Example Class 4

LIU Chen

Department of Statistics and Actuarial Science,
The University of Hong Kong

October 11, 2022

Bayesian Approach: Prior to posterior

Bayesian approach

- X: observable random variate with probability function $f(x \mid \theta)$
- θ: unobservable random variate with a specified prior probability function $\pi(\theta)$:

$$\int_{\Theta} \pi(\theta) d\theta = 1 \quad (\text{ continuous } \theta), \qquad \text{or}$$

$$\sum_{\theta \in \Theta} \pi(\theta) = 1 \quad (\text{ discrete } \theta)$$

Bayesian Approach: Prior to posterior

Posterior Probability Function

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

$$\begin{aligned} \pi(\theta \mid x) &= \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{\int_\Theta f\left(x \mid \theta'\right) \pi\left(\theta'\right) d\theta'} \\ &\propto f(\boldsymbol{x} \mid \theta)\pi(\theta) \end{aligned}$$

Bayesian Approach: **Expected Posterior Loss**

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

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

Bayesian Approach: **Bayesian decision**

Definition.

A Bayesian decision is to take an action $a \in \mathcal{A}$ which minimises the expected posterior loss $\mathbb{E}[L(\theta, a) \mid \boldsymbol{x}]$

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

$$\begin{aligned}
\mathbb{E}[L(\theta, a) \mid \boldsymbol{x}] &= \int_\Theta L(\theta, a) \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{f(\boldsymbol{x})} d\theta \\
&= \frac{1}{f(\boldsymbol{x})} \int_\Theta L(\theta, a) f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta
\end{aligned}$$

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

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

Consider two examples of loss function L:

1. $L(\boldsymbol{\theta}, \boldsymbol{a}) = \|\boldsymbol{\theta} - \boldsymbol{a}\|_2^2 \left(\boldsymbol{a} \in \mathbb{R}^k\right)$ is minimized by
$$\boldsymbol{a} = \mathbb{E}[\boldsymbol{\theta} \mid \boldsymbol{x}] = \begin{bmatrix} \mathbb{E}\left[\theta_1 \mid \boldsymbol{x}\right] \\ \vdots \\ \mathbb{E}\left[\theta_k \mid \boldsymbol{x}\right] \end{bmatrix} = \text{posterior mean of } \boldsymbol{\theta}$$
2. $L(\boldsymbol{\theta}, \boldsymbol{a}) = \|\boldsymbol{\theta} - \boldsymbol{a}\|_1 \left(\boldsymbol{a} \in \mathbb{R}^k\right)$ is minimized by $\boldsymbol{a}$ satisfying $\mathbb{P}\left(\theta_i \leq a_i \mid \boldsymbol{x}\right) = 1/2, \quad i = 1, \ldots, k$. Thus, the Bayesian decision is to set a_i to be the posterior median of θ_i

Hypothesis testing about θ

Consider testing:

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1 \equiv \Theta \backslash \Theta_0$$

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**

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

Given x, the **expected posterior loss**

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

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

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

i.e.

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

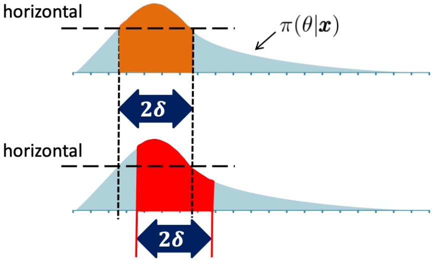
Interval estimation of θ I

(a) Fix length $= 2\delta$ $\rightarrow$ maximise *posterior coverage probability*

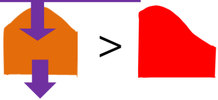
Any candidate can be represented by a

Maximise: $\underbrace{\int_{a-\delta}^{a+\delta} \pi(\theta|\boldsymbol{x})\, d\theta}_{\text{posterior coverage probability of } [a-\delta, a+\delta]} \propto \int_{a-\delta}^{a+\delta} \pi(\theta) f(\boldsymbol{x}|\theta)\, d\theta$ w.r.t. a

Special case: unimodal $\pi(\theta|x)$



optimal solution



$[a-\delta, a+\delta]$ satisfying:

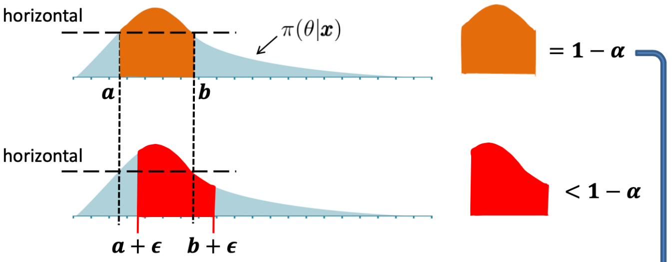
$$\boxed{\pi(a-\delta|\boldsymbol{x}) = \pi(a+\delta|\boldsymbol{x})}$$

(assuming $\pi(\theta|\boldsymbol{x})$ continuous
at $\theta = a-\delta,\ a+\delta$)

Interval estimation of θ II

(b) Desire posterior coverage probability $\geq 1-\alpha \quad \rightarrow \quad$ minimise *length*

Special case: unimodal $\pi(\theta|x)$



Optimal solution:

$[a, b]$ satisfying: $\boxed{\pi(a|\boldsymbol{x}) = \pi(b|\boldsymbol{x}) \quad \& \quad \int_a^b \pi(\theta|\boldsymbol{x})\, d\theta = 1-\alpha}$

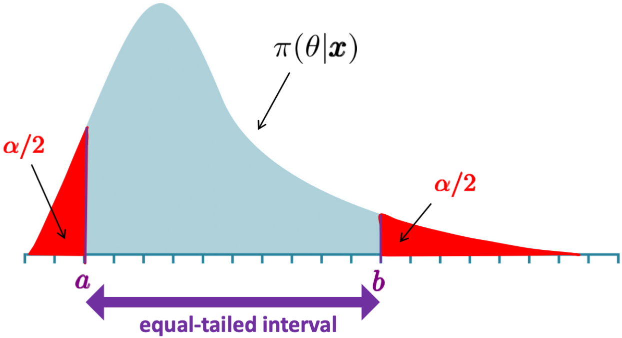
(assuming $\pi(\theta|\boldsymbol{x})$ continuous at $\theta = a,\, b$)

Interval estimation of θ III

(c) Fix posterior coverage probability $= 1 - \alpha$ & require "*equal-tailed*"

i.e.

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



Predictive distribution

$\boldsymbol{\theta} \sim \pi(\theta)$ (prior)

Observation

$\boldsymbol{X} \sim f(\boldsymbol{x}|\theta)$

$\Rightarrow$ (posterior) $\pi(\theta|\boldsymbol{X}) \propto \pi(\theta) f(\boldsymbol{X}|\theta)$

unobserved/future $Y|(\theta, \boldsymbol{X}) \sim g(y|\theta, \boldsymbol{X})$

(conditional probability function of Y given $(\theta, \boldsymbol{X})$)

$$(\theta, Y)|\boldsymbol{X} \sim g(y|\theta, \boldsymbol{X})\pi(\theta|\boldsymbol{X})$$

(*interested in predicting* Y)

$$g^*(y|\boldsymbol{X}) \triangleq \int_\Theta g(y|\theta, \boldsymbol{X})\, \pi(\theta|\boldsymbol{X})\, d\theta$$

posterior predictive distribution of Y

Linear Model Basics

- A linear model includes an $n \times 1$ response vector $\boldsymbol{y} = (y_1, \ldots, y_n)$ and an $n \times p$ design matrix (regressors) $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p]$. The relationship between $\boldsymbol{y}$ and $\boldsymbol{X}$ has the form,

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}, \quad \epsilon \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}).$$

- From standard statistical analysis, the classical unbiased and least-square estimates of the regression parameter β and σ^2 are

$$\hat{\beta} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{y}; \quad \hat{\sigma}^2 = \frac{1}{n-p}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})^\intercal (\boldsymbol{y} - \boldsymbol{X}\hat{\beta})$$

- The predicted value of $\boldsymbol{y}_0$ given $\boldsymbol{X}_0$ is

$$\hat{\boldsymbol{y}}_0 = \boldsymbol{X}_0 \hat{\beta} = \boldsymbol{X}_0 (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{y}$$

Hypothesis testing under linear model

$$H_0 : \beta_j = c \quad \text{versus} \quad H_1 : \beta_j \neq c$$

- **Wald t-test**

$$T = \frac{\hat{\beta}_j - c}{\sqrt{var(\hat{\beta}_j)}} \overset{\text{under } H_0}{\sim} t(n-p)$$

- **Likelihood ratio test**

$$LR = -2 \ln \left(\frac{L(\hat{\beta}_{H_0})}{L(\hat{\beta})} \right) \overset{\text{under } H_0}{\sim} \chi^2(1)$$

Bayesian linear model

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}, \quad \epsilon \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

$$\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2 \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$$

$$\beta_j \sim N(0, \sigma_0^2),\ j = 1, \dots, p$$

$$\sigma^2 \sim InvGamma(\xi, \xi)$$

Bayesian linear model (cont'd)

$$
\begin{aligned}
f(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{X}, \boldsymbol{y}) \propto & f(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) \pi(\boldsymbol{\beta}) \pi(\sigma^2) \\
\propto & (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\intercal (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2} \right) \exp\left(-\frac{\boldsymbol{\beta}^\intercal \boldsymbol{\beta}}{2\sigma_0^2} \right) \\
& (\sigma^2)^{-\xi - 1} \exp(-\frac{\xi}{\sigma^2}) \\
f(\sigma^2 | \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\beta}) \propto & (\sigma^2)^{-\xi - n/2 - 1} \exp\left(-\frac{\xi + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\intercal (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{\sigma^2} \right) \\
f(\boldsymbol{\beta} | \boldsymbol{X}, \boldsymbol{y}, \sigma^2) \propto & \exp\left(-\frac{\boldsymbol{\beta}^\intercal \boldsymbol{\beta}}{2\sigma_0^2} - \frac{\boldsymbol{\beta}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{y}^\intercal \boldsymbol{X}\boldsymbol{\beta}}{2\sigma^2} \right)
\end{aligned}
$$

Bayesian linear model (cont'd)

Sometimes prior distributions might not be valid distributions. For example,

$$\pi(\boldsymbol{\beta}) \propto 1 \longleftrightarrow N(0, \sigma^2), \sigma^2 \to \infty$$

$$\pi(\sigma^2) \propto 1/\sigma^2 \longleftrightarrow InvGamma(\xi, \xi), \xi \to 0$$

Even if the priors are improper, as long as the resulting posterior distribution are valid we can still conduct legitimate statistical inference on them

Predicting from linear models

- New covariates $\tilde{\boldsymbol{X}}$, and wish to predict the corresponding outcome $\tilde{y}$.
- If $\boldsymbol{\beta}$ and σ^2 are known, then $\tilde{y} \sim N(\tilde{\boldsymbol{X}}\beta, \sigma^2)$
- When parameters are unknown,

$$f(\tilde{y}|\boldsymbol{y}) = \int f(\tilde{y}|\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) d\boldsymbol{\beta} d\sigma^2$$

- For each posterior draw of $(\boldsymbol{\beta}_{(i)}, \sigma^2_{(i)})_{i=1}^M$, draw $\tilde{y}_i$ from $N(\tilde{\boldsymbol{X}}\beta_{(i)}, \sigma^2_{(i)})$. The resulting samples $(\tilde{y}_{(i)})_{i=1}^M$ represent the predictive distribution.

Bayesian Linear Regression (Example)

- Dataset: "data.csv"
- Linear regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$.

- Frequentist MLE:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y} \approx (1.085, 0.384, 2.445)^T,$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}})^2 \approx 0.326.$$

Bayesian Linear Regression (Example)

- Bayesian paradigm: $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_0, \sigma_0^2 \boldsymbol{I}_3)$ and $\sigma^2 \sim IG(\xi_0, \xi_0)$ where $\boldsymbol{I}_n$ denote the n-dimensional identity matrix, $\boldsymbol{\mu}_0 = \mathbf{0}$, $\sigma_0 = 10$, $\xi_0 = 0.1$
- Gibbs sampler:

$$
\begin{aligned}
p(\sigma^2 | \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{y}) &\propto \left(\frac{1}{\sigma^2}\right)^{n/2+\xi_0+1} e^{-\frac{1}{\sigma^2}[(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})/2+\xi_0]} \\
&\sim IG(n/2+\xi_0, (\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})/2+\xi_0) \\
p(\boldsymbol{\beta} | \boldsymbol{\mu}_0, \sigma_0, \sigma, \boldsymbol{X}, \boldsymbol{y}) &\propto e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\Sigma}\boldsymbol{\eta})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}-\boldsymbol{\Sigma}\boldsymbol{\eta})} \\
&\sim N(\boldsymbol{\Sigma}\boldsymbol{\eta}, \boldsymbol{\Sigma}), \\
\boldsymbol{\eta} &= \boldsymbol{\mu}_0/\sigma_0^2 + \boldsymbol{X}^T\boldsymbol{y}/\sigma^2, \\
\boldsymbol{\Sigma} &= (\boldsymbol{I}_3/\sigma_0^2 + \boldsymbol{X}^T\boldsymbol{X}/\sigma^2)^{-1}
\end{aligned}
$$

Bayesian Linear Regression (Example)

	β_0	β_1	β_2	σ^2
mean	1.085	0.382	2.445	0.334
variance	0.00733	0.02331	0.00184	0.00116

Table 1: Posterior means and variances of unknown parameters.

Bayesian Linear Regression (Example, with an interaction)

- MLE results. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$
- $\hat{\beta}_0 = 1.007, \quad \hat{\beta}_1 = 0.518, \quad \hat{\beta}_2 = 2.037, \quad \hat{\beta}_3 = 0.754, \quad \hat{\sigma}^2 = 0.295$
- Posterior means and variances of unknown parameters are listed below.

	β_0	β_1	β_2	β_3	σ^2
mean	1.007	0.518	2.035	0.756	0.305
variance	0.00693	0.02193	0.00966	0.02741	0.00098

Table 2: Posterior means and variances of unknown parameters.

Bayesian Linear Regression (Example, with an interaction)

- Two-sided hypothesis test:

$$H_0 : \beta_3 = 0 \quad \text{versus} \quad H_1 : \beta_3 \neq 0.$$

- Likelihood ratio test (LRT):

$$T_{\mathrm{LRT}} = (n-4)(\Lambda^{-2/n} - 1) = \frac{\boldsymbol{y}^T(\boldsymbol{H} - \boldsymbol{H}_{-4})\boldsymbol{y}}{\boldsymbol{y}^T(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}/(n-4)} \sim F(1, n-4), \tag{1}$$

  where $\boldsymbol{H}$ and $\boldsymbol{H}_{-4}$ denote the hat matrices with and without the interaction term.

- Wald test:

$$T_{\mathrm{Wald}} = \frac{\hat{\beta}_3^2}{\hat{\sigma}^2 \boldsymbol{e}_4^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{e}_4}, \tag{2}$$

  where $\boldsymbol{e}_4 = (0, 0, 0, 1)^T$.

Bayesian Linear Regression (Example, with an interaction)

$$T_{\mathrm{Wald}} = \frac{\hat{\beta}_3}{\hat{\sigma}\sqrt{(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}_{3,3}}} \sim t(200-4) = t(196)$$

$$p\text{-value}_{\mathrm{Wald}} = 1.002 \times 10^{-5}$$

$$-2\ln\left(\frac{L(\hat{\beta}_{H_0})}{L(\hat{\beta})}\right) \overset{\text{under } H_0}{=} -2\ln\frac{L(\hat{\boldsymbol{\beta}}_{H_0}, \hat{\sigma}_{H_0})}{L(\hat{\boldsymbol{\beta}}, \hat{\sigma})} \sim \chi^2(1)$$

$$L(\boldsymbol{\beta}, \sigma) = (2\pi)^{-n/2}\sigma^{-n}\exp\left(-\frac{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\intercal(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2}\right)$$

$$p\text{-value}_{\mathrm{LRT}} = 7.926 \times 10^{-6}$$

Bayesian Linear Regression (Example, with an interaction)

- The p-value under the two-sided hypothesis test is

$$p\text{-value}_2 = 2 - 2\Phi(|Z|) = 2[1 - \max\{\Phi(Z), \Phi(-Z)\}].$$

We define the two-sided posterior probability (PoP_2) as

$$\mathrm{PoP}_2 = 2[1 - \max\{\Pr(\beta_3 > 0|y), \Pr(\beta_3 < 0|y)\}].$$

- $\mathrm{PoP}_2 = 0$

Coin Tossing Problem

- We consider an experiment in which a coin was tossed 12 times, with 9 heads and 3 tails observed.
- Let θ be the probability of observing a head for a toss of the coin
- Suppose that we conduct a one-sided hypothesis test,

$$H_0 : \theta \leq 0.5 \quad \text{versus} \quad H_1 : \theta > 0.5.$$

- Based on the binomial or negative binomial likelihood, the frequentist hypothesis test yields conflicting results under the significance level of $\alpha = 0.05$: The null hypothesis is accepted under the binomial distribution, but it is rejected under the negative binomial distribution.

Coin Tossing Problem (Binomial)

- For $Y \sim \mathrm{Bin}(n, \theta)$, we have

$$\begin{aligned}
\Pr(Y \geq y \mid \theta) &= \sum_{k=y}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} \\
&= \frac{\Gamma(n+1)}{\Gamma(n-y+1)\Gamma(y)} \int_0^\theta t^{y-1}(1-t)^{n-y} dt = I_\theta(y, n-y+1)
\end{aligned}$$

  where $I_x(a, b)$ is the regularized incomplete beta function defined as

$$\begin{aligned}
I_x(a, b) &= \frac{B(x; a, b)}{B(a, b)} \\
B(x; a, b) &= \int_0^x t^{a-1}(1-t)^{b-1} dt \\
B(a, b) &= \int_0^1 t^{a-1}(1-t)^{b-1} dt
\end{aligned}$$

- $$p\text{−value}_{\mathrm{Bin}} = I_{0.5}(y, n-y+1)$$

Coin Tossing Problem (Negative-Binomial)

- For $Y \sim \mathrm{NB}(r, \theta)$, we have

$$\begin{aligned}
\Pr(Y \geq y \mid r, \theta) &= \sum_{k=y}^{\infty} \binom{k+r-1}{k} \theta^k (1-\theta)^r \\
&= \frac{\Gamma(y+r)}{\Gamma(r)\Gamma(y)} \int_0^{\theta} t^{y-1}(1-t)^{r-1} dt \\
&= I_\theta(y, r)
\end{aligned}$$

-

$$p\mathrm{-value}_{\mathrm{NB}} = I_{0.5}(y, r) = I_{0.5}(y, n-y)$$

Coin Tossing Problem (Example)

- Under Bayesian paradigm, we assume a symmetric beta prior ($\alpha = \beta$) for θ, i.e., $\theta \sim \mathrm{Beta}(\alpha, \beta)$.
- For $\alpha = 2, 1, 0.5, 0.1, 0.01, 0.001, 0.0001, 0.00001$, calculate $P(H_0|n = 12, y = 3)$ and comment on your findings in comparison with the p-values obtained under the binomial and negative binomial likelihood.
- Plot (i) p-value$_{\mathrm{B}}$ against $P(H_0|n, y)$ (ii) p-value$_{\mathrm{NB}}$ against $P(H_0|n, y)$ as sample size n increases while fixing $y/n = 0.75$. For the Bayesian paradigm, we set $\alpha = \beta = 0.01$.

Coin Tossing Problem (Example)

$\alpha = \beta$	$P(H_0 \mid n, y)$
2	0.0592346
1	0.0461426
0.5	0.0394446
0.1	0.0340598
0.01	0.0328493
0.001	0.0327283
0.0001	0.0327162
0.00001	0.0327150

Coin Tossing Problem (Example)

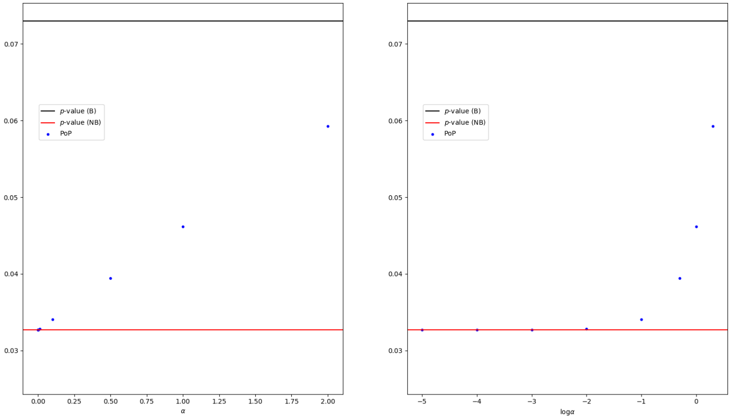


Figure 1: Plots of $P(H_0|n, y)$ against α=$(2, 1, 0.5, 0.1, 0.01, 0.001, 0.0001, 0.00001)$.

Coin Tossing Problem (Example)

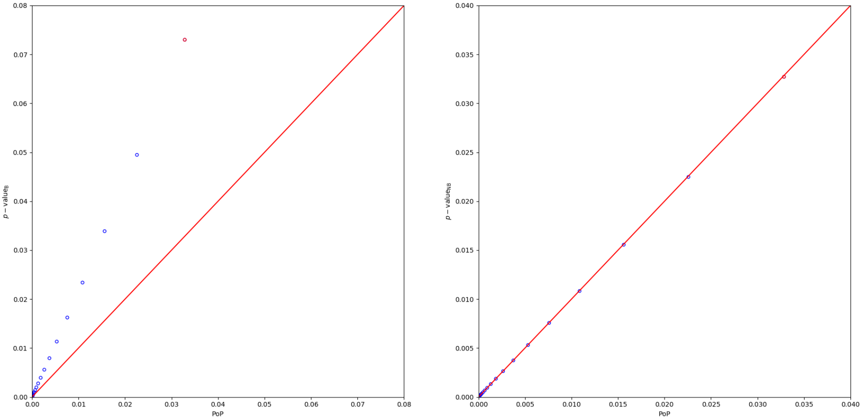


Figure 2: Plots of $P(H_0|n, y)$ against $p-\mathrm{value}_{\mathrm{B}}$ (left) and $p-\mathrm{value}_{\mathrm{NB}}$ (right)