The Method of Moments

Bruce M. Boghosian

Tufts University School of Arts and Sciences

Department of Mathematics

Tufts University

Outline

Motivation

Bruce M. Boghosian

- Distributions are often written so that the mean and the variance are parameters,
$$f_X(x) = \frac{1}{\sqrt{2\pi v}} \exp\left[-\frac{(x-\mu)^2}{2v}\right]$$
- Using MLE, we noticed that the estimate of the mean was often equal to the sample mean, and the estimate of the variance was often equal to the sample variance.
- Might it be possible to determine estimators in this way, by demanding that the moments of the posited distribution are equal to those of the sample?
- We match as many moments as there are parameters in our posited distribution.
- From our explorations last time, we know that this will work in certain cases.

General methodology

- Make n measurements of Y, resulting in $Y_j = y_j$ for $j = 1, \ldots, n$
- Suppose our posited density function has s parameters, $f_Y(y; \theta_1, \ldots, \theta_s)$
- Find s *moments*, and set these equal to corresponding *sample moments*

$$E(Y) = \int dy\ f_Y(y; \theta_1, \ldots, \theta_s) y = \frac{1}{n} \sum_{j=1}^{n} y_j$$

$$E(Y^2) = \int dy\ f_Y(y; \theta_1, \ldots, \theta_s) y^2 = \frac{1}{n} \sum_{j=1}^{n} y_j^2$$

$$\vdots$$

$$E(Y^s) = \int dy\ f_Y(y; \theta_1, \ldots, \theta_s) y^s = \frac{1}{n} \sum_{j=1}^{n} y_j^s$$

- Yields s simultaneous equations for the s parameters.

General Bernoulli trial

- Define random variable for each coin toss,

$$X := \begin{cases} 1 & \text{if toss results in heads (with probability } p) \\ 0 & \text{if toss results in tails (with probability } 1-p). \end{cases}$$

- Discrete probability function for one toss, where $k \in \{0, 1\}$,

$$p_X(k) = \text{Prob}(X = k) = p^k (1-p)^{1-k}$$

- Normalization: $\sum_{k=0}^{1} p_X(k) = (1-p) + p = 1$
- Mean: $\sum_{k=0}^{1} p_X(k)k = (1-p)0 + p1 = p$
- One parameter, so we estimate p by the sample mean

$$\hat{p}(\vec{k}) = \frac{1}{n}\sum_{j=1}^{n} k_j.$$

- This is identical to our result from MLE.

The Poisson distribution

- We have

$$p_X(k) = \text{Prob}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

- Normalization: $\sum_{k=0}^{\infty} p_X(k) = 1$
- Mean: $E(X) = \sum_{k=0}^{\infty} p_X(k)k = \lambda$
- One parameter, so we estimate λ by the sample mean

$$\hat{\lambda}(\vec{k}) = \frac{1}{n}\sum_{j=1}^{n} k_j.$$

- This is identical to our result from MLE.

The standard normal distribution

- We have

$$f_X(x) = \frac{1}{\sqrt{2\pi v}} \exp\left[-\frac{(x-\mu)^2}{2v}\right],$$

- Normalization: $\int_{\mathbb{R}} dx\ f_X(x) = 1$
- Mean: $E(X) = \int_{\mathbb{R}} dx\ f_X(x) x = \mu$
- Mean square: $E(X^2) = \int_{\mathbb{R}} dx\ f_X(x) x^2 = \mu^2 + v$
- Two parameters, so set

$$\hat{\mu}(\vec{x}) = \frac{1}{n}\sum_{j=1}^{n} x_j \quad \text{and} \quad [\hat{\mu}(\vec{x})]^2 + \hat{v}(\vec{x}) = \frac{1}{n}\sum_{j=1}^{n} x_j^2$$

- This again yields the same result as MLE for both $\hat{\mu}(\vec{x})$ and $\hat{v}(\vec{x})$.

The standard normal distribution (continued)

- To verify the last point for $\hat{v}(\vec{x})$, note

$$
\begin{aligned}
\hat{v}(\vec{x}) &= \frac{1}{n}\sum_{j=1}^{n} x_j^2 - [\hat{\mu}(\vec{x})]^2 \\
&= \frac{1}{n}\sum_{j=1}^{n} x_j^2 - 2\,[\hat{\mu}(\vec{x})]^2 + [\hat{\mu}(\vec{x})]^2 \\
&= \frac{1}{n}\sum_{j=1}^{n} x_j^2 - 2\,[\hat{\mu}(\vec{x})]\,\frac{1}{n}\sum_{j=1}^{n} x_j + \frac{1}{n}\sum_{j=1}^{n} [\hat{\mu}(\vec{x})]^2 \\
&= \frac{1}{n}\sum_{j=1}^{n} \left\{ x_j^2 - 2\hat{\mu}(\vec{x})x_j + [\hat{\mu}(\vec{x})]^2 \right\} \\
&= \frac{1}{n}\sum_{j=1}^{n} \left[x_j - \hat{\mu}(\vec{x}) \right]^2 \\
&= \frac{1}{n}\sum_{j=1}^{n} \left[x_j - \left(\frac{1}{n}\sum_{k=1}^{n} x_k \right) \right]^2
\end{aligned}
$$

Continuous random variable: The uniform distribution

- Suppose $X \in \mathbb{R}$ has the continuous *probability density function*,

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } X \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

- Normalization: $\int_{\mathbb{R}} dx\ f_X(x) = \int_a^b dx\ \frac{1}{b-a} = 1$
- Mean: $E(X) = \int_{\mathbb{R}} dx\ f_X(x)x = \int_a^b dx\ \frac{x}{b-a} = \frac{b+a}{2}$
- Mean square: $E(X^2) = \int_{\mathbb{R}} dx\ f_X(x)x^2 = \frac{a^2+ab+b^2}{3}$

Method of moments for uniform distribution

Bruce M. Boghosian

- There are two parameters, so set first two moments equal to sample moments,

$$\begin{aligned}\frac{a+b}{2} =& M_1 := E(X) \\ \frac{a^2+ab+b^2}{3} =& M_2 := E(X^2)\end{aligned}$$

- Solve for a and b to obtain estimators

$$\begin{aligned}\hat{a}(\vec{x}) =& M_1 - \sqrt{3}\sqrt{M_2 - M_1^2} \\ \hat{b}(\vec{x}) =& M_1 - \sqrt{3}\sqrt{M_2 - M_1^2}\end{aligned}$$

- Mean plus or minus $\sqrt{3}$ times standard deviation.
- This is very different from the MLE estimates (min and max).

Comparison of both methods

- Sample n random points uniformly on interval $[3, 5]$

	MLE		MM	
n	a	b	a	b
10	3.0711	4.96605	3.44151	4.93097
50	3.03101	4.93344	3.11829	4.99247
100	3.00624	4.99622	2.91091	4.97592
500	3.01072	4.99533	2.94429	4.96859
1000	3.00406	4.99752	3.03049	5.03114
5000	3.00084	4.9993	3.01044	5.00693
10000	3.00013	4.99969	3.00128	5.00224

- Show Mathematica code

Method of moments for uniform distribution

Bruce M. Boghosian

- There are two parameters, so set first two moments equal to sample moments,

$$
\begin{aligned}
\frac{a+b}{2} &= M_1 := E(X) \\
\frac{a^2+ab+b^2}{3} &= M_2 := E(X^2)
\end{aligned}
$$

- Solve for a and b to obtain estimators

$$
\begin{aligned}
\hat{a}(\vec{x}) &= M_1 - \sqrt{3}\sqrt{M_2 - M_1^2} \\
\hat{b}(\vec{x}) &= M_1 - \sqrt{3}\sqrt{M_2 - M_1^2}
\end{aligned}
$$

- Mean plus or minus $\sqrt{3}$ times standard deviation.
- This is very different from the MLE estimates (min and max).

Wealth distributions

- Pareto distribution for wealth ($\theta > 1$ so mean exists):

$$f_W(w) = \begin{cases} \theta k^\theta w^{-\theta-1} & \text{for } w \geq k \\ 0 & \text{otherwise} \end{cases}$$

- Moments

$$E(W) = \frac{k\theta}{\theta - 1} = M_1$$
$$E(W^2) = \frac{k^2\theta}{\theta - 2} = M_2.$$

- Most of the time M_2 does not exist, so a hybrid approach may be best.

$$\hat{k} = \min_j w_j$$
$$\hat{\theta} = \frac{M_1}{M_1 - \min_j w_j}.$$

Summary

- We have learned the *method of moments* for finding estimators.
- We have compared it to *maximum likelihood estimation*.
- We have seen examples where the two methods give the same estimator.
- We have seen examples where they don't.
- We have seen examples with one and two parameters.
- We have seen a "hybrid" example, where both methods may be used.