Composition Methods

(Chapter 5)

Michael Tsiang

Stats 102C: Introduction to Monte Carlo Methods

Acknowledgements: Qing Zhou

Outline

Introduction

- The inverse CDF method is a way of transforming the uniform random variable $U \sim \mathrm{Unif}(0,1)$ into another random variable $X \sim F^{-1}(U)$ in order to sample from X.

- In addition to the inverse CDF transform, there are other types of transformations that can be applied in order to simulate random variables.

- These transformation (or composition) methods allow us a way to leverage sampling from simpler distributions to sample from more complicated distributions.

Examples: Well Known Compositions

- If $Z \sim \mathcal{N}(0,1)$, then $Z \sim N(0,1) \quad Z^2 \sim \chi^2(1)$

$$V = Z^2 \sim \chi^2(1)$$

has a chi-square distribution with 1 degree of freedom.

- If $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$ are independent, then

$Z_1 \sim \chi^2(m) \quad Z_2 \sim \chi^2(n)$

$$F = \frac{U/m}{V/n} \sim F \text{ distribution}(m,n)$$

$\frac{Z_1/m}{Z_2/n} \sim F(m,n)$

has an F-distribution with (m,n) degrees of freedom.

- If $Z \sim \mathcal{N}(0,1)$ and $V \sim \chi^2(n)$ are independent, then

$T = \frac{Z}{\sqrt{V/n}} \sim T(n)$

$$T = \frac{Z}{\sqrt{V/n}} \sim T \text{ distribution } (n)$$

has a Student's t-distribution with n degrees of freedom.

Outline

The Box-Muller Transform

- The **Box-Muller transform** (George Box and Mervin Muller, 1958) is a method to transform two uniform random variables into a pair of independent standard normal random variables.

- The main idea is to change coordinates from Cartesian to polar coordinates.

The Box-Muller Transform

$\vec{X} = (x_1, x_2 \cdots x_n)$ $\vec{y} = (y_1 \cdots y_n)$

- Let X and Y be independent standard normal random variables: $X, Y \sim \mathcal{N}(0,1)$ and $X \perp Y$. (independent)

- The joint PDF of X and Y is given by

$$
\begin{aligned}
f_{XY}(x,y) &= f(x)f(y) \quad \text{(independent)} \\
\text{(joint pdf)} \quad &= \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}}e^{-y^2/2} \\
&= \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}.
\end{aligned}
$$

$f(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}} \sim \mathcal{N}(0,1)$

The Box-Muller Transform

$xy \longrightarrow r\theta$

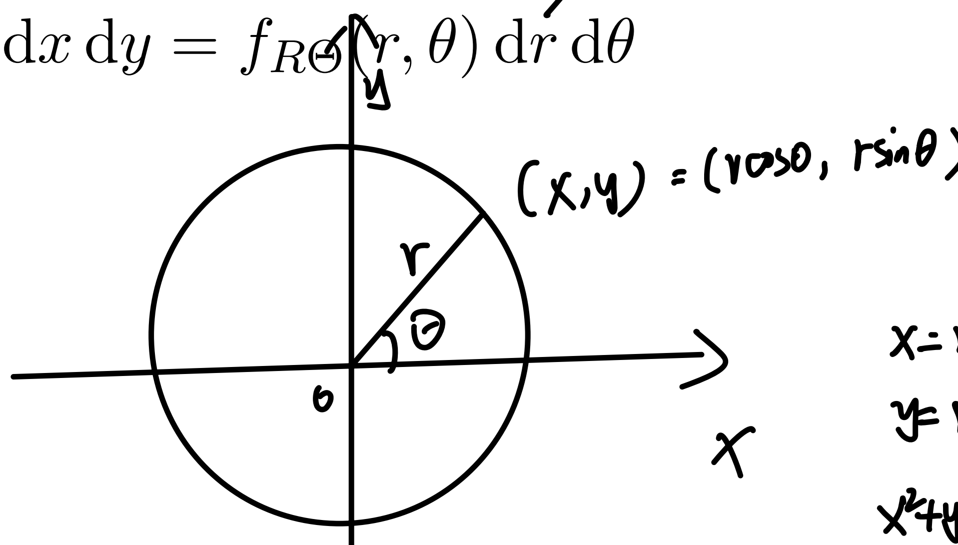
- The relationship between Cartesian coordinates (x, y) and polar coordinates (r, θ) is

$$
\begin{aligned}
x &= r\cos\theta \\
y &= r\sin\theta
\end{aligned}
$$

- Change $f_{XY}(x, y)$ to polar coordinates:

polar

$$\underbrace{f_{XY}(x, y)\,\mathrm{d}x\,\mathrm{d}y}_{\text{cartesian}} = f_{R\Theta}(r, \theta)\,\mathrm{d}r\,\mathrm{d}\theta$$



$x = r\cos\theta$

$y = r\sin\theta$

$x^2 + y^2 = r^2$

From multivariable calculus, the change of variables is given by

cartisan

polar

$$f_{R\Theta}(r,\theta) = f_{XY}(x,y)\frac{\mathrm{d}x\,\mathrm{d}y}{\mathrm{d}r\,\mathrm{d}\theta}$$

$$= f_{XY}(x,y)\underbrace{\left|\frac{\partial(x,y)}{\partial(r,\theta)}\right|}_{=J}$$

How x, y change, affect r θ change

$x = r\cos\theta \quad \frac{\partial x}{\partial r} = \cos\theta, \quad \frac{\partial x}{\partial \theta} = -r\sin\theta$

$y = r\sin\theta \quad \frac{\partial y}{\partial r} = \sin\theta, \quad \frac{\partial y}{\partial \theta} = r\cos\theta$

where J is the Jacobian

Jacobin = r

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

$\cos\theta\; r\cos\theta - (-r\sin\theta)\cdot\sin\theta$

Since $x^2+y^2=r^2$, then $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$

$$f_{R\Theta}(r,\theta) = f_{XY}(x,y)\cdot r = \frac{1}{2\pi}e^{-\frac{r^2}{2}}\cdot r.$$

polar

$f(x,y) = \frac{1}{2\pi}e^{-\frac{r^2}{2}} = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$

The Box-Muller Transform

$f(x,y) \Rightarrow f_{r\theta}(r\theta) = f(x,y)\cdot r = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \cdot r$

transformed $= \frac{1}{2\pi} e^{-\frac{r^2}{2}} \cdot r$

$\frac{1}{2}\, dr^2$

For $r \geq 0$ and $\theta \in [0, 2\pi)$, we have

变成 dr^2 ?

$$f_{R\Theta}(r,\theta)\,\mathrm{d}r\,\mathrm{d}\theta = \frac{1}{2\pi} e^{-\frac{r^2}{2}}\, r\,\mathrm{d}r\,\mathrm{d}\theta.$$

$\downarrow r^2$?

Applying another change of variables from (r, θ) to (r^2, θ), we have

$$\mathrm{d}r^2 = 2r\,\mathrm{d}r, \text{ so } r\,\mathrm{d}r = \frac{1}{2}\,\mathrm{d}r^2.$$

$f_{r\theta}(r\theta) = \frac{1}{2\pi} e^{-\frac{r^2}{2}}\, \frac{1}{2} dr^2\, d\theta$

$f(r^2)$ $\quad$ $f(\theta)$

$\rightarrow f_{r\theta}(r\theta) = \left(\frac{1}{2} e^{-\frac{r^2}{2}} dr^2\right) \cdot \left(\frac{1}{2\pi} d\theta\right)$

The Box-Muller Transform

$[0, \infty)$ $[0, 2\pi)$

The joint PDF $f_{R\Theta}(r, \theta)\,\mathrm{d}r\,\mathrm{d}\theta$, for $r \geq 0$ and $\theta \in [0, 2\pi)$, can now be written as

$$
\begin{aligned}
f_{R\Theta}(r, \theta)\,\mathrm{d}r\,\mathrm{d}\theta &= f_{R^2\Theta}(r^2, \theta)\,\mathrm{d}r^2\,\mathrm{d}\theta \\
\frac{1}{2\pi}e^{-\frac{r^2}{2}} \cdot r\,\mathrm{d}r\,\mathrm{d}\theta &= \frac{1}{2\pi}e^{-\frac{r^2}{2}} \cdot \frac{1}{2}\,\mathrm{d}r^2\,\mathrm{d}\theta \\
&= \left(\frac{1}{2}e^{-\frac{r^2}{2}}\,\mathrm{d}r^2\right)\left(\frac{1}{2\pi}\,\mathrm{d}\theta\right) \\
&= f_{R^2}(r^2)\,\mathrm{d}r^2 \cdot f_{\Theta}(\theta)\,\mathrm{d}\theta,
\end{aligned}
$$

$r^2 \sim \exp(\lambda = \frac{1}{2})$

θ distribution $\mathrm{Unif}(0, 2\pi)$

$\Theta \sim \mathrm{Uniform}(0, 2\pi)$

r^2 distribution

which shows:

$\exp(\frac{1}{2})$ → inverse generate from uniform

- $R^2 \perp \Theta$ (i.e., R^2 and Θ are independent) $r^2 \in [0, +\infty)$
- $\Theta \sim \mathrm{Unif}(0, 2\pi)$ → $\theta\ [0, 2\pi)$
- $R^2 \sim \mathrm{Exp}(\lambda = \frac{1}{2})$

The Box-Muller Transform

1. Generate $\Theta \sim \mathrm{Unif}(0, 2\pi)$. ✓
2. Generate $V \sim \mathrm{Exp}(\lambda = \frac{1}{2})$ (i.e., $V = R^2$) and compute
$$R = \sqrt{V}.$$ $f(r^2) = \frac{1}{2} e^{-\frac{1}{2} r^2}$ $(\lambda = \frac{1}{2})$
3. Compute

exp uniform

$$X = R\cos\Theta$$
$$Y = R\sin\Theta.$$

Then $X, Y \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. $f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$. (CDF)

Note that we can use $U \sim \mathrm{Unif}(0, 1)$ to sample from $\mathrm{Exp}(\lambda = \frac{1}{2})$ using the inverse CDF method:

let $r^2 = v$

$$-\frac{1}{\lambda}\log U = -2\log U \sim \mathrm{Exp}(\lambda = \tfrac{1}{2}).$$

$f(v) = \frac{1}{2} e^{-\frac{1}{2}v^2} \rightarrow F(v) \Rightarrow F^{-1}(v) = U \Rightarrow V = -\frac{1}{\lambda}\log(U)$ where $U \sim \mathrm{unif}(0,1)$

The Box-Muller Transform

Box-Muller algorithm:

$\Theta = \text{Unifor}(0, 2\pi)$

$U = \text{unifor}(0,1)$

1. Generate $U \sim \text{Unif}(0,1)$ and compute
$$\Theta = 2\pi U.$$

$\therefore \Theta = 2\pi U$

$\rightarrow$ for exp$(\frac{1}{2})$

2. Generate $V \sim \text{Unif}(0,1)$ and compute
$$R = \sqrt{-2\log V}.$$

$R^2 = -2\log V$

$V \sim \text{unifor}(0,1)$

$V = \sqrt{-2\log V}$

3. Compute
$$\begin{aligned} X &= R\cos\Theta \\ Y &= R\sin\Theta. \end{aligned}$$

Then $X, Y \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$.

R Code to sample from $X, Y \sim \mathcal{N}(0, 1)$:

```
> set.seed(9999) # Set the seed for reproduceability
> n <- 1000 # Specify the number of points to generate

> # Generate n points from Unif(0, 1)
> U <- runif(n, 0, 1)
```

$U \begin{cases} \theta \\ V \end{cases}$ $U \to \theta$

```
> # Compute Theta
> Theta <- 2 * pi * U
```

$\theta = 2 \cdot \pi U$

θ

```
> # Generate n points from Unif(0, 1)
> V <- runif(n, 0, 1)
```

V $V \to R$

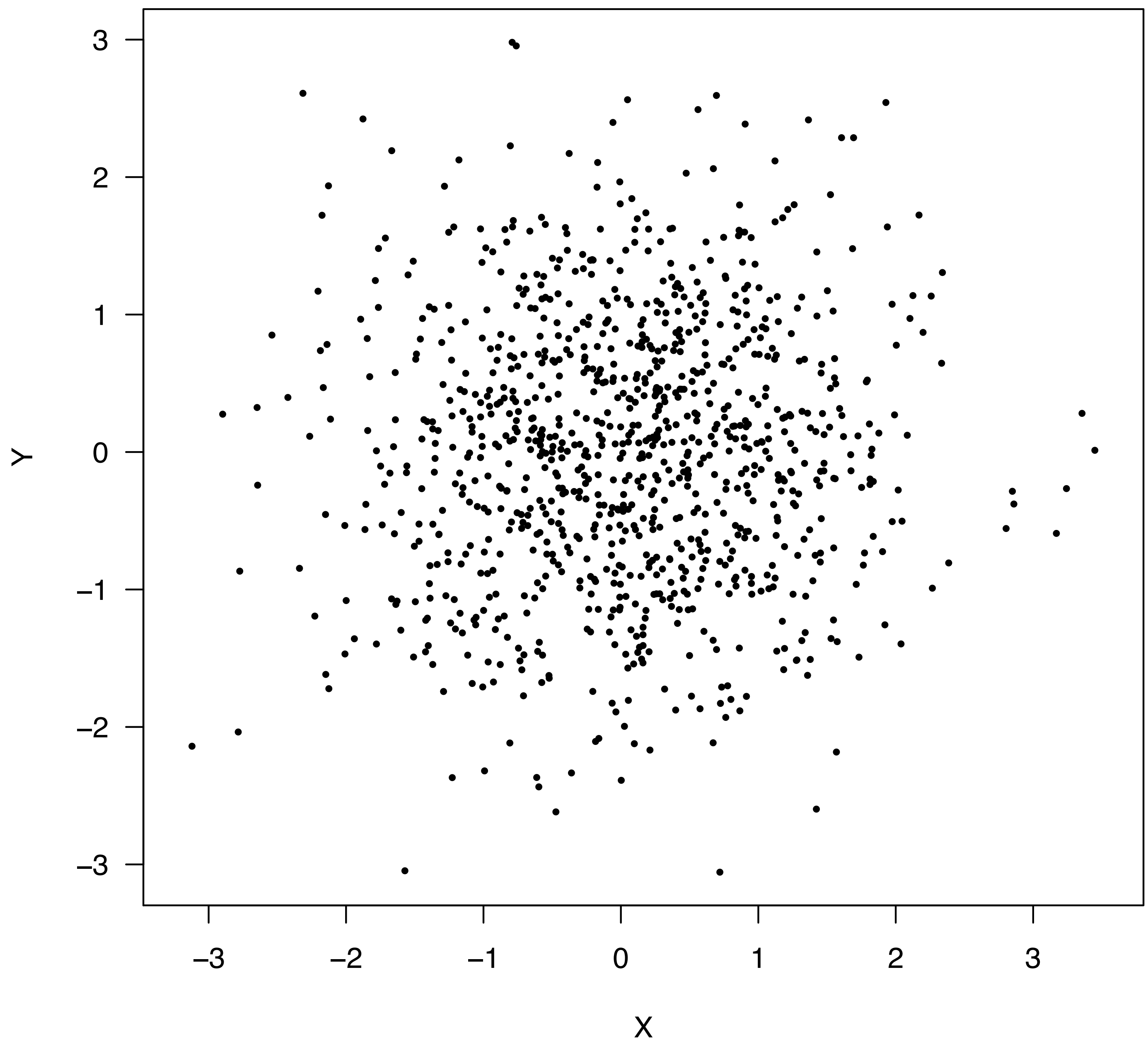
```
> # Compute R
> R <- sqrt(-2 * log(V))
```

$R = \sqrt{-2\log(V)}$

```
> # Compute X and Y
> X <- R * cos(Theta)
> Y <- R * sin(Theta)
```

generate X and y

The Box-Muller Transform



R Code for the plot:

```
> plot(X, Y, pch = 19, cex = 0.4, asp = 1, las = 1)
```

generate from $f(x,y) \sim$ brivate normal (standard)

$\rightarrow f(\theta, V)$

Normal Distribution $\mathcal{N}(\mu, \sigma^2)$

- We can now assume we can generate samples from $\mathcal{N}(0,1)$.
- How can we use samples from $\mathcal{N}(0,1)$ to sample from

$$W \sim \mathcal{N}(\mu, \sigma^2), \text{ for any } \mu \in \mathbb{R}, \sigma^2 > 0?$$

from standard normal to normal

single normal :

$$\boxed{\begin{array}{ll} \mu(AX) = A\mu(X) & \mu(AX+b) = A\mu(X) + b \\ var(AX) = A^2 var(X) & var(AX+b) = A^2 var(X) \end{array}}$$

Normal Distribution $\mathcal{N}(\mu, \sigma^2)$

1. Generate $Z \sim \mathcal{N}(0, 1)$. standard normal

2. Then $W = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$.

Why?

- W is normally distributed.
- $E(W) = E(\mu + \sigma Z) = \mu + \sigma \overset{0}{E(Z)} = \mu$
- $\mathrm{Var}(W) = \mathrm{Var}(\mu + \sigma Z) = \mathrm{Var}(\sigma Z) = \sigma^2 \underset{1^2}{\mathrm{Var}(Z)} = \sigma^2$

Outline

Bivariate Normal Distribution $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$

marginal

- Suppose $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.
- $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$ has a **bivariate normal distribution** with mean vector

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}, = \begin{pmatrix} Var(X_1) & cov(X_1, X_2) \\ cov(X_2, X_1) & Var(X_2) \end{pmatrix}$$

where

$$\sigma_{12} = \mathrm{Cov}(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)].$$

Bivariate Normal Distribution $\mathcal{N}_2(\mu, \Sigma)$

Bivariate

- Let $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{pmatrix}$.

- Let A be a 2×2 matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. Suppose $= \begin{pmatrix} \text{var}X_1 & \sigma_{12} \\ \sigma_{21} & \text{var}X_2 \end{pmatrix}$

  Constant

$$Y = AX = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}. \sim \text{Normal}$$

  Then $Y \sim \mathcal{N}_2(\boldsymbol{\mu}_Y, \Sigma_Y)$ has a bivariate normal distribution, with

$$\boldsymbol{\mu}_Y = A\boldsymbol{\mu} \text{ and } \Sigma_Y = A\Sigma A^T.$$

- How can we use samples from $\mathcal{N}(0, 1)$ to sample from

$$X \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma), \text{ for any } \boldsymbol{\mu} \in \mathbb{R}^2, \Sigma > 0?$$

$u(Y) = u(AX) = Au(X) = Au$ $\quad$ $\text{cov}(AX) = A\,\text{cov}(x)A^T = A\Sigma A^T$

Bivariate Normal Distribution $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$

We want to sample from $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$. $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$

- Generate $Z_1, Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. Then

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}_2(\mathbf{0}, I), \quad \text{where } \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

identity matrix $\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & & & \vdots \\ 0 & & 1 & \cdots & \vdots \\ 0 & \cdots & \cdots & \cdot & 1 \end{bmatrix}$

- Let

$$X = \boldsymbol{b} + AZ \sim \mathcal{N}_2(\boldsymbol{\mu}_X, \Sigma_X),$$

for some vector $\boldsymbol{b}$ and matrix A. $\mu(X) = \mu(Z) = b$

$\mathrm{cov}(X) = A \Sigma A^T = A I A^T = AA^T$

- We want to find $\boldsymbol{b}$ and A such that $\boldsymbol{\mu}_X = \boldsymbol{\mu}$ and $\Sigma_X = \Sigma$.

Bivariate Normal Distribution $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$

By the linearity of expectation,

$$
\begin{aligned}
\boldsymbol{\mu}_X &= E(X) \\
&= E(\boldsymbol{b} + AZ) \\
&= \boldsymbol{b} + E(AZ) \\
&= \boldsymbol{b} + AE(Z) \\
&= \boldsymbol{b} + A \cdot \mathbf{0} \\
&= \boldsymbol{b} \quad \text{(New)}
\end{aligned}
$$

So if $\boldsymbol{b} = \boldsymbol{\mu}$, then $\boldsymbol{\mu}_X = \boldsymbol{b} = \boldsymbol{\mu}$. (new)

Bivariate Normal Distribution $\mathcal{N}_2(\mu, \Sigma)$

By the properties of covariance,

$$\begin{aligned}\Sigma_X &= \mathrm{Cov}(X)\\ &= \mathrm{Cov}(\boldsymbol{b} + AZ)\\ &= \mathrm{Cov}(AZ)\\ &= A\mathrm{Cov}(Z)A^T\\ &= AIA^T\\ &= AA^T, \quad \text{(new cov mat)}\end{aligned}$$

so $\Sigma_X = AA^T$.

We want to find a matrix A such that $AA^T = \Sigma$.

$X \sim (b, AA^T)$

from $\underset{\sim\sim\sim}{X = b + AZ}$

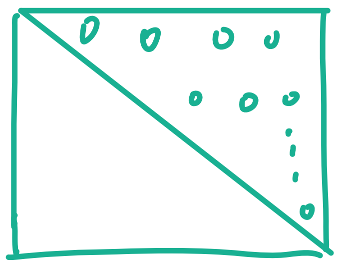
cov(A X) = A cov(X) A^T

- For any symmetric, positive definite matrix Σ, there exists a unique lower triangular matrix A such that

$$\Sigma = AA^T.$$

  This form of Σ is called the $\boxed{\textbf{Cholesky decomposition}.}$

- How do we find the the lower triangular matrix A?



The lower triangular matrix A has the form

$$A = \begin{pmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{pmatrix} \cdot A^T \begin{bmatrix} t_{11} & t_{21} \\ 0 & t_{22} \end{bmatrix}$$

We want to find t_{11}, t_{21}, and t_{22} such that

$$AA^T = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \cdot = \begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \left[\begin{matrix} t_{11}^2 & t_{11}t_{21} \\ t_{11}t_{21} & t_{21}^2+t_{22}^2 \end{matrix}\right. & \left.\right] \\ \sigma_{21} & \sigma_2^2 \end{array}$$

$t_{11} t_{21} = \sigma_{12}$ $\qquad$ $t_{11}^2 = \sigma_1^2 \longrightarrow t_{11} = \sigma_1$

$= \sigma_1 \cdot t_{21} = \sigma_{12} \longrightarrow t_{21} = \dfrac{\sigma_{12}}{\sigma_1}$

$t_{22}^2 + t_{21}^2 = \sigma_2^2$

$t_{22}^2 + \left(\dfrac{\sigma_{12}}{\sigma_1}\right)^2 = \sigma_2^2 \longrightarrow t_{22} = \sqrt{\sigma_2^2 - \left(\dfrac{\sigma_{12}}{\sigma_1}\right)^2}$

Bivariate Normal Distribution $\mathcal{N}_2(\mu, \Sigma)$

We want to find t_{11}, t_{21}, and t_{22} such that

$$
\begin{aligned}
AA^T &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \\
\begin{pmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{pmatrix} \begin{pmatrix} t_{11} & t_{21} \\ 0 & t_{22} \end{pmatrix} &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \\
\begin{pmatrix} t_{11}^2 & t_{11}t_{21} \\ t_{11}t_{21} & t_{21}^2 + t_{22}^2 \end{pmatrix} &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.
\end{aligned}
$$

Setting terms equal to each other, we have

$$
\left\{
\begin{aligned}
t_{11} &= \sqrt{\sigma_1^2} &&= \sigma_1 \\
t_{21} &= \frac{\sigma_{12}}{t_{11}} &&= \frac{\sigma_{12}}{\sigma_1} \\
t_{22} &= \sqrt{\sigma_2^2 - t_{21}^2} &&= \sqrt{\sigma_2^2 - \left(\frac{\sigma_{12}}{\sigma_1}\right)^2}.
\end{aligned}
\right.
$$

Bivariate Normal Distribution $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$

Thus we have shown the following result:

Let $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ be any vector in $\mathbb{R}^2$, and let Σ be a symmetric, positive definite 2×2 matrix.

If $Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}_2(\mathbf{0}, I)$, where $\mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, then

$$X = \overset{(b)}{\boldsymbol{\mu}} + AZ = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \underset{A}{\begin{pmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{pmatrix}} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma),$$

where $\Sigma = AA^T$.

Bivariate Normal Distribution $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$

Algorithm to sample from $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$:

1. Generate $Z_1, Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. Then
$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}_2(\mathbf{0}, I), \quad \text{where } \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$
2. Compute the Cholesky decomposition of Σ: Find a lower triangular matrix A such that
$$\Sigma = AA^T.$$
3. Then
$$X = \boldsymbol{\mu} + AZ \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma).$$

Note: This algorithm generalizes to sample from any multivariate normal distribution $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ for any dimension d.

R Code to sample from $X \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$:

```
> set.seed(9999) # Set the seed for reproduceability
> n <- 1000 # Specify the number of points to generate

> # Set values for mu and Sigma
> mu <- c(1, 2)
> Sigma <- cbind(c(1, 0.7), c(0.7, 1))
```

Z

$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$

$\Sigma = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}$

```
> # Generate n vectors from N_2(0, I)
> Z <- matrix(rnorm(2 * n, mean = 0, sd = 1),
+             nrow = 2, ncol = n
+             )
```

2000

rnorm(0,1)

$\begin{array}{cc} z_1 & z_2 \end{array}$
$\begin{bmatrix} & \\ & \end{bmatrix}$

```
> # Compute Cholesky decomposition of Sigma
> # chol() outputs the upper triangular matrix t(A)
> A <- t(chol(Sigma))

> # Compute X = mu + AZ
> X <- mu + A %*% Z
```

$X = u + AZ$

lower triangular mat

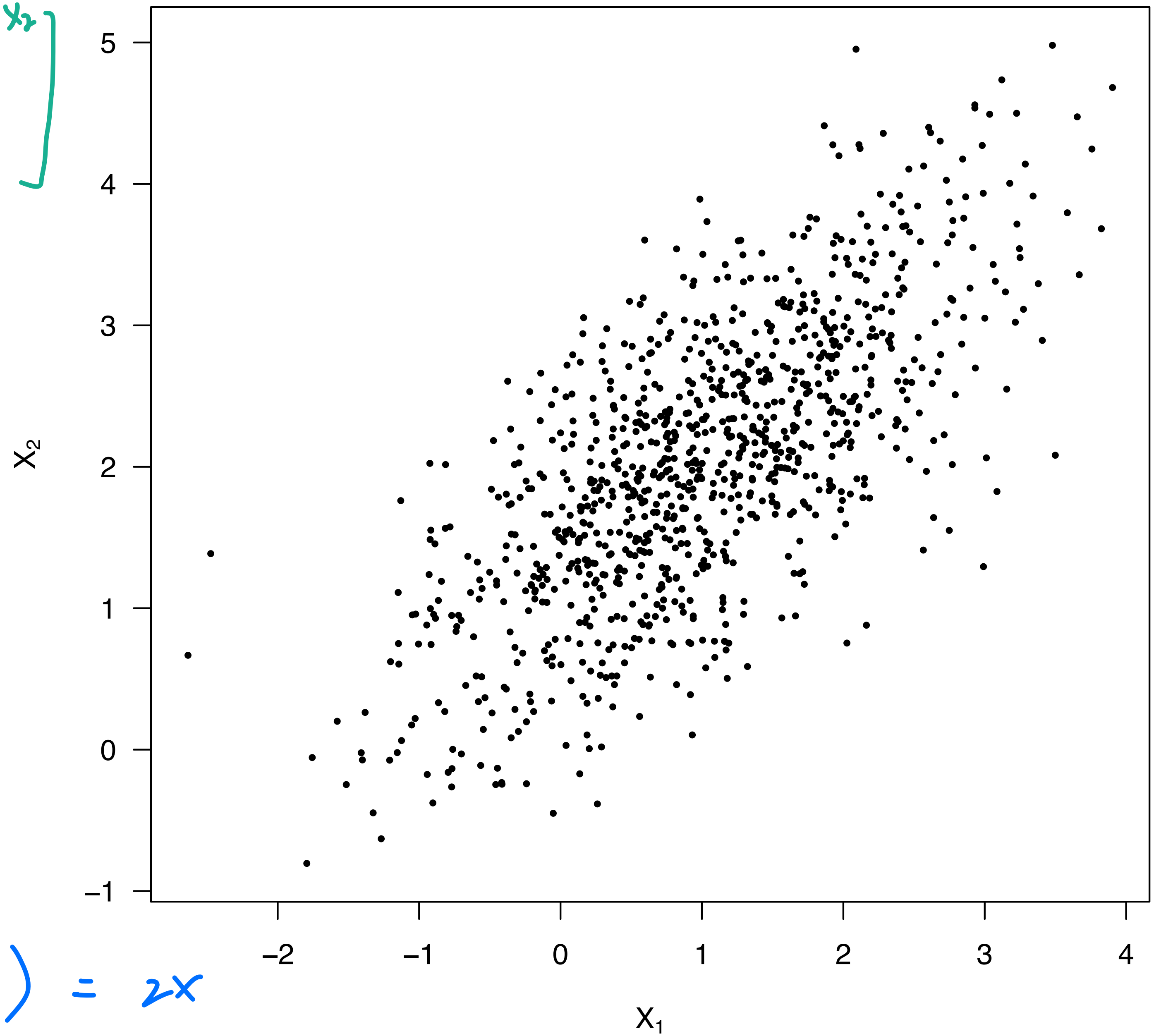
normal distributio

from initial u

(1,2)

Bivariate Normal Distribution $\mathcal{N}_2(\mu, \Sigma)$

$$\begin{array}{cc} X_1 & X_2 \end{array}$$
$$\begin{bmatrix} \quad & \quad \\ & \\ & \end{bmatrix}$$



$(X_1 \; X_2) = 2x$

R Code for the plot:

```
> plot(X[1, ], X[2, ],
+      pch = 19, cex = 0.4, asp = 1, las = 1,
+      xlab = expression(X[1]), ylab = expression(X[2])
+      )
```

Outline

- Let $X_1, X_2, \ldots, X_m$ be independent random variables. The **convolution** of $X_1, X_2, \ldots, X_m$ is the sum

$$S = \sum_{i=1}^{m} X_i = X_1 + X_2 + \cdots + X_m.$$

- Many common random variables can be expressed as a convolution.

- To simulate from a convolution, we can generate samples from $X_1, X_2, \ldots, X_m$ and compute the sum.

Sum Bernoulli = Binomial

- **Binomial**: Let $X_1, X_2, \ldots, X_m \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$ be Bernoulli random variables with parameter p, i.e., for any i,

$$X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$$

Then

$$S = \sum_{i=1}^{m} X_i = \underbrace{\overset{1}{X_1} + \overset{1}{X_2} + \overset{0}{\cdots} \overset{0}{+} \overset{1 \cdots}{\cdots} + X_m}_{\text{Sum = \# success}} \sim \text{Bin}(m, p)$$

has a binomial distribution with parameters m and p.

Convolutions

Sum poisson = poisson ($\Sigma\lambda_i$)

- **Poisson**: If $X_i \sim \mathrm{Pois}(\lambda_i)$, $i = 1, 2, \ldots, m$, are independent Poisson random variables, for $\lambda_i > 0$ for all i, then

$$S = \sum_{i=1}^{m} X_i = X_1 + X_2 + \cdots + X_m \sim \mathrm{Pois}\left(\sum_{i=1}^{m} \lambda_i\right)$$

Sum(poss) = poss

has a Poisson distribution with mean parameter $\sum_{i=1}^{m} \lambda_i$.

- **Negative Binomial**: If $X_1, X_2, \ldots, X_m \overset{\text{iid}}{\sim} \mathrm{Geom}(p)$, then

Sum(geo) ~ Negative Bin (m, p)

$$S = \sum_{i=1}^{m} X_i = X_1 + X_2 + \cdots + X_m \sim \mathrm{NegBin}(m, p)$$

has a negative binomial distribution with parameters m and p.

Sum geo = NegBin (m p)

Convolutions

sum normal^2 = chi (m)

- **Chi-square**: If $Z_1, Z_2, \ldots, Z_m \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, then

sum(normal^2) ~ chi (m)

$$S = \sum_{i=1}^{m} Z_i^2 = Z_1^2 + Z_2^2 + \cdots + Z_m^2 \sim \chi^2(m)$$

sum(chi) ~ chi

has a chi-square distribution with m degrees of freedom.

- **Gamma**: If $X_1, X_2, \ldots, X_m \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$, for $\lambda > 0$, then

sum(exp) ~ gamma (m, λ)

$$S = \sum_{i=1}^{m} X_i = X_1 + X_2 + \cdots + X_m \sim \text{Gamma}(m, \lambda)$$

has a gamma distribution with parameters m and λ.

sum exp = gamma (λ)

Convolutions

- Using convolutions, we can apply more general transforms to generate more complicated distributions.

- If $X_1, X_2, \ldots, X_a, X_{a+1}, \ldots, X_{a+b} \overset{\text{iid}}{\sim} \text{Exp}(\lambda = 1)$, then

$$Y = \frac{\sum_{i=1}^{a} X_i}{\sum_{i=1}^{a+b} X_i} \sim \text{Beta}(a, b)$$

$\frac{\Sigma \exp}{\Sigma \exp} = \text{Beta}$

has a beta distribution with parameters a and b, for $a, b \in \mathbb{N}$.

- If $U \sim \text{Gamma}(\alpha, \lambda)$ and $V \sim \text{Gamma}(\beta, \lambda)$, for $\alpha, \beta, \lambda > 0$, are independent, then

$$X = \frac{U}{U + V} \sim \text{Beta}(\alpha, \beta)$$

$\frac{\text{gamma}(\alpha, \lambda)}{\text{gamma}(\alpha, \lambda) + \text{gamma}(\beta, \lambda)}$

has a beta distribution with parameters α and β. $= \text{Beta}(\alpha, \beta)$

Outline

$a+b=1$ $a+c+e=1$

Mixtures

$af_1(x) + bf_2(x)$ | $af_1(x)^2 + cf_2(x) + ef_1(x)$

marginal briarivate

- Let $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(3, 2^2)$, with respective PDFs

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{and} \quad f_2(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-3}{2}\right)^2}.$$

- Consider the **mixture normal distribution**

$$\begin{aligned} f(x) &= 0.5 f_1(x) + 0.5 f_2(x) \\ &= 0.5 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} + 0.5 \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-3}{2}\right)^2}. \end{aligned}$$

  Does this define a proper PDF?

- How can we sample from $f(x)$?

$$f(x) = \begin{cases} f_1(x) & p=0.5 \\ f_2(x) & p=0.5 \end{cases}$$

Mixtures

Verify that $f(x)$ is a PDF:

- $f(x) \geq 0$ for all $x \in \mathbb{R}$:

  For any $x \in \mathbb{R}$, $f_1(x) \geq 0$ and $f_2(x) \geq 0$, so

  $$f(x) = 0.5 f_1(x) + 0.5 f_2(x) \geq 0.$$

- $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$:

  $\frac{1}{2}$ time generate from $f_1(x)$

  $\frac{1}{2}$ time generate from $f_2(x)$

  $$\begin{aligned}\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x &= \int_{-\infty}^{\infty} [0.5 f_1(x) + 0.5 f_2(x)]\,\mathrm{d}x \\ &= 0.5 \underbrace{\int_{-\infty}^{\infty} f_1(x)\,\mathrm{d}x}_{1} + 0.5 \underbrace{\int_{-\infty}^{\infty} f_2(x)\,\mathrm{d}x}_{1} \\ &= 0.5 + 0.5 \\ &= 1\end{aligned}$$

Mixtures

Algorithm to sample from the mixture $0.5\,\mathcal{N}(0,1) + 0.5\,\mathcal{N}(3,2^2)$:

1. Generate a value K from the PMF given by

K	1	2
$P(K=k)$	0.5	0.5

$p(X=1)=0.5=p(X=2)$

2. Generate $X \sim f_K(x)$. In other words,

$$X \sim \begin{cases} \mathcal{N}(0,1) & \text{if } K = 1 \\ \mathcal{N}(3,2^2) & \text{if } K = 2. \end{cases}$$

$$\to f(x) \begin{cases} f_1(x) & 0.5 \\ f_2(x) & 0.5 \end{cases}$$

Then $X \sim f(x) = 0.5 f_1(x) + 0.5 f_2(x)$.

Mixtures

A random variable X has a **mixture distribution** if its PDF is a weighted sum

weight Sum(θ_i)=1

$$f(x) = \sum_{i=1}^{m} \theta_i f_i(x) = \theta_1 f_1(x) + \theta_2 f_2(x) + \cdots + \theta_m f_m(x),$$

or proporition

for some sequence of PDFs $f_1(x), f_2(x), \ldots, f_m(x)$ and **mixing weights** $\theta_i > 0$ such that $\sum_{i=1}^{m} \theta_i = 1$.

Mixtures

Algorithm to sample from the mixture $f(x) = \sum_{i=1}^{m} \theta_i f_i(x)$:

1. Generate a value K from the PMF given by

dice roll

K	1	2	$\cdots$	m
$P(K = k)$	θ_1	θ_2	$\cdots$	θ_m

discrete

2. Generate $X \sim f_K(x)$. In other words,

$$X \sim \begin{cases} f_1(x) & \text{if } K = 1 \\ f_2(x) & \text{if } K = 2 \\ \vdots & \vdots \\ f_m(x) & \text{if } K = m. \end{cases}$$

Then $X \sim f(x) = \sum_{i=1}^{m} \theta_i f_i(x)$.