Bayesian Statistics III/IV (MATH3361/4071) Easter term 2020

Revision Handout: Bayesian Statistics

Lecturer: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Bayesian statistics III/IV material term 1 2019 [1]

- Lecture handouts [2], and note [3]
- Exercise sheet [4]

References:

- Berger, J. O. (2013). Statistical decision theory and Bayesian analysis. Springer Science & Business Media.
- Robert, C. (2007). The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media.
- DeGroot, M. H. (2005). Optimal statistical decisions (Vol. 82). John Wiley & Sons.
- Raiffa, H., & Schlaifer, R. (1961). Applied statistical decision theory.
- O'Hagan, A., & Forster, J. J. (2004). Kendall's advanced theory of statistics, volume 2B: Bayesian inference (Vol. 2). Arnold.

[1] https://github.com/georgios-stats/Bayesian_Statistics

[2] https://github.com/georgios-stats/Bayesian_Statistics/tree/master/LectureHandouts

[3] https://duo.dur.ac.uk/bbcswebdav/pid-407343-dt-announcement-rid-20622014_1/users/mffk55/Bayesian-Statistics-MATH3341-4031/Logs/SuggestionsForTheExam.pdf

[4] https://github.com/georgios-stats/Bayesian_Statistics/tree/master/Homework

Bayesian calculations

Exercise 1. ($\star$)Consider an i.i.d. sample $y_1, \ldots, y_n$ from the skew-logistic distribution with PDF

$$f(y_i|\theta) = \frac{\theta e^{-y_i}}{(1+e^{-y_i})^{\theta+1}}$$

with parameter $\theta \in (0, \infty)$. To account for the uncertainty about θ we assign a Gamma prior distribution with PDF

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} 1(\theta \in (0, \infty)),$$

and fixed hyper parameters a, b specified by the researcher's prior info.

1. Derive the posterior distribution of θ.
2. Derive the predictive PDF for a future $z = y_{n+1}$.

Solution. It is

$$f(y_i|\theta) = \frac{\theta e^{-y_i}}{(1+e^{-y_i})^{\theta+1}} = \frac{\theta e^{-y_i}}{(1+e^{-y_i})} \exp\left(-\theta \log(1+e^{-y_i})\right)$$

1. By using the Bayes theorem

$$\begin{aligned}
\pi(\theta|y) \propto & f(y|\theta)\pi(\theta) \qquad \propto \prod_{i=1}^{n} f(y_i|\theta)\pi(\theta) \; = \prod_{i=1}^{n} \frac{\theta e^{-y_i}}{(1+e^{-y_i})^{\theta+1}} \, \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} 1(\theta \in (0, \infty)) \\
& \propto \cancel{\prod_{i=1}^{n} \frac{e^{-y_i}}{(1+e^{-y_i})}} \theta^n \prod_{i=1}^{n} \exp(-\theta \log(1+e^{-y_i})) \, \cancel{\frac{b^a}{\Gamma(a)}} \theta^{a-1} e^{-b\theta} 1(\theta \in (0, \infty)) \\
& \propto \theta^{n+a-1} \exp\left(-\theta \left[\sum_{i=1}^{n} \log(1+e^{-y_i}) + b\right]\right) 1(\theta \in (0, \infty)) \propto \text{Ga}(\theta | a+n, b + \sum_{i=1}^{n} \log(1+e^{-y_i}))
\end{aligned}$$

So

$$\theta|y \sim \text{Ga}\left(\underbrace{a+n}_{=a^*}, \underbrace{b + \sum_{i=1}^{n} \log(1+e^{-y_i})}_{=b^*}\right)$$

2. By using the definition for the predictive PDF, it is

$$\begin{aligned}
f(z|y) &= \int_{\mathbb{R}} f(z|\theta)\pi(\theta|y) \mathrm{d}\theta \\
&= \int_{\mathbb{R}_+} \frac{e^{-z}}{(1+e^{-z})} \theta \exp(-\theta \log(1+e^{-z})) \frac{(b^*)^{a^*}}{\Gamma(a^*)} \theta^{a^*-1} \exp(-\theta b^*) \mathrm{d}\theta \\
&= \frac{(b^*)^{a^*}}{\Gamma(a^*)} \frac{e^{-z}}{(1+e^{-z})} \int_{\mathbb{R}_+} \theta^{a^*+1-1} \exp(-\theta(b^* + \log(1+e^{-y}))) \mathrm{d}\theta \\
&= \frac{(b^*)^{a^*}}{\Gamma(a^*)} \frac{e^{-z}}{(1+e^{-z})} \frac{\Gamma(a^*+1)}{(b^* + \log(1+e^{-z}))^{a^*+1}} = \frac{e^{-z}}{(1+e^{-z})} \frac{(b^*)^{a^*}}{(b^* + \log(1+e^{-z}))^{a^*+1}} a^*
\end{aligned}$$

Exponential distribution family & Conjugate prior

Exercise 2. $(\star)$

1. Show that the skew-logistic family of distributions, with

$$f(x|\theta) = \frac{\theta e^{-x}}{(1+e^{-x})^{\theta+1}} \tag{1}$$

for $x \in \mathbb{R}$, labeled by $\theta > 0$, is a member of the exponential family and identify the factors u, g, h, ϕ, θ, c.

2. Show that the Gamma distribution

$$f(\theta|\alpha_0, \beta_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0 - 1} e^{-\beta_0 \theta} \tag{2}$$

with hyperparameters $w_0 := (\alpha_0, \beta_0)$ (where $\alpha_0 > 1$ and $\beta_0 > 0$) is conjugate for i.i.d. sampling from for the skew-logistic distribution. Relate the hyperparameters (τ_0, τ_1) to the standard parameters (α_0, β_0) of the gamma distribution.

Solution.

1. It is

$$\begin{aligned}
f(x|\theta) &= \frac{\theta e^{-x}}{(1+e^{-x})^{\theta+1}} \\
&= e^{-x}\theta \frac{1}{(1+e^{-x})} \exp\left(-\theta \log(1+e^{-x})\right) \\
&= e^{-x}\theta \exp\left(-\theta \log(1+e^{-x})\right) \\
&= \frac{e^{-x}}{(1+e^{-x})} \theta \exp\left(-\theta \log(1+e^{-x})\right)
\end{aligned}$$

So it is a member of the exponential distribution family

$$\text{Ef}_k(x|u, g, h, \phi, \theta, c) = u(x) g(\theta) \exp\left(\sum_{j=1}^{k} c_j \phi_j(\theta) h_j(x)\right)$$

with $k = 1$, and

$$u(x) = \frac{e^{-x}}{(1+e^{-x})}, \qquad g(\theta) = \theta, \qquad h(x) = \log(1+e^{-x}), \qquad \phi(\theta) = \theta, \qquad c = -1.$$

2. Following the corresponding theorem,

$$\begin{aligned}
\pi(\theta|\tau) &\propto g(\theta)^{\tau_0} \exp\left(\sum_{j=1}^{k} c_j \phi_j(\theta) \tau_j\right) \\
&\propto \theta^{\tau_0} \exp(-1\theta\tau_1) \ \propto \theta^{(\tau_0+1)-1} \exp(-\theta\tau_1)
\end{aligned}$$

where we recognize the Gamma PDF which identifies a Gamma distribution $\theta|\tau \sim \text{Ga}(\tau_0+1, \tau_1)$.

Also

$$K(\tau) = \int_{\mathbb{R}_+} g(\theta)^{\tau_0} \exp\left(\sum_{j=1}^{k} \phi_j(\theta)\tau_j\right) \mathrm{d}\theta = \int_{\mathbb{R}_+} \theta^{\tau_0} \exp(-\theta\tau_1) \mathrm{d}\theta = \frac{\Gamma(\tau_0+1)}{\tau_1^{\tau_0+1}}$$

since $\int_{\mathbb{R}_+} \text{Ga}(\theta|\tau_0+1, \tau_1) \mathrm{d}\theta = 1$. To easy the notation with $\theta|\tau \sim \text{Ga}(a, b)$, we can set $(a, b) = (\tau_0+1, \tau_1)$.

Conjugate prior

Exercise 3. ($\star$)Let $y = (y_1, ..., y_n)$ be observables drawn iid from sampling distribution $y_i \overset{\text{iid}}{\sim} \mathrm{U}(0, \theta)$ for all $i = 1, ..., n$. Specify the conjugate prior for θ.

Pareto distribution: If $x \sim \mathrm{Pa}(a, b)$, then it has a pdf $f(x) = ab^a(\frac{1}{\theta})^{a+1}1(b < \theta)$

Solution. The likelihood $f(y|\theta)$ can be factorized as

$$f(y|\theta) = \prod_{i=1}^{n} \mathrm{U}(y_i|0, \theta) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} 1\left(y_i \in [0, \theta]\right) = \underbrace{\left(\frac{1}{\theta}\right)^n 1\left(\max_{\forall i=1:n}(y_i) \in [0, \theta]\right)}_{=k(t(y)|\theta)}$$

with sufficient statistic $t = (n, \max_{\forall i=1:n}(y_i))$. Hence, I set

$$\pi(\theta) := \pi(\theta|\tau) \propto \left(\frac{1}{\theta}\right)^{\tau_0} 1\left(\tau_1 \in [0, \theta]\right) \propto \left(\frac{1}{\theta}\right)^{\tau_0} 1\left(\tau_1 < \theta\right) \propto \mathrm{Pa}(\theta|a = \tau_0 - 1, b = \tau_1)$$

By Bayes theorem the posterior is

$$\begin{aligned}\pi(\theta|y) &\propto \prod_{i=1}^{n} \mathrm{Un}(y_i|0, \theta)\mathrm{Pa}(\theta|\tau_0 - 1, \tau_1) \propto \overbrace{\left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} 1(y_i < \theta)}^{=\prod_{i=1}^{n}\mathrm{Un}(y_i|0,\theta)} \times \overbrace{\left(\frac{1}{\theta}\right)^{\tau_0} 1(\theta > \tau_1)}^{\propto \mathrm{Pa}(\theta|\tau_0-1,\tau_1)} \\ &\propto \left(\frac{1}{\theta}\right)^{n+\tau_0} \underbrace{\prod_{i=1}^{n} 1(\theta > x_i) 1(\theta > \tau_1)}_{=1(\theta > \max(\tau_1, x_{(n)}))} \propto \mathrm{Pa}(\theta|a^* = n + \tau_0 - 1, b^* = \max(\tau_1, x_{(n)})).\end{aligned}$$

where $\theta > \max(\tau_1, x_{(n)})$.

Jeffreys' priors

Exercise 4. ($\star\star$)Consider the Bayesian model

$$\begin{cases} x_i|\theta & \overset{\text{iid}}{\sim} \text{Ga}(\alpha,\beta),\ \forall i=1,...,n \\ (\alpha,\beta) & \sim \Pi(\alpha,\beta) \end{cases}$$

where $\text{Ga}(a,\beta)$ is the Gamma distribution with expected value α/β. Specify a Jeffrey's prior for $\theta=(\alpha,\beta)$.

Hint-1: Gamma distr.: $x \sim \text{Ga}(a,b)$ has pdf $f(x)=\frac{b^a}{\Gamma(a)}x^{a-1}\exp(-bx)1_{(0,+\infty)}(x)$, and Expected value $\text{E}_{\text{Ga}}(x|a,b)=\frac{a}{b}$

Hint-2: You may also need that the second derivative of the logarithm of a Gamma function is the 'polygamma function of order 1'. Ie,

- $\digamma^{(0)}(\alpha)=\frac{\text{d}}{\text{d}\alpha}\log(\Gamma(a))$
- $\digamma^{(1)}(\alpha)=\frac{\text{d}^2}{\text{d}\alpha^2}\log(\Gamma(a))$

Hint-3: You may leave your answer in terms of function $\digamma^{(1)}(\alpha)$.

Solution. It is $\pi(\alpha,\beta)\propto\sqrt{\det(\mathscr{I}(\alpha,\beta))}\propto\sqrt{\det(\mathscr{I}_1(\alpha,\beta))}$ where

$$\mathscr{I}_1(\alpha,\beta)=-\text{E}_{\text{F}(x|\alpha,\beta)}\begin{bmatrix}\frac{\text{d}^2}{\text{d}\alpha^2}\log(f(x|\alpha,\beta)) & \frac{\text{d}^2}{\text{d}\alpha\text{d}\beta}\log(f(x|\alpha,\beta)) \\ \frac{\text{d}^2}{\text{d}\alpha\text{d}\beta}\log(f(x|\alpha,\beta)) & \frac{\text{d}^2}{\text{d}\beta^2}\log(f(x|\alpha,\beta))\end{bmatrix},\ \text{with}$$

So

$$f(x|\alpha,\beta)=\text{Ga}(x|\alpha,\beta)=\frac{\beta^\alpha}{\Gamma(a)}x^{\alpha-1}\exp(-\beta x)\Longrightarrow$$
$$\log(f(x|\alpha,\beta))=a\log(\beta)-\log(\Gamma(\alpha))-\beta x+(\alpha-1)\log(x)$$

So

$$\frac{\text{d}}{\text{d}\alpha}\log(f(x|\alpha,\beta))=\log(\beta)-\frac{\text{d}}{\text{d}\alpha}\log(\Gamma(\alpha))+\log(x)$$
$$\frac{\text{d}}{\text{d}\alpha^2}\log(f(x|\alpha,\beta))=-\frac{\text{d}^2}{\text{d}\alpha^2}\log(\Gamma(\alpha))=-\digamma^{(1)}(\alpha)$$
$$\frac{\text{d}}{\text{d}\beta}\log(f(x|\alpha,\beta))=\frac{\alpha}{\beta}-x$$
$$\frac{\text{d}^2}{\text{d}\beta^2}\log(f(x|\alpha,\beta))=-\frac{\alpha}{\beta^2}$$
$$\frac{\text{d}^2}{\text{d}\alpha\text{d}\beta}\log(f(x|\alpha,\beta))=\frac{1}{\beta}$$

and

$$\text{E}_{\text{Ga}(a,b)}\left(\frac{\text{d}}{\text{d}\alpha^2}\log(f(x|\alpha,\beta))\right)=-\frac{\text{d}^2}{\text{d}\alpha^2}\log(\Gamma(\alpha))=-\digamma^{(1)}(\alpha)$$
$$\text{E}_{\text{Ga}(a,b)}\left(\frac{\text{d}^2}{\text{d}\beta^2}\log(f(x|\alpha,\beta))\right)=-\frac{\alpha}{\beta^2}$$
$$\text{E}_{\text{Ga}(a,b)}\left(\frac{\text{d}^2}{\text{d}\alpha\text{d}\beta}\log(f(x|\alpha,\beta))\right)=\frac{1}{\beta}$$

Hence

$$\mathscr{I}_1(\alpha,\beta) = -\mathrm{E}_{\mathrm{Ga}(a,b)}\begin{bmatrix} -F^{(1)}(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & -\frac{\alpha}{\beta^2} \end{bmatrix} = \begin{bmatrix} F^{(1)}(\alpha) & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{bmatrix}$$

Therefore

$$\pi(\alpha,\beta) \propto \sqrt{\det(\mathscr{I}(\alpha,\beta))} \propto \sqrt{\det(\mathscr{I}_1(\alpha,\beta))} = \sqrt{F^{(1)}(\alpha)\frac{\alpha}{\beta^2} + \frac{1}{\beta^2}} = \frac{1}{\beta}\sqrt{F^{(1)}(\alpha)\alpha + 1}$$

Bayesian point estimation

Exercise 5. ($\star\star$)Consider a Bayesian model

$$\begin{cases} y|\theta & \sim \mathrm{N}(\theta, 1) \\ \theta & \sim \mathrm{N}(0, 1) \end{cases}$$

1. Compute the Bayes point estimate $\hat{\delta}$ of θ under the loss $\ell(\theta, \delta) = \exp\left(\frac{3}{4}\theta^2\right)(\theta - \delta)^2$
2. Show that $\hat{\delta}$ is inadmissible, and discuss why this happens according to the Theorems in Handout 10.

Solution. The posterior of θ given y is $\theta|y \sim \mathrm{N}\left(\mu = \frac{1}{2}y, \sigma^2 = \frac{1}{2}\right)$ –the derivation is easy.

1. According to Proposition in the Handout, because $\mathrm{E}_\pi\left(\exp\left(\frac{3}{4}\theta^2\right)|y\right) > 0$, it is

$$\hat{\delta} = \frac{\mathrm{E}_\pi\left(\theta \exp\left(\frac{3}{4}\theta^2\right)|y\right)}{\mathrm{E}_\pi\left(\exp\left(\frac{3}{4}\theta^2\right)|y\right)}$$

I compute $\Delta_j = \mathrm{E}_\pi\left(\theta^j \exp\left(\frac{3}{4}\theta^2\right)|y\right)$ up to a multiplicative on j. It is $\hat{\delta} = \frac{\Delta(1)}{\Delta(0)} = 2y$, so for $j = 0, 1$ I have

$$\begin{aligned}
\Delta_j = \mathrm{E}(\theta^j \exp\left(\frac{3}{4}\theta^2\right)|y) &= \int_\Theta \theta^j \exp\left(\frac{3}{4}\theta^2\right) \mathrm{N}(\theta|\frac{1}{2}y, \frac{1}{2})\mathrm{d}\theta \\
&= \int_\Theta \exp(\frac{3}{4}\theta^2)\theta^j \frac{1}{\sqrt{2\pi 1/2}} \exp(-\frac{1}{2}\frac{(\frac{1}{2}y - \theta)^2}{1/2})\mathrm{d}\theta \\
&= \int_\Theta \theta^j \frac{1}{\sqrt{2\pi 1/2}} \exp(-\frac{1}{2}\frac{(\frac{1}{2}y - \theta)^2}{1/2} + \frac{3}{4}\theta^2)\mathrm{d}\theta \\
&= \int_\Theta \theta^j \frac{1}{\sqrt{2\pi 1/2}} \exp(-\frac{1}{2}\frac{(\theta - 2y)^2}{2} + \frac{3}{2}y^2)\mathrm{d}\theta \\
&= \frac{1}{2}\exp(\frac{3}{2}y^2) \int_\Theta \theta^j \frac{1}{\sqrt{2\pi 2}} \exp(-\frac{1}{2}\frac{(\theta - 2y)^2}{2})\mathrm{d}\theta \;\; = \begin{cases} \frac{1}{2}\exp(\frac{3}{2}y^2)2y & , j = 1 \\ \frac{1}{2}\exp(\frac{3}{2}y^2) & , j = 0 \end{cases}
\end{aligned}$$

So, I get $\hat{\delta} = \frac{\Delta(1)}{\Delta(0)} = 2y$.

$$\begin{aligned}
\Delta(j) = \mathrm{E}_\pi\left(\theta^j \exp\left(\frac{3}{4}\theta^2\right)|y\right) &= \int \theta^j \exp\left(\frac{3}{4}\theta^2\right) \mathrm{N}\left(\mu|\theta, \sigma^2\right)\mathrm{d}\theta \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int \theta^j \exp\left(-\frac{1}{2}\left[\theta^2 + \mu^2 - 2\theta\mu - \frac{3}{2}\theta^2\right]\right)\mathrm{d}\theta \\
&\propto \int \theta^j \exp\left(-\frac{1}{2}\left[\theta^2\left(\frac{1}{\sigma^2} - \frac{3}{2}\right) - 2\theta\mu\right]\right)\mathrm{d}\theta \\
&\propto \int \theta^j \exp\left(-\frac{1}{2}\left(\theta - \frac{\mu/\sigma^2}{\frac{1}{\sigma^2} - \frac{3}{2}}\right)^2 \Big/ \left(\frac{1}{\sigma^2} - \frac{3}{2}\right)^{-1}\right)\mathrm{d}\theta \\
&\propto \int \theta^j \frac{1}{\sqrt{2\pi\left(\frac{1}{\sigma^2} - \frac{3}{2}\right)}} \exp\left(-\frac{1}{2}\left(\theta - \frac{\mu/\sigma^2}{\frac{1}{\sigma^2} - \frac{3}{2}}\right)^2 \Big/ \left(\frac{1}{\sigma^2} - \frac{3}{2}\right)^{-1}\right)\mathrm{d}\theta \\
&\propto \int \theta^j \mathrm{N}\left(\mu|\frac{\mu/\sigma^2}{\frac{1}{\sigma^2} - \frac{3}{2}}, \frac{1}{\frac{1}{\sigma^2} - \frac{3}{2}}\right)\mathrm{d}\theta
\end{aligned}$$

So, by setting $\mu = \frac{1}{2}y, \sigma^2 = \frac{1}{2}$, I get $\hat{\delta} = \frac{\Delta(1)}{\Delta(0)} = 2y$.

2. Assume $\delta_c(y) = cy$ where my estimator is a member; i.e. $\hat{\delta} = \delta_2$. The risk for $\delta_c(y)$ is

$$R(\theta, \delta_c) = \mathrm{E}_{\mathrm{N}(\theta,1)}\left(\exp\left(\frac{3}{4}\theta^2\right)(\theta - cy)^2\right) = \exp\left(\frac{3}{4}\theta^2\right)\left(c^2 + \theta^2(c-1)^2\right)$$

because:

$$\begin{aligned}
R(\theta, \delta_c) &= \mathrm{E}_F(\ell(\theta, \delta_c(y))|\theta) = \mathrm{E}_{\mathrm{N}(\theta,1)}\left(\exp(\frac{3}{4}\theta^2)(\theta - cy)^2\right) \\
&= \exp\left(\frac{3}{4}\theta^2\right)\mathrm{E}_{\mathrm{N}(\theta,1)}(\theta - cy)^2 \\
&= \exp\left(\frac{3}{4}\theta^2\right)\mathrm{E}_{\mathrm{N}(\theta,1)}(cy - \theta)^2 = \exp\left(\frac{3}{4}\theta^2\right)\mathrm{E}_{\mathrm{N}(\theta,1)}([cy - c\theta] + [c\theta - \theta])^2 \\
&= \exp\left(\frac{3}{4}\theta^2\right)\mathrm{E}_{\mathrm{N}(\theta,1)}([cy - c\theta]^2 + [c\theta - \theta]^2 - 2[cy - c\theta]) \\
&= \exp\left(\frac{3}{4}\theta^2\right)\left(c^2\mathrm{E}_{\mathrm{N}(\theta,1)}[y - \theta]^2 + \mathrm{E}_{\mathrm{N}(\theta,1)}[(c-1)\theta]^2 - 2[c\mathrm{E}_{\mathrm{N}(\theta,1)}(y) - c\theta]\right) \\
&= \exp\left(\frac{3}{4}\theta^2\right)\left(c^2 + \theta^2(c-1)^2\right)
\end{aligned}$$

I observe that $\hat{\delta} = \delta_2$ is dominated by δ_1. It is

$$R(\theta, \delta_2) = \exp\left(\frac{3}{4}\theta^2\right)(4 + \theta)^2$$

and

$$R(\theta, \delta_1) = \exp\left(\frac{3}{4}\theta^2\right)$$

where one can see that $R(\theta, \delta_2) = R(\theta, \delta_1)$ for $\theta \in \{0, -4\}$ and $R(\theta, \delta_2) > R(\theta, \delta_1)$ for all the rest θ. Hence $\hat{\delta} = \delta_2$ is inadmissible.

I observe that $\hat{\delta}$ does not produces a finite Bayes risk

$$\begin{aligned}
r(\pi, \hat{\delta}) &= \int R(\theta, \hat{\delta})\pi(\theta)\mathrm{d}\theta = \int R(\theta, \hat{\delta})\mathrm{N}(\theta|0,1)\mathrm{d}\theta \\
&\propto \int (4 + \theta)^2 \exp\left(\frac{3}{4}\theta^2\right)\exp\left(-\frac{1}{2}\theta^2\right)\mathrm{d}\theta \\
&= \int (4 + \theta)^2 \exp\left(\frac{1}{4}\theta^2\right)\mathrm{d}\theta > \int \exp\left(\frac{1}{4}\theta^2\right)\mathrm{d}\theta = \infty
\end{aligned}$$

and hence Bayesian point estimate $\hat{\delta}$ is not necessarily admissive.

Credible intervals

Exercise 6. ($\star\star$)Assume an 1-dimensional random quantity $x \sim Q(x|y)$.

Hint: In the Lecture Handout (Handout 11: Bayesian point estimation), we can say that:

The Bayes estimate $\hat{\delta}$ of x under the linear loss function

$$\ell(x, \delta; \varpi) = (1 - \varpi)(\delta - x)1_{x \leq \delta}(\delta) + \varpi(x - \delta)1_{x > \delta}(\delta),$$

where $\varpi \in [0, 1]$, is the ϖ-th quantile of distribution Q, let's denote it as x_{ϖ}.

Do the following:

1. Derive the $(1 - a)$-credible interval $C_a = [L, U]$ for x as a Bayesian rule C_a under the loss function

$$\ell(x, C_a; \varpi_L, \varpi_U) = \ell(x, L; \varpi_L) + \ell(x, U; \varpi_U) \tag{3}$$

   by computing L and U.

2. Your client is worried the same both for under-estimation and over-estimation; derive a suitable $(1 - a)$-credible interval $C_a = [L, U]$ based on (3) by computing L,and U.

3. Your client is worried only for over-estimation; derive a suitable $(1 - a)$-credible interval $C_a = [L, U]$ based on (3) by computing L and U.

Solution. It is given that

$$\begin{aligned} 0 = & \frac{\mathrm{d}}{\mathrm{d}\delta} \mathrm{E}_Q \left(\ell(x, \delta; \varpi) |y \right) \Big|_{\delta = \hat{\delta}} = \frac{\mathrm{d}}{\mathrm{d}\delta} \int \ell(x, \delta; \varpi) \mathrm{d}Q(x|y) \Big|_{\delta = \hat{\delta}} \\ & = (1 - \varpi)\mathsf{P}_Q \left(\{x \leq \hat{\delta}\} |y \right) - \varpi \mathsf{P}_Q \left(\{x \leq \hat{\delta}\}^{\complement} |y \right) \implies \hat{\delta} = x_{\varpi} \end{aligned}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\delta^2} \mathrm{E}_Q \left(\ell(x, \delta; \varpi) |y \right) \Big|_{\delta = \hat{\delta}} = q(\hat{\delta}|y) > 0$$

1. The decision space is $\mathcal{D} = \{C_a = [L, U] \, : \, \mathsf{P}_Q(x \in C_a|y) = 1 - a\}$. Therefore,

$$0 = \frac{\mathrm{d}}{\mathrm{d}L} \mathrm{E}_Q \left(\ell(x, C_a; \varpi_L, \varpi_U) |y \right) \Big|_{C_a = [\hat{L}, \hat{U}]} = \frac{\mathrm{d}}{\mathrm{d}L} \mathrm{E}_Q \left(\ell(x, L; \varpi_L) |y \right) \Big|_{L = \hat{L}} \implies \hat{L} = x_{\varpi_L}$$

   based on the Handout result, and likewise

$$0 = \frac{\mathrm{d}}{\mathrm{d}U} \mathrm{E}_Q \left(\ell(x, C_a; \varpi_L, \varpi_U) |y \right) \Big|_{C_a = [\hat{L}, \hat{U}]} = \frac{\mathrm{d}}{\mathrm{d}U} \mathrm{E}_Q \left(\ell(x, U; \varpi_U) |y \right) \Big|_{U = \hat{U}} \implies \hat{U} = x_{\varpi_U}$$

   So $x \in [x_{\varpi_L}, x_{\varpi_U}]$ where $\varpi_U + \varpi_L = 1 - a$.

   It is the minimum because

$$\frac{\mathrm{d}^2}{\mathrm{d}U^2} \mathrm{E}_Q \left(\ell(x, C_a; \varpi_L, \varpi_U) |y \right) \Big|_{C_a = [\hat{L}, \hat{U}]} = q(\hat{U}|y) > 0$$

$$\frac{\mathrm{d}^2}{\mathrm{d}L^2} \mathrm{E}_Q \left(\ell(x, C_a; \varpi_L, \varpi_U) |y \right) \Big|_{C_a = [\hat{L}, \hat{U}]} = q(\hat{L}|y) > 0$$

$$\frac{\mathrm{d}}{\mathrm{d}U} \frac{\mathrm{d}}{\mathrm{d}L} \mathrm{E}_Q \left(\ell(x, C_a; \varpi_L, \varpi_U) |y \right) \Big|_{C_a = [\hat{L}, \hat{U}]} = 0$$

   and hence the determinant of the Hessian in positive.

2. Then I can use the equi-tail interval: $x \in [x_{a/2}, x_{1-a/2}]$ with $\varpi_L = c$ and $\varpi_U = 1$
3. Then I can use the lower-tail interval: $x \in (-\infty, x_{1-a}]$ with $\varpi_L = 0$ and $\varpi_U = 1 - a$.

HPD Credible intervals

Exercise 7. ($\star\star$)Consider the Bayesian model

$$y_i|\theta \overset{\text{iid}}{\sim} \mathrm{N}\left(\theta, \sigma^2\right),\ i = 1, ..., n$$
$$\theta \sim \mathrm{N}\left(\mu_0, \sigma_0^2\right)$$

where μ_0, σ_0^2 are fixed hyper-parameters, and θ unknown.

1. Derive the $1 - \frac{a}{2}$ HPD credible posterior interval for θ.

   Hint-1: It is
   $$\sum_{i=1}^{n} \frac{(x - \mu_i)^2}{\sigma_i^2} = \frac{(x - \hat{\mu})^2}{\hat{\sigma}^2} + \text{const ind of } x$$
   where $\hat{\sigma}^2 = \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\right)^{-1}$ and $\hat{\mu} = \hat{\sigma}^2 \left(\sum_{i=1}^{n} \frac{\mu_i}{\sigma_i^2}\right)$.

   Hint-2: The 97.5% quantile of the standard Normal distribution is 1.959964.

2. What size your dataset need to have in order to satisfy a 0.95% HPD credible posterior interval for θ which has length of 1 unit? Consider that $\sigma^2 = 4$ and $\sigma_0^2 = 9$.

Solution. Let $y = (y_1, ..., y_n)$.

1. The posterior pdf of θ is

$$\pi(\theta|y) \propto f(y|\theta) \times \pi(\theta) = \prod_{i=1}^{n} f(y_i|\theta \times) \pi(\theta) \propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\frac{(y_i - \theta)^2}{\sigma^2}\right) \times \exp\left(-\frac{1}{2}\frac{(\theta - \mu_0)^2}{\sigma_0^2}\right)$$
$$= \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left[\frac{(\theta - y_i)^2}{\sigma^2} + \frac{(\theta - \mu_0)^2}{\sigma_0^2}\right]\right) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(\theta - \mu_n)^2}{\sigma_n^2} + \text{const...}\right)$$

   with
   $$\sigma_n^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1} \text{ and } \mu_n = \sigma_n^2\left(\frac{\bar{y}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)$$

   Hence
   $$\theta|y \sim \mathrm{N}\left(\mu_n, \sigma_n^2\right)$$

   To find the 2-sides the $1 - \frac{a}{2}$ HPD credible posterior interval for θ, aka $[L, U]$, I consider the theorem in the Handouts. Namely:

$$1 - a = \int_{L}^{U} \mathrm{N}(\theta|\mu_n, \sigma_n^2)\mathrm{d}\theta = \mathsf{P}_{\mathrm{N}(\mu_n, \sigma_n^2)}(\theta < U) - \mathsf{P}_{\mathrm{N}(\mu_n, \sigma_n^2)}(\theta < L) = \Phi\left(\frac{U - \mu_n}{\sigma_n}\right) - \Phi\left(\frac{L - \mu_n}{\sigma_n}\right)$$

   where $\Phi(\cdot)$ denotes the CDF of $\mathrm{N}(0, 1)$. Also, it has to be
   $$\pi(U|y) = \pi(L|y)$$
   and because the PDF of $\mathrm{N}\left(\mu_n, \sigma_n^2\right)$ is symmetric around μ_n
   $$L - \mu_n = \mu_n - U \implies L = 2\mu_n - U$$

So

$$1-a=\Phi\left(\frac{U-\mu_n}{\sigma_n}\right)-\Phi\left(\frac{L-\mu_n}{\sigma_n}\right)=\Phi\left(\frac{U-\mu_n}{\sigma_n}\right)-\Phi\left(\frac{2\mu_n-U-\mu_n}{\sigma_n}\right)=2\Phi\left(\frac{U-\mu_n}{\sigma_n}\right)-1 \implies$$

$$1-\frac{a}{2}=\Phi\left(\frac{U-\mu_n}{\sigma_n}\right) \implies z_{1-\frac{a}{2}}=\frac{U-\mu_n}{\sigma_n} \implies U=\mu_n+z_{1-\frac{a}{2}}\sigma_n^2$$

and hence $L=\mu_n-z_{1-\frac{a}{2}}\sigma_n^2$.

So the $1-\frac{a}{2}$ HPD credible posterior interval for θ is

$$\left[\mu_n-z_{1-\frac{a}{2}}\sigma_n\ ,\ \mu_n+z_{1-\frac{a}{2}}\sigma_n\right]=\left[\mu_n-z_{1-\frac{a}{2}}\Big/\sqrt{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}\ ,\ \mu_n+z_{1-\frac{a}{2}}\Big/\sqrt{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}\ \right]$$

2. The length of the $1-\frac{a}{2}$ HPD credible posterior interval for θ is

$$\ell_n=2z_{1-\frac{a}{2}}\Big/\sqrt{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}$$

So

$$1=\ell_n=2z_{1-\frac{a}{2}}\Big/\sqrt{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}=2\times 1.959964\Big/\sqrt{\left(\frac{n}{4}+\frac{1}{9}\right)}$$

and hence $n\approx 62$.

Hypothesis test

Exercise 8. ($\star\star$)Let $y = (y_1, ..., y_n)$ a sequence of observables, and assume that $n = 5$, and $y_* = \sum_{i=1}^{5} y_i = 3$. Assume a sampling distribution $y_i|\theta \overset{\text{iid}}{\sim} \text{Br}(\theta)$, with unknown parameter $\theta \in [0, 1]$, a priori following a uniform distribution.

1. By using Jeffreys' scaling rule, perform the following hypothesis test for $\theta_0 = 1/2$

$$\text{H}_0 : \theta = \theta_0 \quad \text{vs} \quad \text{H}_1 : \theta \neq \theta_0$$

   (a) Compute the posterior probability of the NULL hypothesis.

Solution. This is a simple vs. general alternative hypothesis. I specify the overall prior with pdf

$$\pi(\theta) = \pi_0 1(\theta = \theta_0) + (1 - \pi_0)\text{U}(\theta|0, 1)$$

for some $\pi_0 > 0$. I leave π_0 abstract, however the usual choice (but maybe not the best) is $\pi_0 = 1/2$.

1. The Bayes factor is

$$\text{B}_{01}(y) = \frac{f_0(y)}{f_1(y)}$$

   where

$$f_0(y) = \prod_{i=1}^{n} \text{Br}(y_i|\theta_0) = \prod_{i=1}^{n} \theta_0^{y_i}(1 - \theta_0)^{1-y_i} = \theta_0^{y_*}(1 - \theta_0)^{n-y_*}$$

   and

$$\begin{aligned} f_1(y) &= \int_{(0,1)} \prod_{i=1}^{n} \text{Br}(y_i|\theta)\text{U}(\theta|0, 1)\text{d}\theta = \int_{(0,1)} \prod_{i=1}^{n} \theta^{y_i}(1 - \theta)^{1-y_i}\text{d}\theta \\ &= \int_{(0,1)} \theta^{y_*}(1 - \theta)^{n-y_*}\text{d}\theta = \text{B}(y_* + 1, n - y_* + 1) \end{aligned}$$

   So

$$\text{B}_{01}(y) = \frac{f_0(y)}{f_1(y)} = \frac{\theta_0^{y_*}(1 - \theta_0)^{n-y_*}}{\text{B}(y_* + 1, n - y_* + 1)} = \frac{(1/2)^5}{\text{B}(4, 3)} = \frac{15}{8}$$

   Then $B_{01}(y) = \frac{15}{8} \approx 2$, and $\log_{10}(\text{B}_{01}(y)) \approx 0.27$. According to Jeffreys' scaling rule, H_0 is supported. We can accept the null hypothesis.

2. The posterior probability of H_0 is

$$\text{P}_\Pi(\theta = \theta_0|y) = \text{P}_\Pi(\theta \in \Theta_0|y) = \left[1 + \frac{1 - \pi_0}{\pi_0}\text{B}_{01}(y)^{-1}\right]^{-1} = \left[1 + \frac{1/2}{1 - 1/2}(\frac{15}{8})^{-1}\right]^{-1} = \frac{15}{23} \approx 0.65$$

   and hence the posterior distribution tends to support H_0.

Hypothesis test, and differences from classical stats results

Exercise 9. ($\star\star\star$)

1. Consider the Single vs. General alternative Bayesian hypothesis test

$$\mathrm{H}_0 : \theta = \theta_0 \qquad \text{vs} \qquad \mathrm{H}_1 : \theta \neq \theta_0.$$

   of more formally

$$\mathrm{H}_0 : \begin{cases} y|\theta_0 & \sim F(y|\theta_0) \\ \theta_0 \text{ is fixed} \end{cases} \qquad \text{vs} \qquad \mathrm{H}_1 : \begin{cases} y|\theta & \sim F(y|\theta) \\ \theta & \sim \Pi_1(\theta),\ \theta \in \Theta_1 \end{cases} \tag{4}$$

   Show that the Bayes factor B_{01} of H_0 and H_1 can be computes as

$$\mathrm{B}_{01} = \frac{\pi_1\left(\theta_0|y\right)}{\pi_1\left(\theta_0\right)}$$

2. In the above hypothesis test, assume that $y|\theta \sim \mathrm{Bin}(n, \theta)$ has Binomial sampling distribution with unknown parameter θ.

   (a) Find an approximation of B_{01}, when n is large.

   (b) For large n, show that $\mathrm{B}_{01} > 1$ when $z_0 = \frac{p-\theta_0}{\sqrt{u}}$, with $p = \frac{y}{n}$ and $u = \frac{p(1-p)}{n}$, satisfies $|z| < \max\left(\sqrt{k}, 0\right)$ for some k that depends of n, p, and $\pi_1\left(\theta_0\right)$.

   (c) Let the conditional prior $\Pi_1(\theta)$ be a Uniform distribution with positive mass above the interval $[0, 1]$. Show that this choice of the conditional prior $\Pi_1(\theta)$ can create a "paradox" when compared with fixed size tests.

Solution.

1. Based on the Lecture discussions we know that for Hipothesis test such as

$$\mathrm{H}_0 : \theta = \theta_0 \qquad \text{vs} \qquad \mathrm{H}_1 : \theta \neq \theta_0.$$

   the Bayes factor B_{01} of H_0 and H_1 is

$$\mathrm{B}_{01}(y) = \frac{f_0(y)}{f_1(y)}$$

   where $f_0(y) = f(y|\theta_0)$ and $f_1(y) = \int f(y|\theta)\pi_1(\theta)\mathrm{d}\theta$.

   From Bayesian theorem it is

$$\pi_1(\theta|y) = \frac{f(y|\theta)\pi_1(\theta)}{f_1(y)}$$

   So

$$\pi_1(\theta_0|y) = \frac{f(y|\theta_0)\pi_1(\theta_0)}{f_1(y)} = \frac{f_0(y)\pi_1(\theta_0)}{f_1(y)} = \mathrm{B}_{01}\pi_1(\theta_0)$$

   because

$$\int_{\{\theta=\theta_0\}} f(y|\theta)1\left(\theta \in \{\theta_0\}\right)\mathrm{d}\theta = f_0(y) = \int_{\{\theta=\theta_0\}} f(y|\theta)1\left(\theta \in \{\theta_0\}\right)\mathrm{d}\theta = f(y|\theta_0)$$

   Hence

$$\mathrm{B}_{01} = \frac{\pi_1\left(\theta_0|y\right)}{\pi_1\left(\theta_0\right)}$$

2.

(a) For Stats 2, we can find say about the density of the posterior of $\theta|y$ that

$$\pi_1(\theta|y) \approx \mathrm{N}\left(\theta|\hat{\theta}_{\text{MLE}}, 1/\mathscr{I}(\hat{\theta}_{\text{MLE}})\right)$$

where $\hat{\theta}_{\text{MLE}} = \frac{y}{n} = p$ and $1/\mathscr{I}(\hat{\theta}_{\text{MLE}}) = p(1-p)/n = u$.
So

$$\mathrm{B}_{01} = \frac{\pi_1(\theta_0|y)}{\pi_1(\theta_0)} \approx \frac{(2\pi u)^{-1/2}\exp\left(-\frac{1}{2}\frac{(\theta_0-p)^2}{u}\right)}{\pi_1(\theta_0)}$$

(b) For

$$\begin{aligned}\mathrm{B}_{01} > 1 \iff -\log(\mathrm{B}_{01}) < 0 \iff \\ \frac{(\theta_0-p)^2}{u} <& \underbrace{-\log(2\pi u)}_{k^*} - 2\log(\pi_1(\theta_0)) \iff \\ \frac{(\theta_0-p)^2}{u} <& k^* - 2\log(\pi_1(\theta_0))\end{aligned}$$

So re result follows in the required form with

$$k = k^* - 2\log(\pi_1(\theta_0))$$

where

$$k^* = -\log(2\pi u) = -\log\left(2\pi\frac{p(1-p)}{n}\right)$$

(c) I have $\pi_1(\theta_0) = 1$. Then $\mathrm{B}_{01} > 1$ when $|z| < \sqrt{k^*}$.

- Assume I choose $\sqrt{k^*} = 3$, and I investigate what I get here:
- In frequentist stasts, I suppose that the classical hypothesis test would reject H_0 whenever $|z| < \sqrt{k^*} = \sqrt{3}$ (this corresponds to in sig. level 1%). So if I get $|z| = 3$, I would reject H_0.
- In Bayesian stats, $|z| = 3$ corresponds to

$$\mathrm{B}_{01} \approx \frac{(2\pi u)^{-1/2}\exp\left(-\frac{1}{2}\frac{(\theta_0-p)^2}{u}\right)}{\pi_1(\theta_0)} = \frac{\exp\left(\frac{1}{2}3\right)\exp\left(-\frac{1}{2}3\right)}{1} = 1$$

- For the record, I can get $\sqrt{k^*} = 2$ for a sample size $n = 2\pi p(1-p)\exp(9) \approx 16206\pi p(1-p)$

Bayesian paradigm, Exponential family of distributions, Conjugate priors, predictions, HPD credible intervals, hypothesis test

Exercise 10. ($\star\star$)Consider two categorical random variables X with I levels, and Y with J levels. Consider associated joint and marginal probabilities,

$$\pi_{i,j} = \mathsf{P}(X=i, Y=j)\;; \quad \pi_{i,+} = \sum_{j=1}^{J} \pi_{i,j} = \mathsf{P}(X=i)\;; \quad \pi_{+,j} = \sum_{i=1}^{I} \pi_{i,j} = \mathsf{P}(Y=j)\;; \quad \pi_{+,+} = 1\;;$$

for $i = 1, ..., I$, and $j = 1, ..., J$. We use vectoriced notation

$$\begin{aligned}
\boldsymbol{\pi}_{X,Y} &= (\pi_{i,j})_{I,J} = (\pi_{1,1}, ..., \pi_{1,J}, ..., \pi_{i,j}, ..., \pi_{I,1}, ..., \pi_{I,J}) \\
\boldsymbol{\pi}_{X} &= (\pi_{i,+})_{I} = (\pi_{1,+}, ..., \pi_{I,+}) \\
\boldsymbol{\pi}_{Y} &= (\pi_{+,j})_{J} = (\pi_{+,1}, ..., \pi_{+,J})
\end{aligned}$$

Assume that there is available a sample classified in a 2 way $I \times J$ contingency table according to the classifier variables X and Y. Let $n_{i,j}$ denote the number of the observations int he (i,j)th cell, and let $n_{i,+} = \sum_{j=1}^{J} n_{i,j}$, $n_{+,j} = \sum_{i=1}^{I} n_{i,j}$, and $n_{+,+} = \sum_{i,j=1,1}^{I,J} n_{i,j}$ for $i = 1, ..., I$, and $j = 1, ..., J$. In a similar manner as above we define

$$\boldsymbol{n}_{X,Y} = (n_{i,j})_{I,J}\;; \qquad \boldsymbol{n}_{X} = (n_{i,+})_{I}\;; \qquad \boldsymbol{n}_{Y} = (n_{+,j})_{J}\,.$$

Hint-1: The Multinomial distribution $x \sim \text{Mult}(n, \pi)$ has pmf

$$f(y|\pi) = \binom{n}{y} \prod_{i=1}^{k} \pi_i^{y_i} 1\,(y_i \in \mathcal{Y})\,; \binom{n}{y} = \frac{n!}{y_1! \cdots y_n!}; \qquad \mathcal{Y} = \left\{ y \in \{0, ..., n\}^k : \sum_{i=1}^{k} y_i = n \right\}$$

Hint-2: The Dirichlet distribution $x \sim \text{Dir}(a)$ has pdf

$$f(x|a) = \frac{1}{B(a)} \prod_{i=1}^{k} x_i^{a_i - 1} 1\,(x_i \in \mathcal{X})\,; B(a) = \frac{\Gamma(\prod_{i=1}^{k} a_i)}{\prod_{i=1}^{k} \Gamma(a_i)} \qquad \mathcal{X} = \left\{ x \in (0,1)^k : \sum_{i=1}^{k} x_i = 1 \right\}$$

Hint-3: The Beta distribution $x \sim \text{Be}(a, b)$ has pdf

$$f(x|a,b) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} 1\,(x \in (0,1))\,; B(a) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \qquad \mathcal{X} = [0,1]$$

with $a > 0$ and $b > 0$. It is $\text{E}(x|a,b) = a/(a+b)$, and $\text{Var}(x|a,b) = ab/\left[(a+b)^2(a+b+1)\right]$. When $a = b$, then the PDF is symmetric.

Address the following:

1. Assume that the sampling distribution of $(n_{i,j})_{I,J}$ is a Multinomial, and consider the Bayesian model

$$\mathscr{B}_1 : \begin{cases} \boldsymbol{n}_{X,Y} | \boldsymbol{\pi}_{X,Y} & \sim \text{Mult}\,(n_{+,+}, \boldsymbol{\pi}_{X,Y}) \\ \boldsymbol{\pi}_{X,Y} & \sim \Pi(\boldsymbol{\pi}_{X,Y}) \end{cases}$$

where $n_{+,+}$ is fixed, and $\Pi(\boldsymbol{\pi}_{X,Y})$ is currently unspecified.

   (a) Write down the likelihood of $\boldsymbol{n}_{X,Y}$ given $\boldsymbol{\pi}_{X,Y}$
   (b) Show that the sampling distribution is an Exponential family of distributions.
   (c) Derive the conjugate prior family of distributions, and specify the prior distribution $\Pi(\boldsymbol{\pi}_{X,Y})$

(d) Derive the posterior distribution of $\boldsymbol{\pi}_{X,Y}$ given $\boldsymbol{n}_{X,Y}$, and write down its density.

(e) Compute the predictive distribution of a future observation $m_{X,Y} = (m_{i,j})_{I,J}$ where $m_{i,j} \in \{0, 1\}$ for $i = 1, ..., I$ and $j = 1, ..., J$ and $\sum_{i=1,j=1}^{I,J} m_{i,j} = m_{+,+}$ as fixed hyper-parameter. $m_{+,+}$ is fixed with $m_{+,+} = 1$.

2. Assume that the sampling distribution of $(n_{i,j})_{I,J}$ is a Multinomial, and consider the Bayesian model

$$\mathscr{B}_2 : \begin{cases} \boldsymbol{n}_{X,Y}|\boldsymbol{\pi}_{X,Y} & \sim \text{Mult}\left(n_{+,+}, \boldsymbol{\pi}_{X,Y}\right) \\ \boldsymbol{\pi}_{X,Y} & = \lambda(\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y) \\ (\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y) & \sim \Pi(\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y) \end{cases}$$

which assumes that X an Y are independent, and where $\Pi(\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y)$ and $\lambda(\cdot, \cdot)$ are currently unspecified. Here $n_{+,+}$ is fixed.

(a) Specify the function $h(\cdot, \cdot)$ and re-write Bayesian model $\mathscr{B}_2$ by applying your $h(\cdot, \cdot)$.

(b) Write down the likelihood of $\boldsymbol{n}_{X,Y}$ given $\boldsymbol{\pi}_X$ and $\boldsymbol{\pi}_Y$.

(c) Show that the sampling distribution is an Exponential family of distributions.

(d) Derive the conjugate prior family of distributions, and specify the prior distribution $\Pi(\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y)$

(e) Derive the posterior distribution of $\boldsymbol{\pi}_{X,Y}$ given $\boldsymbol{n}_{X,Y}$, and write down its density.

(f) Compute the predictive distribution of a future observation $m_{X,Y} = (m_{i,j})_{I,J}$ where $m_{i,j} \in \mathbb{N}$ for $i = 1, ..., I$ and $j = 1, ..., J$ and $\sum_{i=1,j=1}^{I,J} m_{i,j} = m_{+,+} = 1$ as fixed hyper-parameter.

(g) Compute the 95% HPD credible interval for $\pi_{1,+}$. Assume $I = J = 2$ and assume that $\text{E}(\pi_{1,+}|n_{X,Y}) = 1/2$ and $\text{Var}(\pi_{1,+}|n_{X,Y}) = 1/140$. Consider that the quantiles of the Beta distribution are known.

3. Derive a Hypothesis test where you compare the following hypothesis (here written down in words as)

$$\begin{aligned} \text{H}_0 &: X \text{ and } Y \text{ are independent} \\ \text{H}_1 &: X \text{ and } Y \text{ are not independent} \end{aligned}$$

Consider that priors $\pi_0 = \pi(\text{H}_0)$ and $\pi_1 = \pi(\text{H}_1)$ are already specified and known/.

(a) Re-write the pair of hypothesis as two Bayesian models under comparison by using priors leading to tractable Bayesian inference.

(b) Compute the Bayes factor

(c) Write down the decision rule based on the $c_I - c_{II}$-loss function.

Solution.

1.

(a) The likelihood is

$$f(n_{XY}|\pi_{XY}) = \frac{n_{++}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{i,j}!} \prod_{i=1}^{I}\prod_{j=1}^{J} \pi_{i,j}^{n_{i,j}}$$

(b) The multinational sampling distribution can be written in the form

$$
\begin{aligned}
f(n_{XY}|\pi_{XY}) =& \frac{n_{++}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{i,j}!}\prod_{i=1}^{I}\prod_{j=1}^{J}\pi_{i,j}^{n_{i,j}} = \frac{n_{++}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{i,j}!}\prod_{(i,j)\neq(I,J)}\pi_{i,j}^{n_{i,j}}\pi_{I,J}^{n_{I,J}} \\
=& \frac{n_{++}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{i,j}!}\exp\left(\sum_{(i,j)\neq(I,J)} n_{i,j}\log\left(\pi_{i,j}\right)\right. \\
& \left. + \left(n_{++} - \sum_{(i,j)\neq(I,J)} n_{i,j}\right)\log\left(1-\sum_{(i,j)\neq(I,J)}\pi_{i,j}\right)\right) \\
=& \frac{n_{++}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{i,j}!}\exp\left(\sum_{(i,j)\neq(I,J)} n_{i,j}\log\left(\frac{\pi_{i,j}}{1-\sum_{(i,j)\neq(I,J)}\pi_{i,j}}\right)\right. \\
& \left. + n_{++}\log\left(1-\sum_{(i,j)\neq(I,J)}\pi_{i,j}\right)\right) \\
=& \frac{n_{++}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{i,j}!}\left(1-\sum_{(i,j)\neq(I,J)}\pi_{i,j}\right)^{n_{++}}\exp\left(\sum_{(i,j)\neq(I,J)} n_{i,j}\log\left(\frac{\pi_{i,j}}{1-\sum_{(i,j)\neq(I,J)}\pi_{i,j}}\right)\right)
\end{aligned}
$$

and hence it is a $(IJ-1)$-exponential family of distributions with

$$
u(n) = \frac{n_{++}!}{\prod_{i=1}^{I}\prod_{j=1}^{J} n_{i,j}!}n_{++}, \quad g(\pi) = \left(1-\sum_{(i,j)\neq(I,J)}\pi_{i,j}\right)^{n_{++}}, \quad \phi_{i,j}(\pi) = \log\left(\frac{\pi_{i,j}}{1-\sum_{(i,j)\neq(I,J)}\pi_{i,j}}\right),
$$

$h(n) = 1$, and $c_{i,j} = n_{i,j}$, for $i = 1, ..., I-1$, and $j = 1, ..., J-1$.

(c) The sampling distribution is recognized as an exponential family of distributions then the conjugate prior has pdf

$$
\begin{aligned}
p(\pi|\tau) \propto & g(\pi)^{\tau_0}\exp\left(\sum_{(i,j)\neq(I,J)}\tau_{i,j}\phi_{i,j}(\pi)\right) \\
=& \left(1-\sum_{(i,j)\neq(I,J)}\pi_{i,j}\right)^{\tau_0}\exp\left(\sum_{(i,j)\neq(I,J)}\tau_{i,j}\log\left(\frac{\pi_{i,j}}{1-\sum_{(i,j)\neq(I,J)}\pi_{i,j}}\right)\right) \\
=& \prod_{(i,j)\neq(I,J)}\pi_{i,j}^{\tau_{i,j}}\pi_{I,J}^{\tau_0-\sum_{(i,j)\neq(I,J)}\tau_{i,j}} \propto \prod_{i=1}^{I}\prod_{j=1}^{J}\pi_{i,j}^{a_{i,j}-1}
\end{aligned}
$$

for some $\tau_0, \tau_{1,1}, ..., \tau_{I-1,J-1}$, and for some $(a_{i,j})_{I,J}$. Hence, it is the Dirichlet distribution

$$
\pi_{X,Y}|a \sim \text{Dir}\left(a\right)
$$

(d) By using the Bayes theorem, I compute the pdf of the posterior distribution as

$$
p(\pi_{X,Y}|n_{X,Y}) \propto f(n_{X,Y}|\pi)p(\pi_{X,Y}) \propto \prod_{i,j=1}^{I,J}\pi_{i,j}^{n_{i,j}}\prod_{i,j=1}^{I,J}\pi_{i,j}^{a_{i,j}-1} = \prod_{i,j=1}^{I,J}\pi_{i,j}^{n_{i,j}+a_{i,j}-1} \propto \frac{1}{B(a^*)}\prod_{i,j=1}^{I,J}\pi_{i,j}^{a_{i,j}^*-1}
$$

and hence the posterior distribution is a Dirichlet distribution as

$$
\pi_{X,Y}|n_{X,Y} \sim \text{Dir}\left(a^*\right)
$$

where $a^* = \left(a^*_{i,j}\right)_{I,J}$ with $a^*_{i,j} = n_{i,j} + a_{i,j}$ for $i = 1, ..., I$ and $j = 1, ..., J$.

(e) I compute the predictive distribution has pmf

$$\begin{aligned}
g(m_{X,Y}|n_{X,Y}) &= \int f(m_{X,Y}|\pi_{X,Y})p(\pi_{X,Y}|n_{X,Y})\mathrm{d}\pi_{X,Y} \\
&= \int \binom{m_{++}}{\boldsymbol{m}_{XY}} \prod_{i=1}^{I}\prod_{j=1}^{J} \pi_{i,j}^{m_{i,j}} \frac{1}{B(a^*)} \prod_{i,j=1}^{I,J} \pi_{i,j}^{a^*_{i,j}-1} \mathrm{d}\pi_{X,Y} \\
&= \binom{m_{++}}{\boldsymbol{m}_{XY}} \frac{1}{B(a^*)} \int \prod_{i=1}^{I}\prod_{j=1}^{J} \pi_{i,j}^{m_{i,j}+a^*_{i,j}-1} \mathrm{d}\pi_{X,Y} \\
&= \binom{m_{++}}{\boldsymbol{m}_{XY}} \frac{B(a^{**})}{B(a^*)}
\end{aligned}$$

where $a^{**} = \left(a^{**}_{i,j}\right)_{I,J}$ with $a^{**}_{i,j} = m_{i,j} + n_{i,j} + a_{i,j}$ for $i = 1, ..., I$ and $j = 1, ..., J$.

2.

(a) It is

$$\pi_{i,j} = \mathsf{P}(X = i, Y = j) = \mathsf{P}(X = i)\,\mathsf{P}(Y = j) = \pi_{i,+}\pi_{+,j}$$

So $\lambda(\boldsymbol{\pi}_Y, \boldsymbol{\pi}_Y) = \boldsymbol{\pi}_{XY}$, is such that $\boldsymbol{\pi}_{XY} = (\pi_{i,j})_{I,J} = (\pi_{i,+}\pi_{+,j})_{I,J}$. Hence, it is

$$\mathscr{B}_2 : \begin{cases} \boldsymbol{n}_{X,Y}|\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y \sim \text{Mult}\left(n_{+,+}, \boldsymbol{\pi}_{XY}\left(\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y\right)\right) \\ \pi_{i,j} = \pi_{i,+}\pi_{+,j}; & i = 1, ..., I, \;\; j = 1, ..., J \\ (\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y) \sim \Pi(\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y) \end{cases}$$

(b) The likelihood is

$$f(\boldsymbol{n}_{XY}|\boldsymbol{\pi}_{XY}) = \binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{\forall i,j} \pi_{i,j}^{n_{i,j}} = \binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{i=1}^{I}\prod_{j=1}^{J} \pi_{i,+}^{n_{i,j}} \pi_{+,j}^{n_{i,j}} = \binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{i=1}^{I} \pi_{i,+}^{n_{i,+}} \prod_{j=1}^{J} \pi_{+,j}^{n_{+,j}}$$

(c) A distribution whose pdf can be formulated as the product of two exponential family distributions is an exponential distribution family. This is obvious from the definition of the exponential family of distributions. Namely,

$$\begin{aligned}
f(\boldsymbol{n}_{XY}|\boldsymbol{\pi}_{XY}) &= \binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{\forall i,j} \pi_{i,j}^{n_{i,j}} = \binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{i=1}^{I}\prod_{j=1}^{J} \pi_{i,+}^{n_{i,j}} \pi_{+,j}^{n_{i,j}} = \binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{i=1}^{I} \pi_{i,+}^{n_{i,+}} \prod_{j=1}^{J} \pi_{+,j}^{n_{+,j}} \\
&= \binom{n_{++}}{\boldsymbol{n}_{XY}} \frac{1}{\binom{n_{++}}{\boldsymbol{n}_X}} \text{Dir}(n_X|\boldsymbol{\pi}_X) \frac{1}{\binom{n_{++}}{\boldsymbol{n}_Y}} \text{Dir}(n_Y|\boldsymbol{\pi}_Y)
\end{aligned}$$

anyway... Note that

$$\begin{aligned}
\frac{1}{\binom{n_{++}}{\boldsymbol{n}_X}} \text{Dir}(n_X|\boldsymbol{\pi}_X) = \prod_{i=1}^{I} \pi_{i,+}^{n_{i,+}} &= \exp\left(\sum_{i\neq I} n_{i,+}\log(\pi_{i,+}) + \left(n_{++} - \sum_{i\neq I} n_{i,+}\right)\log\left(1 - \sum_{i\neq I}\pi_{i,+}\right)\right) \\
&= \exp\left(n_{++}\log\left(1 - \sum_{i\neq I}\pi_{i,+}\right) + \sum_{i\neq I} n_{i,+}\log\left(\frac{\pi_{i,+}}{1 - \sum_{i\neq I}\pi_{i,+}}\right)\right) \\
&= \left(1 - \sum_{i\neq I}\pi_{i,+}\right)^{n_{++}} \exp\left(\sum_{i\neq I} n_{i,+}\log\left(\frac{\pi_{i,+}}{1 - \sum_{i\neq I}\pi_{i,+}}\right)\right)
\end{aligned}$$

and similar for

$$\frac{1}{\binom{n_{++}}{\boldsymbol{n}_Y}}\mathrm{Dir}(n_Y|\boldsymbol{\pi}_Y) = \left(1-\sum_{j\neq J}\pi_{+,j}\right)^{n_{++}} \exp\left(\sum_{j\neq J} n_{+,j}\log\left(\frac{\pi_{+,j}}{1-\sum_{j\neq J}\pi_{+,j}}\right)\right)$$

We put everything together ...

$$\begin{aligned}
f(\boldsymbol{n}_{XY}|\boldsymbol{\pi}_{XY}) =& \binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{i=1}^{I}\pi_{i,+}^{n_{i,+}}\prod_{j=1}^{J}\pi_{+,j}^{n_{+,j}} \\
=& \binom{n_{++}}{\boldsymbol{n}_{XY}} \exp\left(\sum_{i\neq I} n_{i,+}\log(\pi_{i,+}) + \left(n_{++}-\sum_{i\neq I}n_{i,+}\right)\log\left(1-\sum_{i\neq I}\pi_{i,+}\right)\right) \\
&\times \exp\left(\sum_{j\neq J} n_{+,j}\log(\pi_{+,j}) + \left(n_{++}-\sum_{j\neq J}n_{+,j}\right)\log\left(1-\sum_{j\neq J}\pi_{+,j}\right)\right) \\
=& \binom{n_{++}}{\boldsymbol{n}_{XY}} \left(1-\sum_{i\neq I}\pi_{i,+}\right)^{n_{++}}\left(1-\sum_{j\neq J}\pi_{+,j}\right)^{n_{++}} \\
&\times \exp\left(\sum_{i\neq I} n_{i,+}\log\left(\frac{\pi_{i,+}}{1-\sum_{i\neq I}\pi_{i,+}}\right) + \sum_{j\neq J} n_{+,j}\log\left(\frac{\pi_{+,j}}{1-\sum_{j\neq J}\pi_{+,j}}\right)\right)
\end{aligned}$$

So it is an $I+J-2$ exponential family of distributions with parameters

$$u(n) = \binom{n_{++}}{\boldsymbol{n}_{XY}}, \qquad g(\pi) = \left(1-\sum_{i\neq I}\pi_{i,+}\right)^{n_{++}}\left(1-\sum_{j\neq J}\pi_{+,j}\right)^{n_{++}},$$

$$\phi(\pi) = \left(\log\left(\frac{\pi_{1,+}}{1-\sum_{i\neq I}\pi_{i,+}}\right),...,\log\left(\frac{\pi_{I-1,+}}{1-\sum_{i\neq I}\pi_{i,+}}\right),\log\left(\frac{\pi_{+,1}}{1-\sum_{j\neq J}\pi_{+,j}}\right),...,\log\left(\frac{\pi_{+,J-1}}{1-\sum_{j\neq J}\pi_{+,j}}\right)\right)$$

$h(n_{X,Y}) = (n_{1,+},...,n_{i-1,+},n_{+,1},...,n_{+,J-1})$, and $c=1$, for $i=1,...,I-1$, and $j=1,...,J-1$.
The sufficient statistic is

$$h(n_{X,Y}) = (n_{1,+},...,n_{i-1,+},n_{+,1},...,n_{+,J-1})$$

(d) I can use the standard procedure, where I substitute the data quantities in the likelihood kernel, aka

$$\begin{aligned}
f(\boldsymbol{n}_{XY}|\boldsymbol{\pi}_{XY}) =& \binom{n_{++}}{\boldsymbol{n}_{XY}} \left(1-\sum_{i\neq I}\pi_{i,+}\right)^{n_{++}}\left(1-\sum_{j\neq J}\pi_{+,j}\right)^{n_{++}} \\
&\times \exp\left(\sum_{i\neq I} n_{i,+}\log\left(\frac{\pi_{i,+}}{1-\sum_{i\neq I}\pi_{i,+}}\right) + \sum_{j\neq J} n_{+,j}\log\left(\frac{\pi_{+,j}}{1-\sum_{j\neq J}\pi_{+,j}}\right)\right)
\end{aligned}$$

with a kernel

$$k(t(\boldsymbol{n}_{XY})) = \exp\left(\sum_{i\neq I} n_{i,+}\log\left(\frac{\pi_{i,+}}{1-\sum_{i\neq I}\pi_{i,+}}\right) + \sum_{j\neq J} n_{+,j}\log\left(\frac{\pi_{+,j}}{1-\sum_{j\neq J}\pi_{+,j}}\right)\right)$$

sufficient statistic

$$h(n_{X,Y}) = (n_{1,+},...,n_{i-1,+},n_{+,1},...,n_{+,J-1})$$

and residual

$$\varrho(n_{X,Y}) = \binom{n_{++}}{\boldsymbol{n}_{XY}} \left(1 - \sum_{i \neq I} \pi_{i,+}\right)^{n_{++}} \left(1 - \sum_{j \neq J} \pi_{+,j}\right)^{n_{++}}$$

So the conjugate prior is

$$\begin{aligned} k(\boldsymbol{\tau}) &= \exp\left(\sum_{i \neq I} \tau_{i,+} \log\left(\frac{\pi_{i,+}}{1 - \sum_{i \neq I} \pi_{i,+}}\right) + \sum_{j \neq J} \tau_{+,j} \log\left(\frac{\pi_{+,j}}{1 - \sum_{j \neq J} \pi_{+,j}}\right)\right) \\ &\propto \prod_{i=1}^{I} \pi_{i,+}^{b_{i,+}} \prod_{j=1}^{J} \pi_{+,j}^{b_{+,j}} \end{aligned}$$

for some hyper-parameters $(b_{i,+})$ and $(b_{+,j})$ where I can recognize two independent Dirichlet distributions, as

$$p\left(\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y\right) = \text{Dir}\left(\boldsymbol{\pi}_X | n_X, b_X\right) \text{Dir}\left(\boldsymbol{\pi}_Y | n_Y, b_Y\right)$$

where $b_X = (b_{i,+})$ and $b_Y = (b_{+,j})$. Hence,

$$\begin{aligned} \boldsymbol{\pi}_X &\sim \text{Dir}\left(n_X, b_X\right) \\ \boldsymbol{\pi}_Y &\sim \text{Dir}\left(n_Y, b_Y\right) \\ \boldsymbol{\pi}_X &\perp \boldsymbol{\pi}_Y \end{aligned}$$

(e) The posterior is

$$\begin{aligned} p(\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y | n_{X,Y}) &\propto f(n_{X,Y} | \boldsymbol{\pi}_X, \boldsymbol{\pi}_Y) p\left(\boldsymbol{\pi}_X, \boldsymbol{\pi}_Y\right) \\ &= \binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{i=1}^{I} \prod_{j=1}^{J} \pi_{i,+}^{n_{i,j}} \pi_{+,j}^{n_{i,j}} \frac{1}{B\left(b_X\right)} \prod_{i=1}^{I} \pi_{i,+}^{b_{i,+}-1} \frac{1}{B\left(b_Y\right)} \prod_{j=1}^{J} \pi_{+,j}^{b_{+,j}-1} \\ &\propto \prod_{i=1}^{I} \pi_{i,+}^{n_{i,+}} \prod_{j=1}^{J} \pi_{+,j}^{n_{+,j}} \prod_{i=1}^{I} \pi_{i,+}^{b_{i,+}} \prod_{j=1}^{J} \pi_{+,j}^{b_{+,j}} = \prod_{i=1}^{I} \pi_{i,+}^{n_{i,+}+b_{i,+}-1} \prod_{j=1}^{J} \pi_{+,j}^{n_{+,j}+b_{+,j}-1} \\ &\propto \left[\frac{1}{B\left(b_X^*\right)} \prod_{i=1}^{I} \pi_{i,+}^{n_{i,+}+b_{i,+}-1}\right] \left[\frac{1}{B\left(b_Y^*\right)} \prod_{j=1}^{J} \pi_{+,j}^{n_{+,j}+b_{+,j}-1}\right] \end{aligned}$$

where $b_X^* = (n_{i,+} + b_{i,+})_I$ and $b_Y^* = (n_{+,j} + b_{+,j})_J$. Hence I can recognize that the posterior distribution is a Dirichlet distribution. So

$$\begin{aligned} \boldsymbol{\pi}_X | n_X &\sim \text{Dir}\left(\boldsymbol{\pi}_X | n_X, b_X\right) \\ \boldsymbol{\pi}_Y | n_Y &\sim \text{Dir}\left(\boldsymbol{\pi}_Y | n_Y, b_Y\right) \\ \boldsymbol{\pi}_X &\perp \boldsymbol{\pi}_Y | n_{X,Y} \end{aligned}$$

aka $\boldsymbol{\pi}_X$ and $\boldsymbol{\pi}_Y$ are a posteriori independent.

(f) It is $m_{X,Y}=\left(m_{i,j}\right)_{i,j}$. I compute the predictive distribution has pmf

$$
\begin{aligned}
g(m_{X,Y}|n_{X,Y}) &= \int f(m_{X,Y}|\pi_{X,Y})p(\pi_{X,Y}|n_{X,Y})\mathrm{d}\pi_{X,Y} \\
&= \int \binom{m_{++}}{\boldsymbol{m}_{XY}} \prod_{i=1}^{I}\pi_{i,+}^{m_{i,+}} \prod_{j=1}^{J}\pi_{+,j}^{m_{+,j}} \frac{1}{B\left(b_X^*\right)} \prod_{i=1}^{I}\pi_{i,+}^{n_{i,+}+b_{i,+}-1} \frac{1}{B\left(b_Y^*\right)} \prod_{j=1}^{J}\pi_{+,j}^{n_{+,j}+b_{+,j}-1}\mathrm{d}\pi_{X,Y} \\
&= \binom{m_{++}}{\boldsymbol{m}_{XY}} \frac{1}{B\left(b_X^*\right)} \frac{1}{B\left(b_Y^*\right)} \left[\int \prod_{i=1}^{I}\pi_{i,+}^{m_{i,+}+n_{i,+}+b_{i,+}-1}\mathrm{d}\pi_X\right] \left[\int \prod_{j=1}^{J}\pi_{+,j}^{m_{+,j}+n_{+,j}+b_{+,j}-1}\mathrm{d}\pi_Y\right] \\
&= \binom{m_{++}}{\boldsymbol{m}_{XY}} \frac{1}{B\left(b_X^*\right)} \frac{1}{B\left(b_Y^*\right)} B\left(b_X^{**}\right) B\left(b_Y^{**}\right) \quad = \binom{m_{++}}{\boldsymbol{m}_{XY}} \frac{B\left(b_X^{**}\right)}{B\left(b_X^*\right)} \frac{B\left(b_Y^{**}\right)}{B\left(b_Y^*\right)}
\end{aligned}
$$

where:

$$b_X^{**}=\left(b_{i,+}^{**}\right)_I \quad \text{with } b_{i,+}^{**}=m_{i,+}+n_{i,+}+b_{i,+} \quad \text{for} \quad i=1,...,I$$

and

$$b_Y^{**}=\left(b_{+,j}^{**}\right)_J \quad \text{with} \quad b_{+,j}^{**}=m_{+,j}+n_{+,j}+b_{+,j} \quad \text{for} \quad j=1,...,J.$$

(g) Well, for $I=J=2$, it is

$$p(\pi_{1,+}|n_X)=\frac{1}{B\left(b_X^*\right)}\prod_{i=1}^{2}\pi_{i,+}^{n_{i,+}+b_{i,+}-1}=\frac{1}{B\left(b_{1,+},b_{2,+}\right)}\pi_{1,+}^{n_{1,+}+b_{1,+}-1}\pi_{2,+}^{n_{2,+}+b_{2,+}-1}$$

so in simple words

$$\pi_{1,+}|n_X \sim \text{Be}\left(b_{1,+}^*,b_{2,+}^*\right)$$

where

$$b_{1,+}^*=n_{1,+}+b_{1,+}$$

and

$$b_{2,+}^*=n_{2,+}+b_{2,+}$$

This is obvious as Beta distribution is a specific case of the Dirichlet distribution. So now we will proceed with the Beta distribution with which we are most familiar, ok.

Also

$$\begin{cases}\text{E}\left(\pi_{1,+}|n_X\right)=\frac{1}{2}\\ \text{Var}\left(\pi_{1,+}|n_X\right)=\frac{1}{140}\end{cases} \implies \begin{cases}\frac{b_{1,+}^*}{b_{1,+}^*+b_{2,+}^*}=\frac{1}{2}\\ \frac{b_{1,+}^*b_{2,+}^*}{\left(b_{1,+}^*+b_{2,+}^*\right)^2\left(b_{1,+}^*+b_{2,+}^*+1\right)}=\frac{1}{140}\end{cases} \implies \begin{cases}b_{1,+}^*=17\\ b_{2,+}^*=17\end{cases}$$

To find the 2-sides C_a parametric HPD credible interval for $\pi_{1,+}$, I use the Theorem in the Handout about 1D HPD credit intervals. So:

$$1-a=\int_L^U \text{Be}(\theta|17,17)\mathrm{d}\theta=\mathsf{P}_{\text{Be}(17,17)}(\theta<U)-\mathsf{P}_{\text{Be}(17,17)}(\theta<L)$$

The posterior is symmetric around 0.5 because $a_n=b_n$, hence

$$\text{Be}(U|17,17)=\text{Be}(L|17,17) \implies 0.5-L=U-0.5 \implies L=1-U$$

Hence,

$$1-a=\mathsf{P}_{\text{Be}(17,17)}(\theta<U)-\left(1-\mathsf{P}_{\text{Be}(17,17)}(\theta<U)\right)=2\mathsf{P}_{\text{Be}(17,17)}(\theta<U)-1$$

so U is such that $\mathsf{P}_{\mathrm{Be}(17,17)}(\theta < U) = 1 - a/2$ and $L = 1 - U$. For $a = 0.95$, the 95% posterior credible interval for θ is

$$[L, U] = [0.36, 0.64].$$

3. I will use the results from the previous parts. To denote the pdf of prior and posteriors I ll use the notation $p(\cdot)$ instead of $\pi(\cdot)$ to avoid confusion.

   (a) He compare two models, the hypothesis H_0 is the Bayesian model

$$\mathscr{B}_1 : \begin{cases} \boldsymbol{n}_{X,Y} | \boldsymbol{\pi}_{X,Y} & \sim \text{Mult}\left(n_{+,+}, \boldsymbol{\pi}_{X,Y}\right) \\ \boldsymbol{\pi}_{X,Y} & \sim \text{Dir}\left(a\right) \end{cases}$$

   while the hypothesis H_1 is the Bayesian model

$$\mathscr{B}_2 : \begin{cases} \boldsymbol{n}_{X,Y} | \boldsymbol{\pi}_X, \boldsymbol{\pi}_Y & \sim \text{Mult}\left(n_{+,+}, \boldsymbol{\pi}_X \boldsymbol{\pi}_Y\right) \ \text{ where } \pi_{i,j} = \pi_{i,+}\pi_{+,j};\ i = 1, ..., I, \ \ j = 1, ..., J \\ \boldsymbol{\pi}_X & \sim \text{Dir}\left(n_X, b_X\right) \\ \boldsymbol{\pi}_Y & \sim \text{Dir}\left(n_Y, b_Y\right) \\ \boldsymbol{\pi}_X \perp \boldsymbol{\pi}_Y & \end{cases}$$

   (b) For model $\mathscr{B}_1$ the marginal likelihood is

$$\begin{aligned} f_1(\boldsymbol{n}_{X,Y}) &= \int f_1(\boldsymbol{n}_{XY} | \boldsymbol{\pi}_{XY}) p_1(\boldsymbol{\pi}_{X,Y}) \mathrm{d}\boldsymbol{\pi}_{XY} = \\ &= \int \left[\binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{i=1}^{I} \prod_{j=1}^{J} \pi_{i,j}^{n_{i,j}} \right] \left[\frac{1}{B(a)} \prod_{i=1}^{I} \prod_{j=1}^{J} \pi_{i,j}^{a_{i,j}-1} \right] \mathrm{d}\boldsymbol{\pi}_{XY} \\ &= \binom{n_{++}}{\boldsymbol{n}_{XY}} \frac{1}{B(a)} \int \prod_{i=1}^{I} \prod_{j=1}^{J} \pi_{i,j}^{n_{i,j}+a_{i,j}-1} \mathrm{d}\boldsymbol{\pi}_{XY} \\ &= \binom{n_{++}}{\boldsymbol{n}_{XY}} \frac{B\left(a^*\right)}{B\left(a\right)} \end{aligned}$$

   $a^* = \left(a^*_{i,j}\right)_{I,J}$ with $a^*_{i,j} = n_{i,j} + a_{i,j}$ for $i = 1, ..., I$ and $j = 1, ..., J$.
   For model $\mathscr{B}_2$ the marginal likelihood is

$$\begin{aligned} f_2(\boldsymbol{n}_{X,Y}) &= \int f_2(\boldsymbol{n}_{XY} | \boldsymbol{\pi}_{XY}) p_2(\boldsymbol{\pi}_{X,Y}) \mathrm{d}\boldsymbol{\pi}_{XY} = \\ &= \int \left[\binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{i=1}^{I} \prod_{j=1}^{J} \pi_{i,+}^{n_{i,j}} \pi_{+,j}^{n_{i,j}} \right] \left[\frac{1}{B(b_X)} \prod_{i=1}^{I} \pi_{i,+}^{b_{i,+}-1} \frac{1}{B(b_Y)} \prod_{j=1}^{J} \pi_{+,j}^{b_{+,j}-1} \right] \mathrm{d}\boldsymbol{\pi}_{XY} \\ &= \int \left[\binom{n_{++}}{\boldsymbol{n}_{XY}} \prod_{i=1}^{I} \pi_{i,+}^{n_{i,+}} \prod_{j=1}^{J} \pi_{+,j}^{n_{+,j}} \right] \left[\frac{1}{B(b_X)} \prod_{i=1}^{I} \pi_{i,+}^{b_{i,+}-1} \frac{1}{B(b_Y)} \prod_{j=1}^{J} \pi_{+,j}^{b_{+,j}-1} \right] \mathrm{d}\boldsymbol{\pi}_{XY} \\ &= \binom{n_{++}}{\boldsymbol{n}_{XY}} \frac{1}{B(b_X) B(b_Y)} \left[\int \prod_{i=1}^{I} \pi_{i,+}^{n_{i,+}+b_{i,+}-1} \mathrm{d}\boldsymbol{\pi}_{XY} \right] \left[\int \prod_{j=1}^{J} \pi_{+,j}^{n_{+,j}+b_{+,j}-1} \mathrm{d}\boldsymbol{\pi}_{Y} \right] \\ &= \binom{n_{++}}{\boldsymbol{n}_{XY}} \frac{B(b_X^*) B(b_Y^*)}{B(b_X) B(b_Y)} \end{aligned}$$

   where $b_X^* = \left(n_{i,+} + b_{i,+}\right)_{i=1,...,I}$ and $b_Y^* = \left(n_{+,j} + b_{+,j}\right)_{j=1,...,J}$.

The Bayes factor $\mathrm{B}_{01}(\boldsymbol{n}_{X,Y})$ comparing hypotheses H_0 and H_1 is

$$\mathrm{B}_{01}(\boldsymbol{n}_{X,Y}) = \frac{f_1(\boldsymbol{n}_{X,Y})}{f_2(\boldsymbol{n}_{X,Y})} = \frac{\binom{n_{++}}{\boldsymbol{n}_{XY}}\frac{B(a^*)}{B(a)}}{\binom{n_{++}}{\boldsymbol{n}_{XY}}\frac{B(b_X^*)B(b_Y^*)}{B(b_X)B(b_Y)}} = ...$$

(c) We accept model $\mathscr{B}_1$ if $\mathrm{B}_{01}(\boldsymbol{n}_{X,Y}) > \frac{c_{II}}{c_I}\frac{\pi_1}{\pi_0}$, and we accept $\mathscr{B}_2$ if $\mathrm{B}_{01}(\boldsymbol{n}_{X,Y}) < \frac{c_{II}}{c_I}\frac{\pi_1}{\pi_0}$.
