WILEY

Applied Statistics and Probability for Engineers

Sixth Edition

Douglas C. Montgomery   George C. Runger

Chapter 3

Discrete Random Variables and Probability Distributions

3 Discrete Random Variables and Probability Distributions

CHAPTER OUTLINE

3-1 Discrete Random Variables
3-2 Probability Distributions and Probability Mass Functions
3-3 Cumulative Distribution Functions
3-4 Mean and Variance of a Discrete Random Variable
3-5 Discrete Uniform Distribution
3-6 Binomial Distribution
3-7 Geometric and Negative Binomial Distributions
3-8 Hypergeometric Distribution
3-9 Poisson Distribution

WILEY

Learning Objectives for Chapter 3

After careful study of this chapter, you should be able to do the following:

1. Determine probabilities from probability mass functions and the reverse.
2. Determine probabilities and probability mass functions from cumulative distribution functions and the reverse.
3. Calculate means and variances for discrete random variables.
4. Understand the assumptions for discrete probability distributions.
5. Select an appropriate discrete probability distribution to calculate probabilities.
6. Calculate probabilities, means and variances for discrete probability distributions.

WILEY

Discrete Random Variables

Physical systems can be modeled by the same or similar random experiments and random variables. The distribution of the random variable involved in each of these common systems can be analyzed. The results can be used in different applications and examples.

We often omit a discussion of the underlying sample space of the random experiment and directly describe the distribution of a particular random variable.

Example 3-2: Camera Flash Tests

The time to recharge the flash is tested in three cell-phone cameras. The probability that a camera passes the test is 0.8, and the cameras perform independently. See Table 3-1 for the sample space for the experiment and associated probabilities. For example, because the cameras are independent, the probability that the first and second cameras pass the test and the third one fails, denoted as *ppf, is*

$$P(ppf) = (0.8)(0.8)(0.2) = 0.128$$

The random variable X denotes the number of cameras that pass the test. The last column of the table shows the values of X assigned to each outcome of the experiment.

Table 3-1 Camera Flash Tests

Outcome				
Camera #				
1	2	3	Probability	X
Pass	Pass	Pass	0.512	3
Fail	Pass	Pass	0.128	2
Pass	Fail	Pass	0.128	2
Fail	Fail	Pass	0.032	1
Pass	Pass	Fail	0.128	2
Fail	Pass	Fail	0.032	1
Pass	Fail	Fail	0.032	1
Fail	Fail	Fail	0.008	0
			1.000	

Probability Distributions

A random variable is a function that assigns a real number to each outcome in the sample space of a random experiment.

The probability distribution of a random variable X gives the probability for each value of X.

Example 3-4: Digital Channel

- There is a chance that a bit transmitted through a digital transmission channel is received in error.
- Let *X* equal the number of bits received in error in the next 4 bits transmitted.
- The associated probability distribution of *X* is shown in the table.
- The probability distribution of *X* is given by the possible values along with their probabilities.

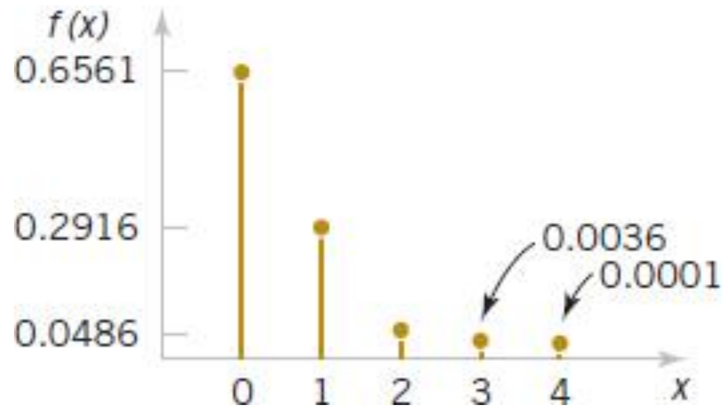

Figure 3-1 Probability distribution for bits in error.

$P(X=0)$ =	0.6561
$P(X=1)$ =	0.2916
$P(X=2)$ =	0.0486
$P(X=3)$ =	0.0036
$P(X=4)$ =	0.0001
	1.0000

Probability Mass Function

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a probability mass function is a function such that:

$$(1)\quad f(x_i) \geq 0$$

$$(2)\quad \sum_{i=1}^{n} f(x_i) = 1$$

$$(3)\quad f(x_i) = P(X = x_i)$$

Example 3-5: Wafer Contamination

- Let the random variable X denote the number of wafers that need to be analyzed to detect a large particle of contamination. Assume that the probability that a wafer contains a large particle is 0.01, and that the wafers are independent. Determine the probability distribution of X.
- Let p denote a wafer in which a large particle is **p**resent & let a denote a wafer in which it is absent.
- The sample space is: $S = \{p, ap, aap, aaap, \ldots\}$
- The range of the values of X is: $x = 1, 2, 3, 4, \ldots$

Probability Distribution		
$P(X = 1) =$	0.01	0.01
$P(X = 2) =$	$(0.99)*0.01$	0.0099
$P(X = 3) =$	$(0.99)^2*0.01$	0.0098
$P(X = 4) =$	$(0.99)^3*0.01$	0.0097

Cumulative Distribution Functions

Example 3-6: Consider the probability distribution for the digital channel example.

x	$P(X=x)$
0	0.6561
1	0.2916
2	0.0486
3	0.0036
4	0.0001
	1.0000

Find the probability of three or fewer bits in error.

- The event ($X \le 3$) is the total of the events: ($X = 0$), ($X = 1$), ($X = 2$), and ($X = 3$).
- From the table:

$$P(X \le 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0.9999$$

Cumulative Distribution Function and Properties

The cumulative distribution function, is the probability that a random variable *X* with a given probability distribution will be found at a value less than or equal to *x*.

Symbolically,

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

For a discrete random variable *X*, *F*(*x*) satisfies the following properties:

$$(1)\quad F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

$$(2)\quad 0 \le F(x) \le 1$$

$$(3)\quad \text{If } x \le y, \text{ then } F(x) \le F(y)$$

Example 3-8: Sampling without Replacement

A day's production of 850 parts contains 50 defective parts. Two parts are selected at random without replacement. Let the random variable *X* equal the number of defective parts in the sample. Find the cumulative distribution function of *X*.

The probability mass function is calculated as follows:

$$P(X=0)=\tfrac{800}{850}\cdot\tfrac{799}{849}=0.886$$

$$P(X=1)=2\cdot\tfrac{800}{850}\cdot\tfrac{50}{849}=0.111$$

$$P(X=2)=\tfrac{50}{850}\cdot\tfrac{49}{849}=0.003$$

Therefore,

$$F(0)=P(X\le 0)=0.886$$

$$F(1)=P(X\le 1)=0.997$$

$$F(2)=P(X\le 2)=1.000$$

$$F(x)=\begin{cases}0 & x<0\\ 0.886 & 0\le x<1\\ 0.997 & 1\le x<2\\ 1 & 2\le x\end{cases}$$

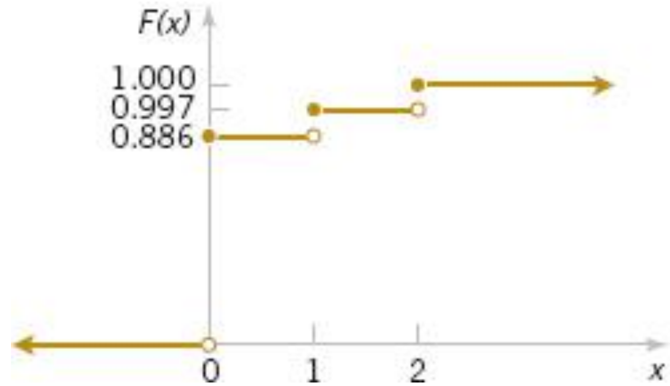

Figure 3-4 Cumulative Distribution Function

Variance Formula Derivations

$$V(X) = \sum_x (x-\mu)^2 f(x)$$ is the definitional formula

$$= \sum_x \left(x^2 - 2\mu x + \mu^2\right) f(x)$$

$$= \sum_x x^2 f(x) - 2\mu \sum xf(x) + \mu^2 \sum_x f(x)$$

$$= \sum_x x^2 f(x) - 2\mu^2 + \mu^2$$

$$= \sum_x x^2 f(x) - \mu^2$$ is the computational formula

The computational formula is easier to calculate manually.

Example 3-9: Digital Channel

In Example 3-4, there is a chance that a bit transmitted through a digital transmission channel is received in error. X is the number of bits received in error of the next 4 transmitted. The probabilities are

$$P(X = 0) = 0.6561,\ P(X = 2) = 0.0486,\ P(X = 4) = 0.0001,$$
$$P(X = 1) = 0.2916,\ P(X = 3) = 0.0036$$

Use table to calculate the mean & variance.

x	$f(x)$	$x \cdot f(x)$	$(x\text{-}0.4)^2$	$(x\text{-}0.4)^2 \cdot f(x)$	$x^2 \cdot f(x)$
0	0.6561	0.0000	0.160	0.1050	0.0000
1	0.2916	0.2916	0.360	0.1050	0.2916
2	0.0486	0.0972	2.560	0.1244	0.1944
3	0.0036	0.0108	6.760	0.0243	0.0324
4	0.0001	0.0004	12.960	0.0013	0.0016
	Total =	0.4000		0.3600	0.5200
		= Mean		= Variance (σ^2)	$= E(x^2)$
		$= \mu$		$\sigma^2 = E(x^2) - \mu^2 =$	0.3600
				Computational formula	

Expected Value of a Function of a Discrete Random Variable

If *X* is a discrete random variable with probability mass function *f* (*x*),

$$E\left[h(X)\right] = \sum_{x} h(x) f(x)$$

If $h(x) = (X - \mu)^2$, then its expectation is the variance of *X.*

Example 3-12: Digital Channel

In Example 3-9, X is the number of bits in error in the next four bits transmitted. What is the expected value of the square of the number of bits in error?

X	0	1	2	3	4
$f(X)$	0.6561	0.2916	0.0486	0.0036	0.0001

Here $h(X) = X^2$

Answer:

$E(X^2) = X^2 \cdot f(X) = 0^2 \text{ x } 0.6561 + 1^2 \text{ x } 0.2916 + 2^2 \text{ x } 0.0486 + 3^2 \text{ x } 0.036 + 4^2 \text{ x } 0.0001$

$= 0.5200$

Discrete Uniform Distribution

If the random variable X assumes the values x_1, x_2, …, x_n, with equal probabilities, then the discrete uniform distribution is given by

$$f(x_i) = 1/n$$

Discrete Uniform Distribution

- Let X be a discrete random variable ranging from $a, a+1, a+2, \ldots, b,$ for $a \le b$. There are $b - (a-1)$ values in the inclusive interval. Therefore:

$$f(x) = 1/(b-a+1)$$

- Its measures are:

$$\mu = E(x) = (b+a)/2$$

$$\sigma^2 = V(x) = [(b-a+1)^2 - 1]/12$$

Note that the mean is the midpoint of a & b.

WILEY

Example 3-14: Number of Voice Lines

Let the random variable X denote the number of the 48 voice lines that are in use at a particular time. Assume that X is a discrete uniform random variable with a range of 0 to 48. Find $E(X)$ & σ.

Answer:

$$\mu = \frac{48+0}{2} = 24$$

$$\sigma_X = \sqrt{\frac{(48-0+1)^2-1}{12}} = \sqrt{\frac{2400}{12}} = 14.14$$

Binomial Distribution

The random variable X that equals the number of trials that result in a success is a binomial random variable with parameters $0 < p < 1$ and $n = 1, 2, \ldots$.

The probability mass function is:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \ldots n$$

For constants *a* and *b*, the binomial expansion is

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

WILEY

Example 3-17: Binomial Coefficient

Exercises in binomial coefficient calculation:

$$\binom{10}{3} = \frac{10!}{3!7!} = \frac{10\cdot 9\cdot 8\cdot 7!}{3\cdot 2\cdot 1\cdot 7!} = 120$$

$$\binom{15}{10} = \frac{15!}{10!5!} = \frac{15\cdot 14\cdot 13\cdot 12\cdot 11.10!}{10!.5\cdot 4\cdot 3\cdot 2\cdot 1} = 3,003$$

$$\binom{100}{4} = \frac{100!}{4!96!} = \frac{100\cdot 99\cdot 98\cdot 97.96!}{4\cdot 3\cdot 2\cdot 1.96!} 3,921,225$$

Exercise 3-18: Organic Pollution-1

Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Find the probability that, in the next 18 samples, exactly 2 contain the pollutant.

Answer:

Let X denote the number of samples that contain the pollutant in the next 18 samples analyzed. Then X is a binomial random variable with $p = 0.1$ and $n = 18$

$$P(X=2)=\binom{18}{2}(0.1)^2(0.9)^{16}=153(0.1)^2(0.9)^{16}=0.2835$$

0.2835	=BINOMDIST(2,18,0.1,FALSE)

WILEY

Exercise 3-18: Organic Pollution-2

Determine the probability that at least 4 samples contain the pollutant.

Answer:

$$P(X \geq 4) = 1 - P(X < 4)$$

$$= 1 - \sum_{x=0}^{3} \binom{18}{x} (0.1)^x (0.9)^{18-x}$$

$$= 1 - [0.150 + 0.300 + 0.284 + 0.168]$$

$$= 0.098$$

0.0982	= 1 - BINOMDIST(3,18,0.1,TRUE)

Exercise 3-18: Organic Pollution-3

Now determine the probability that $3 \le X < 7$.

Answer:

$$P(3 \le X < 7) = \sum_{x=3}^{6} \binom{18}{x} (0.1)^x (0.9)^{18-x}$$

$$= 0.168 + 0.070 + 0.022 + 0.005$$

$$= 0.265$$

0.2660	= BINOMDIST(7,18,0.1,TRUE) - BINOMDIST(2,18,0.1,TRUE)

Appendix A, Table II (pg. 705) is a cumulative binomial table for selected values of *p* and *n*.

Binomial Mean and Variance

If X is a binomial random variable with parameters p and n,

$$\mu = E(X) = np$$

and

$$\sigma^2 = V(X) = np(1-p)$$

Example 3-19:

For the number of transmitted bit received in error in Example 3-16, n = 4 and p = 0.1. Find the mean and variance of the binomial random variable.

Answer:

$$\mu = E(X) = np = 4*0.1 = 0.4$$

$$\sigma^2 = V(X) = np(1-p) = 4*0.1*0.9 = 0.36$$

$$\sigma = SD(X) = 0.6$$

Geometric Distribution

- Binomial distribution has:
  - Fixed number of trials.
  - Random number of successes.

- Geometric distribution has reversed roles:
  - Random number of trials.
  - Fixed number of successes, in this case 1.

- The probability density function of Geometric distribution is

$$f(x) = p(1-p)^{x-1}$$

$x = 1, 2, \ldots \infty$, the number of failures until the 1st success. $0 < p < 1$, the probability of success.

Example 3.21: Wafer Contamination

The probability that a wafer contains a large particle of contamination is 0.01. Assume that the wafers are independent. What is the probability that exactly 125 wafers need to be analyzed before a particle is detected?

Answer:

Let X denote the number of samples analyzed until a large particle is detected. Then X is a geometric random variable with parameter $p = 0.01$.

$$P(X=125) = (0.99)^{124}(0.01) = 0.00288.$$

Geometric Mean & Variance

If *X* is a geometric random variable with parameter *p*,

$$\mu = E(X) = \frac{1}{p}$$

and

$$\sigma^2 = V(X) = \frac{(1-p)}{p^2}$$

Exercise 3-22: Mean and Standard Deviation

The probability that a bit transmitted through a digital transmission channel is received in error is 0.1. Assume that the transmissions are independent events, and let the random variable X denote the number of bits transmitted until the first error. Find the mean and standard deviation.

Answer:

$$\text{Mean} = \mu = E(X) = 1 / p = 1 / 0.1 = 10$$

$$\text{Variance} = \sigma^2 = V(X) = (1-p) / p^2 = 0.9 / 0.01 = 90$$

$$\text{Standard deviation} = \sqrt{90} = 9.49$$

Lack of Memory Property

For a geometric random variable, the trials are independent. Thus the count of the number of trials until the next success can be started at any trial without changing the probability distribution of the random variable.

The implication of using a geometric model is that the system presumably will not wear out. For all transmissions the probability of an error remains constant. Hence, the geometric distribution is said to lack any memory.

Example 3-23: Lack of Memory Property

In Example 3-21, the probability that a bit is transmitted in error is 0.1. Suppose 50 bits have been transmitted. What is the mean number of bits transmitted until the next error?

Answer:

The mean number of bits transmitted until the next error, after 50 bits have already been transmitted, is

$$1 / 0.1 = 10.$$

the same result as the mean number of bits until the first error.

Negative Binomial Distribution

In a series of independent trials with constant probability of success p, the random variable X which equals the number of trials until r successes occur is a negative binomial random variable with parameters $0 < p < 1$ and r = 1, 2, 3, ...

The probability mass function is:

$$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad \text{for } x = r, r+1, r+2 \ldots$$

Mean & Variance of Negative Binomial

If X is a negative binomial random variable with parameters p and r,

$$\mu = E(X) = \frac{r}{p}$$

and

$$\sigma^2 = V(X) = \frac{r(1-p)}{p^2}$$

Example 3-25: Camera Flashes

The probability that a camera passes a particular test is 0.8, and the cameras perform independently. What is the probability that the third failure is obtained in five or fewer tests?

Let X denote the number of cameras tested until three failures have been obtained. The requested probability is $P(X \leq 5)$. Here X has a negative binomial distribution with $p = 0.2$ and $r = 3$. Therefore,

$$
\begin{aligned}
P(X \leq 5) &= \sum_{x=3}^{5} \binom{x-1}{2} (0.2)^3 (0.8)^{x-3} \\
&= 0.2^3 + \binom{3}{2} 0.2^3 (0.8) + \binom{4}{2} 0.2^3 (0.8)^2 \\
&= 0.056
\end{aligned}
$$

Hypergeometric Distribution

- A set of N objects contains:

  K objects classified as success

  $N - K$ objects classified as failures

- A sample of size n objects is selected without replacement from the N objects randomly, where $K \le N$ *and* $n \le N$.
- Let the random variable X denote the number of successes in the sample. Then X is a hypergeometric random variable with probability density function

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \text{ where } x = \max(0, n+K-N) \text{ to } \min(K, n)$$

Example 3-27: Parts from Suppliers-1

A batch of parts contains 100 parts from supplier A and 200 parts from Supplier B. If 4 parts are selected randomly, without replacement, what is the probability that they are all from Supplier A?

Answer:

Let X equal the number of parts in the sample from Supplier A.

$$P(X=4)=\frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}=0.0119$$

In Excel	
0.01185	=HYPGEOMDIST(4,100,4,300)

WILEY

Example 3-27: Parts from Suppliers-2

What is the probability that two or more parts are from supplier A?

Answer:

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4)$$

$$= \frac{\binom{100}{2}\binom{200}{2}}{\binom{300}{4}} + \frac{\binom{100}{3}\binom{200}{1}}{\binom{300}{4}} + \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}$$

$$= 0.298 + 0.098 + 0.0119$$

$$= 0.408$$

In Excel	
0.40741	= HYPGEOMDIST(2,100,4,300)
	+ HYPGEOMDIST(3,100,4,300)
	+ HYPGEOMDIST(4,100,4,300)

Example 3-27: Parts from Suppliers-3

What is the probability that at least one part in the sample is from Supplier A?

Answer:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{100}{0}\binom{200}{4}}{\binom{300}{4}} = 0.804$$

In Excel	
0.80445	= 1 - HYPGEOMDIST(0,100,4,300)

Hypergeometric Mean & Variance

If *X* is a hypergeometric random variable with parameters *N*, *K*, and *n*, then

$$\mu = E(X) = np \quad \text{and} \quad \sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$$

where $p = K/N$

and $\left(\frac{N-n}{N-1}\right)$ is the finite population correction factor.

σ^2 approaches the binomial variance as n/N becomes small.

Example 3-29: Customer Sample-1

A list of customer accounts at a large company contains 1,000 customers. Of these, 700 have purchased at least one of the company's products in the last 3 months. To evaluate a new product, 50 customers are sampled at random from the list. What is the probability that more than 45 of the sampled customers have purchased from the company in the last 3 months?

Let X denote the number of customers in the sample who have purchased from the company in the last 3 months. Then X is a hypergeometric random variable with N = 1,000, K = 700, n = 50.

$$P(X > 45) = \sum_{x=46}^{50} \frac{\binom{700}{x}\binom{300}{50-x}}{\binom{1,000}{50}}$$

This a lengthy problem! ☹

Example 3-29: Customer Sample-2

Using the binomial approximation to the distribution of *X results in*

$$P(X > 45) = \sum_{x=46}^{50} \binom{50}{x} 0.7^x (1-0.7)^{50-x} = 0.00017$$

In Excel	
0.000172	= 1 - BINOMDIST(45, 50, 0.7, TRUE)

WILEY

Poisson Distribution

The random variable *X* that equals the number of events in a Poisson process is a Poisson random variable with parameter λ > 0, and the probability density function is:

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

Example 3-31: Calculations for Wire Flaws-1

For the case of the thin copper wire, suppose that the number of flaws follows a Poisson distribution With a mean of 2.3 flaws per mm. Find the probability of exactly 2 flaws in 1 mm of wire.

Answer:

Let *X* denote the number of flaws in 1 mm of wire

$$P(X=2)=\frac{e^{-2.3}2.3^2}{2!}=0.265$$

In Excel	
0.26518	= POISSON(2, 2.3, FALSE)

WILEY

Example 3-31: Calculations for Wire Flaws-2

Determine the probability of 10 flaws in 5 mm of wire.

Answer :

Let *X* denote the number of flaws in 5 mm of wire.

$$E(X) = \lambda = 5\text{ mm} \cdot\ 2.3\text{ flaws/mm} = 11.5\text{ flaws}$$

$$\mathrm{P}(X = 10) = e^{-11.5}\frac{11.5^{10}}{10!} = 0.113$$

In Excel	
0.1129	= POISSON(10, 11.5, FALSE)

Example 3-31: Calculations for Wire Flaws-3

Determine the probability of at least 1 flaw in 2 mm of wire.

Answer :

Let X denote the number of flaws in 2 mm of wire.

Note that P(X ≥ 1) requires ∞ terms. ☹

$$E(X) = \lambda = 2 \text{ mm} \cdot \ 2.3 \text{ flaws/mm} = 4.6 \text{ flaws}$$

$$\mathrm{P}(X \geq 1) = 1 - P(X = 0) = 1 - e^{-4.6}\frac{4.6^0}{0!} = 0.9899$$

In Excel	
0.989948	= 1 - POISSON(0, 4.6, FALSE)

Poisson Mean & Variance

If X is a Poisson random variable with parameter λ, then

$$\mu = E(X) = \lambda \quad \text{and} \quad \sigma^2 = V(X) = \lambda$$

The mean and variance of the Poisson model are the same.

For example, if particle counts follow a Poisson distribution with a mean of 25 particles per square centimeter, the variance is also 25 and the standard deviation of the counts is 5 per square centimeter.

If the variance of a data is much greater than the mean, then the Poisson distribution would not be a good model for the distribution of the random variable.

Important Terms & Concepts of Chapter 3

Bernoulli trial

Binomial distribution

Cumulative probability distribution – discrete random variable

Discrete uniform distribution

Expected value of a function of a random variable

Finite population correction factor

Geometric distribution

Hypergeometric distribution

Lack of memory property – discrete random variable

Mean – discrete random variable

Mean – function of a discrete random variable

Negative binominal distribution

Poisson distribution

Poisson process

Probability distribution – discrete random variable

Probability mass function

Standard deviation – discrete random variable

Variance – discrete random variable