CS 107, Probability, Spring 2019
Lecture 43

Michael Poghosyan

AUA

- 3 lectures

Ta-da-da-daaaam! Quiz time!

Content

- The Central Limit Theorem
- Expectation and Variance for Important Distributions
- Intro to Markov Chains

The Central Limit Theorem

CLT gives more info about the Distribution of S_n and $\overline{X}_n$:

The Central Limit Theorem, CLT

Assume $X_1, X_2, ..., X_n$ are IID with finite Expectation $\mu = \mathbb{E}(X_1)$ and Variance $\sigma^2 = Var(X_1)$. We Standardize S_n (or $\overline{X}_n$):

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma}$$

$$\left(Z_n = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} = \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \right).$$

Then, for any subset $A \subset \mathbb{R}$,

$$\mathbb{P}(Z_n \in A) \to \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Example:

Example: Assume I have a Piggy Bank.

Example:

Example: Assume I have a Piggy Bank.

Example:

Example: Assume I have a Piggy Bank.
I am collecting coins worth 500 AMD in my Piggy.

Example:

Example: Assume I have a Piggy Bank.
I am collecting coins worth 500 AMD in my Piggy. The mean weight of a 500 AMD coin is 5 gr with Standard Deviation 0.1 gr.

Example:

Example: Assume I have a Piggy Bank.
I am collecting coins worth 500 AMD in my Piggy. The mean weight of a 500 AMD coin is 5 gr with Standard Deviation 0.1 gr. I have collected 250 coins.

Example:

Example: Assume I have a Piggy Bank.
I am collecting coins worth 500 AMD in my Piggy. The mean weight of a 500 AMD coin is 5 gr with Standard Deviation 0.1 gr. I have collected 250 coins. Assume W is the weight, in grams, of all that 250 coins.

Example:

Example: Assume I have a Piggy Bank.
I am collecting coins worth 500 AMD in my Piggy. The mean weight of a 500 AMD coin is 5 gr with Standard Deviation 0.1 gr. I have collected 250 coins. Assume W is the weight, in grams, of all that 250 coins. Clearly, the average weight of all coins is $250 \cdot 5gr = 1250gr$, i.e., $\mathbb{E}(W) = 1250gr$.

Example:

Example: Assume I have a Piggy Bank.
I am collecting coins worth 500 AMD in my Piggy. The mean weight of a 500 AMD coin is 5 gr with Standard Deviation 0.1 gr. I have collected 250 coins. Assume W is the weight, in grams, of all that 250 coins. Clearly, the average weight of all coins is $250 \cdot 5gr = 1250gr$, i.e., $\mathbb{E}(W) = 1250gr$.

- Calculate, approximately,

$$\mathbb{P}(W > 1200);$$

Example:

Example: Assume I have a Piggy Bank.
I am collecting coins worth 500 AMD in my Piggy. The mean weight of a 500 AMD coin is 5 gr with Standard Deviation 0.1 gr. I have collected 250 coins. Assume W is the weight, in grams, of all that 250 coins. Clearly, the average weight of all coins is $250 \cdot 5gr = 1250gr$, i.e., $\mathbb{E}(W) = 1250gr$.

- Calculate, approximately,

$$\mathbb{P}(W > 1200);$$

- Calculate, approximately,

$$\mathbb{P}(1100 \leq W < 1300);$$

Example:

Example: Assume I have a Piggy Bank.
I am collecting coins worth 500 AMD in my Piggy. The mean weight of a 500 AMD coin is 5 gr with Standard Deviation 0.1 gr. I have collected 250 coins. Assume W is the weight, in grams, of all that 250 coins. Clearly, the average weight of all coins is $250 \cdot 5gr = 1250gr$, i.e., $\mathbb{E}(W) = 1250gr$.

- Calculate, approximately,
$$\mathbb{P}(W > 1200);$$
- Calculate, approximately,
$$\mathbb{P}(1100 \leq W < 1300);$$
- Guess what I could do with the collected amount $\ddot\smile$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) =$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) =$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$
- If $X \sim Binom(n, p)$, then $\mathbb{E}(X) =$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$
- If $X \sim Binom(n, p)$, then $\mathbb{E}(X) = np$, $Var(X) =$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$
- If $X \sim Binom(n, p)$, then $\mathbb{E}(X) = np$, $Var(X) = np(1-p)$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$
- If $X \sim Binom(n, p)$, then $\mathbb{E}(X) = np$, $Var(X) = np(1-p)$
  Idea: $X = X_1 + X_2 + ... + X_n$, where $X_k \sim Bernoulli(p)$ are IID

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$
- If $X \sim Binom(n, p)$, then $\mathbb{E}(X) = np$, $Var(X) = np(1-p)$
  Idea: $X = X_1 + X_2 + ... + X_n$, where $X_k \sim Bernoulli(p)$ are IID
- If $X \sim Geom(p)$, then $\mathbb{E}(X) =$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$
- If $X \sim Binom(n, p)$, then $\mathbb{E}(X) = np$, $Var(X) = np(1-p)$
  Idea: $X = X_1 + X_2 + ... + X_n$, where $X_k \sim Bernoulli(p)$ are IID
- If $X \sim Geom(p)$, then $\mathbb{E}(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$
- If $X \sim Binom(n, p)$, then $\mathbb{E}(X) = np$, $Var(X) = np(1-p)$
  Idea: $X = X_1 + X_2 + ... + X_n$, where $X_k \sim Bernoulli(p)$ are IID
- If $X \sim Geom(p)$, then $\mathbb{E}(X) = \dfrac{1}{p}$, $Var(X) = \dfrac{1-p}{p^2}$
- If $X \sim Poisson(\lambda)$, then $\mathbb{E}(X) =$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$
- If $X \sim Binom(n, p)$, then $\mathbb{E}(X) = np$, $Var(X) = np(1-p)$
  Idea: $X = X_1 + X_2 + ... + X_n$, where $X_k \sim Bernoulli(p)$ are IID
- If $X \sim Geom(p)$, then $\mathbb{E}(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$
- If $X \sim Poisson(\lambda)$, then $\mathbb{E}(X) = \lambda$, $Var(X) =$

Expectations and Variances for Important RVs

Discrete Distributions

- If $X \sim Bernoulli(p)$, then $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$
- If $X \sim Binom(n, p)$, then $\mathbb{E}(X) = np$, $Var(X) = np(1-p)$
  Idea: $X = X_1 + X_2 + ... + X_n$, where $X_k \sim Bernoulli(p)$ are IID
- If $X \sim Geom(p)$, then $\mathbb{E}(X) = \dfrac{1}{p}$, $Var(X) = \dfrac{1-p}{p^2}$
- If $X \sim Poisson(\lambda)$, then $\mathbb{E}(X) = \lambda$, $Var(X) = \lambda$

Continuous Distributions

- If $X \sim Unif[a, b]$, then $\mathbb{E}(X) =$

Expectations and Variances for Important RVs

Continuous Distributions

- If $X \sim Unif[a, b]$, then $\mathbb{E}(X) = \dfrac{a+b}{2}$, $Var(X) =$

Expectations and Variances for Important RVs

Continuous Distributions

- If $X \sim Unif[a, b]$, then $\mathbb{E}(X) = \dfrac{a+b}{2}$, $Var(X) = \dfrac{(b-a)^2}{12}$;

Expectations and Variances for Important RVs

Continuous Distributions

- If $X \sim Unif[a, b]$, then $\mathbb{E}(X) = \dfrac{a+b}{2}$, $Var(X) = \dfrac{(b-a)^2}{12}$;
- If $X \sim Exp(\lambda)$, then $\mathbb{E}(X) =$

Expectations and Variances for Important RVs

Continuous Distributions

- If $X \sim Unif[a, b]$, then $\mathbb{E}(X) = \dfrac{a+b}{2}$, $Var(X) = \dfrac{(b-a)^2}{12}$;
- If $X \sim Exp(\lambda)$, then $\mathbb{E}(X) = \dfrac{1}{\lambda}$, $Var(X) = \dfrac{1}{\lambda^2}$;

Expectations and Variances for Important RVs

Continuous Distributions

- If $X \sim Unif[a, b]$, then $\mathbb{E}(X) = \dfrac{a+b}{2}$, $Var(X) = \dfrac{(b-a)^2}{12}$;
- If $X \sim Exp(\lambda)$, then $\mathbb{E}(X) = \dfrac{1}{\lambda}$, $Var(X) = \dfrac{1}{\lambda^2}$;
- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathbb{E}(X) =$

Expectations and Variances for Important RVs

Continuous Distributions

- If $X \sim Unif[a, b]$, then $\mathbb{E}(X) = \dfrac{a+b}{2}$, $Var(X) = \dfrac{(b-a)^2}{12}$;
- If $X \sim Exp(\lambda)$, then $\mathbb{E}(X) = \dfrac{1}{\lambda}$, $Var(X) = \dfrac{1}{\lambda^2}$;
- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathbb{E}(X) = \mu$, $Var(X) =$

Expectations and Variances for Important RVs

Continuous Distributions

- If $X \sim Unif[a, b]$, then $\mathbb{E}(X) = \dfrac{a+b}{2}$, $Var(X) = \dfrac{(b-a)^2}{12}$;
- If $X \sim Exp(\lambda)$, then $\mathbb{E}(X) = \dfrac{1}{\lambda}$, $Var(X) = \dfrac{1}{\lambda^2}$;
- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$;

Expectations and Variances for Important RVs

Continuous Distributions

- If $X \sim Unif[a, b]$, then $\mathbb{E}(X) = \dfrac{a+b}{2}$, $Var(X) = \dfrac{(b-a)^2}{12}$;
- If $X \sim Exp(\lambda)$, then $\mathbb{E}(X) = \dfrac{1}{\lambda}$, $Var(X) = \dfrac{1}{\lambda^2}$;
- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$;
- If $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$, then $\mathbb{E}(\mathbf{X}) = \mu$, $Cov(\mathbf{X}) = \Sigma$
  Note: $Cov(\mathbf{X})$ is the Covariance Matrix of $\mathbf{X}$.

Intro to Markov Chains

Intro To Markov Chains

Up to this point we have considered sequences of r.v. X_k that are IID.

Intro To Markov Chains

Up to this point we have considered sequences of r.v. X_k that are IID. This kind of sequences appear naturally in many phenomena, and the IID structure simplifies calculations, and gives us the celebrated LLN and CLT.

Intro To Markov Chains

Up to this point we have considered sequences of r.v. X_k that are IID. This kind of sequences appear naturally in many phenomena, and the IID structure simplifies calculations, and gives us the celebrated LLN and CLT.

But in many other real-life phenomena, we cannot model things as Independent r.v.s. Say, in Stock Markets, if S_n is the daily closing price of some stock, then $S_0, S_1, S_2, \dots$ are not Independent.

Intro To Markov Chains

Up to this point we have considered sequences of r.v. X_k that are IID. This kind of sequences appear naturally in many phenomena, and the IID structure simplifies calculations, and gives us the celebrated LLN and CLT.

But in many other real-life phenomena, we cannot model things as Independent r.v.s. Say, in Stock Markets, if S_n is the daily closing price of some stock, then $S_0, S_1, S_2, ...$ are not Independent. Or, if $W_1, W_2, W_3, ...$ are daily weather condition r.v.s (say, $W_i = 1$, if the i-th day was Sunny, $W_i = 0$, if Rainy), then W_k-s are not Independent.

Intro To Markov Chains

Up to this point we have considered sequences of r.v. X_k that are IID. This kind of sequences appear naturally in many phenomena, and the IID structure simplifies calculations, and gives us the celebrated LLN and CLT.

But in many other real-life phenomena, we cannot model things as Independent r.v.s. Say, in Stock Markets, if S_n is the daily closing price of some stock, then $S_0, S_1, S_2, \dots$ are not Independent. Or, if $W_1, W_2, W_3, \dots$ are daily weather condition r.v.s (say, $W_i = 1$, if the i-th day was Sunny, $W_i = 0$, if Rainy), then W_k-s are not Independent. So we want to be able to model Dependent Data sequences, and one of the simplest and very important Models is the Markov Chain Model.

Intro To Markov Chains

Up to this point we have considered sequences of r.v. X_k that are IID. This kind of sequences appear naturally in many phenomena, and the IID structure simplifies calculations, and gives us the celebrated LLN and CLT.

But in many other real-life phenomena, we cannot model things as Independent r.v.s. Say, in Stock Markets, if S_n is the daily closing price of some stock, then $S_0, S_1, S_2, \ldots$ are not Independent. Or, if $W_1, W_2, W_3, \ldots$ are daily weather condition r.v.s (say, $W_i = 1$, if the i-th day was Sunny, $W_i = 0$, if Rainy), then W_k-s are not Independent. So we want to be able to model Dependent Data sequences, and one of the simplest and very important Models is the Markov Chain Model.

Recall also that we have considered a Markov Chain - when talking about the Language Modeling!

Intro To Markov Chains

Intro To Markov Chains

Here we will consider Stochastic Processes $X_t : t \in T$, where X_t is a r.v. for each t.

Intro To Markov Chains

Here we will consider Stochastic Processes $X_t : t \in T$, where X_t is a r.v. for each t. Usually, we refer to t as time, and distinguish between

Intro To Markov Chains

Here we will consider Stochastic Processes $X_t : t \in T$, where X_t is a r.v. for each t. Usually, we refer to t as time, and distinguish between

- **Discrete Time Stochastic Process**: here $T = \{0, 1, 2, ...\}$, so t can take values 0, 1, 2, ..., and we will have just a sequence of r.v.s

$$X_0, X_1, X_2, ...$$

Intro To Markov Chains

Here we will consider Stochastic Processes $X_t : t \in T$, where X_t is a r.v. for each t. Usually, we refer to t as time, and distinguish between

- **Discrete Time Stochastic Process**: here $T = \{0, 1, 2, ...\}$, so t can take values 0, 1, 2, ..., and we will have just a sequence of r.v.s

$$X_0, X_1, X_2, ...$$

- **Continuous Time Stochastic Process**: here $T = [0, +\infty)$ or $T = [0, A]$, so t can take any real values from the specified interval. Say, we can have X_1, $X_{1.3}$, $X_{\sqrt{2}}$ etc.

Intro To Markov Chains

Here we will consider Stochastic Processes $X_t : t \in T$, where X_t is a r.v. for each t. Usually, we refer to t as time, and distinguish between

- **Discrete Time Stochastic Process**: here $T = \{0, 1, 2, ...\}$, so t can take values 0, 1, 2, ..., and we will have just a sequence of r.v.s

$$X_0, X_1, X_2, ...$$

- **Continuous Time Stochastic Process**: here $T = [0, +\infty)$ or $T = [0, A]$, so t can take any real values from the specified interval. Say, we can have X_1, $X_{1.3}$, $X_{\sqrt{2}}$ etc.

We will consider only Discrete Time Markov Chains, which is an example of Discrete Time Stochastic Processes.

Definition of the Markov Chain: State Space

First, we fix a set $S =$ the State Space.

Definition of the Markov Chain: State Space

First, we fix a set S = the State Space. S is the set of possible values of X_k.

Definition of the Markov Chain: State Space

First, we fix a set S = the State Space. S is the set of possible values of X_k. We will assume here again that S is Finite: in this case we talk about **Finite State Discrete Time Markov Chains**.

Definition of the Markov Chain: State Space

First, we fix a set $S =$ the State Space. S is the set of possible values of X_k. We will assume here again that S is Finite: in this case we talk about **Finite State Discrete Time Markov Chains**. For simplicity, we can take $S = \{1, 2, 3, ..., N\}$ (we just enumerate the States, we can start also from 0).

Definition of the Markov Chain: State Space

First, we fix a set $S =$ the State Space. S is the set of possible values of X_k. We will assume here again that S is Finite: in this case we talk about **Finite State Discrete Time Markov Chains**. For simplicity, we can take $S = \{1, 2, 3, ..., N\}$ (we just enumerate the States, we can start also from 0).

Assume now we are modeling a (Probabilistic, Stochastic) System, which can be in any of states from S.

Definition of the Markov Chain: State Space

First, we fix a set $S =$ the State Space. S is the set of possible values of X_k. We will assume here again that S is Finite: in this case we talk about **Finite State Discrete Time Markov Chains**. For simplicity, we can take $S = \{1, 2, 3, ..., N\}$ (we just enumerate the States, we can start also from 0).
Assume now we are modeling a (Probabilistic, Stochastic) System, which can be in any of states from S. We denote the state of our System at time t by X_t, $t = 0, 1, 2, ...$ So X_t can take values from S, and X_t is a r.v., since we do not know in which State will be our X_t.

Definition of the Markov Chain: State Space

First, we fix a set $S =$ the State Space. S is the set of possible values of X_k. We will assume here again that S is Finite: in this case we talk about **Finite State Discrete Time Markov Chains**. For simplicity, we can take $S = \{1, 2, 3, ..., N\}$ (we just enumerate the States, we can start also from 0).

Assume now we are modeling a (Probabilistic, Stochastic) System, which can be in any of states from S. We denote the state of our System at time t by X_t, $t = 0, 1, 2, ...$ So X_t can take values from S, and X_t is a r.v., since we do not know in which State will be our X_t.

Example: Say, we are modeling the Weather. The States are $Rainy, Sunny$, or, say, $0, 1$. W_t is the weather at the t-th day, started from today ($t = 0$). So W_0 is today's weather, W_1 is the tomorrow's weather etc.

Definition of the Markov Chain

Now, we give the Definition:

Definition of the Markov Chain

Now, we give the Definition:

Markov Chain

Assume $X_0, X_1, X_2, ...$ is a sequence of r.v.s (Discrete Stochastic Process), which take values from $S = \{1, 2, ..., N\}$. We say that X_n, $n = 0, 1, 2, ...$ is a (Finite State Discrete Time) **Markov Chain**, if

$$\mathbb{P}(X_{t+1} = j | X_t = i, X_{t-1} = k, ..., X_0 = m) = \mathbb{P}(X_{t+1} = j | X_t = i)$$

for any time t, for any state $j, i, k, ..., m$.

Definition of the Markov Chain

Now, we give the Definition:

Markov Chain

Assume $X_0, X_1, X_2, \ldots$ is a sequence of r.v.s (Discrete Stochastic Process), which take values from $S = \{1, 2, \ldots, N\}$. We say that X_n, $n = 0, 1, 2, \ldots$ is a (Finite State Discrete Time) **Markov Chain**, if

$$\mathbb{P}(X_{t+1} = j | X_t = i, X_{t-1} = k, \ldots, X_0 = m) = \mathbb{P}(X_{t+1} = j | X_t = i)$$

for any time t, for any state $j, i, k, \ldots, m$.

Interpretation: Given today's State, tomorrow's State is independent of the past States.

Definition of the Markov Chain

Now, we give the Definition:

Markov Chain

Assume $X_0, X_1, X_2, \ldots$ is a sequence of r.v.s (Discrete Stochastic Process), which take values from $S = \{1, 2, \ldots, N\}$. We say that X_n, $n = 0, 1, 2, \ldots$ is a (Finite State Discrete Time) **Markov Chain**, if

$$\mathbb{P}(X_{t+1} = j | X_t = i, X_{t-1} = k, \ldots, X_0 = m) = \mathbb{P}(X_{t+1} = j | X_t = i)$$

for any time t, for any state $j, i, k, \ldots, m$.

Interpretation: Given today's State, tomorrow's State is independent of the past States. Or, in other words, today's information is enough to completely determine the probabilities of Tomorrow's states.