Lecture 7: Monte Carlo Methods

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

April 30, 2024

Outline

1. History of Monte Carlo
2. Sampling: Random Variable Generation
3. Sampling: Random Vector Generation
4. Monte Carlo Integration
5. Asymptotic Analysis: Law of Large Numbers
6. Non-asymptotic Analysis: Inequalities

Outline

Motivation I

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.

Motivation II

Probability

Math

Statistics

Science

Monte Carlo

Computing

Monte Carlo Methods

- One of the top ten algorithms for science and engineering in 20th century
- Monte Carlo Methods, Simplex Method, Fast Fourier Transform, Quicksort, QR Algorithm...

Widely Applications

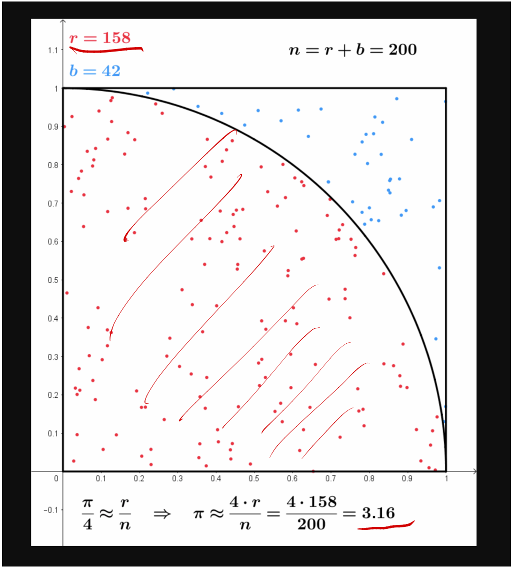
Monte Carlo methods have been used in various tasks, including

- Sampling from the underlying probability distribution $f(x)$ and simulating a random system
- Sampling from posterior distribution for bayesian inference
- Estimation through numerical integration

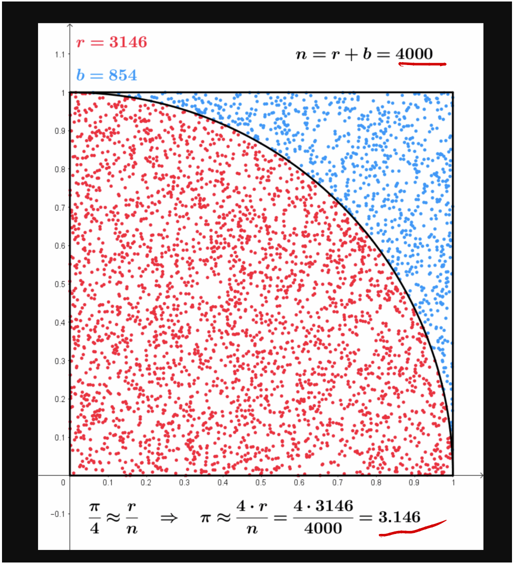
$$c = E_{\pi}(h(x)) = \int f(x)h(x)dx.$$

- Optimizing a target function to find its maxima or minima

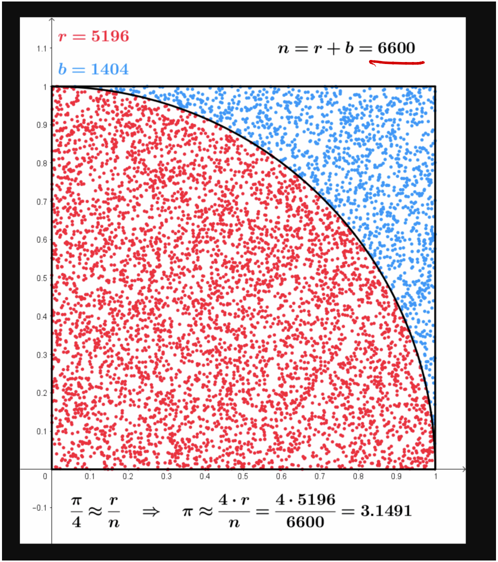
Classical Example: Estimation of π



Classical Example: Estimation of π



Classical Example: Estimation of π



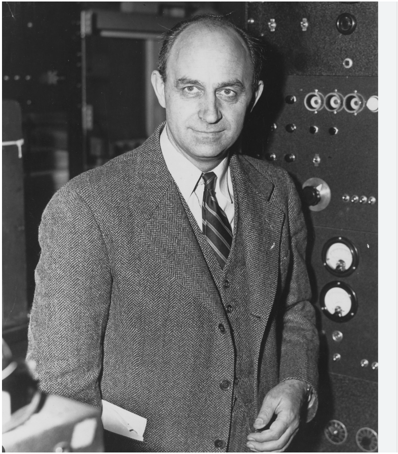
History

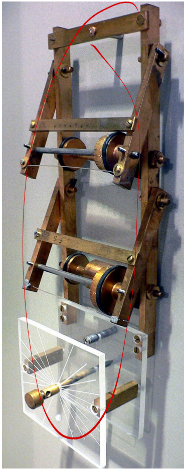
Monte Carlo Methods

- Basic Monte Carlo methods: formally proposed by Stanislaw Ulam & John Von Neumann in 1940s at Los Alamos National Lab (Named after a casino in Monaco)

Monte Carlo Trolley

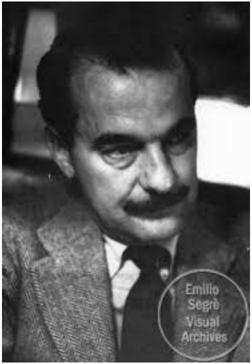
- Analog computer invented by Enrico Fermi in 1946





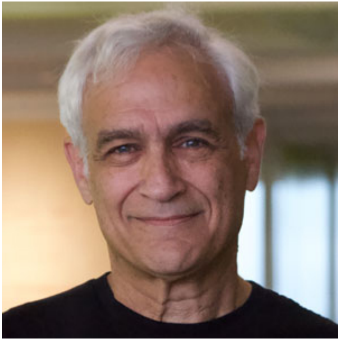
Markov Chain Monte Carlo Methods

- Metropolis-Hastings Algorithm: formally proposed by Nicholas Metropolis et al in 1950s at Los Alamos National Lab, then extended in 1970 by Wilfred Keith Hastings



Markov Chain Monte Carlo Methods

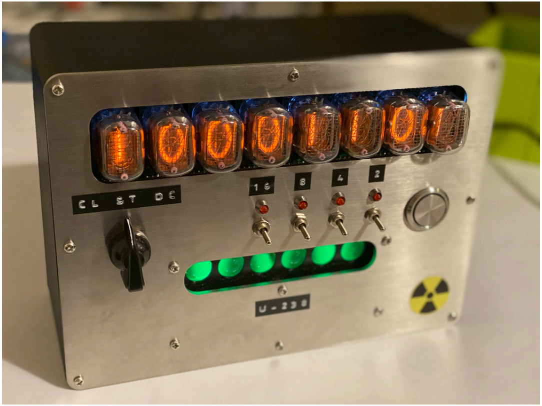
- Gibbs Sampling Algorithm: proposed in 1984 by brothers Stuart Geman (1949-) and Donald Geman (1943-).
- Gibbs sampling is named after the physicist Josiah Willard Gibbs (1839-1903), in reference to an analogy between the sampling algorithm and statistical physics.



Outline

Randomness Generation

- Earlier days: manual techniques including coin flipping, dice rolling, card shuffling, and roulette spinning
- Early days: physical devices including noise diodes and Geiger counters (`https://github.com/nategri/chernobyl_dice`)



Randomness Generation

- The prevailing belief: only mechanical or electronic devices could produce truly random sequences
- The book: *A Million Random Digits With 100,000 Normal Deviates* (based on Uranium radiation) RAND
- Current days: computer simulation with deterministic algorithms, also called pseudorandom number generator

Unif(0,1)

Sampling

- Assuming an algorithm is available for generating Unif(0, 1) random numbers
- Two elementary methods for generating random variables (or samples)
  - Inverse-transform method: operates on the CDF
  - The acceptance-rejection method: operates on the PDF (or PMF)

Inverse Transform Method

- 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

Inverse Transform Method: Recall

Theorem

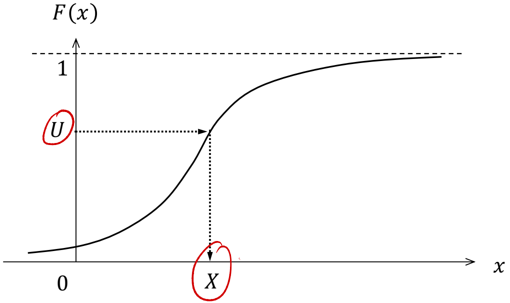
Let F be a CDF which is a continuous function and strictly increasing on the support of the distribution. This ensures that the inverse function F^{-1} exists, as a function from $(0,1)$ to $\mathbb{R}$. We then have the following results.

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 $F(X) \sim \mathrm{Unif}(0,1)$.*

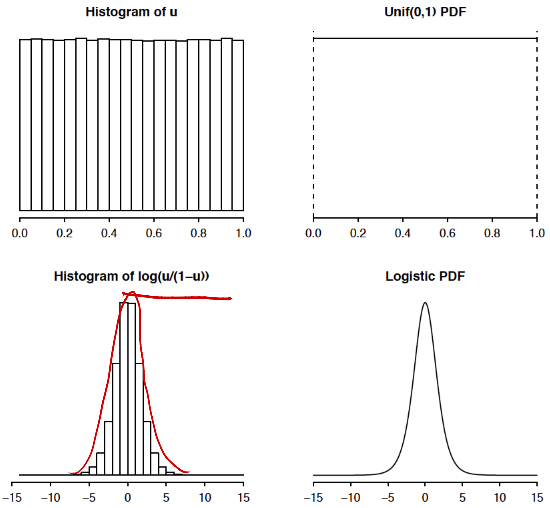
Algorithm Inverse-Transform Method: PDF Case

input: Cumulative distribution function F.
output: Random variable X distributed according to F.

1: Generate U from $\mathrm{Unif}(0, 1)$.
2: $X \leftarrow F^{-1}(U)$
3: **return** X



Histogram & PDF: Example



Box-Muller Method: Recall

1°. $U \sim Unif(0, 2\pi) = 2\pi\, Unif(0,1)$

$U_2 \sim Unif(0,1)$; $U \stackrel{d}{=} 2\pi U_2$

Let $U \sim \text{Unif}(0, 2\pi)$, and let $T \sim \text{Expo}(1)$ be independent of U. Define $X = \sqrt{2T}\cos U$ and $Y = \sqrt{2T}\sin U$. Then X and Y are independent, and their marginal distributions are standard normal distribution. 2°. $F_T(t) = 1 - e^{-t}, t > 0$ $\Rightarrow F_T^{-1}(u) = -\ln(1-u)$, $0<u<1$

Algorithm Normal Random Variable Generation: Box-Muller Approach

$U_1 \sim Unif(0,1)$; $-\ln(1-U_1) \sim Expo(1)$

---

output: Independent standard normal random variables X and Y.

1: Generate two independent random variables, U_1 and U_2, from *Unif*$(0, 1)$.

2: $X \leftarrow (-2\ln U_1)^{1/2}\cos(2\pi U_2)$ $\quad \sqrt{-2\ln(1-U_1)}$

3: $Y \leftarrow (-2\ln U_1)^{1/2}\sin(2\pi U_2)$

4: **return** X, Y

$U_1 \sim 1 - U_1$

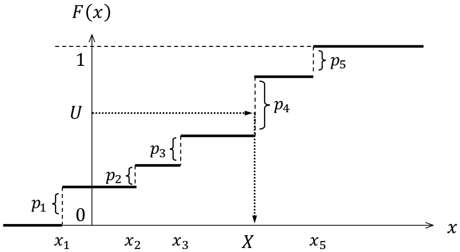
$T := -\ln U_1$

Inverse Transform Method for Discrete Distribution

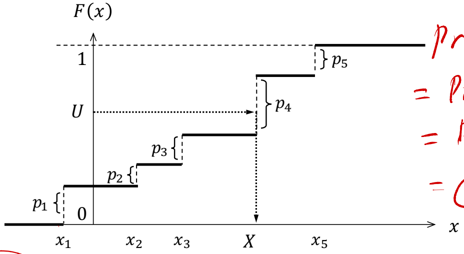
- Example: PMF with $P(X = x_j) = p_j$, $j = 1, 2, 3, 4, 5$, $x_1 < x_2 < x_3 < x_4 < x_5$.
- CDF with

$$F(x_k) = P(X \leq x_k) = \sum_{j=1}^{k} P(X = x_j) = \sum_{j=1}^{k} p_j, k = 1, 2, 3, 4, 5.$$

$p_1 + p_2 + p_3 + p_4 + p_5 = 1$



Inverse Transform Method for Discrete Distribution



$Pr(p_1 < U \le p_1 + p_2)$
$= Pr(U \le p_1 + p_2) - Pr(U \le p_1)$
$= p_1 + p_2 - p_1$
$= p_2$

$Pr(0 < U \le p_1) = Pr(U \le p_1)$
$= p_1 = p(X = x_1)$

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

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

Inverse Transform Method for Discrete Distribution

Generalization of Inverse Transform

$F^{-1}(u) = \inf\{x : F(x) \geq u\}$

CDF

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

$k-1$ $\quad U > F(x_{k-1})$

$Pr(F(x_{k-1}) < U \leq F(x_k))$

$= Pr(U \leq F(x_k)) - Pr(U \leq F(x_{k-1}))$

$= F(x_k) - F(x_{k-1}) = \sum_{j=1}^{k} p_j - \sum_{j=1}^{k-1} p_j = p_k$

Bernoulli Distribution

$X_1 = 0$, $X_2 = 1$

- Bernoulli distribution Bern(p) with PMF:

CDF $F(X_1) = 1-p$; $F(X_2) = 1$;

$$P(X = 1) = p, P(X = 0) = 1 - p, 0 < p < 1.$$

Algorithm Inverse-Transform Method: PMF Case

---

input: $p \in (0, 1)$
output: Discrete random variable $X \sim \text{Bern}(p)$
1: Generate $U \sim \text{Unif}(0, 1)$.
2: If $U \leq 1 - p$, then $X \leftarrow 0$
3: Else $X \leftarrow 1$
4: **return** X

Inverse Transform Method for Discrete Distribution

Algorithm Inverse-Transform Method: PMF Case

input: PMF $\{p_j\}$ for distribution with non-monotonic sequence $\{x_j\}$

output: Discrete random variable X distributed according to PMF $\{p_j\}$.

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

2: Find the positive integer, k, such that

$$\sum_{j=1}^{k-1} p_j < U \le \sum_{j=1}^{k} p_j.$$

p_k

3: **return** $X = x_k$

Bernoulli Distribution

$X_1 = 1$, $X_2 = 0$

- Bernoulli distribution Bern(p) with PMF:

$$P(X = 1) = p, P(X = 0) = 1 - p, 0 < p < 1.$$

Algorithm Inverse-Transform Method: PMF Case

input: $p \in (0, 1)$

output: Discrete random variable $X \sim \text{Bern}(p)$

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

2: If $U \leq p$, then $X \leftarrow 1$

3: Else $X \leftarrow 0$

4: **return** X

Acceptance-Rejection Method

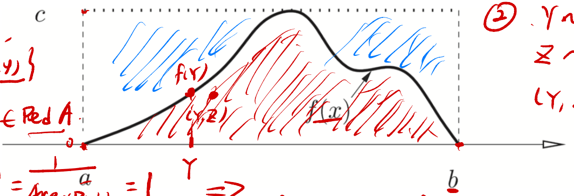
① objective PDF f: $x \in [a,b]$, $c \geq \sup_x f(x)$

$\int_a^b f(x)dx = 1$; Area(Red) = 1

② $Y \sim \text{Unif}(a,b)$ independent
$Z \sim \text{Unif}(0,c)$
(Y, Z) unif over the Rectangle

③ Acceptance-Rejection: if $Z \leq f(Y)$; Accept (Y, Z).

Red $A = \{(y,z) : a \leq y \leq b,\ 0 \leq z \leq f(y)\}$



④ $(Y^*, Z^*) \in$ Red A

$f_{Y^*,Z^*}(y,z) = \frac{1}{\text{Area(Red)}} = 1 \Rightarrow$

$f_{Y^*}(y) = \int_0^{f(y)} f_{Y^*,Z^*}(y,z)\,dz = \int_0^{f(y)} 1 \cdot dz = f(y)$

$\Rightarrow Y^* \sim f$

Algorithm Acceptance-Rejection Algorithm

Step 1: Generate $Y \sim \text{Unif}(a, b)$.
Step 2: Generate $Z \sim \text{Unif}(0, c)$.
Step 3: If $Z \leq f(Y)$, set $X = Y$. Otherwise go back to step 1.

Acceptance-Rejection Method

① $X \in [a,b]$, PDF f: desired

② $Y \sim g$; $Z \sim \text{Unif}(0, c \cdot g(Y))$.

$\Rightarrow f_{Y,Z}(y,z) = f_Y(y) \cdot f_{Z|Y}(z|y)$

$= g(y) \cdot \frac{1}{c \cdot g(y)} = \frac{1}{c}$

$= \frac{1}{\text{Area(Triangle)}}$.

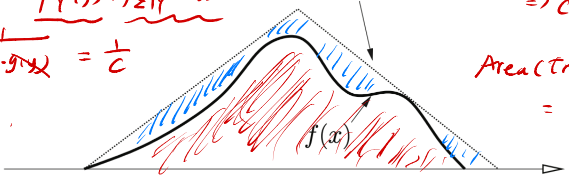
$\Rightarrow (Y, Z) \sim \text{Unif (Triangle)}$

④: $\phi(x) = c\, g(x) \geq f(x)$.

$\Rightarrow c \geq \frac{f(x)}{g(x)}$, $\forall x \in [a,b]$

$\Rightarrow c = \sup_x \frac{f(x)}{g(x)}$.

PDF $[a,b]$



③ $(Y^*, Z^*) \in$ Red A. $\sim \text{Unif}(A)$.

Area(Triangle)

$= \int_a^b \phi(x)\,dx$

$= \int_a^b c \cdot g(x)\,dx$

$= c \int_a^b g(x)\,dx$

$= c$

Algorithm Acceptance-Rejection Algorithm

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

$Z | Y{=}y \sim \text{Unif}(0, c \cdot g(y))$

$\text{Unif}(0, c \cdot g(Y)) \leq f(Y) \Leftrightarrow c \cdot g(Y) \cdot \text{Unif}(0,1) \leq f(Y)$

$\Leftrightarrow \text{Unif}(0,1) \leq \frac{f(Y)}{c \cdot g(Y)}$

Acceptance-Rejection Method

$f \sim g$
Supporting set

- 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 $U \leq \frac{f(Y)}{c \cdot g(Y)}$, set $X = Y$. Otherwise go back to step 1.

Acceptance-Rejection Method

(ii)

$Y \sim g \xrightarrow{A} X$, $\bar{A}$

\# of iterations

$N \sim FS(p)$

$p = P(A) = \frac{1}{c}$

Theorem

$\Rightarrow E(N) = \frac{1}{p} = c$

(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 c.
(iii) $c \geq 1$ ✓

Proof

(i'). event A = " $U \le \frac{f(Y)}{c\cdot g(Y)}$ ", $\underline{f_Y(y|A)} \overset{\downarrow \text{desired}}{=} \underline{f(y)}$ PDF

$$f_Y(y|A) = \frac{P(A|Y=y)}{P(A)} f_Y(y).$$

$U \sim \text{unif}(0,1)$

1°. $P(A|Y=y) = P\left(U \le \frac{f(Y)}{c\cdot g(Y)} \middle| Y=y\right) = P\left(U \le \frac{f(y)}{c\cdot g(y)} \middle| Y=y\right)$

$Y \sim g$

$$= P\left(U \le \frac{f(y)}{c\cdot g(y)}\right) = \frac{f(y)}{c\cdot g(y)}$$

$\left[c \ge \sup_y \frac{f(y)}{g(y)}\right]$

$\Rightarrow \frac{f(y)}{c\cdot g(y)} \le 1$

2°. $\underline{P(A)} \overset{\text{LOTP}}{=} \int \underline{P(A|Y=y)} \cdot g(y)\, dy = \int \frac{f(y)}{c\cdot g(y)} g(y)\, dy$

$$= \frac{1}{c}\int f(y)\, dy = \underline{\frac{1}{c}} \le 1 \quad \Rightarrow \underline{c \ge 1}$$

$$\Rightarrow f_Y(y|A) = \frac{P(A|Y=y)}{P(A)} \underline{f_Y(y)} = \frac{\frac{f(y)}{\cancel{c}\cdot \cancel{g(y)}}}{\cancel{\frac{1}{c}}} \cdot \cancel{g(y)} = f(y).$$

$$\Rightarrow X \sim f.$$

Proof

Example: Beta Distribution

$X \sim Beta(2,4)$, $f(x) = 20x(1-x)^3$, $0<x<1$

- An r.v. X is said to have the *Beta distribution* with parameters a and b, $a > 0$ and $b > 0$, if its PDF is

$$f(x) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}, \ 0 < x < 1,$$

  where the constant $\beta(a,b)$ is chosen to make the PDF integrate to 1. We write this as $X \sim \mathrm{Beta}(a,b)$. $a=b=1$; $f(x) \propto$ constant.
- Beta distribution is a generalization of uniform distribution.
- Use the Acceptance-Rejection Method to generate a random variable with distribution $Beta(2,4)$

Solution

$f(x) = 20x(1-x)^3, \ 0<x<1$

① g : Unif(0,1), $g(x) = 1, \ 0<x<1$.

$$c \geq \sup_y \frac{f(y)}{g(y)} = \sup_{y\in(0,1)} \frac{20y(1-y)^3}{1} = \sup_{y\in(0,1)} 20y(1-y)^3. \quad \Rightarrow y^* = \frac{1}{4}.$$

$\Rightarrow c \geq \frac{135}{64}$. choose $c = \frac{135}{64}$.

② $\Rightarrow 0<y<1, \ \frac{f(y)}{c\cdot g(y)} = \frac{20y(1-y)^3}{\frac{135}{64}\cdot 1} = \frac{256}{27}y(1-y)^3.$

Step 1 : Generate $Y \sim \text{Unif}(0,1)$. (g)

Step 2 : Generate $U \sim \text{unif}(0,1)$

Step 3 : If $U \leq \frac{f(Y)}{c\cdot g(Y)} = \frac{256}{27}Y(1-Y)^3$, set $X = Y$.

otherwise reject Y, Go back to Step 1.

Solution

Example: Normal Distribution

① $Z \sim N(0,1)$. $(-\infty, +\infty)$

$X = |Z|$ $(0, +\infty)$

C $P(X \le x) = P(|Z| \le x) = 2P(0 \le Z \le x) = 2\int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$

$= \int_0^x \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}z^2} dz \Rightarrow f_X(x) = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}x^2}, \ 0 < x < \infty$.

② . choose $g \sim \mathrm{Expo}(1)$. $g(x) = e^{-x}$, $0 < x < \infty$.

- Use the Acceptance-Rejection Method to generate a random variable with distribution $N(0,1)$

$c \ge \sup_y \frac{f(y)}{g(y)} = \sup_y \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}y^2 + y} = \sup_y \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}(y-1)^2 + \frac{1}{2}} = \sqrt{\frac{2e}{\pi}}$.

$\Rightarrow y^* = 1$, choose $c = \sqrt{\frac{2e}{\pi}}$.

$\Rightarrow \frac{f(y)}{c \cdot g(y)} = e^{\{y - \frac{1}{2}y^2 - \frac{1}{2}\}} = e^{-\frac{1}{2}(y-1)^2}$

Solution ③

Step 1 : Generate $Y \sim \text{Expo}(1)$

2 : ⋯ $U \sim \text{unif}(0,1)$

3 : If $U \le e^{-\frac{1}{2}(Y-1)^2}$, set $X = Y$.

otherwise return to Step 1.

$X = |Z|$

Step 4 : generate $U' \sim \text{unif}(0,1)$

$$Z = \begin{cases} X & \text{if } U' \le \frac{1}{2} \\ -X & \text{otherwise.} \end{cases}$$

$$\underline{Z \sim N(0,1)}$$

Box-Muller ;

Acceptance-Rejection ;

Solution

Outline

Change of Variables

Theorem

Let $\mathbf{X} = (X_1, ..., X_n)$ *be a continuous random vector with joint PDF* $f_{\mathbf{X}}(x)$*, and let* $\mathbf{Y} = g(\mathbf{X})$ *where* g *is an invertible function from* $\mathbb{R}^n$ *to* $\mathbb{R}^n$*. Let* $y = g(\mathbf{x})$ *and suppose that all the partial derivatives* $\frac{\partial x_i}{\partial y_j}$ *exists and are continuous, so we can form the* **Jacobian matrix**

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}$$

Also assume that the determinant of the Jacobian matrix is never 0. Then the joint PDF of $\mathbf{Y}$ *is*

$$f_{\mathbf{Y}}(y) = f_{\mathbf{X}}(x)|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}|$$

Example: Generate Uniform Distribution over An Ellipse

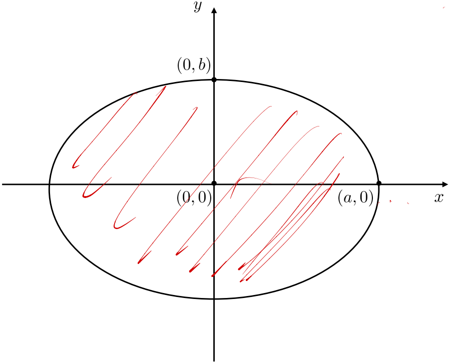
objective PDF $f_{X,Y}(x,y) = \frac{1}{\pi \cdot a \cdot b}$ $\theta\ (x,y) \in E_2(a,b)$

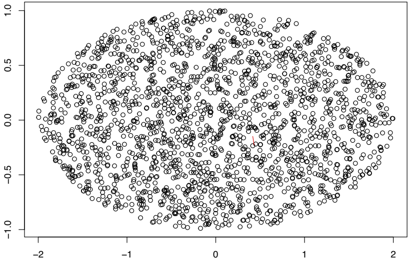
- Ellipse:

$x = \rho \cdot a \cdot \cos\theta \qquad \rho \in [0,1] \ ; \ \theta \in [0, 2\pi]$

$y = \rho \cdot b \cdot \sin\theta$

$$E_2(a,b) = \left\{ (x,y) \in \mathbb{R}^2 : \left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 \leq 1 \right\}.$$





Solution

① Jacobi Matrix $\begin{bmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} a\cos\theta & -\rho a\sin\theta \\ b\sin\theta & \rho b\cos\theta \end{bmatrix}$

$J = \det(\quad) = \rho ab.$

$\Rightarrow f_{R,\Theta}(\rho,\theta) = f_{X,Y}(x,y)\cdot|J| = \frac{1}{\pi ab}\rho ab = \frac{\rho}{\pi}, \quad \rho\in[0,1], \quad \theta\in[0,2\pi].$

$\Rightarrow f_R(\rho) = \int_0^{2\pi}\frac{\rho}{\pi}\,d\theta = 2\rho, \quad 0\le\rho\le 1 \quad \Rightarrow \underset{\text{CDF}}{F_R(\rho)} = \rho^2, \quad 0\le\rho\le 1.$

$f_\Theta(\theta) = \int_0^1\frac{\rho}{\pi}\,d\rho = \frac{1}{2\pi}, \quad 0\le\theta\le 2\pi. \quad \Theta \sim \mathrm{Unif}(0,2\pi)$

$\Rightarrow f_{R,\Theta}(\rho,\theta) = f_R(\theta)\cdot f_\Theta(\theta), \quad R, \Theta$ independent.

R, $F_R^{-1}(z) = \sqrt{z}, \quad 0\le z\le 1$;

Solution ②

U_1, U_2 independent Unif(0,1)

$X = \rho a \cos\theta$

$Y = \rho b \sin\theta$

$$X \leftarrow a\sqrt{U_1}\cos(2\pi U_2)$$

$$Y \leftarrow b\sqrt{U_1}\sin(2\pi U_2)$$

(X, Y)

Solution

Change of Variables

A : $n \times n$

$\det(A^T A) = \det^2(A)$

Theorem

Let $\mathbf{X} = (X_1, ..., X_n)$ be a continuous random vector with joint PDF $f_{\mathbf{X}}(x)$, and let $\mathbf{Y} = g(\mathbf{X})$ where g is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Let $y = g(\mathbf{x})$ and we have the **Jacobian matrix** $\frac{\partial \mathbf{x}}{\partial \mathbf{y}}$. *The corresponding Gram matrix is*

$$\mathbf{G} = \left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right)^T \frac{\partial \mathbf{x}}{\partial \mathbf{y}}.$$

Then the joint PDF of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(y) = f_{\mathbf{X}}(x)\sqrt{det(\mathbf{G})}$$

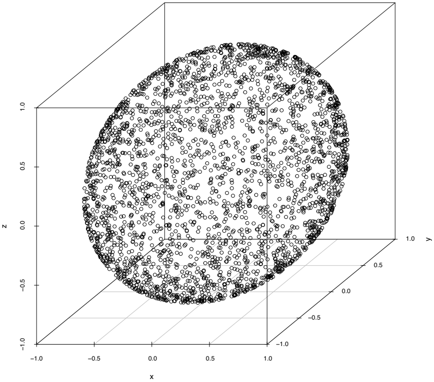
Example: Generate Uniform Distribution over A Sphere

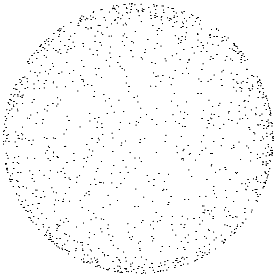
Ball $B_3(r) = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 \le r^2\}$

- Sphere:

$\frac{4}{3}\pi r^3$

$$S_2(r) = \left\{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = r^2\right\}. \quad 4\pi r^2$$





Solution

① $f_{X,Y,Z}(x,y,z) = \frac{1}{4\pi r^2} \quad \forall (x,y,z) \in S_2(r).$

$$\begin{cases} x = r\sin\theta\cos\phi \\ y = r\sin\theta\sin\phi \\ z = r\cos\theta \end{cases} \qquad \begin{aligned} &\theta : \theta \in [0,\pi]; \\ &\phi : \phi \in [0,2\pi]; \end{aligned} \qquad (x,y,z) \to (\theta,\phi)$$

Jacobi Matrix $M = \begin{bmatrix} \frac{\partial x}{\partial\theta} & \frac{\partial x}{\partial\phi} \\ \frac{\partial y}{\partial\theta} & \frac{\partial y}{\partial\phi} \\ \frac{\partial z}{\partial\theta} & \frac{\partial z}{\partial\phi} \end{bmatrix} = \begin{bmatrix} r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ r\cos\theta\sin\phi & r\sin\theta\cos\phi \\ -r\sin\theta & 0 \end{bmatrix}$

Gram matrix $G = M^T M = \begin{bmatrix} r^2 & 0 \\ 0 & r^2\sin^2\theta \end{bmatrix} \Rightarrow \det(G) = r^4\sin^2\theta.$

$$\Rightarrow f_{\Theta,\Phi}(\theta,\phi) = f_{X,Y,Z}(x,y,z)\cdot\sqrt{\det(G)} = \frac{1}{4\pi r^2}\cdot r^2\sin\theta = \frac{1}{4\pi}\sin\theta$$

$\theta \in [0,\pi]$

$\phi \in [0,2\pi]$

Solution

② $f_\Theta(\theta) = \int_0^{2\pi} f_{\Theta,\Phi}(\theta,\phi)\,d\phi = \frac{1}{2}\sin\theta \;;\; 0 \le \theta \le \pi.$

$\underline{F_\Theta(\theta)} = \int_0^\theta f_\Theta(s)\,ds = \underline{\frac{1-\cos\theta}{2}} \,,\; 0 \le \theta \le \pi.$

$f_\Phi(\phi) = \frac{1}{2\pi} \;;\; 0 \le \phi \le 2\pi \;;\; \Rightarrow \underline{f_{\Theta,\Phi}(\theta,\phi)} = \underline{f_\Theta(\theta)} \cdot \underline{f_\Phi(\phi)}$

Θ and ϕ independent. ; $\phi \sim \text{Unif}(0,2\pi) = 2\pi\,\text{Unif}(0,1)$;

$$\underline{F^{-1}_\Theta(s) = \arccos(1-2s) \,,\; 0 \le s \le 1}$$

Note : if $\theta = \arccos(1-2s) \Rightarrow \underline{\cos}\theta = 1-2s$

$\sin^2\theta = 1-\cos^2\theta = 1-(1-2s)^2 = 4s(1-s)$

$\Rightarrow \sin\theta > 0 \;;\; \Rightarrow \underline{\sin\theta} = \underline{2\sqrt{s(1-s)}}$

Solution ③

independently generate U_1, U_2 Unif(0,1)

$$
\begin{cases}
X \leftarrow r \cdot 2\sqrt{U_1(1-U_1)} \cdot \cos(2\pi U_2) \\
Y \leftarrow r \cdot 2\sqrt{U_1(1-U_1)} \cdot \sin(2\pi U_2) \\
Z \leftarrow r \cdot (1-2U_1)
\end{cases}
$$

Outline

Monte Carlo Integration

$E(X)$ $E[g(X)]$

$X_1, \dots X_n$: $\frac{1}{n}(X_1 + \dots + X_n)$

$g(X_1), \dots g(X_n)$: $\frac{1}{n}(g(X_1) + \dots + g(X_n))$

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

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

$X \sim \mathrm{Unif}(a,b)$

- Now we have integration

$f(x)$. PDF

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

$= (b-a) \int_a^b g(x) f(x) dx$

$= (b-a) \cdot E[g(X)]$

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

$$X_1, X_2, \dots, X_n \sim \mathrm{Unif}(a, b).$$

$\approx (b-a) \cdot \frac{1}{n}(g(X_1) + \dots + g(X_n))$

- Monte Carlo Integration:

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

Monte Carlo Integration

Example: π as An Integration

Evaluate the integration

$$\int_0^1 \frac{4}{1+x^2}dx.$$

- $g(x) = 4/(1+x^2), 0 < x < 1.$ $a=0, b=1$
- $X_1, \ldots, X_n$: samples from Unif$(0,1)$.
- Monte Carlo Integration:

$$\int_0^1 \frac{4}{1+x^2}dx \approx \frac{1}{n}\sum_{i=1}^{n} \frac{4}{1+X_i^2}$$

Example

Evaluate the integration

$$\int_0^4 \sqrt{x+\sqrt{x+\sqrt{x+\sqrt{x}}}}\,dx.$$

- Corresponding

$$g(x) = \sqrt{x+\sqrt{x+\sqrt{x+\sqrt{x}}}}.$$

- $X_1, \ldots, X_n$: samples from Unif$(0, 4)$.

  $a=0,\ b=4$

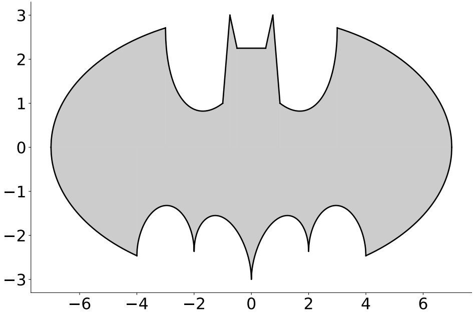
  $b-a=4$

- Monte Carlo Integration:

$$\int_0^4 \sqrt{x+\sqrt{x+\sqrt{x+\sqrt{x}}}}\,dx \approx \frac{4}{n}\sum_{i=1}^{n} \sqrt{X_i+\sqrt{X_i+\sqrt{X_i+\sqrt{X_i}}}}$$

Example: Area of Batman Curve

- Challenging and Fun
- https://mathworld.wolfram.com/BatmanCurve.html



Example: Estimation of Probability

- 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:

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

$X_1, \dots, X_n \sim X$

Example: Estimation of π

generate n points $(x_1, y_1), \dots (x_n, y_n)$

$-1 \le x_i \le 1$

$-1 \le y_i \le 1$

① event A_i : "the ith point lands within the circle", $\Leftrightarrow \{x_i^2 + y_i^2 \le 1\}$.

② $I_{A_i} = Z_i$

$P(Z_i = 1) = P(A_i) = \frac{\pi \cdot 1}{4} = \frac{\pi}{4}$

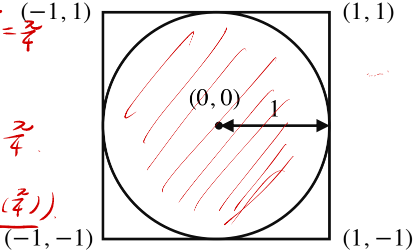
$P(Z_i = 0) = 1 - \frac{\pi}{4}$

$\Rightarrow E(Z_i) = P(Z_i = 1) = \frac{\pi}{4}$.

$Z_1, \dots Z_n \sim Z \ (\mathrm{Bern}(\frac{\pi}{4}))$

$\Rightarrow E(Z) = \frac{\pi}{4}$

$\Rightarrow \pi = 4E(Z) \approx 4 \cdot \frac{1}{n}(Z_1 + \dots + Z_n)$



Example: Estimation of π

Example: Estimation of π

Useful Tools: Importance Sampling

- Standard Monte Carlo integration is great if you can sample from the target distribution (i.e. the desired distribution)
- But what if you can't sample from the target?
- **Importance Sampling**: draw the sample from a proposal distribution and re-weight the integral using importance weights so that the correct distribution is targeted

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.

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

Importance Sampling

g is a PDF.

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

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

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

Example: Gaussian Tail Probability

$P(-3 < Y < 3) = 0.997$

Method 1: $C = P(Y>8) = E[I(Y>8)]$

$f \sim N(0,1)$, $Y_1, \dots, Y_n$

$C \approx 0$

$h(y) = I(y>8) = \begin{cases} 1 & \text{if } y>8 \\ 0 & \text{otherwise} \end{cases}$

$\approx \frac{1}{n} \sum_{j=1}^{n} I(Y_j > 8)$

Evaluate the probability of rare event $c = \mathbb{P}(Y > 8)$, where $Y \sim N(0, 1)$.

choose $g \sim N(8,1)$, $Y_1, \dots Y_n \sim g$.

Method 2: importance sampling

$$C \approx \frac{1}{n} \sum_{j=1}^{n} \frac{h(Y_j) f(Y_j)}{g(Y_j)} = \frac{1}{n} \sum_{j=1}^{n} I(Y_j > 8) \cdot \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Y_j^2}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Y_j-8)^2}}$$

$$= \frac{1}{n} \sum_{j=1}^{n} I(Y_j > 8) \cdot e^{-8Y_j + 32}$$

$n = 50000$, $C \approx 6.25 \times 10^{-16}$

Solution

Outline

Sample Mean: Recall

Definition

Let $X_1, \ldots, X_n$ be i.i.d. random variables with finite mean μ and finite variance σ^2. The *sample mean* $\bar{X}_n$ is defined as follows:

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j .$$

The sample mean $\bar{X}_n$ is itself an r.v. with mean μ and variance σ^2/n.

$n \to \infty$

$\sigma^2/n \to 0$

Strong Law of Large Numbers (SLLN)

$\int_a^b g(x)dx \approx \frac{b-a}{n}\sum_{i=1}^n g(X_i)$ $\Rightarrow$ $\frac{b-a}{n}\sum_{i=1}^n g(X_i) \xrightarrow[n\to\infty]{w.p.1} \int_a^b g(x)dx.$

$X_1 \dots X_n$ i.i.d. unif(a,b).

<1> $X_1 \dots X_n$ i.i.d. g : continuous function.

$g(X_1) \dots g(X_n)$ i.i.d.

$E[g(X_i)] = \int_a^b g(x)\cdot\frac{1}{b-a}dx.$

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.

<2> By SLLN $\frac{g(X_1)+\dots+g(X_n)}{n} \xrightarrow[n\to\infty]{w.p.1} E[g(X)] = \int_a^b g(x)\frac{1}{b-a}dx$

$\Rightarrow \frac{(b-a)}{n}\sum_{i=1}^n g(X_i) \xrightarrow[n\to\infty]{w.p.1} \int_a^b g(x)dx.$

Weak Law of Large Numbers (WLLN)

$X_n \xrightarrow{a.s} 0$

$X_n \xrightarrow{w.p.1} 0$

$P(\lim_{n\to\infty} X_n = 0) = 1$

$X_n \xrightarrow{P} 0.$

$\lim_{n\to\infty} P(|X_n - 0| > \varepsilon) = 0.$

$\forall \varepsilon > 0.$

$X_n = \begin{cases} 1 & w.p.\ \frac{1}{n} \\ 0 & w.p.\ 1-\frac{1}{n}. \end{cases}$

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).

$\lim_{n\to\infty} P(|X_n - 0| > \varepsilon) = P(X_n > \varepsilon)$

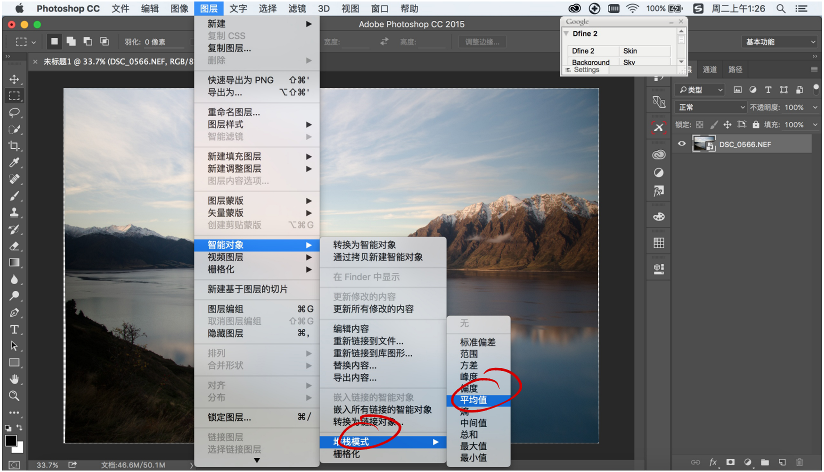
$\varepsilon > 1$: $= 0$

$0 < \varepsilon < 1$: $= P(X_n = 1) = \frac{1}{n} \to 0.$

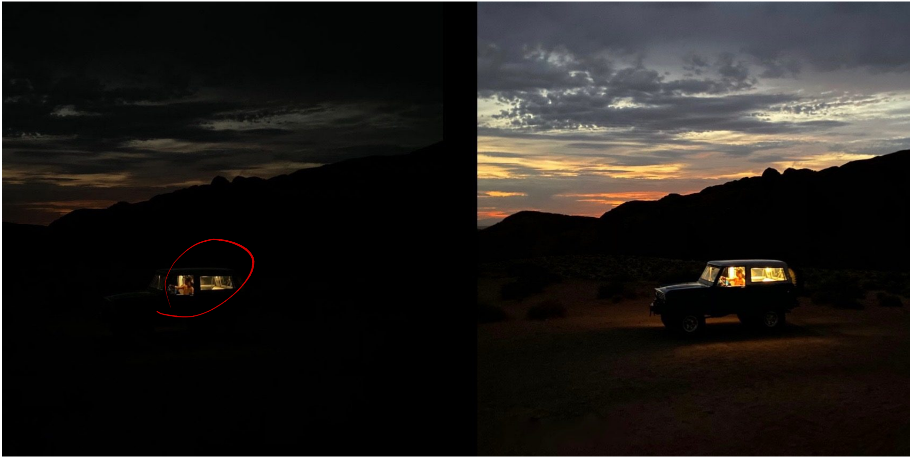
$X_n \xrightarrow{P} 0.$

Widely Applications: Photo Stacking with PC

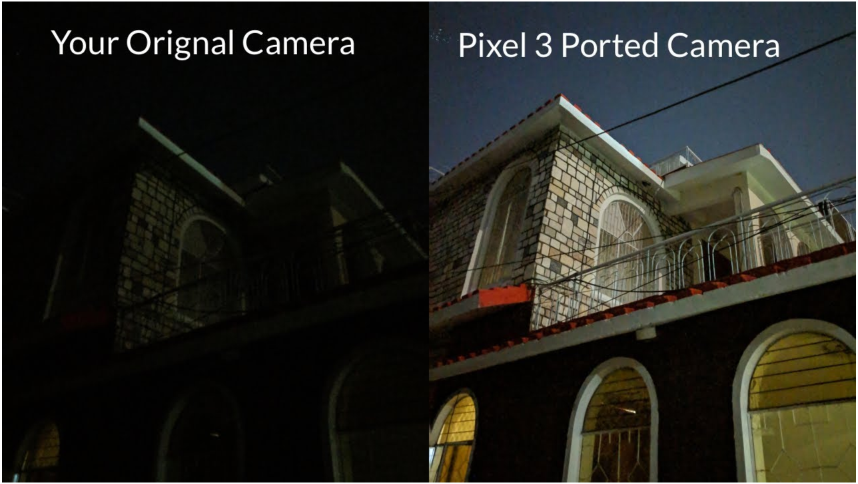
Widely Applications: Photo Stacking with PC



Widely Applications: Night Model with Smart Phone



Widely Applications: Photo Stacking with Smart Phone



Widely Applications: Photo Stacking with Smart Phone

Widely Applications: Photo Stacking with Smart Phone

Outline

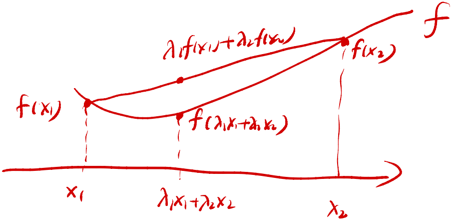
Cauchy-Schwarz Inequality: Recall

Theorem

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

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

Jensen's Inequality



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

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

Jensen's Inequality

Theorem

Let X be a random variable. If g is a convex function, then $E(g(X)) \geq g(E(X))$. If g is a concave function, then $E(g(X)) \leq g(E(X))$. 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.

Quick Examples

g is convex ; $E[g(X)] \geq g[E(X)]$ | $g'' \geq 0$
concave ; $\leq$ | convex.

1°. $g(x) = x^2$, $x \in R$; convex ; $\Rightarrow E[X^2] \geq (E(X))^2$ ✓

$Var(X) = E(X^2) - E^2(X) \geq 0$.

2°. $g(x) = \frac{1}{x}$, $x > 0$; convex ; $\Rightarrow E[\frac{1}{X}] \geq \frac{1}{E(X)}$ ✓

3°. $g(x) = \log x$, $x > 0$; concave ; $\Rightarrow E[\log X] \leq \log(E(X))$.

Entropy

- Let X be a discrete r.v. whose distinct possible values are $a_1, a_2, ..., a_n$, with probabilities $p_1, p_2 ..., p_n$ respectively (so $p_1 + p_2 + \cdots + p_n = 1$).
- The *entropy* of X is defined as follows: $H(X) = \sum_{j=1}^{n} p_j \log_2 (1/p_j)$.
- Using Jensen's inequality, show that the maximum possible entropy for X is when its distribution is uniform over $a_1, a_2, \ldots, a_n$, i.e., $p_j = 1/n$ for all j.
- This makes sense intuitively, since learning the value of X conveys the most information on average when X is equally likely to take any of its values, and the least possible information if X is a constant.

Proof

① Construct a random variable Y. s.t

$$Y = \begin{cases} \frac{1}{p_1} & \text{w.p. } p_1 \\ \frac{1}{p_2} & \text{w.p. } p_2 \\ \vdots & \\ \frac{1}{p_n} & \text{w.p. } p_n \end{cases} \Rightarrow E(Y) = \frac{1}{p_1}\cdot p_1 + \frac{1}{p_2}\cdot p_2 + \cdots + \frac{1}{p_n}\cdot p_n = n$$

② $H(X) \triangleq \sum_{j=1}^{n} p_j \log_2 \frac{1}{p_j} = \underline{E[\log_2 Y]} \leq \log_2 E[Y] = \log_2 n$.

$\forall p_1 \dots p_n \Rightarrow \underline{\max_{p_1 \dots p_n} H(X) \leq \log_2 n}$

③ when $X \sim \text{Dunif}(\frac{1}{n})$, $p_1 = p_2 = p_n = \frac{1}{n}$; $H(X) = \sum_{j=1}^{n} \frac{1}{n} \cdot \log_2 n = \log_2 n$.

$\Rightarrow \underline{\max_{p_1 \dots p_n} H(X) \geq \log_2 n}$ $\Rightarrow \max_{p_1 \dots p_n} H(X) = \log_2 n$

Kullback-Leibler Divergence

Let $\mathbf{p} = (p_1, ..., p_n)$ and $\mathbf{r} = (r_1, ..., r_n)$ be two probability vectors (so each is nonnegative and sums to 1). Think of each as a possible PMF for a random variable whose support consists of n distinct values. The *Kullback-Leibler* divergence between $\mathbf{p}$ and $\mathbf{r}$ is defined as

$$D(\mathbf{p}, \mathbf{r}) = \sum_{j=1}^{n} p_j \log_2 (1/r_j) - \sum_{j=1}^{n} p_j \log_2 (1/p_j).$$

Show that the Kullback-Leibler divergence is nonnegative.

Proof

① $D(p,r) = \sum_{j=1}^{n} p_j \log_2 \frac{1}{r_j} - \sum_{j=1}^{n} p_j \log_2 \frac{1}{p_j} = \sum_{j=1}^{n} p_j \log_2 \frac{p_j}{r_j}$

$$= \underline{- \sum_{j=1}^{n} p_j \log_2 \frac{r_j}{p_j}}$$

② Construct a random variable Y. s.t.

$$P\left(Y = \frac{r_j}{p_j}\right) = p_j, \quad j = 1, 2, \ldots n.$$

$$\Rightarrow \quad E(Y) = \sum_{j=1}^{n} \frac{r_j}{p_j} \cdot p_j = \sum_{j=1}^{n} r_j = \underline{1}$$

③ $D(p,r) = -E[\log_2 Y] \geq -\log_2 E[Y] = -\log_2 1 = 0$

Markov's Inequality

Con centration Inequality

Chebyshev

Markov Lyapunov

$p(|X - E(X)| \geq a)$

a↑

$\frac{1}{a}$

$\frac{1}{a^2}$

$\frac{1}{e^a}$

Theorem

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

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

Proof

$$P(|X| \geq a) \leq \frac{1}{a} E|X| \quad , \quad a > 0$$

① $Y = \frac{1}{a}|X| \geq 0$: $I(Y \geq 1) \leq Y$

	LHS		RHS.
$Y \geq 1$	1	$\leq$ ✓	Y
$0 \leq Y < 1$	0	$\leq$ ✓	Y

$\Rightarrow E[I(Y \geq 1)] \leq E[Y]$

$\Rightarrow P(Y \geq 1) \leq E[Y] = E[\frac{1}{a}|X|] = \frac{1}{a} E[|X|].$

$\Updownarrow$

$P(\frac{|X|}{a} \geq 1)$ $\leq$

$\Updownarrow$

$P(|X| \geq a)$

Chebyshev's Inequality

$$P(|X-\mu| \geq a) = P(|X-\mu|^2 \geq a^2)$$

Markov's Inequality

$$\leq \frac{1}{a^2} E(|X-\mu|^2) = \frac{1}{a^2} Var(X) = \frac{1}{a^2}\sigma^2$$

Theorem

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

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

$\boxed{O(\frac{1}{a^2})}$

Application: $\bar{X}_n$ Sample mean : $X_1, \dots, X_n \overset{iid}{\sim} (\mu, \sigma^2)$ $E(\bar{X}_n) = \mu$; $Var(\bar{X}_n) = \frac{1}{n}\sigma^2$

$$P(|\bar{X}_n - \mu| \geq a) \leq \frac{1}{a^2} Var(\bar{X}_n) = \frac{\sigma^2}{n \cdot a^2} \xrightarrow{n\to\infty} 0$$

$O(\frac{1}{n})$

$$\bar{X}_n \xrightarrow{P} \mu.$$

Proof

Chernoff's Inequality

$\forall t > 0$, $P(X \geq a) = P(e^{tX} \geq e^{ta})$

Markov's Inequality

$\leq \frac{E[e^{tX}]}{e^{ta}}$ $f(t)$

Theorem

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

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

$f(t)$

$\rightarrow$ MGF.

$\forall t > 0$, $P(X \geq a) \leq f(t)$

$\Rightarrow P(X \geq a) \leq \inf_{t>0} f(t)$

Proof

Chernoff's Technique

$\forall t<0$; $P(X \leq a) = P(tX \geq ta)$

$= P(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}} \cdot f(t)$

Theorem

For any r.v. X and constants a,

$$P(X \geq a) \leq \inf_{t>0} \frac{E\left(e^{tX}\right)}{e^{ta}}$$

$$P(X \leq a) \leq \inf_{t<0} \frac{E\left(e^{tX}\right)}{e^{ta}}.$$

Proof

Example: Normal Distribution

① MGF of X : $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

② $P(X>a) \le \inf_{t>0} \frac{E[e^{tX}]}{e^{ta}} = \inf_{t>0} f(t)$

Given $X \sim \mathcal{N}(\mu, \sigma^2)$, for arbitrary constant $a > \mu$, find the Chernoff bound on $P(X > a)$.

$f(t) = \frac{E[e^{tX}]}{e^{ta}} = \frac{M_X(t)}{e^{ta}} = e^{\frac{1}{2}\sigma^2 t^2 + (\mu - a)t}$

$= e^{\frac{1}{2}\sigma^2 [(t + \frac{\mu-a}{\sigma^2})^2 - \frac{(\mu-a)^2}{\sigma^4}]}$ $t^* = \frac{a-\mu}{\sigma^2} > 0$

$\Rightarrow P(X>a) \le f(t^*) = e^{-\frac{(a-\mu)^2}{2\sigma^2}}$

$a = \mu + \varepsilon$ $O(e^{-\varepsilon^2})$

$\Rightarrow P(X > \mu + \varepsilon) \le e^{-\frac{\varepsilon^2}{2\sigma^2}}$ $\Rightarrow P(X - \mu > \varepsilon) \le e^{-\frac{\varepsilon^2}{2\sigma^2}}$

Solution

Hoeffding Bound

$E(\frac{1}{n}\sum_{i=1}^{n} x_i) = \mu$

Theorem

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

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \ge \epsilon\right) \le 2e^{-\frac{2n\epsilon^2}{(b-a)^2}}.$$

$\varepsilon \uparrow \quad \downarrow$

$n \uparrow \quad \downarrow$

$P(|\bar{X}_n - \mu| \geq \varepsilon)$ $\quad o(e^{-\varepsilon^2})$ $\quad o(e^{-n})$

Application: Parameter Estimation

$p \in [\hat{p}-\varepsilon, \hat{p}+\varepsilon] \Leftrightarrow \hat{p}-\varepsilon \le p \le \hat{p}+\varepsilon$

$\Leftrightarrow -\varepsilon \le p-\hat{p} \le \varepsilon$

$\Leftrightarrow -\varepsilon \le \hat{p}-p \le \varepsilon \Leftrightarrow |\hat{p}-p| \le \varepsilon$

Instead of predicting a single value $\hat{p}$ for the parameter p, we given an interval that is likely to contain the parameter:

Definition

$\delta = 0.05$

A $1-\delta$ confidence interval for a parameter p is an interval $[\hat{p}-\epsilon, \hat{p}+\epsilon]$ such that

$$Pr\left(p \in [\hat{p}-\epsilon, \hat{p}+\epsilon]\right) \ge 1-\delta.$$

$$Pr(|\hat{p}-p| \le \varepsilon) \ge 1-\delta$$

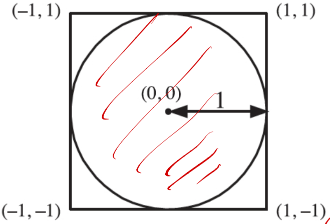
$$\Leftrightarrow Pr(|\hat{p}-p| > \varepsilon) \le \delta$$

Application Example: Monte Carlo Method for Estimation π

① (x_i, y_i) $(-1 \le x_i \le 1,\ -1 \le y_i \le 1)$

Circle: $\{(x, y): x^2 + y^2 \le 1\}$

Z_i: indicator of the event "the ith point landing within the circle".



$P(Z_i = 1) = \frac{\pi}{4}$

$E(Z_i) = \frac{\pi}{4}$.

② $W \triangleq \frac{1}{n}\sum_{i=1}^{n} Z_i$

$E(W) = \frac{\pi}{4}$.

- A point chosen uniformly at random in the square has probability $\pi/4$ of landing in the circle

③. $\hat{\pi} = 4W = 4 \cdot \frac{1}{n}\sum_{i=1}^{n} Z_i$

Confidence interval of π.

Example: Monte Carlo Method for Estimation π

3 $n \to \infty$, $\hat{\pi} \to \pi$ (w.p. 1).

$z_1 \dots z_n$, z. i.i.d. $\text{Bern}(\frac{\pi}{4})$

n is finite

$$Pr(|\hat{\pi} - \pi| \geq \varepsilon) = Pr(|4W - \pi| \geq \varepsilon)$$

$$= Pr\left(\left|W - \frac{\pi}{4}\right| \geq \frac{\varepsilon}{4}\right) = Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} z_i - E(z)\right| \geq \frac{\varepsilon}{4}\right)$$

~~1° Chebyshev's inequality~~ $\cancel{\leq \frac{Var(\frac{1}{n}\sum_{i=1}^{n} z_i)}{(\frac{\varepsilon}{4})^2} = \frac{\frac{\pi}{4}(1-\frac{\pi}{4})}{(\frac{\varepsilon}{4})^2 \cdot n}}$

Hoeffding's inequality bound

$$\leq 2 \cdot e^{-\frac{2n(\frac{\varepsilon}{4})^2}{(1-0)^2}} = 2e^{-\frac{1}{8}n\varepsilon^2} = \delta$$

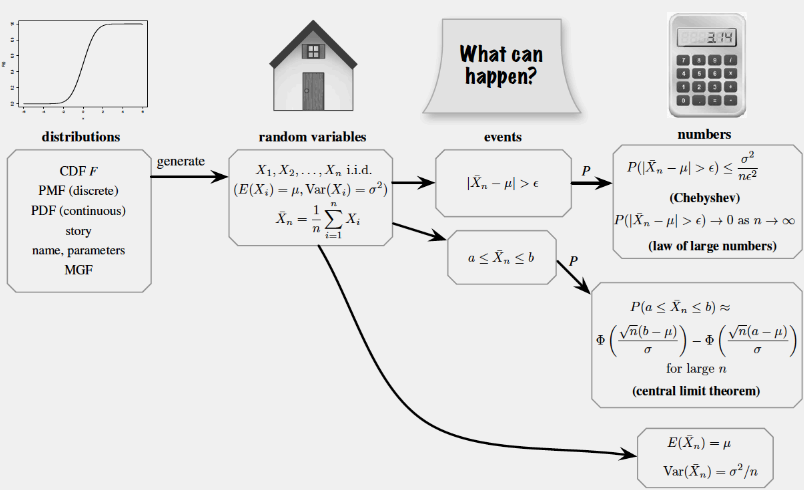
$$\Rightarrow \varepsilon = \sqrt{\frac{8\ln(\frac{2}{\delta})}{n}}$$

$\delta = 0.05$

$$\Rightarrow Pr\left(\pi \in \left(\hat{\pi} - \sqrt{\frac{8\ln(\frac{2}{\delta})}{n}},\ \hat{\pi} + \sqrt{\frac{8\ln(\frac{2}{\delta})}{n}}\right)\right) \geq 1 - \delta$$

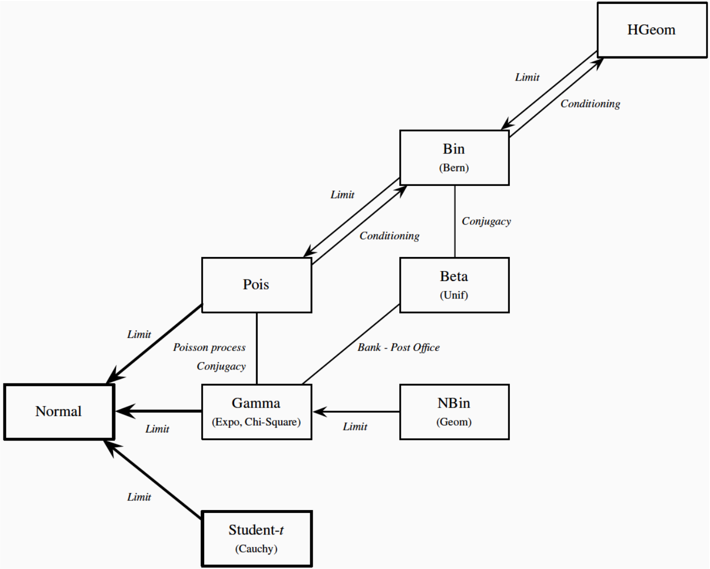
Example: Monte Carlo Method for Estimation π

Summary 1



What can happen?
distributions
random variables
events
numbers
CDF F
PMF (discrete)
PDF (continuous)
story
name, parameters
MGF
generate
$X_1, X_2, \dots, X_n$ i.i.d.
$(E(X_i) = \mu, \mathrm{Var}(X_i) = \sigma^2)$
$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
$|\bar{X}_n - \mu| > \epsilon$
$a \le \bar{X}_n \le b$
P
$P(|\bar{X}_n - \mu| > \epsilon) \le \frac{\sigma^2}{n\epsilon^2}$
(Chebyshev)
$P(|\bar{X}_n - \mu| > \epsilon) \to 0$ as $n \to \infty$
(law of large numbers)
P
$P(a \le \bar{X}_n \le b) \approx$
$\Phi\left(\frac{\sqrt{n}(b-\mu)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}(a-\mu)}{\sigma}\right)$
for large n
(central limit theorem)
$E(\bar{X}_n) = \mu$
$\mathrm{Var}(\bar{X}_n) = \sigma^2/n$

Summary 2



HGeom
Limit
Conditioning
Bin
(Bern)
Limit
Conditioning
Conjugacy
Pois
Beta
(Unif)
Limit
Poisson process
Conjugacy
Bank - Post Office
Normal
Gamma
(Expo, Chi-Square)
NBin
(Geom)
Limit
Limit
Limit
Student-t
(Cauchy)

References

- Chapter 10 of **BH**
- Chapter 5 of **BT**