Discrete Random Variables

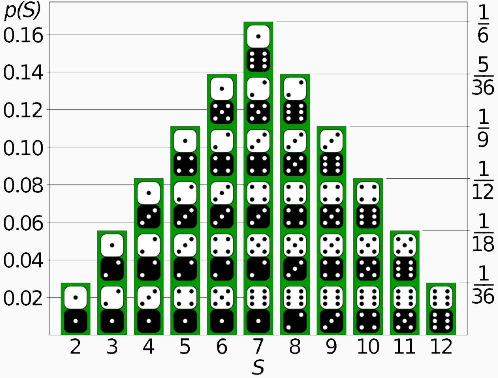
Gonzalo G. Peraza Mues

February 13, 2017

Definition: Random Variable

A random variable is a function that maps events from a random experiments to numerical values.

Example



Example: Sexes of 3 children

$$S = \{(b,b,b),(b,b,g),(b,g,b),(b,g,g),\\ (g,b,b),(g,b,g),(g,g,b),(g,g,g)\}$$

Let X be the number of female children.

$$P(X=0) = \frac{1}{8}$$
$$P(X=1) = \frac{3}{8}$$
$$P(X=2) = \frac{3}{8}$$
$$P(X=3) = \frac{1}{8}$$

Discrete Random Variable

A random variable is said to be discrete if its possible values constitute a sequence of separated points on the number line.

$P(X = x_i)$ is the probability that $X = x_i$.

The collection of these probabilities is called the probability distribution of X

$$\sum_{i=1}^{n} P(X = x_i) = 1$$

Example

Suppose that X is a random variable that takes on one of the values 1, 2, or 3. If

$$P(X = 1) = 0.4 \quad \text{and} \quad P(X = 2) = 0.1$$

what is $P(X = 3)$?

Example

A saleswoman has scheduled two appointments to sell encyclopedias. She feels that her first appointment will lead to a sale with probability 0.3. She also feels that the second will lead to a sale with probability 0.6 and that the results from the two appointments are independent. What is the probability distribution of X, the number of sales made?

X can take on values 0, 1, 2.

Expected Value

The expected value of X is a weighted average of the possible values of X:

$$\mathrm{E}[X] = \sum_{i=1}^{n} x_i P(X = x_i)$$

If X is a random variable associated with some experiment, then the average value of X over a large number of replications of the experiment is approximately $\mathrm{E}[X]$.

Example

Suppose we roll a die that is equally likely to have any of its 6 sides appear face up. Find $E[X]$, where X is the side facing up.

Example

Consider a random variable X that takes on either the value 1 or 0 with respective probabilities p and $1 - p$. That is,

$$P(X = 1) = p \quad \text{and} \quad P(X = 0) = 1 - p$$

Find $\mathrm{E}[X]$.

Example

An insurance company sets its annual premium on its life insurance policies so that it makes an expected profit of 1 percent of the amount it would have to pay out upon death. Find the annual premium on a \$200,000 life insurance policy for an individual who will die during the year with probability 0.02.

Properties of expected values

$$\mathrm{E}[cX] = c\,\mathrm{E}[X]$$

$$\mathrm{E}[X + c] = \mathrm{E}[X] + c$$

$$\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]$$

$$\mathrm{E}\left[\sum_{i=1}^{k} X_i\right] = \sum_{i=1}^{k} \mathrm{E}[X_i]$$

Example

A married couple works for the same employer. The wife's Christmas bonus is a random variable whose expected value is \$1500.

1. If the husband's bonus is set to equal 80 percent of his wife's, find the expected value of the husband's bonus.
2. If the husband's bonus is set to equal \$1000 more than his wife's, find its expected value.

Example

The following table lists the number of civilian full-time law enforcement employees in eight cities in 1990.

City	Civilian law enforcement employees
Minneapolis, MN	105
Newark, N J	155
Omaha, NE	149
Portland, OR	195
San Antonio, TX	290
San Jose, CA	357
Tucson, AZ	246
Tulsa, OK	178

Suppose that two of the cities are to be randomly chosen and all the civilian law enforcement employees of these cities are to be interviewed. Find the expected number of people who will be interviewed.

Example

A building contractor has sent in bids for three jobs. If the contractor obtains these jobs, they will yield respective profits of 20, 25, and 40 (in units of \$1000). On the other hand, for each job the contractor does not win, he will incur a loss (due to time and money already spent in making the bid) of 2. If the probabilities that the contractor will get these jobs are, respectively, 0.3, 0.6, and 0.2, what is the expected total profit?

Variance and standard deviation of random variables

If X is a random variable with expected value μ, then the variance of X, denoted by $\mathrm{Var}(X)$, is defined by

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - \mu^2$$

The standard deviation of X, denoted by $SD(X)$, is

$$SD(X) = \sqrt{\mathrm{Var}(X)}$$

Example

Find $\mathrm{Var}(X)$ when the random variable X is such that

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

Example

The return from a certain investment (in units of \$1000) is a random variable X with probability distribution

$$P(X = -1) = 0.7 \quad P(X = 4) = 0.2 \quad P(X = 8) = 0.1$$

Find $\mathrm{Var}(X)$, the variance of the return.

Properties of variances

$$\mathrm{Var}(cX) = c^2 \, \mathrm{Var}(X)$$

$$\mathrm{Var}(X + c) = \mathrm{Var}(X)$$

Independent random variables

Random variables X and Y are independent if knowing the value of one of them does not change the probabilities of the other.

If X and Y are independent random variables, then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

More generally

$$\mathrm{Var}\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} \mathrm{Var}(X_i)$$

Example

Determine the variance of the sum obtained when a pair of fair dice is rolled.

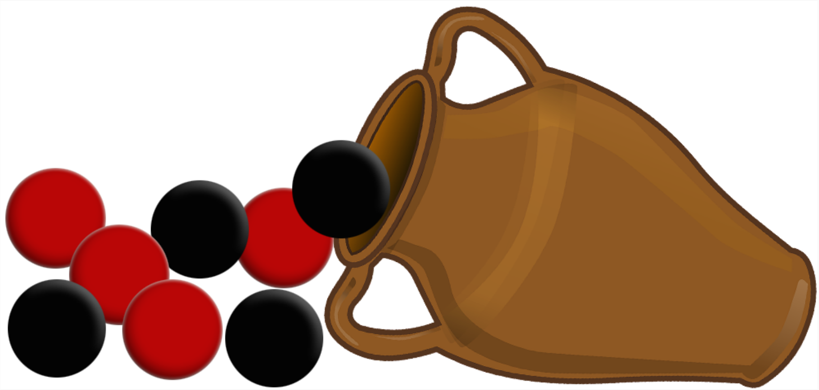
Example

The annual gross earnings of a certain rock singer are a random variable with an expected value of \$400,000 and a standard deviation of \$80,000. The singer's manager receives 15 percent of this amount. Determine the expected value and standard deviation of the amount received by the manager.

The urn problem

Pick a ball, return the ball. Repeat n times.

What is the probability of picking X red balls?



Binomial Random Variables

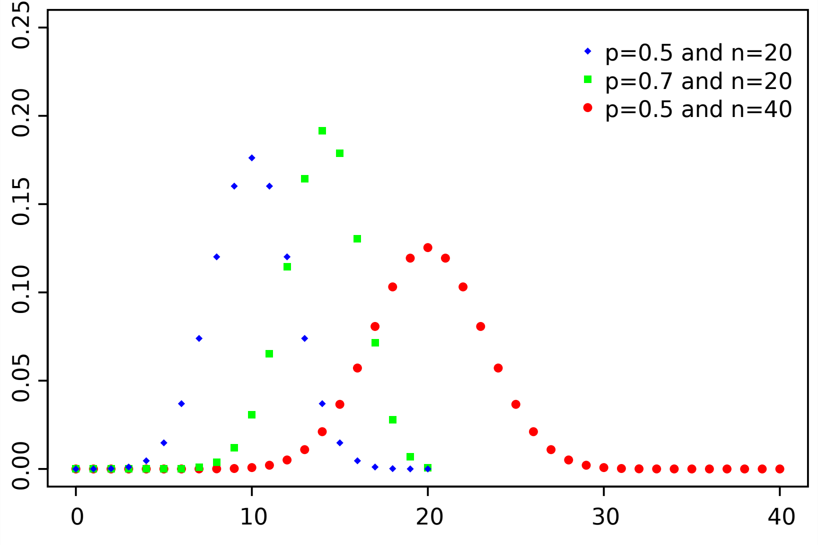
A binomial random variable with parameters n and p represents the number of successes in n independent trials, when each trial is a success with probability p. If X is such a random variable, then for $i = 0, \ldots, n$, the probability mass function is

$$P(X = i) = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

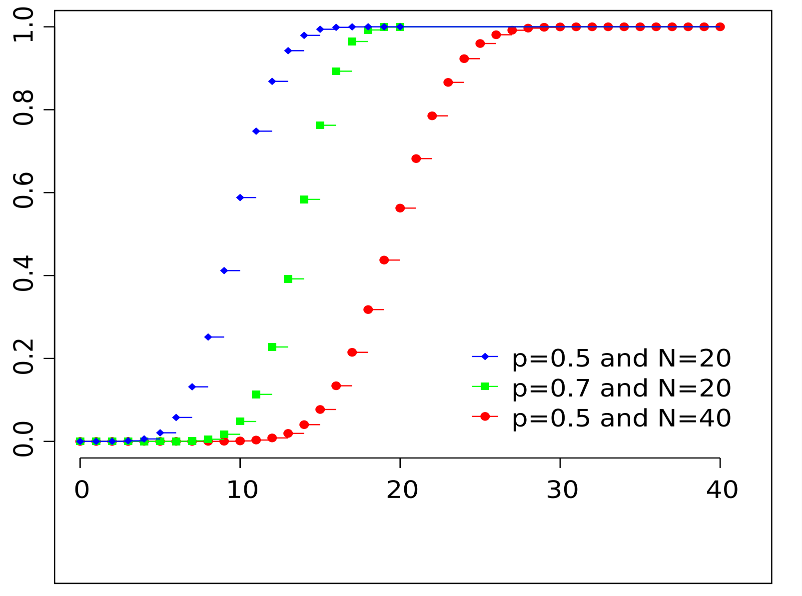
and the cumulative distribution function is

$$P(X \leq j) = \sum_{i=0}^{j} \binom{n}{i} p^i (1-p)^{n-i}$$

Probability mass function



Cumulative distribution function



Example

Three fair coins are flipped. If the outcomes are independent, determine the probability that there are a total of i heads, for $i = 0, 1, 2, 3$.

Example

Suppose that a particular trait (such as eye color or handedness) is determined by a single pair of genes, and suppose that d represents a dominant gene and r a recessive gene. A person with the pair of genes (d, d) is said to be pure dominant, one with the pair (r, r) is said to be pure recessive, and one with the pair (d, r) is said to be hybrid. The pure dominant and the hybrid are alike in appearance. When two individuals mate, the resulting offspring receives one gene from each parent, and this gene is equally likely to be either of the parent's two genes.

1. What is the probability that the offspring of two hybrid parents has the opposite (recessive) appearance?
2. Suppose two hybrid parents have 4 offsprings. What is the probability 1 of the 4 offspring has the recessive appearance?

Example

1. Determine $P(X \leq 12)$ when X is a binomial random variable with parameters 20 and 0.4.
2. Determine $P(Y \leq 10)$ when Y is a binomial random variable with parameters 16 and 0.5.

Use the R function `pbinom`

Expected Value and Variance of a Binomial Random Variable

If X is binomial with parameters n and p, then

$$
\begin{aligned}
\mathrm{E}[X] &= np \\
\mathrm{Var}(X) &= np(1-p)
\end{aligned}
$$

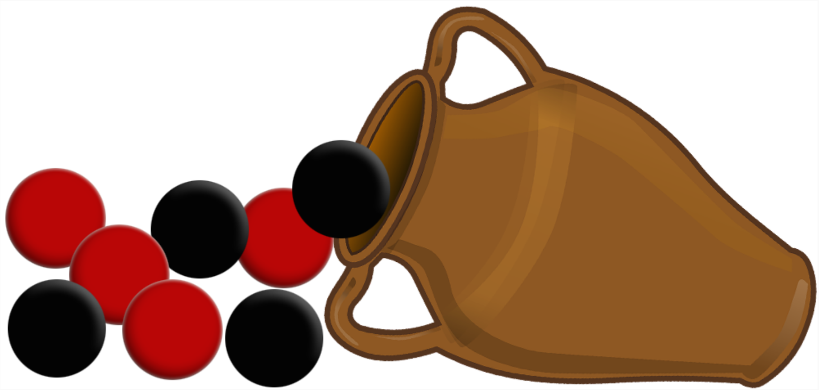
Example

Suppose that each screw produced is independently defective with probability 0.01. Find the expected value and variance of the number of defective screws in a shipment of size 1000.

The urn problem, again

Pick a ball, keep the ball. Repeat n times.

What is the probability of picking X red balls?



Hypergeometric Distribution

The hypergeometric distribution describes the probability of X successes in n draws, without replacement, from a finite population of size N that contains exactly K successes. The probability mass function is

$$P(X = i) = \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}$$

When N is large in relation to n, and p=K/N, a hypergeometric random variable with parameters n, N, K approximately has a binomial distribution with parameters n and p.

Expected Value and Variance for a Hypergeometric distribution

$$
\begin{aligned}
\mathrm{E}[X] &= np \\
\mathrm{Var}(X) &= \frac{N-n}{N-1} np(1-p)
\end{aligned}
$$

Example

If 6 people are randomly selected from a group consisting of 12 men and 8 women, then the number of women chosen is a hypergeometric random variable with parameters $n = 6$, $N = 20$, $p = 8/20 = 0.4$. Similarly, the number of men chosen is a hypergeometric random variable with parameters $n = 6$, $N = 20$, $p = 0.6$. What is the mean and variance of both groups?

Poisson Random Variables

A random variable X is called a Poisson random variable with parameter λ if

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

Binomial random variables with large values of n and small values of p can be approximated by a Poisson random variable with $\lambda = np$.

If X is a Poisson random variable with parameter $\lambda > 0$, then

$$\begin{aligned} \mathrm{E}[X] =& \lambda \\ \mathrm{Var}(x) =& \lambda \end{aligned}$$

Example

Suppose that items produced by a certain machine are independently defective with probability 0.1. What is the probability that a sample of 10 items will contain at most 1 defective item? What is the Poisson approximation for this probability?

Example

Suppose the average number of accidents occurring weekly on a particular highway is equal to 1.2. Approximate the probability that there is at least one accident this week.

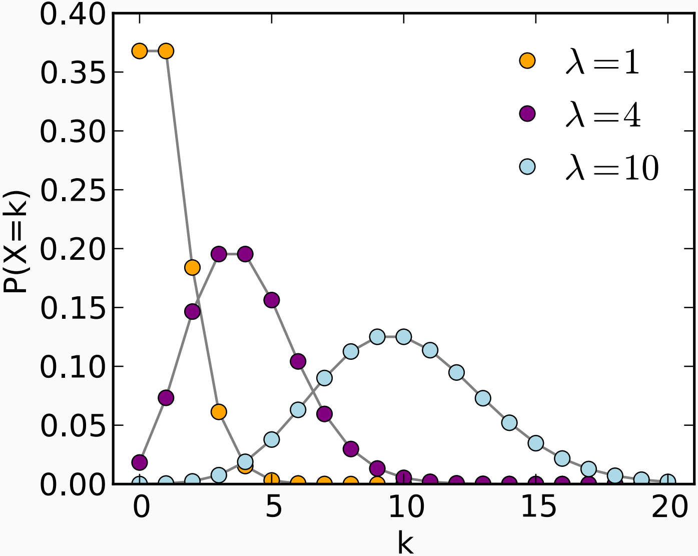
The Poisson Distribution

It expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate (λ) and independently of the time since the last event.

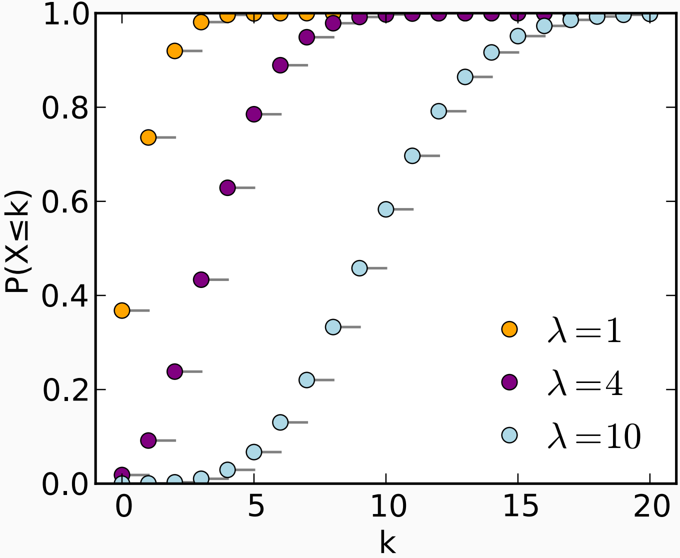
Examples:

- The number of meteors greater than 1 meter diameter that strike earth in a year.
- The number of patients arriving in an emergency room between 11 and 12 pm.

Probability mass function



Cumulative distribution function



Example

On a particular river, overflow floods occur once every 100 years on average. Calculate the probability of k = 0, 1, 2, 3, 4, 5, or 6 overflow floods in a 100-year interval, assuming the Poisson model is appropriate.

Example

Ugarte and colleagues report that the average number of goals in a World Cup soccer match is approximately 2.5 and the Poisson model is appropriate.

Derivation

The Poisson distribution is the limit of the Binomial distribution as a function of $\lambda = np$ when $n \to \infty$. Hint: $e^x = \lim_{n\to\infty}(1 + x/n)^n$.