Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

‘**Probabilistic**’ stressed

From the previous lectures it should be clear that I stick to

- **Probability theory**:

'Probabilistic' stressed

From the previous lectures it should be clear that I stick to

- **Probability theory**:
  - basic rules;
  - logic ('math');

‘Probabilistic’ stressed

From the previous lectures it should be clear that I stick to

- **Probability theory**:
  - basic rules;
  - logic ('math');
  - ‘tricks’ to perform calculations
    (e.g. integrals by Monte Carlo);

‘Probabilistic’ stressed

From the previous lectures it should be clear that I stick to

- **Probability theory**:
  - basic rules;
  - logic ('math');
  - ‘tricks’ to perform calculations
    (e.g. integrals by Monte Carlo);
  - approximations under suitable conditions.

'Probabilistic' stressed

From the previous lectures it should be clear that I stick to

- **Probability theory**:
  - basic rules;
  - logic ('math');
  - 'tricks' to perform calculations
    (e.g. integrals by Monte Carlo);
  - approximations under suitable conditions.

At every step 'probability' means...

‘Probabilistic’ stressed

From the previous lectures it should be clear that I stick to

- **Probability theory**:
  - basic rules;
  - logic ('math');
  - ‘tricks’ to perform calculations (e.g. integrals by Monte Carlo);
  - approximations under suitable conditions.

At every step ‘probability’ means... **probability**.

Although other ‘methods’ might be more popular

Although other ‘methods’ might be more popular

Although other ‘methods’ might be more popular

Although other ‘methods’ might be more popular

Although other 'methods' might be more popular

Although other 'methods' might be more popular

[Plus other **prescriptions** you might imagine. . .]

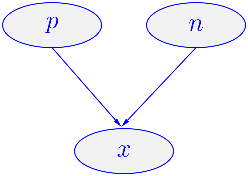
n independent Bernoulli processes

General case

n independent Bernoulli processes

General case

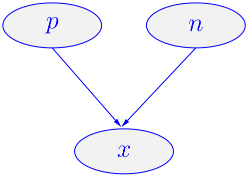
Model



n independent Bernoulli processes

General case

Model



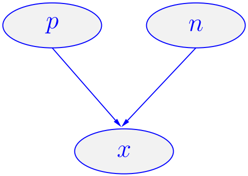
Joint pdf (omitting background condition I):

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

n independent Bernoulli processes

General case

Model



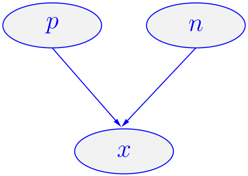
Joint pdf (omitting background condition I):

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

n independent Bernoulli processes

General case

Model



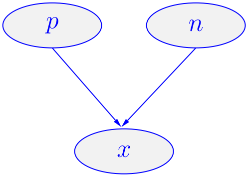
Joint pdf (omitting background condition I):

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

n independent Bernoulli processes

General case

Model



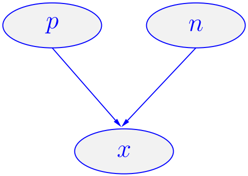
Joint pdf (omitting background condition I):

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

n independent Bernoulli processes

General case

Model



Joint pdf (omitting background condition I):

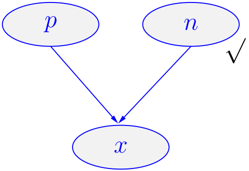
$$
\begin{aligned}
f(x,p,n) &= f(x\,|\,p,n)\cdot f(p,n) \\
&= f(x\,|\,p,n)\cdot f(p\,|\,n)\cdot f(n) \\
&= f(x\,|\,p,n)\cdot f(n\,|\,p)\cdot f(p) \\
&= f(x\,|\,p,n)\cdot f(p)\cdot f(n)
\end{aligned}
$$

(n and p are independent)

n independent Bernoulli processes

Usual case → n fixed (for the moment)

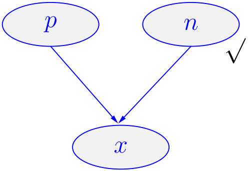
Model



n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



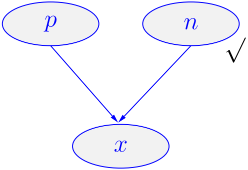
Joint pdf

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

n independent Bernoulli processes

Usual case $\rightarrow$ n fixed (for the moment)

Model



Joint pdf

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

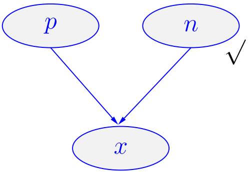
Typical problems

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

n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



Joint pdf

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

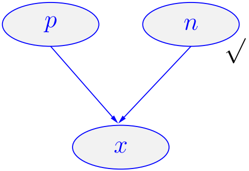
Typical problems

- 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

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

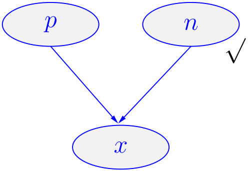
Typical problems

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

n independent Bernoulli processes

Usual case → n fixed (for the moment)

Model



Joint pdf

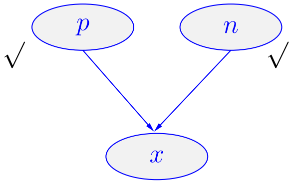
$$f(x, p \,|\, n) \;=\; f(x \,|\, p, n) \cdot f(p)$$

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

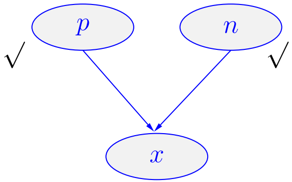
Graphical models of the typical problems



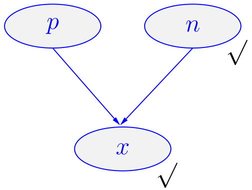
$$\rightarrow\ f(x \,|\, n, p)$$

n independent Bernoulli processes

Graphical models of the typical problems



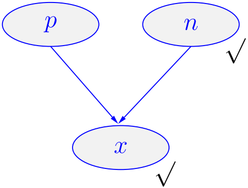
$\rightarrow\ f(x \mid n, p)$



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

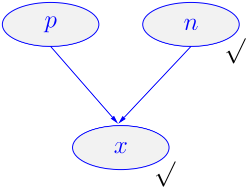
n independent Bernoulli processes

Inferring p



n independent Bernoulli processes

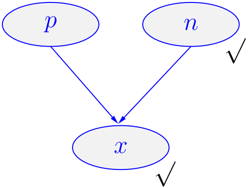
Inferring p



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

n independent Bernoulli processes

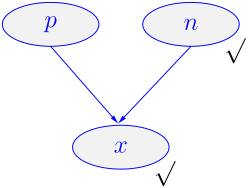
Inferring p



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

n independent Bernoulli processes

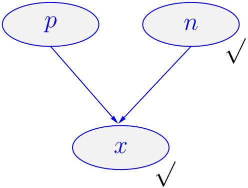
Inferring p



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

n independent Bernoulli processes

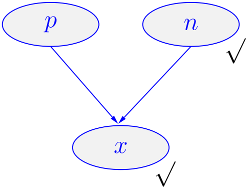
Inferring p



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

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

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

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:

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

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:

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

(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"

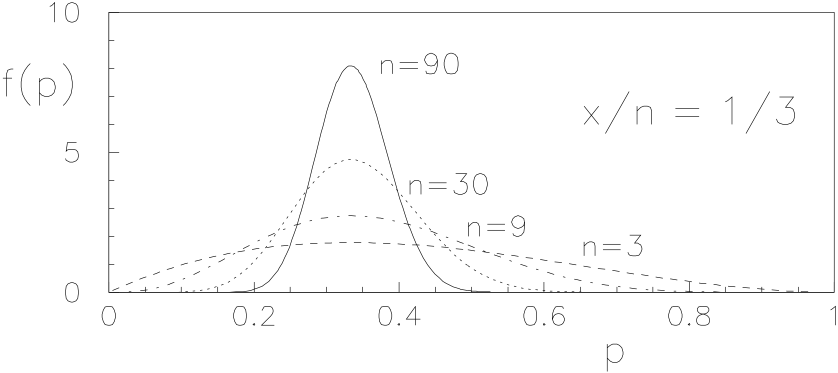
Solution for uniform prior (think to **Bayes' billard**)

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

Inferring "Bernoulli's p"

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

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



Inferring "Bernoulli's p"

Summaries of the **posterior distribution**

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

Inferring "Bernoulli's p"

Summaries of the **posterior distribution**

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

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

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**

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

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

"recursive Laplace formula"

("Laplace's rule of succession")

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

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:

$$P(E_i \mid p) = p$$

- 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 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.

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

Inferring the “Bernoulli's p”

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

- 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.

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

Inferring the "Bernoulli's p"

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

- 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:

$$P(E_i \mid p) = p$$

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

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

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 $p \leftrightarrow q$):

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

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 $p \leftrightarrow q$):

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

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 $p \leftrightarrow q$):

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

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:

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

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 $\sigma_p \rightarrow 0$,

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;
- and $(n-x)$ 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

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

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

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

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

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

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,
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;
- <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**

— **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).
- Proability 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).

- Proability 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).

- Proability 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).

- Proability 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 $n_s \to n$)

- $N - n = 124$
  $\to$ 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 $n_s \rightarrow n$)

- $N - n = 124$

  $\rightarrow$ 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:

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

- 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:

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

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

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:

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

- 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 $\surd$
  $\rightarrow \sigma_\epsilon = 0.0049 \approx 0.005$

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?

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

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:

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

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:

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

- 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:

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

- 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 $x = \mathbf{0}$

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

Inferring "Bernoulli's p"

Observing $x = \mathbf{0}$

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

Inferring "Bernoulli's p"

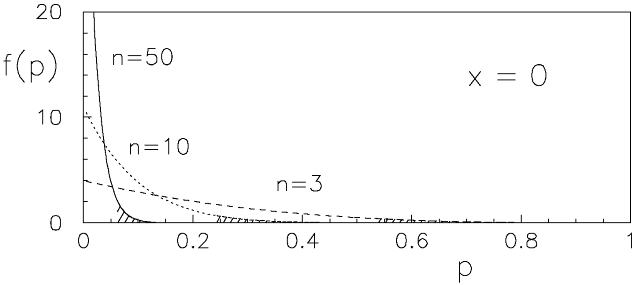
Observing $x = \mathbf{0}$

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

Inferring "Bernoulli's p"

Observing $x = \mathbf{0}$

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



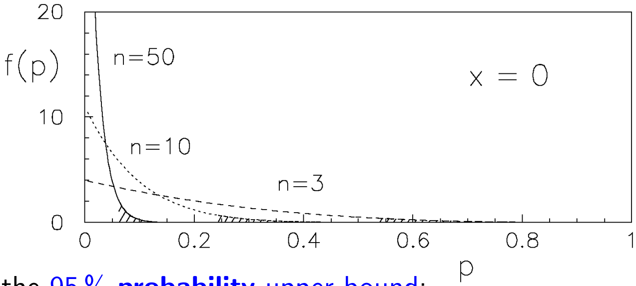
Inferring "Bernoulli's p"

Observing $x = \mathbf{0}$

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

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

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



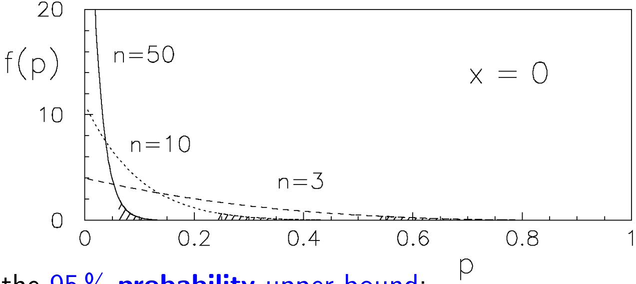
To get the 95 % **probability** upper bound:

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

Inferring “Bernoulli’s p”

Observing $x = \mathbf{0}$

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



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

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

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

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

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

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

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 $p \sim \mathcal{O}\left(10^{-6}\right)$.

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

Inferring the "Bernoulli's p"

Mathematically convenient priors

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

→ **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,

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

Beta distribution

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

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

- 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,

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

- 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,

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

- 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

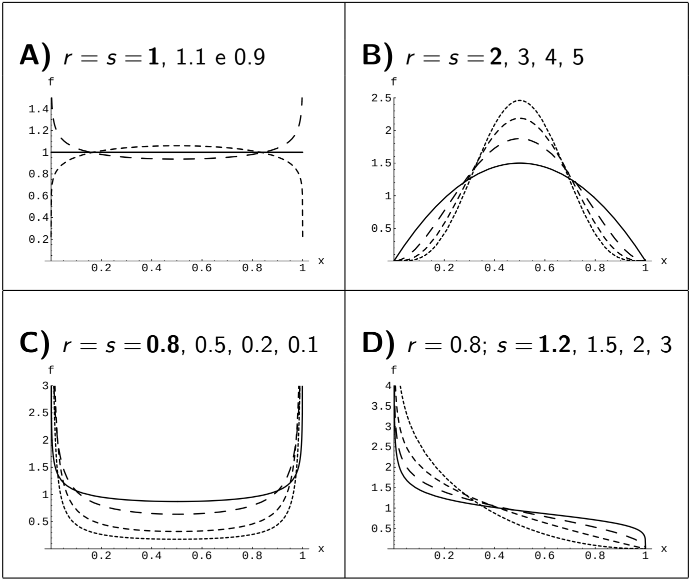
$$\beta(r,s) = \frac{\Gamma(r)\,\Gamma(s)}{\Gamma(r+s)}$$

Try e.g.

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

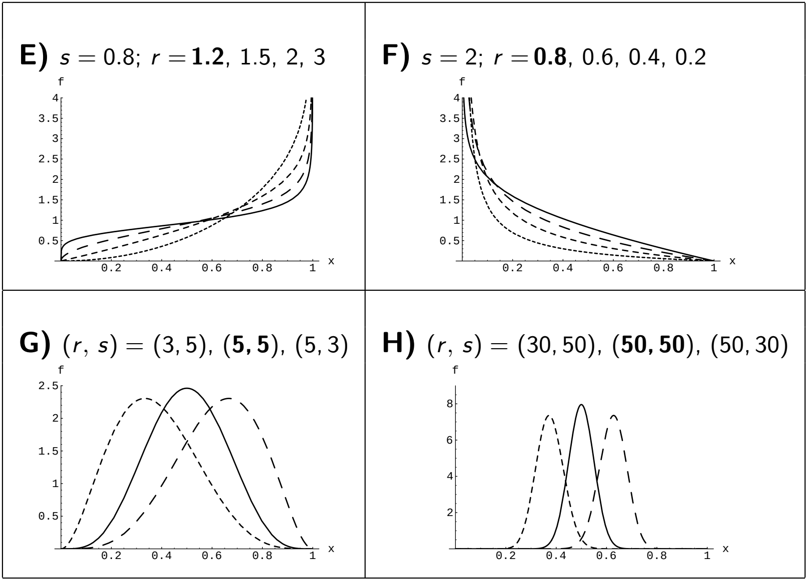
Beta distribution

Some examples



Beta distribution

Some examples



Beta distribution

Summaries

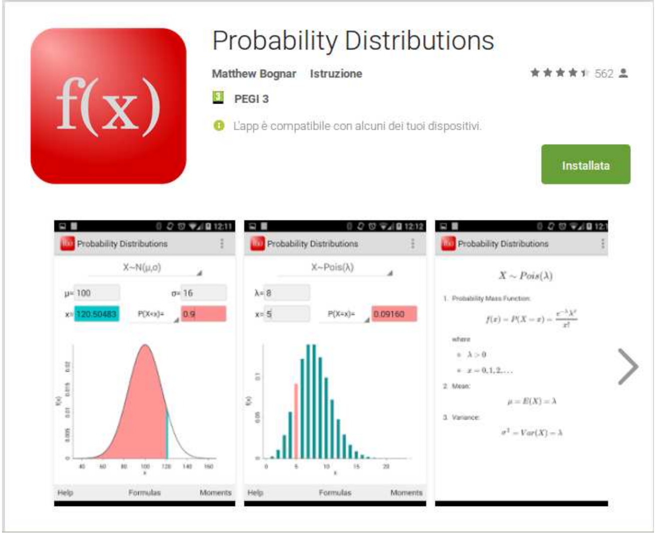
$$\begin{aligned} \mathrm{E}(X) &= \frac{r}{r+s} \\ \mathrm{Var}(X) &= \frac{rs}{(r+s+1)\,(r+s)^2}\,. \end{aligned}$$

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

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

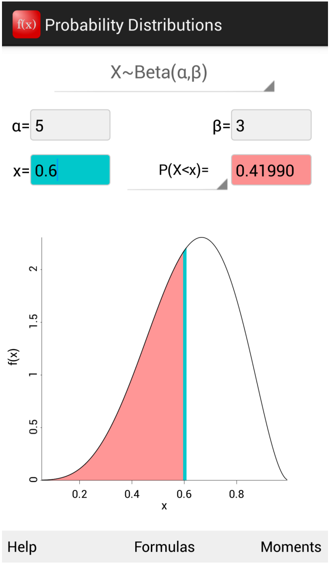
A useful app

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

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

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

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

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

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

Simple updating rule:

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

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

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

Beta distribution as prior

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

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

Simple updating rule:

$$r_f = r_i + x$$
$$s_f = s_i + (n-x)$$

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

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

$$\mathrm{Var}(X) = \frac{r_f\, s_f}{(r_f + s_f + 1)\,(r_f + s_f)^2} = \frac{(x+1)(n-x+1)}{(n+3)\,(n+2)^2}$$

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

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

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

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 $\mathcal{B}_{10,\,1/2}$.

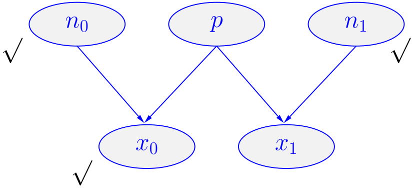
Predictive distribution

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 $f(p)$

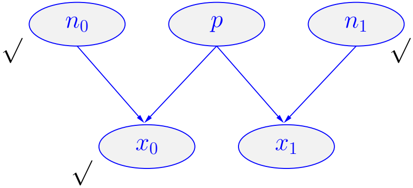
Predictive distribution

Predicting future nr. of successes and future frequences



Predictive distribution

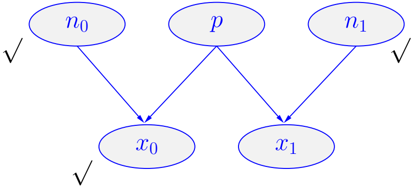
Predicting future nr. of successes and future frequences



- 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

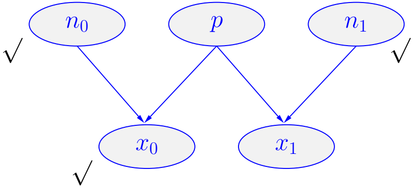
Predicting future nr. of successes and future frequences



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

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 $f(p)$
- $f(x) = \int_0^1 f(x \mid p)\, f(p)\, dp$.

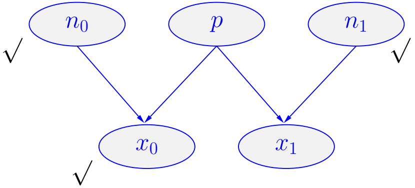
- More precisely,

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

- $X_1 \to f_1$

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 $f(p)$
- $f(x) = \int_0^1 f(x \,|\, p)\, f(p)\, dp$.

- More precisely,

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

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

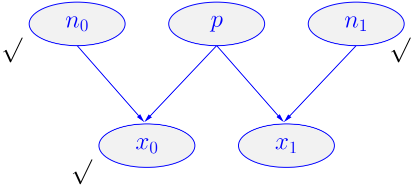
Predictive distribution

Some examples

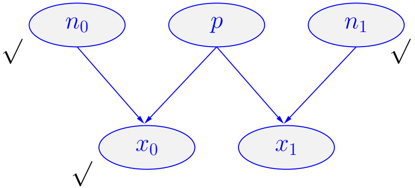
$f(x_1 \mid n_0, x_0, n_1 = 10)$ in %

X_1	$\frac{X_1}{n_1}$	$\begin{cases} x_0 = 1 \\ n_0 = 2 \end{cases}$	$\begin{cases} x_0 = 10 \\ n_0 = 20 \end{cases}$	$\begin{cases} x_0 = 100 \\ n_0 = 200 \end{cases}$	$\begin{cases} x_0 = 1000 \\ n_0 = 2000 \end{cases}$
0	0	3.85	0.42	0.12	0.10
1	0.1	6.99	2.29	1.11	0.99
2	0.2	9.44	6.51	4.67	4.42
3	0.3	11.19	12.54	11.88	11.74
4	0.4	12.24	18.07	20.21	20.48
5	0.5	12.59	20.33	24.02	24.55
6	0.6	12.24	18.07	20.21	20.48
7	0.7	11.19	12.54	11.88	11.74
8	0.8	9.44	6.51	4.67	4.42
9	0.9	6.99	2.29	1.11	0.99
10	1	3.84	0.42	0.12	0.10
$\mathrm{E}(X_1)$		5	5	5	5
$\sigma[X_1]$		2.64	1.87	1.62	1.58

Joint inference and prediction



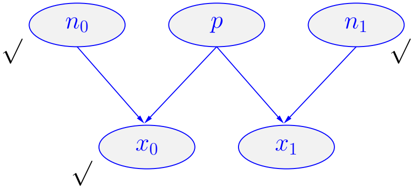
Joint inference and prediction



In reality the general solution starts from

$$f(n_0, p, n_1, x_0, x_1)$$

Joint inference and prediction



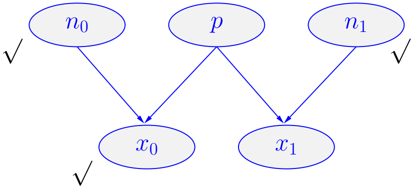
In reality the general solution starts from

$$f(n_0, p, n_1, x_0, x_1)$$

conditioning on what it is 'known' (or 'assumed'):

$$f(p, x_1 \mid n_0, x_0, n_1) = \frac{f(n_0, p, n_1, x_0, x_1)}{f(n_0, x_0, n_1)}$$

Joint inference and prediction



In reality the general solution starts from

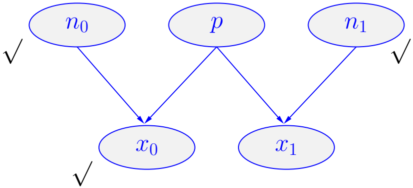
$$f(n_0, p, n_1, x_0, x_1)$$

conditioning on what it is 'known' (or 'assumed'):

$$f(p, x_1 \mid n_0, x_0, n_1) = \frac{f(n_0, p, n_1, x_0, x_1)}{f(n_0, x_0, n_1)}$$

$\Rightarrow$ p **and** x_1 **are correlated!**

Joint inference and prediction



In reality the general solution starts from

$$f(n_0, p, n_1, x_0, x_1)$$

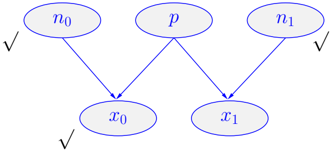
conditioning on what it is 'known' (or 'assumed'):

$$f(p, x_1 \mid n_0, x_0, n_1) = \frac{f(n_0, p, n_1, x_0, x_1)}{f(n_0, x_0, n_1)}$$

$\Rightarrow$ p **and** x_1 **are correlated!**

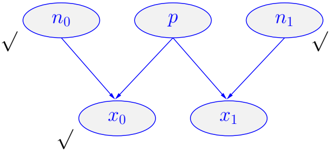
$$\rho(p, x_1) > 0$$

Joint inference and prediction



Let's do the math.

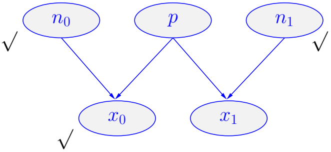
Joint inference and prediction



Let's do the math.

- Three **observed variables**

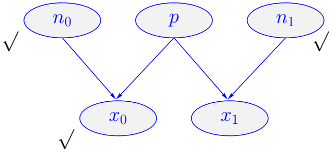
Joint inference and prediction



Let's do the math.

- Three **observed variables** (no uncertainty): n_0, x_0 and n_1.

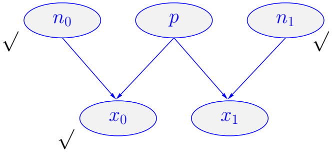
Joint inference and prediction



Let's do the math.

- Three **observed variables** (no uncertainty): n_0, x_0 and n_1.
- Two **unobserved variables** (uncertain value)

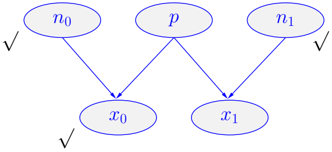
Joint inference and prediction



Let's do the math.

- Three **observed variables** (no uncertainty): n_0, x_0 and n_1.
- Two **unobserved variables** (uncertain value): p and x_1.

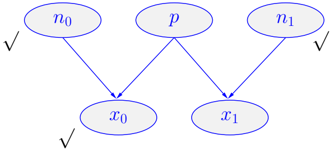
Joint inference and prediction



Let's do the math.

- Three **observed variables** (no uncertainty): n_0, x_0 and n_1.
- Two **unobserved variables** (uncertain value): p and x_1.
- $f(n_0, x_0, n_1)$ **is a number**, given the model.

Joint inference and prediction

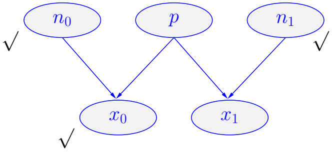


Let's do the math.

- Three **observed variables** (no uncertainty): n_0, x_0 and n_1.
- Two **unobserved variables** (uncertain value): p and x_1.
- $f(n_0, x_0, n_1)$ **is a number**, given the model.
  It might be difficult to calculate, but it is a number.

$$f(p, x_1 \,|\, n_0, x_0, n_1) \;=\; \frac{f(p, x1, n_0, n_1, x_0)}{f(n_0, x_0, n_1)}$$

Joint inference and prediction



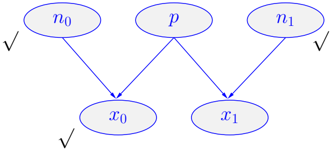
Let's do the math.

- Three **observed variables** (no uncertainty): n_0, x_0 and n_1.
- Two **unobserved variables** (uncertain value): p and x_1.
- $f(n_0, x_0, n_1)$ **is a number**, given the model.
  It might be difficult to calculate, but it is a number.

$$f(p, x_1 \,|\, n_0, x_0, n_1) \;=\; \frac{f(p, x1, n_0, n_1, x_0)}{f(n_0, x_0, n_1)}$$

$$\propto \; f(p, x1, n_0, n_1, x_0)$$

Joint inference and prediction



Let's do the math.

- Three **observed variables** (no uncertainty): n_0, x_0 and n_1.
- Two **unobserved variables** (uncertain value): p and x_1.
- $f(n_0, x_0, n_1)$ **is a number**, given the model.
  It might be difficult to calculate, but it is a number.

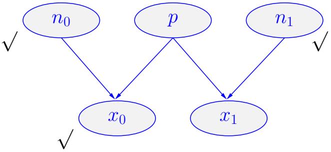
$$f(p, x_1 \,|\, n_0, x_0, n_1) = \frac{f(p, x1, n_0, n_1, x_0)}{f(n_0, x_0, n_1)}$$

$$\propto f(p, x1, n_0, n_1, x_0)$$

$$\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1) = f(p, x1, n_0, n_1, x_0)$$

$\tilde{f}()$: unnormalized pdf.

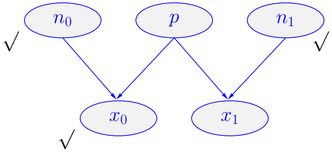
Joint inference and prediction



Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

$$f(p, x_1 \mid n_0, x_0, n_1) \;\propto\; f(x_0 \mid n_0, p) \cdot f(x_1 \mid p, n_1) \cdot f_0(p)$$

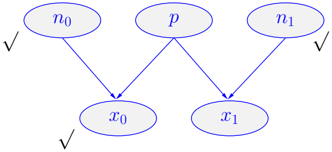
Joint inference and prediction



Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

$$
\begin{aligned}
f(p, x_1 \,|\, n_0, x_0, n_1) &\propto f(x_0 \,|\, n_0, p) \cdot f(x_1 \,|\, p, n_1) \cdot f_0(p) \\
&\propto p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{(n_1-x_1)!\, x_1!} \cdot f_0(p)
\end{aligned}
$$

Joint inference and prediction



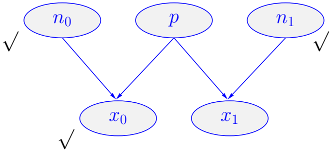
Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

$$f(p, x_1 \mid n_0, x_0, n_1) \propto f(x_0 \mid n_0, p) \cdot f(x_1 \mid p, n_1) \cdot f_0(p)$$

$$\propto p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{(n_1-x_1)!\, x_1!} \cdot f_0(p)$$

$$\tilde{f}(p, x_1 \mid n_0, x_0, n_1) = \frac{p^{x_0+x_1}(1-p)^{n_0+n_1-x_0-x_1}}{(n_1-x_1)!\, x_1!} \cdot f_0(p)$$

Joint inference and prediction



Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

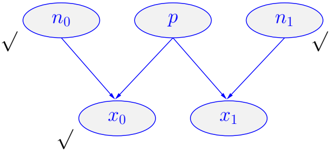
$$f(p, x_1 \mid n_0, x_0, n_1) \propto f(x_0 \mid n_0, p) \cdot f(x_1 \mid p, n_1) \cdot f_0(p)$$

$$\propto p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{(n_1-x_1)!\,x_1!} \cdot f_0(p)$$

$$\tilde{f}(p, x_1 \mid n_0, x_0, n_1) = \frac{p^{x_0+x_1}(1-p)^{n_0+n_1-x_0-x_1}}{(n_1-x_1)!\,x_1!} \cdot f_0(p)$$

Problem almost solved

Joint inference and prediction



Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

$$f(p, x_1 \,|\, n_0, x_0, n_1) \;\propto\; f(x_0 \,|\, n_0, p) \cdot f(x_1 \,|\, p, n_1) \cdot f_0(p)$$

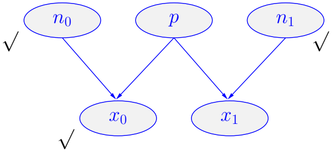
$$\propto\; p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{(n_1-x_1)!\,x_1!} \cdot f_0(p)$$

$$\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1) \;=\; \frac{p^{x_0+x_1}(1-p)^{n_0+n_1-x_0-x_1}}{(n_1-x_1)!\,x_1!} \cdot f_0(p)$$

Problem almost solved

- Possibly calculate the normalization, then all moments and probability intervals of interest.

Joint inference and prediction



Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

$$f(p, x_1 \mid n_0, x_0, n_1) \propto f(x_0 \mid n_0, p) \cdot f(x_1 \mid p, n_1) \cdot f_0(p)$$

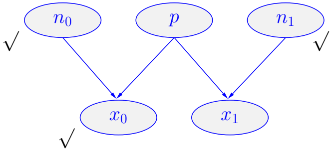
$$\propto p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{(n_1-x_1)!\,x_1!} \cdot f_0(p)$$

$$\tilde{f}(p, x_1 \mid n_0, x_0, n_1) = \frac{p^{x_0+x_1}(1-p)^{n_0+n_1-x_0-x_1}}{(n_1-x_1)!\,x_1!} \cdot f_0(p)$$

Problem almost solved

- Possibly calculate the normalization, then all moments and probability intervals of interest.
- Do it numerically,

Joint inference and prediction



Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

$$f(p, x_1 \,|\, n_0, x_0, n_1) \;\propto\; f(x_0 \,|\, n_0, p) \cdot f(x_1 \,|\, p, n_1) \cdot f_0(p)$$

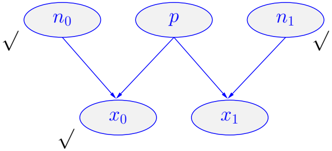
$$\propto\; p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{(n_1-x_1)!\, x_1!} \cdot f_0(p)$$

$$\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1) \;=\; \frac{p^{x_0+x_1}(1-p)^{n_0+n_1-x_0-x_1}}{(n_1-x_1)!\, x_1!} \cdot f_0(p)$$

Problem almost solved

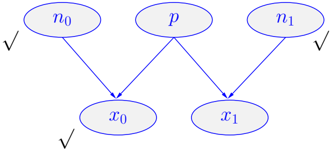
- Possibly calculate the normalization, then all moments and probability intervals of interest.
- Do it numerically,
- or by by sampling.

Joint inference and prediction



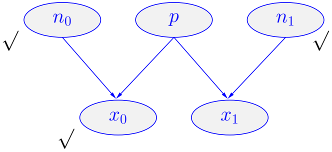
$\Rightarrow$ sample $\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1)$

Joint inference and prediction



$\Rightarrow$ sample $\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1)$ using Monte Carlo techniques

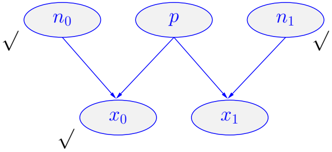
Joint inference and prediction



$\Rightarrow$ sample $\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1)$ using Monte Carlo techniques

$\Rightarrow$ **Markov Chain Monte Carlo (MCMC)**

Joint inference and prediction

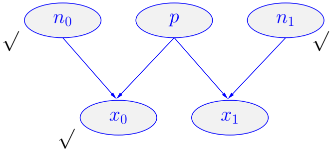


$\Rightarrow$ sample $\tilde{f}(p, x_1 \mid n_0, x_0, n_1)$ using Monte Carlo techniques

$\Rightarrow$ **Markov Chain Monte Carlo (MCMC)**

$\Rightarrow$ JAGS does it for us

Joint inference and prediction



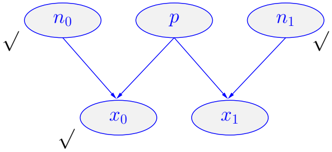
$\Rightarrow$ sample $\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1)$ using Monte Carlo techniques

$\Rightarrow$ **Markov Chain Monte Carlo (MCMC)**

$\Rightarrow$ JAGS does it for us

- by Gibbs sampler (JAGS: Just Another Gibbs Sampler) if pdf's involved allow it;

Joint inference and prediction



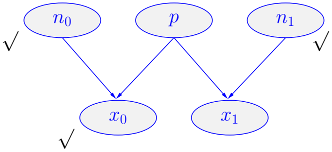
$\Rightarrow$ sample $\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1)$ using Monte Carlo techniques

$\Rightarrow$ **Markov Chain Monte Carlo (MCMC)**

$\Rightarrow$ JAGS does it for us

- by Gibbs sampler (JAGS: Just Another Gibbs Sampler) if pdf's involved allow it;
- by Metropolis

Joint inference and prediction



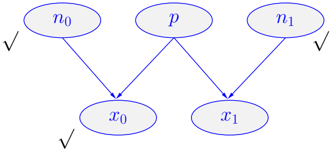
⇒ sample $\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1)$ using Monte Carlo techniques

⇒ Markov Chain Monte Carlo (MCMC)

⇒ JAGS does it for us

- by Gibbs sampler (JAGS: Just Another Gibbs Sampler) if pdf's involved allow it;
- by Metropolis ("when the going gets tough, the tough get going" – J. Belushi)

Joint inference and prediction



$\Rightarrow$ sample $\tilde{f}(p, x_1 \mid n_0, x_0, n_1)$ using Monte Carlo techniques

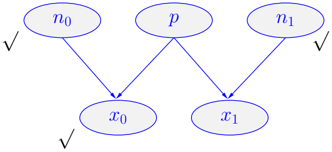
$\Rightarrow$ Markov Chain Monte Carlo (MCMC)

$\Rightarrow$ JAGS does it for us

- by Gibbs sampler (JAGS: Just Another Gibbs Sampler) if pdf's involved allow it;
- by Metropolis ("when the going gets tough, the tough get going" – J. Belushi)

JAGS called from R using the package rjags.

Joint inference and prediction



$\Rightarrow$ sample $\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1)$ using Monte Carlo techniques

$\Rightarrow$ Markov Chain Monte Carlo (MCMC)

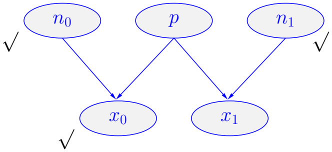
$\Rightarrow$ JAGS does it for us

- by Gibbs sampler (JAGS: Just Another Gibbs Sampler) if pdf's involved allow it;
- by Metropolis ("when the going gets tough, the tough get going" – J. Belushi)

JAGS called from R using the package rjags.

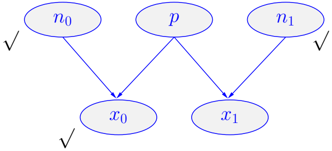
(No details on MCMC provided $\longrightarrow$ see references on the web site)

Graphical models: some terminology



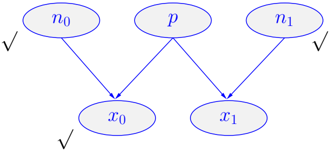
- nodes (observed/unobserved);
- child/childred;
- parent(s).

Graphical models: some terminology



- nodes (observed/unobserved);
- child/childred;
- parent(s).
- **A node without parents needs a prior**
  (node *p* in this case)

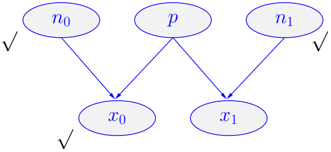
Joint inference and prediction in JAGS



Model

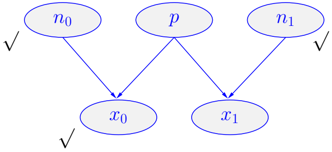
```
model{
    x0 ~ dbin(p, n0);
    x1 ~ dbin(p, n1);
     p ~ dbeta(1, 1);
}
```

Joint inference and prediction in JAGS



Then the model has to be in a file.

Joint inference and prediction in JAGS



Then the model has to be in a file.
For such a small model we can write it directly from R on a temporary file:

```
model = "tmp_model.bug"
write("
model{
    x0 ~ dbin(p, n0);
    x1 ~ dbin(p, n1);
     p ~ dbeta(1, 1);
}
", model)
```

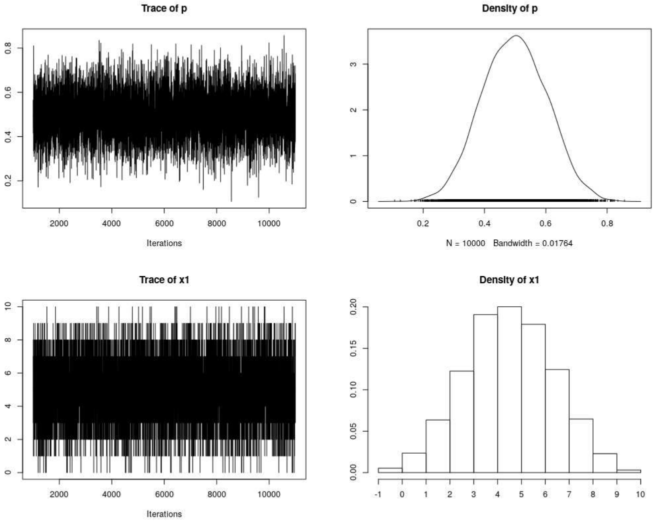
Use of JAGS from R via rjags

Second part of the R script (⇒ inf_p_pred.R)

```
library(rjags)

data = list(n0=20, x0=10, n1=10)

jm <- jags.model(model, data)
chain <- coda.samples(jm, c("p", "x1"), n.iter=10000)

plot(chain)
print(summary(chain))
```

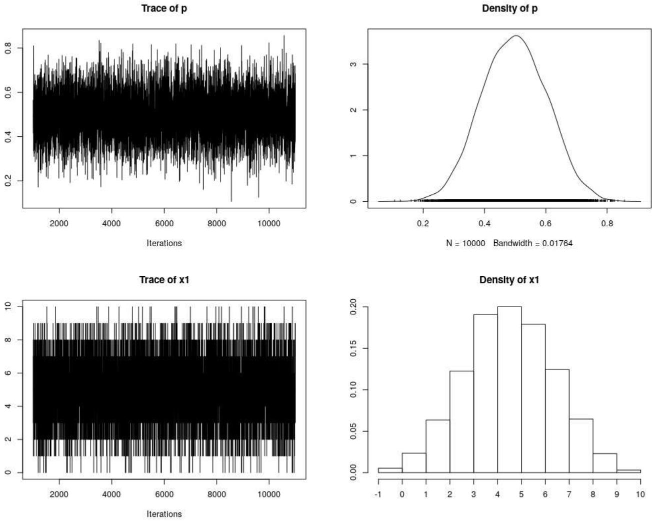
Use of JAGS from R via rjags

$(n0 = 20,\ x0 = 10,\ n1 = 10)$



Use of JAGS from R via rjags

$(n0 = 20,\ x0 = 10,\ n1 = 10)$



$p = 0.498 \pm 0.105$; $x_1 = 4.98 \pm 1.86$ (10000 samples).

Inference and prediction with JAGS/rjags

Comparison with exact result of $f(x_1 \mid n_0, x_0, n_1)$

$f(x_1 \mid n_0, x_0, n_1 = 10)$ in %

X_1	$\frac{X_1}{n_1}$	$\begin{cases} x_0 = 1 \\ n_0 = 2 \end{cases}$	$\begin{cases} x_0 = 10 \\ n_0 = 20 \end{cases}$	$\begin{cases} x_0 = 100 \\ n_0 = 200 \end{cases}$	$\begin{cases} x_0 = 1000 \\ n_0 = 2000 \end{cases}$
0	0	3.85	0.42	0.12	0.10
1	0.1	6.99	2.29	1.11	0.99
2	0.2	9.44	6.51	4.67	4.42
3	0.3	11.19	12.54	11.88	11.74
4	0.4	12.24	18.07	20.21	20.48
5	0.5	12.59	20.33	24.02	24.55
6	0.6	12.24	18.07	20.21	20.48
7	0.7	11.19	12.54	11.88	11.74
8	0.8	9.44	6.51	4.67	4.42
9	0.9	6.99	2.29	1.11	0.99
10	1	3.84	0.42	0.12	0.10
$E(X_1)$		5	5	5	5
$\sigma[X_1]$		2.64	1.87	1.62	1.58

Inference and prediction with JAGS/rjags

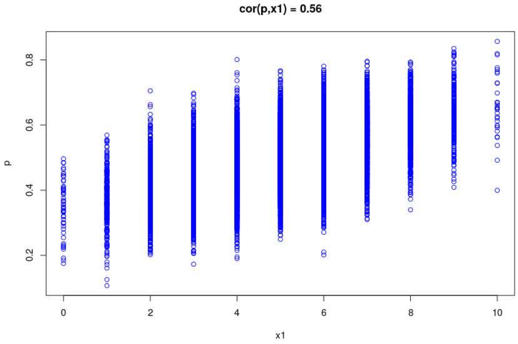
Scatter plot of sampled $f(p, x_1 \mid n_0, x_0, n_1)$

```
p  <- as.vector(chain[[1]][,1])
x1 <- as.vector(chain[[1]][,2])
plot(x1, p, col='blue',
     main=sprintf("cor(p,x1) = %.2f", cor(p,x1)))
print( table(x1)/10000 )
```

Inference and prediction with JAGS/rjags

Scatter plot of sampled $f(p, x_1 \mid n_0, x_0, n_1)$

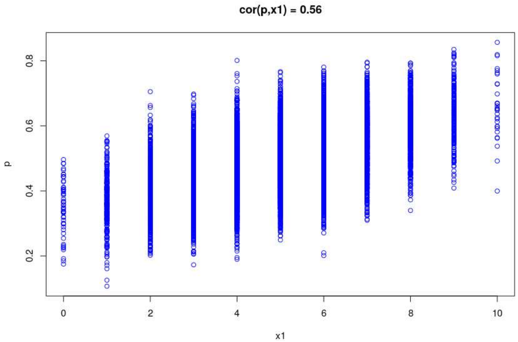
```
p  <- as.vector(chain[[1]][,1])
x1 <- as.vector(chain[[1]][,2])
plot(x1, p, col='blue',
     main=sprintf("cor(p,x1) = %.2f", cor(p,x1)))
print( table(x1)/10000 )
```



Inference and prediction with JAGS/rjags

Scatter plot of sampled $f(p, x_1 \mid n_0, x_0, n_1)$

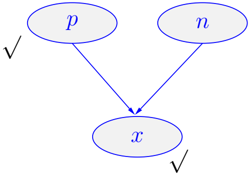
```
p  <- as.vector(chain[[1]][,1])
x1 <- as.vector(chain[[1]][,2])
plot(x1, p, col='blue',
     main=sprintf("cor(p,x1) = %.2f", cor(p,x1)))
print( table(x1)/10000 )
```



(The last command, print(. . .), produces the relative frequencies of occurrance of x_1 → **try it**)

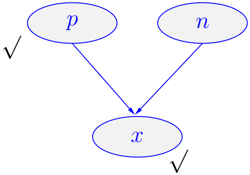
n independent Bernoulli processes

Inferring n



n independent Bernoulli processes

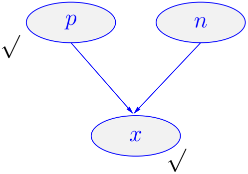
Inferring *n*



Think at a detector having a *well known efficiency* ($\epsilon \equiv p$)

n independent Bernoulli processes

Inferring n

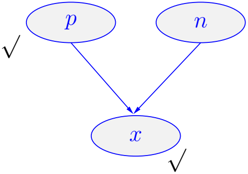


Think at a detector having a *well known efficiency* ($\epsilon \equiv p$):

- we have recorded x 'signals';

n independent Bernoulli processes

Inferring *n*

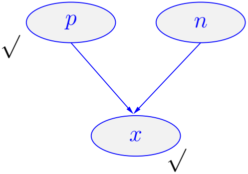


Think at a detector having a *well known efficiency* ($\epsilon \equiv p$):

- we have recorded x 'signals';
- how many particles impinged the detector?

n independent Bernoulli processes

Inferring n

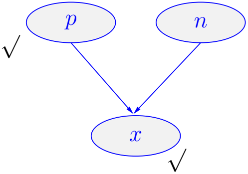


Think at a detector having a *well known efficiency ($\epsilon \equiv p$)*:

- we have recorded x 'signals';
- how many particles impinged the detector? $\longrightarrow f(n\,|\,x,p)$ **?**

n independent Bernoulli processes

Inferring *n*



Think at a detector having a *well known efficiency* $(\epsilon \equiv p)$:

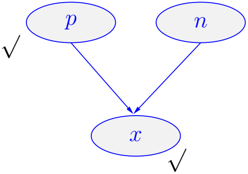
- we have recorded x 'signals';
- how many particles impinged the detector? $\longrightarrow f(n \,|\, x, p)$ **?**

Not to be confused with a different problem:

- a Poisson process has produced x in the measuring time T;
- what is λ of the related Poisson distribution?

n independent Bernoulli processes

Inferring n



Think at a detector having a *well known efficiency* ($\epsilon \equiv p$):

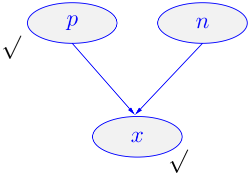
- we have recorded x 'signals';
- how many particles impinged the detector? $\longrightarrow f(n \,|\, x, p)$ **?**

Not to be confused with a different problem:

- a Poisson process has produced x in the measuring time T;
- what is λ of the related Poisson distribution? $\longrightarrow f(\lambda \,|\, x)$ **?**

n independent Bernoulli processes

Inferring *n*



Think at a detector having a *well known efficiency* ($\epsilon \equiv p$):

- we have recorded x 'signals';
- how many particles impinged the detector? $\longrightarrow f(n \mid x, p)$ **?**

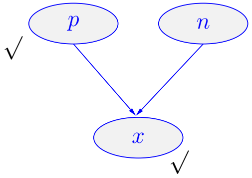
Not to be confused with a different problem:

- a Poisson process has produced x in the measuring time T;
- what is λ of the related Poisson distribution? $\longrightarrow f(\lambda \mid x)$ **?**
  [or, more precisely, what is the rate r? $\longrightarrow f(r \mid x, T)$ **?**]

n independent Bernoulli processes

Extending the model

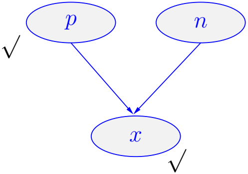
Our problem (but in Physics it is often not so simple)



n independent Bernoulli processes

Extending the model

Our problem (but in Physics it is often not so simple)

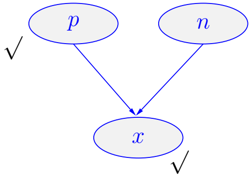


But we need some (usually indirect) knowledge about p

n independent Bernoulli processes

Extending the model

Our problem (but in Physics it is often not so simple)

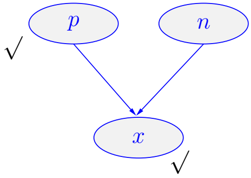


But we need some (usually indirect) knowledge about p
(Usually we do not calculate p from the fraction of white balls!)

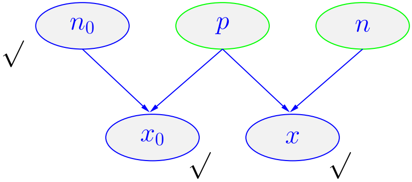
n independent Bernoulli processes

Extending the model

Our problem (but in Physics it is often not so simple)



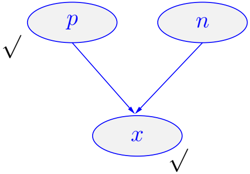
But we need some (usually indirect) knowledge about p
(Usually we do not calculate p from the fraction of white balls!)



n independent Bernoulli processes

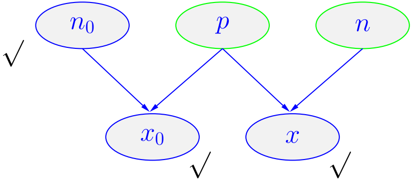
Extending the model

Our problem (but in Physics it is often not so simple)



But we need some (usually indirect) knowledge about p

(Usually we do not calculate p from the fraction of white balls!)



But what is n?

Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have 'physical meaning'.

Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have ‘physical meaning’.

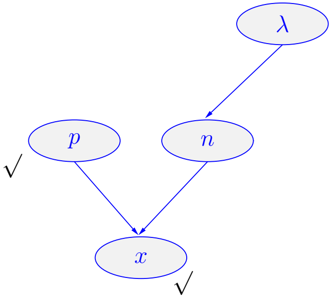
- When we say “we are uncertain on numbers”, we do not mean that we are uncertain on the numbers we ‘see’ in our detector, but to ‘other numbers’.
- Typically $n \longleftrightarrow \lambda$.

Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have 'physical meaning'.

- When we say "we are uncertain on numbers", we do not mean that we are uncertain on the numbers we 'see' in our detector, but to 'other numbers'.
- Typically $n \longleftrightarrow \lambda$.

Assuming for a while p well known and focusing on 'n':

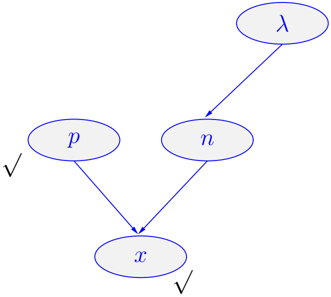


Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have 'physical meaning'.

- When we say "we are uncertain on numbers", we do not mean that we are uncertain on the numbers we 'see' in our detector, but to 'other numbers'.
- Typically $n \longleftrightarrow \lambda$.

Assuming for a while p well known and focusing on 'n':



But, as we have seen studying the **Poisson process**,

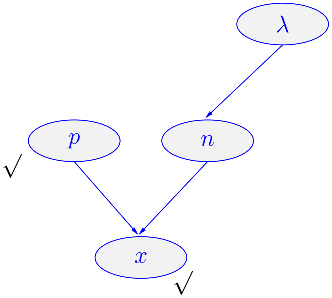
λ is not really physical

Extending the model

In Physics we are usually not interested in the numbers we do see, but in those which have 'physical meaning'.

- When we say "we are uncertain on numbers", we do not mean that we are uncertain on the numbers we 'see' in our detector, but to 'other numbers'.
- Typically $n \longleftrightarrow \lambda$.

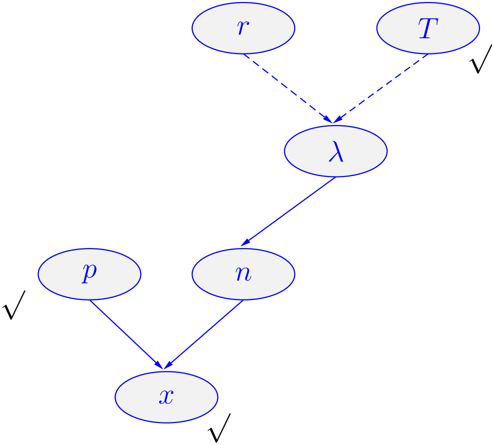
Assuming for a while p well known and focusing on 'n':



But, as we have seen studying the **Poisson process**,
λ is not really physical $\longrightarrow$

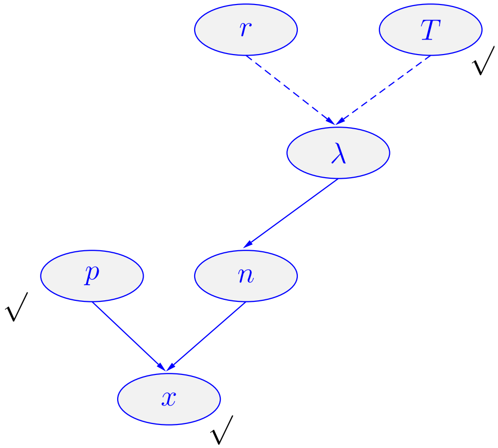
Extending the model

$\boldsymbol{\lambda} = r \cdot T$:



Extending the model

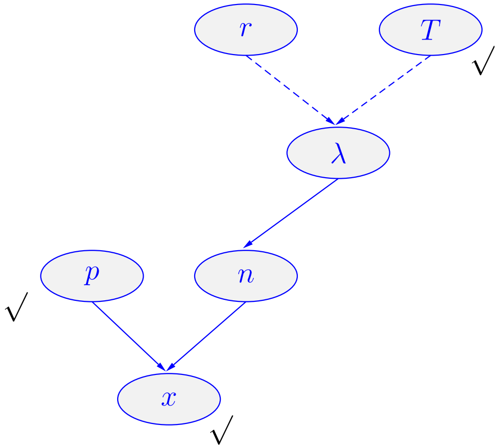
$\boldsymbol{\lambda} = r \cdot T$:



(Dashed arrows used in literature for deterministic links)

Extending the model

$\boldsymbol{\lambda} = r \cdot T$:

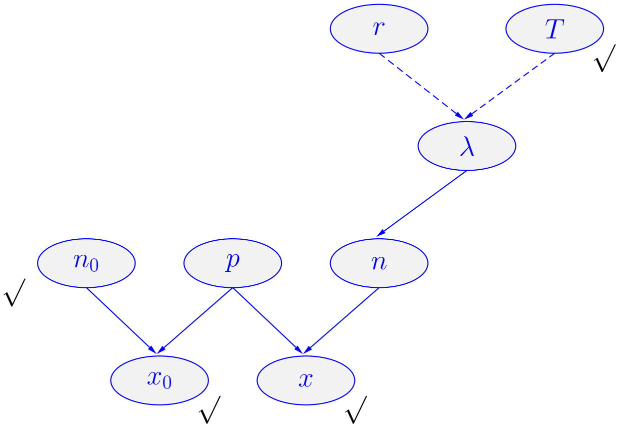


(Dashed arrows used in literature for deterministic links)

In JAGS, e.g., `lambda <- r * T;`

Extending the model

Remembering that p was got from a measurement:

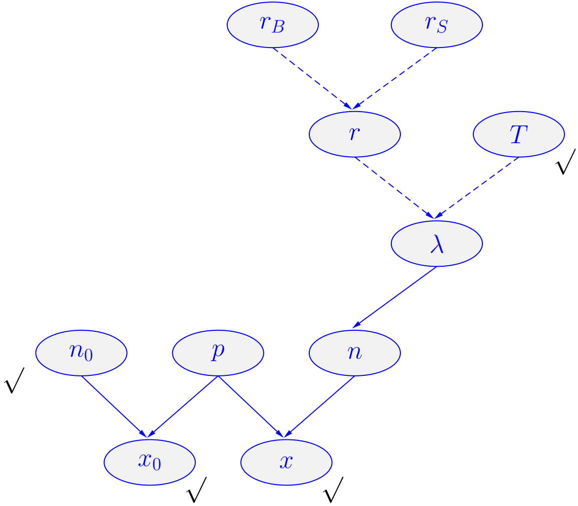


Extending the model

The rate r gets contributions from signal and background

Extending the model

The rate *r* gets contributions from signal and background



Extending the model

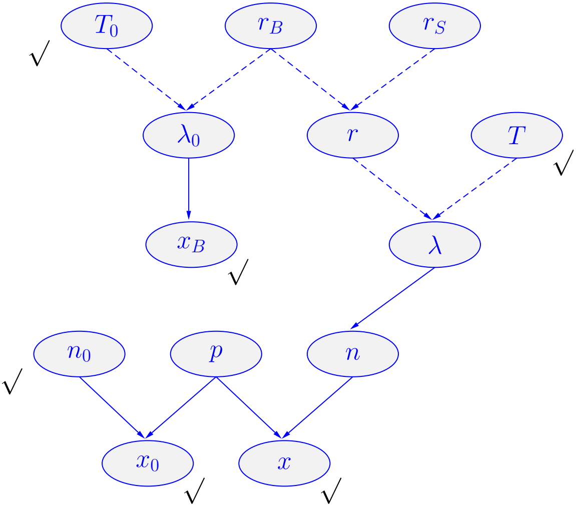
But, since $r = r_S + r_B$,

we need some independent knowledge of the background

Extending the model

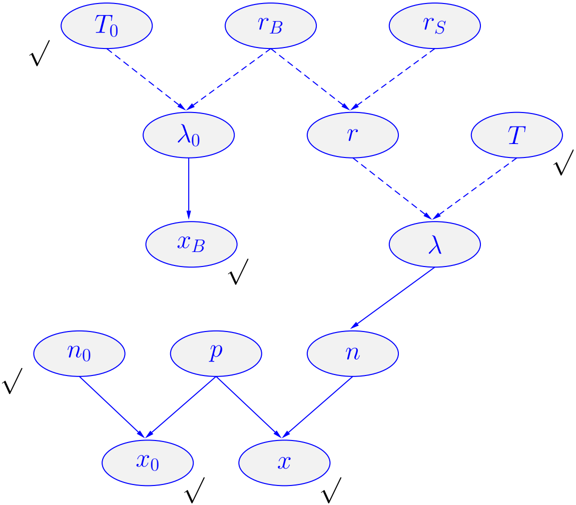
But, since $r = r_S + r_B$,

we need some independent knowledge of the background



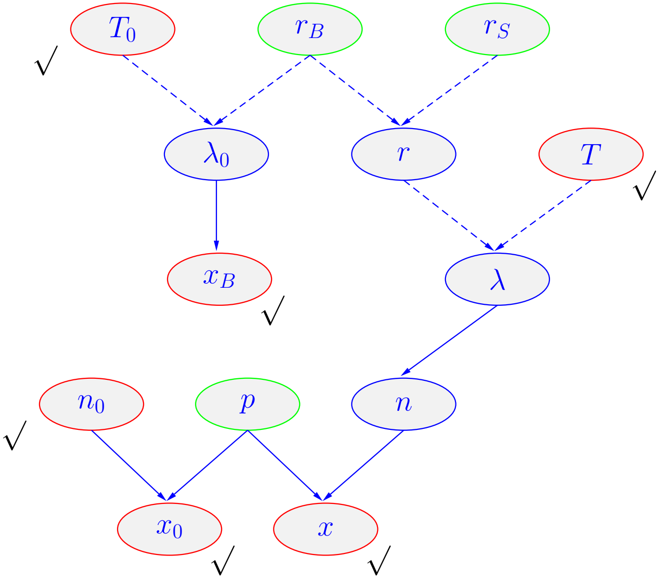
Extending the model

But, since $r = r_S + r_B$,
we need some independent knowledge of the background

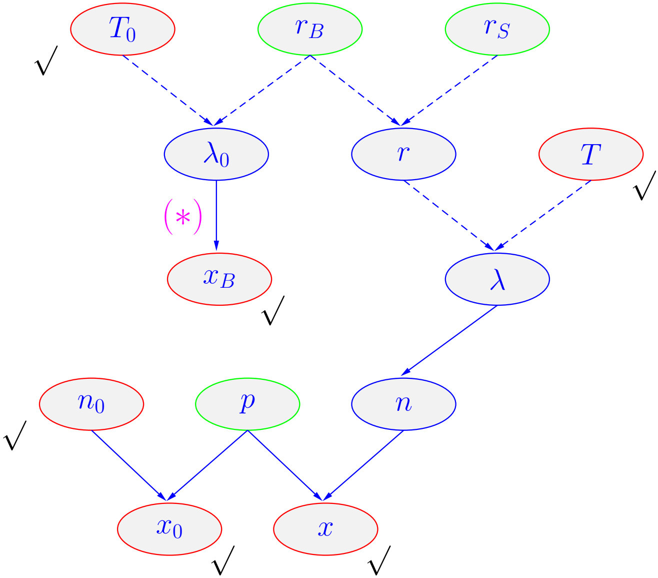


(T_0 and T assumed to be measured with sufficient accuracy)

Extending the model



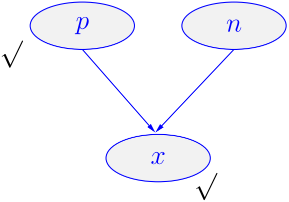
Extending the model



(*) Assuming unity efficiency

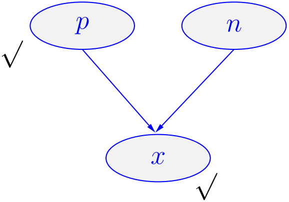
Inferring n 'assuming' p and x

Back to our initial problem



Inferring n 'assuming' p and x

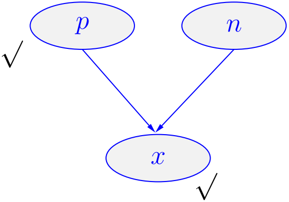
Back to our initial problem



$$f(n \,|\, p, x) \;\; \propto \;\; f(x \,|\, n, p) \cdot f_0(n)$$

Inferring n 'assuming' p and x

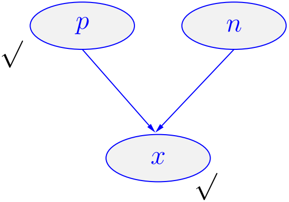
Back to our initial problem



$$f(n\,|\,p,x) \;\propto\; f(x\,|\,n,p)\cdot f_0(n)$$
$$\propto\; f(x\,|\,n,p) \qquad [\,\text{uniform prior}\,]$$

Inferring n 'assuming' p and x

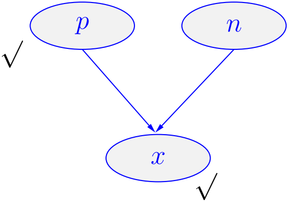
Back to our initial problem



$$
\begin{aligned}
f(n \,|\, p, x) &\propto f(x \,|\, n, p) \cdot f_0(n) \\
&\propto f(x \,|\, n, p) \qquad [\,\text{uniform prior}\,] \\
&\propto \frac{n!}{x!\,(n-x)!}\, p^x \cdot (1-p)^{n-x}
\end{aligned}
$$

Inferring n 'assuming' p and x

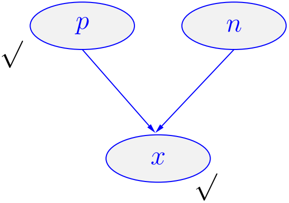
Back to our initial problem



$$
\begin{aligned}
f(n\,|\,p,x) &\propto f(x\,|\,n,p)\cdot f_0(n) \\
&\propto f(x\,|\,n,p) \qquad\qquad [\,\text{uniform prior}\,] \\
&\propto \frac{n!}{x!\,(n-x)!}\,p^x\cdot(1-p)^{n-x} \\
&\propto \frac{n!}{x!\,(n-x)!}\,p^x\cdot\frac{(1-p)^n}{(1-p)^x}
\end{aligned}
$$

Inferring n 'assuming' p and x

Back to our initial problem



$$
\begin{aligned}
f(n \mid p, x) &\propto f(x \mid n, p) \cdot f_0(n) \\
&\propto f(x \mid n, p) \qquad [\,\text{uniform prior}\,] \\
&\propto \frac{n!}{x!\,(n-x)!}\, p^x \cdot (1-p)^{n-x} \\
&\propto \frac{n!}{x!\,(n-x)!}\, p^x \cdot \frac{(1-p)^n}{(1-p)^x} \\
&\propto \frac{n!}{(n-x)!}\,(1-p)^n
\end{aligned}
$$

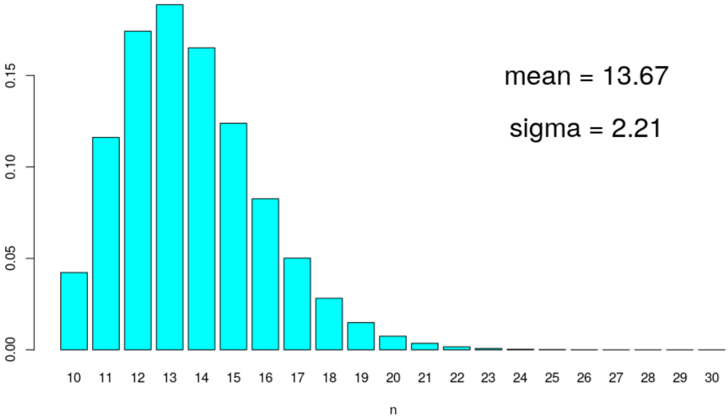
Inferring n 'assuming' p and x

Example in R with $p = 0.75$ and $x = 10$

```
p = 0.75; x = 10
n.max = 30

n  = x:n.max
fn = factorial(n)/factorial(n-x)*(1-p)^n
fn = fn/sum(fn)

media.n  = sum(fn*n)
media.n2 = sum(fn*n^2)
sigma.n  = sqrt(media.n2 - media.n^2)

barplot(fn, names=n, col='cyan', xlab='n')
text(20,0.15, sprintf("mean = %.2f", media.n), cex=2)
text(20,0.12, sprintf("sigma = %.2f", sigma.n),cex=2)
```

Inferring n 'assuming' p and x

$$f(n \,|\, x = 10, p = 0.75)$$



Inferring n 'assuming' p and x

Or we can feed JAGS with the following simple model

```
model{
    x ~ dbin(p, n);
    n ~ dnegbin(0.001, 1) I(nmin,);
}
```

Inferring n 'assuming' p and x

Or we can feed JAGS with the following simple model

```
model{
    x ~ dbin(p, n);
    n ~ dnegbin(0.001, 1) I(nmin,);
}
```

Remarks

- `dnegbin(0.001, 1)` is a 'negative binomial',
  in practice a geometric distribution 'from 0';

Inferring n 'assuming' p and x

Or we can feed JAGS with the following simple model

```
model{
    x ~ dbin(p, n);
    n ~ dnegbin(0.001, 1) I(nmin,);
}
```

Remarks

- `dnegbin(0.001, 1)` is a 'negative binomial', in practice a geometric distribution 'from 0';
- being $p = 0.001$, it has expected value 999

Inferring n 'assuming' p and x

Or we can feed JAGS with the following simple model

```
model{
    x ~ dbin(p, n);
    n ~ dnegbin(0.001, 1) I(nmin,);
}
```

Remarks

- dnegbin(0.001, 1) is a 'negative binomial', in practice a geometric distribution 'from 0';
- being $p = 0.001$, it has expected value 999 ($\frac{1}{p} - 1$, because it starts from 0)

Inferring n 'assuming' p and x

Or we can feed JAGS with the following simple model

```
model{
    x ~ dbin(p, n);
    n ~ dnegbin(0.001, 1) I(nmin,);
}
```

Remarks

- `dnegbin(0.001, 1)` is a 'negative binomial', in practice a geometric distribution 'from 0';
- being $p = 0.001$, it has expected value 999 ($\frac{1}{p} - 1$, because it starts from 0) and standard deviation 1000

Inferring n 'assuming' p and x

Or we can feed JAGS with the following simple model

```
model{
    x ~ dbin(p, n);
    n ~ dnegbin(0.001, 1) I(nmin,);
}
```

Remarks

- `dnegbin(0.001, 1)` is a 'negative binomial', in practice a geometric distribution 'from 0';
- being $p = 0.001$, it has expected value 999 ($\frac{1}{p} - 1$, because it starts from 0) and standard deviation 1000
  $\Rightarrow$ In practice, it is uniform in the region of interest
- `I(nmin,)` means that `n` cannot be smaller than `nmin`

Inferring n 'assuming' p and x

Or we can feed JAGS with the following simple model

```
model{
    x ~ dbin(p, n);
    n ~ dnegbin(0.001, 1) I(nmin,);
}
```

Remarks

- `dnegbin(0.001, 1)` is a 'negative binomial',
  in practice a geometric distribution 'from 0';
- being $p = 0.001$, it has expected value 999
  ($\frac{1}{p} - 1$, because it starts from 0)
  and standard deviation 1000
  $\Rightarrow$ In practice, it is uniform in the region of interest
- `I(nmin,)` means that `n` cannot be smaller than `nmin`
  (`nmin` is indeed equal to the observed `x`, for obvious reasons, but JAGS needs a separate constant)

Inferring n 'assuming' p and x

Or we can feed JAGS with the following simple model

```
model{
    x ~ dbin(p, n);
    n ~ dnegbin(0.001, 1) I(nmin,);
}
```

Remarks

- `dnegbin(0.001, 1)` is a 'negative binomial', in practice a geometric distribution 'from 0';
- being $p = 0.001$, it has expected value 999 ($\frac{1}{p} - 1$, because it starts from 0) and standard deviation 1000
  $\Rightarrow$ In practice, it is uniform in the region of interest
- `I(nmin,)` means that `n` cannot be smaller than `nmin` (`nmin` is indeed equal to the observed `x`, for obvious reasons, but JAGS needs a separate constant)
  The remaining R code is left as **exercise**

Inferring Poisson's λ

- set up the problem;

Inferring Poisson's λ

- set up the problem;
- solution for uniform prior;

Inferring Poisson's λ

- set up the problem;
- solution for uniform prior;
- the case of no events observed;

Inferring Poisson's λ

- set up the problem;
- solution for uniform prior;
- the case of no events observed;
- prior conjugate;

Inferring Poisson's λ

- set up the problem;
- solution for uniform prior;
- the case of no events observed;
- prior conjugate;
- predictive distribution;

Inferring Poisson's λ

- set up the problem;
- solution for uniform prior;
- the case of no events observed;
- prior conjugate;
- predictive distribution;
- from λ to r (not covered, since it is straightforward; but remember that the 'physical quantity' is r)

Inferring Poisson's λ

$$f(\lambda \,|\, x, \mathcal{P}) \;=\; \frac{\frac{\lambda^x\, e^{-\lambda}}{x!}\, f_\circ(\lambda)}{\int_0^\infty \frac{\lambda^x\, e^{-\lambda}}{x!}\, f_\circ(\lambda)\,\mathrm{d}\lambda}\,.$$

Inferring Poisson's λ

$$f(\lambda \,|\, x, \mathcal{P}) \;=\; \frac{\frac{\lambda^x\, e^{-\lambda}}{x!}\, f_\circ(\lambda)}{\int_0^\infty \frac{\lambda^x\, e^{-\lambda}}{x!}\, f_\circ(\lambda)\, \mathrm{d}\lambda}\,.$$

Assuming $f_\circ(\lambda)$ constant up to a certain $\lambda_{max} \gg x$ and making the integral by parts we obtain

$$f(\lambda \,|\, x, \mathcal{P}) \;=\; \frac{\lambda^x\, e^{-\lambda}}{x!}$$

$$F(\lambda \,|\, x, \mathcal{P}) \;=\; 1 - e^{-\lambda}\left(\sum_{n=0}^{x} \frac{\lambda^n}{n!}\right)$$

Inferring Poisson's λ

$$f(\lambda \,|\, x, \mathcal{P}) \;=\; \frac{\frac{\lambda^x\, e^{-\lambda}}{x!}\, f_\circ(\lambda)}{\int_0^\infty \frac{\lambda^x\, e^{-\lambda}}{x!}\, f_\circ(\lambda)\, \mathrm{d}\lambda}\,.$$

Assuming $f_\circ(\lambda)$ constant up to a certain $\lambda_{max} \gg x$ and making the integral by parts we obtain

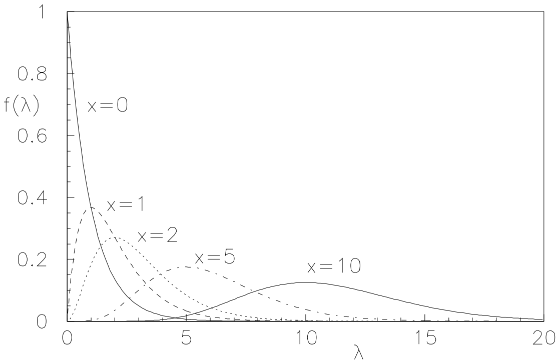
$$f(\lambda \,|\, x, \mathcal{P}) \;=\; \frac{\lambda^x\, e^{-\lambda}}{x!}$$

$$F(\lambda \,|\, x, \mathcal{P}) \;=\; 1 - e^{-\lambda}\left(\sum_{n=0}^{x} \frac{\lambda^n}{n!}\right)$$

Summaries

$$\mathrm{E}(\lambda) \;=\; x+1,$$
$$\mathrm{Var}(\lambda) \;=\; x+1,$$
$$\lambda_m \;=\; x$$

Some examples of $f(\lambda)$

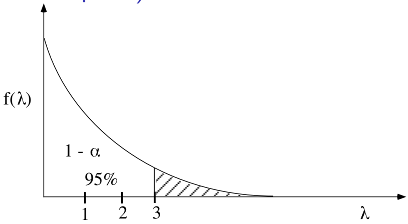


For 'large' x $f(\lambda)$ it becomes Gaussian with expected value x and standard deviation $\sqrt{x}$.

The difference between the most probable λ and its expected value for small x is due to the asymmetry of $f(\lambda)$.

Inferring λ from $x = 0$

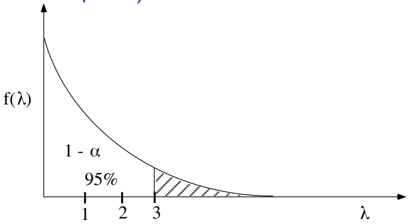
(From a flat prior!)



$$
\begin{aligned}
f(\lambda \,|\, x = 0, \mathcal{P}) &= e^{-\lambda} \\
F(\lambda \,|\, x = 0, \mathcal{P}) &= 1 - e^{-\lambda}
\end{aligned}
$$

Inferring λ from $x = 0$

(From a flat prior!)



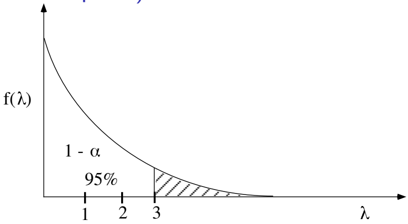
$$f(\lambda\,|\,x=0,\mathcal{P}) \;=\; e^{-\lambda}$$
$$F(\lambda\,|\,x=0,\mathcal{P}) \;=\; 1-e^{-\lambda}$$

Upper probabilistic limit (e.g. at 95% probability):

$$P(\lambda \le \lambda_u\,|\,x=0) = F(\lambda_u\,|\,x=0) \;=\; 0.95$$

Inferring λ from $x = 0$

(From a flat prior!)



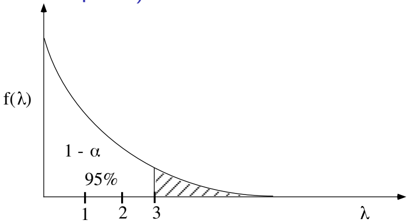
$$f(\lambda\,|\,x=0,\mathcal{P}) = e^{-\lambda}$$
$$F(\lambda\,|\,x=0,\mathcal{P}) = 1-e^{-\lambda}$$

Upper probabilistic limit (e.g. at 95% probability):

$$P(\lambda \le \lambda_u\,|\,x=0) = F(\lambda_u\,|\,x=0) = 0.95$$
$$1 - F(\lambda_u\,|\,x=0) = e^{-\lambda_u} = 0.05$$

Inferring λ from $x = 0$

(From a flat prior!)



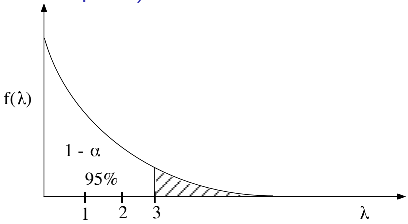
$$
\begin{aligned}
f(\lambda \,|\, x=0, \mathcal{P}) &= e^{-\lambda} \\
F(\lambda \,|\, x=0, \mathcal{P}) &= 1 - e^{-\lambda}
\end{aligned}
$$

Upper probabilistic limit (e.g. at 95% probability):

$$
\begin{aligned}
P(\lambda \le \lambda_u \,|\, x=0) = F(\lambda_u \,|\, x=0) &= 0.95 \\
1 - F(\lambda_u \,|\, x=0) = e^{-\lambda_u} &= 0.05 \\
\lambda_u &= 3 \text{ at } 95\,\% \text{ probability}\,.
\end{aligned}
$$

Inferring λ from $x = 0$

(From a flat prior!)



$$\begin{aligned} f(\lambda \,|\, x=0, \mathcal{P}) &= e^{-\lambda} \\ F(\lambda \,|\, x=0, \mathcal{P}) &= 1 - e^{-\lambda} \end{aligned}$$

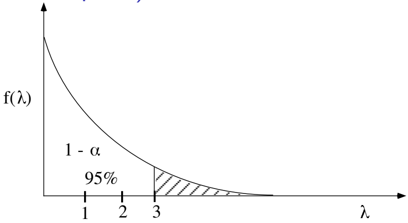
Upper probabilistic limit (e.g. at 95% probability):

$$\begin{aligned} P(\lambda \le \lambda_u \,|\, x=0) = F(\lambda_u \,|\, x=0) &= 0.95 \\ 1 - F(\lambda_u \,|\, x=0) = e^{-\lambda_u} &= 0.05 \\ \lambda_u &= 3 \text{ at } 95\,\% \text{ probability}\,. \end{aligned}$$

But not because $f(x=0 \,|\, \lambda=3) = e^{-3} = 0.05$!

Inferring λ from $x = 0$

(From a flat prior!)



$$f(\lambda \,|\, x=0, \mathcal{P}) = e^{-\lambda}$$
$$F(\lambda \,|\, x=0, \mathcal{P}) = 1 - e^{-\lambda}$$

Upper probabilistic limit (e.g. at 95% probability):

$$P(\lambda \le \lambda_u \,|\, x=0) = F(\lambda_u \,|\, x=0) = 0.95$$
$$1 - F(\lambda_u \,|\, x=0) = e^{-\lambda_u} = 0.05$$
$$\lambda_u = 3 \text{ at } 95\,\% \text{ probability}\,.$$

But not because $f(x=0 \,|\, \lambda=3) = e^{-3} = 0.05$!

In this case it works just by chance

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Do you remember? (From first lecture)

In general $P(A \,|\, B) \neq P(B \,|\, A)$

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

Do you remember? (From first lecture)

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

- $P(\text{Positive} \mid \overline{HIV}) \neq P(\overline{HIV} \mid \text{Positive})$
- $P(\text{Win} \mid \text{Play}) \neq P(\text{Play} \mid \text{Win})$ [Lotto]
- $P(\text{Pregnant} \mid \text{Woman}) \neq P(\text{Woman} \mid \text{Pregnant})$

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Do you remember? (From first lecture)

$$\text{In general } P(A \,|\, B) \neq P(B \,|\, A)$$

- $P(\text{Positive} \,|\, \overline{HIV}) \neq P(\overline{HIV} \,|\, \text{Positive})$
- $P(\text{Win} \,|\, \text{Play}) \neq P(\text{Play} \,|\, \text{Win})$ [Lotto]
- $P(\text{Pregnant} \,|\, \text{Woman}) \neq P(\text{Woman} \,|\, \text{Pregnant})$

Everyone was laughing, but this is *more or less* the 'logic' behind frequentistic CL upper/lower bounds

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Do you remember? (From first lecture)

In general $P(A \,|\, B) \neq P(B \,|\, A)$

- $P(\text{Positive} \,|\, \overline{HIV}) \neq P(\overline{HIV} \,|\, \text{Positive})$
- $P(\text{Win} \,|\, \text{Play}) \neq P(\text{Play} \,|\, \text{Win})$ [Lotto]
- $P(\text{Pregnant} \,|\, \text{Woman}) \neq P(\text{Woman} \,|\, \text{Pregnant})$

Everyone was laughing, but this is *more or less* the 'logic' behind frequentistic CL upper/lower bounds

Very little to laugh...

Inferring Poisson's λ

Conjugate prior

$$f(\lambda \,|\, x) \quad \propto \quad \lambda^x \, e^{-\lambda} \cdot f_\circ(\lambda)$$

Inferring Poisson's λ

Conjugate prior

$$
\begin{aligned}
f(\lambda \,|\, x) &\propto \lambda^x \, e^{-\lambda} \cdot f_\circ(\lambda) \\
&\propto \lambda^x \, e^{-\lambda} \cdot \lambda^a \, e^{-b\,\lambda}
\end{aligned}
$$

Inferring Poisson's λ

Conjugate prior

$$
\begin{aligned}
f(\lambda\,|\,x) &\propto \lambda^x\, e^{-\lambda} \cdot f_\circ(\lambda) \\
&\propto \lambda^x\, e^{-\lambda} \cdot \lambda^a\, e^{-b\,\lambda} \\
&\propto \lambda^{x+a}\, e^{-(1+b)\,\lambda}
\end{aligned}
$$

Does such a probability function 'exist'?

Inferring Poisson's λ

Conjugate prior

$$
\begin{aligned}
f(\lambda\,|\,x) &\propto \lambda^x\, e^{-\lambda} \cdot f_\circ(\lambda) \\
&\propto \lambda^x\, e^{-\lambda} \cdot \lambda^a\, e^{-b\,\lambda} \\
&\propto \lambda^{x+a}\, e^{-(1+b)\,\lambda}
\end{aligned}
$$

Does such a probability function 'exist'?

$\Rightarrow$ Gamma distribution

Gamma distribution

$X \sim \text{Gamma}(c, r)$:

$$f(x \,|\, \text{Gamma}(c, r)) = \frac{r^c}{\Gamma(c)} x^{c-1} e^{-r\,x} \qquad \begin{cases} r,\; c > 0 \\ x \geq 0 \end{cases} ,$$

where

$$\Gamma(c) = \int_0^\infty x^{c-1} e^{-x} \mathrm{d}x$$

(for n integer, $\Gamma(n+1) = n!$).

Gamma distribution

$X \sim \text{Gamma}(c, r)$:

$$f(x \,|\, \text{Gamma}(c, r)) = \frac{r^c}{\Gamma(c)} x^{c-1} e^{-r\,x} \qquad \left\{ \begin{array}{l} r,\ c > 0 \\ x \geq 0 \end{array} \right. ,$$

where

$$\Gamma(c) = \int_0^\infty x^{c-1} e^{-x} \mathrm{d}x$$

(for n integer, $\Gamma(n+1) = n!$).

c is called *shape* parameter, while $1/r$ is the *scale* parameter.

Gamma distribution

$X \sim \text{Gamma}(c, r)$:

$$f(x \,|\, \text{Gamma}(c, r)) = \frac{r^c}{\Gamma(c)} x^{c-1} e^{-r\,x} \qquad \begin{cases} r,\ c > 0 \\ x \geq 0 \end{cases} \quad ,$$

where

$$\Gamma(c) = \int_0^\infty x^{c-1} e^{-x} \mathrm{d}x$$

(for n integer, $\Gamma(n+1) = n!$).

c is called *shape* parameter, while $1/r$ is the *scale* parameter.

- If c is integer, the distribution is also known as Erlang, describing the time to wait before observing the c-th event in a Poisson process of intensity ('rate') r.

Gamma distribution

$X \sim \text{Gamma}(c, r)$:

$$f(x \,|\, \text{Gamma}(c, r)) = \frac{r^c}{\Gamma(c)} x^{c-1} e^{-r\,x} \qquad \left\{ \begin{array}{l} r,\ c > 0 \\ x \geq 0 \end{array} \right. ,$$

where

$$\Gamma(c) = \int_0^{\infty} x^{c-1} e^{-x} \mathrm{d}x$$

(for n integer, $\Gamma(n+1) = n!$).

c is called *shape* parameter, while $1/r$ is the *scale* parameter.

- If c is integer, the distribution is also known as Erlang, describing the time to wait before observing the c-th event in a Poisson process of intensity ('rate') r.
- For $c = 1$ the Gamma distribution recovers the exponential.

Gamma distribution

$X \sim \text{Gamma}(c, r)$:

$$f(x\,|\,\text{Gamma}(c,r)) = \frac{r^c}{\Gamma(c)} x^{c-1} e^{-r\,x} \qquad \begin{cases} r,\ c > 0 \\ x \geq 0 \end{cases} \quad ,$$

where

$$\Gamma(c) = \int_0^\infty x^{c-1} e^{-x} \mathrm{d}x$$

(for n integer, $\Gamma(n+1) = n!$).

c is called *shape* parameter, while $1/r$ is the *scale* parameter.

- If c is integer, the distribution is also known as Erlang, describing the time to wait before observing the c-th event in a Poisson process of intensity ('rate') r.
- For $c = 1$ the Gamma distribution recovers the exponential.
- Finally, the χ^2 distribution is just a particular Gamma:

$$f(x\,|\,\chi^2_\nu) \;=\; f(x\,|\,\text{Gamma}(\nu/2,\, 1/2))$$

Gamma distribution

$X \sim \text{Gamma}(c, r)$:

$$f(x\,|\,\text{Gamma}(c,r)) = \frac{r^c}{\Gamma(c)} x^{c-1} e^{-r\,x} \qquad \begin{cases} r,\ c > 0 \\ x \geq 0 \end{cases} ,$$

where

$$\Gamma(c) = \int_0^\infty x^{c-1} e^{-x} \mathrm{d}x$$

(for n integer, $\Gamma(n+1) = n!$).

c is called *shape* parameter, while $1/r$ is the *scale* parameter.

- If c is integer, the distribution is also known as Erlang, describing the time to wait before observing the c-th event in a Poisson process of intensity ('rate') r.
- For $c = 1$ the Gamma distribution recovers the exponential.
- Finally, the χ^2 distribution is just a particular Gamma:

$$f(x\,|\,\chi^2_\nu) \;=\; f(x\,|\,\text{Gamma}(\nu/2,\, 1/2))$$

- **The Gamma is a key distribution!**

Gamma distribution

$X \sim \text{Gamma}(c, r)$:

$$f(x\,|\,\text{Gamma}(c,r)) = \frac{r^c}{\Gamma(c)} x^{c-1} e^{-r\,x} \qquad \left\{ \begin{array}{l} r,\ c > 0 \\ x \geq 0 \end{array} \right. ,$$

where

$$\Gamma(c) = \int_0^\infty x^{c-1} e^{-x} \mathrm{d}x$$

(for n integer, $\Gamma(n+1) = n!$).

c is called *shape* parameter, while $1/r$ is the *scale* parameter.

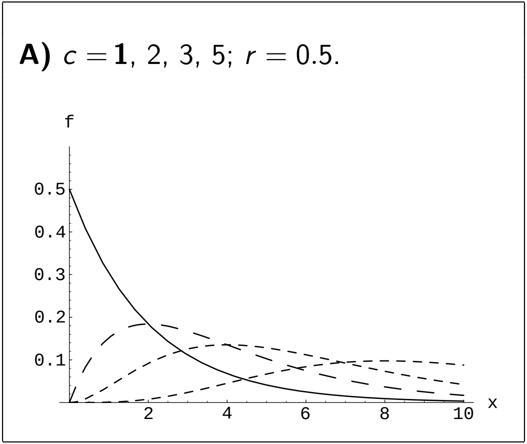
- If c is integer, the distribution is also known as Erlang, describing the time to wait before observing the c-th event in a Poisson process of intensity ('rate') r.
- For $c = 1$ the Gamma distribution recovers the exponential.
- Finally, the χ^2 distribution is just a particular Gamma:

$$f(x\,|\,\chi^2_\nu) \quad = \quad f(x\,|\,\text{Gamma}(\nu/2,\, 1/2))$$

- **The Gamma is a key distribution!**
  The Erlang distribution is important to get a physical intuition of the properties of Gamma and then of the χ^2!

Gamma distribution

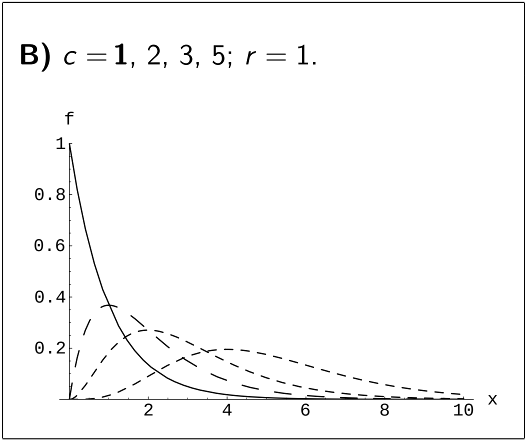
Some examples



r: rate (if the variable is a time, then r is Poisson rate).

Gamma distribution

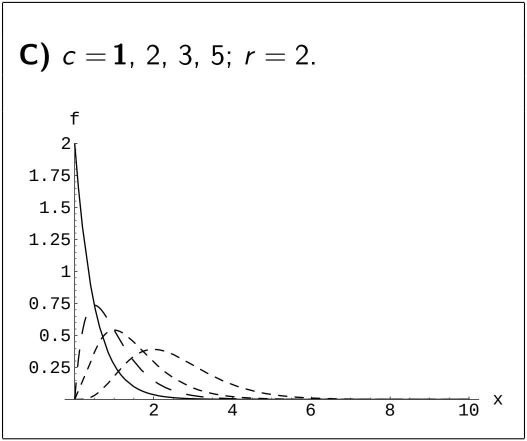
Some examples



r: rate (rate increases $\rightarrow$ distributions squized)

Gamma distribution

Some examples



r: rate (rate increases $\rightarrow$ distributions squized)

Gamma (and χ^2) distribution

Summaries

$$
\begin{aligned}
E(X) &= \frac{c}{r} \\
\mathrm{Var}(X) &= \frac{c}{r^2} = \frac{E(X)}{r} \\
\mathrm{mode}(X) &= \begin{cases} 0 & \text{if } c \le 1 \\ \frac{c-1}{r} & \text{if } c > 1 \end{cases}
\end{aligned}
$$

Gamma (and χ^2) distribution

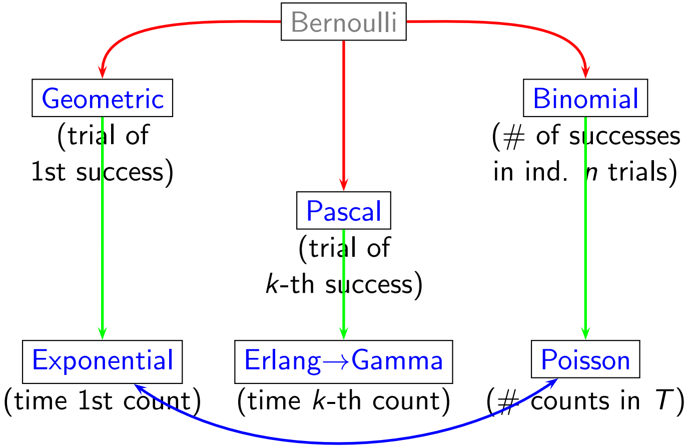
Summaries

$$\begin{aligned}
\mathrm{E}(X) &= \frac{c}{r} \\
\mathrm{Var}(X) &= \frac{c}{r^2} = \frac{\mathrm{E}(X)}{r} \\
\mathrm{mode}(X) &= \begin{cases} 0 & \text{if } c \leq 1 \\ \frac{c-1}{r} & \text{if } c > 1 \end{cases}
\end{aligned}$$

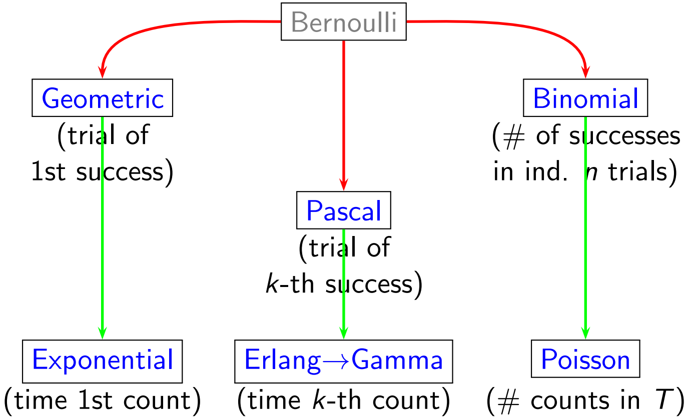
Therefore, for the χ^2 ($\rightarrow c = \nu/2$, $r = 1/2$)

$$\begin{aligned}
\mathrm{E}(\chi^2) &= \nu \\
\mathrm{Var}(\chi^2) &= 2\,\nu \\
\mathrm{mode}(\chi^2) &= \begin{cases} 0 & \text{if } \nu \leq 2 \\ \nu - 2 & \text{if } \nu > 2 \end{cases}
\end{aligned}$$

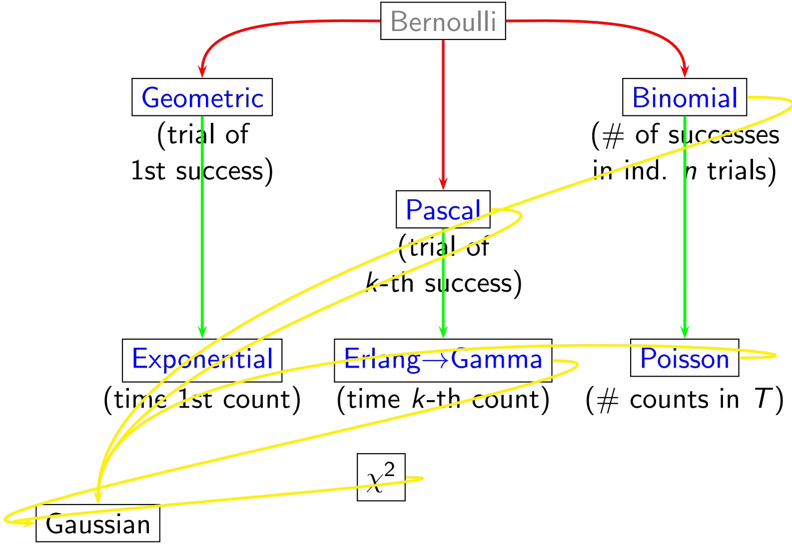
Distributions derived from the Bernoulli process



Distributions derived from the Bernoulli process



Distributions derived from the Bernoulli process



Inferring the Poisson's λ

Use of gamma conjugate prior

- $$f(\lambda \,|\, x, \text{Gamma}(c_i, r_i)) \;\propto\; \left[\lambda^x e^{-\lambda}\right] \times \left[\lambda^{c_i-1} e^{-r_i\,\lambda}\right]$$

Inferring the Poisson's λ

Use of gamma conjugate prior

- $$\begin{aligned} f(\lambda \,|\, x, \text{Gamma}(c_i, r_i)) &\propto \left[\lambda^x e^{-\lambda}\right] \times \left[\lambda^{c_i - 1} e^{-r_i\,\lambda}\right] \\ &\propto \lambda^{x+c_i-1} e^{-(r_i+1)\,\lambda}\,, \end{aligned}$$

  where c_i and r_i are the initial parameters of the gamma distribution.

Inferring the Poisson's λ

Use of gamma conjugate prior

- $$\begin{aligned} f(\lambda \,|\, x, \text{Gamma}(c_i, r_i)) &\propto \left[\lambda^x e^{-\lambda}\right] \times \left[\lambda^{c_i-1} e^{-r_i\,\lambda}\right] \\ &\propto \lambda^{x+c_i-1} e^{-(r_i+1)\,\lambda}\,, \end{aligned}$$

  where c_i and r_i are the initial parameters of the gamma distribution.

- Updating rule

$$\begin{aligned} c_f &= c_i + x \\ r_f &= r_i + 1 \end{aligned}$$

Inferring the Poisson's λ

Use of gamma conjugate prior

- $$\begin{aligned} f(\lambda \,|\, x, \text{Gamma}(c_i, r_i)) &\propto \left[\lambda^x e^{-\lambda}\right] \times \left[\lambda^{c_i - 1} e^{-r_i \lambda}\right] \\ &\propto \lambda^{x + c_i - 1} e^{-(r_i + 1)\lambda}, \end{aligned}$$

  where c_i and r_i are the initial parameters of the gamma distribution.
- Updating rule
  $$\begin{aligned} c_f &= c_i + x \\ r_f &= r_i + 1 \end{aligned}$$
- A "flat conjugate" prior (not just academic!):

Inferring the Poisson's λ

Use of gamma conjugate prior

- $$\begin{aligned} f(\lambda \,|\, x, \text{Gamma}(c_i, r_i)) &\propto \left[\lambda^x e^{-\lambda}\right] \times \left[\lambda^{c_i-1} e^{-r_i\,\lambda}\right] \\ &\propto \lambda^{x+c_i-1} e^{-(r_i+1)\,\lambda}\,, \end{aligned}$$

  where c_i and r_i are the initial parameters of the gamma distribution.
- Updating rule
  $$\begin{aligned} c_f &= c_i + x \\ r_f &= r_i + 1 \end{aligned}$$
- A "flat conjugate" prior (not just academic!):
  $\rightarrow$ exponential with very large τ (or vanishing r)

Inferring the Poisson's λ

Use of gamma conjugate prior

- $$\begin{aligned} f(\lambda \,|\, x, \text{Gamma}(c_i, r_i)) &\propto \left[\lambda^x e^{-\lambda}\right] \times \left[\lambda^{c_i - 1} e^{-r_i\,\lambda}\right] \\ &\propto \lambda^{x + c_i - 1} e^{-(r_i + 1)\,\lambda}, \end{aligned}$$

  where c_i and r_i are the initial parameters of the gamma distribution.
- Updating rule
  $$\begin{aligned} c_f &= c_i + x \\ r_f &= r_i + 1 \end{aligned}$$
- A "flat conjugate" prior (not just academic!):
  → exponential with very large τ (or vanishing r)
- $c = 1$, $r \to 0$
  $$f(\lambda \,|\, x, \text{Gamma}(c_i = 1, r_i \to 0)) \propto \lambda^x e^{-\lambda}$$

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

What shall we get in a future measurement?

(assuming same r and T)

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

What shall we get in a future measurement?

(assuming same r and T)

$$f(x_f \,|\, x_p) \;=\; \int_0^{\infty} f(x_f \,|\, \lambda) \cdot f(\lambda \,|\, x_p)\, \mathrm{d}\lambda$$

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

What shall we get in a future measurement?

(assuming same r and T)

$$f(x_f\,|\,x_p) \;=\; \int_0^\infty f(x_f\,|\,\lambda)\cdot f(\lambda\,|\,x_p)\,\mathrm{d}\lambda$$

Left as exercise, e.g. numerically or with JAGS

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

What shall we get in a future measurement?

(assuming same r and T)

$$f(x_f \,|\, x_p) \;=\; \int_0^\infty f(x_f \,|\, \lambda) \cdot f(\lambda \,|\, x_p) \, \mathrm{d}\lambda$$

Left as exercise, e.g. numerically or with JAGS

Just intuitive arguments for large number behaviour (e.g. $x_p = 100$)

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

What shall we get in a future measurement?

(assuming same r and T)

$$f(x_f \,|\, x_p) \;=\; \int_0^{\infty} f(x_f \,|\, \lambda) \cdot f(\lambda \,|\, x_p)\, \mathrm{d}\lambda$$

Left as exercise, e.g. numerically or with JAGS

Just intuitive arguments for large number behaviour (e.g. $x_p = 100$)

- λ will be ≈ 100, with 'standard uncertainty' ≈ 10:
  $\rightarrow \lambda = 100 \pm 10$;

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

What shall we get in a future measurement?

(assuming same r and T)

$$f(x_f\,|\,x_p) \;=\; \int_0^\infty f(x_f\,|\,\lambda)\cdot f(\lambda\,|\,x_p)\,\mathrm{d}\lambda$$

Left as exercise, e.g. numerically or with JAGS

Just intuitive arguments for large number behaviour (e.g. $x_p = 100$)

- λ will be ≈ 100, with 'standard uncertainty' ≈ 10: $\rightarrow \lambda = 100 \pm 10$;
- <u>if we were</u> sure that λ was 100, <u>then</u> $x_f = 100 \pm 10$.

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

What shall we get in a future measurement?

(assuming same r and T)

$$f(x_f\,|\,x_p) \;=\; \int_0^\infty f(x_f\,|\,\lambda)\cdot f(\lambda\,|\,x_p)\,\mathrm{d}\lambda$$

Left as exercise, e.g. numerically or with JAGS

Just intuitive arguments for large number behaviour (e.g. $x_p = 100$)

- λ will be ≈ 100, with 'standard uncertainty' ≈ 10: $\rightarrow \lambda = 100 \pm 10$;
- if we were sure that λ was 100, then $x_f = 100 \pm 10$.
- but we have to 'convolute' our uncertainty concerning λ

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

What shall we get in a future measurement?

(assuming same r and T)

$$f(x_f \,|\, x_p) \;=\; \int_0^{\infty} f(x_f \,|\, \lambda) \cdot f(\lambda \,|\, x_p)\, \mathrm{d}\lambda$$

Left as exercise, e.g. numerically or with JAGS

Just intuitive arguments for large number behaviour (e.g. $x_p = 100$)

- λ will be ≈ 100, with 'standard uncertainty' ≈ 10: $\rightarrow \lambda = 100 \pm 10$;
- if we were sure that λ was 100, then $x_f = 100 \pm 10$.
- but we have to 'convolute' our uncertainty concerning λ $\rightarrow$ uncertainty about x_f has to increase;

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

What shall we get in a future measurement?

(assuming same r and T)

$$f(x_f \,|\, x_p) = \int_0^\infty f(x_f \,|\, \lambda) \cdot f(\lambda \,|\, x_p)\, \mathrm{d}\lambda$$

Left as exercise, e.g. numerically or with JAGS

Just intuitive arguments for large number behaviour (e.g. $x_p = 100$)

- λ will be ≈ 100, with 'standard uncertainty' ≈ 10:
  $\rightarrow \lambda = 100 \pm 10$;
- if we were sure that λ was 100, then $x_f = 100 \pm 10$.
- but we have to 'convolute' our uncertainty concerning λ
  $\rightarrow$ uncertainty about x_f has to increase;
- by how much?

Predictive distribution

We have seen how to learn about λ given the observed x (hereafter x_p)

What shall we get in a future measurement?

(assuming same r and T)

$$f(x_f \,|\, x_p) \;=\; \int_0^\infty f(x_f \,|\, \lambda)\cdot f(\lambda \,|\, x_p)\,\mathrm{d}\lambda$$

Left as exercise, e.g. numerically or with JAGS

Just intuitive arguments for large number behaviour (e.g. $x_p = 100$)

- λ will be ≈ 100, with 'standard uncertainty' ≈ 10: $\rightarrow \lambda = 100 \pm 10$;
- if we were sure that λ was 100, then $x_f = 100 \pm 10$.
- but we have to 'convolute' our uncertainty concerning λ $\rightarrow$ uncertainty about x_f has to increase;
- by how much? $\rightarrow$ Left as exercise

Example with JAGS

```
# inf_lambda_pred.bug
model {
   X ~ dpois(lambda);
   lambda ~ dexp(0.00001)
   Y ~ dpois(lambda);
}

# inf_lambda_pred.R
library(rjags)
modello = "inf_lambda_pred.bug"  # file con il modello
dati <- NULL  # oggetto con i dati
dati$X  <- 100
jm <- jags.model(modello, dati)  # definisce il modello
update(jm, 100)                  # burn in
catena <- coda.samples(jm, c("lambda","Y"), n.iter=10000)
print(summary(catena))
plot(catena)
```

The End