IEOR 165 – Lecture 2
Method of Moments

1 Estimating Mean and Variance

Imagine a scenario in which we are building an engineering model for a telephone call center to handle airline reservations. For the purpose of analysis, we need to determine (i) the average time spent handling a single call, and (ii) the variance of time spent handling a single call. To determine these quantities, we decide to conduct an experiment in which we record the length of time spent handling n randomly chosen calls over the time span of one month. The question is how can we determine the average and variance using this data?

1.1 Abstract Model

We can abstract this scenario into the following mathematical setting: Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) random variables from some unknown distribution with cdf $F_X(u)$. We should think of the X_i as n independent measurements from a single unknown distribution, and this corresponds to the length of time spent handling the randomly chosen calls in our scenario above. The mathematical question we are interested in answering is how to determine $\mu(X)$ and $\sigma^2(X)$?

1.2 Law of Large Numbers

One useful insight comes from the law of large numbers (lln). Roughly speaking, the lln states that if $U_1, \ldots, U_n$ are iid random variables, then the sample average of the U_i converges to to true mean $\mu(U)$ as n goes to infinity:

$$\frac{1}{n} \sum_{i=1}^{n} U_i \to \mu(U).$$

This is a useful insight because it means that we can use the sample average $\frac{1}{n} \sum_{i=1}^{n} U_i$ as an *estimate* of the true mean $\mu(U)$.

1.3 Mean and Variance Estimates

In fact, this insight provides an approach for estimating both the mean and variance of X. In particular, observe that

$$\frac{1}{n} \sum_{i=1}^{n} X_i \to \mu(X).$$

Thus, the sample average $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is an estimate of the mean of X. (Note that we use a "hat" over variables to denote an estimated quantity, as opposed to the true quantity.)

Estimating the variance requires a little more work. First, note that variance is defined as $\sigma^2(X) = \mathbb{E}((X - \mu)^2)$. Performing algebra on this expression, we have

$$\sigma^2(X) = \mathbb{E}((X - \mu)^2) = \mathbb{E}(X^2 - 2\mu X + \mu^2) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 = $$
$$\mathbb{E}(X^2) - 2\mu^2 + \mu^2 = \mathbb{E}(X^2) - \mu^2,$$

where we have used the following properties of expectation:

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;

- $\mathbb{E}(kX) = k\mathbb{E}(X)$.

Second, recall that

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \to \mathbb{E}(X^2).$$

This means we can use $\frac{1}{n} \sum_{i=1}^{n} X_i^2$ as an estimate of $\mathbb{E}(X^2)$, and we have already chosen $\frac{1}{n} \sum_{i=1}^{n} X_i$ as our estimate of the mean μ. Combining this, we conclude that an estimate of the variance is given by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \hat{\mu}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^2.$$

1.4 Example: Call Center Estimation

Suppose the data from the experiments is $X_1 = 3$, $X_2 = 1$, $X_3 = 20$, $X_4 = 6$, $X_5 = 5$. (In a real experiment, we would have more than $n = 5$ measurements.) Then using the above formulas, our estimate of the mean is

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{5} \cdot (3 + 1 + 20 + 6 + 5) = 7,$$

and our estimate of the variance is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \hat{\mu}^2 = \frac{1}{5} \cdot (3^2 + 1^2 + 20^2 + 6^2 + 5^2) - 7^2 = 45.2.$$

2 Method of Moments

In the above call center scenario, we assumed that the distribution for the time spent handling a single call was unknown. However, it is the case that the distribution for the service time in a call center is often well represented by an exponential distribution. A random variable X with exponential distribution is denoted by $X \sim \mathcal{E}(\lambda)$, where $\lambda > 0$ is the *rate*, and it is the distribution with pdf

$$f_X(u) = \begin{cases} \lambda \exp(-\lambda u), & \text{if } u \geq 0, \\ 0 & \text{otherwise} \end{cases}.$$

The cdf is given by

$$F_X(u) = \begin{cases} 1 - \exp(-\lambda u), & \text{if } u \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

and so $\mathbb{P}(X > u) = \exp(-\lambda u)$ for $u \geq 0$. The mean is $\mu = \frac{1}{\lambda}$, and the variance is $\sigma^2 = \frac{1}{\lambda^2}$.

One important property of an exponential distribution is that $\sigma^2 = \mu^2$. However, the *estimators* for mean and variance that we derived above are not guaranteed to satisfy this relationship. In fact – for the data in the above example – we have that $\hat{\mu}^2 = 49$ while $\hat{\sigma}^2 = 45.2$. This is not a good situation, because it means that we are not using the full statistical knowledge available to us when making the estimates of mean and variance. Given this mismatch, we turn our attention towards the question of how we can develop mean and variance estimators that make better use of our statistical knowledge.

2.1 Abstract Model

We can abstract this scenario into the following mathematical setting: Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) random variables from a *known* distribution with cdf $F_X(u; \theta_1, \ldots, \theta_p)$, but with p unknown parameters $\theta_1, \ldots, \theta_p$ that affect the shape of the distribution. Two simple examples are:

- $X \sim \mathcal{N}(\theta_1, \theta_2)$, where θ_1 is an unknown mean of the Gaussian and θ_2 is the unknown variance of the Gaussian;

- $X \sim \mathcal{E}(\theta_1)$, where θ_1 is the rate parameter of the exponential distribution.

The mathematical question we are interested in answering is how to determine estimates of $\theta_1, \ldots, \theta_p$?

2.2 Method of Moments Estimator

Note that $\mathbb{E}(X^j)$ is the j-th moment of X, and this is where the name of the method comes from. To simplify the notation, let $\mu_j = \mathbb{E}(X^j)$. The methods of moments estimator is derived using the following steps:

1. Symbolically compute the first p moments and express them as functions of the θ parameters:

$$\mu_1 = \mathbb{E}(X) = g_1(\theta_1, \ldots, \theta_p)$$
$$\mu_2 = \mathbb{E}(X^2) = g_2(\theta_1, \ldots, \theta_p)$$
$$\vdots$$
$$\mu_p = \mathbb{E}(X^p) = g_p(\theta_1, \ldots, \theta_p).$$

2. Symbolically invert the system of equations $g_1, \ldots, g_p$, which gives a new set of functions

$$\theta_1 = h_1(\mu_1, \ldots, \mu_p)$$
$$\theta_2 = h_2(\mu_1, \ldots, \mu_p)$$
$$\vdots$$
$$\theta_p = h_p(\mu_1, \ldots, \mu_p).$$

3. Motivated by the lln, compute the following estimates of the moments

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

$$\vdots$$

$$\hat{\mu}_p = \frac{1}{n} \sum_{i=1}^{n} X_i^p$$

4. Compute estimates of the parameters by substituting $\hat{\mu}_j$ for μ_j in the set of functions $h_1, \ldots, h_p$, which gives the following estimators of the parameters:

$$\hat{\theta}_1 = h_1(\hat{\mu}_1, \ldots, \hat{\mu}_p)$$
$$\hat{\theta}_2 = h_2(\hat{\mu}_1, \ldots, \hat{\mu}_p)$$
$$\vdots$$
$$\hat{\theta}_p = h_p(\hat{\mu}_1, \ldots, \hat{\mu}_p).$$

2.3 Example: Exponential Distribution

We do not consider the example of a Gaussian distribution, because it turns out that the derivation for the method of moments estimator is equivalent to the derivation we did in the first section. Instead, we compute the method of moments estimator for an exponential distribution $X \sim \mathcal{E}(\theta_1)$. Since an exponential distribution has a single parameter, we have $p = 1$. Proceeding with the four steps, gives

1. Note $\mu_1 = \frac{1}{\theta_1}$, which we know by the properties of an exponential distribution;

2. Inverting this function gives $\theta_1 = \frac{1}{\mu_1}$;

3. An estimator of the first moment is $\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$;

4. Substituting this estimator gives $\hat{\theta}_1 = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} X_i}$.

Next, recall the data from the call center experiments: $X_1 = 3$, $X_2 = 1$, $X_3 = 20$, $X_4 = 6$, $X_5 = 5$. Substituting these values into the equation we have derived, we get

$$\hat{\theta}_1 = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} X_i} = \frac{1}{\frac{1}{5} \cdot (3 + 1 + 20 + 6 + 5)} = \frac{1}{7}.$$

To estimate the mean and variance, we substitute the estimated parameter into the equations for the mean and variance. Thus, we get

$$\hat{\mu} = \frac{1}{\hat{\theta}_1} = 7$$

$$\hat{\sigma}^2 = \frac{1}{\hat{\theta}_1^2} = 49.$$

Comparing this estimated mean and variance to the previous estimates, we see that the estimated mean remains the same but the estimated variance is larger. However, this set of estimates is arguably better because these estimates satisfy the relationships we would expect if the data is drawn from an exponential distribution.

2.4   Example: Gamma Distribution

Suppose $X_1, \ldots, X_n$ are iid and drawn from a Gamma distribution, which is a distribution with pdf

$$f_X(u) = \frac{1}{\Gamma(k)\theta^k} u^{k-1} \exp(-u/\theta),$$

where $\Gamma(\cdot)$ is the Gamma function. The mean of a Gamma distribution is given by $\mu = k\theta$, and its variance is $\sigma^2 = k\theta^2$. What are the equations for the method of moments estimator for k, θ? There are two useful hints for this problem. The first is that, in this example, the parameters are $k = \theta_1$ and $\theta = \theta_2$. The second is that the relationship $\mathbb{E}(X^2) = \mu_2 = \sigma^2 + \mu^2$ is true for every distribution where these quantities are finite.

Proceeding with the four steps, gives

1. From the properties of a Gamma distribution, we have

$$\mu_1 = k\theta$$
$$\mu_2 = k\theta^2 + k^2\theta^2;$$

2. To invert these functions, first note that substituting the equation for μ_1 into the equation for μ_2 gives

$$\mu_2 = \mu_1 \theta + \mu_1^2 \qquad \Rightarrow \qquad \theta = \frac{\mu_2 - \mu_1^2}{\mu_1}.$$

Substituting this into the equation for μ_1 gives

$$\mu_1 = \frac{k(\mu_2 - \mu_1^2)}{\mu_1} \qquad \Rightarrow \qquad k = \frac{\mu_1^2}{\mu_2 - \mu_1^2}.$$

3. An estimator of the first and second moments is

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2;$$

4. Substituting this estimator into the equations for k, θ gives

$$\hat{k} = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2} = \frac{\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}$$

$$\hat{\theta} = \frac{\hat{\mu}_2 - \hat{\mu}_1^2}{\hat{\mu}_1} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n} X_i}.$$

2.5 Example: Linear Model of Building Energy

The amount of energy consumed in a building on the i-th day E_i heavily depends on the outside temperature on the i-th day T_i. Generally, there are two situations. When the outside temperature is low, decreasing temperature leads to increased energy consumption because the building needs to provider greater amounts of heating. When the outside temperature is high, increasing temperature leads to increased energy consumption because the building needs to provide greater amounts of cooling. This is a general phenomenon, and is empirically observed with the real data shown in Figure 1, which was collected from Sutardja Dai Hall.

Based on the figure, we might guess that when the outside temperature is below 59°F, the relationship between energy consumption and outside temperature is given by

$$E = a \cdot T + b + \epsilon,$$

where a, b are unknown constants, and ϵ is a zero-mean/finite-variance random variable that is independent of T. What is the method of moments estimate of $a, b, \sigma^2(\epsilon)$ if we have n measurements of (T_i, E_i)? There are two useful hints for this problem. The first is to use the random variable
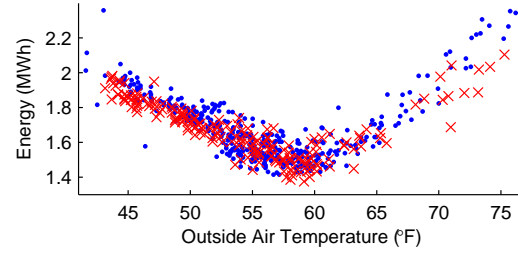
$$\epsilon = E - a \cdot T - b.$$

Figure 1: Real data (right) from Sutardja Dai Hall (left) is shown. The data marked with dots and crosses denote measurements made running two different versions of software on the heating, ventilation, and air-conditioning (HVAC) system in Sutardja Dai Hall.

The second is to use the moments $\mu_1(\epsilon)$ and $\mu_1(\epsilon \cdot T)$ in step 1, instead of the moments $\mu_1(\epsilon)$ and $\mu_2(\epsilon)$.

Proceeding with the four steps, gives

1. Recall that $\mu_1(\epsilon) = 0$ since ϵ is zero-mean, and $\mu_1(\epsilon \cdot T) = 0$ since ϵ is zero-mean and independent of T. The moments are:

$$0 = \mu_1(\epsilon) = \mu_1(E) - a\mu_1(T) - b$$
$$0 = \mu_1(\epsilon \cdot T) = \mu_1(ET) - a\mu_2(T) - b\mu_1(T);$$

2. This is a linear system of equations, and so in matrix notation this is

$$\begin{bmatrix} \mu_1(T) & 1 \\ \mu_2(T) & \mu_1(T) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \mu_1(E) \\ \mu_1(ET) \end{bmatrix}$$

Solving for a, b gives

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{(\mu_1(T))^2 - \mu_2(T)} \begin{bmatrix} \mu_1(T) & -1 \\ -\mu_2(T) & \mu_1(T) \end{bmatrix} \begin{bmatrix} \mu_1(E) \\ \mu_1(ET) \end{bmatrix}$$

Simplifying this yields

$$a = \frac{\mu_1(E)\mu_1(T) - \mu_1(ET)}{(\mu_1(T))^2 - \mu_2(T)}$$
$$b = \frac{\mu_1(ET)\mu_1(T) - \mu_1(E)\mu_2(T)}{(\mu_1(T))^2 - \mu_2(T)}.$$

3. An estimator of the remaining moments is

$$\hat{\mu}_1(E) = \frac{1}{n} \sum_{i=1}^{n} E_i$$

$$\hat{\mu}_1(T) = \frac{1}{n} \sum_{i=1}^{n} T_i$$

$$\hat{\mu}_1(ET) = \frac{1}{n} \sum_{i=1}^{n} E_i T_i$$

$$\hat{\mu}_2(T) = \frac{1}{n} \sum_{i=1}^{n} T_i^2;$$

4. Substituting this estimator gives

$$\hat{a} = \frac{(\frac{1}{n} \sum_{i=1}^{n} E_i)(\frac{1}{n} \sum_{i=1}^{n} T_i) - \frac{1}{n} \sum_{i=1}^{n} E_i T_i}{(\frac{1}{n} \sum_{i=1}^{n} T_i)^2 - \frac{1}{n} \sum_{i=1}^{n} T_i^2}$$

$$\hat{b} = \frac{(\frac{1}{n} \sum_{i=1}^{n} E_i T_i)(\frac{1}{n} \sum_{i=1}^{n} T_i) - (\frac{1}{n} \sum_{i=1}^{n} E_i)(\frac{1}{n} \sum_{i=1}^{n} T_i^2)}{(\frac{1}{n} \sum_{i=1}^{n} T_i)^2 - \frac{1}{n} \sum_{i=1}^{n} T_i^2}.$$

However, this is a clumsy derivation for an estimator of a, b. In coming lectures, we will learn simpler approaches for constructing such *linear models*.