Math Statistics
Weekly HW 7

Ozaner Hansha

October 30, 2020

Note that we use $\bar{M}_k$ to denote the kth sample moment of some understood RV X. Also note that we use the following estimator to clean up our calculations:

$$\widehat{\sigma^2} = \bar{M}_2 - \bar{M}_1^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2 \qquad \text{(i.e. the MLE sample variance)}$$

Question 1

Problem: Consider a negative binomial distribution $NB(p,r)$. Use the method of moments to find estimators for p and r.

Solution: Recall that the method of moments estimator for k parameters is to set the first k moments of the distribution equal to the first k sample moments and solve for each parameter:

$$\begin{aligned}
\begin{cases}\bar{M}_1 = E[X] \\ \bar{M}_2 = E[X^2]\end{cases} &\implies \begin{cases}\bar{M}_1 = \frac{pr}{1-p} \\ \bar{M}_2 = E[X^2]\end{cases} && \text{(mean of NB distribution)} \\
&\implies \begin{cases}\bar{M}_1 = \frac{pr}{1-p} \\ \bar{M}_2 = \operatorname{Var}(X) + E[X]^2\end{cases} && (\operatorname{Var}(X) = E[X^2] - E[X]^2) \\
&\implies \begin{cases}\bar{M}_1 = \frac{pr}{1-p} \\ \bar{M}_2 = \frac{pr}{(1-p)^2} + \left(\frac{pr}{1-p}\right)^2\end{cases} && \text{(mean \& variance of NB distribution)} \\
&\implies \begin{cases}\bar{M}_1 = \frac{pr}{1-p} \\ \bar{M}_2 = \frac{pr(pr+1)}{(1-p)^2}\end{cases} \implies \begin{cases}\bar{M}_1 = \frac{pr}{1-p} \\ \frac{\bar{M}_2}{\bar{M}_1} = \frac{pr+1}{1-p}\end{cases} \\
&\implies \begin{cases}\bar{M}_1 = \frac{pr}{1-p} \\ \frac{\bar{M}_2}{\bar{M}_1} - \bar{M}_1 = \frac{1}{1-p}\end{cases} \implies \begin{cases}\bar{M}_1 = \frac{pr}{1-p} \\ \frac{1}{\frac{\bar{M}_2}{\bar{M}_1} - \bar{M}_1} = 1-p\end{cases} \\
&\implies \begin{cases}\bar{M}_1 = \frac{pr}{1-p} \\ \frac{\bar{M}_1}{\bar{M}_2 - \bar{M}_1^2} = 1-p\end{cases} \\
&\implies \begin{cases}\bar{X} = \frac{pr}{1-p} \\ \frac{\bar{X}}{\widehat{\sigma^2}} = 1-p\end{cases} && \text{(def. of } \bar{X} \text{ \& } \widehat{\sigma^2}) \\
&\implies \begin{cases}\bar{X} = \frac{pr}{1-p} \\ 1 - \frac{\bar{X}}{\widehat{\sigma^2}} = p\end{cases} \implies \begin{cases}\bar{X} = \frac{\left(1-\frac{\bar{X}}{\widehat{\sigma^2}}\right)r}{1-\left(1-\frac{\bar{X}}{\widehat{\sigma^2}}\right)} \\ 1 - \frac{\bar{X}}{\widehat{\sigma^2}} = p\end{cases} \\
&\implies \begin{cases}\bar{X} = \frac{\left(1-\frac{\bar{X}}{\widehat{\sigma^2}}\right)r}{\frac{\bar{X}}{\widehat{\sigma^2}}} \\ 1 - \frac{\bar{X}}{\widehat{\sigma^2}} = p\end{cases} \implies \begin{cases}\frac{\bar{X}^2}{\widehat{\sigma^2}} = \left(1-\frac{\bar{X}}{\widehat{\sigma^2}}\right)r \\ 1 - \frac{\bar{X}}{\widehat{\sigma^2}} = p\end{cases} \\
&\implies \begin{cases}\frac{\bar{X}^2}{\widehat{\sigma^2}\left(1-\frac{\bar{X}}{\widehat{\sigma^2}}\right)} = r \\ 1 - \frac{\bar{X}}{\widehat{\sigma^2}} = p\end{cases} \implies \begin{cases}\frac{\bar{X}^2}{\widehat{\sigma^2} - \bar{X}} = r \\ 1 - \frac{\bar{X}}{\widehat{\sigma^2}} = p\end{cases}
\end{aligned}$$

And so we are left with the following estimators:

$$\hat{p}_{\text{MM}} = 1 - \frac{\bar{X}}{\widehat{\sigma^2}} = 1 - \frac{\bar{X}}{\bar{M}_2 - \bar{X}^2}$$
$$\hat{r}_{\text{MM}} = \frac{\bar{X}^2}{\widehat{\sigma^2} - \bar{X}} = \frac{\bar{X}^2}{\bar{M}_2 - \bar{X}^2 - \bar{X}}$$

Question 2

Problem: Consider a uniform distribution $\mathcal{U}(a, b)$. Use the method of moments to find estimators for $\theta_1 = b - a$ and $\theta_2 = a + b$.

Solution: Just as above we set the moments equal to the sample moments:

$$\begin{cases} \bar{M}_1 = E[X] \\ \bar{M}_2 = E[X^2] \end{cases} \implies \begin{cases} \bar{M}_1 = E[X] \\ \bar{M}_2 = \text{Var}(X) + E[X]^2 \end{cases} \qquad (\text{Var}(X) = E[X^2] - E[X]^2)$$
$$\implies \begin{cases} \bar{M}_1 = \frac{a+b}{2} \\ \bar{M}_2 = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4} \end{cases} \qquad (\text{mean \& variance of uniform distribution})$$
$$\implies \begin{cases} \bar{M}_1 = \frac{\theta_2}{2} \\ \bar{M}_2 = \frac{\theta_1^2}{12} + \frac{\theta_2^2}{4} \end{cases} \qquad (\text{def. of } \theta_1 \ \& \ \theta_2)$$
$$\implies \begin{cases} 2\bar{M}_1 = \theta_2 \\ \bar{M}_2 = \frac{\theta_1^2}{12} + \frac{\theta_2^2}{4} \end{cases} \implies \begin{cases} 2\bar{M}_1 = \theta_2 \\ \bar{M}_2 = \frac{\theta_1^2}{12} + \frac{4\bar{M}_1^2}{4} \end{cases}$$
$$\implies \begin{cases} 2\bar{M}_1 = \theta_2 \\ \bar{M}_2 - \bar{M}_1^2 = \frac{\theta_1^2}{12} \end{cases}$$
$$\implies \begin{cases} 2\bar{M}_1 = \theta_2 \\ \widehat{\sigma^2} = \frac{\theta_1^2}{12} \end{cases} \qquad (\text{def. of } \bar{X} \ \& \ \widehat{\sigma^2})$$
$$\implies \begin{cases} 2\bar{X} = \theta_2 \\ \sqrt{12\widehat{\sigma^2}} = \theta_1 \end{cases}$$

And so we are left with the following estimators:

$$\hat{\theta}_{1\text{MM}} = \sqrt{12\widehat{\sigma^2}} = \sqrt{12(M_2 - \bar{X}^2)}$$
$$\hat{\theta}_{2\text{MM}} = 2\bar{X}$$

Question 3

Problem: Consider a uniform distribution $\mathcal{U}(a, b)$. Give the MLEs of both a and b, given an i.i.d. sample of size n.

Solution: The MLE of a parameter is the one that maximizes the likelihood of observing the sample. In our case we have:

$$\begin{aligned}
(\hat{a}_{\text{MLE}}, \hat{b}_{\text{MLE}}) &= \operatorname*{arg\,max}_{\substack{a,b \\ a<b}} p_X(\mathbf{x}; a, b) && \text{(def. of MLE)} \\
&= \operatorname*{arg\,max}_{\substack{a,b \\ a<b}} \prod_{i=1}^{n} p_{X_i}(x_i; a, b) && \text{(independent observations)} \\
&= \operatorname*{arg\,max}_{\substack{a,b \\ a<b}} \prod_{i=1}^{n} \frac{1}{b-a}[x_i \in [a, b]] && \text{(pdf of uniform distribution)} \\
&= \operatorname*{arg\,max}_{\substack{a,b \\ a<b}} \frac{1}{(b-a)^n} \prod_{i=1}^{n} [x_i \in [a, b]] \\
&= \operatorname*{arg\,max}_{\substack{a,b \\ a<b}} \frac{1}{(b-a)^n} [(\forall i \in (1..n)) \ x_i \in [a, b]] && \text{(product of indicators is } \wedge\text{)} \\
&= \operatorname*{arg\,max}_{\substack{a,b \\ a<b}} \underbrace{\frac{1}{b-a} [(\forall i \in (1..n)) \ x_i \in [a, b]]}_{f(a,b)} && (\tfrac{1}{b-a} > 0\text{, same as maximizing } n\text{th root})
\end{aligned}$$

Now let us start with $\hat{a}_{\text{MLE}}$. Taking b to be constant, the function $\frac{1}{b-a}$ over the interval $(-\infty, b)$ (which are all the possible values of a) is both monotonically increasing and always greater than 0.

As such, maximizing a is a matter of picking a value as close to b as possible, while still satisfying the indicator function. Because if a did not satisfy it then $f(a, b) = 0$ which is lower than $f(a', b)$ for any other a' which *does* satisfy the indicator function.

In other words, a is given by the minimum of the sample. If it was any higher, it would not satisfy the indicator and thus not maximize f. If it was any lower, then it would not be as large since $\frac{1}{b-a}$ is monotonically increasing:

$$\hat{a}_{\text{MLE}} = \min\{x_i \mid \forall i \in (1..n)\}$$

Similar reasoning holds for $\hat{b}_{\text{MLE}}$. Taking a to be constant, the function $\frac{1}{b-a}$ over the interval (a, ∞) (which are all the possible values of b) is both monotonically decrasing and always greater than 0.

As such, maximizing b is a means picking a value as close to a as possible, while still satisfying the indicator function. Because if b did not satisfy it then $f(a, b) = 0$ which is lower than $f(a, b')$ for any other b' which *does* satisfy the indicator function.

In other words, b is given by the maximum of the sample. If it was any lower, it would not satisfy the indicator and thus not maximize f. If it was any higher, then it would not be as large since $\frac{1}{b-a}$ is monotonically decreasing:

$$\hat{b}_{\text{MLE}} = \max\{x_i \mid \forall i \in (1..n)\}$$