Example 3.18

$$\Omega = \{00, 01, 10, 11\}$$

$$P(00) = P(01) = P(10) = P(11) = \frac{1}{4}$$

$X(\omega)$ = Number of ones in ω.

$X(00) = 0$

$X(10) = X(01) = 1$

$X(11) = 2$

$$E(X) = X(00) \cdot P(00) + X(01) \cdot P(01) + X(10) \cdot P(10) + X(11) \cdot P(11) = 1$$

$$E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2)$$

$$0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.$$

Let's generalize to the biased-coin case.

1 = tail with prob p

0 = head with prob. $1-p$

2 tosses

$$P(00) = (1-p)^2$$
$$P(01) = p(1-p)$$
$$P(10) = p(1-p)$$
$$P(11) = p^2$$

$X(\omega)$ = # ones in ω
= # tails.

$$E(X) = X(00) \cdot P(00) + X(01) \cdot P(01) + X(10) \cdot P(10) + X(11) \cdot P(11) =$$
$$0 \cdot (1-p)^2 + 1 \cdot p \cdot (1-p) + 1 \cdot p(1-p) + 2 \cdot p^2$$
$$= 2p.$$

$$E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=1)$$
$$= 0 \cdot (1-p)^2 + 1 \cdot 2p(1-p) + 2 \cdot p^2 = 2p.$$

The expectation can be determined from the distribution:

Prop 3.19 For a random variable $X: \Omega \to \mathbb{R}$ on a finite probability space (Ω, P) can be determined from the distribution $x \in X(\Omega) \mapsto P(X=x)$ via

$$E(X) = \sum_{x \in X(\Omega)} x \cdot P(X=x)$$

Proof:

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$$

$$= \sum_{x \in X(\Omega)} \sum_{\substack{\omega \in \Omega: \\ X(\omega)=x}} x \cdot P(\omega)$$

$$= \sum_{x \in X(\Omega)} x \underbrace{\sum_{\substack{\omega \in \Omega: \\ X(\omega)=x}} P(\omega)}_{= P(X=x)}$$

$\square$

Proposition 3.20 (Markov's Inequality – the probability of going (much) higher than the average)

For $X: \Omega \to \mathbb{R}_{\geq 0}$ and $a > E(X)$, one has $\boxed{P(X \geq a) \leq \frac{E(X)}{a}}$

Proof: We show $a \cdot P(X \geq a) \leq E(X)$:

$$a \cdot P(X \geq a) = a \cdot \sum_{\substack{\omega \in \Omega: \\ X(\omega) \geq a}} P(\omega)$$

$$= \sum_{\substack{\omega \in \Omega: \\ X(\omega) \geq a}} a \cdot P(\omega) \leq \sum_{\substack{\omega \in \Omega: \\ X(\omega) \geq a}} X(\omega) \cdot P(\omega) \leq \underbrace{\sum_{\omega \in \Omega} X(\omega) P(\omega)}_{= E(X)}$$

$\square$

What is the probability (an estimate on the probability) that X deviates significantly from its expectation $E(X)$? We want an estimate assuming we know the variance $V(X)$.

$$V(X) = E\left((X - E(X))^2\right).$$

Theorem 3.21. (Chebyshev's inequality)

For $X: \Omega \to \mathbb{R}$ and $b > 0$, one has

$$\boxed{P(|X - E(X)| \geq b) \leq \frac{V(X)}{b^2}}$$

Proof:

$$P(|X - E(X)| \geq b) = P\left((X - E(X))^2 \geq b^2\right)$$

$$\overset{\text{Prop. 3.20}}{\leq} \frac{E\left((X - E(X))^2\right)}{b^2} = \frac{V(X)}{b^2}. \qquad \square$$

Question: Assume you have n independent tosses of a biased coin.

Tail $\hat{=} 1$ with prob. p

Head $\hat{=} 0$ with prob. $1-p$.

Let X = # tails thrown.

$E(X) = np$ (intuitively clear, but can also be proven).

$V(X) = ???$

How likely is it to deviate significantly from $E(X)$?

Prop. 3.22 For events $A_1, \dots, A_m \subseteq \Omega$ one has $P\left(\bigcup_{i=1}^{m} A_i\right) \leq \sum_{i=1}^{m} P(A_i)$.

Proof: easy ← try doing it yourselves properly.

Theorem 3.23 Let $X_1, \dots, X_m$ be independent real valued random variables. Then

$$E(X_1 \cdot \ldots \cdot X_m) = E(X_1) \cdot \ldots \cdot E(X_m)$$

and

$$V(X_1 + \ldots + X_m) = V(X_1) + \ldots + V(X_m).$$

$$E(X+Y) = E(X) + E(Y) \quad ?$$

$$\begin{array}{c} \| \end{array}$$

$$\sum_{\omega \in \Omega} \left(X(\omega) + Y(\omega)\right) \cdot P(\omega) = \sum_{\omega \in \Omega} X(\omega) P(\omega) + \sum_{\omega \in \Omega} Y(\omega) P(\omega)$$

Example 3.24. n independent tosses of a biased coin with the prob. of the tail $\hat{=} 1$ being p. Let X be the number of tails thrown.

$X = X_1 + \ldots + X_n$ where $X_i \in_R \{0, 1\}$ is the outcome in the i-th toss.

$$E(X) = E(X_1 + \dots + X_n)$$
$$= E(X_1) + \dots + E(X_n)$$
$$= n \cdot E(X_1)$$
$$= n\big(0 \cdot P(X_1 = 0) + 1 \cdot P(X_1 = 1)\big)$$
$$= n \cdot P(X_1 = 1) = np.$$

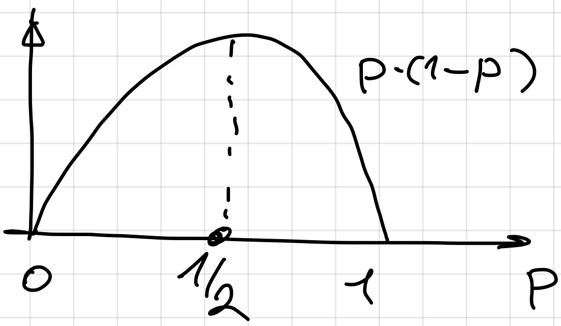
$$V(X) = V(X_1 + \dots + X_n) \overset{\text{Th. 3.2.3.}}{=} V(X_1) + \dots + V(X_n)$$

indep. tosses.

$$= n \cdot V(X_1)$$

$$= n \cdot p \cdot (1-p).$$

$$V(X_1) = E\big((X_1 - E(X_1))^2\big)$$
$$= E\big((X_1 - p)^2\big)$$
$$= (0-p)^2 P(X_1 = 0) + (1-p)^2 P(X_1 = 1)$$
$$= p^2 \cdot (1-p) + (1-p)^2 \cdot p$$
$$= p \cdot (1-p)\big(p + (1-p)\big)$$
$$= p \cdot (1-p).$$



"the most uncertain coin is the unbiased coin".

Example 3.25 $n=100$ indep. rounds of a game. Win with prob. 0.1 per round.

$X = \#$ wins.

$$E(X) = n \cdot p = 100 \cdot 0{,}1 = 10.$$

$$V(X) = 100 \cdot 0{,}1 \cdot 0{,}9 = 9.$$

$$P(X \geq 20) = P(|X - \underbrace{10}_{E(X)}| \geq 10)$$

$$\underset{\text{Chebyshev}}{\leq} \frac{V(X)}{10^2} = \frac{9}{10^2} = 0{,}09$$

$\boxed{\text{LQ:}}$ Prove 3.22 and 3.23, at least in special cases.

Remark 3.26 Chebyshev's inequality can also be written in this way:

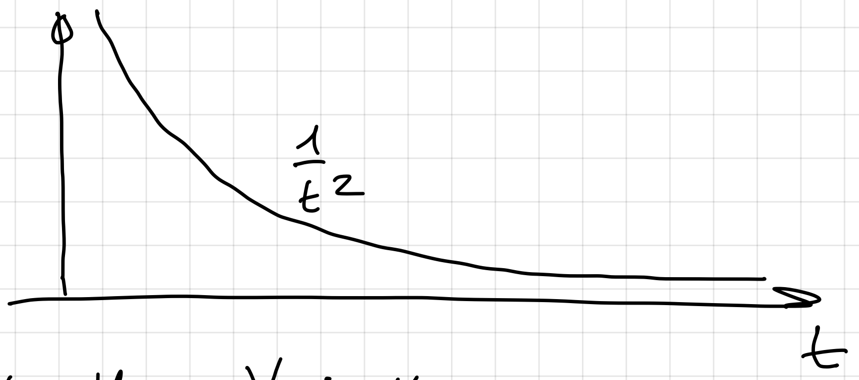
$$\text{Prob}\left(|X - E(X)| \geq t \cdot \sqrt{V(X)} \right) \leq \frac{1}{t^2}$$

for every $t > 0$.

Let's apply this to $X = \#$ tails thrown in n tosses of a biased coin with the prob. of the tail equal to p. We obtain:

$$\text{Prob}\left(|X - np| \geq t \sqrt{np(1-p)} \right) \leq \frac{1}{t^2}.$$

Prob...

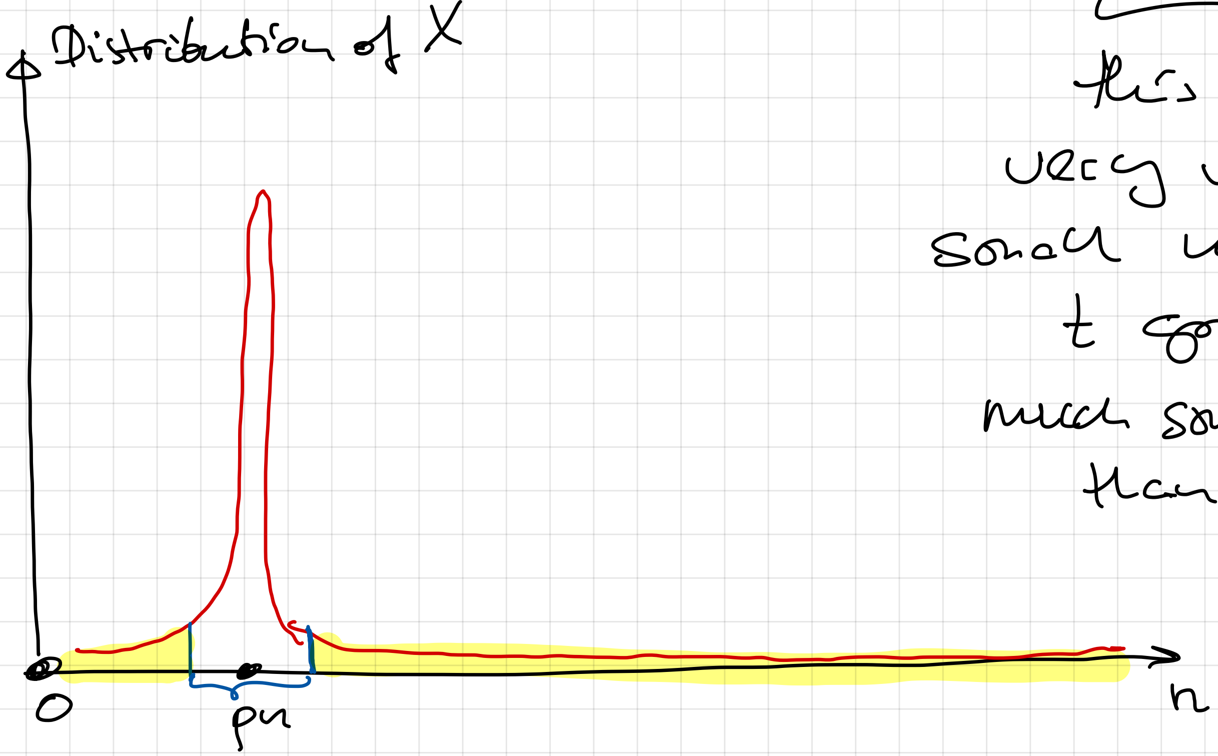


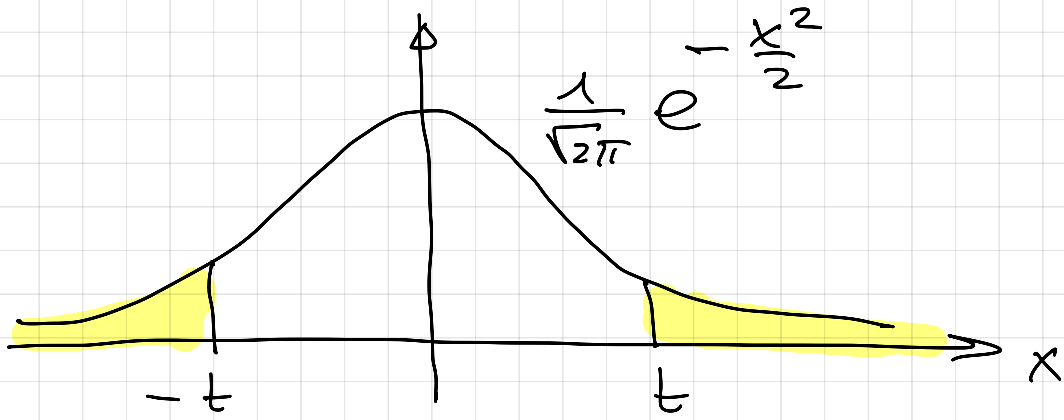
Remember that X is the sum of independent identically distributed variables.

$$X = X_1 + X_2 + \dots + X_n \text{, with}$$

$X_i \in_R \{0,1\}$ being the outcome in the ith toss. The central limit theorem (De Moivre-Laplace Theorem in our particular case) tells us:

$$\lim_{n\to\infty} P\left(|X - np| \geq t\sqrt{np(1-p)}\right) = 2 \cdot \underbrace{\frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-\frac{x^2}{2}}\,dx}_{\substack{\text{this gets very very small when } t \text{ grows,} \\ \text{much smaller than } \frac{1}{t^2}}}$$





$\boxed{3.2}$ Stochastic channels

Def. 3.27. For a probability space (Ω, P) (a finite one) a random function $F: V \xrightarrow{R} W$ from V to W is a function

$$F: V \times \Omega \longrightarrow W$$

so $F(v, \omega)$ depends on $v \in V$ (the input) and $\omega \in \Omega$ (randomness or chance).

In other words, if $V = \{v_1, \dots, v_m\}$ we can interpret F as a system of random variables

$$F(v_1) \in_R W, \dots, F(v_m) \in_R W.$$

Def 3.28 Let W be a finite word space, say $W = K^n$, then a stochastic channel CH on W is a random function

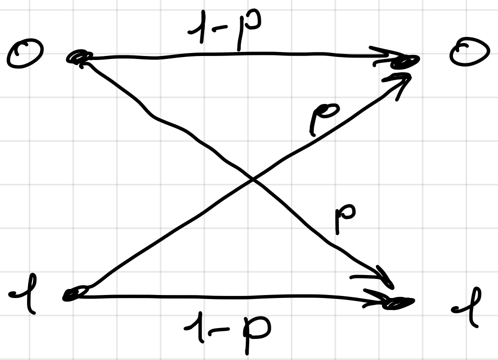
$$CH: W \xrightarrow{R} W.$$

For CH we can define for each $x, y \in W$

$$P_{x \to y} := P(CH(x) = y).$$

this is the probability that CH transmits $x \in W$ as y. Whenever we know the values $p_{x \to y}$, we can analyze the properties of the channel.

Ex 3.29 $W = \{0,1\}$



	$y=0$	$y=1$
$x=0$	$1-p$	p
$x=1$	p	$1-p$

table for $p_{x \to y}$.

Rem 3.30 Observe that

$$\sum_{y \in W} p_{x \to y} = 1 \quad \text{because}$$

the values $p_{x \to y}$ with $y \in W$ form the distribution of $CH(x)$.

Def 3.31 The binary symmetric channel for words of length n with the cross-over probability $p \in (0,1)$ is defined as follows: $CH : \{0,1\}^n \xrightarrow{R} \{0,1\}^n$ with

$$CH(x_1, \dots, x_n) = (x_1 \oplus N_1, \dots, x_n \oplus N_n)$$

where $\oplus$ is addition mod 2 (so, XOR) and $N_1, \dots, N_n \in_R \{0,1\}$ are the outcomes of

The n independent tosses of a biased coin with the tail probability equal to p.

$$\underbrace{\begin{array}{c|cccc} & y=00 & y=01 & y=10 & y=11 \\ \hline x=00 & (1-p)^2 & p(1-p) & p(1-p) & p^2 \\ x=01 & p(1-p) & (1-p)^2 & p^2 & p(1-p) \\ x=10 & p(1-p) & p^2 & (1-p)^2 & p(1-p) \\ x=11 & p^2 & p(1-p) & p(1-p) & (1-p)^2 \end{array}}_{P_{x \to y}}$$

Double check: $(1-p)^2 + 2p(1-p) + p^2 =$
$(1-p+p)^2 = 1$

(each row sums up to one).

Def. 3.32 For $q \in \mathbb{N}$, $q \geq 2$, the q-ary symmetric channel on the words of length $n \in \mathbb{N}$ with the cross-over prob. $p \in (0,1)$ is introduced as

$$CH : K^n \xrightarrow{R} K^n$$

defined as follows (K is an alphabet of size q)

$$CH(x_1, \dots, x_n) = (Y_1, \dots, Y_n)$$

where $Y_1, \dots, Y_n \in_R K$ are independent random variables such that

$$P(Y_i = x_i) = 1-p \quad \text{and}$$

$$P(Y_i = y_i) \text{ where } y_i \in K \setminus \{x_i\}$$

is equal to

$$P(Y_i = y_i) = \frac{p}{q-1}$$

Ex 3.33 $K = \{0,1,2\}$, $q = 3$, $\boxed{n=1}$

$$\underbrace{\begin{array}{c|ccc} & y=0 & y=1 & y=2 \\ \hline x=0 & 1-p & \frac{p}{2} & \frac{p}{2} \\ x=1 & p/2 & 1-p & p/2 \\ x=2 & p/2 & p/2 & 1-p \end{array}}_{P_{x \to y}}$$

Take $p = \frac{1}{5}$

$$P(CH(x) = x) = 1-p = 1 - \frac{1}{5} = \frac{4}{5} = 80\%.$$

$$1-p = \frac{p}{2}$$

$$1 = \frac{3p}{2}$$

$$\boxed{p = \frac{2}{3}} \Longrightarrow \begin{array}{|ccc} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{array} \Rightarrow$$ we cannot guess x from $CH(x)$.

Proposition 3.34 Let K be a q-ary alphabet with $q \in \mathbb{N}$, $q \geq 2$ and $CH: K^n \xrightarrow{R} K^n$ be a q-ary symmetric channel with the crossover probability $p \in (0,1)$. Then

$$P_{x \to y} = (1-p)^{n-d(x,y)} \cdot \left(\frac{p}{q-1}\right)^{d(x,y)},$$

where $d(x,y)$ is the Hamming distance of x and y, is valid for all $x,y \in K^n$.

Proof: For $x=(x_1,\dots,x_n)$ and $y=(y_1,\dots,y_n)$

$$P_{x\to y} = P(A_1 \cap A_2 \cap \dots \cap A_n), \text{ where}$$

A_i is the probability that x_i is transmitted as y_i. The events $A_1,\dots,A_n$ are independent by the def. of a q-ary stochastic channel.

$$\Rightarrow \quad P_{x\to y} = \prod_{i=1}^{n} P(A_i).$$

$$P(A_i) = \begin{cases} 1-p & \text{if } x_i = y_i \\ \frac{p}{q-1} & \text{else} \end{cases}$$

$$\boxed{\begin{array}{l} x = 0\ 0\ 1\ 0\ 0\ 2\ 3\ 3 \\ \downarrow \\ y = 1\ 0\ 3\ 2\ 1\ 0\ 3\ 1 \end{array}} \leftarrow \left(\frac{p}{q-1}\right)^6 \cdot (1-p)^2$$

$$\Rightarrow \quad P_{x\to y} = (1-p)^{n-d(x,y)} \cdot \left(\frac{p}{q-1}\right)^{d(x,y)}.$$

$\square$

3.3 Maximal likelyhood decoding

Setting: you've got a channel

$$CH : W \xrightarrow{R} W$$ and you want to guess x from the knowledge of $CH(x)$.

Frequently, one also considers an input from W to be a random variable independent on the randomness in the channel.

Assume, we have a code $\emptyset \neq C \subseteq W$ and we have a random variable $X \in_R C$ (a random codeword) with some known distribution

$$p_x := P(X = x) \text{ for each}$$

$x \in C$. Let Y be the output of the stochastic channel on the input X. Assume that we know $p_{x \to y} = P(Y = y \mid X = x)$.

We want to guess x from y:

$$P(X = x \mid Y = y)$$

$$= \frac{P(Y = y \mid X = x)}{P(Y = y)} P(X = x)$$

$$= \frac{p_{x \to y}}{\sum_{x \in C} P(Y=y \mid X=x)\, P(X=x)} p_x$$

$$= \frac{p_{x \to y}}{\sum_{x' \in C} p_{x'} p_{x' \to y}} p_x$$

We arrive at:

$$P(X=x \mid Y=y) = p_x \frac{p_{x \to y}}{\sum_{x' \in C} p_{x'} p_{x' \to y}},$$

which is true when $P(Y=y) > 0$ for all $y \in W$.

In particular, if X is uniformly distributed in C, meaning

$p_x = \frac{1}{|C|}$ for each $x \in C$,

we have

$$P(X=x \mid Y=y) = \frac{P_{x\to y}}{\sum_{x'\in C} P_{x'\to y}}$$

$\boxed{LQ}$

$C=\{0,1\}$ $W=\{0,1,2\}$

$P_{x\to y}$	$y=0$	$y=1$	$y=2$
$x=0$	0,8	0.2	0
$x=1$	0,1	0.8	0.1
$x=2$	0	0.2	0.8

$p_0 = \frac{1}{2}$, $p_1 = \frac{1}{2}$

You receive	You decode it to.....?
0	....
1	....
2	....