Sampling Random Variates

- X random variable.

Generate SAMPLES/OBSERVATIONS/REALIZATIONS

$x_1, \dots, x_N$ of R.V. X

EXAMPLES: THROW A COIN

ROLL A DIE

DO IT ON A COMPUTER

1) ADD A RANDOMNESS SOURCE (EXTERNAL)

2) PSEUDO-RANDOM NUMBER GENERATOR

generate a sequence of numbers which looks random. → passes some statistical test.

ALL PRNGs generate samples from UNIFORM DISTRIBUTION in [0,1]

$$X \sim U(0,1) \quad \text{iff} \quad p(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & \text{otherwise} \end{cases}$$

- PRNGs in pills
- SAMPLING OTHER R.V.

⋮

- McMC

Pseudo-Random Number Generator (PRNG)

<u>def</u>: $(\mathcal{S}, \mathcal{U}, s_0, f, g)$

- $\mathcal{S}$ state space
- $\mathcal{U}$ output space
- $s_0 \in \mathcal{S}$ initial seed
- $f : \mathcal{S} \to \mathcal{S}$
- $g : \mathcal{S} \to \mathcal{U}$

$\mathcal{S} = \mathbb{Z} \bmod m = \{0, \dots, m-1\}$

m LARGE

$\mathcal{U} = [0, 1]$

$g(s) = \frac{s}{m}$

PRNG works by producing a sequence

$s_0, \ s_1 = f(s_0), \ s_2 = f(s_1), \ \dots, \ s_i = f(s_{i-1}) \dots$

$u_1 = g(s_1) \quad u_2 = g(s_2) \quad \dots \quad u_i = g(s_i) \quad \dots$

- m is VERY BIG
- $\mathcal{S}$ is FINITE

$S = \{s_i\}_{i \in \mathbb{N}} \quad s_i = f(s_{i-1})$

$\exists K, p \geq 0 : \forall i \geq K$

$\underline{s_{i+p} = s_i}$

- We want $u_1, u_2, \dots, u_i, \dots$ looks <u>RANDOM</u>.

LINEAR CONGRUENTIAL GENERATORS (LCG)

$$f(s_i) = (\underbrace{a}_{\text{multiplier}} \cdot s_i + \underbrace{c}_{\text{increment}}) \bmod m \qquad S = \mathbb{Z} \bmod m$$

ADDITIVE LCG

if $c=0$, MULTIPLICATIVE LCG

- A LCG has FULL PERIOD $(p=m)$ iff

1) c and m are RELATIVELY PRIME

2) q is prime and $\underbrace{q \mid m}_{q \text{ divides } m} \Rightarrow q \mid a-1$

3) $4 \mid m \Rightarrow 4 \mid a-1$

Example: $a=5$, $c=3$, $m=2^4$ ✓

$m = 2^{63}$ for 64-bit precision for double.

COMPOSITE LCG

1) $\mathcal{S} = \mathbb{Z}_m^{q+1} = \{0, \dots, m-1\}^{q+1}$. $|\mathcal{S}| = m^{q+1}$

$(s_{i+1}, s_i, \dots, s_{i-q+1}) = f(s_i, \dots, s_{i-q})$ $\boxed{a_0 s_i + \dots + a_q s_{i-q} \bmod m \overset{!!}{=} s_{i+1}}$

2) q different LCG

$s_{1,i}, \dots, s_{q,i}$ are states of q different generators.

$s_i = (\delta_1 s_{1,i} + \dots + \delta_q s_{q,i}) \bmod m$

We can have very large p, e.g. $p = 2^{191} \simeq 3 \cdot 10^{57}$ (rand.c)

$s_{j,i} \in \mathbb{Z}_{m_j}$ $\qquad \mathcal{S} = \prod_j \mathbb{Z}_{m_j}$ $\quad |\mathcal{S}| = m_1 \times \dots \times m_q$

TESTING PRNGs

Consider $\underbrace{x_1, \dots, x_n}$ of SAMPLES

$$\boxed{H_0: \ x_1, \dots, x_n \sim \mathcal{U}(0,1) \text{ and are independent}}$$

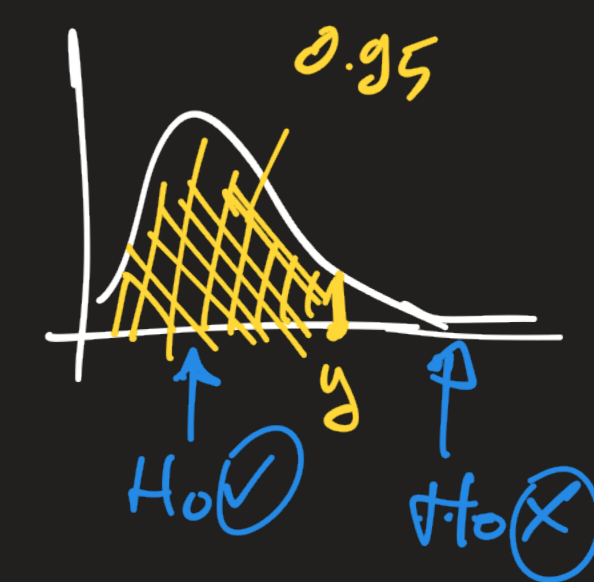
> STRONGER TEST
> BY CONSIDERING BLOCKS OF
> d SAMPLES AND TESTING
> THAT THEY COME FROM
> $\mathcal{U}([0,1]^d)$

STATISTICAL TEST : χ^2-TEST

1) DIVIDE $[0,1]$ in K subintervals of equal length $1/K$. I_j

2) Count: $N_j = \sum_{i=1}^{n} \underbrace{\mathbb{1}(x_i \in I_j)}_{\begin{cases} 1 & \text{if } x_i \in I_j \\ 0 & \text{otherwise} \end{cases}}$

3) CHI2 $= \sum_{j=1}^{K} \left(N_j - \frac{n}{K}\right)^2 \Big/ \frac{n}{K} = \boxed{\frac{K}{n} \cdot \sum_{j=1}^{K} \left(N_j - \frac{n}{K}\right)^2}$



FOR LARGE n, CHI2 IS APPROX. χ^2-distributed with $K-1$ $\left[\underbrace{Y \sim \chi^2(K-1)} \text{ iff } Y = \sum_{j=1}^{K-1} Z_j^2, \ Z_j \sim \mathcal{N}(0,1) \right]$ degrees of freedom

FIND y s.t. $P(Y \leq y) = 0.95$ any confidence level

ACCEPT H_0 iff CHI2 $\leq y$

INVERSION METHOD

X real valued R.V. ($\mathbb{R}$) with p density $p(x)$

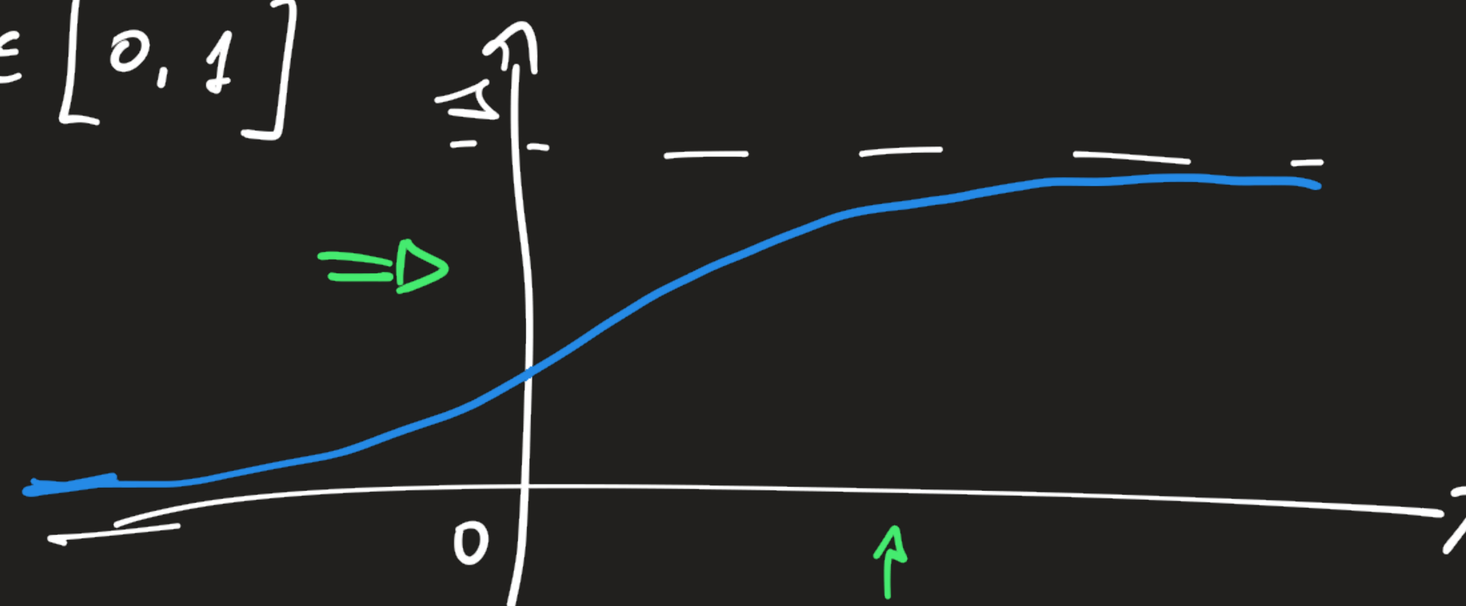
CUMULATIVE DISTRIBUTION FUNCTION $F(x) = P(X \leq x) = \int_{-\infty}^{x} p(y)\,dy$

F is MONOTONE, $F(-\infty) = 0$ $\left(\lim_{x \to -\infty} F(x) = 0\right)$

$F(+\infty) = 1$

X is abs. continuous, then F is continuous and strictly monotonic

$\Rightarrow$ F is invertible, $F^{-1}(u)$. $u \in [0,1]$



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

1) Compute F^{-1}, maybe analytically

2) Sample $u \in U(0,1)$

3) Compute $x = F^{-1}(u)$, then $x \sim X$

$X \sim Exp(\lambda)$. $p(x) = \lambda e^{-\lambda x}$
$x \in [0, +\infty)$

$F(x) = 1 - \exp(-\lambda x)$

$F(x) = U \Leftrightarrow$

$x = -\frac{1}{\lambda} \log(U)$

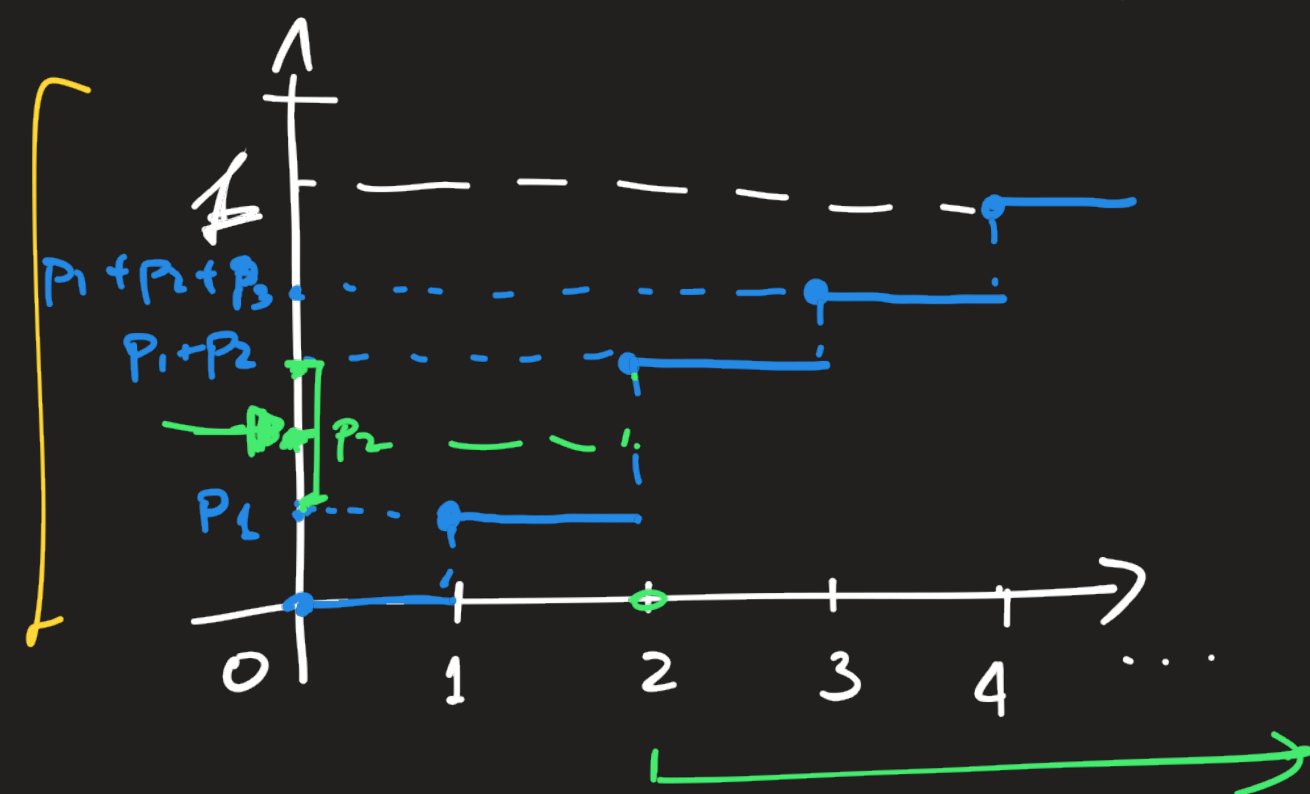
$u \sim U$

$1 - u \sim U$

INVERSION METHOD FOR DISCRETE RANDOM VARIATES

$\mathcal{X} = \{1, \dots, K\}$ $\quad X \sim p : \mathcal{X} \to [0,1] \quad p(i) = p_i \quad \sum_{i=1}^{K} p_i = 1$

$F(x) : P(X \le i) = \sum_{j \le i} p_j$

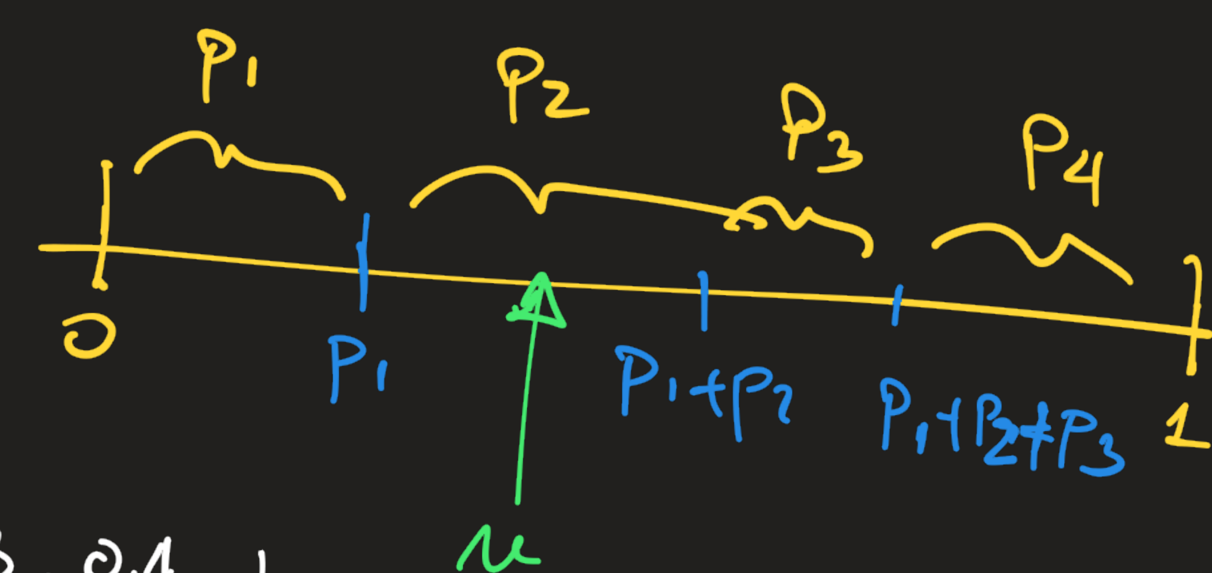


PSEUDO-INVERSE

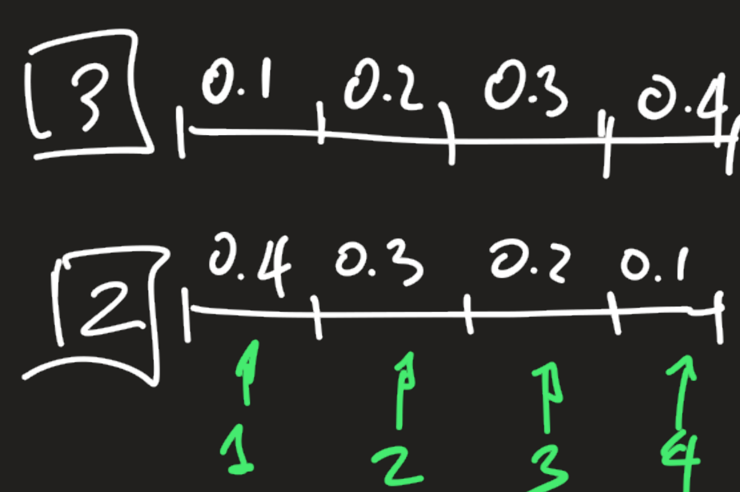
$$F^{-1}(u) = \inf_{x \in \mathbb{R}} F(x) \ge u$$

1) sample $u \sim U(0,1)$

2) $F^{-1}(u)$ and return it.




```
S ← 0
FOR i ← 1 to K
    S += P[i]
    IF u ≤ S
        RETURN i
```

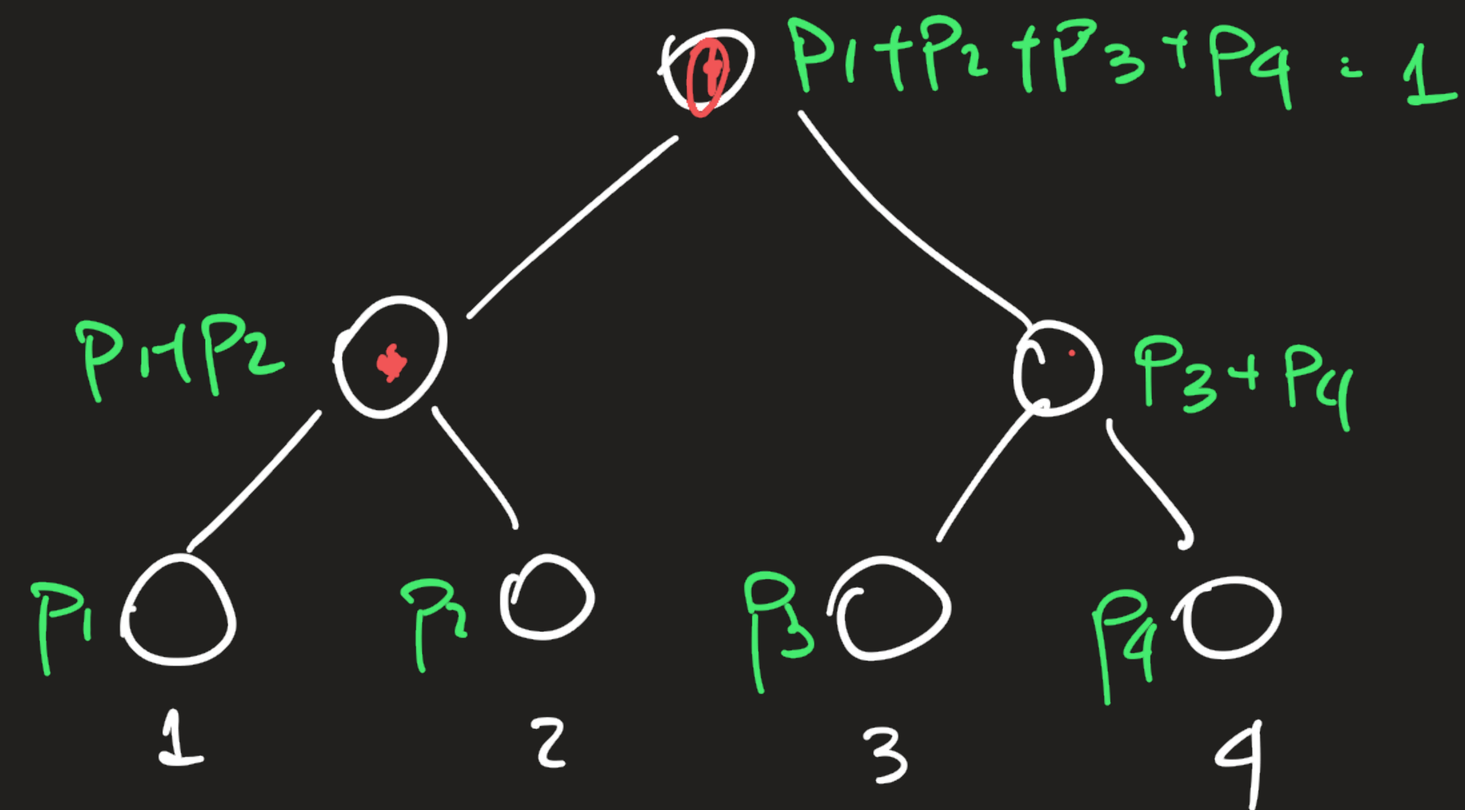


two scenarios. Which is the expected number of comparisons / complexity of the code.

- Alternative speed up, BINARY SEARCH TREE

BINARY SEARCH TREE

$u \in [0,1]$



① IF $u \leq p_1 + p_2$

go left

else

$u = u - (p_1 + p_2)$

go right

$u \leftarrow \text{UNIFORM}(0,1)$

BT-SAMPLE (node, u)

    IF node is a leaf

        Return node.id

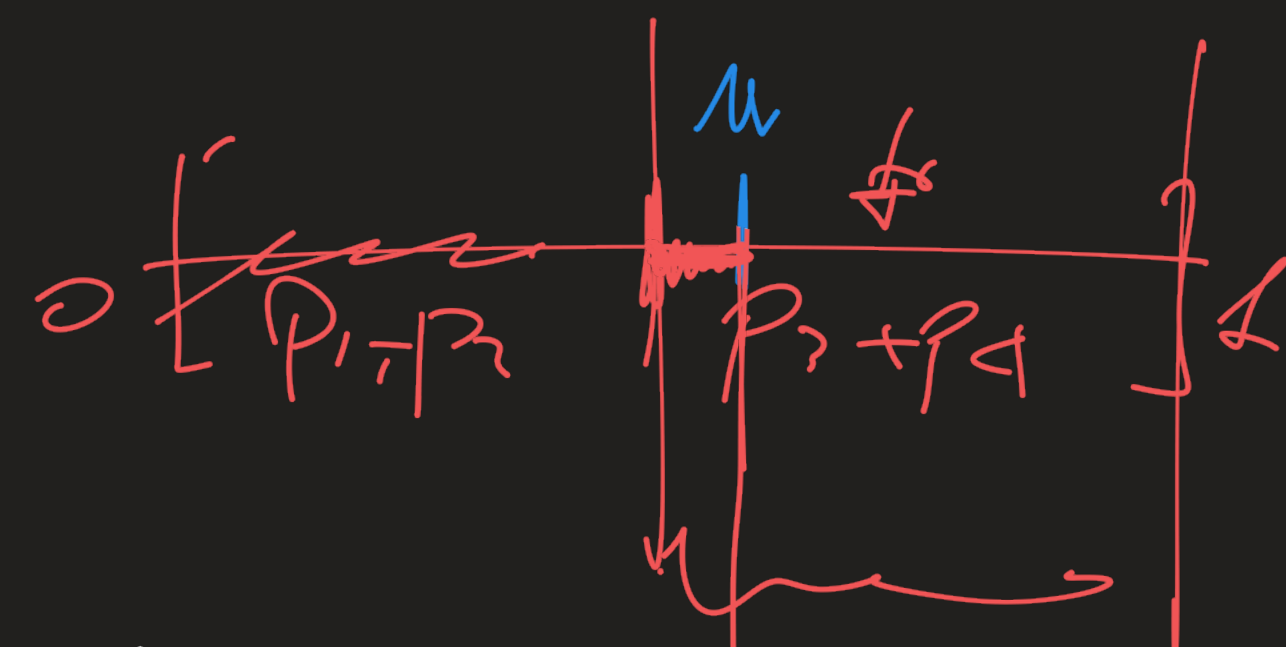
    IF $u \leq$ node.left.value

        BT_SAMPLE (node.left)

    ELSE

        $u = u -$ node.left.value

        BT-SAMPLE (node.right)



$O(\log_2 K)$

BOX MULLER TRANSFORM

$X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x)$ is not analytically tractable

$X, Y \sim \mathcal{N}(0, 1)$ independent

$$p(x,y) = p(x)\,p(y) = \frac{1}{2\pi} \exp\left(-(x^2+y^2)/2\right)$$ / joint density

CHANGE INTO POLAR COORDINATES

$$\begin{cases} R = X^2 + Y^2 \\ \vartheta = \tan^{-1}(Y/X) \end{cases}$$

$(r, \vartheta) \sim (x, y)$ $\quad r = x^2 + y^2$, $\vartheta = \tan^{-1}(y/x)$ $\leftarrow$ $\frac{\partial}{\partial x} \tan^{-1}(x) = \frac{1}{1+x^2}$

$p(x,y) \sim \underline{p(\vartheta, r)} \cdot |\det(J)|$ $\quad J =$ jacobian

$$J = \begin{pmatrix} \frac{\partial \vartheta}{\partial x} & \frac{\partial \vartheta}{\partial y} \\ \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \end{pmatrix} = \begin{pmatrix} -y/(x^2+y^2) & x/(x^2+y^2) \\ 2x & 2y \end{pmatrix}$$

$\det J = -2$

$$P(\vartheta, r) = \frac{1}{4\pi} e^{-r/2} = \frac{1}{2\pi} \cdot \frac{1}{2} e^{-r/2}$$

unif. $[0, 2\pi]$ $\quad$ exponential rate $\frac{1}{2}$

$$p(\vartheta, r) = p(\vartheta)\, p(r)$$

$$p(\vartheta) = \frac{1}{2\pi} \sim \mathrm{Unif}(0, 2\pi)$$

$$p(r) = \frac{1}{2} e^{-r/2} \sim \mathrm{Exp}\left(\frac{1}{2}\right)$$

Sample $u_1, u_2 \sim \mathcal{U}(0,1)$

$$\begin{cases} R = -2 \log u_1 \\ \vartheta = 2\pi \cdot u_2 \end{cases}$$

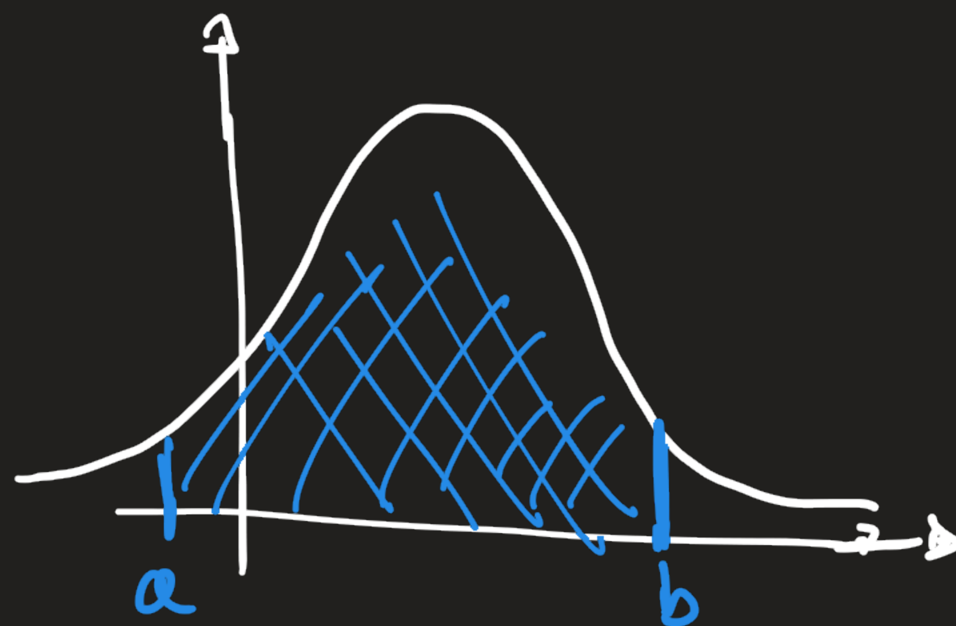
$$\begin{cases} X = \sqrt{R} \cos\vartheta \\ Y = \sqrt{R} \sin\vartheta \end{cases}$$ or Box Muller transform

$X \sim \mathcal{N}(0,1)$ $\qquad$ $X' \sim \underline{\mathcal{N}(\mu, \sigma^2)}$

$$X' = \sigma X + \mu$$

TRUNCATED DISTRIBUTIONS

INVERSION METHOD



$$x = F^{-1}(u)$$

$$F(b) - F(a) = P(X \in [a,b])$$

CONSTRAINT
$x \in [a,b]$

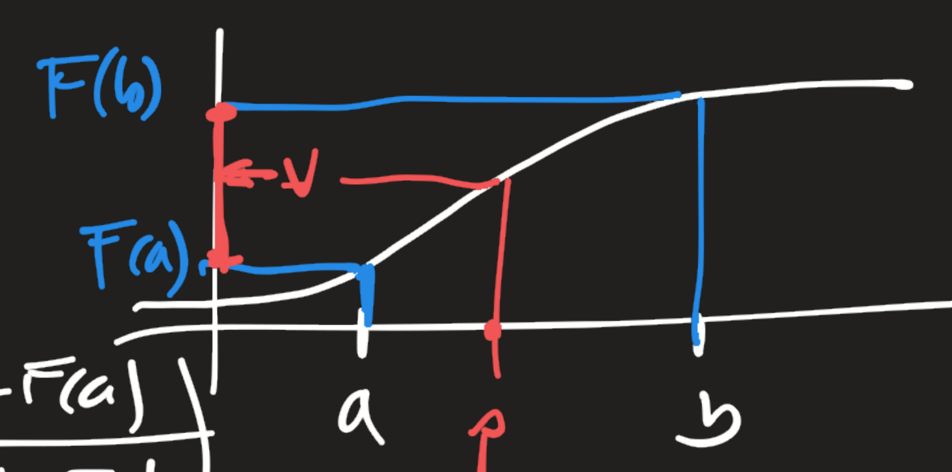
$$p^*(x) = \begin{cases} \dfrac{p(x)}{F(b)-F(a)} & , x \in [a,b] \\ 0 & , \text{otherwise} \end{cases}$$

$$F^*(x) = \begin{cases} \dfrac{F(x) - F(a)}{F(b) - F(a)} & , x \in [a,b] \\ 0 & , x < a \\ 1 & , x > b \end{cases}$$

$$u \sim \text{Unif}(0,1)$$

$$V = F(a) + (F(b) - F(a)) \cdot u$$

RETURN $F^{-1}(V)$



$$P(X \le x) = P(F^{-1}(V) \le x) = P(V \le F(x)) =$$

$$= P(F(a) + (F(b) - F(a))u \le F(x)) = P\left(u \le \frac{F(x)-F(a)}{F(b)-F(a)}\right)$$

$$= P(u \le F^*(x)) = F^*(x)$$

MIXTURE OF R.V.

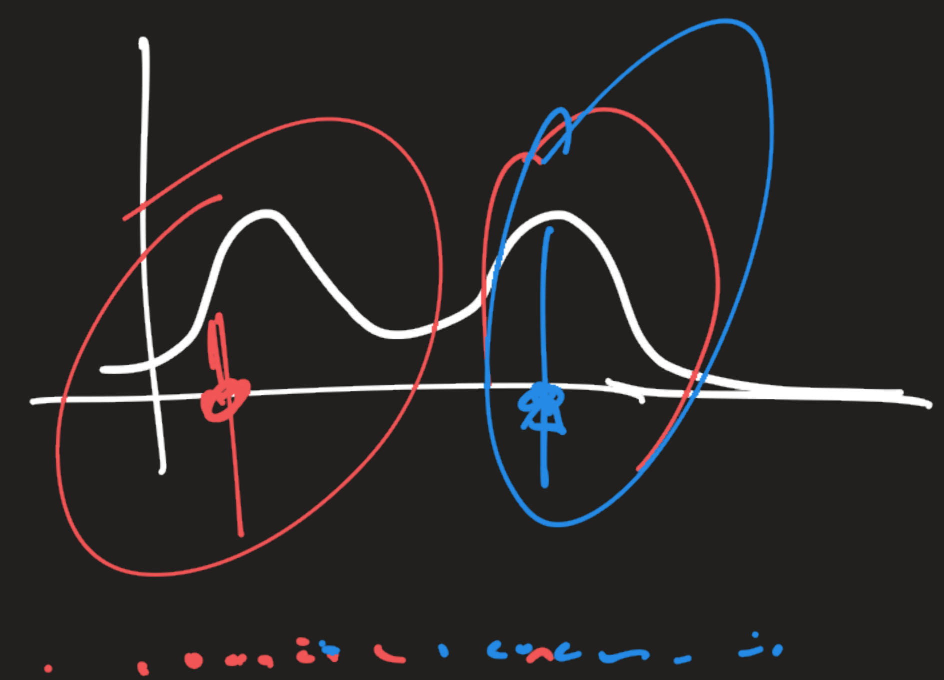
$$\underbrace{p(x) = \sum_{j=1}^{K} \pi_j \, p_j(x)}$$

$\pi_j \in [0,1]$ $\sum_{j=1}^{K} \pi_j = 1$

$p_j(x)$ density and so is $p(x)$

Suppose we can sample from p_j
How to sample from $p(x)$?

$p_j(x) = \mathcal{N}(x \mid \mu_j, \sigma_j^2)$

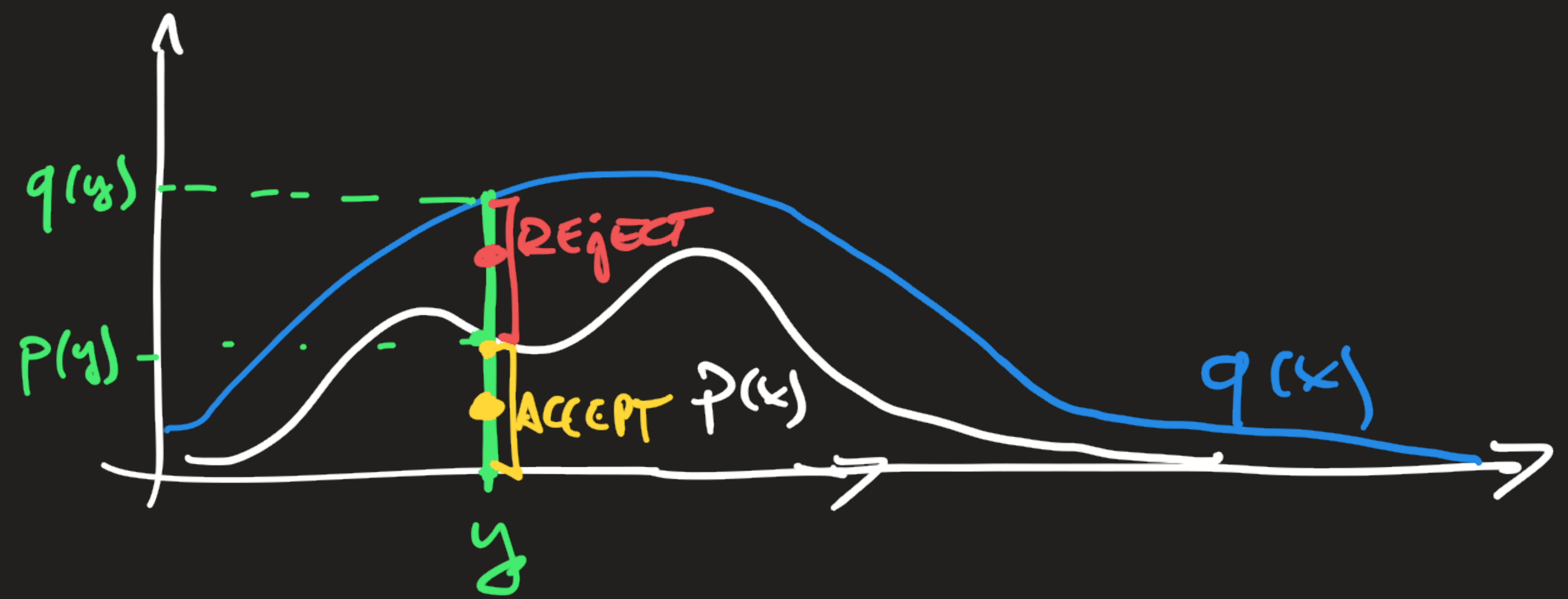


ANCESTRAL SAMPLING.

1) Sample from $\pi_j \to \bar{j}$

2) Sample x from $p_{\bar{j}}(x)$

REJECTION SAMPLING.



$p(x) \leq q(x)$

$$c = \int_{\mathcal{X}} q(x) \geq \int_{\mathcal{X}} p(x) = 1$$

$$c \geq 1$$

$\pi(x) = \dfrac{q(x)}{c}$ is a density, a nice density

1) sample y from $\pi(x)$
2) sample $u \sim U(0,1)$ indep. from y
3) IF $u \leq \dfrac{p(y)}{q(y)}$ RETURN y, OTHERWISE REJECT AND REPEAT

$q(y) \cdot u \leq p(y)$

AVERAGE ACCEPTANCE PROB.

$$\int \underbrace{p(\text{accept} \mid y)}_{p(y)/q(y)} \cdot \pi(y)\, dy = \int \frac{p(y)}{q(y)} \cdot \pi(y)\, dy = \frac{1}{c}$$