# Homework 3

Tim Vigers

25 September 2020

## BD 1.4.5

Give an example in which the best linear predictor of $Y$ given $Z$ is a constant (has no predictive value) whereas the best predictor of $Y$ given $Z$ predicts Y perfectly.

The unique best linear predictor $\mu_L(Z) = E[Y] - \frac{Cov(Z,Y)}{Var(Z)}E[Z] + \frac{Cov(Z,Y)}{Var(Z)}Z$ and the best MSPE predictor of $Y$ given $Z$ is $E[Y|Z]$. First, take $Y = Z^2$ and calculate the covariance of $Y$ and $Z$:

$$Cov(Z,Y) = E[ZY] - E[Z]E[Y] = E[Z^3] - E[Z]E[Z^2]$$

If we restrict $Z$ so that $E[Z^3]$ and $E[Z]$ both equal 0, then:

$$\mu_L(Z) = E[Y] - 0 + 0 = E[Y]$$

The expected value of $Y$ is a constant, so this satisfies the first part of the question. Next, check the best MSPE predictor of $Y$ given $Z$:

$$\mu(Z) = E[Y|Z] = E[Z^2|Z] = Z^2 = Y$$

Thus, $\mu(Z)$ perfectly predicts $Y$ and this satisfies the second part of the question.

## BD 1.4.14

Let $Z_1$ and $Z_2$ be independent and have exponential distributions with density $\lambda e^{-\lambda z}$, $z > 0$. Define $Z = Z_2$ and $Y = Z_1 + Z_1 Z_2$. Find:

### a)

The best MSPE predictor $E[Y|Z = z]$ of $Y$ given $Z = z$:

First find $E[Y|Z = z] = E[Z_1 + Z_1 Z_2 | Z_2 = z]$. Because $Z_1$ and $Z_2$ are independent, this simplifies to $E[Z_1] + E[Z_1]E[Z_2|Z_2 = z]$, which is $\frac{1}{\lambda} + \frac{1}{\lambda}z = \frac{z+1}{\lambda}$.

### b)

$E[E[Y|Z]]$:

First find $E[Y|Z] = E[Z_1 + Z_1 Z_2 | Z_2] = \frac{Z+1}{\lambda}$ (see above). This contains the random variable $Z$, so take the expectation again:

$$E[\frac{Z+1}{\lambda}] = \frac{E[Z]+1}{\lambda} = \frac{\frac{1}{\lambda}+1}{\lambda} = \frac{1}{\lambda^2} + \frac{1}{\lambda}$$

## c)

$Var(E[Y|Z])$:

From above we know that $E[Y|Z] = \frac{Z+1}{\lambda}$. So, we find the variance of this using $Var(\frac{Z+1}{\lambda}) = \frac{Var(Z+1)}{\lambda^2}$. Because the variance of a RV plus a constant is the same as the variance of the RV, this simplifies to $\frac{Var(Z)}{\lambda^2} = \frac{\frac{1}{\lambda^2}}{\lambda^2} = \frac{1}{\lambda^4}$

## d)

$Var(Y|Z = z)$:

First we write $Y$ in terms of $Z_1$ and $Z_2$ to get $Var(Y|Z = z) = Var(Z_1 + Z_1Z_2|Z = z)$. Then we can plug in $Z_2 = z$ to get $Var(Y|Z = z) = Var(Z_1 + Z_1z)$ and rearrange and simplify to get $Var((z + 1)Z_1) = (z + 1)^2Var(Z_1)$. So, $Var(Y|Z = z) = (\frac{z+1}{\lambda})^2$.

## e)

$E[Var(Y|Z)]$:

From above we know that $Var(Y|Z) = Var(Z_1 + Z_1Z|Z) = (Z + 1)^2Var(Z_1) = \frac{(Z+1)^2}{\lambda^2}$. By expanding the numerator we get $E[Var(Y|Z)] = E[\frac{Z^2+2Z+1}{\lambda^2}]$. To find $E[Z^2]$ we rearrange the formula for variance to get $E[Z^2] = Var(Z) + E[Z]^2 = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2}$. So, plugging this back in we get:

$$E[Var(Y|Z)] = E[\frac{Z^2 + 2Z + 1}{\lambda^2}] = \frac{E[Z^2] + E[2Z] + 1}{\lambda^2} = \frac{\frac{2}{\lambda^2} + \frac{2}{\lambda} + 1}{\lambda^2}$$

This could be further rearranged, but I kind of like this form.

## f)

The best linear MSPE predictor of $Y$ based on $Z = z$:

Given $Z = z$, $Cov(Z, Y) = 0$ because $z$ is a constant. Therefore, the best linear predictor $\mu_L(Z) = E[Y|Z = z]$ (see equations in problem 1). So this is the same as part a).

# BD 1.6.4

Which of the following families of distributions are exponential families? (Prove or disprove.)

## b)

$p(x, \theta) = exp[-2log\theta + log(2x)]1[x \in (0, \theta)]$

This is not an exponential family because the indicator function depends on both $x$ and $\theta$, so the support depends on the parameter.

## d)

$\mathcal{N}(\theta, \theta^2)$

See scanned pages for the remainder of these problems.

Homework 3 (continued) Tim Vigess

BD 1.6.4 d)

The normal distribution can be written:

$$N(\theta, \theta^2) = \frac{1}{\sqrt{2\pi}\,\theta} \exp\left(-\frac{1}{2}\left(\frac{x-\theta}{\theta}\right)^2\right) 1(x \in \mathbb{R})$$

I haven't forgotten about this but suppress it in the following.

Expanding the numerator in the exponent gives you

$$N(\theta, \theta^2) = \frac{1}{\sqrt{2\pi}\,\theta} \exp\left(\frac{-x^2 + 2x\theta - \theta^2}{2\theta^2}\right)$$, which further simplifies to:

$$\frac{e^{-\log(\theta)}}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\theta^2} + \frac{x}{\theta} - \frac{1}{2}\right) = \frac{e^{-1/2}}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\theta^2} + \frac{x}{\theta} - \log(\theta)\right)$$

The above format clearly shows how to write the density in exponential family form:

$$N(\theta, \theta^2) = h(x) \exp\left(\sum_{j=1}^{2} \eta_j(\theta) T_j(x) - B(\theta)\right)$$

where $h(x) = \dfrac{1(x \in \mathbb{R}) e^{-1/2}}{\sqrt{2\pi}}$ $\qquad B(\theta) = \log(\theta)$

$\eta_1(\theta) = \dfrac{-1}{2\theta^2}$ $\qquad T_1(x) = x^2$

$\eta_2(\theta) = \dfrac{1}{\theta}$ $\qquad T_2(x) = x$

f) The unconditional distribution of $X$ is

$$p(x,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$, so the conditional is:

$$p(x,\theta \mid X>0) = \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}}{P(X>0)} = \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}}{1-(1-\theta)^n}$$

by the laws of condition probability. This can be rewritten as:

$$\binom{n}{x}\exp\left(x\log(\theta)+(n-x)\log(1-\theta)-\log(1-(1-\theta)^n)\right)$$

Which can be re-arranged to be:

$$\binom{n}{x}\exp\left(x(\log(\theta)-\log(1-\theta))+n\log(1-\theta)-\log(1-(1-\theta)^n)\right)$$

So the parts of the EF notation are:

$h(x)=\binom{n}{x}$ $\eta(\theta)=\log(\theta/1-\theta)$ $T(x)=x$

$B(\theta)=n\log(1-\theta)-\log(1-(1-\theta)^n)$

1.6.11. a) $N(\mu, \sigma^2)$

This problem is mostly a book-keeping one, so start by expanding and rearranging the PDF:

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-x^2}{2\sigma^2} + \frac{x\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right)$$

Then $h(x)$ absorbs $\frac{1}{\sqrt{2\pi}}$, and we can move $\frac{1}{\sigma}$ into the exponent by exponeniting the log:

$$N(\mu, \sigma^2) = h(x) \exp\left(\frac{-x^2}{2\sigma^2} + \frac{x\mu}{\sigma^2} - \left(\frac{\mu^2}{2\sigma^2} + \log(\sqrt{2\pi}\,\sigma)\right)\right)$$

From here is easy to see that:

$$\eta_1 = \frac{-1}{2\sigma^2} \qquad T_1(x) = x^2 \qquad A(\theta) = \frac{\mu^2}{\sigma^2} + \log(\sigma)$$

$$\eta_2 = \frac{\mu}{\sigma^2} \qquad T_2(x) = x$$

Next we get $A$ in terms of $\eta$:

$$\mu^2 = \eta_2^2 \sigma^4 \qquad 2\sigma^2 = \frac{-1}{\eta_1} \qquad \sigma = \frac{1}{\sqrt{-2\eta_1}} = (-2\eta_1)^{-\frac{1}{2}}$$

$$\mu^2 = \frac{\eta_2^2}{4\eta_1^2} \qquad \sigma^4 = \frac{1}{4\eta_1^2} \qquad \sqrt{2\pi}\,\sigma = \frac{\sqrt{2\pi}}{\sqrt{-2\eta_1}} = \left(\frac{2\pi}{-2\eta_1}\right)^{1/2}$$

Plugging the relevant pieces into A gives us:

$$A(\eta) = \frac{\eta_2^2(-\eta_1)}{4\eta_1^2} + \frac{1}{2}\log\left(\frac{\pi}{-\eta_1}\right)$$

$$= \frac{-\eta_2^2}{4\eta_1} + \frac{1}{2}\log\left(\frac{\pi}{-\eta_1}\right) \qquad \eta_1 \in \mathbb{R}^- \qquad \eta_2 \in \mathbb{R}$$

So by Theorem 1.6.2 the moment generating function for $T(X)$ is:

$$M(s) = \exp\left(A(\eta+s) - A(\eta)\right)$$ for $s$ in some neighborhood of $0$

And $E[T(X)] = A'(\eta)$ and $Var(T(X)) = A''(\eta)$.

To find the vector of expected values (because $T(X) = \{x^2, x\}$) we take the first derivative of $A(\eta)$ w.r.t. $\eta_1$ and then $\eta_2$:

$$\frac{\partial}{\partial \eta_1} A(\eta) = \frac{-\eta_2^2}{4}\left(\frac{-1}{\eta_1^2}\right) + \frac{1}{2}\left(\frac{1}{-\eta_1}\right) = \frac{\eta_2^2}{4\eta_1^2} - \frac{1}{2\eta_1}$$

$$\frac{\partial}{\partial \eta_2} A(\eta) = \frac{-1}{4\eta_1}\left(2\eta_2\right) = \frac{-\eta_2}{2\eta_1}$$

The can be written in terms of $\mu$ and sigma:
$E(T_1(X) = x^2) = \mu^2 + \sigma^2$, $E(T_2(X) = x) = \mu$.

To find the variance of $(T_1(X), T_2(X))$, we create a variance-covariance matrix of the form:

$$\begin{bmatrix} \frac{\partial^2 A(\eta)}{\partial \eta_1^2} & \frac{\partial^2 A(\eta)}{\partial \eta_1 \partial \eta_2} \\ \frac{\partial^2 A(\eta)}{\partial \eta_1 \partial \eta_2} & \frac{\partial^2 A(\eta)}{\partial \eta_2^2} \end{bmatrix}$$

where $\frac{\partial^2 A(\eta)}{\partial \eta_1^2} = \frac{\eta_2^2}{4}\left(\frac{-1}{\eta_1^3}\right) - \frac{1}{2}\left(\frac{-1}{\eta_1^2}\right) = \frac{-\eta_2^2}{4\eta_1^3} + \frac{1}{2\eta_1^2}$

$$\frac{\partial^2 A(\eta)}{\partial \eta_2^2} = \frac{-1}{2\eta_1}$$

$$\frac{\partial^2 A(\eta)}{\partial \eta_1 \partial \eta_2} = \frac{\partial}{\partial \eta_2} \frac{\eta_2^2}{4\eta_1^2} - \frac{1}{2\eta_1} = \frac{2\eta_2}{4\eta_1^2}$$

and $\frac{\partial}{\partial \eta_1} \frac{-\eta_2}{2\eta_1} = \frac{\eta_2}{2\eta_1^2}$

These can then be put in terms of $\mu$ and $\sigma$:

$$\begin{bmatrix} \frac{-\eta_2^2}{4\eta_1^3} + \frac{1}{2\eta_1^2} & \frac{-\eta_2}{2\eta_1^2} \\ \frac{\eta_2}{2\eta_1^2} & \frac{-1}{2\eta_1} \end{bmatrix} = \begin{bmatrix} \frac{\mu^2\sigma^2}{2} + 2\sigma^4 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & \sigma^2 \end{bmatrix}$$

b) $\Gamma(p, \lambda)$ with $p$ fixed

$$\Gamma(p,\lambda) = \frac{1}{\Gamma(p)\,\lambda^p} x^{p-1} \exp\left(\frac{-x}{\lambda}\right)$$

We can move $1/\lambda^p$ into the exponent and put $x^{p-1}/\Gamma(p)$ into $h(x)$ to obtain:

$$\Gamma(p, \lambda) = h(x) \exp\left(\frac{-x}{\lambda} - p\log(\lambda)\right)$$

Therefore $\eta = \frac{-1}{\lambda}$ $\quad T(x) = x$

Putting $p\log(\lambda)$ in terms of $\eta$ gives us:

$$p\log(\lambda) = p\log\left(\frac{1}{-\eta}\right) = p(\log(1) - \log(-\eta))$$

So $A(\eta) = -p\log(-\eta)$.

To find $E[T(x)]$ we again take $A'(\eta)$:

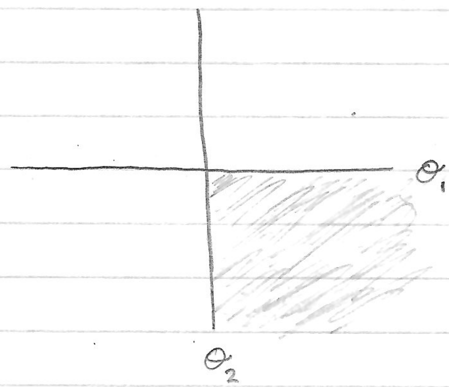
$$A'(\eta) = \frac{-p}{\eta} = \frac{-p}{-1/\lambda} = \boxed{p\lambda}$$

Next to find the variance of $T(x)$ we take

$$A''(\eta) = \frac{p}{\eta^2} = \frac{p}{(-1/\lambda)^2} = \boxed{p\lambda^2}$$

2.2.8) Find the least-squares estimator for the model $Y_i = \theta_1 + \theta_2 z_i + \varepsilon_i$ with $\varepsilon_i$ satisfying the Gauss-Markov assumptions and restrictions $\theta_1 \geq 0$, $\theta_2 \leq 0$.

First, it's helpful to plot $\theta_1$ and $\theta_2$ to visualize the restrictions:




The lower right quadrant is where $\theta_1$ and $\theta_2$ are defined. For a point in this space, the contrast $\sum_{i=1}^{n}(Y_i - g(\theta, z_i))^2$ leads to the least squares equation:

$$Q = \sum_{i=1}^{n}(Y_i - (\theta_1 + \theta_2 z_i))^2$$

(From here on I'm going to use $\Sigma$ to represent $\sum_{i=1}^{n}$).

To minimize this, we take partial derivatives with respect to $\theta_1$ and $\theta_2$ and solve the system of equations when both are set equal to 0:

$$d/d\theta_1\, Q = -2\,\Sigma(Y_i - \theta_1 - \theta_2 z_i)$$

$$d/d\theta_2\, Q = -2\,\Sigma(Y_i - \theta_1 - \theta_2 z_i) z_i$$

First the partial derivative w.r.t. $\theta_1$:

$$\sum(y_i - \hat{\theta}_1 - \hat{\theta}_2 z_i) = 0$$
$$\sum y_i - n\hat{\theta}_1 - \hat{\theta}_2 \sum z_i = 0$$

$$\sum y_i - \hat{\theta}_2 z_i = n\hat{\theta}_1$$

So, $\hat{\theta}_1 = \dfrac{\sum y_i - \hat{\theta}_2 z_i}{n} = \boxed{\bar{y} - \hat{\theta}_2 \bar{z}}$

Then the partial derivative w.r.t. $\theta_2$, substituting the above formula in for $\hat{\theta}_1$:

$$\sum(y_i - (\bar{y} - \hat{\theta}_2\bar{z}) - \hat{\theta}_2 z_i)z_i = 0$$
$$\sum(y_i - \bar{y} + \hat{\theta}_2(\bar{z} - z_i))z_i = 0$$
$$\sum(y_i - \bar{y})z_i + \hat{\theta}_2 \sum(\bar{z} - z_i)z_i = 0$$

$$\sum z_i(y_i - \bar{y}) = \hat{\theta}_2 \sum z_i(z_i - \bar{z})$$

So, $\boxed{\hat{\theta}_2 = \dfrac{\sum z_i(y_i - \bar{y})}{\sum z_i(z_i - \bar{z})}}$

Because of the restrictions on the model, we also need to consider situations in which the minimum falls outside the bottom-right quadrant. For example, if the minimum falls in bottom-left quadrant, then the restricted minimum will occur at $\theta_2 = 0$ and some point in $\theta_1$:

If $\theta_2 = 0$, $\quad Q_1 = \Sigma(Y_i - \theta_1)^2$

and $\quad d/d\theta_1 \; Q = -2\Sigma(Y_i - \theta_1)$

setting this to 0 and solving for $\hat{\theta}_1$ gives us:

$$\Sigma Y_i - n\hat{\theta}_1 = 0, \quad \text{and} \quad \boxed{\hat{\theta}_1 = \frac{\Sigma Y_i}{n} = \bar{Y}}$$

Using the same approach for when the minimum falls in the upper right quadrant we see that if $\theta_1 = 0$,

$$Q_2 = \Sigma(Y_i - \theta_2 z_i)^2 \quad \text{and}$$

$$d/d\theta_2 \; Q_2 = -2\Sigma(Y_i - \theta_2 z_i).$$ setting this to 0 and

solving for $\hat{\theta}_2$ gives us:

$$\Sigma(Y_i - \theta_2 z_i) = 0$$
$$\Sigma Y_i = \hat{\theta}_2 \Sigma z_i$$

So $\quad \boxed{\hat{\theta}_2 = \frac{\Sigma Y_i}{\Sigma z_i}}$

If the minimum is in the top-left quadrant, the function is minimized at $\hat{\theta}_1 = 0$ and $\hat{\theta}_2 = 0$, because it is a quadratic function. So if you imagine climbing out of a bowl which has its lowest point in the top left quadrant, the lowest point you could reach in the restricted space would be the origin.

2.2.35) Let $g(x) = \dfrac{1}{\pi(1+x^2)}$ $x \in \mathbb{R}$ and

$X_1$ and $X_2$ be iid with density $g(x-\theta)$ $\theta \in \mathbb{R}$. Let $x_1$ and $x_2$ be the observations and $\Delta = \frac{1}{2}(x_1 - x_2)$.

a & b) First take the log of the likelihood $(\ell_x(\theta))$

$$L_x(\theta) = \prod_{i=1}^{2} \frac{1}{\pi(1+(x_i-\theta)^2)} = \frac{1}{\pi^2} \prod_{i=1}^{2} \frac{1}{(1+(x_i-\theta)^2)}$$

$$\ell_x(\theta) = -2\log(\pi) + \sum_{i=1}^{2} \frac{1}{(1+(x_i-\theta)^2)}$$

Take the first derivative w.r.t. $\theta$:

$$\frac{d}{d\theta} \ell_x(\theta) = \sum_{i=1}^{2} \frac{2(x_i-\theta)}{(1+(x_i-\theta)^2)} = \frac{2(x_1-\theta)}{(1+(x_1-\theta)^2)} + \frac{2(x_2-\theta)}{(1+(x_2-\theta)^2)}$$

Next, because there are only two observations, we can write $x_1$ and $x_2$ in terms of $\bar{x}$ and $\Delta$:

$$x_1 = \frac{x_1+x_2}{2} + \frac{x_1-x_2}{2} = \bar{x} + \Delta \text{ and it follows } x_2 = \bar{x} - \Delta.$$

Plugging these values into the $d/d\theta\ \ell_x(\theta)$ gives us:

$$\frac{2(\bar{x}+\Delta-\theta)}{(1+(\bar{x}+\Delta-\theta)^2)} + \frac{2(\bar{x}-\Delta-\theta)}{(1+(\bar{x}-\Delta-\theta)^2)}$$

This can be rearranged to:

$$\frac{2(\bar{x}-\theta)((\bar{x}-\theta)^2-\Delta^2+1)}{(1+(\bar{x}+\Delta-\theta)^2)(1+(\bar{x}-\Delta-\theta)^2)}$$

Setting this equal to 0 and solving for $\theta$ provides three potential MLEs:

$$2(\bar{x}-\theta)((\bar{x}-\theta)^2-\Delta^2+1) = 0$$

$$\bar{x}-\theta = 0 \quad \text{with} \quad \boxed{\hat{\theta} = \bar{x}}$$

or

$$(\bar{x}-\theta)^2-\Delta^2+1 = 0$$

$$\bar{x}-\theta = \sqrt{\Delta^2-1} \quad \text{with} \quad \boxed{\hat{\theta} = \bar{x} \pm \sqrt{\Delta^2-1}}$$

If $|\Delta| > 1$, then $\pm\sqrt{\Delta^2-1}$ exists and there are two values for the MLE. However, if $|\Delta| \leq 1$, then $\pm\sqrt{\Delta^2-1}$ does not exist and the MLE is $\hat{\theta} = \bar{x}$, which is unique.

The second derivative of $\ell_x(\theta)$ is

$$\frac{2((\bar{x}+\Delta-\theta)^2-1)}{(1+(\bar{x}+\Delta-\theta)^2)^2} + \frac{2((\bar{x}-\Delta-\theta)^2-1)}{(1+(\bar{x}-\Delta-\theta)^2)^2} =$$

Both pieces are negative or 0 for $\bar{x} = \hat{\theta}$ as long as $|\Delta| < 1$, so in this case $\bar{x}$ is indeed an MLE. If $|\Delta| > 1$, both pieces are positive.