

Solutions to Tutorial Week 9

STAT3023: Statistical Inference

Semester 2, 2022

Lecturers: Linh Nghiem, Michael Stewart and Rachel Wang

1. For the simple prediction problem where Y has a strictly increasing, continuous CDF $F(\cdot)$ and $\mu = E(Y)$ exists and is finite and the decision space is $\mathcal{D} = \mathbb{R}$, determine the decision d that minimises the risk

$$R(d) = E[L(d|Y)]$$

for the asymmetric piecewise-linear loss function given by

$$L(d|y) = \begin{cases} p(y-d) & \text{for } d < y, \\ (1-p)(d-y) & \text{for } d > y \end{cases}$$

and some $0 < p < 1$ (**hint**: we have already seen the case $p = 0.5$).

Solution: Write $f(y) = \frac{d}{dy}F(y)$ for the density (PDF) of Y . Then

$$\begin{aligned} R(d) &= (1-p) \int_{-\infty}^d (d-y)f(y) dy + p \int_d^{\infty} (y-d)f(y) dy \\ &= (1-p)dF(d) - (1-p) \int_{-\infty}^d yf(y) dy + p \int_d^{\infty} yf(y) dy - pd[1-F(d)] \\ &= (1-p)dF(d) - pd[1-F(d)] + p\mu - [(1-p)+p] \int_{-\infty}^d yf(y) dy \\ &= dF(d) - pd + p\mu - \int_{-\infty}^d yf(y) dy. \end{aligned}$$

The derivative is

$$R'(d) = df(d) + F(d) - p - [df(d)] = F(d) - p.$$

This is negative (so $R(d)$ decreases) for d such that $F(d) < p$. $R(d)$ increases for d such that $F(d) > p$; it is thus minimised at $d = F^{-1}(p)$, the p -th quantile of $F(\cdot)$; the case $p = 0.5$ (which we saw in lectures) gives the “population median”.

2. Determine the optimal decision $d \in \mathcal{D} = \mathbb{R}$ for the simple prediction problem where Y has a continuous distribution on $(0, \infty)$ with density $f(\cdot)$ satisfying

- $f(x) = 0$ for $x \leq 0$;
- $f(x) > 0$ and decreasing in x for $x > 0$

and the loss function $L(d|y)$ is given by

$$L(d|y) = \begin{cases} 0 & \text{if } |d-y| \leq C \\ 1 & \text{if } |d-y| > C, \end{cases}$$

for some known $0 < C < \infty$.

Solution: The risk is simply the *non-coverage* probability of the “prediction interval” $d \pm C$. The optimal choice is $d = C$, yielding the prediction interval $[0, 2C]$. To see why, note that the risk is

$$\begin{aligned} R(d) &= 1 - P\{|d-Y| \leq C\} \\ &= 1 - P(d-C \leq Y \leq d+C) \\ &= P(Y < d-C) + P(Y > d+C) \\ &= F(d-C) + 1 - F(d+C) \end{aligned}$$

The derivative is then

$$R'(d) = f(d - C) - f(d + C).$$

So long as $0 < d - C < d + C < \infty$, i.e. $d > C$, this difference is positive (due to the fact that $f(\cdot)$ is decreasing) and so $R(d)$ increases in $d > C$. For $-C < d < C$, the difference is negative (the first term $f(d - C)$ is zero in that case, the second term $f(d + C)$ is positive) and so $R(d)$ decreases in $-C < d < C$. For $d \leq -C$, both terms are zero and so the risk is constant. The risk is therefore minimised at $d = C$.

3. Suppose $Z \sim N(0, 1)$.

(a) Show that for any constant c ,

$$E\{|c + Z|\} = c[1 - 2\Phi(-c)] + \frac{2e^{-\frac{1}{2}c^2}}{\sqrt{2\pi}}.$$

where $\Phi(\cdot)$ is the cdf of $N(0, 1)$.

Solution:

$$\begin{aligned} \int_{-\infty}^{\infty} |c + z| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz &= \int_{-\infty}^{-c} [-(c + z)] \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz + \int_{-c}^{\infty} (c + z) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= -c\Phi(-c) - \int_{-\infty}^{-c} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &\quad + c[1 - \Phi(-c)] + \int_{-c}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= c[1 - 2\Phi(-c)] + \frac{1}{\sqrt{2\pi}} \left\{ \left[-e^{-\frac{1}{2}z^2} \right]_{-c}^{\infty} - \left[-e^{-\frac{1}{2}z^2} \right]_{-\infty}^{-c} \right\} \\ &= c[1 - 2\Phi(-c)] + \frac{2e^{-\frac{1}{2}c^2}}{\sqrt{2\pi}}. \end{aligned}$$

(b) Suppose $c_n \rightarrow 0$ as $n \rightarrow \infty$. Determine $\lim_{n \rightarrow \infty} E\{|c_n + Z|\}$.

Solution: As $c_n \rightarrow 0$,

- $c_n [1 - 2\Phi(-c_n)] \rightarrow 0$;
- $e^{-\frac{1}{2}c_n^2} \rightarrow 1$.

So

$$\lim_{n \rightarrow \infty} E\{|c_n + Z|\} = \lim_{n \rightarrow \infty} \left\{ c_n [1 - 2\Phi(-c_n)] + \frac{2e^{-\frac{1}{2}c_n^2}}{\sqrt{2\pi}} \right\} = 0 + \frac{2}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}}.$$

4. Suppose $\mathbf{X} = (X_1, \dots, X_n)$ consists of iid $N(\theta, 1)$ random variables and that it is desired to determine Bayes procedures using the weight function/prior is given by $w(\theta) \equiv 1$ (the “flat prior”). Show that the resultant posterior density is the normal density with mean $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and variance $\frac{1}{n}$.

Solution: The likelihood is

$$\begin{aligned} f_{\theta}(\mathbf{X}) &= \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_i - \theta)^2} \right] \\ &= (2\pi)^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n (X_i - \theta)^2} \\ &= (2\pi)^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n X_i^2 + \theta \sum_{i=1}^n X_i - \frac{n\theta^2}{2}} \\ &= (2\pi)^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n X_i^2 + \frac{n\bar{X}^2}{2} - \frac{n\bar{X}^2}{2} + \theta \sum_{i=1}^n X_i - \frac{n\theta^2}{2}} \\ &= (2\pi)^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n (X_i - \bar{X})^2} e^{-\frac{n\bar{X}^2}{2} + \theta \sum_{i=1}^n X_i - \frac{n\theta^2}{2}} \\ &= (2\pi)^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n (X_i - \bar{X})^2} e^{-\frac{n}{2}(\theta - \bar{X})^2} \\ &= n^{-1/2} (2\pi)^{-(n-1)/2} e^{-\frac{1}{2} \sum_{i=1}^n (X_i - \bar{X})^2} \underbrace{\frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n}{2}(\theta - \bar{X})^2}}_{\text{integrates to 1}}. \end{aligned}$$

As a function of θ , this is the $N(\bar{X}, \frac{1}{n})$ density, times a “constant” (another expression involving the X_i s and n , but not θ). Thus multiplying by the weight function/prior $w(\theta) \equiv 1$ and then “integrating out” θ gives

$$m(\mathbf{X}) = \int_{-\infty}^{\infty} w(\theta) f_{\theta}(\mathbf{X}) d\theta = n^{-1/2} (2\pi)^{-(n-1)/2} e^{-\frac{1}{2} \sum_{i=1}^n (X_i - \bar{X})^2}$$

and thus the posterior density is

$$p(\theta|\mathbf{X}) = \frac{w(\theta) f_{\theta}(\mathbf{X})}{m(\mathbf{X})} = \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n}{2}(\theta - \bar{X})^2}.$$

5. Suppose $\mathbf{X} = (X_1, \dots, X_n)$ consists of iid $N(\theta, 1)$ random variables and that it is desired to determine Bayes procedures using the weight function/prior $w(\cdot)$ given by the $N(\mu_0, \sigma_0^2)$ density, that is

$$w(\theta) = \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(\theta - \mu_0)^2}.$$

Show that the resultant posterior density is the normal density with mean

$$\left(\frac{1}{1 + n\sigma_0^2} \right) \mu_0 + \left(\frac{n\sigma_0^2}{1 + n\sigma_0^2} \right) \bar{X}$$

and variance

$$\frac{\sigma_0^2}{1 + n\sigma_0^2}.$$

Solution: The product of the weight function (prior) and the likelihood is (writing “const.” for an expression not involving θ),

$$\begin{aligned} w(\theta) f_{\theta}(\mathbf{X}) &= \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(\theta - \mu_0)^2} (2\pi)^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n (X_i - \theta)^2} \\ &= \text{const.} \exp \left\{ -\frac{\theta^2}{2\sigma_0^2} + \frac{\theta\mu_0}{\sigma_0^2} + \theta \sum_{i=1}^n X_i - \frac{n\theta^2}{2} \right\} \\ &= \text{const.} \exp \left\{ -\frac{\theta^2}{2} \left[\frac{1 + n\sigma_0^2}{\sigma_0^2} \right] + \theta \left[\frac{\mu_0 + n\bar{X}\sigma_0^2}{\sigma_0^2} \right] \right\} \\ &= \text{const.} \exp \left\{ -\frac{1}{2} \left[\frac{1 + n\sigma_0^2}{\sigma_0^2} \right] \left(\theta^2 - 2\theta \left[\frac{\mu_0 + n\bar{X}\sigma_0^2}{1 + n\sigma_0^2} \right] \right) \right\} \\ &= \text{const.} \exp \left\{ -\frac{1}{2} \left[\frac{1 + n\sigma_0^2}{\sigma_0^2} \right] \left(\theta^2 - 2\theta \left[\frac{\mu_0 + n\bar{X}\sigma_0^2}{1 + n\sigma_0^2} \right] + \left[\frac{\mu_0 + n\bar{X}\sigma_0^2}{1 + n\sigma_0^2} \right]^2 \right) \right\} \\ &= \text{const.} \exp \left\{ -\frac{1}{2} \left[\frac{1 + n\sigma_0^2}{\sigma_0^2} \right] \left(\theta - \left[\frac{\mu_0 + n\bar{X}\sigma_0^2}{1 + n\sigma_0^2} \right] \right)^2 \right\} \end{aligned}$$

which is a constant multiple of the desired normal density, so when renormalised that normal density becomes the posterior density.

6. Suppose $\mathbf{X} = (X_1, \dots, X_n)$ consists of iid $N(\theta, 1)$ random variables. We are interested in finding Bayes decisions/procedures for various loss functions using each of the two weight functions/priors used in questions 4 and 5 above: the “flat prior” and the “normal prior” respectively.

- (a) When the loss function is $L(d|\theta) = (d - \theta)^2$, the Bayes procedure in each case is the posterior mean. Determine for both decisions $d(\cdot)$,
- (i) the risk $R(\theta|d) = E_{\theta} [L(d(\mathbf{X})|\theta)]$;

Solution: When using the flat prior, $d(\mathbf{X}) = \bar{X}$, the risk is just $\text{Var}_{\theta}(\bar{X}) = \frac{1}{n}$.

For the normal prior,

$$d(\mathbf{X}) = \left(\frac{1}{1 + n\sigma_0^2} \right) \mu_0 + \left(\frac{n\sigma_0^2}{1 + n\sigma_0^2} \right) \bar{X},$$

since the risk is the mean-squared error (MSE) we use the identity

$$\text{MSE} = \text{Var} + (\text{Bias})^2.$$

The bias is

$$E_\theta [d(\mathbf{X})] - \theta = \frac{\mu_0 + n\sigma_0^2\theta - [1 + n\sigma_0^2]\theta}{1 + n\sigma_0^2} = \frac{\mu_0 - \theta}{1 + n\sigma_0^2},$$

and the variance is

$$\text{Var}_\theta [d(\mathbf{X})] = \left(\frac{n\sigma_0^2}{1 + n\sigma_0^2} \right)^2 \text{Var}_\theta (\bar{X}) = \left(\frac{n\sigma_0^2}{1 + n\sigma_0^2} \right)^2 \frac{1}{n} = \frac{n\sigma_0^4}{(1 + n\sigma_0^2)^2}.$$

Thus the MSE is

$$\frac{n\sigma_0^4 + (\mu_0 - \theta)^2}{(1 + n\sigma_0^2)^2}.$$

- (ii) the *limiting* risk $\lim_{n \rightarrow \infty} nE_\theta [L(d(\mathbf{X})|\theta)]$.

Solution: Multiplying both risks by n gives, for $d(\mathbf{X}) = \bar{X}$:

$$nE_\theta [(\bar{X} - \theta)^2] \equiv 1.$$

For the second estimator we get

$$n \left[\frac{n\sigma_0^4 + (\mu_0 - \theta)^2}{(1 + n\sigma_0^2)^2} \right] = \frac{n^2\sigma_0^4 \left[1 + \frac{(\mu_0 - \theta)^2}{n\sigma_0^4} \right]}{n^2\sigma_0^4 \left(\frac{1}{n\sigma_0^2} + 1 \right)^2} = \frac{1 + \frac{(\mu_0 - \theta)^2}{n\sigma_0^4}}{\left(\frac{1}{n\sigma_0^2} + 1 \right)^2} \rightarrow 1,$$

since both numerator and denominator tend to 1. *Note that both procedures have the same limiting risk. Also they don't depend on θ .*

- (b) When the loss function is $L(d|\theta) = |d - \theta|$, the Bayes procedure in each case is the posterior median. Determine for both decisions $d(\cdot)$

- (i) the risk $R(\theta|d) = E_\theta [L(d(\mathbf{X})|\theta)]$;

Solution: For the flat prior, $d(\mathbf{X}) = \bar{X}$, and the risk is $E_\theta (|\bar{X} - \theta|)$. Since $\bar{X} \sim N(\theta, \frac{1}{n})$, $\sqrt{n}(\bar{X} - \theta) \sim N(0, 1)$. Therefore using question 3 with $c = 0$, the risk is

$$E_\theta (|\bar{X} - \theta|) = \frac{1}{\sqrt{n}} E_\theta (\sqrt{n} |\bar{X} - \theta|) = \frac{1}{\sqrt{n}} \sqrt{\frac{2}{\pi}}.$$

For the normal prior,

$$d(\mathbf{X}) = \frac{n\sigma_0^2 \bar{X} + \mu_0}{1 + n\sigma_0^2},$$

and the risk is

$$\begin{aligned} E_\theta \left[\left| \frac{\mu_0 + n\sigma_0^2 \bar{X} - [1 + n\sigma_0^2] \theta}{1 + n\sigma_0^2} \right| \right] &= \frac{E_\theta |(\mu_0 - \theta) + n\sigma_0^2(\bar{X} - \theta)|}{1 + n\sigma_0^2} \\ &= \frac{\sigma_0^2 \sqrt{n}}{1 + n\sigma_0^2} E_\theta \left\{ \left| \frac{\mu_0 - \theta}{\sigma_0^2 \sqrt{n}} + \sqrt{n}(\bar{X} - \theta) \right| \right\}. \end{aligned}$$

This is of the form $k_n E(|c_n + Z|)$ for $Z \sim N(0, 1)$ and so by question 3 this is

$$k_n \left\{ c_n [1 - 2\Phi(-c_n)] + \frac{2e^{-\frac{1}{2}c_n^2}}{\sqrt{2\pi}} \right\}$$

where

$$k_n = \frac{\sigma_0^2 \sqrt{n}}{1 + n\sigma_0^2}$$

$$c_n = \frac{\mu_0 - \theta}{\sigma_0^2 \sqrt{n}}.$$

- (ii) the *limiting* risk $\lim_{n \rightarrow \infty} \sqrt{n} E_\theta [L(d(\mathbf{X})|\theta)]$.

Solution: For $d(\mathbf{X}) = \bar{X}$,

$$\sqrt{n} E_\theta \{|\bar{X} - \theta|\} = \sqrt{\frac{2}{\pi}}.$$

For the second case, since

$$\sqrt{n} k_n = \frac{n\sigma_0^2}{1 + n\sigma_0^2} \rightarrow 1 \quad \text{and}$$

$$c_n = \frac{\mu_0 - \theta}{\sigma_0^2 \sqrt{n}} \rightarrow 0$$

using part (b) of question 3, the limiting risk

$$\sqrt{n} E_\theta \{|d(\mathbf{X}) - \theta|\} = \sqrt{\frac{2}{\pi}}.$$

Again, both procedures have the same limiting risk, which do not depend on θ !

Hint: in each case write the risk in the form $k_n E_\theta \{|c_n + \sqrt{n}(\bar{X} - \theta)|\}$ for sequences $\{k_n\}$ and $\{c_n\}$ and use question 3 above.

- (c) When the loss function is $L(d|\theta) = 1 \{|d - \theta| > C/\sqrt{n}\}$ the Bayes procedure in each case is the level set of the posterior density of width $\frac{2C}{\sqrt{n}}$. Because the posterior density is symmetric about the posterior mean/median (and unimodal) in each case, this is simply of the form

$$\text{posterior mean} \pm \frac{C}{\sqrt{n}}.$$

Determine for both decisions $d(\cdot)$

- (i) the risk $R(\theta|d) = E_\theta [L(d(\mathbf{X})|\theta)]$;

Solution: The risk is the probability of non-coverage. For either choice of $d(\mathbf{X})$ this can be written as

$$E_\theta [L(d(\mathbf{X})|\theta)] = 1 - P_\theta \{|d(\mathbf{X}) - \theta| \leq C/\sqrt{n}\}$$

$$= P_\theta \left\{ d(\mathbf{X}) < \theta - \frac{C}{\sqrt{n}} \right\} + P_\theta \left\{ d(\mathbf{X}) > \theta + \frac{C}{\sqrt{n}} \right\}. \quad (1)$$

When $d(\mathbf{X}) = \bar{X}$, by symmetry this can be written as

$$2P_\theta \left\{ \bar{X} > \theta + \frac{C}{\sqrt{n}} \right\} = 2P_\theta \{ \sqrt{n}(\bar{X} - \theta) > C \}$$

$$= 2[1 - \Phi(C)].$$

Since this doesn't depend on n , it is also the limiting risk.

In a similar way, for the second choice

$$d(\mathbf{X}) = \left(\frac{1}{1 + n\sigma_0^2} \right) \mu_0 + \left(\frac{n\sigma_0^2}{1 + n\sigma_0^2} \right) \bar{X},$$

we have

$$d(\mathbf{X}) - \theta = \frac{\mu_0 - \theta + n\sigma_0^2(\bar{X} - \theta)}{1 + n\sigma_0^2}$$

Then we can write the first probability in (1) as

$$\begin{aligned} P_\theta \left\{ d(\mathbf{X}) - \theta < -\frac{C}{\sqrt{n}} \right\} &= P_\theta \left\{ n\sigma_0^2 (\bar{X} - \theta) < -\frac{C}{\sqrt{n}} (1 + n\sigma_0^2) - (\mu_0 - \theta) \right\} \\ &= P_\theta \left\{ \sqrt{n} (\bar{X} - \theta) < -C - \frac{C}{n\sigma_0^2} - \frac{\mu - \theta}{\sigma_0^2 \sqrt{n}} \right\}. \end{aligned}$$

In a similar way the second probability in (1) may be written as

$$P_\theta \left\{ \sqrt{n} (\bar{X} - \theta) > C + \frac{C}{n\sigma_0^2} - \frac{\mu - \theta}{\sigma_0^2 \sqrt{n}} \right\}.$$

Thus the *exact* risk is

$$\begin{aligned} &\Phi \left(-C - \frac{C}{n\sigma_0^2} - \frac{\mu - \theta}{\sigma_0^2 \sqrt{n}} \right) + 1 - \Phi \left(C + \frac{C}{n\sigma_0^2} - \frac{\mu - \theta}{\sigma_0^2 \sqrt{n}} \right) \\ &= \left[1 - \Phi \left(C + \frac{C}{n\sigma_0^2} + \frac{\mu - \theta}{\sigma_0^2 \sqrt{n}} \right) \right] + \left[1 - \Phi \left(C + \frac{C}{n\sigma_0^2} - \frac{\mu - \theta}{\sigma_0^2 \sqrt{n}} \right) \right] \end{aligned}$$

(ii) the *limiting* risk $\lim_{n \rightarrow \infty} E_\theta [L(d(\mathbf{X})|\theta)]$.

Solution: See above for $d(\mathbf{X}) = \bar{X}$. For the second choice note that since

$$C + \frac{C}{n\sigma_0^2} \pm \frac{\mu - \theta}{\sigma_0^2 \sqrt{n}} \rightarrow C$$

as $n \rightarrow \infty$, the limiting risk in the second case is also $2[1 - \Phi(C)]$; so *again* both methods have the same *limiting* risks which are free of θ .