ECO 520 - FALL 2018 FINAL EXAM DECEMBER 7, 2018

When a question asks to *show* a result, it means you have to give a formal proof. Try to give a sufficient amount of details to support your computations and make use of the knowledge you acquired during lectures and recitations. Needless to say, cheating will not be tolerated.

Name:____

1. Let X_n denote a random variable with mean μ and variance σ^2/n^p, where $p > 0$, μ, and σ^2 are constants (not functions of n). Show that X_n converges in probability to μ. (*Hint*: Use Chebyshev?s inequality.)

   Solution. *For a sequence of random variables X_n, the Chebyshev's inequality claims that*

   $$P\left(|X_n - E(X_n)| > \varepsilon\right) \leq \frac{Var(X_n)}{\varepsilon^2},$$

   with $\mu = E(X_n)$ and $Var(X_n) = \sigma^2/n^p$. For a fixed ε, μ, σ^2 that do not depend on n, and $p > 0$, we have that

   $$\lim_{n\to\infty} \frac{\sigma^2}{n^p\varepsilon^2} = 0,$$

   which implies that

   $$\lim_{n\to\infty} P\left(|X_n - E(X_n)| > \varepsilon\right) = 0.$$

   This is enough to claim that $X_n \xrightarrow{p} \mu$.

2. Let $W_n \sim \chi^2_n$. Then the moment generating function of W_n is given by

   $$M_{W_n}(t) = (1-2t)^{-n/2}, \text{ for } t < 0.5.$$

   We would like to investigate the limiting distribution of the random variable

   $$Y_n = \frac{W_n - n}{\sqrt{2n}}.$$

   Follow these steps

   a) Derive the Moment Generating Function of Y_n. Show that this is equal to

   $$M_{Y_n}(t) = \left(e^{t\sqrt{2/n}} - t\sqrt{\frac{2}{n}}e^{t\sqrt{2/n}}\right)^{-n/2}, \text{ for } t < \sqrt{\frac{n}{2}}.$$

   Solution. *Recall that*

   $$\begin{aligned} M_{Y_n}(t) =& E\left[e^{tY_n}\right] = E\left[e^{t\frac{W_n-n}{\sqrt{2n}}}\right] = e^{-t\sqrt{n/2}}E\left[e^{t\frac{W_n}{\sqrt{2n}}}\right] \\ =& e^{-t\sqrt{n/2}}E\left[e^{\zeta W_n}\right] = e^{-t\sqrt{n/2}}(1-2\zeta)^{-n/2}, \end{aligned}$$

   where $\zeta = t/\sqrt{2n}$, with $\zeta < 0.5$, which implies,

   $$t/\sqrt{2n} < 1/2 \quad \Rightarrow \quad t < \sqrt{\frac{n}{2}}.$$

   Thus

   $$M_{Y_n}(t) = e^{-t\sqrt{n/2}}(1 - t\sqrt{\frac{2}{n}})^{-n/2} = \left(e^{t\sqrt{2/n}} - t\sqrt{\frac{2}{n}}e^{t\sqrt{2/n}}\right)^{-n/2}.$$

b) Use a Taylor expansion of the exponential function up to the third order to finally show that

$$\lim_{n\to\infty} M_{Y_n}(t) = e^{t^2/2}.$$

Solution. *Let*

$$e^{t\sqrt{2/n}} = 1 + \frac{\sqrt{2}t}{\sqrt{n}} + \frac{t^2}{n} + \frac{\sqrt{2}t^3}{3n\sqrt{n}},$$

which gives

$$\begin{aligned} e^{t\sqrt{2/n}} & - t\sqrt{\frac{2}{n}}e^{t\sqrt{2/n}} \\ &= 1 + \frac{\sqrt{2}t}{\sqrt{n}} + \frac{t^2}{n} + \frac{\sqrt{2}t^3}{3n\sqrt{n}} - t\sqrt{\frac{2}{n}} - t^2\frac{2}{n} - t^3\frac{2\sqrt{2}}{n\sqrt{n}} - t^4\frac{4}{n^2} \\ &= 1 - \frac{t^2}{n} + o(t^2/n) = 1 - \frac{t^2}{2n/2} + o(t^2/n). \end{aligned}$$

Notice that

$$\lim_{n\to\infty} \left(1 - \frac{t^2}{2n/2}\right)^{n/2} = e^{-t^2/2},$$

so that

$$\lim_{n\to\infty} \left(1 - \frac{t^2}{2n/2}\right)^{-n/2} = e^{t^2/2}.$$

c) What is then the asymptotic distribution of the random variable Y_n?

Solution. *Given the form of the MGF, Y_n converges to a standard normal distribution.*

3. Consider the following random variable $X^* \sim N(\mu, 1)$, with $\{X_i^*, i = 1, \ldots, n\}$, an IID sample from this distribution. In some cases, it is not possible to directly observe X_i^*, and we only have access to a (nonlinear) transformation of X_i^*, which we denote X_i. Derive the asymptotic properties of the maximum likelihood estimator of μ in the three following cases, where $\mathbb{1}$ is the indicator function,

a) $X_i = X_i^*$. That is, we directly observe the random variable X_i^*.

Solution. *Here, you could have replied without any computation. The MLE of μ, $\hat{\mu}_n$, is $\bar{X}$, the sample mean. As the population variable is normally distributed with mean μ and variance 1, we know that, both in finite sample and asymptotically,*

$$\sqrt{n}\left(\bar{X} - \mu\right) \sim N(0, 1),$$

directly using the properties of the normal distribution. Therefore, the MLE inherits these properties.

b) $X_i = X_i^* \mathbb{1}\left(X_i^* > 0\right)$. In this case, we only observe X_i^* when it is positive, and 0 otherwise. The distribution is a truncated normal distribution at 0. Its pdf is

$$f_X(x; \mu) = \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(X-\mu)^2}{2}\right)}{1 - \Phi\left(-\mu\right)},$$

where Φ is the CDF of a standard normal random variable. (*Hint*: You are not going to be able to find a closed form solution here. Obtain directly the second derivative and infer the properties of the asymptotic distribution. Let $\phi(\cdot)$ be the pdf of a standard normal distribution. The mean and variance of a truncated normal at 0 are given by

$$\begin{aligned} E(X) &= \mu + \frac{\phi(-\mu)}{1 - \Phi(-\mu)} \\ Var(X) &= 1 - \frac{\mu\phi(-\mu)}{1 - \Phi(-\mu)} - \left(\frac{\phi(-\mu)}{1 - \Phi(-\mu)}\right)^2) \end{aligned}$$

Cont.

Solution.

Remark (Common mistake). Notice that the parameter of interest is still μ. The issue that we have now is that $\bar{X}$ is not anymore an estimator of μ, because

$$E\left[\bar{X}\right] = \mu + \frac{\phi(-\mu)}{1-\Phi(-\mu)},$$

the mean of a truncated normal distribution, which is not equal to μ.

I write the log-likelihood function directly (ignoring any constant term)

$$\ell_X(\mu) = -\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{2} - n\log(1-\Phi(-\mu))$$

The first order condition is

$$\frac{d\ell_X(\mu)}{d\mu} = \sum_{i=1}^{n} X_i - n\mu - \frac{n\phi(-\mu)}{1-\Phi(-\mu)} = 0,$$

so that the MLE $\hat{\mu}_n$ satisfies,

$$\hat{\mu}_n + \frac{\phi(-\hat{\mu}_n)}{1-\Phi(-\hat{\mu}_n)} = \bar{X}.$$

You cannot solve this analytically, so to find the asymptotic distribution, we have to use ML theory. Taking the second derivative we get

$$\frac{d^2\ell_X(\mu)}{d\mu^2} = -n\left(1 - \frac{\phi'(-\mu)}{1-\Phi(-\mu)} - \left(\frac{\phi(-\mu)}{1-\Phi(-\mu)}\right)^2\right).$$

Notice that

$$\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}},$$

which implies

$$\phi'(z) = \frac{-z}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} = -z\phi(z).$$

This gives that $\phi'(-\mu) = \mu\phi(-\mu)$, and

$$\frac{d^2\ell_X(\mu)}{d\mu^2} = -n\left(1 - \frac{\mu\phi(-\mu)}{1-\Phi(-\mu)} - \left(\frac{\phi(-\mu)}{1-\Phi(-\mu)}\right)^2\right) = -nVar(X) < 0,$$

that does not depend on μ, and implies that the maximum is well defined for this optimization problem. The sample information matrix

$$I_n(\mu, X) = -\frac{1}{n}\frac{d^2\ell_X(\mu)}{d\mu^2} = Var(X).$$

Directly using ML theory (asymptotic unbiasedness, consistency and asymptotic normality) this gives,

$$\sqrt{n}\left(\hat{\mu}_n - \mu\right) \xrightarrow{d} N\left(0, Var(X)^{-1}\right)$$

c) $X_i = \mathbb{1}\left(X_i^* > 0\right)$. (*Hint*: You should know what the distribution of X_i is in this case.)

Solution. *In this case the distribution of X is a Bernoulli distribution. The probability that $X = 1$ is given by*

$$p = P(X = 1) = P(X^* > 0) = P(Z > -\mu) = P(Z < \mu) = \Phi(\mu).$$

The likelihood function is given by

$$L_X(\mu) = \prod_{i=1}^{n} \Phi(\mu)_i^X (1 - \Phi(\mu))^{1-X_i} = \Phi(\mu)^{\sum_{i=1}^n X_i} (1 - \Phi(\mu))^{n - \sum_{i=1}^n X_i}.$$

The log-likelihood is

$$\ell_X(\mu) = \sum_{i=1}^{n} X_i \log(\Phi(\mu)) + \left(n - \sum_{i=1}^{n} X_i\right) \log(1 - \Phi(\mu)).$$

The first order condition directly gives (replacing $\sum_{i=1}^n X_i = n\bar{X}$)

$$\begin{aligned}\frac{d\ell_X(\mu)}{d\mu} &= -\, n\phi(\mu) \left(-\frac{1}{\Phi(\mu)}\bar{X} + \frac{1}{1 - \Phi(\mu)}(1 - \bar{X})\right) \\ &= -\, n \frac{\phi(\mu)}{\Phi(\mu)(1 - \Phi(\mu))} \left(\Phi(\mu) - \bar{X}\right) = 0.\end{aligned}$$

For $|\mu| < \infty$, $n\phi(\mu) > 0$, and the first order condition is satisfied when the term inside parenthesis is equal to 0. Therefore,

$$\Phi(\mu) = \bar{X} \quad \Rightarrow \quad \hat{\mu}_n = \Phi^{-1}(\bar{X}),$$

where the inverse exists, as the CDF of a normal distribution is strictly increasing in $[0, 1]$ and therefore invertible. The second derivative is given by

$$\begin{aligned}\frac{d^2\ell_X(\mu)}{d\mu^2} &= -\, n \left[\left(\frac{\phi(\mu)}{\Phi(\mu)(1 - \Phi(\mu))}\right)' \left(\Phi(\mu) - \bar{X}\right) + \frac{(\phi(\mu))^2}{\Phi(\mu)(1 - \Phi(\mu))}\right] \\ \frac{d^2\ell_X(\hat{\mu}_n)}{d\mu^2} &= -\, n \frac{(\phi(\hat{\mu}_n))^2}{\Phi(\hat{\mu}_n)(1 - \Phi(\hat{\mu}_n))} < 0,\end{aligned}$$

where the last condition holds as the ratio is strictly positive. Therefore, the sample information matrix at the ML estimator of μ is equal to

$$I_n(\hat{\mu}_n, X) = -\frac{1}{n}\frac{d^2\ell_X(\hat{\mu}_n)}{d\mu^2} = \frac{(\phi(\hat{\mu}_n))^2}{\Phi(\hat{\mu}_n)(1 - \Phi(\hat{\mu}_n))}.$$

Directly using ML theory (asymptotic unbiasedness, consistency and asymptotic normality) this gives,

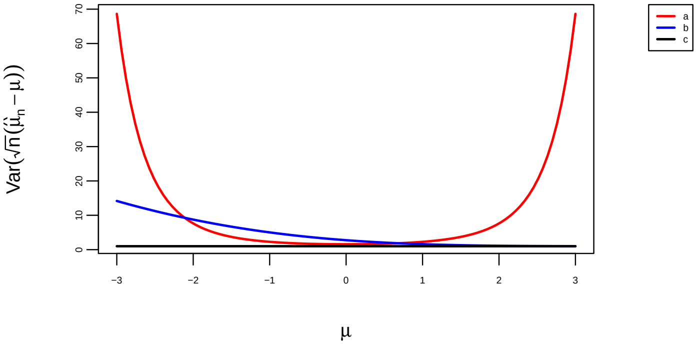
$$\sqrt{n}\,(\hat{\mu}_n - \mu) \xrightarrow{d} N\left(0, \frac{\Phi(\mu)(1 - \Phi(\mu))}{(\phi(\mu))^2}\right).$$

d) Are you able to say anything about the relative asymptotic efficiency of these estimators? What happens when our information about X_i^* decreases?

Solution. *All estimators are asymptotically unbiased, so that we can run a fair comparison using their asymptotic properties. The asymptotic variance of the distribution of the full information estimator (the one in part a) is equal to $1/n$, so that we can take it as a benchmark.*

Without doing any computations, we can intuitively argue that, as our information about X^* decreases, the variance of our estimator should increase. Therefore, the estimator in part a) is relatively more efficient than the estimators in parts b) and c). Similarly, the estimator in part *b) is more efficient than the estimator in part c). This intuitive argument was enough to get full points on this question. However, as it sometimes turns out with intuition, the argument is not fully correct. When the mean is positive and sufficently far from* 0*, then the efficiency of the estimator in part b) can get arbitrarly close to the estimator in part a). However, if the mean is negative, the further it is from* 0*, the more the variance of the estimator in part b) would increase. This is shown in Figure 3.*

```
mu <- seq(-3,3,length.out=100)
Phim <- pnorm(mu)
phim <- dnorm(mu)
varc <- Phim*(1-Phim)/phim^2
varb <- (1 - mu*phim/Phim - (phim/Phim)^2)^(-1)
vara <- array(1,c(length(mu),1))
```



Let me give a more formal argument. I will omit the division by n in the following, as it does not matter (all estimators converge at $\sqrt{n}$ rate).
First of all, notice that

$$\Phi(\mu)(1-\Phi(\mu)) > \phi(\mu)^2.$$

Taking logs on both sides (log is monotone, so it does not affect the ordering)

$$\begin{aligned} \log(\Phi(\mu)) + \log(1-\Phi(\mu)) >& 2\log(\phi(\mu)) = -\log(2\pi) - \mu^2, \\ \mu^2 + \log(\Phi(\mu)) + \log(1-\Phi(\mu)) >& -\log(2\pi). \end{aligned}$$

For $|\mu| < \infty$, $0 < \Phi(\mu) < 1$, so that the logarithm on the left hand side is not diverging to infinity. The part on the left-hand side (lhs) is a strictly concave function of μ, and it reaches its minimum at 0*, where the function is equal to $2\log(0.5) = -1.386$. The right-hand side is always equal to $-\log(2\pi) = -1.838$, and therefore lower than the lhs for every value of μ. This strict inequality is always satisfied. We can conclude that the variance of the estimator in part c) is always larger than the variance of the estimator in part a).*
For the variance of the estimator in part b) to be larger than 1*, we need*

$$1 - \frac{\mu\phi(\mu)}{\Phi(\mu)} - \left(\frac{\phi(\mu)}{\Phi(\mu)}\right)^2 \leq 1, \quad -\left[\frac{\mu\phi(\mu)}{\Phi(\mu)} + \left(\frac{\phi(\mu)}{\Phi(\mu)}\right)^2\right] \leq 0.$$

For positive μ, the inequality is obvious, as the terms on the lhs are strictly negative. For negative μ, the inequality holds as $\Phi(\mu)$ becomes smaller, and the quadratic term dominates the difference. Therefore, we have proven that the variance of the estimator in part a) is smaller than the variance of the estimator in part b).

Cont.

For the last part of the proof, let's focus on the ratio of the variances between the estimator in b) and the estimator in a). We have

$$\frac{1-\frac{\mu\phi(\mu)}{\Phi(\mu)}-\left(\frac{\phi(\mu)}{\Phi(\mu)}\right)^2}{\frac{(\phi(\mu))^2}{\Phi(\mu)(1-\Phi(\mu))}} = \frac{\Phi(\mu)^2-\mu\phi(\mu)\Phi(\mu)-\phi(\mu)^2}{\phi(\mu)^2}\frac{1-\Phi(\mu)}{\Phi(\mu)}$$

$$\left(\left(\frac{\Phi(\mu)}{\phi(\mu)}\right)^2-\mu\left(\frac{\Phi(\mu)}{\phi(\mu)}\right)-1\right)\frac{1-\Phi(\mu)}{\Phi(\mu)}.$$

The part in parenthesis is strictly increasing in μ, while the function $(1-\Phi(\mu))/\Phi(\mu)$ is strictly decreasing in μ. Let us find the points in which they intersect. That is, the point in which

$$\left(\left(\frac{\Phi(\mu)}{\phi(\mu)}\right)^2-\mu\left(\frac{\Phi(\mu)}{\phi(\mu)}\right)-1\right)=\frac{\Phi(\mu)}{1-\Phi(\mu)}.$$

This is not easy to solve analytically. Let's use **R** *to find these roots.*

```r
fun1 <- function(mu) ( pnorm(mu)/dnorm(mu))^2 -mu*pnorm(mu)/dnorm(mu) -1
fun2 <- function(mu) pnorm(mu)/(1 - pnorm(mu))

diff_fun <- function(mu) fun1(mu)/fun2(mu) - 1
zero1 <- uniroot(diff_fun,c(0,1))
zero2 <- uniroot(diff_fun,c(-3,-2))
```

It turns out these points are $\mu_1 \approx 0.6815$, and $\mu_2 \approx -2.1123$. Thus, for $\mu > \mu_1$, and $\mu < \mu_2$, the estimator in part b) is relatively more efficient than the estimator in part c). While for $\mu_2 \leq \mu \leq \mu_1$, the estimator in part c) is more efficient.
Let me conclude with two simulation studies to show the theoretical result. In the first, I am taking $\mu = 0$, so that the estimator in part c) is more efficient; and in the second I take $\mu = 2$, so that the estimator is part b) is more efficient.

```r
rm(list=ls())
library(xtable)

set.seed(123)
mu_sim1 <- 0
mu_sim2 <- 2
tot_sim <- 10000
n <- 1000
nmulti <- 10

muhat_a1 <- integer(tot_sim)
muhat_a2 <- integer(tot_sim)
muhat_b1 <- integer(tot_sim)
muhat_b2 <- integer(tot_sim)
muhat_c1 <- integer(tot_sim)
muhat_c2 <- integer(tot_sim)

for(j in 1:tot_sim){
  x1 <- rnorm(n,mean=mu_sim1,sd =1)
  x2 <- rnorm(n,mean=mu_sim2,sd =1)

  muhat_a1[j] <- mean(x1)
  muhat_a2[j] <- mean(x2)

  x1astb <- x1*(x1 > 0)
  x2astb <- x2*(x2 > 0)

  #Log-Likelihood
  llik_b <- function(mu) 0.5*n*log(2*pi) + 0.5*sum((x1astb - mu)^2) + n*log(1 - pnorm(-mu))
  #First order condition
  scor_b <- function(mu) -mean(x1astb) + (mu + pnorm(-mu)/(1 - pnorm(-mu)))

  temp1val <- integer(nmulti)
  temp1par <- integer(nmulti)
  for(l in 1:nmulti){
      temp1      <- optim(mean(x1astb)+rnorm(1),fn = llik_b,gr=scor_b,method = "CG")
      temp1par[l]<- temp1$par
      temp1val[l]<- temp1$val
  }
  muhat_b1[j]<- temp1par[which.min(temp1val)]

  #Log-Likelihood
  llik_b <- function(mu) 0.5*n*log(2*pi) + 0.5*sum((x2astb - mu)^2) + n*log(1 - pnorm(-mu))
  #First order condition
  scor_b <- function(mu) -mean(x2astb) + (mu + pnorm(-mu)/(1 - pnorm(-mu)))

  temp2val <- integer(nmulti)
  temp2par <- integer(nmulti)
  for(l in 1:nmulti){
      temp2      <- optim(mean(x2astb)+rnorm(1),fn = llik_b, gr = scor_b,method = "CG")
      temp2par[l]<- temp2$par
      temp2val[l]<- temp2$val
  }

  muhat_b2[j]<- temp2par[which.min(temp2val)]

  x1astc <- (x1 >= 0)
  x2astc <- (x2 >= 0)

  muhat_c1[j]<- qnorm(mean(x1astc))
  muhat_c2[j]<- qnorm(mean(x2astc))
}
```

For every simulation I am taking $S = 10^4$ simulations from a normal random variable. Variances for the three estimators in this Monte-Carlo simulation are reported in Table 1.

	a	b	c
$\mu = 0$	0.99181	21.18614	1.58082
$\mu = 2$	1.01529	1.28124	8.14051

Table 1: Simulation Results

Notice that the table confirms our theoretical result. The full information estimator is the most efficient; while the estimator in part c) is more efficient than the one in part b) when $\mu = 0$, and viceversa when $\mu = 2$.

4. Consider the uniform distribution with density function $f_{X|\Theta}(x|\theta) = 1/\theta$, $0 \leq x \leq \theta$, and θ unknown.

a) Show that the Pareto distribution,

$$\pi_\Theta(\theta) = \begin{cases} ak^a\theta^{-(a+1)}, & \theta \geq k, a > 0 \\ 0, & \text{otherwise} \end{cases},$$

is a conjugate prior for the uniform distribution.

Solution. *Using Bayes' formula*

$$\begin{aligned}\pi_{\Theta|X}(\theta|x) \propto & \pi_\Theta(\theta) f_{X|\Theta}(x|\theta) = ak^a\theta^{-(a+1)}\theta^{-1} \\ =& ak^a\theta^{-(a+2)},\end{aligned}$$

with $\theta \geq k$ is the kernel of a Pareto distribution with parameters $(k, a+1)$.

b) Show that $\hat{\theta} = max\{X_1, \dots, X_n\}$ is the Maximum Likelihood Estimator of θ, where $\{X_1, \dots, X_n\}$ is an IID sample from $f_X(x;\theta)$.

Solution. *The likelihood and log-likelihood are*

$$L_X(\theta) = \frac{1}{\theta^n}, \quad \ell_X(\theta) = -n\log(\theta),$$

respectively, for $\theta \geq X_{(n)}$, where $X_{(n)}$ is the largest value of X in our sample. Both functions are strictly decreasing in θ, so that the ML estimator is $\hat{\theta}_n = X_{(n)}$.

c) Find the posterior distribution. (*Hint*: It is convenient in this case to find the exact expression of the posterior, so you may not want to ignore the denominator in the Bayes' formula this time.)

Cont.

Solution. *The posterior distribution is*

$$\pi_{\Theta|X}(\theta|x) = \frac{\pi_\Theta(\theta)L_X(\theta)}{f_X(x)} = \frac{\pi_\Theta(\theta)L_X(\theta)}{\int_\Theta \pi_\Theta(\theta)L_X(\theta)d\theta}.$$

We have that

$$\pi_\Theta(\theta)L_X(\theta) = ak^a\theta^{-(a+n+1)},$$

and

$$\begin{aligned}\int_\Theta ak^a\theta^{-(a+n+1)}d\theta =& ak^a\int_M^\infty \theta^{-(a+n+1)}d\theta \\ =& -\frac{ak^a}{a+n}\left[\theta^{-(a+n)}\right]_M^\infty = ak^a\frac{M^{-(a+n)}}{a+n}\end{aligned}$$

where $M = max(X_{(n)}, k)$. Putting everything together, we have that

$$\pi_{\Theta|X}(\theta|x) = (a+n)M^{a+n}\theta^{-(a+n+1)},$$

with $\theta \geq M$, which is a Pareto distribution with parameters $(M, a+n)$.

d) Find the bayesian point estimator for the quadratic cost function.

Solution. *When we choose the quadratic cost function, the Bayesian estimator is simply the posterior mean. We thus have,*

$$\hat{\theta} = E\left[\theta|X\right] = \frac{(a+n)M}{a+n-1}.$$

e) Find the MAP estimator and compare it to the MLE.

Solution. *The MAP estimator is defined as*

$$\hat{\theta}_{MAP} = \arg\max_\theta\left[\ell_X(\theta) + \log\pi_\Theta(\theta)\right].$$

As for the maximization of the likelihood, this problem does not have an interior solution, so we need to look at solutions at the boundaries. Since $\theta \geq M$, and the objective function is strictly decreasing in θ, we have $\hat{\theta}_{MAP} = M = max(X_{(n)}, k)$. Notice that for $k < X_{(n)}$, this estimator is equal to the MLE.

The End.