Problem Sheet 8

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

   We consider the so-called sparse sequence model

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

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

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

   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?

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

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

   (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.

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

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

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 $X = (X_1, \ldots, X_n)$ 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 $\bar{X}$ the empirical mean of the X_i,

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

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

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

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

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

(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?

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

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

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.

(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?

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

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

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

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

3. **Credible intervals, confidence interval**

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

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

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

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

(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.

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

(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}.

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