3.5 Binomial "Relatives"

Negative Binomial - Criteria

An experiment is said to be a negative binomial experiment provided:

1. Each repetition is called a trial.
2. For each trial there are two mutually exclusive (disjoint) outcomes: success or failure
3. The probability of success is the same for each trial of the experiment
4. The trials are independent
5. The trials are repeated until r successes are observed, where r is specified in advance.

103

Negative Binomial Notation

When we studied the Binomial distribution, we were only interested in the probability for a success or a failure to happen. The negative binomial distribution addresses the number of trials necessary ***before the k^th^ success***. If the trials are repeated *x* times until the k^{th} success, we will have had *x – k* failures. If *p* is the probability for a success and *(1 – p)* the probability for a failure, the probability for the r^{th} success to occur at the x^{th} trial will be

$$P(x) = {}_{x-1}C_{k-1}\, p^k(1-p)^{x-k} \qquad x = k, k+1, k+2, \ldots$$

where k is the number of successes observed in x trials of a binomial experiment with success rate of p.

104

Mean and Standard Deviation of a Negative Binomial RV

A negative binomial experiment with probability of success p has

Mean $\mu_x = k/p$

Standard Deviation $\sigma_x = \sqrt{k(1-p)/p^2}$

where k is the number of successes observed in x trials of a binomial experiment with success rate of p

Note that the geometric distribution is a special case of the negative binomial distribution with *k* = 1.

105

Examples of Negative Binomial PDF

- Number of cars arriving at a service station until the fourth one that needs brake work
- Flipping a coin until the fourth tail is observed
- Number of planes arriving at an airport until the second one that needs repairs
- Number of house showings before an agent gets her third sale
- Length of time (in days) until the second sale of a large computer system

106

Example 1

The drilling records for an oil company suggest that the probability the company will hit oil in productive quantities at a certain offshore location is 0.3 . Suppose the company plans to drill a series of wells looking for three successful wells. $P(x) = {}_{x-1}C_{r-1}\, p^k(1-p)^{x-k}$ $\quad p = 0.3$

a) What is the probability that the third success will be achieved with the 8th well drilled?

$P(8) = {}_{8-1}C_{3-1}\, p^3(1-p)^{8-3} = {}_7C_2\,(0.3)^3(0.7)^5$

$= (21)(0.027)(0.16807) = 0.0953$

b) What is the probability that the third success will be achieved with the 20th well drilled?

$P(20) = {}_{20-1}C_{3-1}\, p^3(1-p)^{20-3} = {}_{19}C_2\,(0.3)^3(0.7)^{17}$

$= (171)(0.027)(0.00233) = 0.0107$

107

Example 1 cont

The drilling records for an oil company suggest that the probability the company will hit oil in productive quantities at a certain offshore location is 0.3 . Suppose the company plans to drill a series of wells looking for three successful wells.

c) Find the mean and standard deviation of the number of wells that must be drilled before the company hits its third productive well.

Mean $\quad \mu_x = k/p \quad = 3 / 0.3 = 10$

Standard Deviation $\quad \sigma_x = \sqrt{k(1-p)/p^2} \quad = \sqrt{3(0.7)/(0.3)^2}$

$= 4.8305$

108

Example 2

A standard, fair die is thrown until 3 aces occur. Let Y denote the number of throws.

Find the mean of Y

$E(Y) = k/p = 3/(1/6) = 18$

Find the variance of Y

$V(Y) = k(1-p)/p^2 = 3(5/6)/(1/6)^2 = 90$

Find the probability that at least 20 throws will needed

$P(\text{at least } 20) = P(Y\geq 20) = 1 - P(Y<20)$

$= 1 - [P(3) + P(4) + \ldots + P(18) + P(19)]$

$P(Y\geq 20) = 0.3643$

109

Other Definitions of the Negative Binomial

Some textbooks define the negative binomial as the "number of failures observed prior to the k^{th} success".

This is equivalent, but requires changing the formulas for the mean and variance.

110

Hypergeometric distribution

Sampling with replacement

If we sample with replacement and the trials are all independent, the binomial distribution applies.

Sampling without replacement

If we sample without replacement, a different probability distribution applies.

111

Hypergeometric Distribution conditions

It is a discrete distribution.

Sampling is done ***without replacement***.

The number of objects in the population, *N,* is finite and known.

Each trial has exactly two possible outcomes: success and failure.

Trials are not independent

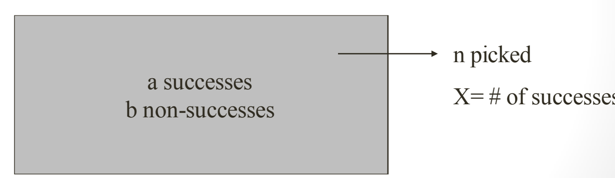
X is the number of successes in the *n* trials

112

Example

Pick up n balls from a box without replacement. The box contains a white balls and b black balls

X=# of white balls picked



113

In the box: M successes, N - M non-successes

The probability of getting x successes (white balls):

$$p(x)=\frac{\text{\# of ways to pick n balls with x successes}}{\text{total \# of ways to pick n balls}}$$

\# of ways to pick x successes

=(# of ways to choose x successies)*(# of ways to choose n-x non-successes)

$$=\binom{M}{x}\binom{N-M}{n-x}$$

$$p(x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}},x=0,1,2,\ldots,a$$

$$\max(0,n-N+M)\le X\le \min(n,M)$$

Hypergeometric Distribution

$$\text{mean}=\mu_x=n\frac{M}{N}$$

$$\text{variance}=\sigma_x^2=n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Example

52 cards. Pick n=5. X=# of aces, then a=4, b=48

$$P(X=2)=\frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}}$$

116

Example

A box has 100 batteries.

a = 98 good ones; b = 2 bad ones

n = 10

X=# of good ones

$$P(X=8)=\frac{\binom{98}{8}\binom{2}{2}}{\binom{100}{10}}$$

117

Continued

$$P(\text{at least 1 bad one}) = 1 - P(\text{all good})$$
$$= 1 - P(X = 10)$$
$$= 1 - \frac{\binom{98}{10}\binom{2}{0}}{\binom{100}{10}}$$
$$= 1 - \frac{98}{100} \bullet \frac{97}{99} \bullet \frac{96}{98} \bullet \frac{95}{97} \bullet \frac{94}{96} \bullet \frac{93}{95} \bullet \frac{92}{94} \bullet \frac{91}{93} \bullet \frac{90}{92} \bullet \frac{89}{91}$$

118

Binomial Approximation to the Hypergeometric Distribution

Sometimes we know that the proper model is hypergeometric, but we do not have the information we need to specify the proper model

In this case, the binomial is an acceptable approximation, if the sample size n is less than 5% of the population size N. Otherwise it is not.

(Note: some texts say 10%)

119

Binomial Approximation to the Hypergeometric Distribution

e.g. Sample 20 items from a very large number (several thousand) of items where 5% are defective. What is the probability we get more than 2 defectives?

We would like to use the Hypergeometric distribution, but we do not know a and b. Since 20 is less than 5% of "several thousand", we can use the binomial approximation.

The 5% condition works because the probability of a success does not change much depending on the successes or failures in previous draws.

120

Slide 5-120

Poisson Distribution

Describes discrete occurrences over a continuum or interval

A discrete distribution

Describes rare events

Each occurrence is independent any other occurrences.

The number of occurrences in each interval can vary from zero to infinity.

The expected number of occurrences must hold constant throughout the experiment.

121

Poisson Distribution

Probability function

$$P(X) = \frac{\lambda^X e^{-\lambda}}{X!} \quad \text{for} \quad X = 0,1,2,3,\ldots$$

where:

$\lambda = long - run\ average$

$e = 2.718282\ldots$ (the base of natural logarithms)

$$\text{mean} = \mu_x = \lambda \qquad \text{variance} = \sigma_x^2 = \frac{1}{\lambda^2}$$

122

Poisson distribution

Events happen independently in time or space with, on average, λ events per unit time or space.

- Radioactive decay; λ=2 particles per minute
- Lightening strikes; λ=0.01 strikes per acre

123

Poisson probabilities

Radioactive decay

x = # of particles/min

λ = 2 particles per minute (average rate)

$$P(x=3) = \frac{2^3 e^{-2}}{3!}, \; x = 0,\ 1,\ 2,\ \ldots$$

124

Poisson probabilities

Radioactive decay

X = # of particles/hour

λ = 2 particles/min * 60min/hour=120 particles/hr

$$P(x=125) = \frac{120^{125} e^{-120}}{125!}, \; \text{x=0, 1, 2, ...}$$

125