ECON2125/4021/8013

Lecture 12

John Stachurski

Semester 1, 2015

Background Reading on Prob Theory

Most relevant

- The lecture slides
- The course notes PDF file

Least useful

- Simon and Blume
- Most other intermediate math econ books

If you really want something else

- Google for related PDFs
- Takashi Amemiya, Introduction to Statistics and Econometrics, first 6 chapters

Random Variables

What is a **random variable** (RV)?

- Bad definition: A value X that "changes randomly"
- Good definition: a function X from Ω into $\mathbb{R}$

Interpretation: RVs convert sample space outcomes into numerical outcomes

General idea:

- "nature" picks out ω in Ω
- random variable gives numerical summary $X(\omega)$

Note: Some technical details omitted — see course notes

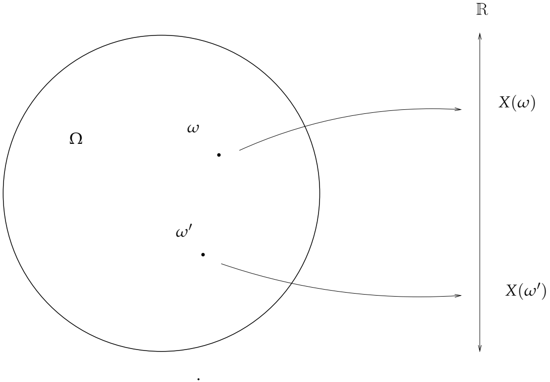
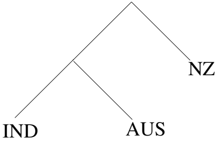


Figure : A random variable $X\colon \Omega \to \mathbb{R}$

Example. NZ in final of WC and IND, AUS in semi



Sample space for winner is

$$\Omega = \{\mathsf{AUS}, \mathsf{IND}, \mathsf{NZ}\}$$

My payoffs

$$X(\omega) = \begin{cases} 39.95 & \text{if } \omega = \mathsf{AUS} \\ -39.95 & \text{if } \omega = \mathsf{NZ} \\ -39.95 & \text{if } \omega = \mathsf{IND} \end{cases}$$

Example

Suppose Ω is set of infinite binary sequences

$$\Omega := \{(b_1, b_2, \ldots) : b_n \in \{0,1\} \text{ for each } n\}$$

We can create different random variables mapping $\Omega \to \mathbb{R}$:

- Number of "flips" till first "heads":

$$X(\omega) = X(b_1, b_2, \ldots) = \min\{n : b_n = 1\}$$

- Number of "heads" in first 10 "flips":

$$Y(\omega) = Y(b_1, b_2, \ldots) = \sum_{n=1}^{10} b_n$$

Notational Conventions for RVs

First, note that

$$\{X \text{ has some property}\} := \{\omega \in \Omega : X(\omega) \text{ has some property}\}$$

Example

$$\{X \leq 2\} := \{\omega \in \Omega : X(\omega) \leq 2\}$$

This helps us understand how to evaluate $\mathbb{P}\{X \leq 2\}$

$\mathbb{P}$ assigns probability to events, so

$$\mathbb{P}\{X \leq 2\} = \mathbb{P}\{\omega \in \Omega : X(\omega) \leq 2\}$$

Example. Recall the prob space associated with rolling a dice twice:

$$\Omega := \{(i,j) : i,j \in \{1,\ldots,6\}\} \quad \text{and} \quad \mathbb{P}(E) := \#E/36$$

If $X(\omega) = X((i,j)) = i + j$, what is $\mathbb{P}\{X \leq 3\}$?

We have

$$\begin{aligned} \{X \leq 3\} &:= \{\omega \in \Omega : X(\omega) \leq 3\} \\ &= \{(i,j) : i,j \in \{1,\ldots,6\},\ i + j \leq 3\} \\ &= \{(1,1),(1,2),(2,1)\} \end{aligned}$$

$$\therefore \quad \mathbb{P}\{X \leq 3\} = \frac{\#\{X \leq 3\}}{36} = \frac{3}{36} = \frac{1}{12}$$

Example

Let $\mathbb{P}$ be any probability on some sample space Ω

Given random variable X and scalars $a \leq b$, we claim that

$$\mathbb{P}\{X \leq a\} \leq \mathbb{P}\{X \leq b\}$$

This holds because

$$\{X \leq a\} := \{\omega \in \Omega : X(\omega) \leq a\}$$
$$\subset \{\omega \in \Omega : X(\omega) \leq b\} := \{X \leq b\}$$

Now apply monotonicity: $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

Example

As before, let $\mathbb{P}$ be any probability and X any RV

Given scalars $a \leq b$, we claim that

$$\mathbb{P}\{a < X < b\} = \mathbb{P}\{a < X \leq b\} - \mathbb{P}\{X = b\}$$

Ex. Show that

- $\{X = b\} \subset \{a < X \leq b\}$
- $\{a < X < b\} = \{a < X \leq b\} \setminus \{X = b\}$

(Translate into statments about ω as in previous slide)

Now apply $A \subset B \implies \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$

Pointwise Interpretation

In probability theory we often see statements like

- "Since $X \leq Y$, we know that..", or
- "Letting $Z := \alpha X + \beta Y$, we have..."

Such statements about RVs should be interpreted pointwise

Thus,

$$X \leq Y \iff X(\omega) \leq Y(\omega), \quad \forall \omega \in \Omega$$

$$Z := \alpha X + \beta Y \iff Z(\omega) = \alpha X(\omega) + \beta Y(\omega), \quad \forall \omega \in \Omega$$

$$X = Y \iff X(\omega) = Y(\omega), \quad \forall \omega \in \Omega$$

etc.

Types of Random Variables

There is a hierarchy of random variables, from simple to complex

1. binary random variables — take only two values
2. finite random variables — take only finitely many values
3. general random variables — range can be infinite

RVs of types 1 and 2

- are useful in practice
- are great for building intuition

Type 3 RVs are often technically demanding

But results for cases 1–2 usually carry over to case 3

A **binary random variable** is an RV taking values in $\{0,1\}$

Example. Let Ω be the sample space for rolling a dice twice

$$\Omega := \{(i,j) : i,j \in \{1,\ldots,6\}\}$$

and let

$$X(\omega) = X((i,j)) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are even} \\ 0 & \text{otherwise} \end{cases}$$

Example. Let Ω be set of infinite binary sequences and let X be existence of heads in first 5 flips

$$X(\omega) = X(b_1, b_2, \ldots) = \begin{cases} 1 & \text{if } \exists\, i \leq 5 \text{ s.t. } b_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

Indicator Functions

A useful piece of notation for binary RVs is indicator functions

Type 1: Let Q be a statement, such as "X is greater than 3"

Then the **indicator function** for Q is

$$\mathbb{1}\{Q\} := \begin{cases} 1 & \text{if } Q \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

Example. Bet payoffs from WC example

$$X(\omega) = 39.95\,\mathbb{1}\{\omega = \text{AUS}\} - 39.95\,\mathbb{1}\{\omega = \text{IND or NZ}\}$$

Type 2: Given $C \in \mathcal{F}$, the **indicator function** for C is the function

$$\mathbb{1}_C \colon \Omega \to \{0,1\}, \qquad \mathbb{1}_C(\omega) = \begin{cases} 1 & \text{if } \omega \in C \\ 0 & \text{otherwise} \end{cases}$$

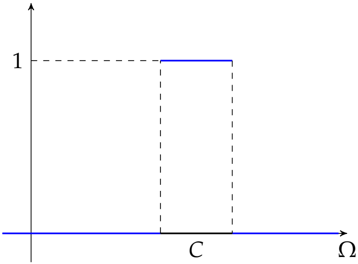


Figure : Visualization when $\Omega = \mathbb{R}$

Fact. Every binary RV is of the form $\mathbb{1}_C$ for some $C \in \mathcal{F}$

Proof: Fixing $C \in \mathcal{F}$, note that $\mathbb{1}_C$ is a binary random variable because

1. $\mathbb{1}_C$ is a map from Ω to $\mathbb{R}$ — and hence an RV
2. $\mathbb{1}_C$ takes values in $\{0, 1\}$ — and hence binary

To see that every binary RV has this form, let X be any binary random variable

Define

$$C := \{\omega \in \Omega : X(\omega) = 1\}$$

Then $X(\omega) = \mathbb{1}_C(\omega)$ for all $\omega \in \Omega$ (check it)

That is, $X = \mathbb{1}_C$

Finite Random Variables

A **finite random variable** is an RV that takes only finitely many values

- That is, X is finite $\iff$ $\mathrm{rng}(X)$ is finite

Example. Let

- Ω be set of infinite binary sequences
- X be number of heads in first N flips

That is

$$X(\omega) = X(b_1, b_2, \ldots) = \sum_{i=1}^{N} b_i$$

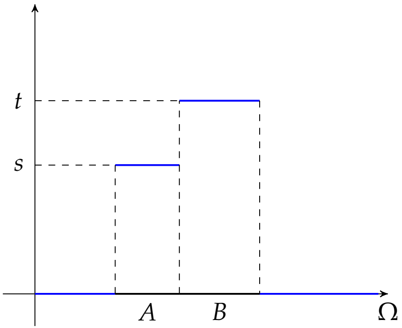
Finite RVs can be formed by taking "linear combinations" of binary RVs

Example. From WC example,

$$X(\omega) = 39.95\,\mathbb{1}\{\omega = \text{AUS}\} - 39.95\,\mathbb{1}\{\omega = \text{IND or NZ}\}$$

Example. $X(\omega) = s\,\mathbb{1}_A(\omega) + t\,\mathbb{1}_B(\omega)$ with A and B disjoint means

$$X(\omega) = \begin{cases} s & \text{if } \omega \in A \\ t & \text{if } \omega \in B \\ 0 & \text{if } \omega \in (A \cup B)^c \end{cases}$$



$$X(\omega) = s\mathbb{1}_A(\omega) + t\mathbb{1}_B(\omega) \text{ when } \Omega = \mathbb{R}$$

Fact. Every finite RV can be expressed as a linear combination of binary RVs

To see this let X be finite with $\mathrm{rng}(X) = \{s_1, \ldots, s_J\}$

Letting $A_j := \{\omega \in \Omega : X(\omega) = s_j\}$, X can be expressed as

$$X(\omega) = \sum_{j=1}^{J} s_j \mathbb{1}_{A_j}(\omega)$$

With the pointwise notational convention, also written as

$$X = \sum_{j=1}^{J} s_j \mathbb{1}_{A_j}$$

Thus, a general expression for a finite RV is

$$X = \sum_{j=1}^{J} s_j \mathbb{1}_{A_j}$$

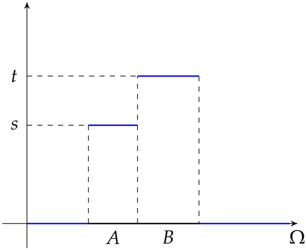
With this expression we always assume that

- the s_j's are distinct
- the A_j's are a partition of Ω

Ex. Using these assumptions, show that

1. $X(\omega) = s_j$ if and only if $\omega \in A_j$
2. $\{X = s_j\} = A_j$
3. $\mathbb{P}\{X = s_j\} = \mathbb{P}(A_j)$

Example. Recall $X = s\mathbb{1}_A + t\mathbb{1}_B$



We actually want the sets to form a partition of Ω

To do this, rewrite as

$$X = s\mathbb{1}_A + t\mathbb{1}_B + 0\mathbb{1}_{(A\cup B)^c}$$

Expectations

Roughly speaking, for a random variable X, the expectation is

$$\mathbb{E}[X] := \text{the "sum" of all possible values of } X, \\ \text{weighted by their probabilities}$$

- scare quotes because range might be uncountable

Example. Recall WC example

$$X(\omega) = 39.95\, \mathbb{1}\{\omega = \text{AUS}\} - 39.95\, \mathbb{1}\{\omega = \text{IND or NZ}\}$$

From previous lectures numbers I get $\mathbb{P}\{\omega = \text{AUS}\} = 0.39$ so

$$\mathbb{E}[X] = 39.95 \times 0.39 - 39.95 \times (1 - 0.39) = -8.79$$

Formally, for a finite RV X with range $s_1, \ldots, s_J$ we define its **expectation** $\mathbb{E}[X]$ to be

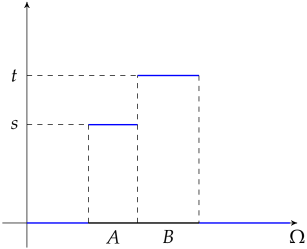
$$\mathbb{E}[X] = \sum_{j=1}^{J} s_j \mathbb{P}\{X = s_j\}$$

Fact.

$$X = \sum_{j=1}^{J} s_j \mathbb{1}_{A_j} \quad \implies \quad \mathbb{E}[X] = \sum_{j=1}^{J} s_j \mathbb{P}(A_j)$$

Proof: True because $A_j = \{X = s_j\}$

Example. Let $X = s\mathbb{1}_A + t\mathbb{1}_B + 0\mathbb{1}_{(A\cup B)^c}$



Applying the definition gives

$$\mathbb{E}[X] = s\mathbb{P}(A) + t\mathbb{P}(B) + 0 \times \mathbb{P}\{(A \cup B)^c\}$$
$$= s\mathbb{P}(A) + t\mathbb{P}(B)$$

Expectations of Binary Random Variables

Fact. If $A \in \mathcal{F}$ then

$$\mathbb{E}\left[\mathbb{1}_A\right] = \mathbb{P}(A)$$

Proof: We can write

$$\mathbb{1}_A = 1 \times \mathbb{1}_A + 0 \times \mathbb{1}_{A^c}$$

Applying the definition gives

$$\mathbb{E}\left[\mathbb{1}_A\right] = 1 \times \mathbb{P}(A) + 0 \times \mathbb{P}(A^c) = \mathbb{P}(A)$$

Fact. The expectation of a constant α is α

True meaning:

- α is the constant random variable $\alpha \mathbb{1}_\Omega$
- $\mathbb{E}[\alpha]$ is short for $\mathbb{E}[\alpha \mathbb{1}_\Omega]$

Proof: From the definition we have

$$\begin{aligned} \mathbb{E}[\alpha] &= \mathbb{E}[\alpha \mathbb{1}_\Omega] && \text{(true meaning)} \\ &= \alpha \mathbb{P}(\Omega) && \text{(by def of } \mathbb{E}\text{)} \\ &= \alpha \end{aligned}$$

Expectations of General RVs

How about the expectation of an RV with infinite range?

The idea: any RV X can be approximated by a sequence of finite-valued random variables X_n.

The expectation of X is then defined as

$$\mathbb{E}[X] := \lim_{n \to \infty} \mathbb{E}[X_n]$$

Loosely speaking, we are replacing sums with integrals

The full definition involves measure theory, so we skip it

Later we'll learn how to calculate $\mathbb{E}[X]$ in specific situations

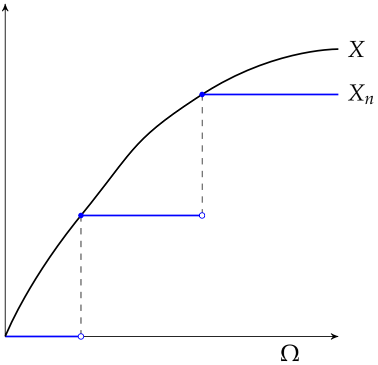


Figure : Approximation of general X with finite X_n

Monotonicity of Expectations

Fact. If X and Y are RVs with $X \leq Y$, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$

- Recall that $X \leq Y$ should be interpreted pointwise

Proof for the case $X = \mathbb{1}_A$ and $Y = \mathbb{1}_B$:

Observe that $\mathbb{1}_A \leq \mathbb{1}_B \implies A \subset B$

- To see this pick any $\omega \in A$
- Since $\mathbb{1}_A(\omega) \leq \mathbb{1}_B(\omega)$ we must have $\omega \in B$ (why?)

Now we apply monotonicity of $\mathbb{P}$ to obtain

$$\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A) \leq \mathbb{P}(B) = \mathbb{E}[\mathbb{1}_B]$$

Linearity of Expectations

Fact. If X and Y are RVs and α and β are constants, then

$$\mathbb{E}\left[\alpha X + \beta Y\right] = \alpha \mathbb{E}\left[X\right] + \beta \mathbb{E}\left[Y\right]$$

Proof for the case $\beta = 0$ and X finite

We aim to show that $\mathbb{E}\left[\alpha X\right] = \alpha \mathbb{E}\left[X\right]$ for $X := \sum_{j=1}^{J} s_j \mathbb{1}_{A_j}$

Let $Y := \alpha X$ we have

$$Y = \alpha X = \alpha \left[\sum_{j=1}^{J} s_j \mathbb{1}_{A_j}\right] = \sum_{j=1}^{J} \alpha s_j \mathbb{1}_{A_j}$$

$$\therefore \quad \mathbb{E}\left[\alpha X\right] = \mathbb{E}\left[Y\right] = \sum_{j=1}^{J} \alpha s_j \mathbb{P}(A_j) = \alpha \left[\sum_{j=1}^{J} s_j \mathbb{P}(A_j)\right] = \alpha \mathbb{E}\left[X\right]$$

Variance and Covariance

The ***k*-th moment of *X*** is defined as $\mathbb{E}\left[X^k\right]$ for $k \in \mathbb{N}$

The **variance** of X is defined as

$$\mathrm{var}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

The **standard deviation** of X is $\sqrt{\mathrm{var}[X]}$

- Measure the dispersion of X

The **covariance** of random variables X and Y is defined as

$$\mathrm{cov}[X, Y] := \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$$

- All of these might or might not exist (be finite)

Fact. If α and β are constants and X and Y are random variables, then

1. $\mathrm{var}[X] \geq 0$
2. $\mathrm{var}[\alpha] = 0$
3. $\mathrm{var}[\alpha + \beta X] = \beta^2 \, \mathrm{var}[X]$
4. $\mathrm{var}[\alpha X + \beta Y] = \alpha^2 \, \mathrm{var}[X] + \beta^2 \, \mathrm{var}[Y] + 2\alpha\beta \, \mathrm{cov}[X, Y]$

Ex. Check all these facts using the properties of $\mathbb{E}$

Correlation

Let X and Y be RVs with variances σ_X^2 and σ_Y^2

The **correlation** of X and Y is defined as

$$\operatorname{corr}[X, Y] := \frac{\operatorname{cov}[X, Y]}{\sigma_X \, \sigma_Y}$$

If $\operatorname{corr}[X, Y] = 0$, we say that X and Y are **uncorrelated**

Fact. Given RVs X and Y, constants $\alpha, \beta > 0$, we have

1. $-1 \leq \operatorname{corr}[X, Y] \leq 1$
2. $\operatorname{corr}[\alpha X, \beta Y] = \operatorname{corr}[X, Y]$

CDFs

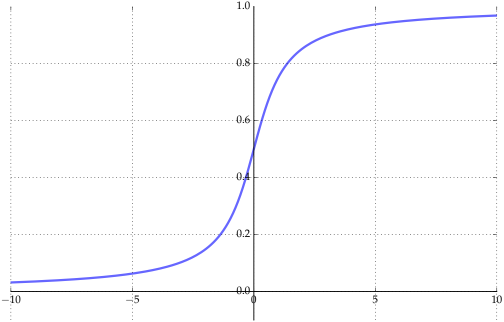
A **cumulative distribution function** (cdf) on $\mathbb{R}$ is a function $F\colon \mathbb{R} \to [0,1]$ that is

- right-continuous
- monotone increasing
- satisfies $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to \infty$

Here

- right continuity means $x_n \downarrow x$ implies $F(x_n) \downarrow F(x)$
- monotonicity $x \leq x'$ implies $F(x) \leq F(x')$

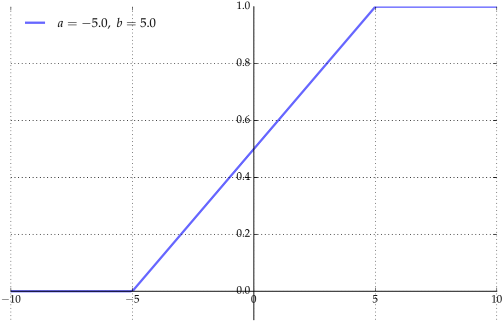
Example. The function $F(x) = \arctan(x)/\pi + 1/2$ is a cdf called the **Cauchy cdf**



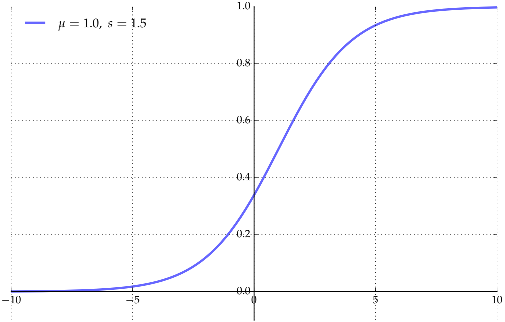
Example. Given $a < b$, the function

$$F(x) = \frac{x-a}{b-a}\mathbb{1}\{a \leq x < b\} + \mathbb{1}\{b \leq x\}$$

is a cdf called the **uniform cdf** on $[a, b]$



Example. The function $F(x) = \tanh((x - \mu)/2s)/2 + 1/2$ is a cdf for each $\mu \in \mathbb{R}$ and $s \in (0, \infty)$, called the **logistic cdf**



Distributions

Let

- Ω be any sample space
- X be any random variable on Ω
- $\mathbb{P}$ be any probability on Ω

Consider the function $F \colon \mathbb{R} \to [0,1]$ defined by

$$F(x) = \mathbb{P}\{X \leq x\}$$

This function is called the **distribution function** generated by X

We write $X \sim F$

Summarizes lots of useful information about X

Fact. The distribution function of any random variable is a cdf

Partial proof: Fix X and let F be its distribution

Let's just show that F is increasing

To see this, pick any $x \leq x'$

Note that $\{X \leq x\} \subset \{X \leq x'\}$

As a result we have

$$F(x) := \mathbb{P}\{X \leq x\} \leq \mathbb{P}\{X \leq x'\} =: F(x')$$

(Further details omitted—see course notes for related exercises)

Here's an example of how F summarizes useful info about X

Fact. If $X \sim F$ and $a \leq b$, then $\mathbb{P}\{a < X \leq b\} = F(b) - F(a)$

Proof: Recall that

$$A \subset B \implies \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$$

Also, if $a \leq b$, then

- $\{X \leq a\} \subset \{X \leq b\}$
- $\{a < X \leq b\} = \{X \leq b\} \setminus \{X \leq a\}$

$$\therefore \quad \mathbb{P}\{a < X \leq b\} = \mathbb{P}\{X \leq b\} - \mathbb{P}\{X \leq a\} = F(b) - F(a)$$

Densities and Probability Mass Functions

There are two special cases where cdfs can be reduced to simpler objects

The two cases are

1. The cdf increases only with jumps — the **discrete** case
2. The cdf is smooth with no jumps — the **density** case

Not every cdf fits into one of these categories

But when it does things are simpler

Remark: The density case is sometimes called the "continuous" case, but this is a misnomer

The Density Case

A **density function** on $\mathbb{R}$ is a function $p\colon \mathbb{R} \to [0, \infty)$ such that

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

Fact. If p is a density and F is defined by

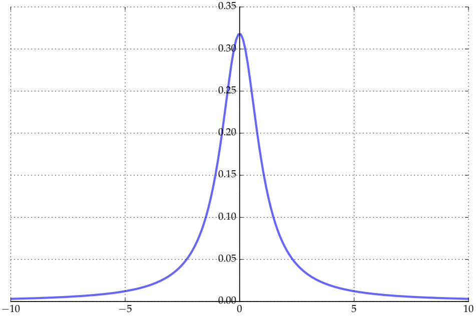
$$F(x) = \int_{-\infty}^{x} p(s)ds$$

then F is a cdf — called the cdf **generated by** p

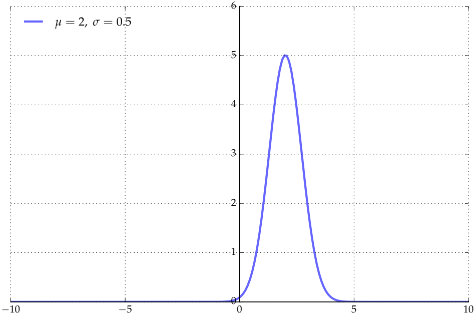
If $X \sim F$ and F is generated by density p, then we say that

- p **is the density of** X, or X **has density** p

Example. The function $p(x) = 1/(\pi + \pi x^2)$ is a density called the **Cauchy density**



Example. $p(x) = (2\pi\sigma^2)^{-1/2} \exp(-(x-\mu)^2/(2\sigma^2))$ is a density called the **normal density** and written $N(\mu, \sigma^2)$



Fact. If F is a cdf generated by density p, then F is differentiable and $F'(x) = p(x)$

Ex. Verify using the Fundamental Theorem of Calculus

Example. Recall the Cauchy cdf

$$F(x) = \arctan(x)/\pi + 1/2$$

Since

$$\frac{d}{dx} \arctan(x) = \frac{1}{1+x^2}$$

we have

$$F'(x) = \frac{1}{\pi(1+x^2)} = \text{Cauchy density } p(x)$$

Fact. If $X \sim F$ where F is generated by a density, then

$$\mathbb{P}\{X = x\} = 0 \quad \text{for every} \quad x \in \mathbb{R}$$

Proof: Suppose instead that $\exists\, \bar{x} \in \mathbb{R}$ with $\mathbb{P}\{X = \bar{x}\} = \delta > 0$

Then, for any $\epsilon > 0$,

$$\mathbb{P}\{\bar{x} - \epsilon < X \leq \bar{x}\} \geq \mathbb{P}\{X = \bar{x}\} = \delta$$

Hence

$$\frac{F(\bar{x}) - F(\bar{x} - \epsilon)}{\epsilon} \geq \frac{\delta}{\epsilon}$$

Hence F is not differentiable at $\bar{x}$

Hence F is not generated by a density — contradiction

Visualization:

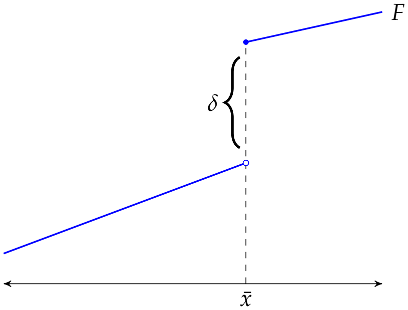


Figure : If $\mathbb{P}\{X = \bar{x}\} = \delta$, then F jumps by δ at $\bar{x}$

Fact. If X is a random variable with density p and $a \leq b$, then all of the following are true

$$\mathbb{P}\{a < X \leq b\} = \int_a^b p(s)ds$$

$$\mathbb{P}\{a \leq X < b\} = \int_a^b p(s)ds$$

$$\mathbb{P}\{a \leq X \leq b\} = \int_a^b p(s)ds$$

$$\mathbb{P}\{a < X < b\} = \int_a^b p(s)ds$$

Let's check that if X is a random variable with density p, then

$$\mathbb{P}\{a < X < b\} = \int_a^b p(s) ds$$

Proof: Letting F be the cdf generated by p,

$$\begin{aligned}
\mathbb{P}\{a < X < b\} &= \mathbb{P}\{a < X \leq b\} - \mathbb{P}\{X = b\} \\
&= \mathbb{P}\{a < X \leq b\} \\
&= F(b) - F(a) \\
&= \int_{-\infty}^b p - \int_{-\infty}^a p \\
&= \int_{-\infty}^a p + \int_a^b p - \int_{-\infty}^a p = \int_a^b p
\end{aligned}$$

The Discrete Case

A **probability mass function** (**pmf**) is a pair $\mathbf{p} := (p_1, \ldots, p_J)$ and $\mathbf{s} := (s_1, \ldots, s_J) \in \mathbb{R}^J$ with

$$0 \leq p_j \leq 1 \text{ for each } j \quad \text{and} \quad \sum_{j=1}^{J} p_j = 1$$

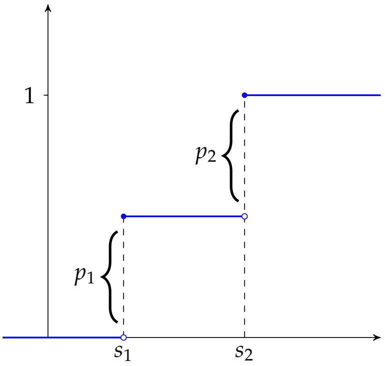
Fact. If $(\mathbf{p}, \mathbf{s})$ is a pmf, and F is defined by

$$F(x) = \sum_{j=1}^{J} \mathbb{1}\{s_j \leq x\} p_j$$

then F is a cdf — called the cdf **generated by** $(\mathbf{p}, \mathbf{s})$

- Visually, F is a step function with jump p_j at s_j

Example. $F(x) = \mathbb{1}\{s_1 \leq x\}p_1 + \mathbb{1}\{s_2 \leq x\}p_2$



If $X \sim F$ and F is generated by pmf $(\mathbf{p}, \mathbf{s})$, then we say that

- $(\mathbf{p}, \mathbf{s})$ **is the pmf of** X, or X **has pmf** $(\mathbf{p}, \mathbf{s})$

Intuitively, such an X takes value s_j with probability p_j

In particular, if

$$X = \sum_{j=1}^{J} s_j \mathbb{1}_{A_j}$$

is an RV with pmf $(\mathbf{p}, \mathbf{s})$ then

$$p_j = \mathbb{P}(A_j) \quad \text{for each } j$$

You can check these details from the definitions if you like

Neither Density nor PMF

Some cdfs fit neither the density nor the discrete case

- mixes jumps and smooth increases

There exist many results on

- decomposing such cdfs into pure jump and pure density components
- working with "measures" — objects that generalize cdfs, pmfs, densities, etc.

This is part of a field called "measure theory"

Would be a natural next step after finishing this course...

Expectations from Distributions

Let $h \colon \mathbb{R} \to \mathbb{R}$

How to calculate expectation of $Y = h(X)$?

We can use our definition of expectations but often it's not helpful

On the other hand, if $X \sim F$ and we know something about F, this can help us compute the expectation

This is true particularly when F is generated by a density or pmf

The details follow

Fact. If X is a finite RV with pmf $(\mathbf{p}, \mathbf{s})$, then

$$\mathbb{E}\left[h(X)\right] = \sum_{j=1}^{J} h(s_j) p_j$$

Proof: Let $X = \sum_{j=1}^{J} s_j \mathbb{1}_{A_j}$ with $\mathbb{P}(A_j) = p_j$

Fixing $h \colon \mathbb{R} \to \mathbb{R}$ we have

$$h(X(\omega)) = \sum_{j=1}^{J} h(s_j) \mathbb{1}_{A_j}(\omega)$$

Ex. Check it

By definition, the expectation of this discrete RV is

$$\sum_{j=1}^{J} h(s_j) \mathbb{P}(A_j) = \sum_{j=1}^{J} h(s_j) p_j$$

Fact. If X has density p, then

$$\mathbb{E}\left[h(X)\right] = \int_{-\infty}^{\infty} h(x)p(x)dx$$

Example. If we write $X \sim N(\mu, \sigma^2)$ we mean that X has density

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

In this case it's well known that

$$\mathbb{E}\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx$$

$$= \sigma^2$$

Convenient notation to unify:

$$\text{If } X \sim F, \text{ we write } \mathbb{E}\left[h(X)\right] = \int h(x)F(dx)$$

Meaning:

- In the density case,

$$\int h(x)F(dx) := \int_{-\infty}^{\infty} h(x)p(x)dx$$

- In the discrete case,

$$\int h(x)F(dx) := \sum_{j=1}^{J} h(s_j)p_j$$

Note: $\int h(x)F(dx)$ is actually the L-S integral—see course notes