Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

Back to the exponential distribution

Short reminder

We have seen how the exponential distribution arises from the Poisson process.

Back to the exponential distribution

Short reminder

We have seen how the exponential distribution arises from the Poisson process.

In particular

- Poisson distribution and Exponential distributions are two different points of view on the same process;

Back to the exponential distribution

Short reminder

We have seen how the exponential distribution arises from the Poisson process.

In particular

- Poisson distribution and Exponential distributions are two different points of view on the same process;
- the parameters of the two distributions are hence related, depending on the physical one, r:

$$\begin{aligned} \tau &= 1/r \\ \lambda &= r\,T \\ &= \frac{T}{\tau} \end{aligned}$$

Back to the exponential distribution

Short reminder

We have seen how the exponential distribution arises from the Poisson process.

In particular

- Poisson distribution and Exponential distributions are two different points of view on the same process;
- the parameters of the two distributions are hence related, depending on the physical one, r:

$$\begin{aligned} \tau &= 1/r \\ \lambda &= r\,T \\ &= \frac{T}{\tau} \end{aligned}$$

- The Exponential is a Geometric in the continuum (it makes no sense to speak about the "precise trial", but we can talk about "time the occurrence")

Exponential random number generator

Exercise: use the algoritm of *inverting the cumulative* to write an *exponential random number generator*

- reminder:

$$
\begin{aligned}
f(t) &= \frac{1}{\tau}\, e^{-t/\tau} \\
&= r\, e^{-r\,t}
\end{aligned}
$$

$$
\begin{aligned}
F(t) &= 1 - e^{-t/\tau} \\
&= 1 - e^{-r\,t}
\end{aligned}
$$

Property of “no memory”

Both the Exponential and the Geometric have the property of ‘no memory’:

If the ’success’ is not occurred up to a certain ‘point’ (trial or instant), all probabilistic considerations restart from that that ‘point’.

$$P(X > x + x_\circ \,|\, X > x_\circ) \;\;=\;\; P(X > x)$$

Property of “no memory”

Both the Exponential and the Geometric have the property of ‘no memory’:

If the ’success’ is not occurred up to a certain ‘point’ (trial or instant), all probabilistic considerations restart from that that ‘point’.

$$P(X > x + x_\circ \,|\, X > x_\circ) \;\; = \;\; P(X > x)$$

If we imagine precesses that go on also after each success (imagine a coin repeatedly tossed, or a Poisson process whose intensity remains constant over a long time)

Property of "no memory"

Both the Exponential and the Geometric have the property of 'no memory':

If the 'success' is not occurred up to a certain 'point' (trial or instant), all probabilistic considerations restart from that that 'point'.

$$P(X > x + x_\circ \,|\, X > x_\circ) \;=\; P(X > x)$$

If we imagine precesses that go on also after each success (imagine a coin repeatedly tossed, or a Poisson process whose intensity remains constant over a long time)

- the 'zero' of the counting (or of the time) can be taken when a 'success' has occurred;

Property of "no memory"

Both the Exponential and the Geometric have the property of 'no memory':

If the 'success' is not occurred up to a certain 'point' (trial or instant), all probabilistic considerations restart from that that 'point'.

$$P(X > x + x_\circ \,|\, X > x_\circ) \;\; = \;\; P(X > x)$$

If we imagine precesses that go on also after each success (imagine a coin repeatedly tossed, or a Poisson process whose intensity remains constant over a long time)

- the 'zero' of the counting (or of the time) can be taken when a 'success' has occurred; hence
- the Geometric distribution describes the number of trials between consecutive successes (and in this case it is convenient to make X starting from 0, instead than from 1);

Property of "no memory"

Both the Exponential and the Geometric have the property of 'no memory':

If the 'success' is not occurred up to a certain 'point' (trial or instant), all probabilistic considerations restart from that that 'point'.

$$P(X > x + x_\circ \,|\, X > x_\circ) \;=\; P(X > x)$$

If we imagine precesses that go on also after each success (imagine a coin repeatedly tossed, or a Poisson process whose intensity remains constant over a long time)

- the 'zero' of the counting (or of the time) can be taken when a 'success' has occurred; hence
- the Geometric distribution describes the number of trials between consecutive successes (and in this case it is convenient to make X starting from 0, instead than from 1);
- the Exponential distribution describes the time interval between two consecutive events (as long as r remains 'practically' constant).

Decay life time and half time

(An interesting exercise)

Imagine we have, at $t = 0$, $N(0) = N_0$ nuclei.

- Probability that one nucleus decays in the time interval between 0 and ΔT:

Decay life time and half time

(An interesting exercise)

Imagine we have, at $t = 0$, $N(0) = N_0$ nuclei.

- Probability that one nucleus decays in the time interval between 0 and ΔT:

$$
\begin{aligned}
P(\text{“1 decay in } \Delta t\text{”}) &= \int_0^{\Delta t} \frac{1}{\tau} e^{-t/\tau}\, dt \\
&= 1 - e^{-\Delta t/\tau} \\
&\qquad\qquad (\text{If } \Delta t \ll \tau) \\
&\approx 1 - \left(1 - \frac{\Delta t}{\tau}\right) = \frac{\Delta t}{\tau}
\end{aligned}
$$

Decay life time and half time

(An interesting exercise)

Imagine we have, at $t = 0$, $N(0) = N_0$ nuclei.

- Probability that one nucleus decays in the time interval between 0 and ΔT:

$$\begin{aligned} P(\text{``1 decay in } \Delta t\text{''}) &= \int_0^{\Delta t} \frac{1}{\tau} e^{-t/\tau}\, dt \\ &= 1 - e^{-\Delta t/\tau} \\ &\qquad (\text{If } \Delta t \ll \tau) \\ &\approx 1 - \left(1 - \frac{\Delta t}{\tau}\right) = \frac{\Delta t}{\tau} \\ \Rightarrow P(\text{``1 d. in } \Delta t\text{''}) &= \frac{\Delta t}{\tau}. \end{aligned}$$

Decay life time and half time

(An interesting exercise)

Imagine we have, at $t = 0$, $N(0) = N_0$ nuclei.

- Probability that one nucleus decays in the time interval between 0 and ΔT:

$$
\begin{aligned}
P(\text{“1 decay in } \Delta t\text{”}) &= \int_0^{\Delta t} \frac{1}{\tau} e^{-t/\tau}\, dt \\
&= 1 - e^{-\Delta t/\tau} \\
&\qquad \text{(If } \Delta t \ll \tau\text{)} \\
&\approx 1 - \left(1 - \frac{\Delta t}{\tau}\right) = \frac{\Delta t}{\tau} \\
\Rightarrow P(\text{“1 d. in } \Delta t\text{”}) &= \frac{\Delta t}{\tau}.
\end{aligned}
$$

- If we have at a given instant N nuclei, how many will decay in Δt?

$\longrightarrow$

Decay life time and half time

(An interesting exercise – cont.d 1)

$$
\begin{aligned}
X &\sim \mathcal{B}(N, \frac{\Delta t}{\tau}) \\
\mathrm{E}[X] &= N \cdot \frac{\Delta t}{\tau} \;\; [= N \cdot r \cdot \Delta t] \\
\sigma(X) &= \sqrt{\frac{\Delta t}{\tau} \cdot \left(1 - \frac{\Delta t}{\tau}\right) \cdot N} \approx \sqrt{\frac{\Delta t}{\tau} \cdot N}
\end{aligned}
$$

Decay life time and half time

(An interesting exercise – cont.d 1)

$$\begin{aligned} X &\sim \mathcal{B}(N, \frac{\Delta t}{\tau}) \\ \mathrm{E}[X] &= N \cdot \frac{\Delta t}{\tau} \;\; [= N \cdot r \cdot \Delta t] \\ \sigma(X) &= \sqrt{\frac{\Delta t}{\tau} \cdot \left(1 - \frac{\Delta t}{\tau}\right) \cdot N} \approx \sqrt{\frac{\Delta t}{\tau} \cdot N} = \sqrt{\mathrm{E}[X]} \end{aligned}$$

- Relative uncertainty ($\mathrm{E}[X] > 0$):

$$v = \frac{\sigma(X)}{\mathrm{E}[X]} \;\; = \;\; \frac{1}{\sqrt{\mathrm{E}[X]}}$$

Decay life time and half time

(An interesting exercise – cont.d 1)

$$\begin{aligned} X &\sim \mathcal{B}(N, \frac{\Delta t}{\tau}) \\ \mathrm{E}[X] &= N \cdot \frac{\Delta t}{\tau} \quad [= N \cdot r \cdot \Delta t] \\ \sigma(X) &= \sqrt{\frac{\Delta t}{\tau} \cdot \left(1 - \frac{\Delta t}{\tau}\right) \cdot N} \approx \sqrt{\frac{\Delta t}{\tau} \cdot N} = \sqrt{\mathrm{E}[X]} \end{aligned}$$

- Relative uncertainty ($\mathrm{E}[X] > 0$):

$$v = \frac{\sigma(X)}{\mathrm{E}[X]} = \frac{1}{\sqrt{\mathrm{E}[X]}}$$

- When the number of decays in Δt is rather 'large', that is
  - N 'large'
  - Δt 'not too small'

$v \rightarrow 0$

Decay life time and half time

(An interesting exercise – cont.d 1)

$$
\begin{aligned}
X &\sim \mathcal{B}(N, \frac{\Delta t}{\tau}) \\
\mathrm{E}[X] &= N \cdot \frac{\Delta t}{\tau} \;\; [= N \cdot r \cdot \Delta t] \\
\sigma(X) &= \sqrt{\frac{\Delta t}{\tau} \cdot \left(1 - \frac{\Delta t}{\tau}\right) \cdot N} \approx \sqrt{\frac{\Delta t}{\tau} \cdot N} = \sqrt{\mathrm{E}[X]}
\end{aligned}
$$

- Relative uncertainty ($\mathrm{E}[X] > 0$):

$$v = \frac{\sigma(X)}{\mathrm{E}[X]} \quad = \quad \frac{1}{\sqrt{\mathrm{E}[X]}}$$

- When the number of decays in Δt is rather 'large', that is
  - N 'large'
  - Δt 'not too small'

  $v \to 0$
- The process can be seen as deterministic: $\longrightarrow$

Decay life time and half time

(An interesting exercise – cont.d 2)

$$\Delta N = -\mathrm{E}[X] \;\; = \;\; -N\,\frac{\Delta t}{\tau}$$

Decay life time and half time

(An interesting exercise – cont.d 2)

$$\Delta N = -\mathrm{E}[X] \;\; = \;\; -N\,\frac{\Delta t}{\tau}$$

$$\frac{\Delta N}{\Delta t} \;\; = \;\; -\frac{N}{\tau}$$

Decay life time and half time

(An interesting exercise – cont.d 2)

$$\Delta N = -\mathrm{E}[X] \quad = \quad -N\,\frac{\Delta t}{\tau}$$

$$\frac{\Delta N}{\Delta t} \quad = \quad -\frac{N}{\tau}$$

that we can 'conveniently extend' to the continuum as

$$\frac{dN}{dt} \quad = \quad -\frac{N}{\tau}\,,$$

resulting in

$$N(t) \quad = \quad N_0 \cdot e^{-t/\tau}\,.$$

Decay life time and half time

(An interesting exercise – cont.d 3)

Twofold meaning of τ:

Decay life time and half time

(An interesting exercise – cont.d 3)

Twofold meaning of τ:

- expected time to disintegrate for each nucleus;

Decay life time and half time

(An interesting exercise – cont.d 3)

Twofold meaning of τ:

- expected time to disintegrate for each nucleus;
- time constant of the decreasing numbers of nuclei, seen as a continuous process ($\sim$ deterministic law, like capacitor discharge in a RC circuit).

Decay life time and half time

(An interesting exercise – cont.d 3)

Twofold meaning of τ:

- expected time to disintegrate for each nucleus;
- time constant of the decreasing numbers of nuclei, seen as a continuous process ($\sim$ deterministic law, like capacitor discharge in a RC circuit).
- Half live:

$$N(t_{1/2}) = \frac{N_0}{2} \quad \rightarrow \quad t_{1/2} = \tau \ln 2$$

Decay life time and half time

(An interesting exercise – cont.d 3)

Twofold meaning of τ:

- expected time to disintegrate for each nucleus;
- time constant of the decreasing numbers of nuclei, seen as a continuous process ($\sim$ deterministic law, like capacitor discharge in a RC circuit).
- Half live:

$$N(t_{1/2}) = \frac{N_0}{2} \quad \rightarrow \quad t_{1/2} = \tau \ln 2$$

- For a single nucleus $t_{1/2}$ is the median of the p.d.f. of the time to decay.

Decay life time and half time

(An interesting exercise – cont.d 3)

Twofold meaning of τ:

- expected time to disintegrate for each nucleus;
- time constant of the decreasing numbers of nuclei, seen as a continuous process ($\sim$ deterministic law, like capacitor discharge in a RC circuit).
- Half live:

$$N(t_{1/2}) = \frac{N_0}{2} \quad \rightarrow \quad t_{1/2} = \tau \ln 2$$

- For a single nucleus $t_{1/2}$ is the median of the p.d.f. of the time to decay.
- Experimentally
  - $\{\Delta N,\, \Delta t,\, N\} \;\rightarrow r \;\rightarrow\; \tau$

  (without having to observe the instant of 'birth' and of dead of single objects)

Decay life time and half time

(An interesting exercise – cont.d 3)

Twofold meaning of τ:

- expected time to disintegrate for each nucleus;
- time constant of the decreasing numbers of nuclei, seen as a continuous process ($\sim$ deterministic law, like capacitor discharge in a RC circuit).
- Half live:

$$N(t_{1/2}) = \frac{N_0}{2} \quad \rightarrow \quad t_{1/2} = \tau \ln 2$$

- For a single nucleus $t_{1/2}$ is the median of the p.d.f. of the time to decay.
- Experimentally
  - $\{\Delta N,\, \Delta t,\, N\} \;\rightarrow r \;\;\rightarrow\;\; \tau$

  (without having to observe the instant of 'birth' and of dead of single objects)

Quite not an easy concept for the general public $\longrightarrow$

Decay life time and half time

How life times are perceived...

Many years ago there was a claim of a proton decay observed in an underground experiment:

- The 'observed' lifetime was about 10^{25} years (order of magnitude – details are irrelevant);

Decay life time and half time

How life times are perceived. . .

Many years ago there was a claim of a proton decay observed in an underground experiment:

- The ‘observed’ lifetime was about 10^{25} years
(order of magnitude – details are irrelevant);
- This is how the new was reported by a major Italian newspaper (Corriere della Sera):

Decay life time and half time

How life times are perceived...

Many years ago there was a claim of a proton decay observed in an underground experiment:

- The 'observed' lifetime was about 10^{25} years
  (order of magnitude – details are irrelevant);
- This is how the new was reported by a major Italian newspaper (Corriere della Sera):

Observed how a proton dies.
It was 10^{25} years old.

The very venerably aged proton

Corriere della Sera, 1st June 1984

Venerdì 1 giugno 1984 **INTERNO**

L'eccezionale avvenimento scientifico illustrato al convegno di cosmofisica dell'Aquila

«Abbiamo visto come muore un protone»

Gli scienziati italiani spiegano: aveva 10 mila triliardi di anni

La comunicazione è stata fatta dagli studiosi che operano nel laboratorio sotto il Monte Bianco - La particella che hanno visto disintegrarsi il 17 maggio scorso era la seconda al cui «decesso» assistevano in diretta - Questo evento rarissimo ha un'enorme importanza per le moderne teorie unificate della fisica

DAL NOSTRO INVIATO SPECIALE

L'AQUILA — Il protone dà segni di stanchezza. I cosmofisici che lavorano nel laboratorio sotto il monte Bianco ne hanno visto disintegrarsi uno il 17 maggio scorso. Aveva un'età venerabilissima: circa diecimila triliardi di anni. E' il secondo protone la cui morte viene colta in diretta dagli scienziati italiani da quando è cominciato l'esperimento NUSEX, dalle iniziali di «Nucleon stability experiment».

La comunicazione di questo importante risultato scientifico è stata data l'altro ieri, con la prudenza necessaria, al secondo «Convegno nazionale di fisica cosmica», che si tiene all'Aquila, dal fisico sperimentale Pio Picchi.

NUSEX è una collaborazione fra l'Istituto di cosmofisica del CNR di Torino e i laboratori INFN di Milano e di Frascati

per le moderne teorie unificate della fisica, ha accresciuto l'interesse per un altro e più grande laboratorio sotterraneo dove si studieranno questi fenomeni: il laboratorio sotto il Gran Sasso.

Durante un intermezzo dei lavori congressuali, i cento fisici riuniti a L'Aquila hanno visitato le gallerie dove, entro due anni, sorgeranno i nuovi laboratori dell'INFN. Il parlamento ha da poco approvato il secondo finanziamento di sessanta miliardi per il completamento delle opere pubbliche, e ora gli studiosi aspettano che venga dato il via all'acquisto della strumentazione scientifica.

«Sotto il Gran Sasso inizieremo, molto probabilmente, con tre esperimenti — spiega il professor Renato Scrimaglio, direttore del costruendo laboratorio —: 1) lo studio del decadimento del protone; 2) lo studio dei neutrini solari; 3) la ricerca di una particella chiamata monopolo magnetico. Si tratta di ricerche di frontiera aperte alla collaborazione dei fisici di tutto il mondo».

I programmi dettagliati degli esperimenti saranno definiti al più presto da una commissione presieduta dal fisico Antonino Zichichi, che è stato anche il proponente del laboratorio sotto il Gran Sasso. La commissione si riunirà per la prima volta il 4 giugno prossimo. I finanziamenti degli esperimenti scientifici saranno assicurati nell'ambito del bilancio 1985-1990 dell'INFN di cui si aspetta l'approvazione. Esso prevede una spesa complessiva di mille miliardi dei quali un centinaio da destinare alla strumentazione del laboratorio sotto il Gran Sasso.

Sotto 1.400 metri di roccia, ad ammirare la grande camera alta venti metri in cui si studieranno le particelle, si è avventurato anche il professor Bruno Rossi, 81 anni, uno dei padri fondatori della fisica cosmica. «E' un paesaggio dantesco — ha esclamato —. Se quando ero giovane ricercatore mi avessero detto che per studiare la radiazione cosmica sarebbe stato necessario ficcarsi qui dentro non ci avrei creduto».

— E se potesse ricominciare da capo — gli abbiamo chiesto — verrebbe a lavorarci, qui sotto?

«Ah no, certamente no», ha risposto il professore scuotendo la testa.

Franco Foresta Martin

In una conferenza mondiale a Roma si proporranno nuove strategie per le attività ittiche

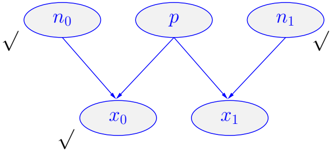
La FAO punta al raddoppio della pesca

Back to the

inferential/predictive problem

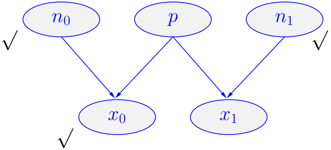
related to the binomial model

Joint inference and prediction



Let's do the math.

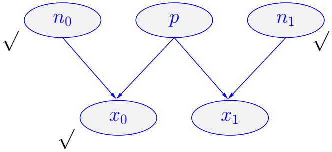
Joint inference and prediction



Let's do the math.

- Three **observed variables**

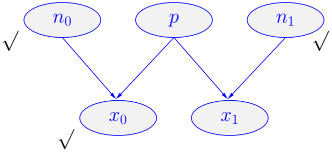
Joint inference and prediction



Let's do the math.

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

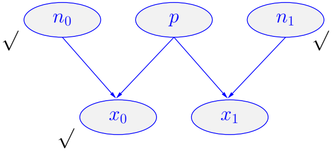
Joint inference and prediction



Let's do the math.

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

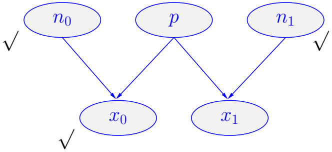
Joint inference and prediction



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.

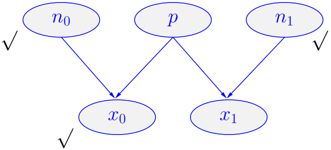
Joint inference and prediction



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

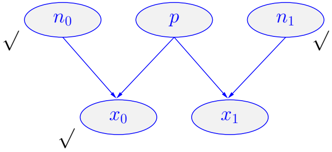


Let's do the math.

- 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.

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

Joint inference and prediction



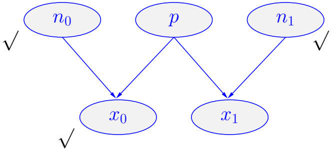
Let's do the math.

- 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.

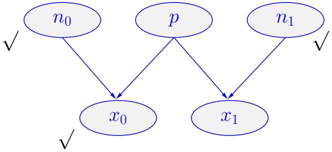
$$f(p, x_1 \,|\, n_0, x_0, n_1) = \frac{f(p, x1, n_0, n_1, x_0)}{f(n_0, x_0, n_1)}$$

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

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

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

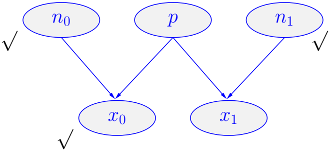
Joint inference and prediction



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

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

Joint inference and prediction

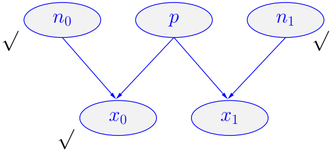


Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

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

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

Joint inference and prediction



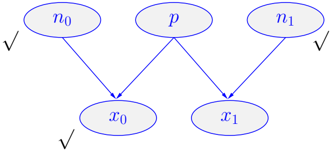
Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

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

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

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

Joint inference and prediction



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

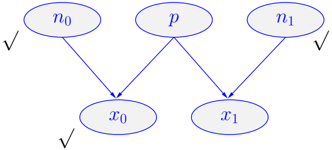
$$f(p, x_1 \,|\, n_0, x_0, n_1) \propto f(x_0 \,|\, n_0, p) \cdot f(x_1 \,|\, p, n_1) \cdot f_0(p)$$

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

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

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):

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

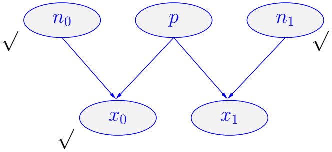
$$\propto p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{x_1!\,(n_1-x_1)!}\, f_0(p)$$

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

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):

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

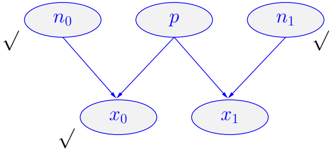
$$\propto\; p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{x_1!\,(n_1-x_1)!}\, f_0(p)$$

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

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):

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

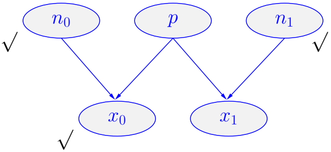
$$\propto\; p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{x_1!\,(n_1-x_1)!}\, f_0(p)$$

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

Problem almost solved

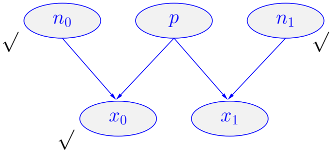
- Possibly calculate the normalization, then all moments and probability intervals of interest.
- Do it numerically or by by sampling.

Joint inference and prediction



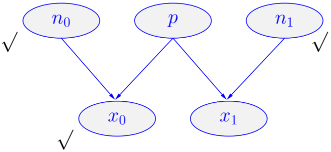
$\Rightarrow$ sample $\tilde{f}(p, x_1 \mid n_0, x_0, n_1)$

Joint inference and prediction



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

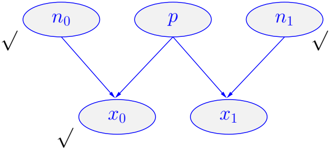
Joint inference and prediction



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

$\Rightarrow$ Markov Chain Monte Carlo (MCMC)

Joint inference and prediction

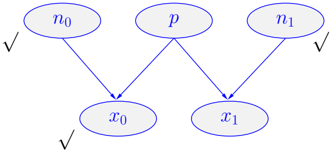


$\Rightarrow$ sample $\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1)$ using Monte Carlo techniques

$\Rightarrow$ Markov Chain Monte Carlo (MCMC)

$\Rightarrow$ JAGS does it for us

Joint inference and prediction



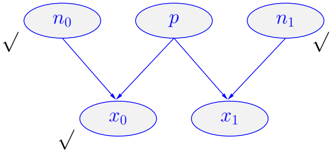
$\Rightarrow$ sample $\tilde{f}(p, x_1 \mid n_0, x_0, n_1)$ using Monte Carlo techniques

$\Rightarrow$ **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;

Joint inference and prediction



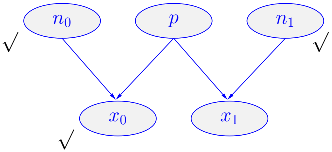
$\Rightarrow$ sample $\tilde{f}(p, x_1 \mid n_0, x_0, n_1)$ using Monte Carlo techniques

$\Rightarrow$ **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

Joint inference and prediction



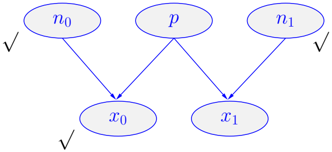
$\Rightarrow$ sample $\tilde{f}(p, x_1 \,|\, n_0, x_0, n_1)$ using Monte Carlo techniques

$\Rightarrow$ **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)

Joint inference and prediction



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

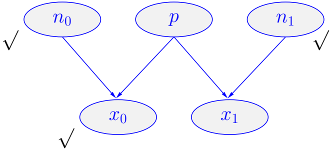
$\Rightarrow$ **Markov Chain Monte Carlo (MCMC)**

$\Rightarrow$ 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)

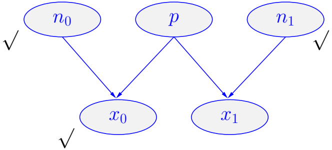
JAGS called from R using the package rjags.

Graphical models: some terminology



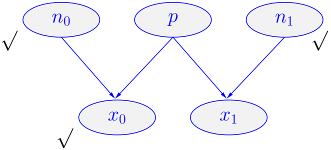
- nodes (observed/unobserved);
- child/childred;
- parent(s).

Graphical models: some terminology



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

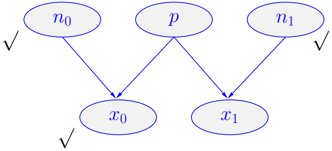
Joint inference and prediction in JAGS



Model

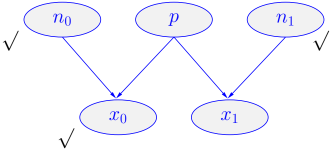
```
model{
    x0 ~ dbin(p, n0);
    x1 ~ dbin(p, n1);
     p ~ dbeta(1, 1);
}
```

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)
```

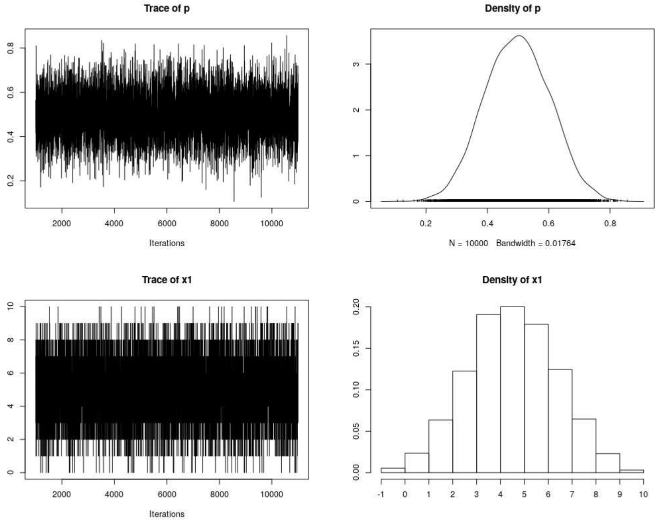
Use of JAGS from R via rjags

Second part of the R script (⇒ inf_p_pred_jags.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))
```

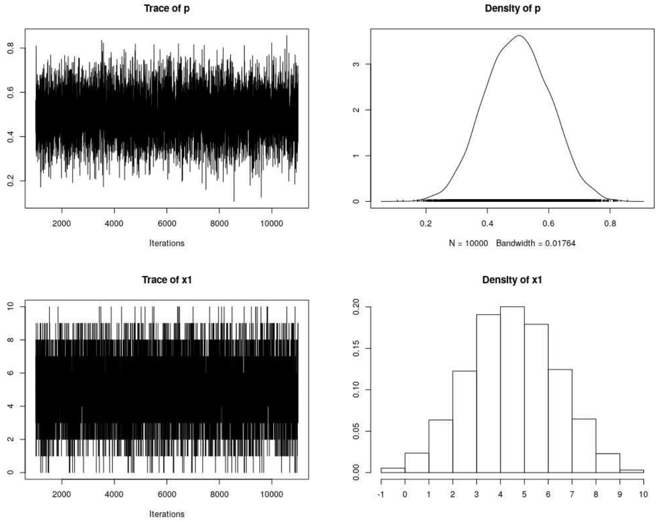
Use of JAGS from R via rjags

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



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
$\mathrm{E}(X_1)$		5	5	5	5
$\sigma[X_1]$		2.64	1.87	1.62	1.58

Inference and prediction with JAGS/rjags

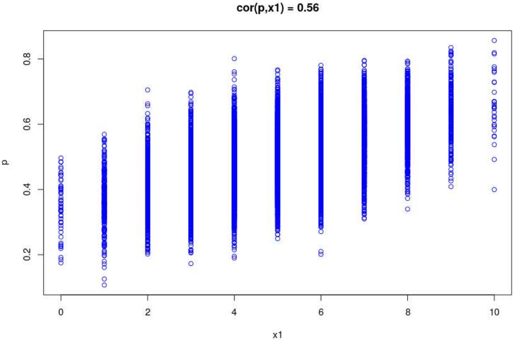
Scatter plot of sampled $f(p, x_1 \mid n_0, x_0, n_1)$

```
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 $f(p, x_1 \mid n_0, x_0, n_1)$

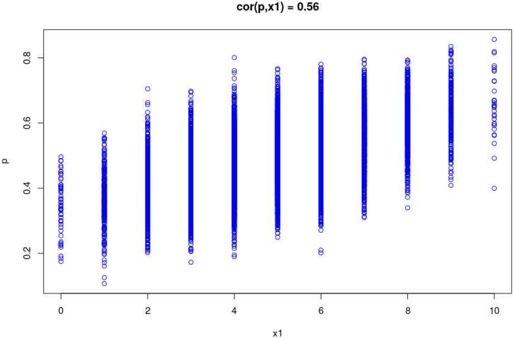
```
p  <- as.vector(chain[[1]][,1])
x1 <- as.vector(chain[[1]][,2])
plot(x1, p, col='blue',
     main=sprintf("cor(p,x1) = %.2f", cor(p,x1)))
print( table(x1)/10000 )
```



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 $\longrightarrow$ **try it**)

Self made Gibbs sampler

Take, as in the JAGS example, a uniform $f_0(p) \;\rightarrow\; f_0(p) = 1$:

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

Self made Gibbs sampler

Take, as in the JAGS example, a uniform $f_0(p) \rightarrow f_0(p) = 1$:

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

In the **first step** ($i = 0$)

- extract $p^{(i=0)}$ at random: `p = runif(1, 0, 1)`;

Self made Gibbs sampler

Take, as in the JAGS example, a uniform $f_0(p) \rightarrow f_0(p) = 1$:

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

In the **first step** ($i = 0$)

- extract $p^{(i=0)}$ at random: `p = runif(1, 0, 1);`
  (or fix a value by hand)

- extract $x_1^{(i=0)}$ conditioned also by $p = p^{(i=0)}$:

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

Self made Gibbs sampler

Take, as in the JAGS example, a uniform $f_0(p)$ $\rightarrow$ $f_0(p) = 1$:

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

In the **first step** ($i = 0$)

- extract $p^{(i=0)}$ at random: `p = runif(1, 0, 1);`
  (or fix a value by hand)

- extract $x_1^{(i=0)}$ conditioned also by $p = p^{(i=0)}$:

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

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

Self made Gibbs sampler

Take, as in the JAGS example, a uniform $f_0(p) \;\rightarrow\; f_0(p) = 1$:

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

In the **first step** ($i=0$)

- extract $p^{(i=0)}$ at random: `p = runif(1, 0, 1);`
  (or fix a value by hand)

- extract $x_1^{(i=0)}$ conditioned <u>also</u> by $p = p^{(i=0)}$:

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

$\rightarrow$ `x1 = rbinom(1, n1, p)`

Self made Gibbs sampler (cont.d)

First step:

→ `p = runif(1, 0, 1)`
(or fix a value)

→ `x1 = rbinom(1, n1, p)`

Self made Gibbs sampler (cont.d)

First step:

→ `p = runif(1, 0, 1)`
(or fix a value)

→ `x1 = rbinom(1, n1, p)`

Then

1. $i = i + 1$; extract $p^{(i)}$ conditioned also by $x_1^{(i-1)}$:

$$
\begin{aligned}
f(p \mid n_0, x_0, n_1, x_1) &\propto p^{x_0}(1-p)^{n_0-x_0} \cdot p^{x_1}(1-p)^{n_1-x_1} \\
&\propto p^{x_0+x_1} \cdot (1-p)^{n_0+n_1-x_0-x_1}
\end{aligned}
$$

Self made Gibbs sampler (cont.d)

First step:

→ `p = runif(1, 0, 1)`
(or fix a value)

→ `x1 = rbinom(1, n1, p)`

Then

1. $i = i + 1$; extract $p^{(i)}$ conditioned also by $x_1^{(i-1)}$:

$$f(p \,|\, n_0, x_0, n_1, x_1) \;\propto\; p^{x_0}(1-p)^{n_0-x_0} \cdot p^{x_1}(1-p)^{n_1-x_1}$$
$$\propto\; p^{x_0+x_1} \cdot (1-p)^{n_0+n_1-x_0-x_1} \qquad [\rightarrow \text{Beta()}]$$

Self made Gibbs sampler (cont.d)

First step:

→ `p = runif(1, 0, 1)`
(or fix a value)

→ `x1 = rbinom(1, n1, p)`

Then

1. $i = i + 1$; extract $p^{(i)}$ conditioned also by $x_1^{(i-1)}$:

$$
\begin{aligned}
f(p \,|\, n_0, x_0, n_1, x_1) &\propto p^{x_0}(1-p)^{n_0-x_0} \cdot p^{x_1}(1-p)^{n_1-x_1} \\
&\propto p^{x_0+x_1} \cdot (1-p)^{n_0+n_1-x_0-x_1} \qquad [\to \text{Beta()}]
\end{aligned}
$$

2. Extract $x_1^{(i)}$, conditioned by $p^{(i)}$, as in Step 1

Self made Gibbs sampler (cont.d)

First step:

→ `p = runif(1, 0, 1)`
(or fix a value)

→ `x1 = rbinom(1, n1, p)`

Then

1. $i = i + 1$; extract $p^{(i)}$ conditioned also by $x_1^{(i-1)}$:

$$f(p \mid n_0, x_0, n_1, x_1) \propto p^{x_0}(1-p)^{n_0-x_0} \cdot p^{x_1}(1-p)^{n_1-x_1}$$
$$\propto p^{x_0+x_1} \cdot (1-p)^{n_0+n_1-x_0-x_1} \quad [\rightarrow \text{Beta}()]$$

2. Extract $x_1^{(i)}$, conditioned by $p^{(i)}$, as in Step 1;
⇒ then `go to 1.`

Self made Gibbs sampler (cont.d)

Summarizing, for $i \geq 1$

Self made Gibbs sampler (cont.d)

Summarizing, for $i \geq 1$

1. $i = i + 1$

   then extract $p^{(i)}$:

Self made Gibbs sampler (cont.d)

Summarizing, for $i \geq 1$

1. $i = i + 1$
   then extract $p^{(i)}$:

$$p^{(i)} \sim \text{Beta}(r = x_0 + x_1 + 1,\, s = n_0 + n_1 - x_0 - x_1 + 1)$$

Self made Gibbs sampler (cont.d)

Summarizing, for $i \geq 1$

1. $i = i + 1$

   then extract $p^{(i)}$:

$$p^{(i)} \sim \text{Beta}(r = x_0 + x_1 + 1,\, s = n_0 + n_1 - x_0 - x_1 + 1)$$

2. extract $x_1^{(i)}$:

$$x_1^{(i)} \sim \mathcal{B}_{n_1,p};$$

Self made Gibbs sampler (cont.d)

Summarizing, for $i \geq 1$

1. $i = i + 1$
   then extract $p^{(i)}$:
$$p^{(i)} \sim \text{Beta}(r = x_0 + x_1 + 1,\, s = n_0 + n_1 - x_0 - x_1 + 1)$$

2. extract $x_1^{(i)}$:
$$x_1^{(i)} \sim \mathcal{B}_{n_1,p};$$

Then do loop through steps 1 and 2 'many times'.

Self made Gibbs sampler (cont.d)

Summarizing, for $i \geq 1$

1. $i = i + 1$

   then extract $p^{(i)}$:

$$p^{(i)} \sim \text{Beta}(r = x_0 + x_1 + 1, \, s = n_0 + n_1 - x_0 - x_1 + 1)$$

2. extract $x_1^{(i)}$:

$$x_1^{(i)} \sim \mathcal{B}_{n_1,p};$$

Then do loop through steps 1 and 2 'many times'.

At the end, we have the histories of the variables of interest

Self made Gibbs sampler (cont.d)

Summarizing, for $i \geq 1$

1. $i = i + 1$
   then extract $p^{(i)}$:
   $$p^{(i)} \sim \text{Beta}(r = x_0 + x_1 + 1,\, s = n_0 + n_1 - x_0 - x_1 + 1)$$
2. extract $x_1^{(i)}$:
   $$x_1^{(i)} \sim \mathcal{B}_{n_1,p};$$

Then do loop through steps 1 and 2 'many times'.

At the end, we have the histories of the variables of interest

- the points $(p^{(i)}, x_1^{(i)})$ will 'visit' the (p, x_1) space according to $f(p, x_1)$.

Self made Gibbs sampler (cont.d)

Summarizing, for $i \geq 1$

1. $i = i + 1$

   then extract $p^{(i)}$:

$$p^{(i)} \sim \text{Beta}(r = x_0 + x_1 + 1,\, s = n_0 + n_1 - x_0 - x_1 + 1)$$

2. extract $x_1^{(i)}$:

$$x_1^{(i)} \sim \mathcal{B}_{n_1,p};$$

   Then do loop through steps 1 and 2 'many times'.

At the end, we have the histories of the variables of interest

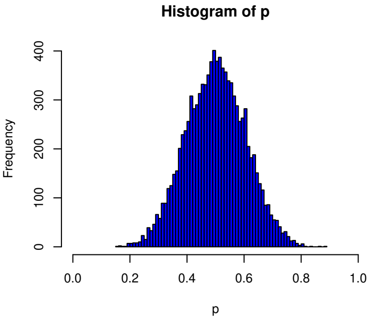
- the points $(p^{(i)}, x_1^{(i)})$ will 'visit' the (p, x_1) space according to $f(p, x_1)$.
  (There are theorems...)

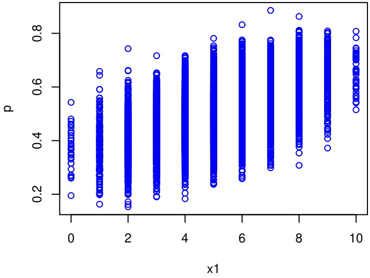
Gibbs sampler, implemented in R

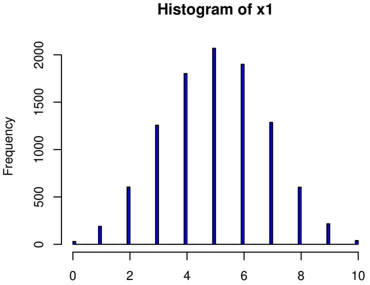
```
# initialize the history vectors
N = 10000; p = x1 = rep(0, N)

# observed nodes
n0 = 20; x0 = 10; n1 = 10

# initial p and n1
p[1] = rbeta(1,1,1)             # uniform
x1[1] = rbinom(1, n1, p[1])
# cat(sprintf("p = %f,  x1 = %d\n", p[1], x1[1]))

for (i in 2:N) {
   p[i] = rbeta(1, x0+x1[i-1]+1, n0+n1-x0-x1[i-1]+1)
   x1[i] = rbinom(1, n1, p[i])
  # cat(sprintf("p = %f,  x1 = %d\n", p[i], x1[i]))
}
```

⟶ inf_p_pred_gibbs.R

Inference/prediction by a Gibbs sampler implemented in R







mean(p) = 0.501
sigma(p) = 0.104

mean(x1) = 5.031
sigma(x1) = 1.817

rho(p,x1) = 0.548

Same problem solved using hit/miss

```
# model parameters, N and uf()
# as in the previous script

# find the maximum
p  <- seq(0,1,len=101)
x1 <- 0:n1
f.max = 0
for (i in 1:length(p)) {
    for (j in 1:length(x1)) {
      f =  uf(p[i], x1[j], n0, x0, n1)
    if( f > f.max ) f.max = f
  }
}
f.max <- f.max * 1.1   # exaggerated safety factor
```

Same problem solved using hit/miss (cont.d)

```
# sample in 2D
#1. random choice in the plane (p,x1)
p.r  <- runif(N)
x1.r <- sample(0:n1, N, rep=TRUE)

# 2. Calculate the function in correspondence of the points
f <- numeric(N)
for (i in 1:N) {
  f[i] <- uf(p.r[i], x1.r[i], n0, x0, n1)
}

# 3. accept events
hit <-  runif(N)*f.max <= f  # accepted events ('hit')
p   <- p.r[hit]
x1  <- x1.r[hit]

##   => inf_p_pred_hit-miss.R
```

Same problem solved numerically

As an **exercise**, for completeness, try to solve the problem numerically (no sampling):

Same problem solved numerically

As an **exercise**, for completeness, try to solve the problem numerically (no sampling):

- discretize the problem making a grid in the (p, x_1) plane;

Same problem solved numerically

As an **exercise**, for completeness, try to solve the problem numerically (no sampling):

- discretize the problem making a grid in the (p, x_1) plane;
- calculate the unnormalized pdf in each elements of the grid;

Same problem solved numerically

As an **exercise**, for completeness, try to solve the problem numerically (no sampling):

- discretize the problem making a grid in the (p, x_1) plane;
- calculate the unnormalized pdf in each elements of the grid;
- calculate the normalization constant;

Same problem solved numerically

As an **exercise**, for completeness, try to solve the problem numerically (no sampling):

- discretize the problem making a grid in the (p, x_1) plane;
- calculate the unnormalized pdf in each elements of the grid;
- calculate the normalization constant;
- calculate all moments of interest, including correlation the coefficient;

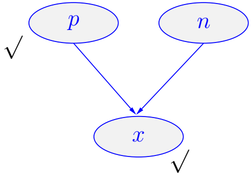
Same problem solved numerically

As an **exercise**, for completeness, try to solve the problem numerically (no sampling):

- discretize the problem making a grid in the (p, x_1) plane;
- calculate the unnormalized pdf in each elements of the grid;
- calculate the normalization constant;
- calculate all moments of interest, including correlation the coefficient;
- (optional) make a lego plot, or a false-color plot to visualize the shape of the distribution.

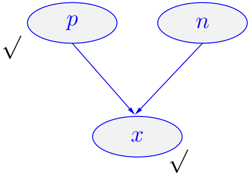
n independent Bernoulli processes

Inferring *n*



n independent Bernoulli processes

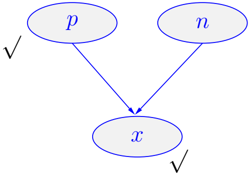
Inferring n



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

n independent Bernoulli processes

Inferring *n*

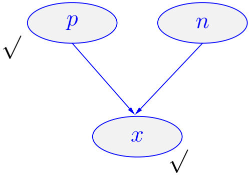


Think at a detector having a *well known efficiency ($\epsilon \equiv p$)*:

- we have recorded x 'signals';

n independent Bernoulli processes

Inferring *n*

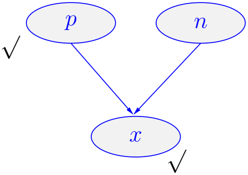


Think at a detector having a *well known efficiency* ($\epsilon \equiv p$):

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

n independent Bernoulli processes

Inferring n

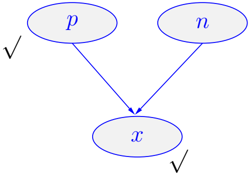


Think at a detector having a *well known efficiency* ($\epsilon \equiv p$):

- 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* ($\epsilon \equiv p$):

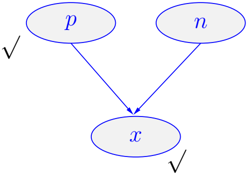
- we have recorded x 'signals';
- how many particles impinged the detector? $\longrightarrow f(n\,|\,x, p)$ **?**

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 ($\epsilon \equiv p$)*:

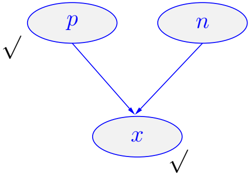
- we have recorded x 'signals';
- how many particles impinged the detector? $\longrightarrow f(n\,|\,x,p)$ **?**

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? $\longrightarrow f(n\,|\,x, p)$ **?**

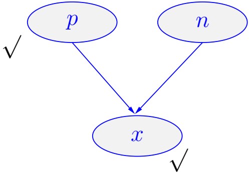
Not to be confused with a different problem:

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

n independent Bernoulli processes

Extending the model

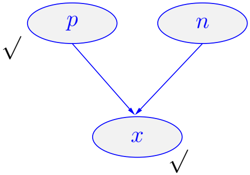
Our problem (but in Physics it is often not so simple)



n independent Bernoulli processes

Extending the model

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

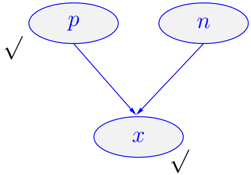


But we need some (usually indirect) knowledge about *p*

n independent Bernoulli processes

Extending the model

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

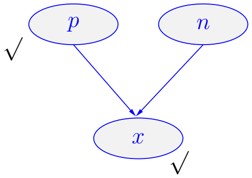


But we need some (usually indirect) knowledge about *p*
(Usually we do not calculate *p* from the fraction of white balls!)

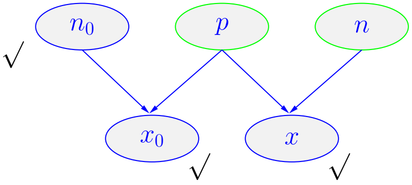
n independent Bernoulli processes

Extending the model

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



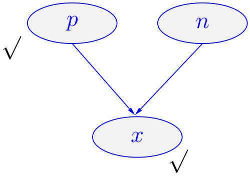
But we need some (usually indirect) knowledge about p
(Usually we do not calculate p from the fraction of white balls!)



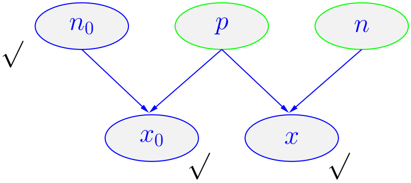
n independent Bernoulli processes

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 we are uncertain on the numbers we ‘see’ in our detector, but to ‘other numbers’**.

Extending the model

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

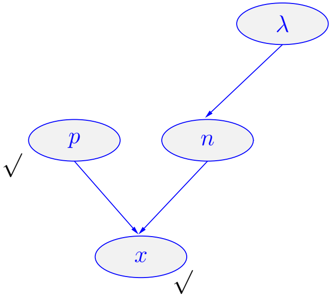
- When we say "we are uncertain on numbers", **we do not mean we are uncertain on the numbers we 'see' in our detector, but to 'other numbers'**.
- Typically $n \longleftrightarrow \lambda$.

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 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':

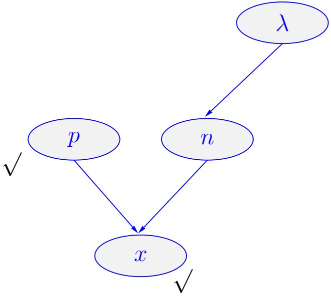


Extending the model

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 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':



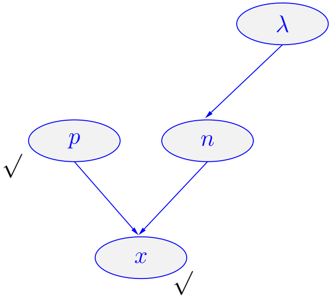
But, as we have seen studying the **Poisson process**,
λ is not really physical

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 we are uncertain on the numbers we 'see' in our detector, but to 'other numbers'**.
- Typically $n \longleftrightarrow \lambda$.

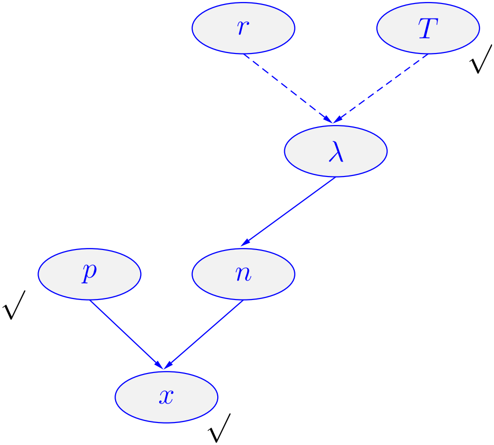
Assuming for a while p well known and focusing on 'n':



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

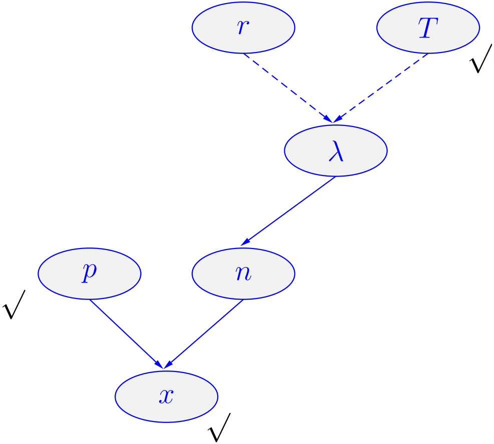
Extending the model

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



Extending the model

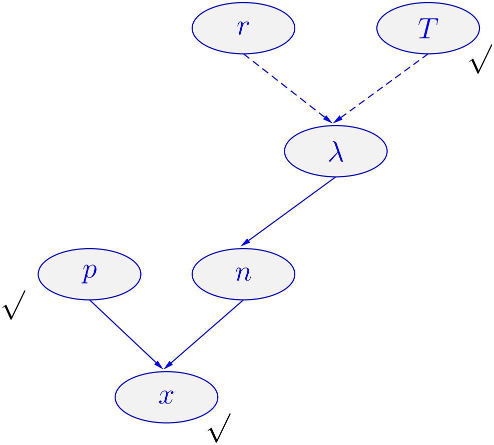
$\boldsymbol{\lambda} = r \cdot T$:



(Dashed arrows used in literature for deterministic links)

Extending the model

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

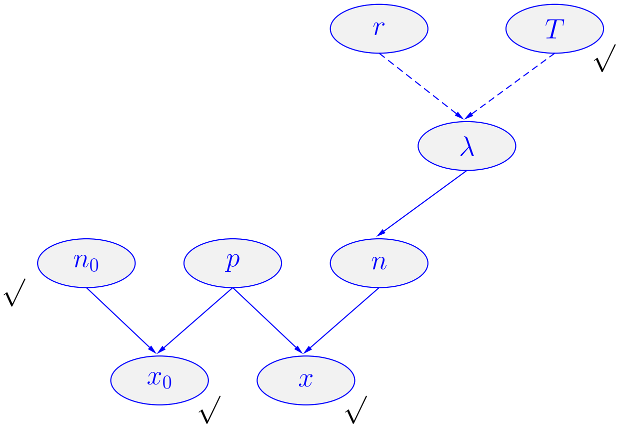


(Dashed arrows used in literature for deterministic links)

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

Extending the model

Remembering that p was got from a measurement:

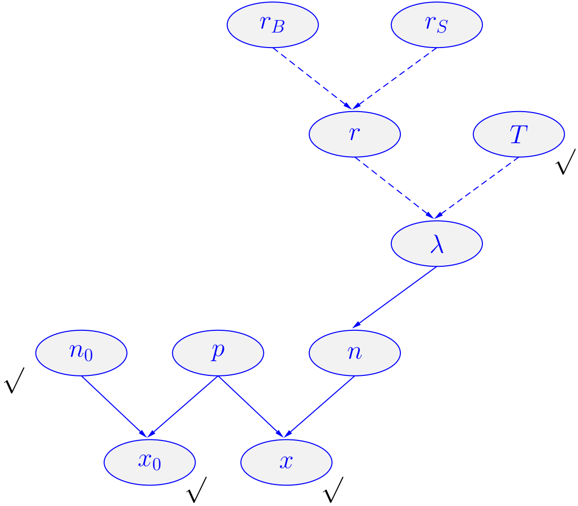


Extending the model

The rate r gets contributions from signal and background

Extending the model

The rate *r* gets contributions from signal and background



Extending the model

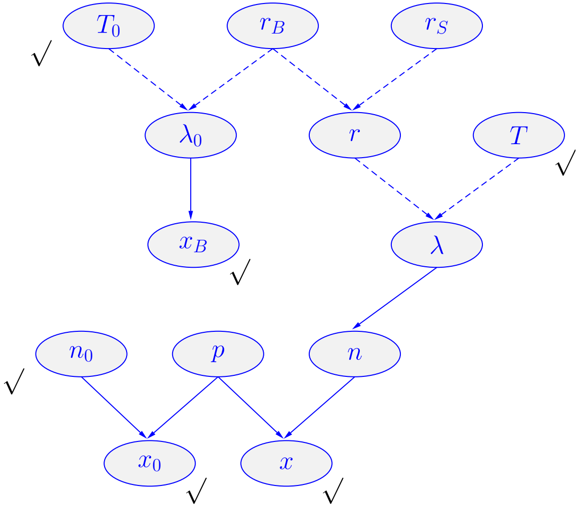
But, since $r = r_S + r_B$,

we need some independent knowledge of the background

Extending the model

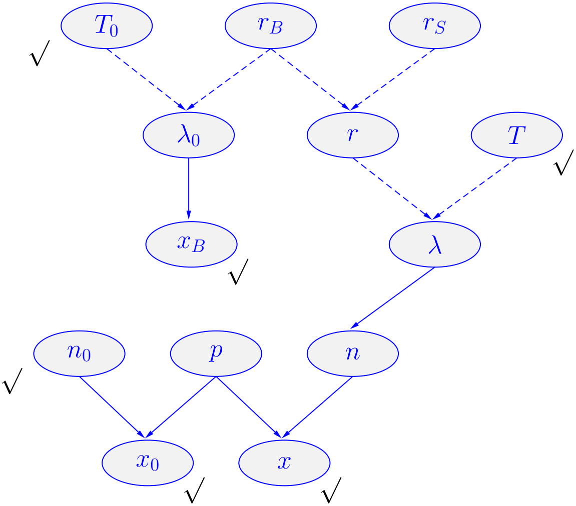
But, since $r = r_S + r_B$,

we need some independent knowledge of the background



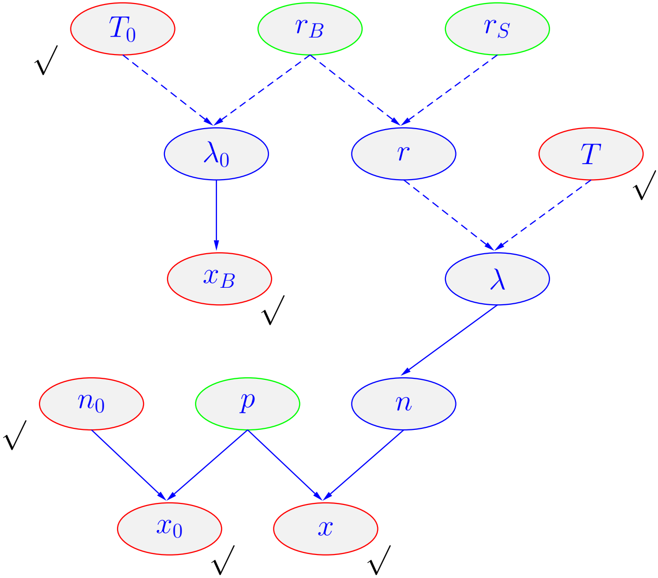
Extending the model

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

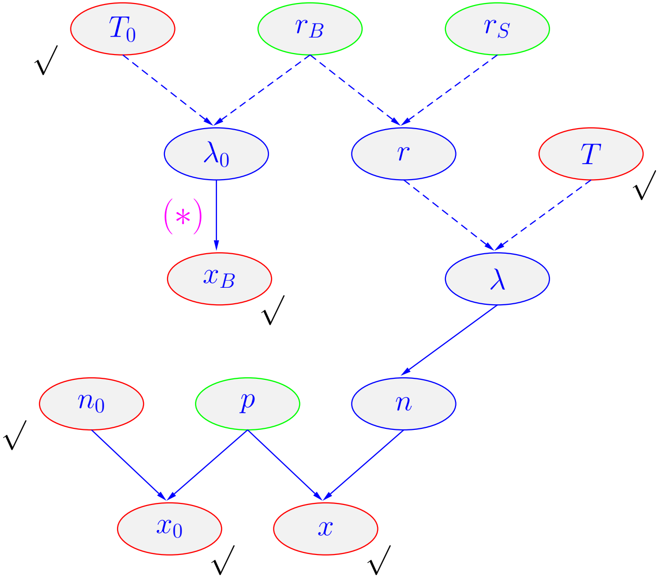


(T_0 and T assumed to be measured with sufficient accuracy)

Extending the model



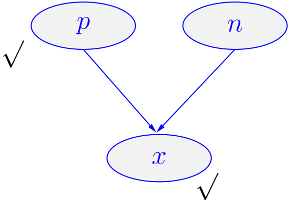
Extending the model



(*) Assuming unity efficiency

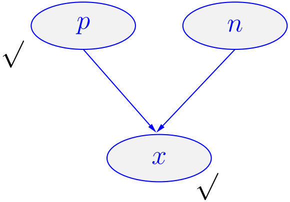
Inferring n ‘assuming’ p and x

Back to our initial problem



Inferring n 'assuming' p and x

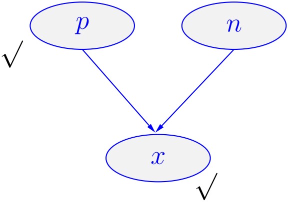
Back to our initial problem



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

Inferring n 'assuming' p and x

Back to our initial problem

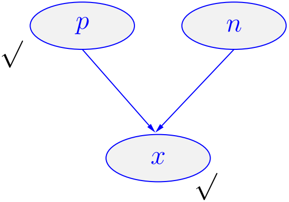


$$f(n\,|\,p,x) \;\propto\; f(x\,|\,n,p)\cdot f_0(n)$$

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

Inferring n 'assuming' p and x

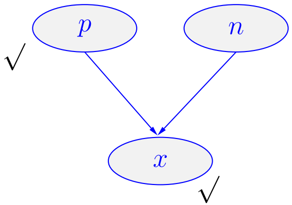
Back to our initial problem



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

Inferring n 'assuming' p and x

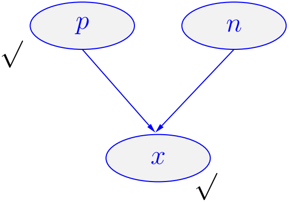
Back to our initial problem



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

Inferring n 'assuming' p and x

Back to our initial problem



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

Exercise

This time the unnormalized probability function

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

is so simple that we can simply solve it numerically.

Exercise

This time the unnormalized probability function

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

is so simple that we can simply solve it numerically.

Using $p = 0.75$ and $x = 10$ and $n_{max} = 30$

- calculate the *vector* of $\tilde{f}(n\,|\,p,x)$ for $x \leq n \leq n_{max}$;

Exercise

This time the unnormalized probability function

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

is so simple that we can simply solve it numerically.

Using $p = 0.75$ and $x = 10$ and $n_{max} = 30$

- calculate the *vector* of $\tilde{f}(n \,|\, p, x)$ for $x \leq n \leq n_{max}$;
- calculate the normalization factor;

Exercise

This time the unnormalized probability function

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

is so simple that we can simply solve it numerically.

Using $p = 0.75$ and $x = 10$ and $n_{max} = 30$

- calculate the *vector* of $\tilde{f}(n \mid p, x)$ for $x \leq n \leq n_{max}$;
- calculate the normalization factor;
- calculate $\mathrm{E}(n)$, $\sigma(n)$;

Exercise

This time the unnormalized probability function

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

is so simple that we can simply solve it numerically.

Using $p = 0.75$ and $x = 10$ and $n_{max} = 30$

- calculate the *vector* of $\tilde{f}(n \mid p, x)$ for $x \leq n \leq n_{max}$;
- calculate the normalization factor;
- calculate $\mathrm{E}(n)$, $\sigma(n)$;
- make a barplot of the distribution;

Exercise

This time the unnormalized probability function

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

is so simple that we can simply solve it numerically.

Using $p = 0.75$ and $x = 10$ and $n_{max} = 30$

- calculate the *vector* of $\tilde{f}(n \,|\, p, x)$ for $x \leq n \leq n_{max}$;
- calculate the normalization factor;
- calculate $\mathrm{E}(n)$, $\sigma(n)$;
- make a barplot of the distribution;
- calculate $P(n \geq 20)$

Inferring n 'assuming' p and x

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

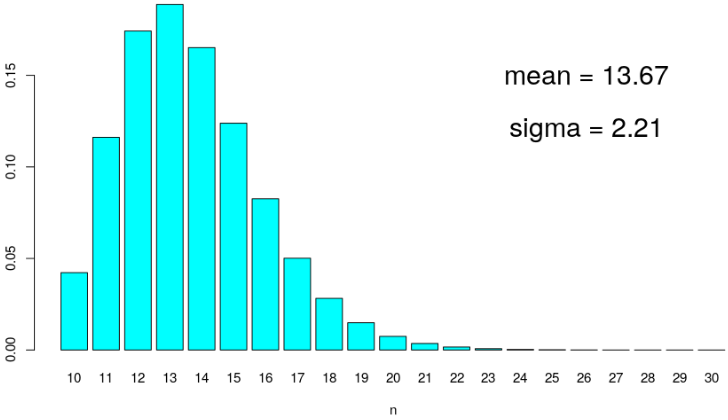
Left as exercise

(1-D version of that proposed at the end of last lecture)

Result $\Rightarrow$

Inferring n 'assuming' p and x

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



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**

Moving to the Poisson model

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

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

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

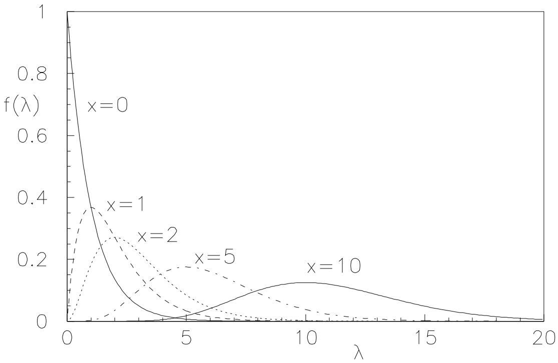
$$f(\lambda\,|\,x,\mathcal{P}) \;=\; \frac{\lambda^x\,e^{-\lambda}}{x!}$$

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

Summaries

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

Some examples of $f(\lambda)$

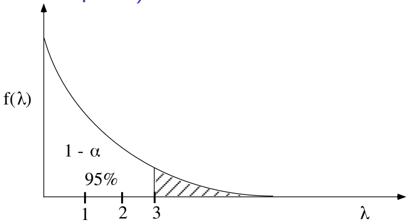


For 'large' x $f(\lambda)$ it becomes Gaussian with expected value x and standard deviation $\sqrt{x}$.

> 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 $x = 0$

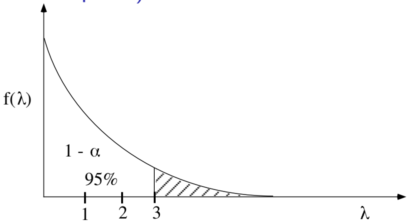
(From a flat prior!)



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

Inferring λ from $x = 0$

(From a flat prior!)



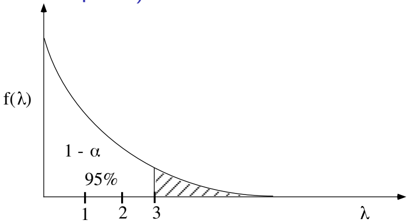
$$f(\lambda \,|\, x = 0, \mathcal{P}) = e^{-\lambda}$$
$$F(\lambda \,|\, x = 0, \mathcal{P}) = 1 - e^{-\lambda}$$

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

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

Inferring λ from $x = 0$

(From a flat prior!)



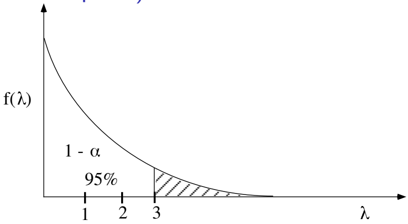
$$
\begin{aligned}
f(\lambda \,|\, x=0, \mathcal{P}) &= e^{-\lambda} \\
F(\lambda \,|\, x=0, \mathcal{P}) &= 1 - e^{-\lambda}
\end{aligned}
$$

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

$$
\begin{aligned}
P(\lambda \leq \lambda_u \,|\, x=0) = F(\lambda_u \,|\, x=0) &= 0.95 \\
1 - F(\lambda_u \,|\, x=0) = e^{-\lambda_u} &= 0.05
\end{aligned}
$$

Inferring λ from $x = 0$

(From a flat prior!)



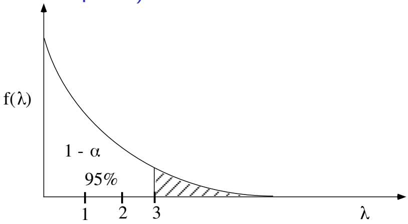
$$
\begin{aligned}
f(\lambda\,|\,x=0,\mathcal{P}) &= e^{-\lambda}\\
F(\lambda\,|\,x=0,\mathcal{P}) &= 1-e^{-\lambda}
\end{aligned}
$$

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

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

Inferring λ from $x = 0$

(From a flat prior!)



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

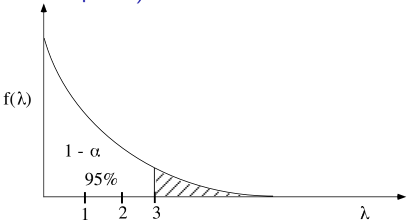
Upper probabilistic limit (e.g. at 95% probability):

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

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

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

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

Do you remember? (From first lecture)

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

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

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)

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

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

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

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.

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

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

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

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.

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:

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

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.

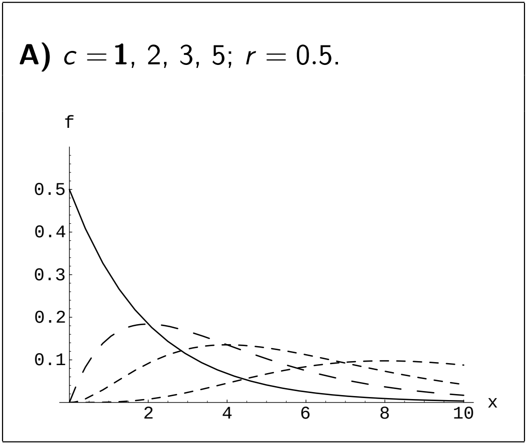
- If c is integer, the distribution is also known as Erlang, describing the time to wait before observing the c-th event in a Poisson process of intensity ('rate') r.
- For $c = 1$ the Gamma distribution recovers the exponential.
- Finally, the χ^2 distribution is just a particular Gamma:

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

- **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

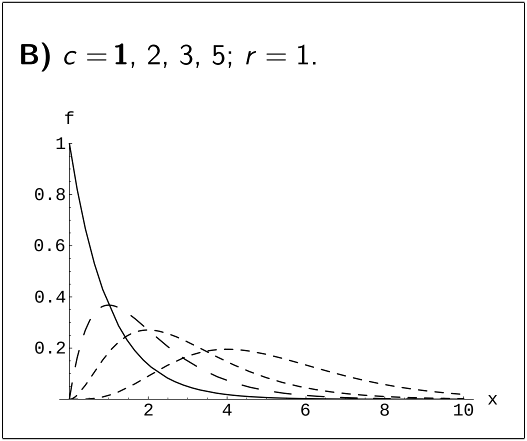
Some examples



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

Gamma distribution

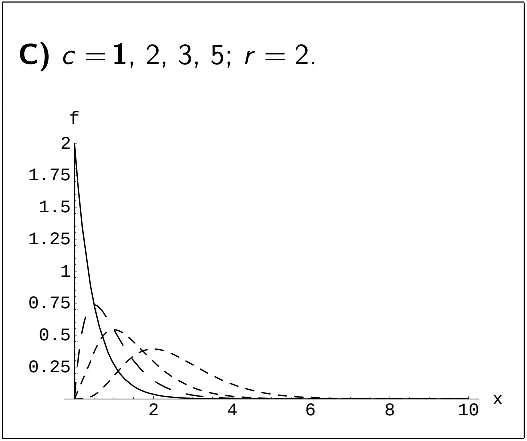
Some examples



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

Gamma distribution

Some examples



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

Gamma (and χ^2) distribution

Summaries

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

Gamma (and χ^2) distribution

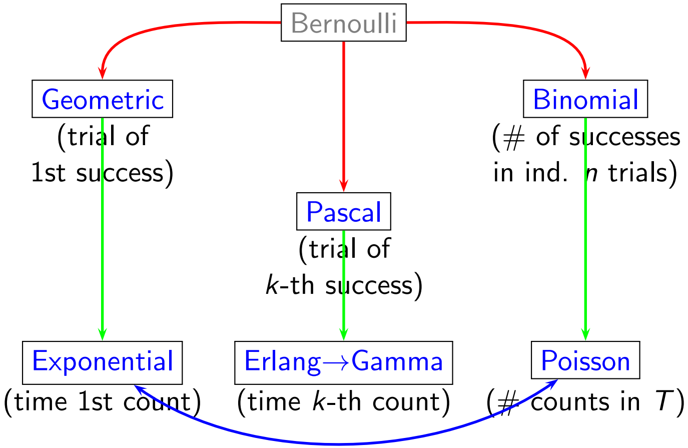
Summaries

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

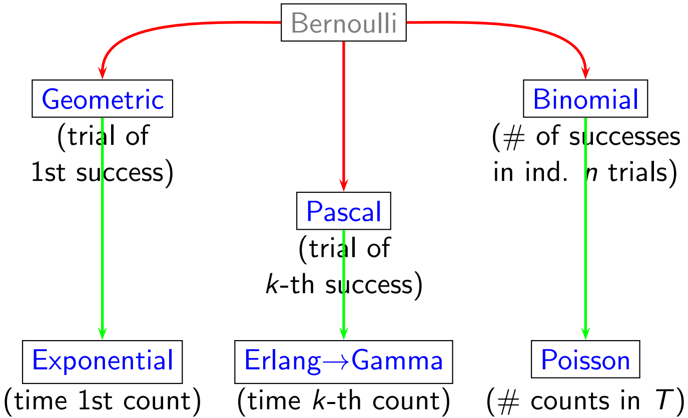
Therefore, for the χ^2 ($\rightarrow c = \nu/2$, $r = 1/2$)

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

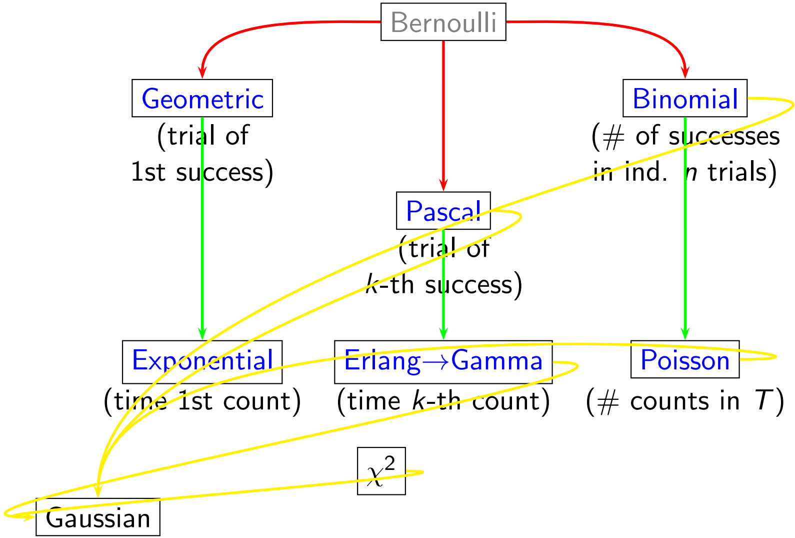
Distributions derived from the Bernoulli process



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

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

  where c_i and r_i are the initial parameters of the gamma distribution.
- Updating rule
  $$c_f = c_i + x$$
  $$r_f = r_i + 1$$
- 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!):
  $\rightarrow$ exponential with very large τ (or vanishing r)
- $c = 1$, $r \rightarrow 0$

$$f(\lambda \,|\, x, \text{Gamma}(c_i = 1, r_i \rightarrow 0)) \propto \lambda^x e^{-\lambda}$$

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: $\rightarrow \lambda = 100 \pm 10$;
- <u>if we were</u> sure that λ was 100, <u>then</u> $x_f = 100 \pm 10$.
- 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:
  $\rightarrow \lambda = 100 \pm 10$;
- <u>if we were</u> sure that λ was 100, <u>then</u> $x_f = 100 \pm 10$.
- but we have to 'convolute' our uncertainty concerning λ
  $\rightarrow$ uncertainty about x_f <u>has</u> 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

Inferring λ and predicting future nr of counts with JAGS

Model file (`inf_lambda_pred.bug`)

```
model {
   X ~ dpois(lambda);
   lambda ~ dexp(0.00001)
   Y ~ dpois(lambda);
}
```

Inferring λ and predicting future nr of counts with JAGS

Model file (`inf_lambda_pred.bug`)

```
model {
   X ~ dpois(lambda);
   lambda ~ dexp(0.00001)
   Y ~ dpois(lambda);
}
```

R stearing script:

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

Adding the background

Adding background – a practical introduction with Jags

- Just an extra, independent, Poisson process in the production of events in the observation time T:

Adding background – a practical introduction with Jags

- Just an extra, independent, Poisson process in the production of events in the observation time T:

$$r = r_s + r_B$$

Adding background – a practical introduction with Jags

- Just an extra, independent, Poisson process in the production of events in the observation time T:

$$
\begin{aligned}
r &= r_s + r_B \\
\lambda = r\,T &= r_s\,T + r_B\,T
\end{aligned}
$$

Adding background – a practical introduction with Jags

- Just an extra, independent, Poisson process in the production of events in the observation time T:

$$\begin{aligned} r &= r_s + r_B \\ \lambda = r\,T &= r_s\,T + r_B\,T \end{aligned}$$

$$X \sim \mathcal{P}_\lambda$$

Adding background – a practical introduction with Jags

- Just an extra, independent, Poisson process in the production of events in the observation time T:

$$\begin{aligned} r &= r_s + r_B \\ \lambda = r\,T &= r_s\,T + r_B\,T \end{aligned}$$

$$\begin{aligned} X &\sim \mathcal{P}_\lambda \\ f(r\,|\,x, r_B, T) &\propto f(x\,|\,r, r_B, T)\cdot f_0(r) \end{aligned}$$

Adding background – a practical introduction with Jags

- Just an extra, independent, Poisson process in the production of events in the observation time T:

$$\begin{aligned} r &= r_s + r_B \\ \lambda = r\,T &= r_s\,T + r_B\,T \end{aligned}$$

$$\begin{aligned} X &\sim \mathcal{P}_\lambda \\ f(r\,|\,x, r_B, T) &\propto f(x\,|\,r, r_B, T)\cdot f_0(r) \end{aligned}$$

Uncertainty on r_B?

Adding background – a practical introduction with Jags

- Just an extra, independent, Poisson process in the production of events in the observation time T:

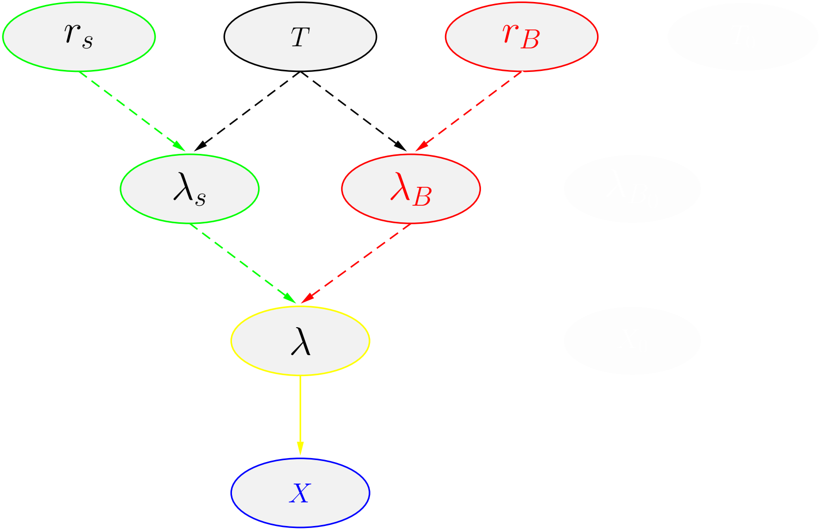
$$\begin{aligned} r &= r_s + r_B \\ \lambda = r\,T &= r_s\,T + r_B\,T \end{aligned}$$

$$\begin{aligned} X &\sim \mathcal{P}_\lambda \\ f(r\,|\,x, r_B, T) &\propto f(x\,|\,r, r_B, T)\cdot f_0(r) \end{aligned}$$

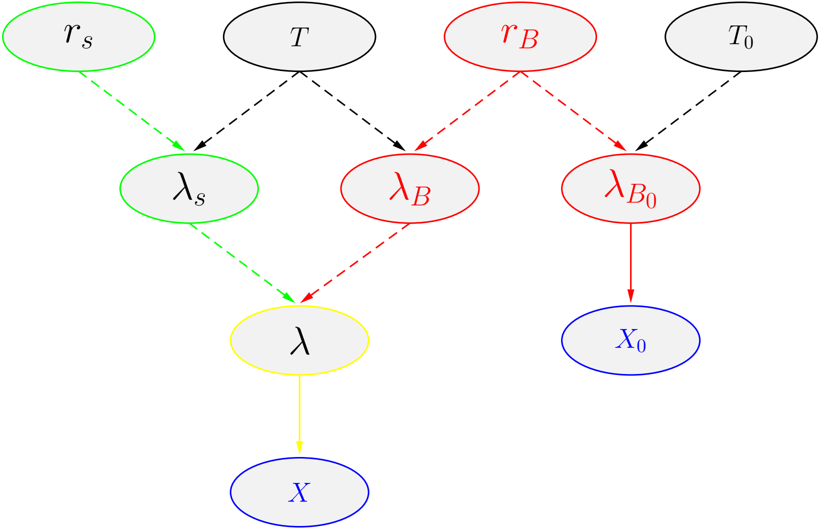
Uncertainty on r_B? Usual way: integrate over all possible values

$$f(r\,|\,x, T) = \int_0^\infty f(r\,|\,x, r_B, T)\cdot f(r_B)\,dr_B$$

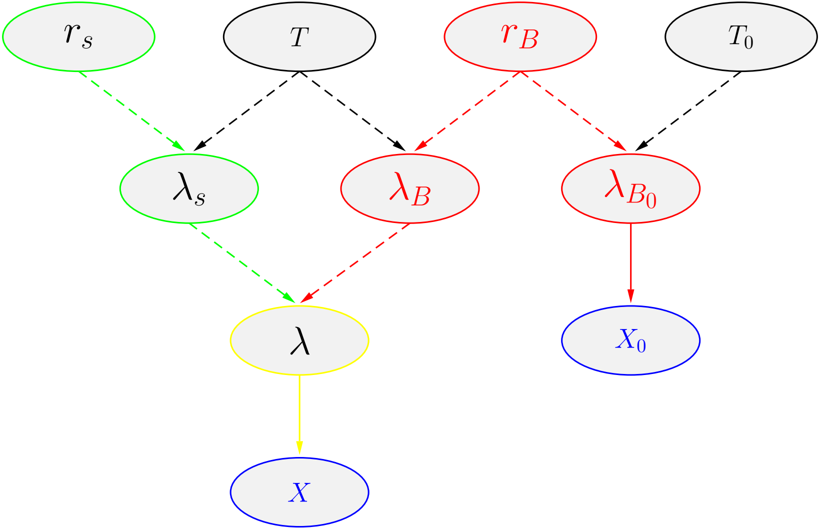
Signal and background



Signal and background



Signal and background



⇒ inf_r_bck_measured.R

⇒ inf_r_bck_measured.bug

Inferring signal and background with JAGS

```
model {
   X ~ dpois(lambda)
   lambda <- ls + lB
   ls <- r * T
   r ~ dgamma(1, 0.00001) # gamma, but indeed dexp(0.00001)
   lB <- rB * T

   # experiment with background only
   lB0 <- rB * TB
   XB ~ dpois(lB0)
   rB ~ dgamma(1, 0.00001) # vague priors also on the bkgd
}
```

Inferring signal and background with JAGS

```
model = "inf_r_bck_measured.bug"  # model file

data <- NULL   # R list containing data
data$X <- 100  # observed nr of counts from signal+bkgd
data$T <- 10   # time of measurement signal+background
data$TB <- 4   # time of measurement of background alone
data$XB <- 20  # observed nr of counts from bkgd alone

jm <- jags.model(model, data)  # define the model

update(jm, 100)    # "burn in": history not recorded,
                   # just to get rid of initial position
                   # (exaggerated in this case!)
chain <- coda.samples(jm, c("r","rB"), n.iter=10000)

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

The End