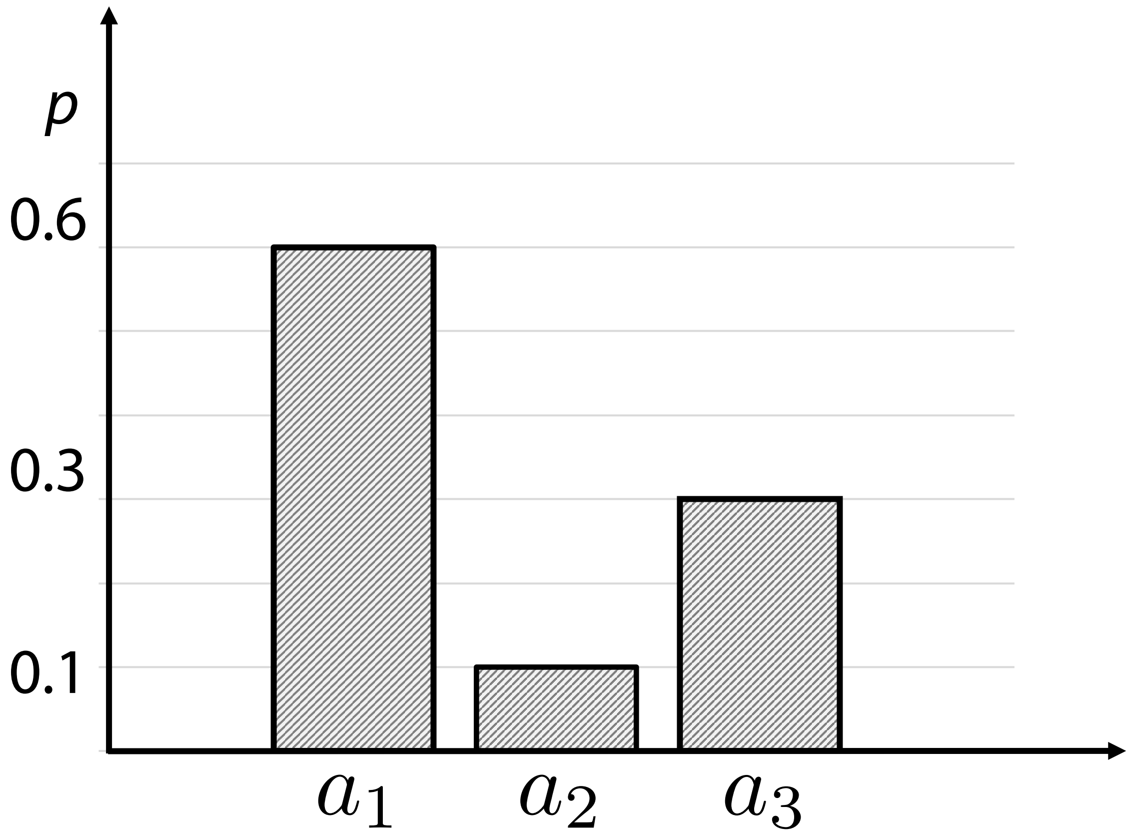
Sampling from 1d distributions

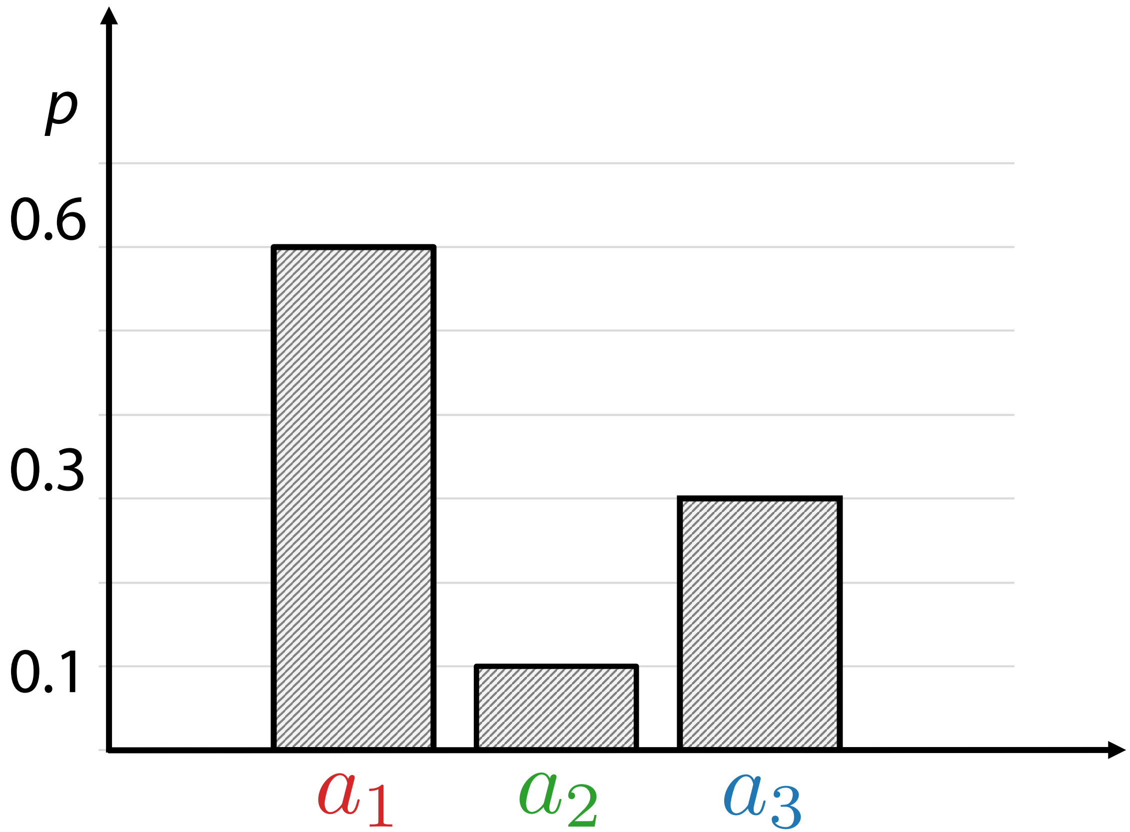
1d sampling (discrete)



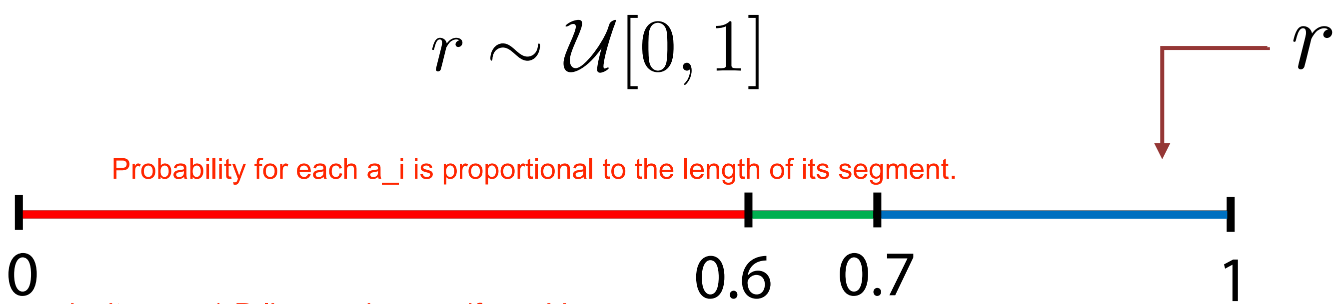
We can always sample from uniform $\mathcal{U}[0,1]$

Uniform is quite easy to emulate.

1d sampling (discrete)



$$r \sim \mathcal{U}[0,1]$$



make it some 1-D line, and use uniform. Very easy.

Summary

1d discrete distributions with finite number of values are easy

At least then number of values is < 100 000

If we have 1 million discrete classes (values here mean class), then this will be hard.
We have to use techniques for continuous distributions.

Continuous sampling

1d sampling (continuous)

Use central limit theorem. The sum of the independent x_i variables will eventually become gaussian

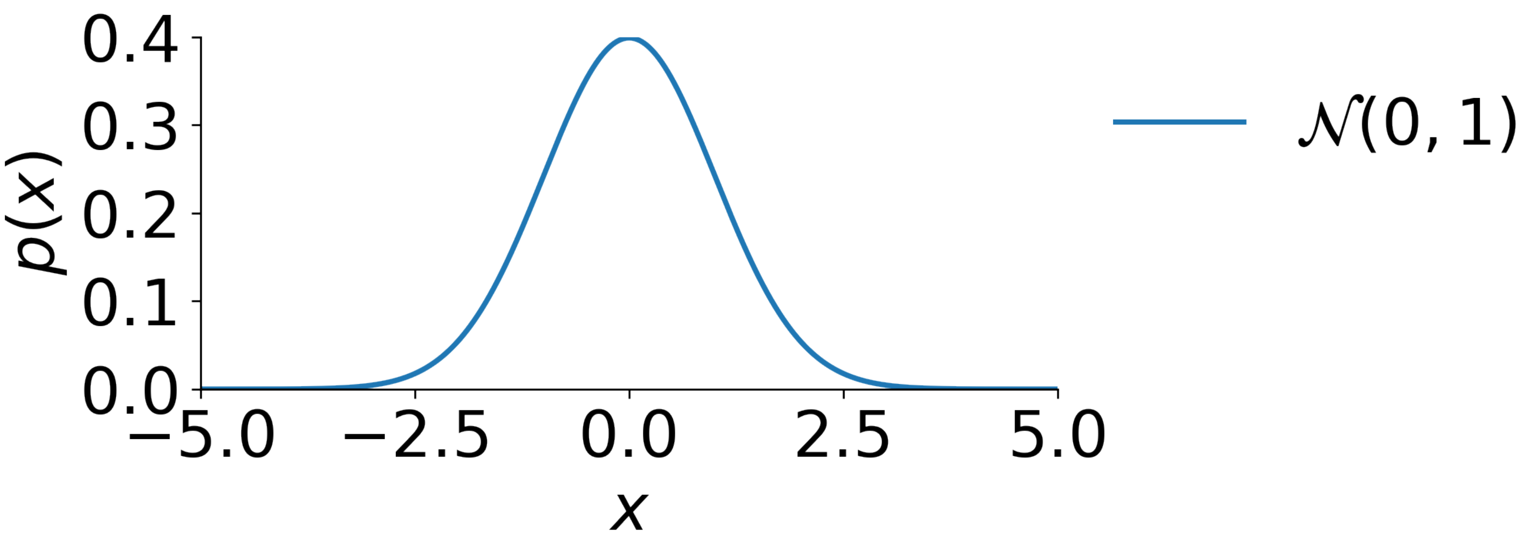
Sampling from Gaussian distribution

$$z = \sum_{i=1}^{12} x_i - 6, \quad x_i \sim \mathcal{U}[0, 1]$$

x_i is uniform.
Expectation of each x_i is 0.5, so 0.5 x 12 = 6.
Subtract 6 to make E(z) = 0.

However, we can't get values outside the range [-6,6].
This can happen in a gaussian (with small probability).

$$p(z) \approx \mathcal{N}(0, 1)$$

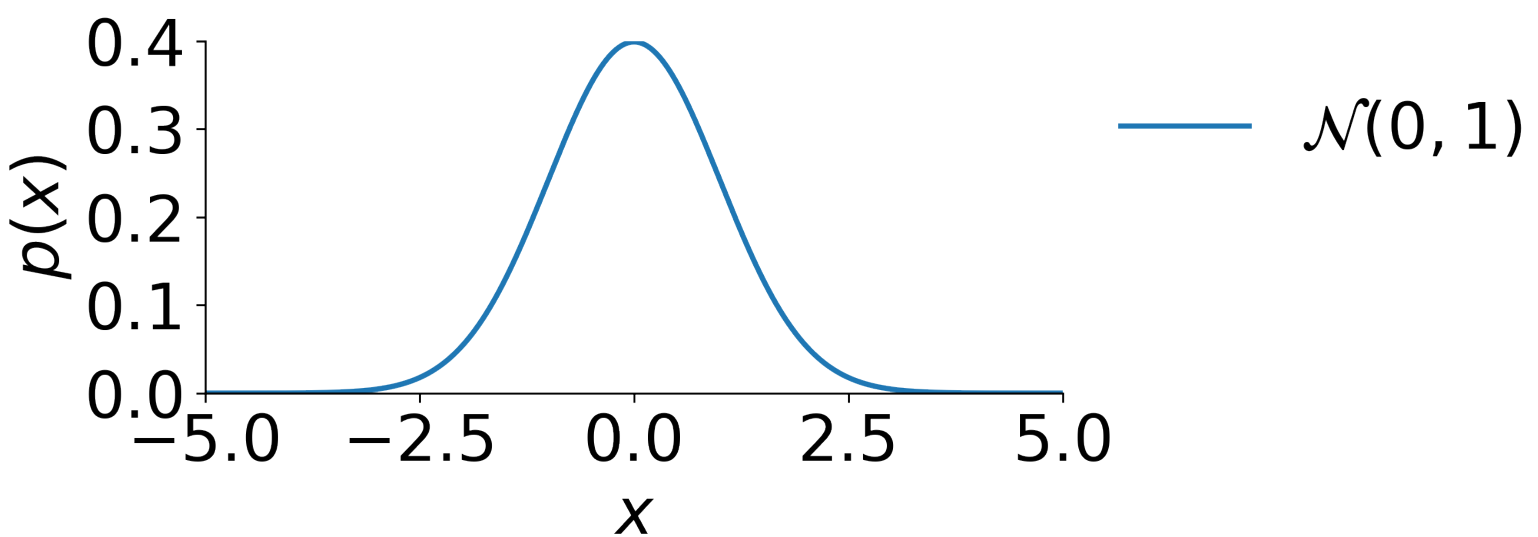


1d sampling (continuous)

Sampling from Gaussian distribution

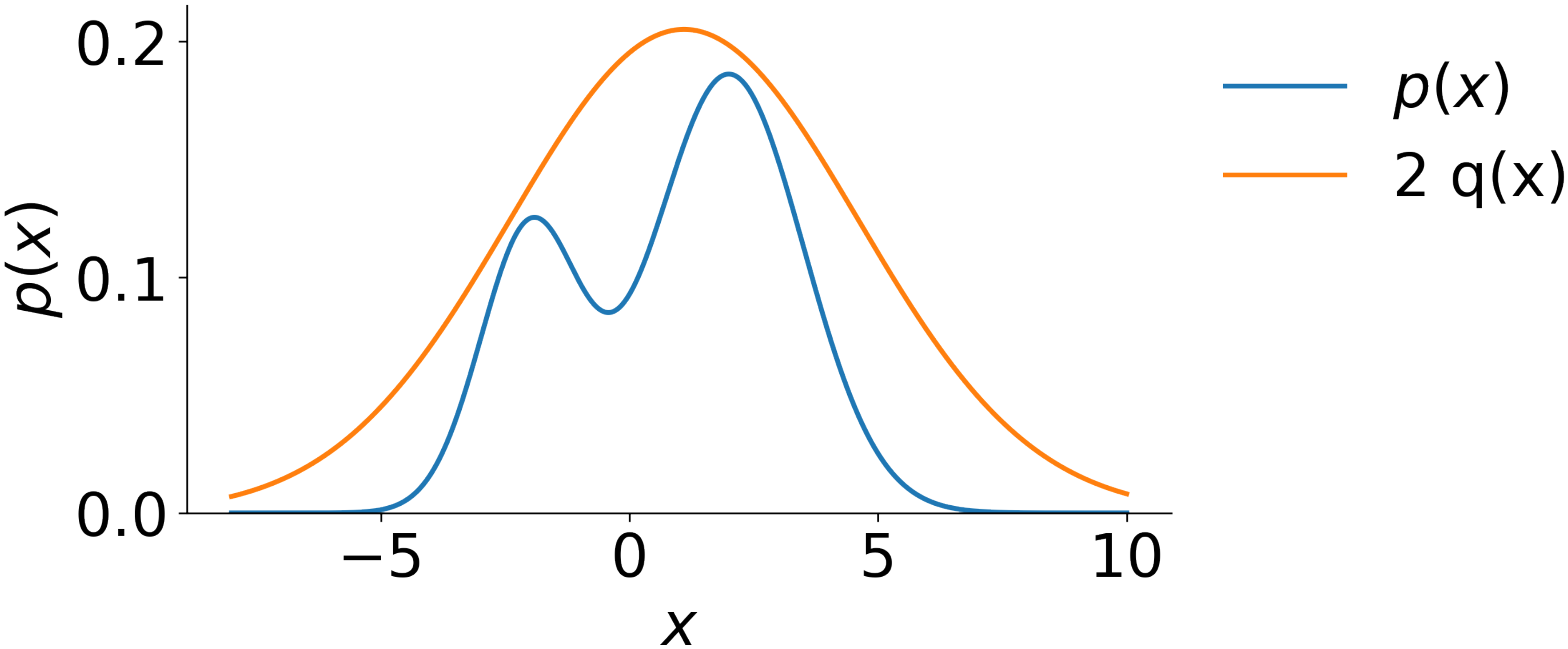
Or call library function ☺

z = numpy.random.randn()



1d sampling (continuous)

How do we sample from p(x)?



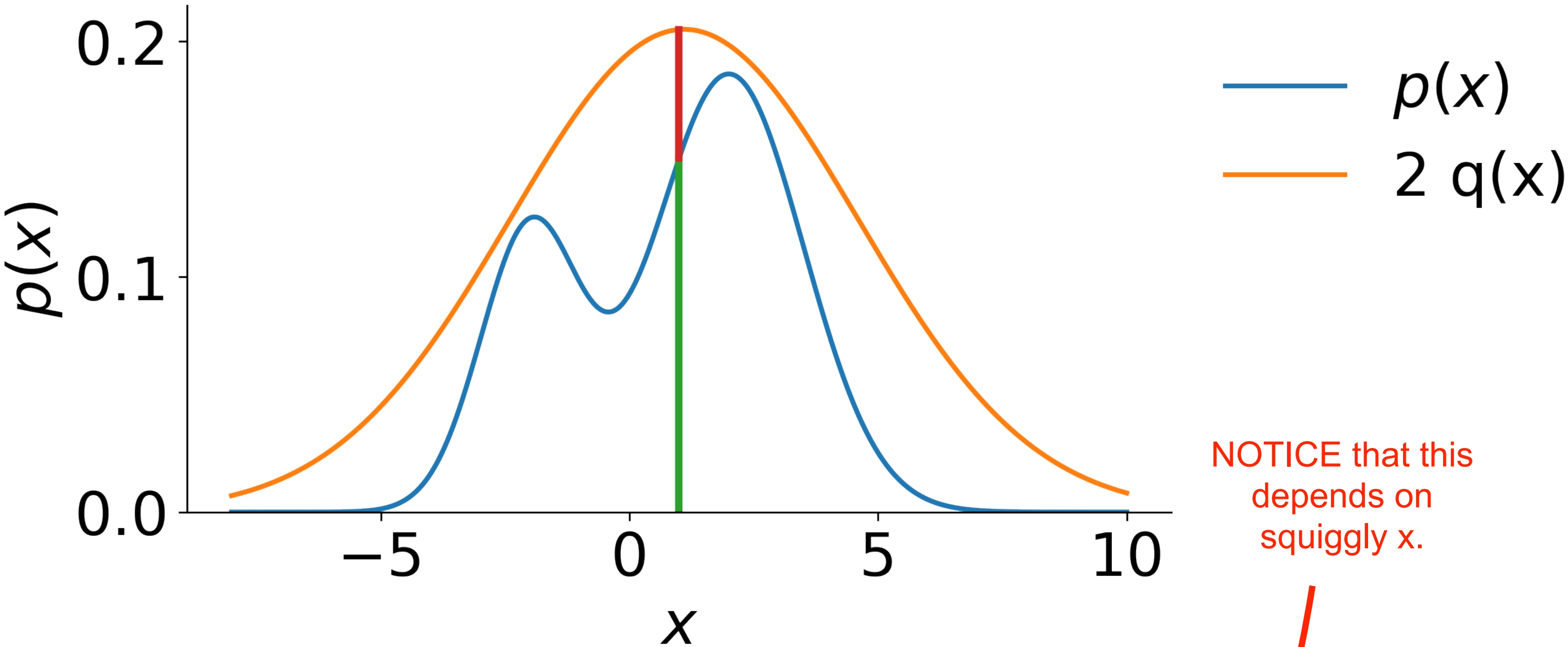
Let's upper bound our distribution with some gaussian times so constant.

$$q(x) = \mathcal{N}(1, 3^2)$$

$$p(x) \leq 2q(x)$$

Why?
Because we know how to sample from gaussians.

1d sampling (continuous)



NOTICE that this depends on squiggly x.

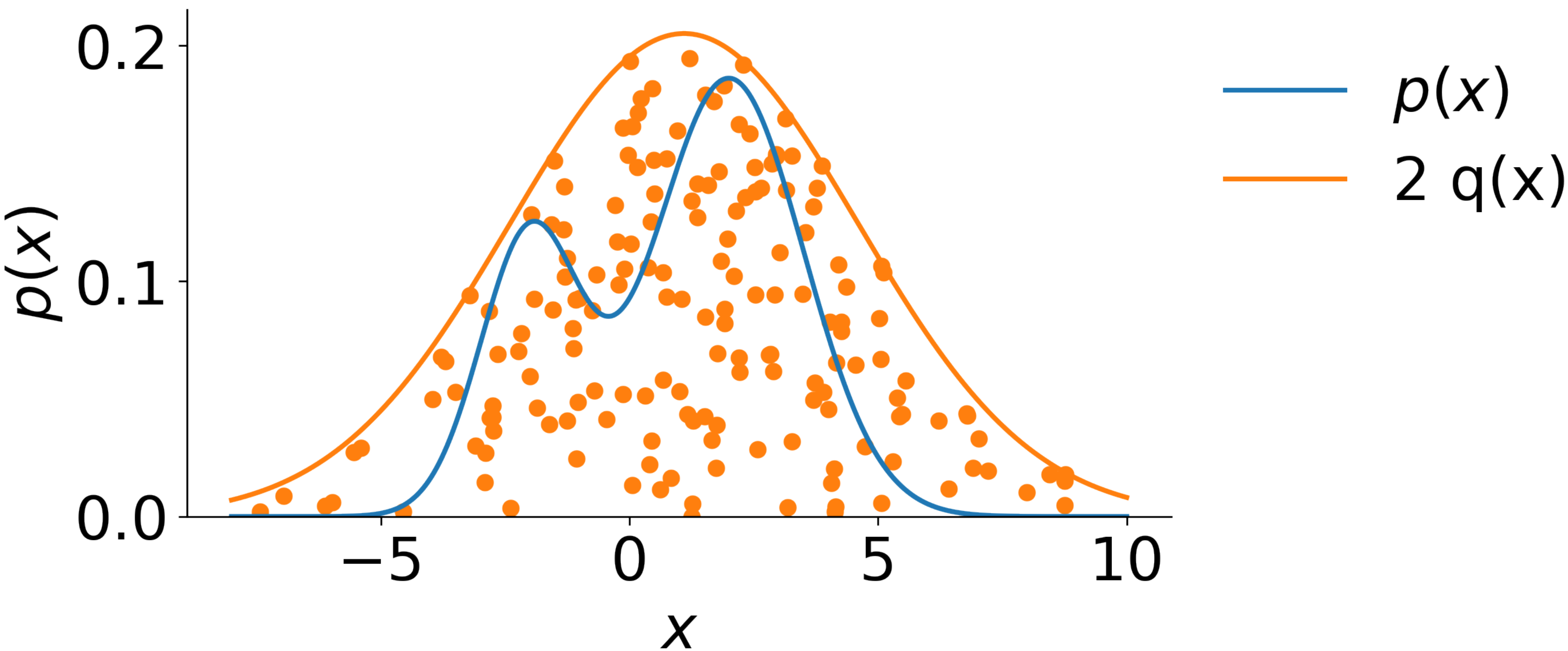
$$\widetilde{x} \sim q(x) \qquad\qquad y \sim \mathcal{U}[0, 2q(\widetilde{x})]$$

Aha. If their heights are the same, accept 100%!

Accept $\widetilde{x}$ with probability $\dfrac{p(x)}{2q(x)}$: if $y \leq p(x)$

if y is greater than p(x), then don't accept at all.

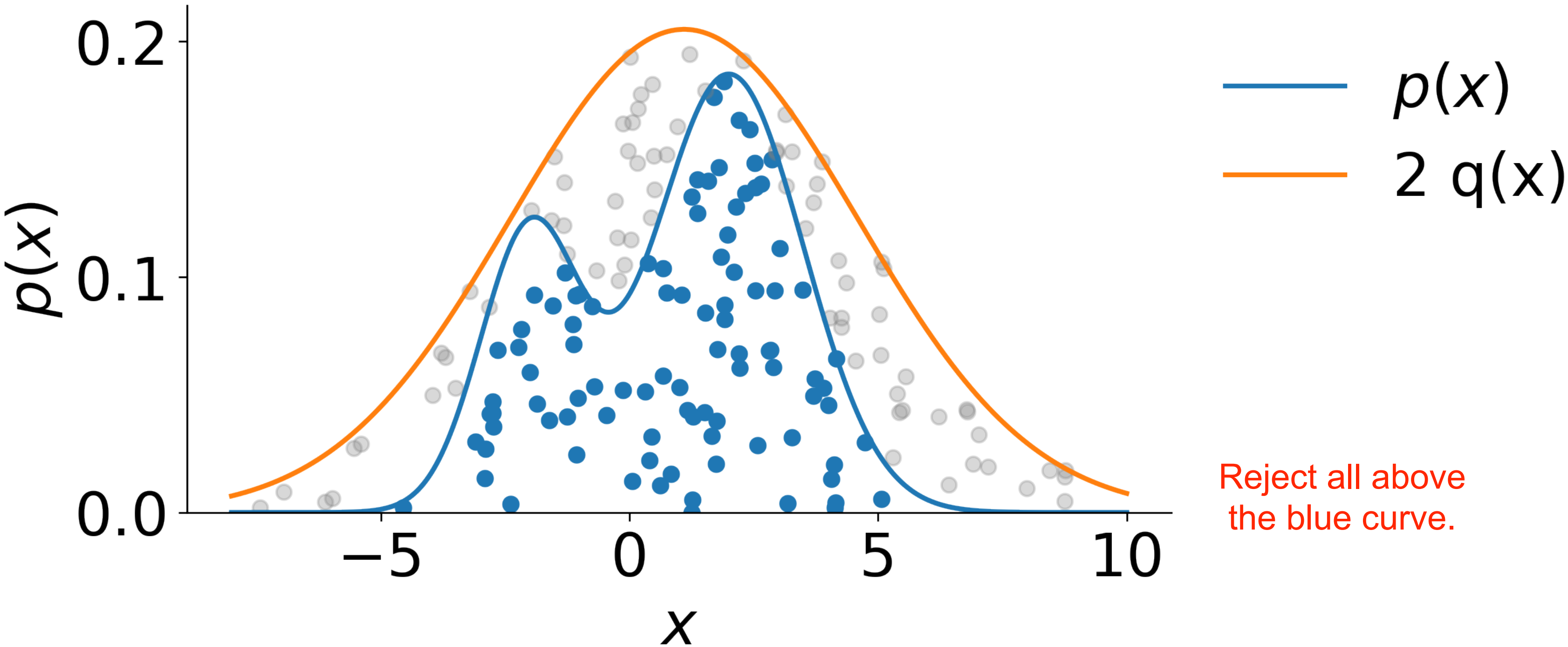
1d sampling (continuous)



$$\widetilde{x} \sim q(x) \qquad\qquad y \sim \mathcal{U}[0, 2q(\widetilde{x})]$$

Accept $\widetilde{x}$ with probability $\dfrac{p(x)}{2q(x)}$: if $y \leq p(x)$

1d sampling (continuous)

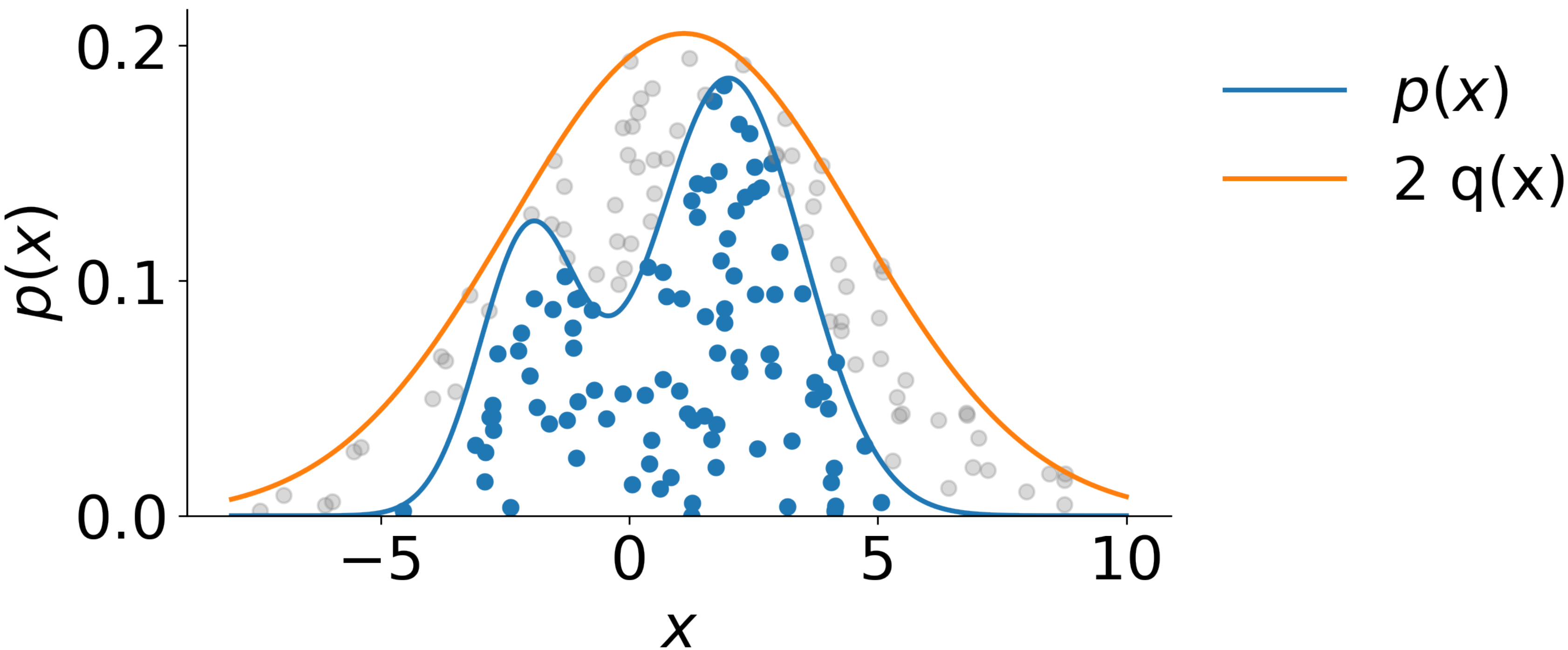


Reject all above the blue curve.

$$\widetilde{x} \sim q(x) \qquad\qquad y \sim \mathcal{U}[0, 2q(\widetilde{x})]$$

Accept $\widetilde{x}$ with probability $\dfrac{p(x)}{2q(x)}$: if $y \leq p(x)$

1d sampling (continuous)



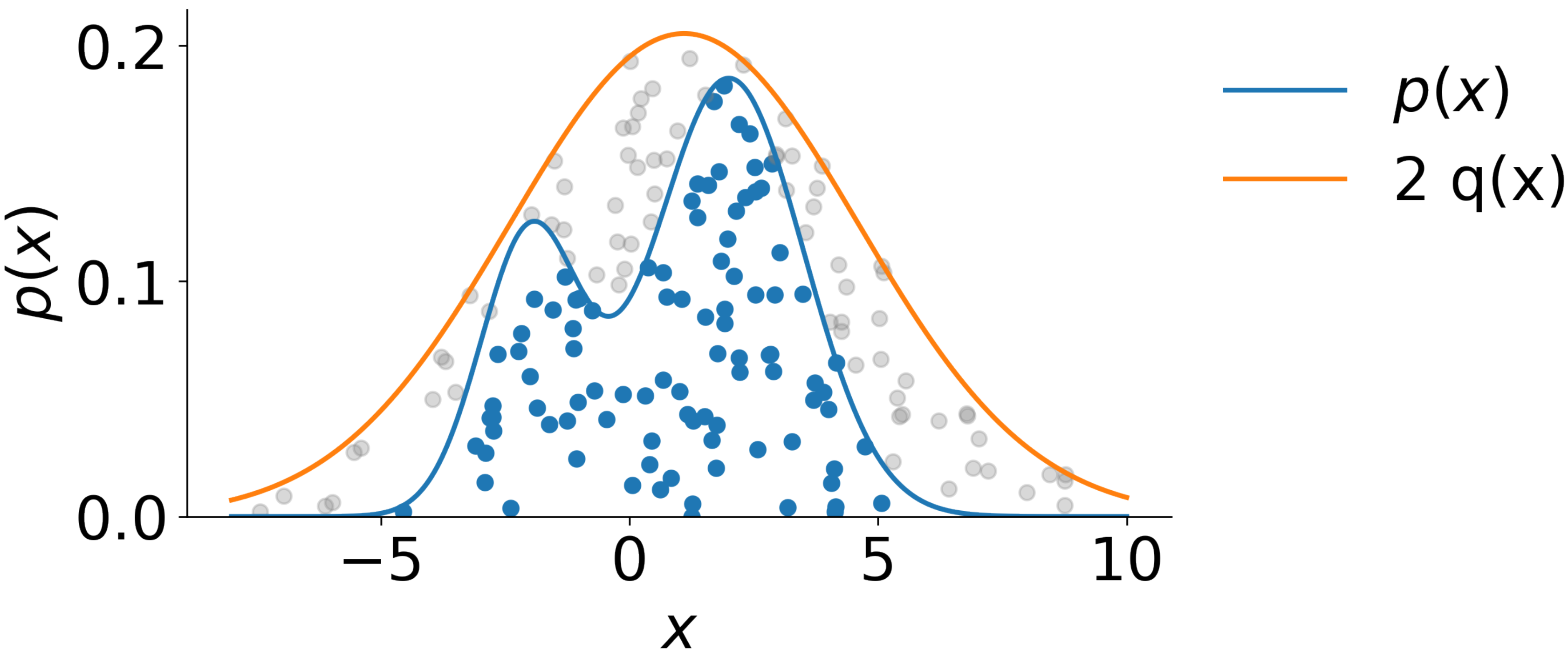
M is the constant.

Presumably, q(x) will converge to p(x) so that p(x) / Mq(x) \rightarrow 1 / M.

$$p(x) \leq Mq(x)$$

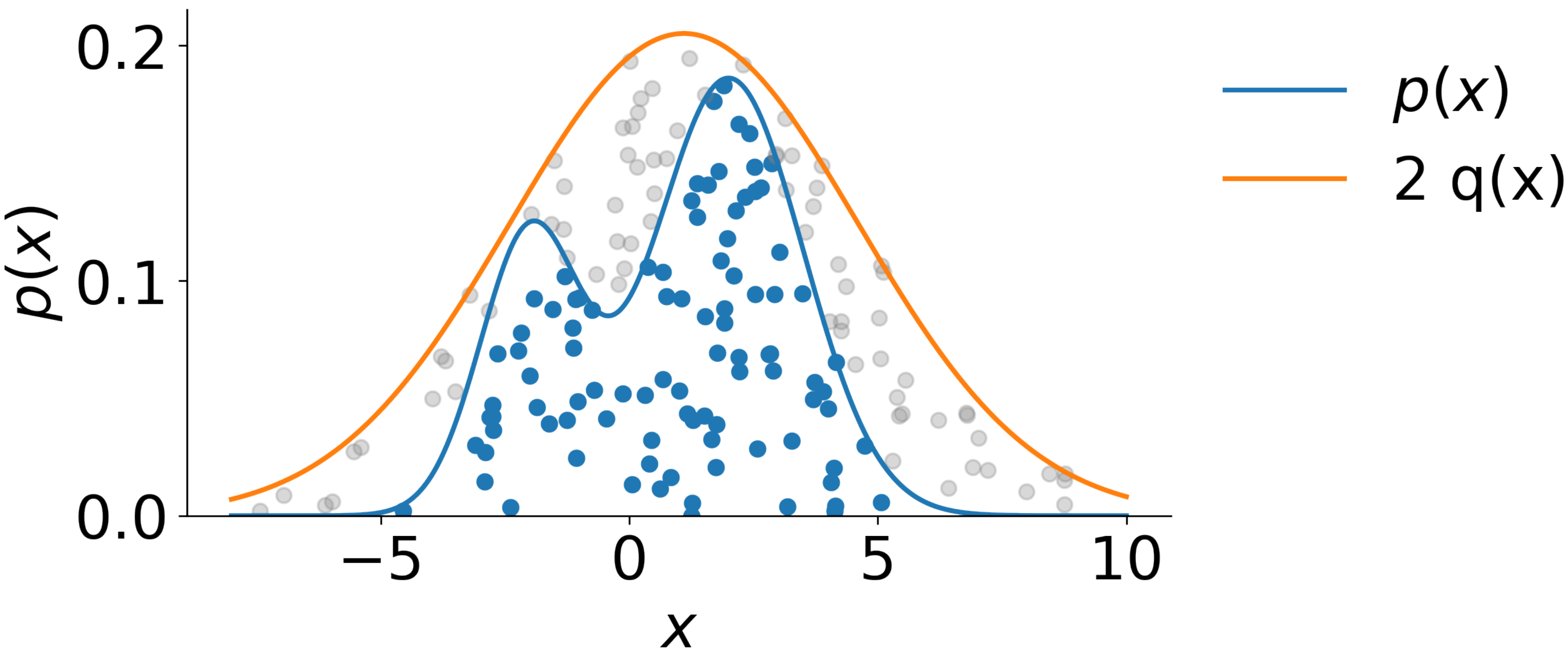
Accepts $\frac{1}{M}$ points on average

1d sampling (continuous)



$$\frac{\widehat{p}(x)}{Z} \leq M q(x)$$

1d sampling (continuous)



$$\widehat{p}(x) \leq \underbrace{ZM}_{\widetilde{M}} q(x)$$

Summary

Pros:

- Works for most distributions (even unnormalized)

Cons:

- If q and p are too different (M is large), rejects most of the points
- M is large for d-dimensional distributions