# Statistical methods

*Tomaž Podobnik*
*Oddelek za fiziko, FMF, UNI-LJ*
*Odsek za eksperimentalno fiziko delcev, IJS*

## Contents:

# 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.

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

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

$$s_n^2 \equiv \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$$

**Proposition.** Consider i.i.d. $\{X_1, \ldots, X_n\}$, and suppose that $\langle x \rangle, \langle x^2 \rangle, \langle x^3 \rangle, \langle x^4 \rangle$ all exist and are finite. Then,

$$\langle s_n^2 \rangle = Var(x) \, .$$

# III. Frequency Interpretation of Probability Distributions:

"In order to make the theory operational, we must introduce a concept of probability that links the mathematics to an external world of measurable phenomena." (A. Stuart, J. K. Ord (1994), § 8.8, p. 290.)

"The most striking achievement of the physical sciences is prediction." (G. Pólya (1954), Chap. XIV, § 4, p. 64.)

"The pure mathematician can do what he pleases, but the applied mathematician must be at least partially sane." (M. Kline (1980). *Mathematics: The Loss of Certainty,* Chap. XIII, p. 285.)

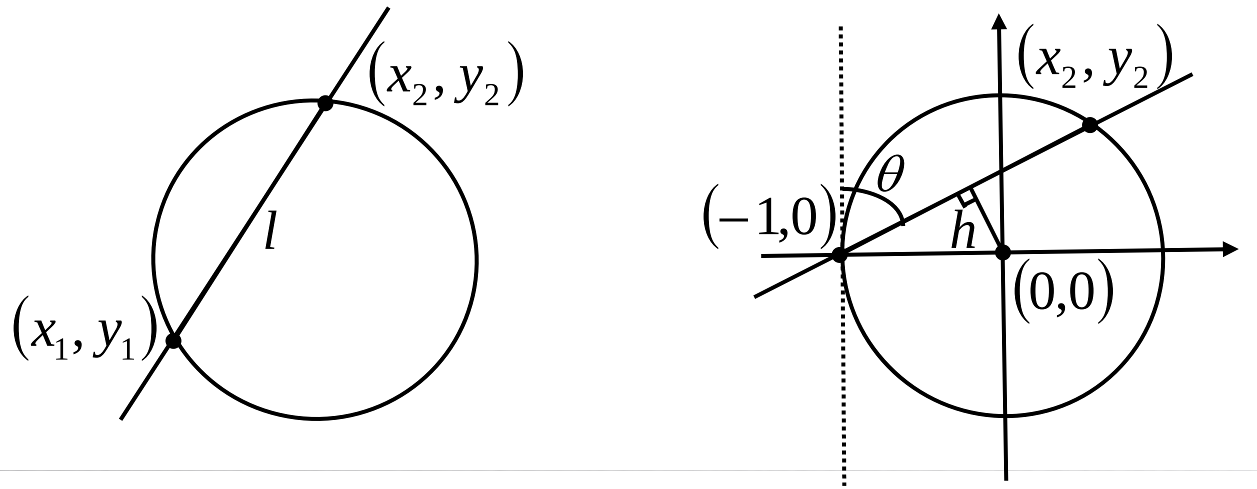
# III. Frequency Interpretation of Probability Distributions:

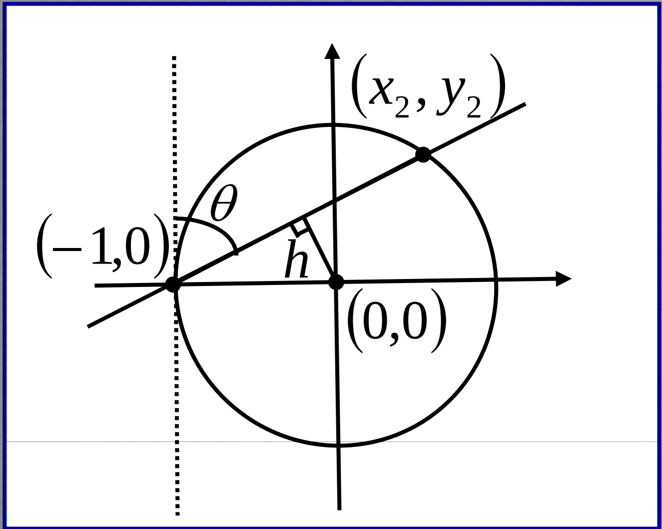
1. Example,
2. Binary random sequence,
3. Random sequence of real numbers,
4. Monte Carlo methods.

# 1. Example 1. (Bertrand's paradox).

A straw is tossed at random so that the line determined by the straw intersects the unit circle. What is the expected length $\langle l \rangle$ of the chord thus defined?

J.L. Bertrand (1889), *Calcul des Probilités*, pp. 4-5.
J.B. Paris (1994), *The Uncertain's Reasoner Companion*, Chap. 6, pp. 71-72.
E.T. Jaynes (2003), *Probability Theory,* § 12.4.4, pp. 386-394.






$$a)\quad f(h)=\begin{cases}1\,;\,0\le h\le 1\\ 0\,;\,\text{otherwise}\end{cases}\quad\Rightarrow\langle l\rangle=\frac{\pi}{2}\approx 1.57\,;$$

$$b)\quad f(\theta)=\begin{cases}\dfrac{2}{\pi}\,;\,0\le\theta\le\dfrac{\pi}{2}\\ 0\,;\,\text{otherwise}\end{cases}\quad\Rightarrow\langle l\rangle=\frac{4}{\pi}\approx 1.27\,;$$

$$c)\quad f(x_2)=\begin{cases}\dfrac{1}{2}\,;\,-1\le x_2\le 1\\ 0\,;\,\text{otherwise}\end{cases}\quad\Rightarrow\langle l\rangle=\frac{4}{3}\approx 1.33\,.$$

## 2. Binary random sequences.

Consider an **infinite** binary sequence $1,0,1,1,0,1,0,0,0,1,0,1,1,0,1,...$ with equal relative frequencies of appearance of 1's and 0's,

$$\nu_1 = \nu_0 = \frac{1}{2};$$

or more precisely,

$$\lim_{n\to\infty} P\left(\left|\frac{n_1}{n} - \frac{1}{2}\right| < \varepsilon\right) = 1.$$

We say that $\nu_1=\nu_0=1/2$ is true almost everywhere with respect to the Bernoulli measure $Bn(1/2)$ on the space of infinite binary sequences, called Cantor space ($Bn(1/2)$ on the Cantor space is isomorphic to the Lebesgue measure on the interval $[0,1]$).

For a $Bn(1/2)$ *-typical* binary sequence we would further expect that

$$\nu_{1,1} = \nu_{1,0} = \nu_{0,1} = \nu_{0,0} = \frac{1}{4},$$

$$\nu_{1,1,1} = \nu_{0,1,1} = \nu_{1,0,1} = \nu_{1,1,0} = \nu_{0,0,1} = \nu_{0,1,0} = \nu_{1,0,0,} = \nu_{0,0,0} = \frac{1}{8},$$

$$\vdots$$

holds $Bn(1/2)$-almost everywhere.

That is, from a $Bn(1/2)$ -typical binary sequence we would naively expect to satisfy all properties true $Bn(1/2)$-almost everywhere. Unfortunately, such a definition is vacuous.

**Definition 1 ($Bn(1/2)$- random binary sequence).** *An infinite binary sequence is called (*Martin-Löf*)* $Bn(1/2)$- random *iff it is not rejected by the Martin-Löf test (i.e., if it satisfies a (special) countable sequence of properties true* $Bn(1/2)$-almost everywhere).

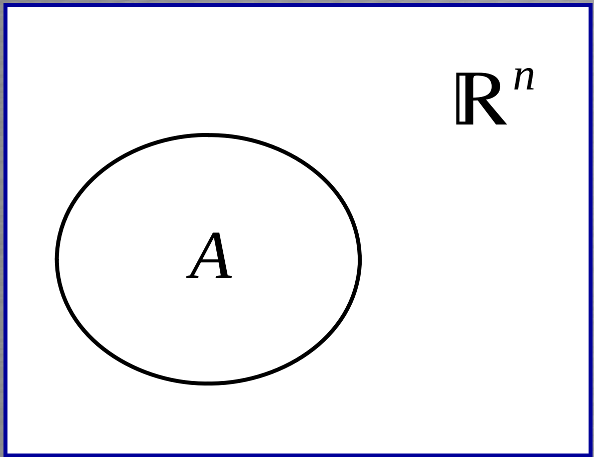
P. Martin-Löf (1966), Inform. Control **9**, 602-619.

The limiting frequencies $\nu_1$ and $\nu_0$ need not be the same, e.g., $\nu_1=2/3$ and $\nu_0=1/3$.

**Definition 2 ($Bn(\nu_1)$- random binary sequence).** *An infinite binary sequence is called (*Martin-Löf*)* $Bn(\nu_1)$- random *iff it is not rejected by the Martin-Löf test (i.e., if it satisfies a countable sequence of properties true* $Bn(\nu_1)$-almost everywhere).

**Remark 1.** *No finite binary sequence is random.*

# 3. Real random sequences.




Given a probability space $(\mathbb{R}^n, \mathcal{B}^n, \Pr_{\mathbf{X}})$, a set $A \in \mathcal{B}^n$ and an infinite sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$, $(\mathbf{x}_i \in \mathbb{R}^n)$ give rise to a binary sequence $b_1, b_2, \ldots$, where

$$b_i = \begin{cases} 1\,;\ \mathbf{x}_i \in A \\ 0\,;\ \text{otherwise} \end{cases}.$$

**Definition 3 ($\Pr_{\mathbf{X}}$-random sequence).** *Given a probability space* $(\mathbb{R}^n, \mathcal{B}^n, \Pr_{\mathbf{X}})$, *an infinite sequence* $\mathbf{x}_1, \mathbf{x}_2, \ldots$, $(\mathbf{x}_i \in \mathbb{R}^n)$ *is* $\Pr_{\mathbf{X}}$-random *iff for every* $A \in \mathcal{B}^n$ *the corresponding binary sequence* $b_1, b_2, \ldots$ *is* $Bn[\Pr_{\mathbf{X}}(A)]$-*random.*

In this way, the probability distribution $\Pr_{\mathbf{X}}$ on $\mathcal{B}^n$ coincides with the (frequency) distribution of the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$, which is characteristic of the frequency interpretation of probability.

**Remark 2.** Every finite sequence is **non**-random. Consequently, the randomness of QM cannot be verified, it can only be postulated.

**Remark 3.** Every (possibly infinite) sequence that results from an algorithm is **non**-random. Consequently, none of the numbers from random number generators, based on algorithms, is truly random. Rather, they are pseudo-random numbers.

There are random number generators based on QM processes such as, for example, radioactive decays. The numbers from these generators may be (parts of) truly random sequences.

# 4. Monte Carlo methods.

**Basis:** Generator of (pseudo-) random numbers, uniformly distributed on an interval, often $[0,1]$.

MC integration:

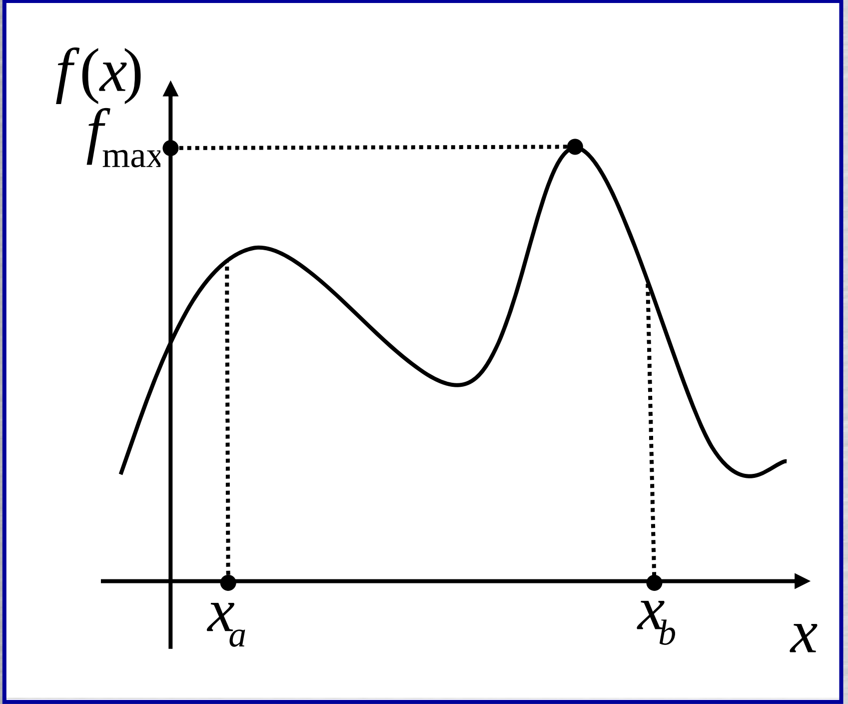
1. $x_i = x_a + \mathrm{rndm}_i \times (x_b - x_a)$

2. $y_i = \mathrm{rndm}'_i \times f_{\max}$

3. $y_i \le f(x_i) \Rightarrow N_{\mathrm{acc}} = N_{\mathrm{acc}} + 1$

– – – – – – – – – – – – – – – – –

$$\int_{x_a}^{x_b} f(x)\,dx = \frac{N_{\mathrm{acc}}}{N_{\mathrm{gen}}} \times (x_b - x_a) \times f_{\max}$$



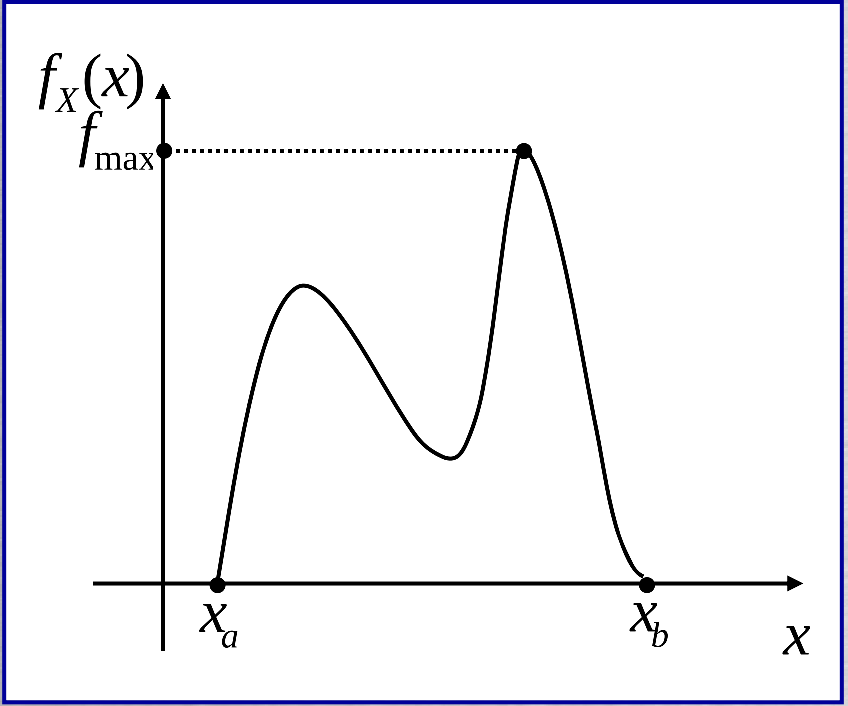
## (Pseudo-) Random numbers for arbitrary $f_X(x)$:

$$V_X = [x_a, x_b]$$

1. $x_i = x_a + \text{rndm}_i \times (x_b - x_a)$
2. $y_i = \text{rndm}'_i \times f_{\max}$
3. $y_i \le f_X(x_i) \Rightarrow \text{accept } x_i$

$$- - - - - - - - - - - - - - - - -$$

accepted $\{x_i\} \sim f_X(x)$



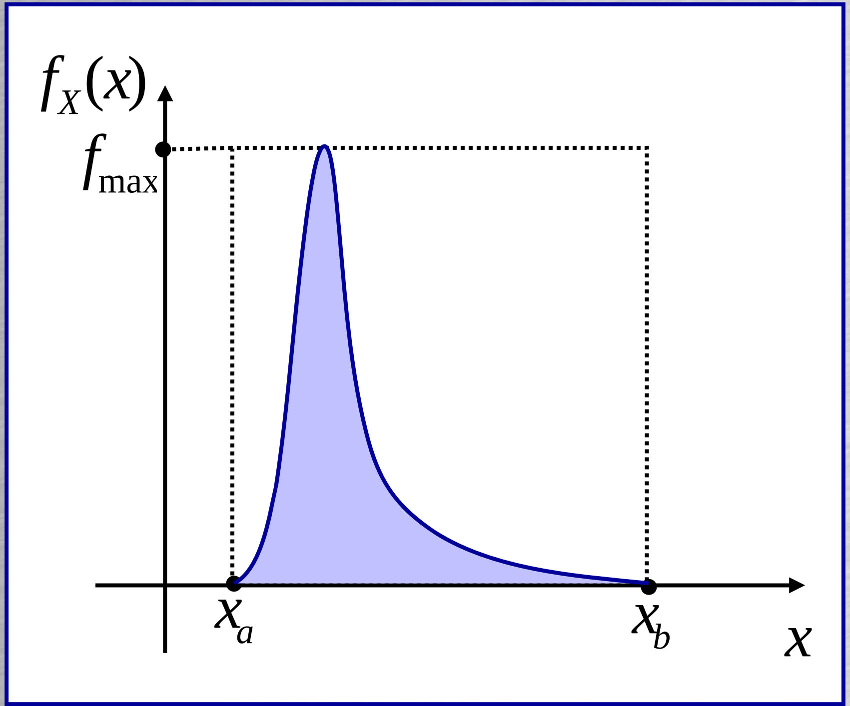
Low efficiencies may represent a serious problem:

$$V_X = [x_a, x_b]$$

1. $x_i = x_a + \mathrm{rndm}_i \times (x_b - x_a)$
2. $y_i = \mathrm{rndm}'_i \times f_{\mathrm{max}}$
3. $y_i \le f_X(x_i) \Rightarrow$ accept $x_i$

$$S_{\mathrm{rec}} = [x_a, x_b] \times f_{\mathrm{max}}$$

$$\frac{N_{\mathrm{acc}}}{N_{\mathrm{gen}}} = \frac{S_{\mathrm{shad}}}{S_{\mathrm{rec}}}$$



Solution if $F_X(x)$ simple (analytic) expression (e.g., for Exponential distr.):

$$y \equiv F_X(x) \Rightarrow f_Y(y) = \begin{cases} 1\,;\, 0 \le y \le 1 \\ 0\,;\, \text{otherwise} \end{cases};$$

$1.\ y_i = \mathrm{rndm}_i$ (100% efficiency)

$2.\ x_i = F_X^{-1}(\mathrm{y_i})$

Solutions for Normal distributions:

a) sum of $n$ uniform i.i.d. variables,

b) 2D Normal distribution……..: $\Rightarrow$

$$\begin{cases} f_{X,Y}(x,y) = f_Y(y)\, f_Y(y) \Rightarrow f_{R,\Phi}(r,\phi) = f_R(r)\, f_\Phi(\phi)\,; \\ f(\phi) = \begin{cases} \dfrac{1}{2\pi};0 \le \phi \le 2\pi \\ 0\ \ ;\text{otherwise} \end{cases},\ f_R(r) = r\exp\{-r^2/2\}\,; \\ \Rightarrow F_R(r) = 1 - \exp\{-r^2/2\}, r \ge 0\,; \Rightarrow z \equiv F_R(r) \\ \Rightarrow f_{Z,\Phi}(z,\phi) = f_Z(z)\, f_\Phi(\phi)\,;\, f_Z(z) = \begin{cases} 1;0 \le z \le 1 \\ 0;\text{otherwise} \end{cases}. \end{cases}$$

# IV. (Classical) Confidence Intervals:

1. Motivation,
2. Construction,
3. Intervals based on likelihood-ratio ordering,
4. Intervals for constrained parameters,
5. Confidence intervals for discrete distributions,
6. On the shortest confidence intervals.

# 1. Motivation.

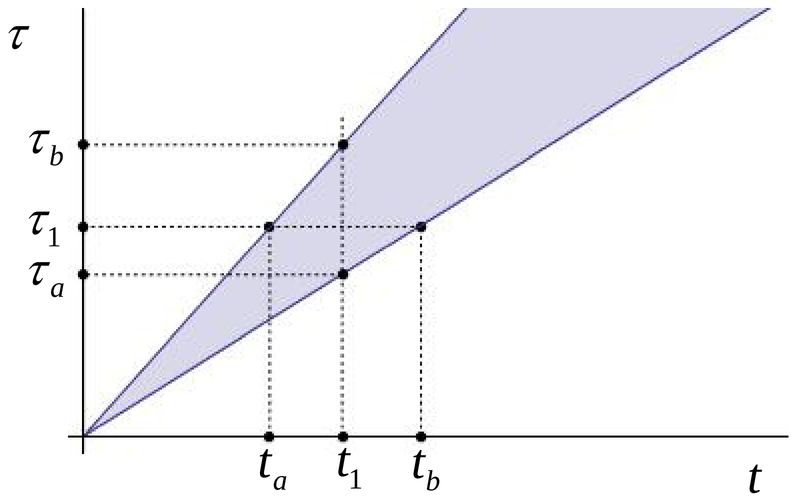
**Example 1 (Prolog):** given $t_1$, can we say anything about $\tau$ ?

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

For verifiable predictions we must turn to interval estimations.

J. Neyman (1937). Phil. Trans. R. Soc., **A 236**, 333-380.

## 2. Construction.




1) $\alpha \in [0, 1-\gamma]$

2) $\tau_1$

3) $t_a : F_I(t_a \mid \tau_1) = \alpha$

4) $t_b : F_I(t_b \mid \tau_1) = \alpha + \gamma$

$\Rightarrow \Pr_I(t_a < t \le t_b \mid \tau)$

$= F_I(t_b \mid \tau_1) - F_I(t_a \mid \tau_1)$

$= \gamma$

5) $\tau \in (0, \infty)$

6) $t_1$; $\tau_1$ true value

7) $\tau_1 \in (\tau_a, \tau_b) \Leftrightarrow t_1 \in (t_a, t_b)$

**Remark 4.** $$\left.\begin{aligned} F_I(t_1 \mid \tau_b) &= \alpha \\ F_I(t_1 \mid \tau_a) &= \alpha + \gamma \end{aligned}\right\} \Rightarrow \gamma = F_I(t_1 \mid \tau_a) - F_I(t_1 \mid \tau_b).$$

**Remark 5.** $\alpha = \alpha(\tau)$, as long as $t_a(\tau)$ and $t_b(\tau)$ strictly monotone in $\tau$.

# 3. Confidence intervals based on likelihood-ratio ordering.

G.J. Feldman, R.D. Cousins (1998), Phys. Rev. **D 57**, 3873-3889.

May be regarded as a definition of $\alpha(\tau)$ (of $\alpha(\theta)$).

Given $\theta : R(x,\theta) \equiv \dfrac{f_I(x\,|\,\theta)}{f_I(\hat{x}\,|\,\theta)}$ ; $x = \hat{x} : f_I(x\,|\,\theta) = \text{max.}$,

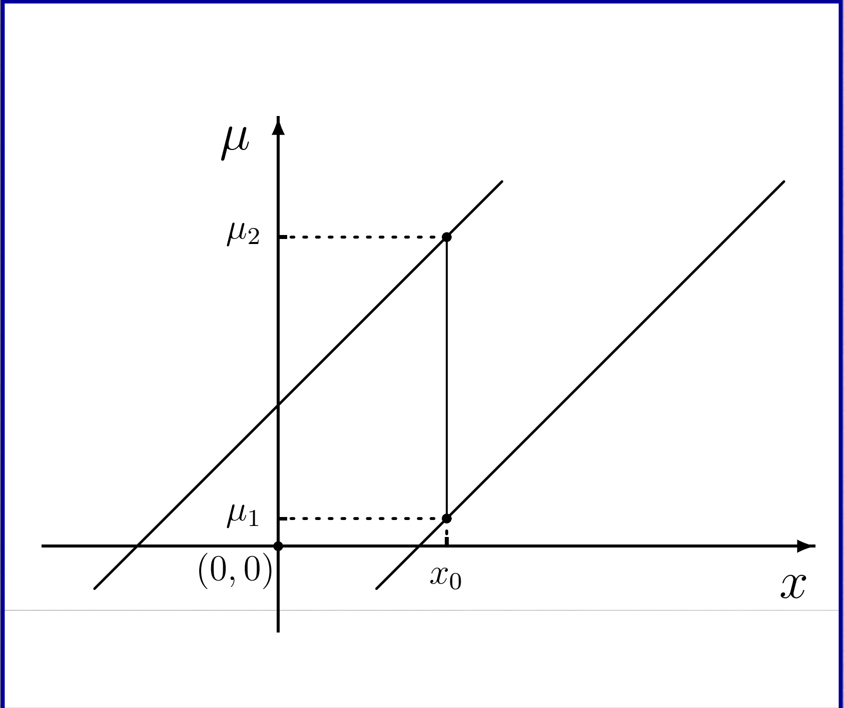
$$A = (x_a, x_b) = \{x \in V_X : R(x,\theta) \geq R_0 \text{ and } \Pr_X(A\,|\,\theta) = \gamma\}.$$

The rest of the procedure is identical to the one on the previous slide.

**Remark 6.** These confidence intervals are equivariant under ono - to - one reparametrizations $y \equiv s(x)$ and $\nu \equiv \bar{s}(\theta)$:

$$(\nu_a[s(x_1)], \nu_b[s(x_1)]) = (\bar{s}[\theta_a(x_1)], \bar{s}[\theta_b(x_1)]).$$

## **Example 2.** Confidence intervals for $x \sim N(\mu, \sigma = 1)$:




$\gamma = 0.9$

a) $\alpha = 0.05 = \text{const.}$,

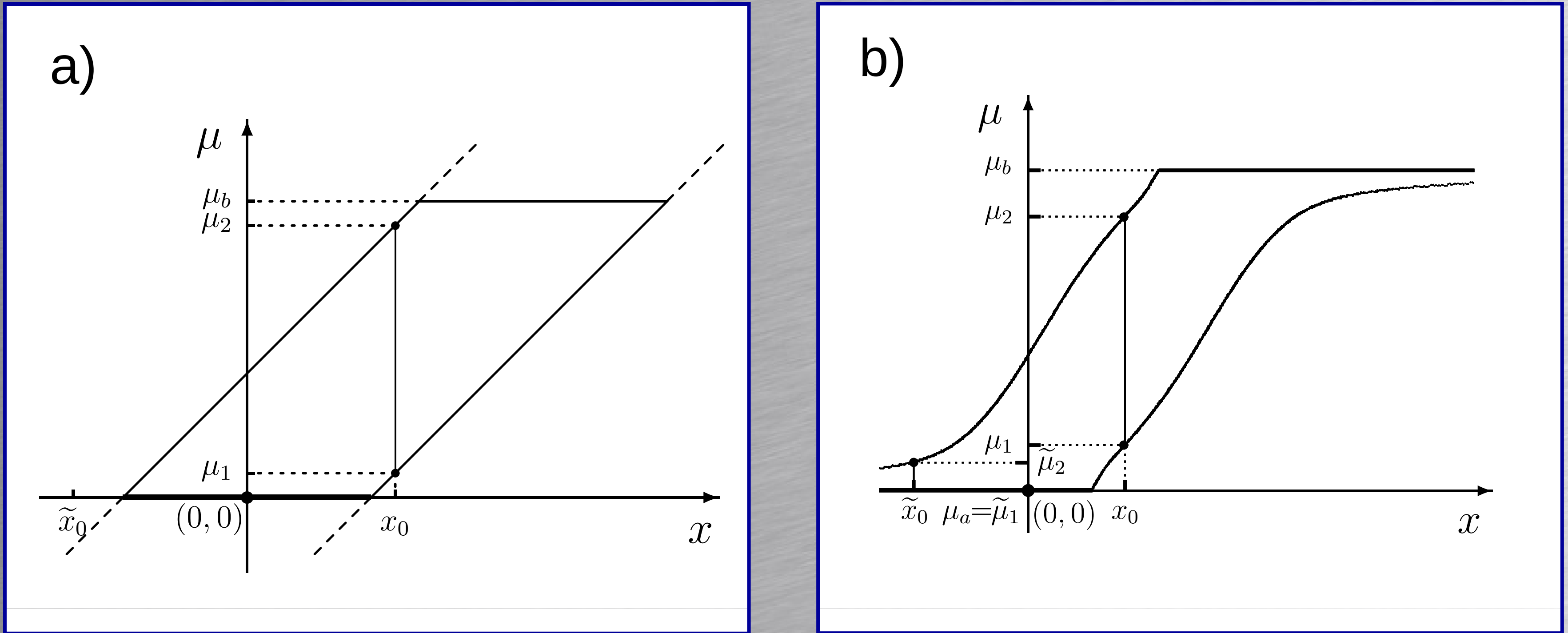
b) intervals from the likelihood - ratio ordering principle.

Important because of CLT :

$$n \to \infty : \bar{x}_n \sim N\left(\langle x \rangle, \sqrt{\frac{Var(x)}{n}}\right) \approx N\left(\langle x \rangle, \sqrt{\frac{s_n^2}{n}}\right)$$

## 4. Intervals for constrained parameters.

**Example 3. Confidence intervals for** $x \sim N(0 \le \mu \le 3.92, \sigma = 1)$:



## 5. Confidence intervals for discrete distributions.

For discrete $F_I(n \mid \mu)$, equations

$$F_I(n_a \mid \mu) = \alpha\,,\; F_I(n_b \mid \mu) = \alpha + \gamma$$

do not have solutions for all $\mu$.

Construct the shortest CI's whose coverage $\geq \gamma$ (conservative CI's).

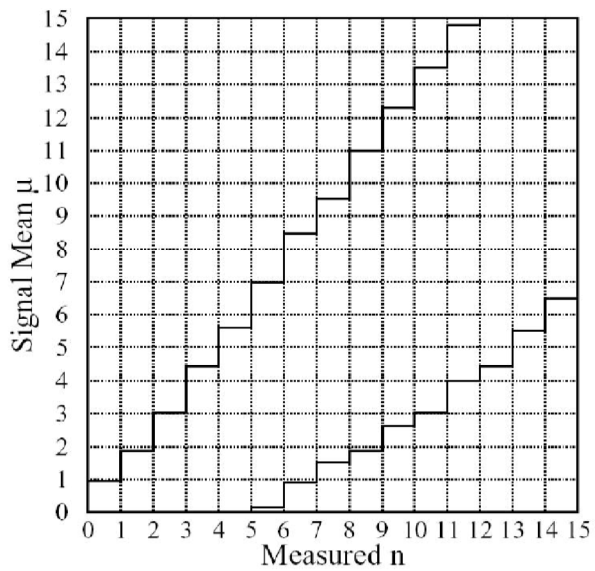



FIG. 7. Confidence belt based on our ordering principle, for 90% C.L. confidence intervals for unknown Poisson signal mean $\mu$ in the presence of Poisson background with known mean $b = 3.0$.

# 6. On the shortest confidence intervals.

**Example 4 (Exponential family):** Given $t_1$, chose $\alpha = \text{const.}$ such that the lenght of the confidence interval $(\tau_a, \tau_b)$ will be minimal.

For $\gamma = 0.2 : \alpha = 0.740976 \Rightarrow (\tau_a, \tau_b) = (0.353 t_1, 0.740 t_1)$.

**Note :** $(\tau_a, \tau_b)$ does not contain $\hat{\tau} = t_1$, but contains $t_1 / 2$.

**Example 4 (cont'd):** $x \equiv \ln x, \mu \equiv \ln \tau \Rightarrow f_{I'}(x \mid \mu) = e^{x-\mu} \exp\{-e^{x-\mu}\}$.

For $\gamma = 0.2 : \alpha = 0.527573 \Rightarrow (\mu_a, \mu_b) = (\ln t_1 - 0.263, \ln t_1 + 0.288)$.

**Note :** $(\mu_a, \mu_b) \neq (\ln \tau_a, \ln \tau_b)$.

**Example 5 (Hypothesis testing):** $H : \tau = \tau_1$. $H$ rejected at confidence (significance) level $\gamma$ if $\tau_1$ is outside the shortest confidence interval $(\tau_a, \tau_b)$ whose coverage is $\gamma$.

The choice of parametrization depends on what you want, accept or reject $H$ (ideology!).

## V. Hypothesis testing:

1. Basic definitions,
2. Errors of the first and the second kind,
3. Neyman-Pearson Lemma,
4. Uniformly most powerful tests.

# 1. Basic definitions.

H. Frank, S.C. Althoen (1994), *Statistics*, Chaps. 9-11, pp. 326-480.

**Inference :** what is the value of parameter $\theta$?

**Test of a hypothesis :** is $\theta_0$ the value of parameter $\theta$?

**Test (null) hypothesis** $H_0 : \theta = \theta_0$.

**Alternative hypothesis** $H_1 : \theta = \theta_1$ $(\theta_1 \neq \theta_0)$ or $\theta > \theta_0$.

**Test** $W$ **:** a numerical index that is expected to take the value $w_0$ if $H_0$ is correct and is expected some other value if $H_1$ is correct.

**Test statistic** $W$ **:** $W = W(x_1, x_{2,} \ldots)$ .

**Rejection (critical) region** $R_C$ **:** the region of the values of a test $W$ that are unlikely (??) if $H_0$ is correct but relatively likely (??) if $H_1$ is correct.

**Critical value (significance level, size of** $R_C$**) :** $\alpha \equiv \Pr_I(R_C \mid \theta_0)$ .

**Confidence level** $Cl \equiv 1-\alpha$ .

**Decision :** if $w \in R_C$ abandon $H_0$ in favor of $H_1$ at confidence level $Cl$;
if $w \notin R_C$, $H_1$ is rejected.

## 2. Errors of the first and second kind.

**Error I (false positive):** rejecting $H_0$ when it is correct.

**Error II (false negative):** rejecting $H_1$ when it is correct (i.e., fail to reject $H_0$ when it is indeed false).

$\alpha = P(\text{Error I})$, $\beta \equiv P(\text{Error II})$.

**Power of the test:** $1-\beta$ (probability that a test will reject a false $H_0$).

## 3. Neyman-Pearson Lemma.

**A best** $R_C$ **of size** $\alpha$ **:** $P(R_C \mid H_0) = \alpha$ , $P(R_C \mid H_1) \geq P(Q_C \mid H_1)$
for all $Q_C$ for which $P(Q_C \mid H_0) = \alpha$ .

**Lemma 1 (Neyman - Pearson).** $H_0 : \theta = \theta_0,\ H_1 : \theta = \theta_1$ ,

$$W(x_1,\ldots,x_n;\theta_0,\theta_1) = \frac{L(x_1,\ldots,x_n;\theta_0)}{L(x_1,\ldots,x_n;\theta_1)},$$

$$R_C = \{(x_1,\ldots,x_n) : W(x_1,\ldots,x_n;\theta_0,\theta_1) \leq \eta\}$$

$\Rightarrow R_C$ is a best critical region of size $\alpha$, i.e.,
$W(x_1,\ldots,x_n;\theta_0,\theta_1)$ is a **most powerful test**
of size $\alpha = P(R_C \mid H_0)$.

## 4. Uniformly most powerful test.

**Uniformly most powerful test of size $\alpha$.**

$H_0 : \theta = \theta_0, H_1 : \theta \in A; (\theta_0 \notin A)$,

$W\left(x_1, \ldots, x_n; \theta_0, \theta_1\right)$ most powerfull for all $\theta_1 \in A$

$\Rightarrow W\left(x_1, \ldots, x_n; \theta_0, \theta_1\right)$ is a uniformly most powerful test of size $\alpha$ for alternatives in $A$.

**Example 6 (UMPT for Normal distr.) :** $\{X_1, \ldots, X_n\}$ i.i.d., $X_i \sim N(\mu, 1)$,

$$H_0 : \mu = \mu_0, H_1 : \mu > \mu_0 \Rightarrow \frac{L(x_1, \ldots, x_n; \mu_0)}{L(x_1, \ldots, x_n; \mu_1)}; \mu_1 > \mu_0, \text{ is UMPT.}$$

**Example 7 :** There is no UMPT if $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$.

Text box
With shadow