

2018, Montpellier

# Bayesian statistics: a short introduction

David Makowski

[david.makowski@inra.fr](mailto:david.makowski@inra.fr)

# Some references

- Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. *Philos. Trans. Roy. Soc. London*, 53, 370-418.
- Carlin, B.P. and Louis, T.A. 2000. *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall, London.
- Chib, S., Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. *American Statistician* 49, 327-335.
- Gelman A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin. 2013. *Bayesian data analysis*. Third edition. Chapman & Hall, London.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J. 1995. *Markov Chain Monte Carlo in practice*. Chapman & Hall, London.
- Laplace...
- Low Choy, S., O'Leary, R., and Mengerson, K. 2009. Elicitation by Design for Ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology* 90: 265-277
- Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325-337
- Robert, C. P., Casella, G. 2004. *Monte Carlo Statistical Methods*. Springer, New York.
- Smith, A. F. M., Gelfand, A. E. 1992. Bayesian statistics without tears: A sampling-resampling perspective. *American Statistician* 46, 84-88.

## Some references

Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. *Philos. Trans. Roy. Soc. London*, 53, 370-418.

Carlin, B.P. and Louis, T.A. 2000. *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall, London.

Chib, S., Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. *American Statistician* 49, 327-335.

Gelman A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin. 2013. *Bayesian data analysis*. Third edition. Chapman & Hall, London.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J. 1995. *Markov Chain Monte Carlo in practice*. Chapman & Hall, London.

Low Choy, S., O'Leary, R., and Mengersen, K. 2009. Elicitation by Design for Ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology* 90: 265-277

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325-337

Robert, C. P., Casella, G. 2004. *Monte Carlo Statistical Methods*. Springer, New York.

Smith, A. F. M., Gelfand, A. E. 1992. Bayesian statistics without tears: A sampling-resampling perspective. *American Statistician* 46, 84-88.

*PHILOSOPHICAL*  
TRANSACTIONS:

---

**An Essay towards Solving a  
Problem in the Doctrine of  
Chances. By the Late Rev. Mr.  
Bayes, F. R. S. Communicated by  
Mr. Price, in a Letter to John  
Canton, A. M. F. R. S.**

Mr. Bayes and Mr. Price

*Phil. Trans.* 1763 **53**, 370-418, published 1 January  
1763

---

quodque solum, certa nitri signa præbere, sed plura concurrere debere, ut de vero nitro producto dubium non relinquatur.

---

LII. *An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763. I Now send you an essay which I have found among the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved. Experimental philosophy, you will find, is nearly interested in the subject of it; and on this account there seems to be particular reason for thinking that a communication of it to the Royal Society cannot be improper.

[ 376 ]

P R O B L E M.

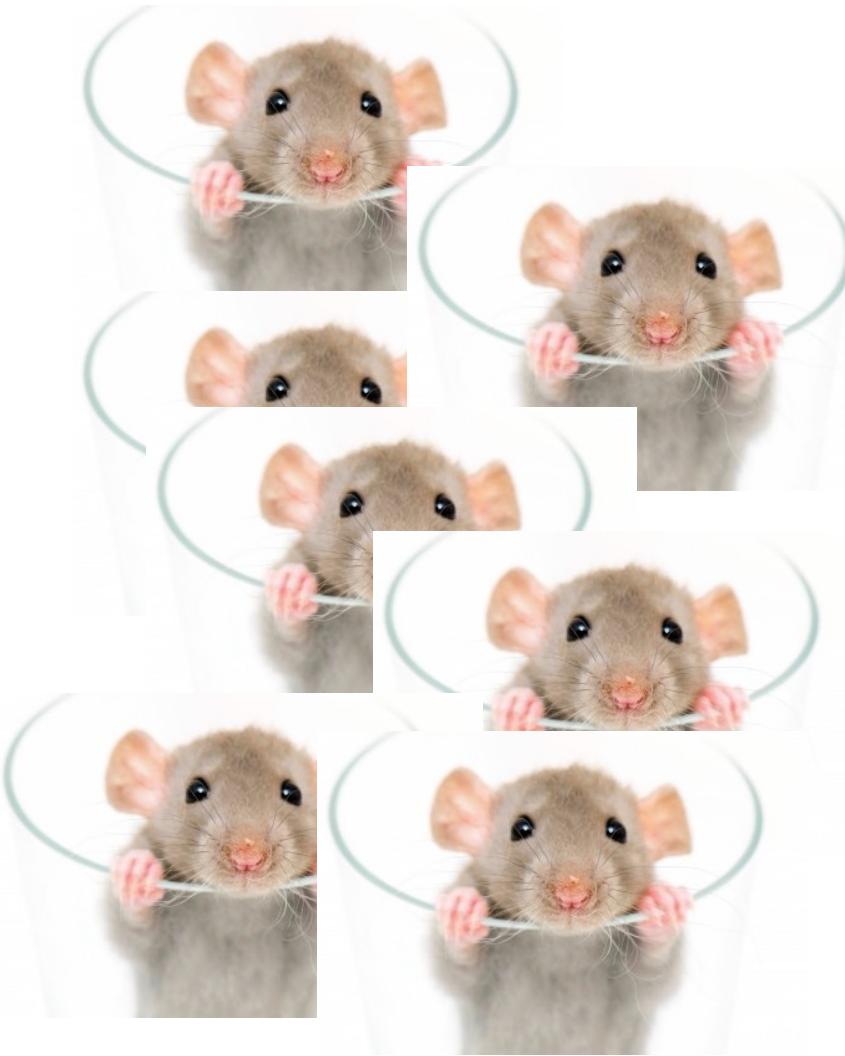
*Given* the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.

« Given that  $n$  independent Bernoulli trials with unknown constant success probability  $p$  have brought  $k$  successes, what is the probability that the parameter  $p$  lies between two given bounds? »

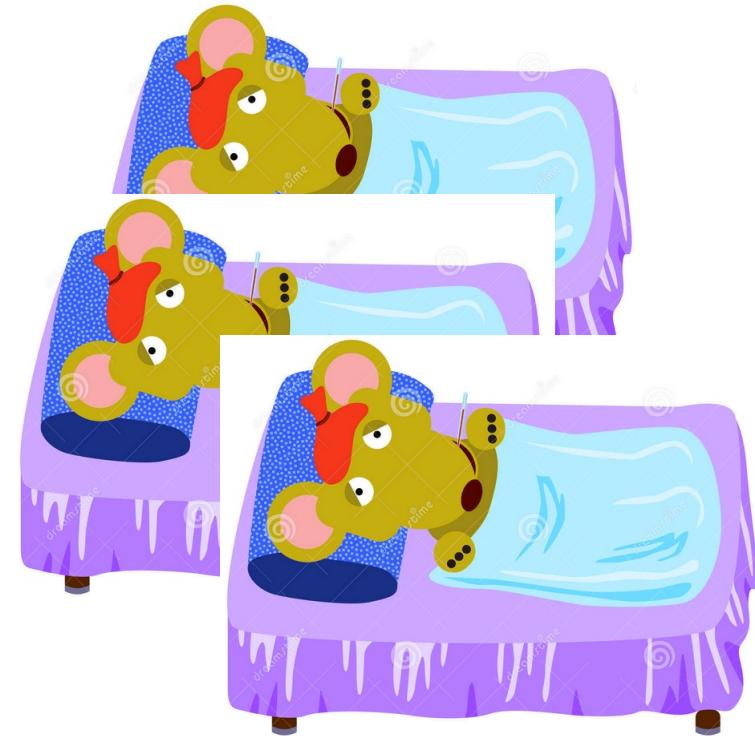
$n$  rats



$n$  rats

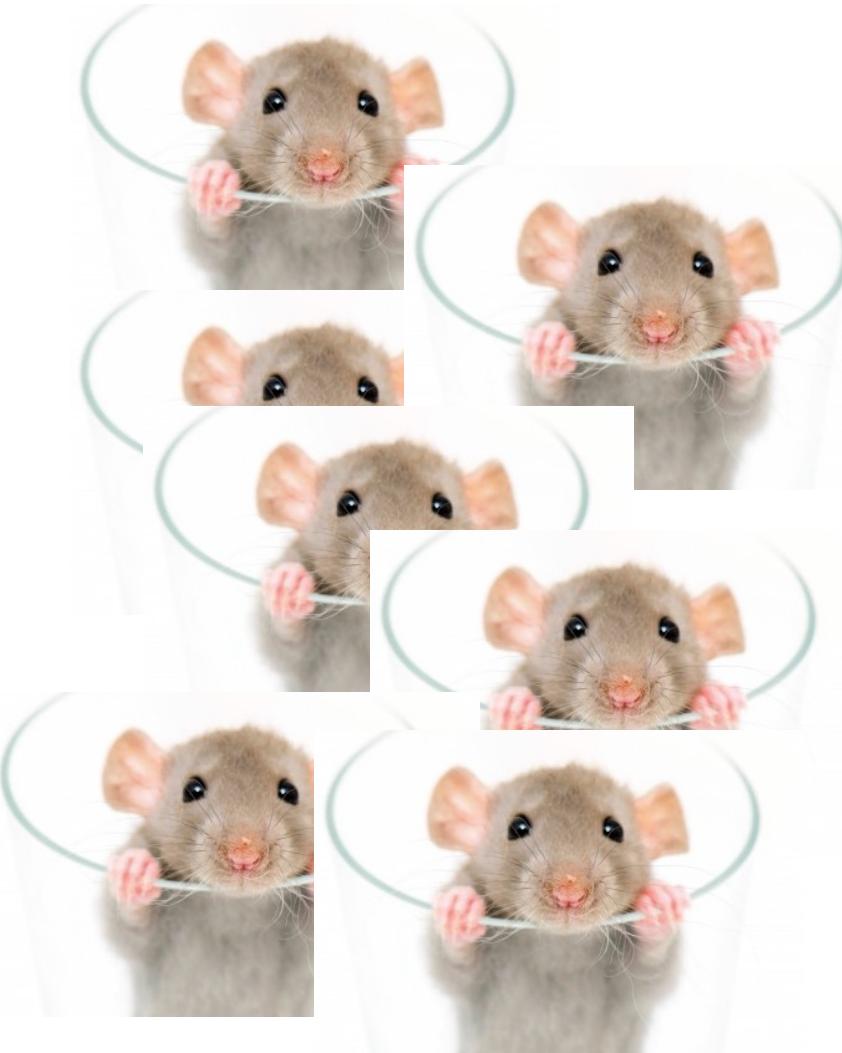


$k$  contaminated rats

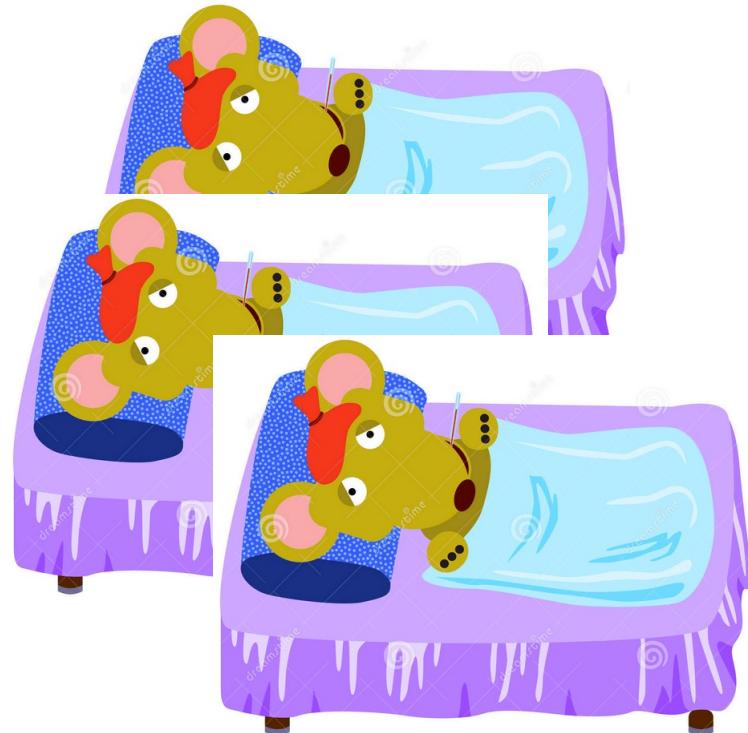


# Probability $p$ of contamination?

$n$  rats



$k$  contaminated rats



# Some references

Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. Philos. Trans. Roy. Soc. London, 53, 370-418.

Carlin, B.P. and Louis, T.A. 2000. *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall, London.

**Chib, S., Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. American Statistician 49, 327-335.**

Gelman A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin. 2013. *Bayesian data analysis*. Third edition. Chapman & Hall, London.

**Gilks, W.R., Richardson, S., Spiegelhalter, D.J. 1995. *Markov Chain Monte Carlo in practice*. Chapman & Hall, London.**

Low Choy, S., O'Leary, R., and Mengersen, K. 2009. Elicitation by Design for Ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology* 90: 265-277

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325-337

Robert, C. P., Casella, G. 2004. *Monte Carlo Statistical Methods*. Springer, New York.

**Smith, A. F. M., Gelfand, A. E. 1992. Bayesian statistics without tears: A sampling-resampling perspective. American Statistician 46, 84-88.**

# Some references

Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. Philos. Trans. Roy. Soc. London, 53, 370-418

Carlin, B.P. and Louis, T.A. 2000. *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall, London.

Chib, S., Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. American Statistician 49, 327-335.

Gelman A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin. 2013. *Bayesian data analysis*. Third edition. Chapman & Hall, London.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J. 1995. *Markov Chain Monte Carlo in practice*. Chapman & Hall, London.

Low Choy, S., O'Leary, R., and Mengersen, K. 2009. Elicitation by Design for Ecology: using expert opinion to inform priors for Bayesian statistical models. Ecology 90: 265-277

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10:325-337

Robert, C. P., Casella, G. 2004. *Monte Carlo Statistical Methods*. Springer, New York.

Smith, A. F. M., Gelfand, A. E. 1992. Bayesian statistics without tears: A sampling-resampling perspective. American Statistician 46, 84-88.

# Some references

- Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. *Philos. Trans. Roy. Soc. London*, 53, 370-418.
- Carlin, B.P. and Louis, T.A. 2000. *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall, London.
- Chib, S., Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. *American Statistician* 49, 327-335.
- Gelman A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin. 2013. *Bayesian data analysis*. Third edition. Chapman & Hall, London.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J. 1995. *Markov Chain Monte Carlo in practice*. Chapman & Hall, London.

**Low Choy, S., O'Leary, R., and Mengersen, K. 2009. Elicitation by Design for Ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology* 90: 265-277**

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325-337

Robert, C. P., Casella, G. 2004. *Monte Carlo Statistical Methods*. Springer, New York.

Smith, A. F. M., Gelfand, A. E. 1992. Bayesian statistics without tears: A sampling-resampling perspective. *American Statistician* 46, 84-88.

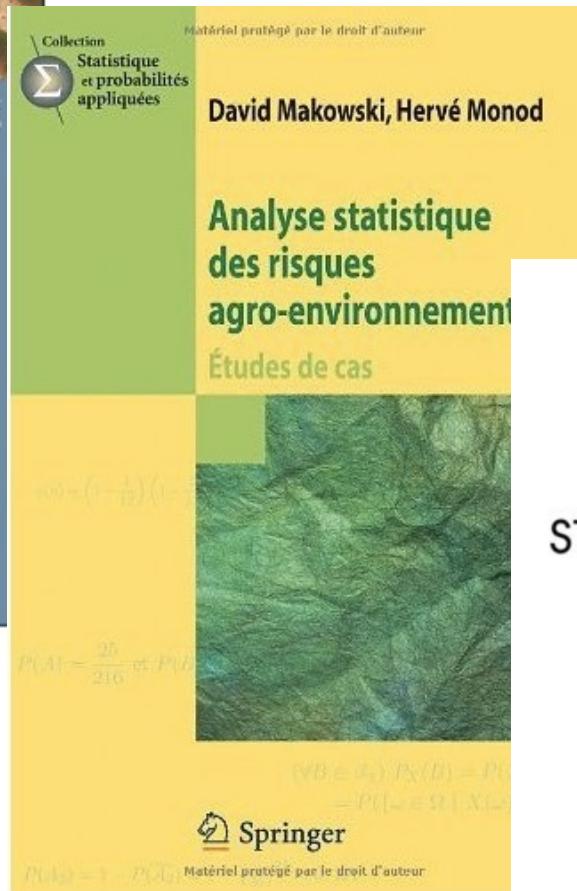
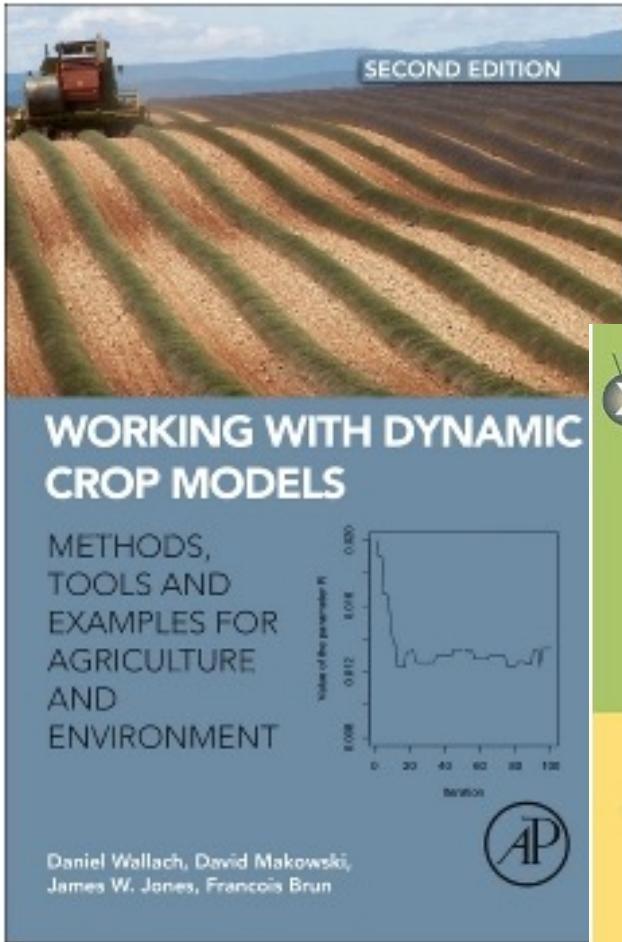
# Some references

- Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. *Philos. Trans. Roy. Soc. London*, 53, 370-418.
- Carlin, B.P. and Louis, T.A. 2000. *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall, London.
- Chib, S., Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. *American Statistician* 49, 327-335.
- Gelman A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin. 2013. *Bayesian data analysis*. Third edition. Chapman & Hall, London.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J. 1995. *Markov Chain Monte Carlo in practice*. Chapman & Hall, London.
- Low Choy, S., O'Leary, R., and Mengerson, K. 2009. Elicitation by Design for Ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology* 90: 265-277

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325-337

Robert, C. P., Casella, G. 2004. *Monte Carlo Statistical Methods*. Springer, New York.

Smith, A. F. M., Gelfand, A. E. 1992. Bayesian statistics without tears: A sampling-resampling perspective. *American Statistician* 46, 84-88.



Collectif BIOBAYES

## Initiation à la statistique bayésienne

Bases théoriques et applications en alimentation, environnement, épidémiologie et génétique

Préface de Jean-Michel Marin



ellipses

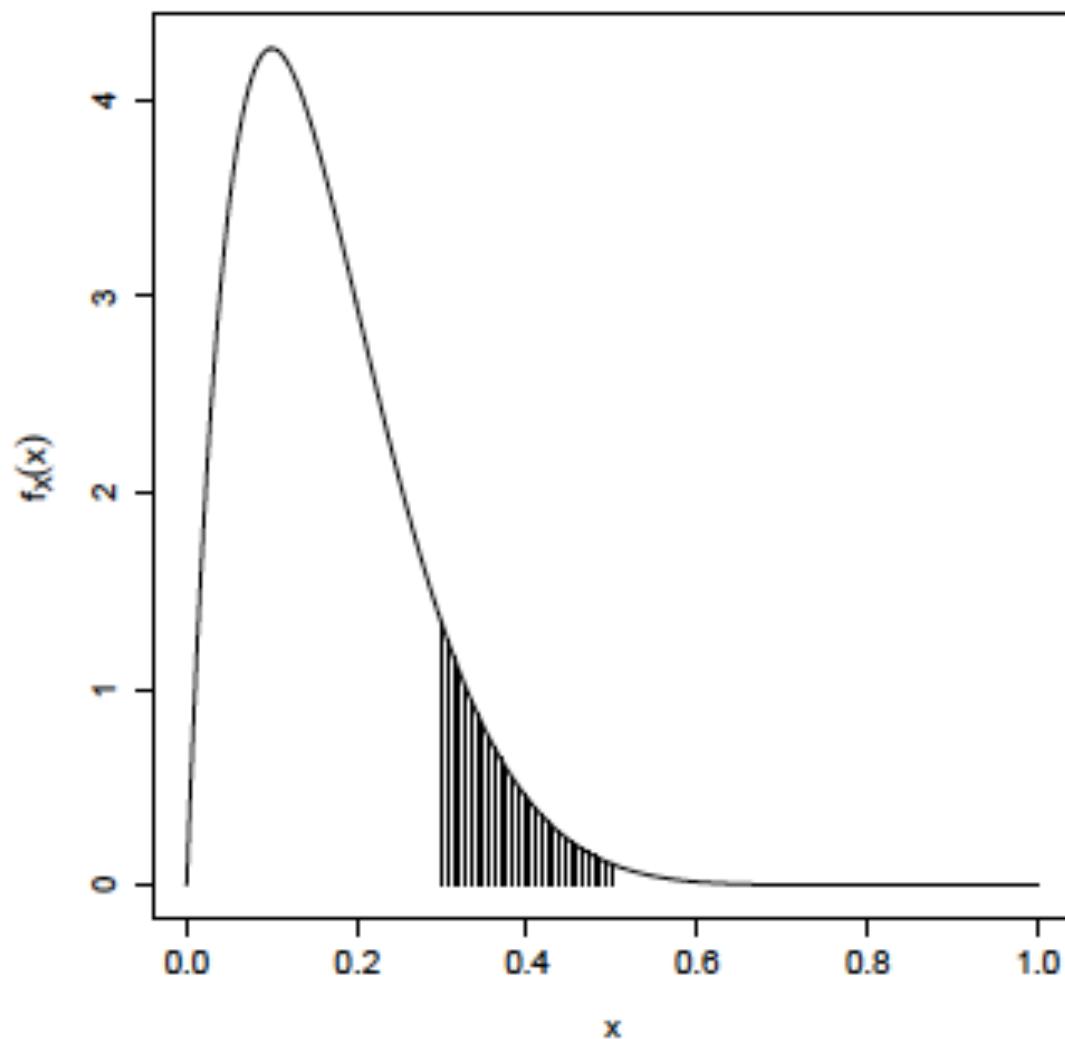
# Estimation of parameters ( $\theta$ )

**Parameter = Unknown numerical value**

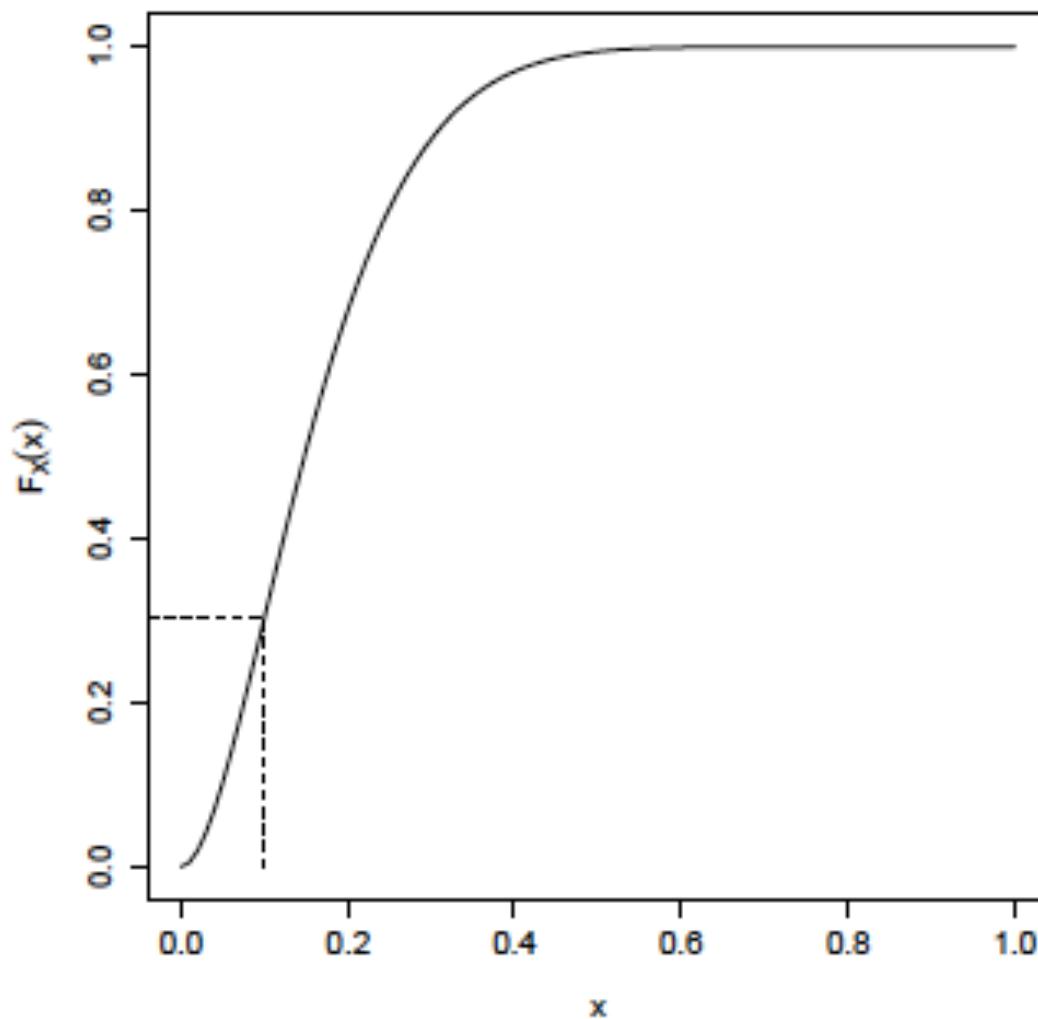
**Information available to estimate parameters**

- A dataset ( $Y$ ).
- Prior knowledge about parameter values.

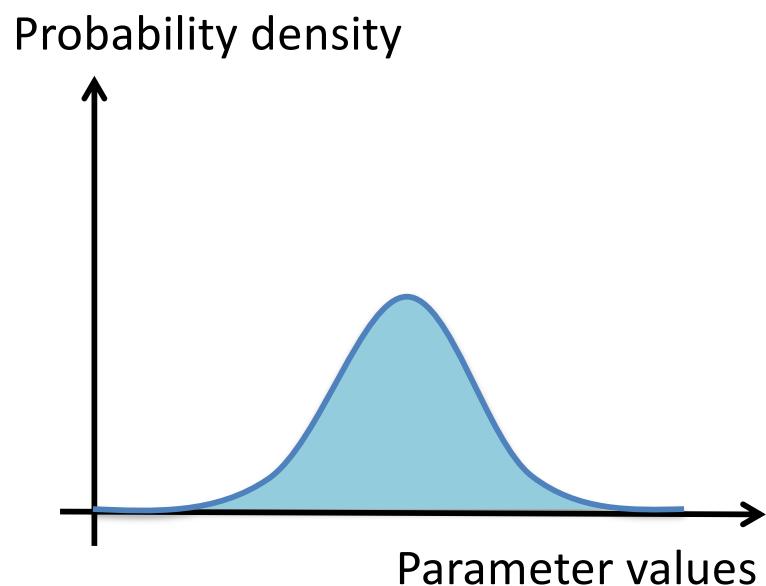
# Probability density



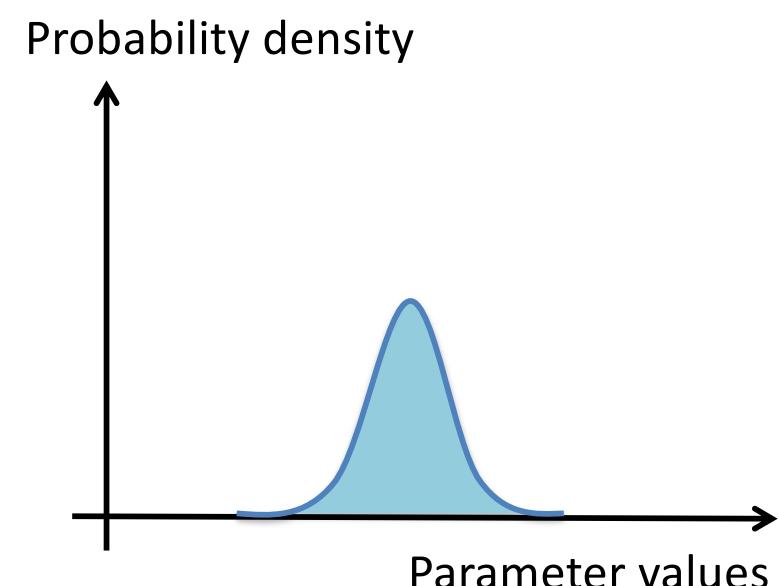
## Cumulative probability distribution



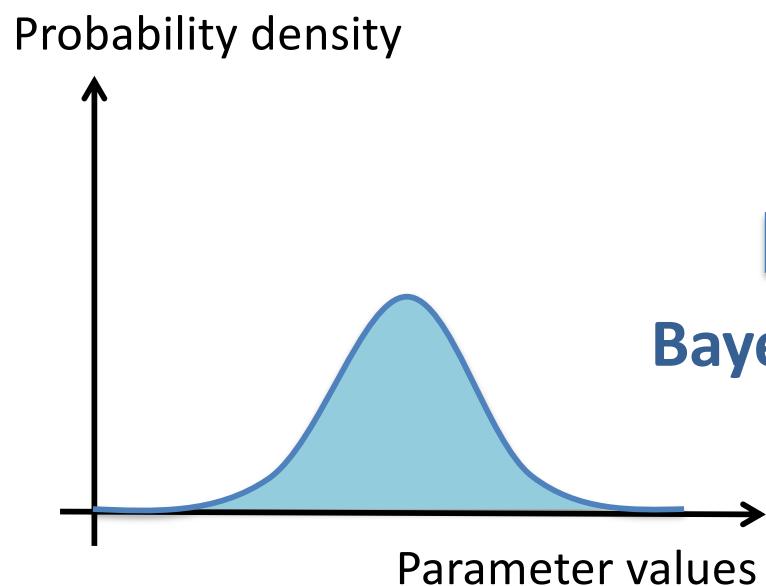
## PRIOR probability distribution



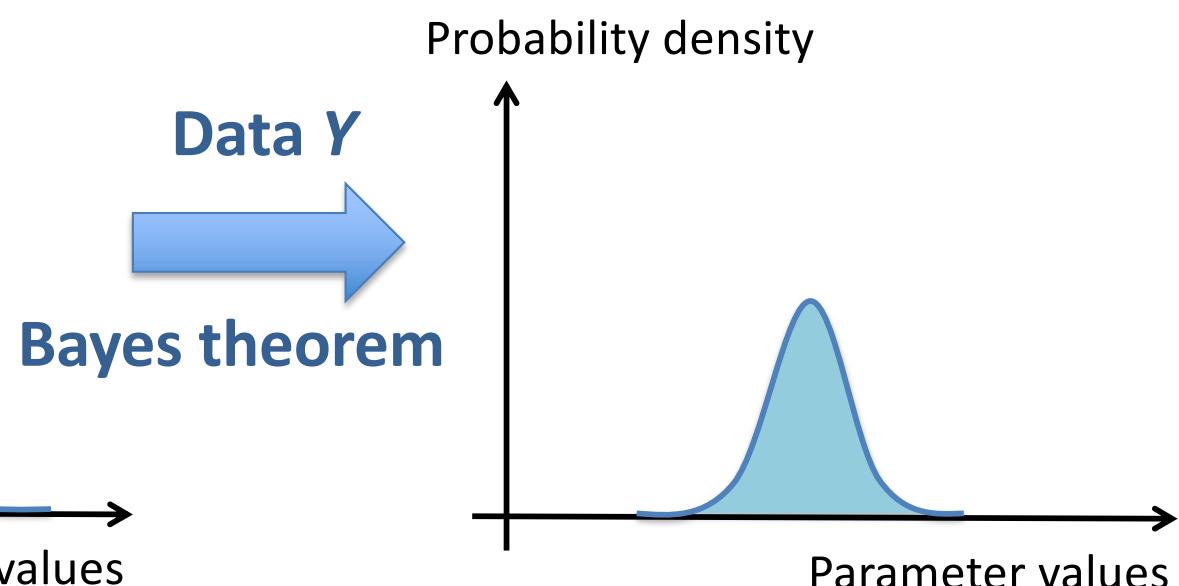
## POSTERIOR probability distribution



## PRIOR probability distribution



## POSTERIOR probability distribution



Data  $Y$   
Bayes theorem

	<b>Classic methods</b>	<b>Bayesian methods</b>
<b>Parameters</b>	Fixed	Random

	<b>Classic methods</b>	<b>Bayesian methods</b>
Parameters	Fixed	Random
Prior knowledge about parameter values	Not taken into account	Taken into account

	<b>Classic methods</b>	<b>Bayesian methods</b>
Parameters	Fixed	Random
Prior knowledge about parameter values	Not taken into account	Taken into account
Likelihood	Used	Used

	<b>Classic methods</b>	<b>Bayesian methods</b>
Parameters	Fixed	Random
Prior knowledge about parameter values	Not taken into account	Taken into account
Likelihood	Used	Used
Computation	Easier	More difficult

# Estimation of parameters ( $\theta$ )

**Parameter = Unknown numerical value**

**Information available to estimate parameters**

- A set of observations ( $Y$ ).
- Prior knowledge about parameter values.

## Two probability distributions

- **Prior parameter distribution** = probability distribution describing our initial knowledge about parameter values.

$$P(\theta)$$

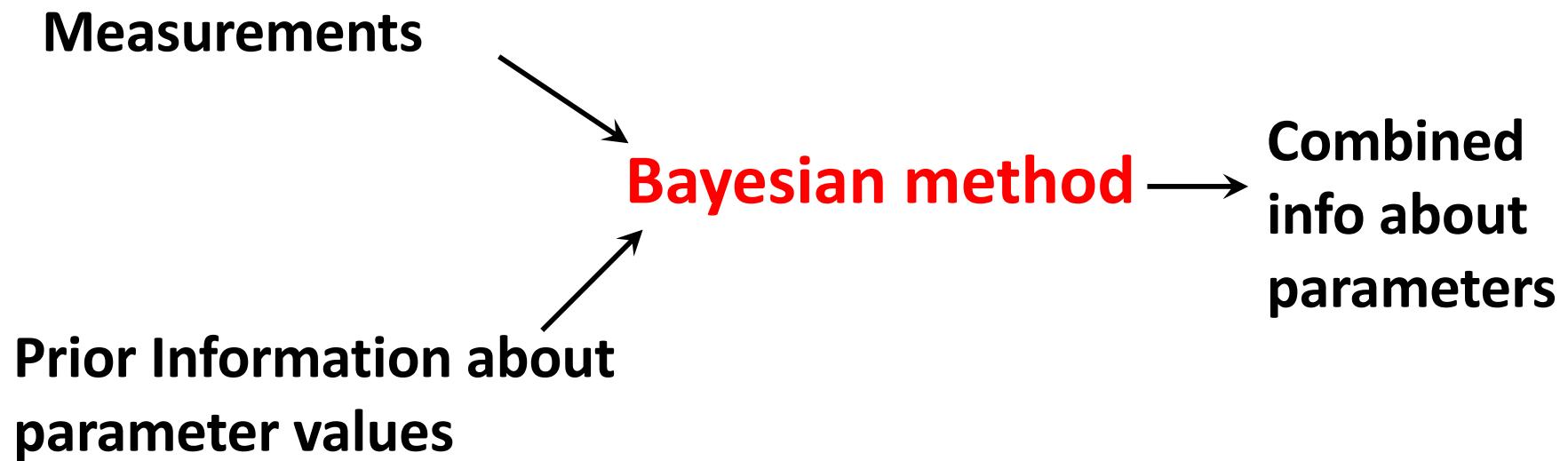
- **Likelihood function** = function relating data to parameters.

$$P(Y|\theta)$$

# The Bayes theorem

**Posterior  $\propto$  Likelihood x Prior**

$$P(\theta | Y) = \frac{P(Y | \theta)P(\theta)}{P(Y)}$$



## Example 1

**Probability of disease in function of a diagnosis test**

$$P(Diseased | Test+) = \frac{P(Test+ | Diseased) P(Diseased)}{P(Test+)}$$

## Example 1

Probability of disease in function of a diagnosis test

$$P(\text{Diseased} | \text{Test}+) = \frac{P(\text{Test+} | \text{Diseased}) P(\text{Diseased})}{P(\text{Test}+)}$$

Data    Prior

$$P(\theta | Y) = \frac{P(Y | \theta) P(\theta)}{P(Y)}$$

## Example 1

### Probability of disease in function of a diagnosis test

$$P(Diseased | Test+) = \frac{P(Test+ | Diseased)P(Diseased)}{P(Test+ | Diseased)P(Diseased) + P(Test+ | Not Diseased)P(Not Diseased)}$$

## Example 1

### Probability of disease in function of a diagnosis test

$$P(Diseased | Test+) = \frac{P(Test+ | Diseased)P(Diseased)}{P(Test+ | Diseased)P(Diseased) + P(Test+ | Not Diseased)P(Not Diseased)}$$

$$P(Diseased | Test+) = \frac{\text{Sensitivity} \times \text{Prevalence}}{\text{Sensitivity} \times \text{Prevalence} + (1-\text{Specificity})(1-\text{Prevalence})}$$

## Example 1

### Probability of disease in function of a diagnosis test

$$P(Diseased | Test+) = \frac{P(Test+ | Diseased)P(Diseased)}{P(Test+ | Diseased)P(Diseased) + P(Test+ | Not Diseased)P(Not Diseased)}$$

$$P(Diseased | Test+) = \frac{\text{Sensitivity} \times \text{Prevalence}}{\text{Sensitivity} \times \text{Prevalence} + (1-\text{Specificity})(1-\text{Prevalence})}$$

**Perfect test →**  $P(Diseased | Test+) = \frac{1 \times \text{Prevalence}}{1 \times \text{Prevalence} + (1-1)(1-\text{Prevalence})} = 1$

## Example 1

### Probability of disease in function of a diagnosis test

$$P(Diseased | Test+) = \frac{P(Test+ | Diseased)P(Diseased)}{P(Test+ | Diseased)P(Diseased) + P(Test+ | Not Diseased)P(Not Diseased)}$$

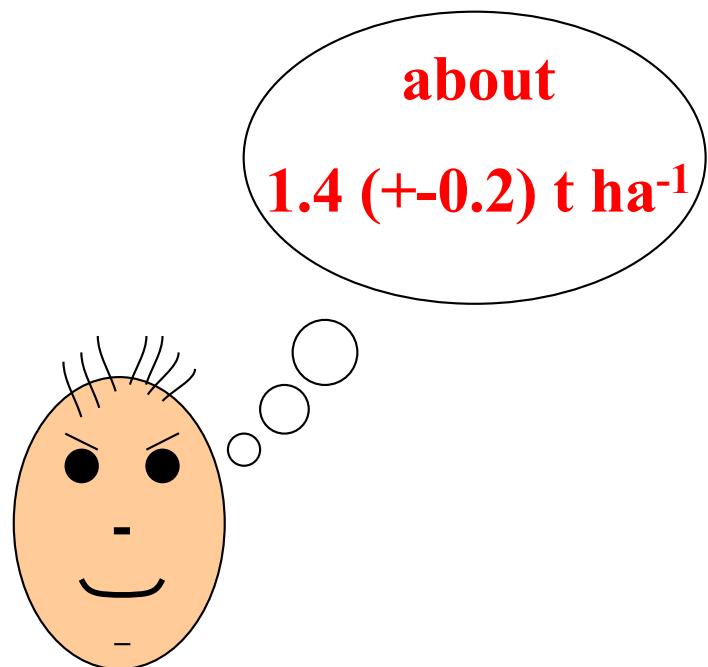
$$P(Diseased | Test+) = \frac{\text{Sensitivity} \times \text{Prevalence}}{\text{Sensitivity} \times \text{Prevalence} + (1-\text{Specificity})(1-\text{Prevalence})}$$

**Useless test →**

$$P(Diseased | Test+) = \frac{0.5 \times \text{Prevalence}}{0.5 \times \text{Prevalence} + (1-0.5)(1-\text{Prevalence})}$$
$$= \text{Prevalence}$$

## Example 2

Estimation of yield  $\theta$  in a crop field by combining a measurement with expert knowledge



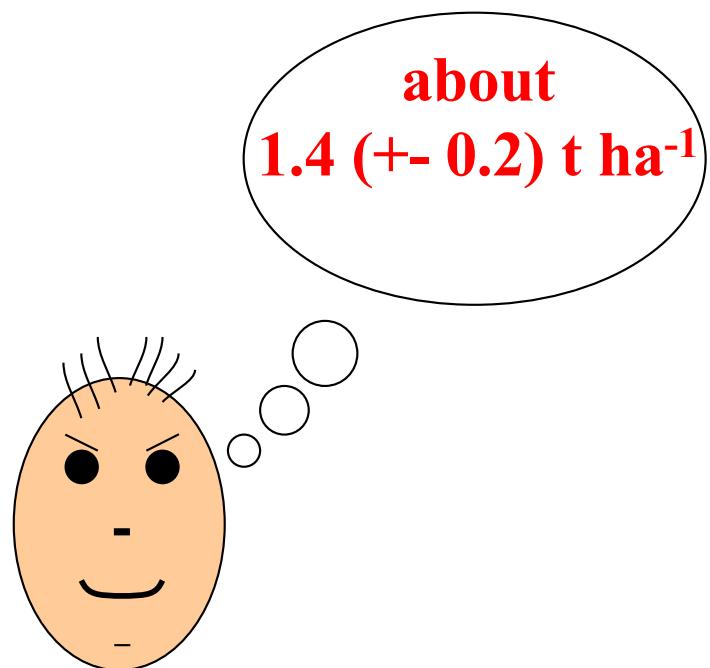
*Expert*

## Example 2

Estimation of yield  $\theta$  in a crop field by combining a measurement with expert knowledge



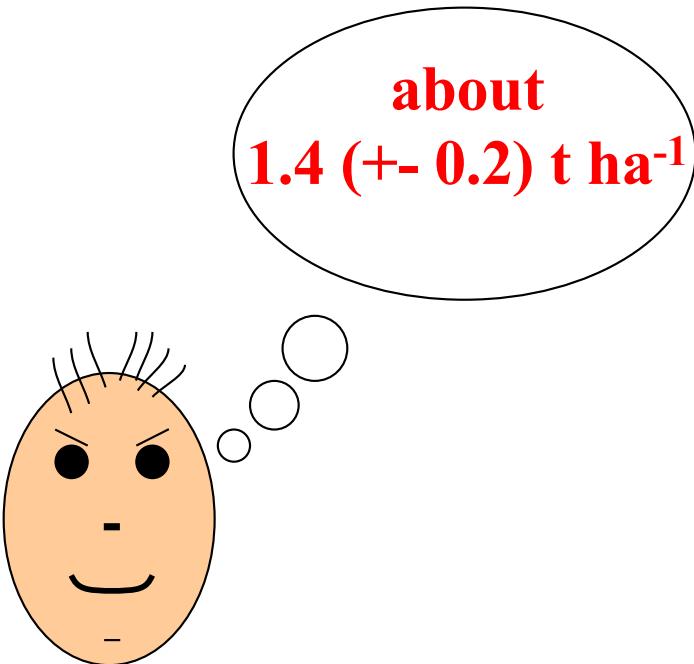
$$Y = 1.9 (+-0.2) \text{ t ha}^{-1}$$



*Expert*

## Prior distribution

- It describes our belief about the parameter values **before** we observe the measurements.
- It is based on past studies, expert knowledge, and litterature.



## Example 2 (continued)

### Definition of a prior distribution

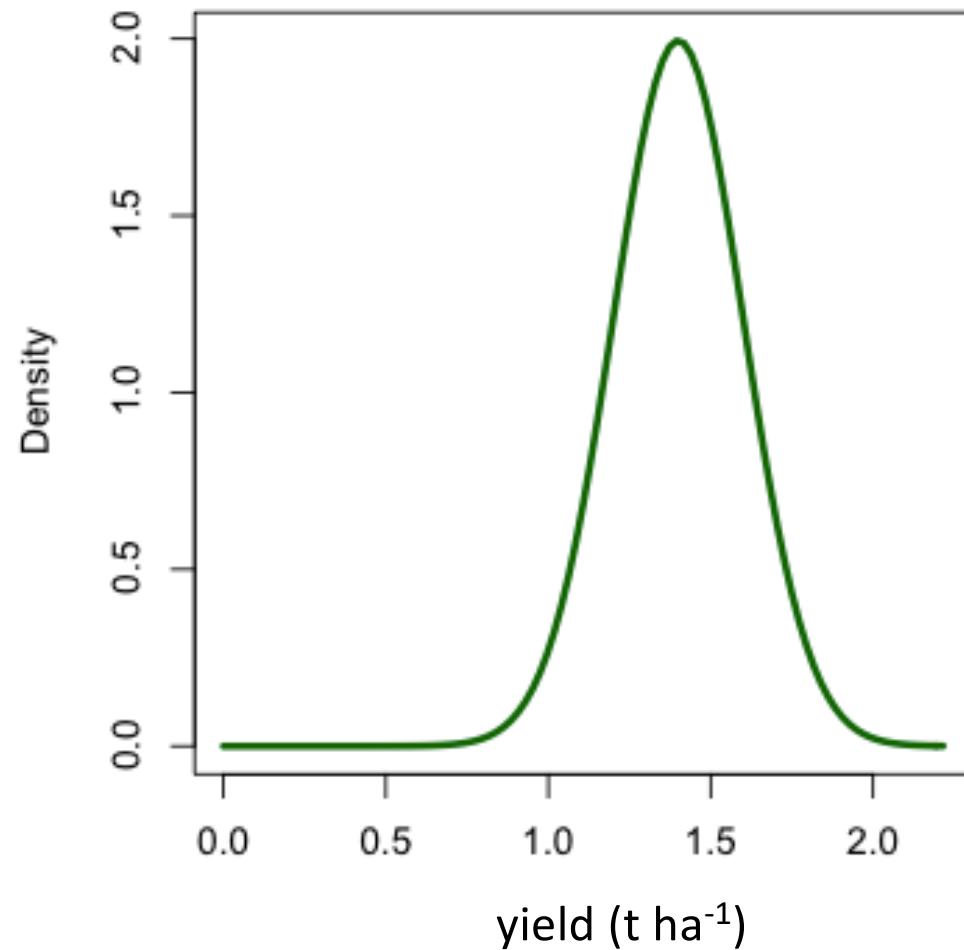
$$\theta \sim N(\mu, \tau^2)$$

$$P(\theta) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left[-\frac{(\theta - \mu)^2}{2\tau^2}\right] = \frac{1}{\sqrt{2\pi0.2^2}} \exp\left[-\frac{(\theta - 1.4)^2}{2 \times 0.2^2}\right]$$

- Normal probability distribution.
- Expected value equal to 1.4
- Standard error equal to 0.2

## Example 2 (continued)

### Plot of the prior distribution



## Likelihood function

- A **likelihood function** is a function relating data to parameters.
- It is equal to the probability that the **measurements** would have been observed **given some parameter values**.
- Notation:  $P(Y | \theta)$

## Example 2 (continued)

### Statistical model

$$\begin{aligned} Y &= \theta + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2) \end{aligned}$$

$$Y | \theta \sim N(\theta, \sigma^2)$$

## Example 2 (continued)

### Definition of a likelihood function

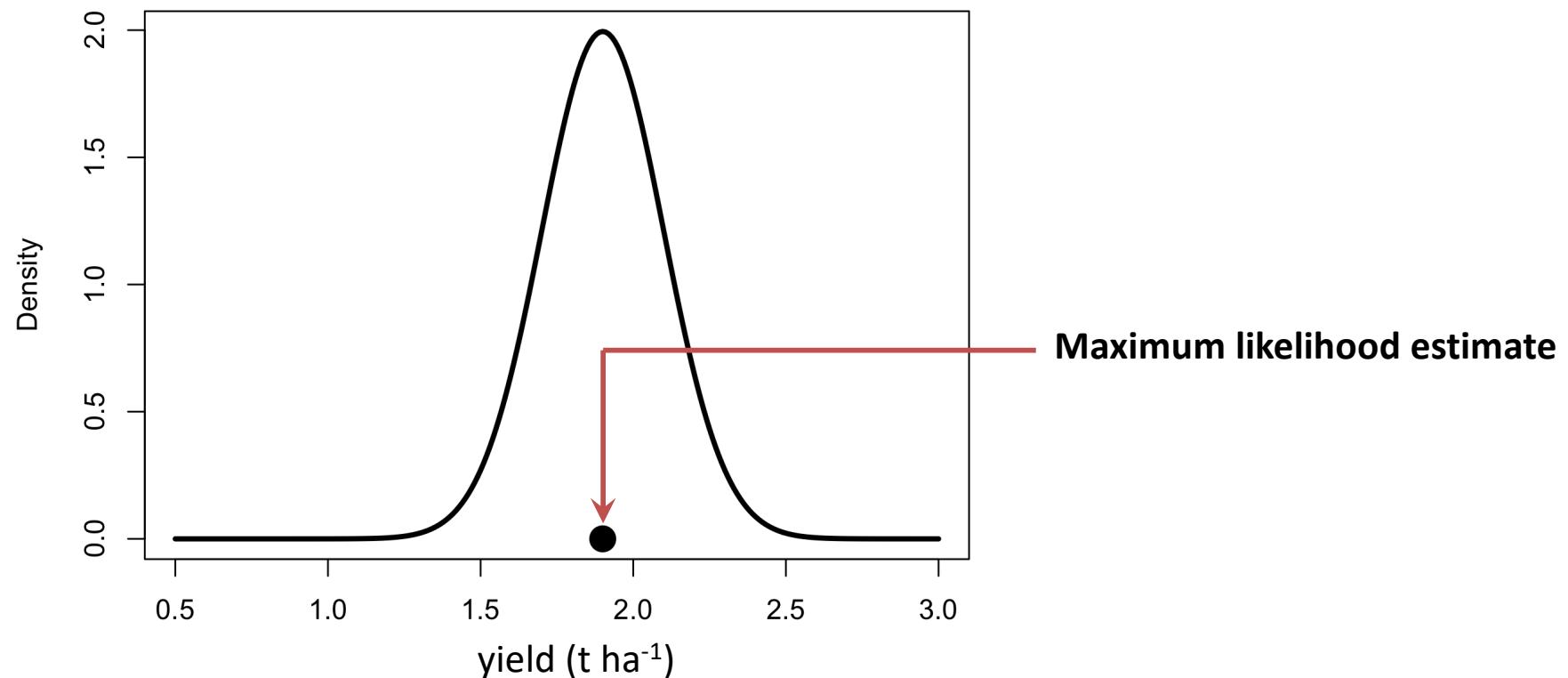
$$Y | \theta \sim N(\theta, \sigma^2)$$

$$P(Y|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(Y-\theta)^2}{2\sigma^2}\right] = \frac{1}{\sqrt{2\pi}0.2} \exp\left[-\frac{(1.9-\theta)^2}{2\times0.2^2}\right]$$

- Normal probability distribution.
- Measurement  $y$  assumed unbiased and equal to 1.9
- Standard error  $\sigma$  assumed equal to 0.2

## Example 2 (continued)

### Definition of a likelihood function

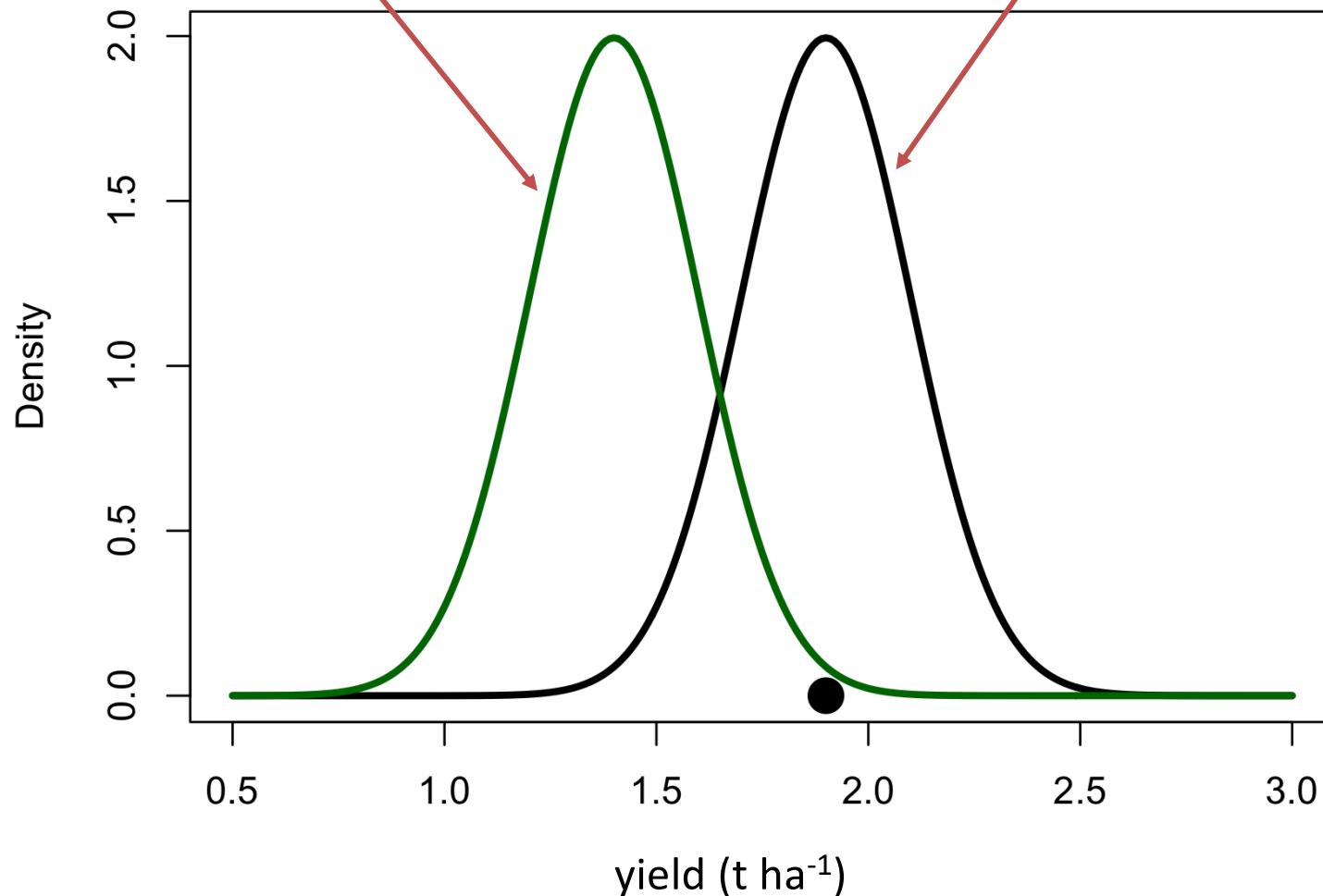


## Maximum likelihood

Likelihood functions are also used by frequentist to implement the *maximum likelihood method*.

The maximum likelihood estimator is the value of  $\theta$  maximizing  $P(Y | \theta)$ .

**Prior probability distribution**      **Likelihood function**



## Example 2 (continued)

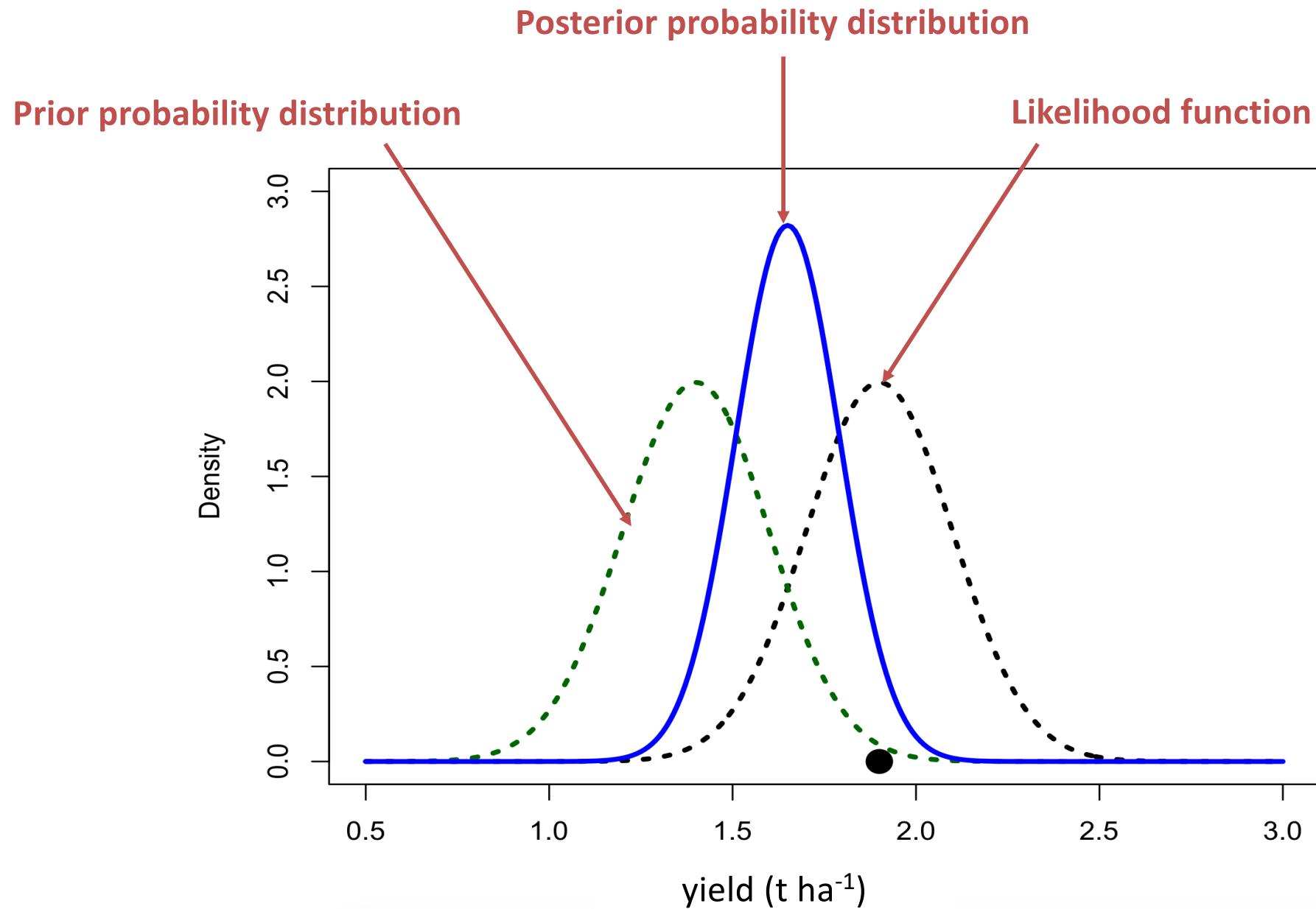
### Analytical expression of the posterior distribution

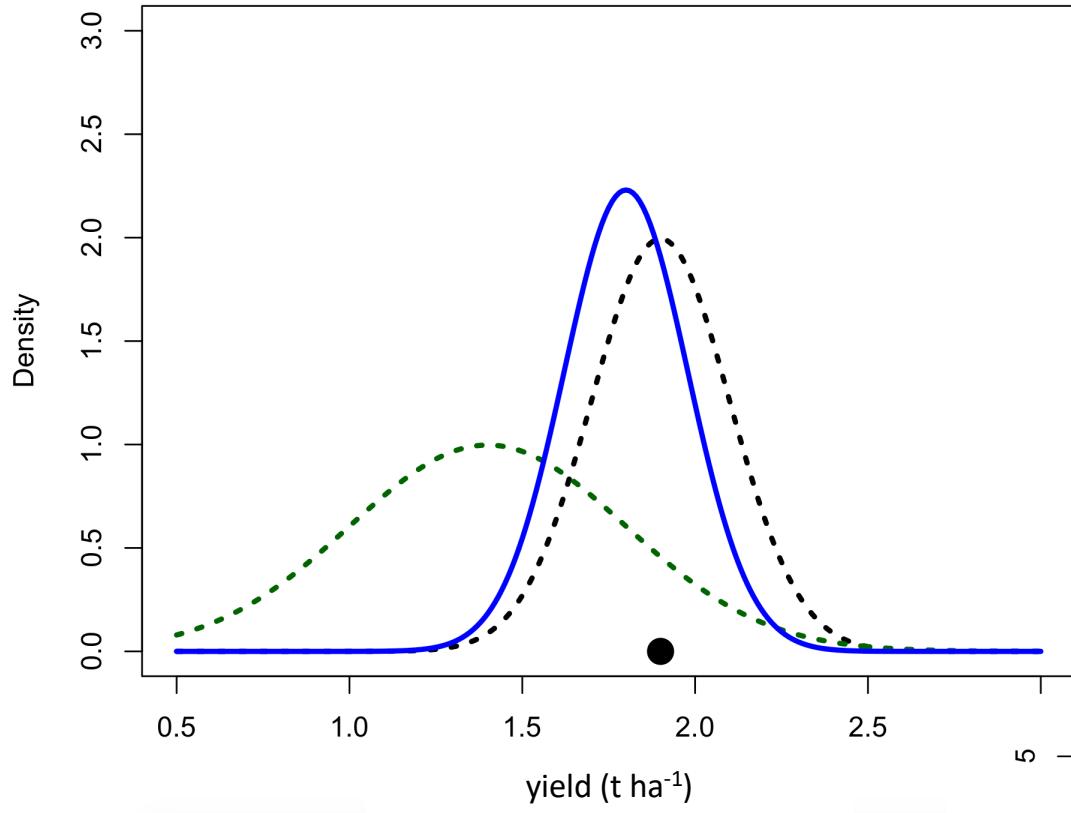
$$\theta | Y \sim N(\mu_{post}, \sigma_{post}^2)$$

$$\mu_{post} = (1 - B) \times \mu + B \times Y = 1.65$$

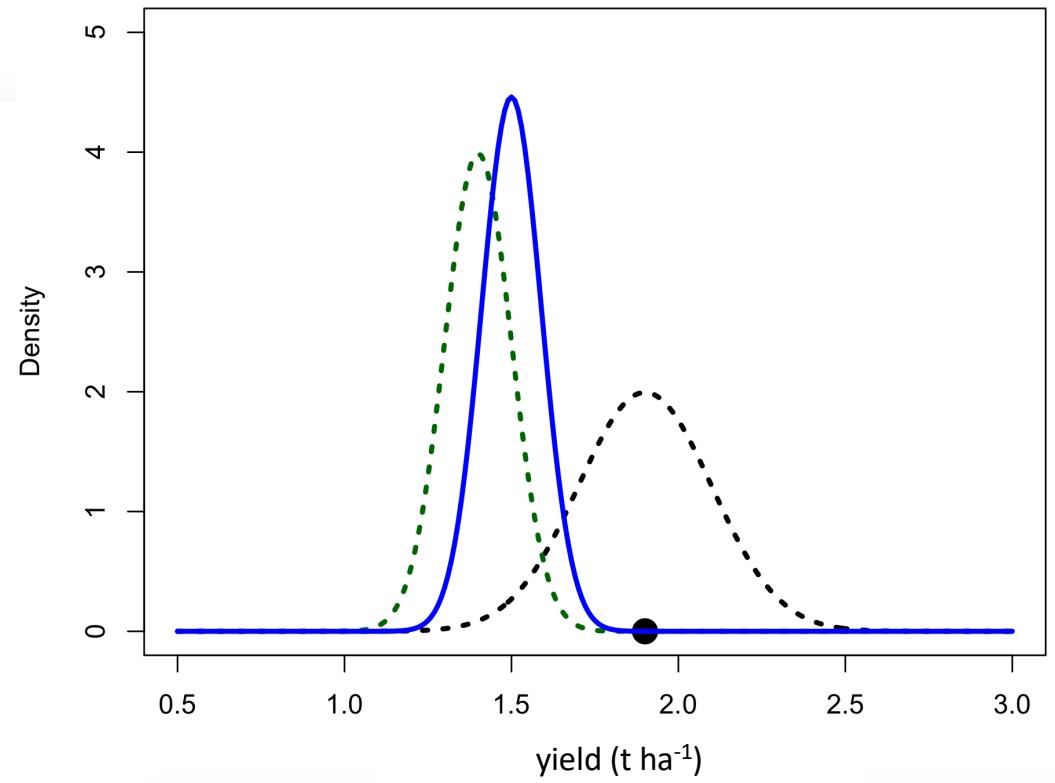
$$\sigma_{post}^2 = (1 - B) \times \tau^2 = 0.02$$

$$B = \frac{\tau^2}{\tau^2 + \sigma^2} = 0.5$$





Higher prior uncertainty



Lower prior uncertainty

## Example 2 (continued)

### Discussion of the posterior distribution

1. Result is a probability **distribution** (posterior distr.)
2. Posterior mean is **intermediate** between prior mean and observation.
3. Weight of each depends on prior variance and measurement error.
4. Posterior variance is **lower** than both prior variance and measurement error variance.
5. Used just **one data point** and still got estimator.

## Frequentist *versus* Bayesian

Bayesian analysis introduces an element of **subjectivity**:  
*the prior distribution.*

But its representation of the uncertainty is **easy** to understand

- the uncertainty is assessed conditionally to the observations,
- the calculations are straightforward when the posterior distribution is known.

## Which is better?

**Bayesian methods often lead to**

- more **realistic** estimated parameter values,
- in some cases, more **accurate** model predictions.

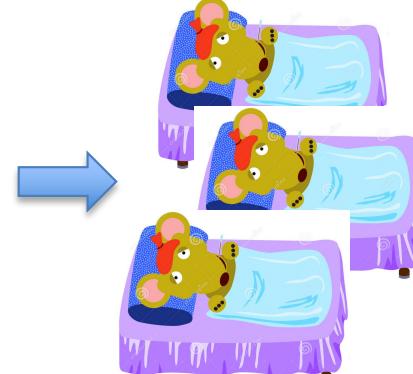
**Problems when prior information is wrong and when one has a strong confidence in it.**

## Example 3

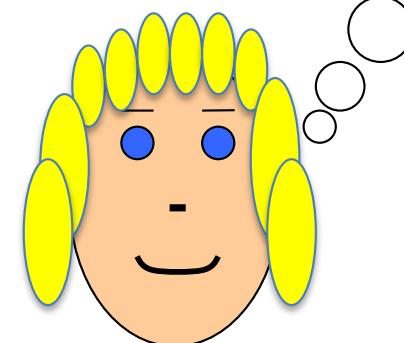
**Estimation of a probability of contamination  
by combining a measurement with expert knowledge**



$Y$  diseased rats



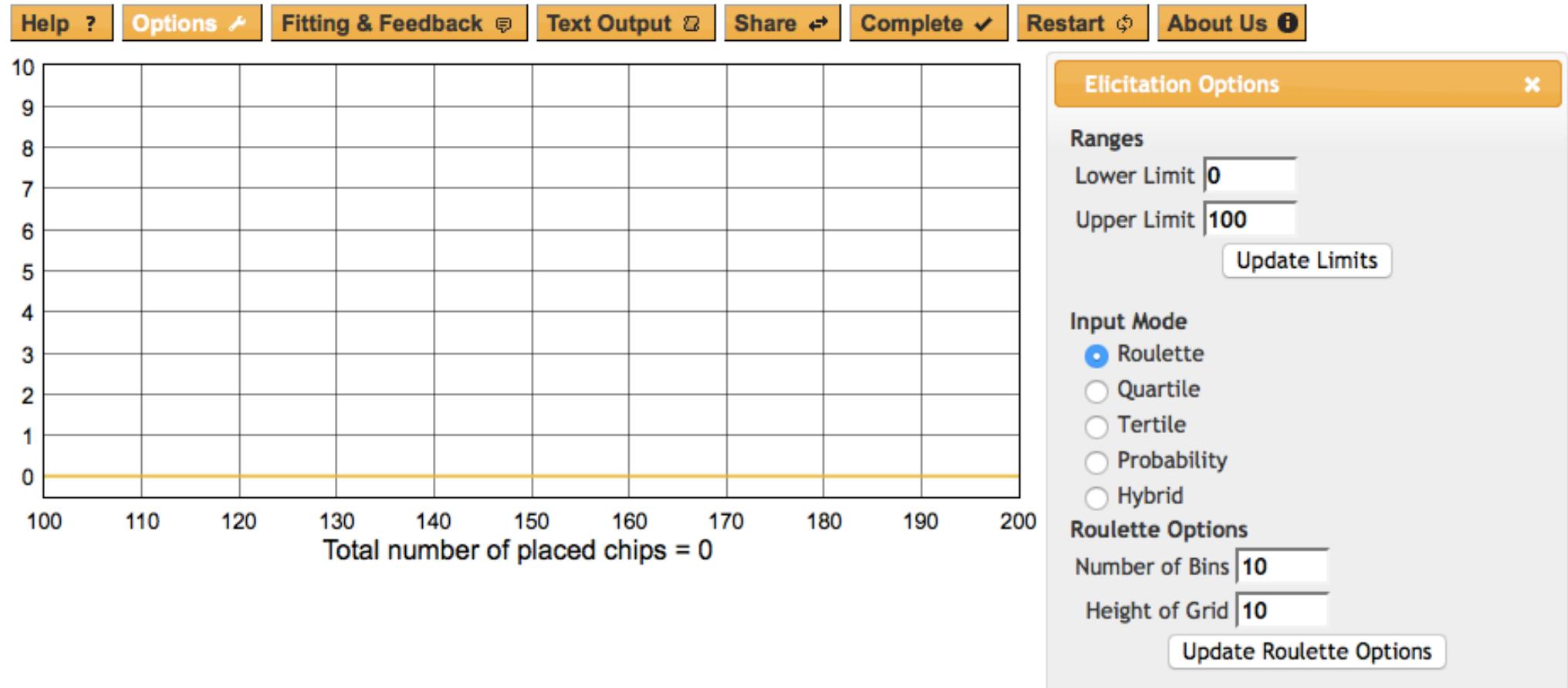
*Prob.=?*



Probability of contamination?

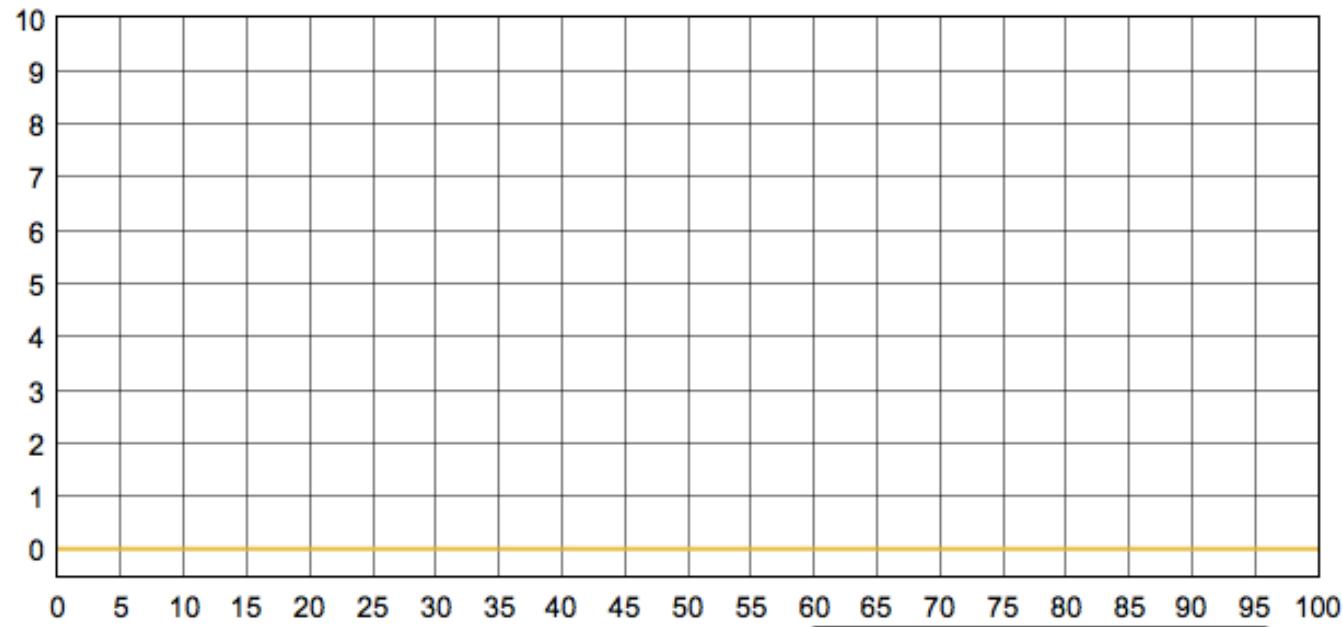
*Expert*

# Expert elicitation



Match Uncertainty Elicitation Tool  
Shelf package

Help ? Options ↗ Fitting & Feedback ↗ Text Output ↗ Share ↗ Complete ✓ Restart ↗ About Us ⓘ



Click to allocate chips to bins.  
The probability within a bin is  
represented by the proportion of  
chips allocated to that bin.

### Elicitation Options ✖

#### Ranges

Lower Limit

Upper Limit

#### Input Mode

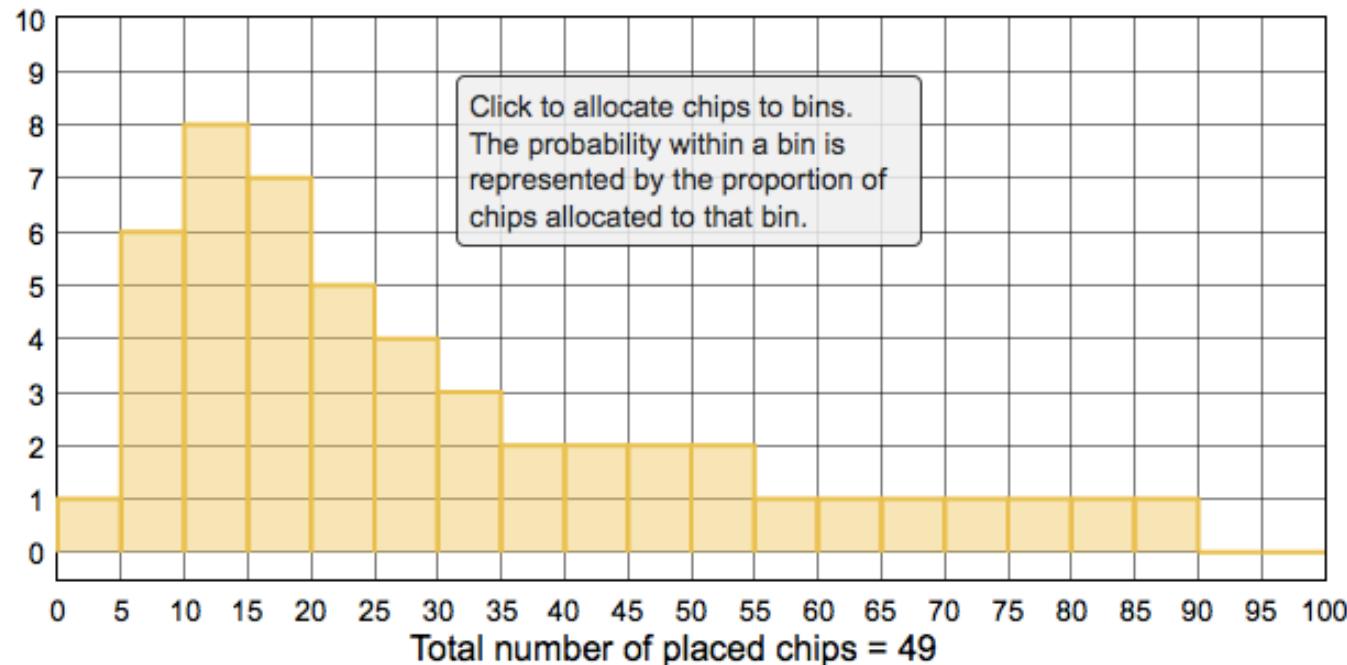
- Roulette
- Quartile
- Tertile
- Probability
- Hybrid

#### Roulette Options

Number of Bins

Height of Grid

Help ? Options ↗ Fitting & Feedback ↗ Text Output ↗ Share ↗ Complete ✓ Restart ↗ About Us ⓘ



### Elicitation Options ✖

#### Ranges

Lower Limit

Upper Limit

#### Input Mode

- Roulette
- Quartile
- Tertile
- Probability
- Hybrid

#### Roulette Options

Number of Bins

Height of Grid

Help ?

Options ↗

Fitting &amp; Feedback 📈

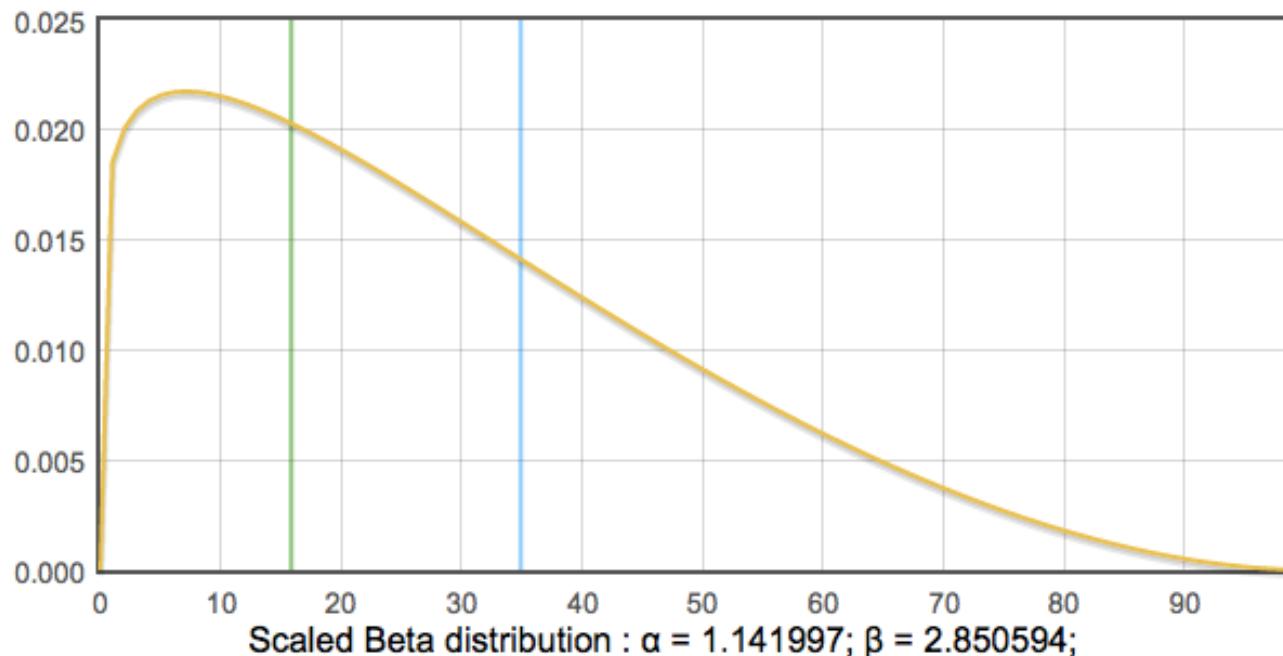
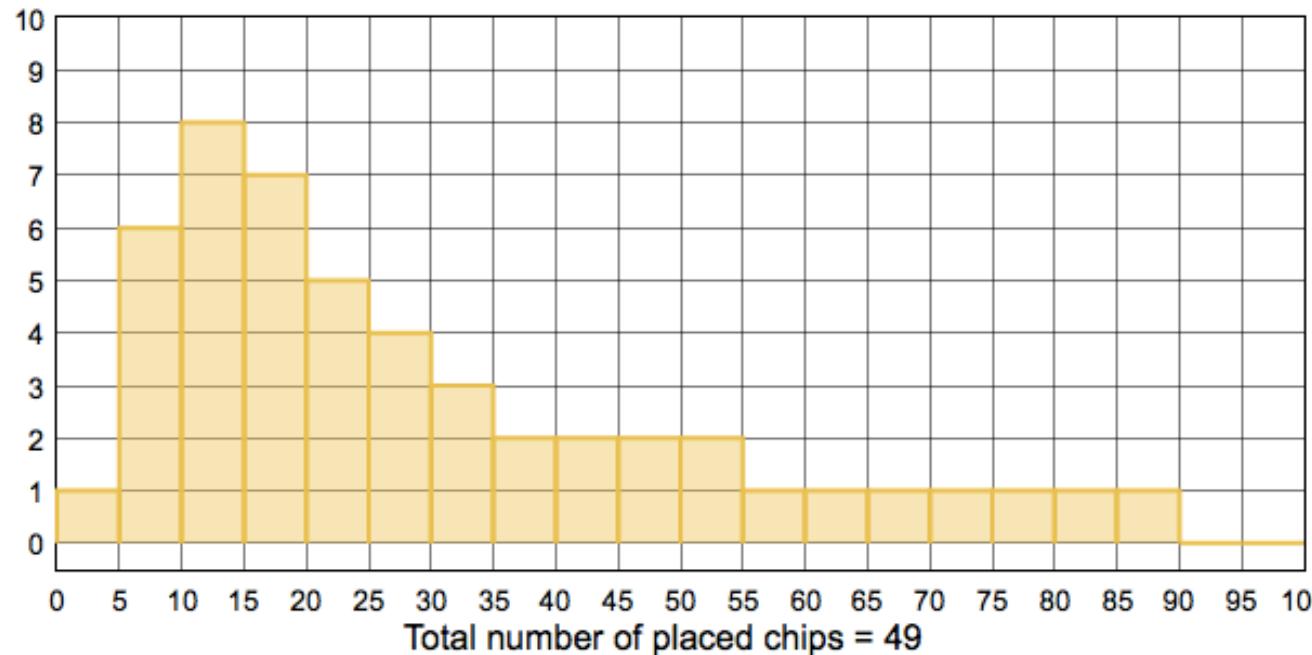
Text Output ↗

Share ↖

Complete ✓

Restart ⏪

About Us ⓘ

Elicitation Options ✖

## Ranges

Lower Limit Upper Limit 

## Input Mode

- Roulette
- Quartile
- Tertile
- Probability
- Hybrid

## Roulette Options

Number of Bins Height of Grid Fitting & Feedback ✖

## Distribution

- Normal
- Student-t
- Scaled Beta
- Gamma
- Log Normal
- Log Student-t
- Auto-select best fit

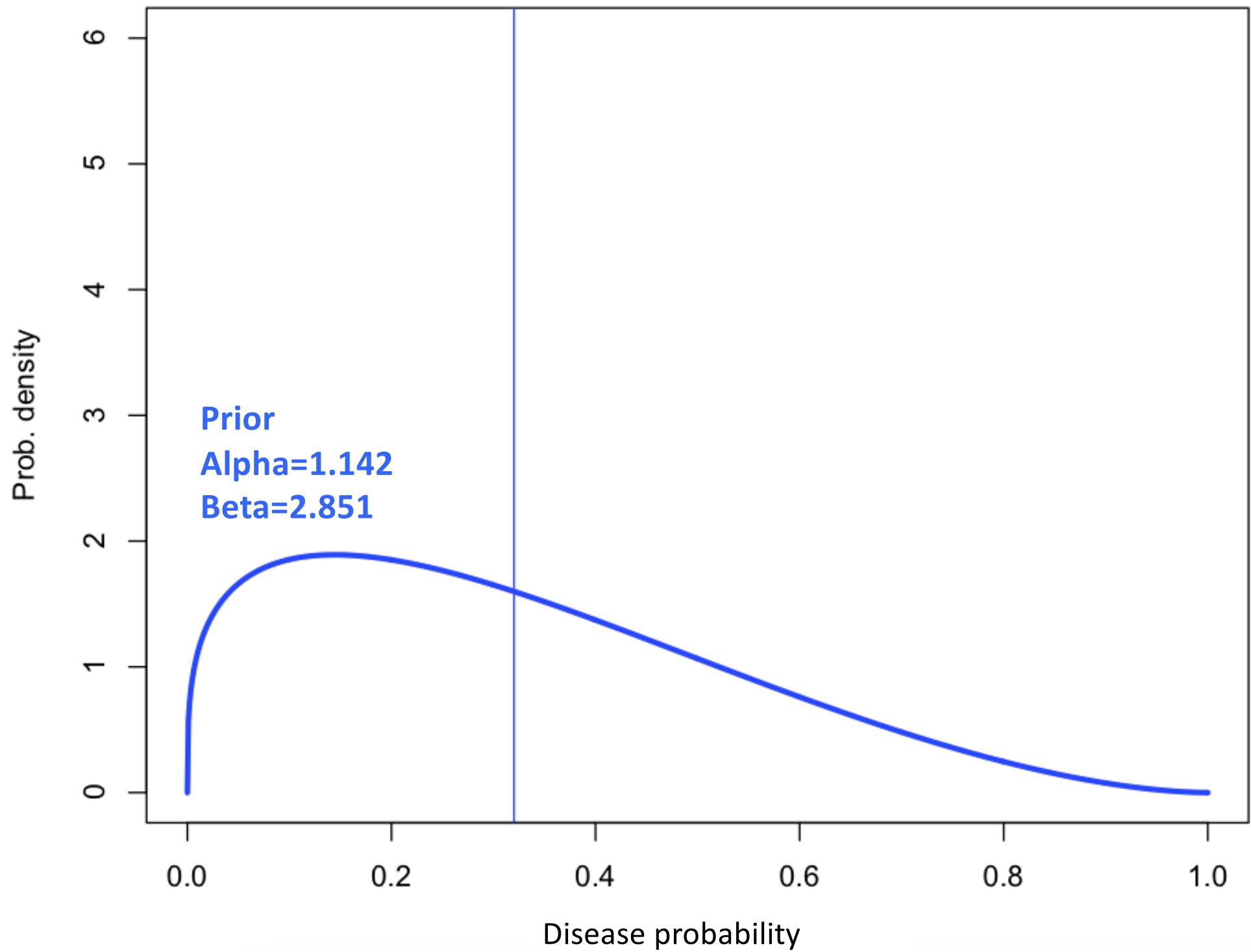
## Feedback Percentiles

33 rd percentile

= 33 33 rd percentile 15.9

66 th percentile

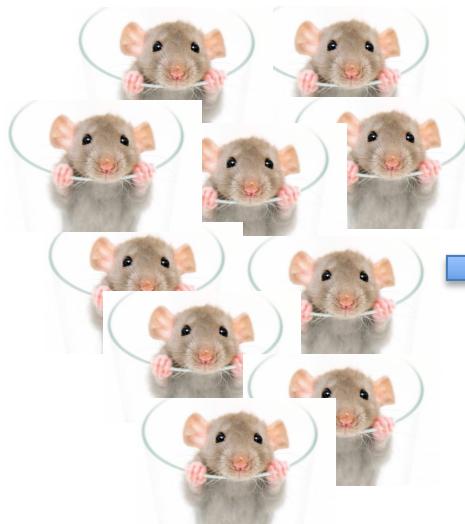
= 66 66 th percentile 34.9



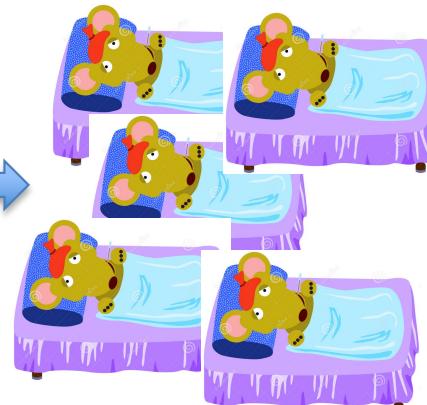
## Example 3

**Estimation of a probability of contamination  
by combining a measurement with expert knowledge**

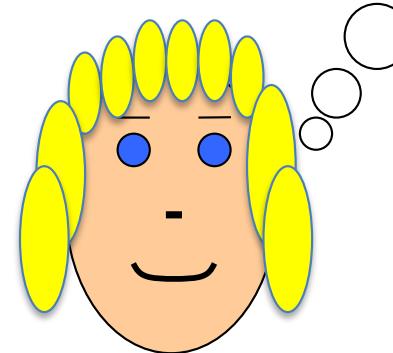
$N=10$  rats



$\gamma=5$  diseased rats

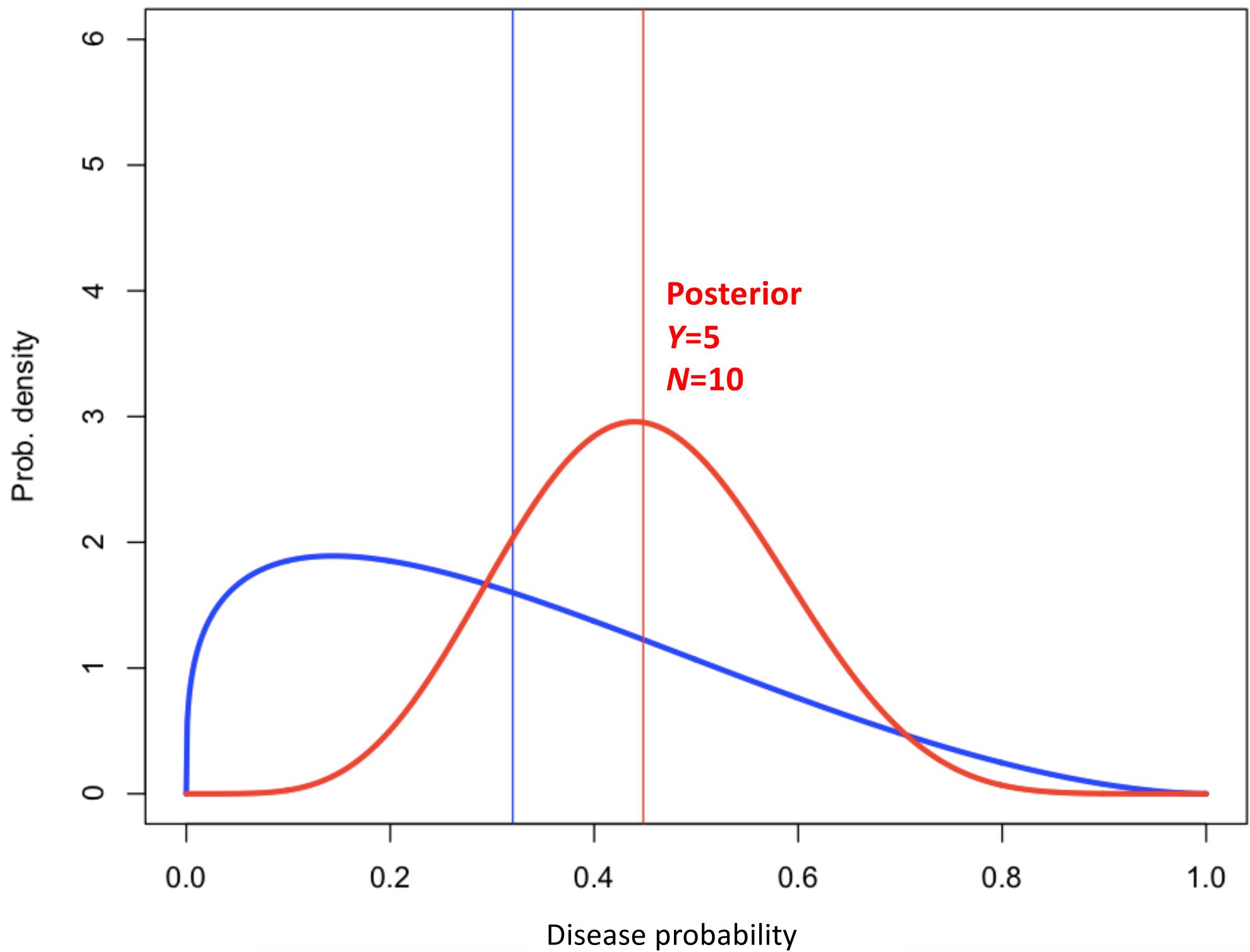


Alpha=1.142  
Beta=2.851



Probability of contamination?

*Expert*



Prior

$$\text{Beta}(\alpha, \beta)$$

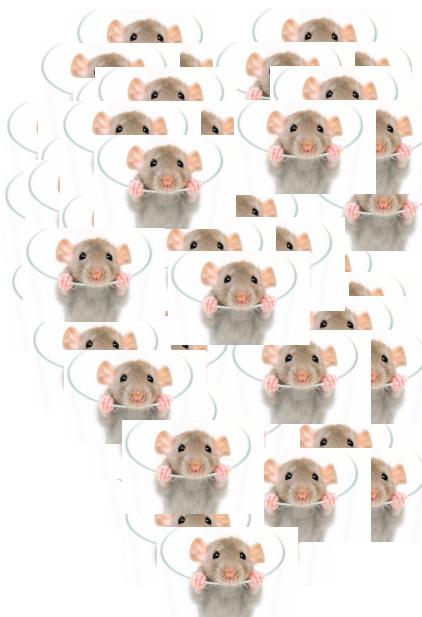
Posterior

$$\text{Beta}(\alpha + Y \beta + N - Y)$$

## Example 3

**Estimation of a probability of contamination  
by combining a measurement with expert knowledge**

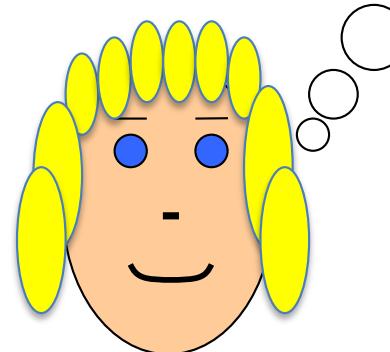
$N=40$  rats



$\gamma=20$  diseased rats

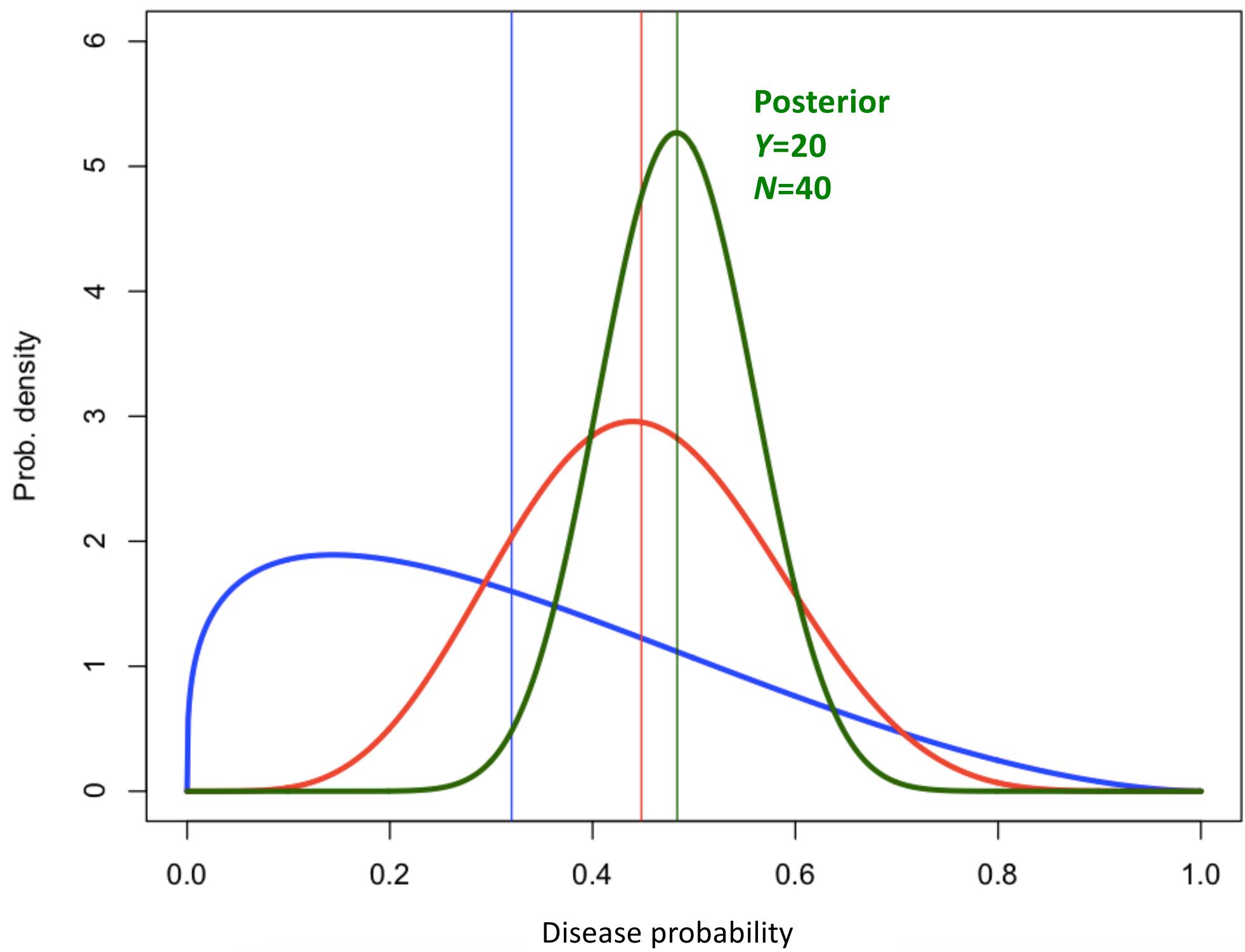


Alpha=1.142  
Beta=2.851



*Expert*

Probability of contamination?

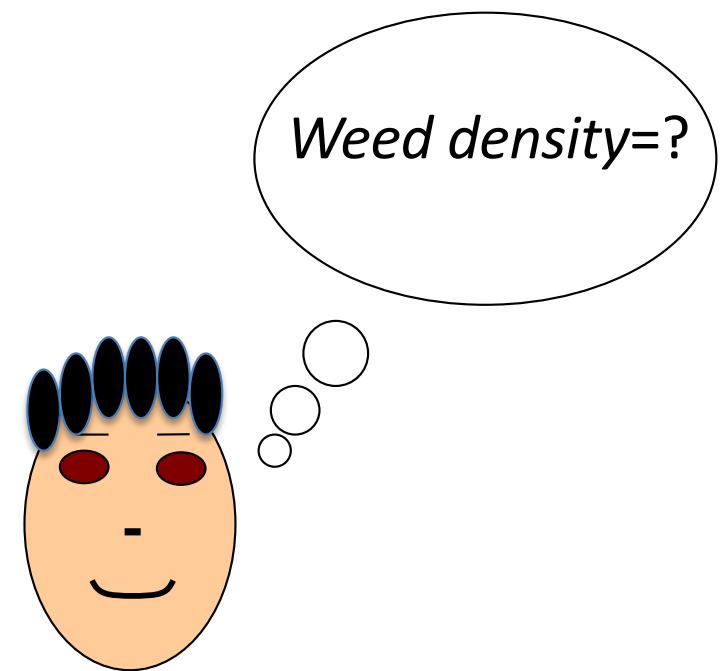


# Exercise

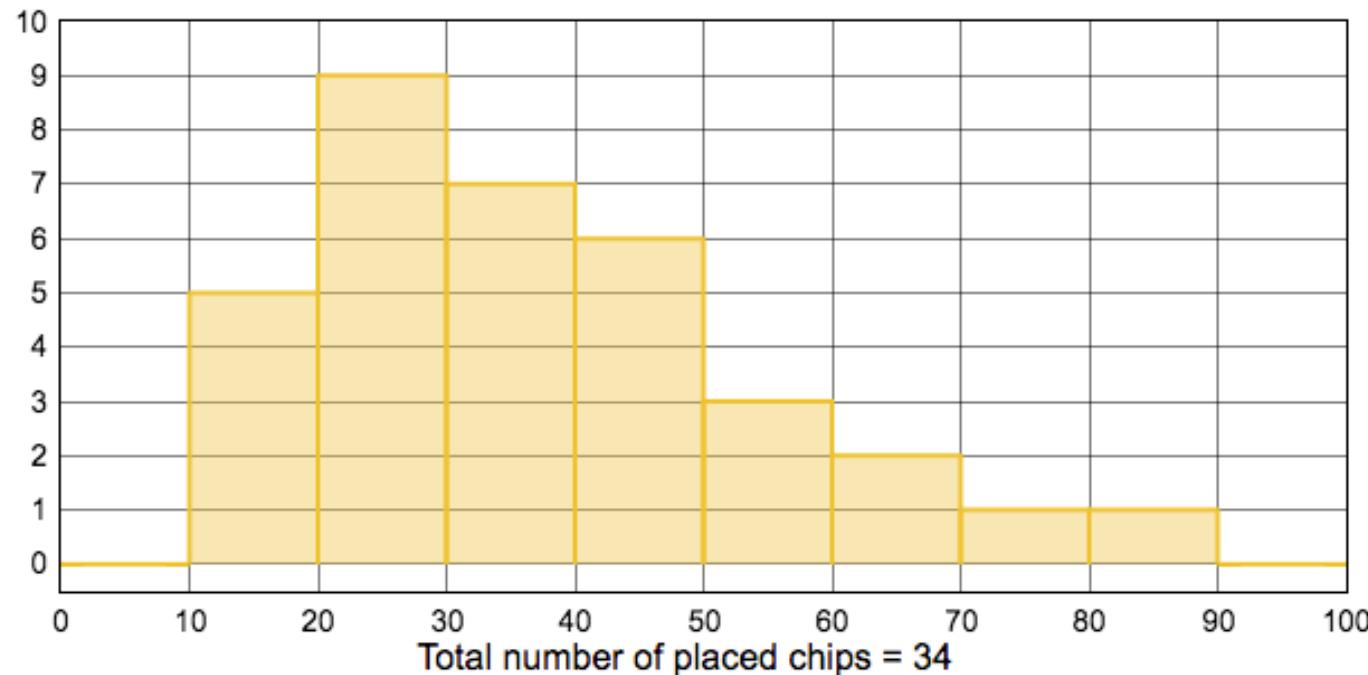
- Compute the posterior distribution for
  - $N=10, Y=0$
  - $N=20, Y=0$
- Calculate the posterior mean
- Plot the results
- Conclude

## Example 4

**Estimation of a plant density  
by combining a measurement with expert knowledge**



*Expert*



### Elicitation Options X

#### Ranges

Lower Limit

Upper Limit

#### Input Mode

- Roulette
- Quartile
- Tertile
- Probability
- Hybrid

### Fitting & Feedback X

#### Distribution

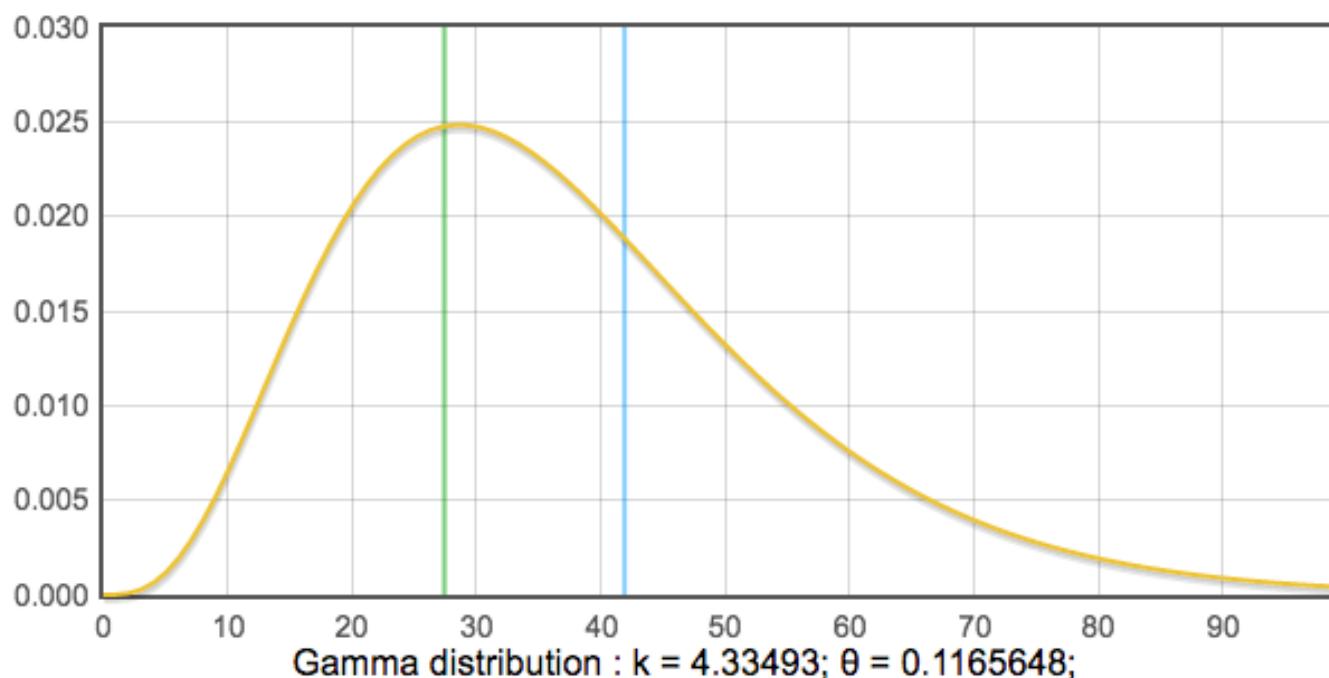
- Normal
- Student-t
- Scaled Beta
- Gamma
- Log Normal
- Log Student-t
- Auto-select best fit

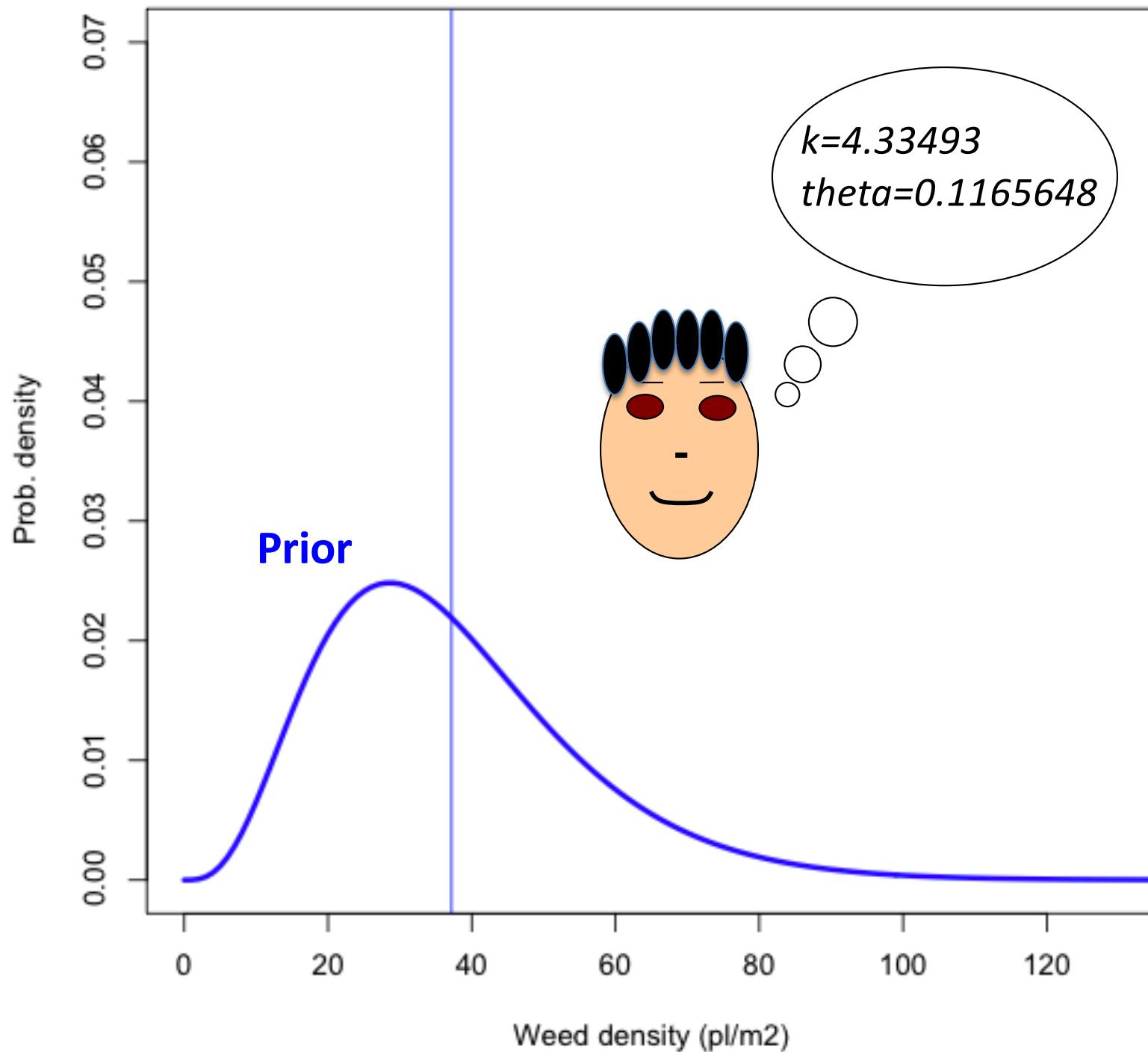
#### Feedback Percentiles

27.4

=

41.9



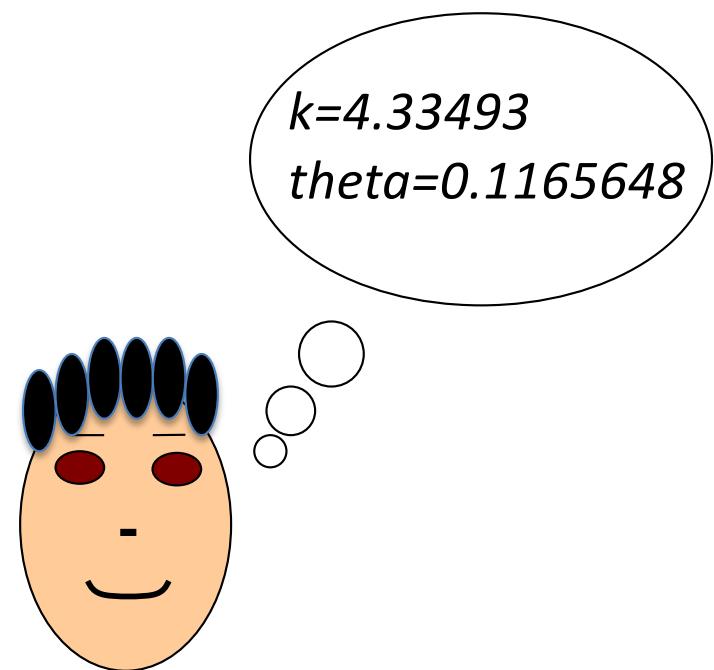


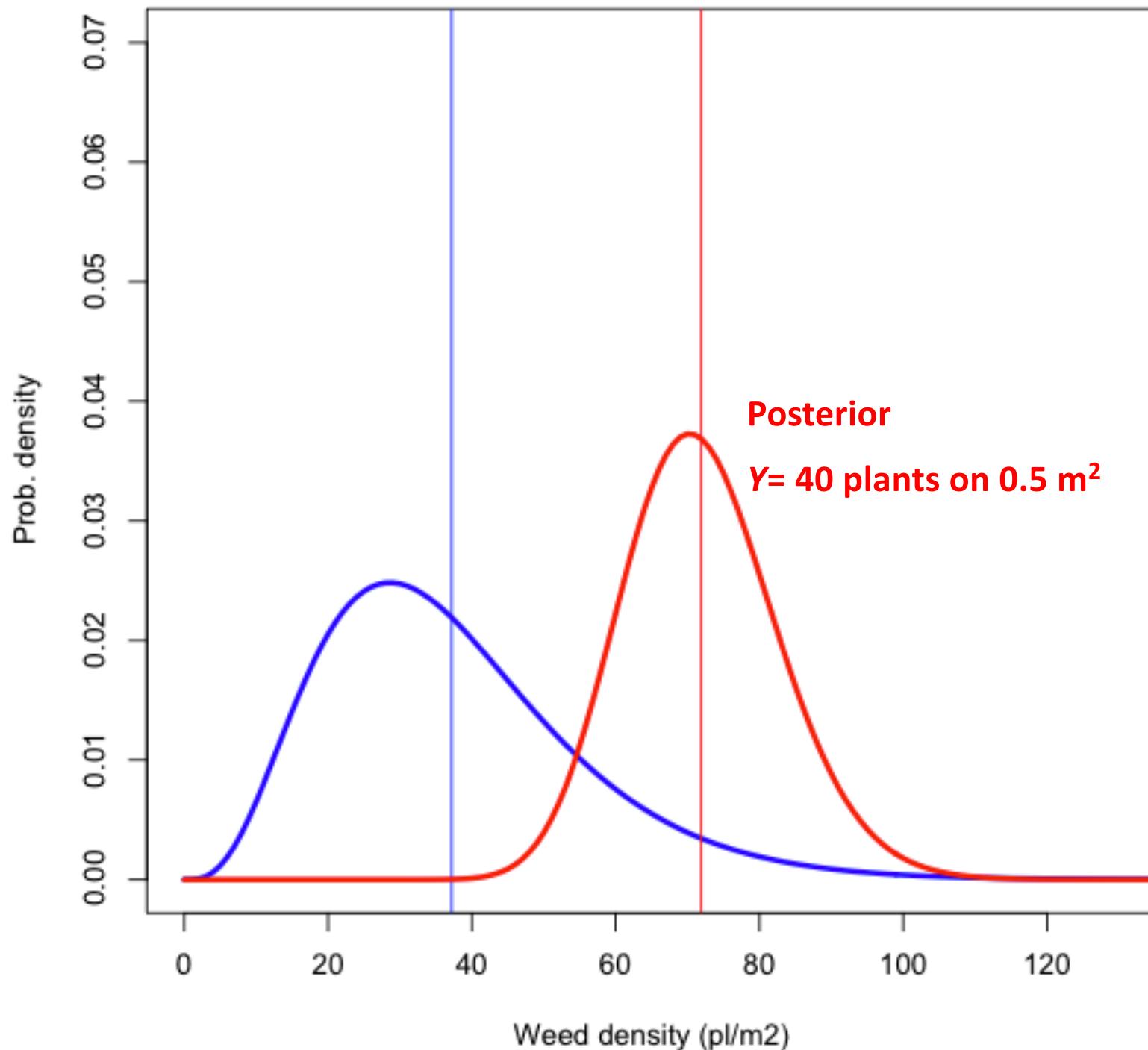
## Example 4

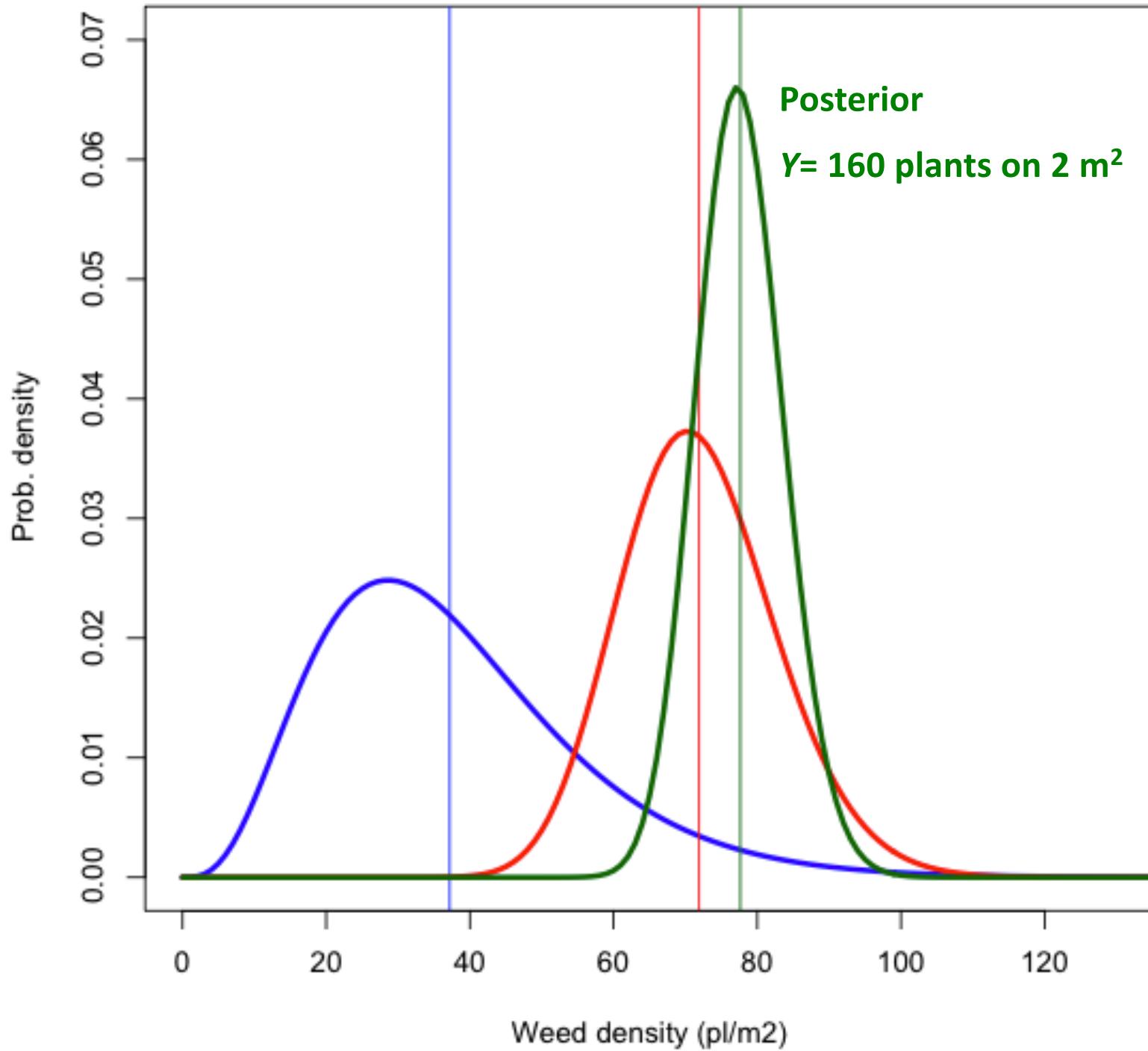
### Estimation of a plant density by combining a measurement with expert knowledge



40 plants on  $0.5 \text{ m}^2$







# Practical considerations

- The analytical expression of the posterior distribution can be derived for simple applications :

Prior	Likelihood	Posterior
Gaussian	Gaussian	Gaussian
Beta	Binomial	Beta
Gamma	Poisson	Gamma

# Practical considerations

- The analytical expression of the posterior distribution can be derived for simple applications :

Prior	Likelihood	Posterior
Gaussian	Gaussian	Gaussian
Beta	Binomial	Beta
Gamma	Poisson	Gamma

**For complex problems, the posterior distribution must be approximated.**

# Algorithms

- Importance sampling
  - Metropolis-Hastings
  - Gibbs sampling
  - Integrated Nested Laplace approximation
- ...
- Markov chain Monte Carlo  
(MCMC)**

# Algorithms

- Importance sampling
  - Metropolis-Hastings
  - Gibbs sampling
  - Integrated Nested Laplace approximation
- Markov chain Monte Carlo  
(MCMC)

...

# The importance sampling algorithm

- **Step 0.** Choose a « proposal » distribution  $g(\theta)$ . This can in principle be ANY (well, almost any) distribution.
- **Step 1.** Generate a sample of size  $N$  (for example,  $N=50000$ ) from  $g(\theta)$ . This gives  $\theta_1, \theta_2, \dots, \theta_N$ .
- **Step 2.** Calculate weights  $w_i = P(\theta_i|y) / g(\theta_i)$
- **Step 3.** Calculate normalized weights

$$w_i^* = \frac{w_i}{\sum_{i=1}^N w_i}$$

- **Step 4.** The weighted sample  $(\theta_1, w_1^*), (\theta_2, w_2^*), \dots, (\theta_N, w_N^*)$  provides an approximation to the posterior distribution
- **Step 5.** Resampling

# Algorithms

- Importance sampling
- Metropolis-Hastings
- Gibbs sampling
- Integrated Nested Laplace approximation

...

# The Metropolis-Hastings algorithm

**Step 0.** Choose a starting value  $\theta_1$ . Define a proposal distribution  $P_p(\theta_c|\theta_i)$ .  
(For example, use a normal distribution with mean equal to  $\theta_i$  ).

Repeat steps 1-3 for  $i=1,\dots,N$

**Step 1.** Generate a candidate parameter value  $\theta_c$  from  $P_p(\theta_c|\theta_i)$ .

**Step 2.** Calculate

$$T = \frac{P(Y|\theta_c)P(\theta_c)P_p(\theta_i|\theta_c)}{P(Y|\theta_i)P(\theta_i)P_p(\theta_c|\theta_i)}$$

**Step 3.** If  $T \geq 1$ , then  $\theta_{i+1} = \theta_c$ . If  $T < 1$ , then draw  $u$  from a uniform distribution on the interval  $(0, 1)$ . If  $u < T$  then  $\theta_{i+1} = \theta_c$  otherwise  $\theta_{i+1} = \theta_i$ .

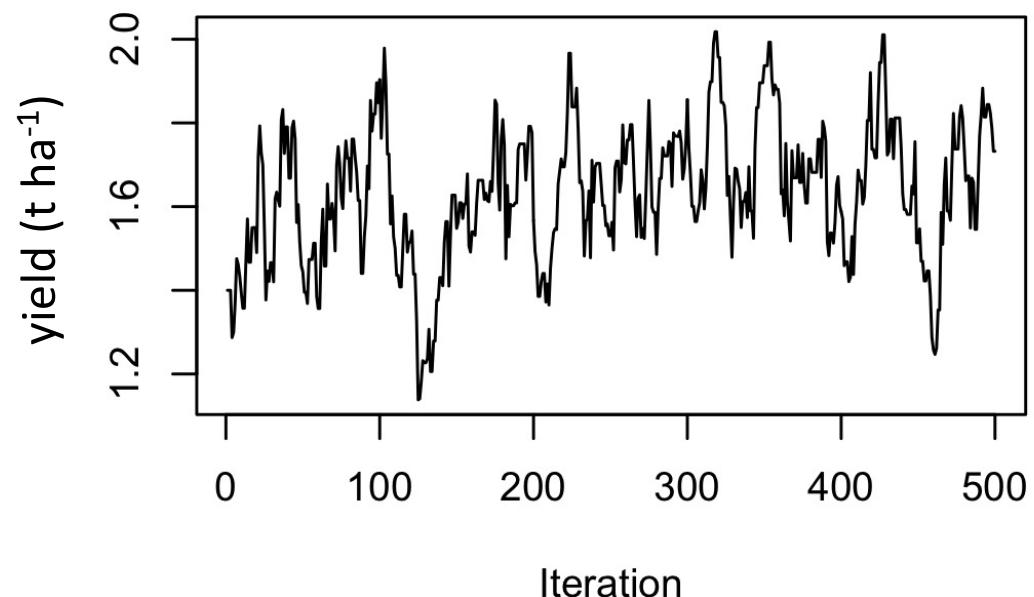
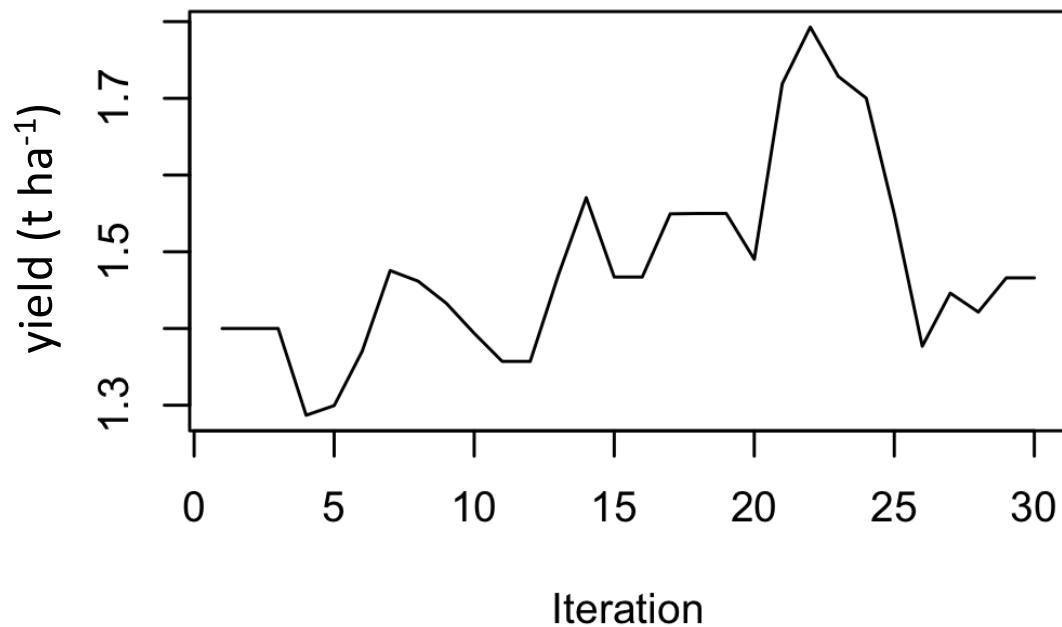
The result of the algorithm is a list of  $N$  parameter values. The same value may be repeated several times.

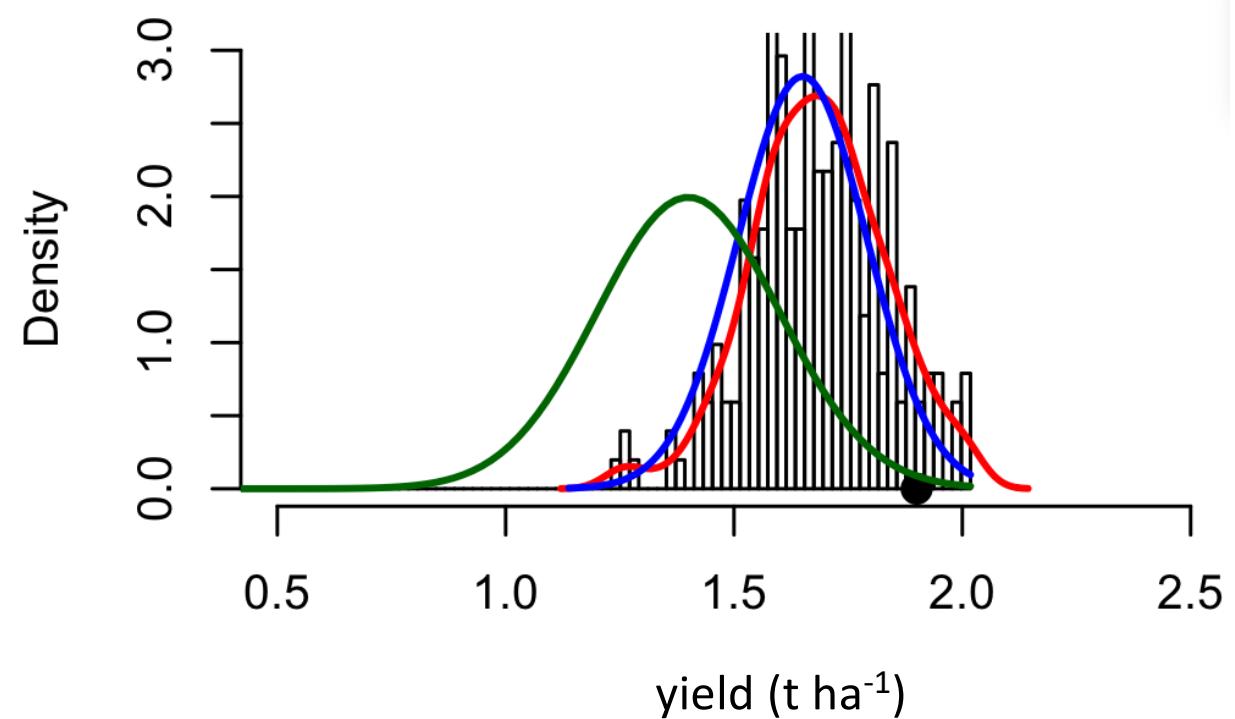
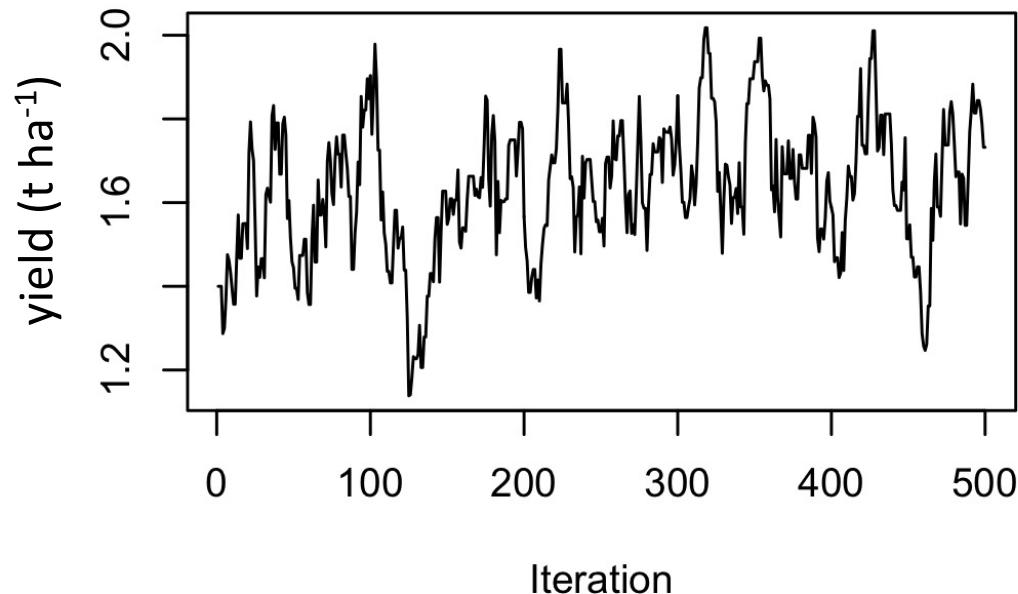
# The Metropolis-Hastings algorithm with symmetric proposal distribution

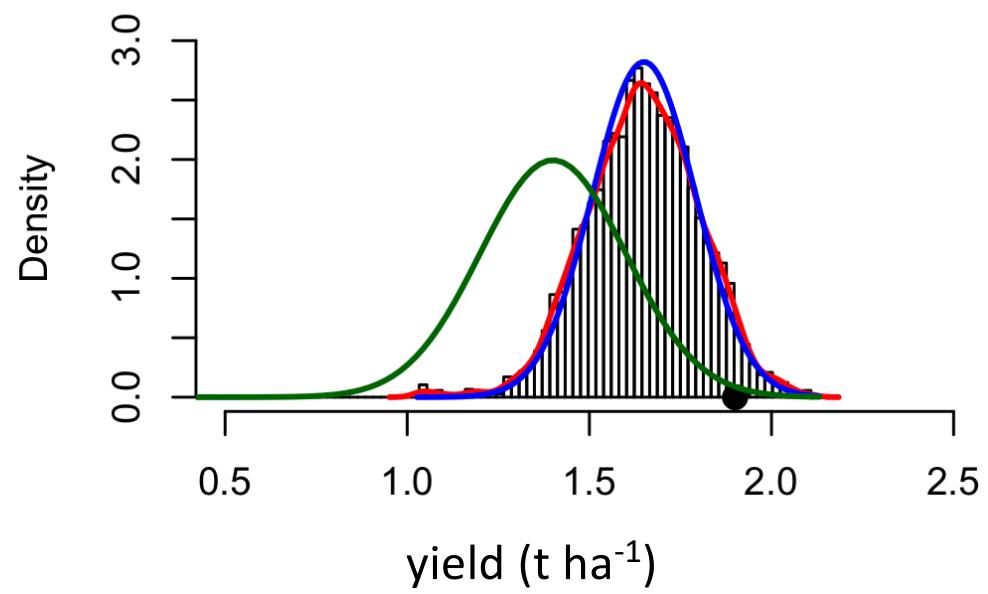
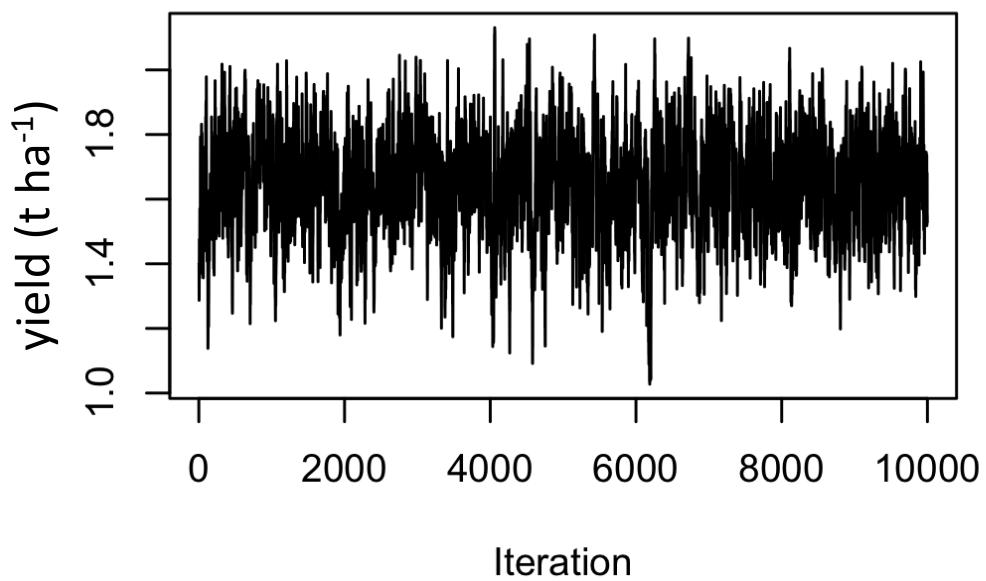
- A common choice for the proposal distribution  $P(\theta_c|\theta_i)$  is a normal distribution with mean equal to  $\theta_i$  and constant variance.
- In this case  $P(\theta_c|\theta_i) = P(\theta_i|\theta_c)$  and the expression for  $T$  simplifies:

$$T = \frac{P(Y|\theta_c)P(\theta_c)}{P(Y|\theta_i)P(\theta_i)}$$

Likelihood                          Prior density







# Exercise

- Open the file MHyield.R
- Run the algorithm with 100 iterations
- Increase the number of iterations to 10000
- Double the standard deviation of the measurement and compute the posterior.
- What is the effect of reducing the accuracy of the measurement?

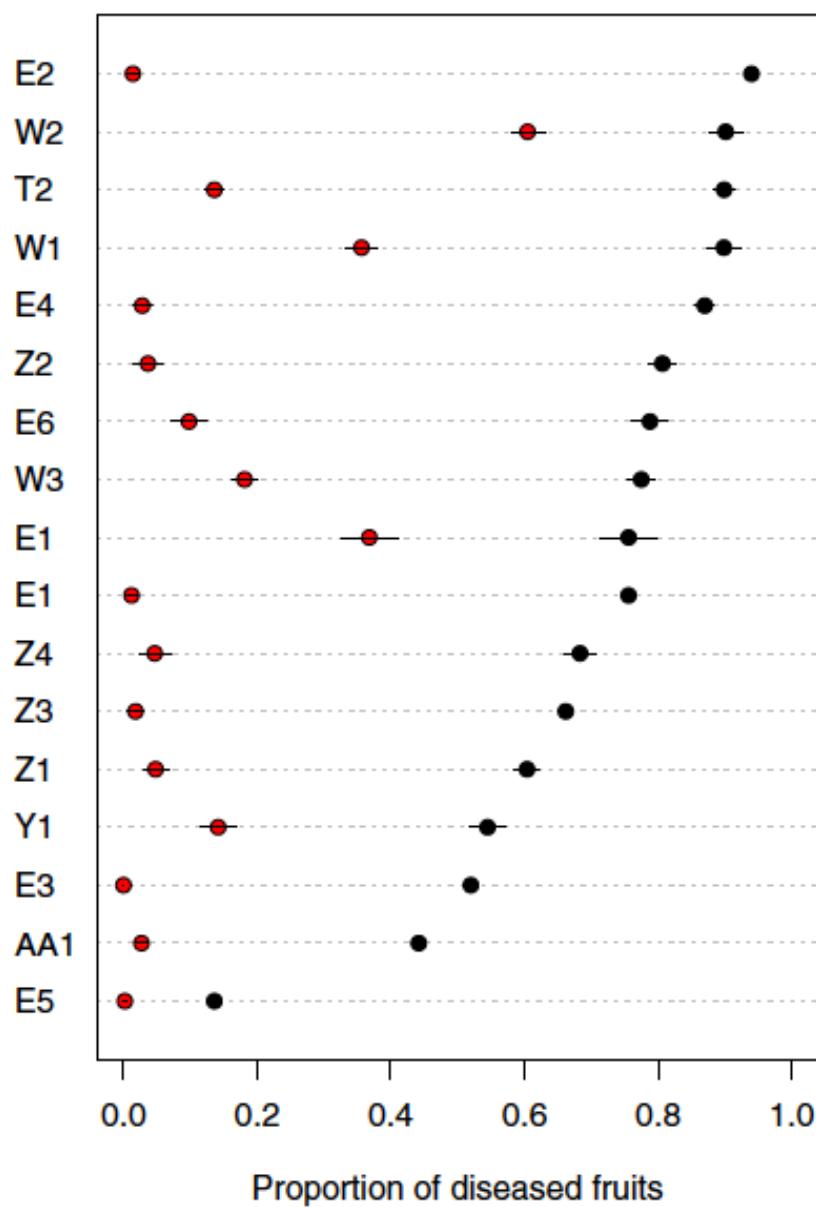
# The Bayesian boom

- New methods for estimating posterior probability distributions
    - ✓ Markov chain Monte Carlo (MCMC) Late 1990s
    - ✓ Importance sampling Late 1990s
    - ✓ Approximate Bayesian Computation (ABC) Early 2000s
    - ✓ Bayesian melding Early 2000s
    - ✓ Bayesian Model Averaging (BMA) Early 2000s
    - ✓ Integrated Nested Laplace Approximations ~ 2010
  - Strong decrease of computational time
  - Dedicated packages (Rjags, MCMCglmm, brms, etc.)
- **It is now possible to apply Bayesian techniques to complex problems**

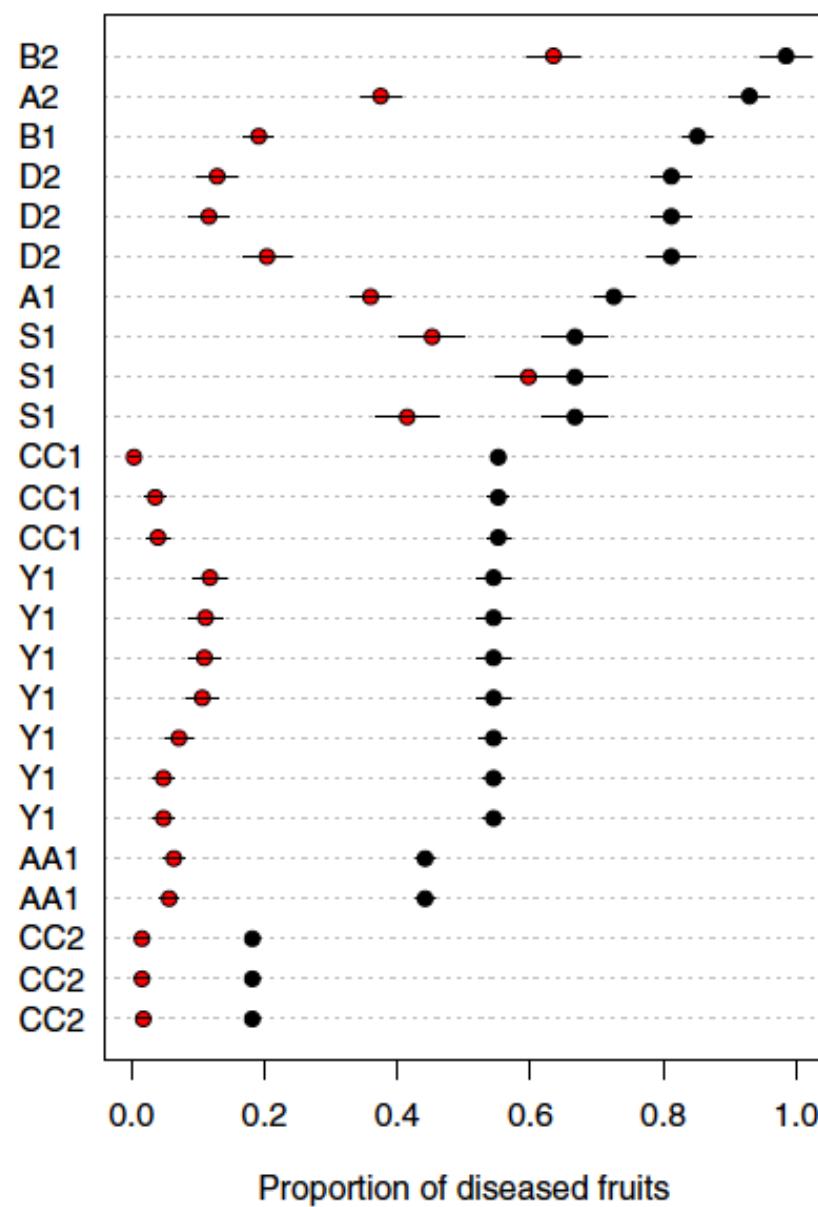
# **Comparison of statistical models in a meta-analysis of fungicide treatments for the control of citrus black spot caused by *Phyllosticta citricarpa***

**D. Makowski • A. Vicent • M. Pautasso •  
G. Stanganelli • T. Rafoss**

dit



cu



$$Y_{ij} \sim Binomial \quad (n_{ij}, \pi_{ij}) \quad \begin{array}{l} i = \text{study index} \\ j = \text{treatment index} \end{array}$$

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_{0i} + \alpha_{Ti}X_{ij}$$

$$\begin{pmatrix} \alpha_{0i} \\ \alpha_{Ti} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_0 \\ \mu_T \end{pmatrix}, \Sigma\right] \quad \Sigma = \begin{bmatrix} \sigma_0^2 & c \\ c & \sigma_T^2 \end{bmatrix}$$

$$\mu_0, \mu_T \sim N(0, 10^6).$$

$$\Sigma \sim InvWish(\bar{\psi}, v)$$

Type of fungicide	Disease incidence	
	Untreated fruits	Treated fruits
“cu”	0.74 (0.53-0.87)	0.16 (0.072-0.32)
“dit”	0.74 (0.62-0.83)	0.053 (0.022-0.12)

# Summary

- Concepts are relatively simple
- Bayesian methods are useful for
  - combining expert knowledge and data
  - analyzing uncertainty
  - dealing with complex data structure
- Software are available: many R packages
- Computation can be long and complex