Statistics 301: Probability and Statistics

Discrete Random Variables

Module 4

Updated 2019

Terminology

random variable: describes the outcomes of a statistical experiment; a valued function

discrete variable: values that are countable; whole number values (i.e. chapters in a book, number of classes you can take)

Rules:

(1) $0 \leq P(A) \leq 1$, probabilities must be between 0 and 1 (0% and 100%)

(2) $\sum P(A_i) = 1 = S$, the sum of the probabilities for an experiment must sum to 1 (the sample space)

example random variable

Suppose Nancy has classes three days a week. She attends classes three days a week 80% of the time, two days 15% of the time, one day 4% of the time, and no days 1% of the time. Suppose one week is randomly selected.

Table 9: Nancy's classes

X=# classes attended/week	0	1	2	3
$P(x_i)$	0.01	0.04	0.15	0.8

Rules of expectation I (mean, variance, standard deviation)

Since not all values of a random variable have the same probability, to calculate the mean, we have to approach it in a slightly different way. The mean is called an Expected Value (E(X)). It is a weighted mean (weighted average); meaning some values have a greater chance of happening than others.

$$E(X) = \sum x \cdot p(x_i)$$

$$V(X) = \sum (x_i - \mu)^2 p(x_i)$$

$$SD(X) = \sqrt{V(X)}$$

Nancy's classes

Find the mean, variance, and standard deviation of the number of days in a week Nancy attends classes. Examples like Nancy's classes, flipping two coins, etc. are referred to as "generic" discrete probability distributions. There are special distributions that are "named" and used often to model situations.

$E(X) = \mu = \sum xp(x) = 0(0.01) + 1(0.04) + 2(0.15) + 3(0.8) = 2.74$ days on average

$V(X) = \sigma^2 = \sum(x-\mu)^2 p(x) = (0-2.74)^2(0.01) + (1-2.74)^2(0.04) + (2-2.74)^2(0.15) + (3-2.74)^2(0.8)$ $= 0.3324$

$SD(X) = \sigma = \sqrt{\sigma^2} = \sqrt{0.3324} = 0.576541$

Rules of expectation II

$$E(X+b) = E(X) + E(b) = E(X) + b$$
$$V(X+b) = V(X) + V(b) = V(X) + 0 = V(X)$$
$$E(aX) = aE(X)$$
$$V(aX) = a^2 V(X)$$

Rules of expectation III

If X and Y are independent random variables:

$$E(X \pm Y) = EX \pm EY$$
$$V(X \pm Y) = VX + VY$$
$$E(aX \pm bY) = aEX \pm bEY$$
$$V(aX \pm bY) = a^2 VX + b^2 VY$$

(1) Binomial distribution

There are three assumptions for an experiment to have a binomial distribution
(a) n, the number of trials, are fixed
(b) Only two outcomes are possible: success with probability p, and failure with probability $q = 1 - p$
(c) The n trials are independent

The probability is the probability of x successes out of n trials; any experiment that meets assumptions 2 and 3 when $n = 1$ is called a *Bernoulli Trial*, a binomial experiment happens when the number of successes is counted in one or more Bernoulli Trials.

Shorthand notation:

$$X \sim bin(n, p)$$

Binomial formulas I

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, \cdots, n$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

is a combination and reads as "n choose x." It is the number of ways that x things can be chosen from n trials (number of ways to get a sum of 7 when rolling two 6-sided dice, as an example).

Note that the exponents of p and q must sum to n.

Binomial formulas II

$$EX = np \quad VX = npq \quad SDX = \sqrt{npq}$$

Most scientific calculators have a ${}_nC_r$ command (same as "n choose x") for combinations

Binomial example

Suppose that 20% of all copies of a particular textbook fail a certain binding strength test. Let X denote the number that fail the test among 15 randomly selected copies. Find the probability that exactly 4 fail the test, at most 4 fail the test, at least 4 fail the test.

$$X \sim bin(15, 0.20)$$

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

$$q = 1 - p = 1 - 0.2 = 0.8$$

Binomial example

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

$$P(X = 3) = \binom{15}{3} (0.2)^3 (0.8)^{15-3}$$

$$= 455(0.2)^3(0.8)^{12} = 0.250139$$

Binomial example

$$P(X \leq 3) = \sum_0^3 bin(0:3, 15, 0.2)$$

$= P(0) + P(1) + \cdots + P(3) = 455(0.2)^3(0.8)^{12} + 105(0.2)^2(0.8)^{13}$
$+15(0.2)^1(0.8)^{14} + 1(0.2)^0(0.8)^{15}$
$= 0.648162$

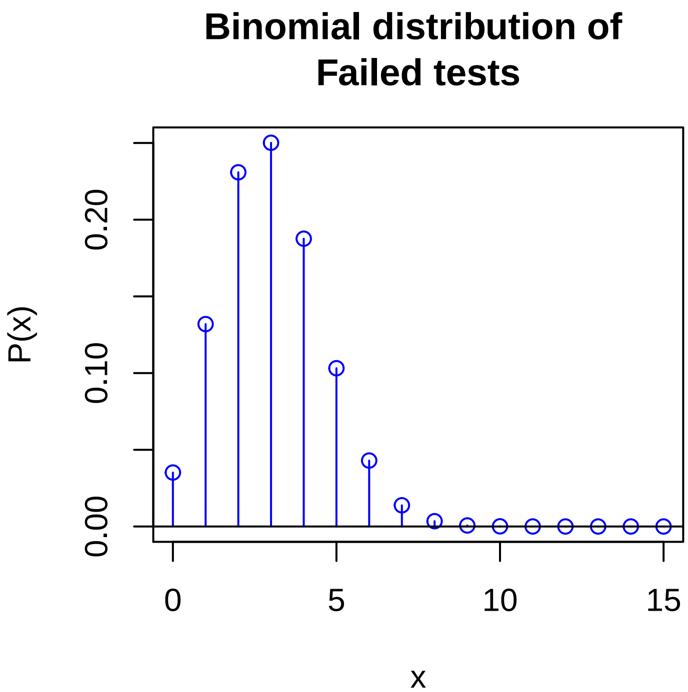
Binomial EX, VX, SDX

$$EX = np = 15(0.2) = 3$$

$$VX = npq = 15(0.2)(0.8) = 2.4$$

$$SDX = \sqrt{npq} = \sqrt{2.4} = 1.549193$$

Binomial Graph



(2) Geometric distribution

There are three assumptions for an experiment to have a geometric distribution
(a) The experiment continues until the first success happens (which is preceded by $n-1$ failures)
(b) Only two outcomes are possible: success with probability p, and failure with probability $q = 1 - p$
(c) The n trials are independent, and since n is not fixed, there needs to be at least one trial

The probability is the probability of $n-1$ failures until the first success (on the last trial)

$$\text{Shorthand notation: } X \sim geo(p)$$

Geometric formulas

$$P(X = x) = q^{x-1}p \qquad x = 1, 2, \cdots, \infty$$

$$EX = \frac{1}{p}, \quad VX = \frac{q}{p^2}, \quad SDX = \sqrt{\frac{q}{p^2}}$$

Geometric example

The probability of a defective steel rod is 0.01. Steel rods are selected at random. Find the probability that the first defect occurs on the ninth steel rod; at least nine before the first defect?

$$X \sim geo(0.01)$$

$$P(X = x) = q^{x-1}p$$

$$q = 1 - p = 1 - 0.01 = 0.99$$

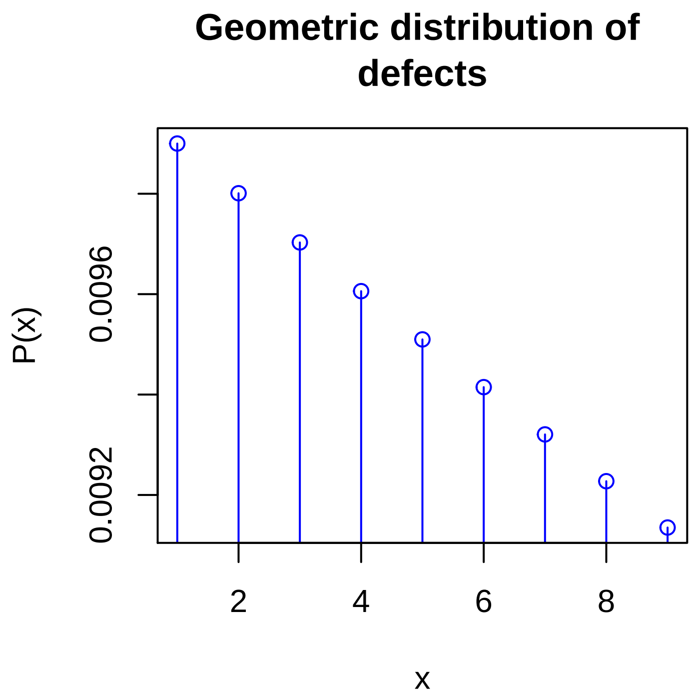
Geometric example continued

$$P(X = 9) = (0.99)^{9-1}(0.01) = 0.009135$$

$$P(X \geq 9) = \sum_{9}^{\infty} geom(9 : \infty, 20, 0.35)$$
$$= 1 - P(X < 9) = 1 - P(X \leq 8)$$

$= 1 - [P(1) + \cdots + P(8)] = 1 - [(0.99)^{8-1}(0.01) + (0.99)^{7-1}(0.01) + (0.99)^{7-1}(0.01) + (0.99)^{6-1}(0.01) + (0.99)^{5-1}(0.01) + (0.99)^{4-1}(0.01) + (0.99)^{3-1}(0.01) + (0.99)^{2-1}(0.01) + (0.99)^{1-1}(0.01)] = 1 - 0.076483 = 0.923517$

Geometric Graph



Geometric EX, VX, SDX

$$EX = \frac{1}{p} = \frac{1}{0.01} = 100$$

$$VX = \frac{q}{p^2} = \frac{0.99}{0.01^2} = 9900$$

$$SDX = \sqrt{VX} = \sqrt{9900} = 99.498744$$

(3) Hypergeometric distribution

There are three assumptions for an experiment to have a hypergeometric distribution
(a) The population or set to be sampled consists of N individuals (a finite population)
(b) Each individual is either a success or failure, and there are M successes in the population
(c) Sampling is done as swor from the combined groups; each pick is not independent (because of swor)

The probability is the probability of the number of items *from the group of interest*

Hyper Logistics

The probability is the probability of the number of items *from the group of interest*

$$\text{Shorthand notation: } X \sim hyper(M, n, N)$$

Use the ${}_nC_r$ function for all three parts of the calculation.

Hypergeometric formulas

$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \quad x = 0, 1, \ldots, n$$

M = sample size of things of interest
x = argument of interest
$N - M$ = second group of things
$n - x$ = number of things from second group (dependent on how many things from the 1st group)
N = the total number of things

Hyper EX, VX, SDX

$$EX = \frac{Mn}{N}$$

$$VX = \left(\frac{N-n}{N-1}\right)\left(\frac{Mn}{N}\right)\left(1 - \frac{M}{N}\right)$$

$$SDX = \sqrt{VX}$$

Hypergeometric example

An animal population thought to be near extinction in a certain region had five individuals that were caught, tagged, and released to mix back into the population. A random sample of 10 of these animals is selected after a time. Let X = the number of tagged animals in the second sample. Say there are actually 25 animals of this type in the region, what is the probability that there are 2, at most 2?

$$X \sim hyper(M = 5, n = 10, N = 25)$$

$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Hyper example continued

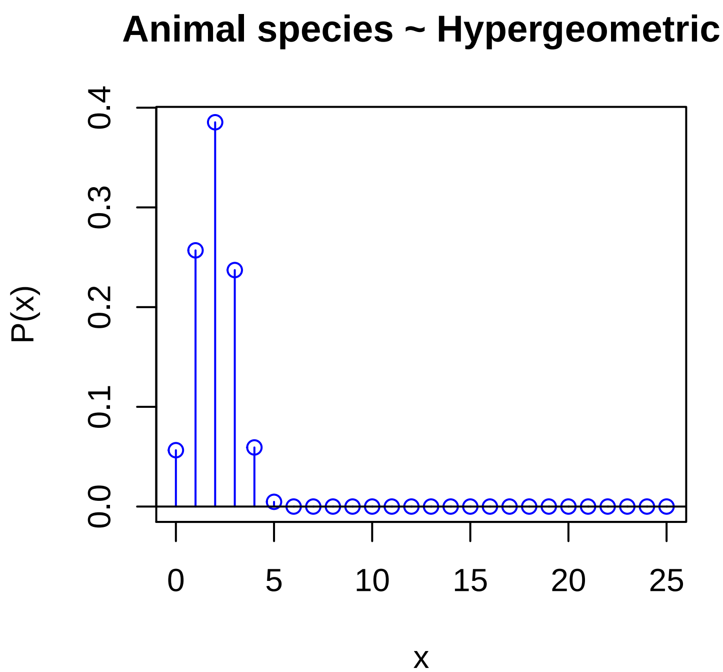
$$P(X = 2) = \frac{\binom{5}{2}\binom{25-5}{10-2}}{\binom{25}{10}} = 0.385375$$

$$P(X \leq 2) = \sum_{0}^{2} hyper(0:2, 5, 10, 25) = P(0) + P(1) + P(2)$$

$$= \sum(0.056522, 0.256917, 0.385375)$$

$$= \frac{\binom{5}{2}\binom{25-5}{10-2}}{\binom{25}{10}} + \frac{\binom{5}{1}\binom{25-5}{10-1}}{\binom{25}{10}} + \frac{\binom{5}{0}\binom{25-5}{10-0}}{\binom{25}{10}} = 0.698814$$

Hypergeometric Graph



Hyper EX, VX, SDX

$$EX = \frac{Mn}{N} = \frac{5(10)}{25} = 2$$

$$VX = \left(\frac{N-n}{N-1}\right)\left(\frac{Mn}{N}\right)\left(1 - \frac{M}{N}\right)$$
$$= \left(\frac{25-10}{25-1}\right)\left(\frac{5(10)}{25}\right)\left(1 - \frac{5}{25}\right)$$

$$= 1$$

$$SDX = \sqrt{VX} = \sqrt{1} = 1$$

(4) Poisson distribution

There are three assumptions for an experiment to have a poisson distribution
(a) Used in modelling rare events
(b) Events are independent
(c) Can be used to estimate binomial when n is "large" and p is "small"

The probability of the number of events happening *within a specified period* This distribution came about to model the number of horse kicks Prussian soldiers received (seriously, go look it up!)

Poisson logistics

Poisson is pronounced as "pwa-son". If you have seen The Little Mermaid, recall the song that the chef sings when Sebastian is running from him in the kitchen (Little Mermaid)[https://www.youtube.com/watch?v=Iq-UaXQ1c1M]

$$\text{Shorthand notation: } X \sim pois(\mu)$$

μ is the average or a rate (use $\mu = np$, EX from binomial)

Poisson formulas

$$P(X = x) = \frac{e^{-\mu}\mu^x}{x!} \quad x = 0, 1, \cdots, \infty$$

$$EX = \mu$$

$$VX = \mu$$

$$SDX = \sqrt{\mu}$$

Poisson example

Consider an experiment that consists of counting the number of α-particles given off in a 1-second time interval by 1 gram of radioactive material. If the average number of α-particles given off is 3.2, what is the probability of exactly 2 α-particles given off in the next 1-second interval? What is the probability that no α-particles are given off in the next 1-second interval? More than 2 α-particles?

$$X \sim pois(3.2) \ or \ X \sim P(3.2)$$

$$P(X = x) = \frac{e^{-\mu}\mu^x}{x!}$$

Poisson example continued

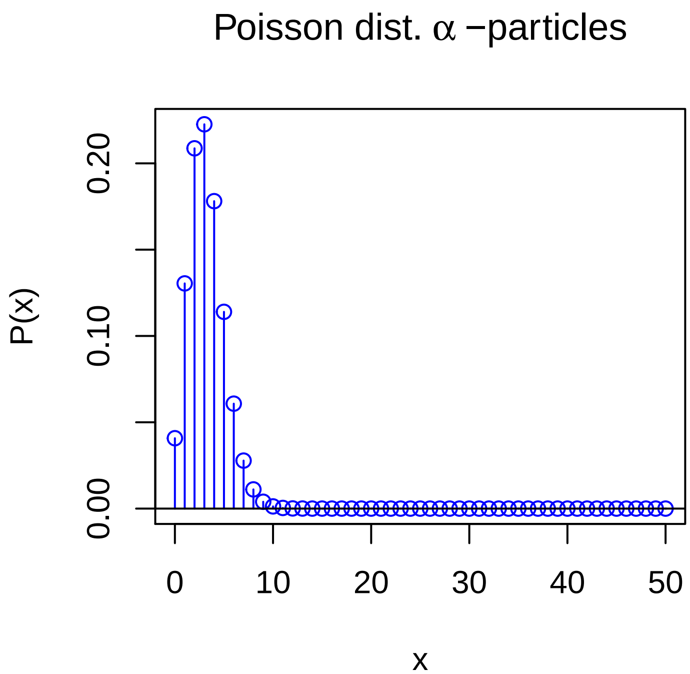
$$P(X = 2) = \frac{e^{-3.2}3.2^2}{2!} = 0.208702$$

$$P(X = 0) = \frac{e^{-3.2}3.2^0}{0!} = 0.040762$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) = 1 - [P(X = 1) + P(X = 0)]$$

$$= 1 - \left(\frac{e^{-3.2}3.2^1}{1!} + \frac{e^{-3.2}3.2^0}{0!}\right)$$
$$= 1 - 0.171201 = 0.828799$$

Poisson graph



Poisson EX, VX, SDX

$$EX = \mu = 3.2$$
$$VX = \mu = 3.2$$
$$SDX = \sqrt{\mu} = \sqrt{3.2} = 1.788854$$

(5) Negative binomial distribution

The assumptions for an experiment to have a negative binomial distribution
(a) The n trials are independent
(b) Only two outcomes are possible: success with probability p, and failure with probability $q = 1 - p$
(c) Probability of success, p, is constant from trial to trial
(d) The experiment continues until a total of r successes have been observed, where r is a *specified* positive integer

The probability of the number of failures which occur before a target number of successes is reached

Example situation for negative binomial distribution

Consider the following statistical experiment[1]. You flip a coin repeatedly and count the number of times the coin lands on heads. You continue flipping the coin until it has landed 5 times on heads. This is a negative binomial experiment because: (1) The experiment consists of repeated trials. We flip a coin repeatedly until it has landed 5 times on heads (2) Each trial can result in just two possible outcomes - heads or tails (3) The probability of success is constant: 0.5 on every trial (4) The trials are independent; that is, getting heads on one trial does not affect whether we get heads on other trials. The experiment continues until a fixed number of successes have occurred; in this case, 5 heads.

[1] https://stattrek.com/probability-distributions/negative-binomial.aspx

Negative binomial logistics

$$\text{Shorthand notation: } X \sim nb(r,p) \ or \ X \sim NB(r,p)$$

r is the target number of successes
p is the probability of success

Because the formula is basically identical to the binomial, you are allowed to use your calculator to do as much of the calculation as you want. There are instructions on how to use the command if you have a TI graphing calculator on page 254. Most scientific calculators have a $_nC_r$ command (same as "n choose x" for binomial and "$x-1$ choose $r-1$" for the negative binomial) for combinations

Negative binomial formulas

$$P(X=x) = \binom{x-1}{r-1} p^r q^{x-r} \quad x = 0, 1, \cdots, \infty$$

$$EX = \frac{rq}{p}$$

$$VX = \frac{rq}{p^2}$$

$$SDX = \sqrt{\frac{rq}{p^2}} = \sqrt{VX}$$

Negative binomial example

A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen. Let $p =$ the probability that a randomly selected couple agrees to participate. If $p = 0.2$, what is the probability that 15 couples must be asked before 5 are found who agree to participate (i.e. what is the chance that 10 couples refuse before the 5^{th} couple agrees)? What is the probability that the first 5 couples asked agree? What is the probability that at most 10 couples refused before the 5^{th} couple agrees? What is the expected number of refusals (failures) until the 5^{th} couple agrees (the 5^{th} success)? What is the variance and standard deviation?

Negative binomial example continued

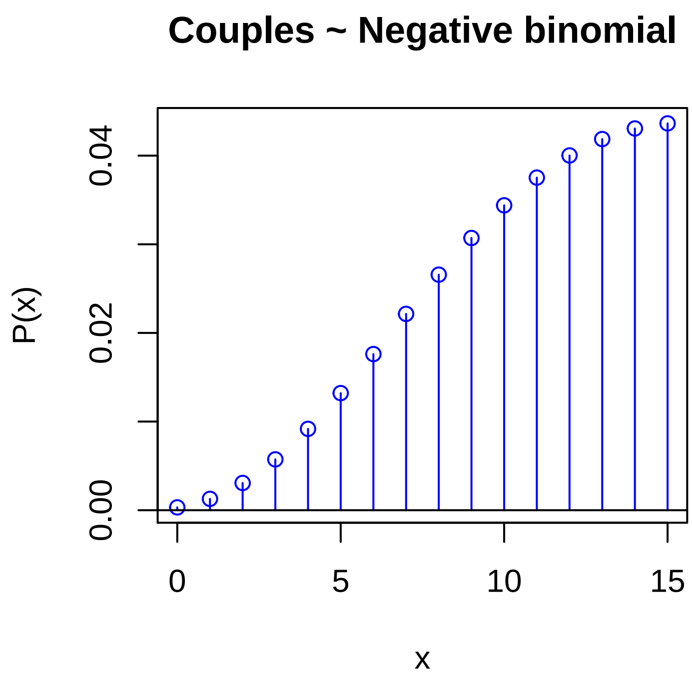
$$X \sim nb(r,p) \Rightarrow X \sim nb(5, 0.2)$$

$$P(X=10) = \binom{10+5-1}{5-1}(0.2)^5(0.8)^{10} = \binom{14}{4}(0.2)^5(0.8)^{10} = 0.034394$$

$$P(X=0) = \binom{0+5-1}{5-1}(0.2)^5(0.8)^0 = \binom{4}{4}(0.2)^5(0.8)^0 = 0.00032$$

$P(X \leq 10) = P(0) + P(1) + \cdots + P(10) = \binom{0+5-1}{5-1}(0.2)^5(0.8)^0 + \binom{1+5-1}{5-1}(0.2)^5(0.8)^1 + \cdots + \binom{10+5-1}{5-1}(0.2)^5(0.8)^{10} = 0.164234$

Negative binomial graph



Negative binomial EX, VX, SDX

$$EX = \frac{rq}{p} = \frac{5(0.2)}{0.8} = 20$$

$$VX = \frac{rq}{p^2} = \frac{5(0.8)}{(0.2)^2} = 100$$

$$SDX = \sqrt{\frac{rq}{p^2}} = \sqrt{VX} = sqrt100 = 10$$