Generative Modeling Questions

Jayin Khanna

1 Questions Answered

1. How to generate 100 data points from a Gaussian distribution with mean 3 and variance 2.
2. How to generate 100 data points from a Laplacian distribution with mean 2 and spread 2.
3. Given 100 points from an unknown distribution, how to generate 10 more points from the same distribution.

2 Question 1: How to generate 100 data points from a Gaussian distribution with mean 3 and variance 2.

2.1 High level overview of Algorithm

1. Generate two independent random numbers $U_1, U_2 \sim \text{Uniform}(0, 1)$.
2. Convert these uniform random numbers into standard normal samples Z_0, Z_1 using the Box-Muller formula:
$$Z_0 = \sqrt{-2 \ln U_1} \cos(2\pi U_2), \quad Z_1 = \sqrt{-2 \ln U_1} \sin(2\pi U_2).$$
3. Rescale the standard normal samples by multiplying with the square root of the variance $\sqrt{2}$:
$$X_i = \sqrt{2} \cdot Z_i, \quad i = 0, 1.$$
4. Shift the rescaled values by adding the mean 3:
$$Y_i = X_i + 3, \quad i = 0, 1.$$
5. The final values Y_i are samples from the Gaussian distribution $\mathcal{N}(3, 2)$.

We require a bit of theory for this question.

2.2 Cartesian and Polar Coordinate Parametrization

A point in the 2D plane can be represented in two common coordinate systems:

- **Cartesian coordinates:** (x, y)
- **Polar coordinates:** (r, θ)

2.3 Conversion between Cartesian and Polar coordinates

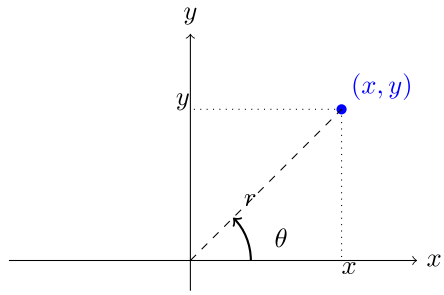
Given a point (x, y) in Cartesian coordinates, the corresponding polar coordinates are:

$$r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1}\left(\frac{y}{x}\right)$$

Conversely, given (r, θ) in polar coordinates, we can compute the Cartesian coordinates:

$$x = r \cos\theta, \quad y = r \sin\theta$$

Illustration



In this diagram, the point can be seen both as:

- A position (x, y) along the x and y axes (Cartesian)
- A radius r and angle θ from the origin (Polar)

2.4 Transformation of RV

A classical method to generate samples from $N(0,1)$ is the Box-Muller transformation method. Here, we will draw random variables (R^2, Θ) from a certain distribution in the polar coordinate system, and then use a transformation h, so that $h(R^2, \Theta) \sim N(0,1)$.

First, we will need some theory for this.
Let X and Y be i.i.d. $\sim N(0,1)$. The joint density of (X, Y) is

$$f(x,y) = \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2}, \quad x \in R,\ y \in R.$$

Let (R^2, Θ) denote the polar coordinates of (X, Y) so that

$$X = R\cos\Theta \quad \text{and} \quad Y = R\sin\Theta;$$

here the support of R is $(0, \infty)$ and the support of Θ is $(0, 2\pi)$.
Then,

$$R^2 = X^2 + Y^2, \quad \tan\Theta = \frac{Y}{X}.$$

Notationally, we denote a realization from (R^2, Θ) as (d, θ) and find the joint density of $f(d, \theta)$. Thus, let $d = x^2 + y^2$ and $\theta = \tan^{-1}(y/x)$. We know that the density for (d, θ) can be found by

$$f(d,\theta) = |J|\, f(x,y),$$

where J is the Jacobian matrix:

$$J = \begin{vmatrix} \frac{\partial x}{\partial d} & \frac{\partial y}{\partial d} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix}.$$

Solving for the Jacobian J

$$J = \begin{vmatrix} \frac{\partial\sqrt{d}\cos\theta}{\partial d} & \frac{\partial\sqrt{d}\sin\theta}{\partial d} \\ \frac{\partial\sqrt{d}\cos\theta}{\partial \theta} & \frac{\partial\sqrt{d}\sin\theta}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \frac{1}{2}\frac{\cos\theta}{\sqrt{d}} & \frac{1}{2}\frac{\sin\theta}{\sqrt{d}} \\ -\sqrt{d}\sin\theta & \sqrt{d}\cos\theta \end{vmatrix} = \frac{1}{2}$$

Joint Density of (R^2, Θ)

Since $d = x^2 + y^2$, the joint density of (R^2, Θ) is $f(d, \theta)$ with:

$$
\begin{aligned}
f(d,\theta) &= \frac{1}{2} \cdot \frac{1}{2\pi} e^{-d/2}, \qquad 0 < d < \infty, \quad 0 < \theta < 2\pi \\
&= \underbrace{\frac{1}{2\pi} I(0 < \theta < 2\pi)}_{U(0,2\pi)} \cdot \underbrace{\frac{1}{2} e^{-d/2} I(0 < d < \infty)}_{\text{Exp}(2)}
\end{aligned}
$$

This is a separable density, so R^2 and Θ are independent, and

$$\Theta \sim U[0, 2\pi], \quad R^2 \sim \text{Exp}(2).$$

To generate from Exp(2), we can use an inverse transform method. If $U \sim U(0,1)$, then by the inverse transform method,

$$-2 \log U \sim \text{Exp}(2) \quad \text{(verify for yourself)}.$$

To generate from $U(0, 2\pi)$, we know that if $U \sim U(0,1)$, then $2\pi U \sim U(0, 2\pi)$.

The Box-Muller algorithm is given below and produces X and Y from $N(0,1)$ independently.

Algorithm 1 Box-Muller algorithm for $N(0,1)$

1: Generate U_1 and U_2 from $U[0,1]$ independently
2: Set $R^2 = -2 \log U_1$ and $\Theta = 2\pi U_2$
3: Set $X = R\cos(\Theta) = \sqrt{-2\log U_1}\cos(2\pi U_2)$
4: Set $Y = R\sin(\Theta) = \sqrt{-2\log U_1}\sin(2\pi U_2)$

3 Question 2: Sampling from the Laplace Distribution via Inverse Transform Sampling

Background: Laplace Distribution

The Laplace distribution (also called the double exponential distribution) with parameters mean μ and scale $b > 0$ has the probability density function (pdf):

$$f(x) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right)$$

This distribution is symmetric around μ, with a peak at μ and heavier tails than a Gaussian.

Step 1: CDF of the Laplace Distribution

The cumulative distribution function (CDF), $F(x) = P(X \leq x)$, integrates the pdf from $-\infty$ up to x.

Due to the absolute value in the pdf, the CDF splits into two cases:

$$
F(x) = \begin{cases} \frac{1}{2} \exp\left(\frac{x - \mu}{b}\right), & x < \mu \\ 1 - \frac{1}{2} \exp\left(-\frac{x - \mu}{b}\right), & x \geq \mu \end{cases}
$$

Step 2: Inverse Transform Sampling (ITS)

Inverse transform sampling is a method to generate samples from any distribution given its CDF and its inverse.

1. Generate $U \sim \text{Uniform}(0, 1)$.
2. Set $X = F^{-1}(U)$, where F^{-1} is the inverse CDF.

Step 3: Find the Inverse CDF of the Laplace Distribution

Let $u \in (0, 1)$ be a sample from the uniform distribution.

Case 1: $u < \frac{1}{2}$
From the CDF for $x < \mu$:

$$u = \frac{1}{2} \exp\left(\frac{x - \mu}{b}\right)$$

Multiply both sides by 2:

$$2u = \exp\left(\frac{x - \mu}{b}\right)$$

Take natural logarithm:

$$\ln(2u) = \frac{x - \mu}{b}$$

Solve for x:

$$x = \mu + b \ln(2u)$$

Case 2: $u \geq \frac{1}{2}$
From the CDF for $x \geq \mu$:

$$u = 1 - \frac{1}{2} \exp\left(-\frac{x - \mu}{b}\right)$$

Rearranged:

$$\frac{1}{2} \exp\left(-\frac{x - \mu}{b}\right) = 1 - u$$

Multiply both sides by 2:

$$\exp\left(-\frac{x - \mu}{b}\right) = 2(1 - u)$$

Take natural logarithm:

$$-\frac{x - \mu}{b} = \ln\left(2(1 - u)\right)$$

Multiply both sides by $-b$:

$$x - \mu = -b \ln\left(2(1 - u)\right)$$

Solve for x:

$$x = \mu - b \ln\left(2(1 - u)\right)$$

Step 4: Combine Both Cases Using the Sign Function

Define the sign function:

$$\operatorname{sgn}(u-0.5) = \begin{cases} -1, & u < 0.5 \\ +1, & u \geq 0.5 \end{cases}$$

Also note

$$|u-0.5| = \begin{cases} 0.5-u, & u < 0.5 \\ u-0.5, & u \geq 0.5 \end{cases}$$

Then the inverse CDF can be expressed as a single formula:

$$X = \mu - b \cdot \operatorname{sgn}(u-0.5) \cdot \ln\left(1 - 2|u-0.5|\right)$$

This works because:

- For $u < 0.5$:
$$1 - 2|u-0.5| = 1 - 2(0.5-u) = 2u$$
so
$$X = \mu - (-1) \cdot b \cdot \ln(2u) = \mu + b\ln(2u)$$
- For $u \geq 0.5$:
$$1 - 2|u-0.5| = 1 - 2(u-0.5) = 2(1-u)$$
so
$$X = \mu - 1 \cdot b \cdot \ln(2(1-u)) = \mu - b\ln(2(1-u))$$

which matches the cases found earlier.

Summary

To sample a Laplace-distributed random variable:

1. Generate $u \sim \text{Uniform}(0,1)$.
2. Compute
$$X = \mu - b \cdot \operatorname{sgn}(u-0.5) \cdot \ln\left(1 - 2|u-0.5|\right).$$

This X is a sample from $\text{Laplace}(\mu, b)$.

What is Inverse Transform Sampling (ITS)?

Inverse Transform Sampling is a fundamental technique to generate random samples from **any** probability distribution, **provided** you know its cumulative distribution function (CDF) and can find its inverse.

Intuition: Why does ITS work?

Imagine you want to sample a random variable X that follows some distribution with CDF $F(x)$.

- The CDF $F(x) = P(X \leq x)$ is a function that maps a value x to a probability between 0 and 1.
- For continuous distributions, $F(x)$ is strictly increasing from 0 (at $x = -\infty$) to 1 (at $x = +\infty$).
- The inverse CDF, $F^{-1}(u)$, "undoes" this mapping: it takes a probability $u \in [0, 1]$ and returns the corresponding value x such that $F(x) = u$.

Step-by-step reasoning

1. Start with a Uniform(0,1) random variable U

- This is the key building block: generating a random number U uniformly distributed on the interval $[0, 1]$.
- Uniform(0,1) means every value between 0 and 1 is equally likely.

2. Use the inverse CDF to transform U into X

- For a given $u \in [0, 1]$, find $x = F^{-1}(u)$ — the value x such that the probability $P(X \leq x) = u$.
- Intuitively, u represents a probability threshold, and $F^{-1}(u)$ gives the quantile or cutoff value x at that probability.

Why does this produce samples from the desired distribution?

We want to prove that if $U \sim \text{Uniform}(0, 1)$, then

$$X = F^{-1}(U)$$

has distribution F.

Proof:

Let's find the CDF of X:

$$P(X \leq x) = P(F^{-1}(U) \leq x)$$

Since F is monotonically increasing:

$$P(F^{-1}(U) \leq x) = P(U \leq F(x))$$

Because F^{-1} is the inverse function of F, the event $F^{-1}(U) \leq x$ is equivalent to $U \leq F(x)$.
Since $U \sim \text{Uniform}(0, 1)$:

$$P(U \leq F(x)) = F(x)$$

This means the CDF of X is exactly $F(x)$, so X has the distribution with CDF F.

Summary

- Start with $U \sim \text{Uniform}(0, 1)$.
- Compute $X = F^{-1}(U)$.
- Then X follows the distribution defined by F.

This means **generating a uniform random number and passing it through the inverse CDF transforms a uniform sample into a sample from any desired distribution**.

4 Question 3: Given 100 points from an unknown distribution, how to generate 10 more points from the same distribution.

A lot of methods

1. Gaussian Mixture Models (GMMs)
2. Variational autoencoders(VAEs)
3. Normalizing flows
4. Generative Adeversarial Networks (GANs)

Why GANs over statistical methods?

Step 1: Classical Sampling vs. Learning to Sample Classical setting:
For known distributions (e.g., Gaussian, Laplace), you can directly sample using formulas or inverse transform sampling.
For unknown distributions (Question 3), classical approaches either:

- Estimate the distribution explicitly (e.g., Kernel Density Estimation (KDE), Gaussian Mixture Models (GMM), parametric fits),
- Or resample from data (bootstrap).

Step 2: What if you can't write down the distribution or it's too complex?
Real-world data (e.g., images, speech, text) come from extremely complex, high-dimensional distributions.
Estimating explicit density functions $p_{\text{data}}(x)$ is nearly impossible.
Classical density estimation methods fail in high dimensions.

Step 3: Enter GANs — learning an implicit generative model
GANs do not try to estimate $p_{\text{data}}(x)$ explicitly. Instead, they learn to:

- Map simple noise variables $z \sim p_z(z)$ (e.g., Gaussian noise) through a neural network $G(z; \theta)$,
- To produce samples $x = G(z; \theta)$ that "look like" they came from the unknown data distribution $p_{\text{data}}(x)$.

This defines an implicit model p_g induced by G, but without an explicit density form.

For a mathematical overview of GANs, please refer to my GAN notes.