malot@unice.fr

II) The discrete r.v.

Def: let X be a discrete r.v.

The probability function that characterizes X is given by:

- $X(\Omega) = \{$all the possible issues$\}$
- $\forall i \in X(\Omega), P(X = i)$

properties:

- $\displaystyle\sum_{i \in X(\Omega)} P(X = i) = 1$
- $\forall i \in X(\Omega),\ P(X = i) \geq 0$

When we consider a r.v:

- discrete
- continuous

mean = expecta-
-tion

Expectation and Variance

<u>def</u>: expectation = (probability) mean

If $\sum_{i \in X(\Omega)} i \times P(X=i)$ exists

we say that X has an expectation

$$\Rightarrow \mathbb{E}[X] = \sum_{i \in X(\Omega)} i \times P(X=i)$$

Example

Let X be the issue when you throw a dice.

- $X(\Omega) = \{1, 2, 3, 4, 5, 6\} = [\![1, 6]\!]$
- $\forall i \in X(\Omega),\ P(X = i) = \frac{1}{6}$

Assumption:

x_1 is associated to X_1

$\vdots$

x_n is associated to X_n

$X_1, \dots, X_n$ are independant and identically distributed (iid)

Example:

$$\mathbb{E}[X] = \sum_{i=1}^{6} i \times \frac{1}{6}$$

$$= \frac{1}{6} \times \frac{6 \times 7}{2} = 3.5$$

Chapter 1 : Introduction to probability

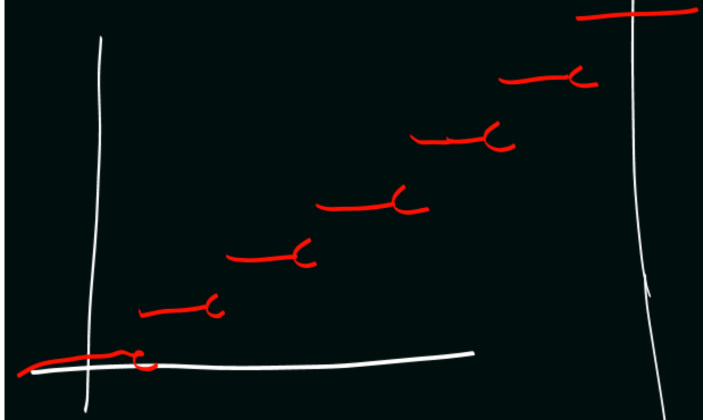
I Notations :

X : random variable (r.v.)

$x_1, \dots, x_n$: n observations of X

Example:

$$F_X(t) = \begin{cases} 0 & \text{if } t < 1 \\ \frac{1}{6} & \text{if } t \in [1;2[\\ \frac{2}{6} & \text{if } t \in [2;3[\\ \frac{3}{6} & \text{if } t \in [3;4[\\ \frac{4}{6} & \text{if } t \in [4;5[\\ 5/6 & \text{if } t \in [5;6[\\ 1 & \text{if } t \geq 6 \end{cases}$$



III Common discrete r.v

- uniform
- bernoulli
- binomial
- geometric
- poisson

def: variance

If $\sum_{i \in X(\Omega)} i^2 P(X=i)$ exists

then the variance of X exists and

$$V[X] = \sum_{i \in X(\Omega)} (i - \mathbb{E}[X])^2 P(X=i)$$

$=$

distribution function:

let X be a discrete r.v:

$\forall t \in \mathbb{R}$, $F_X(t) = P(X \leq t)$

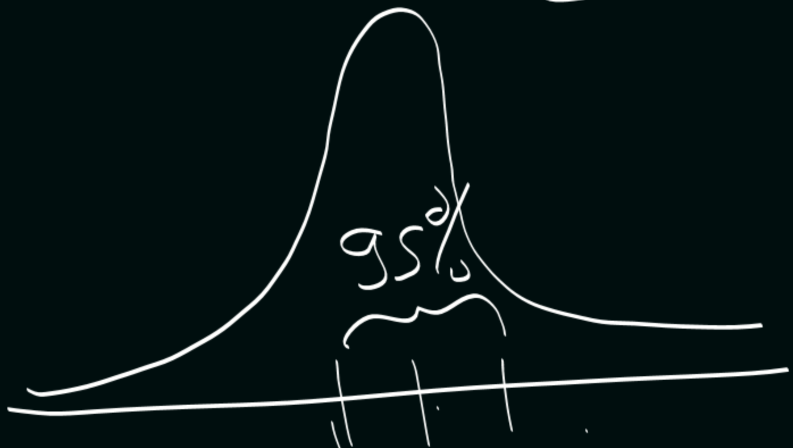
$= \sum_{\substack{i \in X(\Omega) \\ i \leq t}} P(X = i)$

sometimes

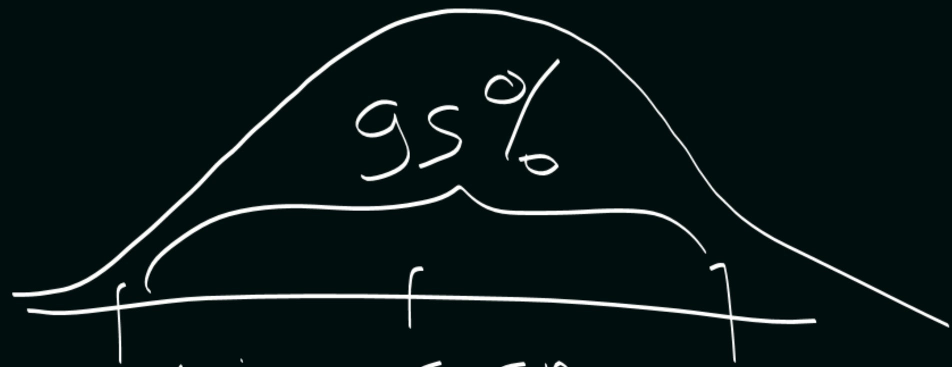
$F_X(t) = P(X < t)$

Rq: In general

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



small variance



bigger one

Ex :

Let F be the function

$$F(t) = \begin{cases} 0 & \text{if } t < 3 \\ 1/2 & \text{if } t \in [3 ; 5[\\ 3/4 & \text{if } t \in [5 ; 6[\\ 1 & \text{if } t \in [6 , +\infty[\end{cases}$$

Let X be the associated r.v.

Then $X(\Omega) = \{3, 5, 6\}$

$$P(X=3) = 1/2$$

$$P(X=5) = 1/4 = (3/4 - 1/2)$$

$$P(X=6) = 1/4 = (1 - 3/4)$$

properties:

- constant piecewise function
- increasing function
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to +\infty} F_X(x) = 1$

Rq: $V[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X]$

property: $V[X] \geq 0$

standard deviation: $\sigma(X) = \sqrt{V[X]}$

c) Binomial

$X \sim \mathcal{B}(n; p)$ $\quad n \in \mathbb{N}^*$

$p \in [0; 1]$

- $X(\Omega) = \{0; 1; \dots; n\}$

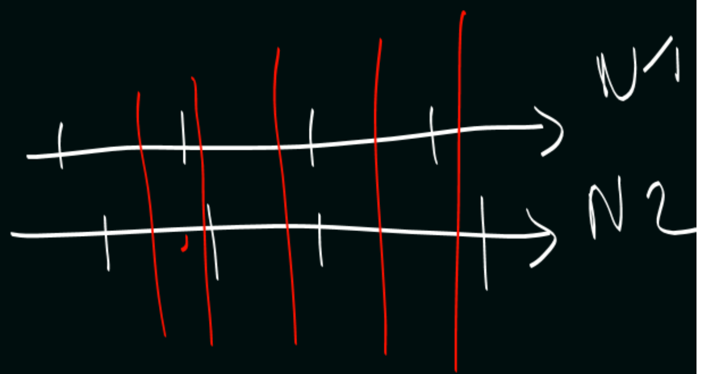
- $\forall k \in X(\Omega)$, $P(X = k) = C_n^k \, p^k (1-p)^{n-k}$

$= \binom{n}{k} p^k (1-p)^{n-k}$

Example:

You throw a coin n times and I count the number of time I get P.

$$X \sim B(n; 0.5)$$



Simulation

a) Uniform

Let $X \sim \mathcal{U}(\{1, \dots, K\})$ $\quad$ K an integer

$K > 1$

- $X(\Omega) = \{1, \dots, K\}$
- $\forall i \in X(\Omega)$, $P(X = i) = \dfrac{1}{K} = \dfrac{1}{\# X(\Omega)}$

In practice:

$$X \sim \mathcal{B}(n; p)$$

$$X = Y_1 + Y_2 + \dots + Y_n$$

with $Y_i \sim \mathcal{B}(p)$

and $Y_1, Y_2, \dots, Y_n$ are independant

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$: n^b of combinations of k elements among n (without order)

Geometric

$X \sim \mathcal{G}(p) \quad p \in [0, 1]$

$X(\Omega) = [\![1, +\infty[\![= \{1, 2, \dots, +\infty\}$

$\forall k \in X(\Omega),\ P(X = k) = p(1-p)^{k-1}$

a real interval $[1, 3]$

an integer interval $[\![1, 3]\!] = \{1, 2, 3\}$

b) Bernoulli

$X \sim \mathcal{B}(p) \qquad p \in [0;1]$

$X(\Omega) = \{0;1\}$

(the important thing is that we have 2 issues)

$P(X=1) = p$

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

$$\frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{(n+1)(2n+1)}{6}$$

$$1^2 + 2^2 + \cdots + n^2$$

2) $X \sim \mathcal{B}(p)$

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

$$= 1 \times P(X=1)$$

$$\boxed{E[X] = p}$$

$$V[X] = \left(0^2 \times P(X=0) + 1^2 P(X=1)\right) - p^2$$

$$\boxed{V[X] = p \times (1-p) = pq}$$

$$V[X] = \frac{(K+1)(2K+1)}{6} - \frac{(K+1)^2}{4}$$

$$\boxed{V[X] = \frac{K^2-1}{12}}$$

$$V[X] = E[X^2] - E[X]^2$$

$$E[X^2] = \sum_{i=1}^{K} i^2 \times P(X=i)$$

$$= \frac{1}{K} \sum_{i=1}^{K} i^2 = \frac{1}{K} \times \frac{K(K+1)(2K+1)}{6}$$

Rk:

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

idea:

$$\begin{array}{ccccc} 1 & 2 & 3 & \cdots & n \\ n & n-1 & n-2 & & 1 \\ \hline (n+1) & (n+1) & (n+1) & & (n+1) \end{array}$$

1) $X \sim \mathcal{U}(\{1, \dots, K\})$

$$\mathbb{E}[X] = \sum_{i=1}^{K} i \times P(X = i)$$

$$= \sum_{i=1}^{K} \frac{1}{K} \times i = \frac{1}{K} \sum_{i=1}^{K} i = \frac{1}{K} \frac{K \times (K+1)}{2}$$

$$\Rightarrow \boxed{\mathbb{E}[X] = \frac{K+1}{2}}$$

Rk :

Let X and Y two r.v

Let F_X the distribution f^{ct} for X

F_Y ——— Y

If $\forall t \in \mathbb{R}$, $F_X(t) = F_Y(t)$

then X and Y have the same law

$X \sim \mathcal{G}(p)$: the first time of succes when we perform bernoulli expperimentations.

? Can you compute the expectation and the variance of X when:

1) $X \sim \mathcal{U}(\{1, \dots, K\})$

2) $X \sim \mathcal{B}(p)$

3) $X \sim \mathcal{B}(n, p)$

$$V[X] = V[Y_1 + \cdots + Y_n]$$

$$= V[Y_1] + \cdots + V[Y_n]$$

the variables are indepen- - dant

$$\boxed{V[X] = n \times p \times q}$$

Property:

$$X \sim \mathcal{G}(p)$$

$$E[X] = 1/p$$

$$V[X] = \frac{1-p}{p^2}$$

RR:

$$E[X] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} k$$

$$= \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \times k \times p^k (1-p)^{n-k}$$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

3) $X \sim \mathcal{B}(n; p)$

$X = Y_1 + Y_2 + \cdots + Y_n$ with $Y_i \sim \mathcal{B}(p)$

$Y_1, \ldots, Y_n$: independant

$$\mathbb{E}[X] = \mathbb{E}[Y_1 + \cdots + Y_n]$$
$$= \mathbb{E}[Y_1] + \mathbb{E}[Y_2] + \cdots + \mathbb{E}[Y_n]$$

$$\boxed{\mathbb{E}[X] = np}$$

Poisson :

$$X \sim \mathcal{P}(\lambda) \qquad \lambda > 0$$

$$X(\Omega) = \mathbb{N}$$

$$\forall k \in \mathbb{N}, \quad P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Rk:

Let X and Y two discrete r.v.

$\forall i \in X(\Omega)$
$\forall j \in Y(\Omega)$

$$P(X=i, Y=j) = P(X=i) \times P(Y=j)$$

Properties:

let X and Y be two r.v

let $\lambda \in \mathbb{R}$

- $\mathbb{E}[X + \lambda Y] = \mathbb{E}[X] + \lambda \mathbb{E}[Y]$
- $V[\lambda X] = \lambda^2 V[X]$
- if X and Y are independant $V[X+Y] = V[X] + V[Y]$

[redacted]

Rq : $\mathcal{B}(p) = \mathcal{B}(1, p)$

Google : CRAN

Comprehensive R Archive Network

IV) Continuous r.v.

<u>Rk:</u> Let X be a continuous r.v.

$$P(X=a)=0$$

$\to$ density function

$f : \mathbb{R} \to \mathbb{R}$

- $\forall x \in \mathbb{R}, f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

Distribution function:

Let X be a continuous r.v.

$$F_X(t) = P(X \leq t) = P(X < t)$$

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

Let $a < b$

$$P(X \in [a,b]) = \int_a^b f(t)\,dt$$

where f is the density f^{ct} associated to X.

Rk:

$$P(X \in [a,b]) = P(X \in [a,b[)$$

$$= P(X \in]a,b])$$

$$= P(X \in]a,b[)$$

Expectation and variance:

let X be a continuous r.v.

If $\int_{-\infty}^{+\infty} |x| f(x)\, dx$ exists

then the expectation of X exists and

$$E[X] = \int_{-\infty}^{+\infty} x f(x)\, dx$$

If $\int_{-\infty}^{+\infty} x^2 f(x)\,dx$ exists

then the variance of X exists and

$$V[X] = \int_{-\infty}^{+\infty} \left(x - E[X]\right)^2 f(x)\,dx$$

$$= \int_{-\infty}^{+\infty} x^2 f(x)\,dx - E[X]^2$$

Properties

$X \sim P(\lambda)$

$\mathbb{E}[X] = \lambda$

$V[X] = \lambda$

properties:

The distribution f^{ct}:

- is a increasing f^{ct}
- $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to +\infty} F_X(x) = 1$
- is a continuous f^{ct}

Rk: Let X and Y be 2 continuous r.v

X and Y are independant if

$$f_{(X,Y)}(x,y) = f_X(x) \times f_Y(y)$$

For friday:

proof of $\mathbb{E}[X]$

for $X \sim \mathcal{E}(\lambda)$

Exponential

$X \sim \mathcal{E}(\lambda)$ $\quad \lambda > 0$

density fct

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

proof:

$$E[X] = \int_a^b \frac{1}{b-a} \times t \, dt$$

$$= \frac{1}{b-a} \int_a^b t \, dt$$

$$= \frac{1}{b-a} \times \left[\frac{t^2}{2}\right]_a^b = \frac{b^2-a^2}{(b-a)\times 2}$$

$$= \frac{b+a}{2}$$

Gaussian

$$X \sim \mathcal{N}(\underset{\text{expectation}}{\mu}, \underset{\text{variance}}{\sigma^2})$$

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

properties:

let $X \sim \mathcal{U}([a,b])$

- $\mathbb{E}[X] = \dfrac{a+b}{2}$

- $V[X] = \dfrac{(b-a)^2}{12}$

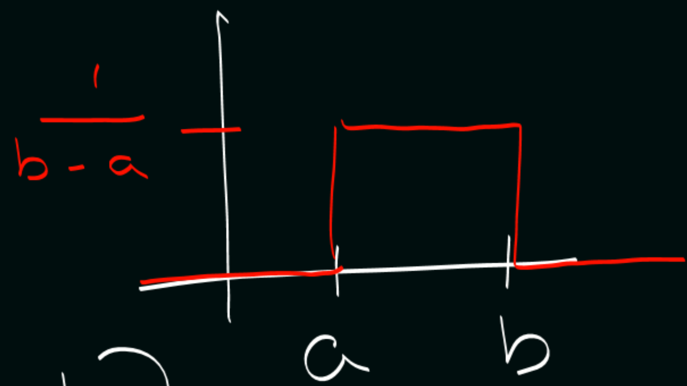
- $F_X(t) = \begin{cases} 0 & \text{if } t < a \\ \dfrac{t-a}{b-a} & \text{if } t \in [a,b] \\ 1 & \text{if } t > b \end{cases}$

IV Famous r.v

- uniform
- exponential
- gaussian = normal

a) Uniform

$X \sim \mathcal{U}([a,b])$



density f^{ct}:

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

properties:

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

$$V[X] = \frac{1}{\lambda^2}$$

Rk:

On the table, we get

$F_X(u)$ for $u \geq 0$

and $X \sim \mathcal{N}(0,1)$

How to compute $\mathbb{P}(X \in [a,b])$?

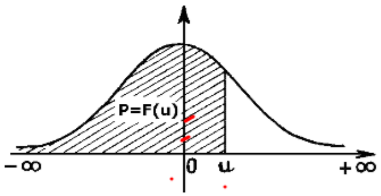
Properties:

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

$$Y = \frac{X-\mu}{\sigma} \sim \underbrace{\mathcal{N}(0,1)}_{\text{standard gaussian}}$$

FONCTION DE RÉPARTITION DE LA LOI NORMALE RÉDUITE

(Probabilité de trouver une valeur inférieure à u)



u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7290	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Table pour les grandes valeurs de u

u	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
F(u)	0,99865	0,99904	0,99931	0,99952	0,99966	0,99976	0,999841	0,999928	0,999968	0,999997

Exercise:

Compute for $X \sim \mathcal{N}(0;1)$

- $P(X \in [2;2.4])$
- $P(X \in [-1;1])$
- $P(X \geq -2)$

$$P(X \in [a,b]) = P(X \leq b) - P(X \leq a)$$
$$= F(b) - F(a)$$

? How to compute $F_X(u)$ for $u \leq 0$?

Let $Y \sim \mathcal{N}(0,1)$

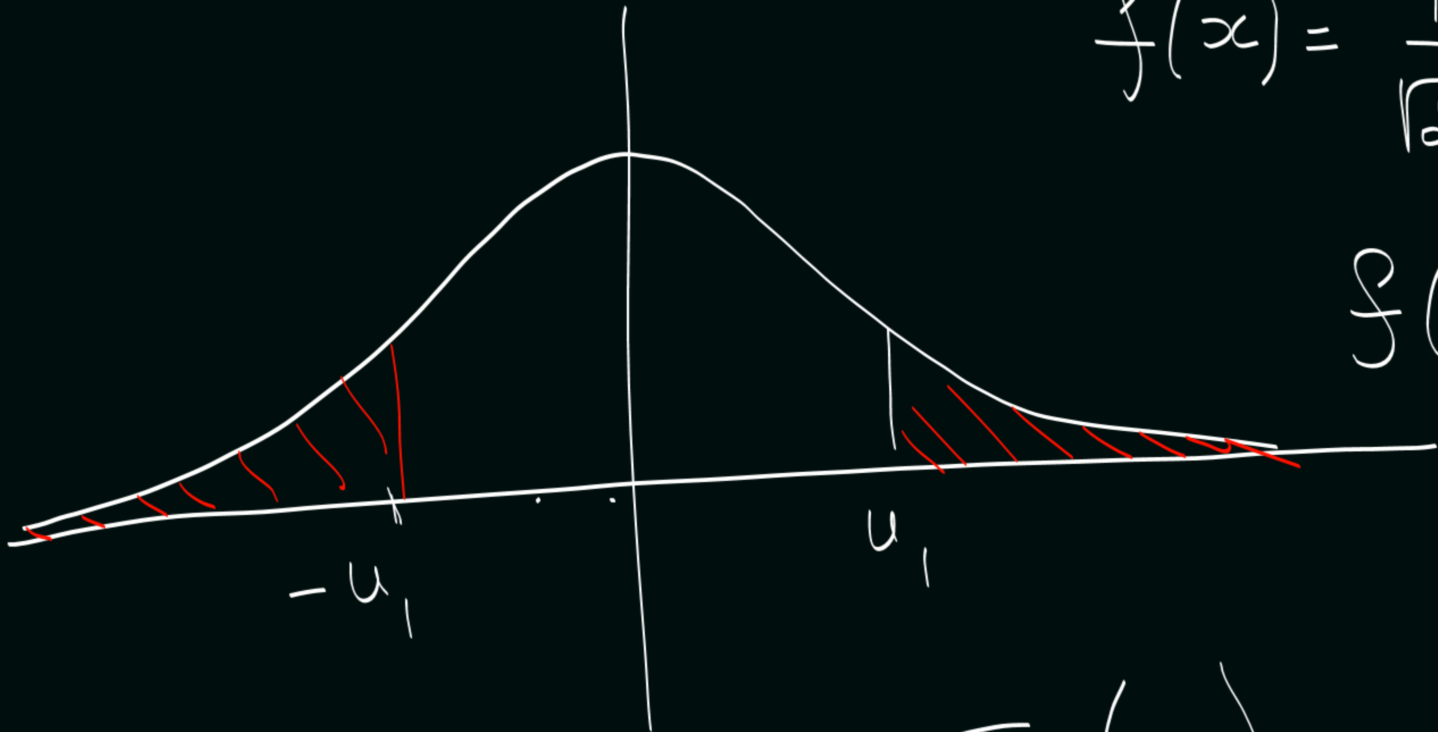
Let $\sigma > 0$ and $\mu \in \mathbb{R}$

$$X = \mu + \sigma Y$$

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

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}$$

$$f(x) = f(-x)$$



$$F_X(-u) = 1 - F_X(u)$$

We are not able to compute

$$\int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx$$

$\Rightarrow$ this quantity can be approximated

$\Rightarrow$ the approximations for a standard gaussian are in a table

How to make the computations with the R software?

for uniform: runif, punif, dunif

simulations of $\mathcal{U}([0;1])$

$$
\begin{aligned}
P(X \geq -2) &= 1 - P(X < -2) \\
&= 1 - F_X(-2) \\
&= 1 - (1 - F_X(2)) \\
&= F_X(2) \\
&= 0.9772
\end{aligned}
$$

$$P(X \in [2, 2.4])$$

$$= F_X(2.4) - F_X(2)$$

$$= 0.9918 - 0.9772$$

$$= 0.0146$$

$$P(X \in [-1,1])$$

$$= F_X(1) - F_X(-1)$$

$$= F_X(1) - (1 - F_X(1))$$

$$= 2 \cdot F_X(1) - 1$$

$$= 2 \times 0.8413 - 1 = 0.6826$$

Compute for $X \sim \mathcal{N}(2;4)$

- $P(X \leq 2) = 0.5$
- $P(X \in [2;4])$

$$X \sim \mathcal{N}(2;4) \Rightarrow Y = \frac{X-2}{2} \sim \mathcal{N}(0;1)$$

$$P(X \in [2;4]) = P(Y \in [0;1]) = F_Y(1) - F_Y(0)$$

$$= 0.8413 - 0.5 = 0.3413$$