CS70: Jean Walrand: Lecture 27.

CLT + Random Thoughts

1. Review: Continuous Probability
2. Normal Distribution
3. Central Limit Theorem
4. Random Thoughts

Continuous Probability

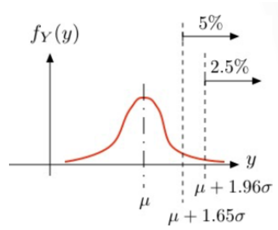
1. pdf: $Pr[X \in (x, x+\delta]] = f_X(x)\delta$.
2. CDF: $Pr[X \leq x] = F_X(x) = \int_{-\infty}^{x} f_X(y)dy$.
3. $U[a,b]$, $Expo(\lambda)$, target.
4. Expectation: $E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$.
5. Expectation of function: $E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x)dx$.
6. Variance: $var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$.
7. Gaussian: $\mathcal{N}(\mu, \sigma^2) : f_X(x) = ...$ "bell curve"

Normal Distribution.

For any μ and σ, a **normal** (aka **Gaussian**) random variable Y, which we write as $Y = \mathcal{N}(\mu, \sigma^2)$, has pdf

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}.$$

Standard normal has $\mu = 0$ and $\sigma = 1$.



Note: $Pr[|Y - \mu| > 1.65\sigma] = 10\%$; $Pr[|Y - \mu| > 2\sigma] = 5\%$.

Scaling and Shifting

Theorem Let $X = \mathcal{N}(0,1)$ and $Y = \mu + \sigma X$. Then

$$Y = \mathcal{N}(\mu, \sigma^2).$$

Proof: $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$. Now,

$$\begin{aligned} f_Y(y) &= \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right) \quad \text{(See Lec. 26, slide 19.)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(y-\mu)^2}{2\sigma^2}\}. \quad \square \end{aligned}$$

Expectation, Variance.

Theorem If $Y = \mathcal{N}(\mu, \sigma^2)$, then

$$E[Y] = \mu \text{ and } var[Y] = \sigma^2.$$

Proof: It suffices to show the result for $X = \mathcal{N}(0,1)$ since $Y = \mu + \sigma X, ...$

Thus, $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$.

First note that $E[X] = 0$, by symmetry.

$$\begin{aligned} var[X] &= E[X^2] = \int x^2 \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\} dx \\ &= -\frac{1}{\sqrt{2\pi}} \int x d\exp\{-\frac{x^2}{2}\} = \frac{1}{\sqrt{2\pi}} \int \exp\{-\frac{x^2}{2}\} dx \text{ by IBP}^1 \\ &= \int f_X(x)dx = 1. \quad \square \end{aligned}$$

[1] Integration by Parts: $\int_a^b f dg = [fg]_a^b - \int_a^b g df$.

Central Limit Theorem.

Law of Large Numbers: For any set of independent identically distributed random variables, X_i, $A_n = \frac{1}{n} \sum X_i$ "tends to the mean."

Say X_i have expecation $\mu = E(X_i)$ and variance σ^2.

Mean of A_n is μ, and variance is σ^2/n.

Let

$$S_n := \frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then,

$$E(S_n) = \frac{1}{\sigma/\sqrt{n}}(E(A_n) - \mu) = 0$$

$$Var(S_n) = \frac{1}{\sigma^2/n} Var(A_n) = 1.$$

Central limit theorem: As n goes to infinity the distribution of S_n approaches the standard normal distribution.

Central Limit Theorem

Central Limit Theorem

Let $X_1, X_2, \ldots$ be i.i.d. with $E[X_1] = \mu$ and $var(X_1) = \sigma^2$. Define

$$S_n := \frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then,

$$S_n \to \mathcal{N}(0,1), \text{as } n \to \infty.$$

That is,

$$Pr[S_n \leq \alpha] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2} dx.$$

Proof: See EE126.

CI for Mean

Let $X_1, X_2, \ldots$ be i.i.d. with mean μ and variance σ^2. Let

$$A_n = \frac{X_1 + \cdots + X_n}{n}.$$

The CLT states that

$$\frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1) \text{ as } n \to \infty.$$

Thus, for $n \gg 1$, one has

$$Pr[-2 \leq |\frac{A_n - \mu}{\sigma/\sqrt{n}}| \leq 2] \approx 95\%.$$

Equivalently,

$$Pr[\mu \in [A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}]] \approx 95\%.$$

That is,

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu.$$

CI for Mean

Let $X_1, X_2, \ldots$ be i.i.d. with mean μ and variance σ^2. Let

$$A_n = \frac{X_1 + \cdots + X_n}{n}.$$

The CLT states that

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1) \text{ as } n \to \infty.$$

Also,

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu.$$

Recall: Using Chebyshev, we found that (see Lec. 22, slide 6)

$$[A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu.$$

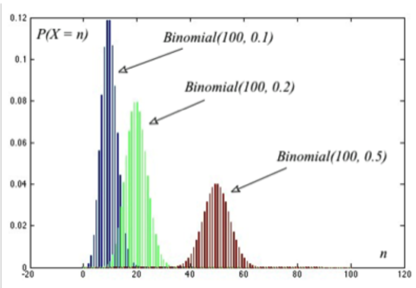
Thus, the CLT provides a smaller confidence interval.

Coins and normal.

Let $X_1, X_2, \ldots$ be i.i.d. $B(p)$. Thus, $X_1 + \cdots + X_n = B(n,p)$.

Here, $\mu = p$ and $\sigma = \sqrt{p(1-p)}$. CLT states that

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{p(1-p)n}} \to \mathcal{N}(0,1).$$



Coins and normal.

Let $X_1, X_2, \ldots$ be i.i.d. $B(p)$. Thus, $X_1 + \cdots + X_n = B(n,p)$.
Here, $\mu = p$ and $\sigma = \sqrt{p(1-p)}$. CLT states that

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{p(1-p)n}} \to \mathcal{N}(0,1)$$

and

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu$$

with $A_n = (X_1 + \cdots + X_n)/n$.
Hence,

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } p.$$

Since $\sigma \leq 0.5$,

$$[A_n - 2\frac{0.5}{\sqrt{n}}, A_n + 2\frac{0.5}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } p.$$

Thus,

$$[A_n - \frac{1}{\sqrt{n}}, A_n + \frac{1}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } p.$$

Summary

$\boxed{\text{Gaussian and CLT}}$

1. Gaussian: $\mathcal{N}(\mu, \sigma^2) : f_X(x) = \ldots$ "bell curve"
2. CLT: X_n i.i.d. $\implies \frac{A_n - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1)$
3. CI: $[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] = 95\%$-CI for μ.

Confusing Statistics: Simpson's Paradox

College	F. Appl.	F. Adm.	% F. Adm.	M. Appl.	M. Adm.	% M. Adm.
A	980	490	50%	200	80	40%
B	20	20	100%	800	720	90%
Total	1000	510	51%	1000	800	80%

The numbers are applications and admissions of males and females to the two colleges of a university.

Overall, the admission rate of male students is 80% whereas it is only 51% for female students.

A closer look shows that the admission rate is larger for female students in both colleges....

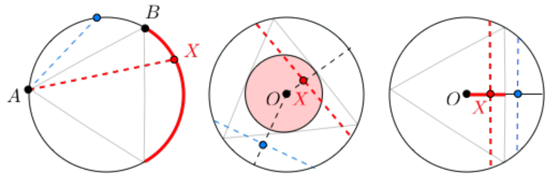
Female students happen to apply more to the college that admits fewer students.

More on Confusing Statistics

Statistics are often confusing:

- The average household annual income in the US is $72k. Yes, but the median is $52k.
- The false alarm rate for prostate cancer is only 1%. Great, but only 1 person in 8,000 has that cancer. So, there are 80 false alarms for each actual case.
- The Texas sharpshooter fallacy. Look at people living close to power lines. You find clusters of cancers. You will also find such clusters when looking at people eating kale.
- False causation. Vaccines cause autism. Both vaccination and autism rates increased....
- Beware of statistics reported in the media!

Choosing at Random: Bertrand's Paradox



The figures corresponds to three ways of choosing a chord "at random." The probability that the chord is larger than the side $|AB|$ of an inscribed equilateral triangle is

- $1/3$ if you choose a point A, then another point X uniformly at random on the circumference (left).
- $1/4$ if you choose a point X uniformly at random in the circle and draw the chord perpendicular to the radius that goes through X (center).
- $1/2$ if you choose a point X uniformly at random on a given radius and draw the chord perpendicular to the radius that goes through X (right).

Confirmation Bias

Confirmation bias is the tendency to search for, interpret, and recall information in a way that confirms one's beliefs or hypotheses, while giving disproportionately less consideration to alternative possibilities.

Confirmation biases contribute to overconfidence in personal beliefs and can maintain or strengthen beliefs in the face of contrary evidence.

Three aspects:

- Biased search for information. E.g., ignoring articles that dispute your beliefs.
- Biased interpretation. E.g., putting more weight on confirmation than on contrary evidence.
- Biased memory. E.g., remembering facts that confirm your beliefs and forgetting others.

Confirmation Bias: An experiment

There are two bags. One with 60% red balls and 40% blue balls; the other with the opposite fractions.

One selects one of the two bags.

As one draws balls one at time, one asks people to declare whether they think one draws from the first or second bag.

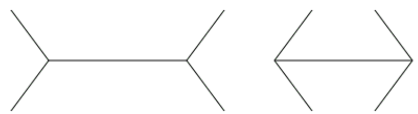
Surprisingly, people tend to be reinforced in their original belief, even when the evidence accumulates against it.

Being Rational: 'Thinking, Fast and Slow'

In this book, Daniel Kahneman discusses examples of our irrationality.

Here are a few examples:

- A judge rolls a die in the morning. In the afternoon, he has to sentence a criminal. Statistically, the sentence tends to be heavier if the outcome of the morning roll was high.
- People tend to be more convinced by articles printed in Times Roman instead of Computer Modern Sans Serif.
- Perception illusions: Which horizontal line is longer?



It is difficult to think clearly!

What to Remember?

Professor, what should I remember about probability from this course? I mean, after the final.

Here is what the prof. remembers:

- Given the uncertainty around us, we should understand some probability.
- One key idea - what we learn from observations: the role of the prior; Bayes' rule; Estimation; confidence intervals... quantifying our degree of certainty.
- This clear thinking invites us to question vague statements, and to convert them into precise ideas.

Key Ideas in CS70 Probability

- Descriptive: $Pr[A], E[X], E[h(X,Y)], \pi_n, \beta(i), \alpha(i)$
- Inference: $Pr[A|B], Pr[A_n|B], L[Y|X], E[Y|X], A_n \pm 2\frac{\sigma}{\sqrt{n}}$
- Prescriptive: How to play a game, how to design, ...

What's Next?

Professors, I loved this course so much! I want to learn more about discrete math and probability!

Funny you should ask! How about

- CS170: Efficient Algorithms and Intractable Problems a.k.a. Introduction to CS Theory: Graphs, Dynamic Programming, Complexity.
- EE126: Probability in EECS: An Application-Driven Course: PageRank, Digital Links, Tracking, Speech Recognition, Planning, etc. Hands on labs with python experiments (GPS, Shazam, ...).
- CS188: Artificial Intelligence: Hidden Markov Chains, Bayes Networks, Neural Networks.
- CS189: Introduction to Machine Learning: Regression, Neural Networks, Learning, etc. Programming experiments with real-world applications.
- EE121: Digital Communication: Coding for communication and storage.
- EE223: Stochastic Control.
- EE229A: Information Theory; EE229B: Coding Theory.

Final Thoughts

More precisely: Some thoughts about the final ....

How to study for the final?

- Lecture Slides; Notes; Discussion Problems; HW
- TA Office Hours, Prof. Office Hours, Reviews by TAs
- Next week: reviews during normal lecture hours:
  - Discrete Math (Tuesday);
  - Probability (Thursday).

Parting Thoughts

You have learned a lot in this course!

Proofs, Graphs, Mod(p), RSA, Reed-Solomon, Decidability, Probability, ... , how to handle stress, how to sleep less, how to keep smiling, ...
Difficult course? Yes! Useful? You bet!
Finally,

Thanks for taking the course!

Thanks to the CS70 Staff:

- The Terrific Tutors
- The Rigorous Readers
- The Thrilling TAs
- The Amazing Assistants

See you on Tuesday.