STAT 830

Problems: Assignment 4

1. Suppose we have data $X_1, \ldots, X_n$ and our model is that the data are iid with a Normal$(\mu, 1)$ distribution. Let $Y_i = 1(X_i > 0)$. Let $\psi = P_\mu(Y_i = 1)$.

   (a) Find the MLE, $\hat{\psi}$, of ψ and an approximate 95% confidence interval for ψ.

   The parameter of interest is

   $$\psi = P(X > 0) = P(X - \mu > -\mu) = P(N(0,1) < \mu) = \Phi(\mu)$$

   so the mle is
   $$\hat{\psi} = \Phi(\bar{X}).$$

   The estimated standard error is

   $$|\psi'|/\sqrt{n} = \phi(\bar{X})/\sqrt{n}.$$

   (b) Let $\tilde{\psi}$ be the sample mean of the Y_i. Show that $\tilde{\psi}$ is a consistent estimator of ψ.

   Averages converge to their expected values and the expected value of Y_i is ψ so

   $$\tilde{\psi} \to \psi.$$

   (c) Use the delta method to compute an approximate standard error of $\hat{\psi}$.

   This is really already in part a).

   (d) Compute the ratio of the standard error of the MLE of ψ and the approximate SE of $\tilde{\psi}$ and take the limit of this ratio as $n \to \infty$. The SE of $\tilde{\psi}$ comes from fact that it is an average – an estimate of a binomial probability.

The SE of $\tilde{\psi}$ comes from fact that it is an average – an estimate of a binomial probability. That is,

$$\sqrt{\Phi(\mu)[1-\Phi(\mu)]}/\sqrt{n}.$$

The (approximate) standard error of $\hat{\psi}$ is, from a),

$$\frac{\phi(\mu)}{\sqrt{n}}.$$

The ratio is

$$\frac{\Phi(\mu)[1-\Phi(\mu)]}{\phi(\mu)}.$$

which is the limit asked for.

(e) Suppose that with some large sample size n you got a certain SE for the MLE of ψ. How many times larger would n have to be to get the same SE, approxim ately, for $\tilde{\psi}$? (One over this number is called the Asymptotic Relative Efficiency of $\tilde{\psi}$.)

The relative efficiency you are asked for is

$$\frac{\phi^2(\mu)}{\Phi(\mu)[1-\Phi(\mu)]}$$

The question asks for the inverse this number. At $\mu = 0$ this is $4/(2\pi) < 1$. It is even smaller for $\mu \neq 0$.

(f) Suppose the data really have the distribution

$$F(x) = \exp(-\exp(-x)).$$

What are the limits of $\tilde{\psi}$ and the limit of the MLE you found in part a)? (The limits I mean are limits as the sample size n goes to ∞.) How do these compare with $P(X > 0)$ for this distribution?

In part e) the estimator $\hat{\psi}$ is still $1-\Phi(-\bar{X})$ which converges to $1-\Phi(-\mu)$. But $\psi = P(X_i > 0) = 1 - F_X(0)$ so unless

$$\Phi(-\mu) = 1 - \Phi(\mu) = F_X(0)$$

this estimate is not consistent. (It is possible for this equality to hold for some non-normal CDFs.)

2. Let $X_1, \ldots, X_n$ be a sample from the Normal$(\mu, 1)$ distribution and let $\theta = e^{\mu}$. Let $\hat\theta$ be the MLE of θ. In this problem I want you to compare several approximations to the distribution of $\hat\theta$.

   The function $f(x) = \exp(x)$ has $f' = f$ so the approximate standard error is

   $$|f'(\mu)|\sqrt{\text{Var}(\bar X)} = e^{\mu}/\sqrt{n} = \theta/\sqrt{n}.$$

   You replace θ by $\exp\{-barX\}$ to get intervals. Here is R code for a simulation study

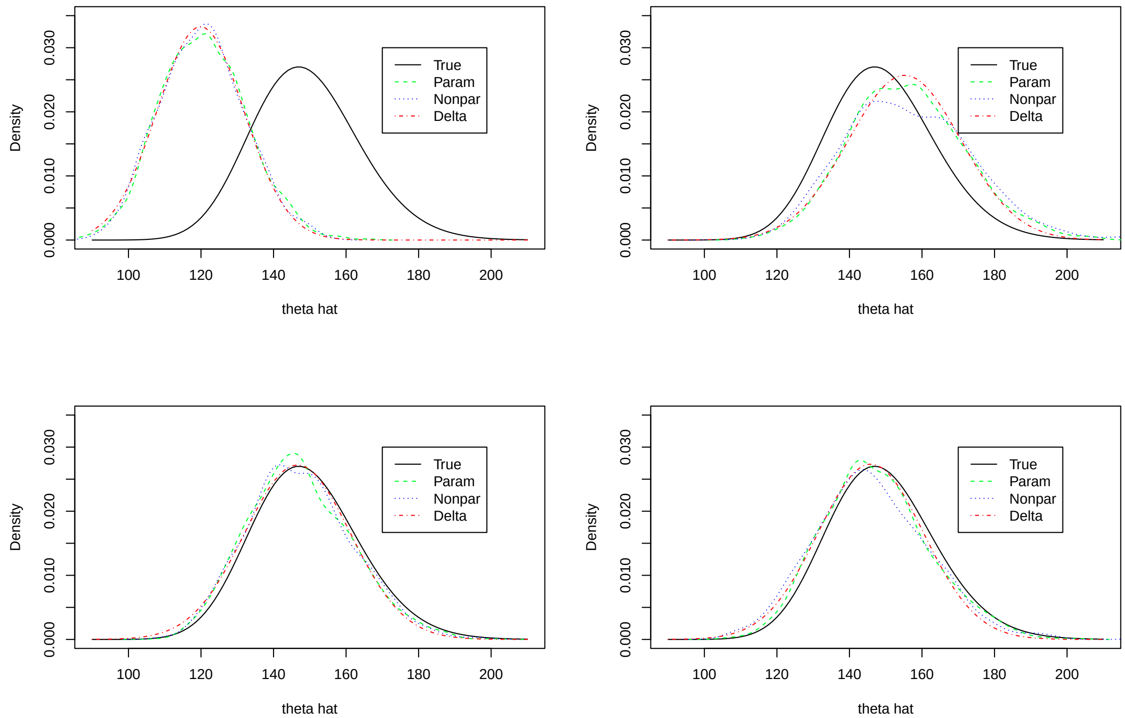
```
pdf("Asst4Q.pdf")
nmc=2000
par(mfrow=c(2,2),mai=c(0.3,0.2,0.2,0.2))
pb =rep(0,nmc)
npb =rep(0,nmc)
for(iframe in 1:4){
 x = rnorm(100)+5
 xbar =mean(x)
 sd=sd(x)
 thetahat = exp(xbar)
 sehat = thetahat/10
# Parametric bootstrap -- 2000 bootstrap samples
 for(i in 1:nmc) pb[i]=exp(mean(rnorm(100)+xbar))
# Nonparametric bootstrap
 for(i in 1:nmc) npb[i]=exp(mean(sample(x,n,replace=T)))
  lower = c(thetahat-1.96*sehat,
       quantile(pb,0.025),quantile(npb,0.025))
  upper = c(thetahat+1.96*sehat,
       quantile(pb,0.975),quantile(npb,0.975))
  cat(round(cbind(lower,rep(exp(5),3),upper),4),fill=28)
  cat("\n")
  cat("\n")
  curve(dnorm(log(x),5,1/10)/x,xlim=c(90,210),
    ylim=c(0,0.035), ylab="Density",
    xlab="theta hat",lty=1)
  lines(density(pb),lty=2,col="green")
```

```
  lines(density(npb),lty=3,col="blue")
  curve(dnorm(x,thetahat,sehat),add=T,lty=4,col="red")
  legend(170,0.03,lty=1:4,
     col=c("black","green","blue","red"),
     legend=c("True","Param","Nonpar","Delta"))
  }
dev.off()
```

The black curve is the true (log-normal) density of $\hat{\theta}$. The red curve is the delta method estimate (a normal density) of that. A density estimate based on the nonparametric bootstrap samples is in blue; a similar estimate for the parametric bootstrap is in green. I made 4 pictures for 4 data sets to show that the conclusions are somewhat sensitive to the actual data set. I should add that this is a tiny Monte Carlo study. I should really repeat this not 4 times but thousands of times and summarize the average behaviour of the methods.

The various approximations are all about equally good in that they are equally far from the truth – much more similar to each other than to the truth. This is just sampling variability and the nature of the statistic – essentially still a simple function of a sample mean.

I asked for cdfs, not densities, but the conclusion is pretty well the same.



3. Suppose $X_1, \ldots, X_n$ are a sample of size n from the density

$$f_{\alpha,\beta}(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp(-x/\beta)\, 1(x > 0).$$

In the following question the digamma function ψ is defined by $\psi(\alpha) = \frac{d}{d\alpha} \log(\Gamma(\alpha))$ and the trigamma function ψ' is the derivative of the digamma function. From the identity $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ you can deduce recurrence relations for the digamma and trigamma functions. You can consult Lecture 6 material where a lot of this is redone. You don't have to redo anything I already told you in the videos or the notes.

(a) If α is known find the mle for β.

The log likelihood for the full model is

$$\ell(\alpha, \beta) = -n \log(\Gamma(\alpha)) - n\alpha \log(\beta) - \sum X_i/\beta + (\alpha - 1) \sum \log(X_i)$$

Replacing α by α_0 and differentiating wrt β gives the likelihood equation

$$n\alpha_0/\beta = \sum X_i/\beta^2$$

so that

$$\hat{\beta} = \bar{X}/\alpha_0$$

(b) When both α and β are unknown what equation must be solved to find $\hat{\alpha}$, the mle of α?

The α derivative of the log likelihood is

$$-n\psi(\alpha) - n\log(\beta) + \sum \log(X_i).$$

Replace β with, $\bar{X}/\alpha$ as in a) to get

$$n\psi(\alpha) = n[\log(\alpha) - \log(\bar{X}) + \overline{\log(X)}].$$

(c) Evaluate the Fisher information matrix.

$$\mathcal{I} = \begin{bmatrix} n\psi'(\alpha) & n/\beta \\ n/\beta & n\alpha/\beta^2 \end{bmatrix}$$

(d) Here is a sample of size 40

```
3.547 1.228 2.052 1.556 2.487 0.469 2.707 0.395
0.770 0.666 4.242 1.474 1.277 2.519 0.578 2.989
1.900 1.422 3.701 1.278 2.820 0.224 0.482 1.426
2.146 2.975 2.792 0.846 3.190 1.680 0.686 1.634
0.969 4.010 1.792 1.287 0.730 0.849 2.447 2.147
```

Use this data in the following questions. First take $\alpha = 1$ and find the mle of β subject to this restriction.

I get $\hat{\beta} = \bar{X} = 1.810$.

(e) Now use $\mathrm{E}(X) = \alpha\beta$ and $\mathrm{Var}(X) = \alpha\beta^2$ to get method of moments estimates $\tilde{\alpha}$ and $\tilde{\beta}$ for the parameters. (This was done in class so I just mean get numbers.)

I get $\hat{\alpha}_{\text{mom}} = 2.830$ and $\hat{\beta}_{\text{mom}} = 0.639$.

Estimate	α	β
Method of Moments	2.8303245	0.6394055
First Iteration	2.4989479	0.7142676
Second Iteration	2.5335954	0.7142909
Limit of Iterations	2.5351893	0.7138422

(f) Do two steps of Newton Raphson to get MLEs.

I record here only iterates in a table.

(g) Compute standard errors for the MLEs and compare the difference between the estimates in the previous 2 questions to the SEs.

The standard errors are 0.534 and 0.166. The differences between the the two iterative schemes are much smaller than these standard errors. Notice that the iterative schemes both converge to the MLEs. Notice that after 2 iterations nothing important is happening.

(h) Do a likelihood ratio test of $H_o : \alpha = 1$.

For $\alpha = 1$ the mle of β is $\bar{X}/\alpha = \bar{X}$. The likelihood ratio statistic works out to be

$$2\left[\ell(\hat{\alpha}, \hat{\beta}) - \ell(1, \hat{\beta}(1))\right] \approx 15.3$$

This is converted to an approximate asymptotic P-value by computing

$$P(\chi^2_1 > 15.3).$$

I used the function `pchisq(15.3,1,lower=F)` in `R` to get

$$P = 9 \times 10^{-5}.$$

There is no credibility to the idea that $\alpha = 1$. I didn't say much about P-values. My own view is that in problems where a firm decision does not have to be made they are much more informative that the phrase "rejected at the 5% level". You never see a scientific article with that sort of bare statement. The P-value allows other people to substitute their own α for yours without doing any further arithmetic.

The preceding words are *controversial* these days because the value of P-values is much diminished by the process of computing many and picking out the interesting ones with no multiple comparisons correction.

4. Suppose $X_1, \ldots, X_n$ are Uniform$[0, \theta]$ and let $T_n = \max\{X_1, \ldots, X_n\}$ be the MLE of θ. Let

$$\phi(X_1, \ldots, X_n) = 1(T_n > c)$$

be a test function to test $H_o : \theta = 1$.

(a) Find the power function of ϕ.

The power function $\pi(\theta)$ is, for $c > 0$,

$$\begin{aligned}\pi(\theta) = P_\theta(Y > c) \\ &= 1 - P(X_1 \le c, \ldots, X_n \le c) \\ &= \begin{cases} 1 - (c/\theta)^n & c < \theta \\ 0 & c \ge \theta \end{cases}\end{aligned}$$

(b) Find a choice of c so that the level of this test is $\alpha = 0.05$.

Set the power at $\theta = 1/2$ equal to α, the desired size to get

$$1 - (2c)^n = 0.05$$

so that

$$c = \frac{(0.95)^{1/n}}{2}.$$

(c) The P-value corresponding to this test is a certain function of T_n. Plot this function of T_n against T_n over the range $0.9 \le T_n \le 1.1$ for $n = 20$.

5. For $n = 1, 2, \ldots$ let $\lambda_n = 1/n$ and suppose that $X_n \sim$Poisson(λ_n).

(a) Prove that X_n converges to 0 in probability.

Since X_n has a Poisson distribution its values are non-negative integers. So for $\epsilon < 1$ the event $|X_n| > \epsilon$ is the same as the event $X_n \neq 0$ which as probability $1 - \exp(-\lambda_n)$ which clearly converges to 0.

(b) Prove that even $Y_n = nX_n$ converges to 0 in probability.

For $\epsilon < 1$ the event $|Y_n| > \epsilon$ is the same as the event $X_n > \epsilon/n$ which is still the same event as $X_n \neq 0$ which as probability $1 - \exp(-\lambda_n)$ which clearly converges to 0.

(c) Show that Y_n does not converge to 0 in pth mean for either $p = 1$ or $p = 2$.

The mean of Y_n is 1 and this is $\mathrm{E}(|Y_n - 0|)$. This does not go to 0 so Y_n does not converge in pth mean for $p = 1$. For $p = 2$ we do the same with $\mathrm{E}(Y_n^2) = n$.

6. Suppose that for each positive integer n we have

$$P\left(X_n = \frac{1}{n}\right) = 1 - \frac{1}{n^2} = 1 - P(X_n = n^2).$$

(a) Does X_n converge in probability and if so to what limit?

The sequence converges to 0 in probability. Fix $\epsilon > 0$ and consider the event $|X_n| > \epsilon$. Consider the case $\epsilon < 1$. Then the event in question cannot happen if $X_n = n^2$. It $X_n = 1/n$ then the event happens if and only if $1/n > \epsilon$. Let N be the largest integer for which $1/N > \epsilon$ (it is the so-called floor of $1/\epsilon$. Then

$$P(|X_n| > \epsilon) = \begin{cases} 1/n & n \leq N \\ 0 n > N \end{cases}.$$

Thus this sequence converges to 0 proving $X_n \to 0$ in probability.

(b) Does X_n converge in quadratic mean and if so to what limit?

The sequence converges to x in quadratic mean if $\mathrm{E}\left\{(X_n - x)^2\right\} \to 0$ which requires

$$\lim_{n\to\infty} \mathrm{E}(X_n) = x$$

and

$$\lim_{n\to\infty} \mathrm{Var}(X_n) = 0.$$

The mean of X_n is

$$n^2(1/n^2) + 1/n(1 - 1/n^2)$$

which converges to 1 so we would have to have $x = 1$. The second moment of X_n is

$$\mathrm{E}(X_n^2) = n^4/n^2 - (1/n)^2(1 - 1/n^2)$$

which converges to ∞. Thus

$$\mathrm{Var}(X_n) = \mathrm{E}(X_n^2) - \left\{\mathrm{E}(X_n)\right\}^2$$

converges to ∞ so this sequence does not converge in quadratic mean.

(c) BONUS question only: does X_n converge almost surely and if so to what limit?

The sequence converges to 0 almost surely.