Bayesian Modelling – Problem Sheet 1 (Solutions)

Problem 1

By definition,

$$
\begin{aligned}
\pi(\theta|x) &\propto f(x|\theta)\pi(\theta)\\
&\propto \frac{1}{|\Sigma|^{n/2}} \exp\Big(-\frac{1}{2}\sum_{i=1}^n (x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\Big)\\
&\times \frac{1}{|\Sigma|^{1/2}} \exp\Big(-\frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)\Big)|\Sigma|^{-\frac{\nu_0+d+1}{2}} \exp\Big(-\frac{1}{2}\mathrm{tr}(\Psi_0\Sigma^{-1})\Big)\\
&\propto |\Sigma|^{-\frac{n+\nu_0+d}{2}-1} \exp\Big(-\frac{1}{2}\Big\{\sum_{i=1}^n (x_i-\mu)^T\Sigma^{-1}(x_i-\mu)+\kappa_0(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)\Big\}\Big)\\
&\times \exp\Big(-\frac{1}{2}\mathrm{tr}(\Psi_0\Sigma^{-1})\Big)\\
&\propto |\Sigma|^{-\frac{n+\nu_0+d}{2}-1} \exp\Big(-\frac{1}{2}\Big\{(n+\kappa_0)\mu^T\Sigma^{-1}\mu-2\mu^T\Sigma^{-1}(n\,\bar{x}_n+\kappa_0\mu_0)\Big\}\Big)\\
&\times \exp\Big(-\frac{1}{2}\Big\{\sum_{i=1}^n x_i^T\Sigma^{-1}x_i+\kappa_0\mu_0^T\Sigma^{-1}\mu_0+\mathrm{tr}(\Psi_0\Sigma^{-1})\Big\}\Big)\\
&\propto |\Sigma|^{-\frac{1}{2}} \exp\Big(-\frac{n+\kappa_0}{2}\Big(\mu-\frac{n\,\bar{x}_n+\kappa_0\mu_0}{n+\kappa_0}\Big)^T\Sigma^{-1}\Big(\mu-\frac{n\,\bar{x}_n+\kappa_0\mu_0}{n+\kappa_0}\Big)\Big)\\
&\times |\Sigma|^{-\frac{\nu_0+n+d+1}{2}} \exp\Big(-\frac{1}{2}\Big\{\sum_{i=1}^n x_i^T\Sigma^{-1}x_i+\kappa_0\mu_0^T\Sigma^{-1}\mu_0+\mathrm{tr}(\Psi_0\Sigma^{-1})-(n+\kappa_0)\mu_n^T\Sigma^{-1}\mu_n\Big\}\Big)
\end{aligned}
$$

with μ_n as in the statement of the problem.

Note that the first term of this last expression is proportional to the density of the $\mathcal{N}_d(\mu_n,\kappa_n^{-1}\Sigma)$ distribution, with κ_n as in the statement of the problem.

For the second term, we have

$$
\begin{aligned}
&\exp\Big(-\frac{1}{2}\Big\{\sum_{i=1}^n x_i^T\Sigma^{-1}x_i+\kappa_0\mu_0^T\Sigma^{-1}\mu_0-(n+\kappa_0)\mu_n^T\Sigma^{-1}\mu_n\Big\}\Big)\\
&=\exp\Big(-\frac{1}{2}\Big\{\sum_{i=1}^n (x_i-\bar{x}_n)^T\Sigma^{-1}(x_i-\bar{x}_n)+n\bar{x}_n^T\Sigma^{-1}\bar{x}_n+\kappa_0\mu_0^T\Sigma^{-1}\mu_0-(n+\kappa_0)\mu_n^T\Sigma^{-1}\mu_n\Big\}\Big)
\end{aligned}
$$

where, using the definition of μ_n,

$$\begin{aligned}
n\bar{x}_n^T\Sigma^{-1}\bar{x}_n{+}\kappa_0\mu_0'\Sigma^{-1}\mu_0 - (n+\kappa_0)\mu_n^T\Sigma^{-1}\mu_n&\\
&= n\bar{x}_n^T\Sigma^{-1}\bar{x}_n + \kappa_0\mu_0^T\Sigma^{-1}\mu_0 - (\kappa_0\mu_0 + n\bar{x}_n)^T\Sigma^{-1}\frac{(\kappa_0\mu_0 + n\bar{x}_n)}{n+\kappa_0}\\
&= \frac{n\kappa_0}{n+\kappa_0}\mu_0^T\Sigma^{-1}\mu_0 + \frac{n\kappa_0}{n+\kappa_0}\bar{x}_n^T\Sigma^{-1}\bar{x}_n - 2\frac{n\kappa_0}{n+\kappa_0}\bar{x}_n^T\Sigma^{-1}\mu_0\\
&= \frac{n\kappa_0}{n+\kappa_0}\big(\mu_0 - \bar{x}_n\big)^T\Sigma^{-1}\big(\mu_0 - \bar{x}_n\big).
\end{aligned}$$

Consequently, using the fact that for any matrices A, B, C of appropriate dimensions we have $\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$, we obtain

$$\begin{aligned}
&\exp\Big(-\frac{1}{2}\Big\{\sum_{i=1}^n x_i^T\Sigma^{-1}x_i + \kappa_0\mu_0^T\Sigma^{-1}\mu_0 + \mathrm{tr}(\Psi_0\Sigma^{-1}) - (n+\kappa_0)\mu_n^T\Sigma^{-1}\mu_n\Big\}\Big)\Big\}\\
&= \exp\Big(-\frac{1}{2}\Big\{\sum_{i=1}^n (x_i - \bar{x}_n)^T\Sigma^{-1}(x_i - \bar{x}_n) + \frac{n\kappa_0}{n+\kappa_0}\big(\mu_0 - \bar{x}_n\big)^T\Sigma^{-1}\big(\mu_0 - \bar{x}_n\big) + \mathrm{tr}(\Psi_0\Sigma^{-1})\Big\}\Big)\\
&= \exp\Big(-\frac{1}{2}\Big\{\mathrm{tr}\Big(\sum_{i=1}^n (x_i - \bar{x}_n)^T\Sigma^{-1}(x_i - \bar{x}_n)\Big) + \frac{n\kappa_0}{n+\kappa_0}\mathrm{tr}\Big(\big(\mu_0 - \bar{x}_n\big)^T\Sigma^{-1}\big(\mu_0 - \bar{x}_n\big)\Big) + \mathrm{tr}(\Psi_0\Sigma^{-1})\Big\}\Big)\\
&= \exp\Big(-\frac{1}{2}\Big\{\mathrm{tr}\Big(\sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)^T\Sigma^{-1}\Big) + \frac{n\kappa_0}{n+\kappa_0}\mathrm{tr}\Big(\big(\mu_0 - \bar{x}_n\big)\big(\mu_0 - \bar{x}_n\big)^T\Sigma^{-1}\Big) + \mathrm{tr}(\Psi_0\Sigma^{-1})\Big\}\Big)\\
&= \exp\Big(-\frac{1}{2}\mathrm{tr}\Big\{\Big(\sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)^T + \frac{n\kappa_0}{n+\kappa_0}\big(\mu_0 - \bar{x}_n\big)\big(\mu_0 - \bar{x}_n\big)^T + \Psi_0\Big)\Sigma^{-1}\Big\}\Big)\\
&= \exp\Big(-\frac{1}{2}\mathrm{tr}(\Psi_n\Sigma^{-1})\Big)
\end{aligned}$$

with Ψ_n as in the statement of the question. Therefore, the quantity

$$|\Sigma|^{-\frac{\nu+n+d+1}{2}}\exp\Big(-\frac{1}{2}\Big\{\sum_{i=1}^n x_i^T\Sigma^{-1}x_i + \kappa_0\mu_0^T\Sigma^{-1}\mu_0 + \mathrm{tr}(\Psi_0\Sigma^{-1}) - (n+\kappa_0)\mu_n^T\Sigma^{-1}\mu_n\Big\}\Big)$$

is proportional to the density of the $\mathcal{W}_d^{-1}(\Psi_n, \nu_n)$ distribution, with ν_n as in the statement of the question, and the result follows.

Problem 2 (The Bayesian linear regression model)

1. For $z \in \mathbb{R}^d$ let $\tilde{f}(\cdot|z,\theta)$ be the density of the $\mathcal{N}_d(\beta^T z, \sigma^2)$ distribution. Then,

$$
\begin{aligned}
&\pi(\theta|x)\\
&\propto \Big(\prod_{i=1}^n \tilde{f}(y_i|z_i,\theta)\Big)\pi(\theta)\\
&\propto (\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\beta^T z_i)^2\Big)(\sigma^2)^{-\frac{d}{2}}\exp\Big(-\frac{1}{2\sigma^2}(\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0)\Big)\\
&\times (\sigma^2)^{-a_0-1}e^{-b_0/\sigma^2}\\
&= (\sigma^2)^{-\frac{n+d+2a_0+2}{2}}\\
&\times \exp\Big(-\frac{1}{2\sigma^2}\big(y^Ty - 2y^TZ\beta + \beta^TZ^TZ\beta + \beta^T\Sigma_0^{-1}\beta - 2\beta^T\Sigma_0^{-1}\beta_0 + \beta_0^T\Sigma^{-1}\beta_0 + 2b_0\big)\Big).
\end{aligned}
$$

Next, note that

$$
\begin{aligned}
-2y^TZ\beta + \beta^TZ^TZ\beta + \beta^T\Sigma_0^{-1}\beta - 2\beta^T\Sigma_0^{-1}\beta_0&\\
&= \beta^T(Z^TZ+\Sigma_0^{-1})\beta - 2\beta^T(Z^Ty + \Sigma_0^{-1}\beta_0)\\
&= \beta^T\Sigma_n^{-1}\beta - 2\beta^T\Sigma_n^{-1}\mu_n + \mu_n^T\Sigma_n^{-1}\mu_n - \mu_n^T\Sigma_n^{-1}\mu_n\\
&= (\beta-\mu_n)^T\Sigma_n^{-1}(\beta-\mu_n) - \mu_n^T\Sigma_n^{-1}\mu_n
\end{aligned}
$$

where Σ_n and μ_n are as in the statement of the problem. Consequently,

$$
\begin{aligned}
\pi(\theta|x) &\propto (\sigma^2)^{-\frac{d}{2}}\exp\Big(-\frac{1}{2\sigma^2}(\beta-\mu_n)^T\Sigma_n^{-1}(\beta-\mu_n)\Big)\\
&\quad\times(\sigma^2)^{-\frac{n+2a_0}{2}-1}\exp\Big(-\frac{1}{2\sigma^2}\big(y^Ty - \mu_n^T\Sigma_n^{-1}\mu_n + \beta_0^T\Sigma_0^{-1}\beta_0 + 2b_0\big)\Big)
\end{aligned}
$$

where the first term is proportional to the density of the $\mathcal{N}_d(\mu_n, \sigma^2\Sigma_n)$ distribution while the second term is proportional to the density of the $\Gamma^{-1}(a_n, b_n)$ distribution, where a_n and b_n as in the statement of the problem.

2. Remark first that $Z^Ty = Z^TZ\hat{\beta}$ and thus

$$
\begin{aligned}
\mu_n &= (Z^TZ+\Sigma_0^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + Z^TZ\hat{\beta})\\
&= M_0\mu_0 + (Z^TZ+\Sigma_0^{-1})^{-1}Z^TZ\hat{\beta}\\
&= M_0\mu_0 + (I_d - M_0)\hat{\beta}.
\end{aligned}
$$

3. In this case we have

$$
M_0 = \frac{1}{c_0}(Z^TZ + c_0^{-1}\tilde{\Sigma}_0^{-1})^{-1}\tilde{\Sigma}_0^{-1}
$$

so that, as $c_0 \to +\infty$, all the elements of M_0 converges to 0. Hence, as $c_0 \to +\infty$, $\mathbb{E}_\pi[\beta|x] \to \hat{\beta}$.

When c_0 increases, the prior variance "increases" in the sense that, for $c_0' > c_0$, the matrix $(c_0' - c_0)\tilde{\Sigma}_0$ is positive definite. Therefore, as c_0 increases the prior distribution becomes less and less informative about β and, in the limiting case $c_0 = +\infty$, has no impact on the posterior mean of β.

Problem 3

1. For $\theta > 0$ the joint likelihood is

$$f(x^{(n)}|\theta) = \prod_{k=1}^{n} \tilde{f}(x_k|\theta) = \frac{\theta^{n\lambda}}{\Gamma(\lambda)^n}\Big(\prod_{k=1}^{n} x_k\Big)^{\lambda-1} e^{-\theta n\bar{x}_n}, \quad \bar{x}_n := \frac{1}{n}\sum_{k=1}^{n} x_k$$

while the density of the prior distribution is

$$\pi(\theta) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\theta^{\alpha_0-1}e^{-\beta_0\theta}.$$

Then, the Bayes formula implies

$$\pi(\theta|x^{(n)}) \propto f(x^{(n)}|\theta)\pi(\theta) \propto \theta^{\alpha_0+n\lambda-1}e^{-\theta(\beta_0+n\bar{x}_n)}$$

so that $\theta|x^{(n)} \sim \text{Gamma}(\alpha_n, \beta_n)$ where

$$\alpha_n = \alpha_0 + n\lambda, \quad \beta_n = \beta_0 + n\bar{x}_n.$$

2. Let $k \in \mathbb{N}_{>0}$. Then, using the result of part 1,

$$\begin{aligned}
\mathbb{E}_\pi[\theta^k|x^{(n)}] &= \int_0^\infty \theta^k \pi(\theta|x^{(n)})\mathrm{d}\theta \\
&= \int_0^\infty \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\theta^{(\alpha_n+k)-1}e^{-\beta_n\theta}\mathrm{d}\theta \\
&= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\frac{\Gamma(\alpha_n+k)}{\beta_n^{\alpha_n+k}}\int_0^\infty \frac{\beta_n^{\alpha_n+k}}{\Gamma(\alpha_n+k)}\theta^{(\alpha_n+k)-1}e^{-\beta_n\theta}\mathrm{d}\theta \\
&= \frac{\Gamma(\alpha_n+k)}{\Gamma(\alpha_n)\beta_n^k}.
\end{aligned}$$

Using the *hint*, $\Gamma(\alpha_n+k) = (\alpha_n+k-1)\cdots\alpha_n\Gamma(\alpha_n)$ and hence we deduce that

$$\mathbb{E}_\pi[\theta^k|x^{(n)}] = \frac{(\alpha_n+k-1)\cdots\alpha_n}{\beta_n^k}, \quad \forall k \in \mathbb{N}_{>0}.$$

Consequently,

$$\mathbb{E}_\pi[\theta|x^{(n)}] = \frac{\alpha_n}{\beta_n} = \frac{\frac{\alpha_0}{n}+\lambda}{\frac{\beta_0}{n}+\bar{x}_n}, \quad \mathrm{Var}_\pi(\theta|x^{(n)}) = \frac{\alpha_n}{\beta_n^2} = \frac{\mathbb{E}_\pi[\theta|x]}{\beta_n}.$$

3. Since $\mathbb{E}[X_1] = \lambda/\theta_0 < +\infty$ we have, by the law of large numbers,

$$\lim_{n\to+\infty} \bar{X}_n = \mathbb{E}[X_1] = \frac{\lambda}{\theta_0}, \quad \text{almost surely}$$

and thus, using the results of part 2, we deduce that

$$\lim_{n\to+\infty} \mathbb{E}_\pi[\theta|X^{(n)}] = \theta_0, \quad \text{almost surely.}$$

This also implies that $\lim_{n\to+\infty} \beta_n = +\infty$, almost surely, and thus

$$\lim_{n\to+\infty} \mathrm{Var}_\pi\big(\theta|X^{(n)}\big) = 0, \quad \text{almost surely.}$$

4. Since $e^{-\theta} > 0$ for all $\theta \in \Theta$ it is sensible to take $\mathcal{D} = \mathbb{R}_{>0}$. Then, we can for instance consider the quadratic loss function $L : \Theta \times \mathcal{D} \to [0, +\infty)$ defined by

$$L(\theta, d) = (e^{-\theta} - d)^2, \quad (\theta, d) \in \Theta \times \mathcal{D}.$$

As seen during the lectures, for the quadratic loss function the minimizer of the posterior expected loss is simply the posterior mean. Therefore,

$$\begin{aligned}
\gamma^\pi(x^{(n)}) = \mathbb{E}_\pi[e^{-\theta}|x^{(n)})] &= \int_0^\infty e^{-\theta}\pi(\theta|x^{(n)}) \\
&= \int_0^\infty \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\theta^{\alpha_n-1}e^{-(\beta_n+1)\theta}\mathrm{d}\theta \\
&= \Big(\frac{\beta_n}{\beta_n+1}\Big)^{\alpha_n}\int_0^\infty \frac{(\beta_n+1)^{\alpha_n}}{\Gamma(\alpha_n)}\theta^{\alpha_n-1}e^{-(\beta_n+1)\theta}\mathrm{d}\theta \\
&= \Big(\frac{\beta_n}{\beta_n+1}\Big)^{\alpha_n}.
\end{aligned}$$

5. By definition, $\delta^\pi(x^{(n)})$ minimizes the posterior expected loss; that is, $\delta^\pi(x^{(n)})$ minimizes the function $\rho : \Theta \to [0, +\infty)$ defined by

$$\rho(d) = \int_\Theta L(\theta, d)\pi(\theta|x^{(n)})\mathrm{d}\theta, \quad d \in \Theta.$$

Using the definition of L, we have

$$\begin{aligned}
\rho(d) &= e^{\kappa\, d}\int_\Theta e^{-\kappa\theta}\pi(\theta|x^{(n)})\mathrm{d}\theta - \kappa d + \kappa\mathbb{E}_\pi[\theta|x^{(n)}] - 1 \\
&= e^{\kappa\, d}\mathbb{E}_\pi[e^{-\kappa\theta}|x^{(n)}] - \kappa d + \kappa\mathbb{E}_\pi[\theta|x^{(n)}] - 1.
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\mathrm{d}\rho}{\mathrm{d}d} &= \kappa\, e^{\kappa\, d}\mathbb{E}_\pi[e^{-\kappa\theta}|x^{(n)}] - \kappa \\
\frac{\mathrm{d}^2\rho}{\mathrm{d}d^2} &= \kappa^2\, e^{\kappa\, d}\mathbb{E}[e^{-\kappa\theta}|x^{(n)}] > 0, \quad \forall d \in \Theta
\end{aligned}$$

and thus $\delta^\pi(x^{(n)})$ is uniquely defined by the condition

$$e^{\kappa\,\delta^\pi(x^{(n)})}\mathbb{E}_\pi[e^{-\kappa\theta}|x^{(n)}]-1=0$$

implying that

$$\delta^\pi(x^{(n)})=-\frac{1}{\kappa}\log\mathbb{E}_\pi[e^{-\kappa\theta}|x^{(n)}].$$

Lastly, since

$$\begin{aligned}\mathbb{E}_\pi[e^{-\kappa\theta}|x^{(n)}]&=\int_\Theta\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\theta^{\alpha_n-1}e^{-(\beta_n+\kappa)\theta}\\&=\Big(\frac{\beta_n}{\beta_n+\kappa}\Big)^{\alpha_n}\int_\Theta\frac{(\beta_n+\kappa)^{\alpha_n}}{\Gamma(\alpha_n)}\theta^{\alpha_n-1}e^{-(\beta_n+\kappa)\theta}\\&=\Big(\frac{\beta_n}{\beta_n+\kappa}\Big)^{\alpha_n}\end{aligned}$$

it follows that

$$\delta^\pi(x^{(n)})=\frac{\alpha_n}{\kappa}\log\Big(1+\frac{\kappa}{\beta_n}\Big).$$

Problem 4

1. By definition, δ^π is such that

$$\delta^\pi(x)\in\operatorname*{argmin}_{d\in\mathbb{R}}\rho(\pi,d|x),\quad\forall x\in\mathcal{X}.$$

Next, if $f(\cdot|\theta)$ is the pdf of the $\mathcal{N}_1(\theta,1)$ distribution and if $\theta\sim\mathcal{N}_1(0,1)$ then it is easy to check that $\theta|x\sim\mathcal{N}_1(0.5x,0.5)$. In addition, remark that for any $x\in\mathcal{X}$ we have

$$\begin{aligned}e^{\frac{3}{4}\theta^2}\pi(\theta|x)&=\frac{1}{\sqrt{\pi}}\exp\Big(-\frac{1}{4}\theta^2+x\theta-\frac{x^2}{4}\Big)\\&=\frac{1}{\sqrt{\pi}}\exp\Big(-\frac{1}{4}\Big(\theta^2-4x\theta+x^2\Big)\Big)\\&=\frac{1}{\sqrt{\pi}}\exp\Big(-\frac{1}{4}(\theta-2x)^2\Big)\Big)e^{\frac{3}{4}x^2}\end{aligned}$$

and thus

$$\operatorname*{argmin}_{d\in\mathbb{R}}\rho(\pi,d|x)=\operatorname*{argmin}_{d\in\mathbb{R}}\int_\Theta(\theta-d)^2\tilde{\pi}(\theta|x)\mathrm{d}\theta$$

where $\tilde{\pi}(\cdot|x)$ is the density of the $\mathcal{N}_1(2x,2)$ distribution. This shows that, for any $(d,x)\in\mathbb{R}^2$, $\rho(\pi,d|x)<+\infty$ so that δ^π is unique and defined by (see Chapter 2, Theorem 2.3)

$$\delta^\pi(x)=\mathbb{E}_{\tilde{\pi}}[\theta|x]=\int_\Theta\theta\,\tilde{\pi}(\theta|x)\mathrm{d}\theta=2x,\quad x\in\mathcal{X}.$$

(Say differently, for $x\in\mathcal{X}$, $\delta^\pi(x)$ is the expected value of θ under $\tilde{\pi}(\theta|x)$.)

2. We have

$$\begin{aligned} r(\pi) &= \int_\Theta R(\theta, \delta^\pi)\pi(\theta)\mathrm{d}\theta \\ &= \int_\Theta e^{\frac{3}{4}\theta^2}\left\{\int_{\mathcal{X}}(\theta - 2x)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}\Big)\mathrm{d}x\right\}\frac{1}{\sqrt{2\pi}}e^{-\frac{\theta^2}{2}}\mathrm{d}\theta \\ &= \int_\Theta \left\{\int_{\mathcal{X}}(\theta - 2x)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}\Big)\mathrm{d}x\right\}\frac{1}{\sqrt{2\pi}}e^{\frac{\theta^2}{4}}\mathrm{d}\theta \end{aligned}$$

where, for $\theta \in \Theta$,

$$\frac{1}{\sqrt{2\pi}}\int_{\mathcal{X}}(\theta - 2x)^2 e^{-\frac{(x-\theta)^2}{2}}\mathrm{d}x = \theta^2 - 4\theta^2 + 4(\theta^2+1) = \theta^2 + 4.$$

Therefore

$$r(\pi) = \int_\Theta (\theta^2+4)e^{\frac{\theta^2}{4}}\mathrm{d}\theta = +\infty.$$

3. It suffices to note that, for any $\theta \in \Theta$,

$$\begin{aligned} R(\theta, \delta_0) - R(\theta, \delta^\pi) &= \int_{\mathcal{X}} \Big(L(\theta, \delta_0(x)) - L(\theta, \delta^\pi(x))\Big) f(x|\theta)\mathrm{d}x \\ &= \int_{\mathcal{X}} \Big(L(\theta, x) - L(\theta, 2x)\Big) f(x|\theta)\mathrm{d}x \\ &= e^{\frac{3}{4}\theta^2}\int_{\mathcal{X}} \Big((\theta - x)^2 - (\theta-2x)^2\Big) f(x|\theta)\mathrm{d}x \\ &= e^{\frac{3}{4}\theta^2}\int_{\mathcal{X}} (2\theta x - 3x^2) f(x|\theta)\mathrm{d}x \\ &= e^{\frac{3}{4}\theta^2}\big(2\theta^2 - 3(\theta^2+1)\big) \\ &= -e^{\frac{3}{4}\theta^2}(\theta^2+3) \\ &< 0. \end{aligned}$$

Problem 5

1. Let $x \in \mathcal{X}$, $\lambda \in (0,1)$ and $d_1, d_2 \in \Theta$ be such that $d_1 \neq d_2$. Then,

$$\begin{aligned}
\lambda\rho(\pi, d_1|x) + (1-\lambda)\rho(\pi, d_2|x) \\
&= \int_\Theta \Big(\lambda L(\theta, d_1) + (1-\lambda)L(\theta, d_2)\Big)\pi(\theta|x)\mathrm{d}\theta \\
&= \int_\Theta \Big(\lambda \tilde{L}(\theta - d_1) + (1-\lambda)\tilde{L}(\theta - d_2)\Big)\pi(\theta|x)\mathrm{d}\theta \\
&\geq \int_\Theta \Big(\tilde{L}\Big(\lambda(\theta - d_1) + (1-\lambda)(\theta - d_2)\Big)\Big)\pi(\theta|x)\mathrm{d}\theta \qquad (1) \\
&= \int_\Theta \Big(\tilde{L}\Big(\theta - \big(\lambda d_1 + (1-\lambda)d_2\big)\Big)\Big)\pi(\theta|x)\mathrm{d}\theta \\
&= \rho(\pi, \lambda d_1 + (1-\lambda)d_2|x)
\end{aligned}$$

   where the inequality holds because $\tilde{L}$ is convex on $\mathbb{R}^d$. This shows that the mapping $d \mapsto \rho(\pi, d|x)$ is convex for every $x \in \mathcal{X}$.

2. If $\tilde{L}$ is strictly convex on $\mathbb{R}^d$ then inequality (1) is strict (because $d_1 \neq d_2$ and $\lambda \in (0,1)$) and the mapping $d \mapsto \rho(\pi, d|x)$ is strictly convex.

   To prove that δ^π is unique we proceed by contradiction. Assume that there exist two estimators δ_1^π and δ_2^π such that

$$\delta_1^\pi(x), \delta_2^\pi(x) \in \operatorname*{argmin}_{d \in \Theta} \rho(\pi, d|x), \quad \forall x \in \mathcal{X} \qquad (2)$$

   and such that, for some $x \in \mathcal{X}$, we have $\delta_1^\pi(x) \neq \delta_2^\pi(x)$. Note that (2) implies that

$$\rho(\pi, \delta_1^\pi(x)|x) = \rho(\pi, \delta_2^\pi(x)|x), \quad \forall x \in \mathcal{X}.$$

   Let $x^* \in \mathcal{X}$ be such that $\delta_1^\pi(x^*) \neq \delta_2^\pi(x^*)$ and let δ_3^π be the estimator defined by

$$\delta_3^\pi(x) = \frac{\delta_1^\pi(x) + \delta_2^\pi(x)}{2}, \quad x \in \mathcal{X}.$$

   Then, using the fact that the mapping $d \mapsto \rho(\pi, d|x^*)$ is strictly convex we have

$$\rho(\pi, \delta_3^\pi(x^*)|x^*) < \frac{1}{2}\rho(\pi, \delta_1^\pi(x^*)|x^*) + \frac{1}{2}\rho(\pi, \delta_2^\pi(x^*)|x^*) = \rho(\pi, \delta_1^\pi(x^*)|x^*)$$

   which contradicts (2). Hence, δ^π is unique.