EE 381 Probability & Statistic with Applications to Computing (Fall 2020)

Lecture 7
Discrete Random Variables (p4)

Duc H. Tran, PhD

Discrete Random Variable’s topics

- Discrete probability distributions
- Joint distributions
- Mathematical expectation
- Variance, Standardized random variables
- Moments and Moment generating function
- Covariance & correlation
- **Special Probability Distributions**

The Binomial Distribution

- Two teams, the Yankees and the Giants, play a *7*-game series. If the Yankees win each game with probability $p=0.5$ independently of any other game, what is the probability the Yankees win the series (win 4 games)?
- We need to use the binomial probability to calculate this probability.
- Binomial probabilities arise in numerous application, not just sports.

The Binomial Distribution

- Suppose that we have an experiment: tossing a coin or die **repeatedly**. Then each toss is called a *trial*.
- In any single *trial*, there will be a probability associated with a particular event. In some cases, this probability will not change from one trial to the next.

→ These trials are said to be *independent* and *identical*, and are often called **Bernoulli trials**.

The Binomial Distribution

- Let *p* be the probability that an event will happen in any single Bernoulli trial (called *the probability of success*)
- Then $q = 1 - p$ is the probability that the event will fail to happen in any single trial (called *the probability of failure*).
- The probability that the event will happen exactly *x* times in *n* trials is given by the following probability function:

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x} \qquad (*)$$

where the random variable *X* denotes the number of successes in *n* trials and $x = 0, 1, \ldots, n$.

The Binomial Distribution

- The discrete probability function (*) in previous slide is often called the binomial distribution since for $x = 0, 1, 2, \ldots, n$, it correspond to successive terms in the binomial expansion

$$(q+p)^n = q^n + \binom{n}{1} q^{n-1} p + \binom{n}{2} q^{n-2} p^2 + \cdots + p^n = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}$$

- The special case of a binomial distribution with $n=1$ is also called the Bernoulli distribution.

The Binomial Distribution

Example: the probability of getting exactly 2 heads in 6 tosses of a fair coin is:

$$P(X = 2) = \binom{6}{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{6-2} = \frac{6!}{2!4!}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{6-2} = \frac{15}{64}$$

The Binomial Distribution

Example: Two teams, the Yankees and the Giants, play a *7*-game series. If the Yankees win each game with probability *p=0.5* (equal ability) independently of any other game, what is the probability the Yankees win the series (win 4 games)?

Recall: $P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$

$$P(X = 4) = \binom{7}{4} p^4 q^{7-4} = \frac{7!}{4!\,(7-4)!} 0.5^4 0.5^{7-4} = 0.2734$$

b) If they are playing 9-game series, what is the probability of winning 5 games out of 9?

$$P(X = 5) = \binom{9}{5} p^5 q^{9-5} = \frac{9!}{5!\,(9-5)!} 0.5^5 0.5^{9-5} = 0.2461$$

Winning 4 out of 7 is more probable than winning 5 out of 9.

Some Properties of the Binomial Distribution

Mean	$\mu = np$
Variance	$\sigma^2 = npq$
Standard Deviation	$\sigma = \sqrt{npq}$
Moment Generating Function	$M(t) = (q + pe^t)^n$

Example:

In 100 tosses of a fair coin, the expected or mean number of heads is $\mu = np = 100 * 0.5 = 50$, while the standard deviation is $\sigma = \sqrt{npq} = \sqrt{100 * 0.5 * 0.5} = 5$.

The law of large numbers for Bernoulli trials

Theorem 4-1: (Law of large numbers for Bernoulli trials)

Let X be the random variable giving the number of successes in n Bernoulli trials, so that X/n is the proportion of successes. Then if p is the probability of success and ϵ is any positive number,

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - p\right| \geq \epsilon\right) = 0$$

In other words, in the long run, it becomes extremely likely that the proportion of successes, X/n, will be as close as you like to the probability of success in a single trial, p.

Example: The probability of getting a 3 when tossing a fair die is 1/6.

The law of large numbers states that the probability of the proportion of 3 in n tosses differing from 1/6 by more than any value $\epsilon > 0$ approaches zero as $n \to \infty$.

The Multinomial Distribution

- The *multinomial distribution* is a generalization of the *binomial distribution*.
- Consider the following example: Rolling a fair die 5 times, what is the probability of getting **two** 3's and **three** 4's?

We are taking an event with multiple possible outcomes (a finite no.), and repeating that event a given number of times.

The required probability is:

$$\frac{5!}{2!\,3!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^3$$

# of ways to get two 3's and three 4's

Probability of two 3's

Probability of three 4's

The Multinomial Distribution

- Suppose that events $A_1, A_2, \ldots, A_k$ are mutually exclusive, and can occur with respective probabilities $p_1, p_2, \ldots, p_k$ where $p_1 + p_2 + \cdots + p_k = 1$. If $X_1, X_2, \ldots, X_k$ are the random variables respectively giving the number of times that $A_1, A_2, \ldots, A_k$ occur in a total of *n* trials, so that $X_1 + X_2 + \cdots + X_k = n$, then

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{n}{n_1!\, n_2! \cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

is the joint probability function for the random variable $X_1, X_2, \ldots, X_k$.

Note: $n_1 + n_2 + \cdots + n_k = n$.

This distribution, which is a generalization of the binomial distribution, is called *the multinomial distribution*, because it is the general term in the multinomial expansion of $(p_1 + p_2 + \cdots + p_k)^n$.

- The expected number of times that $A_1, A_2, \ldots, A_k$ will occur in n trials are $np_1, np_2, \ldots, np_k$ respectively, i.e.,

$$E(X_1) = np_1, \qquad E(X_2) = np_2, \qquad \ldots, \qquad E(X_k) = np_k$$

The Multinomial Distribution

Example: If a fair die is tossed 12 times, the probability of getting 1, 2, 3, 4, 5, and 6 points exactly twice each is:

$$P(X_1 = 2, X_2 = 2, \ldots, X_6 = 2) = \frac{12!}{2!2!2!2!2!2!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2 = \frac{1925}{559{,}872} = 0.00344$$

The Multinomial Distribution

Example: A box contains 5 red balls, 4 white balls, and 3 blue balls. A ball is selected at random from the box, its color is noted, and then the ball is replaced. Find the probability that out of 6 balls selected in this manner, 3 are red, 2 are white, and 1 is blue.

Method 1 (using combinatorial analysis):

The prob. of choosing any red ball is 5/12, then the probability of choosing 3 red balls is $(\frac{5}{12})^3$. Similarly, the probability of choosing 2 white balls is $(\frac{4}{12})^2$, and choosing 1 blue ball is $(\frac{3}{12})^1$.

The probability of choosing 3 red, 2 white, and 1 blue in that order is:

$$\left(\frac{5}{12}\right)^3 * \left(\frac{4}{12}\right)^2 * \left(\frac{3}{12}\right)^1$$

But the same selection can be achieved in various other orders, and the number of these different ways is: $\frac{6!}{3!2!1!}$

The required probability is:

$$\frac{6!}{3!\,2!\,1!} * \left(\frac{5}{12}\right)^3 * \left(\frac{4}{12}\right)^2 * \left(\frac{3}{12}\right)^1 = \frac{625}{5184}$$

The Multinomial Distribution

Example: A box contains 5 red balls, 4 white balls, and 3 blue balls. A ball is selected at random from the box, its color is noted, and then the ball is replaced. Find the probability that out of 6 balls selected in this manner, 3 are red, 2 are white, and 1 is blue.

Method 2 (using multinomial distribution):

$$P(red\ at\ any\ drawing) = \frac{5}{12}$$

$$P(white\ at\ any\ drawing) = \frac{4}{12}$$

$$P(blue\ at\ any\ drawing) = \frac{3}{12}$$

The required probability is:

$$P(3\ red, 2\ white, 1\ blue) = \frac{6!}{3!\,2!\,1!} * \left(\frac{5}{12}\right)^3 * \left(\frac{4}{12}\right)^2 * \left(\frac{3}{12}\right)^1 = \frac{625}{5184}$$

The Hypergeometric Distribution

- Suppose that a box contains *b* blue marbles and *r* red marbles. Let us perform *n* trials of an experiment in which a marble is chosen at random, its color is observed, and then the marble is put back in the box. (*sampling with replacement*)
- If *X* is the random variable denoting the number of blue marbles chosen (successes) in *n* trials, then using the binomial distribution, the probability of exactly *x* successes is:

$$P(X = x) = \binom{n}{x} \frac{b^x r^{n-x}}{(b + r)^n}, \quad x = 0, 1, \ldots, n$$

since $p = \frac{b}{b+r}$, and $\mathrm{q} = 1 - \mathrm{p} = \frac{r}{b+r}$

The Hypergeometric Distribution

- If we modify the event so that *sampling is without replacement*, i.e., the marbles are not replaced after being chosen, then:

$$P(X = x) = \binom{n}{x} \frac{\binom{b}{x}\binom{r}{n-x}}{\binom{b+r}{n}}, \quad x = max\ (0, n\text{-}r), \ldots, min\ (n, b)$$

This is the *hypergeometric distribution*.

- The *mean* and *variance* for this distribution are:

$$\mu = \frac{nb}{b+r}, \qquad \sigma^2 = \frac{nbr(b+r-n)}{(b+r)^2(b+r-1)}$$

Relation btw Hypergeometric and Binomial Distribution

- Let the total number of blue and red marbles be *N*, while the proportions of blue and red marbles are *p* and *q=1-p* respectively, then:

$$p = \frac{b}{b+r} = \frac{b}{N}, \mathrm{q} = \frac{r}{b+r} = \frac{r}{N} \qquad \text{or } \mathrm{b} = Np, \mathrm{r} = N\mathrm{q}$$

The *hypergeometric distribution*, its *mean* and *variance* will become:

$$P(X = x) = \binom{n}{x} \frac{\binom{b}{x}\binom{r}{n-x}}{\binom{b+r}{n}}, \qquad (*)$$

$$\mu = np, \qquad \sigma^2 = \frac{nbq(N-n)}{N-1} \qquad (**)$$

- Note that as $N \to \infty$ (or *N* is large compared with *n*), (*) and (**) will become:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

$$\mu = np, \qquad \sigma^2 = npq$$

This indicates that for large *N*, *sampling without replacement* is practically identical to *sampling with replacement*. In other words, *hypergeometric distribution* is identical with *binomial distribution* when *N* is large.

Reference

Notes, equations, and figures in the lecture are based on or taken from materials in the course textbook:

“Probability and Statistics”, by Spiegel, Schiller and Srinivasan, ISBN 987-007-179557-9 (McGraw-Hill/Schaun’s)