Math 275 Project

Math 275, St Clair

Introduction:

The problem that we are exploring in this project is understanding estimation options for a "half normal" distribution. We are using distance data from Bergen lake. The idea is that the further out a mussle is from the transect, the lower its probability of detection. This is what the "half normal" distribution models.

Methods:

Derivation Results:

More information about the derivation can be found in the appendix. **Sampling Distribution of $\hat{\sigma}^2_{MLE}$:**

$$\hat{\sigma}^2_{MLE} \sim Gamma(\frac{n}{2}, \frac{n}{2\sigma^2})$$

MoM Estimator $\hat{\sigma}^2_{MoM}$ for σ^2:

$$\sigma^2 = \frac{\pi \bar{X}^2}{2}$$

Bias of MLE and MoM estimators of σ^2:

$$Bias(\sigma^2_{MLE}) = 0 \qquad Bias(\sigma^2_{MoM}) = \frac{\pi\sigma^2 - 2\sigma^2}{2n}$$

Find the exact 95% confidence interval of σ^2:

$$L(x) = \frac{\bar{X}^2}{q_{0.975}} \qquad U(x) = \frac{\bar{X}^2}{q_{0.025}}$$

q = quantiles from $Gamma(\frac{n}{2}, \frac{n}{2})$

Simulation:

```
> summary(lake$distance)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.1800  0.3500  0.3851  0.5400  0.9800
> sd(lake$distance)

[1] 0.2449113
```

```
> distance <- lake$distance

> simsize <- 100 # simulations being run
> n <- 30 #sample size
> sigma <- 0.5 # value of sigma for ML estimate
> N <- 1000 #bootstrap
> mle_boot_lower <- mle_lower <- rep(NA,simsize)
> mle_boot_upper <- mle_upper <- rep(NA,simsize)
> mom_boot_lower <- mle_lower <- rep(NA,simsize)
> mom_boot_upper <- mle_upper <- rep(NA,simsize)
> x_boot_MLE <- rep(NA,N)
> x_boot_MOM <- rep(NA,N)
>
> for (i in 1:simsize)
+ {
+   x <- abs(rnorm(n,0,sigma))   # generating half normal random sample, repeat this
+   est_mle <- sum(x^2)/n # MLE formula
+   est_mom <- (pi*mean(x)^2)/2 # MoM formula
+
+   mle_lower[i] <- (sum(x^2)/n)/qgamma(0.975, n/2, n/2) # based on exact MLE CI (part d)
+   mle_upper[i] <- (sum(x^2)/n)/qgamma(0.0275, n/2, n/2)# based on exact MLE CI (part d)
+
+   for (j in 1:N)
+   {
+     x_boot <- sample(x, replace=TRUE)
+     x_boot_MLE[j] <- sum(x_boot^2)/n
+     x_boot_MOM[j] <- (pi*mean(x_boot)^2)/2
+   }
+   # CI based on bootstrapping the MLE
+   mle_boot_lower[i] <- (mean(x)-quantile(x_boot_MLE,0.975)*sqrt(var(x)/n))
+   mle_boot_upper[i] <- (mean(x)-quantile(x_boot_MLE,0.025)*sqrt(var(x)/n))
+   # CI based on bootstrapping the MoM
+   mom_boot_lower[i] <- (mean(x)-quantile(x_boot_MOM,0.975)*sqrt(var(x)/n))
+   mom_boot_upper[i] <- (mean(x)-quantile(x_boot_MOM,0.025)*sqrt(var(x)/n))
+ }
>
> mean(mle_boot_lower)

[1] 0.3784905

> mean(mle_boot_upper)

[1] 0.3918117

> mean(mom_boot_lower)

[1] 0.3767209

> mean(mom_boot_upper)

[1] 0.3918915

> mle_bias <- mean(est_mle) - sigma^2
> perc_bias <- 100*mle_bias/sigma^2   # bias as a percentage of sigma^2
> mse <- mean((est_mle - sigma^2)^2) # mean of squared deviations (est - truth)^2
>
> mom_bias <- mean(est_mom) - sigma^2
```

```
> perc_bias <- 100*mom_bias/sigma^2   # bias as a percentage of sigma^2
> mse <- mean((est_mom - sigma^2)^2) # mean of squared deviations (est - truth)^2

> mle_cover <- mean((mle_boot_lower <= sigma^2) & (sigma^2 <= mle_boot_upper))
> mle_mean.length <- mean(mean(mle_boot_upper) - mean(mle_boot_lower))
>
> mom_cover <- mean((mom_boot_lower <= sigma^2) & (sigma^2 <= mom_boot_upper))
> mom_mean.length <- mean(mean(mom_boot_upper) - mean(mom_boot_lower))
```

Estimation for Bergen

```
> mle_berg <- (sum(distance^2))/length(distance)
> mle_berg_lower <- sum(distance^2)/qgamma(0.975, shape = length(distance)/2,rate = 1/2)
> mle_berg_upper <- sum(distance^2)/qgamma(0.025, shape = length(distance)/2,rate = 1/2)
>
> #Distance of 0.5 meters
> #lower
> exp(-(0.5^2)/(2*mle_berg_lower^2))
```

```
[1] 0.00546522
```

```
> #upper
> exp(-(0.5^2)/(2*mle_berg_upper^2))
```

```
[1] 0.2320951
```

```
> #p value
> exp(-(0.5^2)/(2*mle_berg^2))
```

```
[1] 0.05489094
```

```
> #Distance of 1m
> #lower
> exp(-(1^2)/(2*mle_berg_lower^2))
```

```
[1] 8.921349e-10
```

```
> #upper
> exp(-(1^2)/(2*mle_berg_upper^2))
```

```
[1] 0.002901778
```

```
> #p value
> exp(-(1^2)/(2*mle_berg^2))
```

```
[1] 9.078264e-06
```

Results:

After running these simulations we opted for a maximum likelihood estimator. We felt that based on the setting of the problem as well as general distributions it was better to have that. We are also operating based off the sample data so for future references this may be difficult to achieve.

Discussion:

The difficulty in this model for us in particular was the lack of understanding with the use of it. While it may be particular to wildlife tracking or observation, it was relatively novel for us and we struggled with bigger picture application. However the statistical numbers associated with this were a little bit clearer. We evaluated the different estimators and the Bergen estimates that it produced.

Appendix:

Derivations:

Sampling Distribution of $\hat{\sigma}^2_{MLE}$:

$$X \sim HalfNormal(\sigma^2) \quad W = X^2 \quad F_W(w) = P(W \leq w) = P(X^2 \leq w) = P(X \leq \sqrt{w}) = F_X(\sqrt{w})$$

We don't have to worry about $\sqrt{w}$ being negative.

PDF:

$$f_W(w) = \frac{d}{dw}F_X(\sqrt{w}) = f_X(\sqrt{w}) * \frac{d}{dw}\sqrt{w} = \frac{1}{\sqrt{2\pi w}\sigma}e^{-w/2\sigma^2} \implies Gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$$

$$X_i^2 = W_i \sim Gamma(\frac{1}{2}, \frac{1}{2\sigma^2}) \quad \hat{\sigma}^2_{MLE} = \frac{\sum X_i^2}{n} = \frac{\sum W_i}{n} \implies Gamma(n*\frac{1}{2}, \frac{\frac{1}{2\sigma^2}}{\frac{1}{n}}) \implies Gamma(\frac{n}{2}, \frac{n}{2\sigma^2})$$

MoM Estimator $\hat{\sigma}^2_{MoM}$ for σ^2:

$$E(X) = \sigma\sqrt{\frac{2}{\pi}} \implies \sigma = \frac{\bar{X}}{\sqrt{\frac{2}{\pi}}} \implies \sigma^2 = \frac{\pi\bar{X}^2}{2}$$

Bias of MLE and MoM estimators of σ^2:

$$E(X^2) = Var(X) + E(X)^2 = \sigma^2(1 - \frac{2}{\pi}) + \sigma^2\frac{2}{\pi} = \sigma^2 \quad E(\sigma^2_{MLE}) = E(\sum\frac{X_i^2}{n}) = \frac{1}{n}\sum E(X_i^2) = \frac{1}{n}\sum\sigma^2 = \frac{1}{n}*n\sigma^2 = \sigma^2 \quad Bias(\sigma^2_M$$

$$E(\bar{X}^2) = Var(\bar{X}) + E(\bar{X})^2 = \frac{Var(X)}{n} + E(X)^2 = \frac{\sigma^2 - \frac{2}{\pi}}{n} + \sigma^2\frac{2}{\pi} \quad E(\sigma^2_{MoM}) = E(\frac{\pi\bar{X}^2}{2}) = \frac{\pi}{2}E(\bar{X}^2) = \frac{\pi}{2}(\frac{\sigma^2 - \frac{2}{\pi}}{n} + \sigma^2\frac{2}{\pi}) = \frac{\pi\sigma^2}{2}$$

Find the exact 95% confidence interval of σ^2:

$$X_i^2 = W_i \sim Gamma(\frac{1}{2}, \frac{1}{2\sigma^2}) \quad \frac{\sum X_i^2}{n} = \frac{\sum W_i}{n} \implies \bar{W} \sim Gamma(\frac{n}{2}, \frac{n}{2\sigma^2})$$

The Pivotal Statistic needs to be a function of σ^2 and the distribution does not involve σ^2:

$$\frac{1}{\sigma^2}\bar{W} \sim Gamma(\frac{n}{2}, \frac{n}{2})$$

Confidence Interval:

$$1 - \alpha = P(q_{\alpha/2} \leq \frac{\bar{W}}{\sigma^2} \leq q_{1-\alpha/2}) = P(\frac{q_{\alpha/2}}{\bar{W}} \leq \frac{1}{\sigma^2} \leq \frac{q_{1-\alpha/2}}{\bar{W}}) = P(\frac{\bar{W}}{q_{\alpha/2}} \geq \sigma^2 \geq \frac{\bar{W}}{q_{1-\alpha/2}}) = P(\frac{\bar{X}^2}{q_{\alpha/2}} \geq \sigma^2 \geq \frac{\bar{X}^2}{q_{1-\alpha/2}}) \quad L(x) = \frac{\bar{X}^2}{q_{0.97}}$$