1 A dash on moments: moments, moment-generating functions, and method-of-moments estimators

moments.tex and moments.pdf, October 13, 2010

What is the mean and variance of a distribution? They are ways to characterize the distribution. That is, ways to summarize important information about a distribution.

Obviously, one can find out anything about a distribution if one knows its density function; however, summary statistics can tell a lot. For example, knowing the mean and variance of a distribution tells us a lot about the distribution.

Put simply, the mean and variance of a distribution are its first and second *moments*

More formally, If X is a random variable with some density $f(x)$, the r^{th} moment of X, typically denoted μ'_r, is **defined** as

$$\mu'_r = E[x^r]$$

For example,

$$\text{if } r = 1,\ \mu'_1 = E[x] \equiv \mu_x$$

$$\text{if } r = 2,\ \mu'_2 = E[x^2]$$

μ'_1 and μ'_2 are the first and second (raw) moments of the distribution

If X is a random variable with some density $f(x)$, the rth **central** moment of X about μ_x, typically denoted μ_r, is **defined** as

$$\begin{aligned}\mu_r &= E[(x - \mu'_1)^r] \\ &= E[(x - \mu_x)^r]\end{aligned}$$

For example,

$$\text{if } r = 1,\ \mu_1 = E[x - E[X]] = 0$$

$$\text{if } r = 2,\ \mu_2 = E[(x - E[X])^2] = var[X]$$

μ_1 and μ_2 are the first and second central moments of the distribution

What is the third central moment?

$$\mu_3 = E[(x - \mu_x)^3]$$

What do we call this moment? *skewness*. It is a measure, sort of, of how skewed the distribution is. Can you find this moment for the uniform distribution on the 0 to 1 interval.[1]

What is the forth central moment?

$$\mu_4 = E[(x - \mu_x)^4]$$

What do we call this moment? *Kurtosis*: a measure of the flatness of the distribution near its center. Can you find this moment for the uniform distribution on the 0 to 1 interval.

If one knows all of a density function's moments, one knows the density function.

[1] But as MGB point out, page 76 third edition, the third central moment is often a poor guide to how shewed is the distribution. It is the case that for symetrical distributions the third central moment is zero (if the distribution is symetrical all odd central moments are zero), but $\mu_3 = 0$ does not imply the distribution is symmetrical. $\mu_3 = 0$ for Figure 10 on page 76 of MGB and the distribution looks skewed, at least to me.

1.1 some examples of moments

Consider the moment of the Bernoulli distribution

$$f(x;p) = \begin{cases} p^x(1-p)^{1-x} & \text{if} \quad x=0 \\ p^x(1-p)^{1-x} & \text{if} \quad x=1 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p \equiv \theta < 1$. For the Bernoulli

$$p = \mu_x = \text{ first population moment around } 0$$

What is the second, central moment around μ_x?

For the normal distribution

$$f(x; \mu_x, \sigma_x^2) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(\frac{1}{2\sigma_x^2})(x-\mu_x)^2}$$

the first population moment around 0 is μ_x and the second, central moment is σ_x^2.

What are the third and forth central moments?

For the Poisson distribution λ is the first population moment and the second central moment.

Note that population moments do not always correspond to the parameters in the density function, but often they do.

1.2 Moment generating functions

In spite of the fact that $E[(x - E[X])^r] = \int_{-\infty}^{+\infty} [(x - E[X])^r] f(x) dx$ can, in theory, be directly used to derive central moments, often it is very difficult to directly solve this integral.

For this reason, someone, Doctor Moment, created (discovered) *moment generating functions*; moment generating functions help us derive the moments of a distribution (if someone has not already done it for us)

The moment-generating function, $m_X(t)$, is defined as[2]

$$m_X(t) \equiv E[e^{tX}]$$

where X is a random variable.

If X is a continuous random variable

$$m_X(t) = E[e^{tX}] = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx$$

If X is a discrete random variable

$$m_X(t) = E[e^{tX}] = \sum_j e^{tx_j} f_X(x_j)$$

[2] it derives moments, not central moments. Note that $e^{tx} = 1 + tx + \frac{t^2 x^2}{2!} + \frac{t^3 x^3}{3!} + ...$ so $m_X(t) = E[e^{tX}] = 1 + t\mu_1 + \frac{t^2 \mu_2}{2!} + \frac{t^3 \mu_3}{3!} + ...$

If one knows $m_X(t)$, one can obtain $E[X] = \mu_x$ and $E[X^2]$. One can then use $E[X]$ and $E[X^2]$ to obtain σ_x^2 because $\sigma_x^2 = E[X^2] - (E[X])^2$

How does one derive $E[X]$ and $E[X^2]$ from $m_X(t)$?

It can be shown that

$$E[X] = \frac{\partial m_X(0)}{\partial t} = m'_X(0)$$

That is, the first moment around zero is the derivative of the moment-generating function wrt t evaluated at 0, and

$$E[X^2] = m''_X(0)$$

And $E[X^3] = m'''_X(0)$

Note that not all density functions have moment-generating functions since $m_X(t)$ is the integral $\int_{-\infty}^{+\infty} e^{tx} f_X(x)dx$, $m_X(t)$ only exists if the integral is well defined.

One can think of $m_X(t)$ as an alternative way (in comparision to $f_X(x)$ and $F_X(x)$) to specify the density function for a rv X; that is, with knowledge of $m_X(t)$ one can, at least in theory, determine $F_X(x)$.[3]. And, for some distributional problems in statistics, starting with $m_X(t)$ might be the way to go.

Choose some density function, derive its moment-generating function, and then use your moment-generating function to derive three moments of your density function.

I will choose the Poisson, which is discretely distributed, so $m_X(t) = E[e^{tX}] = \sum_j e^{tx_j} f_X(x_j)$. Since for the Poisson, $f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

$$\begin{aligned} m_X(t) &= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{e^{tx}\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} \end{aligned}$$

but, $\sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} = e^{\lambda e^t}$. (Why? because my math software says so), so

$$\begin{aligned} m_X(t) &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} \end{aligned}$$

The first derivative is

$$\frac{dm_X(t)}{dt} = \frac{d(e^{-\lambda}e^{\lambda e^t})}{dt} = \lambda e^t e^{\lambda e^t} e^{-\lambda}$$

evaluated at $t = 0$ is $\lambda = \mu_X \equiv E[X]$, the correct answer. The second derivative is

[3]Theorem 7 in MGB, third edition, page 76, states, "Let X and Y be two random variables with densities $f_X(.)$ and $f_Y(.)$, respectively. Suppose that $m_X(t)$ and $m_Y(t)$ both exist and are equal for all t in the interval $-h < t < h$ for some $h > 0$. Then the two cummulative density functions $F_X(.)$ and $F_Y(.)$ are equal."

$$
\begin{aligned}
\frac{d^2 m_X(t)}{dt^2} &= \frac{d(\lambda e^t e^{\lambda e^t} e^{-\lambda})}{dt} = \lambda e^t e^{\lambda e^t} e^{-\lambda} + \lambda^2 e^{\lambda e^t} e^{2t} e^{-\lambda} \\
&= \lambda \left(e^{\lambda e^t}\right) \left(e^{-\lambda}\right) \left(e^t + \lambda e^{2t}\right)
\end{aligned}
$$

evaluated at $t = 0$, $\lambda \left(e^{\lambda e^t}\right) \left(e^{-\lambda}\right) \left(e^t + \lambda e^{2t}\right) = \lambda \left(e^{\lambda e^0}\right) \left(e^{-\lambda}\right) \left(e^0 + \lambda e^0\right) = \lambda e^{\lambda} (\lambda + 1) e^{-\lambda} = E[x^2] \equiv \mu_2'$. But, we know that the second central moment $\mu_X \equiv \sigma_x^2 = E[X^2] - (E[X])^2$, so $\mu_X \equiv \sigma_x^2 = \lambda e^{\lambda} (\lambda + 1) e^{-\lambda} - \lambda^2 = \lambda$. WOW, it worked.

Skating on thin ice, let's look for the third moment of the Poisson

$$
\begin{aligned}
\frac{d^3 m_X(t)}{dt^3} &= \frac{d(\lambda \left(e^{\lambda e^t}\right) \left(e^{-\lambda}\right) \left(e^t + \lambda e^{2t}\right))}{dt} = \lambda e^t e^{\lambda e^t} e^{-\lambda} + 3\lambda^2 e^{\lambda e^t} e^{2t} e^{-\lambda} + \lambda^3 e^{\lambda e^t} e^{3t} e^{-\lambda} \\
&= \lambda e^t e^{\lambda e^t} e^{-\lambda} + 3\lambda^2 e^{\lambda e^t} e^{2t} e^{-\lambda} + \lambda^3 e^{\lambda e^t} e^{\frac{3}{2}(2t)} e^{-\lambda} \\
&= \lambda \left(e^{\lambda e^t}\right) \left(e^{-\lambda}\right) \left(e^t + 3\lambda e^{2t} + \lambda^2 e^{3t}\right)
\end{aligned}
$$

evaluated at $t = 0$, $\lambda \left(e^{\lambda e^t}\right) \left(e^{-\lambda}\right) \left(e^t + 3\lambda e^{2t} + \lambda^2 e^{3t}\right) = \lambda \left(e^{\lambda e^0}\right) \left(e^{-\lambda}\right) \left(e^0 + 3\lambda e^0 + \lambda^2 e^0\right) = \lambda e^{\lambda} e^{-\lambda} \left(3\lambda + \lambda^2 + 1\right) = \mu_3' \equiv E[X^3]$, the third moment of the Poisson.

So, the third central moment of the Poisson, μ_3, is $E[(X - E[X])^3] = E[X^3] - (E[X])^3 - 3E[X]E[X^2] + 3(E[X])^2 E[X] = \lambda e^{\lambda} e^{-\lambda} \left(3\lambda + \lambda^2 + 1\right) - \lambda^3 - 3\lambda\lambda e^{\lambda} (\lambda + 1) e^{-\lambda} + 3\lambda^2 \lambda = \lambda = \mu_3$. WOW, so the third central moment is also λ. Are you surprised? Is every moment of the Poisson λ?[4]

1.2.1 Some additional stuff on moments:

An interesting, and difficult, question is whether one can, in theory, derive $f_X(.)$ if one only knows the sequence of moments for the random variable X. Recollect that we noted above that one can derive $f_X(.)$ with only knowledge of $m_X(t)$, but that is different than the question asked here. Another way of asking the current question is can one recover $m_X(t)$ only knowing the sequence of moments generated by it. Here, I provide only an incomplete answer.

First off, knowing only a few of the moments of how X is distributed doesn't tell all; for example, many many density function all have the same first four moments.

MGB say the following cryptic things: "the **entire** [bold added] set of moments $(\mu_1', \mu_2', ...)$ will **ordinarily** [bold added] determine the distribution exactly, ..." Later they say, "there exists a sequence of moments for which there is

[4] No. The forth central moment is $\lambda + 3\lambda^2$(MGB page 539)

an infinite (nondenumerable) collection of different distribution functions having these same moments. In **general**, [bold added] a sequence of moments $\mu_1^{'}, \mu_2^{'}, ...$ does not determine a unique distribution function. The critical words seem to be "ordinarly" and "in general". I take all this to mean that if we know $\mu_1^{'}, \mu_2^{'}, ...$, this will often be enough to identify the distribution, but not always.

An issue that I am still unclear on is how many moments are there? I believe, today, that the number is infinite but countable; that is $E[(X - E[X])^r] = \mu_r$ will exist for all integer r. However, often many of the central moments, the μ_r,will be zero. For example, MGB report that for the univariate normal, $\mu_r = 0 \forall r$ odd, and $\mu_r = \frac{r!}{(r/2)!} \frac{\sigma^r}{2^{r/2}} \; \forall r$ even, so the normal has an infinite number of well-defined central moments, but half of them are zero.

1.3 Estimating moments

Make sure you undertand the distinction between estimating the parameters of a distribution and estimating the moments of a distribution.

Often the estimation problem is one of estimating the moments of a distribution. For example, we estimate the mean and variance of a distribution.

Assume X is a random variable from some family of distributions

$$f_X(x)$$

Consider two cases:

1. We know the form (family) of $f_X(x : \theta)$, but do not know the specific values of the θ parameters in the population, so want to estimate θ.

2. We don't know anything except that X is a random variable with finite mean and variance. We want to estimate μ_x and σ_x^2.

The method of moments estimators, unlike many other estimators, are applicable to both of these situations

1.3.1 Method of moment estimators

One possibility for an estimator for the r^{th} population moment is the r^{th} sample moment.

This estimator is called the *method of moments* estimator because one uses sample moments to estimate population moments.

That is, use the first sample moment around zero, $\bar{x}$, as an estimate of μ_x (the sample mean is the *method of moments* estimator of the population mean)

One could use the use the second sample moment around $\bar{x}$,$\tilde{s}_x^2$, as an estimator of σ_x^2.

Notationally: M for sample moment and m for population moment

$$M_1^{'} = E[X_i^1] = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

as an estimate of μ_x, where $M_1^{'}$ denotes the first sample moment around zero. (The 1 superscript here on X denotes raised to the power of one.)

Use

$$M_2 = E[(X_i - \bar{X})^2] = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \tilde{s}_x^2$$

as an estimator for σ_x^2, where M_2 denotes the second sample moment around the sample mean.

One can show, maybe se showed it in some earlier notes, that

$$E[\tilde{s}_x^2] \neq \sigma_x^2$$

so the method of moments estimator of $\tilde{s}_x^2$ is biased, whereas, $\bar{X}$ is the method of moments estimator of μ_x, which is an unbiased estimator.

Further note that the method of moments method of estimation does not require knowledge of the form of $f_X(x;\theta)$, making it usable in a wide range of situations.

However, to convert estimates of the population moments into estimates of θ, one needs to know the form of $f_X(x;\theta)$.