ELEMENTS OF PROBABILTY

… A. N. Kolmogorov (1933) + …

$$(\Omega, B_{\Omega}, P)$$

Carlos Mana
Benasque-2014
Astrofísica y Física de Partículas
CIEMAT

1) Sample Space (Ω)

Event: *Object of questions that we make about the result of the experiment such that the possible answers are: "it occurs" or "it does not occur"*

elementary = those that can not be decomposed in others of lesser entity

Sample Space: $\Omega =$ *{Set of all the possible elementary results of a random experiment}*

The elementary events have to be:

exclusive: if one happens, no other occurs

exhaustive: any possible elemental result has to be included in Ω

$\{e_k\}$ **is a partition of** Ω $\longrightarrow$ $\Omega = \bigcup_{\forall k} e_k \qquad e_k \bigcap e_j = \emptyset \quad ; \forall k, j \quad k \neq j$

---

sure: *get any result contained in* Ω

impossible: *to get a result that is not contained in* Ω

random event: *any event that is neither impossible nor sure*

$$\dim(\Omega) = \begin{cases} \text{Finite} \\ \text{drawing a die} & \Omega = \{e_1, e_2, e_3, e_4, e_5, e_6\} \\ \text{denumerable} \\ \text{throw a coin and stop when we get head} & \Omega = \{c, xc, xxc, xxxc, \ldots\} \\ \text{non-denumerable} \\ \text{decay time of a particle} & \Omega = \{t \in \Re^+\} \end{cases}$$

2) Measurable Space (Ω, B_Ω)

Ω *Sample Space*

B_Ω *σ-algebra (Boole)*

Class closed under complementation and denumerable union

Why algebra B_Ω ? *We are interested in a class of events that:*

1) *Contains all possible results of the experiment on which we are interested*

2) *Is closed under union and complementation*

$$\forall A_1, A_2 \in B_\Omega \quad \rightarrow \quad \overline{A}_1 \in B_\Omega \quad ; \quad A_1 \bigcup A_2 \in B_\Omega$$

$$\longrightarrow \quad \Omega \in B_\Omega; \quad \emptyset \in B_\Omega; \quad A_1 \bigcap A_2 \in B_\Omega; \quad \overline{A}_1 \bigcup \overline{A}_2 \in B_\Omega; \quad \overline{A}_1 \bigcup \overline{A}_2 \in B_\Omega; \ldots$$

So now:

1) Ω *has all the elementary events (to which we shall assign probabilities)*

2) B_Ω *has all the events we are interested in (deduce their probabilities)*

EXAMPLE:

$$\Omega = \{e_1, e_2, e_3, e_4, e_5, e_6\}$$

Several possible algebras:

Minimal: $B_{\min} = \{E, \emptyset\}$

Interest in evenness: ***A={get an even number}***

Ā={get an odd number}

$$B = \{E, \emptyset, A, \overline{A}\}$$

Maximal: $B_{\max} = \{E, \emptyset, all\ \ possible\ subsets\ of\ \ \Omega\}$

$\dim(\Omega) = n$ $\qquad \binom{n}{k}$ ***Subsets with k elements*** $\qquad \sum_{k=0}^{n} \binom{n}{k} = 2^n$ ***elements***

$$\emptyset : \emptyset = \binom{n}{0} = 1 \qquad \Omega : \binom{n}{n} = 1$$

$\dim(\Omega)$ ***finite*** $\longrightarrow$ B_Ω ***has the structure of Boole algebra***

$\dim(\Omega)$ ***denumerable***

Generalize the Boole algebra such that $\bigcup$ ***and*** $\bigcap$ ***can be performed infinite number of times resulting on events of the same class (closed)***

$$\{A_i\}_{i=1}^{\infty} \in B_\Omega \quad \rightarrow \quad \bigcup_{i=1}^{\infty} A_i \in B_\Omega$$

$$\forall A \in B_\Omega \quad \rightarrow \quad A^c \in B_\Omega$$

$$\left(\bigcap_{i=1}^{\infty} A_i \in B_\Omega, \ldots \right)$$

$\longrightarrow$ B_Ω ***has structure of σ-algebra***

Remember that: ***1) All σ-algebras are Boole algebras***

2) Not all Boole algebras are σ-algebras

$\dim(\Omega)$ *non-denumerable*

In general, in non-denumerable topological spaces there are subsets that can not be considered as events

Which are the "elementary" events?

We are mainly interested in R^n

R : ***linear set of points***

Among its possible subsets are the intervals

$[\quad) \; (\quad] \; (\quad) \; [\quad]$

points (degenerated interval) $\{a\}=[a,a]$ $[a,a)=(a,a]=(a,a)=\emptyset$	$R=(-\infty,\infty)$	***any interval, denumerable or not, is a subset of*** R

…but…

$\bigcap$ **finite or denumerable of intervals is an interval**

$\cancel{\bigcup}$ **finite or denumerable of intervals is not, in general, an interval**

Generate a σ-algebra, for instance, from half open intervals on the right

1) Initial Set (Ω): Contains all half-open intervals on the right $[a,b)$

2) Form the set A by adding their denumerable unions and complements

$$(a,b)=\bigcap_{n=1}^{\infty}\left[a+\tfrac{1}{n},b\right)\quad (a,b]=\bigcap_{n=1}^{\infty}\left(a,b+\tfrac{1}{n}\right)\quad [a,b]=\bigcap_{n=1}^{\infty}\left[a,b+\tfrac{1}{n}\right)\quad \{a\}=[a,a]\quad \ldots$$

It has all intervals and points (degenerated intervals)

3.1) There is at least one σ-algebra containing A

(add sets to close under denumerable union and complementation)

3.2) The intersection of any number of σ-algebras is a σ-algebra

→ There exists a smallest σ-algebra containing A

Borel σ-algebra (B_R) **: Minimum σ-algebra of subsets of** R **generated by** $[a,b)$

(May do with $(a,b],(a,b),[a,b]$ as well) Its elements are **Borel sets (borelians)**

Ω: intervals → assigns P to intervals $N,Z,Q\subset B_R$

3) Measure Space (Ω, B_Ω, μ)

3.1) Measure

Set function $\mu : A \in B_\Omega \to R$ **(univoque)**

i) σ-additive **For any succession of disjoint of sets of B**

$$\{A_1, A_2, \ldots\}; \quad A_i \bigcap_{\substack{i,j=1 \\ i \neq j}}^{\infty} A_j = \emptyset \qquad \mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

ii) Non-negative $\mu : A \in B \to \mu(A) \in \Re^+ + \{0\}$

► **Measure Space** (Ω, B_Ω, μ)

3.2) Probability Measure

$$\mu : A \in B_\Omega \to [0,1] \in R$$

(notation $\mu \to P$ **)**

$\mu(\Omega) = 1$ **(certainty)**

$\mu(A^c) = 1 - \mu(A), \ldots$

► **Probability Space** (Ω, B_Ω, P)

Remember that:
Any bounded measure can be converted in a probability measure

All Borel Sets of R^n are Lebesgue Measurable

There are non-denumerable subsets of R with zero Lebesgue measure (Cantor Ternary Set)

If axiom of choice (Solovay 1970) not all subsets of R are measurable $\to$ are not Borel
(Ex.: Vitali's set)

Random variables

Results of the experiment not necessarily numeric,…

Associate to each elemental event of the sample space Ω one, and only one, real number through a function

(misfortunately called "random variable")

$$X(w): w \in \Omega \to X(w) \in R$$

To keep the structure of the σ-algebra it is necessary that X(w) be Lebesgue (…Borel) measurable

$$X^{-1}(A) \in B; \quad \forall A \in B_R$$

Induced Space

$$(\Omega, B) \xrightarrow{X} (\Omega_I, B_I)$$

$$(\Omega, B, P) \to (\Omega_I, B_I, P_I)$$

$$X^{-1}(A) \in B; \quad \forall A \in B_I$$

(f(w):Ω→Δ Borel measurable ←→ measurable wrt the σ-algebra associated to Ω)

Usually: $(E_I, B_I) = (R, B_R)$

$P(X = k)$ ***or*** $P(X \in (a,b))$

Is neither variable nor random

What is random is the outcome of the experiment before it is done; our knowledge on the result before observation,…

EXAMPLE:

$$\Omega=\{e_1,e_2,e_3,e_4,e_5,e_6\}$$

Interest in evenness: **A={get an even number}**

Ā={get an odd number}

$$B=\{E,\emptyset,A,\bar{A}\}$$

$$X(w): w\in\Omega \to X(w)\in R \qquad X^{-1}(-\infty,a]$$

---

1)

$$X(e_1)=X(e_2)=X(e_3)=1$$

$$X(e_4)=X(e_5)=X(e_6)=-1$$

$$a<-1\to\emptyset\in B$$

$$-1\le a<1\to\{e_1,e_2,e_3\}\notin B$$

$$1\le a\to\Omega\in B$$

---

2)

$$X(e_1)=X(e_3)=X(e_5)=1$$

$$X(e_2)=X(e_4)=X(e_6)=-1$$

$$a<-1\to\emptyset\in B$$

$$-1\le a<1\to\{e_2,e_4,e_4\}=A\in B$$

$$1\le a\to\Omega\in B$$

Types of Random Quantities

$$(\Omega, B, Q) \xrightarrow{X(w):w\in\Omega\to R} (R, B_R, P)$$

According to the Range of X(w) …

finite or denumerable set

Discrete

► **simple random quantity:**

$\{A_k; k=1,\ldots,n\}$ finite partition of Ω

simple function: $X(\omega)=\sum_{k=1}^{n} x_k \mathbf{1}_{A_k}(\omega)$

$\Omega_X = \{x_k \in R; k=1,\ldots,n\} \subset R$

► **elementary random quantity:**

$\{A_k; k=1,\ldots\}$ countable partition of Ω

elementary function: $X(\omega)=\sum_{k=1}^{\infty} x_k \mathbf{1}_{A_k}(\omega)$

$\Omega_X = \{x_k \in R; k=1,\ldots\} \subset R$

$$\boxed{P(X=x_k)=P(A_k)=\int_\Omega \mathbf{1}_{A_k}(\omega)dQ(\omega)}$$

non-denumerable set

Continuous

$\Omega_X \subseteq R$ non-denumerable set

$$P(X\in A)=\int_R \mathbf{1}_A(\omega)dP(\omega)$$

► $X(\omega)$ **absolutely continuous:**

$$P(X\in A)=\int_R \mathbf{1}_A(\omega)dP(\omega)=$$

$$=\int_A dP(\omega)=\int_A f(\omega)dw$$

(Radon-Nikodym Theorem)

► **singular**

Radon-Nikodym Theorem (1913;1930)

ν, μ ***two σ-finite measures over the measurable space*** (Ω, B)

If $\nu << \mu$ $\quad \left(absolutely\ \ continuous: \quad \mu(A)=0 \Rightarrow \nu(A)=0 \qquad \forall A \in B \right)$

$\exists\, f(w)$ ***measurable function over*** B

with range in $[0,\infty)$ ***(non-negative)***

unique ***(if*** $g(w)$ ***same properties as*** $f(w) \longrightarrow \mu\{x \mid f(x) \neq g(x)\} = 0$***)***

such that $\forall A \in B$

$$\nu(A) = \int_A d\nu(w) = \int_A \frac{d\nu}{d\mu} d\mu(w) = \int_A f(w)\, d\mu(w)$$

$$\left(\Leftarrow if\ \exists\, f(w)\ then\ \ \nu << \mu \right)$$

Probability density function:

$$(\Omega, B, Q) \xrightarrow{X(w):\, w \in \Omega \rightarrow R} (R, B_R, P)$$

$$Q(A) = \int_{\Delta \in B_R \mid X^{-1}(\Delta) = A} p(x)\, dx$$

$\mu(w)$ *Lebesgue measure*

Remember that:

The set of points of R with finite probabilities is denumerable

Set of points of R with finite probabilities $W = \{\forall x \in R \mid P(x) > 0\}$

$\{W_k\}$ *partition of* $W = \bigcup_{K=1}^{\infty} W_K$

$$W_1 = \{\forall x \in \Re \mid 1/2 < P(x) \leq 1\}$$
$$W_2 = \{\forall x \in \Re \mid 1/3 < P(x) \leq 1/2\}$$
$$\cdots$$
$$W_K = \{\forall x \in \Re \mid 1/(k+1) < P(x) \leq 1/k\}$$
$$\cdots$$

If $x \in W$, *then* $x \in$ *some* W_k

If $x \in W_k$, *then* $x \in W$

Each set W_k *has, at most, k points for otherwise* $\sum_i P_k(x_i) > 1$

W if the denumerable union of finite sets $\longrightarrow$ ***W is a denumerable set***

$\sum_{\forall i} P(x_i) = 1$ ***so if it is*** ∞***-denumerable, it is not possible that all the points have the same probability***

If X is AC $\longrightarrow \mu([a]) = 0 \longrightarrow P(X = a) = 0$ ***but {X=a} is not an impossible result***

DISTRIBUTION FUNCTION

Distribution Function

Gen.Def: ***One-dimensional DF*** $\forall F : x \in \Omega_X \subset R \to R$

1) Continuous on the right: $\lim_{\varepsilon \to 0^+} F(x+\varepsilon) = F(x) \quad ; \quad \forall x \in R$

2) Monotonous non-decreasing: $if \quad x_1, x_2 \in R \quad and \quad x_1 \le x_2$

$$\to F(x_1) \le F(x_2)$$

3) Limits: $\lim_{x \to -\infty} F(x) = 0 \quad ; \quad \lim_{x \to +\infty} F(x) = 1$

$$F(x+0^+) = F(x) \qquad F(-\infty) = 0 \quad ; \quad F(+\infty) = 1$$

Distribution Function of a Random Quantity $X(w)$

Def.- ***DF associated to the Random Quantity*** X ***is the function***

$$F(x) = P(X \le x) = P\big(X \in (-\infty, x]\big) \quad ; \quad \forall x \in R$$

The Distribution Function of a Random Quantity has all the information needed to describe the properties of the random process.

Properties

$$F(x) = P(X \le x) = P\big(X \in (-\infty, x]\big) \quad ; \quad \forall x \in R$$

$$P(X < x) = F(x - \varepsilon)$$

$$P(X > x) = 1 - P\big(X \le x\big) = 1 - F(x)$$

$$P(x_1 < X \le x_2) = P\big(X \in (x_1, x_2]\big) = F(x_2) - F(x_1)$$

Properties of the DF (some)

DF defined $\forall x \in R$

If X **takes values in** $[a,b] \in R \longrightarrow F(x) = \begin{cases} 0 & \forall x < a \\ 1 & \forall x \geq b \end{cases}$

Set of points of discontinuity of the DF is finite or denumerable

$$F(x+\varepsilon) \neq F(x-\varepsilon)$$

$$D = \{\forall x \in R / F(x+\varepsilon) \neq F(x-\varepsilon)\}$$

1) monotonous non-decreasing $F(x-\varepsilon) < F(x+\varepsilon)$

2) $\forall x \in D \rightarrow r(x) \in Q$ **such that** $F(x-\varepsilon) < r(x) < F(x+\varepsilon)$

3) $x_1, x_2 \in D / x_1 < x_2$ $\quad r(x_1) < F(x_1+\varepsilon) \leq F(x_2-\varepsilon) < r(x_2)$

$r(x)$ **associated is different for each x** $\longrightarrow \forall x \in D \rightarrow r(x) \in Q$

is a one-to-one relation

At each point of discontinuity …

$$P(x-\varepsilon < X \leq x) = P(X \in (x-\varepsilon, x]) = F(x-\varepsilon) - F(x)$$

$$\lim_{\varepsilon \to 0^+} [F(x) - F(x-\varepsilon)] = P(X = x)$$

$F(x)$ **has a jump of amplitude** $P(X = x)$

Distribution Function ⟷ **Probability Measure**

- **For each DF there exists a unique probability measure defined over Borel Sets that assigns the probability** $F(x_2)-F(x_1)$ **to each half-open interval** $(x_1, x_2] \in R$

- **Reciprocally, to each probability measure defined on the measurable space** (Ω, B) **, corresponds a DF**

Random Quantity ⟷ **discrete / continuous** ⟷ **Distribution Function**

singular or absolutely continuous

Discrete Random Quantity

$$X \quad (\Omega, B, Q) \xrightarrow{X(w):\Omega \to R} (R, B_R, P)$$

Range of $X(w): \Omega_X \subseteq R$ ***finite or denumerable set***

takes values $\Omega_X = \{x_1, x_2, \ldots\}$

with probabilities $\{p_1, p_2, \ldots\}$

$$P(X = x_i) = p_i$$

p_k ***real, non-negative and*** $\sum_{\forall k} p_k = 1$

DF:

$$F(x) = P(X \le x) = \sum_{\forall k} p_k \mathbf{1}_{(-\infty, x]}(x_k)$$

$$F(-\infty) = 0 \quad ; \quad F(+\infty) = 1$$

1) Step-wise and monotonous non-decreasing

2) Constant everywhere but on points of discontinuity where it has a jump

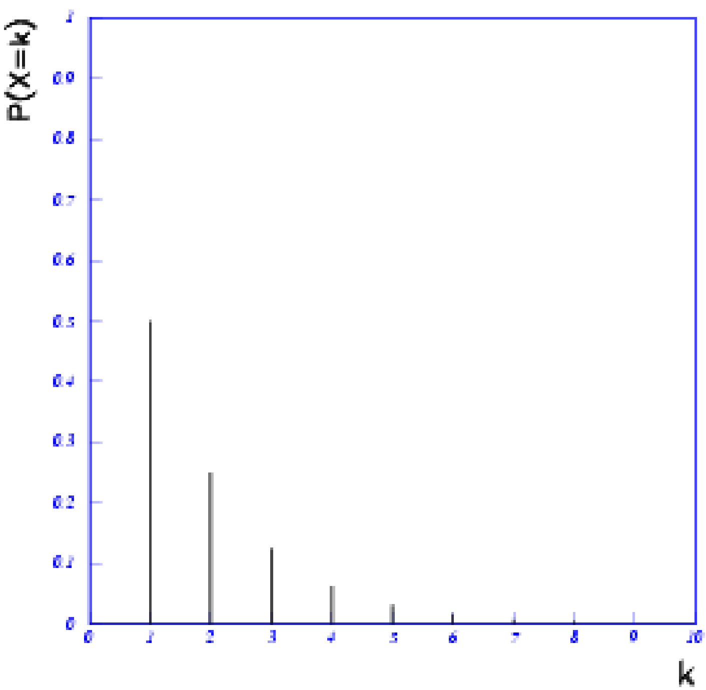
$$F(x_k) - F(x_k - \in) = P(X = x_k) = p_k$$

EXAMPLE:

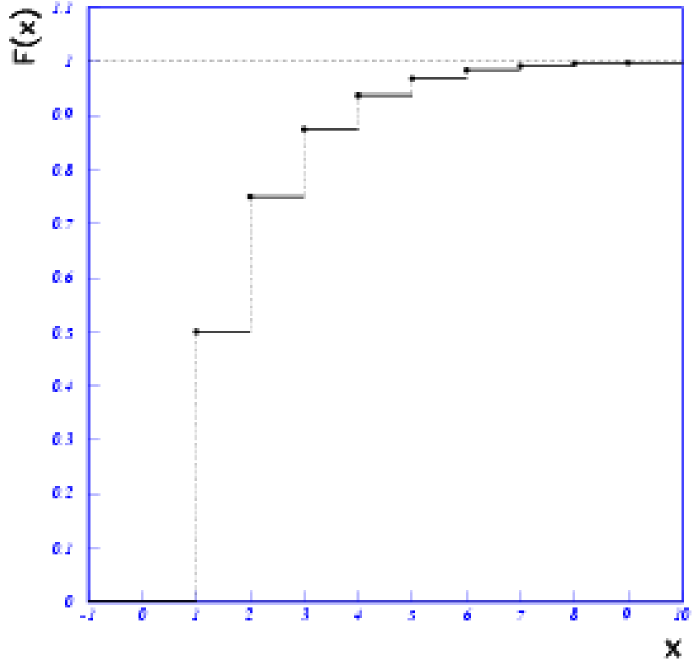
$$\Omega_X = \{1,2,\ldots\}$$

$$p_k = P(X = k) = 1/2^k$$

$$\sum_{\forall i} p_i = 1$$



$$F(x) = P(X \leq x) = \sum_{\forall k} p_k \mathbf{1}_{(-\infty,x]}(x_k)$$



$$F(x_k) - F(x_k - \in) = P(X = x_k) = p_k$$

$$F(-\infty) = 0 \quad ; \quad F(+\infty) = 1$$

Continuous Random Quantity

$$(\Omega, B, Q) \xrightarrow{X(w):\Omega \to R} (R, B_R, P)$$

Range of $X(w): \Omega_X \subseteq R$ **non-denumerable set**

- $F(x)$ **continuous everywhere in R**

$$F(x+\varepsilon) = F(x)$$

$$F(x-\varepsilon) = F(x) - \underbrace{P(X=x)}_{=0} = F(x)$$

- **AC: Radon-Nikodym** $\quad P(A) = \int_A dP = \int_A \frac{dP}{d\mu} d\mu = \int_A p(w)dw$

μ **Lebesgue measure**

$p(x)$ **Probability Density Function** $\quad P(X \le x) = F(x) = \int_{-\infty}^{x} p(u)du$

$$p(x) = \frac{dF(x)}{dx}$$ ***uniquely a.e.***

1) $p(x) \ge 0 \quad ; \quad \forall x \in R$

2) bounded in every bounded interval of R **and Riemann integrable on it**

3) $\int_{-\infty}^{\infty} p(x)dx = 1$

EXAMPLE:

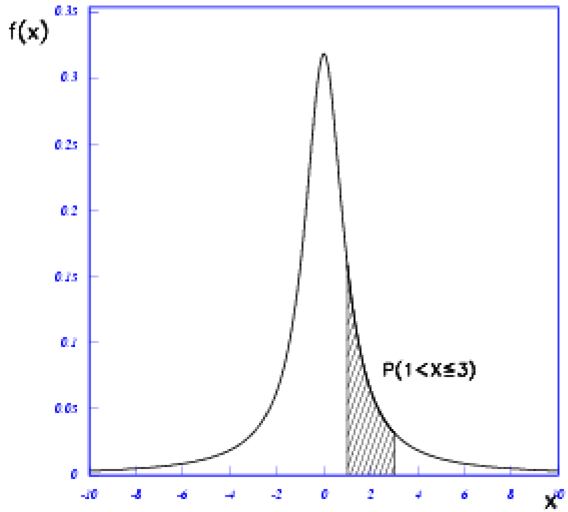


Figura 4.3.- Función densidad de probabilidad (ver el ejemplo 4.3). El area marcada corresponde a la probabilidad $P(1 < X \leq 3)$.

$$p(x) = \frac{1}{\pi} \frac{1}{1 + x^2}$$

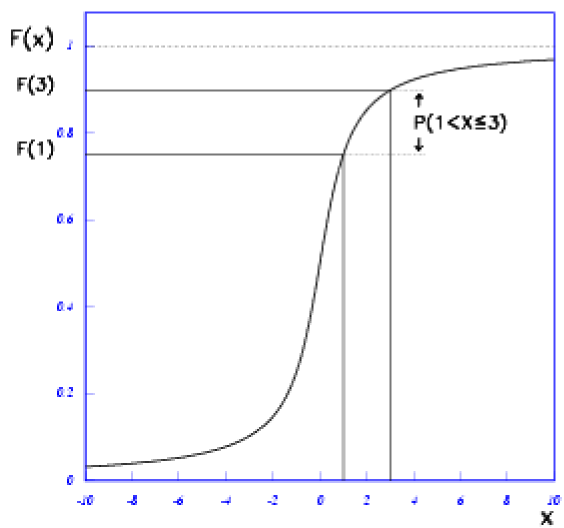


Figura 4.4.- Función de distribución (ver el ejemplo 4.3). La diferencia de ordenadas $F(3) - F(1)$ corresponde a la probabilidad $P(1 < X \leq 3)$.

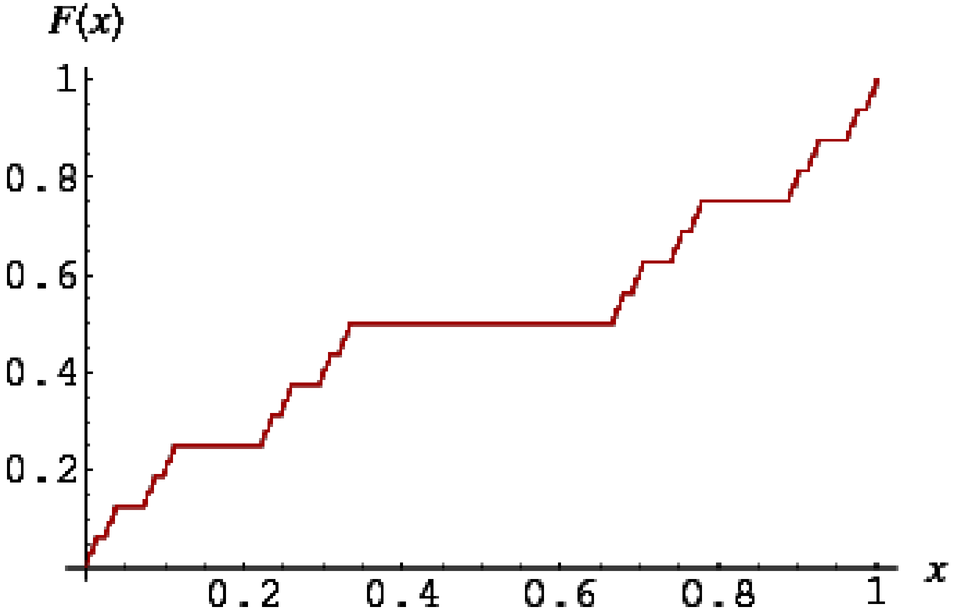
$$F(x) = \int_{-\infty}^{x} p(u)du =$$

$$= \frac{1}{2} + \frac{1}{\pi} \arctan(x)$$

EXAMPLE: $X \sim Cs(0,1)$

$$X = \sum_{n=1}^{\infty} \frac{X_n}{3^n} \qquad \operatorname{supp}[X_n] = \{0,2\}$$

$$P(X_n = 0) = P(X_n = 2) = \frac{1}{2}$$



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

$$\cancel{F(x) = \int_{-\infty}^{x} p(u)\,du}$$

General Distribution Function (Lebesgue Decomposition)

$$F(x) = \underbrace{\sum_{i=1}^{N_D} a_i\, F_i^{D}(x)}_{} + \underbrace{\sum_{j=1}^{N_C} b_j\, F_j^{AC}(x) + \sum_{k=1}^{N_S} a_k\, F_k^{D}(x)}_{}$$

discrete

Step Function

(simple or elementary) with denumerable number of jumps

$$P(X = x_n)$$

$$\sum_n P(X = x_n) = 1$$

(Poisson, Binomial,…)

Abs. continuous

$$F(x) = \int_{-\infty}^{x} p(u)du$$

$$p(x) = F'(x)$$

almost everywhere

pdf: $p(x) \mid \int_{-\infty}^{\infty} p(x)dx = 1$

(Normal, Gamma,…)

Singular

$F(x)$ ***continuous***

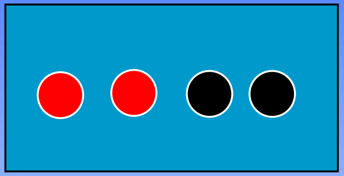
$F'(x) = 0$ ***almost everywhere***

(Dirac Delta, Cantor,…)

CONDITIONAL PROBABILITY and BAYES THEOREM

Given a probability space (Ω, B, P)

- *The information assigned to an event* $A \in B$ *depends on the information we have*



Two consecutive extractions without replacement:

What is the probability to get a red ball in the second extraction?

1) I do not know the outcome of the first : P(r)=1/2

2) It was black: P(r)=2/3

All probabilities are conditional

Conditional Probability

Statistical Independence

Consider (Ω, B_Ω, P)

and two not disjoint sets

$$A, B \subset B_\Omega \qquad A \bigcap B \neq \emptyset$$

$$P(A) \equiv P(A \bigcap \overbrace{E}^{E = B \bigcup \overline{B}}) = \underbrace{P(A \bigcap B)}_{\text{Probability to happen } A \text{ and } B} + \underbrace{P(A \bigcap \overline{B})}_{A \text{ and not } B}$$

$$= P(A, B) + P(A, \overline{B})$$

What is the probability for A to happen

if we know that B has already occured? $\equiv P(A|B)$

$$\left.\begin{array}{l} P(A|B) = C \times P(A \bigcap B) \\ P(B|B) = 1 = C \times P(B \bigcap B) = C \times P(B) \\ \longrightarrow \qquad C^{-1} = P(B) \end{array}\right\}$$

$$\boxed{P(A|B) \equiv \frac{P(A, B)}{P(B)}}$$

$$P(B) \neq 0$$

Notation: $P(A \bigcap B \bigcap C \bigcap \ldots) \equiv P(A, B, C, \ldots)$

Generalization: $P(A_1, A_2, \ldots, A_n) =$

$$= P(A_1 | A_2, \ldots, A_n) P(A_2, \ldots, A_n) =$$

$$= P(A_1 | A_2, \ldots, A_n) P(A_2 | A_3, \ldots, A_n) \cdots P(A_n)$$

$n!$ ***possible arrangements***

$$P(A_1, A_2, \ldots, A_n) = P(A_2 | A_1, \ldots, A_n) P(A_1 | A_3, \ldots, A_n) \cdots P(A_n)$$

Statistical Independence

$P(A|B) = P(A) \longrightarrow A$ ***does not depend on*** B

That B has already happened does not change the probability of ocurrence of A

$P(A|B) \neq P(A) \longrightarrow$ **Correlation** $\begin{cases} +: & P(A|B) > P(A) \\ -: & P(A|B) < P(A) \end{cases}$

Caution !

For a finite collection of n events $A = \{A_1, A_2, \ldots, A_n\} \subset B$

are independent iff: $P(A_p, \ldots, A_m) = P(A_p) \cdots P(A_m)$

for each subset $\{A_p, \ldots, A_m\} \subset A$

$$P(A_i, A_j) = P(A_i)P(A_j) \qquad i, j = 1, \ldots, n \qquad i \neq j$$

$$P(A_i, A_j, A_k) = P(A_i)P(A_j)P(A_k) \qquad i, j, k = 1, \ldots, n \qquad i \neq j \neq k$$

• • •

Conditional dependence

$P(A|B) = P(A)$ ⟶ A ***independent of*** B ***…***

…should say "inconditionally" independent

It may happen that A ***depends on*** B ***through*** C

$P(A, B) = P(A)P(B)$ ***but*** $P(A, B|C) \neq P(A|C)P(B|C)$

Theorem of Total Probability

Partition of the Sample Space

$$\{B_k, k = 1, \ldots n\}$$

$$\Omega = \bigcup_{j=1}^{n} B_j \qquad B_i \bigcap_{i \neq j} B_j = \emptyset$$

$$P(A) \equiv P(A \cap \Omega) = P\left(A \cap \left\{\bigcup_{k=1}^{n} B_k\right\}\right) = P\left(\bigcup_{k=1}^{n} \{A \cap B_k\}\right) =$$

$$= \sum_{k=1}^{n} P(A \cap B_k) = \sum_{k=1}^{n} P(A|B_k)\, P(B_k)$$

$$\boxed{P(A) = \sum_{k=1}^{n} P(A, B_k) = \sum_{k=1}^{n} P(A|B_k)\, P(B_k)}$$

Theorem of Total Probability with Conditional Probabilities

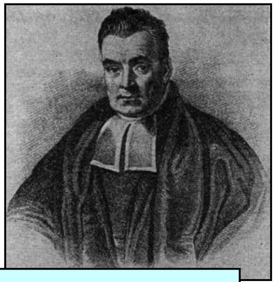
$$P(A, B, C) = P(A|B, C)P(B, C) = P(A|B, C)P(C|B)P(B)$$

$$P(A, B) = \sum_{C} P(A, B, C)$$

$$\boxed{P(A|B) = \sum_{C} P(A|C, B) \cdot P(C|B)}$$

Bayes Theorem

The Reverend Thomas Bayes, F.R.S.
(1701?-1761)



LII. An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S.

Dear Sir,

Read Dec. 23, 1763. I now send you an essay which I have found among the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved. Experimental philosophy, you will find, is nearly interested in the subject of it; and on this account there seems to be particular reason for thinking that a communication of it to the Royal Society cannot be improper.

… to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.

…some rule could be found, according to which we ought to estimate the chance that the probability for the happening of an event perfectly unknown, should lie between any two named degrees of probability, antecedently to any experiments made about it; …

Common sense is indeed sufficient to shew us that, form the observation of what has in former instances been the consequence of a certain cause or action, one may make a judgement what is likely to be the consequence of it another time.

$$P(A,B) = P(A|B)\cdot P(B) = P(B|A)\cdot P(A) \rightarrow P(B|A) = \frac{P(A|B)\cdot P(B)}{P(A)}$$

Partition of the Sample Space $\{H_k, k = 1,\ldots n\}$

Probability of occurrence of event A having occurred (cause, hypothesis) H_i

Probability of occurrence of the event (cause, hypothesis) H_i "a priori", before we know if event A has happened

$$P(H_i|A) = \frac{P(A|H_i)\,P(H_i)}{P(A)}$$

$$i = 1,\ldots n$$

normalisation

Probability ("a posteriori") fo event H_i to happen having observed the occurrence of event (efect) A

Probability that H_i be the cause (hypothesis) of the observed effect A

Other forms of Bayes Theorem:

Partition of the Sample Space

$$\{H_k, k = 1, \ldots n\}$$

BT+ Total Probability Theorem

$$P(A) = \sum_{k=1}^{n} P(A|H_k)\, P(H_k)$$

$$P(H_i|A) = \frac{P(A|H_i)\, P(H_i)}{\sum_{k=1}^{n} P(A|H_k)\, P(H_k)}$$

+ general hypothesis (H_0) ***(all probabilities are condicional)***

$$P(H_i|A, H_0) = \frac{P(A|H_i, H_0)\, P(H_i|H_0) P(H_0)}{\sum_{k=1}^{n} P(A|H_k)\, P(H_k|H_0) P(H_0)}$$

$$P(A, H_0)$$

… and many more to come…

Marginal and Conditional Densities

(Stieltjes-Lebesgue integral $\sum \rightarrow \int$)

$$F(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2)$$

$$F(x_1, x_2) = \int_{-\infty}^{x_1} dw_1 \int_{-\infty}^{x_2} p(w_1, w_2) dw_2 \longrightarrow p(x_1, x_2)$$

$$p(x_1) = \int_{-\infty}^{\infty} p(x_1, x_2) dx_2 \qquad p(x_2) = \int_{-\infty}^{\infty} p(x_1, x_2) dx_1$$

$$p(x_1, x_2) = p(x_2 \mid x_1) p(x_1) = p(x_1 \mid x_2) p(x_2)$$

$$p(x_2 \mid x_1) \overset{Def}{=} \frac{p(x_1, x_2)}{p(x_1)}$$

$$p(x_1 \mid x_2) \overset{Def}{=} \frac{p(x_1, x_2)}{p(x_2)}$$

$$(p(x) \ne 0)$$

$$p(x_2 \mid x_1) = p(x_2)$$

$$p(x_1 \mid x_2) = p(x_1)$$

EXAMPLE: CAUSE-EFFECT

Certain disease occurs in 1 out of 1000 individuals

There is a diagnostic test such that:

If a person is sic, gives positive in the 99% of the cases

If a person is healthy, gives positive in 2% of the cases

A person has given positive in the test. What are the chances that he is sic?

Hypothesis o causes to analyze

exclusive y exhaustive

$$\begin{cases} H_1 : & \textit{is sic} \\ H_2 = \overline{H}_1 : & \textit{is healthy} \end{cases}$$

Data

incidence in the population

"a priori" probabilities

$$P(H_1) = \frac{1}{1000} \qquad P(H_2) = 1 - P(H_1) = \frac{999}{1000}$$

if T denotes the event

T={give positive in the test }

$$P(T|H_1) = \frac{99}{100} \qquad P(T|\overline{H}_1) = \frac{2}{100}$$

A person has given positive.

What are the chances that he is sic?

$$P(H_1|T) = ???$$

H_1 : ***be sic*** ***T={give positive in the test }***

Bayes Theorem:

$$P(H_1|T) = \frac{P(T|H_1)\,P(H_1)}{P(T)}$$

Total Probability Theorem:

$$P(T) = \sum_{k=1}^{2} P(T|H_k)\,P(H_k)$$

$$P(H_1|T) = \frac{P(T|H_1)\,P(H_1)}{P(T) = \sum_{k=1}^{n} P(T|H_k)\,P(H_k)} = \frac{\frac{99}{100}\,\frac{1}{1000}}{\frac{99}{100}\,\frac{1}{1000} + \frac{2}{100}\,\frac{999}{1000}} = 0.047$$

$$P(T|H_1) = 0.99 \;\;!!$$

The test is costly, agresive,… if a person gives positive…

What are the chances that he is healty?

$$P(\overline{H}_1|T) = 1 - P(H_1|T) = 0.953$$

The disease is serious…

What are the chances to be sic giving negative in the test?

$$P(H_1|\overline{T}) = \frac{P(\overline{T}|H_1)\,P(H_1)}{1-P(T)} = \frac{(1-P(T|H_1))\,P(H_1)}{1-P(T)} \approx 10^{-5}$$

Probabilities of interest as function of known data and incidence of the disease in the population

Incidence of the disease in the population $P(H_1) = x$

Probability to be sic giving positive

$$P(H_1|T) = \frac{P(T|H_1)\,x}{P(T|H_1)\,x + P(T|\overline{H}_1)\,(1-x)}$$

Probability to be healthy giving positive

$$P(\overline{H}_1|T) = \frac{P(T|\overline{H}_1)\,(1-x)}{P(T|H_1)\,x + P(T|\overline{H}_1)\,(1-x)}$$

Probability to be sic giving negative

$$P(H_1|\overline{T}) = \frac{P(\overline{T}|H_1)\,x}{1-P(T)} = \frac{(1-P(T|H_1))\,x}{1-P(T)} =$$

$$= \frac{(1-P(T|H_1))\,x}{1-P(T|H_1)\,x - P(T|\overline{H}_1)\,(1-x)}$$

From the given data:

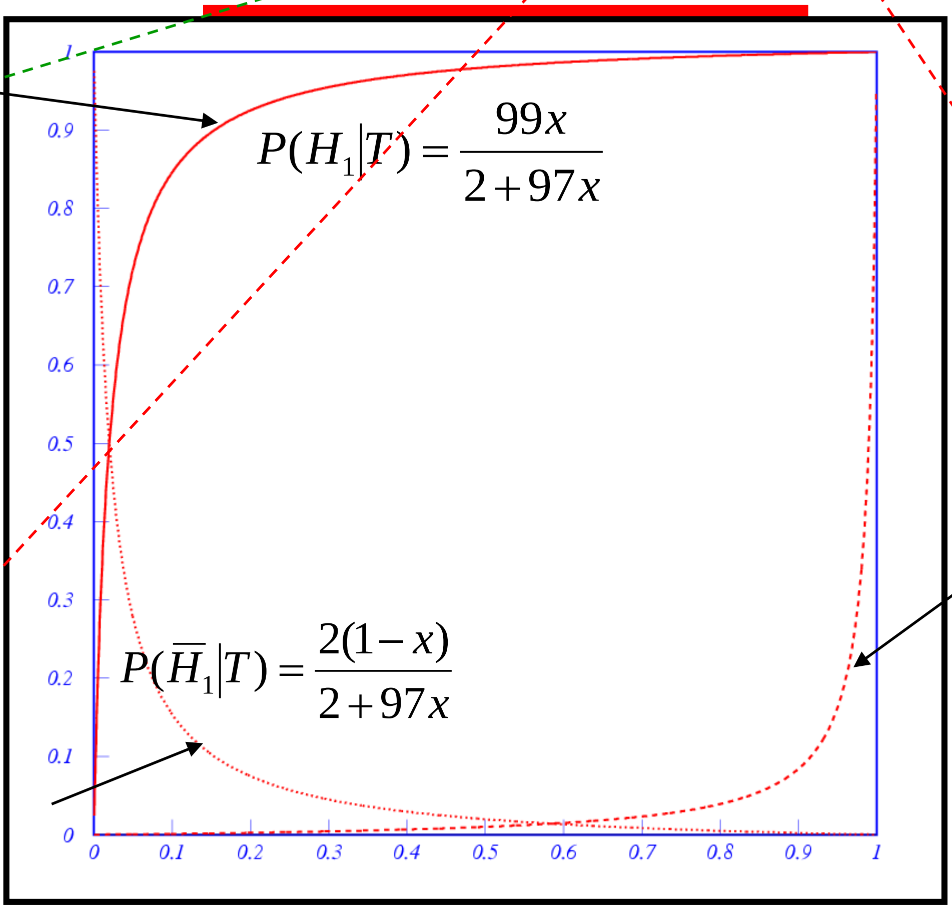
Interest to minimise $P(\overline{H}_1|T)$ $P(H_1|\overline{T})$

maximize $P(H_1|T)$

optimum

eficiency

correctly detected sic



sic undetected

$$P(H_1|\overline{T}) = \frac{x}{102-101x}$$

healthy giving positive

undergo costly and agresive treatment

$x = P(H_1)$ *incidence in the population*

Receiver Operating Characteristic (ROC curve)

For each cut value c

	H	$\overline{H}$
$+$	$P(+\vert H,c)$	$P(+\vert \overline{H},c)$
$-$	$P(-\vert H,c)$	$P(-\vert \overline{H},c)$

$$P(+|H,c) = F_1(c) = \int_{-\infty}^{c} f_1(u)du$$

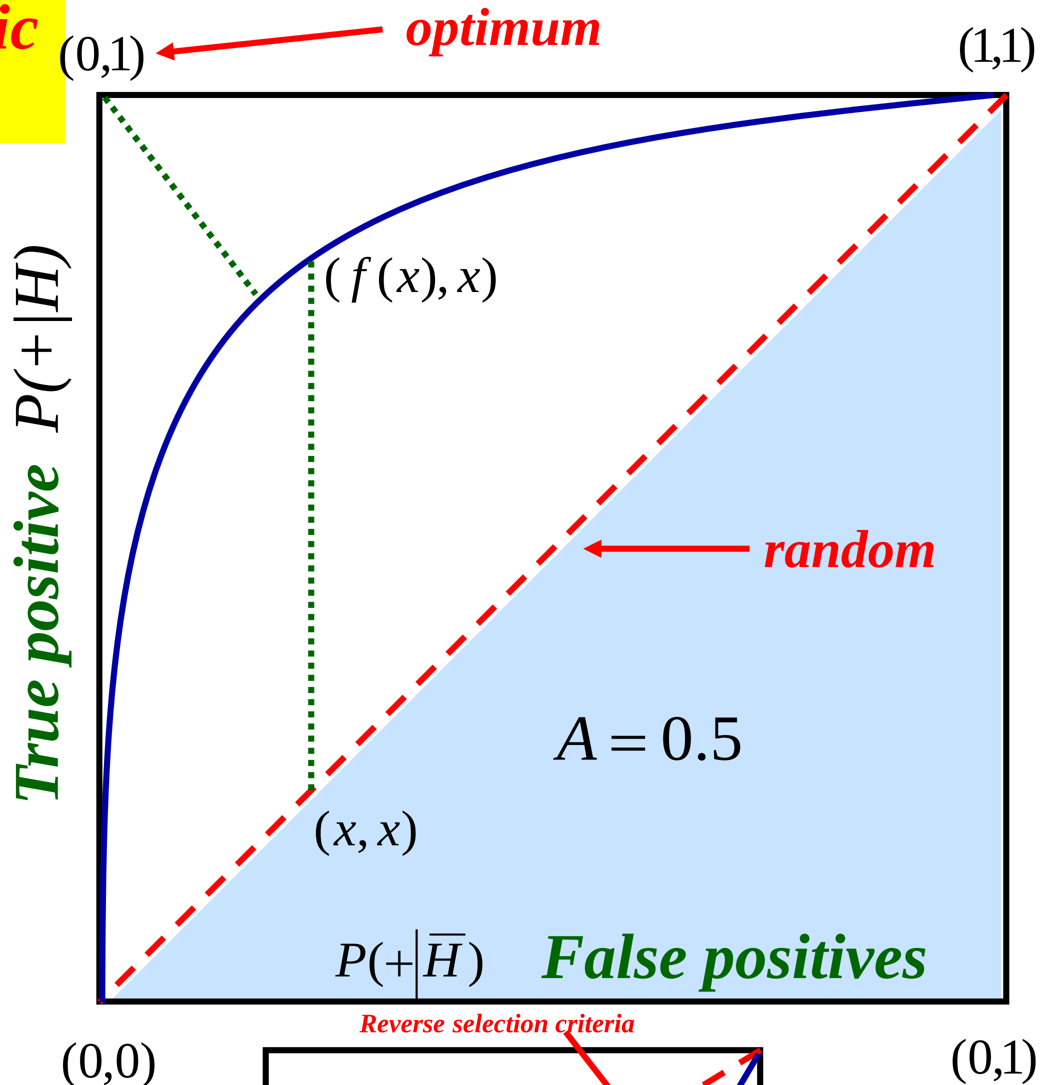
$$P(+|\overline{H},c) = F_2(c) = \int_{-\infty}^{c} f_2(u)du$$

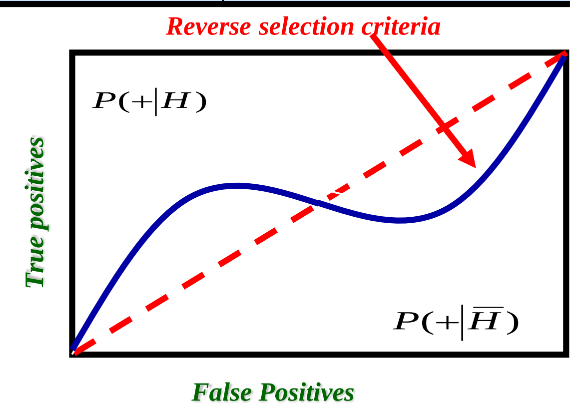
$$x = F_2(c) \rightarrow f(x) = F_1(F_2^{-1}(x))$$

$$A = \int_0^1 f(x)dx = \int_0^1 F_1(F_2^{-1}(x))dx$$

$$A = \int_{-\infty}^{\infty} F_1(u) f_2(u)du = \int_{-\infty}^{\infty} f_2(u)du \int_{-\infty}^{u} f_1(x)dx$$

$$d^2 = (f(x) - x)^2 \rightarrow \max(d)$$





STOCHASTIC CHARACTERISTICS

Mathematical Expectation

$$(\Omega, \mathcal{B}_\Omega, Q) \xrightarrow{X(\omega):\Omega \to R} (R, B, P)$$

$X(\omega)$ **Discrete**

$$\begin{cases} X(\omega) = \sum_{k=1}^{n} x_k \mathbf{1}_{A_k}(\omega) \\ X(\omega) = \sum_{k=1}^{\infty} x_k \mathbf{1}_{A_k}(\omega) \end{cases}$$

$$P(X = x_k) = P(A_k) = \int_R \mathbf{1}_{A_k}(\omega) dP(\omega)$$

$X(\omega)$ **Absolutely continuous**

$$P(X \in A) = \int_R \mathbf{1}_A(\omega) dP(\omega) = \int_A dP(\omega) = \int_A f(\omega) dw$$

$$Y = g[X(\omega)]$$

Def.: Expectation:

$$E[Y] = E[g(X)] \equiv \int_R g[X(\omega)] dP(\omega) = \begin{cases} \sum_k g(x_k) P(X = x_k) \\ \int_R g(x) dF(x) = \int_R g(x) f(x) dx \end{cases}$$

(Stieltjes-Lebesgue integral $\sum \to \int$ ***)***

Moments (wrt origin)

$$\alpha_n \equiv E[X^n] = \int_R x^n p(x)dx$$

$$x^n p(x) \in L_1(R)$$

$\alpha_0 = 1$ $\quad \exists \alpha_n \rightarrow \exists \alpha_{m<n}$ $\quad \nexists \alpha_n \rightarrow \nexists \alpha_{m>n}$ $\quad if\ \exists \alpha_{2n} \geq 0$

Mean: $\mu \equiv E[X] = \int_R xp(x)dx$

- **Linear operator** $\quad X = c_0 + \sum_i c_i X_i \xrightarrow[c_i \in R]{} E[X] = c_0 + \sum_i c_i E[X_i]$
- $\{X_i\}_{i=1}^n$ **independent** $\quad X = \prod_i X_i \longrightarrow E[X] = \prod_i E[X_i]$

Moments wrt point $c \in R$

$$E[(X-c)^n] = \int_R (x-c)^n p(x)dx$$

$\min_{c \in R} E[(X-c)^2]$ $\quad c = \mu$

…Moments wrt Mean

Moments wrt Mean

$$\mu_n = E[(X-\mu)^n] = \int_R (x-\mu)^n p(x)dx$$

Variance: $$\sigma^2 \equiv V[X] \equiv E[(X-\mu)^2] = \int_R (x-\mu)^2 p(x)dx \quad (>0)$$

▶ ***NOT Linear*** $Y = c_0 + c_1 X \longrightarrow V[Y] = \sigma_Y^2 = c_1^2 \sigma_X^2$

Skewness: $$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

Kurtosis: $$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

Watch!! $\boxed{x^n p(x) \in L_1(R)}$

Poisson $P(X=k) = e^{-\mu}\frac{\mu^k}{k!}$

$k = 0,1,2,\ldots$

$$\alpha_n = \sum_{k=0}^{\infty} X^n P(X=k) = e^{-\mu}\sum_{k=0}^{\infty} k^n \frac{\mu^k}{k!}$$

$$\lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| = \lim_{k\to\infty}\left|\frac{1}{k}\left(1+\frac{1}{k}\right)^{n-1}\right| = 0$$

Abs. Conv. $\longrightarrow \exists \alpha_n$

$$P(X=n) = \frac{6}{\pi^2 n^2}$$

$n = 1,2,\ldots$

$$\nexists \alpha_k \equiv E[X^k];\ k \geq 1$$

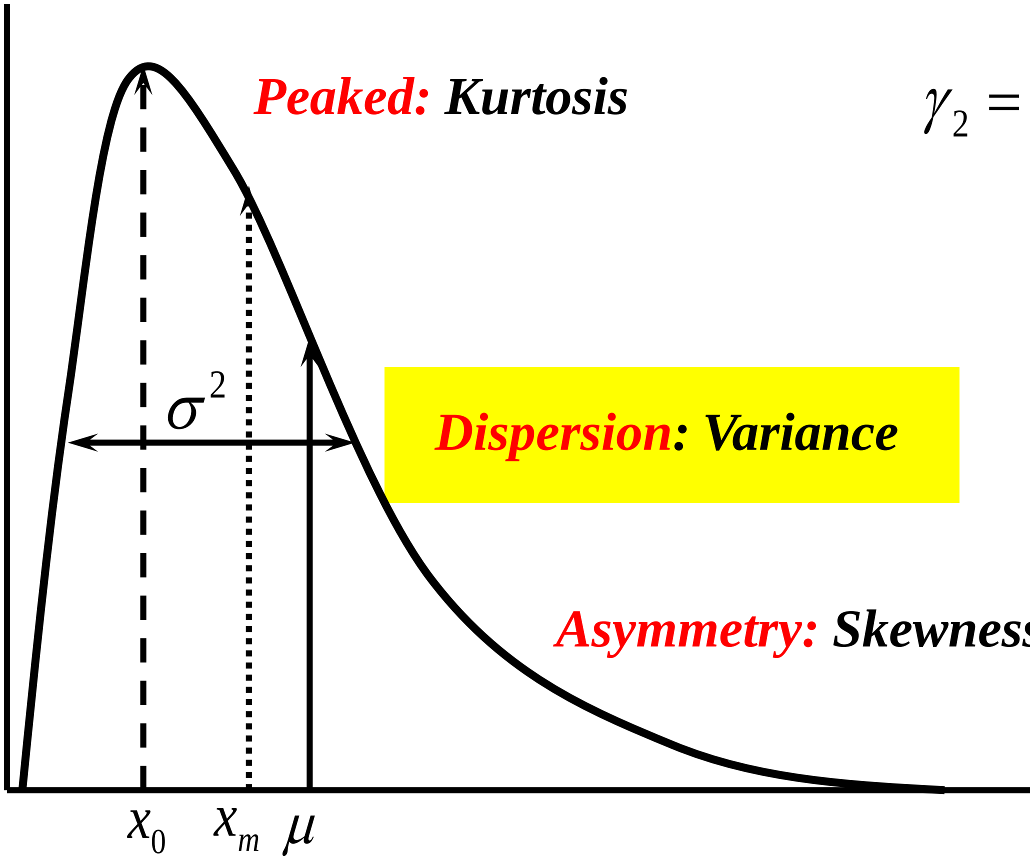
Cauchy $p(x) = \frac{1}{\pi(1+x^2)}$

$$\nexists \int_{-\infty}^{\infty} |x^n p(x)|\, dx \qquad n \geq 1$$

(CPV for $n=1$)

No moments (no mean, no variance,…)

Global Picture



Peaked: Kurtosis

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \begin{cases} > 0 & \\ = 0 & \textbf{Normal} \\ < 0 & \end{cases}$$

Dispersion: Variance

Asymmetry: Skewness

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \begin{cases} > 0 & \textbf{Right} \\ = 0 & \textbf{Symmetric} \\ < 0 & \textbf{Left} \end{cases}$$

Position: Mean: $E[X]$, …

Mode: $x_0 = \sup_{x \in \Omega} p(x)$

Median: $F(x_m) = P(X \le x_m) = 1/2$

quantile: $F(x_\alpha) = P(X \le x_\alpha) = q_\alpha$

$\gamma_1 > 0$ ***Mode < Median < Mean***

$\gamma_1 < 0$ ***Mode > Median > Mean***

Covariance (and "Linear Correlation")

$$V[X_1, X_2] = E[(X_1 - \mu_1)(X_2 - \mu_2)] = E[X_1 X_2] - \mu_1 \mu_2$$

$\{X_1, X_2\}$ ***independent*** $\Rightarrow$ / $\not\Leftarrow$ $V[X_1, X_2] = 0$

$$-1 \leq \rho_{12} = \frac{V[X_1, X_2]}{\sigma_1 \sigma_2} \leq 1$$

Holder inequality: $|\rho_{12}| \leq 1$

Comments

Linear relation: $X_2 = aX_1 + b$ $\rho_{12} = \pm 1$

Quadratic: $X_2 = a + cX_1^2$ $\quad if \quad for \quad X_1 \quad is \quad \gamma_1 = \dfrac{-2\mu}{\sigma}, \quad then \quad \rho_{12} = 0$

Useful expression:

Taylor Expansion for the Variance of $Y = g(X_1, X_2, \ldots) \quad E(X_i) = \mu_i$

$$Y = g(X_1, X_2) = g(\mu_1, \mu_2) + \left[\frac{\partial g}{\partial x_1}\right]_{(\mu_1,\mu_2)} (x_1 - \mu_1) + \left[\frac{\partial g}{\partial x_2}\right]_{(\mu_1,\mu_2)} (x_2 - \mu_2) + O(D_{ij}^2)$$

$$E[Y] = E[g(X_1, X_2)] = g(\mu_1, \mu_2) + O(D_{ij}^2)$$

$$\longrightarrow \quad Y - E[Y] = \left[\frac{\partial g}{\partial x_1}\right]_{(\mu_1,\mu_2)} (x_1 - \mu_1) + \left[\frac{\partial g}{\partial x_2}\right]_{(\mu_1,\mu_2)} (x_2 - \mu_2) + \ldots$$

$$V[Y] \equiv E\left[\left(Y - E[Y]\right)^2\right] \equiv \sigma_Y^2 \approx$$

$$= \left[\frac{\partial g}{\partial x_1}\right]^2_{(\mu_1,\mu_2)} V[X_1] + \left[\frac{\partial g}{\partial x_2}\right]^2_{(\mu_1,\mu_2)} V[X_2] + 2\left[\frac{\partial g}{\partial x_1}\frac{\partial g}{\partial x_2}\right]_{(\mu_1,\mu_2)} V[X_1 X_2] + \ldots$$

$$= \left[\frac{\partial g}{\partial x_1}\right]^2_{(\mu_1,\mu_2)} \sigma_1^2 + \left[\frac{\partial g}{\partial x_2}\right]^2_{(\mu_1,\mu_2)} \sigma_2^2 + 2\left[\frac{\partial g}{\partial x_1}\frac{\partial g}{\partial x_2}\right]_{(\mu_1,\mu_2)} \sigma_1 \sigma_2 \rho_{12} + \ldots$$

(mind for the re-maind-er…)

INTEGRAL TRANSFORMS

Fourier (… Laplace) Transform

Mellin Transform

Fourier Transform (Characteristic Function)

$$f : R \to C \qquad f \in L_1(R) \qquad \Phi(t) = \int_{-\infty}^{\infty} e^{ixt} f(x)dx \qquad t \in R$$

$$\Phi : t \in R \to C$$

Inversion Theorem (Lévy,1925)

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \Phi(t)dt$$

Discrete: $p(X = x_k) = \dfrac{1}{2\pi} \displaystyle\int_{-\infty}^{\infty} e^{-itx_k} \Phi(t)dt$

Reticular: $x_k = a + bn$

$a, b \in R, \quad b \neq 0, \quad n \in Z$

$$p(X = x_k) = \frac{b}{2\pi} \int^{\pi/b} e^{-itx_k} \Phi(t)dt$$

Probability Density…

$$\Phi(t) = E[e^{ixt}]$$

Exists for all $X(\omega)$

Properties:

- ► $\Phi(0) = 1$
- ► ***bounded*** $|\Phi(t)| \leq 1$
- ► ***Schwarz symmetry*** $\Phi(t) = \overline{\Phi}(-t)$
- ► ***Uniformly continuous in R***

$$\forall\, \varepsilon > 0, \exists \delta \mid |\Phi(t+\delta) - \Phi(t)| \leq \varepsilon$$

(all necessary but not sufficient)

► *One-to-one correspondence between DF and CF*

► *Two DF with same CF are the same a.e.*

Useful Relations:

- $$Y = g(X) \quad \to \quad \Phi_Y(t) = E[e^{iYt}] = E[e^{ig(X)t}]$$

$$Y = a + bX \qquad \Phi_Y(t) = e^{iat}\Phi_X(bt)$$
$$a, b \in R$$

- ***If*** $\{X_i \sim p_i(x_i)\}_{i=1}^{n}$ ***are n independent random quantities***

$$X = X_1 + \cdots + X_n \quad \Rightarrow \quad \Phi_X(t) = E[e^{it(X_1 + \cdots + X_n)}] = \Phi_1(t)\cdots\Phi_n(t)$$

$$X = X_1 - X_2 \qquad \Phi_X(t) = \Phi_1(t)\Phi_2(-t) = \Phi_1(t)\overline{\Phi}_2(t)$$

- ***If distribution of*** X ***is symmetric, then*** $\Phi_X(t)$ ***is a real function***

$$\Phi_X(t) = \Phi_{-X}(t) = \Phi_X(-t) = \overline{\Phi}_X(t)$$

Example

$$X_1 \sim Po(n_1 \mid \mu_1)$$
$$X_2 \sim Po(n_2 \mid \mu_2)$$
$$X = X_1 - X_2$$

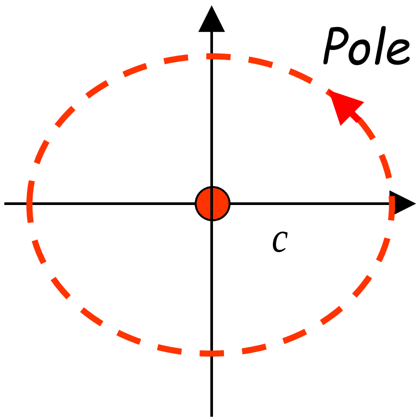
$$\Phi_i(t) = e^{-\mu_i(1-e^{it})}$$

$$\Phi_X(t) = \Phi_1(t)\overline{\Phi}_2(t) = e^{-(\mu_1+\mu_2)} e^{(\mu_1 e^{it} + \mu_2 e^{-it})}$$

X: Discrete reticular: a=0, b=1

$$P(X=n) = \left(\frac{\mu_1}{\mu_2}\right)^{n/2} \frac{e^{-\mu_S}}{2\pi i} \oint_C z^{-(n+1)} e^{-\frac{w}{2}(z+1/z)} dz$$

$$C: \left\{ |z| = \left(\frac{\mu_1}{\mu_2}\right)^{1/2} ; \theta \in (-\pi, \pi] \right\}$$



Pole of order n+1 at z=0

$$\mathrm{Res}\{f(z), z=0] = \sum_{p=0}^{\infty} \frac{(\mu_1\mu_2)^{n/2+p}}{\Gamma(p+n+1)\Gamma(p+1)}$$

$$P(X=n) = \left(\frac{\mu_1}{\mu_2}\right)^{n/2} e^{-(\mu_1+\mu_2)} I_{|n|}(2\mu_1\mu_2)$$

Some Useful Cases: $X = X_1 + \cdots + X_n$

$X_k \sim Po(x_k \mid \mu_k)$	$X \sim Po(x \mid \mu_S)$	$\mu_S = \mu_1 + \cdots + \mu_n$
$X_k \sim N(x_k \mid \mu_k, \sigma_k)$	$X \sim N(x \mid \mu_S, \sigma_S)$	$\mu_S = \mu_1 + \cdots + \mu_n$ $\sigma_S^2 = \sigma_1^2 + \cdots + \sigma_n^2$
$X_k \sim Ca(x_k \mid a_k, b_k)$	$X \sim Ca(x \mid a_S, b_S)$	$a_S = a_1 + \cdots + a_n$ $b_S = b_1 + \cdots + b_n$
$X_k \sim Ga(x_k \mid a, b_k)$	$X \sim Ga(x \mid a, b_S)$	$b_S = b_1 + \cdots + b_n$

Moments of a Distribution

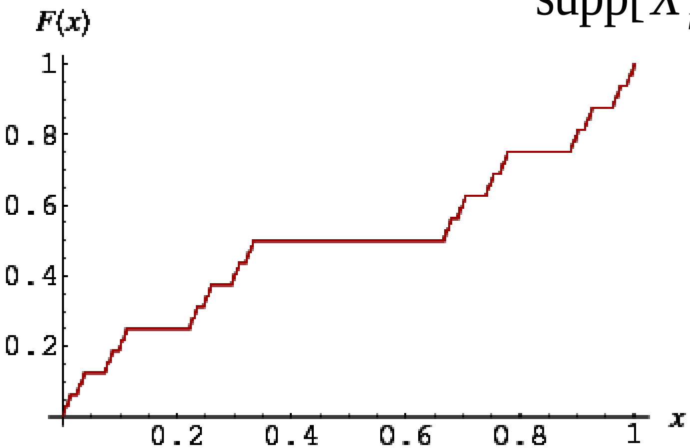
(F-LT usually called "moment generating functions)

$$\Phi(t) = E[e^{iXt}] \quad \rightarrow \quad E[X^k] = (-i)^k \left[\frac{\partial^k}{\partial^k t} \Phi(t) \right]_{t=0}$$

$$\Phi(t_1,\ldots,t_n) = E[e^{i(X_1 t_1 + \ldots + X_1 t_n)}] \longrightarrow E[X_i^{k_i} X_j^{k_j}] = (-i)^{k_i + k_j} \left[\frac{\partial^{k_i}}{\partial^{k_i} t_i} \frac{\partial^{k_j}}{\partial^{k_j} t_j} \Phi(t_1,\ldots,t_n) \right]_{t_1,\ldots,t_n=0}$$

$$X \sim Cs(0,1) \qquad X = \sum_{n=1}^{\infty} \frac{X_n}{3^n}$$

$$\mathrm{supp}[X_n] = \{0,2\}$$



$$P(X_n = 0) = P(X_n = 2) = \frac{1}{2}$$

$$\Phi(t) = E[e^{iXt}] = \frac{1}{2}\left(1 + e^{2it}\right)$$

$$\Phi^{1)}(0) = \frac{i}{2} \quad \rightarrow \quad E[X] = \frac{1}{2}$$

$$\Phi^{2)}(0) = -\frac{3}{8} \quad \rightarrow \quad E[X^2] = \frac{3}{8} \quad \rightarrow \sigma^2 = \frac{1}{8}$$

Mellin Transform

$f: R^+ \to C \qquad f \in L_1(R^+)$

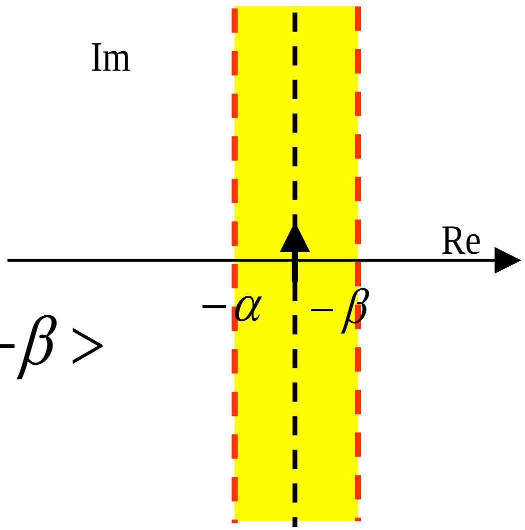
$$M(f;s) = \int_0^\infty f(x) x^{s-1} dx \qquad f(x) = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} M(f,s) x^{-s} ds$$

Obviously, if exists … $\quad s \in \Lambda \subseteq C$

Probability Densities… $\quad M(f;s) = E[X^{s-1}]$

Strip of holomorphy

$< -\alpha, -\beta >$



$$\lim_{x \to 0^+} f(x) = O(x^\alpha)$$

$$\lim_{x \to \infty} f(x) = O(x^\beta)$$

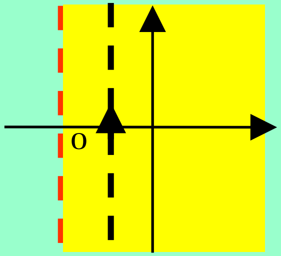
$$-\alpha < \mathrm{Re}(s) < -\beta$$

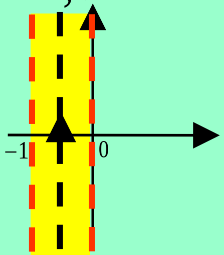
$$\mathrm{Re}(\sigma) \in < -\alpha, -\beta >$$

$f_1(x) = e^{-x} \qquad f_2(x) = e^{-x} - 1$

$M(f_{1,2};s) = \Gamma(s)$

$< 0, \infty > \qquad < -1, 0 >$





$$M(f;s) \oplus < -\alpha, -\beta >$$

Useful Relations:

- $Y = aX^{b}$ $\quad M_Y(s) = a^{s-1} M_X(bs-b+1)$

  $a, b \in R, \quad a > 0$

  $a=1, b=-1 \quad Y = X^{-1} \quad M_Y(s) = M_X(2-s)$

- $\{X_i \sim p_i(x_i)\}_{i=1}^{n}$ $\quad X = X_1 X_2 \cdots X_n \quad M_X(s) = M_1(s) \cdots M_n(s)$

  independent

  $x_i \in [0, \infty)$ $\quad X = X_1 X_2^{-1} \quad M_X(s) = M_1(s) M_2(2-s)$

  Non- negative

Example

$$X_1 \sim Ex(x_1 \mid a_1)$$

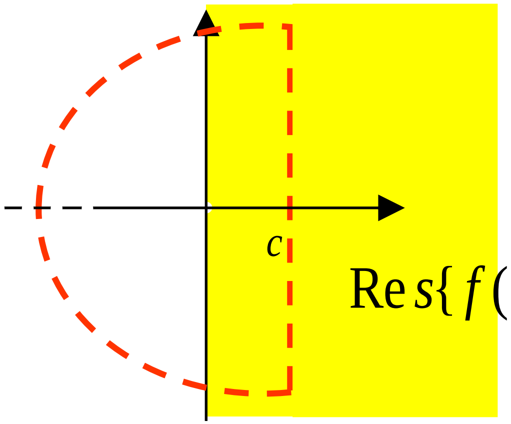
$$X_2 \sim Ex(x_2 \mid a_2)$$

$$X = X_1 X_2$$

$$M_i(s) = \frac{\Gamma(s)}{a_i^{\,s-1}} \qquad <0,\infty>$$

$$M_X(s) = \frac{\Gamma(s)^2}{(a_1 a_2)^{s-1}} \qquad p(x) = \frac{a_1 a_2}{2\pi i}\int_{c-i\infty}^{c+i\infty} (a_1 a_2 x)^{-z}\,\Gamma(z)^2\, dz$$

$$c > 0 \quad \textbf{\textit{real}}$$



$$\mathrm{Res}\{f(z), z_n = -n] = \frac{(a_1 a_2 x)^n}{\Gamma(n+1)^2}\left(2\Psi(n+1) - \ln(a_1 a_2 x)\right)$$

Newman series…

$$p(x) = 2 a_1 a_2 K_0(2\sqrt{a_1 a_2 x})$$

Some Useful Cases:

$$X = X_1 X_2 \qquad\qquad X = X_1 X_2^{-1}$$

$X_1 \sim Ex(x_1 \mid a_1)$

$X_2 \sim Ex(x_2 \mid a_2)$

$$\sim 2a_1 a_2 K_0(2\sqrt{a_1 a_2 x}) \qquad\qquad \sim a_1 a_2 (a_2 + a_1 x)^{-2}$$

$X_1 \sim Ga(x_1 \mid a_1, b_1)$

$X_2 \sim Ga(x_2 \mid a_2, b_2)$

$$\sim \frac{2 a_1^{b_1} a_2^{b_2}}{\Gamma(b_1)\Gamma(b_2)} \left(\frac{a_2}{a_1}\right)^{\upsilon/2} x^{(b_1+b_2)/2-1} K_\nu(2\sqrt{a_1 a_2 x}) \qquad \sim \frac{\Gamma(b_1+b_2)}{\Gamma(b_1)\Gamma(b_2)} \frac{a_1^{b_1} a_2^{b_2} x^{b_1-1}}{(a_1 + a_2 x)^{b_1+b_2}}$$

$$\upsilon = b_2 - b_1 > 0$$

obviously…

$$p(x) = \int_0^\infty p_1(w)\, p_2(x/w) \, {}^1\!/_w \, dw \qquad\qquad p(x) = \int_0^\infty p_1(xw)\, p_2(w) dw$$

…Densities with support in R…

$X_1 \sim N(x_1 \mid 0, \sigma_1)$

$X_2 \sim N(x_2 \mid 0, \sigma_2)$

$$X = X_1 X_2 \sim \left(\frac{a}{\pi}\right) K_0(a \mid x \mid)$$

$$a = (\sigma_1 \sigma_2)^{-1}$$

DISTRIBUTIONS AND GENERALIZED FUNCTIONS

Distribution

Functional: $T:\phi(x)\in D \to < T,\phi > C$

$$D = C_c^{\infty}$$

Linear functional: $< T,\alpha\phi_1+\beta\phi_2 >= \alpha < T,\phi_1 > +\beta < T,\phi_2 >$ $\qquad \alpha,\beta \in C$

Continuous functional: $\{\phi_n\} \xrightarrow{n\to\infty} \phi$

$$< T,\phi_n > \xrightarrow{n\to\infty} < T,\phi >$$

$T \in D'$

is a Distribution

$f:\Omega \subseteq R \to R$

Locally Lebesgue integrable (LLI) defines a distribution

$$< T_f,\phi >= \int_{-\infty}^{\infty} f(x)\phi(x)dx$$

***"regular"* distributions (*"singular"* the rest)**

Some Basic Properties

$$T, G \in D' \qquad \alpha T + \beta G \in D' \qquad \alpha, \beta \in C$$

$$\boxed{T = G \quad \Leftrightarrow \quad < T, \phi > \overset{(ae)}{=} < G, \phi >}$$

$$\text{supp}\, \{T\} \bigcap \text{supp}\, \{\phi\} = \emptyset \quad \Rightarrow \quad < T, \phi >= 0$$

$$\phi \in D; \phi\psi \in D \quad \Rightarrow \quad < \psi T, \phi >=< T, \psi\phi >$$

$$\boxed{< T', \phi >= - < T, \phi' > \qquad < D^p T, \phi >= (-1)^p < T, D^p \phi >}$$

$$\{T_n\} \xrightarrow{n \to \infty} T \quad iff \quad \forall \phi \quad < T_n, \phi > \xrightarrow{n \to \infty} < T, \phi >$$

$$\boxed{< \tilde{T}, \phi >=< T, \tilde{\phi} > \qquad \textbf{Fourier Transform}}$$

$$< S_a T, \phi > \equiv < T(x-a), \phi >=< T, \phi(x+a) >$$

$$< P_a T, \phi > \equiv < T(ax), \phi >= \frac{1}{|a|} < T, \phi(x/a) >$$

Two examples

$$\boxed{<\delta,\phi>=\phi(0)}$$

$$<\delta,\alpha\phi_1+\beta\phi_2>=\alpha\phi_1(0)+\beta\phi_2(0)=\alpha<\delta,\phi_1>+\beta<\delta,\phi_2>$$

$$\lim_{n\to\infty}|<\delta,\phi_n>-<\delta,\phi>|=\lim_{n\to\infty}|\phi_n(0)-\phi(0)|=0$$

$$<S_a\delta,\phi>\equiv<\delta_a,\phi>=<\delta,\phi(x+a)>=\phi(a)$$

$$<P_a\delta,\phi>\equiv<\delta_{ax},\phi>=\frac{1}{|a|}<\delta,\phi(x/a)>=\frac{1}{|a|}\phi(0)$$

$$<\delta',\phi>=-<\delta,\phi'>=-\phi'(0)$$

$$f_n=\frac{n}{2}\mathbf{1}_{[-1/n,1/n]}(x) \qquad \forall\phi \quad <T_n,\phi>\xrightarrow{n\to\infty}<\delta,\phi>$$

$$\boxed{H(x)=\mathbf{1}_{[0,\infty)}(x)}$$

LLI: defines a distribution

$$<T_H,\phi>=\int_{-\infty}^{\infty}\mathbf{1}_{[0,\infty)}(x)\phi(x)dx=\int_0^{\infty}\phi(x)dx$$

$$<H',\phi>=-<H,\phi'>=\phi(0)=<\delta,\phi>$$

Tempered Distributions

$$D = C_c^\infty$$

"rapidly decreasing" (Schwartz Space)

$$S = \{\phi : R \to C \quad | \quad \phi \in C^\infty \; and \quad \lim_{|x|\to\infty} | x^m D^n \phi | = 0 \quad \forall n, m \in N_0\}$$

EXAMPLE: $f(x) = e^{-x}$ $\quad x \in R \quad \notin S$

$x \in R^+ \quad \in S \qquad f(x) = e^{-x} \mathbf{1}_{(0,\infty)}(x)$

EXAMPLE: $f_n = \frac{n}{2} \mathbf{1}_{[-1/n,1/n]}(x) \qquad \forall \phi \quad |< T_n, \phi > - < \delta, \phi >| \xrightarrow{n\to\infty} 0$

δ **admissible** $\forall \phi \in C^0 \qquad \delta'$ **admissible** $\forall \phi \in C^1$

....

EXAMPLE: $f \in L_1(R) \qquad \int_R \frac{f(x)}{(a + x^2)^m} = C < \infty$ **for some** $m \in N_0$

defines a Tempered Distribution T_f

Convergence to zero $\quad \{\phi_n(x)\} \quad \phi_n \in S$

$$\max_{x\in R} | x^k D^m \phi_n(x) | \xrightarrow{n\to\infty} 0 \qquad \forall k, m \in N_0$$

$$< T, \phi_n(x) > \xrightarrow{n\to\infty} 0 \quad \Longrightarrow \quad T$$

Probability Distributions

$$(\Omega, B_\Omega, Q) \qquad X(\omega): \Omega \to R$$

$$F(x) = P(X \le x) = Q(X^{-1}(-\infty, x])$$ ***LLI: defines a distribution*** $< T_F, \phi >$

In general T is a Probability Distribution if: $< T, \phi > \ge 0 \quad \forall \phi \ge 0$

$$< T, 1 > = 1$$

$$\phi(x) = 1 \notin D, \notin S, \ldots$$

T does not have to be generated by a LLI function

…but any LLI function defines a probability distribution if

$$< T_f, \phi > = \int_{-\infty}^{\infty} f(x)\phi(x)dx \ge 0 \quad \forall \phi(x) \ge 0 \qquad < T_f, 1 > = \int_{-\infty}^{\infty} f(x)dx = 1$$

Probability Density "Distribution" $< T, \phi > = < T_F', \phi > = - < T_F, \phi' >$

Delta Distribution unifies discrete and AC random quantities:

$$X(\omega)\begin{cases}\textbf{Discrete:} & rec(X)=\{x_1,x_2,\ldots\} \quad p_k=P(X=x_k) \quad T_D=\sum_{k=1}^{\infty}p_k\delta(x_k)\\ \textbf{Continuous:} & rec(X)=R \quad p(x)\mathbf{1}_{A\subseteq R}(x)\end{cases}$$

CF:

$$\psi(t)=\int_{-\infty}^{\infty}e^{itx}p(x)dx \longrightarrow \psi(t)=\sum_{n=0}^{\infty}\frac{\mu_n}{n!}(it)^n$$

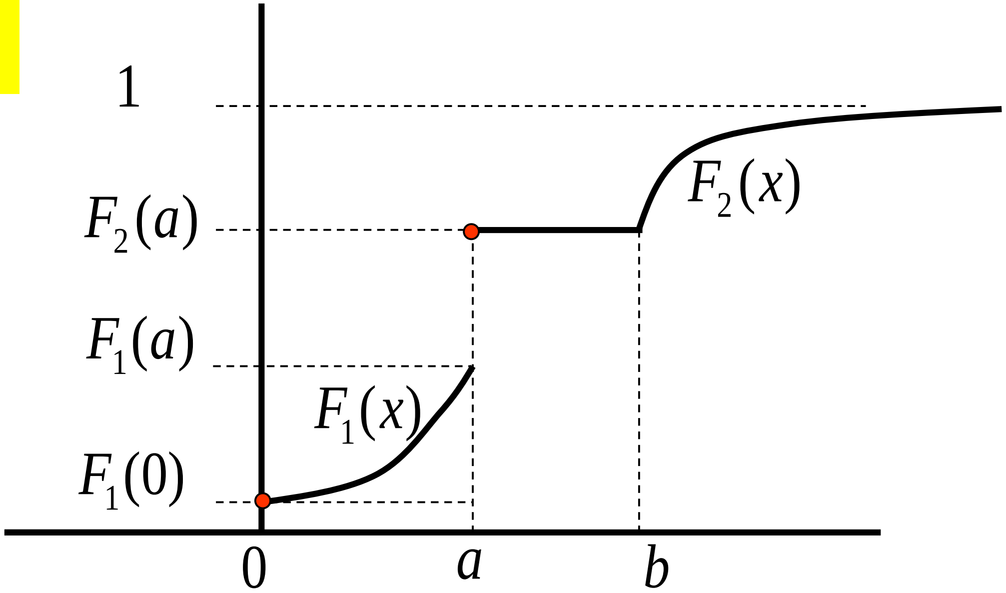
$$p(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\psi(t)dt=\sum_{n=0}^{\infty}\frac{\mu_n}{n!}\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}(it)^n dt$$

delta distribution $\delta(x,0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}dt \qquad T_p=\sum_{n=0}^{\infty}(-1)^n\frac{\mu_n}{n!}\delta^{n)}(x,0)$

$$<T_p,\phi>=\sum_{n=0}^{\infty}(-1)^n\frac{\mu_n}{n!}<\delta^{n)}(x,0),\phi>=\sum_{n=0}^{\infty}\frac{\mu_n}{n!}\phi^{n)}(0)$$

$$\boxed{T_{PROB}=\alpha T_D+(1-\alpha)T_p}$$

Example



$$F(x) = F_1(x)\left[H(x) - H(x-a)\right] + F_2(a)\left[H(x-a) - H(x-b)\right] + F_2(x)H(x-b)$$

$$< T_F', \phi > = < Df, \phi > = - < T_F, \phi' >$$

$$Df = F_1(0)\delta(0) + \left[F_2(a) - F_1(a)\right]\delta(a) + f_1(x)\mathbf{1}_{[0,a)}(x) + f_2(x)\mathbf{1}_{[b,\infty)}(x)$$

$$E\left[X^m\right] = < Df, X^m > = F_1(0)\delta_{m0} + \left[F_2(a) - F_1(a)\right]a^m + \int_0^a f_1(x)x^m dx + \int_b^\infty f_2(x)x^m dx$$

$$1 = F_1(0) + \left[F_2(a) - F_1(a)\right] + \int_0^a f_1(x)dx + \int_b^\infty f_2(x)dx$$

LIMIT THEOREMS and CONVERGENCE

General Problem:

Find the limit behaviour of a sequence of random quantities

Example:

$$\{X_1, X_2, \ldots, X_n, \ldots\} \longrightarrow \left\{Z_1 = X_1, Z_2 = \frac{X_1 + X_2}{2}, \ldots, Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \ldots\right\}$$

How is Z_n ***distributed when*** $n >> (\to \infty)$ ***?***

~"distance" $\longrightarrow$ ***convergence criteria***

1) More or less strong,

2) May have convergence for some criteria and not for others

Convergence in:

Distribution $\longrightarrow$ $\left\{\begin{array}{l}\textit{Central Limit Theorem}\\ \textit{Glivenko-Cantelli Theorem}\end{array}\right.$

Probability $\longrightarrow$ ***Weak Law of Large Numbers***

Almost Sure $\longrightarrow$ ***Strong Law of Lage Numbers***

$L_p(R)$ ***Norm*** $\longrightarrow$ ***Convergence in Quadratic Mean***

Uniform $\longrightarrow$ ***Glivenko-Cantelli Theorem***

Logarithmic

Chebyshev Theorem

$$X \sim F(x) \longrightarrow Y = g(X) \geq 0$$

$$P\left(g(X) \geq k\right) \leq \frac{E[g(X)]}{k}$$

$$P\left(Y \geq k\right)?$$

$$\Omega_X = \Omega_1 \bigcup \Omega_2 \qquad \Omega_1 = \left\{X \middle| g(X) < k\right\} \qquad \Omega_2 = \left\{X \middle| g(X) \geq k\right\}$$

$$E[Y] = \underbrace{\int_{\Omega_1} g(x)dF(x)}_{\boxed{g(X) \geq 0}} + \underbrace{\int_{\Omega_2} g(x)dF(x)}_{\boxed{g(X) \geq k}}$$

$$\geq 0 \qquad \geq \int_{\Omega_2} kdF(x) = kP\left(X \in \Omega_2\right) = kP(g(X) \geq k)$$

Bienaymé-Chebyshev Inequality

X ***with finite mean and variance*** $\left(\mu, \sigma^2\right)$

$$g(X) = (X - \mu)^2 \longrightarrow P\left(|X - \mu| \geq k\sigma\right) \leq \frac{1}{k^2}$$

Convergence in Probability

Let $\{X_1, X_2, \dots, X_n, \dots\}$ **and** $\{F_1(x_1), F_2(x_2), \dots, F_n(x_n), \dots\}$

Def.: X_n ***converges in probability to*** X ***if, and only if***

$$\lim_{n\to\infty} P\left(\left|X_n(x) - X\right| \geq \varepsilon\right) = 0 \quad ; \forall \varepsilon > 0$$

lim(Prob)

o, equivalently, $\lim_{n\to\infty} P\left(\left|X_n(x) - X\right| < \varepsilon\right) = 1 \quad ; \forall \varepsilon > 0$

Weak Law of Large Numbers (J. Bernouilli…)

Let $\{X_1, X_2, \dots, X_n, \dots\}$ ***be r.q. with the same distribution and finite mean*** (μ)

The sequence $\left\{Z_1 = X_1, Z_2 = \frac{X_1 + X_2}{2}, \dots, Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \dots\right\}$ ***converges in Probability to*** μ

$$\lim_{n\to\infty} P\left(\left|Z_n - \mu\right| \geq \varepsilon\right) = 0 \quad ; \forall \varepsilon > 0$$

LLN in practice:…

WLLN: When n is very large, the probability that Z_n differs from μ by a small amount is very small

→ Z_n gets more and more concentrated around the real number μ

But "very small " is not zero: it may happen that for some k>n, Z_k differs from μ by more than ε…

Convergence Almost Sure

Let $\{X_1, X_2, \ldots, X_n, \ldots\}$

Def.: X_n ***converges "almost sure" to*** X ***if, and only if***

$$P\left(\lim_{n\to\infty}\left|X_n(x)-X\right|\geq\varepsilon\right)=0 \quad ;\forall\varepsilon>0$$

Prob(lim)

o, equivalently, $P\left(\lim_{n\to\infty}\left|X_n(x)-X\right|<\varepsilon\right)=1 \quad ;\forall\varepsilon>0$

Strong Law of Large Numbers (E.Borel, A.N. Kolmogorov,…)

Let $\{X_1, X_2, \ldots, X_n, \ldots\}$ ***be r.q. with the same distribution and finite mean*** (μ)

The sequence $\left\{Z_1 = X_1, Z_2 = \frac{X_1+X_2}{2}, \ldots, Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \ldots\right\}$ ***converges Amost Sure to*** μ

$$P\left(\lim_{n\leftarrow\infty}\left|Z_n(x)-\mu\right|\geq\varepsilon\right)=0 \quad ;\forall\varepsilon>0$$

LLN in practice:…

WLLN: When n is very large, the probability that Z_n differs from μ by a small amount is very small

→ Z_n gets more and more concentrated around the real number μ

But "very small " is not zero: it may happen that for some k>n, Z_k differs from μ by more than ε…

SLLN: as n grows, the probability for this to happen tends to zero

Convergence in Distribution

Let $\{X_1, X_2, \ldots, X_n, \ldots\}$ **and their corresponding** $\{F_1(x_1), F_2(x_2), \ldots, F_n(x_n), \ldots\}$

Def.: X_n **tends to be distributed as** $X \sim F(x)$ **if, and only if**

$$\lim_{n\to\infty} F_n(x) = F(x) \equiv \lim_{n\to\infty} P(X_n \le x) = P(X \le x) \quad ; \forall x \in C(F)$$

o, equivalently, $$\lim_{n\to\infty} \phi_n(x) = \phi(x) \quad ; \forall t \in R$$

Central Limit Theorem (Lindberg-Levy,…)

Sequence of independent r.q. $\{X_1, X_2, \ldots, X_n, \ldots\}$ **same distribution**, **finite mean and variance** (μ, σ^2)

Form the sequence $$\left\{ Z_1 = X_1, Z_2 = \frac{X_1 + X_2}{2}, \ldots, Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \ldots \right\}$$

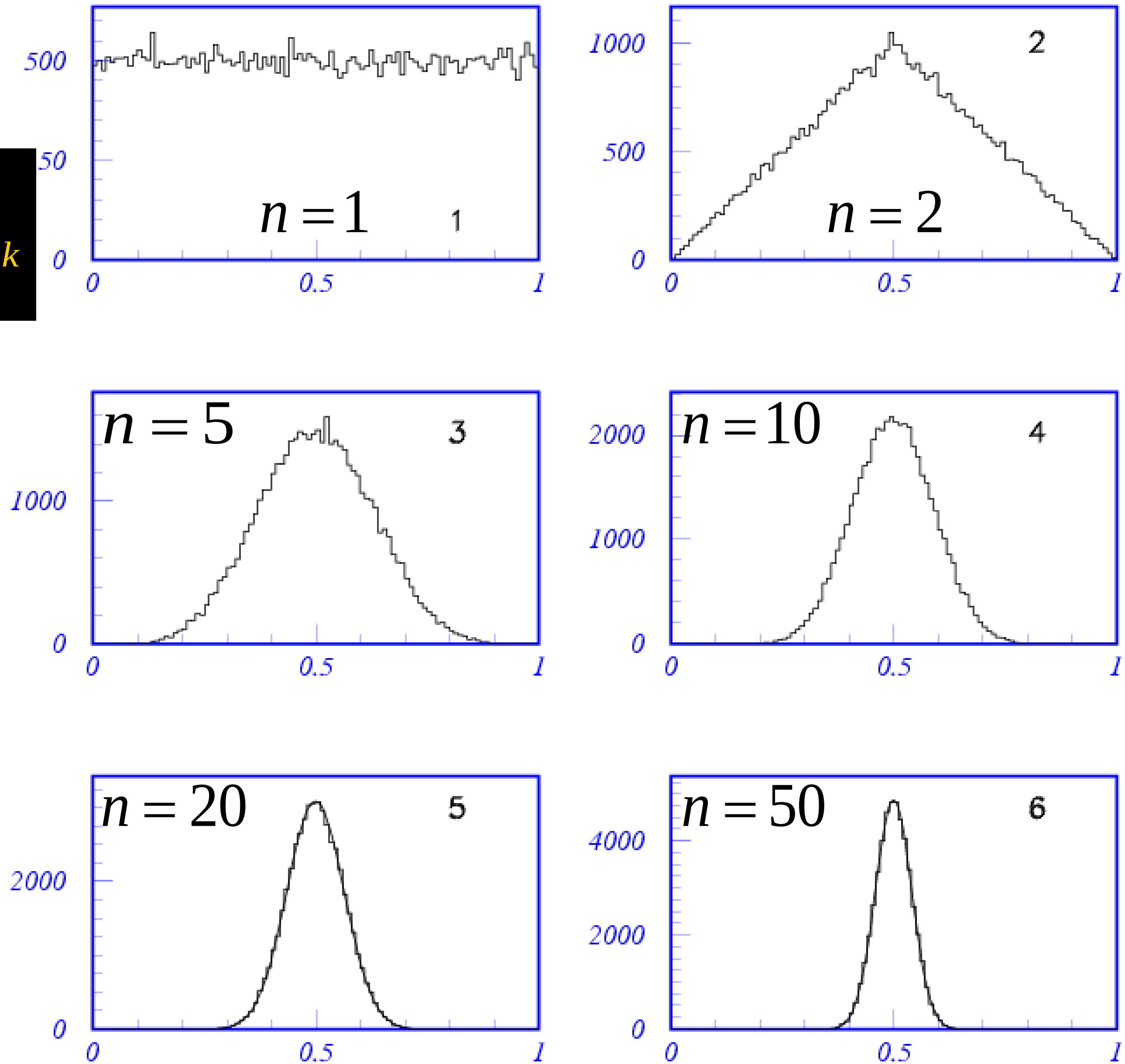
How is Z_n **distributed when** $n >> (\to \infty)$ **?** ***easy dem. from CF***

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k \sim N\left(z \,\middle|\, \mu, \sigma/\sqrt{n}\right)$$

$$\tilde{Z}_n = \frac{Z_n - \mu}{\sigma/\sqrt{n}} \sim N(z|0,1)$$ **standarized**

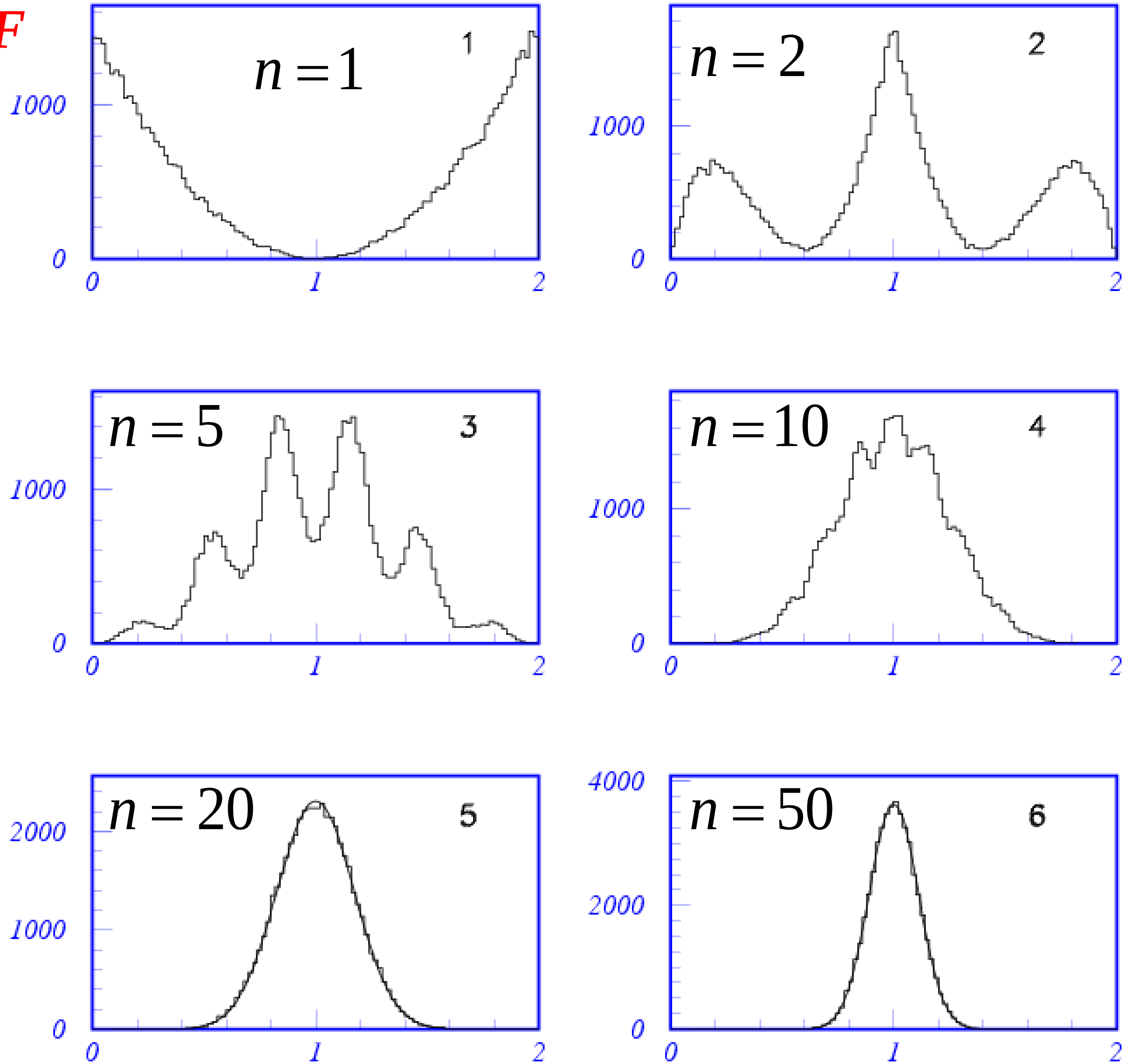
Unifom DF

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$



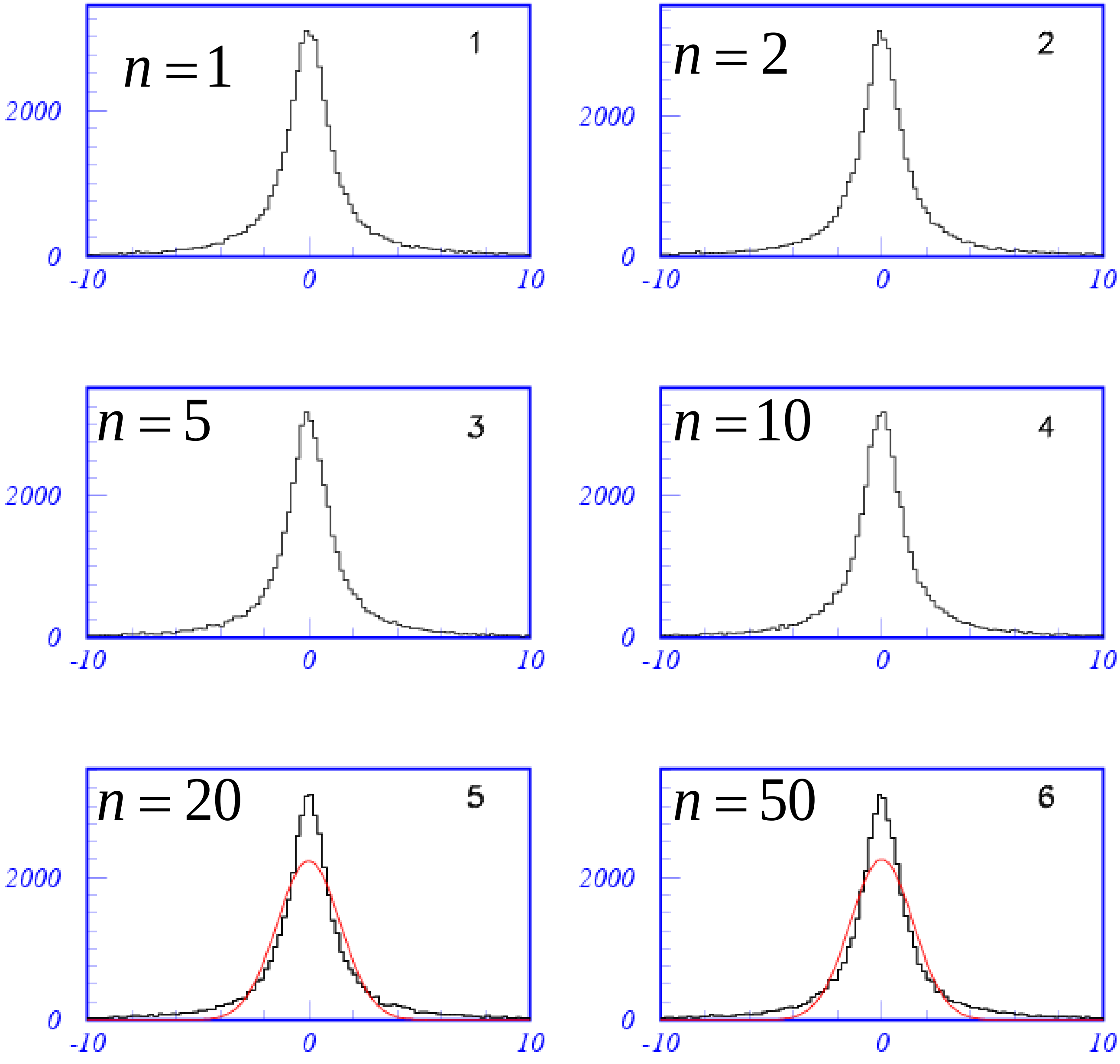
Parabolic DF

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$



Cauchy DF

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$



Uniform Convergence

$$f_n, f : S \to R$$

Def.: *The sequence* $\{f_n(x)\}_{n=1}^{\infty}$ *converges uniformly to* $f(x)$ *if, and only if*

$$\lim_{n\to\infty} \sup_{\substack{x \\ \forall x \in S}} \left| f_n(x) - f(x) \right| = 0$$

(example of aplication of Chebishev Th.)

Glivenko-Cantelli Theorem

experiment $e(1)$ *one observation of* X $\longrightarrow$ $\{x_1\}$

$e(n)$ *independent, identically distributed* $\longrightarrow$ $\{x_1, x_2, \ldots, x_n, \ldots\}$

Empiric Distribution Function

$$\boxed{F_n(x) = \frac{1}{n} \sum_{k=1}^{n} I_{(-\infty, x]}(x_k)}$$

$$\left(\frac{\text{number of values } x_i \text{ lower or equal to } x}{n} \right)$$

If observations are iid: $\lim_{n\to\infty} P\left(\lim_{n\to\infty} \sup_x \left| F_n(x) - F(x) \right| = 0\right) = 1$

The Empiric Distribution Function converges uniformly to the Distribution Function $F(x)$ *of the r.q. X*

Demonstration of Convergence in Probability:

Empiric Distribution Function $$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} I_{(-\infty,x]}(x_k)$$

1) $A = (-\infty, x_0]$

$Y = I_A(x)$ ***is a Bernouilli random quantity***

$$P(Y=1) = P(X \in A) = F(x_0)$$

$$P(Y=0) = P(X \notin A) = 1 - F(x_0)$$

with Characteristic Function $$\phi_Y(t) = e^{it}F(x_0) + \left(1 - F(x_0)\right)$$

2) Sum of iid Bernouilli random quantities $$Z_n = \sum_{k=1}^{n} I_{(-\infty,x]}(x_k) = nF_n(x)$$

$W = Z_n = nF_n(x)$ ***is a Binomial r.q.*** $$P(W=k) = \binom{n}{k} F^k (1-F)^{n-k}$$

$$E[W] = nF \xrightarrow{F_n(x) = \frac{W}{n}} E[F_n] = F$$

$$V[W] = nF(1-F) \longrightarrow V[F_n] = \frac{F(1-F)}{n}$$

3) Bienaymé-Chebyshev Inequality

$$P\left(|X - \mu| \geq k\sigma\right) \leq \frac{1}{k^2} \longrightarrow \boxed{P\left[|F_n(x) - F(x)| \geq \varepsilon\right] \leq \frac{F(1-F)}{n\varepsilon^2}}$$

Convergence in $L_p(R)$ Norm

Def.: $\{X_n(w)\}_{n=1}^{\infty}$ **Converges in** $L_p(R)$ **norm to** $X(w)$

if, and only if,

$$X_n(w) \in L_p(R) \quad ;\forall n \qquad X(w) \in L_p(R) \qquad and \qquad \lim_{n\to\infty} E\left[\left|X_n(w) - X(w)\right|^p\right] = 0$$

$$\left(\textbf{that is:} \qquad \forall \varepsilon > 0 \quad \exists n_0(\varepsilon) \quad | \quad \forall n \geq n_0(\varepsilon) \quad E\left[\left|X_n(w) - X(w)\right|^p\right] < \varepsilon\right)$$

p=2 convergence in quadratic mean

Logarithmic Convergence

Kullback-Leibler "Discrepancy" ***(see Lect. on Information)***

Logarithmic Divergence of a pdf $\tilde{p}(\boldsymbol{x}); \quad \boldsymbol{x} \in X$

from its true pdf $p(\boldsymbol{x})$

$$D_{KL}[\tilde{p} \mid p] \equiv \int_X p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{\tilde{p}(\boldsymbol{x})} d\boldsymbol{x}$$

A sequence $\{p_i(\boldsymbol{x})\}_{i=1}^{\infty}$ ***of pdf "Converges Logarithmically" to a density*** $p(\boldsymbol{x})$ ***iff***

$$\lim_{k\to\infty} D_{KL}[p \mid p_k] = 0$$