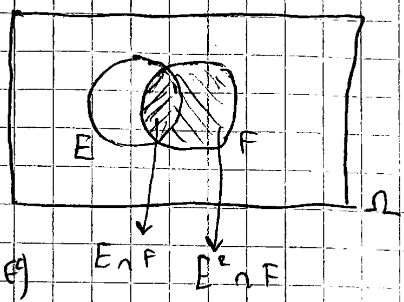
Total Probability

E, F events

$F = E \cap F \cup E^c \cap F$

$P(F) = P(E \cap F) + P(E^c \cap F)$

$\quad = P(E) \cdot P(F|E) + P(E^c) \cdot P(F|E^c)$



<u>e.g. 2 Fair Coins</u>

H_i - coin i is h

$\exists H$ - at least one h $\qquad P(\exists H) = \frac{|\exists H|}{|\Omega|} = \frac{3}{4}$ $\qquad \begin{cases} hh \\ ht \\ th \\ tt \end{cases}$

or $P(\exists H) = P(H_1 \cap \exists H) + P(H_1^c \cap \exists H)$

$\quad = P(H_1) \cdot P(\exists H | H_1) + P(H_1^c) \cdot P(\exists H | H_1^c)$

$\quad = 1/2 \cdot 1 + \frac{1}{2} \cdot 1/2 = \frac{3}{4}$

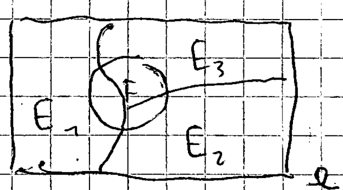
k - Conditions

Let $E_1, E_2, \dots E_n$ partition Ω

$F = \biguplus_{i=1}^{n} E_i \cap F$

$P(F) = \sum_{i=1}^{n} P(E_i \cap F)$

$\quad = \sum_{i=1}^{n} P(E_i) \cdot P(F | E_i)$



<u>e.g.: 2 Dice</u>

D_i - outcome of die i $\qquad S = D_1 + D_2$ - sum of 2 dice.

$P(S=5) = ?$

$P(S=5) = \sum_{i=1}^{4} P(D_1 = i) \cdot P(D_2 = 5-i \mid D_1 = i)$

$\quad = \sum_{i=1}^{4} P(D_1 = i) \cdot P(D_2 = 5-i)$ $\quad \rightarrow$ dice are independent

$\quad = 4 \cdot \frac{1}{36} = \frac{1}{9}$

<u>exercise:</u> 3 factories produce 50%, 30% and 20% of iPhones.
Their defective rates are 4%, 10% and 5% respectively.
What is the overall fraction of defective iPhones?

→

$$P(D) = P(F_1 \cap D) + P(F_2 \cap D) + P(F_3 \cap D)$$
$$= P(F_1)P(D|F_1) + P(F_2)P(D|F_2) + P(F_3)P(D|F_3)$$
$$= 0{,}5 \cdot 0{,}04 + 0{,}3 \cdot 0{,}1 + 0{,}2 \cdot 0{,}05$$
$$= 0{,}02 + 0{,}03 + 0{,}01$$
$$= \underline{\underline{0{,}06}}$$

Bayes' Rule

Asymmetry is quite frequent

e.g. If you're alive today, that implies you were born after 1800
But if you were born after 1800 that doesn't necessarily mean you are still alive

Foward - Backward

at times $P(F|E)$ - easy $P(E|F)$ - hard

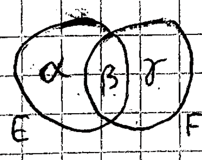
2 dice $P(S=5 \mid D_1 = 2) = P(D_2 = 3) = \frac{1}{6}$ $P(D_1 = 2 \mid S = 5) = ?$

Bayes' Rule is a method for converting $P(F|E)$ to $P(E|F)$

$$\boxed{P(E|F) = \frac{P(E) \cdot P(F|E)}{P(F)}}$$

proof: - $P(E|F) = \frac{P(E \cap F)}{P(F)} \underset{\uparrow}{=} \frac{P(E) \cdot P(F|E)}{P(F)}$
Product Rule

$$- P(F) \cdot P(E|F) = P(E \cap F) = P(E) \cdot P(F|E)$$



$$P(F|E) = \frac{\beta}{\alpha + \beta} \qquad P(E|F) = \frac{\beta}{\beta + \gamma}$$

$$P(E|F) = \frac{\beta}{\beta+\gamma} = \frac{\beta}{\cancel{\alpha+\beta}} \cdot \frac{\cancel{\alpha+\beta}}{\beta+\gamma} = \frac{P(F|E) \cdot P(E)}{P(F)}$$

e.g. 2 coins $P(H_1 \mid \exists H) = ?$ $\exists H$ - at least one H

$$P(H_1 \mid \exists H) = P(\exists H \mid H_1) \cdot \frac{P(H_1)}{P(\exists H)}$$
$$= 1 \cdot \frac{1/2}{3/4}$$
$$= \frac{2}{3}$$

$$\exists H \begin{cases} h & h \\ h & t \\ t & h \end{cases} \quad \left.\begin{matrix} h\ h \\ h\ t \end{matrix}\right\} H_1$$
t t

e.g.: 2 Dice $\quad D_i$ - outcome of die $i \quad S = D_1 + D_2$

$$P(D_1 = 2 \mid S = 5) = \frac{P(S=5 \mid D_1 = 2) \cdot P(D_1 = 2)}{P(S=5)} \overset{*}{=} \frac{1/6 \cdot 1/6}{1/9} = 1/4$$

$$P(S=5 \mid D_1 = 2) = P(D_2 = 3 \mid D_1 = 2) = P(D_2 = 3) = 1/6$$

$* \; P(D_1 = 2) = 1/6$

$* \; P(S = 5) = 1/9$

$$P(D_1 = 2 \mid S = 5) = \frac{|D_1 = 2 \cap S = 5|}{|S = 5|}$$

$$S=5 \begin{cases} 1 & 4 \\ 2 & 3 \\ 3 & 2 \\ 4 & 1 \end{cases} \qquad \left.\begin{matrix} 1 \; 4 \\ 2 \; 3 \end{matrix}\right\} D_1 = 2$$

eg: follow up on earlier iphone production

$$P(F_1 \mid D) = \frac{P(D \mid F_1) \cdot P(F_1)}{P(D)} = \frac{0{,}04 \cdot 0{,}5}{0{,}06} = \frac{1}{3}$$

$P(D \mid F_1) = 0{,}04$
$P(F_1) = 0{,}5$
$P(D) = 0{,}06$

↳ Given an iphone is defective, the probability for it to be produced at factory 1 is 1/3.

$$P(F_2 \mid D) = \frac{0{,}1 \cdot 0{,}3}{0{,}06} = 1/2$$

$$P(F_3 \mid D) = \frac{0{,}05 \cdot 0{,}2}{0{,}06} = 1/6$$

→ Conditional probabilities add to 1

→ Even though factory 1 produces more iphones than factory 2, a defective phone is less likely to be made from factory 1.

Exercise:

A fair coin with P(heads) = 0,5 and a biased coin with P(heads) = 0,75 are placed in an urn.
One of the two coins is picked at random and tossed twice.
Find the probability:
- of observing two heads:

Let F, B and T, be the events that the coin is Fair, Biased and we observe Two Heads respectively.

By the law of total probability $P(T) = P(F) \cdot P(T \mid F) + P(B) \cdot P(T \mid B)$

$$= \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \left(\frac{3}{4}\right)^2$$

$$= \frac{1}{8} + \frac{9}{32}$$

$$= \frac{13}{32}$$

- that the biased coin was picked if two heads are observed

By Bayes' rule:

$$P(B|T) = \frac{P(B) \cdot P(T|B)}{P(T)} = \frac{\frac{1}{2} \cdot \left(\frac{3}{4}\right)^2}{\frac{13}{32}} = \frac{9 \cdot \cancel{32}}{\cancel{32} \cdot 13} = \underline{\underline{\frac{9}{13}}}$$

Monty Hall Problem

Suppose you're on a game show, and you're given the choice of three doors. Behind one of the doors is your prize. The host knows what is behind each door. You pick a door, say door 1. The host reveals another door, say door 3, which doesn't have the prize. You are then asked whether you want to change your selection to door 2.
Is it to your advantage to switch your choice?

L = you lose   W = you win

prize

$\boxed{1}$ $\boxed{2}$ $\boxed{3}$

you don't switch:

$P(W) = \frac{1}{3}$

$P(L) = \frac{2}{3}$

always switch:

$P(W) = \frac{2}{3}$

→ By swapping you always get the "opposite" of your original choice and two thirds of the time your original choice will be a goat.

7. Random Variables, Expectation & Variance

A random variable is a variable whose possible values are numerical outcomes of a random experiment.

There are two types: (depends on sample space)

- Ω is finite $\{1,2,3\}$ $\{e, \pi\}$ or countably infinite
  → discrete numbers

- Ω is uncountably infinite $[0,2]$ $\mathbb{R}$
  → continuous numbers

and a combination $[0,2] \cup \{e, \pi\}$
→ mixed

eg: for discrete random variable: # Heads

3 fair coins $X \hat{=}$ # heads

x	Outcomes	$p(x)$
0	ttt	$1/8$
1	tth, tht, htt	$3/8$
2	thh, hth, hht	$3/8$
3	hhh	$1/8$

Probability Mass Function (pmf)

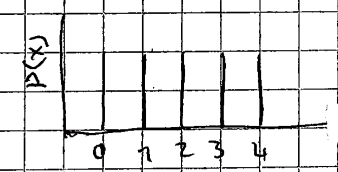
$p : \Omega \to \mathbb{R}$ specify Ω and p

Ω - random variable $\to \subseteq \mathbb{R}$

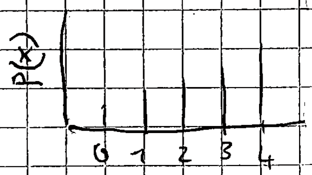
p - $p(x) \geq 0 \quad \forall x \in \Omega \quad \sum_{x \in \Omega} p(x) = 1$

Types of Discrete Distributions

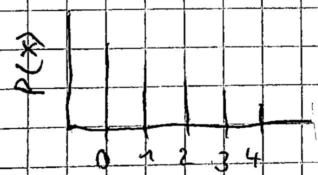
- finite $|\Omega| = n \in \mathbb{P}$
  - uniform: $p_1 = p_2 = \dots = p_n = \frac{1}{n}$



  - increasing: $p_1 \leq p_2 \leq \dots \leq p_n$



  - decreasing:



- infinite $|\Omega| = \infty$
  - one-sided infinite $p_1, p_2, p_3 \dots$
    - > cannot be uniform
      $p = 0 \to \sum = 0$
      $p > 0 \to \sum = \infty$
    - > cannot increase
      $p_i > 0 \to p_{i+1}, p_{i+2}, \dots > 0 \to \sum = \infty$
    - > can decrease
      $1/2, 1/4, 1/8, 1/16, \dots$
      $\sum_{i=1}^{n} \frac{1}{2^i} = 1 - \frac{1}{2^n}$
      $\sum_{i=1}^{\infty} \frac{1}{2^i} = 1$
  - doubly infinite $\dots, p_{-2}, p_{-1}, p_0, p_1, p_2, \dots$
    $\underbrace{\dots, 1/8, 1/4}_{\sum = 1/2}, 0, \underbrace{1/4, 1/8, \dots}_{\sum = 1/2}$

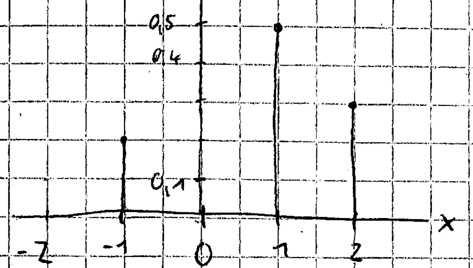
Cumulative Distribution Function (cdf)

↳ probability of intervals

$$F(x) = P(X \leq x) = \sum_{u \leq x} p(u)$$

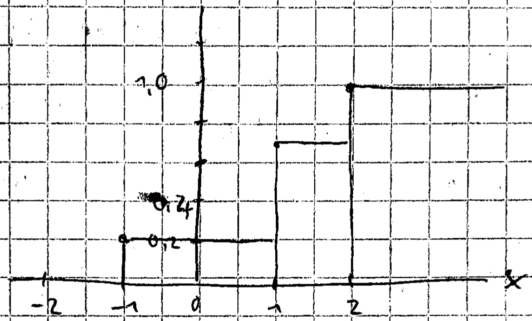
example:

PMF $p(x) = \begin{cases} 0{,}2 & -1 \\ 0{,}5 & 1 \\ 0{,}3 & 2 \end{cases}$



CDF $F(x) = P(X \leq x)$

$$= \sum_{u \leq x} p(u)$$



Notice that even though x is discrete, we define F of x for every real value of x

Properties

- nondecreasing $x \leq y \rightarrow F(x) \leq F(y)$
- limits $\lim_{x \to -\infty} F(x) = 0$

  $\lim_{x \to \infty} F(x) = 1$

- right-continuous $\lim_{x \searrow a} F(x) = F(a)$

Interval Probabilities

$P(X \leq a) = F(a)$

$P(X > a) = 1 - P(X \leq a) = 1 - F(a)$

$$P(a < X \leq b) = P((X \leq b) - (X \leq a))$$
$$= P(X \leq b) - P(X \leq a)$$
$$= F(b) - F(a)$$

Expectation (or "Mean")

The expectation of a random variable can be interpreted as the long-run average of many independent samples from the given distribution.

More precisely, it is defined as the probability-weighted sum of all possible values.

e.g. Fair Die

Roll a fair die $n \to \infty$ times
What is the average of the observed values?

- each value $\sim \frac{n}{6}$ times

$\to$ average: $\frac{\frac{n}{6}\cdot 1 + \frac{n}{6}\cdot 2 + \ldots + \frac{n}{6}\cdot 6}{n} = \frac{1+\ldots+6}{6}$

$= \frac{1}{6}\cdot\frac{(1+6)\cdot 6}{2} = \underline{3{,}5}$ *

In $n \to \infty$ samples x will appear $p(x)\cdot n$ times

$$\boxed{E(X) = \frac{\sum_x [p(x)\cdot n]\cdot x}{n} = \sum p(x)\cdot x}$$

also denoted μ

* for uniform variables $E(X)$ is the arithmetic average of elements in Ω:

X uniform over Ω $\to$ $p(x) = \frac{1}{|\Omega|}$

$$E(X) = \sum_{x\in\Omega} p(x)\cdot x = \sum_{x\in\Omega}\frac{1}{|\Omega|}\cdot x = \frac{1}{|\Omega|}\sum_{x\in\Omega} x$$

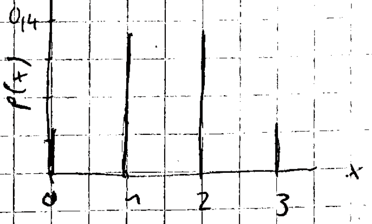
Symmetry

[A distribution p is symmetric around a if for all $x > 0$ $p(a+x) = p(a-x)$]

If p is symmetric around a, then $E(X) = a$

eg: 3 coins

x	outcomes	p(x)
0	ttt	1/8
1	tth, tht, htt	3/8
2	thh, hth, hht	3/8
3	hhh	1/8



Is the Expectation expected?
- not necessarily, we may never see it!

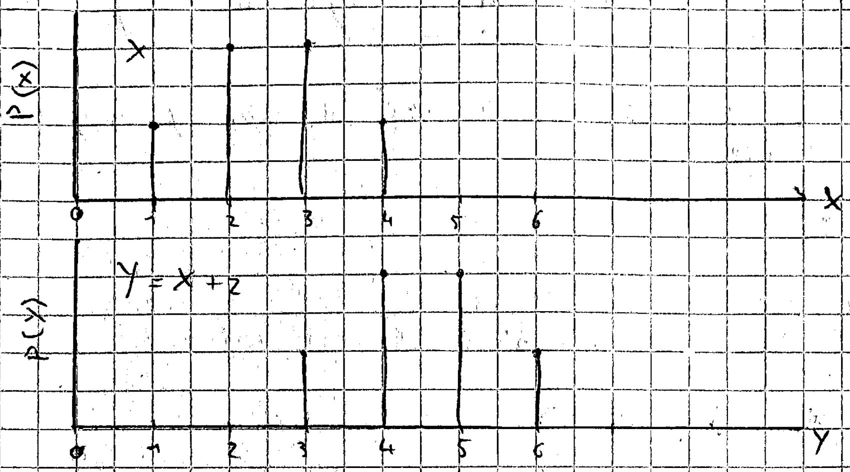
coin: $X \in \{0,1\}$ $p_0 = p_1 = 0{,}5$

$E(X) = 0\cdot p_0 + 1\cdot p_1 = \frac{1}{2}$

↯ $\frac{1}{2}$ will never happen (→ $E(x)$ doesn't have to be a number random var. can take)

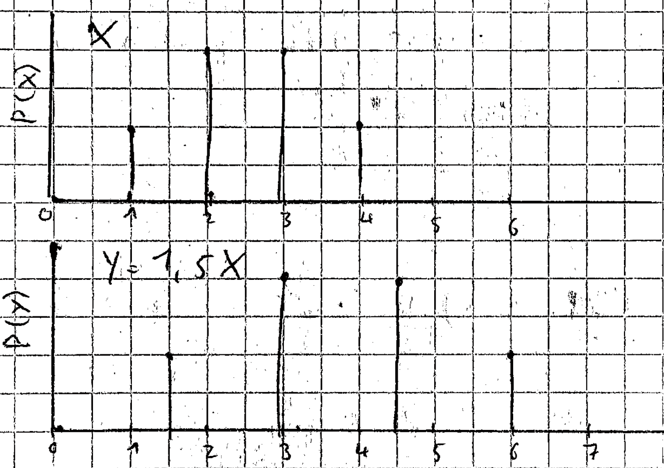
Variable Modifications:

- **translation**: add constant b to X



$P(Y=y) = P(X+b=y) = P(X=y-b)$

- **scaling**: multiply X by a constant b



$P(Y=y) = P(bX=y) = P(X=\frac{y}{b})$

These two examples mapped the same probabilities to the modified X, i.e. the functions were one to one

many to one example: square

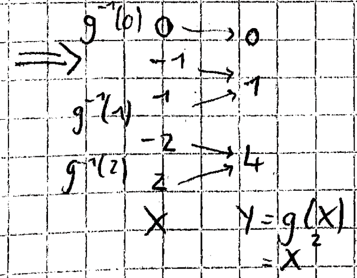
X:

x	-2	-1	0	1	2
$p(X=x)$	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$

$Y = X^2$:

y	0	1	4
$p(Y=y)$	$1/5$	$2/5$	$2/5$

$\Longrightarrow$



$p(Y=y) = P(g(X)=y) = p(X \in g^{-1}(y)) = \sum_{x \in g^{-1}(y)} P(X=x)$

Expectation of Modified Variables

$$E(Y) = \sum_{y} y \cdot P(Y=y)$$

$$= \sum_{y} y \cdot P(X \in g^{-1}(y))$$

$$= \sum_{y} y \sum_{x \in g^{-1}(y)} p(x)$$

$$= \sum_{y} \sum_{x \in g^{-1}(y)} y \cdot p(x)$$

$$= \sum_{y} \sum_{x \in g^{-1}(y)} g(x) \cdot p(x)$$

$$= \sum_{x} g(x) \cdot p(x)$$

X

x	-2	-1	0	1	2
$p(x)$	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$

Y

y	0	1	4
$p(y)$	$1/5$	$2/5$	$2/5$

$$E(Y) = \sum_{y=0,1,4} y \cdot p(Y=y) = 1/5 \cdot 0 + 2/5 \cdot 1 + 2/5 \cdot 4 = \frac{10}{5} = \underline{2}$$

$$E(Y) = \sum_{x} x^2 \cdot p(x)$$

$$= (-2)^2 \cdot 1/5 + (-1)^2 \cdot 1/5 + 0^2 \cdot 1/5 + 1^2 \cdot 1/5 + 2^2 \cdot 1/5$$

$$= \frac{4}{5} + 1/5 + 1/5 + \frac{4}{5} = \underline{2}$$

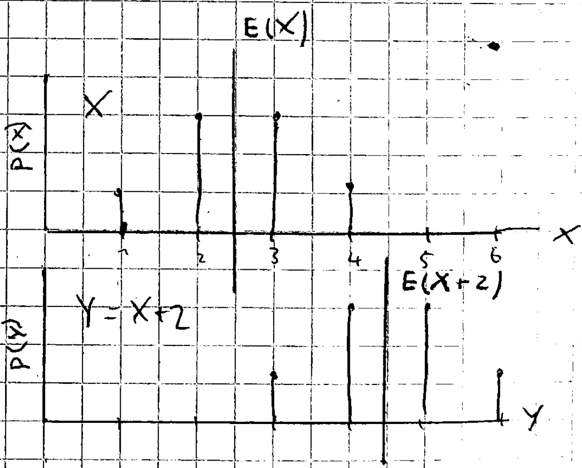
e.g.: Constant addition

$$E(X+b) = \sum p(x) \cdot (x+b)$$

$$= \sum p(x) \cdot x + \sum p(x) \cdot b$$

$$= E(X) + b \cdot \sum p(x)$$

$$= E(X) + b$$



Constant Multiplication: $\sum p(x) \cdot (ax)$
$= a \sum p(x) \cdot x$

Linearity of Expectation

$$E(aX+b) = E(aX) + b$$

$$= a\,E(X) + b$$

Variance

Whereas expectation provides a measure of centrality, the variance of a random variable quantifies the spread of that random variable's distribution.

It is the expected squared difference between X and its mean

constants

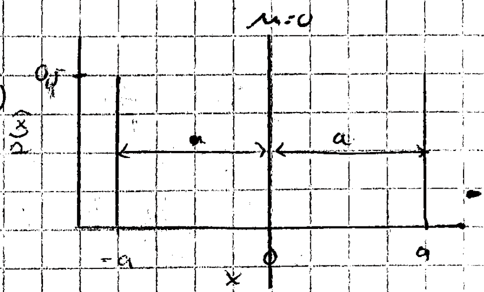
$$V(X) = E\left[(X-\mu)^2\right]$$

$\sigma_x = \sqrt{V(X)}$ ⟶ standard deviation (always positive root)

<u>examples:</u>

x	p_x	$x-\mu$	$(x-\mu)^2$
$-a$	$1/2$	$-a$	a^2
a	$1/2$	a	a^2

$\mu = 0$



$V(X) = 1/2 \cdot a^2 + 1/2 \cdot a^2 = a^2$

$\sigma_x = a$ – average distance from mean

– <u>fair die</u> $\mu = 3,5$

$$V(X) = E(X-\mu)^2 = \frac{2\,(6{,}25 + 2{,}25 + 0{,}25)}{6} = 2{,}92$$

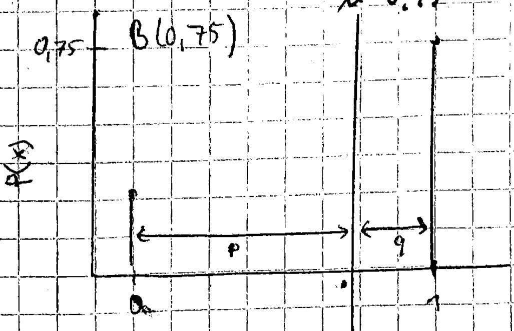
x	p_x	$x-\mu$	$(x-\mu)^2$
1	$1/6$	$-2,5$	$6,25$
2	$1/6$	$-1,5$	$2,25$
3	$1/6$	$-0,5$	$0,25$
4	$1/6$	$0,5$	$0,25$
5	$1/6$	$1,5$	$2,25$
6	$1/6$	$2,5$	$6,25$

$\sigma = \sqrt{2{,}92} = 1{,}71$

– <u>Bernoulli p</u>

x	p_x	$x-\mu$	$(x-\mu)^2$
0	q	$0-p=p$	p^2
1	p	$1-p=q$	q^2

$\mu = p$



$V(X) = q \cdot p^2 + p \cdot q^2 = pq(p+q) = pq$

Different formula for variance

$V(X) = E(X-\mu)^2$ $\qquad E(X) = \mu$

$= E(X^2 - 2\mu X + \cancel{E}\mu^2)$

$= E(X^2) - E(2\mu X) + E(\mu^2)$ constant

$= E(X^2) - 2\mu E(X) + \mu^2$

$= E(X^2) - 2\mu^2 + \mu^2$

$= E(X^2) - \mu^2$

Properties of Variance

addition: X - random variable b - constant

$\mu_{X+b} = \mu_X + b$

$V(X+b) = E[(X+b-\mu_{X+b})^2]$

$= E[(X+b-\mu_X-b)^2]$

$= E[X-\mu_X)^2$

$= V(X)$ $\rightarrow$ Variance stays the same

scaling: $V(aX) = E(aX-\mu_{aX})^2$ $\qquad \mu_{aX} = a\mu_X$

$= E(aX - a\mu_X)^2$

$= E[a^2(X-\mu_X)^2]$

$= a^2 E(X-\mu_X)^2$

$= a^2 V(X)$ $\rightarrow$ difference from mean grew by a^2

$\sigma_{aX} = \sqrt{V(aX)} = \sqrt{a^2 V(X)} = |a|\sigma_X$

$\rightarrow V(aX+b) = V(aX) = a^2 V(X)$

exercise: Let X and Y be independent random v. with expectations 1 and 2, and variances 3 and 4, respectively. Find the variance of $V(XY)$

$V(X) = E(X^2) - E(X)^2$

$E(X^2) = V(X) + E(X)^2 = 3 + 1^2 = 4$

$E(Y^2) = V(Y) + E(Y)^2 = 4 + 2^2 = 8$

$\rightarrow$

$$V(XY) = E((XY)^2) - E(XY)^2$$
$$= E(X^2Y^2) - (E(X)^2 \cdot E(Y)^2)$$
$$= E(X^2) \cdot E(Y^2) - E(X)^2 \cdot E(Y)^2$$
$$= 4 \cdot 8 - 1^2 \cdot 2^2$$
$$= 32 - 4$$
$$= \underline{\underline{28}}$$

Two Variables

→ experiments often result in multiple observations
- wheather: temperature and precipitation
- economy: unemployment and inflation

eg <u>Two Coins</u>

$U, V \sim B(1/2)$ Bernoulli   U and V are independent

u	v	$P(u,v)$
0	0	1/4
0	1	1/4
1	0	1/4
1	1	1/4

$P(u,v) \equiv P(U=u, V=v) = 1/4 \quad \forall \{u,v\} \in \{0,1\}$

$U \backslash V$	0	1
0	1/4	1/4
1	1/4	1/4

Min - Max

$U, V \sim B(1/2) \quad X = \min(U,V) \quad Y = \max(U,V)$

u	v	min	max	
0	0	0	0	1/4
0	1	0	1	1/4 } 1/2
1	0	0	1	1/4 } 1/2
1	1	1	1	1/4

X=min \ Y=max	0	1
0	$\frac{1}{4}$	$\frac{1}{2}$
1	0	1/4

min has to be smaller than max

Product - Sum

$X = U \cdot V \quad Y = U + V$

X \ Y	0	1	2
0	1/4	1/2	0
1	0	0	1/4

General B(p)

$U \sim B(p), \; V \sim B(p) \quad X = \min(U,V) \quad Y = \max(U,V)$

U \ V		
	$\bar{p}\bar{q}$	$\bar{p}q$
	$p\bar{q}$	pq

X=min \ Y=max	0	1
0	$\bar{p}\bar{q}$	$p\bar{q}+\bar{p}q$
1	0	pq

Joint Distribution

X, Y – random variables

It's the probability of every possible (x,y) pair

$$p(x,y) = P(X=x,\ Y=y), \quad \forall x,y \ p(x,y) \geq 0$$

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

eg:

$x \backslash y$	0	1
0	0,1	0,2
1	0,3	0,4

$$P(X \leq Y) = P(X=0, Y=0) + P(X=0, Y=1) + P(X=1, Y=1)$$
$$= P(0,0) + P(0,1) + P(1,1)$$
$$= 0,1 + 0,2 + 0,4$$
$$= 0,7$$

Marginals

Marginal of X $\quad P(x) \equiv P(X=x) = \sum_y p(x,y)$

Marginal of Y $\quad P(y) \equiv P(Y=y) = \sum_x p(x,y)$

$x \backslash y$	0	1
0	0,1	0,2
1	0,3	0,4

$$P(X=0) = P(X=0, Y=0) + P(X=0, Y=1)$$
$$= P(0,0) + P(0,1) = 0,1 + 0,2 = 0,3$$

Conditionals:

$$P(x \mid y) = \frac{p(x,y)}{p(y)} \qquad P(y \mid x) = \frac{p(x,y)}{p(x)}$$

$$P(Y=0 \mid X=0) = \frac{P(X=0, Y=0)}{P(X=0)} = \frac{0,1}{0,3} = \frac{1}{3}$$

$$P(Y=1 \mid X=0) = \frac{P(X=0, Y=1)}{P(X=0)} = \frac{0,2}{0,3} = \frac{2}{3}$$

$$P(X=0 \mid Y=0) = \frac{P(X=0, Y=0)}{P(Y=0)} = \frac{0,1}{0,4} = \frac{1}{4}$$

$$P(X=1 \mid Y=0) = 1 - P(X=0 \mid Y=0) = 1 - \frac{1}{4} = \frac{3}{4}$$

Independence:

X, Y independent $\quad X \perp\!\!\!\perp Y$

$\forall x, y \quad p(y \mid x) = p(y)$

$p(x \mid y) = p(x)$

$p(x, y) = p(x) \cdot p(y)$

eg:

		$P(y)$: 0,2	0,8
	$x \backslash y$		
$P(x)$	0,6	0,12	0,48
	0,4	0,08	0,32

Linearity of Expectation

$$E(X+Y) = \sum_x \sum_y (x+y) \cdot p(x,y)$$

$$= \sum_x \sum_y x \cdot p(x,y) + \sum_x \sum_y y \cdot p(x,y)$$

$$= \sum_x x \sum_y p(x,y) + \sum_y y \sum_x p(x,y)$$

$$= \sum_x x \cdot p(x) + \sum_y y \cdot p(y)$$

$$= EX + EY$$

$\rightarrow$ Expectation of sum = sum of expectations

Covariance

Do Expectations multiply?

$E(XY) = \sum_{x,y} xy \cdot p(x,y)$ $\qquad E(XY) \overset{?}{=} EX \cdot EY$

eg: $X = Y = \begin{cases} -1 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{cases}$

x \ y	-1	1
-1	1/2	0
1	0	1/2

$EX = EY = 0$

$\rightarrow EX \cdot EY = 0$

$E(XY) = EX^2 = E(1) = 1 \rightarrow E(XY) \neq EX \cdot EY$

! Expectations don't always multiply

„centralize" X & Y $\rightarrow$ 0-mean

$$\sigma_{X,Y} \triangleq Cov(X,Y) \triangleq E\left[(X-\mu_X)(Y-\mu_X)\right]$$

$$= E(XY) - E(X\mu_Y) - E(\mu_X Y) + E(\mu_X \mu_Y)$$

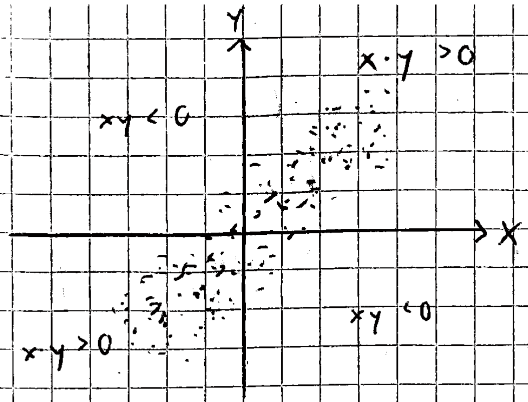
$$= E(XY) - E(X)\mu_Y - \mu_X E(Y) + \mu_X \mu_Y$$

$$= E(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y$$

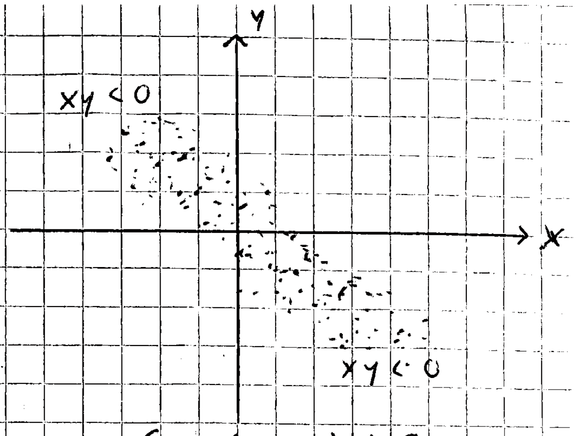
$$= E(XY) - \mu_X\mu_Y$$

Covariance is the degree to which random variables vary similarly

$$Cov(X,X) = E\left[(X-E(X))^2\right] = var(X)$$

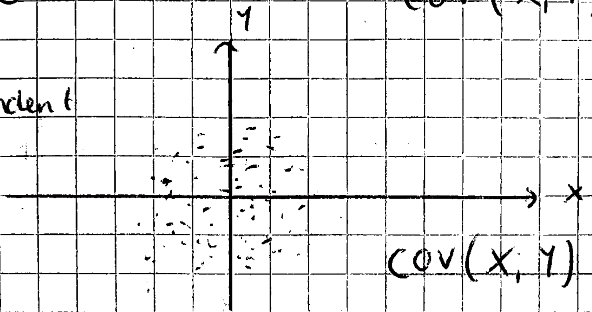


$Cov(X, Y) > 0$



$Cov(X, Y) < 0$

variables independent
$\Downarrow$
$Cov(X, Y) = 0$



$cov(X, Y) \approx 0$

The normalized covariance is called Correlation Coefficient:

$$\rightarrow \rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

properties:

$$\rho_{X,X} = 1 \qquad \rho_{X,-X} = -1$$

$$\rho_{X,Y} = \rho_{Y,X}$$

$$\rho_{aX+b,\, cY+d} = sign(ac) \cdot \rho_{X,Y} \qquad sign(x) = \begin{cases} 1 & x > 0 \\ 0 & 0 \\ -1 & x < 0 \end{cases}$$

real world example of correlation:

- A real-estate investment company invests \$10M in each of 10 states.
  At each state i, the return on its investment is a random variable X_i, with mean 1 and standard deviation 1,3 (in millions).

  If you look at one state in isolation, it would be a pretty risky investment because σ is comparable to the mean. It's not an unlikely event to have a return that's one standard deviation below the mean and if that happens your return would be negative and you would lose money.

  But then you argue that you're investing in ten different states, where you might lose money in some states but overall you would expect ending up with a positive return. with high confidence

? Is that correct?

variance of total return

$$var(X_1 + \dots + X_{10}) = \sum_{i=1}^{10} var(X_i) + \sum_{\{(i,j):\, i \neq j\}} cov(X_i, X_j)$$

- if the X_i are uncorrelated, then: ($\rightarrow$ state markets are independent)

$$var(X_1 + \dots + X_{10}) = 10 \cdot (1{,}3)^2 = 16{,}9$$

$$\sigma(X_1 + \dots + X_{10}) = 4{,}1$$

$$E(X_1 + \dots + X_{10}) = \underline{10}$$

$\rightarrow$ You would only lose money if the outcome is two and a half standard deviations below the mean, which is very unlikely, so you're very confident

- if for $i \neq j$ $\rho(X_i, X_j) = 0{,}9$ $\triangleq$ X_i are correlated

$$cov(X_i, X_j) = \rho\, \sigma_{X_i} \sigma_{X_j} = 0{,}9 \cdot 1{,}3 \cdot 1{,}3$$

$$= 1{,}52$$

$$\rightarrow var(X_1 + \dots + X_{10}) = 10 \cdot (1{,}3)^2 + \underset{\substack{\downarrow \\ ij \text{ permutations}}}{90} \cdot 1{,}52 = 154$$

$$\rightarrow \sigma(X_1 + \dots + X_{10}) = 12{,}4$$

$\rightarrow$ If you happen to be one standard deviation below the expectation which isn't unlikely your return will be negative $\rightarrow$ risky

During the financial crisis investment companies thought that they were secure by diversifying and investing in different housing markets in different states. But it turned out that there were high correlations during markets, hence large losses occured

8. Discrete Distribution Families

Bernoulli

B_p $\quad 0 \leq p \leq 1$

two values	0 (failure)	1 (success)
probability	$1-p$ or q	p

$p(0) + p(1)$

$= (1-p) + p = \underline{1}$ ✓

$$P(X=1) = p = 1 - P(X=0) = 1 - q$$

The Bernoulli distribution can be thought of as a model for the set of possible outcomes of any single experiment that asks a yes-no question, where the outcomes are boolean-valued.

e.g. products: 80 good, 20 defective $\rightarrow \sim B_{0,8}$

next child will be a boy $\rightarrow \sim B_{0,5}$

Mean

$X \sim B_p \quad p(0) = 1-p \quad p(1) = p$

$E(X) = \sum p(x) \cdot x = (1-p) \cdot 0 + p \cdot 1 = p \rightarrow EX = P(X=1)$

$X \sim B_{0,8} \rightarrow E(X) = 0{,}8$

→ fraction of times outcome is $x=1$

Variance

because $0^2 = 0$ and $1^2 = 1 \rightarrow x^2 = x \rightarrow E(x^2) = E(x) = p$

$\rightarrow V(X) = E(X^2) - (EX)^2 = p - p^2 = p(1-p) = \underline{\underline{pq}}$

Standard Deviation

$\sigma = \sqrt{pq}$

p	EX	$V(X)$	σ
0	0	0	0
1	1	0	0
1/2	1/2	1/4	1/2

Independent Trials

- much of B_p importance stems from multiple trials

$0 \leq p \leq 1 \quad X_1, X_2, X_3 \sim B_p \perp\!\!\!\perp$

$q = 1-p \quad P(110) = p^2 q = P(101) = P(011)$

→ Generally: $X_1, X_2, \ldots, X_n \sim B_p \perp\!\!\!\perp$

$x^n = x_1, x_2, \ldots, x_n \in \{0,1\}^n \quad n_0$ 0's and n_1 1's

$P(x_1, \ldots, x_n) = p^{n_1} q^{n_0} \quad P(10101) = p^3 q^2$

Typical samples

distribution	typical seq.	description	probability
B_0	0000000000	constant 0	$1^{10} = 1$
B_1	1111111111	constant 1	$1^{10} = 1$
$B_{0,8}$	1110111011	80% 1's	$0{,}8^8 \cdot 0{,}2^2$
$B_{0,5}$	1011010010	50% 1's	$0{,}5^{10}$

Binomial Distribution

The binomial distribution concerns several Bernoulli trials
It counts the number of successes in n Bernoulli trials

n independent Bernoulli experiments $\quad p$ - success prob

$\rightarrow B_{n,p}$ is the distribution of # successes

for small n

$b_{p,n}(k)$ - probability of k successes

$n=0$

k	$b_{p,0}(k)$
0	1

$n=1$

k	$b_{p,1}(k)$
0	q
1	p

$p+q=1$

$n=2$

k	seq's	$b_{p,2}(k)$
0	00	q^2
1	01, 10	$2pq$
2	11	p^2

$$p^2+2pq+q^2=(p+q)^2=1^2=1$$

general n and k

$n \perp\!\!\!\perp B_p$ experiments $\quad$ # successes $0 \le k \le n$

every k-success sequence: $n-k$ failures, probability $p^k \cdot q^{n-k}$

$\binom{n}{k}$ such sequences

$$\rightarrow b_{p,n}(k) = \binom{n}{k} p^k q^{n-k}$$

Will it add to 1?

$$\sum_{k=0}^{n} b_{p,n}(k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

binomial theorem

$$= (p+q)^n$$

$$= 1^n = 1 \checkmark$$

e.g.: Multiple Choice

Exam has 6 multiple-choice questions, each with 4 possible answers
For each question, student selects one of the 4 answers randomly

X = # correct answers $\sim B_{1/4,6}$

Passing: ≥ 4 correct answers

$P(4) = \binom{6}{4} \frac{1}{4}^4 \frac{3}{4}^2 \approx 0{,}0329 \qquad P(6) = \binom{6}{6} \frac{1}{4}^6 \cdot \frac{3}{4}^0 = 0{,}000244$

$P(5) = \binom{6}{4} \frac{1}{4}^5 \frac{3}{4}^1 \approx 0{,}00439 \qquad P(\ge 4) = P(4)+P(5)+P(6) \approx 3{,}8\%$

Interpretation of the binomial distribution as the sum of n Bernoulli p random variables.

$\rightarrow B_{p,n}$ a sum of n B_p

$X_1, \dots, X_n \sim B_p \perp\!\!\!\perp \qquad X = \sum_{i=1}^{n} X_i$

$P(X=k) = P(\text{exactly } k \text{ of } X_1, \dots, X_n \text{ are } 1)$

$= \binom{n}{k} p^k q^{n-k} = b_{p,n}(k)$

Mean and Variance

$X \sim B_{p,n} \quad X = \sum_{i=1}^{n} X_i$

$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum E X_i = \sum p = np$

$V(X) = V\left(\sum_{i=1}^{n} X_i\right) \overset{\perp\!\!\!\perp}{=} \sum V(X_i) = \sum pq = npq$

$\sigma = \sqrt{npq}$

e.g. earlier multiple choice

$E X = np = 6 \cdot 1/4 = 1{,}5$

$\sigma = \sqrt{npq} = \sqrt{6 \cdot \frac{1}{4} \cdot \frac{3}{4}} = \frac{\sqrt{18}}{4}$

exercise: election; What is the probability that the voter makes a difference?

for simplicity odd # voters: $2n+1$
and Democrats vs. Republican equally likely

$P(\text{voter makes a difference}) = P(\text{other } 2n \text{ voters equally split})$

$b_{p,n}(k) = \binom{n}{k} p^k q^{n-k}$

$$= \binom{2n}{n} \underbrace{\left(\tfrac{1}{2}\right)^n}_{n \text{ will vote D}} \cdot \underbrace{\left(\tfrac{1}{2}\right)^n}_{n \text{ will vote R}}$$

Stirling Approximation

$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

$$= \frac{(2n)!}{n! \cdot n! \cdot 2^n \cdot 2^n}$$

$$\approx \frac{\sqrt{2\pi \cdot 2n}\left(\frac{2n}{e}\right)^{2n}}{\left(\sqrt{2\pi n}\left(\frac{n}{e}\right)^n\right)^2 2^{2n}}$$

$$= \frac{1}{\sqrt{\pi n}}$$

exercise: Alice solves every puzzle with probability 0,6, and Bob with probability 0,5. They are given 7 puzzles and each chooses 5 out of the 7 puzzles randomly and solves them independently. A puzzle is considered solved if at least one of them solves it. What is the probability that all the 7 puzzles are solved?

The probability that all the 7 puzzles are chosen is the probability that Bob chooses the two puzzles Alice did not pick, namely

$$\frac{\binom{5}{3}\binom{2}{2}}{\binom{7}{5}} = \frac{10}{21}$$

The probability that they both fail the same puzzle is $0,4 \cdot 0,5 = 0,2$, hence at least one solves it with probability $1-0,2 = 0,8$

It follows that all puzzles are solved with probability

$$\frac{10}{21} \cdot 0,6^2 \cdot 0,5^2 \cdot 0,8^3 = 0,0219.$$

Poisson distribution

- extension of the binomial distribution

also called distribution of rare events

parameter $\lambda \geq 0$, support $\mathbb{N}$

PMF: $P_\lambda(k) = e^{-\lambda} \frac{\lambda^k}{k!}$

It approximates $B_{p,n}$ for large n and small p so that $np = \lambda$ is moderate

mean

daily applications:
- people clicking ads
- responses to spam
- rare disease infections

for small k:

λ	$P_\lambda(k)$	k: 0	1	2	3
general	$e^{-\lambda}\frac{\lambda^k}{k!}$	$\frac{1}{e^\lambda}$	$\frac{\lambda}{e^\lambda}$	$\frac{\lambda^2}{2e^\lambda}$	$\frac{\lambda^3}{6e^\lambda}$
1	$\frac{1}{ek!}$	$\frac{1}{e}$	$\frac{1}{e}$	$\frac{1}{2e}$	$\frac{1}{6e}$
2	$\frac{2^k}{e^2k!}$	$\frac{1}{e^2}$	$\frac{2}{e^2}$	$\frac{2}{e^2}$	$\frac{4}{3e^2}$
0	$\frac{0^k}{k!}$	1	0	0	0

The Poisson distribution is predominately used to predict the probability of events that will occur based on how often the event had happened in the past

Binomial Approximation - let's derive the Poisson formula!

$$B_{p,n}(k) = \binom{n}{k} p^k q^{n-k} \qquad q = 1-p, \quad p = \frac{\lambda}{n}$$

$$= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n^k}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{\overbrace{\left(1-\frac{\lambda}{n}\right)^n}^{\to e^{-\lambda}}}{\underbrace{\left(1-\frac{\lambda}{n}\right)^k}_{\to 1}} \longrightarrow e^{-\lambda}\frac{\lambda^k}{k!} \quad ✓$$

fix k and λ while $n \to \infty$

Limit of Binomial

1 $\quad \frac{n^k}{n^k} = \underbrace{\frac{n}{n}}_{\to 1} \cdot \underbrace{\frac{(n-1)}{n}}_{\to 1} \cdots \underbrace{\frac{(n-k+1)}{n}}_{\to 1} \to 1$

2 $\quad \left(1-\frac{\lambda}{n}\right)^k \to 1 \qquad$ fixed # (k) terms, each $\to 1$

3 $\quad \left(1-\frac{\lambda}{n}\right)^n = \left(\left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}}\right)^\lambda \to \left(e^{-1}\right)^\lambda \qquad$ note: $\left(1-\frac{1}{m}\right)^m \to e^{-1}$

↳ increasing # terms, each $\to 1$

* (next page)

Is Poisson really a distribution, i.e. will it add to 1?

Taylor expansion: $\quad e^\lambda = \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}$

$$\sum_{k=0}^{\infty} P_\lambda(k) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = e^{-\lambda}e^{\lambda} = 1 \quad ✓$$

Mean and Variance:

P_λ approximates $B_{p,n}$ for $\lambda = np$ when $n \gg 1 \gg p$

	μ	V
$B_{p,n}$	np	npq
P_λ	λ	λ

$q = (1-p) \to 1$

Approximation Example:

Factory produces 200 items, each defective with probability 1% (n quite large, p quite small), $P(3 \text{ defective})$?

precise: $\quad B_{0,01,\,200}(3) = \binom{200}{3}(0{,}01)^3(0{,}99)^{197} \approx 0{,}181$

approximation: Poisson $\quad \lambda = 200 \cdot 0{,}01 = 2 \qquad P_2(3) = e^{-2}\frac{2^3}{3!} \approx 0{,}18 \quad ✓$

→

P(some defective)?

- $B_{0{,}01,\,200}(0) = \binom{200}{0}(0{,}99)^{200} \approx 0{,}134$

  $B_{0{,}01,\,200}(\geq 1) = 1 - 0{,}134 \approx \underline{0{,}866}$

- $P_2(0) = e^{-2}\frac{1}{0!} = e^{-2} \approx 0{,}135$

  $P_2(\geq 1) = 1 - 0{,}135 \approx \underline{0{,}865}$ ✓

Exercise:

1) A vendor sells merchandise through Amazon and Ebay. On Ebay she sells an average of 2 items per day, while on Amazon the daily average is 3. The sales are independent of each other. What is the probability that she sells 5 items on a given day?

$\lambda = \lambda_1 + \lambda_2 = 2 + 3 = 5$

$P_5(5) = \frac{e^{-5} \cdot 5^5}{5!} \approx 0{,}175$

* Intuition behind the transition between the binomial distribution and the poisson distribution:

You choose to calculate the average number of cars that pass by a certain spot per hour by conducting one trial per minute on 60 occasions. But, let's suppose the road is extremely busy, and in reality, at least ten cars pass per minute. According to the Binomial model $E(X) = 60$
in reality $E(X) > 600$
$\longrightarrow$ binomial model is off

The solution is to make your trials more frequent (i.e. increase n). Eventually, in the limit, the approximation you get defines a different model - namely, the Poisson

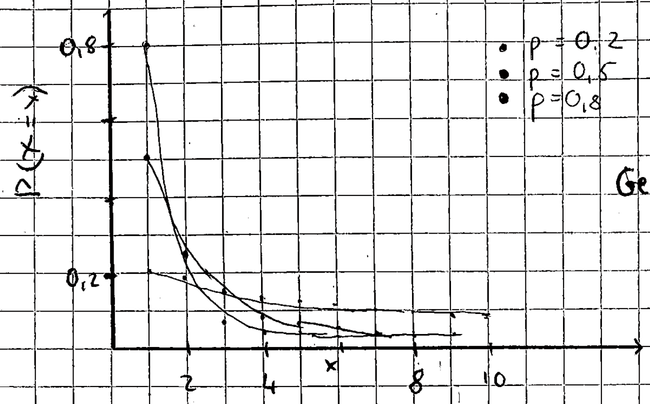
Geometric Distribution

coinflip: G_p # flips till first 1

flips	X
10101	1
01011	2
00011	4

n	$X_1, \dots, X_n$	p(n)
1	$X_1 = 1$	p
2	$X_1 = 0 \quad X_2 = 1$	qp
3	$X_1 = X_2 = 0 \quad X_3 = 1$	$q^2 p$
n	$X_1 = \dots = X_{n-1} = 0 \quad X_n = 1$	$q^{n-1} p$

$g_p(n) = q^{n-1} p$ $\quad < p \leq 1$, $p \neq 0$ $\quad n \geq 1$, n arbitrary high



Geometric $P(X=n) = P($first success at n'th trial$)$

Will it add to 1?

$P(n) = pq^{n-1}$ $\quad q = 1-p$

$(1 + q + q^2 + \dots)(1-q) = 1 + \cancel{q} + q^2 + \dots - \cancel{q} - q^2 - \dots$

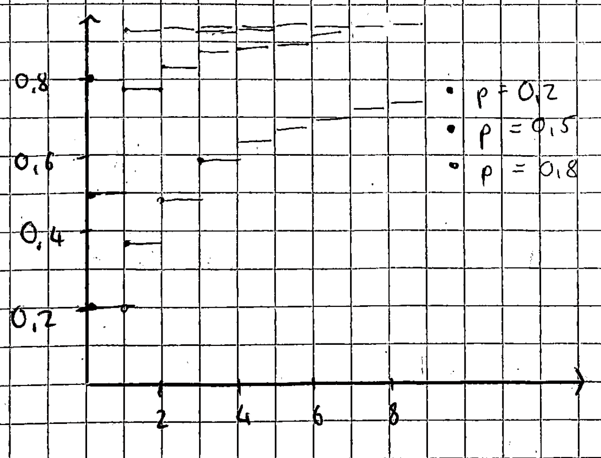
$$\sum_{i=0}^{\infty} q^i = \frac{1}{1-q}$$

$$\sum_{n=1}^{\infty} p(1-p)^{n-1} = p \sum_{i=0}^{\infty} (1-p)^i = p \cdot \frac{1}{1-(1-p)} \quad \frac{p}{p} = 1$$

CDF: $n \in \mathbb{N}$

$P(X > n) = P(X_1 = \dots = X_n = 0) = q^n$

$P(X \leq n) = 1 - P(X > n) = 1 - q^n$



Expectation via "Right" CDF

$$E(X) = \sum_{k=0}^{\infty} k\, p_k$$

$$= p_1 + 2p_2 + 3p_3 + \dots \longrightarrow \begin{array}{cccc} p_1 & & & \\ p_2 & p_2 & & \\ p_3 & p_3 & p_3 & \\ p_4 & p_4 & p_4 & p_4 \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

$$= P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \dots$$

Geometric distribution

$$\rightarrow E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{i=0}^{\infty} P(X > i) = \sum_{i=0}^{\infty} q^i = \frac{1}{1-q} = \frac{1}{p}$$

Variance

$$EX(X-1) = \sum_{n=1}^{\infty} n(n-1) \cdot P(X=n) = p \sum_{n=2}^{\infty} n(n-1)\, q^{n-1}$$

$$= pq \sum_{n=2}^{\infty} \frac{d^2}{dq^2} q^n = pq \frac{d^2}{dq^2} \sum_{n=2}^{\infty} q^n$$

$$\left(\frac{1}{1-q}\right)' = \frac{1}{(1-q)^2}$$

$$= pq \frac{d^2}{dq^2} \left(\frac{1}{1-q} - 1 - q\right)$$

$$\left(\frac{1}{(1-q)^2}\right)' = \frac{2}{(1-q)^3}$$

$$= pq \frac{2}{(1-q)^3} = \frac{2q}{p^2}$$

$$EX^2 = EX(X-1) + EX = \frac{2q}{p^2} + \frac{1}{p} = \frac{2q+p}{p^2} = \frac{1+q}{p^2}$$

$$V(X) = EX^2 - (EX)^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$$

$$\sigma = \frac{\sqrt{q}}{p}$$

<u>example: fair coin</u>

$$X \sim G_{1/2}$$

$$P(X=k) = g_{0,5}(k) = \left(\frac{1}{2}\right)^{k-1} \cdot \frac{1}{2} = \frac{1}{2^k}$$

$$EX = \frac{1}{p} = \underline{\underline{2}}$$

$$VX = \frac{q}{p^2} = 2$$

<u>Memoryless</u>

A distribution over $\mathbb{P} = \{1, 2, \dots\}$ is memoryless if for all $n \geq 0$, $m \geq 1$

$$P(X = n+m \mid X > n) = P(X = m)$$

$$\rightarrow P(X = 12 \mid X > 10) = P(X = 2)$$

Geometric distribution $\rightarrow$ memoryless AND memoryless $\rightarrow$ geometric distribution r Successes

→ What is the probability that the rth success, not the first, but maybe the second success or the third success is at the n^{th} trial?

$$P(r^{th} \text{ sucess at } n\text{'th trial}) = P(r-1 \text{ successes at } n-1 \text{ trials}) \cdot P(n^{th} \text{ trial is success})$$

$n \geq r$

$$b_{n-1,p}(r-1) \qquad p$$

$$= \binom{n-1}{r-1} p^{r-1} q^{n-r} p$$

$$= \binom{n-1}{r-1} p^{r} q^{n-r}$$

$r = 1 \rightarrow pq^{n-1} = g_p(n)$

geometric

<u>example:</u> Startup Statistics

- P(startup success) = 20%, independent of previous attempts

Expected number of startups till first success?

$$X \sim G_{0,2} \qquad E(X) = \frac{1}{p} = \frac{1}{0,2} = 5$$

- as the entrepreneur what is your expected fraction of the company you get to keep?

X - time to first success $\quad p = 0,2$
r^X - fraction of company you keep $\quad r = 0,5$

$$E(r^X) = \sum_{k=1}^{\infty} r^k P(X=k) = \sum_{k=1}^{\infty} pq^{k-1} r^k = pr \sum_{i=0}^{\infty} (qr)^i$$

$$= \frac{pr}{1-qr} = \frac{0,2 \cdot 0,5}{1 - 0,8 \cdot 0,5} = \frac{0,1}{1-0,4} = \frac{0,1}{0,6} \approx \underline{16,67\%}$$

9 Continuous Distributions

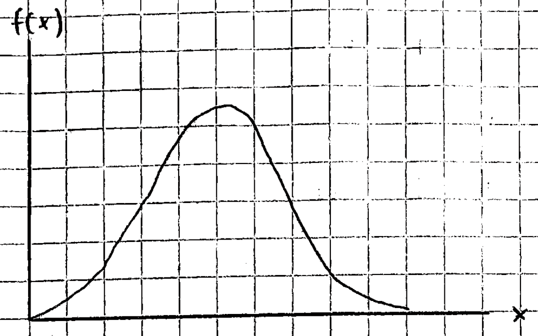
for uncountable # values, intervals

eg. anything in physics
- time (flight, delivery)
- space (height, storm area)
- mass
- temperature

"nearly" continuous
- cost (stock)
- rates (interest, exchange)

Instead of pmf we use probability density function →

$f(x) \geq 0$ relative likelihood of x



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

(area under the curve)

Event Probability

	Discrete	Continuous
$P(A)$	$\sum_{x \in A} p(x)$	$\int_{x \in A} f(x)\,dx$

! typically we are interested in interval probability $P(a \leq X \leq b)$

→ $P(X \leq b) - P(X \leq a)$

→ cumulative distribution function

Cumulative Distribution Function

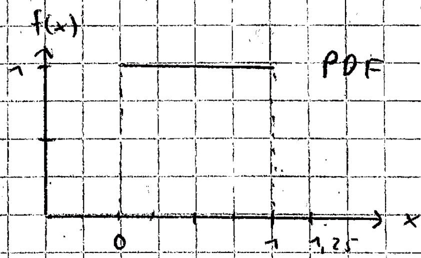
$$F(x) := P(X \leq x)$$

	Discrete	Continuous	
PF → CDF	$\sum_{u \leq x} p(u)$	$\int_{-\infty}^{x} f(u)\,du$	
CDF → PF	$p(x) = F(x) - F(x^*)$	$f(x) = F'(x)$	*

x^* – element preceding x

Uniform

$0 \leq x \leq 1$ $\quad f(x) = \begin{cases} 1 \\ 0 \end{cases}$

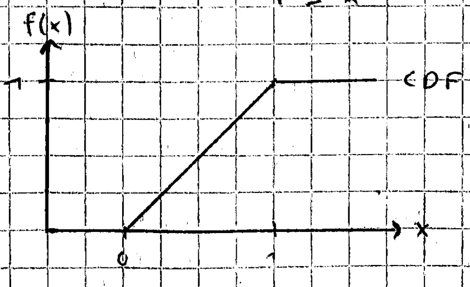


will it $\sum$?

area under the curve: $1 \cdot 1 = 1$

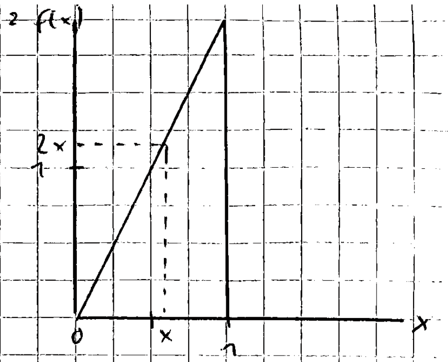
CDF

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \int_{-\infty}^{x} f(u)\,du = \int_{0}^{x} 1\,du = u \Big|_{0}^{x} = x & 0 \leq x \leq 1 \\ 1 & 1 \leq x \end{cases}$$



* $F'(x) = (x)' = 1 = f(x)$ ✓

Triangle:



$$f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Will it $\sum$?

Area under curve: $2 \cdot 1 \cdot \frac{1}{2} = 1$ ✓

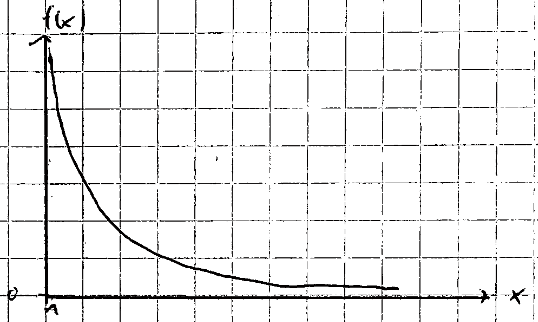
or $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 2x\,dx = x^2 \Big|_0^1 = 1 - 0 = 1$

CDF

$$F(x) = \begin{cases} 0 & x \le 0 \\ \int_{-\infty}^{x} f(u)\,du = \int_0^x 2u\,du = u^2 \Big|_0^x = x^2 & 0 \le x \le 1 \\ 1 & 1 \le x \end{cases}$$

$F'(x) = (x^2)' = 2x = f(x) \qquad 0 \le x \le 1$

Infinite Support:



$$f(x) = \begin{cases} \frac{1}{x^2} & x \ge 1 \\ 0 & x < 1 \end{cases}$$

Will it $\sum$?

$\int_{-\infty}^{\infty} f(x)\,dx = \int_1^{\infty} \frac{1}{u^2}\,du = \frac{-1}{u} \Big|_1^{\infty} = 1$

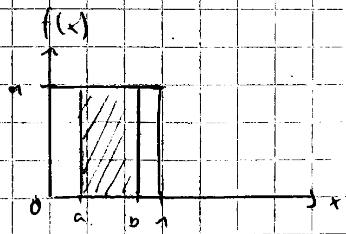
CDF

$$F(x) = \begin{cases} 0 & x < 1 \\ \int_1^x \frac{1}{u^2}\,du = \frac{-1}{u} \Big|_1^x = 1 - \frac{1}{x} & x \ge 1 \end{cases}$$

$F'(x) = \left(1 - \frac{1}{x}\right)' = \frac{1}{x^2} = f(x)$

Interval Probability:

Uniform



$0 \le a \le b \le 1$

$$P(a \le X \le b) = \begin{cases} (b-a) \cdot 1 = b - a \\ \int_a^b f(x)\,dx = \int_a^b 1\,dx = x \Big|_a^b = b - a \\ F(b) - F(a) = b - a \end{cases}$$

e.g.: $P(0{,}6 \le X \le 1{,}3) = P(0{,}6 \le X \le 1) = 0{,}4$

Differences between discrete and continuous

discrete	continuous
$p(x) \leq 1$	$f(x)$ can be > 1
$p(x) \neq 0$	$p(x) = 0$
$P(X \leq a) \neq P(X < a)$	$P(X \leq a) = P(X < a) = F(a)$
	$P(X \geq a) = P(X > a) = 1 - F(a)$
	$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$

Expectation:

	discrete	continuous
EX	$\sum x \cdot p(x)$	$\int_{-\infty}^{\infty} x f(x)\,dx$

examples:

uniform $\quad EX = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x \cdot 1\,dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$ (symmetric around $\frac{1}{2}$)

triangle $\quad EX = \int_0^1 x \cdot 2x\,dx = \frac{2x^3}{3}\Big|_0^1 = \frac{2}{3}$

infinite ("Power law") $\quad EX = \int_1^{\infty} x \frac{1}{x^2}\,dx = \int_1^{\infty} \frac{1}{x}\,dx = \ln x \Big|_1^{\infty} = \infty$

Variance:

	discrete	continuous
$V(X) \hat{=} E[(X-\mu)^2]$	$\sum_x p(x)(x-\mu)^2$	$\int_{-\infty}^{\infty} f(x)(x-\mu)^2\,dx$

examples:

- uniform $\quad E(X^2) = \int_0^1 x^2\,dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$

$V(X) = E(X^2) - EX^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

$\rightarrow \sigma = \frac{1}{2\sqrt{3}}$

- triangle $\quad EX^2 = \int_0^1 x^2 2x\,dx = \frac{1}{2}x^4\Big|_0^1 = \frac{1}{2}$

$V(X) = E(X^2) - E(X)^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{9-8}{18} = \frac{1}{18}$

$\rightarrow \sigma = \frac{1}{3\sqrt{2}}$