= P_Y(Y \leq y) = P_X(X^2 \leq y) = P_X(-y^{1/2} \leq X \leq y^{1/2}) \\ &= \int_{-y^{1/2}}^{y^{1/2}} f(x) dx = y^{1/2} \end{aligned} \quad (24)$$

$$\text{and } f_Y(y) = \frac{\partial}{\partial y} F_Y(y) = \frac{1}{2y^{1/2}}$$

Suppose  $Y = g(X) = aX + b$ ,  $a > 0$ ,  $b \in \mathbb{R}$ . Then

$$P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y - b}{a}\right) = F_X\left(\frac{y - b}{a}\right) \quad (25)$$

# Affine transformations

Suppose  $Y = g(X) = aX + b$ ,  $a > 0$ ,  $b \in \mathbb{R}$ . Then

$$P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y - b}{a}\right) = F_X\left(\frac{y - b}{a}\right) \quad (25)$$

If  $a < 0$ , then

$$P(Y \leq y) = P(aX + b \leq y) = P\left(X \geq \frac{y - b}{a}\right) = 1 - F_X\left(\frac{y - b}{a}\right) \quad (26)$$

# Affine transformations

Suppose  $Y = g(X) = aX + b$ ,  $a > 0$ ,  $b \in \mathbb{R}$ . Then

$$P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y - b}{a}\right) = F_X\left(\frac{y - b}{a}\right) \quad (25)$$

If  $a < 0$ , then

$$P(Y \leq y) = P(aX + b \leq y) = P\left(X \geq \frac{y - b}{a}\right) = 1 - F_X\left(\frac{y - b}{a}\right) \quad (26)$$

In general, as long as the transformation  $Y = g(X)$  is monotonic, then

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\partial}{\partial y} g^{-1}(y) \right| \quad (27)$$



- ▶ Grinstead & Snell Chapters 1,2,4
- ▶ DeGroot & Schervish Chapters 1,2,3