Probability and Inferential Statistics

Lecture SS 16

Expectation, Variance, Covariance and Correlation

Prof. Dr. Rolf Steyer

Expectation of a Discrete Random Variable

The expectation of a random variable is the most important parameter to describe the location of a random variable. It may also be called the "true mean". It is a theoretical parameter estimated by a sample mean. The expectation is also used in the definitions of the variance, the covariance, and the correlation. The values of the random variable considered have to be numbers.

Definition 1. Let X be a discrete real-valued random variable on a probability space$(\Omega, \mathscr{A}, P)$ with a finite number of values $x_1, \ldots, x_n$, and let $P(X{=}x_i)$, $i = 1, \ldots, n$, denote the probability that X takes on the value x_i. Then the *expectation* $E(X)$ *of* X is defined by

$$E(X) \ := \ \sum_{i=1}^{n} x_i \cdot P(X{=}x_i). \tag{1}$$

If X is discrete with a countable number of values $x_1, x_2, \ldots$, then the expectation of X is defined by

$$E(X) \ := \ \sum_{i=1}^{\infty} x_i \cdot P(X{=}x_i) \ := \ \lim_{n\to\infty} \sum_{i=1}^{n} x_i \cdot P(X{=}x_i). \tag{2}$$

Example: Flipping a Single Coin

Example

Let $(\Omega, \mathscr{A}, P)$ represent the random experiment of flipping a coin. Then $\Omega = \{h, t\}$. Now suppose that $X\colon (\Omega, \mathscr{A}) \to (\Omega', \mathscr{A}')$ is defined by

$$X(\omega) = \begin{cases} 1, & \text{if } \omega = h, \\ 0, & \text{otherwise,} \end{cases}$$

with $\Omega' = \{0, 1\}$ and $\mathscr{A}' = \mathscr{P}(\Omega')$. Then Equation (40) yields

$$E(X) \ = \ \sum_{i=1}^{2} x_i \cdot P(X{=}x_i) \ = \ 0 \cdot P(X{=}0) + 1 \cdot P(X{=}1) \ = \ P(X{=}1). \tag{3}$$

Hence, for a dichotomous random variable X with values 0 and 1, the expectation of X is also the probability that X takes on the value 1.

If the coin flip is *fair*, then $E(X) = P(X{=}1) = 1/2$.

If $A \in \mathscr{A}$ and 1_A denotes the indicator variable of A, then

$$E(1_A) \ = \ 0 \cdot P(1_A{=}0) + 1 \cdot P(1_A{=}1) \ = \ 0 \cdot P(A^c) + 1 \cdot P(A) \ = \ P(A). \tag{4}$$

Example: Joe and Ann With Random Variables

Elements of Ω			Random variables			
Unit	Treatment	Success	Observational-unit variable U	Treatment variable X	Outcome variable Y	Probabilities of elementary events $P(\{\omega\})$
(*Joe*,	*no*,	$-$)	*Joe*	0	0	.09
(*Joe*,	*no*,	$+$)	*Joe*	0	1	.21
(*Joe*,	*yes*,	$-$)	*Joe*	1	0	.04
(*Joe*,	*yes*,	$+$)	*Joe*	1	1	.16
(*Ann*,	*no*,	$-$)	*Ann*	0	0	.24
(*Ann*,	*no*,	$+$)	*Ann*	0	1	.06
(*Ann*,	*yes*,	$-$)	*Ann*	1	0	.16
(*Ann*,	*yes*,	$+$)	*Ann*	1	1	.04

The expectation $E(X)$ of X is

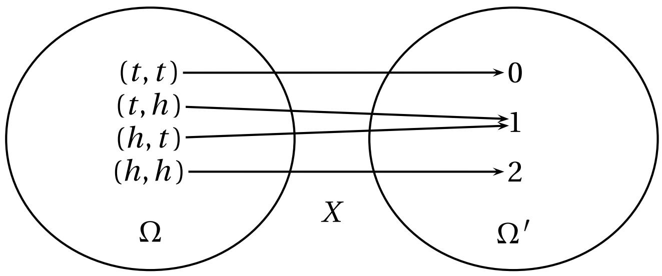
$$
\begin{aligned}
E(X) &= \sum_{i=1}^{2} x_i \cdot P(X{=}x_i) = 0 \cdot P(X{=}0) + 1 \cdot P(X{=}1) \\
&= P(X{=}1) = P\big(\{\omega_3, \omega_4, \omega_7, \omega_8\}\big) = .04 + .16 + .16 + .04 = .40.
\end{aligned}
$$

Example: Number of heads when tossing two coins

Consider flipping two fair coins. Then the set of all possible outcomes is

$$\Omega = \Big\{(h,h), (h,t), (t,h), (t,t)\Big\}$$

and $P(\{\omega\}) = 1/4$ for all $\omega \in \Omega$.



$$
\begin{aligned}
E(X) &= \sum_{i=1}^{3} x_i \cdot P(X{=}x_i) = 0 \cdot P(X{=}0) + 1 \cdot P(X{=}1) + 2 \cdot P(X{=}2) \\
&= 0 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4} = 1.
\end{aligned}
$$

Example: Number of points when tossing a dice: Expectation

Consider tossing a fair dice and let X be the "number of points". Then

$$\begin{aligned} E(X) &= \sum_{i=1}^{6} x_i \cdot P(X{=}x_i) = \frac{1}{6}\sum_{i=1}^{6} x_i \\ &= \frac{1}{6}\cdot(1+2+3+4+5+6) = \frac{21}{6} = 3.5. \end{aligned}$$

Expectation: General Definition

In the following definition we use the concept of an integral of a measurable function with respect to a measure. An introduction to this concept is chapter 3 of Steyer and Nagel (2017).

Definition 2. Let X be a numerical random variable on $(\Omega, \mathscr{A}, P)$. Then

$$E(X) := \int X \, dP \tag{5}$$

is called the *expectation of* X *(with respect to* P*)*, provided that this integral, and with it $E(X)$, exists. If $-\infty < \int X \, dP < \infty$, then we say that $E(X)$ is *finite*.

An alternative notation for the integral with respect to a measure P is

$$E(X) := \int X(\omega) \, P(d\omega). \tag{6}$$

Expectation and Mean Squared Error

Theorem 1. The *mean squared error function MSE*: $\mathbb{R} \rightarrow \mathbb{R}$

$$MSE(a) \;:=\; E\big((X-a)^2\big)$$

has its minimum at $a = E(X)$. Hence, the expectation is the number that minimizes the mean squared error function.

Proof: According to the binomial formula and the rules of computation for expectations [see Eq. (23)]:

$$E\big((X-a)^2\big) \;=\; E\big(X^2 - 2aX + a^2\big) \;=\; E(X^2) - 2aE(X) + a^2.$$

Therefore, the first derivative of $MSE(a)$ is $-2E(X) + 2a$. Fixing this first derivative to 0, we obtain: $a = E(X)$. The second derivative of the function $MSE(a)$ is greater than 0. Therefore, this function has its minimum at $a = E(X)$.

Transformation Theorem

If $g(X)$ is the composition of X and g, then we also call it *a function of* X, and if $\sigma[g(X)] \subset \sigma(X)$, then we say that $g(X)$ *is* X*-measurable*.

Theorem 2. Let $X\colon (\Omega, \mathscr{A}, P) \rightarrow (\Omega', \mathscr{A}')$ be a random variable on $(\Omega, \mathscr{A}, P)$ with distribution P_X, let $g\colon \Omega' \rightarrow \mathbb{R}$ be a real-valued function, and let $g(X)$ denote the composition of X and g such that $\sigma[g(X)] \subset \sigma(X)$. If the expectation of $g(X)$ exists, then

$$E\big[g(X)\big] \;=\; \int g(x)\; P_X\;(dx). \tag{7}$$

Let X be continuous, $g(X)$ an X-measurable function, and f_X a density of X. If the expectation of $g(X)$ exists, then

$$E\big[g(X)\big] \;=\; \int_{-\infty}^{\infty} g(x)\; f_X(x)\; dx, \tag{8}$$

where the right side of this equation denotes the Riemann integral.

Transformation Theorem for a Discrete Random Variable

Let X be a discrete random variable on $(\Omega, \mathscr{A}, P)$ with a finite number of values $x_1, \ldots, x_n$, and let $g(X)$ be a real-valued X-measurable function. If the expectation of $g(X)$ exists, then

$$E\big[g(X)\big] \;=\; \sum_{i=1}^{n} g(x_i) \cdot P(X{=}x_i). \tag{9}$$

Let (X, Y) be a bivariate discrete random variable on $(\Omega, \mathscr{A}, P)$, where X and Y take on the values $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$, respectively. Furthermore, assume that $g(X, Y)$ is a real-valued (X, Y)-measurable function, that $I := \{1, \ldots, n\}$ and $J := \{1, \ldots, m\}$. If the expectation of $g(X, Y)$ exists, then

$$\begin{aligned} E\big[g(X,Y)\big] \;&=\; \sum_{(i,j)\in I\times J} g(x_i, y_j) \cdot P(X{=}x_i, Y{=}y_j) \\ &=\; \sum_{i=1}^{n}\sum_{j=1}^{m} g(x_i, y_j) \cdot P(X{=}x_i, Y{=}y_j). \end{aligned} \tag{10}$$

Variance and Standard Deviation

Definition 3. Let X be a numerical random variable on $(\Omega, \mathscr{A}, P)$ with finite second moment, i.e., with $E(X^2) < \infty$.

(i) The *variance of* X is defined by

$$Var(X) \;:=\; E\Big(\big[X - E(X)\big]^2\Big). \tag{11}$$

(ii) The *standard deviation of* X is the positive square root of the variance, i.e.,

$$SD(X) \;:=\; +\sqrt{Var(X)}. \tag{12}$$

Covariance

Definition 4. Let X and Y be numerical random variables on $(\Omega, \mathscr{A}, P)$ with finite second moments, i.e., with $E(X^2)$, $E(Y^2) < \infty$. Then

$$Cov(X,Y) := E\Big([X - E(X)] \cdot [Y - E(Y)]\Big) \tag{13}$$

is called the *covariance of X and Y*.

Correlation

Definition 5. Let X and Y be numerical random variables on $(\Omega, \mathscr{A}, P)$ with $E(X^2)$, $E(Y^2) < \infty$. Then

$$Corr(X,Y) := \begin{cases} \dfrac{Cov(X,Y)}{SD(X) \cdot SD(Y)}, & \text{if } SD(X), SD(Y) > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

is called the *correlation of X and Y*.

If both standard deviations are greater than 0, then

$$Corr(X,Y) = E\left(\frac{X - E(X)}{SD(X)} \cdot \frac{Y - E(Y)}{SD(Y)}\right). \tag{15}$$

If $X^* = \alpha_0 + \alpha_1 \cdot X$ and $Y^* = \beta_0 + \beta_1 \cdot Y$, with $\alpha_0, \alpha_1, \beta_0, \beta_1 \in \mathbb{R}$, then

$$Corr(X,Y) = Corr(X^*, Y^*). \tag{16}$$

Hence, the correlation is invariant under linear transformations of the two random variables involved.

Example for Computing the Covariance

Table 1: Numerical example for the computation of a covariance

X	Y	$P(X{=}x, Y{=}y)$	$X-E(X)$	$Y-E(Y)$	$[X-E(X)]\cdot[Y-E(Y)]$
0	0	$\frac{5}{40}$	$-\frac{32}{40}$	$-\frac{15}{40}$	$\frac{48}{160}$
0	1	$\frac{3}{40}$	$-\frac{32}{40}$	$\frac{25}{40}$	$-\frac{80}{160}$
1	0	$\frac{20}{40}$	$\frac{8}{40}$	$-\frac{15}{40}$	$-\frac{12}{160}$
1	1	$\frac{12}{40}$	$\frac{8}{40}$	$\frac{25}{40}$	$\frac{20}{160}$

$$Cov(X,Y) \;=\; E\Big(\big[X-E(X)\big]\cdot\big[Y-E(Y)\big]\Big) \;=\; \frac{48}{160}\cdot\frac{5}{40} - \frac{80}{160}\cdot\frac{3}{40} - \frac{12}{160}\cdot\frac{20}{40} + \frac{20}{160}\cdot\frac{12}{40} \;=\; 0.$$

Example: Joe and Ann With Random Variables

Elements of Ω			Random variables							
Unit	Treatment	Success	Observational-unit variable U	Treatment variable X	Outcome variable Y	$X-E(X)$	$Y-E(Y)$	$[X-E(X)]\cdot[Y-E(Y)]$	Probabilities of elementary events $P(\{\omega\})$
(*Joe, no,* −)			*Joe*	0	0	−.40	−.47	.188	.09
(*Joe, no,* +)			*Joe*	0	1	−.40	.53	−.212	.21
(*Joe, yes,* −)			*Joe*	1	0	.60	−.47	−.282	.04
(*Joe, yes,* +)			*Joe*	1	1	.60	.53	.318	.16
(*Ann, no,* −)			*Ann*	0	0	−.40	−.47	.188	.24
(*Ann, no,* +)			*Ann*	0	1	−.40	.53	−.212	.06
(*Ann, yes,* −)			*Ann*	1	0	.60	−.47	−.282	.16
(*Ann, yes,* +)			*Ann*	1	1	.60	.53	.318	.04

$$E(X) \;=\; P(X{=}1) \;=\; P\big(\{\omega_3,\omega_4,\omega_7,\omega_8\}\big) \;=\; .04+.16+.16+.04 \;=\; .40.$$

$$E(Y) \;=\; P(Y{=}1) \;=\; P\big(\{\omega_2,\omega_4,\omega_6,\omega_8\}\big) \;=\; .21+.16+.06+.04 \;=\; .47.$$

$$Cov(X,Y) \;=\; E\Big(\big[X-E(X)\big]\cdot\big[Y-E(Y)\big]\Big)$$

$$=\; .188\cdot(.09+.24) - .212\cdot(.021+.06) - .282\cdot(.04+.16) + .318\cdot(.16+.04) \;=\; .052068.$$

Further Exercises With the Joe-Ann Example

Consider the Joe-Ann example on the last slide and compute the variances of X and Y.
Solution: $Var(X) = .24$ and $Var(Y) = .0.2491$.

Compute the covariance $Cov(X, 1_{U=Joe})$, where $1_{U=Joe}$ denotes the indicator variable of the event that Joe is drawn.
Solution: $Cov(X, 1_{U=Joe}) = 0$.

Implications of Independence

Let X and Y be numerical random variables with $E(X^2), E(Y^2) < \infty$. If X and Y are independent, then

$$E(X \cdot Y) = E(X) \cdot E(Y), \tag{17}$$

$$Cov(X, Y) = 0, \tag{18}$$

and

$$Corr(X, Y) = 0. \tag{19}$$

Let X and Y be dichotomous real-valued random variables with $E(X^2), E(Y^2) < \infty$. Then X and Y are independent if and only if $Cov(X, Y) = 0$.

Rules of Computation for Expectations

Let X be a numerical random variable on $(\Omega, \mathscr{A}, P)$ with expectation $E(X)$ and let $\alpha \in \mathbb{R}$. Then:

$$E(\alpha) = \alpha. \tag{20}$$

$$E(\alpha + X) = \alpha + E(X). \tag{21}$$

$$E(\alpha \cdot X) = \alpha \cdot E(X). \tag{22}$$

Let X and Y be numerical random variables with finite expectations, and let $\alpha, \beta \in \mathbb{R}$. Then

$$E(\alpha \cdot X + \beta \cdot Y) = \alpha \cdot E(X) + \beta \cdot E(Y). \tag{23}$$

Example Binomial Distribution

Let $(\Omega, \mathscr{A}, P)$ be a probability space and consider $X = \sum_{i=1}^{n} 1_{A_i}$, where 1_{A_i} is the indicator variable of $A_i \in \mathscr{A}$, $i = 1, \ldots, n$. Assume

(a) $P(A_i) = p$ for all $i = 1, \ldots, n$, and
(b) the events $A_1, \ldots, A_n$ are independent.

Then X has a binomial distribution and its expectation is

$$\begin{aligned} E(X) &= E\Big(\sum_{i=1}^{n} 1_{A_i}\Big) && [\text{def. of } X] \\ &= \sum_{i=1}^{n} E(1_{A_i}) && [(23)] \\ &= \sum_{i=1}^{n} P(A_i) && [E(1_{A_i}) = P(A_i)] \\ &= \sum_{i=1}^{n} p = n\,p. && [P(A_i) = p] \end{aligned}$$

Rules of Computation for Variances

Let X and Y be random variables with finite second moments $E(X^2)$ and $E(Y^2)$, and let α, $\beta \in \mathbb{R}$. Then

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 && (24)\\ Var(X) &= 0, \qquad \text{if } X = \alpha && (25)\\ Var(\alpha + X) &= Var(X) && (26)\\ Var(\alpha \cdot X) &= \alpha^2 \cdot Var(X). && (27) \end{aligned}$$

If X_i are random variables with finite second moments $E(X_i^2)$ and $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, n$, then

$$Var\left(\sum_{i=1}^{n} \alpha_i X_i\right) = \sum_{i=1}^{n} \alpha_i^2 \; Var(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \alpha_i \, \alpha_j \; Cov(X_i, X_j). \tag{28}$$

For $n = 2$ this equation simplifies to

$$Var(\alpha_1 \cdot X_1 + \alpha_2 \cdot X_2) = \alpha_1^2 \; Var(X_1) + \alpha_2^2 \; Var(X_2) + 2 \, \alpha_1 \, \alpha_2 \; Cov(X_1, X_2). \tag{29}$$

Example: Indicator Variable

Let $(\Omega, \mathscr{A}, P)$ be a probability space and $A \in \mathscr{A}$. Then

$$\begin{aligned} Var(1_A) &= E({1_A}^2) - E(1_A)^2 && [(24)] \qquad (30)\\ &= E(1_A) - E(1_A)^2 && [{1_A}^2 = 1_A]\\ &= E(1_A) \cdot [1 - E(1_A)] &&\\ &= P(A) \cdot [1 - P(A)]. && [(4)] \end{aligned}$$

Example Binomial Distribution

Let $(\Omega, \mathscr{A}, P)$ be a probability space and consider $X = \sum_{i=1}^{n} 1_{A_i}$, where 1_{A_i} is the indicator variable of $A_i \in \mathscr{A}$, $i = 1, \ldots, n$. Assume

(a) $P(A_i) = p$ for all $i = 1, \ldots, n$, and

(b) the events $A_1, \ldots, A_n$ are independent.

Then X has a binomial distribution and its variance is

$$\begin{aligned}
Var(X) &= Var\big(\sum_{i=1}^{n} 1_{A_i}\big) && \text{[def. of } X\text{]} \\
&= \sum_{i=1}^{n} Var(1_{A_i}) && \text{[(28), (18), } Cov(1_{A_i}, 1_{A_j}) = 0 \text{, if } i \neq j\text{]} \\
&= \sum_{i=1}^{n} p(1-p) \;=\; n\,p(1-p). && \text{[(30), (a)]}
\end{aligned}$$

Rules of Computation for Covariances

If X and Y are random variables with finite second moments, then

$$\begin{aligned}
Cov(X, Y) &= E(X \cdot Y) - E(X) \cdot E(Y) && (31) \\
Cov(X, Y) &= 0, \qquad \text{if } X := \alpha && (32) \\
Cov(\alpha + X, \beta + Y) &= Cov(X, Y), \qquad \alpha, \beta \in \mathbb{R} && (33) \\
Cov(\alpha X, \beta Y) &= \alpha \beta\, Cov(X, Y), \qquad \alpha, \beta \in \mathbb{R}. && (34)
\end{aligned}$$

If X_i and Y_j are random variables with finite second moments and $\alpha_i, \beta_j \in \mathbb{R}$, $i, j = 1, \ldots, n$, then

$$Cov\left(\sum_{i=1}^{n} \alpha_i X_i, \sum_{j=1}^{m} \beta_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i\, \beta_j\, Cov(X_i, Y_j). \tag{35}$$

For $n = m = 2$ this yields

$$\begin{aligned}
&Cov(\alpha_1 X_1 + \alpha_2 X_2,\ \beta_1 Y_1 + \beta_2 Y_2) \\
&= \alpha_1 \beta_1\, Cov(X_1, Y_1) + \alpha_1 \beta_2\, Cov(X_1, Y_2) + \\
&\quad\ \alpha_2 \beta_1\, Cov(X_2, Y_1) + \alpha_2 \beta_2\, Cov(X_2, Y_2).
\end{aligned} \tag{36}$$

Definition of the Linear Quasi-Regression

Definition 6. Let X and Y be random variables with finite second moments and $Var(X) > 0$. Then the function $f: \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \alpha_0 + \alpha_1 x, \quad \forall x \in \mathbb{R}, \tag{37}$$

where the pair (α_0, α_1) minimizes the function $MSE: \mathbb{R}^2 \to \mathbb{R}$ with

$$MSE(a_0, a_1) = E\big([Y - (a_0 + a_1 X)]^2\big), \quad \forall a_0, a_1 \in \mathbb{R}, \tag{38}$$

is called the *linear quasi-regression of Y on X*. The composition of X and f is denoted by $Q_{lin}(Y|X)$, i. e.,

$$Q_{lin}(Y|X) = f(X) = \alpha_0 + \alpha_1 X. \tag{39}$$

The linear quasi-regression describes that kind of dependence of Y on X that is quantified by a covariance and a correlation.

Three Characterizations of the Linear Quasi-Regression

Theorem 3. Let $X, Y: (\Omega, \mathscr{A}, P) \to (\mathbb{R}, \mathscr{B})$ be two real-valued random variables with $E(X^2)$, $E(Y^2) < \infty$, and $Var(X) > 0$. Furthermore, let α_0, $\alpha_1 \in \mathbb{R}$, $f(X) = \alpha_0 + \alpha_1 X$ be the composition of X and $f: \mathbb{R} \to \mathbb{R}$, and define $\epsilon := Y - f(X)$. Then the following three propositions are equivalent to each other:

(i) $E(\epsilon) = Cov(X, \epsilon) = 0$.

(ii) $\alpha_0 = E(Y) - \alpha_1 E(X)$ and $\alpha_1 = \dfrac{Cov(X,Y)}{Var(X)}$.

(iii) $f(X) = Q_{lin}(Y|X)$, i. e., α_0, α_1 minimize the function $MSE(a_0, a_1)$ defined by Equation (38).

The proof is found in Steyer and Nagel (2017) (see Th. 7.14).

Conditional Expectation Value

Definition 7. Let Y be a discrete real-valued random variable on a probability space $(\Omega, \mathscr{A}, P)$ with a finite number of values $y_1, \ldots, y_n$, let X be a random variable on $(\Omega, \mathscr{A}, P)$, and x be a value of X such that $P(X{=}x) > 0$. Finally, let $P(Y{=}y_i \,|\, X{=}x)$, $i = 1, \ldots, n$, denote the $(X{=}x)$-conditional probability that Y takes on the value y_i. Then the $(X{=}x)$-*conditional expectation value of* Y is defined by

$$E(Y \,|\, X{=}x) \;:=\; \sum_{i=1}^{n} y_i \cdot P(Y{=}y_i \,|\, X{=}x). \tag{40}$$

If Y is discrete with a countable number of values $y_1, y_2, \ldots$, then the $(X{=}x)$-conditional expectation value of Y is defined by

$$E(Y \mid X{=}x) \;:= \sum_{i=1}^{\infty} y_i \cdot P(Y{=}y_i \,|\, X{=}x) \;:= \lim_{n\to\infty} \sum_{i=1}^{n} y_i \cdot P(Y{=}y_i \,|\, X{=}x)\,. \tag{41}$$

Note that $E(Y \mid X{=}x) = E^{X=x}(Y)$, where $E^{X=x}(Y)$ denotes the expectation of Y with respect to the $(X{=}x)$-conditional-probability measure $P^{X=x}$ on $\mathscr{A}$. Hence, the rules of computation for expectations also apply to $E(Y \mid X{=}x)$. Furthermore, the general definition is analog as well:

$$E(Y \mid X{=}x) := E^{X=x}(Y) = \int Y \, dP^{X=x}. \tag{42}$$