Introduction to Bayesian statistics

- Example (Thomas Bayes, 1763):
  - A billiard ball is dropped on the interval $[0, 1]$
    - it stops at p
    - assume p is uniformly distributed on $[0, 1]$
  - Drop the billiard ball n new times
    - record $y_i = 1$ if ball stops to the left of p
    - $y_i = 0$ otherwise
    - set $x = \sum_{i=1}^{n} y_i$

Introduction to Bayesian statistics

- Example (Thomas Bayes, 1763):
  - A billiard ball is dropped on the interval $[0,1]$
    - it stops at p
    - assume p is uniformly distributed on $[0,1]$
  - Drop the billiard ball n new times
    - record $y_i = 1$ if ball stops to the left of p
    - $y_i = 0$ otherwise
    - set $x = \sum_{i=1}^{n} y_i$
    - thus $x|p \sim \text{bin}(n,p)$,

$$\mathrm{P}(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x} \ , \ x = 0, 1, \ldots, n$$

Introduction to Bayesian statistics

- Example (Thomas Bayes, 1763):
  - A billiard ball is dropped on the interval $[0, 1]$
    - it stops at p
    - assume p is uniformly distributed on $[0, 1]$
  - Drop the billiard ball n new times
    - record $y_i = 1$ if ball stops to the left of p
    - $y_i = 0$ otherwise
    - set $x = \sum_{i=1}^{n} y_i$
    - thus $x|p \sim \text{bin}(n, p)$,

$$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x} \text{ , } x = 0, 1, \ldots, n$$

  - want to estimate p from observed x
  - standard estimator for p in binomial distr.:

$$\widehat{p} = \frac{X}{n}$$

Introduction to Bayesian statistics

- Example (Thomas Bayes, 1763):
  - A billiard ball is dropped on the interval $[0, 1]$
    - it stops at p
    - assume p is uniformly distributed on $[0, 1]$
  - Drop the billiard ball n new times
    - record $y_i = 1$ if ball stops to the left of p
    - $y_i = 0$ otherwise
    - set $x = \sum_{i=1}^{n} y_i$
    - thus $x|p \sim \text{bin}(n, p)$,

$$P(X = x|p) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} \ , \ x = 0, 1, \ldots, n$$

  - want to estimate p from observed x
  - standard estimator for p in binomial distr.:

$$\widehat{p} = \frac{X}{n}$$

  - but we know $p \sim \text{Uniform}[0, 1]$,

$$f(p) = \begin{cases} 1 & \text{for } x \in [0, 1], \\ 0 & \text{otherwise} \end{cases}$$

Introduction to Bayesian statistics

- Recall

$$f(p) = \begin{cases} 1 & \text{for } x \in [0,1], \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{P}(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x} \ , \ x = 0, 1, \ldots, n$$

Introduction to Bayesian statistics

- Recall

$$f(p) = \begin{cases} 1 & \text{for } x \in [0,1], \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{P}(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x} \;,\; x = 0, 1, \ldots, n$$

- Thus

$$\begin{aligned} f(p|x) &= \frac{f(p,x)}{\mathrm{P}(X=x)} = \frac{f(p)\mathrm{P}(X=x|p)}{\int_0^1 \mathrm{P}(X=x|\tilde{p}) f(\tilde{p}) \mathrm{d}\tilde{p}} \\ &= \frac{p^x(1-p)^{n-x}}{\int_0^1 \tilde{p}^x (1-\tilde{p})^{n-x} \mathrm{d}\tilde{p}} = \frac{p^x(1-p)^{n-x}}{B(x+1, n-x+1)} \end{aligned}$$

Introduction to Bayesian statistics

- Recall

$$f(p) = \begin{cases} 1 & \text{for } x \in [0,1], \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{P}(X = x|p) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} \ , \ x = 0, 1, \ldots, n$$

- Thus

$$\begin{aligned} f(p|x) &= \frac{f(p,x)}{\mathrm{P}(X=x)} = \frac{f(p)\mathrm{P}(X=x|p)}{\int_0^1 \mathrm{P}(X=x|\tilde{p})f(\tilde{p})\mathrm{d}\tilde{p}} \\ &= \frac{p^x(1-p)^{n-x}}{\int_0^1 \tilde{p}^x(1-\tilde{p})^{n-x}\mathrm{d}\tilde{p}} = \frac{p^x(1-p)^{n-x}}{B(x+1, n-x+1)} \end{aligned}$$

- This is a beta-distribution, $\mathcal{B}(x+1, n-x+1)$, with

$$\mathrm{E}[p|x] = \frac{x+1}{n+2}$$

- A natural estimator for p

$$\widehat{p} = \frac{X+1}{n+2}$$

Bayesian statistics

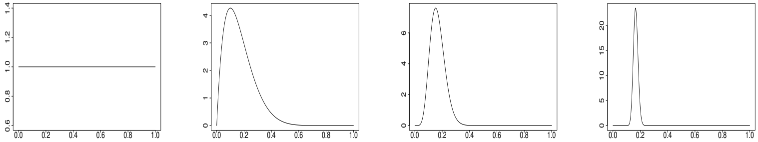
- In previous example: p is a stochastic variable because it is the result of a stochastic experiment
- Bayesian modelling: consider parameters as stochastic variables also when their value is not the result of a stochastic experiment

Bayesian statistics

- In previous example: p is a stochastic variable because it is the result of a stochastic experiment
- Bayesian modelling: consider parameters as stochastic variables also when their value is not the result of a stochastic experiment
- Another (toy) example:
  - I have a dice, let p: probability of getting a six
  - Consider p as a stochastic variable, you don't know it is a proper dice
  - what distribution would you assign to p?

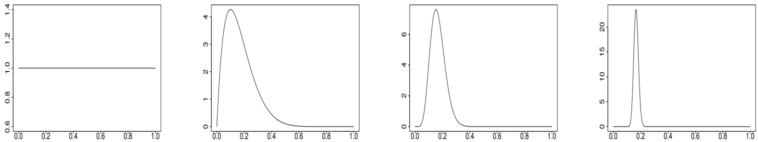
Bayesian statistics

- In previous example: p is a stochastic variable because it is the result of a stochastic experiment
- Bayesian modelling: consider parameters as stochastic variables also when their value is not the result of a stochastic experiment
- Another (toy) example:
  - I have a dice, let p: probability of getting a six
  - Consider p as a stochastic variable, you don't know it is a proper dice
  - what distribution would you assign to p?



Bayesian statistics

- In previous example: p is a stochastic variable because it is the result of a stochastic experiment
- Bayesian modelling: consider parameters as stochastic variables also when their value is not the result of a stochastic experiment
- Another (toy) example:
  - I have a dice, let p: probability of getting a six
  - Consider p as a stochastic variable, you don't know it is a proper dice
  - what distribution would you assign to p?



- we roll the dice n times, let x: number of sixes

$$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x} \text{ , for } x = 0, 1, \ldots, n$$

Bayesian statistics

- Recall
$$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x} \text{ , for } x = 0, 1, \ldots, n$$
- Assume $p \sim \mathcal{B}(\alpha, \beta)$:
$$f(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1}$$
- This gives:
$$f(p|x) = \frac{f(p, x)}{P(X = x)}$$

Bayesian statistics

- Recall
$$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x} \text{ , for } x = 0, 1, \ldots, n$$
- Assume $p \sim \mathcal{B}(\alpha, \beta)$:
$$f(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1}$$
- This gives:
$$f(p|x) = \frac{f(p,x)}{P(X = x)} \propto f(p)P(X = x|p)$$

Bayesian statistics

- Recall
$$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x} \text{ , for } x = 0, 1, \ldots, n$$
- Assume $p \sim \mathcal{B}(\alpha, \beta)$:
$$f(p) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$
- This gives:
$$\begin{aligned} f(p|x) &= \frac{f(p,x)}{P(X=x)} \propto f(p)P(X = x|p) \\ &\propto p^{\alpha-1}(1-p)^{\beta-1} p^x (1-p)^{n-x} \end{aligned}$$

Bayesian statistics

- Recall

$$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x} \text{ , for } x = 0, 1, \ldots, n$$

- Assume $p \sim \mathcal{B}(\alpha, \beta)$:

$$f(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

- This gives:

$$\begin{aligned} f(p|x) &= \frac{f(p,x)}{P(X = x)} \propto f(p)P(X = x|p) \\ &\propto p^{\alpha-1}(1-p)^{\beta-1} p^x (1-p)^{n-x} \\ &= p^{\alpha+x-1}(1-p)^{\beta+n-x-1} \end{aligned}$$

Bayesian statistics

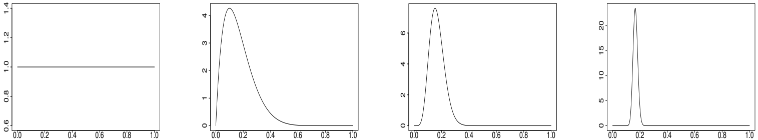
- Recall
$$P(X = x|p) = \binom{n}{x} p^x(1-p)^{n-x} \text{ , for } x = 0, 1, \ldots, n$$
- Assume $p \sim \mathcal{B}(\alpha, \beta)$:
$$f(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1}$$
- This gives:
$$\begin{aligned} f(p|x) &= \frac{f(p,x)}{P(X = x)} \propto f(p)P(X = x|p) \\ &\propto p^{\alpha-1}(1-p)^{\beta-1}p^x(1-p)^{n-x} \\ &= p^{\alpha+x-1}(1-p)^{\beta+n-x-1} \end{aligned}$$
- Thus $p|x \sim \mathcal{B}(\alpha + x, \beta + n - x)$, so
$$E[p|x] = \frac{\alpha + x}{\alpha + \beta + n}$$

Bayesian statistics

- Recall
$$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x} \text{ , for } x = 0, 1, \ldots, n$$
- Assume $p \sim \mathcal{B}(\alpha, \beta)$:
$$f(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}$$
- This gives:
$$\begin{aligned} f(p|x) &= \frac{f(p,x)}{P(X = x)} \propto f(p) P(X = x|p) \\ &\propto p^{\alpha-1}(1-p)^{\beta-1} p^x (1-p)^{n-x} \\ &= p^{\alpha+x-1}(1-p)^{\beta+n-x-1} \end{aligned}$$
- Thus $p|x \sim \mathcal{B}(\alpha + x, \beta + n - x)$, so
$$E[p|x] = \frac{\alpha + x}{\alpha + \beta + n}$$
- Observed $n = 100$, $x = 26$:

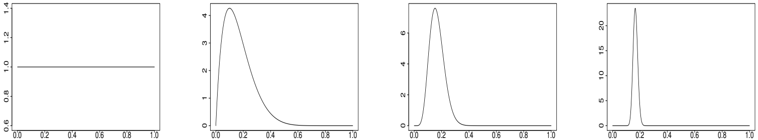
Bayesian statistics — an example

- Before observing the value of x, $f(p)$:

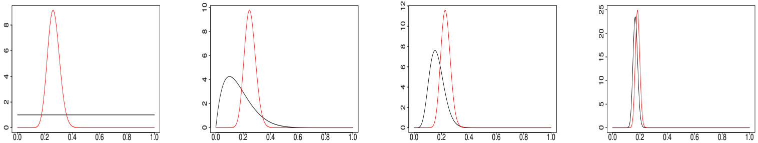


Bayesian statistics — an example

- Before observing the value of x, $f(p)$:

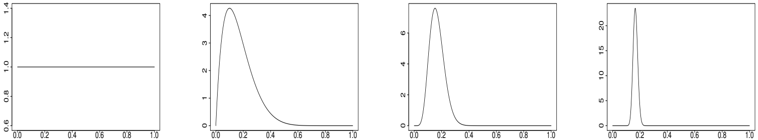


- After observing $n = 100$ and $x = 26$, $f(p|x)$:

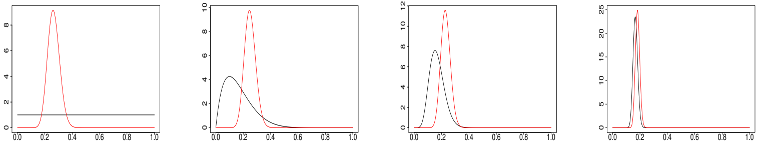


Bayesian statistics — an example

- Before observing the value of x, $f(p)$:



- After observing $n = 100$ and $x = 26$, $f(p|x)$:



$\mathrm{E}[p|x] = 0.265$ $\mathrm{E}[p|x] = 0.255$ $\mathrm{E}[p|x] = 0.230$ $\mathrm{E}[p|x] = 0.183$

Interpretation of probability

- Frequentist (objective): Probability of event A is

$$\mathrm{P}(A) = \lim_{n \to \infty} \frac{m}{n}$$

  where m: # times A occurres in n identical and independent trials

- Bayesian (subjective): Probability of event A, $\mathrm{P}(A)$, is a measure of someone's degree of belief in the occurrence of A.
  - different persons may have different $\mathrm{P}(A)$

Prior and posterior distribution

- Prior distribution: $f(\theta)$
  - a measure of our belief about the value of θ before we have observed the data, based on prior information/experience
- Observation and Likelihood: $f(x|\theta)$
  - observed value x, and its probability distribution given θ
- Posterior distribution: $f(\theta|x)$
  - a measure of our belief about the of value of θ after we have observed the data x, based on prior information/experience *and* the observed data x
  - Bayes theorem

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

Conjugate priors

- In examples: posteriors are all available on closed form
  - this is because we have used *conjugate* priors
- binomial conjugate prior
  - $x|p \sim \text{binomial}(n, p)$
  - $p \sim \text{beta}(\alpha, \beta)$
  - $p|x \sim \text{beta}(\cdot, \cdot)$

Conjugate priors

- In examples: posteriors are all available on closed form
  - this is because we have used *conjugate* priors
- binomial conjugate prior
  - $x|p \sim \text{binomial}(n, p)$
  - $p \sim \text{beta}(\alpha, \beta)$
  - $p|x \sim \text{beta}(\cdot, \cdot)$
- normal (mean) conjugate prior
  - $x_1, \ldots, x_n|\mu \sim \text{N}(\mu, \sigma_0^2)$
  - $\mu \sim \text{N}(\mu_0, \tau^2)$
  - $\mu|x_1, \ldots, x_n \sim \text{N}(\cdot, \cdot)$

Conjugate priors

- In examples: posteriors are all available on closed form
  - this is because we have used *conjugate* priors
- binomial conjugate prior
  - $x|p \sim \text{binomial}(n, p)$
  - $p \sim \text{beta}(\alpha, \beta)$
  - $p|x \sim \text{beta}(\cdot, \cdot)$
- normal (mean) conjugate prior
  - $x_1, \ldots, x_n|\mu \sim \text{N}(\mu, \sigma_0^2)$
  - $\mu \sim \text{N}(\mu_0, \tau^2)$
  - $\mu|x_1, \ldots, x_n \sim \text{N}(\cdot, \cdot)$
- normal (variance) conjugate prior
  - $x_1, \ldots, x_n|\sigma^2 \sim \text{N}(\mu_0, \sigma^2)$
  - $\sigma^2 \sim \text{IG}(\alpha, \beta)$
  - $\sigma^2|x_1, \ldots, x_n \sim \text{IG}(\cdot, \cdot)$

Conjugate priors

- In examples: posteriors are all available on closed form
  - this is because we have used *conjugate* priors
- binomial conjugate prior
  - $x|p \sim \text{binomial}(n, p)$
  - $p \sim \text{beta}(\alpha, \beta)$
  - $p|x \sim \text{beta}(\cdot, \cdot)$
- normal (mean) conjugate prior
  - $x_1, \dots, x_n|\mu \sim \text{N}(\mu, \sigma_0^2)$
  - $\mu \sim \text{N}(\mu_0, \tau^2)$
  - $\mu|x_1, \dots, x_n \sim \text{N}(\cdot, \cdot)$
- normal (variance) conjugate prior
  - $x_1, \dots, x_n|\sigma^2 \sim \text{N}(\mu_0, \sigma^2)$
  - $\sigma^2 \sim \text{IG}(\alpha, \beta)$
  - $\sigma^2|x_1, \dots, x_n \sim \text{IG}(\cdot, \cdot)$
- Conjugate priors makes analytical evaluations easier
  - and may make sampling from the posterior easier ...

Hierarchical Bayesian modeling — a simple example

- A simple example (from George et al., 1993)
  - Analysis of 10 power plant pumps
  - x_i, t_i: number of failures for pump i and length of operation time on that pump (in 1000 hours)
  - Modelling:

Hierarchical Bayesian modeling — a simple example

- A simple example (from George et al., 1993)
  - Analysis of 10 power plant pumps
  - x_i, t_i: number of failures for pump i and length of operation time on that pump (in 1000 hours)
  - Modelling:
    - $x_i|\theta_i \sim \text{Poisson}(\theta_i t_i)$
    - conjugate prior for θ_i: $\theta_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta)$

Hierarchical Bayesian modeling — a simple example

- A simple example (from George et al., 1993)
  - Analysis of 10 power plant pumps
  - x_i, t_i: number of failures for pump i and length of operation time on that pump (in 1000 hours)
  - Modelling:
    - $x_i|\theta_i \sim \text{Poisson}(\theta_i t_i)$
    - conjugate prior for θ_i: $\theta_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta)$
    - hyper-prior distribution on α and β

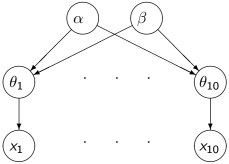
$$\alpha \sim \text{Exp}(1.0)\ ,\ \beta \sim \text{Gamma}(0.1, 1.0)$$

Hierarchical Bayesian modeling — a simple example

- A simple example (from George et al., 1993)
  - Analysis of 10 power plant pumps
  - x_i, t_i: number of failures for pump i and length of operation time on that pump (in 1000 hours)
  - Modelling:
    - $x_i|\theta_i \sim \text{Poisson}(\theta_i t_i)$
    - conjugate prior for θ_i: $\theta_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta)$
    - hyper-prior distribution on α and β

$$\alpha \sim \text{Exp}(1.0) \ , \ \beta \sim \text{Gamma}(0.1, 1.0)$$

  - graphical model:

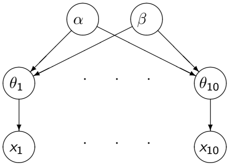


Hierarchical Bayesian modeling — a simple example

- A simple example (from George et al., 1993)
  - Analysis of 10 power plant pumps
  - x_i, t_i: number of failures for pump i and length of operation time on that pump (in 1000 hours)
  - Modelling:
    - $x_i|\theta_i \sim \text{Poisson}(\theta_i t_i)$
    - conjugate prior for θ_i: $\theta_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta)$
    - hyper-prior distribution on α and β

$$\alpha \sim \text{Exp}(1.0) \ , \ \beta \sim \text{Gamma}(0.1, 1.0)$$

  - graphical model:



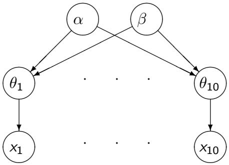
  - observed: $x_1, \dots, x_n$

Hierarchical Bayesian modeling — a simple example

- A simple example (from George et al., 1993)
  - Analysis of 10 power plant pumps
  - x_i, t_i: number of failures for pump i and length of operation time on that pump (in 1000 hours)
  - Modelling:
    - $x_i|\theta_i \sim \text{Poisson}(\theta_i t_i)$
    - conjugate prior for θ_i: $\theta_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta)$
    - hyper-prior distribution on α and β

$$\alpha \sim \text{Exp}(1.0) \; , \; \beta \sim \text{Gamma}(0.1, 1.0)$$

  - graphical model:



  - observed: $x_1, \ldots, x_n$
  - posterior distribution of interest:

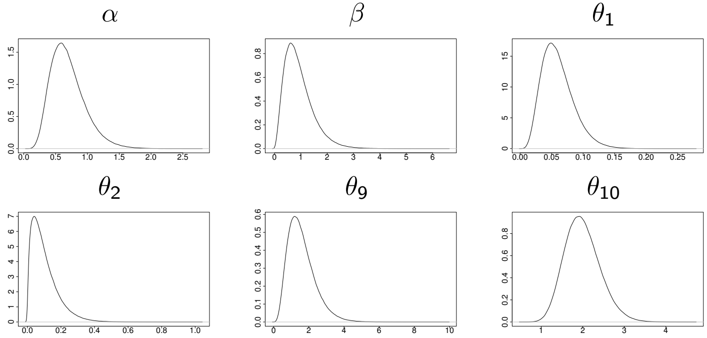
$$f(\alpha, \beta, \theta_1, \ldots, \theta_{10}|x_1, \ldots, x_{10})$$

Hierarchical Bayesian modeling — a simple example

- Data:

Pump	1	2	3	4	5	6	7	8	9	10
t_i	94.3	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
x_i	5	1	5	14	3	19	1	1	4	22

- Posterior density plots:



Hierarchical Bayesian modeling — a simple example

- Data:

Pump	1	2	3	4	5	6	7	8	9	10
t_i	94.3	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
x_i	5	1	5	14	3	19	1	1	4	22

- Posterior mean for θ_i compared to x_i/t_i

parameter	posterior mean	x_i/t_i
θ_1	0.0598	0.0530
θ_2	0.1017	0.0636
θ_3	0.0892	0.0795
θ_4	0.1157	0.1111
θ_5	0.6011	0.5725
θ_6	0.6095	0.6051
θ_7	0.8910	0.9524
θ_8	0.8928	0.9524
θ_9	1.5867	1.9047
θ_{10}	1.9901	2.0952