1. (10 pt) Suppose we take one observation, X, from the discrete distribution,

<table>
<thead>
<tr>
<th>x</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr(X = x</td>
<td>θ)</td>
<td>(1−θ)/4</td>
<td>θ/12</td>
<td>1/2</td>
<td>(3−θ)/12</td>
</tr>
</tbody>
</table>

Find an unbiased estimator of θ. Obtain the maximum likelihood estimator (MLE) \( \hat{θ}(x) \) of θ and show that it is not unique. Is any choice of MLE unbiased?

solutions:
(a) Let \( T(2) = 4, T(x) = 0 \) when \( x \neq 2 \). Then

\[
E[T(X)] = 4 \times \frac{θ}{4} + 0 \times \left( \frac{1-θ}{4} + \frac{θ}{12} + \frac{1}{2} + \frac{3-θ}{12} \right) = θ
\]

, which is an unbiased estimator of θ.

(b) Denote the MLE \( \hat{θ} \). Since there’s only one observation:

\[
\begin{align*}
\hat{θ}(−2) &= \text{arg max}((1−θ)/4)) = 0 \\
\hat{θ}(−1) &= \text{arg max}(θ/12) = 1 \\
\hat{θ}(0) &= \text{arg max}(1/2) = \forall c ∈ (0, 1) \\
\hat{θ}(1) &= \text{arg max}((3−θ)/12) = 0 \\
\hat{θ}(2) &= \text{arg max}(\frac{θ}{4}) = 1
\end{align*}
\]

Since \( \hat{θ}(0) \) can be multiple values, it’s not unique and all the MLE’s are biased since

\[
E[\hat{θ}] = 0 \times (1−θ)/4 + 1 \times θ/12 + \hat{θ}(0) \times \frac{1}{2} + 0 \times (3−θ)/12 + 1 \times \frac{θ}{4} ≠ θ
\]

2. (10 pt) Bickel & Doksum pg 153, problem 2.3.8

solutions:
(a) It is enough to show that the negative log likelihood function \( \ell(\theta) = - \log \prod f(x_i | \theta) \) is a strictly convex function of \( \theta \in \mathbb{R}^p \). Since it’s the sum of the negative log likelihoods for each \( X_j \), and a sum of strictly convex functions is strictly convex, it’s enough to consider a single observation \( n = 1 \).

To show the negative log likelihood is strictly convex it is enough to show that its Hessian, the matrix \( H_{ij} = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\theta) \) of second derivatives, is positive-definite everywhere (except possibly at \( \hat{\theta} = x \)). Let’s compute the necessary derivatives:

\[
\ell(\theta) = - \log c + \left( \sum_{i=1}^{p} (x_i - \theta_i)^2 \right)^{\alpha/2}
\]

\[
\frac{\partial}{\partial \theta_k} \ell(\theta) = -\alpha \left( \sum_{i=1}^{p} (x_i - \theta_i)^2 \right)^{(\alpha-2)/2} (x_k - \theta_k)
\]

\[
\frac{\partial^2}{\partial \theta_k^2} \ell(\theta) = \alpha \left( \sum_{i=1}^{p} (x_i - \theta_i)^2 \right)^{(\alpha-4)/2} + \alpha(\alpha - 2) \left( \sum_{i=1}^{p} (x_i - \theta_i)^2 \right)^{(\alpha-4)/2} (x_k - \theta_k)^2
\]

\[
\frac{\partial^2}{\partial \theta_i \partial \theta_k} \ell(\theta) = \alpha \left( \sum_{i=1}^{p} (x_i - \theta_i)^2 \right)^{(\alpha-4)/2} (\alpha - 2)(x_i - \theta_i)(x_k - \theta_k)
\]

The Hessian \( H = \alpha | x - \theta |^{\alpha-4} A \) is the product of a constant positive factor \( \alpha | x - \theta |^{\alpha-4} \) and a matrix \( A \) whose on- and off-diagonal entries are:

\[
A_{kk} = \sum_{i=1}^{p} (x_i - \theta_i)^2 + (\alpha - 2)(x_k - \theta_k)^2
\]

\[
A_{kj} = (\alpha - 2)(x_k - \theta_k)(x_j - \theta_j)
\]

If we introduce the notation \( \Delta = (x - \theta) \in \mathbb{R}^p \) for the vector with components \( \Delta_i = (x_i - \theta_i) \), and \( I_p \) for the \( p \times p \) identity matrix, we can write \( A \) in the form

\[
A = |\Delta|^2 I_p + (\alpha - 2)\Delta \Delta^T
\]

The matrix \( A \) staisfies \( A\Delta = \lambda \Delta \) with eigenvalue \( \lambda = |\Delta|^2(\alpha - 1) \), strictly positive since \( \alpha > 1 \) (except at \( \Delta = 0 \), i.e., \( \theta = \hat{\theta} = x \), which is okay). The other eigenvectors are orthogonal to \( \Delta \), all with eigenvalues \( \lambda' = |\Delta|^2 \), which are also strictly positive. Thus \( A \) is positive-definite and so is the Hessian, \( H = \alpha |\Delta|^{\alpha-4} A \).
(b) First consider the case where $\alpha = 1$ in dimension $p = 1$, with $n = 2m$ even. WLOG order the data $x_1 \leq x_2 \leq \cdots \leq x_n$. The log likelihood function is given by

$$
\ell(\theta) = n \log c(\alpha) - \sum |x_i - \theta|,
$$
a continuous function whose derivative does not exist at the data points $\theta \in \{x_i\}$ and which elsewhere satisfies

$$
\frac{d}{d\theta} \ell(\theta) = -\sum \frac{d}{d\theta} |x_i - \theta| = \left[ \sum_{x_i < \theta} (-1) \right] + \left[ \sum_{x_i > \theta} (+1) \right]
$$

(note that the derivative of $|x - \theta|$ is $-1$ on the interval $\theta \in (-\infty, x)$, $+1$ on the interval $\theta \in (x, \infty)$, and undefined at the point $\theta = x$). Thus $\ell(\theta)$ is increasing when $\theta < x_m$, when more than half the $\{x_i\}$ exceed $\theta$, and is decreasing when $\theta > x_{m+1}$, when fewer than half the $\{x_i\}$ exceed $\theta$. In the interval $x_m < \theta < x_{m+1}$ the derivative is zero, so $\ell(\theta)$ is constant there and equal to its maximum value.

In case $n = 2m - 1$ is odd, the same argument shows that $\ell(\theta)$ achieves a unique maximum at the median $x_m$.

In dimension $p > 2$ a similar argument holds, only $\hat{\theta}$ now should be any value in the “median rectangle” (or “median block”) where each of its components is a median of the corresponding components of the $x_i$’s.

3. (10 pt) Let $(X_1, \cdots, X_n)$ be a random sample from the uniform distribution on the interval $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, where $\theta \in \mathbb{R}$ is unknown. Let $X_{(j)}$ be the $j^{th}$ order statistic.

(a) Show that $(X_{(1)} + X_{(n)})/2$ is strongly consistent for $\theta$, i.e., that

$$
\lim_{n \to \infty} (X_{(1)} + X_{(n)})/2 = \theta \text{ a.s.}
$$

(b) Show that $\overline{X}_n := (X_{(1)} + \cdots + X_{(n)})/n$ is $L^2$ consistent.

solution:

(a) For any $\epsilon > 0$,

$$
P(|X_{(1)} - (\theta - \frac{1}{2})| > \epsilon) = P(X_{(1)} > \epsilon + (\theta - \frac{1}{2}))
$$

$$
= [P(X_1 > \epsilon + (\theta - \frac{1}{2}))]^n
$$

$$
= (1 - \epsilon)^n
$$
and

\[ P(|X(n) - (\theta + \frac{1}{2})| > \epsilon) = P(X(n) < -\epsilon + (\theta + \frac{1}{2})) \]
\[ = [P(X_1 < -\epsilon + (\theta + \frac{1}{2}))]^n \]
\[ = (1 - \epsilon)^n \]

Since \( \sum_{i=1}^n (1 - \epsilon)^n < \infty \), by the corollary of Borel-Cantelli lemma, we conclude that \( \lim_{n \to \infty} X(1) = \theta - \frac{1}{2} \) a.s. and \( \lim_{n \to \infty} X(n) = \theta + \frac{1}{2} \) a.s. Hence \( \lim_{n \to \infty} (X(1) + X(n))/2 = \theta \) a.s.

(b) \( \overline{X}_n := (X(1) + \cdots + X(n))/n = (X_1 + \cdots + X_n)/n \) since \( E(X_1) = \theta, E[\overline{X}_n] = \theta \)
and

\[ \text{Var}[\overline{X}_n] = E[\overline{X}_n - \theta^2] = \frac{1}{n} E[(X_1 - \theta)^2] = \frac{1}{12n} \xrightarrow{n \to \infty} 0. \]

Therefore, \( \overline{X}_n \) is \( L^2 \) consistent.

4. (10 pt) Let \( \{X_i\} \text{ i.i.d } \text{N}(\mu, \sigma^2) \) be a random sample from the normal distribution with unknown mean \( \mu \in \mathbb{R} \) and known variance \( \sigma^2 > 0 \). For fixed \( t \neq 0 \), find the Uniform Minimum-Variance Unbiased Estimator (UMVUE) of \( e^{t\mu} \) and show that its variance is larger than the Cramér-Rao lower bound. Show that the ratio of its variance to the Cramér-Rao lower bound converges to 1 as \( n \to \infty \).

solutions:

(a) The sample mean \( \overline{X} \) is complete and sufficient for \( \mu \). Since \( E[e^{t\overline{X}}] = e^{\mu t + \sigma^2 t^2/(2n)} \), the UMVUE of \( e^{t\mu} \) is \( T(X) = e^{-\sigma^2 t^2/(2n) + t\overline{X}} \).

The Fisher information \( I(\mu) = n/\sigma^2 \). Then the Cramér-Rao lower bound is \( (\frac{1}{2n} e^{t\mu})^2 / I(\mu) = \sigma^2 t^2 e^{2t\mu} / n \). On the other hand,

\[ \text{Var}(T) = e^{-\sigma^2 t^2/n} E[e^{2t\overline{X}} - e^{2t\mu}] = (e^{\sigma^2 t^2/n} - 1)e^{2t\mu} > \frac{\sigma^2 t^2 e^{2t\mu}}{n} \]

, the Cramér-Rao lower bound.

(b) The ratio of the variance of the UMVUE over the Cramér-Rao lower bound is \( (e^{\sigma^2 t^2/n} - 1)/\sigma^2 t^2/n \), which converges to 1 as \( n \to \infty \), since \( \lim_{x \to 0} (e^x - 1)/x = 1 \)
5. (10 pt) Let \( X \) have density function \( f(x|\theta), x \in \mathbb{R}^d \), and let \( \theta \) have (proper) prior density \( \pi(\theta) \). Let \( \delta^\pi(x) \) denote the Bayes estimate of \( \theta \) under \( \pi(\theta) \) for squared-error loss (i.e., \( \delta^\pi(x) := E[\theta | X = x] \)) and suppose \( \delta^\pi \) has finite Bayes risk \( r(\pi, \delta^\pi) = E[(\delta^\pi(X) - \theta)^2] \). Show that for any other estimator \( \delta(x) \),

\[
r(\pi, \delta) - r(\pi, \delta^\pi) = \int (\delta(x) - \delta^\pi(x))^2 f(x)dx
\]

where \( f(x) \) is the marginal density of \( X \). If \( f(x|\theta) \) is normal \( \mathcal{N}(\theta, 1) \), consider the collection of estimators of the form \( \delta(X) := cX + d \). Show that whenever \( 0 \leq c < 1 \) these estimators are all proper Bayes estimators, and hence admissible [Hint: Find a prior \( \pi \) for which \( \delta^\pi(X) = cX + d \)]. Show that if \( c > 1 \) the resulting estimator is inadmissible.

solutions:

(a) \[
\begin{align*}
r(\pi, \delta) - r(\pi, \delta^\pi) &= \int \int [(\theta - \delta(x))^2 - (\theta - \delta^\pi(x))^2] f(x|\theta) \pi(\theta) d\theta dx \\
&= \int \int [(\theta - \delta(x))^2 - (\theta - \delta^\pi(x))^2] \pi(\theta|x) f(x) d\theta dx \\
&= \int f(x) \left[ \int [(\theta - \delta(x))^2 - (\theta - \delta^\pi(x))^2] \pi(\theta|x) f(x) d\theta \right] dx \\
&= \int f(x) [\delta^2(x) - \delta^\pi^2(x) + 2\delta^\pi^2(x) - 2\delta \delta^\pi] dx \\
&= \int (\delta(x) - \delta^\pi(x))^2 f(x) dx
\end{align*}
\]

(b) Take \( \pi(\theta) = \mathcal{N}(\mu, \tau^2) \), where \( \tau \) is the precision parameter. Then

\[ E[\theta|x] = \frac{\tau^2}{\tau^2 + 1} x + \frac{1}{\tau^2 + 1} \mu \]

Let \( c = \frac{\tau^2}{\tau^2 + 1}, d = \frac{\mu}{\tau^2 + 1} \). Whenever \( 0 \leq c < 1 \), we can always get \( \mu, \tau^2 \) and \( cX + d \) is a Bayes estimator with proper prior \( \mathcal{N}(\mu, \tau^2) \).

(c) When \( c > 1 \), \( \text{MSE}(cX + d) = R(\theta, \delta_c) = (c\theta + d - \theta)^2 + c^2 \text{Var}(X) > 1 \), while \( \text{MSE} \) for \( \delta_0(X) = X \) is 1, so the estimator \( \delta_0(X) = X \) is uniformly better than \( cX + d \). Similarly when \( c = 1, d \neq 0 \), \( X \) is uniformly better than \( X + d \).

6. (10 pt) Bickel & Doksum pg 197, problem 3.2.1

solutions:
(a) One way to show this is to notice that it is a limiting case of homework problem 1.2.14 from the first homework set, taking the limit as \( \tau \) goes to infinity. Then the posterior distribution of \( \mu \) approaches a normal distribution with mean \( \bar{x} \) and standard deviation \( \sigma^2 \). If we were "bad" and interpreted a frequentist confidence interval for \( \mu \) in a Bayesian way, we would be implicitly assuming that our prior knowledge of \( \mu \) was utterly nil—that all real values were equally likely in our minds (a questionable assumption, since it is not a proper prior). Here is a slightly more detailed development of the solution.

We know that if \( X_j \overset{iid}{\sim} \text{No}(\theta, \sigma^2) \), then the sufficient statistic \( \bar{X} \sim \text{No}(\theta, \sigma^2/n) \). So the likelihood function for \( \theta \) is given by:

\[
\ell(\theta) = f(\bar{x}|\theta) = \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left(\frac{-1}{2\sigma^2/n}(\bar{x} - \theta)^2\right)
\]

The posterior distribution of \( \theta \) given \( X \) is given by:

\[
\pi(\theta|X) \propto \pi(\theta)f(X|\theta)
\]

\[
\pi(\theta|X) \propto 1 \cdot \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left(\frac{-1}{2\sigma^2/n}(\bar{x} - \theta)^2\right)
\]

This function, which came about because it was the pdf of \( \bar{x} \), is also the kernel of a normal density for \( \theta \), and it is easy to see that the mean is \( \bar{x} \) and the variance is \( \sigma^2/n \). So \( \theta|X \sim \text{No}(\bar{x}, \sigma^2/n) \).

Under squared error loss, the Bayes estimator is the posterior mean, which in this case is \( \bar{x} \).

7. (10 pt) Bickel & Doksum pg 197, problem 3.2.4

solutions:

(a) We begin with \( \lambda = \frac{\theta}{1-\theta} \). If we take the prior on \( \lambda \) to be the improper prior \( \pi_\lambda(\lambda) = 1 \), we find the induced prior on \( \theta \) as follows:

\[
\pi_\theta(\theta) \propto \pi_\lambda\left(\frac{\theta}{1-\theta}\right) \frac{d}{d\theta}\left[\frac{\theta}{1-\theta}\right] = 1 \cdot \frac{1}{(1-\theta)^2} = (1-\theta)^{-2}
\]

Using this prior for \( \theta \), we find the posterior distribution of \( \theta \) thus:

\[
\pi(\theta|X) \propto \pi(\theta)\pi(X|\theta)
= (1-\theta)^{-2} (\theta^S(1-\theta)^{n-S})
= \theta^S(1-\theta)^{n-S-2}
\]
This is the kernel of a $\text{Be}(S + 1, n - S - 1)$ density, so long as the conditions are met for such a density to exist, which are that $S + 1 > 0$ and $n - S - 1 > 0$. The first of these will always be true, but the second will only be true when $n - S > 1$, i.e., when at least two “failures” are observed. (Should we have $n - S = 1$ or $n - S = 0$, then the expression $\theta^S(1 - \theta)^{n-S-2}$ would not have a finite integral, so the posterior distribution on $\theta$ would be improper, a decided no-no.)

Under the squared error loss function, the Bayes estimator is the posterior mean of $\theta$, which for a $\text{Be}(S + 1, n - S - 1)$ density is known to be $\frac{S+1}{n}$. A finite when $S < n - 2$.

- Note that some students calculated the Bayes rule for $\lambda$ instead of for $\theta$, which I also marked correct. The Bayes rule for $\lambda$ is $\frac{S+1}{n-S-2}$, exists when $S < n - 1$, finite when $S < n - 2$. 

\[ \frac{S+1}{n} \]