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)

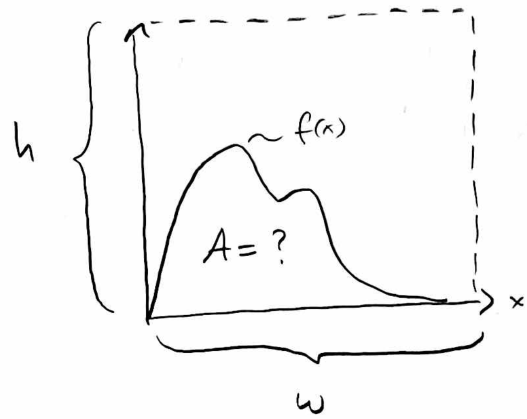
- Look at two methods:

    1) Rejection sampling

    2) Inverse transform sampling

Rejection sampling ("measure area by throwing darts")

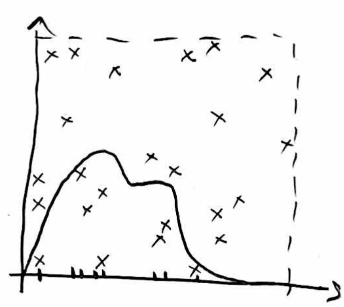
- Class exercise! (no, not really!)
- Everyone gets one piece of chalk
- You are not allowed to leave your seat



- 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

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



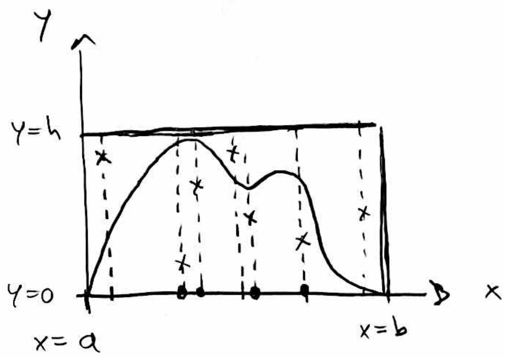
$\Rightarrow$ Primitive MC integration of function $f(x)$

$\Rightarrow$ 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



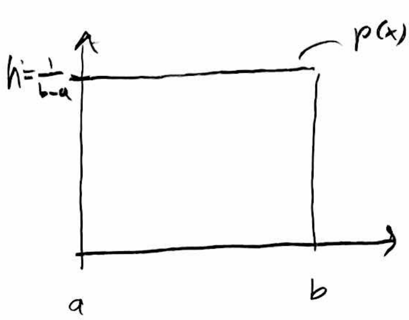
Algo :

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

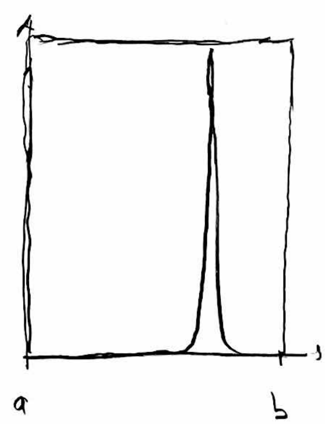
Repeat for as many samples you need

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

- Naive method less efficient for peaked pdfs



decreasing efficiency $\longrightarrow$



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

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

- Say we have this in 1D :

Prob(accept) = 0,5

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

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

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

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

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

$$= (0.5)^n$$

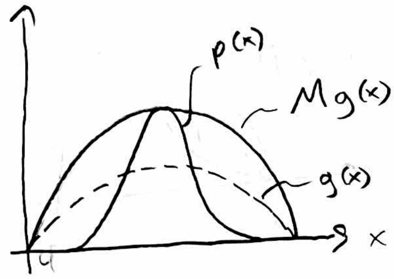
<u>Ex</u>: $n = 10 \Rightarrow \text{Prob}(\text{accept } \bar{x}) \lessapprox 0.001 = 0.1\%$

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

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

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

- Improvement with <u>importance sampling</u>



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...]