1. (a) The log-likelihood function is

$$\ln \mathcal{L}(\lambda) = n\ln(\lambda) - \lambda\sum_{i=1}^n x_i$$

and its first two derivatives are

$$\frac{d}{d\lambda}\ln\mathcal{L}(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i$$
$$\frac{d^2}{d\lambda^2}\ln\mathcal{L}(\lambda) = -\frac{n}{\lambda^2}$$

Setting the first derivative to 0, we obtain $\widehat{\lambda}_n = 1/\bar{X}$ and from the second derivative it follows that $I(\lambda) = 1/\lambda^2$. Therefore, $\sqrt{n}(\widehat{\lambda}_n - \lambda) \xrightarrow{d} \mathcal{N}(0, \lambda^2)$.

(b) Find a and b so that

$$P(a \le \chi^2(2n) \le b) = p$$

where p is the desired confidence level. Then

$$p = P\left(a \le 2\lambda\sum_{i=1}^n X_i \le b\right) = P\left(\frac{a}{2\sum_{i=1}^n X_i} \le \lambda \le \frac{b}{2\sum_{i=1}^n X_i}\right)$$

and so the interval

$$\left[\frac{a}{2\sum_{i=1}^n X_i}, \frac{b}{2\sum_{i=1}^n X_i}\right]$$

is a $100p\%$ confidence interval for λ.

(c) From part (b), we know that under H_0, $T \sim \chi^2(2n)$. Using the N-P Lemma, if $\lambda_1 > 1$ we have

$$\frac{f(x_1,\cdots,x_n;\lambda_1)}{f(x_1,\cdots,x_n;1)} = \lambda_1^n \exp\left((1-\lambda_1)\sum_{i=1}^n x_i\right),$$

which (since $1-\lambda_1 < 0$) is a decreasing function of $\sum_{i=1}^n x_i$ and hence of $2\sum_{i=1}^n x_i$. Therefore the MP α level test of $\lambda = 1$ versus $\lambda = \lambda_1 > 1$ rejects H_0 when $T \le c_\alpha$ where c_α is the α quantile of the $\chi^2(2n)$ distribution. We get the same test for any $\lambda_1 > 1$ so this is a UMP test of H_0 versus H_1.

2. See notes for more details. The key points are given below.

(a) The Hill estimator: Bias increases with k, variance decreases with k. Choice of k can be facilitated by Hill plot but interpretation of this is not easy.

(b) Kernel density estimation: Bias and variance depend on the bandwidth and kernel but the bandwidth parameter is much more important. The bias increases as the bandwidth increases and the variance decreases. In addition, the bias depends on the unknown density, which makes the choice of the bandwidth more complicated.

(c) Non-parametric regression using kernel (weighted average) smoothing: Similar issues as for kernel density estimation with respect to bias and variance depending on the bandwidth and weights (kernel) – the bias depends on the unknown function. We have somewhat more flexibility in kernel smoothing since we can allow the kernel to have negative weights.

3. (a) The posterior distribution of $\{f_0, f_1\}$ $\pi(f_k|x_1,\cdots,x_n)$ is proportional to $\pi_k f_k(x_1,\cdots,x_n)$. Since $\pi(f_0|x_1,\cdots,x_n) + \pi(f_1|x_1,\cdots,x_n) = 1$, it follows that

$$\pi(f_k|x_1,\cdots,x_n) = \frac{\pi_k f_k(x_1,\cdots,x_n)}{\pi_0 f_0(x_1,\cdots,x_n) + \pi_1 f_1(x_1,\cdots,x_n)}$$

(b) $\pi(f_0|x_1,\cdots,x_n) > \pi(f_1|x_1,\cdots,x_n)$ if $f_0(x_1,\cdots,x_n) > f_1(x_1,\cdots,x_n)$ when $\pi_0 = \pi_1 = 1/2$.

(c) **This is not easy!** Following the hint,

$$\frac{1}{n}\ln\left(\frac{\pi(f_0|x_1,\cdots,x_n)}{\pi(f_1|x_1,\cdots,x_n)}\right) = \frac{1}{n}\ln(\pi_0/pi_1) + \frac{1}{n}\sum_{i=1}^n \ln\left(\frac{g_0(x_i)}{g_1(x_i)}\right)$$

If $\pi_0 = 1$ then $\pi(f_0|x_1,\cdots,x_n) = 1$ and so we will assume that $0 < \pi_0 < 1$. Under this assumption, $\ln(\pi_0/pi_1)/n \to 0$ and

$$\frac{1}{n}\sum_{i=1}^n \ln\left(\frac{g_0(X_i)}{g_1(X_i)}\right) \xrightarrow{p} E_0\left[\ln\left(\frac{g_0(X_i)}{g_1(X_i)}\right)\right] = a > 0$$

From this, it follows that for large n,

$$\frac{\pi(f_0|x_1,\cdots,x_n)}{\pi(f_1|x_1,\cdots,x_n)} \approx \exp(na) \to \infty$$

which means that $\pi(f_0|x_1,\cdots,x_n) \to 1$.

4. (a) Trimmed mean: $(x_{(r+1)} + \cdots + x_{(n-r)})/(n-2r)$; Winsorized mean: $(x_{(r+1)} + \cdots + x_{(n-r)} + rx_{(r+1)} + rx_{(n-r)})/n$.

(b) 95% confidence interval: $0.289 \pm 1.96\sqrt{0.1052193}$.

(c) Take $\alpha = r/n$. Then

$$\theta(F) = \alpha F^{-1}(\alpha) + \int_\alpha^{1-\alpha} F^{-1}(t)\,dt + \alpha F^{-1}(1-\alpha)$$

5. Note that the empirical distribution function increases most rapidly for x close to 0.5 and increases more slowly as x increases. This corresponds to a density estimate that is highest around $x = 0.5$ and decreasing as x increases. The only density estimate of the four given that satisfies these criteria is (d).

1. Suppose that $X_1, X_2, \cdots$ are independent random variables with common density function

$$f(x) = \alpha x^{-\alpha-1} \quad \text{for } x \ge 1$$

where $\alpha > 0$.
(a) Let $V_n = \max(X_1,\cdots,X_n)$. Find the density function of V_n. (Be sure to specify the range of V_n.)
(b) Show that $n^{-1/\alpha}V_n \xrightarrow{d} V$ where the random variable V has distribution function

$$F_V(x) = \exp\left(-x^{-\alpha}\right) \quad \text{for } x > 0.$$

2. Suppose that $X_1, X_2, \cdots$ are independent random variables with $E(X_i) = 0$ and $E(X_i^2) = 2$.
(a) Define

$$U_n = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)^{1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\right).$$

Find the limiting distribution of U_n as $n \to \infty$.
(b) Define

$$V_n = \frac{1}{n}\sum_{k=1}^n (-1)^k X_k$$

Show that $V_n \xrightarrow{p} 0$.

3. Suppose that $X_1,\cdots,X_n$ are independent discrete random variables with mass function

$$f(x;\theta) = \left(\frac{\theta}{4}\right)^{|x|}\left(1-\frac{\theta}{2}\right)^{1-|x|} \quad \text{for } x = -1, 0, 1$$

where $0 < \theta < 1$.
Show that the maximum likelihood estimator of θ is

$$\widehat{\theta}_n = \frac{2}{n}\sum_{i=1}^n |X_i|$$

and find the limiting distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$. (You may assume that all the appropriate regularity conditions are satisfied.)

4. Suppose that $X_1,\cdots,X_n$ are independent Bernoulli random variables whose probability mass function is

$$f(x;\theta) = \theta^x(1-\theta)^{1-x} \quad \text{for } x = 0, 1$$

1. [15 marks] Suppose that $X_1,\cdots,X_n$ are independent exponential random variables with density

$$f(x;\lambda) = \lambda\exp(-\lambda x) \quad \text{for } x \ge 0$$

where $\lambda > 0$ is an unknown parameter.
(a) Find the maximum likelihood estimator $\widehat{\lambda}_n$ of λ and find the limiting distribution of $\sqrt{n}(\widehat{\lambda}_n - \lambda)$.
(b) A pivot for λ is $2\lambda\sum_{i=1}^n X_i$, which has a χ^2 distribution with $2n$ degrees of freedom. Show (giving as much detail as possible) how you can use this pivot to construct a confidence interval for λ.
(c) Suppose we want to test the null hypothesis $H_0: \lambda = 1$ versus $H_1: \lambda > 1$ using the test statistic $T = 2\sum_{i=1}^n X_i$. For an α level test, for what values of T would you reject H_0? Give as much detail as possible.

2. [10 marks] Discuss bias/variance tradeoff in the context of **one** of the following estimation procedures.
(a) The Hill estimator
(b) Kernel density estimation
(c) Non-parametric regression using kernel (weighted average) smoothing

3. [15 marks] Suppose that $(X_1,\cdots,X_n)$ have a joint density $f(x_1,\cdots,x_n)$ where f is either f_0 or f_1 (where both f_0 and f_1 have no unknown parameters). We put a prior distribution on the possible densities $\{f_0, f_1\}$: $\pi(f_0) = \pi_0 > 0$ and $\pi(f_1) = \pi_1 > 0$ where $\pi_0 + \pi_1 = 1$.

(a) Show that the posterior distribution of $\{f_0, f_1\}$ is

$$\pi(f_k|x_1,\cdots,x_n) = \tau(x_1,\cdots,x_n)\pi_k f_k(x_1,\cdots,x_n) \quad \text{for } k = 0, 1$$

and give the value of the normalizing constant $\tau(x_1,\cdots,x_n)$.
(b) Suppose that $\pi_0 = \pi_1 = 1/2$. When will $\pi(f_0|x_1,\cdots,x_n) > \pi(f_1|x_1,\cdots,x_n)$?
(c) Suppose now that $X_1,\cdots,X_n$ are independent random variables with common density g where g is either g_0 or g_1 so that

$$f_k(x_1,\cdots,x_n) = g_k(x_1)g_k(x_2)\times\cdots\times g_k(x_n) \quad \text{for } k = 0, 1.$$

If g_0 is the true density of $X_1,\cdots,X_n$ and $\pi_0 > 0$, show that

$$\pi(f_0|x_1,\cdots,x_n) \xrightarrow{p} 1 \quad \text{as } n \to \infty.$$

(Hint: Look at $n^{-1}\ln(\pi(f_0|x_1,\cdots,x_n)/\pi(f_1|x_1,\cdots,x_n))$.)

4. [15 marks] Suppose that $X_1,\cdots,X_n$ are independent random variables with common density function $f(x-\theta)$ (where $f(x) = f(-x)$). An estimator of the parameter θ is the Winsorized mean

$$\widehat{\theta} = \frac{1}{n}\left\{r(X_{(r+1)} + X_{(n-r)}) + \sum_{i=r+1}^{n-r} X_{(i)}\right\}$$

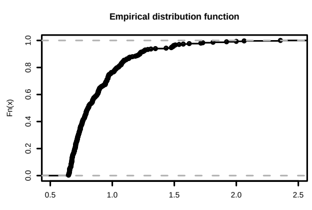
where $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ are the order statistics.
(a) How does the Winsorized mean differ from the trimmed mean?
(b) The R function `winsor(x,r)` computes the Winsorized mean of data `x` where `r` is the parameter r defined above. The R code below computes a jackknife variance estimate of $\widehat{\theta}$. Assuming that the data come from a symmetric distribution, give an approximate 95% confidence interval for θ. (The 0.975 quantile of a standard normal distribution is 1.96.)

```
> m10 <- winsor(x,r=10)
> m10
[1] 0.2894375
> mi <- NULL # leave-one-out estimates
> for (i in 1:100) {
+     mi <- c(mi,winsor(x[-i],r=10))
+     }
> mdot <- mean(mi)

> jackvar <- 99*sum((mi-mdot)^2)/100
> jackvar
[1] 0.1052193
```

(c) Suppose that the density $f(x)$ is not symmetric. What parameter $\theta(F)$ is the Winsorized mean estimating in general?

5. [10 marks] Suppose we have data consisting of 300 observations from an unknown distribution. The plot below is the empirical distribution function of these observations $x_1,\cdots,x_{300}$. On the following page, there are four density estimates. Which one of the four ((a), (b), (c), or (d)) corresponds to the density estimate? Justify your answer.



Let $T = \sum_{i=1}^n X_i$ and recall that

$$P_\theta(T = t) = \binom{n}{t}\theta^t(1-\theta)^{n-t} \quad \text{for } t = 0,\cdots,n$$

(a) State why T is sufficient for θ. (Hint: write the joint probability mass function as a 1 parameter exponential family.)
(b) Find an unbiased estimator of $\theta(1-\theta)$ based on T when $n = 2$.

5. Let X and Y be independent random variables with $E(X) = E(Y) = \theta$, $\text{Var}(X) = \sigma_1^2$ and $\text{Var}(Y) = \sigma_2^2$. Assume that σ_1^2 and σ_2^2 are known and consider estimators of θ of the form

$$\widehat{\theta} = aX + (1-a)Y.$$

Find the value of a that minimizes $\text{Var}(\widehat{\theta})$.

6. Suppose that $X_1,\cdots,X_n$ are independent random variables with common distribution whose mean is θ and variance $k\theta^2$ where k is a known constant. Let

$$V_n = \frac{1}{n(k+1)}\sum_{i=1}^n X_i^2$$

and

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i.$$

(a) Show that V_n is a consistent estimator of θ^2. Is it unbiased? Justify your answer.
(b) Show that $\bar{X}_n^2$ is a consistent estimator of θ^2 and find the limiting distribution of $\sqrt{n}(\bar{X}_n^2 - \theta^2)$.

7. Let $X_1,\cdots,X_n$ be independent random variables where the density function of X_i is

$$f_i(x;\beta) = \frac{1}{\beta t_i}\exp(-x/(\beta t_i)) \quad \text{for } x \ge 0$$

1. (a) The likelihood function is

$$\mathcal{L}(\theta) = \frac{2^n}{\theta^{2n}}\prod_{i=1}^n x_i \quad \text{for } \theta \ge \max\{x_1,\cdots,x_n\}$$

with $\mathcal{L}(\theta) = 0$ when $\theta < \max\{x_1,\cdots,x_n\}$. Thus $\widehat{\theta} = X_{(n)} = \max\{X_1,\cdots,X_n\}$.

(b) The posterior density is

$$\pi(\theta|x_1,\cdots,x_n) \propto \pi(\theta)\mathcal{L}(\theta)$$

where $\mathcal{L}(\theta) = \pi(\theta|x_1,\cdots,x_n) = 0$ for $\theta < T(x_1,\cdots,x_n) = \max\{x_1,\cdots,x_n\}$. For $\theta \ge T(x_1,\cdots,x_n)$, we have

$$\pi(\theta|x_1,\cdots,x_n) = k(x_1,\cdots,x_n)\theta^{-2n}(1+\theta^2)^{-1}$$

where

$$k(x_1,\cdots,x_n) = \left\{\int_{T(x_1,\cdots,x_n)}^\infty \theta^{-2n}(1+\theta^2)^{-1}\,d\theta\right\}^{-1}$$

(c) Note that $\pi(\theta|x_1,\cdots,x_n)$ is decreasing for $\theta \ge T(x_1,\cdots,x_n)$. Thus the HPD interval has the form $[T(x_1,\cdots,x_n), S(x_1,\cdots,x_n)]$ where

$$\int_{T(x_1,\cdots,x_n)}^{S(x_1,\cdots,x_n)} \pi(\theta|x_1,\cdots,x_n)\,d\theta = 0.95.$$

2. (i) By the Delta Method, $g(\bar{X})$ is approximately normal with mean $g(\mu)$ and variance $[g'(\mu)]^2\text{Var}_F(X_i)/n$. Using the substitution principle to estimate $g'(\mu)$ and $\text{Var}_F(X_i)$, we obtain the estimate

$$\widehat{\text{se}}(g(\bar{X})) = \left\{\frac{[g'(\bar{X})]^2}{n(n-1)}\sum_{i=1}^n (X_i - \bar{X})^2\right\}^{1/2}$$

Note that we could replace the $n(n-1)$ in the denominator by n^2.

(ii) The second approach to use the jackknife. If we define $\theta = g(\mu)$ and $\widehat{\theta} = g(\bar{X})$ then the leave-one-out estimates are $\widehat{\theta}_{-i} = g(\bar{X}_{-i})$ where

$$\bar{X}_{-i} = \frac{1}{n-1}\sum_{j\ne i} X_j$$

Then

$$\widehat{\text{se}}(g(\bar{X})) = \left\{\frac{n-1}{n}\sum_{i=1}^n (\widehat{\theta}_{-i} - \widehat{\theta}_\bullet)^2\right\}^{1/2}$$

where $\widehat{\theta}_\bullet$ is the average of $\{\widehat{\theta}_{-i}\}$.

3. (a) The log-likelihood function is

$$\ln\mathcal{L}(\theta) = n\ln(\theta) + n\ln(1+\theta) + (\theta-1)\sum_{i=1}^n \ln(x_i) + \sum_{i=1}^n \ln(1-x_i)$$

and its derivative is

$$\frac{d}{d\theta}\ln\mathcal{L}(\theta) = \frac{n}{\theta} + \frac{n}{1+\theta} + \sum_{i=1}^n \ln(x_i)$$

and so the MLE is the positive solution of

$$\frac{n}{\widehat{\theta}} + \frac{n}{1+\widehat{\theta}} + \sum_{i=1}^n \ln(X_i) = 0$$

(This can be determined by the quadratic formula.)

(b) The observed Fisher information is

$$-\frac{d^2}{d\theta^2}\ln\mathcal{L}(\widehat{\theta}) = \frac{n}{\widehat{\theta}^2} + \frac{n}{(1-\widehat{\theta})^2}$$

which gives

$$\widehat{\text{se}}(\widehat{\theta}) = \left\{\frac{n}{\widehat{\theta}^2} + \frac{n}{(1-\widehat{\theta})^2}\right\}^{-1/2}$$

(c) The form of the test statistic can be determined by

$$\ln\mathcal{L}(\theta) - \ln\mathcal{L}(1) = (\theta-1)\sum_{i=1}^n \ln(X_i) + \text{terms independent of } X_1,\cdots,X_n$$

We will reject H_0 when $\ln\mathcal{L}(\theta) - \ln\mathcal{L}(1)$ is large for $\theta > 1$, which is equivalent to $T = \sum_{i=1}^n \ln(X_i) > k$ for some k satisfying

$$P_{\theta=1}(T > k) = \alpha.$$

(Note that T is a negative random variable and so the constant k will also be negative.)

4. (a) We have

$$E[\widehat{g}(x)] = \frac{1}{2nh}\sum_{i=1}^n I(x-h \le x_i \le x+h)E(Y_i) = \frac{1}{2nh}\sum_{i=1}^n I(x-h \le x_i \le x+h)g(x_i)$$

and

$$\text{Var}[\widehat{g}(x)] = \frac{1}{4n^2h^2}\sum_{i=1}^n I(x-h \le x_i \le x+h)^2\text{Var}(Y_i) = \frac{\sigma^2}{4n^2h^2}\sum_{i=1}^n I(x-h \le x_i \le x+h).$$

since $\text{Var}(Y_i) = \sigma^2$ and $I(x-h \le x_i \le x+h)^2 = I(x-h \le x_i \le x+h)$.

(b) The MSE is

$$\text{bias}^2 + \text{variance} \approx \left(\frac{h^2}{3}\right)^2 + \frac{\sigma^2}{2nh}$$

Its derivative with respect to h is

$$\frac{4h^3}{9} - \frac{\sigma^2}{2nh^2}$$

Setting this to 0 and solving, we get

$$h = \left(\frac{9\sigma^2}{8n}\right)^{1/5}.$$

(Note that h above does truly minimize the approximate MSE since the second derivative is positive.)

5. (a) Applying the chain rule, we get

$$g(x) = G'(x) = \theta F(x)^{\theta-1}F'(x) = \theta F(x)^{\theta-1}f(x)$$

(b) There are a number of approaches here. The obvious one is to use the fact that $\ln G(x) = \theta\ln F(x)$ and estimate F and G by their empirical stributions so that

$$\ln\widehat{G}(x) \approx \theta\ln\widehat{F}(x)$$

and plot $\ln\widehat{G}(x_j)$ versus $\ln\widehat{F}(x_j)$ at some points $x_1,\cdots,x_k$, estimating θ by the slope of these points. Alternatively, we could estimate θ by "pseudo"-maximum likelihood, i.e. maximizing

$$\ln\mathcal{L}(\theta) = n\ln(\theta) + (\theta-1)\sum_{i=1}^n \ln\widehat{F}(y_i)$$

where we may need to modify $\widehat{F}$ so that $\widehat{F}(y_i) > 0$ for all i.

and $t_1,\cdots,t_n$ are known constants. (Note that X_i has an Exponential distribution.)
(a) Show that

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^n X_i/t_i$$

is an unbiased estimator of β.
(b) Show that $(\widehat{\beta} - \beta)/\beta$ is an approximate pivot for β and find an approximate 95% confidence interval for β.

8. Suppose that $X_1, X_2, \cdots$ are independent uniform random variables on $[0, 1]$. Suppose that $\{\widehat{\theta}_n\}$ is a sequence of estimators with $\widehat{\theta}_n > 0$ satisfying

$$\frac{1}{\widehat{\theta}_n} + \frac{1}{\widehat{\theta}_n + 1} = -\frac{1}{n}\sum_{i=1}^n \ln(X_i)$$

(a) Show that $\widehat{\theta}_n \xrightarrow{p} \theta_0$ and find the value of θ_0.

1. [15 marks] Suppose that $X_1,\cdots,X_n$ are independent random variables with common density function

$$f(x;\theta) = \begin{cases} 2x/\theta^2 & \text{for } 0 \le x \le \theta \\ 0 & \text{otherwise}\end{cases}$$

where $\theta > 0$ is an unknown parameter.

(a) Find the maximum likelihood estimator (MLE) of θ.

(b) Suppose that the prior density for θ is

$$\pi(\theta) = \frac{2}{\pi(1+\theta^2)} \quad \text{for } \theta > 0$$

Given data $x_1,\cdots,x_n$, show that the posterior density of θ is

$$\pi(\theta|x_1,\cdots,x_n) = \begin{cases} k(x_1,\cdots,x_n)\theta^{-2n}(1+\theta^2)^{-1} & \text{for } \theta \ge T(x_1,\cdots,x_n) \\ 0 & \text{otherwise}\end{cases}$$

and give expressions for $k(x_1,\cdots,x_n)$ and $T(x_1,\cdots,x_n)$. (You can write $k(x_1,\cdots,x_n)$ as an integral.)

(c) Give the form of a 95% highest posterior density (HPD) interval for θ. (You don't need an explicit formula, just state how you would compute the left and right endpoints of the interval.)

2. [10 marks] Suppose that $X_1,\cdots,X_n$ are independent random variables with unknown distribution function F. The sample mean $\bar{X} = n^{-1}(X_1+\cdots+X_n)$ is a substitution principle estimator of $\mu = E_F(X_i)$ and $g(\bar{X})$ is a substitution principle estimator of $g(\mu)$.

Assuming that g is differentiable with derivative g', describe two approaches to estimating the standard error of $g(\bar{X})$. Give as much detail as possible.

3. [15 marks] Suppose that $X_1,\cdots,X_n$ are independent continuous random variables with density

$$f(x;\theta) = \theta(\theta+1)x^{\theta-1}(1-x) \quad \text{for } 0 \le x \le 1$$

where $\theta > 0$.

(a) Find the MLE of θ based on $X_1,\cdots,X_n$. (You do **not** need to show that your estimator maximizes the likelihood function.)

(b) Give an estimator of the standard error of the MLE from part (a).

(c) Suppose we want to test the null hypothesis $H_0: \theta = 1$ versus the alternative hypothesis $H_1: \theta > 1$ using the test statistic $T = \sum_{i=1}^n \ln(X_i)$. Does the uniformly most powerful (UMP) α-level test reject for $T > k$ or for $T < k$? What condition must the constant k satisfy?

4. [10 marks] Consider the non-parametric regression model

$$Y_i = g(x_i) + \varepsilon_i \quad \text{for } i = 1,\cdots,n$$

where $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2$ and $x_i = i/(n+1)$. In this case, we can estimate $g(x)$ for $x \in (0, 1)$ by

$$\widehat{g}(x) = \frac{1}{2nh}\sum_{i=1}^n I(x-h \le x_i \le x+h)Y_i.$$

(a) Show that

$$E[\widehat{g}(x)] = \frac{1}{2nh}\sum_{i=1}^n I(x-h \le x_i \le x+h)g(x_i)$$
$$\text{and} \quad \text{Var}[\widehat{g}(x)] = \frac{\sigma^2}{4n^2h^2}\sum_{i=1}^n I(x-h \le x_i \le x+h).$$

(b) Suppose that $g(x) = x^2$. In this case, we can approximate the bias and variance of $\widehat{g}(x)$ by

$$E[\widehat{g}(x)] - g(x) \approx \frac{h^2}{3} \quad \text{and} \quad \text{Var}[\widehat{g}(x)] \approx \frac{\sigma^2}{2nh}.$$

(You do **not** need to prove these!) Find the value of h that minimizes the (approximate) mean square error of $\widehat{g}(x)$.

5. [10 marks] Suppose that $X_1,\cdots,X_m$, $Y_1,\cdots,Y_n$ are two independent samples where $X_1,\cdots,X_m$ are independent with distribution function $F(x)$ and density $f(x)$, and $Y_1,\cdots,Y_n$ are independent with distribution function $G(x) = F(x)^\theta$ for some $\theta > 0$.

(a) Show that the common density function of $\{Y_i\}$ is

$$g(x) = \theta F(x)^{\theta-1}f(x).$$

(b) Given $\{X_i\}$ and $\{Y_i\}$, how might you estimate θ if $F(x)$ is unknown? (Hint: There is no single right answer to this question but a good starting point is to note that

$$\ln G(x) = \theta\ln F(x)$$

where F and G can be estimated by the empirical distribution functions of $\{X_i\}$ and $\{Y_i\}$, respectively. Alternatively, you might propose a Bayesian approach where you put prior distributions on both θ and F.)

(b) Show that $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \sigma^2)$ and find the value of σ^2.

9. Suppose that $X_1, X_2, \cdots$ are indpendent random variables such that

$$P(X_i = 0) = p \quad \text{and} \quad P(X_i > x) = (1-p)\exp(-\lambda x) \quad \text{for } x \ge 0.$$

(a) Define $Y = X_1 + \cdots + X_n$. How would you compute the distribution of Y if n is small? If n is large?
(b) Suppose that $p = \exp(-\lambda)$ where λ is an unknown parameter. Find a consistent estimator of λ based on $\sum_{i=1}^n X_i$.
(c) Find an approximate standard error for the estimator in part (b).
(d) Assuming that $p = \exp(-\lambda)$, the likelihood function for λ is

$$\mathcal{L}(\lambda) = \exp\left(-\lambda\sum_{i=1}^n I(x_i = 0)\right)\prod_{i:x_i>0}\{\lambda(1-\exp(-\lambda))\exp(-\lambda x_i)\}$$

Find an expression for the MLE of λ.
(e) How might you estimate the standard error of the MLE in part (d)?

10. Suppose that $X_1,\cdots,X_n$ are independent uniform random variables on $[0, \theta]$ and we test

$$H_0: \theta = 1 \quad \text{versus} \quad H_1: \theta \ne 1$$

at level α. We will reject H_0 if $T > 1$ or $T < k$ where $T = \max(X_1,\cdots,X_n)$.
(a) Find an expression for k to make this a level α test.
(b) Find the power of this test as a function of θ. For what value of θ is the power minimized? (Hint: consider the cases $\theta < 1$ and $\theta > 1$ separately.)

11. Suppose that $X_1,\cdots,X_n$ are independent random variables with density

$$f(x;\lambda) = \lambda\exp(-\lambda x) \quad \text{for } x \ge 0$$

and $Y_1,\cdots,Y_m$ are independent random variables (which are also independent of the X_i's) with density

$$g(y;\theta) = \theta\exp(-\theta y) \quad \text{for } y \ge 0.$$

(a) Show that $2\lambda\sum_{i=1}^n X_i$ and $2\theta\sum_{i=1}^m Y_i$ are independent χ^2 random variables with $2n$ and $2m$ degrees of freedom respectively.
(b) Using the pivot

$$\frac{\lambda\bar{X}}{\theta\bar{Y}}$$

describe in detail how you would construct a 95% confidence interval for λ/θ. What is the distribution of the pivot?

12. Suppose that $X_1,\cdots,X_n$ be independent, identically distributed random variables with frequency function

$$f(x;\theta) = \theta(1-\theta)^x \quad \text{for } x = 0, 1, 2, \cdots$$

where $0 < \theta < 1$.
(a) Show that $\sum_{i=1}^n X_i$ is a sufficient statistic for θ and give its expected value and variance.

1. [20 marks] Suppose that $X_1,\cdots,X_n$ are independent continuous random variables with density

$$f(x;\theta) = \theta x^{\theta-1} \quad \text{for } 0 \le x \le 1$$

where $\theta > 0$.
(a) Find the MLE of θ and give an estimate of its standard error based on the observed Fisher information.
(b) Another estimator of θ is $\widetilde{\theta} = \bar{X}/(1-\bar{X})$. Show that as the sample size $n \to \infty$, $\widetilde{\theta} = \widetilde{\theta}_n \xrightarrow{p} \theta$.
(c) An estimator of the standard error of $\widetilde{\theta}$ defined in part (b) is

$$\widehat{\text{se}}(\widetilde{\theta}) = \frac{1}{(1-\bar{X})^2\sqrt{n}}\left\{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2\right\}^{1/2}.$$

Justify its use as an estimator of the standard error. (Hint: Use the delta method to approximate the distribution of $\widetilde{\theta}$.)

2. [10 marks] Suppose that $X_1,\cdots,X_n$ are independent uniform random variables on the interval $[0, \theta]$; that is, the density is

$$f(x;\theta) = \begin{cases} 1/\theta & \text{if } 0 \le x \le \theta \\ 0 & \text{otherwise}\end{cases}$$

(a) The density of $X_{(n)} = \max(X_1,\cdots,X_n)$ is

$$g(x;\theta) = \frac{n}{\theta^n}x^{n-1} \quad \text{for } 0 \le x \le \theta.$$

(You do not need to show this.) Show that $X_{(n)}/\theta$ is a pivot for θ.
(b) Use the pivot $X_{(n)}/\theta$ to find an exact $100p\%$ confidence interval of the form $[X_{(n)}, aX_{(n)}]$ where $a > 1$ will depend on p and n.

3. [15 marks] The following R code computes the 10% trimmed mean (that is, the largest 10% and smallest 10% of the observations are removed and the middle 80% averaged) of a sample of size 100 and computes a jackknife variance estimate. (Assume that the vector `x` contains 100 observations.)

```
> m10 <- mean(x,trim=0.1)
> m10
[1] 9.888978
> mi <- NULL # leave-one-out estimates
> for (i in 1:100) {
+    mi <- c(mi,mean(x[-i],trim=0.1))
+    }
> mdot <- mean(mi)
> jackvar <- 99*sum((mi-mdot)^2)/100
> jackvar
[1] 0.03750031
```

(a) Assuming that the data $x_1,\cdots,x_{100}$ come from a density $f(x-\theta)$ where $f(x)$ is symmetric around 0, give an approximate 95% confidence interval for θ. (The 0.975 quantile of a standard normal distribution is 1.96.)
(b) For the confidence interval in part (a), what is the approximate pivot that is being used?
(c) Suppose that we cannot assume a symmetric distribution as in part (a). In this case, the interval in part (a) is a confidence interval for some $\theta(F)$. Give an expression for $\theta(F)$. (You may assume that F is a continuous distribution although this is not absolutely necessary.)

4. [10 marks] Describe one of the methods below (either (a) or (b)) giving as much detail as possible.
(a) Kernel density estimation.
(b) Normal quantile-quantile plot.

1. [20 marks] Suppose that $X_1,\cdots,X_n$ are independent continuous random variables with density

$$f(x;\theta) = \theta x^{\theta-1} \quad \text{for } 0 \le x \le 1$$

where $\theta > 0$.
(a) The log-likelihood function is

$$\ln L(\theta) = n\ln(\theta) + (\theta-1)\sum_{i=1}^n \ln(x_i)$$

and its derivatives are

$$\frac{d}{d\theta}\ln L(\theta) = \frac{n}{\theta} + \sum_{i=1}^n \ln(x_i)$$
$$\frac{d^2}{d\theta^2}\ln L(\theta) = -\frac{n}{\theta^2}$$

Thus the MLE is

$$\widehat{\theta} = -\left(\frac{1}{n}\sum_{i=1}^n \ln(X_i)\right)^{-1}$$

and its estimated standard error is

$$\widehat{\text{se}}(\widehat{\theta}) = \left(-\frac{d^2}{d\theta^2}\ln L(\widehat{\theta})\right)^{-1/2} = \frac{\widehat{\theta}}{\sqrt{n}}$$

(b) By the WLLN,

$$\bar{X} \xrightarrow{p} E_\theta(X_i) = \int_0^1 x\theta x^{\theta-1}\,dx = \frac{\theta}{\theta+1}$$

The estimator $\widetilde{\theta} = \bar{X}/(1-\bar{X}) = g(\bar{X})$ where $g(x) = x/(1-x)$ is a continuous function. Thus

$$g(\bar{X}) \xrightarrow{p} g(\theta/(1+\theta)) = \theta.$$

(c) Using the delta method

$$\sqrt{n}(\widetilde{\theta} - \theta) = \sqrt{n}(g(\bar{X}) - g(\theta/(1+\theta))) \xrightarrow{d} N(0, [g'(\theta/(1+\theta))]^2\sigma^2)$$

where $\sigma^2 = \text{Var}_\theta(X_i)$. Now σ^2 can be estimated by the sample variance

$$\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

while $g'(\theta/(1+\theta))$ can be estimated by $g'(\bar{X}) = 1/(1-\bar{X})^2$. Thus

$$\widehat{\text{se}}(\widetilde{\theta}) = \frac{1}{(1-\bar{X})^2\sqrt{n}}\left\{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2\right\}^{1/2}.$$

2. (a) Note that $X_{(n)}/\theta = \max(X_1/\theta,\cdots,X_n/\theta)$ and $X_1/\theta,\cdots,X_n/\theta$ are independent uniform random variables on $[0, 1]$. Therefore the density of $X_{(n)}/\theta$ is $h(x) = nx^{n-1}$ for $0 \le x \le 1$ which does not depend on θ. Therefore, $X_{(n)}/\theta$ is a pivot for θ.
(b) For $0 \le b_1 < b_2 \le 1$,

$$P(b_1 \le X_{(n)}/\theta \le b_2) = b_2^n - b_1^n$$

and so if $b_2^n - b_1^n = p$ then $[X_{(n)}/b_2, X_{(n)}/b_1]$ is an exact $100p\%$ confidence interval for θ. Taking $b_2 = 1$, $a = 1/b_1$ satifies

$$1 - a^{-n} = p$$

and so $a = (1-p)^{-1/n}$.

3. (a) 95% confidence interval for θ: $9.888978 \pm 1.96 \times \sqrt{0.03750031}$.
(b) The approximate pivot here is

$$\frac{\widehat{\theta} - \theta}{\widehat{\text{se}}(\widehat{\theta})}$$

(b) Find the maximum likelihood estimator of θ based on $X_1,\cdots,X_n$. Find the limiting distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ where $\widehat{\theta}_n$ is the maximum likelihood estimator.
(c) Define a statistic Y to be the number of X_i's which are even $(0, 2, 4, \cdots)$; that is,

$$Y = \sum_{i=1}^n I(X_i \text{ is even}) = \sum_{i=1}^n\sum_{k=0}^\infty I(X_i = 2k)$$

(where $I(A)$ is 1 if A is true and 0 if A is false). Find an estimator $\widetilde{\theta}_n$ based on Y such that $\widetilde{\theta}_n \xrightarrow{p} \theta$.

13. Suppose that $Y_1,\cdots,Y_n$ are independent discrete random variable with the probability mass function of Y_i given by

$$f(y|\mu_i) = \frac{\mu_i^y}{(1+\mu_i)^{y+1}} \quad \text{for } y = 0, 1, 2, \cdots$$

where $\mu_i = E(Y_i)$. Suppose that $\ln(\mu_i) = \beta_0 + \beta_1 x_i$ where $x_1,\cdots,x_n$ are known constants.
(a) Show that $\text{Var}(Y_i) = \mu_i(1+\mu_i)$.
(b) Show that the maximum likelihood estimators, $\widehat{\beta}_0$ and $\widehat{\beta}_1$, of β_0 and β_1 satisfy the equations

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n (1+Y_i)\frac{\exp(\widehat{\beta}_0 + \widehat{\beta}_1 x_i)}{1+\exp(\widehat{\beta}_0 + \widehat{\beta}_1 x_i)}$$

and

$$\sum_{i=1}^n x_i Y_i = \sum_{i=1}^n x_i(1+Y_i)\frac{\exp(\widehat{\beta}_0 + \widehat{\beta}_1 x_i)}{1+\exp(\widehat{\beta}_0 + \widehat{\beta}_1 x_i)}.$$

14. Suppose that $X_1,\cdots,X_m$, $Y_1,\cdots,Y_n$ are two independent samples where $X_1,\cdots,X_m$ are independent with distribution function $F(x)$ and density $f(x)$, and $Y_1,\cdots,Y_n$ are independent with distribution function $G(x) = F(x)^\theta$ for some $\theta > 0$.
(a) Show that the common density function of $\{Y_i\}$ is

$$g(x) = \theta F(x)^{\theta-1}f(x)$$

where $\widehat{\theta}$ and $\widehat{\text{se}}(\widehat{\theta})$ is the jackknife standard error estimator.
(c) The 10% trimmed mean trims the largest 10% and smallest 10% of the observations with the middle 80% averaged. It is estimating a parameter $\theta(F)$ that averages the middle 80% of the distribution, that is, $F^{-1}(\tau)$ for $0.1 \le \tau \le 0.9$. Thus

$$\theta(F) = \frac{1}{0.8}\int_{0.1}^{0.9} F^{-1}(\tau)\,d\tau.$$

1. (a) The log-likelihood function is

$$\ln \mathcal{L}(\lambda) = n\ln(\lambda) - \lambda \sum_{i=1}^n x_i$$

and its first two derivatives are

$$\frac{d}{d\lambda}\ln\mathcal{L}(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i$$
$$\frac{d^2}{d\lambda^2}\ln\mathcal{L}(\lambda) = -\frac{n}{\lambda^2}$$

Thus the MLE is $\widehat{\lambda} = 1/\bar{X}$ and its estimated standard error is

$$\widehat{\mathrm{se}}(\widehat{\lambda}) = \left(-\frac{d^2}{d\lambda^2}\ln\mathcal{L}(\widehat{\lambda})\right)^{-1/2} = \frac{\widehat{\lambda}}{\sqrt{n}}.$$

(b) The sample median M_n converges in probability to the population median m, which satsifies

$$\int_0^m \lambda\exp(-\lambda x)\,dx = 1-\exp(-\lambda m) = \frac{1}{2}.$$

Solving for m, we get $m = \ln(2)/\lambda$ or $\lambda = \ln(2)/m$. Thus

$$\widetilde{\lambda}_n = \frac{\ln(2)}{M_n} \xrightarrow{p} \frac{\ln(2)}{m} = \lambda.$$

(c) From part (b), a $100p\%$ confidence interval for λ is

$$\left[\frac{\ln(2)}{T(X_1,\cdots,X_n)}, \frac{\ln(2)}{S(X_1,\cdots,X_n)}\right].$$

2. [12 marks] Suppose that $X_1,\cdots,X_n$ are independent continuous random variables with density function

$$f(x;\theta) = \frac{1}{\theta}g(x/\theta) \quad \text{for } x \ge 0$$

where $\theta > 0$ is an unknown parameter but the function g (which is itself a density function) is known.
(a) The density of X_i/θ is $\theta f(\theta x;\theta) = g(x)$ for $x \ge 0$. Since the distribution of X_i/θ is independent of θ so must the density of $(X_1+\cdots+X_n)/\theta$ and so this is an exact pivot for θ.
(b) By the CLT,

$$\frac{1}{\sigma_g\sqrt{n}}\left(\frac{X_1+\cdots+X_n}{\theta} - n\mu_g\right) \xrightarrow{d} \mathcal{N}(0,1)$$

If $P(-z \le \mathcal{N}(0,1) \le z) = p$ then

$$P\left\{-z \le \frac{1}{\sigma_g\sqrt{n}}\left(\frac{X_1+\cdots+X_n}{\theta} - n\mu_g\right) \le z\right\} \approx p$$

and so we obtain a $100p\%$ confidence interval for θ of the form

$$\left[\frac{X_1+\cdots+X_n}{n\mu_g + \sigma_g z\sqrt{n}}, \frac{X_1+\cdots+X_n}{n\mu_g - \sigma_g z\sqrt{n}}\right].$$

3. (a) Sample mean: $w_i = 1/n$ for all i. Sample median: If n is odd then $w_{(n+1)/2} = 1$ and $w_i = 0$ otherwise; if n is odd then $w_{n/2} = w_{n/2+1} = 1/21$ and $w_i = 0$ otherwise. Both estimators satisfy the two conditions.
(b) 95% CI for θ: $0.07854802 \pm 1.96 \times \sqrt{0.01301084}$.
(c) We can think of $\widehat{\theta}$ as

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n h(i/(n+1))\widehat{F}^{-1}(i/n) \xrightarrow{p} \int_0^1 h(\tau)F^{-1}(\tau)\,d\tau = \theta(F)$$

where

$$h(\tau) = k\left\{\frac{1}{2} - \left|\frac{1}{2}-\tau\right|\right\}$$

where

$$1 = \frac{1}{n}\sum_{i=1}^n h(i/(n+1)) \approx \int_0^1 h(\tau)\,d\tau = k/4.$$

Therefore, $k = 4$ and so

$$\theta(F) = 4\int_0^1 \left\{\frac{1}{2} - \left|\frac{1}{2}-\tau\right|\right\} F^{-1}(\tau)\,d\tau.$$

(To get full marks for part (c), you did not need to give the full solution given here.)
4. Note that the empirical distribution function is nearly flat for $2.5 \le x \le 3.5$ and so the corresponding density estimate should be relatively close to 0 for these values of x. Of the 4 density estimates, the only one that meets this criterion is (a).

1. [18 marks] Suppose that $X_1,\cdots,X_n$ are independent exponential random variables with density function

$$f(x;\lambda) = \lambda\exp(-\lambda x) \quad \text{for } x \ge 0$$

where $\lambda > 0$.
(a) Find the MLE of λ and give an estimate of its standard error based on the observed Fisher information.
(b) Suppose we are given only the sample median M_n of $X_1,\cdots,X_n$ and asked to estimate λ. (For simplicity, assume that M_n is equal to an order statistic $X_{(k)}$ where $k/n \approx 1/2$.) Define $\widetilde{\lambda}_n = \ln(2)/M_n$. Show that

$$\widetilde{\lambda}_n \xrightarrow{p} \lambda.$$

(c) In lecture, we showed how to construct a distribution-free confidence interval for the population median. If the interval $[S(X_1,\cdots,X_n), T(X_1,\cdots,X_n)]$ is a $100p\%$ confidence interval for the population median, find a $100p\%$ confidence interval for λ.

2. [12 marks] Suppose that $X_1,\cdots,X_n$ are independent continuous random variables with density function

$$f(x;\theta) = \frac{1}{\theta}g(x/\theta) \quad \text{for } x \ge 0$$

where $\theta > 0$ is an unknown parameter but the function g (which is itself a density function) is known.
(a) Show that $(X_1+\cdots+X_n)/\theta$ is an exact pivot for θ. (Hint: Find the density function of X_i/θ.)
(b) Even though the random variable in part (a) is a pivot, its distribution is not necessarily easy to compute. Define

$$\mu_g = \int_0^\infty xg(x)\,dx \quad\text{and}\quad \sigma_g^2 = \int_0^\infty (x-\mu_g)^2 g(x)\,dx.$$

Show that

$$\frac{1}{\sigma_g\sqrt{n}}\left(\frac{X_1+\cdots+X_n}{\theta} - n\mu_g\right)$$

is approximately normally distributed with mean 0 and variance 1 when n is sufficiently large and use this to construct an approximate $100p\%$ confidence interval for θ, giving all appropriate details.

3. [15 marks] Suppose that $X_1,\cdots,X_n$ are independent random variables with common density function $f(x-\theta)$ (where $f(x) = f(-x)$). θ can be estimated using the L-estimator

$$\widehat{\theta} = \sum_{i=1}^n w_i X_{(i)}$$

where $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ are the order statistics and the weights $w_1,\cdots,w_n$ satisfy the conditions $w_1+\cdots+w_n = 1$ and $w_k = w_{n-k+1}$ for $k = 1,\cdots,n$.
(a) Are the sample mean and sample median special cases of the estimator $\widehat{\theta}$ defined above? Justify your answer.
(b) Suppose we define $w_1,\cdots,w_n$ to be

$$w_i = k\left\{\frac{1}{2} - \left|\frac{1}{2} - \frac{i}{n+1}\right|\right\}$$

where the constant k is such that $w_1+\cdots+w_n = 1$. We are given an R function `lestimate` where `lestimate(x)` computes the estimate of θ for data `x`. The R code below computes a jackknife variance estimate of $\widehat{\theta}$ based on 100 observations. Assuming that the data come from a symmetric distribution, give an approximate 95% confidence interval for θ. (The 0.975 quantile of a standard normal distribution is 1.96.)

```
> thetahat <- lestimate(x)
> thetahat
[1] 0.07854802
> thetai <- NULL # leave-one-out estimates
> for (i in 1:100) {
+    thetai <- c(thetai,lestimate(x[-i]))
+    }
> thetadot <- mean(thetai)
> jackvar <- 99*sum((thetai-thetadot)^2)/100
> jackvar
[1] 0.01301084
```

(c) Suppose that $X_1,\cdots,X_n$ are independent random variables from some distribution F with quantiles $F^{-1}(\tau)$ for $0 < \tau < 1$. If we define $\widehat{\theta}_n$ to be the L-estimator in part (b) then

$$\widehat{\theta}_n \xrightarrow{p} \theta(F) = \int_0^1 g(\tau)F^{-1}(\tau)\,d\tau.$$

Find the form of the function g above. (Hint: Write w_i in part (b) as $w_i = h(i/(n+1))$ for some function h.)

1. (a) The likelihood and log-likelihood functions are

$$\mathcal{L}(\theta) = \prod_{i=1}^n\left\{\theta x_i^{\theta-1}\right\}$$
$$\ln\mathcal{L}(\theta) = \sum_{i=1}^n\{\ln\theta + (\theta-1)\ln(x_i)\}$$

Differentiating wrt θ, we get

$$\frac{d}{d\theta}\ln\mathcal{L}(\theta) = \frac{n}{\theta} + \sum_{i=1}^n\ln(x_i)$$

and solving $\frac{d}{d\theta}\ln\mathcal{L}(\widehat{\theta}) = 0$ gives the MLE

$$\widehat{\theta} = -\frac{n}{\sum_{i=1}^n \ln(X_i)}.$$

(Note that the second derivative of $\ln\mathcal{L}(\theta)$ is $-n/\theta^2 < 0$ so $\widehat{\theta}$ does indeed maximize the likelihood function.)

(b) Note that

$$\widetilde{\theta} = \frac{\frac{2}{n}\sum_{i=1}^n X_i^2}{1 - \frac{1}{n}\sum_{i=1}^n X_i^2}$$

where (by the WLLN),

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} E_\theta(X_1^2) = \int_0^1 x^2\theta x^{\theta-1}\,dx = \frac{\theta}{2+\theta}.$$

Thus

$$\widetilde{\theta} \xrightarrow{p} \frac{2\theta/(2+\theta)}{1-\theta/(2+\theta)} = \theta.$$

(c) By the CLT,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{\theta}{2+\theta}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{4\theta}{(\theta+4)(\theta+2)^2}\right)$$

Applying the Delta Method with $g(x) = 2x/(1-x)$ and $g'(x) = 1/(1-x)^2$, we get

$$\sqrt{n}(\widetilde{\theta}-\theta) = \sqrt{n}\left(g\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - g\left(\frac{\theta}{2+\theta}\right)\right)$$
$$\xrightarrow{d} \mathcal{N}\left(0, \left[g'\left(\frac{\theta}{2+\theta}\right)\right]^2 \frac{4\theta}{(\theta+4)(\theta+2)^2}\right)$$

and

$$\left[g'\left(\frac{\theta}{2+\theta}\right)\right]^2 \frac{4\theta}{(\theta+3)(\theta+2)^2} = \frac{(\theta+2)^4}{4}\frac{4\theta}{(\theta+4)(\theta+2)^2} = \frac{\theta(\theta+2)^2}{\theta+4}.$$

Substituting $\widetilde{\theta}$ for θ, our estimate of the standard error is

$$\widehat{\mathrm{se}}(\widetilde{\theta}) = \frac{1}{\sqrt{n}}\left\{\frac{\widetilde{\theta}(\widetilde{\theta}+2)^2}{\widetilde{\theta}+4}\right\}^{1/2}.$$

Alternatively, we could estimate $\mathrm{Var}_\theta(X_i^2)$ by

$$\frac{1}{n-1}\sum_{i=1}^n\left\{X_i^2 - \frac{1}{n}\sum_{j=1}^n X_j^2\right\}$$

and apply the Delta Method as above to obtain a standard error estimate.
2. (a) We need to show that the distribution of $X_{(n)}/\theta$ is independent of θ:

$$P_\theta(X_{(n)}/\theta \le y) = P_\theta(X_{(n)} \le \theta y) = \frac{(\theta y)^{2n}}{\theta^{2n}} = y^{2n}$$

for $0 \le y \le 1$.

(b) For $0 \le \gamma_1 < \gamma_2 \le 1$, we have

$$P_\theta\left[\gamma_1 \le \frac{X_{(n)}}{\theta} \le \gamma_2\right] = \gamma_2^{2n} - \gamma_1^{2n}$$

which gives a confidence interval with coverage $\gamma_2^{2n} - \gamma_1^{2n}$ of the form $[X_{(n)}/\gamma_2, X_{(n)}/\gamma_1]$. Here $\gamma_2 = 1$ and so γ_1 must satisfy $1-\gamma_1^{2n} = p$ so that $\gamma_1 = (1-p)^{1/(2n)}$. Thus $a = 1/\gamma_1 = (1-p)^{-1/(2n)}$.

3. (a) From the R output, $\widehat{\theta} = 9.888978$ and $\widehat{\mathrm{se}}(\widehat{\theta}) = \sqrt{0.04001} \approx 0.20$. Thus a 95% confidence interval is $9.89 \pm 1.96 \times 0.20$.

(b) The approximate pivot is $(\widehat{\theta}-\theta)/\widehat{\mathrm{se}}(\widehat{\theta})$ which is assumed to have approximately a $\mathcal{N}(0,1)$ distribution.

4. (a) We need to show that (i) $\widehat{f}_h(x) \ge 0$ for all x, and (ii) $\int_{-\infty}^\infty \widehat{f}_h(x)\,dx = 1$. (i) is trivial since $\widehat{f}_h(x)$ is a sum of non-negative terms. To show (ii), note that

$$\int_{-\infty}^\infty \widehat{f}_h(x)\,dx = \int_{-\infty}^\infty \frac{1}{2nh}\sum_{i=1}^n I(X_i - h \le x \le X_i + h)\,dx$$
$$= \frac{1}{2nh}\sum_{i=1}^n \underbrace{\int_{-\infty}^\infty I(X_i - h \le x \le X_i + h)\,dx}_{=2h}$$
$$= \frac{2nh}{2nh}$$
$$= 1.$$

(b) Note that $\sum_{i=1}^n I(x-h \le X_i \le x+h)$ has a Binomial distribution with parameters n and

$$p = P(x-h \le X_i \le x+h) = \int_{x-h}^{x+h} 2t\,dt = 4xh$$

(assuming that $h \le x \le 1-h$). Thus

$$E[\widehat{f}_h(x)] = \frac{1}{2nh}\times n \times 4xh = 2x$$

and

$$\mathrm{Var}[\widehat{f}_h(x)] = \frac{1}{(2nh)^2}\times n \times 4xh(1-4xh) = \frac{x(1-4xh)}{nh}.$$

Note that, in this special case, $\widehat{f}_h(x)$ is an unbiased estimate of $f(x) = 2x$ for $h \le x \le 1-h$; however, it is biased for $x < h$ and $x > 1-h$ (edge effects) and there is a tradeoff between choosing h large (which makes the variance small but increases the edge effects) and chooseing h small (which minimizes the edge effects but otherwise increases the variance.

1. [15 marks] Suppose that $X_1,\cdots,X_n$ are independent continuous random variables with density

$$f(x;\theta) = \theta x^{\theta-1} \quad \text{for } 0 \le x \le 1$$

where $\theta > 0$.

(a) Find the MLE of θ based on $X_1,\cdots,X_n$. (You do **not** need to show that your estimator maximizes the likelihood function.)

(b) Another estimator of θ is

$$\widetilde{\theta} = \frac{2\sum_{i=1}^n X_i^2}{n - \sum_{i=1}^n X_i^2}$$

Show that as the sample size $n\to\infty$, $\widetilde{\theta} = \widetilde{\theta}_n \xrightarrow{p} \theta$. (Hint: Compute $E_\theta(X_i^2)$ and apply the Weak Law of Large Numbers.)

(c) Use the Delta Method to derive an estimator of the standard error in part (b). You may use the fact that

$$\mathrm{Var}_\theta(X_i^2) = E_\theta(X_i^4) - [E_\theta(X_i^2)]^2 = \frac{4\theta}{(\theta+4)(\theta+2)^2}$$

in your derivation.

2. [10 marks] Suppose that $X_1,\cdots,X_n$ are independent uniform random variables on the interval $[0,\theta]$; that is, the density is

$$f(x;\theta) = \begin{cases} 2x/\theta^2 & \text{if } 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

(a) The distribution function of $X_{(n)} = \max(X_1,\cdots,X_n)$ is

$$G(x;\theta) = P_\theta(X_{(n)} \le x) = \frac{x^{2n}}{\theta^{2n}} \quad \text{for } 0 \le x \le \theta.$$

(You do not need to show this.) Show that $X_{(n)}/\theta$ is a pivot for θ.

(b) Use the pivot $X_{(n)}/\theta$ to find an exact $100p\%$ confidence interval of the form $[X_{(n)}, aX_{(n)}]$ where $a > 1$ will depend on p and n.

3. [10 marks] The following R code computes the 10% trimmed mean (that is, the largest 10% and smallest 10% of the observations are removed and the middle 80% averaged) of a sample of size 100 and computes a jackknife variance estimate. (Assume that the vector `x` contains 100 observations.)

```
> m10 <- mean(x,trim=0.1)
> m10
[1] 9.888978
> mi <- NULL # leave-one-out estimates
> for (i in 1:100) {
+    mi <- c(mi,mean(x[-i],trim=0.1))
+    }
> mdot <- mean(mi)
> jackvar <- 99*sum((mi-mdot)^2)/100
> jackvar
[1] 0.04001
```

(a) Assuming that the data $x_1,\cdots,x_{100}$ come from a density $f(x-\theta)$ where $f(x)$ is symmetric around 0, give an approximate 95% confidence interval for θ using the output given above. (The 0.975 quantile of a standard normal distribution is 1.96.)

(b) For the confidence interval in part (a), what is the approximate pivot that is being used?

4. [15 marks] Suppose that $X_1,\cdots,X_n$ are independent continuous random variables with unknown density function f and define the kernel density estimator

$$\widehat{f}_h(x) = \frac{1}{2nh}\sum_{i=1}^n I(X_i - h \le x \le X_i + h) = \frac{1}{2nh}\sum_{i=1}^n I(x-h \le X_i \le x+h)$$

where $h > 0$ is the bandwidth. (Recall that $I(A) = 1$ if the condition A is true and $I(A) = 0$ if A is false.)

(a) Show that $\widehat{f}_h$ is a density function for any $h > 0$.

(b) Suppose that the true density of X_i is

$$f(x) = \begin{cases} 2x & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

If x lies between h and $1-h$, find the mean and variance of $\widehat{f}_h(x)$. You may use (without proof) the fact that the distribution function is $F(x) = P(X_i \le x) = x^2$ for $0 \le x \le 1$.

1. (a) $\widehat{\mathrm{se}}(\widehat{\theta}) = \sqrt{0.03750031} = 0.194$.
(b) 95% CI: $9.889 \pm 1.96 \times 0.194 = [9.509, 10.269]$.
(c) We are assuming that the distribution of $(\widehat{\theta}-\theta)/\widehat{\mathrm{se}}(\widehat{\theta})$ is approximately $\mathcal{N}(0,1)$, and therefore an approximate pivot for θ.

2. (a) The likelihood function is

$$L(\lambda) = \prod_{i=1}^n\left\{\lambda^2 x_i\exp(-\lambda x_i)\right\} = \lambda^{2n}\exp\left(-\lambda\sum_{i=1}^n x_i\right)\prod_{i=1}^n x_i$$

Taking logs and differentiating, we get

$$\frac{d}{d\lambda}\ln L(\lambda) = \frac{2n}{\lambda} - \sum_{i=1}^n x_i.$$

Setting this to 0 and solving, we get that the MLE is

$$\widehat{\lambda} = \left(\frac{1}{2n}\sum_{i=1}^n X_i\right)^{-1}$$

(Also note that

$$\frac{d^2}{d\lambda^2}\ln L(\lambda) = -\frac{2n}{\lambda^2} < 0$$

so that $\widehat{\lambda}$ does maximize the likelihood function.)
(b) From above, the observed Fisher information is

$$-\frac{d^2}{d\lambda^2}\ln L(\widehat{\lambda}) = \frac{2n}{\widehat{\lambda}^2}$$

and so $\widehat{\mathrm{se}}(\widehat{\lambda}) = \widehat{\lambda}/\sqrt{2n}$.

3. (a) Note that $P_\theta(X_i - \theta \le x) = P_\theta(X_i \le x+\theta)$ and so the density of $X_i - \theta$ is

$$\frac{d}{dx}P_\theta(X_i \le x+\theta) = f(x+\theta;\theta) = g(x),$$

which is independent of θ. Therefore the distribution of $\sum_{i=1}^n(X_i-\theta)$ is also independent of θ and is an exact pivot.
(b) We use the CLT to approximate the distribution of $\sum_{i=1}^n(X_i-\theta)$; by symmetry, its mean is 0 and its variance is $n\sigma^2$ where

$$\sigma^2 = \int_{-1}^1 \frac{3}{4}x^2(1-x^2)\,dx = \frac{1}{5}.$$

Therefore,

$$\sqrt{5/n}\sum_{i=1}^n(X_i-\theta) = \sqrt{5n}(\bar{X}-\theta) \overset{d}{\approx} \mathcal{N}(0,1).$$

Thus if $P(-z_p \le \mathcal{N}(0,1) \le z_p) = p$, a $100p\%$ CI has the form $\bar{X} \pm z_p/\sqrt{5n}$.

4. There are a number of possible answers to this question. For example, using (i) and the method of moments, we can define

$$\sum_{i=1}^k X_{(i)} = \frac{1}{\lambda}\sum_{i=1}^k a_i \Longrightarrow \widehat{\lambda} = \frac{\sum_{i=1}^k a_i}{\sum_{i=1}^k X_{(i)}} \tag{1}$$

or

$$\frac{1}{k}\sum_{i=1}^k X_{(i)}/a_i = \frac{1}{\lambda} \Longrightarrow \widehat{\lambda} = \left(\frac{1}{k}\sum_{i=1}^k X_{(i)}/a_i\right)^{-1} \tag{2}$$

or (thinking of $1/\lambda$ as a slope and estimating it via least squares)

$$\frac{\sum_{i=1}^k X_{(i)}a_i}{\sum_{i=1}^k a_i^2} = \frac{1}{\lambda} \Longrightarrow \widehat{\lambda} = \frac{\sum_{i=1}^k a_i^2}{\sum_{i=1}^k X_{(i)}a_i}. \tag{3}$$

Using (ii), we can use maximum likelihood estimation since the spacings are independent Exponential random variables. The MLE of λ in this case is

$$\widehat{\lambda} = \left\{\frac{1}{k}\left(nX_{(1)} + (n-1)(X_{(2)}-X_{(1)}) + \cdots + (n-k+1)(X_{(k)}-X_{(k-1)})\right)\right\}^{-1}. \tag{4}$$

The estimators defined in (2) and (4) are actually the same.

1. [15 marks] The following R code computes the 10% trimmed mean (that is, the largest 10% and smallest 10% of the observations are removed and the middle 80% averaged) of a sample of size 100 and computes a jackknife variance estimate. (Assume that the vector `x` contains 100 observations.)

```
> m10 <- mean(x,trim=0.1)
> m10
[1] 9.888978
> mi <- NULL # leave-one-out estimates
> for (i in 1:100) {
+    mi <- c(mi,mean(x[-i],trim=0.1))
+    }
> mdot <- mean(mi)
> jackvar <- 99*sum((mi-mdot)^2)/100
> jackvar
[1] 0.03750031
```

(a) What is the jackknife estimate of the standard error of the trimmed mean?
(b) Assuming that the data $x_1,\cdots,x_{100}$ come from a density $f(x-\theta)$ where $f(x)$ is symmetric around 0, give an approximate 95% confidence interval for θ. (The 0.975 quantile of a standard normal distribution is 1.96.)
(c) What assumptions are you using in computing this confidence interval in part (b)? (Hint: What is the pivot you are using to compute this confidence interval?)

2. [10 marks] Suppose that $X_1,\cdots,X_n$ are independent random variables with density function

$$f(x;\lambda) = \lambda^2 x\exp(-\lambda x) \quad \text{for } x \ge 0$$

where $\lambda > 0$.
(a) Show that the maximum likelihood estimator (MLE) of λ is

$$\widehat{\lambda} = \left(\frac{1}{2n}\sum_{i=1}^n X_i\right)^{-1}$$

(b) Give an estimate of the standard error of the MLE $\widehat{\lambda}$ in part (a) based on the observed Fisher information.
3. [12 marks] Suppose that $X_1,\cdots,X_n$ are independent continuous random variables with density function

$$f(x;\theta) = g(x-\theta) \quad \text{for } -\infty < x < \infty$$

where θ is an unknown parameter but the function g (which is itself a density function) is known.
(a) Show that

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

is an exact pivot for θ. (Hint: Find the density function of $X_i - \theta$.)
(b) Even though the random variable in part (a) is an exact pivot, its distribution is not necessarily easy to compute. However, we can appeal to the Central Limit Theorem to approximate its distribution. Suppose that

$$g(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{for } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Assuming that n is large enough so that the pivot in part (a) is approximately normal, construct an approximate $100p\%$ confidence interval for θ, giving all appropriate details.

4. [8 marks] Suppose that $X_1,\cdots,X_n$ are independent Exponential random variables with density

$$f(x;\lambda) = \lambda\exp(-\lambda x) \quad \text{for } x \ge 0$$

where $\lambda > 0$. However, rather than observing all n observations, suppose that you are given only the smallest k values of $X_1,\cdots,X_n$, that is, the order statistics $X_{(1)} \le X_{(2)} \le \cdots \le X_{(k)}$. Using **one** of the following 2 facts [(i) or (ii)], find an estimator of λ based only on $X_{(1)}, X_{(2)},\cdots,X$

(i) For $i = 1,\cdots,k$,

$$E[X_{(i)}] = \frac{1}{\lambda}\sum_{j=1}^i \frac{1}{n-j+1} = \frac{1}{\lambda}a_i;$$

note that $a_1,\cdots,a_k$ are known constants.

(ii) The random variables $nX_{(1)}, (n-1)(X_{(2)}-X_{(1)}),\cdots,(n-k+1)(X_{(k)}-X_{(k-1)})$ are independent Exponential random variables with parameter λ.

1. (a) Note that

$$F(x) = \int_1^x \alpha t^{-\alpha-1}\,dt = 1 - x^{-\alpha}$$

for $x \ge 1$. Then

$$P(V_n \le x) = F(x)^n = (1-x^{-\alpha})^n$$

and the density of V_n is given by

$$\frac{d}{dx}P(V_n \le x) = n\alpha(1-x^{-\alpha})^{n-1}x^{-\alpha-1}$$

for $x \ge 1$.
(b) The distribution function of $n^{-1/\alpha}V_n$ is given by

$$P(n^{-1/\alpha}V_n \le x) = (1-(n^{1/\alpha}x)^{-\alpha})^n$$
$$= \left(1 - \frac{x^{-\alpha}}{n}\right)^n$$
$$\to \exp(-x^{-\alpha})$$

as $n\to\infty$.
2. (a) By the WLLN and CLT,

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} 2$$
$$\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \xrightarrow{d} \mathcal{N}(0,2).$$

Therefore $U_n \xrightarrow{d} \sqrt{2}\times\mathcal{N}(0,2) = \mathcal{N}(0,4)$.
(b) $\mathrm{Var}(V_n) = 2/n \to 0$ as $n\to\infty$ and $E(V_n) = 0$. Therefore $V_n \xrightarrow{p} 0$ (using Chebyshev's inequality).

3. The log-likelihood function is

$$\ln L(\theta) = \sum_{i=1}^n |x_i|\{\ln(\theta/4) - \ln(1-\theta/2)\} + n\ln(1-\theta/2)$$

and its derivative is

$$\frac{d}{d\theta}\ln L(\theta) = \frac{1}{\theta}\sum_{i=1}^n|x_i| + \frac{1}{2-\theta}\sum_{i=1}^n|x_i| - \frac{n}{2-\theta}$$

Setting this derivative to 0 and solving, we get

$$\widehat{\theta} = \frac{2}{n}\sum_{i=1}^n|X_i|.$$

In addition,

$$-\frac{d^2}{d\theta^2}\ln f(x;\theta) = \frac{|x|}{\theta^2} - \frac{|x|}{(2-\theta)^2} + \frac{1}{(2-\theta)^2}$$

and using $E(X_i) = \theta/2$, we have

$$\sqrt{n}(\widehat{\theta}-\theta) \xrightarrow{d} \mathcal{N}(0, 1/I(\theta))$$

where

$$I(\theta) = E\left[\frac{|X_i|}{\theta^2} - \frac{|X_i|}{(2-\theta)^2} + \frac{1}{(2-\theta)^2}\right] = \frac{1}{\theta(2-\theta)}.$$

4. (a) The joint probability mass function of $(X_1,\cdots,X_n)$ is

$$f(x_1,\cdots,x_n;\theta) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$$

and so $T = X_1+\cdots+X_n$ is sufficient for θ by the Factorization Criterion.
(b) For $n = 2$, we need to find $g(t)$ so that $E_\theta[g(T)] = \theta(1-\theta)$. However, for any g

$$E_\theta[g(T)] = g(0)(1-\theta)^2 + 2g(1)\theta(1-\theta) + g(2)\theta^2$$

and by inspection if we take $g(1) = 1/2$ and $g(0) = g(2) = 0$ then $E_\theta[g(T)] = \theta(1-\theta)$.

5. $\mathrm{Var}(\widehat{\theta}) = h(a)$ where

$$h(a) = a^2\mathrm{Var}(X) + (1-a)^2\mathrm{Var}(Y)$$
$$= a^2\sigma_1^2 + (1-a)^2\sigma_2^2$$

$h'(a) = 2a\sigma_1^2 - 2(1-a)\sigma_2^2$ and $h'(a) = 0$ when $a = \sigma_2^2/(\sigma_1^2+\sigma_2^2)$.
6. (a) $E(V_n) = E(X_{\mathbf{2}})/(k+1) = (\mathrm{Var}(X_i)+E(X_i)_2)/(k+1) = \theta_2$. By the WLLN, $V_n \xrightarrow{p} \theta_2$.
Thus V_n is a consistent and unbiased estimator of θ^2.
(b) $\bar{X}_n \xrightarrow{p} \theta$ by the WLLN and therefore $\bar{X}_n^2 \xrightarrow{p} \theta^2$. By the CLT and Delta Method,

$$\sqrt{n}(\bar{X}_n^2 - \theta^2) \xrightarrow{d} \mathcal{N}(0, (2\theta)^2k\theta^2) = \mathcal{N}(0, 4k\theta^4).$$

7. (a) X_i/t_i has an exponential distribution with mean β and so

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^n X_i/t_i$$

is an unbiased estimator of β.
(b) By the CLT,

$$\sqrt{n}(\widehat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0,\beta^2)$$

and so $\sqrt{n}(\widehat{\beta}-\beta)/\beta$ is an approximate pivot for β (with an approximate $\mathcal{N}(0,1)$ distribution). Thus an approximate 95% confidence interval for β is

$$\left[\frac{\widehat{\beta}}{1+1.96/\sqrt{n}}, \frac{\widehat{\beta}}{1-1.96/\sqrt{n}}\right].$$

8. (a) Start by defining $Y_i = -\ln(X_i)$ (which has an exponential distribution with mean 1). Then

$$\frac{1}{\widehat{\theta}_n} + \frac{1}{\widehat{\theta}_n+1} = \bar{Y}_n$$

Then (solving a quadratic equation and taking the positive root), we have

$$\widehat{\theta}_n = \frac{2-\bar{Y}_n+\sqrt{4+\bar{Y}_n^2}}{2\bar{Y}_n}$$

and so (since $\bar{Y}_n \xrightarrow{p} 1$), it follows that

$$\widehat{\theta}_n \xrightarrow{p} \frac{1+\sqrt{5}}{2} = \theta_0$$

(b) By the CLT, $\sqrt{n}(\bar{Y}_n - 1) \xrightarrow{d} \mathcal{N}(0,1)$. $\widehat{\theta}_n = g(\bar{Y}_n)$ for g given in part (a). Thus by the delta method, $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, [g'(1)]^2)$ with $[g'(1)]^2 = (5+2\sqrt{5})^2/25$.

9. (a) (This part is somewhat beyond the scope of the course but ...) First of all, $P(Y = 0) = P(X_1 = X_2 = \cdots = X_n = 0) = p^n$. In general, we can write

$$Y = \sum_{i=1}^n \Delta_i E_i$$

where $\{\Delta_i\}$ and $\{E_i\}$ are independent random variables where $P(\Delta_i = 0) = p$, $P(\Delta_i = 1) = 1-p$ and $\{E_i\}$ are exponential with mean $1/\lambda$. Define $S = \Delta_1+\cdots+\Delta_n$, then given $S = k$ for $k \ge 1$, Y has a Gamma distribution with shape parameter k. For $y > 0$, we have

$$P(Y > y) = \sum_{k=1}^n P(S=k)P(Y > y|S=k)$$
$$= \sum_{k=1}^n \binom{n}{k}(1-p)^k p^{n-k}P(\mathrm{Gamma}(k,\lambda) > y)$$

For larger values of n, we can use the CLT to approximate the distribution of Y (in terms of the mean and variance of X_i).
(b) If $p = \exp(-\lambda)$ then $E(X_i) = (1-\exp(-\lambda))/\lambda$. Thus $\bar{X}$ is a consistent estimator of $(1-\exp(-\lambda))/\lambda$ and so we can estimate λ by

$$\bar{X} = (1-\exp(-\widehat{\lambda}))/\widehat{\lambda}.$$

(c) $\bar{X} = g(\widehat{\lambda})$ where $g(t) = (1-\exp(-t))/t$ and so $\widehat{\lambda} = g^{-1}(\bar{X})$ where the derivative of $g^{-1}(x)$ is $1/g'(g^{-1}(x))$ with $g'(t) = \{(1+t)\exp(-t)-1\}/t^2$. If S^2 is the sample variance of $X_1,\cdots,X_n$ then an estimator of the standard error of $\widehat{\lambda}$ is

$$\widehat{\mathrm{se}}(\widehat{\lambda}) = \frac{S}{\sqrt{n}g'(\widehat{\lambda})}$$

(d) The log-likelihood function is

$$\ln\mathcal{L}(\lambda) = -\lambda\sum_{i=1}^n I(x_i = 0) + \sum_{i:x_i>0}\{\ln(\lambda) + \ln(1-\exp(-\lambda) - \lambda x_i\}$$

Taking the partial derivative with respect to λ, we get the MLE satisfying the equation

$$\sum_{i=1}^n I(X_i = 0) + \sum_{i:X_i>0} X_i = \sum_{i:X_i>0}\left\{\frac{1}{\widehat{\lambda}} + \frac{\exp(-\widehat{\lambda})}{1-\exp(-\widehat{\lambda})}\right\}.$$

There is not a closed-form expression for the MLE although we can compute it numerically.
(e) The observed Fisher information is

$$\widehat{\mathcal{I}(\theta)} = \sum_{i:X_i>0}\left\{\frac{1}{\widehat{\lambda}^2} + \frac{\exp(-\widehat{\lambda})}{1-\exp(-\widehat{\lambda})^2}\right\}$$

An estimate of the standard error is $\widehat{\mathcal{I}(\theta)}^{-1/2}$.
10. Suppose that $X_1,\cdots,X_n$ are independent uniform random variables on $[0,\theta]$ and we test

$$H_0: \theta = 1 \quad \text{versus} \quad H_1: \theta \ne 1$$

at level α. We will reject H_0 if $T > 1$ or $T < k$ where $T = \max(X_1,\cdots,X_n)$.
(a) When $\theta = 1$, $P_1(T > 1) = 0$ and so k satsifies

$$P_1(T < k) = k^n = \alpha$$

and so $k = \alpha^{1/n}$.
(b) $\mathrm{power}(\theta) = P_\theta(T < \alpha^{1/n}) + P_\theta(T > 1)$. First of all, if $\theta < 1$ then $P_\theta(T > 1) = 0$ and so

$$\mathrm{power}(\theta) = \min(\alpha/\theta^n, 1)$$

while for $\theta > 1$,

$$\mathrm{power}(\theta) = \alpha/\theta^n + 1 - 1/\theta^n$$

Note that $\mathrm{power}(\theta)$ increases as θ decreases from 1. For $\theta > 1$, note that

$$\mathrm{power}'(\theta) = n(1-\alpha)/\theta^{n+1} > 0$$

(b) Consider testing the null hypothesis

$$H_0: \theta = 1$$

versus the alternative

$$H_1: \theta > 1.$$

Suppose that $F(x)$ is known (which may be assumed true). Using the Neyman-Pearson paradigm, find a "good" test of H_0 versus H_1. (Hint: Look at the ratio of joint densities under H_0 and H_1.)