Solutions to Problem Sheet 8

1. **Sparse Bayesian estimation by *spike and slab* prior**

   We consider the so-called sparse sequence model

$$X_i = \theta_i + \varepsilon_i, \quad 1 \le i \le n$$

   and $\varepsilon_i \sim \mathcal{N}(0,1)$ independent, with $\theta = (\theta_1, \dots, \theta_n) \in \ell_0[s_n]$ unknown, where we denote

$$\ell_0[s_n] := \{\theta \in \mathbb{R}^n, \mathrm{Card}\{i : \theta_i \neq 0\} \le s_n\}$$

   the set of vectors in $\mathbb{R}^n$ that have at most s_n non-zero coordinates. In the following, s_n, which is unknown, will be assumed small compared to n in the sense that $s_n/n = o(1)$ as $n \to \infty$. We are interested in the estimation of θ for the loss function $\ell(\theta, \theta') = \|\theta - \theta'\|^2$, where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^n$.

   This is one of the simplest and most studied so-called high-dimensional models. Note that the unknown parameter θ is in principle of dimension n equal to the number of observations. With the hypothesis $\theta \in \ell_0[s_n]$, this dimension is actually 'reduced to s_n' (it is a bit more complex than that, because we do not know a priori the indices of the non-zero coefficients). In practice, s_n may itself be quite large and tend to infinity.

   We adopt a Bayesian approach for the estimation of θ in this model. Let Π be the prior distribution on $\mathbb{R}^n$ defined by, for $\alpha \in [0,1]$,

$$\Pi_\alpha = \bigotimes_{i=1}^{n} (1-\alpha)\delta_0 + \alpha Q,$$

   where Q is a probability distribution on $\mathbb{R}$ with density γ with respect to the Lebesgue measure on $\mathbb{R}$.

   (a) How does the prior distribution Π_α reflect the nature of the considered problem?

   Solution: The problem assumes that the vector θ is sparse, meaning that many of its components are exactly zero (at most s_n are non-zero). The prior Π_α reflects this by using a mixture for each independent component θ_i: it places a discrete point mass at zero (the "spike" $(1-\alpha)\delta_0$), which forces the parameter to be strictly zero with probability $1-\alpha$, and an absolutely continuous distribution (the "slab" αQ) to model the non-zero coefficients with probability α.

   (b) What is the prior distribution induced on the number of non-zero coefficients of θ?

Solution: Since the components θ_i are independent under the prior and each has a probability α of being non-zero (drawn from Q), the number of non-zero coefficients $K = \sum_{i=1}^n \mathbf{1}_{\{\theta_i \neq 0\}}$ follows a Binomial distribution $\mathcal{B}(n, \alpha)$.

(c) If s_n were known, what choice of α would seem reasonable to you?

Solution: The expected number of non-zero coefficients under the prior is $\mathbb{E}[K] = n\alpha$. To match the prior expectation with the known sparsity level s_n, a reasonable choice would be $\alpha = \frac{s_n}{n}$. This ensures that, on average, the prior generates vectors with s_n active components.

(d) Determine the posterior distribution $\Pi_\alpha[\cdot \mid X]$ using Bayes' formula and the function $g = \phi * \gamma$, the convolution of the $\mathcal{N}(0,1)$ density and the density γ. One may note that $\Pi_\alpha[\cdot \mid X]$ is a product distribution.

Solution: Since the prior Π_α factorizes and the likelihood of the independent observations also factorizes, the posterior distribution is a product distribution:

$$\Pi_\alpha(d\theta \mid X) = \bigotimes_{i=1}^n \Pi_\alpha(d\theta_i \mid X_i).$$

For a single coordinate i, the marginal density of X_i is obtained by integrating the likelihood over the prior:

$$\begin{aligned} m_\alpha(X_i) &= \int_{\mathbb{R}} \phi(X_i - \theta)\Pi_\alpha(d\theta) \\ &= (1-\alpha)\phi(X_i) + \alpha \int_{\mathbb{R}} \phi(X_i - \theta)\gamma(\theta)d\theta \\ &= (1-\alpha)\phi(X_i) + \alpha g(X_i), \end{aligned}$$

where $g = \phi * \gamma$.
Using Bayes' formula, the posterior distribution for θ_i is:

$$\begin{aligned} \Pi_\alpha(d\theta_i \mid X_i) &= \frac{\phi(X_i - \theta_i)\Pi_\alpha(d\theta_i)}{m_\alpha(X_i)} \\ &= \frac{(1-\alpha)\phi(X_i)\delta_0(d\theta_i) + \alpha\phi(X_i - \theta_i)\gamma(\theta_i)d\theta_i}{(1-\alpha)\phi(X_i) + \alpha g(X_i)}. \end{aligned}$$

This shows that the posterior for each coordinate remains a mixture of a Dirac mass at 0 and an updated continuous distribution.

(e) We propose to determine α by the marginal maximum likelihood method. Show that the corresponding estimator $\hat{\alpha}$ is

$$\hat{\alpha} = \operatorname*{argmax}_{\alpha \in [0,1]} \prod_{i=1}^n \big((1-\alpha)\phi(X_i) + \alpha g(X_i)\big).$$

Solution: The empirical Bayes approach consists of estimating hyperparameters by maximizing the marginal likelihood of the data. The joint marginal distribution of the observations $X = (X_1, \dots, X_n)$ given α is the product of their individual marginal distributions because they are independent:

$$L(\alpha \mid X) = \prod_{i=1}^{n} m_\alpha(X_i).$$

Using the expression for $m_\alpha(X_i)$ derived in the previous question, we have:

$$L(\alpha \mid X) = \prod_{i=1}^{n} \big((1-\alpha)\phi(X_i) + \alpha g(X_i)\big).$$

The marginal maximum likelihood estimator $\hat{\alpha}$ is obtained by maximizing this likelihood over the valid parameter space $[0, 1]$, which directly gives the stated formula.

Johnstone and Silverman (2004, *Annals of Statistics*) demonstrated that for a slightly modified version of $\hat{\alpha}$ (restricting slightly the set of α in the definition of $\hat{\alpha}$), the vector of medians $\hat{\theta}_i^{med}(X)$ of the posterior distribution $\Pi_{\hat{\alpha}}[\cdot \mid X]$ on each coordinate achieves the following rate, for s_n any sequence such that $s_n/n = o(1)$ and $(\log n)^2/s_n = o(1)$,

$$\sup_{\theta \in \ell_0[s_n]} E_\theta \|\hat{\theta}^{med} - \theta\|^2 \leq C s_n \log(n/s_n),$$

for $n \geq N_0$, with N_0 and C being universal constants.

2. **Convergence**

We consider the *fundamental model* $\mathcal{P} = \{P_\theta^{(n)}, \theta \in \mathbb{R}\}$, for which

$$P_\theta^{(n)} = P_\theta^{\otimes n}, \quad P_\theta \sim \mathcal{N}(\theta, 1).$$

We have n observations $X = (X_1, \dots, X_n)$ and we place ourselves in the Bayesian framework

$$\begin{aligned} X \mid \theta &\sim P_\theta^{(n)} \\ \theta &\sim \Pi. \end{aligned}$$

We form the posterior distribution $\Pi[\cdot \mid X]$, the distribution of $\theta \mid X$. We study this distribution from the frequentist point of view: we assume there exists a 'true' value $\theta_0 \in \mathbb{R}$ of the parameter θ and we study $\Pi[\cdot \mid X]$ in probability under $X \sim P_{\theta_0}^{(n)}$. We denote by E_{θ_0} the expectation under this distribution.

PART A. In this part, we choose $\Pi = \mathcal{N}(a, 1)$, where a is a fixed real number.

(a) Show that, for $\overline{X}$ the empirical mean of the X_i,

$$\Pi[\cdot \mid X] = \mathcal{N}\left(\overline{\theta}, \frac{1}{n+1}\right), \quad \overline{\theta} = \frac{n\overline{X}+a}{n+1}.$$

Solution: This posterior distribution was obtained in class and several times in exercises.

(b) From the explicit expression above for the posterior distribution, demonstrate directly that, as $n \to \infty$, for all $M_n \to \infty$,

$$E_{\theta_0}\Pi\left[\left\{\theta : |\theta - \theta_0| \leq \frac{M_n}{\sqrt{n}}\right\} \,\middle|\, X\right] \to 1.$$

Solution: Let $A_n = \left\{\theta : |\theta - \theta_0| \leq \frac{M_n}{\sqrt{n}}\right\}$. We want to show that, as $n \to \infty$, for all $M_n \to \infty$,

$$E_{\theta_0}\Pi[A_n \mid X] \to 1.$$

Let $\mathcal{E}_n$ be the event defined below, which corresponds to requiring that $\overline{\theta}_X$ is sufficiently close to θ_0:

$$\mathcal{E}_n = \left\{|\overline{\theta}_X - \theta_0| \leq \frac{M_n}{2\sqrt{n}}\right\}.$$

Let's first note that $E_{\theta_0}[\mathbf{1}_{\mathcal{E}_n}] = P_{\theta_0}[\mathcal{E}_n] \to 1$ as $n \to \infty$. To see this,

$$\begin{aligned}\sqrt{n}(\overline{\theta}_X - \theta_0) &= \sqrt{n}\left(\frac{n\overline{X}+a}{n+1} - \theta_0\right) \\ &= \frac{n}{n+1}\sqrt{n}(\overline{X} - \theta_0) + \frac{\sqrt{n}}{n+1}(a - \theta_0).\end{aligned}$$

According to Slutsky's lemma, this last quantity converges in distribution to a $\mathcal{N}(0,1)$, since $\sqrt{n}(\overline{X} - \theta_0)$ converges in distribution to a $\mathcal{N}(0,1)$ by the Central Limit Theorem. We deduce

$$P[\mathcal{E}_n^c] = P\left[M_n^{-1}\sqrt{n}(\overline{X} - \theta_0) > \frac{1}{2} + o(1)\right] = o(1),$$

because $M_n^{-1}\sqrt{n}(\overline{X} - \theta_0)$ converges to 0 in probability as $M_n \to \infty$ (Slutsky's lemma). Let us now write

$$\Pi[A_n \mid X] = \Pi[A_n \mid X]\mathbf{1}_{\mathcal{E}_n} + \Pi[A_n \mid X]\mathbf{1}_{\mathcal{E}_n^c} = (i) + (ii).$$

From the above $E_{\theta_0}[(ii)] \leq E_{\theta_0}[\mathbf{1}_{\mathcal{E}_n^c}] = o(1)$ (by bounding $\Pi[A_n \mid X]$, which is a probability, by 1). For term (i), we write

$$E_{\theta_0}[(i)] \geq E_{\theta_0}\left[\Pi\left[|\theta - \overline{\theta}_X| \leq \frac{M_n}{2\sqrt{n}} \,\middle|\, X\right]\mathbf{1}_{\mathcal{E}_n}\right].$$

Indeed, on the event $\mathcal{E}_n$, if $|\theta - \overline{\theta}_X| \leq \frac{M_n}{2\sqrt{n}}$ then

$$|\theta - \theta_0| \leq |\theta - \overline{\theta}_X| + |\overline{\theta}_X - \theta_0| \leq \frac{M_n}{\sqrt{n}}.$$

But by definition of the posterior distribution,

$$\begin{aligned}\Pi\left[|\theta - \overline{\theta}_X| \leq \frac{M_n}{2\sqrt{n}} \;\middle|\; X\right] &= P\left[\left|\mathcal{N}\left(0, \frac{1}{n+1}\right)\right| \leq \frac{M_n}{2\sqrt{n}}\right] \\ &= P\left[\frac{1}{M_n}\sqrt{\frac{n+1}{n}}|\mathcal{N}(0,1)| \leq \frac{1}{2}\right].\end{aligned}$$

Now $\frac{1}{M_n}\sqrt{\frac{n+1}{n}}|\mathcal{N}(0,1)|$ converges in probability to 0 (Slutsky's lemma), so the previous quantity tends to 1 as $n \to \infty$. We conclude that $E_{\theta_0}[(i)] \to 1$ as $n \to \infty$, which was to be demonstrated.

(c) Similarly, still starting from the explicit expression, show that for all $m_n \to 0$,

$$E_{\theta_0}\Pi\left[\left\{\theta : |\theta - \theta_0| \leq \frac{m_n}{\sqrt{n}}\right\} \;\middle|\; X\right] \to 0.$$

Solution: Let $m_n \to 0$ as $n \to \infty$. We write

$$\Pi\left[\left\{\theta : |\theta - \theta_0| \leq \frac{m_n}{\sqrt{n}}\right\} \;\middle|\; X\right] = \int_{\theta_0 - \frac{m_n}{\sqrt{n}}}^{\theta_0 + \frac{m_n}{\sqrt{n}}} f_{\theta|X}(\theta)d\theta,$$

where $f_{\theta|X}$ is the posterior density. The posterior density is Gaussian centered at $\overline{\theta}_X$ according to 1, thus it is unimodal and symmetric around $\overline{\theta}_X$, which implies that the mass it places on any interval of length L is always less than or equal to the mass it places on the interval $[\overline{\theta}_X - L/2, \overline{\theta}_X + L/2]$. We deduce

$$\begin{aligned}\int_{\theta_0 - \frac{m_n}{\sqrt{n}}}^{\theta_0 + \frac{m_n}{\sqrt{n}}} f_{\theta|X}(\theta)d\theta \leq \int_{\overline{\theta}_X - \frac{m_n}{\sqrt{n}}}^{\overline{\theta}_X + \frac{m_n}{\sqrt{n}}} f_{\theta|X}(\theta)d\theta &= P\left[\left|\mathcal{N}\left(0, \frac{1}{n+1}\right)\right| \leq \frac{m_n}{\sqrt{n}}\right] \\ &\leq P\left[\left|\sqrt{\frac{n}{n+1}}\mathcal{N}(0,1)\right| \leq m_n\right] \\ &\leq \int_{-2m_n}^{2m_n} \frac{e^{-u^2/2}}{\sqrt{2\pi}}du \leq \frac{4m_n}{\sqrt{2\pi}} = o(1).\end{aligned}$$

We conclude

$$E_{\theta_0}\Pi\left[\left\{\theta : |\theta - \theta_0| \leq \frac{m_n}{\sqrt{n}}\right\} \;\middle|\; X\right] \to 0.$$

(d) Interpret the results of (b) and (c) from the point of view of convergence rates. If the model and the prior are no longer Gaussian, can we recover the previous results?

Solution: These results are interpreted as follows: $M_n/\sqrt{n}$ is an upper bound for the convergence rate of the posterior distribution. The result of 3. shows that this rate cannot be much improved, since $m_n/\sqrt{n}$, with $m_n = o(1)$, is too fast for the ball of this radius to contain posterior mass asymptotically.

PART B. In this part, we choose $\Pi = \mathrm{Unif}[0,1]$. Besides, we call a *truncated normal distribution* on the interval $J = [a,b]$ a distribution with density on $\mathbb{R}$ proportional to

$$\varphi_{\mu,\sigma^2}(x)\mathbf{1}_J(x),$$

where φ_{μ,σ^2} is the density of a $\mathcal{N}(\mu,\sigma^2)$.

(a) Show that the posterior distribution $\Pi[\cdot \mid X]$ is a truncated Gaussian distribution.

Solution: Bayes' formula gives

$$\begin{aligned} f_{\theta|X}(\theta) &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2\right\}\mathbf{1}_{[0,1]}(\theta) \\ &\propto \exp\left\{-\frac{n}{2}(\theta-\overline{X})^2\right\}\mathbf{1}_{[0,1]}(\theta). \end{aligned}$$

It is therefore a Gaussian distribution $\mathcal{N}(\overline{X}, n^{-1})$ truncated to the interval $[0,1]$.

(b) We now study the behavior of $\Pi[\cdot \mid X]$ under $P_{\theta_0}^{(n)}$.

i. In this question $\theta_0 \in (0,1)$. Is the posterior distribution consistent?

Solution: Let $\theta_0 \in (0,1)$ and let $\varepsilon > 0$. Let us show that the posterior distribution is consistent at θ_0, that is to say when $n \to \infty$,

$$E_{\theta_0}\Pi[|\theta-\theta_0| \le \varepsilon \mid X] \to 1. \tag{1}$$

We proceed as in part A: we introduce the event

$$\mathcal{E}_n = \left\{|\overline{X}-\theta_0| \le \frac{\varepsilon}{2}\right\}.$$

The law of large numbers gives that $P_{\theta_0}[\mathcal{E}_n^c]$ tends to 0 as $n \to \infty$. We have

$$
\begin{aligned}
\Pi[|\theta-\theta_0| \le \varepsilon \mid X]\mathbf{1}_{\mathcal{E}_n} &= \mathbf{1}_{\mathcal{E}_n} \int_{\theta_0-\varepsilon}^{\theta_0+\varepsilon} f_{\theta|X}(\theta)d\theta \\
&= \mathbf{1}_{\mathcal{E}_n} \frac{\int_{\theta_0-\varepsilon}^{\theta_0+\varepsilon} \exp\left\{-\frac{n}{2}(\theta-\overline{X})^2\right\}\mathbf{1}_{[0,1]}(\theta)d\theta}{\int \exp\left\{-\frac{n}{2}(\theta-\overline{X})^2\right\}\mathbf{1}_{[0,1]}(\theta)d\theta} \\
&= \mathbf{1}_{\mathcal{E}_n} \frac{\int_{\sqrt{n}(\theta_0-\varepsilon-\overline{X})}^{\sqrt{n}(\theta_0+\varepsilon-\overline{X})} e^{-u^2/2}du}{\int_{-\sqrt{n}\overline{X}}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2}du} \\
&\ge \mathbf{1}_{\mathcal{E}_n} \frac{\int_{-\sqrt{n}\varepsilon/2}^{\sqrt{n}\varepsilon/2} e^{-u^2/2}du}{\int_{-\sqrt{n}\overline{X}}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2}du} \\
&\ge \mathbf{1}_{\mathcal{E}_n} \frac{\int_{-\sqrt{n}\varepsilon/2}^{\sqrt{n}\varepsilon/2} e^{-u^2/2}du}{\int_{-\infty}^{\infty} e^{-u^2/2}du}
\end{aligned}
$$

where we used the definition of $\mathcal{E}_n$ to include the interval in the numerator in the third line into the interval of the fourth line, and where we bounded the denominator in the last line. The quotient of the last expression tends to 1 as $n \to \infty$. We conclude

$$E_{\theta_0}\Pi[|\theta-\theta_0| \le \varepsilon \mid X] = P(\mathcal{E}_n)(1+o(1)) = 1+o(1),$$

which proves (1).

ii. Assume $|\theta_0 - 1/2| > 1/2$. Is the posterior distribution consistent?

Solution: If $\theta_0 \notin [0,1]$, then since $[0,1]^c$ is open, we can find an open interval V containing θ_0 such that $V \subset [0,1]^c$. But $\Pi(V) = 0$ because Π puts mass on $[0,1]$ only by definition. We deduce that $\Pi[V \mid X] = 0$ (this results immediately from Bayes' formula, see remark made in class). So Π cannot be consistent at θ_0.

iii. If $\theta_0 \ge 1$, show that for all $\varepsilon > 0$, as $n \to \infty$,

$$E_{\theta_0}\Pi[A_\varepsilon \mid X] \to 1, \quad \text{where } A_\varepsilon := \{\theta, 1-\varepsilon \le \theta \le 1\}.$$

Solution: We proceed as in question a), with $(\theta_0-\varepsilon, \theta_0+\varepsilon)$ replaced by $(1-\varepsilon, 1]$, and we give ourselves a little more margin by setting

$$\mathcal{E}_n = \left\{|\overline{X}-1| \le \frac{\varepsilon}{4}\right\}.$$

$$
\begin{aligned}
\Pi[A_\varepsilon \mid X]\mathbf{1}_{\mathcal{E}_n} &= \mathbf{1}_{\mathcal{E}_n} \int_{1-\varepsilon}^{1} f_{\theta|X}(\theta) d\theta \\
&= \mathbf{1}_{\mathcal{E}_n} \frac{\int_{1-\varepsilon}^{1} \exp\left\{-\frac{n}{2}(\theta - \overline{X})^2\right\} \mathbf{1}_{[0,1]}(\theta) d\theta}{\int \exp\left\{-\frac{n}{2}(\theta - \overline{X})^2\right\} \mathbf{1}_{[0,1]}(\theta) d\theta} \\
&= \mathbf{1}_{\mathcal{E}_n} \frac{\int_{\sqrt{n}(1-\varepsilon-\overline{X})}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2} du}{\int_{-\sqrt{n}\overline{X}}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2} du}.
\end{aligned}
$$

We then note that the denominator is written

$$
\int_{-\sqrt{n}\overline{X}}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2} du = \int_{-\sqrt{n}\overline{X}}^{\sqrt{n}(1-\varepsilon-\overline{X})} e^{-u^2/2} du + \int_{\sqrt{n}(1-\varepsilon-\overline{X})}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2} du.
$$

Denoting ϕ the density of a $\mathcal{N}(0,1)$, we have, on $\mathcal{E}_n$,

$$
\begin{aligned}
\frac{1}{\sqrt{2\pi}} \int_{-\sqrt{n}\overline{X}}^{\sqrt{n}(1-\varepsilon-\overline{X})} e^{-u^2/2} du &\leq \sqrt{n}(1-\varepsilon)\phi(\sqrt{n}(1-\varepsilon-\overline{X})) \\
&\leq \sqrt{n}(1-\varepsilon)\phi\left(-\sqrt{n}\frac{3\varepsilon}{4}\right)
\end{aligned}
$$

while, similarly, on $\mathcal{E}_n$,

$$
\begin{aligned}
\frac{1}{\sqrt{2\pi}} \int_{\sqrt{n}(1-\varepsilon-\overline{X})}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2} du &\geq \frac{1}{\sqrt{2\pi}} \int_{\sqrt{n}(1-\frac{\varepsilon}{2}-\overline{X})}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2} du \\
&\geq \sqrt{n}\frac{\varepsilon}{4}\phi\left(-\sqrt{n}\frac{\varepsilon}{2}\right).
\end{aligned}
$$

We conclude that on $\mathcal{E}_n$,

$$
\frac{\int_{\sqrt{n}(1-\varepsilon-\overline{X})}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2} du}{\int_{-\sqrt{n}\overline{X}}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2} du} = \frac{1}{1 + \frac{\int_{-\sqrt{n}\overline{X}}^{\sqrt{n}(1-\varepsilon-\overline{X})} e^{-u^2/2} du}{\int_{\sqrt{n}(1-\varepsilon-\overline{X})}^{\sqrt{n}(1-\overline{X})} e^{-u^2/2} du}} \geq \frac{1}{1 + \frac{(1-\varepsilon)\phi\left(-\sqrt{n}\frac{3\varepsilon}{4}\right)}{\frac{\varepsilon}{4}\phi\left(-\sqrt{n}\frac{\varepsilon}{2}\right)}}.
$$

This last quantity tends to 1 when $n \to \infty$. We deduce

$$
E_{\theta_0}\Pi[A_\varepsilon \mid X]\mathbf{1}_{\mathcal{E}_n} \geq (1 + o(1))P_{\theta_0}(\mathcal{E}_n) = 1 + o(1),
$$

and therefore $E_{\theta_0}\Pi[A_\varepsilon \mid X] \to 1$, which was to be demonstrated.

iv. If $\theta_0 = 0$ or $\theta_0 = 1$, is the posterior distribution consistent?

Solution: If $\theta_0 = 0$ or $\theta_0 = 1$, it follows from question c) that the posterior distribution is consistent at these points. Indeed, every open neighborhood V of θ_0 contains a set A_ε for ε small enough and therefore $E_{\theta_0}\Pi[V \mid X] \to 1$, which shows consistency.

3. **Credible intervals, confidence interval**

Let $X = (X_1, \ldots, X_n)$ with X_i i.i.d. following a Bernoulli distribution $\mathrm{Be}(\theta)$. We place a prior $\mathrm{Beta}(a, b)$ on θ, with $a > 0$ and $b > 0$. We are given

$$\mathbb{E}[\mathrm{Beta}(a, b)] = \frac{a}{a+b}, \quad \mathrm{Var}[\mathrm{Beta}(a, b)] = \frac{ab}{(a+b)^2(a+b+1)}.$$

(a) Give the posterior distribution $\Pi[\cdot \mid X]$. Let m_X denote its mean and v_X its variance.

Solution: Bayes' formula gives

$$f_{\theta|X}(\theta) \propto \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i}\theta^{a-1}(1-\theta)^{b-1} \propto \theta^{n\overline{X}+a-1}(1-\theta)^{n(1-\overline{X})+b-1}.$$

Thus the posterior distribution $\Pi[\cdot \mid X]$ is a $\mathrm{Beta}(n\overline{X}+a, n(1-\overline{X})+b)$ distribution, with mean and variance

$$m_X = \frac{n\overline{X}+a}{n+a+b}, \quad v_X = \frac{(n\overline{X}+a)(n(1-\overline{X})+b)}{(n+a+b)^2(n+a+b+1)}.$$

(b) Construct a credible interval $I^T(X)$ of level at least $1-\alpha$, ($\alpha > 0$), centered at m_X, using Chebyshev's inequality.

Solution: Chebyshev's inequality gives, for any $\delta > 0$,

$$\Pi[|\theta - m_X| > \delta \mid X] \leq \delta^{-2}\,\mathbb{E}[(\theta - m_X)^2 \mid X] = \frac{v_X}{\delta^2}.$$

We choose α such that $\alpha = v_X/\delta^2$. We deduce that

$$I^T(X) = \left[m_X \pm \sqrt{\frac{v_X}{\alpha}}\right]$$

is a credible interval of level at least $1-\alpha$.

(c) Show that $I^T(X)$ is an asymptotic confidence interval under P_{θ_0}, for which we will lower bound the level as a function of α.

Solution: We start by noting, thanks to Slutsky's lemma, that $D_n := \sqrt{n}(m_X - \overline{X}) = o_P(1)$ while nv_X converges in probability to $\theta_0(1-\theta_0) = I_0^{-1}$ (denoting I_0 the Fisher information at point θ_0) which can be written $v_X = \frac{1}{nI_0}(1 + o_P(1))$. We deduce

$$\begin{aligned} P_{\theta_0}[\theta_0 \in I_T(X)] &= P_{\theta_0}\left[|\theta_0 - m_X| \leq \sqrt{\frac{v_X}{\alpha}}\right] \\ &\geq P_{\theta_0}\left[|\theta_0 - \overline{X}| - |\overline{X} - m_X| \leq \sqrt{\frac{v_X}{\alpha}}\right] \\ &\geq P_{\theta_0}\left[\sqrt{n}|\theta_0 - \overline{X}| - \sqrt{n}|\overline{X} - m_X| \leq \sqrt{\frac{nv_X}{\alpha}}\right] \end{aligned}$$

Slutsky's lemma gives, using that $\sqrt{n}(\overline{X} - \theta_0) \to \mathcal{N}(0, I_0^{-1})$ by the Central Limit Theorem, that

$$\frac{\sqrt{n}(\theta_0 - \overline{X}) - \sqrt{n}(\overline{X} - m_X)}{\sqrt{nv_X}} \xrightarrow{\mathcal{L}} \sqrt{I_0}\mathcal{N}(0, I_0^{-1}) - 0 \overset{\mathcal{L}}{=} \mathcal{N}(0,1).$$

We conclude that, for $\Phi(u) = P[\mathcal{N}(0,1) \leq u]$, with $P[|\mathcal{N}(0,1)| \leq t] = 2\Phi(t) - 1$,

$$\liminf_{n\to\infty} P_{\theta_0}[\theta_0 \in I_T(X)] \geq P\left[|\mathcal{N}(0,1)| \leq \frac{1}{\sqrt{\alpha}}\right] = 2\Phi\left(\frac{1}{\sqrt{\alpha}}\right) - 1.$$

and we have indeed obtained a lower bound of the level of $I_T(X)$ as a function of α.

(d) Using the BvM theorem, show that the posterior distribution converges in total variation (towards what distribution?).

Solution: According to the BvM theorem, the posterior distribution converges in total variation, under P_{θ_0}, as follows

$$\left\| \Pi[\cdot \mid X] - \mathcal{N}\left(\hat{\theta}^{MV}(X), \frac{1}{nI(\theta_0)}\right)(\cdot) \right\|_1 \xrightarrow{P_{\theta_0}} 0.$$

Here, it is easily verified that $\hat{\theta}^{MV}(X) = \overline{X}$ and $I(\theta_0)^{-1} = \theta_0(1-\theta_0)$.

(e) Let $I^B(X) = [a_n(X), b_n(X)]$ be the interval defined by the quantiles $\alpha/2$ and $1-\alpha/2$ of the posterior distribution. Give the asymptotic expression of $a_n(X)$ and $b_n(X)$ under P_{θ_0}.

Solution: We have seen in class that the interval defined by the quantiles, if BvM is satisfied, is to an $o_P(1)$ the optimal interval in the frequentist sense, which would be obtained by taking the quantiles of the $\mathcal{N}(\hat{\theta}^{MV}(X), \frac{1}{nI(\theta_0)})$ distribution, let $z_\alpha = q_{1-\alpha/2}^{\mathcal{N}(0,1)}$

$$I^B(X) = [a_n(X), b_n(X)] = \left[\hat{\theta}^{MV} - \frac{z_\alpha}{\sqrt{nI(\theta_0)}}(1 + o_P(1)), \hat{\theta}^{MV} + \frac{z_\alpha}{\sqrt{nI(\theta_0)}}(1 + o_P(1))\right],$$

with $\hat{\theta}^{MV}$ and $I(\theta_0)$ as above.

(f) Compare I^T and I^B. What other type of inequality could have been used in question (b)?

Solution: As seen in class, we have that

$$P_{\theta_0}[\theta_0 \in I^B(X)] \to 1-\alpha.$$

We can compare I^T and I^B from three points of view

- their size: the diameter of I^B is of the order of

  $$\mathrm{Diam}(I^B) = C(1+o_P(1))z_\alpha/\sqrt{n},$$

  while I^T has diameter

  $$\mathrm{Diam}(I^T) = 2\sqrt{v_X/\alpha} = C(1+o_P(1))/\sqrt{n\alpha}.$$

  They have the same dependence on n. However, we can show that (admitted) z_α is of the order of $\sqrt{2\log(1/\alpha)}$ for $\alpha \to 0$, and thus this dependence is much better than $1/\sqrt{\alpha}$.

- their credible level: $1-\alpha$ exactly for I^B and at least $1-\alpha$ for I^T.

- their asymptotic confidence level: it is $1-\alpha$ for I^B and at least $1-2(1-\Phi(1/\sqrt{\alpha}))$ for I^T. We can verify that, if ϕ is the standard normal Gaussian density, when $x \to +\infty$,

  $$1-\Phi(x) \sim \phi(x)/x.$$

  We deduce, when $\alpha \to 0$,

  $$1-\Phi(1/\sqrt{\alpha}) \sim \sqrt{\alpha}\phi(1/\sqrt{\alpha}) \asymp \sqrt{\alpha}e^{-\frac{1}{2\alpha}}.$$

  This quantity tends to 1 much faster than $1-\alpha$. This is logical, because the interval I^T is significantly larger than I^B for small α.

One could also use a Markov-type inequality in question 2, with another power than 2, for example 4, but the calculations would then have been more complicated, it would then be necessary to use the formula for moments up to order 4 of Beta distributions.