Statistical methods

Tomaž Podobnik

Oddelek za fiziko, FMF, UNI-LJ

Odsek za eksperimentalno fiziko delcev, IJS

I. Prologue:

1. Literature,
2. Basic Desiderata,
3. Example.

1. Literature:

- no single textbook covers all topics appropriately,
- relevant references will be given for each of the topics separately.

2. Basic Desiderata:

A Every system must be (self-) consistent.

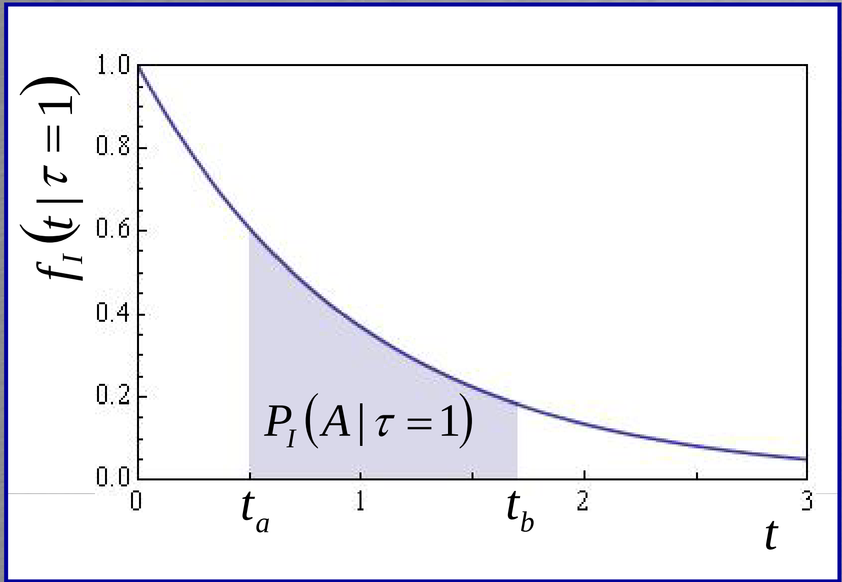
B Every scientific (empirical) theory must be operational
(i.e., it must specify operations that ensure falsifiability
of its predictions).

K. Popper (1959), *The Logic of Scientific Discovery,* § 24, pp. 91-92.
J. Neyman (1937), Phil. Trans. R. Soc., **A 236**, 333-380.
G. Pólya (1954), *Mathematics and Plausible Reasoning, Vol. 2 –*
Patterns of Plausible Inference, Chap. XIV, § 4, p. 64,
and Chap. XV, § 4, p. 117.

Definition 1 (Inconsistency). *A system is* inconsistent *iff, within the rules of the system, a conclusion can be reasoned out in more than one ways, and not all the ways lead to the same result.*

Definition 2 (Consistency). *A system that is not inconsistent (i.e., a system without a single demonstrated inconsistency) is said to be* consistent.

3. Example 1. A decay of an atom (nucleus, particle) at time t_1:



$$f_I(t_1|\tau) = \frac{1}{\tau}\exp\left\{-\frac{t_1}{\tau}\right\};$$

$$I = \left\{ f(t|\tau) = \tau^{-1}\exp[-t/\tau];\ \tau \in \mathbb{R}^+ \right\}$$

$$P_I(A|\tau) = \int_{t_a}^{t_b} f_I(t|\tau)\,dt\,; \quad A = (t_a, t_b)$$

Given t_1, what is the value of the parameter τ?

The Maximum Likelihood Method (MLM):

R. A. Fisher (1922), Phil. Trans. R. Soc., **A 222**, 309-368.

Likelihood function: $L(t_1;\tau) \equiv f_I(t_1 \mid \tau)$.

"The principle of Maximum Likelihood states that, when confronted with a choice of τ we choose that one (if any) which maximizes *L*."

A. Stuart, J. K. Ord (1994), *Kendall's Advanced Theory of Statistics, Vol. 1 – Distribution Theory,* § 8.22, p. 300.

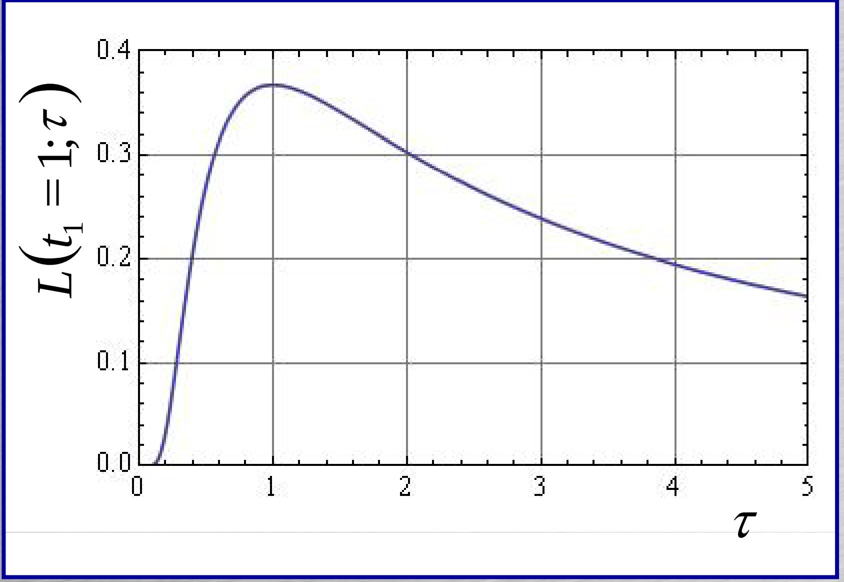
I. Kuščer, A. Kodre (1994), *Matematika v fiziki in tehniki,* § 11.8, p. 309.

R. Jamnik (1975), *Verjetnostni račun in statistika,* § 24, p. 529-530. In I. Vidav (ed.), *Višja matematika II*.

General method:

$$\hat{\tau}: \frac{\partial}{\partial \tau} \ln L(t_1;\tau)\Big|_{\tau=\hat{\tau}} = 0$$

$$\hat{\tau} = t_1$$



What justifies MLM?

(The fallacy of) The limiting property of $\hat{\tau}$:

Set of measurements $\{t_1,\ldots,t_n\}$: i.i.d.

$$L(t_1,\ldots,t_n;\tau) = f_I(t_1,\ldots,t_n \mid \tau) = \prod_{i=1}^{n} f_I(t_i \mid \tau)$$

$$\hat{\tau} = \hat{\tau}(t_1,\ldots,t_n) : \frac{\partial}{\partial\tau} \ln L(t_1,\ldots,t_n;\tau)\Big|_{\tau=\hat{\tau}} = 0 \quad \Rightarrow \quad \hat{\tau} = \bar{t}_n = \frac{1}{n}\sum_{i=1}^{n} t_i$$

$$\lim_{n\to\infty} : f(\hat{\tau} \mid \tau) \sim N\left(\tau, \frac{\tau}{\sqrt{n}}\right)$$

$$N_C = 3 \approx \infty , \qquad n = 1 \approx \infty$$

The conditional probability fallacy:

Interested in $\tau : f(\tau \mid t_1) = \max.$, calculate $\tau : f(t_1 \mid \tau) = \max.$

$\Rightarrow$ an implicit assumption: $f_I(\tau \mid t_1) = f_I(t_1 \mid \tau)$

$$\int_0^\infty f_I(t_1 \mid \tau)\, d\tau = \infty$$

Example (L. Lyons):

A: "The first person I'll meet is female."
B: "The first person I'll meet is pregnant."

$$P(A \mid B) \neq P(B \mid A)$$

The fallacy of point estimates:

The parameter τ may take on every value in a continuum $\mathbb{R}^+ \Longrightarrow$ a measure of a single point $\hat{\tau}$ in the continuum is 0.

For verifiable predictions we must turn to interval estimations.

"Ignorance is preferable to error and he is less remote from the truth who belives nothing than he who believes what is wrong."

(T. Jefferson (1781). *Notes on Virginia*.)

Contents:

I. Prologue,

II. Mathematical Preliminaries,

III. Frequency Interpretation of Probability Distributions,

IV. Confidence Intervals,

V. Testing of Hypotheses,

VI. Inverse Probability Distributions,

VII. Interpretation of Inverse Probability Distributions,

VIII. Time Series and Dynamical Models.

II. Mathematical Preliminaries:

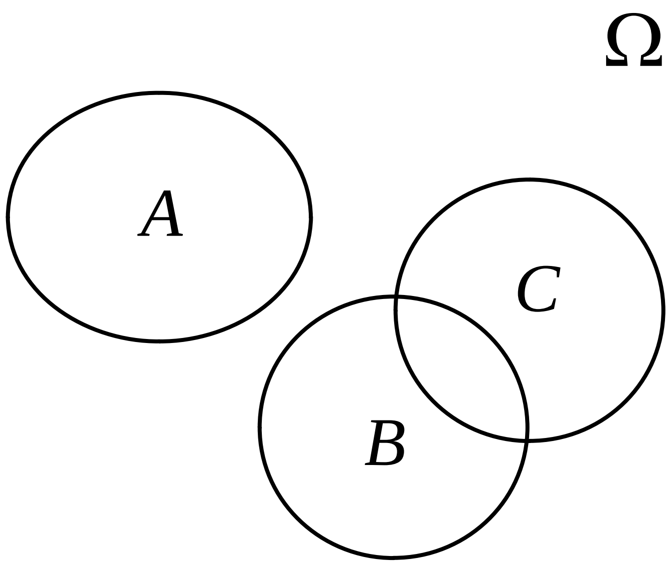
1. Motivation,
2. Probability spaces,
3. Conditional probabilities,
4. Random variables,
5. Probability distributions,
6. Transformations of probability distributions,
7. Conditional distributions,
8. Parametric families of (direct) probability distributions,
9. The Central limit Theorem,
10. Invariant parametric families.

1. Motivation :

- (self-) consistency,
- “Jede axiomatische (abstrakte) Theorie läßt bekanntlicht unbegrentzt viele konkrete Interpretationen zu.” (A. N. Kolmogorov (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung,* Chap. I, p. 1.)

A. Stuart, J. K. Ord (1994), *Kendall’s Advanced Theory of Statistics, Vol. 1 – Distribution Theory*.
M.M.Rao (1993), *Conditional Measures and Applications.*

2. Probability spaces:



$\Omega =$ (abstract, non - empty) universal set

$A, B, C \ldots \subseteq \Omega$

$\Sigma = \{A, B, C, \ldots\}$

Definition 3 (σ-algebra). *Σ is called* σ-algebra *(*σ-field) on Ω *iff:*

i) $\Omega \in \Sigma$,

ii) $\Omega \setminus A;\ A \in \Sigma$,

iii) $\sum_{i=1}^{\infty} A_i \in \Sigma;\ A_i \in \Sigma.$

Definition 4 (Measurable space). *An ordered pair* (Ω, Σ), *consisting of a universal set* Ω *and a* σ*-algebra* Σ *on* Ω *is called a* measurable space.

Example 2 (Measurable space). (Ω,Σ), where $\Sigma=\{\Omega,\emptyset\}$.

Example 3 (Measurable space). $(\mathbb{R}^n,\mathcal{B}^n)$,where $\mathcal{B}^n$ is the Borel σ-algebra (the smallest σ-algebra containing all n-dimensional open rectangles).

Definition 5 (Probability measure). *A real-valued function* P *on a* σ*-algebra* Σ *is called* Probability measure *iff:*

i) $P(A)\geq 0;\ A\in\Sigma$,

ii) $P(\Omega)=1$,

iii) $P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i);\ A_i\cap A_{j\neq i}=\emptyset$.

Definition 6 (Probability space). *An ordered triple* (Ω,Σ,P), *consisting of a universal set* Ω, *of a* σ*-algebra* Σ *on* Ω, *and of a probability (measure)* P *on* Σ *is called* probability space.

3. Conditional probability:

Definition 7 (Conditional probability). *Consider a probability space* (Ω,Σ,P) *and sets* $A,B\in\Sigma$, $P(B)>0$. *Then, the* conditional probability $P(A|B)$ *is defined as*

$$P(A|B) \equiv \frac{P(A\cap B)}{P(B)}.$$

Definition 8 (Independent sets). *Consider a probability space* (Ω,Σ,P) *and sets* $A,B\in\Sigma$. *The sets are (*P*-) independent iff*

$$P(A\cap B) = P(A)\,P(B).$$

Definition 9 (Mutually exclusive and exhaustive sets). *Consider a space* (Ω,Σ,P) *and a set* $\{B_1,\ldots,B_n\}$, $B_i\in\Sigma$ and $P(B_i)>0$. *When*

$$B_i \cap B_{j\neq i} = \varnothing \quad \text{and} \quad \bigcup_{i=1}^{n} B_i = \Omega,$$

the sets B_i *are called* mutually exclusive and $(\Omega$-$)$exhaustive.

Proposition 1 (Law of Total Probability, LTP). *Let* (Ω,Σ,P) *be a probability space and* $\{B_j\}$ *mutually exclusive and* Ω*-exhaustive sets. Then,* $P(A), A\in\Sigma$, *decomposes as*

$$P(A)=\sum_{j=1}^{n} P(B_j)P(A\mid B_j).$$

Proof. $P(A)=P(A\cap\Omega)=P\left(A\cap\left(\bigcup_{j=1}^{n} B_j\right)\right)=P\left(\bigcup_{j=1}^{n}(A\cap B_j)\right)$

$$=\sum_{j=1}^{n} P(A\cap B_j)=\sum_{j=1}^{n} P(B_j)P(A\mid B_j).$$

□

Theorem 1 (Bayes). *Let* (Ω,Σ,P) *be a probability space,* $\{B_j\}$ *mutually exclusive and* Ω*-exhaustive sets, and* $P(A)>0$, $A\in\Sigma$. Then,

$$P(B_i\mid A)=\frac{P(B_i)P(A\mid B_i)}{\sum_{j=1}^{n} P(B_j)P(A\mid B_j)}.$$

Proof. $P(A\cap B_i)=P(B_i)P(A\mid B_i)=P(A)P(B_i\mid A)$ follows from Def. 7, whereas $P(A)$ decomposes according to LTP. □

4. Random variable:

Definition 10 (Scalar random variable). *Given a probability space* (Ω,Σ,P), *let a function* $X:\Omega\to\mathbb{R}$ *be* Σ*-measurable,*

$$A_{X\le x}=\{\omega\in\Omega : X(\omega)\le x\}\in\Sigma\,;\,x\in\mathbb{R}\ .$$

Then, X *is called a (*real-valued, scalar*)* random variable*, while* x *is called a* realization *of* X.

Definition 11 (Random vector). *Given a probability space* (Ω,Σ,P), *let a function* $\mathbf{X}:\Omega\to\mathbb{R}^n$ *be* Σ*-measurable,*

$$A_{\mathbf{X}\le\mathbf{x}}=\{\omega\in\Omega : \mathbf{X}(\omega)\le\mathbf{x}\}\in\Sigma\,;\,\mathbf{x}\in\mathbb{R}^n\ .$$

Then, $\mathbf{X}$ *is called a (*real-valued*)* random vector*, while* $\mathbf{x}$ *is called a* realization *of* $\mathbf{X}$.

Remark 1. *The notion of chance ('physical randomness') is avoided in Defs. 10 and 11.*

5. Probability distributions:

Definition 12 (Probability distribution). *A function* $\Pr_X : \mathfrak{B} \to [0,1]$, *called* probability distribution, *is defined as the image measure of* P *by the random variable* X, $\Pr_X \equiv P \circ X^{-1}$, *such that*

$$\Pr_X(S) = P\left[X^{-1}(S)\right],\ S \in \mathfrak{B} \equiv \mathfrak{B}^1,\ X^{-1}(S) \in \Sigma.$$

Remark 2. Probability vs. Quantum Mechanics.

$$(\mathbb{R}, \mathfrak{B}, \Pr_X) \qquad \psi(x) \equiv \langle x | \Psi \rangle, \langle \psi(x) | \psi(x) \rangle = \int_{\mathbb{R}} \psi^*(x)\psi(x)\,dx = 1$$

$$\uparrow X \qquad \uparrow \hat{X}, |x\rangle; \hat{X}|x\rangle = x|x\rangle$$

$$(\Omega, \Sigma, P) \qquad \Psi, \Phi, \ldots; \langle \Phi | \Psi \rangle; \langle \Psi | \Psi \rangle = 1$$

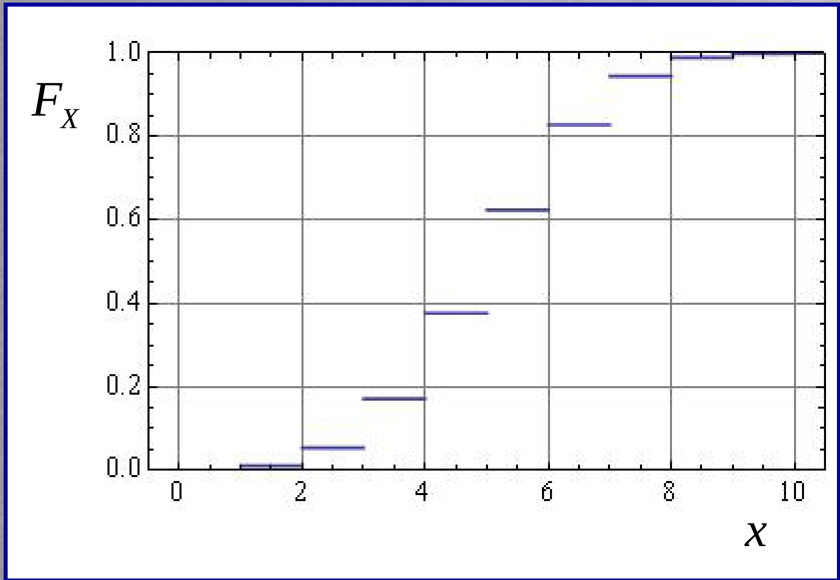
Definition 13 (Cumulative distribution function, cdf). *Given a probability space* (Ω,Σ,P), *a random variable* X, *and a set* $A_{X\leq x}=\{\omega\in\Omega : X(\omega)\leq x\}$, the (cumulative) distribution function $F_X(x):\mathbb{R}\to[0,1]$ *is defined as*

$$F_X(x)\equiv P(A_{X\leq x}).$$

Properties of $F_X(x)$**:**

- non-decreasing,
- $F_X(-\infty)=0,\ F_X(+\infty)=1$,
- $\Pr_X(S)=F_X(x_2)-F_X(x_1);\ S=(x_1,x_2]\in\mathfrak{B}$.

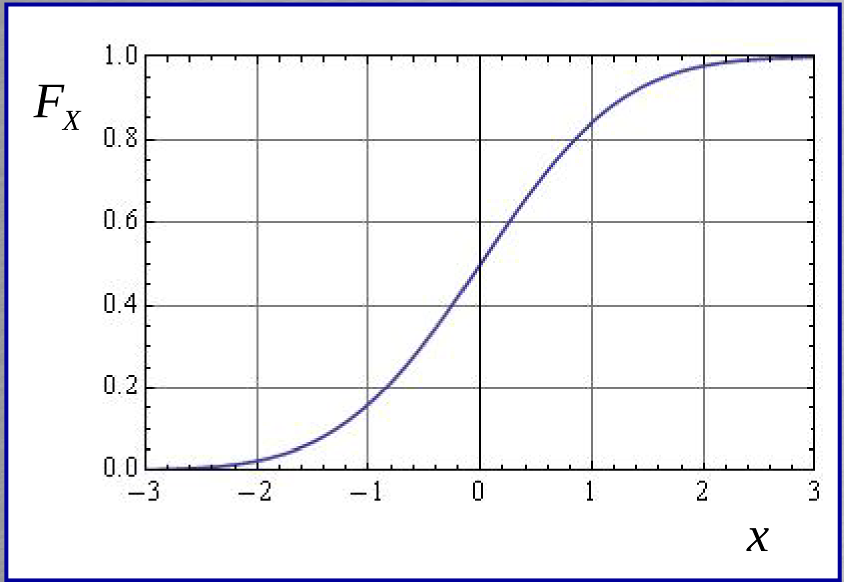
Example 4 (Discrete distribution). $F_X(x)$ discontinuous at $x=x_i$.



Definition 14 (Probability mass function, pmf). *Given a discrete distribution with* $F_X(x)$ *discontinuous at* $x=x_i$*, the* probability mass function $p_X(x)$ *is defined as*

$$p_X(x) \equiv \Pr_X(X = x).$$

Example 5 (Continuous distribution). $F_X(x)$ continuous function.



Definition 15 (Probability density function, pmf). *Cdf of a continuous distribution is expressible as a (Lebesgue) integral of a (non-negative)* probability density function $f_X(x)$,

$$F_X(x) = \int_{-\infty}^{x} f_X(x')dx' .$$

Also, $f_X(x) = dF_X(x)/dx \equiv [F_X(x)]'$ *and* $\mathrm{Pr}_X(S) = \int_S f_X(x)dx;\ S \in \mathfrak{B}.$

Definition 16 (Mean of a distribution).

$$\langle x \rangle \equiv \begin{cases} \int_{\mathbb{R}} x\, f_X(x)\,dx \\ \sum_i x_i\, p(x_i) \end{cases}.$$

Definition 17 (Variance of a distribution).

$$Var(x) \equiv \langle (x - \langle x \rangle)^2 \rangle \equiv \begin{cases} \int_{\mathbb{R}} (x - \langle x \rangle)^2 f_X(x)\,dx \\ \sum_i (x_i - \langle x \rangle)^2 p(x_i) \end{cases}.$$

Definition 18 (Support of a continuous distribution).

$$V_X \equiv \{x \in \mathbb{R} : f(x) > 0\}.$$

Defs. 12 – 18 extend without change to random vectors:

$\Pr_{\mathbf{X}}, F_{\mathbf{X}}(\mathbf{x}), p_{\mathbf{X}}(\mathbf{x}), f_{\mathbf{X}}(\mathbf{x}), \langle \mathbf{x} \rangle, Var(\mathbf{x}), V_{\mathbf{X}}$.

Definition 19 (Independent random variables). *Components of a random vector* $\mathbf{X} = (X_1, X_2)$ *are* called independent random variables *iff*

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1}(x_1) F_{X_2}(x_2).$$

If, in addition, the functions F_{X_1} *and* F_{X_2} *are the same,* X_1 *and* X_2 *are called* independent identically distributed random variables *(*i.i.d.*).*

Definition 20 (Marginal distributions). *Consider a random vector* $\mathbf{X} = (X_1, X_2)$ *. The* marginal *cdf* $F_{\mathbf{X}}(x_1)$ *and pdf* $f_{\mathbf{X}}(x_1)$ *are defined as*

$$F_{\mathbf{X}}(x_1) \equiv F_{\mathbf{X}}(x_1, x_2 = +\infty) \text{ and } f_{\mathbf{X}}(x_1) \equiv \int_{\mathbb{R}} f_{\mathbf{X}}(x_1, x_2) dx_2 .$$

6. Transformations of probability distributions.

Proposition 2 . *Let X and Y be continuous random variables, defined on a probability space (Ω,Σ,P), and let s* be a function on V_X, *such that* $Y = s \circ X$,

$$\left(\mathbb{R},\mathcal{B},\Pr_X\right) \xrightarrow{\ s\ } \left(\mathbb{R},\mathcal{B},\Pr_Y\right)$$

$$\left(\mathbb{R},\mathcal{B},\Pr_X\right) \xleftarrow{\ X\ } \left(\Omega,\Sigma,P\right) \xrightarrow{\ Y\ } \left(\mathbb{R},\mathcal{B},\Pr_Y\right)$$

Let, in addition, $s'(x) \neq 0; \forall x \in V_X$. *Then,*

$$F_Y(y) = \begin{cases} F_X[s^{-1}(y)] & ; s'(x) > 0 \\ 1 - F_X[s^{-1}(y)]; & s'(x) < 0 \end{cases}, \qquad (1)$$

$$f_Y(y) = f_X[s^{-1}(y)]\left|[s^{-1}(y)]'\right| . \qquad (2)$$

Proof. $$A_{Y\le y}=\{\omega\in\Omega:Y(\omega)\le y\}$$
$$=\{\omega\in\Omega:[s\circ X](\omega)\le y\}$$
$$=\{\omega\in\Omega:s[X(\omega)]\le y\}$$
$$=\left\{\omega\in\Omega:\begin{cases}X(\omega)\le s^{-1}(y); s'(x)>0\\ X(\omega)\ge s^{-1}(y); s'(x)<0\end{cases}\right\}$$
$$=\begin{cases}A_{X\le s^{-1}(y)}; s'(x)>0\\ A_{X\ge s^{-1}(y)}; s'(x)<0\end{cases}.$$

□

Eq. (2) extends to random vectors:

$$f_{\mathbf{Y}}(\mathbf{y})=f_{\mathbf{X}}[\mathbf{s}^{-1}(\mathbf{y})]\left|\partial_{\mathbf{y}}\mathbf{s}^{-1}(\mathbf{y})\right|.$$

Example 6 (Convolution).

$$\mathbf{X} = (X_1, X_2),\ f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2);$$

$$\mathbf{X}' \equiv (X_1, S),\ S \equiv X_1 + X_2,\ s \equiv x_1 + x_2\ ;$$

$$f_{\mathbf{X}'}(x_1, s) = f_{\mathbf{X}}(x_1, s - x_1) = f_{X_1}(x_1) f_{X_2}(s - x_1);$$

$$f_{\mathbf{X}'}(s) = \int_{\mathbb{R}} f_{\mathbf{X}'}(x_1, s)\, dx_1 = \int_{\mathbb{R}} f_{X_1}(x_1) f_{X_2}(s - x_1)\, dx_1\ .$$

Properties of convolutions:

1. $\langle s \rangle = \langle x_1 \rangle + \langle x_2 \rangle$;
2. $Var(s) = Var(x_1) + Var(x_2)$;
3. $X_1 \sim N(\mu_1, \sigma_1),\ X_2 \sim N(\mu_2, \sigma_2) \Rightarrow S \sim N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$

7. Conditional probability distributions.

Let: $\mathbf{X} = (Y, Z),\ f_{\mathbf{X}}(y, z);$

$A \equiv (-\infty, y] \times \mathbb{R},\ B \equiv \mathbb{R} \times (z, z+h];\ h > 0,\ \Pr_{\mathbf{X}}(B) = \int_B f_{\mathbf{X}}(y', z')\,dy'\,dz' > 0;$

Recall: $$\Pr_{\mathbf{X}}(A \mid B) = \frac{\Pr_{\mathbf{X}}(A \cap B)}{\Pr_{\mathbf{X}}(B)} = \int_{-\infty}^{y} \left[\frac{\int_z^{z+h} f_{\mathbf{X}}(y', z')\,dz'}{\int_z^{z+h} f_{\mathbf{X}}(z')\,dz'} \right] dy'$$

$\lim_{h \to 0}$:

$$a)\ \int_{-\infty}^{y} \left[\frac{\int_z^{z+h} f_{\mathbf{X}}(y', z')\,dz'}{\int_z^{z+h} f_{\mathbf{X}}(z')\,dz'} \right] dy' \to \int_{-\infty}^{y} \frac{f_{\mathbf{X}}(y', z)}{f_{\mathbf{X}}(z)}\,dy'$$

$$b)\ A \mid B \to (Y, Z) \in (-\infty, y] \times \mathbb{R} \mid (Y, Z) \in \mathbb{R} \times z$$

$$= Y \in (-\infty, y] \mid Z = z$$

$$\Rightarrow \Pr_{\mathbf{X}}(A \mid B) \to F_{\mathbf{X}}(y \mid Z = z) = F_{\mathbf{X}}(y \mid z) = \int_{-\infty}^{y} f_{\mathbf{X}}(y' \mid z)\,dy'$$

$$\Rightarrow f(y \mid z) = \frac{f(y,z)}{f(z)}\,;\; f(z) \equiv \int_{-\infty}^{\infty} f(y,z)\,dy > 0$$

R. Jamnik (1975). § 11, p. 475-477.
I. Kuščer, A. Kodre (1994). § 11.3, p. 281-283.

Borel-Kolmogorov paradox:
A. N. Kolmogorov (1933). *Chap*. V, § 2, p. 44-45.
M. M. Rao (1993). § 3.2, p. 65-66.

Consistent definition (Kolmogorov):
M. M. Rao (1993). § 2.1, pp. 25-26 and pp. 29-30,
§ 2.4, pp. 51-54.

Example 7 (Transformation of a conditional probability distribution).

$$\mathbf{X} = (X_1, X_2),\ f_{\mathbf{X}}(x_1, x_2);$$

$$\mathbf{Y} = \mathbf{s} \circ \mathbf{X} = (s_1 \circ X_1, s_2 \circ X_2) = (Y_1, Y_2);$$

$$[s_{1,2}(x_{1,2})]' \neq 0\,;$$

$$\Rightarrow f_{\mathbf{Y}}(y_1 \mid y_2) = f_{\mathbf{X}}[s_1^{-1}(y_1) \mid s_2^{-1}(y_2)]\left|[s_1^{-1}(y_1)]'\right|.$$

8. Parametric families:

Definition 21 (Parametric family). *The term* parametric family *stands for a collection* $I = \{\Pr_{I,\theta} : \theta \in V_\Theta\}$ *of probability distributions that differ only in the value* θ *of a parameter* Θ.

Example 8 (Reparametrization). Let a probability distribution for a continuous random variable X belong to a family I, $f_X(x) = f_{I,\theta}(x)$, and let $y \equiv s(x)$ and $\lambda \equiv \bar{s}(\theta)$ with $s'(x), \bar{s}'(\theta) \neq 0$. *Then,*

$$f_{I',\lambda}(y) = f_{I,\bar{s}^{-1}(\lambda)}[s^{-1}(y)] \left| [s^{-1}(y)]' \right| .$$

Remark 3. Due to a complete analogy between the transformations in Examples 7 and 8 we define

$$f_I(x \mid \theta) \equiv f_{I,\theta}(x).$$

Example 10 (Families of discrete distributions).

1. Binomial:

$$p_I(n \mid \theta, n_0) = \binom{n_0}{n} \theta^n (1-\theta)^{n_0 - n}\,;\ \theta \in [0,1],\ n_0 \in \mathbb{N},\ n \in \mathbb{N}_0\,;\ n \le n_0\ .$$

2. Poisson:

$$p_I(n \mid \mu) = \frac{\mu^n}{n!} e^{-\mu}\,;\ \mu \in \mathbb{R}^+,\ n \in \mathbb{N}_0\ .$$

3. Multinomial:

$$p_I(n_1, \ldots, n_k \mid \theta_1, \ldots, \theta_k, n_0) = \frac{n_0!}{n_1! \cdots n_k!} \theta_1^{n_1} \cdots \theta_k^{n_k}\,;$$

$$\theta_i \in [0,1],\ n_0 \in \mathbb{N},\ n_i \in \mathbb{N}_0\,;\ k \ge 2\,;\ \sum_{i=1}^{k} \theta_i = 1,\ \sum_{i=1}^{k} n_i = n_0\ .$$

Example 11 (Families of continuous distributions).

$$x, \mu \in \mathbb{R}; \;\; \sigma, \tau \in \mathbb{R}^{+}$$

1. Uniform:

$$f_I(x \mid \mu, \sigma) = \begin{cases} \dfrac{1}{2\sigma}; -1 \le \dfrac{x-\mu}{\sigma} \le 1 \\ 0 \;\; ; \quad \text{otherwise} \end{cases};$$

2. Triangular:

$$f_I(x \mid \mu, \sigma) = \begin{cases} \dfrac{1}{2\sigma}\left(1+\dfrac{x-\mu}{\sigma}\right); -1 \le \dfrac{x-\mu}{\sigma} \le 0 \\ \dfrac{1}{2\sigma}\left(1-\dfrac{x-\mu}{\sigma}\right); 0 < \dfrac{x-\mu}{\sigma} \le 1 \quad ; \\ 0 \;\; ; \quad \text{otherwise} \end{cases}$$

Example 11 (Continued).

3. Normal (Gauss):

$$f_I(x\,|\,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\};\ x\sim N(\mu,\sigma)$$

4. Cauchy:

$$f_I(x\,|\,\mu,\sigma)=\frac{1}{\pi\,\sigma}\left[1+\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-1};$$

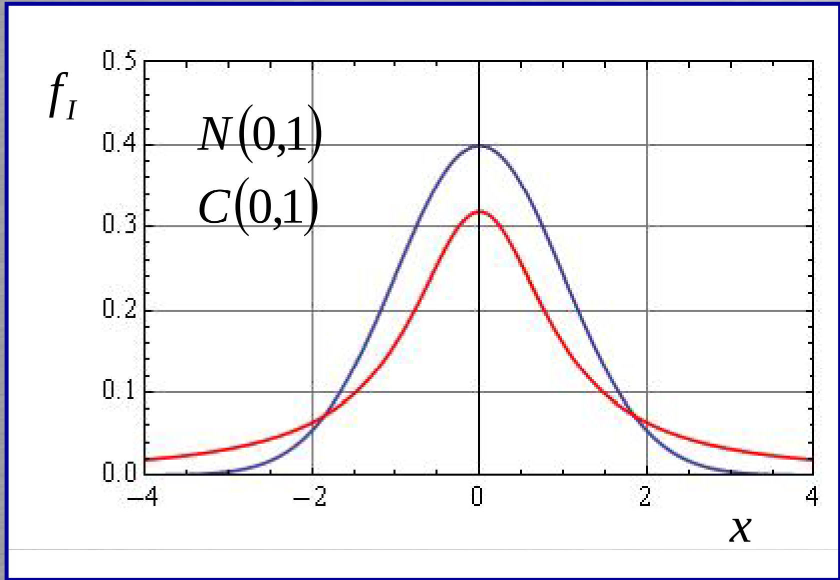
5. Exponential:

$$f_I(t\,|\,\tau)=\frac{1}{\tau}\exp\left\{-\frac{t}{\tau}\right\};\ t\in\mathbb{R}_0^+;$$

6. Weibull:

$$f_I(t\,|\,\tau,\theta)=\frac{\theta}{\tau}\left(\frac{\theta}{\tau}\right)^{\theta-1}\exp\left\{-\left(\frac{t}{\tau}\right)^{\theta}\right\};\ t\in\mathbb{R}_0^+,\ \theta\in\mathbb{R}^+;$$

$$x\equiv\ln t,\ \mu\equiv\ln\tau,\ \sigma\equiv\theta^{-1}.$$



Definition 22 (Location and scale parameters). When

$$F_I(x\,|\,\mu,\sigma) \equiv \int_{-\infty}^{x} f_I(x'\,|\,\mu,\sigma)\,dx' = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

μ *is called* location *and* σ scale parameter*.*

Example 12 (Exponential family revisited).

$$f_I(t\,|\,\tau)=\frac{1}{\tau}\exp\left\{-\frac{t}{\tau}\right\};t\in\mathbb{R}_0^+\Rightarrow F_I(t\,|\,\tau)=1-\exp\left\{-\frac{t}{\tau}\right\},$$

$$x\equiv\ln t,\ \mu\equiv\ln\tau\Rightarrow F_{I'}(x\,|\,\mu)=1-\exp\left\{-e^{x-\mu}\right\}=\Phi(x-\mu).$$

Example 11 (Continued).

7. Multidimensional Normal:

$$f_I(\mathbf{x}\,|\,\boldsymbol{\mu},\mathbf{V})=\frac{1}{(2\pi)^{n/2}\sqrt{\det|\mathbf{V}|}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\};$$

$\mathbf{x},\boldsymbol{\mu}\in\mathbb{R}^n$; $\mathbf{V}$ symmetric, positively definite, $n\times n$ matrix.

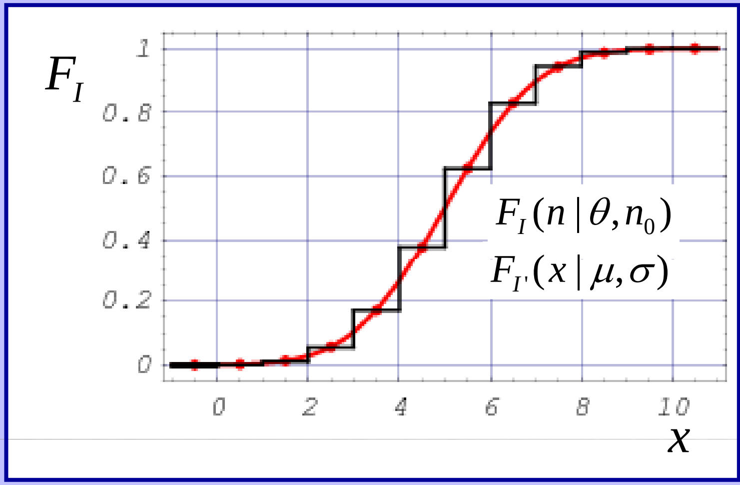
9. The Central Limit Theorem (CLT):

Example 13 (Binomial vs. Normal distribution).

$$p_I(n\,|\,\theta,n_0) = \binom{n_0}{n}\theta^n(1-\theta)^{n_0-n}\,; \qquad \begin{array}{l} n_0 \in \mathbb{N} \\ n \in \mathbb{N}_0 ;\, n \le n_0 \end{array}$$

$$n_0\theta,\, n_0(1-\theta) \gg 1: \quad F_I(n\,|\,\theta,n_0) = \sum_{i=0}^{n} p(i\,|\,\theta,n_0) \simeq \int_{-\infty}^{n+0.5} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x'-\mu)^2}{2\sigma^2}\right\}dx'$$

$$\mu = n_0\theta\,, \quad \sigma = \sqrt{n_0\theta(1-\theta)}$$



Theorem 2 (CLT, Lévy). *Consider i.i.d.* $X_1, \dots, X_n$ *with* $\langle x \rangle = \langle x_i \rangle$ *and* $Var(x) = Var(x_i) < \infty$. *Then,*

$$\lim_{n\to\infty} \bar{x}_n \sim N\left(\langle x \rangle, \sqrt{\frac{Var(x)}{n}} \right), \quad \bar{x}_n \equiv \frac{1}{n} \sum_{i=1}^{n} x_i \, .$$

Proof. By invoking characteristic functions.

(Counter-)Example 14 (CLT vs. Cauchy family).

$$Var(x_i) \textit{ does not exist} \Rightarrow FWHM(\bar{x}_n) = FWHM(x_i),$$
$$\bar{x}_n \sim C\big(\langle x \rangle, FWHM(x)/2\big).$$

10. Invariant families:

Example 15 (Invariance of a location scale family).

$$F_I(x \mid \mu, \sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

$$\left.\begin{aligned} l[(a,b),x] &\equiv ax+b \\ \bar{l}[(a,b),(\mu,\sigma)] &\equiv (a\mu+b, a\sigma) \end{aligned}\right\} (a,b) \in \mathbb{R}^{+} \times \mathbb{R} = G\,,$$

$$F_{I'}\left(l[(a,b),x] \mid \bar{l}[(a,b),(\mu,\sigma)]\right) = F_I\left(l[(a,b),x] \mid \bar{l}[(a,b),(\mu,\sigma)]\right)$$

Example 16 (Invariance under inversion). The Normal and the Cauchy family are invariant under simultaneous inversions of x and μ. We say that they have *positive parity.*

Example 17 (Invariance of the exponential family).

$$F_I(t \mid \tau) = 1 - \exp\left\{-\frac{t}{\tau}\right\},$$

$$\left.\begin{array}{l} l[a,t] \equiv at \\ \bar{l}[a,\tau] \equiv a\tau \end{array}\right\} \; a \in \mathbb{R}^+ = G\,,$$

Proposition 3 (Invariance under Lie group). *A family of probability distributions of a scalar random variable that is invariant under a scalar Lie group is reducible to a location family.*

Text box
With shadow