§2 Decision Problem: Bayesian Approach

§2.1 Prior to posterior

2.1.1 Key features of Bayesian approach:

- The concept of "probability" is employed to interpret an investigator's degree of belief about the unknown parameter θ. Thus, θ is treated as if it were **random**, i.e. had a distribution over the parameter space Θ.
- Statistical inference consists of "revision" of one's belief about the true value of θ in the light of empirical evidence found in data $\boldsymbol{x}$.

2.1.2 *Frequentist approach* —

$\boldsymbol{X}$: observable random variate with probability function $f(\boldsymbol{x}|\theta)$
θ: fixed parameter with unknown value.

Bayesian approach —

$\boldsymbol{X}$: observable random variate with probability function $f(\boldsymbol{x}|\theta)$
θ: unobservable random variate with a specified *prior* probability function $\pi(\theta)$:

$$\int_\Theta \pi(\theta)\,d\theta = 1 \;\; \text{(continuous } \theta\text{)}, \quad \text{or} \quad \sum_{\theta\in\Theta} \pi(\theta) = 1 \;\; \text{(discrete } \theta\text{)}.$$

Note: To fix ideas we focus on the continuous case in our general discussion. The discrete case can be handled analogously.

2.1.3 The Bayesian approach is criticised as being highly sensitive to the **subjective** choice of $\pi(\theta)$. Two statisticians holding opposite prior beliefs may arrive at opposite conclusions from the same observation $\boldsymbol{x}$ and the same model.

2.1.4 Given $\pi(\theta)$ and $f(\boldsymbol{x}|\theta)$, we have

$$\begin{aligned}
\text{"joint probability function" of } (\boldsymbol{X},\theta):&\quad f(\boldsymbol{x}|\theta)\pi(\theta)\\
\text{"marginal probability function" of } \boldsymbol{X}:&\quad \textstyle\int_\Theta f(\boldsymbol{x}|\theta')\,\pi(\theta')\,d\theta' = f_{\boldsymbol{X}}(\boldsymbol{x}), \text{ say}\\
\text{"conditional probability function" of } \theta \text{ given } \boldsymbol{X}=\boldsymbol{x}:&\quad f(\boldsymbol{x}|\theta)\,\pi(\theta)/f_{\boldsymbol{X}}(\boldsymbol{x})
\end{aligned}$$

Definition. The *posterior* probability function of θ given the observed data $\boldsymbol{x}$ is defined to be the conditional probability function of θ given $\boldsymbol{X} = \boldsymbol{x}$, that is

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)\,\pi(\theta)}{\displaystyle\int_\Theta f(\boldsymbol{x}|\theta')\,\pi(\theta')\,d\theta'}.$$

Note: The denominator depends only on $\boldsymbol{x}$ but not on θ. It serves as a **normalising constant** for the posterior probability function $\pi(\cdot|\boldsymbol{x})$. Commonly we write

$$\pi(\theta|\boldsymbol{x}) \propto f(\boldsymbol{x}|\theta)\pi(\theta) \quad \text{(as a function of } \theta\text{)}.$$

2.1.5 **Example §2.1.1** Suppose $X|\theta \sim N(\theta, 1)$; $\theta \sim N(\mu_0, \sigma_0^2)$ (prior); μ_0, σ_0 given.

$$\pi(\theta|x) \;\propto\; f(x|\theta)\,\pi(\theta) \;\propto\; \exp\left\{-\frac{(x-\theta)^2}{2} - \frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right\} \;\propto\; \exp\left\{-\frac{(\theta-\mu_1)^2}{2\sigma_1^2}\right\},$$

where

$$\mu_1 = \frac{\mu_0/\sigma_0^2 + x}{1/\sigma_0^2 + 1}, \qquad \frac{1}{\sigma_1^2} = 1 + \frac{1}{\sigma_0^2}.$$

As $\pi(\theta|x)$ is a pdf, hence

$$\pi(\theta|x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{(\theta-\mu_1)^2}{2\sigma_1^2}\right\},$$

i.e. $\theta|x \sim N(\mu_1, \sigma_1^2)$.

Note:

- μ_1 is a convex combination (i.e. weighted average) of μ_0 and x (location of θ adjusted by x).
- $\sigma_1^2 < \sigma_0^2$ (variability of θ reduced after considering information provided by x).

§2.2 Bayesian decision

2.2.1 Let the prior $\pi(\theta)$ be given for $\theta \in \Theta$. Consider a decision problem with loss function $L(\theta, a)$ for $\theta \in \Theta$ and action $a \in \mathcal{A}$ (action space).

Definition. The *expected posterior loss* given data $\boldsymbol{x}$, incurred by taking action a, is

$$\mathbb{E}[L(\theta, a)|\boldsymbol{x}] = \int_\Theta L(\theta, a)\pi(\theta|\boldsymbol{x})\,d\theta.$$

Note: By contrast, the risk function $R(\theta, d) = \mathbb{E}\big[L(\theta, d(\boldsymbol{X}))\big|\theta\big]$ is the *expected* loss given θ, incurred by adopting decision rule d.

2.2.2 **Definition.** A *Bayesian decision* takes an action $a \in \mathcal{A}$ which minimises the expected posterior loss $\mathbb{E}[L(\theta, a)|\boldsymbol{x}]$.

2.2.3 Writing $f_{\boldsymbol{X}}(\boldsymbol{x}) = \int_\Theta \pi(\theta') f(\boldsymbol{x}|\theta')\, d\theta'$, we have

$$\mathbb{E}[L(\theta, a)|\boldsymbol{x}] = \int_\Theta L(\theta, a) \frac{f(\boldsymbol{x}|\theta)\, \pi(\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x})}\, d\theta = \frac{1}{f_{\boldsymbol{X}}(\boldsymbol{x})} \int_\Theta L(\theta, a)\, f(\boldsymbol{x}|\theta)\, \pi(\theta)\, d\theta.$$

Thus, minimising $\mathbb{E}[L(\theta, a)|\boldsymbol{x}]$ w.r.t. $a \in \mathcal{A}$ is equivalent to minimising $\int_\Theta L(\theta, a) f(\boldsymbol{x}|\theta)\pi(\theta)\, d\theta$ w.r.t. $a \in \mathcal{A}$.

Important remarks:

- If we repeat the above minimisation for **each** possible outcome $\boldsymbol{x} \in \mathcal{S}$, the respective Bayesian decision can be interpreted as a function $d(\boldsymbol{x})$ on the sample space $\mathcal{S}$.

  We have seen in §1.8.8 that this function $d(\boldsymbol{x})$ is a Bayes rule.
- Bayesians are more concerned with the Bayesian decision $d(\boldsymbol{x})$ for the particular $\boldsymbol{x}$ that has been observed, but do not bother to derive the **complete** Bayes rule d.

2.2.4 **Example §1.1**: (cont'd)

Consider a prior pdf proportional to the prior weight function specified in §1.9.5(iii):

$$\pi(\theta) = c\theta^{c-1}, \quad 0 < \theta < 1,$$

for some constant $c > 0$.

Given observation $\boldsymbol{x} = (x_1, x_2) \in \{1, \ldots, 8\}^2$, posterior pdf:

$$\pi(\theta|\boldsymbol{x}) \propto \pi(\theta)\, f(x_1, x_2|\theta) \propto \theta^{c-1+(x_1^2+x_2^2-x_1-x_2)/2}\big(1 - \mathbf{1}\{x_1 < 8\}\theta^{x_1}\big)\big(1 - \mathbf{1}\{x_2 < 8\}\theta^{x_2}\big).$$

The expected posterior loss is given by

$$\mathbb{E}\big[L(\theta, a_0)\big|\,\boldsymbol{x}\big] = 6 \quad \text{and} \quad \mathbb{E}\big[L(\theta, a_1)\big|\,\boldsymbol{x}\big] = 11 - 6\int_0^1 \theta^6 \pi(\theta|\boldsymbol{x})\, d\theta.$$

The Bayesian decision is to take action a_1 if

$$\int_0^1 \theta^6 \pi(\theta|\boldsymbol{x})\, d\theta = \frac{\int_0^1 \theta^{c+5+(x_1^2+x_2^2-x_1-x_2)/2}\big(1 - \mathbf{1}\{x_1 < 8\}\theta^{x_1}\big)\big(1 - \mathbf{1}\{x_2 < 8\}\theta^{x_2}\big)\, d\theta}{\int_0^1 \theta^{c-1+(x_1^2+x_2^2-x_1-x_2)/2}\big(1 - \mathbf{1}\{x_1 < 8\}\theta^{x_1}\big)\big(1 - \mathbf{1}\{x_2 < 8\}\theta^{x_2}\big)\, d\theta} > \frac{5}{6}.$$

Note: If we treat the above Bayesian decision based on (x_1, x_2) as a function, d say, of (x_1, x_2), then d becomes the Bayes rule found in §1.9.5(iii).

2.2.5 **Example §2.2.1**

An experiment has two possible outcomes: failure or success, with probabilities θ and $1-\theta$ respectively. The experiment has been repeated n times independently, and a total of x failures are observed.

It is given that a successful experiment yields a gain of G units and a failure incurs a loss of L units, where $G, L > 0$. Before the n repetitions of the experiment, the prior distribution of θ is chosen to be the beta (a, b) distribution, for a given pair of constants $a, b > 0$.

After having observed the x failures among the n trials, should we still continue the experiment, or should we put a stop to it?

The number X of failures among the n repetitions has the binomial (n, θ) distribution, so that

$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \ldots, n.$$

$$\theta \sim \textit{beta}\,(a,b) \ \textit{(prior)} \quad \Rightarrow \quad \pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, \theta^{a-1}(1-\theta)^{b-1}, \quad \theta \in [0,1].$$

Posterior pdf:

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \theta^x(1-\theta)^{n-x}\theta^{a-1}(1-\theta)^{b-1} \propto \theta^{a+x-1}(1-\theta)^{b+n-x-1}.$$

Hence, $\theta|X = x \ \sim \ \textit{beta}\,(a+x, b+n-x)$, i.e.

$$\pi(\theta|x) = \frac{\Gamma(a+b+n)}{\Gamma(a+x)\,\Gamma(b+n-x)}\, \theta^{a+x-1}(1-\theta)^{b+n-x-1}.$$

If we stop the experiment (action a_S), there will be zero loss. If we continue the experiment another time (action a_C), we have an expected net loss given by $\theta L - (1-\theta)G$. The loss function is therefore

$$L(\theta, a_S) = 0, \qquad L(\theta, a_C) = \theta L - (1-\theta)G.$$

Given observation x, the expected posterior losses incurred by the two actions are given by

$$\mathbb{E}[L(\theta, a_S)|x] = 0,$$
$$\mathbb{E}[L(\theta, a_C)|x] = \mathbb{E}[\theta|x]L - \big(1 - \mathbb{E}[\theta|x]\big)G = \frac{(a+x)L - (b+n-x)G}{a+b+n}.$$

The Bayesian decision is to **continue** *the experiment if*

$$\mathbb{E}[L(\theta, a_C)|x] < 0 = \mathbb{E}[L(\theta, a_S)|x],$$

or equivalently,

$$\frac{L}{G} < \frac{b+n-x}{a+x} = \frac{b + \textit{number of observed successes}}{a + \textit{number of observed failures}}.$$

Remarks:

- If we have strong prior belief in the success rate of the experiment (b large compared to a), then the Bayesian decision inclines towards continuation of the experiment, unless there are too many observed failures compared to successes (x large compared to $n-x$) among the previous n trials.
- If we have very little prior belief in the success rate of the experiment (b small compared to a), then the Bayesian decision inclines towards termination of the experiment, unless there are very few observed failures compared to successes (x small compared to $n-x$) among the previous n trials.
- If the experiment has been repeated a very large number n of times, then $x/n \approx \theta$ by the Strong Law of Large Numbers, so that
$$\frac{b+n-x}{a+x} \approx \frac{1-\theta}{\theta},$$
irrespective of the prior values of a and b. The Bayesian decision approaches the trivial instruction:

  continue *the experiment if* $L(\theta, a_C) < L(\theta, a_S)$,

  as if we had known θ.

§2.3 Bayesian statistical inference

2.3.1 *Point estimation of* $\boldsymbol{\theta} \in \mathbb{R}^k$ —

Consider two examples of loss function L:

(a) $L(\boldsymbol{\theta}, \boldsymbol{a}) = \|\boldsymbol{\theta} - \boldsymbol{a}\|_2^2$ ($\boldsymbol{a} \in \mathbb{R}^k$) $\Rightarrow$

$$\mathbb{E}\big[\|\boldsymbol{\theta} - \boldsymbol{a}\|_2^2|\boldsymbol{x}\big] = \mathbb{E}\big[\|\boldsymbol{\theta} - \mathbb{E}[\boldsymbol{\theta}|\boldsymbol{x}]\|_2^2 \,\big|\, \boldsymbol{x}\big] + \|\mathbb{E}[\boldsymbol{\theta}|\boldsymbol{x}] - \boldsymbol{a}\|_2^2$$

minimised by

$$\boldsymbol{a} = \mathbb{E}[\boldsymbol{\theta}|\boldsymbol{x}] = \begin{bmatrix} \mathbb{E}[\theta_1|\boldsymbol{x}] \\ \vdots \\ \mathbb{E}[\theta_k|\boldsymbol{x}] \end{bmatrix} = \textbf{posterior mean} \text{ of } \boldsymbol{\theta}.$$

(b) $L(\boldsymbol{\theta}, \boldsymbol{a}) = \|\boldsymbol{\theta} - \boldsymbol{a}\|_1$ ($\boldsymbol{a} \in \mathbb{R}^k$) $\Rightarrow$

$$\begin{aligned}\mathbb{E}\big[\|\boldsymbol{\theta} - \boldsymbol{a}\|_1|\boldsymbol{x}\big] &= \sum_{i=1}^k \Big\{\mathbb{E}\big[(a_i - \theta_i)\,\mathbf{1}\{\theta_i \le a_i\} \,\big|\, \boldsymbol{x}\big] + \mathbb{E}\big[(\theta_i - a_i)\,\mathbf{1}\{\theta_i > a_i\} \,\big|\, \boldsymbol{x}\big]\Big\} \\ &= 2\sum_{i=1}^k \Big(a_i\big\{\mathbb{P}(\theta_i \le a_i|\boldsymbol{x}) - 1/2\big\} - \mathbb{E}\big[\theta_i(\mathbf{1}\{\theta_i \le a_i\} - 1/2) \,\big|\, \boldsymbol{x}\big]\Big)\end{aligned}$$

minimised by $\boldsymbol{a}$ satisfying

$$\mathbb{P}(\theta_i \le a_i|\boldsymbol{x}) = 1/2, \quad i = 1, \ldots, k,$$

Thus, the Bayesian decision is to set a_i to be the **posterior median** of θ_i.

2.3.2 *Hypothesis testing about θ —*

Consider testing

$$H_0 : \theta \in \Theta_0 \qquad \text{vs} \qquad H_1 : \theta \in \Theta_1 \equiv \Theta \setminus \Theta_0.$$

Action space: $\mathcal{A} = \{a_0(\leftrightarrow \text{accept } H_0),\ a_1(\leftrightarrow \text{reject } H_0)\}$.
Define, for some $L_0, L_1 > 0$, the loss function

$$L(\theta, a_0) = L_1 \mathbf{1}\{\theta \in \Theta_1\}, \qquad L(\theta, a_1) = L_0 \mathbf{1}\{\theta \in \Theta_0\}.$$

Given $\boldsymbol{x}$, the expected posterior loss associated with action a_j $(j = 0, 1)$ is

$$\mathbb{E}[L(\theta, a_j)|\boldsymbol{x}] = L_{1-j}\, \mathbb{E}[\mathbf{1}\{\theta \in \Theta_{1-j}\}|\boldsymbol{x}] = L_{1-j}\, \mathbb{P}(\theta \in \Theta_{1-j}|\boldsymbol{x}).$$

The Bayesian decision is to choose a_j such that $L_{1-j}\, \mathbb{P}(\theta \in \Theta_{1-j}|\boldsymbol{x})$ is minimised, or equivalently, that

$$\frac{L_{1-j}}{L_j} < \frac{\mathbb{P}(\theta \in \Theta_j|\boldsymbol{x})}{1 - \mathbb{P}(\theta \in \Theta_j|\boldsymbol{x})},$$

i.e.

$$\textit{reject } H_0 \textit{ if } \ \mathbb{P}(H_1|\boldsymbol{x}) = \mathbb{P}(\theta \in \Theta_1|\boldsymbol{x}) = \int_{\Theta_1} \pi(\theta|\boldsymbol{x})\, d\theta > \frac{L_0}{L_0 + L_1}.$$

2.3.3 *Interval estimation of θ —*

(a) *Construct an interval of fixed length 2δ having maximum posterior coverage probability.*

Write a for the mid-point of this interval. We want to choose a to maximise the *posterior coverage probability*

$$\mathbb{P}(\theta \in [\, a - \delta, a + \delta\,] \mid \boldsymbol{x}).$$

If $\pi(\theta|\boldsymbol{x})$ is **unimodal**, then optimal a is given by

$$\pi(a - \delta \mid \boldsymbol{x}) = \pi(a + \delta \mid \boldsymbol{x}).$$

In practice, this gives a close to the mode of $\pi(\cdot|\boldsymbol{x})$.

(b) *Construct an interval of shortest length subject to some fixed posterior coverage probability* $\geq 1-\alpha$.

Let $[a, b]$ be the required interval. Find $a \leq b$ to minimise $b-a$, subject to

$$\mathbb{P}(a \leq \theta \leq b \mid \boldsymbol{x}) = \int_a^b \pi(\theta|\boldsymbol{x})\, d\theta \geq 1-\alpha.$$

If $\pi(\theta|\boldsymbol{x})$ is **unimodal** and continuous, then optimal a and b are solved by the simultaneous equations

$$\pi(a|\boldsymbol{x}) = \pi(b|\boldsymbol{x}) \quad \text{and} \quad \int_a^b \pi(\theta|\boldsymbol{x})\, d\theta = 1-\alpha.$$

(c) *Construct an equal-tailed interval of some fixed posterior coverage probability* $1-\alpha$.

Let $[a, b]$ be the required interval. Find $a \leq b$ which solve the simultaneous equations

$$\begin{cases} \mathbb{P}(\theta > b \mid \boldsymbol{x}) = \displaystyle\int_b^\infty \pi(\theta|\boldsymbol{x})\, d\theta = \alpha/2, \\ \mathbb{P}(\theta < a \mid \boldsymbol{x}) = \displaystyle\int_{-\infty}^a \pi(\theta|\boldsymbol{x})\, d\theta = \alpha/2. \end{cases}$$

2.3.4 Example §2.3.1

Let $\theta \in [0,1]$ be an unknown parameter, which is assigned a prior beta(a, b) distribution for given constants $a, b > 0$.

The observation x is generated from the binomial(n, θ) distribution for a known n.

We have shown in Example §2.2.1 that given x, θ has a beta$(a+x, b+n-x)$ posterior distribution with pdf

$$\pi(\theta|x) = \frac{\Gamma(a+b+n)}{\Gamma(a+x)\,\Gamma(b+n-x)}\,\theta^{a+x-1}(1-\theta)^{b+n-x-1}\,\mathbf{1}\{0 \leq \theta \leq 1\}.$$

We are interested in drawing Bayesian inference about the odds $\psi = \theta/(1-\theta)$.

1. *Point estimation of* ψ —

   Posterior mean:

$$\begin{aligned} \mathbb{E}[\psi|x] &= \int_0^1 \left(\frac{\theta}{1-\theta}\right)\pi(\theta|x)\, d\theta = \frac{\Gamma(a+b+n)}{\Gamma(a+x)\,\Gamma(b+n-x)} \int_0^1 \theta^{a+x}(1-\theta)^{b+n-x-2}\, d\theta \\ &= \frac{\Gamma(a+b+n)}{\Gamma(a+x)\,\Gamma(b+n-x)}\,\frac{\Gamma(a+x+1)\,\Gamma(b+n-x-1)}{\Gamma(a+b+n)} = \frac{a+x}{b+n-x-1}, \end{aligned}$$

provided that $b+n-x>1$. Posterior median M satisfies

$$1/2 = \mathbb{P}(\psi \le M|x) = \mathbb{P}\left(\theta \le \frac{M}{M+1}\middle| x\right) = \int_0^{M/(M+1)} \pi(\theta|x)\, d\theta.$$

2. *Test $H_0 : \psi \le \psi_0$ vs $H_1 : \psi > \psi_0$* —

   Assume losses C_1 and C_2 for types I and II errors, respectively.

   Reject H_0 if

$$\mathbb{P}(\psi > \psi_0|x) = \int_{\psi_0/(\psi_0+1)}^{1} \pi(\theta|x)\, d\theta > \frac{C_1}{C_1+C_2}.$$

3. *Interval estimation of ψ* —

   Posterior pdf of ψ:

$$\pi^*(\psi|x) = \pi\left(\frac{\psi}{\psi+1}\middle| x\right) \frac{d}{d\psi}\left(\frac{\psi}{\psi+1}\right) = \frac{\Gamma(a+b+n)}{\Gamma(a+x)\,\Gamma(b+n-x)}\, \frac{\psi^{a+x-1}}{(\psi+1)^{a+b+n}}$$

   is unimodal in ψ, with *posterior mode* at $\psi = \max\left\{0, \dfrac{a+x-1}{1+b+n-x}\right\}$.

   (a) $\boxed{\text{Interval for } \psi \text{ of fixed length } 2\delta \text{ and maximum posterior coverage probability}}$

   If $a+x-1 \le 0$, then the interval is $[0, 2\delta]$.
   If $a+x-1 > 0$, then the interval is $[c-\delta, c+\delta]$, where c satisfies

$$\pi^*(c-\delta|x) = \pi^*(c+\delta|x), \quad \text{i.e.} \quad \left(\frac{c-\delta}{c+\delta}\right)^{a+x-1} = \left(\frac{c-\delta+1}{c+\delta+1}\right)^{a+b+n}.$$

   (b) $\boxed{\text{Shortest interval for } \psi \text{ of fixed posterior coverage probability} \ge 1-\alpha}$

   If $a+x-1 \le 0$, then the interval is $[0, v]$, where v satisfies

$$\int_0^v \pi^*(\psi|x)\, d\psi = 1-\alpha, \quad \text{or equivalently,} \quad \int_0^{v/(v+1)} \pi(\theta|x)\, d\theta = 1-\alpha.$$

   If $a+x-1 > 0$, then the interval is $[u, v]$, where u, v satisfy

$$\begin{cases} \pi^*(u|x) = \pi^*(v|x), \\ \displaystyle\int_u^v \pi^*(\psi|x)\, d\psi = 1-\alpha, \end{cases} \quad \text{or equivalently,} \quad \begin{cases} \left(\dfrac{u}{v}\right)^{a+x-1} = \left(\dfrac{u+1}{v+1}\right)^{a+b+n}, \\ \displaystyle\int_{u/(u+1)}^{v/(v+1)} \pi(\theta|x)\, d\theta = 1-\alpha. \end{cases}$$

(c) $\boxed{\text{Equal-tailed interval for } \psi \text{ of fixed posterior coverage probability } 1-\alpha}$

The required interval is $[u, v]$ where

$$\int_0^u \pi^*(\psi|x)\,d\psi = \int_v^\infty \pi^*(\psi|x)\,d\psi = \alpha/2,$$

or equivalently,

$$\int_0^{u/(u+1)} \pi(\theta|x)\,d\theta = \int_{v/(v+1)}^1 \pi(\theta|x)\,d\theta = \alpha/2.$$

2.3.5 Example §2.3.2

Consider the parameter space $\Theta = \big\{(\theta_1,\theta_2) : \theta_1,\theta_2 \ge 0,\ \theta_1+\theta_2 \le 1\big\}$ and a hypothesis H which states that the true parameter $\boldsymbol{\theta}$ is in $\Theta_H = \big\{\boldsymbol{\theta}\in\Theta : \theta_1 > \max\{\theta_2, 1-\theta_1-\theta_2\}\big\}$.

Let X follow the binomial $\big(n;\theta_1\big)$ distribution, with the mass function

$$f(x|\boldsymbol{\theta}) = \frac{n!}{x!\,(n-x)!}\,\theta_1^x(1-\theta_1)^{n-x}, \quad x = 0,1,\ldots,n.$$

We are interested in the *odds* in favour of H.

Let $\beta > 0$ be fixed. Assume a prior pdf

$$\pi(\boldsymbol{\theta}) \propto \beta\,\mathbf{1}\{\boldsymbol{\theta}\in\Theta_H\} + \mathbf{1}\{\boldsymbol{\theta}\notin\Theta_H\}, \quad \theta\in\Theta.$$

Noting that H requires "$1-2\theta_1 \le \theta_2 \le \theta_1$", the *prior odds* of H is given by

$$\frac{\mathbb{P}(H)}{1-\mathbb{P}(H)} = \frac{\mathbb{P}(\boldsymbol{\theta}\in\Theta_H)}{\mathbb{P}(\boldsymbol{\theta}\notin\Theta_H)} = \frac{\beta\big\{\int_{1/3}^{1/2}(3\theta_1-1)\,d\theta_1 + \int_{1/2}^1(1-\theta_1)\,d\theta_1\big\}}{\int_0^1(1-\theta_1)\,d\theta_1 - \int_{1/3}^{1/2}(3\theta_1-1)\,d\theta_1 - \int_{1/2}^1(1-\theta_1)\,d\theta_1}$$
$$= \frac{\beta/6}{1/2-1/6} = \beta/2.$$

Based on data $X=x$, $\boldsymbol{\theta}$ has the posterior pdf

$$\pi(\boldsymbol{\theta}|x) \propto \pi(\boldsymbol{\theta})f(x|\boldsymbol{\theta}) \propto \big[\beta\,\mathbf{1}\{\boldsymbol{\theta}\in\Theta_H\} + \mathbf{1}\{\boldsymbol{\theta}\notin\Theta_H\}\big]\theta_1^x(1-\theta_1)^{n-x},$$

so that the *posterior odds* of H becomes

$$\frac{\mathbb{P}(H|x)}{1-\mathbb{P}(H|x)} = \frac{\beta\big\{\int_{1/3}^{1/2}\theta_1^x(1-\theta_1)^{n-x}(3\theta_1-1)\,d\theta_1 + \int_{1/2}^1\theta_1^x(1-\theta_1)^{n-x+1}\,d\theta_1\big\}}{\int_0^{1/2}\theta_1^x(1-\theta_1)^{n-x+1}\,d\theta_1 - \int_{1/3}^{1/2}\theta_1^x(1-\theta_1)^{n-x}(3\theta_1-1)\,d\theta_1}.$$

Note: The ratio $\dfrac{\text{posterior odds}}{\text{prior odds}}$ is generally known as the *Bayes factor*.

2.3.6 *Predictive distribution* —

Suppose we wish to predict some future observation Y which has the conditional probability function $g(y|\theta, \boldsymbol{x})$ given the observed data $\boldsymbol{x}$.

The posterior joint probability function of (Y, θ) is $g(y|\theta, \boldsymbol{x})\, \pi(\theta|\boldsymbol{x})$. Integrating out θ yields the probability function of the *posterior predictive distribution* of Y:

$$g^*(y|\boldsymbol{x}) = \int_\Theta g(y|\theta, \boldsymbol{x})\, \pi(\theta|\boldsymbol{x})\, d\theta = \mathbb{E}\big[g(y|\theta, \boldsymbol{x})\big|\boldsymbol{x}\big].$$

With $g^*(y|\boldsymbol{x})$ available, we can predict Y using the mean, median, mode or any other location parameter of $g^*(\cdot|\boldsymbol{x})$, possibly derived from a loss function.

2.3.7 **Example §2.3.3**

Consider a total of $\theta_1 + \theta_2$ tasks, among which θ_1 are of the K-type and θ_2 of the H-type. It is known that $\theta_1 \in \{0, 1, \ldots, m\}$ and $\theta_2 \in \{0, 1\}$, but the true values of θ_1 and θ_2 are unknown. Note that there can be at most one H-type task. It is known that each task has a probability $(1 - \theta_3)^{i-1}\theta_3$ to be completed on day i, for $i = 1, 2, \ldots$, and for some unknown $\theta_3 \in (0, 1)$. Completion times of the tasks are independent of each other.

For each $i = 1, 2, \ldots$, let K_i and H_i denote respectively the numbers of K-type and H-type tasks completed on day i. Thus, a total of $\sum_{i=1}^n K_i$ K-type tasks and $\sum_{i=1}^n H_i$ H-type tasks have been completed by the end of day n.

Suppose that our prior belief specifies that $\theta_1 \sim$ binomial$\,(m, q)$, $\theta_2 \sim$ Bernoulli$\,(q)$ and $\theta_3 \sim U(0, 1)$, for some given $q \in (0, 1)$, and that $\theta_1, \theta_2, \theta_3$ are independent of each other. Then the prior joint probability function of $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ is given by

$$\pi(\boldsymbol{\theta}) = \binom{m}{\theta_1} q^{\theta_1}(1-q)^{m-\theta_1} \times q^{\theta_2}(1-q)^{1-\theta_2} \times \mathbf{1}\big\{\theta_1 \in \{0, 1, \ldots, m\},\ \theta_2 \in \{0, 1\},\ 0 < \theta_3 < 1\big\}.$$

Given observation of the data $(K_1, H_1), \ldots, (K_n, H_n)$ on the first n days, we wish to predict H_{n+1} based on its posterior predictive distribution.

Note that the data $(K_1, H_1), \ldots, (K_n, H_n)$ follow the joint mass function

$$\begin{aligned} & f\big((K_1, H_1), \ldots, (K_n, H_n)\big|\boldsymbol{\theta}\big) \\ &= \frac{\theta_1!}{K_1! \cdots K_n!\,(\theta_1 - \sum_{i=1}^n K_i)!}\,(1-\theta_3)^{n\{\theta_1+\theta_2-\sum_{i=1}^n(H_i+K_i)\}} \prod_{i=1}^n \big\{(1-\theta_3)^{i-1}\theta_3\big\}^{H_i+K_i} \\ &\quad \times \mathbf{1}\Big\{\sum_{i=1}^n K_i \le \theta_1,\ \sum_{i=1}^n H_i \le \theta_2\Big\}. \end{aligned}$$

The posterior joint probability function of $\boldsymbol{\theta}$ is given by

$$
\begin{aligned}
&\pi\big(\boldsymbol{\theta}\big|(K_1,H_1),\ldots,(K_n,H_n)\big) \propto \pi(\boldsymbol{\theta}) f\big((K_1,H_1),\ldots,(K_n,H_n)\big|\boldsymbol{\theta}\big)\\
&\propto \binom{m-\sum_{i=1}^n K_i}{m-\theta_1}\Big(\frac{q}{1-q}\Big)^{\theta_1+\theta_2}\theta_3^{\sum_{i=1}^n (H_i+K_i)}(1-\theta_3)^{n(\theta_1+\theta_2)-\sum_{i=1}^n (n-i+1)(H_i+K_i)}\\
&\quad\times \mathbf{1}\Big\{\sum_{i=1}^n K_i \le \theta_1 \le m,\ \sum_{i=1}^n H_i \le \theta_2 \le 1,\ 0<\theta_3<1\Big\}.
\end{aligned}
$$

If $\sum_{i=1}^n H_i = 1$, then $H_{n+1}=0$ for sure.

Consider henceforth the nontrivial case where $\sum_{i=1}^n H_i = 0$, i.e. $H_1=\cdots=H_n=0$. The conditional mass function of H_{n+1} is then given by

$$
g\big(y\big|\boldsymbol{\theta},(K_1,0),\ldots,(K_n,0)\big) = (\theta_2\theta_3)^y(1-\theta_2\theta_3)^{1-y}, \quad y\in\{0,1\}.
$$

Thus, the posterior predictive distribution of H_{n+1} has the mass function:

$$
\begin{aligned}
&g^*\big(y\big|(K_1,0),\ldots,(K_n,0)\big)\\
&= \sum_{\theta_1=\sum_{i=1}^n K_i}^{m}\ \sum_{\theta_2=0}^{1}\int_0^1 g\big(y\big|\boldsymbol{\theta},(K_1,0),\ldots,(K_n,0)\big)\pi\big(\boldsymbol{\theta}\big|(K_1,H_1),\ldots,(K_n,H_n)\big)\,d\theta_3\\
&\propto \sum_{\theta_1=\sum_{i=1}^n K_i}^{m}\binom{m-\sum_{i=1}^n K_i}{m-\theta_1}\Big(\frac{q}{1-q}\Big)^{\theta_1}\Big[(1-y)\int_0^1 \theta_3^{\sum_{i=1}^n K_i}(1-\theta_3)^{n\theta_1-\sum_{i=1}^n (n-i+1)K_i}d\theta_3\\
&\quad+\Big(\frac{q}{1-q}\Big)\int_0^1 \theta_3^{y+\sum_{i=1}^n K_i}(1-\theta_3)^{1-y+n\theta_1+n-\sum_{i=1}^n (n-i+1)K_i}d\theta_3\Big].
\end{aligned}
$$

In particular, we have

$$
\begin{aligned}
&g^*\big(1\big|(K_1,0),\ldots,(K_n,0)\big) = \Big(\frac{q}{1-q}\Big)\Big(\frac{1+\sum_{i=1}^n K_i}{2+\sum_{i=1}^n K_i}\Big)\\
&\times \sum_{\theta_1=\sum_{i=1}^n K_i}^{m}\binom{m-\sum_{i=1}^n K_i}{m-\theta_1}\Big(\frac{q}{1-q}\Big)^{\theta_1}\binom{2+n\theta_1+n-\sum_{i=1}^n (n-i)K_i}{2+\sum_{i=1}^n K_i}^{-1}\\
&\times\Bigg\{\sum_{\theta_1=\sum_{i=1}^n K_i}^{m}\binom{m-\sum_{i=1}^n K_i}{m-\theta_1}\Big(\frac{q}{1-q}\Big)^{\theta_1}\\
&\quad\times\Bigg[\Big(\frac{q}{1-q}\Big)\binom{1+n\theta_1+n-\sum_{i=1}^n (n-i)K_i}{1+\sum_{i=1}^n K_i}^{-1}+\binom{1+n\theta_1-\sum_{i=1}^n (n-i)K_i}{1+\sum_{i=1}^n K_i}^{-1}\Bigg]\Bigg\}^{-1}.
\end{aligned}
$$

§2.4 Marginal posterior distribution

2.4.1 If $\boldsymbol{\theta}$ is a vector of more than one components, i.e. $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$, but interest is only in a subset of components, such as $(\theta_1, \ldots, \theta_r)$ $(r < k)$, we can find the *marginal* posterior probability function of $(\theta_1, \ldots, \theta_r)$ by calculating

$$\pi_{1,\ldots,r}(\theta_1, \ldots, \theta_r | \boldsymbol{x}) = \int \cdots \int \pi(\theta_1, \theta_2, \ldots, \theta_k | \boldsymbol{x})\, d\theta_{r+1} \cdots d\theta_k,$$

and draw Bayesian inference about $(\theta_1, \ldots, \theta_r)$ accordingly.

2.4.2 **Example §2.3.3** (cont'd)

Bayesian inference is to be made about θ_2 based on its posterior (marginal) distribution, given observation of the data $(K_1, H_1), \ldots, (K_n, H_n)$ on the first n days.

We have earlier derived an expression proportional to $\pi\big(\boldsymbol{\theta}\big|(K_1, H_1), \ldots, (K_n, H_n)\big)$. The posterior marginal mass function of θ_2 is then given by

$$\pi_2\big(\theta_2\big|(K_1, H_1), \ldots, (K_n, H_n)\big) = \sum_{\theta_1 = \sum_{i=1}^n K_i}^{m} \int_0^1 \pi\big(\boldsymbol{\theta}\big|(K_1, H_1), \ldots, (K_n, H_n)\big)\, d\theta_3$$

$$\propto \sum_{\theta_1 = \sum_{i=1}^n K_i}^{m} \binom{m - \sum_{i=1}^n K_i}{m - \theta_1} \left(\frac{q}{1-q}\right)^{\theta_1 + \theta_2} \int_0^1 \theta_3^{\sum_{i=1}^n (H_i + K_i)} (1 - \theta_3)^{n(\theta_1 + \theta_2) - \sum_{i=1}^n (n-i+1)(H_i + K_i)} d\theta_3$$

$$\propto \sum_{\theta_1 = \sum_{i=1}^n K_i}^{m} \binom{m - \sum_{i=1}^n K_i}{m - \theta_1} \left(\frac{q}{1-q}\right)^{\theta_1 + \theta_2} \binom{n(\theta_1 + \theta_2) + 1 - \sum_{i=1}^n (n-i)(H_i + K_i)}{1 + \sum_{i=1}^n (H_i + K_i)}^{-1}.$$

for $\sum_{i=1}^n H_i \le \theta_2 \le 1$.

Thus, the posterior odds *for* the existence of a H-type task $(\theta_2 = 1)$ is

$$\frac{\pi_2\big(1\big|(K_1, H_1), \ldots, (K_n, H_n)\big)}{\pi_2\big(0\big|(K_1, H_1), \ldots, (K_n, H_n)\big)}$$

$$= \frac{\dfrac{q}{1-q} \displaystyle\sum_{\theta_1 = \sum_{i=1}^n K_i}^{m} \binom{m - \sum_{i=1}^n K_i}{m - \theta_1} \left(\frac{q}{1-q}\right)^{\theta_1} \binom{n(\theta_1 + 1) + 1 - \sum_{i=1}^n (n-i)(H_i + K_i)}{1 + \sum_{i=1}^n (H_i + K_i)}^{-1}}{\mathbf{1}\Big\{\displaystyle\sum_{i=1}^n H_i = 0\Big\} \displaystyle\sum_{\theta_1 = \sum_{i=1}^n K_i}^{m} \binom{m - \sum_{i=1}^n K_i}{m - \theta_1} \left(\frac{q}{1-q}\right)^{\theta_1} \binom{n\theta_1 + 1 - \sum_{i=1}^n (n-i) K_i}{1 + \sum_{i=1}^n K_i}^{-1}}.$$

Note: If $H_j = 1$ for some $j \in \{1, \ldots, n\}$, then the above posterior odds for $\theta_2 = 1$ becomes ∞, which is an obvious result as there must be one H-type task if one is completed on day j.