Discrete distributions

Li Li

Department of Mathematics and Statistics

Random variables

Bernoulli distribution

Binomial distribution

Geometric distribution

Poisson distribution

Bernoulli distribution

- The Bernoulli distribution, named after Swiss mathematician Jacob Bernoulli, is the discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $1 - p$.
- For example, the Bernoulli distribution can be used as the probability distribution of
  - any single experiment that asks a yes-no question;
  - the question results in a boolean-valued outcome;
  - a (possibly biased) coin toss where 1 and 0 would represent "head" and "tail" (or vice versa).
- We will generally call the two outcomes "success" and "failure". Assign the value 1 to success and 0 to failure.

Bernoulli distribution

- A random variable X that follows a Bernoulli distribution assumes

$$X = \begin{cases} 1, & \text{if "success"} \\ 0, & \text{if "failure"} \end{cases}$$

The pmf (probability mass function) of X is

x	0	1
$f(x)$	$1-p$	p

where p is often called "success rate" or "probability of success".

- X is said to be a Bernoulli random variable.
- The experiment is also called a Bernoulli trial.

Mean and Variance of a Bernoulli distribution

$$\mu = p$$

$$\sigma^2 = p \cdot (1-p)$$

$$M(t) = (1-p) + e^t \cdot p$$

Where we use Bernoulli distributions?

- To estimate the proportion of the products with defects, we will take a sample, and model whether each product in the sample has defects using a Bernoulli distribution, i.e. we treat the outcome of the product as if it is a realization from a Bernoulli random variable. We also assume the all the Bernoulli distributions here have the same success rate p, which is also the proportion of the products with defects.
- Machine learning: to classify an object to be a pedestrian or a non-pedestrian, we model each observation in our sample as if it is a realization of a Bernoulli random variable and let its success rate be a function of many inputs.

Estimation of p

- To estimate the probability of "head" of a possibly biased coin, we can flip the coin n times, for example 100 times.
- Let 1 represent "head" and "0" represent "tail". We obtain a sample $\{x_1, x_2, \ldots, x_n\}$ where each x_i is either 0 or 1.
- Chapter 7: An estimate of p based on the sample is

$$\hat{p} = \frac{\sum_{i=1}^{m} x_i}{n}.$$

: the proportion of 1's

- Chapter 8: An interval estimate of p with 95% confidence is

$$\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

Related distributions

- If $X_1, X_2, \ldots, X_n$ are independent, identically distributed (i.i.d.) random variables, all Bernoulli trials with success probability p, then their sum is distributed according to a binomial distribution with parameters n and p.
- The geometric distribution models the number of independent and identical Bernoulli trials needed to get one success.

Binomial distribution

- The binomial distribution with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent and identical Bernoulli experiments:
- All the Bernoulli experiments have the same success/yes/true probability p and failure/no/false probability $q = 1 - p$.
- All the Bernoulli experiments are independent of each other.

Probability mass function

x	0	1	2	3
$f(x)$	$(1-p)^3$	$3p(1-p)^2$	$3p^2\cdot(1-p)$	p^3

Binomial$(3, p)$

Consider tossing three identical coins independently. Each coin has a probability p to show up "head". Define X as the number of heads obtained in the three tosses.

A : head for 1st toss

B : head for 2nd toss

C : head for 3rd toss

$X = \{0, 1, 2, 3\}$

1) $\{X = 0\} = A^c \cap B^c \cap C^c$

1) $\{X = 1\} = (A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C)$

$f(0) = (1-p)\cdot(1-p)(1-p) = (1-p)^3$

$f(1) = 3\cdot \underline{p\cdot(1-p)\cdot(1-p)} = 3p(1-p)^2$

Probability mass function

- The probability mass function of a Binomial distribution is

$$\binom{n}{x}$$

$$f(x) = C_x^n p^x (1-p)^{n-x}, x = 0, 1, 2, \ldots, n$$

where n, p are fixed constants for the experiment. In this case, the random variable X that follows a Binomial distribution is said to be a Binomial random variable.

Cumulative distribution function

The cumulative distribution function can be expressed as:

$$F(x) = P(X \leq x) = \sum_{i=0}^{\lfloor x \rfloor} C_i^n p^i (1-p)^{n-i}, -\infty < x < +\infty$$

where $\lfloor x \rfloor$ is a floor function, i.e. greatest integer less than or equal to x. For example, if $n \geq 2$,
$F(1.5) = C_0^n p^0 (1-p)^n + C_1^n p(1-p)^{n-1}$.

Example

The blood types of successive children born to the same parents are

- independent and have fixed probabilities that depend on the genetic makeup of the parents.
- Each child born to a certain set of parents has probability 0.25 of having blood type O.

If these parents have five children, what is the probability that less than 2 of them have type O blood? Denote X as the number of children that has type O blood.

Mean and Variance

The mean and variance of the Binomial random variable X are

$$E(X) = np$$

$$Var(X) = np(1 - p).$$

Proof for mean and variance

First obtain the moment generating function for the distribution:

Proof for mean and variance

Where we use Binomial distributions?

- To estimate the proportion of the products with defects, we will take a sample, and model the total number of products in the sample that have defects using a Binomial distribution. We also assume the success rate p is the proportion of the all products with defects.
- Binomial distributions are used in genetics to determine the probability that k out of n individuals will have a certain genotype or phenotype.
- Binomial distributions are used in categorical data analysis where you observe contingency tables.

Does X follow a Binomial distribution?

A music distributor inspects an SRS of 10 CDs from a shipment of 100 music CDs. Suppose that (unknown to the distributor) 10% of the CDs in the shipment have defective copy-protection schemes that will harm personal computers. Count the number X of bad CDs in the sample. What if this is a shipment of 100,000 music CDs?

When there are 100 CDs total, X does not follow a Binomial distribution since whether each of the 10 CDs has defective are not independent. However, if there are 100,000 CDs total, whether each of the 10 CDs has defective are nearly independent.

Does X follow a Binomial distribution?

Boxes of six-inch slate flooring tile contain 40 tiles per box. The count X is the number of cracked tiles in a box. You have noticed that most boxes contain no cracked tiles, but if there are cracked tiles in a box, then there are usually several.

X does not follow a Binomial distribution since whether each tile has a crack are not independent.

Does X follow a Binomial distribution?

You are rolling a pair of balanced six-faced dice in a board game. Rolls are independent. You land in a danger zone that requires you to roll doubles (both faces show the same number of spots) before you are allowed to play again. X is the number of rolls until you can play again.

X does not follow a Binomial distribution since X is not the number of doubles out of a total number of rolls.

Estimation of p

- We can estimate p using a sample of a single observation x.
  - Chapter 7: An estimate of p based on the sample is
    $$\hat{p} = \frac{x}{n}.$$
  - Chapter 8: An interval estimate of p with 95% confidence is
    $$\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

Definition

In probability theory and statistics, the geometric distribution is either of two discrete probability distributions:

- The probability distribution of the number X of independent and identical Bernoulli trials needed to get one success, supported on the set $\{1, 2, 3, \dots\}$.
- The probability distribution of the number $Y = X - 1$ of failures before the first success, supported on the set $\{0, 1, 2, 3, \dots\}$.

To avoid ambiguity, it is considered wise to indicate which is intended, by mentioning the support explicitly.

Probability mass function of Geometric distribution

If the probability of success on each trial is p,

- The random variable X that equals the number of trials until the first success has a probability mass function as

  $$f(x) = (1-p)^{x-1}p \text{ where } x = 1, 2, \ldots$$

  For example, $f(X = 3) = (1-p)^2 p$.

- The random variable Y that equals the number of failures before the first success has a probability mass function as

  $$f(y) = (1-p)^{y}p \text{ where } y = 0, 1, 2, \ldots$$

  For example, $f(Y = 3) = (1-p)^3 p$.

Example

You are rolling a pair of balanced six-faced dice in a board game. Rolls are independent. You land in a danger zone that requires you to roll doubles (both faces show the same number of spots) before you are allowed to play again. How long will you wait to play again?

(a) What is the probability that you can play in the second roll?

(b) What is the probability that you need to wait at least 4 rounds of rolling to play again?

Example

Denote X is the number of rolling till a double. The probability of rolling a double is 1/6, hence p is 1/6.

Proof of Mean and Variance

Proof of Mean and Variance

Proof of Mean and Variance

Similarly, we can obtain the mean and variance of Y:

- $\mu = \frac{1-p}{p}$
- $\sigma^2 = \frac{1-p}{p^2}$.

Estimation of p

- To estimate p, we need a sample $x_1, x_2, \ldots, x_n$,
  - Chapter 7: An estimate of p based on the sample is
$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i}.$$
  - An interval estimate of p with 95% confidence is complicated. Refer to paper "A new short exact geometric confidence interval".

Probability mass function

- In probability theory and statistics, the Poisson distribution is a discrete probability distribution for the number of events occurring in a fixed interval of time.
- If X is a random variable following a Poisson distribution with an average number of events in an unit of interval λ in a given continuous interval of length T, then the probability mass function is:

$$f(x) = \frac{e^{-\lambda T}(\lambda T)^x}{x!} \text{ where } x = 0, 1, 2, \ldots$$

For example, $f(0) = e^{-\lambda T}$. Denote this distribution by Poisson(λT).

Definition

Assumptions to be met for assuming a Poisson distribution for the the number of events:

- event can occur any time in the interval,
- event occurs with a constant rate in the interval,
- whether an event occur at a new time is independent from events that occurred at previous times .

Time can be replaced by distance, space, or other conceptual units.

Example

The number of mechanical failures happing in the first five years of a new car can be assumed to follow a Poisson distribution:

- Failures can occur any time in the five years.
- The risk of failing is approximately constant.
- Failures happening at different times are independent events.

Example

Suppose the average number of mechanical failures per year is 3 for a part in the system and assume the mechanical failures follow a Poisson process.

(a) What is the probability that a randomly selected year has at least one failure?

(b) What is the probability that three years have at least four failures?

Example

Example

Where do we use Poisson distributions?

- Reliability analysis where the number of mechanical failures happing in an interval is assumed to follow a Poisson distribution, the rate parameter is a function of inputs, and T is the length of the time interval.
- Spatial analysis where the number of events in a region is assumed to follow a Poisson distribution and the rate is a function of inputs, and T is the number of areal units.
- Epidemiology analysis where the number of influenza over a year is assumed to follow a Poisson distribution and the rate is a function of inputs.

Proof of Mean and Variance

First obtain the moment generating function for the distribution of X:

Proof of Mean and Variance

Poisson approximation to the Binomial

For binomial distribution with large n, we have the following approximation using Poisson distribution:
when $n \to \infty$, $p \to 0$, and $np \to \lambda T$, then

$$\text{Binomial}(n, p) \to \text{Poisson}(\lambda T),$$

i.e.

$$C_x^n p^x (1-p)^{n-x} \to \frac{e^{-\lambda T}(\lambda T)^x}{x!}.$$

So for those “large” Binomials ($n \geq 100$) and for small p (usually ≤ 0.01), we can use a Poisson with $\lambda T = np$ to approximate it!

Estimation of λ

- To estimate λ, we need a sample $x_1, x_2, \ldots, x_n$,
  - Chapter 7: An estimate of λ based on the sample is
  $$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{nT}.$$
  - An interval estimate of λ with 95% confidence (not covered in this class) is
  $$(\hat{\lambda} - 1.96\sqrt{\hat{\lambda}}, \hat{\lambda} + 1.96\sqrt{\hat{\lambda}})$$