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:

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

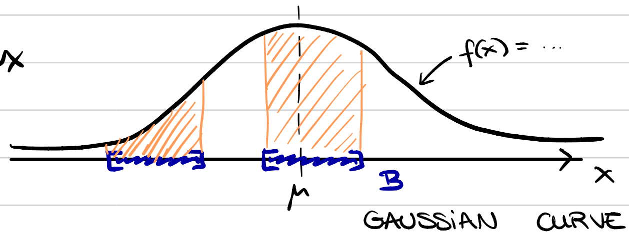
Its symbol is $\mathcal{N}(\mu, \sigma^2)$, although sometimes $\mathcal{N}(\mu, \sigma)$.

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 $\sigma Y + \mu \sim \mathcal{N}(\mu, \sigma^2)$

$\overbrace{\text{Cumulative Distribution Function}}^{\text{CDF}}$ 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:

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

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

* $\operatorname{Var} X = E((X - EX)^2)$

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

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

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

Let $X_1, X_2, \ldots, X_n$ 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 $\operatorname{Var} X_1 (= \operatorname{Var} X_2 = \cdots) = \sigma^2 > 0$.

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.

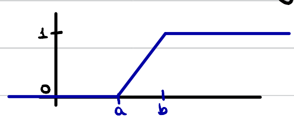
We say that (μ_n) converges (weakly) to μ iff and only if

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

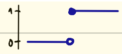
*

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

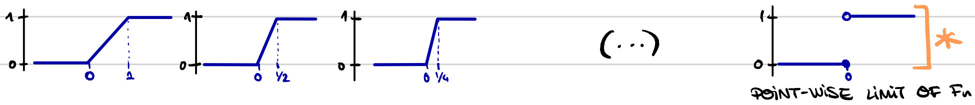
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a,b) \\ 0 & \text{if } x \notin (a,b) \end{cases} \qquad \underset{\text{CDF}}{F(x)} = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } b \leq x \end{cases}$$

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

EX: Let $\mu_n = \text{Unif}(0, 1/n)$. 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; $X \sim \delta_0 \iff X = 0$ almost surely, as $P(X=0)=1$). CDF of δ_0 is 

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



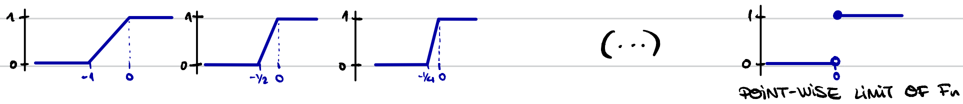
We can see the necessity of the condition *. This function is not CDF of δ_0. It is not CDF of any probability distribution because it is not right-continuous.

It does not converge.

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

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

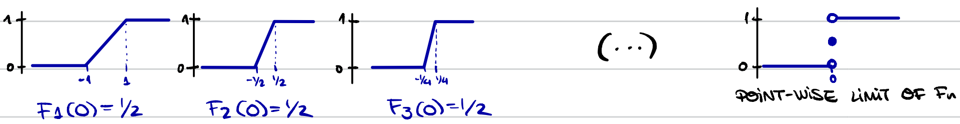
Let us see the CDF of μ_n, $n = 1, 2, \ldots$



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

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

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



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

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 $X, X_1, X_2 \dots$ 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:

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

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

... coming back to CTL. Let $X_1, X_2, \dots$ 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

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

→ We do not need this condition because Φ is cont, as it is the integral of a cont. function. 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:

$$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[n\to\infty]{\text{almost surely}} \mu \iff \frac{\sum_{i=1}^{n} X_i - n\mu}{\underbrace{n}_{\text{in CTL we have } \sqrt{n\sigma^2} \text{ here (smaller number)}}} \xrightarrow[n\to\infty]{\text{almost surely}} 0$$

in CTL we have different conv.

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

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

(CTL)

→ Why those formulas? Let's see:

Let X random variable with expectation EX and variance $\operatorname{Var} X \neq 0$. Then:

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

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

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

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

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

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

$E * = 0$
$\operatorname{Var} * = 1$
$*$ 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 $(X_1, \dots, X_n)$ has the n-dimensional normal distribution if, for every $a_1, \dots, a_n \in \mathbb{R}$, the r.v. $a_1 X_1 + \dots + a_n X_n$ has one-dimensional normal dist.

An equivalent definition could be:
$\mathbb{X} = (X_1, \dots, X_n)$ 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
$\mathcal{N}(\underset{\in\mathbb{R}}{\mu}, \underset{\in\mathbb{R}^+}{\sigma^2})$	$\mathcal{N}(?, ?)$
expectation, variance	"expectation", "variance"

2.1. EXPECTATION AND VARIANCE IN MULTIPLE DIM.

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

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

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

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

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

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

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

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,

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. $\mathrm{Cov}(X, Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$

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. $\mathrm{Cov}(aX,Y) = E((aX - \underbrace{Ea X}_{a EX})(Y-EY)) = E(a(X-EX)(Y-EY)) =$
   $= a\,E((X-EX)(Y-EY)) = a\,\mathrm{Cov}(X,Y)$

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

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

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 $\operatorname{Var} \mathbb{X}$ exists, then:

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

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

PROP: Let X, Y be r.v. Assume that $\operatorname{Var} X$, $\operatorname{Var} Y$ exist. Then, if X and Y are independent $\Rightarrow \operatorname{cov}(X,Y) = 0$ (X, Y are not related)

$$\Leftrightarrow \operatorname{Var}(X \pm Y) = \operatorname{Var} X + \operatorname{Var} Y$$

In general, $\operatorname{cov}(X,Y) = 0 \not\Rightarrow X, Y$ independent.

NOTE: A very important special case is when $\operatorname{cov}(X,Y) = 0 \Rightarrow X, Y$ independent.

Assume that $\binom{X}{Y}$ has 2-dimensional normal distribution, then:

$$\operatorname{cov}(X,Y) = 0 \Leftrightarrow X, Y \text{ independent}$$

Remark: It is not enough to assume that X, Y have normal distribution separately. We need to assume that the joint (2-dim) distribution of $(X,Y)^T$ is normal.

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

$$X_1, \dots, X_n \text{ are independent} \Leftrightarrow \forall i, j \in \{1, \dots, n\}^{i \neq j}, \operatorname{cov}(X_i, X_j) = 0$$

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

THM: Assume that $(X_1, \dots, X_n, Y_1, \dots, Y_m)^T$ is a random vector with normal dist. Then,

$$\begin{array}{l} \mathbb{X} = (X_1, \dots, X_n)^T \\ \mathbb{Y} = (Y_1, \dots, Y_m)^T \end{array} \text{ are independent } \iff \begin{array}{l} \forall\, i \in \{1, \dots, n\} \\ \forall\, j \in \{1, \dots, m\} \end{array} \operatorname{Cov}(X_i, Y_j) = 0.$$

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

$$\begin{array}{cc} \begin{matrix} n \\ m \end{matrix} & \begin{pmatrix} \operatorname{Var}\mathbb{X} & 0 \\ 0 & \operatorname{Var}\mathbb{Y} \end{pmatrix} \\ & \begin{matrix} n & \quad m \end{matrix} \end{array}$$

15 NOV 2023

Recall: EXPONENTIAL DISTRIBUTION $X \sim \mathcal{E}(\beta)$ $\quad \beta > 0$

$EX = 1/\beta$

c.d.f. $F_X(t) = \begin{cases} 0 & t < 0 \\ 1 - e^{-\beta t} & t \geq 0 \end{cases}$

density $f(x) = \begin{cases} 0 & t < 0 \\ \beta e^{-\beta t} & t \geq 0 \end{cases}$

THM: Let $Z_1, \dots, Z_n$ independent, $Z_i \sim \mathcal{E}(\beta_i)$ $\quad i = 1, 2, \dots, n$

Then, $\min(Z_1, \dots, Z_n) \sim \mathcal{E}\left(\sum_{i=1}^{n} \beta_i\right)$

Proof: We will find c.d.f. of $\min(Z_1, \dots, Z_n)$.

- If $t < 0$, $F(t) = P(\overbrace{\min(Z_1, \dots, Z_n)}^{\geq 0} \leq \overset{<0}{t}) = 0$
- If $t \geq 0$, $F(t) = P(\underbrace{\min(Z_1, \dots, Z_n) \leq t}_{\Updownarrow \atop \exists i \in \{1, \dots, n\} \text{ s.t. } Z_i \leq t}) =$ we want to get rid of the min

possible idea, but not useful.

$$= 1 - P(\min(Z_1, \dots, Z_n) > t) = 1 - P(Z_1 > t \overset{\text{AND}}{,} Z_2 > t, \dots, Z_n > t) =$$

independence

$$= 1 - \prod_{i=1}^{n} P(Z_i > t) = 1 - \prod_{i=1}^{n} (1 - \overbrace{P(Z_i < t)}^{F_{Z_i}(t) = 1 - e^{-\beta_i t}}) = 1 - \prod_{i=1}^{n} e^{-\beta_i t} =$$

$$= 1 - e^{-\sum_{i=1}^{n} \beta_i t}$$

22 NOV 2023

THM: FISHER THEOREM

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

1. $\bar{X} := 1/n \sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \sigma^2/n)$
2. $1/\sigma^2 \sum_{i=1}^{n} (X_i - \bar{X}) \sim \chi^2_{n-1}$
3. $\bar{X}$ and $\sum_{i=1}^{n} (X_i - \bar{X})^2$ are independent

Before proving it, let's see some definitions to help us understand the theorem.

DEF: If $U_1, \dots, U_n \sim N(0,1)$ are independent, then $U_1^2 + \dots + U_n^2 \sim \chi_n^2$ (the chi-square distribution with n degrees of freedom)

DEF: If $X \sim \mathcal{N}(0,1)$, $Y \sim \chi^2_n$ are independent, then $\frac{X}{\sqrt{Y/n}} \sim t_n$ (the t-student distribution with n degrees of freedom)

DEF: If $X \sim \chi^2_m$, $Y \sim \chi^2_n$ are independent, then $\frac{X_m/m}{Y_n/n} \sim F_{m,n}$ (the Fischer Snedecor's distribution with m degrees of freedom of the numerator on n degrees of freedom of the denominator).

DEF: Chi-square distribution is a special case of the gamma distribution. Namely, $\chi^2_n = \Gamma(n/2, 1/2)$

Remark: $\alpha, \beta > 0$, $\Gamma(\alpha, \beta)$ is the distribution with the density

$$f(x) = \begin{cases} \frac{\beta^\alpha \cdot x^{\alpha-1}}{\Gamma(\alpha)} \cdot e^{-\beta x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

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

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

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

Proof:

1. If $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ 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

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

3. The random vector $\overbrace{(\bar{X}, X_1 - \bar{X}, \dots, X_n - \bar{X})^T}^{Y \text{ (linear transformation of } X)}$ has a normal distribution.

We will check that $\bar{X}$ and Y are independent. It is enough to show that $\forall i \in \{1, \dots, n\}$, $\operatorname{Cov}(\bar{X}, X_i - \bar{X}) = 0$:

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

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

Finally, we have that $\bar{X}$ and $(X_1 - \bar{X}, \dots, X_n - \bar{X})$ are independent, so, taking $f(z_1, \dots, z_n) = \sum_{i=1}^n z_i^2$, we have that:

$\bar{X}$ and $f(X_1 - \bar{X}, \dots, X_n - \bar{X}) = \sum_{i=1}^n (X_i - \bar{X})^2$ 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)

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

Remark: About 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 $\sqrt{\sigma^2} = \sigma$

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

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

The bias of estimator $\widehat{g(\theta)} \overset{\text{is}}{:=} E_\theta(\widehat{g(\theta)}(\mathbb{X})) - g(\theta)$

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

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 $(L(\mathbb{X}), R(\mathbb{X}))$ or $(\underline{\theta}(\mathbb{X}), \overline{\theta}(\mathbb{X}))$, another notation
where $L(\mathbb{X}) = \underline{\theta}(\mathbb{X}) = \underline{\theta}$ and $R(\mathbb{X}) = \overline{\theta}(\mathbb{X}) = \overline{\theta}$ are two statistics ($\equiv$ functions of the sample $\mathbb{X}$), which satisfy

$$\forall \theta \in \Theta, \quad P_\theta(\theta \in (\underline{\theta}(\mathbb{X}), \overline{\theta}(\mathbb{X}))) \geq 1-\alpha$$

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

→ We would like to have the interval $(\underline{\theta}, \overline{\theta})$ 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 conf. int.
If $\bar{\theta} := +\infty$, then we obtain $(\underline{\theta}, +\infty)$ left-sided conf. int.

Let $(X_1, \dots, X_n)$ be a simple sample of $\mathcal{N}(\mu, \sigma^2)$ 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

-) Point estimator: $\bar{X}$. Since $E[\bar{X}] = \mu$, we know that $\bar{X}$ is an unbiased estimator of μ. $\operatorname{Var}(\bar{X}) = \sigma^2/n$ } for any dist.
$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.

-) Interval estimator: Fix $1-\alpha \in (0,1)$. We have two subcases:

a) σ^2 is known

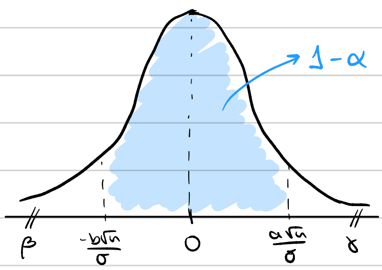
The idea is to consider the interval of the form $(\bar{X}-a, \bar{X}+b)$ for some $a, b > 0$. We want to find approximate a and b.

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

$$P(\bar{X}-a < \mu < \bar{X}+b) = P(a - \bar{X} > -\mu > -b - \bar{X}) =$$

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

So, we have:



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

We note $\frac{-b\sqrt{n}}{\sigma} = u_\beta$ and $\frac{a\sqrt{n}}{\sigma} = u_{1-\gamma}$.

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

$\phi(u_p) = p$ is the cummulative function of $\mathcal{N}(0,1)$.

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

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

Special cases:

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

left-sided c.i. $\rightarrow \left(\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}, +\infty\right)$

a.2) If $\gamma=0$ and $\beta=\alpha$, then $u_{1-\gamma}=u_1=-\infty$ and we obtain:

right-sided c.i. $\rightarrow \left(-\infty, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}\right)$

a.3) The shortest interval is obtained for $\beta=\gamma=\alpha/2$:

symetric with respect to $\bar{X}$ $\rightarrow \left(\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1+\alpha/2}\right)$

b) σ^2 is unknown.

We can't use the formula $\left(\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\gamma}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\beta}\right)$ because we don't know σ. We will replace σ by

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

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

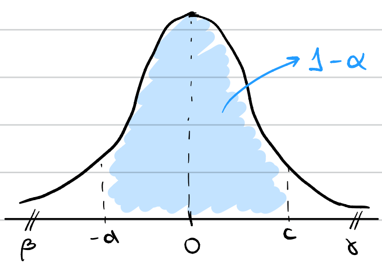
We want $P\left(\mu \in \left(\bar{X} - \frac{S}{\sqrt{n}} c, \bar{X} + \frac{S}{\sqrt{n}} d\right)\right) \overset{(\geq)}{=} 1-\alpha$.

$$\| $$

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

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

independent because of Fisher's thm



$\leftarrow$ density of t_{n-1}

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

·) $c = t_{n-1}(1-\gamma)$ 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

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

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

Special cases:

a.1) left sided c.i. $\rightarrow \left(\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(1-\alpha), +\infty\right)$ $(\gamma = \alpha, \beta = 0)$

a.2) Right sided c.i. $\rightarrow \left(-\infty, \bar{X} + \frac{S}{\sqrt{n}} t_{n-1}(1-\alpha)\right)$ $(\gamma = 0, \beta = \alpha)$

a.3) The shortest interval, symetrical with respect to $\bar{X}$:

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

29 NOV 2023

Parameter σ^2

·) Point estimator: We have two possible cases.

a) μ is known

Let us consider the statistic $\sum_{i=1}^{n}(X_i - \mu)^2$. We have:

$$E\left(\sum_{i=1}^{n}(X_i-\mu)^2\right) = \sum_{i=1}^{n} E((X_i-\mu)^2) = \sum_{i=1}^{n}\left(Var(X_i-\mu) + (E(X_i-\mu))^2\right) =$$

$$= \sum_{i=1}^{n}\left(\underbrace{Var\, X_i}_{\sigma^2} + \underbrace{(\overset{\mu}{E X_i} - \mu)^2}_{0}\right) = n\sigma^2$$

It follows that: $\hat{\sigma}^2 = 1/n \sum_{i=1}^{n}(X_i-\mu)^2$ is an unbiased estimator of σ^2. (We didn't use the assumption about the normality at $X_1, \dots, X_n$ or about the independence).

b) μ is unknown

Then, from Fisher's theorem, we obtain that:

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is an unbiased estimator of σ^2 and

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 \sim \chi^2_{n-1} \quad \text{(Fisher's theorem)}$$

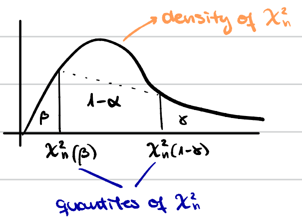
•) Interval estimator: We have two possible cases.

a) μ is known

the unbiased estimator from before

these summands are different

Note that $\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} \underbrace{\left(\frac{X_i - \mu}{\sigma}\right)}_{\mathcal{N}(0,1)}^2 \sim \chi^2_n$



We want to build the confidence interval on the confidence level $1-\alpha \in (0,1)$

We choose two numbers $\gamma, \beta \geq 0$ s.t. $\gamma + \beta = \alpha$.

We have $P\left(\chi^2_n(\beta) < \frac{n\hat{\sigma}^2}{\sigma^2} < \chi^2_n(1-\gamma)\right) \overset{(\geq)}{=} 1-\alpha$

$$= P\left(\frac{1}{\chi^2_n(1-\gamma)} < \frac{\sigma^2}{n\hat{\sigma}^2} < \frac{1}{\chi^2_n(\beta)}\right) = P\left(\sigma^2 \in \left(\frac{n\hat{\sigma}^2}{\chi^2_n(1-\gamma)}, \frac{n\hat{\sigma}^2}{\chi^2_n(\beta)}\right)\right) = 1-\alpha$$

We obtain that, $\forall \beta, \gamma \geq 0$ s.t. $\beta + \gamma = \alpha$, the interval:

$$\left(\frac{n\hat{\sigma}^2}{\chi^2_n(1-\gamma)}, \frac{n\hat{\sigma}^2}{\chi^2_n(\beta)}\right) = \left(\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\chi^2_n(1-\gamma)}, \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\chi^2_n(\beta)}\right)$$

is a confidence interval for σ^2 on the level $1-\alpha$.

Special cases:

a.1) When $\beta=0$ and $\gamma=\alpha$, $\left(\frac{n\hat{\sigma}^2}{\chi^2_n(1-\alpha)}, +\infty\right) = \left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_n(1-\alpha)}, +\infty\right)$

left-sided c.i.

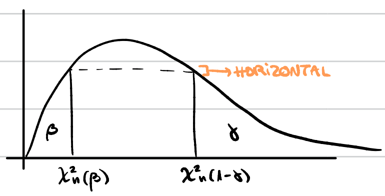
NOT $(-\infty, -)$!!! because $\sigma^2 \geq 0$.

a.2) When $\beta=\alpha$ and $\gamma=0$, $\left[0, \frac{n\hat{\sigma}^2}{\chi^2_n(\alpha)}\right) = \left[0, \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_n(\alpha)}\right)$

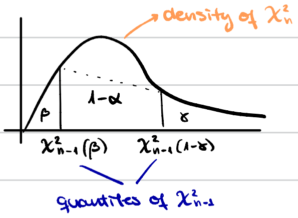
right-sided c.i.

a.3) When $\beta=\gamma=\alpha/2$, $\left(\frac{n\hat{\sigma}^2}{\chi^2_n(1-\alpha/2)}, \frac{n\hat{\sigma}^2}{\chi^2_n(\alpha/2)}\right) = \left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_n(1-\alpha/2)}, \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_n(\alpha/2)}\right)$

a.4) If we want to obtain the shortest two-sided confidence interval, then we need to choose $\beta, \gamma \geq 0$ such that $\beta+\gamma=\alpha$ and:



b) μ is unknown



We have that $P\left(\chi^2_{n-1}(\beta) < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1}(1-\gamma)\right) \overset{(\geq)}{=} 1-\alpha$

$$\| $$

$$P\left(\frac{1}{\chi^2_{n-1}(1-\gamma)} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi^2_{n-1}(\beta)}\right) = 1-\alpha$$

$$\| $$

$$P\left(\sigma^2 \in \left(\frac{(n-1)S^2}{\chi^2_{n-1}(1-\gamma)}, \frac{(n-1)S^2}{\chi^2_{n-1}(\beta)}\right)\right) = 1-\alpha$$

We obtain that $\forall \beta, \gamma \geq 0$ s.t. $\beta+\gamma=\alpha$, the interval

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1}(1-\gamma)}, \frac{(n-1)S^2}{\chi^2_{n-1}(\beta)}\right) = \left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\chi^2_{n-1}(1-\gamma)}, \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\chi^2_{n-1}(\beta)}\right)$$

is a confidence interval for σ^2 on the level $1-\alpha$.

Special cases:

a.1) When $\beta=0$ and $\gamma=\alpha$, $\left(\frac{(n-1)S^2}{\chi^2_{n-1}(1-\gamma)}, +\infty\right) = \left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\chi^2_{n-1}(1-\gamma)}, +\infty\right)$

left-sided c.i.

NOT $-\infty$!

a.2) When $\beta=\alpha$ and $\gamma=0$, $\left[0, \frac{(n-1)S^2}{\chi^2_{n-1}(\beta)}\right) = \left[0, \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\chi^2_{n-1}(\beta)}\right)$

right-sided c.i.

a.3) When $\beta=\gamma=\alpha/2$, $\left(\frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)}, \frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)}\right) = \left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\chi^2_{n-1}(1-\alpha/2)}, \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\chi^2_{n-1}(\alpha/2)}\right)$

a.4) If we want to obtain the shortest two-sided confidence interval, then we need to choose $\beta, \gamma \geq 0$ s.t. $\beta+\gamma=\alpha$ and: