Topic 12: Method of Moments

October 29, 2009

1 Introduction

Let $M_!, M_2, \ldots$ be independent random variables having a common distribution possessing a mean μ. The law of large numbers states that the sample means converge to the population mean

$$\bar{M}_n = \frac{1}{n}\sum_{i=1}^n M_i \to \mu_M \quad \text{as } n \to \infty.$$

For independent random variables $X_1, X_2, \ldots$ chosen according to the probability distribution derived from the parameter value θ and m a real valued function. If $k(\theta) = E_\theta m(X_1)$, then

$$\frac{1}{n}\sum_{i=1}^n m(X_i) \to k(\theta) \quad \text{as } n \to \infty.$$

The choices $m(x) = x^m$ is called the **method of moments**. Write

$$\mu_m = EX^m = k_m(\theta).$$

Our estimation procedure follows from these 4 steps.

Step 1. If the model has d parameters, we compute the functions k_m for the first d moments,

$$\mu_1 = k_1(\theta_1, \theta_2 \ldots, \theta_d), \quad \mu_2 = k_2(\theta_1, \theta_2 \ldots, \theta_d), \quad \ldots, \quad \mu_d = k_d(\theta_1, \theta_2 \ldots, \theta_d),$$

obtaining d equations in d unknowns.

Step 2. We, then solve for the d parameters as a function of the moments.

$$\theta_1 = g_1(\mu_1, \mu_2, \cdots, \mu_d), \quad \theta_2 = g_2(\mu_1, \mu_2, \cdots, \mu_d), \quad \ldots, \quad \theta_d = g_d(\mu_1, \mu_2, \cdots, \mu_d). \tag{1}$$

Step 3. Now, based on the data $X_1, X_2, \ldots, X_n$ we compute the first d **sample moments**,

$$\overline{X}, \overline{X^2}, \ldots, \overline{X^d}.$$

Using the law of large numbers, we have that $\mu_m \approx \overline{X^m}$.

Step 4. We replace the distributional moments μ_m by the sample moments $\overline{X^m}$, then the solutions in (1) give us formulas for the **method of moment estimators** $(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_d)$

$$\hat{\theta}_1 = g_1(\bar{X}, \overline{X^2}, \cdots, \overline{X^d}), \quad \hat{\theta}_2 = g_2(\bar{X}, \overline{X^2}, \cdots, \overline{X^d}), \quad \hat{\theta}_d = g_d(\bar{X}, \overline{X^2}, \cdots, \overline{X^d}).$$

How this abstract description works in practice can be best seen through examples.

2 Examples

Example 1. *Let $X_1, X_2, \ldots, X_n$ be a simple random sample of Pareto random variables with density*

$$f_X(x|\beta) = \frac{\beta}{x^{\beta+1}}, \quad x > 1.$$

The cumulative distribution function is

$$F_X(x) = 1 - x^{-\beta}, \quad x > 1.$$

The mean and the variance

$$\mu = \frac{\beta}{\beta - 1}, \qquad \sigma^2 = \frac{\beta}{(\beta-1)^2(\beta-2)}.$$

In this situation, we have one parameter, namely β. Thus, in step 1, we will only need to determine the first moment

$$\mu_1 = \mu = k_1(\beta) = \frac{\beta}{\beta - 1}$$

to find the method of moments estimator $\hat{\beta}$ for β.

For step 2, we solve for β as a function of the mean μ.

$$\beta = g_1(\mu) = \frac{\mu}{\mu - 1}.$$

Consequently, a method of moments estimate for β is obtained by replacing the distributional mean μ by the sample mean $\bar{X}$.

$$\hat{\beta} = \frac{\bar{X}}{\bar{X} - 1}.$$

A good estimator should have a small variance . We can use the delta method to estimate the variance of $\hat{\beta}$. We begin with the fact that $\bar{X}$ has

$$\textit{mean} \quad \frac{\beta}{\beta-1} \quad \textit{and} \quad \textit{variance} \quad \frac{\beta}{n(\beta-1)^2(\beta-2)}$$

We compute

$$g_1'(\mu) = -\frac{1}{(\mu-1)^2}, \quad \textit{giving} \quad g_1'\left(\frac{\beta}{\beta-1}\right) = -(\beta-1)^2$$

and find that $\hat{\beta}$ has mean approximately equal to β and variance

$$\sigma^2_{\hat{\beta}} \approx g_1'(\mu)^2 \frac{\sigma^2}{n} = (\beta-1)^4 \frac{\beta}{n(\beta-1)^2(\beta-2)} = \frac{\beta(\beta-1)^2}{n(\beta-2)}$$

We simulate the situation with $\beta = 3$ and $n = 100$. Then,

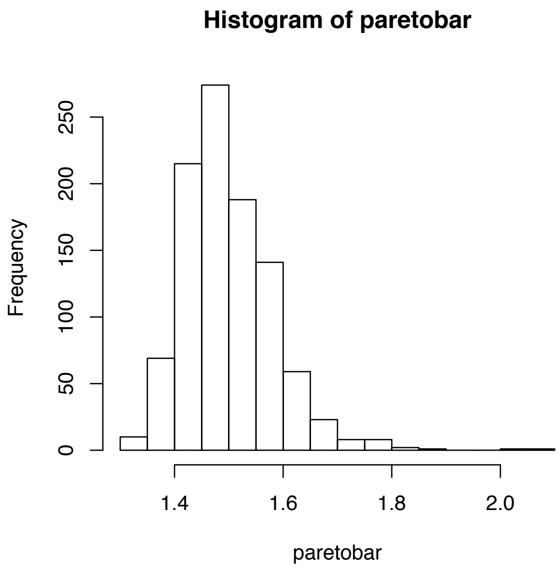
$$\sigma^2_{\hat{\beta}} \approx \frac{3 \cdot 2^2}{100 \cdot 1} = \frac{12}{100} = \frac{3}{25}, \quad \sigma_{\hat{\beta}} \approx 0.346.$$

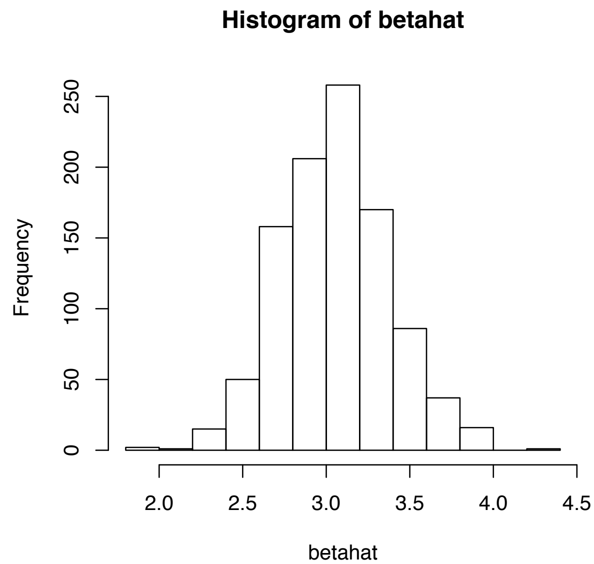
The probability transform states that if the X_i are independent Pareto random variables, then $U_i = F_X(X_i)$ are independent uniform random variables on the interval $[0,1]$. Thus, we can simulate X_i with $F_X^{-1}(U_i)$. If

$$u = F_X(x) = 1 - x^{-3}, \quad \textit{then} \quad x = (1-u)^{-1/3} = v^{-1/3}, \quad \textit{where } v = 1 - u.$$

Note that the $V_i = 1 - U_i$ are also uniform random variables on the interval $[0,1]$ and, consequently, $1/\sqrt[\beta]{V_1}, 1/\sqrt[\beta]{V_2}, \cdots$ have the appropriate Pareto distribution.

```
> paretobar<-rep(0,1000)
> for (i in 1:1000){v<-runif(100);pareto<-1/v^(1/3);paretobar[i]=mean(pareto)}
> hist(paretobar)
> betahat<-paretobar/(paretobar-1)
> hist(betahat)
> mean(betahat)
[1] 3.053254
> sd(betahat)
[1] 0.3200865
```





The sample mean for the estimate for β at 3.053 is close to the simulated value of 3. In this example, the estimator $\hat{\beta}$ is **biased upward***, In other words, on average $\hat{\beta} > \beta$. The sample standard deviation value of 0.320 is close to the value 0.346 estimated by the delta method.*

Example 2. *Fitness is a central concept in the theory of evolution. Relative fitness is quantified as the average number of surviving progeny of a particular genotype compared with average number of surviving progeny of competing genotypes after a single generation.*

Consequently, the distribution of fitness effects, that is, the distribution of fitness for newly arising mutations is a basic question in evolution. A basic understanding of the distribution of fitness effects of newly arising mutations is still in its early stages . Eyre-Walker (2006) examined one particular distribution of fitness effects, namely, deleterious amino acid changing mutations in humans. He approach used a gamma-familty of random variables and gave the estimate of $\hat{\alpha} = 0.23$ and $\hat{\beta} = 5.35$.

A $\Gamma(\alpha, \beta)$ random variable has mean α/β and variance α/β^2. Because we have two parameters, we will need to determine the first two moments.

$$E_{(\alpha,\beta)}X_1 = \frac{\alpha}{\beta} \quad \textit{and} \quad E_{(\alpha,\beta)}X_1^2 = \mathit{Var}_{(\alpha,\beta)}(X_1) + E_{(\alpha,\beta)}[X_1]^2 = \frac{\alpha}{\beta^2} + \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha(1+\alpha)}{\beta^2}.$$

Thus, for step 1, we find that

$$\mu_1 = k_1(\alpha, \beta) = \frac{\alpha}{\beta}, \quad \mu_2 = k_2(\alpha, \beta) = \frac{\alpha(1+\alpha)}{\beta^2}.$$

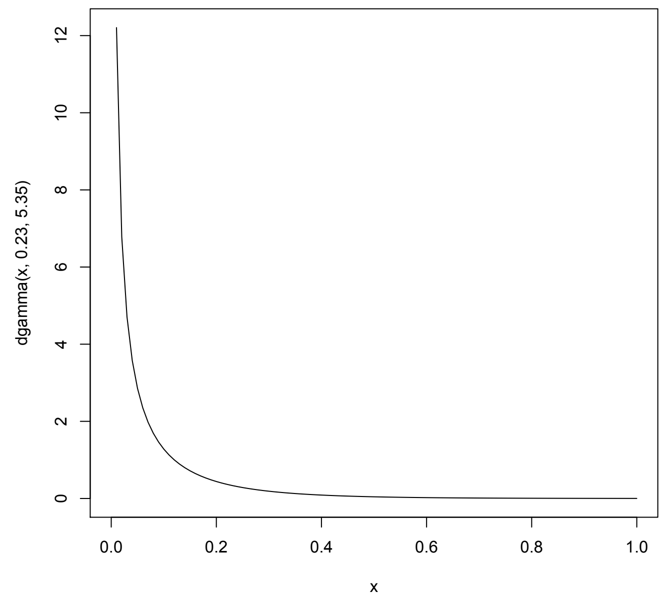


Figure 1: The density of a $\Gamma(0.23, 5.35)$ random variable.

For step 2, we solve for α and β. Note that

$$\mu_2 - \mu_1^2 = \frac{\alpha}{\beta^2},$$

$$\frac{\mu_1}{\mu_2 - \mu_1^2} = \frac{\alpha/\beta}{\alpha/\beta^2} = \beta,$$

and

$$\mu_1 \cdot \frac{\mu_1}{\mu_2 - \mu_1^2} = \frac{\alpha}{\beta} \cdot \beta = \alpha, \quad \textit{or} \quad \alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}.$$

So set

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \quad \textit{and} \quad \overline{X^2} = \frac{1}{n}\sum_{i=1}^n X_i^2$$

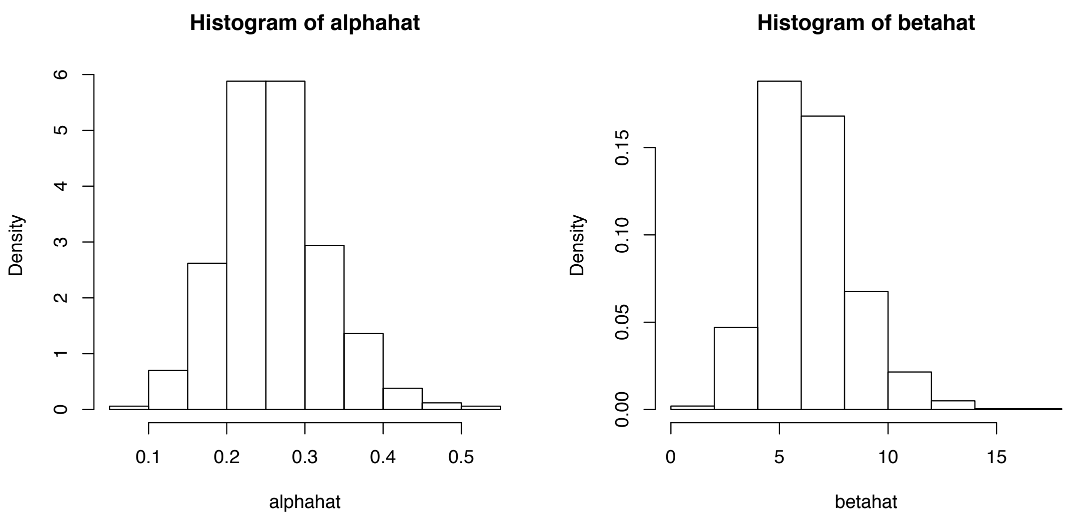
to obtain estimators

$$\hat{\beta} = \frac{\bar{X}}{\overline{X^2} - (\bar{X})^2} \quad \textit{and} \quad \hat{\alpha} = \hat{\beta}\bar{X} = \frac{(\bar{X})^2}{\overline{X^2} - (\bar{X})^2}.$$

To investigate the method of moments on simulated data using R, *we consider 1000 repetitions of 100 independent observations of a $\Gamma(0.23, 5.35)$ random variables.*

```
> xbar <- rep(0,1000)
> x2bar <- rep(0,1000)
> for (i in 1:1000){x<-rgamma(100,0.23,5.35);xbar[i]=mean(x);x2bar[i]=mean(x^2)}
> betahat <- xbar/(x2bar-(xbar)^2)
> alphahat <- betahat*xbar
> mean(alphahat)
[1] 0.2599513
> sd(alphahat)
```

```
[1] 0.06628046
> mean(betahat)
[1] 6.331623
> sd(betahat)
[1] 2.046528
> hist(alphahat,probability=TRUE)
> hist(betahat,probability=TRUE)
```



As we see, the variance in the estimate of β *is quite large. We will revisit this example using another estimation method in the hopes of reduing this variance.*