Monte Carlo (No Markov Chains. . . Yet)

Goals of this lecture:

- Define the Monte Carlo method generally
- Explain why Monte Carlo is useful
- Understand variance of Monte Carlo estimators

all without talking about Markov chains. . .

Journal of the
AMERICAN STATISTICAL ASSOCIATION

SEPTEMBER 1949

A general definition of the Monte Carlo method:

DEFINITION: *Monte Carlo is the art of approximating an expectation by the sample mean of a function of simulated random variables.*

This definition is general enough to encompass everything that has been called "Monte Carlo," yet also makes clear its essence in familiar terms: Monte Carlo is about invoking laws of large numbers to approximate expectations.

REVIEW: *Expectations*.

$$\mathbb{E}[g(X)] = \int_{x \in \mathcal{X}} g(x)p(x)dx \qquad \text{or} \qquad \mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x)p(x)$$

Breaking this down with a discrete example:

Imagine an organism with a distribution of family sizes (i.e., number of offspring), s:

s	$p(s)$
0	0.0184
1	0.0735
2	0.1469
3	0.1959
4	0.1959
5	0.1567
6	0.1045
7	0.0597
8	0.0299
9	0.0133
10	0.0053

The expected family size is:

$$\mathbb{E}[S] = \sum_{s=0}^{10} sp(s).$$

But, perhaps you are interested in the expected number of sibling pairs in a family. If there are s offspring, there are $g(s) = s(s-1)/2$ sibling pairs in it.

$$\mathbb{E}[g(S)] = \sum_{s=0}^{10} [s(s-1)/2]p(s).$$

Important: For every Monte Carlo approximation there is an underlying expectation and a $g(x)$. Get in the habit of identifying those.

Laws of large numbers::

REVIEW: *Weak Law of Large Numbers*.
Let $X_1, X_2, \ldots, X_n$ be iid r.v.'s with $\mathbb{E}|X_i| < \infty$, and $\bar{X}_n = \frac{1}{n}\sum_1^n X_i$.
Then

$$\lim_{n\to\infty} P(|\bar{X}_n - \mathbb{E}X_i| > \epsilon) = 0 \quad \text{for any } \epsilon > 0$$

Basically: if you take a sample mean of random variables, as your sample size gets very large, the sample mean gets very close to the expectation.

This suggests: an expectation might be approximated by the sample mean of n random variables (and it will work better if n is large).

$$\mathbb{E}[g(X)] \approx \frac{1}{n}\sum_{i=1}^{n} g(x^{(i)}) \quad \text{with} \quad X^{(i)} \sim p(x)$$

Why estimating expectations is useful:

Usually, any quantity of interest may be expressed as the expected value of a function of some random variable. Importantly:

Probabilities:

$$P(X \in \mathcal{A}) = \mathbb{E}[I_{\{\mathcal{A}\}}(X)]$$

where $I_{\{\mathcal{A}\}}(X)$ is the *indicator function* taking the value 1 when $X \in \mathcal{A}$ and 0 otherwise.

Integrals: For a simple example, let U be a uniform r.v. on the interval $[a, b)$ with pdf $p(u) = 1/(b-a)$. Hence

$$\int_a^b q(x)dx = (b-a)\int_a^b q(x)\frac{1}{b-a}dx = (b-a)\mathbb{E}[q(U)]$$

We see here a case where Monte Carlo applies to a purely deterministic problem.

Discrete Sums: In the same vein as above, just as any integral can be approximated by Monte Carlo, so can any sum. For another simple, uniform example, let W be a discrete random variable that takes all values w in the set $\mathcal{A}$ with equal probability P. Then, the sum $\sum_{w\in\mathcal{A}} q(w)$ is easily approximated by Monte Carlo:

$$\sum_{w\in\mathcal{A}} q(w) = \frac{1}{P}\sum_{w\in\mathcal{A}} q(w)P = \frac{1}{P}\mathbb{E}[q(W)].$$

This is particularly useful in statistical genetics, because many probabilities of interest may be expressed as an intractable sum over latent variables.

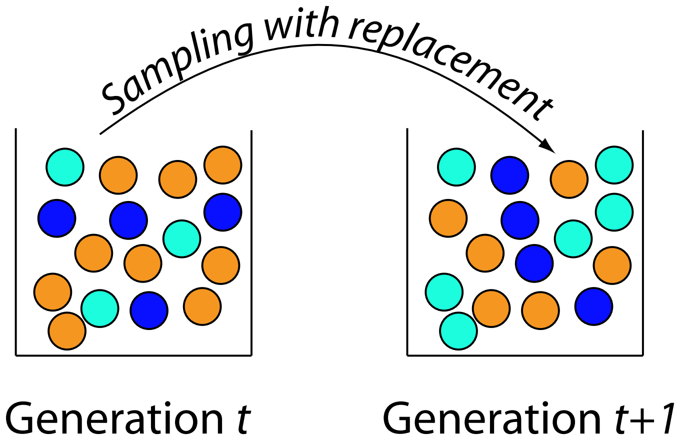
The bottom line is that, since you can cast any quantity as an expectation; you can (in theory) approximate any quantity by Monte Carlo. In the simplest case, when X is distributed according to $p(x)$.

$$\mathbb{E}[g(X)] \approx \frac{1}{n}\sum_{i=1}^{n} g(x^{(i)}) \quad \text{with} \quad X^{(i)} \sim p(x)$$

Genetics example I: Estimating probabilities in the Wright-Fisher model:

The Wright-Fisher model underlies most population genetics theory that deals with the descent of genes in a finite population from one generation to the next.

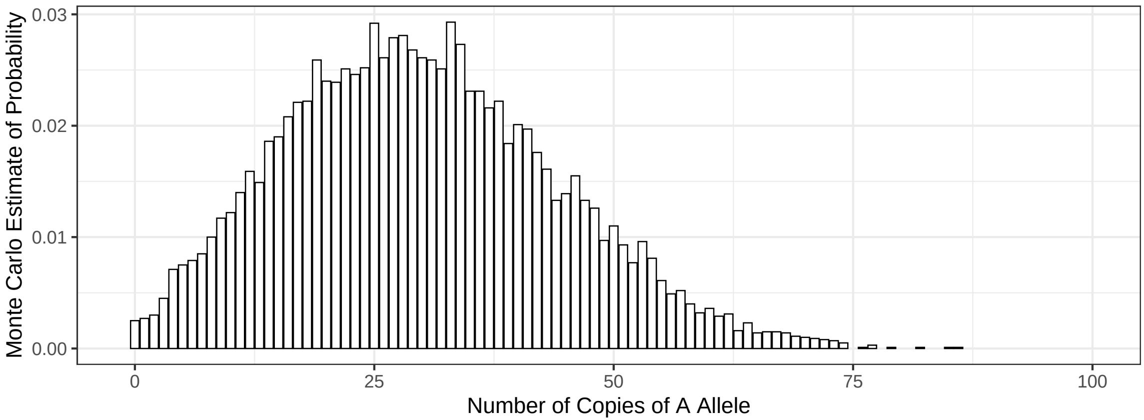
A schematic of its underlying assumptions:



Genetics example I: Estimating probabilities in the Wright-Fisher model:

Let X be the frequency of the A allele in a Wright-Fisher population of size $N = 100$ after 10 generations of drift, having started from a frequency of 30 out of 100.

The distribution of X can be approximated by simulating $n = 10,000$ instances:



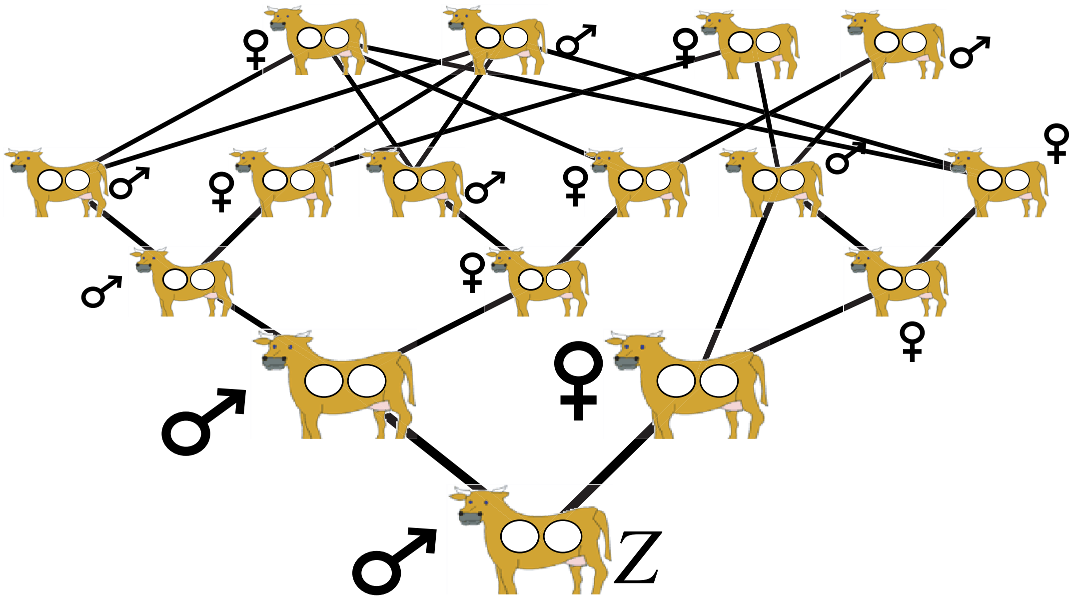
Each histogram column is an approximation of an expectation:

$$\begin{aligned} P(X = a) &= \mathbb{E}[I_{\{x=a\}}(X)] \\ &\approx \frac{1}{n}\sum_{i=1}^{n} I_{\{x=a\}}(x^{(i)}) \end{aligned}$$

for $a = 0, 1, \ldots, 100$, where each $x^{(i)}$ is an independent realization of the number of A alleles at $t = 10$ in the Wright-Fisher model.

Genetics example II: estimating a sum over latent variables:

Wright and McPhee (1925) estimating inbreeding of individuals within cattle pedigrees.



In terms of latent variables:

- The maternal (m) and paternal (p) genes in individual z are direct descendants from an *unobserved* lineages of ancestral gene copies, A_m and A_p from z up to a founder in the pedigree.
- The individual is inbred at a locus if the ancestors of the maternal gene and the paternal gene are the same at any point in their lineages, *i.e.*, if the lineages A_m and A_p hit one another going back up the pedigree.
- The probability that an individual is inbred at a locus is then the sum over latent variables:

$$P(\text{inbred}|\text{pedigree}) = \sum_{A_m, A_p} [I_{\{A_m \text{ ``hits'' } A_p\}}(A_m, A_p)] P(A_m, A_p)$$

where $P(A_m, A_p)$ follows from Mendel's laws. Note that A_m and A_p are independent up until they intersect, then follow the same lineage back in time.

- This sum is an expectation of the indicator function, with respect to the joint distribution of A_m and A_p, given the pedigree:

$$P(\text{inbred}|\text{pedigree}) = \mathbb{E}[I_{\{A_m \text{ ``hits'' } A_p\}}]$$

so it may be estimated by Monte Carlo:

$$\begin{aligned} P(\text{inbred}|\text{pedigree}) &= \mathbb{E}[I_{\{A_m \text{ ``hits'' } A_p\}}] \\ &\approx \frac{1}{n}\sum_{i=1}^{n} I_{\{A_m \text{ ``hits'' } A_p\}} \end{aligned}$$

where $A_m^{(i)}$ and $A_p^{(i)}$ are simulated from their respective distributions, which can be done by flipping a coin, until they intersect (and hence add 1 to the sum) or reach pedigree founders without intersecting (adding 0 to the sum).

Sampling From Posterior Distributions:

- In Bayesian statistics, Monte Carlo sampling is often done from the posterior distribution.
- Imagine a posterior distribution for Q that is a $\mathsf{Beta}(30, 70)$ distribution.
- If you want to know the posterior probability that $Q > 0.35$ that probability is an expectation:

$$P(Q > 0.35) = \mathbb{E}[I_{\{q>0.35\}}(Q)]$$

- Here, $g(Q) = I_{\{q>0.35\}}(Q)$, a function that returns a 1 if $Q > 0.35$ and a 0 otherwise. This expectation can be approximated by Monte Carlo with

$$P(Q > 0.35) = \mathbb{E}[I_{\{q>0.35\}}(Q)] \approx \frac{1}{n}\sum_{i=1}^{n} I_{\{q>0.35\}}(q^{(i)})$$

Variance of Monte Carlo estimators—iid case:

A Monte Carlo estimator is simply a random variable itself—a sum of random variables:

$$G_n = \frac{1}{n}\sum_{i=1}^{n} g(X^{(i)})$$

So, if the X_i are independent[1] the variance of G_n is easily computed as the variance of a sum of independent R.V.'s:

$$\begin{aligned}\text{Var}(G_n) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} g(X^{(i)})\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}[g(X^{(i)})] \\ &= \frac{\text{Var}[g(X^{(i)})]}{n}\end{aligned}$$

Var $\downarrow$ when $n\uparrow$ or if Var$[g(X^{(i)})]$ can be reduced[2].

[1] Note that throughout most of the remainder of the course, we will deal with samples of correlated, non-independent X_i's. This is just a warm-up.

[2] We'll take this up later in our discussion importance sampling