Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"As far as the laws of mathematics refer to reality, they are not certain,

Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality. "
(A. Einstein)

Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality. "
(A. Einstein)

"It is scientific only to say what is more likely and what is less likely"
(R. Feynman)

Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

"As far as the laws of mathematics refer to reality, they are not certain,
and as far as they are certain, they do not refer to reality. "
(A. Einstein)

"It is scientific only to say what is more likely
and what is less likely"
(R. Feynman)

"Probability is good sense reduced to a calculus"
(S. Laplace)

Inferring 'proportions'

Let's turn the toy experiment to a 'serious' physics case:

- Inferring H_j is the same as inferring the proportion of white balls:

$$H_j \quad \longleftrightarrow \quad j \quad \longleftrightarrow \quad p = \frac{j}{5}$$

Inferring ‘proportions’

Let’s turn the toy experiment to a ‘serious’ physics case:

- Inferring H_j is the same as inferring the proportion of white balls:

$$H_j \quad \longleftrightarrow \quad j \quad \longleftrightarrow \quad p = \frac{j}{5}$$

- Increase the number of balls

$$n: \quad 6 \rightarrow \infty$$

$\Rightarrow$ p continuous in $[0, 1]$

Inferring ‘proportions’

Let’s turn the toy experiment to a ‘serious’ physics case:

- Inferring H_j is the same as inferring the proportion of white balls:

$$H_j \quad \longleftrightarrow \quad j \quad \longleftrightarrow \quad p = \frac{j}{5}$$

- Increase the number of balls

$$n: \quad 6 \rightarrow \infty$$

  $\Rightarrow$ p continuous in $[0, 1]$

- Generalize White/Black $\longrightarrow$ Success/Failure

$\Rightarrow$ efficiencies, branching ratios, ...

Uncertain numbers

What matters is **uncertainty**

Uncertain numbers

What matters is **uncertainty**

A number respect to which we are in condition of uncertainty

Uncertain numbers

What matters is **uncertainty**

A number respect to which we are in condition of uncertainty

- The first number rolling a die

Uncertain numbers

What matters is **uncertainty**

A number respect to which we are in condition of uncertainty

- The first number rolling a die
- The temperature at the Geneva airport tomorrow at 7:00 am

Uncertain numbers

What matters is **uncertainty**

A number respect to which we are in condition of uncertainty

- The first number rolling a die
- The temperature at the Geneva airport tomorrow at 7:00 am
- The rate Euro/Dollar the 23 April 2023

Uncertain numbers

What matters is **uncertainty**

A number respect to which we are in condition of uncertainty

- The first number rolling a die
- The temperature at the Geneva airport tomorrow at 7:00 am
- The rate Euro/Dollar the 23 April 2023
- The integrated luminosity provided by LHC in 2008

Uncertain numbers

What matters is **uncertainty**

A number respect to which we are in condition of uncertainty

- The first number rolling a die
- The temperature at the Geneva airport tomorrow at 7:00 am
- The rate Euro/Dollar the 23 April 2023
- The integrated luminosity provided by LHC in 2008
- The number of signatures of the first LHC physics paper

Uncertain numbers

What matters is **uncertainty**

A number respect to which we are in condition of uncertainty

- The first number rolling a die
- The temperature at the Geneva airport tomorrow at 7:00 am
- The rate Euro/Dollar the 23 April 2023
- The integrated luminosity provided by LHC in 2008
- The number of signatures of the first LHC physics paper
- The mass of the Higgs boson

Uncertain numbers

What matters is **uncertainty**

A number respect to which we are in condition of uncertainty

- The first number rolling a die
- The temperature at the Geneva airport tomorrow at 7:00 am
- The rate Euro/Dollar the 23 April 2023
- The integrated luminosity provided by LHC in 2008
- The number of signatures of the first LHC physics paper
- The mass of the Higgs boson

They could be referred to future, past, model parameters or even hyper-parameters.

Uncertain numbers

What matters is **uncertainty**

A number respect to which we are in condition of uncertainty

- The first number rolling a die
- The temperature at the Geneva airport tomorrow at 7:00 am
- The rate Euro/Dollar the 23 April 2023
- The integrated luminosity provided by LHC in 2008
- The number of signatures of the first LHC physics paper
- The mass of the Higgs boson

They could be referred to future, past, model parameters or even hyper-parameters.

What matters is uncertainty!

Uncertain numbers

What matters is **uncertainty**

A number respect to which we are in condition of uncertainty

- The first number rolling a die
- The temperature at the Geneva airport tomorrow at 7:00 am
- The rate Euro/Dollar the 23 April 2023
- The integrated luminosity provided by LHC in 2008
- The number of signatures of the first LHC physics paper
- The mass of the Higgs boson

They could be referred to future, past, model parameters or even hyper-parameters.
What matters is uncertainty!

But it must be a well defined number

- any uncertainty on its definition will increase our uncertainty about it ($\rightarrow$ ISO)

Bernoulli process

A basic simple example, although conceptually very important.

Bernoulli process

A basic simple example, although conceptually very important.

- Bernoulli process
  - X : $0, 1$ (failure/success)
    $f(0) = 1 - p$
    $f(1) = p$
  - it seems of practical irrelevance,
    $\rightarrow$ but of primary importance

Bernoulli process

A basic simple example, although conceptually very important.

- Bernoulli process
  - $X : 0, 1$ (failure/success)
    $f(0) = 1 - p$
    $f(1) = p$
  - it seems of practical irrelevance,
    $\rightarrow$ but of primary importance
- The drunk man problem
  - Eight keys
  - After each trial he 'loses memory'
  - We watch him and – cynically – bet on the attempt on which he will succeed:
  - $X = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, \ldots$?

Bernoulli process

A basic simple example, although conceptually very important.

- Bernoulli process
  - X : 0, 1 (failure/success)
    $f(0) = 1 - p$
    $f(1) = p$
  - it seems of practical irrelevance,
    → but of primary importance
- The drunk man problem
  - Eight keys
  - After each trial he 'loses memory'
  - We watch him and – cynically – bet on the attempt on which he will succeed:
  - $X = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, \ldots ?$

  → On which number would you bet?

Propagating probability values

We cannot say any number "at random",

"because All attempts are equally likely"

Propagating probability values

We cannot say any number "at random",

"because All attempts are equally likely"

$\rightarrow$ 'half true', i.e. wrong. . .

- what is constant is $P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$,
  where $E_i \rightarrow X = i$.

Propagating probability values

We cannot say any number "at random",

"because All attempts are equally likely"

$\rightarrow$ 'half true', i.e. wrong...

- what is constant is $P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, where $E_i \rightarrow X = i$.

$\Rightarrow$ Beliefs are framed in a network!

- Once we assess something, we are implicitly making an infinity of assessments concerning logically connected events!
- We only need to make them explicit, using logic (trivial in principle, though it can be sometimes hard)

Propagating probability values

We cannot say any number "at random",

"because All attempts are equally likely"

$\rightarrow$ 'half true', i.e. wrong...

- what is constant is $P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, where $E_i \rightarrow X = i$.

$\Rightarrow$ Beliefs are framed in a network!

- Once we assess something, we are implicitly making an infinity of assessments concerning logically connected events!
- We only need to make them explicit, using logic (trivial in principle, though it can be sometimes hard)

(Observation at the basis of "propagation of uncertainties", the so called error propagation.)

Building up $f(x)$ of the drunk man problem

$P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, with $p = 1/8$:

Building up $f(x)$ of the drunk man problem

$P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, with $p = 1/8$:

$$f(1) \quad = \quad P(E_1) = p$$

Building up $f(x)$ of the drunk man problem

$P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, with $p = 1/8$:

$$\begin{aligned} f(1) &= P(E_1) = p \\ f(2) &= P(E_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)\,p \end{aligned}$$

Building up $f(x)$ of the drunk man problem

$P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, with $p = 1/8$:

$$
\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(E_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)\,p \\
f(3) &= P(E_3 \,|\, \overline{E}_1 \cap \overline{E}_1) \cdot P(\overline{E}_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)^2\,p
\end{aligned}
$$

Building up $f(x)$ of the drunk man problem

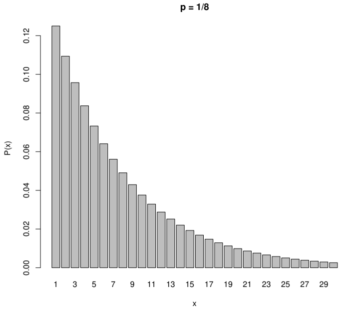
$P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, with $p = 1/8$:

$$\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(E_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)\,p \\
f(3) &= P(E_3 \,|\, \overline{E}_1 \cap \overline{E}_1) \cdot P(\overline{E}_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)^2\,p \\
f(x) &= p\,(1-p)^{x-1}
\end{aligned}$$

Building up $f(x)$ of the drunk man problem

$P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, with $p = 1/8$:

$$
\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(E_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)\,p \\
f(3) &= P(E_3 \,|\, \overline{E}_1 \cap \overline{E}_1) \cdot P(\overline{E}_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)^2\,p \\
f(x) &= p\,(1-p)^{x-1}
\end{aligned}
$$



Beliefs decrease geometrically

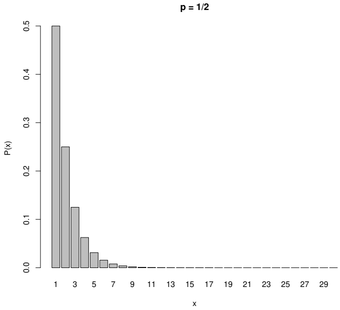
$\Rightarrow$ Geometric distribution

Building up $f(x)$ of the drunk man problem

$P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, with $p = 1/8$:

$$
\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(E_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)\,p \\
f(3) &= P(E_3 \,|\, \overline{E}_1 \cap \overline{E}_1) \cdot P(\overline{E}_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)^2\,p \\
f(x) &= p\,(1-p)^{x-1}
\end{aligned}
$$

$p = 1/2 \rightarrow$ tossing a coin

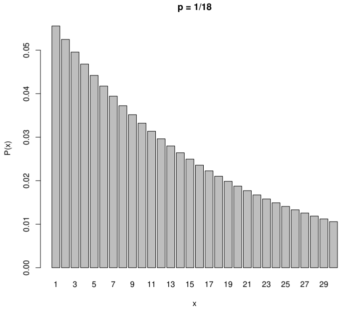


Building up $f(x)$ of the drunk man problem

$P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, with $p = 1/8$:

$$
\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(E_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)\, p \\
f(3) &= P(E_3 \,|\, \overline{E}_1 \cap \overline{E}_1) \cdot P(\overline{E}_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)^2\, p \\
f(x) &= p\,(1-p)^{x-1}
\end{aligned}
$$

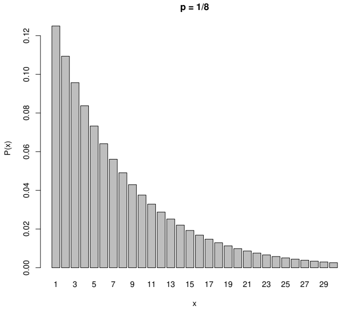
$p = 1/18 \rightarrow$ a particular number at the Italian lotto ($p = 5/90$)



Building up $f(x)$ of the drunk man problem

$P(E_i \,|\, \overline{\bigcup_{j<i} E_j}) = p$, with $p = 1/8$:

$$
\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(E_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)\,p \\
f(3) &= P(E_3 \,|\, \overline{E}_1 \cap \overline{E}_1) \cdot P(\overline{E}_2 \,|\, \overline{E}_1) \cdot P(\overline{E}_1) = (1-p)^2\,p \\
f(x) &= p\,(1-p)^{x-1}
\end{aligned}
$$



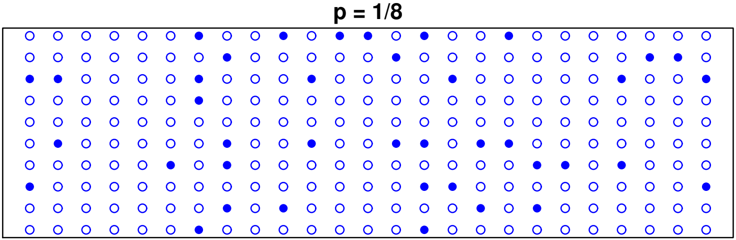
Most probable value does not depend on p.

Not a suitable indicator to state our expectation

The same is true for the range of possibilities: $X: 1, 2, \ldots, \infty$

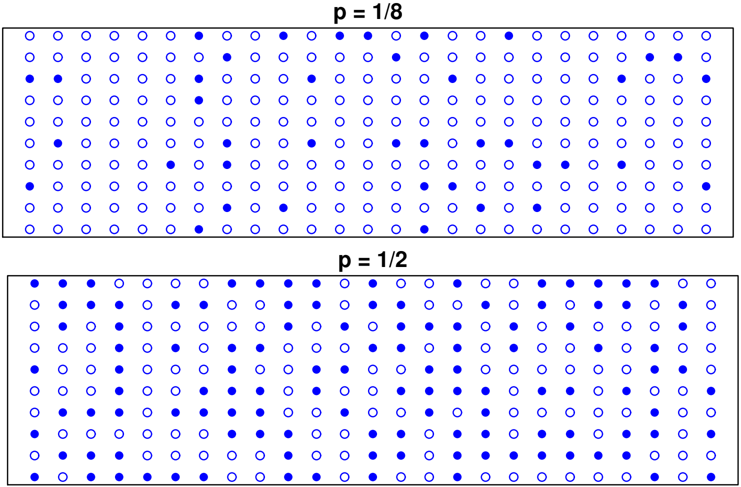
Sequencees of Bernoulli trials

The smaller is p the longer we have to wait to get a 'success'.



Sequencees of Bernoulli trials

The smaller is p the longer we have to wait to get a 'success'.



Prevision and prevision uncertainty

More suitable quantity two summarize in two numbers the our probabilistic 'expectation' and its uncertainty:

Prevision and prevision uncertainty

More suitable quantity two summarize in two numbers the our probabilistic ‘expectation’ and its uncertainty:

$$\mathrm{E}[X] \quad = \quad \sum_x x\, f(x)$$

Prevision and prevision uncertainty

More suitable quantity two summarize in two numbers the our probabilistic 'expectation' and its uncertainty:

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_x x\, f(x) \\
\mathrm{Variance}(X) &= \sum_x (x - \mathrm{E}[X])^2\, f(x) \quad \longrightarrow \sigma^2(X) \rightarrow \sigma = \sqrt{\sigma^2}
\end{aligned}
$$

Prevision and prevision uncertainty

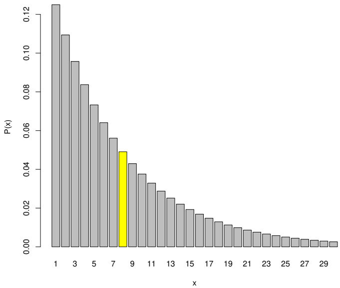
More suitable quantity two summarize in two numbers the our probabilistic 'expectation' and its uncertainty:

$$\begin{aligned} \mathrm{E}[X] &= \sum_x x\, f(x) \\ \mathrm{Variance}(X) &= \sum_x (x - \mathrm{E}[X])^2\, f(x) \quad \longrightarrow \sigma^2(X) \rightarrow \sigma = \sqrt{\sigma^2} \end{aligned}$$

$\mathrm{E}[X] = 1/p$
$\sigma(X) = \sqrt{1-p}/p$

$p = 1/8$:

$$\begin{aligned} \mathrm{E}[X] &= 8 \\ \sigma(X) &= 7.5 \end{aligned}$$



Prevision and prevision uncertainty

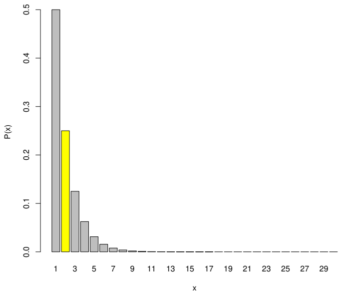
More suitable quantity two summarize in two numbers the our probabilistic 'expectation' and its uncertainty:

$$\begin{aligned} \mathrm{E}[X] &= \sum_x x\, f(x) \\ \mathrm{Variance}(X) &= \sum_x (x - \mathrm{E}[X])^2\, f(x) \quad \longrightarrow \sigma^2(X) \rightarrow \sigma = \sqrt{\sigma^2} \end{aligned}$$

$\mathrm{E}[X] = 1/p$
$\sigma(X) = \sqrt{1-p}/p$

$p = 1/2$:

$$\begin{aligned} \mathrm{E}[X] &= 2 \\ \sigma(X) &= 1.4 \end{aligned}$$



Prevision and prevision uncertainty

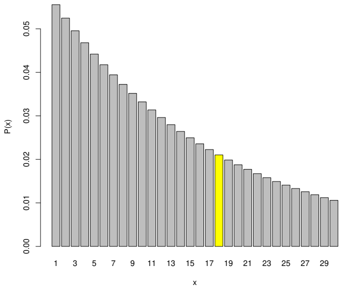
More suitable quantity two summarize in two numbers the our probabilistic ‘expectation’ and its uncertainty:

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_x x\, f(x) \\
\mathrm{Variance}(X) &= \sum_x (x - \mathrm{E}[X])^2\, f(x) \quad \longrightarrow \sigma^2(X) \rightarrow \sigma = \sqrt{\sigma^2}
\end{aligned}
$$

$\mathrm{E}[X] = 1/p$
$\sigma(X) = \sqrt{1-p}/p$

$p = 1/18$:

$$
\begin{aligned}
\mathrm{E}[X] &= 18 \\
\sigma(X) &= 17.5
\end{aligned}
$$



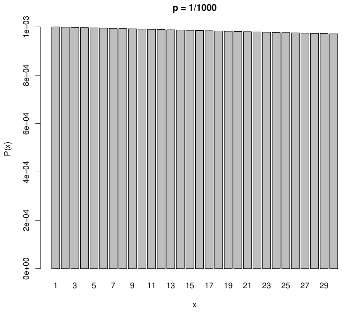
Prevision and prevision uncertainty

More suitable quantity two summarize in two numbers the our probabilistic 'expectation' and its uncertainty:

$$\begin{aligned} \mathrm{E}[X] &= \sum_x x\, f(x) \\ \mathrm{Variance}(X) &= \sum_x (x-\mathrm{E}[X])^2\, f(x) \;\longrightarrow \sigma^2(X) \rightarrow \sigma = \sqrt{\sigma^2} \end{aligned}$$

$\mathrm{E}[X] = 1/p$
$\sigma(X) = \sqrt{1-p}/p \xrightarrow[p\to 0]{} 1/p$

$\rightarrow$ rare events might happen at any moment!
(Though they have 'zero' probability to happen at any given moment!)



Mechanical analogies of E[X] and Var[X]

$\Longrightarrow$ Center of mass and momentum of inertia

Mechanical analogies of E[X] and Var[X]

$\Longrightarrow$ Center of mass and momentum of inertia

$$\sigma \neq \text{RMS!}$$

Mechanical analogies of E[X] and Var[X]

$\Longrightarrow$ Center of mass and momentum of inertia

$$\sigma \neq \text{RMS!}$$

- Nasty comment of a non-HEP colleague during the discussion of a Tesi di Laurea in Rome (the poor student had presented several ROOT plots. . .):

Mechanical analogies of E[X] and Var[X]

$\Longrightarrow$ Center of mass and momentum of inertia

$$\sigma \neq \text{RMS!}$$

- Nasty comment of a non-HEP colleague during the discussion of a Tesi di Laurea in Rome (the poor student had presented several ROOT plots. . .):
  "If he doesn't even understand the difference between *standard deviation* and *root mean square*, how can we trust that he understands the sophisticated things he is talking about?"

Some details on the geometric distribution

Normalization and cumulative distribution

$$\begin{aligned}\sum_{x=1}^{\infty} f(x\,|\,\mathcal{G}_p) &= \sum_{x=1}^{\infty}(1-p)^{x-1}\,p \\ &= p\sum_{x=1}^{\infty} q^{x-1} \\ &= p\,\frac{1}{1-q} = 1\end{aligned}$$

Some details on the geometric distribution

Normalization and cumulative distribution

$$\begin{aligned}
\sum_{x=1}^{\infty} f(x\,|\,\mathcal{G}_p) &= \sum_{x=1}^{\infty}(1-p)^{x-1}\,p \\
&= p\sum_{x=1}^{\infty} q^{x-1} \\
&= p\,\frac{1}{1-q} = 1
\end{aligned}$$

$$\begin{aligned}
F(x\,|\,\mathcal{G}_p) \equiv P(X \le x) &= 1 - P(X > x) \\
&= 1 - P(\overline{E}_1 \cap \overline{E}_2 \cap \cdots \cap \overline{E}_x) \\
&= 1 - (1-p)^x \quad \text{for } x = 1, 2, \ldots
\end{aligned}$$

Some details on the geometric distribution

Expected value

$$
\begin{aligned}
\mathrm{E}(X \mid \mathcal{G}_p) &= \sum_{x=1}^{\infty} x\, f(x) = \sum_{x=1}^{\infty} x\, p\,(1-p)^{x-1} \\
&= p \sum_{x=1}^{\infty} x\, q^{x-1} \\
&= p \frac{d}{dq} \sum_{x=1}^{\infty} q^{x} \\
&= p \frac{d}{dq} \left(\frac{1}{1-q} \right) = p \frac{1}{(1-q)^2} \\
&= \frac{1}{p}
\end{aligned}
$$

Some details on the geometric distribution

Expected value

$$
\begin{aligned}
\mathrm{E}(X \mid \mathcal{G}_p) &= \sum_{x=1}^{\infty} x\, f(x) = \sum_{x=1}^{\infty} x\, p\, (1-p)^{x-1} \\
&= p \sum_{x=1}^{\infty} x\, q^{x-1} \\
&= p \frac{d}{dq} \sum_{x=1}^{\infty} q^x \\
&= p \frac{d}{dq} \left(\frac{1}{1-q} \right) = p \frac{1}{(1-q)^2} \\
&= \frac{1}{p}
\end{aligned}
$$

Variance: calculation a bit more complicate $\rightarrow \mathrm{Var}(X \mid \mathcal{G}_p) = q/p^2$

Some details on the geometric distribution

As implemented in packages

Usually the variable X is not defined the trial of the first success, but rather the number of trials before it.

Some details on the geometric distribution

As implemented in packages

Usually the variable X is not defined the trial of the first success, but rather the number of trials before it.

- X is shifted by -1;

Some details on the geometric distribution

As implemented in packages

Usually the variable X is not defined the trial of the first success, but rather the number of trials before it.

- X is shifted by -1;
- similarly the expected value;

Some details on the geometric distribution

As implemented in packages

Usually the variable X is not defined the trial of the first success, but rather the number of trials before it.

- X is shifted by -1;
- similarly the expected value;
- variance invariant.

Some details on the geometric distribution

As implemented in packages

Usually the variable X is not defined the trial of the first success, but rather the number of trials before it.

- X is shifted by -1;
- similarly the expected value;
- variance invariant.
- Example in R:

```
> x <- rgeom(1000000, 1/8)
> mean(x)
 [1] 6.993153
> sd(x)
 [1] 7.474005
```

Expected value and 'standard uncertainty'

The detail on the uncertainty is provided by $f(x)$.

- E$[X]$ and $\sigma(X)$ are just convenient summaries.
- In the general case they do not convey a precise confidence that X will occur in the range E$[X] \pm \sigma(X)$, though this probability is rather 'high' for typical $f(x)$ of interest.

Expected value and 'standard uncertainty'

The detail on the uncertainty is provided by $f(x)$.

- E$[X]$ and $\sigma(X)$ are just convenient summaries.
- In the general case they do not convey a precise confidence that X will occur in the range E$[X] \pm \sigma(X)$, though this probability is rather 'high' for typical $f(x)$ of interest.
- Another location summary (that statisticians like much) is given by the median, while the 'quantiles' provide (left open) intervals in which the variable is expected to fall with some probability (typically 10%, 20%, etc.).

Expected value and ‘standard uncertainty’

The detail on the uncertainty is provided by $f(x)$.

- $\mathrm{E}[X]$ and $\sigma(X)$ are just convenient summaries.
- In the general case they do not convey a precise confidence that X will occur in the range $\mathrm{E}[X] \pm \sigma(X)$, though this probability is rather ‘high’ for typical $f(x)$ of interest.
- Another location summary (that statisticians like much) is given by the median, while the 'quantiles' provide (left open) intervals in which the variable is expected to fall with some probability (typically 10%, 20%, etc.).
- Anyway, it is important to be prepared to $f(x)$ of any kind, because – fortunately! – Nature is not boring...

Expected value and ‘standard uncertainty’

The detail on the uncertainty is provided by $f(x)$.

- $\mathrm{E}[X]$ and $\sigma(X)$ are just convenient summaries.
- In the general case they do not convey a precise confidence that X will occur in the range $\mathrm{E}[X] \pm \sigma(X)$, though this probability is rather ‘high’ for typical $f(x)$ of interest.
- Another location summary (that statisticians like much) is given by the median, while the ’quantiles’ provide (left open) intervals in which the variable is expected to fall with some probability (typically 10%, 20%, etc.).
- Anyway, it is important to be prepared to $f(x)$ of any kind, because – fortunately! – Nature is not boring...
- In particular, $f(x)$ might be asymmetric or, ‘multinomial’, i.e. with more than one local maximum.

Probability distributions and ‘statistical’ distributions

It is important to stress the difference between

- Probability distribution
  - To each possible outcome we associate how much we are confident on it:

$$x \longleftrightarrow f(x)$$

- Statistical distribution
  - To each observed outcome we associated its (relative) frequency

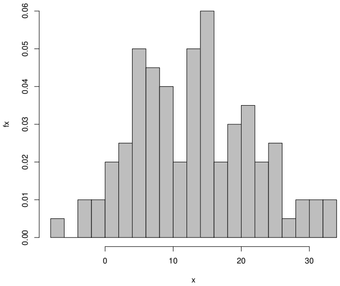
$$x \longleftrightarrow f_x$$

(e.g. an histogram of experimental observations)

Summaries (‘mean’, variance, ’σ’, ’skewness’, etc) have similar names and analogous definitions, but conceptual different meaning.

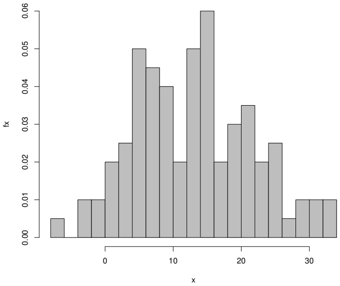
A histogram is not, usually, a probability distribution

In particular a histogram of experimental data <u>is not</u> a probability distribution (unless one reshuffles those events, and extracts one of them at random).



A histogram is not, usually, a probability distribution

In particular a histogram of experimental data is not a probability distribution (unless one reshuffles those events, and extracts one of them at random).



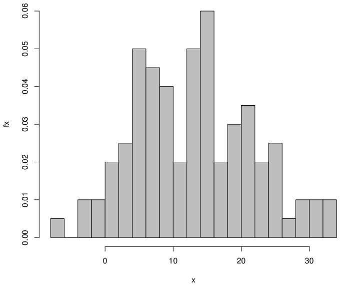
Average and variance

$$\overline{x} = \sum_x x\, f_x$$

$$\sigma^2 = \sum_x (x - \overline{x})^2\, f_x$$

A histogram is not, usually, a probability distribution

In particular a histogram of experimental data is not a probability distribution (unless one reshuffles those events, and extracts one of them at random).



Average and variance

$$\overline{x} = \sum_x x\, f_x$$

$$\sigma^2 = \sum_x (x - \overline{x})^2\, f_x$$

→ Just a rough empirical description of the shape
⇒ center of mass and momentum of inertia!
(Famous '$n/(n-1)$' correction: interference descriptive ↔ inferential statistics.)

Number of successes in n trials

Well known and understood Binomial distribution

- Each sequence of x successes and $(n-x)$ failures has probability $P_s(x, n, p) = p^x(1-p)^{(n-x)}$.

Number of successes in n trials

Well known and understood Binomial distribution

- Each sequence of x successes and $(n-x)$ failures has probability $P_s(x,n,p) = p^x(1-p)^{(n-x)}$.
- If we are only interested into the number of successes, independently of the order, we have to sum up all probabilities, that is multiply $P_s(x,n,p)$ by the number of sequences.

Number of successes in n trials

Well known and understood Binomial distribution

- Each sequence of x successes and $(n-x)$ failures has probability $P_s(x, n, p) = p^x(1-p)^{(n-x)}$.
- If we are only interested into the number of successes, independently of the order, we have to sum up all probabilities, that is multiply $P_s(x, n, p)$ by the number of sequences.
- The latter is the number of combinations of x elements taken from n "objects", and is equal to the binomial coefficient (and hence the name).

Number of successes in n trials

Well known and understood Binomial distribution

- Each sequence of x successes and $(n-x)$ failures has probability $P_s(x,n,p) = p^x(1-p)^{(n-x)}$.
- If we are only interested into the number of successes, independently of the order, we have to sum up all probabilities, that is multiply $P_s(x,n,p)$ by the number of sequences.
- The latter is the number of combinations of x elements taken from n "objects", and is equal to the binomial coefficient (and hence the name).
- We get finally

$$f(x\,|\,\mathcal{B}_{n,p}) = \binom{n}{x} p^x\,(1-p)^{n-x} = \frac{n!}{x!\,(n-x)!}\,p^x\,(1-p)^{n-x}\,.$$

Number of successes in n trials

Well known and understood Binomial distribution

- Each sequence of x successes and $(n-x)$ failures has probability $P_s(x,n,p) = p^x(1-p)^{(n-x)}$.
- If we are only interested into the number of successes, independently of the order, we have to sum up all probabilities, that is multiply $P_s(x,n,p)$ by the number of sequences.
- The latter is the number of combinations of x elements taken from n “objects”, and is equal to the binomial coefficient (and hence the name).
- We get finally

$$f(x|\mathcal{B}_{n,p}) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!\,(n-x)!}\, p^x (1-p)^{n-x}\,.$$

- Summaries (we shall see how to get them very easily)

$$\mu = \mathrm{E}[X \,|\, \mathcal{B}_{n,p}] = n\,p$$
$$\sigma[X \,|\, \mathcal{B}_{n,p}] = \sqrt{n\,p\,(1-p)}$$
$$v \equiv \frac{\sigma}{|\mu|} \propto \frac{1}{\sqrt{n}}$$

Binomial distribution

Barplot with R

```
> n=10; p=0.3; x=0:n; P=dbinom(x, n, p)
```

Binomial distribution

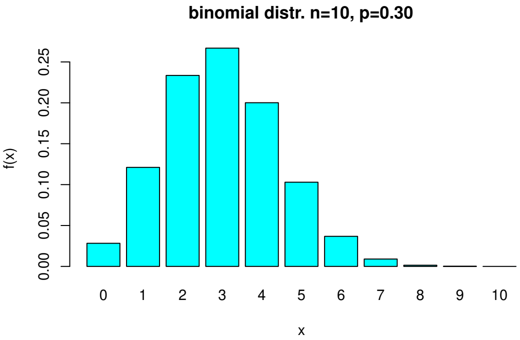
Barplot with R

```
> n=10; p=0.3; x=0:n; P=dbinom(x, n, p)
> barplot(P, names=x, col='cyan', xlab='x', ylab='f(x)',
+ main=sprintf("binomial distr.  n=%d, p=%.2f",n,p))
```

Binomial distribution

Barplot with R

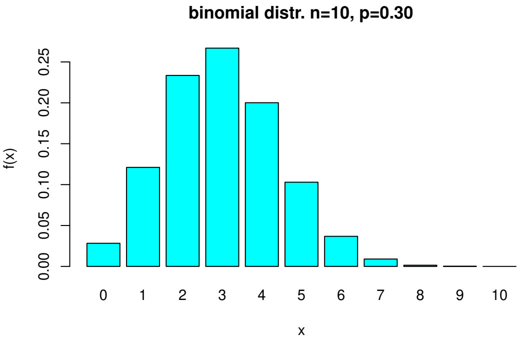
```
> n=10; p=0.3; x=0:n; P=dbinom(x, n, p)
> barplot(P, names=x, col='cyan', xlab='x', ylab='f(x)',
+ main=sprintf("binomial distr.  n=%d, p=%.2f",n,p))
```



Binomial distribution

Barplot with R

```
> n=10; p=0.3; x=0:n; P=dbinom(x, n, p)
> barplot(P, names=x, col='cyan', xlab='x', ylab='f(x)',
+ main=sprintf("binomial distr.  n=%d, p=%.2f",n,p))
```

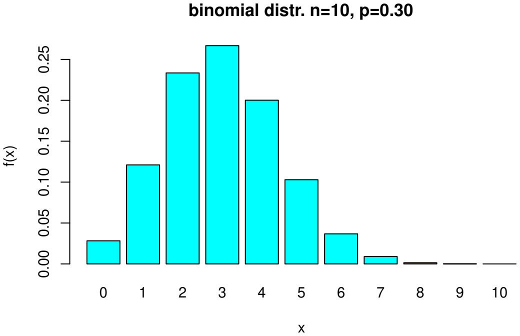



```
> ( E.X <- sum(x*P) )
 [1] 3
```

Binomial distribution

Barplot with R

```
> n=10; p=0.3; x=0:n; P=dbinom(x, n, p)
> barplot(P, names=x, col='cyan', xlab='x', ylab='f(x)',
+ main=sprintf("binomial distr.  n=%d, p=%.2f",n,p))
```




```
> ( E.X <- sum(x*P) )
 [1] 3
> ( sigma.X <- sqrt( sum(x^2*P) - E.X^2 ) )
 [1] 1.449138
```

Binomial for $n \to \infty$ and $p \to 0$

$m = n\,p = 1$						
x	$f(x)$					
	$\mathcal{B}_{2,\frac{1}{2}}$	$\mathcal{B}_{4,\frac{1}{4}}$	$\mathcal{B}_{10,\frac{1}{10}}$	$\mathcal{B}_{50,\frac{1}{50}}$	$\mathcal{B}_{1000,\frac{1}{1000}}$	$\mathcal{B}_{10^6,10^{-6}}$
0	0.25	0.316	0.349	0.364	0.368	0.368
1	0.50	0.422	0.387	0.372	0.368	0.368
2	0.25	0.211	0.194	0.186	0.184	0.184
3		0.047	0.057	0.061	0.061	0.061
4		0.004	0.011	0.015	0.015	0.015
5			0.001	0.003	0.003	0.003
6			...	...	0.001	0.001
...			...	...	...	...
10			10^{-10}	...	...	...
...				...	...	...
50				$\approx 10^{-85}$	...	...
...					...	...
1000					$\approx 10^{-3000}$	...
...						...
1000000						$\approx 10^{-6\times 10^6}$

Calculating extreme probabilities with R

```
> n=10; dbinom(n, n, 1/n)
 [1] 1e-10
```

Calculating extreme probabilities with R

```
> n=10; dbinom(n, n, 1/n)
 [1] 1e-10
> n=10; dbinom(n, n, 1/n, log=TRUE)/log(10)
 [1] -10
```

Calculating extreme probabilities with R

```
> n=10; dbinom(n, n, 1/n)
 [1] 1e-10
> n=10; dbinom(n, n, 1/n, log=TRUE)/log(10)
 [1] -10
> n=50; dbinom(n, n, 1/n, log=TRUE)/log(10)
 [1] -84.9485
> n=1000; dbinom(n, n, 1/n, log=TRUE)/log(10)
 [1] -3000
> n=1000000; dbinom(n, n, 1/n, log=TRUE)/log(10)
 [1] -6e+06
```

Pascal distribution

$\Rightarrow$ Trial number at which the k-th 'success' (exactly) occurs.
(For $k = 1$ it recovers the Geometric distribution.)

Pascal distribution

$\Rightarrow$ Trial number at which the k-th 'success' (exactly) occurs.
(For $k = 1$ it recovers the Geometric distribution.)

- Exercise: get $f(x \,|\, \mathcal{P}a_{k,p})$

Pascal distribution

$\Rightarrow$ Trial number at which the k-th 'success' (exactly) occurs.
(For $k = 1$ it recovers the Geometric distribution.)

- Exercise: get $f(x \,|\, \mathcal{P}a_{k,p})$

- we shall see in the sequel how to evaluate simply expected value and standard deviation:

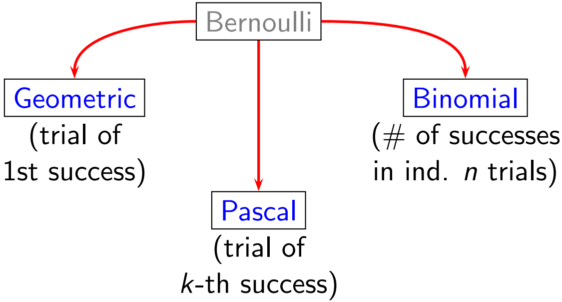
Pascal distribution

$\Rightarrow$ Trial number at which the k-th 'success' (exactly) occurs.
(For $k = 1$ it recovers the Geometric distribution.)

- Exercise: get $f(x \,|\, \mathcal{P}a_{k,p})$

- we shall see in the sequel how to evaluate simply expected value and standard deviation:
  - Guesses?

Distributions derived from the Bernoulli process



(Binomial well known. We shall not use the Pascal)

Poisson distribution

One of the best known distributions by physicists.

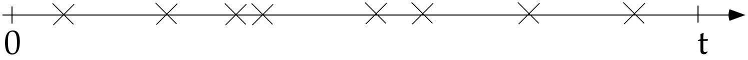
For a while, just take a mathematical approach:

$$f(x\,|\,\mathcal{P}_\lambda) = \frac{\lambda^x}{x!}\,e^{-\lambda} \qquad \left\{ \begin{array}{l} 0 < \lambda < \infty \\ x = 0, 1, \ldots, \infty \end{array} \right. .$$

Reminding also the well known property

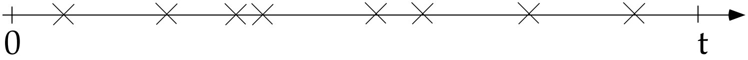
$$\mathcal{B}_{n,p} \xrightarrow[\substack{n \to \infty \\ p \to 0 \\ (n\,p = \lambda)}]{} P_\lambda\,.$$

Poisson process



Let us consider some phenomena that might happen at a give instant

Poisson process



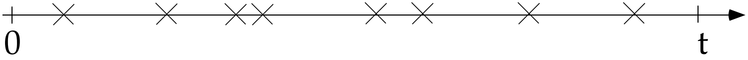
Let us consider some phenomena that might happen at a give instant, such that

- Probability of 1 count in ΔT is proportional to ΔT, with ΔT ‘small’.

$$p = P(\text{“1 count in } \Delta T\text{''}) = r\,\Delta T$$

where r is the intensity of the process’

Poisson process



Let us consider some phenomena that might happen at a give instant, such that

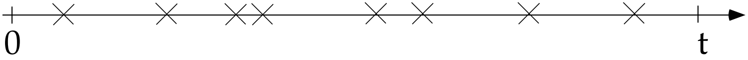
- Probability of 1 count in ΔT is proportional to ΔT, with ΔT 'small'.

$$p = P(\text{“1 count in } \Delta T\text{''}) = r\,\Delta T$$

 where r is the intensity of the process'

- $P(\geq 2 \text{ counts}) \ll P(1 \text{ count})$ (OK if ΔT is small enough)

Poisson process



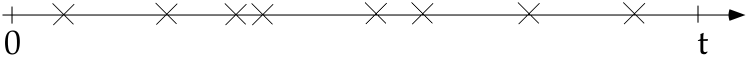
Let us consider some phenomena that might happen at a give instant, such that

- Probability of 1 count in ΔT is proportional to ΔT, with ΔT 'small'.

$$p = P(\text{“1 count in } \Delta T\text{''}) = r\,\Delta T$$

  where r is the intensity of the process'
- $P(\geq 2 \text{ counts}) \ll P(1 \text{ count})$ (OK if ΔT is small enough)
- What happens in one interval does not depend on what happened (or 'will happen') in other intervals (if disjoints)

Poisson process



Let us consider some phenomena that might happen at a give instant, such that

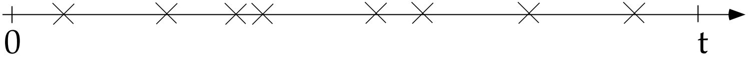
- Probability of 1 count in ΔT is proportional to ΔT, with ΔT 'small'.

$$p = P(\text{“1 count in } \Delta T\text{''}) = r\,\Delta T$$

  where r is the intensity of the process'
- $P(\geq 2 \text{ counts}) \ll P(1 \text{ count})$ (OK if ΔT is small enough)
- What happens in one interval does not depend on what happened (or 'will happen') in other intervals (if disjoints)

Let us divide a finite interval T in n small intervals, i.e. $T = n\,\Delta T$, and $\Delta T = T/n$.

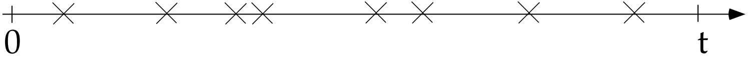
Poisson process → Poisson distribution



Considering the possible occurrence of a count in each small interval ΔT an independent Bernoulli trial, of probability

$$p = r\,\Delta T = r\,\frac{T}{n}$$

Poisson process → Poisson distribution

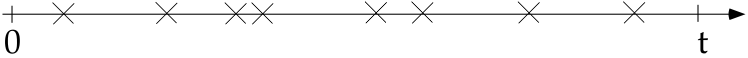


Considering the possible occurrence of a count in each small interval ΔT an independent Bernoulli trial, of probability

$$p = r\,\Delta T = r\,\frac{T}{n}$$

If we are interested in the number of counts in T, independently from the order: $\rightarrow$ Binomial : $\mathcal{B}_{n,p}$

Poisson process → Poisson distribution



Considering the possible occurrence of a count in each small interval ΔT an independent Bernoulli trial, of probability

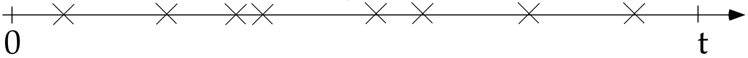
$$p = r\,\Delta T = r\,\frac{T}{n}$$

If we are interested in the number of counts in T, independently from the order: $\rightarrow$ Binomial : $\mathcal{B}_{n,p}$

But $n \rightarrow \infty$ and $p \rightarrow 0$ $\Rightarrow \mathcal{B}_{n,p} \rightarrow \mathcal{P}_\lambda$ where $\lambda = n\,p = r\,T$

$\Rightarrow$ λ depends only on the intensity of the process and on the finite time of observation.

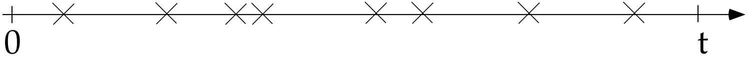
Poisson process → waiting time



Another interesting problem: how long do we have to wait for the first count? (Starting from any arbitrary time)

Problem analogous to the Geometric, but now it makes no sense to talk about the *i*-th small interval the counts will occur!

Poisson process → waiting time



Another interesting problem: how long do we have to wait for the first count? (Starting from any arbitrary time)

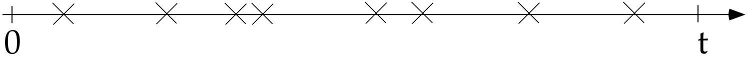
Problem analogous to the Geometric, but now it makes no sense to talk about the *i*-th small interval the counts will occur!

Let us restart from the Geometric and calculate $P(X > x)$:

$$P(X > x) = \sum_{i>x} f(i \,|\, \mathcal{G}_p) = (1-p)^x$$

(The count will not occur in the first x trials).

Poisson process → waiting time



Another interesting problem: how long do we have to wait for the first count? (Starting from any arbitrary time)

Problem analogous to the Geometric, but now it makes no sense to talk about the i-th small interval the counts will occur!

Let us restart from the Geometric and calculate $P(X > x)$:

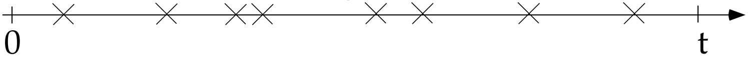
$$P(X > x) = \sum_{i>x} f(i \,|\, \mathcal{G}_p) = (1-p)^x$$

(The count will not occur in the first x trials).

In the domain of time, using $p = r\,t/n$ and then making the limit:

$$P(T > t) = (1-p)^n \quad = \quad \left(1 - \frac{r\,t}{n}\right)^n$$

Poisson process → waiting time



Another interesting problem: how long do we have to wait for the first count? (Starting from any arbitrary time)

Problem analogous to the Geometric, but now it makes no sense to talk about the *i*-th small interval the counts will occur!

Let us restart from the Geometric and calculate $P(X > x)$:

$$P(X > x) = \sum_{i>x} f(i\,|\,\mathcal{G}_p) = (1-p)^x$$

(The count will not occur in the first x trials).

In the domain of time, using $p = r\,t/n$ and then making the limit:

$$P(T > t) = (1-p)^n \;=\; \left(1 - \frac{r\,t}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-r\,t}$$

Poisson process → Exponential distribution

Knowing $P(T > t)$ we get easily the cumulative $F(t)$:

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-r\,t}\,.$$

$F(t)$ is now a continuous function!

Poisson process $\rightarrow$ Exponential distribution

Knowing $P(T > t)$ we get easily the cumulative $F(t)$:

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-r\,t}\,.$$

$F(t)$ is now a continuous function!

In some region of t there is a concentration of probability more than in other regions.

Poisson process $\rightarrow$ Exponential distribution

Knowing $P(T > t)$ we get easily the cumulative $F(t)$:

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-r\,t}\,.$$

$F(t)$ is now a continuous function!

In some region of t there is a concentration of probability more than in other regions.

$\rightarrow$ This leads us to define a probability density function (pdf) for continuous variables:

$$f(t) = \frac{d\,F(t)}{d\,t}\,.$$

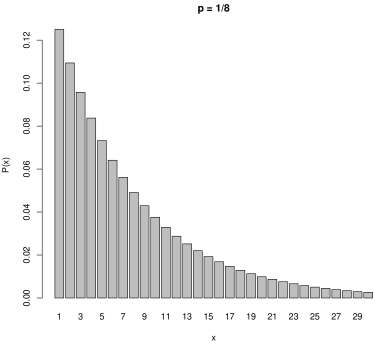
- In this case $f(t) = r\,e^{-r\,t} = \frac{1}{\tau}\,e^{-t/\tau}$

$\rightarrow$ Exponential distribution $(\tau = 1/r)$: $\mathrm{E}[T] = \sigma(T) = \tau$.

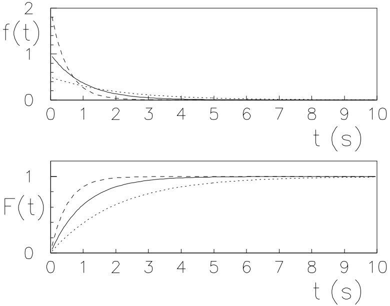
(⇒ Properties of pdf assumed to be known for the moment.)

Geometric ↔ Exponential

Geometric



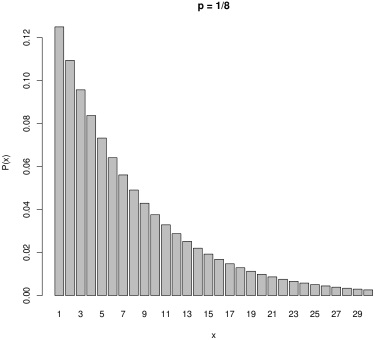
Exponential



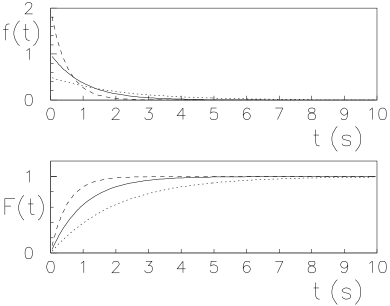
Exponential is just the limit to the continuum of the Geometric.

Geometric $\leftrightarrow$ Exponential

Geometric



Exponential

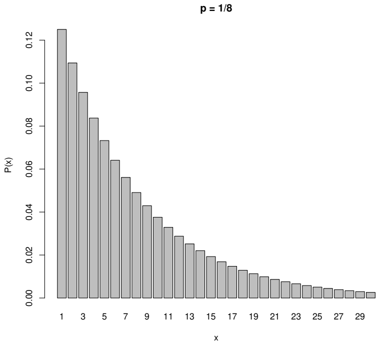


Exponential is just the limit to the continuum of the Geometric.

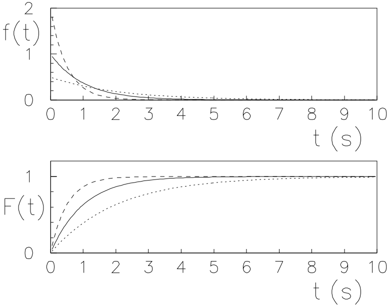
'No memory' property for both:

Geometric ↔ Exponential

Geometric



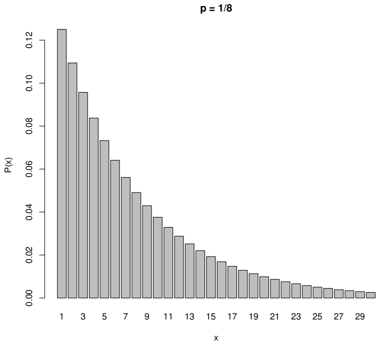
Exponential



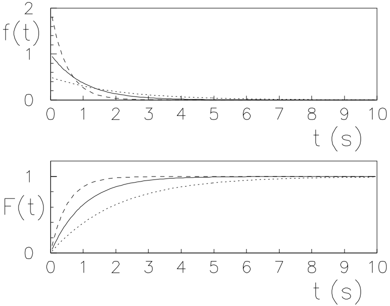
Exponential is just the limit to the continuum of the Geometric.
'No memory' property for both: Assuming that a success (or a count) has not happened until a certain trial (or time), the distributions restart from there.

Geometric $\leftrightarrow$ Exponential

Geometric

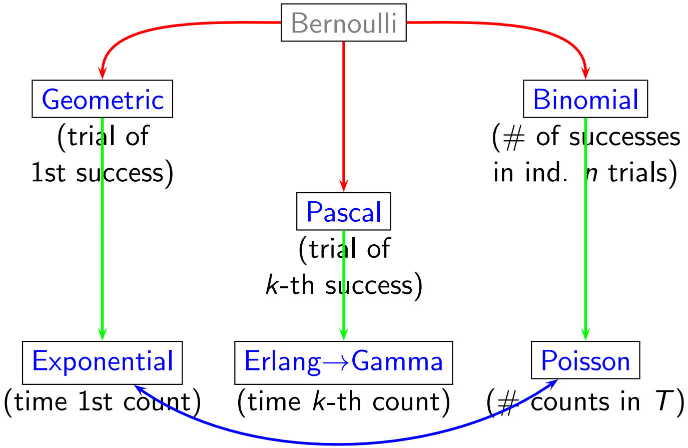


Exponential

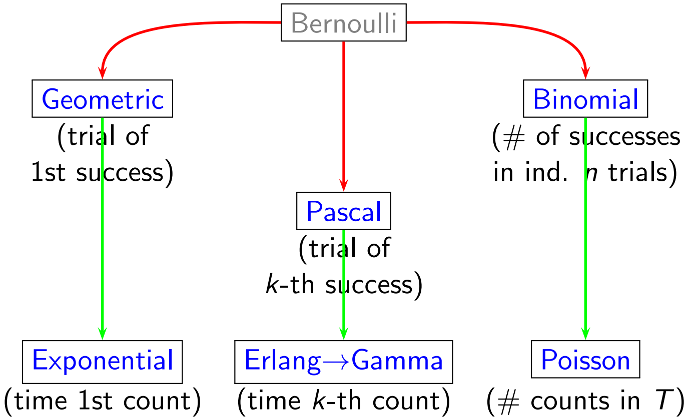


Exponential is just the limit to the continuum of the Geometric.
'No memory' property for both: Assuming that a success (or a count) has not happened until a certain trial (or time), the distributions restart from there. No need to know the instant of particle creation to measure 'life time' ($\rightarrow$ the "10^{25} year old" proton!).

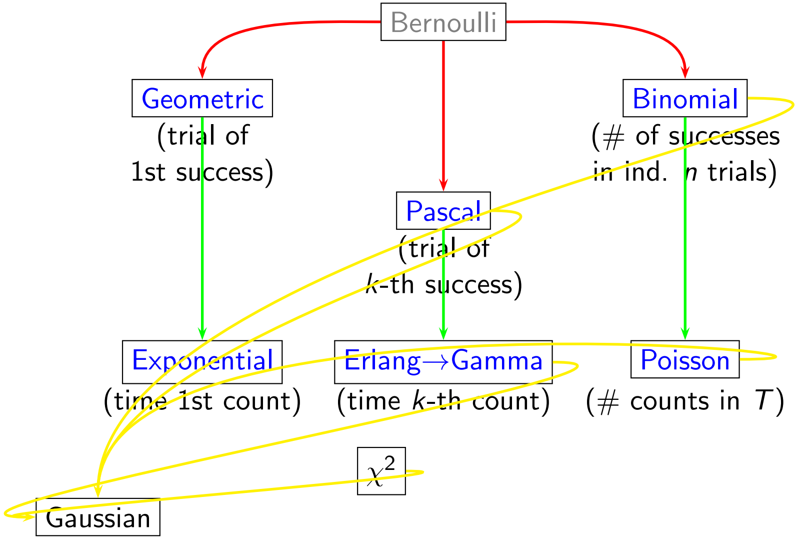
Distributions derived from the Bernoulli process



Distributions derived from the Bernoulli process



Distributions derived from the Bernoulli process



Important properties of probability distributions

(Assumed known)

$\mathrm{E}(\cdot)$ is a linear operator:

$$\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b\,.$$

Transformation properties of variance and standard deviation:

$$\begin{aligned} \mathrm{Var}(aX + b) &= a^2\,\mathrm{Var}(X)\,, \\ \sigma(aX + b) &= |a|\,\sigma(X)\,. \end{aligned}$$

From probability to future frequencies

Let us think to n independent Bernoulli trials that
have to be made.

Number of successes $X \sim \mathcal{B}_{n,p}$, with p.

We might be interested to the relative frequency of successes, i.e.
$f_n = X/n$: $f_n = 0, 1/n, 2/n, \ldots, 1$

What do we expect for f_n?

From probability to future frequencies

Let us think to n independent Bernoulli trials that have to be made.

Number of successes $X \sim \mathcal{B}_{n,p}$, with p.

We might be interested to the relative frequency of successes, i.e. $f_n = X/n$: $f_n = 0,\, 1/n,\, 2/n,\, \ldots,\, 1$

What do we expect for f_n? $f(f_n)$ can be obtained from $f(x)$.

$$
\begin{aligned}
\mathrm{E}(f_n) &\equiv \frac{1}{n}\,\mathrm{E}(X\,|\,\mathcal{B}_{n,p}) = \frac{n\,p}{n} = p \\
\sigma(f_n) &\equiv \frac{1}{n}\,\sigma(X\,|\,\mathcal{B}_{n,p}) = \frac{\sqrt{p\,(1-p)}}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0
\end{aligned}
$$

We expect p, with uncertainty that decreases with $\sqrt{n}$:
→ *Bernoulli's theorem*, the most known, misunderstood and misused theorem of probability theory.

From probability to future frequencies

Let us think to n independent Bernoulli trials that have to be made.

Number of successes $X \sim \mathcal{B}_{n,p}$, with p.

We might be interested to the relative frequency of successes, i.e. $f_n = X/n$: $f_n = 0, 1/n, 2/n, \ldots, 1$

What do we expect for f_n? $f(f_n)$ can be obtained from $f(x)$.

$$\begin{aligned} \mathrm{E}(f_n) &\equiv \frac{1}{n}\,\mathrm{E}(X \,|\, \mathcal{B}_{n,p}) = \frac{n\,p}{n} = p \\ \sigma(f_n) &\equiv \frac{1}{n}\,\sigma(X \,|\, \mathcal{B}_{n,p}) = \frac{\sqrt{p\,(1-p)}}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0 \end{aligned}$$

In particular, it justifies neither the ‘increased probability‘ of 'late numbers' at lotto, nor frequency based definition of probability (Circular: cannot define probability from a theorem resulting from Probability Theory!)

About the misuse of Bernoulli theorem

"For those who seek to connect the notion of probability with that of frequency, results which relate probability and frequency in some way (and especially those results like the 'law of large numbers') play a pivotal rôle, providing support for the approach and for the identification of the concepts.

About the misuse of Bernoulli theorem

"For those who seek to connect the notion of probability with that of frequency, results which relate probability and frequency in some way (and especially those results like the 'law of large numbers') play a pivotal rôle, providing support for the approach and for the identification of the concepts. Logically speaking, however, one cannot escape from the dilemma posed by the fact that the same thing cannot both be assumed first as a definition and then proved as a theorem;

About the misuse of Bernoulli theorem

"For those who seek to connect the notion of probability with that of frequency, results which relate probability and frequency in some way (and especially those results like the 'law of large numbers') play a pivotal rôle, providing support for the approach and for the identification of the concepts. Logically speaking, however, one cannot escape from the dilemma posed by the fact that the same thing cannot both be assumed first as a definition and then proved as a theorem; nor can one avoid the contradiction that arises from a definition which would assume as certain something that the theorem only states to be very probable."

(de Finetti)

Probabilistic inference

applied
to the ‘binomial case’

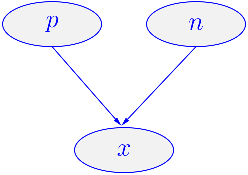
n independent Bernoulli processes

General case

n independent Bernoulli processes

General case

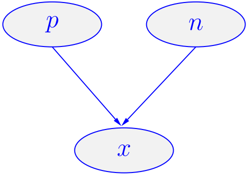
Model



n independent Bernoulli processes

General case

Model



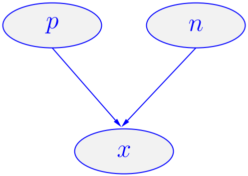
Joint pdf (omitting background condition I):

$$f(x, p, n) = f(x \mid p, n)$$

n independent Bernoulli processes

General case

Model



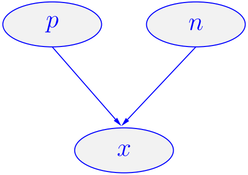
Joint pdf (omitting background condition I):

$$f(x, p, n) \quad = \quad f(x \mid p, n) \cdot f(p, n)$$

n independent Bernoulli processes

General case

Model



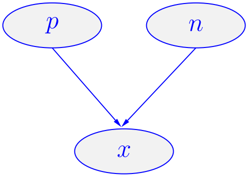
Joint pdf (omitting background condition I):

$$
\begin{aligned}
f(x, p, n) &= f(x \mid p, n) \cdot f(p, n) \\
&= f(x \mid p, n) \cdot f(p \mid n) \cdot f(n)
\end{aligned}
$$

n independent Bernoulli processes

General case

Model



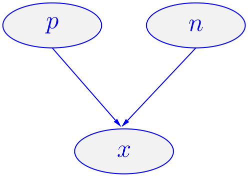
Joint pdf (omitting background condition I):

$$\begin{aligned} f(x, p, n) &= f(x \,|\, p, n) \cdot f(p, n) \\ &= f(x \,|\, p, n) \cdot f(p \,|\, n) \cdot f(n) \\ &= f(x \,|\, p, n) \cdot f(n \,|\, p) \cdot f(p) \end{aligned}$$

n independent Bernoulli processes

General case

Model



Joint pdf (omitting background condition I):

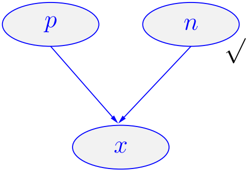
$$
\begin{aligned}
f(x,p,n) &= f(x\,|\,p,n)\cdot f(p,n) \\
&= f(x\,|\,p,n)\cdot f(p\,|\,n)\cdot f(n) \\
&= f(x\,|\,p,n)\cdot f(n\,|\,p)\cdot f(p) \\
&= f(x\,|\,p,n)\cdot f(p)\cdot f(n)
\end{aligned}
$$

(n and p are independent)

n independent Bernoulli processes

Usual case → n fixed (for the moment)

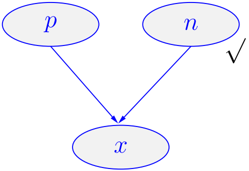
Model



n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



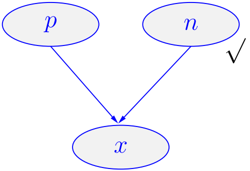
Joint pdf

$$f(x, p \mid n) = f(x \mid p, n) \cdot f(p)$$

n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



Joint pdf

$$f(x, p \,|\, n) \;=\; f(x \,|\, p, n) \cdot f(p)$$

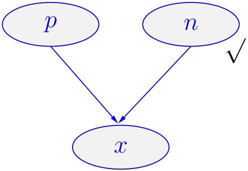
Typical problems

- p is assumed → interested in $f(x \,|\, n, p)$
  → well known binomial;

n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



Joint pdf

$$f(x, p \mid n) \;=\; f(x \mid p, n) \cdot f(p)$$

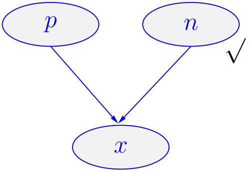
Typical problems

- p is assumed → interested in $f(x \mid n, p)$
  → well known binomial;
- x is assumed ('observed')

n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



Joint pdf

$$f(x, p \mid n) = f(x \mid p, n) \cdot f(p)$$

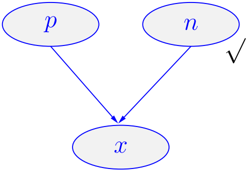
Typical problems

- p is assumed → interested in $f(x \mid n, p)$
  → well known binomial;
- x is assumed ('observed') → $f(p \mid n, x)$

n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



Joint pdf

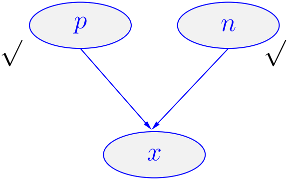
$$f(x, p \,|\, n) \;=\; f(x \,|\, p, n) \cdot f(p)$$

Typical problems

- p is assumed → interested in $f(x \,|\, n, p)$
  → well known binomial;
- x is assumed ('observed') → $f(p \,|\, n, x)$:
  → ?

n independent Bernoulli processes

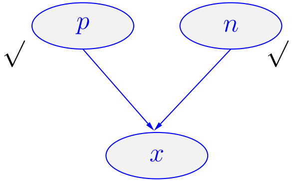
Graphical models of the typical problems



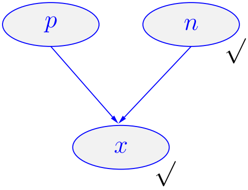
$\rightarrow\ f(x\,|\,n,p)$

n independent Bernoulli processes

Graphical models of the typical problems



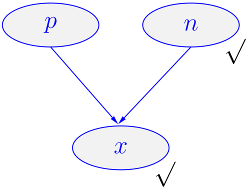
$\rightarrow\ f(x\,|\,n,p)$



$\rightarrow\ f(p\,|\,n,x)$

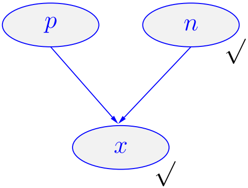
n independent Bernoulli processes

Inferring p



n independent Bernoulli processes

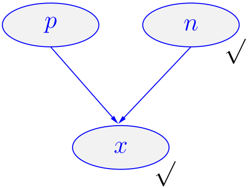
Inferring p



$$f(p\,|\,x,n) \;=\; \frac{f(p,x\,|\,n)}{f(x\,|\,n)}$$

n independent Bernoulli processes

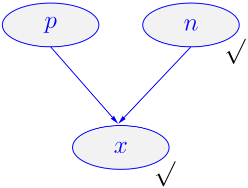
Inferring p



$$f(p\,|\,x,n) = \frac{f(p,x\,|\,n)}{f(x\,|\,n)} = \frac{f(x\,|\,n,p)\cdot f_0(p)}{f(x\,|\,n)}$$

n independent Bernoulli processes

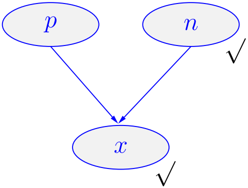
Inferring p



$$
\begin{aligned}
f(p \,|\, x, n) &= \frac{f(p, x \,|\, n)}{f(x \,|\, n)} \\
&= \frac{f(x \,|\, n, p) \cdot f_0(p)}{f(x \,|\, n)} \\
&= \frac{f(x \,|\, n, p) \cdot f_0(p)}{\int_0^1 f(x \,|\, n, p) \cdot f_0(p)\, dp}
\end{aligned}
$$

n independent Bernoulli processes

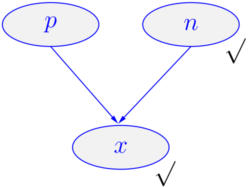
Inferring p



$$
\begin{aligned}
f(p\,|\,x,n) &= \frac{f(p,x\,|\,n)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{\int_0^1 f(x\,|\,n,p)\cdot f_0(p)\,dp} \\
&\propto f(x\,|\,n,p)\cdot f_0(p)
\end{aligned}
$$

n independent Bernoulli processes

Inferring p



$$
\begin{aligned}
f(p\,|\,x,n) &= \frac{f(p,x\,|\,n)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{\int_0^1 f(x\,|\,n,p)\cdot f_0(p)\,dp} \\
&\propto f(x\,|\,n,p)\cdot f_0(p)
\end{aligned}
$$

(denominator just normalization!)

Inferring "Bernoulli's p"

We just need to make explicit $f(x \,|\, n, p)$:

$$f(x \,|\, n, p) \;=\; \binom{n}{x}\, p^x \,(1-p)^{n-x}$$

Inferring "Bernoulli's p"

We just need to make explicit $f(x \mid n, p)$:

$$f(x \mid n, p) \;=\; \binom{n}{x} p^x (1-p)^{n-x} \;=\; \frac{n!}{(n-x)!\, x!}\, p^x (1-p)^{n-x}$$

Inferring "Bernoulli's p"

We just need to make explicit $f(x \,|\, n, p)$:

$$f(x \,|\, n, p) \;=\; \binom{n}{x}\, p^x\,(1-p)^{n-x} \;=\; \frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}$$

We get then, including normalization:

$$f(p \,|\, x, n) \;=\; \frac{\frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}\, f_\circ(p)}{\int_0^1 \frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}\, f_\circ(p)\,\mathrm{d}p}$$

Inferring "Bernoulli's p"

We just need to make explicit $f(x \,|\, n, p)$:

$$f(x \,|\, n, p) \;=\; \binom{n}{x}\, p^x\,(1-p)^{n-x} \;=\; \frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}$$

We get then, including normalization:

$$f(p \,|\, x, n) \;=\; \frac{\frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}\, f_\circ(p)}{\int_0^1 \frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}\, f_\circ(p)\,\mathrm{d}p}$$

$$= \frac{p^x\,(1-p)^{n-x}\, f_\circ(p)}{\int_0^1 p^x\,(1-p)^{n-x}\, f_\circ(p)\,\mathrm{d}p}$$

Inferring "Bernoulli's p"

We just need to make explicit $f(x \,|\, n, p)$:

$$f(x \,|\, n, p) \;=\; \binom{n}{x}\, p^x\,(1-p)^{n-x} \;=\; \frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}$$

We get then, including normalization:

$$f(p \,|\, x, n) \;=\; \frac{\frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}\, f_\circ(p)}{\int_0^1 \frac{n!}{(n-x)!\,x!}\, p^x\,(1-p)^{n-x}\, f_\circ(p)\,\mathrm{d}p}$$

$$\;=\; \frac{p^x\,(1-p)^{n-x}\, f_\circ(p)}{\int_0^1 p^x\,(1-p)^{n-x}\, f_\circ(p)\,\mathrm{d}p}$$

(The binomial coefficient is irrelevant, not depending on p)

Inferring "Bernoulli's p"

$$f(p \,|\, x, n) \;=\; \frac{p^x\,(1-p)^{n-x}\, f_\circ(p)}{\int_0^1 p^x\,(1-p)^{n-x}\, f_\circ(p)\,\mathrm{d}p}$$

Inferring "Bernoulli's p"

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}\,f_\circ(p)}{\int_0^1 p^x\,(1-p)^{n-x}\,f_\circ(p)\,\mathrm{d}p}$$

For teaching purposes we start from a **uniform prior**, i.e. $\boldsymbol{f_\circ(p) = 1}$:

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}}{\int_0^1 p^x\,(1-p)^{n-x}\,\mathrm{d}p}$$

Inferring "Bernoulli's p"

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}\,f_\circ(p)}{\int_0^1 p^x\,(1-p)^{n-x}\,f_\circ(p)\,\mathrm{d}p}$$

For teaching purposes we start from a **uniform prior**, i.e. $\boldsymbol{f_\circ(p) = 1}$:

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}}{\int_0^1 p^x\,(1-p)^{n-x}\,\mathrm{d}p}$$

- The integral at the denominator is the special function "β" (also defined for real values of x and n).

Inferring "Bernoulli's p"

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}\,f_\circ(p)}{\int_0^1 p^x\,(1-p)^{n-x}\,f_\circ(p)\,\mathrm{d}p}$$

For teaching purposes we start from a **uniform prior**, i.e. $\boldsymbol{f_\circ(p) = 1}$:

$$f(p\,|\,x,n) \;=\; \frac{p^x\,(1-p)^{n-x}}{\int_0^1 p^x\,(1-p)^{n-x}\,\mathrm{d}p}$$

- The integral at the denominator is the special function "β" (also defined for real values of x and n).
- In our case these two numbers are integer and the integral becomes equal to

$$\frac{x!\,(n-x)!}{(n+1)!}$$

Inferring "Bernoulli's p"

Solution for uniform prior (think to **Bayes' billard**)

$$f(p\,|\,x,n) \;=\; \frac{(n+1)!}{x!\,(n-x)!}\,p^x\,(1-p)^{n-x}$$

Inferring "Bernoulli's p"

Solution for uniform prior (think to **Bayes' billard**)

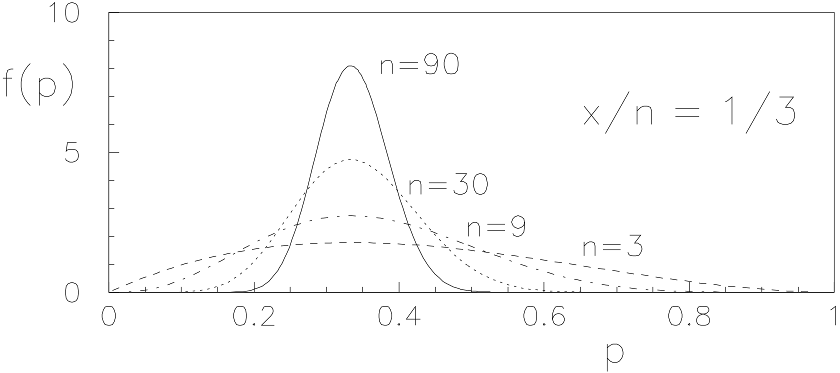
$$
\begin{aligned}
f(p\,|\,x,n) &= \frac{(n+1)!}{x!\,(n-x)!}\,p^x\,(1-p)^{n-x} \\
&= (n+1)\cdot\frac{n!}{x!\,(n-x)!}\,p^x\,(1-p)^{n-x}
\end{aligned}
$$

Inferring "Bernoulli's p"

Solution for uniform prior (think to **Bayes' billard**)

$$f(p \,|\, x, n) = \frac{(n+1)!}{x!\,(n-x)!}\, p^x\,(1-p)^{n-x}$$

$$= (n+1)\cdot\frac{n!}{x!\,(n-x)!}\, p^x\,(1-p)^{n-x}$$



Inferring "Bernoulli's p"

Summaries of the **posterior distribution**

$$p_m = \text{mode}(p) = \frac{x}{n}$$

Inferring "Bernoulli's p"

Summaries of the **posterior distribution**

$$
\begin{aligned}
p_m = \mathrm{mode}(p) &= \frac{x}{n} \\
\mathrm{E}(p) &= \frac{x+1}{n+2}
\end{aligned}
$$

Inferring "Bernoulli's p"

Summaries of the **posterior distribution**

$$p_m = \text{mode}(p) = \frac{x}{n}$$

$$E(p) = \frac{x+1}{n+2}$$

"recursive Laplace formula"

("Laplace's rule of succession")

Inferring “Bernoulli's p”

Summaries of the **posterior distribution**

$$p_m = \text{mode}(p) = \frac{x}{n}$$

$$E(p) = \frac{x+1}{n+2}$$

“recursive Laplace formula”

(“Laplace's rule of succession”)

$$\begin{aligned}
\text{Var}(p) &= \frac{(x+1)(n-x+1)}{(n+3)(n+2)^2} \\
&= \frac{x+1}{n+2}\left(\frac{n+2}{n+2} - \frac{x+1}{n+2}\right)\frac{1}{n+3} \\
&= E(p)\,(1-E(p))\,\frac{1}{n+3}
\end{aligned}$$

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".

  [$^{(*)}$ "First" does not imply time order, but just order in usage.]

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".

  [$^{(*)}$ "First" does not imply time order, but just order in usage.]

- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p,

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i \,|\, p) = p$$

Inferring the "Bernoulli's p"

About the meaning of $E(p)$

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i \,|\, p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

Inferring the "Bernoulli's p"

About the meaning of $\mathrm{E}(p)$

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- <u>If</u> we were sure about p, <u>then</u> p would be our probability:

$$P(E_i \,|\, p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

$$P(E_{i>n} \,|\, x, n) = \int_0^1 P(E_i \,|\, p) \cdot f(p \,|\, x, n)\, \mathrm{d}p$$

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i \,|\, p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

$$\begin{aligned} P(E_{i>n} \,|\, x, n) &= \int_0^1 P(E_i \,|\, p) \cdot f(p \,|\, x, n)\, \mathrm{d}p \\ &= \int_0^1 p \cdot f(p \,|\, x, n)\, \mathrm{d}p \end{aligned}$$

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i \,|\, p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

$$\begin{aligned} P(E_{i>n} \,|\, x, n) &= \int_0^1 P(E_i \,|\, p) \cdot f(p \,|\, x, n)\, \mathrm{d}p \\ &= \int_0^1 p \cdot f(p \,|\, x, n)\, \mathrm{d}p = \mathrm{E}(p) \quad (!!) \end{aligned}$$

Inferring the "Bernoulli's p"

About the meaning of $\mathrm{E}(p)$

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i\,|\,p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

$$\begin{aligned} P(E_{i>n}\,|\,x,n) &= \int_0^1 P(E_i\,|\,p)\cdot f(p\,|\,x,n)\,\mathrm{d}p \\ &= \int_0^1 p\cdot f(p\,|\,x,n)\,\mathrm{d}p = \mathrm{E}(p) \quad (!!) \end{aligned}$$

$\mathrm{E}(p)$ (and not the mode!) is the probability of every 'future' event which is believed to have the **same p** of the 'previous' ones.

Inferring the "Bernoulli's p"

About the meaning of E(p)

- We have used the "first"$^{(*)}$ n trials to learn about "p".
  [$^{(*)}$ "First" does not imply time order, but just order in usage.]
- What will be the probability of other trials?

$$P(E_{i>n}) = ??$$

- If we were sure about p, then p would be our probability:

$$P(E_i \,|\, p) = p$$

- But since we are uncertain about it, we have to take into account all possible values, weighing them with our degree of belief.

$$\begin{aligned} P(E_{i>n} \,|\, x, n) &= \int_0^1 P(E_i \,|\, p) \cdot f(p \,|\, x, n)\, \mathrm{d}p \\ &= \int_0^1 p \cdot f(p \,|\, x, n)\, \mathrm{d}p = \mathrm{E}(p) \quad (!!) \end{aligned}$$

E(p) (and not the mode!) is the probability of every 'future' event which is believed to have the **same *p*** of the 'previous' ones.
(But keep in mind the **inductivist turkey**!)

Inferring the "Bernoulli's p"

Large number behaviour

When the number of successes and the number of failures become 'large' (x large is not enough, as it can be easily understood from the simmetric properties of the binomial $p \leftrightarrow q$):

Inferring the "Bernoulli's p"

Large number behaviour

When the number of successes and the number of failures become 'large' (x large is not enough, as it can be easily understood from the simmetric properties of the binomial $p \leftrightarrow q$):

$$\mathrm{E}(p) \approx p_m = \frac{x}{n}$$

Inferring the "Bernoulli's p"

Large number behaviour

When the number of successes <u>and</u> the number of failures become 'large' (x large is not enough, as it can be easily understood from the simmetric properties of the binomial $p \leftrightarrow q$):

$$\begin{aligned} \mathrm{E}(p) &\approx p_m = \frac{x}{n} \\ \mathrm{Var}(p) &\approx \frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n} = \frac{p_m\,(1-p_m)}{n} \end{aligned}$$

Inferring the "Bernoulli's p"

Large number behaviour

When the number of successes and the number of failures become 'large' (x large is not enough, as it can be easily understood from the simmetric properties of the binomial $p \leftrightarrow q$):

$$\begin{aligned} \mathrm{E}(p) &\approx p_m = \frac{x}{n} \\ \mathrm{Var}(p) &\approx \frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n} = \frac{p_m\,(1-p_m)}{n} \\ \sigma(p)(=\sigma_p) &\approx \sqrt{\frac{p_m\,(1-p_m)}{n}} \propto \frac{1}{\sqrt{n}} \end{aligned}$$

Inferring the "Bernoulli's p"

Large number behaviour

When the number of successes and the number of failures become 'large' (x large is not enough, as it can be easily understood from the simmetric properties of the binomial $p \leftrightarrow q$):

$$\begin{aligned} \mathrm{E}(p) &\approx p_m = \frac{x}{n} \\ \mathrm{Var}(p) &\approx \frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n} = \frac{p_m\,(1-p_m)}{n} \\ \sigma(p)(=\sigma_p) &\approx \sqrt{\frac{p_m\,(1-p_m)}{n}} \propto \frac{1}{\sqrt{n}} \end{aligned}$$

Moreover $f(p)$ tends to a Gaussian distribution:

$$p \sim \mathcal{N}(p_m, \sigma_p)$$

When the number of successes and the number of failures become 'large' (x large is not enough, as it can be easily understood from the simmetric properties of the binomial $p \leftrightarrow q$):

$$\begin{aligned} \mathrm{E}(p) &\approx p_m = \frac{x}{n} \\ \mathrm{Var}(p) &\approx \frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n} = \frac{p_m\,(1-p_m)}{n} \\ \sigma(p)(=\sigma_p) &\approx \sqrt{\frac{p_m\,(1-p_m)}{n}} \propto \frac{1}{\sqrt{n}} \end{aligned}$$

Moreover $f(p)$ tends to a Gaussian distribution:

$$p \sim \mathcal{N}(p_m, \sigma_p)$$

When $n \to \infty$, then $\sigma_p \to 0$,

Inferring the "Bernoulli's p"

Large number behaviour

When the number of successes and the number of failures become 'large' (x large is not enough, as it can be easily understood from the simmetric properties of the binomial $p \leftrightarrow q$):

$$\begin{aligned} \mathrm{E}(p) &\approx p_m = \frac{x}{n} \\ \mathrm{Var}(p) &\approx \frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n} = \frac{p_m\,(1-p_m)}{n} \\ \sigma(p)(=\sigma_p) &\approx \sqrt{\frac{p_m\,(1-p_m)}{n}} \propto \frac{1}{\sqrt{n}} \end{aligned}$$

Moreover $f(p)$ tends to a Gaussian distribution:

$$p \sim \mathcal{N}(p_m, \sigma_p)$$

When $n \to \infty$, then $\sigma_p \to 0$, $\to$ and hence

$$P(E_{i>n}\,|\,x,n) \text{ “}\longrightarrow\text{” } \frac{x}{n}$$

Inferring the "Bernoulli's p"

Large number behaviour

When the number of successes and the number of failures become 'large' (x large is not enough, as it can be easily understood from the simmetric properties of the binomial $p \leftrightarrow q$):

$$\begin{aligned} \mathrm{E}(p) &\approx p_m = \frac{x}{n} \\ \mathrm{Var}(p) &\approx \frac{x}{n}\left(1-\frac{x}{n}\right)\frac{1}{n} = \frac{p_m\,(1-p_m)}{n} \\ \sigma(p)(=\sigma_p) &\approx \sqrt{\frac{p_m\,(1-p_m)}{n}} \propto \frac{1}{\sqrt{n}} \end{aligned}$$

Moreover $f(p)$ tends to a Gaussian distribution:

$$p \sim \mathcal{N}(p_m, \sigma_p)$$

When $n \rightarrow \infty$, then $\sigma_p \rightarrow 0$, $\rightarrow$ and hence

$$P(E_{i>n}\,|\,x, n) \text{ “}\longrightarrow\text{” } \frac{x}{n}$$

(Similarly to Bernoulli's theorem, it is not a 'mathematical' limit!)

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- and $(n - x)$ large

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- <u>and</u> $(n - x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1 - p$)

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- and $(n-x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1-p$),

then

$$\mathrm{E}(p) \approx \frac{x}{n}$$

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- and $(n - x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1 - p$),

then

$$\begin{aligned} \mathrm{E}(p) &\approx \frac{x}{n} \\ \sigma(p) &\approx \frac{1}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1 - \frac{x}{n}\right)} \end{aligned}$$

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- <u>and</u> $(n-x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1-p$),

then

$$\begin{aligned} \mathrm{E}(p) &\approx \frac{x}{n} \\ \sigma(p) &\approx \frac{1}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)} \end{aligned}$$

— $f(p\,|\,x, n)$ tends to Gaussian

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- and $(n-x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1-p$),

then

$$\begin{aligned} \mathrm{E}(p) &\approx \frac{x}{n} \\ \sigma(p) &\approx \frac{1}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)} \end{aligned}$$

— $f(p\,|\,x, n)$ tends to Gaussian,
a reflection of the Gaussian limit of $f(x\,|\,p, n)$

Inferring the "Bernoulli's p"

When

- n large;
- x large;
- <u>and</u> $(n - x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1 - p$),

then

$$\begin{aligned} \mathrm{E}(p) &\approx \frac{x}{n} \\ \sigma(p) &\approx \frac{1}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)} \end{aligned}$$

— $f(p \,|\, x, n)$ tends to Gaussian,
a reflection of the Gaussian limit of $f(x \,|\, p, n)$

— **The probability of a future events is evaluated from the relative frequency of the past events**

Inferring the "Bernoulli's p"

Large number behaviour: summary

When

- n large;
- x large;
- and $(n-x)$ large
  (remember: in the binomial what is 'success' and what is 'failure' is not absolute: $p \longleftrightarrow q = 1-p$),

then

$$\mathrm{E}(p) \approx \frac{x}{n}$$

$$\sigma(p) \approx \frac{1}{\sqrt{n}}\sqrt{\frac{x}{n}\left(1-\frac{x}{n}\right)}$$

— $f(p\,|\,x,n)$ tends to Gaussian,
a reflection of the Gaussian limit of $f(x\,|\,p,n)$

— **The probability of a future events is evaluated from the relative frequency of the past events**

— **No need of 'frequentistic definition'!**

Probability vs frequency

Frequency and probability are **related** in probability theory:

Probability vs frequency

Frequency and probability are **related** in probability theory:

- Relative frequencies of successes in future trials can be 'forecasted' from *p*

Probability vs frequency

Frequency and probability are **related** in probability theory:

- Relative frequencies of successes in future trials can be 'forecasted' from p (Bernoulli theorem).

Probability vs frequency

Frequency and probability are **related** in probability theory:

- Relative frequencies of successes in future trials can be 'forecasted' from p (Bernoulli theorem).

- Probability p can be evaluated from past frequencies, under some assumptions

Probability vs frequency

Frequency and probability are **related** in probability theory:

- Relative frequencies of successes in future trials can be 'forecasted' from p (Bernoulli theorem).

- Probability p can be evaluated from past frequencies, under some assumptions ('Bayes theorem' $\rightarrow$ **Laplace's rule**)

Probability vs frequency

Frequency and probability are **related** in probability theory:

- Relative frequencies of successes in future trials can be 'forecasted' from p (Bernoulli theorem).

- Probability p can be evaluated from past frequencies, under some assumptions ('Bayes theorem' $\rightarrow$ **Laplace's rule**)

BUT

- There is no need to identify the two concepts.

Probability vs frequency

Frequency and probability are **related** in probability theory:

- Relative frequencies of successes in future trials can be 'forecasted' from p (Bernoulli theorem).
- Probability p can be evaluated from past frequencies, under some assumptions ('Bayes theorem' $\rightarrow$ **Laplace's rule**)

BUT

- There is no need to identify the two concepts.
- It does not justify the frequentistic definition.

Propagation of **errors** in the evaluation of efficiency

From a recent 'tesi di laurea' in Rome ('quadriennale') (undergraduate thesis)

Da questa analisi si ottengono

$$\mathrm{N} = (82502 \pm 287) \qquad \mathrm{n_S} = (82378 \pm 287).$$

dove $\sigma_{N(n)} = \sqrt{N(n)}$.
Da N e $\mathrm{n_S}$ si ricava il valore dell'efficienza in Pos 1:

$$\epsilon_{\mathrm{S(Pos1)}} = \frac{\mathrm{n_S}}{\mathrm{N}} = (99.847)\%$$

Propagation of **errors** in the evaluation of efficiency

From a recent 'tesi di laurea' in Rome ('quadriennale') (undergraduate thesis)

Da questa analisi si ottengono

$$N = (82502 \pm 287) \qquad n_S = (82378 \pm 287).$$

dove $\sigma_{N(n)} = \sqrt{N(n)}$.
Da N e n_S si ricava il valore dell'efficienza in Pos 1:

$$\epsilon_{S(Pos1)} = \frac{n_S}{N} = (99.847)\%$$

σ_N: ???
σ_n: ???
(hereafter $n_s \rightarrow n$)

Propagation of **errors** in the evaluation of efficiency

From a recent 'tesi di laurea' in Rome ('quadriennale') (undergraduate thesis)

Da questa analisi si ottengono

$$N = (82502 \pm 287) \qquad n_S = (82378 \pm 287).$$

dove $\sigma_{N(n)} = \sqrt{N(n)}$.

Da N e n_S si ricava il valore dell'efficienza in Pos 1:

$$\epsilon_{S(Pos1)} = \frac{n_S}{N} = (99.847)\%$$

σ_N: ???

σ_n: ???

(hereafter $n_s \rightarrow n$)

- $N - n = 124$

  $\rightarrow$ with $\sigma_N = \sigma_n = 287$:

Propagation of **errors** in the evaluation of efficiency

From a recent 'tesi di laurea' in Rome ('quadriennale') (undergraduate thesis)

Da questa analisi si ottengono

$$N = (82502 \pm 287) \qquad n_S = (82378 \pm 287).$$

dove $\sigma_{N(n)} = \sqrt{N(n)}$.

Da N e n_S si ricava il valore dell'efficienza in Pos 1:

$$\epsilon_{S(Pos1)} = \frac{n_S}{N} = (99.847)\%$$

σ_N: ???

σ_n: ???

(hereafter $n_s \to n$)

- $N - n = 124$

  $\to$ with $\sigma_N = \sigma_n = 287$: efficiency could be > 1:

Propagation of **errors** in the evaluation of efficiency

From a recent 'tesi di laurea' in Rome ('quadriennale') (undergraduate thesis)

Da questa analisi si ottengono

$$\mathrm{N} = (82502 \pm 287) \qquad \mathrm{n_S} = (82378 \pm 287).$$

dove $\sigma_{N(n)} = \sqrt{N(n)}$.
Da N e $\mathrm{n_S}$ si ricava il valore dell'efficienza in Pos 1:

$$\epsilon_{\mathrm{S(Pos1)}} = \frac{\mathrm{n_S}}{\mathrm{N}} = (99.847)\%$$

σ_N: ???
σ_n: ???
(hereafter $n_s \rightarrow n$)

- $N - n = 124$
  $\rightarrow$ with $\sigma_N = \sigma_n = 287$: efficiency could be > 1: **???**

Propagation of **errors** in the evaluation of efficiency

- Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:

$$\sigma = \frac{1}{N}\sqrt{n + \frac{n^2}{N}} \qquad (4.51)$$

Propagation of **errors** in the evaluation of efficiency

- Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:

$$\sigma = \frac{1}{N}\sqrt{n + \frac{n^2}{N}} \qquad (4.51)$$

- Statistical ‘error’

Propagation of **errors** in the evaluation of efficiency

- Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:

$$\sigma = \frac{1}{N}\sqrt{n + \frac{n^2}{N}} \tag{4.51}$$

- Statistical 'error' (meant as 'uncertainty') obtained propagating the **errors**

Propagation of **errors** in the evaluation of efficiency

- Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:

$$\sigma = \frac{1}{N}\sqrt{n + \frac{n^2}{N}} \qquad (4.51)$$

- Statistical 'error' (meant as 'uncertainty') obtained propagating the **errors** (this time they are really errors)...

Propagation of **errors** in the evaluation of efficiency

- Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:

$$\sigma = \frac{1}{N}\sqrt{n + \frac{n^2}{N}} \qquad (4.51)$$

- Statistical 'error' (meant as 'uncertainty') obtained propagating the **errors** (this time they are really errors)...
- ...from two independent experimental values (N, n)

Propagation of **errors** in the evaluation of efficiency

- Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:

$$\sigma = \frac{1}{N}\sqrt{n + \frac{n^2}{N}} \qquad (4.51)$$

- Statistical 'error' (meant as 'uncertainty') obtained propagating the **errors** (this time they are really errors)...
- ...from two independent experimental values (N, n) **???**

Propagation of **errors** in the evaluation of efficiency

> • Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:
>
> $$\sigma = \frac{1}{N}\sqrt{n + \frac{n^2}{N}} \qquad (4.51)$$

- Statistical 'error' (meant as 'uncertainty') obtained propagating the **errors** (this time they are really errors)...
- ...from two independent experimental values (N, n) **???**
- Eq. (4.51) correctly follows

Propagation of **errors** in the evaluation of efficiency

- Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:

$$\sigma = \frac{1}{N}\sqrt{n + \frac{n^2}{N}} \qquad (4.51)$$

- Statistical 'error' (meant as 'uncertainty') obtained propagating the **errors** (this time they are really errors)...
- ...from two independent experimental values (N, n) **???**
- Eq. (4.51) correctly follows from the bad reasoning $\surd$

Propagation of **errors** in the evaluation of efficiency

- Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:

$$\sigma = \frac{1}{\mathrm{N}}\sqrt{\mathrm{n} + \frac{\mathrm{n}^2}{\mathrm{N}}} \tag{4.51}$$

- Statistical 'error' (meant as 'uncertainty') obtained propagating the **errors** (this time they are really errors)...
- ...from two independent experimental values (N, n) **???**
- Eq. (4.51) correctly follows from the bad reasoning $\surd$
  $\rightarrow \sigma_\epsilon = 0.0049 \approx 0.005$

Propagation of **errors** in the evaluation of efficiency

- Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:

$$\sigma = \frac{1}{N}\sqrt{n + \frac{n^2}{N}} \qquad (4.51)$$

- Statistical 'error' (meant as 'uncertainty') obtained propagating the **errors** (this time they are really errors)...
- ...from two independent experimental values (N, n) **???**
- Eq. (4.51) correctly follows from the bad reasoning $\surd$
  $\rightarrow \sigma_\epsilon = 0.0049 \approx 0.005$
- How much is it wrong?

$$\frac{\sigma(\epsilon)^{wrong}}{\sigma(\epsilon)^{correct}} = \frac{1/\sqrt{N}\sqrt{n/N\cdot(1+n/N)}}{1/\sqrt{N}\sqrt{n/N\cdot(1-n/N)}}$$

Propagation of **errors** in the evaluation of efficiency

> • Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:
>
> $$\sigma = \frac{1}{\mathrm{N}}\sqrt{\mathrm{n} + \frac{\mathrm{n}^2}{\mathrm{N}}} \qquad (4.51)$$

- Statistical 'error' (meant as 'uncertainty') obtained propagating the **errors** (this time they are really errors)...
- ...from two independent experimental values (N, n) **???**
- Eq. (4.51) correctly follows from the bad reasoning $\surd$
  $\rightarrow \sigma_\epsilon = 0.0049 \approx 0.005$
- How much is it wrong?

$$\frac{\sigma(\epsilon)^{wrong}}{\sigma(\epsilon)^{correct}} = \frac{1/\sqrt{N}\sqrt{n/N\cdot(1+n/N)}}{1/\sqrt{N}\sqrt{n/N\cdot(1-n/N)}} = \sqrt{\frac{1+\epsilon_m}{1-\epsilon_m}}$$

Propagation of **errors** in the evaluation of efficiency

- Errore statistico. La stima di questo errore si ottiene con la propagazione degli errori di una funzione di due variabili sperimentali indipendenti (N,n) e si ricava dalla seguente espressione:

$$\sigma = \frac{1}{\mathrm{N}}\sqrt{\mathrm{n} + \frac{\mathrm{n}^2}{\mathrm{N}}} \qquad (4.51)$$

- Statistical 'error' (meant as 'uncertainty') obtained propagating the **errors** (this time they are really errors)...
- ...from two independent experimental values (N, n) **???**
- Eq. (4.51) correctly follows from the bad reasoning $\surd$
  $\rightarrow \sigma_\epsilon = 0.0049 \approx 0.005$
- How much is it wrong?

$$\frac{\sigma(\epsilon)^{wrong}}{\sigma(\epsilon)^{correct}} = \frac{1/\sqrt{N}\sqrt{n/N\cdot(1+n/N)}}{1/\sqrt{N}\sqrt{n/N\cdot(1-n/N)}} = \sqrt{\frac{1+\epsilon_m}{1-\epsilon_m}}$$
$$= 36$$

Propagation of errors. . . and of mistakes

Eseguendo queste operazioni otteniamo il seguente risultato:

$$\epsilon_{\mathrm{S(Pos1)}} = (99.847 \pm 0,005^{(\mathrm{stat})} \pm 0,010^{(\mathrm{sist})})\%.$$

Propagation of errors. . . and of mistakes

Eseguendo queste operazioni otteniamo il seguente risultato:

$$\epsilon_{\mathrm{S(Pos1)}} = (99.847 \pm 0,005^{(\mathrm{stat})} \pm 0,010^{(\mathrm{sist})})\%.$$

- Finally,
  the wrong $\sigma_\epsilon = 0.005 = 0.5\%$ becomes 0.005%

Propagation of errors. . . and of mistakes

> Eseguendo queste operazioni otteniamo il seguente risultato:
>
> $$\epsilon_{\mathrm{S(Pos1)}} = (99.847 \pm 0,005^{(\mathrm{stat})} \pm 0,010^{(\mathrm{sist})})\%.$$

- Finally,
  the wrong $\sigma_\epsilon = 0.005 = 0.5\%$ becomes $0.005\% = 0.00005$

Propagation of errors. . . and of mistakes

Eseguendo queste operazioni otteniamo il seguente risultato:

$$\epsilon_{S(Pos1)} = (99.847 \pm 0,005^{(stat)} \pm 0,010^{(sist)})\%.$$

- Finally,
  the wrong $\sigma_\epsilon = 0.005 = 0.5\%$ becomes $0.005\% = 0.00005$

Propagation of errors. . . and of mistakes

Eseguendo queste operazioni otteniamo il seguente risultato:

$$\epsilon_{\mathrm{S(Pos1)}} = (99.847 \pm 0,005^{\mathrm{(stat)}} \pm 0,010^{\mathrm{(sist)}})\%.$$

- Finally,
  the wrong $\sigma_\epsilon = 0.005 = 0.5\%$ becomes $0.005\% = 0.00005$

Good luck to the experiment!

Inferring the "Bernoulli's p"

Approximate solution using the 'Gaussian trick'

Exercise

- Given $f(p) \propto p^x\,(1-p)^{n-x}$,
- define $\varphi(p) = -\ln f(p)$
- and evaluate
  - $\frac{d\varphi}{dp}$
  - $\frac{d^2\varphi}{dp^2}$
- Then estimate
  - $\mathrm{E}(p) \approx p_m$ from minimum;
  - $\sigma^2(p)$ from second derivative at the minimum.

Inferring "Bernoulli's p"

Observing $x = 0$

$$f(0\,|\,\mathcal{B}_{n,p}) \;=\; (1-p)^n,$$

Inferring "Bernoulli's p"

Observing $x = 0$

$$f(0\,|\,\mathcal{B}_{n,p}) = (1-p)^n,$$

$$f(p\,|\,x=0,n,\mathcal{B}) = \frac{(1-p)^n}{\int_0^1 (1-p)^n\,dp} = (n+1)\,(1-p)^n,$$

Inferring "Bernoulli's p"

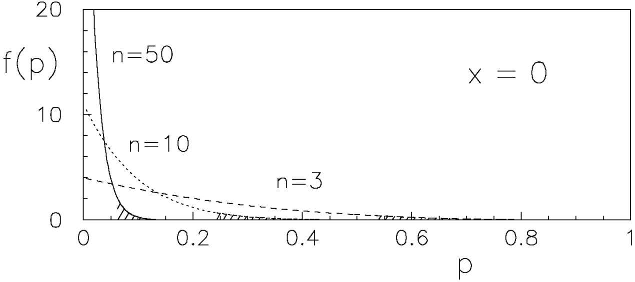
Observing $x = 0$

$$
\begin{aligned}
f(0\,|\,\mathcal{B}_{n,p}) &= (1-p)^n, \\
f(p\,|\,x=0,n,\mathcal{B}) &= \frac{(1-p)^n}{\int_0^1 (1-p)^n\,dp} = (n+1)\,(1-p)^n, \\
F(p\,|\,x=0,n,\mathcal{B}) &= 1-(1-p)^{n+1}\,.
\end{aligned}
$$

Inferring "Bernoulli's p"

Observing $x = 0$

$$
\begin{aligned}
f(0\,|\,\mathcal{B}_{n,p}) &= (1-p)^n, \\
f(p\,|\,x=0,n,\mathcal{B}) &= \frac{(1-p)^n}{\int_0^1 (1-p)^n\,dp} = (n+1)\,(1-p)^n, \\
F(p\,|\,x=0,n,\mathcal{B}) &= 1-(1-p)^{n+1}\,.
\end{aligned}
$$



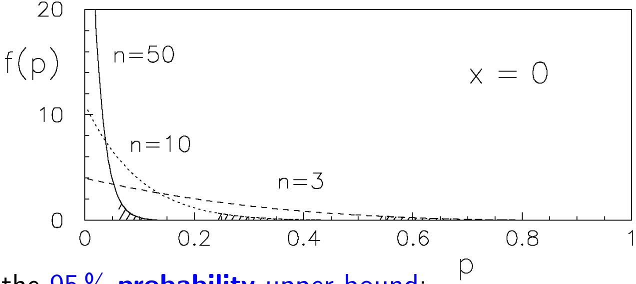
Inferring "Bernoulli's p"

Observing $x = 0$

$$f(0\,|\,\mathcal{B}_{n,p}) = (1-p)^n,$$

$$f(p\,|\,x=0,n,\mathcal{B}) = \frac{(1-p)^n}{\int_0^1 (1-p)^n\,dp} = (n+1)\,(1-p)^n,$$

$$F(p\,|\,x=0,n,\mathcal{B}) = 1-(1-p)^{n+1}\,.$$



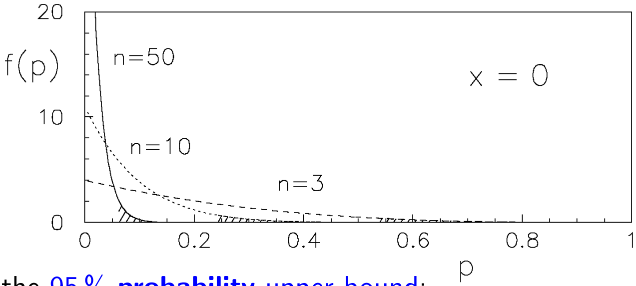
To get the 95 % **probability** upper bound:

$$F(p_\circ\,|\,x=0,n,\mathcal{B}) = 0.95,$$

Inferring "Bernoulli's p"

Observing $x = 0$

$$\begin{aligned} f(0\,|\,\mathcal{B}_{n,p}) &= (1-p)^n, \\ f(p\,|\,x=0,n,\mathcal{B}) &= \frac{(1-p)^n}{\int_0^1 (1-p)^n\,dp} = (n+1)\,(1-p)^n, \\ F(p\,|\,x=0,n,\mathcal{B}) &= 1-(1-p)^{n+1}\,. \end{aligned}$$



To get the 95 % **probability** upper bound:

$$\begin{aligned} F(p_\circ\,|\,x=0,n,\mathcal{B}) &= 0.95, \\ p_\circ &= 1-\sqrt[n+1]{0.05}\,. \end{aligned}$$

Inferring "Bernoulli's p"

Observing $x = n$

$$f(n \,|\, \mathcal{B}_{n,p}) \;=\; p^n$$

Inferring "Bernoulli's p"

Observing $x = n$

$$
\begin{aligned}
f(n \,|\, \mathcal{B}_{n,p}) &= p^n \\
f(p \,|\, x = n, \mathcal{B}) &= \frac{p^n}{\int_0^1 p^n \, \mathrm{d}p} = (n+1)\, p^n
\end{aligned}
$$

Inferring "Bernoulli's p"

Observing $x = n$

$$
\begin{aligned}
f(n \,|\, \mathcal{B}_{n,p}) &= p^n \\
f(p \,|\, x = n, \mathcal{B}) &= \frac{p^n}{\int_0^1 p^n \, \mathrm{d}p} = (n+1)\, p^n \\
F(p \,|\, x = n, \mathcal{B}) &= p^{n+1}
\end{aligned}
$$

95% **probability** lower bound

$$
\begin{aligned}
F(p_\circ \,|\, x = n, \mathcal{B}) &= 0.05\,, \\
p_\circ &= \sqrt[n+1]{0.05}\,.
\end{aligned}
$$

Inferring the "Bernoulli's p"

A glance to upper/lower probabilistic limits

	Probability level = 95 %		
n	$x = n$	$x = 0$	
	binomial	binomial	Poisson approx. ($p_\circ = 3/n$)
3	$p \geq 0.47$	$p \leq 0.53$	$p \leq 1$
5	$p \geq 0.61$	$p \leq 0.39$	$p \leq 0.6$
10	$p \geq 0.76$	$p \leq 0.24$	$p \leq 0.3$
50	$p \geq 0.94$	$p \leq 0.057$	$p \leq 0.06$
100	$p \geq 0.97$	$p \leq 0.029$	$p \leq 0.03$
1000	$p \geq 0.997$	$p \leq 0.003$	$p \leq 0.003$

Upper/lower probability bounds – caveat!

The probabilistic upper/lower bounds of the previous slides depend on the assumption $f(p) = 1$

Upper/lower probability bounds – caveat!

The probabilistic upper/lower bounds of the previous slides depend on the assumption $f(p) = 1$

- A flat prior seems armless. . .

Upper/lower probability bounds – caveat!

The probabilistic upper/lower bounds of the previous slides depend on the assumption $f(p) = 1$

- A flat prior seems armless...
- ...but **it isn't!**

Upper/lower probability bounds – caveat!

The probabilistic upper/lower bounds of the previous slides depend on the assumption $f(p) = 1$

- A flat prior seems armless. . .
- . . . but **it isn't!**
- Imagine that p refers to a branching ratio:
  $f_0(p) = 1$ implies

$$P(p \leq 0.1) = P(p \geq 0.9)$$

$$P(p \leq 0.01) \, [= P(p \geq 0.99)]$$

Upper/lower probability bounds – caveat!

The probabilistic upper/lower bounds of the previous slides depend on the assumption $f(p) = 1$

- A flat prior seems armless...
- ...but it isn't!
- Imagine that p refers to a branching ratio:
  $f_0(p) = 1$ implies

$$\begin{aligned} P(p \leq 0.1) &= P(p \geq 0.9) \\ P(p \leq 0.01)\,[= P(p \geq 0.99)] &= \frac{1}{10}\,P(p \leq 0.1) \\ \ldots &\quad \ldots \end{aligned}$$

Really do you believe so?

Upper/lower probability bounds – caveat!

The probabilistic upper/lower bounds of the previous slides depend on the assumption $f(p) = 1$

- A flat prior seems armless…
- …but it isn't!
- Imagine that p refers to a branching ratio: $f_0(p) = 1$ implies

$$
\begin{aligned}
P(p \leq 0.1) &= P(p \geq 0.9) \\
P(p \leq 0.01) \, [= P(p \geq 0.99)] &= \frac{1}{10} P(p \leq 0.1) \\
\ldots &\quad \ldots
\end{aligned}
$$

Really do you believe so?

Exercise: try to plot $f(p \,|\, x = 0, n = 100)$ in log-log scale

```
> p=10^seq(-5,-1,len=100);
> plot(p, (1-p)^100, ty='l', log='xy'); grid()
```

(and think about it!)

Very rare processes

Sensitivity bounds: some hints for **self study**

Let us restart from the Bayes' rule

$$f(p \,|\, x, n) \quad \propto \quad p^x \, (1-p)^{n-x} \, f_\circ(p)$$

Very rare processes

Sensitivity bounds: some hints for **self study**

Let us restart from the Bayes' rule

$$f(p\,|\,x,n) \quad \propto \quad p^x\,(1-p)^{n-x}\,f_\circ(p)$$

If you believe that p has to be very small, because you are dealing with a rather rare decay, just model $f_\circ(p)$ with something reasonable and do the math.

Very rare processes

Sensitivity bounds: some hints for **self study**

Let us restart from the Bayes' rule

$$f(p \,|\, x, n) \;\propto\; p^x\,(1-p)^{n-x}\,f_\circ(p)$$

If you believe that p has to be very small, because you are dealing with a rather rare decay, just model $f_\circ(p)$ with something reasonable and do the math.

For example, you might thing that $p \sim \mathcal{O}\left(10^{-6}\right)$.

Very rare processes

Sensitivity bounds: some hints for **self study**

Let us restart from the Bayes' rule

$$f(p\,|\,x,n) \quad \propto \quad p^x\,(1-p)^{n-x}\,f_\circ(p)$$

If you believe that p has to be very small, because you are dealing with a rather rare decay, just model $f_\circ(p)$ with something reasonable and do the math.

For example, you might thing that $p \sim \mathcal{O}\left(10^{-6}\right)$.
Then, e.g., $f_\circ(p) = 10^6\,\exp\left[-10^6\,p\right]$
with $\mathrm{E}(p) = 10^{-6}$ and $\sigma(p) = 10^{-6}$.

Very rare processes

Sensitivity bounds: some hints for **self study**

Let us restart from the Bayes' rule

$$f(p\,|\,x,n) \quad \propto \quad p^x\,(1-p)^{n-x}\,f_\circ(p)$$

If you believe that p has to be very small, because you are dealing with a rather rare decay, just model $f_\circ(p)$ with something reasonable and do the math.

For example, you might thing that $p \sim \mathcal{O}\left(10^{-6}\right)$.
Then, e.g., $f_\circ(p) = 10^6\,\exp\left[-10^6\,p\right]$
with $\mathrm{E}(p) = 10^{-6}$ and $\sigma(p) = 10^{-6}$.

- Do the math and calculate the posterior.

Very rare processes

Sensitivity bounds: some hints for **self study**

Let us restart from the Bayes' rule

$$f(p \,|\, x, n) \;\; \propto \;\; p^x \, (1-p)^{n-x} \, f_\circ(p)$$

If you believe that p has to be very small, because you are dealing with a rather rare decay, just model $f_\circ(p)$ with something reasonable and do the math.

For example, you might thing that $p \sim \mathcal{O}\left(10^{-6}\right)$.
Then, e.g., $f_\circ(p) = 10^6 \, \exp\left[-10^6\, p\right]$
with $\mathrm{E}(p) = 10^{-6}$ and $\sigma(p) = 10^{-6}$.

- Do the math and calculate the posterior.
- Anticipation of the result
  - if the prior is not updated at all, or if it is not changed significantly, than the experimental information is irrelevant.

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented (like what we have called the "Gaussian trick").

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented (like what we have called the "Gaussian trick").
Here is a very elegant one, particularly suitable useful to infer Bernoulli's p.

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented (like what we have called the "Gaussian trick").

Here is a very elegant one, particularly suitable useful to infer Bernoulli's p.

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a (1-p)^b$$

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented (like what we have called the "Gaussian trick").

Here is a very elegant one, particularly suitable useful to infer Bernoulli's p.

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a\,(1-p)^b$$

- Then the inference becomes

$$f(p\,|\,x,n) \quad \propto \quad p^x\,(1-p)^{n-x}\cdot p^a\,(1-p)^b$$

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented (like what we have called the "Gaussian trick").
Here is a very elegant one, particularly suitable useful to infer Bernoulli's p.

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a\,(1-p)^b$$

- Then the inference becomes

$$\begin{aligned} f(p\,|\,x,n) &\propto p^x\,(1-p)^{n-x} \cdot p^a\,(1-p)^b \\ &\propto p^{a+x}\,(1-p)^{b+(n-x)} \end{aligned}$$

Inferring the “Bernoulli’s p”

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace’ ideas (“Bayesian”) has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented (like what we have called the “Gaussian trick”).
Here is a very elegant one, particularly suitable useful to infer Bernoulli’s p.

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a\,(1-p)^b$$

- Then the inference becomes

$$\begin{aligned} f(p\,|\,x,n) &\propto p^x\,(1-p)^{n-x}\cdot p^a\,(1-p)^b \\ &\propto p^{a+x}\,(1-p)^{b+(n-x)} \\ &\propto p^{a'}\,(1-p)^{b'} \end{aligned}$$

Beta distribution

Indeed, such a pdf exists ($a = r - 1$; $b = s - 1$).

Beta distribution

Indeed, such a pdf exists ($a = r - 1$; $b = s - 1$).
In general, given the generic uncertain number X,

$$f(x \,|\, \text{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \le x \le 1 \end{cases}$$

Beta distribution

Indeed, such a pdf exists ($a = r - 1$; $b = s - 1$).
In general, given the generic uncertain number X,

$$f(x \,|\, \text{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \leq x \leq 1 \end{cases}$$

- The denominator is just for normalization, i.e.

$$\beta(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\, \mathrm{d}x$$

Beta distribution

Indeed, such a pdf exists ($a = r - 1$; $b = s - 1$).
In general, given the generic uncertain number X,

$$f(x\,|\,\mathrm{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \le x \le 1 \end{cases}$$

- The denominator is just for normalization, i.e.

$$\beta(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\,\mathrm{d}x$$

Indeed this integral defines the beta function, resulting in

$$\beta(r,s) = \frac{\Gamma(r)\,\Gamma(s)}{\Gamma(r+s)}$$

Beta distribution

Indeed, such a pdf exists ($a = r - 1$; $b = s - 1$).
In general, given the generic uncertain number X,

$$f(x\,|\,\mathrm{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \le x \le 1 \end{cases}$$

- The denominator is just for normalization, i.e.

$$\beta(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\,\mathrm{d}x$$

Indeed this integral defines the beta function, resulting in

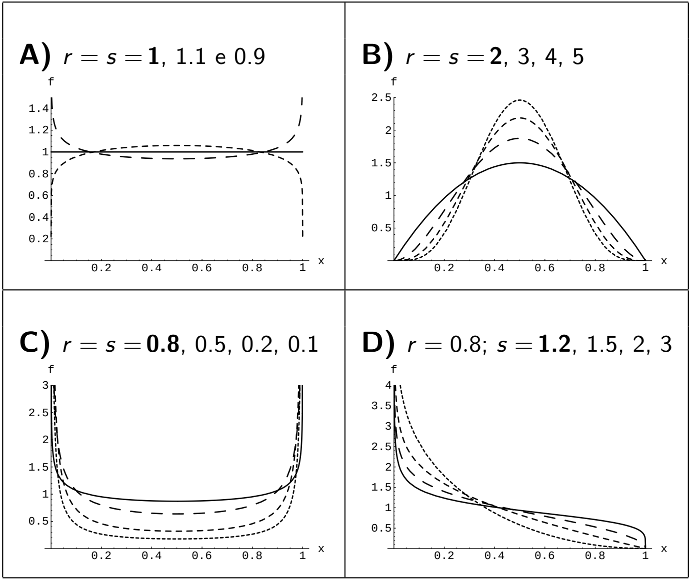
$$\beta(r,s) = \frac{\Gamma(r)\,\Gamma(s)}{\Gamma(r+s)}$$

Try e.g.

```
> p<-seq(0,1,by=0.01)
> plot(p, dbeta(p, 3, 5), ty='l', col='blue')
```

Beta distribution

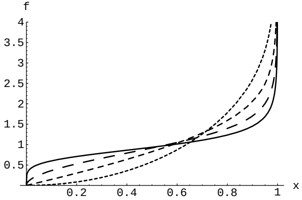
Some examples



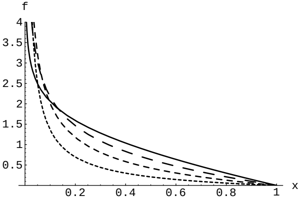
Beta distribution

Some examples

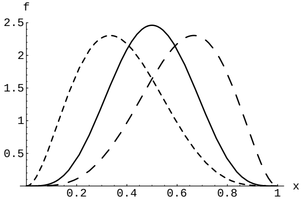
E) $s = 0.8$; $r = \mathbf{1.2}$, 1.5, 2, 3



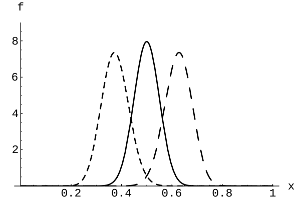
F) $s = 2$; $r = \mathbf{0.8}$, 0.6, 0.4, 0.2



G) $(r, s) = (3, 5)$, $\mathbf{(5, 5)}$, $(5, 3)$



H) $(r, s) = (30, 50)$, $\mathbf{(50, 50)}$, $(50, 30)$



Beta distribution
Summaries

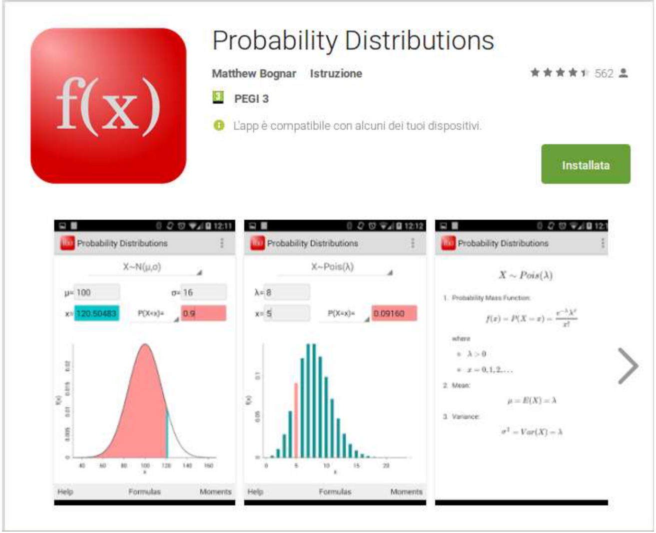
$$
\begin{aligned}
\mathrm{E}(X) &= \frac{r}{r+s} \\
\mathrm{Var}(X) &= \frac{rs}{(r+s+1)\,(r+s)^2}\,.
\end{aligned}
$$

Mode, unique if $r > 1$ and $s > 1$:

$$\frac{r-1}{r+s-2}$$

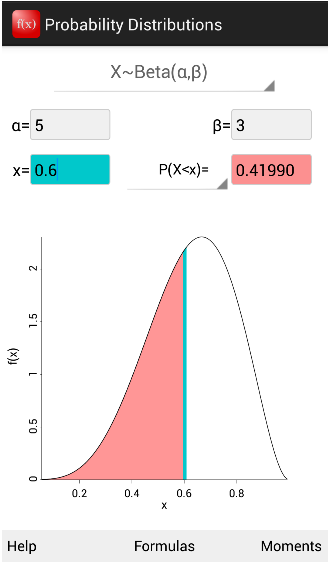
A useful app

https://play.google.com/store/apps/details?id=com.mbognar.probdist



A useful app

An example



Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$
\begin{aligned}
f(p \mid n, x, \text{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\
&\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1} \,.
\end{aligned}
$$

Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$\begin{aligned} f(p \mid n, x, \text{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\ &\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1} \,. \end{aligned}$$

Simple updating rule:

$$\begin{aligned} r_f &= r_i + x \\ s_f &= s_i + (n-x) \end{aligned}$$

Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$
\begin{aligned}
f(p \,|\, n, x, \mathrm{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\
&\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1}\,.
\end{aligned}
$$

Simple updating rule:

$$
\begin{aligned}
r_f &= r_i + x \\
s_f &= s_i + (n-x)
\end{aligned}
$$

Check the case of **uniform prior** $(r_i = s_i = 1)$

$$
\mathrm{E}(X) = \frac{r_f}{r_f + s_f} = \frac{x+1}{n+2}
$$

Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$\begin{aligned} f(p \,|\, n, x, \text{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\ &\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1} \,. \end{aligned}$$

Simple updating rule:

$$\begin{aligned} r_f &= r_i + x \\ s_f &= s_i + (n-x) \end{aligned}$$

Check the case of **uniform prior** ($r_i = s_i = 1$)

$$\begin{aligned} \text{E}(X) &= \frac{r_f}{r_f + s_f} = \frac{x+1}{n+2} \\ \text{Var}(X) &= \frac{r_f\, s_f}{(r_f + s_f + 1)\,(r_f + s_f)^2} = \frac{(x+1)(n-x+1)}{(n+3)\,(n+2)^2} \end{aligned}$$

Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$
\begin{aligned}
f(p \,|\, n, x, \mathrm{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\
&\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1}\,.
\end{aligned}
$$

Simple updating rule:

$$
\begin{aligned}
r_f &= r_i + x \\
s_f &= s_i + (n-x)
\end{aligned}
$$

Check the case of **uniform prior** ($r_i = s_i = 1$)

$$
\begin{aligned}
\mathrm{E}(X) &= \frac{r_f}{r_f + s_f} = \frac{x+1}{n+2} \\
\mathrm{Var}(X) &= \frac{r_f\, s_f}{(r_f + s_f + 1)\,(r_f + s_f)^2} = \frac{(x+1)(n-x+1)}{(n+3)\,(n+2)^2} \\
\mathrm{mode}(X) &= \frac{r_f - 1}{r_f + s_f - 2} = \frac{x}{n}
\end{aligned}
$$

Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$\begin{aligned} f(p \mid n, x, \text{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\ &\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1}\,. \end{aligned}$$

Simple updating rule:

$$\begin{aligned} r_f &= r_i + x \\ s_f &= s_i + (n-x) \end{aligned}$$

Check the case of **uniform prior** $(r_i = s_i = 1)$

$$\begin{aligned} \text{E}(X) &= \frac{r_f}{r_f + s_f} = \frac{x+1}{n+2} \\ \text{Var}(X) &= \frac{r_f\, s_f}{(r_f + s_f + 1)\,(r_f + s_f)^2} = \frac{(x+1)(n-x+1)}{(n+3)\,(n+2)^2} \\ \text{mode}(X) &= \frac{r_f - 1}{r_f + s_f - 2} = \frac{x}{n} \qquad \surd \end{aligned}$$

Conjugate priors

The Beta distribution is an example of conjugate prior:

Conjugate priors

The Beta distribution is an example of conjugate prior:

- a pdf such that prior and posterior belong to the same family;
- its parameters are updated by the the 'likelihood'.

Conjugate priors

The Beta distribution is an example of conjugate prior:

- a pdf such that prior and posterior belong to the same family;
- its parameters are updated by the the 'likelihood'.

Note:

- not all *conjugate priors* are as flexible as the Beta.

Conjugate priors

The Beta distribution is an example of conjugate prior:

- a pdf such that prior and posterior belong to the same family;
- its parameters are updated by the the 'likelihood'.

Note:

- not all *conjugate priors* are as flexible as the Beta.

  (In particular, the Gaussian is self-conjugate,
  which is not so great...)

More on priors

Data dominated inference

Let's look again at how the prior gets updated

$$\begin{aligned} f(p \mid n, x, r_i, s_i) &\propto \left[p^x(1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\ &\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1} \end{aligned}$$

More on priors

Data dominated inference

Let's look again at how the prior gets updated

$$
\begin{aligned}
f(p \mid n, x, r_i, s_i) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\
&\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1} \\
r_f &= r_i + x \\
s_f &= s_i + (n-x)
\end{aligned}
$$

More on priors

Data dominated inference

Let's look again at how the prior gets updated

$$
\begin{aligned}
f(p \mid n, x, r_i, s_i) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\
&\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1} \\
r_f &= r_i + x \\
s_f &= s_i + (n-x) \\
\mathrm{E}(p) &= \frac{r_f}{r_f+s_f} = \frac{r_i+x}{r_i+s_i+n} \\
\mathrm{Var}(p) &= \frac{r_f\, s_f}{(r_f+s_f+1)\,(r_f+s_f)^2} \\
&= \frac{(r_i+x)\cdot(s_i+n-x)}{(r_i+s_i+n+1)\,(r_i+s_i+n)^2}
\end{aligned}
$$

More on priors

Data dominated inference

Let's look again at how the prior gets updated

$$
\begin{aligned}
f(p \mid n, x, r_i, s_i) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\
&\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1} \\
r_f &= r_i + x \\
s_f &= s_i + (n-x) \\
\mathrm{E}(p) &= \frac{r_f}{r_f+s_f} = \frac{r_i+x}{r_i+s_i+n} \\
\mathrm{Var}(p) &= \frac{r_f\, s_f}{(r_f+s_f+1)\,(r_f+s_f)^2} \\
&= \frac{(r_i+x)\cdot(s_i+n-x)}{(r_i+s_i+n+1)\,(r_i+s_i+n)^2}
\end{aligned}
$$

If $x \gg r_i$ and $(n-x) \gg s_i$

$$
\begin{aligned}
r_f &\approx x \\
s_f &\approx (n-x)
\end{aligned}
$$

Predictive distribution

Predicting future nr. of successes and future frequences

- Imagine we have have got 5 successes in 10 trials.

Predictive distribution

Predicting future nr. of successes and future frequences

- Imagine we have have got 5 successes in 10 trials.
- Imagine that we want to make another 10 trials:
  what is the probability to get 0, 1, . . . , 10 successes?

Predictive distribution

Predicting future nr. of successes and future frequences

- Imagine we have have got 5 successes in 10 trials.
- Imagine that we want to make another 10 trials:
  what is the probability to get 0, 1, . . . , 10 successes?
- From the past data (and assuming a flat prior), we 'know' that $p \approx 0.5$.

Predictive distribution

Predicting future nr. of successes and future frequences

- Imagine we have have got 5 successes in 10 trials.
- Imagine that we want to make another 10 trials:
  what is the probability to get 0, 1, . . . , 10 successes?
- From the past data (and assuming a flat prior), we 'know' that $p \approx 0.5$.
- If we were sure that p was $1/2$, then we could simply use $\mathcal{B}_{10,\,1/2}$.

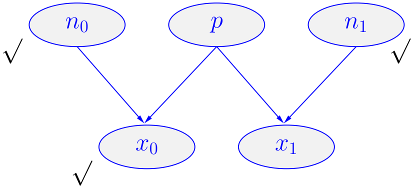
Predictive distribution

Predicting future nr. of successes and future frequences

- Imagine we have have got 5 successes in 10 trials.
- Imagine that we want to make another 10 trials:
  what is the probability to get 0, 1, ..., 10 successes?
- From the past data (and assuming a flat prior), we 'know' that $p \approx 0.5$.
- If we were sure that p was $1/2$, then we could simply use $\mathcal{B}_{10,\,1/2}$.
- But we are not sure about it: we need to take into account all possible values, each weighted by $f(p)$

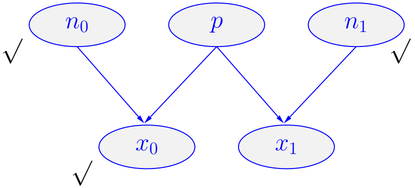
Predictive distribution

Predicting future nr. of successes and future frequences



Predictive distribution

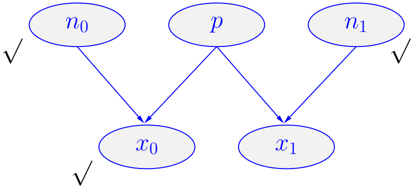
Predicting future nr. of successes and future frequences



- We need to take into account all possible values of p, each weighted by how much we believe it, i.e. by $f(p)$

Predictive distribution

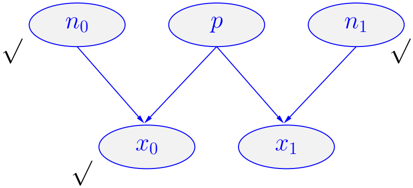
Predicting future nr. of successes and future frequences



- We need to take into account all possible values of p, each weighted by how much we believe it, i.e. by $f(p)$
- $f(x) = \int_0^1 f(x \mid p)\, f(p)\, dp$.

Predictive distribution

Predicting future nr. of successes and future frequences



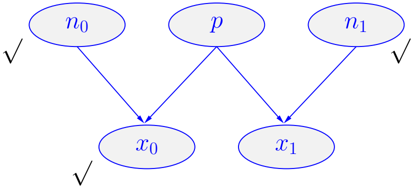
- We need to take into account all possible values of p, each weighted by how much we believe it, i.e. by $f(p)$
- $f(x) = \int_0^1 f(x \,|\, p)\, f(p)\, dp$.
- More precisely,

$$f(x_1 \,|\, n_1, n_0, x_0) = \int_0^1 f(x_1 \,|\, n_1, p)\, f(p \,|\, x_0, n_0)\, dp$$

- $X_1 \to f_1$

Predictive distribution

Predicting future nr. of successes and future frequences



- We need to take into account all possible values of p, each weighted by how much we believe it, i.e. by $f(p)$
- $f(x) = \int_0^1 f(x \mid p)\, f(p)\, dp$.

- More precisely,

$$f(x_1 \mid n_1, n_0, x_0) = \int_0^1 f(x_1 \mid n_1, p)\, f(p \mid x_0, n_0)\, dp$$

- $X_1 \rightarrow f_1$ (Predicting a **future frequency from a past frequency**)

Applications to Covid-19 tests and vaccines

(Left to self-reading)

- GdA, A. Esposito, Checking individuals and sampling populations with imperfect tests,
  `arXiv:2009.04843 [q-bio.PE]`

Applications to Covid-19 tests and vaccines

(Left to self-reading)

- GdA, A. Esposito, Checking individuals and sampling populations with imperfect tests,
  `arXiv:2009.04843 [q-bio.PE]`

- GdA, A. Esposito, What is the probability that a vaccinated person is shielded from Covid-19? A Bayesian MCMC based reanalysis of published data with emphasis on what should be reported as 'efficacy',
  `arXiv:2102.11022 [stat.AP]`
  (Only limited to data published in 2020)

Applications to Covid-19 tests and vaccines

(Left to self-reading)

- GdA, A. Esposito, Checking individuals and sampling populations with imperfect tests,
  `arXiv:2009.04843 [q-bio.PE]`

- GdA, A. Esposito, What is the probability that a vaccinated person is shielded from Covid-19? A Bayesian MCMC based reanalysis of published data with emphasis on what should be reported as 'efficacy',
  `arXiv:2102.11022 [stat.AP]`
  (Only limited to data published in 2020)

Both written (also) with teaching purposes,
also providing R and JAGS code.

Applications to Covid-19 tests and vaccines

(Left to self-reading)

- GdA, A. Esposito, Checking individuals and sampling populations with imperfect tests,
  arXiv:2009.04843 [q-bio.PE]

- GdA, A. Esposito, What is the probability that a vaccinated person is shielded from Covid-19? A Bayesian MCMC based reanalysis of published data with emphasis on what should be reported as 'efficacy',
  arXiv:2102.11022 [stat.AP]
  (Only limited to data published in 2020)

Both written (also) with teaching purposes,
also providing R and JAGS code.
(The R code can be used as a kind of *pseudocode*
by those who prefer to call JAGS from Python.)

The End