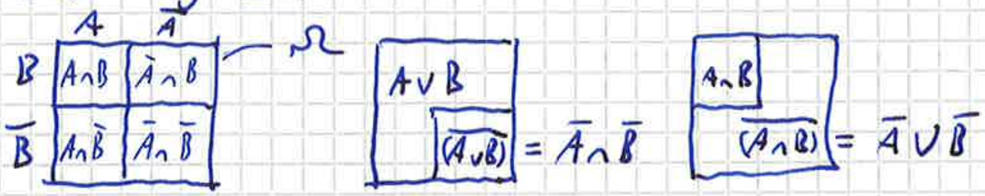
RECAP L1:

Event $A = \{\omega_1, \omega_2\}$ (elements)   e.g. "odd die" $= \{1, 3, 5\}$ $|A| = 3$

if $A \cap B = \{\}$ → "mutually exclusive"   then $\underbrace{P(A \cup B) = P(A) + P(B) \cancel{- P(A \cap B)}}_{\text{if mutually exclusive.}}$

$A \cap B = B \cap A$

De Morgan's laws (see exercise):



Conditional Probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$ "A given B"

A and B independent if $P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$

"knowledge about B has no influence of the probability of A"

Mutually Exclusive ⇒ Dependent

for $P(A) > 0$, $P(B) > 0$ if $P(A \cap B) = P(\{\}) = 0$

$P(A) \cdot P(B) > 0$ ⇒ dependent.
$\neq 0 = P(A \cap B)$

Laplace Probabilities:

$P(\{\omega_i\}) = \frac{1}{|\Omega|}$ for all i

MSE-213 L2

Book REF: Ross CH 3.1 → 4.6

2.1. PROBABILITY TREES

CHAIN RULE: $P(A \cap B) = P(B|A) \cdot P(A) = P(A|B) \cdot P(B)$

(from the definition of $P(B|A)$ ...)

also $P(A \cap B \cap C) = P(C|A \cap B) \cdot P(A \cap B)$

$= P(C|A \cap B) \cdot P(B|A) \cdot P(A)$ etc...

TOTAL PROBABILITY: $P(B) = P(A \cap B) + P(\bar{A} \cap B)$

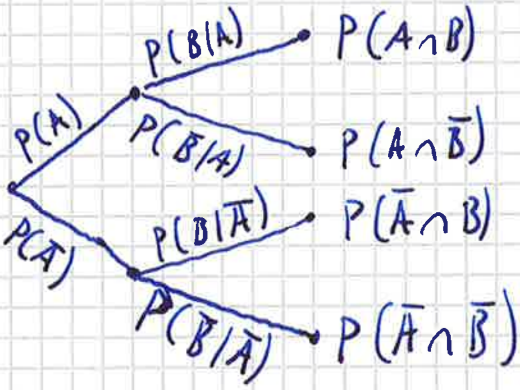
$= P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})$

generally: $P(B) = \sum_{i=1}^{n} P(C_i \cap B)$

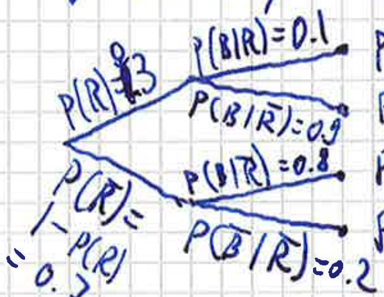
if $C_1 \cup C_2 \cup \ldots \cup C_n = \Omega$ (complete)

and $C_i \cap C_j = \{\}$ for $i \neq j$ (mutually exclusive)

Tree Diagram:



EX: There is a 30% probability that it rains in the morning. If it rains (R), Prof. J. goes to EPFL by bike with 10% probability. If it does not rain, he goes by bike with 80% probability. What is the total probability for going by bike (B)?



$P(R \cap B) = 0.3 \cdot 0.1 = 0.03$

$P(R \cap \bar{B}) = 0.3 \cdot 0.9 = 0.27$

$P(\bar{R} \cap B) = 0.7 \cdot 0.8 = 0.56$

$P(\bar{R} \cap \bar{B}) = 0.7 \cdot 0.2 = 0.14$

$P(B) = P(R \cap B) + P(\bar{R} \cap B)$

$= 0.59 = 59\%$

2.2. Bayes' Rule

remember: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \cap A)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)}$

$$\Rightarrow \boxed{P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}} = P(B|A) \cdot \frac{P(A)}{P(A \cap B) + P(\bar{A} \cap B)}$$

EX: You see Prof. J. came by bike (B is true), what is the probability that it was raining (P(R|B))?

$$P(R|B) = P(B|R) \cdot \frac{P(R)}{P(B)}$$

$$= 0.1 \cdot \frac{0.3}{0.3 \cdot 0.1 + (1-0.3) \cdot 0.8} = \frac{0.03}{0.59} = \frac{3}{59} \approx 5.1\%$$

2.3 Random Variables

The <u>function</u> X assigns a numerical value x_i to each element of the sample space ω_i. Usually this value is a real number, then $X: \Omega \to \mathbb{R}$ (real numbers)

$$\omega_i \to x_i$$

$$\text{or } X(\omega_i) = x_i$$

Ex: $X(\text{"heads"}) = 1 \quad X(\text{"tails"}) = 0$

so $\omega_1 = \text{"heads"} \quad \omega_2 = \text{"tails"}$

$x_1 = 1 \quad x_2 = 0$

Confusingly, the <u>function</u> X is known as a <u>random variable</u>

We also have $P(\omega_i)$ the probability of some element.

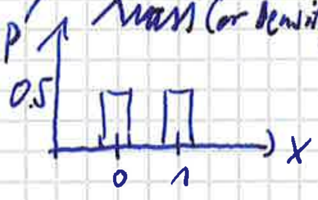
We write the shorthand $P(\omega_i) = p_i = p_{\omega_i}$

EX: $P(\text{"heads"}) = p_1 = p_{\text{heads}} = 1/2$

and hence $P(X=1) = \frac{1}{2} \quad P(X=0) = \frac{1}{2}$

is our "Probability Mass (or Density) Function" for a <u>discrete</u> r.

We can plot it:



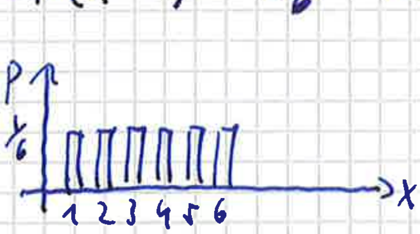
("bar plots" are typical)

We use the shorthand: $p(x_i) = p_i$

EX: throw die, $X(⚀) = 1$, $X(⚁) = 2$ etc...

$$P(⚀) = P(X=1) = 1/6 \quad \text{etc...}$$

ω_i	⚀	⚁	⚂	⚃	⚄	⚅
x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6



The connection between elements and assigned value can be very different. For example we could have, for an unordered 2-coin toss:

ω_1 = "2 heads" $\quad$ ω_2 = "2 tails" $\quad$ ω_3 = "1 head 1 tail"

$x_1 = 100$ $\quad$ $x_2 = -5$ $\quad$ $x_3 = 0$

$p_1 = \frac{1}{4}$ $\quad$ $p_2 = \frac{1}{4}$ $\quad$ $p_3 = \frac{1}{2}$

But note that $\sum_{i=1}^{N} p_i = 1$ always.

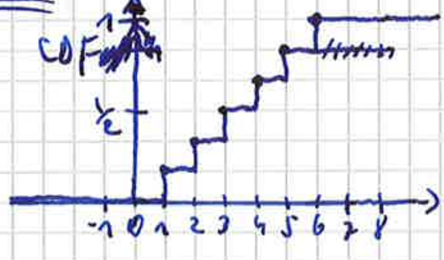
2.4 CUMULATIVE DISTRIBUTION FUNCTION:

What is the probability for $X \leq$ some value? This is called Cumulative distribution function, $CDF(x) = F(x)$.

We have $CDF(-\infty) = 0$, $CDF(\infty) = 1$, CDF is non-decreasing (i.e. $CDF(x) \geq CDF(y)$ if $x \geq y$)

$$CDF(x) = \sum_{i=-\infty}^{x} P(X) \quad \text{for discrete distributions.}$$

EX: die throw where ⚀ → 1 etc



2.5 Expectation Value

If you sample a random variable infinitely many times, which arithmetic mean ("average") do you expect?

$$\mathbb{E}(X) = \sum_{i=1}^{N} P(w_i) \cdot X(w_i) = \sum_{i=1}^{N} x_i \cdot \underbrace{p_i}_{\text{"weight"}} = \underbrace{\mu}_{\text{mean of the random variable}}$$

<u>Ex</u> for a die throw, as above, we have:

$$\mathbb{E}(X) = \frac{1}{6} \cdot (1+2+3+4+5+6) = \frac{21}{6} = 3.5$$

for the coin game above we have

$$\mathbb{E}(X) = \frac{1}{4} \cdot 100 + \frac{1}{4} \cdot (-5) + \frac{1}{2} \cdot 0 = 23.75 \approx 24$$

A game is considered fair if $\mathbb{E}(X) = 0$. $\longrightarrow$ SLIDES

We can have $\mathbb{E}(f(X))$, e.g. $\mathbb{E}(X^2)$ or $\mathbb{E}(\cos(X))$.

Then: $\mathbb{E}(f(X)) = \sum_{i=1}^{N} f(x_i) \cdot p(x_i)$

<u>EX</u>: $\mathbb{E}(a + bX) = \mathbb{E}(a) + \mathbb{E}(bX)$ (DISTRIBUTIVE, see definition ...)

(a, b; constants)

$$= \sum_i a\, p_i + \sum_i b\, x_i\, p_i$$

$$= a \sum_i p_i + b \sum_i x_i\, p_i$$

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

<u>EX</u>: Single coin toss with 0 for tails, 1 for heads:

$$\mathbb{E}(X) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = 0.5$$

$$\mathbb{E}(X^2) = \frac{1}{2} \cdot 0^2 + \frac{1}{2} 1^2 = 0.5 \neq [\mathbb{E}(X)]^2 = 0.25$$ ↯

$$\mathbb{E}(\cos(X)) \approx \frac{1}{2} \cdot 1 + \frac{1}{2} 0.54 = 0.77 \neq \cos(E(X)) = 0.88$$

2.6 Variance

The mean $\mu = \mathbb{E}(X)$ gives the "central tendency" of X.

To characterize the spread we define the variance:

$$\boxed{\mathbb{V}(X) = Var(X) = \mathbb{E}((X-\mu)^2)}$$ → units of X^2

this is equivalent to $= \sum_i (x_i - \mu)^2 p_i$ → as $()^2 \geq 0$ always and $p_i \geq 0$ always $\Rightarrow \underline{\mathbb{V}(X) \geq 0}$

$$= \sum_i (x_i^2 + \mu^2 - 2\mu x_i) p_i$$

$$= \sum_i x_i^2 p_i + \mu^2 \sum_i p_i - 2\mu \sum x_i p_i$$

$$= \mathbb{E}(X^2) + \mu^2 \cdot 1 - 2\mu \cdot (\mu)$$

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

$$\boxed{\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2}$$

EX: dice: $\mathbb{V}(X) = \left(\frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)\right) - \left(\frac{21}{6}\right)^2$

$$= \frac{91}{6} - \frac{441}{36} = \underline{\underline{\frac{35}{12} \simeq 2.9}}$$

Coin: $\mathbb{V}(X) = \left(\frac{1}{2}(0^2+1^2)\right) - \left(\frac{1}{2}\right)^2$

$$= \frac{1}{2} - \frac{1}{4} = \underline{\underline{\frac{1}{4}}}$$

Property of $\mathbb{V}(X)$:

$$\mathbb{V}(a + bX) = b^2\,\mathbb{V}(X) \qquad \text{So } \frac{\mathbb{V}(bX)}{\mathbb{E}(bX)} = b\,\frac{\mathbb{V}(X)}{\mathbb{E}(X)}$$

2.7 Standard deviation:

$\sigma = SD(X) = \sqrt{\mathbb{V}(X)}$ → same units as X

Note $SD(bX) = \sqrt{b^2} \cdot SD(X) = |b|\, SD(X)$

$$\text{So } \frac{SD(bX)}{\mathbb{E}(bX)} = \frac{|b|}{b}\,\frac{SD(X)}{\mathbb{E}(X)} = \text{sign}(b)\,\frac{SD(X)}{\mathbb{E}(X)}$$

2.8 Bernoulli Distributions

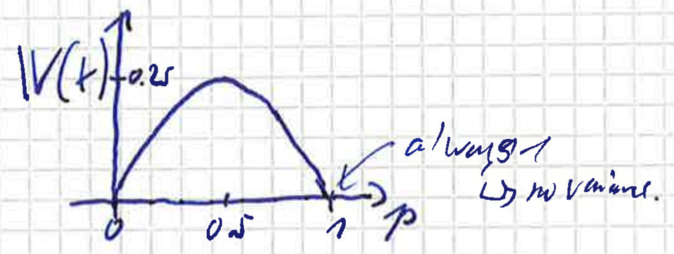
- two possible outcomes $\Omega = \{\omega_0, \omega_1\} = \{\text{"fail"}, \text{"win"}\}$
- $X(\omega_0) = 0 \quad X(\omega_1) = 1$
- $P(\omega_0) = 1-p \quad P(\omega_1) = p$

$\Rightarrow \underline{\mathbb{E}(X)} = 0\cdot(1-p) + 1\cdot p = \underline{p}$

$\underline{\mathbb{V}(X)} = \mathbb{E}(X^2) - E(X)^2$

$= (0^2(1-p) + 1^2 p) - p^2$

$= p - p^2 = \underline{p(1-p)}$



2.9 Binomial Distribution

Repeat Bernoulli experiment n times, ignore order.

Probability of k total "wins" $P(X=k)$

$X = X_1 + X_2 + \ldots + X_n \rightarrow$ total number of "win"s

$\Omega = \{(n\times f, 0\times w), (n-1\times f, 1\times w), (\ldots), (0\times f, n\times w)\}$

$|\Omega| = n+1$ (one option)

$$P(X=k) = \underset{k \text{ "win"s}}{p^k} \cdot \underset{(n-k) \text{ fails}}{(1-p)^{(n-k)}} \binom{n}{k}$$

$\binom{n}{k}$: "n choose k" possible combinations

$= \dfrac{n!}{k!(n-k)!}$ where $n! = n\cdot(n-1)\cdot\ldots\cdot 1$

binomial coefficient.

$\binom{n}{k} = \binom{n}{n-k}$

1
1 1
1 2 1
1 3 3 1

<u>EX</u> $P(X=k)$ for $n=2$

$P(X=0) = (1-p)^2$

$P(X=2) = p^2$

$P(X=1) = p(1-p)\cdot 2$

$\mathbb{E}(X)$ ... complicated via $P(X)$ ...

easy via $\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + \dots X_n)$

$$= \sum_{j=1}^{n} \underset{\substack{\uparrow \\ \text{Bernoulli}}}{\mathbb{E}(X_j)} = n \cdot p$$

→ 1000 coins... 50% chance... 500 expected...

$$\mathbb{V}(X) = \sum_{j=1}^{n} \mathbb{V}(X_j) = n \cdot p(1-p)$$

Because X_i, X_j are independent (will learn this later...)

$$\frac{\sigma}{\mu} = \frac{\sqrt{\mathbb{V}(X)}}{\mathbb{E}(X)} = \frac{\sqrt{n\, p(1-p)}}{n \cdot p} = \frac{1}{\sqrt{n}} \frac{\sqrt{p(1-p)}}{p}$$

→ relative width decreases...