Statistical_Inference

Laura Bennett

June 29, 2017

Statistical Inference for Data Science: Chapter 7 Exercises

1. I simulate 1,000,000 standard normals. The LLN says that their sample average must be close to?

- **As they are standard normals, each will have an average of $\mu = 0$; the LLN states that their sample average must then be equal to 0 as well.**

2. About what is the probability of getting 45 or fewer heads out of 100 flips of a **fair coin**? (Use the CLT, not the exact binomial calculation)

- **According to the CLT, as you collect more sample averages, the distribution you are measuring will eventually converge to a normal distribution. In this case, the distribution will have a mean of $\mu = 50$; the standard deviation of the coin was originally $\sqrt{np\,(1-p)} = \sqrt{100 \cdot 0.5\,(0.5)} = 5$ as per the calculation for a binomial distribution, but the standard deviation for the assumed normal distribution will be $\frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{100}} = 0.5$. Now, using a normal distribution, we can calculate the probability of getting 45 or fewer heads:**

```
r   pnorm(45, mean=50, sd=0.5, lower.tail=TRUE)
```

```
## [1] 7.619853e-24
```

3. Consider the father.son data. Using the CLT and assuming that the fathers are a random sample from a population of interest, what is a 95% confidence mean height in inches?

- **The 95% confidence interval for a normal distribution may be calculated as $\bar{X} \pm \frac{2\sigma}{\sqrt{n}}$. Thus, we may calculate the confidence interval for this dataset as:**

```
r   library(UsingR)
```

```
## Loading required package: MASS

## Loading required package: HistData

## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##    ## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':   ##    ##     format.pval,
round.POSIXt, trunc.POSIXt, units

##    ## Attaching package: 'UsingR'

## The following object is masked from 'package:survival':   ##    ##     cancer
```

```
r   data(father.son)   xbar <- mean(father.son[['fheight']])   sd <- sd(father.son[['fheight']])
n <- length(father.son[['fheight']])   upper <- (xbar + (2*sd)/sqrt(n))   lower <- (xbar
- (2*sd)/sqrt(n))   print(paste('(',lower,',',upper,')'))
```

```
## [1] "( 67.5198946040033 , 67.8542991251247 )"
```

4. The goal of a confidence interval having coverage 95% is to imply what?

- **The probability that the sample mean is in the interval is 95%**

5. The rate of search entries into a web site was 10 per minute when monitoring for an hour. Use R to calculate the exact Poisson interval for the rate of events per minute.

- **We may use the r function poisson.test() to calculate the exact Poisson interval:**

```
r   poisson.test(600,T=60)
```

```
##    ##    Exact Poisson test   ##    ## data:  600 time base: 60   ## number of events
= 600, time base = 60, p-value < 2.2e-16   ## alternative hypothesis: true event rate is
not equal to 1   ## 95 percent confidence interval:   ##   9.215749 10.833152   ## sample
estimates:   ## event rate    ##         10
```

6. Consider a uniform distribution. If we were to sample 100 draws from a uniform distribution (which has a mean of 0.5 and a variance of 1/12) and take their mean $\bar{X}$, what is the approximate probability of getting as large as 0.51 or larger?

- **This is similar to problem #2, as we are modeling a uniform distribution with a normal distribution with mean $\mu = 0.5$ and standard deviation $\sigma = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{\frac{1}{12}}{100}}$**

```
r   sd <- sqrt((1/12)/100)   pnorm(0.51,0.5,sd,lower.tail = FALSE)
```

```
## [1] 0.3645172
```