1 Method of Moments Estimation

1.1 The Technique

- Coming up with estimators can be challenging.
- Multiple techniques have been developed for finding "good" estimators
- One of these techniques is called the "Method of Moments"
  - By definition we have the kth moment of a distribution
  $$\mu_k = E[X^k]$$
  - We also have what is called the kth *sample* moment
  $$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$
  - If there are k unknown parameter, then we derive the formulas for k different moments such that each unkown parameter appears in one of the moment formulas
  - We pair the distribution moment formula (in terms of the unkon parameters) and set them equal to the corresponding sample moments
  - We then will have some number of equations with an equal number of unknown values
  - This means there will be a unique solution for our estimators, in terms of the data
  - The resulting solution(s) gives us our estimator(s)
  - NOTE: We assume that our sample is all coming from the same distribution, and that the sample is independent

1.2 Examples

1. Let $X_1, ..., X_n$ be i.i.d. $Bern(p)$ find the MOM estimator for p
   Solution:
   $$\begin{aligned} \mu_1 = E[X_1] &= p \\ m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i &= \bar{X} \end{aligned}$$
   So, we set $\mu_1 = m_1$ and solve the equation in terms of p. In this case the solution is trivially $\hat{p} = \bar{X}$

2. Let $X_1, ..., X_n$ be i.i.d. $U(0, \theta)$ find the MOM estimator for θ
   Solution:

$$\begin{aligned}
\mu_1 = E[X_1] &= \frac{\theta}{2} \\
m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i &= \bar{X}
\end{aligned}$$

   So, we set $\mu_1 = m_1$ and solve the equation in terms of θ. In this case the solution is $\hat{\theta} = 2\bar{X}$

3. Let $X_1, ..., X_n$ be i.i.d. $NBinomial(r, p)$ find the MOM estimators for r and p
   Solution:

$$\begin{aligned}
\mu_1 = E[X_1] &= \frac{r}{p} \\
\mu_2 = E[X_1^2] &= V[X_1] + E^2[X1] \\
&= \frac{r(1-p)}{p^2} + (\frac{r}{p})^2 \\
&= \frac{r(1-p+r)}{p^2} \\
m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i &= \bar{X} \\
m_2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\
\rightarrow \text{set moments equal} & \\
\Rightarrow \frac{r}{p} &= m_1 \\
\frac{r(1-p+r)}{p^2} &= m_2 \\
\Rightarrow r &= m_1 p \\
\Rightarrow m_2 &= \frac{(m_1 p)(1-p+pm_1)}{p^2} \\
&= \frac{m_1(1-p+pm_1)}{p} \\
&= \frac{m_1}{p} - m_1(1-m_1) \\
\Rightarrow m_2 + m_1(1-m_1) &= \frac{m_1}{p} \\
\Rightarrow \hat{p} &= \frac{m_1}{m_2 + m_1(1-m_1)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\bar{X}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 + \bar{X}(1-\bar{X})} \\
\Rightarrow \hat{r} &= m_1\hat{p} \\
&= \frac{m_1^2}{m_2 + m_1(1-m_1)} \\
&= \frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 + \bar{X}(1-\bar{X})}
\end{aligned}
$$

1.3 Exercises

1. Let $X1, ..., X_n$ be i.i.d. $Poisson(\lambda)$. Find the MOM estimator for λ
   Solution:

   $$
   \begin{aligned}
   \mu_1 = E[X_1] &= \lambda \\
   m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i &= \bar{X}
   \end{aligned}
   $$

   So, we set $\mu_1 = m_1$ and solve the equation in terms of λ. In this case the solution is trivially $\hat{\lambda} = \bar{X}$

2. Let $X1, ..., X_n$ be i.i.d. $Geo(p)$. Find the MOM estimator for p
   Solution:

   $$
   \begin{aligned}
   \mu_1 = E[X_1] &= \frac{1}{p} \\
   m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i &= \bar{X}
   \end{aligned}
   $$

   So, we set $\mu_1 = m_1$ and solve the equation in terms of p. In this case the solution is trivially $\hat{p} = \frac{1}{\bar{X}}$

3. Let $X1, ..., X_n$ be i.i.d. $\Gamma(\alpha, \beta)$. Find the MOM estimators for α, β
   Solution:

   $$
   \begin{aligned}
   \mu_1 = E[X_1] &= \alpha\beta \\
   \mu_2 = E[X_1^2] &= V[X_1] + E^2[X1] \\
   &= \alpha\beta^2 + (\alpha\beta)^2 \\
   &= (1+\alpha)\alpha\beta^2 \\
   m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i &= \bar{X}
   \end{aligned}
   $$

$$
\begin{aligned}
m_2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\
\rightarrow \text{set moments equal} & \\
\Rightarrow \alpha\beta &= m_1 \\
(1+\alpha)\alpha\beta^2 &= m_2 \\
\Rightarrow \beta &= \frac{m_1}{\alpha} \\
\Rightarrow m_2 &= (1+\alpha)\alpha(\frac{m_1}{\alpha})^2 \\
&= \frac{m_1^2}{\alpha} + m_1^2 \\
\Rightarrow m_2 - m_1^2 &= \frac{m_1^2}{\alpha} \\
\Rightarrow \hat{\alpha} &= \frac{m_1^2}{m_2 - m_1^2} \\
&= \frac{\bar{X}^2}{(\frac{1}{n}\sum_{i=1}^{n} X_i^2) - \bar{X}^2} \\
\Rightarrow \hat{\beta} &= \frac{m_1}{\hat{\alpha}} \\
&= \frac{m_1}{m_1^2/(m_2 - m_1^2)} \\
&= \frac{m_2 - m_1^2}{m_1} \\
&= \frac{(\frac{1}{n}\sum_{i=1}^{n} X_i^2) - \bar{X}^2}{\bar{X}}
\end{aligned}
$$

2 MLE

2.1 The Technique

- Another technique to creating "good" estimators is called *Maximum likelihood estimation* (MLE for short)
- As the name suggests, we examine what is called the *likelihood* of the data, and then maximize the likelihood function with respect to θ, the unkown parameter(s).

Definition 1 (Likelihood Function)**.** *Let $x_1, ..., x_n$ be observations that correspond to a sample $X_1, ..., X_n$ whose distribution depends on parameter θ. Then, the* likelihood function, *$L(\theta|x_1, ..., x_n)$ is defined to be the joint probability (probability density) of $x_1, ..., x_n$ if $X_1, ..., X_n$ are discrete (continuous).*

- The $|x_1, ..., x_n$ part of the likelihood statement simply is meant to specify that the likelihood is considered a function of the parameter θ and that the data represented as $x_1, ..., x_n$ are assumed observed and thus fixed.

- When $X_1, ..., X_n$ are i.i.d. the Likelihood function becomes very simple, namely:
  When the data is continuous

$$\begin{aligned} L(x_1, ..., x_n|\theta) &= f(x_1, ..., x_n|\theta) \\ &= f(x_1|\theta) \times ... \times f(x_n|\theta) \end{aligned}$$

  and when the data is discrete

$$\begin{aligned} L(x_1, ..., x_n|\theta) &= p(x_1, ..., x_n|\theta) \\ &= p(x_1|\theta) \times ... \times p(x_n|\theta) \end{aligned}$$

2.2 Examples

1. Let $X_1, ..., X_n$ be i.i.d. $Exp(\delta)$. find the MLE of δ
   Solution:

$$\begin{aligned} X_1, ..., X_n \quad \text{i.i.d.} &\quad Exp(\delta) \\ \Rightarrow L(x_1, ..., x_n|\delta) &= \prod_{i=1}^{n} \frac{1}{\delta} e^{-\frac{x_i}{\delta}} I(0 < x_i) \\ &= \frac{1}{\delta^n} e^{-\sum_{i=1}^{n} \frac{x_i}{\delta}} \prod_{i=1}^{n} I(0 < x_i) \\ &= \frac{1}{\delta^n} e^{-\sum_{i=1}^{n} \frac{x_i}{\delta}} I(0 < x_{(1)}) \leftarrow \text{Since } \prod_{i=1}^{n} I(0 < x_i) \equiv I(0 < x_{(1)}) \end{aligned}$$

   We wish to maximize $L(\delta)$, so one technique we can use is to find a critical point of $L(\delta)$ and confirm that it is a global maximum. Usually Likelihood functions are not great for taking derivatives, so what we can usually do is recognize that $\ln(L(\delta))$ is monotonically increasing in δ and therefore $L(\delta)$ will be maximized for the same value of δ that $\ln(L(\delta))$, and then maximize $\ln(L(\delta))$

$$\begin{aligned} \ln(L\delta)) = \ell(\delta) &= -n\ln(\delta) - \sum_{i=1}^{n} \frac{x_i}{\delta} + \ln\Big(I(0 < x_{(1)})\Big) \\ \Rightarrow \frac{\mathrm{d}\ell(\delta)}{\mathrm{d}\delta} &= \frac{-n}{\delta} + \frac{\sum_{i=1}^{n} x_i}{\delta^2} \\ \text{set } \tfrac{\mathrm{d}\ell(\delta)}{\mathrm{d}\delta} = 0 & \\ \Rightarrow 0 &= \frac{-n}{\delta} + \frac{\sum_{i=1}^{n} x_i}{\delta^2} \\ \Rightarrow n &= \frac{\sum_{i=1}^{n} x_i}{\delta} \end{aligned}$$

$$
\begin{aligned}
\Rightarrow \hat{\delta} &= \bar{X}
\end{aligned}
$$

check to make sure critical point is maximum

$$
\begin{aligned}
L(0) &= 0 \\
L(\infty) &= 0 \\
\frac{\mathrm{d}^2\ell(\delta)}{\mathrm{d}\delta^2} &= \frac{n}{\delta^2} - 2\frac{\sum_{i=1}^n x_i}{\delta^3} \\
\Rightarrow \frac{\mathrm{d}^2\ell(\delta)}{\mathrm{d}\delta^2}@(\delta = \bar{X}) &= \frac{n}{\bar{X}^2} - \frac{2n\bar{X}}{\bar{X}^3} \\
&= \frac{n - 2n}{\bar{X}^2} \\
&< 0
\end{aligned}
$$

Since the second derivative is negative,
we have a local, and therefore global max

Therefore $\hat{\delta} = \bar{X}$ is our MLE.

Note, we perform the second derivative test to confirm that the critical point is in fact a maximum, but this is not the only way that this can be done. Though more cumbersome, one could compare the value of the likelihood function as the critical point to the value of the likelihood function at the boundaries of the parameter to show that the likelihood is largest at the critical point, thus making that critical point a maximum.

2. Let $X_1, ..., X_n$ be i.i.d. $U(0, \theta)$. Find the MLE of θ.
Solution:

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^n \frac{1}{\theta} I(0 < x_i < \theta) \\
&= \frac{1}{\theta^n} \prod_{i=1}^n I(0 < x_i < \theta) \\
&= \frac{1}{\theta^n} I(0 < x_{(1)} < x_{(n)} < \theta) \\
&= \frac{1}{\theta^n} I(0 < x_{(1)}) I(x_{(n)} < \theta) \\
L(\theta) &= \begin{cases} 0 & \text{if } \theta \leq x_{(n)} \\ \frac{1}{\theta^n} & \text{if } x_{(n)} < \theta \end{cases}
\end{aligned}
$$

We see that $L(\theta)$ is a monotonically decreasing function of θ when $\theta > x_{(n)}$ so on that interval $L(\theta)$ is maximized when $\theta = x_{(n)}$ and $L(\theta) = 0$ if $\theta \leq x_{(n)}$ thus, $L(\theta)$ is globally maximized when $\theta = x_{(n)}$ thus our MLE is $\hat{\theta} = x_{(n)}$

2.3 Exercises

1. Let $X_1, ..., X_n$ be i.i.d. $Poisson(\lambda)$. Find the MLE of λ.

2. Let $X_1, ..., X_n$ be i.i.d. $N(0, \sigma^2)$. Find the MLE of σ^2.

2.4 Solutions

1. Let $X_1, ..., X_n$ be i.i.d. $Poisson(\lambda)$. Find the MLE of λ.
   Solution:

$$\begin{aligned}
L(\lambda) &= \prod_{i=1}^{n} (\frac{\lambda^{x_i}}{x_i!} e^{-\lambda}) \\
&= \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \prod_{i=1}^{n} \frac{1}{x_i!} \\
\Rightarrow \ln(L(\lambda)) = \ell(\lambda) &= (\sum_{i=1}^{n} x_i) \ln(\lambda) - n\lambda + \sum_{i=1}^{n} - \ln(x_i!) \\
\Rightarrow \frac{\mathrm{d}\ell(\lambda)}{\mathrm{d}\lambda} &= \frac{(\sum_{i=1}^{n} x_i)}{\lambda} - n \\
\text{Set } \tfrac{\mathrm{d}\ell(\lambda)}{\mathrm{d}\lambda} = 0 & \\
\Rightarrow \frac{(\sum_{i=1}^{n} x_i)}{\lambda} - n &= 0 \\
\Rightarrow \hat{\lambda} &= \bar{X} \\
\text{second derivative test to confirm max} & \\
\rightarrow \frac{\mathrm{d}^2\ell(\lambda)}{\mathrm{d}\lambda^2} &= -\frac{(\sum_{i=1}^{n} x_i)}{\lambda^2} \\
&< 0 \text{ for all } \lambda > 0
\end{aligned}$$

   Thus $\hat{\lambda} = \bar{X}$ is a maximal value, and thus is our MLE.

2. Let $X_1, ..., X_n$ be i.i.d. $N(0, \sigma^2)$. Find the MLE of σ^2.

   Solution:

$$\begin{aligned}
L(\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i)^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i)^2}
\end{aligned}$$

$$
\begin{aligned}
\Rightarrow \ln(L(\sigma^2)) = \ell(\sigma^2) &= -\frac{n}{2}\ln(2\pi\sigma^2) + -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i)^2 \\
\Rightarrow \frac{\mathrm{d}\ell(\sigma^2)}{\mathrm{d}\sigma^2} &= -\frac{n}{2}\frac{2\pi}{2\pi\sigma^2} + -\frac{\sum_{i=1}^n x_i^2}{2}(-1)\frac{1}{(\sigma^2)^2} \\
&\quad \uparrow \text{Note, we differentiate with } \uparrow \\
&\quad\quad \text{respect to } \sigma^2 \text{ not } \sigma \\
\text{Set } \tfrac{\mathrm{d}\ell(\sigma^2)}{\mathrm{d}\sigma^2} = 0 & \\
\Rightarrow -\frac{n}{2}\frac{1}{\sigma^2} + \frac{\sum_{i=1}^n x_i^2}{2}\frac{1}{(\sigma^2)^2} &= 0 \\
\Rightarrow \frac{\sum_{i=1}^n x_i^2}{2}\frac{1}{\sigma^2} &= \frac{n}{2} \\
\Rightarrow \frac{\sum_{i=1}^n x_i^2}{n} &= \sigma^2 \\
\text{second derivative test to confirm max} & \\
\rightarrow \frac{\mathrm{d}^2\ell(\sigma^2)}{\mathrm{d}(\sigma^2)^2} &= -\frac{n}{2}\frac{-1}{(\sigma^2)^2} + \frac{\sum_{i=1}^n x_i^2}{2}\frac{-2}{(\sigma^2)^3} \\
&= \frac{n\sigma^2}{2(\sigma^2)^3} + \frac{-2\sum_{i=1}^n x_i^2}{2(\sigma^2)^3} \\
&= \frac{n\sigma^2 - 2\sum_{i=1}^n x_i^2}{2(\sigma^2)^3} \\
\text{check second derivative at } \sigma^2 = \tfrac{\sum_{i=1}^n x_i^2}{n} & \\
\rightarrow \frac{\mathrm{d}^2\ell(\sigma^2)}{\mathrm{d}(\sigma^2)^2}\Big|_{\sigma^2 = \frac{\sum_{i=1}^n x_i^2}{n}} &= \frac{n\frac{\sum_{i=1}^n x_i^2}{n} - 2\sum_{i=1}^n x_i^2}{2((\frac{\sum_{i=1}^n x_i^2}{n})^2)^3} \\
&< 0 \\
\text{Since} \sum_{i=1}^n x_i^2 - 2\sum_{i=1}^n x_i^2 &< 0
\end{aligned}
$$

Thus $\hat{\sigma^2} = \frac{\sum_{i=1}^n x_i^2}{n}$ is a maximal value and therefore is our MLE.

2.5 Invariance Property

- One neat feature of Maximum Likelihood estimation is the Invariance propery of MLEs

Theorem 1. *Let θ be an unknown parameter, $\hat{\theta}$ be the MLE for θ, and let g be a function that is defined over the range of possible values of θ. Then the MLE of $g(\theta)$ is $g(\hat{\theta})$*

2.6 Examples

Let $X_1, ..., X_n$ be i.i.d. $Exp(\delta)$. We have shown that $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ is the MLE for δ. Therefore, by the invariance property:

1. $\bar{X}^2$ the MLE for δ^2 (the variance of the underlying population).
2. The MLE for $\sqrt{\delta^2 + \ln(\delta)}$ is $\sqrt{\bar{X}^2 + \ln(\bar{x})}$