COMP2610 / COMP6261 - Information Theory

Lecture 9: Probabilistic Inequalities

Assignment Project Exam Help

https://eduassistpro.github.i

Add WeChat edu_assist_pr

Australian National University

20 August, 2018

Last time

Assignment Project Exam Help

Mutual information chain rule

Jensen'

https://eduassistpro.github.i

"Informa

Data processing inequality

Add WeChat edu_assist_pr

Review: Data-Processing Inequality

Theorem

if $X \to Y \to Z$ then $I(X;Y) \geq I(X;Z)$

Assignment Project Exam Help

- X is t … s the pro …

https://eduassistpro.github.i

- No “clever” manipulation of the data can impro … can be made from the data

Add WeChat edu_assist_pr

- No processing of Y, deterministic or random, can increase the information that Y contains about X

This time

Assignment Project Exam Help

- Markov's inequality
- Ch https://eduassistpro.github.i
Add WeChat edu_assist_pr
- Law of large numbers

Outline

1. Properties of expectation and variance
2. Mark
3. Cheb
4. Law of large numbers
5. Wrapping Up

Assignment Project Exam Help
https://eduassistpro.github.i
Add WeChat edu_assist_pr

1 Properties of expectation and variance

Assignment Project Exam Help

2 Markov's Inequality

3 Chebyshev's Inequality

https://eduassistpro.github.io

4 Law of large numbers

Add WeChat edu_assist_pro

5 Wrapping Up

Expectation and Variance

Let X be a random variable over $\mathcal{X}$, with probability distribution p

Assignment Project Exam Help

Expected value:

Varianc

https://eduassistpro.github.i

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}$$

$$= \mathbb{E}[X^2] - ($$

Add WeChat edu_assist_pr

Standard deviation is $\sqrt{\mathbb{V}[X]}$

Properties of expectation

A key property of expectations is linearity:

Assignment Project Exam Help

$$\mathbb{E}\left[\sum^{n} X_i\right] = \sum^{n} \mathbb{E}\left[X_i\right]$$

https://eduassistpro.github.i

$x_1 \in \mathcal{X}_1$ $x_n \in \mathcal{X}_n$ $i=1$

This holds even if the variables are dependent!

Add WeChat edu_assist_pr

We have for any $a \in \mathbb{R}$,

$$\mathbb{E}[aX] = a \cdot \mathbb{E}[X].$$

Properties of variance

We have linearity of variance for independent random variables:

Assignment Project Exam Help

$$\mathbb{V}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{V}[X_i].$$

Does not h

https://eduassistpro.github.i

(prove this: expand the definition of variance and rely upo $X_i)\mathbb{E}(X_j)$ when $X_i \perp X_j$)

Add WeChat edu_assist_pr

We have for any $a \in \mathbb{R}$,

$$\mathbb{V}[aX] = a^2 \cdot \mathbb{V}[X].$$

1. Properties of expectation and variance

Assignment Project Exam Help

2. Markov's inequality

3. Cheb

https://eduassistpro.github.i

4. Law of large numbers

Add WeChat edu_assist_pr

5. Wrapping Up

Markov's Inequality

Motivation

Assignment Project Exam Help

1000 school students sit an examination

The busy p

https://eduassistpro.github.i

The principal wants to estimate the maximum poss
who scored more than 80

Add WeChat edu_assist_pr

- A question about the *minimum* number of s

Markov's Inequality

Motivation

Call x the number of students who score > 80

Assignment Project Exam Help

Call S is the total score of students who score ≤ 80

We know:

$$40 \cdot 1 \text{[illegible]} \} > 80x$$

https://eduassistpro.github.i

Exam scores are nonnegative, so certainly

Add WeChat edu_assist_pr

Thus, $80x < 40 \cdot 1000$, or

$$x < 500.$$

Can we formalise this more generally?

Markov's Inequality

Assignment Project Exam Help

Theorem

Let X be a nonnegative random variable. Then, for any $\lambda > 0$,

https://eduassistpro.github.i

Bounds probability of observing a large outcome

Add WeChat edu_assist_pr

Vacuous if $\lambda < \mathbb{E}[X]$

Assignment Project Exam Help

Corollary

Let X be

https://eduassistpro.github.i

Observations of nonnegative random variable u
than expected value

Add WeChat edu_assist_pr

Vacuous if $\lambda < 1$

Markov's Inequality

Proof

Assignment Project Exam Help

$$\mathbb{E}[X] = \quad x \cdot p(x)$$

https://eduassistpro.github.i

$$\geq \sum_{x \geq \lambda} x \cdot p(x)$$ nonn

Add WeChat edu_assist_pr

$$\geq \sum_{x \geq \lambda} \lambda \cdot p(x)$$

$$= \lambda \cdot p(X \geq \lambda).$$

Markov's Inequality

Illustration from

https://justindomke.wordpress.com/2008/06/19/markovs-inequality/

Assignment Project Exam Help

https://eduassistpro.github.i

Add WeChat edu_assist_pr

Markov's Inequality

Illustration from

https://justindomke.wordpress.com/2008/06/19/markovs-inequality/

Assignment Project Exam Help

https://eduassistpro.github.i

Add WeChat edu_assist_pr

Markov's Inequality

Illustration from

https://justindomke.wordpress.com/2008/06/19/markovs-inequality/

Assignment Project Exam Help

https://eduassistpro.github.i

Add WeChat edu_assist_pr

Markov's Inequality

Illustration from

https://justindomke.wordpress.com/2008/06/19/markovs-inequality/

Assignment Project Exam Help

https://eduassistpro.github.i

Add WeChat edu_assist_pr

Assignment Project Exam Help
https://eduassistpro.github.i
Add WeChat edu_assist_pr

Chebyshev's Inequality

Motivation

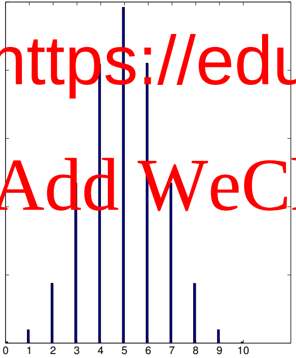
Markov's inequality only uses the mean of the distribution

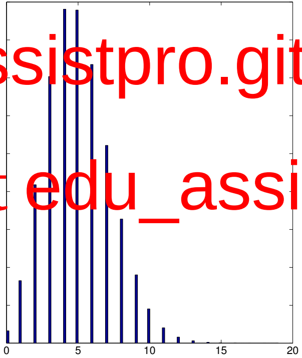
Assignment Project Exam Help

What about the spread of the distribution (variance)?

https://eduassistpro.github.i

Add WeChat edu_assist_pr





Chebyshev's Inequality

Assignment Project Exam Help

Theorem

Let X be a random variable with $\mathbb{E}[X] < \infty$. Then, for any $\lambda > 0$,

https://eduassistpro.github.i

Bounds the probability of observing an "unexpect

Add WeChat edu_assist_pr

Does not require non negativity

Two-sided bound

Chebyshev's Inequality

Alternate Statement

Assignment Project Exam Help

Corollar

Let X be 0.

https://eduassistpro.github.i

$$(| \quad - \quad [\quad]| \geq \lambda \cdot \quad \overline{[\quad]}) \quad -.$$

Observatio s are unlikely to occur several standa the mean

Add WeChat edu_assist_pr

Chebyshev's Inequality

Proof

Define

$$Y = (X - \mathbb{E}[X])^2.$$

Assignment Project Exam Help

Then, by Markov's inequality, for any $\nu > 0$,

https://eduassistpro.github.i

But,

$$\mathbb{E}[Y] = \mathbb{V}[X$$

Add WeChat edu_assist_pr

Also,

$$Y \geq \nu \iff |X - \mathbb{E}$$

Thus, setting $\lambda = \sqrt{\nu}$,

$$p(|X - \mathbb{E}[X]| \geq \lambda) \leq \frac{\mathbb{V}[X]}{\lambda^2}.$$

Chebyshev's Inequality

Illustration

For a binomial X with N trials and success probability θ, we have e.g.

$$p\left(|X - N\theta| \geq \sqrt{2N\theta(1-\theta)}\right) \leq \frac{1}{2}.$$

Assignment Project Exam Help

https://eduassistpro.github.i

Add WeChat edu_assist_pr

Chebyshev's Inequality

Example

Suppose we have a coin with bias θ, i.e. $p(X = 1) = \theta$

Assignment Project Exam Help

Say we flip the coin n times, and observe $x_1, \ldots, x_n \in \{0, 1\}$

We use th

https://eduassistpro.github.i

Estimate how large n should be such that

Add WeChat edu_assist_pr

$$p(|\hat{\theta}_n - \theta| \geq 0.05)$$

1% probability of a 5% error

(Aside: the need for two parameters here is generic: "Probabably Approximately Correct")

Observe that

$$\mathbb{E}[\hat{\theta}_n] = \frac{\sum_{i=1}^{n} \mathbb{E}[x_i]}{n} = \theta$$

$$\mathbb{V}[\hat{\theta}_n] = \frac{\sum_{i=1}^{n} \mathbb{V}[x_i]}{n^2} = \frac{\theta(1-\theta)}{n}$$

Thus, app[illegible]

$$p(|\hat{\theta}_n - \theta| \geq 0.05) \leq [illegible]$$

We are guaranteed this is less than 0.01 if

$$n \geq \frac{\theta(1-\theta)}{(0.05)^2(0.01)}.$$

When $\theta = 0.5$, $n \geq 10,000$ (!)

Assignment Project Exam Help
https://eduassistpro.github.i
Add WeChat edu_assist_pr

Assignment Project Exam Help

https://eduassistpro.github.i

Add WeChat edu_assist_pr

Independent and Identically Distributed

Let $X_1, \ldots, X_n$ be random variables such that:

Assignment Project Exam Help

- Each X_i is independent of X_j
- The

https://eduassistpro.github.i

Then, we say that $X_1, \ldots, X_n$ are independ
(or iid)

Add WeChat edu_assist_pr

Example: For n independent flips of an unbiase $_1$ $_n$ re iid from Bernoulli$\left(\frac{1}{2}\right)$

Law of Large Numbers

Theorem

Let $X_1, \ldots, X_n$ be a sequence of iid random variables, with

Assignment Project Exam Help

$$\mathbb{E}[X_i] = \mu$$

and $\mathbb{V}[X_i$ [illegible]

https://eduassistpro.github.i

$${}_n = \frac{[illegible]}{n}.$$

Then, for any $\epsilon > 0$,

Add WeChat edu_assist_pr

$$\lim_{n \to \infty} p(|\bar{X}_n - \mu| > \epsilon) = 0.$$

Given enough trials, the empirical "success frequency" will be close to the expected value

Since the X_i's are identically distributed,

Assignment Project Exam Help

$$\mathbb{E}[\bar{X}_n] = \mu.$$

Since the

https://eduassistpro.github.i

$$= \frac{\mathbb{V}[X_1 + \dots}{n^2}$$

Add WeChat edu_assist_pr

$$= \frac{n\sigma^2}{n^2}$$

$$= \frac{\sigma^2}{n}.$$

Law of Large Numbers

Proof

Assignment Project Exam Help

Applying Chebyshev's inequality to $\bar{X}_n$,

$$\frac{\mathbb{V}[\bar{X}_n]}{}$$

https://eduassistpro.github.i

For any fixed $\epsilon > 0$, as $n \to \infty$, the right hand s

Add WeChat edu_assist_pr

Thus,

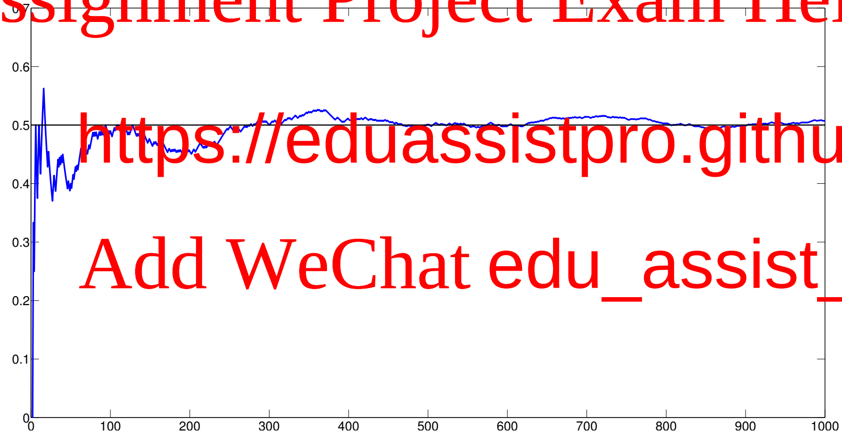
$$p(|\bar{X}_n - \mu| < \epsilon) \to 1.$$

Law of Large Numbers

Illustration

$N = 1000$ trials with Bernoulli random variable with parameter $\frac{1}{2}$

Assignment Project Exam Help



https://eduassistpro.github.i

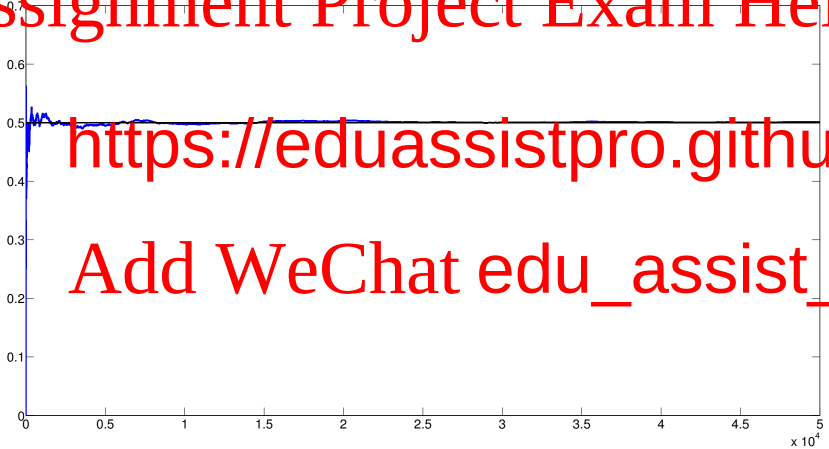
Add WeChat edu_assist_pr

Law of Large Numbers

Illustration

$N = 50000$ trials with Bernoulli random variable with parameter $\frac{1}{2}$

Assignment Project Exam Help
https://eduassistpro.github.i
Add WeChat edu_assist_pr



Assignment Project Exam Help

https://eduassistpro.github.i

Add WeChat edu_assist_pr

Summary & Conclusions

Assignment Project Exam Help

- Markov's inequality

https://eduassistpro.github.i

- Ch

Add WeChat edu_assist_pr

- Law of large numbers

Next time

Assignment Project Exam Help

- Ensembles and sequences

https://eduassistpro.github.i

- Typ

Add WeChat edu_assist_pr

- Approximation Equipartition (AEP)