Mathematics for Machine Learning

Vazgen Mikayelyan

September 12, 2020

Sequence of Functions

Sequence of Functions

Let $f_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Sequence of Functions

Let $f_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Definition

We will say that f_n converges to f everywhere on X, if $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$.

Sequence of Functions

Let $f_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Definition

We will say that f_n converges to f everywhere on X, if $\lim\limits_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$.

Definition

We will say that f_n converges to f uniformly on X, if for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ holds $|f_n(x) - f(x)| < \varepsilon$ for all $x \in X$. We will write $f_n(x) \rightrightarrows f(x), x \in X$.

Sequence of Functions

Let $f_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Definition

We will say that f_n converges to f everywhere on X, if $\lim\limits_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$.

Definition

We will say that f_n converges to f uniformly on X, if for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ holds $|f_n(x) - f(x)| < \varepsilon$ for all $x \in X$. We will write $f_n(x) \rightrightarrows f(x), x \in X$.

Theorem

f_n converges uniformly on X if and only if for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $m \in \mathbb{N}$ holds $|f_{n+m}(x) - f_n(x)| < \varepsilon$ for all $x \in X$.

Series of Functions

Series of Functions

Let $u_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Series of Functions

Let $u_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Definition

We will say that the series $\sum_{n=1}^{\infty} u_n(x)$ *converges to* f *everywhere on* X*, if the sequence* $f_n(x) = \sum_{k=1}^{n} u_k(x)$ *converges to* f *everywhere on* X*.*

Series of Functions

Let $u_n : X \to \mathbb{R}, n \in \mathbb{N}$ be a sequence of functions.

Definition

We will say that the series $\sum_{n=1}^{\infty} u_n(x)$ *converges to* f *everywhere on* X, *if the sequence* $f_n(x) = \sum_{k=1}^{n} u_k(x)$ *converges to* f *everywhere on* X.

Definition

We will say that the series $\sum_{n=1}^{\infty} u_n(x)$ *converges to* f *uniformly on* X, *if the sequence* $f_n(x) = \sum_{k=1}^{n} u_k(x)$ *converges to* f *uniformly on* X.

Series of Functions

Theorem

If there exists a sequence of real numbers λ_n *such that* $|u_n(x)| \leq \lambda_n$ *for all* $x \in X$ *and* $n \in \mathbb{N}$ *and* $\sum_{n=1}^{\infty} \lambda_n < +\infty$, *then* $\sum_{n=1}^{\infty} u_n(x)$ *is uniformly convergent.*

Series of Functions

Theorem

If there exists a sequence of real numbers λ_n *such that* $|u_n(x)| \leq \lambda_n$ *for all* $x \in X$ *and* $n \in \mathbb{N}$ *and* $\sum_{n=1}^{\infty} \lambda_n < +\infty$, *then* $\sum_{n=1}^{\infty} u_n(x)$ *is uniformly convergent.*

Example

Prove that the series $\sum_{n=1}^{\infty} \frac{\sin nx}{1+n^2}$ *is uniformly convergent on* $\mathbb{R}$.

Sequence of Functions

Theorem

If $f_n \in C(X), n \in \mathbb{N}$ *and* $f_n(x) \rightrightarrows f(x), x \in X$*, then* $f \in C(X)$*.*

Sequence of Functions

Theorem

If $f_n \in C(X), n \in \mathbb{N}$ *and* $f_n(x) \rightrightarrows f(x), x \in X$*, then* $f \in C(X)$*.*

Theorem

If $f_n(x) \rightrightarrows f(x), x \in X$ *and* $\lim\limits_{x \to a} f_n(x) = c_n$ *for all* $n \in \mathbb{N}$*, then there exists finite limits* $\lim\limits_{n \to \infty} c_n = c$ *and* $\lim\limits_{x \to a} f(x) = c$*. In other words*

$$\lim_{x \to a} \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} \lim_{x \to a} f_n(x).$$

Sequence of Functions

Theorem

If $f_n \in C[a,b], n \in \mathbb{N}$ *and* $f_n(x) \rightrightarrows f(x), x \in [a,b]$, *then*

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b f_n(x)\,dx$$

Sequence of Functions

Theorem

If $f_n \in C[a,b], n \in \mathbb{N}$ *and* $f_n(x) \rightrightarrows f(x), x \in [a,b]$, *then*

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b f_n(x)\,dx$$

Theorem

If $f_n' \in C[a,b], n \in \mathbb{N}$, $f_n(x) \to f(x)$, $x \in [a,b]$ *and* $f_n'(x) \rightrightarrows g(x)$, $x \in [a,b]$, *then* $f'(x) = g(x)$.

Sequence of Functions

Theorem

If $f_n \in C[a,b], n \in \mathbb{N}$ *and* $f_n(x) \rightrightarrows f(x), x \in [a,b]$, *then*

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b f_n(x)\,dx$$

Theorem

If $f_n' \in C[a,b], n \in \mathbb{N}$, $f_n(x) \to f(x)$, $x \in [a,b]$ *and* $f_n'(x) \rightrightarrows g(x)$, $x \in [a,b]$, *then* $f'(x) = g(x)$.

Theorem

If $f \in C[a,b]$, *then there exists a sequence of polynomials* P_n, *such that*

$$P_n(x) \rightrightarrows f(x), x \in [a,b].$$

Power Series

Definition

Power series are the series of the form $\sum_{n=0}^{\infty} a_n (x - x_0)^n$.

Power Series

Definition

Power series are the series of the form $\sum_{n=0}^{\infty} a_n (x - x_0)^n$.

Theorem

If $R = \frac{1}{\overline{\lim\limits_{n\to\infty}} \sqrt[n]{|a_n|}}$, *then the series* $\sum_{n=0}^{\infty} a_n x^n$ *is convergent if* $|x| < R$ *and is divergent for* $|x| > R$.

Power Series

Definition

Power series are the series of the form $\sum_{n=0}^{\infty} a_n (x - x_0)^n$.

Theorem

If $R = \frac{1}{\overline{\lim\limits_{n\to\infty}} \sqrt[n]{|a_n|}}$, *then the series* $\sum_{n=0}^{\infty} a_n x^n$ *is convergent if* $|x| < R$ *and is divergent for* $|x| > R$.

Theorem

If $R > 0$ *then* $\sum_{n=0}^{\infty} a_n x^n \in C(-R, R)$.

Power Series

Theorem

If $R \in (0, \infty)$ *and the series* $\sum_{n=0}^{\infty} a_n R^n$ *is convergent then the series* $\sum_{n=0}^{\infty} a_n x^n$ *is uniformly convergent on* $[0, R]$*.*

Power Series

Theorem

If $f(x) = \sum_{n=0}^{\infty} a_n x^n$ *and* $R > 0$, *then* $f \in \mathcal{R}[0, a]$ *for all* $a \in (0, R)$ *and*

$$\int_0^a f(x)\, dx = \sum_{n=0}^{\infty} \frac{a_n}{n+1} a^{n+1}$$

Power Series

Theorem

If $f(x)=\sum_{n=0}^{\infty} a_n x^n$ *and* $R>0$, *then* $f\in\mathcal{R}[0,a]$ *for all* $a\in(0,R)$ *and*

$$\int_0^a f(x)\,dx=\sum_{n=0}^{\infty}\frac{a_n}{n+1}a^{n+1}$$

Theorem

If $f(x)=\sum_{n=0}^{\infty} a_n x^n$ *and* $R>0$, *then* f *is differentiable in* $(-R,R)$ *and*

$$f'(x)=\sum_{n=1}^{\infty} n a_n x^{n-1}$$

Convergence of Random Variables

Convergence of Random Variables

Definition

A sequence of random variables X_n converges in distribution to X, if their CDFs F_n converge to the CDF of X (F) at any point of continuity of F.

Convergence of Random Variables

Definition

A sequence of random variables X_n converges in distribution to X, if their CDFs F_n converge to the CDF of X (F) at any point of continuity of F.

Definition

A sequence of random variables X_n converges to X almost surely, if

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

Convergence of Random Variables

Definition

A sequence of random variables X_n *converges in probability to* X*, if for any* $\varepsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0.$$

Convergence of Random Variables

Definition

A sequence of random variables X_n *converges in probability to* X*, if for any* $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0.$$

Definition

A sequence of random variables X_n *converges to* X *in the r-th mean* $(r > 0)$*, if* $\mathbb{E}\left[|X_n|^r\right] < +\infty$ *for all* $n \in \mathbb{N}$, $\mathbb{E}\left[|X|^r\right] < +\infty$ *and*

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0.$$

Examples

Let X_n be a exponential random variable with parameter n, i.e. CDF of X_n is

$$F_n(x) = \begin{cases} 1 - e^{-nx}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

Examples

Let X_n be a exponential random variable with parameter n, i.e. CDF of X_n is

$$F_n(x) = \begin{cases} 1 - e^{-nx}, \text{ if } x \geq 0, \\ 0, \text{ if } x < 0. \end{cases}$$

Then for all $\varepsilon > 0$ we obtain

$$\lim_{n \to \infty} \mathbb{P}\left(|X_n| \geq \varepsilon\right) =$$

Examples

Let X_n be a exponential random variable with parameter n, i.e. CDF of X_n is

$$F_n(x) = \begin{cases} 1 - e^{-nx}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

Then for all $\varepsilon > 0$ we obtain

$$\lim_{n \to \infty} \mathbb{P}(|X_n| \geq \varepsilon) = \lim_{n \to \infty} (1 - F_n(\varepsilon)) =$$

Examples

Let X_n be a exponential random variable with parameter n, i.e. CDF of X_n is

$$F_n(x) = \begin{cases} 1 - e^{-nx}, \text{ if } x \geq 0, \\ 0, \text{ if } x < 0. \end{cases}$$

Then for all $\varepsilon > 0$ we obtain

$$\lim_{n\to\infty} \mathbb{P}(|X_n| \geq \varepsilon) = \lim_{n\to\infty} (1 - F_n(\varepsilon)) = \lim_{n\to\infty} e^{-n\varepsilon} =$$

Examples

Let X_n be a exponential random variable with parameter n, i.e. CDF of X_n is

$$F_n(x) = \begin{cases} 1 - e^{-nx}, \text{ if } x \geq 0, \\ 0, \text{ if } x < 0. \end{cases}$$

Then for all $\varepsilon > 0$ we obtain

$$\lim_{n\to\infty} \mathbb{P}(|X_n| \geq \varepsilon) = \lim_{n\to\infty} (1 - F_n(\varepsilon)) = \lim_{n\to\infty} e^{-n\varepsilon} = 0.$$

Examples

Let X_n be a exponential random variable with parameter n, i.e. CDF of X_n is

$$F_n(x) = \begin{cases} 1 - e^{-nx}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

Then for all $\varepsilon > 0$ we obtain

$$\lim_{n\to\infty} \mathbb{P}(|X_n| \geq \varepsilon) = \lim_{n\to\infty} (1 - F_n(\varepsilon)) = \lim_{n\to\infty} e^{-n\varepsilon} = 0.$$

So $X_n \xrightarrow{\mathbb{P}} 0$.

Theorem

If $X_n \xrightarrow{(r)} X$, *then* $X_n \xrightarrow{\mathbb{P}} X$.

Convergence of Random Variables

Theorem

If $X_n \xrightarrow{(r)} X$, *then* $X_n \xrightarrow{\mathbb{P}} X$.

Theorem

If $X_n \xrightarrow{a.s.} X$, *then* $X_n \xrightarrow{\mathbb{P}} X$.

Convergence of Random Variables

Theorem

If $X_n \xrightarrow{(r)} X$, *then* $X_n \xrightarrow{\mathbb{P}} X$.

Theorem

If $X_n \xrightarrow{a.s.} X$, *then* $X_n \xrightarrow{\mathbb{P}} X$.

Theorem

If $X_n \xrightarrow{\mathbb{P}} X$ *then there exists* $k_n \in \mathbb{N}$ *such that* $X_{k_n} \xrightarrow{a.s.} X$.

Convergence of Random Variables

Theorem

If $X_n \xrightarrow{(r)} X$, *then* $X_n \xrightarrow{\mathbb{P}} X$.

Theorem

If $X_n \xrightarrow{a.s.} X$, *then* $X_n \xrightarrow{\mathbb{P}} X$.

Theorem

If $X_n \xrightarrow{\mathbb{P}} X$ *then there exists* $k_n \in \mathbb{N}$ *such that* $X_{k_n} \xrightarrow{a.s.} X$.

Theorem

If $X_n \xrightarrow{\mathbb{P}} X$, *then* $X_n \xrightarrow{D} X$.

Convergence of Random Variables

Theorem

If $X_n \xrightarrow{(r)} X$, *then* $X_n \xrightarrow{\mathbb{P}} X$.

Theorem

If $X_n \xrightarrow{a.s.} X$, *then* $X_n \xrightarrow{\mathbb{P}} X$.

Theorem

If $X_n \xrightarrow{\mathbb{P}} X$ *then there exists* $k_n \in \mathbb{N}$ *such that* $X_{k_n} \xrightarrow{a.s.} X$.

Theorem

If $X_n \xrightarrow{\mathbb{P}} X$, *then* $X_n \xrightarrow{D} X$.

Theorem

$X_n \xrightarrow{\mathbb{P}} c$ *iff* $X_n \xrightarrow{D} c$.

Examples

Let $\Omega = [0, 1]$, P is the Lebesgue measure and $\mathcal{F}$ is the σ-algebra of all measurable sets.

Examples

Let $\Omega = [0,1]$, P is the Lebesgue measure and $\mathcal{F}$ is the σ-algebra of all measurable sets. Denote

$$X_n(\omega) = \begin{cases} e^n, \text{ if } 0 \le \omega \le \frac{1}{n}, \\ 0, \text{ if } \frac{1}{n} < \omega \le 1. \end{cases}$$

Examples

Let $\Omega = [0,1]$, P is the Lebesgue measure and $\mathcal{F}$ is the σ-algebra of all measurable sets. Denote

$$X_n(\omega) = \begin{cases} e^n, & \text{if } 0 \leq \omega \leq \frac{1}{n}, \\ 0, & \text{if } \frac{1}{n} < \omega \leq 1. \end{cases}$$

For any $\omega \in (0,1]$ there exists $n_0 \in \mathbb{N}$ such that $X_n(\omega) = 0$ for $n \geq n_0$.

Examples

Let $\Omega = [0, 1]$, P is the Lebesgue measure and $\mathcal{F}$ is the σ-algebra of all measurable sets. Denote

$$X_n(\omega) = \begin{cases} e^n, \text{ if } 0 \leq \omega \leq \frac{1}{n}, \\ 0, \text{ if } \frac{1}{n} < \omega \leq 1. \end{cases}$$

For any $\omega \in (0, 1]$ there exists $n_0 \in \mathbb{N}$ such that $X_n(\omega) = 0$ for $n \geq n_0$. Therefore $X_n(\omega) \to 0$, for all $\omega \in (0, 1]$, i.e $X_n \xrightarrow{a.s.} 0$.

Examples

Let $\Omega = [0,1]$, P is the Lebesgue measure and $\mathcal{F}$ is the σ-algebra of all measurable sets. Denote

$$X_n(\omega) = \begin{cases} e^n, \text{ if } 0 \le \omega \le \frac{1}{n}, \\ 0, \text{ if } \frac{1}{n} < \omega \le 1. \end{cases}$$

For any $\omega \in (0,1]$ there exists $n_0 \in \mathbb{N}$ such that $X_n(\omega) = 0$ for $n \ge n_0$.
Therefore $X_n(\omega) \to 0$, for all $\omega \in (0,1]$, i.e $X_n \xrightarrow{a.s.} 0$.
For any $\varepsilon > 0$ we have

$$\mathbb{P}(X_n \ge \varepsilon) \le \frac{1}{n}, n \in \mathbb{N},$$

Examples

Let $\Omega = [0,1]$, P is the Lebesgue measure and $\mathcal{F}$ is the σ-algebra of all measurable sets. Denote

$$X_n(\omega) = \begin{cases} e^n, \text{ if } 0 \leq \omega \leq \frac{1}{n}, \\ 0, \text{ if } \frac{1}{n} < \omega \leq 1. \end{cases}$$

For any $\omega \in (0,1]$ there exists $n_0 \in \mathbb{N}$ such that $X_n(\omega) = 0$ for $n \geq n_0$.
Therefore $X_n(\omega) \to 0$, for all $\omega \in (0,1]$, i.e $X_n \xrightarrow{a.s.} 0$.
For any $\varepsilon > 0$ we have

$$\mathbb{P}(X_n \geq \varepsilon) \leq \frac{1}{n}, n \in \mathbb{N},$$

therefore $X_n \xrightarrow{\mathbb{P}} 0$.

Examples

Let $\Omega = [0, 1]$, P is the Lebesgue measure and $\mathcal{F}$ is the σ-algebra of all measurable sets. Denote

$$X_n(\omega) = \begin{cases} e^n, \text{ if } 0 \le \omega \le \frac{1}{n}, \\ 0, \text{ if } \frac{1}{n} < \omega \le 1. \end{cases}$$

For any $\omega \in (0, 1]$ there exists $n_0 \in \mathbb{N}$ such that $X_n(\omega) = 0$ for $n \ge n_0$.
Therefore $X_n(\omega) \to 0$, for all $\omega \in (0, 1]$, i.e $X_n \xrightarrow{a.s.} 0$.
For any $\varepsilon > 0$ we have

$$\mathbb{P}(X_n \ge \varepsilon) \le \frac{1}{n}, n \in \mathbb{N},$$

therefore $X_n \xrightarrow{\mathbb{P}} 0$. Note that

$$\mathbb{E}[|X_n|^r] = \frac{e^{rn}}{n} \to +\infty, r > 0.$$

Examples

Let $\Omega = [0, 1]$, P is the Lebesgue measure and $\mathcal{F}$ is the σ-algebra of all measurable sets. Denote

$$X_n(\omega) = \begin{cases} e^n, \text{ if } 0 \le \omega \le \frac{1}{n}, \\ 0, \text{ if } \frac{1}{n} < \omega \le 1. \end{cases}$$

For any $\omega \in (0, 1]$ there exists $n_0 \in \mathbb{N}$ such that $X_n(\omega) = 0$ for $n \ge n_0$.
Therefore $X_n(\omega) \to 0$, for all $\omega \in (0, 1]$, i.e $X_n \xrightarrow{a.s.} 0$.
For any $\varepsilon > 0$ we have

$$\mathbb{P}(X_n \ge \varepsilon) \le \frac{1}{n}, n \in \mathbb{N},$$

therefore $X_n \xrightarrow{\mathbb{P}} 0$. Note that

$$\mathbb{E}[|X_n|^r] = \frac{e^{rn}}{n} \to +\infty, r > 0.$$

therefore $X_n \overset{(r)}{\not\to} 0$.

Examples

Let $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space, such that $\mathbb{P}\left(A\right) = \mathbb{P}\left(\overline{A}\right) = 0.5$ for some $A \in \mathcal{F}$.

Examples

Let $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space, such that $\mathbb{P}(A) = \mathbb{P}\left(\overline{A}\right) = 0.5$ for some $A \in \mathcal{F}$. Denote

$$X_n(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ -1, & \text{if } \omega \in \overline{A} \end{cases}$$

and $X = -X_1$.

Examples

Let $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space, such that $\mathbb{P}(A) = \mathbb{P}\left(\overline{A}\right) = 0.5$ for some $A \in \mathcal{F}$. Denote

$$X_n(\omega) = \begin{cases} 1, \text{ if } \omega \in A, \\ -1, \text{ if } \omega \in \overline{A} \end{cases}$$

and $X = -X_1$. It is not difficult to see that

$$F_n(x) = F(x) = \begin{cases} 0, \text{ if } x < -1, \\ 0.5, \text{ if } x \in [-1, 1), \\ 1, \text{ if } x \geq 1. \end{cases}$$

Examples

Let $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space, such that $\mathbb{P}(A) = \mathbb{P}\left(\overline{A}\right) = 0.5$ for some $A \in \mathcal{F}$. Denote

$$X_n(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ -1, & \text{if } \omega \in \overline{A} \end{cases}$$

and $X = -X_1$. It is not difficult to see that

$$F_n(x) = F(x) = \begin{cases} 0, & \text{if } x < -1, \\ 0.5, & \text{if } x \in [-1, 1), \\ 1, & \text{if } x \geq 1. \end{cases}$$

However $|X_n - X| = 2$.

Examples

Let $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space, such that $\mathbb{P}(A) = \mathbb{P}\left(\overline{A}\right) = 0.5$ for some $A \in \mathcal{F}$. Denote

$$X_n(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ -1, & \text{if } \omega \in \overline{A} \end{cases}$$

and $X = -X_1$. It is not difficult to see that

$$F_n(x) = F(x) = \begin{cases} 0, & \text{if } x < -1, \\ 0.5, & \text{if } x \in [-1, 1), \\ 1, & \text{if } x \geq 1. \end{cases}$$

However $|X_n - X| = 2$. So $X_n \xrightarrow{D} X$, but $X_n \overset{\mathbb{P}}{\not\to} X$.

The Weak Law of Large Numbers

The Weak Law of Large Numbers

Markov's theorem

Let X_n be a sequence of random variables, each having a finite mean $(\mathbb{E}[X_k] = a_k)$ and variance. If

$$\lim_{n\to\infty} \frac{1}{n^2} Var\left(\sum_{k=1}^{n} X_k\right) = 0, \text{ then } \frac{1}{n}\sum_{k=1}^{n}(X_k - a_k) \xrightarrow{\mathbb{P}} 0.$$

The Weak Law of Large Numbers

Markov's theorem

Let X_n be a sequence of random variables, each having a finite mean $(\mathbb{E}[X_k] = a_k)$ and variance. If

$$\lim_{n\to\infty} \frac{1}{n^2} Var\left(\sum_{k=1}^{n} X_k\right) = 0, \text{ then } \frac{1}{n}\sum_{k=1}^{n}(X_k - a_k) \xrightarrow{\mathbb{P}} 0.$$

Corollary

If X_n is a sequence of independent and identically distributed random variables, each having a finite mean $(\mathbb{E}[X_1] = a)$ and variance, then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\mathbb{P}} a.$$