Introduction to Machine Learning

Week 1

Lecture 2: Random Variables

Prof. Subrahmanya Swamy Peruru
Department of Electrical Engineering
IIT Kanpur

Recap

Probability $(\Omega, \mathcal{F}, P)$

Sample Space $\Omega = \{\mathrm{H}, \mathrm{T}\}$

Collection of Events $\mathcal{F} = \{\emptyset, \Omega, \{H\}, \{T\}\}$

Probability Measure

$P : \mathcal{F} \to [0, 1]$

- $P(\emptyset) = 0$
- $P(\Omega) = 1$
- $P(\{H\}) = 0.5$
- $P(\{T\}) = 0.5$

Axioms

- Normalization: $P(\Omega) = 1$
- Countable Additivity

Properties

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

Inclusion-Exclusion: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Union Bound: $P(A \cup B) \leq P(A) + P(B)$

Recap

Conditional Probability

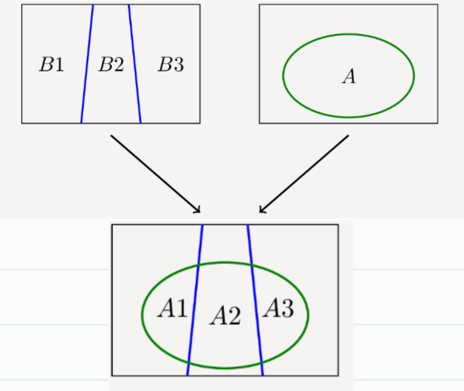
A given B: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

Independence: $P(A \cap B) = P(A)\,P(B)$

(or)

$P(A \mid B) = P(A)$

LoTP: $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$



Recap

Bayes Rule

- Consider our two-coin experiment discussed before

Question: Given the outcome is a head, what is the probability that coin 1 was chosen?

$$P(Coin\ 1 \mid Head) = \frac{P(Coin\ 1 \cap Head)}{P(Head)}$$

$$= \frac{P(Head \mid Coin\ 1)\, P(Coin\ 1)}{P(Head \mid Coin\ 1)\, P(Coin\ 1) + P(Head \mid Coin\ 2)\, P(Coin\ 2)} \quad \textbf{(LoTP)}$$

Prior: $P(Coin\ 1)$

Likelihood: $P(Head \mid Coin\ 1)$

Posterior: $P(Coin\ 1 \mid Head)$

Recap

Conditional Independence

- When you toss a coin twice
- Outcome of **first toss is independent of the second** toss
  - H_1: Head in *first* toss
  - H_2: Head in *second* toss
  - $P(H_1 \cap H_2) = P(H_1)P(H_2)$

Definition

Let $P(C) > 0$.

Then A, B are conditionally independent given C if

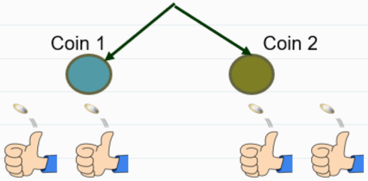
$$P(A \cap B|C) = P(A\,|C)\;P(B\,|C)$$

Another Experiment:

Suppose there are **two coins**:

- **Coin 1** lands heads with probability 0.9
- **Coin 2** lands heads with probability 0.1

Select one of the two coins uniformly at random



Toss the selected coin twice

Exercise:

Are H_1 and H_2 independent in this new experiment?

Are H_1 and H_2 conditionally independent given that Coin 1 was chosen?

Lecture Content

1 Random Variables

- Discrete RV
- Continuous RV
- CDF, PMF, PDF

2 Expectation

- Mean
- Variance
- Linearity of Expectation

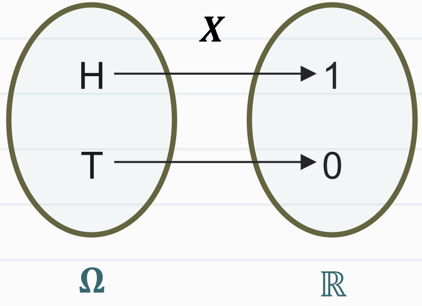
3 Multiple Random Variables

- Joint, Marginal, Conditional Dist.
- Independence
- Correlation

Random variable

- A random variable **maps** the **outcomes** of random experiment **to real numbers**

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



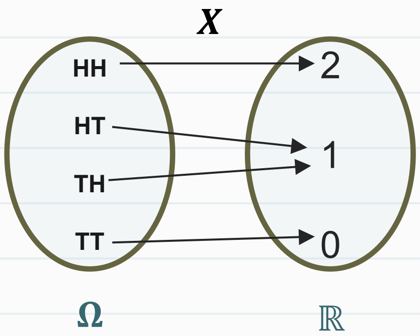
- We often write $X(\omega)$ as the value of the RV when outcome ω occurs.

$$X(H) = 1$$
$$X(T) = 0$$

Example:

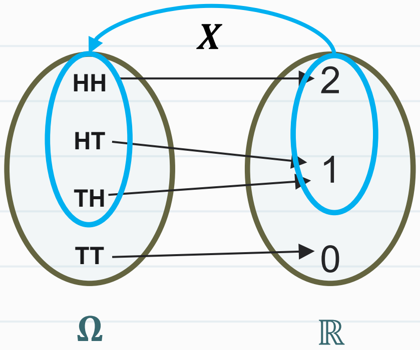
Experiment: Toss a coin twice
Random Variable: Number of Heads

$$\Omega = \{HH, HT, TH, HH\}$$



How to compute probabilities of RV ?

X = # Heads when a fair coin tossed twice



$$\mathbb{P}(X \in \{1,2\}) = \mathbb{P}(\{HH, HT, TH\}) = \frac{3}{4}$$

In General:

$$\mathbb{P}(\boldsymbol{X} \in \boldsymbol{A}) = \mathbb{P}(\boldsymbol{\omega} \mid \boldsymbol{X}(\boldsymbol{\omega}) \in \boldsymbol{A})$$

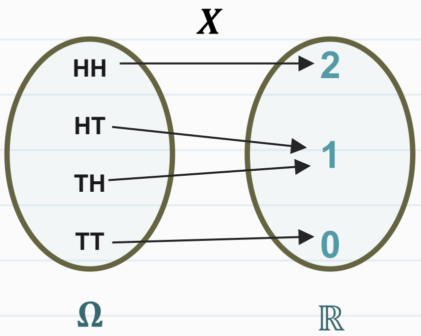
Types of RV:

- Discrete
- Continuous

Discrete RV

The **range** (support) of X is the set of values it can take:

$$\mathcal{X} = \{X(\omega) \colon \omega \in \Omega\}$$



Range $\mathcal{X} = \{0, 1, 2\}$

Discrete RV: range is countable (finite or countably infinite).

PMF: probability mass function

If $\boldsymbol{X}$ is **discrete**, define the **PMF**:

$$p_X(x) = \mathbb{P}(X = x), \quad x \in \mathcal{X}.$$

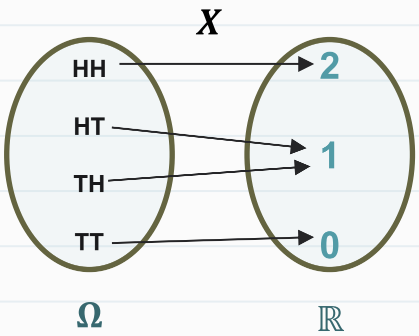
Properties:

$$p_X(x) \geq 0, \qquad \sum_{x \in \mathcal{X}} p_X\,(x) = 1.$$

For any event $\{X \in A\}$ with $A \subseteq \mathcal{X}$:

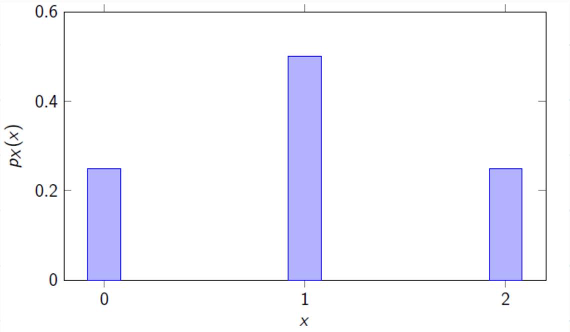
$$\mathbb{P}(X \in A) = \sum_{x \in A} p_X\,(x).$$

Example: PMF



Range $\mathcal{X} = \{0, 1, 2\}$

$$p_X(0) = \mathbb{P}(TT) = \frac{1}{4}, \quad p_X(1) = \mathbb{P}(HT \text{ or } TH) = \frac{1}{2}, \quad p_X(2) = \mathbb{P}(HH) = \frac{1}{4}$$



Verify: $\sum_{x \in \mathcal{X}} p_X(x) = 1$

$$\frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$$

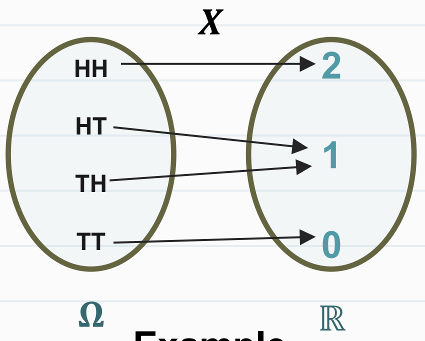
PMF is a function on discrete points; total mass sums to 1

CDF: Cumulative Distribution Function

PMF: Only for Discrete RV,

CDF: For any RV

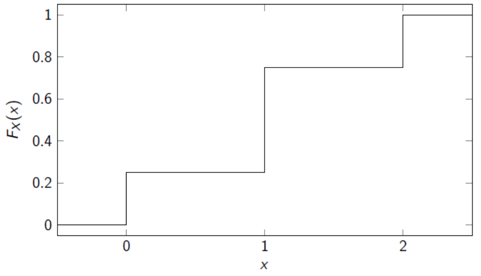
$$\boldsymbol{F_X(\mathrm{x}) = \mathbb{P}(\mathrm{X} \leq \mathrm{x})}$$



Example

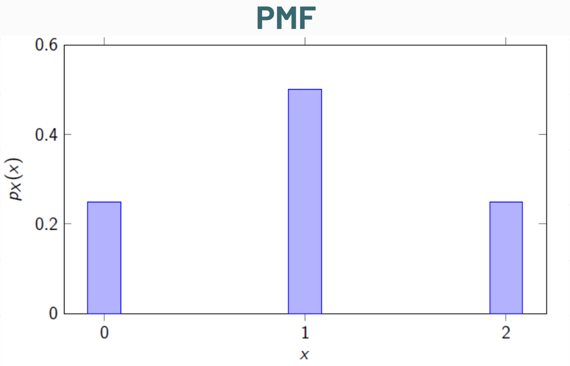
- $F_X(-1) = \mathbb{P}(X \leq -1) = 0$
- $F_X(0) = \mathbb{P}(X \leq 0) = \frac{1}{4}$
- $F_X(1) = \mathbb{P}(X \leq 1) = \frac{3}{4}$
- $F_X(2) = \mathbb{P}(X \leq 2) = 1$
- $F_X(3) = \mathbb{P}(X \leq 3) = 1$

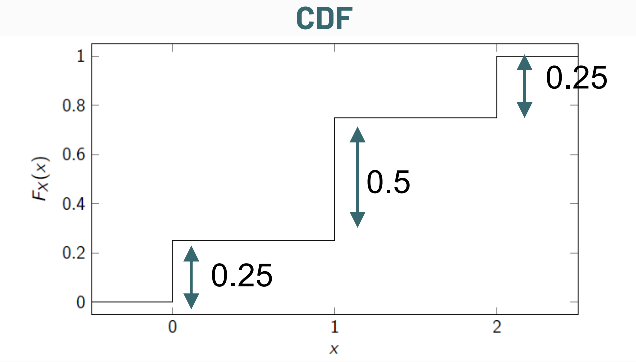
CDF



- $F_X(x)$ is non-decreasing in x.
- $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to+\infty} F_X(x) = 1$.
- For any $a < b$: $\mathbb{P}(\mathbf{a < X \leq b}) = \boldsymbol{F_X(\mathrm{b}) - F_X(\mathrm{a})}$.

Discrete RVs: PMF and CDF





Discrete CDF is a **step function**

Jumps equal the PMF masses

$$\mathbb{P}(X \leq x) = \boldsymbol{F_X(x)} = \sum_{\boldsymbol{y=-\infty}}^{\boldsymbol{x}} \boldsymbol{p_X(y)}$$

Moving to Continuous RVs

Suppose X takes one of the three values in $\{2,\ 2.25,\ 2.5\}$ uniformly at random

Discrete

$$\mathbb{P}(X = 2) = \frac{1}{3}$$

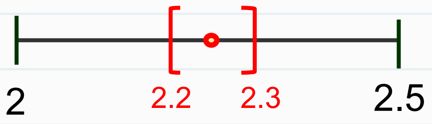
$$\mathbb{P}(X = 2.25) = \frac{1}{3}$$

$$\mathbb{P}(X = 2.5) = \frac{1}{3}$$

PMF

Suppose X takes a real number in the interval $[2, 2.5]$ uniformly at random

What is $\mathbb{P}(X = 2.25)? = 0$



What is $\mathbb{P}(X \in [2.2, 2.3])$ $\longrightarrow$ **Can be answered using PDF**

Continuous RVs: PDF

PDF is the derivative of CDF:

The probability density function (PDF) is given by

$$f_X(x) = \frac{dF_X(x)}{dx}, \qquad \text{(wherever } F_X(x) \text{ is differentiable)}$$

Properties:

$$f_X(x) \geq 0$$

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(t)\, dt$$

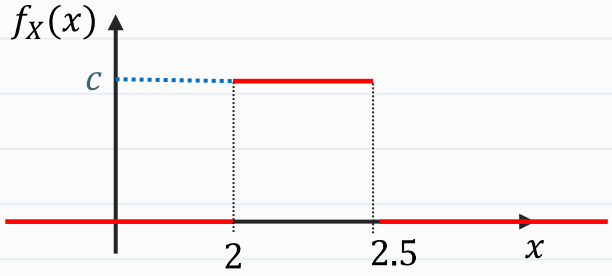
$$\int_{-\infty}^{\infty} f_X(t)\, dt = 1$$

CDF is an integral of PDF:

$$\mathbb{P}(X \leq x) = F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

Example: Continuous RV

Example: $X \sim Unif[2, 2.5]$



$$\int_{-\infty}^{+\infty} f_X(x)\, dx = 1 \quad \Rightarrow \int_{2}^{2.5} c\, dx = 1$$

$$\Rightarrow c = 2$$

Important Observation:

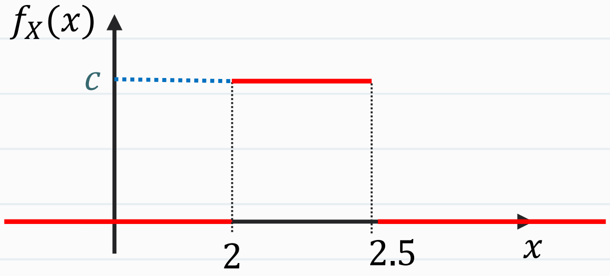
- **The value of a PDF can be greater than 1.**
- This is because the PDF itself is **not a probability.**
- **Probabilities are obtained by integrating the PDF.**

Compute $\mathbb{P}(2.25 \leq X \leq 2.5)$

$$\int_{2.25}^{2.5} f_X(x)\, dx = \int_{2.25}^{2.5} 2\, dx = 2 \times (2.5 - 2.25) = 0.5$$

Example: Continuous RV

Exercise: Compute **CDF** of $X \sim Unif[2, 2.5]$



Integrate the PDF

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt$$

Expectation of a random variable

For discrete X:

$$\mu(X) = \mathbb{E}[X] = \sum_x x \; p_X(x)$$

For continuous X:

$$\mu(X) = \mathbb{E}[X] = \int_{-\infty}^{\infty} x \; f_X(x) \, dx$$

Discrete → Continuous:
PMF → PDF
Sum → Integral

Example (Bernoulli RV): Find $E[X]$ when the PMF is given by

$$p_X(1) = p$$
$$p_X(0) = 1 - p$$

$$E[X] = 1 \times p + 0 \times (1 - p) = p$$

Exercise: For the above random variable X, **Find $\boldsymbol{E[X^3 + 2X + 3]}$**

Function of a RV: $Y = g(X)$ (For example, $Y = X^2$)

How to find $\mathbb{E}[Y]$?

Approach 1: Compute the PMF $p_Y(y)$ and then compute $\mathbb{E}[\mathrm{Y}] = \sum_{\mathrm{y}} y \; p_Y(\mathrm{y})$

Approach 2 (Easier): Directly compute $\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum_x g(x) \, p_X(x)$

Note: Expectation exists only when the absolute-value integrals/sums are finite

Expectation is Linear

For any $a, b \in \mathbb{R}$:

$$\mathbb{E}[a\,X + b] = a\,\mathbb{E}[X] + b$$

For any set of RVs $X_1, X_2, \ldots, X_n$:

$$\mathbb{E}[X_1 + X_2 + \cdots X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots \mathbb{E}[X_n]$$

Example: $X \sim Binomial\,(n, p)$. Find $\mathbb{E}[X]$.

- Consider a coin whose **probability of heads is $\boldsymbol{p}$**
- Let $\boldsymbol{X}$ be the **number of heads** observed in $\boldsymbol{n}$ **independent tosses of the coin**.

Using Standard Approach:

- Start with the Binomial PMF $p_X(x) = \binom{n}{x}\,p^x\,(1-p)^{n-x}$
- Compute $\mathbb{E}[X] = \sum_x x\,p_X(x)$

Using Linearity of Expectation:

- Let $X_1, X_2, \ldots X_n$ be $Bernoulli(p)$ RVs
- taking a value $1/0$ depending on whether the corresponding toss is a head/tail
- $X = X_1 + X_2 + \cdots X_n$
- $\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots \mathbb{E}[X_n] = p + p + \cdots p = np$

Variance and standard deviation

Variance: $\mathrm{Var}(\boldsymbol{X}) = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])^{2}\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

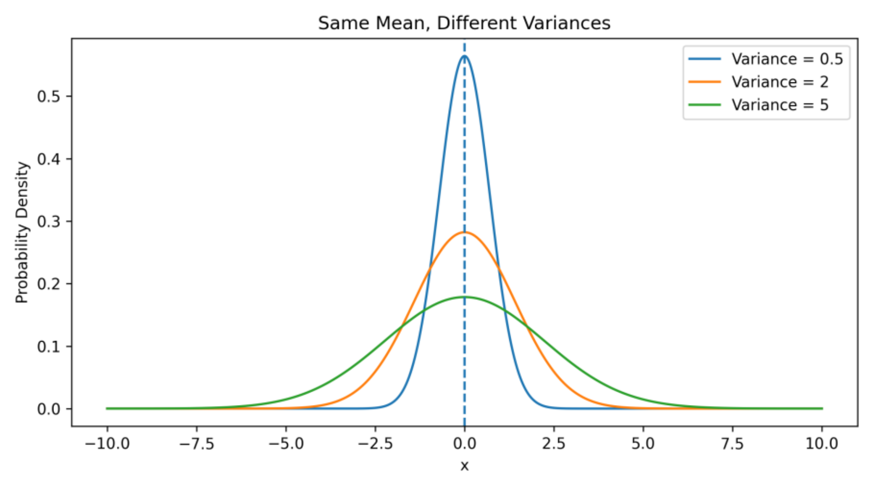
Standard deviation: $\boldsymbol{\sigma_X} = \sqrt{\mathbf{Var(X)}}$

Example: Gaussian Distribution

$X \sim \mathcal{N}(\mu, \sigma^2)$:

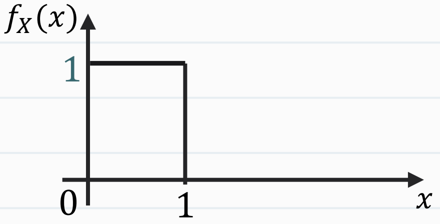
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\,\sigma^2}\right)$$

Variance measures how spread out the values are around the mean



Examples : $\mathbb{E}[X]$ and $\mathrm{Var}(X)$

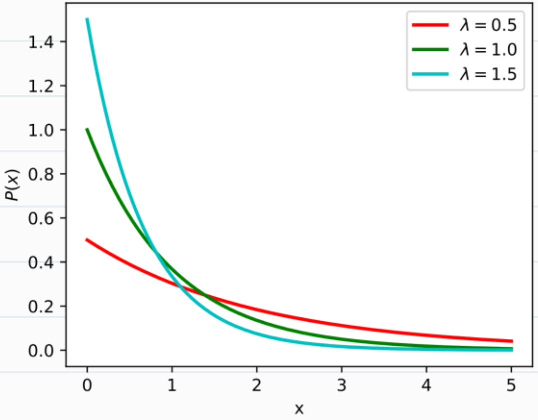
Uniform RV $X \sim Unif(0,1)$



- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x) dx = \int_0^1 x\ dx = \frac{1}{2}$
- $\mathbb{E}[X^2] = \int_0^1 x^2\ dx = \frac{1}{3}$
- $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

Exercise: Solve this problem for
$X \sim Unif(a, b)$

Exponential RV $X \sim \mathrm{Exp}(\lambda)$



- $f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$
- $\mathbb{E}[X] = \int_0^{\infty} x\, \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}$
- $Var(X) = \frac{1}{\lambda^2}$

Multiple RVs: Joint PMF (discrete)

Suppose I want to characterize two random variables (X, Y)

Is it sufficient to know their PMFS: $p_X(x), p_Y(y)$?

NO

Example: Predict rain or not based on

$$X = Temperature$$
$$Y = Humidity$$

For discrete X, Y, **joint PMF**:

$$p_{X,Y}(x, y) = \mathbb{P}(X = x,\ Y = y).$$

$$\sum_{x,y} p_{XY}(x, y) = 1$$

Example:

$$X = Cointoss\ \{0\ , 1\},$$
$$Y = 3 - Faced\ die\ \{1, 2, 3\}$$

	$y = 0$	$y = 1$	$y = 2$
$x = 0$	$p_{XY}(0, 0)$	$p_{XY}(0, 1)$	$p_{XY}(0,2)$
$x = 1$	$p_{XY}(1, 0)$	$p_{XY}(1, 1)$	$p_{XY}(1, 2)$

Marginal PMF

Marginals: $p_X(x),\ p_Y(y)$

Relation to Joint PMF:

$$p_X(x) = \sum_y p_{X,Y}(x,y), \qquad p_Y(y) = \sum_x p_{X,Y}(x,y).$$

Marginal Probability Example:

Suppose (X,Y) has joint PMF:

	$y=0$	$y=1$	$y=2$
$x=0$	0.10	0.10	0.05
$x=1$	0.20	0.25	0.10
$x=2$	0.05	0.10	0.05

Check total probability: sum of all entries $= 1$.

Compute marginals:

$$p_X(1) = 0.20 + 0.25 + 0.10 = 0.55,$$
$$p_Y(1) = 0.10 + 0.25 + 0.10 = 0.45.$$

Can we find the joint PMF from the marginals PMFs ?

No (unless they are independent)

Independence of RVs

Discrete case: X and Y are independent if

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \quad \forall \mathbf{x}, \mathbf{y}$$

Example

	$y = 0$	$y = 1$	$y = 2$
$x = 0$	0.10	0.10	0.05
$x = 1$	0.20	0.25	0.10
$x = 2$	0.05	0.10	0.05

$$p_{X,Y}(1,1) = 0.25.$$

$$p_X(1) = 0.20 + 0.25 + 0.10 = 0.55,$$

$$p_Y(1) = 0.10 + 0.25 + 0.10 = 0.45.$$

But

$$p_X(1)p_Y(1) = 0.55 \times 0.45 = 0.2475 \neq 0.25.$$

So X and Y are **not independent**.

Conditional PMF (discrete)

If $p_Y(y) > 0$, define:

$$\boldsymbol{p}_{\mathbf{X}|Y}(\mathbf{x}|\mathbf{y}) = \frac{\boldsymbol{p}_{\mathbf{X},Y}(\mathbf{x},\mathbf{y})}{\boldsymbol{p}_Y(\mathbf{y})}.$$

Using the table:

$$p_{X|Y}(1|1) = \frac{p_{X,Y}(1,1)}{p_Y(1)} = \frac{0.25}{0.45} = \frac{5}{9} \approx 0.556.$$

Joint PMF

	$y = 0$	$y = 1$	$y = 2$
$x = 0$	0.10	0.10	0.05
$x = 1$	0.20	0.25	0.10
$x = 2$	0.05	0.10	0.05

Expectation with two RVs

$$\mathbb{E}[X] = \sum_x x\ p_X(x), \qquad \mathbb{E}[Y] = \sum_y y\ p_Y(y).$$

A function of two RVs $\boldsymbol{g(X, Y)}$ $(Example: g(X, Y) = XY)$

$$\mathbb{E}[\boldsymbol{g}(\mathbf{X}, \mathbf{Y})] = \sum_{\boldsymbol{x}} \sum_{\boldsymbol{y}} \boldsymbol{g}(\boldsymbol{x},\ \mathbf{y})\ \boldsymbol{p}_{\mathbf{X},\boldsymbol{Y}}(\mathbf{x}, \mathbf{y}).$$

Useful identity (if independent):

$$X \perp Y \quad \Rightarrow \quad \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Recall: (**Linearity of Expectation even if X, Y are not independent**)

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

Conditional Expectation

$$\mathbb{E}[\mathbf{Y}|\mathbf{X}=\mathbf{x}] = \sum_{y} y \; p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$$

Example: Compute $\mathbb{E}[\mathbf{Y} \mid \mathbf{X} = \mathbf{1}]$

$$p_{Y|X}(y|1) = \frac{p_{X,Y}(1,y)}{p_X(1)} = \frac{p_{X,Y}(1,y)}{0.55}.$$

So:

$$\mathbb{E}[\boldsymbol{Y} \mid \mathbf{X} = \mathbf{1}] = \sum_{y} y \; p_{Y|X}(y|\mathbf{1})$$

$$= 0 \times p_{Y|X}(0 \mid \mathbf{1}) + 1 \times p_{Y|X}(1 \mid \mathbf{1}) + 2 \times p_{Y|X}(2 \mid \mathbf{1})$$

$$= 0 \times \frac{0.20}{0.55} + 1 \times \frac{0.25}{0.55} + 2 \times \frac{0.10}{0.55}$$

$$= 0.82$$

Joint PMF

	$y = 0$	$y = 1$	$y = 2$
$x = 0$	0.10	0.10	0.05
$x = 1$	0.20	0.25	0.10
$x = 2$	0.05	0.10	0.05

Covariance and correlation

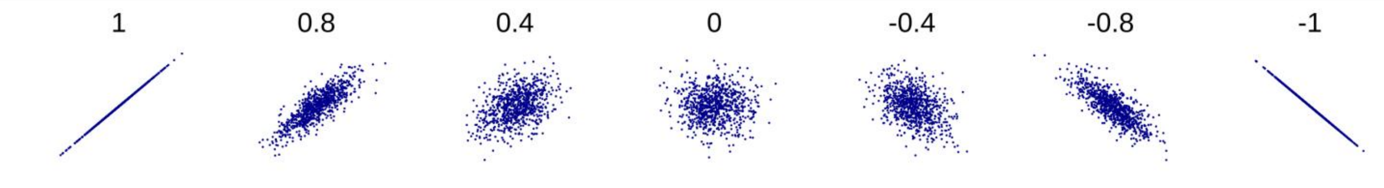
Covariance:

$$\mathrm{Cov}(\mathbf{X},\mathbf{Y}) = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[\mathbf{XY}] - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{Y}].$$

Correlation Coefficient:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \in [-1,1].$$

Several sets of (x, y) points, with the correlation coefficient of x and y for each set.



The correlation captures

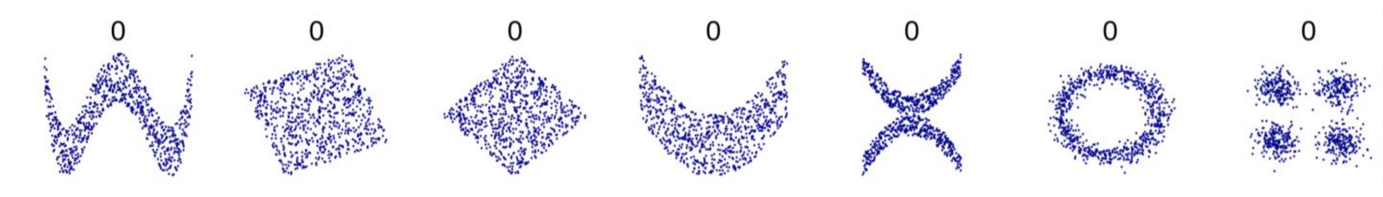
- the presence of a **linear relationship**
- and how **noisy** that relationship is

Uncorrelated RVs: X and Y are uncorrelated if $\boldsymbol{Cov(X,Y) = 0}$ (or equivalently) $\boldsymbol{\rho_{X,Y} = 0}$

Figure Source: https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

Zero Correlation?

Correlation does not capture non-linear relationships



Several sets of (x, y) points, with the correlation coefficient of x and y for each set.

Independence ⇒ zero covariance, but **not conversely**

Independent $\Rightarrow$ Uncorrelated

Uncorrelated $\not\Rightarrow$ Independent

Figure Source: https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

Example: uncorrelated but dependent

Let $\mathbf{X} \sim \mathbf{Unif}(-\mathbf{1}, \mathbf{1})$ and define $\mathbf{Y} = \boldsymbol{X}^{\mathbf{2}}$.

Then $\mathbb{E}[X] = 0$ by symmetry.

Compute:

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X^3] - 0.$$

But X^3 is an odd function and X is symmetric on $[-1,1]$, so $\mathbb{E}[X^3] = 0$.

Thus $\mathbf{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$ (uncorrelated), ***yet*** **$\mathbf{Y}$ is a deterministic function of $\mathbf{X}$** $\Rightarrow$ dependent.

References

- *Chapters 3-5* of the book **H. Pishro-Nik, "Introduction to probability, statistics, and random processes",** available at https://www.probabilitycourse.com, Kappa Research LLC, 2014.

Thank You