Advanced Simulation - Lecture 2

George Deligiannidis

Outline

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

$$I = \int_{\mathbb{X}} \varphi(x)\,\pi(x)\,dx$$

where π is a pdf on $\mathbb{X}$ and $\varphi : \mathbb{X} \to \mathbb{R}$.

Monte Carlo method:

sample n independent copies $X_1, \ldots, X_n$ of $X \sim \pi$,

$$\widehat{I}_n = \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i).$$

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

$$\widehat{\pi}_n(dx) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(dx)$$

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* $\widehat{I}_n$ *is a strongly consistent estimator of* I.

Proposition (CLT)

If

$$\sigma^2 = \mathbb{V}(\varphi(X)) = \int_{\mathbb{X}} [\varphi(x) - I]^2 \pi(x)\, dx < \infty$$

then

$$\mathbb{E}\left(\left(\widehat{I}_n - I\right)^2\right) = \mathbb{V}\left(\widehat{I}_n\right) = \frac{\sigma^2}{n}$$

and

$$\frac{\sqrt{n}}{\sigma}\left(\widehat{I}_n - I\right) \xrightarrow{D} \mathcal{N}(0, 1).$$

Monte Carlo Integration: Variance Estimation I

Proposition

Assume $\sigma^2 = \mathbb{V}(\varphi(X)) < \infty$ *then*

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

is an unbiased sample variance estimator of σ^2.

Monte Carlo Integration: Variance Estimation II

Proof.

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

$$\mathbb{E}(S_n^2) = \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}((Y_i - \overline{Y})^2) = \frac{1}{n-1}\mathbb{E}\Big(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\Big)$$

$$\mathbb{E}(\overline{Y}^2) = \frac{1}{n^2}\mathbb{E}\Big[\sum Y_i^2 + \sum_{i\neq j} Y_i Y_j\Big] = \frac{1}{n}(\mathbb{V}(Y) + I^2) + \frac{n-1}{n}I^2$$

$$= \frac{\mathbb{V}(Y)}{n} + I^2$$

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

$$= \mathbb{V}(Y) = \mathbb{V}(\varphi(X)).$$

□

Monte Carlo Integration: Error Estimates

Chebyshev's inequality: exact but possibly rough

$$\mathbb{P}\left(\left|\widehat{I}_n - I\right| > c\frac{\sigma}{\sqrt{n}}\right) \leq \frac{\mathbb{V}\left(\widehat{I}_n\right)}{c^2\sigma^2/n} = \frac{1}{c^2}.$$

CLT: much tighter but approximate and for large n

$$\mathbb{P}\left(\left|\widehat{I}_n - I\right| > c\frac{\sigma}{\sqrt{n}}\right) \approx 2\left(1 - \Phi\left(c\right)\right) = \mathcal{O}\left(\frac{e^{-c^2/2}}{c}\right).$$

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

$$\left(\widehat{I}_n \pm c_\alpha\frac{\sigma}{\sqrt{n}}\right) \approx \left(\widehat{I}_n \pm c_\alpha\frac{S_n}{\sqrt{n}}\right)$$

and the rate is in $1/\sqrt{n}$ whatever $\mathbb{X}$.

Toy Example

Consider the case where we have a square $\mathcal{S} \subseteq \mathbb{R}^2$, sides of length 2, with inscribed disk $\mathcal{D}$ of radius 1.

Use Monte Carlo to compute the area I of $\mathcal{D}$.

$$
\begin{aligned}
I = \pi &= \iint_{\mathcal{D}} \mathrm{d}x_1 \mathrm{d}x_2 \\
&= \iint_{\mathcal{S}} \mathbb{I}_{\mathcal{D}}(x_1, x_2)\, \mathrm{d}x_1 \mathrm{d}x_2 \text{ as } \mathcal{D} \subset \mathcal{S} \\
&= 4 \iint_{\mathbb{R}^2} \mathbb{I}_{\mathcal{D}}(x_1, x_2)\, \pi(x_1, x_2)\, \mathrm{d}x_1 \mathrm{d}x_2
\end{aligned}
$$

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

$$
\pi(x_1, x_2) = \frac{1}{4} \mathbb{I}_{\mathcal{S}}(x_1, x_2)
$$

is the uniform density on the square $\mathcal{S}$.

Toy Example

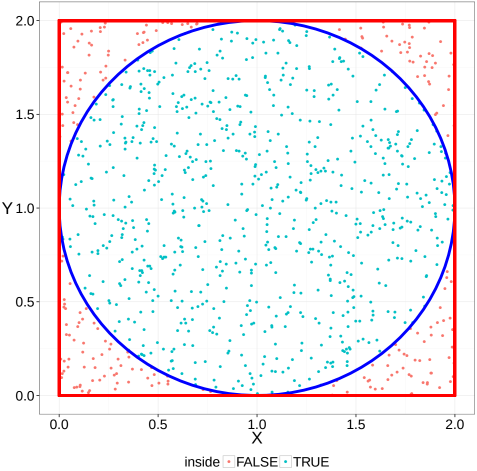


Figure: $\widehat{I}_n = 4\frac{n_{\mathcal{D}}}{n}$ where $n_{\mathcal{D}}$ is the number of samples which fell within the disk.

Toy Example

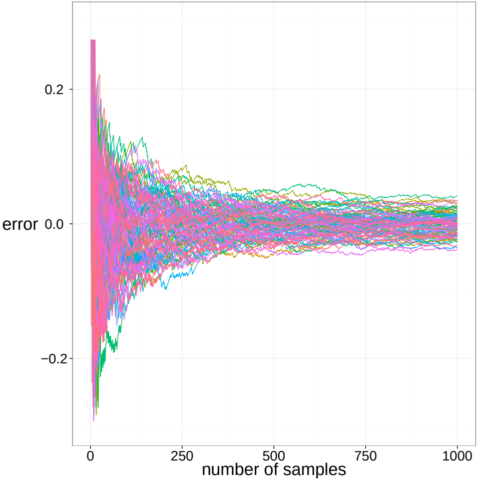


Figure: Relative error of $\widehat{I}_n$ against the number of samples.

Drawing random numbers

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.

Pseudo-Random Number Generation

Start off with a "seed" x_0.

Given x_n, produce

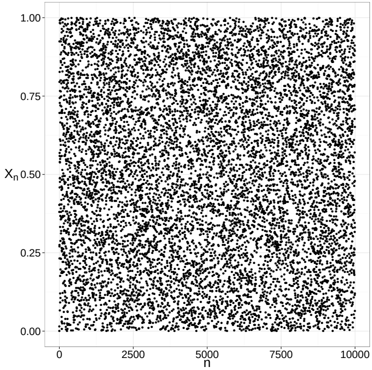
$$x_{n+1} = (ax_n + c) \mod M,$$

for integers a, c, and M.

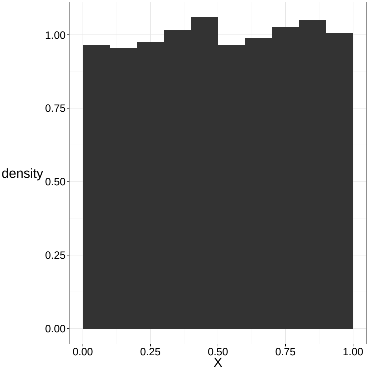
Maximum period M.

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

Then $U_n = X_n/M$ behaves similarly to $\mathcal{U}[0,1]$ random variable, despite not being random at all.



(a) Figure A



(b) Figure B

Figure: **Left:** 10,000 pseudo random numbers in $[0, 1]$;
Right: histogram.

Drawing random numbers

Assumption: we have access to i.i.d. $(U_i, i \geq 1) \sim \mathcal{U}_{[0,1]}$.

To simulate from $\pi(x_1, x_2) = \frac{1}{4}\mathbb{I}_{\mathcal{S}}(x_1, x_2)$, 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 $\mathcal{S}$.

We will see how to use the above to simulate many different random variables.

Galton's machine to draw normal samples

Inversion Method

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

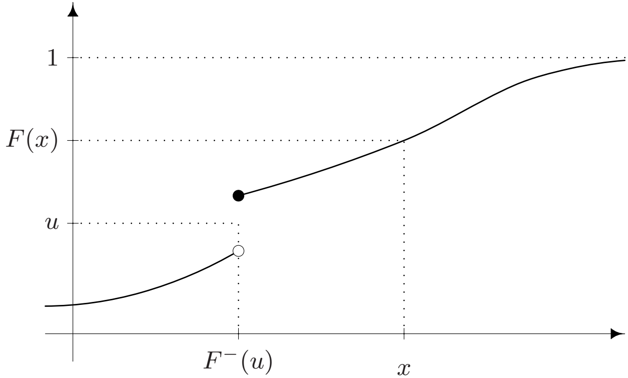


Figure: Cumulative distribution function F and representation of the inverse cumulative distribution function.

Inversion Method

Proposition

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

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

Proof.

Fact: $F^{-}(u) \leq x \Leftrightarrow u \leq F(x)$.
Thus for $U \sim \mathcal{U}_{[0,1]}$, we have

$$\mathbb{P}\left(F^{-}(U) \leq x\right) = \mathbb{P}\left(U \leq F(x)\right) = F(x). \qquad \square$$

Examples

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

Thus when $U \sim \mathcal{U}_{[0,1]}$,

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

Discrete distribution. Assume X takes values $x_1 < x_2 < \cdots$ with probability $p_1, p_2, \ldots$ so

$$F(x) = \sum_{x_k \leq x} p_k,$$

$$F^{-}(u) = x_k \text{ for } p_1 + \cdots + p_{k-1} < u \leq p_1 + \cdots + p_k.$$

Transformation Method

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 $\varphi : \mathbb{Y} \to \mathbb{X}$ such that

$$Y \sim q \Longrightarrow X = \varphi(Y) \sim \pi.$$

Inversion is a special case of this idea.

Transformation Method-Example

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

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

Proof. The moment generating function of X is

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

which is the MGF of the Gamma density with param's α and β

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

Beta distribution. See Exercise sheet 1.

Transformation Method: pushforward

Let $\mathbb{Y}, \mathbb{X}$ be two topological spaces equipped with their Borel σ-algebras.

Suppose that $f : \mathbb{Y} \mapsto \mathbb{X}$ is Borel measurable;

Suppose that q is a Borel probability measure on $\mathbb{Y}$ and let $Y \sim q$.

Write π for the distribution of $X = f(Y)$, a Borel probability measure on $\mathbb{X}$.

Then π is the *push-forward of q under f*, written

$$\pi = f_* \mu.$$

It's defined as

$$\pi(B) = (f_*)\mu(B) = q\left(f^{-1}(B)\right), \qquad \text{for all } B \in \mathcal{B}(\mathbb{X}).$$

In terms of expectations

$$\int h \circ \varphi \mathrm{d}q = \int h \mathrm{d}f_* \mu.$$

Transformation Method: change of variables formula

When $\mathrm{d}q(x) = q(x)\mathrm{d}x$, and φ is a bijection, then π also has a density given by the change of variables formula

$$\pi(x) = q \circ \varphi^{-1}(x) \left| \det(D\varphi^{-1})(x) \right| .$$

Transformation Method - Box-Muller Algorithm

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, ϑ independent and $R^2 \sim \mathrm{Exp}(1/2)$, $\vartheta \sim \mathcal{U}_{[0,2\pi]}$ 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:

$$\begin{aligned} f_{X,Y}(x,y) &= f_{R^2,\vartheta}(r^2(x,y),\theta(x,y))\Big|\det\frac{\partial(r^2,\vartheta)}{\partial(x,y)}\Big| \\ &= f_{R^2,\vartheta}(r^2(x,y),\theta(x,y))\Big|\det\frac{\partial(x,y)}{\partial(r^2,\vartheta)}\Big|^{-1} \\ &= \frac{1}{2}\frac{1}{2\pi}\exp\big[-\frac{x^2+y^2}{2}\big]2 = \frac{1}{2\pi}\exp\big[-\frac{x^2+y^2}{2}\big], \end{aligned}$$

since

$$\det\frac{\partial(x,y)}{\partial(r^2,\vartheta)}| = \begin{vmatrix} \frac{\cos(\vartheta)}{2r} & -r\sin\vartheta \\ \frac{\sin(\vartheta)}{2r} & r\cos\vartheta \end{vmatrix} = \frac{1}{2}.$$

thus (X,Y) are independent standard normal.

Transformation Method - Multivariate Normal

Let $Z = (Z_1, ..., Z_d) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.
Let L be a real invertible $d \times d$ matrix satisfying $L\ L^T = \Sigma$, and $X = LZ + \mu$. Then $X \sim \mathcal{N}(\mu, \Sigma)$.

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

$$\pi(x) = q(z) \left|\det \partial z/\partial x\right|$$

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

$$\begin{aligned} z^T z &= (x-\mu)^T \left(L^{-1}\right)^T L^{-1} (x-\mu) \\ &= (x-\mu)^T \Sigma^{-1} (x-\mu). \end{aligned}$$

In practice, use a Cholesky factorization $\Sigma = L\ L^T$ where L is a lower triangular matrix.

Sampling via Composition

Assume we have a joint pdf $\overline{\pi}$ with marginal π; i.e.

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

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

$$\overline{\pi}_{X,Y}(x,y) = \overline{\pi}_Y(y)\,\overline{\pi}_{X|Y}(x|y).$$

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

In this case, it is sufficient to sample

$$Y \sim \overline{\pi}_Y \text{ then } X|Y \sim \overline{\pi}_{X|Y}(\cdot|Y)$$

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

Latent variable models; HMMs;

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

$$\overline{\pi}_{X,Y}(x,y) = \alpha_y \times \pi_y(x).$$

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

$$X | (Y = y) \sim \pi_y.$$

Rejection Sampling

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

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

1. Draw $X \sim q$, draw $U \sim \mathcal{U}_{[0,1]}$.
2. Accept $X = x$ as a sample from π 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 π.

Rejection Sampling

Proof.

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

where

$$\begin{aligned}
&\mathbb{P}\left(X \in A, X \text{ accepted}\right) \\
&= \int_{\mathbb{X}} \int_0^1 \mathbb{I}_A(x)\, \mathbb{I}\left(u \leq \frac{\pi(x)}{M\, q(x)}\right) q(x)\, du dx \\
&= \int_{\mathbb{X}} \mathbb{I}_A(x)\, \frac{\pi(x)}{M\, q(x)} q(x)\, dx \\
&= \int_{\mathbb{X}} \mathbb{I}_A(x)\, \frac{\pi(x)}{M} dx = \frac{\pi(A)}{M}.
\end{aligned}$$

$\square$