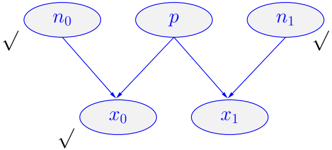
Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

Back to the ‘binomial’ model

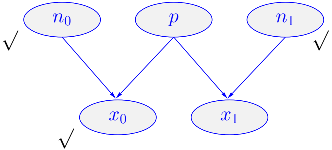
Joint inference and prediction



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

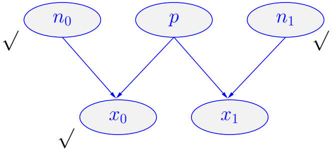


Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

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

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

Joint inference and prediction



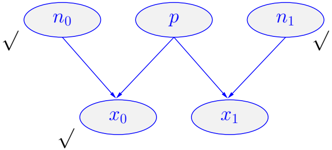
Using the chain rule ('bottom-up')
(and neglecting all factors that do not depend on p and x_1):

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

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

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

Joint inference and prediction



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

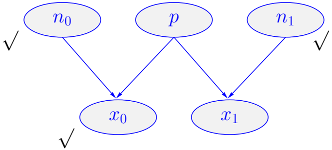
$$f(p, x_1 \mid n_0, x_0, n_1) \propto f(x_0 \mid n_0, p) \cdot f(x_1 \mid p, n_1) \cdot f_0(p)$$

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

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

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):

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

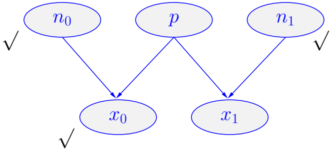
$$\propto p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{(n_1-x_1)!\,x_1!} \cdot f_0(p)$$

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

Problem almost solved

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

Joint inference and prediction



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

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

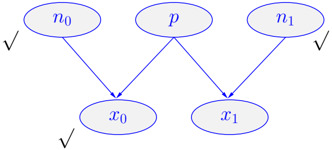
$$\propto\; p^{x_0}(1-p)^{n_0-x_0} \cdot \frac{p^{x_1}(1-p)^{n_1-x_1}}{(n_1-x_1)!\,x_1!} \cdot f_0(p)$$

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

Problem almost solved

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

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,
- or by by sampling.

Details of the numeric solution (→ inf_p_pred_numeric.R)

```
# prior on p (unnormalized beta)
uf0p <- function(p) {
   r0 = 1
   s0 = 1
   ifelse(pp >= 0 && pp <= 1, p^(r0-1)*(1-p)^(s0-1), 0)
}

# unnormalized distribution
uf <- function(p, x1, n0, x0, n1) {
   if(p<0 || p>1)     return(0)
   if(x1 <0 || x1>n1) return(0)
   return( p^(x0+x1)*(1-p)^(n0+n1-x0-x1) /
          (factorial(x1)*factorial(n1-x1)) * uf0p(p) )
}

# normalized distribution (after 'norm' has been evaluated)
f <- function(p, x1, n0, x0, n1) {
     uf(p, x1, n0, x0, n1)/norm}
```

Details of the numeric solution – cont.d

```
# grid in the (p,x1) space
x1 <- 0:n1
N.x1 = n1+1
N.p <- 50                         # nr of intervals in p
Dp = 1/N.p
p <- seq(Dp/2, 1-Dp/2, Dp)       # centers of intervals!

# normalization
norm = 0
for (i in 1:N.p) {
    for (j in 1:N.x1) {
     norm <- norm + uf(p[i], x1[j], n0, x0, n1)
    }
}
```

Details of the numeric solution – cont.d

```
# marginal of p (discretized)
f.p <- rep(0, N.p)
for (i in 1:N.p) {
   for (j in 1:N.x1) {
      f.p[i] <- f.p[i] + f(p[i], x1[j], n0, x0, n1)
   }
}

# marginal of x1
f.x1 <- rep(0,N.x1)
for (j in 1:N.x1) {
   for (i in 1:N.p) {
       f.x1[j] <- f.x1[j] + f(p[i], x1[j], n0, x0, n1)
   }
}
```

Details of the numeric solution – cont.d

```
# moments of p
E.p     <- sum(p*f.p)
sigma.p <- sqrt( sum(p^2*f.p) - E.p^2 )
cat(sprintf("\n  p = %.3f +. %.3f\n", E.p, sigma.p))

# moments of x1
E.x1     <- sum(x1*f.x1)
sigma.x1 <- sqrt( sum(x1^2*f.x1) - E.x1^2 )
cat(sprintf("\n x1 = %.3f +. %.3f\n", E.x1, sigma.x1))

# covariance and correlation coefficient
E.p.times.x1 <- 0 # 1. expected value of the product
for (i in 1:N.p) {
   for (j in 1:N.x1) {
      E.p.times.x1 <- E.p.times.x1 +  p[i] * x1[j] *
                      f(p[i], x1[j], n0, x0, n1)
   }
}
```

Details of the numeric solution – cont.d

```
# covariance and correlation coefficient

# 1. expected value of the product
E.p.times.x1 <- 0
for (i in 1:N.p) {
   for (j in 1:N.x1) {
      E.p.times.x1 <- E.p.times.x1 +  p[i] * x1[j] *
                      f(p[i], x1[j], n0, x0, n1)
   }
}
# 2. covariance
Cov <- E.p.times.x1 - E.p*E.x1

# 3. correlation coefficient
rho <- Cov / (sigma.p * sigma.x1)

cat(sprintf("\n rho(p,x1) = %.3f\n", rho))
```

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Imagine we have a *non trivial $f(x)$*, for a generic x,
and we are interested in sampling it:

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Imagine we have a *non trivial* $f(x)$, for a generic x,
and we are interested in sampling it:
$\Rightarrow$ get a sample of x, such that their relative frequency
(under proper discretization) is proportional to $f(x)$.

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Imagine we have a *non trivial* $f(x)$, for a generic x,
and we are interested in sampling it:
$\Rightarrow$ get a sample of x, such that their relative frequency (under proper discretization) is proportional to $f(x)$.

- if the distribution is 'well known', then we can just use a function of a suitable scientific library,
  e.g., in R, `rnorm()`, `runif()`, `rbeta()`, etc.

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Imagine we have a *non trivial* $f(x)$, for a generic x,
and we are interested in sampling it:
$\Rightarrow$ get a sample of x, such that their relative frequency
(under proper discretization) is proportional to $f(x)$.

- if the distribution is 'well known', then we can just use a function of a suitable scientific library, e.g., in R, `rnorm()`, `runif()`, `rbeta()`, etc.
- if the cumulative distribution, $F(x)$, can be analytically inverted, then just use the method of inverting the cumulative

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Imagine we have a *non trivial $f(x)$*, for a generic x,
and we are interested in sampling it:
$\Rightarrow$ get a sample of x, such that their relative frequency
(under proper discretization) is proportional to $f(x)$.

- if the distribution is 'well known', then we can just use a function of a suitable scientific library,
  e.g., in R, `rnorm()`, `runif()`, `rbeta()`, etc.
- if the cumulative distribution, $F(x)$, can be analytically inverted, then just use the method of inverting the cumulative
  e.g. in R

```
tau=5; hist(-tau*log(runif(100000)), nc=50)
```

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Imagine we have a *non trivial $f(x)$*, for a generic x,
and we are interested in sampling it:
$\Rightarrow$ get a sample of x, such that their relative frequency (under proper discretization) is proportional to $f(x)$.

- if the distribution is 'well known', then we can just use a function of a suitable scientific library,
  e.g., in R, `rnorm()`, `runif()`, `rbeta()`, etc.
- if the cumulative distribution, $F(x)$, can be analytically inverted, then just use the method of inverting the cumulative
  e.g. in R
  ```
  tau=5; hist(-tau*log(runif(100000)), nc=50)
  ```
- if $f(x)$ is defined in a limited region, the we can use the simple hit/miss method.

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Imagine we have a *non trivial $f(x)$*, for a generic x,
and we are interested in sampling it:
$\Rightarrow$ get a sample of x, such that their relative frequency (under proper discretization) is proportional to $f(x)$.

- if the distribution is 'well known', then we can just use a function of a suitable scientific library, e.g., in R, `rnorm()`, `runif()`, `rbeta()`, etc.
- if the cumulative distribution, $F(x)$, can be analytically inverted, then just use the method of inverting the cumulative e.g. in R

  ```
  tau=5; hist(-tau*log(runif(100000)), nc=50)
  ```
- if $f(x)$ is defined in a limited region, the we can use the simple hit/miss method.

Otherwise Markov Chain Monte Carlo, via a suitable algoritm

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Imagine we have a *non trivial* $f(x)$, for a generic x,
and we are interested in sampling it:
⇒ get a sample of x, such that their relative frequency
(under proper discretization) is proportional to $f(x)$.

- if the distribution is 'well known', then we can just use a function of a suitable scientific library,
  e.g., in R, `rnorm()`, `runif()`, `rbeta()`, etc.
- if the cumulative distribution, $F(x)$, can be analytically inverted, then just use the method of inverting the cumulative
  e.g. in R

  ```
  tau=5; hist(-tau*log(runif(100000)), nc=50)
  ```
- if $f(x)$ is defined in a limited region, the we can use the simple hit/miss method.

Otherwise Markov Chain Monte Carlo, via a suitable algoritm

⇒ The Metropolis algorithm is one of the most powerful one.

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Metropolis algorithm

[Note how we are using the variable t, instead of the usual i, to order the steps, to remind of an evolution in time of *system*.]

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Metropolis algorithm

[Note how we are using the variable t, instead of the usual i, to order the steps, to remind of an evolution in time of *system*.]

- Imagine that at a given '$t-1$' the system is in $x^{(t-1)}$:

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Metropolis algorithm

[Note how we are using the variable t, instead of the usual i, to order the steps, to remind of an evolution in time of *system*.]

- Imagine that at a given '$t-1$' the system is in $x^{(t-1)}$:
  - propose a new position x^* chosen at random among the possible states with a *symmetric* proposing function $q()$, i.e. $q(x_i \,|\, x_j) = q(x_j \,|\, x_i)$;

Metropolis algorithm

[Note how we are using the variable t, instead of the usual i, to order the steps, to remind of an evolution in time of *system*.]

- Imagine that at a given '$t-1$' the system is in $x^{(t-1)}$:
  - propose a new position x^* chosen at random among the possible states with a *symmetric* proposing function $q()$, i.e. $q(x_i \,|\, x_j) = q(x_j \,|\, x_i)$;
  - calculate the acceptance probability as

$$A = \min\left(1, \frac{f(x^*)}{f(x^{(t-1)})}\right)$$

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Metropolis algorithm

[Note how we are using the variable t, instead of the usual i, to order the steps, to remind of an evolution in time of *system*.]

- Imagine that at a given '$t-1$' the system is in $x^{(t-1)}$:
  - propose a new position x^* chosen at random among the possible states with a *symmetric* proposing function $q()$, i.e. $q(x_i \,|\, x_j) = q(x_j \,|\, x_i)$;
  - calculate the acceptance probability as

$$A = \min\left(1, \frac{f(x^*)}{f(x^{(t-1)})}\right)$$

  - Move to x^* with probability A, otherwise stay in $x^{(t-1)}$ (in the sense $x^{(t)} = x^{(t-1)}$).

Metropolis algorithm

[Note how we are using the variable t, instead of the usual i, to order the steps, to remind of an evolution in time of *system*.]

- Imagine that at a given '$t-1$' the system is in $x^{(t-1)}$:
  - propose a new position x^* chosen at random among the possible states with a *symmetric* proposing function $q()$, i.e. $q(x_i \,|\, x_j) = q(x_j \,|\, x_i)$;
  - calculate the acceptance probability as

$$A = \min\left(1, \frac{f(x^*)}{f(x^{(t-1)})}\right)$$

  - Move to x^* with probability A, otherwise stay in $x^{(t-1)}$ (in the sense $x^{(t)} = x^{(t-1)}$).

Metropolis algorithm

[Note how we are using the variable t, instead of the usual i, to order the steps, to remind of an evolution in time of *system*.]

- Imagine that at a given '$t-1$' the system is in $x^{(t-1)}$:
  - propose a new position x^* chosen at random among the possible states with a *symmetric* proposing function $q()$, i.e. $q(x_i\,|\,x_j) = q(x_j\,|\,x_i)$;
  - calculate the acceptance probability as

$$A = \min\left(1, \frac{f(x^*)}{f(x^{(t-1)})}\right)$$

  - Move to x^* with probability A, otherwise stay in $x^{(t-1)}$ (in the sense $x^{(t)} = x^{(t-1)}$).

Mostly important: normalization irrelevant!

Metropolis algorithm for MCMC

A flash/practical introduction in 1D

Metropolis algorithm

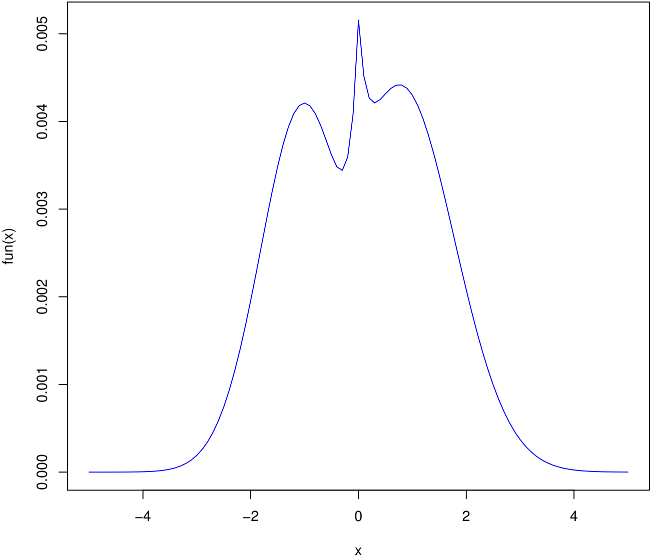
[Note how we are using the variable t, instead of the usual i, to order the steps, to remind of an evolution in time of *system*.]

- Imagine that at a given '$t-1$' the system is in $x^{(t-1)}$:
  - propose a new position x^* chosen at random among the possible states with a *symmetric* proposing function $q()$, i.e. $q(x_i \,|\, x_j) = q(x_j \,|\, x_i)$;
  - calculate the acceptance probability as
$$A = \min\left(1, \frac{f(x^*)}{f(x^{(t-1)})}\right)$$
  - Move to x^* with probability A, otherwise stay in $x^{(t-1)}$ (in the sense $x^{(t)} = x^{(t-1)}$).

Mostly important: normalization irrelevant!

Technically, it can be proved that the algorithm has the *desired properties* to produce a Markov Chain.

Metropolis applied to unnormalized pdf's



Metropolis applied to unnormalized pdf's

⇒ metropolis.R

```
prop <- function(x) {d=1; x + runif(1, -d, d)}

metropolis <- function(n, x0) {
  x = rep(0,n)
  x[1] = x0
  for (i in 2:n) {
    x.p <- prop(x[i-1])
    A <- fun(x.p)/fun(x[i-1])
    x[i] <- ifelse (runif(1) <= A, x.p, x[i-1])
  }
  return(x)
}
```

Metropolis applied to unnormalized pdf's

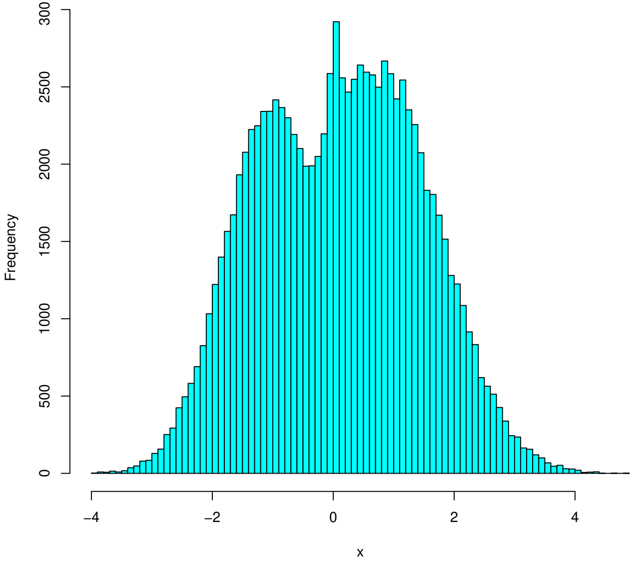
⇒ metropolis.R

```
prop <- function(x) {d=1; x + runif(1, -d, d)}

metropolis <- function(n, x0) {
  x = rep(0,n)
  x[1] = x0
  for (i in 2:n) {
    x.p <- prop(x[i-1])
    A <- fun(x.p)/fun(x[i-1])
    x[i] <- ifelse (runif(1) <= A, x.p, x[i-1])
  }
  return(x)
}
```

```
> n <- 10000
> x0 <- 0
> x <- metropolis(n, x0, fun, prop)
> hist(x, nc=100, col='cyan')
```

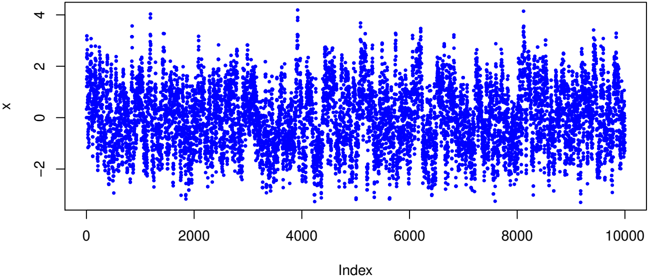
Metropolis applied to unnormalized pdf's

Uniform proposal with $d = 1$



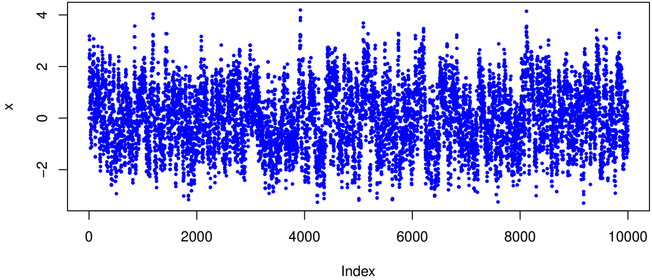
Metropolis applied to unnormalized pdf's

Uniform proposal with $d = 1$

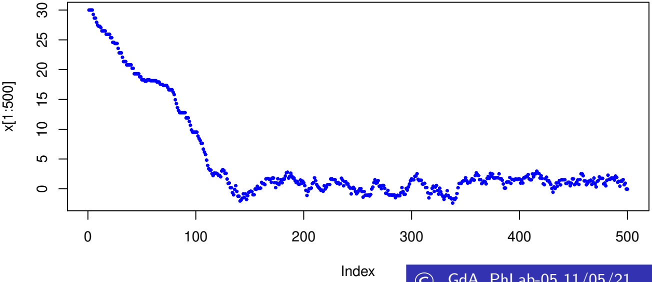


Metropolis applied to unnormalized pdf's

Uniform proposal with $d = 1$

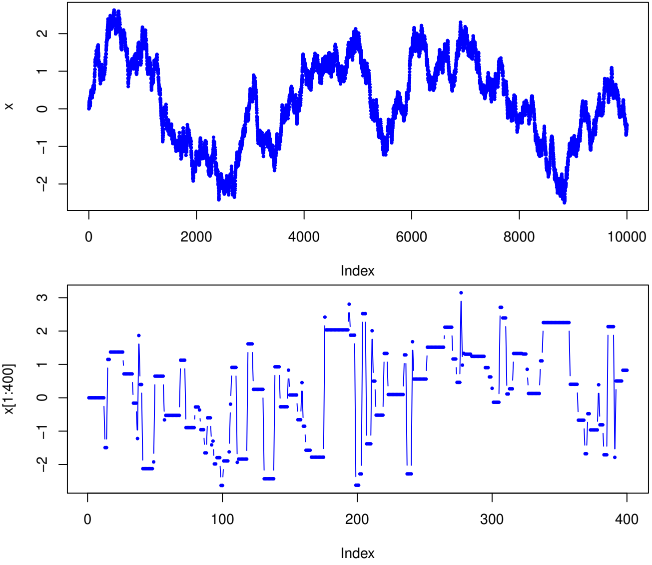


Starting from $x = 30$



Metropolis applied to unnormalized pdf's

Uniform proposal with $d = 0.1$ and $d = 10$



Understanding Metropolis

Proposed exercise Try to use Metropolis in order to make a random walk inside a square, with uniform distribution

Metropolis applied to our binomial problem ($\rightarrow \{p, x_1\}$)

⇒ inf_p_pred_metropolis.R

```
#model parameters
n0 = 20; x0 = 10; n1 = 10

# Metropolis parameters
N = 10000
Dx = 2; Dp = 0.1

# unnormalized distr. (you might want to add a prior)
uf <- function(p, x1, n0, x0, n1) {
  if(p<0 || p>1)     return(0)
  if(x1 <0 || x1>n1) return(0)
  return( p^x0*(1-p)^(n0-x0) * p^x1*(1-p)^(n1-x1) /
          (factorial(x1)*factorial(n1-x1)) )
}

# proposal functions
pr.p  <- function(p.o, Dp) p.o + runif(1, -Dp, +Dp)
pr.x1 <- function(x1.o, Dx) x1.o + sample(-Dx:Dx)[1]
```

Metropolis applied to our binomial problem ($\rightarrow \{p, x_1\}$)

```
# inits  (just empty vectors)
p <- x1 <- numeric(N)

# initial p and x1  (not to confused with the prior on p!!)
p[1] = rbeta(1,1,1)          # uniform (or anything you like...)
x1[1] = rbinom(1, n1, p[1])

# Metropolis
for (i in 2:N) {
   p.p   <- pr.p(p[i-1], Dp)                           # proposals
   x1.p  <- pr.x1(x1[i-1], Dx)
   A <- min(1, uf(p.p, x1.p, n0, x0, n1) /        # acceptance
               uf(p[i-1], x1[i-1], n0, x0, n1)  )
   if ( runif(1) <= A ) {
      p[i]  <- p.p
      x1[i] <- x1.p
   } else {
      p[i]  <- p[i-1]
      x1[i] <- x1[i-1]
   }
}
```

More on Poisson processes

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:

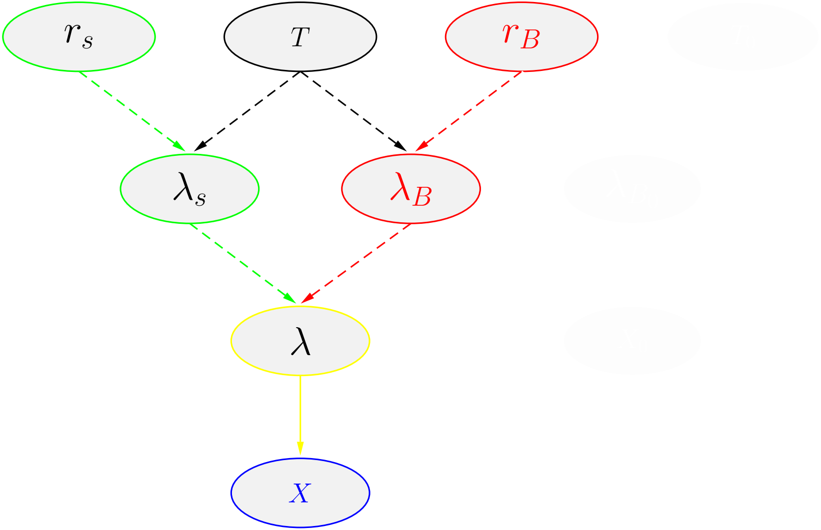
$$
\begin{aligned}
r &= r_s + r_B \\
\lambda = r\,T &= r_s\,T + r_B\,T
\end{aligned}
$$

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

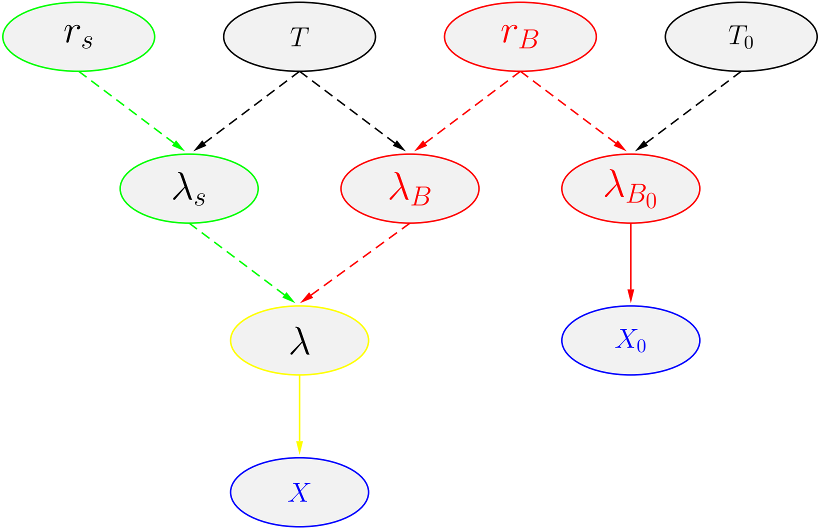
Uncertainty on r_B? Usual way: integrate over all possible values

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

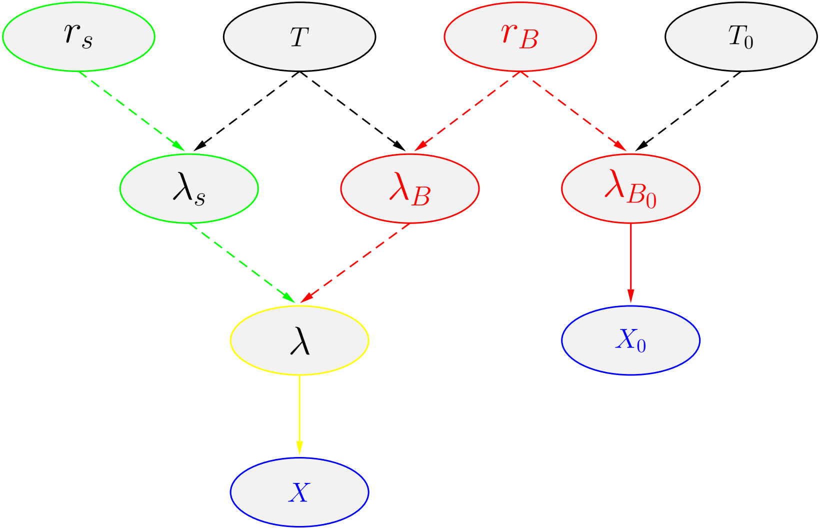
Signal and background



Signal and background



Signal and background



⇒ inf_r_bck_measured.R

⇒ inf_r_bck_measured.bug

Signal + background

JAGS model (⇒ inf_r_bck_measured.bug)

```
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 prior also on the background
}
```

Signal + background

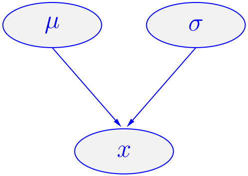
Stearing code (⇒ inf_r_bck_measured.R)

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

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

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

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

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

Gaussian model

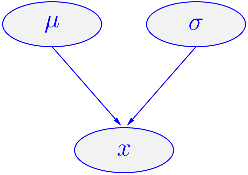
Inferring μ of the normal distribution

Setting up the problem



Inferring μ of the normal distribution

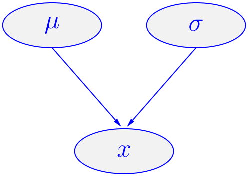
Setting up the problem



- In general $f(x, \mu, \sigma \mid I)$

Inferring μ of the normal distribution

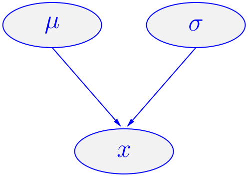
Setting up the problem



- In general $f(x, \mu, \sigma \,|\, I)$
- We start assuming σ well known, that we call here σ_e to remember that it is the standard deviation which describes statistical errors.

Inferring μ of the normal distribution

Setting up the problem

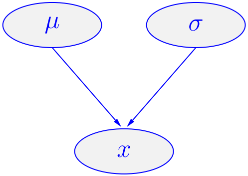


- In general $f(x, \mu, \sigma \mid I)$

- We start assuming σ well known, that we call here σ_e to remember that it is the standard deviation which describes statistical errors.
- And let us start from having observed the 'first' value x_1

Inferring μ of the normal distribution

Setting up the problem



- In general $f(x, \mu, \sigma \mid I)$

- We start assuming σ well known, that we call here σ_e to remember that it is the standard deviation which describes statistical errors.
- And let us start from having observed the 'first' value x_1 (remember that time order is not important; what matters is the order in which the information is used)

Inferring μ of the normal distribution

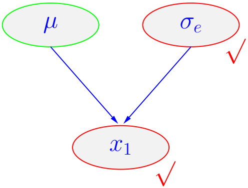
- σ_e assumed perfectly known;
- x_1 observed

Inferring μ of the normal distribution

- σ_e assumed perfectly known;
- x_1 observed ($\equiv$ 'assumed perfectly known')

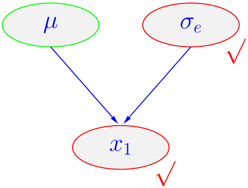
Inferring μ of the normal distribution

- σ_e assumed perfectly known;
- x_1 observed ($\equiv$ 'assumed perfectly known')



Inferring μ of the normal distribution

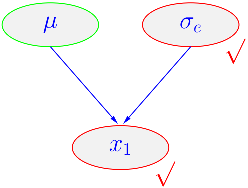
- σ_e assumed perfectly known;
- x_1 observed ($\equiv$ 'assumed perfectly known')



- Our task: $f(\mu \,|\, x_1, \sigma_e)$

Inferring μ of the normal distribution

- σ_e assumed perfectly known;
- x_1 observed ($\equiv$ 'assumed perfectly known')

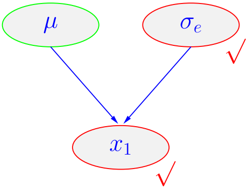


- Our task: $f(\mu \,|\, x_1, \sigma_e)$

- In general: $f(\mu \,|\, \text{data}, I)$

Inferring μ of the normal distribution

- σ_e assumed perfectly known;
- x_1 observed ($\equiv$ 'assumed perfectly known')

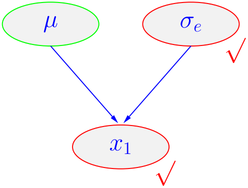


- Our task: $f(\mu \,|\, x_1, \sigma_e)$

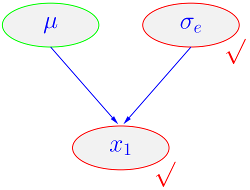
- In general: $f(\mu \,|\, \text{data}, I)$

'data' can be a set of observations

Inferring μ of the normal distribution



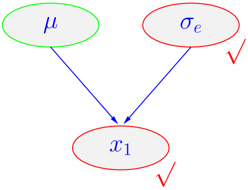
Inferring μ of the normal distribution



(Considering implicit the condition σ_e as well as I)

$$f(\mu \,|\, x_1) \propto f(x_1 \,|\, \mu) \cdot f_0(\mu)$$

Inferring μ of the normal distribution

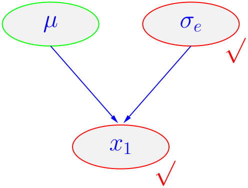


(Considering implicit the condition σ_e as well as I)

$$f(\mu \,|\, x_1) \propto f(x_1 \,|\, \mu) \cdot f_0(\mu)$$

$$f(\mu \,|\, x_1) = \frac{f(x_1 \,|\, \mu) \cdot f_0(\mu)}{f(x_1)}$$

Inferring μ of the normal distribution



(Considering implicit the condition σ_e as well as I)

$$f(\mu \,|\, x_1) \propto f(x_1 \,|\, \mu) \cdot f_0(\mu)$$

$$\begin{aligned} f(\mu \,|\, x_1) &= \frac{f(x_1 \,|\, \mu) \cdot f_0(\mu)}{f(x_1)} \\ &= \frac{f(x_1 \,|\, \mu) \cdot f_0(\mu)}{\int_{-\infty}^{+\infty} f(x_1 \,|\, \mu) \cdot f_0(\mu)\, d\mu} \end{aligned}$$

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

$$f(\mu\,|\,x_1) = \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}}{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}\,\mathrm{d}\mu}$$

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

$$f(\mu \,|\, x_1) \;=\; \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}}{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}\,\mathrm{d}\mu}$$

In the denominator, the exponential depends on $(x_1-\mu)^2$:
$\rightarrow$ the integral over μ is equal to the integral over x_1

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

$$f(\mu \,|\, x_1) \;=\; \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}\,\mathrm{d}\mu}$$

In the denominator, the exponential depends on $(x_1 - \mu)^2$:
$\rightarrow$ the integral over μ is equal to the integral over x_1: $\longrightarrow 1$

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

$$f(\mu\,|\,x_1) \;=\; \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}}{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}\,\mathrm{d}\mu}$$

In the denominator, the exponential depends on $(x_1-\mu)^2$:
$\rightarrow$ the integral over μ is equal to the integral over x_1: $\longrightarrow 1$

$$f(\mu\,|\,x_1) \;=\; \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(\mu-x_1)^2}{2\,\sigma_e^2}}$$

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

$$f(\mu\,|\,x_1) \;=\; \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}}{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}\,\mathrm{d}\mu}$$

In the denominator, the exponential depends on $(x_1-\mu)^2$:
$\rightarrow$ the integral over μ is equal to the integral over x_1: $\longrightarrow 1$

$$f(\mu\,|\,x_1) \;=\; \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(\mu-x_1)^2}{2\,\sigma_e^2}}$$

Note the swap of μ and x_1 at the exponent,
to emphasize that they have now different roles:

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

$$f(\mu\,|\,x_1) \;=\; \frac{\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}}{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}}\,\mathrm{d}\mu}$$

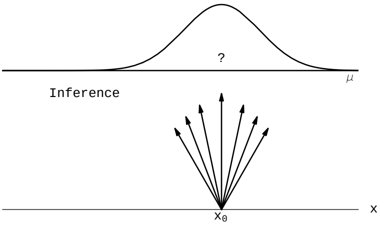
In the denominator, the exponential depends on $(x_1-\mu)^2$:
$\rightarrow$ the integral over μ is equal to the integral over x_1: $\longrightarrow 1$

$$f(\mu\,|\,x_1) \;=\; \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(\mu-x_1)^2}{2\,\sigma_e^2}}$$

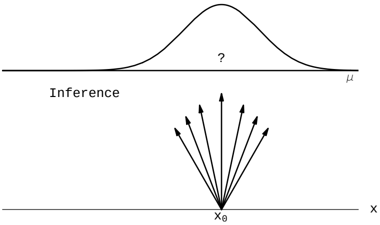
Note the swap of μ and x_1 at the exponent,
to emphasize that they have now different roles:

- μ is the variable;
- x_1 is a parameter

Inferring μ of the normal distribution

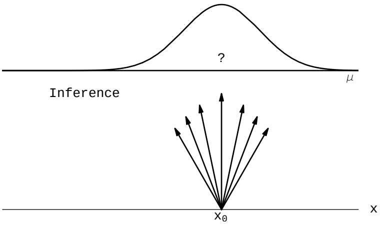


Inferring μ of the normal distribution



$$f(\mu \,|\, x_1) \;=\; \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(\mu-x_1)^2}{2\,\sigma_e^2}}$$

Inferring μ of the normal distribution

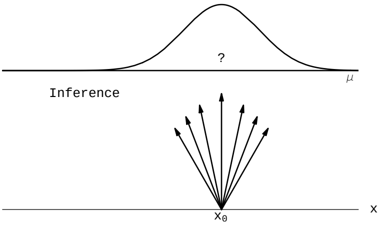


$$f(\mu \,|\, x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_e}\, e^{-\frac{(\mu - x_1)^2}{2\,\sigma_e^2}}$$

Summaries:

$$\begin{aligned} \mathrm{E}[\mu] &= x_1 \\ \sigma(\mu) &= \sigma_e \end{aligned}$$

Inferring μ of the normal distribution



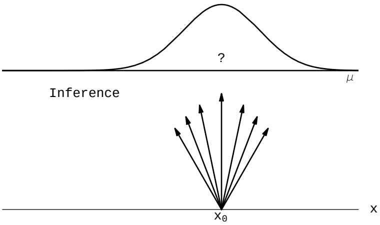
$$f(\mu \,|\, x_1) \;=\; \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\, e^{-\frac{(\mu - x_1)^2}{2\,\sigma_e^2}}$$

Summaries:

$$\mathrm{E}[\mu] \;=\; x_1$$
$$\sigma(\mu) \;=\; \sigma_e$$

All probability intervals calculated from the pdf.

Inferring μ of the normal distribution



$$f(\mu\,|\,x_1) \;=\; \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(\mu-x_1)^2}{2\,\sigma_e^2}}$$

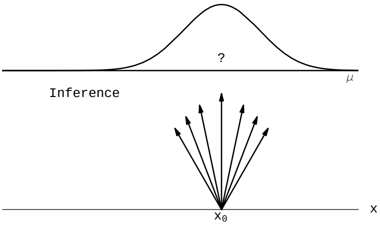
Summaries:

$$\begin{aligned} \mathrm{E}[\mu] &= x_1 \\ \sigma(\mu) &= \sigma_e \end{aligned}$$

All probability intervals calculated from the pdf.

⇒ really **probability intervals**, and not 'confidence intervals'

Inferring μ of the normal distribution



$$f(\mu\,|\,x_1) = \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(\mu-x_1)^2}{2\,\sigma_e^2}}$$

Summaries:

$$\mathrm{E}[\mu] = x_1$$
$$\sigma(\mu) = \sigma_e$$

All probability intervals calculated from the pdf.

⇒ really **probability intervals**, and not 'confidence intervals'(∗)

(∗)The expressions "confidence interval" and "confidence limits" are jeopardized having often little to do with 'confidence' – sic!

Role of the prior

Yes, but the prior?

Think about it. . .

Next time $\Longrightarrow$

Inference and prediction related to Gaussian errors

We have seen the simple case

$$(x_1, \sigma_e) \longrightarrow f(\mu \,|\, x_1, \sigma_e, f_0(\mu) = k)$$

Inference and prediction related to Gaussian errors

We have seen the simple case

$$(x_1, \sigma_e) \;\longrightarrow\; f(\mu \,|\, x_1, \sigma_e, f_0(\mu) = k)$$

Further (important) cases:

- several measurements, each with precisely known σ_i;

Inference and prediction related to Gaussian errors

We have seen the simple case

$$(x_1, \sigma_e) \;\longrightarrow\; f(\mu \,|\, x_1, \sigma_e, f_0(\mu) = k)$$

Further (important) cases:

- several measurements, each with precisely known σ_i;
- taking into account of priors;

Inference and prediction related to Gaussian errors

We have seen the simple case

$$(x_1, \sigma_e) \longrightarrow f(\mu \,|\, x_1, \sigma_e, f_0(\mu) = k)$$

Further (important) cases:

- several measurements, each with precisely known σ_i;
- taking into account of priors;
- getting μ and σ ('σ_e') from a sample

Inference and prediction related to Gaussian errors

We have seen the simple case

$$(x_1, \sigma_e) \;\longrightarrow\; f(\mu \,|\, x_1, \sigma_e, f_0(\mu) = k)$$

Further (important) cases:

- several measurements, each with precisely known σ_i;
- taking into account of priors;
- getting μ and σ ('σ_e') from a sample
  ('small samples' next lecture);

Inference and prediction related to Gaussian errors

We have seen the simple case

$$(x_1, \sigma_e) \;\longrightarrow\; f(\mu \,|\, x_1, \sigma_e, f_0(\mu) = k)$$

Further (important) cases:

- several measurements, each with precisely known σ_i;
- taking into account of priors;
- getting μ and σ ('σ_e') from a sample
  ('small samples' next lecture);
- predicting a new ('future') value of x;

Inference and prediction related to Gaussian errors

We have seen the simple case

$$(x_1, \sigma_e) \longrightarrow f(\mu \,|\, x_1, \sigma_e, f_0(\mu) = k)$$

Further (important) cases:

- several measurements, each with precisely known σ_i;
- taking into account of priors;
- getting μ and σ (‘σ_e’) from a sample
  (‘small samples’ next lecture);
- predicting a new (‘future’) value of x;
- taking into account of systematics

Inference and prediction related to Gaussian errors

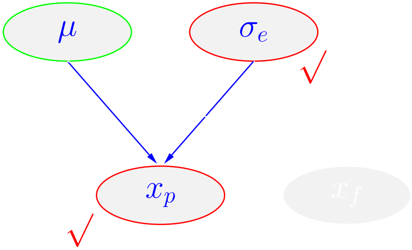
We have seen the simple case

$$(x_1, \sigma_e) \longrightarrow f(\mu \,|\, x_1, \sigma_e, f_0(\mu) = k)$$

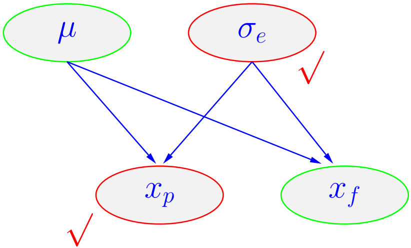
Further (important) cases:

- several measurements, each with precisely known σ_i;
- taking into account of priors;
- getting μ and σ ('σ_e') from a sample
  ('small samples' next lecture);
- predicting a new ('future') value of x;
- taking into account of systematics – general introduction and exact solution for an important case in Physics.

Predictive distribution

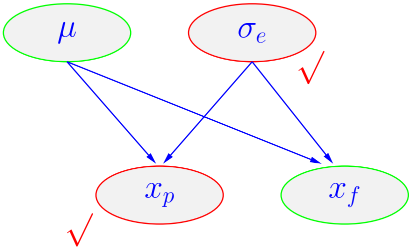


Predictive distribution



What shall we observe in a next measurement x_f ('f' as 'future'), given our knowledge on μ based on the previous observation x_p?

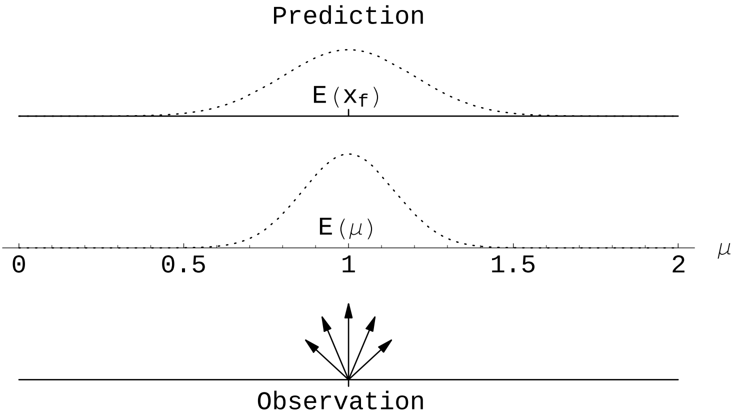
Predictive distribution



What shall we observe in a next measurement x_f ('f' as 'future'), given our knowledge on μ based on the previous observation x_p?
(Note the new evocative name for the observation, instead of x_1)

Predictive distribution

$$x_p \longrightarrow \mu \longrightarrow x_f$$



Predictive distribution

Probability theory teaches us how to include the uncertainty concerning μ:

$$f(x\,|\,I) = \int_{-\infty}^{+\infty} f(x\,|\,\mu, I)\, f(\mu\,|\,I)\, \mathrm{d}\mu\,.$$

Predictive distribution

Probability theory teaches us how to include the uncertainty concerning μ:

$$f(x\,|\,I) = \int_{-\infty}^{+\infty} f(x\,|\,\mu, I)\, f(\mu\,|\,I)\,\mathrm{d}\mu\,.$$

Thus, in our case (considering the possibility that $\sigma_p \neq \sigma_f$)

$$f(x_f\,|\,x_p) \;=\; \int_{-\infty}^{+\infty} f(x_f\,|\,\mu)\cdot f(\mu\,|\,x_p)\,\mathrm{d}\mu$$

Predictive distribution

Probability theory teaches us how to include the uncertainty concerning μ:

$$f(x\,|\,I) = \int_{-\infty}^{+\infty} f(x\,|\,\mu, I)\, f(\mu\,|\,I)\,\mathrm{d}\mu\,.$$

Thus, in our case (considering the possibility that $\sigma_p \neq \sigma_f$)

$$\begin{aligned} f(x_f\,|\,x_p) &= \int_{-\infty}^{+\infty} f(x_f\,|\,\mu)\cdot f(\mu\,|\,x_p)\,\mathrm{d}\mu \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\,\pi}\,\sigma_f}\,\exp\left[-\frac{(x_f-\mu)^2}{2\,\sigma_f^2}\right] \frac{1}{\sqrt{2\,\pi}\,\sigma_p}\,\exp\left[-\frac{(\mu-x_p)^2}{2\,\sigma_p^2}\right]\mathrm{d}\mu \end{aligned}$$

Predictive distribution

Probability theory teaches us how to include the uncertainty concerning μ:

$$f(x\,|\,I) = \int_{-\infty}^{+\infty} f(x\,|\,\mu, I)\, f(\mu\,|\,I)\, \mathrm{d}\mu\,.$$

Thus, in our case (considering the possibility that $\sigma_p \neq \sigma_f$)

$$\begin{aligned}
f(x_f\,|\,x_p) &= \int_{-\infty}^{+\infty} f(x_f\,|\,\mu)\cdot f(\mu\,|\,x_p)\,\mathrm{d}\mu \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\,\pi}\,\sigma_f}\,\exp\left[-\frac{(x_f-\mu)^2}{2\,\sigma_f^2}\right]\frac{1}{\sqrt{2\,\pi}\,\sigma_p}\,\exp\left[-\frac{(\mu-x_p)^2}{2\,\sigma_p^2}\right]\mathrm{d}\mu \\
&= \frac{1}{\sqrt{2\,\pi}\,\sqrt{\sigma_p^2+\sigma_f^2}}\,\exp\left[-\frac{(x_f-x_p)^2}{2\,(\sigma_p^2+\sigma_f^2)}\right]
\end{aligned}$$

Predictive distribution

Probability theory teaches us how to include the uncertainty concerning μ:

$$f(x\,|\,I) = \int_{-\infty}^{+\infty} f(x\,|\,\mu, I)\, f(\mu\,|\,I)\,\mathrm{d}\mu\,.$$

Thus, in our case (considering the possibility that $\sigma_p \neq \sigma_f$)

$$\begin{aligned}
f(x_f\,|\,x_p) &= \int_{-\infty}^{+\infty} f(x_f\,|\,\mu)\cdot f(\mu\,|\,x_p)\,\mathrm{d}\mu \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\,\pi}\,\sigma_f}\,\exp\left[-\frac{(x_f-\mu)^2}{2\,\sigma_f^2}\right]\frac{1}{\sqrt{2\,\pi}\,\sigma_p}\,\exp\left[-\frac{(\mu-x_p)^2}{2\,\sigma_p^2}\right]\mathrm{d}\mu \\
&= \frac{1}{\sqrt{2\,\pi}\,\sqrt{\sigma_p^2+\sigma_f^2}}\,\exp\left[-\frac{(x_f-x_p)^2}{2\,(\sigma_p^2+\sigma_f^2)}\right]
\end{aligned}$$

In particular, if $\sigma_p = \sigma_f = \sigma$, then

$$f(x_f\,|\,x_p, \sigma_p = \sigma_f = \sigma) = \frac{1}{\sqrt{2\,\pi}\,\sqrt{2}\,\sigma}\,\exp\left[-\frac{(x_f-x_p)^2}{2\,(\sqrt{2}\,\sigma)^2}\right]$$

Problem on the expected $\overline{x}_f$ having observed $\overline{x}_p$

Data: $\overline{x}_p = 8.1234$, $s = 0.7234$, $n = 10000$

Problem on the expected $\overline{x}_f$ having observed $\overline{x}_p$

Data: $\overline{x}_p = 8.1234$, $s = 0.7234$, $n = 10000$

$$\mu = \overline{x}_p \pm \frac{s}{\sqrt{n}}$$

Problem on the expected $\overline{x}_f$ having observed $\overline{x}_p$

Data: $\overline{x}_p = 8.1234$, $s = 0.7234$, $n = 10000$

$$\mu = \overline{x}_p \pm \frac{s}{\sqrt{n}} = 8.1234 \pm 0.0072$$

Problem on the expected $\overline{x}_f$ having observed $\overline{x}_p$

Data: $\overline{x}_p = 8.1234$, $s = 0.7234$, $n = 10000$

$$\mu = \overline{x}_p \pm \frac{s}{\sqrt{n}} = 8.1234 \pm 0.0072$$

(based on standard knowledge, including the fact that $\sigma_e \approx s$ with rather good approximation – we shall return on this point later)

Also the question concerning x_f (meant a single observation) is rather easy to answer:

$$x_f = \overline{x}_p \pm s$$

Problem on the expected $\overline{x}_f$ having observed $\overline{x}_p$

Data: $\overline{x}_p = 8.1234$, $s = 0.7234$, $n = 10000$

$$\mu \;=\; \overline{x}_p \pm \frac{s}{\sqrt{n}} = 8.1234 \pm 0.0072$$

(based on standard knowledge, including the fact that $\sigma_e \approx s$ with rather good approximation – we shall return on this point later)

Also the question concerning x_f (meant a single observation) is rather easy to answer:

$$x_f \;=\; \overline{x}_p \pm s = 8.12 \pm 0.72 \qquad \text{(Gaussian)}$$

Problem on the expected $\overline{x}_f$ having observed $\overline{x}_p$

Data: $\overline{x}_p = 8.1234$, $s = 0.7234$, $n = 10000$

$$\mu = \overline{x}_p \pm \frac{s}{\sqrt{n}} = 8.1234 \pm 0.0072$$

(based on standard knowledge, including the fact that $\sigma_e \approx s$ with rather good approximation – we shall return on this point later)

Also the question concerning x_f (meant a single observation) is rather easy to answer:

$$x_f = \overline{x}_p \pm s = 8.12 \pm 0.72 \qquad \text{(Gaussian)}$$

More interesting was question concerning $\overline{x}_f$, remembering that an aritmethic average can be considered an equivalent measurement with 'σ_e' $= \sigma(\overline{x}) = \sigma(x_i)/\sqrt{n}$:

Problem on the expected $\overline{x}_f$ having observed $\overline{x}_p$

Data: $\overline{x}_p = 8.1234$, $s = 0.7234$, $n = 10000$

$$\mu \;\; = \;\; \overline{x}_p \pm \frac{s}{\sqrt{n}} = 8.1234 \pm 0.0072$$

(based on standard knowledge, including the fact that $\sigma_e \approx s$ with rather good approximation – we shall return on this point later)

Also the question concerning x_f (meant a single observation) is rather easy to answer:

$$x_f \;\; = \;\; \overline{x}_p \pm s = 8.12 \pm 0.72 \qquad \text{(Gaussian)}$$

More interesting was question concerning $\overline{x}_f$, remembering that an aritmethic average can be considered an equivalent measurement with 'σ_e' $= \sigma(\overline{x}) = \sigma(x_i)/\sqrt{n}$:

$$\overline{x}_f \;\; = \;\; \overline{x}_p \pm \sqrt{2}\,\frac{s}{\sqrt{n}} = 8.123 \pm 0.010 \qquad \text{(Gaussian)}$$

Expected $\overline{x}_f$ having observed $\overline{x}_p$

However, the factor $\sqrt{2}$ is usually 'forgotten'

Expected $\overline{x}_f$ having observed $\overline{x}_p$

However, the factor $\sqrt{2}$ is usually 'forgotten'

For example, suppose one has n observations of a random variable x and a hypothesis for the p.d.f. $f(x;\theta)$ which contains an unknown parameter θ. From the sample $x_1, \ldots, x_n$ a function $\hat{\theta}(x_1, \ldots, x_n)$ is constructed (e.g. using maximum likelihood) as an estimator for θ. Using one of the techniques discussed in Chapters 5–8 (e.g. analytic method, RCF bound, Monte Carlo, graphical) the standard deviation of $\hat{\theta}$ can be estimated. Let $\hat{\theta}_{\rm obs}$ be the value of the estimator actually observed, and $\hat{\sigma}_{\hat{\theta}}$ the estimate of its standard deviation. In reporting the measurement of θ as $\hat{\theta}_{\rm obs} \pm \hat{\sigma}_{\hat{\theta}}$ one means that repeated estimates all based

on n observations of x would be distributed according to a p.d.f. $g(\hat{\theta})$ centered around some true value θ and true standard deviation $\sigma_{\hat{\theta}}$, which are estimated to be $\hat{\theta}_{\rm obs}$ and $\hat{\sigma}_{\hat{\theta}}$.

Expected $\overline{x}_f$ having observed $\overline{x}_p$

However, the factor $\sqrt{2}$ is usually 'forgotten'

For example, suppose one has n observations of a random variable x and a hypothesis for the p.d.f. $f(x;\theta)$ which contains an unknown parameter θ. From the sample $x_1,\ldots,x_n$ a function $\hat{\theta}(x_1,\ldots,x_n)$ is constructed (e.g. using maximum likelihood) as an estimator for θ. Using one of the techniques discussed in Chapters 5–8 (e.g. analytic method, RCF bound, Monte Carlo, graphical) the standard deviation of $\hat{\theta}$ can be estimated. Let $\hat{\theta}_{\rm obs}$ be the value of the estimator actually observed, and $\hat{\sigma}_{\hat{\theta}}$ the estimate of its standard deviation. In reporting the measurement of θ as $\hat{\theta}_{\rm obs} \pm \hat{\sigma}_{\hat{\theta}}$ one means that repeated estimates all based

Classical confidence intervals (exact method) **119**

on n observations of x would be distributed according to a p.d.f. $g(\hat{\theta})$ centered around some true value θ and true standard deviation $\sigma_{\hat{\theta}}$, which are estimated to be $\hat{\theta}_{\rm obs}$ and $\hat{\sigma}_{\hat{\theta}}$.

(Glen Cowan, *Statistical Data Analysis*)

Remark on 'conventional statistics'

Objection:

"A method which is 'classical' and 'exact' cannot be wrong"

Remark on ‘conventional statistics’

Objection:

“A method which is ‘classical’ and ‘exact’ cannot be wrong”

Uhm. . .

Remark on ‘conventional statistics’

Objection:

“A method which is ‘classical’ and ‘exact’ cannot be wrong”

Uhm. . .

- Frequentist ‘gurus’ are champions in misusing terminonology, thus confusing people (“CL”, “confidence intervals”).

Remark on 'conventional statistics'

Objection:

"A method which is 'classical' and 'exact' cannot be wrong"

Uhm...

- Frequentist 'gurus' are champions in misusing terminonology, thus confusing people ("CL", "confidence intervals").
- Details in GdA, *About the proof of the so called exact classical confidence intervals. Where is the trick?*, https://arxiv.org/abs/physics/0605140

Remark on 'conventional statistics'

Objection:

"A method which is 'classical' and 'exact' cannot be wrong"

Uhm. . .

- Frequentist 'gurus' are champions in misusing terminonology, thus confusing people ("CL", "confidence intervals").
- Details in GdA, *About the proof of the so called exact classical confidence intervals. Where is the trick?*, https://arxiv.org/abs/physics/0605140

If you like, the method is exact not because it provides precisely the correct answer to our problem

Remark on 'conventional statistics'

Objection:

"A method which is 'classical' and 'exact' cannot be wrong"

Uhm. . .

- Frequentist 'gurus' are champions in misusing terminonology, thus confusing people ("CL", "confidence intervals").
- Details in GdA, *About the proof of the so called exact classical confidence intervals. Where is the trick?*, `https://arxiv.org/abs/physics/0605140`

If you like, the method is exact not because it provides precisely the correct answer to our problem, but because it results from an exact prescription.

Remark on ‘conventional statistics’

Objection:

“A method which is ‘classical’ and ‘exact’ cannot be wrong”

Uhm. . .

- Frequentist ‘gurus’ are champions in misusing terminonology, thus confusing people (“CL”, “confidence intervals”).
- Details in GdA, *About the proof of the so called exact classical confidence intervals. Where is the trick?*, https://arxiv.org/abs/physics/0605140

If you like, the method is exact not because it provides precisely the correct answer to our problem, but because it results from an exact prescription.

Q. Does the method always produce wrong results?

Remark on 'conventional statistics'

Objection:

"A method which is 'classical' and 'exact' cannot be wrong"

Uhm. . .

- Frequentist 'gurus' are champions in misusing terminonology, thus confusing people ("CL", "confidence intervals").
- Details in GdA, *About the proof of the so called exact classical confidence intervals. Where is the trick?*, `https://arxiv.org/abs/physics/0605140`

If you like, the method is exact not because it provides precisely the correct answer to our problem, but because it results from an exact prescription.

Q. Does the method always produce wrong results?

A. In most routine cases the answer is *'numerically' OK*.

Remark on 'conventional statistics'

Objection:

"A method which is 'classical' and 'exact' cannot be wrong"

Uhm. . .

- Frequentist 'gurus' are champions in misusing terminonology, thus confusing people ("CL", "confidence intervals").
- Details in GdA, *About the proof of the so called exact classical confidence intervals. Where is the trick?*, `https://arxiv.org/abs/physics/0605140`

If you like, the method is exact not because it provides precisely the correct answer to our problem, but because it results from an exact prescription.

Q. Does the method always produce wrong results?

A. In most routine cases the answer is *'numerically' OK*.
In Frontier Physics cases this is often not the case (!).

Remark on 'conventional statistics'

Objection:

"A method which is 'classical' and 'exact' cannot be wrong"

Uhm. . .

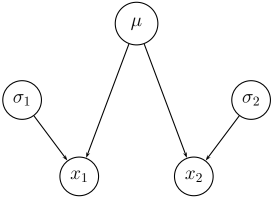
- Frequentist 'gurus' are champions in misusing terminonology, thus confusing people ("CL", "confidence intervals").
- Details in GdA, *About the proof of the so called exact classical confidence intervals. Where is the trick?*, `https://arxiv.org/abs/physics/0605140`

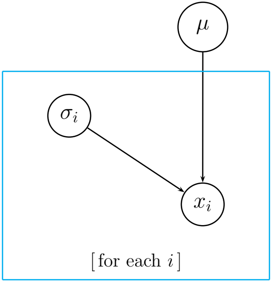
If you like, the method is exact not because it provides precisely the correct answer to our problem, but because it results from an exact prescription.

Q. Does the method always produce wrong results?

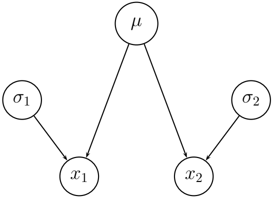
A. In most routine cases the answer is *'numerically' OK*.
In Frontier Physics cases this is often not the case (!).
GdA, *Bayesian reasoning versus conventional statistics in High Energy Physics*,
`https://arxiv.org/abs/physics/9811046`

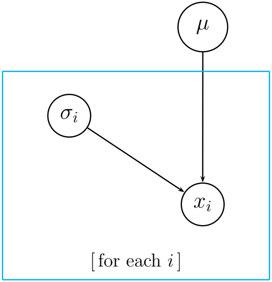
Inferring μ from several *'independent'* measurements





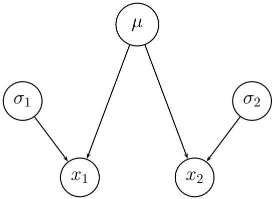
Inferring μ from several *'independent'* measurements

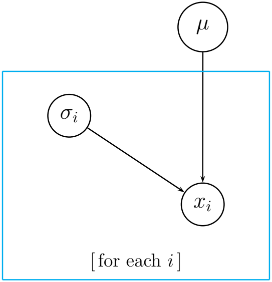




$x_i \sim \mathcal{N}(\mu, \sigma_i)$ with σ_i assumed **precisely known**

Inferring μ from several *'independent'* measurements

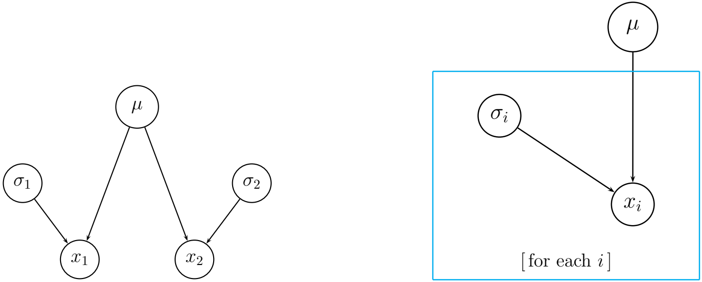




$x_i \sim \mathcal{N}(\mu, \sigma_i)$ with σ_i assumed **precisely known**

$$f(\mu \,|\, x_1, x_2, \ldots, \sigma_1, \sigma_2, \ldots, I) \equiv f(\mu \,|\, \underline{x}, \underline{\sigma} \ldots, I).$$

Inferring μ from several *'independent'* measurements

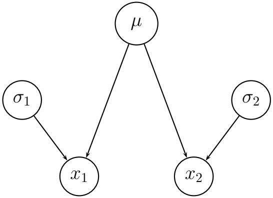


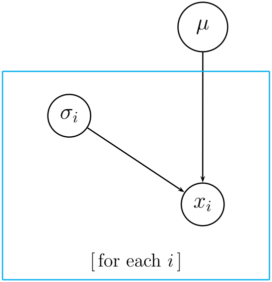
$x_i \sim \mathcal{N}(\mu, \sigma_i)$ with σ_i assumed **precisely known**

$$f(\mu \,|\, x_1, x_2, \ldots, \sigma_1, \sigma_2, \ldots, I) \equiv f(\mu \,|\, \underline{x}, \underline{\sigma} \ldots, I).$$

$$f(\mu \,|\, \underline{x}, \underline{\sigma}, I) \;=\; \frac{f(\mu, \underline{x} \,|\, \underline{\sigma}, I)}{f(\underline{x} \,|\, \underline{\sigma}, I)}$$

Inferring μ from several *'independent'* measurements





$x_i \sim \mathcal{N}(\mu, \sigma_i)$ with σ_i assumed **precisely known**

$$f(\mu \,|\, x_1, x_2, \ldots, \sigma_1, \sigma_2, \ldots, I) \equiv f(\mu \,|\, \underline{x}, \underline{\sigma} \ldots, I).$$

$$\begin{aligned} f(\mu \,|\, \underline{x}, \underline{\sigma}, I) &= \frac{f(\mu, \underline{x} \,|\, \underline{\sigma}, I)}{f(\underline{x} \,|\, \underline{\sigma}, I)} \\ &\propto f(\mu, \underline{x} \,|\, \underline{\sigma}, I)\,. \end{aligned}$$

Independency vs **conditional independency**

Important remark:

Are the x_i of this model really *independent*?

Independency vs **conditional independency**

Important remark:

Are the x_i of this model really *independent*?

- **NO**

Independency vs **conditional independency**

Important remark:

Are the x_i of this model really *independent*?

- **NO** $\Rightarrow$ if they were independent it would be impossible to learn from each other

Independency vs **conditional independency**

Important remark:

Are the x_i of this model really *independent*?

- **NO** $\Rightarrow$ if they were independent it would be impossible to learn from each other

$$\longrightarrow \quad f(x_f \,|\, x_p)$$

Independency vs **conditional independency**

Important remark:

Are the x_i of this model really *independent*?

- **NO** $\Rightarrow$ if they were independent it would be impossible to learn from each other

$$\begin{aligned} & \longrightarrow & f(x_f \,|\, x_p) \\ f(x_1, x_2) & = & f(x_1 \,|\, x_2) \cdot f(x_2) \end{aligned}$$

- They are, more precisely, conditionally independent:
  - **given $\boldsymbol{\mu}$** they become independent

$$f(x_1, x_2 \,|\, \mu) \;=\; f(x_1 \,|\, x_2, \mu) \cdot f(x_2 \,|\, \mu)$$

Independency vs **conditional independency**

Important remark:

Are the x_i of this model really *independent*?

- **NO** $\Rightarrow$ if they were independent it would be impossible to learn from each other

$$\begin{aligned} & \longrightarrow && f(x_f \,|\, x_p) \\ f(x_1, x_2) &= && f(x_1 \,|\, x_2) \cdot f(x_2) \end{aligned}$$

- They are, more precisely, conditionally independent:
  - **given $\boldsymbol{\mu}$** they become independent

$$\begin{aligned} f(x_1, x_2 \,|\, \mu) &= f(x_1 \,|\, x_2, \mu) \cdot f(x_2 \,|\, \mu) \\ &= f(x_1 \,|\mu) \cdot f(x_2 \,|\, \mu) \end{aligned}$$

Independency vs **conditional independency**

Important remark:

Are the x_i of this model really *independent*?

- **NO** $\Rightarrow$ if they were independent it would be impossible to learn from each other

$$
\begin{aligned}
 & \longrightarrow & f(x_f \,|\, x_p) \\
f(x_1, x_2) & = & f(x_1 \,|\, x_2) \cdot f(x_2)
\end{aligned}
$$

- They are, more precisely, conditionally independent:
  - **given $\boldsymbol{\mu}$** they become independent

$$
\begin{aligned}
f(x_1, x_2 \,|\, \mu) & = & f(x_1 \,|\, x_2, \mu) \cdot f(x_2 \,|\, \mu) \\
 & = & f(x_1 \,|\mu) \cdot f(x_2 \,|\, \mu)
\end{aligned}
$$

Similar to the **six boxes**:

– Composition perfectly known $\rightarrow$ $E^{(i)}$ and $E^{(j)}$ independent;

Independency vs **conditional independency**

Important remark:

Are the x_i of this model really *independent*?

- **NO** $\Rightarrow$ if they were independent it would be impossible to learn from each other

$$
\begin{aligned}
& \longrightarrow && f(x_f \,|\, x_p) \\
f(x_1, x_2) &= && f(x_1 \,|\, x_2) \cdot f(x_2)
\end{aligned}
$$

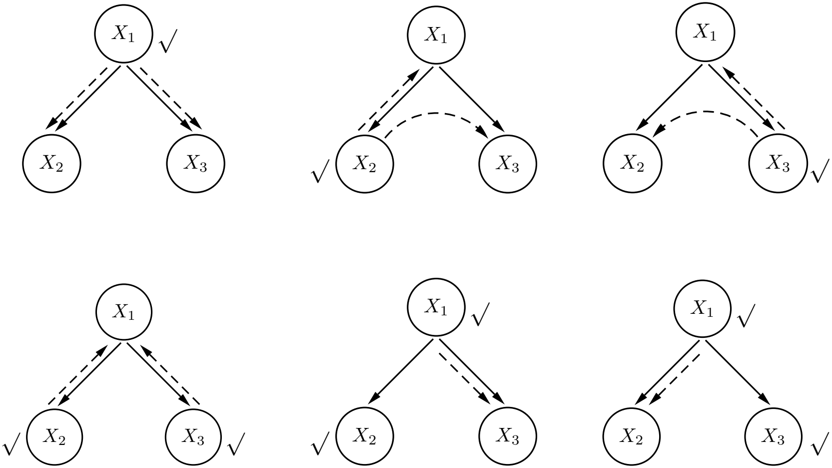
- They are, more precisely, conditionally independent:
  - **given $\boldsymbol{\mu}$** they become independent

$$
\begin{aligned}
f(x_1, x_2 \,|\, \mu) &= f(x_1 \,|\, x_2, \mu) \cdot f(x_2 \,|\, \mu) \\
&= f(x_1 \,|\mu) \cdot f(x_2 \,|\, \mu)
\end{aligned}
$$

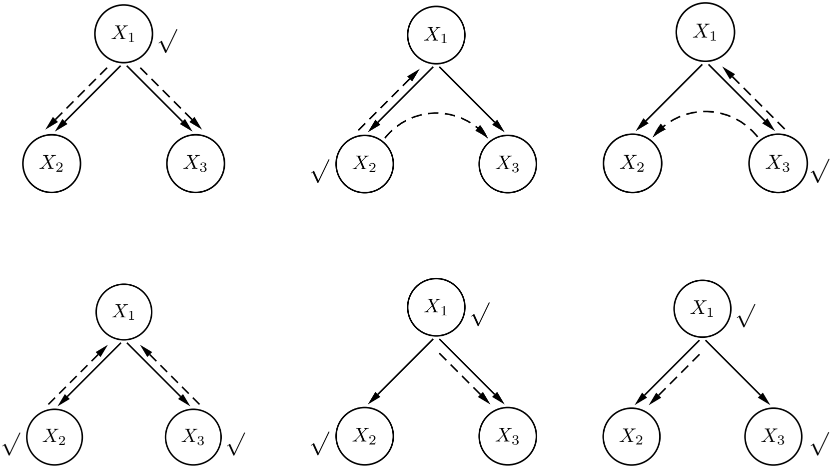
Similar to the **six boxes**:

– Composition perfectly known $\rightarrow E^{(i)}$ and $E^{(j)}$ independent;
– Uncertain composition $\rightarrow E^{(i)}$ changes $P(E^{(j)})$.

Propagation of evidence in a 'divergent connection'



Propagation of evidence in a ‘divergent connection’



More in arXiv:1504.02065 (*“Learning about probabilistic inference and forecasting by playing with multivariate normal distributions”*)

From two observations to n observations

Using the chain rule

$$f(x_1, x_2, \mu, \,|\, \sigma_1, \sigma_2, I) \;=\; f(x_1 \,|\, x_2, \mu, \sigma_1, \sigma_2, I)$$

From two observations to n observations

Using the chain rule

$$f(x_1, x_2, \mu, \,|\, \sigma_1, \sigma_2, I) \;=\; f(x_1 \,|\, x_2, \mu, \sigma_1, \sigma_2, I) \cdot f(x_2 \,|\, \mu, \sigma_1, \sigma_2, I) \cdot$$

From two observations to n observations

Using the chain rule

$$\begin{aligned} f(x_1, x_2, \mu, \,|\, \sigma_1, \sigma_2, I) &= f(x_1 \,|\, x_2, \mu, \sigma_1, \sigma_2, I) \cdot \\ & \quad f(x_2 \,|\, \mu, \sigma_1, \sigma_2, I) \cdot \\ & \quad f(\mu \,|\, \sigma_1, \sigma_2, I) \end{aligned}$$

From two observations to n observations

Using the chain rule

$$
\begin{aligned}
f(x_1, x_2, \mu, \mid \sigma_1, \sigma_2, I) &= f(x_1 \mid x_2, \mu, \sigma_1, \sigma_2, I) \cdot \\
&\quad\; f(x_2 \mid \mu, \sigma_1, \sigma_2, I) \cdot \\
&\quad\; f(\mu \mid \sigma_1, \sigma_2, I) \\
&= f(x_1 \mid \mu, \sigma_1, I) \cdot f(x_2 \mid \mu, \sigma_2, I) \cdot f(\mu \mid I)
\end{aligned}
$$

From two observations to n observations

Using the chain rule

$$\begin{aligned} f(x_1, x_2, \mu, \,|\, \sigma_1, \sigma_2, I) &= f(x_1 \,|\, x_2, \mu, \sigma_1, \sigma_2, I) \cdot \\ &\quad f(x_2 \,|\, \mu, \sigma_1, \sigma_2, I) \cdot \\ &\quad f(\mu \,|\, \sigma_1, \sigma_2, I) \\ &= f(x_1 \,|\, \mu, \sigma_1, I) \cdot f(x_2 \,|\, \mu, \sigma_2, I) \cdot f(\mu \,|\, I) \end{aligned}$$

$$\Rightarrow \; f(\underline{x}, \mu, \,|\, \underline{\sigma}) = \left[\prod_i f(x_i \,|\, \mu, \sigma_i)\right] \cdot f_0(\mu)$$

Using (for the moment) a uniform prior (*practically flat* distribution in the region of interest):

$$f(\mu \,|\, \underline{x}, \underline{\sigma}, f_0(\mu) = k) \;\propto\; f(\underline{x}, \mu, \,|\, \underline{\sigma}) \;=\; \prod_i f_{\mathcal{N}}(x_i \,|\, \mu, \sigma_i)$$

Details of the calculations

$$f(\mu \,|\, \underline{x}, \underline{\sigma}, f_0(\mu) = k) \;\propto\; \prod_i \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]$$

Details of the calculations

$$f(\mu\,|\,\underline{x},\underline{\sigma},f_0(\mu)=k) \;\propto\; \prod_i \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]$$

$$\propto\; \exp\left[-\sum_i \frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]$$

Details of the calculations

$$
\begin{aligned}
f(\mu\,|\,\underline{x},\underline{\sigma},f_0(\mu)=k) &\propto \prod_i \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto \exp\left[-\sum_i \frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto \exp\left[-\frac{1}{2}\sum_i \frac{x_i^2-2x_i\mu+\mu^2}{\sigma_i^2}\right]
\end{aligned}
$$

Details of the calculations

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \underline{\sigma}, f_0(\mu) = k) &\propto \prod_i \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto \exp\left[-\sum_i \frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto \exp\left[-\frac{1}{2}\sum_i \frac{x_i^2 - 2x_i\mu + \mu^2}{\sigma_i^2}\right] \\
&\propto \exp\left[-\frac{1}{2}\sum_i \left(\frac{x_i^2}{\sigma_i^2} - 2\,\frac{x_i}{\sigma_i^2}\,\mu + \frac{\mu^2}{\sigma_i^2}\right)\right]
\end{aligned}
$$

Details of the calculations

$$\begin{aligned}
f(\mu\,|\,\underline{x},\underline{\sigma},f_0(\mu)=k) &\propto \prod_i \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto \exp\left[-\sum_i \frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto \exp\left[-\frac{1}{2}\sum_i \frac{x_i^2-2x_i\mu+\mu^2}{\sigma_i^2}\right] \\
&\propto \exp\left[-\frac{1}{2}\sum_i \left(\frac{x_i^2}{\sigma_i^2}-2\,\frac{x_i}{\sigma_i^2}\,\mu+\frac{\mu^2}{\sigma_i^2}\right)\right]
\end{aligned}$$

$$\propto \exp\left[-\frac{1}{2}\cdot\frac{\sum_i 1/\sigma_i^2}{\sum_i 1/\sigma_i^2}\cdot\left(\sum_i \frac{x_i^2}{\sigma_i^2}-2\left(\sum_i \frac{x_i}{\sigma_i^2}\right)\mu+\left(\sum_i \frac{1}{\sigma_i^2}\right)\mu^2\right)\right]$$

Details of the calculations

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \underline{\sigma}, f_0(\mu) = k) \;&\propto\; \prod_i \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto\; \exp\left[-\sum_i \frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto\; \exp\left[-\frac{1}{2}\sum_i \frac{x_i^2 - 2x_i\mu + \mu^2}{\sigma_i^2}\right] \\
&\propto\; \exp\left[-\frac{1}{2}\sum_i \left(\frac{x_i^2}{\sigma_i^2} - 2\,\frac{x_i}{\sigma_i^2}\,\mu + \frac{\mu^2}{\sigma_i^2}\right)\right]
\end{aligned}
$$

$$
\propto\; \exp\left[-\frac{1}{2}\cdot\frac{\sum_i 1/\sigma_i^2}{\sum_i 1/\sigma_i^2}\cdot\left(\sum_i \frac{x_i^2}{\sigma_i^2} - 2\left(\sum_i \frac{x_i}{\sigma_i^2}\right)\mu + \left(\sum_i \frac{1}{\sigma_i^2}\right)\mu^2\right)\right]
$$

$$
\propto\; \exp\left[-\frac{1}{2}\cdot\left(\sum_i 1/\sigma_i^2\right)\cdot\left(\overline{x^2} - 2\,\overline{x}\,\mu + \mu^2\right)\right]
$$

Details of the calculations

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \underline{\sigma}, f_0(\mu) = k) &\propto \prod_i \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto \exp\left[-\sum_i \frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto \exp\left[-\frac{1}{2}\sum_i \frac{x_i^2 - 2x_i\mu + \mu^2}{\sigma_i^2}\right] \\
&\propto \exp\left[-\frac{1}{2}\sum_i \left(\frac{x_i^2}{\sigma_i^2} - 2\,\frac{x_i}{\sigma_i^2}\,\mu + \frac{\mu^2}{\sigma_i^2}\right)\right]
\end{aligned}
$$

$$
\propto \exp\left[-\frac{1}{2}\cdot\frac{\sum_i 1/\sigma_i^2}{\sum_i 1/\sigma_i^2}\cdot\left(\sum_i \frac{x_i^2}{\sigma_i^2} - 2\left(\sum_i \frac{x_i}{\sigma_i^2}\right)\mu + \left(\sum_i \frac{1}{\sigma_i^2}\right)\mu^2\right)\right]
$$

$$
\propto \exp\left[-\frac{1}{2}\cdot\left(\sum_i 1/\sigma_i^2\right)\cdot\left(\overline{x^2} - 2\,\overline{x}\,\mu + \mu^2\right)\right]
$$

$$
\propto \exp\left[-\frac{\overline{x^2} - 2\,\overline{x}\,\mu + \mu^2}{2/(\sum_i 1/\sigma_i^2)}\right]
$$

Details of the calculations

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \underline{\sigma}, f_0(\mu) = k) \;&\propto\; \prod_i \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto\; \exp\left[-\sum_i \frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \\
&\propto\; \exp\left[-\frac{1}{2}\sum_i \frac{x_i^2 - 2x_i\mu + \mu^2}{\sigma_i^2}\right] \\
&\propto\; \exp\left[-\frac{1}{2}\sum_i \left(\frac{x_i^2}{\sigma_i^2} - 2\,\frac{x_i}{\sigma_i^2}\,\mu + \frac{\mu^2}{\sigma_i^2}\right)\right]
\end{aligned}
$$

$$
\propto\; \exp\left[-\frac{1}{2}\cdot\frac{\sum_i 1/\sigma_i^2}{\sum_i 1/\sigma_i^2}\cdot\left(\sum_i \frac{x_i^2}{\sigma_i^2} - 2\left(\sum_i \frac{x_i}{\sigma_i^2}\right)\mu + \left(\sum_i \frac{1}{\sigma_i^2}\right)\mu^2\right)\right]
$$

$$
\propto\; \exp\left[-\frac{1}{2}\cdot\left(\sum_i 1/\sigma_i^2\right)\cdot\left(\overline{x^2} - 2\,\overline{x}\,\mu + \mu^2\right)\right]
$$

$$
\propto\; \exp\left[-\frac{\overline{x^2} - 2\,\overline{x}\,\mu + \mu^2}{2/(\sum_i 1/\sigma_i^2)}\right] \;\propto\; \exp\left[-\frac{-2\,\overline{x}\,\mu + \mu^2}{2\,\sigma_C^2}\right]
$$

Inferring μ from n 'independent' measurements

$$f(\mu\,|\,\underline{x},\underline{\sigma},f_0(\mu)=k) \quad\propto\quad \exp\left[-\frac{-2\,\overline{x}\,\mu+\mu^2}{2\,\sigma_C^2}\right]$$

Inferring μ from n 'independent' measurements

$$f(\mu \,|\, \underline{x}, \underline{\sigma}, f_0(\mu) = k) \;\propto\; \exp\left[-\frac{-2\,\overline{x}\,\mu + \mu^2}{2\,\sigma_C^2}\right]$$

$$\propto\; \exp\left[-\frac{\overline{x}^2 - 2\,\overline{x}\,\mu + \mu^2}{2\,\sigma_C^2}\right]$$

Inferring μ from n 'independent' measurements

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \underline{\sigma}, f_0(\mu) = k) &\propto \exp\left[-\frac{-2\,\overline{x}\,\mu + \mu^2}{2\,\sigma_C^2}\right] \\
&\propto \exp\left[-\frac{\overline{x}^2 - 2\,\overline{x}\,\mu + \mu^2}{2\,\sigma_C^2}\right] \\
&\propto \exp\left[-\frac{(\mu - \overline{x})^2}{2\,\sigma_C^2}\right]
\end{aligned}
$$

(having used the technique of complementing the exponential) with

$$
\begin{aligned}
\overline{x} &= \frac{\sum_i x_i/\sigma_i^2}{\sum_i 1/\sigma_i^2} \\
\overline{x^2} &= \frac{\sum_i x_i^2/\sigma_i^2}{\sum_i 1/\sigma_i^2} \\
\sigma_C^2 &= \frac{1}{\sum_i 1/\sigma_i^2}
\end{aligned}
$$

Yes, but the prior?

Yes, but the prior?

Remember that (writing σ_e again)

$$f(\mu\,|\,x_1\,\sigma_e) \propto f(x_1\,|\,\mu,\sigma_e)\cdot f_0(\mu)$$

Yes, but the prior?

Remember that (writing σ_e again)

$$f(\mu \,|\, x_1\, \sigma_e) \propto f(x_1 \,|\, \mu, \sigma_e) \cdot f_0(\mu)$$

- The first factor in the r.h.s. ('likelihood') prefers a region a few σ_e's around x_1.

Yes, but the prior?

Remember that (writing σ_e again)

$$f(\mu \,|\, x_1\, \sigma_e) \propto f(x_1 \,|\, \mu, \sigma_e) \cdot f_0(\mu)$$

- The first factor in the r.h.s. ('likelihood') prefers a region a few σ_e's around x_1.
- If $f_0(\mu)$ is 'practically flat' in that region, then it is irrelevant.

Yes, but the prior?

Remember that (writing σ_e again)

$$f(\mu \,|\, x_1\, \sigma_e) \propto f(x_1 \,|\, \mu, \sigma_e) \cdot f_0(\mu)$$

- The first factor in the r.h.s. ('likelihood') prefers a region a few σ_e's around x_1.
- If $f_0(\mu)$ is 'practically flat' in that region, then it is irrelevant.
- Otherwise model it at best and do the math (e.g. by MCMC).

Yes, but the prior?

Remember that (writing σ_e again)

$$f(\mu \,|\, x_1\, \sigma_e) \propto f(x_1 \,|\, \mu, \sigma_e) \cdot f_0(\mu)$$

- The first factor in the r.h.s. ('likelihood') prefers a region a few σ_e's around x_1.
- If $f_0(\mu)$ is 'practically flat' in that region, then it is irrelevant.
- Otherwise model it at best and do the math (e.g. by MCMC).
- And, please, **remember Gauss** (well aware of the limitations)

Yes, but the prior?

Remember that (writing σ_e again)

$$f(\mu \,|\, x_1\, \sigma_e) \propto f(x_1 \,|\, \mu, \sigma_e) \cdot f_0(\mu)$$

- The first factor in the r.h.s. ('likelihood') prefers a region a few σ_e's around x_1.
- If $f_0(\mu)$ is 'practically flat' in that region, then it is irrelevant.
- Otherwise model it at best and do the math (e.g. by MCMC).
- And, please, **remember Gauss** (well aware of the limitations)
  ... and that
  "All models are wrong, but some are useful"
  (G. Box)

Yes, but the prior?

Remember that (writing σ_e again)

$$f(\mu \,|\, x_1\, \sigma_e) \propto f(x_1 \,|\, \mu, \sigma_e) \cdot f_0(\mu)$$

- The first factor in the r.h.s. ('likelihood') prefers a region a few σ_e's around x_1.
- If $f_0(\mu)$ is 'practically flat' in that region, then it is irrelevant.
- Otherwise model it at best and do the math (e.g. by MCMC).
- And, please, **remember Gauss** (well aware of the limitations)
  ...and that
  "All models are wrong, but some are useful"
  (G. Box)

  And Gauss was the first to realize that
  the Gaussian is indeed 'wrong'!

Use of a conjugate prior

As we have already, a ‘trick’ developped in order to simplify the calculations is the use of conjugate priors:

Use of a conjugate prior

As we have already, a ‘trick’ developped in order to simplify the calculations is the use of conjugate priors:

- Binomial distribution: Beta distribution.
- Poisson distribution: Gamma distribution.

Use of a conjugate prior

As we have already, a 'trick' developped in order to simplify the calculations is the use of conjugate priors:

- Binomial distribution: Beta distribution.
- Poisson distribution: Gamma distribution.
- Gaussian distribution: Gaussian distribution.

Use of a conjugate prior

As we have already, a 'trick' developped in order to simplify the calculations is the use of conjugate priors:

- Binomial distribution: Beta distribution.
- Poisson distribution: Gamma distribution.
- Gaussian distribution: Gaussian distribution.

Not very flexible,

Use of a conjugate prior

As we have already, a ‘trick’ developped in order to simplify the calculations is the use of conjugate priors:

Binomial distribution: Beta distribution.

Poisson distribution: Gamma distribution.

Gaussian distribution: Gaussian distribution.

Not very flexible, but a a good starting point

- in order to avoid to blindly accept ‘crazy results’

Use of a conjugate prior

As we have already, a 'trick' developped in order to simplify the calculations is the use of conjugate priors:

Binomial distribution: Beta distribution.

Poisson distribution: Gamma distribution.

Gaussian distribution: Gaussian distribution.

Not very flexible, but a a good starting point

- in order to avoid to blindly accept 'crazy results' (imagine someone mixing up orders of magnitudes);

Use of a conjugate prior

As we have already, a 'trick' developped in order to simplify the calculations is the use of conjugate priors:

Binomial distribution: Beta distribution.

Poisson distribution: Gamma distribution.

Gaussian distribution: Gaussian distribution.

Not very flexible, but a a good starting point

- in order to avoid to blindly accept 'crazy results' (imagine someone mixing up orders of magnitudes);
- our prior could come from a previous experimental result, providing a Gaussian pdf about 'μ';

Use of a conjugate prior

As we have already, a 'trick' developped in order to simplify the calculations is the use of conjugate priors:

Binomial distribution: Beta distribution.

Poisson distribution: Gamma distribution.

Gaussian distribution: Gaussian distribution.

Not very flexible, but a a good starting point

- in order to avoid to blindly accept 'crazy results' (imagine someone mixing up orders of magnitudes);
- our prior could come from a previous experimental result, providing a Gaussian pdf about 'μ';
- making the Gaussian prior broad enough we can recover a $\approx$ uniform distribution.

Use of a conjugate prior

As we have already, a 'trick' developped in order to simplify the calculations is the use of conjugate priors:

Binomial distribution: Beta distribution.

Poisson distribution: Gamma distribution.

Gaussian distribution: Gaussian distribution.

Not very flexible, but a a good starting point

- in order to avoid to blindly accept 'crazy results' (imagine someone mixing up orders of magnitudes);
- our prior could come from a previous experimental result, providing a Gaussian pdf about 'μ';
- making the Gaussian prior broad enough we can recover a $\approx$ uniform distribution.

From the second item it is clear that we can make use of what we have developed in the previous slides for 'combinations'.

Use of a conjugate prior

As we have already, a 'trick' developped in order to simplify the calculations is the use of conjugate priors:

Binomial distribution: Beta distribution.

Poisson distribution: Gamma distribution.

Gaussian distribution: Gaussian distribution.

Not very flexible, but a a good starting point

- in order to avoid to blindly accept 'crazy results' (imagine someone mixing up orders of magnitudes);
- our prior could come from a previous experimental result, providing a Gaussian pdf about 'μ';
- making the Gaussian prior broad enough we can recover a $\approx$ uniform distribution.

From the second item it is clear that we can make use of what we have developed in the previous slides for 'combinations'. (Just measurement nr 1, or nr 0, if you prefer...)

Adding a prior piece of information

$$f(\mu\,|\,\underline{x},\underline{\sigma}) \;\propto\; \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]\cdot f_0(\mu)$$

Adding a prior piece of information

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \underline{\sigma}) &\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \cdot f_0(\mu) \\
&\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right]
\end{aligned}
$$

Adding a prior piece of information

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \underline{\sigma}) &\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \cdot f_0(\mu) \\
&\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right] \\
&\propto \prod_{i=0}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right].
\end{aligned}
$$

Adding a prior piece of information

$$
\begin{aligned}
f(\mu\,|\,\underline{x},\underline{\sigma}) &\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]\cdot f_0(\mu) \\
&\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]\cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right] \\
&\propto \prod_{i=0}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]\cdot \\
&\propto \exp\left[-\frac{(\mu-\overline{x})^2}{2\,\sigma_C^2}\right]\cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right]
\end{aligned}
$$

Adding a prior piece of information

$$
\begin{aligned}
f(\mu\,|\,\underline{x},\underline{\sigma}) &\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]\cdot f_0(\mu) \\
&\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]\cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right] \\
&\propto \prod_{i=0}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]\cdot \\
&\propto \exp\left[-\frac{(\mu-\overline{x})^2}{2\,\sigma_C^2}\right]\cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right] \\
&\propto \exp\left[-\frac{(\mu-\overline{x}_p)^2}{2\,\sigma_{C_p}^2}\right]
\end{aligned}
$$

Adding a prior piece of information

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \underline{\sigma}) &\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]\cdot f_0(\mu) \\
&\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]\cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right] \\
&\propto \prod_{i=0}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right]\cdot \\
&\propto \exp\left[-\frac{(\mu-\overline{x})^2}{2\,\sigma_C^2}\right]\cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right] \\
&\propto \exp\left[-\frac{(\mu-\overline{x}_p)^2}{2\,\sigma_{C_p}^2}\right]
\end{aligned}
$$

with

$$
\overline{x}_p = \frac{\overline{x}/\sigma_C^2 + x_0/\sigma_0^2}{1/\sigma_C^2 + 1/\sigma_0^2}\,;
$$

Adding a prior piece of information

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \underline{\sigma}) &\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \cdot f_0(\mu) \\
&\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right] \\
&\propto \prod_{i=0}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \cdot \\
&\propto \exp\left[-\frac{(\mu-\overline{x})^2}{2\,\sigma_C^2}\right] \cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right] \\
&\propto \exp\left[-\frac{(\mu-\overline{x}_p)^2}{2\,\sigma_{C_p}^2}\right]
\end{aligned}
$$

with

$$
\overline{x}_p = \frac{\overline{x}/\sigma_C^2 + x_0/\sigma_0^2}{1/\sigma_C^2 + 1/\sigma_0^2}\,; \qquad \frac{1}{\sigma_{C_p}^2} = \frac{1}{\sigma_C^2} + \frac{1}{\sigma_0^2}
$$

Adding a prior piece of information

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \underline{\sigma}) &\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \cdot f_0(\mu) \\
&\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right] \\
&\propto \prod_{i=0}^{n} \exp\left[-\frac{(x_i-\mu)^2}{2\,\sigma_i^2}\right] \cdot \\
&\propto \exp\left[-\frac{(\mu-\overline{x})^2}{2\,\sigma_C^2}\right] \cdot \exp\left[-\frac{(\mu-x_0)^2}{2\,\sigma_0^2}\right] \\
&\propto \exp\left[-\frac{(\mu-\overline{x}_p)^2}{2\,\sigma_{C_p}^2}\right]
\end{aligned}
$$

with

$$
\overline{x}_p = \frac{\overline{x}/\sigma_C^2 + x_0/\sigma_0^2}{1/\sigma_C^2 + 1/\sigma_0^2}\,; \qquad \frac{1}{\sigma_{C_p}^2} = \frac{1}{\sigma_C^2} + \frac{1}{\sigma_0^2}
$$

$\sigma_0^2 \gg \sigma_C^2 \quad \Rightarrow$ irrelevant ($\approx$ 'uniform').

Joint inference of μ and τ ($\rightarrow \sigma$) with JAGS/rjags

Model (to be written in the model file)

```
model{
    for (i in 1:length(x)) {
      x[i] ~ dnorm(mu, tau);
    }
    mu ~ dnorm(0.0, 1.0E-6);
    tau ~ dgamma(1.0, 1.0E-6);
    sigma <- 1.0/sqrt(tau);
}
```

Joint inference of μ and τ ($\rightarrow \sigma$) with JAGS/rjags

Model (to be written in the model file)

```
model{
    for (i in 1:length(x)) {
      x[i] ~ dnorm(mu, tau);
    }
    mu ~ dnorm(0.0, 1.0E-6);
    tau ~ dgamma(1.0, 1.0E-6);
    sigma <- 1.0/sqrt(tau);
}
```

Simulated data

```
mu.true = 3; sigma.true = 2; sample.n = 20
x = rnorm(sample.n, mu.true, sigma.true)
```

Joint inference of μ and τ ($\rightarrow \sigma$) with JAGS/rjags

Model (to be written in the model file)

```
model{
    for (i in 1:length(x)) {
      x[i] ~ dnorm(mu, tau);
    }
    mu ~ dnorm(0.0, 1.0E-6);
    tau ~ dgamma(1.0, 1.0E-6);
    sigma <- 1.0/sqrt(tau);
}
```

Simulated data

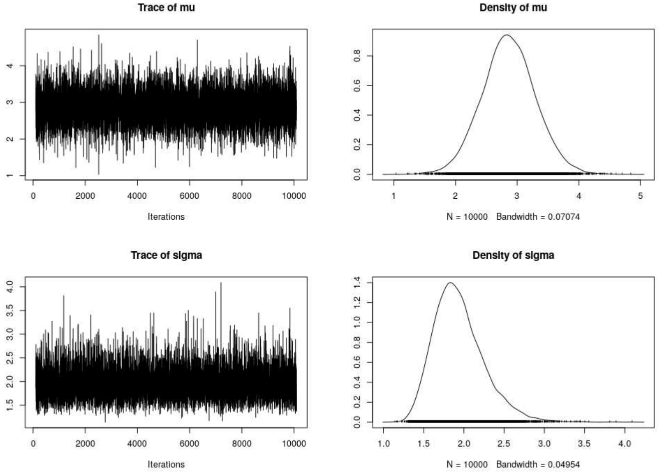
```
mu.true = 3; sigma.true = 2; sample.n = 20
x = rnorm(sample.n, mu.true, sigma.true)
```

JAGS calls

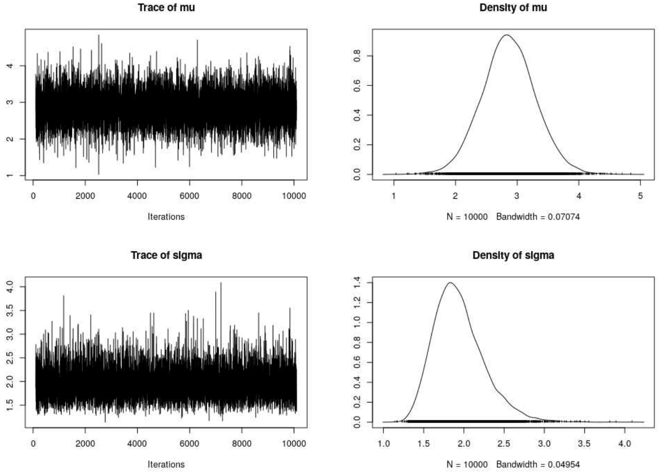
```
data = list(x=x)
inits = list(mu=mean(x), tau=1/var(x))
jm <- jags.model(model, data, inits)
update(jm, 100)
chain <- coda.samples(jm, c("mu","sigma"), n.iter=10000)
```

Joint inference of μ and τ ($\rightarrow \sigma$) with JAGS/rjags

$\Rightarrow$ inf_mu_sigma.R



Joint inference of μ and τ ($\rightarrow \sigma$) with JAGS/rjags
$\Rightarrow$ inf_mu_sigma.R



$\overline{\text{mu}} = 2.87$, std(mu) $= 0.44$; $\overline{\text{sigma}} = 1.94$, std(sigma) $= 0.31$

Proposed exercise

Try to reproduce the results of `inf_mu_sigma.R`
by a self-written Metropolis algorithm

Remark: the priors about μ and σ (there is no need to use τ) can be simply uniform, but, obviously, the unnormalized joint distribution has to return 0 for $\sigma \leq 0$, such that the MC chain cannot make a jump into such uphysical region.

The End