S722 HW9

John Koo

To save on typing, I will denote $\frac{\partial^k}{\partial x^k} f(x) = \partial_x^k f(x)$.

5.43

Taylor expansion of $g(Y_n)$ around θ:

$g(Y_n) \approx g(\theta) + g'(\theta)(Y_n - \theta)$
$\implies \sqrt{n}(g(Y_n) - g(\theta)) = g'(\theta)\sqrt{n}(Y_n - \theta)$

Let $Z_n = \sqrt{n}(Y_n - \theta)$. Then $Z_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

Let Z be the limit of Z_n. Then $g'(\theta)Z \sim \mathcal{N}(0, \sigma^2(g'(\theta))^2)$, so $g'(\theta)Z_n \xrightarrow{d} \mathcal{N}(0, \sigma^2(g'(\theta))^2)$.

Then $\sqrt{n}(g(Y_n) - g(\theta)) = g'(\theta)\sqrt{n}(Y_n - \theta) = Z_n \xrightarrow{d} g'(\theta)Z \sim \mathcal{N}(0, \sigma^2(g'(\theta))^2)$.

a

$P(|Y_n - \mu| < \epsilon) = P(\sqrt{n}|Y_n - \mu| < \sqrt{n}\epsilon) \to P(|Z| < \infty$ as $n \to \infty$.

b

(Done before part a)

Theorem 5.5.24

(Covered in problem 5.43)

5.44

a

Since $X_i \overset{iid}{\sim} Bernoulli(p)$, $E[X_i] = p$ and $Var(X_i) = p(1-p)$. Then this follows from the central limit theorem.

b

Let $g(y) = y(1-y)$. Then $g'(y) = 1-2y$. Then $\sqrt{n}(g(Y_n) - g(p)) \xrightarrow{d} \mathcal{N}(0, \sigma_X^2(g'(p))^2) = \mathcal{N}(0, p(1-p)(1-2p)^2)$

c

Again, letting $g(y) = y(1-y)$, we get $g''(y) = -2$. Then $n(g(Y_n) - g(1/2)) \xrightarrow{d} \frac{1}{4}\frac{-2}{2}\chi_1^2 = -\frac{1}{4}\chi_1^2$.

8.31

a

$T(X) = \sum X_i$ is a sufficient statistic for λ, and $T \sim Poisson(n\lambda)$.

$\frac{f(t||lambda_1)}{f(t|\lambda_2)} = e^{-n(\lambda_1-\lambda_2)}(\lambda_1/\lambda_2)^t$ is monotonic in t, so a test of the form $T > k \iff \phi(T) = 1$ is UMP.

a

$T_n \xrightarrow{d} \mathcal{N}(n\lambda, n\lambda)$ by CLT. Under H_0, this is $T_n \xrightarrow{d} \mathcal{N}(n, n)$. Then $\frac{T_n - n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$.

Similarly, when $\lambda = 2$, we get $\frac{T_n - 2n}{\sqrt{2n}} \xrightarrow{d} \mathcal{N}(0, 1)$.

So our system of equations becomes:

- $\frac{T_n - n}{\sqrt{n}} = z_{.05}$
- $\frac{T_n - 2n}{\sqrt{2n}} = -z_{.9}$

Because I don't want to do this out, I'll just use R:

```
z.05 <- qnorm(.95)
z.9 <- qnorm(.1)

f <- function(x) {
  c((x[1] - x[2]) / sqrt(x[2]) - z.05,
    (x[1] - 2 * x[2]) / sqrt(2 * x[2]) - z.9)
}

pracma::fsolve(f, c(1, 1))
```

```
$x
[1] 17.63917 11.95252

$fval
[1] 6.811696e-10 6.311125e-10
```

$n = 12$

10.1

$E[X] = \int_{-1}^{1} x\frac{1}{2}(1 + \theta x)dx = \int_{-1}^{1} \frac{x}{2} + \frac{\theta}{2}x^2 dx$ (the first part is an odd function so we can neglect it ...)
$= \int_{-1}^{1} \frac{\theta}{2}x^2 = \frac{\theta}{3}$

$E[X^2] = \int_{-1}^{1} \frac{x^2}{2} + \frac{\theta}{2}x^3 dx$
(the second part is odd this time ...)
$= \int_{-1}^{1} x^2/2dx = 1/3$

Then $Var(X) = \frac{3-\theta^2}{9}$.

We can see that $3\bar{X}_n$ is an unbiased estimator for θ. Furthermore, $Var(3\bar{X}_n) = \frac{9}{n^2}n(\frac{3-\theta^2}{9}) = \frac{3-\theta^2}{n} \to 0$ as $n \to \infty$. So $3\bar{X}_n$ is a consistent estimator of θ.

10.3

a

$\ell(\theta) = -\frac{n}{2}\log 2\pi\theta - \frac{1}{2\theta}\sum(x_i - \theta)^2 = -\frac{n}{2}\log\theta - \frac{\sum x_i^2}{2\theta} - \frac{n\theta}{2} + C$

Then taking the derivative and setting it to 0, we get

$0 = -\frac{n}{2\theta} + \frac{\sum x_i^2}{2\theta^2} - \frac{n}{2}$
$\implies 0 = n\theta - \sum x_i^2 + n\theta^2$
$\implies 0 = \theta + \theta^2 - \frac{1}{n}\sum x_i^2$

Solving this, we get $\theta = \frac{-1\pm\sqrt{1+4W}}{2}$, and since $\theta > 0$, $\hat{\theta} = \frac{-1+\sqrt{1+4W}}{2}$.

b

$\ell''(\theta) = \frac{n}{2\theta^2} - \frac{\sum x_i^2}{\theta^3}$

So $I(\theta) = -E[\log f(X|\theta)]$
$= -\frac{n}{2\theta^2} + \frac{1}{\theta^3}\sum E[X_i^2]$
$= -\frac{n}{2\theta^2} + \frac{n(\theta+\theta^2)}{\theta^3}$
$= -\frac{n}{2\theta^2} + \frac{n+n\theta}{\theta^2} = \frac{2n\theta+n}{2\theta^2}$

So the asymptotic variance is $\frac{2\theta^2}{2n\theta+n}$.

10.4

a

$\sum X_i Y_i = \sum X_i(\beta X_i + \epsilon_i) = \beta\sum X_i^2 + \sum X_i\epsilon_i$

Then the expression becomes $\beta + \frac{\sum X_i\epsilon_i}{\sum X_i^2}$.

The expected value is β since in the second part, we can separate $E[X_i\epsilon_i] = E[X_i]E[\epsilon_i] = 0$.

From a table of normal moments, $Var(X_i^2) = 2\tau^2(2\mu^2 + \tau^2)$.

$Var(X_i\epsilon_i) = E[X_i^2]E[\epsilon_i^2] = (\mu^2 + \tau^2)\sigma^2$

So the variance is $\frac{n\sigma^2(\mu^2+\tau^2)}{n^2(\mu^2+\tau^2)^2} = \frac{\sigma^2}{n(\mu^2+\tau^2)}$ since $E[X_i\epsilon_i] = 0$.

b

$\frac{\sum Y_i}{\sum X_i} = \frac{\sum\beta X_i + \sum\epsilon_i}{\sum X_i} = \beta + \frac{\sum\epsilon_i}{\sum X_i}$

As before, since $E[\epsilon_i] = 0$, the expectation is β.

And as before, we only have to consider the first part of the formula, so the variance is $\frac{n\sigma^2}{n^2\mu^2} = \frac{\sigma^2}{n\mu^2}$.

c

$\frac{1}{n}\sum Y_i/X_i = \frac{1}{n}\sum \frac{\beta X_i + \epsilon_i}{X_i}$
$= \beta + \sum \frac{1}{n} \sum \frac{\epsilon_i}{X_i}$

Again, since $E[\epsilon_i] = 0$, the expectation is just β.

The variance is $\frac{1}{n^2}\sum \frac{\sigma^2}{\mu^2} = \frac{\sigma^2}{n\mu^2}$

10.8

a

$\ell'(\theta) = (\sum \log f(x_i|\theta))' = \sum \frac{\partial_\theta f(x_i|\theta)}{f(x_i|\theta)}$

And then just multiply the left and right sides by $1/\sqrt{n}$.

As $n \to \infty$, $\hat{\theta} \to \theta_0$, so $\ell'(\hat{\theta}) \to \ell(\theta_0)$ and $\ell'(\hat{\theta}) = 0$ for all n.

By definition, the expected value of the square of this quantity is the Fisher information, which is also the variance since the expected value is 0.

Then by the central limit theorem, $\frac{1}{n}\sum W_i \xrightarrow{d} \mathcal{N}(0, I(\theta_0))$

b

$\partial_\theta \left(\frac{\partial_\theta f(x_i|\theta)}{f(x_i|\theta)} \right) = \frac{(\partial_\theta^2 f(x_i|\theta)) f(x_i|\theta) - (\partial_\theta f(x_i|\theta))^2}{(f(x_i|\theta))^2}$
$= \frac{\partial_\theta^2 f(x_i|\theta)}{f(x_i|\theta)} - \left(\frac{\partial_\theta f(x_i|\theta)}{f(x_i|\theta)}\right)^2$

And the second term is W_i^2.

So $\ell''(\theta_0|X) = -\sum W_i^2 + \sum \frac{\partial_\theta^2 f(X_i|\theta)}{f(X_i|\theta)}$

Similar to part (a), the mean of W_i^2 is the Fisher information for a single observation. Since the sample is iid, its average is the Fisher information for the sample.

For the second part, if we assume regularity,
$E\left[\frac{\partial_\theta^2 f(x_i|\theta)}{f(x_i|\theta)}\right] = \int \frac{\partial_\theta^2 f(x|\theta)}{f(x|\theta)} f(x|\theta)dx = \partial_\theta^2 \int f(x|\theta)dx = \partial_\theta^2(1) = 0$

Since $\frac{1}{n}\sum W_i$ is a sample mean, it converges in probability to its expected value, $I(\theta_0)$.

Theorem 10.1.12

Statement of the theorem:

- given
  - $X_i \overset{iid}{\sim} f(x|\theta)$
  - $\hat{\theta}$ the MLE of θ
  - $\tau(\theta)$
- then
  - $\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{d} \mathcal{N}(0, v(\theta))$
    where $v(\theta)$ is the CRLB

Proof:

First, show for $\tau(\theta) = \theta$.

By Taylor expansion, we can write
$\ell'(\theta|x) \approx \ell'(\theta_0|x) + (\theta - \theta_0)\ell''(\theta_0|x)$,
and take θ_0 as the true value.

Then rearranging some terms, we have
$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{-\ell'(\theta_0)/\sqrt{n}}{\ell''(\theta_0)/n}$

From problem 10.8, we saw that the numerator converges in distribution to $\mathcal{N}(0, I(\theta_0))$ and the denominator converges in probability to $I(\theta_0)$. Then using Slutsky's theorem, the entire fraction must converge in distribution to $\mathcal{N}(0, 1/I(\theta_0))$.

Then applying theorem 5.5.24, this extends to any continuous transformation of $\hat{\theta}$.

10.9

a

Let $T = 1$ if $X_1 = 0$ and 0 otherwise. Then T is an unbiased estimator of $e^{-\lambda}$ and $E[T|\sum X_i]$ is UMVUE for λ. We can use the fact that T is Bernoulli.

$E[T|\sum X_i] = P(X_1 = 0|\sum X_i = y)$
$= \frac{P(X_1=0, \sum X_i=y)}{P(\sum X_i=y)}$
$= \frac{e^{-\lambda}((n-1)\lambda)^y e^{-(n-1)\lambda}/y!}{(n\lambda)^y e^{-n\lambda}/y!}$
$= (1 - 1/n)^{\sum X_i}$

b

Similar to above, let $T = 1$ if $X_1 = 1$ and 0 otherwise. Then again, T is Bernoulli and unbiased, so we have
$E[T|\sum X_i] = P(X_1 = 1|\sum X_i = y)$
$= \bar{X}_n(1 - 1/n)^{\sum X_i - 1}$

c

First, note that $Var(\hat{\lambda}) = \lambda/n$

part a

We can see that the UMVUE is a function of the MLE. $g(x) = (1-1/n)^{nx}$. Then $g'(x) = n(1-1/n)^{nx}\log(1-1/n)$.

By the delta method, as $n \to \infty$, the variance of $g(\hat{\lambda})$ goes to $\frac{\lambda}{n}n^2(1 - 1/n)^{2n\lambda}(\log(1 - 1/n))^2(1/n) = \lambda(1 - 1/n)^{2n\lambda}(\log(1 - 1/n))^2$.

On the other hand, the MLE for $e^{-\lambda}$ is just $e^{-\hat{\lambda}}$, and so its asymptotic variance by the delta method is $(\lambda/n)e^{-2\lambda}(1/n)$.

Then the ratio is $\frac{e^{-2\lambda}}{n^2(1-1/n)^{2n\lambda}(\log(1-1/n))^2} = \frac{e^{-2\lambda}}{(1-1/n)^{2n\lambda}(\log(1-1/n)^n)^2}$

As $n \to \infty$, $(\log(1-1/n)^n)^2 \to (-1)^2 = 1$, so we can drop this term. In addition, $(1-1/n)^{2n\lambda} \to e^{-2\lambda}$, so this cancels out with the numerator, and the entire thing goes to 1. Asymptotically, the two estimators are equivalent.

part b

This time, $g(\lambda) = \lambda(1-1/n)^{n\lambda-1}$, so $g'(\lambda) = (1-1/n)^{n\lambda-1}\big(1+\lambda\log(1-1/n)^n\big) = \frac{n}{n-1}(1-1/n)^{n\lambda}(1+\lambda\log(1-1/n)^n)$.
For large n:

- the first term goes to 1
- the second term goes to $e^{-\lambda}$
- the third term goes to $1-\lambda$

So we are left with $\approx e^{-\lambda}(1-\lambda)$

The MLE of $\lambda e^{-\lambda}$ is also a function of the MLE of λ, and its derivative is $(\lambda-1)e^{-\lambda}$.

Then we can see that taking the ratio of these two yields -1, and squaring it yields 1. Again, asymptotically, the two estimators are equivalent.

d

```
x <- c(10, 7, 8, 13, 8,
       9, 5, 7, 6, 8,
       3, 6, 6, 3, 5)
n <- length(x)

umvue.0 <- (1 - 1 / n) ** sum(x)
umvue.1 <- mean(x) * (1 - 1 / n) ** (sum(x) - 1)

mle.0 <- exp(-mean(x))
mle.1 <- mean(x) * exp(-mean(x))
```

	UMVUE	MLE
$P(X=0)$	7.65×10^{-4}	9.75×10^{-4}
$P(X=1)$	0.005685	0.006758

Weak law of large numbers

Central limit theorem

5.35

5.41