Lorenzo Zambelli
web: lorenzozambelli.it

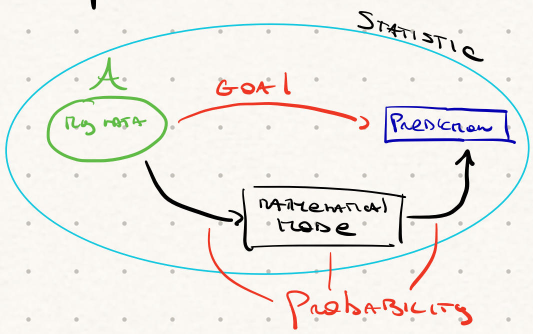
Probability Theory

Tutorial 1

Material: - Combinatorics

→ Extra Intro

What is probability?



At its core → "Measure theory"

Probability	Statistic
logically self contained	Messier, an art
A few rules for computing prob.	seek to make probability inference from data
One correct answer	No single correct answer
"The skeleton"	"The human"

Probability is just a set of rules for counting and computing "measure" (probability) of a set.

We will learn:
- combinatorics
- Basic set theory / Probability space
- Distributions and some major ones

→ Review

Can you count?

Numbers of possible arrangements of size r from n object

	Without Replacement	With Replacement
Ordered	$\frac{n!}{(n-r)!}$	n^r
Unordered	$\binom{n}{r}$	$\binom{n+r-1}{r}$

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

Binomial coefficient

1. How many ways we can choose a sequence r from n objects?
   n^r

2. How many ways are there to organize n object?
   $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$

3. How many subset $\{1, \ldots, n\}$ of cardinality k are there?
   $\binom{n}{k}$

4. How many ways are there to draw k balls out of n with replacement but without order?
   $\binom{k+n-1}{k}$

lets consider a lock with k digits

1	2	3	.....	k

(1) how many combinations does the lock has?

we have 10 values for each digit, hence

$$\underbrace{10 \cdot 10 \cdot 10 \cdot \dots \cdot 10}_{k} \Rightarrow 10^k$$

(2) If we can't have repetition and we have 10 digits?

-> 1st: 10 -> 3rd: 10 − 2

-> 2nd: 10 − 1 -> 10th: 1

If each digit must be ≠ from those next to it, how many combination?

-> 1st: 10 possibilities,

-> 2nd: 10 − 1

-> 3rd: $(10-1)-1+\underbrace{1}_{\text{the one from first}}$

-> 4th: 10 − 1

⋮ ⋮ $\Rightarrow 10^2 \cdot (10-1)^{k-2}$

-> kth: 10

(3) how many ways can we have a value X appearing r times?

If r are taken, then the remaining k − r digits can be filled with the remaining 9 values.

Thus

1	2	3	...	k

r have X

$$\overbrace{\underbrace{X \cdots X}_{r!}\ \underbrace{k{-}r{+}1 \cdots k}_{(k-r)!}}^{k!}$$

r! := # of ways to organize a sequence of r elem

(k−r)! := # " of k−r elements

$$\Rightarrow \frac{k!}{r!(k-r)!} = \binom{k}{r}$$ otherwise we count twice!

-> Extra Exercise week 1

① You are supervising 250 students. You assign them at random into 50 groups consisting of 5 students each.
What is the probability that in at least one of the 50 groups all students have a BSN starting with the same two digits (assuming all digits combinations are equally likely)?

② You are presented with a box containing 9 locks and 9 associated keys. You take 4 locks and 4 keys at random. What is the probability that you have picked at least one lock together with the correct key?

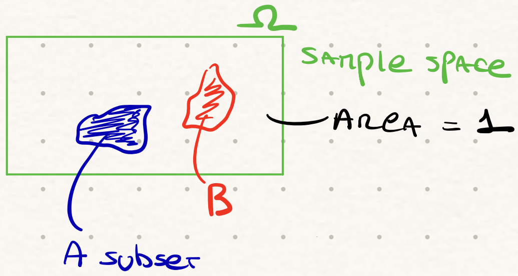
③ (a)
Anna, Ben, Hiske and Jelmer play a card game. The deck consist of 52 cards containing precisely 4 Queens and 4 Kings. The cards are shuffled at random and then each player receives exactly 13 cards. What is the probability that each player receives exactly one Queen and exactly one King?

(b) Assume that Anna receives indeed precisely one Queen and precisely one King. Ana holds all the cards in her hands in a random order. What is the probability that the Queen and King are adjacent positions?

Lorenzo Zambelli
web: lorenzozambelli.it

Tutorial 2

→ Review and extra



— Area = 1 ← Normalize s.t. we can compare with different random experiment

$\mathbb{P} : \mathcal{A} \longrightarrow \mathbb{R}$

$\mathbb{P}$ — measure function → function to compute the size "measure" of a set.

$\mathcal{A}$ — collection of subset

We want:

1. $\mathbb{P}(\Omega) = 1$ & $\mathbb{P}(\phi) = 0$
2. $\mathbb{P}(A) \in [0,1]$
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ if A, B disjoint
4. $\mathbb{P}(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$ if we have pairwise disjoint sets

} Kolmogorov's Axioms

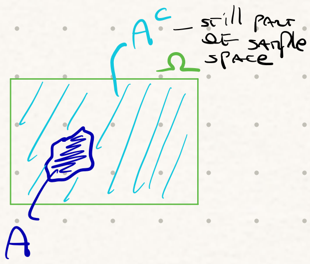
└→ we need a σ-Algebra

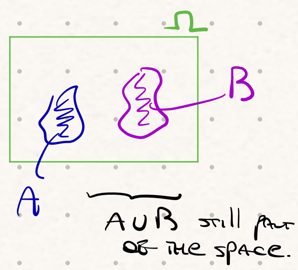
Definition: A collection $\mathcal{A}$ of subsets of a set Ω is an algebra if

(1) $\Omega \in \mathcal{A}$ (2) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ (3) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$

σ-algebra if $A_i \in \mathcal{A} \Rightarrow \bigcup_{j=1}^{\infty} A_i \in \mathcal{A}$ ← pairwise disjoint

we want to be in the sample space





Thus, the probability function can be defined by the following definition:

Definition: Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a σ-algebra. A map $\mu : \mathcal{A} \to \mathbb{R}$ is called a measure if

($\mathcal{P}(\Omega)$ ← power set)

(1) $\mu(\phi) = 0$

(2) $A_i \in \mathcal{A}$ pairwise disjoint $\Rightarrow \mu(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mu(A_j)$

Definition: A probability measure is a measurable map $\mathbb{P} : \mathcal{A} \longrightarrow [0,1]$ where $\mathbb{P}(\Omega) = 1$.

A probability space is a triade $(\Omega, \mathcal{A}, \mu)$ where Ω is the sample space, $\mathcal{A}$ a σ-algebra and μ a probability measure.

Lorenzo Zambelli
web: lorenzozambelli.it

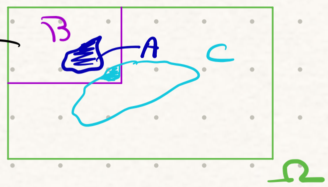
Tutorial 3

– Review & extra

Material:
- Conditional probability
- Low of total probability
- Bayers formula

Conditional probability: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space

subset $B \in \mathcal{A}$ with $\mathbb{P}(B) \neq 0$



$\Rightarrow$ new probability space $(B, \tilde{\mathcal{A}}, \mu)$ where $\mu(A) = \dfrac{\mathbb{P}(A)}{\mathbb{P}(B)}$

new because we do not care to "focus" on $B \Rightarrow$ only $A \in \mathcal{A}$ s.t. $A \subseteq B$

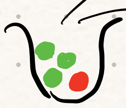
$\Rightarrow$ new probability space $(\Omega, \mathcal{A}, \mathbb{P}_B)$ where $\dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

Def: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space, $B \in \mathcal{A}$ with $\mathbb{P}(B) \neq 0$ then

$\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ is called conditional probability of A under B

$\mathbb{P}(\cdot | B) : \mathcal{A} \to [0,1]$ is called conditional probability measure given B

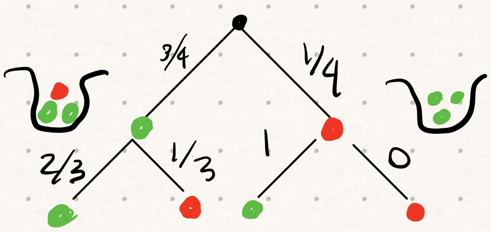
→ **Example**



possible sample $(\text{green}, \text{red})$

$G := \{g, r\}, \quad \Omega = G \times G, \quad \mathcal{A} = \mathcal{P}(\Omega)$

$\mathbb{P}$ given by probability mass function



$\Rightarrow \mathbb{P}(\{(g,g)\}) = \frac{3}{4} \cdot \frac{2}{3}$

$\mathbb{P}(\{(g,r)\}) = \frac{3}{4} \cdot \frac{1}{3}$

$\mathbb{P}(\{(r,g)\}) = \frac{1}{4} \cdot 1$

$\mathbb{P}(\{(r,r)\}) = \frac{1}{4} \cdot 0$

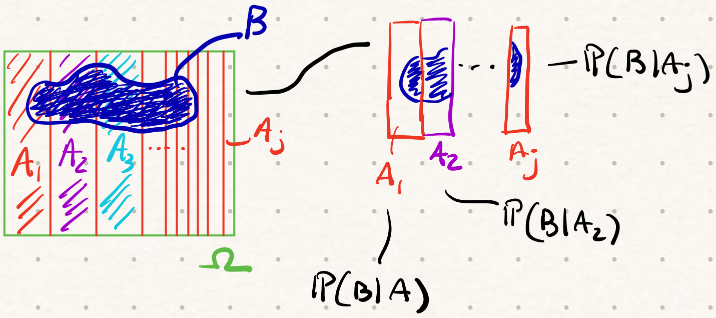
event: $B :=$ First ball Green $= \{(g,g), (g,r)\}$

$\mathbb{P}(\{(g,r)\} | B) = \dfrac{\mathbb{P}(\{(g,r)\})}{\mathbb{P}(B)} = 1/3$

Law of total probability

Def: Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space. Assume $B \in \mathcal{A}$ have $\mathbb{P}(B) > 0$ and $A_1, A_2, \dots$ are pairwise disjoint and $\bigcup A_j = \Omega$. Then,

$$\mathbb{P}(B) = \sum \mathbb{P}(B \mid A_j) \cdot \mathbb{P}(A_j)$$



measure of B w.r.t A_j

$\mathbb{P}(B \mid A_j) \cdot \mathbb{P}(A_j)$

measure of B w.r.t Ω

$$\Rightarrow \mathbb{P}(B) = \sum \mathbb{P}(B \mid A_j) \cdot \mathbb{P}(A_j)$$

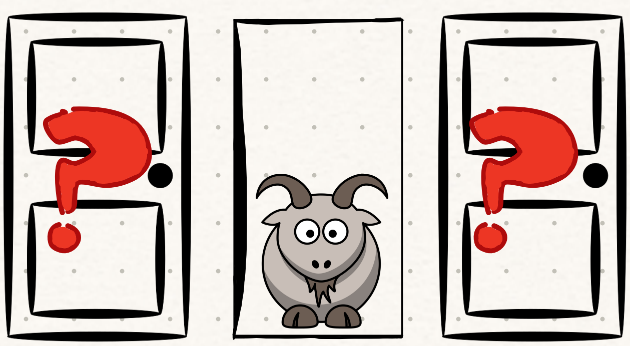
Bayes formula

Def: Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space. Assume $B \in \mathcal{A}$ have $\mathbb{P}(B) > 0$ and $A_1, A_2, \dots$ are pairwise disjoint and $\bigcup A_j = \Omega$. Then,

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A_i) \cdot \mathbb{P}(A_i)}{\sum \mathbb{P}(B \mid A_j) \cdot \mathbb{P}(A_j)}$$

Remark: really important in statistic. All field called Bayes inference and it is all based upon this formula.

Example: Monty Hall problem → 3 doors: 2 goats, 1 winning price



1) you choose a door
2) The conductor open a door with a goat, and ask you if you want to change.
3) Do you change?

→ Extra problems

Lorenzo Zambelli
web: lorenzozambelli.it

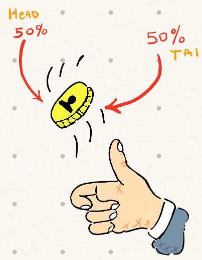
Week 2

-> review & extra

material:
- Random Variable
- Distribution
- CDF
- PMF : Discrete
- PDF : continuos
- Exp & Variance

Random Variable $X : \Omega \to \mathbb{R}$

Example: tossing a coin 3 times

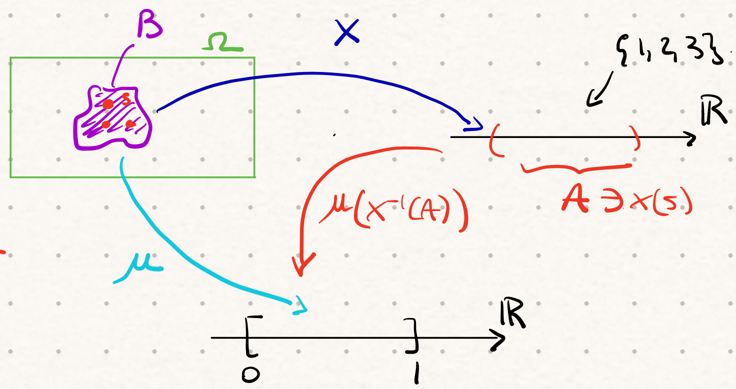


let $X := \#$ of heads obtain in 3 tossed

sample space

s	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
X(s)	3	2	2	2	1	1	1	0

The range of X is $\mathcal{R} := \{1, 2, 3\}$ insted of using a Ω of 8.



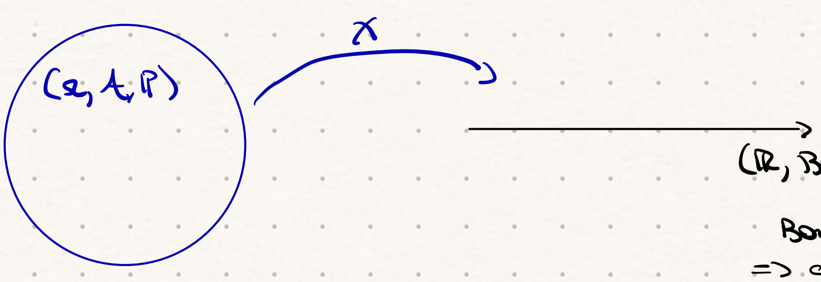
more generally a measurable space $(\tilde{\Omega}, \tilde{\mathcal{A}})$

$$\mu(\{s \in B : X(s) \in A\}) = \mu(X^{-1}(A))$$

DEF: Let $(\Omega, \mathcal{A})$ and $(\tilde{\Omega}, \tilde{\mathcal{A}})$ be measurable spaces. A map $X : \Omega \to \tilde{\Omega}$ is called a random variable if

$$X^{-1}(B) \in \mathcal{A} \quad \text{for all } A \in \tilde{\mathcal{A}}$$

Distributions



$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$

Borel σ-Algebra
=> collection of sets generated by open sets

DEF: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $X : \Omega \to \mathbb{R}$ be a random variable. Then $\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \to [0,1]$ defined by $\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B))$ is called a probability distribution of X.

Lorenzo Zambelli

web: lorenzozambelli.it

→ Exaple

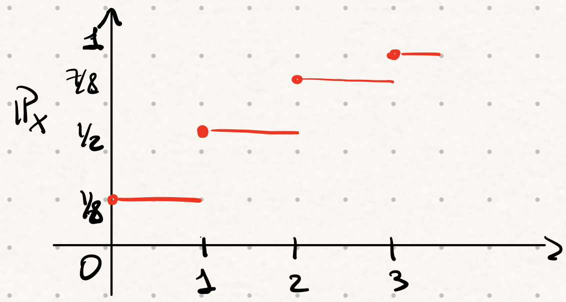
n tosses of the same coin $(\Omega, \mathcal{A}, \mathbb{P})$

$\Omega = \{0,1\}^n$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}(\{\omega\}) = p^{\#1s} \cdot (1-p)^{\#0s}$

$X : \Omega \to \mathbb{R}$, $X(\omega) :=$ number of 1s in ω $\Rightarrow X \sim Bin(n, p)$ ← Binomial distribution

If we going back and we toss 3 coins again we would get

$$\mathbb{P}_X(x) = \begin{cases} 0 & \text{if } -\infty < x < 0 \\ 1/8 & \text{if } 0 \le x < 1 \\ 1/2 & \text{if } 1 \le x < 2 \\ 7/8 & \text{if } 2 \le x < 3 \\ 1 & \text{if } 3 \le x < \infty \end{cases}$$

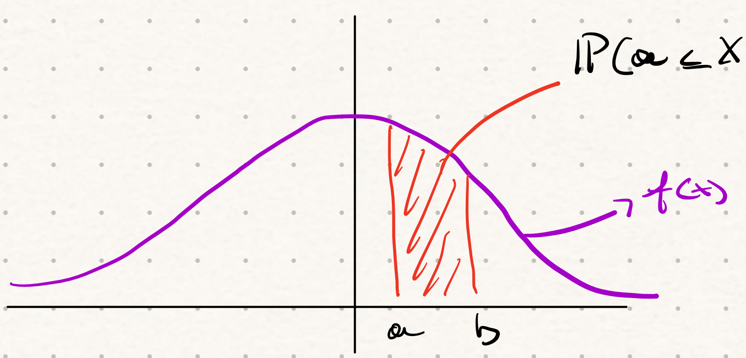


CDF

Discrete & continuos

Discrete	Continuos
- Finite many outcomes	- uncountably many outcomes
σ-Algebra $\mathcal{A} = \mathcal{P}(\Omega)$	$\mathcal{A} = \mathcal{B}(\Omega)$
prob. measure $\mathbb{P} : \mathcal{A} \to [0,1]$	$\mathbb{P} : \mathcal{A} \to [0,1]$
probability mass function $(p_\omega)_{\omega \in \Omega}$, $p_\omega \ge 0$, $\sum_{\omega \in \Omega} p_\omega = 1$	probability density $f : \Omega \to \mathbb{R}$ with $f(x) \ge 0$, $\int_\Omega f(x)\,dx = 1$
↳ $\mathbb{P}(A) := \sum_{\omega \in A} p_\omega$	↳ $\mathbb{P}(A) = \int_A f(x)\,dx$

→ Example normal curve



$$\mathbb{P}(a < X < b) = \int_a^b f(x)\,dx$$

$$= \mathbb{P}_X(b) - \mathbb{P}_X(a)$$

↑ CDF ↑ CDF

Lorenzo Zambelli

web: lorenzozambelli.it

Expected value & Variance

Let a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a random variable $X: \Omega \to \mathbb{R}$. Then we denote $\mathbb{E}(X) \in \mathbb{R}$ the expected value of X.

DEF: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space, $X: \Omega \to \mathbb{R}$ random variable

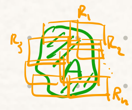
$$\mathbb{E}(X) := \int_\Omega X \, d\mathbb{P} \quad := \text{Lebesgue measure}$$

$\int_\Omega \mathbb{1}_A \, d\mathbb{P} = \mathbb{P}(A)$ ← indicator function

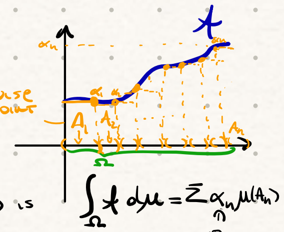
DEF: Let $A \subset \mathbb{R}^d$ then $\mathbb{P}(A) \in [0, \infty)$ is

$$\mathbb{P}(A) = \inf \left\{ \sum_{n=1}^{\infty} \ell(R_n) : R_n \subset \mathbb{R}^d \text{ closed rectangle}, \; A \subset \bigcup_{n=1}^{\infty} R_n \right\}$$

volume of R_n; they form a cover

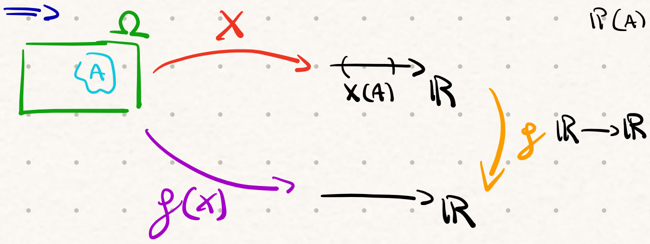


~ R_n cover of A and $\mathbb{P}(A)$ infimum of the total volumes of all possible covers R_n



$$\int_\Omega f \, d\mu = \sum_{n \in \mathbb{R}} \alpha_n \mu(A_n)$$

Change of variable ⟹



$x^{-1}(x) = \omega$

$$\int_A g(\underbrace{X(\omega)}_{x}) \, d\mathbb{P}(\omega) = \int_{X(A)} g(x) \, d\underbrace{(\mathbb{P} \circ X^{-1})}_{\text{density function}}(x)$$

only reals that we are computing the integral with respect to the measure $\mathbb{P}(\cdot)$

is equivalent to

Discrete

$$\sum_{x \in X(A)} g(x) \underbrace{p_X(x)}_{\text{PMF}}$$

Continuous

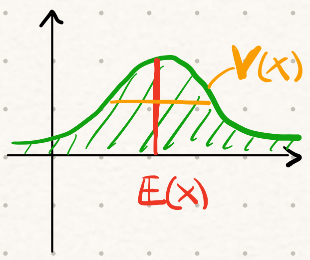
$$\int_{X(A)} g(x) \underbrace{f_X(x)}_{\text{PDF}} \, dx$$

~ **Note**

The $\mathbb{E}$ is a linear function,

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

Expected Value → Variance (minimize distance)



How spread we are?

~ The expected value of the distance between a quantity b (i.e., $\mathbb{E}(X)$) & the value of our random variable.

$$\mathrm{Var}(X) = \min_b \mathbb{E}(X - b)^2 = \mathbb{E}(X - \mathbb{E}(X))^2 = \dots = \boxed{\mathbb{E}(X^2) - \mathbb{E}(X)^2}$$

Note $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$

Just a traslation the distance remain unchange

$$\sim \mathbb{E}(aX + b - a\mathbb{E}(X) - b)^2$$

$\mathbb{E}$ linear

$$= a^2 \mathbb{E}(X - \mathbb{E}(X))^2$$
$$= a^2 \mathrm{Var}(X)$$

Lorenzo Zambelli
web: lorenzozambelli.it

Common Distributions
~ summary

$\mathbb{P}(X = x \mid p) = p^x (1-p)^{1-x}$

$X = \begin{cases} 1 & \text{success} \quad p \\ 0 & \text{failure} \quad 1-p \end{cases} \qquad 0 \le p \le 1$

only 2 outcomes 1 trial

n Trials

"What is $\mathbb{P}$ of getting x success in n trials?"

$\mathbb{P}(X = x \mid n, p) = \boxed{\binom{n}{x}} p^x (1-p)^{n-x}$

ways of ordering
x success out n

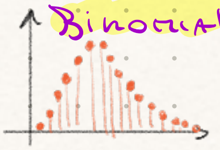
Binomial

Bernoulli

r Success

"how many trials do we need for r success?"

Negative Binomial



$y :=$ The Trials at which r_{th} success occurs

$\mathbb{P}(y = y \mid r, p) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$

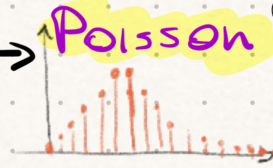
$r = 1$

Discrete

$\mathbb{P}(X = x \mid N) = 1/N$

Uniform

Poisson



$\mathbb{P}(X = x \mid \lambda) = \dfrac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, \dots$

The most used for waiting times
Like $X :=$ # calls in a minute
$\lambda := \mathbb{E}$ value of calls
As $t \to 1$ minute $\mathbb{P} \downarrow 0$
$t \in B(0)$ ~ almost uniform distribution i.e. same probability

Geometric

$\mathbb{P}(X = x \mid p) = p(1-p)^{x-1}$

$X :=$ Trial at which the first success occurs

Memoryless
$\mathbb{P}(X > s \mid X > t) = \mathbb{P}(X > s - t)$

Waiting for a success

Used to model "lifetimes", "Time until failure"

Lorenzo Zambelli

web: lorenzozambelli.it

Common Distributions

~ summary

Continuous

Normal

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

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

Normalize Area

Center at μ

symmetry

Properties

① $(x+b) \sim \mathcal{N}(\mu+b, \sigma^2)$ translation

② $a \cdot X \sim \mathcal{N}(ab, a^2\sigma^2)$

③ If n is big & p not close to 0 or 1 then $X \sim Bin(n,p) \approx Y \sim \mathcal{N}(np, np(1-p))$

Normal & Binomial

Beta

$$f(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad x \in (0,1)$$

$\alpha > 0, \beta > 0$

many shapes → Great to approx data

symmetric → more concentrated $\alpha \to \infty$

$\alpha = \beta = 1$ uniform distribution

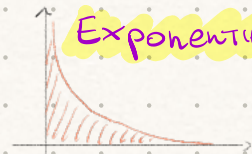
Beta function

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$$

Gamma function relationship

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Exponential



$$f(x|\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Exponential Families

memoryless

particular case of Gamma distribution

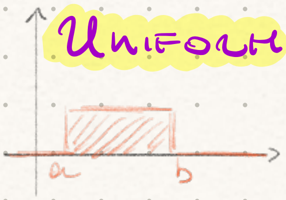
analogue to Geometric Distribution

A pdf/pmf is called an Exponential Family iff

$$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right)$$

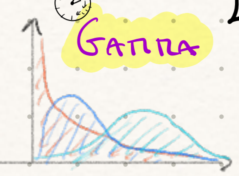
$h(x) \geq 0$, $c(\theta) \geq 0$, w_i & t_i realvalued

Uniform



$$f(x|a,b) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

Gamma



$\alpha = 1$

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha) \Rightarrow \Gamma(n) = (n-1)!$$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

Gamma function

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

$x \in (0, \infty)$, $\alpha, \beta > 0$

Shape Rate

memoryless

If $X \sim Gamma(\alpha, \beta)$

$$\mathbb{P}(X \leq x) = \mathbb{P}(Y \geq \alpha)$$

$$Y \sim Poisson(x/\beta)$$

Lorenzo Zambelli
web: lorenzozambelli.it

Moments & Moment Generating Functions

Def

For each integer n, the nth moment of X, μ'_n is
$$\mu'_n = \mathbb{E}(X^n)$$

The nth central moment of X, μ_n is $\mu_n = \mathbb{E}(X-\mu)^n$

Variance is the 2nd central moment

The moment generating function of X is defined by
$$M_X(t) = \mathbb{E}(e^{tX})$$

$$M_X(t) = \underbrace{\sum_x e^{tx} p_X(x)}_{\text{Discrete}} \qquad \underbrace{\int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx}_{\text{Continuous}}$$

weight average of Exp. function

Properties

$$M_X(0) = 1$$

$$M'_X(0) = \mathbb{E}(X)$$
$$\Downarrow$$
$$M^{(n)}_X(0) = \mathbb{E}(X^n)$$

$$e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \dots$$

$$\Rightarrow \mathbb{E}(e^{tX}) = 1 + t\,m_1 + \frac{t^2 m^2}{2!} + \dots$$

1st moment; 2nd moment

Independences random variables $X_1, \dots, X_n$ then
$$M_{X_1,\dots,X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

$$M_{aX}(t) = M_X(at)$$

$$M_{X+b}(t) = e^{bt} M_X(t)$$

Multivariate Distributions

Def

A n-dimensional random vector $X : \Omega \to \mathbb{R}^n$

Joint

CDF
$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) = \mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n)$$
$$\mathbb{P}(X \in A) = \int \cdots \int_A f_{X_1,\dots,X_n}(x_1,\dots,x_n)\,dx_1 \cdots dx_n$$

PDF
$$\mathbb{P}(X \in A) = \sum_A f_{X_1,\dots,X_n}(x_1,\dots,x_n)$$

Independence
$$f_{X_1,\dots,X_n}(x_1,\dots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

Bayes Theorem
$$f_{X_1,\dots,X_m | X_{m+1},\dots,X_n} = \frac{f_{X_1,\dots,X_n}}{f_{X_{m+1},\dots,X_n}}$$

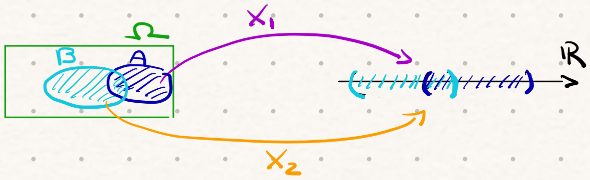
Joint mgf
$$M_{X_1,\dots,X_n}(t) = \begin{cases} \sum e^{\sum_{i}^{n} t_i x_i} f_X(x) \\ \int_{-\infty}^{\infty} e^{\sum_{i}^{n} t_i x_i} f_X(x)\,dx_1 \cdots dx_n \end{cases}$$

Lorenzo Zambelli
web: lorenzozambelli.it

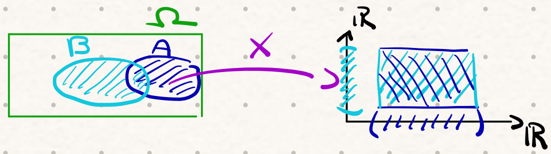
Random Vectors

A set of 2 or more random variables constitutes a **Random Vector**.

Ex: If $X_i : \Omega \longrightarrow \mathbb{R}$ $i = 1, 2$ is a random variable



Then,



$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : \Omega \longrightarrow \mathbb{R}^2$$

~ Characterization

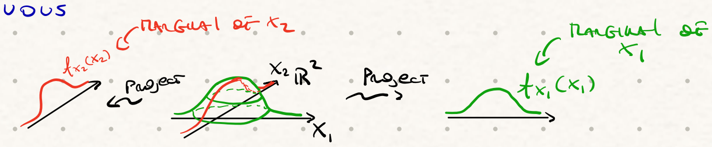
Same idea as 1-D !

	Continuous	Discrete
PDF/PMF	$f_{x_1,x_2}(x_1,x_2) = \frac{\partial^2 F_x(x)}{\partial x_1 \partial x_2}$	$\mathbb{P}(\{x_1 = x_1\} \cap \{x_2 = x_2\})$
CDF	$F_x(x) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_x(u)\, du_1\, du_2$	$\mathbb{P}(\{x_1 \le x_1\} \cap \{x_2 \le x_2\}) = \sum_{u_1 \le x_1} \sum_{u_2 \le u_2} f_{x_1,x_2}(u_1,u_2)$

~ Marginal distribution

Let $X : \Omega \longrightarrow \mathbb{R}^n$ be a random vector

Continuous



$$f_{x_1}(t) = \int_{\mathbb{R}^{n-1}} f_x(t, x_2, \dots, x_n)\, d(x_2, \dots, x_n)$$

Discrete

$$f_{x_1}(t) = \sum_{x_2, \dots, x_n \in \mathbb{R}} f_x(t, x_2, \dots, x_n)$$

~ Joint table

$f_{x_1}(b_i) = \sum_{i=1}^{n} f_x(b_i, a_i)$

$x_1 \backslash x_2$	a_1	a_2	----	a_n	f_{x_1}
b_1	$f_{x_1,x_2}(b_1,a_1)$				
⋮					
h_k					
f_{x_2}					1

$\underbrace{\qquad\qquad}_{\text{marginal of } x_2}$

$$\sum_{i=1}^{k} f_x(b_i, a_1) = f_{x_2}(a_1)$$

$\mathbb{P}(\Omega) = 1$

$$\sum_{(x_1,x_2) \in \mathbb{R}^2} f_{x_1,x_2}(x_1,x_2) = 1$$

~ Independance & Conditional

Ind: $f_{x_1,x_2}(x_1,x_2) = f_{x_1}(x_1) f_{x_2}(x_2)$ } Independance of events

Cond: $f_{x_1|x_2}(x_1|x_2) = \dfrac{f_{x_1,x_2}(x_1,x_2)}{f_{x_2}(x_2)}$ } Bayes Theorem

The multidimension case have the same idea of 1-D => We are still computing the "area" of a set $A \in \Omega$

Lorenzo Zambelli
web: lorenzozambelli.it

Covariance & Correlation

Def: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space $X, Y : \Omega \to \mathbb{R}$ random variables

$$Cov(X,Y) = \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y)))$$

$$= \underbrace{\mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X)}_{\text{"how close are the two variable from independance"}}$$

$\rightsquigarrow$ X, Y ind $\Rightarrow$ $Cov(X,Y) = 0$ ~ X, Y uncorrelated

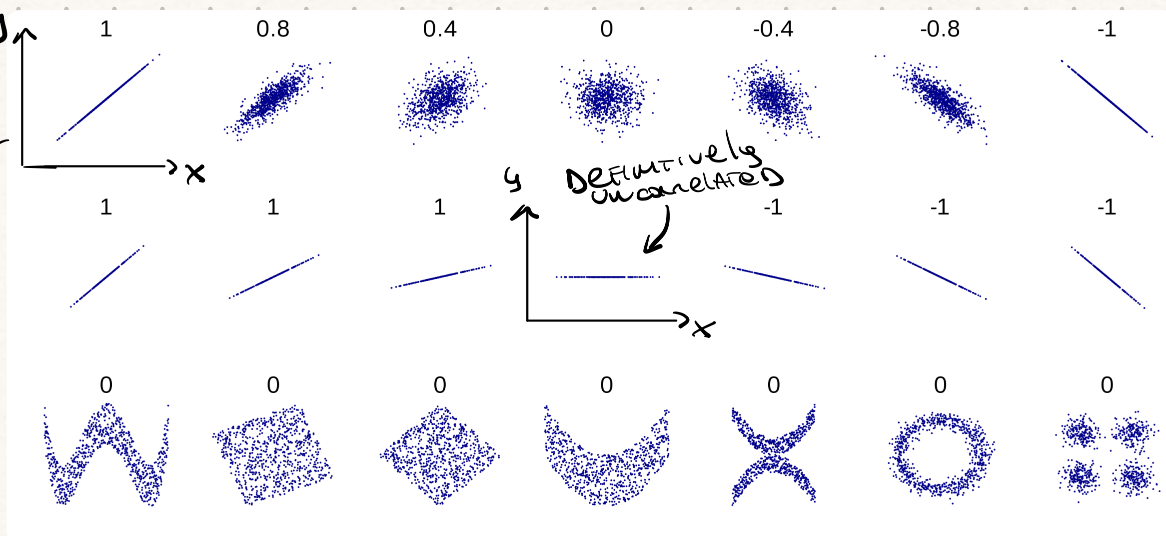
$\cancel{\Leftarrow}$ usually False

Def: Correlation coefficient $\rho_{X,Y} = \dfrac{Cov(X,Y)}{\sigma(X)\,\sigma(Y)} \in [-1, 1]$

$\rightsquigarrow$ If $\rho_{X,Y}$ closer to ± 1 $\Rightarrow$ correlated.

Ex: Consider the following sets (x, y) with different values of correlation

linear correlation



In application is often hard to find data that are completely uncorrelated, thus there exists methods to study "how strong" is the relationship. One of the methods uses the Bayes Theorem basically you construct an Graph called DAG through means of the BT.

$$X \longrightarrow Y$$

↳ Y depends on X