

Stat01-Geometric.wmxm

TABLE OF CONTENTS

Preface	1
References	1
Discrete Distributions Defined	3
What Is a Discrete Random Variable?	3
Mean and Variance of Discrete Data Set	5
The Discrete Geometric (p) distribution	5
pdf_geometric (k, p)	6
Statology 1: Coin Tosses	6
cdf_geometric (k, p)	9
Maxima Function geomCalc (k, p)	10
Statology 2: Number of Bankruptcies	12
PSU Example 11.1	12
mean_geometric (p)	13
var_geometric (p)	14
std_geometric (p)	14
Statology 3: Supporters of a Law	15
Statology 4: Chances of Network Failures	16
Calcworks Ex. 1	17
Stat-howto Ex. 1	18
Inverse type of problem	19
quantile_geometric (q, p)	19
Kidney Donor Scenario	20
Rolling a Fair Dice	21
Winning a Prize from a Claw Machine	22
random_geometric (p), random_geometric (p, n)	23
100 Simulated Bernoulli Trials with $p = 0.2$	24
discrete_freq (data)	24
1000 Simulated Bernoulli Trials with $p = 0.2$	26

1 *Preface*

In Stat01-Geometric.wmxm we discuss the discrete Geometric (p) probability distribution and its application, using Maxima tools and methods.

Edwin L. (Ted) Woollett
<https://home.csulb.edu/~woollett/>
 April 1, 2024

2 *References*

In our series Statistics with Maxima we have used some examples and explanations (with much editing and additions) from:

Ch. 3 Reagle & Salvatore [RS], Statistics and Econometrics, 2nd ed, Schaum's Outlines, 2011, McGraw Hill,

Ch. 8 Fred Senese [FS], Symbolic Mathematics for Chemists: A Guide for Chemists, 2019, Wiley,

Louis Lyons, Statistics for Nuclear and Particle Physics, 1986, Cambridge Univ. Press,

Luca Lista, 'Statistical Methods for Data Analysis in Particle Physics',
Lecture Notes in Physics 909, 2016, Springer-Verlag,

Frederick James, 'Statistical Methods in Experimental Physics', 2nd ed.,
2006, World Scientific.

<https://www.statology.org/geometric-distribution-real-life-examples/>

<https://online.stat.psu.edu/stat414/lesson/11/11.1>

<https://calcworkshop.com/discrete-probability-distribution/geometric-distribution/>

<https://www.statisticshowto.com/geometric-distribution/>

https://minisham.redbrick.dcu.ie/CA/Notes/CA266/10_Geometric_Distribution.pdf

<https://www.studysmarter.co.uk/explanations/math/statistics/geometric-distribution/>

Homemade functions fill, head, tail, Lsum are useful for looking at long lists.

```
(%i4) load (descriptive);  
load (distrib);  
fpprintprec : 6$  
ratprint : false$
```

```
(%o1)
```

```
C:/maxima-5.43.2/share/maxima/5.43.2/share/descriptive/descriptive.mac
```

```
(%o2)
```

```
C:/maxima-5.43.2/share/maxima/5.43.2/share/distrib/distrib.mac
```

```
(%i12) fill ( aL ) := [ first (aL), last (aL), length (aL) ]$
declare (fill, evfun)$
head(L) := if listp (L) then rest (L, - (length (L) - 3) ) else
error("Input to 'head' must be a list of expressions ")$
declare(head, evfun)$
tail (L) := if listp (L) then rest (L, length (L) - 3 ) else
error("Input to 'tail' must be a list of expressions ")$
declare(tail, evfun)$
Lsum (aList) := apply ("+", aList)$
declare (Lsum, evfun)$
```

3 *Discrete Distributions Defined*

From www.investopedia.com:

"A discrete probability distribution counts occurrences that have countable or finite outcomes.

Discrete distributions contrast with continuous distributions, where outcomes can fall anywhere on a continuum.

Common examples of discrete distribution include the binomial, Poisson, and Bernoulli distributions.

These distributions often involve statistical analyses of "counts" or "how many times" an event occurs.

In finance, discrete distributions are used in options pricing and forecasting market shocks or recessions."

From <http://www.stat.yale.edu/Courses/1997-98/101/ranvar.htm>:

"If a random variable can take only a finite number of distinct values, then it must be discrete. Examples of discrete random variables include the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery, the number of defective light bulbs in a box of ten."

4 *What Is a Discrete Random Variable?*

Quoting Luca Lista (Sec. 1.1):

"Many processes in nature have uncertain outcomes. This means that their result cannot be predicted before the process occurs. A random process is a process that can be reproduced, to some extent, within some given boundary and initial conditions, but whose outcome is uncertain. This situation may be due to insufficient information about the process intrinsic dynamics which prevents to predict its outcome, or lack of sufficient accuracy in reproducing the initial conditions in order to ensure its exact reproducibility. Some processes like quantum mechanics phenomena have intrinsic randomness. This will lead to possibly different outcomes if the experiment is repeated several times, even if each time the initial conditions are exactly reproduced, within the possibility of control of the experimenter. Probability is a measurement of how favored one of the possible outcomes of such a random process is compared with any of the other possible outcomes."

A coin toss is "random" because we are ignorant of the 'initial conditions'. Repeated trials tell us something about how those initial conditions vary between trials

For the purposes of calculating things for experimental physics, we need physical probability. In particular we need 'frequentist probability':

"Probability is the frequency with which a particular outcome occurs in repeated trials."

$P = (\text{number of occasions on which that outcome occurs}) / (\text{total number of measurements})$.

Quoting L. Lyons, Sec. 2.1:

"In many situations we deal with experiments in which the essential circumstances are kept constant, and yet repetitions of the experiment produce different results. Thus the result of an individual measurement or trial may be unpredictable, and yet the possible results of a series of such measurements have a well defined distribution."

What about events that can't be repeated? They don't have probabilities.

Quoting [RS] Sec. 3.3:

"A random variable is a variable whose values are associated with some probability of being observed. A discrete (as opposed to continuous) random variable is one that can assume only finite and distinct values. The set of all possible values of a random variable and its associated probabilities is called a probability distribution. The sum of all probabilities equals 1."

Quoting

https://saylordotorg.github.io/text_introductory-statistics/s08-discrete-random-variables.html,

"The probability distribution of a discrete random variable X is a listing of each possible value x taken by X along with the probability $P(x)$ that X takes that value in one trial of the experiment.

The mean μ of a discrete random variable X is a number that indicates the average value of X over numerous trials of the experiment. It is computed using the formula $\mu = \sum x P(x)$.

The variance σ^2 and standard deviation σ of a discrete random variable X are numbers that indicate the variability of X over numerous trials of the experiment. They may be computed using the formula $\sigma^2 = (\sum x^2 P(x)) - \mu^2$, taking the square root to obtain σ ."

5 Mean and Variance of Discrete Data Set

Consider a data set that contains M unique discrete values x_k , and assume the value x_k occurs with frequency f_k . Let N equal the sum of the frequencies.

$$N = \sum (f_k, k, 1, M).$$

The mean

$$\langle x \rangle = \sum (f_k x_k, k, 1, M) / N.$$

The variance

$$\text{Var}(x) = \sum (f_k (x_k - \langle x \rangle)^2, k, 1, M) / N.$$

The standard deviation is the square root of the variance.

6 The Discrete Geometric (p) distribution

Quoting

<https://www.statology.org/geometric-distribution-real-life-examples/>

"The Geometric distribution is a probability distribution that is used to model the probability of experiencing a certain amount of failures before experiencing the first success in a series of Bernoulli trials."

"A Bernoulli trial is an experiment with only two possible outcomes – “success” or “failure” – and the probability of success is the same each time the experiment is conducted. An example of a Bernoulli trial is a coin flip. The coin can only land on two sides (we could call heads a “success” and tails a “failure”) and the probability of success on each flip is 0.5, assuming the coin is fair."

"If a random variable X follows a geometric distribution, then the probability of experiencing k failures before experiencing the first success can be found by the following formula:

$$P(k) = p \cdot (1-p)^k$$

where:

k : number of failures before first success

p : probability of success on each trial"

Note that a probability of an event is a number p such that $0 < p < 1$.

The percent probability is $100 \cdot p$.

The theoretical mean μ of a Geometric(p) random variable is $(1-p)/p$.

The theoretical variance of a Geometric(p) random variable is $(1-p)/p^2$.

One standard deviation σ of a Geometric(p) random variable is the square root of the variance.

Quoting

https://stats.libretexts.org/Bookshelves/Probability_Theory/

[Probability_Mathematical_Statistics_and_Stochastic_](https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/11%3A_Bernoulli_Trials/11.03%3A_The_Geometric_Distribution)

[Processes_\(Siegrist\)/11%3A_Bernoulli_Trials/11.03%3A_The_Geometric_Distribution](https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/11%3A_Bernoulli_Trials/11.03%3A_The_Geometric_Distribution)

"..... if the first success has not occurred by trial number m , then the remaining number of trials needed to achieve the first success has the same distribution as the trial number of the first success in a fresh sequence of Bernoulli trials. In short, Bernoulli trials have no memory. This fact has implications for a gambler betting on Bernoulli trials (such as in the casino games roulette or craps). No betting strategy based on observations of past outcomes of the trials can possibly help the gambler."

6.1 pdf_geometric (k, p)

The Maxima function `pdf_geometric (k, p)` returns the value at k of the probability function of a Geometric(p) random variable, with $0 < p \leq 1$.

The probability function is defined as $p \cdot (1 - p)^k$. This is interpreted as the probability of k failures before the first success.

6.2 Statology 1: Coin Tosses

This example is from:

<https://www.statology.org/geometric-distribution-real-life-examples/>

"Suppose we want to know how many times we'll have to flip a fair coin until it lands on heads. We can use the following formulas to determine the probability of experiencing 0, 1, 2, 3 failures, etc. before the coin lands on heads; the coin can experience 0 "failures" if it lands on heads on the first flip." We are defining "success" to be: the coin lands on heads. Which means "failure" is defined as tails.

For each coin flip, $p = P(\text{success})$, $(1 - p) = P(\text{failure})$, the joint probability of k failures followed by one success factors into the product of the probability of each event (whether failure or success), since the flipping of the coin event is completely independent of the outcome of any of the other flipping events.

Hence $P(k \text{ failures and 1 success}) = P(F)*P(F)* \dots *P(F)*P(S) = (1-p)^k * p$.

Let's define a formal Maxima function of this form.

```
(%i13) P(k, p) := p*(1 - p)^k$
```

For a balanced unbiased coin, $p = 0.5$.

We will compare the results produced by `pdf_geometric(k,p)` with $P(k,p)$.

```
(%i16) p:0.5$
mypoints :makelist ([k, pdf_geometric (k, p)], k, 0, 4);
makelist ( [k, P(k, p)], k, 0, 4);
(myoints) [[0,0.5],[1,0.25],[2,0.125],[3,0.0625],[4,0.03125]]
(%o16) [[0,0.5],[1,0.25],[2,0.125],[3,0.0625],[4,0.03125]]
```

Note that $p*(1 - p)^0 = p = 0.5$ for any value of p .

Since $(1-p) = 0.5 = p$ for this example, starting with $P(0) = 0.5$, we get $P(1) = 0.5*p = 0.5^2 = 0.25$, and so on.

Let k = number of failures before the first success, x = number of trials needed to finally find success
 $x = 1, 2, 3, \dots$, $k = x - 1 = 0, 1, 2, 3, \dots$

The probability of experiencing 0 failures before coin lands on heads is 50%. (H) ($x = 1, k = 0$)

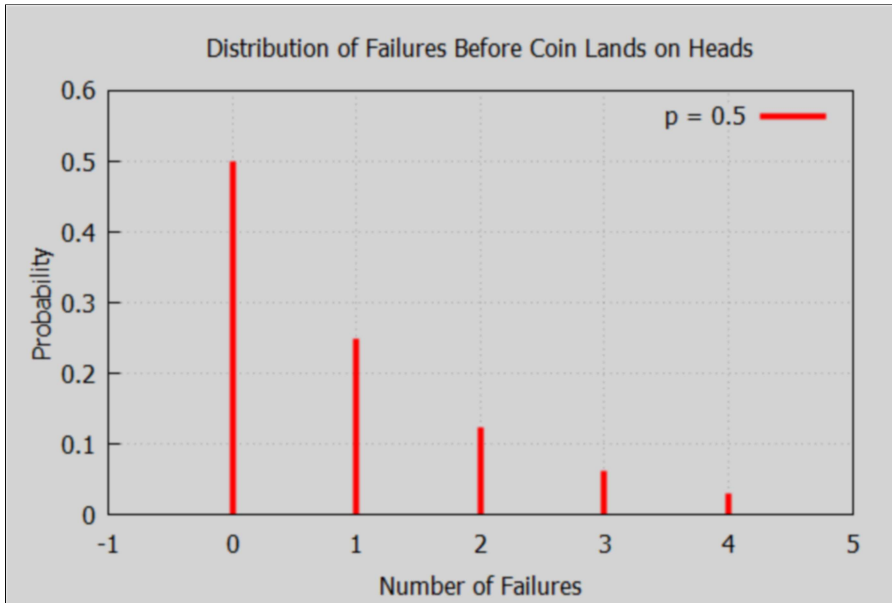
The probability of experiencing 1 failure before coin lands on heads is 25%. (T, H) ($x = 2, k = 1$)

The probability of experiencing 2 failures before coin lands on heads is 12.5%. (T,T,H) ($x = 3, k = 2$)

The probability of experiencing 3 failures before coin lands on heads is 6.25%. (T,T,T,H)

```
(%i17) wxdraw2d ( xrange = [-1, 5], yrange = [0, 0.6], points_joined = impulses,
  xlabel = "Number of Failures", ylabel = "Probability",
  title = "Distribution of Failures Before Coin Lands on Heads",
  background_color = light_gray, grid = true, key = "p = 0.5",
  line_width = 4, color = red, points (mypoints) )$
```

(%t17)



Let x be the number of independent trials required to finally get a success.
 Let k be the number of failures required to finally get success on the next trial.
 Then $x = 1, 2, 3, \dots$ and $k = 0, 1, 2, 3, \dots$ and $k = x - 1$.

What is the probability $P(k \geq 0)$? Answer: 1 (100%). $P(k \geq 0)$ should be interpreted as the probability we will need a value of k greater than or equal to 0 before getting a success, which is the "area under the curve" from $k = 0$ to $k = \text{inf}$. For a discrete distribution, the continuous distribution concept "area under the curve" becomes the sum of the discrete probability values given by `pdf_geometric(k, p)` in some range $[k1, k2]$.

The probability $x \geq 1$ is unity: $P(x \geq 1) = 1$. Note $P(x \geq 1)$ should be interpreted as the probability we will need a value of trials (x) greater than or equal to 1, but x starts at $x = 1$ and continues to arbitrarily large integral values.

Now $x \geq 1 \implies x - 1 \geq 0 \implies k \geq 0$. So $P(k \geq 0) = P(x \geq 1)$

What is the probability $P(k > 0)$? What is the probability we will need one or more failures before the next trial will result in success?

Answer: $P(k > 0) = P(k \geq 0) - P(k = 0) = 1 - 0.5 = 0.5$ (50%).

What is the probability $P(k < 2)$? What is the probability we will need less than two failures before the next trial will result in success? This is "the area under the curve" from $k = 0$ to $k = 1$, ie., the sum of the discrete probabilities $P(0)$ and $P(1)$.

Answer: $P(k < 2) = P(0) + P(1) = 0.5 + 0.25 = 0.75$ (75%).

6.3 `cdf_geometric(k, p)`

The 'cumulative distribution function' (`cdf`): `cdf_geometric(k, p)` returns the sum of the values $P(k)$ for a Geometric (p) random variable, with $0 < p < 1$, beginning at $k=0$ and ending at k .

$P(k \leq n) = \text{cdf_geometric}(n, p)$

Continuing with the fair coin toss example, we then get $P(k < 2) = P(0) + P(1) = 0.5 + 0.25 = 0.75$ from `cdf_geometric(1, 0.5)`.

```
(%i18) cdf_geometric(1, 0.5);
```

```
(%o18) 0.75
```

What is the probability $P(k \leq 2)$? Answer: $P(0) + P(1) + P(2) = 0.75 + 0.125 = 0.875$ (87.5%).

```
(%i19) cdf_geometric(2, 0.5);
```

```
(%o19) 0.875
```

What is the probability $P(k > 2)$? Answer: $1 - P(k \leq 2) = 1 - 0.875 = 0.125$ (12.5%).

```
(%i20) 1 - cdf_geometric(2, 0.5);
```

```
(%o20) 0.125
```

What is the probability $P(k \geq 2)$? Answer: $1 - P(k < 2) = 1 - 0.75 = 0.25$ (25%).

```
(%i21) 1 - cdf_geometric (1, 0.5);
```

```
(%o21) 0.25
```

In summary, we have:

```
P(k = 2): 0.12500 pdf_geometric (2, 0.5);
```

```
P(k < 2): 0.75000 cdf_geometric (1, 0.5);
```

```
P(k ≤ 2): 0.87500 cdf_geometric (2, 0.5);
```

```
P(k > 2): 0.12500 1 - cdf_geometric (2, 0.5);
```

```
P(k ≥ 2): 0.25000 1 - cdf_geometric (1, 0.5);
```

6.4 Maxima Function geomCalc (k, p)

We can incorporate the methods in the above calculation to create a Maxima function `geomCalc (k, p)`. We could, of course, have multiplied by 100, to list percent probabilities instead of probabilities.

```
(%i22) geomCalc (kk, pp) :=
block (
  if not numberp (pp) then (
    print ("pp must be a number such that 0 < pp < 1"),
    return (false) ),
  if pp <= 0 or pp >= 1 then (
    print ("Need 0 < pp < 1"),
    return (false) ),
  pp : float (pp),
  print (" "),
  print (sconcat (" For p = ", pp, " the probabilities are:")),
  print (" "),
  print (sconcat (" P ( k = ", kk, " ) : ", pdf_geometric (kk, pp) )),
  print (sconcat (" P ( k < ", kk, " ) : ", cdf_geometric (kk - 1, pp) )),
  print (sconcat (" P ( k <= ", kk, " ) : ", cdf_geometric (kk, pp) )),
  print (sconcat (" P ( k > ", kk, " ) : ", 1 - cdf_geometric (kk, pp) )),
  print (sconcat (" P ( k >= ", kk, " ) : ", 1 - cdf_geometric (kk - 1, pp) )),
  done )$
```

(%i23) `geomCalc (2, 0.5)$`

For $p = 0.5$ the probabilities are:

$$P(k = 2) : 0.125$$

$$P(k < 2) : 0.75$$

$$P(k \leq 2) : 0.875$$

$$P(k > 2) : 0.125$$

$$P(k \geq 2) : 0.25$$

(%i24) `geomCalc (1, 0.5)$`

For $p = 0.5$ the probabilities are:

$$P(k = 1) : 0.25$$

$$P(k < 1) : 0.5$$

$$P(k \leq 1) : 0.75$$

$$P(k > 1) : 0.25$$

$$P(k \geq 1) : 0.5$$

(%i25) `geomCalc (0, 0.5)$`

For $p = 0.5$ the probabilities are:

$$P(k = 0) : 0.5$$

$$P(k < 0) : 0$$

$$P(k \leq 0) : 0.5$$

$$P(k > 0) : 0.5$$

$$P(k \geq 0) : 1$$

(%i26) `geomCalc (2, x)$`

pp must be a number such that $0 < pp < 1$

(%i27) `geomCalc (2, 3)$`

Need $0 < pp < 1$

These results agree with the results returned by

<https://www.statology.org/geometric-distribution-calculator/>

Instead of using `cdf_geometric` is various ways to construct `calcGeom`, we could have used formulas directly related to p , such as $P(k \leq m) = 1 - (1 - p)^{(m+1)}$, $P(k < m) = 1 - (1 - p)^m$, $P(k \geq m) = (1 - p)^m$, $P(k > m) = (1 - p)^{(m+1)}$. Note that `cdf_geometric (m, p) = P(k <= m)`.

6.5 Statology 2: Number of Bankruptcies

This example is from:

<https://www.statology.org/geometric-distribution-real-life-examples/>

"Suppose it's known that 4% of individuals who visit a certain bank are visiting to file bankruptcy.

Suppose a banker wants to know the probability that he will meet with 10 people before encountering someone who is filing for bankruptcy. We are defining "success" to be: the banker meets someone wishing to file bankruptcy. We are defining "failure" to be: the banker meets someone not wishing to file for bankruptcy. So in this possible scenario, the 10th person the banker meets wants to file for bankruptcy, meaning the preceding 9 persons are "failures". What is the probability of that happening?

```
(%i28) cdf_geometric (9, 0.04), numer;
```

```
(%o28) 0.335167
```

Thus there is a 33.5% probability the banker will meet with 10 people before encountering someone wishing to file for bankruptcy.

```
(%i29) geomCalc (9, 0.04)$
```

For $p = 0.04$ the probabilities are:

$$P(k = 9) : 0.0277014$$

$$P(k < 9) : 0.307466$$

$$P(k \leq 9) : 0.335167$$

$$P(k > 9) : 0.664833$$

$$P(k \geq 9) : 0.692534$$

6.6 PSU Example 11.1

We adapt a problem from:

<https://online.stat.psu.edu/stat414/lesson/11/11.1>

A reporter is looking on the streets of Kansas City, selecting people until he finds the first person who attended the most recent Kansas City Chiefs home game. Suppose the probability of selecting such a person is $p = 0.20 = P(\text{success})$ for each person met. Then the probability of failure for each person met is $(1 - p) = P(\text{failure}) = 0.80$.

Let x = number selected (with the last one success), and k = number of failures before success.

$x = 1, 2, 3, \dots$, and $k = 0, 1, 2, \dots$

So $k = x - 1$. The condition $x > 6 \implies x - 1 > 6 - 1 \implies k > 5$, etc.

A.) What is the probability he must select 4 people before finding such a person. Selecting 4 people one after the other before finding success means the first three people selected represent failures. $x = 4 \implies k = 3$. (F,F,F,S)

```
(%i30) pdf_geometric (3, 0.20), numer;
```

```
(%o30) 0.1024
```

Hence about a 10% chance the reporter must select 4 people before finding one who attended the last home game.

B.) What is the probability that the reporter must select more than 6 people before he finds one who attended the last home football game?

Ans: $P(x > 6) = 1 - P(x \leq 6) = 1 - P(k \leq 5) = 1 - \text{cdf_geometric}(5, 0.2)$

```
(%i31) 1 - cdf_geometric (5, 0.2), numer;
```

```
(%o31) 0.262144
```

There is about a 26% chance that the reporter would have to select more than 6 people before finding one who attended the last home game.

```
(%i32) geomCalc (5, 0.2)$
```

For $p = 0.2$ the probabilities are:

$P(k = 5) : 0.065536$

$P(k < 5) : 0.67232$

$P(k \leq 5) : 0.737856$

$P(k > 5) : 0.262144$

$P(k \geq 5) : 0.32768$

6.7 mean_geometric (p)

The Maxima function `mean_geometric (p)` returns the mean of a Geometric(p) random variable, with $0 < p \leq 1$. The theoretical mean of Geometric(p) [with k = number of failures before success, $k = 0, 1, 2, \dots$] is given by $E(k) = (1 - p)/p$, which is returned by Maxima's `mean_geometric` function.

The theoretical mean number of **trials** x required (the last being success) is $E(x) = 1/p$.

Calculating $E(k)$ by hand as $(1 - p)/p$, with $p = 0.2$, $E(k) = 4$, $E(x) = 1/p = 5$.

```
(%i33) 0.8/0.2;
```

```
(%o33) 4.0
```

Using Maxima's `mean_geometric` function:

```
(%i34) mean_geometric (0.2);
```

```
(%o34) 4.0
```

The mean value of a geometric distribution with $p = 0.2$ is $k = 4$ failures, or $x = 5$ independent trials. We should expect the reporter to have to select 5 people before he finds one who attended the last home game. Of course, on any given try, it may take 1 person or it may take 10, but 5 is the average number.

6.8 var_geometric (p)

The Maxima function `var_geometric (p)` returns the variance of a Geometric(p) random variable, with $0 < p \leq 1$. The theoretical variance of a Geometric (p) random variable [with k = number of failures before success, $k = 0, 1, 2, \dots$] is $\text{Var} (k) = (1 - p)/p^2$. This is also $\text{Var} (x)$.

Calculating the variance in the number of failures for $p = 0.2$ by hand:

```
(%i35) 0.8/ (0.2)^2;
```

```
(%o35) 20.0
```

```
(%i36) var_geometric (0.2);
```

```
(%o36) 20.0
```

The square root of the variance $V(k)$ is one standard deviation about the mean.

```
(%i37) sqrt (%);
```

```
(%o37) 4.47214
```

6.9 std_geometric (p)

The Maxima function `std_geometric (p)` returns the standard deviation of a Geometric(p) random variable, with $0 < p \leq 1$. This is the square root of the variance, or $\sqrt{(1-p)/p}$.

```
(%i38) sqrt(0.8)/0.2;
```

```
(%o38) 4.47214
```

```
(%i39) std_geometric (0.2);
```

```
(%o39) 4.47214
```

6.10 Statology 3: Supporters of a Law

This example is from:

<https://www.statology.org/geometric-distribution-real-life-examples/>

"Suppose a researcher is waiting outside of a library to ask people if they support a certain law. The probability that a given person supports the law is $p = 0.2$."

"What is "the probability of interviewing 0, 1, 2 people, etc. before the researcher speaks with someone who supports the law?" We are defining success as finding a person who supports the particular law.

x = number of people interviewed. $P(x = 1) = P(k = 0)$

Let k be the number of failures before finding a person who supports the law when the chance that any given person interviewed supports the law is 20%.

The probability of 0 failures before finding a person who supports the law is then

```
(%i40) pdf_geometric (0,0.2);
```

```
(%o40) 0.2
```

$P(k = 0) = P(x = 1) = 0.2$ (20% chance).

$P(x = 2) = P(k = 1)$.

```
(%i41) pdf_geometric (1, 0.2);
```

```
(%o41) 0.16
```

and so on.

```
(%i42) mypoints : makelist ([k, pdf_geometric (k, 0.2)], k, 0, 4);
```

```
(mypoints) [[0,0.2],[1,0.16],[2,0.128],[3,0.1024],[4,0.08192]]
```

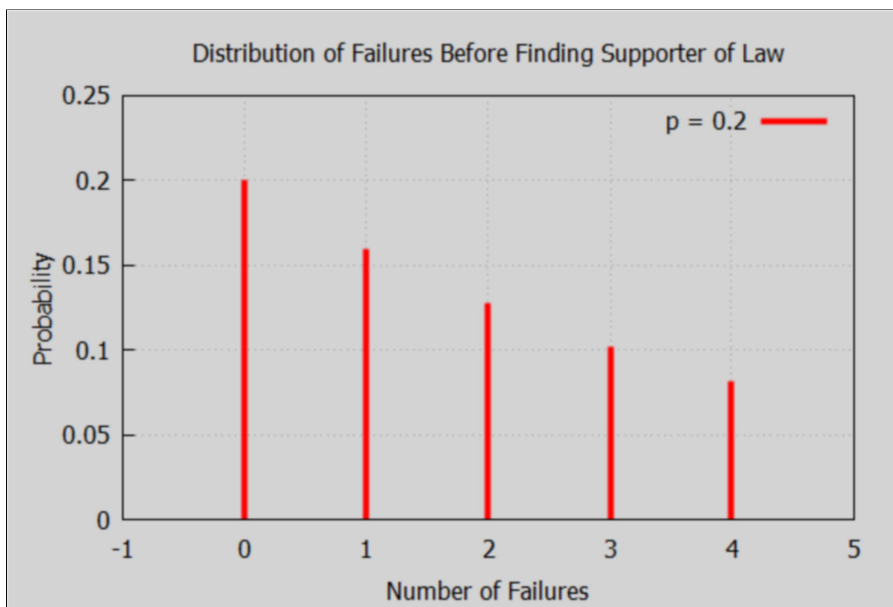
The probability that the first person (0 failures) the researcher speaks to supports the law is 20% (1 in 5).

The probability that the researcher must interview 2 people (one failure) before finding someone who supports the law is 16%.

The probability that the researcher must interview 3 people (two failures) before finding someone who supports the law is 12.8%.

```
(%i43) wxdraw2d ( xrange = [-1, 5], yrange = [0, 0.25], points_joined = impulses,
  xlabel = "Number of Failures", ylabel = "Probability",
  title = "Distribution of Failures Before Finding Supporter of Law",
  key = "p = 0.2",
  background_color = light_gray, grid = true,
  line_width = 4, color = red, points (mypoints) )$
```

(%t43)



We can use our Maxima function `geomCalc (k, p)` defined in Example 1:

6.11 Statology 4: Chances of Network Failures

This example is from:

<https://www.statology.org/geometric-distribution-real-life-examples/>

"Suppose it's known that the chance that a certain company experiences a network failure in a given week is 10%. Suppose the CEO of the company would like to know the chance that the company can go 5 weeks or longer without experiencing a network failure."

$p = P(\text{success}) = 0.1 = \text{probability that a week will go by with a network failure.}$

$(1-p) = P(\text{failure}) = 0.9 = \text{probability that a week will go by without a network failure.}$

We need to calculate $P(k \geq 5)$.


```
(%i44) geomCalc (5, 0.1)$
```

For $p = 0.1$ the probabilities are:

$$P(k = 5) : 0.059049$$

$$P(k < 5) : 0.40951$$

$$P(k \leq 5) : 0.468559$$

$$P(k > 5) : 0.531441$$

$$P(k \geq 5) : 0.59049$$

Thus the chance the company can go five weeks or longer WITHOUT experiencing a network failure is 59%.

6.12 Calcwork Ex. 1

We summarize a problem from:

<https://calcworkshop.com/discrete-probability-distribution/geometric-distribution/>

"Suppose Max owns a lightbulb manufacturing company and determines that 3 out of every 75 bulbs are defective. What is the probability that Max will find the first faulty lightbulb on the 6th one that he tested?"

If $P(\text{success}) = P(\text{finding defective lightbulb}) = p = 3/75 = 0.04$, the $P(\text{failure}) = (1 - p) = 0.96$.
 $P(x = 6) = P(k = 5) = \text{pdf_geometric}(5, 0.04)$.

```
(%i45) pdf_geometric (5, 0.04), numer;
```

```
(%o45) 0.0326149
```

About a 3% chance of finding the first faulty lightbulb on the 6th one tested.

"Now, what if Max wants to know the likelihood that it takes at least six trials until he finds the first defective lightbulb?"

$P(x \geq 6) = P(k \geq 5) = 1 - \text{cdf_geometric}(4, 0.04)$.

```
(%i46) 1 - cdf_geometric (4, 0.04);
```

```
(%o46) 0.815373
```

```
(%i47) geomCalc (5, 0.04)$
```

For $p = 0.04$ the probabilities are:

$P(k = 5) : 0.0326149$

$P(k < 5) : 0.184627$

$P(k \leq 5) : 0.217242$

$P(k > 5) : 0.782758$

$P(k \geq 5) : 0.815373$

"... there is an 0.815 chance of Max needing at least six trials until he finds the first defective lightbulb."

"And using this same example, let's determine the number lightbulbs we would expect Max to inspect until he finds his first defective, as well as the standard deviation."

Expectation value for the number of lightbulbs one must test before finding a faulty one is $E(x) = 1/p$
 Expectation value for the number of failures before the next lightbulb tested is faulty is
 $E(k) = (1 - p)/p$

```
(%i49) 1/0.04;  
0.96/0.04;
```

```
(%o48) 25.0
```

```
(%o49) 24.0
```

```
(%i50) mean_geometric (0.04);
```

```
(%o50) 24.0
```

```
(%i51) std_geometric (0.04);
```

```
(%o51) 24.4949
```

"This shows us that we would expect Max to inspect 25 lightbulbs before finding his first defective, with a standard error of 24.49."

6.13 Stat-howto Ex. 1

A reporter is looking for people who voted in the last election and voted as an independent. The probability of finding such a voter is $P(\text{success}) = p = 0.2$. What is the probability you would meet an independent voter on your third try ($x = 3, k = 2$)

```
(%i52) pdf_geometric (2, 0.2), numer;
```

```
(%o52) 0.128
```

6.14 Inverse type of problem.

We summarize Ex. 3 from

https://minisham.redbrick.dcu.ie/CA/Notes/CA266/10_Geometric_Distribution.pdf

"It is known that 20% of products on a production line are defective. Products are inspected until first defective is encountered. Let x = number of inspections to obtain first defective.

What is the minimum number of inspections that would be necessary so that the probability of observing a defective is more than 75%?

Choose n so that $P(x \leq n) \geq .75$

6.14.1 quantile_geometric (q, p)

`quantile_geometric (q, p)` returns the q -quantile (an integer) of a Geometric (p) random variable, with $0 < q < 1$, $0 < p < 1$.

This is the inverse of `cdf_geometric (k, p)`.

The probability from which the quantile is derived is defined as $p \cdot (1 - p)^k$, interpreted as the probability of k failures before the first success.

`quantile_geometric (q, p)` is not strictly the inverse of `cdf_geometric (k, p)`, since `quantile_geometric (q, p)` is only defined for values of q which belong to the set defined by `cdf_geometric (k, p)` with $k = 0, 1, \dots$ (non-negative integers). In the following we construct that set and display floating point values of qq as the second element of `[k, cdf_geometric (k, 0.2), quantile_geometric (q, 0.2)]`.

Returning to the problem framed above: Choose n so that $P(x \leq n) \geq .75$

```
(%i53) quantile_geometric (0.75, 0.2);
```

```
(%o53) 6
```

```
(%i54) for k:4 thru 7 do (
    qq : float (cdf_geometric (k, 0.2)),
    print ([k, qq, quantile_geometric (qq, 0.2)] ) )$
[4,0.67232,4]
[5,0.737856,5]
[6,0.790285,6]
[7,0.832228,7]
```

We see that the minimum number of inspections such that the probability of finding a defective product is equal to or greater than 0.75 is 6 (which value has 0.79 rather than 0.75).

Continuing with the example, What is the average number of inspections to obtain the first defective?

The average number of trials (inspections) $E(x) = 1/p$ and the average number of failures before finding a defective is $E(k) = (1-p)/p$

```
(%i56) 1/0.2;
        0.8/0.2;
```

```
(%o55) 5.0
```

```
(%o56) 4.0
```

```
(%i57) mean_geometric (0.2);
```

```
(%o57) 4.0
```

The average number of inspections before finding a defective is 5. ($x = 5$, $k = 4$ failures).

6.15 Kidney Donor Scenario

A problem from

<https://www.studysmarter.co.uk/explanations/math/statistics/geometric-distribution/>

"A patient suffers kidney failure and requires a transplant from a suitable donor. The probability that a random donor will match this patient's requirements is 0.2.

1.) Suppose that no donor matches the patient's requirements until a fifth donor comes in. What is the probability of this scenario?

$P(\text{success}) = p = 0.2 =$ probability a random donor will match patient's need. $P(x = 5) = P(k = 4)$.

```
(%i58) pdf_geometric (4, 0.2), numer;
```

```
(%o58) 0.08192
```

The probability the fifth donor will provide the needed kidney is about 8.2%

2.) Find the probability of the patient requiring 10 or fewer donors until a match is found.

$P(x \leq 10) = P(k \leq 9)$

(%i59) geomCalc (9, 0.2)\$

For $p = 0.2$ the probabilities are:

$$P(k = 9) : 0.0268435$$

$$P(k < 9) : 0.865782$$

$$P(k \leq 9) : 0.892626$$

$$P(k > 9) : 0.107374$$

$$P(k \geq 9) : 0.134218$$

About a 90% chance a match will be found after 10 donors.

3.) What is the expected number of donors required to get a match?

The expected number of donors $E(x) = 1/p = 1/0.2 = 5$

4.) Find the standard deviation of this scenario.

(%i60) std_geometric (0.2);

(%o60) 4.47214

6.16 Rolling a Fair Dice

A problem from

<https://www.studysmarter.co.uk/explanations/math/statistics/geometric-distribution/>

"Suppose you roll a fair dice until you get a three as a result."

The probability of rolling a 3 is $P(\text{success}) = p = 1/6 = 0.1667$, $P(\text{failure}) = 1-p = 5/6 = 0.8333$.

1.) "What is the probability that you don't roll a three until your fourth roll?"

$$P(x=4) = P(k = 3)$$

(%i61) pdf_geometric (3, 1/6), numer;

(%o61) 0.0964506

9.6% ~ 10% chance won't get a three until the fourth try.

2.) "Find the probability of getting the three you need in less than 10 rolls."

$$P(x < 10) = P(k < 9).$$

```
(%i62) geomCalc (9, 1/6);
```

For $p = 0.166667$ the probabilities are:

$$P(k = 9) : 0.0323011$$

$$P(k < 9) : 0.806193$$

$$P(k \leq 9) : 0.838494$$

$$P(k > 9) : 0.161506$$

$$P(k \geq 9) : 0.193807$$

```
(%o62) done
```

About 81% chance you will get the three you need in less than 10 rolls.

3.) "What is the expected number of rolls required to get your desired outcome?"

$$E(x) = 1/p = 1/(1/6) = 6 \text{ rolls.}$$

4.) "Find the variance of this experiment."

```
(%i63) var_geometric (1/6), numer;
```

```
(%o63) 30.0
```

```
(%i64) sqrt(%);
```

```
(%o64) 5.47723
```

6.17 Winning a Prize From a Claw Machine

A problem from

<https://www.studysmarter.co.uk/explanations/math/statistics/geometric-distribution/>

"Suppose that the probability of winning an item from a claw machine is 0.05."

$p = 0.05 = P(\text{success})$ on each attempt to extract a prize using a claw machine.

1.) "What is the probability of winning an item on your first try?"

$$P(x = 1) = P(k = 0).$$

```
(%i65) pdf_geometric (0, 0.05), numer;
```

```
(%o65) 0.05
```

2.) What is the probability of winning an item in less than 20 tries?"

$$P(x < 20) = P(k < 19).$$

```
(%i66) geomCalc (19, 0.05);
```

For $p = 0.05$ the probabilities are:

$$P(k = 19) : 0.0188677$$

$$P(k < 19) : 0.622646$$

$$P(k \leq 19) : 0.641514$$

$$P(k > 19) : 0.358486$$

$$P(k \geq 19) : 0.377354$$

```
(%o66) done
```

About a 62% chance you will win a prize in less than 20 tries.

3.) Suppose you need to use a quarter for each try. What is the expected amount of money spent for getting a prize?"

The average number of tries before winning a prize is $E(x) = 1/p = 1/0.05 = 20$.
 $(20 \text{ tries}) \cdot (0.25\$/\text{try}) = 20 \cdot (1/4) \$ = 5\$$. Expect the pay about \$5 per prize won.

6.18 random_geometric (p), random_geometric (p, n)

The Maxima function `random_geometric (p)` returns the number of failures (k) in one Bernoulli trial in which the probability of "success" is p .

The Maxima function `random_geometric (p,n)` returns a list of n calls to `random_geometric (p)`.

The probability from which the random sample is derived is defined as $p(1-p)^k$. This is interpreted as the probability of k failures before the first success.

Suppose $p = 0.2$; the probability of success in each trial is 20%.

```
(%i67) random_geometric (0.2);
```

```
(%o67) 5
```

```
(%i68) for j thru 5 do print (random_geometric (0.2))$
4
1
1
0
5
```

To obtain the number of failures before success in each of 10 Bernoulli trials, given that the chance of success in each such trial is 20%, we would use `random_geometric (0.2, 10)`.

```
(%i69) for j thru 5 do print (random_geometric (0.2, 10))$
[6,3,5,3,0,4,5,0,2,3]
[23,6,5,0,3,4,8,6,1,2]
[0,5,3,2,5,4,4,0,7,0]
[1,14,0,4,2,0,14,2,5,1]
[4,10,1,2,11,1,10,15,2,4]
```

6.18.1 100 Simulated Bernoulli Trials, $p = 0.2$

Let `data` be a list of the number of failures before success in 100 simulated Bernoulli trials, assuming $p = 0.2$.

```
(%i73) data : random_geometric (0.2, 100)$
fill (data);
head (data);
tail (data);
(%o71) [13,0,100]
(%o72) [13,1,0]
(%o73) [0,5,0]
```

6.18.2 discrete_freq (Data)

The Maxima function `discrete_freq(aList)` counts the number of unique discrete "readings" of some instrument recorded in the list `aList` and returns a new list [list-of-unique-readings, list-of-frequency-of-each-unique-reading]. The elements of the list 'frequencies' corresponds to the elements of the list 'uniqueData', element by element.

```
(%i76) [uniqueData, frequencies] : discrete_freq (data)$
uniqueData;
frequencies;
(%o75) [0,1,2,3,4,5,6,7,8,9,10,12,13,14,15,16,19]
(%o76) [21,21,10,8,9,7,3,2,4,3,1,2,3,1,2,2,1]
```



```
(%i77) length (uniqueData);
```

```
(%o77) 17
```

Out of a list of 100 semi-random integers, there is a lot of overlap, resulting in only a few unique integers.

```
(%i78) nfrequencies : frequencies/Lsum (frequencies), numer;
```

```
(nfrequencies) [0.21,0.21,0.1,0.08,0.09,0.07,0.03,0.02,0.04,0.03,0.01,0.02,0.03  
,0.01,0.02,0.02,0.01]
```

```
(%i79) Lsum (frequencies);
```

```
(%o79) 100
```

```
(%i80) Lsum (nfrequencies);
```

```
(%o80) 1.0
```

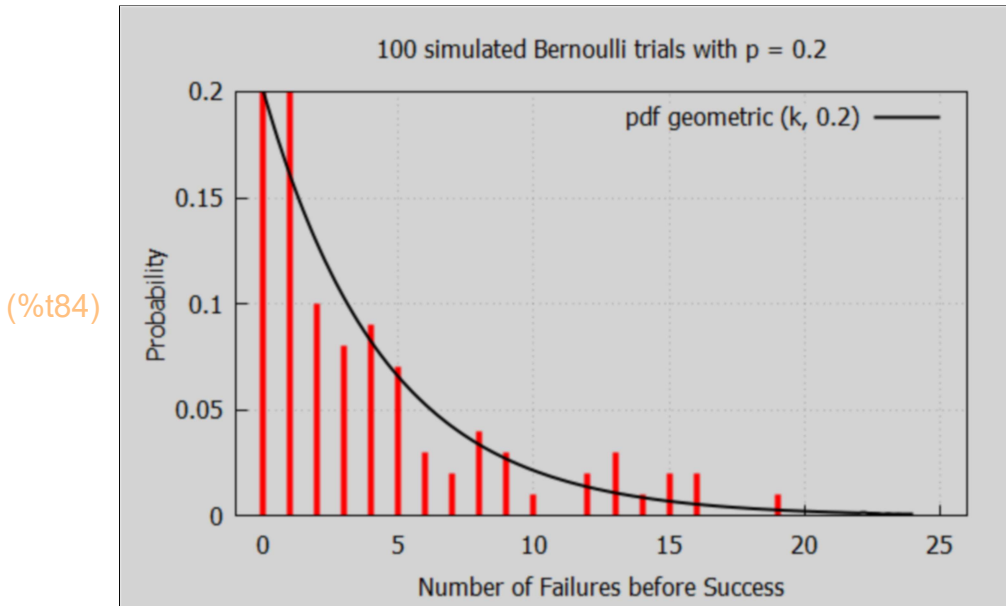
```
(%i81) lmax (nfrequencies);
```

```
(%o81) 0.21
```

```
(%i83) mypoints : makelist ([uniqueData[j], nfrequencies[j]], j, 1, length (uniqueData))$  
fill (mypoints);
```

```
(%o83) [[0,0.21],[19,0.01],17]
```

```
(%i84) wxdraw2d ( xrange = [-1, 26], yrange = [0, 0.2], points_joined = impulses,
  xlabel = "Number of Failures before Success", ylabel = "Probability",
  title = "100 simulated Bernoulli trials with p = 0.2",
  background_color = light_gray, grid = true,
  line_width = 4, color = red, points (mypoints),
  color = black, line_width = 2,
  key = "pdf geometric (k, 0.2)",
  explicit (pdf_geometric (kk, 0.2), kk, 0, 24) )$
```



6.18.3 1000 Simulated Bernoulli Trials, $p = 0.2$

We should get results closer to the theoretical prediction if we simulate 1000 Bernoulli trials instead of 100. Let data be a list of the number of failures before success in 1000 simulated Bernoulli trials, assuming $p = 0.2$.

```
(%i86) data : random_geometric (0.2, 1000)$
fill (data);
```

```
(%o86) [4,2,1000]
```

```
(%i89) [uniqueData, frequencies] : discrete_freq (data)$
uniqueData;
frequencies;
```

```
(%o88) [0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,
24,27]
```

```
(%o89) [181,169,128,108,87,45,53,54,45,26,27,14,24,6,8,7,4,3,1,2,
3,2,2,1]
```

```
(%i90) nfreqencies : frequencies/Lsum (frequencies), numer;
```

```
(nfrequencies) [0.181,0.169,0.128,0.108,0.087,0.045,0.053,0.054,0.045,0.026,  
0.027,0.014,0.024,0.006,0.008,0.007,0.004,0.003,0.001,0.002,0.003,  
0.002,0.002,0.001]
```

```
(%i91) Lsum (nfrequencies);
```

```
(%o91) 1.0
```

```
(%i92) lmax (nfrequencies);
```

```
(%o92) 0.181
```

```
(%i94) mypoints : makelist ([uniqueData[j], nfreqencies[j]], j, 1, length (uniqueData))$  
fill (mypoints);
```

```
(%o94) [[0,0.181],[27,0.001],24]
```

```
(%i95) wxdraw2d ( xrange = [-1, 17], yrange = [0, 0.2], points_joined = impulses,  
xlabel = "Number of Failures before Success", ylabel = "Probability",  
title = "random geometric (0.2, 1000)",  
background_color = light_gray, grid = true,  
line_width = 4, color = red, points (mypoints),  
color = black, line_width = 2,  
key = "pdf geometric (k, 0.2)",  
explicit (pdf_geometric (kk, 0.2), kk, 0, 17) )$
```

```
(%t95)
```

