Confidence Intervals

Michael J. Keim

February 21, 2007

1 What is a confidence interval?

Consider the following definitions from Casella and Berger's *Statistical Inference, 2nd ed.* (2002), but don't spend too much time trying to parse their exact meaning at this point:

> An *interval estimate* of a real-valued parameter θ is any pair of functions $L(x_1, \ldots, x_n)$ and $U(x_1, \ldots, x_n)$ of a sample that satisfy $L(\boldsymbol{x}) \leq U(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{X}$. If $\boldsymbol{X} = \boldsymbol{x}$ is observed, the inference $L(\boldsymbol{x}) \leq \theta \leq U(\boldsymbol{x})$ is made. The random interval $[L(\boldsymbol{x}), U(\boldsymbol{x})]$ is called an *interval estimator.*
>
> ...
>
> For an *interval estimator* $[L(\boldsymbol{x}), U(\boldsymbol{x})]$ of a parameter θ, the *coverage probability* of $[L(\boldsymbol{x}), U(\boldsymbol{x})]$ is the probability that the random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ covers the true parameter, θ. In symbols, it is denoted by either $P_\theta(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})])$ or $P(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})]|\theta)$.
>
> For an *interval estimator* $[L(\boldsymbol{x}), U(\boldsymbol{x})]$ of a parameter θ, the *confidence coefficient* of $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is the infimum of the coverage probabilities, $\inf_\theta P_\theta(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})])$.
>
> ...
>
> Interval estimators, together with a measure of confidence (usually a confidence coefficient), are sometimes known as confidence intervals. (p.418-419)

Notice two things in these definitions: First, they are written in great generality indicating there are many flavors of interval estimators. The basic idea (from a frequentist point of view) is that we observe a realization (i.e. data) $\boldsymbol{x}$ of a random variable $\boldsymbol{X}$ whose distribution involves θ. From the data, we want an estimate of θ that reflects our uncertainty about the exact true value of θ. So, we use two functions that give lower, $L(\boldsymbol{x})$, and upper, $U(\boldsymbol{x})$, "bounds" on our estimate of θ, and then conclude (infer) with some level of confidence that θ is in fact between $L(\boldsymbol{x})$ and $U(\boldsymbol{x})$. The distance between $L(\boldsymbol{x})$ and $U(\boldsymbol{x})$ reflects how confident we are in our estimate of θ: wider intervals indicate greater uncertainty.

Second, a careful distinction is made between random variables (capital letters) and particular realizations of those random variables (lower-case letters). This illustrates the importance placed on how interval estimators are interpreted in the frequentist paradigm, which we will address later. These notes will only consider the frequentist paradigm, which is a failing because the Bayesian paradigm is important. Confidence intervals, however, are a frequentist idea, as opposed to *credible intervals*, a Bayesian analog not discussed here.

There will be many kinds of functions $L(\cdot)$ and $U(\cdot)$. Rather than mulling over the precise mathematical definitions, we will learn by doing. We consider here some of the most frequently used confidence intervals.

2 CI for the mean: Gaussian data

Here are some data:

$$\boldsymbol{x} = 11.9, 13.2, 14.6, 16.1, 16.3, 18.0, 18.6, 19.1, 19.5, 20.5$$

These have been given to Mike and Eli by a statistical genie, who tells them the data came from a Gaussian distribution with variance $\sigma^2 = 4$, but she refuses to tell them the mean μ. She promises a good-luck charm for the statistical theory qualifier in June if they can give her an interval estimate of μ that contains the true value.

Skeptical but having copious free time, Eli and Mike accept her challenge. They begin with her claim that the data are Gaussian and run the following code in `R`:

```
x<-c(11.9,13.2, 14.6, 16.1, 16.3, 18.0, 18.6, 19.1, 19.5, 20.5)
qqnorm(x)
qqline(x)
```

They accept the Gaussian assumption at this point. A point estimate for μ is straightforward: $\hat{\mu} = \bar{x} = 16.78$. Now they know by X Gaussian (with $\sigma = 2$ and $n = 10$) that:

$$\frac{\bar{x} - \mu}{2/\sqrt{10}} \sim \mathrm{N}(0, 1)$$

By symmetry of Gaussian distributions, they make the following probability statements:

$$\begin{aligned} 1 - \alpha &= P(-z_{1-\alpha/2} \le \frac{\bar{x} - \mu}{2/\sqrt{10}} \le z_{1-\alpha/2}) \\ &= P(-z_{1-\alpha/2} \cdot 2/\sqrt{10} \le \bar{x} - \mu \le z_{1-\alpha/2} \cdot 2/\sqrt{10}) \\ &= P(\bar{x} - z_{1-\alpha/2} \cdot 2/\sqrt{10} \le \mu \le \bar{x} + z_{1-\alpha/2} \cdot 2/\sqrt{10}) \\ &= P(L(\boldsymbol{x}) \le \mu \le U(\boldsymbol{x})). \end{aligned}$$

where $P(Z \le z_{1-\alpha/2}) = 1 - \alpha/2$ defines the standard Gaussian quantile $z_{1-\alpha/2}$.

At this point a vigorous debate erupts over the appropriate α level to choose. Setting α close to 0 will increase the our confidence that the interval includes the true parameter, but the interval will be wider as a result because $z_{1-\alpha/2}$ will grow larger. Dialogue ensues:

Eli: I'm sick of this genie and I don't need her help on the stat qualifier. I think we should choose $\alpha = .5$ just to tick her off.
Mike: I need all the help I can get, even from a genie, so I choose $\alpha = .001$

Plugging in to find appropriate quantiles, the following confidence intervals result:

```
# Eli's CI
x.bar<-mean(x)
z<-qnorm(.75)
x.bar-z*2/sqrt(10)
x.bar+z*2/sqrt(10)
```

Eli's 50% confidence interval: [16.35,17.2]

```
# Mike's CI x.bar<-mean(x)
z<-qnorm(.9995)
x.bar-z*2/sqrt(10)
x.bar+z*2/sqrt(10)
```

Mike's 99.9% confidence interval: [14.7,18.9].

The genie returns and reveals that the true mean is $\mu = 16$, which is contained in Mike's very conservative interval, but not in Eli's. The genie is offended by Eli's cheek but finds Mike's desperation off-putting. Eli leaves the room to pursue his own investigations, while the following dialogues ensues:

Genie: You're pathetic. Look how much bigger your interval was than Eli's just because you were afraid of not capturing μ. I'm going to give you some more Gaussian data, but this time I'm not telling you the variance and I'm telling you to use $\alpha = .05$ since it's a common convention.
Mike: < expletive deleted >

Here are the new data:

$$x = 5.0, 5.8, 6.3, 6.6, 6.7, 6.8, 6.9, 7.5, 8.7, 10.2$$

Not knowing the variance isn't a big deal because Gossett showed:

$$\frac{\bar{x} - \mu}{s/\sqrt{10}} \sim t_9$$

where t_η is a t-distribution with η degrees of freedom and s^2 is the sample variance. The probability statements given earlier are easily modified:

$$\begin{aligned}
1-\alpha &= P\left(-t_{\eta,1-\alpha/2} \le \frac{\bar{x}-\mu}{s/\sqrt{10}} \le t_{\eta,1-\alpha/2}\right) \\
&= P(-t_{\eta,1-\alpha/2} \cdot s/\sqrt{10} \le \bar{x}-\mu \le t_{\eta,1-\alpha/2} \cdot s/\sqrt{10}) \\
&= P(\bar{x} - t_{\eta,1-\alpha/2} \cdot s/\sqrt{10} \le \mu \le \bar{x} + t_{\eta,1-\alpha/2} \cdot s/\sqrt{10}).
\end{aligned}$$

where $P(t_\eta \le t_{\eta,1-\alpha/2}) = 1-\alpha/2$ defines the quantile $t_{\eta,1-\alpha/2}$.

Mike implements this in `R` with the following code:

```
x<-c(5.0,5.8,6.3,6.6,6.7,6.8,6.9,7.5,8.7,10.2)
qqnorm(x)
qqline(x) # maybe Gaussianity is a bad assumption?
x.bar<-mean(x)
t<-qt(.975,df=9)
x.bar-t*sd(x)/sqrt(10)
x.bar+t*sd(x)/sqrt(10
# Here is an easy shortcut
t.test(x,conf.level=.95)
```

Mike's 95% confidence interval: [6.0,8.1].

The genie reappears at this point to reveal that the true mean is $\mu = 7$, but I think we're all getting just a little tired of that genie by now. She is dispatched to sing back-up vocals for the Broadway production of *Alladin.*

Meanwhile, back at the Batcave, Eli is investigating the true nature of the confidence interval he calculated earlier. His internal monologue is vibrant and conducted in complete sentences.

Eli: What does it mean that the parameter is contained in this interval with probability $1-\alpha$? It appears that the upper and lower limits are random variables but give fixed quantities once we observe the data. Maybe a simulation will help me understand this better.

I will simulate 1000 data sets of size $n = 20$ from a N(3,5) distribution and for each data set I will calculate a 90% confidence interval. I will then record whether the true mean $\mu = 3$ is contained in that interval.

Here is Eli's `R` code:

```
num.sims<-1000
ci.lower<-rep(NA,num.sims) # store lower CI limits
ci.upper<-rep(NA,num.sims) # store upper CI limits
ci.success<-rep(0,num.sims) # store whether mu is in CI
```

```
mu<-3
sigma<-sqrt(5)
n<-20
for(i in 1:num.sims){
   x<-rnorm(n,mean=mu,sd=sigma)
   x.bar<-mean(x)
   z<-qnorm(.95)
   ci.lower[i]<-x.bar-z*sigma/sqrt(n)
   ci.upper[i]<-x.bar+z*sigma/sqrt(n)
   # record successful coverage as 1, failure as 0
   # note that storage vector was created full of 0's
   if(mu > ci.lower[i] & mu < ci.upper[i]) ci.success[i]<-1
} # close loop
sum(ci.success)/num.sims # is this close to 90%?
```

Eli wants to plot this, so he draws the first 100 confidence intervals, color-coding the successful intervals as black, and the failed intervals as red. Here is his code:

```
plot(x,type="n",xlim=c(1,5),ylim=c(0,100),
   ylab="",yaxt="n",
   xlab="confidence intervals, true mu = 3")
for(i in 1:100){
   if(ci.success[i]==1) segments(ci.lower[i],i,ci.upper[i],i,col="black")
   else segments(ci.lower[i],i,ci.upper[i],i,col="red")
}
abline(v=mu,col="blue")
```

and the results can be seen in figure 1. He has 8/100 confidence intervals fail in the figure so his success rate of 92% was close to the nominal rate of 90%. For all 1000 simulations, his success rate was 89.7%, which is quite close to 90%.

First exercise

Alter Eli's code to consider a N(3,2) distribution, keeping $n = 20$. Re-run the simulation and plot the first 100 confidence intervals. How do the interval widths compare? Can you explain why this happens? Now keep a N(3,2) distribution and re-run the code using $n = 40$. How do the widths compare? Can you explain why this happens?

Second exercise - extension of first

When we talked about t-tests, we identified the α-level, the sample size n and the process variance σ^2 as factors affecting the power of the test. Explain in words how each of these

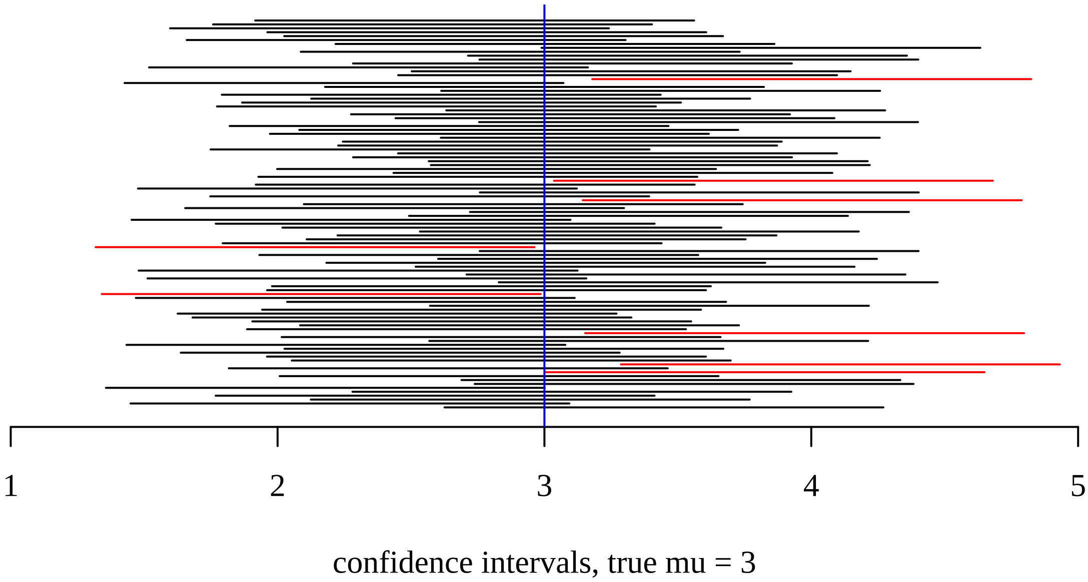

Figure 1: The first 100 confidence intervals from Eli's simulation study. Of these 100 intervals, 92 were successful at including $\mu = 3$.

factors impacts the width of a confidence interval for the mean. Address both the case where σ^2 is known and unkown.

How does this relate to Casella and Berger's definitions? We have $L(\boldsymbol{X}) = \overline{\boldsymbol{X}} - z_{1-\alpha/2} * \sigma/\sqrt{(n)}$ and $U(\boldsymbol{X}) = \overline{\boldsymbol{X}} + z_{1-\alpha/2} * \sigma/\sqrt{(n)}$ as our functions in the normal case. A common notation for this is:

$$\bar{x} \pm z_{1-\alpha/2} * s/\sqrt{n} \quad \text{or} \quad \bar{x} \pm z_{1-\alpha/2} * se(\bar{x})$$

where $se(\bar{x}) = s/\sqrt{n}$ is the standard error for the sample mean. We can set our confidence level by adjusting the α level and consequently changing $z_{1-\alpha/2}$. Once we observe $\boldsymbol{X} = \boldsymbol{x}$, then $L(\boldsymbol{x}) = \overline{\boldsymbol{x}} - z_{1-\alpha/2} * \sigma/\sqrt{(n)}$ and $U(\boldsymbol{x}) = \overline{\boldsymbol{x}} + z_{1-\alpha/2} * \sigma/\sqrt{(n)}$ become fixed quantities. The true mean μ is either in the interval or it isn't; it no longer makes sense to talk about the probability of it being in the interval. Instead, we talk about $1-\alpha\%$ confidence, because if the experiment were run over and over again, this *method* of calculating a confidence interval gives a successful interval $1-\alpha\%$ of the time. This is similar to "likelihood", which looks very much like a probability, but is merely a product of density functions. The different name is supposed to help you keep these straight. Note however, that notation like: $P(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})]|\theta)$ can make you think of θ as a random variable which, from the frequentist point of view, it is not.

As an aside, note that $se(\bar{x})$ and our estimate s of the error standard deviation are not the same thing. This distinction can take a while to get used to. Usually, standard errors are functions of s and n and they generaly decrease in n. Learn to keep these two ideas straight.

Third exercise

What are $L(\cdot)$ and $U(\cdot)$ for the case where the variance is unknown, i.e. the t-distribution case?

Fourth exercise

Explain why

$$[\bar{x} - z_{.99} \cdot s/n, \bar{x} + z_{.96} \cdot s/n]$$

is still a valid 95% confidence interval. How does this interval compare to the form given earlier? Are there any situations where we might want to alter that original form in this way, or some other way?

Fifth exercise

If we don't know the variance σ^2 but want to calculate a confidence interval for the mean, under what circumstances can we still use the normal quantiles instead of the t-quantiles?

3 CI for the regression slope: Gaussian data

Confidence intervals are not restricted to the mean. All we need is a distribution (either exact or approximate/asymptotic) for our estimator. We can then take the upper $1 - \alpha/2$ and lower $\alpha/2$ quantiles and we have a $1 - \alpha$ confidence interval for our parameter of interest. How well this confidence interval works will depend on whether our presumed model is a good description of the data, the quality of our estimator (i.e. its bias and variance properties) and the amount of data we have.

Here's a recent example in the regression context: Recall the Great QERM Insomnia Experiment in which the zone-out duration (ZOD) during a quantitative fisheries seminar was regressed on sleep the previous night. (This is the experiment that does not involve pie at all.) I'm going to change the postulated model notation slightly to avoid confusion about α as a parameter and α as our Type I error rate. Our model is

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \text{with} \quad \epsilon_i \sim \quad \text{iid} \quad \mathrm{N}(0, \sigma^2).$$

On page 15 of the lecture notes, we see the following about our estimated slope:

$$t_0 = \hat{\beta_1}/se(\hat{\beta_1}) \sim t_{n-2}$$

where β_1 is the slope paramter, $se(\hat{\beta_1})$ is the standard error for $\hat{\beta}$.

We have $\hat{\beta_1} = -2.12$, $se(\hat{\beta_1}) = \sqrt{0.147}$ and $n = 15$, hence a $1 = \alpha\%$ confidence interval for β_1 is given by:

$$\hat{\beta} \pm \sqrt{0.147} \cdot t_{13,1-\alpha/2}.$$

Using $\alpha = .05$ gives [-2.95,-1.29].

It is not a coincidence that our hypothesis test at the $\alpha = .05$ level rejected the null hypothesis that $\beta_1 = 0$ and our 95% confidence interval for β_1 does not contain 0! At the beginning of this example we stated that all we require for a confidence interval is an estimator and a distribution for that estimator (or at least the ability to determine the quantiles as required by our α level). This is akin to hypothesis testing requiring only a test statistic and a null distribution, or at least the ability to find critical values under the null based on α.

Here is relevant `R` code:

```
Sleep.df<-read.table(‘‘filedirectory\\Sleep.dat’’)
model<-lm(ZOD Sleep,data=Sleep.df)
confint(model,conf.level=.95)
```

Sixth exercise

Recall the shoe material experient used to demonstrate the paired-sample t-test. What question are we answering there? Write both in words and in mathematical notation. Calculate a sensible confidence interval based on the given data, assuming a Gaussian distribution for the error terms. How does this confidence interval reflect the testing answer given earlier? An answer to this particular exercise is given at the end of these notes. Please spend at least one full day trying before reading it.

4 CIs for MLEs

Thus far we have considered exact confidence intervals, i.e. those where an exact distribution was used to calculate quantiles used in the confidence interval. Maximum likelihood estimators (MLEs) give a good example of aymptotic confidence intervals, where we have an approximate distribution, and how we need to be careful when using them. For MLEs, if a rather long list of regularity conditions is satisfied, then we have the following results:

Theorem 1 (Casella and Berger p.472) *Let $X_1, \ldots, X_n$ be iid $f(x|\theta)$, let $\hat{\theta}$ denote the MLE of θ, and let $\tau(\theta)$ be a continuous function of θ (including the identity function). Then if the regularity conditions are satisfied:*

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \rightarrow_d \mathrm{N}(0, \nu(\theta)) \quad \text{or} \quad \tau(\hat{\theta}) \dot{\sim} N(\tau(\theta), \nu(\theta)/n)$$

where $\nu(\theta)$ is the Cramér-Rao lower bound.

The regularity conditions, in this case, are not something we can safely assume are always satisfied. They are listed on page 516 of Casella and Berger, and many of them *are* usually satisfied in typical models. They include (but are not limited too) conditions on interchanging differentiation and integration and a condition that the full parameter space Ω contains an open set ω of which the true parameter θ is an interior point, i.e. the true parameter can't be a boundary point of the parameter space. This actually arises as a practical problem in some environmental problems. Finally, note that even when all the regularity assumptions are met, this is an *approximate* result based on asymptotics, so the sample size needs to be large enough to justify the inference. Determining how large is "large enough" is not always straightforward task.

In this section we will consider one case each where

1. regularity conditions are met and asymptotics work well
   (Poisson, n big, θ big)
2. regularity conditions are met and asymptotics work poorly,
   (Poisson, n small, θ small) and
3. regularity conditions are not met
   (Uniform, estimating range).

Poisson, n big, θ big

Suppose we observe count data that are assumed to be 20 iid realizations of a Poisson process with unknown rate parameter θ.

$$x = 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 7, 10, 10$$

Seventh exercise

Show that the MLE for θ is $\bar{x}$ and the Fisher Information for a sample of size $n = 1$ is $1/\theta$. Why is the Cramér-Rao lower bound met by the Fisher Information here? (Aside: you should get in the habit of verifying why the Fisher information makes sense for a given model. Here, recall that the variance of a Poisson(θ) r.v. is θ, so the data becomes more variable as θ grows, hence the information in a sample of size 1 will be smaller.)

Based on the previous exercise and the fact that the Fisher information for n iid samples is n times the Fisher information for a sample of size 1., we should have

$$\hat{\theta} = \bar{x} \mathrel{\dot\sim} \mathrm{N}(\theta, \theta/n)$$

Hence our 95% confidence interval is given by:

$$\hat{\theta} \pm \sqrt{\hat{\theta}/n} \cdot z_{1-\alpha/2}$$

which, for the given data with $\hat{\theta} = 5.3$ and $n = 20$ gives [4.3,6.3]. In actuality, I used $\theta = 5$ to generate this data. Here is the `R` code:

```
x = c(2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 7, 10, 10)
n<-length(x)
theta.hat<-mean(x)
se.theta.hat<-sqrt(theta.hat/n)
theta.hat-se.theta.hat*qnorm(.975)
theta.hat+se.theta.hat*qnorm(.975)
```

So, how do we know that this actually works well? When $\theta = 5$ and $n = 20$, can we assume that the asymptotics are kicking in? Here is a simulation study that should allay our fears:

```
n<-20
theta<-5
num.sims<-1000
theta.hats<-rep(NA,num.sims)
ci.success<-rep(0,num.sims)
for(i in 1:num.sims){
   x<-rpois(n,theta)
   theta.hats[i]<-mean(x)
   ci.lower<-mean(x)-sqrt(mean(x)/n)*qnorm(.975)
   ci.upper<-mean(x)+sqrt(mean(x)/n)*qnorm(.975)
   if(ci.lower<theta & ci.upper>theta) ci.success[i]<-1
}
sum(ci.success)/num.sims # is this near 95%?
# are our estimators normally distributed?
qqnorm(theta.hats)
qqline(theta.hats)
# Looks good!
```

Note that our true parameter $\theta = 5$ can be contained in an open set in the interior of $\Omega = (0, \infty)$ and we can swap differentiation and integration as many times as required (the verification of these conditions is left as an exercise.)

Poisson, n small, θ small

Here's an example where we might be wary of the asympotic result. Suppose we have $n = 4$ and $\theta = 1/3$. Can we expect good performance of our asymptotic confidence intervals? Here is a simulation study (a simple modification of that given above) that may give us pause:

```
n<-4
theta<-1/3
num.sims<-1000
theta.hats<-rep(NA,num.sims)
ci.success<-rep(0,num.sims)
for(i in 1:num.sims){
   x<-rpois(n,theta)
   theta.hats[i]<-mean(x)
   ci.lower<-mean(x)-sqrt(mean(x)/n)*qnorm(.975)
   ci.upper<-mean(x)+sqrt(mean(x)/n)*qnorm(.975)
   if(ci.lower<theta & ci.upper>theta) ci.success[i]<-1
}
sum(ci.success)/num.sims # Well below 95%?
# are our estimators normally distributed?
qqnorm(theta.hats)
qqline(theta.hats)
# Blech!
```

There are two problems with the Gaussian approximation: the distibution of $\hat{\theta}$ is highly discete due to small n, and there is large mass at $\hat{\theta} = 0$ because of many 0's observed due to small θ. Again, however, all the regularity conditions are met; we just have an unrealistic sample size to assume an asymptotic result will hold.

Uniform

Finally, consider the example given on page 339 of Casella and Berger: We observe n values of $X \sim U(0, \theta)$ and want to estimate θ. Here we violate the assumption of swapping diffentiation and integration (se Casella and Berger p. 340 for proof), so the asymptotic result does not hold. Suppose we went ahead heedlessly and calculated a CI using the MLE $\hat{\theta} = \max(x_i)$ and the "false Fisher Information" given by the expected value of the squared log-likelihood $= n/\theta^2$. Then we'd expect

$$\hat{\theta} \dot{\sim} \mathrm{N}(\theta, \theta^2/n)$$

Here's a simulation study that shows this won't work even for large sample sizes (you can easily modify this from the Poisson case).

```
n<-30
theta<-5
num.sims<-1000
theta.hats<-rep(NA,num.sims)
ci.success<-rep(0,num.sims)
```

```
for(i in 1:num.sims){
   x<-runif(n,min=0,max=theta)
   theta.hats[i]<-max(x)
   ci.lower<-max(x)-max(x)/sqrt(n)*qnorm(.975)
   ci.upper<-max(x)+max(x)/sqrt(n)*qnorm(.975)
   if(ci.lower<theta & ci.upper>theta) ci.success[i]<-1
}
sum(ci.success)/num.sims # is this near 95%?
# are our estimators normally distributed?
qqnorm(theta.hats)
qqline(theta.hats)
# Pretty far off!
```

Don't be comforted by the fact that the coverage probability is 1 or near 1; this is just as bad as the coverage probability being too small as it indicates your confidence intervals are wider than they need to be.

All is not lost in the Uniform case. I've included some notes at the end related to a homework problem from STAT 513 involving estimation for the Uniform(0,θ) distribution that might be illuminating.

5 What you should know and be able to do

1. Be able to interpret a confidence interval and state its meaning in words.
   Suggested practice: Construct confidence intervals for the treatment means and their differences in the Great Pie Zone-Out ANOVA example and write down what they mean. Calculate a confidence interval for the difference in needle length based on a t-distribution for the white pine data given in the notes from the first/second week of class.

2. Given an estimator, its distribution (or density) and an α-level, calculate a confidence interval.
   Suggested practice: revisit the problem for estimating the upper-range for a Uniform (0,θ) from the 513 homework. The distribution of the given estimator is *not* based on maximum likelihood so it should perform well.

3. Derive the probability statements for confidence intervals for the mean in the Gaussian case when the variance is known or unknown. These are helpful derivations to help illustrate what the confidence level means.
   Suggested practice: write these out 3 times a day for a week.

4. Given a simple model, simulate data from that model to evaluate a confidence interval method. (Think of evaluating our Poisson model).
   Suggested practice: Normal approximation to the Binomial with small and large n. Also, simulate data to evaluate the estimator given in the 513 homework for the Uniform (0,θ) problem (this is done for you in Appendix B, but you can use it for practice).

5. Explain the difference between an exact and an asymptotic confidence interval, and use simulations to evaluate the reliability of an asymptotic confidence interval.

Please realize that this is not a comprehensive treatment of confidence intervals and how they are generated. In addition to skipping Bayesian credible intervals, these notes do not mention bootstrap methods for generating confidence intervals, which is an important topic in its own right.

Appendix A. Answer to sixth exercise

Research question: Does material A wear out at the same rate as material B? In mathematical terms, we are testing $H_0 : d = \mu_A - \mu_B = 0$ vs. $H_1 : d = \mu_A - \mu_B \neq 0$. We have paired data here, so we want to take advantage of the blocking effect to reduce variability. Hence we consider $Y_i = A_i - B_i$ where A_i is the amount of wear for material A in individual i. Our statistical model is:

$$Y_i = \mu_A - \mu_B + \epsilon_i \quad \text{with} \quad \epsilon_i \sim \quad \text{iid} \;\; \text{N}(0, \sigma^2),$$

which can also be written as:

$$Y_i \sim \text{N}(\mu_A - \mu_B, \sigma^2).$$

Because we have normal data with small $n = 10$ and unknown variance, we'll use a t_η distribution. Our parameter of interest here isn't μ_A or μ_B; it is the difference $d = \mu_A - \mu_B$. Our $1 - \alpha\%$ confidence interval is given by:

$$\hat{d} \pm se(\hat{d}) \cdot t_{9,1-\alpha/2}$$

where $\hat{d} = \bar{y} = 0.41$ and $se(\hat{d}) = s/\sqrt{n} = .387$ give a 95% CI of [0.13,0.69]. Our paired sample test rejected the null hypothesis that $d = 0$, which agrees with our confidence interval excluding this value.

Here is some `R` code for analyzing this data, which includes generating a confidence interval. Note that you can set the α level desired in the options for the `t.test()` function.

```
A<-c(14,8.8,11.2,14.2,11.8,6.4,9.8,11.3,9.3,13.6)
B<-c(13.2,8.2,10.9,14.3,10.7,6.6,9.5,10.8,8.8,13.3)
t.test(A,B,paired=T)
```

A naive analysis of this data that ignored the blocking would look like this:

```
t.test(A,B,var.equal=T)
```

This test fails to reject the null and corresponds to generating two separate confidence intervals for μ_A and μ_B and seeing if they overlap. These intervals are given by:

$$\hat{\mu}_A \pm se(\hat{\mu}_A) \cdot t_{9,1-\alpha/2}$$

and likewise for μ_B. Here is the relevant `R` code:

```
a.hat<-mean(A)
b.hat<-mean(B)
se.a<-sd(A)/sqrt(10)
se.b<-sd(B)/sqrt(10)
# CI for A
a.hat-se.a*qt(.975,df=9)
```

```
a.hat+se.a*qt(.975,df=9)
# CI for B
b.hat-se.b*qt(.975,df=9)
b.hat+se.b*qt(.975,df=9)
```

Note that these confidence intervals overlap. This is a special case related to a point made by Gerald Van Belle on page 39 of his *Statistical Rules of Thumb*:

> Confidence intervals associated with statistics can overlap as much as 29% and the statistics can still be significanly different.

In the shoe data case, this overlap arises from ignoring the blocking effect of the paired data. In general, however, this can happen when we calculate confidence intervals separately for different means (e.g. treatment means in ANOVA) instead of confidence intervals for differences in the means. The take-home message is: Don't be glib about comparing confidence intervals instead of testing; you might overlook a significant difference of interest.

Appendix B. More on Uniform problem

Suppose $\boldsymbol{X} = X_1, \ldots, X_n$ consists of iid Uniform (0,θ) random variables and we want to estimate θ from a sample $x_1, \ldots, x_n$ and calculate a 1-α confidence interval. We've already established that maximum likelihood is not helpful here.

A sensible beginning is to consider $X_{(n)} = \max\{X_1, \ldots, X_n\}$. (I'm assuming Galen's notation for order statistics remains consistent over the years.) What happens if we use $\hat{\theta} = T(\boldsymbol{X}) = X_{(n)}$ as our estimator of θ? We know

$$P(T \le t) = \left(\frac{t}{\theta}\right)^n = F(t) \;\Rightarrow\; f(t) = \frac{dF(t)}{dt} = \frac{nt^{n-1}}{\theta^n}$$

give the distribution and density functions. We can use the density to evaluate the bias properties of our proposed estimator $\hat{\theta}$:

$$\mathrm{E}\,\hat{\theta} = \mathrm{E}\,T \int_0^\theta \frac{nt^{n-1}}{\theta^n} t\,dt = \left[\frac{n}{(n+1)}\frac{t^{n+1}}{\theta^n}\right]_0^\theta = \frac{n}{n+1}\,\theta$$

which suggests

$$\tilde{\theta} = \frac{n+1}{n} T$$

as a fix-up because our first-cut estimator is biased. We will continue to work with T directly as $\tilde{\theta}$ is a simple function of T. Modifications can be made later.

The problem remains of finding a confidence interval for θ; we want a and b satisfying

$$P(a \le \theta \le b) = 1 - \alpha.$$

We have a density for T that involves θ so that is where we should begin. This step requires a leap of faith: even though our beginning statement

$$P(a' \leq T \leq b') = 1 - \alpha$$

does not include θ explicitly, it seems likely that if we muddle along for a bit we'll begin to sniff out a θ somewhere. So, to find a' and b', we need to solve:

$$\int_0^{a'} f(t)\,dt = \alpha/2 \quad \text{and} \quad \int_0^{b'} f(t)\,dt = 1 - \alpha/2$$

to find our first bounds. It is left as an exercise to show that the solutions are $a' = \theta \cdot (\alpha/2)^{1/n}$ and $b' = \theta \cdot (1-\alpha/2)^{1/n}$. We've found our θs; now we just move things around:

$$\begin{aligned} 1 - \alpha &= P\left(\theta \cdot (\alpha/2)^{1/n} \leq T \leq \theta \cdot (1-\alpha/2)^{1/n}\right) \\ &= P\left(T \cdot (1-\alpha/2)^{-1/n} \leq \theta \leq T \cdot (\alpha/2)^{-1/n}\right) \end{aligned}$$

So we have found $a = T \cdot (1-\alpha/2)^{-1/n}$ and $b = T \cdot (\alpha/2)^{-1/n}$ to define our confidence interval. Let's evaluate this with a simulation:

```
n<-10
alpha<-.05
theta<-5
num.sims<-1000
ci.success<-rep(0,num.sims)
for(i in 1:num.sims){
x<-runif(n,min=0,max=theta)
T<-max(x)
ci.lower<-T*(1-alpha/2)^(-1/n)
ci.upper<-T*(alpha/2)^(-1/n)
if(ci.lower < theta & ci.upper > theta) ci.success[i]<-1
}
# Is this near 95%?
sum(ci.success)/num.sims
```

We see that this exact confidence interval works quite well. It is possible to modify our expressions of a and b to put them in terms of $\tilde{\theta}$ instead of T, but it actually looks messier. Notice that this confidence interval is not symmetric and does not have the form $T\pm$ something. This makes sense because $T < \theta$ with probability 1, so none of our estimators should be less than T. The take-home lessons here are don't assume every confidence interval will have a ± form to it (or be symmetric) and just because maximum likelihood isn't working doesn't mean you have no recourse to other methods.