class poll: inversion for sampling from unit disc

want to generate (X, Y) uniform over the unit disc, i.e., over $\{(x, y) | x^2 + y^2 \leq 1\}$ given $U, V \sim U[0, 1]$ i.i.d rvs, which of the following gives the correct sample?

(a) $R = U, \Theta = 2\pi V$ and $(X, Y) = (R \cos \Theta, R \sin \Theta)$

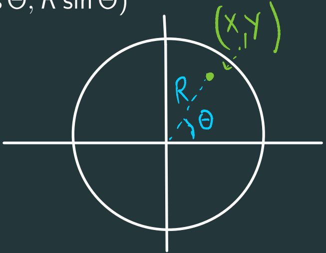
(b) $R = \sqrt{U}, \Theta = 2\pi V$ and $(X, Y) = (R \cos \Theta, R \sin \Theta)$

(c) $R = U^2, \Theta = 2\pi V$ and $(X, Y) = (R \cos \Theta, R \sin \Theta)$

(d) $R = 2U - 1, \underline{\Theta = \pi V}$ and $(X, Y) = (R \cos \Theta, R \sin \Theta)$

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

(e) None of the above



solution: inversion for sampling from unit disc

want to generate (X, Y) uniform over the unit disc, i.e., over $\{(x, y)|x^2 + y^2 \leq 1\}$
given $U, V \sim U[0, 1]$ i.i.d rvs, which of the following gives the correct sample?

$$F_R(r) = \mathbb{P}[R \leq r]$$

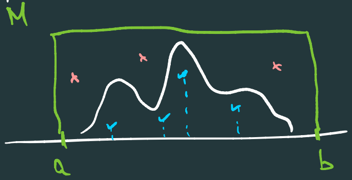
$$= \frac{\text{Area of } \odot \text{ of radius } r}{\text{Area of full } \odot}$$

$$= \frac{\pi r^2}{\pi} = r^2 \text{ for } r \in [0,1]$$

<u>Inversion</u> - Want $R = F^{-1}(U) = \sqrt{U}$

Last lecture

- (inversion review - see recorded lecture)
- Acceptance - Rejection - main 'primitive' for simulation
  - Vanilla AR $\equiv$ bound $f(x)$ in a rectangle

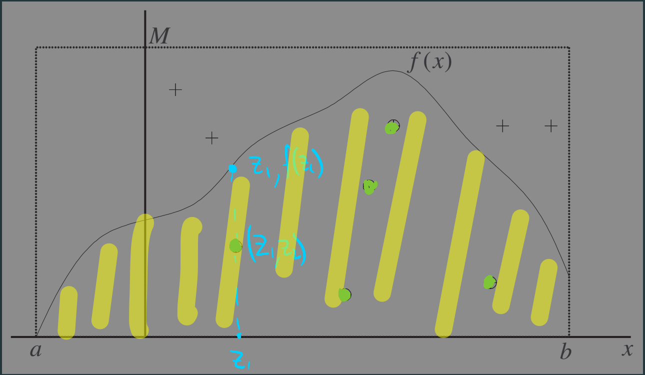


- Sample pts uniformly in rect∠
- Accept if below $f(x)$
- Output x-coordinate

acceptance-rejection

Easy claim - Accepted pts (Z_1, Z_2) are uniformly distributed 'between $y=0$ and $y=f(x)$'

want samples of a rv $X \in [a, b]$, with pdf $f(x) \leq M$



acceptance-rejection sampling

1. generate $U_1, U_2 \sim U[0,1]$
2. set $Z_1 = a + (b-a)U_1$, $Z_2 = MU_2$
3. if $Z_2 \leq f(Z_1)$, return $X_o = Z_1$; else, reject and repeat

$Z_1 \sim \text{Unif}[a,b]$

$Z_2 \sim \text{Unif}[0,M]$

$(Z_1, Z_2) \sim$ Unif in rectangle

AR sampling: running time

how many $U[0,1]$ samples do we need for one sample of X?

Area under $f(z)$

- $\mathbb{P}[\underbrace{(Z_1,Z_2)}_{\text{'trial'}} \text{ accepted}] = \mathbb{P}[Z_2 \leq f(Z_1)] = \dfrac{1}{M(b-a)}$

Area of rectL

Q: Define $N \equiv$ # of $U[0,1]$ samples required to get a single X

$= 2R$, where $R =$ # of 'trials' till we accept a sample

$=$ # of trials till 'success'

$\sim \text{Geom}(1/M(b-a))$

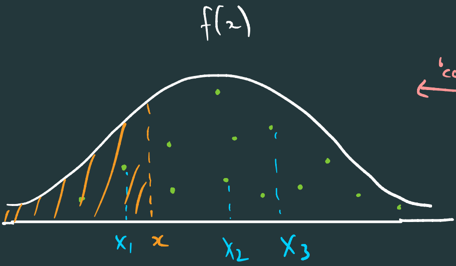
$$\Rightarrow \boxed{N \sim 2 \cdot \text{Geom}(1/M(b-a))}$$

$$\Rightarrow \mathbb{E}[N] = 2 \cdot M(b-a)$$

Note - bigger rectangle $\Rightarrow$ more samples needed

Claim - If you give me pt (X, Y) which is uniformly distributed between $y = 0$ and $y = f(x)$ (for some pdf $f(\cdot)$), then $X \sim F(x) = \int_{-\infty}^{x} f(z)\,dz$

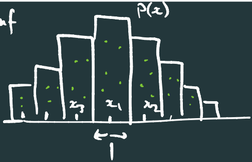
distributed according to CDF F



'calculus'

'Baby proof' - Pmf

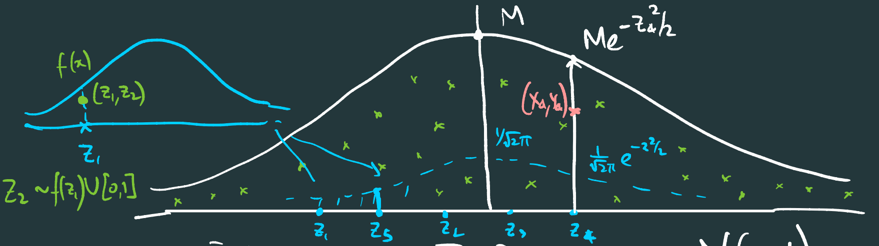
# of pts in rectangle $x \propto p(x)$



Actual proof . $\mathbb{P}[X \leq x] = \dfrac{\text{Area to the left of } x}{\text{Total area}}$

$$\Rightarrow \mathbb{P}[X \leq x] = \frac{\int_{-\infty}^{x} f(z)\,dz}{\int_{-\infty}^{\infty} f(z)\,dz} = F(x)$$

Related question - Consider $g(x) = Me^{-x^2/2}, \ x \in \mathbb{R}$



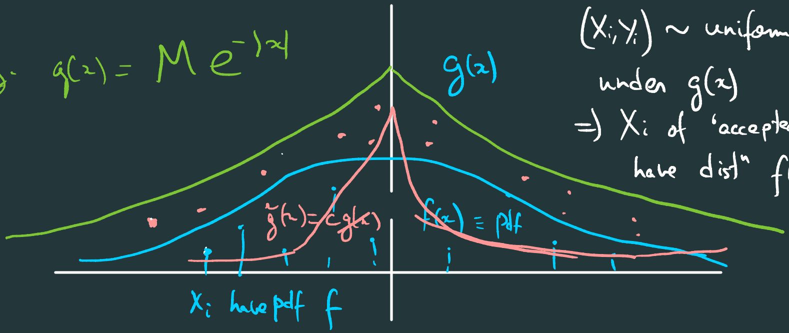
- Suppose I have samples $Z_1, Z_2, \dots \sim N(0,1)$
and $U_1, U_2, \dots \sim U[0,1]$

• Set $\boxed{X_i = Z_i, \quad Y_i = Me^{-Z_i^2/2}\, U_i}$

Then (X_i, Y_i) are uniformly distributed between $y=0$ and $y=g(x)$

Generalized AR

Eg. $g(x) = M e^{-|x|}$



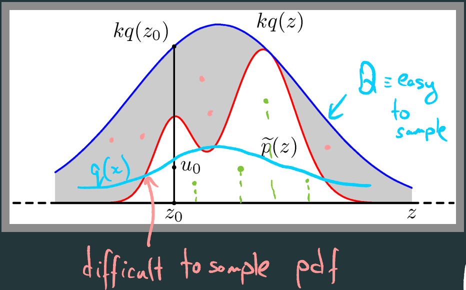
- If we have pts $(X_i, Y_i) \sim$ uniform under $g(x)$
$\Rightarrow$ X_i of 'accepted pts' have dist^n $f(x)$

- How do we get uniform samples under $g(x)$?
- Sample $X_i \sim \dfrac{g(x)}{\int_{-\infty}^{\infty} g(z)\,dz}$, $Y_i \sim g(X_i) \cdot U_i$

find a pdf which has the 'same shape' as $g(x)$ (ie $\tilde{g}(x) = c\,g(x)$)

in example, $\tilde{g}(z) = \frac{1}{2} e^{-|z|}$ 'Laplace dist^n' ie, 2-sided exponential

generalized AR sampling



General process

- Sample $X_i \sim Q$ ← easy to sample distn
- Find some k such that

$$\frac{kq(x)}{\tilde{p}(x)} \geq 1 \text{ for all } x$$

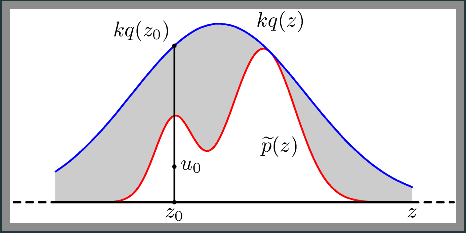
(ie, $kq(x)$ 'covers' $\tilde{p}(x)$)

(ie, set $k \geq \max_x \tilde{p}(x)/q(x)$)

- Sample $Y_i \sim kq(X_i) \cdot U_i \Rightarrow (X_i, Y_i) \sim$ Unif under $kq(x)$
- Accept X_i if $Y_i \leq \tilde{p}(X_i) \rightarrow kq(X_i)\, U_i \leq \tilde{p}(X_i)$

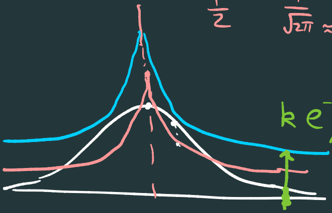
ie, accept X_i with probability $\tilde{p}(X_i) / kq(X_i)$

generalized AR sampling



$$k \geqslant \max_x \frac{\tilde{p}(x)}{q(x)}$$

$\frac{1}{2}$ $\quad$ $\frac{1}{\sqrt{2\pi}} \approx 6.2$



$k e^{-z}/2$

$E_1 \sim Lap(1)$

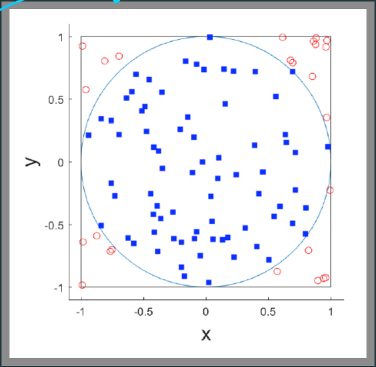
$\frac{1}{2} e^{-z}$

(even more) generalized AR sampling

Given set A which is easy to sample from, and set $B \subseteq A$ which is hard to sample from

given pts $X_1, X_2, \ldots X_n \sim \text{Unif}(A)$

X_i in this example is in $[-1,1]^2$



Acceptance-Rejection Principle

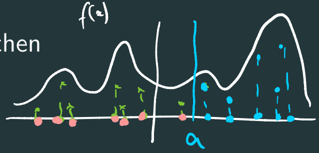
$$\text{dist}^n \text{ of } X_i \mid X_i \in B \equiv \text{Unif over } B$$

unif samples from A

have fn which tests membership in B

class poll: ordering conditional expectations

consider rv $X \in \mathbb{R}$ with cdf F, and any $a \in \mathbb{R}$; then



(a) $\mathbb{E}[X] \geq \mathbb{E}[X|X \geq a]$

✓ (b) $\mathbb{E}[X] \leq \mathbb{E}[X|X \geq a]$

(c) depends on if a is positive or negative

(d) depends both on a and F

- AR way of thinking

$X_1, X_2, X_3 \ldots X_n \sim F$

$(\mathbb{E}[X] \overset{n\to\infty}{\approx} \frac{1}{n}\sum X_i)$

$X_1, X_2, \ldots, X_4 \quad X_{n-1} \ldots$

$Y_1 \quad Y_2 \quad Y_3 \quad Y_m \ldots Y_n$

via AR principle

$$\mathbb{E}[X|X \geq a] = \mathbb{E}[Y] \overset{n\to\infty}{\approx} \frac{1}{n}\sum Y_i$$