Statistics for Data Science

MSc Data Science WiSe 2019/20

Prof. Dr. Dirk Ostwald

(5) Expectation, covariance, inequalities, limits

Bibliographic remarks

The material presented in this section is not to be understood as a comprehensive introduction to the respective topics, but merely serves as a collection of results that will be put to use use in later sections. The treatment of expectations, variances, and covariances follows Wasserman (2004, Sections 3.1 - 3.3) and **?**, Sections 4.1 - 4.3, 4.6. The discussion of the Markov and Chebyshev inequalities follows **?**, Section 6.2. The discussion of Jensen's inequality, as well as the Weak and Strong Laws of Large Number follow Casella and Berger (2002, Section 4.7 and 5.5), respectively. The discussion of the Central limit theorem follows **?**, Section 6.3.

Expectation, covariance, inequalities, limits

- Expectation
- Variance
- Covariance
- Inequalities
- Limits

Expectation, covariance, inequalities, limits

- **Expectation**
- Variance
- Covariance
- Inequalities
- Limits

Definition

The *expected value* or *expectation* of a random variable X is defined as

$$\mathbb{E}(X) := \int x \, d\mathbb{P}(x) = \begin{cases} \sum_{x \in \mathcal{X}} x \, p(x), & \text{if } X \text{ is discrete} \\ \int_{\mathcal{X}} x \, p(x) \, dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{1}$$

We say that $\mathbb{E}(X)$ exists, if and only if $\int |x| \, d\mathbb{P}(x) < \infty$.

Remarks

- The expectation is a one-number summary of a distribution.
- Intuitively, $\mathbb{E}(X) \approx \frac{1}{n} \sum_{i=1}^{n} X_i$ for a large number n of i.i.d. draws X_i.
- $\int x \, d\mathbb{P}(x)$ serves as a summary notation for $\sum_{x \in \mathcal{X}} x \, p(x)$ and $\int_{\mathcal{X}} x \, p(x) \, dx$.
- The formal basis for $\int x \, d\mathbb{P}(x)$ is the *Lebesgue integral*.

Example (Bernoulli variable expectation)

Let $X \sim \mathrm{Bern}(\mu)$. Then

$$\mathbb{E}(X) = \mu. \tag{2}$$

<u>Proof</u>

X is discrete with $\mathcal{X} = \{0, 1\}$. Thus,

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x \in \{0,1\}} x \, \mathrm{Bern}(x; \mu) \\ &= 0 \cdot \mu^0 (1-\mu)^{1-0} + 1 \cdot \mu^1 (1-\mu)^{1-1} \\ &= 1 \cdot \mu^1 (1-\mu)^0 \\ &= \mu. \end{aligned} \tag{3}$$

$\square$

Example (Gaussian variable expectation)

Let $X \sim \mathrm{N}(\mu, \sigma^2)$. Then

$$\mathbb{E}(X) = \mu. \tag{4}$$

Proof

We assume that

$$\int_{\mathbb{R}} x \exp\left(-x^2\right) \, dx = 0 \text{ and } \int_{\mathbb{R}} \exp\left(-x^2\right) \, dx = \sqrt{\pi} \tag{5}$$

are known. We have

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \, \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \, dx. \tag{6}$$

Setting $\xi := (x - \mu)/\sqrt{2}\sigma$ and thus $x = \sqrt{2}\sigma\xi + \mu$ then yields

$$\begin{aligned}
\mathbb{E}(X) &= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} \left(\sqrt{2}\sigma\xi + \mu\right) \exp\left(-\xi^2\right) \, d\xi \\
&= \frac{1}{\sqrt{\pi}} \left(\sqrt{2}\sigma \int_{\mathbb{R}} \xi \exp\left(-\xi^2\right) \, d\xi + \mu \int_{\mathbb{R}} \exp\left(-\xi^2\right) \, d\xi\right) \\
&= \frac{1}{\sqrt{\pi}} \left(0 + \mu\sqrt{\pi}\right) \\
&= \frac{\mu\sqrt{\pi}}{\sqrt{\pi}} \\
&= \mu.
\end{aligned} \tag{7}$$

□

Theorem (Properties of expectations)

- *Linearity.* If $X_1, ..., X_n$ are random variables and $a_1, ..., a_n$ are constants, then

$$\mathbb{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i). \tag{8}$$

- *Factorization under independence.* If $X_1, ..., X_n$ are independent random variables, then

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i). \tag{9}$$

Remarks

- These properties are helpful when evaluating expectations.
- Proofs of these properties are discussed in the Übung.

Expectation, covariance, inequalities, limits

- Expectation
- **Variance**
- Covariance
- Inequalities
- Limits

Definition (Variance and standard deviation)

Let X be a random variable with expectation $\mathbb{E}(X)$. The variance of X is defined by

$$\mathbb{V}(X) := \mathbb{E}\left((X - \mathbb{E}(X))^2\right), \tag{10}$$

assuming that this expectation exists. The standard deviation is defined as

$$\mathbb{S}(X) := \sqrt{\mathbb{V}(X)}. \tag{11}$$

Remarks

- The variance measures the spread of a distribution.
- The square is necessitated by $\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(X) = 0$.
- An alternative measure of distribution spread is $\mathbb{E}(|X - \mathbb{E}(X)|)$.
- Another alternative measure is the entropy of a distribution.

Example (Bernoulli variable variance)

Let $X \sim \text{Bern}(\mu)$. Then the variance of X is

$$\mathbb{V}(X) = \mu(1-\mu). \tag{12}$$

Proof

X is discrete and we have $\mathbb{E}(X) = \mu$. Thus

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}\left((X-\mu)^2\right) \\
&= \sum_{x \in \{0,1\}} (x-\mu)^2 \text{Bern}(x;\mu) \\
&= (0-\mu)^2 \mu^0 (1-\mu)^{1-0} + (1-\mu)^2 \mu^1 (1-\mu)^{1-1} \\
&= \mu^2(1-\mu) + (1-\mu)^2 \mu \\
&= \left(\mu^2 + (1-\mu)\mu\right)(1-\mu) \\
&= \left(\mu^2 + \mu - \mu^2\right)(1-\mu) \\
&= \mu(1-\mu).
\end{aligned} \tag{13}$$

□

Example (Gaussian variable variance)

Let $X \sim \mathrm{N}(\mu, \sigma^2)$. Then

$$\mathbb{V}(X) = \sigma^2. \tag{14}$$

Proof

See https://proofwiki.org/wiki/Variance_of_Gaussian_Distribution/Proof_1.

$\square$

Theorem (Variance translation theorem)

It holds that

$$\mathbb{V}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2. \tag{15}$$

Proof

With the definition of the variance and the linearity of expectations, we have

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \\ &= \mathbb{E}\left(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2\right) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2. \end{aligned} \tag{16}$$

□

Remark

- The theorem is useful, if computing $\mathbb{E}\left(X^2\right)$ and $\mathbb{E}(X)$ is easy.

Theorem (Variance properties)

- *Variance of linear-affine functions.* If X is a random variable and a and b are constants, then

$$\mathbb{V}(aX + b) = a^2\mathbb{V}(X). \tag{17}$$

- *Variance of linear combinations.* If $X_1, ..., X_n$ are independent and $a_1, ..., a_n$ are constants, then

$$\mathbb{V}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathbb{V}(X_i). \tag{18}$$

Remark

- Proofs of these properties are discussed in the Übung.

Expectation, covariance, inequalities, limits

- Expectation
- Variance
- **Covariance**
- Inequalities
- Limits

Definition (Covariance and correlation)

The covariance of two random variables X and Y with finite expectations is defined as

$$\mathbb{C}(X,Y) = \mathbb{E}\left(\left(X - \mathbb{E}(X)\right)\left(Y - \mathbb{E}(Y)\right)\right). \tag{19}$$

if this expectation exists. The correlation of two random variables X and Y with finite expectations is defined as

$$\rho(X,Y) = \frac{\mathbb{C}(X,Y)}{\sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}}. \tag{20}$$

Remarks

- The covariance of X with itself is the variance of X.
- Correlation is a normalized measure of stochastic dependence.
- $\rho(X,Y) \in [-1,1]$.
- Correlation measures the degree of linear dependence.

Theorem (Covariance translation theorem)

It holds that

$$\mathbb{C}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \tag{21}$$

Proof

With the definition of the covariance, we have

$$\begin{aligned}
\mathbb{C}(X,Y) &= \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right) \\
&= \mathbb{E}\left(XY - X\mathbb{E}(Y) + \mathbb{E}(X)Y - \mathbb{E}(X)\mathbb{E}(Y)\right) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned} \tag{22}$$

□

Remarks

- The theorem is useful, if computing $\mathbb{E}(XY), \mathbb{E}(X)$ and $\mathbb{E}(Y)$ is easy.
- For $Y = X$, we recover $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
- For independent X, Y we have $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ and hence $\mathbb{C}(X,Y) = 0$.

Theorem (Variances of sums of random variables)

For two random variables X and Y it holds that

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathbb{C}(X,Y) \tag{23}$$

and

$$\mathbb{V}(X-Y) = \mathbb{V}(X) + \mathbb{V}(Y) - 2\mathbb{C}(X,Y) \tag{24}$$

Moreover, for n random variables $X_1, ..., X_n$ it holds that

$$\mathbb{V}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathbb{V}(X_i) + 2\sum_{j=1}^{n}\sum_{i\leq j} \mathbb{C}(X_i, X_j). \tag{25}$$

Remark

- Proofs of these properties are discussed in the Übung.

Expectation, covariance, inequalities, limits

- Expectation
- Variance
- Covariance
- **Inequalities**
- Limits

Theorem (Markov inequality)

Let X denote a random variable with $\mathbb{P}(X \geq 0) = 1$. Then for all $x \in \mathbb{R}$ it holds that

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}(X)}{x}. \tag{26}$$

<u>Proof</u>

We consider the case of a continuous X with PDF p. We first note that

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \xi\, p(\xi)\, d\xi = \int_{-\infty}^{x} \xi\, p(\xi)\, d\xi + \int_{x}^{\infty} \xi\, p(\xi)\, d\xi. \tag{27}$$

With $\mathbb{P}(X \geq 0) = 1$ it then follows that

$$\mathbb{E}(X) \geq \int_{x}^{\infty} \xi\, p(\xi)\, d\xi \geq \int_{x}^{\infty} x\, p(\xi)\, d\xi = x \int_{x}^{\infty} p(\xi)\, d\xi = x\, \mathbb{P}(X \geq x). \tag{28}$$

Hence

$$\mathbb{E}(X) \geq x\, \mathbb{P}(X \geq x) \Leftrightarrow \mathbb{P}(X \geq x) \leq \frac{\mathbb{E}(X)}{x}. \tag{29}$$

□

Remarks

- The Markov inequality relates exceedance probabilities and expectations.
- For example, if $\mathbb{E}(X) = 1$, then $\mathbb{P}(X \geq 100) \leq 0.01$.

Theorem (Chebychev inequality)

Let X denote a random variable with variance $\mathbb{V}(X)$. Then for all $x \in \mathbb{R}$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq x) \leq \frac{\mathbb{V}(X)}{x^2}. \tag{30}$$

Proof

We first note that for $a, b \in \mathbb{R}$, it holds that

$$a^2 \geq b^2 \Leftrightarrow |a| \geq b \tag{31}$$

Next, set $Y := X - \mathbb{E}(X)$. Then with the Markov inequality

$$\mathbb{P}\left(Y \geq x^2\right) \leq \frac{\mathbb{E}(Y)}{x^2} \Leftrightarrow \mathbb{P}\left((X - \mathbb{E}(X))^2 \geq x^2\right) \leq \frac{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}{x^2} \Leftrightarrow \mathbb{P}(|X - \mathbb{E}(X)| \geq x) \leq \frac{\mathbb{V}(X)}{x^2}.$$

□

Remarks

- The Chebychev inequality relates deviations from expectations to variances.
- Note that

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq 3\sqrt{\mathbb{V}(X)}\right) \leq \frac{\mathbb{V}(X)}{\left(3\sqrt{\mathbb{V}(X)}\right)^2} = \frac{1}{9}.$$

Theorem (Jensen's inequality)

Let X be a random variable and g be a convex function, i.e.

$$g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda)g(x_2). \tag{32}$$

Then

$$\mathbb{E}(g(X)) \geq g(\mathbb{E}(X)). \tag{33}$$

Conversely, let g be a concave function, i.e.

$$g(\lambda x_1 + (1-\lambda)x_2) \geq \lambda g(x_1) + (1-\lambda)g(x_2). \tag{34}$$

Then

$$\mathbb{E}(g(X)) \leq g(\mathbb{E}(X)). \tag{35}$$

Remarks

- For convex g the function's graph lies below the straight line $g(x_1)$ to $g(x_2)$.
- For concave g the function's graph lies above the straight line $g(x_1)$ to $g(x_2)$.
- The logarithm is a concave function, hence $\mathbb{E}(\ln X) \leq \ln \mathbb{E}(X)$.

Proof

By adapting the proof of Casella and Berger (2002, Theorem 4.7.8), we show the inequality for the concave case. Let f be a tangent line at the point $g(\mathbb{E}(X))$, i.e. is a linear-affine function of the form $f(X) := aX + b$ for some $a, b \in \mathbb{R}$ with $f(\mathbb{E}(X)) = g(\mathbb{E}(X))$. Because g is concave, we have $g(x) \leq ax + b$ for all $x \in \mathbb{R}$ and thus als $g(X) \leq aX + b$. Hence,

$$\mathbb{E}(g(X)) \leq \mathbb{E}(aX + b) = a\mathbb{E}(X) + b = f(\mathbb{E}(X)) = g(\mathbb{E}(X)). \tag{36}$$

□

Theorem (Cauchy-Schwarz inequality)

Let X, Y be two random variables such that $\mathbb{E}(X, Y)$ exists. Then

$$\mathbb{E}(XY)^2 \leq \mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right) \tag{37}$$

Remark

- This is $|\langle x, y\rangle|^2 \leq \langle x, x\rangle \cdot \langle y, y\rangle$ for random variables.

Proof

We consider the case that $0 < \mathbb{E}(X^2) < \infty$ and $0 < \mathbb{E}(Y^2) < \infty$. Then for all $a, b \in \mathbb{R}$, it holds that

$$0 \leq \mathbb{E}\left((aX + bY)^2\right) \text{ and } 0 \leq \mathbb{E}\left((aX - bY)^2\right). \tag{38}$$

Hence,

$$0 \leq a^2\mathbb{E}(X^2) + b^2\mathbb{E}(Y^2) + 2ab\mathbb{E}(XY) \text{ and } 0 \leq a^2\mathbb{E}(X^2) + b^2\mathbb{E}(Y^2) - 2ab\mathbb{E}(XY). \tag{39}$$

Setting $a := \sqrt{\mathbb{E}(Y^2)}$ and $b := \sqrt{\mathbb{E}(X^2)}$ then yields

$$\begin{aligned} 0 &\leq \mathbb{E}(Y^2)\mathbb{E}(X^2) + \mathbb{E}(X^2)\mathbb{E}(Y^2) + 2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) \\ \Leftrightarrow -2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq 2\mathbb{E}(X^2)\mathbb{E}(Y^2) \\ \Leftrightarrow -\sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)} \\ \Leftrightarrow -\mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}, \end{aligned} \tag{40}$$

and similarly

$$\begin{aligned} 0 &\leq \mathbb{E}(Y^2)\mathbb{E}(X^2) + \mathbb{E}(X^2)\mathbb{E}(Y^2) - 2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) \\ \Leftrightarrow 2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq 2\mathbb{E}(X^2)\mathbb{E}(Y^2) \\ \Leftrightarrow \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)} \\ \Leftrightarrow \mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}. \end{aligned} \tag{41}$$

Together, the above imply the Cauchy-Schwarz inequality. See **?**, Theorem 4.2.6 for a full proof.

Theorem (Correlation inequality)

Let X and Y denote two random variables with $\mathbb{V}(X), \mathbb{V}(Y) > 0$. Then

$$\rho(X,Y)^2 = \frac{\mathbb{C}(X,Y)^2}{\mathbb{V}(X)\mathbb{V}(Y)} \leq 1. \tag{42}$$

Proof

According to the Cauchy-Schwarz inequality for random variables U, V, it holds that

$$(\mathbb{E}(UV))^2 \leq \mathbb{E}\left(U^2\right)\mathbb{E}\left(V^2\right). \tag{43}$$

Set $U := X - \mathbb{E}(X)$ and $V := Y - \mathbb{E}(Y)$. Then according to the Cauchy-Schwarz inequality

$$\mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)\right)^2 \leq \mathbb{E}\left((X - \mathbb{E}(X))^2\right)\mathbb{E}\left((Y - \mathbb{E}(Y))^2\right) \tag{44}$$

Hence,

$$\mathbb{C}(X,Y)^2 \leq \mathbb{V}(X)\mathbb{V}(Y) \Leftrightarrow \frac{\mathbb{C}(X,Y)^2}{\mathbb{V}(X)\mathbb{V}(Y)} \leq 1. \tag{45}$$

□

Remark

- $\rho(X,Y)^2 \leq 1 \Leftrightarrow |\rho(X,Y)| \leq 1$, that is $\rho(X,Y) \in [-1,1]$.

Expectation, covariance, inequalities, limits

- Expectation
- Variance
- Covariance
- Inequalities
- **Limits**

Overview

- Intuitively, both the Weak and the Strong Law of Large Numbers state that for i.i.d. random samples of a distribution, the sample mean approximates the expectation of that distribution for large sample sizes. The "Weak" and the "Strong" form of the Law of Large Numbers differ in the form of random variable convergence considered.
- Intuitively, the Central Limit Theorem states that the sum of very many independent and arbitrarily distributed random variables is normally distributed. The "Lindenberg and Lévy" form assumes independent and identically distributed random variables and is easier to proof than the "Liapunov" form, which only assumes independent random variables.

Definition (Convergence in probability)

A sequence $X_1, X_2, \ldots$ of random variables *converges to a random variable X in probability*, written as

$$X_n \xrightarrow[n\to\infty]{P} X, \tag{46}$$

if for every $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \epsilon) = 1 \Leftrightarrow \lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0 \tag{47}$$

Remarks

- $X \xrightarrow[n\to\infty]{\text{P}} X$ means that the probability that X_n lies inside the random interval $]X - \epsilon, X + \epsilon[$, no matter how small this interval may be, approaches 1 as $n \to \infty$.
- Intuitively, for a constant random variable $X := x$, this means that the probability distribution of X_i becomes increasingly concentrated around x as $n \to \infty$.

Theorem (Weak Law of Large Numbers)

Let $X_1, ..., X_n$ denote a random sample from a distribution with expectation μ. Let

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \tag{48}$$

denote the sample mean. Then $\bar{X}_n$ converges to μ in probability,

$$\bar{X}_n \xrightarrow[n\to\infty]{P} \mu. \tag{49}$$

Remark

- $\bar{X}_n \xrightarrow[n\to\infty]{P} \mu$ means that the probability that the sample mean is identical to the expectation of $X_i, i = 1, ..., n$ approaches 1 as the sample size $n \to \infty$.

Definition (Almost sure convergence)

A sequence $X_1, X_2, \ldots$ of random variables *converges almost surely to a random variable* X, written as

$$X_n \xrightarrow[n \to \infty]{\text{a.s}} X, \tag{50}$$

if for every $\varepsilon > 0$

$$\mathbb{P}\left(\lim_{n \to \infty} |X_n - \mu| < \varepsilon \right) = 1. \tag{51}$$

Remarks

- Recall that for $(\Omega, \mathcal{A}, \mathbb{P})$ random variables are functions $X : \Omega \to \mathcal{X}$.
- Let $N \subset \Omega$ be a null set, i.e. $\mathbb{P}(N) = 0$.
- A.s. convergence implies $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega \setminus N$.
- A.s. convergence corresponds to pointwise convergence of function sequences
- A.s. convergence implies convergence in probability, but not vice versa.
- A.s. convergence is a strong form of random variable convergence.

Theorem (Strong Law of Large Numbers)

Let $X_1, ..., X_n$ denote a random sample from a distribution with expectation μ. Let

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \tag{52}$$

denote the sample mean. Then $\bar{X}_n$ converges almost surely to μ,

$$\bar{X}_n \xrightarrow{\text{a.s.}} \mu. \tag{53}$$

Remark

- $\bar{X}_n \xrightarrow{\text{a.s}} \mu$ means that the probability that $|X_n - X|$ is smaller than some arbitrarily small $\epsilon > 0$ is 1 as n goes to infinity.

Definition (Convergence in distribution and asymptotic distribution)

A sequence $X_1, X_2, \ldots$ of random variables *converges to a random variable* X *in distribution*, written as

$$X_n \xrightarrow[n\to\infty]{\text{D}} X, \tag{54}$$

if

$$\lim_{n\to\infty} P_{X_n}(x) = P_X(x). \tag{55}$$

at all points where P_X is continuous. If $X_n \xrightarrow[n\to\infty]{\text{D}} X$, then the distribution of X is referred to as the *asymptotic distribution of* X_n.

Remarks

- $X \xrightarrow[n\to\infty]{\text{D}} X$ is a statement about the convergence of CDFs.
- Convergence in probability implies convergence in distribution.
- Almost sure convergence implies convergence in distribution.
- Convergence in distribution is a weak form of convergence.

Theorem (Central limit theorem (Lindenberg and Lévy))

Let the random variables $X_1, ..., X_n$ form an independent and identically distributed random sample of size n from a given distribution with expectation μ and variance $0 < \sigma^2 < \infty$. Then for each fixed number x

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/n^{1/2}} \leq x\right) = \Phi(x) \tag{56}$$

where Φ is the CDF of the standard normal distribution.

Remarks

- For large n, the distribution of $n^{1/2}(\bar{X}_n - \mu)/\sigma$ is approximately $N(0,1)$.
- For large n, the distribution of $\bar{X}_n$ is approximately $N(\mu, \sigma^2/n)$.
- For large n, the distribution of $\sum_{i=1}^{n} X_i$ is approximately $N(n\mu, n\sigma^2)$.

Theorem (Central limit theorem (Liapounov))

Let $X_1, X_2, \ldots$ be a sequence of independent, but not necessarily identically, distributed random variables, such that

$$\mathbb{E}(|X_i - \mu_i|^3) < \infty \text{ and } \lim_{n\to\infty} \frac{\sum_{i=1}^n \mathbb{E}\left(|X_i - \mu_i|^3\right)}{(\sum_{i=1}^n \sigma_i^2)^{3/2}} = 0. \tag{57}$$

Let $\mu_i := \mathbb{E}(X_i)$ and $\sigma_i^2 = \mathbb{V}(X_i)$ for $i = 1, \ldots, n$ and define

$$Y_n := \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}, \tag{58}$$

such that $\mathbb{E}(Y_n) = 0$ and $\mathbb{V}(Y_n) = 1$. Then, for each fixed number x,

$$\lim_{n\to\infty} \mathbb{P}(Y_n \le x) = \Phi(x), \tag{59}$$

where Φ is the CDF of the standard normal distribution.

Remarks

- For large n, the distribution of $\sum_{i=1}^n X_i$ is approximately $N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$.
- The sum of many independent random factors is approximately normally distributed.
- This justifies the ubiquitous assumption of normally distributed observation errors.

References

Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Held, L. and Sabanés Bové, D. (2014). *Applied statistical inference*, volume 10. Springer.

Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships. *The Annals of Mathematical Statistics*, 14(3):217–226.

Slutsky, E. (1926). Über stochastische asymptoten und grenzwerte (abdruck). *Tohoku Mathematical Journal, First Series*, 27:67–70.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer.