Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

An important clarification

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

An important clarification

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

An important clarification

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

Objection: "n, x are certain $\Rightarrow$ $f(n, x) = 1$"

An important clarification

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

Objection: "n, x are certain $\Rightarrow$ $f(n,x) = 1$"

FALSE!

An important clarification

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

Objection: "n, x are certain $\Rightarrow$ $f(n,x) = 1$"

FALSE!

Remember the six boxes (**Att:**:

$$P(H_i\,|\,W) = \frac{P(H_i \cap W)}{P(W)}$$

An important clarification

$$f(p \mid n, x) = \frac{f(p, n, x)}{f(n, x)}$$
$$\propto f(p, n, x)$$

Objection: "n, x are certain $\Rightarrow$ $f(n, x) = 1$"

FALSE!

Remember the six boxes (**Att:**:

$$P(H_i \mid W) = \frac{P(H_i \cap W)}{P(W)} = \frac{P(W \mid H_i) \cdot P(H_i)}{P(W)}$$

An important clarification

$$f(p \mid n, x) = \frac{f(p, n, x)}{f(n, x)} \propto f(p, n, x)$$

Objection: "n, x are certain $\Rightarrow$ $f(n, x) = 1$"

FALSE!

Remember the six boxes (**Att:**:

$$P(H_i \mid W) = \frac{P(H_i \cap W)}{P(W)} = \frac{P(W \mid H_i) \cdot P(H_i)}{P(W)}$$

In this case $P(W) < 1$: probability of W before it was observed!

Reconditioning on a certain event?

$$P(H_i \,|\, \Omega) \;=\; \frac{P(H_i \cap \Omega)}{P(\Omega)}$$

Reconditioning on a certain event?

$$P(H_i \mid \Omega) = \frac{P(H_i \cap \Omega)}{P(\Omega)}$$

$$= \frac{P(H_i)}{1} = P(H_i)$$

Reconditioning on a certain event?

$$\begin{aligned} P(H_i \,|\, \Omega) &= \frac{P(H_i \cap \Omega)}{P(\Omega)} \\ &= \frac{P(H_i)}{1} = P(H_i) \end{aligned}$$

$\Rightarrow$ Monthy Hall ('three boxes, version 2')

Reconditioning on a certain event?

$$P(H_i \,|\, \Omega) = \frac{P(H_i \cap \Omega)}{P(\Omega)} = \frac{P(H_i)}{1} = P(H_i)$$

⇒ Monthy Hall ('three boxes, version 2')

- if you trust the quiz master, then "no prize in the box that he is going to open" is a certain event;

Reconditioning on a certain event?

$$P(H_i \,|\, \Omega) = \frac{P(H_i \cap \Omega)}{P(\Omega)} = \frac{P(H_i)}{1} = P(H_i)$$

⇒ Monthy Hall (‘three boxes, version 2’)

- if you trust the quiz master, then “no prize in the box that he is going to open” is a certain event;
- ⇒ the probability of finding the prize in your box remains 1/3.

Reconditioning on a certain event?

$$\begin{aligned} P(H_i \,|\, \Omega) &= \frac{P(H_i \cap \Omega)}{P(\Omega)} \\ &= \frac{P(H_i)}{1} = P(H_i) \end{aligned}$$

⇒ Monthy Hall ('three boxes, version 2')

- if you trust the quiz master, then "no prize in the box that he is going to open" is a certain event;
- ⇒ the probability of finding the prize in your box remains $1/3$.

The situation is quite different when there are two players

Reconditioning on a certain event?

$$P(H_i \,|\, \Omega) = \frac{P(H_i \cap \Omega)}{P(\Omega)} = \frac{P(H_i)}{1} = P(H_i)$$

$\Rightarrow$ Monthy Hall ('three boxes, version 2')

- if you trust the quiz master, then "no prize in the box that he is going to open" is a certain event;
- $\Rightarrow$ the probability of finding the prize in your box remains $1/3$.

The situation is quite different when there are two players

- *there was* $2/3$ probability that the other player would not find the prize in the box;

Reconditioning on a certain event?

$$
\begin{aligned}
P(H_i \mid \Omega) &= \frac{P(H_i \cap \Omega)}{P(\Omega)} \\
&= \frac{P(H_i)}{1} = P(H_i)
\end{aligned}
$$

$\Rightarrow$ Monthy Hall ('three boxes, version 2')

- if you trust the quiz master, then "no prize in the box that he is going to open" is a certain event;
- $\Rightarrow$ the probability of finding the prize in your box remains $1/3$.

The situation is quite different when there are two players

- *there was* $2/3$ probability that the other player would not find the prize in the box;
- $\Rightarrow$ the probability of finding the prize in your box rises to $1/2$.

Reconditioning on a certain event?

$$P(H_i\,|\,\Omega) = \frac{P(H_i\cap\Omega)}{P(\Omega)} = \frac{P(H_i)}{1} = P(H_i)$$

⇒ Monthy Hall ('three boxes, version 2')

- if you trust the quiz master, then "no prize in the box that he is going to open" is a certain event;
- ⇒ the probability of finding the prize in your box remains 1/3.

The situation is quite different when there are two players

- *there was* 2/3 probability that the other player would not find the prize in the box;
- ⇒ the probability of finding the prize in your box rises to 1/2.

And if you do not trust the quiz master?

Reconditioning on a certain event?

$$P(H_i \,|\, \Omega) = \frac{P(H_i \cap \Omega)}{P(\Omega)} = \frac{P(H_i)}{1} = P(H_i)$$

⇒ Monthy Hall ('three boxes, version 2')

- if you trust the quiz master, then "no prize in the box that he is going to open" is a certain event;
- ⇒ the probability of finding the prize in your box remains $1/3$.

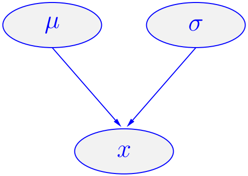
The situation is quite different when there are two players

- *there was* $2/3$ probability that the other player would not find the prize in the box;
- ⇒ the probability of finding the prize in your box rises to $1/2$.

And if you do not trust the quiz master?
Add this hypothesis in the model and apply probability theory!

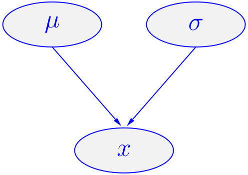
Inferring μ of the normal distribution

Setting up the problem



Inferring μ of the normal distribution

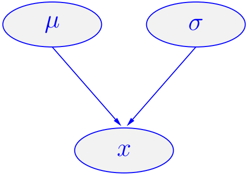
Setting up the problem



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

Inferring μ of the normal distribution

Setting up the problem

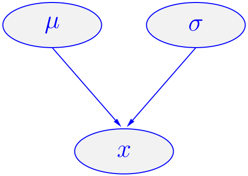


- In general $f(x, \mu, \sigma \,|\, I)$

- We start assuming σ well known, that we call here σ_e to remember that it is the standard deviation which describes statistical errors.

Inferring μ of the normal distribution

Setting up the problem

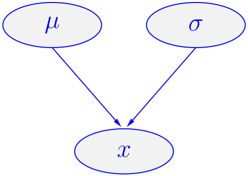


- In general $f(x, \mu, \sigma \mid I)$

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

Inferring μ of the normal distribution

Setting up the problem



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

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

Inferring μ of the normal distribution

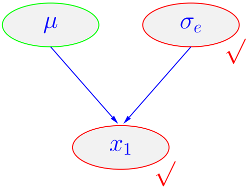
- σ_e assumed perfectly known;
- x_1 observed

Inferring μ of the normal distribution

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

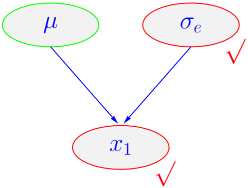
Inferring μ of the normal distribution

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



Inferring μ of the normal distribution

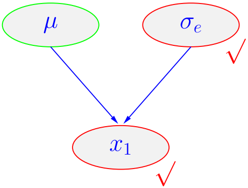
- σ_e assumed perfectly known;
- x_1 observed ($\equiv$ 'assumed perfectly known')



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

Inferring μ of the normal distribution

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

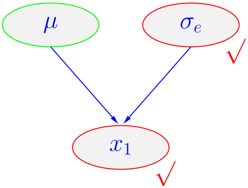


- Our task: $f(\mu \,|\, x_1, \sigma_e)$

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

Inferring μ of the normal distribution

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

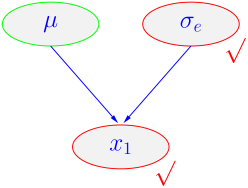


- Our task: $f(\mu \,|\, x_1, \sigma_e)$

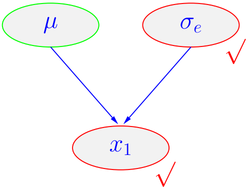
- In general: $f(\mu \,|\, \text{data}, I)$

'data' can be a set of observations

Inferring μ of the normal distribution



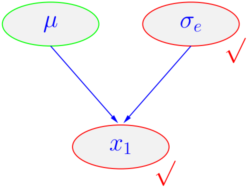
Inferring μ of the normal distribution



(Considering implicit the condition σ_e as well as I)

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

Inferring μ of the normal distribution

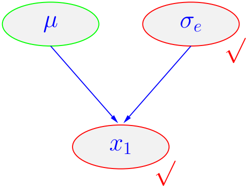


(Considering implicit the condition σ_e as well as I)

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

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

Inferring μ of the normal distribution



(Considering implicit the condition σ_e as well as I)

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

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

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

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

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

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

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

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

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

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

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

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

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

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

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

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

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

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

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

Inferring μ of the normal distribution

Solution for a flat prior

Starting as usual from a flat prior

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

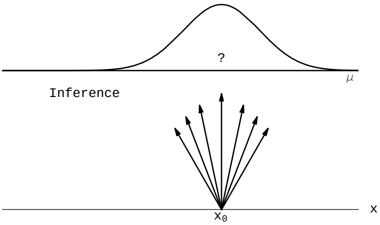
In the denominator, the exponential depends on $(x_1-\mu)^2$:
$\rightarrow$ the integral over μ is equal to the integral over x_1: $\longrightarrow 1$

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

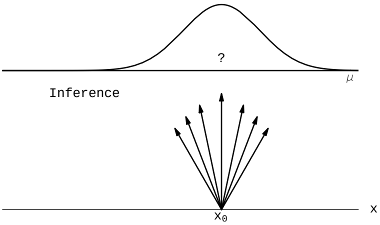
Note the swap of μ and x_1 at the exponent,
to emphasize that they have now different roles:

- μ is the variable;
- x_1 is a parameter

Inferring μ of the normal distribution

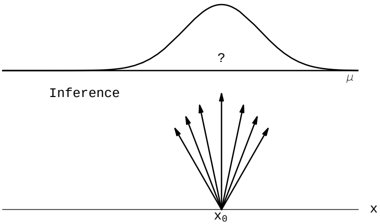


Inferring μ of the normal distribution



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

Inferring μ of the normal distribution

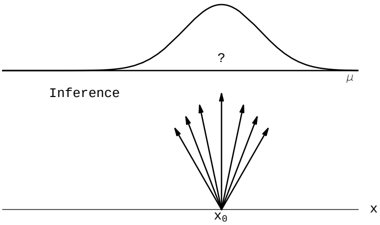


$$f(\mu\,|\,x_1) \;=\; \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(\mu-x_1)^2}{2\,\sigma_e^2}}$$

Summaries:

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

Inferring μ of the normal distribution



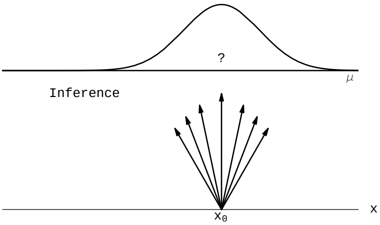
$$f(\mu\,|\,x_1) = \frac{1}{\sqrt{2\,\pi}\,\sigma_e}\,e^{-\frac{(\mu-x_1)^2}{2\,\sigma_e^2}}$$

Summaries:

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

All probability intervals calculated from the pdf.

Inferring μ of the normal distribution



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

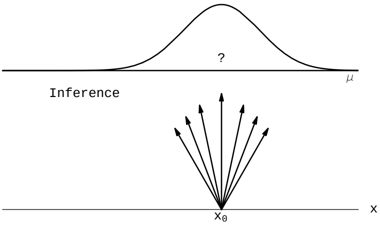
Summaries:

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

All probability intervals calculated from the pdf.

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

Inferring μ of the normal distribution



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

Summaries:

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

All probability intervals calculated from the pdf.

⇒ really **probability intervals**, and not 'confidence intervals'$^{(*)}$

$^{(*)}$The expressions "confidence interval" and "confidence limits" are jeopardized having often little to do with 'confidence' – sic!

Yes, but the prior?

Yes, but the prior?

Remember that (writing σ_e again)

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

Yes, but the prior?

Remember that (writing σ_e again)

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

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

Yes, but the prior?

Remember that (writing σ_e again)

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

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

Yes, but the prior?

Remember that (writing σ_e again)

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

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

Yes, but the prior?

Remember that (writing σ_e again)

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

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

Yes, but the prior?

Remember that (writing σ_e again)

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

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

Yes, but the prior?

Remember that (writing σ_e again)

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

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

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

Use of a conjugate prior

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

Use of a conjugate prior

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

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

Use of a conjugate prior

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

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

Use of a conjugate prior

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

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

Imagine that our initial prior was of the kind

$$\mu \sim \mathcal{N}(\mu_\circ, \sigma_\circ)$$

Use of a conjugate prior

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

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

Imagine that our initial prior was of the kind

$$\mu \sim \mathcal{N}(\mu_\circ, \sigma_\circ)$$

then

$$f(\mu \,|\, x_1, \sigma_e, \mu_\circ, \sigma_\circ) \propto e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}} \cdot e^{-\frac{(\mu-\mu_\circ)^2}{2\,\sigma_\circ^2}},$$

Use of a conjugate prior

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

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

Imagine that our initial prior was of the kind

$$\mu \sim \mathcal{N}(\mu_\circ, \sigma_\circ)$$

then

$$f(\mu \,|\, x_1, \sigma_e, \mu_\circ, \sigma_\circ) \;\; \propto \;\; e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}} \cdot e^{-\frac{(\mu-\mu_\circ)^2}{2\,\sigma_\circ^2}},$$

resulting into (technical details in next slide)

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

Use of a conjugate prior

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

Binomial distribution: Beta distribution.

Poisson distribution: Gamma distribution.

Gaussian distribution: Gaussian distribution.

Imagine that our initial prior was of the kind

$$\mu \sim \mathcal{N}(\mu_\circ, \sigma_\circ)$$

then

$$f(\mu\,|\,x_1, \sigma_e, \mu_\circ, \sigma_\circ) \;\propto\; e^{-\frac{(x_1-\mu)^2}{2\,\sigma_e^2}} \cdot e^{-\frac{(\mu-\mu_\circ)^2}{2\,\sigma_\circ^2}}\,,$$

resulting into (technical details in next slide)

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

with

$$\mu_A = \frac{x_1/\sigma_e^2 + \mu_\circ/\sigma_\circ^2}{1/\sigma_e^2 + 1/\sigma_\circ^2}$$

$$\frac{1}{\sigma_A^2} = \frac{1}{\sigma_e^2} + \frac{1}{\sigma_\circ^2}\,.$$

Other ‘Gaussian tricks’

Here are the details of our to get the previous result

$$
\begin{aligned}
f(\mu) &\propto \exp\left[-\frac{1}{2}\left(\frac{-2\,\mu\,x_1\sigma_\circ^2+\mu^2\sigma_\circ^2+-2\,\mu\,\mu_\circ\sigma_e^2+\mu^2\sigma_e^2}{\sigma_e^2+\sigma_\circ^2}\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{\mu^2-2\,\mu\left(\frac{x_1\,\sigma_\circ^2+\mu_\circ\,\sigma_e^2}{\sigma_e^2+\sigma_\circ^2}\right)}{(\sigma_e^2\cdot\sigma_\circ^2)/(\sigma_e^2+\sigma_\circ^2)}\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{\mu^2-2\mu\,\mu_A}{\sigma_A^2}\right)\right] \\
&\propto \exp\left[-\frac{(\mu-\mu_A)^2}{2\,\sigma_A^2}\right]
\end{aligned}
$$

In particolular, in the last step the trick of complementing the exponential has been used, since adding/removing constant terms in the exponential is equivalent to multiply/devide by factors. Once we recognize the structure, the normalization is automatic.

Use of a conjugate prior to infer μ of a Gaussian

- Unfortunately, the conjugate prior of a Gaussian is not that flexible.

Use of a conjugate prior to infer μ of a Gaussian

- Unfortunately, the conjugate prior of a Gaussian is not that flexible.
- It results on the well known formula to 'combine results' by a weighted average, with weights equal to the inverses of the variances

Use of a conjugate prior to infer μ of a Gaussian

- Unfortunately, the conjugate prior of a Gaussian is not that flexible.
- It results on the well known formula to 'combine results' by a weighted average, with weights equal to the inverses of the variances
- In particular

$$\sigma_A < \min(\sigma_0, \sigma_e)$$

Use of a conjugate prior to infer μ of a Gaussian

- Unfortunately, the conjugate prior of a Gaussian is not that flexible.
- It results on the well known formula to 'combine results' by a weighted average, with weights equal to the inverses of the variances
- In particular

$$\sigma_A < \min(\sigma_0, \sigma_e)$$

$\rightarrow$ a measurement improves our knowledge about μ

Use of a conjugate prior to infer μ of a Gaussian

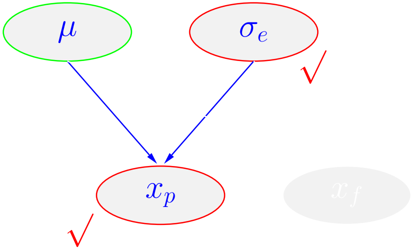
- Unfortunately, the conjugate prior of a Gaussian is not that flexible.
- It results on the well known formula to 'combine results' by a weighted average, with weights equal to the inverses of the variances
- In particular

$$\sigma_A < \min(\sigma_0, \sigma_e)$$

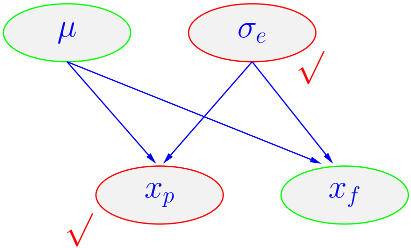
$\rightarrow$ a measurement improves our knowledge about μ

- A flat prior is recovered for $\sigma_o^2 \gg \sigma_e^2$ (and x_0 'reasonable').

Predictive distribution

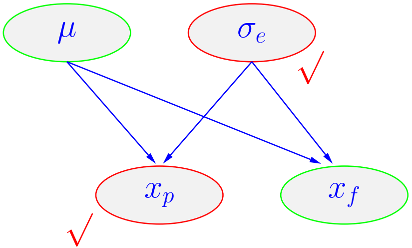


Predictive distribution



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

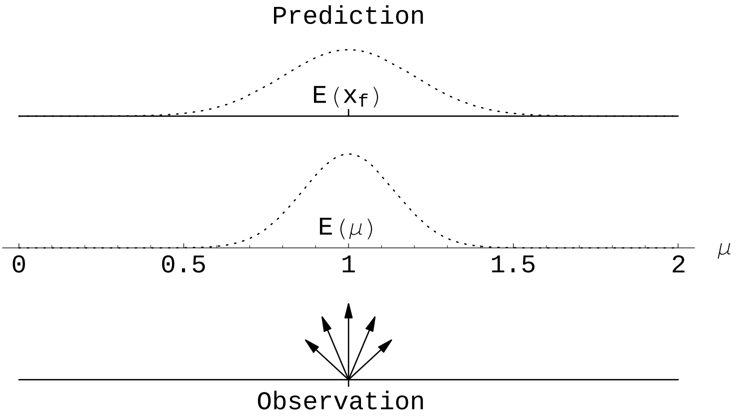
Predictive distribution



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

Predictive distribution

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



Predictive distribution

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

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

Predictive distribution

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

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

Thus, in our case

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

Predictive distribution

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

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

Thus, in our case

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

Predictive distribution

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

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

Thus, in our case

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

Predictive distribution

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

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

Thus, in our case

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(Glen Cowan, *Statistical Data Analysis*)

Remark on ‘conventional statistics’

Objection:

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

Remark on 'conventional statistics'

Objection:

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

Uhm. . .

Remark on 'conventional statistics'

Objection:

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

Uhm. . .

- Frequentist 'gurus' are champions in misusing terminonology, thus confusing people ("CL", "confidence intervals").

Remark on 'conventional statistics'

Objection:

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

Uhm. . .

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

Remark on 'conventional statistics'

Objection:

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

Uhm. . .

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

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

Remark on ‘conventional statistics’

Objection:

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

Uhm. . .

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

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

Remark on ‘conventional statistics’

Objection:

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

Uhm. . .

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

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

Q. Does the method always produce wrong results?

Remark on ‘conventional statistics’

Objection:

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

Uhm. . .

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

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

Q. Does the method always produce wrong results?

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

Remark on 'conventional statistics'

Objection:

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

Uhm. . .

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

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

Q. Does the method always produce wrong results?

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

Remark on 'conventional statistics'

Objection:

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

Uhm. . .

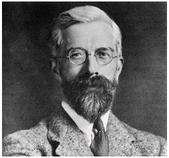
- Frequentist 'gurus' are champions in misusing terminonology, thus confusing people ("CL", "confidence intervals").
- Details in GdA, *About the proof of the so called exact classical confidence intervals. Where is the trick?*, https://arxiv.org/abs/physics/0605140

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

Q. Does the method always produce wrong results?

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

Prescriptions?



Objective prescriptions?

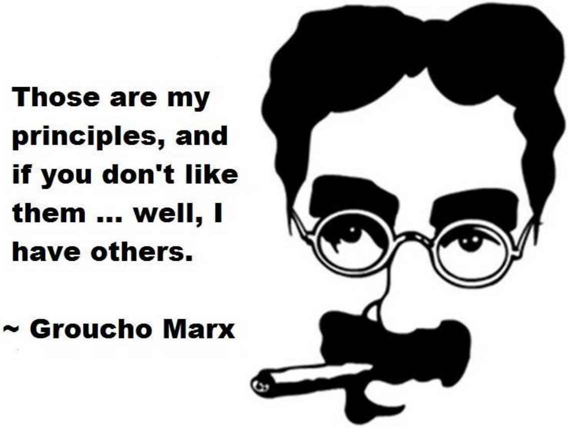
Mistrust those who promise you 'objective' methods to form up your confidence about the physical world!

Principles?

Too many unnecessary ‘principles’ on the market.

Principles?

Too many unnecessary ‘principles’ on the market.



Introducing systematics

Influence quantities

By influence quantities we mean:

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

Introducing systematics

Influence quantities

By influence quantities we mean:

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

$\rightarrow$ all calibration constants;

Introducing systematics

Influence quantities

By influence quantities we mean:

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

Introducing systematics

Influence quantities

By influence quantities we mean:

- $\rightarrow$ all kinds of external factors which may 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).

Introducing systematics

Influence quantities

By influence quantities we mean:

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

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

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 observables 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 $\boldsymbol{h}$

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

- We can use Bayes' theorem to make an 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 $\boldsymbol{h}$

Conditional inference

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

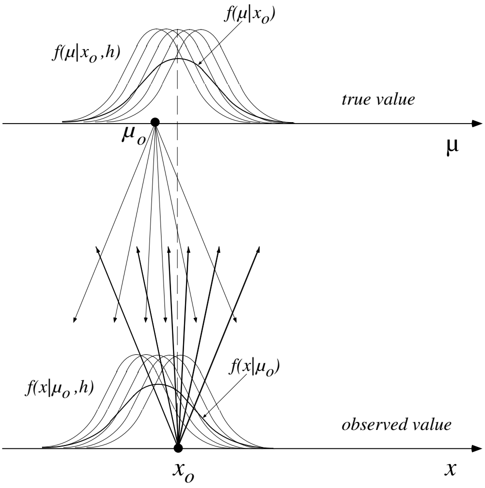
$$\boldsymbol{x} \Rightarrow f(\boldsymbol{\mu} \,|\, \boldsymbol{x}, \boldsymbol{h})\,.$$

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 $\boldsymbol{h}$

Propagation of uncertainties

- Essentially, one applies the propagation of uncertainty, whose most general case has been illustrated in the previous section, 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 (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.

The third form is particularly convenient to make linear expansions which lead to approximate solutions.

Systematics due to uncertain offset

Model:

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

Systematics due to uncertain offset

Model:

- the "zero" of the instrument is not usually known exactly, owing to calibration uncertainty.
- This can be parametrized assuming that its 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 not usually known exactly, owing to calibration uncertainty.
- This can be parametrized assuming that its 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 initial 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 not usually known exactly, owing to calibration uncertainty.
- This can be parametrized assuming that its 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 initial 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 $\mu + Z$:

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

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

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

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

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

- $f(\mu)$ is still a 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 a 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

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 a 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.

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:

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

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

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

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:

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

Systematics due to uncertain offset

Summary:

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

Systematics due to uncertain offset

Summary:

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

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

Systematics due to uncertain offset

Summary:

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

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

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

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:

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

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

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

As more or less intuitively expected from an offset!

An exercise

Two samples of data have been collected with the same instrument. These are the numbers, as they result from a printout (homogeneous quantities, therefore measurement unit omitted):

- $n_1 = 1000$, $\overline{x}_1 = 10.4012$, $s_1 = 5.7812$;
- $n_2 = 2000$, $\overline{x}_2 = 10.2735$, $s_2 = 5.9324$.

We know that the instrument has an offset uncertainty of 0.15.

1. Report the results on μ_1, μ_2, σ_1 and σ_2.
2. If you consider the σ's of the two samples consistent you might combine the result.
3. Calculate the correlation coefficient between μ_1 and μ_2.
4. Give also the result on $s = \mu_1 + \mu_2$ and $s = \mu_1 - \mu_2$, including $\rho(s, d)$.
5. Give also the result on $z_1 = \mu_1\,\mu_2^2$ and $z_2 = \mu_1/\mu_2$, including $\rho(z_1, z_2)$.
6. Consider also a third data sample, recorded with the same instrument:
   $n_3 = 4$, $\overline{x}_3 = 13.8931$, $s_3 = 4.5371$.

Inferring μ from a sample

(Gaussian, independent observations, $\boldsymbol{\sigma}$ **perfectly known**)

$$f(\mu\,|\,\underline{x},\sigma) \quad\propto\quad f(\underline{x}\,|\,\mu,\sigma)\cdot f_0(\mu)$$

Inferring μ from a sample

(Gaussian, independent observations, $\boldsymbol{\sigma}$ **perfectly known**)

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \sigma) &\propto f(\underline{x} \,|\, \mu, \sigma) \cdot f_0(\mu) \\
&\propto \prod_i f(x_i \,|\, \mu, \sigma) \cdot f_0(\mu) =
\end{aligned}
$$

Inferring μ from a sample

(Gaussian, independent observations, $\boldsymbol{\sigma}$ **perfectly known**)

$$
\begin{aligned}
f(\mu \,|\, \underline{x}, \sigma) &\propto f(\underline{x} \,|\, \mu, \sigma) \cdot f_0(\mu) \\
&\propto \prod_i f(x_i \,|\, \mu, \sigma) \cdot f_0(\mu) = \prod_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\,\sigma^2}} \cdot f_0(\mu)
\end{aligned}
$$

Inferring μ from a sample

(Gaussian, independent observations, $\boldsymbol{\sigma}$ **perfectly known**)

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

Inferring μ from a sample

(Gaussian, independent observations, $\boldsymbol{\sigma}$ **perfectly known**)

$$f(\mu\,|\,\underline{x},\sigma) \propto f(\underline{x}\,|\,\mu,\sigma)\cdot f_0(\mu)$$

$$\propto \prod_i f(x_i\,|\,\mu,\sigma)\cdot f_0(\mu) = \prod_i \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\,\sigma^2}}\cdot f_0(\mu)$$

$$\propto \exp\left[-\frac{\sum_i(x_i-\mu)^2}{2\,\sigma^2}\right]\cdot f_0(\mu)$$

$$\propto \exp\left[-\frac{(\sum_i x_i^2 - 2\mu\sum_i x_i + n\,\mu^2)}{2\,\sigma^2}\right]\cdot f_0(\mu)$$

Inferring μ from a sample

(Gaussian, independent observations, $\boldsymbol{\sigma}$ **perfectly known**)

$$f(\mu\,|\,\underline{x},\sigma) \propto f(\underline{x}\,|\,\mu,\sigma)\cdot f_0(\mu)$$

$$\propto \prod_i f(x_i\,|\,\mu,\sigma)\cdot f_0(\mu) = \prod_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\,\sigma^2}}\cdot f_0(\mu)$$

$$\propto \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\,\sigma^2}\right]\cdot f_0(\mu)$$

$$\propto \exp\left[-\frac{(\sum_i x_i^2 - 2\mu\sum_i x_i + n\,\mu^2)}{2\,\sigma^2}\right]\cdot f_0(\mu)$$

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

Inferring μ from a sample

(Gaussian, independent observations, $\boldsymbol{\sigma}$ **perfectly known**)

$$f(\mu\,|\,\underline{x},\sigma) \propto f(\underline{x}\,|\,\mu,\sigma)\cdot f_0(\mu)$$

$$\propto \prod_i f(x_i\,|\,\mu,\sigma)\cdot f_0(\mu) = \prod_i \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\,\sigma^2}}\cdot f_0(\mu)$$

$$\propto \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\,\sigma^2}\right]\cdot f_0(\mu)$$

$$\propto \exp\left[-\frac{(\sum_i x_i^2 - 2\mu\sum_i x_i + n\,\mu^2)}{2\,\sigma^2}\right]\cdot f_0(\mu)$$

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

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

Inferring μ from a sample

(Gaussian, independent observations, $\boldsymbol{\sigma}$ **perfectly known**)

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

Inferring μ from a sample

(Gaussian, independent observations, **σ perfectly known**)

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

Trick: complementing of exponential

Inferring μ from a sample

(Gaussian, independent observations, σ perfectly known)

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

Inferring μ from a sample

(Gaussian, independent observations, σ perfectly known)

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

In the case of $f_0(\mu)$ irrelevant (but we know how to act otherwise!)
we recognize by eye a Gaussian

Inferring μ from a sample

(Gaussian, independent observations, σ perfectly known)

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

In the case of $f_0(\mu)$ irrelevant (but we know how to act otherwise!)
we recognize by eye a Gaussian

$$f(\mu \,|\, \underline{x}, \sigma) \quad = \quad \frac{1}{\sqrt{2\pi}\,\sigma/\sqrt{n}} \exp\left[-\frac{(\mu-\overline{x})^2}{2\,(\sigma/\sqrt{n})^2}\right]$$

Inferring μ from a sample

(Gaussian, independent observations, σ perfectly known)

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

In the case of $f_0(\mu)$ irrelevant (but we know how to act otherwise!) we recognize by eye a Gaussian

$$f(\mu\,|\,\underline{x},\sigma) \quad=\quad \frac{1}{\sqrt{2\pi}\,\sigma/\sqrt{n}}\,\exp\left[-\frac{(\mu-\overline{x})^2}{2\,(\sigma/\sqrt{n})^2}\right]$$

μ is Gaussian around arithmetic average, with standard deviation $\sigma/\sqrt{n}$

$$\mu\sim\mathcal{N}(\overline{x},\frac{\sigma}{\sqrt{n}})$$

Inferring μ from a sample

(Gaussian, independent observations, σ perfectly known)

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

In the case of $f_0(\mu)$ irrelevant (but we know how to act otherwise!)
we recognize by eye a Gaussian

$$f(\mu \,|\, \underline{x}, \sigma) \quad = \quad \frac{1}{\sqrt{2\pi}\,\sigma/\sqrt{n}}\,\exp\left[-\frac{(\mu-\overline{x})^2}{2\,(\sigma/\sqrt{n})^2}\right]$$

μ is Gaussian around arithmetic average, with standard deviation $\sigma/\sqrt{n}$

$$\mu \sim \mathcal{N}(\overline{x}, \frac{\sigma}{\sqrt{n}})$$

- $\overline{x}$ is a *sufficient statistic* (**very important concept!**)

Inferring μ from a sample

(Gaussian, independent observations, σ perfectly known)

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

In the case of $f_0(\mu)$ irrelevant (but we know how to act otherwise!)
we recognize by eye a Gaussian

$$f(\mu\,|\,\underline{x},\sigma) \quad = \quad \frac{1}{\sqrt{2\pi}\,\sigma/\sqrt{n}}\exp\left[-\frac{(\mu-\overline{x})^2}{2\,(\sigma/\sqrt{n})^2}\right]$$

μ is Gaussian around arithmetic average, with standard deviation $\sigma/\sqrt{n}$

$$\mu \sim \mathcal{N}(\overline{x},\frac{\sigma}{\sqrt{n}})$$

- $\overline{x}$ is a *sufficient statistic* (**very important concept!**)
  $\Rightarrow$ $\overline{x}$ it provides the same information about μ
  contained in detailed knowledge of $\underline{x}$

Inferring μ from a sample

(Gaussian, independent observations, σ perfectly known)

Exercise

- In the last steps we have used the technique of *complementing the exponential*.
- Restart, using a *flat prior*, from

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

and use the 'Gaussian tricks' (first and second derivatives of $\varphi(\mu)$) to find $\mathrm{E}(\mu)$ and $\mathrm{Var}(\mu)$.

Inferring μ from a sample

(Gaussian, independent observations, σ perfectly known)

Exercise

- In the last steps we have used the technique of *complementing the exponential*.
- Restart, using a *flat prior*, from

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

  and use the 'Gaussian tricks' (first and second derivatives of $\varphi(\mu)$) to find $\mathrm{E}(\mu)$ and $\mathrm{Var}(\mu)$.
- In this case the result is exact, because $f(\mu \,|\, \underline{x}, \sigma)$ is indeed Gaussian.
  (A hint is that $\frac{d^2\varphi(\mu)}{d\mu^2}$ is constant $\forall\,\mu$)

Joint inference of μ and σ from a sample

$$f(\mu,\sigma\,|\,\underline{x}) \;\propto\; \prod_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\,\sigma^2}} \cdot f_0(\mu,\sigma)$$

Joint inference of μ and σ from a sample

$$f(\mu,\sigma\,|\,\underline{x}) \;\propto\; \prod_i \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\,\sigma^2}}\cdot f_0(\mu,\sigma)$$

$$\propto\; \frac{1}{\sigma^n}\;\exp\left[-\frac{\sum_i(x_i-\mu)^2}{2\,\sigma^2}\right]\cdot f_0(\mu,\sigma),$$

Joint inference of μ and σ from a sample

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

Joint inference of μ and σ from a sample

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

Joint inference of μ and σ from a sample

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

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

Joint inference of μ and σ from a sample

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

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

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

Joint inference of μ and σ from a sample

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

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

- the inference on μ and σ depends only on $\overline{x}$ and s (and on the priors, as it has to be!).
  $\Rightarrow$ $\overline{x}$ and s are **sufficient statistics**

Joint inference of μ and σ from a sample

In practice

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

Joint inference of μ and σ from a sample

In practice

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

Then

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

Joint inference of μ and σ from a sample

In practice

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

Then

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

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

Joint inference of μ and σ from a sample

In practice

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

Then

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

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

Details in the Appendix, but some remarks are in order:

Joint inference of μ and σ from a sample

In practice

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

Then

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

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

Details in the Appendix, but some remarks are in order:

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

Joint inference of μ and σ from a sample

In practice

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

Then

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

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

Details in the Appendix, but 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

Joint inference of μ and σ from a sample

In practice

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

Then

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

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

Details in the Appendix, but 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')

Joint inference of μ and σ from a sample

In practice

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

Then

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

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

Details in the Appendix, but 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'

Joint inference of μ and σ from a sample

In practice

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

Then

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

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

Details in the Appendix, but 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'

$$\Rightarrow n \to \infty$$

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)

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

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)

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

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)

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

$^{(*)}$ 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 σ $\Rightarrow$ inducing 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)

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

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

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

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

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

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

[(*)] The most sensitive is the prior on σ $\Rightarrow$ inducing abstract speculations in mathematicians and statisticians who often have little idea of what they are talking about (**Gauss was Gauss!**).

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)

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

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

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

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

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

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

(*) The most sensitive is the prior on σ $\Rightarrow$ inducing abstract speculations in mathematicians and statisticians who often have little idea of what they are talking about (**Gauss was Gauss!**).
$\Rightarrow$ **See references and links**

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning μ on a precise value of $\sigma = \sigma_*$:

$$f(\mu \,|\, \overline{x}, s, \sigma_*)$$

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning μ on a precise value of $\sigma = \sigma_*$:

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

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning μ on a precise value of $\sigma = \sigma_*$:

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

All factors not depending on μ absorbed in '$\propto$'.

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning μ on a precise value of $\sigma = \sigma_*$:

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

All factors not depending on μ absorbed in '$\propto$'.
In the case of uniform $f_0(\mu)$ it turns out that μ is Gaussian around $\overline{x}$ with standard deviation equal to $\sigma_*/\sqrt{n}$.

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning μ on a precise value of $\sigma = \sigma_*$:

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

All factors not depending on μ absorbed in '$\propto$'.
In the case of uniform $f_0(\mu)$ it turns out that μ is Gaussian around $\overline{x}$ with standard deviation equal to $\sigma_*/\sqrt{n}$.

"Obviously!": this is equivanent to the choice $f_o(\sigma) = \delta(\sigma - \sigma_*)$

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning σ on a precise value of $\mu = \mu_*$:

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

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning σ on a precise value of $\mu = \mu_*$:

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

with $K^2 = n\,(s^2 + (\mu_* - \overline{x})^2)/2$, just a positive constant.

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning σ on a precise value of $\mu=\mu_*$:

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

with $K^2=n\,(s^2+(\mu_*-\overline{x})^2)/2$, just a positive constant.

Change of variable: $\sigma\to\tau=1/\sigma^2$ (technically convenient):

$$f(\tau\,|\,\overline{x},s,\mu_*) \quad\propto\quad \tau^{n/2}\,\exp\left[-K^2\,\tau\right]\cdot f_0(\tau)$$

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning σ on a precise value of $\mu = \mu_*$:

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

$$\propto\;\; \sigma^{-n} \exp\left[-\frac{K^2}{\sigma^2}\right] \cdot f_0(\sigma)$$

with $K^2 = n\,(s^2 + (\mu_* - \overline{x})^2)/2$, just a positive constant.

Change of variable: $\sigma \rightarrow \tau = 1/\sigma^2$ (technically convenient):

$$f(\tau \,|\, \overline{x}, s, \mu_*) \;\;\propto\;\; \tau^{n/2} \exp\left[-K^2\,\tau\right] \cdot f_0(\tau)$$

We 'easily' recognize in $\tau^{n/2} \exp\left[-K^2\,\tau\right]$

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning σ on a precise value of $\mu = \mu_*$:

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

$$\propto\; \sigma^{-n} \exp\left[-\frac{K^2}{\sigma^2}\right] \cdot f_0(\sigma)$$

with $K^2 = n\,(s^2 + (\mu_* - \overline{x})^2)/2$, just a positive constant.

Change of variable: $\sigma \to \tau = 1/\sigma^2$ (technically convenient):

$$f(\tau \,|\, \overline{x}, s, \mu_*) \;\propto\; \tau^{n/2} \exp\left[-K^2\,\tau\right] \cdot f_0(\tau)$$

We ‘easily’ recognize in $\tau^{n/2} \exp\left[-K^2\,\tau\right]$ a **Gamma distribution**:

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning σ on a precise value of $\mu=\mu_*$:

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

$$\propto\; \sigma^{-n}\exp\left[-\frac{K^2}{\sigma^2}\right]\cdot f_0(\sigma)$$

with $K^2=n\,(s^2+(\mu_*-\overline{x})^2)/2$, just a positive constant.
Change of variable: $\sigma\rightarrow\tau=1/\sigma^2$ (technically convenient):

$$f(\tau\,|\,\overline{x},s,\mu_*) \;\propto\; \tau^{n/2}\exp\left[-K^2\,\tau\right]\cdot f_0(\tau)$$

We 'easily' recognize in $\tau^{n/2}\exp\left[-K^2\,\tau\right]$ a **Gamma distribution**:
$\rightarrow$ also $f(\tau\,|\,\overline{x},s,\mu_*)$ will be a **Gamma** if a Gamma $f_0(\tau)$ is chosen

Joint inference of μ and σ from a sample

Conditional distributions

Joint distribution:

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

Conditioning σ on a precise value of $\mu = \mu_*$:

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

$$\propto\; \sigma^{-n}\,\exp\left[-\frac{K^2}{\sigma^2}\right]\cdot f_0(\sigma)$$

with $K^2 = n\,(s^2+(\mu_*-\overline{x})^2)/2$, just a positive constant.
Change of variable: $\sigma \rightarrow \tau = 1/\sigma^2$ (technically convenient):

$$f(\tau\,|\,\overline{x},s,\mu_*) \;\propto\; \tau^{n/2}\,\exp\left[-K^2\,\tau\right]\cdot f_0(\tau)$$

We 'easily' recognize in $\tau^{n/2}\,\exp\left[-K^2\,\tau\right]$ a **Gamma distribution**:
$\rightarrow$ also $f(\tau\,|\,\overline{x},s,\mu_*)$ will be a **Gamma** if a Gamma $f_0(\tau)$ is chosen

$\Rightarrow$ Gibbs sampler!

Joint inference of μ and τ ($\rightarrow \sigma$) from a sample

Sampling the posterior by **MCMC** using **Gibbs sampler**

0) Inizialization:
   - $i = 0$
   - choose a suitable Gaussian conjugate for μ ($\sigma_0 \rightarrow \infty$ for 'flat');
   - choose a suitable Gamma conjugate for τ ($c = 1$, $r \rightarrow 0$ for 'flat');

Joint inference of μ and τ ($\rightarrow \sigma$) from a sample

Sampling the posterior by **MCMC** using **Gibbs sampler**

0) Inizialization:
   - $i = 0$
   - choose a suitable Gaussian conjugate for μ ($\sigma_0 \rightarrow \infty$ for ‘flat’);
   - choose a suitable Gamma conjugate for τ ($c = 1$, $r \rightarrow 0$ for ‘flat’);
   - choose an arbitary (but possible ‘reasonable’) μ_i;

Joint inference of μ and τ ($\rightarrow \sigma$) from a sample

Sampling the posterior by **MCMC** using **Gibbs sampler**

0) Inizialization:
   - $i = 0$
   - choose a suitable Gaussian conjugate for μ ($\sigma_0 \rightarrow \infty$ for ‘flat’);
   - choose a suitable Gamma conjugate for τ ($c = 1$, $r \rightarrow 0$ for ‘flat’);
   - choose an arbitary (but possible ‘reasonable’) μ_i;
   - extract at random τ_i from $f(\tau \,|\, \overline{x}, s, \mu_i)$

Joint inference of μ and τ ($\rightarrow \sigma$) from a sample

Sampling the posterior by **MCMC** using **Gibbs sampler**

0) Inizialization:
   - $i = 0$
   - choose a suitable Gaussian conjugate for μ ($\sigma_0 \rightarrow \infty$ for 'flat');
   - choose a suitable Gamma conjugate for τ ($c = 1$, $r \rightarrow 0$ for 'flat');
   - choose an arbitary (but possible 'reasonable') μ_i;
   - extract at random τ_i from $f(\tau \,|\, \overline{x}, s, \mu_i)$

   Then **loop** n **times** :

1) $i = i + 1$;
   extract at random μ_i from $f(\mu \,|\, \overline{x}, s, \tau_{i-1})$;

Joint inference of μ and τ ($\rightarrow \sigma$) from a sample

Sampling the posterior by **MCMC** using **Gibbs sampler**

0) Inizialization:
   - $i = 0$
   - choose a suitable Gaussian conjugate for μ ($\sigma_0 \rightarrow \infty$ for 'flat');
   - choose a suitable Gamma conjugate for τ ($c = 1$, $r \rightarrow 0$ for 'flat');
   - choose an arbitary (but possible 'reasonable') μ_i;
   - extract at random τ_i from $f(\tau \,|\, \overline{x}, s, \mu_i)$

   Then **loop** n **times** :

1) $i = i + 1$;
   extract at random μ_i from $f(\mu \,|\, \overline{x}, s, \tau_{i-1})$;
2) extract at random τ_i from $f(\tau \,|\, \overline{x}, s, \mu_i)$;
   goto 1)

Joint inference of μ and τ ($\rightarrow \sigma$) from a sample

Sampling the posterior by **MCMC** using **Gibbs sampler**

0) Inizialization:
   - $i = 0$
   - choose a suitable Gaussian conjugate for μ ($\sigma_0 \rightarrow \infty$ for 'flat');
   - choose a suitable Gamma conjugate for τ ($c = 1$, $r \rightarrow 0$ for 'flat');
   - choose an arbitary (but possible 'reasonable') μ_i;
   - extract at random τ_i from $f(\tau \,|\, \overline{x}, s, \mu_i)$

   Then **loop *n* times** :

1) $i = i + 1$;
   extract at random μ_i from $f(\mu \,|\, \overline{x}, s, \tau_{i-1})$;
2) extract at random τ_i from $f(\tau \,|\, \overline{x}, s, \mu_i)$;
   goto 1)

Try it!

You only need Gaussian and Gamma random number generators (e.g. in R)

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

Model (to be written in the model file)

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

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

Model (to be written in the model file)

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

Simulated data

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

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

Model (to be written in the model file)

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

Simulated data

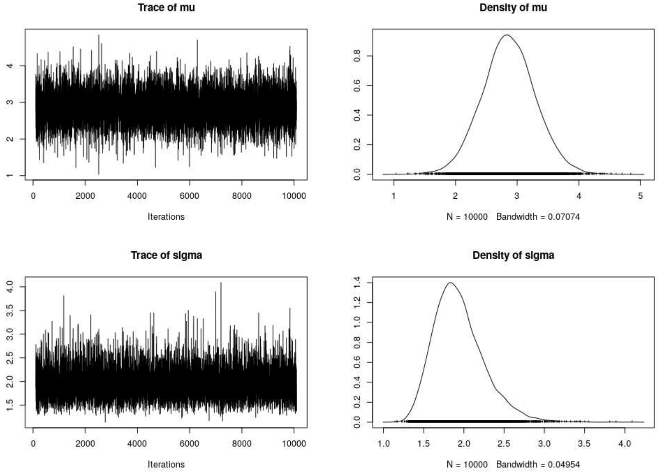
```
mu.true = 3; sigma.true = 2; sample.n = 20
x = rnorm(sample.n, mu.true, sigma.true)
```

JAGS calls

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

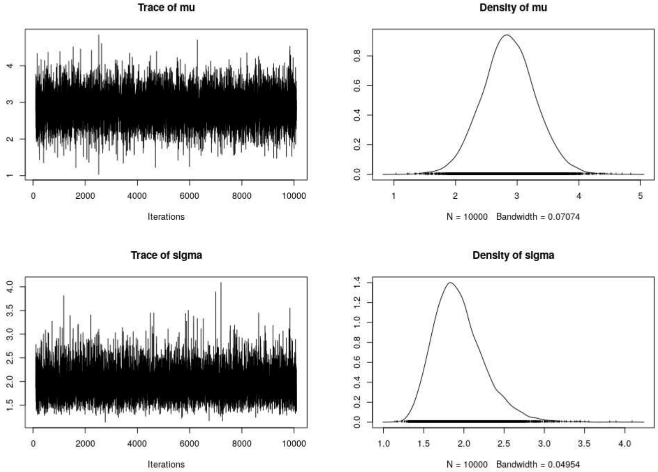
Joint inference of μ and τ ($\rightarrow \sigma$) with JAGS/rjags

$\Rightarrow$ inf_mu_sigma.R



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

$\Rightarrow$ inf_mu_sigma.R



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

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

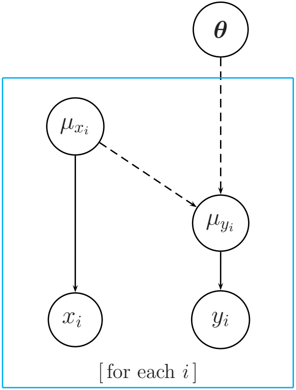
  parametric inference.

- 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.

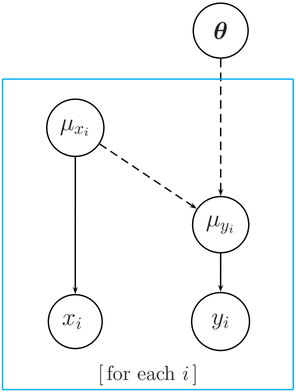


Fits – introduction

- 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.

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

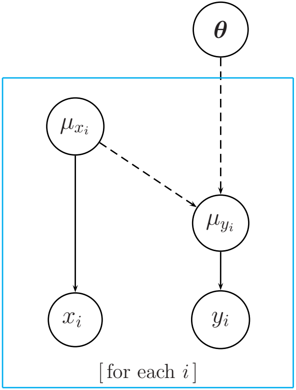


Fits – introduction

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

  parametric inference.

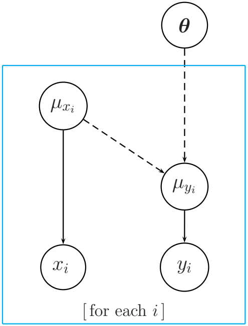
- set up the model,
  e.g. $\mu_{y_i} = m\,\mu_{x_i} + c$
  Note: Linearity is between μ_{y_i} and μ_{x_i}, not between y_i and x_i!
- apply probability rules;
- perform the calculations.



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

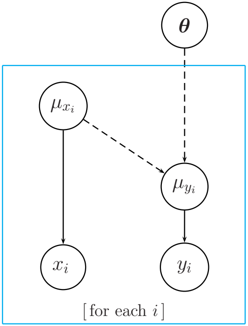
$\rightarrow f(m, c\,|\,\mathrm{x}, \mathrm{y}, \boldsymbol{\sigma})$, in the case of case of linear fit with "σ's known a priori" (!)

Linear fit – introduction



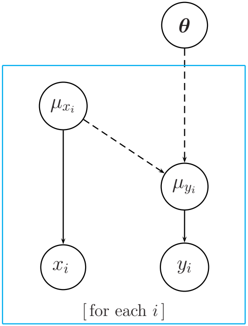
- Deterministic links between μ_x's and μ_y's.

Linear fit – introduction



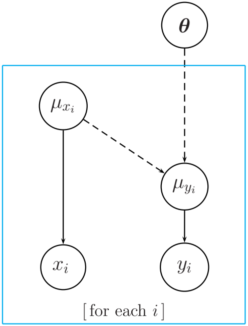
- Deterministic links between μ_x's and μ_y's.
- Probabilistic links between μ_x's and x's, and between μ_y's and y's (errors on both axes)

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 (σ's known): $\{\boldsymbol{x}, \boldsymbol{y}\} \rightarrow \boldsymbol{\theta} = (m, c)$

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 (σ's known): $\{\boldsymbol{x}, \boldsymbol{y}\} \;\rightarrow \boldsymbol{\theta} = (m, c)$
- 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

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

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.

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

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

$\Rightarrow$ 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

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

- 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'.

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'.
- 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

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

  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.

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);
}
```

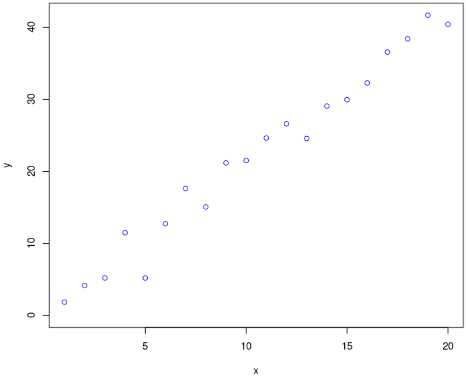
Simulated data

```
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)) )
```

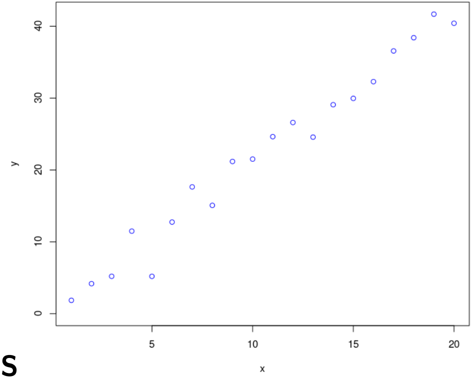
Linear fits with uncertain σ in JAGS

Plot of simulated data



Linear fits with uncertain σ in JAGS

Plot of simulated data



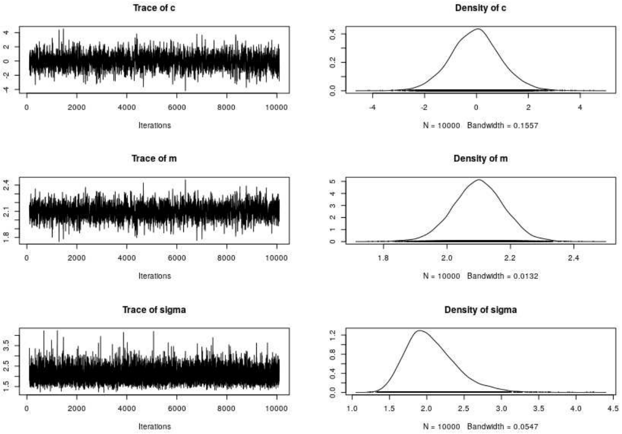
Calling JAGS

```
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

⇒ `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

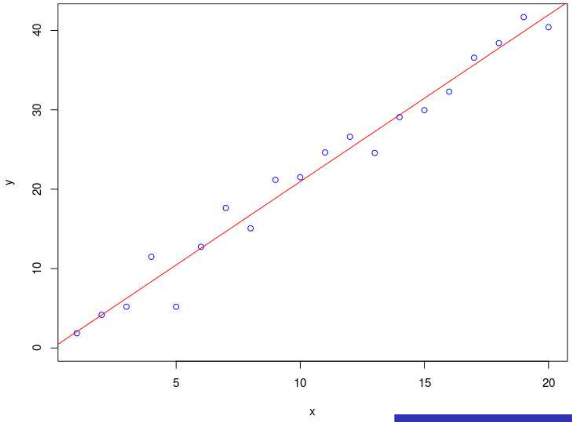
'Check' the result

```
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

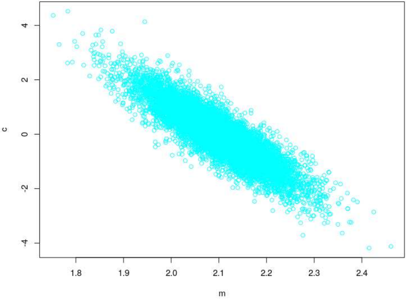
Correlation between m **and** c

```
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

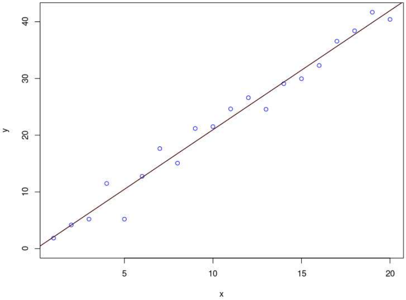
Check with `R` `lm()` (least square)

```
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)

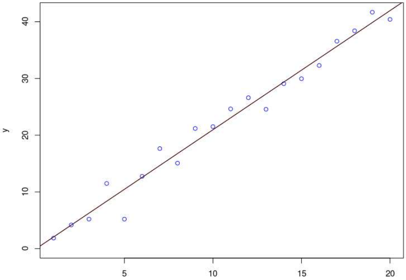
```
plot(x,y, col='blue',ylim=c(0,max(y)) )
abline(mean(c), mean(m), col='red') # JAGS
abline(lm(y~x), col='black')        # least squares
```



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
```

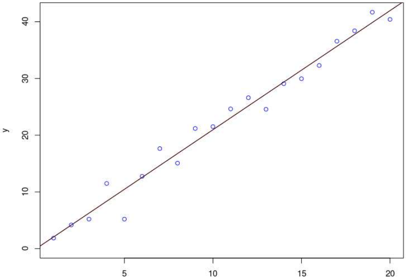


Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result

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
```

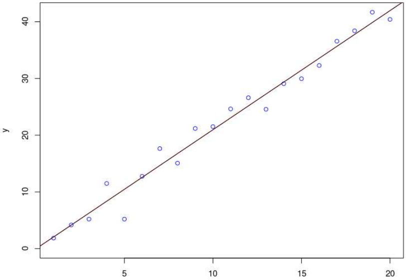


Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time?

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
```

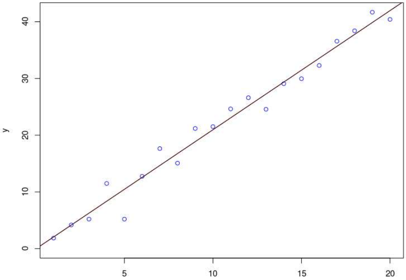


Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends...**

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
```



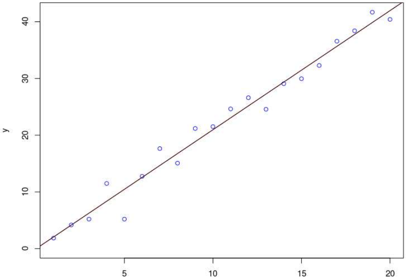
Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends...**

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
```



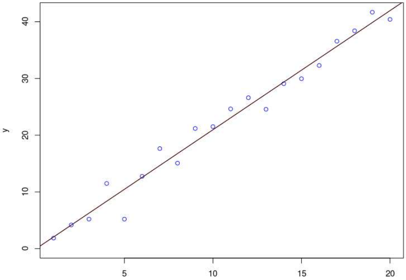
Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends...**

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
```



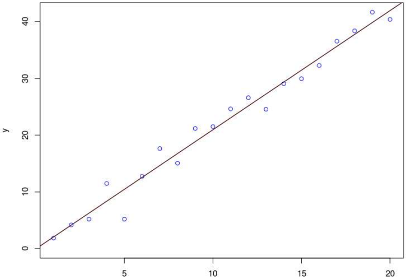
Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends. . .**

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 $\approx$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
```



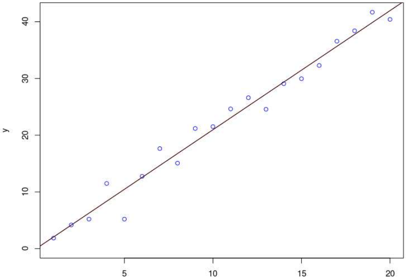
Linear model line ($c = -0.05$, $m = 2.10$) covers perfectly the JAGS result: waste of time? **It all depends. . .**

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 $\approx$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

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

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

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

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

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!)

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

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

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

Our problem

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

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

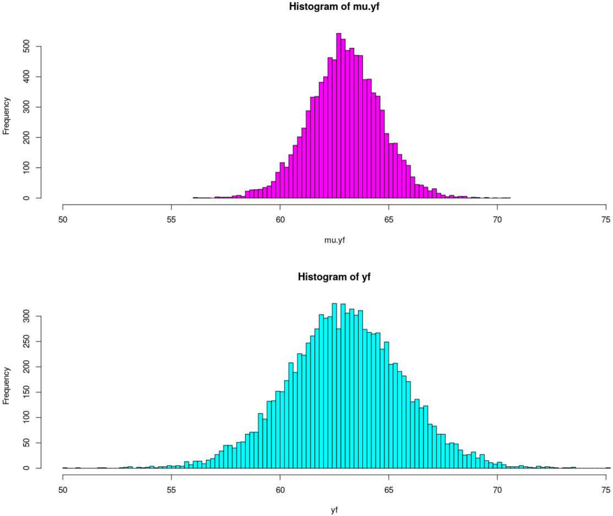
Including prediction in the JAGS model

```
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)

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

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

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)

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

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

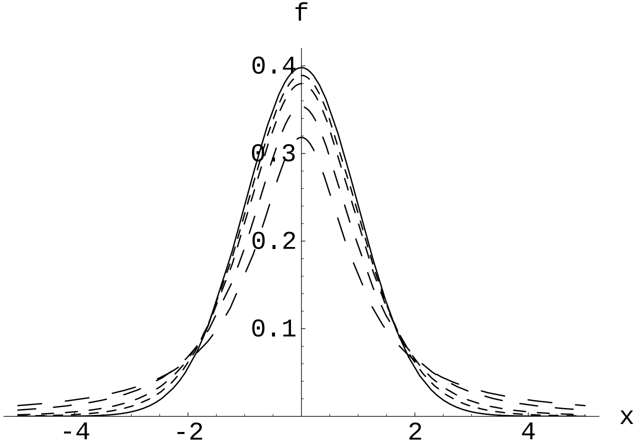
$$
\begin{aligned}
\nu &= n-2\\
t &= \frac{\mu-\overline{x}}{s/\sqrt{n-2}}\,,
\end{aligned}
$$

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,

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

Inferring μ and σ from a sample

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

In summary,

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

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

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

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.

Inferring μ and σ from a sample

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

In summary,

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

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

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

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 $\sigma(\mu) = s/\sqrt{n}$.

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:

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

[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?

[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.

[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?

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:

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

So what?

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

- 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:

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

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)

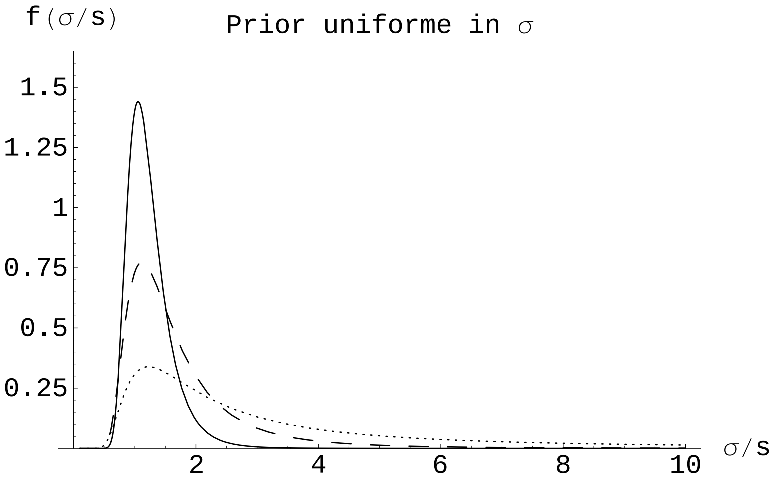
Marginal $f(\sigma)$

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

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

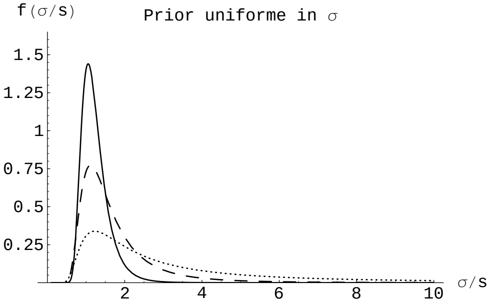
(Gaussian, independent observations – prior uniform on *mu* and σ)



$n = 3$ (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"

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

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"

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

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'):

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

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'):

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

It follows

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

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'):

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

It follows

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

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 $\tau = 1/\sigma^2$)

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

Inferring μ and σ from a sample

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

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

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

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

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 $\overline{x}$)

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

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 $\tau = 1/\sigma^2$)

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

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,

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

$\Rightarrow$ 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.