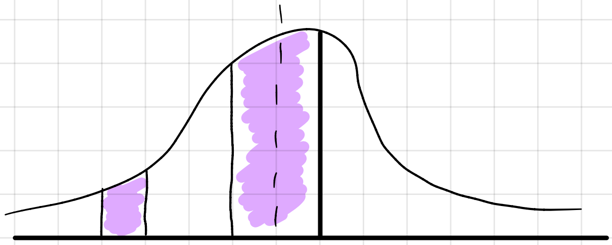
One dimensional normal distribution

Parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$

Density : $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Symbol: $N(\mu,\sigma^2) \Longrightarrow$ It means that the random variable $X \sim N(\mu,\sigma^2)$

For every Borel set $B \subset \mathbb{R}$ we have $P(X \in B) = \int_B \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$



Special case: Standard normal distribution $\mu = 0$ $\sigma^2 = 1$

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

What is standarization?

If $X \sim N(\mu,\sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$ (standard normal distribution)

CDF of standard normal distribution:

$F: \mathbb{R} \longrightarrow [0,1]$ given by $F(x) = P(X \leq x) = P(X \in]-\infty, x])$

$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt$

In the normal distribution ($X \sim N(\mu,\sigma^2)$), $E[X] = \mu$ and $Var(X) = \sigma^2$

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

The normal distribution is important because of the Central Limit Theorem (CLT)

Let $X_1, X_2, X_3, \dots$ be indepent random variables with the same probability distribution (i.i.d) such that $E[X_1] = \dots = E[X_n] = \mu \in \mathbb{R}$ and $Var(X_1) = \dots = Var(X_n) = \sigma^2 > 0$. Then the distribution of

$$\frac{X_1 + X_2 + \dots + X_n - n \cdot \mu}{\sqrt{n \cdot \sigma^2}}$$ converges to the SND.

Example : Uniform distribution

$a, b \in \mathbb{R} \quad a < b$

Density: $f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & x \notin (a,b) \end{cases}$

1. ONE-DIMENSIONAL NORMAL DISTRIBUTION

DEF: One-dimensional Normal distribution with parametres $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ is the probability distribution with the density:

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

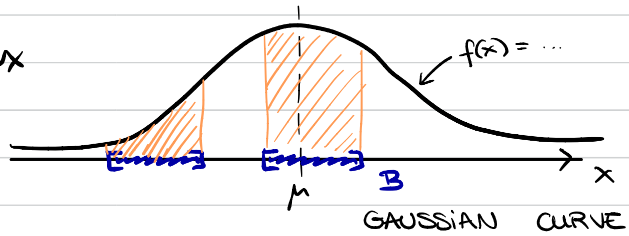
Its symbol is $\mathcal{N}(\mu,\sigma^2)$, although sometimes $\mathcal{N}(\mu,\sigma)$.

It means that the random variable $X \sim \mathcal{N}(\mu,\sigma^2)$ has the distribution

For every Borel set $B \subset \mathbb{R}$ ($\equiv$ for every interval $B \subset \mathbb{R}$) we have:

$$P(X \in B) = \int_B \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

$=$ area of



SPECIAL CASE: Standard Normal Distribution $\mathcal{N}(0,1)$

Where $\mu = 0$, $\sigma^2 = 1$. It is the distribution with the density:

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

→ The process of standarization is the following one:

Let $\mu \in \mathbb{R}$ and $\sigma^2 > 0$:

- if $X \sim \mathcal{N}(\mu,\sigma^2)$, then $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$
- If $Y \sim \mathcal{N}(0,1)$, then $\sigma Y + \mu \sim \mathcal{N}(\mu,\sigma^2)$

$\underbrace{\text{Cumulative Distribution Function}}^{\text{CDF}}$ of standard normal dist.

The CDF of a random variable X is the function $F: \mathbb{R} \to [0,1]$ given by:

we can define it with the inequality "<" (NOT EQUIVALENT)

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

→ Let $X \sim \mathcal{N}(0,1)$. Then, the CDF:

$$\phi(x) := F(x) = P(X \le x) = P(X \in]-\infty, x]) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt$$

→ If $X \sim \mathcal{N}(0,1)$, then $E(X) = \mu$ and $Var(X) = \sigma^2$

* $Var X = E((X - EX)^2)$

→ In particular, if $\mathcal{N}(0,1)$, then $EX = 0$ and $Var X = 1$.

Why the normal distribution is important?
Because of the Central Limit Theorem

THM: Central Limit Theorem (CLT) (one of many possible forms)

Let $X_1, X_2, \dots, X_n$ be independent random variables with the same probability distribution (i.i.d.) independent identically distributed
such that $EX_1 (= EX_2 = \cdots) = \mu \in \mathbb{R}$ and $Var X_1 (= Var X_2 = \cdots) = \sigma^2 > 0$.

Then, the distribution of $\dfrac{X_1 + X_2 + \cdots + X_n - n \cdot \mu}{\sqrt{n \cdot \sigma^2}} \to \mathcal{N}(0,1)$.

DEF: Convergence of probability distributions (weak conv.)

Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of probability distributions and let μ a probability distribution.

Let F_n be CDF of μ_n and F be CDF of μ.

We say that (μ_n) converges (weakly) to μ if and only if

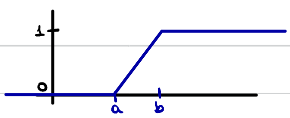
$\forall x \in \mathbb{R}$ where F is continuous at x, $\lim_{n \to \infty} F_n(x) = F(x)$

*

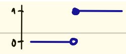
DEF: Let $a, b \in \mathbb{R}$, $a < b$. Uniform Distribution, noted by $\mathcal{U}(a,b)$, is the distribution with the density

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a,b) \\ 0 & \text{if } x \notin (a,b) \end{cases}$$

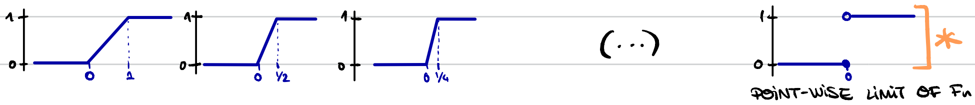
CDF of $\mathcal{U}nif(a,b)$ is given by



EX: Let $\mu_n = \mathrm{Unif}(0, 1/n)$. If $n \to \infty$, then $(0, 1/n)$ gets shorter.

We expect that (μ_n) converges (weakly) to $\mu = \delta_0$ (one-point probability distribution cumulated at 0; $X \sim \delta_0 \Leftrightarrow X = 0$ almost surely, as $P(X=0) = 1$). CDF of δ_0 is 

Let us see the CDF of μ_n, $n = 1, 2, \dots$



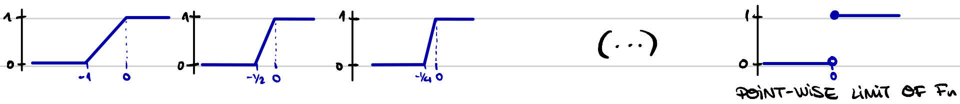
We can see the necessity of the condition *. This function is not CDF of δ_0. It is not CDF of any probability distribution because it is <u>not</u> right-continuous.

It does <u>not</u> converge.

EX: Let $\mu_n = \mathrm{Unif}(-1/n, 0)$. If $n \to \infty$, then $(-1/n, 0)$ gets shorter.

We expect that (μ_n) converges (weakly) to $\mu = \delta_0$.

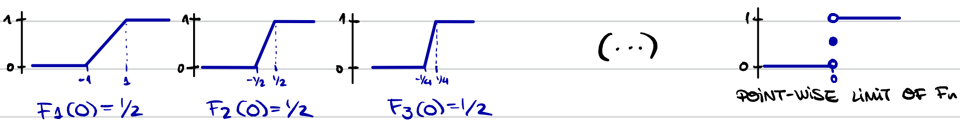
Let us see the CDF of μ_n, $n = 1, 2, \dots$



Surprisingly, this time, $f(x) = F(x)$ for every $x \in \mathbb{R}$.

EX: Let $\mu_n = \mathrm{Unif}(-1/n, 1/n)$. If $n \to \infty \Rightarrow (-1/n, 1/n)$ gets shorter.

Let us see the CDF of μ_n, $n = 1, 2, \dots$



Now, we have that $\forall x \neq 0$, $\lim_{n\to\infty} F_n(x) = F(x)$, but $F(0) = 1$ is different from $\lim_{n\to\infty} F_n(0) = 1/2$.

EX: Let $\mu_n = \begin{cases} \text{unif}\,(0, 1/n) & \text{for } n \text{ odd} \\ \text{unif}\,(-1/n, 0) & \text{for } n \text{ even} \end{cases}$

We have $\forall x \neq 0$, $\lim_{n\to\infty} F_n(x) = F(x)$. Anyway, $\nexists \lim_{n\to\infty} F_n(0)$.

DEF: Two other (equivalent) definitions of weak convergence of p.d.:

Let (μ_n) be a sequence of p.d. and μ a p.d.. Let $X, X_1, X_2 \ldots$ be random variables such as $X_n \sim \mu_n$ and $X \sim \mu$.

The sequence μ_n converges (weakly) to μ if and only if:

- $\underset{\substack{B \subset \mathbb{R} \\ \text{borel set}}}{\forall} P(X \in \overset{\text{boundary}}{\partial B}) = 0 \Rightarrow \lim_{n\to\infty} \underbrace{P(X_n \in B)}_{= \mu_n(B)} = \underbrace{P(X \in B)}_{= \mu(B)}$

  $\partial B = \bar{B} \setminus \mathring{B}$

- $\underset{\substack{f:\mathbb{R}\to\mathbb{R} \\ \text{bounded} \\ \text{continuous}}}{\forall} \lim_{n\to\infty} Ef(X_n) = Ef(X)$

... coming back to CTL. Let $X_1, X_2, \ldots$ be i.i.d.
$EX_1 (= EX_2 = \cdots) = \mu \in \mathbb{R}$
$\text{Var}\,X_1 (= \text{Var}\,X_2 = \cdots) = \sigma^2 > 0$.

The sequence of distributions of $\dfrac{\sum_{i=n}^{n} X_i - \mu n}{\sqrt{n\sigma^2}}$ is convergent to $\mathcal{N}(0,1)$. What does this mean?

It means that

$$\underset{x \in \mathbb{R}}{\forall} \text{ ~~where } \phi \text{ is cont~~} \Rightarrow P\left(\frac{\sum_{i=n}^{n} X_i - \mu n}{\sqrt{n\sigma^2}} \leq x\right) \xrightarrow{n\to\infty} \underbrace{\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt}_{\text{CDF of } \mathcal{N}(0,1)}$$

We do not need this condition because ϕ is cont, as it is the integral of a cont. fonction. Every point $x \in \mathbb{R}$ is a continuous point of ϕ.

This means that the above convergence is the point-wise convergence. It can be shown that we can have <u>uniform</u> convergence here.

Some comments about CTL:

→ From the Strong Law of Large Numbers, we know that:

$$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[n\to\infty]{\text{almost surely}} \mu \iff \frac{\sum_{i=1}^{n} X_i - n\mu}{n} \xrightarrow[n\to\infty]{\text{almost surely}} 0$$

in CTL we have different conv.

in CTL we have $\mathcal{N}(0,1)$

in CTL we have $\sqrt{n\sigma^2}$ here (smaller number)

→ $$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow[n\to\infty]{\text{weakly}} \mathcal{N}(0,1) \iff \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{weakly}} \mathcal{N}(0,\sigma^2)$$

(CTL)

→ Why those formulas? Let's see:

Let X random variable with expectation EX and variance Var X ≠ 0. Then:

- $$E\,\frac{X-EX}{\sqrt{\operatorname{Var}X}} = \frac{1}{\sqrt{\operatorname{Var}X}}\,E(\underset{\text{random}}{X} - \underset{\text{constant}}{EX}) = \frac{1}{\sqrt{\operatorname{Var}X}}\overbrace{(EX - \underbrace{E(EX)}_{EX})}^{\to 0} = 0$$

- $$\operatorname{Var}\frac{X-EX}{\sqrt{\operatorname{Var}X}} = \left(\frac{1}{\sqrt{\operatorname{Var}X}}\right)^2 \operatorname{Var}(X - \underset{\text{constant (we can omit it)}}{EX}) = \frac{1}{\operatorname{Var}X}\operatorname{Var}X = 1$$

Now, let $X = \sum_{i=1}^{n} X_i$. Then:

- $EX = E(\sum X_i) = \sum EX_i \overset{\text{identical}}{=} \sum \mu = n\mu$

- $\operatorname{Var} X = \operatorname{Var}(\sum X_i) \overset{\text{independent}}{=} \sum \operatorname{Var} X_i = \sum \sigma^2 = n\sigma^2$

It follows that $$\frac{X-EX}{\sqrt{\operatorname{Var}X}} = \underbrace{\frac{\sum X_i - n\mu}{\sqrt{n\sigma^2}}}_{*}$$

E * = 0
Var * = 1
* is the expression of CTL.

According to the CTL, for large n: $$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \approx \mathcal{N}(0,1)$$

$$\sum_{i=1}^{n} X_i - n\mu \approx \mathcal{N}(0, n\sigma^2)$$

$$\sum_{i=1}^{n} X_i \approx \mathcal{N}(n\mu, n\sigma^2).$$

COR: We see that if a random variable is the sum of large number of "small" independent summands, then the distribution of this random variable is (approx.) normal.

2. MULTIDIMENSIONAL NORMAL DISTRIBUTION

DEF: We say that a random vector $(X_1, \ldots, X_n)$ has the n-dimensional normal distribution if, for every $a_1, \ldots, a_n \in \mathbb{R}$, the r.v. $a_1 X_1 + \cdots + a_n X_n$ has one-dimensional normal dist.

An equivalent definition could be:
$\mathbb{X} = (X_1, \ldots, X_n)$ has n-dimensional normal distribution if, for each linear functional $\varphi: \mathbb{R}^n \to \mathbb{R}$, the r.v. $\varphi(\mathbb{X})$ has one-dimensional normal distribution.

Remark: In the above definitions, we assume that $\mathcal{N}(\mu, 0) := \delta_\mu$ has also a normal dist.

$X \sim \mathcal{N}(\mu, 0) = \delta_\mu \iff P(X=\mu) = 1$ (one-point probability dist.)

PROP: $\mathcal{N}(\mu, 0)$ is the degenerated normal distribution.

Let's see a comparison table:

One dimension	n-dimension
$\mathcal{N}(\underset{\text{expectation}}{\overset{\in\mathbb{R}}{\mu}}, \underset{\text{variance}}{\overset{\in\mathbb{R}^+}{\sigma^2}})$	$\mathcal{N}(\underset{\text{"expectation"}}{?}, \underset{\text{"variance"}}{?})$

2.1. EXPECTATION AND VARIANCE IN MULTIPLE DIM.

Let $\mathbb{X} = (X_1, \ldots, X_n)^T$.

DEF: The expectated value $E\mathbb{X}$ is defined as follows:

$$E\mathbb{X} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix} \qquad E\mathbb{X} \text{ exists} \iff \exists\, E\mathbb{X}_1, \ldots, E\mathbb{X}_n$$

What about $\operatorname{Var} \mathbb{X}$? Unfortunately, $(\operatorname{Var} X_1, \ldots, \operatorname{Var} X_n)^T$ is <u>not</u> what we need.

DEF: $\operatorname{Var} \mathbb{X}$

$$\operatorname{Var} \mathbb{X} = \begin{pmatrix} \operatorname{Cov}(X_1, X_1) & \operatorname{Cov}(X_1, X_2) & \cdots & \operatorname{Cov}(X_1, X_n) \\ \operatorname{Cov}(X_2, X_1) & \operatorname{Cov}(X_2, X_2) & \cdots & \operatorname{Cov}(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ \operatorname{Cov}(X_n, X_1) & \operatorname{Cov}(X_n, X_2) & \cdots & \operatorname{Cov}(X_n, X_n) \end{pmatrix} =$$

$$= \left(\operatorname{Cov}(X_i, X_j)\right)_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}}$$

Remember that, for X and Y r.v.:

$$\mathrm{Var}\, X = E((X-EX)^2) = E(X^2) - (EX)^2$$

$$\mathrm{Cov}(X,Y) = E((X-EX)(Y-EY)) = E(XY) - E(X)E(Y)$$

PROP: Properties of the Covariance and the Cov matrix

1. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X) \Rightarrow \mathrm{Var}(\underline{X}) = \mathrm{Var}(\underline{X})^T$
2. $\mathrm{Cov}(X,X) = E((X-EX)(X-EX)) = \mathrm{Var}\, X$

   → The diagonal of $\mathrm{Var}\,\underline{X}$ consists of $\mathrm{Var}\, X_1, \dots, \mathrm{Var}\, X_n$.

Let $a \in \mathbb{R}$, X, Y, Z be r.v. Then,

$$\left.\begin{aligned}
&3.\ \mathrm{Cov}(aX,Y) = \mathrm{Cov}(X,aY) = a\,\mathrm{Cov}(X,Y)\\
&4.\ \mathrm{Cov}(X+Y,Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)\\
&5.\ \mathrm{Cov}(X,Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)
\end{aligned}\right\}\ \text{Covariance is 2-linear}$$

6. $\mathrm{Cov}(X,a) = \mathrm{Cov}(a,X) = 0$.
7. $\mathrm{Cov}(X+a,Y) = \mathrm{Cov}(X,Y+a) = \mathrm{Cov}(X,Y)$
8. $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X,Y)$

   ↘ generally:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\, X_i + 2\sum_{i<j} \mathrm{Cov}(X_i,X_j)$$

Some proofs:

3. $\mathrm{Cov}(aX,Y) = E((aX - \overbrace{EaX}^{a\,EX})(Y-EY)) = E(a(X-EX)(Y-EY)) =$
$= a\,E((X-EX)(Y-EY)) = a\,\mathrm{Cov}(X,Y)$

4. $\mathrm{Cov}(X+Y,Z) = E((X+Y-\overbrace{E(X+Y)}^{EX-EY})(Z-EZ)) =$
$= E([(X-EX)+(Y-EY)](Z-EZ)) = E((X-EX)(Z-EZ)+(Y-EY)(Z-EZ)) =$
$= \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$.

6. $\mathrm{Cov}(X,a) = E((X-EX)(a-\overbrace{Ea}^{a})) = E((X-EX)\cdot 0) = E0 = 0$

7. $\mathrm{Cov}(X, Y+a) \overset{4}{=} \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,a) \overset{6}{=} \mathrm{Cov}(X,Y)$

8. $\mathrm{Var}(X \pm Y) = \mathrm{Cov}(X\pm Y, X \pm Y) = \mathrm{Cov}(X, X\pm Y) \pm \mathrm{Cov}(Y, X \pm Y) =$
$= [\mathrm{Cov}(X,X) \pm \mathrm{Cov}(X,Y)] \pm [\mathrm{Cov}(Y,X) \pm \mathrm{Cov}(Y,Y)] =$
$= \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X,Y)$.

COR: 1 - $\mathrm{Var}(X+a) = \mathrm{Cov}(X+a, X+a) = \mathrm{Cov}(X,X) = \mathrm{Var}(X)$

2 - $\mathrm{Var}(aX) = \mathrm{Cov}(aX, aX) = a^2 \mathrm{Cov}(X,X) = a^2 \mathrm{Var}(X)$

More about the covariance matrix, $\mathrm{Var}\,\mathbb{X}$:

Knowing that for X a r.v., $\mathrm{Var}\,X = E((X-EX)^2)$, then, we have:

vector x vector = matrix

$$\mathrm{Var}\,\mathbb{X} = E((\mathbb{X} - E\mathbb{X})(\mathbb{X} - E\mathbb{X})^T)$$

Indeed:

$$E((\mathbb{X}-E\mathbb{X})(\mathbb{X}-E\mathbb{X})^T) = E\left(\begin{pmatrix} X_1 - EX_1 \\ X_2 - EX_2 \\ \vdots \\ X_n - EX_n \end{pmatrix} (X_1 - EX_1, \dots, X_n - EX_n)\right) =$$

$$= E\left(\begin{pmatrix} (X_1-EX_1)(X_1-EX_1), & \dots, & (X_1-EX_1)(X_n-EX_n) \\ \vdots & & \vdots \\ (X_n-EX_n)(X_1-EX_1), & \dots, & (X_n-EX_n)(X_n-EX_n) \end{pmatrix}\right) = \mathrm{Var}\,\mathbb{X}.$$

THM: If $\mathrm{Var}\,\mathbb{X}$ is the covariance matrix of a r.vector $\mathbb{X}$, then $\mathrm{Var}\,\mathbb{X}$ is positive semidefinite.

Proof: We need to show that for each $\mathbb{t} = (t_1, \dots, t_n)^T$, we have $\mathbb{t}^T \cdot \mathrm{Var}\,\mathbb{X} \cdot \mathbb{t} \geq 0$.

$$\mathbb{t}^T \cdot \mathrm{Var}\,\mathbb{X} \cdot \mathbb{t} = (t_1, \dots, t_n) \cdot \mathrm{Var}\,\mathbb{X} \cdot \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} = \sum_{i=1}^n \sum_{j=1}^n t_i \,\mathrm{Cov}(X_i, X_j)\, t_j =$$

$$= \sum_{i=1}^n \mathrm{Cov}\left(t_i X_i, \sum_{j=1}^n t_j X_j\right) = \mathrm{Cov}\left(\sum_{i=1}^n t_i X_i, \sum_{j=1}^n t_j X_j\right) = \mathrm{Var}\left(\sum_{i=1}^n t_i X_i\right) \geq 0$$

$\dagger$

PROP: We see that, if $\mathbb{X} = (X_1, \dots, X_n)^T$ is a r.vector such that the covariance matrix $\text{Var}\,\mathbb{X}$ exists, then:

1. $\text{Var}\,\mathbb{X}$ is a matrix of dimensions $n \times n$
2. $\text{Var}\,\mathbb{X}$ is symetric: $(\text{Var}\,\mathbb{X})^T = \text{Var}\,\mathbb{X}$
3. $\text{Var}\,\mathbb{X}$ is positive semidefinite.

It can be shown that for each matrix A with the above properties, there exists a r.vector $\mathbb{X}$ satisfying $\text{Var}\,\mathbb{X} = A$.

Problem. Let X_1 and X_2 be two independent random variables such that $E[X_1] = E[X_2] = m$. The value of m is unknown. We know $Var(X_1) = \sigma_1^2 > 0$ and $Var(X_2) = \sigma_2^2 > 0$. Find numbers $a_1, a_2, b \in \mathbb{R}$ such that the random variable $\tilde{X} = a_1X_1 + a_2X_2 + b$ satisfies:

$E[\tilde{X}] = m$

$Var(\tilde{X})$ is as small as possible

$$E[a_1X_1 + a_2X_2 + b] = \underbrace{a_1E[X_1] + a_2E[X_2] + b}_{\underbrace{m(a_1+a_2)+b = m}_{\underbrace{m(a_1+a_2-1)}\ =\ \underbrace{-b}}} = m$$

They don't depend on m

$m(a_1+a_2-1)$ is a linear function of m
$-b$ is const $\Bigg\}$ So $\underbrace{a_1+a_2-1=0}_{\substack{a_1+a_2=1\\ b=0}}$

$$Var\tilde{X} = Var(a_1X_1 + a_2X_2) \overset{X_1, X_2 \text{ independent}}{=} Var(a_1X_1) + Var(a_2X_2) =$$

$$= a_1^2 Var(X_1) + a_2^2 Var(X_2) = \underbrace{a_1^2\sigma_1^2 + a_2^2\sigma_2^2}_{\text{Minimize}}$$

$$a_1 + a_2 = 1 \Rightarrow a_2 = 1 - a_1$$

$$a_1^2\sigma_1^2 + a_2^2\sigma_2^2 = \underbrace{a_1^2\sigma_1^2 + (1-a_1)^2\sigma^2}_{(\sigma_1^2+\sigma_2^2)a_1^2 - 2\sigma_2^2a_1 + \sigma_2^2}$$

The minimum is obtained for $x = -\frac{b}{2a}$

$$x = a_1, \quad -\frac{b}{2a} = -\frac{-2\sigma_2^2}{2(\sigma_1^2+\sigma_2^2)} = \frac{\sigma_2^2}{(\sigma_1^2+\sigma_2^2)}$$

And finally, $a_2 = \frac{\sigma_1^2}{(\sigma_1^2+\sigma_2^2)}$

$$Var\tilde{X} = \frac{\sigma_1^2\sigma_2^4}{(\sigma_1^2+\sigma_2^2)^2} + \frac{\sigma_1^4\sigma_2^2}{(\sigma_1^2+\sigma_2^2)^2} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

Generalization of the obtained result

n independent random variables $X_1, \dots, X_n$ satisfying $E[X_1] = \dots = E[X_n] = m$ (m unknown). We know $\operatorname{Var} X_i = \sigma_i^2 > 0$ from $i = 1, 2, \dots, n$. Find $a_1, \dots, a_n, b \in \mathbb{R}$ such that $\tilde{X} = \sum_{i=1}^{n} a_i X_i + b$, $E[\tilde{X}] = m$ and $\operatorname{Var}\tilde{X}$ is as small as possible

Using a similar method we obtain that $b = 0$ and

$$a_i = \frac{\frac{1}{\operatorname{Var} X_i}}{\sum_{k=1}^{n} \frac{1}{\operatorname{Var} X_k}}$$

$\tilde{X}$ is the weighted mean of $X_1, \dots, X_n$ with the weights proportional to reciprocals of variances

Moreover, $\operatorname{Var}\tilde{X} = \dfrac{1}{\sum_{k=1}^{n} \frac{1}{\operatorname{Var} X_k}}$

$X_1, \dots, X_n$ i.r.v. with the same distribution ($E[X_i] = m$ for $i = 1, \dots, n$) and $\operatorname{Var} X_i = \sigma^2$ for $i = 1, \dots, n$). We want to know m and σ^2. The unbiased estimation of m is $\hat{m} = \overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$

$\hookrightarrow E[\overline{X}] = m$

The unbiased estimation of σ^2 are

$$E[S^2] = E[S^2] = \sigma^2 \qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \qquad \left(\begin{matrix} n > 0 \\ \text{we do not know } m \end{matrix}\right)$$

$$\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - m)^2 \qquad \text{(we know } m\text{)}$$

Problem. Let $X_1, X_2, \dots, X_n$ be independent r.v.s such that $E[X_1] = E[X_2] = \dots = E[X_n] = m$ (m unknown) and $\operatorname{Var} X_i = \frac{\sigma^2}{w_i}$, where $\sigma^2 > 0$ is unknown and $w_1, w_2, \dots, w_n > 0$ are known.

Give examples of "good" unbiased estimators of m and σ^2.

We will start with m. We can try to use the answer to the previous problem

$$\hat{m} = \tilde{X} = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

$$a_i = \frac{\frac{1}{\operatorname{Var} X_i}}{\sum_{k=1}^{n} \frac{1}{\operatorname{Var} X_k}} \leftarrow \text{Unfortunately, we do not know } \operatorname{Var} X_i = \frac{\sigma^2}{w_i}$$

$$= \frac{\frac{w_i}{\sigma^2}}{\sum_{k=1}^{n} \frac{w_k}{\sigma^2}} = \frac{w_i}{\sum_{k=1}^{n} w_k} \leftarrow \text{It does not depend on } \sigma^2$$

It follows that $\hat{\omega} = \tilde{X} = \dfrac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$ is the best (Var ~ min) linear (of the form $\sum a_i X_i + b$) and unbiased ($E[\hat{\omega}] = \omega$) estimator of ω

Remark: If $\operatorname{Var} X_1 = \dots = \operatorname{Var} X_n = \sigma^2$ ($w_1 = \dots = w_n = 1$), then we obtain

$$\hat{\omega} = \tilde{X} = \frac{\sum_{i=1}^{n} 1 \cdot X_i}{\sum_{i=1}^{n} 1} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i = \bar{X}$$

Let's try to estimate σ. We want to mimic

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Let's try to find the constant c, such that $\widehat{\sigma^2} = c \cdot \sum_{i=1}^{n} w_i \cdot (X_i - \tilde{X})^2$ is an unbiased estimation of σ^2.

We want $E[\widehat{\sigma^2}] = \sigma^2$

$$c \cdot \underbrace{E\left(\sum_{i=1}^{n} w_i (X_i - \tilde{X})^2 \right)}_{\sum_{i=1}^{n} w_i E[(X_i - \tilde{X})^2]} = \sigma^2$$

$$\sum_{i=1}^{n} w_i \underbrace{E[(X_i - \tilde{X})^2]}_{\operatorname{Var}(X_i - \tilde{X}) + (E[X_i - \tilde{X}])^2}$$

$$\underbrace{\operatorname{Var}(X_i - \tilde{X})}_{\operatorname{Var} X_i + \operatorname{Var} \tilde{X} - 2 \operatorname{Cov}(X_i, \tilde{X})} + \big(\cancel{\underbrace{E[X_i - \tilde{X}]}_{E[X_i] - E[\tilde{X}]}}\big)^2 \to 0, \quad \underbrace{E[X_i]}_{m} - \underbrace{E[\tilde{X}]}_{m}$$

$$\underbrace{\operatorname{Var} X_i + \operatorname{Var} \tilde{X} - 2 \operatorname{Cov}(X_i, \tilde{X})}_{\frac{\sigma^2}{w_i} + \frac{\sigma^2}{\sum_{i=1}^{n} w_i} - 2 \frac{\sigma^2}{\sum_{i=1}^{n} w_i}} \quad {*}\ \Delta$$

$$\frac{\sigma^2}{w_i} + \frac{\sigma^2}{\sum_{i=1}^{n} w_i} - 2 \frac{\sigma^2}{\sum_{i=1}^{n} w_i} = \frac{\sigma^2}{w_i} - \frac{\sigma^2}{\sum_{i=1}^{n} w_i}$$

$$* \ \operatorname{Cov}(X_i, \tilde{X}) = \operatorname{Cov}\left(X_i, \frac{\sum_{k=1}^{n} w_k X_k}{\sum_{k=1}^{n} w_k}\right) = \frac{\sum_{k=1}^{n} w_k \operatorname{Cov}(X_i, X_k)}{\sum_{k=1}^{n} w_k} = \begin{cases} \operatorname{Var} X_i = \frac{\sigma^2}{w_i} & \text{if } k = i \\ 0 & \text{if } k \neq i \end{cases}$$

$$\Delta \ \operatorname{Var} \tilde{X} = \operatorname{Var} \frac{\sum_{k=1}^{n} w_k X_k}{\sum_{k=1}^{n} w_k} = \frac{1}{(\sum w_k)^2} \cdot \operatorname{Var}\left(\sum w_k X_k\right) \underset{\text{Independent}}{=} \frac{1}{(\sum w_k)^2} \cdot \operatorname{Var} \sum_{i=1}^{n} w_k^2 \operatorname{Var} X_k = \frac{1}{(\sum w_k)^2} \sum_{k=1}^{n} w_k^2 \cdot \frac{\sigma^2}{w_k} =$$

$$= \frac{\sigma^2 \sum_{k=1}^{n} w_k}{\left(\sum_{i=1}^{n} w_k\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} w_k}$$

$$E\left(\sum_{i=1}^{n} w_i (X_i-\tilde{X})^2\right) = \sum_{i=1}^{n} w_i E[(X_i-\tilde{X})^2] = \sum_{i=1}^{n} w_i\left(\frac{\sigma^2}{w_i} - \frac{\sigma^2}{\sum_{k=1}^{n} w_k}\right) = \sum_{i=1}^{n}\sigma^2 - \frac{\sigma^2 \cdot \sum_{i=1}^{n} w_i}{\sum_{k=1}^{n} w_k} = n\sigma^2-\sigma^2$$

So $c(n-1)\sigma^2 = \sigma^2 \Rightarrow c = \frac{1}{n-1} \Rightarrow$ The unbiased estimation of σ^2 is $\hat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n} w_i (X_i-\tilde{X})^2$

Remark: If $\mathrm{Var}X_1 = \dots = \mathrm{Var}X_n = \sigma^2$ $(\equiv w_1 = w_2 = \dots = w_n = 1)$, then we obtain

$$\hat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n} 1\cdot(X_i-\tilde{X})^2 = S^2$$

Let X, Y – random variables

Assume that $\mathrm{Var}X$ and $\mathrm{Var}Y$ exist. Then

If X,Y are independent $\underset{\not\Leftarrow}{\Rightarrow}$ $\mathrm{Cov}(X,Y) = 0$ (X,Y are not correlated)

$\iff \mathrm{Var}(X \pm Y) = \mathrm{Var}X \pm \mathrm{Var}Y$

Special case, when $\mathrm{Cov}(X,Y) = 0 \Rightarrow X,Y$ indep.

Assume that $\binom{X}{Y}$ has 2-dimensional normal distribution.

Then $\mathrm{Cov}(X,Y) = 0 \iff X,Y$ independent

It is not enough to assume that X,Y have normal distributions separately. We need to assume that the joint (two-dimensional) distribution of (X,Y) is normal

Problem

Assume that X and Y follow the normal distribution N(0,1) and are independent. Find real numbers a,b such that X+2Y+3 and X+aY+b are independent random variables.

Let's try to see that $(X+2Y+3, X+aY+b)^T$ has 2-dim. normal distribution.

X,Y have normal distributions and are independent

$\Downarrow$

random vector $\binom{X}{Y}$ has normal distribution

$\Downarrow$

$$\begin{pmatrix} X+2Y+3 \\ X+aY+b \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 2 \\ 1 & a \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} 3 \\ b \end{pmatrix}}_{\text{affine transformation of } \binom{X}{Y}}$$

It follows that

$X+2Y+3$ and $X+aY+b$ are independent $\iff \mathrm{Cov}(X+2Y+3, X+aY+b)=0$

$$\mathrm{Cov}(X+2Y+3, X+aY+b) = \mathrm{Cov}(X+2Y, X+aY) =$$

Constant

$$= \underbrace{\mathrm{Cov}(X,X)}_{\substack{\mathrm{Var}X \\ \parallel \\ 1}} + a\cancel{\mathrm{Cov}(X,Y)}^{0} + 2\cancel{\mathrm{Cov}(Y,X)}^{0} + 2a\underbrace{\mathrm{Cov}(Y,Y)}_{\substack{\mathrm{Var}(Y) \\ \parallel \\ 1}} =$$

$$= 1 + 2a$$

So we need $a = -\frac{1}{2}$ and $b \in \mathbb{R}$

Thm. Assume that $(X_1, X_2, \dots, X_n)^T$ has normal distribution. Then

$X_1, \dots, X_n$ are independent $\iff \forall_{\substack{i,j \in \{1,2,\dots,n\} \\ i \neq j}} \mathrm{Cov}(X_i, X_j) = 0$

$\Updownarrow$

$\mathrm{Var}(X_1, \dots, X_n)^T$ is a diagonal matrix

Thm. Assume that $(X_1, X_2, \dots, X_n, Y_1, \dots, Y_m)^T$ is a random vector with normal distribution. Then

$X = (X_1, X_2, \dots, X_n)^T$
$Y = (Y_1, Y_2, \dots, Y_m)^T$ are independent

$\Updownarrow$

$\forall_{i=1,2,\dots,n} \forall_{j=1,\dots,m} \mathrm{Cov}(X_i, Y_j) = 0 \iff$

$\mathrm{Var}(X_1, \dots, X_n, Y_1, \dots, Y_m)$ has the form

$$\begin{array}{c} n\{ \\ m\{ \end{array} \left(\begin{array}{c|c} \text{sth} & 0 \\ \hline 0 & \text{sth} \end{array}\right)$$
$$\qquad\quad \underbrace{}_{n} \underbrace{}_{m}$$

Problem

The random vector $(X,Y,Z)^T$ has distribution

$\mathcal{N}\left(\begin{pmatrix}1\\0\\3\end{pmatrix},\begin{pmatrix}2&-1&0\\-1&4&0\\0&0&3\end{pmatrix}\right)$. Are there constants

$a,b\in\mathbb{R}$ such that the random variable $X+Y+aZ+3$ is

independent of the random vector $(X+2Z+4,\, X-bY+Z)$? If so,

determine these constants.

First, we will try to show that $\begin{pmatrix}X+Y+aZ+3\\X+2Z+4\\X-bY+Z\end{pmatrix}$ has normal distribution:

$$\begin{pmatrix}X+Y+aZ+3\\X+2Z+4\\X-bY+Z\end{pmatrix}=\underbrace{\begin{pmatrix}1&1&a\\1&0&2\\1&-b&1\end{pmatrix}\begin{pmatrix}X\\Y\\Z\end{pmatrix}+\begin{pmatrix}3\\4\\0\end{pmatrix}}_{\text{affine transformation of }\begin{pmatrix}X\\Y\\Z\end{pmatrix}}$$

As a consequence we see that has a normal distribution.
It follows that $X+Y+aZ+3$ is independent of $\begin{pmatrix}X+2Z+4\\X-bY+Z\end{pmatrix}$ if and only if $\operatorname{Cov}(X+Y+aZ+3,\,X+2Z+4)=0$
and
$\operatorname{Cov}(X+Y+aZ+3,\,X-bY+Z)=0$

$$\operatorname{Cov}(X+Y+aZ+3,X+2Z+4)=\operatorname{Cov}(X+Y+aZ,X+2Z)=\underbrace{\operatorname{Var}X}_{2}+\cancelto{0}{2\operatorname{Cov}(X,Z)}+\underbrace{\operatorname{Cov}(Y,X)}_{-1}$$
$$+\cancelto{0}{2\operatorname{Cov}(Y,Z)}+\cancelto{0}{a\operatorname{Cov}(Z,X)}+2a\underbrace{\operatorname{Var}Z}_{3}=1+6a \;\hookrightarrow\; a=-\frac{1}{6}$$

$$\operatorname{Cov}(X+Y+aZ+3,X-bY+Z)=\operatorname{Cov}(X+Y+aZ,X-bY+Z)=\underbrace{\operatorname{Var}X}_{2}-b\underbrace{\operatorname{Cov}(X,Y)}_{-1}+$$
$$+\underbrace{\operatorname{Cov}(Y,X)}_{-1}-b\underbrace{\operatorname{Var}Y}_{4}+a\underbrace{\operatorname{Var}Z}_{3}=1-3b+3\cdot\underbrace{\left(-\tfrac{1}{6}\right)}_{a}$$

$$=\frac{1}{2}-3b\implies b=\frac{1}{6}$$

The random variable $X+Y+aZ-3$ is independent of $\begin{pmatrix}X+2Z+4\\X-bY+Z\end{pmatrix}$

if and only if $a=-\frac{1}{6}$ and $b=\frac{1}{6}$

Problem

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & -3 \\ -3 & 5 \end{pmatrix}\right)$$

Compute $Cov(X,Y)$, $E[XY]$, $Cov(X^2,Y^2)$

$Cov(X,Y) = -3$

Since $Cov(X,Y) = E[XY] - E[X]\cdot E[Y]$

$$\hookrightarrow E[XY] = \underbrace{Cov(X,Y)}_{-3} + \underset{0}{\underset{\parallel}{E[X]}}\cdot \underset{0}{\underset{\parallel}{E[Y]}} = -3$$

$$Cov(X^2,Y^2) = E[X^2Y^2] - \underset{(\underset{2}{VarX}+\underset{0^2}{E[X]^2})}{\underset{\parallel}{E[X^2]}}\cdot \underset{(\underset{5}{VarY}+\underset{0^2}{E[Y]^2})}{\underset{\parallel}{E[Y^2]}} = E[X^2Y^2] - 10$$

We also know that $Var(XY) = E[(XY)^2] - E[XY]^2 =$

$= E[X^2Y^2] - (-3)^2 = E[X^2Y^2] - 9$. But $E[X^2Y^2] = ?$

Let's try to standarize the random vector $\begin{pmatrix} X \\ Y \end{pmatrix}$. Let's find a matrix L (2×2 real matrix) and a vector b such that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{L}\cdot\begin{pmatrix} u \\ v \end{pmatrix} + b \text{ where } u,v \sim \mathcal{N}(0,1) \text{ independent}$$

We have $b = E\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\quad \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{L}\cdot\begin{pmatrix} u \\ v \end{pmatrix}$

$Var\begin{pmatrix} X \\ Y \end{pmatrix} = \mathcal{L}\cdot\mathcal{L}^T$

Let $\mathcal{L} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We want to have $\underbrace{\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} 2 & -3 \\ -3 & 5 \end{pmatrix}}$

$a^2 + b^2 = 2$

$ac + bd = -3$

$c^2 + d^2 = 5$

We want any of the solutions.
We do not need all of them.
Let $b = 0$. $a = \pm\sqrt{2} \Rightarrow a = \sqrt{2}$
$\sqrt{2}\cdot c + 0\cdot d = -3 \Rightarrow c = \frac{-3}{\sqrt{2}}$

$$\left(\frac{-3}{\sqrt{2}}\right)^2 + d^2 = 5 \Rightarrow d^2 = 5 - \frac{9}{4} \Rightarrow d = \frac{1}{\sqrt{2}}$$

or, for example, $a = b = -1$ $\quad c = 1 \quad d = 2$ $\quad \mathcal{L} = \begin{pmatrix} -1 & -1 \\ 1 & 2 \end{pmatrix}$

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \begin{pmatrix} -1 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} -U-V \\ U+2V \end{pmatrix}$$, where $U, V \sim \mathcal{N}(0,1)$ independent

$$E[X^2Y^2] = E[(-U-V)^2(U+2V)^2] = E[(U^2+2UV+V^2)(U^2+4UV+4V^2)] =$$

$$= E[U^4 + 4V^4 + 13U^2V^2 + 12UV^3 + 6U^3V) = E[U^4] + 4E[V^4] + 13\underbrace{E[U^2]E[V^2]}_{U,V \text{ independent}}$$

$$+ 12\cancel{E[U]E[V^3]}^{0} + 6\cdot \cancel{E[U^3]\cdot E[V]}^{0} \overset{*}{=} 3 + 4\cdot 3 + 13\cdot 1\cdot 1 = 28$$

$$E[U^{2n+1}] = \int_{-\infty}^{+\infty} x^{2n+1} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

Assume that $U \sim \mathcal{N}(0,1)$. Then $E[U^k] = \begin{cases} 0 & \text{when } k \text{ is odd} \\ 1\cdot 3\cdot 5\cdot 7 \cdot \ldots \cdot (k-1) = \dfrac{k!}{\left(\frac{k}{2}\right)! \cdot 2^{k/2}} & \text{when } k \text{ is even} \end{cases}$

So $E[U^4] = 1\cdot 3 = 3 = E[V^4]$

$Cov(X^2, Y^2) = 28 - 10 = 18$

$Var(XY) = 28 - 9 = 19$

Problem. Let $X_1, X_2, \dots, X_{20}$ be independent random variables such that $Var(X_i) = \sigma^2$ for $i = 1, \dots, 20$, $E[X_i] = \mu_1$ $i = 1, \dots, 10$, $E[X_i] = \mu_2$ for $i = 11, \dots, 20$. Let $\overline{X}_1 = \frac{1}{10}\sum_{i=1}^{10} X_i$, $\overline{X}_2 = \frac{1}{10}\sum_{i=11}^{20} X_i$, $\overline{X} = \frac{1}{20}\sum_{i=1}^{20} X_i$

Find numbers $\alpha, \beta \in \mathbb{R}$ such that $\hat{\sigma}^2 = \alpha \sum_{i=1}^{20}(X_i - \overline{X})^2 + \beta(\overline{X}_1 - \overline{X}_2)^2$ is an unbiased estimator of σ^2.

We need to find α, β such that

$$\sigma^2 = \sum \hat{\sigma}^2 = \alpha \cdot E\left[\sum_{i=1}^{20}(X_i - \overline{X})^2\right] + \beta \cdot E[(\overline{X}_1 - \overline{X}_2)^2]$$

(for every μ_1, μ_2, σ^2)

$$E[(\overline{X}_1 - \overline{X}_2)^2] = Var(\overline{X}_1 - \overline{X}_2) + \underbrace{E[\overline{X}_1 - \overline{X}_2]^2}_{\underbrace{\underbrace{E[\overline{X}_1]}_{\mu_1} - \underbrace{E[\overline{X}_2]}_{\mu_2}}_{(\mu_1 - \mu_2)^2}} \overset{*}{=}$$

We are using that $\overline{X}_1$ and $\overline{X}_2$ are independent

$*$

$$= \underbrace{Var \overline{X}_1}_{} + \underbrace{Var \overline{X}_2}_{\text{Same as } Var(\overline{X}_1)} + (\mu_1 - \mu_2)^2 = \frac{\sigma^2}{5} + (\mu_1 - \mu_2)^2$$

$$Var\left(\frac{1}{10}\sum_{i=1}^{10} X_i\right) = \frac{1}{100}\sum_{i=1}^{10} Var(X_i) = \frac{\sigma^2}{10}$$

$$E\left[\sum_{i=1}^{20}(X_i - \overline{X})^2\right] = \sum_{i=1}^{20} E[(X_i - \overline{X})^2] = 10 E[(X_1 - \overline{X})^2] + 10 E[(X_{11} - \overline{X})^2] =$$

$$= 10\Big(Var(X_1 - \overline{X}) + Var(X_{11} - \overline{X}) + \underbrace{E[X_1 - \overline{X}]^2}_{} + \underbrace{E[X_{11} - \overline{X}]^2}_{\frac{\mu_2 - \mu_1}{2}}\Big) =$$

$$E[X_1] - E[\overline{X}], \quad E[X_1] = \mu_1, \quad E[\overline{X}] = E\left[\frac{\overline{X}_1 + \overline{X}_2}{2}\right] = \frac{\mu_1 + \mu_2}{2}, \quad \underbrace{\mu_1 - \frac{\mu_1 + \mu_2}{2}}_{\frac{\mu_1}{2} - \frac{\mu_2}{2}}$$

$$= 10\left(Var(X_1 - \overline{X}) + Var(X_{11} - \overline{X}) + 2 \cdot \left(\frac{\mu_1 - \mu_2}{2}\right)^2\right) = 10\left(2 \cdot \frac{19}{20}\sigma^2 + 2 \cdot \left(\frac{\mu_1 - \mu_2}{2}\right)^2\right)$$

$$* \; Var(X_1 - \overline{X}) = Var X_1 - 2\underbrace{Cov(X_1, \overline{X})} + \underbrace{Var \overline{X}} = \sigma^2 - \frac{\sigma^2}{10} + \frac{\sigma^2}{20} = \frac{19\sigma^2}{20}$$

$$\shortparallel$$

$$Var(X_{11} - \overline{X})$$

$$Var\left(\frac{1}{20}\sum_{i=1}^{20} X_i\right) = \frac{1}{400}\sum_{i=1}^{20} Var(X_i) = \frac{\sigma^2}{20}$$

$$Cov\left(X_1, \frac{1}{20}\sum_{i=1}^{20} X_i\right) = \frac{1}{20}\sum_{i=1}^{20} Cov(X_1, X_i) = \begin{cases} 0, & i \neq 1 \\ \sigma^2, & i = 1 \end{cases} = \frac{\sigma^2}{20}$$

We are looking for $\alpha, \beta \in \mathbb{R}$ such that

$$\sigma^2 = \alpha\left(19\sigma^2 + 5(\mu_1 - \mu_2)^2\right) + \beta\left(\frac{\sigma^2}{5} + (\mu_1 - \mu_2)^2\right) =$$

for all σ^2, μ_1, μ_2

$$= \sigma^2 \underbrace{\left(19\alpha + \frac{\beta}{5}\right)}_{1} + (\mu_1 - \mu_2)^2 \underbrace{(5\alpha + \beta)}_{0} \Longrightarrow \alpha = \frac{1}{18} \qquad \beta = -\frac{5}{18}$$

Problem. Let $X_1, X_2 \sim Unif(0,1)$ independent. Compute $\mu = E|X_1 - X_2|$, $\sigma^2 = Var|X_1 - X_2|$

A) $\mu = 1/3 \quad \sigma^2 = 1/36$

B) $\mu = 1/2 \quad \sigma^2 = 1/12$

C) $\mu = 1/2 \quad \sigma^2 = 1/24$

D) $\mu = 1/3 \quad \sigma^2 = 1/24$

E) $\mu = 1/3 \quad \sigma^2 = 1/6$

$X \sim Unif(a,b)$

$E[X] = \frac{a+b}{2}$

$Var(X) = \frac{(b-a)^2}{12}$

$$\sigma^2 = Var|X_1 - X_2| = E(|X_1-X_2|^2) - (\underbrace{E|X_1-X_2|}_{\mu})^2 =$$

$$= E((X_1-X_2)^2) - \mu^2 = Var(X_1 - X_2) + (\underbrace{E(X_1-X_2)}_{0})^2 - \mu^2 =$$

Indep.

$$= \underbrace{Var X_1}_{1/12} + \underbrace{Var X_2}_{1/12} - \mu^2 = \frac{1}{6} - \mu^2$$

So $\sigma^2 + \mu^2 = \frac{1}{6}$

Problem. Let $X_1, \dots, X_{n+m}$ be a simple sample from $N(\mu, \sigma^2)$. We know $X_1, X_2, \dots, X_n$ and $\overline{X} = \frac{1}{n+m} \sum_{i=1}^{n+m} X_i$. Find $\alpha \in \mathbb{R}$ such that $\hat{\sigma}^2 = \alpha \sum_{i=1}^{n} (X_i - \overline{X})^2$ is an unbiased estimator of σ^2.

We are looking for $\alpha \in \mathbb{R}$ such that $\sigma^2 = E\hat{\sigma}^2 = \alpha E\left[\sum_{i=1}^{n} (X_i - \overline{X})^2\right] =$

$$= \alpha \sum_{i=1}^{n} E[(X_i - \overline{X})^2] = \alpha \sum_{i=1}^{n} \left(Var(X_i - \overline{X}) + \underbrace{E[X_i - \overline{X}]}_{\substack{E[X_i] - E[\overline{X}] \\ \mu - \mu \\ 0}}{}^2\right) = \alpha \cdot n \left(\underbrace{Var X_1}_{\sigma^2} - 2\underbrace{Cov(X_1, \overline{X})}_{} + \underbrace{Var \overline{X}}_{\frac{\sigma^2}{n+m}}\right) \overset{(*)}{=}$$

$$Cov\left(X_1, \frac{1}{n+m} \sum_{i=1} X_i\right) = \frac{1}{n+m} \sum_{i=1}^{n+m} Cov(X_1, X_i) = \frac{\sigma^2}{n+m}$$

$$(*) = \alpha \cdot n \left(\sigma^2 - \frac{\sigma^2}{n+m}\right) = \alpha \cdot \frac{n(n+m-1)}{n+m} \sigma^2 \Longrightarrow \alpha = \frac{n+m}{n(n+m-1)}$$

Problem. Let $X_1, X_2, \dots, X_{100}$ - independent
$E[X_i] = \mu$, $Var X_i = \sigma^2$ for $i = 1, \dots, 100$
We do not know $X_1, X_2, \dots, X_{100}$. We know

$Y_1 = X_1 + X_2 + \dots + X_{10}$

$Y_2 = X_{11} + \dots + X_{20}$

$\vdots$

$Y_{10} = X_{91} + \dots + X_{100}$

Find $\alpha \in \mathbb{R}$ such that $\hat{\sigma}^2 = \alpha \sum_{i=1}^{10} (Y_i - \overline{Y})^2$ is an unbiased estimator of σ^2.

$E[Y_i] = 10\mu \qquad Var(Y_i) = 10\sigma^2 \quad$ for $i = 1, 2, \dots, 10$

$$E\left(\frac{1}{10-1}\sum_{i=1}^{10}(Y_i - \bar{Y})^2\right) = 10\sigma^2$$

$$E\Big[\underbrace{\frac{1}{90}}_{=\alpha}\sum_{i=1}^{10}(Y_i - \bar{Y})^2\Big) = \sigma^2$$

$Y \sim Unif(0,1)$

Conditional distribution of $X/Y=y \sim Unif(0,y)$. Compute $P(X<\frac{1}{2})$

A) 0.5 B) 0.622 C) 0.75 D) 0.847 E) 0.911

If $Y < 1/2 \Rightarrow X < 1/2$ because $X/Y=y \sim Unif(0,y)$ $(0 \le X \le Y)$

So $P(X<1/2) = \underbrace{P(Y<1/2)}_{=0.5} + \underbrace{P(Y>1/2, X<1/2)}_{>0} > 0.5$

So A is not correct.

$$P(X<1/2) = E[P(X<1/2/Y)] = \int_0^1 \underbrace{P(X<1/2/Y=y)}_{X|Y=y \sim Unif(0,y)} \cdot \overset{\text{density of } Y \text{ on } (0,1)}{1}\, dy$$

$$P(X<1/2/Y=y) = \begin{cases} 1 & y < 1/2 \\ \frac{1/2}{y} = \frac{1}{2y} & \frac{1}{2} \le y < 1 \end{cases}$$

FISHER'S THEOREM

Let $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ be independent. Then:

- $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
- $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2_{n-1}$
- $\overline{X}$ and $\sum_{i=1}^{n}(X_i - \overline{X})^2$ are independent

Definition. If $U_1, U_2, \dots, U_n \sim \mathcal{N}(0,1)$ are independent, then $U_1^2 + U_2^2 + \dots + U_n^2 \sim \chi^2_n$ (the chi-square distribution with n degrees of freedom)

Definition. If $X \sim \mathcal{N}(0,1)$, $Y \sim \chi^2_n$ are independent, then

$\frac{X}{\sqrt{Y/n}} \sim t_n$ (t-student's distribution with n degrees of freedom.

Definition. If $X \sim \chi^2_m$, $Y \sim \chi^2_n$ are independent, then

$\frac{\chi_m / m}{\chi_n / n} \sim F_{m,n}$ (the Fischer Snedecor's distribution with m degrees of freedom of the numerator on n degrees of freedom of the denominator).

Definition. Chi-square distribution is a special case of the gamma distribution. Namely, $\chi^2_n = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$

Remark. $\alpha, \beta > 0$ $\Gamma(\alpha, \beta)$ is the distribution with the density $f(x) = \begin{cases} \frac{\beta^\alpha \cdot x^{\alpha-1}}{\Gamma(\alpha)} \cdot e^{-\beta x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$

$\Gamma(\alpha) = \{\text{gamma function}\} = \int_0^\infty x^{\alpha-1} \cdot e^{-x}\, dx$

$\Gamma(1,\beta) = \mathcal{E}xp(\beta)$

In particular $\chi^2_2 = \Gamma(1,\frac{1}{2}) = \mathcal{E}xp(\frac{1}{2})$

Proof. If $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu,\sigma^2)$ and are independent, then $X = (X_1,\dots,X_n)$ has a normal distribution

$$\mathcal{N}\left(\begin{pmatrix}\mu\\ \vdots \\ \mu\end{pmatrix}, \begin{pmatrix}\sigma^2 & & 0\\ & \ddots & \\ 0 & & \sigma^2\end{pmatrix}\right).$$ It follows that

$$\overline{X} = \left(\frac{1}{n},\dots,\frac{1}{n}\right)\begin{pmatrix}X_1\\ \vdots \\ X_n\end{pmatrix} \sim \mathcal{N}\left(\left(\frac{1}{n},\dots,\frac{1}{n}\right)\begin{pmatrix}\mu\\ \vdots\\ \mu\end{pmatrix}, \left(\frac{1}{n},\dots,\frac{1}{n}\right)\begin{pmatrix}\sigma^2 & & 0\\ & \ddots & \\ 0 & & \sigma^2\end{pmatrix}\begin{pmatrix}\frac{1}{n}\\ \vdots\\ \frac{1}{n}\end{pmatrix}\right) = \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

- Random vector

$$(\overline{X}, \underbrace{X_1 - \overline{X}, \dots, X_n - \overline{X}}_{Y \text{ (linear transformation of } X)})^T \text{ has a normal distribution}$$

We will check that $\overline{X}$ and Y are independent. It is enough to show that $\forall_i \; \mathrm{Cov}(\overline{X}, X_i - \overline{X}) = 0$:

$$\mathrm{Cov}(\overline{X}, X_i - \overline{X}) = \mathrm{Cov}(\overline{X}, X_i) - \mathrm{Cov}(\overline{X},\overline{X}) = \mathrm{Cov}\left(\frac{1}{n}\sum_{j=1}^n X_j, X_i\right) - \mathrm{Var}\,\overline{X} =$$

$$= \frac{1}{n}\cdot \sum_{j=1}^n \underbrace{\mathrm{Cov}(X_j, X_i)}_{=\begin{cases}0 & \text{if } i\neq j\\ \sigma^2 & \text{if } j=i\end{cases}} - \frac{\sigma^2}{n} = \frac{1}{n}\cdot\sigma^2 - \frac{\sigma^2}{n} = 0 \text{ (as we wanted)}$$

Finally, we have that $\overline{X}$ and $(X_1 - \overline{X}, \dots, X_n - \overline{X})$ are independent

$\downarrow \qquad \downarrow f(z_1,\dots,z_n) = \sum_{i=1}^n z_i^2$

$\overline{X}$ and $\sum_{i=1}^n (X_i - \overline{X})^2$ are independent

Estimation of parameters

Assume that we observe a random sample X with the (unknown) distribution P_θ, where $\theta \in \Theta$ is unknown. (e.g. if $P_\theta = \mathcal{N}(\mu,\sigma^2)$ then $\theta = (\mu,\sigma^2)$ $\mu\in\mathbb{R}, \sigma^2 > 0$). We want to know θ (and P_θ)

Estimation (point estimation)

Point estimation of the parameter θ on the parametric function $g(\theta)$ is any statistic ($\equiv$ a function of the sample $\mathbb{X}$), which approximates θ (or $g(\theta)$).

(Abt. parametric function

If e.g. $P_\theta = \mathcal{N}(\mu, \sigma^2)$ and $\theta = (\mu, \sigma^2)$
then we can be interested in $g(\theta) = g(\mu, \sigma^2) = \mu$
Similarly for σ^2 or $\sigma = \sqrt{\sigma^2}$)

Usually we denote: $\hat{\theta}$, $\widehat{g(\theta)}$, $\hat{\theta}(\mathbb{X})$, $\widehat{g(\theta)}(X)$. We say that the estimator is unbiased if $\underset{\theta\in\Theta}{\forall} E_\theta(\hat{\theta}(\mathbb{X})) = \theta$ $\left(\underset{\theta\in\Theta}{\overset{\text{or}}{\forall}} E_\theta(\widehat{g(\theta)}(X)) = g(\theta)\right)$

The bias of estimator $\hat{\theta} := E(\hat{\theta}(\mathbb{X})) - \theta$

" " $\widehat{g(\theta)} := E_\theta(\widehat{g(\theta)}(\mathbb{X})) - g(\theta)$

MSE (Mean Square Error) of the estimator $E_\theta((\underset{\widehat{g(\theta)}}{\hat{\theta}} - \underset{g(\theta)}{\theta})^2)$

Interval estimation

Confidence intervals. We assume that $\theta \in \mathbb{R}$ (or $g(\theta) \in \mathbb{R}$)

Confidence interval of the parameter θ (or $g(\theta)$) on the confidence level $1-\alpha \in (0,1)$ is an interval of the form

(usually $1-\alpha \approx 1$, e.g. $1-\alpha = 0.95$ or 0.9 or 0.99)

$(\mathcal{L}(\mathbb{X}), R(\mathbb{X}))$ another notation $(\underline{\theta}(\mathbb{X}), \bar{\theta}(\mathbb{X}))$

where $\mathcal{L}(\mathbb{X}) = \underline{\theta}(\mathbb{X}) = \underline{\theta}$ and $R(\mathbb{X}) = \bar{\theta}(\mathbb{X}) = \bar{\theta}$ are two statistics ($\equiv$ functions of the sample $\mathbb{X}$), which satisfies $\underset{\theta\in\Theta}{\forall} P_\theta(\theta \in (\underline{\theta}(\mathbb{X}), \bar{\theta}(\mathbb{X}))) \geq 1-\alpha$

We would like to have the interval $(\underline{\theta}, \bar{\theta})$ as short as possible (for the given $1-\alpha$)

One-sided confidence intervals:

If $\underline{\theta} := -\infty$ then we obtain $(-\infty, \bar{\theta})$ (right-sided confidence interval)

If $\bar{\theta} := +\infty$ then we obtain $(\underline{\theta}, +\infty)$ (left-sided confidence interval)

Let $(X_1, \dots, X_n)$ be a simple sample from $\mathcal{N}(\mu, \sigma^2)$ i.e. $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ are independent. We want to estimate the parameters μ and $\sigma^2 > 0$.

Parameter μ

The point estimator: $\overline{X}$ since $E[\overline{X}] = \mu$, we know that $\overline{X}$ is an unbiased estimator of μ.

$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$

for any distribution

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Interval estimator: Two subcases: · σ^2 is known · σ^2 is unknown

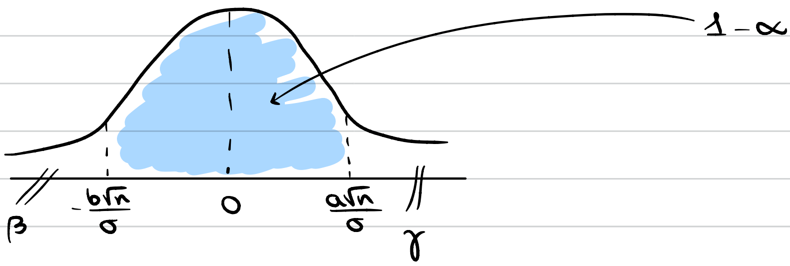
1) If σ^2 is known. $\quad 1-\alpha \in (0,1)$ is fixed

The idea: Consider the interval of the form $(\overline{X}+a, \overline{X}+b)$ for some $a, b > 0$. Find approximate a and b.

We want $P(\mu \in (\overline{X}-a, \overline{X}+b)) \geq 1-\alpha$ and the length is as small as possible

$$P(\mu \in (\overline{X}-a, \overline{X}+b)) = P(\overline{X}-a < \mu < \overline{X}+b)$$

$$= P(a - \overline{X} > -\mu > -b - \overline{X}) = P(a > \overline{X} - \mu > -b) = P\Big(\frac{a\sqrt{n}}{\sigma} > \underbrace{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}_{\mathcal{N}(0,1)} > \frac{-b\sqrt{n}}{\sigma}\Big)$$



$$(1-\alpha) + \beta + \gamma = 1 \iff \beta + \gamma = \alpha$$

$$-\frac{b\sqrt{n}}{\sigma} = u_\beta \quad \text{and} \quad \frac{a\sqrt{n}}{\sigma} = u_{1-\gamma}$$

u_p is the p-th quantile of $\mathcal{N}(0,1)$

$\phi(u_p) = p$ $\quad$ Cumulative distribution function of $\mathcal{N}(0,1)$

We obtain $a = \frac{\sigma}{\sqrt{n}} u_{1-\gamma} \qquad b = \frac{\sigma}{\sqrt{n}} u_{1-\beta}$

Finally, we obtain that for every $\beta, \gamma \geq 0$ satisfying $\beta + \gamma = \alpha$ the interval $\left(\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\gamma}, \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\beta}\right)$ is a confidence interval for μ on the confidence level $1-\alpha$.

Special cases:

If $\beta = 0$ and $\gamma = \alpha$, then $u_{1-\beta} = u_1 = +\infty$ and we obtain

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}, +\infty\right) \text{ - left-sided interval}$$

If $\gamma = 0$ and $\beta = \alpha$, then we obtain

$$\left(-\infty, \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}\right) \text{ - right-sided interval}$$

The shortest interval is obtained for $\beta = \gamma = \frac{\alpha}{2}$

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right)$$

symetric v.r.t. $\overline{X}$

Another case: σ^2 is unknown. We can't use formula $\left(\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\gamma}, \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\beta}\right)$ because we don't know σ.

We will replace σ by $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}$

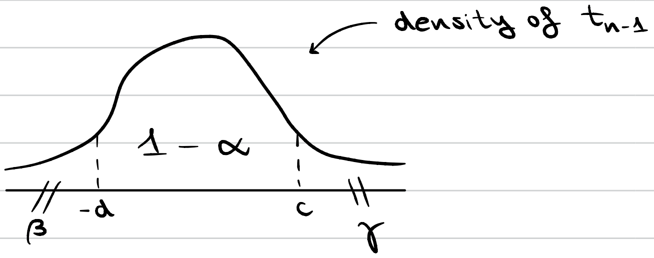
We consider $\left(\overline{X} - \frac{S}{\sqrt{n}} c, \overline{X} + \frac{S}{\sqrt{n}} d\right)$

We want $P\left(\mu \in \left(\overline{X} - \frac{S}{\sqrt{n}} c, \overline{X} + \frac{S}{\sqrt{n}} d\right)\right) \overset{(\geq)}{=} 1-\alpha$

$$= P\left(-d < \frac{\overline{X} - \mu}{S/\sqrt{n}} < c\right)$$

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{S/\sigma} = \frac{\overbrace{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}}^{\sim N(0,1)}}{\sqrt{\underbrace{\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2}_{\sim \chi^2_{n-1}} / n-1}} \sim t_{n-1}$$

Indep. because of Fisher's theorem



$$-d = t_{n-1}(\beta) \Rightarrow d = t_{n-1}(1-\beta)$$

$c = t_{n-1}(1-\gamma)$ quantile of order $1-\gamma$ from t_{n-1}

Finally we obtain: For every $\beta, \gamma \geq 0$, s.t. $\beta + \gamma = \alpha$
the interval

$$\left(\overline{X} - \frac{S}{\sqrt{n}} t_{n-1}(1-\gamma), \overline{X} + \frac{S}{\sqrt{n}} t_{n-1}(1-\beta)\right)$$

is a confidence interval for μ with the confidence level $1-\alpha$

$$S := \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2} \qquad n > 1$$

Special cases:

Left-sided confidence interval $\left(\overline{X} - \frac{S}{\sqrt{n}} t_{n-1}(1-\alpha), +\infty\right)$

Right-sided " " $\left(-\infty, \overline{X} + \frac{S}{\sqrt{n}} t_{n-1}(1-\alpha)\right)$

The shortest, symetric w.r.t. $\overline{X}$ (with respect to)

$$\left(\overline{X} - \frac{S}{\sqrt{n}} t_{n-1}\left(1-\frac{\alpha}{2}\right), \overline{X} + \frac{S}{\sqrt{n}} t_{n-1}\left(1-\frac{\alpha}{2}\right)\right)$$