

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 $Y = Z^3$ 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:

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

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 $\frac{Var(Z)}{\lambda^2} = \frac{1}{\lambda^2} = \frac{1}{\lambda^4}$ ✓

d)

$Var(Y|Z = z)$:

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

e)

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

From above we know that $Var(Y|Z) = Var(Z_1 + Z_1 Z|Z) = (Z+1)^2 Var(Z_1) = \frac{(Z+1)^2}{\lambda^2}$. 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)

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

This is not an exponential family because the indicator function depends on both x and θ , 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 Vigers

BD 1.6.4 d)

The normal distribution can be written:

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

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

Expanding the numerator in the exponent gives you

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

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

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

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

where $h(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}}$ $B(\theta) = \log(\theta)$

$$n_1(\theta) = \frac{-1}{2\theta^2}$$

$$T_1(x) = x^2$$

$$n_2(\theta) = \frac{1}{\theta}$$

$$T_2(x) = x \quad \checkmark$$

f) The unconditional distribution of X is

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

is:

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

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:

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

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

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 exponentiating 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:

$$n_1 = \frac{-1}{2\sigma^2} \quad T_1(x) = x^2$$

$$n_2 = \frac{\mu}{\sigma^2} \quad T_2(x) = x \quad A(\sigma) = \frac{\mu^2}{\sigma^2} + \log(\sigma)$$

Next we get A in terms of n :

$$\mu^2 = n_2^2 \sigma^4 \quad 2\sigma^2 = \frac{-1}{n_1} \quad \checkmark \quad \sigma = \frac{1}{\sqrt{-2n_1}} = (-2n_1)^{\frac{1}{2}}$$

$$\mu^2 = \frac{n_2^2}{4n_1^2}$$

$$\sigma^4 = \frac{1}{4n_1^2}$$

$$\sqrt{2\pi}\sigma = \frac{\sqrt{2\pi}}{\sqrt{-2n_1}} = \left(\frac{2\pi}{-2n_1}\right)^{\frac{1}{2}}$$

Plugging the relevant pieces into A gives us:

$$A(n) = \frac{n_2^2(-n_1)}{4n_1^2} + \frac{1}{2} \log\left(\frac{n_1}{-n_1}\right)$$

$$= -\frac{n_2^2}{4n_1} + \frac{1}{2} \log\left(\frac{n_1}{-n_1}\right) \quad n_1 \in \mathbb{R}^- \quad n_2 \in \mathbb{R}$$

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

$$M(s) = \exp(A(n+s) - A(n)) \quad \text{for } s \text{ in some neighborhood of 0}$$

And $E[T(X)] = A'(n)$ and $\text{Var}(T(X)) = A''(n)$.

To find the vector of expected values (because $T(X) = \{x^2, x\}$) we take the first derivative of $A(n)$ w.r.t. n_1 and then n_2 :

$$\frac{\partial}{\partial n_1} A(n) = -\frac{n_2^2}{4} \left(-\frac{1}{n_1^2}\right) + \frac{1}{2} \left(\frac{1}{-n_1}\right) = \frac{n_2^2}{4n_1^2} - \frac{1}{2n_1}$$

$$\frac{\partial}{\partial n_2} A(n) = \frac{-1}{4n_1} \left(2n_2\right) = \frac{-n_2}{2n_1}$$

The can be written in terms of μ and sigma:

$$E(T_1(X)=x^2) = \mu^2 + \sigma^2, \quad E(T_2(X)=x) = \mu.$$

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

$$\begin{bmatrix} \frac{\partial^2 A(n)}{\partial n_1^2} & \frac{\partial^2 A(n)}{\partial n_1 \partial n_2} \\ \frac{\partial^2 A(n)}{\partial n_1 \partial n_2} & \frac{\partial^2 A(n)}{\partial n_2^2} \end{bmatrix}$$

where $\frac{\partial^2 A(n)}{\partial n_1^2} = \frac{n_2^2}{4} \left(\frac{-1}{n_1^3} \right) - \frac{1}{2} \left(\frac{-1}{n_1^2} \right) = \frac{-n_2^2}{4n_1^3} + \frac{1}{2n_1^2}$

$$\frac{\partial^2 A(n)}{\partial n_2^2} = -\frac{1}{2n_1}$$

$$\frac{\partial^2 A(n)}{\partial n_1 \partial n_2} = \frac{\partial}{\partial n_2} \frac{n_2^2}{4n_1^2} - \frac{1}{2n_1} = \frac{2n_2}{4n_1^2}$$

and $\frac{\partial}{\partial n_1} -\frac{n_2}{2n_1} = \frac{n_2}{2n_1^2}$

These can then be put in terms of μ and σ :

$$\begin{bmatrix} \frac{-n_2^2}{4n_1^3} + \frac{1}{2n_1^2} & \frac{n_2}{2n_1^2} \\ \frac{n_2}{2n_1^2} & -\frac{1}{2n_1} \end{bmatrix} = \begin{bmatrix} \frac{\mu^2 \sigma^2}{2} + 2\sigma^4 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & \sigma^2 \end{bmatrix} \checkmark$$

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 $\frac{1}{\lambda^p}$ 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 $n = \frac{-1}{\lambda}$ $T(x) = x$

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

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

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

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

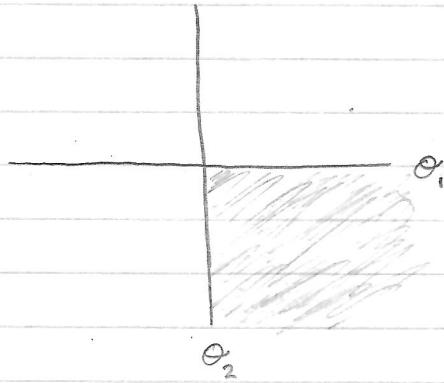
$$A'(n) = -\frac{p}{n} = -\frac{p}{-\frac{1}{\lambda}} = p\lambda$$

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

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

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

First, it is helpful to plot θ_1 and θ_2 to visualize the restrictions:



The lower right quadrant is where θ_1 and θ_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

$$Q = \sum_{i=1}^n (y_i - (\theta_1 + \theta_2 z_i))^2 \quad \begin{matrix} \text{(from here on I'm going} \\ \text{to use } \Sigma \text{ to represent } \sum_{i=1}^n \end{matrix}$$

To minimize this, we take partial derivatives with respect to θ_1 and θ_2 and solve the system of equations when both are set equal to 0:

$$\frac{d}{d\theta_1} Q = -2 \sum (y_i - \theta_1 - \theta_2 z_i)$$

$$\frac{d}{d\theta_2} Q = -2 \sum (y_i - \theta_1 - \theta_2 z_i) z_i$$

First the partial derivative w.r.t. θ_1 :

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

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

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

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

Then the partial derivative w.r.t. θ_2 , substituting the above formula in for $\hat{\theta}_1$:

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

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

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

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

$$\text{So, } \hat{\theta}_2 = \frac{\sum z_i (y_i - \bar{y})}{\sum z_i (\bar{z} - z_i)} \quad \checkmark$$

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 θ_1 :

If $\theta_2 = 0$, $Q_1 = \sum (y_i - \theta_1)^2$

and $\frac{d}{d\theta_1} Q = -2 \sum (y_i - \theta_1)$

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

$$\sum y_i - n \hat{\theta}_1 = 0, \text{ and } \hat{\theta}_1 = \frac{\sum y_i}{n} = \bar{y}$$

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

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

$\frac{d}{d\theta_2} Q_2 = -2 \sum (y_i - \theta_2 z_i)$. setting this to 0 and solving for $\hat{\theta}_2$ gives us:

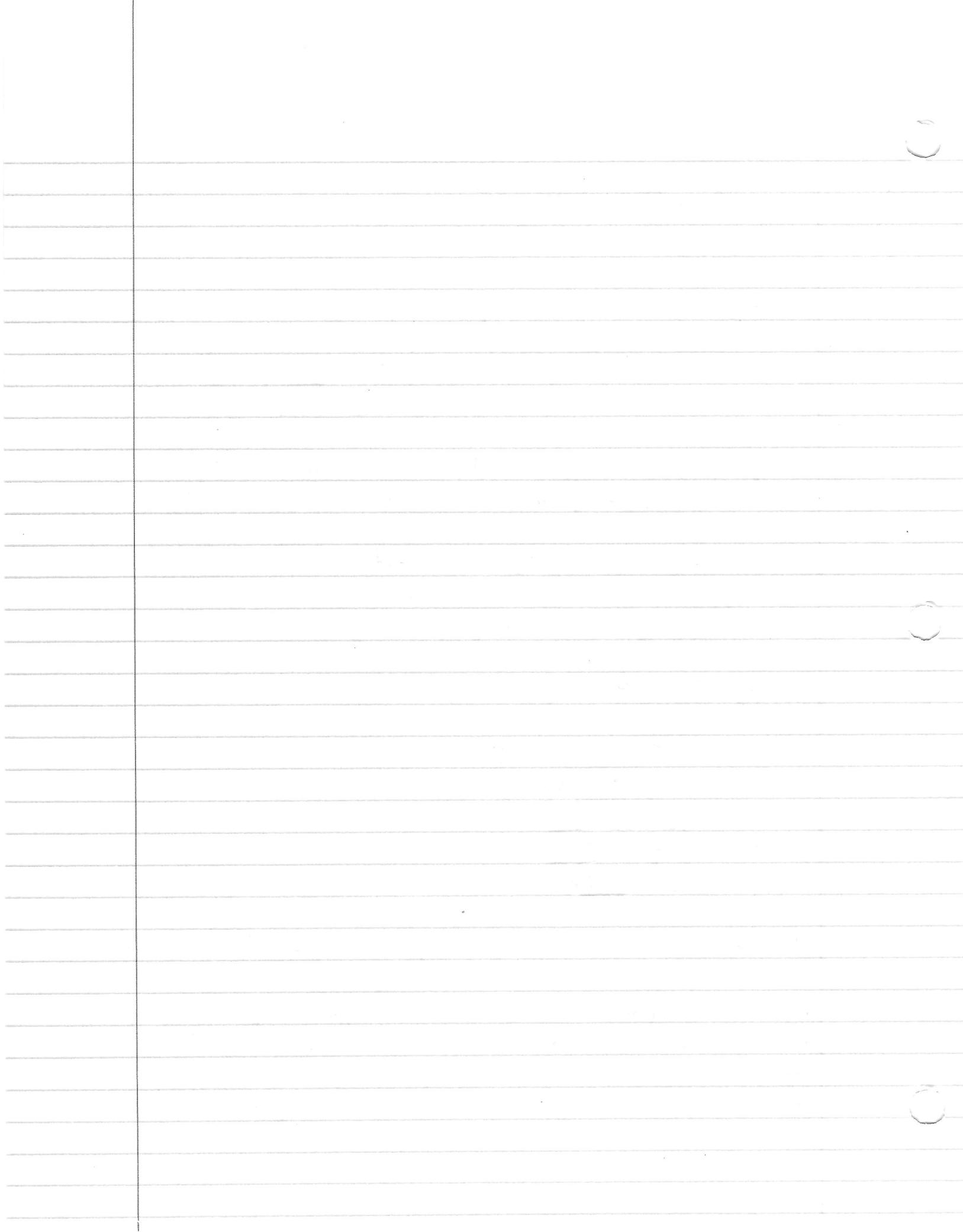
$$\sum (y_i - \theta_2 z_i) = 0$$

$$\sum y_i = \hat{\theta}_2 \sum z_i$$

so
$$\hat{\theta}_2 = \frac{\sum y_i}{\sum z_i}$$

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.





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

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 ($\ell_x(\theta)$)

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

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

Take the first derivative w.r.t. θ :

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

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

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

Plugging these values into the $\frac{\partial}{\partial \theta} \ell_x(\theta)$ 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 θ provides three potential MLEs:

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

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

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

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



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 $l_x(\theta)$ 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.