Bayesian statistics

Exercise 1 (Bernoulli model): 1. Model $p(x|\theta) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}$
Prior $\pi(\theta) = 1_{[0,1]}(\theta)$
Posterior $\pi(\theta|x) \propto \theta^S(1-\theta)^{n-S}1_{[0,1]}(\theta)$ with $S = \sum_{i=1}^n x_i$
This is a Beta distribution with parameter $(S+1, n-S+1)$.

2. $\widehat{\theta}(x) = \mathrm{E}(\theta|x) = \frac{S+1}{n+2}$.

3. Bias $\mathrm{E}_\theta(\widehat{\theta}(X)) = \frac{n\theta+1}{n+2}$ gives $b(\theta,\widehat{\theta}) = \frac{1-2\theta}{n+2}$
Variance $\mathrm{var}_\theta(\widehat{\theta}(X)) = \frac{n}{(n+2)^2}\theta(1-\theta)$
Quadratic risk, $R(\theta,\widehat{\theta}) = b(\theta,\widehat{\theta})^2 + \mathrm{var}_\theta(\widehat{\theta}(X)) = \frac{1}{(n+2)^2}(4(\theta-\frac{1}{2})^2 + n\theta(1-\theta))$
Bayes quadratic risk, $r(\widehat{\theta}) = \mathrm{E}(R(\theta,\widehat{\theta})) = \frac{1}{6(n+2)}$

4. MLE $\widehat{\theta}_{\mathrm{MLE}}(x) = \frac{S}{n}$.
Quadratic risk $R(\theta,\widehat{\theta}_{\mathrm{MLE}}) = \frac{\theta(1-\theta)}{n}$.
Bayes quadratic risk, $r(\widehat{\theta}_{\mathrm{MLE}}) = \frac{1}{6n}$, higher than with Bayes estimator.
Quadratic risk lower for $\theta(1-\theta) < \frac{n}{4(2n+1)}$, that is $|\theta - \frac{1}{2}| > \frac{1}{2}\sqrt{\frac{n+1}{2n+1}}$.

Exercise 2 (Poisson model): 1. Model $p(x|\theta) = \prod_{i=1}^n e^{-\theta}\frac{\theta^{x_i}}{x_i!}$
Prior $\pi(\theta) = \lambda e^{-\lambda\theta}$
Posterior $\pi(\theta|x) \propto e^{-n\theta}\theta^S e^{-\lambda\theta} = \theta^S e^{-(\lambda+n)\theta}$ with $S = \sum_{i=1}^n x_i$
This is a Gamma distribution with parameters $(\lambda+n, S+1)$.

2. $\widehat{\theta}(x) = \mathrm{E}(\theta|x) = \frac{S+1}{\lambda+n}$.

3. Bias $\mathrm{E}_\theta(\widehat{\theta}(X)) = \frac{n\theta+1}{n+\lambda}$ gives $b(\theta,\widehat{\theta}) = \frac{1-\lambda\theta}{\lambda+n}$
Variance $\mathrm{var}_\theta(\widehat{\theta}(X)) = \frac{n\theta}{(\lambda+n)^2}$
Quadratic risk, $R(\theta,\widehat{\theta}) = b(\theta,\widehat{\theta})^2 + \mathrm{var}_\theta(\widehat{\theta}(X)) = \frac{(1-\lambda\theta)^2+n\theta}{(\lambda+n)^2}$
Bayes quadratic risk, $r(\widehat{\theta}) = \mathrm{E}(R(\theta,\widehat{\theta})) = \frac{1}{\lambda(\lambda+n)}$

Exercise 3 (Gaussian model – mean): 1. Model $p(x|\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x_i-\theta)^2}$
Prior $\pi(\theta) \propto e^{-\frac{1}{2}(\theta-\mu)^2}$
Posterior $\pi(\theta|x) \propto e^{-\frac{1}{2}\left((\theta-\mu)^2 + \sum_{i=1}^n \frac{1}{\sigma^2}(x_i-\theta)^2\right)}$
that is $\pi(\theta|x) \propto e^{-\frac{1}{2}\left(\theta^2\left(\frac{n}{\sigma^2}+1\right) - 2\theta\left(\frac{S}{\sigma^2}+\mu\right)\right)}$ with $S = \sum_{i=1}^n x_i$
This is a Gaussian distribution with mean $\frac{\frac{S}{\sigma^2}+\mu}{\frac{n}{\sigma^2}+1}$ and variance $\frac{1}{\frac{n}{\sigma^2}+1}$.

2. $\widehat{\theta}(x) = \mathrm{E}(\theta|x) = \frac{\frac{S}{\sigma^2}+\mu}{\frac{n}{\sigma^2}+1}$.

3. $\mathrm{E}(\widehat{\theta}(X)) = \frac{\frac{n\theta}{\sigma^2}+\mu}{\frac{n}{\sigma^2}+1}$
Bias $b(\theta,\widehat{\theta}) = \frac{\mu-\theta}{\frac{n}{\sigma^2}+1}$

Variance $\text{var}_\theta(\widehat{\theta}(X)) = \frac{\frac{n}{\sigma^2}}{(\frac{n}{\sigma^2}+1)^2}$
Quadratic risk $R(\theta, \widehat{\theta}) = \frac{1}{(\frac{n}{\sigma^2}+1)^2}((\mu-\theta)^2 + \frac{n}{\sigma^2})$
Bayes risk $r(\widehat{\theta}) = \frac{1}{\frac{n}{\sigma^2}+1}$ (using the fact that $\text{E}((\mu-\theta)^2) = \text{var}(\theta) = 1$).
The variance term dominates for large n.

Exercise 4 (Bernoulli model – discrete prior)**:** 1. Model $p(x|\theta) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}$
Prior $\pi(\theta) \propto 1$ for the reference measure $\mu = \delta_{\frac{1}{4}} + \delta_{\frac{1}{3}}$
Posterior $\pi(\theta|x) \propto \theta^S(1-\theta)^{n-S}$ with $S = \sum_{i=1}^n x_i$
that is,

$$\frac{\pi(\frac{1}{3}|x)}{\pi(\frac{1}{4}|x)} = \left(\frac{4}{3}\right)^S \left(\frac{8}{9}\right)^{n-S}$$

and

$$\pi(\frac{1}{4}|x) = \frac{1}{1+\frac{\pi(\frac{1}{3}|x)}{\pi(\frac{1}{4}|x)}}, \quad \pi(\frac{1}{3}|x) = \frac{1}{1+\frac{\pi(\frac{1}{4}|x)}{\pi(\frac{1}{3}|x)}}.$$

2. $\widehat{\theta}(x) = \text{E}(\theta|x) = \frac{1}{4}\pi(\frac{1}{4}|x) + \frac{1}{3}\pi(\frac{1}{3}|x)$

3. When $n \to +\infty$, we have $S/n \to \theta$ by the Strong Law of Large Numbers so that

$$\pi(\frac{1}{4}|x) \approx \frac{1}{1+\left(\left(\frac{4}{3}\right)^\theta \left(\frac{8}{9}\right)^{1-\theta}\right)^n}$$

Now

$$\left(\frac{4}{3}\right)^\theta \left(\frac{8}{9}\right)^{1-\theta} < 1 \quad \iff \quad \theta < \theta_0 = \frac{\log \frac{9}{8}}{\log \frac{3}{2}} \approx 0.29.$$

Thus $\widehat{\theta}(x) \to \frac{1}{4}$ if $\theta < \theta_0$, $\widehat{\theta}(x) \to \frac{1}{3}$ if $\theta > \theta_0$ and $\widehat{\theta}(x) \to \frac{7}{12}$ if $\theta = \theta_0$.

Exercise 5 (Translated exponential)**:** 1. Model $p(x|\theta) = \prod_{i=1}^n e^{\theta - x_i}$ with $x_i \geq \theta$ for all $i = 1, \dots, n$
Prior $\pi(\theta) \propto e^{-\lambda\theta}$
Posterior $\pi(\theta|x) \propto e^{(n-\lambda)\theta} 1_{\theta \leq m}$ with $m = \min_{i=1,\dots,n} x_i$

2. $\widehat{\theta}(x) = \text{E}(\theta|x) = \frac{me^{(n-\lambda)m}}{e^{(n-\lambda)m}-1} - \frac{1}{n-\lambda}$ for $\lambda \neq n$ and $\widehat{\theta}(x) = m/2$ otherwise.

3. For large n, $\widehat{\theta}(x) \to m$, which is the MLE.

4. Since the minimum of n i.i.d. exponential random variables with parameter 1 is an exponential random variable with parameter n (easy to check), the mean and variance of $\widehat{\theta}_{\text{MLE}}(X) = \min(X_1, \dots, X_n)$ are $\theta + \frac{1}{n}$ and $\frac{1}{n^2}$.
Bias $b(\theta, \widehat{\theta}_{\text{MLE}}) = \frac{1}{n}$.
Variance $\text{var}_\theta(\widehat{\theta}_{\text{MLE}}) = \frac{1}{n^2}$.
Quadratic risk $R(\theta, \widehat{\theta}_{\text{MLE}}) = \frac{2}{n^2}$, independent of θ.
Bayes risk is the same, $r(\widehat{\theta}_{\text{MLE}}) = \frac{2}{n^2}$.

Exercise 6 (Exponential model – mean)**:** 1. Model $p(x|\theta) = \prod_{i=1}^n \theta e^{-\theta x_i}$
Prior $\pi(\theta) \propto e^{-\lambda\theta}$
Posterior $\pi(\theta|x) \propto \theta^n e^{-\theta(S+\lambda)}$ with $S = \sum_{i=1}^n x_i$
this is a Gamma distribution with parameters $(S+\lambda, n+1)$.

2. $\widehat{g}(x) = \mathrm{E}(g(\theta)|x) = \frac{S+\lambda}{n}$

3. $\mathrm{E}_\theta(\widehat{g}(X)) = \frac{1}{\theta} + \frac{\lambda}{n}$
Bias $b(\theta, \widehat{\theta}) = \frac{\lambda}{n}$
Variance $\mathrm{var}_\theta(\widehat{g}(X)) = \frac{1}{n\theta^2}$.
Quadratic risk $R(\theta, \widehat{\theta}) = \frac{1}{n}(\lambda + \frac{1}{\theta^2})$.
Infinite Bayes risk since $\mathrm{E}(\frac{1}{\theta^2}) = +\infty$ for θ exponential.

Exercise 7 (Poisson model – Jeffreys prior)**:** 1. Since $I(\theta) = \frac{1}{\theta}$ we get $\pi(\theta) \propto \frac{1}{\sqrt{\theta}}$.
This is *not* a probability measure!

2. Posterior $\pi(\theta|x) \propto \frac{1}{\sqrt{\theta}}\theta^S e^{-n\theta}$ with $S = \sum_{i=1}^n x_i$
this is a Gamma distribution with parameters $(n, S + \frac{1}{2})$.

3. $\widehat{\theta}(x) = \mathrm{E}(\theta|x) = \frac{1}{n}(S + \frac{1}{2})$. Here no parameter needed.

4. Bias $b(\theta, \widehat{\theta}) = \frac{1}{2n}$
Variance $\mathrm{var}_\theta(\widehat{\theta}(X)) = \frac{\theta}{n}$
Quadratic risk, $R(\theta, \widehat{\theta}) = b(\theta, \widehat{\theta})^2 + \mathrm{var}_\theta(\widehat{\theta}(X)) = \frac{1}{n}(\frac{1}{4n} + \theta)$

Exercise 8 (Gaussian vector)**:** 1. Model $p(x|\theta) \propto \prod_{i=1}^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\theta)^T\Gamma^{-1}(x_i-\theta)}$
Prior $\pi(\theta) \propto e^{-\frac{1}{2}\theta^T\theta}$
Posterior $\pi(\theta|x) \propto e^{-\frac{1}{2}\theta^T\theta} e\left(-\frac{1}{2}(n\theta^T\Gamma^{-1}\theta - 2\theta^T\Gamma^{-1}S)\right)$ with $S = \sum_{i=1}^n x_i$
Gaussian with mean $(\Gamma + nI)^{-1}S$ and covariance matrix $(I + n\Gamma^{-1})^{-1}$

2. $\widehat{g}(x) = \mathrm{E}(g(\theta)|x) = w^T(\Gamma + nI)^{-1}S$