Problem 4.1.3. Suppose the number of customers X that enter a store between the hours of 9:00 am and 10:00 am follows a Poisson distribution with parameter θ. Suppose a random sample of the number of customers that enter the store between 9:00 am and 10:00 am for 10 days results in the values

$$9 \quad 7 \quad 9 \quad 15 \quad 10 \quad 13 \quad 11 \quad 7 \quad 2 \quad 12.$$

(a) Determine the maximum likelihood extimate of θ. Show that it is an unbiased estimator.
(b) Based on these data, obtain the realization of your estimator in part (a). Explain the meaning of this estimate in terms of the number of customers.

Solution 4.1.3.
(a) The probability mass function for X is $p(x;\theta) = e^{-\theta}\theta^x/x!$, $x = 0, 1, 2, \ldots$, zero elsewhere; $\theta \in \Omega = (0, \infty)$. The natural log of the likelihood function is therefore

$$\begin{aligned}\ell(\theta) = \ln L(\theta; x_1, x_2, \ldots, x_n) &= \ln \frac{e^{-n\theta}\theta^{x_1+x_2+\cdots+x_n}}{x_1!x_2!\cdots x_n!} \\ &= -n\theta + n\bar{x}\ln(\theta) - \ln(x_1!x_2!\cdots x_n!).\end{aligned}$$

Here we use the observation that $x_1 + x_2 + \cdots + x_n = n\bar{x}$. Differentiation gives $\ell'(\theta) = -n(\theta - \bar{x})/\theta$. Solving $\ell'(\hat{\theta}) = -n(\hat{\theta} - \bar{x})/\hat{\theta} = 0$ gives $\hat{\theta} = \bar{x}$. Since $\ell''(\theta) = -n\bar{x}/\theta^2 < 0$, we see that $\hat{\theta}$ maximizes $\ell(\theta)$. Therefore the Maximum Likelihood Estimator for θ is

$$\hat{\theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Since the parameter θ is the expected value of a Poisson random variable, we see that $E[\hat{\theta}] = E[\overline{X}] = E[X] = \theta$, which shows that the MLE estimator for θ is unbiased.

(b) Our realization of $\hat{\theta}$, based on the observed data, is

$$\hat{\theta} = \frac{9 + 7 + 9 + 15 + 10 + 13 + 11 + 7 + 2 + 12}{10} = 9.5.$$

This means that, on average, we shold expect to see about 9.5 customers come into the store between the hours of 9:00 am and 10:00 am.

Problem 4.2.1. Let the observed value of the mean $\overline{X}$ and of the sample variance of a random sample of size 20 from a distribution that is $N(\mu, \sigma^2)$ be 81.2 and 26.5, respectively. Find respectively 90%, 95% and 99% confidence intervals for μ. Note how the lengths of the confidence intervals increase as the confidence increases.

Solution 4.2.1. We will use the confidence interval found in equation (4.2.3) on page 215. We are given $n = 20$, $\bar{x} = 81.2$, $s = \sqrt{26.5} \doteq 5.148$, and we calculate the standard error $s/\sqrt{n} = \sqrt{26.5/20} \doteq 1.151$. We use Table IV on page 660 to find the upper $\alpha/2$ critical values for the Student's t-distribution with $n - 1 = 19$ degrees of freedom; $P(t > t_{\alpha/2,19}) = \alpha/2$.

α	$t_{\alpha/2,19}$	$\frac{s}{\sqrt{n}} t_{\alpha/2,19}$	$\left(\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2,19}, \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2,19}\right)$
0.10	1.729	1.990	$(79.210, 83.190)$
0.05	2.093	2.409	$(78.791, 83.609)$
0.01	2.861	3.293	$(77.907, 84.493)$

TABLE 1. $(1 - \alpha)100\%$ Confidence Intervals for μ

Problem 4.2.2. Consider the data on the lifetimes of motors given in Exercise 4.1.1. Obtain a large sample confidence interval for the mean lifetime of a motor.

Solution 4.2.2. The lifetime in hours of twenty test motors is given below.

1	4	5	21	22	28	40	42	51	53
58	67	95	124	124	160	202	260	303	363

We calculate the sample mean $\bar{x} = 101.15$, the sample standard deviation $s = 105.41$, the standard error $s/\sqrt{20} = 23.57$, and we use Table III on page 659 to find $z_{0.025} = 1.960$. Then an approximate 95% confidence interval for the mean μ has endpoints $\bar{x} \pm z_{0.025} s/\sqrt{20} = 101.150 \pm 46.197$. The interval is therefore $(54.953, 147.347)$.

Problem 4.2.4. Suppose we assume that $X_1, X_2, \ldots, X_n$ is a random sample from a $\Gamma(1, \theta)$.

(a) Show that the random variable $(2/\theta) \sum_{i=1}^{n} X_i$ has a χ^2-distribution with $2n$ degrees of freedom.
(b) Using the random variable in part (a) as a pivot random variable, find a $(1 - \alpha)100\%$ confidence interval for θ.
(c) Obtain the confidence interval in part (b) for the data of Exercise 4.1.1 and compare it with the interval you obtained in Exercise 4.2.2.

Solution 4.2.4.

(a) The moment generating function for X is $m(t) = E[e^{tX}] = (1 + \theta t)^{-1}$, thus the moment generating function for $2X/\theta$ is $E[e^{(2t/\theta)X}] = m(2t/\theta) = (1 + 2t)^{-1}$. Since this is the mgf for a χ^2-distribution with 2 degrees of freedom, we see that $2X/\theta$ has a χ^2-distribution with 2 degrees of freedom. By Corollary 3.3.1 on page 161, the sum of n independent random variables each having a χ^2-distribution with 2 degrees of freedom is a χ^2-distributed random variable with $2n$ degrees of freedom. Since each term of $\sum_{i=1}^{n} 2X_i/\theta$ has a χ^2 distribution with 2 degrees of freedom, and since $X_1, X_2, \dots, X_n$ are independent random variables, we see that $(2/\theta)\sum_{i=1}^{n} X_i$ has a χ^2-distribution with $2n$ degrees of freedom.

(b) Let W be a random variable that has a χ^2-distribution with r degrees of freedom. Let $w_{\alpha,r}$ be the real number for which $P(W > w_{\alpha,r}) = \alpha$. For example, using Table II on page 658 with $r = 16$ degrees of freedom $P(W \le 28.845) = 0.975$. Thus $P(W > 28.845) = 0.025$ and we see that $w_{0.025,16} = 28.845$.

Let $r = 2n$ and observe that

$$P(w_{1-(\alpha/2),2n} \le W \le w_{\alpha/2,2n}) = 1 - (\alpha/2) - (\alpha/2) = 1 - \alpha.$$

A straightforward calculation shows that

$$w_{1-(\alpha/2),2n} \le \frac{2}{\theta}\sum_{i=1}^{n} X_i \le w_{\alpha/2,2n}$$
$$\iff \frac{2}{w_{\alpha/2,2n}}\sum_{i=1}^{n} X_i \le \theta \le \frac{2}{w_{1-(\alpha/2),2n}}\sum_{i=1}^{n} X_i.$$

By part (a) and the above probability calculation, we conclude that a $(1-\alpha)100\%$ confidence interval for θ is

$$\left(\frac{2}{w_{\alpha/2,2n}}\sum_{i=1}^{n} X_i, \frac{2}{w_{1-(\alpha/2),2n}}\sum_{i=1}^{n} X_i\right).$$

(c) From problem 4.2.2, we recall that $n = 20$ and $\sum_{i=1}^{20} x_i = 20\bar{x} = 2023$. To find a 95% confidence interval ($\alpha = 0.05$), we use a calculator to find the critical values $w_{0.975,40} = 24.433$ and $w_{0.025,40} = 59.342$. Substituting these values we obtain the 95% confidence interval for θ

$$(68.181, 165.595).$$

The approximate large sample confidence interval from problem 4.2.2 is $(54.953, 147.347)$. The length of this interval is 92.394 which is shorter than the length of the other interval, 116.888. On the other hand, we should keep in mind that $\overline{X}$ is only approximately normal, while $(2/\theta)\sum_{i=1}^{20} X_i$ has exactly a χ^2-distribution with $r = 40$ degrees of freedom. We should not be surprised if the confidence intervals based on problem 4.2.2 really have probability less than 95% of covering the mean θ.

Problem 4.2.5. In Exercise 4.1.2, the weights of 26 professional baseball pitchers were given. From the same data set, the weights of 33 professional baseball hitters (not pitchers) are given below. Assume that the data sets are independent of one another.

155	155	160	160	160	166	170	175	175	175	180	185	185
185	185	185	185	185	190	190	190	190	190	195	195	195
195	200	205	207	210	211	230						

Use expression (4.2.13) to find a 95% confidence interval for the difference in mean weights between the pitchers and the hitters. Which group (on the average) appears to be heavier? Why would this be so? (The sample means and variances for the weights of the pitchers and hitters are, respectively, Pitchers 201, 305.68 and Hitters 185.4, 298.13.)

Solution 4.2.5. Let $n_1 = 26$ and $n_2 = 33$ be the respective sample sizes for pitchers and hitters. Similarly, $\bar{x}_1 = 201, s_1^2 = 305.68$ and $\bar{x}_2 = 185.4, s_2^2 = 298.13$ denote the respective sample means and variances for the pitchers and hitters. From these values we compute the pooled sample variance

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(25)(305.68)+(32)(298.13)}{25+32} \doteq 301.4414$$

and the pooled sample standard deviation $s_p \doteq \sqrt{301.4414} \doteq 17.3621$. Our pivotal quantity has a student's t-distribution with $n_1 + n_2 - 2 = 57$ degrees of freedom. Using a calculator we find the critical value $t_{0.025,57} \doteq 2.002$ and we calculate

$$t_{0.025} s_p \sqrt{(1/n_1) + (1/n_2)} \doteq (2.002)(17.362)\sqrt{(1/26)+(1/33)} \doteq 9.12.$$

Adding and subtracting this value to the difference in the mean weights $\bar{x}_1 - \bar{x}_2 = 201 - 185.4 = 15.6$ gives the endpoints for the 95% confidence interval

$$(6.48, 24.72)$$

for the difference in the mean weights of pitchers and hitters. Based on the data from these samples it appears that, on average, pitchers are heavier than hitters. One possible explination for this difference could be that hitters are usually fielders and their daily training may burn more calories than that of a pitcher, who presumably does less running than most other fielders. Other explinations are certainly possible.

Problem 4.2.6. In the baseball data set discussed in the last exercise, it was found that out of the 59 baseball players, 15 were left-handed. Is this odd, since the proportion of left-handed males in America is about 11%? Answer by using (4.2.7) to construct a 95% approximate confidence interval for p, the proportion of left-handed baseball players.

Solution 4.2.6. With the sample size of $n = 59$, we calculate the sample proportion of left-handed players $\hat{p} = 15/59 \doteq 0.2542$ and the estimated standard error $\sqrt{\hat{p}(1-\hat{p})/n} \doteq 0.0567$. Using the critical value $z_{0.025} \doteq 1.9600$ we find $z_{0.025}\sqrt{\hat{p}(1-\hat{p})/n} \doteq 0.1111$, which gives an approximate

95% confidence interval for the proportion of left-handed baseball players (among pitchers and hitters)

$$(0.1431, 0.3653)\,.$$

Since 0.11, which is the proportion of left-handed males in America, lies outside this interval it seems that the proportion of left-handed professional baseball players (at least within pitchers and hitters) is greater than the proportion among all American males.

Problem 4.2.7. Let $\overline{X}$ be the mean of a random sample of size n from a distribution that is $N(\mu, 9)$. Find n such that $P(\overline{X}-1 < \mu < \overline{X}+1) = 0.90$, approximately.

Solution 4.2.7. We know that $\sqrt{n}\left(\overline{X}-\mu\right)/3$ has the standard normal distribution. We also know that $P(-1.645 < Z < 1.645) \doteq 0.90$, and a straightforward calculation shows that

$$\overline{X}-1 < \mu < \overline{X}+1$$
$$\iff -\frac{\sqrt{n}}{3} < \sqrt{n}\frac{\overline{X}-\mu}{3} < \frac{\sqrt{n}}{3}.$$

Thus, to make sure that the probability is at least 0.90 we should take $n = \left\lceil (3\cdot 1.645)^2 \right\rceil = \lceil 24.35 \rceil = 25$.

Problem 4.2.8. Let a random sample of size 17 from the normal distribution $N(\mu, \sigma^2)$ yield $\bar{x} = 4.7$ and $s^2 = 5.76$. Determine a 90% confidence interval for μ.

Solution 4.2.8. Since the sample comes from a normal distribution, we may use the confidence interval (4.2.3) on page 215 with $\alpha = 0.100$. Using a calculator we find the critical value $t_{0.050,16} \doteq 1.7459$ and we calculate $t_{0.050,16}s/\sqrt{n} \doteq (1.7459)\sqrt{5.76/17} \doteq 1.02$. Adding and subtracting this value to $\bar{x} = 4.7$ gives the endpoints of the 90% confidence interval for μ,

$$(3.68, 5.72).$$

Problem 4.2.9. Let $\overline{X}$ denote the mean of a random sample of size n from a distribution that has mean μ and variance $\sigma^2 = 10$. Find n so that the probability is aproximately 0.954 that the random interval $\left(\overline{X}-1/2, \overline{X}+1/2\right)$ includes μ.

Solution 4.2.9. Given the confidence coefficient of 0.954 we calculate $\alpha = 1-0.954 = 0.046$ and $\alpha/2 = 0.023$. Using a calculator we find the critical value $z_{0.023} \doteq 1.9954$. If n is sufficiently large, the the Central Limit Theorem applies to $\overline{X}$ and we may use the standard normal random variable Z as a pivotal quantity with $\overline{X} \approx \sqrt{10/n}Z + \mu$. From this we calculate that n should be large enough for $z_{0.023}\sqrt{10/n} \leq 1/2$. Therefore it suffices to take

$$n = \left\lceil (2\sqrt{10}\,z_{0.023})^2 \right\rceil = \left\lceil 40(1.9954)^2 \right\rceil = \left\lceil 159.3 \right\rceil = 160.$$

Problem 4.2.10. Let $X_1, X_2, \ldots, X_9$ be a random sample of size 9 from a distribution that is $N(\mu, \sigma^2)$.

(a) If σ is known, find the length of a 95% confidence interval for μ if this interval is based on the random variable $\sqrt{9}\left(\overline{X} - \mu\right)/\sigma$.

(b) If σ is unknown, find the expected value of the length of a 95% confidence interval for μ if this interval is based on the random variable $\sqrt{9}\left(\overline{X} - \mu\right)/S$. *Hint:* Write $E[S] = \left(\sigma/\sqrt{n-1}\right) E\left[\left((n-1)S^2/\sigma^2\right)^{1/2}\right]$.

(c) Compare these two answers.

Solution 4.2.10.

(a) Our pivotal random variable $\sqrt{9}\left(\overline{X} - \mu\right)/\sigma$ has the $N(0,1)$ distribution, and the endpoints of a 95% conficence interval for μ are $\overline{X} \pm z_{0.025}\sigma/\sqrt{9}$. Therefore, the length of this confidence interval is

$$2z_{0.025}\sigma/3 \doteq (2)(1.9600)\sigma/3 \doteq 1.3066\sigma.$$

(b) Our pivotal random variable $\sqrt{9}\left(\overline{X} - \mu\right)/S$ has the Student's t-distribution with $r = 8$ degrees of freedom. The endpoints of a 95% confidence interval for μ are $\overline{X} \pm t_{0.025,8}S/\sqrt{9}$. Therefore, the expected length of this confidence interval is

$$E[2t_{0.025,8}S/\sqrt{9}] = 2t_{0.025,8}E[S]/3 \doteq (2)(2.3060)E[S]/3 \doteq 1.5373E[S].$$

To evaluate $E[S]$ we use Student's Theorem (3.6.1 on page 193) to recall that

$$\frac{1}{\sigma^2}\sum_{i=1}^{9}\left(X_i - \overline{X}\right)^2 = \frac{8S^2}{\sigma^2}$$

has a χ^2-distribution with $r = 8$ degrees of freedom. Next we apply Theorem 3.3.1 on page 160 with $r = 8$ and $k = 1/2$ to see that

$$(\sqrt{8}/\sigma)E[S] = E\left[\sqrt{8S^2/\sigma^2}\right] = \frac{2^{1/2}\Gamma(\frac{8}{2} + \frac{1}{2})}{\Gamma(\frac{8}{2})} = \frac{\sqrt{2}\Gamma(9/2)}{\Gamma(4)} = \frac{35\sqrt{2\pi}}{32}.$$

Solving we find that $E[S] = 35\sqrt{\pi}\sigma/64 \doteq 0.9693\sigma$. Finally, the expected length of the 95% confidence interval is approximately

$$1.5373E[S] \doteq (1.5373)(0.9693)\sigma \doteq 1.4901\sigma.$$

(c) The ratio of these lengths is

$$\frac{\frac{2}{3}t_{0.025,8}E[S]}{\frac{2}{3}z_{0.025}\sigma} = \frac{t_{0.025,8}}{z_{0.025}} \cdot \frac{35\sqrt{\pi}}{64} \doteq 1.140447.$$

Thus, the confidence intervals for μ when σ is unknown are, on average, about 14% longer that the confidence intervals for μ when σ is known.

Problem 4.2.12. Let Y be $b(300, p)$. If the observed value of Y is $y = 75$, find an approximate 90% confidence interval for p.

Solution 4.2.12. With $n = 300$ and $y = 75$ we calculate our estimate $\hat{p} = 75/300 = 0.25$. Using a calculator we find the critical value $z_{0.050} \doteq 1.645$ and we calculate $z_{0.050}\sqrt{\hat{p}(1-\hat{p})/n} \doteq 1.645\sqrt{(0.25)(0.75)/300} \doteq 0.041$. Adding and subtracting this value to $\hat{p} = 0.25$ we obtain the 90% confidence interval for p (see equation 4.2.7 on page 217)

$$(0.209, 0.291)\,.$$

Problem 4.4.5. Let $Y_1 < Y_2 < Y_3 < Y_4$ be the order statistics of a random sample of size 4 from the distribution having pdf $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Find $P(Y_4 \geq 3)$.

Solution 4.4.5. Since $Y_4 = \max(X_1, X_2, X_3, X_4)$, we see that $Y_4 < 3$ if and only if $X_i < 3$ for all $i = 1, 2, 3, 4$. Since $P(X_i < 3) = 1 - e^{-3}$ and since the random variables X_1, X_2, X_3, X_4 are independent,

$$P(Y_4 < 3) = P(X_1 < 3, X_2 < 3, X_3 < 3, X_4 < 3) = (1 - e^{-3})^4$$

which implies that $P(Y_4 \geq 3) = 1 - (1 - e^{-3})^4$.