THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics

Lecture 23: Continuous random variables- Inequalities, CLT

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

www.cs.cmu.edu/∼psarkar/teaching

Roadmap

- Examples of Markov and Chebyshev
- Weak law of large numbers and CLT
- Normal approximation to Binomial

Markov's inequality Example

You have 20 independent *Poisson*(1) random variables $X_1, \ldots, X_{20}$. Use the Markov inequality to bound $P(\sum_{i=1}^{20} X_i \geq 15)$

Markov's inequality Example

You have 20 independent *Poisson*(1) random variables $X_1, \ldots, X_{20}$. Use the Markov inequality to bound $P(\sum_{i=1}^{20} X_i \geq 15)$

- $P(\sum_i X_i \geq 15) \leq \frac{E[\sum_i X_i]}{15} = \frac{20}{15} = \frac{4}{3}$

Markov's inequality Example

You have 20 independent $Poisson(1)$ random variables $X_1, \ldots, X_{20}$. Use the Markov inequality to bound $P(\sum_{i=1}^{20} X_i \geq 15)$

- $P(\sum_i X_i \geq 15) \leq \frac{E[\sum_i X_i]}{15} = \frac{20}{15} = \frac{4}{3}$
- How useful is this?

Chebyshev's inequality Example

You have n independent $Poisson(1)$ random variables $X_1, \ldots, X_n$. Use the Chebyshev inequality to bound $P(|\bar{X} - 1| \geq 1)$?

Chebyshev's inequality Example

You have n independent $Poisson(1)$ random variables $X_1, \ldots, X_n$. Use the Chebyshev inequality to bound $P(|\bar{X} - 1| \geq 1)$?

- $$\begin{aligned} P(|\bar{X} - 1| \geq 1) &\leq \frac{\text{var}(X_1)}{n} = \frac{1}{n} \\ &= \frac{1}{10} \qquad \text{When } n = 10 \\ &= \frac{1}{100} \qquad \text{When } n = 100 \\ &\ldots \end{aligned}$$

Weak law of large numbers

The WLLN basically states that the sample mean of a large number of random variables is very close to the true mean with high probability.

- Consider a sequence of i.i.d random variables $X_1, \ldots X_n$ with mean μ and variance σ^2.
- Let $M_n = \dfrac{X_1 + \cdots + X_n}{n}$.

Weak law of large numbers

The WLLN basically states that the sample mean of a large number of random variables is very close to the true mean with high probability.

- Consider a sequence of i.i.d random variables $X_1, \ldots X_n$ with mean μ and variance σ^2.
- Let $M_n = \dfrac{X_1 + \cdots + X_n}{n}$.
- $E[M_n] = \dfrac{E[X_1] + \cdots + E[X_n]}{n} = \mu$

Weak law of large numbers

The WLLN basically states that the sample mean of a large number of random variables is very close to the true mean with high probability.

- Consider a sequence of i.i.d random variables $X_1, \dots X_n$ with mean μ and variance σ^2.
- Let $M_n = \dfrac{X_1 + \cdots + X_n}{n}$.
- $E[M_n] = \dfrac{E[X_1] + \cdots + E[X_n]}{n} = \mu$
- $\text{var}(M_n) = \dfrac{\text{var}[X_1] + \cdots + \text{var}[X_n]}{n^2} = \dfrac{\sigma^2}{n}$

Weak law of large numbers

The WLLN basically states that the sample mean of a large number of random variables is very close to the true mean with high probability.

- Consider a sequence of i.i.d random variables $X_1, \dots X_n$ with mean μ and variance σ^2.
- Let $M_n = \dfrac{X_1 + \dots + X_n}{n}$.
- $E[M_n] = \dfrac{E[X_1] + \dots + E[X_n]}{n} = \mu$
- $\text{var}(M_n) = \dfrac{\text{var}[X_1] + \dots + \text{var}[X_n]}{n^2} = \dfrac{\sigma^2}{n}$
- So $P(|M_n - \mu| \geq \epsilon) \leq \dfrac{\sigma^2}{n\epsilon^2}$

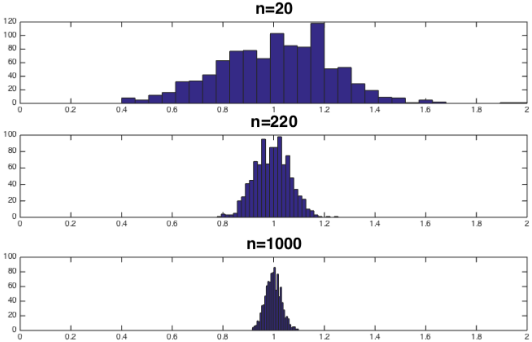
Weak law of large numbers

The WLLN basically states that the sample mean of a large number of random variables is very close to the true mean with high probability.

- Consider a sequence of i.i.d random variables $X_1, \ldots X_n$ with mean μ and variance σ^2.
- Let $M_n = \dfrac{X_1 + \cdots + X_n}{n}$.
- $E[M_n] = \dfrac{E[X_1] + \cdots + E[X_n]}{n} = \mu$
- $\text{var}(M_n) = \dfrac{\text{var}[X_1] + \cdots + \text{var}[X_n]}{n^2} = \dfrac{\sigma^2}{n}$
- So $P(|M_n - \mu| \geq \epsilon) \leq \dfrac{\sigma^2}{n\epsilon^2}$
- For large n this probability is small.

Illustration

Consider the mean of n independent Poisson(1) random variables. For each n, we plot the distribution of the average.



Can we say more? Central Limit Theorem

Turns out that not only can you say that the sample mean is close to the true mean, you can actually predict its distribution using the famous Central Limit Theorem.

- Consider a sequence of i.i.d random variables $X_1, \ldots X_n$ with mean μ and variance σ^2.
- Let $\bar{X}_n = \dfrac{X_1 + \cdots + X_n}{n}$. Remember $E[\bar{X}_n] = \mu$ and $\text{var}(\bar{X}_n) = \sigma^2/n$
- Standardize $\bar{X}_n$ to get $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$
- As n gets bigger, $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ behaves more and more like a $Normal(0, 1)$ random variable.
- $P\left(\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < z\right) \approx \Phi(z)$

Can we say more? Central Limit Theorem

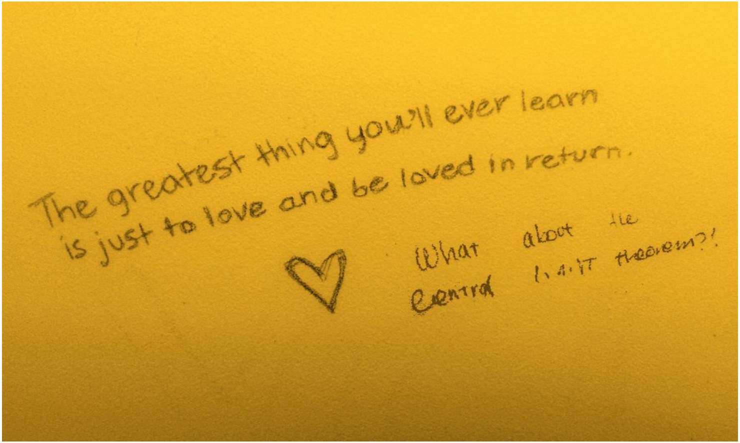


Figure: (Courtesy: Tamara Broderick) You bet!

Example

You have 20 independent $Poisson(1)$ random variables $X_1, \ldots, X_{20}$. Use the CLT to bound $P(\sum_{i=1}^{20} X_i \geq 15)$

Example

You have 20 independent $Poisson(1)$ random variables $X_1, \ldots, X_{20}$. Use the CLT to bound $P(\sum_{i=1}^{20} X_i \geq 15)$

$$P(\sum_i X_i \geq 15) = P(\sum_i X_i - 20 \geq -5)$$

- $$= P(\frac{\bar{X} - 1}{1/\sqrt{20}} \geq -.25\sqrt{20})$$
$$\approx P(Z \geq -1.18) = 0.86$$

Example

You have 20 independent $Poisson(1)$ random variables $X_1, \ldots, X_{20}$. Use the CLT to bound $P(\sum_{i=1}^{20} X_i \geq 15)$

$$P(\sum_i X_i \geq 15) = P(\sum_i X_i - 20 \geq -5)$$

- $$= P(\frac{\bar{X} - 1}{1/\sqrt{20}} \geq -.25\sqrt{20})$$
  $$\approx P(Z \geq -1.18) = 0.86$$
- How useful is this? Better than Markov.

Example

An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that, because of changing atmospheric conditions and normal error, each time a measurement is made it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light-years), how many measurements need he make to 95% sure that his estimated distance is accurate to within $\pm .5$ lightyears?

Example

An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that, because of changing atmospheric conditions and normal error, each time a measurement is made it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light-years), how many measurements need he make to 95% sure that his estimated distance is accurate to within $\pm .5$ lightyears?

- Let $\bar{X}_n$ be the mean of the measurements.
- How large does n have to be so that $P(|\bar{X}_n - d| \leq .5) = 0.95$

$$P\left(\frac{|\bar{X}_n - d|}{2/\sqrt{n}} \leq 0.25\sqrt{n}\right) \approx P(|Z| \leq 0.25\sqrt{n}) = 1 - 2P(Z \leq -0.25\sqrt{n}) = 0.95$$

- $$\begin{aligned} P(Z \leq -0.25\sqrt{n}) &= 0.025 \\ -0.25\sqrt{n} &= -1.96 \\ \sqrt{n} &= 7.84 \\ n &\approx 62 \end{aligned}$$

Normal Approximation to Binomial

The probability of selling an umbrella is 0.5 on a rainy day. If there are 400 umbrellas in the store, whats the probability that the owner will sell at least 180?

- Let X be the total number of umbrellas sold.
- $X \sim Binomial(400, .5)$
- We want $P(X > 180)$. Crazy calculations.

Normal Approximation to Binomial

The probability of selling an umbrella is 0.5 on a rainy day. If there are 400 umbrellas in the store, whats the probability that the owner will sell at least 180?

- Let X be the total number of umbrellas sold.
- $X \sim Binomial(400, .5)$
- We want $P(X > 180)$. Crazy calculations.

- But can we approximate the distribution of X/n?
- $X/n = (\sum_i Y_i)/n$ where $E[Y_i] = 0.5$ and $\text{var}(Y_i) = 0.25$.
- Sure! CLT tells us that for large n, $\frac{X/400 - 0.5}{\sqrt{0.25/400}} \sim N(0, 1)$
- So $P(X > 180) = P((X - 200)/\sqrt{100} > -2) \approx P(Z \geq -2) = 1 - \Phi(-2) = 0.97$

Frequentist Statistics

- The parameter(s) θ is fixed and unknown
- Data is generated through the likelihood function $p(X;\theta)$ (if discrete) or $f(X;\theta)$ (if continuous).
- Now we will be dealing with multiple candidate models, one for each value of θ
- We will use $E_\theta[h(X)]$ to define the expectation of the random variable $h(X)$ as a function of parameter θ

Problems we will look at

- **Parameter estimation**: We want to estimate unknown parameters from data.
  - **Maximum Likelihood estimation (section 9.1):** Select the parameter that makes the observed data most likely.
  - i.e. maximize the probability of obtaining the data at hand.
- **Hypothesis testing:** An unknown parameter takes a finite number of values. One wants to find the best hypothesis based on the data.
  - **Significance testing:** Given a hypothesis, figure out the rejection region and reject the hypothesis if the observation falls within this region.