Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

A simple random walk on a plane

```
> prop.sq <- function(xy.o, d) xy.o + runif(2,-d, d)
```

A simple random walk on a plane

```
> prop.sq <- function(xy.o, d) xy.o + runif(2,-d, d)
> xy <- c(0,0)
```

A simple random walk on a plane

```
> prop.sq <- function(xy.o, d) xy.o + runif(2,-d, d)
> xy <- c(0,0)
> ( xy <- prop.sq(xy, 1) )
 [1] 0.4555056 -0.3119171
```

A simple random walk on a plane

```
> prop.sq <- function(xy.o, d) xy.o + runif(2,-d, d)
> xy <- c(0,0)
> ( xy <- prop.sq(xy, 1) )
[1] 0.4555056 -0.3119171
> ( xy <- prop.sq(xy, 1) )
[1] 0.1603211 -0.2819019
```

A simple random walk on a plane

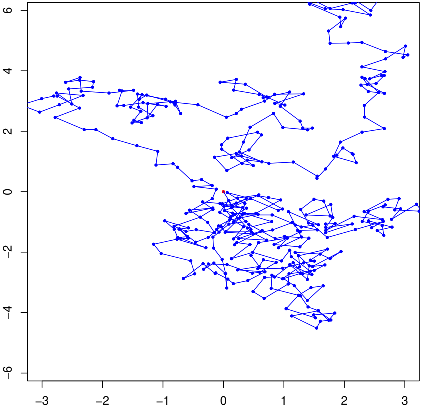
```
> prop.sq <- function(xy.o, d) xy.o + runif(2,-d, d)
> xy <- c(0,0)
> ( xy <- prop.sq(xy, 1) )
[1] 0.4555056 -0.3119171
> ( xy <- prop.sq(xy, 1) )
[1] 0.1603211 -0.2819019
```

⇒ mcmc_unbound.R

A simple random walk on a plane

```
> prop.sq <- function(xy.o, d) xy.o + runif(2,-d, d)
> xy <- c(0,0)
> ( xy <- prop.sq(xy, 1) )
 [1] 0.4555056 -0.3119171
> ( xy <- prop.sq(xy, 1) )
 [1] 0.1603211 -0.2819019
```

⇒ mcmc_unbound.R



Visiting uniformly the points inside a square

- How can we modify the previous example in order that the visited points are uniform inside a square?

Visiting uniformly the points inside a square

- How can we modify the previous example in order that the visited points are uniform inside a square?
- That is we would like to get something similar to

```
> x <- runif(n, -1, 1)
> y <- runif(n, -1, 1)
```

Visiting uniformly the points inside a square

- How can we modify the previous example in order that the visited points are uniform inside a square?
- That is we would like to get something similar to

```
> x <- runif(n, -1, 1)
> y <- runif(n, -1, 1)
```

- How can we bound the random walk to force it to remain inside the square?

Visiting uniformly the points inside a square

- How can we modify the previous example in order that the visited points are uniform inside a square?
- That is we would like to get something similar to

```
> x <- runif(n, -1, 1)
> y <- runif(n, -1, 1)
```

- How can we bound the random walk to force it to remain inside the square?

  $\rightarrow$ Reject the move that leads out of the square.

Visiting uniformly the points inside a square

- How can we modify the previous example in order that the visited points are uniform inside a square?
- That is we would like to get something similar to

```
> x <- runif(n, -1, 1)
> y <- runif(n, -1, 1)
```

- How can we bound the random walk to force it to remain inside the square?

  $\rightarrow$ Reject the move that leads out of the square.

  - Obvious. . .
  - but how to record the points from which there was a rejection?

Visiting uniformly the points inside a square

- How can we modify the previous example in order that the visited points are uniform inside a square?
- That is we would like to get something similar to

```
> x <- runif(n, -1, 1)
> y <- runif(n, -1, 1)
```

- How can we bound the random walk to force it to remain inside the square?

  → Reject the move that leads out of the square.

  - Obvious...
  - but how to record the points from which there was a rejection?

  1. Just count the point once, until the 'chain' moves.

Visiting uniformly the points inside a square

- How can we modify the previous example in order that the visited points are uniform inside a square?
- That is we would like to get something similar to

```
> x <- runif(n, -1, 1)
> y <- runif(n, -1, 1)
```

- How can we bound the random walk to force it to remain inside the square?

  → Reject the move that leads out of the square.
  - Obvious...
  - but how to record the points from which there was a rejection?

  1. Just count the point once, until the 'chain' moves.
  2. Count it many time we "move to it"
     (a rejection can be seen as a move to the same position!).

Visiting uniformly the points inside a square

- How can we modify the previous example in order that the visited points are uniform inside a square?
- That is we would like to get something similar to

```
> x <- runif(n, -1, 1)
> y <- runif(n, -1, 1)
```

- How can we bound the random walk to force it to remain inside the square?

  → Reject the move that leads out of the square.

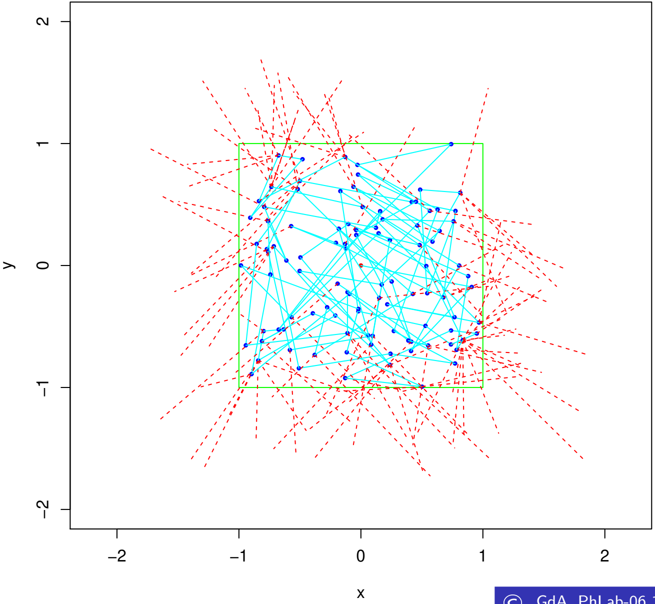
  - Obvious...
  - but how to record the points from which there was a rejection?

  1. Just count the point once, until the 'chain' moves.
  2. Count it many time we "move to it"
     (a rejection can be seen as a move to the same position!).

  ⇒ mcmc_square.R

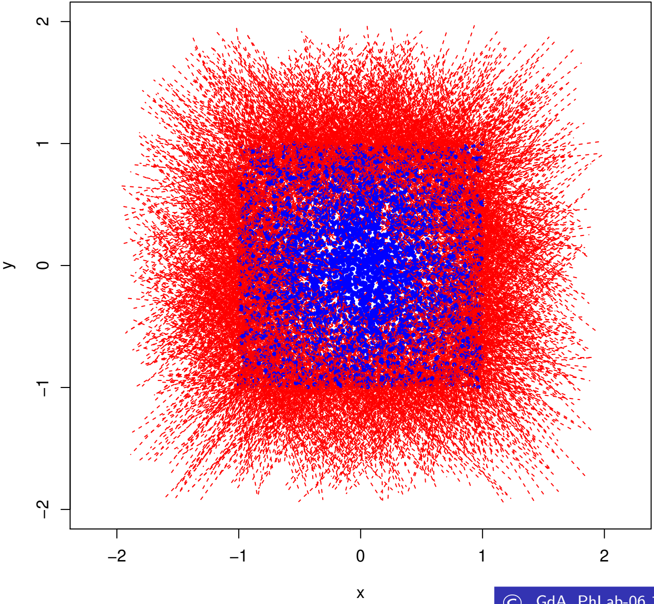
Visiting uniformly the points inside a square

200 attempts



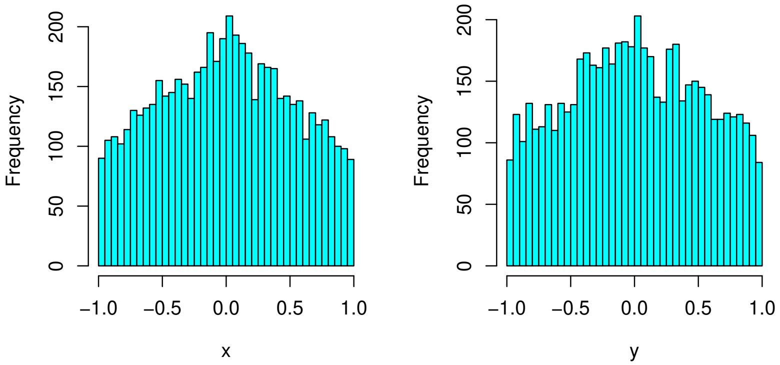
Visiting uniformly the points inside a square

10000 attempts (intermediate cyan 'lines' omitted)



Visiting uniformly the points inside a square

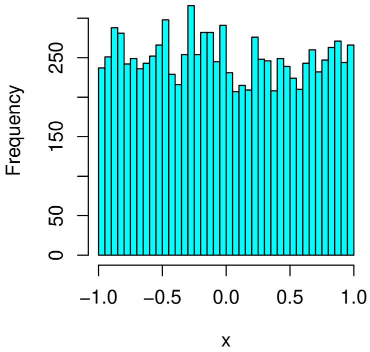
Method 1: count points only once



Not what we wanted!

Visiting uniformly the points inside a square

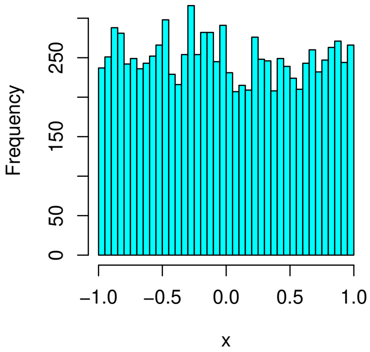
Method 2: count points several times



$\surd$

Visiting uniformly the points inside a square

Method 2: count points several times

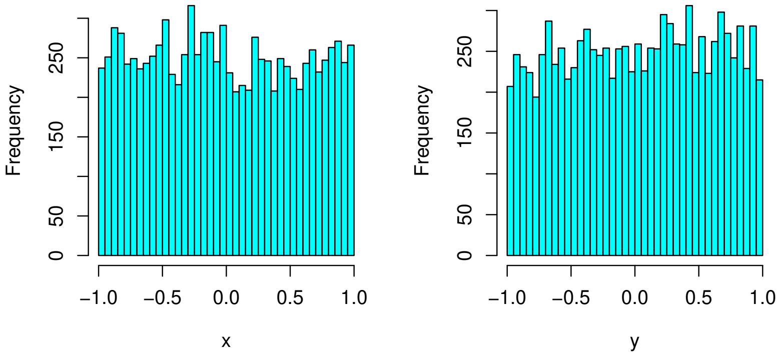


$\surd$

Why?

Visiting uniformly the points inside a square

Method 2: count points several times



√ **Why?**

MCMC theory left to self study

Inferring μ and σ from n 'independent' measurements

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

Inferring μ and σ from n 'independent' measurements

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

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

Inferring μ and σ from n 'independent' measurements

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

Inferring μ and σ from n 'independent' measurements

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

Inferring μ and σ from n 'independent' measurements

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

But this time we cannot use the trick of complementing the exponential, because $\overline{x^2}$ is devided by $2\,\sigma^2/n$

Inferring μ and σ from n 'independent' measurements

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

But this time we cannot use the trick of complementing the exponential, because $\overline{x^2}$ is devided by $2\,\sigma^2/n$

Nevertheless, we can add and remove $\overline{x}^2$, thus getting in the numerator of the exponent

$$\overline{x^2}-\overline{x}^2+\overline{x}^2-2\,\overline{x}\,\mu+\mu^2$$

Inferring μ and σ from n 'independent' measurements

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

But this time we cannot use the trick of complementing the exponential, because $\overline{x^2}$ is devided by $2\,\sigma^2/n$

Nevertheless, we can add and remove $\overline{x}^2$, thus getting in the numerator of the exponent

$$\overline{x^2}-\overline{x}^2+\overline{x}^2-2\,\overline{x}\,\mu+\mu^2=\overline{x^2}-\overline{x}^2+(\overline{x}-\mu)^2$$

Inferring μ and σ from n 'independent' measurements

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

But this time we cannot use the trick of complementing the exponential, because $\overline{x^2}$ is devided by $2\,\sigma^2/n$

Nevertheless, we can add and remove $\overline{x}^2$, thus getting in the numerator of the exponent

$$\overline{x^2}-\overline{x}^2+\overline{x}^2-2\,\overline{x}\,\mu+\mu^2=\overline{x^2}-\overline{x}^2+(\overline{x}-\mu)^2=s^2+(\overline{x}-\mu)^2$$

Inferring μ and σ from n 'independent' measurements

We have then

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

with s^2 equal to the sample variance.

Inferring μ and σ from n 'independent' measurements

We have then

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

with s^2 equal to the sample variance.

That is

$$f(\mu, \sigma \,|\, \underline{x}, f_0(\mu,\sigma)) \quad = \quad f(\mu, \sigma \,|\, \overline{x}, s, f_0(\mu,\sigma))$$

Inferring μ and σ from n 'independent' measurements

We have then

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

with s^2 equal to the sample variance.

That is

$$f(\mu,\sigma\,|\,\underline{x},f_0(\mu,\sigma)) \;=\; f(\mu,\sigma\,|\,\overline{x},s,f_0(\mu,\sigma))$$

$\Rightarrow$ The inference does not depend on the detailed knowledge of $\underline{x}$

Inferring μ and σ from n 'independent' measurements

We have then

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

with s^2 equal to the sample variance.

That is

$$f(\mu,\sigma\,|\,\underline{x}, f_0(\mu,\sigma)) \quad = \quad f(\mu,\sigma\,|\,\overline{x}, s, f_0(\mu,\sigma))$$

$\Rightarrow$ The inference does not depend on the detailed knowledge of $\underline{x}$, but rather only on $\overline{x}$ and s^2.

Inferring μ and σ from n ‘independent’ measurements

We have then

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

with s^2 equal to the sample variance.

That is

$$f(\mu,\sigma \,|\, \underline{x}, f_0(\mu,\sigma)) \quad = \quad f(\mu,\sigma \,|\, \overline{x}, s, f_0(\mu,\sigma))$$

⇒ The inference does not depend on the detailed knowledge of $\underline{x}$, but rather only on $\overline{x}$ and s^2.

⇒ statistical sufficiency

Inferring μ and σ from n 'independent' measurements

Adding the prediction of x_f

Inferring μ and σ from n 'independent' measurements

Adding the prediction of x_f

$$f(x_f, \mu, \sigma \,|\, \ldots) \;=\; f(x_f \,|\, \mu, \sigma, \ldots) \cdot f(\mu, \sigma \,|\, \ldots)$$

Inferring μ and σ from n 'independent' measurements

Adding the prediction of x_f

$$f(x_f, \mu, \sigma \,|\, \ldots) \;=\; f(x_f \,|\, \mu, \sigma, \ldots) \cdot f(\mu, \sigma \,|\, \ldots)$$

that is

$$f(x_f, \mu, \sigma \,|\, \ldots) \;\propto\; \sigma^{-1} \exp\left[-\frac{(x_f - \mu)^2}{2\,\sigma^2}\right] \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right]$$

Inferring μ and σ from n 'independent' measurements

Adding the prediction of x_f

$$f(x_f, \mu, \sigma \,|\, \ldots) \;=\; f(x_f \,|\, \mu, \sigma, \ldots) \cdot f(\mu, \sigma \,|\, \ldots)$$

that is

$$\begin{aligned} f(x_f, \mu, \sigma \,|\, \ldots) &\propto \sigma^{-1} \exp\left[-\frac{(x_f - \mu)^2}{2\,\sigma^2}\right] \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \\ &\propto \sigma^{-(n+1)} \exp\left[-\frac{(x_f - \mu)^2}{2\,\sigma^2}\right] \cdot \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \end{aligned}$$

Inferring μ and σ from n 'independent' measurements

Adding the prediction of x_f

$$f(x_f, \mu, \sigma \,|\, \ldots) \;=\; f(x_f \,|\, \mu, \sigma, \ldots) \cdot f(\mu, \sigma \,|\, \ldots)$$

that is

$$\begin{aligned} f(x_f, \mu, \sigma \,|\, \ldots) &\propto \sigma^{-1} \exp\left[-\frac{(x_f - \mu)^2}{2\,\sigma^2}\right] \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \\ &\propto \sigma^{-(n+1)} \exp\left[-\frac{(x_f - \mu)^2}{2\,\sigma^2}\right] \cdot \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \end{aligned}$$

that we can extend it to $f(\overline{x}_f, \mu, \sigma \,|\, \ldots)$

Inferring μ and σ from n 'independent' measurements

Adding the prediction of x_f

$$f(x_f, \mu, \sigma \,|\, \ldots) \;=\; f(x_f \,|\, \mu, \sigma, \ldots) \cdot f(\mu, \sigma \,|\, \ldots)$$

that is

$$\begin{aligned} f(x_f, \mu, \sigma \,|\, \ldots) &\propto \sigma^{-1} \exp\left[-\frac{(x_f-\mu)^2}{2\,\sigma^2}\right] \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \\ &\propto \sigma^{-(n+1)} \exp\left[-\frac{(x_f-\mu)^2}{2\,\sigma^2}\right] \cdot \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \end{aligned}$$

that we can extend it to $f(\overline{x}_f, \mu, \sigma \,|\, \ldots)$, that is

$$f(\overline{x}_f, \mu, \sigma \,|\, \ldots) \;\propto\; \sigma^{-(n+1)} \exp\left[-\frac{(x_f-\mu)^2}{2\,\sigma^2/n}\right] \cdot \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right]$$

Inferring μ and σ from n 'independent' measurements

Adding the prediction of x_f

$$f(x_f,\mu,\sigma\,|\,\ldots) \;=\; f(x_f\,|\,\mu,\sigma,\ldots)\cdot f(\mu,\sigma\,|\,\ldots)$$

that is

$$\begin{aligned} f(x_f,\mu,\sigma\,|\,\ldots) &\propto \sigma^{-1}\exp\left[-\frac{(x_f-\mu)^2}{2\,\sigma^2}\right]\sigma^{-n}\exp\left[-\frac{s^2+(\mu-\overline{x})^2}{2\,\sigma^2/n}\right] \\ &\propto \sigma^{-(n+1)}\exp\left[-\frac{(x_f-\mu)^2}{2\,\sigma^2}\right]\cdot\exp\left[-\frac{s^2+(\mu-\overline{x})^2}{2\,\sigma^2/n}\right] \end{aligned}$$

that we can extend it to $f(\overline{x}_f,\mu,\sigma\,|\,\ldots)$, that is

$$f(\overline{x}_f,\mu,\sigma\,|\,\ldots) \;\propto\; \sigma^{-(n+1)}\exp\left[-\frac{(x_f-\mu)^2}{2\,\sigma^2/n}\right]\cdot\exp\left[-\frac{s^2+(\mu-\overline{x})^2}{2\,\sigma^2/n}\right]$$

Try to do it with Metropolis!

Back to $f(\mu, \sigma)$ from a sample

In practice

$$f(\mu, \sigma \,|\, \overline{x}, s, \ldots) \quad \propto \quad \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

Back to $f(\mu, \sigma)$ from a sample

In practice

$$f(\mu, \sigma \,|\, \overline{x}, s, \ldots) \;\;\propto\;\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

Then

$$f(\mu \,|\, \overline{x}, s, \ldots) \;\;=\;\; \int_0^\infty f(\mu, \sigma \,|\, \overline{x}, s)\, d\sigma$$

Back to $f(\mu, \sigma)$ from a sample

In practice

$$f(\mu, \sigma \,|\, \overline{x}, s, \ldots) \;\propto\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

Then

$$f(\mu \,|\, \overline{x}, s, \ldots) \;=\; \int_0^{\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\sigma$$

$$f(\sigma \,|\, \overline{x}, s, \ldots) \;=\; \int_{-\infty}^{+\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\mu$$

Back to $f(\mu,\sigma)$ from a sample

In practice

$$f(\mu,\sigma\,|\,\overline{x},s,\ldots) \;\propto\; \sigma^{-n}\exp\left[-\frac{s^2+(\mu-\overline{x})^2}{2\,\sigma^2/n}\right]\cdot f_0(\mu,\sigma)$$

Then

$$f(\mu\,|\,\overline{x},s,\ldots) \;=\; \int_0^{\infty} f(\mu,\sigma\,|\,\overline{x},s)\,d\sigma$$

$$f(\sigma\,|\,\overline{x},s,\ldots) \;=\; \int_{-\infty}^{+\infty} f(\mu,\sigma\,|\,\overline{x},s)\,d\mu$$

Some remarks are in order:

Back to $f(\mu, \sigma)$ from a sample

In practice

$$f(\mu, \sigma \,|\, \overline{x}, s, \ldots) \propto \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

Then

$$f(\mu \,|\, \overline{x}, s, \ldots) = \int_0^{\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\sigma$$

$$f(\sigma \,|\, \overline{x}, s, \ldots) = \int_{-\infty}^{+\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\mu$$

Some remarks are in order:

- $f(\mu \,|\, \overline{x}, s)$ is in general not Gaussian (**not even starting from a flat prior!**)

Back to $f(\mu, \sigma)$ from a sample

In practice

$$f(\mu, \sigma \,|\, \overline{x}, s, \ldots) \;\propto\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

Then

$$f(\mu \,|\, \overline{x}, s, \ldots) \;=\; \int_0^{\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\sigma$$

$$f(\sigma \,|\, \overline{x}, s, \ldots) \;=\; \int_{-\infty}^{+\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\mu$$

Some remarks are in order:

- $f(\mu \,|\, \overline{x}, s)$ is in general not Gaussian (**not even starting from a flat prior!**) due to the uncertainty on σ

Back to $f(\mu, \sigma)$ from a sample

In practice

$$f(\mu, \sigma \,|\, \overline{x}, s, \ldots) \;\propto\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

Then

$$f(\mu \,|\, \overline{x}, s, \ldots) \;=\; \int_0^{\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\sigma$$

$$f(\sigma \,|\, \overline{x}, s, \ldots) \;=\; \int_{-\infty}^{+\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\mu$$

Some remarks are in order:

- $f(\mu \,|\, \overline{x}, s)$ is in general not Gaussian (**not even starting from a flat prior!**) due to the uncertainty on σ ('convolution over all possible values')

Back to $f(\mu, \sigma)$ from a sample

In practice

$$f(\mu, \sigma \,|\, \overline{x}, s, \ldots) \;\propto\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

Then

$$f(\mu \,|\, \overline{x}, s, \ldots) \;=\; \int_0^{\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\sigma$$

$$f(\sigma \,|\, \overline{x}, s, \ldots) \;=\; \int_{-\infty}^{+\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\mu$$

Some remarks are in order:

- $f(\mu \,|\, \overline{x}, s)$ is in general not Gaussian (**not even starting from a flat prior!**) due to the uncertainty on σ ('convolution over all possible values')
- It tends to Gaussian when 'σ is precisely measured'

Back to $f(\mu, \sigma)$ from a sample

In practice

$$f(\mu, \sigma \,|\, \overline{x}, s, \ldots) \;\propto\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

Then

$$f(\mu \,|\, \overline{x}, s, \ldots) \;=\; \int_0^{\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\sigma$$

$$f(\sigma \,|\, \overline{x}, s, \ldots) \;=\; \int_{-\infty}^{+\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\mu$$

Some remarks are in order:

- $f(\mu \,|\, \overline{x}, s)$ is in general not Gaussian (**not even starting from a flat prior!**) due to the uncertainty on σ ('convolution over all possible values')
- It tends to Gaussian when 'σ is precisely measured'

$$\Rightarrow n \rightarrow \infty$$

Joint inference of μ and σ from a sample

Large sample behaviour starting from uniform priors$^{(*)}$
(with '**std**' for **standard deviation** to avoid confusion with unkown σ)

Joint inference of μ and σ from a sample

Large sample behaviour starting from uniform priors$^{(*)}$
(with '**std**' for **standard deviation** to avoid confusion with unkown σ)

$$\mathrm{E}(\mu) \xrightarrow{'n\rightarrow\infty'} \overline{x}$$

$$\mathrm{std}(\mu) \xrightarrow{n\rightarrow\infty} \frac{s}{\sqrt{n}}$$

$$\mu \xrightarrow{n\rightarrow\infty} \sim \mathcal{N}(\overline{x}, \frac{s}{\sqrt{n}})$$

Joint inference of μ and σ from a sample

Large sample behaviour starting from uniform priors[(*)]
(with '**std**' for **standard deviation** to avoid confusion with unkown σ)

$$
\begin{aligned}
\mathrm{E}(\mu) &\xrightarrow{'n\to\infty'} \overline{x} \\
\mathrm{std}(\mu) &\xrightarrow{n\to\infty} \frac{s}{\sqrt{n}} \\
\mu &\xrightarrow{n\to\infty} \sim \mathcal{N}(\overline{x}, \frac{s}{\sqrt{n}}) \\
\mathrm{E}(\sigma) &\xrightarrow[n\to\infty]{} s \\
\mathrm{std}(\sigma) &\xrightarrow[n\to\infty]{} \frac{s}{\sqrt{2\,n}} \\
\sigma &\xrightarrow[n\to\infty]{} \sim \mathcal{N}(s, \frac{s}{\sqrt{2\,n}})
\end{aligned}
$$

Joint inference of μ and σ from a sample

Large sample behaviour starting from uniform priors[(*)]
(with '**std**' for **standard deviation** to avoid confusion with unkown σ)

$$
\begin{aligned}
E(\mu) &\xrightarrow{'n\to\infty'} \overline{x} \\
\text{std}(\mu) &\xrightarrow{n\to\infty} \frac{s}{\sqrt{n}} \\
\mu &\xrightarrow{n\to\infty} \sim \mathcal{N}(\overline{x}, \frac{s}{\sqrt{n}}) \\
E(\sigma) &\xrightarrow[n\to\infty]{} s \\
\text{std}(\sigma) &\xrightarrow[n\to\infty]{} \frac{s}{\sqrt{2\,n}} \\
\sigma &\xrightarrow[n\to\infty]{} \sim \mathcal{N}(s, \frac{s}{\sqrt{2\,n}})
\end{aligned}
$$

(*) The most sensitive is the prior on σ

Joint inference of μ and σ from a sample

Large sample behaviour starting from uniform priors$^{(*)}$
(with '**std**' for **standard deviation** to avoid confusion with unkown σ)

$$
\begin{aligned}
E(\mu) &\xrightarrow{'n\to\infty'} \overline{x} \\
\text{std}(\mu) &\xrightarrow{n\to\infty} \frac{s}{\sqrt{n}} \\
\mu &\xrightarrow{n\to\infty} \sim \mathcal{N}(\overline{x}, \frac{s}{\sqrt{n}}) \\
E(\sigma) &\xrightarrow[n\to\infty]{} s \\
\text{std}(\sigma) &\xrightarrow[n\to\infty]{} \frac{s}{\sqrt{2\,n}} \\
\sigma &\xrightarrow[n\to\infty]{} \sim \mathcal{N}(s, \frac{s}{\sqrt{2\,n}})
\end{aligned}
$$

$^{(*)}$ The most sensitive is the prior on $\sigma \Rightarrow$ it induces abstract speculations in mathematicians and statisticians who often have little idea of what they are talking about

Joint inference of μ and σ from a sample

Large sample behaviour starting from uniform priors$^{(*)}$
(with '**std**' for **standard deviation** to avoid confusion with unkown σ)

$$
\begin{aligned}
\mathrm{E}(\mu) &\xrightarrow{'n\rightarrow\infty'} \overline{x} \\
\mathrm{std}(\mu) &\xrightarrow{n\rightarrow\infty} \frac{s}{\sqrt{n}} \\
\mu &\xrightarrow{n\rightarrow\infty} \sim \mathcal{N}(\overline{x}, \frac{s}{\sqrt{n}}) \\
\mathrm{E}(\sigma) &\xrightarrow[n\rightarrow\infty]{} s \\
\mathrm{std}(\sigma) &\xrightarrow[n\rightarrow\infty]{} \frac{s}{\sqrt{2\,n}} \\
\sigma &\xrightarrow[n\rightarrow\infty]{} \sim \mathcal{N}(s, \frac{s}{\sqrt{2\,n}})
\end{aligned}
$$

$^{(*)}$ The most sensitive is the prior on σ $\Rightarrow$ it induces abstract speculations in mathematicians and statisticians who often have little idea of what they are talking about (**Gauss was Gauss!**).
$\rightarrow$ See **appendix** with details on small numbers

Including uncertainties due to systematics

Exact solution in a special (important) case

Introducing systematics
Influence quantities (ISO)

Introducing systematics

Influence quantities (ISO)

By influence quantities we mean:

$\rightarrow$ all kinds of external factors which can influence the result (temperature, atmospheric pressure, etc.);

Introducing systematics

Influence quantities (ISO)

By influence quantities we mean:

→ all kinds of external factors which can influence the result (temperature, atmospheric pressure, etc.);

→ all calibration constants;

Introducing systematics

Influence quantities (ISO)

By influence quantities we mean:

- → all kinds of external factors which can influence the result (temperature, atmospheric pressure, etc.);
- → all calibration constants;
- → all possible hypotheses upon which the results may depend (e.g. Monte Carlo parameters).
- → '***h***'

Introducing systematics

Influence quantities (ISO)

By influence quantities we mean:

- $\rightarrow$ all kinds of external factors which can influence the result (temperature, atmospheric pressure, etc.);
- $\rightarrow$ all calibration constants;
- $\rightarrow$ all possible hypotheses upon which the results may depend (e.g. Monte Carlo parameters).
- $\rightarrow$ '$\boldsymbol{h}$'

From a probabilistic point of view, there is no distinction between $\boldsymbol{\mu}$ and $\boldsymbol{h}$:

Introducing systematics

Influence quantities (ISO)

By influence quantities we mean:

- $\rightarrow$ all kinds of external factors which can influence the result (temperature, atmospheric pressure, etc.);
- $\rightarrow$ all calibration constants;
- $\rightarrow$ all possible hypotheses upon which the results may depend (e.g. Monte Carlo parameters).
- $\rightarrow$ '$\boldsymbol{h}$'

From a probabilistic point of view, there is no distinction between $\boldsymbol{\mu}$ and $\boldsymbol{h}$: they are all conditional hypotheses for the $\boldsymbol{x}$, i.e. causes which produce the observed effects.

Introducing systematics

Influence quantities (ISO)

By influence quantities we mean:

→ all kinds of external factors which can influence the result (temperature, atmospheric pressure, etc.);

→ all calibration constants;

→ all possible hypotheses upon which the results may depend (e.g. Monte Carlo parameters).

→ '$\boldsymbol{h}$'

From a probabilistic point of view, there is no distinction between $\boldsymbol{\mu}$ and $\boldsymbol{h}$: they are all conditional hypotheses for the $\boldsymbol{x}$, i.e. causes which produce the observed effects. The difference is simply that we are interested in $\boldsymbol{\mu}$ rather than in $\boldsymbol{h}$.

Introducing systematics

Influence quantities (ISO)

By influence quantities we mean:

- $\rightarrow$ all kinds of external factors which can influence the result (temperature, atmospheric pressure, etc.);
- $\rightarrow$ all calibration constants;
- $\rightarrow$ all possible hypotheses upon which the results may depend (e.g. Monte Carlo parameters).
- $\rightarrow$ '$\boldsymbol{h}$'

From a probabilistic point of view, there is no distinction between $\boldsymbol{\mu}$ and $\boldsymbol{h}$: they are all conditional hypotheses for the $\boldsymbol{x}$, i.e. causes which produce the observed effects. The difference is simply that we are interested in $\boldsymbol{\mu}$ rather than in $\boldsymbol{h}$.

(But remember that many 'h's are Physics:

Introducing systematics

Influence quantities (ISO)

By influence quantities we mean:

- → all kinds of external factors which can influence the result (temperature, atmospheric pressure, etc.);
- → all calibration constants;
- → all possible hypotheses upon which the results may depend (e.g. Monte Carlo parameters).
- → '$\boldsymbol{h}$'

From a probabilistic point of view, there is no distinction between $\boldsymbol{\mu}$ and $\boldsymbol{h}$: they are all conditional hypotheses for the $\boldsymbol{x}$, i.e. causes which produce the observed effects. The difference is simply that we are interested in $\boldsymbol{\mu}$ rather than in $\boldsymbol{h}$.

(But remember that many 'h's are Physics:
for example you will not be able to get meaningful results on the primordial Universe if you do not first understand the Physics of Cosmic Dust)

Introducing systematics

Several approaches (within probability theory – no adhocheries!)

Uncertainty due to systematic effects is also included in a natural way in this approach. Let us first define the notation (i is the generic index):

Introducing systematics

Several approaches (within probability theory – no adhocheries!)

Uncertainty due to systematic effects is also included in a natural way in this approach. Let us first define the notation (i is the generic index):

- $\boldsymbol{x} = \{x_1, x_2, \ldots x_{n_x}\}$ is the 'n-tuple' (vector) of data X_i;

Introducing systematics

Several approaches (within probability theory – no adhocheries!)

Uncertainty due to systematic effects is also included in a natural way in this approach. Let us first define the notation (*i* is the generic index):

- $\boldsymbol{x} = \{x_1, x_2, \ldots x_{n_x}\}$ is the 'n-tuple' (vector) of data X_i;
- $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots \mu_{n_\mu}\}$ is the n-tuple of *(true) values* μ_i;

Introducing systematics

Several approaches (within probability theory – no adhocheries!)

Uncertainty due to systematic effects is also included in a natural way in this approach. Let us first define the notation (i is the generic index):

- $\boldsymbol{x} = \{x_1, x_2, \ldots x_{n_x}\}$ is the 'n-tuple' (vector) of data X_i;
- $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots \mu_{n_\mu}\}$ is the n-tuple of *(true) values* μ_i;
- $\boldsymbol{h} = \{h_1, h_2, \ldots h_{n_h}\}$ is the n-tuple of influence quantities H_i. (see ISO GUM).

Taking into account of uncertain $\boldsymbol{h}$

1. Global inference on $f(\boldsymbol{\mu}, \boldsymbol{h})$

- We can use Bayes' theorem to make a joint inference on $\boldsymbol{\mu}$ and $\boldsymbol{h}$.

Taking into account of uncertain $\boldsymbol{h}$

1. Global inference on $f(\boldsymbol{\mu}, \boldsymbol{h})$

- We can use Bayes' theorem to make a joint inference on $\boldsymbol{\mu}$ and $\boldsymbol{h}$. A subsequent marginalization over $\boldsymbol{h}$ yields the p.d.f. of interest:

$$\boldsymbol{x} \Rightarrow f(\boldsymbol{\mu}, \boldsymbol{h} \,|\, \boldsymbol{x}) \Rightarrow f(\boldsymbol{\mu} \,|\, \boldsymbol{x})\,.$$

Taking into account of uncertain $\boldsymbol{h}$

1. Global inference on $f(\boldsymbol{\mu}, \boldsymbol{h})$

- We can use Bayes' theorem to make a joint inference on $\boldsymbol{\mu}$ and $\boldsymbol{h}$. A subsequent marginalization over $\boldsymbol{h}$ yields the p.d.f. of interest:

$$\boldsymbol{x} \Rightarrow f(\boldsymbol{\mu}, \boldsymbol{h} \,|\, \boldsymbol{x}) \Rightarrow f(\boldsymbol{\mu} \,|\, \boldsymbol{x}) \,.$$

  This method, depending on the joint prior distribution $f_\circ(\boldsymbol{\mu}, \boldsymbol{h})$, can even model possible correlations between $\boldsymbol{\mu}$ and $\boldsymbol{h}$.

Taking into account of uncertain $\boldsymbol{h}$

1. Global inference on $f(\boldsymbol{\mu}, \boldsymbol{h})$

- We can use Bayes' theorem to make a joint inference on $\boldsymbol{\mu}$ and $\boldsymbol{h}$. A subsequent marginalization over $\boldsymbol{h}$ yields the p.d.f. of interest:

$$\boldsymbol{x} \Rightarrow f(\boldsymbol{\mu}, \boldsymbol{h} \,|\, \boldsymbol{x}) \Rightarrow f(\boldsymbol{\mu} \,|\, \boldsymbol{x}) \,.$$

  This method, depending on the joint prior distribution $f_\circ(\boldsymbol{\mu}, \boldsymbol{h})$, can even model possible correlations between $\boldsymbol{\mu}$ and $\boldsymbol{h}$.(*)

(*) Usually this is not the case, but in some field of reaserch it is not impossible that e.g. the physical properties of the instrument depend on the physical properties (of materials) you are studying.

Taking into account of uncertain $\boldsymbol{h}$

2. Conditional inference

- Given the observed data, one has a joint distribution of $\boldsymbol{\mu}$ for all possible configurations of $\boldsymbol{h}$:

$$\boldsymbol{x} \Rightarrow f(\boldsymbol{\mu} \,|\, \boldsymbol{x}, \boldsymbol{h})\,.$$

Taking into account of uncertain $\boldsymbol{h}$

2. Conditional inference

- Given the observed data, one has a joint distribution of $\boldsymbol{\mu}$ for all possible configurations of $\boldsymbol{h}$:

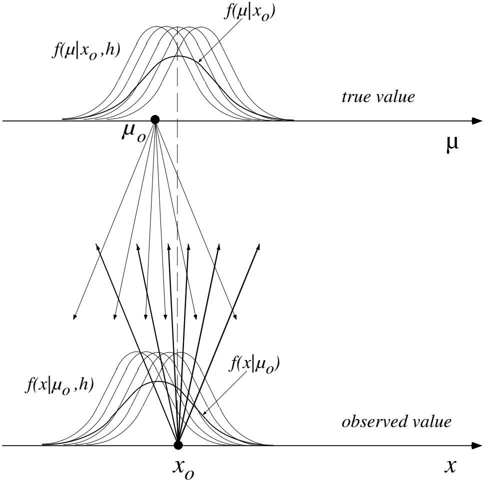
$$\boldsymbol{x} \Rightarrow f(\boldsymbol{\mu} \,|\, \boldsymbol{x}, \boldsymbol{h})\,.$$

Each conditional result is reweighed with the distribution of beliefs of $\boldsymbol{h}$, using the well-known law of probability:

$$f(\boldsymbol{\mu} \,|\, \boldsymbol{x}) = \int f(\boldsymbol{\mu} \,|\, \boldsymbol{x}, \boldsymbol{h})\, f(\boldsymbol{h})\, \mathrm{d}\boldsymbol{h}\,.$$

Taking into account of uncertain *h*

Conditional inference



Taking into account of uncertain $\boldsymbol{h}$

3.Propagation of uncertainties

- Essentially, one applies the propagation of uncertainty, making use of the following model:

Taking into account of uncertain $\boldsymbol{h}$

3.Propagation of uncertainties

- Essentially, one applies the propagation of uncertainty, making use of the following model: One considers a 'raw result' on raw values $\boldsymbol{\mu}_R$ for some nominal values of the influence quantities, i.e.

$$f(\boldsymbol{\mu}_R \,|\, \boldsymbol{x}, \boldsymbol{h}_\circ)\,.$$

Taking into account of uncertain $\boldsymbol{h}$

- Essentially, one applies the propagation of uncertainty, making use of the following model: One considers a 'raw result' on raw values $\boldsymbol{\mu}_R$ for some nominal values of the influence quantities, i.e.

$$f(\boldsymbol{\mu}_R \,|\, \boldsymbol{x}, \boldsymbol{h}_\circ)\,.$$

  Then the (corrected) true values are obtained as a function of the raw ones and of the possible values of the influence quantities

Taking into account of uncertain $\boldsymbol{h}$

3.Propagation of uncertainties

- Essentially, one applies the propagation of uncertainty, making use of the following model: One considers a 'raw result' on raw values $\boldsymbol{\mu}_R$ for some nominal values of the influence quantities, i.e.

$$f(\boldsymbol{\mu}_R \,|\, \boldsymbol{x}, \boldsymbol{h}_\circ)\,.$$

  Then the (corrected) true values are obtained as a function of the raw ones and of the possible values of the influence quantities, i.e.

$$\mu_i = \mu_i(\mu_{i_R}, \boldsymbol{h})\,,$$

  and $f(\boldsymbol{\mu})$ is evaluated by probability rules.

Taking into account of uncertain $\boldsymbol{h}$

- Essentially, one applies the propagation of uncertainty, making use of the following model: One considers a 'raw result' on raw values $\boldsymbol{\mu}_R$ for some nominal values of the influence quantities, i.e.

$$f(\boldsymbol{\mu}_R \,|\, \boldsymbol{x}, \boldsymbol{h}_\circ)\,.$$

  Then the (corrected) true values are obtained as a function of the raw ones and of the possible values of the influence quantities, i.e.

$$\mu_i = \mu_i(\mu_{i_R}, \boldsymbol{h})\,,$$

  and $f(\boldsymbol{\mu})$ is evaluated by probability rules.

This third approach is particularly convenient to make linear expansions which lead to approximate formulae.

Systematics due to uncertain offset

Model:

- the *"zero" of the instrument is usually not exactly known*, owing to calibration uncertainty.

Systematics due to uncertain offset

Model:

- the *"zero" of the instrument is usually not exactly known*, owing to calibration uncertainty.
- This can be parametrized assuming that the (true) value Z is normally distributed around 0 (i.e. the calibration was properly done!) with a standard deviation σ_Z:

$$Z \sim \mathcal{N}(0, \sigma_Z)$$

Systematics due to uncertain offset

Model:

- the *"zero" of the instrument is usually not exactly known*, owing to calibration uncertainty.
- This can be parametrized assuming that the (true) value Z is normally distributed around 0 (i.e. the calibration was properly done!) with a standard deviation σ_Z:

$$Z \sim \mathcal{N}(0, \sigma_Z)$$

- Since the (true) value of μ is usually independent of the true value of Z, the prior joint probability density function can be written as the product of the marginal ones:

$$f_\circ(\mu, z) = f_\circ(\mu)\, f_\circ(z) = k\, \frac{1}{\sqrt{2\,\pi}\,\sigma_Z}\, \exp\left[-\frac{z^2}{2\,\sigma_Z^2}\right].$$

Systematics due to uncertain offset

Model:

- the *"zero" of the instrument is usually not exactly known*, owing to calibration uncertainty.
- This can be parametrized assuming that the (true) value Z is normally distributed around 0 (i.e. the calibration was properly done!) with a standard deviation σ_Z:

$$Z \sim \mathcal{N}(0, \sigma_Z)$$

- Since the (true) value of μ is usually independent of the true value of Z, the prior joint probability density function can be written as the product of the marginal ones:

$$f_\circ(\mu, z) = f_\circ(\mu)\, f_\circ(z) = k\, \frac{1}{\sqrt{2\,\pi}\,\sigma_Z}\, \exp\left[-\frac{z^2}{2\,\sigma_Z^2}\right].$$

- X is no longer Gaussian distributed around μ,

Systematics due to uncertain offset

Model:

- the *"zero" of the instrument is usually not exactly known*, owing to calibration uncertainty.
- This can be parametrized assuming that the (true) value Z is normally distributed around 0 (i.e. the calibration was properly done!) with a standard deviation σ_Z:

$$Z \sim \mathcal{N}(0, \sigma_Z)$$

- Since the (true) value of μ is usually independent of the true value of Z, the prior joint probability density function can be written as the product of the marginal ones:

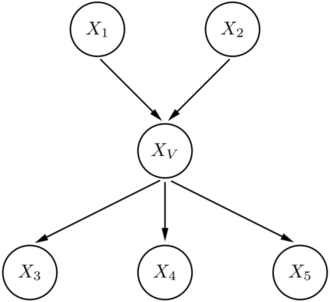
$$f_\circ(\mu, z) = f_\circ(\mu)\, f_\circ(z) = k\, \frac{1}{\sqrt{2\,\pi}\,\sigma_Z}\, \exp\left[-\frac{z^2}{2\,\sigma_Z^2}\right].$$

- X is no longer Gaussian distributed around μ, but around $\mu + Z$:

$$X \sim \mathcal{N}(\mu + Z, \sigma)$$

Graphical model

In terms of the generic variables X_i (arXiv:1504.02065)



$$
\begin{aligned}
X_1 &\longleftrightarrow \mu \\
X_2 &\longleftrightarrow Z \\
X_V &\longleftrightarrow \mu + Z \\
X_{3-\ldots} &\longleftrightarrow X_i \text{ (observations)}
\end{aligned}
$$

Systematics due to uncertain offset

Application to the single (equivalent) measuement X_1, with std σ_1

Likelihood:

$$f(x_1 \,|\, \mu, z) = \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right].$$

Systematics due to uncertain offset

Application to the single (equivalent) measuement X_1, with std σ_1

Likelihood:

$$f(x_1 \,|\, \mu, z) = \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right].$$

$$f(\mu, z \,|\, x_1) \propto \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right]\frac{1}{\sqrt{2\,\pi}\,\sigma_Z}\exp\left[-\frac{z^2}{2\,\sigma_Z^2}\right]$$

Systematics due to uncertain offset

Application to the single (equivalent) measuement X_1, with std σ_1

Likelihood:

$$f(x_1 \,|\, \mu, z) = \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right].$$

$$f(\mu, z \,|\, x_1) \propto \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right]\frac{1}{\sqrt{2\,\pi}\,\sigma_Z}\exp\left[-\frac{z^2}{2\,\sigma_Z^2}\right]$$

After joint inference and marginalization

$$f(\mu \,|\, x_1) = \frac{\int \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right]\frac{1}{\sqrt{2\,\pi}\,\sigma_Z}\exp\left[-\frac{z^2}{2\,\sigma_Z^2}\right]\mathrm{d}z}{\int\!\!\int \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right]\frac{1}{\sqrt{2\,\pi}\,\sigma_Z}\exp\left[-\frac{z^2}{2\,\sigma_Z^2}\right]\mathrm{d}\mu\,\mathrm{d}z}.$$

Systematics due to uncertain offset

Application to the single (equivalent) measuement X_1, with std σ_1

Likelihood:

$$f(x_1\,|\,\mu, z) = \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right].$$

$$f(\mu, z\,|\,x_1) \propto \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right]\frac{1}{\sqrt{2\,\pi}\,\sigma_Z}\,\exp\left[-\frac{z^2}{2\,\sigma_Z^2}\right]$$

After joint inference and marginalization

$$f(\mu\,|\,x_1) = \frac{\int \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right]\frac{1}{\sqrt{2\,\pi}\,\sigma_Z}\,\exp\left[-\frac{z^2}{2\,\sigma_Z^2}\right]\mathrm{d}z}{\iint \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu-z)^2}{2\,\sigma_1^2}\right]\frac{1}{\sqrt{2\,\pi}\,\sigma_Z}\,\exp\left[-\frac{z^2}{2\,\sigma_Z^2}\right]\mathrm{d}\mu\,\mathrm{d}z}.$$

Integrating we get

$$f(\mu) = f(\mu\,|\,x_1, \ldots, f_\circ(z)) = \frac{1}{\sqrt{2\,\pi}\,\sqrt{\sigma_1^2+\sigma_Z^2}}\,\exp\left[-\frac{(\mu-x_1)^2}{2\,(\sigma_1^2+\sigma_Z^2)}\right].$$

Systematics due to uncertain offset

Technical remark

It may help to know that

$$\int_{-\infty}^{+\infty} \exp\left[b\,x - \frac{x^2}{a^2}\right] \mathrm{d}x = \sqrt{a^2\,\pi}\,\exp\left[\frac{a^2\,b^2}{4}\right]$$

Systematics due to uncertain offset

Result

$$f(\mu) = f(\mu\,|\,x_1,\ldots,f_\circ(z)) = \frac{1}{\sqrt{2\,\pi}\,\sqrt{\sigma_1^2+\sigma_Z^2}}\,\exp\left[-\frac{(\mu-x_1)^2}{2\,(\sigma_1^2+\sigma_Z^2)}\right].$$

Systematics due to uncertain offset

Result

$$f(\mu) = f(\mu \,|\, x_1, \ldots, f_\circ(z)) = \frac{1}{\sqrt{2\,\pi}\,\sqrt{\sigma_1^2 + \sigma_Z^2}}\,\exp\left[-\frac{(\mu - x_1)^2}{2\,(\sigma_1^2 + \sigma_Z^2)}\right].$$

- $f(\mu)$ is still Gaussian, but with a larger variance

Systematics due to uncertain offset

Result

$$f(\mu) = f(\mu \,|\, x_1, \ldots, f_\circ(z)) = \frac{1}{\sqrt{2\,\pi}\,\sqrt{\sigma_1^2 + \sigma_Z^2}} \,\exp\left[-\frac{(\mu - x_1)^2}{2\,(\sigma_1^2 + \sigma_Z^2)}\right].$$

- $f(\mu)$ is still Gaussian, but with a larger variance
- The global standard uncertainty is the quadratic combination of that due to the statistical fluctuation of the data sample and the uncertainty due to the imperfect knowledge of the *systematic effect*:

$$\sigma_{tot}^2 = \sigma_1^2 + \sigma_Z^2\,.$$

Systematics due to uncertain offset

$$f(\mu) = f(\mu \,|\, x_1, \ldots, f_\circ(z)) = \frac{1}{\sqrt{2\,\pi}\,\sqrt{\sigma_1^2 + \sigma_Z^2}} \exp\left[-\frac{(\mu - x_1)^2}{2\,(\sigma_1^2 + \sigma_Z^2)}\right].$$

- $f(\mu)$ is still Gaussian, but with a larger variance
- The global standard uncertainty is the quadratic combination of that due to the statistical fluctuation of the data sample and the uncertainty due to the imperfect knowledge of the *systematic effect*:

$$\sigma_{tot}^2 = \sigma_1^2 + \sigma_Z^2\,.$$

- This result (a theorem under well stated conditions!) is often used as a 'prescription', although there are still some "old-fashioned" recipes which require different combinations of the contributions to be performed.

Systematics due to uncertain offset

Measuring two quantities with the same instrument

Measuring μ_1 and μ_2, resulting into x_1 and x_2.

Systematics due to uncertain offset

Measuring two quantities with the same instrument

Measuring μ_1 and μ_2, resulting into x_1 and x_2.

Setting up the model:

$$Z \sim \mathcal{N}(0, \sigma_Z)$$

Systematics due to uncertain offset

Measuring two quantities with the same instrument

Measuring μ_1 and μ_2, resulting into x_1 and x_2.
Setting up the model:

$$\begin{aligned} Z &\sim \mathcal{N}(0, \sigma_Z) \\ X_1 &\sim \mathcal{N}(\mu_1 + Z, \sigma_1) \end{aligned}$$

Systematics due to uncertain offset

Measuring two quantities with the same instrument

Measuring μ_1 and μ_2, resulting into x_1 and x_2.
Setting up the model:

$$
\begin{aligned}
Z &\sim \mathcal{N}(0, \sigma_Z) \\
X_1 &\sim \mathcal{N}(\mu_1 + Z, \sigma_1) \\
X_2 &\sim \mathcal{N}(\mu_2 + Z, \sigma_2)
\end{aligned}
$$

Systematics due to uncertain offset

Measuring two quantities with the same instrument

Measuring μ_1 and μ_2, resulting into x_1 and x_2.
Setting up the model:

$$
\begin{aligned}
Z &\sim \mathcal{N}(0, \sigma_Z) \\
X_1 &\sim \mathcal{N}(\mu_1 + Z, \sigma_1) \\
X_2 &\sim \mathcal{N}(\mu_2 + Z, \sigma_2) \\
f(x_1, x_2 \,|\, \mu_1, \mu_2, z) &= \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\, \exp\left[-\frac{(x_1 - \mu_1 - z)^2}{2\,\sigma_1^2}\right] \\
&\quad \times \frac{1}{\sqrt{2\,\pi}\,\sigma_2}\, \exp\left[-\frac{(x_2 - \mu_2 - z)^2}{2\,\sigma_2^2}\right]
\end{aligned}
$$

Systematics due to uncertain offset

Measuring two quantities with the same instrument

Measuring μ_1 and μ_2, resulting into x_1 and x_2.

Setting up the model:

$$
\begin{aligned}
Z &\sim \mathcal{N}(0, \sigma_Z) \\
X_1 &\sim \mathcal{N}(\mu_1 + Z, \sigma_1) \\
X_2 &\sim \mathcal{N}(\mu_2 + Z, \sigma_2) \\
f(x_1, x_2 \,|\, \mu_1, \mu_2, z) &= \frac{1}{\sqrt{2\,\pi}\,\sigma_1}\,\exp\left[-\frac{(x_1-\mu_1-z)^2}{2\,\sigma_1^2}\right] \\
&\quad \times \frac{1}{\sqrt{2\,\pi}\,\sigma_2}\,\exp\left[-\frac{(x_2-\mu_2-z)^2}{2\,\sigma_2^2}\right] \\
&= \frac{1}{2\,\pi\,\sigma_1\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{(x_1-\mu_1-z)^2}{\sigma_1^2} + \frac{(x_2-\mu_2-z)^2}{\sigma_2^2}\right)\right]
\end{aligned}
$$

Systematics due to uncertain offset

Measuring two quantities with the same instrument

$$f(\mu_1, \mu_2 \,|\, x_1, x_2) \;=\; \frac{\int f(x_1, x_2 \,|\, \mu_1, \mu_2, z)\, f_\circ(\mu_1, \mu_2, z)\mathrm{d}z}{\int \ldots \,\mathrm{d}\mu_1\, \mathrm{d}\mu_2\, \mathrm{d}z}$$

Systematics due to uncertain offset

Measuring two quantities with the same instrument

$$f(\mu_1, \mu_2 \,|\, x_1, x_2) = \frac{\int f(x_1, x_2 \,|\, \mu_1, \mu_2, z)\, f_\circ(\mu_1, \mu_2, z)\mathrm{d}z}{\int \ldots\, \mathrm{d}\mu_1\, \mathrm{d}\mu_2\, \mathrm{d}z}$$

$$= \frac{1}{2\,\pi\,\sqrt{\sigma_1^2+\sigma_Z^2}\,\sqrt{\sigma_2^2+\sigma_Z^2}\,\sqrt{1-\rho^2}}$$

$$\times \exp\left\{-\frac{1}{2\,(1-\rho^2)}\left[\frac{(\mu_1-x_1)^2}{\sigma_1^2+\sigma_Z^2}\right.\right.$$

$$\left.\left. -2\,\rho\,\frac{(\mu_1-x_1)(\mu_2-x_2)}{\sqrt{\sigma_1^2+\sigma_Z^2}\,\sqrt{\sigma_2^2+\sigma_Z^2}} + \frac{(\mu_2-x_2)^2}{\sigma_2^2+\sigma_Z^2}\right]\right\}$$

Systematics due to uncertain offset

Measuring two quantities with the same instrument

$$
\begin{aligned}
f(\mu_1,\mu_2\,|\,x_1,x_2) &= \frac{\int f(x_1,x_2\,|\,\mu_1,\mu_2,z)\,f_\circ(\mu_1,\mu_2,z)\mathrm{d}z}{\int \ldots\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2\,\mathrm{d}z} \\
&= \frac{1}{2\,\pi\,\sqrt{\sigma_1^2+\sigma_Z^2}\,\sqrt{\sigma_2^2+\sigma_Z^2}\,\sqrt{1-\rho^2}} \\
&\quad \times \exp\left\{-\frac{1}{2\,(1-\rho^2)}\left[\frac{(\mu_1-x_1)^2}{\sigma_1^2+\sigma_Z^2}\right.\right. \\
&\quad \left.\left. -2\,\rho\,\frac{(\mu_1-x_1)(\mu_2-x_2)}{\sqrt{\sigma_1^2+\sigma_Z^2}\,\sqrt{\sigma_2^2+\sigma_Z^2}} + \frac{(\mu_2-x_2)^2}{\sigma_2^2+\sigma_Z^2}\right]\right\}
\end{aligned}
$$

where

$$
\rho = \frac{\sigma_Z^2}{\sqrt{\sigma_1^2+\sigma_Z^2}\,\sqrt{\sigma_2^2+\sigma_Z^2}}\,.
$$

Systematics due to uncertain offset

Measuring two quantities with the same instrument

$$
\begin{aligned}
f(\mu_1,\mu_2\,|\,x_1,x_2) &= \frac{\int f(x_1,x_2\,|\,\mu_1,\mu_2,z)\, f_\circ(\mu_1,\mu_2,z)\mathrm{d}z}{\int \ldots\, \mathrm{d}\mu_1\, \mathrm{d}\mu_2\, \mathrm{d}z} \\
&= \frac{1}{2\,\pi\,\sqrt{\sigma_1^2+\sigma_Z^2}\,\sqrt{\sigma_2^2+\sigma_Z^2}\,\sqrt{1-\rho^2}} \\
&\quad \times \exp\left\{-\frac{1}{2\,(1-\rho^2)}\left[\frac{(\mu_1-x_1)^2}{\sigma_1^2+\sigma_Z^2}\right.\right. \\
&\quad \left.\left. -2\,\rho\,\frac{(\mu_1-x_1)(\mu_2-x_2)}{\sqrt{\sigma_1^2+\sigma_Z^2}\,\sqrt{\sigma_2^2+\sigma_Z^2}} + \frac{(\mu_2-x_2)^2}{\sigma_2^2+\sigma_Z^2}\right]\right\}
\end{aligned}
$$

where

$$
\rho = \frac{\sigma_Z^2}{\sqrt{\sigma_1^2+\sigma_Z^2}\,\sqrt{\sigma_2^2+\sigma_Z^2}}\,.
$$

$\Rightarrow$ bivariate normal distribution!

Systematics due to uncertain offset

Summary:

$$\mu_1 \sim \mathcal{N}\left(x_1, \sqrt{\sigma_1^2 + \sigma_Z^2}\right)$$

$$\mu_2 \sim \mathcal{N}\left(x_2, \sqrt{\sigma_2^2 + \sigma_Z^2}\right)$$

$$\rho = \frac{\sigma_Z^2}{\sqrt{\sigma_1^2 + \sigma_Z^2}\,\sqrt{\sigma_2^2 + \sigma_Z^2}}$$

Systematics due to uncertain offset

Summary:

$$\begin{aligned}
\mu_1 &\sim \mathcal{N}\left(x_1, \sqrt{\sigma_1^2 + \sigma_Z^2}\right) \\
\mu_2 &\sim \mathcal{N}\left(x_2, \sqrt{\sigma_2^2 + \sigma_Z^2}\right) \\
\rho &= \frac{\sigma_Z^2}{\sqrt{\sigma_1^2 + \sigma_Z^2}\,\sqrt{\sigma_2^2 + \sigma_Z^2}} \\
\mathrm{Cov}(\mu_1, \mu_2) &= \rho\,\sigma_{\mu_1}\sigma_{\mu_2} \\
&= \rho\,\sqrt{\sigma_1^2 + \sigma_Z^2}\,\sqrt{\sigma_2^2 + \sigma_Z^2} = \sigma_Z^2
\end{aligned}$$

Systematics due to uncertain offset

Summary:

$$
\begin{aligned}
\mu_1 &\sim \mathcal{N}\left(x_1, \sqrt{\sigma_1^2 + \sigma_Z^2}\right) \\
\mu_2 &\sim \mathcal{N}\left(x_2, \sqrt{\sigma_2^2 + \sigma_Z^2}\right) \\
\rho &= \frac{\sigma_Z^2}{\sqrt{\sigma_1^2 + \sigma_Z^2}\,\sqrt{\sigma_2^2 + \sigma_Z^2}} \\
\mathrm{Cov}(\mu_1, \mu_2) &= \rho\,\sigma_{\mu_1}\sigma_{\mu_2} \\
&= \rho\,\sqrt{\sigma_1^2 + \sigma_Z^2}\,\sqrt{\sigma_2^2 + \sigma_Z^2} = \sigma_Z^2
\end{aligned}
$$

Checks, defining $S = \mu_1 + \mu_2$ and $D = \mu_1 - \mu_2$

Systematics due to uncertain offset

Summary:

$$\begin{aligned}
\mu_1 &\sim \mathcal{N}\left(x_1, \sqrt{\sigma_1^2 + \sigma_Z^2}\right) \\
\mu_2 &\sim \mathcal{N}\left(x_2, \sqrt{\sigma_2^2 + \sigma_Z^2}\right) \\
\rho &= \frac{\sigma_Z^2}{\sqrt{\sigma_1^2 + \sigma_Z^2}\,\sqrt{\sigma_2^2 + \sigma_Z^2}} \\
\mathrm{Cov}(\mu_1, \mu_2) &= \rho\,\sigma_{\mu_1}\sigma_{\mu_2} \\
&= \rho\,\sqrt{\sigma_1^2 + \sigma_Z^2}\,\sqrt{\sigma_2^2 + \sigma_Z^2} = \sigma_Z^2
\end{aligned}$$

Checks, defining $S = \mu_1 + \mu_2$ and $D = \mu_1 - \mu_2$

$$D \sim \mathcal{N}\left(x_1 - x_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

Systematics due to uncertain offset

Summary:

$$
\begin{aligned}
\mu_1 &\sim \mathcal{N}\left(x_1, \sqrt{\sigma_1^2+\sigma_Z^2}\right)\\
\mu_2 &\sim \mathcal{N}\left(x_2, \sqrt{\sigma_2^2+\sigma_Z^2}\right)\\
\rho &= \frac{\sigma_Z^2}{\sqrt{\sigma_1^2+\sigma_Z^2}\,\sqrt{\sigma_2^2+\sigma_Z^2}}\\
\mathrm{Cov}(\mu_1,\mu_2) &= \rho\,\sigma_{\mu_1}\sigma_{\mu_2}\\
&= \rho\,\sqrt{\sigma_1^2+\sigma_Z^2}\,\sqrt{\sigma_2^2+\sigma_Z^2} = \sigma_Z^2
\end{aligned}
$$

Checks, defining $S=\mu_1+\mu_2$ and $D=\mu_1-\mu_2$

$$
\begin{aligned}
D &\sim \mathcal{N}\left(x_1-x_2, \sqrt{\sigma_1^2+\sigma_2^2}\right)\\
S &\sim \mathcal{N}\left(x_1+x_2, \sqrt{\sigma_1^2+\sigma_2^2+(2\,\sigma_Z)^2}\right)
\end{aligned}
$$

Systematics due to uncertain offset

Summary:

$$\mu_1 \sim \mathcal{N}\left(x_1, \sqrt{\sigma_1^2 + \sigma_Z^2}\right)$$

$$\mu_2 \sim \mathcal{N}\left(x_2, \sqrt{\sigma_2^2 + \sigma_Z^2}\right)$$

$$\rho = \frac{\sigma_Z^2}{\sqrt{\sigma_1^2 + \sigma_Z^2}\,\sqrt{\sigma_2^2 + \sigma_Z^2}}$$

$$\begin{aligned} \text{Cov}(\mu_1, \mu_2) &= \rho\,\sigma_{\mu_1}\sigma_{\mu_2} \\ &= \rho\,\sqrt{\sigma_1^2 + \sigma_Z^2}\,\sqrt{\sigma_2^2 + \sigma_Z^2} = \sigma_Z^2 \end{aligned}$$

Checks, defining $S = \mu_1 + \mu_2$ and $D = \mu_1 - \mu_2$

$$D \sim \mathcal{N}\left(x_1 - x_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

$$S \sim \mathcal{N}\left(x_1 + x_2, \sqrt{\sigma_1^2 + \sigma_2^2 + (2\,\sigma_Z)^2}\right)$$

As more or less intuitively expected from an offset!

A simple example with JAGS

```
model {
   for (i in 1:length(x)) {
      x[i] ~ dnorm(mu.s, tau)
   }
   mu.s <- mu + z;
   mu ~ dnorm(0.0, 1.0E-6)
   z ~ dnorm(0, 1/sigma.z^2)
   tau ~ dgamma(1.0, 1.0E-4)
   sigma <- 1/sqrt(tau)
   x.f ~ dnorm(mu.s, tau)
}
```

=> `norm_sist_z.R`

A simple example with JAGS

```
model {                                    =>  norm_sist_z.R
   for (i in 1:length(x)) {
      x[i] ~ dnorm(mu.s, tau)
   }
   mu.s <- mu + z;
   mu ~ dnorm(0.0, 1.0E-6)
   z ~ dnorm(0, 1/sigma.z^2)
   tau ~ dgamma(1.0, 1.0E-4)
   sigma <- 1/sqrt(tau)
   x.f ~ dnorm(mu.s, tau)
}
```

- extend it to two data sets measuring μ_1 and μ_2 with common systematics;

A simple example with JAGS

```
model {                                   =>  norm_sist_z.R
   for (i in 1:length(x)) {
      x[i] ~ dnorm(mu.s, tau)
   }
   mu.s <- mu + z;
   mu ~ dnorm(0.0, 1.0E-6)
   z ~ dnorm(0, 1/sigma.z^2)
   tau ~ dgamma(1.0, 1.0E-4)
   sigma <- 1/sqrt(tau)
   x.f ~ dnorm(mu.s, tau)
}
```

- extend it to two data sets measuring μ_1 and μ_2 with common systematics;
- include also derived quantities, like sum and difference;

A simple example with JAGS

```
model {                                        =>  norm_sist_z.R
   for (i in 1:length(x)) {
      x[i] ~ dnorm(mu.s, tau)
   }
   mu.s <- mu + z;
   mu ~ dnorm(0.0, 1.0E-6)
   z ~ dnorm(0, 1/sigma.z^2)
   tau ~ dgamma(1.0, 1.0E-4)
   sigma <- 1/sqrt(tau)
   x.f ~ dnorm(mu.s, tau)
}
```

- extend it to two data sets measuring μ_1 and μ_2 with common systematics;
- include also derived quantities, like sum and difference;
- modify the model in order to describe a systematics affetting the scale

A simple example with JAGS

```
model {                                        =>  norm_sist_z.R
   for (i in 1:length(x)) {
      x[i] ~ dnorm(mu.s, tau)
   }
   mu.s <- mu + z;
   mu ~ dnorm(0.0, 1.0E-6)
   z ~ dnorm(0, 1/sigma.z^2)
   tau ~ dgamma(1.0, 1.0E-4)
   sigma <- 1/sqrt(tau)
   x.f ~ dnorm(mu.s, tau)
}
```

- extend it to two data sets measuring μ_1 and μ_2 with common systematics;
- include also derived quantities, like sum and difference;
- modify the model in order to describe a systematics affetting the scale: $f = 1 \pm \sigma_f$;

A simple example with JAGS

```
model {                                  =>  norm_sist_z.R
   for (i in 1:length(x)) {
      x[i] ~ dnorm(mu.s, tau)
   }
   mu.s <- mu + z;
   mu ~ dnorm(0.0, 1.0E-6)
   z ~ dnorm(0, 1/sigma.z^2)
   tau ~ dgamma(1.0, 1.0E-4)
   sigma <- 1/sqrt(tau)
   x.f ~ dnorm(mu.s, tau)
}
```

- extend it to two data sets measuring μ_1 and μ_2 with common systematics;
- include also derived quantities, like sum and difference;
- modify the model in order to describe a systematics affetting the scale: $f = 1 \pm \sigma_f$;
- then, add other derived quantities, like product and ratio.

Fits – introduction

- In a probabilistic framework the issue of the fits is nothing but

  parametric inference.

- set up the model,
  e.g. $\mu_{y_i} = m\,\mu_{x_i} + c$

Fits – introduction

- In a probabilistic framework the issue of the fits is nothing but
  parametric inference.

- set up the model,
  e.g. $\mu_{y_i} = m\,\mu_{x_i} + c$
  Note: Linearity is between μ_{y_i} and μ_{x_i}, not between y_i and x_i!

Fits – introduction

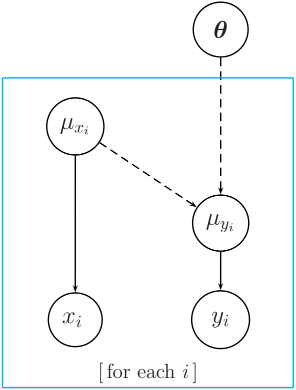
- In a probabilistic framework the issue of the fits is nothing but

  parametric inference.

- set up the model,
  e.g. $\mu_{y_i} = m\,\mu_{x_i} + c$
  Note: Linearity is between μ_{y_i} and μ_{x_i}, not between y_i and x_i!
- apply probability rules;

Fits – introduction

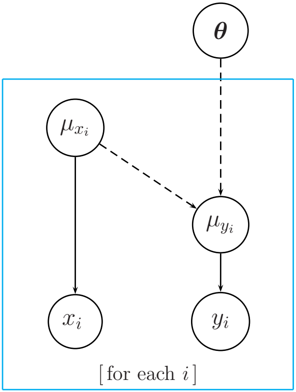
- In a probabilistic framework the issue of the fits is nothing but
  parametric inference.



- set up the model,
  e.g. $\mu_{y_i} = m\,\mu_{x_i} + c$
  Note: Linearity is between μ_{y_i} and μ_{x_i}, not between y_i and x_i!
- apply probability rules;
- perform the calculations.

Fits – introduction

- In a probabilistic framework the issue of the fits is nothing but parametric inference.



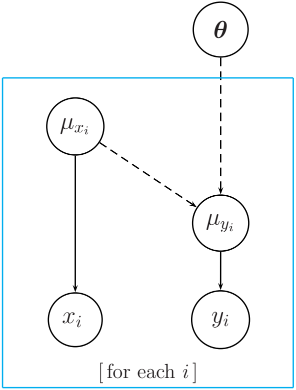
- set up the model,
  e.g. $\mu_{y_i} = m\,\mu_{x_i} + c$
  Note: Linearity is between μ_{y_i} and μ_{x_i}, not between y_i and x_i!
- apply probability rules;
- perform the calculations.

$\rightarrow f(\boldsymbol{\theta}\,|\,\mathrm{x}, \mathrm{y}, I)$

Fits – introduction

- In a probabilistic framework the issue of the fits is nothing but

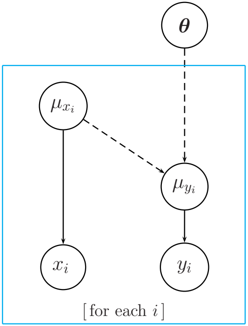
  parametric inference.



- set up the model,
  e.g. $\mu_{y_i} = m\,\mu_{x_i} + c$
  Note: Linearity is between μ_{y_i} and μ_{x_i}, not between y_i and x_i!
- apply probability rules;
- perform the calculations.

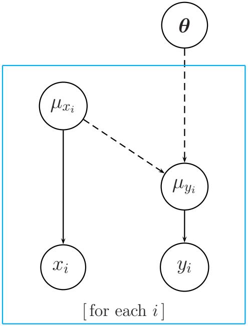
$\rightarrow f(\boldsymbol{\theta}\,|\,\mathrm{x}, \mathrm{y}, I)$
$\rightarrow f(m, c\,|\,\mathrm{x}, \mathrm{y}, \boldsymbol{\sigma})$, in the case of case of linear fit with "σ's known a priori" (!)

Linear fit – introduction



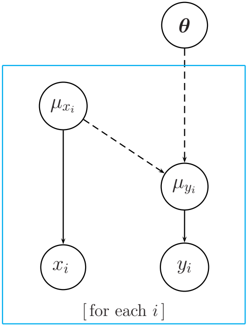
- Deterministic links between μ_x's and μ_y's.

Linear fit – introduction



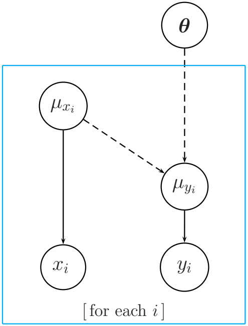
- Deterministic links between μ_x's and μ_y's.
- Probabilistic links between μ_x's and x's, and between μ_y's and y's (errors on both axes)

Linear fit – introduction



- Deterministic links between μ_x's and μ_y's.
- Probabilistic links between μ_x's and x's, and between μ_y's and y's (errors on both axes)
- $\Rightarrow$ aim of fit (<u>σ's known</u>): $\{\boldsymbol{x}, \boldsymbol{y}\} \;\rightarrow \boldsymbol{\theta} = (m, c)$

Linear fit – introduction



- Deterministic links between μ_x's and μ_y's.
- Probabilistic links between μ_x's and x's, and between μ_y's and y's (errors on both axes)
- $\Rightarrow$ aim of fit (<u>σ's known</u>): $\{\boldsymbol{x}, \boldsymbol{y}\} \rightarrow \boldsymbol{\theta} = (m, c)$
- If σ_x's and σ_y's are unkown and assumed all equal $\{\boldsymbol{x}, \boldsymbol{y}\} \rightarrow \boldsymbol{\theta} = (m, c, \sigma_x, \sigma_y)$
- etc...

Linear fit – simplest case

$$f(m, c \,|\, \boldsymbol{x}, \boldsymbol{y}, I) \;\;\propto\;\; f(\boldsymbol{x}, \boldsymbol{y} \,|\, m, c, I) \cdot f_0(m, c)$$

Simplifying hypotheses:

- No error on $\mu_x \Rightarrow \mu_{x_i} = x_i$:
  $\mu_{y_i} = m\,x_i + c$.

Linear fit – simplest case

$$f(m, c\,|\,\boldsymbol{x}, \boldsymbol{y}, I) \;\propto\; f(\boldsymbol{x}, \boldsymbol{y}\,|\,m, c, I) \cdot f_0(m, c)$$

Simplifying hypotheses:

- No error on $\mu_x \Rightarrow \mu_{x_i} = x_i$:
  $\mu_{y_i} = m\,x_i + c$.
- Gaussian errors on y, with $y_i \sim \mathcal{N}(\mu_{y_i}, \sigma_i)$, with σ_i "known somehow" (or "to be determined in some way")

Linear fit – simplest case

$$f(m, c \,|\, \boldsymbol{x}, \boldsymbol{y}, I) \;\propto\; f(\boldsymbol{x}, \boldsymbol{y} \,|\, m, c, I) \cdot f_0(m, c)$$

Simplifying hypotheses:

- No error on $\mu_x \Rightarrow \mu_{x_i} = x_i$:
  $\mu_{y_i} = m\,x_i + c$.
- Gaussian errors on y, with $y_i \sim \mathcal{N}(\mu_{y_i}, \sigma_i)$, with σ_i "known somehow" (or "to be determined in some way")
- Independence of data points.

Linear fit – simplest case

$$f(m, c \,|\, \boldsymbol{x}, \boldsymbol{y}, I) \;\propto\; f(\boldsymbol{x}, \boldsymbol{y} \,|\, m, c, I) \cdot f_0(m, c)$$

Simplifying hypotheses:

- No error on $\mu_x \Rightarrow \mu_{x_i} = x_i$:
  $\mu_{y_i} = m\,x_i + c$.
- Gaussian errors on y, with $y_i \sim \mathcal{N}(\mu_{y_i}, \sigma_i)$, with σ_i "known somehow" (or "to be determined in some way")
- Independence of data points.

$$\begin{aligned} f(m, c \,|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\sigma}) &\propto \exp\left[-\sum_i \frac{(y_i - \mu_{y_i})^2}{2\,\sigma_i^2}\right] \cdot f_0(m, c) \\ &\propto \exp\left[-\frac{1}{2}\sum_i \frac{(y_i - m\,x_i - c)^2}{\sigma_i^2}\right] \cdot f_0(m, c) \end{aligned}$$

Linear fit – simplest case

$$f(m, c \,|\, \boldsymbol{x}, \boldsymbol{y}, I) \;\;\propto\;\; f(\boldsymbol{x}, \boldsymbol{y} \,|\, m, c, I) \cdot f_0(m, c)$$

Simplifying hypotheses:

- No error on $\mu_x \Rightarrow \mu_{x_i} = x_i$:
  $\mu_{y_i} = m\, x_i + c$.
- Gaussian errors on y, with $y_i \sim \mathcal{N}(\mu_{y_i}, \sigma_i)$, with σ_i "known somehow" (or "to be determined in some way")
- Independence of data points.

$$f(m, c \,|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\sigma}) \;\;\propto\;\; \exp\left[-\sum_i \frac{(y_i - \mu_{y_i})^2}{2\,\sigma_i^2}\right] \cdot f_0(m, c)$$

$$\;\;\propto\;\; \exp\left[-\frac{1}{2}\sum_i \frac{(y_i - m\, x_i - c)^2}{\sigma_i^2}\right] \cdot f_0(m, c)$$

$\Rightarrow$ flat priors: inference only depends on $\exp\left[-\frac{1}{2}\sum_i \frac{(y_i - m\, x_i - c)^2}{\sigma_i^2}\right]$.

Least squares and 'Gaussian tricks' on linear fits

$$f(m, c \,|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\sigma}) \;\propto\; \exp\left[-\frac{\sum_i (y_i - m\,x_i - c)^2}{2\,\sigma_i^2}\right] \cdot f_0(m, c)$$

- <u>If</u> the prior is irrelevant and the σ's are all equal, than the maximum of the posterior is obtained when the sum of the squares is minimized:
  $\Rightarrow$ Least Square 'Principle'.

Least squares and 'Gaussian tricks' on linear fits

$$f(m, c \,|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\sigma}) \;\; \propto \;\; \exp\left[-\frac{\sum_i (y_i - m\,x_i - c)^2}{2\,\sigma_i^2}\right] \cdot f_0(m, c)$$

- If the prior is irrelevant and the σ's are all equal, than the maximum of the posterior is obtained when the sum of the squares is minimized:
  $\Rightarrow$ Least Square 'Principle'.
- You might recognize at the exponent: $\chi^2/2$:
  $\Rightarrow$ χ^2 minimization.

Least squares and 'Gaussian tricks' on linear fits

$$f(m, c \,|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\sigma}) \;\;\propto\;\; \exp\left[-\frac{\sum_i (y_i - m\,x_i - c)^2}{2\,\sigma_i^2}\right] \cdot f_0(m, c)$$

- If the prior is irrelevant and the σ's are all equal, than the maximum of the posterior is obtained when the sum of the squares is minimized:
  $\Rightarrow$ Least Square 'Principle'.
- You might recognize at the exponent: $\chi^2/2$:
  $\Rightarrow$ χ^2 minimization.
- As an approximation, one can obtain best fit parameters and covariance matrix by the 'Gaussian trick'
  $\Rightarrow$ $\varphi(m, c) \propto \chi^2$.

Least squares and 'Gaussian tricks' on linear fits

$$f(m, c \,|\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\sigma}) \;\;\propto\;\; \exp\left[-\frac{\sum_i (y_i - m\,x_i - c)^2}{2\,\sigma_i^2}\right] \cdot f_0(m, c)$$

- If the prior is irrelevant and the σ's are all equal, than the maximum of the posterior is obtained when the sum of the squares is minimized:
  $\Rightarrow$ Least Square 'Principle'.
- You might recognize at the exponent: $\chi^2/2$:
  $\Rightarrow$ χ^2 minimization.
- As an approximation, one can obtain best fit parameters and covariance matrix by the 'Gaussian trick'
  $\Rightarrow$ $\varphi(m, c) \propto \chi^2$.

$\Rightarrow$ same result of the detailed one is achieved, simply because the problem is linear!
(No garantee in general!)

Uncertain standard deviation

In the probabilistic approach it is rather simple: just add σ in $\boldsymbol{\theta}$ to infer.

- For example, if we have good reasons to belief that the σ's are all equal, then

$$f(m, c, \sigma \,|\, \boldsymbol{x}, \boldsymbol{y}) \;\propto\; \sigma^{-n} \exp\left[-\frac{\sum_i (y_i - m\,x_i - c)^2}{2\,\sigma^2}\right] \cdot f_0(m, c, \sigma)$$

Uncertain standard deviation

In the probabilistic approach it is rather simple: just add σ in $\boldsymbol{\theta}$ to infer.

- For example, if we have good reasons to belief that the σ's are all equal, then

$$f(m, c, \sigma \,|\, \boldsymbol{x}, \boldsymbol{y}) \;\; \propto \;\; \sigma^{-n} \exp\left[-\frac{\sum_i (y_i - m\,x_i - c)^2}{2\,\sigma^2}\right] \cdot f_0(m, c, \sigma)$$

  Even if the prior is flat in all parameters

  - methods "based only on the properties of the argument of the exponent" fail, because they miss the contribution from σ^{-n}!

Uncertain standard deviation

In the probabilistic approach it is rather simple: just add σ in $\boldsymbol{\theta}$ to infer.

- For example, if we have good reasons to belief that the σ's are all equal, then

$$f(m, c, \sigma \,|\, \boldsymbol{x}, \boldsymbol{y}) \;\; \propto \;\; \sigma^{-n} \exp\left[-\frac{\sum_i (y_i - m\, x_i - c)^2}{2\,\sigma^2}\right] \cdot f_0(m, c, \sigma)$$

  Even if the prior is flat in all parameters

  - methods "based only on the properties of the argument of the exponent" fail, because they miss the contribution from σ^{-n}!
  - The Gaussian trick applied to the full posterior perfoms better.

Uncertain standard deviation

In the probabilistic approach it is rather simple: just add σ in $\boldsymbol{\theta}$ to infer.

- For example, if we have good reasons to belief that the σ's are all equal, then

$$f(m, c, \sigma \,|\, \boldsymbol{x}, \boldsymbol{y}) \quad \propto \quad \sigma^{-n} \exp\left[-\frac{\sum_i (y_i - m\,x_i - c)^2}{2\,\sigma^2}\right] \cdot f_0(m, c, \sigma)$$

  Even if the prior is flat in all parameters

  - methods "based only on the properties of the argument of the exponent" fail, because they miss the contribution from σ^{-n}!
  - The Gaussian trick applied to the full posterior perfoms better.

Residuals? Ok if there are many points, otherwise we do not take into account the uncertainty on σ and its effect on the probability function of m and c.

Uncertain standard deviation

In the probabilistic approach it is rather simple: just add σ in $\boldsymbol{\theta}$ to infer.

- For example, if we have good reasons to belief that the σ's are all equal, then

$$f(m, c, \sigma \,|\, \boldsymbol{x}, \boldsymbol{y}) \;\propto\; \sigma^{-n} \exp\left[-\frac{\sum_i (y_i - m\, x_i - c)^2}{2\,\sigma^2}\right] \cdot f_0(m, c, \sigma)$$

  Even if the prior is flat in all parameters

  - methods "based only on the properties of the argument of the exponent" fail, because they miss the contribution from σ^{-n}!
  - The Gaussian trick applied to the full posterior perfoms better.

Residuals? Ok if there are many points, otherwise we do not take into account the uncertainty on σ and its effect on the probability function of m and c.

Note: as long as σ is constant (although unknown) and the prior flat in m and c the best estimates of m and c do not depend in σ.

Linear fits with uncertain σ in JAGS

Model

```
var mu.y[N];
model{
    for (i in 1:N) {
      y[i] ~ dnorm(mu.y[i], tau);
      mu.y[i] <- x[i]*m + c;
    }
    c ~ dnorm(0, 1.0E-6);
    m ~ dnorm(0, 1.0E-6);
    tau ~ dgamma(1.0, 1.0E-6);
    sigma <- 1.0/sqrt(tau);
}
```

Linear fits with uncertain σ in JAGS

Model

```
var mu.y[N];
model{
    for (i in 1:N) {
      y[i] ~ dnorm(mu.y[i], tau);
      mu.y[i] <- x[i]*m + c;
    }
    c ~ dnorm(0, 1.0E-6);
    m ~ dnorm(0, 1.0E-6);
    tau ~ dgamma(1.0, 1.0E-6);
    sigma <- 1.0/sqrt(tau);
}
```

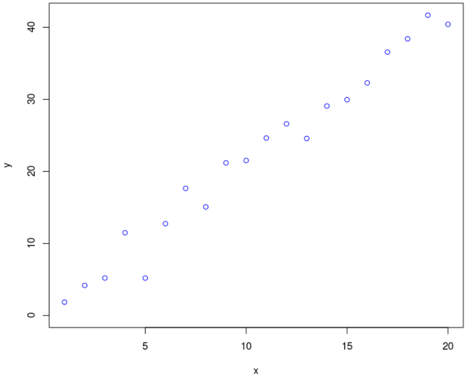
Simulated data

```
m.true = 2; c.true = 1; sigma.true=2
x = 1:20
y = m.true * x + c.true + rnorm(length(x), 0, sigma.true)

plot(x,y, col='blue',ylim=c(0,max(y)) )
```

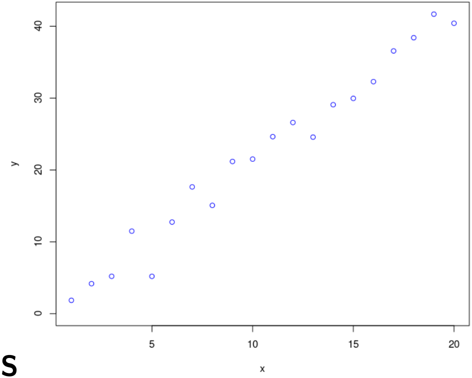
Linear fits with uncertain σ in JAGS

Plot of simulated data



Linear fits with uncertain σ in JAGS

Plot of simulated data



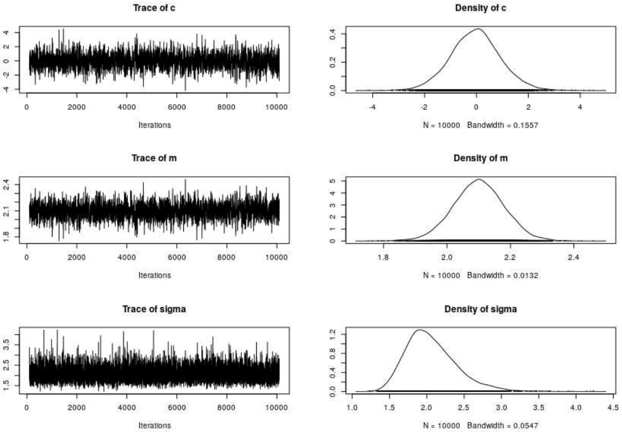
Calling JAGS

```
ns=10000
jm <- jags.model(model, data, inits)
update(jm, 100)
chain <- coda.samples(jm, c("c","m","sigma"), n.iter=ns)
```

Linear fits with uncertain σ in JAGS

$\Rightarrow$ `linear_fit.R`

JAGS summary



$c = -0.04 \pm 0.96$; $m = 2.10 \pm 0.08$; $\sigma = 2.06 \pm 0.34$

Linear fits with uncertain σ in JAGS

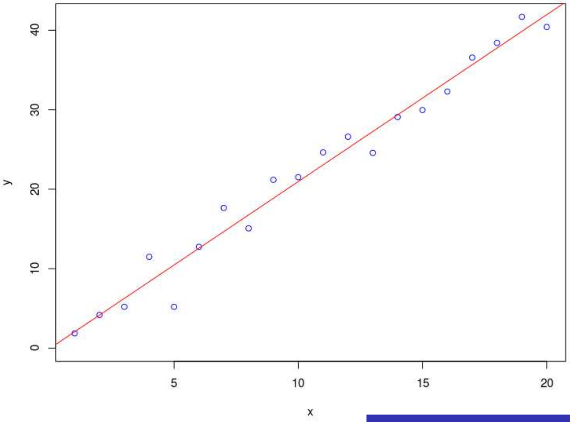
‘Check’ the result

```
c     <- as.vector(chain[[1]][,1])
m     <- as.vector(chain[[1]][,2])
sigma <- as.vector(chain[[1]][,3])
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red')
```

Linear fits with uncertain σ in JAGS

'Check' the result

```
c     <- as.vector(chain[[1]][,1])
m     <- as.vector(chain[[1]][,2])
sigma <- as.vector(chain[[1]][,3])
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red')
```



Linear fits with uncertain σ in JAGS

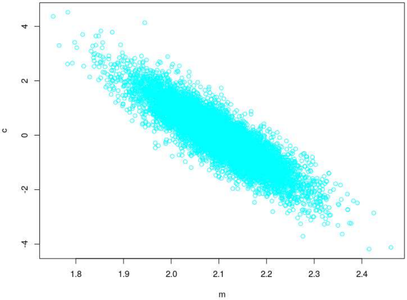
Correlation between m and c

```
plot(m,c,col='cyan')
cat(sprintf("rho(m,x) = %.3f\n", cor(m,c) ))
```

Linear fits with uncertain σ in JAGS

Correlation between m and c

```
plot(m,c,col='cyan')
cat(sprintf("rho(m,x) = %.3f\n", cor(m,c) ))
```



$$\rho(m, c) = -0.88$$

Linear fits with uncertain σ in JAGS

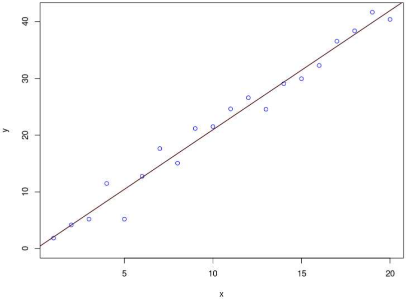
Check with R `lm()` (least square)

```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```

Linear fits with uncertain σ in JAGS

Check with `R` `lm()` (least square)

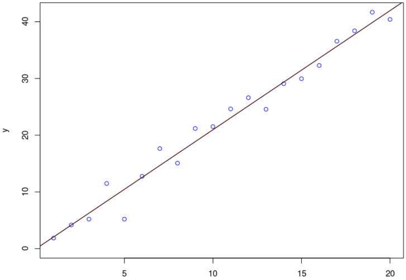
```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```



Linear fits with uncertain σ in JAGS

Check with R `lm()` (least square)

```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```

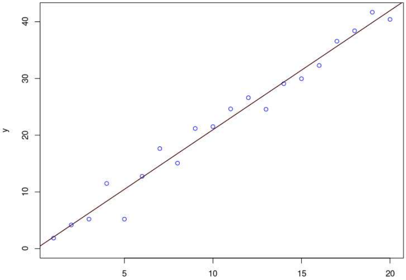


Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result

Linear fits with uncertain σ in JAGS

Check with `R` `lm()` (least square)

```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```

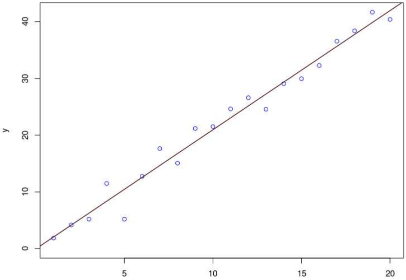


Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time?

Linear fits with uncertain σ in JAGS

Check with `R lm()` (least square)

```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```

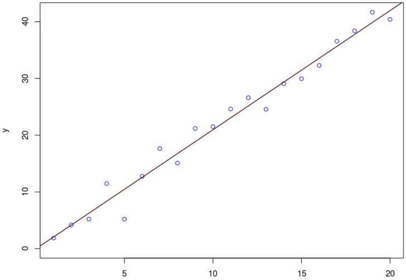


Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends. . .**

Linear fits with uncertain σ in JAGS

Check with R `lm()` (least square)

```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```



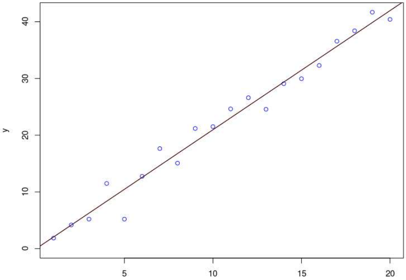
Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends. . .**

If the purpose was just to get an idea of the trend, then drawing a line with pencil and ruler would have been enough

Linear fits with uncertain σ in JAGS

Check with `R lm()` (least square)

```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```



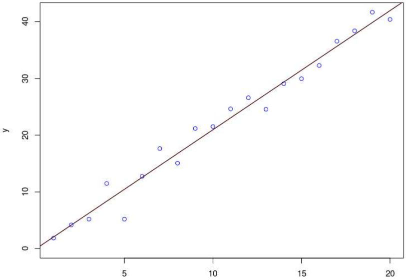
Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends...**

If the purpose was just to get an idea of the trend, then drawing a line with pencil and ruler would have been enough (as suggested to students of Circuit Lab)

Linear fits with uncertain σ in JAGS

Check with `R` `lm()` (least square)

```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```



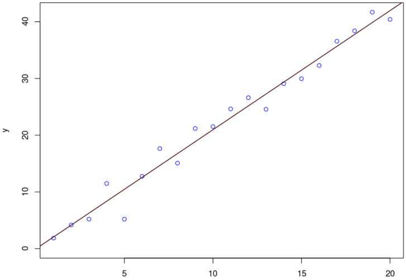
Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends...**

If the purpose was just to get an idea of the trend, then drawing a line with pencil and ruler would have been enough (as suggested to students of Circuit Lab): m and c ≈OK

Linear fits with uncertain σ in JAGS

Check with `R lm()` (least square)

```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```



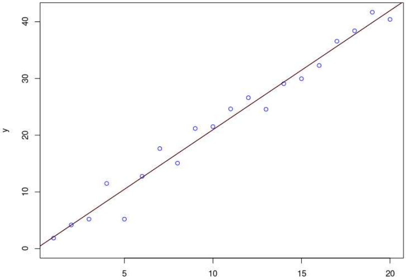
Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends...**

If the purpose was just to get an idea of the trend, then drawing a line with pencil and ruler would have been enough (as suggested to students of Circuit Lab): m and c ≈OK: **NO FIT: focus on circuits!**

Linear fits with uncertain σ in JAGS

Check with `R lm()` (least square)

```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```



Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends...**

If the purpose was just to get an idea of the trend, then drawing a line with pencil and ruler would have been enough (as suggested to students of Circuit Lab): m and c ≈OK: **NO FIT: focus on circuits!**

Otherwise: $\Rightarrow f(c, m, \sigma \,|\, \text{data points})$

Forecasting new μ_y and new y

Imagine we are interested at "y at $x_f = 30$" (referring to our 'data').

Forecasting new μ_y and new y

Imagine we are interested at "y at $x_f = 30$" (referring to our 'data').

- First at all it is important to distinguish

$$
\begin{aligned}
\mu_y(x_f) &\;\rightarrow\; \mu_y(\mu_{x_f}) \qquad \text{(no error on } x\text{)} \\
y(x_f) &\;\rightarrow\; y(\mu_{x_f})
\end{aligned}
$$

Forecasting new μ_y and new y

Imagine we are interested at "y at $x_f = 30$" (referring to our 'data').

- First at all it is important to distinguish

$$\begin{array}{rcll} \mu_y(x_f) & \rightarrow & \mu_y(\mu_{x_f}) & \text{(no error on } x\text{)} \\ y(x_f) & \rightarrow & y(\mu_{x_f}) & \end{array}$$

- Then we have to take into account all uncertainties, including correlations (not only the covariance matrix!)

Forecasting new μ_y and new y

Imagine we are interested at "y at $x_f = 30$" (referring to our 'data').

- First at all it is important to distinguish

$$\begin{aligned} \mu_y(x_f) &\rightarrow \mu_y(\mu_{x_f}) \quad \text{(no error on } x\text{)} \\ y(x_f) &\rightarrow y(\mu_{x_f}) \end{aligned}$$

- Then we have to take into account all uncertainties, including correlations (not only the covariance matrix!)

Our problem

$$f(\mu_{y_f}\,|\,\text{data}, x_f) = \int f(\mu_{y_f}\,|\,m, c, x_f) \cdot f(m, c\,|\,\text{data})\,\mathrm{d}c\,\mathrm{d}m$$

Forecasting new μ_y and new y

Imagine we are interested at "y at $x_f = 30$" (referring to our 'data').

- First at all it is important to distinguish

$$\begin{aligned} \mu_y(x_f) &\rightarrow \mu_y(\mu_{x_f}) \quad \text{(no error on } x\text{)} \\ y(x_f) &\rightarrow y(\mu_{x_f}) \end{aligned}$$

- Then we have to take into account all uncertainties, including correlations (not only the covariance matrix!)

Our problem

$$f(\mu_{y_f}\,|\,\text{data}, x_f) = \int f(\mu_{y_f}\,|\,m, c, x_f) \cdot f(m, c\,|\,\text{data})\,\mathrm{d}c\,\mathrm{d}m$$

$$f(y_f\,|\,\text{data}, x_f) = \int f(y_f\,|\,\mu_{y_f}) \cdot f(\mu_{y_f}\,|\,\text{data}, x_f)\,\mathrm{d}\mu_{y_f}$$

Forecasting new μ_y and new y

Including prediction in the JAGS model

```
var mu.y[N];
model{
    for (i in 1:N) {
      y[i] ~ dnorm(mu.y[i], tau);
      mu.y[i] <- x[i] * m + c;
    }
    mu.yf <- xf * m + c;      # future 'true value' for x=xf
    yf ~ dnorm(mu.yf, tau);   # future 'observation for x=xf
    c ~ dnorm(0, 1.0E-6);
    m ~ dnorm(0, 1.0E-6);
    tau ~ dgamma(1.0, 1.0E-6);
    sigma <- 1.0/sqrt(tau);
}
```

Forecasting new μ_y and new y

Including prediction in the JAGS model

```
var mu.y[N];
model{
    for (i in 1:N) {
      y[i] ~ dnorm(mu.y[i], tau);
      mu.y[i] <- x[i] * m + c;
    }
    mu.yf <- xf * m + c;      # future 'true value' for x=xf
    yf ~ dnorm(mu.yf, tau);  # future 'observation for x=xf
    c ~ dnorm(0, 1.0E-6);
    m ~ dnorm(0, 1.0E-6);
    tau ~ dgamma(1.0, 1.0E-6);
    sigma <- 1.0/sqrt(tau);
}
```

Or we can do the 'integral' by sampling, using the MCMC histories of the quantities of interest
(see previous model, without prediction)

Forecasting new μ_y and new y

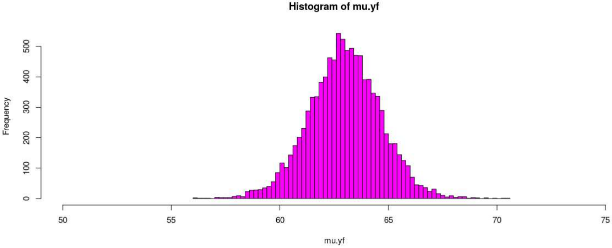
Including prediction in the JAGS model

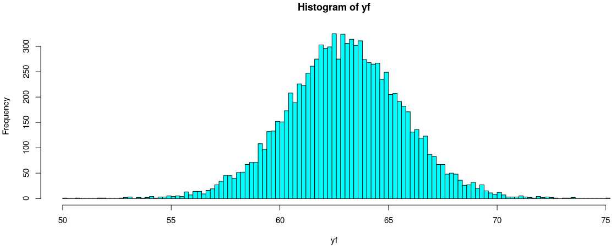
```
var mu.y[N];
model{
    for (i in 1:N) {
      y[i] ~ dnorm(mu.y[i], tau);
      mu.y[i] <- x[i] * m + c;
    }
    mu.yf <- xf * m + c;      # future 'true value' for x=xf
    yf ~ dnorm(mu.yf, tau);   # future 'observation for x=xf
    c ~ dnorm(0, 1.0E-6);
    m ~ dnorm(0, 1.0E-6);
    tau ~ dgamma(1.0, 1.0E-6);
    sigma <- 1.0/sqrt(tau);
}
```

Or we can do the 'integral' by sampling, using the MCMC histories of the quantities of interest
(see previous model, without prediction)

$\Rightarrow$ Left as exercise

Forecasting new μ_y and new y with JAGS





$\mu_y(x = 30) = 63.0 \pm 1.7$; $\quad y(x = 30) = 63.0 \pm 2.7$

Try with Root ;-) ['data' on the web site]

The End

Appendix on small samples

Inferring μ and σ from a sample

(Gaussian, independent observations)

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

Inferring μ and σ from a sample

(Gaussian, independent observations)

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

Inferring μ and σ from a sample

(Gaussian, independent observations)

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

with $s^2 = \overline{x^2} - \overline{x}^2$, variance of the sample.

Inferring μ and σ from a sample

(Gaussian, independent observations)

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

$$\propto\; \sigma^{-n}\exp\left[-\frac{\overline{x^2}-\overline{x}^2+\overline{x}^2-2\,\mu\,\overline{x}+\mu^2}{2\,\sigma^2/n}\right]\cdot f_0(\mu,\sigma)$$

$$\propto\; \sigma^{-n}\exp\left[-\frac{s^2+(\mu-\overline{x})^2}{2\,\sigma^2/n}\right]\cdot f_0(\mu,\sigma)$$

with $s^2=\overline{x^2}-\overline{x}^2$, variance of the sample.

- the inference on μ and σ depends only on s^2 and $\overline{x}$ (and on the priors, as it has to be!).

Inferring μ and σ from a sample

(Gaussian, independent observations)

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

$$\propto\; \sigma^{-n}\exp\left[-\frac{\overline{x^2}-\overline{x}^2+\overline{x}^2-2\,\mu\,\overline{x}+\mu^2}{2\,\sigma^2/n}\right]\cdot f_0(\mu,\sigma)$$

$$\propto\; \sigma^{-n}\exp\left[-\frac{s^2+(\mu-\overline{x})^2}{2\,\sigma^2/n}\right]\cdot f_0(\mu,\sigma)$$

with $s^2 = \overline{x^2}-\overline{x}^2$, variance of the sample.

- the inference on μ and σ depends only on s^2 and $\overline{x}$ (and on the priors, as it has to be!).
- Evaluate $f(\mu,\sigma\,|\,\overline{x},s)$ and then

$$f(\mu\,|\,\overline{x},s) \;=\; \int_0^\infty f(\mu,\sigma\,|\,\overline{x},s)\,d\sigma$$

Inferring μ and σ from a sample

(Gaussian, independent observations)

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

with $s^2=\overline{x^2}-\overline{x}^2$, variance of the sample.

- the inference on μ and σ depends only on s^2 and $\overline{x}$ (and on the priors, as it has to be!).
- Evaluate $f(\mu,\sigma\,|\,\overline{x},s)$ and then

$$
\begin{aligned}
f(\mu\,|\,\overline{x},s) &= \int_0^{\infty} f(\mu,\sigma\,|\,\overline{x},s)\,d\sigma\\
f(\sigma\,|\,\overline{x},s) &= \int_{-\infty}^{+\infty} f(\mu,\sigma\,|\,\overline{x},s)\,d\mu
\end{aligned}
$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on σ)

$$f(\mu, \sigma \,|\, \underline{x}) \;\; \propto \;\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right]$$

Marginalizing[1]

$$f(\mu \,|\, \underline{x}) \;\; = \;\; \int_0^{\infty} f(\mu, \sigma \,|\, \underline{x})\, \mathrm{d}\sigma$$

[1]The integral of interest is

$$\int_0^{\infty} z^{-n} \exp\left[-\frac{c}{2\,z^2}\right] \mathrm{d}z = 2^{(n-3)/2}\, \Gamma\left[\frac{1}{2}(n-1)\right] c^{-(n-1)/2}.$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on σ)

$$f(\mu, \sigma \,|\, \underline{x}) \;\propto\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right]$$

Marginalizing[1]

$$\begin{aligned} f(\mu \,|\, \underline{x}) &= \int_0^\infty f(\mu, \sigma \,|\, \underline{x})\, \mathrm{d}\sigma \\ &\propto \left((\overline{x} - \mu)^2 + s^2\right)^{-(n-1)/2} \end{aligned}$$

[1] The integral of interest is

$$\int_0^\infty z^{-n} \exp\left[-\frac{c}{2\,z^2}\right] \mathrm{d}z = 2^{(n-3)/2}\,\Gamma\left[\frac{1}{2}(n-1)\right] c^{-(n-1)/2}.$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on σ)

$$f(\mu, \sigma \,|\, \underline{x}) \;\propto\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right]$$

Marginalizing[1]

$$\begin{aligned} f(\mu \,|\, \underline{x}) &= \int_0^\infty f(\mu, \sigma \,|\, \underline{x})\, \mathrm{d}\sigma \\ &\propto \left((\overline{x} - \mu)^2 + s^2\right)^{-(n-1)/2} \\ &\propto \left(1 + \frac{(\mu - \overline{x})^2}{s^2}\right)^{-(n-1)/2} \end{aligned}$$

[1] The integral of interest is

$$\int_0^\infty z^{-n} \exp\left[-\frac{c}{2\,z^2}\right] \mathrm{d}z = 2^{(n-3)/2}\,\Gamma\left[\frac{1}{2}(n-1)\right] c^{-(n-1)/2}.$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on σ)

$$f(\mu\,|\,\underline{x}) \;\propto\; \left(1+\frac{(\mu-\overline{x})^2}{s^2}\right)^{-(n-1)/2} \quad ??$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on σ)

$$
\begin{aligned}
f(\mu \,|\, \underline{x}) &\propto \left(1 + \frac{(\mu - \overline{x})^2}{s^2}\right)^{-(n-1)/2} \quad ?? \\
&\propto \left(1 + \frac{(\mu - \overline{x})^2}{(n-2)\, s^2/(n-2)}\right)^{-((n-2)+1)/2}
\end{aligned}
$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on σ)

$$
\begin{aligned}
f(\mu \,|\, \underline{x}) &\propto \left(1 + \frac{(\mu - \overline{x})^2}{s^2}\right)^{-(n-1)/2} \quad ?? \\
&\propto \left(1 + \frac{(\mu - \overline{x})^2}{(n-2)\, s^2/(n-2)}\right)^{-((n-2)+1)/2} \\
&\propto \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}
\end{aligned}
$$

with

$$
\begin{aligned}
\nu &= n - 2 \\
t &= \frac{\mu - \overline{x}}{s/\sqrt{n-2}}\,,
\end{aligned}
$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on σ)

$$
\begin{aligned}
f(\mu\,|\,\underline{x}) &\propto \left(1+\frac{(\mu-\overline{x})^2}{s^2}\right)^{-(n-1)/2} \quad ?? \\
&\propto \left(1+\frac{(\mu-\overline{x})^2}{(n-2)\,s^2/(n-2)}\right)^{-((n-2)+1)/2} \\
&\propto \left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}
\end{aligned}
$$

with

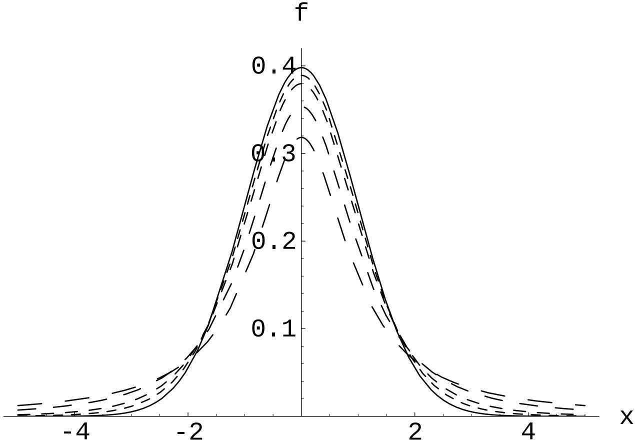
$$
\begin{aligned}
\nu &= n-2 \\
t &= \frac{\mu-\overline{x}}{s/\sqrt{n-2}}\,,
\end{aligned}
$$

that is

$$
\mu = \overline{x} + \frac{s}{\sqrt{n-2}}\,t\,,
$$

where t is a "Student t" with $\nu = n-2$

Student t



Examples of Student t for ν equal to 1 , 2, 5, 10 and 100 ($\approx$ "∞").

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

In summary,

$$\frac{\mu - \overline{x}}{s/\sqrt{n-2}} \quad \sim \quad \text{Student}(\nu = n-2)$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

In summary,

$$\frac{\mu - \overline{x}}{s/\sqrt{n-2}} \quad \sim \quad \text{Student}(\nu = n-2)$$

$$\text{E}(\mu) \quad \overset{(n\geq 3)}{=} \quad \overline{x}$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

In summary,

$$
\begin{aligned}
\frac{\mu - \overline{x}}{s/\sqrt{n-2}} \quad &\sim \quad \text{Student}(\nu = n-2) \\
\text{E}(\mu) \quad &\stackrel{(n\geq 3)}{=} \quad \overline{x} \\
\sigma(\mu) \quad &\stackrel{(n\geq 4)}{=} \quad \frac{s}{\sqrt{n-4}}\,.
\end{aligned}
$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

In summary,

$$\begin{aligned}\frac{\mu-\overline{x}}{s/\sqrt{n-2}} &\sim \text{Student}(\nu=n-2)\\ \text{E}(\mu) &\stackrel{(n>3)}{=} \overline{x}\\ \sigma(\mu) &\stackrel{(n>4)}{=} \frac{s}{\sqrt{n-4}}\,.\end{aligned}$$

The uncertainty on σ increases the probability of the values of μ far from $\overline{x}$:

- not only the standard uncertainty increases, but the distribution itself changes and, as 'well know' the t distribution has 'higher' tails.

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

In summary,

$$\frac{\mu - \overline{x}}{s/\sqrt{n-2}} \sim \text{Student}(\nu = n-2)$$

$$\text{E}(\mu) \stackrel{(n>3)}{=} \overline{x}$$

$$\sigma(\mu) \stackrel{(n>4)}{=} \frac{s}{\sqrt{n-4}}\,.$$

The uncertainty on σ increases the probability of the values of μ far from $\overline{x}$:

- not only the standard uncertainty increases, but the distribution itself changes and, as 'well know' the t distribution has 'higher' tails.

However, when n is very large the Gaussian distribution is recovered (the t-distribution tends to a gaussian), with $\sigma(\mu) = s/\sqrt{n}$.

Expected value and variance only exist above certain values of n:

$$\begin{aligned} \mathrm{E}(\mu) &\overset{(n>3)}{=} \overline{x} \\ \sigma(\mu) &\overset{(n>4)}{=} \frac{s}{\sqrt{n-4}}\,. \end{aligned}$$

[2] Flat priors are good for teaching purposes, but when the result hurts with our beliefs it means we have to use priors that match with previous knowledge.

Inferring μ and σ from a sample

Misunderstandings and 'myths' related to the Student t distribution

Expected value and variance only exist above certain values of n:

$$\begin{aligned} \mathrm{E}(\mu) &\overset{(n>3)}{=} \overline{x} \\ \sigma(\mu) &\overset{(n>4)}{=} \frac{s}{\sqrt{n-4}} \,. \end{aligned}$$

So what?

[2] Flat priors are good for teaching purposes, but when the result hurts with our beliefs it means we have to use priors that match with previous knowledge.

Inferring μ and σ from a sample

Expected value and variance only exist above certain values of n:

$$\begin{aligned} \mathrm{E}(\mu) &\stackrel{(n>3)}{=} \overline{x} \\ \sigma(\mu) &\stackrel{(n>4)}{=} \frac{s}{\sqrt{n-4}}\,. \end{aligned}$$

So what?

It is just a reflex of the fact that we have used, for lazyness,[2] priors which are indeed absurd.

[2] Flat priors are good for teaching purposes, but when the result hurts with our beliefs it means we have to use priors that match with previous knowledge.

Inferring μ and σ from a sample

Expected value and variance only exist above certain values of n:

$$\begin{aligned} \mathrm{E}(\mu) &\overset{(n>3)}{=} \overline{x} \\ \sigma(\mu) &\overset{(n>4)}{=} \frac{s}{\sqrt{n-4}}\,. \end{aligned}$$

So what?

It is just a reflex of the fact that we have used, for lazyness,[2] priors which are indeed absurd.

- In no measurement we beleive that μ and/or σ could be 'infinite'.

[2] Flat priors are good for teaching purposes, but when the result hurts with our beliefs it means we have to use priors that match with previous knowledge.

Inferring μ and σ from a sample

Misunderstandings and 'myths' related to the Student t distribution

Expected value and variance only exist above certain values of n:

$$\begin{aligned} E(\mu) &\overset{(n>3)}{=} \overline{x} \\ \sigma(\mu) &\overset{(n>4)}{=} \frac{s}{\sqrt{n-4}} \,. \end{aligned}$$

So what?

It is just a reflex of the fact that we have used, for lazyness,[2] priors which are indeed absurd.

- In no measurement we beleive that μ and/or σ could be 'infinite'.
- Just plug in some reasonable, although very vagues, proper priors, and the problem disappears.

---

[2]Flat priors are good for teaching purposes, but when the result hurts with our beliefs it means we have to use priors that match with previous knowledge.

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

- Large n limit:

$$
\begin{aligned}
\mathrm{E}(\mu) &\xrightarrow{n\to\infty} \overline{x} \\
\sigma(\mu) &\xrightarrow{n\to\infty} \frac{s}{\sqrt{n}} \\
\mu &\xrightarrow{n\to\infty} \sim \mathcal{N}(\overline{x}, \frac{s}{\sqrt{n}}).
\end{aligned}
$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Marginal $f(\sigma)$

$$f(\sigma\,|\,\overline{x},s) \;=\; \int_{-\infty}^{+\infty} f(\mu,\sigma\,|\,\overline{x},s)\,d\mu$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Marginal $f(\sigma)$

$$
\begin{aligned}
f(\sigma\,|\,\overline{x}, s) &= \int_{-\infty}^{+\infty} f(\mu, \sigma\,|\,\overline{x}, s)\, d\mu \\
&\propto \sigma^{-n} \exp\left[-\frac{n\, s^2}{2\,\sigma^2}\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{n\,(\overline{x}-\mu)^2}{2\sigma^2}\right] \mathrm{d}\mu
\end{aligned}
$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Marginal $f(\sigma)$

$$
\begin{aligned}
f(\sigma\,|\,\overline{x},s) &= \int_{-\infty}^{+\infty} f(\mu,\sigma\,|\,\overline{x},s)\,d\mu \\
&\propto \sigma^{-n}\exp\left[-\frac{n\,s^2}{2\,\sigma^2}\right]\int_{-\infty}^{+\infty}\exp\left[-\frac{n\,(\overline{x}-\mu)^2}{2\sigma^2}\right]\mathrm{d}\mu \\
&\propto \sigma^{-(n-1)}\exp\left[-\frac{n\,s^2}{2\,\sigma^2}\right]
\end{aligned}
$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Marginal $f(\sigma)$

$$
\begin{aligned}
f(\sigma \,|\, \overline{x}, s) &= \int_{-\infty}^{+\infty} f(\mu, \sigma \,|\, \overline{x}, s)\, d\mu \\
&\propto \sigma^{-n} \exp\left[-\frac{n\, s^2}{2\, \sigma^2}\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{n\,(\overline{x}-\mu)^2}{2\sigma^2}\right] \mathrm{d}\mu \\
&\propto \sigma^{-(n-1)} \exp\left[-\frac{n\, s^2}{2\, \sigma^2}\right]
\end{aligned}
$$

That is. . . (no special function)

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Marginal $f(\sigma)$

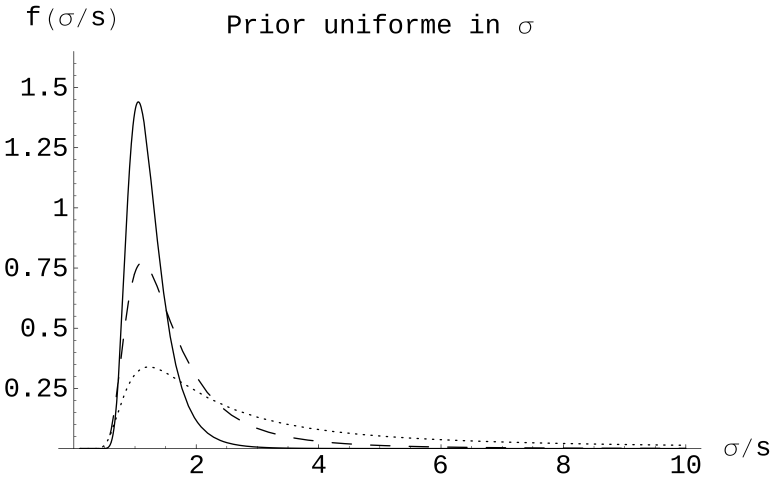
$$
\begin{aligned}
f(\sigma\,|\,\overline{x}, s) &= \int_{-\infty}^{+\infty} f(\mu, \sigma\,|\,\overline{x}, s)\, d\mu \\
&\propto \sigma^{-n} \exp\left[-\frac{n\,s^2}{2\,\sigma^2}\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{n\,(\overline{x}-\mu)^2}{2\sigma^2}\right] \mathrm{d}\mu \\
&\propto \sigma^{-(n-1)} \exp\left[-\frac{n\,s^2}{2\,\sigma^2}\right]
\end{aligned}
$$

That is... (no special function)

[But if we would use $\tau = 1/\sigma^2$ we would recognize a Gamma...]

Inferring μ and σ from a sample

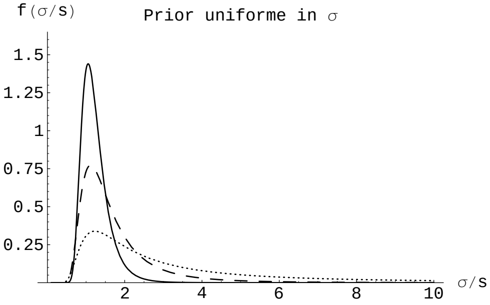
(Gaussian, independent observations – prior uniform on *mu* and σ)



$n = 3$ (dotted), $n = 5$ (dashed) e $n = 10$ (continous).

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)



$$\mathrm{E}(\sigma) \xrightarrow[n\to\infty]{} s$$

$$\mathrm{std}(\sigma) \xrightarrow[n\to\infty]{} \frac{s}{\sqrt{2\,n}}$$

$$\sigma \xrightarrow[n\to\infty]{} \sim \mathcal{N}(s, \frac{s}{\sqrt{2\,n}})\,.$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Using the "Gaussian trick"

$$\varphi(\mu,\sigma) \quad = \quad n \ln\sigma + \frac{s^2 + (\mu - \overline{x})^2)}{2\,\sigma/n}$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Using the "Gaussian trick"

$$\varphi(\mu,\sigma) \;=\; n\ln\sigma + \frac{s^2 + (\mu-\overline{x})^2)}{2\,\sigma/n}$$

First derivatives:

$$\frac{\partial\varphi}{\partial\mu} \;=\; \frac{\mu-\overline{x}}{\sigma/n}$$

$$\frac{\partial\varphi}{\partial\sigma} \;=\; \frac{n}{\sigma} - \frac{n\,s^2}{\sigma^3}$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Using the "Gaussian trick"

$$\varphi(\mu,\sigma) \quad = \quad n \ln \sigma + \frac{s^2 + (\mu - \overline{x})^2)}{2\,\sigma/n}$$

First derivatives:

$$\begin{aligned} \frac{\partial\varphi}{\partial\mu} &= \frac{\mu - \overline{x}}{\sigma/n} \\ \frac{\partial\varphi}{\partial\sigma} &= \frac{n}{\sigma} - \frac{n\,s^2}{\sigma^3} \end{aligned}$$

From which it follows (equating the derivatives to zero)

$$\begin{aligned} \mathrm{E}(\mu) &= \overline{x} \\ \mathrm{E}(\sigma) &= s \end{aligned}$$

(They are indeed the modes!)

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Hessian calculated at $\mu = \overline{x}$ and $\sigma = s$ (hereafter 'm'):

$$\left.\frac{\partial^2\varphi}{\partial\mu^2}\right|_m = \left.\frac{n}{\sigma^2}\right|_m = \frac{n}{s^2}$$

$$\left.\frac{\partial^2\varphi}{\partial\sigma^2}\right|_m = \left.\left(-\frac{n}{\sigma^2} + \frac{3\,(s^2 + (\mu-\overline{x})^2)}{\sigma^4/n}\right)\right|_m = \frac{2\,n}{s^2}$$

$$\left.\frac{\partial^2\varphi}{\partial\mu\partial\sigma}\right|_m = \left.\frac{-2\,(\mu-\overline{x})}{\sigma^3/n}\right|_m = 0$$

$$\left.\frac{\partial^2\varphi}{\partial\sigma\partial\mu}\right|_m = \left.\frac{-2\,(\mu-\overline{x})}{\sigma^3/n}\right|_m = 0$$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Hessian calculated at $\mu = \overline{x}$ and $\sigma = s$ (hereafter 'm'):

$$\begin{aligned}
\left.\frac{\partial^2\varphi}{\partial\mu^2}\right|_m &= \left.\frac{n}{\sigma^2}\right|_m = \frac{n}{s^2} \\
\left.\frac{\partial^2\varphi}{\partial\sigma^2}\right|_m &= \left.\left(-\frac{n}{\sigma^2} + \frac{3\,(s^2 + (\mu-\overline{x})^2)}{\sigma^4/n}\right)\right|_m = \frac{2\,n}{s^2} \\
\left.\frac{\partial^2\varphi}{\partial\mu\partial\sigma}\right|_m &= \left.\frac{-2\,(\mu-\overline{x})}{\sigma^3/n}\right|_m = 0 \\
\left.\frac{\partial^2\varphi}{\partial\sigma\partial\mu}\right|_m &= \left.\frac{-2\,(\mu-\overline{x})}{\sigma^3/n}\right|_m = 0
\end{aligned}$$

It follows

$$\begin{aligned}
\mathrm{std}(\mu) &= \frac{s}{\sqrt{n}} \\
\mathrm{std}(\sigma) &= \frac{s}{\sqrt{2\,n}}\,,
\end{aligned}$$

reobtaining the large number limit.

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Hessian calculated at $\mu = \overline{x}$ and $\sigma = s$ (hereafter 'm'):

$$\left.\frac{\partial^2\varphi}{\partial\mu^2}\right|_m = \left.\frac{n}{\sigma^2}\right|_m = \frac{n}{s^2}$$

$$\left.\frac{\partial^2\varphi}{\partial\sigma^2}\right|_m = \left.\left(-\frac{n}{\sigma^2} + \frac{3\,(s^2 + (\mu - \overline{x})^2)}{\sigma^4/n}\right)\right|_m = \frac{2\,n}{s^2}$$

$$\left.\frac{\partial^2\varphi}{\partial\mu\partial\sigma}\right|_m = \left.\frac{-2\,(\mu - \overline{x})}{\sigma^3/n}\right|_m = 0$$

$$\left.\frac{\partial^2\varphi}{\partial\sigma\partial\mu}\right|_m = \left.\frac{-2\,(\mu - \overline{x})}{\sigma^3/n}\right|_m = 0$$

It follows

$$\mathrm{std}(\mu) = \frac{s}{\sqrt{n}}$$

$$\mathrm{std}(\sigma) = \frac{s}{\sqrt{2\,n}}\,,$$

reobtaining the large number limit. And, notice, $\rho(\mu, \sigma) = 0\,.$

Inferring μ and σ from a sample

(Gaussian, independent observations – prior uniform on μ and σ)

Hessian calculated at $\mu = \overline{x}$ and $\sigma = s$ (hereafter 'm'):

$$\left.\frac{\partial^2 \varphi}{\partial \mu^2}\right|_m = \left.\frac{n}{\sigma^2}\right|_m = \frac{n}{s^2}$$

$$\left.\frac{\partial^2 \varphi}{\partial \sigma^2}\right|_m = \left.\left(-\frac{n}{\sigma^2} + \frac{3\,(s^2 + (\mu - \overline{x})^2)}{\sigma^4/n}\right)\right|_m = \frac{2\,n}{s^2}$$

$$\left.\frac{\partial^2 \varphi}{\partial \mu \partial \sigma}\right|_m = \left.\frac{-2\,(\mu - \overline{x})}{\sigma^3/n}\right|_m = 0$$

$$\left.\frac{\partial^2 \varphi}{\partial \sigma \partial \mu}\right|_m = \left.\frac{-2\,(\mu - \overline{x})}{\sigma^3/n}\right|_m = 0$$

It follows

$$\text{std}(\mu) = \frac{s}{\sqrt{n}}$$

$$\text{std}(\sigma) = \frac{s}{\sqrt{2\,n}}\,,$$

reobtaining the large number limit. And, notice, $\rho(\mu, \sigma) = 0$.

Q.: Are they independent?

Inferring μ and σ from a sample

(Gaussian, independent observations. Expression the Gaussian in terms of $\tau = 1/\sigma^2$)

$$f(\mu, \sigma \,|\, \underline{x}) \;\propto\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

Inferring μ and σ from a sample

(Gaussian, independent observations. Expression the Gaussian in terms of $\tau = 1/\sigma^2$)

$$f(\mu, \sigma \,|\, \underline{x}) \;\propto\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

It is technically convenient to use $\tau = 1/\sigma^2$:

$$f(\mu, \tau \,|\, \underline{x}) \;\propto\; \tau^{n/2} \exp\left[-\frac{n\,\tau}{2}\,\left(s^2 + (\mu - \overline{x})^2\right)\right] \cdot f_0(\mu, \tau)$$

Inferring μ and σ from a sample

(Gaussian, independent observations. Expression the Gaussian in terms of $\tau = 1/\sigma^2$)

$$f(\mu, \sigma \,|\, \underline{x}) \;\; \propto \;\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

It is technically convenient to use $\tau = 1/\sigma^2$:

$$f(\mu, \tau \,|\, \underline{x}) \;\; \propto \;\; \tau^{n/2} \exp\left[-\frac{n\,\tau}{2}\,\left(s^2 + (\mu - \overline{x})^2\right)\right] \cdot f_0(\mu, \tau)$$

For a fixed μ (and observed s and $\overline{x}$)

$$f(\tau \,|\, \underline{x}, \mu) \;\; \propto \;\; \tau^{\alpha}\, e^{-\beta\,\tau} \cdot f_0(\tau)$$

Inferring μ and σ from a sample

(Gaussian, independent observations. Expression the Gaussian in terms of $\tau = 1/\sigma^2$)

$$f(\mu, \sigma \,|\, \underline{x}) \;\;\propto\;\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

It is technically convenient to use $\tau = 1/\sigma^2$:

$$f(\mu, \tau \,|\, \underline{x}) \;\;\propto\;\; \tau^{n/2} \exp\left[-\frac{n\,\tau}{2}\,\left(s^2 + (\mu - \overline{x})^2\right)\right] \cdot f_0(\mu, \tau)$$

For a fixed μ (and observed s and $\overline{x}$)

$$f(\tau \,|\, \underline{x}, \mu) \;\;\propto\;\; \tau^{\alpha}\, e^{-\beta\,\tau} \cdot f_0(\tau)$$

Do you recongnize a *famous* mathematical form?

Inferring μ and σ from a sample

(Gaussian, independent observations. Expression the Gaussian in terms of $\tau = 1/\sigma^2$)

$$f(\mu, \sigma \,|\, \underline{x}) \;\;\propto\;\; \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

It is technically convenient to use $\tau = 1/\sigma^2$:

$$f(\mu, \tau \,|\, \underline{x}) \;\;\propto\;\; \tau^{n/2} \exp\left[-\frac{n\,\tau}{2}\,\left(s^2 + (\mu - \overline{x})^2\right)\right] \cdot f_0(\mu, \tau)$$

For a fixed μ (and observed s and $\overline{x}$)

$$f(\tau \,|\, \underline{x}, \mu) \;\;\propto\;\; \tau^{\alpha}\, e^{-\beta\,\tau} \cdot f_0(\tau)$$

Do you recongnize a *famous* mathematical form?

On the other way around, for a fixed τ,

$$f(\mu \,|\, \underline{x}, \tau) \;\;\propto\;\; \exp\left[-\frac{n\,\tau}{2}\,(\mu - \overline{x})^2\right] \cdot f_0(\mu)$$

Inferring μ and σ from a sample

(Gaussian, independent observations. Expression the Gaussian in terms of $\tau = 1/\sigma^2$)

$$f(\mu, \sigma \,|\, \underline{x}) \quad \propto \quad \sigma^{-n} \exp\left[-\frac{s^2 + (\mu - \overline{x})^2}{2\,\sigma^2/n}\right] \cdot f_0(\mu, \sigma)$$

It is technically convenient to use $\tau = 1/\sigma^2$:

$$f(\mu, \tau \,|\, \underline{x}) \quad \propto \quad \tau^{n/2} \exp\left[-\frac{n\,\tau}{2}\,\left(s^2 + (\mu - \overline{x})^2\right)\right] \cdot f_0(\mu, \tau)$$

For a fixed μ (and observed s and $\overline{x}$)

$$f(\tau \,|\, \underline{x}, \mu) \quad \propto \quad \tau^{\alpha}\, e^{-\beta\,\tau} \cdot f_0(\tau)$$

Do you recongnize a *famous* mathematical form?

On the other way around, for a fixed τ,

$$f(\mu \,|\, \underline{x}, \tau) \quad \propto \quad \exp\left[-\frac{n\,\tau}{2}\,(\mu - \overline{x})^2\right] \cdot f_0(\mu)$$

$\Rightarrow$ Gibbs sampling

Practical introduction to BUGS

- Introducing the *bug* language to build up the models.
- Running the model (including data and 'inits') in the OpenBUGS GUI.
- Analysing the resulting chain in R.