Math Stats

Instructor: Jonathan Niles-Weed

Homework 5
Due: Sunday October 16, 11:59pm via NYU Gradescope

Collaborated with Jonah Potzcobutt, and Andre Chen.

1. A parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ of continuos distri utions with corresponding densities p_θ is known as a exponential family if there eists functions $h : \mathbb{R} \to \mathbb{R}, T : \mathbb{R} \to \mathbb{R}$, and $A : \Theta \to \mathbb{R}$ such that

$$p_\theta(x) = h(x)e^{\theta T(x) - A(\theta)} \quad \forall \theta \in \Theta$$

(a) Show that h and A in the above definition are not unique (so that the model is not identifiable as written).

We can take the integral of the pdf given above to show the desired property:

$$\begin{aligned} p_\theta(x) &= h(x)e^{\theta T(x) - A(\theta)} dx \\ \int p_\theta(x) &= \int h(x)e^{\theta T(x) - A(\theta)} dx \\ 1 &= e^{-A(\theta)} \int h(x)e^{\theta T(x)} dx \end{aligned} \tag{1}$$

We can arbitrarily use new functions $A(\theta)', h(x)'$ such that $e^{-A(\theta)'} = ce^{-A(\theta)}$ and $h(x)' = \frac{1}{c}h(x)$ thus satisfies the above inequality while also being different members of the exponential family, thus showing that without specific information, an exponential model expressed in its general form is not identifiable.

(b) Show that the set of Gaussian distributions with variance 1 forms an exponential family. Identify T, h, and A in this example.

We can show this property by beginning with the definition of the pdf of a uni-variate Gaussian:

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

Since we have $\sigma^2 = 1$ we can simplify the above:

$$\begin{aligned} \mathcal{N}(\mu, 1) &= \frac{1}{\sqrt{2\pi}} \exp -\frac{x^2 + \mu^2 - 2x\mu}{2} \\ &= \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \exp -\frac{\mu^2 - 2x\mu}{2} \end{aligned} \tag{2}$$

We can pick $h(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2), T(x) = x, A(\theta) = \frac{\mu^2}{2}$ to reach our desired formulation of the gaussian as a member of the exponential family.

(c) Show that the set of all distributions on $[1, +\infty)$ with densities of the form $p_\theta(x) = (\theta - 1)x^{-\theta}$ for $\theta > 1$ forms an exponential family. Identify T, h, and A.

We can use the same approach as we did in the last problem starting with the given pdf:

$$p_\theta(x) = (\theta - 1)x^{-\theta}$$

We can begin by exponentiating the equation while simultaneously taking the log at the same time:

$$\begin{aligned} p_\theta(x) &= (\theta - 1)x^{-\theta} \\ &= e^{log(\theta-1)-\theta log(x)} \\ &= e^{\theta(-log(x))-log(\frac{1}{\theta-1})} \end{aligned} \tag{3}$$

We can choose $h(x) = 1, T(x) = e^{-log(x)}, A(\theta) = e^{log(\frac{1}{\theta-1})}$

(d) Show that for any exponential family, $A(\theta) = log \int h(x)e^{(x)}dx$. Conclude that for all $\theta \in \Theta$ $A'(\theta) = \mathbb{E}_\theta T(X)$, where $\mathbb{E}_\theta$ is the expectation when $X \sim P_\theta$. (You may assume that it is valid to interchange differentiation and integration, and that is open.)

We have:

$$\int p_\theta(x)dx = 1$$

We can use the identity of the exponential family:

$$\begin{aligned} \int p_\theta(x)dx &= 1 \\ \int h(x)e^{\theta T(x)-A(\theta)}dx &= 1 \\ e^{-A(\theta)} \int h(x)e^{\theta T(x)}dx &= 1 \\ log \int h(x)e^{\theta T(x)}dx &= A(\theta) \quad \textit{multiply both sides by } e^{A(\theta)} \textit{ and take log} \end{aligned} \tag{4}$$

We can now take the derivative of each side with respect to θ using the chain rule:

$$
\begin{aligned}
A'(\theta) &= \frac{\partial log \int h(x)e^{\theta T(x)}dx}{\partial\theta} \\
&= \frac{\frac{\partial}{\partial\theta}\int h(x)e^{\theta T(x)}dx}{\int h(x)e^{\theta T(x)}dx} \\
&= \frac{\int T(x)h(x)e^{\theta T(x)}dx}{\int h(x)e^{\theta T(x)}dx} \quad \textit{exch. int and diff in num} \\
&= e^{-A(\theta)}\int T(x)h(x)e^{\theta T(x)}dx \\
&= \int T(x)h(x)e^{\theta T(x)-A(\theta)}dx \\
&= \mathbb{E}_\theta T(x) \quad \square
\end{aligned} \tag{5}
$$

(e) Show that in an exponential family, if $X_1, \ldots, X)n \sim P_\theta$ are i.i.d., then any solution to

$$
\frac{1}{n}\sum_{i=1}^{n} T(X_i) = \mathbb{E}_\theta T(X)
$$

is a maximum liklihood estimator. (Hint: you may use the fact the following two facts: A is a differentiable convex function of θ, and if f is a differentiable convex function and $f'(\theta) = 0$, then θ is a global minimum of f.)

We can use the log-likelihood function here:

$$
\mathcal{L}(\theta|x) = \sum_{i=1}^{n} log(p_\theta(x_I)) = \sum_{i=1}^{n} log(h(x_i)) + \theta T(x_i) - A(\theta)
$$

We know that the above expression is concave and differentiable as $log(h(x_i))$ is constant and $\theta T(x_i)$ is linear with respect to θ and $A(\theta)$ is convex, thus $-A(\theta)$ is concave.

We can take the first order derivative and set to 0 to calculate the MLE:

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\mathcal{L}(\theta|x) = \sum_{i=1}^{n} T(x_i) - A'(\theta) = 0 \\
\sum_{i=1}^{n} (T(x_i) - \mathbb{E}_\theta T(x)) = 0 \\
\frac{1}{n}\sum_{i=1}^{n} T(x_i) = \mathbb{E}_\theta T(x)
\end{aligned} \tag{6}
$$

Therefore, any solution to the above is an MLE.

Note the similarity between (5.7) and (5.2). For exponential families, maximum likelihood estimation is equivalent to a method-of-moments-like procedure, with the function $T(X)$ used in place of X^k.

2. Let $X_1, \dots, X_n \sim Unif([0, \theta])$ be i.i.d. for some $\theta > 0$.

(a) Compute the MLE of θ

We have the pdf of a Univorm distribution as: $\frac{1}{b-a}$ so in our case this is $\frac{1}{\theta - 0}$. Therefore the likelihood is:

$$\mathcal{L}(\theta | X_1, \dots, X_n) = \prod_{i=1}^{n} P_\theta(x_i)$$

We can order our observations in ascending fashion such that $x_i, \dots, x_n$ satisfy $0 \leq x_1 \leq \dots \leq x_n \leq \theta$ where $x_n = \max(x_1, \dots, x_n)$ We can then rewrite our likelihood as:

$$\mathcal{L}(\theta | X_1, \dots, X_n) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{0 \leq x \leq \theta} = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{1}_{0 \leq x \leq \theta}$$

Which is the 0 for all $\theta < x_n$

Therefore, we have $\hat{\theta} = argmax \mathcal{L}(\theta | x_n) = x_n$

(b) Use the method of moments to estimate θ

We can use the method of moments with $K = 1$:

$$\hat{m}_k = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Then:

$$\begin{aligned}
\mathbb{E}_\theta(x) &= \int_0^\theta x \frac{1}{\theta} dx \\
&= \frac{1}{2\theta} x^2 0^\theta \\
&= \frac{\theta}{2}
\end{aligned} \tag{7}$$

We can then solve for θ:

$$\frac{1}{n} \sum_{i=1}^{n} x_i = \frac{\theta}{2} \to \hat{\theta} = \frac{2}{n} \sum_{i=1}^{n} x_i$$

(c) Show that if $n \geq 3$ then with positive probability the method of moements applied to this example gives an estimator $\hat{\theta}$ which is unrealistic, in the sense that the statistician can be sure that $Unif([0, \hat{\theta}])$ did not produce the data.

$$\mathbb{P}\left\{\frac{2}{n}\sum_{i=1}^{n} x_i \leq \hat{\theta}_{MLE}\right\} = \mathbb{P}\left\{\frac{2}{n}\sum_{i=1}^{n} x_i \leq x_n\right\} \quad where\ x_n = \max_{i=1,\ldots,n} x_i$$
$$= \mathbb{P}\left\{\sum_{i=1}^{n-1} x_i \leq x_n(1-\frac{2}{n})\right\} \tag{8}$$

We can then investigate cases where $n \geq 3$:

$$\mathbb{P}\left\{\sum_{i=1}^{n-1} x_i \leq x_n(c)\right\} \quad where\ c > 0 \tag{9}$$

Then anytime we have $\mathbb{P}\left\{\hat{\theta}_{MoM} \leq \hat{\theta}_{MLE}\right\} = \mathbb{P}\left\{\hat{\theta}_{Not-Relevant}\right\} > 0$

(d) Show that both the MLE and the method of moments estimators are consistent. (Hint: in the case of the MLE, it may be useful to use the fact that for indepdent random variables $X_1, \ldots, X_n$,

$$\mathbb{P}\left\{\max_{i=1,\ldots,n} X_i \leq t\right\} = \prod_{i=1}^{n} \mathbb{P}\left\{X_i \leq t\right\}$$

which follows form the fact that $max_{i=1,\ldots,n} X_i \leq t$ if and only if $X_i \leq t$ for all i.)

We can start with the following definition to show that the MLE estimator is consistent:

$$\mathbb{P}\left\{|\max_{i=1,\ldots,n} X_i - \theta| \geq t\right\} = 1 - \mathbb{P}\left\{|\max_{i=1,\ldots,n} X_i - \theta| \leq t\right\}$$
$$= 1 - \mathbb{P}\left\{\max_{i=1,\ldots,n} X_i \leq t+\theta\right\} - \mathbb{P}\left\{\max_{i=1,\ldots,n} X_i \leq t-\theta\right\} \tag{10}$$
$$= 1 - \prod_{i=1}^{n} \mathbb{P}\left\{x_i \leq t\theta\right\} - \prod_{i=1}^{n} \mathbb{P}\left\{x_i \leq \theta - t\right\}$$

The probability product that $x_i \leq t+\theta$ is always 1, while the right probability with $x_i \leq t-\theta$ is in $[0,1]$ as $x \in [0,\theta]$ and $t > 0$. Therefore, we can show consistency as $n \to \infty$ then the expression becomes $\mathbb{P}\left\{|\hat{\theta} - \theta| \geq t\right\} \xrightarrow[n\to\infty]{} 1 - 1 + 0 = 0\ \forall t > 0$

Now for the Method of moments:

$$\mathbb{P}\left\{|2\frac{1}{n}\sum_{i=1}^{n} x_i - \theta| \geq t\right\} = 1 - \mathbb{P}\left\{2\frac{1}{n}\sum_{i=1}^{n} x_i - \theta \leq t\right\} - \mathbb{P}\left\{2\frac{1}{n}\sum_{i=1}^{n} x_i + \theta \leq t\right\}$$
$$= 1 - \mathbb{P}\left\{2\frac{1}{n}\sum_{i=1}^{n} x_i \leq t+\theta\right\} - \mathbb{P}\left\{2\frac{1}{n}\sum_{i=1}^{n} x_i \leq t-\theta\right\} \tag{11}$$

We can then use the same logic as we did for the MLE portion of the problem to show that the left probability is always 1, and the right side probability approaaches 0 asymptotically. Therefore, we have shown that both MLE and MoM are consistent estimators, as $\frac{2}{n}\sum_{i=1}^{n} x_i \xrightarrow[n\to\infty]{} \theta$

3. This exercise will explore the notation of "asymptotic efficiency" and give another justification for the use of the maximum likelihood estimator. We assume in this problem that each $\mathbb{P}_\theta \in \mathcal{P}$ is continuous with a smooth, strictly positive density p_θ. We write $s(\theta|x)$ for the score function:

$$s(\theta|x) = \frac{\partial}{\partial\theta}\ell(\theta|x) = \frac{\partial}{\partial\theta} log p_\theta(x)$$

(a) Show that:

$$s(\theta|\omega) = \sum_{i=1}^{n} s(\theta|x_i) = \frac{\frac{\partial}{\partial\theta}\prod_{i=1}^{n} p_\theta(x_i)}{\prod_{i=1}^{n} p_\theta(x_i)}$$

We can manipulate the problem statement to show the desired property:

$$\begin{aligned} s(\theta|\omega) &= \sum_{i=1}^{n} s(\theta|x_i) \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial\theta} log p_\theta(x) \\ &= \frac{\partial}{\partial\theta}\sum_{i=1}^{n} log p_\theta(x) \\ &= \frac{\partial}{\partial\theta} log \prod_{i=1}^{n} p_\theta(x) \\ &= \frac{\frac{\partial}{\partial\theta}\prod_{i=1}^{n} p_\theta(x)}{\prod_{i=1}^{n} p_\theta(x)} \qquad \square \quad \textit{chain rule} \end{aligned} \tag{12}$$

(b) Using part (a), show that for a suitably smooth function f of $\omega = (X_1, \ldots, X_n)$ for $X_1, \ldots, X_n$ i.i.d. we have:

$$\mathbb{E}_\theta f(\omega) s(\theta|\omega) = \frac{d}{d\theta}\{\mathbb{E}_\theta f(\omega)\}$$

Again, using some algebra we can show the desired property, but first we note the following:

$$\begin{aligned} \frac{d}{d\theta}\mathbb{E}_\theta f(\omega) &= \frac{d}{d\theta}\int f(x_i, \ldots, x_n) p_\theta(x_1, \ldots, x_n)\} \\ &= \frac{d}{d\theta}\int f(x_i, \ldots, x_n)\prod_{i=1}^{n} P_\theta(x_i)\} \end{aligned} \tag{13}$$

Which holds as we have a joint pdf of two indepdent random variables. We will start the rest of the problem from the hint (which I did not bother to write)

$$\begin{aligned}
\mathbb{E}_\theta f(\omega)s(\theta|\omega) &= \int f(x_i,\ldots,x_n)\frac{\frac{\partial}{\partial\theta}\prod_{i=1}^n p_\theta(x_i)}{\prod_{i=1}^n p_\theta(x_i)}\prod_{i=1}^n P_\theta(x_i)\} \\
&= \int f(x_i,\ldots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^n P_\theta(x_i)\} \\
&= \frac{\partial}{\partial\theta}\int f(x_i,\ldots,x_n)\prod_{i=1}^n P_\theta(x_i)\} \\
&= \frac{d}{d\theta}\{\mathbb{E}_\theta f(\omega)\} \quad \square
\end{aligned} \tag{14}$$

(c) Let $\hat{\theta}$ be any unbiased estimator of θ. Using the previous part, show that:

$$\mathbb{E}_\theta s(\theta|\omega) = 0 \quad \forall\theta\in\Theta$$

and

$$\mathbb{E}_\theta\left\{\hat{\theta}(\omega)s(\theta|\omega)\right\} = 1 \quad \forall\theta\in\Theta$$

To show the first part:

$$\begin{aligned}
\mathbb{E}_\theta s(\theta|\omega) &= \int \frac{\partial}{\partial\theta}\prod_{i=1}^n p_\theta(x_i)d\omega \\
&= \frac{\partial}{\partial\theta}\int\prod_{i=1}^n p_\theta(x_i)d\omega \\
&= \frac{\partial}{\partial\theta}\int_\omega p_\theta(\omega)d\omega \\
&= \frac{\partial}{\partial\theta}1 = 0
\end{aligned} \tag{15}$$

Now for the second part:

$$\begin{aligned}
\mathbb{E}_\theta\left\{\hat{\theta}(\omega)s(\theta|\omega)\right\} &= \frac{\partial}{\partial\theta}\int\hat{\theta}(\omega)\prod_{i=1}^n p_\theta(x_i)d\omega \\
&= \frac{\partial}{\partial\theta}\mathbb{E}_\theta\hat{\theta}(\omega) \\
&= \frac{\partial}{\partial\theta}\theta \\
&= 1
\end{aligned} \tag{16}$$

The above holds as $\hat{\theta}$ is an unbiased estimator.

(d) Conclude that for any $\lambda \in \mathbb{R}$ and any unbiased estimator $\hat{\theta}$

$$0 \leq \mathbb{E}_\theta(\lambda(\hat{\theta}(\omega) - \theta) - s(\theta|\omega))^2 = \lambda^2 Var_\theta(\hat{\theta}) + Var_\theta(s(\theta|\omega)) - 2\lambda$$

In particular, by choosing $\lambda = Var_\theta(s(\theta|\omega))$ and rearranging, show that any unbiased estimator $\hat{\theta}$ of θ satisfies:

$$Var_\theta(\hat{\theta}) \geq \frac{1}{Var_\theta(s(\theta|\omega))} = \frac{1}{nVar_\theta(s(\theta|X_1))}$$

This is known as the Cramer-Rao bound, and $Var_\theta(s(\theta|X_1))$ is known as the Fisher information and denoted $I(\theta)$. It gives a lower bound on the variance of any unbiased estimator. An unbiased estimator matching this bound is called efficient.

We have:

$$\begin{aligned}\mathbb{E}_\theta(\lambda(\hat{\theta}(\omega) - \theta) - s(\theta|\omega))^2 \geq 0 \quad as\ \mathbb{E}_\theta(x^2) \geq 0 \\ \mathbb{E}_\theta(\lambda^2(\hat{\theta}(\omega) - \theta)^2 - 2\lambda(\hat{\theta}(\omega) - \theta)s(\theta|\omega) + s(\hat{\theta}|\omega)^2) \geq 0 \\ \lambda^2\mathbb{E}_\theta(\hat{\theta}(\omega) - \theta)^2 - 2\lambda 1 + \mathbb{E}_\theta s(\hat{\theta}|\omega)^2) \geq 0\end{aligned} \tag{17}$$

As $\mathbb{E}_\theta(\hat{\theta}(\omega) - \theta)s(\theta|\omega) = 1$. Looking at the $\mathbb{E}_\theta s(\hat{\theta}|\omega)^2)$ term:

$$Var(s(\hat{\theta}|\omega)) = \mathbb{E}_\theta s(\hat{\theta}|\omega)^2) - \mathbb{E}_\theta s(\hat{\theta}|\omega))^2 = \mathbb{E}_\theta s(\hat{\theta}|\omega)^2$$

So we now have:

$$\lambda^2\mathbb{E}_\theta(\hat{\theta}(\omega) - \theta)^2 + Var_\theta(s(\theta|\omega)) - 2\lambda \geq 0$$

Choosing $\lambda = Var_\theta(s(\theta|\omega))$ we can manipulate the above to yield:

$$\begin{aligned}Var_\theta(\theta) &= \frac{1}{Var_\theta(s(\theta|\omega))} \\ &= \frac{1}{Var_\theta(\sum_{i=1}^n s(\theta|x_i))} \\ &= \frac{1}{nVar_\theta(\sum_{i=1}^n s(\theta|x_1))}\end{aligned} \tag{18}$$

4. Suppose that $Y_1, \ldots, Y_n$ satisfy:

$$Y_i = \beta x_i + \epsilon_i$$

Where $x_1, \ldots, x_n$ are fixed and known, and $\epsilon_1, \ldots, \epsilon_n \sim \mathcal{N}(0, \sigma^2)$ are i.i.d. and β and σ^2 are unknown.

(a) Compute the MLE $\hat{\beta}$ of β

We have chosen to model know that $y_i \sim \mathcal{N}(\beta x_i, \sigma^2)$ therefore the log-likelihood is proportional to:

$$log\mathcal{L}(\theta|y_1, \ldots, y_n) \propto \frac{1}{2}\sum_{i=1}^{n}(y_i - \beta x_i)^2$$

We can take the derivative with respect to β and set the resulting expression to 0 to solve for the MLE solution:

$$\begin{aligned}
\frac{\partial log\mathcal{L}(\theta|y_1, \ldots, y_n)}{\partial \beta} &= 0 \\
\sum_{i=1}^{n}(y_i - \beta x_i)(x_i) &= 0 \\
\sum_{i=1}^{n} y_i x_i &= \beta \sum_{i=1}^{n} x_i^2 \\
\frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} &= \hat{\beta} \qquad \square
\end{aligned} \tag{19}$$

(b) Compute $\mathbb{E}\hat{\beta}$ and $Var(\hat{\beta})$

We can compute the expectation for $\mathbb{E}\hat{\beta}$ as follows:

$$\begin{aligned}
\mathbb{E}\hat{\beta} &= \mathbb{E}\frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \\
&= \frac{1}{\sum_{i=1}^{n} x_i^2}\mathbb{E}\sum_{i=1}^{n} y_i x_i \\
&= \frac{1}{\sum_{i=1}^{n} x_i^2}\sum_{i=1}^{n} \mathbb{E} y_i x_i \\
&= \frac{1}{\sum_{i=1}^{n} x_i^2}\sum_{i=1}^{n} \mathbb{E}(\beta x_i + \epsilon_i) x_i \\
&= \frac{1}{\sum_{i=1}^{n} x_i^2}\sum_{i=1}^{n} \mathbb{E}(\beta x_i^2) + \mathbb{E}(\epsilon_i x_i) \\
&= \frac{1}{\sum_{i=1}^{n} x_i^2}\sum_{i=1}^{n} x_i^2\mathbb{E}(\beta) + x_i\mathbb{E}(\epsilon_i) \\
&= \frac{\sum_{i=1}^{n} x_i^2\mathbb{E}(\beta) + 0}{\sum_{i=1}^{n} x_i^2} \\
&= \beta \qquad \square
\end{aligned} \tag{20}$$

We have shown that as $\mathbb{E}\hat{\beta} = \beta$ that $\hat{\beta}$ is an unbiased estimator, which will come into play in problem c. Now we compute for $Var(\hat{\beta})$:

$$\begin{aligned} Var\left(\hat{\beta}\right) &= Var\left(\frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2}\right) \\ &= \frac{1}{(\sum_{i=1}^n x_i^2)^2} Var(\sum_{i=1}^n (\beta x_i^2 + \epsilon_i x_i) \quad as\ Var(kx) = k^2 Var(x) \\ &= \frac{1}{(\sum_{i=1}^n x_i^2)^2} Var(\sum_{i=1}^n (\epsilon_i x_i) \quad \textit{shift invariant} \\ &= \frac{1}{(\sum_{i=1}^n x_i^2)^2} \sum_{i=1}^n x_i^2 Var((\epsilon_i) \\ &= \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \qquad \square \end{aligned} \tag{21}$$

(c) Conclude that $\frac{1}{n}\sum_{i=1}^n x_i^2 \xrightarrow[n\to\infty]{} \infty$ then $\hat{\beta}$ is consistent.

We have shown that $\hat{\beta}$ is an unbiased estimator, that is its expecation is β, $\mathbb{E}\hat{\beta} = \beta$. We also showed that the variance is equal to $Var\left(\hat{\beta}\right) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$, so if $\frac{1}{n}\sum_{i=1}^n x_i^2 \xrightarrow[n\to\infty]{} \infty$ then $Var\left(\hat{\beta}\right) \to 0$. As $\hat{\beta}$ will then have 0 variance, all of its probability mass must be on its expectation, $\mathbb{E}\hat{\beta} = \beta$, thus showing that the estimator is consistent. We can show this more rigorously using Chebichevs bound:

$$\mathbb{P}\left\{|\hat{\beta} - \beta| \geq t\right\} = \mathbb{P}\left\{|\hat{\beta} - \mathbb{E}\hat{\beta}| \geq t\right\} \leq \frac{Var\left(\hat{\beta}\right)}{t^2} \tag{22}$$

Since $Var\left(\hat{\beta}\right) \to 0$ as $n \to \infty$, then $\hat{\beta} \xrightarrow{p} \beta$, showing $\hat{\beta}$ is consistent.

(d) Give an example showing that if $\frac{1}{n}\sum_{i=1}^n x_i^2$ does not approach infinity, then $\hat{\beta}$ can fail to be consistent.

If we take a toy example where all of $X_1, \ldots, X_n = 0$ then $\frac{1}{n}\sum_{i=1}^n x_i^2$ would not converge to infinity. In fact, our quantity would have undefined variance as $Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} = \sigma^2/0 = undefined$, and thus $\hat{\beta}$ wouldn't be a consistent estimator.