Bayesian Statistics III/IV (MATH3341/4031) Michaelmas term, 2019

# Handout 11: Bayesian point estimation [a]

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

**Aim:** To explain and produce point estimators in the Bayesian framework.

**References:**

- Raiffa, H., & Schlaifer, R. (1961; Chapter 6). Applied statistical decision theory
- Berger, J. O. (2013; Section 4.3.1). Statistical decision theory and Bayesian analysis. Springer Science & Business Media.
- DeGroot, M. H. (2005, Sections 11.1-11.4). Optimal statistical decisions (Vol. 82). John Wiley & Sons.
- Robert, C. (2007; Chapter 4.1). The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media.

**Web applets:**

`https://georgios-stats-1.shinyapps.io/demo_pointestimation/`

[a] Author: Georgios P. Karagiannis.

## 1 Set-up and aim

*Notation* 1. Consider a Bayesian model

$$\begin{cases} y|\theta & \sim F(y|\theta) \\ \theta & \sim \Pi(\cdot) \end{cases}$$

where $y := (y_1, ..., y_n) \in \mathcal{Y}$ is a sequence of observables, assumed to be generated from the parametric sampling distribution $F(y|\theta)$ with pdf/pmf $f(y|\theta)$ and labeled by an unknown parameter $\theta \in \Theta$ following a prior distribution $\Pi(\theta)$ with pdf/pmf $\pi(\theta)$.

**The AIM, in parametric (or predictive) point estimation,** is to derive a parametric (or predictive) point estimator $\hat{\delta}(y)$, a quantity summarizing Your believes about the unknown parameter $\theta$ (or unknown future outcome sequence $z := (y_{n+1}, ..., y_{n+m})$) or any function of it in an appropriate manner.

**Point estimation is addressed** in the statistical decision theory framework, where the decision rule is the estimator $\hat{\delta}(y)$, (the decision space results consequently) and the loss function $\ell(\cdot, \cdot)$ is set in a subjective manner as a penalty and according to what You may loss or how You may suffer if you use the produced estimator.

*Note* 2. Although we define the parametric and predictive point estimation as two distinct concepts, it will become clear in Section 4 they are treated in a similar manner.

## 2 Parametric point estimation[1]

*Note* 3. Posterior degree of believe about uncertain parameter $\theta \in \Theta \subseteq \mathbb{R}^d$ is quantified via the posterior distribution $\Pi(\theta|y)$ ,

$$\mathrm{d}\Pi(\theta|y) = \pi(\theta|y)\mathrm{d}\theta$$

with cdf $\Pi(\theta|y)$ and pdf/pmf $\pi(\theta|y)$.

**Definition 4.** Bayes point estimator $\hat{\delta} = \hat{\delta}(y)$ of $\theta$ under the loss function $\ell(\theta, \delta)$ and the posterior distribution $\Pi(\theta|y)$ is an Bayes rule (which minimizes the posterior expected loss $\varrho(\pi, d|y) = \mathrm{E}_\Pi(\ell(\theta, \delta)|y)$); i.e.

$$\hat{\delta} = \arg\min_{\forall \delta \in \mathcal{D}} \mathrm{E}_\Pi(\ell(\theta, \delta)|y) = \arg\min_{\forall \delta \in \mathcal{D}} \underbrace{\int_\Theta \ell(\theta, \delta)\mathrm{d}\Pi(\theta|y)}_{=\varrho(\pi, \delta|y)}$$

*Note* 5. Traditionally the accuracy of a statistical estimator (called standard error) is described by the squared root of the mean square error of the estimator.

**Definition 6.** Let $\hat{\delta} = (\hat{\delta}_1, ..., \hat{\delta}_d)$ be the Bayes point estimator of $\theta \in \Theta \subseteq \mathbb{R}^d$ with posterior distribution $\Pi(\theta|y)$. The standard error of the $j$-th dimension of $\hat{\delta}$ is defined as

$$\mathrm{se}_\Pi\left(\hat{\delta}_j|y\right) = \sqrt{\left[\mathrm{mse}_\Pi(\hat{\delta}|y)\right]_{j,j}}$$

where

$$\mathrm{mse}_\Pi\left(\hat{\delta}|y\right) = \mathrm{E}_\Pi\left((\theta - \hat{\delta})(\theta - \hat{\delta})^\top|y\right)$$

is the mean squared error of $\hat{\delta}$.

**Proposition 7.** *MSE of (any) estimator $\hat{\delta}$ of $\theta$ following posterior distribution $\Pi(\theta|y)$ can be decomposed as*

$$E_\Pi\left((\theta - \hat{\delta})(\theta - \hat{\delta})^\top|y\right) \;=\; Var_\Pi(\theta|y) + \left(E_\Pi(\theta|y) - \hat{\delta}\right)\left(E_\Pi(\theta|y) - \hat{\delta}\right)^\top$$

*Remark* 8. By Definition 6, the mse of 1-dim Bayes point estimator $\delta$ of $\theta$ with posterior distribution $\Pi(\theta|y)$ is

$$\mathrm{se}_\Pi\left(\hat{\delta}|y\right) = \sqrt{\mathrm{mse}_\Pi(\delta|y)}$$

where

$$\mathrm{mse}_\Pi\left(\hat{\delta}|y\right) = \mathrm{E}_\Pi\left((\theta - \hat{\delta})^2|y\right) = \mathrm{Var}_\Pi(\theta|y) + \left(\mathrm{E}_\Pi(\theta|y) - \hat{\delta}\right)^2$$

## 3 Predictive point estimation

*Note* 9. The Bayesian point predictive estimator and its standard error are defined similar to the parametric ones.

*Note* 10. Degree of believe about a future sequence of outcomes $z = (y_{n+1}, ..., y_{n+m}) \in \mathcal{Z}$ is quantified via the predictive distribution

$$\mathrm{d}G(z|y) = g(z|y)\mathrm{d}z$$

with cdf $G(z|y)$ and pdf/pmf $g(z|y)$.

**Definition 11.** Bayes predictive point estimator of $z = (y_{n+1}, ..., y_{n+m}) \in \mathcal{Z}$ under the loss function $\ell(z, \delta)$ and predictive distribution $G(z|y)$ is the decision rule $\delta \in \mathcal{D} = \mathcal{Z}$ which minimizes $\mathrm{E}_G(\ell(z, \delta)|y)$; i.e.

$$\hat{\delta} = \arg\min_{\forall \delta} \mathrm{E}_G(\ell(z, \delta)|y) = \arg\min_{\forall \delta \in \mathcal{D}} \int_{\mathcal{Z}} \ell(y, \delta)\mathrm{d}G(z|y)$$

[1] Web applet: `https://georgios-stats-1.shinyapps.io/demo_pointestimation/`

*Note* 12. The accuracy of the predictive point estimator is traditionally presented by using the squared mean (predictive) square error.

**Definition 13.** Let $\hat{\delta} = (\hat{\delta}_1, ..., \hat{\delta}_m)$ be the Bayes point predictive estimator of $z \in \mathcal{Z} \subseteq \mathbb{R}^m$ with predictive distribution $G(z|y)$. Then the standard error of the $j$-th dimension of $\hat{\delta}$ is defined as

$$\text{se}_G(\hat{\delta}_j|y) = \sqrt{\left[\text{mse}_\Pi(\hat{\delta}|y)\right]_{j,j}}$$

where

$$\text{mse}_G(\hat{\delta}|y) = \text{E}_G\left((z-\hat{\delta})(z-\hat{\delta})^\top|y\right)$$

is the mean squared error of $\hat{\delta}$.

## 4 Popular point estimators[2]

*Note* 14. A number of standard Bayesian point estimators, along with the corresponding loss functions are examined below. These point estimators correspond to summary statistics of the posterior/predictive distribution (mean, median, mode, quantiles, etc.).

*Note* 15. Bayesian point estimators are applied to both parametric and predictive inference likewise and hence presented together in what follows.

*Notation* 16. Consider unknown random quantity $x \in \mathcal{X} \subseteq \mathbb{R}^k$ following a distribution $Q(x|y)$ such as

$$\text{d}Q(x|y) = q(x|y)\text{d}x$$

with cdf $Q(x|y)$ and pdf/pmf $q(x|y)$. These are dummy quantities for the following:

- In parametric inference, we have $x \equiv \theta$, $Q \equiv \Pi$, $q \equiv \pi$, and $k = d$.
- In predictive inference, we have $x \equiv z$, $Q \equiv G$, $q \equiv g$, and $k = m$.
- In more extreme cases, we have $x \equiv (z, \theta)$, $Q \equiv P$, $q \equiv p$, and $k = d + m$.
- Note that $x$ can also be any function of $\theta$ or $z$, or $(z, \theta)$...

**Quadratic loss function**

**Proposition 17.** *The Bayes point estimate $\hat{\delta}$ of $x$ with respect to the quadratic loss function $\ell(x, \delta) = (x-\delta)^\top H(x-\delta)$, where $H > 0$, is*

$$\hat{\delta} = E_Q(x|y) \tag{1}$$

*Proof.* It is

$$\begin{aligned} 0 = \frac{\text{d}}{\text{d}\delta}\int \ell(x,\delta)\text{d}Q(x|y)\Big|_{\delta=\hat{\delta}} &= \frac{\text{d}}{\text{d}\delta}\int (x-\delta)H(x-\delta)^\top \text{d}Q(x|y)\Big|_{\delta=\hat{\delta}} \\ = -2H\int (x-\hat{\delta})\text{d}Q(x|y) \quad &= -2H\int \theta\text{d}Q(x|y) + 2H\hat{\delta} \quad = -2H\text{E}_Q(x|y) + 2H\hat{\delta}. \end{aligned}$$

□

*Remark* 18. If $k = 1$, then $\ell(x, \delta) = H(x-\delta)^2$, with $H > 0$, and the Bayes point estimate is $\hat{\delta} = \text{E}_Q(x|y)$.

*Remark* 19. The loss in Proposition 17 has the same effect as $\ell(x, \delta) = \|x-\delta\|_2^2$ and hence $H$ has no effect.

[2]Web applet: `https://georgios-stats-1.shinyapps.io/demo_pointestimation/`

*Remark* 20. The point estimator in Proposition 17 minimizes the standard error, as

$$\text{se}(\hat{\delta}|y) = \sqrt{\text{mse}_Q(\hat{\delta}|y)} = \sqrt{\text{Var}_Q(x|y) + \cancel{\left(\text{E}_Q(x|y) - \delta\right)^2}} = 0 = \sqrt{\text{Var}_Q(x|y)}$$

**Weighted quadratic loss function**

**Proposition 21.** *The Bayes estimate $\hat{\delta}$ of $x$ under the weighted quadratic loss function $\ell(x,\delta) = w(x)(x-\delta)^2$, where $w(x)$ as a non negative function with $E_Q(w(x)|y) > 0$, is*

$$\delta^\pi(y) = \frac{E_Q(w(x)x|y)}{E_Q(w(x)|y)}. \tag{2}$$

*Proof.* It is

$$\begin{aligned} 0 = \frac{\text{d}}{\text{d}\delta}\int_{\mathcal{X}} \ell(x,\delta)\text{d}Q(x|y)\Big|_{\delta=\hat{\delta}} &= \frac{\text{d}}{\text{d}\delta}\int_{\mathcal{X}} w(x)(x-\hat{\delta})^2\text{d}Q(x|y)\Big|_{\delta=\hat{\delta}} \\ = 2\int_{\mathcal{X}} w(x)(x-\hat{\delta})(-1)\text{d}Q(x|y) &= -2\left[\int_{\mathcal{X}} w(x)x\text{d}Q(x|y)\right] + 2\left[\int_{\mathcal{X}} w(\theta)\text{d}Q(x|y)\right]\hat{\delta}. \\ = -2\text{E}_Q(w(x)x|y) + 2\text{E}_Q(w(x)|y)\hat{\delta} &= 2(\text{E}_Q(w(x)x|y) - \text{E}_Q(w(x)|y)\hat{\delta}) \end{aligned}$$

Also, $\frac{\text{d}^2}{\text{d}\delta^2}\int_{\mathcal{X}} \ell(x,\hat{\delta})\text{d}Q(x|y) = -2\text{E}_Q(w(x)|y) < 0$. This completes the proof. □

*Remark* 22. Weighted quadratic loss allows the discrepancy $(x-\delta)^2$ to vary with $x$. This loss function is appropriate in cases where a given discrepancy in estimation can vary in harm according to what $x$ happens to be.

**Example 23.** Consider there is interest in performing parametric inference for $\theta$. Show that Proposition 21 exhibits a duality between loss and prior distribution, in the sense that it is equivalent to estimate $\theta$ under loss $\ell(\theta,\delta) = w(\theta)(\theta-\delta)^2$ with prior pdf $\pi(\theta)$ (under (2)), or under loss $\tilde{\ell}(\theta,\delta) = (\theta-\delta)^2$ with prior pdf $\tilde{\pi}(\theta) \propto \pi(\theta)$ (under 1).

**Solution.** The estimator of $\theta$ under

- loss $\ell(\theta,\delta) = w(\theta)(\theta-\delta)^2$ and prior pdf $\pi(\theta)$ is $\hat{\delta}(y) = \frac{\text{E}_\Pi(w(\theta)\theta|y)}{\text{E}_\Pi(w(\theta)|y)}$
- loss $\tilde{\ell}(\theta,\delta) = (\theta-\delta)^2$ the prior pdf $\tilde{\pi}(\theta) \propto \pi(\theta)$ is $\tilde{\delta}(y) = \text{E}_{\tilde{\Pi}}(\theta|y)$

But

$$\begin{aligned} \tilde{\delta}(y) = \text{E}_{\tilde{\Pi}}(\theta|y) = \int_\Theta \theta\frac{f(y|\theta)\tilde{\pi}(\theta)}{\int_\Theta f(y|\theta)\tilde{\pi}(\theta)\text{d}\theta}\text{d}\theta &= \int_\Theta \theta\frac{f(y|\theta)w(\theta)\pi(\theta)}{\int_\Theta f(y|\theta)w(\theta)\pi(\theta)\text{d}\theta}\text{d}\theta = \frac{\int_\Theta \theta f(y|\theta)w(\theta)\pi(\theta)\text{d}\theta}{\int_\Theta f(y|\theta)w(\theta)\pi(\theta)\text{d}\theta} \\ = \frac{\int_\Theta \theta w(\theta)\frac{f(y|\theta)\pi(\theta)}{\int_\Theta f(y|\theta)\pi(\theta)\text{d}\theta}\text{d}\theta}{\int_\Theta w(\theta)\frac{f(y|\theta)\pi(\theta)}{\int_\Theta f(y|\theta)\pi(\theta)\text{d}\theta}\text{d}\theta} &= \frac{\int_\Theta \theta w(\theta)\pi(\theta|y)\text{d}\theta}{\int_\Theta w(\theta)\pi(\theta|y)\text{d}\theta} = \frac{\text{E}_\Pi(w(\theta)\theta|y)}{\text{E}_\Pi(w(\theta)|y)} = \hat{\delta}(y) \end{aligned}$$

- This is an example supporting that loss and prior are difficult to separate and may be specified/analysed simultaneously.

**Linear loss function**

**Proposition 24.** *The Bayes estimate $\hat{\delta}$ of $x$ under the linear loss function*

$$\ell(x,\delta) = c_1(\delta-x)I_{\{x\leq\delta\}}(\delta) + c_2(x-\delta)I_{\{x\leq\delta\}^\complement}(\delta)$$

*is the $\frac{c_2}{c_1+c_2}$-th quantile of distribution $Q$, namely*

$$\hat{\delta} \text{ is such that } \mathsf{P}_Q\left(x \leq \hat{\delta}|y\right) = \frac{c_2}{c_1+c_2}.$$

*Proof.* The posterior expected loss is

$$\begin{aligned}\int \ell(x,\delta)\mathrm{d}Q(x|y) &= \int c_1(\delta-\theta)1_{\{x\leq\delta\}}(\delta)\mathrm{d}Q(x|y) + \int c_2(\theta-\delta)1_{\{x\leq\delta\}^{\complement}}(\delta)\mathrm{d}Q(x|y)\\ &= c_1\int_{-\infty}^{\delta}(\delta-x)\mathrm{d}Q(x|y) + c_2\int_{\delta}^{+\infty}(x-\delta)\mathrm{d}Q(x|y)\\ &= c_1\int_{-\infty}^{\delta}\mathsf{P}_Q\left(x\leq\hat{\delta}|y\right)\mathrm{d}x - c_2\int_{\delta}^{+\infty}\mathsf{P}_Q\left(x>\hat{\delta}|y\right)\mathrm{d}x\end{aligned}$$

by transformation as $\xi = -x$ in the second part and integration by parts[3]. We seek for the maximizer of it w.r.t. $\delta$ by

$$\begin{aligned}0 = \frac{\mathrm{d}}{\mathrm{d}\delta}\int \ell(x,\delta)\mathrm{d}Q(x|y)\Big|_{\delta=\hat{\delta}} &\overset{(*)}{=} c_1\mathsf{P}_Q\left(\{x\leq\hat{\delta}\}|y\right) - c_2\mathsf{P}_Q\left(\{x\leq\hat{\delta}\}^{\complement}|y\right)\\ \iff -c_2\mathsf{P}_Q\left(\{x\leq\hat{\delta}\}|y\right) &= c_1\mathsf{P}_Q\left(\{x\leq\hat{\delta}\}|y\right) - c_2\mathsf{P}_Q\left(\{x\leq\hat{\delta}\}^{\complement}|y\right) - c_2\mathsf{P}_Q\left(\{x\leq\hat{\delta}\}|y\right)\\ \iff \mathsf{P}_Q\left(x\leq\hat{\delta}|y\right) &= \frac{c_2}{c_1+c_2}.\end{aligned}$$

the derivative in $(*)$ was computed by the Fundamental Theorem of Calculus. □

*Note* 25. Below, we introduce the absolute loss as a special case of the linear loss which leads to median summaries.

**Proposition 26.** *The Bayes estimate $\hat{\delta}$ of $x$ under the absolute loss function $\ell(x,\delta) = \|x-\delta\|_1$ is the median of distribution $Q(x|y)$*

$$\hat{\delta} = \textit{median}_Q(x|y). \tag{3}$$

*Proof.* This is straightforward by setting $c_1 = c_2$ in Proposition 24, where we get $\mathsf{P}_Q\left(x\leq\hat{\delta}|y\right) = \frac{c_2}{c_1+c_2} = 0.5$. □

*Remark* 27. The linear loss function in Proposition 24, allows the adjustment of the penalty between over-estimating and under-estimating $x$, by adjusting $c_1$and $c_2$. Hence, the absolute loss in Proposition 26, is more appropriate when over-estimation and under-estimation are of the same concern (as penalized the same).

*Remark* 28. Compared to the linear loss functions $\ell(x,\delta) = \|x-\delta\|_1$, the quadratic loss $\ell(x,\delta) = \|x-\delta\|_2^2$ aims at over-penalizing large but unlikely errors. The linear loss functions increase much slower than the quadratic loss, and hence, while remaining convex, they do not penalize so much the large but unlikely errors.

### Zero-one loss functions

**Proposition 29.** *The Bayes estimate $\hat{\delta}$ of $x$ under the zero-one loss function*

$$\ell(x,\delta) = 1 - 1_{B_\epsilon(\delta)}(x);\ \ \textit{where}\ \ B_\epsilon(\delta) = (x\in\mathcal{X}\mid \|x-\delta\|\leq\epsilon)$$

*is*

$$\hat{\delta} = \arg\max_{\forall\delta}\mathsf{P}_Q\left(x\in B_\epsilon(\delta)|y\right). \tag{4}$$

*Proof.* It is

$$\int \ell(x,\delta)\mathrm{d}Q(x|y) = 1 - \int 1_{B_\epsilon(\delta)}(x)\mathrm{d}Q(x|y) \quad = 1 - \mathsf{P}_Q\left(x\in B_\epsilon(\delta)|y\right)$$

[3]Integration by parts: $\int_a^b F(x)\mathrm{d}G(x) + \int_a^b G(x)\mathrm{d}F(x) = G(b)F(b) - G(a)F(a)$

which is minimized where $\mathsf{P}_Q\left(x \in B_\epsilon(\delta)|y\right)$ is maximized. □

**Proposition 30.** *The Bayes estimate $\hat{\delta}$ of $x$ with respect to the zero-one loss $\ell(x,\delta) = 1 - 1_\delta(x)$ is the mode of distribution Q, and it is called Maximum A posteriori (MAP) estimator, i.e.*

$$\hat{\delta} = mode_Q(x|y). \tag{5}$$

*Proof.* Straightforward from Proposition 29, by letting $\epsilon \to 0^+$. □

*Note* 31. Consider that parametric inference about $\theta$ is of interest. In the frequentist sense, MAP estimator can be seen as a penalized maximum likelihood estimator in the sense that it essentially maximizes

$$\log(\pi(\theta|y)) = \log f(y|\theta) + \log \pi(\theta) - [\text{normalising const.}] \ \propto \log f(y|\theta) + \log \pi(\theta) \tag{6}$$

which is the log-likelihood penalized by the log-prior (!!!)

**Example:** Consider the Bayesian model

$$\begin{cases} y|\mu & \sim \mathrm{N}(\mu, \sigma^2) \\ \mu & \sim \mathrm{N}(\mu_0, \sigma_0^2) \end{cases}$$

where $\sigma^2, \mu_0, \sigma_0^2$ are known. Then (6) becomes

$$\log(\pi(\mu|y)) \propto \log(\mathrm{N}(y|\mu,\sigma^2)) - \frac{1}{2\sigma_0^2}(\mu - \mu_0)^2$$

Here maximization of the posterior pdf is a compromise between the maximization of the log-likelihood and minimization of the distance $(\mu - \mu_0)^2$ regulated by $\sigma_0^2$. Hence the prior shrinks (or penalizes) $\mu$ towards $\mu_0$ and this shrinkage is adjusted via $\sigma_0^2$.

**Example:** Consider the Bayesian regression model

$$\begin{cases} y|\beta & \sim \mathrm{N}(\Phi\beta, I\sigma^2) \\ \beta & \sim \mathrm{N}\left([0,...,0]^\top, \mathrm{diag}(v_1,...,v_d)\right) \end{cases}$$

where $\sigma^2, v_1, ..., v_d$ are known. Then (6) becomes

$$\log(\pi(\beta|y)) \propto \log(\mathrm{N}(y|\Phi\beta, I\sigma^2)) - \frac{1}{2}\sum_{j=1}^{d}\frac{1}{v_j}(\beta_j - 0)^2$$

Here, the prior shrinks all $\beta_j$ towards 0 and this shrinkage is adjusted via $v_j$. Hence it can be speculated that the elements of the MAP estimate vector $\hat{\beta}$ at which the likelihood part would assign them small values (in absolute sense) will end up with much smaller values and closer to zero. This has applications to dimension reduction / variable selection / compressive sensing in specific high-dimensional regression problems.

*Remark* 32. Zero-one loss imposes a quite forceful penalization; because the penalty is equal to 0 if $\delta$ is the correct answer, and 1 if it is wrong.

**Example 33.** Consider a Bayesian model

$$\begin{cases} y_i|\theta & \overset{\text{iid}}{\sim} \mathrm{Br}(\theta), \qquad i = 1,...,n \\ \theta & \sim \mathrm{Be}(a,b) \end{cases}$$

where $a > 0$, $b > 0$, and $n > 2$. Find the Bayesian estimator of parameter $\theta$ for the zero-one loss function.

**Hint:** The posterior is $\theta|y \sim \mathrm{Be}(a + n\bar{y}, b + n - n\bar{y})$.

**Hint:** Consider PDF/PMF, for Bernoulli, and Beta distributions

$$f_{\text{Br}(\theta)}(y) = \theta^y(1-\theta)^{1-y}1(y \in \{0,1\})\,; \qquad \pi_{\text{Be}(a,b)}(\theta) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}1(\theta \in [0,1])$$

**Solution.** Given the absolute loss function, the point estimator is the Maximum Aposteriori Estimator (the posterior mode). Then

$$\log(\pi(\theta|y)) \propto (n\bar{y} + a - 1)\log(\theta) + (n - n\bar{y} + b - 1)\log(1-\theta)$$

So, for $a > 0$, $b > 0$

$$0 = \frac{\mathrm{d}}{\mathrm{d}\theta}\log(\pi(\theta|y))\Big|_{\theta=\hat{\delta}} = \frac{n\bar{y} + a - 1}{\theta} - \frac{n - n\bar{y} + b - 1}{1-\theta}\Big|_{\theta=\hat{\delta}} \implies \hat{\delta} = \frac{n\bar{y} + a - 1}{n + a + b - 2}.$$

It is good to notice (although not asked by the exercise) that

- If $a \to 1$, $b \to 1$ (aka $\pi(\theta) \propto 1$), then , like the Frequentists.
- If $a \to 0$, $b \to 0$ (aka $\pi(\theta) \propto \theta^{-1}(1-\theta)^{-1}$), then $\hat{\delta} = \frac{n\bar{y}-1}{n-2}$.
- If $a \to 1/2$, $b \to 1/2$ (aka $\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$), then $\hat{\delta} = \frac{n\bar{y}-1/2}{n-1}$.
- If $n \to \infty$, $a > 0$, $b > 0$, then $\hat{\delta} = \bar{y}$. Like the Frequentists.

**Question 34.** *Practice with Exercise 59 from the Exercise sheet.*