The University of Sydney
School of Mathematics and Statistics

Solutions to Tutorial Week 9

STAT3023: Statistical Inference — Semester 2, 2023

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

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

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

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

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

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

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

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

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

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 $d \le -C$, 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,

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

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

Solution:

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

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

Solution: As $c_n \to 0$,

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

So

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

4. Suppose $\boldsymbol{X} = (X_1,\ldots,X_n)$ 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

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

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

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

and thus the posterior density is

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

5. Suppose $\boldsymbol{X} = (X_1, \ldots, X_n)$ 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),

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

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

6. Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ consists of iid $N(\theta, 1)$ random varibles. 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 $R(\theta|d) = E_\theta\left[L\left(d(\mathbf{X})|\theta\right)\right]$;

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

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

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

and the variance is

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

Thus the MSE is

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

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

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

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

For the second estimator we get

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

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 $R(\theta|d) = E_\theta\left[L\left(d(\mathbf{X})|\theta\right)\right]$;

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

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

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

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

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

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

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 $\lim_{n\to\infty}\sqrt{n}E_\theta\left[L\left(d(\mathbf{X})|\theta\right)\right]$.
Solution: For $d(\mathbf{X})=\bar{X}$,

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

For the second case, since

$$\sqrt{n}k_n = \frac{n\sigma_0^2}{1+n\sigma_0^2} \to 1 \quad \text{and}$$
$$c_n = \frac{\mu_0-\theta}{\sigma_0^2\sqrt{n}} \to 0$$

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

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

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

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

(c) When the loss function is $L(d|\theta)=1\left\{|d-\theta|>C/\sqrt{n}\right\}$ 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 $R(\theta|d)=E_\theta\left[L\left(d(\mathbf{X})|\theta\right)\right]$;
Solution: The risk is the probability of non-coverage. For either choice of $d(\mathbf{X})$ this can be written as

$$\begin{aligned} E_\theta\left[L\left(d(\mathbf{X})|\theta\right)\right] &= 1-P_\theta\left\{|d(\mathbf{X})-\theta|\le C/\sqrt{n}\right\} \\ &= P_\theta\left\{d(\mathbf{X})<\theta-\frac{C}{\sqrt{n}}\right\}+P_\theta\left\{d(\mathbf{X})>\theta+\frac{C}{\sqrt{n}}\right\}\,. \end{aligned} \tag{1}$$

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

$$\begin{aligned} 2P_\theta\left\{\bar{X}>\theta+\frac{C}{\sqrt{n}}\right\} &= 2P_\theta\left\{\sqrt{n}\left(\bar{X}-\theta\right)>C\right\} \\ &= 2\left[1-\Phi(C)\right]\,. \end{aligned}$$

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

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

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

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

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

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

Thus the *exact* risk is

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

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

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

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

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