W1211 Introduction to Statistics
Lecture 8

Wei Wang

Oct 1st, 2012

Variance of Random Variables

- We defined the concept of sample variance in the first chapter. Similarly we can define the variance of random variables, which is still a measure of deviation from the center.
- X has pmf $p(x)$ and expected value μ, then the variance of X, denoted as $V(X)$ or σ^2 is

$$V(X) = E[(X - \mu)^2]$$

  The standard deviation of X is

$$\sigma = \sqrt{\sigma^2}$$

Example

- A reader can check out at most 6 videos from a library at one time. Consider only those who check out videos, let X denote the number of videos checked out to a randomly selected individual. The pmf is

x	1	2	3	4	5	6
p(x)	.3	.25	.15	.05	.1	.15

- How to calculate variance and standard deviation of X?

Shortcut Formula for σ^2

- Similar to sample variance, we have a shortcut formula for variance of random variables too
- $$V(X) = \sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$$
- Proof:

$$\sigma^2 = \sum_D x^2 \cdot p(x) - 2\mu \cdot \sum_D x \cdot p(x) + \mu^2 \sum_D p(x) = E(X^2) - \mu^2$$

Variance of Linear Function of Random Variables

- How would variance change if we add a constant to the RV?

Variance of Linear Function of Random Variables

- How would variance change if we add a constant to the RV?
- How would variance change if we mupltiply it by a constant?

Variance of Linear Function of Random Variables

- How would variance change if we add a constant to the RV?
- How would variance change if we mupltiply it by a constant?
- $$V(aX + b) = a^2 \cdot V(X)$$

  and

  $$\sigma_{aX+b} = |a| \cdot \sigma_X$$

Discrete Probability Models

- We often conduct trials/experiments repeatedly. Today we will discuss probability models that allow us to answer questions regarding repeated trials. In particular we will be interested in series of *n* repeated trials of a random phenomena with two possible outcomes.

- Many popular discrete models are motivated by coin tosses, or more specifically a series of *n* Bernoulli trials. A series of *n* trials are Bernoulli trials if:
    1. The *n* trials are identical.
    2. The trials are independent (the outcome on any particular trial does not influence the outcome on any other trial).
    3. Each trial has two possible outcomes: success or failure.
    4. The probability of success, denoted by *p*, is the same for each trial (identical).

Binomial Experiment

- If in Bernoulli trials, the number of trials n is fixed in advance of the experiment. This experiment is called a binomial experiment.

Ex. The same coin is tossed successively and independently 10 times.

Ex. Suppose there are 50 colored socks in the drawer, of which 16 are red and the other 34 are blue. We are going to randomly draw 10 socks out of the drawer *without replacement*. We label the ith trial as a success if the ith sock is blue. (Is this a binomial experiment? What if it's *with replacement*?)

Ex. The previous example, what if we have 500,000 socks, of which 400,000 are blue. A sample of 10 socks are drawn without replacement.

$$P(\text{success on } 2 \mid \text{success on } 1) = 399{,}999/499{,}999 = .80000$$

$$P(\text{success on } 10 \mid \text{success on first } 9) = 399{,}991/499{,}991 = .799996 \approx .80000$$

A Rule of Thumb

- For drawing without replacement (*hypergeometric*), as the previous example suggests, although the trials are not exactly independent, the conditional probability differ so slightly from one another that for practical purposes the trials can be regarded as independent. Thus, to a very good approximation, the previous experiment is binomial with $n = 10$ and $p = .8$.

- As a rule of thumb: consider sampling without replacement from a dichotomous population of size N. If the sample size (number of trials) n is at most 5% of the population size, the experiment can be analyzed as though it were exactly a binomial experiment.

Binomial RV

- The binomial random variable X associated with a binomial experiment consisting of *n* trials is defined as

  X = the number of successes among the *n* trials.

- The pmf of a binomial rv X depends on the two parameters *n* and *p,* we denote the pmf by $b(x; n, p)$. The cdf will be denoted by

  $$P(X \leq x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p).$$

  Note that *x* can only take values in {0,1,…, n}.