Lecture 3 - Moments

A. The Expectation Operator and Moments

It is often inconvenient to use the entire pmf or pdf of a random variable. As such, we want a few numbers to tell us about the pmf/pdf. Additionally, these numbers, moments of the distributions, often correspond to physical observations.

(i) Expectation Operator

The expectation operator is motivated by the conventional idea of an average. It is defined as:

$$E[X] = \sum_i x_i \, p(x_i) \qquad \text{discrete}$$

$$E[X] = \int_{-\infty}^{\infty} x \, f(x) \, dx \qquad \text{continuous}$$

$m = E[X]$ is the expectation or mean of X

The expectation is a linear operator:

$$E[aX + bY] = a\,E[X] + b\,E[Y]$$

(ii) Moments

The mean is not the only piece of useful information that can be determined using the expectation operator. The n^{th} moment of the pmf/pdf is defined as:

$$m_n = E[x^n] \qquad m = m_1$$

$$E[X^n] = \sum_i x_i^n \, p(x_i) \qquad \text{discrete}$$

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x)\, dx \qquad \text{continuous}$$

Also useful are the n^{th} centered moments:

$$E[(X-m)^n] = \sum_i (x_i - m)^n \, p(x_i) \qquad \text{discrete}$$

$$E[(X-m)^n] = \int_{-\infty}^{\infty} (x-m)^n f(x)\, dx \qquad \text{continuous}$$

The most common centered moments have names:

$n=2$: variance, $\sigma^2 = E[(X-m)^2] = E[X^2] - m^2$

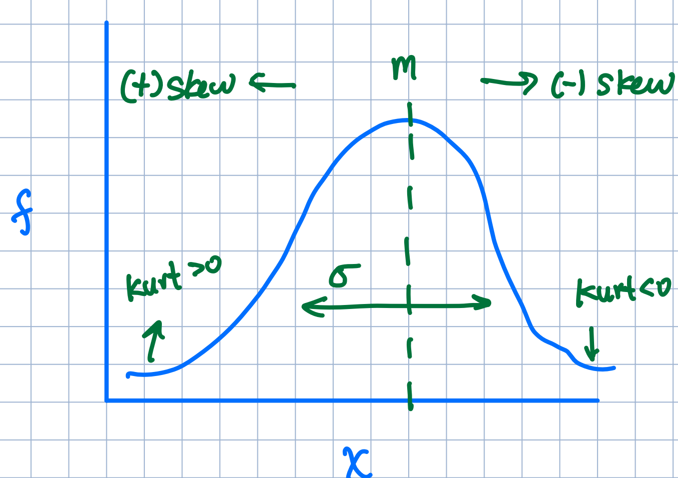
$n=3$: skewness, $E[(X-m)^3]/\sigma^3$

$n=4$: kurtosis, $E[(X-m)^4]/\sigma^4$

skewness & kurtosis are scaled, too

Recall also that $\sqrt{\sigma^2} = \sigma$ is the standard deviation.

Example: Intuition about what centered moments mean



m: center of mass

σ: width

skew: symmetry about m

kurt: "fat" or "thin" tails
kurt = 0 for normal

It is sometimes convenient to define a normalized R.V.

$$Y = \frac{X-m}{\sigma}$$

This has the nice properties that

$$E[Y]=0\ , \quad E[Y^2]=1\ , \quad \text{skew} = E[Y^3]$$
$$\text{kurt} = E[Y^4]$$

B. Characteristic Functions

Sometimes pmfs, pdfs, and cdfs are not very convenient for computing needed quantities. These include the calculation of moments and the calculation of pdfs of functions of RVs.

Characteristic functions (and related) are very useful tools for doing the above. They also provide some insight into moments and some fundamental theorems that make them worth learning. They are also very common in stat thermo.

We define the characteristic function of a random variable to be the Fourier transform of the pdf

$$\phi(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f(x)\, dx\,. \qquad i=\sqrt{-1}$$

Alternatively, $\phi(\omega)$ can be written using the expectation operator

$$\phi(\omega) = E[e^{i\omega x}]$$

The characteristic function has several important properties:

(1) It's inverse is $f(x)$.

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega x}\phi(\omega)\, d\omega$$

There are different FT conventions. Note this one.

The mapping between $f(x)$ and $\phi(\omega)$ is one-to-one, meaning $\phi(\omega)$ contains all the same information as $f(x)$.

(2) The derivatives of $\phi(\omega)$ are related to the moments.

$$\phi^{(n)}(\omega=0) = i^n E[x^n] \qquad \phi^{(n)}(\omega) = \frac{d^n\phi}{d\omega^n}$$

← evaluate at $\omega = 0$ ← not centralized

Quick proof for 1st moment:

$$\frac{d\phi}{d\omega} = \frac{d}{d\omega} E\left[e^{i\omega x}\right] = E\left[ix e^{i\omega x}\right]$$

$$\left.\frac{d\phi}{d\omega}\right|_{\omega=0} = E[ix] = i\,E[x]$$

(i) Other Transforms

The characteristic function is not the only transform that you will see. Let me briefly discuss a few others.

The moment generating function (mgf) is given by

$$M(s) = E\left[e^{sx}\right] = \int_{-\infty}^{\infty} e^{sx} f(x)\,dx$$

← two-sided Laplace transform

- The characteristic function is a "Wick rotation" of the mgf: $\phi(\omega) = M(i\omega)$
- Derivatives of the mgf are the moments.

$$M^{(n)}(s=0) = E[x^n] = m_n$$

← no pesky i

- The Taylor series of the mgf contains the moments.

$$M(s) = 1 + E[x]s + E[x^2]\frac{s^2}{2!} + E[x^3]\frac{s^3}{3!} + \dots$$
$$= 1 + m_1 s + m_2\frac{s^2}{2!} + m_3\frac{s^3}{3!} + \dots$$

If the mgf and the pdf are one-to-one, and the mgf is specified by the moments, then specifying all of the moments is equivalent to specifying the pdf. This is called the "moment problem". There are some important mathematical conditions that must be satisfied though for this to be true. (Carleman's condition)

The cumulant generating function (cgf) is given by

$$K(s) = \ln E[e^{sx}] = \ln M(s)$$

- Derivatives of the cgf are called cumulants

$$\left.\frac{d^n K}{ds^n}\right|_{s=0} = K^{(n)}(0) = k_n$$

$$K(s) = k_1 s + k_2\frac{s^2}{2!} + k_3\frac{s^3}{3!} + \dots$$

cumulants have "decreasing informational content" at higher order.

k_1 is the mean, k_2 is the variance, and k_3 is the 3rd central moment. H.O.T are not in general central moments, but they are polynomial functions of them. Cumulants have a special significance in statistical thermodynamics related to "connectedness" in n-body correlations. We'll hopefully talk about this soon.

The probability generating function (pgf) is the version for discrete RVs. It is defined as:

$$G(z) = E\left[z^{x}\right] = \sum_{n=0}^{\infty} z^{n} \overset{\text{pmf}}{p(X=n)}$$

$G(z^{-1})$ is the z-transform of the pmf.

We wont use the pgf much (at all?). I include it for completeness. Details are in the appendix.

C. Characteristic function of normal distribution

The normal distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^{2}\right]$$

Use the definition of the characteristic function

$$\phi(\omega) = E\left[e^{i\omega x}\right] = \int_{-\infty}^{\infty} e^{i\omega x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^{2}\right] dx$$

change of variables: $y = \frac{x-m}{\sigma} \rightarrow x = y\sigma + m, \; dy = \frac{dx}{\sigma}$

$$\phi(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[i\omega(y\sigma + m)\right] \exp\left(-\frac{y^{2}}{2}\right) dy$$

$$= \frac{1}{\sqrt{2\pi}} e^{i\omega m} \int_{-\infty}^{\infty} \exp(i\omega\sigma y) \exp\left(-y^{2}/2\right) dy$$

using a F.T. table (Mathematica)

$$\int_{-\infty}^{\infty} e^{i\omega\sigma y} \exp\left(-y^{2}/2\right) dy = \sqrt{2\pi}\, \exp\left(-\frac{\omega^{2}\sigma^{2}}{2}\right)$$

putting it together

$$\phi(\omega) = \frac{1}{\cancel{\sqrt{2\pi}}} e^{i\omega m} \cdot \cancel{\sqrt{2\pi}}\, e^{-\omega^2\sigma^2/2}$$

$$\boxed{\phi(\omega) = \exp(i\omega m - \omega^2\sigma^2/2)}$$

(i) what is the mgf?

need to "pull out" an $i\omega = s$ or $\omega = -is$

$$M(s) = \exp\left[sm - (-is)^2\sigma^2/2\right]$$

$$= \exp\left(sm - (-s^2)\sigma^2/2\right) = \exp\left(sm + s^2\sigma^2/2\right)$$

$$\boxed{M(s) = \exp\left(sm + s^2\sigma^2/2\right)}$$

(ii) What are the moments?

$$\frac{dM}{ds} = \exp(sm + s^2\sigma^2/2)\cdot(m + s\sigma^2) = M(m + s\sigma^2)$$

$$\frac{d^2M}{ds^2} = M(m+s\sigma^2)^2 + M\sigma^2$$

$$\frac{d^3M}{ds^3} = M(m+s\sigma^2)^3 + 2M(m+s\sigma^2)\sigma^2 + M(m+s\sigma^2)\sigma^2$$

$$\frac{d^4M}{ds^4} = M(m+s\sigma^2)^4 + 3M(m+s\sigma^2)^2\sigma^2 + 2M\sigma^2(m+s\sigma^2)^2$$
$$+ 2M\sigma^4 + M(m+s\sigma^2)^2\sigma^2 + M\sigma^4$$

$$m_1 = \left.\frac{dM}{ds}\right|_{s=0} = \cancelto{1}{M(0)}\cdot m = \boxed{m}$$

$$m_2 = \left.\frac{d^2M}{ds^2}\right|_{s=0} = \cancelto{1}{M(0)} \cdot m^2 + \cancelto{1}{M(0)}\,\sigma^2 = \boxed{m^2 + \sigma^2}$$

$$m_3 = \left.\frac{d^3M}{ds^3}\right|_{s=0} = M(0)\,m^3 + 2M(0)\,m\sigma^2 + M(0)\,m \cdot \sigma^2$$

$$= m^3 + 2m\sigma^2 + m\sigma^2 = \boxed{m^3 + 3m\sigma^2}$$

$$m_4 = \left.\frac{d^4M}{ds^4}\right|_{s=0} = m^4 + 3m^2\sigma^2 + 2m^2\sigma^2 + 2\sigma^4 + m^2\sigma^2 + \sigma^4$$

$$\boxed{m_4 = m^4 + 6m^2\sigma^2 + 3\sigma^4}$$

Note that these are polynomials of m and σ only

(iii) What is the cgf? What are the cumulants?

$$k(s) = \ln M(s) = \boxed{sm + s^2\sigma^2/2}$$

$$\frac{dk}{ds} = m + s\sigma^2 \qquad \frac{d^2k}{ds^2} = \sigma^2 \qquad \frac{d^nk}{ds^n} = 0 \quad \text{for } n \geq 3$$

$$k_1 = \left.\frac{dk}{ds}\right|_{s=0} = \boxed{m}$$

$$k_2 = \left.\frac{d^2k}{ds^2}\right|_{s=0} = \boxed{\sigma^2}$$

$$\boxed{k_3 = k_4 = k_n = 0 \ , \ n \geq 3}$$

D. Appendix

(i) Additional property of the characteristic function

The sum of independent RVs are products.

$Z = X + Y$ where X and Y are independent RVs "factorization property"

$$\phi_Z(\omega) = \phi_X(\omega)\,\phi_Y(\omega)$$ * See proof below

In real-space, this results in a convolution

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)\, f_Y(y)\, dy$$ ← property of Fourier transforms of products

* Quick Proof:

$$\phi_Z(\omega) = E\left[e^{i\omega z}\right] = E\left[e^{i\omega(x+y)}\right]$$

$$= E\left[e^{i\omega x} e^{i\omega y}\right] \underset{\text{independence}}{=} E\left[e^{i\omega x}\right] E\left[e^{i\omega y}\right]$$

$$= \phi_X(\omega)\,\phi_Y(\omega)$$

(ii) Details of the probability generating function

The pgf can be done using a "z-transform", a discrete analogue of the Laplace transform

$$Z[f_n] = \sum_{n=0}^{\infty} z^{-n} f_n$$ Definition of the z-transform

Derivatives of G give moments

$$\left.\frac{d^nG}{dz^n}\right|_{z=1} = G^{(n)}(1)$$

"n^{th} factorial moment"

$$= E[x(x-1)(x-2)\dots(x-n+1)].$$

One can get probabilities from derivatives, too

$$\frac{1}{n!}\left.\frac{d^nG}{dz^n}\right|_{z=0} = \frac{G^{(n)}(0)}{n!} = p(x_n=n) = p(n)$$

(iii) Why characteristic function?

- Fourier transforms are tabulated.
- The Laplace transform isn't always guaranteed to exist (fat tails), but the Fourier transform is.
- Take one transform → derivatives after that

(iv) Relationship between moments and cumulants

m_i : moments $\quad m_1 = E[x]$, $m_2 = E[x^2]$

k_i : cumulants

$k_1 = m_1$ mean

$k_2 = m_2 - m_1^2$ variance

$k_3 = m_3 - 3m_2m_1 + 2m_1^3$ skewness

$k_4 = m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4$ excess kurtosis

$m_1 = k_1$

$m_2 = k_2 + k_1^2$

$m_3 = k_3 + 3k_2k_1 + k_1^3$

$m_4 = k_4 + 4k_3k_1 + 3k_2^2 + 6k_2k_1^2 + k_4^4$