Monte Carlo methods rely on random numbers to approximate integrals.

In this lecture we’ll see some statistical problems involving integrals, and discuss the properties of the basic Monte Carlo estimator.

We will see some basic methods for sampling from distributions: inversion, transformation, rejection sampling...
Monte Carlo Integration

- We are interested in computing

\[ I = \int_X \varphi(x) \pi(x) \, dx \]

where \( \pi \) is a pdf on \( X \) and \( \varphi : X \to \mathbb{R} \).

- Monte Carlo method:
  - sample \( n \) independent copies \( X_1, \ldots, X_n \) of \( X \sim \pi \),

\[ \hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i). \]

- **Remark:** You can think of it as having the following empirical measure approximation of \( \pi(dx) \)

\[ \hat{\pi}_n(dx) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(dx) \]

where \( \delta_{X_i}(dx) \) is the Dirac measure at \( X_i \).
Monte Carlo Integration: Limit Theorems

Proposition (LLN)
If \( \mathbb{E}(|\varphi(X)|) < \infty \) then \( \hat{I}_n \) is a strongly consistent estimator of \( I \).

Proposition (CLT)
If
\[
\sigma^2 = \mathbb{V}(\varphi(X)) = \int_X [\varphi(x) - I]^2 \pi(x) \, dx < \infty
\]
then
\[
\mathbb{E}\left(\left(\hat{I}_n - I\right)^2\right) = \mathbb{V}(\hat{I}_n) = \frac{\sigma^2}{n}
\]
and
\[
\frac{\sqrt{n}}{\sigma} \left(\hat{I}_n - I\right) \overset{D}{\to} \mathcal{N}(0, 1).
\]
Proposition

Assume $\sigma^2 = V(\varphi(X)) < \infty$ then

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left( \varphi(X_i) - \hat{I}_n \right)^2$$

is an unbiased sample variance estimator of $\sigma^2$. 
Proof.

Let $Y_i = \varphi(X_i)$ then we have

\[
\mathbb{E}(S_n^2) = \frac{1}{n-1} \sum_{i=1}^{n} \mathbb{E}((Y_i - \bar{Y})^2) = \frac{1}{n-1} \mathbb{E} \left( \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 \right)
\]

\[
\mathbb{E}(\bar{Y}^2) = \frac{1}{n^2} \mathbb{E} \left[ \sum_{i=1}^{n} Y_i^2 + \sum_{i \neq j} Y_i Y_j \right] = \frac{1}{n} (V(Y) + I^2) + \frac{n-1}{n} I^2
\]

\[
= \frac{V(Y)}{n} + I^2
\]

\[
\mathbb{E}(S_n^2) = \frac{n}{n-1} V(Y) - \frac{n}{n-1} \frac{V(Y)}{n} + \frac{n}{n-1} I^2 - \frac{n}{n-1} I^2
\]

\[
= V(Y) = V(\varphi(X)).
\]
Monte Carlo Integration: Error Estimates

- **Chebyshev’s inequality**: exact but possibly rough

\[ P \left( \left| \hat{I}_n - I \right| > c \frac{\sigma}{\sqrt{n}} \right) \leq \frac{\text{Var} (\hat{I}_n)}{c^2 \sigma^2 / n} = \frac{1}{c^2}. \]

- **CLT**: much tighter but approximate and for large \( n \)

\[ P \left( \left| \hat{I}_n - I \right| > c \frac{\sigma}{\sqrt{n}} \right) \approx 2 (1 - \Phi (c)) = O \left( \frac{\text{e}^{-c^2/2}}{c} \right). \]

- Choosing \( c = c_\alpha \) s.t. \( 2 (1 - \Phi (c_\alpha)) = \alpha \), an approximate \( (1 - \alpha) 100\% \)-CI for \( I \) is

\[ \left( \hat{I}_n \pm c_\alpha \frac{\sigma}{\sqrt{n}} \right) \approx \left( \hat{I}_n \pm c_\alpha \frac{S_n}{\sqrt{n}} \right) \]

and the rate is in \( 1/\sqrt{n} \) whatever \( X \).
Consider the case where we have a square $S \subseteq \mathbb{R}^2$, sides of length 2, with inscribed disk $D$ of radius 1.

Use Monte Carlo to compute the area $I$ of $D$.

\[
I = \pi = \iint_D d\mathbf{x}_1 d\mathbf{x}_2 = \iint_S \mathbb{1}_D (\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \text{ as } D \subset S = 4 \iint_{\mathbb{R}^2} \mathbb{1}_D (\mathbf{x}_1, \mathbf{x}_2) \pi (\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2
\]

where $S := [-1, 1] \times [-1, 1]$ and

\[
\pi (\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{4} \mathbb{1}_S (\mathbf{x}_1, \mathbf{x}_2)
\]

is the uniform density on the square $S$. 
Figure: \( \hat{I}_n = 4 \frac{n_D}{n} \) where \( n_D \) is the number of samples which fell within the disk.
Figure: Relative error of $\hat{I}_n$ against the number of samples.
Computing intricate high-dimensional integrals boils down to generating random variables from complicated distributions.

How does a computer simulate random variables?

Firstly it can produce a random integer uniformly distributed in \( \{0, \ldots, M - 1\} \) for some large \( M \), often \( M = 2^{32} \) giving 32-bit integers.

These are pseudo-random numbers.

Then various techniques are used to produce all others distributions of interest.
Start off with a “seed” $x_0$.

- Given $x_n$, produce

$$x_{n+1} = (ax_n + c) \pmod{M},$$

for integers $a$, $c$, and $M$.

- Maximum period $M$.

- Hull and Dobell (1962) provide necessary and sufficient conditions for period $M$.

- Then $U_n = X_n / M$ behaves similarly to $U[0, 1]$ random variable, despite not being random at all.
Figure: **Left:** 10,000 pseudo random numbers in [0, 1]; **Right:** histogram.
**Assumption:** we have access to i.i.d. \( (U_i, i \geq 1) \sim U_{[0,1]} \).

To simulate from \( \pi (x_1, x_2) = \frac{1}{4} \mathbb{1}_S (x_1, x_2) \), we draw \( U_1 \) and \( U_2 \) uniformly and define \( X_1 = 2U_1 - 1, \ X_2 = 2U_2 - 1 \). Then the point \( (X_1, X_2) \) is distributed uniformly within \( S \).

We will see how to use the above to simulate many different random variables.
Galton’s machine to draw normal samples
Consider a real-valued random variable $X$ and its associated cumulative distribution function (cdf)

$$F(x) = \mathbb{P}(X \leq x) = F(x).$$

The cdf $F : \mathbb{R} \to [0, 1]$ is

- increasing; i.e. if $x \leq y$ then $F(x) \leq F(y)$,
- right continuous; i.e. $F(x + \varepsilon) \to F(x)$ as $\varepsilon \to 0^+$,
- $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to +\infty$.

We define the generalised inverse

$$F^{-}(u) = \inf \{x \in \mathbb{R}; F(x) \geq u\}$$

also known as the quantile function.
**Inversion Method**

\[ F^{-1}(u) = x \]

**Figure:** Cumulative distribution function $F$ and representation of the inverse cumulative distribution function.
**Proposition**

Let $F$ be a cdf and $U \sim U_{[0,1]}$. Then $X = F^-(U)$ has cdf $F$.

In other words, to sample from a distribution with cdf $F$, we can sample $U \sim U_{[0,1]}$ and then return $F^-(U)$.

**Proof.**

**Fact:** $F^-(u) \leq x \iff u \leq F(x)$.

Thus for $U \sim U_{[0,1]}$, we have

$$
P(F^-(U) \leq x) = P(U \leq F(x)) = F(x).$$
Examples

- **Exponential distribution.** If \( F(x) = 1 - e^{-\lambda x} \), then \( F^{-1}(u) = F^{-1}(u) = -\log (1 - u) / \lambda \).

Thus when \( U \sim U[0,1] \),

\[-\log (1 - U) / \lambda \sim \text{Exp}(\lambda), \quad \text{and} \quad -\log (U) / \lambda \sim \text{Exp}(\lambda).\]

- **Discrete distribution.** Assume \( X \) takes values \( x_1 < x_2 < \cdots \) with probability \( p_1, p_2, \ldots \) so

\[ F(x) = \sum \limits_{x_k \leq x} p_k, \]

\[ F^{-1}(u) = x_k \text{ for } p_1 + \cdots + p_{k-1} < u \leq p_1 + \cdots + p_k. \]
Setting:
- We can simulate $Y \sim q, Y \in \mathbb{Y}$.
- We want to simulate: $X \sim \pi, X \in \mathbb{X}$.
- **Transformation method:** find a function $\varphi : \mathbb{Y} \rightarrow \mathbb{X}$ such that
  
  $$Y \sim q \implies X = \varphi(Y) \sim \pi.$$  

- Inversion is a special case of this idea.
Transformation Method

- **Gamma distribution.** For $\alpha \in \mathbb{N}$, let $Y_i, i = 1, 2, \cdots$, be i.i.d. with $Y_i \sim \text{Exp}(1)$. Then

$$X := \beta^{-1} \sum_{i=1}^{\alpha} Y_i \sim \mathcal{G}(\alpha, \beta).$$

**Proof.** The moment generating function of $X$ is

$$E \left( e^{tX} \right) = \prod_{i=1}^{\alpha} E \left( e^{\beta^{-1}tY_i} \right) = \frac{1}{(1 - t/\beta)^\alpha},$$

which is the MGF of the Gamma density with param’s $\alpha$ and $\beta$

$$\pi(x) \propto x^{\alpha-1} \exp(-\beta x).$$

- **Beta distribution.** See Exercise sheet 1.
■ **Gaussian distribution.** Let $U_1 \sim \mathcal{U}_{[0,1]}$ and $U_2 \sim \mathcal{U}_{[0,1]}$ be independent and set

$$R = \sqrt{-2 \log(U_1)}, \ \vartheta = 2\pi U_2.$$ 

Clearly $R, \vartheta$ independent and $R^2 \sim \text{Exp}(1/2)$, $\vartheta \sim \mathcal{U}_{[0,2\pi]}$ with joint density

$$q(r^2, \vartheta) = \frac{1}{2\pi} \frac{1}{2} \exp(-r^2/2).$$

Set $X = R \cos(\vartheta), Y = R \sin(\vartheta)$ a bijection.
Transformation Method - Box-Muller Algorithm

■ By standard facts:

\[
f_{X,Y}(x,y) = f_{R^2,\theta}(r^2(x,y), \theta(x,y)) \left| \det \frac{\partial(r^2, \theta)}{\partial(x,y)} \right|^{-1} = f_{R^2,\theta}(r^2(x,y), \theta(x,y)) \left| \det \frac{\partial(x,y)}{\partial(r^2, \theta)} \right|^{-1} \]

\[
= \frac{1}{2\pi} \exp \left[ -\frac{x^2 + y^2}{2} \right] 2 = \frac{1}{2\pi} \exp \left[ -\frac{x^2 + y^2}{2} \right],
\]

since

\[
\left| \det \frac{\partial(x,y)}{\partial(r^2, \theta)} \right| = \left| \begin{array}{cc}
\cos(\theta) & -r \sin(\theta) \\
\frac{\sin(\theta)}{2r} & r \cos(\theta)
\end{array} \right| = \frac{1}{2}.
\]

■ thus \((X, Y)\) are independent standard normal.
Let $Z = (Z_1, ..., Z_d)$ i.i.d. $\mathcal{N}(0, 1)$. Let $L$ be a real invertible $d \times d$ matrix satisfying $LL^T = \Sigma$, and $X = LZ + \mu$. Then $X \sim \mathcal{N}(\mu, \Sigma)$.

We have indeed $q(z) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}z^Tz\right)$ and 

$$\pi(x) = q(z) |\det \frac{\partial z}{\partial x}|$$

where $\frac{\partial z}{\partial x} = L^{-1}$ and $\det (L^{-1}) = \det (\Sigma)^{-1/2}$. Additionally,

$$z^Tz = (x - \mu)^T \left(L^{-1}\right)^T L^{-1} (x - \mu)$$

$$= (x - \mu)^T \Sigma^{-1} (x - \mu).$$

In practice, use a Cholesky factorization $\Sigma = LL^T$ where $L$ is a lower triangular matrix.
Assume we have a joint pdf $\pi$ with marginal $\pi$; i.e.

$$\pi(x) = \int \pi_{X,Y}(x,y) \, dy$$

where $\pi(x,y)$ can always be decomposed as

$$\pi_{X,Y}(x,y) = \pi_Y(y) \pi_{X|Y}(x|y).$$

It might be easy to sample from $\pi(x,y)$ whereas it is difficult/impossible to compute $\pi(x)$.

In this case, it is sufficient to sample

$$Y \sim \pi_Y \text{ then } X|Y \sim \pi_{X|Y}(\cdot|Y)$$

so $(X,Y) \sim \pi_{X,Y}$ and hence $X \sim \pi$. 

Finite Mixture of Distributions

- Assume one wants to sample from

\[ \pi(x) = \sum_{i=1}^{p} \alpha_i \pi_i(x) \]

where \( \alpha_i > 0, \sum_{i=1}^{p} \alpha_i = 1 \) and \( \pi_i(x) \geq 0, \int \pi_i(x) \, dx = 1 \).

- We can introduce \( Y \in \{1, \ldots, p\} \) and

\[ \pi_{X,Y}(x,y) = \alpha_y \times \pi_y(x). \]

- To sample from \( \pi(x) \), first sample \( Y \) from a discrete distribution such that \( \mathbb{P}(Y = k) = \alpha_k \) then

\[ X \mid (Y = y) \sim \pi_y. \]
Rejection Sampling

**Basic idea:** Sample from instrumental proposal \( q \neq \pi \); correct through rejection step to obtain a sample from \( \pi \).

**Algorithm (Rejection Sampling).** Given two densities \( \pi, q \) with \( \pi(x) \leq M q(x) \) for all \( x \), we can generate a sample from \( \pi \) by

1. Draw \( X \sim q \), draw \( U \sim U_{[0,1]} \).
2. Accept \( X = x \) as a sample from \( \pi \) if

\[
U \leq \frac{\pi(x)}{M q(x)},
\]

otherwise go to step 1.

**Proposition**

*The distribution of the samples accepted by rejection sampling is \( \pi \).*
Proof.

\[ P \left( X \in A \mid X \text{ accepted} \right) = \frac{P \left( X \in A, X \text{ accepted} \right)}{P \left( X \text{ accepted} \right)} \]

where

\[ P \left( X \in A, X \text{ accepted} \right) = \int_X \int_0^1 I_A(x) I \left( u \leq \frac{\pi(x)}{M q(x)} \right) q(x) \, du \, dx \]

\[ = \int_X I_A(x) \frac{\pi(x)}{M q(x)} q(x) \, dx \]

\[ = \int_X I_A(x) \frac{\pi(x)}{M} \, dx = \frac{\pi(A)}{M}. \]