Advanced statistics and modelling

2024 spring

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

EXPECTATION, INEQUALITIES, CONVERGENCE

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

EXPECTATION AND VARIANCE

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

Expected value

Expected value

The **expected value** or **mean** of a random variable X is corresponding to its average value, formally defined as

$$\mathbb{E}(X) = \langle X \rangle = \begin{cases} \sum_i x_i f(x_i) & \text{if } X \text{ is discrete,} \\ \int x\rho(x)dx & \text{if } X \text{ is continuos.} \end{cases}$$

- If $Y = g(X)$, then the expected value of Y can be calculated as

$$\mathbb{E} = \int y\rho_Y(y)dy = \int g(x)\rho_X(x)dx.$$

- If $Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, where a_i are constants, then

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

- If $X_1, X_2, \ldots, X_n$ are **independent** random variables, then

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

Expected value

Expected value

The **expected value** or **mean** of a random variable X is corresponding to its average value, formally defined as

$$\mathbb{E}(X) = \langle X \rangle = \begin{cases} \sum_i x_i f(x_i) & \text{if } X \text{ is discrete,} \\ \int x\rho(x)dx & \text{if } X \text{ is continuos.} \end{cases}$$

- If $Y = g(X)$, then the expected value of Y can be calculated as
$$\mathbb{E} = \int y\rho_Y(y)dy = \int g(x)\rho_X(x)dx.$$
- If $Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, where a_i are constants, then
$$\mathbb{E}(Z) = \mathbb{E}\left(\sum_{i=1}^n a_i Z_i\right) = \sum_{i=1}^n a_i \mathbb{E}(X_i).$$
- If $X_1, X_2, \ldots, X_n$ are **independent** random variables, then
$$\mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i).$$

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Expected value

Expected value

The **expected value** or **mean** of a random variable X is corresponding to its average value, formally defined as

$$\mathbb{E}(X) = \langle X \rangle = \begin{cases} \sum_i x_i f(x_i) & \text{if } X \text{ is discrete,} \\ \int x\rho(x)dx & \text{if } X \text{ is continuos.} \end{cases}$$

- If $Y = g(X)$, then the expected value of Y can be calculated as
$$\mathbb{E} = \int y\rho_Y(y)dy = \int g(x)\rho_X(x)dx.$$
- If $Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, where a_i are constants, then
$$\mathbb{E}(Z) = \mathbb{E}\left(\sum_{i=1}^n a_i Z_i\right) = \sum_{i=1}^n a_i \mathbb{E}(X_i).$$
- If $X_1, X_2, \ldots, X_n$ are **independent** random variables, then
$$\mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i).$$

Expected value

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

Expected value

The **expected value** or **mean** of a random variable X is corresponding to its average value, formally defined as

$$\mathbb{E}(X) = \langle X \rangle = \begin{cases} \sum_i x_i f(x_i) & \text{if } X \text{ is discrete,} \\ \int x\rho(x)dx & \text{if } X \text{ is continuos.} \end{cases}$$

- If $Y = g(X)$, then the expected value of Y can be calculated as

$$\mathbb{E} = \int y\rho_Y(y)dy = \int g(x)\rho_X(x)dx.$$

- If $Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, where a_i are constants, then

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

- If $X_1, X_2, \ldots, X_n$ are **independent** random variables, then

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

Variance and standard deviation

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

Variance and standard deviation

- The **variance** of a random variable X with expected value $\mathbb{E}(X) = \mu$ is defined as

$$\mathbb{V}(X) = \sigma^2(X) = \mathbb{E}\left([X-\mu]^2\right) = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i), & \text{if } X \text{ is discr.} \\ \int (x-\mu)^2 \rho(x)dx, & \text{if } X \text{ is cont.} \end{cases}$$

- The **standard deviation** of X is $\mathrm{sd}(X) = \sigma(X) = \sqrt{\sigma^2(X)} = \sqrt{\mathbb{V}(X)}$.

- According to the definition $\sigma^2(X) = \mathbb{E}(X^2) - \mu^2$.
- If a and b are constants, then $\sigma^2(aX+b) = a^2\sigma^2(X)$.
- If $X_1, X_2, \ldots, X_n$ are independent and $a_1, a_2, \ldots, a_n$ are constants, then

$$\sigma^2\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \sigma^2(X_i).$$

Variance and standard deviation

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

Variance and standard deviation

- The **variance** of a random variable X with expected value $\mathbb{E}(X) = \mu$ is defined as

$$\mathbb{V}(X) = \sigma^2(X) = \mathbb{E}\left([X-\mu]^2\right) = \begin{cases} \sum_i (x_i-\mu)^2 f(x_i), & \text{if } X \text{ is discr.} \\ \int (x-\mu)^2 \rho(x)dx, & \text{if } X \text{ is cont.} \end{cases}$$

- The **standard deviation** of X is $\mathrm{sd}(X) = \sigma(X) = \sqrt{\sigma^2(X)} = \sqrt{\mathbb{V}(X)}$.

- According to the definition $\sigma^2(X) = \mathbb{E}(X^2) - \mu^2$.
- If a and b are constants, then $\sigma^2(aX+b) = a^2\sigma^2(X)$.
- If $X_1, X_2, \dots, X_n$ are independent and $a_1, a_2, \dots, a_n$ are constants, then

$$\sigma^2\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \sigma^2(X_i).$$

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Variance and standard deviation

Variance and standard deviation

- The **variance** of a random variable X with expected value $\mathbb{E}(X) = \mu$ is defined as

$$\mathbb{V}(X) = \sigma^2(X) = \mathbb{E}\left([X-\mu]^2\right) = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i), & \text{if } X \text{ is discr.} \\ \int (x-\mu)^2 \rho(x)dx, & \text{if } X \text{ is cont.} \end{cases}$$

- The **standard deviation** of X is $\mathrm{sd}(X) = \sigma(X) = \sqrt{\sigma^2(X)} = \sqrt{\mathbb{V}(X)}$.

- According to the definition $\sigma^2(X) = \mathbb{E}(X^2) - \mu^2$.
- If a and b are constants, then $\sigma^2(aX + b) = a^2\sigma^2(X)$.
- If $X_1, X_2, \ldots, X_n$ are independent and $a_1, a_2, \ldots, a_n$ are constants, then

$$\sigma^2\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \sigma^2(X_i).$$

Variance and standard deviation

Variance and standard deviation

- The **variance** of a random variable X with expected value $\mathbb{E}(X) = \mu$ is defined as

$$\mathbb{V}(X) = \sigma^2(X) = \mathbb{E}\left([X-\mu]^2\right) = \begin{cases} \sum_i (x_i-\mu)^2 f(x_i), & \text{if } X \text{ is discr.} \\ \int (x-\mu)^2 \rho(x)dx, & \text{if } X \text{ is cont.} \end{cases}$$

- The **standard deviation** of X is $\mathrm{sd}(X) = \sigma(X) = \sqrt{\sigma^2(X)} = \sqrt{\mathbb{V}(X)}$.

- According to the definition $\sigma^2(X) = \mathbb{E}(X^2) - \mu^2$.
- If a and b are constants, then $\sigma^2(aX+b) = a^2\sigma^2(X)$.
- If $X_1, X_2, \ldots, X_n$ are independent and $a_1, a_2, \ldots, a_n$ are constants, then

$$\sigma^2\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \sigma^2(X_i).$$

Mean and variance of important distributions

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Distribution	Mean	Variance
Bernoulli(p)	p	$p(1-p)$
Binomial(N,p)	Np	$Np(1-p)$
Geometric(p)	$1/p$	$(1-p)/p^2$
Poisson(λ)	λ	λ
Uniform(a,b)	$(a+b)/2$	$(b-a)^2/12$
Normal(μ,σ^2)	μ	σ^2
Exponential(λ)	λ	λ^2
Gamma(q,λ)	$q\lambda$	$q\lambda^2$
χ_n^2	n	$2n$
t_n	0 if $n>1$	$n/(n-2)$ if $n>2$

Sample mean and sample variance

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Sample mean and sample variance

- If $X_1, X_2, \ldots, X_n$ are random variables, then their **sample mean** is defined as

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- The **sample variance** is defined as

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X_n})^2.$$

- If $X_1, X_2, \ldots, X_n$ are IID where $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$, then

$$\begin{aligned} \mathbb{E}(\overline{X_n}) &= \mu, \\ \mathbb{V}(\overline{X_n}) &= \frac{\sigma^2}{n}, \\ \mathbb{E}(S_n^2) &= \sigma^2. \end{aligned}$$

Sample mean and sample variance

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Sample mean and sample variance

- If $X_1, X_2, \ldots, X_n$ are random variables, then their **sample mean** is defined as

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- The **sample variance** is defined as

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

- If $X_1, X_2, \ldots, X_n$ are IID where $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$, then

$$\begin{aligned}\mathbb{E}(\overline{X_n}) &= \mu,\\ \mathbb{V}(\overline{X_n}) &= \frac{\sigma^2}{n},\\ \mathbb{E}(S_n^2) &= \sigma^2.\end{aligned}$$

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Sample mean and sample variance

Sample mean and sample variance

- If $X_1, X_2, \ldots, X_n$ are random variables, then their **sample mean** is defined as

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- The **sample variance** is defined as

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2.$$

- If $X_1, X_2, \ldots, X_n$ are IID where $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$, then

$$\begin{aligned} \mathbb{E}(\overline{X_n}) &= \mu, \\ \mathbb{V}(\overline{X_n}) &= \frac{\sigma^2}{n}, \\ \mathbb{E}(S_n^2) &= \sigma^2. \end{aligned}$$

Covariance and correlation

Covariance and correlation

- The **covariance** between random variables X and Y is defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right].$$

- The **correlation** is simply

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

- According to the definition of the covariance $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
- The correlation satisfies $-1 \leq \mathrm{Corr}(X,Y) \leq 1$.
- For random variables $X_1, X_2, \ldots X_l$ the variance of their linear combination is

$$\mathbb{V}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \mathbb{V}(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j).$$

Covariance and correlation

Covariance and correlation

- The **covariance** between random variables X and Y is defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right].$$

- The **correlation** is simply

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

- According to the definition of the covariance $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
- The correlation satisfies $-1 \le \mathrm{Corr}(X,Y) \le 1$.
- For random variables $X_1, X_2, \ldots X_i$ the variance of their linear combination is

$$\mathbb{V}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \mathbb{V}(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j).$$

Covariance and correlation

Covariance and correlation

- The **covariance** between random variables X and Y is defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right].$$

- The **correlation** is simply

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

- According to the definition of the covariance $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
- The correlation satisfies $-1 \leq \mathrm{Corr}(X,Y) \leq 1$.
- For random variables $X_1, X_2, \ldots X_i$ the variance of their linear combination is

$$\mathbb{V}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \mathbb{V}(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j).$$

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Covariance and correlation

Covariance and correlation

- The **covariance** between random variables X and Y is defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right].$$

- The **correlation** is simply

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

- According to the definition of the covariance $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
- The correlation satisfies $-1 \le \mathrm{Corr}(X,Y) \le 1$.
- For random variables $X_1, X_2, \dots X_i$ the variance of their linear combination is

$$\mathbb{V}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \mathbb{V}(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j).$$

Covariance and correlation

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

Covariance and correlation

- The **covariance** between random variables X and Y is defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right].$$

- The **correlation** is simply

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

- According to the definition of the covariance $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
- The correlation satisfies $-1 \le \mathrm{Corr}(X,Y) \le 1$.
- For random variables $X_1, X_2, \ldots X_i$ the variance of their linear combination is

$$\mathbb{V}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \mathbb{V}(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j).$$

Covariance matrix

Variance-covariance matrix

For a vector of random variables $X = (X_1, X_2, \ldots, X_n)$ the **variance-covariance matrix** is defined as

$$\Sigma(X) = \begin{pmatrix} \mathbb{V}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathbb{V}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathbb{V}(X_n) \end{pmatrix}$$

- For a multinomial distribution $X \sim$Multinomial(N, p) the variance-covariance matrix reads

$$\Sigma(X) = \begin{pmatrix} Np_1(1-p_1) & -Np_1p_2 & \cdots & -Np_1p_n \\ -Np_2p_1 & Np_2(1-p_2) & \cdots & -Np_2p_n \\ \vdots & & \ddots & \vdots \\ -Np_kp_1 & -Np_kp_2 & \cdots & Np_k(1-p_k) \end{pmatrix}$$

- In case of a multivariate Normal distribution, $X \sim N(\mu, \Sigma)$ the parameter Σ is actually the covariance matrix.

Covariance matrix

Variance-covariance matrix

For a vector of random variables $X = (X_1, X_2, \dots, X_n)$ the **variance-covariance matrix** is defined as

$$\Sigma(X) = \begin{pmatrix} \mathbb{V}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathbb{V}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathbb{V}(X_n) \end{pmatrix}$$

- For a multinomial distribution $X \sim$Multinomial(N, p) the variance-covariance matrix reads

$$\Sigma(X) = \begin{pmatrix} Np_1(1-p_1) & -Np_1p_2 & \cdots & -Np_1p_n \\ -Np_2p_1 & Np_2(1-p_2) & \cdots & -Np_2p_n \\ \vdots & & \ddots & \vdots \\ -Np_kp_1 & -Np_kp_2 & \cdots & Np_k(1-p_k) \end{pmatrix}$$

- In case of a multivariate Normal distribution, $X \sim N(\mu, \Sigma)$ the parameter Σ is actually the covariance matrix.

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Covariance matrix

Variance-covariance matrix

For a vector of random variables $X = (X_1, X_2, \ldots, X_n)$ the **variance-covariance matrix** is defined as

$$\Sigma(X) = \begin{pmatrix} \mathbb{V}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathbb{V}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathbb{V}(X_n) \end{pmatrix}$$

- For a multinomial distribution $X \sim$Multinomial(N, p) the variance-covariance matrix reads

$$\Sigma(X) = \begin{pmatrix} Np_1(1-p_1) & -Np_1p_2 & \cdots & -Np_1p_n \\ -Np_2p_1 & Np_2(1-p_2) & \cdots & -Np_2p_n \\ \vdots & & \ddots & \vdots \\ -Np_kp_1 & -Np_kp_2 & \cdots & Np_k(1-p_k) \end{pmatrix}$$

- In case of a multivariate Normal distribution, $X \sim N(\mu, \Sigma)$ the parameter Σ is actually the covariance matrix.

Covariance matrix

- If $\vec{a} = (a_1, a_2, \ldots, a_n)$ is a constant vector, and $\vec{X} = (X_1, X_2, \ldots, X_n)$ is a random vector with $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_n) = (\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_n))$ and a variance-covariance matrix $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j)$, then

$$\mathbb{V}(\vec{a}^T \vec{X}) = \vec{a}^T \Sigma \vec{a}.$$

- If A is an n by n matrix, then

$$\Sigma(A\vec{X}) = A \Sigma A^T.$$

Covariance matrix

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

- If $\vec{a} = (a_1, a_2, \ldots, a_n)$ is a constant vector, and $\vec{X} = (X_1, X_2, \ldots, X_n)$ is a random vector with $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_n) = (\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_n))$ and a variance-covariance matrix $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j)$, then

$$\mathbb{V}(\vec{a}^T \vec{X}) = \vec{a}^T \Sigma \vec{a}.$$

- If A is an n by n matrix, then

$$\Sigma(A\vec{X}) = A \Sigma A^T.$$

Conditional expectation

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Conditional expectation

- The **conditional expectation and variance** of variable X given that variable Y takes the value $Y = y$ is given by

$$\mathbb{E}(X \mid Y = y) = \int x\rho(x \mid Y = y)dx,$$
$$\mathbb{V}(X \mid Y = y) = \int [x - \mathbb{E}(X \mid Y = y)]^2 \rho(x \mid Y = y)dx.$$

(Note that these are functions of y, thus, can be treated also as random variables).

- The conditional expectation of a function $r(x, y)$ is similarly

$$\mathbb{E}(r(X, Y) \mid Y = y) = \int r(x, y)\rho(x \mid Y = y)dx.$$

- Rule of iterated expectations:

$$\mathbb{E}\left(\mathbb{E}(X \mid Y = y)\right) = \mathbb{E}(X).$$
$$\mathbb{E}\left(\mathbb{E}(r(X, Y) \mid Y = y)\right) = \mathbb{E}(r(X, Y)).$$

Conditional expectation

Conditional expectation

- The **conditional expectation and variance** of variable X given that variable Y takes the value $Y = y$ is given by

$$\mathbb{E}(X \mid Y = y) = \int x\rho(x \mid Y = y)dx,$$
$$\mathbb{V}(X \mid Y = y) = \int [x - \mathbb{E}(X \mid Y = y)]^2 \rho(x \mid Y = y)dx.$$

(Note that these are functions of y, thus, can be treated also as random variables).

- The conditional expectation of a function $r(x, y)$ is similarly

$$\mathbb{E}(r(X, Y) \mid Y = y) = \int r(x, y)\rho(x \mid Y = y)dx.$$

- Rule of iterated expectations:

$$\mathbb{E}\left(\mathbb{E}(X \mid Y = y)\right) = \mathbb{E}(X).$$
$$\mathbb{E}\left(\mathbb{E}(r(X, Y) \mid Y = y)\right) = \mathbb{E}(r(X, Y)).$$

Conditional expectation

Conditional expectation

- The **conditional expectation and variance** of variable X given that variable Y takes the value $Y = y$ is given by

$$\mathbb{E}(X \mid Y = y) = \int x\rho(x \mid Y = y)dx,$$
$$\mathbb{V}(X \mid Y = y) = \int [x - \mathbb{E}(X \mid Y = y)]^2 \rho(x \mid Y = y)dx.$$

(Note that these are functions of y, thus, can be treated also as random variables).

- The conditional expectation of a function $r(x, y)$ is similarly

$$\mathbb{E}(r(X, Y) \mid Y = y) = \int r(x, y)\rho(x \mid Y = y)dx.$$

- Rule of iterated expectations:

$$\mathbb{E}\left(\mathbb{E}(X \mid Y = y)\right) = \mathbb{E}(X).$$
$$\mathbb{E}\left(\mathbb{E}(r(X, Y) \mid Y = y)\right) = \mathbb{E}(r(X, Y)).$$

Conditional expectation

Conditional expectation

- The **conditional expectation and variance** of variable X given that variable Y takes the value $Y = y$ is given by

$$\mathbb{E}(X \mid Y = y) = \int x\rho(x \mid Y = y)dx,$$

$$\mathbb{V}(X \mid Y = y) = \int [x - \mathbb{E}(X \mid Y = y)]^2 \rho(x \mid Y = y)dx.$$

 (Note that these are functions of y, thus, can be treated also as random variables).

- The conditional expectation of a function $r(x, y)$ is similarly

$$\mathbb{E}(r(X, Y) \mid Y = y) = \int r(x, y)\rho(x \mid Y = y)dx.$$

- Rule of iterated expectations:

$$\mathbb{E}\left(\mathbb{E}(X \mid Y = y)\right) = \mathbb{E}(X).$$

$$\mathbb{E}\left(\mathbb{E}(r(X, Y) \mid Y = y)\right) = \mathbb{E}(r(X, Y)).$$

Conditional expectation

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

- Variance and conditional variance:

$$\mathbb{V}(X) = \mathbb{E}[\mathbb{V}(X \mid Y = y)] + \mathbb{V}[\mathbb{E}(X \mid Y = y)].$$

Expectation, Inequalities, Convergence

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function

Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality

Convergence
Convergence types
Law of large numbers
Central limit theorems

Characteristic function and moment generating function

- The characteristic function ϕ of a random variable X is defined as

$$\phi_X(t) = \mathbb{E}\left(e^{itX}\right) = \begin{cases} \sum_i e^{itx_i} f_X(x_i) & \text{if } X \text{ is discrete,} \\ \int e^{itx} \rho_X(x) dx & \text{if } X \text{ is continuos.} \end{cases}$$

- The moment generating function ψ is

$$\psi_X(t) = \mathbb{E}\left(e^{tX}\right) = \begin{cases} \sum_i e^{tx_i} f_X(x_i) & \text{if } X \text{ is discrete,} \\ \int e^{tx} \rho_X(x) dx & \text{if } X \text{ is continuos.} \end{cases}$$

- The n-th moment of X can be calculated as

$$\mathbb{E}(X^n) = \frac{1}{i^n} \frac{\partial^n \phi(t)}{\partial t^n}\bigg|_{t=0} = \frac{\partial^n \psi_X(t)}{\partial t^n}\bigg|_{t=0}.$$

Characteristic function

Characteristic function and moment generating function

- The characteristic function ϕ of a random variable X is defined as

$$\phi_X(t) = \mathbb{E}\left(e^{itX}\right) = \begin{cases} \sum_i e^{itx_i} f_X(x_i) & \text{if } X \text{ is discrete,} \\ \int e^{itx} \rho_X(x) dx & \text{if } X \text{ is continuos.} \end{cases}$$

- The moment generating function ψ is

$$\psi_X(t) = \mathbb{E}\left(e^{tX}\right) = \begin{cases} \sum_i e^{tx_i} f_X(x_i) & \text{if } X \text{ is discrete,} \\ \int e^{tx} \rho_X(x) dx & \text{if } X \text{ is continuos.} \end{cases}$$

- The n-th moment of X can be calculated as

$$\mathbb{E}(X^n) = \frac{1}{i^n} \frac{\partial^n \phi(t)}{\partial t^n}\bigg|_{t=0} = \frac{\partial^n \psi_X(t)}{\partial t^n}\bigg|_{t=0}.$$

Characteristic function

Expectation, Inequalities, Convergence

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function

Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality

Convergence
Convergence types
Law of large numbers
Central limit theorems

Characteristic function and moment generating function

- The characteristic function ϕ of a random variable X is defined as

$$\phi_X(t) = \mathbb{E}\left(e^{itX}\right) = \begin{cases} \sum_i e^{itx_i} f_X(x_i) & \text{if } X \text{ is discrete,} \\ \int e^{itx} \rho_X(x)dx & \text{if } X \text{ is continuos.} \end{cases}$$

- The moment generating function ψ is

$$\psi_X(t) = \mathbb{E}\left(e^{tX}\right) = \begin{cases} \sum_i e^{tx_i} f_X(x_i) & \text{if } X \text{ is discrete,} \\ \int e^{tx} \rho_X(x)dx & \text{if } X \text{ is continuos.} \end{cases}$$

- The n-th moment of X can be calculated as

$$\mathbb{E}(X^n) = \frac{1}{i^n} \frac{\partial^n \phi(t)}{\partial t^n}\bigg|_{t=0} = \frac{\partial^n \psi_X(t)}{\partial t^n}\bigg|_{t=0}.$$

Characteristic function

- For discrete variables the **generating function** is defined as

$$G_X(z) = \sum_i z^{x_i} f_X(x_i).$$

→ The characteristic function is $\phi_X(t) = G_X(z = e^{it})$ and the moment generating function is $\psi_X(t) = G_X(z = e^t)$.

- For independent random variables $X_1, X_2, \ldots, X_n$, the characteristic function, the moment generating function (and for discrete variables, the generating function) for their sum $X = \sum_{i=1}^{n} X_i$ is simply

$$\phi_X(t) = \prod_{i=1}^{n} \phi_{X_i}(t), \qquad \psi_X(t) = \prod_{i=1}^{n} \psi_{X_i}(t), \qquad G_X(z) = \prod_{i=1}^{n} G_{X_i}(z).$$

Characteristic function

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

- For discrete variables the **generating function** is defined as

$$G_X(z) = \sum_i z^{x_i} f_X(x_i).$$

→ The characteristic function is $\phi_X(t) = G_X(z = e^{it})$ and the moment generating function is $\psi_X(t) = G_X(z = e^t)$.

- For independent random variables $X_1, X_2, \ldots, X_n$, the characteristic function, the moment generating function (and for discrete variables, the generating function) for their sum $X = \sum_{i=1}^{n} X_i$ is simply

$$\phi_X(t) = \prod_{i=1}^{n} \phi_{X_i}(t), \qquad \psi_X(t) = \prod_{i=1}^{n} \psi_{X_i}(t), \qquad G_X(z) = \prod_{i=1}^{n} G_{X_i}(z).$$

Characteristic function

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

- For discrete variables the **generating function** is defined as

$$G_X(z) = \sum_i z^{x_i} f_X(x_i).$$

→ The characteristic function is $\phi_X(t) = G_X(z = e^{it})$ and the moment generating function is $\psi_X(t) = G_X(z = e^t)$.

- For independent random variables $X_1, X_2, \ldots, X_n$, the characteristic function, the moment generating function (and for discrete variables, the generating function) for their sum $X = \sum_{i=1}^{n} X_i$ is simply

$$\phi_X(t) = \prod_{i=1}^{n} \phi_{X_i}(t), \qquad \psi_X(t) = \prod_{i=1}^{n} \psi_{X_i}(t), \qquad G_X(z) = \prod_{i=1}^{n} G_{X_i}(z).$$

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

INEQUALITIES

Markov's inequality

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

Markov's inequality

Assuming that X is a non-negative random variable for which $\mathbb{E}(X)$ exists, for any $z < 0$ the following inequality holds:

$$P(X > z) \leq \frac{\mathbb{E}(X)}{z}.$$

Chebyshev's inequality

Chebyshev's inequality

Assuming $\mathbb{E}(X) = \mu$ and $\mathbb{V}(X) = \sigma^2$ for a random variable X, the following inequalities hold:

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

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

Hoeffding's inequality

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

Hoeffding's inequality for Bernoulli random variables

- Assume IID Bernoulli distributed random variables $X_1, X_2, \ldots, X_n \sim \text{Bernoulli}(p)$:

$$P(X_i = 1) = p, \qquad P(X_i = 0) = 1 - p, \qquad \forall i.$$

- The sample mean is defined as

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

and **Hoeffding's inequality** states that this is getting "exponentially" close to p if n is increased, since for any $\epsilon > 0$

$$P\left(\left|\overline{X_n} - p\right| \geq \epsilon\right) \leq 2e^{-2n\epsilon^2}.$$

Hoeffding's inequality

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Hoeffding's inequality (in general)

- Assume $X_1, X_2, \ldots, X_n$ are independent random variables, bounded by the intervals $[a_i, b_i]$, (thus, $a_i \le X_i \le b_i$).
- The sample mean is defined as usual,

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

and according to **Hoeffding's inequality** for any $\epsilon > 0$

$$P\left(\left|\overline{X_n} - \mathbb{E}(\overline{X_n})\right| \ge \epsilon\right) \le 2\exp\left(-\frac{2n^2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Hoeffding's inequality

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Proof:

- First we need Hoeffding's Lemma, stating that assuming a random variable X bounded by the interval $[a, b]$ with $\mathbb{E}(X) = 0$, for any λ

$$\mathbb{E}(e^{\lambda X}) \le \exp\left(\frac{\lambda^2 (b-a)^2}{8}\right).$$

- Proof of the lemma:
  - Since e^{sx} is a convex function,

$$\forall x \in [a, b]: \quad e^{sx} \le \frac{b-x}{b-a} e^{sa} + \frac{x-a}{b-a} e^{sb}.$$

  - Applying $\mathbb{E}$ to both sides we obtain

$$\mathbb{E}(e^{sX}) \quad \le \quad \frac{b-\mathbb{E}(X)}{b-a} e^{sa} + \frac{\mathbb{E}(X)-a}{b-a} e^{sb} \underset{\mathbb{E}(X)=0}{=} \frac{b}{b-a} e^{sa} - \frac{a}{b-a} e^{sb} =$$

$$\left(-\frac{a}{b-a}\right) e^{sa}\left[-\frac{b}{a} + e^{sb-sa}\right] =$$

$$\left(-\frac{a}{b-a}\right) e^{sa}\left[-\frac{b-a+a}{a} + e^{s(b-a)}\right] =$$

Hoeffding's inequality

$$\begin{aligned}\mathbb{E}(e^{sX}) \quad &\leq \quad \left(-\frac{a}{b-a}\right) e^{sa} \left[-\frac{b-a+a}{a} + e^{s(b-a)}\right] = \\ &\left(-\frac{a}{b-a}\right) e^{sa} \left[-\frac{b-a}{a} - 1 + e^{s(b-a)}\right].\end{aligned}$$

- Let us define $\theta = -\frac{a}{b-a} > 0$, yielding

$$\mathbb{E}(e^{sX}) \leq \left(1 - \theta + \theta e^{s(b-a)}\right) e^{-s\theta(b-a)}.$$

- We further define $u = s(b-a)$ and the function $\varphi : \mathbb{R} \to \mathbb{R}$ with $\varphi(u) = -\theta u + \ln(1 - \theta + \theta e^u)$. This way

$$\mathbb{E}(e^{sX}) \leq e^{\phi(u)}.$$

- We are going to use Taylor's theorem (the Extended Mean Value Theorem), stating that for every u there exists a v between 0 and u such that

$$\varphi(u) = \varphi(0) + u\varphi'(0) + \frac{u^2}{2}\varphi''(v).$$

Hoeffding's inequality

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

- The derivatives of $\varphi(u) = -\theta u + \ln(1 - \theta + \theta e^u)$ are

$$\begin{aligned}
\varphi(0) &= 0 \\
\varphi'(0) &= -\theta + \left.\frac{\theta e^u}{1-\theta+\theta e^u}\right|_{u=0} = 0 \\
\varphi''(v) &= \underbrace{\frac{\theta e^v}{1-\theta+\theta e^v}}_{t}\left(1 - \frac{\theta e^v}{1-\theta+\theta e^v}\right) = t(1-t) \le \frac{1}{4}.
\end{aligned}$$

- Thus, according to Taylor's theorem

$$\varphi(u) \le 0 + 0 \cdot u + \frac{u^2}{2}\frac{1}{4} = \frac{s^2(b-a)^2}{8},$$

- proving the lemma

$$\mathbb{E}(e^{sX}) \le e^{\varphi(u)} \le \exp\left(\frac{s^2(b-a)^2}{8}\right).$$

Hoeffding's inequality

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

- Back to the main proof:
  We first point out that equivalently to

$$P\left(\left|\overline{X_n} - \mathbb{E}(\overline{X_n})\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n^2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

also

$$P\left(\overline{X_n} - \mathbb{E}(\overline{X_n}) \geq \epsilon\right) \leq \exp\left(-\frac{2n^2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Taking the expression on the left, for any $s > 0$ we can write

$$P\left(\overline{X_n} - \mathbb{E}(\overline{X_n}) \geq \epsilon\right) = P\left(e^{s\left(\overline{X_n} - \mathbb{E}(\overline{X_n})\right)} > e^{s\epsilon}\right).$$

- Since the variable on the left is non-negative, we can use Markov's inequality as

$$P\left(\overline{X_n} - \mathbb{E}(\overline{X_n}) \geq \epsilon\right) = P\left(e^{s\left(\overline{X_n} - \mathbb{E}(\overline{X_n})\right)} > e^{s\epsilon}\right) \leq e^{-s\epsilon}\mathbb{E}\left[e^{s\left(\overline{X_n} - \mathbb{E}(\overline{X_n})\right)}\right] = e^{-s\epsilon}\prod_{i=1}^n \mathbb{E}\left[e^{\frac{s}{n}(X_i - \mathbb{E}(X_i))}\right]$$

Hoeffding's inequality

According to the above

$$P\left(\overline{X_n} - \mathbb{E}(\overline{X_n}) \geq \epsilon\right) \leq e^{-s\epsilon} \prod_{i=1}^{n} \mathbb{E}\left[e^{\frac{s}{n}(X_i - \mathbb{E}(X_i))}\right].$$

- We can use Hoeffding's lemma to the term on the right as

$$\begin{aligned} P\left(\overline{X_n} - \mathbb{E}(\overline{X_n}) \geq \epsilon\right) \quad &\leq \quad e^{-s\epsilon} \prod_{i=1}^{n} \mathbb{E}\left[e^{\frac{s}{n}(X_i - \mathbb{E}(X_i))}\right] \leq \\ & \quad\;\; e^{-s\epsilon} \prod_{i=1}^{n} e^{\frac{s^2 (b_i - a_i)^2}{8n^2}} = \exp\left(-s\epsilon + \frac{s^2}{8n^2} \sum_{i=1}^{n} (b_i - a_i)^2\right). \end{aligned}$$

- To obtain the best possible upper bound we have to find the minimum of $g(s) = -s\epsilon + \frac{s^2}{8n^2} \sum_{i=1}^{n} (b_i - a_i)^2$ as a function of s, which happens to be at $s = 4\epsilon n^2 \left[\sum_{i=1}^{n} (b_i - a_i)^2\right]^{-1}$. By substituting back we obtain that

$$P\left(\overline{X_n} - \mathbb{E}(\overline{X_n}) \geq \epsilon\right) \leq \exp\left(-\frac{2\epsilon^2 n^2}{\sum_{i=1}^{n} (b_i - a_i)^2}\right).$$

Cauchy–Schwarz inequality

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

Cauchy–Schwarz inequality

For any random variables X and Y for which $\mathbb{E}(X)$ exists, the following inequality holds:

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Proof:

- For any real number a we know that $(Y - aX)^2 \geq 0$, thus,

$$0 \leq \mathbb{E}\left[(Y - aX)^2\right] = \mathbb{E}\left[Y^2 - 2aYX + a^2X^2\right] = \mathbb{E}(Y^2) - 2a\mathbb{E}(YX) + a^2\mathbb{E}(X^2).$$

- Let us choose $a = \frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}$:

$$0 \leq \mathbb{E}(Y^2) - \frac{\mathbb{E}(XY)^2}{\mathbb{E}(X^2)} \rightarrow 0 \leq \mathbb{E}(Y^2)\mathbb{E}(X^2) - \mathbb{E}(XY)^2$$

Cauchy–Schwarz inequality

Cauchy–Schwarz inequality

For any random variables X and Y for which $\mathbb{E}(X)$ exists, the following inequality holds:

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Proof:

- For any real number a we know that $(Y - aX)^2 \geq 0$, thus,

$$0 \leq \mathbb{E}\left[(Y - aX)^2\right] = \mathbb{E}\left[Y^2 - 2aYX + a^2X^2\right] = \mathbb{E}(Y^2) - 2a\mathbb{E}(YX) + a^2\mathbb{E}(X^2).$$

- Let us choose $a = \frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}$:

$$0 \leq \mathbb{E}(Y^2) - \frac{\mathbb{E}(XY)^2}{\mathbb{E}(X^2)} \quad \rightarrow \quad 0 \leq \mathbb{E}(Y^2)\mathbb{E}(X^2) - \mathbb{E}(XY)^2$$

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Cauchy–Schwarz inequality

Cauchy–Schwarz inequality

For any random variables X and Y for which $\mathbb{E}(X)$ exists, the following inequality holds:

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Proof:

- For any real number a we know that $(Y - aX)^2 \geq 0$, thus,

$$0 \leq \mathbb{E}\left[(Y - aX)^2\right] = \mathbb{E}\left[Y^2 - 2aYX + a^2X^2\right] = \mathbb{E}(Y^2) - 2a\mathbb{E}(YX) + a^2\mathbb{E}(X^2).$$

- Let us choose $a = \frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}$:

$$0 \leq \mathbb{E}(Y^2) - \frac{\mathbb{E}(XY)^2}{\mathbb{E}(X^2)} \quad \rightarrow \quad 0 \leq \mathbb{E}(Y^2)\mathbb{E}(X^2) - \mathbb{E}(XY)^2$$

Jensen's inequality

Jensen's inequality

If $f(x)$ is a convex function, and $g(x)$ is a concave function then

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[x]), \qquad \mathbb{E}[g(X)] \leq g(\mathbb{E}[x]).$$

- Proof:
  Let $h(x) = ax + b$ be a line tangent to $f(x)$ at the point $x = \mathbb{E}(X)$.
  Since $f(x)$ is convex, it lies above this line. Thus,

$$\mathbb{E}[f(X)] \geq \mathbb{E}[h(x)] = \mathbb{E}[ax+b] = a\mathbb{E}(X) + b = h(x = \mathbb{E}[X]) = f(\mathbb{E}[X]).$$

- Consequence: E.g., since $f(x) = x^2$ and $f(x) = 1/x$ are convex, for any variable in general

$$\mathbb{E}(X^2) \geq (\mathbb{E}[X])^2, \qquad \mathbb{E}(1/X) \geq \frac{1}{\mathbb{E}(X)}$$

Jensen's inequality

Jensen's inequality

If $f(x)$ is a convex function, and $g(x)$ is a concave function then

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[x]), \qquad \mathbb{E}[g(X)] \leq g(\mathbb{E}[x]).$$

- Proof:
  Let $h(x) = ax + b$ be a line tangent to $f(x)$ at the point $x = \mathbb{E}(X)$.
  Since $f(x)$ is convex, it lies above this line. Thus,

$$\mathbb{E}[f(X)] \geq \mathbb{E}[h(x)] = \mathbb{E}[ax + b] = a\mathbb{E}(X) + b = h(x = \mathbb{E}[X]) = f(\mathbb{E}[X]).$$

- Consequence: E.g., since $f(x) = x^2$ and $f(x) = 1/x$ are convex, for any variable in general

$$\mathbb{E}(X^2) \geq (\mathbb{E}[X])^2, \qquad \mathbb{E}(1/X) \geq \frac{1}{\mathbb{E}(X)}$$

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Jensen's inequality

Jensen's inequality

If $f(x)$ is a convex function, and $g(x)$ is a concave function then

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[x]), \qquad \mathbb{E}[g(X)] \leq g(\mathbb{E}[x]).$$

- Proof:
  Let $h(x) = ax + b$ be a line tangent to $f(x)$ at the point $x = \mathbb{E}(X)$. Since $f(x)$ is convex, it lies above this line. Thus,

$$\mathbb{E}[f(X)] \geq \mathbb{E}[h(x)] = \mathbb{E}[ax + b] = a\mathbb{E}(X) + b = h(x = \mathbb{E}[X]) = f(\mathbb{E}[X]).$$

- Consequence: E.g., since $f(x) = x^2$ and $f(x) = 1/x$ are convex, for any variable in general

$$\mathbb{E}(X^2) \geq (\mathbb{E}[X])^2, \qquad \mathbb{E}(1/X) \geq \frac{1}{\mathbb{E}(X)}$$

CONVERGENCE

Convergence

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Convergence in probability

If $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of random variables, then X_n is converging to X in probability (or weakly) if for every $\epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) \to 0.$$

This is usually denoted by $X_n \xrightarrow{P} X$.

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Convergence

Convergence in probability

If $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of random variables, then X_n is converging to X in probability (or weakly) if for every $\epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) \to 0.$$

This is usually denoted by $X_n \xrightarrow{P} X$.

Almost sure convergence

If $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of random variables, then X_n is converging to X almost surely (or almost everywhere, or with probability 1, or strongly) if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

This is usually denoted by $X_n \xrightarrow{\text{a.s.}} X$.

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Convergence

Convergence in distribution

If $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of random variable, where the corresponding CDFs are denoted by F_n, then X_n is converging to X (with CDF given by $F(x)$) in distribution if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

at all x where $F(x)$ is continuous. This is usually denoted by $X_n \xrightarrow{\mathrm{d}} X$

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Convergence

Convergence in distribution

If $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of random variable, where the corresponding CDFs are denoted by F_n, then X_n is converging to X (with CDF given by $F(x)$) in distribution if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

at all x where $F(x)$ is continuous. This is usually denoted by $X_n \xrightarrow{\mathrm{d}} X$

Convergence in mean

If $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of random variable and $r \geq 1$, then X_n is converging to X in the r-th mean (or in the L^r-th norm) if

$$\lim_{n \to \infty} \mathbb{E}(|X_n - X|^r) = 0.$$

This is usually denoted as $X_n \xrightarrow{L^r} X$.

Expectation, Inequalities, Convergence

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function

Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality

Convergence
Convergence types
Law of large numbers
Central limit theorems

Chain of implications:

$$\begin{array}{ccccc} \xrightarrow{L^s} & \underset{s>r\geq 1}{\Longrightarrow} & \xrightarrow{L^r} & & \\ & & \Downarrow & & \\ \xrightarrow{\text{a.s.}} & \Longrightarrow & \xrightarrow{P} & \Longrightarrow & \xrightarrow{d} \end{array}$$

Law of large numbers

Expectation, Inequalities, Convergence

Expectation and Variance
- Expected value
- Variance
- Sample mean and variance
- Covariance and correlation
- Covariance matrix
- Conditional expectation
- Characteristic function

Inequalities
- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality
- Cauchy–Schwarz inequality
- Jensen's inequality

Convergence
- Convergence types
- Law of large numbers
- Central limit theorems

Law of large numbers

The weak law of large number states that if $X_1, X_2, \ldots, X_n$ are IID where $\mathbb{E}(X_i) = \mu$, then

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu.$$

Law of large numbers

Expectation, Inequalities, Convergence

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function

Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality

Convergence
Convergence types
Law of large numbers
Central limit theorems

Law of large numbers

The weak law of large number states that if $X_1, X_2, \ldots, X_n$ are IID where $\mathbb{E}(X_i) = \mu$, then

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu.$$

Proof:
Assuming that $\mathbb{V}(X_i) = \sigma^2 < \infty$ we can use Chebyshev's inequality as

$$P\left(\left|\overline{X_n} - \mu\right| > \epsilon\right) \leq \frac{\mathbb{V}(\overline{X_n})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow[n\to\infty]{} 0$$

Central limit theorem

The Central limit theorem

Let $X_1, X_2, \ldots, X_n, \ldots$ be IID random variables with $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$. Then the transformed variables

$$Z_n = \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0, 1).$$

In other words,

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx.$$

Central limit theorem for heavy tailed distributions

Expectation and Variance
Expected value
Variance
Sample mean and variance
Covariance and correlation
Covariance matrix
Conditional expectation
Characteristic function
Inequalities
Markov's inequality
Chebyshev's inequality
Hoeffding's inequality
Cauchy–Schwarz inequality
Jensen's inequality
Convergence
Convergence types
Law of large numbers
Central limit theorems

Central limit theorem for heavy tailed distributions

If $X_1, X_2, \ldots, X_n$ are IID variables with a PDF such that

$$\rho(x) \approx \frac{C_0}{|x|^{\mu+1}}$$

for $|x| >> 1$ where $1 \leq \mu \leq 2$, then the transformed variable

$$Y_n = \frac{1}{n^{1/\mu}} \sum_{i=1}^{n} (X_i - \mathbb{E}(X_i))$$

will converge in distribution to a Levy distribution, $Y_n \xrightarrow{d} Y \sim \text{Levy}(\mu)$. which can be most easily specified by its characteristic function,

$$\phi_Y(t) = e^{-C|t|^{\mu}},$$

having also a heavy tail,

$$\rho_Y(y) \approx \frac{C_1}{|y|^{\mu+1}}$$

for $y >> 1$.