

Stat02-Binomial.wmx

TABLE OF CONTENTS

Preface	1
References	2
Discrete Distributions Defined	3
What Is a Discrete Random Variable?	4
Mean and Variance of Discrete Data Set	5
Fundamental Probability Laws	5
Permutations and Combinations	7
binomial (n, k)	9
The Binomial (n, p) Distribution	9
pdf_binomial (k, n, p), $n > 0$, $0 < p < 1$, $k = 0, 1, 2, \dots, n$	11
Two Tosses of a Balanced Coin	13
mean_binomial (n, p)	14
var_binomial (n, p)	14
std_binomial (n, p)	15
cdf_binomial (k, n, p)	15
Example 1	16
Example 2	18
quantile_binomial (q, n, p)	18
100 Flips of a Balanced Coin	18
Inverse Type of Problem	22
binomialCalc (k, n, p)	22
confidence (q, m, s)	23
Number of Side Effects from Medications	25
Number of Fraudulent Credit Card Transactions	28
Number of Spam Emails per Day	32
Number of River Overflows	32
Probability of Intercepting All Incoming Missiles	33
Number of Missiles Launched for Even Chance for Defense Penetration	35
random_binomial (n, p), random_binomial (n, p, m)	39
Random sample size $m = 100$ simulations, $n = 100$, $p = 0.5$	40
discrete_freq (data)	41
Random sample size $m = 1000$, $n = 100$, $p = 0.5$	42

1 Preface

In Stat02-Binomial.wmx we discuss the discrete Binomial (n, p) probability distribution and its application, using Maxima tools and methods. This is the second worksheet in my Statistics with Maxima section.

Edwin L. (Ted) Woollett
<https://home.csulb.edu/~woollett/>
April 9, 2024

2 *References*

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

Roger Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, John Wiley, 1989,

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.

References specific to the Binomial distribution:

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

<https://online.stat.psu.edu/stat414/lesson/10>

```
(%i5) load (descriptive);
      load (distrib);
      fpprintprec : 6$
      ratprint : false$
      logexpand : all$

(%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
```

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

```
(%i13) 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 Fundamental Probability Laws

We follow Matthews and Walker, *Mathematical Methods of Physics*, 2nd. Ed (1970) Ch. 14

"Let $P(A)$ be the probability of something (called A) occurring when an experiment is performed. $0 \leq P(A) \leq 1$. If A is certain to happen, then $P(A) = 1$. If A will certainly not happen, then $P(A)=0$."

"To illustrate some more complicated possibilities consider an experiment with n equally likely outcomes, involving two events, A and B . Let

n_1 = number of outcomes in which A occurs, but not B

n_2 = number of outcomes in which B occurs, but not A

n_3 = number of outcomes in which both A and B occur

n_4 = number of outcomes in which neither A nor B occurs.

Since we have exhausted all possibilities,

$$n_1 + n_2 + n_3 + n_4 = n .$$

The probabilities of A and of B are

$$P(A) = (n_1 + n_3)/n, \quad P(B) = (n_2 + n_3)/n.$$

We can define more complicated probabilities.

The probability of either A or B (or both) occurring is

$$P(A + B) = (n_1 + n_2 + n_3)/n.$$

The probability of both A and B occurring is called the joint probability of A and B :

$$P(A, B) = n_3/n.$$

Finally, we may define "conditional probabilities"; the probability that A occurs, given that B occurs, is

$$P(A | B) = n_3/(n_2 + n_3).$$

Similarly, $P(B | A) = n_3 / (n_1 + n_3)$.

Two important rules may be extracted from this simple example, and in fact are easily seen to be true in general:

$$P(A + B) = P(A) + P(B) - P(A, B), \tag{1}$$

$$\text{since } (n_1+n_3)/n + (n_2+n_3)/n - n_3/n = (n_1 + n_2 + n_3)/n.$$

Likewise we have the law, the probability of either A or B or C or any two, or all three occurring is

$$P(A + B + C) = P(A) + P(B) + P(C) - P(A, B, C).$$

If events A, B, C are mutually exclusive events, then the joint probability $P(A, B, C) = 0$, and we are left with $P(A + B + C) = P(A) + P(B) + P(C)$.

$$P(A, B) = P(B) P(A | B) = P(A) P(B | A), \tag{2}$$

$$\text{since } n_3/n = [(n_2+n_3)/n] * [n_3/(n_2+n_3)] = [(n_1+n_3)/n] * [n_3/(n_1+n_3)]."$$

Consider the probability of drawing an ace from a deck of 52 cards. Since a deck contains 4 aces, $P(1 \text{ ace}) = 4/52 = 1/13$. If event A is drawing an ace from deck 1 (of 2 decks) and event B is drawing an ace from deck 2, the probability that when one card is drawn from each of two decks, at least one will be an ace is $P(A + B) = P(A) + P(B) - P(A,B) = 1/13 + 1/13 - (1/13)^2 = (2 \cdot 13 - 1)/13^2 = 25/169$. In calculating $P(A, B) = P(A) \cdot P(B)$ we used the assumption that the result of drawing a card from deck 1 has no influence on the result when drawing a card from deck 2 - the two drawings are "independent".

Recall the probability of both A and B occurring is called the joint probability of A and B, $P(A,B)$. In general, if events A and B are statistically independent, $P(A, B) = P(A) \cdot P(B)$. (3)

If events A and B are mutually exclusive (which means $P(A, B) = 0$) then $P(A + B) = P(A) + P(B)$ as one can see from (1) above.

An example of $P(A, B) = P(B) \cdot P(A | B) = P(A) \cdot P(B | A)$ is the probability of drawing two hearts, when two cards are drawn successively from the same deck of cards without replacement: You have 13 hearts in a 52 card deck to start, so multiply the probabilities of each draw: $P(2 \text{ hearts}) = P(\text{draw 1 heart, draw 2 heart}) = P(1 \text{ heart}) \cdot P(1 \text{ heart} | 1 \text{ heart}) = (13/52) \cdot (12/51) = 1/17$

To find the probability of drawing a heart on the first draw and a club on the second draw from the same deck of cards, you follow a similar process. The probability of drawing a heart on the first draw is $13/52$, and the probability of drawing a club on the second draw is $13/51$ since there are 13 clubs remaining in the deck after removing one heart. The probability of drawing a heart and then a club from the same deck is $(13/52) \times (13/51) = 169/2652$.

$$(\%i14) \quad (13/52) \cdot (13/51);$$

$$(\%o14) \quad \frac{13}{204}$$

7 *Permutations and Combinations*

7.1 Permutations

Quoting: <https://en.wikipedia.org/wiki/Permutation>

"In mathematics, a permutation of a set is, loosely speaking, an arrangement of its members into a sequence or linear order, or if the set is already ordered, a rearrangement of its elements. The word "permutation" also refers to the act or process of changing the linear order of an ordered set."

"For example, there are six permutations (orderings) of the set {1, 2, 3}: written as tuples, they are (1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), and (3, 2, 1). Anagrams of a word whose letters are all different are also permutations: the letters are already ordered in the original word, and the anagram reorders them. The study of permutations of finite sets is an important topic in combinatorics and group theory."

"The number of permutations of n distinct objects is n factorial, usually written as $n!$, which means the product of all positive integers less than or equal to n ."

"The number of arrangements or 'permutations', of n objects is $n!$, since the first position can be occupied by any one of the n objects, the second by any one of the $(n - 1)$ remaining objects, and so on."

"In elementary combinatorics, the k -permutations, or partial permutations, are the ordered arrangements of k distinct elements selected from a set. When k is equal to the size of the set, these are the permutations in the previous sense."

"In older literature and elementary textbooks, a k -permutation of n means an ordered arrangement (list) of a k -element subset of an n -set. The number of such k -permutations (k -arrangements) of n is denoted variously..." . We will use $P(n, k)$ for that number.

$$P(n, k) = \text{product of } k \text{ factors} = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n - (k - 1)) \\ = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n - k + 1) = n! / (n-k)!$$

since the first element of the sequence (permutation) can be chosen from any of n elements of the n -set, and then the second element of the sequence can be chosen from any of the $n-1$ remaining elements of the n -set, etc.

For example consider the example of a 3-set $S = \{1, 2, 3\}$, and form all 2-permutations: (1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2).

The first element of each 2-permutation can be chosen in $n = 3$ ways, the second element in $n-1 = 2$ ways. So the total number of 2-permutations is $n \cdot (n-1) = 3 \cdot 2 = 6$. With $k = 2$, $n - k + 1 = 3 - 2 + 1 = 2$. So we can write (using $(n - k)! = (3 - 2)! = 1! = 1$)

$$P(3, 2) = n \cdot (n-1) \cdot \dots \cdot (n - k + 1) = 3 \cdot 2 = 6 = n! / (n-k)! = 3! / 1! = (1 \cdot 2 \cdot 3) / 1 = 6.$$

7.2 Combinations

Quoting: <https://en.wikipedia.org/wiki/Combination>

"In mathematics, a combination is a selection of items from a set that has distinct members, such that the order of selection does not matter (unlike permutations). For example, given three fruits, say an apple, an orange and a pear, there are three combinations of two that can be drawn from this set: an apple and a pear; an apple and an orange; or a pear and an orange. More formally, a k -combination of a set S is a subset of k distinct elements of S . [This means none of the members of a k -combination can be identical - 'no repetitions'] So, two combinations are identical if and only if each combination has the same members. [The arrangement of the members in each set does not matter.] If the set has n elements, the number of k -combinations, denoted by $C(n,k)$, is equal to the binomial coefficient."

Given the set of 3 objects $\{1, 2, 3\}$, the number of 2-combinations of this set are:
 $(1,2), (1, 3), (2, 3)$, so $C(3, 2) = 3$.

7.2.1 binomial (n, k)

The Maxima function binomial (n, k) can be used here:

```
(%i15) binomial (3, 2);
```

```
(%o15) 3
```

```
(%i16) 3! / (2! * (3 - 2)! );
```

```
(%o16) 3
```

In general each k -combination can be changed by changing the arrangement (order) of the k distinct elements, and there are $k!$ ways of ordering those k elements. Hence the number $P(n,k)$ of k -permutations of n distinct elements equals the number of (unordered) k -combinations of n elements times the number of ways of ordering the k elements:

$$P(n, k) = C(n, k) * k!$$

Solving for $C(n, k)$ we get:

$$C(n, k) = P(n, k) / k!, \text{ or}$$

$$C(n, k) = n! / (n-k)! / k! == n! / (k! * (n - k)!) = \text{binomial}(n, k)$$

8 The Binomial (n, p) Distribution

8.1 Pb (k; n, p)

The binomial distribution is used to find the probability of k 'successes' $P_b(k; n, p)$ in n trials of the same experiment when (1) there are only two possible and mutually exclusive outcomes, (2) the n trials are independent, and (3) the probability of 'success', p , remains constant in each trial. Then

$$P_b(k; n, p) = \left(\frac{n!}{k! \cdot (n - k)!} \right) \cdot p^k \cdot (1 - p)^{(n - k)}.$$

where $n!$ (read 'n factorial') = $n \cdot (n-1) \cdot (n-2) \dots \cdot 3 \cdot 2 \cdot 1$, and $0! = 1$ by definition. This assumes $0 < p < 1$, n is a positive integer, and $k = 1, 2, 3, \dots, n$.

The [theoretical] mean of the binomial distribution is

$$\mu = n \cdot p.$$

The [theoretical] variance of the binomial distribution is $n \cdot p \cdot (1 - p) = \mu \cdot (1 - p)$.

The [theoretical] standard deviation is $\sigma = \sqrt{n \cdot p \cdot (1 - p)} = \sqrt{\mu \cdot (1 - p)}$.

If $p < 0.5$ the distribution is skewed to the right. If $p > 0.5$ the distribution is skewed to the left.

If $p = 0.5$ the distribution is symmetrical.

8.1.1 Origin of $P_b(k; n, p)$ Formula

Adapting the explanation in

https://en.wikipedia.org/wiki/Binomial_distribution,

The formula for $P_b(k; n, p)$ can be understood as follows:

$p^k * (1-p)^{(n-k)}$ is the probability of obtaining the sequence of Bernoulli trials in which the first k trials are "successes" and the remaining (last) $(n - k)$ trials result in "failure". Since the trials are independent with probabilities remaining constant between them, any arrangement of n trials with k successes (and $(n - k)$ failures) has the same probability of being achieved (regardless of positions of successes within the arrangement).

How many distinguishable arrangements of n objects if $n = n_1 + n_2$, where $n_1 = k$ of the objects are identical (type 'success') and the other $n_2 = (n - k)$ objects (type 'failure') are also identical (but different from the n_1 type)? The total number of permutations of n objects is $n!$ but each *distinguishable* permutation appears $n_1! * n_2!$ times, so the number of *distinguishable* permutations is $n!/(n_1!*n_2!) = n! / (k! * (n-k)!) = \text{binomial}(n, k)$.

Here is a simple example: suppose $n = 3$, $n! = 6$, $n = n_1 + n_2$, $n_1 = 1$ success, $n_1! = 1$, $n_2 = 2$, $n_2! = 2$ failures, number of "distinguishable" permutations is $n!/(n_1!*n_2!) = 6/(1*2) = 3$. Distinguishable permutations: {SFF, FSF, FFS}.

Since the n trials are independent with probabilities remaining constant between them, any arrangement of n trials with $n_1 = k$ successes (and $n_2 = (n - k)$ failures) has the same probability of being achieved (regardless of positions of successes within the arrangement). That probability is $p^k * (1 - p)^{(n-k)}$.

Recall from our section on basic probability laws, we had the example:

If events A, B, C are mutually exclusive events, then the joint probability $P(A, B, C) = 0$, and we are left with $P(A + B + C) = P(A) + P(B) + P(C)$.

Let $N =$ number of distinguishable permutations $= n! / (k! * (n-k)!)$ in the above example, and let $\{d_1, d_2, \dots, d_N\}$ denote the distinguishable arrangements of the k successes each mutually exclusive events whose joint probability is equal to zero.

Then

$$P(d_1 + d_2 + d_3 + \dots + d_N) = P(d_1) + P(d_2) + P(d_3) + \dots + P(d_N).$$

In our case each term is equal to $p^k * (1 - p)^{(n-k)}$, which is added $n! / (k! * (n-k)!)$ times to produce $P_b(k; n, p) = \text{binomial}(n, k) * p^k * (1 - p)^{(n-k)}$.

8.2 pdf_binomial (k, n, p), $n > 0$, $0 < p < 1$, $k = 0, 1, 2, \dots, n$

What we have called $P_b(k; n, p)$ is returned by the Maxima function `pdf_binomial(k, n, p)`, with $0 < p < 1$ and with n a positive integer and with $k = 0, 1, 2, \dots, n$. ("pdf" standing for probability distribution function.)

Assume $p = 0.5$ is the probability of success in each identical conditions trial. What is the probability of $k = 0$ success if we only have 1 such trial? 2 such trials? etc?

Answer: 0.5, 0.25, 0.125 etc.

Because $p = 0.5$, $1-p = 1 - 0.5 = 0.5 = p$, and

$P_b(k; n, 0.5) =$

$(n! / (k! * (n - k)!)) * p^k * (1 - p)^{(n - k)} = (n! / (k! * (n - k)!)) * (p^k) * (1 - p)^{(n - k)} = (1 - p)^n = p^n$.

(%i17) `pdf_binomial(k, n, p);`

(%o17) $\binom{n}{k} (1-p)^{n-k} p^k$

(%i18) `grind(%);`

$binomial(n,k)*(1-p)^{(n-k)}*p^k$

(%i19) `pdf_binomial(k, n, 0.5);`

(%o19) $0.5^n \binom{n}{k}$

(%i20) `grind(%);`

$0.5^n * binomial(n,k)$

(%i21) `pdf_binomial(0, n, 0.5);`

(%o21) $1.0 0.5^n$

The probability of zero success in one trial is $0.50 = p$,

(%i22) `pdf_binomial(0, 1, 0.5);`

(%o22) 0.5

The probability of zero successes in two trials is 0.25 since $p^2 = (0.5)^2 = 0.25$.

(%i23) `pdf_binomial(0, 2, 0.5);`

(%o23) 0.25

The probability of zero successes in ten trials is 0.0009765.

(%i24) `0.5^10;`

(%o24) $9.76563 \cdot 10^{-4}$

Convert a probability (which must be in the interval $[0, 1]$) to % chance by multiplying by 100.

```
(%i25) 100*%;
```

```
(%o25) 0.0976563
```

Watch out for the confusing notation - In Maxima the symbol % refers to the previous output, and has nothing to do with percent chance.

We get the same probability using pdf_binomial (0, 10, 0.5).

```
(%i26) pdf_binomial (0, 10, 0.5);
```

```
(%o26) 9.76563 10-4
```

What about the probability of getting zero successes in 100 trials (with $p = 0.5$ for each trial).

```
(%i27) pdf_binomial (0, 100, 0.5);
```

```
(%o27) 7.88861 10-31
```

8.3 [RS: Example 9]: Two Tosses of a Balanced Coin

Sidebar:

See the Oct. 11, 2023 article at: <https://phys.org/news/2023-10-flipped-coins-fair-thought.html> with the title: 'Flipped coins found not to be as fair as thought', by Bob Yirka, Phys.org, about practical conditions required for a 'fair' coin toss.

Quoting [RS]:

"The possible outcomes in two [successive] tosses [two trials] of a balanced coin are TT, TH, HT, and HH. Thus

$$P(0 H) = 1/4, P(1 H) = 1/2, \text{ and } P(2 H) = 1/4.$$

[Note the sum of the probabilities is 1]

The number of heads is therefore a discrete random variable, and the set of all possible outcomes [in two trials] is a 'discrete probability distribution'".

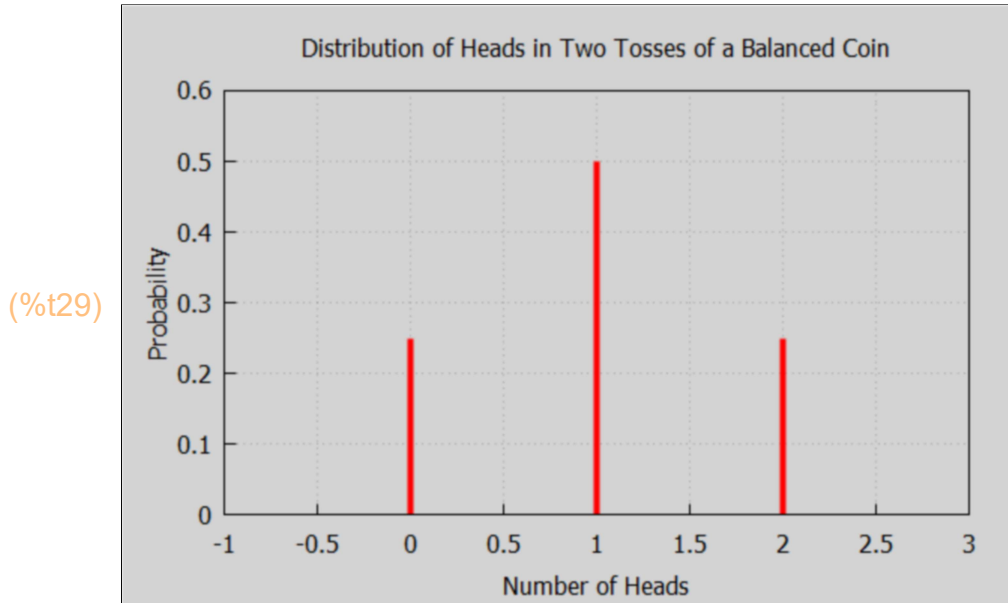
Since "the binomial distribution is used to 'find the probability of k number of occurrences or successes of an event $P(k)$, in n trials of the same experiment when (1) there are only two possible and mutually exclusive outcomes, (2) the $n = 2$ trials are independent, and (3) the probability of occurrence or success, $p = 0.5$ remains constant in each trial', the distribution of successes is described by a Binomial (2, 0.5) distribution.

Applying the Maxima function pdf_binomial (k, n, p) with $p = 0.5$, $n = 2$, $k = 0, 1, 2$,

```
(%i28) mypoints : makelist ([ k, pdf_binomial (k, 2, 0.5 ) ], k, 0, 2 );
```

```
(mypoints) [[0,0.25],[1,0.5],[2,0.25]]
```

```
(%i29) wxdraw2d ( xrange = [-1, 3], yrange = [0, 0.6], points_joined = impulses,
  xlabel = "Number of Heads", ylabel = "Probability",
  title = "Distribution of Heads in Two Tosses of a Balanced Coin",
  background_color = light_gray, grid = true,
  line_width = 4, color = red, points (mypoints) )$
```



8.3.1 mean_binomial (n, p)

What is the 'mean number of heads' in two trials for which $p = 1/2$?

The mean result (the average result) is advertised to be $n \cdot p = 2 \cdot (1/2) = 1$, also given using $\sum (x \cdot P(x), x, \text{xmin}, \text{xmax}) = 0 \cdot P(0) + 1 \cdot P(1) + 2 \cdot P(2) = 0 + 1 \cdot (1/2) + 2 \cdot (1/4) = 1/2 + 1/2 = 1$.

Using our basic definitions, we have a data set with $M = 3$ distinct values: $x_1 = 0$, $x_2 = 1$, and $x_3 = 2$ (heads), which occurs with frequencies $f_1 = 1$, $f_2 = 2$, $f_3 = 1$, and the sum of the frequencies $\sum (f_k, k, 1, M) = N = 4$. Then the mean is given by our basic definition:
 $\text{mean} = \langle x \rangle = \sum (f_k \cdot x_k, k, 1, M) / N = (1 \cdot 0 + 2 \cdot 1 + 1 \cdot 2) / 4 = 4/4 = 1$.

Using Maxima's `mean_binomial (n, p)` function:

```
(%i30) mean_binomial (2, 0.5);
```

```
(%o30) 1.0
```

8.3.2 var_binomial (n, p)

The theoretical standard deviation is advertised to be $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$, and the variance is $\sigma^2 = n \cdot p \cdot (1 - p) = \mu \cdot (1 - p) = 1 \cdot (1/2) = 1/2$.

The Maxima function: `var_binomial (n, p)` returns the variance of a Binomial (n, p) random variable, with $0 < p < 1$ and n a positive integer.

```
(%i31) var_binomial (2, 0.5);
```

```
(%o31) 0.5
```

Using our basic definition $\text{Var}(x) = \sum (f_k * (x_k - \langle x \rangle)^2, k, 1, M) / N$, we get $(1*(0 - 1)^2 + 2*(1 - 1)^2 + 1*(2 - 1)^2) / 4 = 2/4 = 1/2$.

8.3.3 std_binomial (n, p)

The standard deviation is the square root of the variance, so should be $\text{sqrt}(1/2)$.

```
(%i32) sqrt (1/2);
```

```
(%o32)  $\frac{1}{\sqrt{2}}$ 
```

```
(%i33) %, numer;
```

```
(%o33) 0.707107
```

The theoretical standard deviation is $\sigma = \text{sqrt}(n * p * (1 - p)) = \text{sqrt}(\mu * (1 - p)) = \text{sqrt}(1 * (1 - 1/2)) = \text{sqrt}(1/2)$.

The Maxima function `std_binomial (n, p)` returns the standard deviation σ of a Binomial (n, p) random variable, with $0 < p < 1$ and n a positive integer.

```
(%i34) std_binomial (2, 0.5);
```

```
(%o34) 0.707107
```

8.3.4 cdf_binomial (k, n, 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.

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

The 'cumulative distribution function' [cdf] `cdf_binomial (k, n, p)`

returns the sum of the values $P(k)$ for a Binomial (n, p) random variable, with $0 < p < 1$ and n a positive integer, beginning at $k = 0$ and ending at k. If k is n, the returned answer should be 1, since the sum of the probabilities of all possible events is 1.

$P(k \leq k1) = \text{cdf_geometric}(k1, n, p)$

```
(%i35) for j:0 thru 2 do print (j, cdf_binomial (j, 2, 0.5) )$
0 0.25
1 0.75
2 1
```

Here is a summary of the uses of pdf_binomial and cdf_binomial:

$P(k = k_1) = \text{pdf_binomial}(k_1, n, p)$, (probability of finding exact value k_1),
 $P(k < k_1) = \text{cdf_binomial}(k_1 - 1, n, p)$, (probability of finding a value less than k_1),
 $P(k \leq k_1) = \text{cdf_binomial}(k_1, n, p)$, (probability of finding a value less than or equal to k_1),
 $P(k > k_1) = 1 - \text{cdf_binomial}(k_1, n, p)$, (probability of finding a value greater than k_1),
 based on: $P(k \leq k_1) + P(k > k_1) = 1$,
 $P(k \geq k_1) = 1 - \text{cdf_binomial}(k_1 - 1, n, p)$, (probability of finding a value greater than or
 equal to k_1), based on: $P(k < k_1) + P(k \geq k_1) = 1$.

8.3.5 Example 1 of Use of cdf_binomial (k, n, p)

This is Example 10.6 from <https://online.stat.psu.edu/stat414/lesson/10/10.3>

By some estimates, twenty-percent (20%) of Americans have no health insurance. Randomly sample $n = 15$ Americans.

A.) Let k denote the number in the sample with no health insurance. What is the probability that exactly 3 of the 15 sampled have no health insurance?

Solution:

Since 15 is small relative to the population of $N = 300,000,000$ Americans, and all of the other criteria pass muster (two possible outcomes, independent trials, ...), the random variable k can be assumed to follow a binomial distribution with $n = 15$ and $p = 0.2$.

Using the probability mass function for a binomial random variable, the calculation is then relatively straightforward:

$$P(k = 3) = \text{binomial}(15, 3) * (0.2)^3 * (0.8)^{12} = 0.25.$$

```
(%i37) binomial (15, 3)* (0.2)^3 * (0.8)^12;
pdf_binomial (3, 15, 0.2);
(%o36) 0.250139
(%o37) 0.250139
```

That is, there is a 25% chance, in sampling 15 random Americans, that we would find exactly 3 that had no health insurance.

B.) What is the probability that at most one of those sampled has no health insurance?

Solution:

"At most one" means either 0 or 1 of those sampled have no health insurance. That is, we need to find:

$$P(k \leq 1) = P(0) + P(1), \text{ which is given by } \text{cdf_binomial}(1, n, p),$$

```
(%i38) cdf_binomial (1, 15, 0.2);
```

```
(%o38) 0.167126
```

That is, we have a 16.7% chance, in sampling 15 random Americans, that we would find at most one that had no health insurance.

C.) What is the probability that more than seven have no health insurance?

Solution:

Yikes! "More than seven" in the sample means finding one of the values 8, 9, 10, 11, 12, 13, 14, 15

$$P(k \leq 7) + P(k > 7) = 1 \text{ implies } P(k > 7) = 1 - P(k \leq 7) = 1 - \text{cdf_binomial}(7, 15, 0.2).$$

```
(%i39) 1 - cdf_binomial (7, 15, 0.2);
```

```
(%o39) 0.00423975
```

.... the probability that more than 7 in a random sample of 15 would have no health insurance is 0.0042 (the chance of finding that more than 7 in the random sample have no health insurance is about 0.4%).

D.) What is the probability that at least one of the 15 sampled has no health insurance?

$$P(k < 1) + P(k \geq 1) = 1 \text{ implies that } P(k \geq 1) = 1 - P(k < 1) = 1 - \text{cdf_binomial}(0, 15, 0.2)$$

```
(%i41) 1 - cdf_binomial (0, 15, 0.2);
```

```
1 - pdf_binomial (0, 15, 0.2);
```

```
(%o40) 0.964816
```

```
(%o41) 0.964816
```

That is, the probability that at least one person in a random sample of 15 would have no health insurance is 0.9648. (The chance of finding at least one person in a random sample of 15 with no health insurance is 96.5%.)

E.) What is the probability that fewer than 5 have no health insurance?

This asks for the probability $P(k < 5)$, ie., the number found is either zero, one, two, three, or four, which we can find using $\text{cdf_binomial}(4, 15, 0.2)$, [$P(k < 5) = P(k \leq 4)$]:

```
(%i42) cdf_binomial (4, 15, 0.2);
(%o42) 0.835766
```

.... the probability that fewer than 5 people in a random sample of 15 would have no health insurance is 0.8358.

8.3.6 Example 2 of Use of cdf_binomial (k, n, p)

This is Ex. 10.7 in <https://online.stat.psu.edu/stat414/lesson/10/10.3>

Many utility companies promote energy conservation by offering discount rates to consumers who keep their energy usage below certain established subsidy standards. A recent EPA report notes that 70% of the island residents of Puerto Rico have reduced their electricity usage sufficiently to qualify for discounted rates. If ten residential subscribers are randomly selected from San Juan, Puerto Rico, what is the probability that at least four qualify for the favorable rates?

The sample size is $n = 10$, the probability of a random island resident being qualified for the discount is $p = 0.7$. If $k =$ the number in the sample of ten who qualify for the discount, this question asks for $P(k \geq 4)$. Since $P(k < 4) + P(k \geq 4) = 1$, $P(k \geq 4) = 1 - P(k < 4) = 1 - \text{cdf_binomial}(3, 10, 0.7)$.

```
(%i43) 1 - cdf_binomial (3, 10, 0.7);
(%o43) 0.989408
```

.... the probability that at least four people in a random sample of ten would qualify for favorable rates is 0.9894. (~ 99% chance)

8.3.7 quantile_binomial (q, n, p)

The Maxima function `quantile_binomial (q,n,p)` returns the q -quantile of a $\text{Binomial}(n,p)$ random variable, with $0 \leq p \leq 1$ and n a positive integer. This is the inverse of `cdf_binomial`. Argument q must be an element of $[0,1]$ and q should be a value returned by `cdf_binomial`.

```
(%i44) for j:0 thru 2 do (
  qq : cdf_binomial (j, 2, 0.5),
  print (j, qq, quantile_binomial (qq, 2, 0.5) )) $
0 0.25 1
1 0.75 1
2 1 2
```

8.4 Number of Successes in 100 Flips of a Balanced Coin

We continue to assume the probability of success in any one trial is $p = 0.5$ (50% probability).

`mean_binomial (n, p)` produces the mean and `std_binomial (n,p)` calculates one standard deviation.

```
(%i45) mean_binomial (100, 0.5);
```

```
(%o45) 50.0
```

The average (ie., the mean) number of heads out of 100 independent flips of a balanced coin is predicted to be 50. Here is `std_binomial (100, 0.5)` which measures the spread about the mean number of heads (50):

```
(%i46) std_binomial (100, 0.5);
```

```
(%o46) 5.0
```

Let's calculate the probability of observing exactly $k = 50$ successes in $n = 100$ trials, if the probability of success on each single trial is $p = 0.5$.

```
(%i47) pdf_binomial (50, 100, 0.5);
```

```
(%o47) 0.0795892
```

Thus about 8% chance for getting exactly 50 successes in 100 trials.

Next we calculate the probability of observing $k = 0$ heads in 100 flips of a balanced coin.

```
(%i48) pdf_binomial (0, 100, 0.5);
```

```
(%o48) 7.88861 10-31
```

```
(%i49) rat(%);
```

```
(%o49)/R/ 
$$\frac{1}{1267650600228229401496703205376}$$

```

```
(%i50) float(denom(%));
```

```
(%o50) 1.26765 1030
```

The probability of getting zero heads in the course of 100 flips of a balanced coin is one part in about 10 raised to the 30 power.

See: [https://en.wikipedia.org/wiki/Orders_of_magnitude_\(numbers\)](https://en.wikipedia.org/wiki/Orders_of_magnitude_(numbers))
 10^{30} is about one fifth of the estimated number of bacterial cells on planet earth, an incredibly large number.

$$10^{(-30)} = (1,000)^{(-10)} = \text{one quintillionth.}$$

To get the probability that we get somewhere in the range of 45 successes to 55 successes in 100 trials, (mean +/- std) we need to sum over these probabilities from $x = 45$ to $x = 55$.

```
(%i51) sum (pdf_binomial (j, 100, 0.5), j, 45, 55);
(%o51) 0.728747
```

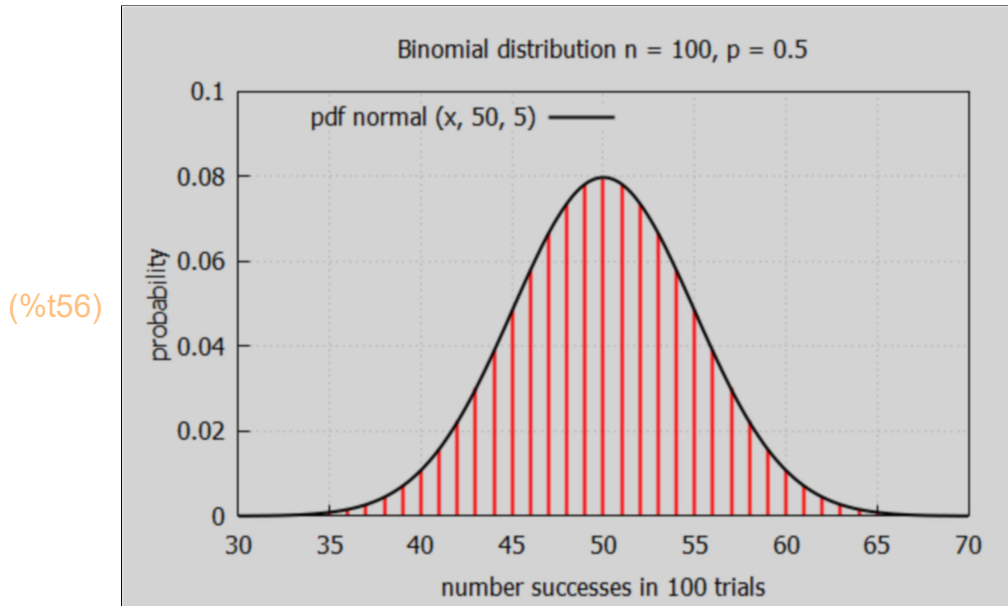
So about 73% chance of getting somewhere in the range of 45 - 55 successes in 100 trials.

We can make a plot of the probability distribution function pdf_binomial (k,100,0.5).

```
(%i55) mypoints : makelist ( [j, pdf_binomial (j, 100, 0.5) ], j, 20, 80)$
fill (mypoints);
head (mypoints);
tail (mypoints);
(%o53) [[20,4.22816 10-10],[80,4.22816 10-10],61]
(%o54) [[20,4.22816 10-10],[21,1.61073 10-9],[22,5.78398 10-9]]
(%o55) [[78,5.78398 10-9],[79,1.61073 10-9],[80,4.22816 10-10]]
```

The *continuous* normal distribution (see pdf_normal (x, m, s)) closely agrees with the discrete binomial distribution, provided $m = \text{mean_binomial}(100, 0.5) = 50$, and $s = \text{std_binomial}(100, 0.5) = 5$.

```
(%i56) wxdraw2d ( xrange = [30, 70], yrange = [0, 0.1], points_joined = impulses,
  xlabel = "number successes in 100 trials", ylabel = "probability", grid = true,
  title = "Binomial distribution n = 100, p = 0.5",
  background_color = light_gray,
  line_width = 2, color = red, points (mypoints),
  color = black, key_pos = top_left,
  key = "pdf normal (x, 50, 5)",
  explicit (pdf_normal (x, 50, 5), x, 30, 70) )$
```



pdf_normal (x, m, s) produces the probability at location x for a normal distribution having mean m and standard deviation s.

The maximum of the normal distribution used is given by:

```
(%i57) pdf_normal (50, 50, 5), numer;
(%o57) 0.0797885
```

which is also the maximum of our binomial distribution.

The area (probability) in the x-interval $[m - 3*s, m + 3*s]$, or within 3 standard deviations of the mean is almost equal to 1.

```
(%i58) integrate (pdf_normal (x, 50, 5), x, 35, 65), numer;
(%o58) 0.9973
```

Summing the values of pdf_binomial (k, n, p) we get very close to the same answer:

```
(%i59) sum (pdf_binomial (j, 100, 0.5), j, 35, 65);
(%o59) 0.99821
```

8.4.1 Inverse Type of Problem

Question: There is a 95% chance that the number of heads (in 100 coin flips) will be in the range $[0, x]$. What is the approximate value of x ?

An *approximate* answer is provided by `quantile_binomial (0.95, 100, 0.5)`, since as yet we don't know if 0.95 is a possible return value of `cdf_binomial (k, 100, 0.5)` with k equal to an integer.

```
(%i60) quantile_binomial (0.95, 100, 0.5);
(%o60) 58
```

We can check this approximate value as follows:

```
(%i61) for j: 57 thru 59 do print (j, cdf_binomial (j, 100, 0.5) )$
57 0.933395
58 0.955687
59 0.971556
```

There is about a 95% chance that the number of heads (in 100 coin flips) will lie in the range of 0 to between 57 and 58.

8.4.2 binomialCalc (k, n, p)

What is the probability (when $n = 100$ trials and $p = 0.5$) that $k < 55$?

That is, what is the probability that the number of successes in 100 trials (flips) is less than 55? One way to calculate this is by summing the values of `pdf_binomial (k, 100, 0.5)` from $k = 0$ to $k = 54$.

```
(%i62) sum (pdf_binomial (k, 100, 0.5), k, 0, 54);
(%o62) 0.815899
```

Another way is to use `cdf_binomial (54, 100, 0.5)`:

```
(%i63) cdf_binomial (54, 100, 0.5);
(%o63) 0.815899
```

What is the probability the number of successes in 100 trials (flips) is less than or equal to 55?

```
(%i65) sum (pdf_binomial (k, 100, 0.5), k, 0, 55);
cdf_binomial (55, 100, 0.5);
(%o64) 0.864373
(%o65) 0.864373
```

What is the probability the number of successes in 100 trials (flips) is greater than 55?

```
(%i67) 1 - sum (pdf_binomial (k, 100, 0.5), k, 0, 55);
1 - cdf_binomial (55, 100, 0.5);
(%o66) 0.135627
(%o67) 0.135627
```

What is the probability the number of successes in 100 trials is greater than or equal to 55?

```
(%i69) 1 - sum (pdf_binomial (k, 100, 0.5), k, 0, 54);
1 - cdf_binomial (54, 100, 0.5);
(%o68) 0.184101
(%o69) 0.184101
```

We can incorporate these methods into a Maxima function `binomialCalc(xx, n, p)` which shows the probabilities $P(k = xx)$, $P(k < xx)$, $P(k \leq xx)$, $P(k > xx)$, and $P(k \geq xx)$ as output.

```
(%i70) binomialCalc (xx, nn, pp) :=
block (
  print (" "),
  print (sconcat (" For n = ", nn, ", p = ", pp, ", the probabilities are:")),
  print (" "),
  print (sconcat (" P ( k = ", xx, " ) : ", pdf_binomial (xx, nn, pp) )),
  print (sconcat (" P ( k < ", xx, " ) : ", cdf_binomial (xx - 1, nn, pp) )),
  print (sconcat (" P ( k <= ", xx, " ) : ", cdf_binomial (xx, nn, pp) )),
  print (sconcat (" P ( k > ", xx, " ) : ", 1 - cdf_binomial (xx, nn, pp) )),
  print (sconcat (" P ( k >= ", xx, " ) : ", 1 - cdf_binomial (xx - 1, nn, pp) )),
  done )$
(%i71) binomialCalc (55, 100, 0.5)$
```

For n = 100, p = 0.5, the probabilities are:

```
P ( k = 55 ) : 0.0484743
P ( k < 55 ) : 0.815899
P ( k <= 55 ) : 0.864373
P ( k > 55 ) : 0.135627
P ( k >= 55 ) : 0.184101
```

These answers agree with the results returned from
<https://www.statology.org/binomial-distribution-calculator/>

8.4.3 confidence (q, m, s)

What is the q confidence interval about the mean m when we can approximate the distribution with a (continuous) Normal (m, s) distribution?

Suppose $q = 0.95$ (in the case we are looking for the 95% confidence interval about the mean). The normal distribution is symmetrical, so we can say that there is 95% confidence that the resulting number of successes in 100 trials lies between x_1 and x_2 , where $x_1 = m - dx$, $x_2 = m + dx$ (for some dx); that is the area (probability) under the normal curve in the interval $[m - dx, m + dx]$ is equal to 0.95. That means that the area left for the tails (left tail and right tail combined) is 0.05 (5%) and the area of the left tail ($x < x_1$) is 0.025 (2.5%) and the area under the right tail ($x > x_2$) is 0.025 (2.5%) as well.

We then generalize to an arbitrary choice for q . We make use of the Maxima function `quantile_normal (q, m, s)` which we will cover in the worksheet Stat04-Normal.wmxm. Because the Normal (m, s) distribution is an example of a continuous probability distribution function such as `pdf_normal (x, m, s)`, where x can assume any real valued number, we can use `quantile_normal (q, m, s)` with any real number q .

With $q = 0.95$, we find x_2 using `quantile_normal (0.975, m, s)` since the combined left tail and the interval $[x_1, x_2]$ area is $0.95 + 0.025 = 0.975$. We find x_1 using `quantile_normal (0.025, m, s)`. We then find $dx = (x_2 - x_1)/2$

```
(%i72) confidence (qq, mm, ss) :=
block ([ddx, xx1, xx2],
  float ( quantile_normal (qq + (1 - qq)/2, mm, ss) -
          quantile_normal ( (1 - qq)/2 , mm, ss) ),
  ddx : %% / 2,
  xx1 : mm - ddx,
  xx2 : mm + ddx,
  print ( "delx = ", ddx ),
  print ( " x1 = ", xx1, ", x2 = ", xx2 ),
  [xx1, xx2])$
```

For our 100 flips of a balanced coin example, the mean $m = 50$, the standard deviation $s = 5$.

```
(%i73) confidence (0.95, 50, 5)$
delx = 9.79982
x1 = 40.2002 , x2 = 59.7998
```

```
(%i74) confidence (0.95, 50, 5);
delx = 9.79982
x1 = 40.2002 , x2 = 59.7998
```

```
(%o74) [40.2002,59.7998]
```


We can thus be 95% confident that the number of successes in 100 flips of a balanced coin will lie in the range (40 - 60).

We can alter confidence (q,m,s) to confidence_integral (q,m,s) which rounds x1 and x2 to the closest integer.

```
(%i75) confidence_integral (qq, mm, ss) :=
  block ([ddx, xx1, xx2],
    float ( quantile_normal (qq + (1 - qq)/2, mm, ss) -
      quantile_normal ( (1 - qq)/2 , mm, ss) ),
    ddx : %% / 2,
    xx1 : mm - ddx,
    xx2 : mm + ddx,
    print ( "delx = ", ddx ),
    print ( " x1 = ", round (xx1), ", x2 = ", round (xx2) ),
    round ( [xx1, xx2] ))$
```

```
(%i76) confidence_integral (0.95, 50, 5)$
delx = 9.79982
x1 = 40 , x2 = 60
```

```
(%i77) confidence_integral (0.95, 50, 5);
delx = 9.79982
x1 = 40 , x2 = 60
```

```
(%o77) [40,60]
```

The following examples are from:

<https://www.statology.org/binomial-distribution-real-life-examples/>
which offers the web based calculator referred to just above.

8.5 Statology 1: Number of Side Effects from Medications

This example is from the webpage

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

"Medical professionals use the binomial distribution to model the probability that a certain number of patients will experience side effects as a result of taking new medications."

"For example, suppose it is known that 5% of adults who take a certain medication experience negative side effects. We can use a Binomial Distribution Calculator to find the probability that more than a certain number of patients in a random sample of 100 will experience negative side effects."

" $P(X > 5 \text{ patients experience side effects}) = 0.38400$

$P(X > 10 \text{ patients experience side effects}) = 0.01147$

$P(X > 15 \text{ patients experience side effects}) = 0.0004$ [we will show this is too large by a factor of 10]. And so on."

"This gives medical professionals an idea of how likely it is that more than a certain number of patients will experience negative side effects."

In this example, each patient taking the medication either experiences negative side effects ('success') or does not experience negative side effects. The probability of 'success' is $p = 0.05$ (5%), and $n = 100$ patients take the same medication (100 trials), each with the same probability p of negative side effects. Then the average number of patients (out of 100) who experience negative side effects is 5 (out of 100) and the standard deviation is about 2.2.

```
(%i79) mean_binomial (100, 0.05);
```

```
std_binomial (100, 0.05);
```

```
(%o78) 5.0
```

```
(%o79) 2.17945
```

Using our homemade Maxima function `binomialCalc (x, n, p)`,

```
(%i80) binomialCalc (5, 100, 0.05)$
```

For $n = 100$, $p = 0.05$, the probabilities are:

$P(k = 5) : 0.180018$

$P(k < 5) : 0.435981$

$P(k \leq 5) : 0.615999$

$P(k > 5) : 0.384001$

$P(k \geq 5) : 0.564019$

The above output shows that the example assertion $P(k > 5) = 0.384$ is correct.

(%i81) `binomialCalc (10, 100, 0.05)`\$

For $n = 100$, $p = 0.05$, the probabilities are:

$$P(k = 10) : 0.0167159$$

$$P(k < 10) : 0.971812$$

$$P(k \leq 10) : 0.988528$$

$$P(k > 10) : 0.0114724$$

$$P(k \geq 10) : 0.0281883$$

The output above shows that the example assertion $P(k > 10) = 0.1147$ is correct.

(%i82) `binomialCalc (15, 100, 0.05)`\$

For $n = 100$, $p = 0.05$, the probabilities are:

$$P(k = 15) : 9.88002e-5$$

$$P(k < 15) : 0.999864$$

$$P(k \leq 15) : 0.999963$$

$$P(k > 15) : 3.70541e-5$$

$$P(k \geq 15) : 1.35854e-4$$

This last calculation shows $P(x > 15) = 3.7 \cdot (10^{-5}) = 3.7e-5 = 0.000037 \sim 0.00004$, which is a factor of 10 smaller than the webpage asserts.

There is a 95% chance that the number of patients out of 100 who experience negative side effects will be in the range $[0, x]$. What is the approximate value of x ?

An approximate answer is provided by `quantile_binomial (0.95, 100, 0.05)`

(%i83) `quantile_binomial (0.95, 100, 0.05);`

(%o83) 9

We can check this estimate as follows:

(%i85) `cdf_binomial (8, 100, 0.05);`

`cdf_binomial (9, 100, 0.05);`

(%o84) 0.93691

(%o85) 0.971812

There is a 95% chance the number of patients who experience negative side effects (out of 100) will lie in the range (0, 8 or 9).

Here we try using `confidence_integral` (q, m, s).

We can have 95% confidence that the number of patients who experience negative side effects (out of 100) will lie in the interval $(m - dx, m + dx)$. We know the mean number is $m = 5$, and one standard deviation $s = 2.2$ approximately.

```
(%i86) confidence_integral (0.95, 5, 2.2)$
```

```
delx = 4.31192
```

```
x1 = 1 , x2 = 9
```

With dx about 4 and the mean equal to 5 we can be 95% confident that the number of patients experiencing negative side effects (out of 100) will be in the approximate range (1, 9).

So both methods give the same approximate range.

8.6 Statology 2: Number of Fraudulent Credit Card Transactions

This example is from the webpage

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

"Banks use the binomial distribution to model the probability that a certain number of credit card transactions are fraudulent."

"For example, suppose it is known that 2% of all credit card transactions in a certain region are fraudulent. If there are 50 transactions per day in a certain region, we can use a Binomial Distribution Calculator to find the probability that more than a certain number of fraudulent transactions occur in a given day:" [assuming a random sample of 50 transactions per day]

" $P(X > 1 \text{ fraudulent transaction}) = 0.26423$

$P(X > 2 \text{ fraudulent transactions}) = 0.07843$

$P(X > 3 \text{ fraudulent transactions}) = 0.01776$ "

"And so on."

"This gives banks an idea of how likely it is that more than a certain number of fraudulent transactions will occur in a given day."

In this example, each credit card transaction either is fraudulent ('success') or is not fraudulent. The probability of 'success' is $p = 0.02$ (2%), and $n = 50$ credit card transactions occur (50 trials), each with the same probability p of being fraudulent.

```
(%i88) mean_binomial (50, 0.02);  
      std_binomial (50, 0.02);  
(%o87) 1.0  
(%o88) 0.989949
```

Let's show a plot of the probability distribution for this case Binomial (50, 0.02).

```
(%i89) mypoints : makelist ( [j, pdf_binomial (j, 50, 0.02) ], j, 0, 5 );  
(mypoints) [[0,0.36417],[1,0.371602],[2,0.185801],[3,0.0606697],[4,  
0.0145483],[5,0.00273152]]
```

The *continuous* normal distribution (see pdf_normal (x, m, s)) closely agrees with the discrete binomial distribution, provided $m = \text{mean_binomial}(50, 0.2) = 1$, and $s = \text{std_binomial}(50, 0.2) = 0.99$, as we just calculated. Recall that pdf_binomial (k, n, p) is only nonzero for $k = 0, 1, 2, \dots$, so pdf_binomial (0.5, n, p) = 0 (see just below).

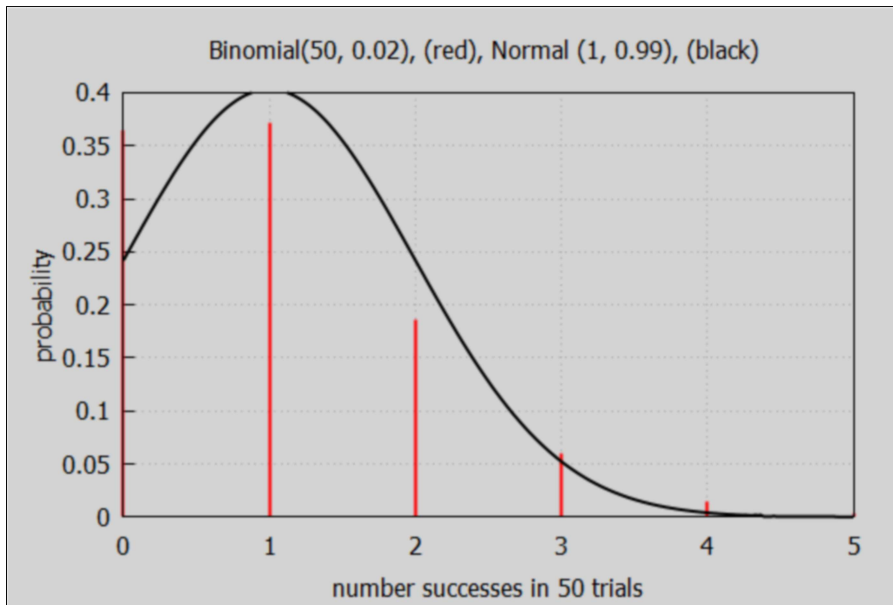
```
(%i90) pdf_binomial (0, 50, 0.02);  
(%o90) 0.36417
```

```
(%i91) pdf_binomial (0.5, 50, 0.02);  
(%o91) 0
```

```
(%i92) pdf_binomial (1, 50, 0.02);  
(%o92) 0.371602
```

```
(%i93) wxdraw2d ( xrange = [0, 5], yrange = [0, 0.4], points_joined = impulses,
  xlabel = "number successes in 50 trials", ylabel = "probability", grid = true,
  title = "Binomial(50, 0.02), (red), Normal (1, 0.99), (black)",
  background_color = light_gray,
  line_width = 2, color = red, points (mypoints),
  color = black, explicit (pdf_normal (x, 1, 0.99), x, 0, 5) )$
```

(%t93)



Assertion: $P(X > 1 \text{ credit card transactions [out of 50] are fraudulent}) = 0.26423$.
Maxima check:

```
(%i94) binomialCalc (1, 50, 0.02)$
```

For $n = 50$, $p = 0.02$, the probabilities are:

$$P(k = 1) : 0.371602$$

$$P(k < 1) : 0.36417$$

$$P(k \leq 1) : 0.735771$$

$$P(k > 1) : 0.264229$$

$$P(k \geq 1) : 0.63583$$

The chance that more than 1 credit card transaction (out of 50) will be fraudulent is about 26%.

Assertion: $P(X > 2 \text{ credit card transactions [out of 50] are fraudulent}) = 0.07843$
Maxima check:

```
(%i95) binomialCalc (2, 50, 0.02)$
```

For $n = 50$, $p = 0.02$, the probabilities are:

```
P ( k = 2 ) : 0.185801
P ( k < 2 ) : 0.735771
P ( k <= 2 ) : 0.921572
P ( k > 2 ) : 0.0784277
P ( k >= 2 ) : 0.264229
```

The chance that more than 2 credit card transactions (out of 50) will be fraudulent is about 8%.

Assertion: $P(X > 3 \text{ credit card transactions [out of 50] are fraudulent}) = 0.01776$
Maxima check:

```
(%i96) binomialCalc (3, 50, 0.02)$
```

For $n = 50$, $p = 0.02$, the probabilities are:

```
P ( k = 3 ) : 0.0606697
P ( k < 3 ) : 0.921572
P ( k <= 3 ) : 0.982242
P ( k > 3 ) : 0.0177581
P ( k >= 3 ) : 0.0784277
```

The chance that more than 3 credit card transactions (out of 50) will be fraudulent is about 2%.

There is a 95% chance that the number of fraudulent transactions will be in the range $[0, x]$.
What is the approximate value of x ?

An approximate answer is provided by `quantile_binomial (0.95, 50, 0.02)`.

```
(%i97) quantile_binomial (0.95, 50, 0.02);
```

```
(%o97) 3
```

We can check this as follows:

```
(%i98) for j:2 thru 4 do print (j, cdf_binomial (j, 50, 0.02) )$
2 0.921572
3 0.982242
4 0.99679
```

There is a 95% chance the number of fraudulent credit card transactions per day will lie in the interval [0, 2 or 3].

8.7 Statology 3: Number of Spam Emails per Day

This example is from

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

"Email companies use the binomial distribution to model the probability that a certain number of spam emails land in an inbox per day."

"For example, suppose it is known that 4% of all emails are spam. If an account receives 20 emails in a given day, we can use a Binomial Distribution Calculator to find the probability that a certain number of those emails are spam:"

"P(X = 0 spam emails) = 0.44200

P(X = 1 spam email) = 0.36834

P(X = 2 spam emails) = 0.14580

And so on."

```
(%i99) makelist ([ j, pdf_binomial (j, 20, 0.04) ], j, 0, 2);
```

```
(%o99) [[0,0.442002],[1,0.368335],[2,0.145799]]
```

Out of 20 emails per day, (assuming a 4% spam rate) the chance of getting exactly no spam is 44%, the chance of getting one spam email is 36%, and the chance of getting two spam emails is about 15%.

8.8 Statology 4: Number of River Overflows

This example is from
<https://www.statology.org/binomial-distribution-real-life-examples/>

"Park systems use the binomial distribution to model the probability that rivers overflow a certain number of times each year due to excessive rain."

"For example, suppose it is known that a given river overflows during 5% of all storms. If there are 20 storms in a given year, we can use a Binomial Distribution Calculator to find the probability that the river overflows a certain number of times:"

"P(X = 0 overflows) = 0.35849
 P(X = 1 overflow) = 0.37735
 P(X = 2 overflows) = 0.18868"

"And so on."

"This gives the parks departments an idea of how many times they may need to prepare for overflows throughout the year."

```
(%i100) makelist ([j, pdf_binomial (j, 20, 0.05)], j, 0, 2);
(%o100) [[0,0.358486],[1,0.377354],[2,0.188677]]
```

Assuming a given river overflows during 5% of all storms, and assuming 20 storms/year, the chance there will be no overflow ($X = 0$) is about 36%, the chance there will be exactly one overflow during the 20 storms is about 38%, and the chance there will be exactly two river overflows during the 20 storms is about 19%.

```
(%i101) binomialCalc (2, 20, 0.05)$
```

For $n = 20$, $p = 0.05$, the probabilities are:

$P(k = 2) : 0.188677$
 $P(k < 2) : 0.73584$
 $P(k \leq 2) : 0.924516$
 $P(k > 2) : 0.0754837$
 $P(k \geq 2) : 0.26416$

```
(%i102) mean_binomial (20, 0.05);
(%o102) 1.0
```

8.9 Barlow Prob. 3.1

A defence system is 99.5% efficient in intercepting ballistic missiles.

A.) What is the probability that it will intercept all of 100 missiles launched against it?

The probability $P(k,n)$ that the system will intercept exactly k missiles out of n incoming if the defence system intercepts each missile with a probability $p = 0.995$, and we treat each missile as an (independent Bernoulli trial) is given by the binomial distribution.

$$P_{\text{intercept}}(k,n) = \text{binomial}(n, k) * p^k * (1-p)^{(n-k)} = \text{pdf_binomial}(k,n,p)$$

For $n = 100$ missiles and $k = 100$ intercepts,

```
(%i103) pdf_binomial (100, 100, 0.995);
```

```
(%o103) 0.60577
```

So about 61% chance of intercepting all 100 out of 100 incoming.

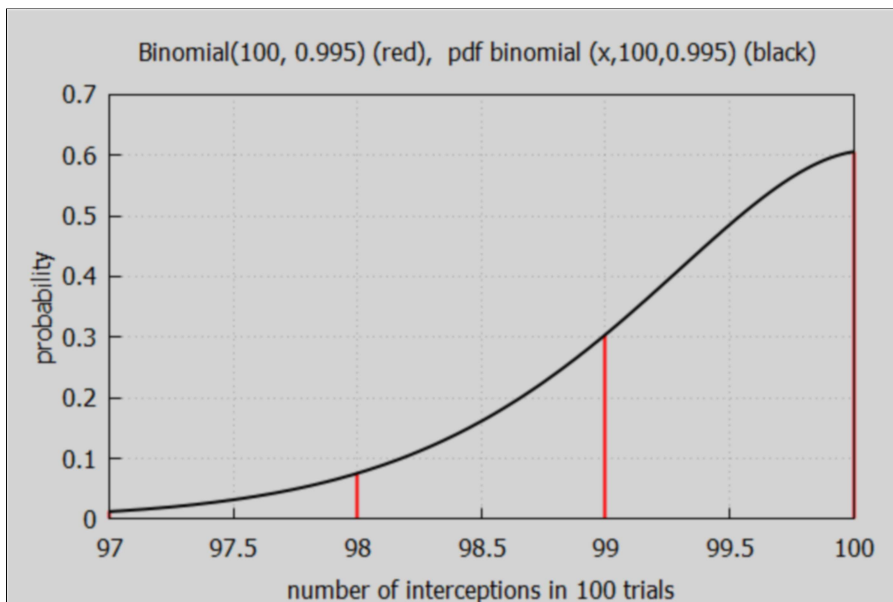
Let's make a plot of the values of the Binomial ($n = 100$, $p = 0.995$) for $k \leq 100$.

```
(%i104) mypoints : makelist ( [k, pdf_binomial (k, 100, 0.995) ], k, 97, 100 );
```

```
(mypoints) [[97,0.0124296],[98,0.0757194],[99,0.304407],[100,0.60577]]
```

```
(%i105) wxdraw2d ( xrange = [97, 100], yrange = [0, 0.7], points_joined = impulses,
  xlabel = "number of interceptions in 100 trials", ylabel = "probability", grid = true,
  title = "Binomial(100, 0.995) (red), pdf binomial (x,100,0.995) (black) ",
  background_color = light_gray,
  line_width = 2, color = red, points (mypoints),
  color = black, explicit (pdf_binomial (x, 100, 0.995), x, 97, 100) )$
```

```
(%t105)
```



B.) How many missiles must an aggressor launch to have a better than even chance of one or more penetrating the defences?

'one or more' means at least one. Let x = number of missiles which evade the defense system. The probability of any given missile evading the defenses is $p_e = 1 - 0.995 = 0.005$.

This question is asking: for what number n of missiles launched do we have $P_{\text{evade}}(x \geq 1) = 0.5$?

Since $P_{\text{evade}}(x < 1) + P_{\text{evade}}(x \geq 1) = 1$, we have the identity:

$$P_{\text{evade}}(x \geq 1) = 1 - P_{\text{evade}}(x < 1) = 1 - \text{cdf_binomial}(0, n, 0.005), \text{ or}$$

$$P_{\text{evade}}(x \geq 1) = 1 - \text{pdf_binomial}(0, n, 0.005).$$

8.9.1 Numerical Solution

```
(%i106) eqn : 1 - pdf_binomial (0, n, 0.005) = 0.5;
```

```
(eqn) 1 - 1.0 0.995^n = 0.5
```

```
(%i107) expr : lhs (eqn) - rhs(eqn);
```

```
(expr) 0.5 - 1.0 0.995^n
```

```
(%i108) find_root (expr, n, 100, 200);
```

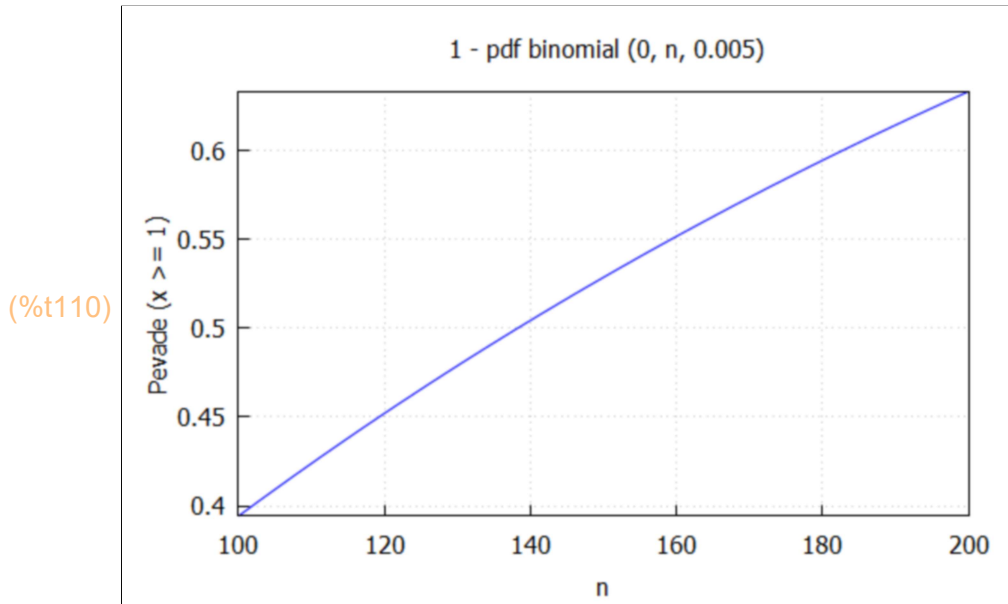
```
(%o108) 138.283
```

8.9.2 Graphical Solution

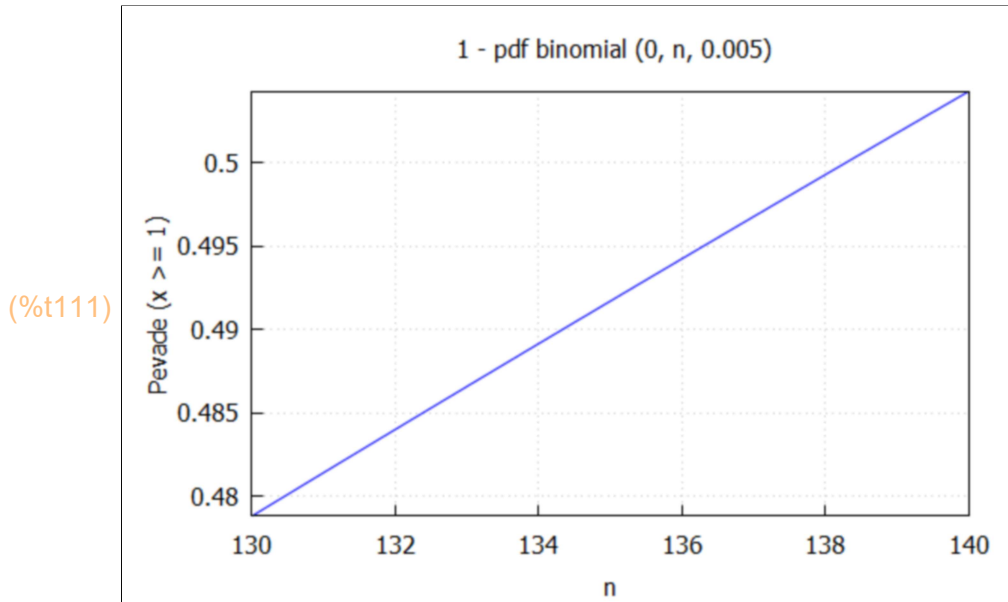
We first survey the feasible solutions graphically. Define `look (n1, n2)` which plots the curve of $1 - \text{pdf_binomial}(0, n, 0.005)$ for $n_1 \leq n \leq n_2$.

```
(%i109) look (n1, n2) := wxdraw2d (xlabel = "n", ylabel = "Pevade (x >= 1)",
    xrange = [n1, n2], title = "1 - pdf binomial (0, n, 0.005)", grid = true,
    explicit (1 - pdf_binomial (0, n, 0.005), n, n1, n2) )$
```

```
(%i110) look(100, 200)$
```



```
(%i111) look (130, 140)$
```



For 138 missiles launched we appear to reach $P_{\text{evade}} = \sim 0.5$. Print out the values near $n = 138$.

```
(%i112) for n : 135 thru 140 do print (n, 1 - pdf_binomial (0, n, 0.005) )$
```

```
135 0.491705
136 0.494246
137 0.496775
138 0.499291
139 0.501795
140 0.504286
```

Thus with 139 missiles launched, the aggressor has an even chance of one or more (at least one) missiles evading the defence system.

8.9.3 Analytic Solution

For an analytic approach to this problem:

```
(%i113) eqn : 1 - pdf_binomial (0, n, 0.005) = 0.5;
```

```
(eqn) 1 - 1.0 0.995^n = 0.5
```

```
(%i114) eqn - 1;
```

```
(%o114) -1.0 0.995^n = -0.5
```

```
(%i115) -1*%;
```

```
(%o115) 1.0 0.995^n = 0.5
```

```
(%i116) log (%);
```

```
(%o116) -0.00501254 n = -0.693147
```

```
(%i117) solve (% , n), numer;
```

```
(%o117) [n = 138.283]
```

The analytic solution thus says the aggressor should launch 138 missiles to have an even chance of one or more evading the defense system.

8.10 Barlow Prob. 3.2

In the previous question, how many missiles would be needed to ensure a better than even chance of more than two missiles evading the defences?

$P(x \leq 2) + P(x > 2) = 1$ implies:

$P(x > 2) = 1 - P(x \leq 2)$, or

$P(x > 2) = 1 - \text{cdf_binomial}(2, n, 0.005)$

8.10.1 numerical solution

```
(%i118) eqn : expand (1 - cdf_binomial (2, n, 0.005) ) = 0.5;
```

```
(eqn) -1.25 10^-5 n^2 0.995^n-2 - 0.0049625 n 0.995^n-2 - 0.990025 0.995^n-2 + 1 = 0.5
```

```
(%i119) expr : lhs (eqn) - rhs (eqn);
```

```
(expr) -1.25 10-5 n2 0.995n-2 - 0.0049625 n 0.995n-2 - 0.990025 0.995n-2 +
0.5
```

find_root (expr, variable, start, stop) looks for the value of the variable for which the expression expr passes through zero in the range [start, stop].

```
(%i120) find_root (expr, n, 100, 800);
```

```
(%o120) 534.475
```

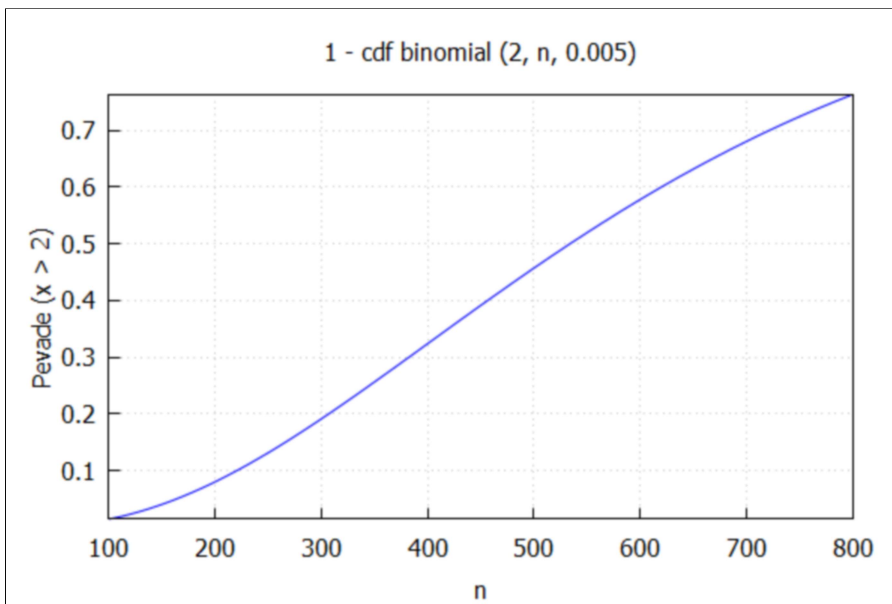
8.10.2 graphical solution

Graphical exploration for various values of n.

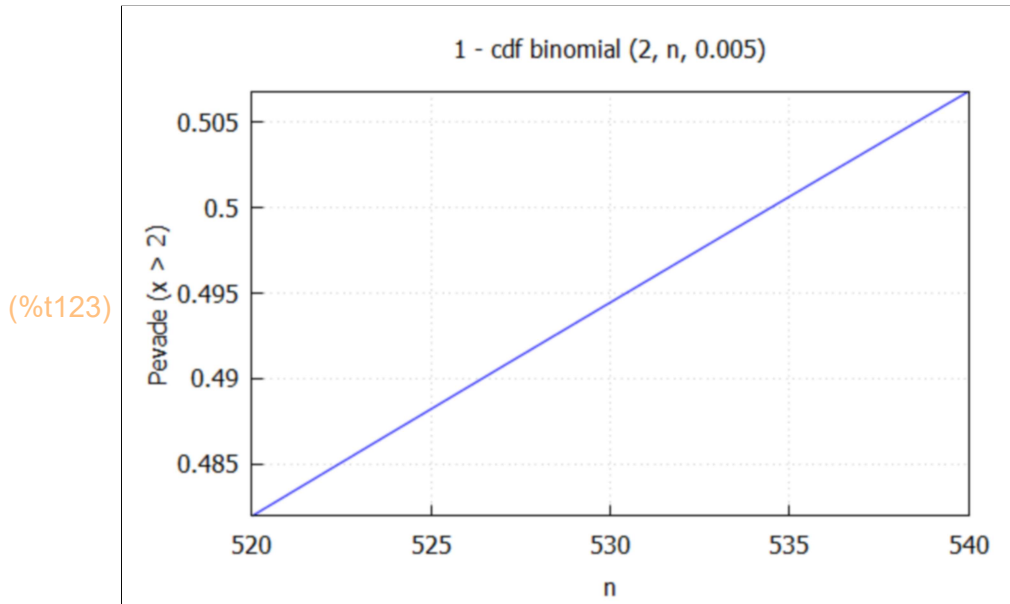
```
(%i121) look (n1, n2) := wxdraw2d (xlabel = "n", ylabel = "Pevade (x > 2)",
xrange = [n1, n2], grid = true,
title = " 1 - cdf binomial (2, n, 0.005) ",
explicit (1 - cdf_binomial (2, n, 0.005), n, n1, n2) )$
```

```
(%i122) look (100, 800)$
```

```
(%t122)
```



```
(%i123)look (520, 540)$
```



Print out values for n in the range [530, 540]:

```
(%i124)for n : 530 thru 540 do print (n, 1 - cdf_binomial (2, n, 0.005) )$
```

```
530 0.494453
531 0.495695
532 0.496936
533 0.498175
534 0.499413
535 0.500649
536 0.501883
537 0.503116
538 0.504347
539 0.505577
540 0.506805
```

The agressor needs 535 missiles launched to have an even chance of more than two missiles (three or more) penetrating the defense system.

8.11 random_binomial (n, p), random_binomial (n, p, m)

The Maxima function `random_binomial (n, p)` returns a Binomial (n, p) random variate, with $0 < p < 1$ and n a positive integer. Each time you run the following command, you will get a different set of five numbers of successes (in general).

Choosing $n = 100$, $p = 0.5$, reminds us of 100 flips of a balanced coin. With $p = 0.5$, the mean number of heads ("success") out of 100 flips is 50, so we expect `random_binomial (100, 0.5)` to return a number fairly close to 50 more often than not.

```
(%i125) mean_binomial (100, 0.5);
(%o125) 50.0
```

```
(%i126) for j thru 5 do print (j, random_binomial (100, 0.5) )$
1 52
2 51
3 46
4 59
5 46
```

8.11.1 Random Sample of size $m = 10$, $n = 100$, $p = 0.5$

The Maxima function `random_binomial (n, p, m)` returns a "random sample of size m ", with each of the m samples being the result of a simulation of n trials of an identical experiment, each trial having the same probability of success p . Each trial is independent of the other trials, and in each trial there are only two possible and mutually exclusive outcomes.

Thus `random_binomial (100, 0.5, 10)` calls `random_binomial (100, 0.5)` ten times and returns a list of ten values, each value being the number of successes found in 100 identical trials. For this example we know that the average (mean) number of successes in 100 identical trials is 50 (which is $n \cdot p$).

```
(%i130) rsample : random_binomial (100, 0.5, 10)$
fill (rsample);
head (rsample);
tail (rsample);
(%o128) [45,49, 10]
(%o129) [45,49,54]
(%o130) [45,57,49]
```

Here is the whole list of 10 values.

```
(%i131) rsample;
(%o131) [45,49,54,55,48,49,49,45,57,49]
```

With such a small sample there are no repeated integral values in this list.

8.11.2 Random sample size $m = 100$ simulations, $n = 100$, $p = 0.5$

We expect to get some integer repetitions in the returned list of values if we use `random_binomial (100, 0.5, 100)`, which returns a list of 100 integers.

```
(%i135) rsample : random_binomial (100, 0.5, 100)$
      fill (rsample);
      head (rsample);
      tail (rsample);
(%o133) [48,52,100]
(%o134) [48,45,49]
(%o135) [46,52,52]
```

8.11.3 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', in the following, element by element.

```
(%i138) [uniqueData, frequencies] : discrete_freq (rsample)$
      uniqueData;
      frequencies;
(%o137) [39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,
      58,59,60]
(%o138) [1,1,1,2,7,3,2,11,5,9,6,2,4,8,10,6,7,5,5,2,2,1]
```

If we add the integers in the list 'frequencies, we will get 100, since `rsample` has 100 elements, and each of the 100 elements (an integer) will contribute unity to one of the `uniqueData` elements.

```
(%i139) Lsum (frequencies);
(%o139) 100
```

Let `nfrequencies` be a list of "normalized" frequencies (so the sum of the elements of `nfrequencies` equals 1 (we need the sum of the probabilities to equal 1).

```
(%i140) nfrequencies : frequencies/100.0;
(nfrequencies) [0.01,0.01,0.01,0.02,0.07,0.03,0.02,0.11,0.05,0.09,0.06,0.02,
      0.04,0.08,0.1,0.06,0.07,0.05,0.05,0.02,0.02,0.01]
```

```
(%i141) Lsum (nfrequencies );
(%o141) 1.0
```

```
(%i142) lmax (nfrequencies);
```

```
(%o142) 0.11
```

To make a plot of probabilities (vertical axis) versus specific numbers of successes in 100 trials (horizontal axis), we let mypoints be a list of 100 sublists, with each sublist containing $[x, y] = [\text{number successes in 100 trials, the corresponding probability}] = [\text{uniqueData}[j], \text{nfrequencies}[j]]$ based on the $m = 100$ simulations.

```
(%i146) mypoints : makelist ([ uniqueData[j], nfrequencies [j] ], j, 1, length (uniqueData))$
```

```
fill (mypoints);
```

```
head (mypoints);
```

```
tail (mypoints);
```

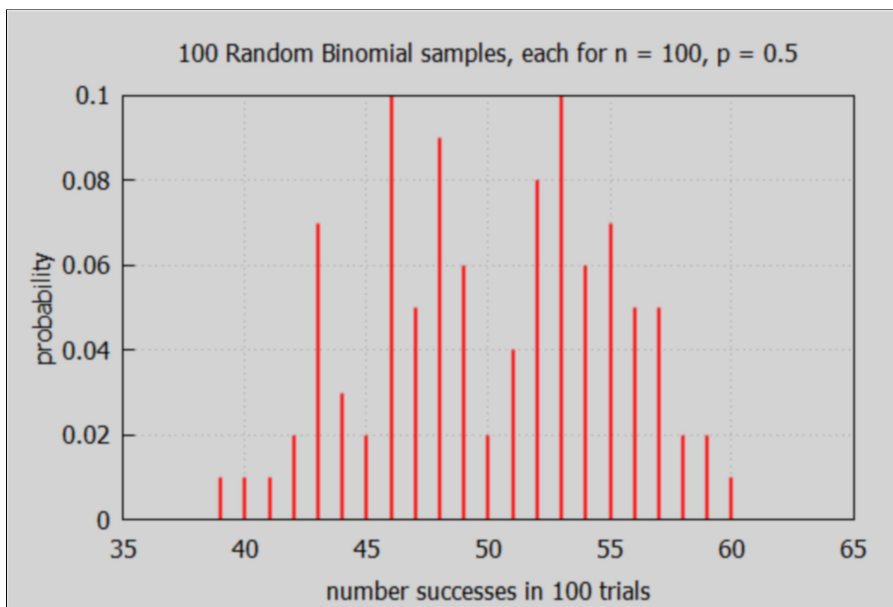
```
(%o144) [[39,0.01],[60,0.01],22]
```

```
(%o145) [[39,0.01],[40,0.01],[41,0.01]]
```

```
(%o146) [[58,0.02],[59,0.02],[60,0.01]]
```

```
(%i147) wxdraw2d ( xrange = [35, 65], yrange = [0, 0.1], points_joined = impulses,
  xlabel = "number successes in 100 trials", ylabel = "probability", grid = true,
  title = "100 Random Binomial samples, each for n = 100, p = 0.5",
  background_color = light_gray,
  line_width = 2, color = red, points (mypoints) )$
```

```
(%t147)
```



When you run this script on your own, your probabilities (vertical axis) will be slightly different, and you can change the vertical range yrange in draw2d, if needed, to see all your points clearly.

8.11.4 Random sample size $m = 1000$, $n = 100$, $p = 0.5$

```
(%i149) mean_binomial (100, 0.5);
      std_binomial (100, 0.5);
```

```
(%o148) 50.0
```

```
(%o149) 5.0
```

```
(%i153) rsample : random_binomial (100, 0.5, 1000)$
      fill (rsample);
      head (rsample);
      tail (rsample);
```

```
(%o151) [53,48,1000]
```

```
(%o152) [53,52,39]
```

```
(%o153) [45,47,48]
```

```
(%i156) [uniqueData, frequencies] : discrete_freq (rsample)$
      uniqueData;
      frequencies;
```

```
(%o155) [34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,
      53,54,55,56,57,58,59,60,61,62,63,64,65]
```

```
(%o156) [2,2,1,2,8,11,8,9,23,29,35,52,54,65,94,77,64,81,66,74,57,
      49,39,29,24,12,18,6,3,2,2,2]
```

Define a list of normalized frequencies, nfrequencies.

```
(%i157) nfrequencies : frequencies/Lsum (frequencies), numer;
```

```
(nfrequencies) [0.002,0.002,0.001,0.002,0.008,0.011,0.008,0.009,0.023,0.029,
      0.035,0.052,0.054,0.065,0.094,0.077,0.064,0.081,0.066,0.074,0.057,
      0.049,0.039,0.029,0.024,0.012,0.018,0.006,0.003,0.002,0.002,0.002]
```

```
(%i158) Lsum (frequencies);
```

```
(%o158) 1000
```

```
(%i159) Lsum (nfrequencies);
```

```
(%o159) 1.0
```

```
(%i160) lmax (nfrequencies);
```

```
(%o160) 0.094
```

```
(%i161) uniqueData;
```

```
(%o161) [34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,
      53,54,55,56,57,58,59,60,61,62,63,64,65]
```

```
(%i162) Lud : length (uniqueData);
(Lud) 32

(%i166) mypoints : makelist ([ uniqueData[j], nfrequencies [j] ], j, 1, Lud)$
fill (mypoints);
head (mypoints);
tail (mypoints);

(%o164) [[34,0.002],[65,0.002],32]
(%o165) [[34,0.002],[35,0.002],[36,0.001]]
(%o166) [[63,0.002],[64,0.002],[65,0.002]]

(%i171) print (" ")$
print (sconcat (Lud,
" normalized frequencies (red) from 1000 random simulations"))$
print (" pdf_binomial (k, 100, 0.5)", " (black) ")$
print (" ")$
wxdraw2d ( xrange = [35, 65], yrange = [0, 0.1], points_joined = impulses,
xlabel = "number successes in 100 trials", ylabel = "probability", grid = true,
title = sconcat (Lud, " normalized frequencies "),
background_color = light_gray,
line_width = 2, color = red, points (mypoints),
color = black, line_width = 2, key_pos = top_left,
key = "pdf binomial (k, 100, 0.5)",
explicit (pdf_binomial (kk, 100, 0.5), kk, 35, 65) )$
```

32 normalized frequencies (red) from 1000 random simulations
pdf_binomial (k, 100, 0.5) (black)

