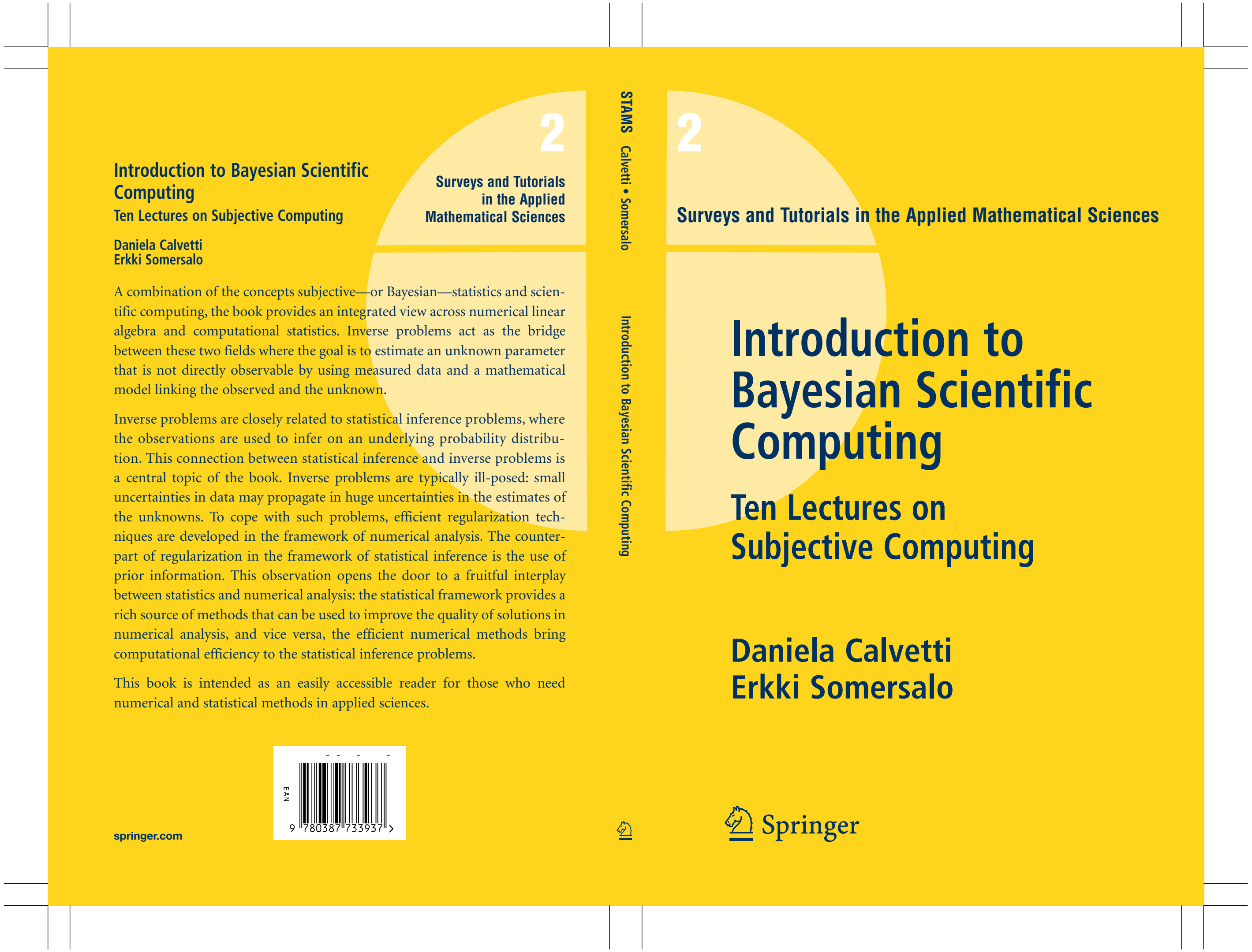
MATERIAL BASED ON...



Inverse Problems and Subjective Computing

Subjective computing = subjective probability + scientific computing

- Inverse problems: Concerns the problem of retrieving information of unknown parameters by indirect observations.
- Statistical inference: Concerns the problem of inferring properties of an unknown distribution from data generated from that distribution.

Quantity is unknown $\longrightarrow$ information incomplete $\longrightarrow$ random variables $\longrightarrow$ probability distributions

Why Statistics?

“Statistics is the science of information gathering, especially when the information arrives in little pieces rather than in one or two big pieces.”

(Bradley Efron)

“Probability is common sense reduced to calculations.”

(Pierre-Simon Laplace, 1813)

BAYESIAN PERSPECTIVE TO INVERSE PROBLEMS

- All unknowns are modelled as random variables.
- Randomness is an expression of the *lack of information*, or *ignorance* of their values.
- Random variables are characterized by their probability distributions.
- INVERSE PROBLEM: Find the probability distribution of the unknowns you are interested in.

NOTE 1: Randomness is not the object's but the *subject's* property[1].

NOTE 2: Computational models predict only observables, i.e., what an observer (subject) can expect.[2]

[1] Bruno de Finetti: "Probability does not exist!"

[2] Niels Bohr: "It is a mistake to think that physics should reveal how the nature is made. Physics deals with what can be said about the nature."

Tossing a coin

Obvious solution:

$$\mathrm{P}(\text{heads}) = \mathrm{P}(\text{tails}) = \frac{1}{2}.$$

No particular justification needed, since generally accepted.

Any competing theory, e.g.,

$$\mathrm{P}(\text{heads}) = 0.6, \quad \mathrm{P}(\text{tails}) = 0.4,$$

falsifiable by empirical evidence.

Cf. testing of scientific theories!

$$\text{Theory} \neq \text{Reality (whatever it means!)}$$

Frequentist statistics

- Probability of an event is its relative frequency of occurrence in an asymptotically infinite series of repeated experiments.

Leaves completely out non-repeatable events.

Example: “The probability of rain tomorrow is 0.7.”

Such statement may be based, *but need not to be* on previous experiments.

De Finetti’s critique, *coherence* and *exchangability*.

Coherence

Betting argument, simplest form:

- Two players, P1 and P2
- Random event, outcomes "A" and "B" possible.
- Winner gets 1$ from the loser.
- P1 decides how much it costs to bet for "A" and for "B".
- P2 decides which side P1 has to take.

P1 has to decide the prices so that he feels comfortable which ever way P2 decides. (*Dutch book argument.*)

Model based probabilities

Example: Tossing of a thumbtack: Which way does it end?

Possible models:

1. Convex hull is a cone. Surface of the cone $S_1 = \pi R\sqrt{R^2 + H^2}$. Surface of the bottom $S_2 = \pi R^2$.

$$\text{Odds of ending on bottom} = \frac{S_2}{S_1} = \frac{R}{\sqrt{R^2 + H^2}}.$$

2. Or maybe the center of mass counts? Consider stability!

3. Does the surface material count?

Judgemental part: Which model do we trust?

Computational frequentism

Example: Growth of bacteria in a petri dish.

Set a probability for having an average bacteria density above a given level.

Stochastic growth model (a life game):

1. Give a probability for the culture to spread from an occupied square to a neighboring empty square.
2. Give a probability for death in a square surrounded by occupied squares (competition).

Simulate: Create a sample based on the model with different initial states.

OBJECTIVE CHANCE

Examples where the probabilities are unarguably (?) set:

- tossing a fair die (by definition od "fair")
- Urn models (white and black balls in an urn, by definition of "random pick")
- Quantum mechanics, quantum information: quantum state does not exist if not observed (compare with coin tossing: the coin is not in a "half heads-half tails" state.) No mess with "multiverses", Schrödinger cats or other useless speculation.

THROUGH THE FORMAL THEORY, LIGHTLY

A. N. Kolmogorov, founding father of probability *theory* (cf. Laplace)

Define Ω to be a *probability space* equipped with a *probability measure* P that measures the probability of *events* $E \subset \Omega$.

We require that

$$0 \leq \mathrm{P}(E) \leq 1.$$

If $A \cap B = \emptyset$, $A, B \subset \Omega$,

$$P(A \cup B) = P(A) + P(B).$$

Since Ω contains all events,

$$\mathrm{P}(\Omega) = 1, \quad (\text{“something happens”})$$

and

$$\mathrm{P}(\emptyset) = 0. \quad (\text{“nothing happens”})$$

Independency, conditional probability

Two events A and B are *independent*, if

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

Conditional probability: Probability that A happens *provided* that B happens,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

For independent events,

$$P(A \mid B) = P(A).$$

The (abstract) probability space Ω is almost never constructed in practice.

Random variables

A real valued random variable X is a mapping

$$X : \Omega \rightarrow \mathbb{R}.$$

We call $x = X(\omega)$, $\omega \in \Omega$, a *realization* of X.

Probability distribution: For $B \subset \mathbb{R}$,

$$\mu_X(B) = \mathrm{P}(X^{-1}(B)) = \mathrm{P}\{X(\omega) \in B\}.$$

Probability density

$$\mu_X(B) = \int_B \pi_X(x)dx.$$

Often, we write simply

$$\pi_X(x) = \pi(x).$$

Expectation, Variance

Given a random variable X, the *expectation* is the center of mass of the probability distribution,

$$\mathrm{E}\{X\} = \int_{\mathbb{R}} x\pi(x)dx = \overline{x}.$$

The *variance* is the expectation of the squared deviation from the expectation,

$$\mathrm{var}\big(X\big) = \mathrm{E}\{(X - \overline{x})^2\} = \int_{\mathbb{R}} (x - \overline{x})^2 \pi(x)dx.$$

The kth moment is defined as

$$\mathrm{E}\{(X - \overline{x})^k\} = \int_{\mathbb{R}} (x - \overline{x})^k \pi(x)dx.$$

The third moment ($k = 3$) is called the *skewness* and the fourth ($k = 4$) is the *kurtosis* of the density.

EXAMPLE

The expectation of a random variable, in spite of its name, is not necessarily the value that we should expect a realization to have.

Let $\Omega = [-1, 1]$, and

$$\mathrm{P}(I) = \frac{1}{2}\int_I dx = \frac{1}{2}|I|, \quad I \subset [-1, 1].$$

Random variables

$$X_1 : [-1, 1] \to \mathbb{R}, \qquad X_1(\omega) = 1 \qquad \forall \omega \in \mathbb{R},$$

and

$$X_2 : [-1, 1] \to \mathbb{R}, \qquad X_2(\omega) = \begin{cases} 2 & \omega \geq 0 \\ 0 & \omega < 0 \end{cases}$$

It is immediate to check that

$$\mathrm{E}\{X_1\} = \mathrm{E}\{X_2\} = 1,$$

although $X_2(\omega) \neq 1$ always.

Covariance, Correlation

Consider two random variables $X, Y : \Omega \to \mathbb{R}$.

Joint probability density

$$\mathrm{P}\{X \in A, Y \in B\} = \mathrm{P}\big(X^{-1}(A) \cap Y^{-1}(B)\big) = \int\int_{A\times B} \pi(x,y)dxdy.$$

The random variables X and Y are *independent* if

$$\pi(x,y) = \pi(x)\pi(y).$$

The *covariance* of X and Y is the *mixed central moment*

$$\mathrm{cov}(X,Y) = \mathrm{E}\big\{(X-\bar{x})(Y-\bar{y})\big\}.$$

It is straightforward to verify that

$$\mathrm{cov}\big(X,Y\big) = \mathrm{E}\{XY\} - \mathrm{E}\{X\}\mathrm{E}\{Y\}.$$

The *correlation* of X and Y is

$$\operatorname{corr}(X,Y) = \mathrm{E}\{XY\}.$$

The *correlation coefficient* of X and Y is

$$\operatorname{corrc}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}}},$$

or, equivalently, the correlation of the centered normalized random variables

$$\tilde{X} = \frac{X - \bar{x}}{\sigma_{\mathrm{X}}}, \qquad \tilde{Y} = \frac{Y - \bar{y}}{\sigma_{\mathrm{Y}}}.$$

It is an easy exercise to verify that

$$\mathrm{E}\{\tilde{X}\} = \mathrm{E}\{\tilde{Y}\} = 0, \quad \operatorname{var}(\tilde{X}) = \operatorname{var}(\tilde{Y}) = 1.$$

The random variables X and Y are *uncorrelated* if their correlation coefficient is zero, i.e.,

$$\mathrm{cov}\big(X, Y\big) = 0$$

If X and Y independent, they are uncorrelated:

$$\mathrm{E}\big\{(X - \overline{x})(Y - \overline{y})\big\} = \mathrm{E}\big\{X - \overline{x}\big\}\mathrm{E}\big\{Y - \overline{y}\big\} = 0,$$

vice versa is not necessarily the case.

X and Y are *orthogonal* if

$$\mathrm{E}\big\{XY\big\} = 0.$$

In that case

$$\mathrm{E}\big\{(X + Y)^2\big\} = \mathrm{E}\big\{X^2\big\} + \mathrm{E}\big\{Y^2\big\}.$$

Marginal denisty

X and Y with joint probability density $\pi(x, y)$

Probability density of X when Y may take *any value*:

$$\pi(x) = \mathrm{P}\{X = x\} = \int_{\mathbb{R}} \pi(x, y) dy.$$

Analogously,

$$\pi(y) = \mathrm{P}\{Y = y\} = \int_{\mathbb{R}} \pi(x, y) dx.$$

Conditional probability density

X and Y with joint probability density $\pi(x, y)$.

$$\pi(x \mid y) = \frac{\pi(x, y)}{\pi(y)}, \quad \pi(y) \neq 0.$$

This is the probability density of X assuming that $Y = y$.

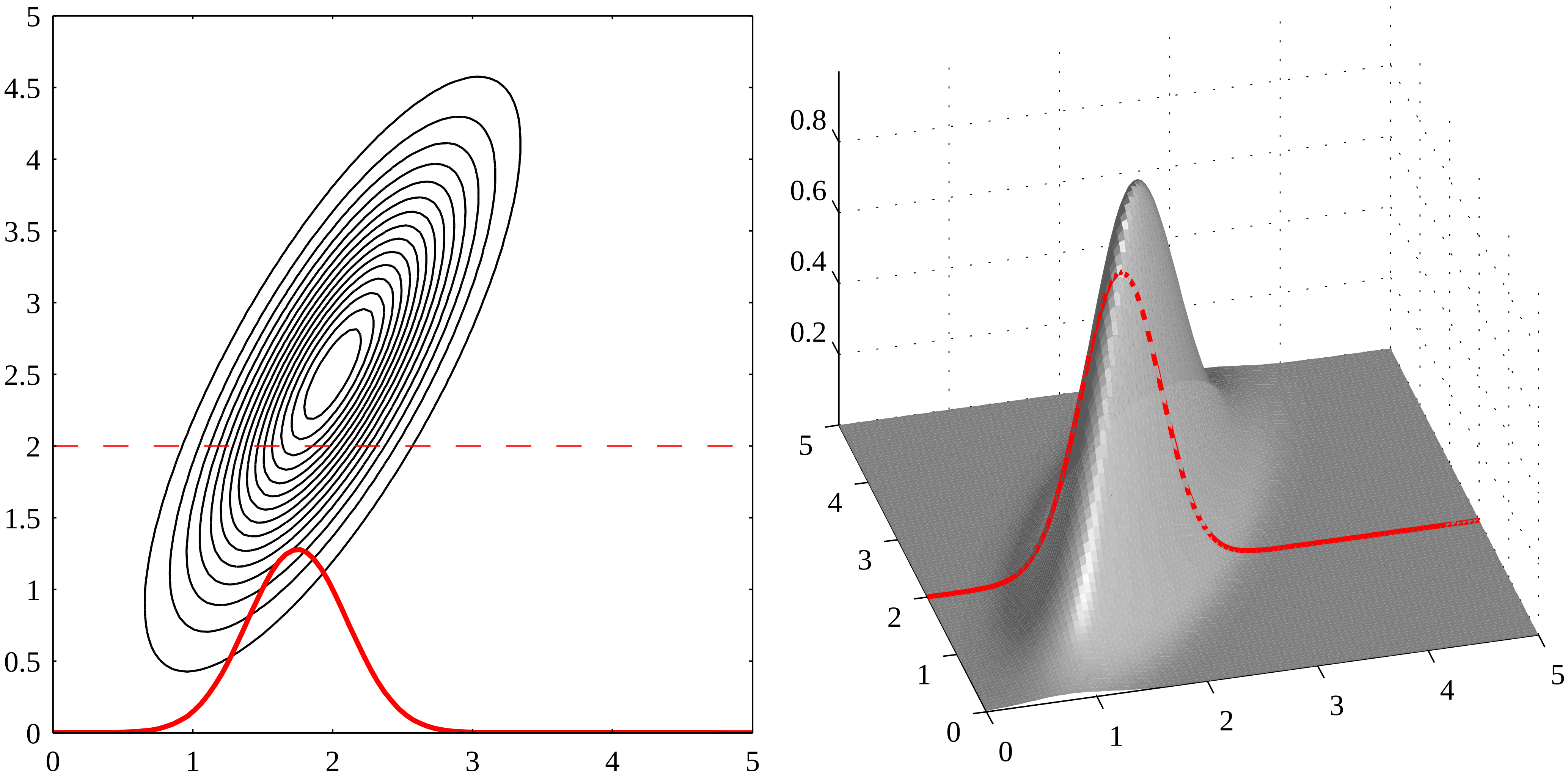
Important identity:

$$\pi(x, y) = \pi(x \mid y)\pi(y) = \pi(y \mid x)\pi(x).$$

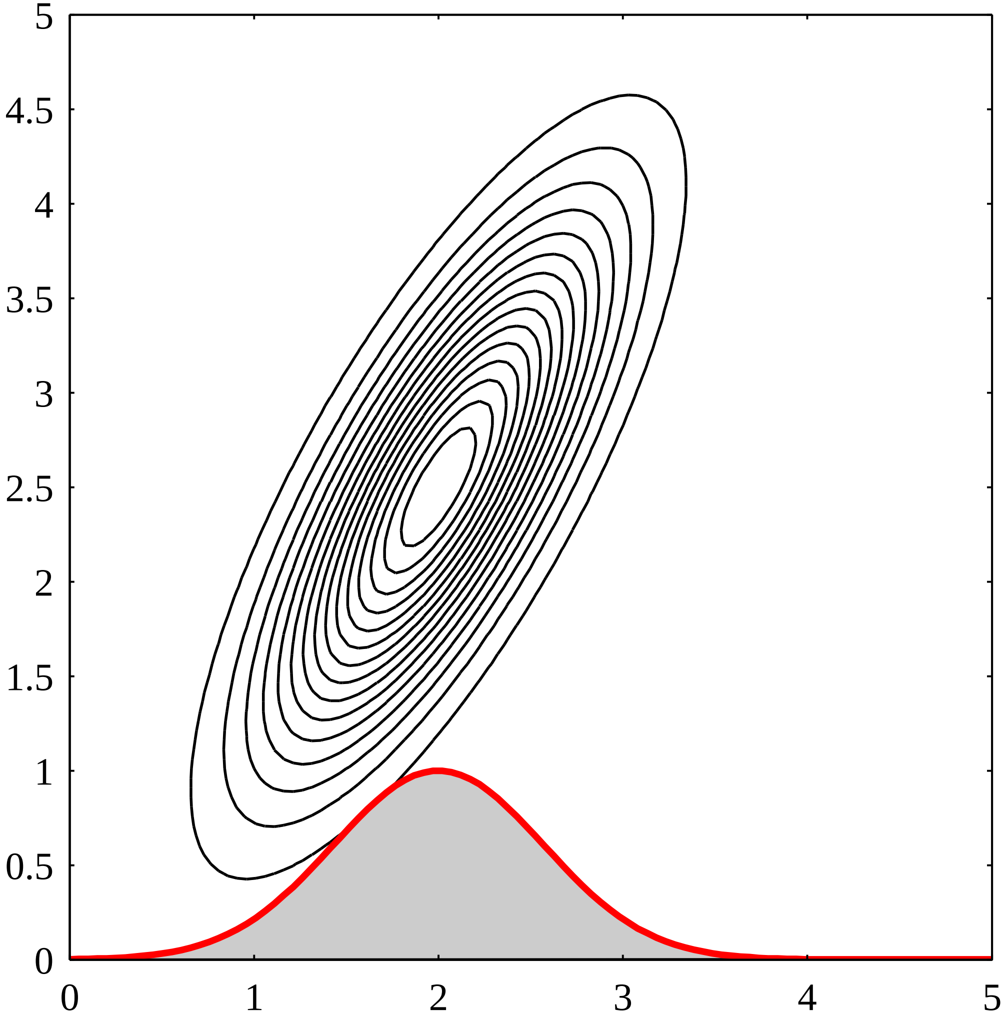
Implication:

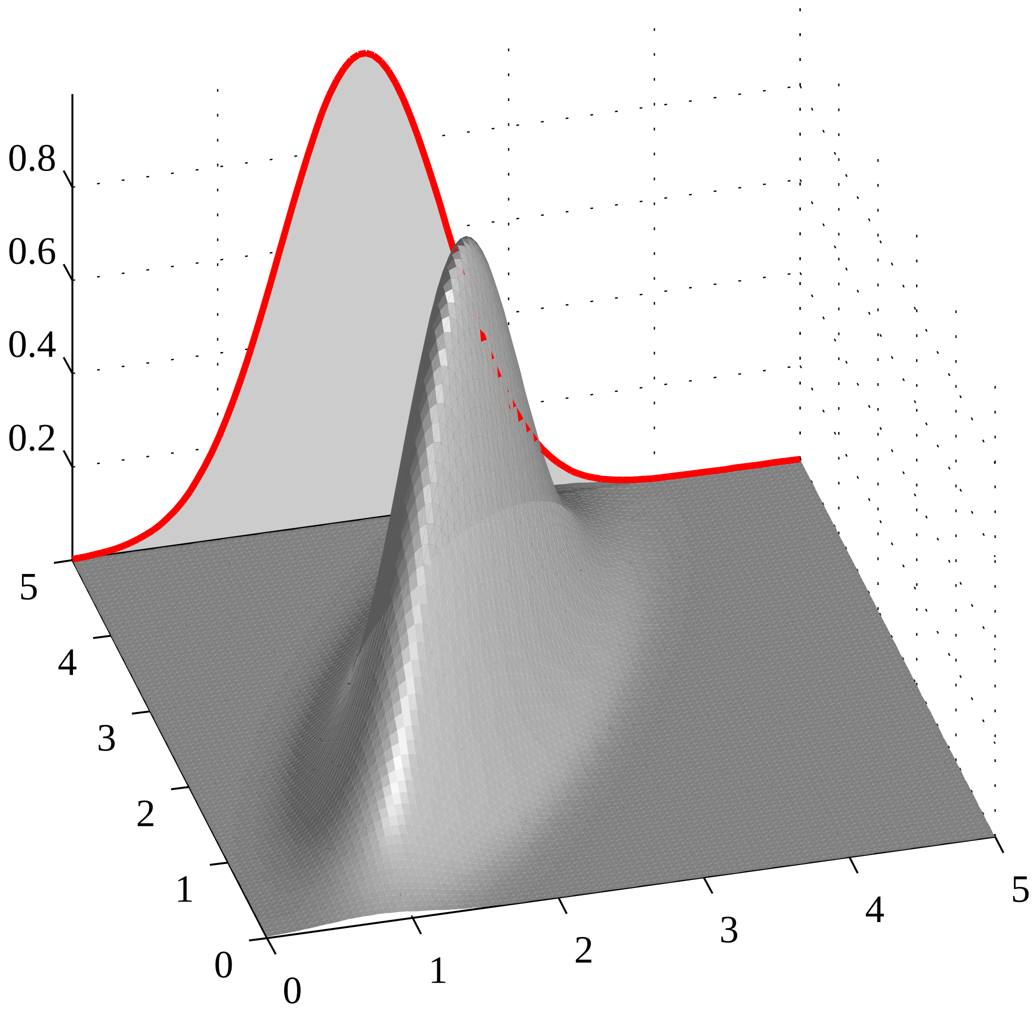
$$\pi(x \mid y) = \frac{\pi(y \mid x)\pi(x)}{\pi(y)} \quad \text{(Bayes formula)}$$

CONDITIONAL DENSITY



Marginal density





CONDITIONAL EXPECTATIONS

$$\mathrm{E}\{X \mid y\} = \int_{\mathbb{R}} x\pi(x \mid y)dx.$$

Expectation of X via conditional expectation:

$$\mathrm{E}\{X\} = \int x\pi(x)dx = \int x\left(\int \pi(x,y)dy\right)dx,$$

Substitute:

$$\begin{aligned} \mathrm{E}\{X\} &= \int x\left(\int \pi(x \mid y)\pi(y)dy\right)dx \\ &= \int \left(\int x\pi(x \mid y)dx\right)\pi(y)dy = \int \mathrm{E}\{X \mid y\}\pi(y)dy. \end{aligned} \tag{1}$$

MULTIVARIATE RANDOM VARIABLE

Define

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} : \Omega \to \mathbb{R}^n,$$

with each component X_i being an $\mathbb{R}$-valued variable.

Probability density of X = joint probability density $\pi = \pi_X : \mathbb{R}^n \to \mathbb{R}+$ of its components.

Expectation is

$$\overline{x} = \int_{\mathbb{R}^n} x\pi(x)dx \in \mathbb{R}^n,$$

or, componentwise,

$$\overline{x}_i = \int_{\mathbb{R}^n} x_i\pi(x)dx \in \mathbb{R}, \quad 1 \leq i \leq n.$$

The *covariance matrix* is defined as

$$\mathrm{cov}(X) = \int_{\mathbb{R}^n} (x - \overline{x})(x - \overline{x})^{\mathrm{T}} \pi(x) dx \in \mathbb{R}^{n \times n},$$

or, componentwise,

$$\mathrm{cov}(X)_{ij} = \int_{\mathbb{R}^n} (x_i - \overline{x}_i)(x_j - \overline{x}_j) \pi(x) dx \in \mathbb{R}^{n \times n}, \quad 1 \leq i, j \leq n.$$

Covariance matrix is *symmetric* and *positive semi-definite*: For any $v \in \mathbb{R}^n$, $v \neq 0$,

$$\begin{aligned} v^{\mathrm{T}} \mathrm{cov}(X) v &= \int_{\mathbb{R}^n} \left[v^{\mathrm{T}} (x - \overline{x}) \right] \left[(x - \overline{x})^{\mathrm{T}} v \right] \pi(x) dx \qquad (2) \\ &= \int_{\mathbb{R}^n} \left(v^{\mathrm{T}} (x - \overline{x}) \right)^2 \pi(x) dx \geq 0. \end{aligned}$$

Variance of X into the direction v.

Diagonal of the covariance matrix gives the variances of the individual components.

Denote by $x_i' \in \mathbb{R}^{n-1}$ the vector x with the ith component deleted:

$$
\begin{aligned}
\mathrm{cov}(X)_{ii} &= \int_{\mathbb{R}^n} (x_i - \overline{x}_i)^2 \pi(x) dx = \int_{\mathbb{R}} (x_i - \overline{x}_i)^2 \underbrace{\left(\int_{\mathbb{R}^{n-1}} \pi(x_i, x_i') dx_i' \right)}_{= \pi(x_i)} dx_i \\
&= \int_{\mathbb{R}} (x_i - \overline{x}_i)^2 \pi(x_i) dx_i = \mathrm{var}(X_i).
\end{aligned}
$$

EXAMPLE

On working days, a train leaves from a station every S minutes. On Sundays, the interval is $2S$.

You arrive to the station with no information of the time table.

Waiting time = random variable T, distribution

$$T \sim \pi(t \mid \text{working day}) = \frac{1}{S}\chi_S(t), \quad \chi_S(t) = \begin{cases} 1, & 0 \le t < S, \\ 0 & \text{otherwise} \end{cases}$$

and

$$T \sim \pi(t \mid \text{Sunday}) = \frac{1}{2S}\chi_{2S}(t).$$

Expected waiting times

$$\mathrm{E}\{T \mid \text{working day}\} = \frac{1}{S}\int_0^S t dt = \frac{S}{2}.$$

Similarly,

$$\mathrm{E}\{T \mid \text{Sunday}\} = S.$$

No idea of the weekday: give equal probability for each week day:

$$\pi(\text{working day}) = \frac{6}{7}, \quad \pi(\text{Sunday}) = \frac{1}{7}.$$

Waiting time, regardless of the day:

$$\begin{aligned} \mathrm{E}\{T\} &= \mathrm{E}\{T \mid \text{working day}\}\pi(\text{working day}) + \mathrm{E}\{T \mid \text{Sunday}\}\pi(\text{Sunday}) \\ &= \frac{3S}{7} + \frac{S}{7} = \frac{4S}{7}. \end{aligned}$$

EXAMPLES OF DISTRIBUTIONS

A weak light source emits photons that are counted with a CCD (*Charged Coupled Device*).

Counting process $N(t)$,

$$N(t) = \text{ number of particles observed in } [0, t] \in \mathbb{N}$$

is an integer-valued random variable.

To set up a statistical model, make the following assumptions:

1. *Stationarity:* Let Δ_1 and Δ_2 be any two time intervals of equal length, n any non-negative integer. Assume that

$$\text{Prob. of } n \text{ photons in } \Delta_1 = \text{Prob. of } n \text{ photons in } \Delta_2.$$

2. *Independent increments:* Let $\Delta_1, \ldots, \Delta_n$ be non-overlapping time intervals, $k_1, \ldots, k_n$ non-negative integers. Denote by A_j the event defined as

$$A_j = k_j \text{ photons arrive in the time interval } \Delta_j.$$

Assume that these events are mutually independent,

$$\mathrm{P}\big\{A_1 \cap \ldots \cap A_n\big\} = \mathrm{P}\big\{A_1\big\} \cdots \mathrm{P}\big\{A_n\big\}.$$

3. *Negligible probability of coincidence:* Assume that the probability of two or more events at the same time is negligible. More precisely, assume that $N(0) = 0$ and

$$\lim_{h \to 0} \frac{\mathrm{P}\big\{N(h) > 1\big\}}{h} = 0.$$

("No faster than linear growth.")

If these assumptions hold, then N is a *Poisson process:*

$$\mathrm{P}\big\{N(t) = n\big\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad \lambda > 0.$$

(Proof, see, e.g., S Ghahramani: *Fundamentals of Probability.* Prentice Hall 1996.)

Fix $t = T$=observation time and define a random variable $N = N(T)$. Write $\theta = \lambda T$. We write

$$N \sim \mathrm{Poisson}(\theta).$$

Expectation

$$\pi(n) = \mathrm{P}\{N = n\} = \frac{\theta^n}{n!} e^{-\theta}, \quad \theta > 0.$$

$$\begin{aligned} \mathrm{E}\{N\} &= \sum_{n=0}^{\infty} n\pi(n) = e^{-\theta} \sum_{n=0}^{\infty} n \frac{\theta^n}{n!} \\ &= e^{-\theta} \sum_{n=1}^{\infty} \frac{\theta^n}{(n-1)!} = e^{-\theta} \sum_{n=0}^{\infty} \frac{\theta^{n+1}}{n!} \\ &= \theta. \end{aligned}$$

Variance

Write first

$$\begin{aligned} \mathrm{E}\{(N-\theta)^2\} &= \mathrm{E}\{N^2\} - 2\theta \underbrace{\mathrm{E}\{N\}}_{=\theta} + \theta^2 \\ &= \mathrm{E}\{N^2\} - \theta^2 \\ &= \sum_{n=0}^{\infty} n^2 \pi(n) - \theta^2, \end{aligned}$$

and substitute

$$\pi(n) = \frac{\theta^n}{n!} e^{-\theta}, \quad \theta > 0.$$

$$
\begin{aligned}
\mathrm{E}\big\{(N-\theta)^2\big\} &= e^{-\theta}\sum_{n=0}^{\infty} n^2\frac{\theta^n}{n!} - \theta^2 = e^{-\theta}\sum_{n=1}^{\infty} n\frac{\theta^n}{(n-1)!} - \theta^2 \\
&= e^{-\theta}\sum_{n=0}^{\infty}(n+1)\frac{\theta^{n+1}}{n!} - \theta^2 \\
&= \theta e^{-\theta}\sum_{n=0}^{\infty} n\frac{\theta^n}{n!} + \theta e^{-\theta}\sum_{n=0}^{\infty}\frac{(\theta)^n}{n!} - \theta^2 \\
&= \theta e^{-\theta}\big((\theta+1)e^{\theta}\big) - \theta^2 \\
&= \theta.
\end{aligned}
$$

GAUSSIAN DISTRIBUTIONS

A random variable $X \in \mathbb{R}$ is normally distributed, or Gaussian,

$$X \sim \mathcal{N}(x_0, \sigma^2),$$

if

$$\mathrm{P}\{X \leq t\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} \exp\left(-\frac{1}{2\sigma^2}(x - x_0)^2\right) dx.$$

Multivariate extension: $X \in \mathbb{R}^n$ is Gaussian, if its probability density is

$$\pi(x) = \left(\frac{1}{(2\pi)^n \mathrm{det}(\Gamma)}\right)^{1/2} \exp\left(-\frac{1}{2}(x - x_0)^{\mathrm{T}}\Gamma^{-1}(x - x_0)\right),$$

where $x_0 \in \mathbb{R}^n$, $\Gamma \in \mathbb{R}^{n\times n}$ is symmetric positive definite.

$$\mathrm{P}\{X \in B\} = \int_B \pi(x) dx.$$

CENTRAL LIMIT THEOREM

Gaussian variables appear, e.g., when *macroscopic* measurements are averages of individual *microscopic* random effects.

EXAMPLES: Pressure, temperature, electric current, luminosity.

CENTRAL LIMIT THEOREM: Assume that random variables $X_1, X_2, \ldots$ are *independent* and *identically distributed* (i.i.d.), each with expectation μ and variance σ^2. Then

$$Z_n = \frac{1}{\sigma\sqrt{n}}\big(X_1 + X_2 + \cdots + X_n - n\mu\big)$$

converges to the distribution of a standard normal random variable,

$$\lim_{n\to\infty} \mathrm{P}\{Z_n \leq x\} = \frac{1}{2\pi}\int_{-\infty}^{x} e^{-t^2/2}dt.$$

In practice, if

$$Y_n = \frac{1}{n}\sum_{j=1}^{n} X_j,$$

and n is large, a good approximation is

$$Y_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

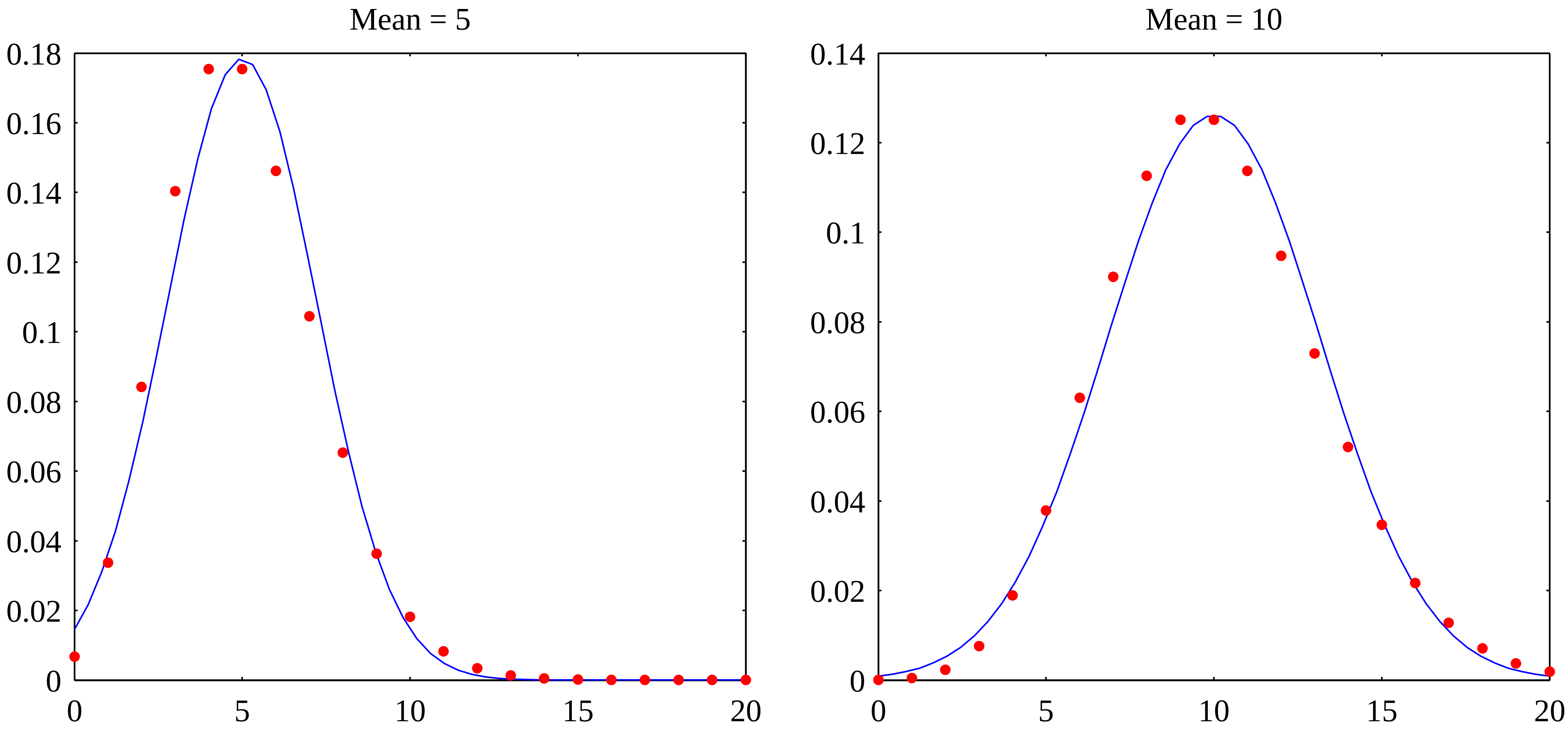
Try with Poisson distribution: Total count of photons is a sum of individual contributions.

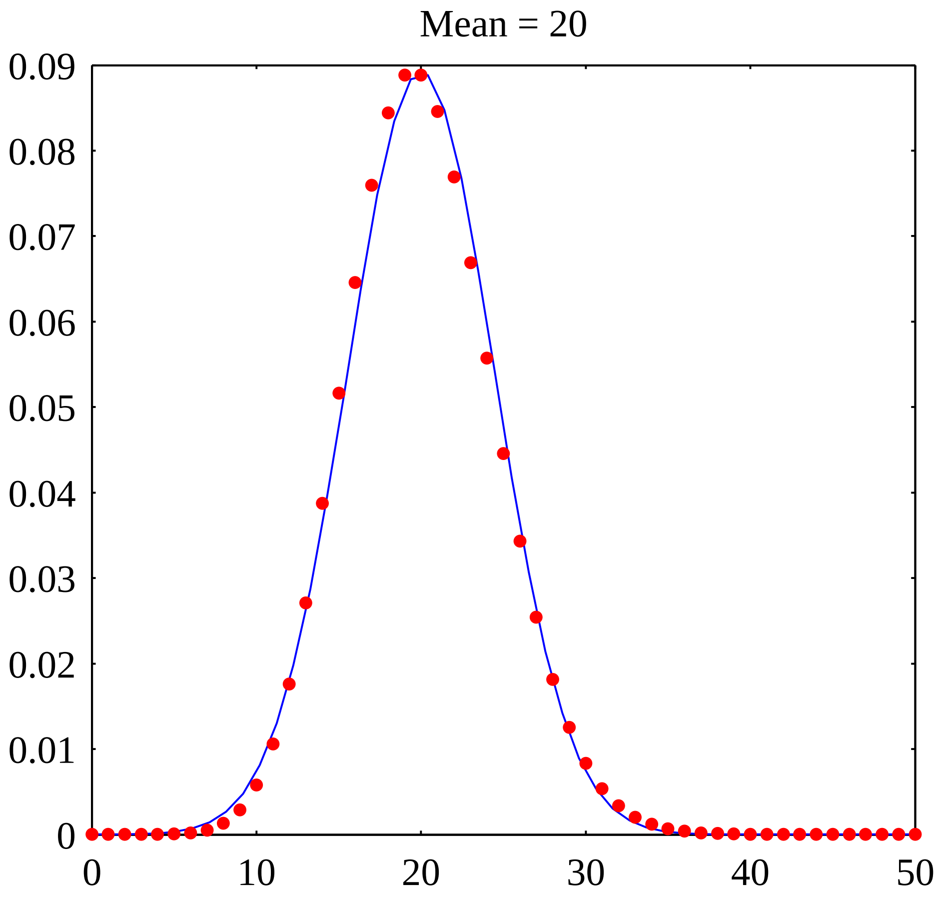
Plot

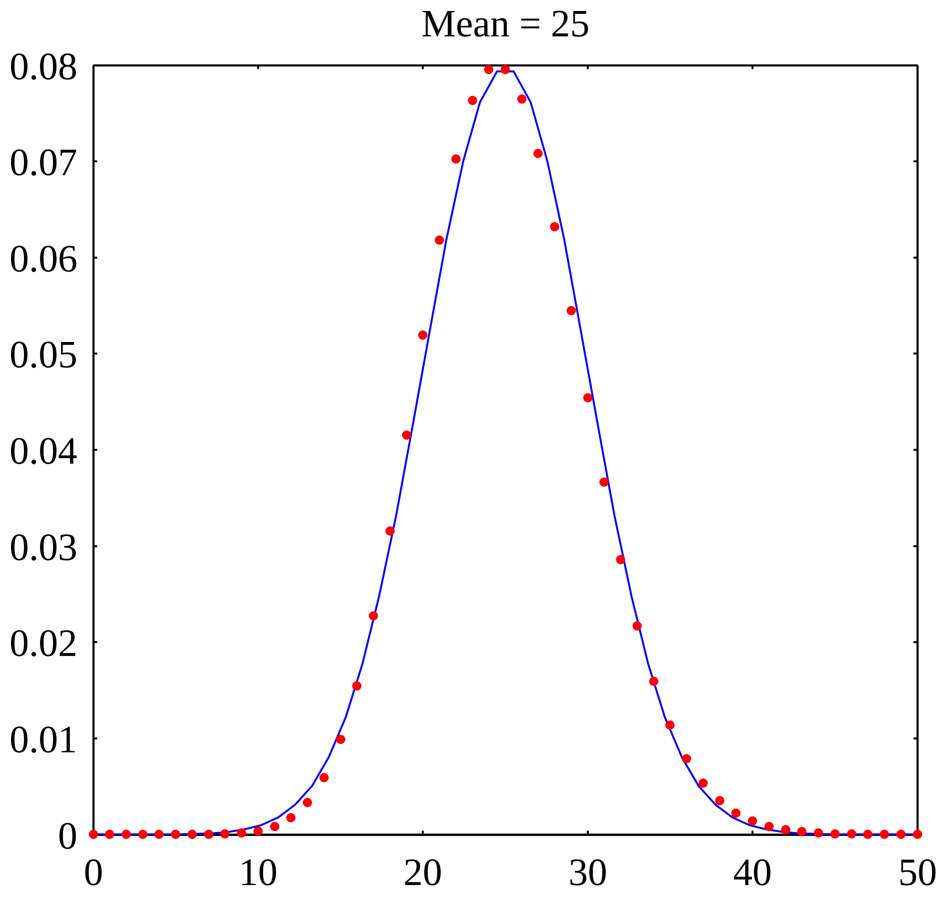
$$n \mapsto \frac{\theta^n}{n!}e^{-\theta} = \pi_{\rm Poisson}(n \mid \theta)$$

versus

$$x \mapsto \frac{1}{\sqrt{2\pi\theta}}\exp\left(-\frac{1}{2\theta}(x-\theta)^2\right) = \pi_{\rm Gaussian}(x \mid \theta, \theta).$$


Mean = 5
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
Mean = 10
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20


Mean = 20
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
10
20
30
40
50


Mean = 25
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
10
20
30
40
50

Measuring the Quality of Approximation

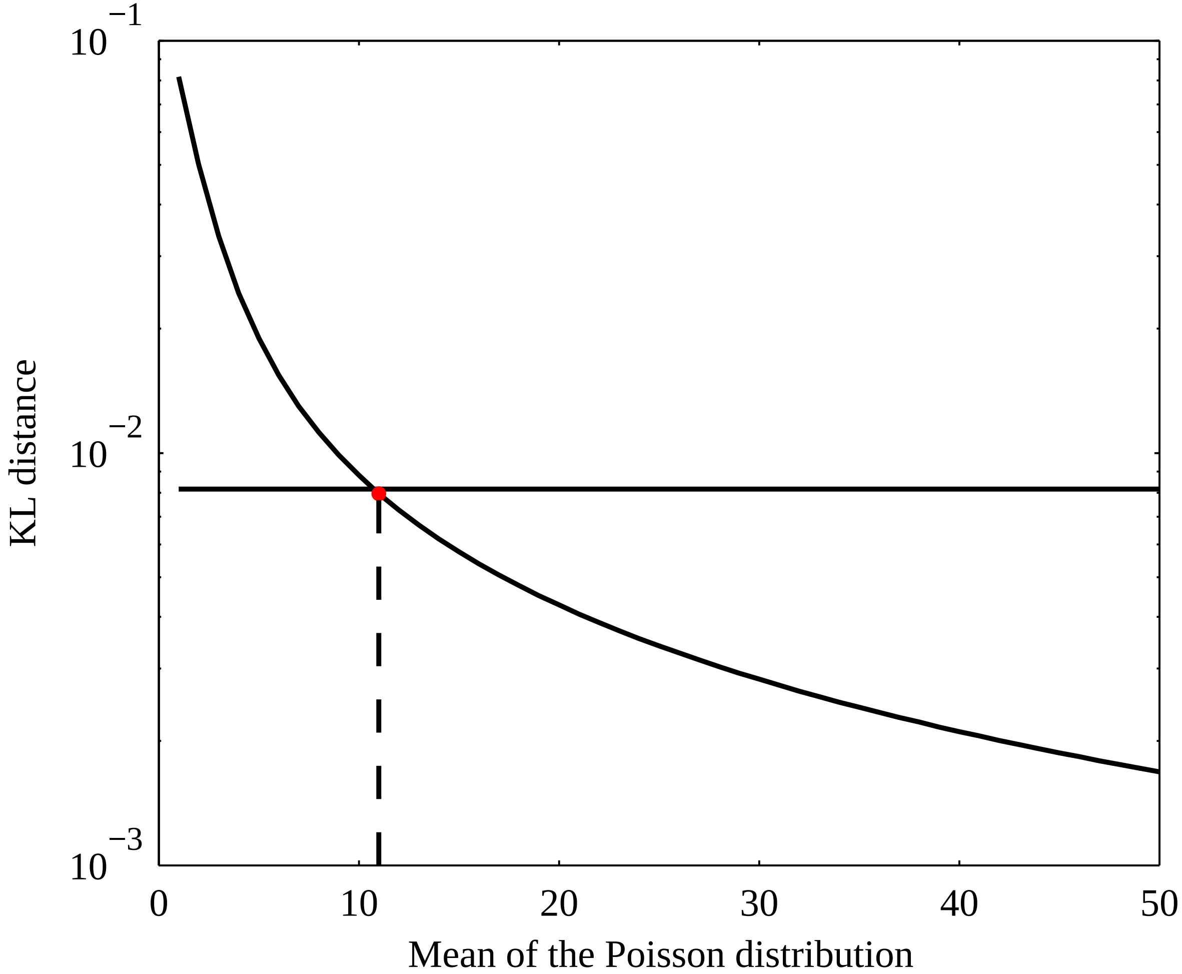
Relative error:

$$e(\theta, n) = \frac{|\pi_{\rm Poisson}(n \mid \theta) - \pi_{\rm Gaussian}(n \mid \theta, \theta)|}{\pi_{\rm Poisson}(n \mid \theta)}.$$

Distance between the two densities:

$$\begin{aligned}
&{\rm dist}_{\rm KL}\big(\pi_{\rm Poisson}(\,\cdot\, \mid \theta), \pi_{\rm Gaussian}(\,\cdot\, \mid \theta, \theta)\big) \\
&\qquad = \sum_{n=0}^{\infty} \pi_{\rm Poisson}(n \mid \theta) \log\left(\frac{\pi_{\rm Gaussian}(n \mid \theta)}{\pi_{\rm Poisson}(n \mid \theta, \theta)}\right),
\end{aligned}$$

known as *Kullback–Leibler distance.*


10^−1
10^−2
10^−3
KL distance
0
10
20
30
40
50
Mean of the Poisson distribution