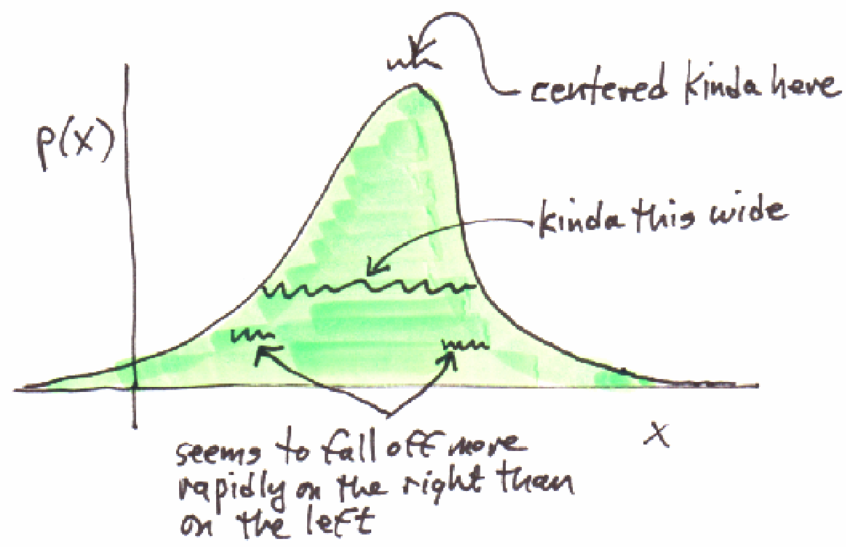
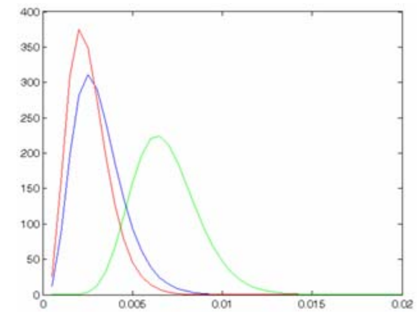
Opinionated Lessons in Statistics

by Bill Press

#7 Central Tendency and Moments

We are often interested in distributions that have some kind of localization (because why would we be interested if they didn't?)





We already saw the beta distribution with α, $\beta > 0$ as an example on the interval [0,1], and the Towne family example (not any simple function). We'll see more examples soon.

Suppose we want to summarize *p(x)* by a single number *a*, its "value". Let's find the value *a* that minimizes the mean-square discrepancy of the "typical" value *x*:

Recall expectation notation:

$$\langle \text{anything} \rangle \equiv \int_x (\text{anything})\, p(x) dx$$

i.e., the weighted average (or mean) of "anything", weighted by the probable values of x. Expectation is linear over "anything" (sums, constants times, etc.).

$$\text{minimize:} \quad \Delta^2 \equiv \left\langle (x-a)^2 \right\rangle = \left\langle x^2 - 2ax + a^2 \right\rangle$$
$$= (\left\langle x^2 \right\rangle - \langle x \rangle^2) + (\langle x \rangle - a)^2$$

This is the variance Var(x), but all we care about here is that it doesn't depend on a.

(in physics this is called the "parallel axis theorem")

The minimum is obviously $a = \langle x \rangle$. (Take derivative wrt a and set to zero if you like mechanical calculations.)

Why mean-square? Why not mean-absolute? Try it!

$$\Delta = \langle |x-a| \rangle = \int_{-\infty}^{\infty} |x-a|\, p(x) dx$$

$$= \int_{-\infty}^{a} (a-x)\, p(x) dx + \int_{a}^{\infty} (x-a)\, p(x) dx$$

So,

$$0 = \frac{d\Delta}{da} = \int_{-\infty}^{a} p(x) dx + 0 - \int_{a}^{\infty} p(x) dx + 0$$

Integrand at a

$$\Rightarrow$$

$$\int_{-\infty}^{a} p(x) dx = \int_{a}^{\infty} p(x) dx = \tfrac{1}{2}$$

$\Rightarrow a$ is the median value

Mean and median are both “measures of central tendency”.

Higher moments, centered moments are conventionally defined by

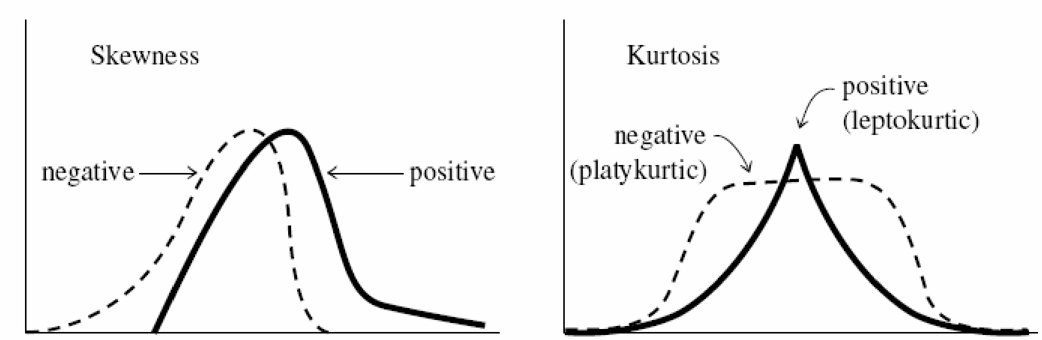
$$\mu_i \equiv \langle x^i \rangle = \int x^i \, p(x) dx$$
$$M_i \equiv \left\langle (x - \langle x \rangle)^i \right\rangle = \int (x - \langle x \rangle)^i \, p(x) dx$$

The centered second moment M_2, the variance, is by far most useful

$$M_2 \equiv \mathrm{Var}(x) \equiv \left\langle (x - \langle x \rangle)^2 \right\rangle = \left\langle x^2 \right\rangle - \langle x \rangle^2$$
$$\sigma(x) \equiv \sqrt{\mathrm{Var}(x)}$$

"standard deviation" summarizes a distribution's half-width (r.m.s. deviation from the mean)

Third and fourth moments also have "names"



But generally wise to be cautious about using high moments. Otherwise perfectly good distributions don't have them at all (divergent). And (related) it can take a lot of data to measure them accurately.

Mean and variance are additive over independent random variables:

$$\overline{(x+y)} = \bar{x} + \bar{y} \qquad \mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(x)$$

note “bar” notation, equivalent to < >

Certain combinations of higher moments are also additive. These are called semi-invariants or cumulants.

$$I_2 = M_2 \qquad I_3 = M_3 \qquad I_4 = M_4 - 3M_2^2$$

$$I_5 = M_5 - 10M_2M_3 \qquad I_6 = M_6 - 15M_2M_4 - 10M_3^2 + 30M_2^3$$

How to derive these? If you are a little bit sophisticated about probability (from a previous course?) look at Wikipedia “Cumulant”. It’s very cool!

Skew and kurtosis are dimensionless combinations of semi-invariants

$$\mathrm{Skew}(x) = I_3 / I_2^{3/2} \qquad \mathrm{Kurt}(x) = I_4 / I_2^2$$

Factoid: A Gaussian has all of its semi-invariants higher than I_2 equal to zero. A Poisson distribution has all of its semi-invariants equal to its mean.