STAT 511

Lecture 6: The Binomial, Hypergeometric, Negative Binomial and Poisson Distributions

Devore: Section 3.4-3.6

Prof. Michael Levine

February 5, 2019

Binomial Experiment

1. The experiment consists of a sequence of n trials, where n is fixed in advance of the experiment.
2. The trials are identical, and each trial can result in one of the same two possible outcomes, which are denoted by success (S) or failure (F).
3. The trials are independent
4. The probability of success is constant from trial to trial and is denoted by p.

- Given a binomial experiment consisting of n trials, the binomial random variable X associated with this experiment is defined as $X =$ the number of Ss among n trials

Example where the experiment is not binomial I

- Consider 50 restaurants to be inspected; 15 of them currently have at least one serious health code violation while the rest have none.
- There are 5 inspectors, each of whom will inspect 1 restaurant during the coming week.
- The restaurant names are sampled as slips of paper **without replacement**; ith trial is a success if the restaurant has no violations where $= 1, \ldots, 5$.
- Then $P(S \text{on the 1st}) = \frac{35}{50} = .70$

Example where the experiment is not binomial II

- Similarly, $P(S \text{on the 2nd}) = P(SS) + P(FS) = .70$
- However, $P(S \text{on the 5h trial}|SSSS) = \frac{31}{46} = .67$ while $P(S \text{on the 5h trial}|FFFF) = \frac{35}{46} = .76$
- If the sample size n is at most 5% of the population size, the experiment can be analyzed as though it were exactly a binomial experiment.

Binomial pmf

- Because 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 binomial pmf is

$$b(x;n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

Expected value and variance of a binomial RV

- Let $X_1, \ldots, X_n$ be mutually independent Bernoulli random variables, each with success probability p. Then, $Y = \sum_{i=1}^{n} X_i$ is a binomial random variable with pmf $b(x; n, p)$.
- The expected value is

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

- The variance is

$$V(Y) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Example

- A card is drawn from a standard 52-card deck. If drawing a club is considered a success, find the probability of
  1. exactly one success in 4 draws (with replacement)
  2. no successes in 5 draws (with replacement).

1. $p = \frac{1}{4}$ and $q = \frac{3}{4}$
2. The probability of just 1 success is

$$\binom{4}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^3 \approx 0.422$$

3. The probability of no successes in 5 draws is

$$\binom{5}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^5 \approx 0.237$$

Binomial Cdf

- For X distributed as Bin(n, p), the cdf is defined as

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

- In the above, $x = 0, 1, 2, \ldots, n$

Example

- If the probability of a student successfully passing this course (C or better) is 0.82, find the probability that given 8 students
- All 8 pass

$$\binom{8}{8}(0.82)^8(0.18)^0 \approx 0.2044$$

- None pass

$$\binom{8}{0}(0.82)^0(0.18)^8 \approx 0.0000011$$

- At least 6 pass

$$\binom{8}{6}(0.82)^6(0.18)^2 + \binom{8}{7}(0.82)^7(0.18)^1$$
$$+\binom{8}{8}(0.82)^8(0.18)^0 \approx 0.8392$$

The Hypergeometric Distribution

1. The population to be sampled consists of N individuals, objects, or elements (a finite population).
2. Each individual can be characterized as a success (S) or failure (F), and there are M successes in the population
3. A sample of n individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen.

- Let X be the number of Ss in a random sample of size n drawn from a population consisting of M Ss and NM Fs.
- The probability distribution of X, called the **hypergeometric distribution**, is given by

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

- In the above, $\max(0, n - N + M) \leq x \leq \min(n, M)$

Example: **capture-recapture model I**

- 5 individuals from an animal population thought to be near extinction in a certain region have been caught, tagged and released.
- Afterwards, a sample of 10 animals is selected. Let X be the number of tagged animals in the second sample.
- If there are actually 25 animals of this type in the region, what is $P(X = 2)$? Or $P(X = 25)$?

Example: **capture-recapture model II**

- The parameter values are $n = 10$, $M = 5$ and $N = 25$
- Denote X the number of tagged animals in the *recapture* sample.
- The pmf of X is

$$h(x; 10, 5, 25) = \frac{\binom{5}{x}\binom{20}{10-x}}{\binom{25}{10}}$$

$$P(X = 2) = h(2; 10, 5, 25) = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} = .385$$

$$P(X \leq 2) = \sum_{x=0}^{2} h(x; 10, 5, 25) = .699$$

Hypergeometric Mean and Variance

- The mean is
$$E(X) = n \cdot \frac{M}{N}$$
- Note that $p = \frac{M}{N}$ is the proportion of successes in the population. Compare this to the binomial mean $E(X) = np$.
- The variance is
$$V(X) = \frac{N-n}{N-1} \cdot np(1-p)$$
- It differs from the binomial variance by the **finite population correction factor** $\frac{N-n}{N-1}$.
- Note that this factor is relatively small when n is small compared to N.

Example

- In the capture-recapture example,

$$E(X) = 10 \cdot (.2) = 2$$

- The variance is

$$V(X) = \frac{15}{24}(10)(.2)(.8) = 1$$

whereas it is equal to 1.6 for sampling with replacement.

Possible Application

- A possible application of hypergeometric distribution is to estimate N when it is unknown.
- We do it by equating sample proportion x/n with the population proportion M/N; the result is

$$\hat{N} = \frac{M \cdot n}{x}$$

Negative Binomial Distribution

- The experiment consists of a sequence of independent trials.
- Each trial can result in a success (S) or a failure (F).
- The probability of success is constant from trial to trial, so

$$P(S \text{ on trial } i) = p$$

  for $i = 1, 2, 3, \ldots$
- The experiment continues until a total of r successes have been observed, where r is a specified positive integer.
- X is the number of failures it takes to achieve r successes

Pmf of a negative binomial

- The pmf of the negative binomial rv X with parameters $r =$ number of Ss and $p = P(S)$ is

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x$$

for $x = 0, 1, 2, \ldots$

Example I

- A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to take part in a new childbirth regimen.
- Let p be the probability of agreement for a couple. If $p = 0.2$, what is the probability that at most 15 couples must be asked before 5 are found who agree to participate?

Example II

- The probability that at most 10 F's are observed (at most 15 couples must be interviewed) is

$$P(X \leq 10) = \sum_{x=0}^{10} nb(x; 5, 0.2) = .164$$

- Note that when $r = 1$ the pfm is $nb(x; 1, p) = (1 - p)^x p$. Thus, geometric distribution is a special case of negative binomial.

- What does the negative binomial distribution model?
- Imagine a sequence of oil drills in the quest for success. Then we are interested in the probability of X being as small as possible...

Poisson Distribution

- Let us count the number of changes (e.g. the number of arriving phone calls) in a system in a given continuous time interval. Assume that
  1. The numbers of changes occurring in non-overlapping intervals are independent
  2. The probability of exactly one change in a sufficiently short interval of length h is approximately λh
  3. The probability of two or more changes in a sufficiently short interval h is $o(h)$ (that is, essentially zero)
- Then the number of changes has the Poisson distribution with the parameter $\lambda > 0$.

Some Properties

- Its pmf is

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

  where $x = 0, 1, 2, 3 \ldots$.
- Clearly, $p(x;\lambda) > 0$. Also,

$$\sum_{x=0}^{\infty} p(x;\lambda) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = 1$$

- The last property is the consequence of the Taylor series expansion of the exponent function $e^{-\lambda}$

Example

- Let X denote the number of traps (defects of a certain kind) in a particular type of metal oxide semiconductor transistor
- Let $X \sim P(\mu)$ with $\mu = 2$.
- The Poisson model is suggested in the article Analysis of Random Telegraph Noise in 45-nm CMOS Using On-Chip Characterization System" from IEEE Trans. on Electronic Devices, 2013
- Then,

$$P(X = 3) = p(3;2) = \frac{e^2 2^3}{3!} = .180$$

- Also,

$$P(X \leq 3) = F(3;2) = \sum_{x=0}^{3} \frac{e^{-2} 2^x}{x!} = .135 + .271 + .271 + .180 = .857$$

- Poisson approximation of the binomial distribution: if $n \to \infty$, $p \to 0$ and $np \to \lambda > 0$ we have

$$b(x; n, p) \to p(x; \lambda)$$

- An acceptable rule of thumb is to use this approximation when $n \geq 50$ and $np < 5$.

Example

- In a certain industrial facility accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other
- What is the probability that in any given period of 400 days there will be an accident on one day? Note that $n = 400$ and $p = 0.005$. Using Poisson approximation with $\lambda = np = 2$, we have

$$P(X = 1) = e^{-2}2^1 = 0.271$$

- What is the probability that there are at most three days with an accident?

$$P(X \leq 3) = \sum_{x=0}^{3} e^{-2}2^x/x! = 0.857$$

Mean and variance of Poisson distribution

- The mean and variance of the Poisson distribution are equal:

$$E(X) = V(X) = \lambda$$

- If events in a Poisson process occur at a mean rate of λ per unit, the expected number of occurrences in an interval of length t is λt.
- For example, if phone calls arrive at a switchboard following a Poisson process at a mean rate of 3 per minute, the expected number of phone calls in a 5-minute interval is $5 \cdot 3 = 15$.
- Moreover, the number of occurrences X in the interval of length t has the Poisson pmf

$$p(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

Example

- Suppose pulses arrive at a counter at an average rate of 6 per minute, so that $\lambda = .6$.
- Define X as the number of pulses in the 30 sec interval - it is Poisson with parameter $\alpha t = 6(0.5) = 3$.
- Then,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{e^{-3}(3^0)}{0!} = .95$$

- λ is called the *rate* of the Poisson process.