# Chapter 2
# Bayesian Inference

Xavi Puig
Departament d'Estadística i I.O.


## Chapter 2. Bayesian Inference

Summarizing……

Bayesiaa Model

$$\boxed{\begin{array}{l} \tilde{y} \mid \theta \sim p(\tilde{y} \mid \theta) \\ \theta \sim \pi(\theta) \end{array}} \longrightarrow p(\tilde{y})$$

$\Downarrow$

**Data** $\Rightarrow$ **Th. Bayes**

$\Downarrow$

Bayesian Model (updated)

$$\boxed{\begin{array}{l} \tilde{y} \mid \theta \sim p(\tilde{y} \mid \theta) \\ \theta \mid y \sim \pi(\theta \mid y) \end{array}} \longrightarrow p(\tilde{y} \mid y)$$

The updated model allows you to calculate point estimations, interval estimations, predictions…

## Chapter 2. Bayesian Inference

After collecting the data and checking the model, Statistical Inference tries to guess the true parameter value, $\theta^*$, that is, it tries to guess the model that generated the observed data,

$$p(\tilde{y} \mid \theta^*) \in M$$

Statistical Inference mainly consists of:

a) Point Estimation

b) Interval Estimation

c) Prediction

d) Hypothesis Test

## Chapter 2. Bayesian Inference

**2.1 Posterior distribution as an estimator**

2.2 Point Estimation

2.3 Interval Estimation

2.4 Prediction

2.5 Hypothesis Test

## 2.1 Posterior Distribution as an Estimator

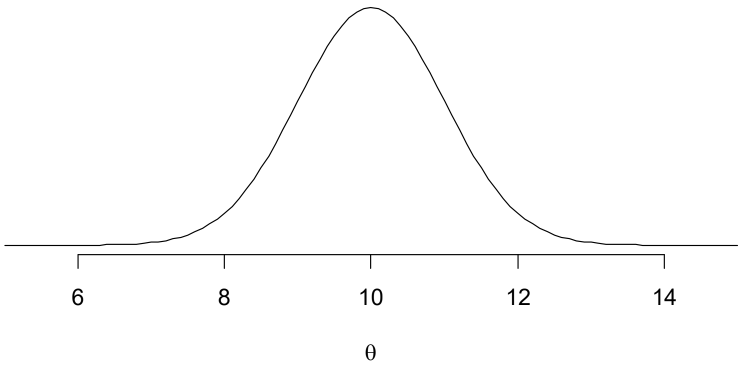
The posterior distribution, $\pi(\theta|y)$, has all the information about the parameter after observing the data,

$$\pi(\theta \mid y) = \frac{p(y \mid \theta)\pi(\theta)}{p(y)} \propto L_y(\theta)\pi(\theta)$$

The posterior distribution is a compromise between the prior distribution (the information before observing the data) and the likelihood function (the data information).

## 2.1 Posterior Distribution as an Estimator

The posterior distribution, $\pi(\theta|y)$, is the natural Bayesian estimator for $\theta$

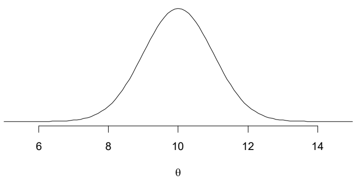


## Chapter 2. Bayesian Inference



## 2.2 Point Estimation

A Point estimator summarizes $\pi(\theta|y)$




It consists of choosing $\hat{\theta}$ in such a way that it is the good approximation to $\theta^*$

## 2.2 Point Estimation

Any measure of the location of $\pi(\theta|y)$ will serve as a point estimate:

- $\widehat{\theta}_{pe} = E(\theta \mid y) = \int \theta \pi(\theta \mid y) \partial\theta$
- $\widehat{\theta}_{pme}$ is such that $\int_{-\infty}^{\widehat{\theta}_{pme}} \pi(\theta \mid y) \partial\theta = 0.5$
- $\widehat{\theta}_{pmo}$ is such that it maximizes $\pi(\theta \mid y)$

## 2.2 Point Estimation

Observation

We can get the point estimates:

a) analytically

b) by simulation

## 2.2 Point Estimation

Example: The posterior expected value

a) Analytically

$$\hat{\theta}_{pe} = E(\theta \mid y) = \int \theta \pi(\theta \mid y) \partial\theta$$

b) By simulation

Be $\theta^{(1)},\ldots,\theta^{(M)}$ simulations of $\pi(\theta \mid y)$ then

$$\hat{\theta}_{pe} \approx \frac{\sum_{j=1}^{M} \theta^{(j)}}{M}$$

The larger $M$ (the number of simulations), the better the approximation

## Chapter 2. Bayesian Inference

2.1 Posterior distribution as an estimator

2.2 Point Estimation

**2.3 Interval Estimation**

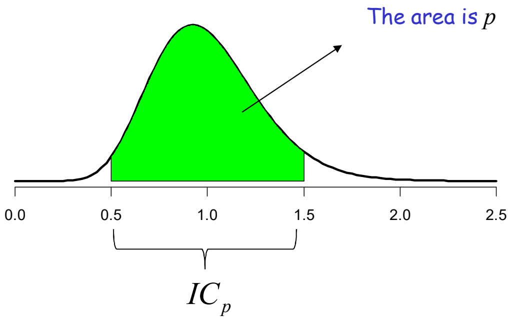
2.4 Prediction

2.5 Hypothesis Test

## 2.3 Interval Estimation

A posterior credibility (or probability ) interval of $p$ for $\theta$ , $IC_p$, is any region in $\Omega$ in such a way that

$$p(\theta \in IC_p \mid y) = \int_{IC_p} \pi(\theta \mid y) \partial\theta = p$$

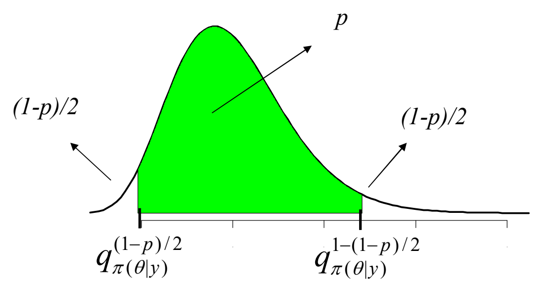



Observation: we can define in an similar way a prior $IC_p$ using the prior distribution instead of the posterior distribution.

## 2.3 Interval Estimation

The main types of credible intervals are the:

**Intervals based on percentiles**



## 2.3 Interval Estimation

Observation

The Credible Intervals based on percentiles can be calculated in two ways:

a) Analytically

b) By simulation

## 2.3 Interval Estimation

Example: A 95% Credible Interval for θ

a) Analytically

$$\int_{-\infty}^{q^{0.025}_{\pi(\theta|y)}} \pi(\theta \mid y)\partial\theta = 0.025$$

$$\int_{q^{0.975}_{\pi(\theta|y)}}^{\infty} \pi(\theta \mid y)\partial\theta = 0.025$$

The larger $M$ (the number of simulations) the better the approximation

b) By simulation

Be $\theta^{(1)},\ldots,\theta^{(M)}$ simulations of $\pi(\theta \mid y)$ then

$$\left(q^{0.025}_{\theta^{(1)},\ldots,\theta^{(M)}},\ q^{0.975}_{\theta^{(1)},\ldots\theta^{(M)}}\right)$$

percentile 2.5% percentile 97.5%

## 2.3 Interval Estimation

Observation

We can do inference about $g(\theta)=\psi$ instead of $\theta$:

a) Getting $\pi(\psi|y)$ analytically, or

b) Approximating $\pi(\psi|y)$ using simulations

---

a) Getting $\pi(\psi|y)$ analytically:

$$\pi_{\psi}(\psi \mid y) = \pi_{\theta}(g^{-1}(\psi) \mid y)\left|\frac{\partial g^{-1}(\psi)}{\partial \psi}\right|$$

## 2.3 Interval Estimation

b) Approximating $\pi(\psi|y)$ by simulation

To approximate $\pi(\psi|y)$ by simulation, first of all, we must set the number of simulations, $M$, (the larger the better) then from $j$=1 to $M$:

1. Simulate $\theta^{(j)}$ from $\pi(\theta \mid y)$
2. calculate $\psi^{(j)} = g(\theta^{(j)})$

These $\psi^{(1)},\ldots,\psi^{(M)}$ simulated values are simulations from the $\pi(\psi|y)$. Hence, using these simulations we can calculate everything we want/need (moments, probabilities, etc.). And by graphing it we can reproduce the shape of its probability distribution.

## 2.3 Interval Estimation

Now, using these simulated values: $\psi^{(1)},\ldots,\psi^{(M)}$

A point estimator for $\psi$ could be:

$$\hat{\psi} = E(\psi \mid y) \approx \frac{\sum_{j=1}^{M} \psi^{(j)}}{M}$$

And, a 95% credible interval for $\psi$ could be:

$$\left(q^{0.025}_{\psi^{(1)},\ldots,\psi^{(M)}},\ q^{0.975}_{\psi^{(1)},\ldots,\psi^{(M)}}\right)$$

percentile 2.5%

percentile 97.5%

## Chapter 2. Bayesian Inference

## 2.4 Prediction

In a similar way the posterior distribution,

$\pi(\theta \mid y)$ represents all that we know about the parameter

The posterior predictive distribution,

$p(\tilde{y} \mid y)$ represents all that we know about future values for $\tilde{y}$

## 2.4 Prediction

The posterior predictive distribution can be calculated in two ways:

a) Analytically

b) By simulation

---

a) Analytically:

$$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \theta)\pi(\theta \mid y)d\theta$$

## 2.4 Prediction

To approximate the posterior predictive distribution using simulations, first of all, you must set the number of simulations, *M* (the larger the better), then for *j*=1 to *M*:

1. Simulate

$$\theta^{(j)} \text{ de } \pi(\theta \mid y), \text{ and}$$

2. Simulate

$$\tilde{y}^{(j)} \text{ de } p(\tilde{y} \mid \theta^{(j)})$$

These $\tilde{y}^{(1)},\ldots,\tilde{y}^{(M)}$ simulated values are simulations from the posterior predictive distribution. Hence, using these simulations we can calculate everything we want/need (moments, probabilities, etc.). And, by graphing it we can reproduce the shape of its probability distribution.

## 2.4 Prediction

We can also calculate point estimates and credibility intervals for the predictions. We can calculate them analytically or by simulation.

Example:

a) Analytically:

- $E(\tilde{y} \mid y) = \int \tilde{y} p(\tilde{y} \mid y) \partial \tilde{y} = \int \tilde{y} \int p(\tilde{y} \mid \theta) \pi(\theta \mid y) d\theta \partial \tilde{y}$

- $CI_{95\%} = (q^{0.025}_{p(\tilde{y}|y)}, q^{0.975}_{p(\tilde{y}|y)})$

  Where $q^{0.025}_{p(\tilde{y}|y)}$ $q^{0.975}_{p(\tilde{y}|y)}$ are such that,

$$\int_{-\infty}^{q^{0.025}_{p(\tilde{y}|y)}} p(\tilde{y} \mid y) \partial \theta = 0.025, \quad \int_{q^{0.975}_{p(\tilde{y}|y)}}^{\infty} \pi(\theta \mid y) \partial \theta = 0.025$$

## 2.4 Prediction

Example: point estimate and credibility interval for a prediction

b) By simulation

Be $\tilde{y}^{(1)},\ldots,\tilde{y}^{(M)}$ simulations from the posterior predictive distribution

- $$E(\tilde{y} \mid y) \approx \frac{\sum_{j=1}^{M} \tilde{y}^{(j)}}{M}$$

- $$IC_{95\%} = \left(q^{0.025}_{\tilde{y}^{(1)},\ldots,\tilde{y}^{(M)}},\ q^{0.975}_{\tilde{y}^{(1)},\ldots,\tilde{y}^{(M)}}\right)$$

percentile 2.5% percentile 97.5%

## Chapter 2. Bayesian Inference

## 2.5 Hypothesis Test

### Example 1: the main idea

Given a Bayesian Model: $M = \{p(\tilde{y} \mid \theta), \theta \in \Omega\}, \quad \pi(\theta)$

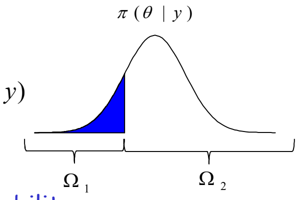
We split the parameter space in two: $\Omega = \Omega_1 \cup \Omega_2$

We want to decide to which subspace the parameter belongs

$$H_1 : \theta \in \Omega_1$$
$$H_2 : \theta \in \Omega_2$$

After observing the data, we update the Bayesian Model and compute the posterior distribution for each hypothesis as:

$$p(H_1 \mid y) = p(\theta \in \Omega_1 \mid y) = \int_{\Omega_1} \pi(\theta \mid y) d\theta$$
$$p(H_2 \mid y) = p(\theta \in \Omega_2 \mid y) = \int_{\Omega_2} \pi(\theta \mid y) d\theta = 1 - p(H_1 \mid y)$$




We choose the hypothesis with the highest probability

## 2.5 Hypothesis Test

### Example 2: Two possible values

We have a manipulated coin, with a probability of 0.7 of getting head or tail (this is what we want to know). We will toss the coin 10 times and count the number of observed heads.

$$M = \{p(\tilde{y} \mid \theta) = Binomial(n = 10, \theta), \theta \in \{0.3, 0.7\}\}$$
$$M = \{Binomial(n = 10, \theta = 0.3), Binomial(n = 10, \theta = 0.7)\}$$

$$H_1 : \theta = 0.3 \qquad \pi(\theta) = \begin{cases} 0.5 & si \quad \theta = 0.3 \\ 0.5 & si \quad \theta = 0.7 \end{cases}$$
$$H_2 : \theta = 0.7$$

We toss the coin $n=10$ times and observe $y$=7 heads.

Now we have to calculate the posterior probability of each hypothesis.

## 2.5 Hypothesis Test

### Example 2 (continued): Two possible values

$$M = \{p(\tilde{y} \mid \theta) = Binomial(n = 10, \theta), \theta \in \{0.3, 0.7\}\} \qquad \pi(\theta) = \begin{cases} 0.5 & si \quad \theta = 0.3 \\ 0.5 & si \quad \theta = 0.7 \end{cases}$$

$$H_1 : \theta = 0.3$$
$$H_2 : \theta = 0.7$$

We toss the coin $n=10$ vtimes and observe $y=7$ heads.

$$p(H_1 \mid y) = \frac{p(H_1)p(y \mid H_1)}{p(y)} = \frac{p(H_1)p(y \mid H_1)}{p(H_1)p(y \mid H_1) + p(H_2)p(y \mid H_2)} =$$

$$= \frac{0.5 \times \binom{10}{7} 0.3^7 (1-0.3)^{10-7}}{0.5 \times \binom{10}{7} 0.3^7 (1-0.3)^{10-7} + 0.5 \times \binom{10}{7} 0.7^7 (1-0.7)^{10-7}} = 0.033$$

$$p(H_2 \mid y) = 1 - p(H_1 \mid y) = 0.967$$

We choose $H_2$

## 2.5 Hypothesis Test

### Example 3: Three hypotheses

The time needed for a specific radioactive particle to disintegrate follows an *Exponencial* Model. Physicists agree to use a *Gamma*($a$=10, $b$=10) as a prior distribution.

The Bayesian model is:

$$\tilde{y} \mid \lambda \sim \exp(\lambda)$$
$$\lambda \sim Gamma(10, 10)$$

$$p(y|\lambda) = \lambda e^{-\lambda y}$$

$$\pi(\lambda) = \frac{b^a \lambda^{(a-1)} e^{-b\lambda}}{\Gamma(a)}$$

To avoid errors related to the parametrizations, it is a good idea to show the expressions for the distributions.

We want to chose among:

$$H_1 : \ \lambda \in (0, 0.5)$$
$$H_2 : \ \lambda \in [0.5, 1.5)$$
$$H_3 : \ \lambda \in [1.5, \infty)$$

## 2.5 Hypothesis Test

### Example 3 (continued): Three hypotheses

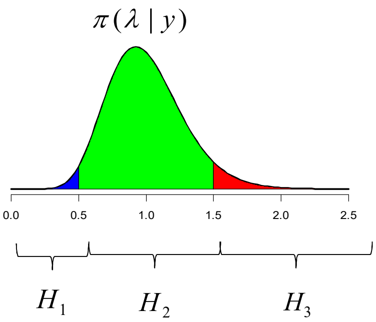
The observed data is: 0.9, 1.1 and 1.

The posterior distribution is: $\lambda \mid y \sim Gamma(13, 13)$

We calculate the posterior distribution for each hypothesis as follows:

$H_1: \ \lambda \in (0, 0.5)$
$H_2: \ \lambda \in [0.5, 1.5)$
$H_3: \ \lambda \in [1.5, \infty)$

$$p(H_1 \mid y) = \int_0^{0.5} \pi(\lambda \mid y) d\lambda = 0.016$$
$$p(H_2 \mid y) = \int_{0.5}^{1.5} \pi(\lambda \mid y) d\lambda = 0.935$$
$$p(H_3 \mid y) = \int_{1.5}^{\infty} \pi(\lambda \mid y) d\lambda = 0.049$$




We choose $H_2$

## Chapter 2. Bayesian Inference

2.1 Posterior distribution as an estimator
2.2 Point Estimation
2.3 Interval Estimation
2.4 Prediction
2.5 Hypothesis Test

# Chapter 2
# Bayesian Inference

Xavi Puig Oriol
Departament d'Estadística i I.O.