AUA CS108, Statistics, Fall 2020

Lecture 19

Michael Poghosyan

09 Oct 2020

Contents

Uniform Distribution

- Parameters: a, b ($a < b$)

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;
- Support: $[a, b]$

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;
- Support: $[a, b]$
- PDF:

$$f(x) = f(x|a, b) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & otherwise \end{cases}$$

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;
- Support: $[a, b]$
- PDF:

$$f(x) = f(x|a, b) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$.

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;
- Support: $[a, b]$
- PDF:

$$f(x) = f(x|a, b) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$.
- Models: Usually we think about the Uniform Distribution when talking about *picking a random number from an interval*

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim \mathit{Unif}[a, b]$;
- Support: $[a, b]$
- PDF:

$$f(x) = f(x|a, b) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & \mathit{otherwise} \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{a+b}{2}$, $\mathit{Var}(X) = \frac{(b-a)^2}{12}$.
- Models: Usually we think about the Uniform Distribution when talking about *picking a random number from an interval*
- **R** name: `unif` with the parameters `min = 0` and `max = 1`

Uniform Distribution

- Parameters: a, b $(a < b)$
- Notation: $X \sim Unif[a, b]$;
- Support: $[a, b]$
- PDF:

$$f(x) = f(x|a, b) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$.
- Models: Usually we think about the Uniform Distribution when talking about *picking a random number from an interval*
- **R** name: `unif` with the parameters `min = 0` and `max = 1`
- Example:

```r
runif(10, min = 2, max = 5)
```

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim \mathit{Exp}(\lambda)$ (or $\widetilde{\mathit{Exp}}(\beta)$);

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $\widetilde{Exp}(\beta)$);
- Support: $[0, +\infty)$

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim \mathit{Exp}(\lambda)$ (or $\widetilde{\mathit{Exp}}(\beta)$);
- Support: $[0, +\infty)$
- PDF:

$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & \mathit{otherwise} \end{cases}$$

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $\widetilde{Exp}(\beta)$);
- Support: $[0, +\infty)$
- PDF:
$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$
- Mean and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $\widetilde{Exp}(\beta)$);
- Support: $[0, +\infty)$
- PDF:
$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$
- Mean and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.
- Models: time elapsed until the occurrence of certain event, or the time between events (waiting times), when that time is random.

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $\widetilde{Exp}(\beta)$);
- Support: $[0, +\infty)$
- PDF:
$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$
- Mean and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.
- Models: time elapsed until the occurrence of certain event, or the time between events (waiting times), when that time is random. λ is the average "arrival rate", the reciprocal of the average time between the events,
$$\lambda = \frac{1}{\text{average time between events}}.$$

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $\widetilde{Exp}(\beta)$);
- Support: $[0, +\infty)$
- PDF:
$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$
- Mean and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.
- Models: time elapsed until the occurrence of certain event, or the time between events (waiting times), when that time is random. λ is the average "arrival rate", the reciprocal of the average time between the events,
$$\lambda = \frac{1}{\text{average time between events}}.$$
- **R** name: `exp` with the parameter `rate = 1`

Exponential Distribution

- Parameter: $\lambda > 0$ (rate) (or, sometimes, $\beta = \frac{1}{\lambda}$, scale)
- Notation: $X \sim Exp(\lambda)$ (or $\widetilde{Exp}(\beta)$);
- Support: $[0, +\infty)$
- PDF:

$$f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

- Mean and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.
- Models: time elapsed until the occurrence of certain event, or the time between events (waiting times), when that time is random. λ is the average "arrival rate", the reciprocal of the average time between the events,
$$\lambda = \frac{1}{\text{average time between events}}.$$
- **R** name: `exp` with the parameter `rate = 1`

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mean and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$.

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mean and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$.
- Models: A lot of things. The idea is the following:

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mean and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$.
- Models: A lot of things. The idea is the following: usually, when we are estimating something, we give that estimate in the form:

$$\text{estimate} = \text{true (unknown) value} + \text{error}$$

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:
$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
- Mean and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$.
- Models: A lot of things. The idea is the following: usually, when we are estimating something, we give that estimate in the form:
$$\text{estimate} = \text{true (unknown) value} + \text{error}$$
And this `error` is, supposedly, close to 0.

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:
$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
- Mean and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$.
- Models: A lot of things. The idea is the following: usually, when we are estimating something, we give that estimate in the form:
$$\text{estimate} = \text{true (unknown) value} + \text{error}$$
And this `error` is, supposedly, close to 0. And the distribution of `error`, which is a random term, is usually taken as Normal (with the Mean $= 0$).

Normal (Gaussian) Distribution

- Parameters: μ (mean) and σ^2 (variance);
- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$;
- Support: $\mathbb{R}$
- PDF:
$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
- Mean and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$.
- Models: A lot of things. The idea is the following: usually, when we are estimating something, we give that estimate in the form:
$$\text{estimate} = \text{true (unknown) value} + \text{error}$$
And this `error` is, supposedly, close to 0. And the distribution of `error`, which is a random term, is usually taken as Normal (with the Mean $= 0$). Also it is important because of the CLT

Normal (Gaussian) Distribution

- **R** name: `norm` with the parameters `mean = 0`, `sd = 1`

Normal (Gaussian) Distribution

- **R** name: `norm` with the parameters `mean = 0`, `sd = 1`
- Example:

```
rnorm(10, mean = 2, sd = 3)
```

```
##  [1]  0.9455538  1.6175023  2.3491199  6.6423999  3.3385
##  [7]  8.5884579  1.2447808 -5.1912542 -1.5678284
```

Normal (Gaussian) Distribution

- **R** name: `norm` with the parameters `mean = 0`, `sd = 1`
- Example:

```
rnorm(10, mean = 2, sd = 3)
```

```
##  [1]  0.9455538  1.6175023  2.3491199  6.6423999  3.3385
##  [7]  8.5884579  1.2447808 -5.1912542 -1.5678284
```

Note:

- **R** is using $\mathcal{N}(mean, sd)$ format
- **In Math** we are using the $\mathcal{N}(mean, variance)$ format

Normal (Gaussian) Distribution

- **R** name: `norm` with the parameters `mean = 0`, `sd = 1`
- Example:

```
rnorm(10, mean = 2, sd = 3)
```

```
##  [1]  0.9455538  1.6175023  2.3491199  6.6423999  3.3385
##  [7]  8.5884579  1.2447808 -5.1912542 -1.5678284
```

Note:

- **R** is using $\mathcal{N}(mean, sd)$ format
- **In Math** we are using the $\mathcal{N}(mean, variance)$ format

So if you want to generate a sample of size 100 from $\mathcal{N}(2, 9)$, use the command `rnorm(100, mean = 2, sd = 3)`.

Normal (Gaussian) Distribution

Additional Properties:

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$ and

$$\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) =$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Normal (Gaussian) Distribution

- If $Z \sim \mathcal{N}(0, 1)$, then $X = \mu + \sigma \cdot Z \sim \mathcal{N}(\mu, \sigma^2)$

Normal (Gaussian) Distribution

- If $Z \sim \mathcal{N}(0, 1)$, then $X = \mu + \sigma \cdot Z \sim \mathcal{N}(\mu, \sigma^2)$
- If $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ for $k = 1, ..., n$ are independent, then $X_1 + X_2 + ... + X_n \sim \mathcal{N}(\mu_1 + \mu_2 + ... + \mu_n, \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2)$.

Normal (Gaussian) Distribution

- If $Z \sim \mathcal{N}(0,1)$, then $X = \mu + \sigma \cdot Z \sim \mathcal{N}(\mu, \sigma^2)$
- If $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ for $k = 1, ..., n$ are independent, then $X_1 + X_2 + ... + X_n \sim \mathcal{N}(\mu_1 + \mu_2 + ... + \mu_n, \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2)$.
- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{P}(\mu - \sigma < X < \mu + \sigma) \approx 0.6827,$$

$$\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.9545$$

and

$$\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.9973.$$

Additions

- See many other Distributions at Wiki or in different textbooks.

Additions

- See many other Distributions at Wiki or in different textbooks.
- Relationships among probability distributions: Wiki

Additions

- See many other Distributions at Wiki or in different textbooks.
- Relationships among probability distributions: Wiki
- Another page for the Relationship: L. Leemis Page

Convergence of a sequence of r.v.s

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

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