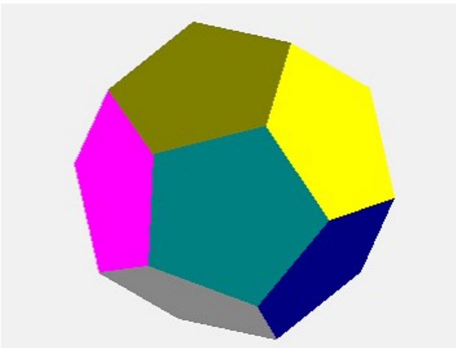
Probability Theory

An introduction and beyond

Chapter 5

Stochastic variables

This is an article from my home page: www.olewitthansen.dk



Ole Witt-Hansen (2017) 1997

Contents

1. What is a stochastic variable?

Let (U,P) be a finite probability field. A stochastic variable is formally defined as a real function defined on U. For the definition set of a function f we shall generally write as $Dm(f)$.
Traditionally stochastic variables are written with capital letters as X, Y and Z. Since the stochastic variables are defined on a finite set, their value set is also finite. There is also a tradition to denote the values that a stochastic variable can take with the letter t.
The value set of a function, we shall denote $Vm(f)$.

If there are n numbers in $Vm(X)$, (value set of X), they are usually denoted: $Vm(X) = \{t_1, t_2, t_3, \ldots, t_n\}$

The concept of a stochastic variable is best illustrated by looking at some examples. We have actually already looked at several examples of stochastic variables without being aware of it.
If the outcomes of an experiment consist of numbers, then the outcomes themselves may be considered as a stochastic variable of the identical function.

Examples

1. When throwing a dice, having the outcome space $U = \{1, 2, 3, 4, 5\ 6\}$ the number of eyes shown may be considered as a (trivial) stochastic variable. Writing $X(1) = 1$, $X(2) = 2$...etc. osv. But you may also define other stochastic variables on U, such as $Y(u) = u-3.5$ or $Z(u) = u^2$.

2. We have already seen another example of a stochastic variable, when we treated the binomial distribution. Repeating n times an experiment having an outcome space U, the outcome space can be written $UxUxUx\ldots xU = U^n$. If we in this outcome space let X denotes the number of "successes", that is, the occurrences of a given event (e.g. the number of times a dice shows 6 eyes) then the value set of X is $Vm(X) = \{0, 1, 2, \ldots, n\}$ and the probability distribution of X is the binomial distribution:

$$P(X = j) = \binom{n}{j} p^j (1-p)^{n-j} \quad ; j = 0, 1, 2, \ldots, n$$

3. Someone is offered a game, where you roll a dice. At each of the outcomes $\{1,2,3,4,5\}$ the player wins 1 \$, but if he throws 6-eyes, he looses 6 \$. If we let $X(u)$ represent the gain then $X(1) = X(2) = X(3) = X(4) = X(5) = 1$ and $X(6) = -6$. The values set of X is $\{1,-6\}$. And the probabilities are $P(X = 1) = 5/6$ and $P(X = -6) = 1/6$.
Is this an equal game or is it less advantageous for the player?

   On a roulette there are 37 fields $\{0,1,2,\ldots 36\}$. If the ball hits the field that you are playing at, you win 36 times your bet. We consider a stochastic variable representing the gain. And let us assume that 23 comes out, then X has the values: $X(0) = X(1) = X(2) = \ldots X(22) = -1$, $X(23) = 35$, $X(24) = X(25) = \ldots = X(36) = -1$. The value set of X is $Vm(X) = \{-1,35\}$. With the probabilities $P(-1) = 36/37$, $P(35) = 1/37$. Is this game an equal game or less advantageous for the player?

2. The mean value of a stochastic variable

The *mean* of a stochastic variable may be conceived as a theoretical model of the average of the occurrence of an event in a random experiment. The mean of a stochastic variable X is written $E(X)$, and often the Greek letter μ (my) is used to denote the theoretical mean. This is most conveniently illustrated with an example.

Example

You may think of a game, where two coins are thrown. If one coin show heads and the other show tails, then the player looses 1 \$. If they both show tails, the player looses 2 \$, but if they both show heads, the player wins 3 \$. We now define a stochastic variable *X*, which is the gain of the game. For simplicity we denote the four outcomes *a, b, c, d*. The outcome space, the probabilities and the stochastic variable are shown in a table.

Outcome u	*a* = (*tail*, *tail*)	*b* = (*tail*,*head*)	*c* = (*head*, *tail*)	*d* = (*head*, *head*)
P(u) =	¼	¼	¼	¼
X(u) =	-2	-1	-1	3

We now contemplate that the games is played a large (n =1000) number of times, and we want to find the collected gain of the n games. Let us therefore assume that (a) has occurred n_a times (b) has occurred n_b times and so on.
Then we have $n_a + n_b + n_c + n_d = n$. The total gain G in the n games is therefore.

$$G = (-2)\cdot n_a + (-1)\cdot n_b + (-1)\cdot n_c + (3)\cdot n_d = X(a)\cdot n_a + X(b)\cdot n_b + X(c)\cdot n_c + X(d)\cdot n_d$$

The gain per game, that is, the average gain, is G/n. Dividing with n in each term in the last expression for G, we find:

$$\frac{G}{n} = X(a)\cdot\frac{n_a}{n} + X(b)\cdot\frac{n_b}{n} + X(c)\cdot\frac{n_c}{n} + X(d)\cdot\frac{n_d}{n}$$

Now the ratios n_a/n , n_b/n, n_c/n, n_d/n are the frequencies (the relative occurrences) of the outcomes *a, b, c , d,* of which the probabilities (in probability theory) are a model for. And they should therefore correspond to the theoretical probabilities *P*(a), *P*(b), *P*(c), *P*(d). So the mean value in probability theory should correspond to the statistical average, which justifies the following definition:

$$E(X) = X(\mathrm{a})\cdot P(\mathrm{a}) + X(\mathrm{b})\cdot P(\mathrm{b}) + X(\mathrm{c})\cdot P(\mathrm{c}) + X(\mathrm{d})\cdot P(\mathrm{d})$$

$$E(X) = (-2)\cdot 1/4 + (-1)\cdot 1/4 + (-1)\cdot 1/4 + 3\cdot 1/4 = -1/4$$

The theoretical average (the mean value) of each game is therefore -25 cent.

Definition:

Let there be given a probability field (U, P), where U is the finite outcome space, and let X be a stochastic variable defined on U. Then the mean value $E(X)$ of the stochastic variable is defined as:

$$E(X) = X(u_1)\cdot P(u_1) + X(u_2)\cdot P(u_2) + X(u_3)\cdot P(u_3) + ... + X(u_n)\cdot P(u_n)$$

Or when written with the summation sign:

$$E(X) = \sum_{u \in U} X(u)\cdot P(u)$$

However, the definition equation of the mean is only rarely uses for practical calculations of $E(X)$.

From the definition it follows trivially some rules for calculation the mean values for stochastic variables. Let X and Y be stochastic variables defined on a finite probability fields (U,P) and let c be an arbitrary real number. The following rules apply.

$$E(X+Y) = E(X) + E(Y) \qquad E(c\cdot X) = c\cdot E(X) \qquad E(c) = c$$

We only settle for proving the first one. The others are equally trivially proved.

$$E(X+Y)=\sum_{u\in U}(X(u)+Y(u))\cdot P(u)=\sum_{u\in U}(X(u))\cdot P(u)+\sum_{u\in U}(Y(u))\cdot P(u)=E(X)+E(Y)$$

The outcomes where a stochastic variable has a constant value is an *event* in the outcome space. If $Vm(X) = \{t_1, t_2, t_3, ..., t_n\}$ is the set of values that *X have*, we can define the event A_k by:

$$A_k = \{u \in U \mid X(u) = t_k\}$$

The probability $P(A_k)$ that this event occur, we shall for obvious reasons write as $P(X = t_k)$.

The sets $A_1, A_2, A_3, ..., A_n$ constitute a class division of the outcome space *U,* since . $U = A_1 \cup A_2 \cup A_3 \cup ... \cup A_n$ and $A_i \cap A_j = \emptyset \quad for \quad i \neq j$, and therefore we may divide the sum, (when we evaluate the mean value) in sums where $X(u)$ have the same value.
Since $X(u) = t_k$ in all outcomes belonging to A_k, we may move it outside the summation sign:

$$E(X)=\sum_{u\in U}X(u)\cdot P(u)=t_1\cdot\sum_{u\in A_1}P(u)+t_2\cdot\sum_{u\in A_2}P(u)+...+t_n\cdot\sum_{u\in A_n}P(u)$$

$$E(X)=\sum_{u\in U}X(u)\cdot P(u)=t_1\cdot P(X=t_1)+t_2\cdot P(X=t_2)+...+t_n\cdot P(X=t_n)$$

We thus find the simpler formula for evaluating $E(X)$, the mean value of a stochastic variable X.

$$E(X)=\sum_{k=1}^{n}t_k P(X=t_k)$$

Examples

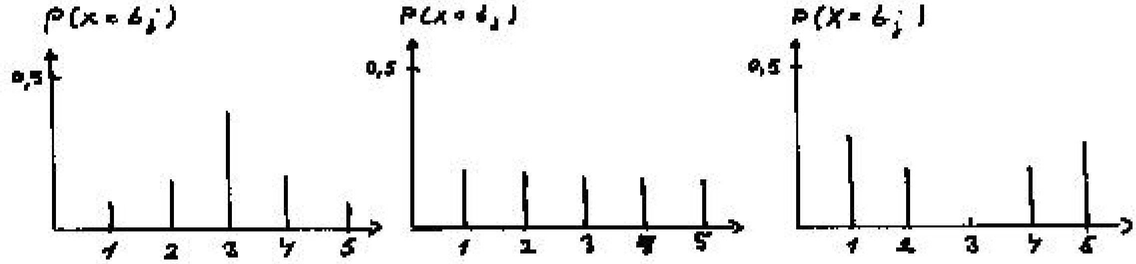
1. In the example above where a dice was thrown, we had $Vm(X) = \{1,-6\}$ and $P(X=1) = 5/6$ and $P(X = 6) = 1/6$. From which we get: $E(X) = 1\cdot 5/6 + (-6)\cdot 1/6 = -1/6$, so the game is not advantageous for the player.

2. In the example with the roulette $Vm(X)=\{-1,35\}$, $P(X = -1) = 36/37$ and $P(X = 35) = 1/37$. From which we get: $E(X) = (-1)\cdot 36/37 + 35\cdot 1/37 = -1/37$. The game is not advantageous for the player.

3. We shall then determine the mean value of the eyes shown, when rolling a dice. The outcome space is $U = \{1, 2, 3, 4, 5, 6\}$ and $P(u) = 1/6$ for all outcomes u. Putting $X = u$ in the formula get for the mean value, we find: $E(X) = 1/6\cdot 1+1/6\cdot 2+…+ 1/6\cdot 6 = 1/6\cdot(1+2+3+4+5+6) = 21/6 = 3½$.

2.1 The variance and the standard deviation

Below are shown the probability distributions for three stochastic variables, all of them having the mean value 3, but nevertheless being quite different from each other.



On way to expressing this is that they have a different *variance*. Loosely we can say that the variance measures how much on an average the individual outcomes differ from the mean value.

This leads to the definition of the *variance* for a stochastic variable.
The mean value is quite easy to comprehend, as a theoretical estimate of the average, but there is not really a similar concept, from daily life experiences, which corresponds to the variance.

If you think of a series of measurements of the same quantity, then due to the uncertainty of the instruments (or other), the different measurements will not come out the same.

What you would like to have is a number that can decide whether the deviations of the measurements from the mean are due only to statistical uncertainties or not.

The *variance* $Var(X)$, or rather the *standard deviation,* which is the square root of $Var(X)$, is such a number.

The variance is most applicable, when the observations gather around the mean value.
If you should want a number indicating the deviation of the observations from the mean value, you could for example choose the mean value of $|\, X(u) - E(X)\,|$, but we do not for various reasons.
Instead the *variance* is defined as the mean value of: $(X(u) - E(X))^2 = (X(u) - \mu)^2$, which is written as $Var(X)$ or $\sigma(X)^2$. The square root of the variance is called the standard deviation: $\sigma(X) = \sqrt{Var(X)}$

$$Var(X) = \sigma^2(X) = E(\,(X - \mu)^2\,)$$

The following important formula follows from the calculation rules for the mean.

$$E(\,(X - \mu)^2\,) = E(X^2 + \mu^2 - 2\cdot X\cdot\mu) = E(X^2) + E(\mu^2) - 2\cdot\mu\cdot E(X) = E(X^2) - \mu^2$$

Thus:

$$Var(X) = \sigma^2(X) = E(\,(X - \mu)^2\,) = E(X^2) - \mu^2$$

In most cases this formula is much easier to use than de definition formula.
$E(X^2)$ is calculated in the same manner as $E(X)$.

$$E(X^2) = \sum_{k=1}^{n} t_k^{\,2} P(X = t_k)$$

Example
We are now able to calculate the mean and the variance for the 3 stick diagrams above, which we denote (1), (2) and (3): The first line are the probabilities.

X_1: $P(1) = 0.1$, $P(2) = 0.2$, $P(3) = 0.4$, $P(4) = 0.2$, $P(5) = 0.1$
$\mu_1 = E(X_1) = 1\cdot 0.1+2\cdot 0.2+3\cdot 0.4+4\cdot 0.2+5\cdot 0.1 = 3.0$
$E(X_1^2) = 1^2\cdot 0.1+2^2\cdot 0.2+3^2\cdot 0.4+4^2\cdot 0.2+5^2\cdot 0.1 = 9.3$
$Var(X_1^2) = E(X_1^2) - \mu_1^2 = 9.3 - 9 = 0.3$

X_2: $P(1) = 0.2$, $P(2) = 0.2$, $P(3) = 0.2$, $P(4) = 0.2$, $P(5) = 0.2$
$\mu_1 = E(X_1) = 1\cdot 0.2+2\cdot 0.2+3\cdot 0.2+4\cdot 0.2+5\cdot 0.2 = 3.0$
$E(X_2^2) = 1^2\cdot 0.2+2^2\cdot 0.2+3^2\cdot 0.2+4^2\cdot 0.2+5^2\cdot 0.2 = 11.0$
$Var(X_2^2) = E(X_2^2) - \mu_2^2 = 11.0 - 9 = 2.0$

X_3: $P(1) = 0.3$, $P(2) = 0.2$, $P(3) = 0.0$, $P(4) = 0.2$, $P(5) = 0.3$
$\mu_3 = E(X_3) = 1\cdot 0.3+2\cdot 0.2+3\cdot 0.0+4\cdot 0.2+5\cdot 0.3 = 3.0$
$E(X_2^2) = 1^2\cdot 0.3+2^2\cdot 0.2+3^2\cdot 0.0+4^2\cdot 0.2+5^2\cdot 0.3 = 11.8$
$Var(X_3^2) = E(X_3^2) - \mu_3^2 = 11.8 - 9 = 2.8$

As you can see, the variances of the 3 distributions are as asserted.

There are yet some calculation rules for the variance and standard deviation, which can be proven in much the same way, as we did for the mean value.

$$Var(c) = 0 \quad , \quad Var(c\cdot X) = c^2\cdot Var(X) \quad , \quad Var(X + c) = Var(X)$$

$$\sigma(c) = 0 \quad , \quad \sigma(c\cdot X) = |c|\cdot\sigma(X) \quad , \quad \sigma(X + c) = \sigma(X)$$

We settle for showing that: $Var(X + c) = Var(X)$. We know that $E(X + c) = \mu + c$
And consequently

$$Var(X + c) = E((X + c) - (\mu + c))^2) = E((X - \mu)^2) = Var(X).$$

However, in general it is *not* valid that: $\sigma(X + Y)^2 = \sigma(X)^2 + \sigma(Y)^2$.

This applies only if $E(Y\cdot X) = E(X)\cdot E(Y)$, and the latter is only the case when X and Y are statistically independent stochastic variables. To account for this, we first have to prove a lemma (a minor theorem).

Lemma
If X and Y are two stochastic variables defined on the same outcome space, and the two events $X = t_i$ and $Y = s_j$ are independent of each other for all $t_i \in Vm(X)$ and $s_j \in Vm(Y)$, that is, $P(X = t_i \wedge Y = s_j) = P(X = t_i)\cdot P(Y = s_j)$, then it follows that $E(X\cdot Y) = E(X)\cdot E(Y)$.

$$E(X\cdot Y) = \sum_{t_i\in Vm(X), s_j\in Vm(Y)} t_i s_j P(X = t_i \wedge Y = s_j) = \sum_{t_i\in Vm(X), s_j\in Vm(Y)} t_i s_j P(X = t_i) P(Y = s_j) =$$
$$\sum_{t_i\in Vm(X)} t_i P(X = t_i) \sum_{s_j\in Vm(Y)} s_j P(Y = s_j) = E(X)\cdot E(Y)$$

Now let X and Y be two *independent* stochastic variables, then we will show that:

$$\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y)$$

$$\begin{aligned}\sigma^2(X+Y) &= E((X+Y)^2) - E(X+Y)^2 = E((X+Y)^2) - (E(X)+E(Y))^2 \\ &= E(X^2+Y^2+2X\cdot Y) - (E(X)^2+E(Y)^2+2E(X)E(Y)) \\ &= E(X^2)+E(Y^2)+2E(X\cdot Y) - E(X)^2 - E(Y)^2 - 2E(X)E(Y) \\ &= E(X^2)+E(Y^2)+2E(X)E(Y) - E(X)^2 - E(Y)^2 - 2E(X)E(Y) = \\ &= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 = \sigma^2(X) + \sigma^2(Y)\end{aligned}$$

Applying this theorem, we may show an important theorem concerning the theoretic mean of a stochastic variable and the average and variance of a series of measurement of this variable.

Let X be a stochastic variable having the mean value μ and the standard deviation σ. If we make n *independent* measurements $\{x_1, x_2, x_3, \ldots, x_n\}$ of X and forming the average:

$$s = \tfrac{1}{n}(x_1 + x_2 + x_3 + \ldots + x_n)$$

Then it corresponds to the stochastic variable

$$X_G = \tfrac{1}{n}(X_1 + X_2 + X_3 + \ldots + X_n)$$

By using the calculation rules for taking the mean, we have:

$$E(X_G) = \tfrac{1}{n}(E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n)) = \tfrac{1}{n} n\mu = \mu$$

This is hardly surprising. For the standard deviation of the mean we find, however:

$$\sigma^2(X_G) = \sigma^2(\tfrac{1}{n}(X_1 + X_2 + X_3 + \ldots + X_n)) = \tfrac{1}{n^2}\sigma^2((X_1 + X_2 + X_3 + \ldots + X_n)) =$$
$$\tfrac{1}{n^2}(\sigma^2(X_1) + \sigma^2(X_2) + \sigma^2(X_3) + \ldots + \sigma^2(X_n)) = \tfrac{1}{n^2}(n\sigma^2) = \tfrac{1}{n}\sigma^2$$

So the standard deviation of the *mean* is equal to the standard deviation of X, divided by the square root of the number of observations.

$$\sigma(X_G) = \frac{1}{\sqrt{n}}\sigma(X).$$

Examples

1. We want to calculate the variance and standard deviation on the number of eyes, when rolling a dice.
   We have already found the mean in an example above that $E(X) = 3.5$
   $\mathrm{Var}(X) = E(X^2) - E(X)^2 = 1^2\cdot 1/6 + 2^2\cdot 1/6 + 3^2\cdot 1/6 + 4^2\cdot 1/6 + 5^2\cdot 1/6 + 6^2\cdot 1/6 - (3.5)^2 = 2.92$
   $\sigma(X) = \sqrt{Var(X)} = 1.7$

3. The mean value and variances of the binomial distribution

We consider an arbitrary binomial distribution with n repetitions of experiment, and we are only interested in whether a certain event A occurs in each experiment. The probability $P(A)$ that the event occurs is called the primary probability, and is set to p.

The mean and the variance are often very important statistical descriptors, when using the binomial distribution.
To obtain an expression for the mean and the variance, we invent a stochastic variable X_i from the definition:
$X_i = 1$, if the primary event occurs the i'th time the experiment is conducted, and $X_i = 0$ otherwise.

For the mean of X_i , we trivially have: $E(X_i) = 1\cdot p + 0\cdot(1\text{-}p) = p$ since the primary event occurs with probability p in the i'th experiment.

If X is the stochastic variable which counts the number of "successes" (the number of times the primary event occur), then obviously: $X = X_1 + X_2 + X_3 + \ldots + X_n$, where X_k is the stochastic variable we introduced above. According to the calculation rules, regarding taking the mean value, we find:

$$E(X) = E(X_1 + X_2 + X_3 + \ldots + X_n) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n) = n\cdot p$$

Examples

1. This should not be surprising at all. If you for example roll a dice a 100 times, you would expect on the average to get 6 eyes $100 \cdot \frac{1}{6} \approx 16$ times.
2. In a family having 5 children ($n = 4$, $p = ½$) the average number of boys is $5\cdot ½ = 2½$.

When we must calculate the variance of a stochastic variable having a binomial distribution, we shall need to calculate the variance of the stochastic variable X_j , introduced above. For this variable applies $X_j^2 = 1^2 = 1$, when "success" in the i'th experiment and otherwise 0, therefore:

$$Var(X_j) = E(X_j^2) - E(X_j)^2 = 1\cdot p - p^2 = p(1 - p)$$

Since $X_1, X_2, X_3, \ldots, X_n$ are independent, we may simply determine the variance of $X = X_1 + X_2 + X_3 + \ldots + X_n$, as the sum of the variances of the individual X_j , and since they all have the same variance, we simply get:

$$Var(X) = n\cdot Var(X_j) = np(1\text{-}p) \quad \text{or} \quad \sigma(X) = \sqrt{np(1-p)}$$

We define the frequency of "succes" in n attempts, corresponding to the stochastic variable $Y = \frac{1}{n} X$. According to the calculation rules for the mean and variance, we find the following expression for the mean of the frequency.

$$E(Y) = \tfrac{1}{n} E(X) = p \quad \text{and} \quad \sigma(Y) = \tfrac{1}{n}\sigma(X) = \tfrac{1}{n}\sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}}$$

Example

If you throw a dice a100 times in the intention of getting 6 eyes, then the standard deviation of the mean becomes:

$$\sigma(X) = \sqrt{100 \cdot \frac{1}{6}(1 - \frac{1}{6})} = \frac{10}{6}\sqrt{5} = 3.73$$

Exercises

The probability distribution for a stochastic variable is shown below

x_i	4	5	6	7	8
$P(X = x_i)$	0.01	0.08	0.26	0.41	0.24

a) Make a stick diagram, and a graph of the distribution function of *X*.
b) Calculate the mean value and the variance of *X*.

1. From a population of 10 elements, of which 4 have defects is taken without replacement a sample of 3 elements. Find the mean of defect elements in the sample. Next find the mean of defect elements if the sample is taken with replacement. Find the standard deviation of the number of defect elements in both cases.

2. In a lottery with 50000 tickets, there is one the first price at 5000 \$, 3 prices at 1000 \$, 10 prices at 100 \$, and 20 prices at 10 \$. When the cost of a ticket I 2 \$, then find the mean value and the standard deviation on the gain on a ticket.

4. The mean and the variance of the Poisson distribution

The Poisson distribution show up, when there is a certain probability λ per unit of time that an event will happen. It may be that a nucleus will decay, or the probability of a call to a (old) telephone central. The probability that the event will happen in an infinitesimal time *dt* is therefore λdt.
(We choose the shortcut of using infinitesimal times *dt*, instead of Δt, and afterwards taking the limit Δt ->0).
The probability that *n* nuclei will decay, (or having *n* calls to a central) during the time *t*, is given by the Poisson distribution:

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

First we shall prove that it meets the condition of being a probability distribution:

$$\sum_{n=0}^{\infty} P_n(t) = 1 \quad \Leftrightarrow \quad \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = 1$$

However

$$\sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} = e^{-\lambda t} e^{\lambda t} = 1$$

Where we have used that $\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$

The mean is then calculated according to its definition: $E(X) = \langle X \rangle = \sum_{i=1}^{n} t_i P(X = t_i)$

$$\langle n \rangle = \sum_{n=0}^{\infty} n P_n(t) = \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

$$= \lambda t e^{-\lambda t} \sum_{n=0}^{\infty} n \frac{(\lambda t)^{n-1}}{n!} = \lambda t e^{-\lambda t} \sum_{n=1}^{\infty} \frac{(\lambda t)^{n-1}}{(n-1)!} = \lambda t e^{-\lambda t} e^{\lambda t} = \lambda t$$

We see that the mean value of the Poisson distribution is $\langle n \rangle = \lambda t$.

The variance $\sigma_n^{\ 2} = E((n - <n>)^2) = E(n^2) - <n>^2$ is found by a similar series of rewritings and reductions. First we do the mean of n^2 : $E(n^2) = \sum_{n=1}^{\infty} n^2 P_n(t)$.

$$\begin{aligned} E(n^2) = <n^2> &= \sum_{n=0}^{\infty} n^2 P_n(t) = \sum_{n=0}^{\infty} n^2 \frac{(\lambda t)^n}{n!} e^{-\lambda t} \\ &= \lambda t \sum_{n=1}^{\infty} n \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} = \lambda t \sum_{k=0}^{\infty} (k+1) \frac{(\lambda t)^k}{k!} e^{-\lambda t} \\ &= \lambda t(\lambda t + 1) = (\lambda t)^2 + \lambda t = <n>^2 + <n> \end{aligned}$$

$$\sigma_n^{\ 2} = E((n - E(n))^2) = E(n^2) - E(n)^2 = (\lambda t)^2 + \lambda t - (\lambda t)^2 = \lambda t$$

Thus

$$\sigma_n^{\ 2} = \lambda t = <n> \quad \text{and} \quad \sigma_n = \sqrt{\lambda t} = \sqrt{<n>}$$

5. Estimating the population mean and population standard deviation

Let us assume that we have a series of independent observations, $x_1, x_2, x_3, \ldots, x_n$, corresponding to the stochastic variables X_1. $X_2, \ldots ,X_n$, having the same mean and variance. Such a series is usually called a population. And let us assume that we do not know the probability distribution of X, and neither the theoretical mean or variance.
We wish to find an estimate of the mean and the variance, based on the experimental measurements. They are then calculated according to the two formulas:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k \qquad and \qquad s^2 = \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \bar{x})^2$$

Corresponding to the stochastic variables:

$$\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k \qquad and \qquad s^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X})^2$$

The choice of these formulas should of course be that: $E(\bar{X}) = \mu$ and $E(s^2) = \sigma^2$, which we shall prove below. The formula for the mean is trivial, since:

$$E(\bar{X}) = \frac{1}{n} \sum_{k=1}^{n} E(X_k) = \frac{1}{n}(n\mu) = \mu$$

The formula for the estimate of the variance is a bit more intriguing, since we to not simply take the mean of $(X_k - \bar{X})^2$, in which case we should divide the sum by n, instead of dividing by n -1.
This has to do with the concept in probability theory called *the degree of freedom.*

Although the variables X_1. $X_2, \ldots X_n$ are considered independent, they are subject to the constraint:

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

So in the formula for the variance, the variables do not have n degrees of freedom but only n -1. This may be seen as a bit abstract, (and it is).

We show the formula for the estimate of the variance by first evaluating $E((X_i - \bar{X})^2)$, remembering that: $\sigma^2 = E(X^2) - \mu^2$ so that $E(X^2) = \sigma^2 + \mu^2$.

$$E((X_i - \bar{X})^2) = E(X_i^2) + E(\bar{X}^2) - 2E(X_i\bar{X}) = \sigma^2 + \mu^2 + \frac{1}{n^2}E((\sum_{i=1}^{n} X_i)^2) - 2\frac{1}{n}E(X_i \sum_{k=1}^{n} X_k)$$

Since X_i and X_k are assumed to be independent, we have:

$$E(X_i X_k) = E(X_i)E(X_k) = \mu^2 \quad for \quad i \neq k \quad og \quad E(X_i^2) = \sigma^2 + \mu^2$$

The sum $(\sum_{i=1}^{n} X_i)$ has n terms, and therefore $(\sum_{i=1}^{n} X_i)^2$ has n^2 terms. Among these, n terms are of the type X_i^2, and therefore $n^2 - n$ of the type $X_i X_k$, where $i \neq k$. Collecting it all we get:

$$E((X_i - \bar{X})^2) = \sigma^2 + \mu^2 + \frac{1}{n^2}((n^2 - n)E(X_i X_k) + nE(X_i^2)) - \frac{2}{n}((n-1)E(X_i X_k) + E(X_i^2))$$

$$E((X_i - \bar{X})^2) = \sigma^2 + \mu^2 + \frac{1}{n^2}((n^2 - n)\mu^2 + n(\sigma^2 + \mu^2)) - \frac{2}{n}((n-1)\mu^2 + \sigma^2 + \mu^2)$$

$$E((X_i - \bar{X})^2) = \sigma^2 + \mu^2 + \frac{1}{n}\sigma^2 + \mu^2 - \frac{2}{n}\sigma^2 - 2\mu^2 = \frac{n-1}{n}\sigma^2$$

From which we finally get:

$$E(s^2) = E\left(\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2\right) = \frac{1}{n-1}\sum_{k=1}^{n} E((X_k - \bar{X})^2) = \frac{1}{n-1} n \frac{n-1}{n}\sigma^2 = \sigma^2$$

As asserted.

6. Chebychev's inequality

Chebychev's inequality is the foundation of what loosely is called: ”The law of large numbers”, and which often erroneously is displayed, as the if the frequency of an event in a large number of repetitions of an experiment converge to the probability of that event.
However this cannot be proven due to the stochastic nature, but it can be proven that the *probability* that the frequency deviates substantially from the probability of that event converges to zero, when the number of repetitions goes to infinity.

Let X is be stochastic variable, having the mean μ and the variance σ^2. We are interested in an estimate of $|X - \mu|$, that is, the deviation from the mean measured in units of σ.
To this purpose we define a stochastic variable Y_c.

$Y_c = c^2$ for $|X - \mu| \geq c$ and otherwise 0.

From the definition of Y_c , it follows immediately: $Y_c \leq (X - \mu)^2$ and therefore:

$$E(Y_c) \leq E((X-\mu)^2) \quad \Leftrightarrow \quad E(Y_c) \leq \sigma^2$$

Thus, when we calculate the mean of Y_c from the definition of the mean value, we have.

$$E(Y_c) = c^2 P(|X-\mu| \geq c) + 0 \cdot P(|X-\mu| < c) = c^2 P(|X-\mu| \geq c)$$

Hold together with: $E(Y_c) \leq \sigma^2$, we find one (crude) version of Chebychev's inequality:

$$c^2 P(|X-\mu| \geq c) \leq \sigma^2 \quad \Leftrightarrow \quad P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Which expresses, that the probability of getting a result which deviates more from the mean than c, is equal to the square of the standard deviation σ^2 divided by c^2.

At the first glance Chebychev's inequality, does not seem very impressive (and it is not), but if we apply it to the frequency $Y = \frac{1}{n}X$, which is the frequency of "succes" in n repetitions of the same experiment, where X is the number of "successes", we have:

$$E(Y) = \tfrac{1}{n}E(X) = p \qquad \text{and} \qquad \sigma(Y) = \sqrt{\frac{p(1-p)}{n}}$$

Then we find:

$$P(|Y-p| \geq c) \leq \frac{\sigma^2}{c^2} = \frac{p(1-p)}{nc^2} \leq \frac{1}{4nc^2}$$

The last inequality is obtained, because $p(1 - p)$ has max ¼ for p = ½

If we choose e.g. c = 0.01, then we see, that the probability that one should have a frequency, that deviates more than 0.01 from the mean p, is less than $\frac{1}{4nc^2}$. If we claim that this probability being less than 0.1, we must solve the inequality:

$$\frac{1}{4nc^2} \leq \frac{1}{10} \Leftrightarrow n \geq \frac{10}{4 \cdot (0.01)^2} \Leftrightarrow n \geq 25000$$

The result is purely of theoretical interest. It is possible to sharpen Chebychev's inequality to obtain a lower estimate of n.

7. The distribution function. Graphic representation

The probability distribution function for a stochastic variable, that is, the probabilities $P(X = t_i)$, where $t_i \in Vm(X)$, (the value set) is also named the frequency function $f(t)$.

$f(t) = P(X = t)$ is only defined for $t \in Vm(X)$. The distribution function $F(t)$ is on the other hand defined for all real numbers.

$$F(t) = P(X \le t) = \sum_{t_i \le t} P(X = t_i)\ ,\ t \in R$$

In a finite probability space , where $Vm(X) = \{t_1, t_2, t_3 \ldots t_n\}$, $F(t)$ will be constant in each of the intervals $t_i \le t < t_{i+1}$,
$F(t) = 0$ for $t < t_1$ and $F(t) = 1$ for $t \ge t_n$..
Below is shown two examples of distribution functions for the binomial distribution.