Introduction to Bayesian statistics

Geogios Karagiannis (Durham University)

2 July 2018

Notation

Consider a probability triple $(\Omega, \mathcal{F}, \pi)$, and

... r.v. $x : \Omega \to \mathbb{R}^{d_x}$, with $x := x(\omega)$, following a prob distr. π_x

$$\pi_x(x \in A) = \pi(\{\omega \in \Omega \,|\, x(\omega) \in A\}).$$

Notation

Consider a probability triple $(\Omega, \mathcal{F}, \pi)$, and

... r.v. $x : \Omega \to \mathbb{R}^{d_x}$, with $x := x(\omega)$, following a prob distr. π_x

$$\pi_x(x \in A) = \pi(\{\omega \in \Omega \,|\, x(\omega) \in A\}).$$

- The expected value of a function $h(x)$ w.r.t. π_x is

$$\mathrm{E}_{\pi_x}(h(x)) = \int_{\mathbb{R}^{d_x}} h(x) \mathrm{d}\pi_x(x) = \int_{\mathbb{R}^{d_x}} h(x) \pi_x(\mathrm{d}x)$$

$$= \begin{cases} \int_{\mathbb{R}^{d_x}} h(x)\pi_x(x)\mathrm{d}x & \text{if x is continuous} \\ \sum_{\forall x \in \mathbb{R}^{d_x}} h(x)\pi_x(x) & \text{if x is discrete} \end{cases}$$

The Bayes theorem (sets, events, etc...)

Consider a probability triple $(\Omega, \mathcal{F}, \pi)$,

The Bayes theorem (sets, events, etc...)

Consider a probability triple $(\Omega, \mathcal{F}, \pi)$,

- If $B \subseteq \Omega$, and $A \subseteq \Omega$, then it is

$$\pi(A|B) = \frac{\pi(B|A)\pi(A)}{\pi(B)}$$

given that $B \neq \varnothing$

The Bayes theorem (sets, events, etc...)

Consider a probability triple $(\Omega, \mathcal{F}, \pi)$,

- If $B \subseteq \Omega$, and $A \subseteq \Omega$, then it is

$$\pi(A|B) = \frac{\pi(B|A)\pi(A)}{\pi(B)}$$

given that $B \neq \varnothing$

- If , $B \subseteq \Omega$, and $\{A_1, ..., A_k\}$ is a partition of Ω, then

$$\pi(A_j|B) = \frac{\pi(B|A_j)\pi(A_j)}{\sum_{j=1}^{k} \pi(B|A_j)\pi(A_j)}, \ \forall j = 1, ..., k$$

given that $B \neq \varnothing$

The Bayes theorem (random variables)

Consider r.v. $x \in \mathbb{R}^{d_x}$ and $y \in \mathbb{R}^{d_y}$

The Bayes theorem (random variables)

Consider r.v. $x \in \mathbb{R}^{d_x}$ and $y \in \mathbb{R}^{d_y}$

- the PDF, or PMF of r.v. $x|y$ is

$$\pi_{x|y}(x|y) = \frac{\pi_{y|x}(y|x)\pi_x(x)}{\int_{\mathbb{R}^{d_x}} \pi_{y|x}(y|x)\pi_x(\mathrm{d}x)}$$

The Bayes theorem (random variables)

Consider r.v. $x \in \mathbb{R}^{d_x}$ and $y \in \mathbb{R}^{d_y}$

- the PDF, or PMF of r.v. $x|y$ is

$$\pi_{x|y}(x|y) = \frac{\pi_{y|x}(y|x)\pi_x(x)}{\int_{\mathbb{R}^{d_x}} \pi_{y|x}(y|x)\pi_x(\mathrm{d}x)}$$

- for any $C \subseteq \mathbb{R}^{d_x}$

$$\begin{aligned}\pi_{x|y}(x \in C|y) &= \int_C \pi_{x|y}(\mathrm{d}x|y) \\ &= \begin{cases} \int_C \frac{\pi_{y|x}(y|x)\pi_x(x)}{\int_{\mathbb{R}^{d_x}} \pi_{y|x}(y|x)\pi_x(x)\mathrm{d}x} \mathrm{d}x & \text{, if } x \text{ is cont} \\ \sum_{x \in C} \frac{\pi_{y|x}(y|x)\pi_x(x)}{\sum_{x \in \mathbb{R}^{d_x}} \pi_{y|x}(y|x)\pi_x(x)} & \text{, if } x \text{ is discr} \end{cases}\end{aligned}$$

The Bayes theorem (random variables)

Consider r.v. $x \in \mathbb{R}^{d_x}$ and $y \in \mathbb{R}^{d_y}$

- the PDF, or PMF of r.v. $x|y$ is

$$\pi_{x|y}(x|y) = \frac{\pi_{y|x}(y|x)\pi_x(x)}{\int_{\mathbb{R}^{d_x}} \pi_{y|x}(y|x)\pi_x(\mathrm{d}x)}$$

- for any $C \subseteq \mathbb{R}^{d_x}$

$$\pi_{x|y}(x \in C|y) = \int_C \pi_{x|y}(\mathrm{d}x|y)$$
$$= \begin{cases} \int_C \frac{\pi_{y|x}(y|x)\pi_x(x)}{\int_{\mathbb{R}^{d_x}} \pi_{y|x}(y|x)\pi_x(x)\mathrm{d}x} \mathrm{d}x & \text{, if } x \text{ is cont} \\ \sum_{x\in C} \frac{\pi_{y|x}(y|x)\pi_x(x)}{\sum_{x\in\mathbb{R}^{d_x}} \pi_{y|x}(y|x)\pi_x(x)} & \text{, if } x \text{ is discr} \end{cases}$$

- for any tiny set $\mathrm{d}x \subseteq \mathbb{R}^{d_x}$

$$\pi_{x|y}(\mathrm{d}x|y) = \frac{\pi_{y|x}(y|x)\pi_x(\mathrm{d}x)}{\int_{\mathbb{R}^{d_x}} \pi_{y|x}(y|x)\pi_x(\mathrm{d}x)}$$

Schools of statistics

- Frequentist school
  - probability of an event is the limiting relative frequency of occurrence of the event in an infinite sequence of trials.

Schools of statistics

- Frequentist school
  - probability of an event is the limiting relative frequency of occurrence of the event in an infinite sequence of trials.

- Subjective Bayesian school
  - probabilities are subjective; they represent the investigator's degrees of belief.

Schools of statistics

- Frequentist school
  - probability of an event is the limiting relative frequency of occurrence of the event in an infinite sequence of trials.
- Subjective Bayesian school
  - probabilities are subjective; they represent the investigator's degrees of belief.
  - probability is the degree of belief in a proposition, based on all the available information.

Schools of statistics

- Frequentist school
  - probability of an event is the limiting relative frequency of occurrence of the event in an infinite sequence of trials.

- Subjective Bayesian school
  - probabilities are subjective; they represent the investigator's degrees of belief.
  - probability is the degree of belief in a proposition, based on all the available information.
  - ...aka, all probabilities are conditionals; e.g., $\pi(A|B)$, or $\pi(x|y)$....

Schools of statistics

- Frequentist school
  - probability of an event is the limiting relative frequency of occurrence of the event in an infinite sequence of trials.

- Subjective Bayesian school
  - probabilities are subjective; they represent the investigator's degrees of belief.
  - probability is the degree of belief in a proposition, based on all the available information.
  - ...aka, all probabilities are conditionals; e.g., $\pi(A|B)$, or $\pi(x|y)$....
  - Statistical analyses based on the same data, and performed by different researchers may be different.

Finite exchangeability

- The random quantities $\{x_1, ..., x_n\}$ are finitely exchangeable under a probability P if the implied joint distribution satisfies

$$P(x_1 \in A_1, ..., x_n \in A_n) = P(x_{\mathfrak{p}(1)} \in A_{\mathfrak{p}(1)}, ..., x_{\mathfrak{p}(n)} \in A_{\mathfrak{p}(n)})$$

for all permutations $\mathfrak{p}$ defined on the set $\{1, ..., n\}$.

  - In terms of the corresponding PDF/PMF:

$$p(x_1, ..., x_n) = p(x_{\mathfrak{p}(1)}, ..., x_{\mathfrak{p}(n)}).$$

Finite exchangeability

- The random quantities $\{x_1, ..., x_n\}$ are finitely exchangeable under a probability P if the implied joint distribution satisfies

$$P(x_1 \in A_1, ..., x_n \in A_n) = P(x_{\mathfrak{p}(1)} \in A_{\mathfrak{p}(1)}, ..., x_{\mathfrak{p}(n)} \in A_{\mathfrak{p}(n)})$$

for all permutations $\mathfrak{p}$ defined on the set $\{1, ..., n\}$.

  - In terms of the corresponding PDF/PMF:

$$p(x_1, ..., x_n) = p(x_{\mathfrak{p}(1)}, ..., x_{\mathfrak{p}(n)}).$$

- The random quantities $x_1, x_2, ...$ are infinitely exchangeable under a probability P if every finite sub-sequence is exchangeable

Representation theorem for 0-1 r.v.

If $x_1, x_2, \ldots$ is an infinitely exchangeable sequence of $0-1$ random quantities, there exists a distribution π such that the join $p(x_1, \ldots, x_n)$ for any $x_1, \ldots, x_n$ has the form

$$p(x_1, \ldots, x_n) = \int_0^1 \prod_{i=1}^n f(x_i|\theta)\pi(\mathrm{d}\theta)$$

where

$$f(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$$

and

$$\pi(\theta \leqslant t) = \lim_{n\to\infty} P(\frac{1}{n}\sum_{i=1}^n x_i \leqslant t)$$

and $\theta = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i$ is the limiting relative frequency of 1s.

... interpretation of $0-1$ representation theorem

It is important from the subjectivist modeling point of view.

... interpretation of $0-1$ representation theorem

It is important from the subjectivist modeling point of view.

- $x_i|\theta \overset{\text{IID}}{\sim} \text{Br}(\theta)$, for all $i = 1, ..., n$
  - the x_i are considered to be independent Bernoulli random quantities, conditional on the random quantity θ.

... interpretation of $0-1$ representation theorem

It is important from the subjectivist modeling point of view.

- $x_i|\theta \overset{\text{IID}}{\sim} \text{Br}(\theta)$, for all $i = 1, ..., n$
  - the x_i are considered to be independent Bernoulli random quantities, conditional on the random quantity θ.

- θ is itself assigned a probability distribution $\pi(\text{d}\theta)$,

... interpretation of $0-1$ representation theorem

It is important from the subjectivist modeling point of view.

- $x_i|\theta \overset{\text{IID}}{\sim} \text{Br}(\theta)$, for all $i = 1, ..., n$
    - the x_i are considered to be independent Bernoulli random quantities, conditional on the random quantity θ.
- θ is itself assigned a probability distribution $\pi(\mathrm{d}\theta)$,
- $\theta = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i$, a.s. by the SLLN,
    - hence $\pi(\mathrm{d}\cdot)$ are one's beliefs about the limiting relative frequency of 1's.

... predictions

Further results

- If $x_1, x_2, \ldots$ is an infinitely exchangeable sequence of random quantities, then

$$p(x_{n+1}|x_{1:n}) = \int_{\Theta} f(x_{n+1}|\theta)\pi(\mathrm{d}\theta|x_{1:n})$$

where

$$\pi(\mathrm{d}\theta|x_{1:n}) = \frac{\prod_{i=1}^{n} f(x_i|\theta)\pi(\mathrm{d}\theta)}{\int_{\Theta}\prod_{i=1}^{n} f(x_i|\theta)\pi(\mathrm{d}\theta)}$$

Definitions needed...

A parametric statistical model

- consists of the observation of a sequence of a r.v. $x_{1:n} = (x_1, ..., x_n)$,

Definitions needed...

A parametric statistical model

- consists of the observation of a sequence of a r.v. $x_{1:n} = (x_1, ..., x_n)$,
    - distributed according to $f(\mathrm{d} \cdot |\theta)$,

Definitions needed...

A parametric statistical model

- consists of the observation of a sequence of a r.v. $x_{1:n} = (x_1, ..., x_n)$,
  - distributed according to $f(\mathrm{d}\cdot|\theta)$,
    - where only the parameter $\theta \in \Theta$ is unknown.

Definitions needed...

A parametric statistical model

- consists of the observation of a sequence of a r.v. $x_{1:n} = (x_1, ..., x_n)$,
  - distributed according to $f(\mathrm{d}\cdot|\theta)$,
    - where only the parameter $\theta \in \Theta$ is unknown.
- denoted as

$$f(x_{1:n}|\theta)$$

Definitions needed...

A parametric statistical model

- consists of the observation of a sequence of a r.v. $x_{1:n} = (x_1, ..., x_n)$,
  - distributed according to $f(\mathrm{d} \cdot |\theta)$,
    - where only the parameter $\theta \in \Theta$ is unknown.
- denoted as

$$f(x_{1:n}|\theta)$$

A Bayesian model

- is made of a parametric statistical model, $f(x_{1:n}|\theta)$, and

Definitions needed...

A parametric statistical model

- consists of the observation of a sequence of a r.v. $x_{1:n} = (x_1, ..., x_n)$,
  - distributed according to $f(\mathrm{d}\cdot|\theta)$,
    - where only the parameter $\theta \in \Theta$ is unknown.
- denoted as

$$f(x_{1:n}|\theta)$$

A Bayesian model

- is made of a parametric statistical model, $f(x_{1:n}|\theta)$, and
  - a prior distr. $\pi(\mathrm{d}\theta)$ on the unknown parameters θ.

Definitions needed...

A parametric statistical model

- consists of the observation of a sequence of a r.v. $x_{1:n} = (x_1, ..., x_n)$,
  - distributed according to $f(\mathrm{d}\cdot|\theta)$,
    - where only the parameter $\theta \in \Theta$ is unknown.
- denoted as

$$f(x_{1:n}|\theta)$$

A Bayesian model

- is made of a parametric statistical model, $f(x_{1:n}|\theta)$, and
  - a prior distr. $\pi(\mathrm{d}\theta)$ on the unknown parameters θ.
- denoted as

$$(f(x_{1:n}|\theta), \pi(\mathrm{d}\theta)) \text{ or } \begin{cases} x_{1:n}|\theta & \sim f(\mathrm{d}\cdot|\theta) \\ \theta & \sim \pi(\mathrm{d}\theta) \end{cases}$$

Quantities involved

The likelihood function of θ ginen the data $x_{1:n}$ denoted as $L(\theta; x_{1:n})$, defined as ,

$$L(\theta; x_{1:n}) = f(x_{1:n}|\theta)$$

Quantities involved

The likelihood function of θ ginen the data $x_{1:n}$ denoted as $L(\theta; x_{1:n})$, defined as ,

$$L(\theta; x_{1:n}) = f(x_{1:n}|\theta)$$

The posterior distribution of θ given the data $x_{1:n}$ has PDF/PMF

$$\pi(\theta|x_{1:n}) = \frac{f(x_{1:n}|\theta)\pi(\theta)}{\int_\Theta f(x_{1:n}|\theta)\pi(\mathrm{d}\theta)} = \frac{L(\theta; x_{1:n})\pi(\theta)}{\int_\Theta L(\theta; x_{1:n})\pi(\mathrm{d}\theta)}$$

$$\pi(\theta|x_{1:n}) \propto f(x_{1:n}|\theta)\pi(\theta)$$

Quantities involved

The likelihood function of θ ginen the data $x_{1:n}$ denoted as $L(\theta; x_{1:n})$, defined as ,

$$L(\theta; x_{1:n}) = f(x_{1:n}|\theta)$$

The posterior distribution of θ given the data $x_{1:n}$ has PDF/PMF

$$\pi(\theta|x_{1:n}) = \frac{f(x_{1:n}|\theta)\pi(\theta)}{\int_\Theta f(x_{1:n}|\theta)\pi(\mathrm{d}\theta)} = \frac{L(\theta; x_{1:n})\pi(\theta)}{\int_\Theta L(\theta; x_{1:n})\pi(\mathrm{d}\theta)}$$

$$\pi(\theta|x_{1:n}) \propto f(x_{1:n}|\theta)\pi(\theta)$$

The predictive distribution of $y := x_{n+1}$ given the data $x_{1:n}$ has PDF/PMF

$$p(y|x_{1:n}) = \underbrace{\int_\Theta f(y|\theta)\pi(\mathrm{d}\theta|x_{1:n})}_{=\mathrm{E}_{\pi_{\theta|x_{1:n}}}(f(y|\theta))}$$

Conjugate prior distr.

If $\mathcal{F} = \{f(\cdot|\theta); \forall\theta \in \Theta\}$ is a family of distributions (sampling distributions),

Conjugate prior distr.

If $\mathcal{F} = \{f(\cdot|\theta); \forall \theta \in \Theta\}$ is a family of distributions (sampling distributions),

and $\mathcal{P} = \{\pi(\cdot)\}$ is a family of prior distributions for θ,

Conjugate prior distr.

If $\mathcal{F} = \{f(\cdot|\theta); \forall\theta \in \Theta\}$ is a family of distributions (sampling distributions),

and $\mathcal{P} = \{\pi(\cdot)\}$ is a family of prior distributions for θ,

then the family $\mathcal{P}$ is conjugate for $\mathcal{F}$ if

$$\forall f(x_{1:n}|\theta) \in \mathcal{F} \text{ and } \pi(\theta) \in \mathcal{P} \implies \pi(\theta|x_{1:n}) \in \mathcal{P}$$

Sufficiency...

- The summary statistic $t_n := t_n(x_{1:n})$ is parametric sufficient for parameter θ for an exchangeable sequence of observables $x_1, ..., x_n$ if and only if

$$\pi(\mathrm{d}\theta|x_{1:n}) = \pi(\mathrm{d}\theta|t_n)$$

Sufficiency...

- The summary statistic $t_n := t_n(x_{1:n})$ is parametric sufficient for parameter θ for an exchangeable sequence of observables $x_1, ..., x_n$ if and only if
$$\pi(\mathrm{d}\theta | x_{1:n}) = \pi(\mathrm{d}\theta | t_n)$$

- Neyman factorization criterion: The summary statistic $t_n = t(x_{1:n})$ is parametric sufficient for $x_{1:n} = (x_1, ..., x_n)$ if and only if, the joint PDF/PMF $x_{1:n}|\theta$ is s.t.
$$f(x_1, ..., x_n|\theta) = h(t_n, \theta) g(x_1, ..., x_n),$$
for some functions $h \geqslant 0$, $g \geqslant 0$.

Exponential family of distributions- definition

A distribution with PDF/PMF $f(x|\theta)$, $\theta \in \Theta$, belongs to the k-parameter exponential family if

$$f(x|\theta) = u(x)g(\theta)\exp(\sum_{j=1}^{k} c_j\phi_j(\theta)h_j(x))$$

for $x \in \mathcal{X}$ where

$$g(\theta)^{-1} = \int_{\mathcal{X}} u(x)\exp(\sum_{j=1}^{k} c_j\phi_j(\theta)h_j(x))\mathrm{d}x < \infty$$

Exponential family of distributions- definition

A distribution with PDF/PMF $f(x|\theta)$, $\theta \in \Theta$, belongs to the k-parameter exponential family if

$$f(x|\theta) = u(x)g(\theta)\exp(\sum_{j=1}^{k} c_j \phi_j(\theta) h_j(x))$$

for $x \in \mathcal{X}$ where

$$g(\theta)^{-1} = \int_{\mathcal{X}} u(x)\exp(\sum_{j=1}^{k} c_j \phi_j(\theta) h_j(x))\mathrm{d}x < \infty$$

Here, $f(x|\theta)$ is denoted as $\mathrm{Ef}_k(x|u, g, h, \phi, \theta, c)$

Exponential family of distributions- priors [Applets: #1]

If $x_1, x_2, x_3, ...x_n$ such that $x_i \in \mathcal{X}$ are IID according to the given regular exponential family $\mathrm{Ef}_k(\cdot|\cdot)$, then :

Exponential family of distributions- priors

[Applets: #1]

If $x_1, x_2, x_3, ...x_n$ such that $x_i \in \mathcal{X}$ are IID according to the given regular exponential family $\mathrm{Ef}_k(\cdot|\cdot)$, then :

- Sufficient statistic is $t_n := (n, \sum_{i=1}^n h_1(x_i), ..., \sum_{i=1}^n h_k(x_i))$.

Exponential family of distributions- priors [Applets: #1]

If $x_1, x_2, x_3, ...x_n$ such that $x_i \in \mathcal{X}$ are IID according to the given regular exponential family $\text{Ef}_k(\cdot|\cdot)$, then :

- Sufficient statistic is $t_n := (n, \sum_{i=1}^n h_1(x_i), ..., \sum_{i=1}^n h_k(x_i))$.
- The conjugate prior distribution for $\theta \in \Theta$ has the form

$$\pi(\theta|\tau) \propto \quad g(\theta)^{\tau_0} \exp(\sum_{j=1}^{k} c_j \phi_j(\theta) \tau_j)$$

Exponential family of distributions- priors [Applets: #1]

If $x_1, x_2, x_3, ...x_n$ such that $x_i \in \mathcal{X}$ are IID according to the given regular exponential family $\mathrm{Ef}_k(\cdot|\cdot)$, then :

- Sufficient statistic is $t_n := (n, \sum_{i=1}^{n} h_1(x_i), ..., \sum_{i=1}^{n} h_k(x_i))$.
- The conjugate prior distribution for $\theta \in \Theta$ has the form

$$\pi(\theta|\tau) \propto \quad g(\theta)^{\tau_0} \exp(\sum_{j=1}^{k} c_j \phi_j(\theta)\tau_j)$$

- The posterior distribution for $\theta \in \Theta$ has the form

$$\begin{aligned}\pi(\theta|x_{1:n}, \tau) &\propto \quad g(\theta)^{\tau_0^*} \exp(\sum_{j=1}^{k} c_j \phi_j(\theta)\tau_j^*) \\ &\propto \quad \pi(\theta|\tau^*)\end{aligned}$$

with $\tau^* = (\tau_0^*, \tau_1^*, ..., \tau_k^*)$, $\tau_0^* = \tau_0 + n$, and $\tau_j^* = \sum_{i=1}^{n} h_j(x_i) + \tau_j$

Bayes point estimator

$\delta^{\pi} := \delta^{\pi}(x_{1:n})$: Bayes point estimator of θ , w.r.t.

- the posterior distribution

$$\pi(\mathrm{d}\theta|x_{1:n})$$

- the loss function $\ell(\theta, \delta)$

$$\ell : \Theta \times \mathcal{D} \to [0, +\infty)$$

is the quantity δ^{π} s.t.

Bayes point estimator

$\delta^{\pi} := \delta^{\pi}(x_{1:n})$: Bayes point estimator of θ , w.r.t.

- the posterior distribution

$$\pi(\mathrm{d}\theta|x_{1:n})$$

- the loss function $\ell(\theta, \delta)$

$$\ell : \Theta \times \mathcal{D} \rightarrow [0, +\infty)$$

is the quantity δ^{π} s.t.

$$\begin{aligned}\delta^{\pi} &= \arg\min_{\forall \delta \in \mathcal{D}} \mathrm{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(\ \ell(\theta, \delta)\ ) \\ &= \arg\min_{\forall \delta \in \mathcal{D}} \int_{\Theta} \ell(\theta, \delta)\pi(\mathrm{d}\theta|x_{1:n})\end{aligned}$$

Bayes estimator standard error (univariate case)

The standard error of δ^{π} of parameter $\theta \in \Theta$ with $\pi(\mathrm{d}\theta|x_{1:n})$ is

$$\mathsf{Se}(\delta|x_{1:n}) = \sqrt{\mathsf{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(\theta - \delta)^2}$$

Bayes estimator standard error (univariate case)

The standard error of δ^{π} of parameter $\theta \in \Theta$ with $\pi(\mathrm{d}\theta|x_{1:n})$ is

$$\mathsf{Se}(\delta|x_{1:n}) = \sqrt{\mathsf{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(\theta-\delta)^2}$$

where

$$\underbrace{\mathsf{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(\theta-\delta)^2}_{=\text{post. MSE}} = \underbrace{\mathsf{Var}_{\pi(\mathrm{d}\theta|x_{1:n})}(\theta)}_{=\text{post. var.}} + (\underbrace{\mathsf{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(\theta)-\delta}_{=\text{bias}})^2$$

is the mean squared error of δ.

Standard parametric point estimators

The Bayes estimate δ^{π} of θ with respect to the :

- weighted quadratic loss function $\ell(\theta, \delta) = w(\theta)(\theta - \delta)^2$, $w(\theta) > 0$ is

$$\delta^{\pi} = \frac{\mathrm{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(w(\theta)\theta)}{\mathrm{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(w(\theta))}. \tag{6.1}$$

Standard parametric point estimators

The Bayes estimate δ^{π} of θ with respect to the :

- weighted quadratic loss function $\ell(\theta, \delta) = w(\theta)(\theta - \delta)^2$, $w(\theta) > 0$ is

$$\delta^{\pi} = \frac{\mathrm{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(w(\theta)\theta)}{\mathrm{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(w(\theta))}. \tag{6.1}$$

- quadratic loss function $\ell(\theta, \delta) = (\theta - \delta)^2$, is

$$\delta^{\pi} = \mathrm{E}_{\pi(\theta|x_{1:n})}(\theta) \tag{6.2}$$

Standard parametric point estimators

The Bayes estimate δ^{π} of θ with respect to the :

- weighted quadratic loss function $\ell(\theta,\delta) = w(\theta)(\theta-\delta)^2$, $w(\theta) > 0$ is

$$\delta^{\pi} = \frac{\mathrm{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(w(\theta)\theta)}{\mathrm{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(w(\theta))}. \tag{6.1}$$

- quadratic loss function $\ell(\theta,\delta) = (\theta-\delta)^2$, is

$$\delta^{\pi} = \mathrm{E}_{\pi(\theta|x_{1:n})}(\theta) \tag{6.2}$$

- linear loss function $\ell(\theta,\delta) = c_1(\delta-\theta)\mathbb{1}_{\theta\leqslant\delta}(\delta) + c_2(\theta-\delta)\mathbb{1}_{\theta>\delta}(\delta)$ is

$$\text{is the } \frac{c_2}{c_1+c_2}\text{-th posterior quantile}$$

Standard parametric point estimators

The Bayes estimate δ^{π} of θ with respect to the :

- weighted quadratic loss function $\ell(\theta,\delta)=w(\theta)(\theta-\delta)^2$, $w(\theta)>0$ is

$$\delta^{\pi}=\frac{\mathrm{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(w(\theta)\theta)}{\mathrm{E}_{\pi(\mathrm{d}\theta|x_{1:n})}(w(\theta))}. \tag{6.1}$$

- quadratic loss function $\ell(\theta,\delta)=(\theta-\delta)^2$, is

$$\delta^{\pi}=\mathrm{E}_{\pi(\theta|x_{1:n})}(\theta) \tag{6.2}$$

- linear loss function $\ell(\theta,\delta)=c_1(\delta-\theta)\mathbb{1}_{\theta\leqslant\delta}(\delta)+c_2(\theta-\delta)\mathbb{1}_{\theta>\delta}(\delta)$ is

$$\text{is the } \frac{c_2}{c_1+c_2}\text{-th posterior quantile}$$

- absolute loss function $\ell(\theta,\delta)=|\theta-\delta|$, is

$$\delta=\mathrm{median}_{\pi(\theta|x_{1:n})}(\theta).$$

Standard predictive point estimators

The Bayes estimate δ^p of $y = x_{n+1}$ with respect to the :

- weighted quadratic loss function $\ell(y,\delta) = w(y)(y-\delta)^2$, $w(y) > 0$ is

$$\delta^p = \frac{\mathrm{E}_{p(\mathrm{d}|x_{1:n})}(w(y)y)}{\mathrm{E}_{p(\mathrm{d}y|x_{1:n})}(w(y))}.$$

- quadratic loss function $\ell(\theta,\delta) = (\theta-\delta)^2$, is

$$\delta^p = \mathrm{E}_{p(\mathrm{d}y|x_{1:n})}(y)$$

- linear loss function $\ell(y,\delta) = c_1(\delta - y)\mathbb{1}_{y\leqslant\delta}(\delta) + c_2(y-\delta)\mathbb{1}_{y>\delta}(\delta)$ is

$$\text{is the } \frac{c_2}{c_1+c_2}\text{-th predictive quantile}$$

- absolute loss function $\ell(y,\delta) = |y-\delta|$, is

$$\delta^p = \mathrm{median}_{p(\mathrm{d}y|x_{1:n})}(y).$$

Posterior credible set

[Applet: #1]

The $100(1-a)\%$ credible set for $\theta \in \Theta$ w.r.t. the posterior $\pi(\mathrm{d}\theta|x_{1:n})$

is any subset C_a of Θ

such that

$$\pi(\theta \in C_a|x_{1:n}) = \int_{C_a} \pi(\mathrm{d}\theta|x_{1:n}) \geqslant 1 - a$$

Posterior highest probability density (HPD) set [Applet: #1]

The $100(1-a)\%$ HPD set for $\theta \in \Theta$ w.r.t. the posterior $\pi(\mathrm{d}\theta|x_{1:n})$

is the subset C_a of Θ

which is of the form

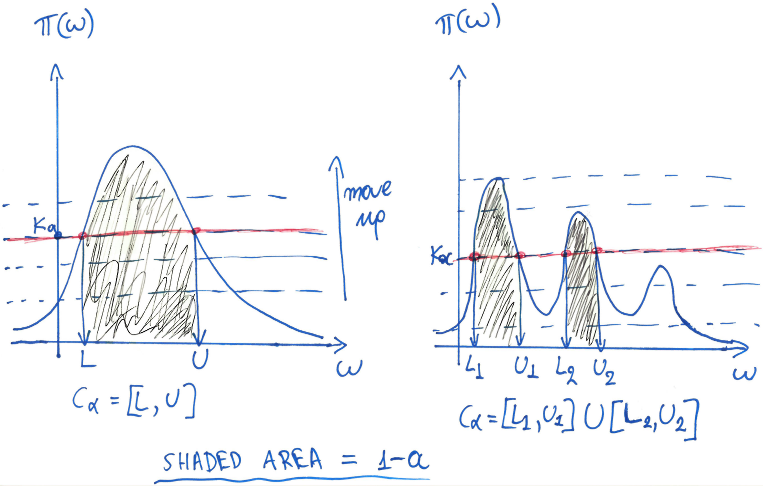
$$C_a = \{\theta \in \Theta : \pi(\theta|x_{1:n}) \geqslant k_a\}$$

such that k_a is the largest constant where

$$\pi(\theta \in C_a|x_{1:n}) = \int_{C_a} \pi(\mathrm{d}\theta|x_{1:n}) \geqslant 1-a$$

Schematic of the (1-a) HPD

[Applet: #1]



Posterior (HPD) interval

[Applet: #1]

Assume $\theta \sim \pi(\mathrm{d}\theta|x_{1:n})$ with unimodal density $\pi(\theta|x_{1:n})$.

If the interval $C_a = [L, U]$ satisfies

1. $\int_L^U \pi(\theta|x_{1:n})\mathrm{d}\theta = 1 - a$,
2. $\pi(U|x_{1:n}) = \pi(L|x_{1:n}) > 0$, and
3. $\theta_{\text{mode}} \in (L, U)$, where θ_{mode} is the mode of $\pi(\theta|x_{1:n})$,

then $C_a = [L, U]$ is the HPD interval of θ w.r.t. $\pi(\mathrm{d}\theta|x_{1:n})$.

Hypothesis test

There is interest to test the hypotheses

$$\mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta \in \Theta_1$$

where $\Theta_0 \cap \Theta_1 = \varnothing$

Hypothesis test

There is interest to test the hypotheses

$$\mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta \in \Theta_1$$

where $\Theta_0 \cap \Theta_1 = \varnothing$

Consider the (overall) Bayesian model

$$\begin{cases} x_i|\theta & \overset{\text{IID}}{\sim} f(\mathrm{d}\cdot|\theta),\ \theta \in \Theta, \qquad i = 1, ..., n \\ \theta & \sim \pi(\mathrm{d}\theta) \end{cases}$$

where Θ is partitioned by $\{\Theta_0, \Theta_1\}$

Hypothesis test

There is interest to test the hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

The overall prior on $\theta \in \Theta$,

$$\pi(\mathrm{d}\theta) = \underbrace{\pi(\theta \in \Theta_0)}_{=\pi_0} \times \underbrace{\pi(\mathrm{d}\theta|\theta \in \Theta_0)}_{=\pi_0(\mathrm{d}\theta)} + \underbrace{\pi(\theta \in \Theta_1)}_{\pi_1}\underbrace{\pi(\mathrm{d}\theta|\theta \in \Theta_1)}_{=\pi_1(\mathrm{d}\theta)}$$

Hypothesis test

There is interest to test the hypotheses

$$\mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta \in \Theta_1$$

The overall prior on $\theta \in \Theta$,

$$\pi(\mathrm{d}\theta) = \underbrace{\pi(\theta \in \Theta_0)}_{=\pi_0} \times \underbrace{\pi(\mathrm{d}\theta|\theta \in \Theta_0)}_{=\pi_0(\mathrm{d}\theta)} + \underbrace{\pi(\theta \in \Theta_1)}_{\pi_1}\underbrace{\pi(\mathrm{d}\theta|\theta \in \Theta_1)}_{=\pi_1(\mathrm{d}\theta)}$$

where

$$\pi_0 = \int_{\Theta_0} \pi(\mathrm{d}\theta), \qquad \pi_0(\theta) = \frac{\pi(\theta)\mathbb{1}_{\Theta_0}(\theta)}{\int_{\Theta_0} \pi(\mathrm{d}\theta)},$$

$$\pi_1 = \int_{\Theta_1} \pi(\mathrm{d}\theta), \qquad \pi_1(\theta) = \frac{\pi(\theta)\mathbb{1}_{\Theta_1}(\theta)}{\int_{\Theta_1} \pi(\mathrm{d}\theta)}.$$

Hypothesis test

There is interest to test the hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

In Bayes paradigm, it becomes

$$\begin{cases} H_0 : & (\, f(x_{1:n}|\theta)\,,\ \pi_0(d\theta)\,) \\ H_1 : & (\, f(x_{1:n}|\theta)\,,\ \pi_1(d\theta)\,) \end{cases}$$

- $\pi_0(d\theta)$ gives $\pi_0(\theta) > 0$ at $\theta \in \Theta_0$ and $\pi_0(\theta) = 0$ at $\theta \in \Theta_1$
- $\pi_1(d\theta)$ gives $\pi_1(\theta) > 0$ at $\theta \in \Theta_1$ and $\pi_1(\theta) = 0$ at $\theta \in \Theta_0$

Hypothesis test

There is interest to test the hypotheses

$$\mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta \in \Theta_1$$

In Bayes paradigm, it becomes

$$\begin{cases} \mathrm{H}_0 : & x_{1:n} \sim p_0(x_{1:n}) = \int_{\Theta_0} \prod_{i=1}^{n} f(x_i|\theta)\pi_0(\mathrm{d}\theta) \\ \mathrm{H}_1 : & x_{1:n} \sim p_1(x_{1:n}) = \int_{\Theta_1} \prod_{i=1}^{n} f(x_i|\theta)\pi_1(\mathrm{d}\theta) \end{cases}$$

- $\pi_0(\mathrm{d}\theta)$ gives $\pi_0(\theta) > 0$ at $\theta \in \Theta_0$ and $\pi_0(\theta) = 0$ at $\theta \in \Theta_1$
- $\pi_1(\mathrm{d}\theta)$ gives $\pi_1(\theta) > 0$ at $\theta \in \Theta_1$ and $\pi_1(\theta) = 0$ at $\theta \in \Theta_0$

Hypothesis test (as point estimation)

The Bayesian parametric point estimator δ^{π} of the indicator function

$$\mathbb{1}_{\Theta_1}(\theta) = \begin{cases} 0 & , \theta \in \Theta_0 \\ 1 & , \theta \in \Theta_1 \end{cases}$$

Hypothesis test (as point estimation)

The Bayesian parametric point estimator δ^{π} of the indicator function

$$\mathbb{1}_{\Theta_1}(\theta) = \begin{cases} 0 & , \theta \in \Theta_0 \\ 1 & , \theta \in \Theta_1 \end{cases}$$

w.r.t. the loss function $c_{\mathrm{I}} - c_{\mathrm{II}}$ loss function

$$\ell(\theta, \delta) = \begin{cases} 0 & \text{, if } \theta \in \Theta_0,\ \delta = 0 \\ 0 & \text{, if } \theta \notin \Theta_0,\ \delta = 1 \\ c_{\mathrm{II}} & \text{, if } \theta \notin \Theta_0,\ \delta = 0 \\ c_{\mathrm{I}} & \text{, if } \theta \in \Theta_0,\ \delta = 1 \end{cases}$$

Hypothesis test (as point estimation)

The Bayesian parametric point estimator δ^{π} of the indicator function

$$\mathbb{1}_{\Theta_1}(\theta) = \begin{cases} 0 & , \theta \in \Theta_0 \\ 1 & , \theta \in \Theta_1 \end{cases}$$

w.r.t. the loss function $c_{\text{I}} - c_{\text{II}}$ loss function

$$\ell(\theta, \delta) = \begin{cases} 0 & \text{, if } \theta \in \Theta_0,\ \delta = 0 \\ 0 & \text{, if } \theta \notin \Theta_0,\ \delta = 1 \\ c_{\text{II}} & \text{, if } \theta \notin \Theta_0,\ \delta = 0 \\ c_{\text{I}} & \text{, if } \theta \in \Theta_0,\ \delta = 1 \end{cases}$$

is

$$\delta^{\pi} = \begin{cases} 0 & \text{, if } \pi(\theta \in \Theta_0 | x_{1:n}) > \frac{c_{\text{II}}}{c_{\text{II}} + c_{\text{I}}} \\ 1 & \text{, otherwise} \end{cases}$$

Bayes factor

For

$$\mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta \in \Theta_1$$

The Bayes factor $\mathrm{B}_{01}(x_{1:n})$ of H_0 over H_1 is

$$\mathrm{B}_{01}(x_{1:n}) = \frac{\pi(\theta \in \Theta_0 | x_{1:n}) / \pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1 | x_{1:n}) / \pi(\theta \in \Theta_1)}$$

where

$$\pi(\theta \in \Theta_0) = \int_{\Theta_0} \pi(\mathrm{d}\theta); \qquad \pi(\theta \in \Theta_0 | x_{1:n}) = \int_{\Theta_0} \pi(\mathrm{d}\theta | x_{1:n});$$
$$\pi(\theta \in \Theta_1) = \int_{\Theta_1} \pi(\mathrm{d}\theta); \qquad \pi(\theta \in \Theta_1 | x_{1:n}) = \int_{\Theta_1} \pi(\mathrm{d}\theta | x_{1:n}).$$

Bayes factor (cases)

For

$$\mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta \in \Theta_1$$

The Bayes factor $\mathrm{B}_{01}(x_{1:n})$ of H_0 over H_1 is

$$\mathrm{B}_{01}(x_{1:n}) = \frac{\pi(\theta \in \Theta_0|x_{1:n})/\pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1|x_{1:n})/\pi(\theta \in \Theta_1)}$$

Bayes factor (cases)

For

$$\mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta \in \Theta_1$$

The Bayes factor $\mathrm{B}_{01}(x_{1:n})$ of H_0 over H_1 is

$$\mathrm{B}_{01}(x_{1:n}) = \frac{\pi(\theta \in \Theta_0 | x_{1:n}) / \pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1 | x_{1:n}) / \pi(\theta \in \Theta_1)}$$

$$= \begin{cases} \frac{\int_{\Theta_0} f(x_{1:n}|\theta)\pi_0(\mathrm{d}\theta)}{\int_{\Theta_1} f(x_{1:n}|\theta)\pi_1(\mathrm{d}\theta)} & ; \mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs} \quad \mathrm{H}_1 : \theta \in \Theta_1 \\ & \\ & \end{cases}$$

Bayes factor (cases)

For

$$\mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta \in \Theta_1$$

The Bayes factor $\mathrm{B}_{01}(x_{1:n})$ of H_0 over H_1 is

$$\mathrm{B}_{01}(x_{1:n}) = \frac{\pi(\theta \in \Theta_0 | x_{1:n}) / \pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1 | x_{1:n}) / \pi(\theta \in \Theta_1)}$$

$$= \begin{cases} \frac{\int_{\Theta_0} f(x_{1:n}|\theta)\pi_0(\mathrm{d}\theta)}{\int_{\Theta_1} f(x_{1:n}|\theta)\pi_1(\mathrm{d}\theta)} & ; \mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs} \quad \mathrm{H}_1 : \theta \in \Theta_1 \\ \frac{f(x_{1:n}|\theta_0)}{\int_{\Theta_1} f(x_{1:n}|\theta)\pi_1(\mathrm{d}\theta)} & ; \mathrm{H}_0 : \theta \in \{\theta_0\} \quad \text{vs} \quad \mathrm{H}_1 : \theta \in \Theta_1 \end{cases}$$

Bayes factor (cases)

For

$$\mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs.} \quad \mathrm{H}_1 : \theta \in \Theta_1$$

The Bayes factor $\mathrm{B}_{01}(x_{1:n})$ of H_0 over H_1 is

$$\mathrm{B}_{01}(x_{1:n}) = \frac{\pi(\theta \in \Theta_0 | x_{1:n}) / \pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1 | x_{1:n}) / \pi(\theta \in \Theta_1)}$$

$$= \begin{cases} \frac{\int_{\Theta_0} f(x_{1:n}|\theta)\pi_0(\mathrm{d}\theta)}{\int_{\Theta_1} f(x_{1:n}|\theta)\pi_1(\mathrm{d}\theta)} & ;\ \mathrm{H}_0 : \theta \in \Theta_0 \quad \text{vs} \quad \mathrm{H}_1 : \theta \in \Theta_1 \\ \frac{f(x_{1:n}|\theta_0)}{\int_{\Theta_1} f(x_{1:n}|\theta)\pi_1(\mathrm{d}\theta)} & ;\ \mathrm{H}_0 : \theta \in \{\theta_0\} \quad \text{vs} \quad \mathrm{H}_1 : \theta \in \Theta_1 \\ \frac{f(x_{1:n}|\theta_0)}{f(x_{1:n}|\theta_1)} & ;\ \mathrm{H}_0 : \theta \in \{\theta_0\} \quad \text{vs} \quad \mathrm{H}_1 : \theta \in \{\theta_1\} \end{cases}$$

Bayes factor: decision

For

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

The hypothesis H_0 is accepted when

$$B_{01}(x_{1:n}) > \frac{c_{II}}{c_I} \frac{\pi_1}{\pi_0}$$

under the loss function

$$\ell(\theta, \delta) = \begin{cases} 0 & \text{, if } \theta \in \Theta_0,\ \delta = 0 \\ 0 & \text{, if } \theta \notin \Theta_0,\ \delta = 1 \\ c_{II} & \text{, if } \theta \notin \Theta_0,\ \delta = 0 \\ c_I & \text{, if } \theta \in \Theta_0,\ \delta = 1 \end{cases}$$