Gaussian vectors and central limit theorem

Samy Tindel

Purdue University

Probability Theory 2 - MA 539

Outline

Outline

Standard Gaussian random variable

Definition: Let

- X be a real valued random variable.

X is called standard Gaussian if its probability law admits the density:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbb{R}.$$

Notation: We denote by $\mathcal{N}_1(0,1)$ or $\mathcal{N}(0,1)$ this law.

Gaussian random variable and expectations

Reminder:

1. For all bounded measurable functions g, we have

$$\mathbf{E}[g(X)] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(x) \exp\left(-\frac{x^2}{2}\right) dx.$$

2. In particular,

$$\int_{\mathbb{R}} \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}.$$

Gaussian moments

Proposition 1.

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

1. For all $z \in \mathbb{C}$, we have
$$\mathbf{E}[\exp(zX)] = \exp(z^2/2).$$
As a particular case, we get
$$\mathbf{E}[\exp(\imath tX)] = e^{-t^2/2}, \quad \forall t \in \mathbb{R}.$$

2. For all $n \in \mathbb{N}$, we have
$$\mathbf{E}[X^n] = \begin{cases} 0 \quad \text{if } n \text{ is odd,} \\ \dfrac{(2m)!}{m!2^m}, \quad \text{if } n \text{ is even, } n = 2m. \end{cases}$$

Proof

(i) Definition of the transform:
$\int_{\mathbb{R}} \exp(zx - \frac{1}{2}x^2)dx$ absolutely convergent for all $z \in \mathbb{C}$
$\hookrightarrow$ the quantity $\varphi(z) = \mathbf{E}[e^{zX}]$ is well defined and,

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(zx - \frac{1}{2}x^2\right) dx.$$

(ii) Real case: Let $z \in \mathbb{R}$.
Decomposition $zx - \frac{1}{2}x^2 = -\frac{1}{2}(x-z)^2 + \frac{z^2}{2}$
and change of variable $y = x - z \Rightarrow \varphi(z) = e^{z^2/2}$

Proof (2)

(iii) Complex case:
φ and $z \mapsto e^{z^2/2}$ are two entire functions
Since those two functions coincide on $\mathbb{R}$, they coincide on $\mathbb{C}$.

(iv) Characteristic function:
In particular, if $z = \imath t$ with $t \in \mathbb{R}$, we have

$$\mathbf{E}[\exp(\imath t X)] = e^{-t^2/2}$$

Proof (3)

(v) Moments: Let $n \geq 1$.

Convergence of $\mathbf{E}[|X^n|]$: easy argument

In addition, we almost surely have

$$e^{\imath tX} = \lim_{n\to\infty} S_n, \quad \text{with} \quad S_n = \sum_{k=0}^{n} \frac{(\imath t)^k}{k!} X^k.$$

However, $|S_n| \leq Y$ with

$$Y = \sum_{k=0}^{\infty} \frac{|t|^k |X|^k}{k!} = e^{|tX|} \leq e^{tX} + e^{-tX}.$$

Since $\mathbf{E}[\exp(aX)] < \infty$, we obtain that Y is integrable

Applying dominated convergence, we end up with

$$\mathbf{E}[\exp(\imath tX)] = \mathbf{E}\left[\sum_{n\geq 0} \frac{(\imath tX)^n}{n!}\right] = \sum_{n\geq 0} \frac{\imath^n t^n}{n!} \mathbf{E}[X^n]. \tag{1}$$

Identifying lhs and rhs, we get our formula for moments

Gaussian random variable

Corollary: Owing to the previous proposition, if $X \sim \mathcal{N}(0,1)$
$\hookrightarrow$ $\mathbf{E}[X] = 0$ and $\mathbf{Var}(X) = 1$

Definition:
A random variable is said to be Gaussian if there exists $X \sim \mathcal{N}(0,1)$ and two constants a and b such that $Y = aX + b$.

Parameter identification: we have

$$\mathbf{E}[Y] = b, \quad \text{and} \quad \mathbf{Var}(Y) = a^2 \, \mathbf{Var}(X) = a^2.$$

Notation: We denote by $\mathcal{N}(m, \sigma^2)$ the law of a Gaussian random variable with mean m and variance σ^2.

Properties of Gaussian random variables

Density: we have

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad \text{is the density of } \mathcal{N}(m,\sigma^2)$$

Characteristic function: let $Y \sim \mathcal{N}(m,\sigma^2)$. Then

$$\mathbf{E}[\exp(\imath t Y)] = \exp\left(\imath t m - \frac{t^2}{2}\sigma^2\right), \quad t \in \mathbb{R}.$$

The formula above also characterizes $\mathcal{N}(m,\sigma^2)$

Gaussian law: illustration

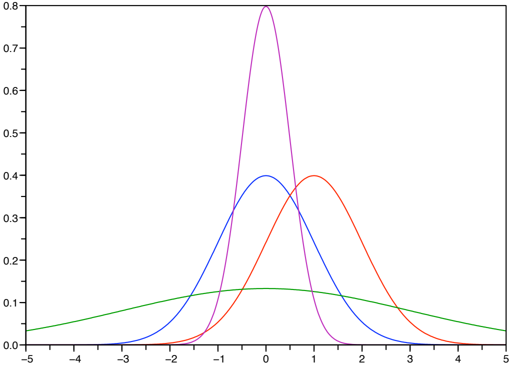

Figure: Distributions $\mathcal{N}(0,1)$, $\mathcal{N}(1,1)$, $\mathcal{N}(0,9)$, $\mathcal{N}(0,1/4)$.

Sum of independent Gaussian random variables

Proposition 2.

Let Y_1 and Y_2 be two independent Gaussian random variables
Assume $Y_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ and $Y_2 \sim \mathcal{N}_1(m_2, \sigma_2^2)$.
Then $Y_1 + Y_2 \sim \mathcal{N}_1(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Proof:
Via characteristic functions

Remarks:

- It is easy to identify the parameters of $Y_1 + Y_2$
- Possible generalization to $\sum_{j=1}^{n} Y_j$

Outline

Matrix notation

Transpose:
If A is a matrix, A^* designates the transpose of A.

Particular case: Let $x \in \mathbb{R}^n$. Then

- x is a column vector in $\mathbb{R}^{n,1}$
- x^* is a row matrix

Inner product:
If x and y are two vectors in $\mathbb{R}^n$, their inner product is denoted by

$$\langle x, y \rangle = x^* y = y^* x = \sum_{i=1}^{n} x_i y_i, \text{ if } x^* = (x_1, \ldots, x_n), \; y^* = (y_1, \ldots, y_n).$$

Vector valued random variable

Definition 3.

1. A random variable X with values in $\mathbb{R}^n$ is given by n real valued random variables $X_1, X_2, \ldots, X_n$.
2. We denote by X the column matrix with coordinates $X_1, X_2, \ldots, X_n$:

$$X^* = (X_1, X_2, \ldots, X_n).$$

Expected value and covariance

Expected value: Let $X \in \mathbb{R}^n$. $\mathbf{E}[X]$ is the vector defined by

$$\mathbf{E}[X]^* = (\mathbf{E}[X_1], \mathbf{E}[X_2] \ldots, \mathbf{E}[X_n]) .$$

Note: here we assume that all the expectations are well-defined.

Covariance: Let $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$.
The covariance matrix $K_{X,Y} \in \mathbb{R}^{n,m}$ is defined by

$$K_{X,Y} = \mathbf{E}\left[(X - \mathbf{E}[X])\,(Y - \mathbf{E}[Y])^*\right]$$

Elements of the covariance matrix: for $1 \le i \le n$ and $1 \le j \le m$

$$K_{X,Y}(i,j) = \mathbf{Cov}(X_i, Y_j) = \mathbf{E}\left[(X_i - \mathbf{E}[X_i])\,(Y_j - \mathbf{E}[Y_j])\right]$$

Simples properties

Linear transforms and Expectation-covariance:
Let $X \in \mathbb{R}^n$, $A \in \mathbb{R}^{m,n}$, $u \in \mathbb{R}^m$. Then

$$\mathbf{E}[u + AX] = u + A\,\mathbf{E}[X], \quad \text{and} \quad K_{u+AX} = K_{AX} = AK_XA^*.$$

Another formula for the covariance:

$$K_{X,Y} = \mathbf{E}\left[XY^*\right] - \mathbf{E}\left[X\right]\,\mathbf{E}\left[Y\right]^*.$$

As a particular case,

$$K_X = \mathbf{E}\left[XX^*\right] - \mathbf{E}\left[X\right]\mathbf{E}\left[X\right]^*$$

Outline

Definition

Definition: Let $X \in \mathbb{R}^n$.
X is a Gaussian random vector iff for all $\lambda \in \mathbb{R}^n$

$$\langle \lambda, X \rangle = \lambda^* X = \sum_{i=1}^{n} \lambda_i X_i \text{ is a real valued Gaussian r.v.}$$

Remarks:
(1) X Gaussian vector
$\Rightarrow$ Each component X_i of X is a real Gaussian r.v

(2) Key example of Gaussian vector:
Independent Gaussian components $X_1, \ldots, X_n$

(3) Easy construction of random vector $X \in \mathbb{R}^2$ such that
(i) X_1, X_2 real Gaussian (ii) X is not a Gaussian vector

Characteristic function

Proposition 4.

Let X Gaussian vector with mean m and covariance K
Then, for all $u \in \mathbb{R}^n$,

$$\mathbf{E}\left[\exp(\imath\langle u, X\rangle)\right] = e^{\imath\langle u, m\rangle - \frac{1}{2} u^* K u},$$

where we use the matrix representation for the vector u

Proof

Identification of $\langle u, X\rangle$:
$\langle u, X\rangle$ Gaussian r.v by assumption, with parameters

$$\mu := \mathbf{E}[\langle u, X\rangle] = \langle u, m\rangle, \quad \text{and} \quad \sigma^2 := \mathbf{Var}(\langle u, X\rangle) = u^* K u \tag{2}$$

Characteristic function of 1-d Gaussian r.v:
Let $Y \sim \mathcal{N}(\mu, \sigma^2)$. Then recall that

$$\mathbf{E}[\exp(\imath t Y)] = \exp\left(\imath t\mu - \frac{t^2}{2}\sigma^2\right), \quad t \in \mathbb{R}. \tag{3}$$

Conclusion: Easily obtained by plugging (2) into (11)

Remark and notation

Remark: According to Proposition 4
$\hookrightarrow$ The law of a Gaussian vector X is characterized by its mean m and its covariance matrix K
$\hookrightarrow$ If X and Y are two Gaussian vectors with the same mean and covariance matrix, their law is the same

Caution: This is only true for Gaussian vectors.
In general, two random variables sharing the same mean and variance are not equal in law

Notation: If X Gaussian vector with mean m and covariance K
We write $X \sim \mathcal{N}(m, K)$

Linear transformations

Proposition 5.

Let

- $X \sim \mathcal{N}(m_X, K_X)$
- $A \in \mathbb{R}^{p,n}$ and $z \in \mathbb{R}^p$

Set

$$Y = AX + z$$

Then

$$Y \sim \mathcal{N}(m_Y, K_Y), \quad \text{with} \quad m_Y = z + Am_X, \quad K_Y = AK_XA^*$$

Proof

Aim: Let $u \in \mathbb{R}^p$.
We wish to prove that $u^* Y$ is a Gaussian r.v.

Expression for $u^* Y$: We have

$$u^* Y = u^* z + u^* A X = u^* z + v^* X,$$

where we have set $v = A^* u$. This is a Gaussian r.v

Conclusion: Y is a Gaussian vector. In addition,

$$m_Y = \mathbf{E}[Y] = z + A\mathbf{E}[X] = z + A m_X, \quad \text{and} \quad K_Y = A K_X A^*.$$

Positivity of the correlation matrix

Proposition 6.

Let X be a random vector with covariance matrix K.
Then K is a symmetric positive matrix.

Proof:

Symmetry: $K(i,j) = \mathbf{Cov}(X_i, X_j) = \mathbf{Cov}(X_j, X_i) = K(j,i)$

Positivity: Let $u \in \mathbb{R}^n$ and $Y = u^* X$. Then

$$\mathbf{Var}(Y) = u^* K u \geq 0$$

Linear algebra lemma

Lemma 7.

Let

- $\Gamma \in \mathbb{R}^{n,n}$, symmetric and positive.

Then there exists a matrix $A \in \mathbb{R}^{n,n}$ such that

$$\Gamma = AA^*$$

Proof

Diagonal form of Γ:

- Γ symmetric $\Rightarrow$ there exists an orthogonal matrix U and $D_1 = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ such that $D_1 = U^* \Gamma U$
- Γ positive $\Rightarrow \lambda_i \geq 0$ for all $i \in \{1, 2, \ldots, n\}$.

Definition of the square root:

- Let $D = \mathrm{Diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$.
- We set $A = UD$.

Conclusion:

- Recall that $U^{-1} = U^*$, therefore $\Gamma = U D_1 U^*$.
- Now $D_1 = D^2 = DD^*$, and thus

$$\Gamma = UDD^* U^* = UD(UD)^* = AA^*.$$

Construction of a Gaussian vector

Theorem 8.

Let

- $m \in \mathbb{R}^n$
- $\Gamma \in \mathbb{R}^{n,n}$ symmetric and positive

Then

There exists a Gaussian vector $X \sim \mathcal{N}(m, \Gamma)$

Proof

Standard Gaussian vector in $\mathbb{R}^n$:
Let $Y_1, Y_2, \ldots, Y_n$, i.i.d with common law $\mathcal{N}_1(0,1)$. We set

$$Y^* = (Y_1, \ldots, Y_n), \quad \text{and therefore} \quad Y \sim \mathcal{N}(0, \mathrm{Id}_n).$$

Definition of X: Let $A \in \mathbb{R}^{n,n}$ such that $AA^* = \Gamma$.
We define X as:

$$X = m + AY.$$

Conclusion:
According to Proposition 5 we have $X \sim \mathcal{N}(m, K_X)$, with

$$K_X = A\,K_Y\,A^* = A\,\mathrm{Id}\,A^* = AA^* = \Gamma.$$

Decorrelation and independence

Theorem 9.

Let X be Gaussian vector, with $X^* = (X_1, \ldots, X_n)$.
Then

The random variables $X_1, \ldots, X_n$ are independent

$$\Longleftrightarrow$$

The covariance matrix K_X is diagonal.

Proof of $\Rightarrow$

Decorrelation of coordinates:
If $X_1, \ldots, X_n$ are independent, then

$$K(i,j) = \mathbf{Cov}(X_i, X_j) = 0, \quad \text{whenever } i \neq j.$$

Therefore K_X is diagonal.

Proof of $\Leftarrow$ (1)

Characteristic function of X: Set $K = K_X$. We have shown that

$$\mathbf{E}[\exp(\imath\langle u, X\rangle)) = e^{\imath\langle u, \mathbf{E}[X]\rangle - \frac{1}{2}u^* K u}, \ \ u \in \mathbb{R}^n. \tag{4}$$

Since K is diagonal, we have :

$$u^* K u = \sum_{l=1}^{n} u_l^2 \, K(l,l) = \sum_{l=1}^{n} u_l^2 \, \mathbf{Var}(X_l). \tag{5}$$

Characteristic function of each coordinate:
Let ϕ_{X_l} be the characteristic function of X_l
We have $\phi_{X_l}(s) = \mathbf{E}[e^{\imath s X_l}]$, for all $s \in \mathbb{R}$.

Taking u such that $u_i = 0$, for all $i \neq l$ in (4) and (5) we get

$$\phi_{X_l}(u_l) = \mathbf{E}\left[\exp(\imath u_l X_l)\right] = e^{\imath u_l \mathbf{E}[X_l] - \frac{1}{2}u_l^2 \mathbf{Var}(X_l)}.$$

Proof of $\Leftarrow$ (2)

Conclusion:

We can recast (4) as follows: for all $u = (u_1, u_2, ..., u_n)$,

$$\prod_{j=1}^{n} \phi_{X_j}(u_j) = E\left[\exp\left(\imath \sum_{l=1}^{n} u_l X_l\right)\right] = \mathbf{E}[\exp(\imath\langle u, X\rangle)],$$

This means that

the random variables $X_1, \ldots, X_n$ are independent.

Lemma about absolutely continuous r.v

Lemma 10.

Let

- $\xi \in \mathbb{R}^n$ a random variable admitting a density.
- H a subspace of $\mathbb{R}^n$, such that $\dim(H) < n$.

Then

$$P(\xi \in H) = 0.$$

Proof

Change of variables:
We can assume $H \subset H'$ with

$$H' = \{(x_1, x_2, ..., x_n);\ x_n = 0\}$$

Conclusion:
Denote by φ the density of ξ. We have:

$$\begin{aligned} P(\xi \in H) &\leq P(\xi \in H') \\ &= \int_{\mathbb{R}^n} \varphi(x_1, x_2, ..., x_n) \mathbf{1}_{\{x_n=0\}} dx_1\, dx_2 \dots dx_n \\ &= 0. \end{aligned}$$

Gaussian density

Theorem 11.

Let $X \sim \mathcal{N}(m, K)$. Then

1. X admits a density iff K is invertible.
2. If K is invertible, the density of X is given by

$$f(x) = \frac{1}{(2\pi)^{n/2}(\det(K))^{1/2}} \exp\left(-\frac{1}{2}(x-m)^* K^{-1}(x-m)\right)$$

Proof

(1) Density and inversion of K: We have seen

$$X \stackrel{(d)}{=} m + AY, \quad \text{where} \quad AA^* = K, \; Y \sim \mathcal{N}(0, \mathrm{Id}_n)$$

(i) Assume A non invertible.

A non invertible $\Rightarrow \mathrm{Im}(A) = H$, with $\dim(H) < n$
$\hookrightarrow \mathbf{P}(AY \in H) = 1$

Contradiction:
X admits a density $\Rightarrow X - m$ admits a density
$\Rightarrow P(X - m \in H) = 0$

However, we have seen that $\mathbf{P}(X - m \in H) = \mathbf{P}(AY \in H) = 1$.

Hence X doesn't admit a density.

Proof (2)

(ii) Assume A invertible.

A invertible
$\Rightarrow$ application $y \to m + Ay$ is a $\mathcal{C}^1$ bijection
$\Rightarrow$ the random variable $m + AY$ admits a density.

(iii) Conclusion.

Since $AA^* = K$, we have

$$\det(A)\,\det(A^*) = (\det(A))^2 = \det(K)$$

and we get the equivalence:

$$A \text{ invertible} \iff K \text{ is invertible}.$$

Proof (3)

(2) Expression of the density: Let $Y \sim \mathcal{N}(0, \mathrm{Id}_n)$. Density of Y:

$$g(y) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\langle y, y\rangle\right).$$

Change of variable: Set

$$X' = AY + m \quad \text{that is} \quad Y = A^{-1}(X' - m)$$

Proof (4)

Jacobian of the transformation: for $x \mapsto A^{-1}(x-m)$ we have

$$\text{Jacobian} = A^{-1}$$

Determinant of the Jacobian:

$$\det(A^{-1}) = [\det(A)]^{-1} = [\det(K)]^{-1/2}$$

Expression for the inner product:

We have $K^{-1} = (AA^*)^{-1} = (A^*)^{-1}A^{-1}$, and

$$\begin{aligned}\langle y, y\rangle &= \langle A^{-1}(x-m), A^{-1}(x-m)\rangle \\ &= (x-m)^*(A^{-1})^*A^{-1}(x-m) = (x-m)^*K^{-1}(x-m).\end{aligned}$$

Thus X' admits the density f.

Since X and X' share the same law, X admits the density f.

Outline

Law of large numbers

Theorem 12.

We consider the following situation:

- $(X_n;\ n \geq 1)$ sequence of i.i.d $\mathbb{R}^k$-valued r.v
- Hypothesis: $\mathbf{E}[|X_1|] < \infty$, and we set $\mathbf{E}[X_1] = m \in \mathbb{R}^k$

We define

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j .$$

Then

$$\lim_{n \to \infty} \bar{X}_n = m, \quad \text{almost surely}$$

Central limit theorem

Theorem 13.

We consider the following situation:

- $\{X_n;\ n \geq 1\}$ sequence of i.i.d $\mathbb{R}^k$-valued r.v
- Hypothesis: $\mathbf{E}[|X_1|^2] < \infty$
- We set $\mathbf{E}[X_1] = m \in \mathbb{R}^k$ and $\mathbf{Cov}(X_1) = \Gamma \in \mathbb{R}^{k,k}$

Then

$$\sqrt{n}\left(\bar{X}_n - m\right) \xrightarrow{(d)} \mathcal{N}_k(0, \Gamma), \quad \text{with} \quad \bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Interpretation: $\bar{X}_n$ converges to m with rate $n^{-1/2}$

Convergence in law, first definition

Remark: For notational sake
$\hookrightarrow$ the remainder of the section will focus on $\mathbb{R}$-valued r.v

Definition 14.

Let

- $\{X_n;\, n \geq 1\}$ sequence of r.v, X_0 another r.v
- F_n distribution function of X_n
- F_0 distribution function of X_0
- We set $\mathcal{C}(F) \equiv \{x \in \mathbb{R};\, F \text{ continuous at point } x\}$

Definition 1: We have

$$\lim_{n\to\infty} X_n \stackrel{(d)}{=} X_0 \text{ if } \lim_{n\to\infty} F_n(x) = F_0(x) \text{ for all } x \in \mathcal{C}(F).$$

Convergence in law, equivalent definition

Proposition 15.

Let

- $\{X_n;\, n \geq 1\}$ sequence of r.v, X_0 another r.v
- We set
  $C_b(\mathbb{R}) \equiv \{\varphi : \mathbb{R} \to \mathbb{R};\, \varphi \text{ continuous and bounded}\}$

Definition 2: We have

$$\lim_{n\to\infty} X_n \overset{(d)}{=} X_0$$

iff

$$\lim_{n\to\infty} \mathbf{E}[\varphi(X_n)] = \mathbf{E}[\varphi(X_0)] \text{ for all } \varphi \in C_b(\mathbb{R}).$$

Central limit theorem in $\mathbb{R}$

Theorem 16.

We consider the following situation:

- $\{X_n;\ n \geq 1\}$ sequence of i.i.d $\mathbb{R}$-valued r.v
- Hypothesis: $\mathbf{E}[|X_1|^2] < \infty$
- We set $\mathbf{E}[X_1] = \mu$ and $\mathbf{Var}(X_1) = \sigma^2$

Then

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{(d)} \mathcal{N}(0, \sigma^2), \quad \text{with} \quad \bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Otherwise stated we have

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma n^{1/2}} \quad \xrightarrow{(d)} \quad \mathcal{N}(0, 1)$$

Application: Bernoulli distribution

Proposition 17.

Let $(X_n;\ n \geq 1)$ sequence of i.i.d $\mathcal{B}(p)$ r.v
Then

$$\sqrt{n}\left(\frac{\bar{X}_n - p}{[p(1-p)]^{1/2}}\right) \xrightarrow{(d)} \mathcal{N}_1(0,1).$$

Remark:
For practical purposes as soon as $np > 15$, the law of

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{np(1-p)}}$$

is approached by $\mathcal{N}_1(0,1)$. Notice that $X_1 + \cdots + X_n \sim \mathrm{Bin}(n,p)$.

Binomial distribution: plot (1)

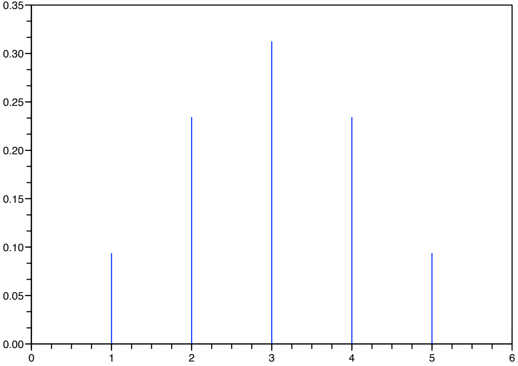

Figure: Distribution $\text{Bin}(6; 0.5)$. x-axis: k, y-axis: $\mathbf{P}(X = k)$

Binomial distribution: plot (2)

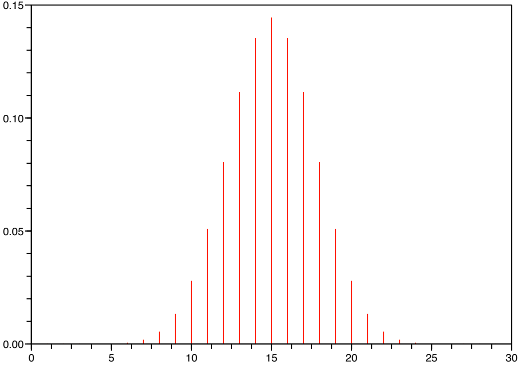

Figure: Distribution $\text{Bin}(30; 0.5)$. x-axis: k, y-axis: $\mathbf{P}(X = k)$

Relation between pdf and chf

Theorem 18.

Let

- F be a distribution function on $\mathbb{R}$
- ϕ the characteristic function of F

Then F is uniquely determined by ϕ

Proof (1)

Setting: We consider

- A r.v X with distribution F and chf ϕ
- A r.v Z with distribution G and chf γ

Relation between chf: We have

$$\int_{\mathbb{R}} e^{-\imath\theta z}\phi(z)\, G(dz) = \int_{\mathbb{R}} F(dx)\, \gamma(x-\theta) \tag{6}$$

Proof (2)

Proof of (6): Invoking Fubini, we get

$$\begin{aligned}\mathbf{E}\left[e^{-\imath\theta Z}\phi(Z)\right] &= \int_{\mathbb{R}} e^{-\imath\theta z}\phi(z)\, G(dz) \\ &= \int_{\mathbb{R}} G(dz)\, e^{-\imath\theta z}\left[\int_{\mathbb{R}} e^{\imath zx}\, F(dx)\right] \\ &= \int_{\mathbb{R}} F(dx)\left[\int_{\mathbb{R}} e^{\imath z(x-\theta)}\, G(dz)\right] \\ &= \mathbf{E}\left[\gamma(X-\theta)\right]\end{aligned}$$

Proof (3)

Particularizing to a Gaussian case: We now consider

- $Z \sim \sigma N$ with $N \sim \mathcal{N}(0,1)$
- In this case, if $n \equiv$ density of $\mathcal{N}(0,1)$, we have

$$G(dz) = \sigma^{-1}\, n(\sigma^{-1} z)\, dz$$

With this setting, relation (6) becomes

$$\int_{\mathbb{R}} e^{-\imath\theta\sigma z} \phi(\sigma z) n(z)\, dz = \int_{\mathbb{R}} e^{-\frac{1}{2}\sigma^2 (z-\theta)^2}\, F(dz) \tag{7}$$

Proof (4)

Integration with respect to θ: Integrating (7) wrt θ we get

$$\int_{-\infty}^{x} d\theta \int_{\mathbb{R}} e^{-\imath\theta\sigma z}\phi(\sigma z)\, n(z)\, dz = A_{\sigma,\theta}(x), \tag{8}$$

where

$$A_{\sigma,\theta}(x) = \int_{-\infty}^{x} d\theta \int_{\mathbb{R}} e^{-\frac{1}{2}\sigma^2(z-\theta)^2}\, F(dz)$$

Proof (5)

Expression for $A_{\sigma,\theta}$: We have

$$A_{\sigma,\theta}(x) \overset{\text{Fubini}}{=} \int_{\mathbb{R}} F(dz) \int_{-\infty}^{x} e^{-\frac{1}{2}\sigma^2(z-\theta)^2}\, d\theta$$

$$\overset{\text{c.v: } s=\theta-z}{=} \left(2\pi\sigma^{-2}\right)^{1/2} \int_{\mathbb{R}} F(dz) \int_{-\infty}^{x-z} n_{0,\sigma^{-2}}(s)\, ds$$

Therefore, considering $N \perp\!\!\!\perp X$ with $N \sim \mathcal{N}(0,1)$ we get

$$A_{\sigma,\theta}(x) = \left(2\pi\sigma^{-2}\right)^{1/2} \mathbf{P}\left(\sigma^{-1}N + X \leq x\right) \tag{9}$$

Proof (6)

Summary: Putting together (8) and (9) we get

$$\int_{-\infty}^{x} d\theta \int_{\mathbb{R}} e^{-\imath\theta\sigma z}\phi(\sigma z)\, n(z)\, dz = \left(2\pi\sigma^{-2}\right)^{1/2} \mathbf{P}\left(\sigma^{-1}N + X \leq x\right)$$

Divide the above relation by $(2\pi\sigma^{-2})^{1/2}$. We obtain

$$\frac{\sigma}{(2\pi)^{1/2}} \int_{-\infty}^{x} d\theta \int_{\mathbb{R}} e^{-\imath\theta\sigma z}\phi(\sigma z)\, n(z)\, dz = \mathbf{P}\left(\sigma^{-1}N + X \leq x\right) \quad (10)$$

Proof (7)

Convergence result: Recall that

$$X_{1,n} \xrightarrow{(d)} X_1 \quad \text{and} \quad X_{2,n} \xrightarrow{(\mathbf{P})} X_2 \quad \Longrightarrow \quad X_{1,n} + X_{2,n} \xrightarrow{(d)} X_1 \qquad (11)$$

Notation: We set

$$\mathcal{C}(F) \equiv \{x \in \mathbb{R};\ F \text{ continuous at point } x\}$$

Proof (8)

Limit as $\sigma \to \infty$:

Thanks to our convergence result, one can take limits in (10)

$$\begin{aligned}
&\lim_{\sigma\to\infty} \frac{\sigma}{(2\pi)^{1/2}} \int_{-\infty}^{x} d\theta \int_{\mathbb{R}} e^{-\imath\theta\sigma z} \phi(\sigma z)\, n(z)\, dz \\
&= \lim_{\sigma\to\infty} \mathbf{P}\left(\sigma^{-1} N + X \le x\right) \\
&= \mathbf{P}\left(X \le x\right) \\
&= F(x),
\end{aligned} \tag{12}$$

for all $x \in \mathcal{C}(F)$

Conclusion:

F is determined by ϕ.

Fourier inversion

Proposition 19.

Let

- F be a distribution function on $\mathbb{R}$, and $X \sim F$
- ϕ the characteristic function of F

Hypothesis:

$$\phi \in L^1(\mathbb{R})$$

Conclusion:

F admits a bounded continuous density f, given by

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\imath yx} \phi(y) \, dy$$

Proof (1)

Density of $\sigma^{-1}N + X$: We set

$$F_\sigma(x) = \mathbf{P}\left(\sigma^{-1}N + X \le x\right)$$

Since both N and X admit a density, F_σ admits a density f_σ

Expression for F_σ: Recall relation (10)

$$\frac{\sigma}{(2\pi)^{1/2}} \int_{-\infty}^{x} d\theta \int_{\mathbb{R}} e^{-\imath\theta\sigma z} \phi(\sigma z)\, n(z)\, dz = F_\sigma(x) \tag{13}$$

Proof (2)

Expression for f_σ: Differentiating the lhs of (13) we get

$$\begin{aligned} f_\sigma(\theta) &= \frac{\sigma}{(2\pi)^{1/2}} \int_{\mathbb{R}} e^{-\imath\theta\sigma z} \phi(\sigma z)\, n(z)\, dz \\ (\text{c.v: } \sigma z = y) &= \frac{\sigma}{(2\pi)^{1/2}} \int_{\mathbb{R}} e^{-\imath\theta y} \phi(y)\, n(\sigma^{-1} y)\, dy \\ (n \text{ is Gaussian}) &= \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} e^{-\imath\theta y} \phi(y)\, e^{-\frac{\sigma^{-2} y^2}{2}}\, dy \end{aligned}$$

Relation (10) on a finite interval: Let $I = [a, b]$. Using f_θ we have

$$\mathbf{P}\left(\sigma^{-1} N + X \in [a, b]\right) = F_\sigma(b) - F_\sigma(a) = \int_a^b f_\sigma(\theta)\, d\theta \qquad (14)$$

Proof (3)

Limit of f_σ: By dominated convergence,

$$\lim_{\sigma\to\infty} f_\sigma(\theta) = \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} e^{-\imath\theta y} \phi(y)\, dy \equiv f(\theta)$$

Domination of f_σ: We have

$$\begin{aligned} f_\sigma(\theta) &= \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} e^{-\imath\theta y} \phi(y)\, e^{-\frac{\sigma^{-2}y^2}{2}}\, dy \\ &\leq \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} |\phi(y)|\, dy \\ &= \frac{1}{(2\pi)^{1/2}} \|\phi\|_{L^1(\mathbb{R})} \end{aligned}$$

Proof (4)

Limits in (14): We use

- On lhs of (14): Convergence result (11)
- On rhs of (14): Dominated convergence (on finite interval I)

We get

$$\mathbf{P}(X \in [a, b]) = F(b) - F(a) = \int_a^b f(\theta)\, d\theta$$

Conclusion:

X admits f (obtained by Fourier inversion) as a density

Convergence in law and chf

Theorem 20.

Let

- $\{X_n;\, n \geq 1\}$ sequence of r.v, X_0 another r.v
- ϕ_n chf of X_n, ϕ_0 chf of X_0

Then

(i) We have

$$\lim_{n\to\infty} X_n \overset{(d)}{=} X_0 \quad \Longrightarrow \quad \lim_{n\to\infty} \phi_n(t) = \phi_0(t) \text{ for all } t \in \mathbb{R}$$

(ii) Assume that

- $\phi_0(0) = 1$ and ϕ_0 continuous at point 0

Then we have

$$\lim_{n\to\infty} \phi_n(t) = \phi_0(t) \text{ for all } t \in \mathbb{R} \quad \Longrightarrow \quad \lim_{n\to\infty} X_n \overset{(d)}{=} X_0$$

Central limit theorem in $\mathbb{R}$ (repeated)

Theorem 21.

We consider the following situation:

- $\{X_n;\ n \geq 1\}$ sequence of i.i.d $\mathbb{R}$-valued r.v
- Hypothesis: $\mathbf{E}[|X_1|^2] < \infty$
- We set $\mathbf{E}[X_1] = \mu$ and $\mathbf{Var}(X_1) = \sigma^2$

Then

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{(d)} \mathcal{N}(0, \sigma^2), \quad \text{with} \quad \bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Otherwise stated we have

$$\frac{S_n - n\mu}{\sigma n^{1/2}} \quad \xrightarrow{(d)} \quad \mathcal{N}(0,1), \quad \text{with} \quad S_n = \sum_{i=1}^{n} X_i \qquad (15)$$

Proof of CLT (1)

Reduction to $\mu = 0$, $\sigma = 1$: Set

$$\hat{X}_i = \frac{X_i - \mu}{\sigma}, \quad \text{and} \quad \hat{S}_n = \sum_{i=1}^{n} \hat{X}_i$$

Then

$$\hat{S}_n = \frac{S_n - n\mu}{\sigma}, \qquad \hat{X}_i \sim \mathcal{N}(0,1)$$

and

$$\frac{S_n - n\mu}{\sigma n^{1/2}} = \frac{\hat{S}_n}{n^{1/2}}$$

Thus it is enough to prove (15) when $\mu = 0$ and $\sigma = 1$

Proof of CLT (2)

Aim: For X_i such that $\mathbf{E}[X_i] = 0$ and $\mathbf{Var}(X_i) = 1$, set

$$\phi_n(t) = \mathbf{E}\left[e^{\imath t \frac{S_n}{n^{1/2}}}\right]$$

We wish to prove that

$$\lim_{n\to\infty} \phi_n(t) = e^{-\frac{1}{2}t^2}$$

According to Theorem 20 -(ii), this yields the desired result

Taylor expansion of the chf

Lemma 22.

Let

- Y be a r.v.
- ψ chf of Y

Hypothesis: for $\ell \geq 1$,

$$\mathbf{E}\left[|Y|^{\ell}\right] < \infty.$$

Conclusion:

$$\left|\psi(s) - \sum_{k=0}^{\ell} \frac{(\imath s)^k}{k!}\,\mathbf{E}[X^k]\right| \leq \mathbf{E}\left[\frac{|s\,X|^{\ell+1}}{(\ell+1)!} \wedge \frac{2|sX|^{\ell}}{\ell!}\right].$$

Proof: Similar to (1).

Proof of CLT (3)

Computation for ϕ_n: We have

$$\phi_n(t) = \left(\mathbf{E}\left[e^{\imath \frac{tX_1}{n^{1/2}}}\right]\right)^n = \left[\phi\left(\frac{t}{n^{1/2}}\right)\right]^n, \quad (16)$$

where

$$\phi \equiv \text{characteristic function of } X_1$$

Proof of CLT (4)

Expansion of ϕ: According to Lemma 22, we have

$$\phi\left(\frac{t}{n^{1/2}}\right) = 1 + \imath t \frac{\mathbf{E}[X_1]}{n^{1/2}} + \imath^2 t^2 \frac{\mathbf{E}[X_1^2]}{2n} + R_n$$
$$= 1 - \frac{t^2}{2n} + R_n, \qquad (17)$$

and R_n satisfies

$$|R_n| \le \mathbf{E}\left[\frac{|t\,X_1|^3}{6n^{3/2}} \wedge \frac{|t\,X_1|^2}{n}\right]$$

Behavior of R_n: By dominated convergence we have

$$\lim_{n\to\infty} n\,|R_n| = 0 \qquad (18)$$

Products of complex numbers

Lemma 23.

Let

- $\{a_i; 1 \leq i \leq n\}$, such that $a_i \in \mathbb{C}$ and $|a_i| \leq 1$
- $\{b_i; 1 \leq i \leq n\}$, such that $b_i \in \mathbb{C}$ and $|b_i| \leq 1$

Then we have

$$\left| \prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i \right| \leq \sum_{i=1}^{n} |a_i - b_i|$$

Proof of Lemma 23

Case $n = 2$: Stems directly from the identity

$$a_1 a_2 - b_1 b_2 = a_1 (a_2 - b_2) + (a_1 - b_1) b_2$$

General case: By induction

Proof of CLT (5)

Summary: Thanks to (16) and (17) we have

$$\phi_n(t) = \left[\phi\left(\frac{t}{n^{1/2}}\right)\right]^n, \quad \text{and} \quad \phi\left(\frac{t}{n^{1/2}}\right) = 1 - \frac{t^2}{2n} + R_n$$

Application of Lemma 23: We get

$$\left|\left[\phi\left(\frac{t}{n^{1/2}}\right)\right]^n - \left(1 - \frac{t^2}{2n}\right)^n\right| \tag{19}$$

$$\leq n\left|\phi\left(\frac{t}{n^{1/2}}\right) - \left(1 - \frac{t^2}{2n}\right)\right| \tag{20}$$

$$= n\,|R_n| \tag{21}$$

Proof of CLT (6)

Limit for ϕ_n: Invoking (18) and (19) we get

$$\lim_{n\to\infty}\left|\phi_n(t)-\left(1-\frac{t^2}{2n}\right)^n\right|=0$$

In addition

$$\lim_{n\to\infty}\left|\left(1-\frac{t^2}{2n}\right)^n-e^{-\frac{t^2}{2}}\right|=0$$

Therefore

$$\lim_{n\to\infty}\left|\phi_n(t)-e^{-\frac{t^2}{2}}\right|=0$$

Conclusion: CLT holds, since

$$\lim_{n\to\infty}\phi_n(t)=e^{-\frac{1}{2}t^2}$$

Outline

Gamma and chi-square laws

Definition 1:

For all $\lambda > 0$ and $a > 0$, we denote by $\gamma(\lambda, a)$ the distribution on $\mathbb{R}$ defined by the density

$$\frac{x^{\lambda-1}}{a^{\lambda}\Gamma(\lambda)} \exp\left(-\frac{x}{a}\right) \mathbf{1}_{\{x>0\}}, \quad \text{where} \quad \Gamma(\lambda) = \int_0^{\infty} x^{\lambda-1} e^{-x} dx$$

This distribution is called gamma law with parameters λ, a.

Definition 2:

Let $X_1, \ldots, X_n$ i.i.d $\mathcal{N}(0,1)$. We set $Z = \sum_{i=1}^{n} X_i^2$.
The law of Z is called
chi-square distribution with n degrees of freedom.
We denote this distribution by $\chi^2(n)$.

Gamma and chi-square laws (2)

Proposition 24.

The distribution $\chi^2(n)$ coincides with $\gamma(n/2, 2)$.

As a particular case, if

- $X_1, \ldots, X_n$ i.i.d $\mathcal{N}(0, 1)$
- We set $Z = \sum_{i=1}^{n} X_i^2$,

then we have

$$Z \sim \gamma(n/2, 2).$$

Empirical mean and variance

Let $X_1, \ldots, X_n$ n real r.v

Definition: we set

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k, \quad \text{and} \quad S_n^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X}_n)^2.$$

$\bar{X}_n$ is called empirical mean.
S_n^2 is called empirical variance.

Property:
Let $X_1, \ldots, X_n$ n i.i.d real r.v
Assume $\mathbf{E}[X_1] = m$ and $\mathbf{Var}(X_1) = \sigma^2$. Then

$$\mathbf{E}\left[\bar{X}_n\right] = m, \quad \text{and} \quad \mathbf{E}\left[S_n^2\right] = \sigma^2$$

Law of $(\bar{X}_n, S_n^2)$ in a Gaussian situation

Theorem 25.

Let X_1, X_2,...,X_n i.i.d with common law $\mathcal{N}_1(m, \sigma^2)$.
Then

1. $\bar{X}_n$ and S_n^2 are independent.
2. $\bar{X}_n \sim \mathcal{N}_1(m, \frac{\sigma^2}{n})$ and $\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$.

Proof (1)

(1) Reduction to $m = 0$ and $\sigma = 1$: we set

$$X_i' = \frac{X_i - m}{\sigma} \Longleftrightarrow X_i = \sigma X_i' + m \quad 1 \leq i \leq n.$$

The r.v $X_1', \ldots, X_n'$ are i.i.d distributed as $\mathcal{N}_1(0, 1)$
$\hookrightarrow$ empirical mean $\bar{X}_n'$, empirical variance $S_n'^2$

Proof (2)

(1) Reduction to $m = 0$ and $\sigma = 1$ (ctd):
It is easily seen (using $X_i - \bar{X}_n = \sigma(X_i' - \bar{X}_n')$) that

$$\bar{X}_n = \sigma \bar{X}_n' + m, \quad \text{and} \quad S_n^2 = \sigma^2 S_n'^2.$$

Thus we are reduced to the case $m = 0$ and $\sigma = 1$

(2) Reduced case:
Consider $X_1, \ldots, X_n$ i.i.d $\mathcal{N}(0, 1)$
Let $u_1^* = n^{-1/2}(1, 1, \ldots, 1)$
We can construct $u_2, \ldots, u_n$ such that $(u_1, \ldots, u_n)$ onb of $\mathbb{R}^n$
Let $A \in \mathbb{R}^{n,n}$ whose columns are $u_1, \ldots, u_n$
We set $Y = A^* X$

Proof (3)

(i) Expression for the empirical mean:

A orthogonal matrix: $AA^* = A^*A = \mathrm{Id}$

$\hookrightarrow$ $Y \sim \mathcal{N}(0, K_Y)$ with

$$K_Y = A^* K_X (A^*)^* = A^* \,\mathrm{Id}\, A = A^* A = \mathrm{Id},$$

because the covariance matrix K_X of X is Id.

Due to the fact that the first row of A^* is

$$u_1^* = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, , \ldots, \frac{1}{\sqrt{n}} \right),$$

we have:

$$Y_1 = \frac{1}{\sqrt{n}} (X_1 + X_2 + \ldots + X_n) = \sqrt{n} \bar{X}_n,$$

or otherwise stated, $\bar{X}_n = \frac{Y_1}{\sqrt{n}}$

Proof (4)

(ii) Expression for the empirical variance:

Let us express S_n^2 in terms of Y:

$$\begin{aligned}(n-1)S_n^2 &= \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \sum_{k=1}^n (X_k^2 - 2X_k\bar{X}_n + \bar{X}_n^2) \\ &= \left(\sum_{k=1}^n X_k^2\right) - 2\bar{X}_n \left(\sum_{k=1}^n X_k\right) + n\bar{X}_n^2.\end{aligned}$$

As a consequence,

$$(n-1)S_n^2 = \left(\sum_{k=1}^n X_k^2\right) - 2\bar{X}_n(n\bar{X}_n) + n\bar{X}_n^2 = \left(\sum_{k=1}^n X_k^2\right) - n\bar{X}_n^2.$$

Proof (5)

(ii) Expression for the empirical variance (ctd): We have

$$Y = A^* X, \ A^* \text{ orthogonal} \Rightarrow \textstyle\sum_{k=1}^n Y_k^2 = \sum_{k=1}^n X_k^2$$

Hence

$$(n-1)S_n^2 = \sum_{k=1}^n X_k^2 - n\bar{X}_n^2 = \sum_{k=1}^n Y_k^2 - Y_1^2 = \sum_{k=2}^n Y_k^2.$$

Proof (6)

Summary: We have seen that

$$\bar{X}_n = \frac{Y_1}{\sqrt{n}}, \quad \text{and} \quad (n-1)S_n^2 = \sum_{k=2}^{n} Y_k^2$$

Conclusion:

1. $Y \sim \mathcal{N}(0, \mathrm{Id}_n) \Rightarrow Y_1, \ldots, Y_n$ i.i.d $\mathcal{N}(0,1)$
$\hookrightarrow$ independence of $\bar{X}_n$ and S_n^2.
2. Furthermore, $\bar{X}_n = \frac{Y_1}{\sqrt{n}} \Rightarrow \bar{X}_n \sim \mathcal{N}_1(0, 1/n)$
3. We also have $(n-1)S_n^2 = \sum_{k=2}^{n} Y_k^2$
$\Rightarrow$ the law of $(n-1)S_n^2$ is $\chi^2(n-1)$.