

# Stats 102C - Welcome

Miles Chen, PhD

Week 0 Friday

## Section 1

Syllabus Stuff

# Syllabus

Please review syllabus

## Section 2

What is this class about?

# Monte Carlo Methods

Monte Carlo methods use repeated random sampling to estimate a numeric answer to a problem that would be difficult to answer analytically.

## A short history (adapted from Wikipedia)

It was invented by Stanislaw Ulam while working on the Manhattan Project at Los Alamos National Laboratory. The scientists had trouble using conventional mathematical methods to solve the calculations needed.

Ulam was inspired by a question he had when playing the card game solitaire. What is the probability that a random deal of the game would be winnable? Solving it using combinatorial calculations would be quite difficult. Perhaps an easier method would be to shuffle the deck, play 100 different random games, and see what percentage were winnable. This would provide an estimate of the win percentage.

Ulam described the method to his colleague, John von Neumann, who coded the method into the ENIAC mainframe computer.

Because the work on the atomic bomb was highly secretive, they used a code name: Monte Carlo, named after the Monte Carlo casino in Monaco.

## Calculations involving random variables

In Stats 100B, you learned how to find the properties of probability distributions.

For example, let's say you have some random variable  $X$  with probability function:

$$X \sim f(x)$$

If continuous, the expected value of this distribution is:

## Calculations involving random variables

In Stats 100B, you learned how to find the properties of probability distributions.

For example, let's say you have some random variable  $X$  with probability function:

$$X \sim f(x)$$

If continuous, the expected value of this distribution is:

$$\mu = E[X] = \int_{\mathcal{X}} xf(x)dx$$

Where  $\mathcal{X}$  represents the support of  $f$ , (e.g.  $-\infty$  to  $\infty$  for the normal distribution)

## Calculations involving random variables

In Stats 100B, you learned how to find the properties of probability distributions.

For example, let's say you have some random variable  $X$  with probability function:

$$X \sim f(x)$$

If continuous, the expected value of this distribution is:

$$\mu = E[X] = \int_{\mathcal{X}} xf(x)dx$$

Where  $\mathcal{X}$  represents the support of  $f$ , (e.g.  $-\infty$  to  $\infty$  for the normal distribution)

If discrete, the expected value of this distribution is:

## Calculations involving random variables

In Stats 100B, you learned how to find the properties of probability distributions.

For example, let's say you have some random variable  $X$  with probability function:

$$X \sim f(x)$$

If continuous, the expected value of this distribution is:

$$\mu = E[X] = \int_{\mathcal{X}} xf(x)dx$$

Where  $\mathcal{X}$  represents the support of  $f$ , (e.g.  $-\infty$  to  $\infty$  for the normal distribution)

If discrete, the expected value of this distribution is:

$$\mu = E[X] = \sum_{i=1}^{\infty} x_i f(x_i)$$

# Variance of a random variables

The variance of the random variable can be calculated with:

$$\sigma^2 = E[(X - \mu)^2] = \int_{\mathcal{X}} (x - \mu)^2 f(x) dx$$

Which can also be calculated with

$$\sigma^2 = E[X^2] - E[X]^2 = \int_{\mathcal{X}} x^2 f(x) dx - \mu^2$$

(For the discrete version replace the integral with a sum)

## Arbitrary calculations involving random variables

The expected value of any arbitrary function ( $g(x)$ ) of the random variable is:

$$E[g(X)] = \int_{\mathcal{X}} g(x)f(x)dx$$

# The Big Picture Monte Carlo Methods

In many cases involving probability distributions, the integral (or sum) is too complicated to solve analytically.

Instead of trying to get the exact analytic answer, we can get an estimate based on simulated randomness.

## Example

We can estimate the mean of a random variable  $X$  by simulating random draws of  $X$  from  $f(x)$ .

Let's say we generate a random sample of  $N$  values  $(x_1, x_2, \dots, x_N)$  drawn from the distribution defined by  $f(x)$ .

The mean of  $X$  can be estimated by finding the sample mean of the random sample.

$$\mu_{\text{estimated}} = \frac{1}{N} \sum_{i=1}^N x_i$$

## Example - Chi squared distribution - Analytic method

What is the mean of a chi-squared random variable with 5 degrees of freedom?

## Example - Chi squared distribution - Analytic method

What is the mean of a chi-squared random variable with 5 degrees of freedom?

[https://en.wikipedia.org/wiki/Chi-square\\_distribution](https://en.wikipedia.org/wiki/Chi-square_distribution)

PDF:

$$\frac{1}{2^{k/2}\Gamma(k/2)}x^{(k/2-1)}e^{-x/2}$$

## Example - Chi squared distribution - Analytic method

This slide is review and you will not be tested on the Moment generating function.

The moment generating function of the chi-squared distribution is:

$$E[e^{tX}] = (1 - 2t)^{\frac{-k}{2}}$$

The mean is equal to the first moment, which is the first derivative of the moment generating function:

$$\begin{aligned}\mu &= \frac{d}{dt}(1 - 2t)^{\frac{-k}{2}} \Big|_{t=0} \\ &= \frac{-k}{2}(1 - 2t)^{\left(\frac{-k}{2} - 1\right)}(-2) \Big|_{t=0} \\ &= \frac{-k}{2}(1)(-2) \\ &= k\end{aligned}$$

Analytic answer: 5

Copyright Miles Chen. For personal use only. Do not distribute.

## Example - Monte Carlo Method

What is the mean of a chi-squared random variable with 5 degrees of freedom?

Estimation via Monte Carlo:

```
set.seed(1)
n <- 10^5
samp <- rchisq(n, df = 5)
mean(samp)

## [1] 4.993096
```

Our Monte Carlo estimate of the mean is close to the analytic answer. With our computer able to generate values from the chi-squared distribution quickly, it is easy to run.

## Example - What about the variance?

What is the variance of a chi-squared random variable with 5 degrees of freedom?

## Example - What about the variance?

What is the variance of a chi-squared random variable with 5 degrees of freedom?

If you go through the analytics (using the moment generating function or some other method), you will find that the variance is equal to degrees of freedom times 2.

Estimation via Monte Carlo:

```
var(samp)
```

```
## [1] 9.978977
```

A pretty good estimate! And very easy!

## Where we are headed:

Monte Carlo methods are extremely useful for Bayesian statistics because we frequently end up with big and messy integrals.

MC methods are also very useful in high dimensional settings where the math gets messy.

Through random simulations, Monte Carlo methods can produce estimates without the need to untangle the hard or impossible math problems.

## Plan for course topics:

- Core Concepts of Bayesian Statistics
- Monte Carlo Integration with R's built-in random number generators
- Generating random values for variables with unusual distributions
- Importance Sampling
- Markov Chains, Ergodic Theory
- Intro to Markov Chain Monte Carlo
- Metropolis Algorithm
- Gibbs Sampler

## Section 3

### Academic Integrity

# Academic Integrity and Plagiarism

Some truths:

- There is a lot of high quality code that does exactly what you need available out on the Internet. Some of it is available in ready-to-install packages and some are available as solutions on places like stackexchange and github.
- If the goal is to accomplish a task, you should probably use the readily available packages or code solutions out there.
- However, the goal of this class is not to get you to accomplish some task. The goal is to help you learn how to write code.

# No Pain, No Gain

Think of the gym. The goal of lifting weights at the gym is not to lift weights. Lifting weights is a means to the real goal of gaining strength.

“No Pain, No Gain”: if your weight training does not result in some muscle soreness, you probably did not exert enough effort to expect muscle gain. Experiencing muscle soreness is a sign that your muscles will go through repairs and get stronger.

Your brain is similar: if your brain does not struggle when writing code, then it has no reason to create additional neuron connections that will improve your abilities as a coder. On the other hand, if your brain struggles with writing code, then your brain will try to create new connections between neurons so the next time will not be as hard. And thus you become a better coder.

# No Pain, No Gain

Plagiarizing code for a difficult assignment is like having a stronger person lift the weights that are too heavy for you.

This would be a good solution if the goal of lifting weights was to lift the weights. But this does not help for the goal of gaining strength.

Similarly, copying, pasting, and modifying a stronger programmer's code works if the goal is to accomplish a task. It does not help towards the goal of creating neuron connections in the brain that will make you a better coder.

## Course Goals

I believe students resort to plagiarism because they have confused the goals of the course.

Students who plagiarize believe the goal for them is to get a good grade (or avoid a bad grade) in the class. For these students, the goal of learning is secondary to the goal of getting the desired grade.

But this is wrong! The goal of the course is your learning.

I will admit, a major conflicting issue here is that I am not able to create individualized grading schemas that evaluate exactly how much each student learned over the course. All students are graded on the same criteria and evaluated on what they turn in for the assignments.

That said, I hope you can judge your performance in a class based on what you learned and not your letter grade.

## Section 4

Grades and Life

# Your grades do not define you

You are here at UCLA. One of the reasons you got into UCLA was because you had good grades in high school and/or at community college. While you are in school, a lot of your energy is poured into your classes and I can understand why grades feel so important. That said,

## **Your grades do not define you**

It feels good to get good grades. Grades do play a role in graduate school admissions. But they are not the most important thing in life. No one on their death bed looks back and says “I wish I got an A- instead of a B+ in that one college class.”

# Work - Life Balance

I like to split where we put our energies of life into three broad categories:

- Work
  - ▶ Jobs and internships
  - ▶ School and academics
  - ▶ Other professional obligations
- Relationships
  - ▶ Family
  - ▶ Friends
  - ▶ Romantic partner
  - ▶ Other social obligations
- Self
  - ▶ Care of physical health (food, sleep, exercise)
  - ▶ Care of mental health (sleep, play, entertainment)
  - ▶ Care of spiritual health (if you are spiritual/religious)

There are 24 hours in a day. It is not possible to give 100% to all categories

# Work - Life Balance

Work-Life balance is achieved by consciously choosing what is important to you and devoting your time and energies accordingly.

In general, the more you put in, the more you get out.

Satisfaction can be found by accepting the natural consequences of what you have chosen to deprioritize.

## Work - Life Balance

Let's say you are part of a group of friends. Let's say that one day you become involved with a romantic partner.

If you choose to invest all of your relationship hours into your romantic partner, you will likely develop a very strong relationship with your romantic partner. However, because you now invest much less into your original group of friends, those relationships will naturally become more distant. When you see distance forming, it can initially feel hostile. This is not (necessarily) the result of your friends being angry that you have a romantic partner but the natural consequence of having less time to spend with them.

As people, we have to make a choice about what is important to us.

When you accept the natural consequences of investing less time into something, you can reduce your own feelings of bitterness and jealousy.

## Work - Life Balance

In the corporate and professional world, people who devote a lot of energy into the goals of the company are rewarded. The company is not necessarily punishing people who choose to have families and a life outside of work.

From the company's perspective, who would they rather promote?

- the person who did everything asked of them and then continued to stay at work and did even more
- the person who did everything asked of them and then immediately left to spend time with their family/friends/romantic partner

You have to choose what is important to you. If climbing the ranks within the company is more important, then you will spend your time accordingly. If spending time with your family/friends/romantic partner is more important, then spend your time accordingly.

# Self care is important

You must not neglect taking care of your physical and mental health.

If you neglect care of self, you will likely operate at less than 100% efficiency and the time you invest in work/school/relationships will not be as productive.

Examples:

- You don't get enough sleep. A friend invites you out. You choose to accept your friend's invitation instead of sleep, but you are a bit 'out of it' and are a drag to hang around. Maybe it would have been better to decline your friend's invitation and get sleep.
- Exams are coming up. You choose to skip a meal and minimize sleep to study. You end up getting sick. Your performance on the exam suffers. Maybe it would have been better to eat properly, sleep well, and study a bit less.

# Self care is important

When I tell you that your physical and mental health is important, I'm encouraging you to choose to invest your time into activities like exercise, sleep, and relaxation that will boost your physical and mental health.

Sometimes this means choosing not to complete your homework to 100%.

The natural consequence of this is a homework grade that is less than 100%.

When you can readily embrace this natural consequence of prioritizing your own physical and mental health over your homework grade, you can enjoy the quarter with less bitterness, more joy and better health.

## Beware of “fruitless” entertainment

Entertainment and fun activities are important for your mental well being. It's important to have fun.

I love hanging out with people I like, watching TV, movies, sports, playing board games, video games, going on hikes, browsing the Internet, reading a book, listening to podcasts, etc.

Participating in an entertainment activity should be a break from work and should give you mental energy so you can return to your work in a good mood.

Some activities can have the opposite effect - they drain you. Some video games, apps, and social media sites are designed to be addictive - giving your brain immediate dopamine pleasure hits while you use them so you play round after round or continue scrolling forever (and keep coming back) . . . but after spending hours of doing the activity, you don't feel good about yourself.

Be mindful and selective about your entertainment activities.