TA Lecture 11 - Monte Carlo Methods

Dec 26 - 28

School of Information Science and Technology,
ShanghaiTech University

上海科技大学
ShanghaiTech University

Outline

Monte Carlo Methods

If you can not calculate a probability or expectation exactly, then you have three powerful strategies:

- Simulations using Monte Carlo Methods
- Approximations using limiting theorems
  - Poisson approximation: The Law of Small Numbers
  - Sample mean limit: The Law of Large Numbers
  - Normal approximation: The Central Limit Theorem
- Bounds (upper and lower bounds) on probability using inequalities.

Inverse Transform

- Given a Unif(0, 1) r.v., we can construct an r.v. with any continuous distribution we want.
- Conversely, given an r.v. with an arbitrary continuous distribution, we can create a Unif(0, 1) r.v.
- Other names:
  - probability integral transform
  - inverse transform sampling
  - the quantile transformation
  - the fundamental theorem of simulation

Theorem

Let F be a CDF which is a continuous function and strictly increasing on the support of the distribution. This ensures that the inverse function F^{-1} exists, as a function from $(0, 1)$ to $\mathbb{R}$. We then have the following results.

1. *Let $U \sim \mathrm{Unif}(0, 1)$ and $X = F^{-1}(U)$. Then X is an r.v. with CDF F.*
2. *Let X be an r.v. with CDF F. Then $F(X) \sim \mathrm{Unif}(0, 1)$.*

Inverse Transform: Continuous

Algorithm Inverse-Transform Method: PDF Case

input: Cumulative distribution function F.

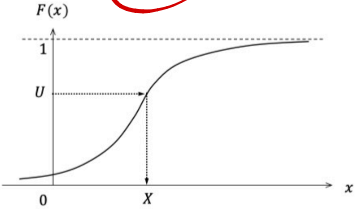
output: Random variable X distributed according to F.

Generate U from $\mathrm{Unif}(0,1)$.

$X \leftarrow F^{-1}(U)$

return X

F^{-1}



Inverse Transform: Discrete

Algorithm Inverse-Transform Method: PMF Case

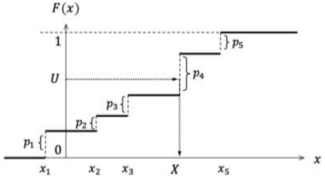
input: Discrete cumulative distribution function F with monotonic sequence $\{x_j\}$

output: Discrete random variable X distributed according to F.

1: Generate $U \sim \text{Unif}(0, 1)$

2: Find the smallest positive integer, k, such that $U \leq F(x_k)$. Let $X \leftarrow x_k$.

3: **return** X



- $U \sim \text{Unif}(0, 1)$:

$$X = \begin{cases} x_1 & \text{if } 0 < U \leq p_1 \\ x_2 & \text{if } p_1 < U \leq p_1 + p_2 \\ x_3 & \text{if } p_1 + p_2 < U \leq p_1 + p_2 + p_3 \\ x_4 & \text{if } p_1 + p_2 + p_3 < U \leq p_1 + p_2 + p_3 + p_4 \\ x_5 & \text{if } p_1 + p_2 + p_3 + p_4 < U \leq 1 \end{cases}$$

Acceptance-Rejection

- Suppose one can generate samples (relatively easily) from PDF g
- How can random samples be simulated from PDF f?

Algorithm Acceptance-Rejection Algorithm

Let c denote a constant such that $c \geq \sup_y \frac{f(y)}{g(y)}$. Then:

Step 1: Generate $Y \sim g$.
Step 2: Generate $U \sim \text{Unif}(0, 1)$.
Step 3: If $\boxed{U \leq \frac{f(Y)}{c \cdot g(Y)}}$ set $X = Y$. Otherwise go back to step 1.

Theorem

(i) The random variable generated by the Acceptance-Rejection method has the desired PDF f.
(ii) The number of iterations of the algorithm that are needed is a first-success random variable with mean $\boxed{c}$.
(iii) $c \geq 1$

Monte Carlo Integration

- We can use the sample mean to approximate the expectation:

$$E[g(X)] \approx \frac{1}{n}\sum_{i=1}^{n} g(X_i).$$

sample average.

- Now we have integration

$$\int_a^b g(x)\,dx = (b-a)\int_a^b g(x)\cdot\frac{1}{b-a}\,dx.$$

- Drawing n samples (empirical samples) from Unif(a, b):

$$X_1, X_2, \ldots, X_n \sim \boxed{\text{Unif}(a, b).}$$

Unif(0,1)

- Monte Carlo Integration:

$$\boxed{\int_a^b g(x)\,dx \approx \frac{1}{n}\sum_{i=1}^{n} g(X_i)(b-a).}$$

$n \to \infty$

$10^6, 10^7$

- Indicator: bridge between expectation and probability
- Given event A:

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{Otherwise} \end{cases}.$$

- For random variable X:

$$\begin{aligned} P(X \in A) &= 1 \cdot P(X \in A) + 0 \cdot P(X \notin A) \\ &= E(I_A(X)) \\ &\approx \boxed{\frac{1}{n} \sum_{i=1}^{n} I_A(X_i).} \end{aligned}$$

Importance Sampling

$$H = E_f[h(Y)] = \int h(y)f(y)dy$$

- h is some function and f is the PDF of random variable Y
- When the PDF f is difficult to sample from, importance sampling can be used
- Rather than sampling from f, you specify a different PDF g, as the proposal distribution.

$$H = \int h(y)f(y)dy = \int h(y)\frac{f(y)}{g(y)}g(y)dy = \int \frac{h(y)f(y)}{g(y)}g(y)dy$$

$$H = E_f[h(Y)] = \int \frac{h(y)f(y)}{g(y)}g(y)dy = E_g\left[\frac{h(Y)f(Y)}{g(Y)}\right]$$

- Hence, given an iid sample $Y_1, \ldots, Y_n$ from PDF g, our estimator of H becomes

$$\hat{H} = \frac{1}{n}\sum_{j=1}^{n}\frac{h(Y_j)f(Y_j)}{g(Y_j)}$$

$P(Y>8)$

$Y \sim N(0,1)$

Law of Large Number

SLLN

Theorem

The sample mean $\bar{X}_n$ converges to the true mean μ pointwise as $n \to \infty$, with probability 1. In other words, the event $\bar{X}_n \to \mu$ has probability 1.

WLLN

Theorem

For all $\epsilon > 0$, $P\left(\left|\bar{X}_n - \mu\right| > \epsilon\right) \to 0$ as $n \to \infty$. (This form of convergence is called convergence in probability).

Cauchy-Schwarz Inequality

Theorem

For any r.v.s X and Y with finite variances,

$$|E(XY)| \leq \sqrt{E(X^2)\,E(Y^2)}.$$

Expectation.

Jensen's Inequality

If f is a convex function, $0 \le \lambda_1, \lambda_2 \le 1$, $\lambda_1 + \lambda_2 = 1$, then for any x_1, x_2,

$$f(\lambda_1 x_1 + \lambda_2 x_2) \le \lambda_1 f(x_1) + \lambda_2 f(x_2).$$

Theorem

Let X be a random variable. If g is a convex function, then $E(g(X)) \ge g(E(X))$. If g is a concave function, then $E(g(X)) \le g(E(X))$. In both cases, the only way that equality can hold is if there are constants a and b such that $g(X) = a + bX$ with probability 1.

$E \Rightarrow \sum \ldots \quad \int$

Concentration Inequalities

Theorem

For any r.v. X and constant $a > 0$,

$$P(|X| \geq a) \leq \frac{E|X|}{a}.$$

x

Theorem

Let X have mean μ and variance σ^2. Then for any $a > 0$,

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Theorem

For any r.v. X and constants $a > 0$ and $t > 0$,

$$P(X \geq a) \leq \frac{E\left(e^{tX}\right)}{e^{ta}}.$$

Hoeffding Inequality

Theorem

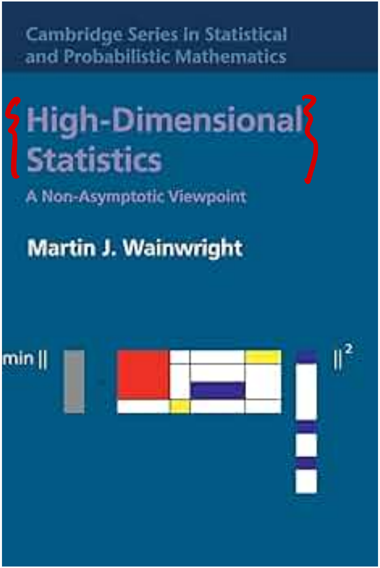
Let the random variables $X_1, X_2, \ldots, X_n$ be independent with $E(X_i) = \mu$, $a \leq X_i \leq b$ for each $i = 1, \ldots, n$, where a, b are constants. Then for any $\epsilon \geq 0$,

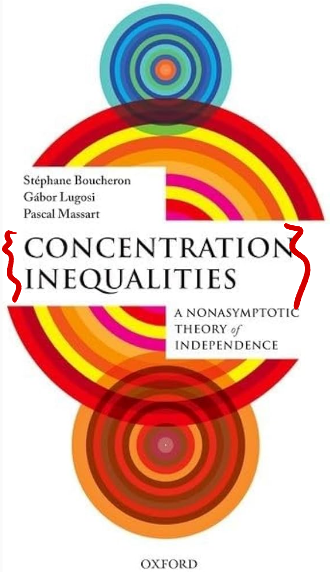
$$\mathbb{P}(|\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\text{linear}} - \mu| \geq \epsilon) \leq \underbrace{\boxed{2e^{-\frac{2n\epsilon^2}{(b-a)^2}}}}_{\text{exponential}}.$$

linear

exponential.

More Concentration Inequality





Outline