Expected value and variance

Josemari Sarasola

Statistics for Business

Gizapedia

Introduction

As for data sets, for probability distributions we can give the center, by the mean (usually), and the dispersion or spread, by the variance (usually).

Expected value

The expected value, expectation or mean value is a measure of the center of a distribution. We write it as $\mu = E[X]$ and calculate it as:

- in discrete distributions,

$$\mu = E[X] = \sum_{x \in \Omega} x \cdot p(x)$$

- in continuous distributions,

$$\mu = E[X] = \int_{\Omega} x f(x) dx$$

Expected value and arithmetic mean

We throw a die. What is the expected value?

x	1	2	3	4	5	6	
p(x)	1/6	1/6	1/6	1/6	1/6	1/6	1
xp(x)							$\mu = \frac{21}{6} = 3.5$

Now, we throw the die three times and take these values: 4, 5 and 6. The arithmetic mean is 5.

Itxaropen matematikoa eta batezbestekoa

Why are different μ and $\overline{x}$? Both are about the same concept (the center) but they are calculated from different points of view.

μ, expected value	$\overline{x}$, arithmetic mean
• constant	• variable, as data are also variable
• from the probability distribution	• from data
• ideal	• concrete
• in the long term $\overline{x}$ gets closer to μ	• in the short term (with few data) different from μ
• it's a parameter	• used as an estimator for μ
• often unknown	• we can always calculate it

Properties of the expected value

(1) *related to the sum of rv-s,*

$$X_1, X_2, \ldots, X_n \; rv - s$$

$$\mathbf{X} = X_1 + X_2 + \ldots + X_n \rightarrow E[\mathbf{X}] = E[X_1] + E[X_2] + \ldots + E[X_n]$$

E.g., the expected value of the points drawn with 10 dice is the sum of expected values of the points of 10 dice: $3.5 + \cdots + 3.5 = 35$

(2) *related to linear transformations,*

$$X \; rv$$

$$Y = a + bX \rightarrow E[Y] = a + bE[X]$$

E.g., for playing with a die they pay us 10€ per point and a fixed amount of 20€ : $money = 20 + 10X \rightarrow E[money] = 20 + 10E[X] = 20 + 10 \times 3.5 = 55$

Moments

They are just formulae for distributions, but sometimes they do mean something: e.g., μ (the expected value) is a moment.

Raw moments: about 0 (or origin)

We define r-th moment about zero as:

$$\alpha_r = E[X^r]$$

- for discrete dist., $\alpha_r = \sum_{x \in \Omega} x^r p(x)$;
- for continuous dist., $\alpha_r = \int_\Omega x^r f(x)$.

Remark: just as to calculate μ, but instead of x, we put x^r.

E.g.:

- 0-th moment about origin is. Proof: $\alpha_0 = \sum x^0 p(x) = \sum p(x) = 1$;
- 1-st moment about origin is μ. Proof: $\alpha_1 = \sum x^1 p(x) = \sum xp(x) = \mu$;

Moments: about the center

We define r-th central moment as:

$$\mu_r = E[(X-\mu)^r]$$

- for discrete dist., $\mu_r = \sum_{x\in\Omega}(x-\mu)^r p(x)$;
- for continuous dist., $\mu_r = \int_\Omega (x-\mu)^r f(x)$.

Remark: just as to calculate μ, but instead of x, we put $(x-\mu)^r$.

E.g.:

- 0-th central moment is 1. Proof: $\mu_0 = \sum(x-\mu)^0 p(x) = \sum p(x) = 1$;
- 1-st central moment is 0. Proof: $\mu_1 = \sum(x-\mu)^1 p(x) = \sum xp(x) - \mu\sum p(x) = \mu - \mu = 0$;

Variance

Variance is the commonly used measure of dispersion for distributions.It's the 2-nd central moment (σ: lower case sigma):

$$\sigma_X^2 = var[x] = \mu_2 = E[(X - \mu)^2]$$

We can also calculate as (much easier):

$$\sigma_X^2 = \alpha_2 - \alpha_1^2$$

(namely, 2-nd moment about the origin minus μ squared)

Variance

- We must distinguish σ_X^2 (variance of probability distributions) eta s_X^2 (variance from data).
- We write measures for probability distributions with greek letters, and with latin letters for measures for data
- *Standard deviation* is the root square of the variance:

$$\sigma_X = \sqrt{\sigma_X^2}$$

Properties of the variance

(1) *related to the sum of rv-s,*

$$X_1, X_2, \ldots, X_n \; rv - s$$

independent to each other, $\rightarrow$

$$\mathbf{X} = X_1 + X_2 + \ldots + X_n \rightarrow \sigma^2_{\mathbf{X}} = \sigma^2_{X_1} + \sigma^2_{X_2} + \ldots + \sigma^2_{X_n}$$

(2) *related to linear transformations,*

$$X \; rv$$

$$Y = a + bX \rightarrow \sigma^2_Y = b^2 \sigma^2_X$$

The expected value as decision criterion

Probability or expected value?

In A production process, the probability of a faulty item is 0.1; in B process, it's 0.05. Which is the optimal process? We should take the process with the smallest probability of a item being faulty. So, the answer is B.

The expected value as decision criterion

Probability or expected value?

We get more information about the latter: price of the items is 4 euros by unit. Cost per unit for A process is 2; and 3 for B process. So, now we can calculate the average return per unit:

x_A	$p(X_A)$	$x_A p(X_A)$	x_B	$p(X_B)$	$x_B p(X_B)$
-2	0.1	-0.2	-3	0.05	-0.15
2	0.9	1.8	1	0.95	0.9
		1.6			0.8

And so, A process would be the best one, given that information.

The expected value as decision criterion

Probability or expected value?

So, given the dilemma, which is the best process? We could conclude that it's A, because that optimal came from a bigger amount of information (probabilities, returns, and expected value), but that's not right. The optimal to be taken depends on our goal, on our own decision criterion. It's we who have to take the decision, and so the criterion, that finally depends on our goals: probability or expected value.

The expected value as decision criterion

Short term vs long term

The expected value holds in the long term (for many data), so it's for the long term (the same decision is to be taken many times) that is a good criterion. But in the short term the thing is different.

The expected value as decision criterion

Short term vs long term

We have to choose between two head or tails games, A and B. Which is the best in the long term? And in the short term (concretely, just to play once)?

x_A	$p(X_A)$	$x_A p(X_A)$	x_B	$p(X_B)$	$x_B p(X_B)$
-10	0.5	-5	-1000	0.5	-500
20	0.5	10	3000	0.5	1500
		5			1000

The expected value as decision criterion

Short term vs long term

In the long term B is better. But would you play it just once? In the short term things are not so clear.

The expected value as decision criterion

In the short term: expected value and risk

In the short term, we must take into account both the expected value (the bigger, the better) and the risk (the smaller, the better). Risk is a short term criterion, since in the short term we may be lucky or unlucky. In the long term risk disappears, as almost surely bad outcomes are balanced with good outcomes. Generally, we play avoiding risk (that is, we have *risk aversion*) in the short term.

The expected value as decision criterion

In the short term: expected value and risk

In economics, we usuallly measure risk by means of the variance. The bigger the variance is, the more dispersed are the outcomes (it's possible to have very big and very small outcomes) and the bigger the risk will be.
Variance of B is bigger than variance of A, so B game is riskier.
So, in the short term, we will choose *games* (so to speak) with smaller variance and bigger expectation.
In finance, risk usually refers to stock values changing frequently and very much. In that context, we call risk *volatility*.

The expected value as decision criterion

Utility functions

In the short term, if an investment has smaller variance and bigger expectation than other one, there's no problem: we choose the former. But what if an investment has bigger expectation and bigger variance (or smaller variance and smaller expectation)? Then we have a dilemma. In order to solve the dilemma and take a decision, we define an utility function, depending on both expectation and variance (or deviation) (and may be other variables) and calculate the utility for each of the alternatives. The alternative with the biggest utility will be the best one.

The expected value as decision criterion

Utility functions

E.g., this is an utility function:

$$U = \frac{\mu}{\sigma}$$

But we can define many others. But all utility functions must hold some conditions, namely *an axiom system*. These are some of those axioms:

- the bigger μ is, the bigger the utility is;
- the bigger σ is, the smaller the utility is.

In the short term (remember!), we will choose always the alternative with the biggest utility.

The expected value as decision criterion

Newsvendor problem

It's common revenue management problem. A newsvendor has to decide how many newspapers he must buy in order to resell them. He buys each paper at 1 euro price, and sells it at a 4 euro price. Not sold papers are thrown away. The demand follows this distribution:

x	1	2	3	4
$p(x)$	0.1	0.3	0.4	0.2

How many papers should he buy in order to maximize the expected return?

The expected value as decision criterion

Newsvendor problem

$$s : sales; \; r : return$$

Buy 1 paper

s	r	$p(r)$	$rp(r)$
1	4-1=3	1	3
			3

Buy 2 papers

s	r	$p(r)$	$rp(r)$
1	$1 \times 4 - 2 \times 1 = 2$	0.1	0.2
2	$2 \times 4 - 2 \times 1 = 6$	0.9	5.4
			5.6

The expected value as decision criterion

Newsvendor problem

Buy 3 papers

s	r	$p(r)$	$rp(r)$
1	$1 \times 4 - 3 \times 1 = 1$	0.1	0.1
2	$2 \times 4 - 3 \times 1 = 5$	0.3	1.5
3	$3 \times 4 - 3 \times 1 = 9$	0.6	5.4
			7

Buy 4 papers

s	r	$p(r)$	$rp(r)$
1	$1 \times 4 - 4 \times 1 = 0$	0.1	0
2	$2 \times 4 - 4 \times 1 = 4$	0.3	1.2
3	$3 \times 4 - 4 \times 1 = 8$	0.4	3.2
4	$4 \times 4 - 4 \times 1 = 12$	0.2	2.4
			6.8

The expected value as decision criterion

Newsvendor problem

So the newsvendor should buy 3 newspapers, in order to maximize the expected return.

The expected value as decision criterion

Newsvendor problem: continuous domain

A business man wants to buy some apple juice to resell it. He buys at a 1 euro price per liter, and sells it at a 4 euro price per liter. Not sold juice is deprecated. The demand follows this distribution:

$$f(x) = \frac{1}{4};\ 1 < x < 5$$

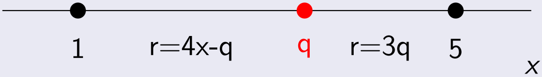
How much juice should he buy?

The expected value as decision criterion

Newsvendor problem: continuous domain

r return is calculated in this manner, given q (amount of juice to be purchased in advance) and x (demand):

- $x < q \rightarrow r = 4x - 1 \times q = 4x - q$
- $x > q \rightarrow r = (4 - 1)q = 3q$



The expected value as decision criterion

Newsvendor problem: continuous domain

Now we calculate the expected value of r:

$$E[m] = \int_1^q \underbrace{(4x - q)}_{r} \overbrace{\frac{1}{4}}^{f(x)} dx + \int_q^5 \underbrace{3q}_{r} \overbrace{\frac{1}{4}}^{f(x)} dx$$

$$= \int_1^q x - \frac{q}{4} dx + \int_q^5 \frac{3q}{4} dx$$

$$= \left[\frac{x^2}{2} - \frac{qx}{4}\right]_1^q + \left[\frac{3qx}{4}\right]_q^5 = 4q - \frac{q^2}{2} - \frac{1}{2}$$

The expected value as decision criterion

Newsvendor problem: continuous domain

We want to calculate q maximizing $E[r]$ so we take the derivative with respect to q, equal it to 0 and solve it for q:

$$\frac{dE[r]}{dq} = 0 \rightarrow 4 - \frac{2q}{2} = 0 \rightarrow q^* = 4\ l$$

Chebyshev's inequality

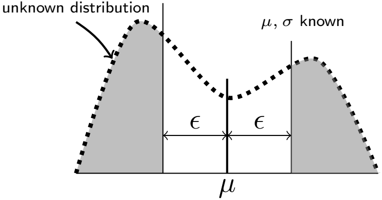
If we don't know the probability distribution of a rv, we cannot calculate probabilities. But known μ and σ, we can approximate them by means of the Chebyshev's inequality.

Chebyshev's inequality

We have two formula for the inequality: the outer formula and the inner formula.

- outer formula:

$$P\left(|X-\mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2}$$



- inner formula:

$$P\left(|X-\mu|<\epsilon\right)\geq 1-\frac{\sigma^2}{\epsilon^2}$$