YSU ASDS, Statistics, Fall 2019

Lecture 12

Michael Poghosyan

28 Sep 2019

Probability Reminder

Contents

Last Lecture ReCap

- Give the definitions and examples for each Distribution

Supplement for Correlations, Just a link

See the How to handle correlated Features? at Kaggle.

Sequences of R.V.s, Examples

Let $X_1, X_2, ..., X_n, ...$ be a sequence of r.v. on the same Probability Space.

Sequences of R.V.s, Examples

Let $X_1, X_2, ..., X_n, ...$ be a sequence of r.v. on the same Probability Space.

Examples:

- We toss a coin, infinitely many times, and let X_k be 0, it the k-th toss resulted in Heads, and $X_k = 1$ otherwise.

Sequences of R.V.s, Examples

Let $X_1, X_2, ..., X_n, ...$ be a sequence of r.v. on the same Probability Space.

Examples:

- We toss a coin, infinitely many times, and let X_k be 0, it the k-th toss resulted in Heads, and $X_k = 1$ otherwise.
- Let X_k be the Closing price for day k calculated from today for the AMZN Stock.

Sequences of R.V.s, Examples

Let $X_1, X_2, ..., X_n, ...$ be a sequence of r.v. on the same Probability Space.

Examples:

- We toss a coin, infinitely many times, and let X_k be 0, it the k-th toss resulted in Heads, and $X_k = 1$ otherwise.
- Let X_k be the Closing price for day k calculated from today for the AMZN Stock.
- Let X_k be the height (in cm) of the k-th person I will meet tomorrow.

Sequences of R.V.s, Examples

Let $X_1, X_2, ..., X_n, ...$ be a sequence of r.v. on the same Probability Space.

Examples:

- We toss a coin, infinitely many times, and let X_k be 0, it the k-th toss resulted in Heads, and $X_k = 1$ otherwise.
- Let X_k be the Closing price for day k calculated from today for the AMZN Stock.
- Let X_k be the height (in cm) of the k-th person I will meet tomorrow.
- Let X_k be the number of downloads for the Supper-Pupper inc. mobile app for the day k.

Sequences of R.V.s, Examples

Let $X_1, X_2, ..., X_n, ...$ be a sequence of r.v. on the same Probability Space.

Examples:

- We toss a coin, infinitely many times, and let X_k be 0, it the k-th toss resulted in Heads, and $X_k = 1$ otherwise.
- Let X_k be the Closing price for day k calculated from today for the AMZN Stock.
- Let X_k be the height (in cm) of the k-th person I will meet tomorrow.
- Let X_k be the number of downloads for the Supper-Pupper inc. mobile app for the day k.
- Let X_k be the blood pressure for the patient k for some clinic.

Sequences of R.V.s, Examples

Let $X_1, X_2, ..., X_n, ...$ be a sequence of r.v. on the same Probability Space.

Examples:

- We toss a coin, infinitely many times, and let X_k be 0, it the k-th toss resulted in Heads, and $X_k = 1$ otherwise.
- Let X_k be the Closing price for day k calculated from today for the AMZN Stock.
- Let X_k be the height (in cm) of the k-th person I will meet tomorrow.
- Let X_k be the number of downloads for the Supper-Pupper inc. mobile app for the day k.
- Let X_k be the blood pressure for the patient k for some clinic.

I know, almost all examples are examples of *finite* sequences, but for theoretical (and practical) reasons we can assume they are infinite.

Sequences of R.V.s, Motivation

In the rest of the course we will work a lot with sequences of r.v.s. We will construct Statistics for the unknown parameters of Distribution, something to estimate that parameters.

Sequences of R.V.s, Motivation

In the rest of the course we will work a lot with sequences of r.v.s. We will construct Statistics for the unknown parameters of Distribution, something to estimate that parameters.

And one of the important questions will be: is our Statistic good enough to estimate the parameter? The point is that since the parameter value is unknown, we need to have some theoretical guarantees that our estimators are working well.

Sequences of R.V.s, Motivation

In the rest of the course we will work a lot with sequences of r.v.s. We will construct Statistics for the unknown parameters of Distribution, something to estimate that parameters.

And one of the important questions will be: is our Statistic good enough to estimate the parameter? The point is that since the parameter value is unknown, we need to have some theoretical guarantees that our estimators are working well.

So we will use different notions of r.v. sequence convergence to assess the quality of our estimator, Statistics.

Sequences of R.V.s, Motivation

In the rest of the course we will work a lot with sequences of r.v.s. We will construct Statistics for the unknown parameters of Distribution, something to estimate that parameters.

And one of the important questions will be: is our Statistic good enough to estimate the parameter? The point is that since the parameter value is unknown, we need to have some theoretical guarantees that our estimators are working well.

So we will use different notions of r.v. sequence convergence to assess the quality of our estimator, Statistics.

Also, we will use a lot the CLT, say, to construct Confidence Intervals and design Tests for Hypotheses, so we need to know the CLT, and CLT is about the limit of a r.v. sequence.

Convergence of a Sequence of r.v.

There are different notions of a convergence for a r.v. sequence.

[1] And we have different notions for the convergence of functional sequences like pointwise, uniform, a.e., L^p, ... convergences

Convergence of a Sequence of r.v.

There are different notions of a convergence for a r.v. sequence.

This is because, a sequence of r.v., besides being just a sequence of functions[1], also encloses randomness behind, and we need to deal with that randomness.

[1] And we have different notions for the convergence of functional sequences like pointwise, uniform, a.e., L^p, ... convergences

Convergence of a Sequence of r.v.

There are different notions of a convergence for a r.v. sequence.

This is because, a sequence of r.v., besides being just a sequence of functions[1], also encloses randomness behind, and we need to deal with that randomness.

Say, what it means for r.v.s X and Y that X is close to Y?

[1] And we have different notions for the convergence of functional sequences like pointwise, uniform, a.e., L^p, ... convergences

Convergence of a Sequence of r.v.

There are different notions of a convergence for a r.v. sequence.

This is because, a sequence of r.v., besides being just a sequence of functions[1], also encloses randomness behind, and we need to deal with that randomness.

Say, what it means for r.v.s X and Y that X is close to Y?

Aha, that's the problem - it is not so easy to define the closedness $\ddot\smile$

[1] And we have different notions for the convergence of functional sequences like pointwise, uniform, a.e., L^p, ... convergences

Convergence a.s.

Assume X_n is a sequence of r.v. and X is a r.v. over the same Probability Space.

Convergence a.s.

Assume X_n is a sequence of r.v. and X is a r.v. over the same Probability Space.

Definition: We will say that $X_n \to X$ **almost sure**, and we will write $X_n \to X$ a.s. or $X_n \xrightarrow{a.s.} X$, if

$$\mathbb{P}\Big(\omega \in \Omega : \lim_{n \to +\infty} X_n(\omega) = X(\omega)\Big) = 1,$$

or, for short,

$$\mathbb{P}\Big(X_n \to X\Big) = 1$$

Convergence a.s.

Assume X_n is a sequence of r.v. and X is a r.v. over the same Probability Space.

Definition: We will say that $X_n \to X$ **almost sure**, and we will write $X_n \to X$ a.s. or $X_n \xrightarrow{a.s.} X$, if

$$\mathbb{P}\Big(\omega \in \Omega : \lim_{n \to +\infty} X_n(\omega) = X(\omega)\Big) = 1,$$

or, for short,

$$\mathbb{P}\Big(X_n \to X\Big) = 1$$

Equivalently, we can write

$$X_n \xrightarrow{a.s.} X \qquad \text{iff} \qquad \mathbb{P}\Big(X_n \not\to X\Big) = 0.$$

Convergence in Probability

Definition: We will say that $X_n \to X$ **in Probability**, and we will write $X_n \xrightarrow{\mathbb{P}} X$, if

$$\text{for any } \varepsilon > 0, \ \mathbb{P}\Big(|X_n - X| \geq \varepsilon\Big) \to 0, \qquad \text{when} \quad n \to \infty.$$

Convergence in Probability

Definition: We will say that $X_n \to X$ **in Probability**, and we will write $X_n \xrightarrow{\mathbb{P}} X$, if

$$\text{for any } \varepsilon > 0, \ \mathbb{P}\Big(|X_n - X| \geq \varepsilon\Big) \to 0, \qquad \text{when} \quad n \to \infty.$$

Equivalently, we can write

$$X_n \xrightarrow{\mathbb{P}} X \qquad \text{iff} \qquad \mathbb{P}\Big(|X_n - X| < \varepsilon\Big) \to 1 \text{ for any } \varepsilon > 0.$$

Convergence in the Mean Square Sence

Definition: We will say that $X_n \to X$ **in the Quadratic Mean Sense or in L^2 (or in the Mean Square Sense)**, and we will write $X_n \xrightarrow{L^2} X$ or $X_n \xrightarrow{qm} X$, if

$$MSE(X_n, X) = \mathbb{E}\Big((X_n - X)^2\Big) \to 0, \qquad \text{when} \quad n \to \infty.$$

Convergence in the Mean Square Sence

Definition: We will say that $X_n \to X$ **in the Quadratic Mean Sense or in L^2 (or in the Mean Square Sense)**, and we will write $X_n \xrightarrow{L^2} X$ or $X_n \xrightarrow{qm} X$, if

$$MSE(X_n, X) = \mathbb{E}\Big((X_n - X)^2\Big) \to 0, \qquad \text{when} \quad n \to \infty.$$

Here $MSE(X_n, X)$ is the *Mean Square Error* (of the approximation of X by X_n).

Convergence in Distributions

Now we assume that X_n and X are arbitrary r.v.'s, not necessarily defined on the same probability space, and $F_{X_n}(x)$ and $F_X(x)$ are their CDF's, respectively.

Convergence in Distributions

Now we assume that X_n and X are arbitrary r.v.'s, not necessarily defined on the same probability space, and $F_{X_n}(x)$ and $F_X(x)$ are their CDF's, respectively.

Definition: We will say that $X_n \to X$ **in Distribution (or in Law)**, and we will write $X_n \xrightarrow{D} X$, if

$F_{X_n}(x) \to F_X(x) \quad \text{as} \quad n \to \infty$ at any point of continuity x of $F_X(x)$.

Convergence in Distributions

Now we assume that X_n and X are arbitrary r.v.'s, not necessarily defined on the same probability space, and $F_{X_n}(x)$ and $F_X(x)$ are their CDF's, respectively.

Definition: We will say that $X_n \to X$ **in Distribution (or in Law)**, and we will write $X_n \xrightarrow{D} X$, if

$F_{X_n}(x) \to F_X(x) \quad \text{as} \quad n \to \infty$ at any point of continuity x of $F_X(x)$.

Remark: This is equivalent to saying that for (almost) any subsets $A \subset \mathbb{R}$

$$\mathbb{P}(X_n \in A) \to \mathbb{P}(X \in A).$$

Example

Example: Assume I am tossing a fair coin infinitely many times (independently), and let X_n be 1 if Head shows in the n-th trial, and 0 otherwise. So the Distribution of X_n is

$$X_n \sim$$

Example

Example: Assume I am tossing a fair coin infinitely many times (independently), and let X_n be 1 if Head shows in the n-th trial, and 0 otherwise. So the Distribution of X_n is

$$X_n \sim Bernoulli(0.5).$$

Example

Example: Assume I am tossing a fair coin infinitely many times (independently), and let X_n be 1 if Head shows in the n-th trial, and 0 otherwise. So the Distribution of X_n is

$$X_n \sim Bernoulli(0.5).$$

- Is X_n convergent in the sense of Distributions ?
- Is X_n convergent in the Probability sense ?
- Is X_n convergent in the MS sense ?
- Is X_n convergent in the a.s. sense?

Example

Example: Assume I am tossing a fair coin infinitely many times (independently), and let X_n be 1 if Head shows in the n-th trial, and 0 otherwise. So the Distribution of X_n is

$$X_n \sim Bernoulli(0.5).$$

- Is X_n convergent in the sense of Distributions ?
- Is X_n convergent in the Probability sense ?
- Is X_n convergent in the MS sense ?
- Is X_n convergent in the a.s. sense?

Solution: OTB. Not a good/correct example, impossible to answer to the questions except to the first one.

Example

Example: Assume X_n is a Discrete r.v. with the following PMF, defined on the same Probability Space:

X_n	$3+\frac{1}{n^2}$	n
$\mathbb{P}(X_n = x)$	$1-\dfrac{1}{n}$	$\dfrac{1}{n}$.

Example

Example: Assume X_n is a Discrete r.v. with the following PMF, defined on the same Probability Space:

X_n	$3 + \frac{1}{n^2}$	n
$\mathbb{P}(X_n = x)$	$1 - \dfrac{1}{n}$	$\dfrac{1}{n}$.

Which of the followings are true (use only the definitions):

- $X_n \xrightarrow{\mathbb{P}} 3$;
- $X_n \xrightarrow{qm} 3$;
- $X_n \xrightarrow{D} 3$?

Solution: OTB

Example

Example: Assume

$$X_n \sim Unif\left[0, \frac{1}{n}\right]$$

and X_n are defined on the same Probability Space.

Example

Example: Assume

$$X_n \sim Unif\left[0, \frac{1}{n}\right]$$

and X_n are defined on the same Probability Space. Which of the followings are true (use only the definitions):

- $X_n \xrightarrow{\mathbb{P}} 0$;
- $X_n \xrightarrow{qm} 0$;
- $X_n \xrightarrow{D} 0$?

Solution: OTB