ST308 Bayesian Inference

Kostas Kalogeropoulos (Office B610)

Week 2: Exercises

1. Consider the vaccination example in the lecture slides.

   (a) Assume that a person is tested positive for immunity. Which of the decision rules have the lower posterior risk?

   (b) Repeat the above for the case that the person was tested negative.

   (c) Combine the two above cases and choose the optimal decision rule. Compare with the Bayes risk outcome.

Answer:

1(a) The prior distribution is

$$\pi(\theta) = \begin{cases} 0.6 & \text{if } \theta = \theta_1 \\ 0.4 & \text{if } \theta = \theta_2 \end{cases}$$

Since a person is being tested positive $x = x_1$. Hence using Bayes theorem we can find the posterior probability for $\theta = \theta_1$

$$\pi(\theta_1|x_1) = \frac{f(x_1|\theta_1)\pi(\theta_1)}{f(x_1|\theta_1)\pi(\theta_1) + f(x_1|\theta_2)\pi(\theta_2)} = \frac{0.65 \times 0.6}{0.65 \times 0.6 + 0.25 \times 0.4} = 39/49,$$

hence $\pi(\theta_2|x_1) = 1 - \pi(\theta_1|x_1) = 10/49$. The posterior risk for the strategy $\delta_1(x)$ is

$$\begin{aligned} \rho(\delta_1(x_1), \pi(\theta)) &= E_{\theta|x=x_1}(L(\delta_1(x_1), \theta)) = \sum_{i=1}^{2} L(\delta_1(x_1), \theta_i)\pi(\theta_i|x_1) \\ &= L(\delta_1(x_1), \theta_1)\pi(\theta_1|x_1) + L(\delta_1(x_1), \theta_2)\pi(\theta_2|x_1) \\ &= 8 \times \frac{39}{49} + 0 \times \frac{10}{49} = 312/49 \end{aligned}$$

Since $\delta_1(x_1) = \delta_2(x_1)$, $\rho(\delta_1(x_1), \pi(\theta)) = \rho(\delta_2(x_1), \pi(\theta)) = \frac{312}{49}$. The posterior risk for $\delta_3(x_1)$ (and obviously for $\delta_4(x_1)$ since $\delta_3(x_1) = \delta_4(x_1)$) is found in a similar way

$$
\begin{aligned}
\rho(\delta_3(x_1), \pi(\theta)) &= E_{\theta|x=x_1}(L(\delta_3(x_1), \theta)) = \sum_{i=1}^{2} L(\delta_3(x_1), \theta_i)\pi(\theta_i|x_1) \\
&= L(\delta_3(x_1), \theta_1)\pi(\theta_1|x_1) + L(\delta_3(x_1), \theta_2)\pi(\theta_2|x_1) \\
&= 0 \times \frac{39}{49} + 20 \times \frac{10}{49} = 200/49
\end{aligned}
$$

Hence the posterior risk is minimized if actions δ_3 or δ_4 are chosen.

1(b) If a person is tested negative, $x = x_2$ and the posterior probability for $\theta = \theta_1$ is

$$
\pi(\theta_1|x_2) = \frac{f(x_2|\theta_1)\pi(\theta_1)}{f(x_2|\theta_1)\pi(\theta_1) + f(x_2|\theta_2)\pi(\theta_2)} = \frac{0.35 \times 0.6}{0.35 \times 0.6 + 0.75 \times 0.4} = 21/51,
$$

hence $\pi(\theta_2|x_2) = 1 - \pi(\theta_1|x_2) = 30/51$. The posterior risk for the strategy $\delta_1(x)$ is

$$
\begin{aligned}
\rho(\delta_1(x_2), \pi(\theta)) &= E_{\theta|x=x_2}(L(\delta_1(x_2), \theta)) = \sum_{i=1}^{2} L(\delta_1(x_2), \theta_i)\pi(\theta_i|x_2) \\
&= L(\delta_1(x_2), \theta_1)\pi(\theta_1|x_2) + L(\delta_1(x_2), \theta_2)\pi(\theta_2|x_2) \\
&= 8 \times \frac{21}{51} + 0 \times \frac{30}{51} = 168/51
\end{aligned}
$$

Since $\delta_1(x_2) = \delta_3(x_2)$, $\rho(\delta_1(x_2), \pi(\theta)) = \rho(\delta_3(x_2), \pi(\theta)) = \frac{168}{51}$. The posterior risk for $\delta_2(x_2)$ and $\delta_4(x_2)$ (since $\delta_2(x_2) = \delta_4(x_2)$) is

$$
\begin{aligned}
\rho(\delta_2(x_2), \pi(\theta)) &= E_{\theta|x=x_1}(L(\delta_2(x_2), \theta)) = \sum_{i=1}^{2} L(\delta_2(x_2), \theta_i)\pi(\theta_i|x_2) \\
&= L(\delta_2(x_2), \theta_1)\pi(\theta_1|x_2) + L(\delta_2(x_2), \theta_2)\pi(\theta_2|x_2) \\
&= 0 \times \frac{21}{51} + 20 \times \frac{30}{51} = 600/51
\end{aligned}
$$

Hence the posterior risk is minimized if actions δ_1 or δ_3 are chosen.

Combining the two possible cases $x = x_1$ or $x = x_2$ we deduce that $\delta_3(x)$ is the optimal decision rule (strategy), which is the same conclusion as with Bayes risk criterion. This is not surprising since

$$
r(\delta(x), \pi(\theta)) = E_X\left[\rho(\delta(x), \pi(\theta))\right].
$$

Note that

$$m(x_1) = f(x_1|\theta_1)\pi(\theta_1) + f(x_1|\theta_2)\pi(\theta_2) = 0.65 \times 0.6 + 0.25 \times 0.4 = 0.49,$$
$$m(x_2) = f(x_2|\theta_1)\pi(\theta_1) + f(x_2|\theta_2)\pi(\theta_2) = 0.35 \times 0.6 + 0.75 \times 0.4 = 0.51,$$

$$\begin{aligned} r(\delta_3(x), \pi(\theta)) = E_X\left[\rho(\delta_3(x), \pi(\theta))\right] &= \sum_{i=1}^{2} \rho(\delta_3(x_i), \theta) m(x_i) \\ &= \frac{200}{49} 0.49 + \frac{168}{51} 0.51 = 3.68 \end{aligned}$$

exactly the same number as calculated in the lecture slides of week 4.

2. Consider the quadratic error, absolute error and $0-1$ loss functions. Find the Bayes estimator for θ in the case of
   (a) A random sample $x = (x_1, ..., x_n)$ from a Normal(θ, 1). Assign a N(μ, τ^2) prior to θ.
   (b) A single observation x from a Binomial(n, θ). Assign a Beta(α, β) prior to θ.

Answer: As was proven in class the Bayes estimator for the quadratic error, absolute error and $0-1$ error functions is the posterior mean, posterior median and posterior mode respectively.

2(a) As shown in class the posterior distribution in this case is

$$\mathrm{N}\left(\frac{\frac{1}{n}\mu + \tau^2 \bar{x}}{\tau^2 + \frac{1}{n}}, \frac{\tau^2 \frac{1}{n}}{\tau^2 + \frac{1}{n}}\right)$$

Since the mean, median and mode are all equal in Normal distribution the Bayes estimator is the following for all 3 loss functions:

$$\frac{\bar{x}\tau^2 + \mu\frac{1}{n}}{(\frac{1}{n} + \tau^2)}$$

2(b) As shown in class, the posterior distribution for θ in this case is

$$\mathrm{Beta}(\alpha + x, n - x + \beta)$$

The posterior mean is the Bayes estimator for θ in the case of quadratic error loss function and corresponds to

$$\frac{x + \alpha}{n + \alpha + \beta}$$

In the case of the absolute error loss function the Bayes estimator for θ is the posterior median which has no closed form and has to be calculated numerically.

Finally for the $0-1$ loss function we get the posterior mode as the Bayes estimator for θ, which is equal to

$$\frac{x+\alpha-1}{n+\alpha+\beta-2}$$

3. Show that the bayes risk $r(\delta(x), \pi(\theta))$ can be written as averaging the posterior risk over x. In other words show that

$$\int R(\delta(x), \theta)\pi(\theta)d\theta = \int \rho(\delta(x), \pi(\theta))m(x)dx,$$

assuming certain regularity conditions under which

$$\int_\Theta \int_\mathcal{X} L(\delta(x), \theta)h(x, \theta)dxd\theta = \int_\mathcal{X} \int_\Theta L(\delta(x), \theta)\pi(\theta|x)m(x)d\theta dx.$$

Answer: We can write

$$\begin{aligned}
r(\delta(x), \pi(\theta)) &= \int_\Theta R(\delta(x), \theta)\pi(\theta)d\theta = \int_\Theta \int_\mathcal{X} L(\delta(x), \theta)f(x|\theta)\pi(\theta)dxd\theta \\
&= \int_\Theta \int_\mathcal{X} L(\delta(x), \theta)h(x, \theta)dxd\theta = \int_\mathcal{X} \int_\Theta L(\delta(x), \theta)h(x, \theta)d\theta dx \\
&= \int_\mathcal{X} \int_\Theta L(\delta(x), \theta)\pi(\theta|x)m(x)d\theta dx = \int_\mathcal{X} \int_\Theta L(\delta(x), \theta)\pi(\theta|x)d\theta m(x)dx \\
&= \int_\mathcal{X} \rho(\delta(x))m(x)dx
\end{aligned}$$

4. **Optional Exercise:** Let $x = (x_1, \ldots, x_n$ be a random sample from a Poisson(λ) distribution. Assign a Gamma(α, β) prior to λ. Consider the LINEX (LINear-EXponential) loss function.

$$L(a, \lambda) = \exp(k(a-\lambda)) - k(a-\lambda) - 1$$

where k is a known positive constant. Find the Bayes estimator λ.

Answer: As shown in class, the posterior distribution for λ in this case is

$$\text{Gamma}(\alpha + \sum x_i, n+\beta)$$

The Bayes estimator is the one that minimises the posterior risk which is equal to

$$\begin{aligned}
\rho(a, \pi(\lambda)) &= \int_\Lambda L(a, \lambda)\pi(\lambda|x)d\lambda \\
&= \int_\Lambda \left(e^{k(a-\lambda)} - k(a-\lambda) - 1\right)\pi(\lambda|x)d\lambda
\end{aligned}$$

To minimise with respect to a we first set the derivative to 0 and solve

$$
\begin{aligned}
\frac{\partial}{\partial a}\rho(a,\pi(\lambda)) = 0 &\rightarrow \int_\Lambda \frac{\partial}{\partial a}\left(e^{k(a-\lambda)} - k(a-\lambda) - 1\right)\pi(\lambda|x)d\lambda = 0 \\
&\rightarrow \int_\Lambda (ke^{k(a-\lambda)} - k)\pi(\lambda|x)d\lambda = 0 \\
&\rightarrow k\left(\int_\Lambda e^{ka-k\lambda}\pi(\lambda|x)d\lambda - \int_\Lambda k\pi(\lambda|x)d\lambda\right) = 0 \\
&\rightarrow e^{ka}\int_\Lambda (e^{-k\lambda}\pi(\lambda|x)d\lambda = \int_\Lambda \pi(\lambda|x)d\lambda \\
&\rightarrow e^{ka}E_{\lambda|x}\left(e^{-k\lambda}\right) = 1 \\
&\rightarrow e^{-ka} = E_{\lambda|x}\left(e^{-k\lambda}\right), \\
&\rightarrow a = -\frac{\log E_{\lambda|x}\left(e^{-k\lambda}\right)}{k}
\end{aligned}
$$

Second, we check if the second derivative is positive

$$
\frac{\partial^2}{\partial a^2}\rho(a,\pi(\lambda)) = \int_\Lambda k^2 e^{k(a-\lambda)}\pi(\lambda|x)d\lambda = k^2 e^{ka}E_{\lambda|x}\left(e^{-k\lambda}\right) > 0
$$

since $e^{-k\lambda} > 0$. In order to find the Bayes estimator for this loss function we first need to find $E_{\lambda|x}\left(e^{-k\lambda}\right)$. In the general case of Gamma(A,B) we get

$$
\begin{aligned}
E_{\lambda|x}\left(e^{-k\lambda}\right) &= \int_0^\infty e^{-k\lambda}\frac{B^A}{\Gamma(A)}\lambda^{A-1}e^{-B\lambda}d\lambda = \frac{B^A}{\Gamma(A)}\int_0^\infty \lambda^{A-1}e^{-(B+k)\lambda}d\lambda \\
&= \frac{B^A}{\Gamma(A)}\frac{\Gamma(A)}{(B+k)^A} = \left(\frac{B}{B+k}\right)^A
\end{aligned}
$$

and

$$
-\frac{1}{k}\log E_{\lambda|x}\left(e^{-k\lambda}\right) = \frac{A}{k}(\log(B+k) - \log(B))
$$

In our case $A = \alpha + \sum x_i$ and $B = n + \beta$, so we get that the Bayes estimator is

$$
\delta(x) = \frac{\alpha + \sum x_i}{k}(\log(n+\beta+k) - \log(n+\beta))
$$