Transformation of random variables

- Let $Y = f(x)$
- Two different, but related questions:

  1) If pdf for y is $U(0,1)$, what is the pdf $P_x(x)$ for x?

  2) If we can generate samples Y_i from $U(0,1)$, how can we use this to generate samples X_i from some pdf $P_x(x)$?

- We focus on 2)

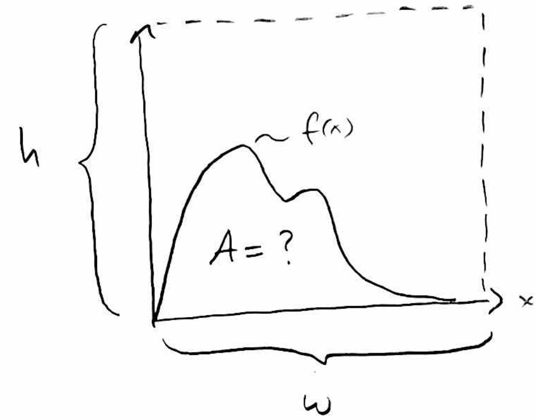
- Look at two methods:

  1) Rejection sampling

  2) Inverse transform sampling

Rejection sampling ("measure area by throwing darts")

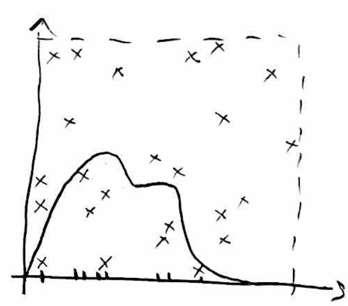
- Class exercise!
  (no, not really!)
- Everyone gets one piece of chalk
- You are not allowed to leave your seat



- Q: How would you estimate area A for the figure on the blackboard?

- A: - Everyone throws chalk at blackboard
  - Assume hit coordinates (x,y) are approximately given by $x \sim U(0,w)$
    $y \sim U(0,h)$
  - Count number of hits inside A: N_A

$$A \simeq \left(\frac{N_A}{N}\right) w h$$

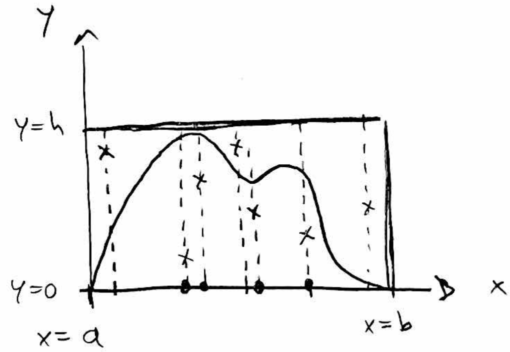


$\Rightarrow$ Primitive MC integration of function $f(x)$

$\Rightarrow$ Also sampling of x values according to a pdf $p_x(x) \propto f(x)$!

- Rejection sampling of x-samples $\sim p_x(x)$:
  - Bound $p_x(x)$ in a box: $x \in [a, b]$, $y \in [0, h]$

Will be generalized beyond box shape



Algo:

1) Sample $x' \sim U(a, b)$
2) Sample $y' \sim U(0, h)$
3) <u>If</u> $y' < p_x(x')$, <u>accept</u> x' as a new x sample

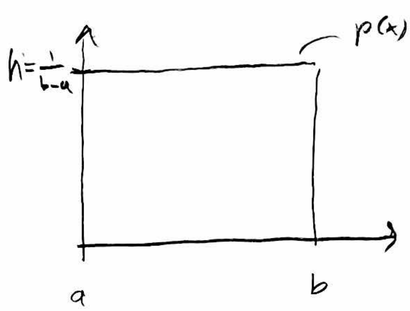
Repeat for as many samples you need

[Cont. from here.]

- Works with any arbitrary pdf shape
  (Also, doesn't have to be normalized)

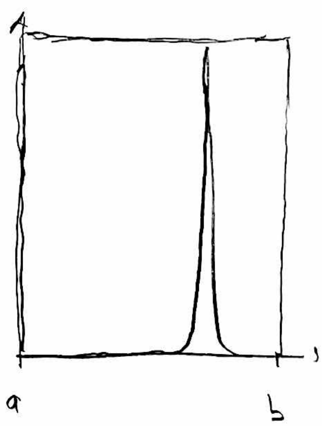
Q: In what cases will this approach be very inefficient?

- Naive method less efficient for peaked pdfs



decreasing efficiency

$\longrightarrow$



[$p_x(x) = U(a,b)$
$\Rightarrow$ 100% acceptance
but completely pointless]

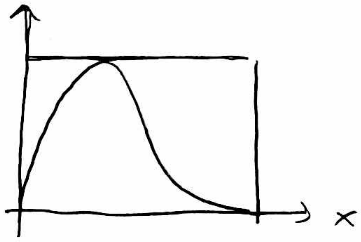
Prob (accept) = 1

[$p_x(x)$ very peaked
$\Rightarrow$ low acceptance]

Prob (accept) ≈ 0

- Problem w/ high dimensions

- Say we have this in 1D :



$\text{Prob(accept)} = 0.5$

- Now we want to sample a sample $\bar{x}$ from the joint pdf $p_{\bar{x}}(\bar{x}) = p_{\bar{x}}(x_1, x_2, x_3, \ldots)$

  where each component x_i has a pdf like this

- Will this work ?

- Will it work <u>well</u> ?

- A : It will <u>work</u> to do rejection sampling, but the efficiency drops quickly with increasing dimensionality

$$\text{Prob(accept } \bar{x}) = \text{Prob}((\bar{x}, y) \text{ within } p_{\bar{x}}(\bar{x}))$$

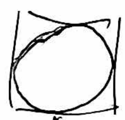
$$= \text{Prob(accept } x_1) \times \text{Prob(accept } x_2) \times \ldots$$

$$= 0.5 \times 0.5 \times \ldots$$

$$= (0.5)^n$$

<u>Ex</u> : $n = 10 \Rightarrow \text{Prob (accept } \bar{x}) \lesssim 0.001 \approx 0.1\%$

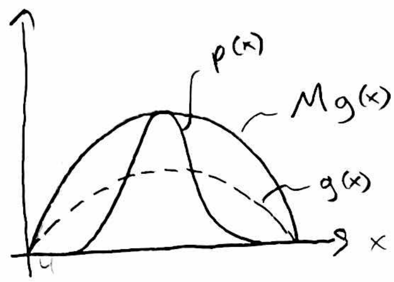
$\Rightarrow$ will need around 1000 attempts to get one sample $\bar{x}$ in 10 D <u>for this pdf</u>



<u>Ex</u> : $\left.\begin{array}{l} \text{Prob(accept } x_i) = 0.8 \\ n = 20 \end{array}\right\} \text{Prob(accept } \bar{x}) \approx 0.011 \approx 1\%$

- In general : Sampling a high-dim box $\rightarrow$ almost always near an edge !

- Improvement with importance sampling



1) Sample $x' \sim g(x)$ (not $x \sim U(a,b)$)

2) Sample $y' \sim U(0, Mg(x'))$	2) Sample $y' \sim U(0,1)$
3) If $y' < p(x')$, accept x' as new x sample	3) If $y' < \frac{p(x')}{Mg(x')}$, accept x' as new x sample

- Compared to "box method", this more rarely samples x regions with low $p_x(x)$, but compensates by increasing the accept. prob. for such samples

- The more similar $g(x)$ is to $p(x)$, the better the efficiency

[But if we could sample $x \sim g(x) = p(x)$ we wouldn't need this method anyway...]

Inverse transform sampling

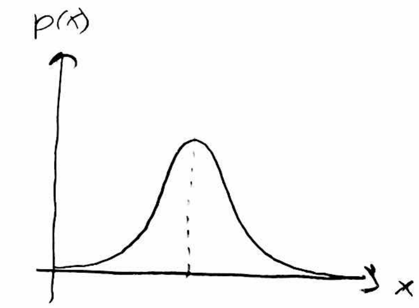
- Recap: The cumulative distr. function (cdf) is defined by

$$F(x) = \text{Prob}(X \leq x) = \int_{-\infty}^{x} p(x')\, dx'$$

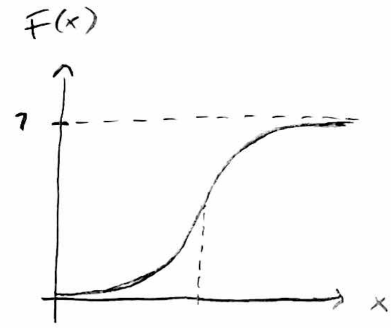
End lect. Nov 12

- Example:

pdf:

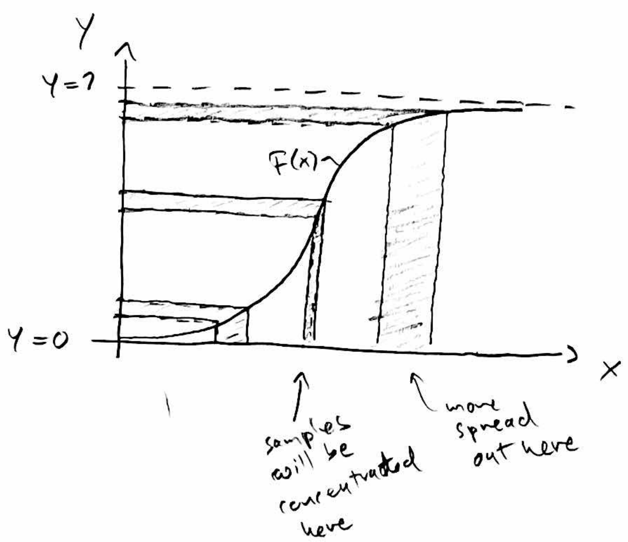


cdf:



$$\boxed{\frac{dF}{dx} = p(x)}$$

- Idea for inverse transform sampling:
  - Assume we know $y = F(x)$ and can find (or evaluate) inverse function F^{-1}
  - Can use this to generate x samples from $p(x)$

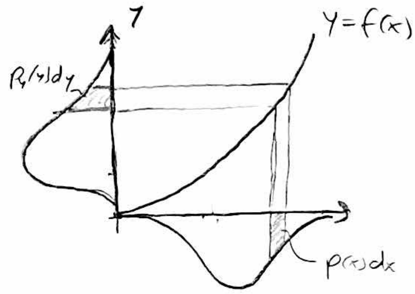


Algo:

1) Sample y from $U(0,1)$
2) Compute $x = F^{-1}(y)$

Repeat for as many samples as needed

- Explanation:
  - In general, if $x \sim p_x(x)$
    and y is some function $y = f(x)$
    we can find $p_y(y)$ from requirement



[Conservation of probability] $\longrightarrow$ $p_y(y)\,dy = p_x(x)\,dx$

  - Now look at special case
    where $y = F(x)$ = cdf for x

- If $x \sim p_x(x)$ and $y = F(x)$, what is $p_y(y)$?

$$p_x(x)\,dx = p_y(y)\,dy$$

$$= p_y(y)\,\frac{dy}{dx}\,dx$$

$$= p_y(y)\,\frac{dF}{dx}\,dx$$

$$\underline{p_x(x)\,dx = p_y(y)\,p_x(x)\,dx}$$

$$\Rightarrow \boxed{p_y(y) = 1 = U(0,1)}$$

- So if we sample $y \sim U(0,1)$
  and compute $x = F^{-1}(y)$
  we get an $x \sim p_x(x)$, which was our goal

- Pro: No rejection sampling needed (efficiency)
- Con: Requires knowledge of <u>inverse cdf</u>

- Similar structure to <random>
  1) Use a <u>RNG</u> to sample from $U(0,1)$
  2) Use a <u>transformation</u> to obtain sample from $p(x)$

- Q: Given a set of samples from some unknown pdf. How to generate more such samples?

  - Rejection sampling from histogram rather than pdf
  - Inverse transf. sampling using cumulative histogram in place of cdf