

ANSWER SHEET 4

Assignment 1. (a) Note that $S^2 = (n-1)^{-1} \sum_{i=1}^n X_i^2 - [n/(n-1)]\bar{X}^2$. By the law of large numbers applied to the i.i.d. sequence $\{X_1, \dots, X_n\}$, it follows that $\bar{X} \xrightarrow{p} \mathbb{E}[X_1] = \mu$ as $n \rightarrow \infty$. Thus, the continuous mapping theorem implies that $\bar{X} \xrightarrow{p} \mu^2$ as $n \rightarrow \infty$. We can also apply the law of large numbers to the i.i.d. sequence $\{X_1^2, \dots, X_n^2\}$. Then, it follows that $n^{-1} \sum_{i=1}^n X_i^2 \xrightarrow{p} \mathbb{E}[X_1^2] = \sigma^2 + \mu^2$ as $n \rightarrow \infty$. Since the real-valued sequence $\{n/(n-1)\}$ converges to one as $n \rightarrow \infty$, it follows from Slutsky's theorem that

$$S^2 = \frac{n}{n-1} \times \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} \times \bar{X}^2 \xrightarrow{p} \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

as $n \rightarrow \infty$.

(b) The central limit theorem implies that $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$ as $n \rightarrow \infty$. Using part (a), and the continuous mapping theorem along with Slutsky's theorem, we now have

$$T_n = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \xrightarrow{d} \sigma^{-1} N(0, \sigma^2) \stackrel{d}{=} N(0, 1)$$

as $n \rightarrow \infty$. Here, $\stackrel{d}{=}$ denotes equality in distribution.

(c) If F is the $N(\mu, \sigma^2)$ distribution, we know that T_n has the t distribution with $(n-1)$ degrees of freedom for each $n \geq 2$.

(d) Part (b) says that the exact distribution of T_n converges to the $N(0, 1)$ distribution as $n \rightarrow \infty$. Using part (c), we can say that the $t_{(n-1)}$ distribution converges to the $N(0, 1)$ distribution as $n \rightarrow \infty$. This is equivalent as saying that the t distribution converges to the standard normal distribution as the degrees of freedom tend to infinity. We saw this phenomenon empirically (using R software) in Exercise 7 in Week 3.

Assignment 2. (a) Since \bar{X} is an unbiased estimator of p , it is easy to see that $\bar{X}(1 - \bar{X})$ is a proxy/estimator of $p(1 - p)$. This is a “plug-in” estimator of $p(1 - p)$.

(b) Note that $n\bar{X} = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$. So, $\mathbb{E}[n\bar{X}] = np$ and $\text{Var}(n\bar{X}) = np(1 - p)$. Now,

$$\begin{aligned} \mathbb{E}[U_n] &= n^{-2} \mathbb{E}[n\bar{X}(n - n\bar{X})] = n^{-1} \mathbb{E}[n\bar{X}] - n^{-2} \mathbb{E}[(n\bar{X})^2] \\ &= n^{-1} \times (np) - n^{-2} \times [np(1 - p) + (np)^2] \\ &= (1 - n^{-1})p(1 - p). \end{aligned}$$

So, U_n is not an unbiased estimator of $p(1 - p)$.

(c) By the weak law of large numbers, we know that $\bar{X} \xrightarrow{p} \mathbb{E}[X_1] = p$ as $n \rightarrow \infty$. Using the continuous mapping theorem with $g(x) = x(1 - x)$, $x \in (0, 1)$, it now follows that $U_n = g(\bar{X}) \xrightarrow{p} g(p) = p(1 - p)$ as $n \rightarrow \infty$. So, U_n is a consistent estimator of $p(1 - p)$.

(d) The central limit theorem implies that $\sqrt{n}(\bar{X} - p) \xrightarrow{d} N(0, p(1 - p))$ as $n \rightarrow \infty$. Let $g(x) = x(1 - x)$, $x \in (0, 1)$. Then, $g'(x) = 1 - 2x$. Using the delta method, it now follows that $\sqrt{n}[U_n - p(1 - p)] \xrightarrow{d} N(0, p(1 - p)(1 - 2p)^2)$ as $n \rightarrow \infty$.

(Note : If $p = 1/2$, the above limiting distribution is degenerate as zero. In fact, in that case, the correct scaling to have a non-degenerate distribution is n instead of \sqrt{n} .)

Assignment 3. (a) Note that $\mathbb{E}[X_1] = p$. So, it is an unbiased estimator of p .

(b) V_n is minimally sufficient for p (see Slide 92).

(c) Recall that $\text{Ber}(p)$ distribution belongs to a one parameter exponential family with $\phi = \ln(p/(1-p))$ and $\tau(X_1, \dots, X_n) = \sum_{i=1}^n X_i = V_n$. So, using the theorem in Slide 91, we only need to show that for each fixed $(y_1, \dots, y_n)^\top$ and $(z_1, \dots, z_n)^\top$, the ratio $f(\mathbf{y}, p)/f(\mathbf{z}, p)$ is independent of p if and only if $\sum_{i=1}^n y_i = \sum_{i=1}^n z_i$. Note that

$$\frac{f(\mathbf{y}, p)}{f(\mathbf{z}, p)} = \frac{\exp\{\ln(p/(1-p)) \sum_{i=1}^n y_i + n \ln(1-p)\}}{\exp\{\ln(p/(1-p)) \sum_{i=1}^n z_i + n \ln(1-p)\}} = \exp\left\{\ln\left(\frac{p}{1-p}\right) \left[\sum_{i=1}^n y_i - \sum_{i=1}^n z_i\right]\right\}.$$

Clearly, the term on the extreme right hand side is a constant (as a function of p) if and only if $\sum_{i=1}^n y_i = \sum_{i=1}^n z_i$. Thus, V_n is minimally sufficient for p . Note that the above proof and the fact that the map $p \mapsto \ln(p/(1-p))$ is a bijection also shows that V_n is minimally sufficient for $\ln(p/(1-p))$, which is the natural parameter.

(d) Recall that $V_n \sim \text{Bin}(n, p)$ and $\sum_{i=2}^n X_i \sim \text{Bin}(n-1, p)$. Now, for $k \geq 1$, we have

$$\begin{aligned} \mathbb{P}[X_1 = 1 \mid V_n = k] &= \frac{\mathbb{P}[X_1 = 1, X_1 + \sum_{i=2}^n X_i = k]}{\mathbb{P}[V_n = k]} \\ &= \frac{\mathbb{P}[X_1 = 1, \sum_{i=2}^n X_i = k-1]}{\mathbb{P}[V_n = k]} \\ &= \frac{p \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{k}{n}. \end{aligned}$$

So, $\mathbb{P}[X_1 = 0 \mid V_n = k] = 1 - (k/n)$. Hence, $W_n = \mathbb{E}[X_1 \mid V_n] = V_n/n$.

(Alternative proof : Let $\psi(V_n) := \mathbb{E}[X_1 \mid V_n] = \mathbb{E}[X_1 \mid \sum_{i=1}^n X_i]$ for a function $\psi(\cdot)$. Since the X_i 's are i.i.d., by symmetry, we have

$$\psi(V_n) = \mathbb{E}\left[X_2 \mid \sum_{i=1}^n X_i\right] = \dots = \mathbb{E}\left[X_n \mid \sum_{i=1}^n X_i\right].$$

Thus,

$$n\psi(V_n) = \sum_{j=1}^n \mathbb{E}\left[X_j \mid \sum_{i=1}^n X_i\right] = \mathbb{E}\left[\sum_{j=1}^n X_j \mid \sum_{i=1}^n X_i\right] = \sum_{i=1}^n X_i,$$

where the last equality follows from the fact that $\mathbb{E}[Z \mid Z] = Z$ for any random variable Z . Thus, $\mathbb{E}[X_1 \mid V_n] = \psi(V_n) = n^{-1} \sum_{i=1}^n X_i = V_n/n$.

(e) $\mathbb{E}[W_n] = \mathbb{E}[V_n/n] = (np)/n = p$. Alternatively, $\mathbb{E}[W_n] = \mathbb{E}[\mathbb{E}[X_1 \mid V_n]] = \mathbb{E}[X_1] = p$.

(f) $\text{Var}(W_n) = \text{Var}(V_n/n) = np(1-p)/n^2 = p(1-p)/n \leq p(1-p) = \text{Var}(X_1)$ for all $n \geq 1$. Equality holds if and only if $n = 1$. So, the inequality is strict for all $n \geq 2$, i.e., for all “practical” sample sizes.

(g) Note that

$$\begin{aligned} \log f(\mathbf{X}, p) &= V_n(\ln p) + (n - V_n)(\ln(1-p)) \\ \Rightarrow \frac{\partial^2}{\partial p^2} \log f(\mathbf{X}, p) &= -\frac{V_n}{p^2} - \frac{(n - V_n)}{(1-p)^2} \\ \Rightarrow \mathcal{I}_n(p) &= \mathbb{E}\left[-\frac{\partial^2}{\partial p^2} \log f(\mathbf{X}, p)\right] = \frac{np}{p^2} + \frac{n - np}{(1-p)^2} = \frac{n}{p(1-p)}. \end{aligned}$$

Thus, the Cramer-Rao lower bound for the variance of an unbiased estimator of p is given by $p(1-p)/n$.

This lower bound is attained by the estimator W_n .

Assignment 4. (a) The random variable will equal 2 with probability $p = 0.49$ and 0 with probability $1 - p = 0.51$. Therefore $X = 2Y$ with $Y \sim \text{Ber}(p)$.

(b) Here we have X_1, \dots, X_{1000} independent realisations of X , and we are interested in their sum S_{1000} . By the above, $S/2 \sim \text{Bin}(1000, p)$. Therefore

$$\mathbb{P}(S \geq 1000) = \mathbb{P}(S/2 \geq 500) = \sum_{k=500}^{1000} \binom{1000}{k} p^k (1-p)^{1000-k}.$$

(c) We know that X has expectation $2p = 0.98$ and variance $4p(1-p) = 4 * 0.2499 = 0.9996$. The central limit theorem tells us that

$$\sqrt{n} \frac{S_n/n - 0.98}{\sqrt{.9996}} \rightarrow N(0, 1)$$

as $n \rightarrow \infty$. Therefore

$$\mathbb{P}(S \geq 1000) = \mathbb{P}\left(\sqrt{\frac{1000}{.9996}}[S/1000 - 0.98] \geq 0.02\sqrt{\frac{1000}{.9996}}\right) \approx 1 - \Phi\left(0.2\sqrt{\frac{10}{.9996}}\right) \approx 0.26.$$

(d) This is carried out with the following code. Try changing the parameters p , n and t .

```
p <- 0.49
n <- 1000
mu <- 2*p
s2 <- 4*p*(1-p)
t <- 1000
1 - pnorm(sqrt(n/s2) * (t/n - mu))
f1 <- function(x) pbinom((x*sqrt(s2/n) + mu)*n/2 , size = n, prob = p)
curve(f1, from = -3, to = 3)
curve(pnorm, add=TRUE, col = "blue")
```

The black and blue curves are nearly identical, so the approximation is very good.

(e) This is because here p is fixed, whereas in the law of rare events p goes to zero with n , with $np \rightarrow \lambda$.

Assignment 5. (a) The support $[0, \infty)$ does not depend on λ , so we have a 1-parameter exponential family :

$$f(x; \lambda) = \exp(-\lambda x + 5 \log \lambda + 4 \log x - \log \Gamma(5)) = \exp(\lambda T(x) - \gamma(\lambda) + S(x)).$$

We obtain (be careful with the signs!)

$$\mathbb{E} X = -\mathbb{E} T(X) = -\gamma'(\lambda) = \frac{5}{\lambda}; \quad \text{Var } X = \text{Var } -T(X) = \text{Var } T(X) = \gamma''(\lambda) = \frac{5}{\lambda^2}.$$

This is in accordance with the formulae of slide 64.

(b) This is a 1-parameter exponential family because the support $[1, \infty)$ does not depend on the parameter and

$$f(x; \theta) = \exp(-\theta \log x + \log \theta - \log x) = \exp(\theta T(x) - \gamma(\theta) + S(x)).$$

Consequently (beware of the signs!)

$$\mathbb{E} \log X = -\mathbb{E} T(X) = -\gamma'(\theta) = \frac{1}{\theta}; \quad \text{Var} \log X = \text{Var} -T(X) = \text{Var} T(X) = \gamma''(\theta) = \frac{1}{\theta^2}.$$

Finally $\mathbb{E}[\log^2 X] = [\mathbb{E} \log X]^2 + \text{Var} \log X = 2\theta^{-2}$.

Assignment 6. (i) Let $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} \mathcal{B}(p)$ with $p \in (0, 1)$. The X_i 's are discrete, so the likelihood function

$$V(p) = f_1(x_1; p) \times f_2(x_2; p) \times \dots \times f_n(x_n; p),$$

where $f_i(x_i; p) = P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}$ is the frequency function for each X_i . We have

$$V(p) = p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \dots p^{x_n}(1-p)^{1-x_n} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}.$$

The m.l.e. is the value of p that maximise $L(p) = \log(V(p))$. We have

$$L(p) = \sum_{i=1}^n x_i \log p + \left(n - \sum_{i=1}^n x_i \right) \log(1-p).$$

To find the maximum we solve

$$\begin{aligned} L'(p) &= 0 \\ \Rightarrow \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p} &= 0 \\ \Rightarrow (1-p) \sum_{i=1}^n x_i - p \left(n - \sum_{i=1}^n x_i \right) &= 0 \\ \Rightarrow \sum_{i=1}^n x_i &= p \left(n - \sum_{i=1}^n x_i + \sum_{i=1}^n x_i \right) \\ \Rightarrow p &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n. \end{aligned}$$

To check that is indeed a maximum notice that

$$L''(p) = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n - \sum_{i=1}^n x_i}{(1-p)^2} < 0,$$

for every $p \in (0, 1)$. Hence the value $p = \bar{x}_n$ maximise the function $V(p)$ and \bar{X}_n is the m.l.e, $\hat{p}_{ML} = \bar{X}_n$.

(ii)

– We write the log likelihood function for λ

$$L(\lambda) = \log(\lambda^n e^{-\lambda \sum_{i=1}^n x_i}) = n \log \lambda - \lambda \sum_{i=1}^n x_i,$$

setting it the derivative to zero we find

$$\hat{\lambda}_n = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{X}_n}.$$

The function ℓ_n is concave, hence we have found a maximum.

(iii)

– The likelihood for (μ, σ^2) is

$$V(\mu, \sigma^2) = \prod_{i=1}^n p(x_i; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right).$$

So the log likelihood is

$$\begin{aligned} L(\mu, \sigma^2) = \log V(\mu, \sigma^2) &= \sum_{i=1}^n \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x_i - \mu)^2 \right) \\ &= -\frac{1}{2} \left(n \log(2\pi\sigma^2) + \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right). \end{aligned}$$

Write $w = \sigma^2$. We have

$$\begin{aligned} \frac{\partial \ell}{\partial \mu} &= \frac{1}{w} \sum_{i=1}^n (x_i - \mu), \\ \frac{\partial \ell}{\partial w} &= -\frac{n}{2w} + \frac{1}{2w^2} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

The first partial derivative vanishes when

$$\sum_{i=1}^n (x_i - \mu) = 0$$

hence

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

The second partial derivative vanishes when

$$-n + \frac{1}{w} \sum_{i=1}^n (x_i - \mu)^2 = 0 \tag{1}$$

hence

$$\hat{w} = \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

By direct computation we find that the hessian matrix at $(\hat{\mu}, \hat{w})$ is

$$H|_{(\mu,w)=(\hat{\mu},\hat{w})} = \begin{bmatrix} -\frac{n}{\hat{w}} & 0 \\ 0 & -\frac{n}{2\hat{w}^2} \end{bmatrix}.$$

The matrix is negative definite so $(\hat{\mu}, \hat{w}) = (\hat{\mu}, \hat{\sigma}^2)$ is a maximum.

(iv) Let X_1, X_2, \dots, X_n a sample from a uniform $U[0, \theta]$ with $\theta > 0$. The likelihood function is

$$V(\theta) = f_1(x_1; \theta) \times f_2(x_2; \theta) \times \dots \times f_n(x_n; \theta),$$

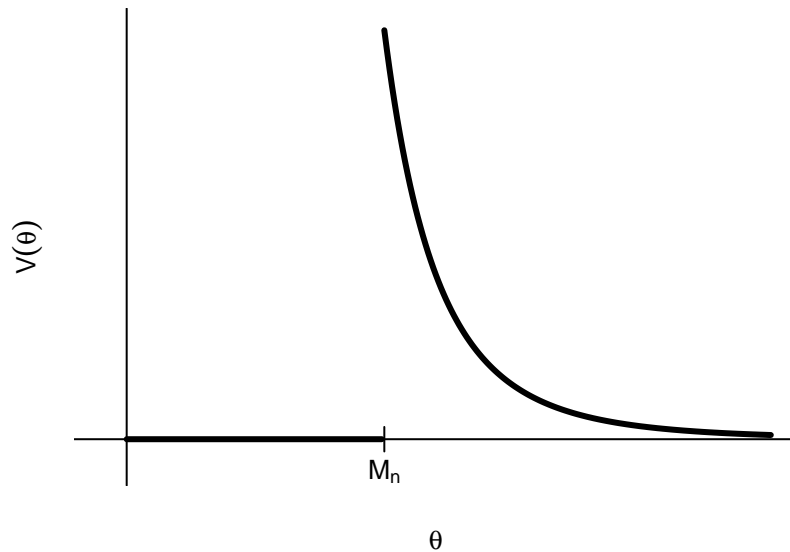
where $f_i(x_i; \theta) = f_i(x_i)$ is the density of each X_i . So

$$V(\theta) = \begin{cases} 1/\theta^n & \text{si } x_i \in [0, \theta] \text{ pour } i \in \{1, \dots, n\} \\ 0 & \text{sinon.} \end{cases}$$

Or else

$$V(\theta) = \begin{cases} 1/\theta^n & \text{si } \max_{i \in \{1, \dots, n\}} x_i \leq \theta \\ 0 & \text{sinon.} \end{cases}$$

Let $M_n = \max(X_1, \dots, X_n)$.



We can see on the figure that the function is maximised for $\theta = M_n$. In particular for $\theta < M_n$ the function equals 0, while for $\theta \geq M_n$ the likelihood is a decreasing function of θ .

Note that $V(\theta)$ is not derivable, hence the maximum cannot be found using $L'(\theta)$ as in the previous exercises.

Assignment 7. (a) By the law of large numbers, $\bar{X}_n \rightarrow \mathbb{E}X = m(\theta)$ in probability as $n \rightarrow \infty$.

(b) Since \bar{X}_n is close to $m(\theta)$, it makes sense to estimate θ by the solution of the equation $m(\tilde{\theta}) = \bar{X}_n$. When m is invertible, this amounts to $\tilde{\theta} = m^{-1}(\bar{X}_n)$. If the inverse is continuous, then by the continuous mapping theorem $\tilde{\theta} = m^{-1}(\bar{X}_n) \rightarrow m^{-1}(m(\theta)) = \theta$ in probability. Therefore, $\tilde{\theta}$ is consistent for θ , a desirable property.

(c) Here we have $m(\lambda) = 1/\lambda$. We obtain the equation

$$\bar{X}_n = m(\tilde{\lambda}) = 1/\tilde{\lambda}$$

so that $\tilde{\lambda} = 1/\bar{X}_n$. The maximum likelihood estimator is the same (see slide 147).

(d) Here we have $\mathbb{E} X = \kappa/2$ and we obtain the equation

$$\bar{X}_n = m(\tilde{\kappa}) = \tilde{\kappa}/2$$

so that $\tilde{\kappa} = 2\bar{X}_n$. The maximum likelihood estimator is $X_{(n)} = \max(X_1, \dots, X_n)$ (see slide 151).

(e) Here we have

$$\mathbb{E} X = \int_1^\infty \theta x^{-\theta} dx = \frac{\theta}{\theta - 1}.$$

(If $\theta \leq 1$ the expectation is infinite.) We obtain the equation

$$\bar{X}_n = m(\tilde{\theta}) = \frac{\tilde{\theta}}{\tilde{\theta} - 1} = 1 + \frac{1}{\tilde{\theta} - 1}$$

so that $\tilde{\theta} = 1 + 1/(\bar{X}_n - 1) = \bar{X}_n/(\bar{X}_n - 1)$.

The log likelihood and its derivatives are

$$\ell(\theta) = n \log \theta - \theta \sum_{i=1}^n \log X_i - \sum_{i=1}^n \log X_i$$

$$\ell'(\theta) = \frac{n}{\theta} - \sum_{i=1}^n \log X_i$$

$$\ell''(\theta) = -\frac{n}{\theta^2} < 0$$

so that $\hat{\theta} = n/\sum \log X_i$ is the maximum likelihood estimator.

(f) For the exponential, the mean squared errors are the same because the estimators are the same.

For the uniform case, $\tilde{\kappa} = 2\bar{X}_n$ is unbiased and has variance $\kappa^2/(3n)$ (slide 62). Its mean squared error is therefore $\kappa^2/(3n)$. The maximum likelihood estimator $\hat{\kappa}$ has density function $n x^{n-1}/\kappa^n$ on $[0, \kappa]$ and 0 otherwise. Thus

$$\mathbb{E} \hat{\kappa} = \int_0^\kappa n \frac{x^n}{\kappa^n} dx = \frac{n\kappa}{n+1}; \quad \mathbb{E} \hat{\kappa}^2 = \int_0^\kappa n \frac{x^{n+1}}{\kappa^n} dx = \frac{n\kappa^2}{n+2}; \quad \text{Var} \hat{\kappa} = \mathbb{E} \hat{\kappa}^2 - \mathbb{E}^2 \hat{\kappa} = \frac{n\kappa^2}{(n+2)(n+1)^2}.$$

We see that $\hat{\kappa}$ is biased with mean squared error

$$[\mathbb{E} \hat{\kappa} - \theta]^2 + \text{Var} \hat{\kappa} = \frac{\kappa^2}{(n+1)^2} + \frac{n\kappa^2}{(n+2)(n+1)^2} = \frac{2\kappa^2}{(n+1)(n+2)}.$$

This behaves like $1/n^2$ whereas the mean squared error of $\tilde{\kappa}$ behaves like $1/n$. Thus, despite being biased, $\hat{\kappa}$ has a smaller mean squared error when $n \geq 3$ is sufficiently large.

(g) For the Pareto case, we have $\mathbb{E} X^2 = \theta/(\theta - 2)$ (infinite if $\theta \leq 2$) and $\text{Var} X = \theta/[(\theta - 2)(\theta - 1)^2]$. By the central limit theorem

$$\sqrt{n} \left(\bar{X}_n - \frac{\theta}{\theta - 1} \right) \rightarrow N \left(0, \frac{\theta}{(\theta - 2)(\theta - 1)^2} \right), \quad n \rightarrow \infty.$$

The function $m^{-1}(x) = 1 + 1/(x - 1)$ is differentiable at $\theta/(\theta - 1) > 1$ with derivative $(\theta - 1)^2$ there. The delta method then gives

$$\sqrt{n}(\tilde{\theta} - \theta) = \sqrt{n}\left(m^{-1}(\bar{X}_n) - m^{-1}\left(\frac{\theta}{\theta - 1}\right)\right) \rightarrow (\theta - 1)^2 N(0, \text{Var} X) = N\left(0, \frac{\theta(\theta - 1)^2}{\theta - 2}\right).$$

The asymptotic variance of the method of moments estimator $\tilde{\theta}$ is $\theta(\theta - 1)^2/[(\theta - 2)n]$. For the maximum likelihood estimator we have $\sqrt{n}(1/\hat{\theta} - 1/\theta) \rightarrow N(0, \theta^{-2})$ and by the delta method the asymptotic variance of $\hat{\theta}$ is θ^2/n . This is smaller than the asymptotic variance of the method of moments estimator because $\theta^2 < \theta(\theta - 1)^2/(\theta - 2)$; the difference is large when θ is close to 2. Both asymptotic variances decay like $1/n$.

Assignment 8. (i). We have seen that the minimal sufficient statistic is $T = \sum_{i=1}^n x_i$ and it's defined over an open set. \bar{T} is the sample mean. From the Theorem in Slide 100 it follows that :

$$\begin{aligned} \mathbb{E}[T] &= n \frac{\partial}{\partial \phi} \gamma(\phi) = ne^\phi & \mathbb{E}[\bar{T}] &= e^\phi \\ \text{Var}[T] &= n \frac{\partial^2}{\partial \phi^2} \gamma(\phi) = ne^\phi & \text{Var}[\bar{T}] &= \frac{e^\phi}{n}. \end{aligned}$$

In particular from the Th. on Slide 114 it follows that

$$\sqrt{n}(\bar{T} - e^\phi) \rightarrow \mathcal{N}(0, e^\phi)$$

in distribution, so \bar{T} is asymptotically $\mathcal{N}(e^\phi, e^\phi/n)$. Consequently T is asymptotically Normal $\mathcal{N}(ne^\phi, ne^\phi)$.

(ii). Here the full code to overlap the two curves. You can play with an increasing N .

```
library(Matrix)
library(stats)

N=10
m=1000
lambda=5
sample <- matrix(rpois(N*m, lambda), N, m)
sample
suff_stat <- colMeans(sample)
suff_stat

# gamma^theta = lambda -> we standardize the sample mean with the true mean and the t

norm_suff_stat <- sqrt(N/lambda)*(suff_stat - lambda)

plot(ecdf(norm_suff_stat), vertical=T)

plot(ecdf(norm_suff_stat), verticals=T)
curve(pnorm, add=TRUE, col = "blue")
```


- (iii). The sum of independent Poisson is again Poisson distributed. We are going to show that $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ are independent for $\lambda, \mu \geq 0$ then $X + Y \sim \text{Poisson}(\lambda + \mu)$. The full proof will follow by induction.

Note that we could also show it directly. For x integer (since X and Y only take values in $\{0\} \cup \mathbb{N}$) we have

$$\mathbb{P}(X+Y = x) = \sum_{k=0}^x \mathbb{P}(X = k, Y = x-k) = \sum_{k=0}^x e^{-\lambda} e^{-\mu} \frac{\lambda^k \mu^{x-k}}{k!(x-k)!} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^x}{x!} \sum_{k=0}^x \binom{x}{k} \frac{\lambda^k \mu^{x-k}}{(\lambda + \mu)^x}.$$

The last formula equals 1 by the binomial formula. Thus $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Assignment 9. (i) Let $X_1, \dots, X_n, Y_1, \dots, Y_n$, be a sample from a Binomial with $P(X_i = x_i) = \binom{n}{x_i} p^{x_i} (1-p)^{1-x_i}$. We have that

$$\frac{f(x; p)}{f(y; p)} = \frac{\binom{n}{x_i} p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}}{\binom{n}{y_i} p^{\sum_{i=1}^n y_i} (1-p)^{n-\sum_{i=1}^n y_i}}$$

Now, we can ignore the factorial constant because we want the ratio to be constant w.r.t. the parameters. We see that the ratio above is independent on p iff $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. Hence by the factorisation theorem $T(y) = \sum_{i=1}^n y_i$ is a minimal sufficient statistics.

- (ii) Let $X_1, \dots, X_n, Y_1, \dots, Y_n$, be a sample from a Poisson distribution. We look at the ratio of Poisson distribution in terms of x and y and we check when the ratio is independent on the parameter λ . In particular

$$\begin{aligned} \frac{f(x; p)}{f(y; p)} &= \frac{e^{n\lambda} \lambda^{\sum_{i=1}^n x_i} \prod_{i=1}^n y_i!}{\prod_{i=1}^n x_i! e^{n\lambda} \lambda^{\sum_{i=1}^n y_i}} \\ &= \lambda^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i} \frac{\prod_{i=1}^n y_i!}{\prod_{i=1}^n x_i!} \end{aligned}$$

The above expression is independent w.r.t. λ iff $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. Hence by the factorisation theorem $T(y) = \sum_{i=1}^n y_i$ is a minimal sufficient statistics.

- (iii) Let X_1, \dots, X_n and Y_1, \dots, Y_n be samples from a Normal distribution, the ratio of densities is

$$\begin{aligned} \frac{f(x)}{f(y)} &= \prod_{i=1}^n \frac{(2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)}{(2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)} \\ &= \exp\left\{ (2\sigma^{-2}) \left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2 + 2\mu \left(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i \right) \right) \right\}. \end{aligned}$$

The above ratio is constant w.r.t. μ if and only if $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$, while it is constant w.r.t. σ^2 iff $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ and $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$ thus $T_1(y) = \sum_{i=1}^n y_i$ is a minimal sufficient statistics for μ and $(T_1(y) = \sum_{i=1}^n y_i, T_2(y) = \sum_{i=1}^n y_i^2)$ is minimal sufficient for σ^2 .

(Note : We could have also taken the sample mean as a minimal sufficient statistics).