Plan

- ☐ We have 10 lectures: Monday-Friday, 2 lectures per day
  - lecture notes
  - slides ="canvas" for note taking
  - videos for refreshing after the course is over
- ☐ Recitations Monday-Thursday
  - problems for the day
  - solving in groups- stop by my office! → 16:30 – 17:30
  - recitation="solving some problems"

Office hours:

- ☐ Take-home problem set
  - work in groups up to 3
  - submit solutions by September 21
  - BDP.EXERCISES@SZGERZENSEE.CH.
- ☐ Ask me clarifying questions!

Lecture 1. Introduction to Probability

Basic Definitions

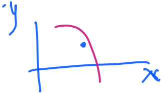
Randomness: happen or not

- ☐ ω is the outcome of experiment (example: roll of die). state of the world
- ☐ *Sample space* Ω = set of all possible outcomes
- ☐ An *event* = a subset of Ω

Exp. 1: roll 3 dices

$\Omega : i_1, i_2, i_3 \; ; \; i_j \in \{1, 2, 3 .. 6\}$ → A = {all rows even} set of outcomes

Exp 2: Throw a ball



(x, y) where ball touches the ground. → we circle radius 1.

Exp 3: Pick a person at random

Ω = {height, weight, edu, sex, ..}

↳ person is taller than 160

events are sets.

Basic Definitions

- ☐ There is a standard set of operations one can do with sets:
  - $a \in A$ (a is contained in A)
  - $A \subset B$ (A is a subset of B)
  - $A \cup B$ (the union of A and B) "or"
  - $A \cap B$ (the intersection of A and B) "and"
  - A^c (the compliment of A in Ω) "not"

Basic Definitions

set of sets that are nice & we can work w it

- A collection of subsets $\mathcal{A}$ is called a *σ-algebra* (or σ-field) if it satisfies the following conditions:
  (a) $\Omega \in \mathcal{A}$
  (b) If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$
  (c) If $A_1, A_2, \ldots, \in \mathcal{A}$ then $(\cup_{i=1}^{\infty} A_i) \in \mathcal{A}$ infinite but countable events

  Sometimes (a) is replaced with $\emptyset \in \mathcal{A}$.

Basic Definitions

we always think some prob. attached to each σ., event

- Probability measure is a real-valued set function defined on a σ-algebra $\mathcal{A}$ such that
  - (a) If $A \in \mathcal{A}$ then $\mathbb{P}(A) \geq 0$
  - (b) $\mathbb{P}(\Omega) = 1$
  - (c) If $\{A_i\}_{i=1}^{\infty}$ is a countable collection of disjoint sets in $\mathcal{A}$, then $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$
- *Probability Space*=triple $(\Omega, \mathcal{A}, \mathbb{P})$.

$\Omega = A \cup A^c$

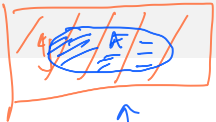
$P(\Omega) = P(A) + P(A^c)$

1 ≥ 0 ≥ 0 (+ve)

Prob. of any event betwn 0 & 1

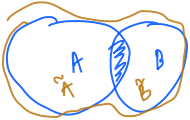
$P(A^c) = 1 - P(A)$

Basic Definitions



sum of each piece

- Some useful facts
  - (a) $0 \leq \mathbb{P}(A) \leq 1$
  - (b) $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$
  - (c) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
  - (d) If $\{A_i\}_{i=1}^{\infty}$ are a set of mutually exclusive and exhaustive subsets of Ω, then $\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap A_i)$.



A & B generally intersecting.

$A \cup B = \tilde{A} \cup \tilde{B} \cup (A \cap B)$

$\therefore P(A \cup B) = P(\tilde{A}) + P(\tilde{B}) + P(A \cap B)$

$\underbrace{P(\tilde{B}) + P(A \cap B)}_{P(B)}$

$\underbrace{P(\tilde{A}) + P(A \cap B)}_{P(A)}$

Conditional Probability

- ☐ Let A and B denote two events in Ω with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$
- ☐ the conditional probability of the event A given B is defined as

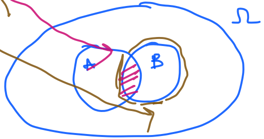
implication of this defn —

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

only possibilities w/in B which means cond.

- ☐ $\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A)$.

flipped but still in intersection



B already happend so for P(A|B) given B.

B: {dice has odd no}
A = {dice < 4}

you can update your prob.

Bayes Rule : mixing up the conditionals.

- We know $\mathbb{P}(B|A)$ and $\mathbb{P}(B|A^c)$ but we really want to know $\mathbb{P}(A|B)$
- Example: B =" a medical test comes up positive" and A = "a patient has a particular disease"
- Bayes Rule:

sick & test +ve.

$$\underbrace{\mathbb{P}(A|B)}_{\text{test positive}} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \underbrace{\mathbb{P}(B|A^c)}_{\text{false +}}\mathbb{P}(A^c)}$$

$$P(B) = P(B \cap A) + P(B \cap A^c)$$

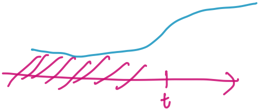
Basic Definitions

- a *cumulative distribution function (cdf)* is $F_X : \mathbb{R} \to [0,1]$:
  $F_X(t) = P\{X \leq t\}$ for all $t \in \mathbb{R}$
- *Discrete* random variable, X, is characterized by a list of possible values, $\{x_1, \ldots, x_n\}$, and their probabilities, $\{p_1, \ldots, p_n\}$
  $F_X(t) = \sum_{j:\, x_j \leq t} p_j$.
- *Continuous* random variable, Y, is characterized by its probability density function (pdf), $f_Y : \mathbb{R} \to \mathbb{R}$ s.t. $P\{a < Y \leq b\} = \int_a^b f_Y(s)ds$.
  on a line
- Properties of pdf: $\int_{-\infty}^{+\infty} f_Y(s)ds = 1$ and $f_Y(s) \geq 0$
- $F_Y(t) = \int_{-\infty}^{t} f_Y(s)ds$.

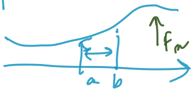
left tail

cdf: can be represented as integral

pdf is the derivation of cdf



how often falls below certain point : non-decreasing



$P[a < x < b]$

$f_x = \frac{d F_X(t)}{dt}$

Functions of Random Variables

constant: a take same value.

- If $Y = X - a$ for some $a \in \mathbb{R}$, then

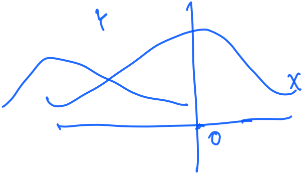
$$\underbrace{F_Y(t)}_{\text{cdf}} = P\{Y \le t\} = P\{X - a \le t\} = P\{X \le t + a\} = F_X(t + a).$$

- for continuous X: $f_Y(t) = f_X(t + a)$
- If $Y = bX$ with $b > 0$, then

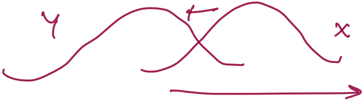
$$F_Y(t) = \underbrace{P\{bX \le t\} = P\{X \le t/b\} =}_{\text{cdf related}} F_X(t/b)$$

- for continuous X: $f_Y(t) = f_X(t/b)/b$

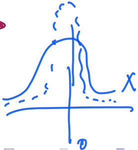
get pdf of Y by shifting X w a constant.



cdf of X



b>1 : wider Y
b<1 : narrow



squeezes.

Functions of Random Variables

- Let X be random variable and g be function
- define random variable $Y = g(X)$

$$\underbrace{F_Y(t)} = P\{Y \leq t\} = \underbrace{P\{g(X) \leq t\}} = P\{X \in g^{-1}(-\infty, t]\},$$

- If g is strictly increasing and continuously differentiable and X is a continuous, then

$$f_Y(t) = f_X(g^{-1}(t)) \left(\frac{dg(s)}{ds} \right)^{-1} \Bigg|_{s=g^{-1}(t)}$$

not use in [illegible]

Expected Value : mean va

x is discrete   realised x. prob.

- $\mathbb{E}[g(X)] = \sum_i g(x_i)p_i$ for discrete random variables
- $\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} \underbrace{g(x)}_{\text{fn } x}\underbrace{f_X(x)}_{\text{prob.}}dx$ for continuous random variables

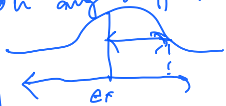
$X \sim$ fair die $\{1, 2, \ldots, 6\}$ $\frac{1}{6}, \frac{1}{6}, \ldots \frac{1}{6}$

$E\,X = \sum_{i=1}^{6} i \cdot \frac{1}{6} = 3.5$

Properties of expectation

- for any constant a (non-random), $\mathbb{E}[a] = a$.
- linearity: for random variables X and Y and constants a and b
  $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$. ← v imp.
- $\mathbb{V}ar(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. ←
- If X is a random variable and a is a constant, then
  $\mathbb{V}ar(aX) = a^2\mathbb{V}ar(X)$ and $\underbrace{\mathbb{V}ar(X + a) = \mathbb{V}ar(X)}_{\text{not depend so not shift in deviation.}}$.

where on any happens.



to measure the spread, get the distance.
(if just take simple diff, can get 0)

$$\mathbb{E}[\underbrace{X^2 - 2X}_{\text{random}}\,\mathbb{E}X + \underbrace{(\mathbb{E}X)^2}_{\text{constants}}]$$

$$\boxed{\begin{aligned} &\mathbb{E}[X^2] - 2\cdot\mathbb{E}[X]\cdot\mathbb{E}[X] + (\mathbb{E}X)^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}X)^2 \end{aligned}}$$

linear fn

$$\mathbb{E}[X^2] > (\mathbb{E}X)^2$$

Examples of distributions

- *Bernoulli*(p): it takes values from $\mathcal{X} = \{0, 1\}$,
- $P\{X = 0\} = 1 - p$ and $P\{X = 1\} = p$

$EX = 1 \cdot p + 0(1-p) = p.$

$EX^2 = p.$ (X^2 takes 0 or 1)

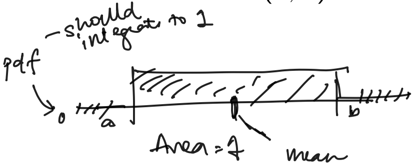
$EX^2 = 1^2 \cdot p + 0^2 \cdot (1-p) = p$

$Var[X] = EX^2 - (EX)^2$?

Examples of distributions

- Uniform(a, b): density $f_X(x) = \frac{1}{b-a}$ for $x \in (a, b)$ and $f_X(x) = 0$ otherwise
- Notation: $X \sim U(a, b)$.

$$\int_a^b f_X(x)\,dx = 1$$



$$EX = \int_a^b \frac{x}{b-a}\,dx = \frac{a+b}{2}$$

$X \sim U[a, b]$ transfer / shift so $a \to 0$ $b \to 1$ $Y = \frac{X-a}{b-a} \sim U[0,1]$

$EY^2 = \int_0^1 x^2\,dx = 1/3$ $\quad Var(Y) = \frac{1}{12}$

$Var[X] = \frac{(b-a)^2}{12}$

Shift forget about in var, only about size / squeeze which comes from squeezing

Examples of distributions

fact: It is density

- $Normal(\mu, \sigma^2)$: density

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

- Notation: $X \sim N(\mu, \sigma^2)$.

$Y = \frac{X - \mu}{\sigma}$

any gaussian by shifting & subtracting μ, squeezing to get normal Y.

cdf $F_Y = P\{Y \le y\} = P\left\{\frac{X-\mu}{\sigma} \le y\right\}$

$= P\{X \le \mu + \sigma y\}$

$= F_X(\mu + \sigma y)$

$f_Y(y) = F_Y(y)' = f_X(\mu + \sigma y)'$

$= f_X(\mu + \sigma y) \cdot \sigma$

$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\mu + \sigma y - \mu)^2}{2\sigma^2}\right\} \cdot \sigma$

$Y \sim N(0, 1)$

μ: mean

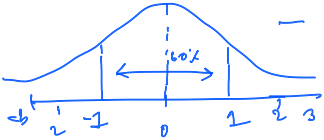
σ^2: var

Y is centered symmetric around 0

Empty slide

$X = \mu + \sigma Y \quad EX = \mu + \sigma EY = \mu$

$\mathrm{Var}(X) = \sigma^2 \, \mathrm{Var}Y = \sigma^2$



— gaussian

— mass between 1 & -1 ≈ 60%

- 2 & -2 : 95%.

- between 3 & -3 : 99%.

Problem $X \sim N(2, 4)$. what is the prob ≤ 5 }

$P(X \leq 5) = P\left\{ \frac{x - 2}{\sqrt{4}} \right. = P\left\{ \frac{X - 2}{\sqrt{4}} \leq \frac{5 - 2}{\sqrt{4}} \right.$

recentering & normalize

$= P\left(N(0, 1) \leq \frac{5 - 2}{\sqrt{4}} \right\} = \Phi\left(\frac{5 - 2}{\sqrt{4}} \right)$