126 Lecture : Random Variables

Passcode: Teapot

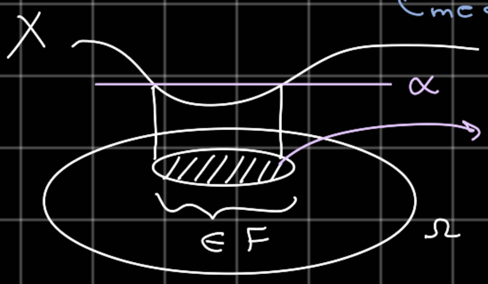
Last time: Probability spaces, conditioning (Bayes' rule, etc)

Random Variables

- def: a random variable X is a function $X: \Omega \to \mathbb{R}$

  satisfying $\{\omega : X(\omega) \leq \alpha\} \in F \;\forall \alpha \in \mathbb{R}$

  measureability



called "measureability"

technical condition ⇒ valid to write things like:

$P\{X \leq 3\} = P(\{\omega : X(\omega) \leq 3\})$

Consider: Say we want to compute $P(\alpha < X < \beta)$

↳ look at the set $\{\omega : X(\omega) < \beta\}$ ← this condition implies

↳ can rewrite as: $↳ = \bigcup_{n \geq 1} \underbrace{\{\omega : X(\omega) \leq \beta - \tfrac{1}{n}\}}_{\in F} \in F$

$\in F$ because it can be constructed with countable unions

↳ also $\{\omega : X(\omega) > \alpha\} = \{\omega : X(\omega) \leq \alpha\}^c \in F$

- the technical condition allows us to compute

  $P(X \in B)$ for pretty much any B (subset of $\mathbb{R}$) of interest.

  → "Borel sets"

- Another consequence of the def of RV's:

  If X, Y are RVs on (Ω, F, P) then:

  → $X + Y$ is a RV

  → $X \cdot Y$ " " "

  → $|X|^p$ $p \in \mathbb{R}$ is a RV

Distributions

↳ similar to a probability measure, but not quite the same; essentially a histogram of probability

↳ Frequency w/ which X takes on values

- For any RV X on probability space (Ω, F, P), we define its distribution (aka its "law") $\mathcal{L}_X$ via:

$$\mathcal{L}_X(B) := P(X \in B) \quad , B \in \mathbb{R}$$

$$\mathcal{L}_X(\{x\}) = P(X = x)$$

$\mathcal{L}_X$ is a probability measure on $\mathbb{R}$

Remark : In practice, we often describe our model for experimental outcomes in terms of distributions. But given this, I can always construct a probability space & a RV X that has this distribution.

ex : Given distribution $\mathcal{L}$, we can consider a probability space $(\mathbb{R}, \mathcal{B}, \mathcal{L})$ and RV :

$X(\omega) = \omega \qquad \omega \in \Omega = \mathbb{R}$

↳ Distribution of X:

$$\mathcal{L}_X(B) = \mathcal{L}(X \in B) = \mathcal{L}\{\omega : X(\omega) \in B\} = \mathcal{L}(B)$$

⇒ we can describe RVs in terms of their distributions & leave the underlying probability space implicit because we can construct a suitable probability space if wanted/needed)

Discrete Random Variables

- important class of RVs
- def : a RV that takes countably many values
- ex :

X = Flip of a p-biased coin

↳ Bernoulli distribution w/parameter p $(\mathrm{Bern}(p))$

X = roll a fair dice

↳ Uniform distribution $(\mathrm{Unif}\{1, 2, \cdots, 6\})$

X = # of coin flips until 1 flip heads

↳ Geometric distribution $(\mathrm{Geom}(p))$

X = # of heads in n flips

↳ Binomial distribution ($Binomial(n, p)$)

- in the discrete RV case w/ RV X, the distribution of X can be summarized by its probability Mass Function (PMF)
- <u>Def</u> (PMF)

$$P_X(x) := P\{X = x\} = P\{\omega : X(\omega) = x\}$$

↳ by the axioms: $P_X(x) \geq 0$ (non-negative)

& $\sum_{x \in \mathcal{X}} P_X(x) = 1$ ($\mathcal{X}$ = countable set of values x takes)

PMFS:

$X \sim Bern(p) \rightarrow P_X(x) = \begin{cases} (1-p) & x = 0 \\ (p) & x = 1 \end{cases}$

$X \sim Unif(\{1, \dots 6\}) \rightarrow P_X(n) = 1/6 \quad n = 1, \dots, 6$

$X \sim Geom(n, p) \rightarrow P_X(n) = \underbrace{(1-p)^{n-1}}_{\text{fail on } n-1 \text{ tosses before success}} p$

$X \sim Bin(n, p) \rightarrow P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$

Q/ what if you have 2 fcns (ie, 2 RVs) defined on a probability space?

A/ Consider joint distributions

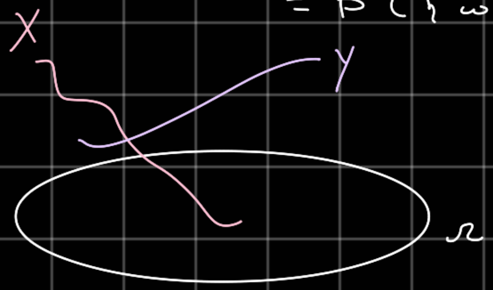
Joint Distributions

- For a pair of Discrete RVs X, Y def'd on a common probability space (Ω, F, P), the "joint distribution" of X, Y summarized by the joint PMF P_{XY}, defined via

$$P_{XY}(x, y) := P(X = x, Y = y)$$

$= P(\{\omega : X(\omega) = x \ \& \ Y(\omega) = y\})$

$= P(\{\omega : X(\omega)\} \cap \{\omega : Y(\omega) = y\})$



↝ can obtain marginal distributions by summing over ???

(by the law of total probability):

$$P_x(x) = \sum_{y \in Y} P_{xy}(x, y)$$

Aside:

X discrete:

$$\mathcal{L}_x(B) = \sum_{x \in \mathcal{X} \cap B} P_x(x)$$ (by the σ-additivity property)

↝ By writing the PMF of joint distribution this way, we can def independent RV

Independent Random Variables

- Def: discrete RVs X, Y are independent if

$$P_{xy}(x, y) = P_x(x)\, P_y(y)$$

($\Leftrightarrow$ $\{\omega : X(\omega) = x\}$ & $\{\omega : Y(\omega) = y\}$ are indep. events $\forall x, y$)

↝ concept of independence that the events that RVs induce

ex joint distributions:

$X_1 = \begin{cases} 0 & \to \text{patient tests negative w/prob } 0.9 \\ 1 & \to \text{" " positive " " } 0.1 \end{cases}$

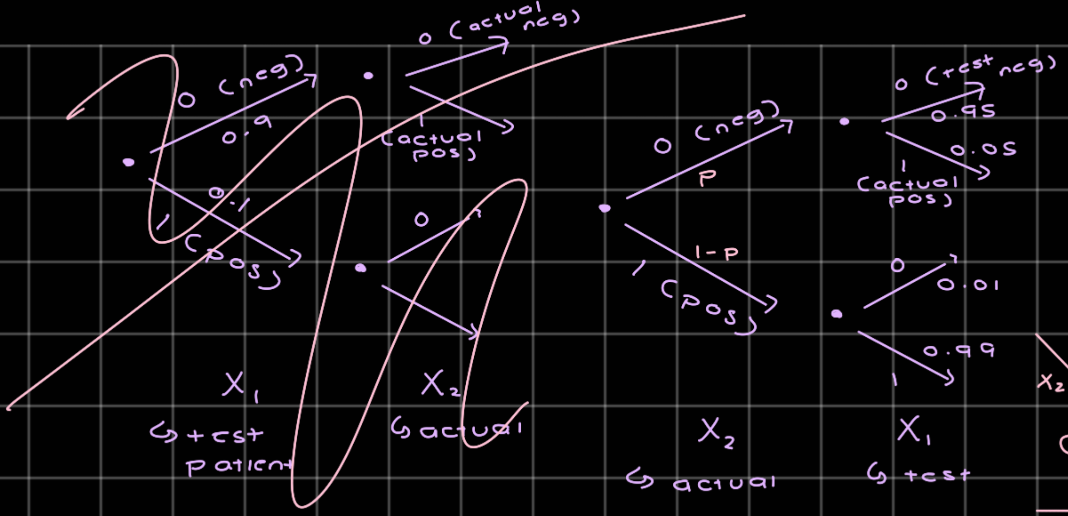
$X_2 = \begin{cases} 0 & \to \text{patient is negative} \\ 1 & \to \text{" " positive} \end{cases}$

False positive rate of test: 5%

" negative " " " : 1%.

Q/ what's joint distribution of X_1 & X_2?

↝ $P_{X_1, X_2} = ?$



$P(0.95) + (1-p)(0.01)$
$0.94p = 0.89$
$p = 0.947$

$(1-p) = 0.053$

$X_2 \backslash X_1$	0	1
0	0.947×0.95	0.947×0.05
1	$0.053 \times 0.01 = 0.00053$	0.053×0.99

Concept of joint distribution & indep. extends to any number of RVs.

Expectation

- typically don't measure outcomes of RVs, we consider the average
- takes in a RV → spits out a #
- for discrete RV X on $(\Omega, \mathcal{F}, P)$, we define its expectation as

$$\mathbb{E}X := \sum_{x \in \mathcal{X}} x \, P\{X = x\}$$

↳ weighted avg of outcomes

↳ can rewrite in terms of PMF → $= \sum_{x \in \mathcal{X}} x \, P_X(x)$

ex: $\Omega = \{0,1\}^n$

$$\underbrace{P(\{\omega\}) = p^{\#\{i : \omega_i = 1\}} (1-p)^{\#\{i : \omega_i = 0\}}}_{\text{model for sequence of indep. p-biased coin flips}} \qquad \mathcal{F} = 2^{\Omega}$$

$X(\omega) := \#\{i : \omega_i = 1\}$ = # of heads

↳ distribution → binomial

$$\mathbb{E}X = \sum_{\omega} \underbrace{\#\{i : \omega_i = 1\}}_{X(\omega)} \underbrace{p^{\#\{i:\omega_i=1\}} (1-p)^{\#\{i : \omega_i = 0\}}}_{P(\{\omega\})}$$

↳ rewrite in terms of PMF $= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$

$= np$

* THE MOST IMPORTANT PROPERTY* of Expectation is that it's linear!

↳ eg: integral: takes a # → spits out a #

↳ integrals are linear, like integral

Linearity of Expectation

- If X, Y rv's def'd on a common probability space, we have

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

↳ Note: No need for X & Y to be independent

Proof:

Lemma: $\mathbb{E}[g(z)] = \sum_{z \in Z} g(z) P_Z(z)$ (LOTUS: Law of the Unconscious Statistician)

$$\mathbb{E}[\underbrace{\alpha X + \beta Y}_{g(X,Y)}] = \sum_{x,y} (\alpha x + \beta y) P_{XY}(x,y)$$

$$= \alpha \sum_{x,y} x P_{XY}(x,y) + \beta \sum_{x,y} y P_{XY}(x,y)$$

$$= \alpha \sum_{x} x P_X(x) + \beta \sum_{y} y P_Y(y)$$

$$= \alpha \mathbb{E}X + \beta \mathbb{E}Y$$