PROBABILITY AND STATISTICS

" They say that understanding ought to work by the rules of right reason. These rules are, or ought to be, contained in Logic; but the actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on.

Therefore, the true logic of this world is the calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind"

J. Clerk Maxwell

Lecture 1: Elements of Probability

Lecture 2: Monte Carlo simulation

Lecture 3: Bayesian Inference

These lectures are just a "guided tour" to the Lecture Notes at: **arXiv:1610.05590v3**

INDEX of Lecture 1:

1) ***Elements of Probability, random quantities, probability densities, …***
2) ***Distribution Function***
3) ***Conditional Probability and Bayes Theorem***
4) ***Stochastic Characteristics (mean, variance, moments,…)***
5) ***Integral Transforms (Fourier, Mellin)***
6) ***Convergence (Laws of Large Numbers, Central Limit Theorem,…)***

1) THE ELEMENTS OF PROBABILTY

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

A. N. Kolmogorov (1933)+…

1) Events and 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 $\longrightarrow$ $\Omega = \bigcup_{\forall k} e_k \qquad e_k \bigcap e_j = \emptyset \quad ;\forall k, j \quad k \neq j$

Types of Events

- *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*

EXAMPLE: $Z \to f\,\bar{f}$

elementary events:

$$e_1 = \{Z \to e^+e^-\}; \quad \dots \quad e_6 = \{Z \to \nu_\tau \bar{\nu}_\tau\}; \quad \dots$$

$$\dots \quad e_7 = \{Z \to u\bar{u}\}; \quad \dots \quad e_{11} = \{Z \to b\bar{b}\}$$

$$\Omega = \{e_1, e_2, \dots, e_{10}, e_{11}\}$$

Sure event: $S = \{Z \to \textit{fermions}\}$

Impossible event: $I = \{Z \to e^- \mu^-\}$

Non-elementary events: $A = \{Z \to \textit{leptons}\} = \bigcup_{i=1}^{6} e_i$ $\quad A^c = \{Z \to \textit{hadrons}\} = \bigcup_{i=7}^{11} e_i$

$$B_1 = \{Z \to \textbf{charged} \quad \textit{leptons}\} = \bigcup_{i=1}^{3} e_{2i-1} \quad \dots \quad B_2 = \{Z \to \textbf{neutral} \quad \textit{leptons}\} = \bigcup_{i=1}^{3} e_{2i}\ ; \quad \dots$$

(Ω) ***… but we are interested in many events (questions) other than the elementary ones…***

$B_1 = \{Z \to \textbf{charged} \quad \textit{leptons}\}$ **?** $\qquad B_2 = \{Z \to \textbf{neutral} \quad \textit{leptons}\}$ **?**

$A = \{Z \to \textit{leptons}\} = B_1 \bigcup B_2$ **?** ***… if not what occurs is*** $A^c = \{Z \to \textit{hadrons}\}$

They are all sets …

and we are about to see that Probability is a measure on sets so we have to single out the sets we are interested in (… sets that we want to "measure") so…

2) Measurable Space (Ω, B_{Ω})

$\oplus$ Ω: Sample Space

B_{Ω} : Algebra ⟶ Class of events closed under union and complements

Why algebra B_{Ω} ?

We are interested in a class of events that:

1) Contains all possible results of the experiment we are interested in

2) Is closed under union and complementation

$$\forall A_1, A_2 \in B_{\Omega} \quad \rightarrow \quad A_1 \bigcup A_2 \in B_{\Omega} \quad ; \quad A_1^{\,c} \in B_{\Omega}$$

➡ $\Omega \in B_{\Omega};\quad \emptyset \in B_{\Omega};\quad A_1 \bigcap A_2 \in B_{\Omega};\quad A_1^{\,c} \bigcup A_2^{\,c} \in B_{\Omega};\quad A_1^{\,c} \bigcap A_2^{\,c} \in B_{\Omega};\ldots$

Morgan's laws : $\left(A_1 \bigcup A_2\right)^c = A_1^{\,c} \bigcap A_2^{\,c} \;;\ldots$

So now we have:

1) Ω has all the elementary events

2) B_{Ω} has all the events we are interested in

We can construct several possible algebras:

EXAMPLE: $Z \to f\,\bar{f}$

$$\Omega = \{e_1, e_2, \ldots, e_{10}, e_{11}\}$$

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

Interest in decay type:

$$\begin{cases} A = \{Z \to leptons\} = \bigcup_{i=1}^{6} e_i \\ A^c = \{Z \to hadrons\} = \bigcup_{i=7}^{11} e_i \end{cases}$$

$$B = \{\Omega, \emptyset, A, A^c\}$$

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

(power set $\mathscr{P}(\Omega)$***)***

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

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

More General Structures of the algebra…

Structure of algebra B_Ω (1)

Dimension of Sample Space

$$\dim(\Omega) = \begin{cases} \textit{Finite} \\ \textit{drawing a die} \\ \textit{Denumerable} \\ \textit{throw a coin and stop when we get head} \\ \textit{Non-denumerable} \\ \textit{decay time of a particle} \end{cases}$$

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

$\Omega = \{h, th, tth, ttth, \dots\}$

$\Omega = \{t \in R^+\}$

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

Structure of algebra B_Ω *(2)*

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

All σ-algebras are Boole algebras but not all Boole algebras are σ-algebras

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

As we shall see, we are interested in R^n

so… What about $\mathscr{P}$***(R)*** $(2^{\aleph_1})$ ***?***

Certainly is an algebra but …

Which are the "basic" events to construct a useful algebra (for us)?

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 collection of intervals, denumerable or not, is a subset of*** R

Nice sets but… a collection of intervals is a σ-algebra ?

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

…but…

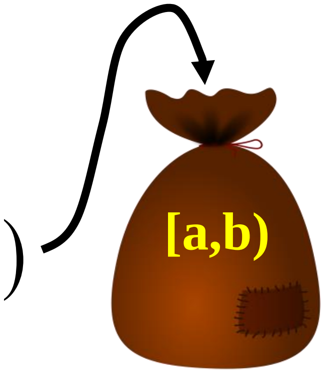
not intervals (in general)

I^c $\bigcup$

Generate a σ-algebra

for instance, from… half open intervals on the right

1) Initial Set (Ω_0): ***Contains all half-open intervals on the right*** $[a,b)$

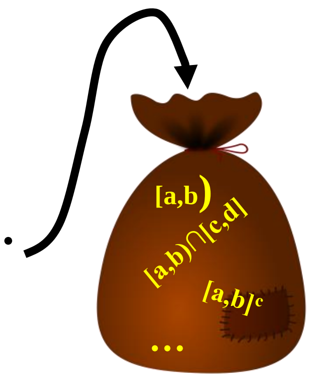


2) Form the set Ω_1 by adding their countable unions and complements

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

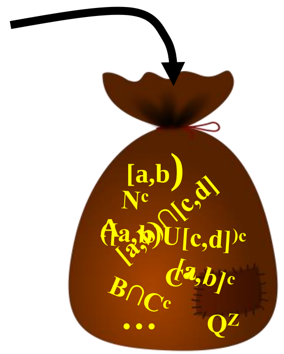
$$[a,b]=\bigcap_{n=1}^{\infty}\left[a,b+\tfrac{1}{n}\right) \qquad \{a\}=[a,a] \quad \ldots$$

It has all intervals and points … among other elements



3) Add sets to close under countable union and complementation

There is at least one σ-algebra containing Ω_1



Borel σ-algebra (B_R) : Smallest σ-algebra of subsets of R that contains intervals ([a,b),…)

$$N,Z,Q\subset B_R$$

May start as well with $(a,b],(a,b),[a,b]$

Its elements are Borel sets (borelians)

$$Next:\quad (\Omega,B_\Omega),\oplus\mu\Rightarrow$$

3) Measure Space (Ω, B_Ω, μ)

Measure:

i) Set function $\mu : A \in B_\Omega \to R$ ***(one and only one real number)***

ii) σ-additive ***For any countable sequence of disjoint sets of*** B_Ω

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

"signed measure" on σ-algebra B_Ω

iii) Non-negative $\mu : A \in B_\Omega \to \mu(A) \in [0, \infty)$ ***… measure***

►***Measure Space*** (Ω, B_Ω, μ)

Two important measures →

Lebesgue Measure in R^n

(R^n, B, λ)

Translation invariant $\oplus$ $\lambda([0,1]^n) = 1$

σ-additive

$\lambda([x_1, x_2]) = x_2 - x_1$

$= \lambda([x_1, x_2)) = \ldots$

- ***All Borel Sets of R^n are Lebesgue Measurable***

- ***There are non-denumerable subsets of R with zero Lebesgue measure***
- ***Not all subsets of R are Lebesgue measurable*** ***(←Axiom of Choice)***
- ***Lebesgue measurable sets (C) not in B*** ***($\lambda(A \epsilon B_R \backslash C)=0$)***

Probability Measure

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

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

(notation $\mu \to P$)

$(\Omega, B_\Omega) \oplus \mu \Rightarrow$ ▶ ***Probability Space*** (Ω, B_Ω, P)

- ***Any bounded measure can be converted in a probability measure***

From ii) (σ-additivity for disjoint sets): $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$

Properties of measure… *eventually properties of probability*

$\forall\ A, B \in B_\Omega :$

$$\boxed{\mu(A^c) = \mu(A \cup A^c) - \mu(A)}$$

$$\boxed{\mu(\emptyset) = 0}$$

$A\,;\, A^c \in B_\Omega$ disjoint sets

$\mu(A \cup A^c) = \mu(A) + \mu(A^c) \rightarrow \mu(A^c) = \mu(A \cup A^c) - \mu(A)$

$\mu(\emptyset) = \mu(A \cap A^c) = \mu(A) + \mu(A^c) - \mu(A \cup A^c) = 0$

$$\boxed{\Omega = A \bigcup A^c \rightarrow P(A^c) = P(\Omega) - P(A) = 1 - P(A), \ldots \quad P(\emptyset) = P(\Omega^c) = 0, \ldots}$$

$$\boxed{\mu(B \setminus A) = \mu(B) - \mu(A) \geq 0}$$

$$\boxed{P(B \setminus A) = P(B) - P(A)}$$

if $A \subseteq B \rightarrow B = A \cup \{B \setminus A\}$ both disjoint

$\mu(B) = \mu(A) + \mu(B \setminus A) \rightarrow \mu(B / A) = \mu(B) - \mu(A) \geq 0$

$(B \setminus A \doteq B \bigcap A^c)$

$$\boxed{\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)}$$

$$\boxed{P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)}$$

$A_1 = A \cap B\,;\, A_2 = A \setminus A_1\,;\, A_3 = B \setminus A_1$ all A_i disjoint sets

$A \cup B = A_1 \cup A_2 \cup A_3 \rightarrow \mu(A \cup B) = \mu(A_1) + \mu(A_2) + \mu(A_3)$

$$\left.\begin{array}{lll} A = A_1 \cup A_2 & \rightarrow & \mu(A) = \mu(A_1) + \mu(A_2) \\ B = A_1 \cup A_3 & \rightarrow & \mu(B) = \mu(A_1) + \mu(A_3) \end{array}\right\}$$

$\rightarrow \mu(A_1) + \mu(A_2) + \mu(A_3) = \mu(A) + \mu(B) - \mu(A_1)$

• • •

EXAMPLE: $Z \to f\ \bar{f}$

$A = \{Z \to leptons\}$ $A^c = \{Z \to hadrons\}$ $B = \{\Omega, \emptyset, A, A^c\}$

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

$\mu(A) = 0.3$

$\mu(\Omega) = 1$ ***(certainty)*** $\mu(A^c) = 1 - \mu(A) = 0.7$

$\mu(\emptyset) = 0$

Now we have the Probability Space (Ω, B_Ω, P) ***…***

… but the results of the experiment are not necessarily numeric, expectations,… →

Random quantities ("variables")

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

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

(misfortunately called "random variable")

Induced Space

$$(\Omega, B) \xrightarrow{X} (\Omega_I, B_I) \longrightarrow (\Omega, B, P) \to (\Omega_I, B_I, P_I)$$

Interest in: (R, B_R) $P_I = P(X = k)$ or $P(X \in (a,b))$

X(w): Is neither random nor variable

What is random is the outcome of the experiment before it is done

Random quantities (2)

$(\Omega, B) \xrightarrow{X} (R, B_R)$ **But B has the events of interest so:**

To keep the structure of the σ-algebra B it is necessary that

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

(i.e. the function X(w) be Lebesgue (…Borel) measurable)

$f(w): w \in \Omega \to \Delta$ is Borel measurable… **wrt the σ-algebra associated to Ω**

EXAMPLE: $Z \to f\,\bar{f}$

$A = \{Z \to leptons\} = \bigcup_{i=1}^{6} e_i \qquad A^c = \{Z \to hadrons\} = \bigcup_{i=7}^{11} e_i \qquad B = \{\Omega, \emptyset, A, A^c\}$

$\Omega = \{e_1, e_2, \ldots, e_{10}, e_{11}\} \qquad X(w): w \in \Omega \to X(w) \in R$

Is the function $X(w)$ an admissible random quantity? $\qquad \forall a \in R \quad X^{-1}(-\infty, a] \in B$ (with $(-\infty, a] \in B_R$) **???**

1) $X(e_1) = X(e_3) = X(e_5) = X(e_7) = X(e_9) = X(e_{11}) = 1$

$X(e_2) = X(e_4) = X(e_6) = X(e_8) = X(e_{10}) = -1$

In this case is simpler: $B_X = \{\{-1,1\}, \emptyset, \{1\}, \{-1\}\}$

so check that $X^{-1}(\{-1\}) \in B$ **and** $X^{-1}(\{1\}) \in B$

$$X^{-1}(\{-1\}) = \left\{\bigcup_{k=1}^{5} e_{2k}\right\} \notin B$$

2) $X(e_1) = \ldots = X(e_6) = 1 \qquad X(e_7) = \ldots = X(e_{11}) = -1$

$a < -1 \to \emptyset \in B \qquad -1 \le a < 1 \to \bigcup_{i=7}^{11} e_i = A^c \in B \qquad 1 \le a \to \Omega \in B$

Indicator function:

$$A \subseteq \Omega; \quad \forall x \in \Omega \quad \rightarrow \quad \mathbf{1}_A(x) \doteq \begin{cases} 1 & if \;\; x \in A \\ 0 & if \;\; x \notin A \end{cases}$$

$$A, B \subseteq \Omega: \quad \mathbf{1}_{A \cap B}(x) = \min\{\mathbf{1}_A(x), \mathbf{1}_B(x)\} = \mathbf{1}_A(x)\mathbf{1}_B(x)$$

$$\mathbf{1}_{A \cup B}(x) = \max\{\mathbf{1}_A(x), \mathbf{1}_B(x)\} = \mathbf{1}_A(x) + \mathbf{1}_B(x) - \mathbf{1}_A(x)\mathbf{1}_B(x)$$

$$\mathbf{1}_{A^c}(x) = 1 - \mathbf{1}_A(x)$$

Types of Random Quantities

Codomain of $X(w): w \in \Omega \to X(w) \in R$ $\begin{cases} \text{Finite / countable set} & \text{Discrete r.q.} \\ \text{Uncountable set} & \text{Continuous r.q.} \end{cases}$

$\{A_k\}_{k\in N}$ **Partition of** $\Omega = \bigcup_k A_k$ **For all elements of Ω that belong to** A_k**,** $X(w)$ **assigns the same value** x_k

Discrete random quantity

▶ $\{A_k\}_{k=1}^{n}$ **finite partition of** $\Omega = \bigcup_{k=1}^{n} A_k$

$$\longrightarrow \quad X(\omega) = \sum_{k=1}^{n} x_k \mathbf{1}_{A_k}(\omega)$$

simple function with codomain

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

simple random quantity

▶ $\{A_k\}_{k=1}^{\infty}$ **countable partition of** $\Omega = \bigcup_{k=1}^{\infty} A_k$

$$\longrightarrow \quad X(\omega) = \sum_{k=1}^{\infty} x_k \mathbf{1}_{A_k}(\omega)$$

elementary function with codomain

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

elementary random quantity

Either case $X(w)$ **takes values on** $\Omega_X = \{x_1, x_2, \ldots\}$ **finite or denumerable set**

with probabilities $\{p_1, p_2, \ldots\}$ $\longrightarrow$ $P(X = x_i) = p_i$ $\begin{cases} \text{real} \\ \text{non-negative} \\ \sum_{\forall k} p_k = 1 \end{cases}$

Continuous random quantity

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

$\Omega_X \subseteq R$ **uncountable set** $\longrightarrow$ $A \subseteq \Omega_X \to P(X \in A) = \int_A dP(x) = \int_R \mathbf{1}_A(x)\,dP(x)$

▶ $X(\omega)$ absolute continuous:

Radon-Nikodym Theorem *(1913, 1930)*

If conditions (*; see notes) satisfied:

$\exists\ p(x)$

- ***unique*** *(if* $g(x)$ *has same properties as* $p(x)$ $\longrightarrow$ $\mu\{x \mid p(x) \neq g(x)\} = 0$ *)*
- ***λ-integrable*** *(in fact Riemann integrable)*
- ***non-negative a.e.*** $(p(x) \geq 0 \quad ae)$
- ***bounded on any bounded interval of R*** $(\Leftarrow$ ***if*** $\exists\ p(w)$ ***then*** $P << \lambda)$

such that $\forall A \in B$

$$P(X \in A) = \int_R \mathbf{1}_A(x)\,dP(x) = \int_A dP(x) = \int_A p(x)\,dx$$

Radon density… Probability Density Function $p(x \mid \theta) = \ldots \times \mathbf{1}_{\Omega_X}(x)$ $\longrightarrow$ $\int_{-\infty}^{\infty} p(x)\,dx = 1$

▶ $X(\omega)$ singular

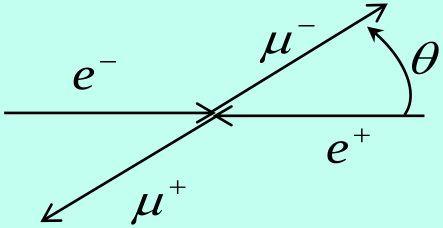
(*): P ***σ-finite measure over the measurable space*** (R, B)

If $P \sim \lambda$ ***(equivalent :*** $\nu << \lambda$ ***and*** $\lambda << \nu$ ***)***

$\nu << \mu$ ***absolute continuous:*** $\mu(A) = 0 \Rightarrow \nu(A) = 0 \quad \forall A \in B$

EXAMPLE: *Random Quantity: Continuous, discrete…*

depends on the codomain of $X(w): w \in \Omega \to X(w) \in R$



$Z = \cos\Theta \qquad Z \sim p(z \mid a) = \frac{3}{8}(1+z^2) + a\,z$

$\Omega_Z = [-1,1] = [-1,0) \cup [0,1] = \Omega_1 \cup \Omega_2 \qquad \Omega_1 \cap \Omega_2 = \emptyset$

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

$X(\Omega_1) = -1 \quad (to\ \ all\ \ w\epsilon\Omega_1\ )$

$X(\Omega_2) = 1 \quad (to\ \ all\ \ w\epsilon\Omega_2\ )$

$P(X(w)=1) = \int_0^1 p(w \mid a)dw = \frac{1}{2}(1+a)$

$P(X(w)=-1) = 1 - P(X(w)=1) = \frac{1}{2}(1-a) \qquad \left(\int_{-1}^{0} p(w \mid a)dw\right)$

Last, remember that:

(see notes for demonstrations)

- ***The set of points of R with finite probabilities*** $W = \{\forall x \in R \mid P(x) > 0\}$ ***is countable***

- $\sum_{\forall i} P(x_i) = 1$ ⟶ ***If Ω is ∞ or denumerable, it is not possible for all the points to have the same probability***

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

P(impossible event)=0 but P(event)=0 ⟶̸ ***event is impossible***

2) THE DISTRIBUTION FUNCTION

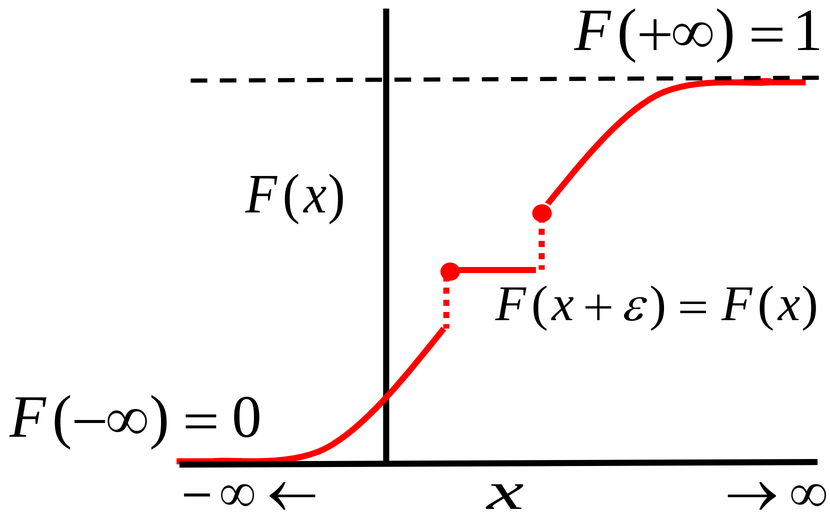
Distribution Function

Def. (gen.): **One-dimensional DF** $\forall F : x \in \Omega_X \subset R \to R$ **such that:**

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

2) Monotonically non-decreasing: $\left.\begin{matrix} if \quad x_1, x_2 \in R \\ and \quad x_1 \leq x_2 \end{matrix}\right\} \to F(x_1) \leq F(x_2)$

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



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

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

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

$$F(x) \equiv P(X \leq x) = P\left(X \in \left(-\infty, x\right]\right) \quad ; \quad \forall x \in R$$

- ***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*** $\left(x_1, x_2\right] \in R$

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

Distribution Function ⟷ ***Probability Measure***

- ***The Distribution Function of a Random Quantity has all the information needed to describe the properties of the random process for a given model.***

Some General Properties of the DF

From definition

(see notes for demonstrations)

$\forall x \in R$

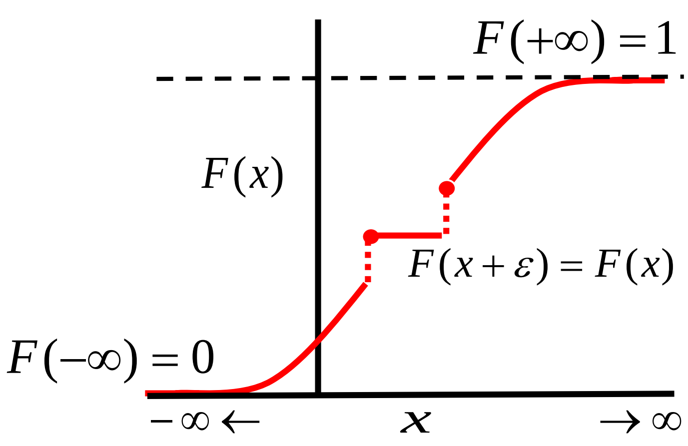
$$F(x) \doteq P(X \le x) = P\big(X \in (-\infty, x]\big) \qquad P(X < x) = F(x-\varepsilon)$$

$$P(X > x) = 1 - P\big(X \le x\big) = 1 - F(x) \qquad P(x_1 < X \le x_2) = P\big(X \in (x_1, x_2]\big) = F(x_2) - F(x_1) \quad \dots$$

DF defined $\forall x \in R$

If X **takes values in** $[a,b] \in R$ $\qquad p(x\,|\,\theta) = \ldots \times \mathbf{1}_{[a,b]}(x)$ **and** $F(x) = \begin{cases} 0 & \forall x < a \\ 1 & \forall x \ge b \end{cases}$

Set of points of discontinuity of the DF $D = \{\forall x \in R \,/\, F(x-\varepsilon) \ne F(x+\varepsilon)\}$ **is finite or countable**



At each point of discontinuity, $F(x)$ **has a jump of amplitude** $P(X = x)$

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

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

Random Quantity ⟶ discrete / continuous (singular or absolutely continuous) ⟶ **Distribution Function**

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

Discrete Random Quantity $X(w)$

Codomain $\Omega_X \subseteq R$ **is finite or countable set**

$X(w)$ **takes values** $\Omega_X = \{x_1, x_2, \ldots\}$

with probabilities $\{p_1, p_2, \ldots\}$ $\longrightarrow$ $P(X = x_i) = p_i$ — **real**, **non-negative**, $\sum_{\forall k} p_k = 1$

Distribution Function:

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

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

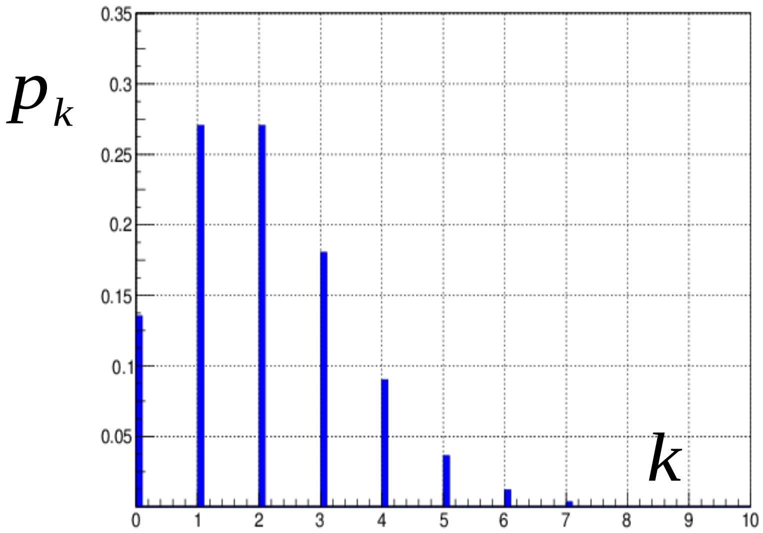
1) Step-wise and monotonous non-decreasing

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

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

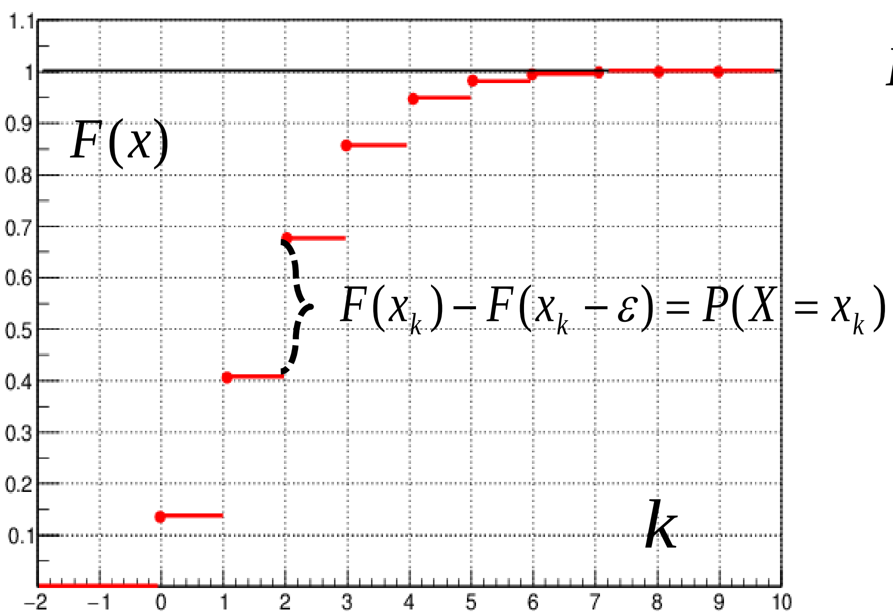
EXAMPLE: **Poisson Distribution: *Po(x|μ)***

$$\Omega_X = \{0,1,2,\ldots\}$$



$$p_k = P(X = k) = e^{-\mu}\frac{\mu^k}{\Gamma(k+1)}$$

$$\sum_{i=1}^{\infty} p_i = 1$$



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

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

Continuous Random Quantity $X(w)$

Codomain $\Omega_X \subseteq R$ is a non-denumerable set

$F(x)$ — ***Continuous on the right:*** $F(x+\varepsilon)=F(x)$

Jump of amplitude P(x) at discontinuity: $F(x-\varepsilon)=F(x)-\underbrace{P(X=x)}_{=0}=F(x)$

***continuous everywhere* in R**

AC:

(… Radon-Nikodym: $P(A)=\int_A dP=\int_A \frac{dP}{d\lambda}d\lambda=\int_A p(t)dt$

Distribution Function: $F(x)=P(X\le x)=\int_{-\infty}^{x} p(t)dt \qquad F(-\infty)=0 \ ; \ F(+\infty)=1$

Probability Density Function (pdf): $p(x)=\frac{dF(x)}{dx}$ ***unique a.e.***

1) $p(x)\ge 0 \ ; \ a.e. \ in \ R$

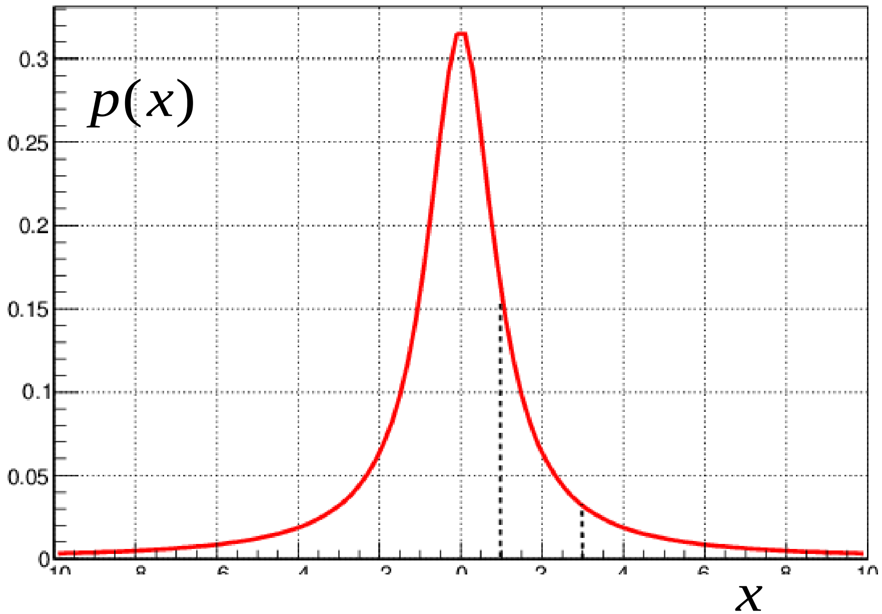
2) *bounded in every bounded interval of* R *and Riemann integrable on it*

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

$p(x|\theta)=\ldots\times \mathbf{1}_{\Omega_X}(x)$

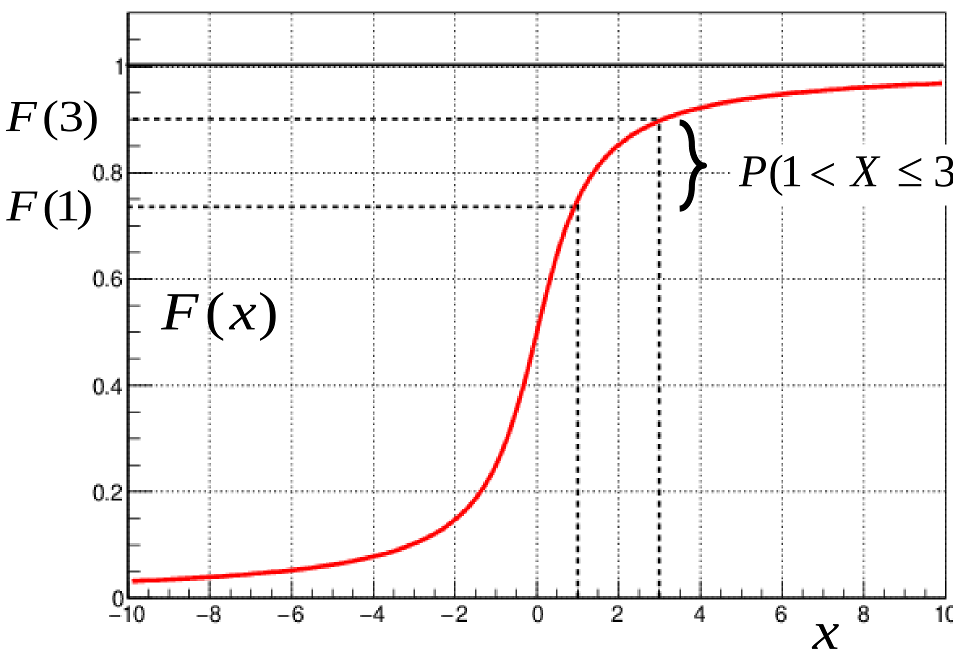
EXAMPLE:

Cauchy Distribution: $Ca(x|0,1)$



$$p(x) = \frac{1}{\pi}\frac{1}{1+x^2}\mathbf{1}_{(-\infty,\infty)}(x)$$

$$F(x) = \int_{-\infty}^{x} p(u)du = \frac{1}{2} + \frac{1}{\pi}\arctan(x)$$



$$P(1 < X \leq 3) = F(3) - F(1)$$

General Distribution Function (Lebesgue Decomposition)

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

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, Cantor,…)

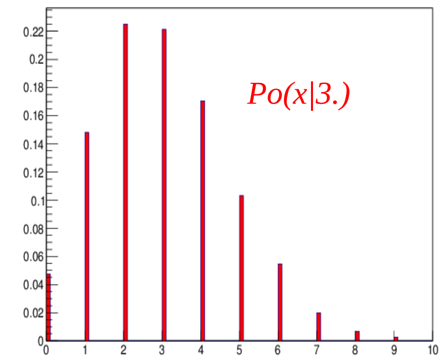
Some Distributions that we shall use frequently:

Discrete Distributions

$$Po(x \mid \mu) = e^{-\mu} \frac{\mu^x}{\Gamma(x+1)} \mathbf{1}_{N_0}(x)$$

Poisson

$$\mu \in R_{>0}$$



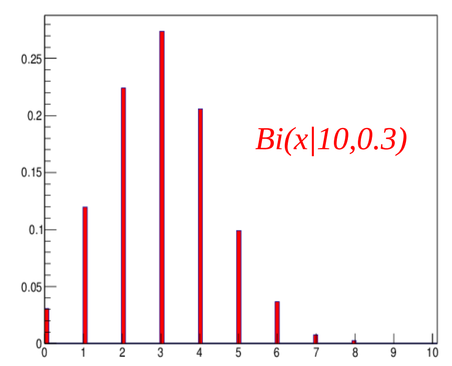
$$Bi(x \mid n, \theta) = \binom{n}{x} \theta^x \, (1-\theta)^{n-x} \, \mathbf{1}_{N_0 \cap [0,n]}(x)$$

Binomial

Bernoulli

$$\theta \in (0,1) \subset R \quad n \in N_{>0}$$

$$Brn(x \mid \theta) = Bi(x \mid 1, \theta)$$



$$+ \quad Mn(\mathbf{x} \mid n, \boldsymbol{\theta}) = \frac{n!}{x_1! \cdots x_k!} (1 - \sum_{j=1}^{k-1} \theta_j)^{(1-\sum_{j=1}^{k-1} x_j)} \prod_{i=1}^{k-1} \theta_i^{x_i} \quad + \ldots$$

Absolute Continuous

$$Ga(x \mid a,b) = C(a,b)\, e^{-ax} x^{b-1}\, \mathbf{1}_{(0,\infty)}(x) \qquad C(a,b) = \frac{a^b}{\Gamma(b)}$$

$$a,b \in R_{>0} \qquad \chi^2(x \mid \nu) = Ga(x \mid 1/2, \nu/2)$$

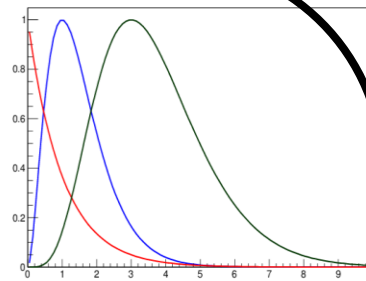
$$Ex(x \mid a) = Ga(x \mid a,1)$$

Gamma

Chi2

Exponential

Ga(x|2,3)
Ga(x|1,1)
Ga(x|1.5,5.5)



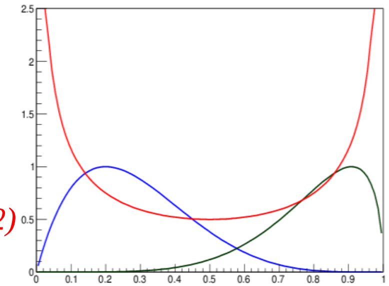
$$Be(x \mid a,b) = C(a,b)\, x^{a-1}(1-x)^{b-1}\, \mathbf{1}_{(0,1)}(x) \qquad C(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

$$a,b \in R_{>0} \qquad Un(x \mid 0,1) = Be(x \mid 1,1)$$

Beta

Uniform

Be(x|2,5)
Be(x|1/2,1/2)
Be(x|6,3/2)



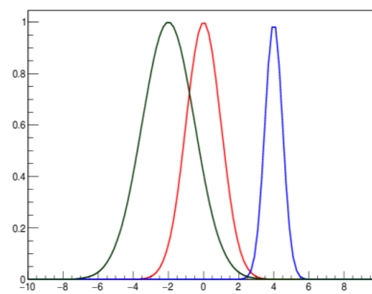
$$N(x \mid \mu,\sigma^2) = C(\sigma)\, \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \mathbf{1}_{(-\infty,\infty)}(x) \qquad C(\sigma) = \left(2\pi\sigma^2\right)^{-1/2}$$

$$\mu \in R$$

$$\sigma \in R_{>0}$$

Normal

N(x|4,1/2)
N(x|0,1)
N(x|-2,3/2



$$St(x \mid \mu,\lambda,\nu) = C(\lambda,\nu)\left(1+\lambda\nu^{-1}(x-\mu)^2\right)^{-(\nu+1)/2} \mathbf{1}_{(-\infty,\infty)}(x) \qquad C(\lambda,\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\left(\frac{\lambda}{\pi\nu}\right)^{1/2}$$

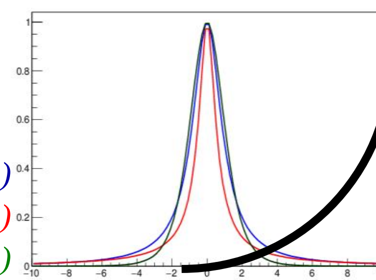
$$\mu \in R \qquad Ca(x \mid \mu,\lambda) = St(x \mid \mu,\lambda,1)$$

$$\lambda,\nu \in R_{>0}$$

Student

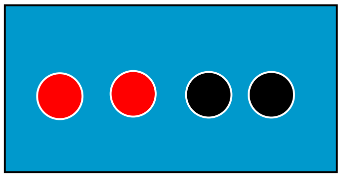
Cauchy

St(x|0,1,1)
St(x|0,2,5)
St(x|0,1,5)



$\ldots +$ *Multivariate Normal, Pareto, Dirichlet, …*

3) CONDITIONAL PROBABILITY and BAYES THEOREM



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

Given a probability space (Ω, B, P)

- *The probability assigned to an event* $A \in B$ *(degree of credibility we have on the occurrence of…) depends on the information we have*

→ *All probabilities are conditional*

Conditional Probability

Statistical Independence

Consider (Ω, B_Ω, P)

and two ***not disjoint*** *sets*

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

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

Probability to happen A and B ***A and not B***

$$= P(A, B) + P(A, B^c)$$

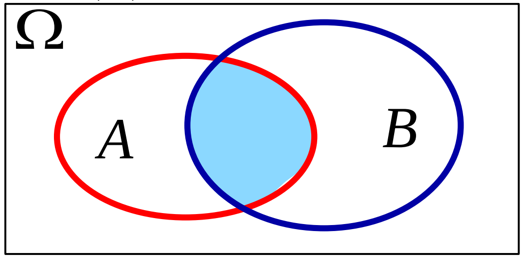
Notation: $P(A \bigcap B \bigcap C \bigcap \ldots) \equiv P(A, B, C, \ldots)$

What is the probability for A to happen

if we know that B has already occurred?

$$\doteq P(A|B)$$

P(A) restricted to B



$$\left.\begin{aligned} P(A|B) &= C \times P(A \bigcap B) \\ P(B|B) &= 1 = C \times P(B \bigcap B) = C \times P(B) \\ C^{-1} &= P(B) \end{aligned}\right\} \qquad \boxed{P(A|B) := \frac{P(A, B)}{P(B)}}$$

$$P(B) \neq 0 \qquad (\textit{Kolmogorov}, \ldots)$$

Statistical Independence

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

$P(A|B) = P(A)$ ***The occurrence of*** A ***does not depend on*** B

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

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

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***

For a finite collection of n events $A = \{A_1, A_2, \ldots, A_n\} \subset B$ ***independendence iff for each subset*** $\{A_p, \ldots, A_m\} \subset A$ ⟶ $P(A_p, \ldots, A_m) = P(A_p) \cdots P(A_m)$

… Conditional independence

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

It could happen that A depends on B through C $P(A,B|C) \neq P(A|C)P(B|C)$

Statistical Independence → Th. Total Probability and Bayes Th. …

Bayes Theorem

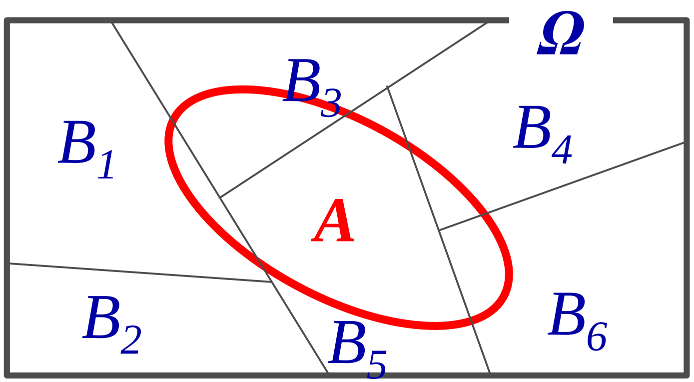
$$P(A,B) = P(A|B)\cdot P(B) = P(B|A)\cdot P(A) \rightarrow P(B|A) = \frac{P(A|B)\cdot P(B)}{P(A)}$$

We shall use that extensively + interpretations/applications in Lecture 3

EXERCISE: Cause(hypothesis)-effect interpretation

Theorem of Total Probability

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



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

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

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

Theorem of Total Probability with Conditional Probabilities $P(A,B,C) \rightarrow P(A|B) = \sum_{C} P(A|C,B)\cdot P(C|B)$

Exercise + Problem:

Cause(hypothesis)-effect interpretation of Bayes Theorem

Event A and partition of hypothesis space $\{H_k, k=1,\ldots n\}$

Probability of occurrence of event A having occurred H_i

Probability of occurrence of the event H_i "a priori", before we know if event A has occurred or not

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

$$i = 1,\ldots n$$

normalization

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

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

+ general hypothesis (H_0) (all probabilities are conditional to…)

$$P(\bullet\,|\,*) \to P(\bullet\,|\,*, H_0) \qquad P(\bullet) \to P(\bullet\,|\,H_0)$$

Problem:

(sic, healty) ↔ (positrons, protons)…

1) Incidence of a rare disease is 1 every 10,000 people

2) There is a test such that

if a person is sic, gives + in 99% of the cases
if a person is healty, test may fail (false positive) and give + in 0.5% of the cases

Hypothesis:	$H_1: \ \text{be sick}$	$H_2 = H_1^c: \ \text{be healthy}$
Test:	$T: \ \text{give positive}$	$T^c: \ \text{give negative}$
Conditional Probabilities:	$P(T \mid H_2) = 0.005$	$P(T \mid H_1) = 0.99$

3) A person is chosen at random (H_0) and gives positive

"The probability of giving positive being healthy is $P(T|H_2)$=0.5%, very small" **(p-value)**

Correct statement, ... but interpretation… and in any case is not what we are interested in.

Find: $\boxed{P(H_1 \mid T)}$ $\qquad \left(P(H_1^c \mid T) \quad ; \quad P(H_1 \mid T^c)\right)$

(ROC curves,… P(A|B) as function of $P(H_1)$,…

n-dimensional random quantity $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$

Marginal and Conditional Densities

$(\sum \leftrightarrow \int)$

$$F(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2) = \int_{-\infty}^{x_1} dw_1 \int_{-\infty}^{x_2} p(w_1, w_2) dw_2 \rightarrow p(x_1, x_2)$$

Joint p.d.f.

Marginal p.d.f.

$$X_1 \sim p(x_1) = \int_{-\infty}^{\infty} p(x_1, x_2) dx_2 \qquad X_2 \sim p(x_2) = \int_{-\infty}^{\infty} p(x_1, x_2) dx_1$$

Definition (pragmatic):
Conditional p.d.f.

$$p(x_2 \mid x_1) := \frac{p(x_1, x_2)}{p(x_1)} \qquad p(x_1 \mid x_2) := \frac{p(x_1, x_2)}{p(x_2)}$$

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

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

Independent:

$$\left.\begin{aligned} p(x_2 \mid x_1) &= p(x_2) \\ p(x_1 \mid x_2) &= p(x_1) \end{aligned}\right\} \rightarrow p(x_1, x_2) = p(x_1) p(x_2)$$

4) STOCHASTIC CHARACTERISTICS

"…when you cannot express it in numbers,
your knowledge is of a meagre and unsatisfactory kind."
(Lord W.T. Kelvin)

Mathematical Expectation

We know already that:

$X(w)$

	Discrete r.q.	***Absolute Continuous r.q***
takes values	$\Omega_X = \{x_1, x_2, \ldots\}$	$\Omega_X \subseteq R$
with probabilities	$P(X = x_i) = p_i$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} p(t)dt$
where	p_i ***: real, non-negative***	$p(x) = \frac{dF(x)}{dx} \geq 0$ ***and unique a.e.***
	$\sum_{\forall k} p_k = 1$	$\int_{-\infty}^{\infty} p(x)dx = 1$

Def.- Math. Expectation of r.q. $Y = g[X(\omega)]$ ***:***

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

$$\left(\sum \leftrightarrow \int \right)$$

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

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

$\alpha_0 = 1 \qquad \exists \alpha_n \to \exists \alpha_{m<n} \qquad \nexists \alpha_n \to \nexists \alpha_{m>n} \qquad \text{if } \exists \alpha_{2n} \text{ then } \alpha_{2n} \geq 0$

Mean: $\mu \equiv \alpha_1 = E[X] = \int_R x p(x)dx$

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

Moments wrt point $c \in R$ $\qquad E[(X-c)^n] = \int_R (x-c)^n p(x)dx$

$\min_{c \in R} E[(X-c)^2] \qquad c = \mu$ → ***…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** $Y = c_0 + c_1 X \xrightarrow[c_i \in R]{} V[Y] = \sigma_Y^2 = c_1^2 \sigma_X^2$

Skewness: $\gamma_1 \equiv \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3}$

Kurtosis: $\gamma_2 \equiv \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$

Cauchy-Schwarz inequality $\gamma_2 \geq 1 + \gamma_1^2$

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

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. $\rightarrow \exists \alpha_n$

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

$n = 1,2,\ldots$

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

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

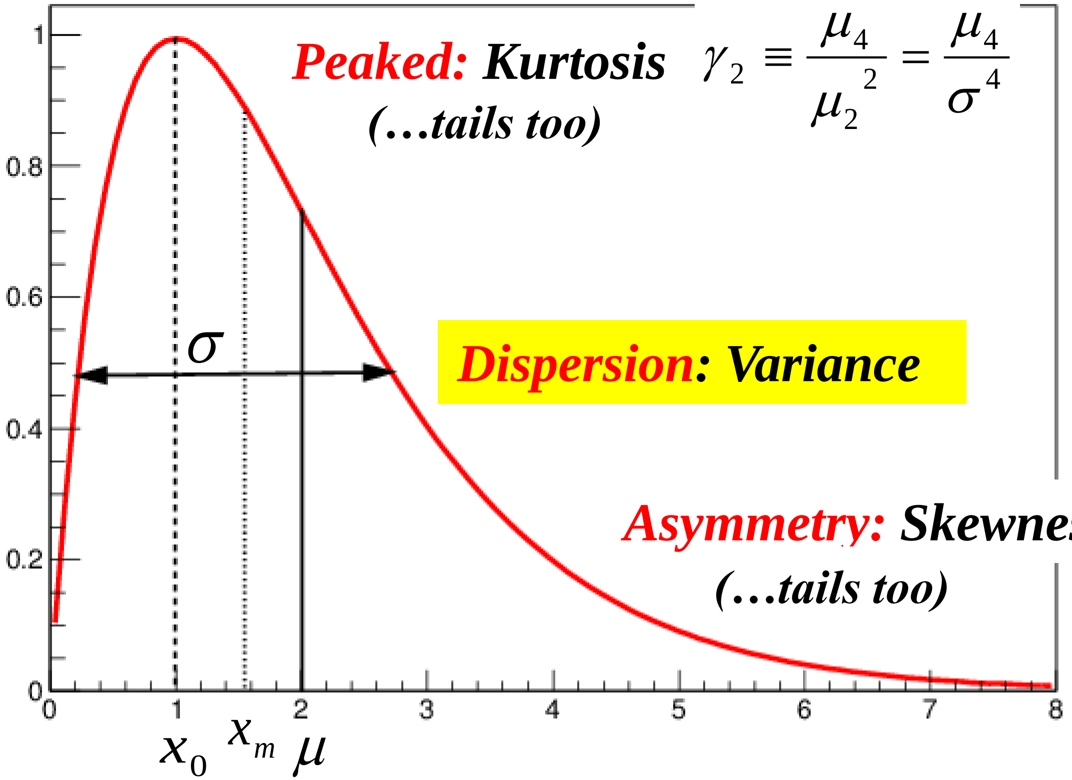
$\nexists \int_{-\infty}^{\infty} |x^n p(x)| dx < +\infty \qquad n \geq 1$

No moments (no mean, no variance,…)

(Cauchy PV for n=1

T. Distributions: Sobolev, Schwarz,…)

Global Picture



Peaked: Kurtosis $\gamma_2 \equiv \frac{\mu_4}{\mu_2^{\,2}} = \frac{\mu_4}{\sigma^4}$

(…tails too)

Extended: $\gamma_2^{ext} = \gamma_2 - 3 \begin{cases} > 0 \\ = 0 \quad \textit{Normal} \\ < 0 \end{cases}$

Dispersion: Variance

Asymmetry: Skewness $\gamma_1 = \frac{\mu_3}{\sigma^3} \begin{cases} > 0 \quad \textit{Right} \\ = 0 \quad \textit{Symmetric} \\ < 0 \quad \textit{Left} \end{cases}$

(…tails too)

***Position:* Mean:** $\mu = E[X]$ **, …**

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

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

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

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

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

$\mathbf{X}=\{X_i\}_{i=1}^{n}$

Covariance (and "Linear Correlation")

$$V[X_1,X_2] \doteq E[(X_1-\mu_1)(X_2-\mu_2)] = E[X_1X_2]-\mu_1\mu_2$$

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

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

$E[X_1X_2]=\mu_1\mu_2$

Cauchy-Schwarz inequality:

$$|\rho_{12}|\le 1$$

Covariance Matrix

$$\boldsymbol{\Sigma}=\left[\Sigma_{ij}\right]\equiv V[X_i,X_j]=\begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{pmatrix}$$

Symmetric $\quad \Sigma_{ij}=\Sigma_{ji}$

Non-negative $\quad \mathbf{x}^t\boldsymbol{\Sigma}\mathbf{x}\ge 0 \quad \forall \mathbf{x}\in R^n$

Correlation Matrix

$$\boldsymbol{\rho}=\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{1n} & \rho_{2n} & \cdots & 1 \end{pmatrix}$$

$$\rho_{ij} := \frac{V[X_i,X_j]}{\sigma_i\sigma_j}$$

Exercise:

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

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

The "error propagation" rule…

Useful but to be used with care!!

$$Y = g(X_1, X_2, \ldots, X_n) = g(\mathbf{X}) \longrightarrow Y = g(\mathbf{X}) = g(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial g}{\partial x_i}\right]_{\boldsymbol{\mu}} (x_i - \mu_i) + O(D_{ij}^2)$$

Taylor Expansion around $E(X_i) = \mu_i$

$$E[Y] = E[g(\mathbf{X})] = g(\boldsymbol{\mu}) + O(D_{ij}^2) \longrightarrow Y - E[Y] = \sum_{i=1}^{n} \left[\frac{\partial g}{\partial x_i}\right]_{\boldsymbol{\mu}} (x_i - \mu_i) + O(D_{ij}^2)$$

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

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

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

(mind for the remainder…) *(do moments exist??)*

Exercise:

1) X_1, X_2 *indep* $X = X_1 X_2$ **Compare** $V[X]$ **with** $V_{ep}[X]$

2) $X_i \sim N(x \mid \mu_i, \sigma_i);\ \mu_i \neq 0$ $X = X_1 X_2^{-1}$ **Think about** $V_{ep}[X]$

5) ORDERED STATISTICS **(see section 6 of the notes)**

6) INTEGRAL TRANSFORMS

Fourier Transform

(Laplace)

$$\Phi(t) = \int_{-\infty}^{\infty} e^{ixt} f(x)dx$$

$t \in R$

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

$$X = X_1 \pm X_2 \pm \ldots$$

Mellin Transform

$$M(f;s) \doteq \int_{0}^{\infty} f(x)x^{s-1}dx$$

$s \in \Lambda \subseteq C$

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

$$X = X_1 X_2 \ldots ; \;\; X_1 X_2^{-1} \ldots$$

Moments of a Distribution

(see back-up slides and notes for details+ useful examples/relations)

7) LIMIT THEOREMS and CONVERGENCE

General Problem: ***Find the limit behaviour of a sequence of random quantities***

Example:

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

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

~"distance" → ***convergence criteria***

1) More or less strong convergence

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

Convergence in:

Distribution	→	***Central Limit Theorem*** ***Glivenko-Cantelli Theorem (weak)***
Probability	→	***Weak Law of Large Numbers***
Almost Sure	→	***Strong Law of Large Numbers***
$L_p(R)$ ***Norm***	→	***Convergence in Quadratic Mean***
Uniform	→	***Glivenko-Cantelli Theorem***
Logarithmic Convergence		

(Probability ↑ Distribution; $L_p(R)$ Norm ↑ Probability; Almost Sure ↑ Probability)

Chebyshev Theorem

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

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

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

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

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

$$\geq 0$$

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

Bienaymé-Chebyshev Inequality

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

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

Convergence in Probability

Consider the sequence $\{X_1, X_2, \ldots, X_n, \ldots\}$

Def.: $\{X_n\}_{n=1}^{\infty}$ converges in probability to X if, and only if

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

lim (Prob)

n dimensions : $|\bullet| \to \|\bullet\|$

Weak Law of Large Numbers (J. Bernouilli…)

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of independent r.q. with the same Distribution Function and first order moment $E[X_i] = \mu$

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 in Probability to μ

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

Problem: show this from Chebyshev Inequality if $V[X_i]=\sigma^2$

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

Consider the sequence $\{X_1, X_2, \ldots, X_n, \ldots\}$

Def.: $\{X_n\}_{n=1}^{\infty}$ **converges "almost sure" to** X **if, and only if**

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

Prob (lim)

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

Let $\{X_i\}_{i=1}^{\infty}$ **be a sequence of independent r.q. with the same Distribution Function and first order moment** $E[X_i] = \mu$

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 in Almost Sure to** μ

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

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

Consider the sequence $\{X_1, X_2, ..., X_n, ...\}$

and their corresponding DF's: $\{F_1(x_1), F_2(x_2), ..., F_n(x_n), ...\}$

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

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

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

Convergence in Distribution determined only by DF

→ RQ do not have to have same support

Central Limit Theorem

(Lindberg-Levy,…)

1) Sequence of independent r.q. $\{X_i\}_{i=1}^{\infty}$ { ***same distribution*** / ***finite mean and variance*** (μ, σ^2)

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

In the limit $n \to \infty$

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k \sim N\left(z \mid \mu, \sigma n^{-1/2}\right)$$

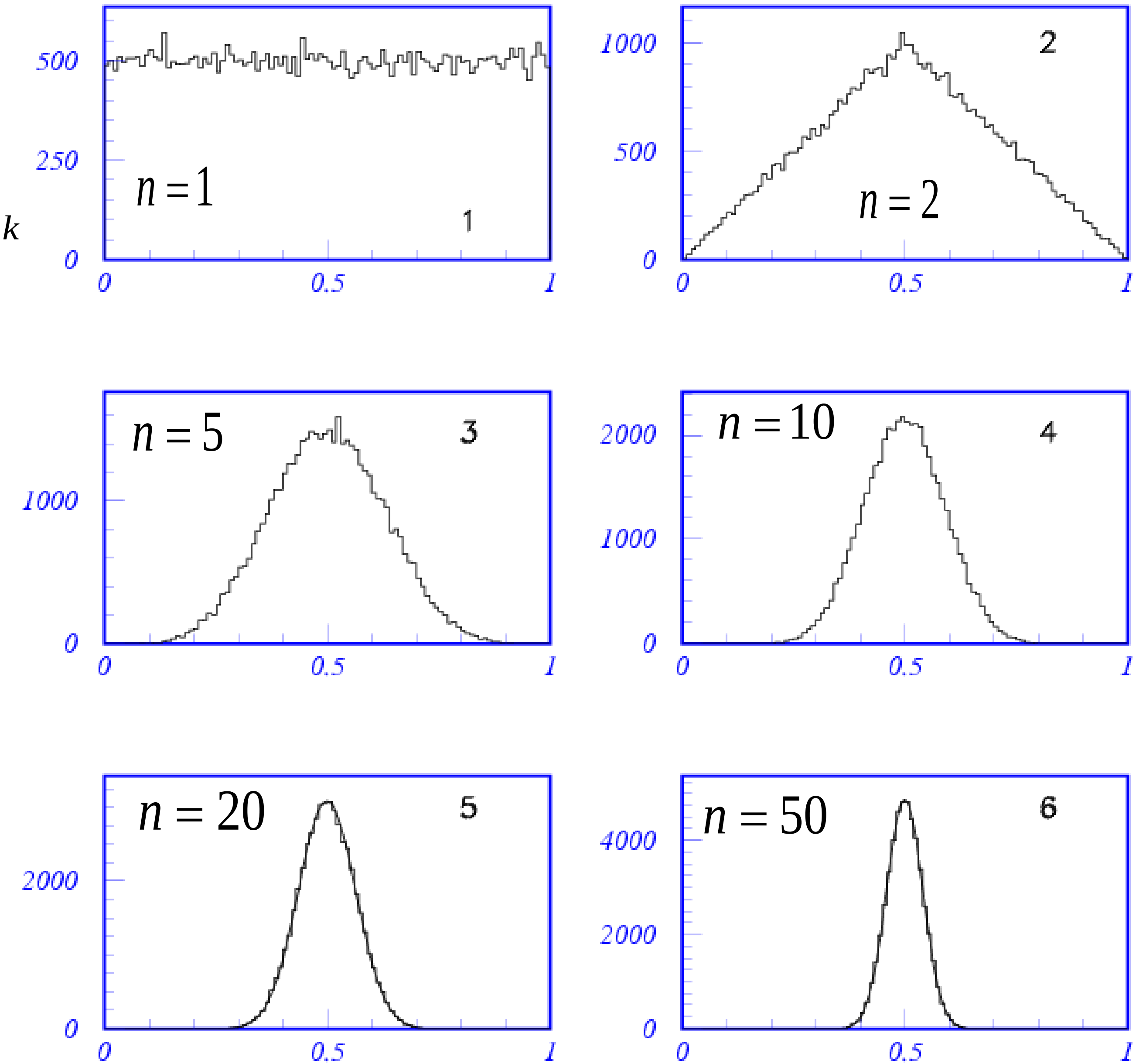
standardized: $\tilde{Z}_n = \frac{Z_n - \mu}{\sigma n^{-1/2}} \sim N(z|0,1)$

Problem: show this from $\lim_{n\to\infty} \phi_n(t)$ $\quad X \sim N(x \mid \mu, \sigma) \to \phi(t) = \exp\left\{it\mu - \frac{1}{2}\sigma^2 t^2\right\}$

Example (CLT: Watch for conditions of applicability!!):

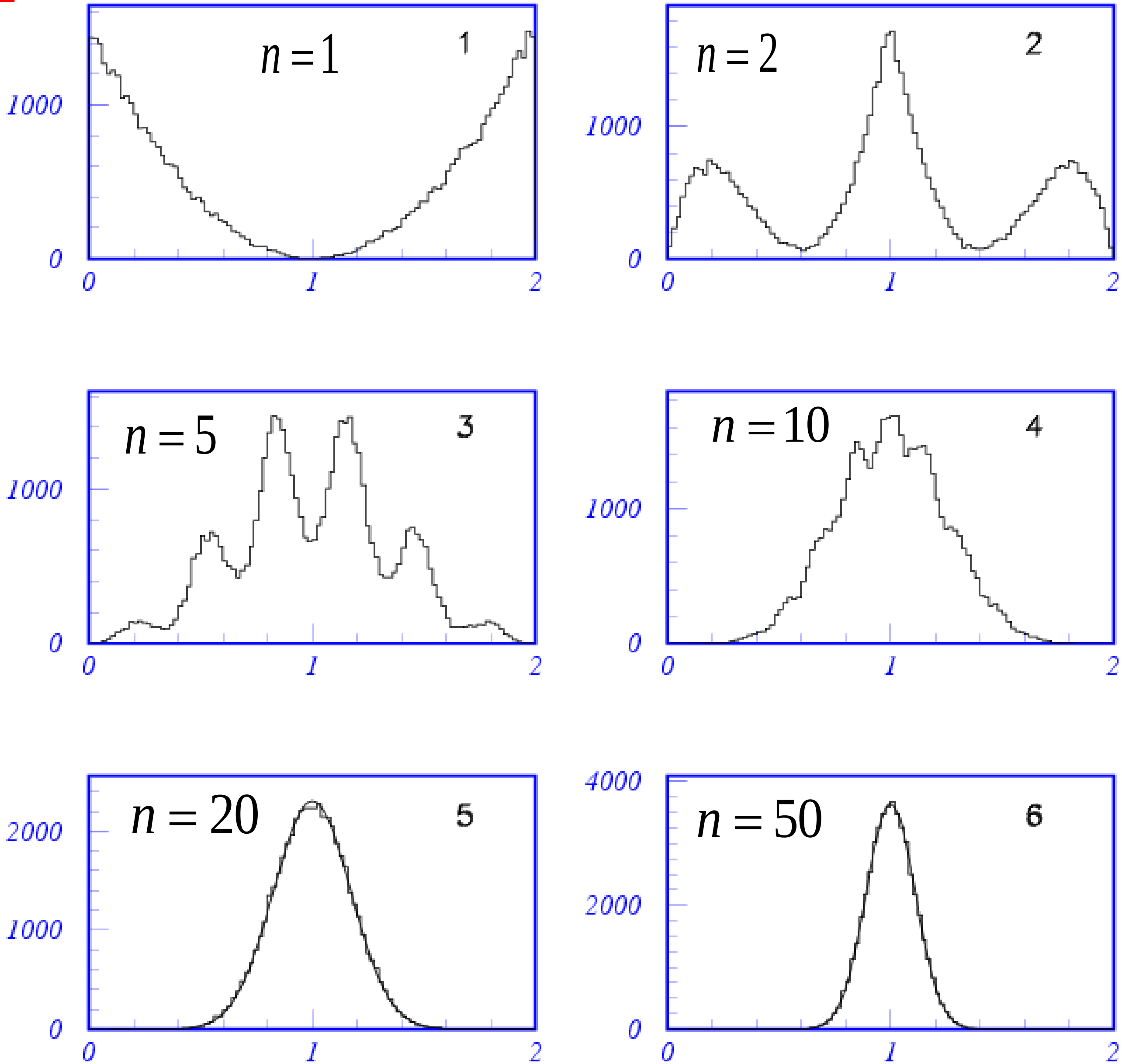
Unifom DF

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



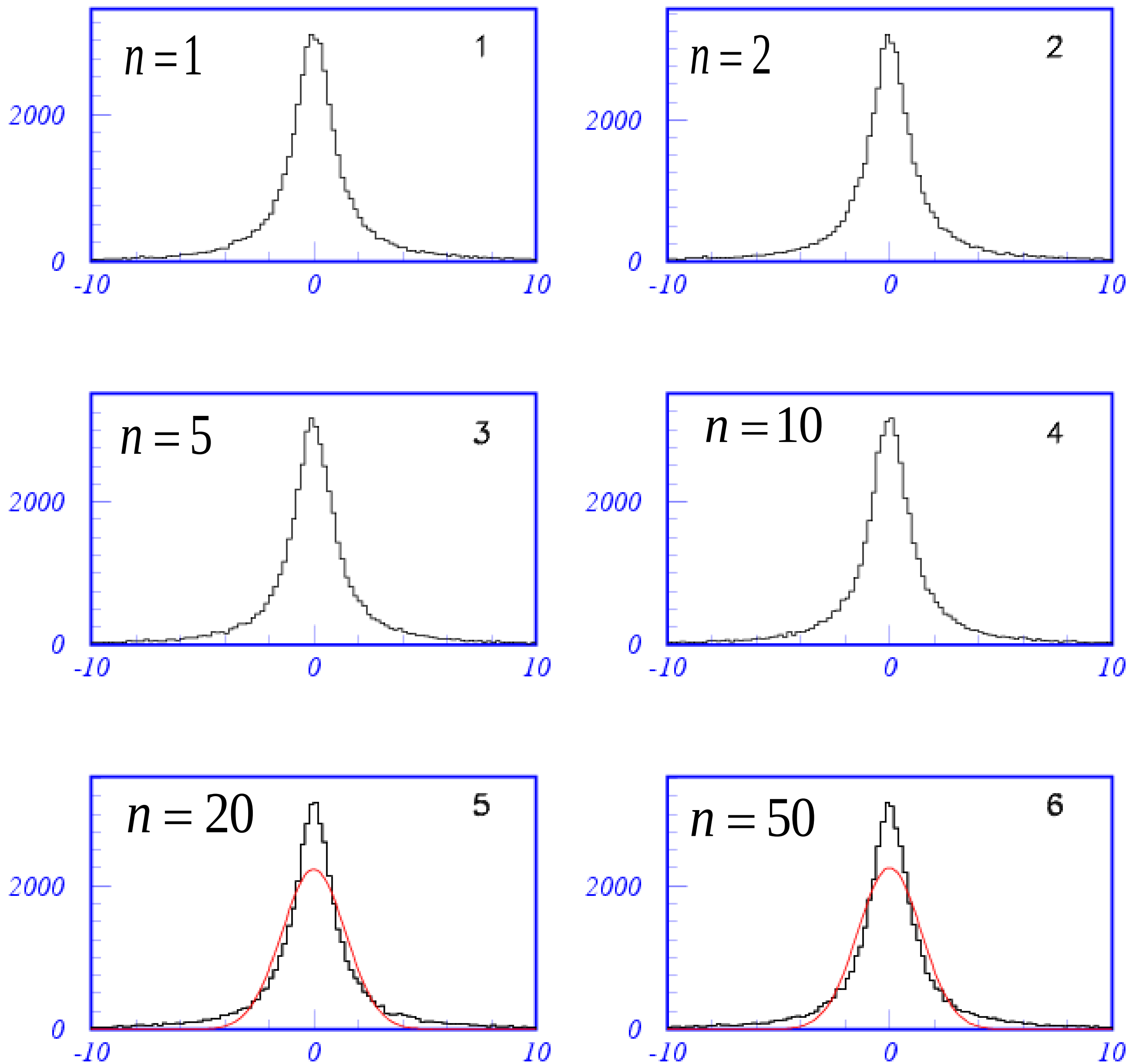
Parabolic DF

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



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***

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

Glivenko-Cantelli Theorem

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

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

Empiric Distribution Function

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

$$\left[\frac{\textit{number of values } x_i \textit{ lower or equal to } x}{n}\right]$$

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

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

Example $X_i \sim Un(x\,|\,0,1)$

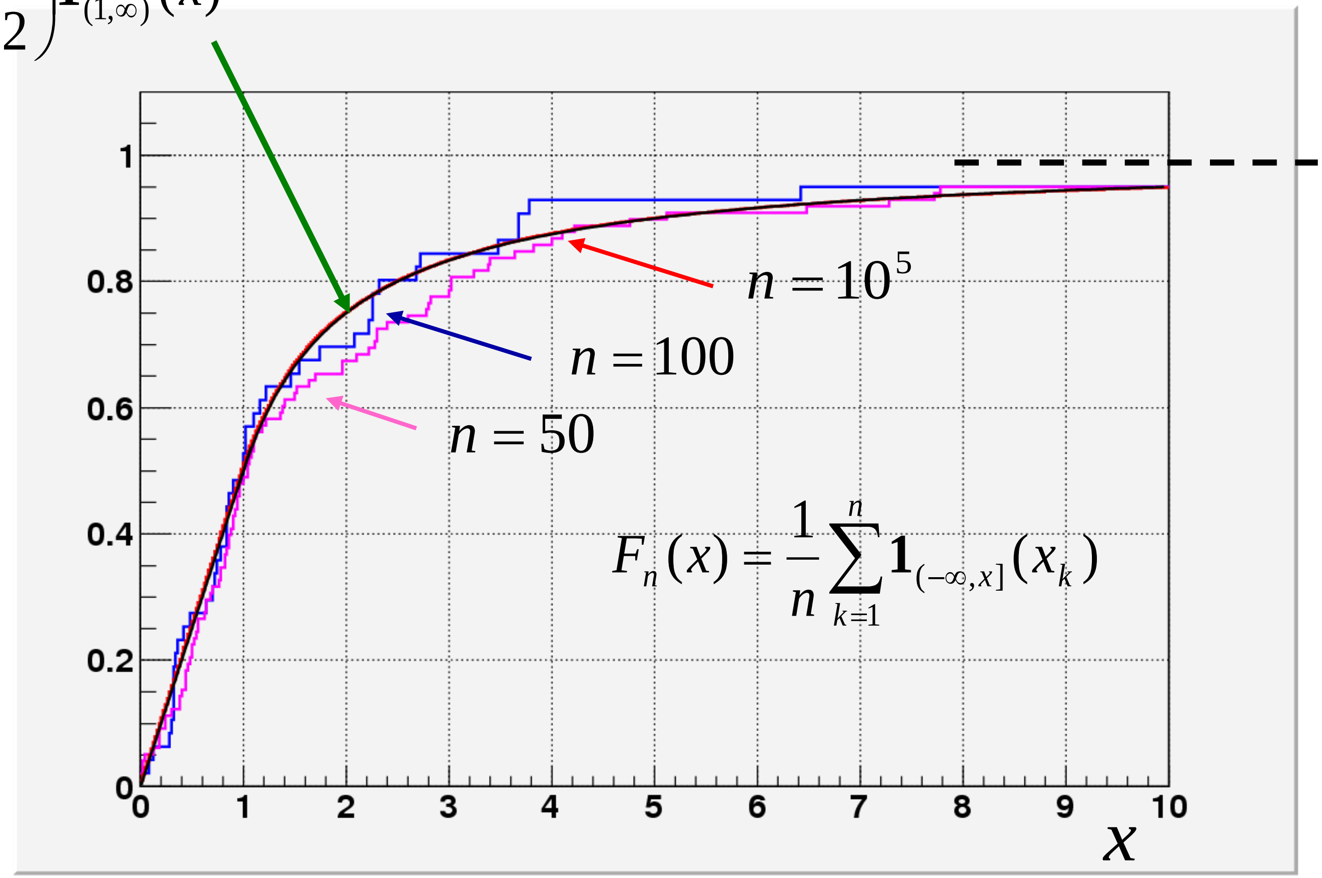
$$X = X_1 X_2^{-1} \sim \frac{1}{2}\left[\mathbf{1}_{(0,1)}(x) + \frac{1}{x^2}\mathbf{1}_{(1,\infty)}(x)\right]$$

$: Un(x\,|\,0,1) + Pa(x\,|\,1,1)$

(show that from Mellin Transform)

$$F(x) = \frac{x}{2}\mathbf{1}_{(0,1]}(x) + \left(1 - \frac{x}{2}\right)\mathbf{1}_{(1,\infty)}(x)$$

No moments $E[x^n]; \quad n \geq 1$



… Bootstrap in Monte Carlo

End of Lecture 1…

Backup slides: ***Notes on Integral Transforms***

6) INTEGRAL TRANSFORMS

“Fourier” Transform… (Characteristic Function)

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

$$\Phi(t) = \int_{-\infty}^{\infty} e^{ixt} f(x)dx$$

$t \in R$

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

Probability Density…

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

Exists for all $X(\omega)$

Properties:

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

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

(all necessary but not sufficient)

Inversion Theorem (Lévy,1925)

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

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

Reticular: $\quad x_k = a + bn \qquad a,b \in R, \quad b \ne 0 \qquad n \in Z$

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

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

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

Useful Relations:

- $$Y = g(X) \quad \rightarrow \quad \Phi_Y(t) = E[e^{iYt}] = E[e^{ig(X)t}] \qquad = \int_{-\infty}^{\infty} e^{ig(x)t}\, p(x)dx$$

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

- ***If*** ***are n independent random quantities***

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

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

- ***If distribution of*** X ***is symmetric:*** $\Phi_X(t) = \Phi_{-X}(t) = \Phi_X(-t) = \overline{\Phi}_X(t)$

then $\Phi_X(t)$ ***is a real function***

Example $X_i \sim Po(n_i \mid \mu_i)$

$X = X_1 - X_2 \qquad \Omega_X = Z$

$$\Phi_i(t) = e^{-\mu_i} \sum_{x=0}^{\infty} \frac{(\mu_i e^{it})^x}{\Gamma(x+1)} = \exp\{-\mu_i (1 - e^{it})\}$$

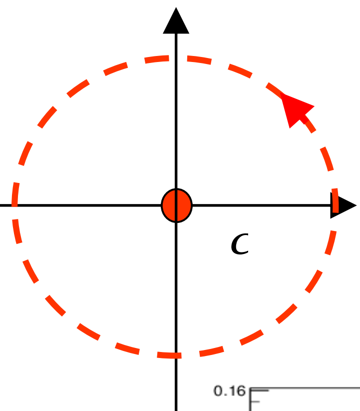
$$\Phi_X(t) = \Phi_1(t)\overline{\Phi}_2(t) = e^{-(\mu_1+\mu_2)} \exp\{\mu_1 e^{it} + \mu_2 e^{-it}\}$$

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

$$p(X = x_k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itx_k} \Phi(t) dt$$

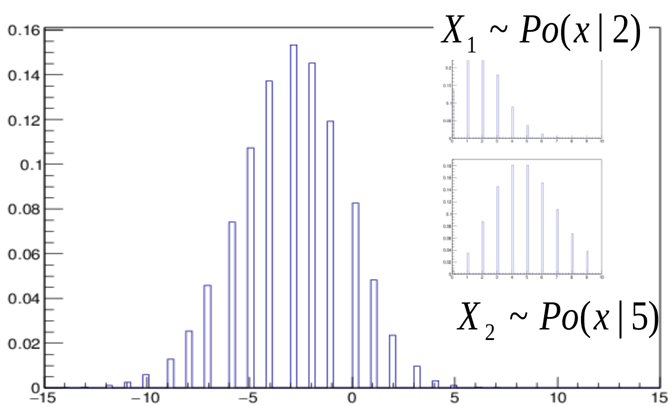
$$P(X = n) = \left(\frac{\mu_1}{\mu_2}\right)^{n/2} \frac{e^{-\mu_S}}{2\pi i} \oint_C z^{-(n+1)} e^{-\frac{w}{2}(z+1/z)} dz$$

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



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 for the Sum of Random Quantities:

$$X = X_1 + \cdots + X_n$$

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

Moments of a Distribution

(Fourier / Laplace Transforms usually called "moment generating functions")

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

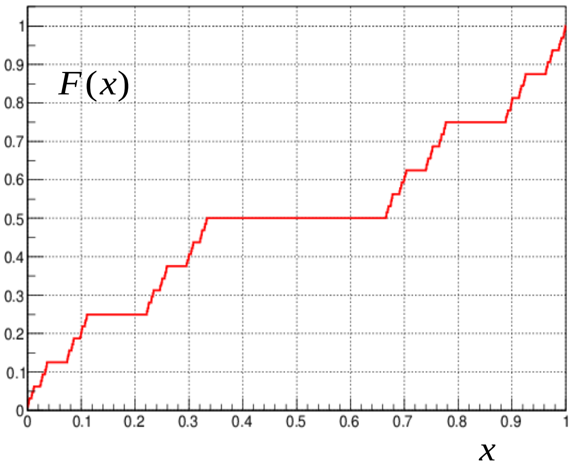
$$\Phi(t_1,\ldots,t_n) = E[e^{i(X_1 t_1 + \ldots + X_1 t_n)}] \longrightarrow E[X_i^{k_i} X_j^{k_j}] = (-i)^{k_i+k_j} \left[\frac{\partial^{k_i}}{\partial^{k_i} t_i} \frac{\partial^{k_j}}{\partial^{k_j} t_j} \Phi(t_1,\ldots,t_n) \right]_{t_1,\ldots,t_n=0}$$

Example

$$X \sim Cs(0,1) \qquad X = \sum_{n=1}^{\infty} \frac{X_n}{3^n} \qquad \mathrm{supp}[X_n] = \{0,2\} \qquad P(X_n = 0) = P(X_n = 2) = \frac{1}{2}$$

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

(the integral $\int_{-\infty}^{x} p(u)du$ is struck through)



$$\Phi_{X_n}(t) = E[e^{iXt}] = \frac{1}{2}\left(1 + e^{2it}\right) \longrightarrow \boxed{\Phi_X(t) = e^{it/2} \prod_{n=1}^{\infty} \cos(t/3^n)}$$

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

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

Mellin Transform

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

$$M(f;s) \doteq \int_0^\infty f(x) x^{s-1} dx$$

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

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

Convergence of integral

Strip of holomorphy

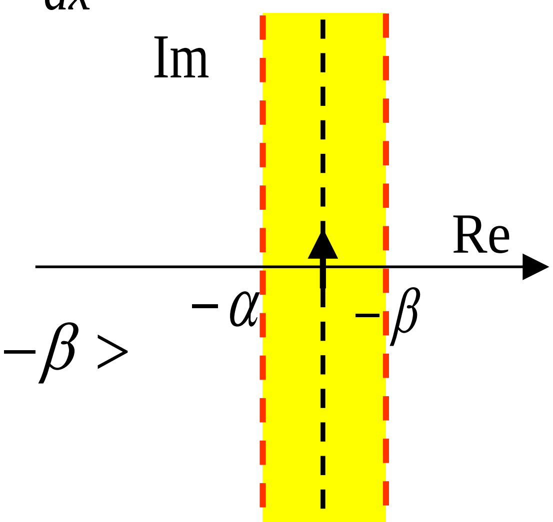
$< -\alpha, -\beta >$

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

$\lim_{x \to 0^+} f(x) = O(x^{\alpha}) \qquad |M(f;s)| \le C_1 \int_0^1 x^{\mathrm{Re}(s)-1+\alpha} dx + C_2 \int_1^\infty x^{\mathrm{Re}(s)-1+\beta} dx$

$\lim_{x \to \infty} f(x) = O(x^{\beta}) \qquad -\alpha < \mathrm{Re}(s) < -\beta$

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

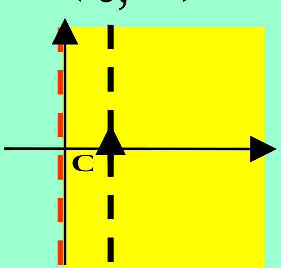


$M(f;s) \oplus < -\alpha, -\beta >$

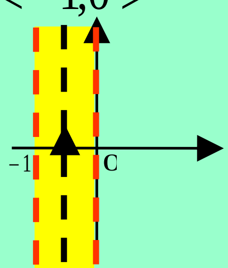
$f_1(x) = e^{-x} \qquad f_2(x) = e^{-x} - 1$

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

$< 0, \infty >$



$< -1, 0 >$



Useful Relations:

▶ $Y = aX^{b}$ $\quad a,b \in R, \quad a > 0$

$$\int y^{s-1} dP(y) = a^{s-1} \int x^{b(s-1)} dP(x) \longrightarrow M_Y(s) = a^{s-1} M_X(bs - b + 1)$$

$a = 1, b = -1$

$$Y = X^{-1} \longrightarrow M_Y(s) = M_X(2 - s)$$

▶ $\{X_i \sim p_i(x_i)\}_{i=1}^{n}$ **Independent** **and non- negative** $x_i \in [0, \infty)$

$$X = X_1 X_2 \cdots X_n \longrightarrow M_X(s) = M_1(s) \cdots M_n(s)$$

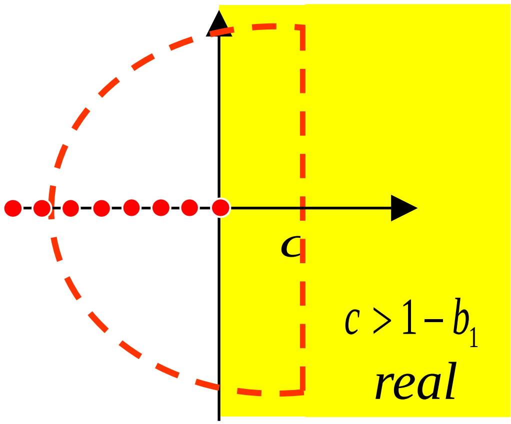
$$X = X_1 X_2^{-1} \longrightarrow M_X(s) = M_1(s) M_2(2 - s)$$

moments too: $M_X(s) = E[X^{s-1}] \to M_X(1) = 1, \quad M_X(n+1) = E[X^{n}]$

$X = X_1 X_2$ $\quad M_X(s) = \left(\int_0^{\infty} x^{s-1} p(x) dx \quad \oplus \quad p(x) = \int_0^{\infty} p_1(w) p_2(x/w) \frac{1}{|w|} dw \right) = M_1(s) M_2(s)$

Example $X_i \sim Ga(x_i \mid a_i, b_i)$ $\quad$ $X = X_1 X_2$

$$Y_i = a_i X_i \sim Ga(y_i \mid 1, b_i) \Rightarrow M_{Y_i}(s) = \frac{\Gamma(b_i + s - 1)}{\Gamma(b_i)} \Rightarrow M_{X_i}(s) = a_i^{1-s} M_Y(s) = a_i^{1-s} \frac{\Gamma(b_i + s - 1)}{\Gamma(b_i)}$$



$$M_X(z) = M_{X_1}(z) M_{X_2}(z) = (a_1 a_2)^{1-z} \frac{\Gamma(b_1 + z - 1)}{\Gamma(b_1)} \frac{\Gamma(b_2 + z - 1)}{\Gamma(b_2)}$$

Without loss of generality, assume $b_2 > b_1$

Strip of holomorphy $< 1 - b_1, \infty >$

$$(a_1 a_2)^{-1} \Gamma(b_1) \Gamma(b_2) p(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} (a_1 a_2 x)^{-z} \Gamma(b_1 + z - 1) \Gamma(b_2 + z - 1) dz$$

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

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

$X = X_1 X_2^{-1}$

$$X = X_1 X_2^{-1} \sim \frac{\Gamma(b_1 + b_2)}{\Gamma(b_1)\Gamma(b_2)} \frac{a_1^{b_1} a_2^{b_2} x^{b_1 - 1}}{(a_2 + a_1 x)^{b_1 + b_2}}$$

$$X_i \sim Ex(x_i \mid a_i) = Ga(x_i \mid a_i, 1)$$

Be careful with strips and integrals! …
(see Notes for more examples)

Example

$X_i \sim Un(x_i \mid 0,1)$

Be careful with strips and integrals …

$$M_X(s) = \int_0^\infty x^{s-1} \mathbf{1}_{(0,1)}(x)dx = \frac{x^s}{s}\Big|_0^1 = \frac{1}{s}$$

$$M_{Y=X^{-1}}(s) = \int_0^\infty y^{s-1} \mathbf{1}_{(1,\infty)}(y) y^{-2} dy = \frac{y^{s-2}}{s-2}\Big|_1^\infty = \frac{1}{2-s}$$

converge for

$s > 0 \rightarrow \langle 0,\infty \rangle$

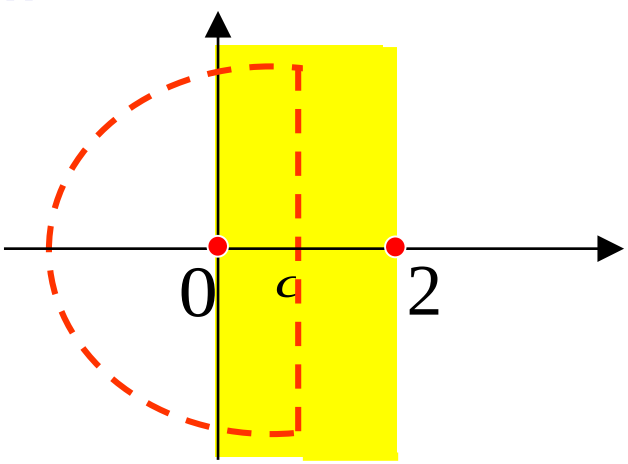
$s < 2 \rightarrow \langle 0,2 \rangle$

$$Z = X_1 X_2^{-1}$$

$$M_Z(s) = M_X(s) M_{X^{-1}}(s) = \frac{1}{s(2-s)}$$

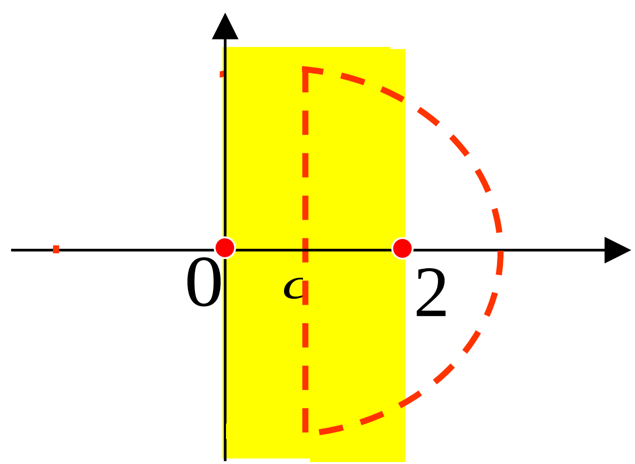
Strip of holomorphy $\langle 0,2 \rangle$

Different Bromwich Contours for x>1 and x<1:



$c > 0$ **real**

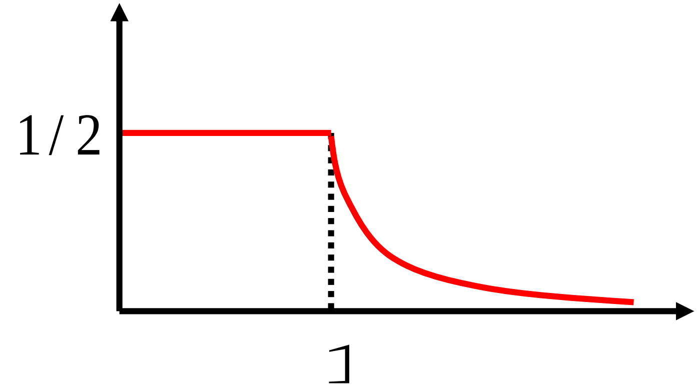
$\ln x < 0 \rightarrow x < 1$



$\ln x > 0 \rightarrow x > 1$

$$X = X_1 X_2^{-1} \sim \frac{1}{2}\left[\mathbf{1}_{(0,1)}(x) + \frac{1}{x^2}\mathbf{1}_{(1,\infty)}(x)\right]$$

$$Un(x\,|\,0,1) + Pa(x\,|\,1,1)$$



$$M(p;s) = \int_0^\infty p(x)x^{s-1}dx \qquad M_X(s) = \frac{1}{s(2-s)}$$

$$M(p;n+1) = \int_0^\infty p(x)x^n dx \longrightarrow E[X^n] = M_X(n+1) = \frac{1}{(n+1)(1-n)}$$

→ ***No moments for n≥1***

$$X_i \sim Un(x_i\,|\,0,1)$$

$$\rightarrow \quad Z = X_1 X_2 \cdots X_n \qquad M_Z(s) = \frac{1}{s^n} \qquad <0,\infty>$$

$z = 0$ ***pole of order n***

$$X = X_1 X_2 \cdots X_n \sim \frac{(-\ln(x))^{n-1}}{\Gamma(n)}\mathbf{1}_{(0,1)}(x)$$

What if supp$\{p(x)\} = R$?

$$M(f;s) = \int_0^\infty f(x) x^{s-1} dx \qquad f : R^+ \to C \qquad f \in L_1(R^+)$$

Partition of supp$\{p_1(x)\} \times$ supp$\{p_2(x)\}$

and change $\{z_1 = \pm x_1 \,;\, z_2 = \pm x_2\}$ **so that supports are on** R^+

Example:

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

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

$$X = X_1 X_2 \sim \left(\frac{a}{\pi}\right) K_0(a\,|x|) \qquad a = (\sigma_1 \sigma_2)^{-1}$$

	$X = X_1 X_2$	$X = X_1 X_2^{-1}$
obviously…	$p(x) = \int_\Omega p(w, x w^{-1})\,\lvert w \rvert^{-1}\,dw$	$p(x) = \int_\Omega p(xw, w)\,\lvert w \rvert\,dw$
independent…	$p(x) = \int_\Omega p_1(w)\,p_2(x/w)\,\frac{1}{\lvert w \rvert}\,dw$	$p(x) = \int_\Omega p_1(xw)\,p_2(w)\,\lvert w \rvert\,dw$

MT: usually more involved but… we have the moments with same effort

Example: Ratio of Normal and χ^2 Distributed r.q.

$$X_1 \sim N(x_1 \mid 0,1) \qquad \mathrm{supp}(X_1) = R$$

$$X_2 \sim \chi^2(x_2 \mid n) \qquad \mathrm{supp}(X_2) = R^+$$

$$X = X_1 (X_2 / n)^{-1/2}$$

- $Z = (X_2/n)^{-1/2}$, $M_{X_2}(s) = \dfrac{2^{s-1}\Gamma(n/2+s-1)}{\Gamma(n/2)}$ $\Bigg\}$ $M_Z(s) = n^{(s-1)/2} M_{X_2}((3-s)/2) = \left(\dfrac{n}{2}\right)^{(s-1)/2} \dfrac{\Gamma((n+1-s)/2)}{\Gamma(n/2)}$

$$0 < \mathrm{Re}(s) < n+1$$

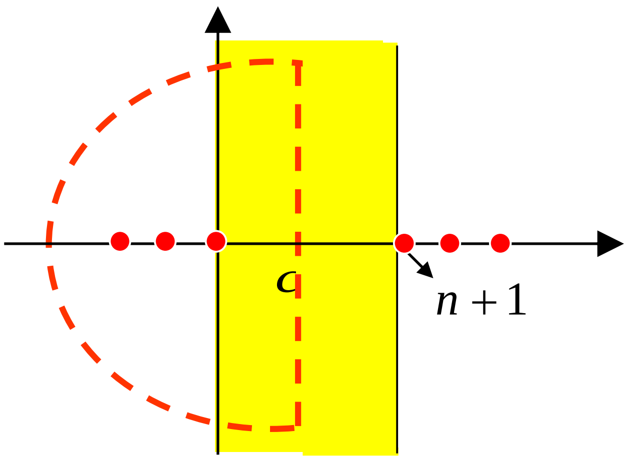
- $X_1 \sim N(x_1 \mid 0,1)$ $\quad p(x_1) = \underbrace{p(x_1)\mathbf{1}_{[0,\infty)}(x_1)}_{p_1^+(x_1)} + \underbrace{p(x_1)\mathbf{1}_{(-\infty,0)}(x_1)}_{p_1^-(x_1)}$

$$M_1^+(s) = \frac{2^{s/2}\Gamma(s/2)}{2\sqrt{2\pi}} \qquad 0 < \mathrm{Re}(s)$$

- $X \sim p(x) = p(x)\mathbf{1}_{[0,\infty)}(x) + p(x)\mathbf{1}_{(-\infty,0)}(x) = p^+(x) + p^-(x)$

$p^+(x)$

$$M_X^+(s) = M_1^+(s) M_Z(s) = \frac{n^{s/2}\Gamma(s/2)\Gamma((n+1-s)/2)}{2\sqrt{n\pi}\,\Gamma(n/2)}$$



Holomorphy: $0 < \mathrm{Re}(s) < n+1$

Poles:

$$s_0 = -2m\ ; \quad m = 0,1,2,\ldots$$

$$s_0 = n+1+2k\ ; \quad k = 0,1,2,\ldots$$

$$p^+(x) = \frac{1}{\sqrt{n\pi}\,\Gamma(n/2)} \sum_{m=0}^{\infty} \frac{(-1)^m}{\Gamma(m+1)} \left(\frac{x^2}{n}\right)^m \Gamma\left(\frac{n+1}{2} + m\right) =$$

$$= \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

$p^-(x)$ ***Same (symmetry)***

$$\left.\begin{array}{l} X_1 \sim N(x_1 \mid 0,1) \\ X_2 \sim \chi^2(x_2 \mid n) \end{array}\right\} \quad X = X_1 \left(\frac{n}{X_2}\right)^{1/2} \sim p(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} = St(X \mid n)$$

Example: ***Ratio of two Normal Distributed r.q.***

$$X_i \sim N(x_i \mid \mu_i, \sigma_i) \qquad X = X_1 X_2^{-1} \qquad p(x) = \int_{-\infty}^{\infty} p_1(xw)\, p_2(w)\, |w|\, dw$$

$$p(x) = \frac{1}{\pi}\frac{a}{1+x^2a^2}\exp\left\{-\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right)\right\}\left\{1+\sqrt{\frac{\pi}{2}}f(x)e^{f(x)^2/2}erf\left(f(x)2^{-1/2}\right)\right\}$$

$$a = \sigma_2\sigma_1^{-1} \qquad f(x) = \frac{\mu_2\sigma_2^{-1} + xa\mu_1\sigma_1^{-1}}{\sqrt{1+x^2a^2}}$$

1) $\mu_1 = \mu_2 = 0 \rightarrow Ca(x \mid 0, a^2)$

2) Has no variance

$$X_1 \sim N(x_1 \mid \mu_1 = -4, \sigma_1 = 1)$$

$$X_2 \sim N(x_2 \mid \mu_2 = 4, \sigma_2 = 3)$$

MC sampling ***Pdf: p(x)***