Statistics

2023 Lectures
Part 4 - Expectations

Institute of Economic Studies
Faculty of Social Sciences
Charles University

Motivation for expected value

Example 29: Consider a random variable X defined on some sample space S. An experiment consists of a random selection of a point $s \in S$ and $X(s)$ can be interpreted as the gain (or loss) of a gambler.

If the experiment is repeated n times, then an outcome $s^{(n)} = (s_1, \ldots, s_n), s_i \in S, i = 1, \ldots, n$ leads to gambler's accumulated gain $X(s_1) + X(s_2) + \cdots + X(s_n)$.

The average gain per gamble then fluctuates less and less as n becomes larger and tends to stabilize at a value which we call expected value of X.

- expected value of a rv can be viewed as the balance point of distribution of probability on a real line

Expectation of discrete rv's

Definition 14: If X is a discrete random variable that assumes values $x_1, x_2, \ldots$ with probabilities $P(X = x_i), i = 1, 2, \ldots,$ then the expected value of X is defined as

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$$

provided

$$E|X| = \sum_{i=1}^{\infty} |x_i| P(X = x_i) < \infty.$$

Example 30: Let A occurs with probability p and you win USD 1 if A occurs otherwise you loose USD 1. Determine expected value of the game.

Example 31: Let A occurs with probability p and you win USD 1 if A occurs otherwise you loose USD 0. Determine expected value of the game.

Examples

Example 32: (Binomial distribution)

$$E(X) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = np.$$

Example 33: (Geometric distribution)

$$E(X) = \sum_{k=0}^{\infty} k \cdot (1-p)^k p = \frac{1-p}{p}.$$

Example 34: (Petersburg paradox)

A fair coin is tossed repeatedly until a head occurs. If the first head occurs on the kth toss then you win 2^k dollars. What is the amount you should be willing to pay to participate in the game?

I will play for fee less then the expected value of the game:

$$E(X) = 2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{2^2} + \cdots = +\infty!$$

$EX = \pm\infty$ vs. EX does not exist

If the condition in Definition 14 is not met then we can consider (for the discrete case)

$$U^+ = \sum(\max(x_i, 0)P(X = x_i))$$
$$U^- = -\sum(\min(x_i, 0)P(X = x_i))$$

Then $E|X| = U^+ + U^-$ and

$$|E(X)| < \infty \Leftrightarrow U^+ < \infty, U^- < \infty$$
$$E(X) = +\infty \Leftrightarrow U^+ = \infty, U^- < \infty$$
$$E(X) = -\infty \Leftrightarrow U^+ < \infty, U^- = \infty$$
$$E(X) \text{ does not exist} \Leftrightarrow U^+ = \infty, U^- = \infty$$

Expectation of continuous rv's

Definition 15: If X is a continuous random variable with the density f_X, then the expected value of X is defined as

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) \mathrm{d}x$$

provided that

$$E|X| = \int_{-\infty}^{\infty} |x| \cdot f_X(x) \mathrm{d}x < \infty.$$

Example 35: Uniform distribution on $[a, b]$:

$$E(X) = \int_a^b \frac{x}{b-a} \mathrm{d}x = \frac{a+b}{2}.$$

More examples

Example 36: (Exponential distribution)

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} \mathrm{d}x = \frac{1}{\lambda}.$$

Example 37: (Normal distribution)

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + z\sigma) e^{-\frac{z^2}{2}} \mathrm{d}z = \\ &= \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} \mathrm{d}z + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} \mathrm{d}z = \\ &= \mu. \end{aligned}$$

Expectation and functions of random variables

Theorem 10: (without proof) Let X be a random variable. In the discrete case, if g is a real function such that $\sum_i |g(x_i)| P(X = x_i) < \infty$ then

$$E(g(X)) = \sum_i g(x_i) P(X = x_i).$$

In the continuous case if $\int_{-\infty}^{\infty} |g(x)| \cdot f_X(x) \mathrm{d}x < \infty$ then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \mathrm{d}x.$$

Properties of the expected value

Theorem 11: (parts c) and d) without proof)

a) If $E(X)$ exists then for all $a, b \in \mathbb{R}$

$$E(aX + b) = aE(X) + b.$$

b) If $X \geq 0$ then $E(X) \geq 0$.

c) For any random variable X

$$|E(X)| \leq E|X|.$$

d) If $|X| < Y$ and $E(Y) < \infty$ then also $E(X) < \infty$.

Expectation and functions of random variables

Theorem 12: (without proof) Let X, Y be random variables. In the discrete case, if $h(\cdot,\cdot)$ is a real function such that $\sum_{i,j} |h(x_i, y_j)| p_{ij} < \infty$ then

$$E(h(X,Y)) = \sum_i \sum_j h(x_i, y_j) p_{ij} = \sum_j \sum_i h(x_i, y_j) p_{ij}.$$

In the continuous case if

$$\iint_{\mathbb{R}^2} |h(x,y)| \cdot f(x,y) \mathrm{d}(x,y) < \infty$$

then

$$\begin{aligned} E(h(X,Y)) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h(x,y) \cdot f(x,y) \mathrm{d}x \right) \mathrm{d}y \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h(x,y) \cdot f(x,y) \mathrm{d}y \right) \mathrm{d}x. \end{aligned}$$

Expectation and functions of random variables

Theorem 13: Assume that $E|X| < \infty$ and $E|Y| < \infty$. Then

$$E(X+Y) = E(X) + E(Y).$$

Consequently,

$$E\left(\sum_i (a_i X_i) + b\right) = \sum_i (a_i E(X_i)) + b$$

and, applying also Theorem 11 b), if $Y \leq Z$ then $E(Y) \leq E(Z)$.

Theorem 14: If X and Y are independent random variables and g, h real valued functions such that $E|g(X)h(Y)| < \infty$ then

$$E(g(X)h(Y)) = E(g(X)) \cdot E(h(Y)).$$

Moments of random variables

Definition 16: For any random variable X, the expectation of X^n (if exists) is called the nth ordinary moment of X

$$m_n = E(X^n).$$

An absolute moment of order n is defined as

$$\beta_n = E|X^n|.$$

Example 38: Poisson distribution with parameter λ:

$$m_1 = E(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda.$$

$$m_2 = E(X^2) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 + \lambda.$$

Existence of moments and Liapunov inequality

Example 39: A degenerate random variable is a random variable taking a constant value. I.e., $X = c$ with $P(X = c) = 1$.

$$E(X) = c \cdot 1 = c, \;\; E(X^2) = c^2 \cdot 1 = c^2.$$

Thus, $m_n = c^n$.

Definition 17: Ordinary moments of $X - E(X)$ are called central moments of X

$$\gamma_n = E(X - m_1)^n.$$

Theorem 15: (without proof) If an absolute moment of order m exists then all moments (ordinary, absolute, central) of orders $n \leq m$ exist.

Theorem 16: (without proof) (Liapunov inequality)
If $0 < \alpha < \beta < \infty$ then

$$[E|X|^\alpha]^{\frac{1}{\alpha}} \leq \left[E|X|^\beta\right]^{\frac{1}{\beta}}.$$

Moment generating function

Definition 18: The function of real variable t defined as

$$m_X(t) = E\left(e^{tX}\right), t \in \mathbb{R},$$

is called a moment generating function (mgf).

- For any random variable, its mgf exists for $t = 0$.

Example 40: (Poisson distribution)

$$m_X(t) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda(e^t - 1)}, \quad t \in \mathbb{R}.$$

Example 41: (Exponential distribution)

$$m_X(t) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} \mathrm{d}x = \frac{\lambda}{\lambda - t}, \quad \text{if } t < \lambda.$$

Theorem 17: (without proof) Let $m_X(t) < \infty$, $|t| < t_0$ for some $t_0 > 0$. Then $E|X|^r < \infty$ for all $r > 0$ and

$$m_n = \left. \frac{\partial^n m_X(t)}{\partial t^n} \right|_{t=0} \left(= m_X^{(n)}(0) \right).$$

Example 42: (Binomial distribution)

$$m_X(t) = \sum_{k=0}^{n} \left[e^{tk} \binom{n}{k} p^k q^{n-k} \right] = (pe^t + q)^n, \quad t \in \mathbb{R}.$$

We also have

$$m'_X(t) = n(pe^t + q)^{n-1} pe^t,$$
$$m''_X(t) = n(n-1)(pe^t + q)^{n-2}(pe^t)^2 + n(pe^t + q)^{n-1} pe^t.$$

Thus $E(X) = m'_X(0) = np$ and $E(X^2) = n(n-1)p^2 + np$.

Properties of mgf's

Theorem 18: If X is a random variable with $m_X(t)$ and $a, b \in \mathbb{R}$ then $Y = aX + b$ has mgf

$$m_Y(t) = e^{bt} m_X(at), \;\; t \in \mathbb{R}.$$

Example 43: $Z \sim N(0,1), X = \mu + \sigma Z, X \sim N(\mu, \sigma^2)$

$$m_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz} e^{-\frac{z^2}{2}} \mathrm{d}z = e^{\frac{t^2}{2}}, t \in \mathbb{R},$$
$$m_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}, t \in \mathbb{R}.$$

Theorem 19: If X and Y are independent random variables with moment generating functions $m_X(t), t \in \mathbb{R}$ and $m_Y(t), t \in \mathbb{R}$ then mgf of $X + Y$ is

$$m_{X+Y}(t) = m_X(t) \cdot m_Y(t), \;\; t \in \mathbb{R}.$$

Properties of mgf's

Theorem 20: (without proof) If X and Y are two random variables such that their moment generating functions $m_X(t), t \in \mathbb{R}$, and $m_Y(t), t \in \mathbb{R}$, coincide in some neighborhood of $t = 0$, then X and Y have the same distribution.

Example 44: Let X and Y be independent and have Poisson distribution with means λ_1 and λ_2. Then

$$m_X(t) = e^{\lambda_1(e^t - 1)}, t \in \mathbb{R}, \;\; m_Y(t) = e^{\lambda_2(e^t - 1)}, t \in \mathbb{R}.$$

By Theorem 19

$$m_{X+Y}(t) = m_X(t)m_Y(t) = e^{(\lambda_1 + \lambda_2)(e^t - 1)}, t \in \mathbb{R},$$

which corresponds to mgf of a Poisson distribution with mean $\lambda_1 + \lambda_2$. By Theorem 20, $X + Y \sim POI(\lambda_1 + \lambda_2)$.

Variance and standard deviation

Definition 19: If $E(X^2) < \infty$ then the second central moment of X

$$VarX = E(X - EX)^2$$

is called the variance of X. Its positive square root is called the standard deviation of X, often denoted by σ.

- since it is expectation of $(X - EX)^2 \geq 0$, variance is always nonnegative
- $$VarX = E(X^2 - 2X \cdot EX + (EX)^2) = E(X^2) - (EX)^2.$$

Theorem 21: If $VarX$ exists then

$$Var(aX + b) = a^2 VarX.$$

Examples

Example 45: $X \sim U[a,b]$
Then $E(X) = \frac{a+b}{2}$ and $E(X^2) = \frac{1}{3}(b^2 + ab + a^2)$. Hence

$$VarX = \frac{1}{12}(b-a)^2.$$

Example 46: $X \sim BIN(n,p)$
Then $E(X) = np$ and $E(X^2) = n(n-1)p^2 + np$. Hence

$$VarX = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p).$$

Interpretation of variance

- Discrete case: $VarX = \sum_i (x_i - E(X))^2 P(X = x_i)$. So, it is small if each term in the sum is small, i.e., when x_i's with large difference $|x_i - E(X)|$ have low probabilities. That is, when the values of X are concentrated closely to $E(X)$.

Theorem 22: The mean square deviation from ξ (i.e. $E(X-\xi)^2$) is minimized at $\xi = E(X)$ and its minimal value equals $VarX$.

Theorem 23: (without proof) The mean absolute deviation from ξ (i.e. $E|X-\xi|$) is minimized at $\xi = m$, where m is median of X.

Definition 20: Let X be a random variable with $E(X) = \mu$, $VarX = \sigma^2 > 0$ and such that the third central moment $\gamma_3 = E(X - \mu)^3$ exists. The ratio $\frac{\gamma_3}{\sigma^3}$ is called the coefficient of skewness.

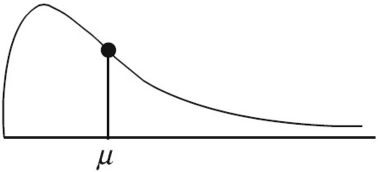
Definition 21: Let X be a random variable with $E(X) = \mu$, $VarX = \sigma^2 > 0$ and such that $\gamma_4 = E(X - \mu)^4$ exists. The ratio $\frac{\gamma_4}{\sigma^4}$ is called the coefficient of kurtosis.

Example 47: Find skewness and kurtosis of the following distributions:

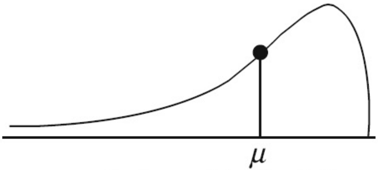
i) $U[0, 1]$,

ii) $BIN(1, p)$.

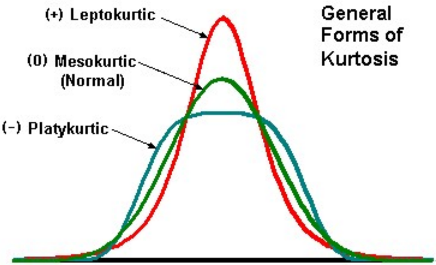
Skewness and kurtosis



$\mu_3 > 0$, skewed to the right



$\mu_3 < 0$, skewed to the left



Chebyshev inequality

Theorem 24: (Chebyshev inequality)
For any rv X with $VarX = \sigma^2 < \infty$, for every $\varepsilon > 0$

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Example 48: We toss a coin 20 times and want to estimate the probability that number of heads will deviate from 10 by 3 or more.
$E(X) = np = 10, VarX = \sigma^2 = np(1-p) = 5$ and

$$P(|X - 10| \geq 3) \leq \frac{5}{9} = 0.5555.$$

However, exact probability equals 0.2632!

- Chebyshev inequality...not impressive, but valid for all rv's.

Beyond Chebyshev inequality

- Among the most important applications are the so called laws of large numbers (cf. Chapter Random samples).

Example 49: Consider the binomial rv S_n (the number of successes in n trials). When p is the probability of a success then

$$E(S_n/n) = p, \;\; Var(S_n/n) = pq/n.$$

Chebychev inequality gives

$$P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leq \frac{pq}{n\varepsilon^2}.$$

For every $\varepsilon > 0$, letting $n \to \infty$:

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) = 0.$$

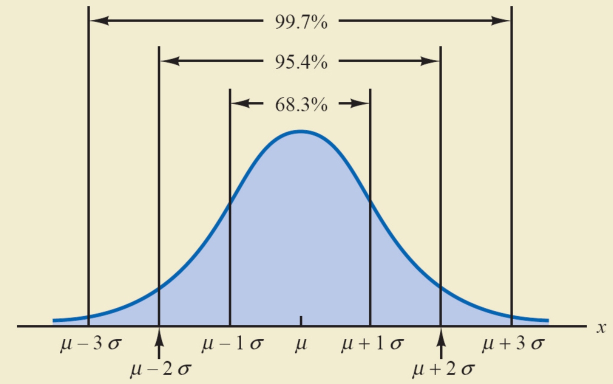
This explains why empirical frequency of an event approaches, as number of trials increases, the probability of the event.

3-σ rule for normal distribution

- Consider $Z \sim N(0, 1)$ then

$$P(|Z| > z) = 2(1 - \Phi(z)).$$

- since $\Phi(3) = 0.9987$ then for $X \sim N(\mu, \sigma^2)$ we have $P(|X - \mu| > 3\sigma) = 0.0026$.



Covariance

Consider random variables X and Y. Then

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

Definition 22: The quantity

$$Cov(X,Y) = E(XY) - E(X) \cdot E(Y)$$

is called the covariance of random variables X and Y, provided $E(X^2) < \infty$ and $E(Y^2) < \infty$ (i.e., $E|XY| < \infty$).

Example 50:

$$Cov(X,X) = E(X^2) - (E(X))^2 = VarX.$$

Theorem 25: If $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are two sets of random variables with finite second moments then

$$Cov\left(\sum_{i=1}^{n} a_i X_i + c, \sum_{j=1}^{m} b_j Y_j + d\right) = \sum_i \sum_j a_i b_j Cov(X_i, Y_j).$$

Theorem 26: Let $X_1, \ldots, X_n$ be random variables with $E(X_i^2) < \infty, i = 1, \ldots, n$. Then

$$Var(a_1 X_1 + \ldots a_n X_n + b) = \sum_{j=1}^{n} a_j^2 Var X_j + 2 \sum\sum_{i<j} a_i a_j Cov(X_i, X_j).$$

Example 51: Let $a_1 = 1, a_2 = -1$ and $b = 0$. Then

$$Var(X - Y) = VarX + VarY - 2Cov(X, Y).$$

Independent vs. uncorrelated rv's

Definition 23: We call rv's X and Y uncorrelated if

$$Cov(X, Y) = 0.$$

- covariance can be understood as a signed measure of the variation of X relative to Y.
- uncorrelated X and Y can be understood as rv's with no linear dependence, but not independent rv's!

Theorem 27: Let $X_1, \ldots, X_n$ be pairwise uncorrelated random variables with $E(X_i^2) < \infty, i = 1, \ldots, n$. Then

$$Var(a_1X_1 + \ldots a_nX_n + b) = \sum_{j=1}^{n} a_j^2 VarX_j.$$

In particular, it holds when $X_1, \ldots, X_n$ are independent.

Averaging of random variables

Example 52: If $X_1, \ldots, X_n$ are independent random variables with the same distribution, then their average

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

has variance

$$Var\bar{X}_n = \frac{VarX_1}{n}.$$

Thus, averaging decreases the variability (in terms of variance) by the factor $1/n$.

Example 53:

Effect of averaging on variability long time before mathematically explained: a 16th century law that defined a length of one foot in Frankfurt

Stand at the door of a church on a Sunday and bid 16 men to stop, tall ones and small ones, as they happen to pass out when the service is finished then make them put their left feet one behind the other, and the length thus obtained shall be a right and lawful rood to measure and survey the land with, and the 16th part of it shall be the right and lawful foot.

i) How much this procedure cuts down the variability (measured by the standard deviation) of the length of feet?

ii) Assume that a man's shoe has an average length of 1 foot and $\sigma = 0.1$ foot. Find the (approximate) probability that the mean of 16 lengths of men's shoes exceed 1 foot by more than 1 inch. Assume normality.

Correlation coefficient

Definition 24: Let X and Y be two random variables with finite second moments. Then for $VarX > 0$ and $VarY > 0$

$$Corr(X, Y) = \rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{VarX \cdot VarY}}$$

is called the coefficient of correlation between X and Y.

Example 54: X and Y discrete rv's has joint values $(-2, 4), (-1, 1), (0, 0), (1, 1), (2, 4)$, each with probability $1/5$. Then $E(X) = 0, E(Y) = 2, E(XY) = 0, Cov(X, Y) = 0, \rho_{X,Y} = 0$. Thus they are not independent, yet they are uncorrelated.

Example 55: X, Y continuous rv's with joint uniform distribution on the square with vertices (-1,0), (0, -1), (1, 0) and (0,1). Then $E(X) = 0, E(Y) = 0, E(XY) = 0, Cov(X, Y) = 0, \rho_{X,Y} = 0$. Yet, X and Y are not independent!

Theorem 28: (Schwarz inequality)
For any random variable X and Y

$$(E(XY))^2 \leq E(X^2)E(Y^2).$$

Theorem 29: The coefficient of correlation $\rho_{X,Y}$ satisfies

a)

$$|\rho_{X,Y}| \leq 1$$

with $|\rho_{X,Y}| = 1$ if and only if there is a constant $a \neq 0$ and b such that $P(Y = aX + b) = 1$.

b) for any a, b, c, d with $ac \neq 0$

$$\rho_{aX+b,cY+d} = \begin{cases} \rho_{X,Y}, & \text{if } ac > 0; \\ -\rho_{X,Y}, & \text{if } ac < 0. \end{cases}$$