Lecture 6D: Continuous Probability II

UC Berkeley CS70
Summer 2023
Nikki Suzani

Gaussian (Normal) Random Variables

A normal (or Gaussian) distribution is a distribution with two parameters, μ – its expectation – and σ^2 – its variance.

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

Some Cool Properties of the Normal Distribution

Gaussian (Normal) Random Variables (cont.)

CDF of the Normal(0, 1) distribution is referred to as Φ(x).

$$\Phi(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{u^2}{2}\right\} du.$$

“Scaling” Normal Random Variables

Lemma 21.1. *If* $X \sim N(\mu, \sigma^2)$, *then* $Y = \frac{X-\mu}{\sigma} \sim N(0,1)$. *Equivalently, if* $Y \sim N(0,1)$, *then* $X = \sigma Y + \mu \sim N(\mu, \sigma^2)$.

“Scaling” Normal Random Variables (cont.)

$$\mathbb{P}[X \le a] = \mathbb{P}[Y \le \frac{a-\mu}{\sigma}]$$

where Y is standard normal.

This allows us to use the CDF!

Sum of Independent Normal RVs

Corollary 21.1. *Let* $X \sim N(\mu_X, \sigma_X^2)$ *and* $Y \sim N(\mu_Y, \sigma_Y^2)$ *be independent normal random variables. Then for any constants* $a, b \in \mathbb{R}$*, the random variable* $Z = aX + bY$ *is also normally distributed with mean* $\mu = a\mu_X + b\mu_Y$ *and variance* $\sigma^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$.

Central Limit Theorem

Let X_1, X_2, … X_n be a sequence of n i.i.d. random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. When **n** is large, both the sample sum and the sample mean can be approximated as normal random variables.

Sample Sum Expectation & Variance

Let X_1, X_2, … X_n be a sequence of n i.i.d. random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Find the distribution of $S_n = X_1 + \ldots + X_n$ for large n.

Scaling to Standard Normal

Sample Mean Expectation and Variance

Let X_1, X_2, … X_n be a sequence of n i.i.d. random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Find the distribution of X-bar = $S_n/n = X_1 + \ldots + X_n$ for large n.

Scaling to Standard Normal

How to Solve CLT Questions

1. Identify that you can use CLT – sample mean/sum of i.i.d. random variables for **large n**.

2. Calculate the Mean and Variance of the sample mean/sum

3. Convert your **value** into the standard normal value (subtract mean, divide by standard deviation)

4. Use Φ(standard normal value)

Dice Example

2. My friend and I gamble on rolls of a die. Each time the die is rolled,

- my friend gives me a dollar if the number of spots is five or six,
- I give my friend a dollar if the number of spots is one or two,
- and no money changes hands if the number of spots is three or four.

If we play this game 400 times, approximately what is the chance that my net gain is more than 20 dollars?

Dice Example (cont.)

Exam Example

N students take an exam where the average is 50, and the variance is 5. Let S_n be the sum of their n scores, and assume all of their scores are independent. What is the probability that their average score is greater than 72.5?

Exam Example

Polling Example

We want to estimate the true proportion **p** of students who like Olivia Rodrigo more than Joshua Bassett. We poll **n** people, for large **n**. How many people do we have to poll to get an estimate with accuracy **0.1** with **95%** probability?

Polling Example

Poisson Example

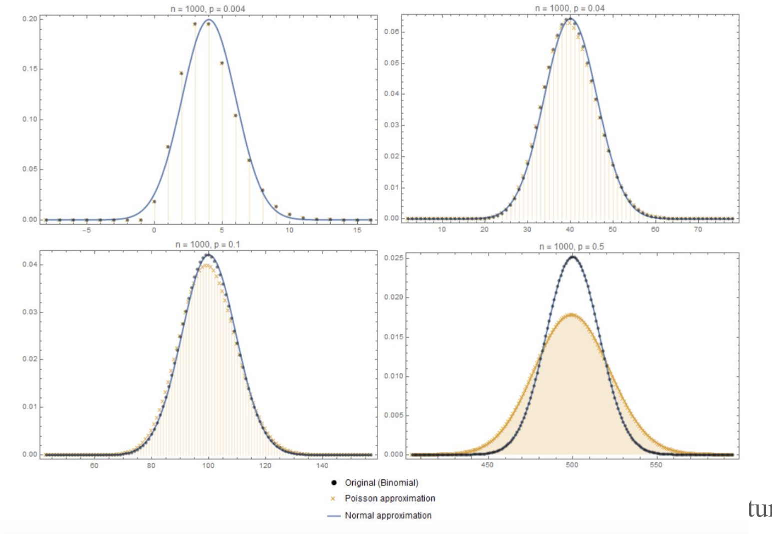
The number of times I listen to Olivia Rodrigo's *Vampire* is modeled by a Poisson(λ) distribution per day, but I don't know λ. All I know about λ is that it is <u>at most</u> 20**.** If I track the number of times I listen to *Vampire* each day, for a large number of days, how many days do I need to track to get a 95% confidence interval for λ of **width 3.**

Poisson Example (cont.)

Poisson Example (cont.)

Normal as a Limiting Distribution of Binomial

When **n** is large, and **p** is not small, the normal distribution is a good approximation to the binomial. Source: https://math.stackexchange.com/questions/3278070/approximation-of-binomial-distribution-poisson-vs-normal-distribution.



(If time)
The following are slides that didn't make it into recording for 6C so I will redo them if there is time, so that they can be in a recording.

Exponential Distribution

Def: For some λ > 0, a continuous random variable X is an exponential random variable with parameter λ if it has the following PDF,

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

We can write X ~ Exp(λ) if is an exponential random variable.

Let's check that f(x) satisfies the two PDF properties

Expectation & Variance of Exponential

For a random variable X ~ Exp(λ),

$$\mathbb{E}[X] = \frac{1}{\lambda} \qquad \textit{and} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Example Exponential

Going back to our terrible alarm clock, we know the behavior is:

Once plugged in, the alarm will randomly ring once after some amount of time, however we know it goes off at a rate of 1 time every 10 minutes.

Let X be the amount of time it takes the alarm to sound. $X \sim Exp(1/10)$, because our rate of rings is 1/10 per minute.

How many minutes should we expect to wait before the alarm rings?

An important property to note

$$\mathbb{P}[X > t] = e^{-\lambda t}$$

Proof:

Exponential Relation to Geometric

Let's try to draw a more rigorous connection between Exponential and Geometric distributions. Take our Geometric r.v. and consider running trials after every d seconds. The probability of a success is p = $\lambda\boldsymbol{\delta}$.

Then, let Y denote the amount of time/seconds before a successful trial:

$$\mathbb{P}[Y > k\delta] = (1-p)^k = (1-\lambda\delta)^k,$$

To translate our trials to a continuum, consider taking the limit of d → 0. Then, for any time t,

$$\mathbb{P}[Y > t] = \mathbb{P}[Y > (\tfrac{t}{\delta})\delta] = (1-\lambda\delta)^{t/\delta} \approx e^{-\lambda t}$$

Recap

Discussed Gaussian (Normal) Random Variables

- Properties of Normal Distributions
- Scaling Normal Distributions
- Summing Independent Random Normal Variables
- The CDF of a Normal Distribution
- Solving Problems using the **Central Limit Theorem (CLT)**