WIIT 7751

Lesson J Guided Notes

Frequently we describe an entire family of distributions that depend upon one or more variable quantities, called parameters.

A Bernoulli probability distribution depends upon the success of the trial. If we call this α then the probability function is

$$p(x;\alpha) = \begin{cases} 1-\alpha & \text{if } x = 0 \\ \alpha & \text{if x} = 1 \\ 0 & \text{otherwise} \end{cases}$$

The notation of $p(x;\alpha)$ means the probability of observing value *x* of the random variable given the parameter α.

Typically, the parameters come after the semi colon. There are some families that have multiple parameters.

Example 1 Suppose we want to model the probability of drawing a card of the hearts suit at random from a standard deck of playing cards. What would the Bernoulli distribution look like?

- What is the probability of drawing a hearts card?

- The probability mass function looks like:

Binomial Experiments

A binomial experiment satisfies the following conditions:

1. The experiment consists of n trials, where n is fixed in advance
2. The trials are identical, and each trial can result in either a success (S) or failure (F)
3. The trials are independent, so the outcome on any particular trial does not affect the other outcomes
4. The probability of a success, called p, does not vary from trial to trial

Binomial pmf

Let X be a random variable that measures the number of successes in n trials where the probability of success on each trial is p.

Then the pmf follows a binomial distribution with parameters n and p and is given by

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

Where $\binom{n}{x}$ represents the number of ways to choose x of n objects, without regard to the order in which they're chosen. $\binom{n}{x}$ is sometimes called a **binomial coefficient**, and is calculated by

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

where $n!=n(n-1)(n-2)\dots(3)(2)(1)$, and $0!=1$

Using R for Binomial Calculations

We could perform calculations by hand, however computers and calculators with the functionality to carry out the calculations are readily available. Since a graphing calculator is not required for the course, we will just use R for these calculations.

In R, each distribution has a set of four function, with prefixes of d, p, q, and r. So for the binomial distribution, we have dbinom, pbinom, qbinom, and rbinom.

Below is a section of the R help. You can see this by typing a question mark (?) before any of the above listed functions.

Binomial {stats} R Documentation

The Binomial Distribution

Description

Density, distribution function, quantile function and random generation for the binomial distribution with parameters `size` and `prob`.

This is conventionally interpreted as the number of 'successes' in `size` trials.

Usage

```
dbinom(x, size, prob, log = FALSE)
pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)
qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
rbinom(n, size, prob)
```

Arguments

`x, q`	vector of quantiles.
`p`	vector of probabilities.
`n`	number of observations. If `length(n) > 1`, the length is taken to be the number required.
`size`	number of trials (zero or more).
`prob`	probability of success on each trial.
`log, log.p`	logical; if TRUE, probabilities p are given as log(p).
`lower.tail`	logical; if TRUE (default), probabilities are $P[X \le x]$, otherwise, $P[X > x]$.

Details

The binomial distribution with `size` = n and `prob` = p has density

$$p(x) = choose(n, x) p^x (1-p)^(n-x)$$

for $x = 0, \ldots, n$. Note that binomial *coefficients* can be computed by `choose` in R.

If an element of `x` is not integer, the result of `dbinom` is zero, with a warning.

$p(x)$ is computed using Loader's algorithm, see the reference below.

The quantile is defined as the smallest value x such that $F(x) \ge p$, where F is the distribution function.

Value

`dbinom` gives the density, `pbinom` gives the distribution function, `qbinom` gives the quantile function and `rbinom` generates random deviates.

Example 2 Tay-Sachs disease is a rare but fatal disease of genetic origin occurring chiefly in infants and children, especially those of Jewish or eastern European extraction. If a couple are both carriers of Tay-Sachs disease, a child of theirs has probability 0.25 of being born with the disease. If such a couple has four children, what is the pmf for the number of children that will have the disease?

Example 3 Sickle-cell anemia is a disease that results when a person has two copies of a certain recessive gene. People with one copy of the gene are called carriers. Carriers do not have the disease, but can pass the gene on to their children. A child born to parents who are both carriers has probability 0.25 of having sickle-cell anemia. A medical study samples 18 children in families where both parents are carriers.

a) What is the probability that four or more of the children have sickle-cell anemia?

b) What is the probability that fewer than three of the children have sickle-cell anemia?

Mean, Variance, and Standard Deviation

Let X be a binomial random variable with n trials and success probability p.

Then the **mean** of X is

$$\mu_x = np$$

The **variance** of X is

$$\sigma_x^2 = np(1-p)$$

The **standard deviation** of X is

$$\sigma_x = \sqrt{np(1-p)}$$

Example 4 The probability that a new car of a certain model will require repairs during the warranty period is 0.15. A particular dealership sells 25 such cars. Let X be the number that will require repairs during the warranty period.

Find the mean and standard deviation of X.