R Homework 4

Robert Garrett

November 29, 2017

Research Question

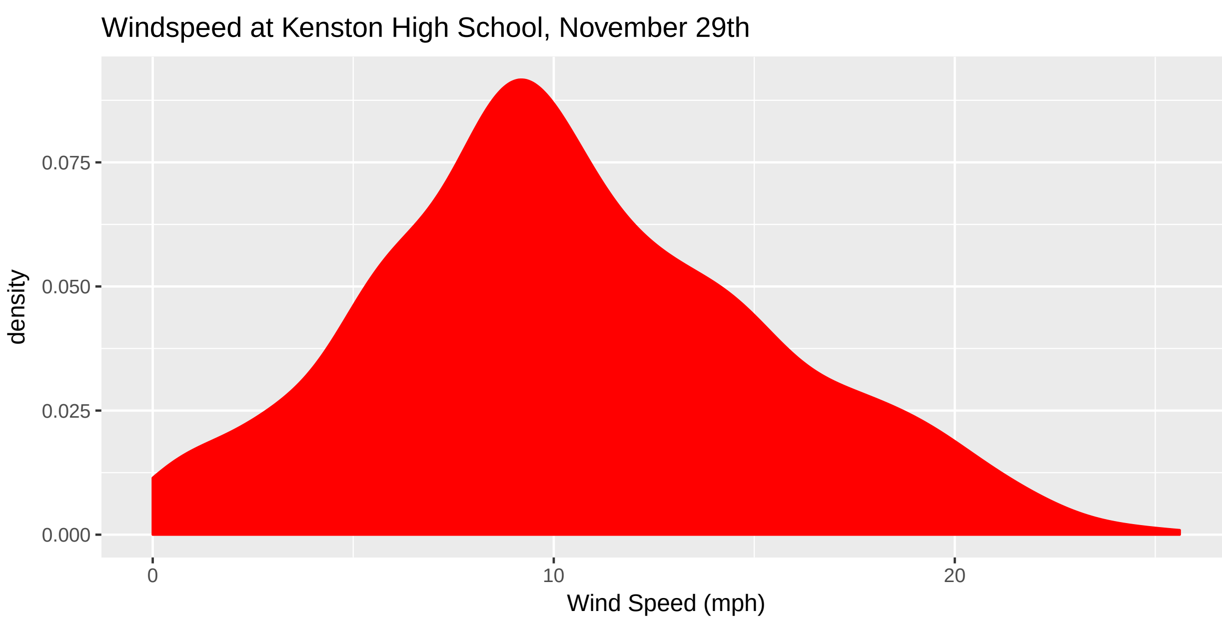
Environmental engineers need to consider many factors when designing and building windmills. One such factor is the windspeed in an area. It is valuable to know the typical windspeeds and the spread of the windspeeds for an area where a windmill is to be constructed. Using Method of Moments and Maximum Likelihood estimation, we can fit distributions to a sample of windspeeds in an area. The fitted distribution can be used in simulating the wind conditions of an area in order for the windmill to be designed optimally and placed in the right location

Dataset

My mother works as a stat teacher at Kenston High School, where a windmill was recently constructed. The data from this windmill is publicly available at http://www.kw4ed.org/projects/kenston/data.php. I scraped November 29th's data from the site, and will fit two different distributions to it after applying the Method of Moments and Maximum Likilihood estimation. This could allow engineers to use my models in order to make sure that the power gathering systems of the windmill are capable of handling the power load produced by the windmill. First, using the Method of Moments I will fit a Gamma distribution to the data. Secondly, I will use MLE to fit a normal distribution to the data. Finally, I can compare the effectiveness of the Gamma and Normal Distributions in modeling the wind speed data.

Here is a histogram of the raw data:

```
ggplot(data=windspeed,aes(Wind.Speed))+
  geom_density(color="red",fill = "red")+
  labs(x="Wind Speed (mph)",title = "Windspeed at Kenston High School, November 29th")
```



Method of Moments

We know the mean and variance, and thus can set these equal to the first moment around 0 and second moment around the mean, respectively:

$$X_i \sim Gamma(\alpha, \beta)$$
$$E(X_i) = \alpha\beta = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$V(X_i) = \alpha\beta^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

From the first equation, we get:

$$\alpha\beta = \overline{X} \Rightarrow \alpha = \frac{\overline{X}}{\beta}$$

And from the second equation we get:

$$\alpha\beta^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \Rightarrow$$
$$\overline{X}\beta = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \Rightarrow$$
$$\hat{\beta}_{MM} = \frac{1}{n\overline{X}}\sum_{i=1}^{n}(X_i - \overline{X})^2 \Rightarrow$$
$$\hat{\alpha}_{MM} = \frac{\overline{X}}{\hat{\beta}_{MM}}$$

Now that these derivations are complete, we can plug the data into the formulas to get our method of moment estimations for alpha and beta.

```
n = nrow(windspeed)
x_bar = sum(windspeed$Wind.Speed)/n
beta_hat_MM = (1/(x_bar*n))*sum((windspeed$Wind.Speed-x_bar)^2)
alpha_hat_MM = x_bar/beta_hat_MM
alpha_hat_MM
```

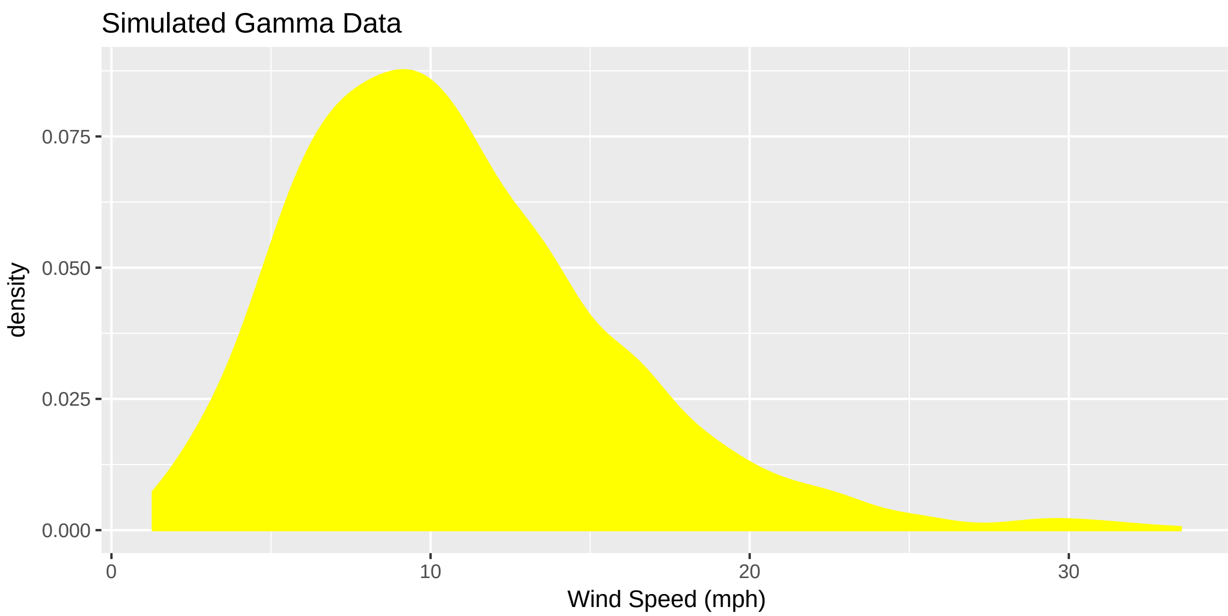
```
## [1] 4.429312
```

```
beta_hat_MM
```

```
## [1] 2.353666
```

Now that we have these values, we can simulate data and plot a histogram of the results:

```
Gamma_MM = rgamma(n,shape=alpha_hat_MM,scale=beta_hat_MM)
ggplot(data=data.frame(Gamma_MM),aes(Gamma_MM))+
  geom_density(color = "yellow",fill="yellow")+
  labs(x="Wind Speed (mph)",title = "Simulated Gamma Data")
```



Maximum Likelihood Estimation

Now to find the maximum likelihood estimators for the normal distribution. We start by finding the log likelihood function:

$$
\begin{aligned}
ln(L(x_1, ..., x_n|\mu, \sigma^2)) &= ln(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} * ... * \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}) \\
&= ln(\frac{1}{\sqrt{2\pi\sigma^2}^n} * e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}) \\
&= \frac{-n}{2} ln(2/pi) - \frac{n}{2} ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2
\end{aligned}
$$

Now to find the MLE for mu, we must take the partial derivative wrt mu and maximize it. So,

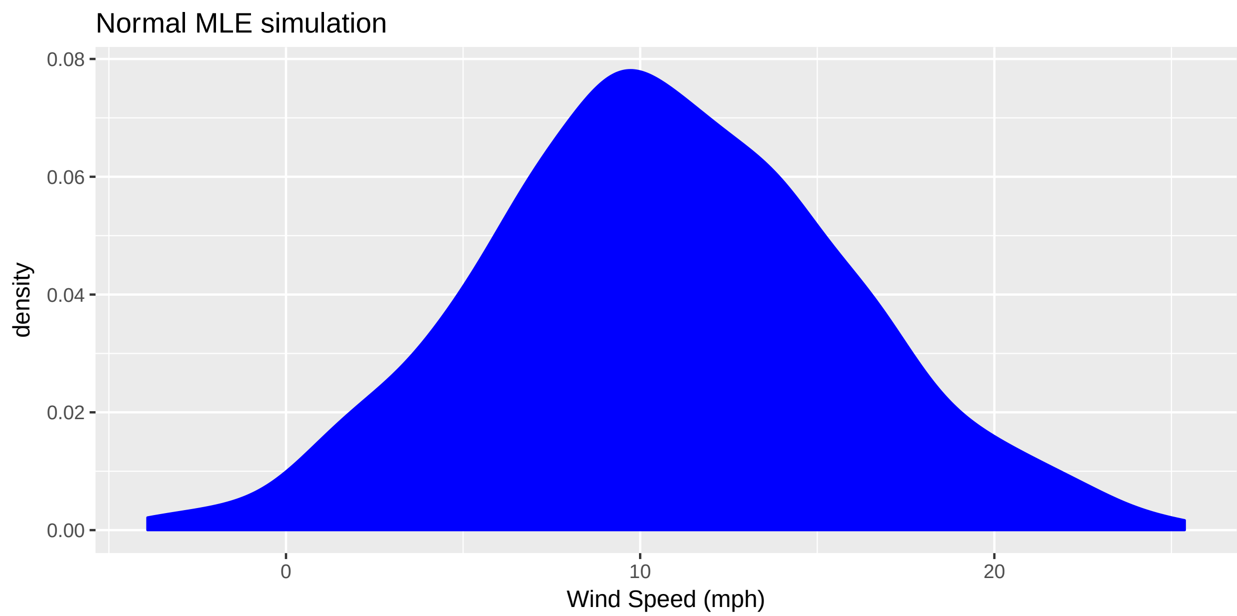
$$\frac{1}{\sigma^2}(\overline{X}) = \frac{1}{\sigma^2}((\sum_{i=1}^{n} x_i) - n\mu) = 0 \Rightarrow \hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Same goes for variance:

$$\frac{1}{2\sigma^2}(\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 - n) = 0 \Rightarrow \hat{\sigma^2}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_{MLE})^2$$

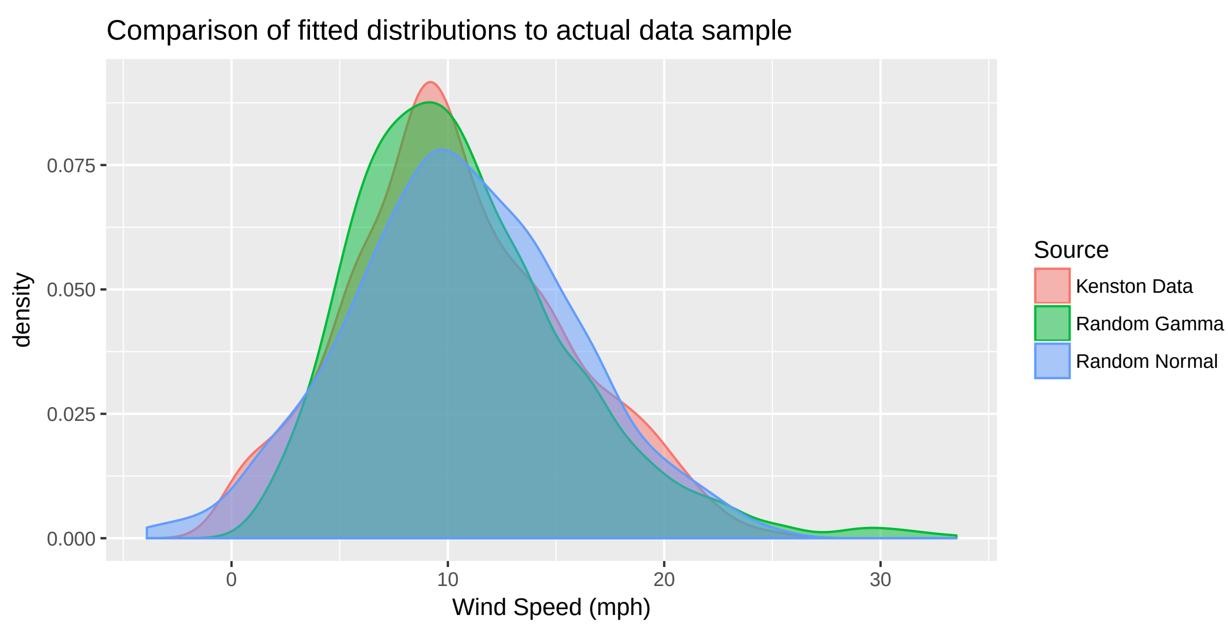
Now that we have finished the derivations, we can find these estimators for our dataset:

```
mu_hat_MLE = sum(windspeed$Wind.Speed)/n
sigmasq_hat_MLE = sum((windspeed$Wind.Speed-mu_hat_MLE)^2)/n

normal_MLE = rnorm(n,mu_hat_MLE,sqrt(sigmasq_hat_MLE))
ggplot(data = data.frame(normal_MLE),aes(normal_MLE)) +
  geom_density(fill = "blue",color = "blue")+
  labs(title = "Normal MLE simulation",x="Wind Speed (mph)")
```



Results

```r
data = data.frame(
  Speed = c(windspeed$Wind.Speed,Gamma_MM,normal_MLE),
  Source = c(rep("Kenston Data",1000),rep("Random Gamma",1000),rep("Random Normal",1000))
)

ggplot(data)+
  geom_density(aes(Speed,fill=Source,color=Source),alpha=.5)+labs(x="Wind Speed (mph)",title = "Comparis
```



Clearly, looking at our comparison histogram the simulated gamma distribution fits the shape, center and skew of the distribution of wind speeds better. Along with this, the gamma distribution accounts for the non negative nature of wind speeds while the normal does not. Further research problems include fitting more types of models to see if another distribution fits well, repeating the same fitting process but using MLE for Gamma, and incorporating more features into the model fitting process to create a more useful simulation.