Solutions to Homework 3
BIOS 7731

1. **BD 1.4.5. Give an example where 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.**

   - Recall that the best linear predictor is $\mu_L = a + bZ$, where $a = E[Y] - bE[Z]$ and $b = \frac{Cov(Z,Y)}{Var(Z)}$. To be constant, then $Cov(Z,Y) = E[ZY] - E[Z]E[Y] = 0$.

   - For the best predictor of Y given Z to predict Y perfectly, then Y must be equal to some function of Z.

   - We need to find a function of Z to set to Y and that has 0 covariance. Use the squared function $Y = Z^2$.

   - Consider the joint density in this discrete example

	Z=1	Z=0	Z=-1	P(Y)
Y=0	0	1/3	0	1/3
Y=1	1/3	0	1/3	2/3
P(Z)	1/3	1/3	1/3	1

     The MSPE is $E[Y|Z]$ and in this case Y is related to Z by $Y = Z^2$, therefore the MSPE $E[Y|Z] = Z^2$ predicts Y perfectly. Also, $Cov(Z,Y) = E[ZY] - E[Z]E[Y] = 0$, since $E[ZY] = 1 * 1/3 + 0 * 1/3 - 1 * 1/3 = 0$, $E[Z] = 0$ and $E[Y] = 2/3$. Therefore, the best linear predictor is a constant and $\mu_L = E[Y] = 2/3$.

   - This can also be shown for a continuous case. For example assume $Z \sim N(0,1)$, and that $Y = Z^2$ so the best predictor of Y given Z predicts Y perfectly. The covariance is $Cov(Z,Y) = E[ZY] - E[Z][Y] = E[Z^3] - 0 = 0$ (3rd moment of $N(0,1)$ is 0), and the best linear predictor is a constant.

2. **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 $Y = Z_1 + Z_1 Z_2$.**

Since Z_1 and Z_2 have exponential distributions, then

$$\begin{aligned} E(Z_1) = E(Z_2) &= \frac{1}{\lambda} \\ E(Z_1^2) = E(Z_2^2) &= \frac{2}{\lambda^2} \\ Var(Z_1) = Var(Z_2) &= \frac{1}{\lambda^2} \end{aligned}$$

These results can be obtained by using integration by parts and using the formula $Var(Z) = E(Z^2) - (E(Z))^2$.

(a) The best MSPE predictor $E(Y|Z = z)$ of Y given $Z = z$.

$$\begin{aligned} E(Y|Z = z) &= E(Z_1 + Z_1 Z_2 | Z_2 = z) \\ &= E(Z_1 + z Z_1 | Z_2 = z) \\ &= E((1 + z) Z_1 | Z_2 = z) \\ &= (1 + z) E(Z_1 | Z_2 = z) \\ &= (1 + z) E(Z_1) \\ &= \frac{1 + z}{\lambda} \end{aligned}$$

Note $E(Z_1|Z_2 = z) = E(Z_1)$ because Z_1 and Z_2 are independent.

(b) $E(E(Y|Z))$

$$\begin{aligned} E(E(Y|Z)) &= E(\frac{1 + Z}{\lambda}) \\ &= \frac{1}{\lambda} E(1 + Z) \\ &= \frac{1}{\lambda}(1 + E(Z)) \\ &= \frac{1}{\lambda}\left(1 + \frac{1}{\lambda}\right) \\ &= \frac{1}{\lambda} + \frac{1}{\lambda^2} \end{aligned}$$

(c) $Var(E(Y|Z))$

$$\begin{aligned} Var(E(Y|Z)) &= Var\left(\frac{1}{\lambda} + \frac{Z}{\lambda}\right) \\ &= \frac{1}{\lambda^2} Var(Z) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\lambda^2}\Big(\frac{1}{\lambda^2}\Big) \\
&= \frac{1}{\lambda^4}
\end{aligned}
$$

(d) $Var(Y|Z=z)$

$$
\begin{aligned}
Var(Y|Z=z) &= Var(Z_1 + Z_1 Z_2 | Z_2 = z) \\
&= Var(Z_1 + zZ_1 | Z_2 = z) \\
&= Var((1+z)Z_1 | Z_2 = z) \\
&= (1+z)^2 Var(Z_1 | Z_2 = z) \\
&= (1+z)^2 Var(Z_1) \\
&= \frac{(1+z)^2}{\lambda^2}
\end{aligned}
$$

Note $Var(Z_1|Z_2 = z) = Var(Z_1)$ because Z_1 and Z_2 are independent.

(e) $E(Var(Y|Z))$

$$
\begin{aligned}
E(Var(Y|Z)) &= E\Big((1+Z)^2 \frac{1}{\lambda^2}\Big) \\
&= \frac{1}{\lambda^2}\Big(E(1 + 2Z + Z^2)\Big) \\
&= \frac{1}{\lambda^2}\Big(E(1) + 2E(Z) + E(Z^2)\Big) \\
&= \frac{1}{\lambda^2}\Big(1 + \frac{2}{\lambda} + \frac{2}{\lambda^2}\Big) \\
&= \frac{1}{\lambda^2} + \frac{2}{\lambda^3} + \frac{2}{\lambda^4}
\end{aligned}
$$

(f) **The best linear MSPE predictor of Y based on $Z = z$.**

Two alternative ways to show this:
- The best MSPE predictor from part a) is $\frac{1}{\lambda} + \frac{z}{\lambda}$, which is linear. Therefore, it is the best linear MSPE predictor.
- Otherwise, derive using Theorem 1.43, where the unique best linear predictor is
$\mu_L(Z) = a_1 + b_1 Z$, where $b_1 = \dfrac{Cov(Z,Y)}{Var(Z)}$ and $a_1 = E[Y] - b_1 E[Z]$.

$$
\begin{aligned}
Cov(Z,Y) &= E\Big[(Z - E[Z])(Y - E[Y])\Big] \\
&= E\Big[\Big(Z - \frac{1}{\lambda}\Big)\Big(Y - \Big(\frac{1}{\lambda} + \frac{1}{\lambda^2}\Big)\Big)\Big] \\
&= E[ZY] - \frac{1}{\lambda}E[Y] - \Big(\frac{1}{\lambda} + \frac{1}{\lambda^2}\Big)E[Z] + \frac{1}{\lambda}\Big(\lambda + \frac{1}{\lambda^2}\Big)
\end{aligned}
$$

$$
\begin{aligned}
&= E[Z(Z_1 + Z_1 Z)] - \frac{1}{\lambda}\Big(\frac{1}{\lambda} + \frac{1}{\lambda^2}\Big) - \frac{1}{\lambda}\Big(\frac{1}{\lambda} + \frac{1}{\lambda^2}\Big) + \frac{1}{\lambda}\Big(\frac{1}{\lambda} + \frac{1}{\lambda^2})\Big) \\
&= E[ZZ_1] + E[Z^2 Z] - \frac{1}{\lambda}\Big(\lambda + \frac{1}{\lambda^2})\Big) \\
&= E[Z]E[Z_1] + E[Z^2]E[Z_1] - \frac{1}{\lambda}\Big(\lambda + \frac{1}{\lambda^2})\Big) \\
&= \frac{1}{\lambda^2} + (Var(Z) + (E[Z])^2)\frac{1}{\lambda} - \frac{1}{\lambda}\Big(\lambda + \frac{1}{\lambda^2})\Big) \\
&= \frac{1}{\lambda^2} + \Big(\frac{1}{\lambda^2} + \frac{1}{\lambda^2}\Big)\frac{1}{\lambda}\Big(\lambda + \frac{1}{\lambda^2})\Big) \\
&= \frac{1}{\lambda^2} + \frac{2}{\lambda^2} - \frac{1}{\lambda^2} - \frac{1}{\lambda^2} \\
&= \frac{1}{\lambda^3}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
b_1 &= \frac{Cov(Z,Y)}{Var(Z)} \\
&= \frac{\frac{1}{\lambda^3}}{\frac{1}{\lambda^2}} \\
&= \frac{1}{\lambda}
\end{aligned}
$$

$$
\begin{aligned}
a_1 &= E[Y] - b_1 E[Z] \\
&= E[E[Y|Z]] - \frac{1}{\lambda}\Big(\frac{1}{\lambda}\Big) \\
&= \frac{1}{\lambda} + \frac{1}{\lambda^2} - \frac{1}{\lambda^2} \\
&= \frac{1}{\lambda}
\end{aligned}
$$

So the unique best linear predictor is $\mu_L(Z) = \frac{1}{\lambda} + \frac{Z}{\lambda} = \frac{1+Z}{\lambda}$.

3. **BD 1.6.4 b) d) f)** Which of the following families of distributions are exponential families? (Prove or disprove.)

- b) $p(x,\theta) = \exp[-2log(\theta) + log(2x)]1[x \in (0,\theta)]$
  The density is
  $$p(x,\theta) \quad = \quad \exp[-2log(\theta) + log(2x)]1[x \in (0,\theta)].$$
  This is not an exponential family since $p(x,\theta)$ can not be written in the form of an exponential family $h(x)\exp[\eta(\theta)T(x) - B(\theta)]$. The indicator $1[x \in (0,\theta)]$ depends on θ and cannot be separated into $\eta(\theta)T(x)$.
- d) $N(\theta,\theta^2)$, $\theta > 0$.
  The density is
  $$p(x,\theta) = \frac{1}{\sqrt{2\pi}\theta}\exp(-\frac{1}{2\theta^2}(x-\theta)^2).$$
  With some re-arranging, this is an exponential family with $\eta(\theta) = [\frac{1}{\theta}, -\frac{1}{2\theta^2}]$, $T(x) = [x, x^2]$, $B(\theta) = \log\theta$, and $h(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2})$. However, the parameter space has dimension $l = 1$, which is lower dimension than η and T, which has dimension $k = 2$. Therefore, this is a curved exponential family.
- f) $p(x,\theta)$ is the conditional frequency function of a binomial $B(n,\theta)$, variable X, given that $X > 0$.
  The density is
  $$p(x,\theta) = \frac{p(X = x, X > 0)}{P(X > 0)} = \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}}{1-(1-\theta)^n}.$$
  With some re-arranging, this is an exponential family with $\eta(\theta) = \log\theta - \log(1-\theta)$, $T(x) = x$, $B(\theta) = -\log(\frac{(1-\theta)^n}{1-(1-\theta)^n})$ and $h(x) = \binom{n}{x}$.

4. **BD 1.6.11** Use Theorems 1.6.2 and 1.6.3 to obtain moment-generating functions for the sufficient statistics, when sampling from the following distributions.

(a) Normal, $\overrightarrow{\theta} = (\mu, \sigma^2)$

$$\begin{aligned} p(x, \overrightarrow{\theta}) &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \\ &= exp\Big(-\frac{1}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2\Big) \\ &= exp\Big(-\frac{1}{2\sigma^2}\Big(x^2 - 2x\mu + \mu^2\Big) - \frac{1}{2}log(2\pi\sigma^2)\Big) \\ &= exp\Big(\frac{\mu x}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{1}{2}\Big(\frac{\mu^2}{\sigma^2} + log(2\pi\sigma^2)\Big)\Big) \end{aligned}$$

This corresponds to a 2-parameter exponential family where $h(x) = 1$, $\theta_1 = \mu$, $\theta_2 = \sigma^2$, $\eta_1 = \frac{\mu}{\sigma^2}$, $T_1(x) = x$, $\eta_2 = -\frac{1}{2\sigma^2}$, $T_2(x) = x^2$, and $B(\theta) = \frac{1}{2}\Big(\frac{\mu^2}{\sigma^2} + log(2\pi\sigma^2)\Big)$, for $\eta_1 \in R$ and $\eta_2 \in R^-$

Rewrite this expression in the following form:
$q(x, \overrightarrow{\eta}) = h(x)\Big(exp\Big(\mathbf{T}^T(x)\overrightarrow{\eta} - A(\overrightarrow{\eta})\Big)\Big)$.
$\mathbf{T} = (x, x^2)$ and $\overrightarrow{\eta} = \Big(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\Big)$ and
$A(\overrightarrow{\eta}) = \frac{1}{2}\Big(-\frac{\eta_1^2}{2\eta_2} + log\Big(\frac{\pi}{-\eta_2}\Big)\Big) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}log(-\eta_2) + \frac{1}{2}log(\pi)$.
By Theorem 1.63, if $\mathcal{E}$ has nonempty interior in $\mathcal{R}^k$ and $\eta_0 \in \mathcal{E}$, then $T(X)$ has under η_0 a moment generating function M given by $M(s) = exp(A(\eta_0 + s) - A(\eta_0))$ valid for all s such that $\eta_0 + s \in \mathcal{E}$.

$$\begin{aligned} M(s) &= exp\Big(\frac{1}{2}\Big(\frac{-(\eta_1+s_1)^2}{2(\eta_2+s_2)} + log(\pi) - log(-(\eta_2+s_2)\Big) - \frac{1}{2}\Big(\frac{-\eta_1^2}{2\eta_2} + log(\pi) - log(-\eta_2)\Big)\Big) \\ &= exp\Big(\frac{1}{2}\Big(\frac{-(\eta_1+s_1)^2}{2(\eta_2+s_2)} - \frac{-\eta_1^2}{2\eta_2}\Big) - \frac{1}{2}\Big(log(-(\eta_2+s_2)) - log(-\eta_2)\Big)\Big) \\ &= \sqrt{\frac{\eta_2}{\eta_2+s_2}} exp\Big(\frac{1}{2}\Big(\frac{-(\eta_1+s_1)^2}{2(\eta_2+s_2)}\Big)\Big)\Big) \end{aligned}$$

<u>Sanity check for $A(\eta)$</u>

Note that $\mu = -\frac{\eta_1}{2\eta_2}$ and $\sigma^2 = -\frac{1}{2\eta_2}$, then taking the first and second derivatives of $A(\eta)$ to find the mean of variance of $T(X)$.

$$\begin{aligned} E_{\eta_0}\mathbf{T}(X) &= A'(\eta_0) \\ &= \Big(\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}\Big) \end{aligned}$$

$$\begin{aligned}
&= \Big(-\frac{\eta_1}{2\eta_2}, \frac{\eta_1^2}{4\eta_2^2} - \frac{1}{2\eta_2}\Big) \\
&= (\mu, \mu^2 + \sigma^2)
\end{aligned}$$

$$\begin{aligned}
Var_{\eta_0}(\mathbf{T}(X)) &= \begin{pmatrix} \frac{\partial^2 A}{\partial \eta_1^2} & \frac{\partial^2 A}{\partial \eta_1 \partial \eta_2} \\ \frac{\partial^2 A}{\partial \eta_1 \partial \eta_2} & \frac{\partial^2 A}{\partial \eta_2^2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{-1}{2\eta_2} & \frac{\eta_1}{2\eta_2^2} \\ \frac{\eta_1}{2\eta_2^2} & \frac{1}{2\eta_2^2} + \frac{\eta_1^2}{2\eta_2^3} \end{pmatrix} \\
&= \begin{pmatrix} \sigma^2 & 2\mu\sigma \\ 2\mu\sigma & 2\sigma^4 + 2\mu^2\sigma^2 \end{pmatrix}
\end{aligned}$$

(b) Gamma, $\Gamma(p, \lambda)$, $\theta = \lambda$, p fixed

$$\begin{aligned}
p(x, \lambda) &= \frac{\lambda^p x^{p-1} \exp(-\lambda x)}{\Gamma(p)} \\
&= \frac{x^{p-1}}{\Gamma(p)} \exp\Big(plog(\lambda) - \lambda x\Big) \\
&= \frac{x^{p-1}}{\Gamma(p)} \exp\Big(-\lambda x - (-plog(\lambda))\Big)
\end{aligned}$$

This corresponds to a one-parameter exponential family where $h(x) = \frac{x^{p-1}}{\Gamma(p)}$, $\eta(\theta) = -\lambda$, $T(x) = x$, and $B(\theta) = -plog(\lambda)$, where $\eta \in R^-$. Rewrite this expression in the following form:

$$q(x, \eta) = h(x)\Big(\exp\Big(\eta T(x) - A(\eta)\Big)\Big).$$

Note $h(x)$, η and $T(x)$ remain the same but now $A(\eta) = -plog(-\eta)$. By Theorem 1.6.2, if X is distributed according to the form of a canonical exponential family, and η is an interior point of $\mathcal{E}$, the moment generating function of $T(X)$ exists and is given by $M(s) = \exp(A(s + \eta) - A(\eta))$ for s in some neighborhood of 0.

$$\begin{aligned}
M(s) &= \exp[-plog(-(\eta + s)) - (-plog(-\eta))] \\
&= \exp[-plog(\frac{-(\eta + s)}{-\eta}] \\
&= \exp[log\Big(\frac{\eta + s}{\eta}\Big)^{-p}] \\
&= \Big(\frac{\eta + s}{\eta}\Big)^{-p} \\
&= \Big(\frac{\eta}{\eta + s}\Big)^{p}
\end{aligned}$$

For the mean and variance, since $\eta = -\lambda$,

$$\begin{aligned} E(T(X)) &= A'(\eta) \\ &= -\frac{p}{\eta} \\ &= \frac{p}{\lambda} \end{aligned}$$

$$\begin{aligned} Var(T(X)) &= A''(\eta) \\ &= \frac{p}{\eta^2} \\ &= \frac{p}{\lambda^2} \end{aligned}$$

5. **BD 2.2.8 Find the least square estimate for the model $Y_i = \theta_1 + \theta_2 x_i + \epsilon_i$ as given in (2.2.4)-(2.2.6) under the restrictions $\theta_1 \geq 0, \theta_2 \leq 0$**

With no restrictions the least squares estimates (LSE) minimize

$$\rho((Y,Z),\beta) = -|y - (\theta_1 - \theta_2 z)|^2.$$

They LSE exist and are $\hat{\theta}_1 = \bar{y} - \hat{\theta}_2 \bar{z}$ and $\hat{\theta}_2 = \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^n (z_i - \bar{z})^2}$ (see BD pg. 100) . The LSE may fall into the restricted region and the problem is completed

- However, if the LSE do not fall in to the restricted region, then they must be on the boundary since $\rho((Y,Z),\beta) = -|y - (\theta_1 - \theta_2 z)|^2$ cannot reach a local minimum in an interior point of the restricted region with a non-vanishing derivative.

- Define the set Γ at the boundary where

$$\Gamma = \{(\theta_1, 0) : \theta_1 \geq 0\} \cup \{(0, \theta_2); \theta_2 \leq 0\}.$$

Minimizing the least square function over Γ,

$$\min_{\Gamma} \sum_{i=1}^n (y_i - \theta_1 - \theta_2 z_i)^2$$

$$= \min\{\min_{\theta_1 \geq 0} \sum_{i=1}^n (y_i - \theta_1)^2, \min_{\theta_2 \leq 0} \sum_{i=1}^n (y_i - \theta_2 z_i)^2\}.$$

Let $\theta_1^* = \bar{y}$ be the minimizer of $\sum_{i=1}^n (y_i - \theta_1)^2$ and $\theta_2^* = \frac{\sum z_i y_i}{\sum z_i^2}$ be the minimizer of $\sum_{i=1}^n (y_i - \theta_2 z_i)^2$ at the respective boundary subspace ($\theta_2 = 0$ or $\theta_1 = 0$).

- There are four cases in the set Γ:

(a) $\theta_1^* \geq 0, \theta_2^* \leq 0$

In this case, both estimates satisify the conditions (under Γ), therefore we check where the minimum is achieved above (either for $\sum_{i=1}^n (y_i - \theta_1)^2$ or $\sum_{i=1}^n (y_i - \theta_2 z_i)^2$).

$$(\hat{\theta}_1, \hat{\theta}_2) = (0, \theta_2^*) \quad if \quad \sum_{i=1}^n (y_i - \theta_1^*)^2 \geq \sum_{i=1}^n (y_i - \theta_2^* z_i)^2$$
$$= (\theta_1^*, 0) \quad o.w.$$

(In other words, we deed to check both boundaries to see where the minimum is achieved).

(b) $\theta_1^* \leq 0, \theta_2^* \leq 0$
In this case, θ_1^* does not satisfy the condition under Γ, therefore it is set to 0.

$$
\begin{aligned}
(\hat{\theta}_1, \hat{\theta}_2) &= (0, \theta_2^*) \quad if \quad \sum_{i=1}^{n} (y_i)^2 \geq \sum_{i=1}^{n} (y_i - \theta_2^* z_i)^2 \\
&= (0, 0) \quad o.w.
\end{aligned}
$$

(c) $\theta_1^* \geq 0, \theta_2^* \geq 0$
In this case, θ_2^* does not satisfy the condition under Γ, therefore it is set to 0.

$$
\begin{aligned}
(\hat{\theta}_1, \hat{\theta}_2) &= (\theta_1^*, 0) \quad if \quad \sum_{i=1}^{n} (y_i - \theta_1^*)^2 \leq \sum_{i=1}^{n} y_i^2 \\
&= (0, 0) \quad o.w.
\end{aligned}
$$

(d) $\theta_1^* \leq 0, \theta_2^* \geq 0$

$$
(\hat{\theta}_1, \hat{\theta}_2) = (0, 0)
$$

To get full credit for the problem, you need to examine the value of the least squares function ρ to examine where the minimum is achieved.

6. **BD 2.3.35 Let $g(x) = \frac{1}{\pi}(1+x^2),\ x \in R$ be the Cauchy density. Let X_1 and X_2 be i.i.d. with density $g(x-\theta),\ \theta \in R$. Let x_1 and x_2 be the observations and set $\Delta = \frac{1}{2}(x_1 - x_2)$. Let $\hat{\theta} = \arg\max L_X(\theta)$ be "the" MLE.**

   The log likelihood is

   $$l_x(\theta) = -2\log(\pi) - \log(1 + (x_1-\theta)^2) - \log(1+(x_2-\theta)^2).$$

   Taking the derivative,

   $$\frac{\partial l_x(\theta)}{\partial \theta} = -2\frac{(x_1-\theta)}{1+(x_1-\theta)^2} - 2\frac{(x_2-\theta)}{1+(x_2-\theta)^2}$$

   $$= 2(x_1-\theta)(1+(x_2-\theta)^2) + 2(x_2-\theta)(1+(x_1-\theta)^2).$$

   Rewriting,

   $$= 2(x_1-\theta) + 2(x_2-\theta) + 2[(x_1-\theta)(x_2-\theta)^2 + (x_2-\theta)(x_1-\theta)^2]$$

   $$= 4(\bar{x}-\theta) + 4[(x_1-\theta)(x_2-\theta)(\bar{x}-\theta)]$$

   $$= 4(\bar{x}-\theta)[1+(x_1-\theta)(x_2-\theta)].$$

   Noting that $x_1-\theta = \bar{x}+\Delta-\theta$ and $x_2-\theta = \bar{x}-\Delta-\theta$, where $\Delta = \frac{1}{2}(x_1-x_2)$, then plugging in above,

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

   Therefore, setting the derivative to zero, the MLE $\hat{\theta}$ solves

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

   - When $|\Delta| \le 1$, the right factor is strictly positive and the only solution is $\hat{\theta} = \bar{x}$. The second derivative with respect to θ is:

     $$-2[(\bar{x}-\theta)^2 - \Delta^2 + 1] - 4(\bar{x}-\theta)^2 = -6(\bar{x}-\theta)^2 + 2\Delta^2 - 2.$$

     The second derivative at $\bar{x}$ is negative since $|\Delta|^2 \le 1$.
     Alternatively, can also show that a maximum is achieved $\bar{x}$ because $l'_x(\theta) > 0$ for $\theta < \bar{x}$ and $l'_x(\theta) < 0$ for $\theta > \bar{x}$

   - When $|\Delta| > 1$, then there are two roots $\hat{\theta} = \bar{x} \pm \sqrt{(\Delta^2 - 1)}$. We need to show that the second derivative is negative at these points. Since $(\bar{x}-\hat{\theta})^2 = (\Delta^2 - 1)$ for both roots, the second derivative is $-4\Delta^2 + 4 < 0$, since $|\Delta| > 1$, and both roots are maximums.

- Not required but graphs can show the behavior of this likelihood. Define $\hat{\theta}_1 = \bar{x} - \sqrt{(\Delta^2 - 1)}$, $\hat{\theta}_2 = \bar{x}$ and $\hat{\theta}_3 = \bar{x} + \sqrt{(\Delta^2 - 1)}$. Graphs below using example parameter values for the likelihood and derivative of the log likelihood. The lines are for $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$. The left graph is an example of the case $|\Delta| \leq 1$, the middle graph is an example of the case $|\Delta| > 1$ and the right graph shows the same example for $|\Delta| > 1$ but for the first derivative (numerator only since the denominator is always positive).