Stats 135, Fall 2019
Lecture 7, Friday, 9/13/2019

CLASS ANNOUNCEMENTS: RStudio Lab demo today in-class, and Quiz today in lab section.

Stat 135 lec 7

Last time

Sec 4.6 Delta Method

Thm (δ-method)

$X_1, \dots, X_n$ iid mean μ, $SD = \sigma$

g smooth around μ, $g'(\mu) \neq 0$

$$Var(g(\bar{X})) \approx (g'(\mu))^2 \frac{\sigma^2}{n}$$

ex *SZ chap 8

Let $X_1, \dots, X_n$ be iid RVs w/ density $f(x|\theta) = (\theta+1)x^\theta$, $0 \le x \le 1$

Find $\hat{\theta}$ and use the δ-method to approx $SE(\hat{\theta})$.

Find $\hat{\theta}$:

$$E(X) = \int_0^1 (\theta+1)x^{\theta+1}\,dx = \frac{(\theta+1)}{(\theta+2)} x^{\theta+2} \Big|_0^1 = \frac{\theta+1}{\theta+2}$$

Note
$$E(X^2) = \frac{\theta+1}{\theta+3}$$

$$\mu = \frac{\theta+1}{\theta+2} \Rightarrow \mu\theta + 2\mu = \theta + 1$$

$$\Rightarrow \mu\theta - \theta = 1 - 2\mu$$

$$\Rightarrow \theta(\mu - 1) = 1 - 2\mu$$

$$\theta = \frac{1-2\mu}{\mu-1}$$

$$\Rightarrow \boxed{\hat{\theta} = \frac{1-2\bar{X}}{\bar{X}-1}}$$

Use δ-method to approx $SE(\hat{\theta})$! $\quad Var(g(\bar{X})) \approx (g'(\mu))^2 \frac{\sigma^2}{n}$

$$g(x) = \frac{1-2x}{x-1}$$

$$g'(x) = \frac{1}{(x-1)^2}$$

$$\Rightarrow g'(\mu) = \frac{1}{(\mu-1)^2}$$

$$\sigma^2 = E(X^2) - E(X)^2 = \left(\frac{\theta+1}{\theta+3}\right) - \left(\frac{\theta+1}{\theta+2}\right)^2$$

$$\Rightarrow Var(\hat{\theta}) \approx (g'(\mu))^2 \frac{\sigma^2}{n}$$

$$= \frac{1}{(\mu-1)^4} \frac{\sigma^2}{n}$$

$$= \frac{1}{\left[\left(\frac{\theta+1}{\theta+2}\right)-1\right]^4} \frac{\left[\left(\frac{\theta+1}{\theta+3}\right) - \left(\frac{\theta+1}{\theta+2}\right)^2\right]}{n}$$

Now plug in $\hat{\theta} = \frac{1-2\bar{X}}{\bar{X}-1}$ for θ

This you can calculate from the data.

<u>Today</u> sec 8.5

① - examples of MLE — by hand / in R

② - property of MLE ~ equivalence

① <u>Sec 8.5</u> Maximum Likelihood estimator (MLE)

Suppose RV $X_1, .., X_n$ have joint density $f(x_1, .., x_n | \theta)$.

Given observed data values

$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ fixed numbers.

We form a likelihood function

$lik(\theta) = f(x_1, .., x_n | \theta)$ ← function of θ for fixed $x_1, .., x_n$

The max value of $lik(\theta)$ represents the value of θ that maximizes the likelihood of observing your data.

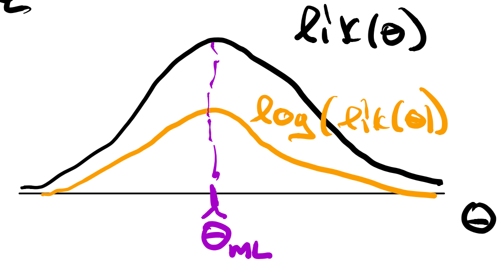
If x_i are iid

$$lik(\theta) = \prod_{i=1}^{n} f(x_i | \theta)$$

Taking the derivative of $lik(\theta)$ will involve the product rule. So we take the log of the likelihood so that we are finding the derivative of a sum.

Note that if $\hat{\theta}_{ML}$ is the value of θ that maximizes $lik(\theta)$, then $\hat{\theta}_{ML}$ also is the value of θ that maximizes $\log(lik(\theta))$ since log is a monotonically increasing function

Picture



$$\ell(\theta) = \log(\text{lik}(\theta)) = \sum_{i=1}^{n} \log(f(x_i|\theta))$$

To maximize $\ell(\theta)$ we take the derivative, set equal to zero, and solve for θ

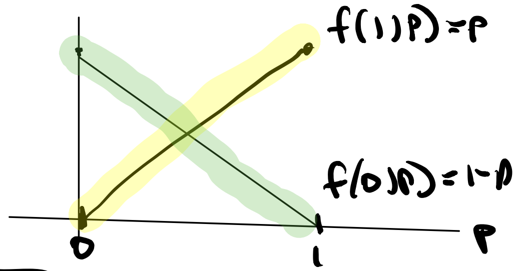
$\hat{\theta}_{ML}$ is the θ that solves $\ell'(\theta)=0$

ex $X_1 \sim \text{Ber}(p)$ (have single observation 0, 1)

$$\text{lik}(p) = f(x|p) = p^x(1-p)^{1-x}$$

$f(1|p)=p$

$f(0|p)=1-p$



$$\boxed{\hat{p}_{ML} = x}$$

ex $X_1, \dots, X_n$ iid $\exp(\lambda)$

$$f(x|\lambda) = \lambda e^{-\lambda x}$$

$$\text{lik}(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$\ell(\lambda) = \log(\lambda^n) - \lambda \sum_{i=1}^{n} x_i$$

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \implies \hat{\lambda}_{ML} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i} = \boxed{\frac{1}{\bar{X}}}$$

$\hat{\lambda}_{ML}$

ex # 16b

Consider an iid sample RV w/ density

$f(x|\sigma) = \frac{1}{2\sigma} \exp\left(\frac{-|x|}{\sigma}\right)$

Find $\hat{\sigma}_{ML}$

$$lik(\sigma) = \prod_{i=1}^{n} f(x_i|\sigma) = \left(\frac{1}{2\sigma}\right)^n \exp\left(-\frac{1}{\sigma}\left(|x_1| + \cdots + |x_n|\right)\right)$$

$$\ell(\sigma) = n\log\left(\frac{1}{2\sigma}\right) - \frac{1}{\sigma}\sum_{i=1}^{n}|x_i|$$

$$= -n\log 2 - n\log\sigma - \frac{1}{\sigma}\sum_{i=1}^{n}|x_i|$$

$$\ell'(\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n}|x_i| = 0$$

$$\Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}|x_i| = \frac{n}{\sigma}$$

$$\Rightarrow \boxed{\hat{\sigma}_{ML} = \frac{\sum_{i=1}^{n}|x_i|}{n}}$$

found $\hat{\sigma}_{MM} = \sqrt{\frac{\sum x_i^2}{n}}$

EXAMPLE C *Gamma Distribution*

Since the density function of a gamma distribution is

$$f(x|\alpha, \lambda) = \frac{1}{\Gamma(\alpha)}\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}, \qquad 0 \leq x < \infty$$

the log likelihood of an i.i.d. sample, $X_i, \ldots, X_n$, is

$$l(\alpha, \lambda) = \sum_{i=1}^{n}[\alpha \log \lambda + (\alpha - 1)\log X_i - \lambda X_i - \log \Gamma(\alpha)]$$

formula we want to optimize

$$= n\alpha \log \lambda + (\alpha - 1)\sum_{i=1}^{n} \log X_i - \lambda \sum_{i=1}^{n} X_i - n \log \Gamma(\alpha)$$

The partial derivatives are

$$\frac{\partial l}{\partial \alpha} = n \log \lambda + \sum_{i=1}^{n} \log X_i - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$\frac{\partial l}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i$$

Setting the second partial equal to zero, we find

$$\hat{\lambda} = \frac{n\hat{\alpha}}{\sum_{i=1}^{n} X_i} = \frac{\hat{\alpha}}{\overline{X}}$$

But when this solution is substituted into the equation for the first partial, we obtain a nonlinear equation for the mle of α:

$$n \log \hat{\alpha} - n \log \overline{X} + \sum_{i=1}^{n} \log X_i - n\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = 0$$

see in-class exercise on bcourses.

This equation cannot be solved in closed form; an iterative method for finding the roots has to be employed. To start the iterative procedure, we could use the initial value obtained by the method of moments.

Computationally finding MLE for Gamma paramaters

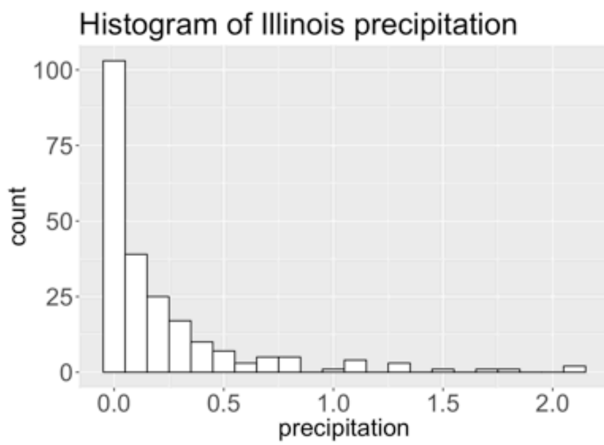
We take the data from 227 storms in Illinois between 1960,and 1964

```
#here is the precip data

precip <- c(0.020, 0.001, 0.001, 0.120, 0.080, 0.420, 1.720, 0.050, 0.010, 0.010,0.003,0.001,0.
003, 0.270, 0.001, 0.060, 0.050, 2.130, 0.040, 1.100, 0.020, 0.001,0.140, 0.080, 0.210, 0.070,
0.320, 0.240, 0.290, 0.001, 0.290, 1.130, 0.003,
0.010, 0.190, 0.002, 0.010, 0.040, 0.002,0.070, 0.450, 0.010, 0.180, 0.670, 0.003, 0.010, 0.040
, 0.002, 0.490, 0.020, 0.020, 0.340, 0.140, 0.370, 0.330,
0.330, 0.350, 0.010, 0.500, 0.760, 1.060, 0.002 ,0.060, 0.160 ,0.270, 0.250,
0.290, 0.020, 0.050, 0.460, 0.070, 0.410, 0.020, 0.080, 0.210, 0.010, 0.440,
0.020, 0.050 ,0.110, 1.500, 0.003, 0.180 ,0.010, 0.002, 0.240, 0.010, 0.750,
0.010, 0.140 ,0.130, 0.010, 0.010 ,0.270, 0.450, 1.780, 0.250, 0.240, 0.004,
0.210, 0.170, 0.830, 0.150, 0.030, 0.030, 0.500 ,0.040, 0.090 ,0.040 ,0.060,
0.060, 0.120, 0.003, 0.003, 0.400, 0.020, 0.510, 0.003, 0.020, 0.020, 0.020,
0.010, 0.001, 0.140, 0.001, 0.100, 0.010 ,1.090 ,0.010, 0.002, 0.001, 0.840,
0.030, 0.350, 0.070, 0.001, 0.002, 0.002, 0.200, 0.060, 0.140, 0.010, 0.020,
0.020, 0.002, 0.001, 0.550, 0.130, 0.190, 2.100, 0.090, 0.350, 0.790, 0.320,
1.350,0.170,0.020,0.002, 0.010,0.250,0.230,0.170,0.010,0.020,0.001,
0.010,0.020,0.110,0.210,1.260,0.010,0.730,0.100,0.090,0.007, 0.360,
0.770,0.210,1.270,0.070,0.080,0.160,0.260,0.010,0.230,0.080,0.020,
0.010,0.290,0.010,0.010,0.070,0.400,0.002, 0.003, 0.010,0.090,0.160,
0.040,0.270,0.730,0.410,0.030,0.120,0.030,1.040,0.060,0.090,0.730,
0.040,0.160,0.590,0.003, 0.002, 0.020,0.004, 0.010,0.001, 0.060,0.620,
0.010,0.520,0.110,0.003, 0.600,0.002, 0.050)

rain <- data.frame(precip)

#make histogram
rain %>% ggplot(aes(x=precip)) +
  geom_histogram(binwidth = .1,col="black",fill="white") +
  labs(title="Histogram of Illinois precipitation",x="precipitation",y="count") + theme(
    axis.text = element_text(size = 20),
    plot.title = element_text(size = 25),
    axis.title=element_text(size=20))
```



First we find the MLE estimate of α and λ. This will be the initial value of our optimization.

```
alpha_hat_MOM <- mean(precip)^2/(mean(precip^2)-mean(precip)^2)
alpha_hat_MOM
```

```
## [1] 0.3779155
```

```
lambda_hat_MOM <- mean(precip)/(mean(precip^2)-mean(precip)^2)
lambda_hat_MOM
```

```
## [1] 1.684175
```

```
fun <-  function(para, x) {
    alpha <-  para[1]
    lambda <-  para[2]
    -sum(alpha * log(lambda) + (alpha - 1) * log(x) - lambda * x - log(gamma(alpha)))
}

mle <-  optim(par = c(alpha_hat_MOM, lambda_hat_MOM),  fn = fun, x=precip)
#par=Initial values for the parameters to be optimized over.
#fn=    A function to be minimized (or maximized), with first argument the vector of parameters
over which minimization is to take place. It should return a scalar result.
#x= further arguements to pass to fn.
#method=the method to be used. The default is a method that only uses values of the function. I
t is very robust to initial values but slow. My favorite is method="BFGS" which is Newton's met
hod which use a combination of the function value and its derivative. It is very fast. Type ?op
tim in the console to learn more.

#Note: the function optim returns a list including "par" the best set of parameters found.


alpha.mle <-  mle$par[1]
lambda.mle <-  mle$par[2]
alpha.mle
```

method = "BFGS"
→ Newton's method.

```
## [1] 0.4408386
```

```
lambda.mle
```

```
## [1] 1.964841
```

Note that the MLE and MOM estimates give different values.