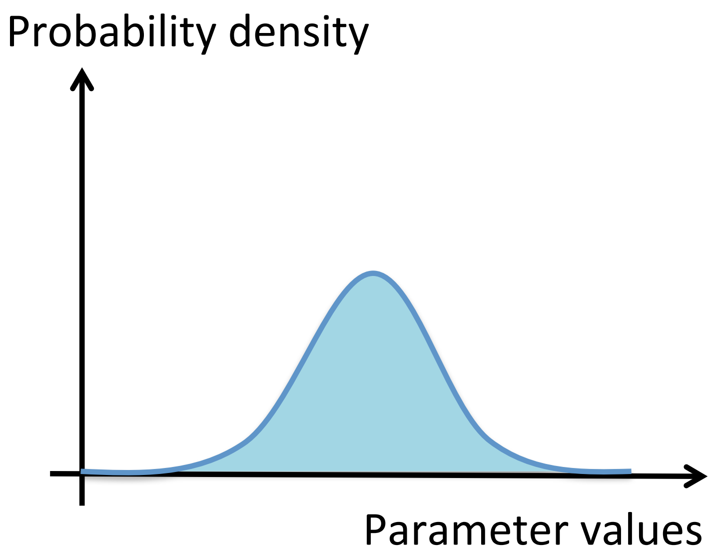
Introduction à la statistique bayésienne

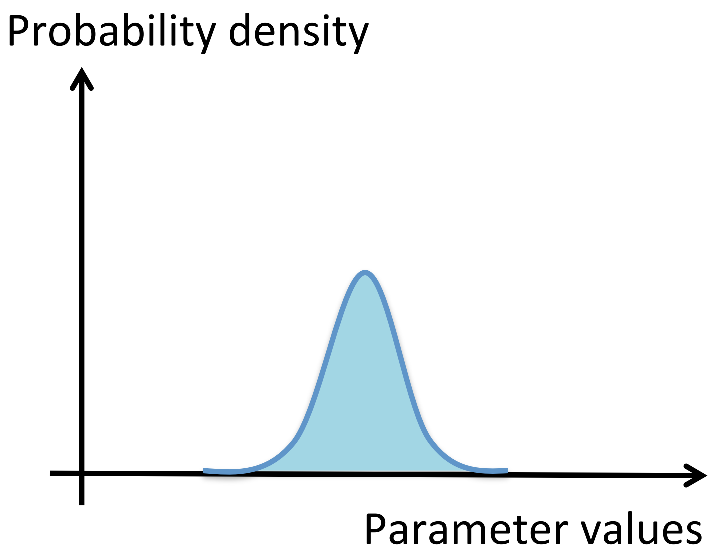
David Makowski

INRA

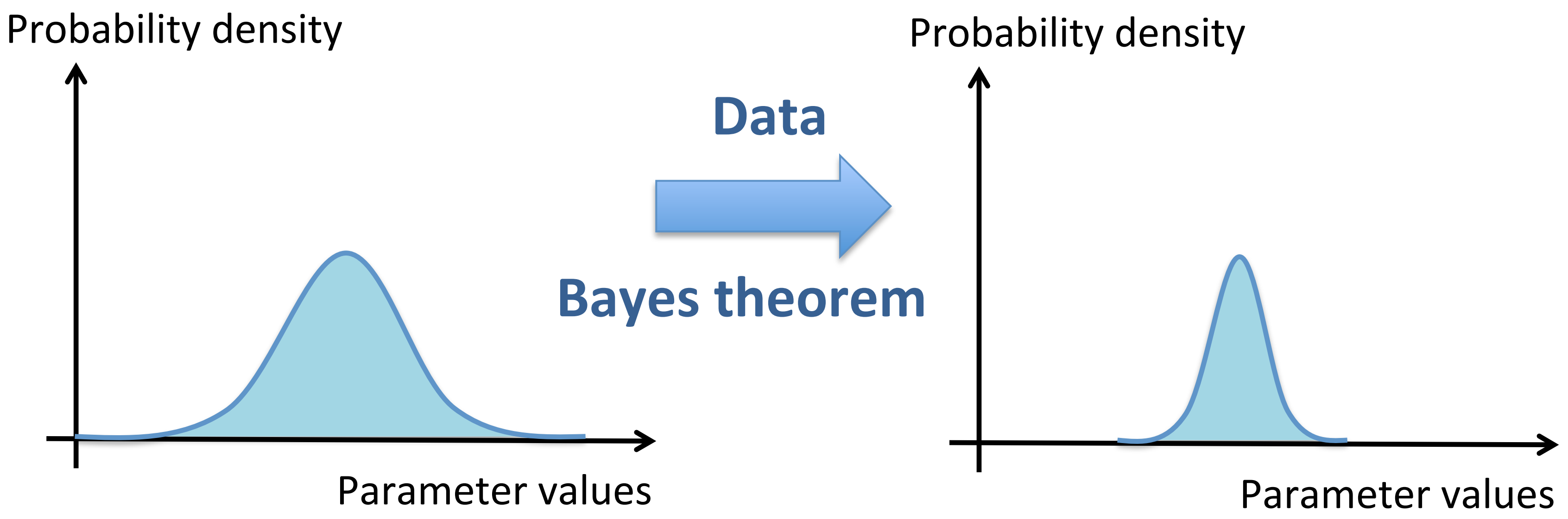
makowski@grignon.inra.fr

PRIOR probability distribution **POSTERIOR probability distribution**





PRIOR probability distribution **POSTERIOR probability distribution**



	Classic methods	Bayesian methods
Parameters	Fixed	Random

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

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

	Classic methods	Bayesian methods
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 (θ)

Parameter = numerical value not calculated by the model and not observed.

Information available to estimate parameters

- A set of observations (y).

- Prior knowledge about parameter values.

Two distributions in Bayes' theorem

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

$$P(\theta)$$

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

$$P(y|\theta)$$

Measurements

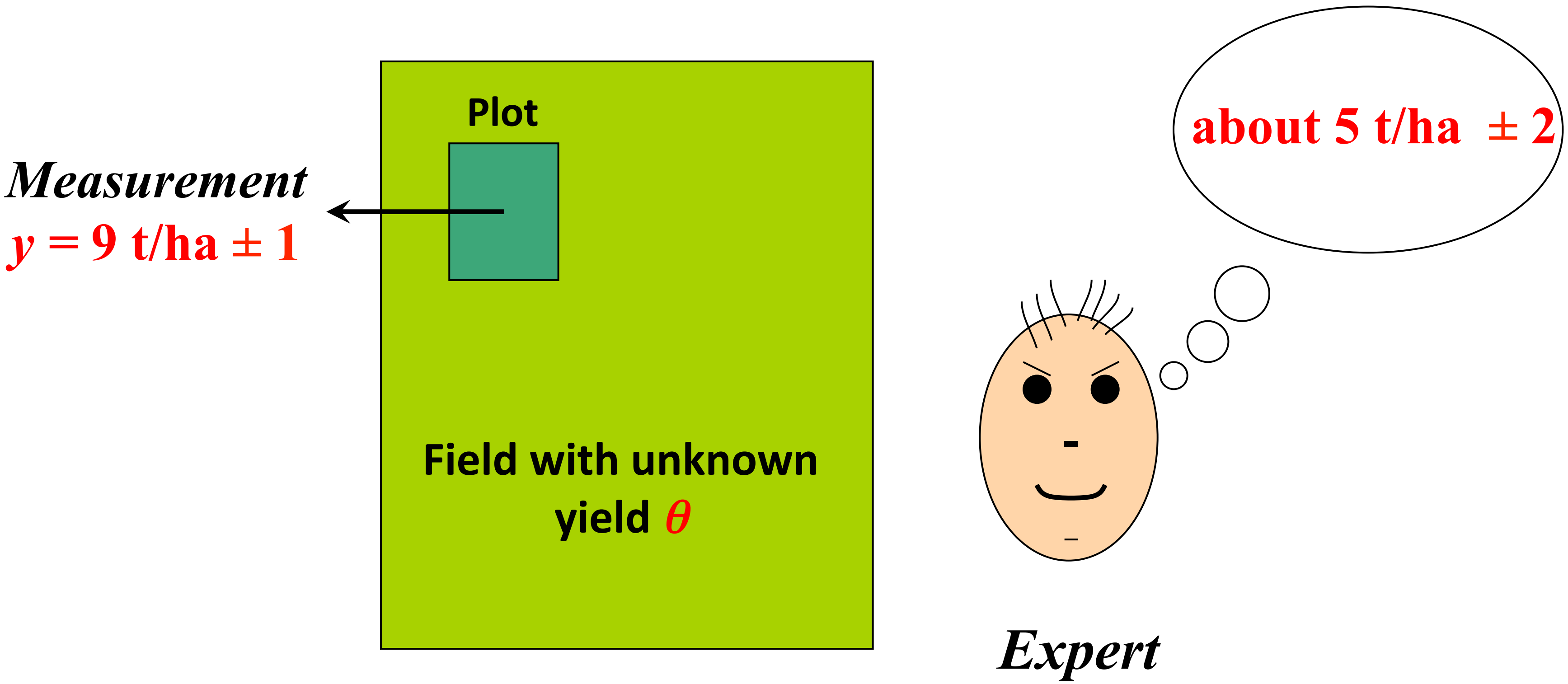
Bayesian method

Prior Information about parameter values

Combined info about parameters

Example 1

Estimation of crop yield θ by combining a measurement with expert knowledge.



Example 1

Estimation of crop yield θ by combining a measurement with expert knowledge.

- One parameter to estimate: the crop yield θ.

- Two types of information available:

  - A measurement equal to 9 t/ha with a standard error equal to 1 t/ha.

  - An estimation provided by an expert equal to 5 t/ha with a standard error equal to 2 t/ha.

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 1 (continued)

Definition of a prior distribution

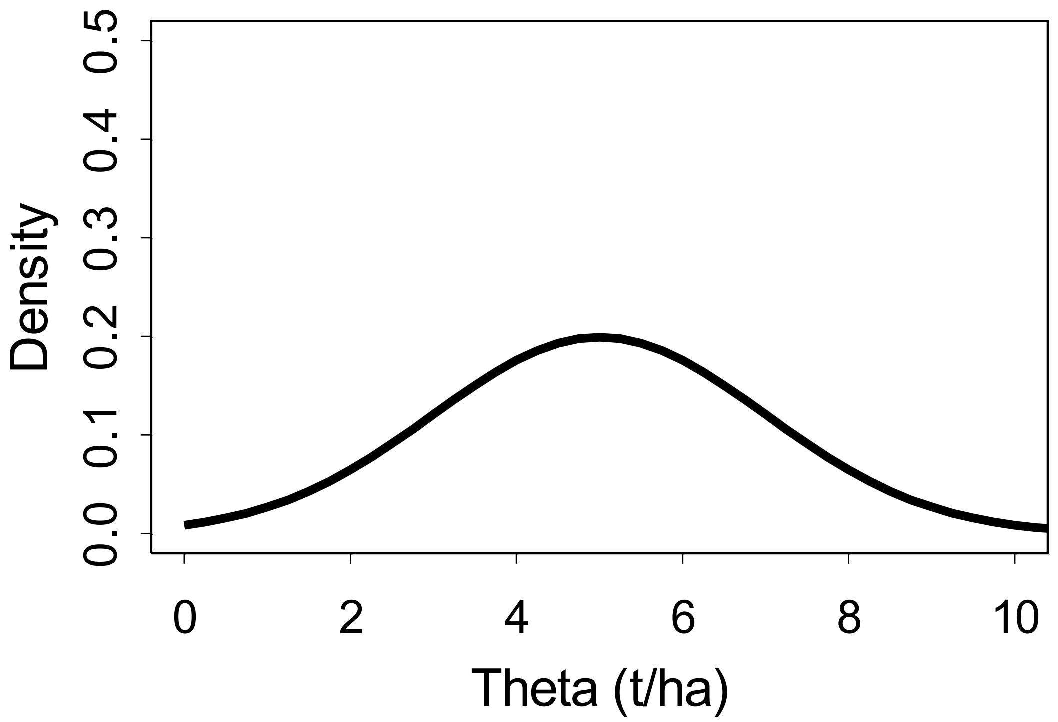
$$\theta \sim N(\mu, \tau^2)$$

$$P(\theta) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left[-\frac{(\theta-\mu)^2}{2\tau^2}\right] = \frac{1}{\sqrt{2\pi 4}} \exp\left[-\frac{(\theta-5)^2}{2\times 4}\right]$$

- Normal probability distribution.
- Expected value equal to 5 t/ha.
- Standard error equal to 2 t/ha

Example 1 (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 | θ)

Example 1 (continued)

Statistical model

$$y = \theta + \varepsilon \quad \text{with } \varepsilon \sim N(0, \sigma^2)$$

$$y \mid \theta \sim N(\theta, \sigma^2)$$

Example 1 (continued)

Definition of a likelihood function

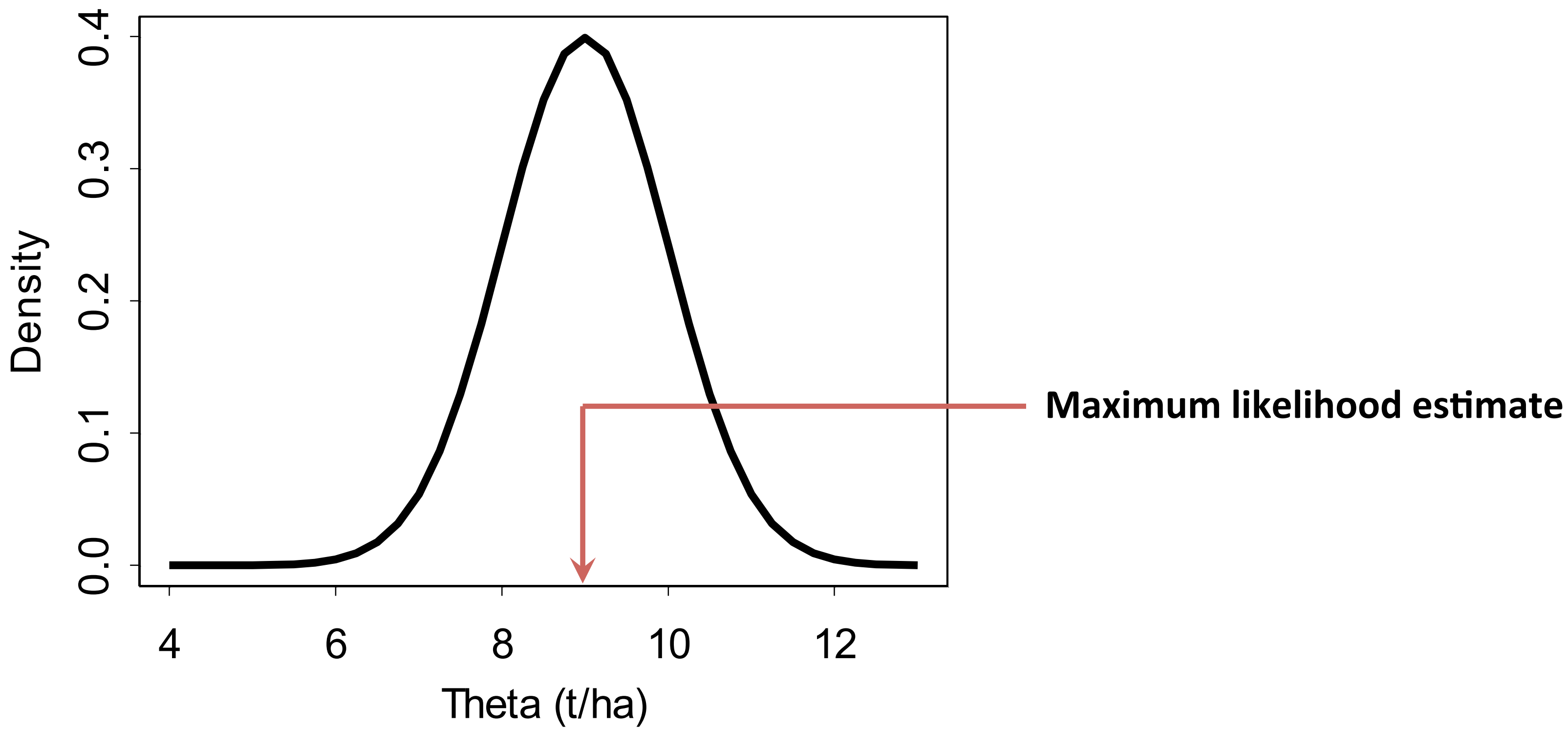
$$y \mid \theta \sim N(\theta, \sigma^2)$$

$$P(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-\theta)^2}{2\sigma^2}\right] = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(9-\theta)^2}{2}\right]$$

- Normal probability distribution.
- Measurement *y* assumed unbiaised and equal to 9 t/ha.
- Standard error *σ* assumed equal to 1 t/ha

Example 1 (continued)

Definition of a likelihood function

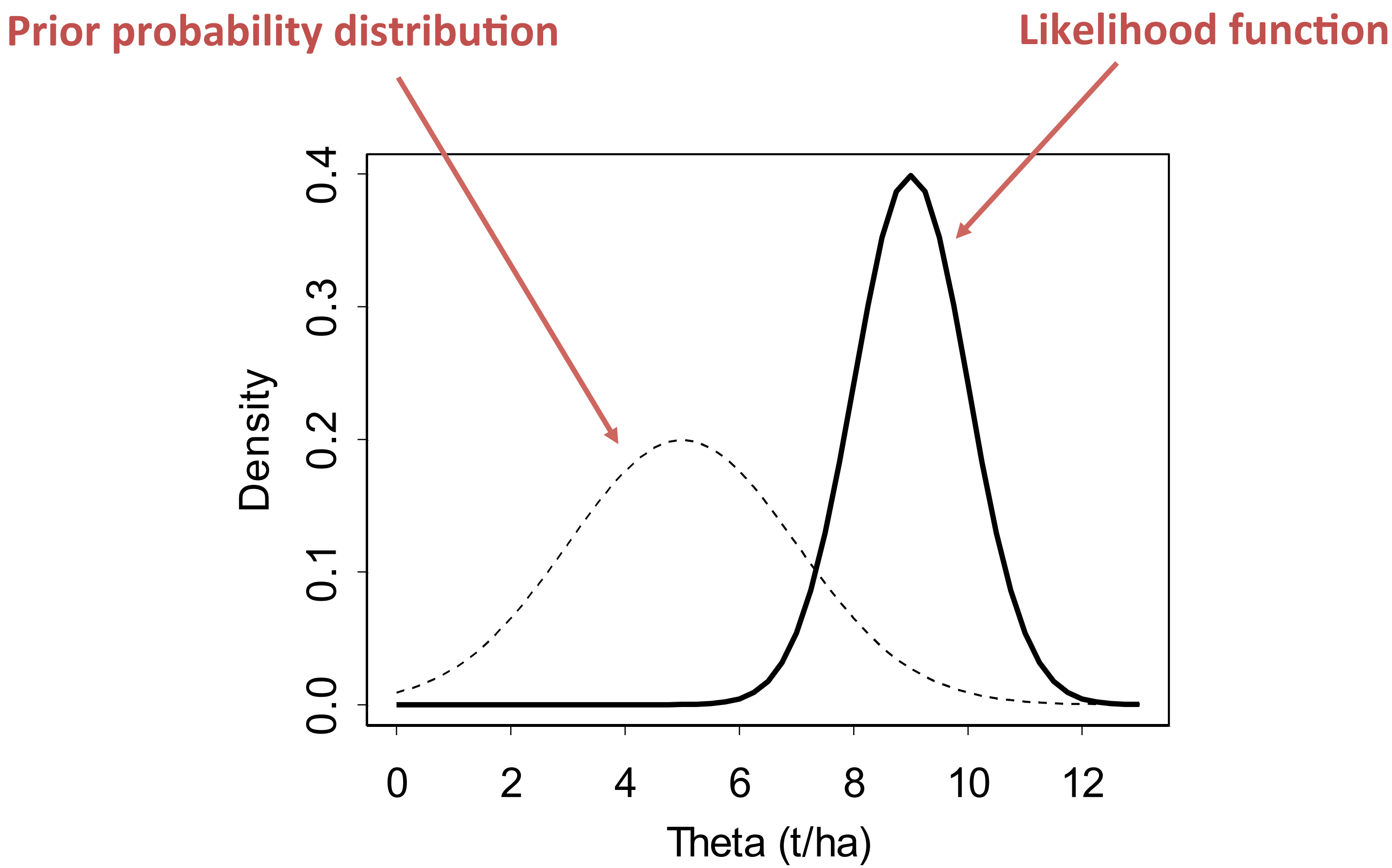


Maximum likelihood

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

The maximum likelihood estimator is the value of θ maximizing

$P(y \mid \theta)$.



Prior probability distribution
Likelihood function
Density
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
12
Theta (t/ha)

Example 1 (continued)

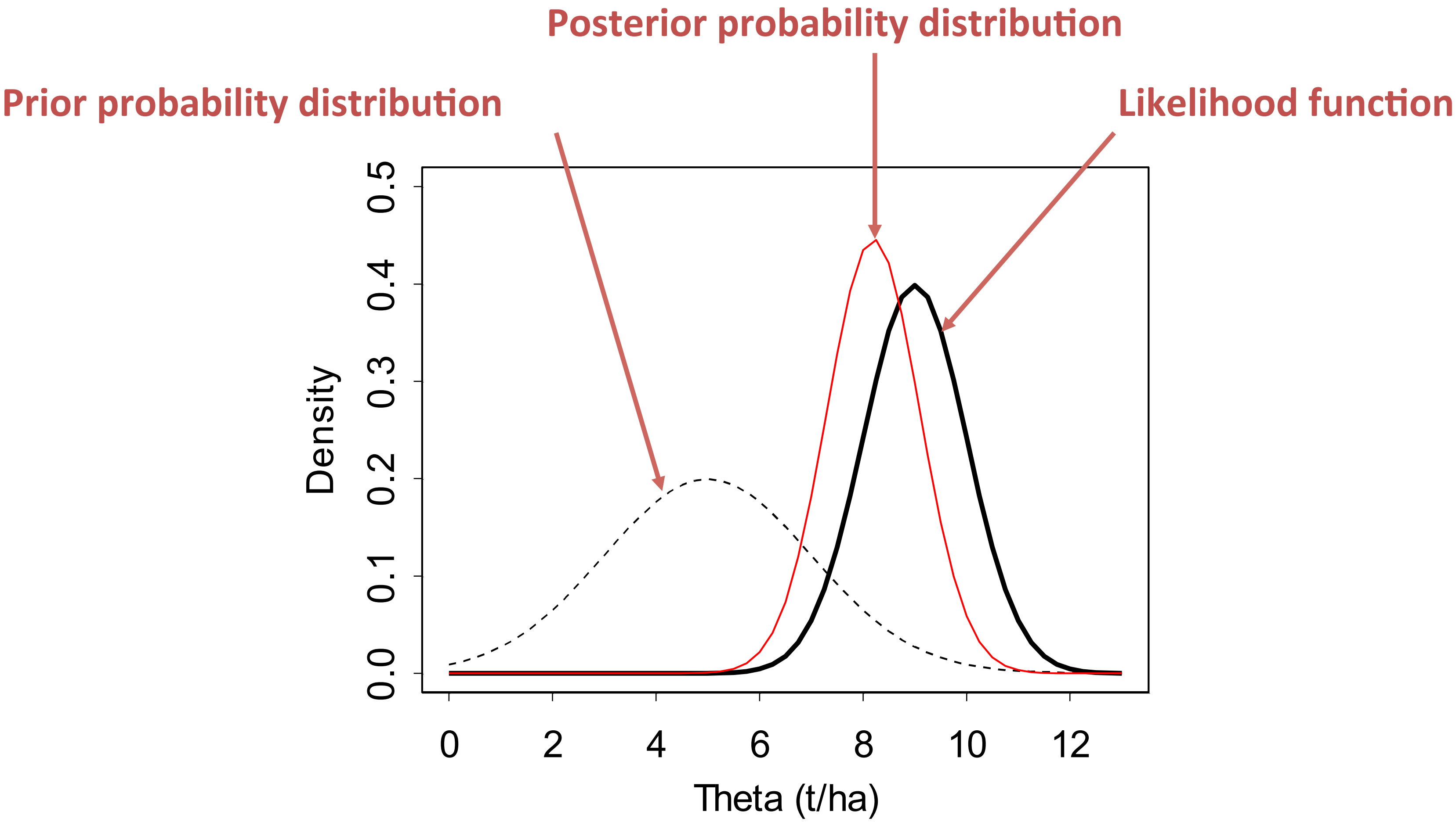
Analytical expression of the posterior distribution

$$\theta \mid y \sim N(\mu_{post}, \tau_{post}^2)$$

$$\mu_{post} = (1 - B) \times \mu + B \times y = 8.2$$

$$\tau^2_{post} = (1 - B) \times \tau^2 = 0.8$$

$$B = \frac{\tau^2}{\tau^2 + \sigma^2} = \frac{4}{5}$$



Posterior probability distribution
Prior probability distribution
Likelihood function
Density
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
10
12
Theta (t/ha)

Example 1 (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

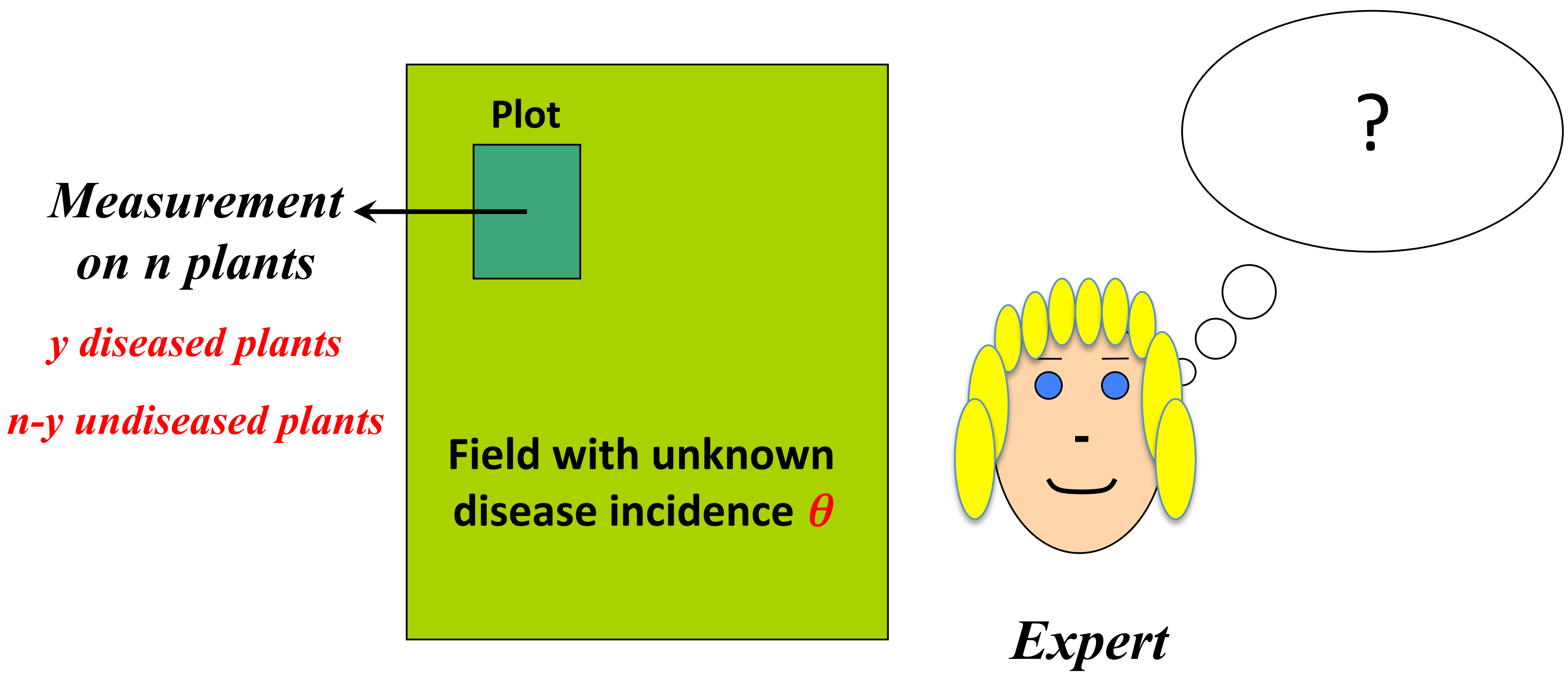
- more realistic estimated parameter values,

- in some cases, more accurate model predictions.

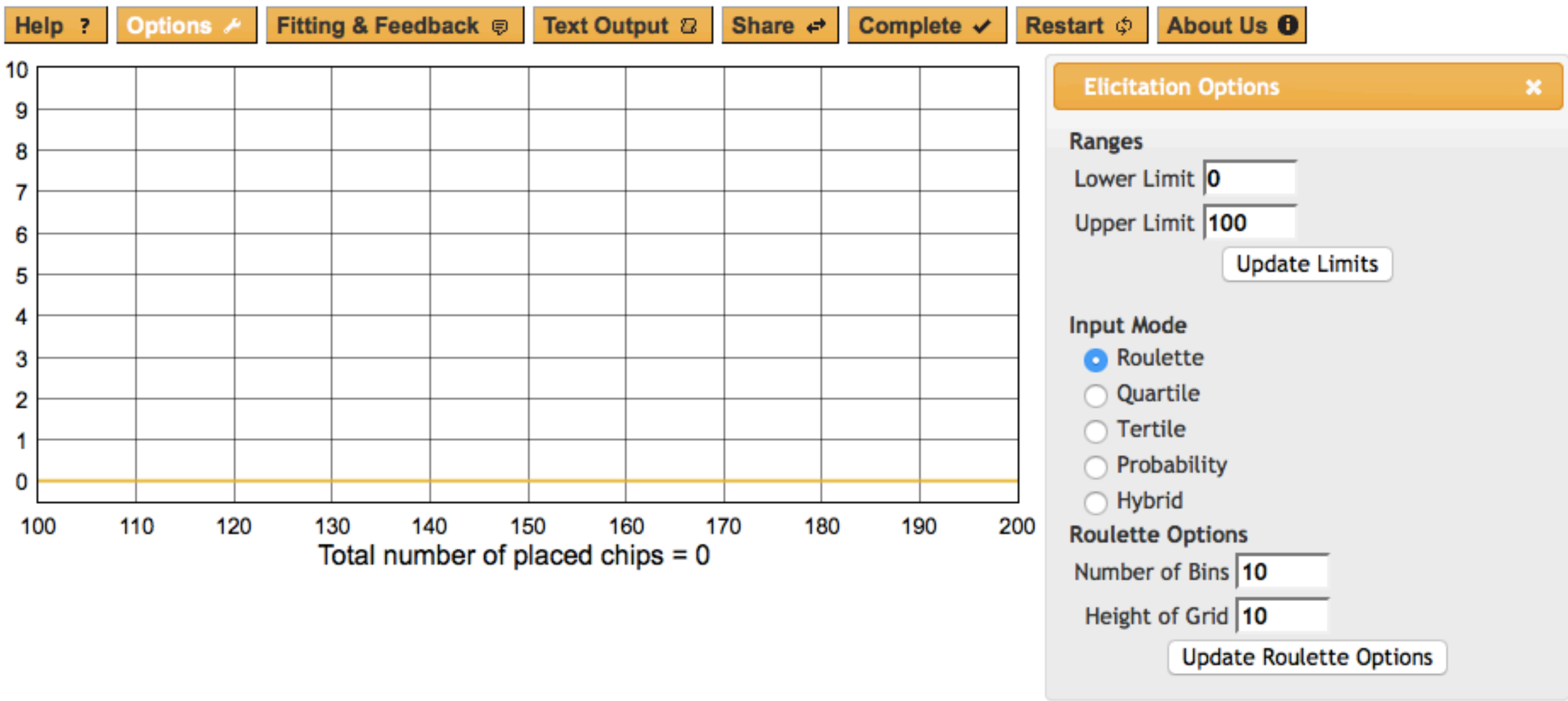
Problems when prior information is wrong and when one has a strong confidence in it.

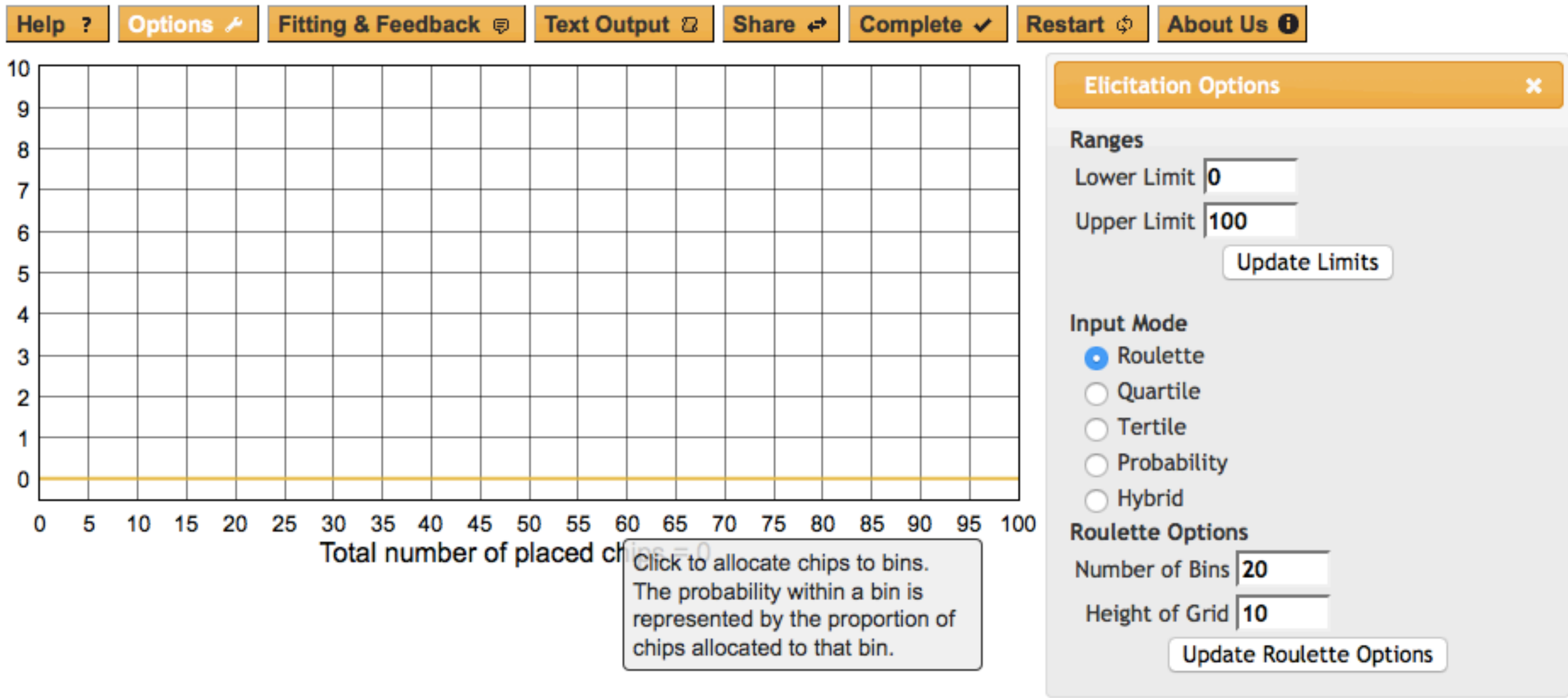
Example 2

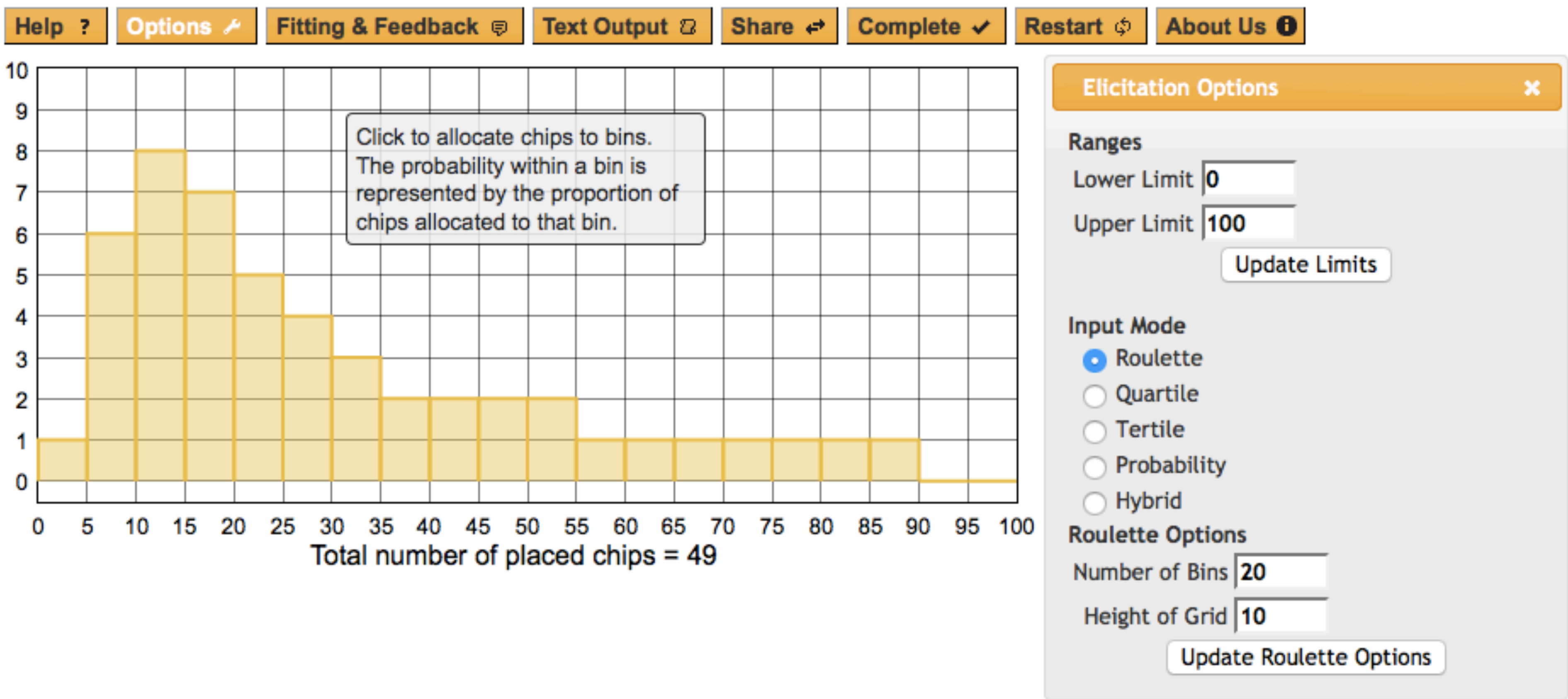
Estimation of a disease incidence θ by combining a measurement with expert knowledge.



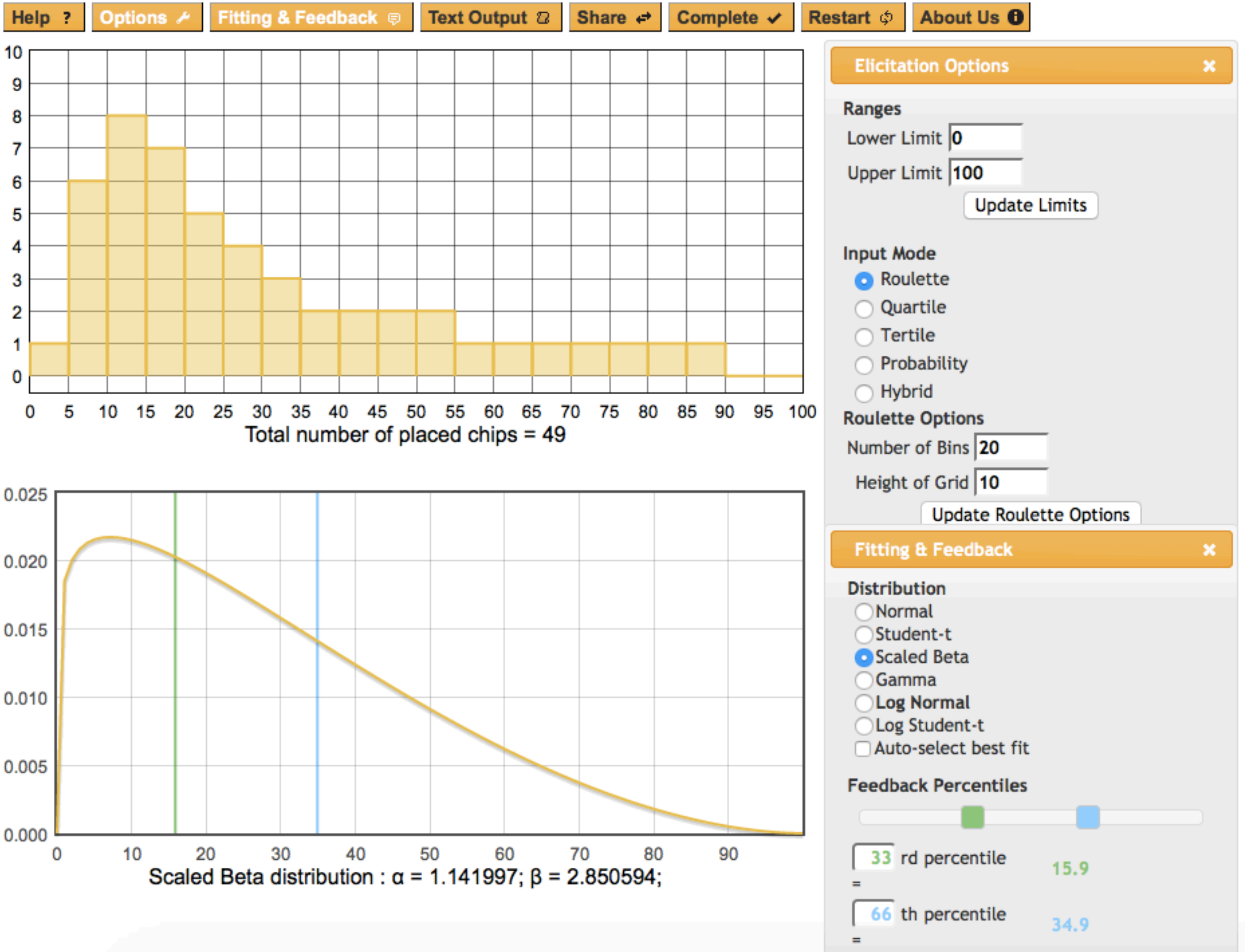
Expert elicitation




Help ?
Options
Fitting & Feedback
Text Output
Share
Complete
Restart
About Us
10
9
8
7
6
5
4
3
2
1
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Total number of placed ch
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 0
Upper Limit 100
Update Limits
Input Mode
Roulette
Quartile
Tertile
Probability
Hybrid
Roulette Options
Number of Bins 20
Height of Grid 10
Update Roulette Options



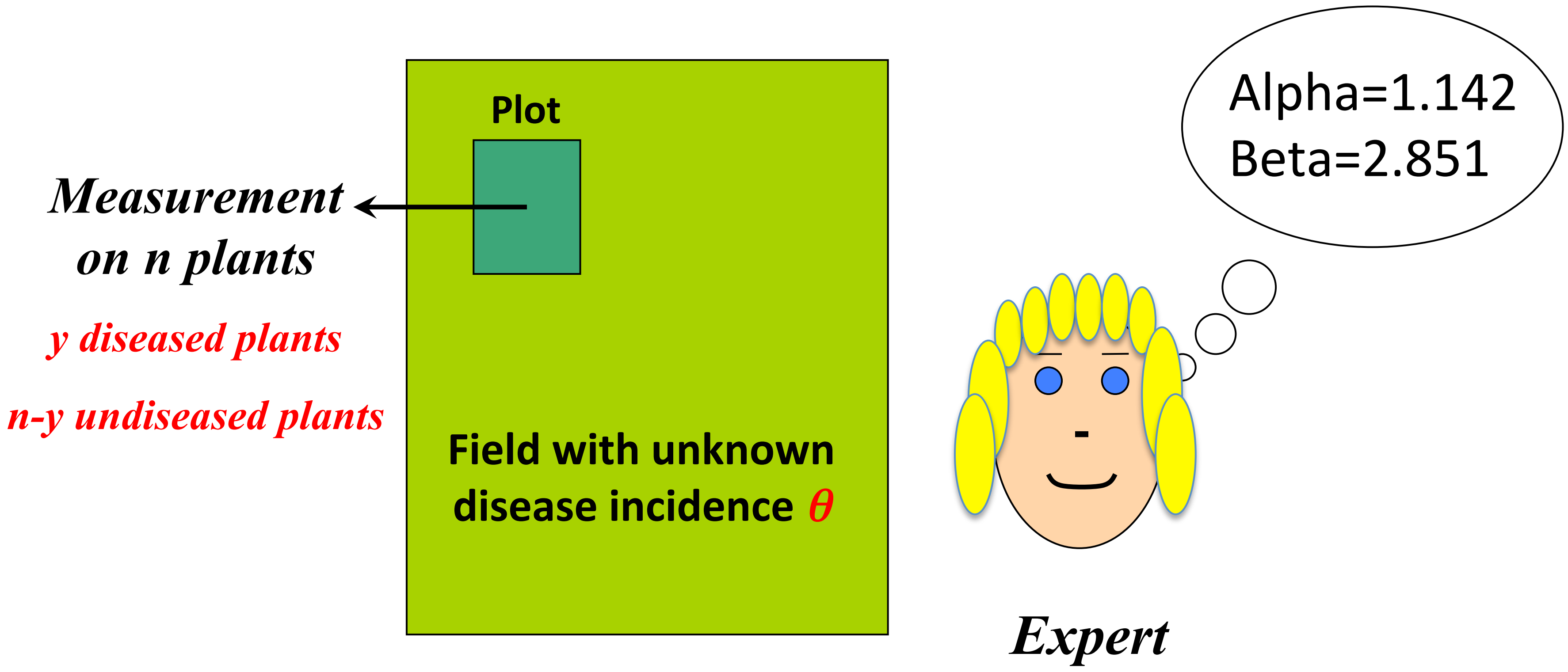
Help ?
Options
Fitting & Feedback
Text Output
Share
Complete
Restart
About Us
Click to allocate chips to bins.
The probability within a bin is
represented by the proportion of
chips allocated to that bin.
0 1 2 3 4 5 6 7 8 9 10
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Total number of placed chips = 49
Elicitation Options
Ranges
Lower Limit 0
Upper Limit 100
Update Limits
Input Mode
Roulette
Quartile
Tertile
Probability
Hybrid
Roulette Options
Number of Bins 20
Height of Grid 10
Update Roulette Options

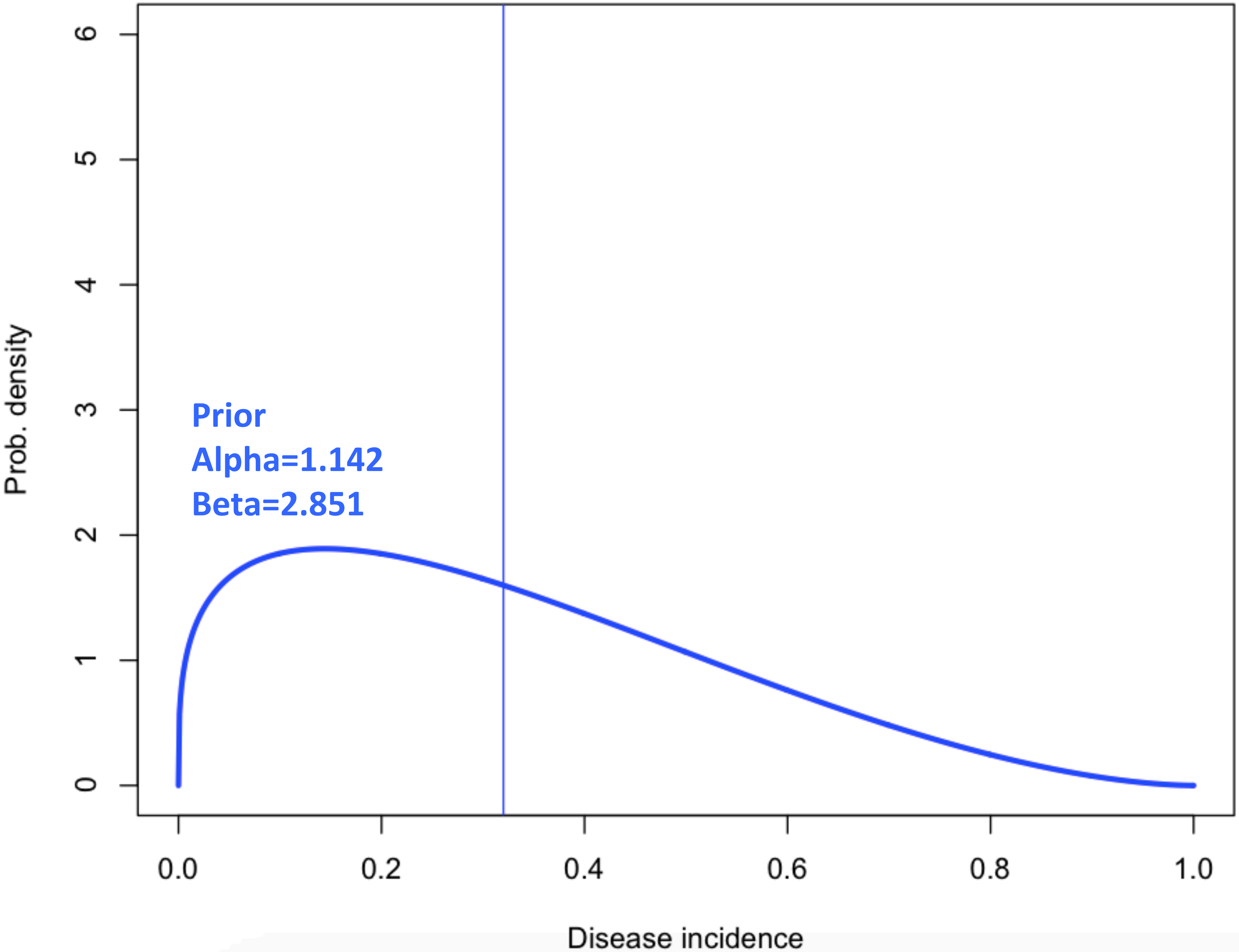


Help ?
Options
Fitting & Feedback
Text Output
Share
Complete
Restart
About Us
Total number of placed chips = 49
Scaled Beta distribution : α = 1.141997; β = 2.850594;
Elicitation Options
Ranges
Lower Limit 0
Upper Limit 100
Update Limits
Input Mode
Roulette
Quartile
Tertile
Probability
Hybrid
Roulette Options
Number of Bins 20
Height of Grid 10
Update Roulette Options
Fitting & Feedback
Distribution
Normal
Student-t
Scaled Beta
Gamma
Log Normal
Log Student-t
Auto-select best fit
Feedback Percentiles
33 rd percentile = 15.9
66 th percentile = 34.9

Example 2

Estimation of a disease incidence θ by combining a measurement with expert knowledge.

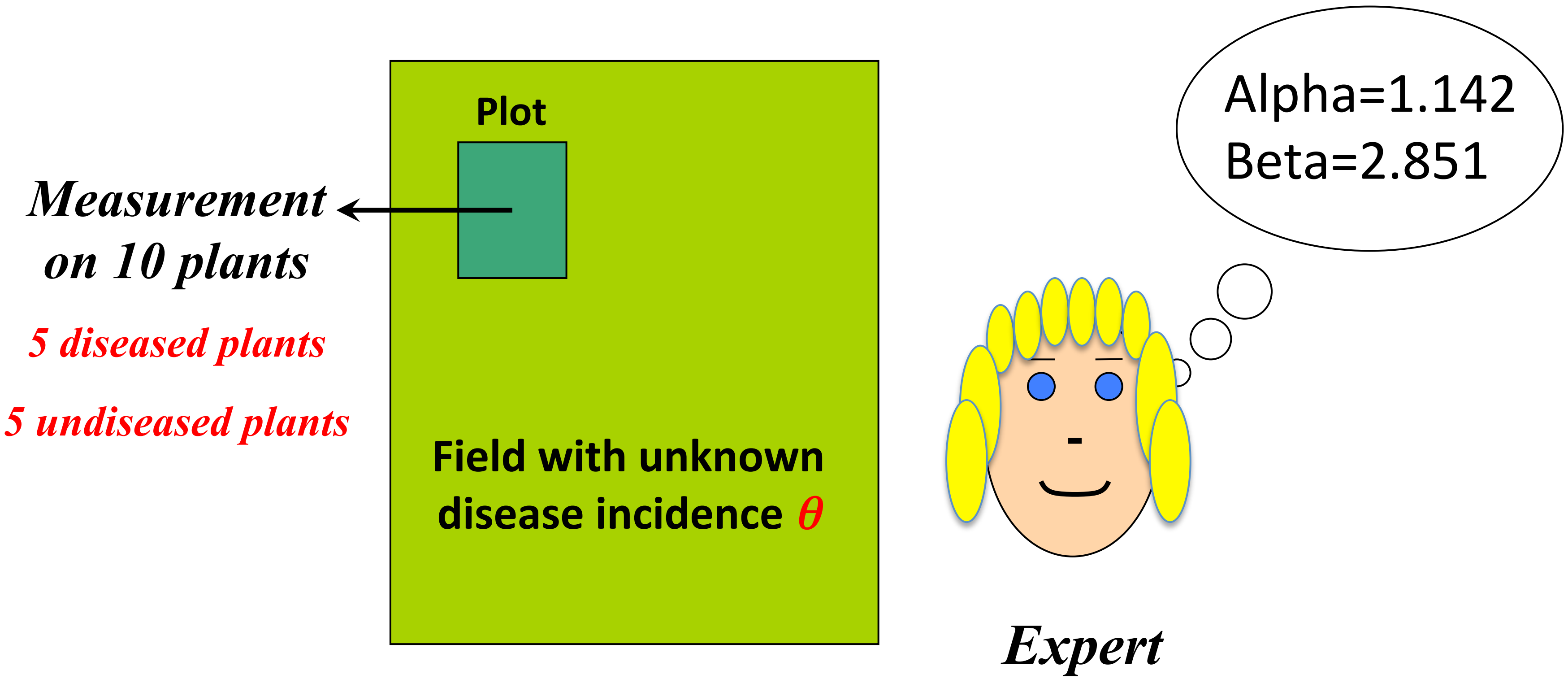


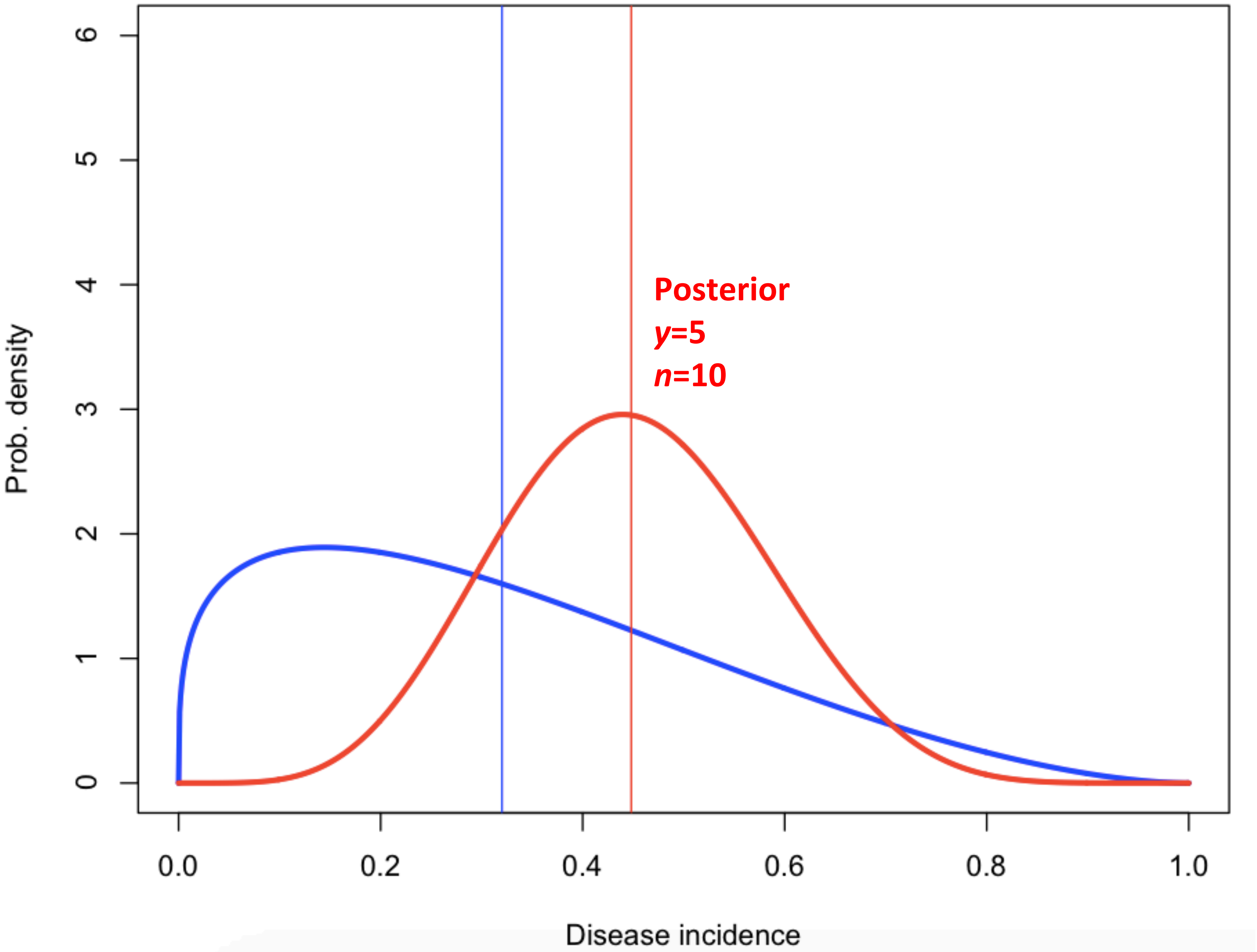


Prior
Alpha=1.142
Beta=2.851
Prob. density
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
Disease incidence

Example 2

Estimation of a disease incidence θ by combining a measurement with expert knowledge.

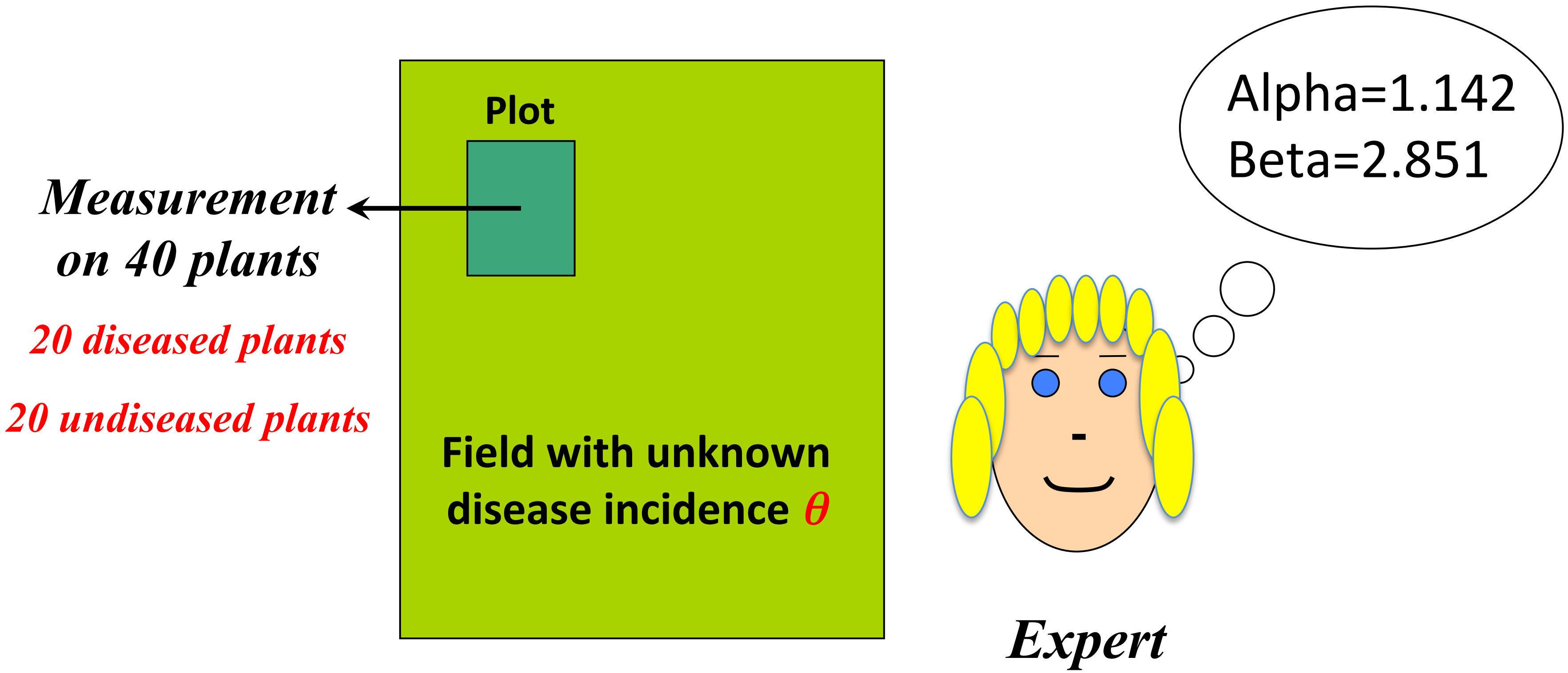


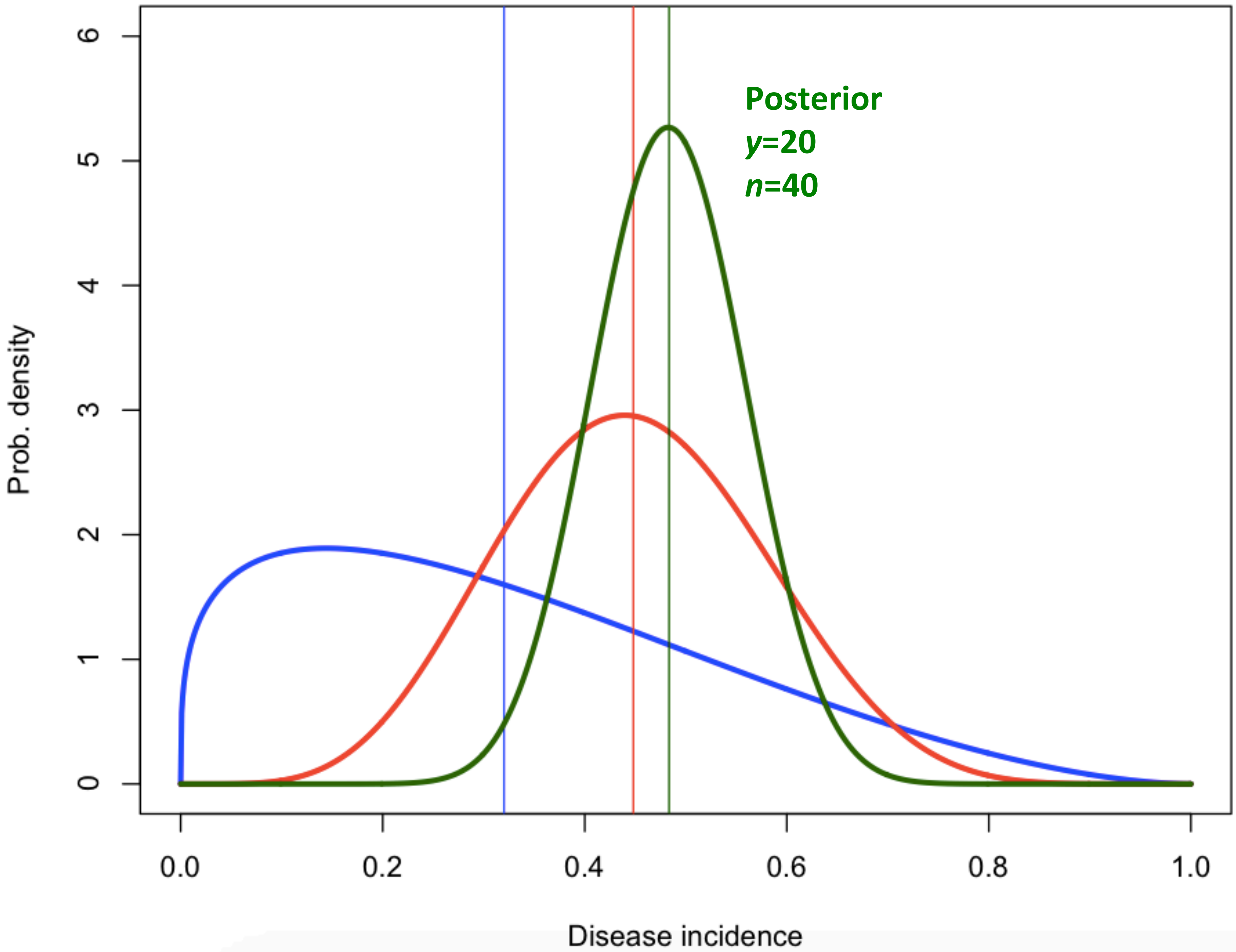


Posterior
y=5
n=10
Prob. density
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
Disease incidence

Example 2

Estimation of a disease incidence θ by combining a measurement with expert knowledge.




Posterior
y=20
n=40
Prob. density
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
Disease incidence

Practical considerations

• The analytical expression of the posterior distribution can be derived for simple applications.

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

Agriculture, Ecosystems and Environment 133 (2009) 208–222

Contents lists available at ScienceDirect

Agriculture, Ecosystems and Environment

journal homepage: www.elsevier.com/locate/agee

Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model

S. Lehuger [a,*], B. Gabrielle [a], M. van Oijen [b], D. Makowski [c], J.-C. Germon [d], T. Morvan [e], C. Hénault [d]

$$N_2O = r\,\mathrm{Da} + c\,\mathrm{Ni}$$

$$\mathrm{Da} = \mathrm{PDR}\ \ F_N\ \ F_W\ \ F_T$$

$$\mathrm{Ni} = \mathrm{MNR}\ \ N_N\ \ N_W\ \ N_T$$

The response functions are unitless and read:

$$F_N = \frac{[NO_3^-]}{Km_{denit} + [NO_3^-]} \tag{7}$$

where F_N is the denitrification response factor to $[NO_3^-]$ the soil nitrate content ($mg\ N\ kg^{-1}$ soil), and Km_{denit} the half-saturation constant ($mg\ N\ kg^{-1}$ soil)

$$F_W = 0,\ WFPS < Tr_{WFPS}$$
$$F_W = \left[\frac{WFPS - Tr_{WFPS}}{1 - Tr_{WFPS}}\right]^{POW},\ WFPS \geq Tr_{WFPS} \tag{8}$$

where F_W is the denitrification response factor to soil WFPS, Tr_{WFPS} is a threshold value below which no denitrification occurs and POW is the exponent of the power law

$$F_T = \exp\left[\frac{(T - TTr_{denit})\ln(Q10_{denit,1}) - 9\ln(Q10_{denit,2})}{10}\right],\ T < TTr_{denit}$$
$$F_T = \exp\left[\frac{(T - 20)\ln(Q10_{denit,2})}{10}\right],\ T \geq TTr_{denit} \tag{9}$$

where F_T is the denitrification response function to soil temperature (T, °C), in the form of two sequential $Q10$ functions below and above a threshold temperature (TTr_{denit}). The two $Q10$ values ($Q10_{denit,1}$ and $Q10_{denit,2}$) correspond to the relative increase in denitrification activity for every 10° C increase in T

$$N_N = \frac{[NH_4^+]}{Km_{int} * Hp + [NH_4^+]} \tag{10}$$

Prior

Parameter vector $\theta = [\theta_1 \cdots \theta_{11}]$				Default value	Prior probability distribution		
θ_i	Symbol	Description	Unit		θ_{min} (i)	θ_{max} (i)	References
θ_1	Tr_{WFPS}	WFPS threshold for denitrification	%	0.62	0.40	0.80	Gabrielle (2006), Hénault et al. (200 Hénault and Germon (2000), Johnss
θ_2	Km_{denit}	Half-saturation constant (denit)	$mg\,N\,kg^{-1}$ soil	22.00	5.00	120.00	Gabrielle (2006), Ding et al. (2007) Parton et al. (2001), Del Grosso et a Parton et al. (1996), Bateman and B Johnsson et al. (2004)
θ_3	TTr_{denit}	Temperature threshold	°C	11.00	10.00	15.00	Gabrielle (2006), Johnsson et al. (20 Renault et al. (1994)
θ_4	$Q10_{denit,1}$	Q10 factor for low temperature	Unitless	89.00	60.00	120.00	Stanford et al. (1975), Maag and Vin
θ_5	$Q10_{denit,2}$	Q10 factor for high temperature	Unitless	2.10	1.00	4.80	Gabrielle (2006), Stanford et al. (19
θ_6	POW_{denit}	Exponent of power function	Unitless	1.74	0.00	2.00	Stanford et al. (1975), Smith et al. (Johnsson et al. (2004), Maag and Vi Maag and Vinther (1996), Skopp et
θ_7	OPT_{WFPS}	Optimum WFPS for nitrification	%	0.60	0.35	0.75	Jambert et al. (1997), Laville et al. (
θ_8	MIN_{WFPS}	Minimum WFPS for nitrification	%	0.10	0.05	0.15	Linn and Doran (1984), Jambert et a Skopp et al. (1990), Ding et al. (200 Parton et al. (2001), Bateman and B
θ_9	MAX_{WFPS}	Maximum WFPS for nitrification	%	0.80	0.80	1.00	Linn and Doran (1984), Parton et al. Bateman and Baggs (2005)
θ_{10}	Km_{int}	Half-saturation constant (nit)	$mg\,N\,kg^{-1}$ soil	10.00	1.00	50.00	Linn and Doran (1984), Jambert et a Pihlatie et al. (2004)
θ_{11}	$Q10_{int}$	Q10 factor for nitrification	Unitless	2.10	1.90	13.00	Maag and Vinther (1996), Laville et Smith (1997), Dobbie and Smith (20

Data

Site	Treatment	Year	Soil texture class	Crop type	N fertiliser (kg N ha^{-1})	Number of observations	Source
Rafidin	N0	1994–1995	Rendzina	Rapeseed	0	7	Gosse et al. (1999)
	N1	1994–1995	Rendzina	Rapeseed	155	8	Gosse et al. (1999)
	N2	1994–1995	Rendzina	Rapeseed	262	9	Gosse et al. (1999)
Villamblain		1998–1999	Loamy Clay	Winter Wheat	230	15	Hénault et al. (2005)
Arrou		1998–1999	Loamy Clay	Winter Wheat	180	18	Hénault et al. (2005)
La Saussaye		1998–1999	Clay Loams	Winter Wheat	200	14	Hénault et al. (2005)
Champnoël	CT	2002–2003	Silt Loam	Maize	0	15	Dambreville et al. (2008)
	AN	2002–2003	Silt Loam	Maize	110	23	Dambreville et al. (2008)
Le Rheu	CT	2004–2005	Silt Loam	Maize	18	24	Dambreville et al. (2008)
	AN	2004–2005	Silt Loam	Maize	180	22	Dambreville et al. (2008)
Grignon		2005	Silt Loam	Maize	140	31	Lehuger et al. (2007)

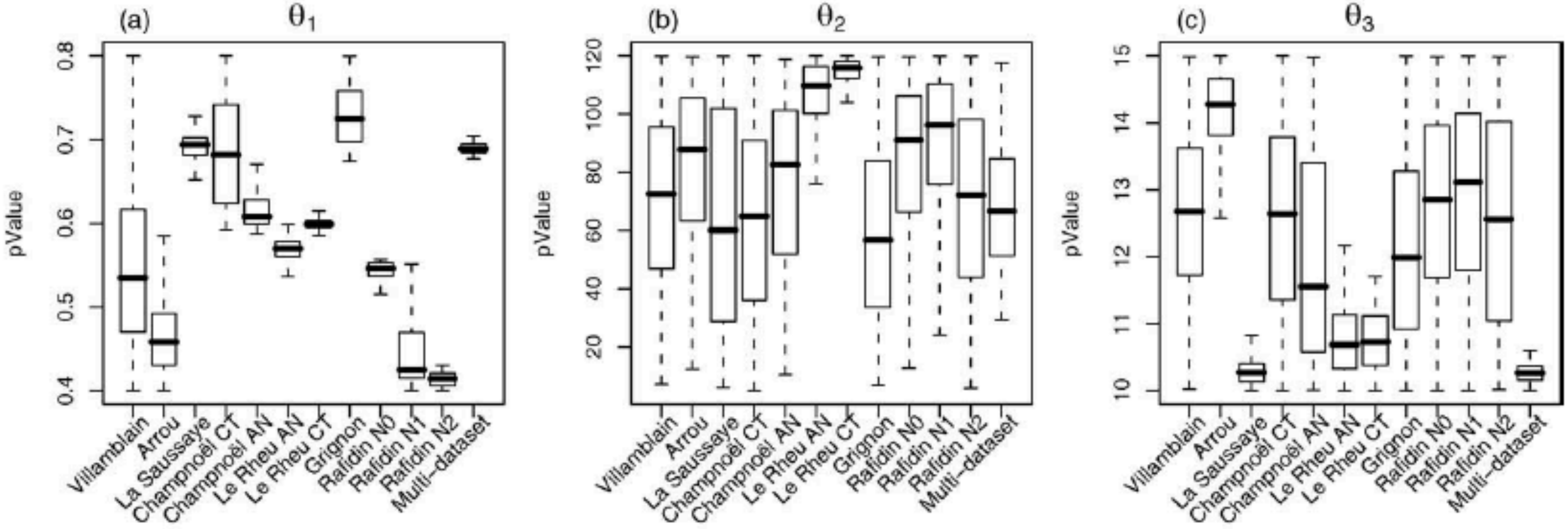
Likelihood

$$\log L_i = \sum_{j=1}^{K} \left(-0.5 \left(\frac{y_j - f(\omega_i; \theta_i)}{\sigma_j} \right)^2 - 0.5 \log(2\pi) - \log(\sigma_j) \right)$$

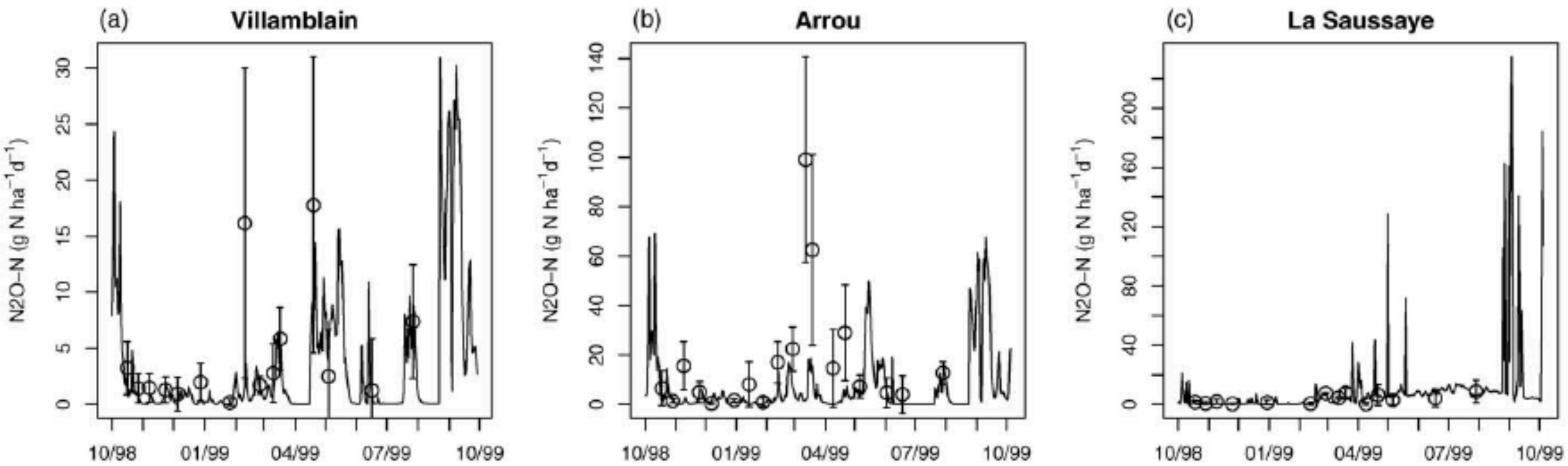
Algorithm

- Metropolis-Hastings implemented in R
- Three chains of 10000 iterations
- Crop model in Fortran encapsulated within R

Posterior parameter distributions



Posterior means of the outputs



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

- Strong decrease of computational time

- Dedicated softwares (BUGS, R packages etc.)

→ **It is now possible to apply Bayesian techniques to complex problems**