Sampling Methods

Nipun Batra

September 11, 2023

IIT Gandhinagar

Topics

1. Monte Carlo Simulation
    - General Form
    - Applications
    - Bias and Variance of Monte Carlo
2. Sampling from common probability distributions
    - PRNG
    - Inverse CDF Sampling
    - Sampling from Normal Distribution
    - Rejection Sampling

The Discovery That Transformed Pi

Monte Carlo Simulation

General Form

The general form of Monte Carlo methods is: The expectation of a function $f(x)$ with respect to a distribution $p(x)$ is given by:

General Form

The general form of Monte Carlo methods is: The expectation of a function $f(x)$ with respect to a distribution $p(x)$ is given by:

$$\mathbb{E}_{x \sim p(x)}[f(x)] = \int f(x)p(x)dx \tag{1}$$

General Form

The general form of Monte Carlo methods is: The expectation of a function $f(x)$ with respect to a distribution $p(x)$ is given by:

$$\mathbb{E}_{x \sim p(x)}[f(x)] = \int f(x)p(x)dx \tag{1}$$

Using Monte Carlo methods, we can estimate the above expectation by sampling x_i from $p(x)$ and computing the average of $f(x_i)$.

General Form

The general form of Monte Carlo methods is: The expectation of a function $f(x)$ with respect to a distribution $p(x)$ is given by:

$$\mathbb{E}_{x \sim p(x)}[f(x)] = \int f(x)p(x)dx \tag{1}$$

Using Monte Carlo methods, we can estimate the above expectation by sampling x_i from $p(x)$ and computing the average of $f(x_i)$.

$$\mathbb{E}_{x \sim p(x)}[f(x)] \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i) \tag{2}$$

where $x_i \sim p(x)$.

Estimating Pi using Monte Carlo (Part 1)

We can estimate the value of pi using Monte Carlo methods by considering a unit square with a quarter circle inscribed within it.

Estimating Pi using Monte Carlo (Part 1)

We can estimate the value of pi using Monte Carlo methods by considering a unit square with a quarter circle inscribed within it.

- Let $p(x)$ be defined over the unit square using the uniform distribution in two dimensions, i.e., $p(x) = U(x) = 1$ for $x \in [0, 1]^2$.

Estimating Pi using Monte Carlo (Part 1)

We can estimate the value of pi using Monte Carlo methods by considering a unit square with a quarter circle inscribed within it.

- Let $p(x)$ be defined over the unit square using the uniform distribution in two dimensions, i.e., $p(x) = U(x) = 1$ for $x \in [0,1]^2$.
- Let $f(x)$ be the indicator function defined as follows:

$$f(x) = \begin{cases} \text{Green}(1), & \text{if } x \text{ falls inside the quarter circle,} \\ \text{Red}(0), & \text{otherwise.} \end{cases}$$

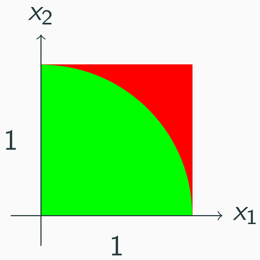
Estimating Pi using Monte Carlo (Part 1)

- Or, we can write $f(x)$ to be the following:

$$f(x) = \begin{cases} 1, & \text{if } x_1^2 + x_2^2 \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

- Or, using the indicator function, we can write $f(x)$ to be the following:

$$f(x) = \mathbb{I}(x_1^2 + x_2^2 \leq 1)$$

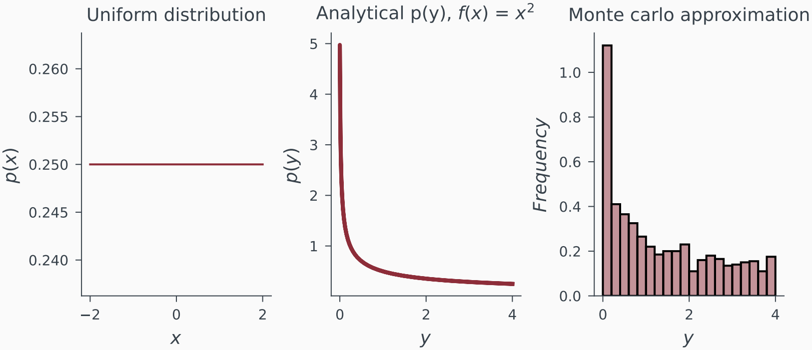


$$\frac{\pi}{4} \approx \frac{\text{Green area}}{\text{Green area} + \text{Red area}}$$

Notebook: `mc_sampling_intro.ipynb`

Estimating a function using Monte Carlo

Let $x \in \mathcal{U}(-1, 1)$ and $y = f(x) = x^2$.



Estimating prior predictive distribution

- Let $p(\theta)$ be the prior distribution of parameter. Say, for example, $p(\theta_i) = \mathcal{N}(0, 1)\ \forall i$ or $p(\theta) = \mathcal{N}(\mu, \Sigma)$.

Estimating prior predictive distribution

- Let $p(\theta)$ be the prior distribution of parameter. Say, for example, $p(\theta_i) = \mathcal{N}(0, 1)\ \forall i$ or $p(\theta) = \mathcal{N}(\mu, \Sigma)$.
- Let $p(y|\theta, x)$ be the likelihood function. Say, for example, $p(y|\theta, x) = \mathcal{N}(x^T\theta, 1)$.

Estimating prior predictive distribution

- Let $p(\theta)$ be the prior distribution of parameter. Say, for example, $p(\theta_i) = \mathcal{N}(0, 1)\ \forall i$ or $p(\theta) = \mathcal{N}(\mu, \Sigma)$.
- Let $p(y|\theta, x)$ be the likelihood function. Say, for example, $p(y|\theta, x) = \mathcal{N}(x^T\theta, 1)$.
- Then, the prior predictive distribution is given by:

$$p(y|x) = \int p(y|\theta, x)p(\theta)d\theta \tag{3}$$

$$p(y|x) \approx \frac{1}{N}\sum_{i=1}^{N} p(y|\theta_i, x) \tag{4}$$

where $\theta_i \sim p(\theta)$.

Estimating prior predictive distribution

Notebook: `mc-linreg-predictive.ipynb`

Estimating posterior predictive distribution

Extending for posterior predictive distribution, we have:

Estimating posterior predictive distribution

Extending for posterior predictive distribution, we have:

$$p(y|x, D) = \int p(y|\theta, x)p(\theta|D)d\theta \tag{5}$$

Estimating posterior predictive distribution

Extending for posterior predictive distribution, we have:

$$p(y|x, D) = \int p(y|\theta, x)p(\theta|D)d\theta \tag{5}$$

$$p(y|x, D) \approx \frac{1}{N}\sum_{i=1}^{N} p(y|\theta_i, x) \tag{6}$$

Estimating posterior predictive distribution

Extending for posterior predictive distribution, we have:

$$p(y|x, D) = \int p(y|\theta, x)p(\theta|D)d\theta \tag{5}$$

$$p(y|x, D) \approx \frac{1}{N} \sum_{i=1}^{N} p(y|\theta_i, x) \tag{6}$$

where $\theta_i \sim p(\theta|D)$.

Estimating marginal likelihood or evidence term for linear regression

[Ref: MML book 9.3.5]

We consider the following generative process:

$$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{S}_0)$$
$$y_n \mid \boldsymbol{x}_n, \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{x}_n^\top \boldsymbol{\theta}, \sigma^2\right),$$

$n = 1, \ldots, N$.

Estimating marginal likelihood or evidence term for linear regression

[Ref: MML book 9.3.5]

We consider the following generative process:

$$\begin{aligned} \boldsymbol{\theta} &\sim \mathcal{N}\left(\boldsymbol{m}_0, \boldsymbol{S}_0\right) \\ y_n \mid \boldsymbol{x}_n, \boldsymbol{\theta} &\sim \mathcal{N}\left(\boldsymbol{x}_n^\top \boldsymbol{\theta}, \sigma^2\right), \end{aligned}$$

$n = 1, \ldots, N$.

The marginal likelihood is given by

$$\begin{aligned} p(\mathcal{Y} \mid \mathcal{X}) &= \int p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \\ &= \int \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\theta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\theta} \mid \boldsymbol{m}_0, \boldsymbol{S}_0\right) \mathrm{d}\boldsymbol{\theta} \end{aligned}$$

Estimating marginal likelihood or evidence term for linear regression

[Ref: MML book 9.3.5]

We consider the following generative process:

$$\boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{m}_0, \boldsymbol{S}_0\right)$$
$$y_n \mid \boldsymbol{x}_n, \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{x}_n^\top \boldsymbol{\theta}, \sigma^2\right),$$

$n = 1, \ldots, N$.

The marginal likelihood is given by

$$\begin{aligned} p(\mathcal{Y} \mid \mathcal{X}) &= \int p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \\ &= \int \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\theta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\theta} \mid \boldsymbol{m}_0, \boldsymbol{S}_0\right) \mathrm{d}\boldsymbol{\theta} \\ &= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{m}_0, \boldsymbol{X}\boldsymbol{S}_0\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right) \end{aligned} \tag{7}$$

Estimating marginal likelihood or evidence term for linear regression

Instead if we used Monte Carlo methods, we would have:

Estimating marginal likelihood or evidence term for linear regression

Instead if we used Monte Carlo methods, we would have:

$$I = p(\mathcal{Y} \mid \mathcal{X}) \approx \frac{1}{N} \sum_{i=1}^{N} p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}_i) \tag{8}$$

where $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta})$.

Estimating Marginal Likelihood in Linear Regression

Generally, we work with log probabilities instead:

$$\log I = \log p(\mathcal{Y} \mid \mathcal{X}) \approx \log \left(\frac{1}{N} \sum_{i=1}^{N} p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}_i) \right) \tag{9}$$

The log-sum-exp trick helps us compute this efficiently.

Log-Sum-Exp Trick

The log-sum-exp trick is a technique to compute $\log\left(\frac{1}{N}\sum_{i=1}^{N} e^{a_i}\right)$ more efficiently.

$$\log\left(\frac{1}{N}\sum_{i=1}^{N} e^{a_i}\right) = \log\left(e^{\max(a_i)}\frac{1}{N}\sum_{i=1}^{N} e^{a_i - \max(a_i)}\right) \tag{10}$$

Log-Sum-Exp Trick

The log-sum-exp trick is a technique to compute $\log\left(\frac{1}{N}\sum_{i=1}^{N} e^{a_i}\right)$ more efficiently.

$$\log\left(\frac{1}{N}\sum_{i=1}^{N} e^{a_i}\right) = \log\left(e^{\max(a_i)}\frac{1}{N}\sum_{i=1}^{N} e^{a_i-\max(a_i)}\right) \tag{10}$$

$$= \max(a_i) + \log\left(\frac{1}{N}\sum_{i=1}^{N} e^{a_i-\max(a_i)}\right) \tag{11}$$

Log-Sum-Exp Trick in Linear Regression

Applying the log-sum-exp trick to linear regression:

$$\log I = \log p(\mathcal{Y} \mid \mathcal{X}) \approx \log \left(\frac{1}{N} \sum_{i=1}^{N} p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}_i) \right) \tag{12}$$

$$= \log \left(\frac{1}{N} \sum_{i=1}^{N} e^{\log p(\mathcal{Y} | \mathcal{X}, \boldsymbol{\theta}_i)} \right) \tag{13}$$

Log-Sum-Exp Trick in Linear Regression

Applying the log-sum-exp trick to linear regression:

$$\log l = \log p(\mathcal{Y} \mid \mathcal{X}) \approx \log \left(\frac{1}{N} \sum_{i=1}^{N} p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}_i) \right) \tag{12}$$

$$= \log \left(\frac{1}{N} \sum_{i=1}^{N} e^{\log p(\mathcal{Y}|\mathcal{X},\boldsymbol{\theta}_i)} \right) \tag{13}$$

$$= \log \left(\frac{1}{N} \sum_{i=1}^{N} e^{\log p(\mathcal{Y}|\mathcal{X},\boldsymbol{\theta}_i) - \max(\log p(\mathcal{Y}|\mathcal{X},\boldsymbol{\theta}_i))} \right) \tag{14}$$

Log-Sum-Exp Trick in Linear Regression

Applying the log-sum-exp trick to linear regression:

$$\log l = \log p(\mathcal{Y} \mid \mathcal{X}) \approx \log \left(\frac{1}{N} \sum_{i=1}^{N} p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}_i) \right) \tag{12}$$

$$= \log \left(\frac{1}{N} \sum_{i=1}^{N} e^{\log p(\mathcal{Y}|\mathcal{X},\boldsymbol{\theta}_i)} \right) \tag{13}$$

$$= \log \left(\frac{1}{N} \sum_{i=1}^{N} e^{\log p(\mathcal{Y}|\mathcal{X},\boldsymbol{\theta}_i) - \max(\log p(\mathcal{Y}|\mathcal{X},\boldsymbol{\theta}_i))} \right) \tag{14}$$

$$= \max(\log p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}_i)) + \log \left(\frac{1}{N} \sum_{i=1}^{N} e^{\log p(\mathcal{Y}|\mathcal{X},\boldsymbol{\theta}_i) - \max(\log p(\mathcal{Y}|\mathcal{X},\boldsymbol{\theta}_i))} \right) \tag{15}$$

Efficient Computation

The log-sum-exp trick allows us to compute $\log I$ more efficiently by:

- Subtracting the maximum value of $\log p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}_i)$ to avoid numerical issues with exponentiation.
- Adding the maximum value back after the sum of exponentials.

This technique helps prevent overflow and underflow issues when dealing with large or small values in the exponentials.

Estimating Marginal Likelihood in Linear Regression

Notebook: `mc-linreg-evidence.ipynb`

Unbiased Estimator?

Is Monte Carlo Sampling a biased or unbiased estimator?

We know:

$$\mathbb{E}_{x\sim p(x)}[f(x)] = \int f(x)p(x)dx = \phi \tag{16}$$

Let $x_i \in 1, \ldots, N$ be i.i.d samples:

$$\hat{\phi} = \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

$$\begin{aligned}\mathbb{E}(\hat{\phi}) &= \int \frac{1}{N}\sum_{i=1}^{N} f(x_i)p(x_i)dx = \frac{1}{N}\sum_{i=1}^{N}\int f(x_i)p(x_i)dx \\ &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(f(x_i)) = \phi\end{aligned}$$

Thus, it is an unbiased estimator!

Sampling converges slowly

The expected square error of the Monte Carlo estimate is given by:

$$\mathbb{E}\left(\hat{\phi} - \mathbb{E}(\hat{\phi})\right)^2 = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}(f(x_i) - \phi)\right]^2$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}(f(x_i)f(x_j)) - \phi\mathbb{E}(f(x_i)) - \mathbb{E}(f(x_j))\phi + \phi^2$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\left(\left(\sum_{i\neq j}\phi^2 - 2\phi^2 + \phi^2\right) + \mathbb{E}(f^2) - \phi^2\right) = \frac{1}{N}\mathbb{V}(f)$$

$$\therefore \mathbb{E}\left(\hat{\phi} - \mathbb{E}(\hat{\phi})\right)^2 = \mathcal{O}(N^{-1})$$

Thus, the expected error drops as $\mathcal{O}(N^{-\frac{1}{2}})$.

Sampling from common probability distributions

Sampling from uniform U(0, 1)

- Question: How can you generate samples from the uniform distribution in $[0, 1]$?

Sampling from uniform U(0, 1)

- Question: How can you generate samples from the uniform distribution in $[0, 1]$?
- Hint: Linear Congruential Generator (LCG) among other PRNGs.

Sampling from uniform U(0, 1)

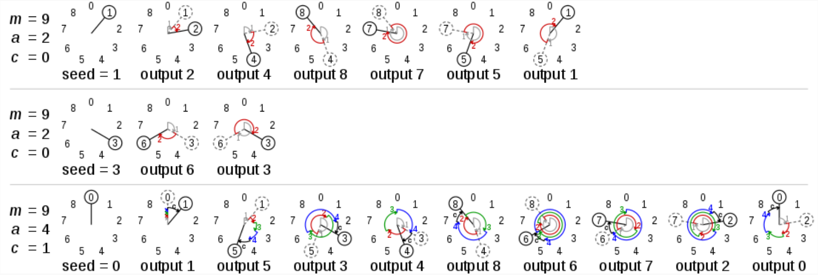
- Question: How can you generate samples from the uniform distribution in $[0, 1]$?
- Hint: Linear Congruential Generator (LCG) among other PRNGs.
- $$x_{n+1} = (ax_n + c) \mod m \tag{17}$$
- where, a, c, m are constants and x_0 is the seed
- x_{n+1} is the next random number between 0 and $m - 1$

Sampling from uniform U(0, 1)

- Question: How can you generate samples from the uniform distribution in $[0, 1]$?
- Hint: Linear Congruential Generator (LCG) among other PRNGs.
- $$x_{n+1} = (ax_n + c) \mod m \tag{17}$$
- where, a, c, m are constants and x_0 is the seed
- x_{n+1} is the next random number between 0 and $m - 1$
- $\frac{x_{n+1}}{m}$ is the next random number between 0 and 1

Sampling from uniform U(0, 1)

From Wikipedia page on LCG



Sampling from uniform U(0, 1)

Notebook: random-uniform.ipynb

Sampling from uniform U(a, b)

- Assume we have $X \sim U(0, 1)$

Sampling from uniform U(a, b)

- Assume we have $X \sim U(0, 1)$
- Then, $Y = a + (b - a)X \sim U(a, b)$

Inverse CDF sampling

[Inspired by content from Ben Lambert and Phillip Hennig]

- Let us try to generate samples from the exponential distribution.

Inverse CDF sampling

[Inspired by content from Ben Lambert and Phillip Hennig]

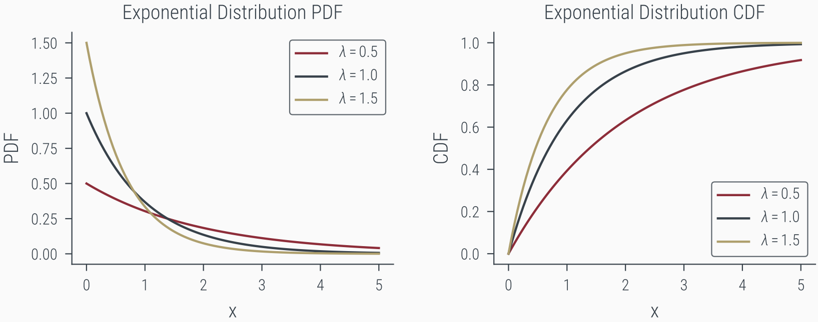
- Let us try to generate samples from the exponential distribution.
- The PDF of the exponential distribution is given by:

Inverse CDF sampling

[Inspired by content from Ben Lambert and Phillip Hennig]

- Let us try to generate samples from the exponential distribution.
- The PDF of the exponential distribution is given by:
-

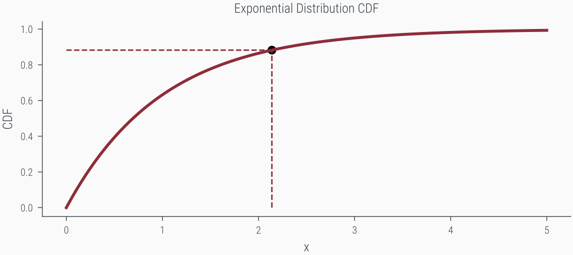
$$p(x) = \lambda e^{-\lambda x} \tag{18}$$



Notebook: inverse-cdf.ipynb

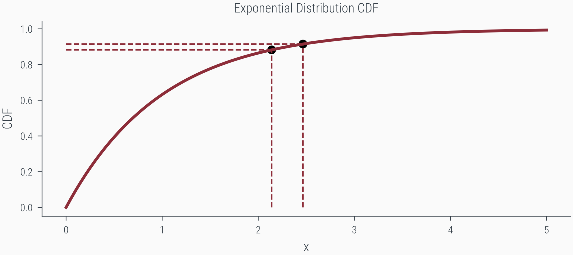
Inverse CDF Sampling for Number of samples = 1

- Let us consider the CDF ($F(x)$) of the exponential distribution ($\lambda = 1$) and try to generate samples from it.
- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.



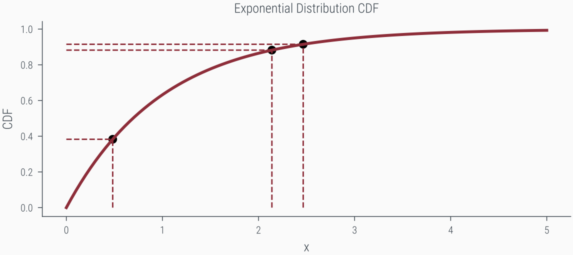
Inverse CDF Sampling for Number of samples = 2

- Let us consider the CDF ($F(x)$) of the exponential distribution ($\lambda = 1$) and try to generate samples from it.
- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.



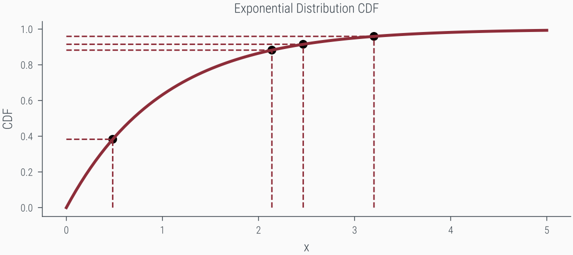
Inverse CDF Sampling for Number of samples = 3

- Let us consider the CDF ($F(x)$) of the exponential distribution ($\lambda = 1$) and try to generate samples from it.
- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.



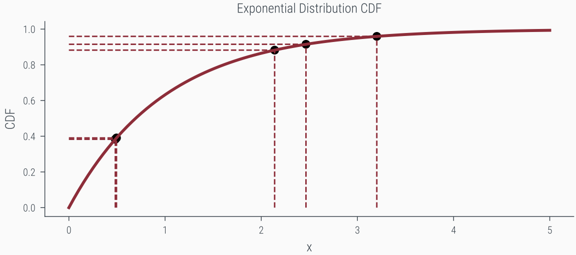
Inverse CDF Sampling for Number of samples = 4

- Let us consider the CDF ($F(x)$) of the exponential distribution ($\lambda = 1$) and try to generate samples from it.
- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.



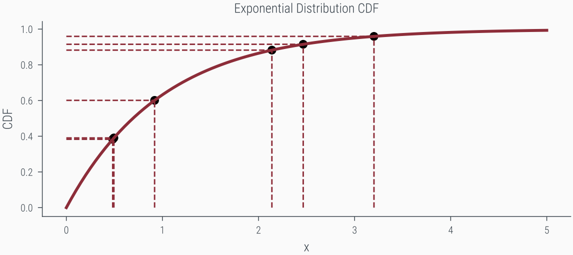
Inverse CDF Sampling for Number of samples = 5

- Let us consider the CDF ($F(x)$) of the exponential distribution ($\lambda = 1$) and try to generate samples from it.
- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.



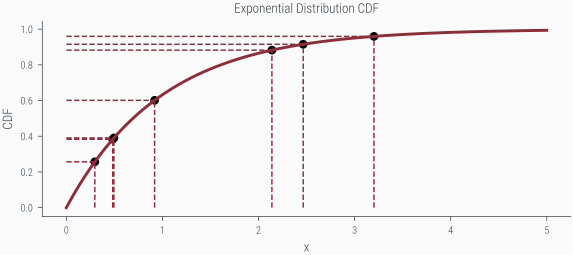
Inverse CDF Sampling for Number of samples = 6

- Let us consider the CDF ($F(x)$) of the exponential distribution ($\lambda = 1$) and try to generate samples from it.
- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.



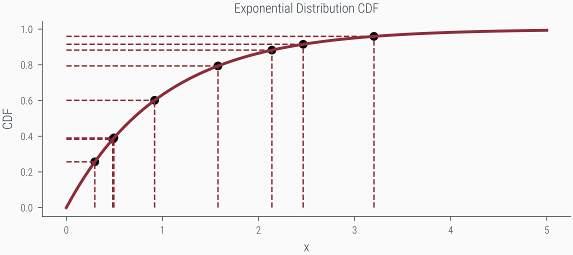
Inverse CDF Sampling for Number of samples = 7

- Let us consider the CDF ($F(x)$) of the exponential distribution ($\lambda = 1$) and try to generate samples from it.
- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.



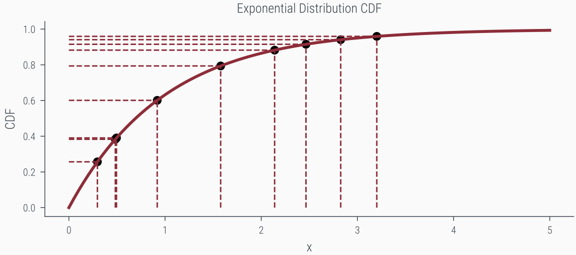
Inverse CDF Sampling for Number of samples = 8

- Let us consider the CDF ($F(x)$) of the exponential distribution ($\lambda = 1$) and try to generate samples from it.
- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.



Inverse CDF Sampling for Number of samples = 9

- Let us consider the CDF ($F(x)$) of the exponential distribution ($\lambda = 1$) and try to generate samples from it.
- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.



Inverse CDF sampling

- We generate a random number $u \sim U(0, 1)$.

Inverse CDF sampling

- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.

Inverse CDF sampling

- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.
- This is equivalent to finding the inverse of the CDF, i.e., $F^{-1}(u)$.

Inverse CDF sampling

- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.
- This is equivalent to finding the inverse of the CDF, i.e., $F^{-1}(u)$.
- For the exponential distribution, let us try to find $F^{-1}(u)$.

Inverse CDF sampling

- We generate a random number $u \sim U(0,1)$.
- We then find the value of x such that $F(x) = u$.
- This is equivalent to finding the inverse of the CDF, i.e., $F^{-1}(u)$.
- For the exponential distribution, let us try to find $F^{-1}(u)$.
-
$$u = 1 - e^{-x} \tag{19}$$

Inverse CDF sampling

- We generate a random number $u \sim U(0, 1)$.
- We then find the value of x such that $F(x) = u$.
- This is equivalent to finding the inverse of the CDF, i.e., $F^{-1}(u)$.
- For the exponential distribution, let us try to find $F^{-1}(u)$.
- $$u = 1 - e^{-x} \tag{19}$$
- $$x = -\log(1 - u) \tag{20}$$

Inverse CDF sampling

[From Wikipedia page on Inverse Transform Sampling] From $U \sim \text{Unif}[0,1]$, we want to generate X with CDF $F_X(x)$. We assume $F_X(x)$ to be a continuous, strictly increasing function, which provides good intuition.

We want to see if we can find some strictly monotone transformation $T : [0,1] \mapsto \mathbb{R}$, such that $T(U) \stackrel{d}{=} X$. We will have

$$F_X(x) = \Pr(X \leq x) = \Pr(T(U) \leq x) = \Pr\left(U \leq T^{-1}(x)\right) = T^{-1}(x), \text{ for}$$

where the last step used that $\Pr(U \leq y) = y$ when U is uniform on $[0,1]$. So we got F_X to be the inverse function of T, or, equivalently $T(u) = F_X^{-1}(u), u \in [0,1]$. Therefore, we can generate X from $F_X^{-1}(U)$.

Generating samples from $\mathcal{N}(0, 1)$ using Box-Muller Transform

[From Wikipedia page on Box-Muller Transform]

- Let $U_1, U_2 \sim U(0, 1)$ be two independent random variables.

Generating samples from $\mathcal{N}(0,1)$ using Box-Muller Transform

[From Wikipedia page on Box-Muller Transform]

- Let $U_1, U_2 \sim U(0,1)$ be two independent random variables.
- Let $Z_0, Z_1 \sim \mathcal{N}(0,1)$ be two independent random variables.

Generating samples from $\mathcal{N}(0,1)$ using Box-Muller Transform

[From Wikipedia page on Box-Muller Transform]

- Let $U_1, U_2 \sim U(0,1)$ be two independent random variables.
- Let $Z_0, Z_1 \sim \mathcal{N}(0,1)$ be two independent random variables.
- Then, $R = \sqrt{-2 \log U_1}$ and $\Theta = 2\pi U_2$ are independent random variables.

Generating samples from $\mathcal{N}(0,1)$ using Box-Muller Transform

[From Wikipedia page on Box-Muller Transform]

- Let $U_1, U_2 \sim U(0,1)$ be two independent random variables.
- Let $Z_0, Z_1 \sim \mathcal{N}(0,1)$ be two independent random variables.
- Then, $R = \sqrt{-2 \log U_1}$ and $\Theta = 2\pi U_2$ are independent random variables.
- Then, $Z_0 = R \cos \Theta$ and $Z_1 = R \sin \Theta$ are independent random variables.

Generating samples from $\mathcal{N}(0,1)$ using Box-Muller Transform

[From Wikipedia page on Box-Muller Transform]

- Let $U_1, U_2 \sim U(0,1)$ be two independent random variables.
- Let $Z_0, Z_1 \sim \mathcal{N}(0,1)$ be two independent random variables.
- Then, $R = \sqrt{-2\log U_1}$ and $\Theta = 2\pi U_2$ are independent random variables.
- Then, $Z_0 = R\cos\Theta$ and $Z_1 = R\sin\Theta$ are independent random variables.
- Z_0 and Z_1 are independent and identically distributed (i.i.d) $\mathcal{N}(0,1)$ random variables.

Notebook: sampling-normal.ipynb

Generating samples from $\mathcal{N}(\mu, \sigma)$

- Let $Z_0 \sim \mathcal{N}(0, 1)$ be independent random variables.

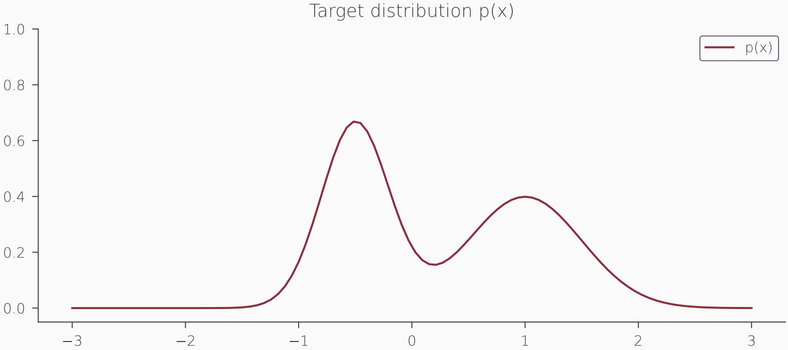
Generating samples from $\mathcal{N}(\mu, \sigma)$

- Let $Z_0 \sim \mathcal{N}(0, 1)$ be independent random variables.
- Then, $X = \mu + \sigma Z_0$ is a random variable with $\mathcal{N}(\mu, \sigma)$ distribution.

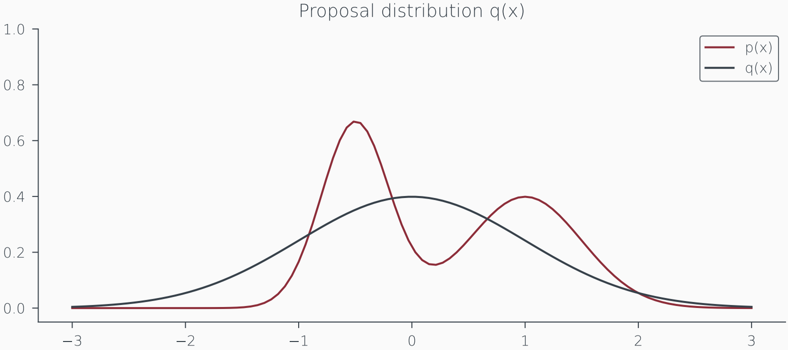
Rejection Sampling

- Let $p(x)$ be the target distribution from which we want to sample.
- Let $q(x)$ be a proposal distribution from which we can sample.
- Let M be a constant such that $M \geq \frac{p(x)}{q(x)} \forall x$.
- Then, we can sample from $p(x)$ by sampling from $q(x)$ and accepting the sample with probability $\frac{p(x)}{Mq(x)}$.

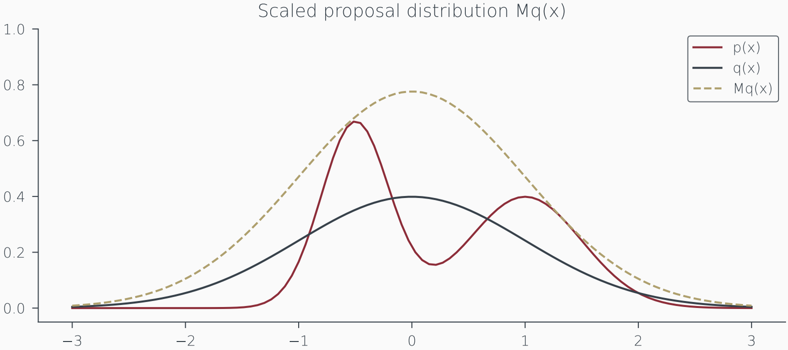
Rejection Sampling



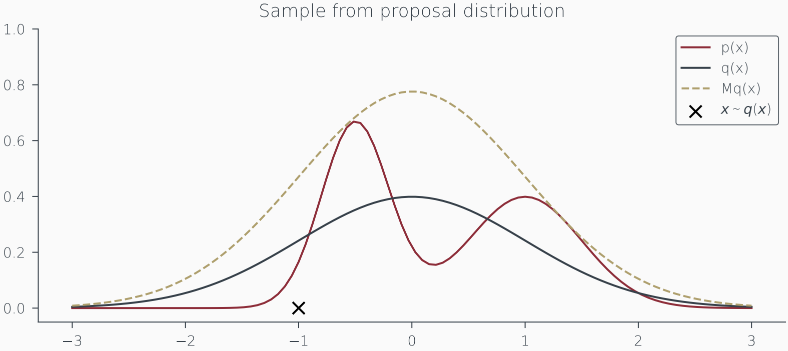
Rejection Sampling



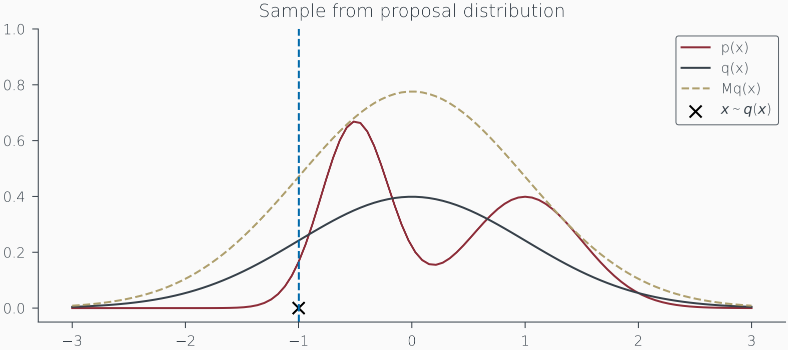
Rejection Sampling



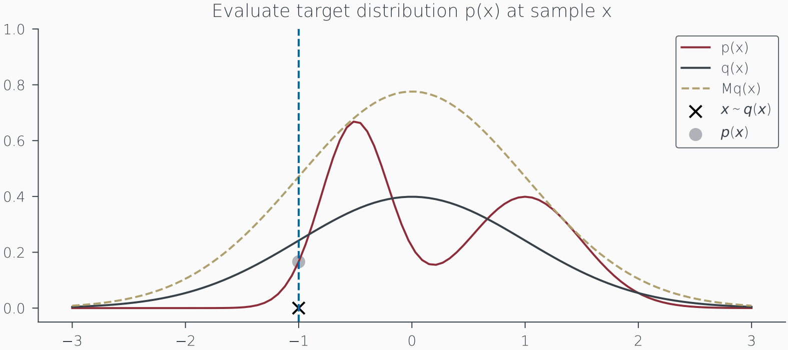
Rejection Sampling



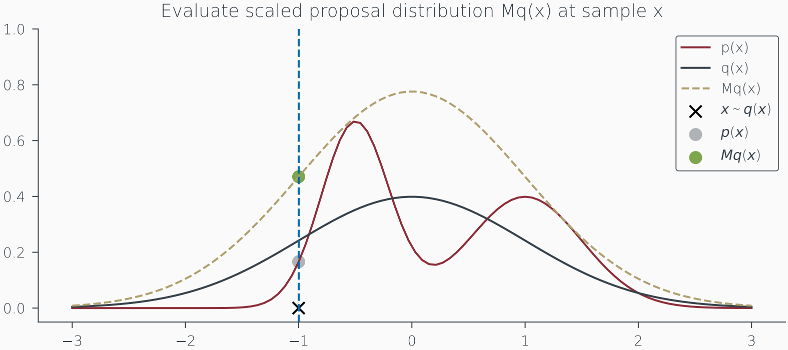
Rejection Sampling



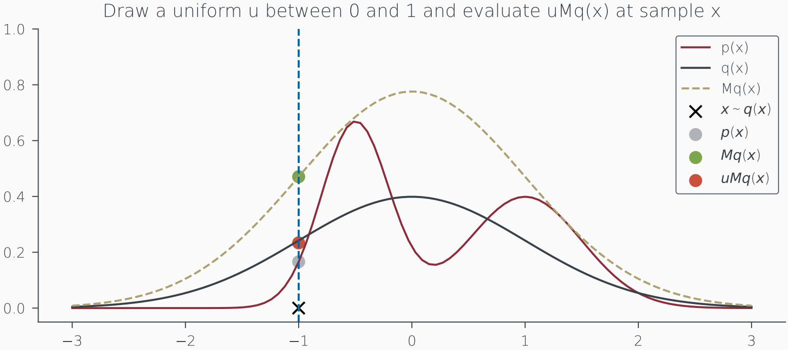
Rejection Sampling



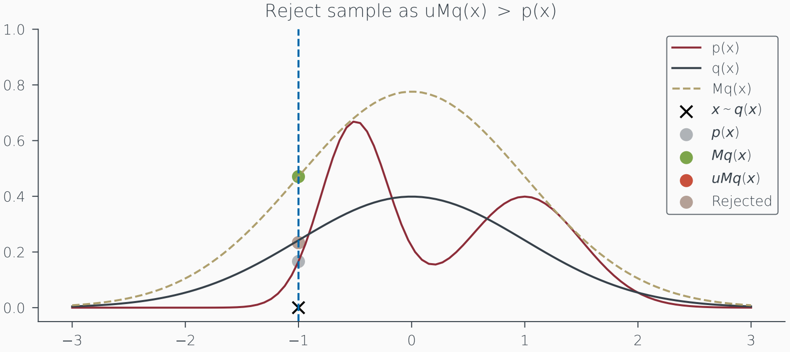
Rejection Sampling



Rejection Sampling



Rejection Sampling



Proof of Rejection Sampling

Acceptance Probability $\alpha(x)$

$$\alpha(x) = \frac{p(x)}{Mq(x)} \tag{21}$$

Bayes Rule for Acceptance

$$P(Sample|Accept) = \frac{P(Accept|Sample)P(Sample)}{P(Accept)} \tag{22}$$

P(Sample)

We draw samples from $q(x)$, so $P(Sample) = q(x)$.

Proof of Rejection Sampling

Further, $P(Accept|Sample) = \alpha(x) = \frac{p(x)}{Mq(x)}$.

Finally, $P(Accept) = \int P(Accept|Sample)P(Sample)dSample = \int \alpha(x)q(x)dx = \frac{1}{M}\int p(x)dx = \frac{1}{M}$.

P(Accept)

$$P(Accept) = \frac{1}{M} \tag{23}$$

Thus, $P(Sample|Accept) = \frac{p(x)}{Mq(x)} \times \frac{q(x)}{1/M} = p(x)$.

Thus, we have shown that the samples we accept are distributed according to $p(x)$.

Rejection Sampling Completed Example

Note: Figures not on github.

Challenges with Rejection Sampling

- Rejection sampling is inefficient when the target distribution is very different from the proposal distribution.
- In this case, we will reject a lot of samples.
- This is a problem when sampling from high-dimensional distributions.
- Acceptance probability $\alpha(x)$ is very low.