Quantitative Genetics, Prediction and Selection Theory

SELECTION INDICES

BLUP and REML

BAYESIAN INFERENCE

GENOMIC SELECTION

Bayesian Methods

1. Bayesian inference

Bayes' theorem

Prior probability

Probability density functions

2. The mixed model in Bayesian inference

Predicting additive values and estimating genetic parameters

Marginalization

Interlude: MCMC

BLUP and REML considered as Bayesian methods

Interlude: Inference on breeding values and genetic parameters under selection

3. The multitrait model

Traits having the same model

Data augmentation

Bayes theorem

A: to be man
B: to be British
N: Total number of individuals
N_A: number of men
N_B: number of British people
N_{AB}: number of British men

$$P(A,B) = \frac{N_{AB}}{N}$$

But if we take only the British people, the probability of being a man is

$$P(A|B) = \frac{N_{AB}}{N_B}$$

Bayes theorem

$$P(A,B) = \frac{N_{AB}}{N} = \frac{N_{AB}}{N_B} \cdot \frac{N_B}{N} =$$

$$= P(A|B) \cdot P(B)$$

Bayes theorem

$$P(A,B) = P(A|B) \cdot P(B)$$
$$= P(B|A) \cdot P(A)$$

$$P(A|B) = P(B|A) \cdot P(A) \, / \, P(B)$$

Bayes theorem

Model: y = Group + e

Group: Selected line (**S**)

Control line (**C**)

Question: Is $S \neq C$?

Bayes theorem

Model: y = Group + e

Group: Selected line (**S**)
Control line (**C**)

Question: Is S>C ?

P(S>C) ? or P(S-C>0) ?

Bayes theorem

$$P(A|B) = P(B|A) \cdot P(A) \,/\, P(B)$$

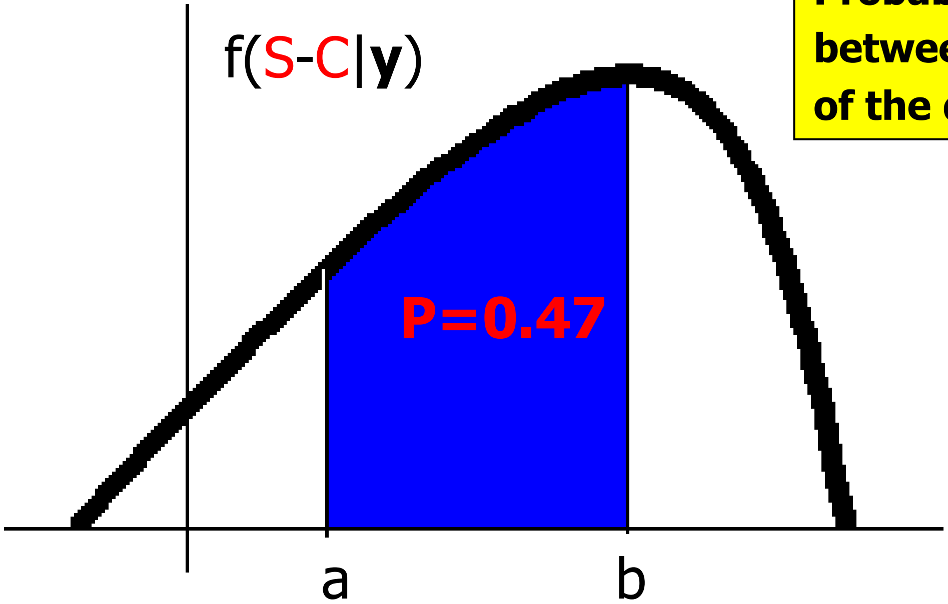
$$P(S-C|y) = P(y|S-C) \cdot P(S-C) \,/\, P(y)$$

S: Selected line

C: Control line

y: data

Density functions



Probability os S-C **being between** a **and** b **is the area of the distribution** P(S-C|**y**)

Prior information

4.1. Exact prior information

4.2. Vague prior information

4.3. No prior information

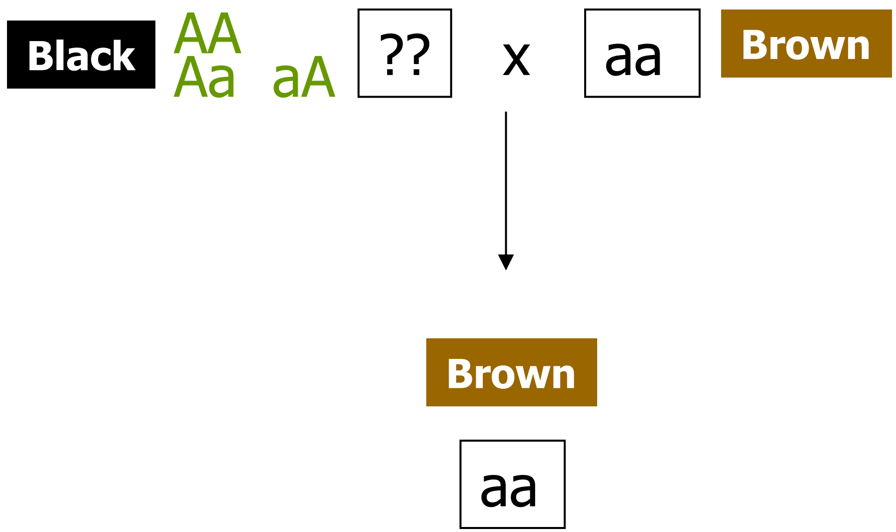
4.4. Improper priors

4.5. The Achilles heel of Bayesian inference

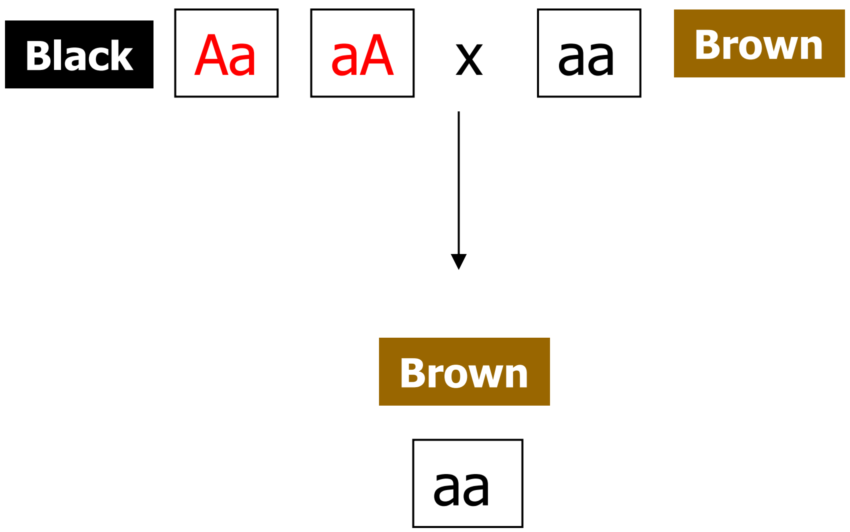
Exact Prior information

Black	AA		
Black	Aa	Black	aA
Brown	aa		

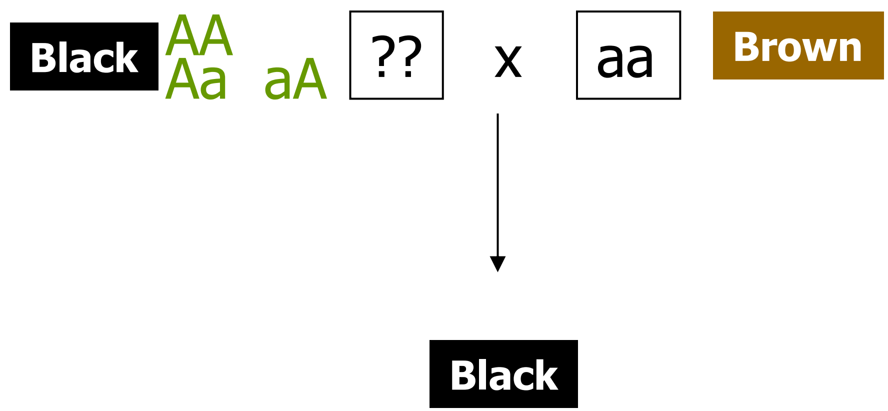
Exact Prior information



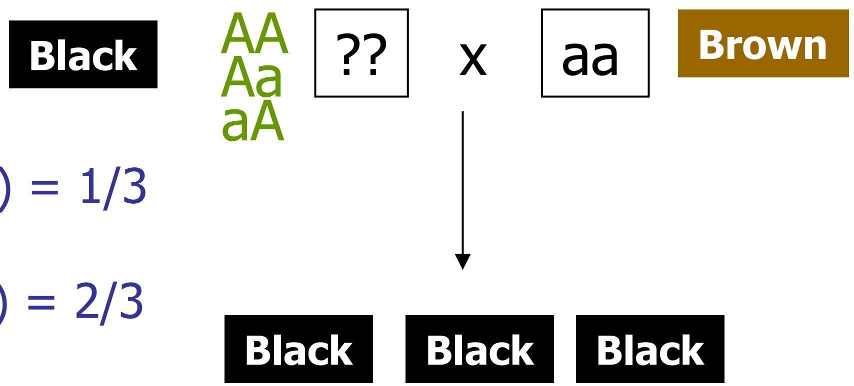
Exact Prior information



Exact Prior information



Exact Prior information



P(AA) = 1/3

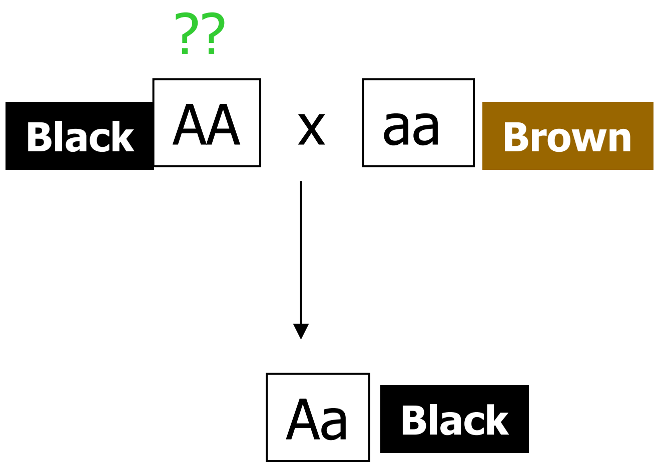
P(Aa) = 2/3

Exact Prior information

P(A|B) = P(B|A) P(A)/P(B)

P(AA | y=3**B**) = P(y=3**B** |AA) · P(AA) / P(y=3**B**)

- P(y=3**B** |AA) = 1
- P(AA) = 1/3
- P(y=3**B**)

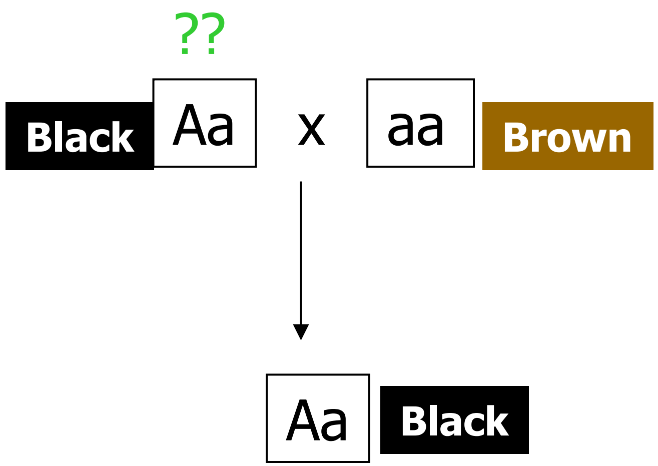


Exact Prior information

$$P(A|B) = P(B|A)\,P(A)/P(B)$$

$$P(Aa \mid y=3\mathbf{B}) = P(y=3\mathbf{B} \mid Aa) \cdot P(Aa) \,/\, P(y=3\mathbf{B})$$

- $P(y=3\mathbf{B} \mid Aa) = (1/2)^3$
- $P(Aa) = 2/3$
- $P(y=3\mathbf{B})$



Exact Prior information

P(A,B) = P(A|B) P(B)

$P(y=3\mathbf{B}) = \quad (y=3\mathbf{B} \ \& \ AA) \quad + \quad P(y=3\mathbf{B} \ \& \ Aa) \quad =$

$P(y=3\mathbf{B} \,|AA)\, P(AA) \ + \ P(y=3\mathbf{B} \,|Aa)\, P(Aa) =$

$1 \cdot 1/3 \quad + \quad (1/2)^3 \cdot 2/3 \quad =$

$= \ 5/12 \ = \ \mathbf{0.42}$

Exact Prior information

P(AA | y=3**B**) = P(y=3**B** |AA) · P(AA) / P(y=3**B**)

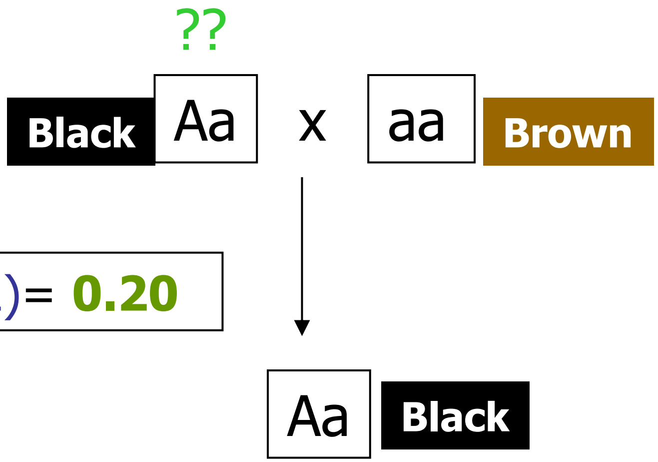
- P(y=3**B** |AA) = 1
- P(AA) = 1/3
- P(y=3**B**) = 5/12

P(AA | y=3**B**) = 1 · (1/3) / (5/12) = **0.80**

Exact Prior information

$P(Aa \mid y=3\mathbf{B}) = P(y=3\mathbf{B} \mid Aa) \cdot P(Aa) / P(y=3\mathbf{B})$

- $P(y= 3\mathbf{B} \mid Aa) = (1/2)^3$
- $P(Aa) = 2/3$
- $P(y=3\mathbf{B}) = 5/12$



$P(Aa \mid y=3\mathbf{B}) = (1/2)^3 \cdot (2/3) / (5/12) = \mathbf{0.20}$

Exact Prior information

Notice that

P(AA | y=3**B**) = 0.80

P(Aa | y=3**B**) = 0.20

1.00

However, the likelihoods

P(y= 3**B** |AA) = 1

P(y= 3**B** |Aa) = $(1/2)^3$= 0.125

By ML we choose AA without a measure of uncertainty

Exact Prior information

WITH FLAT PRIOR INFORMATION

$P(AA) = 1/2$
$P(Aa) = 1/2$

$P(A? \mid y=3\mathbf{B}) = P(y= 3\mathbf{B} \mid A?) \cdot P(A?) / P(y=3\mathbf{B})$

$P(y=3\mathbf{B}) = P(y= 3\mathbf{B} \mid AA)\, P(AA) + P(y= 3\mathbf{B} \mid Aa)\, P(Aa) =$
$= 1 \cdot 1/2 + (1/2)^3 \cdot 1/2 = 9/16 = 0.56$

$P(AA \mid y=3\mathbf{B}) = 1 \cdot (1/2) / (9/16) =$ **0.89**
$P(Aa \mid y=3\mathbf{B}) = (1/2)^3 \cdot (1/2) / (9/16) =$ **0.11**

Exact Prior information

WITH HIGH PRIOR INFORMATION

P(AA) = 0.002

P(Aa) = 0.998

P(A? | y=3**B**) = P(y= 3**B** | A?) · P(A?) / P(y=3**B**)

P(y=3**B**) = P(y= 3**B** |AA) P(AA) + P(y= 3**B** |Aa) P(Aa) =
= 1 · 0.002 + $(1/2)^3$ · 0.998 = 0.13

P(AA | y=3**B**) = 1 · 0.002 / 0.13 = **0.02**

P(Aa | y=3**B**) = $(1/2)^3$ · 0.998 / 0.13 = **0.98**

Exact Prior information

Black Aa x Aa Black

Black AA Aa aA ?? x aa Brown

P(AA) = 0.33

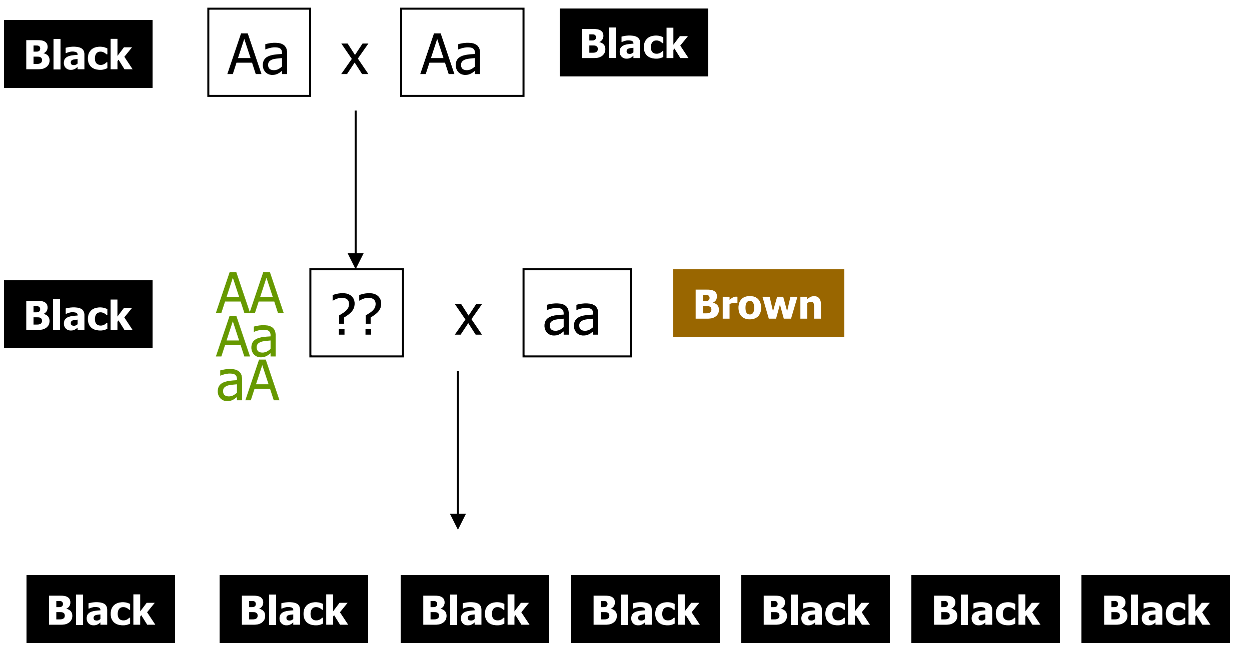
P(Aa) = 0.67

Black Black Black

P(AA|y) = **0.80**

P(Aa|y) = **0.20**

Exact Prior information



Exact Prior information

P(AA) = 0.33
P(Aa) = 0.67

P(AA|y=3**B**) = **0.80**
P(Aa|y=3**B**) = **0.20**

P(AA|y=7**B**) = **0.99**
P(Aa|y=7**B**) = **0.01**

Exact Prior information

P(AA) = 0.33
P(Aa) = 0.67
→
P(AA|y=3**B**) = **0.80**
P(Aa|y=3**B**) = **0.20**

P(AA) = 0.50
P(Aa) = 0.50
→
P(AA|y=3**B**) = **0.89**
P(Aa|y=3**B**) = **0.11**

P(AA) = 0.33
P(Aa) = 0.67
→
P(AA|y=7**B**) = **0.99**
P(Aa|y=7**B**) = **0.01**

P(AA) = 0.50
P(Aa) = 0.50
→
P(AA|y=7**B**) = **0.99**
P(Aa|y=7**B**) = **0.01**

When more data, prior is irrelevant

Exact Prior information

When more data, prior becomes irrelevant

$$f(\theta \mid \mathbf{y}) \propto f(\mathbf{y} \mid \theta) f(\theta) = f(y_1, y_2, \ldots, y_n \mid \theta) f(\theta) =$$

$$= f(y_1 \mid \theta) f(y_2 \mid \theta) \ldots f(y_n \mid \theta) \boxed{f(\theta)}$$

$$\log f(\theta \mid \mathbf{y}) \propto \log f(y_1 \mid \theta) + \log f(y_2 \mid \theta) + \ldots + \log f(y_n \mid \theta) + \boxed{\log f(\theta)}$$

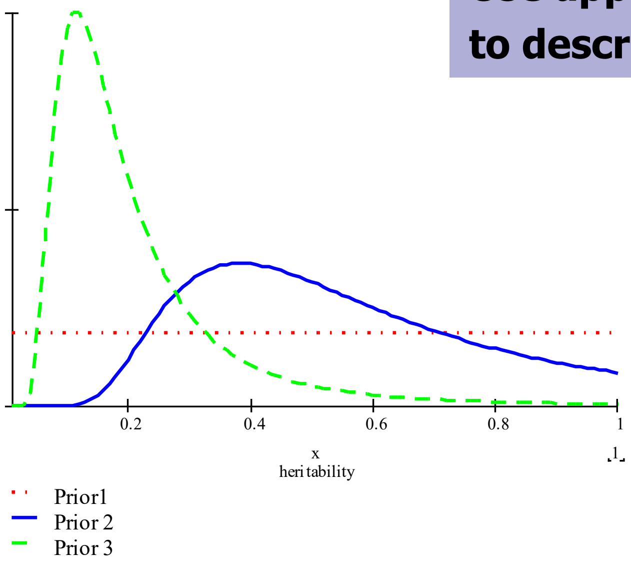
Vague prior information

- PROBABILITY describes BELIEFS
  - Subjective probability is not arbitrary
  - It should be vague (otherwise, no reason to perform an experiment)
  - When not vague, make conditional inferences (avoid problems)

- USE APPROPRIATE PRIOR DENSITIES
  - Linear beliefs (for effects, etc.) are symmetrical: Normal for example.
  - Quadratic beliefs (for variances, h^2, etc.) are assymetrical. I-gamma for example

- TRY SEVERAL PRIORS
  - If posteriors are almost the same, prior information is irrelevant

Vague prior information



Use appropriate functions to describe vague prior knowledge

Blasco et al. 1998
Genetics 143: 301-306

Vague prior information

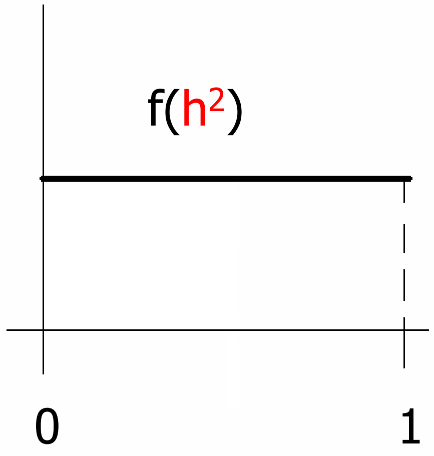
PROBLEMS

- How can be integrated information from other experiments?
  - Is your experiment fully comparable with other experiments?
  - Do you believe in ALL published results?
- How can you define multivariate beliefs?
- The posterior of tod… …morrow's prior

Bayesian propaganda !!

No Prior information

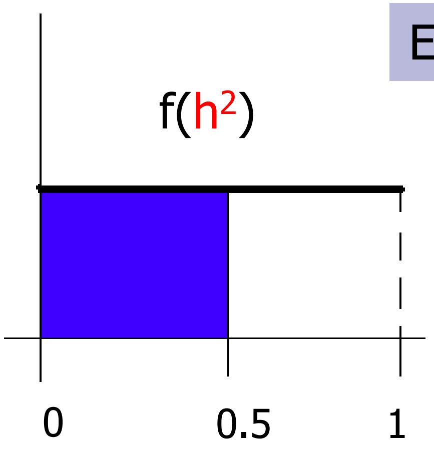
FLAT PRIORS



No Prior information

FLAT PRIORS

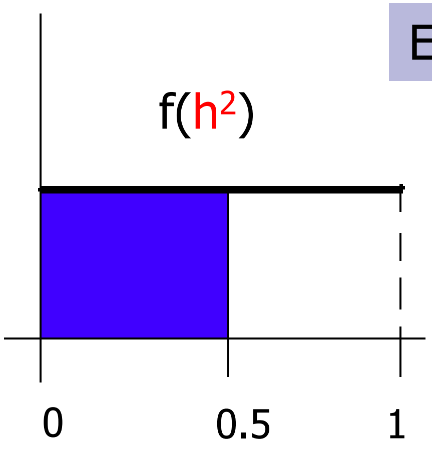
Event A ≡ $h^2 < 0.5$



No Prior information

FLAT PRIORS

Event A ≡ $h^2<0.5$ ≡ $h^4<0.25$



P(A) = $P(h^2<0.5)$ = ½

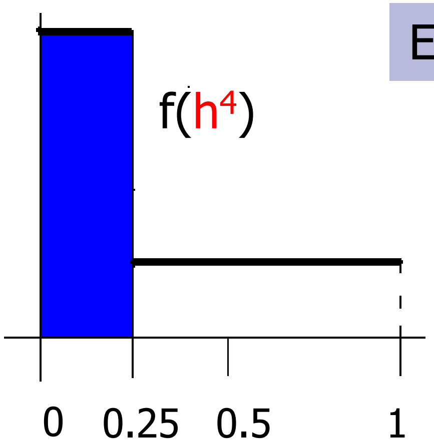
P(A) = $P(h^4<0.25)$ = ½

No Prior information

Flat priors
Are informative !!

FLAT PRIORS

Event A ≡ $h^2<0.5$ ≡ $h^4<0.25$



$P(A) = P(h^2<0.5) = ½$

$P(A) = P(h^4<0.25) = ½$

No Prior information

- Non informative priors are informative
- Thus, we introduce information we do not know where it comes from
- Some non-informative priors minimize the information introduced
- We should avoid (in general) improper priors
- We should check how results are affected by using a prior, even if this prior is non-informative
- … but we do not know how to do this in the multivariate case

No Prior information

- Other alternatives have been proposed:
    - Jeffrey's priors:

      They are invariant to transformations
    - Bernardo's Reference priors:

      Minimum prior information
    - Maximum entropy prior information:

      Minimum prior information with some subjective informative restrictions

- However, all of them have problems in the multivariate case

No Prior information

MULTIVARIATE PRIORS

- We cannot have subjective multivariate priors
    - Subjective priors: hire a psicoanalist
- We cannot have ‘objective’ multivariate priors !!
    - Do not use big flat priors
    - Do not use almost big flat priors!!
    - Be careful with some common priors like inverted Wishart !!
- A practical solution:
    - Flat priors with sound limits
    - Vague Informative priors with sound limits

Improper priors

- Some priors are not densities
  - Example: $f(\theta) = k$ k: arbitrary constant

    $\int f(\theta)\, d\theta = \infty$

- They can produce improper posterior densities
- They lead to proper posterior densities when

  $f(y) = \int f(y|\theta)\, f(\theta)\, d\theta < \infty$

Improper priors

- Sometimes they are innocuous

Example: $y \sim N(\mu,1)$ $\mu \sim k$

$$f(y)=\int_{-\infty}^{\infty} f(y\mid\mu)f(\mu)d\mu=\int_{-\infty}^{\infty} f(y\mid\mu)k\,d\mu=k\int_{-\infty}^{\infty} f(y\mid\mu)d\mu=$$

$$=k\int\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(y-\mu)^2}{2}\right]d\mu=k\int\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(\mu-y)^2}{2}\right]d\mu=k$$

$$f(\mu\mid y)=\frac{f(y\mid\mu)f(\mu)}{f(y)}=\frac{f(y\mid\mu)\cdot k}{k}=f(y\mid\mu)$$

But sometimes they are not !

Improper priors

- Improper priors may lead to improper posteriors
- When using MCMC improper posteriors may not be detected
- Do not use improper priors !!
  - Do not use big flat priors
  - Do not use almost big flat priors**!!** E.g.: $f(\theta) \sim N(0, 10^6)$

  (they behave as impropre priors and give a false sense of safety)
- A practical solution:
  - Flat priors with sound limits
  - Vague Informative priors with sound limits

No Prior information

- Modern Bayesians consider prior information as just a mathematical artefact that allow us to work with probabilities

  … but PROBABILITY x ARTEFACT ≠ PROBABILITY

  PROBABILITY x ARTEFACT = ARTEFACT

- If we **behave** as if it is a probability, the distortion is not high

  … and we can enjoy the advantages of working with probabilities

The Bayesian choice

SOME DISSAPOINTMENTS ALONG YOUR LIFE:

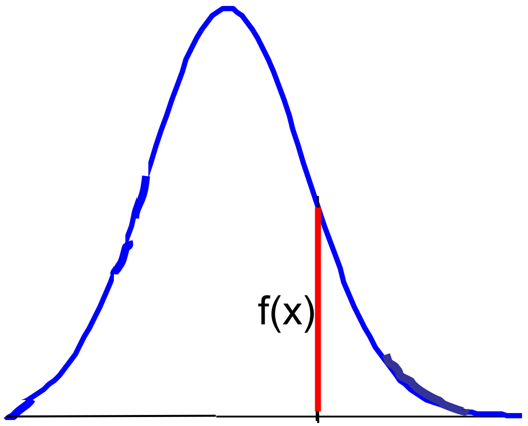
- Father Christmas are mum and dad
- In the improbable case of the existence of the Heaven, nobody has make a reservation for you there
- Bayesian methods have conceptual problems, as the frequentists ones, and it is … or the other one

My opinio… have a more straightforwa… can understand better what y…

The great advantage is to work with Probabilities

Density function

f(x) is **not** a probability

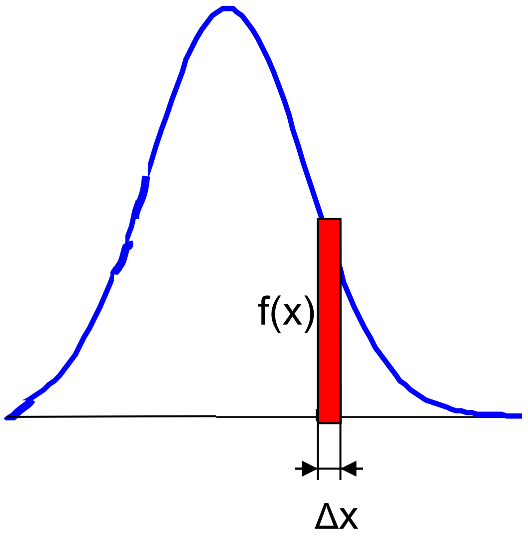


Density function

f(x) is **not** a probability

f(x)·Δx is approx. a probability

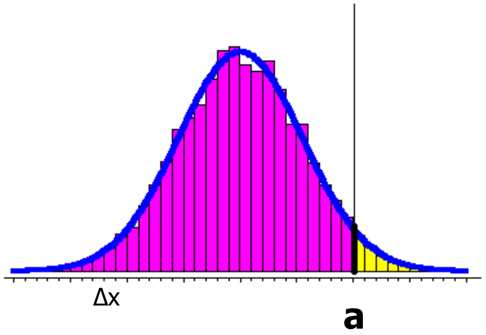
f(x)·dx is a probability



Density function

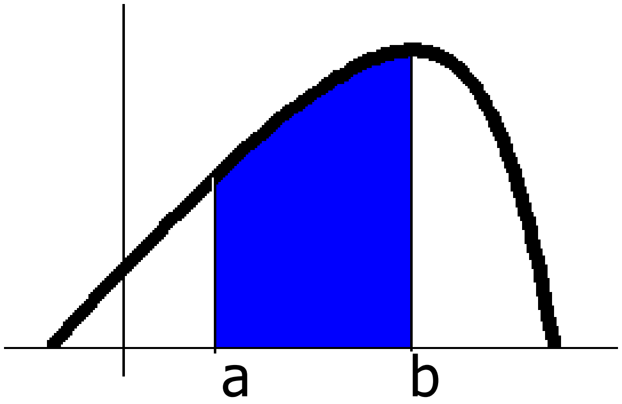
$f(x) \cdot dx$ is a probability

$$\int_{a}^{+\infty} f(x)\,dx$$ is a probability (a sum of probabilities)



Density function

$$P(a \le x \le b) = \int_a^b f(x)\,dx$$



Conditional distribution

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$f(x|y)\Delta x = \frac{f(y|x)\Delta y \cdot f(x)\Delta x}{f(y)\Delta y}$$

$$f(x|y) = \frac{f(y|x) \cdot f(x)}{f(y)}$$

Conditional distribution

$$f(x,y) = f(x|y) \cdot f(y)$$

$$f(x|y) = \frac{f(x,y)}{f(y)} \longrightarrow \boxed{f(x|y) = \frac{f(x,y)}{f(y)}}$$

$$f(x|\,y = y_0) = \frac{f(x, y_0)}{f(y_0)}$$

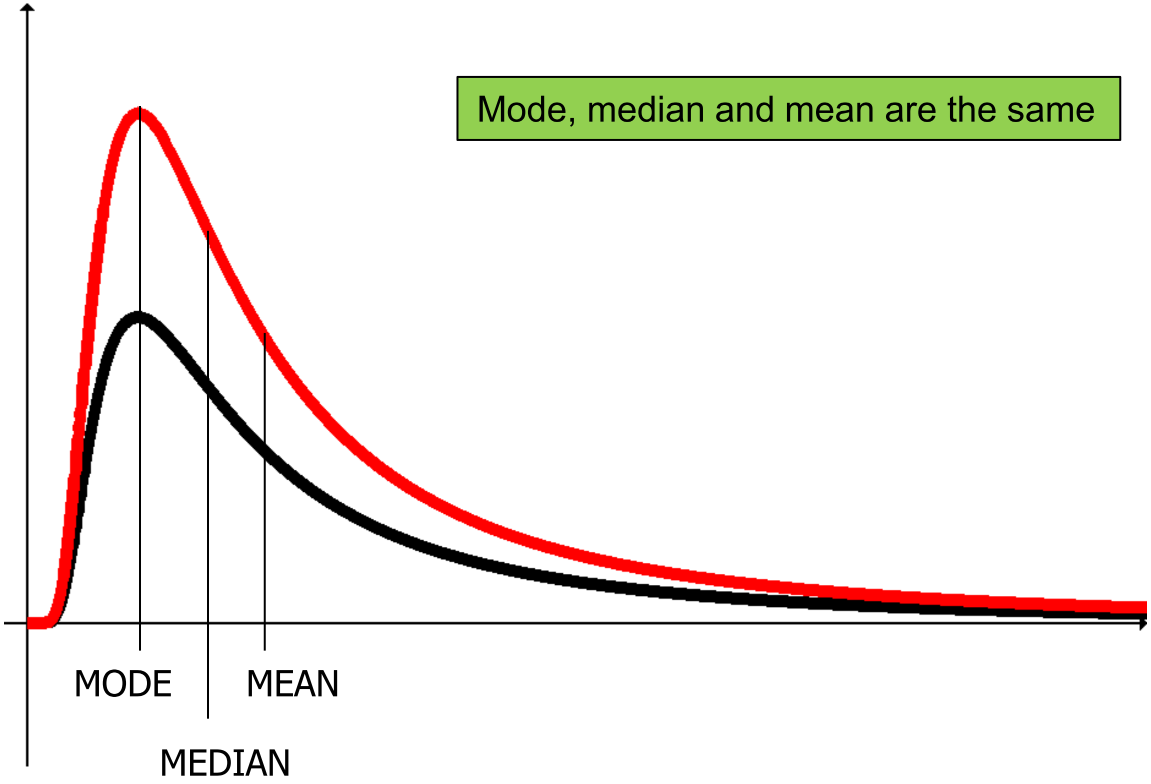
Conditional distribution

$$f(x \mid y) = \frac{f(y \mid x) \cdot f(x)}{f(y)}$$

$$f(x \mid y) = \frac{f(y \mid x) \cdot f(x)}{f(y)}$$

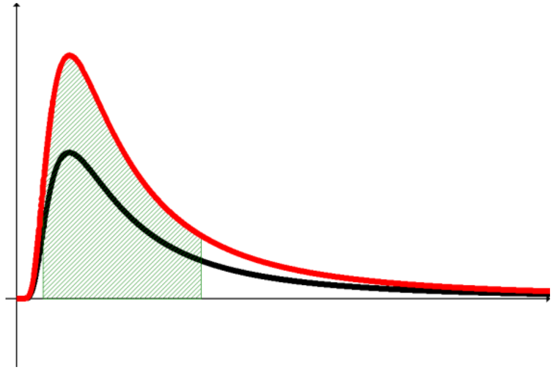
$$f(x \mid y) \propto f(y \mid x) \cdot f(x)$$

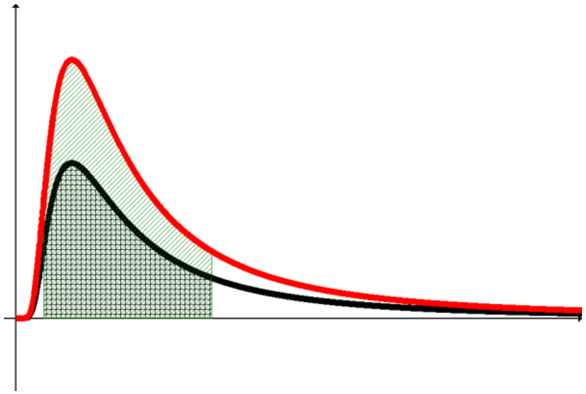
Working proportionally



Working proportionally

Probabilities are the same





Risk, bias and variance

ERROR OF ESTIMATION

$$e = u - \hat{u}$$

LOSS FUNCTION

$$l(\hat{u}, u) = e^2$$

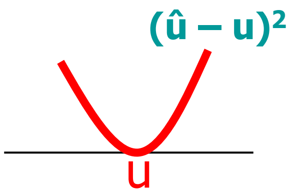
RISK

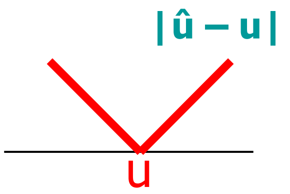
$$R(\hat{u}, u) = E\left[l(\hat{u}, u)\right] = E\left(e^2\right)$$

Features of Bayesian inference

POINT ESTIMATES

- MEAN: minimizes *RISK* = $E(\hat{u} - u)^2$
- MEDIAN: minimizes *RISK* = $E|\hat{u} - u|$
- MODE: is the most probable value





A. Blasco. *Bayesian data análisis of animal scientists*. Appendix 2.1

Features of Bayesian inference

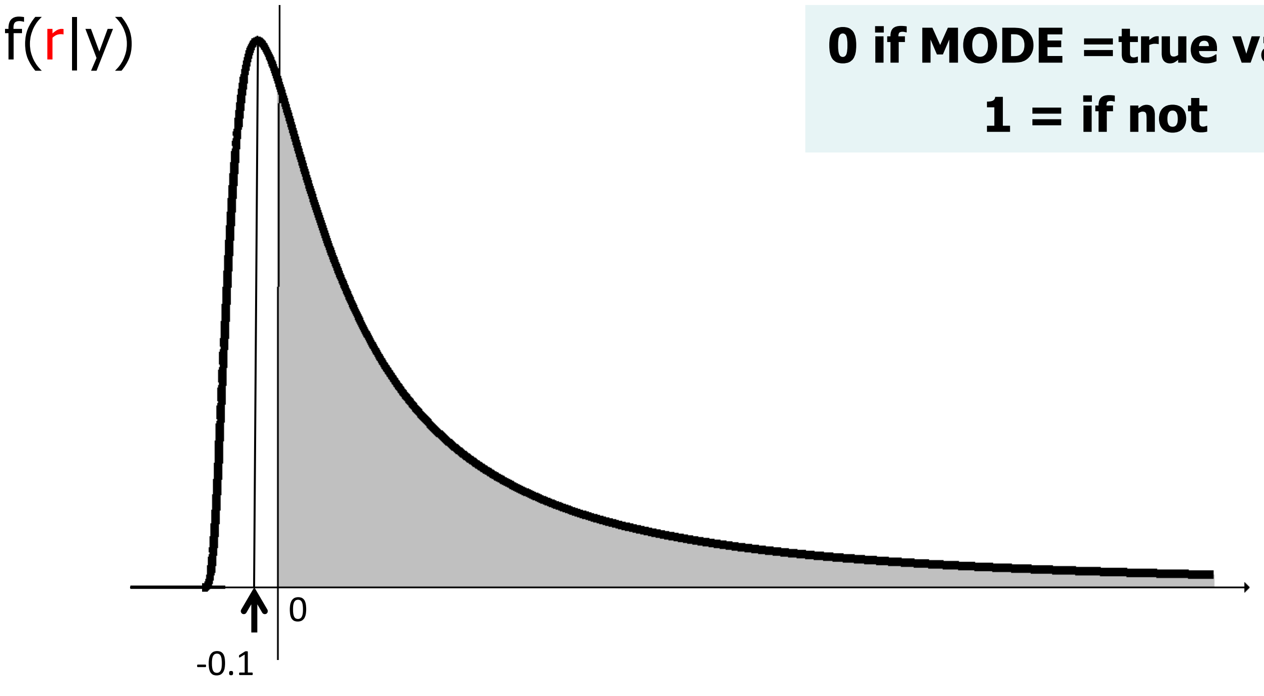
POINT ESTIMATES

… but MODE has a horrible loss function!

0 if MODE =true value
1 = if not

Features of Bayesian inference

POINT ESTIMATES



0 if MODE =true value
1 = if not

Features of Bayesian inference

POINT ESTIMATES

… but MEAN also has a horrible loss function!

	x	x^2
	1	1
	2	4
	3	9
Mean	**2**	**4.7**

$2^2 \neq 4.7$

$(\hat{u} - u)^2$ is NOT invariant to transformations !!

i.e.: the loss of u^2 is not the square of the loss of u !!

i.e.: the MEAN for σ^2 is not the square of the MEAN of σ

Features of Bayesian inference

POINT ESTIMATES

Only the MEDIAN is invariant

x =   1 1 1 2 2 3 4 4 5 5 5

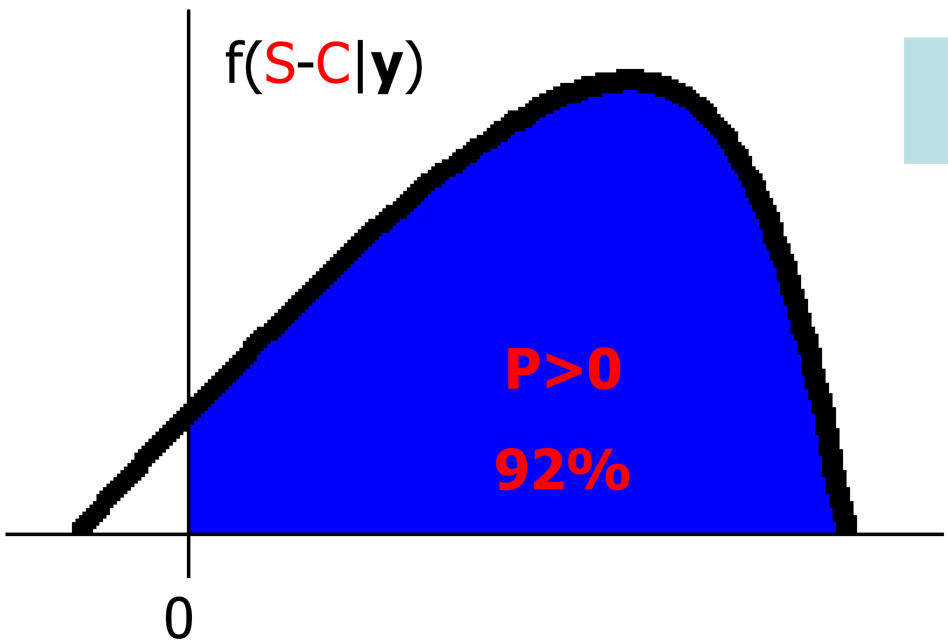
median x = 3

x^2=1 1 1 4 4 9 16 16 25 25 25

median x = 9

Features of Bayesian inference

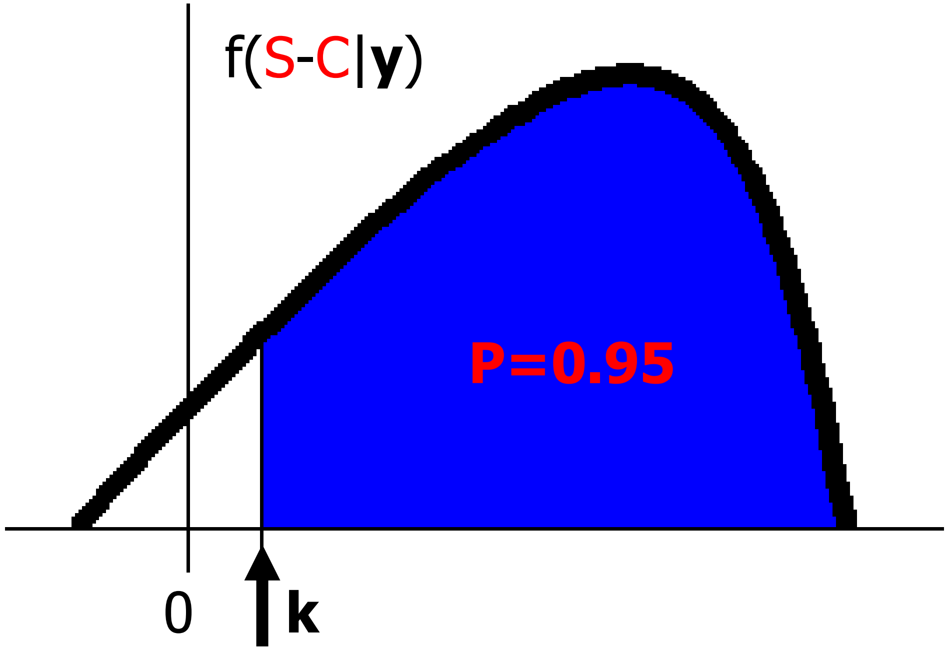
CREDIBILITY INTERVALS



Probability of S >C

Features of Bayesian inference

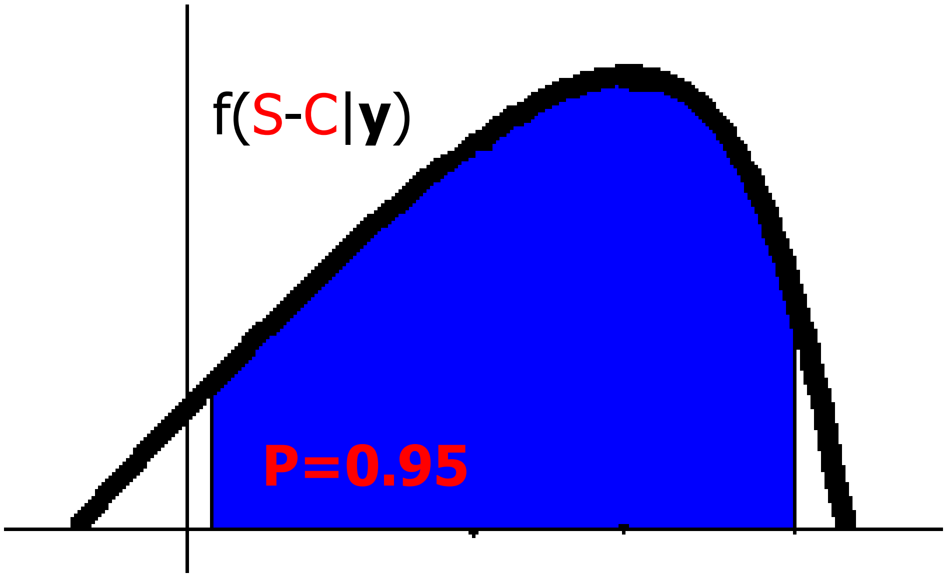
CREDIBILITY INTERVALS



k is a guaranteed value for P=95% or for P= 80%, or for other P

Features of Bayesian inference

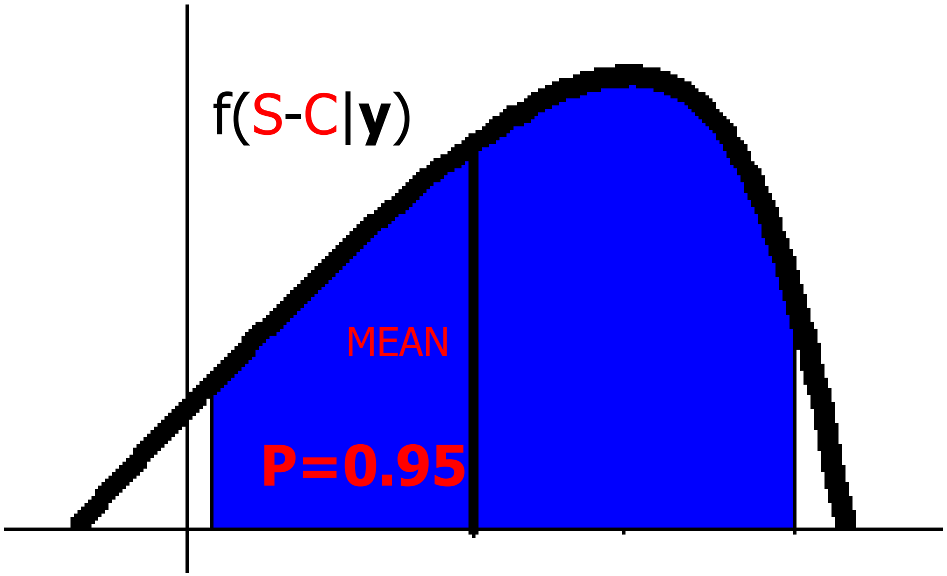
CREDIBILITY INTERVALS



Shortest interval with P=0.95

Features of Bayesian inference

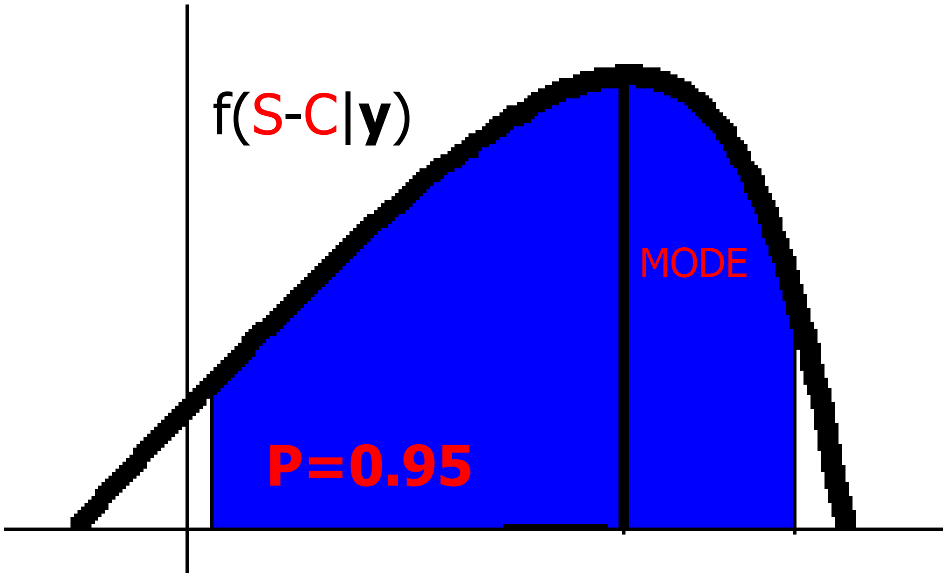
CREDIBILITY INTERVALS



Shortest interval with P=0.95

Features of Bayesian inference

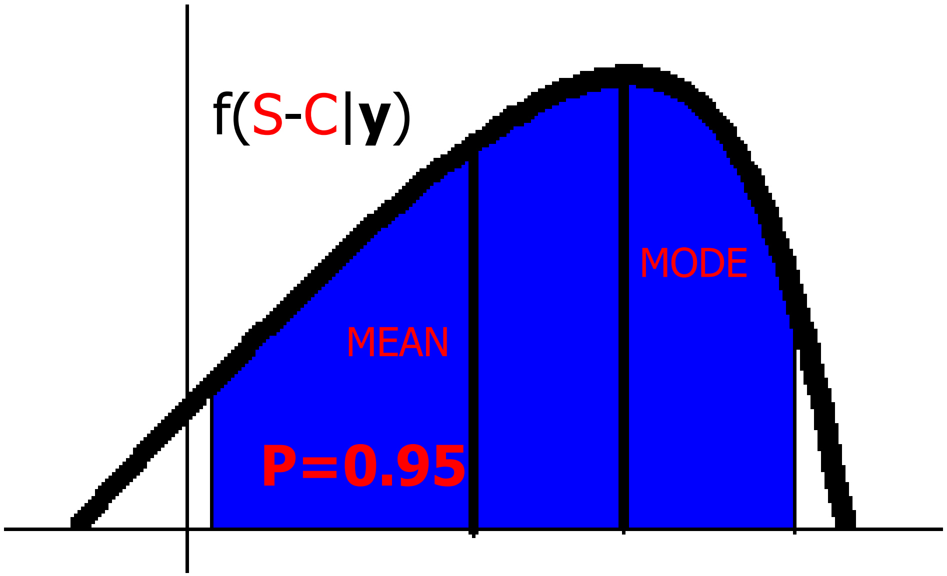
CREDIBILITY INTERVALS



Shortest interval with P=0.95

Features of Bayesian inference

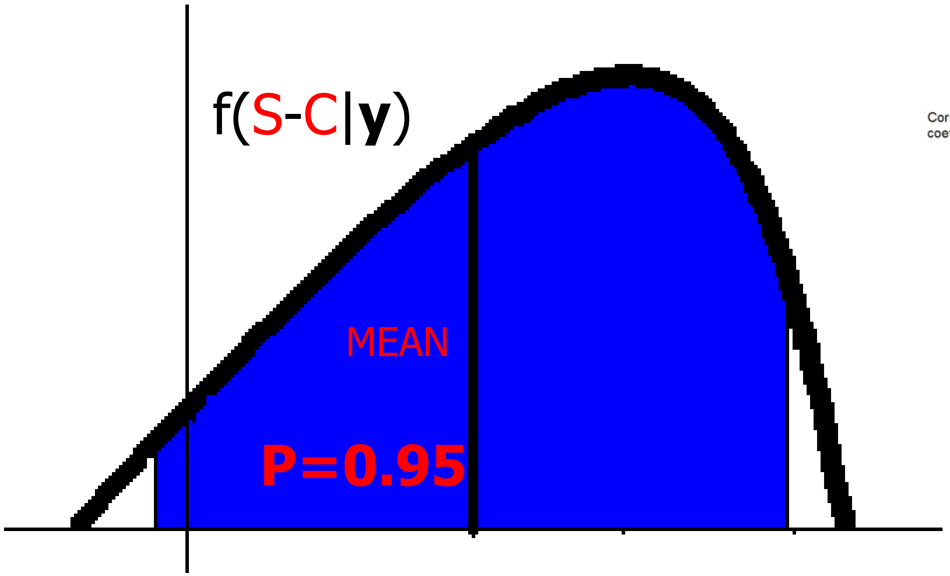
CREDIBILITY INTERVALS

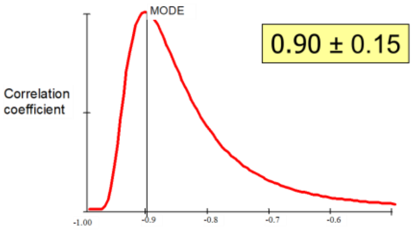


Shortest interval with P=0.95

Features of Bayesian inference

CREDIBILITY INTERVALS





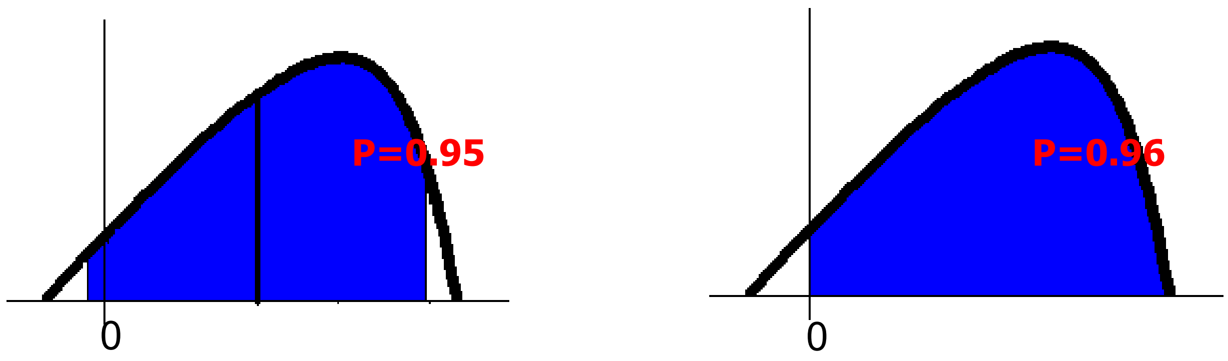
Symmetric interval with P=0.95

Features of Bayesian inference

CREDIBILITY INTERVALS

Notice that zero can be within the confidence interval and still P(S-C>0) can be >0.95

If 0 is within the HPD interval, this does not mean that there are no “significant differences”



Features of Bayesian inference

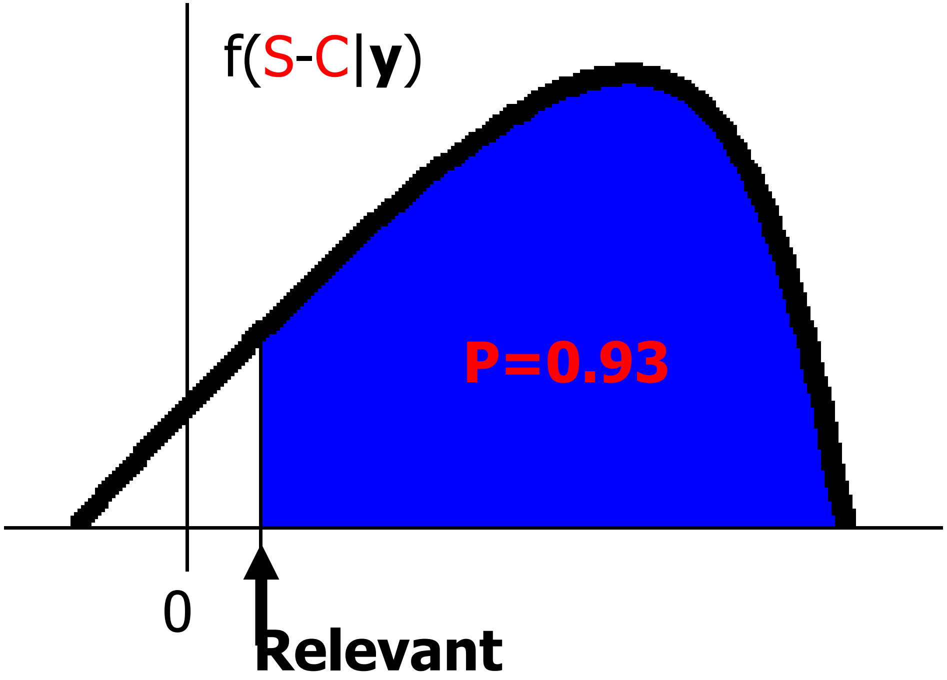
CREDIBILITY INTERVALS

Relevant value: the **minimum difference** between S and C having an economical or biological meaning

- It is the mínimum value from which we take a decision
- It is the value used for experimental designs
- It should be proposed for each trait based on biological or economical arguments
- When no clues, use a fraction of the standard deviation
- In animal production, most economical relevant values go from 1/2 to 1/3 s.d. of the trait

Features of Bayesian inference

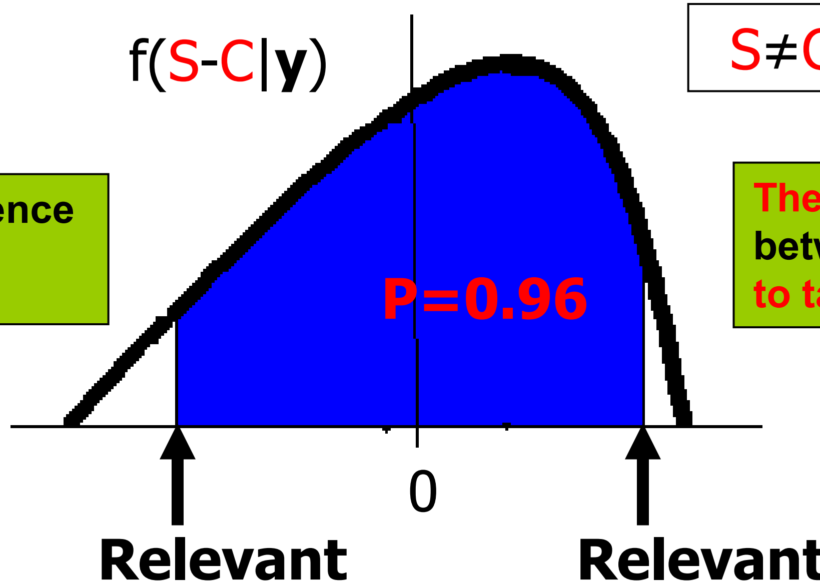
Probability of RELEVANCE



Features of Bayesian inference

Probability of SIMILITUDE

There is no difference
(in practice)
between S and C



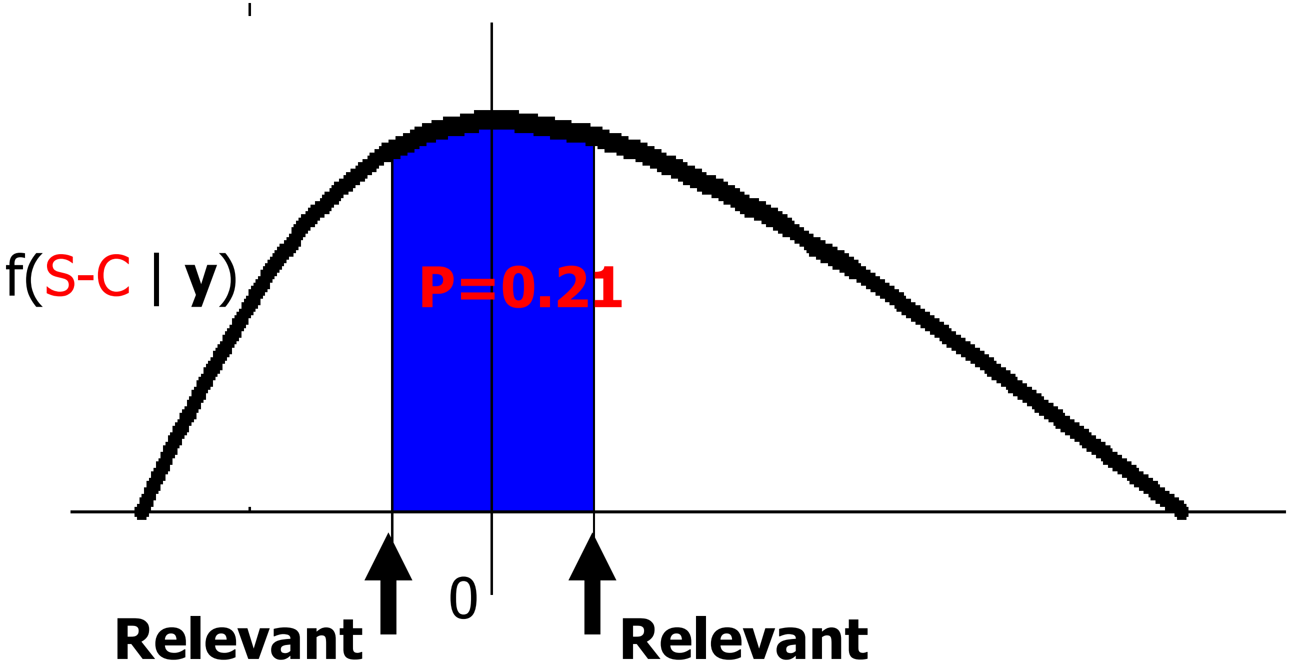
S≠C **always!**

There is no difference
between S and C
to take a decision

Features of Bayesian inference

Probability of SIMILITUDE

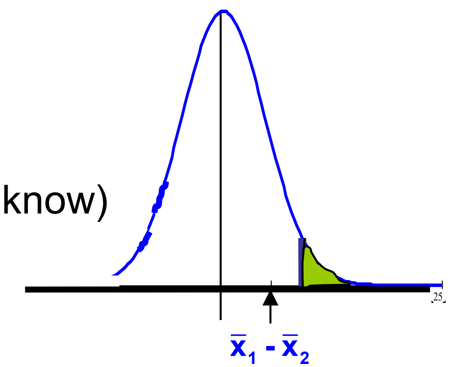
For my decision, I do not know whether $S>C$ or $S<C$



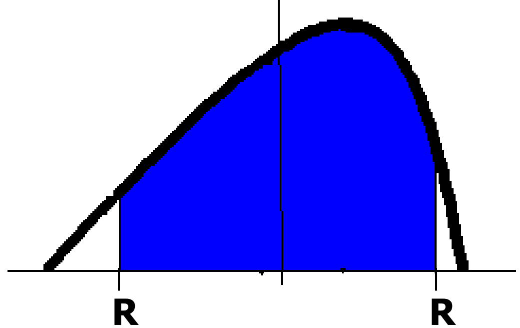
Features of Bayesian inference

Before being Bayesian

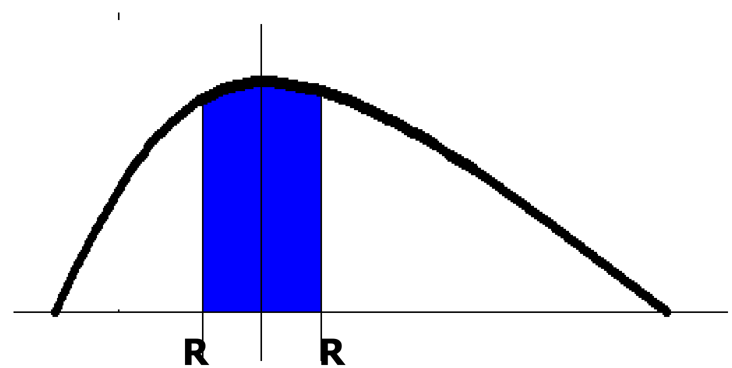
n.s. = no sé (I do not know)



After being Bayesian



There are no differences
(in practice)



I do not know

Features of Bayesian inference

CREDIBILITY INTERVALS

We still have the problem of which difference is “relevant” for many traits:

FLAVOUR: metallic, liver, grass, sweet, etc.
ENZYMES ACTIVITY, WHC, COLOUR, etc.

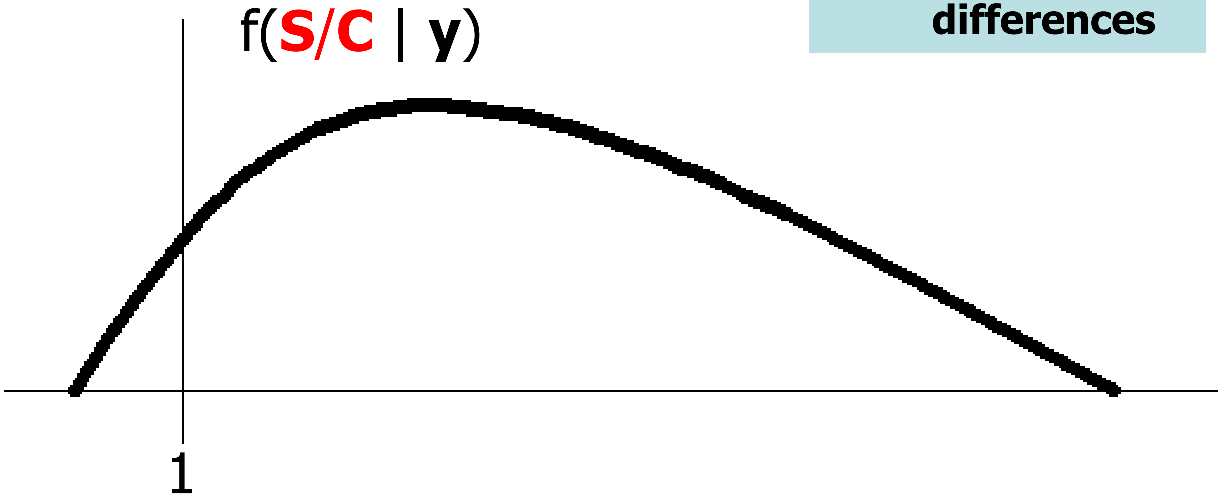
Relevant value: 1/2 or 1/3 SD of the trait

Relevant value: 5% or 10% higher (or lower)

Features of Bayesian inference

CREDIBILITY INTERVALS

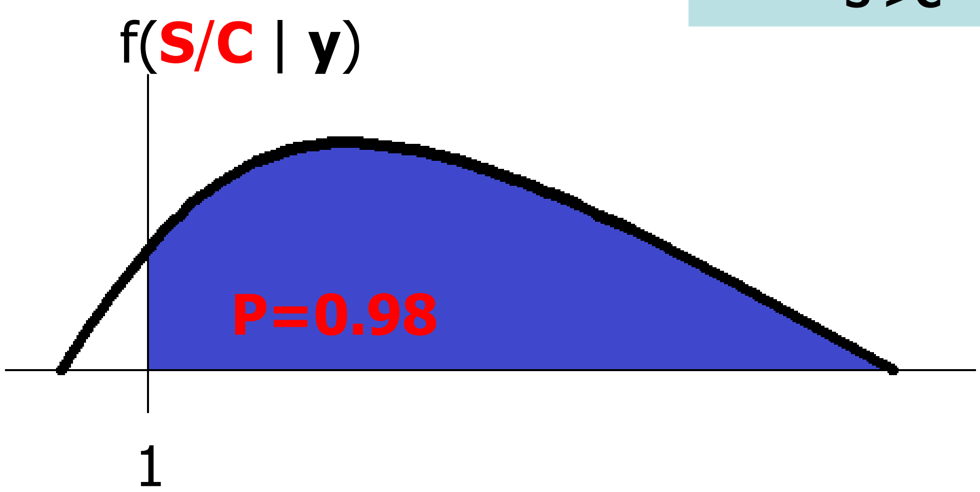
Use ratios instead of differences



Features of Bayesian inference

CREDIBILITY INTERVALS

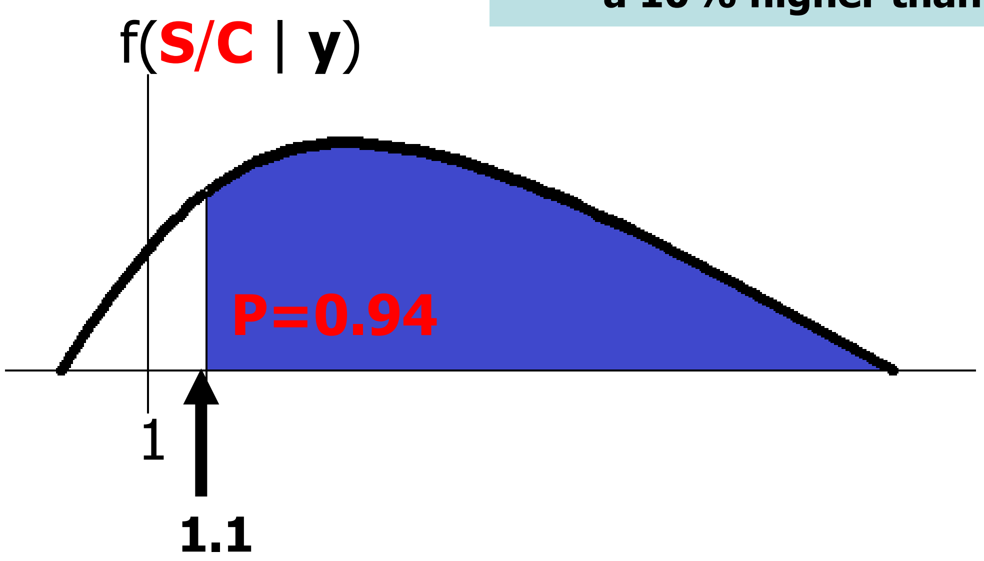
Probability of S >C



Features of Bayesian inference

CREDIBILITY INTERVALS

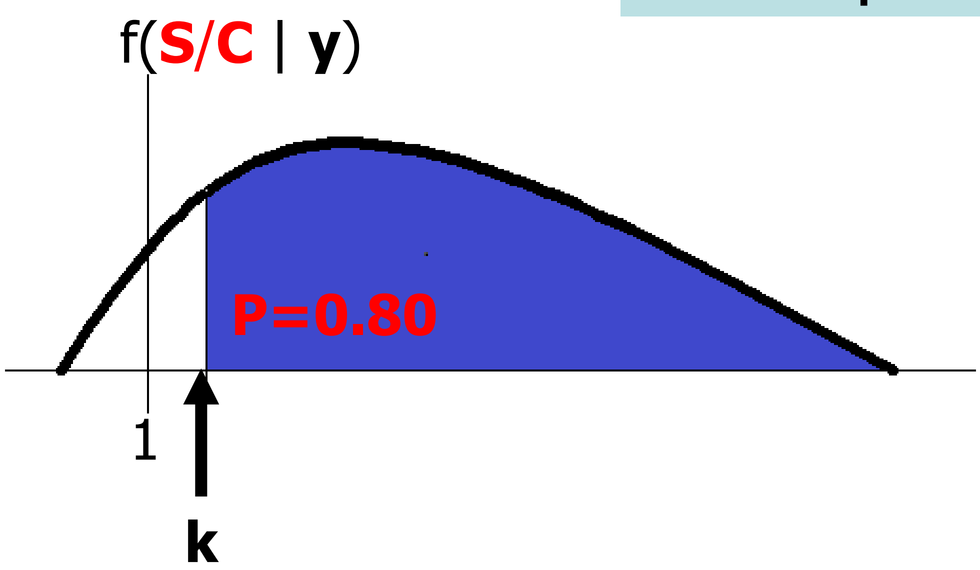
Probability of S being at least a 10% higher than C



Features of Bayesian inference

CREDIBILITY INTERVALS

S is k times higher than C with a probability of 80%



The Bayesian choice

2.1. Bayes theorem

2.2. Features of Bayesian inference

2.3. Marginalisation

2.4. Bayesian Hypothesis tests

2.5. Advantages of Bayesian inference

Marginalisation

SALARY	British **B**	Spanish **S**
Men **M**	36 (40%)	26 (10%)
Women **W**	30 (20%)	20 (30%)

$$f(M,B) = f(M \mid B) \cdot f(B) \rightarrow f(H \mid B) = \frac{f(M,B)}{f(B)} = \frac{0.4}{0.6} = \frac{2}{3}$$

$$f(W,B) = f(W \mid B) \cdot f(B) \rightarrow f(W \mid B) = \frac{f(W,B)}{f(B)} = \frac{0.2}{0.6} = \frac{1}{3}$$

$$\text{British salary} = 36 \cdot \frac{2}{3} + 30 \cdot \frac{1}{3} = 34$$

$$\text{Spanish salary} = 26 \cdot \frac{0.3}{0.3+0.1} + 20 \cdot \frac{0.1}{0.3+0.1} = 24.5$$

Marginalisation

SALARY	British	Spanish
Men	36 (40%)	26 (10%)
Women	30 (20%)	20 (30%)

SALARY	British	Spanish
	34 (60%)	24.5 (40%)

Marginalisation

EXAMPLE

$A = b \cdot y + e = h^2 \cdot y + e$

unknowns: A and h^2 data: **y**

Example: h^2 can only take two values: 0.1 or 0.2

$f(A \mid y) = f(A \mid h^2=0.1, \mathbf{y})\, P(h^2=0.1) + f(A \mid h^2=0.2, \mathbf{y})\, P(h^2=0.2)$

When estimating A, we take into account the error of estimation of h^2 (its probability of being 0.1 or 0.2)

Marginalisation

Summed up

$$f(x) = \int f(x,y)dy = \int f(x \mid y)\, f(y)\, dy$$

Each value of **y**

by its probability

$A = b \cdot y + e = h^2 \cdot y + e$

h^2 can take *any value* between 0 and 1

$$f(A \mid y) = \int_0^1 f(A \mid h^2, y)\, f(h^2) dh^2$$

all possible values of h^2 (which is ‘given’), weighted by their probabilities $f(h^2)dh^2$

The Bayesian choice

2.1. Bayes theorem

2.2. Features of Bayesian inference

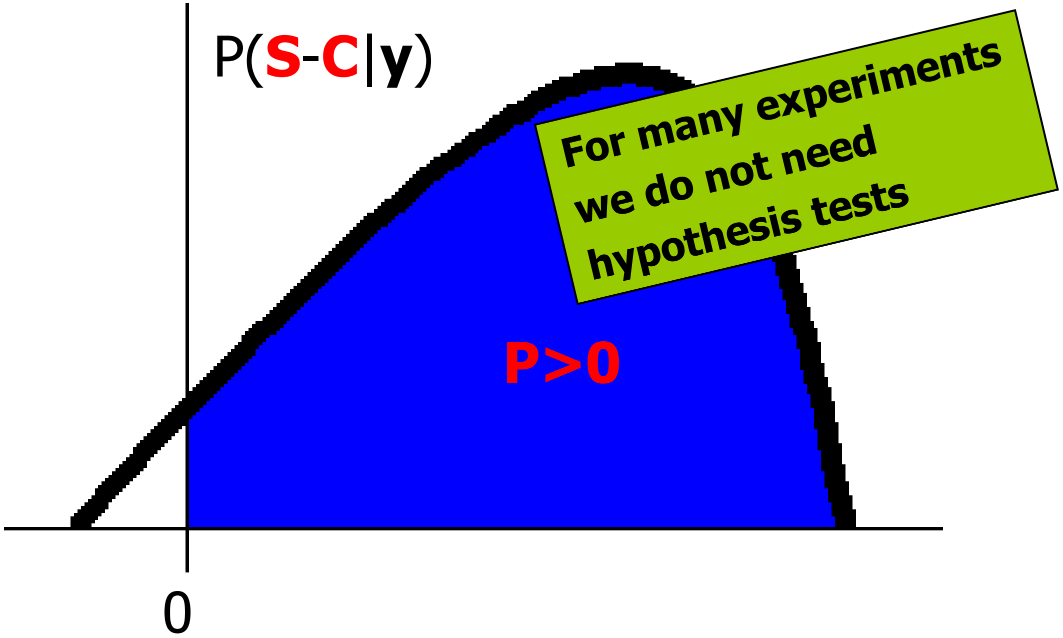
2.3. Marginalisation

2.4. Bayesian Hypothesis tests

2.5. Advantages of Bayesian inference

Hypothesis test

This is **NOT** a hypothesis test



Hypothesis test

Calculate the probability of each hypothesis

$P(M_1|\mathbf{y}), P(M_2|\mathbf{y}), P(M_3|\mathbf{y}), \ldots$

… and choose the M_i more probable

$$P(M_1 \mid \mathbf{y}) = \frac{P(\mathbf{y} \mid M_1) \cdot P(M_1)}{P(\mathbf{y})} = \frac{P(\mathbf{y} \mid M_1) \cdot P(M_1)}{P(\mathbf{y} \mid M_1) + P(\mathbf{y} \mid M_2) + \cdots}$$

$$M_1 : \quad \mathbf{y} = f\left(\boldsymbol{\theta}\right) + \mathbf{e}$$

$$P(\mathbf{y} \mid M_1) = \int f\left(\mathbf{y} \mid \boldsymbol{\theta}\right) \cdot f\left(\boldsymbol{\theta}\right) \cdot d\boldsymbol{\theta}$$

The Bayesian solution

BAYES FACTORS

$$\frac{P(M_0 \mid \mathbf{y})}{P(M_1 \mid \mathbf{y})} = \frac{P(\mathbf{y} \mid M_0) \cdot P(M_0) / P(\mathbf{y})}{P(\mathbf{y} \mid M_1) \cdot P(M_1) / P(\mathbf{y})} = BF \cdot \frac{P(M_0)}{P(M_1)}$$

THEY ARE SENSITIVE TO PRIORS

$$BF = \frac{P(\mathbf{y} \mid M_0)}{P(\mathbf{y} \mid M_1)}$$

$$\frac{P(M_0 \mid \mathbf{y})}{P(M_1 \mid \mathbf{y})} = BF$$

If $P(M_0) = P(M_1)$

Moreover, often $P(M_0) \neq P(M_1)$

Be careful !!

The Bayesian choice

2.1. Bayes theorem

2.2. Features of Bayesian inference

2.3. Marginalisation

2.4. Bayesian Hypothesis tests

2.5. Advantages of Bayesian inference

Advantages of Bayesian Inference

- We are not worried about bias (there is nothing like bias in a Bayesian context)
- We should not decide whether an effect is fixed or random (all of them are random)
- We normally do not need Hypothesis tests
- We have a measure of uncertainty for both hypothesis tests and credibility intervals, we work with Probabilities
- We work with marginal probabilities: i.e., all multivariate problems are converted in univariate.
- We have a method for inferences, a path to be followed.

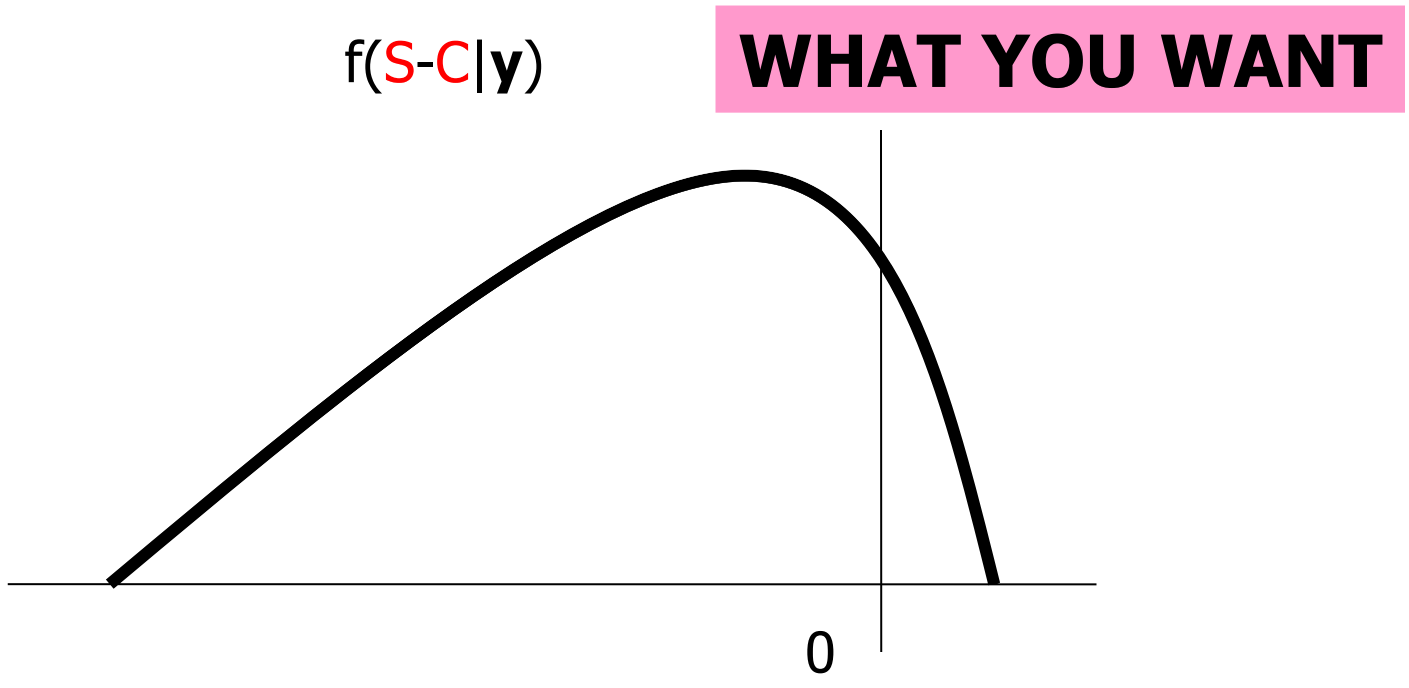
Interlude

MCMC

MCMC light

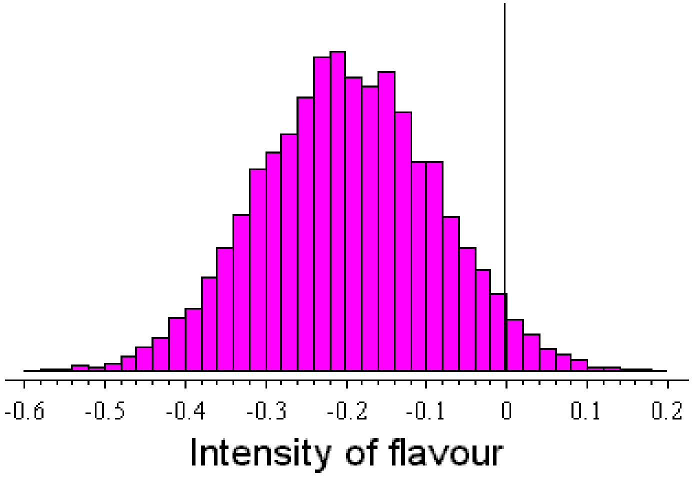
(without MCMC)

MCMC



MCMC

WHAT YOU GET



MCMC

- You get a sample of the marginal posterior distribution for each level of each effect in the model

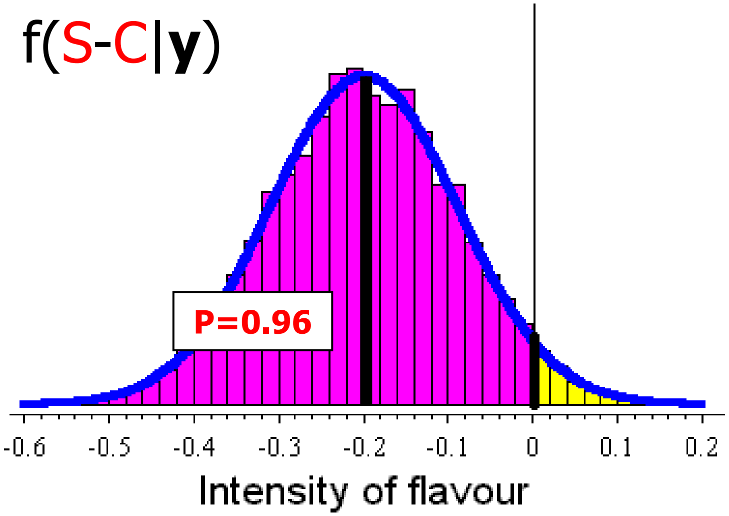
- You can create new samples as functions of the samples

(for example, S–C or S/C)

S: [3.1, 3.3, 4.1, 4.8, 4.9,…..**]**
C: [2.4, 2.6, 2.6, 2.6, 2.8,.....**]**

S–**C: [**0.7, 0.7, 1.5, 2.2, 2.1,………**]**
S/**C: [**1.3, 1.3, 1.6, 1.8, 1.7,………**]**

Marginal posterior distribution



Marginal posterior distribution

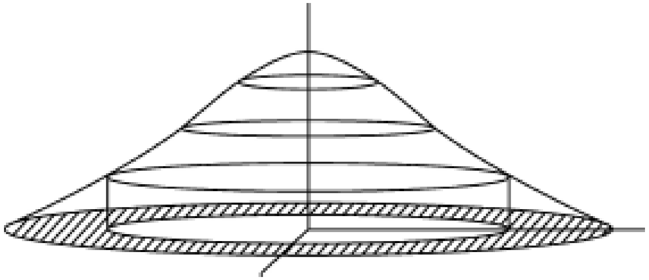
OBJECTIVE → $f(\mu \mid \mathbf{y})$

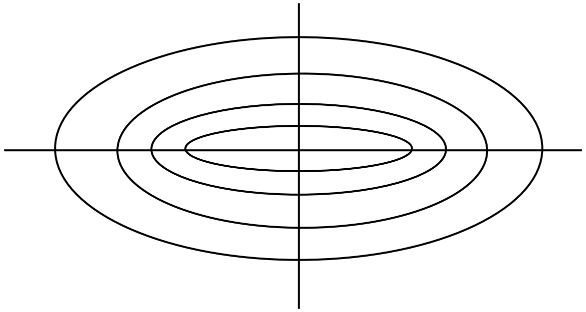
OBJECTIVE → $f(\sigma^2 \mid \mathbf{y})$

OBJECTIVE → random samples of $f(\mu \mid \mathbf{y})$: 12, 11, 15, 9, 6, $\cdots$

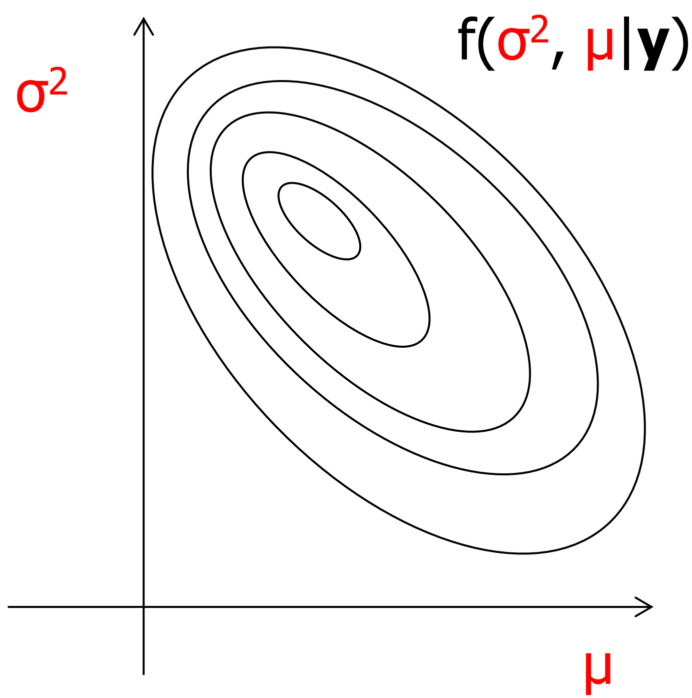
OBJECTIVE → random samples of $f(\sigma^2 \mid \mathbf{y})$: 3, 1.5, 1.8, 2, 3.4, $\cdots$

Gibbs sampling





Gibbs sampling



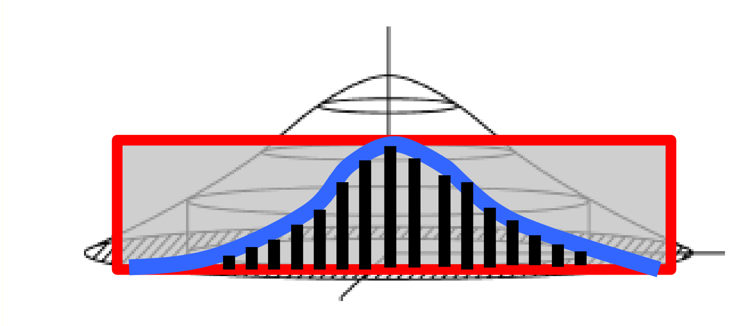
OBJECTIVE:
Random samples from f(σ^2, μ|**y**)

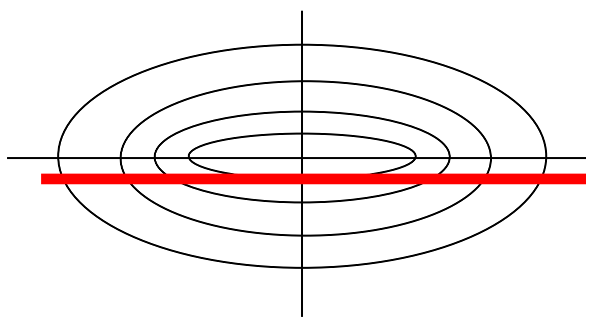
σ^2	μ
3	12
1.5	18
2	11
…	…
…	…

Random samples of f(σ^2|**y**)

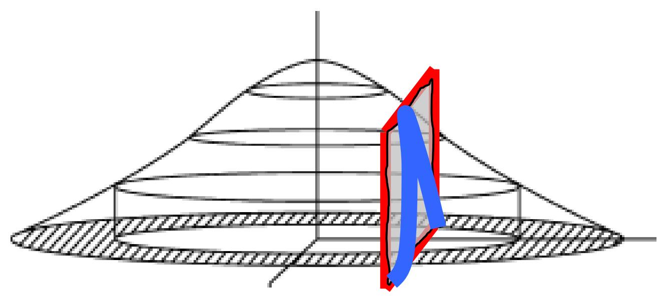
Random samples of f(μ |**y**)

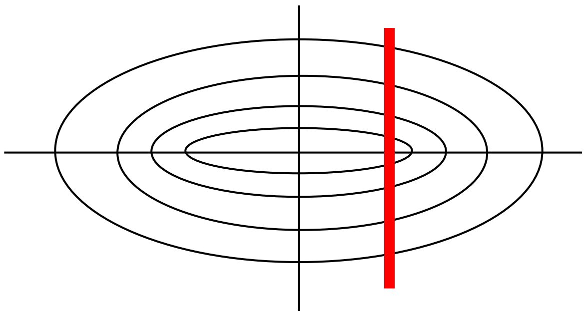
Gibbs sampling



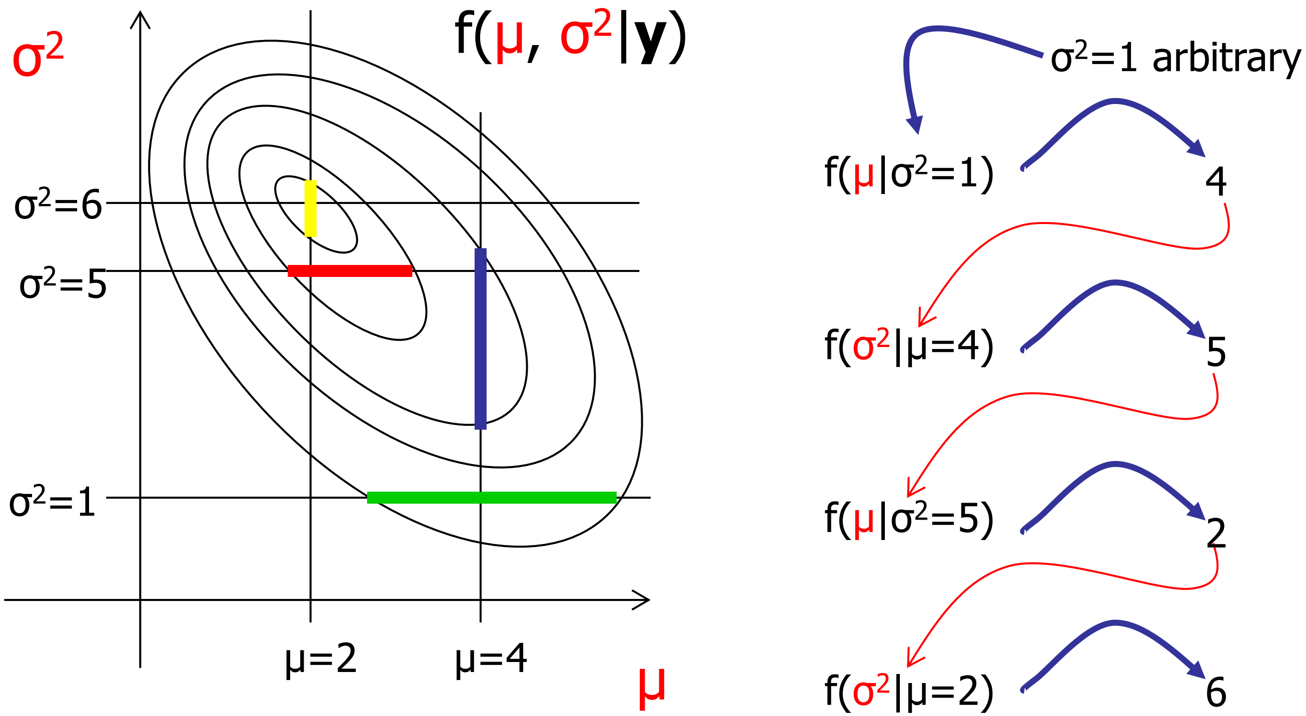


Gibbs sampling

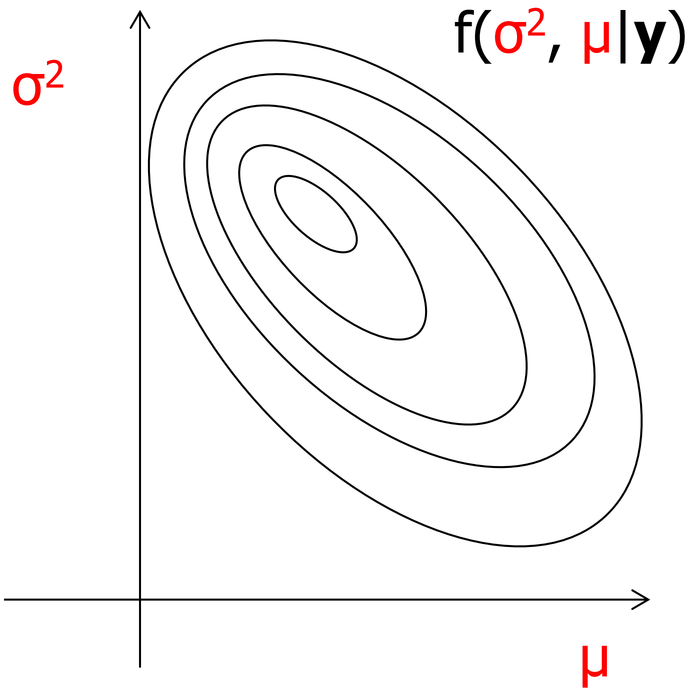




Gibbs sampling



Gibbs sampling



OBJECTIVE:
Random samples from $f(\sigma^2, \mu | \mathbf{y})$

σ^2	μ
3	12
1.5	18
2	11
…	…
…	…

Random samples of $f(\sigma^2 | \mathbf{y})$

Random samples of $f(\mu | \mathbf{y})$

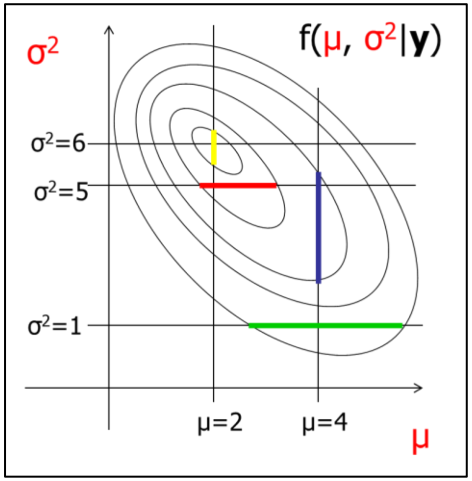
Gibbs sampling

f(μ|σ^2,**y**) : **[**4, 3, 3.3, 4.1, 4.8, 4.9, …**]**

f(μ|**y**): **[**4.8, 4.9,…**]**

f(σ^2|μ,**y**) : **[**2, 5, 3, 3, 2.6, 2.6, 2.8,... **]**

f(σ^2|**y**):**[**2.6, 2.8,...**]**



Inferences

Throw the burn-up & Sort the chain

$f(\mu|\mathbf{y})$ **:** [-5, -4.8, -3.9, -3.9, -1, 0.1, 1, 1.8, 3.3, 4.1, 4.9, …]

$P(\mu>0)$

Find the % of positive samples

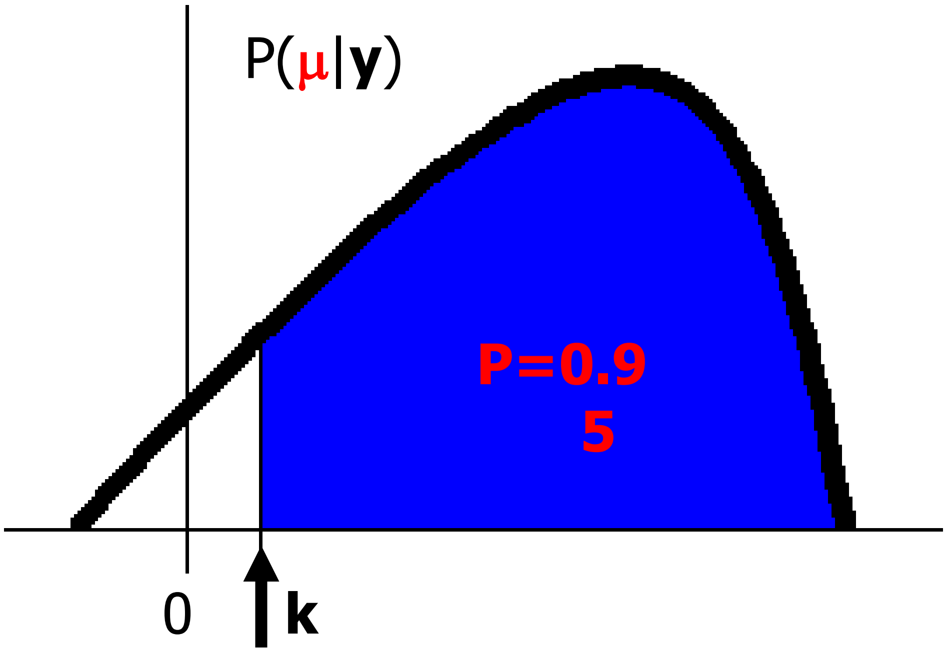
$f(\mu|\mathbf{y})$ **:** [-5, -4.8, -3.9, -3.9, -1, 0.1, 1, 1.8, 3.3, 4.1, 4.9, …]

$P(1\leq \mu \leq 4)$

Find the % of samples between 1 and 4

$f(\mu|\mathbf{y})$ **:** [-5, -4.8, -3.9, -3.9, -1, 0.1, 1, 1.8, 3.3, 4.1, 4.9, …]

Credibility intervals



Inferences

P[k,+∞)=95%

Find the 5 % of the first samples

f(μ|**y**) **:** [-5, -4.8, -3.9, -3.9, -1, 0.1, 1, 1.8| 3.3, 4.1, 4.9, ...]

k=3.3

HPD(95%)

Try intervals with 95 % of samples.
Choose the shortest one

f(μ|**y**) **:** [-5, -4.8, |-3.9, -3.9, -1, 0.1, 1, 1.8, 3.3|, 4.1, 4.9, ...]

Inferences

P[k,+∞)=95%

Find the 5 % of the first samples

f(μ|**y**) **:** [-5, -4.8, -3.9, -3.9, -1, 0.1, 1, 1.8 | 3.3, 4.1, 4.9, ...]

k=3.3

HPD(95%)

Try intervals with 95 % of samples.
Choose the shortest one

f(μ|**y**) **:** [-5, -4.8, -3.9, -3.9, | -1, 0.1, 1, 1.8, 3.3, 4.1, 4.9, | ...]

Inferences

Probability of relevance

Define the minimum relevant quantity R
Find the % of samples higher than R

NO NEED OF INTEGRALS !!

f(μ|**y**) **:** [-5, -4.8, -3.9, -3.9, -1, 0.1, [illegible], 1.8, 3.3, 4.1, 4.9, ...]

P(μ≥R) = 0.89

Relevant quantity R= 0.5

Gibbs sampling
is not a method of estimation
it is a numerical procedure
to obtain
MARGINAL POSTERIOR DISTRIBUTIONS
used for Bayesian estimation

Gibbs sampling

- We transform a multivariate problem in several univariate problems
- We do not accept the first samples because they are not taken at random (burn-in).
- All samples are correlated. We have a Monte Carlo error

Gibbs sampling

- We always accept the samples, but we should know how to sample
- There are algorithms to sample from known functions

FOR EXAMPLE: How to sample from a N(0,1) distribution:

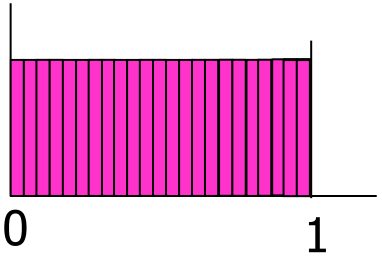
1) Take two random samples $x \in [0,1]$ from a random number generator

2) Calculate $y = \sqrt{-2 \log x_1} \cdot \cos(2\pi x_2)$

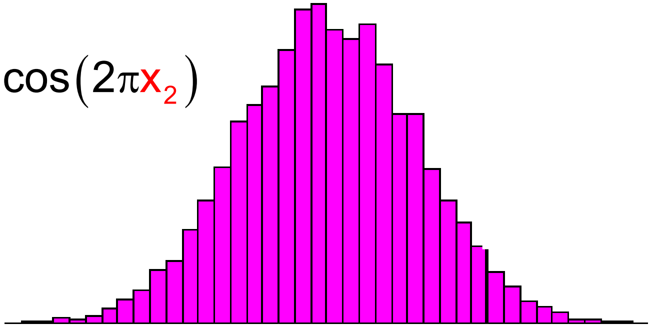
y is then a random sample from a N(0,1)

Gibbs sampling

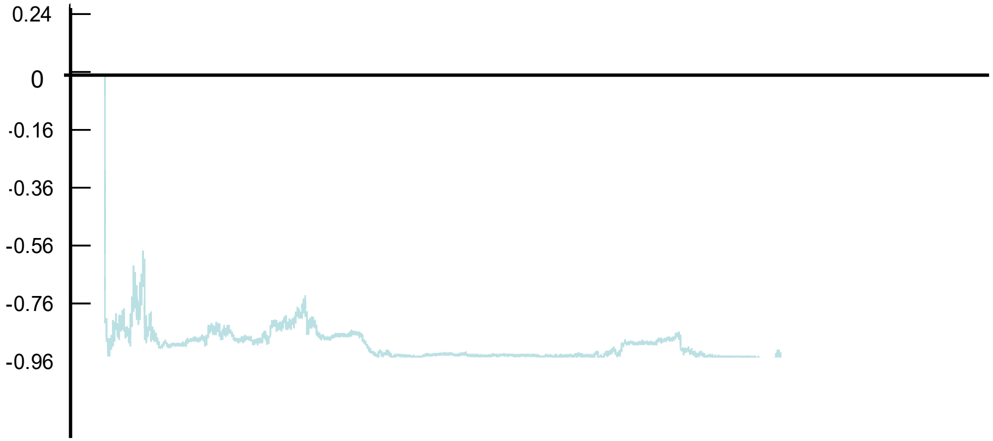
$x \in [0,1]$



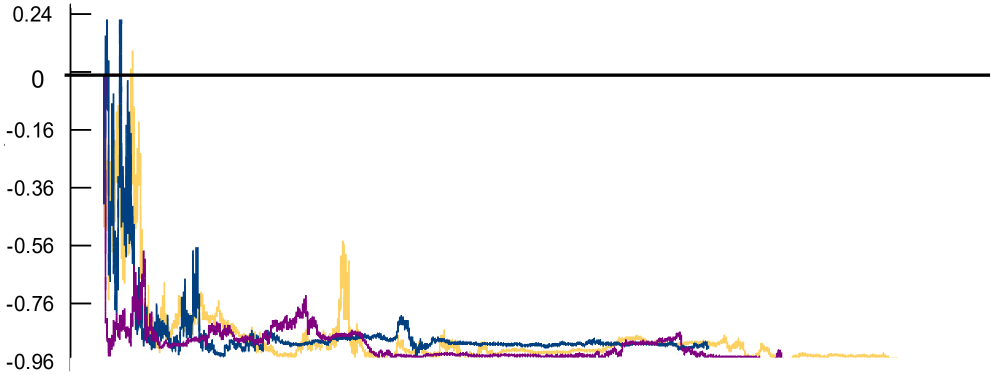
$$y = \sqrt{-2\log x_1} \cdot \cos(2\pi x_2)$$



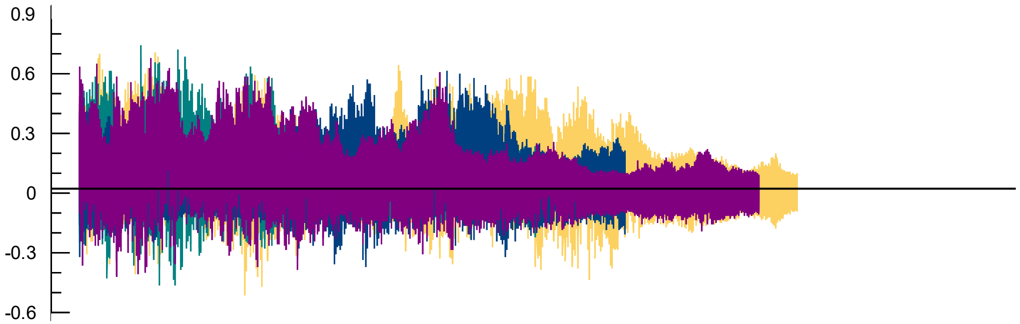
Convergence



Convergence



Gibbs sampling



Convergence tests

- GELMAN & RUBIN
  - The variance between chains is not higher than the variance within chains
- GEWEKE
  - The first half of the chain has the same average as the second part of the chain
- JOHNSON
  - Chains with the same random seed and different initial values should converge

…they never assure convergence !!

Burn-in

- VISUAL INSPECTION
  - Usually it works, but it can give surprises in complicated models
- RAFTERY & LEWIS METHOD
  - It requires your chain to have some properties you do not know whether it has. It is commonly cited, mainly because it gives very low values of burn-in.
- JOHNSON COUPLED CHAINS
  - Chains with the same random seed and different initial values should converge. You decide when.

…they never assure convergence !!

Monte Carlo s.e.

- We estimate the posterior distribution, we have an error of estimation called Monte Carlo s.e.
- We can make it as small as we want, taking more samples
- Samples can be highly correlated. Efective sample size is the size of a uncorrelated sample giving the same Monte Carlo s.e.
- Usually it is not a worth to take two consecutive samples, e.g., we take one each 20 or one each 50. This is called the sampling lag
- We shall calculate autocorrelation between two consecutive samples

Other sampling methods

WE DO NOT KNOW HOW TO SAMPLE

– **ACCEPTANCE-REJECTION SAMPLING**

– **METROPOLIS-HASTINGS**

– **IMPORTANCE SAMPLING**

… etc

End of the Interlude