STA257

Neil Montgomery | HTML is official | PDF versions good for in-class use only
Last edited: 2016-11-28 14:46

key point from end of last class, restated

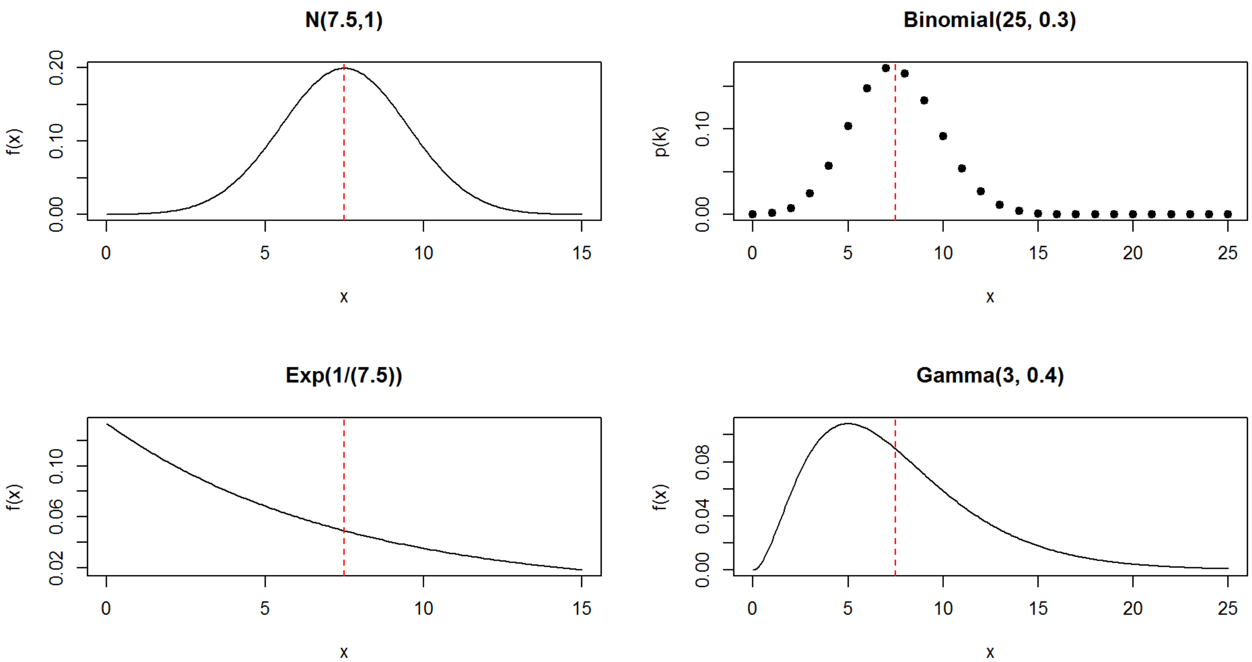
I did: $X_1, \dots, X_n$ i.i.d. implies = .$

A different (better?) approach could have started with a more fundamental:

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

which is true no matter what. (No independence required and all expected values can be different.)

looking back: getting a sense of $E(X) = 7.5$



putting a number on variation

Expected value is a measure of "location", but random variables with the same "location" can be quite different.

Consider the coin tossing game with $E(Y) = 0$:

$$P(Y = y) = \begin{cases} 0.5 & : y = 100 \\ 0.5 & : y = -100 \end{cases}$$

One thing leads to another. Family trees are compared and contrasted, and after more than a few *schnapps* things get interesting:

$$P(Y_2 = y) = \begin{cases} 0.5 & : y = 1000 \\ 0.5 & : y = -1000 \end{cases}$$

Still, $E(Y_2) = 0$. But the values of Y_2 are more spread out.

variance

One way to measure spread is to use the *variance* of X, defined as:
$\mathrm{Var}(X) = E\left[(X - E(X))^2\right]$.

This is a use of $E(g(X))$ with $g(x) = (x - E(X))^2$.

Very useful, and almost always the way to perform the actual calculation.

$$\begin{aligned}\mathrm{Var}(X) &= E\left(X^2 - 2XE(X) + E(X)^2\right)\\ &= E\left(X^2\right) - 2E(X)E(X) + E(X)^2\\ &= E\left(X^2\right) - E(X)^2.\end{aligned}$$

Note: existence of $\mathrm{Var}(X)$ requires existence of both $E(X^2)$ and $E(X)$.

Fun fact: existence of $E(X^2)$ implies the existence of $E(X)$.

examples, sketches, exercises, hints

$X \sim \text{Bernoulli}(p)$ ($p(1-p)$)

$Z \sim N(0,1)$ (1)

$X \sim \text{Poisson}(\lambda)$ (λ) (uses a trick!)

Variance of $X = a$ constant.

Basic examples for exercise (answer): Exp(λ) ($1/\lambda^2$), Gamma (α/λ^2), Geometric ($(1-p)/p^2$) (trick: differentiate power series twice), Binomial ($np(1-p)$)(use Poisson trick).

$\text{Var}(a + bX)$, $\text{Var}(X + Y)$ (independent case)

$\text{Var}(a + bX) = b^2\text{Var}(X)$. Proof…

Example: $X \sim N(\mu, \sigma^2)$

When $X \perp Y$, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. Proof…

Actually independence is stronger than necessary. Only needed $E(XY) = E(X)E(Y)$; to be revisited.

variance of the "sample average"

This is a "grand" example of particular importance.

Suppose again $X_1, \ldots, X_n$ is i.i.d. with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$.. We already know $E\left(\overline{X}\right) = \mu$.

What about $\mathrm{Var}\left(\overline{X}\right)$?

standard deviation; notational conventions

Variance is in the square of the unit of measure of X.

"Standard deviation" is just:

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)},$$

and is a more practical number to use for descriptive purposes (but less practical for theoretical developments.)

A common abbreviation for $\mathrm{Var}(X)$ is σ^2 so that $\mathrm{SD}(X) = \sigma$.

the Russians are coming!

$E(X)$ and $E\left(X^2\right)$ provide information about X that limit its values and probabilities to some extent. Two examples are *Markov's* and *Chebyshev's* inequalities.

Theorem (Markov): If $X \geq 0$ has expected value $E(X)$, then:

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Proof: Much easier than in book…(!)

Theorem (Chebyshev): If $\mathrm{Var}(X) = \sigma^2$ and $E(X) = \mu$:

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

Proof: Apply Markov to the random variable $(X - \mu)^2$ and t^2.

Markov and Chebyshev examples

Consider $X \sim \mathrm{Exp}(5)$ and $t = 0.5$.

$E(X) = 1/5$ and $\mathrm{Var}(X) = 1/25$

$P(X \geq 0.5) = 0.082085 \leq 0.4$ (Markov)

$P(|X - 1/5| \geq 0.5) = 0.0301974 \leq 0.16$ (Chebyshev)

Our Russian friends more useful in theory than in practice.

a quick tour of covariance, correlation, and conditional expectation

covariance, and correlation

The quantity we saw in $\mathrm{Var}(X+Y)$:

$$E(XY) - E(X)E(Y) = E((X - E(X))(Y - E(Y)))$$

is called *covariance* or $\mathrm{Cov}(X, Y)$, and is a measure of *linear* association between the distributions of X and Y.

Note: $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

Note: Covariance is "multi-linear", which means *linear in both variables*.

meaning of "linear association" of X and Y

Essentially: when X exceeds $E(X)$, is Y likelier, or not, to exceed $E(Y)$?

Consider some examples with X and Y jointly uniform over these triangles:

Bounded by (0,0), (0,1), and (1,0). ($\mathrm{Cov}(X, Y) = -3/108$)

Bounded by (0,0), (0,1), and (-1,0). ($\mathrm{Cov}(X, Y) = 3/108$)

correlation

Covariance is in the multiple of the X and Y units. A "unitless" version of covariance is another measure of linear association called "correlation", defined as:

$$\frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Often denoted by ρ and often called "correlation coefficient."

Example (textbook section 4.3 E X A M P L E **F**) the ρ in the definition of the bivariate normal density is the correlation between X and Y.

conditional expectation given $Y = y$

First, recall the definition of *conditional density* for X given $Y = y$:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

This is a valid density, and we can consider the expected value of the random variable $X|Y = y$ with this density:

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx.$$

Analogous definition for the discrete case.

Nothing at all special about this.

recall the two dice example

Considering the sum and absolute difference:

		X_1										
		2	3	4	5	6	7	8	9	10	11	12
	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$
	1	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0
X_2	2	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0
	3	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0
	4	0	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	0
	5	0	0	0	0	0	$\frac{2}{36}$	0	0	0	0	0