Chapter 10

Asymptotic Evaluations

10.1 First calculate some moments for this distribution.

$$\mathrm{E}X = \theta/3, \quad \mathrm{E}\,X^2 = 1/3, \quad \mathrm{Var}X = \frac{1}{3} - \frac{\theta^2}{9}.$$

So $3\bar{X}_n$ is an unbiased estimator of θ with variance

$$\mathrm{Var}(3\bar{X}_n) = 9(\mathrm{Var}X)/n = (3-\theta^2)/n \to 0 \text{ as } n \to \infty.$$

So by Theorem 10.1.3, $3\bar{X}_n$ is a consistent estimator of θ.

10.3 a. The log likelihood is

$$-\frac{n}{2}\log(2\pi\theta) - \frac{1}{2}\sum(x_i-\theta)/\theta.$$

Differentiate and set equal to zero, and a little algebra will show that the MLE is the root of $\theta^2+\theta-W=0$. The roots of this equation are $(-1\pm\sqrt{1+4W})/2$, and the MLE is the root with the plus sign, as it has to be nonnegative.

b. The second derivative of the log likelihood is $(-2\sum x_i^2+n\theta)/(2\theta^3)$, yielding an expected Fisher information of

$$I(\theta) = -\mathrm{E}_\theta \frac{-2\sum X_i^2+n\theta}{2\theta^3} = \frac{2n\theta+n}{2\theta^2},$$

and by Theorem 10.1.12 the variance of the MLE is $1/I(\theta)$.

10.4 a. Write

$$\frac{\sum X_iY_i}{\sum X_i^2} = \frac{\sum X_i(X_i+\epsilon_i)}{\sum X_i^2} = 1 + \frac{\sum X_i\epsilon_i}{\sum X_i^2}.$$

From normality and independence

$$\mathrm{E}X_i\epsilon_i = 0, \quad \mathrm{Var}X_i\epsilon_i = \sigma^2(\mu^2+\tau^2), \quad \mathrm{E}X_i^2 = \mu^2+\tau^2, \quad \mathrm{Var}X_i^2 = 2\tau^2(2\mu^2+\tau^2),$$

and $\mathrm{Cov}(X_i, X_i\epsilon_i) = 0$. Applying the formulas of Example 5.5.27, the asymptotic mean and variance are

$$\mathrm{E}\left(\frac{\sum X_iY_i}{\sum X_i^2}\right) \approx 1 \text{ and } \mathrm{Var}\left(\frac{\sum X_iY_i}{\sum X_i^2}\right) \approx \frac{n\sigma^2(\mu^2+\tau^2)}{[n(\mu^2+\tau^2)]^2} = \frac{\sigma^2}{n(\mu^2+\tau^2)}$$

b.

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

with approximate mean β and variance $\sigma^2/(n\mu^2)$.

c.
$$\frac{1}{n}\sum \frac{Y_i}{X_i} = \beta + \frac{1}{n}\sum \frac{\epsilon_i}{X_i}$$
with approximate mean β and variance $\sigma^2/(n\mu^2)$.

10.5 a. The integral of $\mathrm{E}T_n^2$ is unbounded near zero. We have
$$\mathrm{E}T_n^2 > \sqrt{\frac{n}{2\pi\sigma^2}}\int_0^1 \frac{1}{x^2}e^{-(x-\mu)^2/2\sigma^2}dx > \sqrt{\frac{n}{2\pi\sigma^2}}K\int_0^1 \frac{1}{x^2}dx = \infty,$$
where $K = \max_{0\le x\le 1} e^{-(x-\mu)^2/2\sigma^2}$

b. If we delete the interval $(-\delta,\delta)$, then the integrand is bounded, that is, over the range of integration $1/x^2 < 1/\delta^2$.

c. Assume $\mu > 0$. A similar argument works for $\mu < 0$. Then
$$P(-\delta < X < \delta) = P[\sqrt{n}(-\delta-\mu) < \sqrt{n}(X-\mu) < \sqrt{n}(\delta-\mu)] < P[Z < \sqrt{n}(\delta-\mu)],$$
where $Z \sim \mathrm{n}(0,1)$. For $\delta < \mu$, the probability goes to 0 as $n \to \infty$.

10.7 We need to assume that $\tau(\theta)$ is differentiable at $\theta = \theta_0$, the true value of the parameter. Then we apply Theorem 5.5.24 to Theorem 10.1.12.

10.9 We will do a more general problem that includes *a*) and *b*) as special cases. Suppose we want to estimate $\lambda^t e^{-\lambda}/t! = P(X = t)$. Let
$$T = T(X_1,\dots,X_n) = \begin{cases} 1 & \text{if } X_1 = t \\ 0 & \text{if } X_1 \ne t. \end{cases}$$
Then $\mathrm{E}T = P(T=1) = P(X_1 = t)$, so T is an unbiased estimator. Since $\sum X_i$ is a complete sufficient statistic for λ, $\mathrm{E}(T|\sum X_i)$ is UMVUE. The UMVUE is 0 for $y = \sum X_i < t$, and for $y \ge t$,
$$\begin{aligned} \mathrm{E}(T|y) &= P(X_1 = t|\textstyle\sum X_i = y) \\ &= \frac{P(X_1 = t, \sum X_i = y)}{P(\sum X_i = y)} \\ &= \frac{P(X_1 = t)P(\sum_{i=2}^n X_i = y-t)}{P(\sum X_i = y)} \\ &= \frac{\{\lambda^t e^{-\lambda}/t!\}\{[(n-1)\lambda]^{y-t}e^{-(n-1)\lambda}/(y-t)!\}}{(n\lambda)^y e^{-n\lambda}/y!} \\ &= \binom{y}{t}\frac{(n-1)^{y-t}}{n^y}. \end{aligned}$$

a. The best unbiased estimator of $e^{-\lambda}$ is $((n-1)/n)^y$.

b. The best unbiased estimator of $\lambda e^{-\lambda}$ is $(y/n)[(n-1)/n]^{y-1}$

c. Use the fact that for constants a and b,
$$\frac{d}{d\lambda}\lambda^a b^\lambda = b^\lambda \lambda^{a-1}(a + \lambda\log b),$$
to calculate the asymptotic variances of the UMVUEs. We have for $t = 0$,
$$\mathrm{ARE}\left(\left(\frac{n-1}{n}\right)^{n\hat\lambda}, e^{-\lambda}\right) = \left[\frac{e^{-\lambda}}{\left(\frac{n-1}{n}\right)^{n\lambda}\log\left(\frac{n-1}{n}\right)^n}\right]^2,$$

and for $t = 1$

$$\text{ARE}\left(\frac{n}{n-1}\hat{\lambda}\left(\frac{n-1}{n}\right)^{n\hat{\lambda}}, \hat{\lambda}e^{-\lambda}\right) = \left[\frac{(\lambda-1)e^{-\lambda}}{\frac{n}{n-1}\left(\frac{n-1}{n}\right)^{n\lambda}\left[1+\log\left(\frac{n-1}{n}\right)^n\right]}\right]^2.$$

Since $[(n-1)/n]^n \to e^{-1}$ as $n \to \infty$, both of these AREs are equal to 1 in the limit.

d. For these data, $n = 15$, $\sum X_i = y = 104$ and the MLE of λ is $\hat{\lambda} = \bar{X} = 6.9333$. The estimates are

	MLE	UMVUE
$P(X=0)$	.000975	.000765
$P(X=1)$	.006758	.005684

10.11 a. It is easiest to use the Mathematica code in Example A.0.7. The second derivative of the log likelihood is

$$\frac{\partial^2}{\partial\mu^2}\log\left(\frac{1}{\Gamma[\mu/\beta]\beta^{\mu/\beta}}x^{-1+\mu/\beta}e^{-x/\beta}\right) = \frac{1}{\beta^2}\psi'(\mu/\beta),$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function.

b. Estimation of β does not affect the calculation.

c. For $\mu = \alpha\beta$ known, the MOM estimate of β is $\bar{x}/\alpha$. The MLE comes from differentiating the log likelihood

$$\frac{d}{d\beta}\left(-\alpha n\log\beta - \sum_i x_i/\beta\right) \stackrel{\text{set}}{=} 0 \Rightarrow \beta = \bar{x}/\alpha.$$

d. The MOM estimate of β comes from solving

$$\frac{1}{n}\sum_i x_i = \mu \text{ and } \frac{1}{n}\sum_i x_i^2 = \mu^2 + \mu\beta,$$

which yields $\tilde{\beta} = \hat{\sigma}^2/\bar{x}$. The approximate variance is quite a pain to calculate. Start from

$$\text{E}\bar{X} = \mu, \quad \text{Var}\bar{X} = \frac{1}{n}\mu\beta, \quad \text{E}\hat{\sigma}^2 \approx \mu\beta, \quad \text{Var}\hat{\sigma}^2 \approx \frac{2}{n}\mu\beta^3,$$

where we used Exercise 5.8(b) for the variance of $\hat{\sigma}^2$. Now using Example 5.5.27 and (and assuming the covariance is zero), we have $\text{Var}\tilde{\beta} \approx \frac{3\beta^3}{n\mu}$. The ARE is then

$$\text{ARE}(\hat{\beta}, \tilde{\beta}) = \left[3\beta^3/\mu\right]\left[\text{E}\left(-\frac{d^2}{d\beta^2}l(\mu,\beta|\mathbf{X})\right)\right].$$

Here is a small table of AREs. There are some entries that are less than one - this is due to using an approximation for the MOM variance.

	μ			
β	1	3	6	10
1	1.878	0.547	0.262	0.154
2	4.238	1.179	0.547	0.317
3	6.816	1.878	0.853	0.488
4	9.509	2.629	1.179	0.667
5	12.27	3.419	1.521	0.853
6	15.075	4.238	1.878	1.046
7	17.913	5.08	2.248	1.246
8	20.774	5.941	2.629	1.451
9	23.653	6.816	3.02	1.662
10	26.546	7.704	3.419	1.878

10.13 Here are the 35 distinct samples from $\{2,4,9,12\}$ and their weights.

$\{12,12,12,12\}, 1/256$	$\{9,12,12,12\}, 1/64$	$\{9,9,12,12\}, 3/128$
$\{9,9,9,12\}, 1/64$	$\{9,9,9,9\}, 1/256$	$\{4,12,12,12\}, 1/64$
$\{4,9,12,12\}, 3/64$	$\{4,9,9,12\}, 3/64$	$\{4,9,9,9\}, 1/64$
$\{4,4,12,12\}, 3/128$	$\{4,4,9,12\}, 3/64$	$\{4,4,9,9\}, 3/128$
$\{4,4,4,12\}, 1/64$	$\{4,4,4,9\}, 1/64$	$\{4,4,4,4\}, 1/256$
$\{2,12,12,12\}, 1/64$	$\{2,9,12,12\}, 3/64$	$\{2,9,9,12\}, 3/64$
$\{2,9,9,9\}, 1/64$	$\{2,4,12,12\}, 3/64$	$\{2,4,9,12\}, 3/32$
$\{2,4,9,9\}, 3/64$	$\{2,4,4,12\}, 3/64$	$\{2,4,4,9\}, 3/64$
$\{2,4,4,4\}, 1/64$	$\{2,2,12,12\}, 3/128$	$\{2,2,9,12\}, 3/64$
$\{2,2,9,9\}, 3/128$	$\{2,2,4,12\}, 3/64$	$\{2,2,4,9\}, 3/64$
$\{2,2,4,4\}, 3/128$	$\{2,2,2,12\}, 1/64$	$\{2,2,2,9\}, 1/64$
$\{2,2,2,4\}, 1/64$	$\{2,2,2,2\}, 1/256$	

The verifications of parts $(a)-(d)$ can be done with this table, or the table of means in Example $A.0.1$ can be used. For part (e),verifying the bootstrap identities can involve much painful algebra, but it can be made easier if we understand what the bootstrap sample space (the space of all n^n bootstrap samples) looks like. Given a sample $x_1, x_2, \ldots, x_n$, the bootstrap sample space can be thought of as a data array with n^n rows (one for each bootstrap sample) and n columns, so each row of the data array is one bootstrap sample. For example, if the sample size is $n = 3$, the bootstrap sample space is

x_1	x_1	x_1
x_1	x_1	x_2
x_1	x_1	x_3
x_1	x_2	x_1
x_1	x_2	x_2
x_1	x_2	x_3
x_1	x_3	x_1
x_1	x_3	x_2
x_1	x_3	x_3
x_2	x_1	x_1
x_2	x_1	x_2
x_2	x_1	x_3
x_2	x_2	x_1
x_2	x_2	x_2
x_2	x_2	x_3
x_2	x_3	x_1
x_2	x_3	x_2
x_2	x_3	x_3
x_3	x_1	x_1
x_3	x_1	x_2
x_3	x_1	x_3
x_3	x_2	x_1
x_3	x_2	x_2
x_3	x_2	x_3
x_3	x_3	x_1
x_3	x_3	x_2
x_3	x_3	x_3

Note the pattern. The first column is 9 x_1s followed by 9 x_2s followed by 9 x_3s, the second column is 3 x_1s followed by 3 x_2s followed by 3 x_3s, then repeated, etc. In general, for the entire bootstrap sample,

- The first column is n^{n-1} x_1s followed by n^{n-1} x_2s followed by, ..., followed by n^{n-1} x_ns
- The second column is n^{n-2} x_1s followed by n^{n-2} x_2s followed by, ..., followed by n^{n-2} x_ns, repeated n times
- The third column is n^{n-3} x_1s followed by n^{n-3} x_2s followed by, ..., followed by n^{n-3} x_ns, repeated n^2 times

  ⋮
- The n^{th} column is 1 x_1 followed by 1 x_2 followed by, ..., followed by 1 x_n, repeated n^{n-1} times

So now it is easy to see that each column in the data array has mean $\bar{x}$, hence the entire bootstrap data set has mean $\bar{x}$. Appealing to the $3^3 \times 3$ data array, we can write the numerator of the variance of the bootstrap means as

$$\begin{aligned}
&\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left[\frac{1}{3}(x_i+x_j+x_k)-\bar{x}\right]^2 \\
&= \frac{1}{3^2}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left[(x_i-\bar{x})+(x_j-\bar{x})+(x_k-\bar{x})\right]^2 \\
&= \frac{1}{3^2}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left[(x_i-\bar{x})^2+(x_j-\bar{x})^2+(x_k-\bar{x})^2\right],
\end{aligned}$$

because all of the cross terms are zero (since they are the sum of deviations from the mean). Summing up and collecting terms shows that

$$\frac{1}{3^2}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left[(x_i-\bar{x})^2+(x_j-\bar{x})^2+(x_k-\bar{x})^2\right] = 3\sum_{i=1}^{3}(x_i-\bar{x})^2,$$

and thus the average of the variance of the bootstrap means is

$$\frac{3\sum_{i=1}^{3}(x_i-\bar{x})^2}{3^3}$$

which is the usual estimate of the variance of $\bar{X}$ if we divide by n instead of $n-1$. The general result should now be clear. The variance of the bootstrap means is

$$\begin{aligned}
&\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_n=1}^{n}\left[\frac{1}{n}(x_{i_1}+x_{i_2}+\cdots+x_{i_n})-\bar{x}\right]^2 \\
&= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_n=1}^{n}\left[(x_{i_1}-\bar{x})^2+(x_{i_2}-\bar{x})^2+\cdots+(x_{i_n}-\bar{x})^2\right],
\end{aligned}$$

since all of the cross terms are zero. Summing and collecting terms shows that the sum is $n^{n-2}\sum_{i=1}^{n}(x_i-\bar{x})^2$, and the variance of the bootstrap means is $n^{n-2}\sum_{i=1}^{n}(x_i-\bar{x})^2/n^n = \sum_{i=1}^{n}(x_i-\bar{x})^2/n^2$.

10.15 a. As $B\to\infty$ $\text{Var}^*_B(\hat{\theta}) = \text{Var}^*(\hat{\theta})$.

b. Each $\text{Var}^*_{B_i}(\hat{\theta})$ is a sample variance, and they are independent so the LLN applies and

$$\frac{1}{m}\sum_{i=1}^{m}\text{Var}^*_{B_i}(\hat{\theta}) \overset{m\to\infty}{\to} \text{E}\text{Var}^*_B(\hat{\theta}) = \text{Var}^*(\hat{\theta}),$$

where the last equality follows from Theorem 5.2.6(c).

10.17 a. The correlation is .7781

b. Here is R code (R is available free at http://cran.r-project.org/) to bootstrap the data, calculate the standard deviation, and produce the histogram:

```
cor(law)
n <- 15
theta <- function(x,law){ cor(law[x,1],law[x,2]) }
results <- bootstrap(1:n,1000,theta,law,func=sd)
results[2]
hist(results[[1]])
```

The data "law" is in two columns of length 15, "results[2]" contains the standard deviation. The vector "results[[1]]" is the bootstrap sample. The output is

```
          V1        V2
V1 1.0000000 0.7781716
V2 0.7781716 1.0000000
$func.thetastar
[1] 0.1322881
```

showing a correlation of .7781 and a bootstrap standard deviation of .1323.

c. The R code for the parametric bootstrap is

```
mx<-600.6;my<-3.09
sdx<-sqrt(1791.83);sdy<-sqrt(.059)
rho<-.7782;b<-rho*sdx/sdy;sdxy<-sqrt(1-rho^2)*sdx
rhodata<-rho
for (j in 1:1000)  {
y<-rnorm(15,mean=my,sd=sdy)
x<-rnorm(15,mean=mx+b*(y-my),sd=sdxy)
rhodata<-c(rhodata,cor(x,y))
}
sd(rhodata)
hist(rhodata)
```

where we generate the bivariate normal by first generating the marginal then the condidional, as R does not have a bivariate normal generator. The bootstrap standard deviation is 0.1159, smaller than the nonparametric estimate. The histogram looks similar to the nonparametric bootstrap histogram, displaying a skewness left.

d. The Delta Method approximation is

$$r \sim \mathrm{n}(\rho, (1-\rho^2)^2/n),$$

and the "plug-in" estimate of standard error is $\sqrt{(1-.7782^2)^2/15} = .1018$, the smallest so far. Also, the approximate pdf of r will be normal, hence symmetric.

e. By the change of variables

$$t = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right), \qquad dt = \frac{1}{1-r^2},$$

the density of r is

$$\frac{1}{\sqrt{2\pi}(1-r^2)}\exp\left(-\frac{n}{2}\left[\frac{1}{2}\log\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)\right]^2\right), \quad -1 \le r \le 1.$$

More formally, we could start with the random variable T, normal with mean $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$ and variance $1/n$, and make the transformation to $R = \frac{e^{2T}+1}{e^{2T}-1}$ and get the same answer.

10.19 a. The variance of $\bar{X}$ is

$$\begin{aligned}\mathrm{Var}\bar{X} = \mathrm{E}(\bar{X}-\mu)^2 &= \mathrm{E}\left(\frac{1}{n}\sum_i X_i - \mu\right)^2 \\ &= \frac{1}{n^2}\mathrm{E}\left(\sum_i (X_i-\mu)^2 + 2\sum_{i>j}(X_i-\mu)(X_j-\mu)\right) \\ &= \frac{1}{n^2}\left(n\sigma^2 + 2\frac{n(n-1)}{2}\rho\sigma^2\right) \\ &= \frac{\sigma^2}{n} + \frac{n-1}{n}\rho\sigma^2\end{aligned}$$

b. In this case we have

$$\mathrm{E}\left[\sum_{i>j}(X_i-\mu)(X_j-\mu)\right] = \sigma^2\sum_{i=2}^{n}\sum_{j=1}^{i-1}\rho^{i-j}.$$

In the double sum ρ appears $n-1$ times, ρ^2 appears $n-2$ times, etc.. so

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1}\rho^{i-j} = \sum_{i=1}^{n-1}(n-i)\rho^i = \frac{\rho}{1-\rho}\left(n - \frac{1-\rho^n}{1-\rho}\right),$$

where the series can be summed using (1.5.4), the partial sum of the geometric series, or using Mathematica.

c. The mean and variance of X_i are

$$\mathrm{E}X_i = \mathrm{E}[\mathrm{E}(X_i|X_{i-1})] = \mathrm{E}\rho X_{i-1} = \cdots = \rho^{i-1}\mathrm{E}X_1$$

and

$$\mathrm{Var}X_i = \mathrm{Var}\mathrm{E}(X_i|X_{i-1}) + \mathrm{E}\mathrm{Var}(X_i|X_{i-1}) = \rho^2\sigma^2 + 1 = \sigma^2$$

for $\sigma^2 = 1/(1-\rho^2)$. Also, by iterating the expectation

$$\mathrm{E}X_1X_i = \mathrm{E}[\mathrm{E}(X_1X_i|X_{i-1})] = \mathrm{E}[\mathrm{E}(X_1|X_{i-1})\mathrm{E}(X_i|X_{i-1})] = \rho\mathrm{E}[X_1X_{i-1}],$$

where we used the facts that X_1 and X_i are independent conditional on X_{i-1}. Continuing with the argument we get that $\mathrm{E}X_1X_i = \rho^{i-1}\mathrm{E}X_1^2$. Thus,

$$\mathrm{Corr}(X_1,X_i) = \frac{\rho^{i-1}\mathrm{E}X_1^2 - \rho^{i-1}(\mathrm{E}X_1)^2}{\sqrt{\mathrm{Var}X_1\mathrm{Var}X_i}} = \frac{\rho^{i-1}\sigma^2}{\sqrt{\sigma^2\sigma^2}} = \rho^{i-1}.$$

10.21 a. If any $x_i \to \infty$, $s^2 \to \infty$, so it has breakdown value 0. To see this, suppose that $x_1 \to \infty$. Write

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n-1}\left([(1-\frac{1}{n})x_1 - \bar{x}_{-1}]^2 + \sum_{i=2}^{n}(x_i-\bar{x})^2\right),$$

where $\bar{x}_{-1} = (x_2 + \ldots + x_n)/n$. It is easy to see that as $x_1 \to \infty$, each term in the sum $\to \infty$.

b. If less than 50% of the sample $\to \infty$, the median remains the same, and the median of $|x_i - M|$ remains the same. If more than 50% of the sample $\to \infty$, $M \to \infty$ and so does the MAD.

10.23 a. The ARE is $[2\sigma f(\mu)]^2$. We have

Distribution	Parameters	variance	$f(\mu)$	ARE
normal	$\mu=0,\sigma=1$	1	.3989	.64
logistic	$\mu=0,\beta=1$	$\pi^2/3$	.25	.82
double exp.	$\mu=0,\sigma=1$	2	.5	2

b. If $X_1, X_2, \ldots, X_n$ are iid f_X with $\mathrm{E}X_1 = \mu$ and $\mathrm{Var}X_1 = \sigma^2$, the ARE is $\sigma^2[2 * f_X(\mu)]^2$. If we transform to $Y_i = (X_i - \mu)/\sigma$, the pdf of Y_i is $f_Y(y) = \sigma f_X(\sigma y + \mu)$ with ARE $[2 * f_Y(0)]^2 = \sigma^2[2 * f_X(\mu)]^2$

c. The median is more efficient for smaller ν, the distributions with heavier tails.

ν	Var$X$	$f(0)$	ARE
3	3	.367	1.62
5	5/3	.379	.960
10	5/4	.389	.757
25	25/23	.395	.678
50	25/24	.397	.657
∞	1	.399	.637

d. Again the heavier tails favor the median.

δ	σ	ARE
.01	2	.649
.1	2	.747
.5	2	.895
.01	5	.777
.1	5	1.83
.5	5	2.98

10.25 By transforming $y = x - \theta$,

$$\int_{-\infty}^{\infty} \psi(x-\theta) f(x-\theta)dx = \int_{-\infty}^{\infty} \psi(y) f(y) dy.$$

Since ψ is an odd function, $\psi(y) = -\psi(-y)$, and

$$\begin{aligned}
\int_{-\infty}^{\infty} \psi(y)f(y)dy &= \int_{-\infty}^{0} \psi(y)f(y)dy + \int_{0}^{\infty} \psi(y)f(y)dy \\
&= \int_{-\infty}^{0} -\psi(-y)f(y)dy + \int_{0}^{\infty} \psi(y)f(y)dy \\
&= -\int_{0}^{\infty} \psi(y)f(y)dy + \int_{0}^{\infty} \psi(y)f(y)dy = 0,
\end{aligned}$$

where in the last line we made the transformation $y \to -y$ and used the fact the f is symmetric, so $f(y) = f(-y)$. From the discussion preceding Example 10.2.6, $\hat{\theta}_M$ is asymptotically normal with mean equal to the true θ.

10.27 a.

$$\lim_{\delta\to 0} \frac{1}{\delta}[(1-\delta)\mu + \delta x - \mu] = \lim_{\delta \to 0} \frac{\delta(x-\mu)}{\delta} = x - \mu.$$

b.

$$P(X \le a) = P(X \le a | X \sim F)(1-\delta) + P(x \le a | X = x)\delta = (1-\delta)F(a) + \delta I(x \le a)$$

and

$$(1-\delta)F(a) = \frac{1}{2} \quad \Rightarrow \quad a = F^{-1}\left(\frac{1}{2(1-\delta)}\right)$$
$$(1-\delta)F(a)+\delta = \frac{1}{2} \quad \Rightarrow \quad a = F^{-1}\left(\frac{\frac{1}{2}-\delta}{2(1-\delta)}\right)$$

c. The limit is

$$\lim_{\delta\to 0}\frac{a_\delta - a_0}{\delta} = a'_\delta|_{\delta=0}$$

by the definition of derivative. Since $F(a_\delta) = \frac{1}{2(1-\delta)}$,

$$\frac{d}{d\delta}F(a_\delta) = \frac{d}{d\delta}\frac{1}{2(1-\delta)}$$

or

$$f(a_\delta)a'_\delta = \frac{1}{2(1-\delta)^2} \Rightarrow a'_\delta = \frac{1}{2(1-\delta)^2 f(a_\delta)}.$$

Since $a_0 = m$, the result follows. The other limit can be calculated in a similar manner.

10.29 a. Substituting cl' for ψ makes the ARE equal to 1.

b. For each distribution is the case that the given ψ function is equal to cl', hence the resulting M-estimator is asymptotically efficient by (10.2.9).

10.31 a. By the CLT,

$$\sqrt{n_1}\frac{\hat{p}_1 - p_1}{\sqrt{p_1(1-p_1)}} \to \mathrm{n}(0,1) \quad \text{and} \quad \sqrt{n_2}\frac{\hat{p}_2 - p_2}{\sqrt{p_2(1-p_2)}} \to \mathrm{n}(0,1),$$

so if $\hat{p}_1$ and $\hat{p}_2$ are independent, under $H_0: p_1 = p_2 = p$,

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{p}(1-\hat{p})}} \to \mathrm{n}(0,1)$$

where we use Slutsky's Theorem and the fact that $\hat{p} = (S_1+S_2)/(n_1+n_2)$ is the MLE of p under H_0 and converges to p in probability. Therefore, $T \to \chi^2_1$.

b. Substitute $\hat{p}_i$s for S_i and F_is to get

$$\begin{aligned} T^* &= \frac{n_1^2(\hat{p}_1-\hat{p})^2}{n_1\hat{p}} + \frac{n_2^2(\hat{p}_2-\hat{p})^2}{n_2\hat{p}} \\ &\quad + \frac{n_1^2\left[(1-\hat{p}_1)-(1-\hat{p})\right]^2}{n_1(1-\hat{p})} + \frac{n_2^2\left[(1-\hat{p}_2)-(1-\hat{p})\right]^2}{n_2\hat{p}} \\ &= \frac{n_1(\hat{p}_1-\hat{p})^2}{\hat{p}(1-\hat{p})} + \frac{n_2(\hat{p}_2-\hat{p})^2}{\hat{p}(1-\hat{p})} \end{aligned}$$

Write $\hat{p} = (n_1\hat{p}_1 + n_2\hat{p}_2)/(n_1+n_2)$. Substitute this into the numerator, and some algebra will get

$$n_1(\hat{p}_1-\hat{p})^2 + n_2(\hat{p}_2-\hat{p})^2 = \frac{(\hat{p}_1-\hat{p}_2)^2}{\frac{1}{n_1}+\frac{1}{n_2}},$$

so $T^* = T$.

c. Under H_0,

$$\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)p(1-p)}} \to \mathrm{n}(0,1)$$

and both $\hat{p}_1$ and $\hat{p}_2$ are consistent, so $\hat{p}_1(1-\hat{p}_1) \to p(1-p)$ and $\hat{p}_2(1-\hat{p}_2) \to p(1-p)$ in probability. Therefore, by Slutsky's Theorem,

$$\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \to \mathrm{n}(0,1),$$

and $(T^{**})^2 \to \chi^2_1$. It is easy to see that $T^{**} \neq T$ in general.

d. The estimator $(1/n_1+1/n_2)\hat{p}(1-\hat{p})$ is the MLE of $\mathrm{Var}(\hat{p}_1-\hat{p}_2)$ under H_0, while the estimator $\hat{p}_1(1-\hat{p}_1)/n_1+\hat{p}_2(1-\hat{p}_2)/n_1$ is the MLE of $\mathrm{Var}(\hat{p}_1-\hat{p}_2)$ under H_1. One might argue that in hypothesis testing, the first one should be used, since under H_0, it provides a better estimator of variance. If interest is in finding the confidence interval, however, we are making inference under both H_0 and H_1, and the second one is preferred.

e. We have $\hat{p}_1 = 34/40$, $\hat{p}_2 = 19/35$, $\hat{p} = (34+19)/(40+35) = 53/75$, and $T = 8.495$. Since $\chi^2_{1,.05} = 3.84$, we can reject H_0 at $\alpha = .05$.

10.32 a. First calculate the MLEs under $p_1 = p_2 = p$. We have

$$L(p|x) = p^{x_1}p^{x_2}p^{x_3}\cdots p_{n-1}^{x_{n-1}}\left(1-2p-\sum_{i=3}^{n-1}p_i\right)^{m-x_1-x_2-\cdots-x_{n-1}}.$$

Taking logs and differentiating yield the following equations for the MLEs:

$$\frac{\partial \log L}{\partial p} = \frac{x_1+x_2}{p} - \frac{2\left(m-\sum_{i=1}^{n-1}x_i\right)}{1-2p-\sum_{i=3}^{n-1}p_i} = 0$$

$$\frac{\partial \log L}{\partial p_i} = \frac{x_i}{p_i} - \frac{x_n}{1-2p-\sum_{i=3}^{n-1}p_i} = 0, \quad i = 3,\ldots,n-1,$$

with solutions $\hat{p} = \frac{x_1+x_2}{2m}$, $\hat{p}_i = \frac{x_i}{m}, i = 3,\ldots,n-1$, and $\hat{p}_n = \left(m-\sum_{i=1}^{n-1}x_i\right)/m$. Except for the first and second cells, we have expected = observed, since both are equal to x_i. For the first two terms, expected $= m\hat{p} = (x_1+x_2)/2$ and we get

$$\sum\frac{(observed-expected)^2}{expected} = \frac{\left(x_1-\frac{x_1+x_2}{2}\right)^2}{\frac{x_1+x_2}{2}} + \frac{\left(x_2-\frac{x_1+x_2}{2}\right)^2}{\frac{x_1+x_2}{2}} = \frac{(x_1-x_2)^2}{x_1+x_2}.$$

b. Now the hypothesis is about conditional probabilities is given by H_0: P(change—initial agree)=P(change—initial disagree) or, in terms of the parameters $H_0: \frac{p_1}{p_1+p_3} = \frac{p_2}{p_2+p_4}$. This is the same as $p_1p_4 = p_2p_3$, which is not the same as $p_1 = p_2$.

10.33 Theorem 10.1.12 and Slutsky's Theorem imply that

$$\frac{\hat{\theta}-\theta}{\sqrt{\frac{1}{n}I_n(\hat{\theta})}} \to \mathrm{n}(0,1)$$

and the result follows.

10.35 a. Since $\sigma/\sqrt{n}$ is the estimated standard deviation of $\bar{X}$ in this case, the statistic is a Wald statistic

b. The MLE of σ^2 is $\hat{\sigma}^2_\mu = \sum_i (x_i - \mu)^2/n$. The information number is

$$-\frac{d^2}{d(\sigma^2)^2}\left(-\frac{n}{2}\log\sigma^2 - \frac{1}{2}\frac{\hat{\sigma}^2_\mu}{\sigma^2}\right)\Bigg|_{\sigma^2=\hat{\sigma}^2_\mu} = \frac{n}{2\hat{\sigma}^2_\mu}.$$

Using the Delta method, the variance of $\hat{\sigma}_\mu = \sqrt{\hat{\sigma}^2_\mu}$ is $\hat{\sigma}^2_\mu/8n$, and a Wald statistic is

$$\frac{\hat{\sigma}_\mu - \sigma_0}{\sqrt{\sigma^2_\mu/8n}}.$$

10.37 a. The log likelihood is

$$\log L = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_i (x_i - \mu)^2/\sigma^2$$

with

$$\begin{aligned}\frac{d}{d\mu} &= \frac{1}{\sigma^2}\sum_i (x_i - \mu) = \frac{n}{\sigma^2}(\bar{x} - \mu)\\ \frac{d^2}{d\mu^2} &= -\frac{n}{\sigma^2},\end{aligned}$$

so the test statistic for the score test is

$$\frac{\frac{n}{\sigma^2}(\bar{x} - \mu)}{\sqrt{\sigma^2/n}} = \sqrt{n}\frac{\bar{x} - \mu}{\sigma}$$

b. We test the equivalent hypothesis $H_0 : \sigma^2 = \sigma_0^2$. The likelihood is the same as Exercise 10.35(b), with first derivative

$$-\frac{d}{d\sigma^2} = \frac{n(\hat{\sigma}^2_\mu - \sigma^2)}{2\sigma^4}$$

and expected information number

$$\mathrm{E}\left(\frac{n(2\hat{\sigma}^2_\mu - \sigma^2)}{2\sigma^6}\right) = \frac{n(2\sigma^2 - \sigma^2)}{2\sigma^6} = \frac{n}{2\sigma^4}.$$

The score test statistic is

$$\sqrt{\frac{n}{2}}\frac{\hat{\sigma}^2_\mu - \sigma^2_0}{\sigma^2_0}$$

10.39 We summarize the results for (a) – (c) in the following table. We assume that the underlying distribution is normal, and use that for all score calculations. The actual data is generated from normal, logistic, and double exponential. The sample size is 15, we use 1000 simulations and draw 20 bootstrap samples. Here $\theta_0 = 0$, and the power is tabulated for a nominal $\alpha = .1$ test.

Underlying pdf	Test	θ_0	$\theta_0 + .25\sigma$	$\theta_0 + .5\sigma$	$\theta_0 + .75\sigma$	$\theta_0 + 1\sigma$	$\theta_0 + 2\sigma$
Laplace	Naive	0.101	0.366	0.774	0.957	0.993	1.
	Boot	0.097	0.364	0.749	0.932	0.986	1.
	Median	0.065	0.245	0.706	0.962	0.995	1.
Logistic	Naive	0.137	0.341	0.683	0.896	0.97	1.
	Boot	0.133	0.312	0.641	0.871	0.967	1.
	Median	0.297	0.448	0.772	0.944	0.993	1.
Normal	Naive	0.168	0.316	0.628	0.878	0.967	1.
	Boot	0.148	0.306	0.58	0.836	0.957	1.
	Median	0.096	0.191	0.479	0.761	0.935	1.

Here is Mathematica code:

This program calculates size and power for Exercise 10.39, Second Edition

We do our calculations assuming normality, but simulate power and size under other distributions. We test $H_0 : \theta = 0$.

```
theta_0=0;
Needs["Statistics‘Master‘"]
Clear[x]
f1[x_]=PDF[NormalDistribution[0,1],x];
F1[x_]=CDF[NormalDistribution[0,1],x];
f2[x_]=PDF[LogisticDistribution[0,1],x];
f3[x_]=PDF[LaplaceDistribution[0,1],x];
v1=Variance[NormalDistribution[0,1]];
v2=Variance[LogisticDistribution[0,1]];
v3=Variance[LaplaceDistribution[0,1]];
```

Calculate m-estimate

```
Clear[k,k1,k2,t,x,y,d,n,nsim,a,w1]
ind[x_,k_]:=If[Abs[x]<k,1,0]
rho[y_,k_]:=ind[y,k]*y^2 + (1-ind[y,k])*(k*Abs[y]-k^2)
alow[d_]:=Min[Mean[d],Median[d]]
aup[d_]:=Max[Mean[d],Median[d]]
sol[k_,d_]:=FindMinimum[Sum[rho[d[[i]]-a,k],{i,1,n}],{a,{alow[d],aup[d]}}]
mest[k_,d_]:=sol[k,d][[2]]
```

generate data - to change underlying distributions change the sd and the distribution in the Random statement.

```
n = 15; nsim = 1000; sd = Sqrt[v1];
theta = {theta_0, theta_0 +.25*sd, theta_0 +.5*sd,
        theta_0 +.75*sd, theta_0 + 1*sd, theta_0 +2*sd}
ntheta = Length[theta]
data = Table[Table[Random[NormalDistribution[0, 1]],
       {i, 1, n}],{j, 1,nsim}];
m1 = Table[Table[a /. mest[k1, data[[j]] - theta[[i]]],
       {j, 1, nsim}], {i, 1, n\theta}];
```

Calculation of naive variance and test statistic

```
Psi[x_, k_] =  x*If[Abs[x]<= k, 1, 0]-  k*If[x < -k, 1, 0] +
```

```
              k*If[x > k, 1, 0];
Psi1[x_, k_] = If[Abs[x]  <=  k, 1, 0];
num =Table[Psi[w1[[j]][[i]], k1], {j, 1, nsim}, {i, 1,n}];
den =Table[Psi1[w1[[j]][[i]], k1], {j, 1, nsim}, {i, 1,n}];
varnaive = Map[Mean, num^2]/Map[Mean, den]^2;
naivestat = Table[Table[m1[[i]][[j]] -theta_0/Sqrt[varnaive[[j]]/n],
            {j, 1, nsim}],{i, 1, ntheta}];
absnaive = Map[Abs, naivestat];
N[Table[Mean[Table[If[absnaive[[i]][[j]] > 1.645, 1, 0],
           {j, 1, nsim}]], {i, 1, n\theta}]]
```

Calculation of bootstrap variance and test statistic

```
nboot=20;
u:=Random[DiscreteUniformDistribution[n]]
databoot=Table[data[[jj]][[u]],{jj,1,nsim},{j,1,nboot},{i,1,n}];
m1boot=Table[Table[a/.mest[k1,databoot[[j]][[jj]]],
          {jj,1,nboot}],{j,1,nsim}];
varboot = Map[Variance, m1boot];
bootstat = Table[Table[m1[[i]][[j]] -theta_0/Sqrt[varboot[[j]]],
          {j, 1, nsim}], {i, 1, ntheta}];
absboot = Map[Abs, bootstat];
N[Table[Mean[Table[If[absboot[[i]][[j]] > 1.645, 1,0],
          {j, 1, nsim}]], {i, 1, ntheta}]]\)
```

Calculation of median test - use the score variance at the root density (normal)

```
med = Map[Median, data];
medsd = 1/(n*2*f1[theta_0]);
medstat = Table[Table[med[[j]] + \theta[[i]] - theta_0/medsd,
          {j, 1, nsim}], {i, 1, ntheta}];
absmed = Map[Abs, medstat];
N[Table[Mean[Table[If[\(absmed[[i]][[j]] > 1.645, 1, 0],
          {j, 1, nsim}]], {i, 1, ntheta}]]
```

10.41 a. The log likelihood is

$$\log L = nr \log p + n\bar{x}\log(1-p)$$

with

$$\frac{d}{dp}\log L = \frac{nr}{p} - \frac{n\bar{x}}{1-p} \quad \text{and} \quad \frac{d^2}{dp^2}\log L = -\frac{nr}{p^2} - \frac{n\bar{x}}{(1-p)^2},$$

expected information $\frac{nr}{p^2(1-p)}$ and (Wilks) score test statistic

$$\sqrt{n}\frac{\left(\frac{r}{p} - \frac{n\bar{x}}{1-p}\right)}{\sqrt{\frac{r}{p^2(1-p)}}} = \sqrt{\frac{n}{r}}\left(\frac{(1-p)r + p\bar{x}}{\sqrt{1-p}}\right).$$

Since this is approximately $\mathrm{n}(0,1)$, a $1-\alpha$ confidence set is

$$\left\{p : \left|\sqrt{\frac{n}{r}}\left(\frac{(1-p)r - p\bar{x}}{\sqrt{1-p}}\right)\right| \le z_{\alpha/2}\right\}.$$

b. The mean is $\mu = r(1-p)/p$, and a little algebra will verify that the variance, $r(1-p)/p^2$ can be written $r(1-p)/p^2 = \mu + \mu^2/r$. Thus

$$\sqrt{\frac{n}{r}}\left(\frac{(1-p)r - p\bar{x}}{\sqrt{1-p}}\right) = \sqrt{n}\frac{\mu - \bar{x}}{\sqrt{\mu + \mu^2/r}}.$$

The confidence interval is found by setting this equal to $z_{\alpha/2}$, squaring both sides, and solving the quadratic for μ. The endpoints of the interval are

$$\frac{r(8\bar{x} + z^2_{\alpha/2}) \pm \sqrt{rz^2_{\alpha/2}}\sqrt{16r\bar{x} + 16\bar{x}^2 + rz^2_{\alpha/2}}}{8r - 2z^2_{\alpha/2}}.$$

For the continuity correction, replace $\bar{x}$ with $\bar{x}+1/(2n)$ when solving for the upper endpoint, and with $\bar{x} - 1/(2n)$ when solving for the lower endpoint.

c. We table the endpoints for $\alpha = .1$ and a range of values of r. Note that $r = \infty$ is the Poisson, and smaller values of r give a wider tail to the negative binomial distribution.

r	lower bound	upper bound
1	22.1796	364.42
5	36.2315	107.99
10	38.4565	95.28
50	40.6807	85.71
100	41.0015	84.53
1000	41.3008	83.46
∞	41.3348	83.34

10.43 a. Since

$$P\left(\sum_i X_i = 0\right) = (1-p)^n = \alpha/2 \Rightarrow p = 1 - \alpha^{1/n}$$

and

$$P\left(\sum_i X_i = n\right) = p^n = \alpha/2 \Rightarrow p = \alpha^{1/n},$$

these endpoints are exact, and are the shortest possible.

b. Since $p \in [0,1]$, any value outside has zero probability, so truncating the interval shortens it at no cost.

10.45 The continuity corrected roots are

$$\frac{2\hat{p} + z^2_{\alpha/2}/n \pm \frac{1}{n} \pm \sqrt{\frac{z^2_{\alpha/2}}{n^3}[\pm 2n(1-2\hat{p}) - 1] + (2\hat{p} + z^2_{\alpha/2}/n)^2 - 4\hat{p}^2(1 + z^2_{\alpha/2}/n)}}{2(1 + z^2_{\alpha/2}/n)}$$

where we use the upper sign for the upper root and the lower sign for the lower root. Note that the only differences between the continuity-corrected intervals and the ordinary score intervals are the terms with $\pm$ in front. But it is still difficult to analytically compare lengths with the non-corrected interval - we will do a numerical comparison. For $n = 10$ and $\alpha = .1$ we have the following table of length ratios, with the continuity-corrected length in the denominator

n	0	1	2	3	4	5	6	7	8	9	10
Ratio	0.79	0.82	0.84	0.85	0.86	0.86	0.86	0.85	0.84	0.82	0.79

The coverage probabilities are

p	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1
score	.99	.93	.97	.92	.90	.89	.90	.92	.97	.93	.99
cc	.99	.99	.97	.92	.98	.98	.98	.92	.97	.99	.99

Mathematica code to do the calculations is:

```
Needs["Statistics`Master`"]
Clear[p, x]
pbino[p_, x_] = PDF[BinomialDistribution[n, p], x];
cut = 1.645^2;
n = 10;
```

The quadratic score interval with and without continuity correction

```
slowcc[x_] := p /. FindRoot[(x/n - 1/(2*n) - p)^2 ==
              cut*(p*((1 - p))/n, {p,  .001}]
supcc[x_] :=  p /. FindRoot[(x/n + 1/(2*n) - p)^2 ==
              cut*(p*((1 - p)/n, {p,  .999}]
slow[x_] :=   p /. FindRoot[(x/n - p))^2 ==
              cut*(p*(1 - p))/n, {p,  .001}]
sup[x_] :=    p /. FindRoot[(x/n - p)^2 ==
              cut*(p*(1 - p)/n, {p,  .999}]
scoreintcc=Partition[Flatten[{{0,sup[0]},Table[{slowcc[i],supcc[i]},
              {i,1,n-1}],{slowcc[n],1}},2],2];
scoreint=Partition[Flatten[{{0,sup[0]},Table[{slow[i],sup[i]},
              {i,1,n-1}],{slowcc[n],1}},2],2];
```

Length Comparison

```
Table[(sup[i] - slow[i])/(supcc[i] - slowcc[i]), {i, 0, n}]
```

Now we'll calculate coverage probabilities

```
scoreindcc[p_,x_]:=If[scoreintcc[[x+1]][[1]]<=p<=scoreintcc[[x+1]][[2]],1,0]
scorecovcc[p_]:=scorecovcc[p]=Sum[pbino[p,x]*scoreindcc[p,x],{x,0,n}]
scoreind[p_,x_]:=If[scoreint[[x+1]][[1]]<=p<=scoreint[[x+1]][[2]],1,0]
scorecov[p_]:=scorecov[p]=Sum[pbino[p,x]*scoreind[p,x],{x,0,n}]
{scorecovcc[.0001],Table[scorecovcc[i/10],{i,1,9}],scorecovcc[.9999]}//N
{scorecov[.0001],Table[scorecov[i/10],{i,1,9}],scorecov[.9999]}//N
```

10.47 a. Since $2pY \sim \chi^2_{nr}$ (approximately)

$$P(\chi^2_{nr,1-\alpha/2} \le 2pY \le \chi^2_{nr,\alpha/2}) = 1-\alpha,$$

and rearrangment gives the interval.

b. The interval is of the form $P(a/2Y \le p \le b/2Y)$, so the length is proportional to $b-a$. This must be minimized subject to the constraint $\int_a^b f(y)dy = 1-\alpha$, where $f(y)$ is the pdf of a χ^2_{nr}. Treating b as a function of a, differentiating gives

$$b' - 1 = 0 \quad \text{and} \quad f(b)b' - f(a) = 0$$

which implies that we need $f(b) = f(a)$.