everyday processing of random models can be directly exploited in the implementation of simulation techniques (in particular in the evaluation of the variation of the proposed estimators or of the stationarity of the resulting output), while purely numerical techniques rely on less familiar branches of mathematics. Finally, note that many desirable approaches are those which efficiently combine both perspectives, as in the case of *simulated annealing* (see Section 5.2.3) or Riemann sums (see Section 4.3).

1.5 Problems

1.1 For both the censored data density (1.1) and the mixture of two normal distributions (1.15), plot the probability density function. Use various values for the parameters μ, θ, σ and τ.

1.2 In the situation of Example 1.1, establish that the densities are indeed (1.1) and (1.2).

1.3 In Example 1.1, the distribution of the random variable $Z = \min(X, Y)$ was of interest. Derive the distribution of Z in the following case of *informative censoring*, where $Y \sim \mathcal{N}(\theta, \sigma^2)$ and $X \sim \mathcal{N}(\theta, \theta^2\sigma^2)$. Pay attention to the identifiability issues.

1.4 In Example 1.1, show that the integral

$$\int_\omega^\infty \alpha\beta x^{\alpha-1} e^{-\beta x^\alpha} dx$$

can be explicitly calculated. (*Hint:* Use a change of variables.)

1.5 For the model (1.4), show that the density of $(X_0, \ldots, X_n)$ is given by (1.5).

1.6 In the setup of Example 1.2, derive the moment estimator of the weights $(p_1, \ldots, p_k)$ when the densities f_j are known.

1.7 In the setup of Example 1.6, show that the likelihood equations are given by (1.11) and that their solution is the standard $(\bar{y}, s^2)$ statistic.

1.8 (Titterington et al. 1985) In the case of a mixture of two exponential distributions with parameters 1 and 2,

$$\pi\mathcal{E}xp(1) + (1-\pi)\mathcal{E}xp(2) ,$$

show that $\mathbb{E}[X^s] = \{\pi + (1-\pi)2^{-s}\}\Gamma(s+1)$. Deduce the best (in s) moment estimator based on $t_s(x) = x^s/\Gamma(s+1)$.

1.9 Give the moment estimator for a mixture of k Poisson distributions, based on $t_s(x) = x(x-1)\cdots(x-s+1)$. (*Note:* Pearson 1915 and Gumbel 1940 proposed partial solutions in this setup. See Titterington et al. 1985, pp. 80–81, for details.)

1.10 In the setting of Example 1.9, plot the likelihood based on observing $(x_1, x_2, x_3) = (0, 5, 9)$ from the Student's t density (1.14) with $p = 1$ and $\sigma = 1$ (which is the standard Cauchy density). Observe the effect on multimodality of adding a fourth observation x_4 when x_4 varies.

1.11 The *Weibull distribution* $\mathcal{W}e(\alpha, c)$ is widely used in engineering and reliability. Its density is given by

$$f(x|\alpha, c) = c\alpha^{-1}(x/\alpha)^{c-1}e^{-(x/\alpha)^c} .$$

(a) Show that when c is known, this model is equivalent to a Gamma model.
(b) Give the likelihood equations in α and c and show that they do not allow for explicit solutions.
(c) Consider an iid sample $X_1, \ldots, X_n$ from $\mathcal{W}e(\alpha, c)$ censored from the right in y_0. Give the corresponding likelihood function when α and c are unknown and show that there is no explicit maximum likelihood estimator in this case either.

1.12 (Continuation of Problem 1.11) Show that the cdf of the Weibull distribution $\mathcal{W}e(\alpha, \beta)$ can be written explicitly, and show that the scale parameter α determines the behavior of the hazard rate $h(t) = \frac{f(t)}{1-F(t)}$, where f and F are the density and the cdf, respectively.

1.13 (Continuation of Problem 1.11) The following sample gives the times (in days) at which carcinoma was diagnosed in rats exposed to a carcinogen:

$$143,\ 164,\ 188,\ 188,\ 190,\ 192,\ 206,\ 209,\ 213,\ 216,\ 220,$$
$$227,\ 230,\ 234,\ 246,\ 265,\ 304,\ 216^*,\ 244^*,$$

where the observations with an asterisk are censored (see Pike 1966, for details). Fit a three parameter Weibull $\mathcal{W}e(\alpha, \beta, \gamma)$ distribution to this dataset, where γ is a translation parameter, for (a) $\gamma = 100$ and $\alpha = 3$, (b) $\gamma = 100$ and α unknown and (c) γ and α unknown. (*Note:* Treat the asterisked observations as ordinary data here. See Problem 5.24 for a method of dealing with the censoring.)

1.14 Let $X_1, X_2, \ldots, X_n$ be iid with density $f(x|\theta, \sigma)$, the Cauchy distribution $\mathcal{C}(\theta, \sigma)$, and let $L(\theta, \sigma|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta, \sigma)$ be the likelihood function.
(a) If σ is known, show that a solution to the likelihood equation $\frac{d}{d\theta} L(\theta, \sigma|\mathbf{x}) = 0$ is the root of a $2n - 1$ degree polynomial. Hence, finding *the* likelihood estimator can be challenging.
(b) For $n = 3$, if both θ and σ are unknown, find the maximum likelihood estimates and show that they are unique.
(c) For $n \geq 3$, if both θ and σ are unknown, show that the likelihood is unimodal.
(*Note:* See Copas 1975 and Ferguson 1978 for details.)

1.15 Referring to Example 1.16, show that the posterior distribution (1.22) can be written in the form (1.23).

1.16 Consider a Bernoulli random variable $Y \sim \mathcal{B}([1 + e^\theta]^{-1})$.
(a) If $y = 0$, show that the maximum likelihood estimator of θ is ∞.
(b) Show that the same problem occurs when $Y_1, Y_2 \sim \mathcal{B}([1 + e^\theta]^{-1})$ and $y_1 = y_2 = 0$ or $y_1 = y_2 = 1$. Give the maximum likelihood estimator in the other cases.

1.17 Consider n observations $x_1, \ldots, x_n$ from $\mathcal{B}(k, p)$ where both k and p are unknown.
(a) Show that the maximum likelihood estimator of k, $\hat{k}$, satisfies

$$(\hat{k}(1 - \hat{p}))^n \geq \prod_{i=1}^n (\hat{k} - x_i) \quad \text{and} \quad ((\hat{k} + 1)(1 - \hat{p}))^n < \prod_{i=1}^n (\hat{k} + 1 - x_i),$$

where $\hat{p}$ is the maximum likelihood estimator of p.
(b) If the sample is $16, 18, 22, 25, 27$, show that $\hat{k} = 99$.
(c) If the sample is $16, 18, 22, 25, 28$, show that $\hat{k} = 190$. Discuss the stability of the maximum likelihood estimator.

(*Note:* Olkin et al. 1981 were one of the first to investigate the stability of the MLE for the binomial parameter n; see also Carroll and Lombard 1985, Casella 1986, and Hall 1994.)

1.18 (Robertson et al. 1988) For a sample $X_1, \ldots, X_n$, and a function f on $\mathcal{X}$, the isotonic regression of f with weights ω_i is the solution of the minimization in g of

$$\sum_{i=1}^{n} \omega_i (g(x_i) - f(x_i))^2,$$

under the constraint $g(x_1) \leq \cdots \leq g(x_n)$.

(a) Show that a solution to this problem is obtained by the *pool-adjacent-violators* algorithm:

Algorithm A.1 –Pool-adjacent-violators–

If f is not isotonic, find i such that $f(x_{i-1}) > f(x_i)$, replace $f(x_{i-1})$ and $f(x_i)$ by

$$f^*(x_i) = f^*(x_{i-1}) = \frac{\omega_i f(x_i) + \omega_{i-1} f(x_{i-1})}{\omega_i + \omega_{i-1}},$$

and repeat until the constraint is satisfied. Take $g = f^*$.

(b) Apply this algorithm to the case $n = 4$, $f(x_1) = 23$, $f(x_2) = 27$, $f(x_3) = 25$, and $f(x_4) = 28$, when the weights are all equal.

1.19 (Continuation of Problem 1.18) The simple *tree ordering* is obtained when one compares treatment effects with a control state. The isotonic regression is then obtained under the constraint $g(x_i) \geq g(x_1)$ for $i = 2, \ldots, n$.

(a) Show that the following provides the isotonic regression g^*:

Algorithm A.2 –Tree ordering–

If f is not isotonic, assume w.l.o.g. that the $f(x_i)$'s are in increasing order $(i \geq 2)$. Find the smallest j such that

$$A_j = \frac{\omega_1 f(x_1) + \cdots + \omega_j f(x_j)}{\omega_1 + \cdots \omega_j} < f(x_{j+1}) \,,$$

take $g^*(x_1) = A_j = g^*(x_2) = \cdots = g^*(x_j)$, $g^*(x_{j+1}) = f(x_{j+1})$, ...

(b) Apply this algorithm to the case where $n = 5$, $f(x_1) = 18$, $f(x_2) = 17$, $f(x_3) = 12$, $f(x_4) = 21$, $f(x_5) = 16$, with $\omega_1 = \omega_2 = \omega_5 = 1$ and $\omega_3 = \omega_4 = 3$.

1.20 For the setup of Example 1.8, where $X \sim \mathcal{N}_p(\theta, I_p)$:

(a) Show that the maximum likelihood estimator of $\lambda = \|\theta\|^2$ is $\hat{\lambda}(x) = ||x||^2$ and that it has a constant bias equal to p.

(b) If we observe $Y = \|X\|^2$, distributed as a noncentral chi squared random variable (see Appendix A), show that the MLE of λ is the solution in (1.13). Discuss what happens if $y < p$.

(c) In part (a), if the reference prior $\pi(\theta) = \|\theta\|^{-(p-1)}$ is used, show that the posterior distribution is given by (1.18), with posterior mean (1.19).

1.21 Show that under the specification $\theta_j \sim \pi_j$ and $X \sim f_j(x|\theta_j)$ for model $\mathcal{M}_j$, the posterior probability of model $\mathcal{M}_j$ is given by (1.20).

1.22 Suppose that $X \sim f(x|\theta)$, with prior distribution $\pi(\theta)$, an interest is in the estimation of the parameter $h(\theta)$.

(a) Using the loss function $\mathrm{L}(\delta, h(\theta))$, show that the estimator that minimizes the Bayes risk
$$\int\int \mathrm{L}(\delta, h(\theta)) f(x|\theta)\pi(\theta) dx d\theta$$
is given by the estimator δ that minimizes (for each x)
$$\int \mathrm{L}(\delta, h(\theta))\ \pi(\theta|x)\ d\theta\ .$$
(b) For $\mathrm{L}(\delta,\theta) = \|h(\theta) - \delta\|^2$, show that the Bayes estimator of $h(\theta)$ is $\delta^\pi(x) = \mathbb{E}^\pi[h(\theta)|x]$.
(c) For $\mathrm{L}(\delta,\theta) = |h(\theta) - \delta|$, show that the Bayes estimator of $h(\theta)$ is the median of the posterior distribution.

1.23 For each of the following cases, give the posterior and marginal distributions.
(a) $X|\sigma \sim \mathcal{N}(0, \sigma^2)$, $1/\sigma^2 \sim \mathcal{G}(1, 2)$;
(b) $X|\lambda \sim \mathcal{P}(\lambda)$, $\lambda \sim \mathcal{G}(2, 1)$;
(c) $X|p \sim \mathcal{N}eg(10, p)$, $p \sim \mathcal{B}e(1/2, 1/2)$.

1.24 Let $f(x)$ be a unimodal continuous density, and for a given value of α, let the interval $[a, b]$ satisfy $\int_a^b f = 1 - \alpha$.
(a) Show that the shortest interval satisfying the probability constraint is given by $f(a) = f(b)$, where a and b are on each side of the mode of f.
(b) Show that if f is symmetric, then the shortest interval satisfies $a = -b$.
(c) Find the 90% highest posterior credible regions for the posterior distributions of Problem 1.23.

1.25 (Bauwens 1991) Consider $X_1, \ldots, X_n$ iid $\mathcal{N}(\theta, \sigma^2)$ with prior
$$\pi(\theta, \sigma^2) = \sigma^{-2(\alpha+1)} \exp(-s_0^2/2\sigma^2).$$
(a) Compute the posterior distribution $\pi(\theta, \sigma^2|x_1, \ldots, x_n)$ and show that it depends only on $\bar{x}$ and $s^2 = \sum_{i=1}^n (x_i - \bar{x})^2$.
(b) Derive the posterior expectation $\mathbb{E}^\pi[\sigma^2|x_1, \ldots, x_n]$ and show that its behavior when α and s_0 both converge to 0 depends on the limit of $(s_0^2/\alpha) - 1$.

1.26 In the setting of Section 1.3,
(a) Show that, if the prior distribution is improper, the marginal distribution is also improper.
(b) Show that if the prior $\pi(\theta)$ is improper and the sample space $\mathcal{X}$ is finite, the posterior distribution $\pi(\theta|x)$ is not defined for some value of x.
(c) Consider $X_1, \ldots, X_n$ distributed according to $\mathcal{N}(\theta_j, 1)$, with $\theta_j \sim \mathcal{N}(\mu, \sigma^2)$ $(1 \le j \le n)$ and $\pi(\mu, \sigma^2) = \sigma^{-2}$. Show that the posterior distribution $\pi(\mu, \sigma^2|x_1, \ldots, x_n)$ is not defined.

1.27 Assuming that $\pi(\theta) = 1$ is an acceptable prior for real parameters, show that this generalized prior leads to $\pi(\sigma) = 1/\sigma$ if $\sigma \in \mathbb{R}^+$ and to $\pi(\varrho) = 1/\varrho(1-\varrho)$ if $\varrho \in [0, 1]$ by considering the "natural" transformations $\theta = \log(\sigma)$ and $\theta = \log(\varrho/(1-\varrho))$.

1.28 For each of the following situations, exhibit a conjugate family for the given distribution:
(a) $X \sim \mathcal{G}(\theta, \beta)$; that is, $f_\beta(x|\theta) = \beta^\theta x^{\theta-1} e^{-\beta x}/\Gamma(\theta)$.
(b) $X \sim \mathcal{B}e(1, \theta)$, $\theta \in \mathbb{N}$.

1.29 Show that, if $X \sim \mathcal{B}e(\theta_1, \theta_2)$, there exist conjugate priors on $\theta = (\theta_1, \theta_2)$ but that they do not lead to tractable posterior quantities, except for the computation of $\mathbb{E}^\pi[\theta_1/(\theta_1 + \theta_2)|x]$.

1.30 Consider $X_1, \ldots, X_n \sim \mathcal{N}(\mu + \nu, \sigma^2)$, with $\pi(\mu, \nu, \sigma) \propto 1/\sigma$.

(a) Show that the posterior distribution is not defined for every n.

(b) Extend this result to overparameterized models with improper priors.

1.31 Consider estimation in the *linear model*

$$Y = b_1 X_1 + b_2 X_2 + \epsilon,$$

under the constraint $0 \le b_1,\ b_2 \le 1$, for a sample $(Y_1, X_{11}, X_{21}), \ldots, (Y_n, X_{1n}, X_{2n})$ when the errors ϵ_i are independent and distributed according to $\mathcal{N}(0,1)$. A noninformative prior is

$$\pi(b_1, b_2) = \mathbb{I}_{[0,1]}(b_1)\mathbb{I}_{[0,1]}(b_2)\ .$$

(a) Show that the posterior means are given by $(i = 1, 2)$

$$\mathbb{E}^\pi[b_i|y_1, \ldots, y_n] = \frac{\int_0^1\int_0^1 b_i \prod_{j=1}^n \varphi(y_j - b_1 X_{1j} - b_2 X_{2j})\, db_1\, db_2}{\int_0^1\int_0^1 \prod_{j=1}^n \varphi(y_j - b_1 X_{1j} - b_2 X_{2j})\, db_1\, db_2},$$

where φ is the density of the standard normal distribution.

(b) Show that an equivalent expression is

$$\delta_i^\pi(y_1, \ldots, y_n) = \frac{\mathbb{E}^\pi\left[b_i \mathbb{I}_{[0,1]^2}(b_1, b_2)|y_1, \ldots, y_n\right]}{P^\pi\left[(b_1, b_2) \in [0,1]^2|y_1, \ldots, y_n\right]},$$

where the right-hand term is computed under the distribution

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix}, \left(X^t X\right)^{-1}\right),$$

with $(\hat{b}_1, \hat{b}_2)$ the unconstrained least squares estimator of (b_1, b_2) and

$$X = \begin{pmatrix} X_{11} & X_{21} \\ \vdots & \vdots \\ X_{1n} & X_{2n} \end{pmatrix}.$$

(c) Show that the posterior means cannot be written explicitly, except in the case where $\left(X^t X\right)$ is diagonal.

1.32 (Berger 1985) Consider the hierarchical model

$$\begin{aligned} X|\theta &\sim \mathcal{N}_p(\theta, \sigma^2 I_p), \\ \theta|\xi &\sim \mathcal{N}_p(\xi\mathbf{1}, \sigma_\pi^2 I_p), \\ \xi \sim \mathcal{N}(\xi_0, \tau^2) \quad &, \quad \sigma_\pi^2 \sim \pi_2(\sigma_\pi^2) \end{aligned}$$

where $\mathbf{1} = (1, \ldots, 1)^t \in \mathbb{R}^p$, and σ^2, ξ_0, and τ^2 are fixed.

(a) Show that

$$\delta(x|\xi, \sigma_\pi) = x - \frac{\sigma^2}{\sigma^2 + \sigma_\pi^2}(x - \xi\mathbf{1}),$$

$$\begin{aligned} \pi_2(\xi, \sigma_\pi^2|x) &\propto (\sigma^2 + \sigma_\pi^2)^{-p/2} \exp\left\{-\frac{\|x - \xi\mathbf{1}\|^2}{2(\sigma^2 + \sigma_\pi^2)}\right\} e^{-(\xi - \xi_0)^2/2\tau^2} \pi_2(\sigma_\pi^2) \\ &\propto \frac{\pi_2(\sigma_\pi^2)}{(\sigma^2 + \sigma_\pi^2)^{p/2}} \exp\left\{-\frac{p(\bar{x} - \xi)^2 + s^2}{2(\sigma^2 + \sigma_\pi^2)} - \frac{(\xi - \xi_0)^2}{2\tau^2}\right\} \end{aligned}$$

with $s^2 = \sum_i (x_i - \bar{x})^2$. Deduce that $\pi_2(\xi|\sigma_\pi^2, x)$ is a normal distribution. Give the mean and variance of the distribution.

(b) Show that

$$\delta^\pi(x) = \mathbb{E}^{\pi_2(\sigma_\pi^2|x)}\left[x - \frac{\sigma^2}{\sigma^2+\sigma_\pi^2}(x-\bar{x}\mathbf{1}) - \frac{\sigma^2+\sigma_\pi^2}{\sigma_1^2+\sigma_\pi^2+p\tau^2}(\bar{x}-\xi_0)\mathbf{1}\right]$$

and

$$\pi_2(\sigma_\pi^2|x) \propto \frac{\tau \exp -\frac{1}{2}\left[\frac{s^2}{\sigma^2+\sigma_\pi^2} + \frac{p(\bar{x}-\xi_0)^2}{p\tau^2+\sigma^2+\sigma_\pi^2}\right]}{(\sigma^2+\sigma_\pi^2)^{(p-1)/2}(\sigma^2+\sigma_\pi^2+p\tau^2)^{1/2}}\, \pi_2(\sigma_\pi^2).$$

(c) Deduce the representation

$$\begin{aligned}\delta^\pi(x) = x &- \mathbb{E}^{\pi_2(\sigma_\pi^2|x)}\left[\frac{\sigma^2}{\sigma^2+\sigma_\pi^2}\right](x-\bar{x}\mathbf{1}) \\ &- \mathbb{E}^{\pi_2(\sigma_\pi^2|x)}\left[\frac{\sigma^2+\sigma_\pi^2}{\sigma_1^2+\sigma_\pi^2+p\tau^2}\right](\bar{x}-\xi_0)\mathbf{1}.\end{aligned}$$

and discuss the appeal of this expression from an integration point of view.

1.33 A classical linear regression can be written as $Y \sim \mathcal{N}_p(X\beta, \sigma^2 I_p)$ with X a $p \times q$ matrix and $\beta \in \mathbb{R}^q$.

(a) When X is known, give the natural parameterization of this exponential family and derive the conjugate priors on (β, σ^2).

(b) Generalize to $\mathcal{N}_p(X\beta, \Sigma)$.

1.34 [9] An *autoregressive model* AR(1) connects the random variables in a sample $X_1, \ldots, X_n$ through the relation $X_{t+1} = \varrho X_t + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ is independent of X_t.

(a) Show that the X_t's induce a Markov chain and derive a stationarity condition on ϱ. Under this condition, what is the stationary distribution of the chain?

(b) Give the covariance matrix of $(X_1, \ldots, X_n)$.

(c) If x_0 is a (fixed) starting value for the chain, express the likelihood function and derive a conjugate prior on (ϱ, σ^2). (*Hint:* Note that $X_t|x_{t-1} \sim \mathcal{N}(\varrho x_{t-1}, \sigma^2)$.)

Note: The next four problems involve properties of the exponential family, conjugate prior distributions, and Jeffreys prior distributions. Brown (1986) is a book-length introduction to exponential families, and shorter introductions can be found in Casella and Berger (2001, Section 3.3), Robert (2001, Section 3.2), or Lehmann and Casella (1998, Section 1.5). For conjugate and Jeffreys priors, in addition to Note 1.6.1, see Berger (1985, Section 3.3) or Robert (2001, Sections 3.3 and 3.5).

1.35 Consider $\mathbf{x} = (x_{ij})$ and $\Sigma = (\sigma_{ij})$ symmetric positive-definite $m \times m$ matrices. The *Wishart* distribution, $\mathcal{W}_m(\alpha, \Sigma)$, is defined by the density

$$p_{\alpha,\Sigma}(\mathbf{x}) = \frac{|\mathbf{x}|^{\frac{\alpha-(m+1)}{2}} \exp(-(\mathrm{tr}(\Sigma^{-1}\mathbf{x})/2)}{\Gamma_m(\alpha)|\Sigma|^{\alpha/2}},$$

with $tr(A)$ the trace of A and

[9] This problem requires material that will be covered in Chapter 6. It is put here for those already familiar with Markov chains.

$$\Gamma_m(\alpha) = 2^{\alpha m/2}\pi^{m(m-1)/4}\prod_{i=1}^{m}\Gamma\left(\frac{\alpha - i + 1}{2}\right).$$

(a) Show that this distribution belongs to the exponential family. Give its natural representation and derive the mean of $\mathcal{W}_m(\alpha, \Sigma)$.
(b) Show that, if $Z_1, \ldots, Z_n \sim \mathcal{N}_m(0, \Sigma)$,

$$\sum_{i=1}^{n} Z_i Z_i' \sim \mathcal{W}_m(n, \Sigma).$$

1.36 Consider $X \sim \mathcal{N}(\theta, \theta)$ with $\theta > 0$.
(a) Indicate whether the distribution of X belongs to an exponential family and derive the conjugate priors on θ.
(b) Determine the Jeffreys prior $\pi^J(\theta)$.

1.37 Show that a Student's t distribution $\mathcal{T}_p(\nu, \theta, \tau^2)$ does not allow for a conjugate family, apart from $\mathcal{F}_0$, the (trivial) family that contains all distributions.

1.38 (Robert 1991) The generalized inverse normal distribution $\mathcal{IN}(\alpha, \mu, \tau)$ has the density

$$K(\alpha, \mu, \tau)|y|^{-\alpha} \exp\left\{-\left(\frac{1}{y} - \mu\right)^2 / 2\tau^2\right\},$$

with $\alpha > 0$, $\mu \in \mathbb{R}$, and $\tau > 0$.
(a) Show that this density is well defined and that the normalizing factor is

$$K(\alpha, \mu, \tau)^{-1} = \tau^{\alpha-1} e^{-\mu^2/2\tau^2} 2^{(\alpha-1)/2}\, \Gamma\left(\frac{\alpha-1}{2}\right)\, {}_1F_1\left(\frac{\alpha-1}{2}; 1/2; \frac{\mu^2}{2\tau^2}\right),$$

where ${}_1F_1$ is the *confluent hypergeometric function*

$${}_1F_1(a; b; z) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)}{\Gamma(b+k)} \frac{\Gamma(b)}{\Gamma(a)} \frac{z^k}{k!}$$

(see Abramowitz and Stegun 1964).
(b) If $X \sim \mathcal{N}(\mu, \tau^2)$, show that the distribution of $1/X$ is in the $\mathcal{IN}(\alpha, \mu, \tau)$ family.
(c) Deduce that the mean of $Y \sim \mathcal{IN}(\alpha, \mu, \tau)$ is defined for $\alpha > 2$ and is

$$\mathbb{E}_{\alpha,\mu,\tau}[Y] = \frac{\mu}{\tau^2} \frac{{}_1F_1(\frac{\alpha-1}{2}; 3/2; \mu^2/2\tau^2)}{{}_1F_1(\frac{\alpha-1}{2}; 1/2; \mu^2/2\tau^2)}.$$

(d) Show that $\theta \sim \mathcal{IN}(\alpha, \mu, \tau)$ constitutes a conjugate family for the multiplicative model $X \sim \mathcal{N}(\theta, \theta^2)$.

1.39 Recall the situation of Example 1.8 (see also Example 1.12), where $X \sim \mathcal{N}_p(\theta, I_p)$.
(a) For the prior $\pi(\lambda) = 1/\sqrt{\lambda}$, show that the Bayes estimator of $\lambda = ||\theta||^2$ under quadratic loss can be written as

$$\delta^\pi(x) = \frac{{}_1F_1(3/2; p/2; ||x||^2/2)}{{}_1F_1(1/2; p/2; ||x||^2/2)},$$

where ${}_1F_1$ is the confluent hypergeometric function.

(b) Using the series development of ${}_1F_1$ in Problem 1.38, derive an asymptotic expansion of δ^π (for $||x||^2 \to +\infty$) and compare it with $\delta_0(x) = ||x||^2 - p$.

(c) Compare the risk behavior of the estimators of part (b) under the weighted quadratic loss

$$\mathrm{L}(\delta, \theta) = \frac{(||\theta||^2 - \delta)^2}{2||\theta||^2 + p}.$$

1.40 (Smith and Makov 1978) Consider the mixture density

$$X \sim f(x|p) = \sum_{i=1}^{k} p_i f_i(x),$$

where $p_i > 0$, $\sum_i p_i = 1$, and the densities f_i are known. The prior $\pi(p)$ is a Dirichlet distribution $\mathcal{D}(\alpha_1, \ldots, \alpha_k)$.

(a) Explain why the computing time could get prohibitive as the sample size increases.

(b) A sequential alternative which approximates the Bayes estimator is to replace $\pi(p|x_1, \ldots, x_n)$ by $\mathcal{D}(\alpha_1^{(n)}, \ldots, \alpha_k^{(n)})$, with

$$\alpha_1^{(n)} = \alpha_1^{(n-1)} + P(Z_{n1} = 1|x_n), \ldots, \alpha_k^{(n)} = \alpha_k^{(n-1)} + P(Z_{nk} = 1|x_n),$$

and Z_{ni} ($1 \le i \le k$) is the component indicator vector of X_n. Justify this approximation and compare with the updating of $\pi(\theta|x_1, \ldots, x_{n-1})$ when x_n is observed.

(c) Examine the performances of the approximation in part (b) for a mixture of two normal distributions $\mathcal{N}(0, 1)$ and $\mathcal{N}(2, 1)$ when $p = 0.1, 0.25$, and 0.5.

(d) If $\pi_i^n = P(Z_{ni} = 1|x_n)$, show that

$$\hat{p}_i^{(n)}(x_n) = \hat{p}_i^{(n-1)}(x_{n-1}) - a_{n-1}\{\hat{p}_i^{(n-1)} - \pi_i^n\},$$

where $\hat{p}_i^{(n)}$ is the quasi-Bayesian approximation of $\mathbb{E}^\pi(p_i|x_1, \ldots, x_n)$.

1.6 Notes

1.6.1 Prior Distributions

(i) *Conjugate Priors*

When prior information about the model is quite limited, the prior distribution is often chosen from a parametric family. Families $\mathcal{F}$ that are *closed under sampling* (that is, such that, for every prior $\pi \in \mathcal{F}$, the posterior distribution $\pi(\theta|x)$ also belongs to $\mathcal{F}$) are of particular interest, for both parsimony and invariance motivations. These families are called *conjugate* families. Most often, the main motivation for using conjugate priors is their tractability; however, such choices may constrain the subjective input.

For reasons related to the *Pitman–Koopman Lemma* (see the discussion following Example 1.8), conjugate priors can only be found in exponential families. In fact, if the sampling density is of the form

$$f(x|\theta) = C(\theta)h(x)\exp\{R(\theta) \cdot T(x)\}, \tag{1.27}$$