Why sample?

↳ Estimation of Random Variables, properties of the distribution, etc.

Application: If there is a randomised process with associated random variable X

How to discern $E(X) = \mu$?

Ans: Monte Carlo estimation.

$$E(x) \approx \frac{1}{n} \sum X_i$$

Sample and Estimate

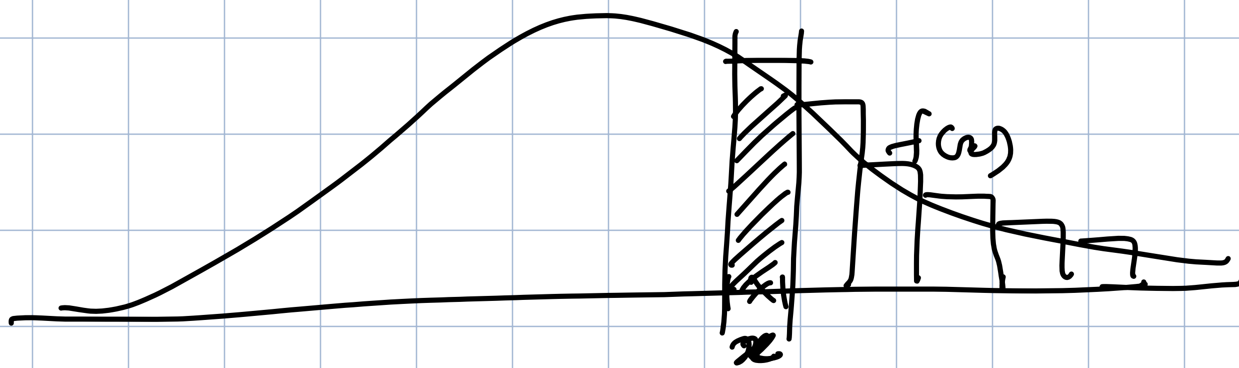
Q: How to sample data from a distribution $f(x)$?

* In physical processes, experiment might be one run of randomised trial

On Computers generally we have access to pseudo random number $U[0,1]$

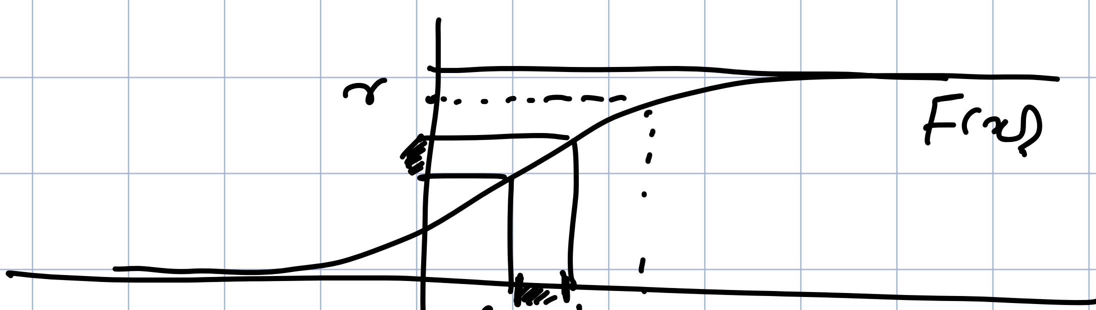
So

Q. If I have acess of $U[0,1]$, how do I create samples from $f(x)$?



Sample from $(x \pm \delta)$ with probability

$$\int_{x-\delta}^{x+\delta} f(x)\,dx$$



$P(a < x < b) = F(b) - F(a)$

$$r \sim U[0,1]$$

$$X = F^{-1}(r)$$

$$r = U[0,1]$$

$$X = F^{-1}(r)$$

$$Pr(X \leq x) = F(x)$$

Thus $\quad Pr(X = x) = f(x)$

Inversion method ! $\uparrow$

(2) Suppose you want to estimate some properties of function $w(x)$ for $x \sim f(x)$ but we do not have access to $\underline{\underline{f(x)}}$.

Examples

* We may know a distribution only up to a constant.

* Even if $f(x)$ is known $F(x)$ or its inverse might be nontrivial to calculate

Suppose you have access to $g(x)$

$$E_f[w(x)] = \int w(x) f(x)\, dx$$

$$= \int w(x) \frac{f(x)}{g(x)} g(x)\, dx$$

$$= E_g\left[\underbrace{\frac{f(x)}{g(x)} w(x)} \right]$$

estimate using Monte Carlo.

$$= \frac{1}{n} \sum_{1}^{n} \frac{f(x_i)}{g(x_i)} \cdot w(x_i)$$

Thus we can estimate $E_f[w(x)]$ using a "proposal distribution".

Q. Are all distribution $g(x)$ good?

Q. What should we look at to know if $g(x)$ is good?

A. How good is the estimate provided by the estimator using $g(x)$.

$$y = \frac{w(x) f(x)}{g(x)} \qquad x \sim g(x)$$

$E(y) = E_f[w(x)] = \mu$

$Var(y) = ?$

$$E_g\left[(\cancel{y} - \mu)^2\right] = E_g\left[\cancel{y}^2\right] - \mu^2$$

What value of g minimises

$$Var_g(y) = \int \frac{f^2 w^2}{g}\, dx - \mu^2$$

$\min Var_g(y)$

$$\int \frac{f^2 w^2}{g}\, dx \geq \Big(\qquad \Big)$$

$$\left(\int a(x)\, b(x)\, dx\right)^2 \underset{\text{t. gm.}}{\leq} \int a^2(x)\, dx \int b^2(x)\, dx.$$

$$\langle a.b\rangle^2 \leq \|a\|_2^2 \|b\|_2^2$$

$$\left(\int f\, |w|\, dx\right)^2 \leq \int \left(\frac{f\, |w|}{\sqrt{g}}\right)^2 dx \cdot \underbrace{\int (\sqrt{g})^2\, dx}_{1}$$

$$\int \frac{f^2 w^2}{g}\,dx \geq \left(\int (f|w|)\,dx\right)^2$$

$$\boxed{\int \frac{f^2 w^2}{g}\,dx \geq E\left[|w|\right]^2}$$

Tight when.

$$\frac{f\,|w|}{\sqrt{g}} = c\sqrt{g}$$

i.e $\quad g = \frac{1}{c} f|w|$

$$\boxed{g \propto f|w|}$$

What if $\quad w(x) > 0 \quad \forall x$

$$g = \frac{1}{c} f(x)\,(w(x))$$

In which case

$$Var = \int \frac{f^2 \omega^2}{f \omega} \cdot c \, dx - \mu^2$$

$$= c \cdot \underbrace{\int f \omega \, dx}_{} - \mu^2$$

$$c \cdot \mu - \mu^2$$

$c = ?$ $\quad c = \int f \cdot \omega \, dx$

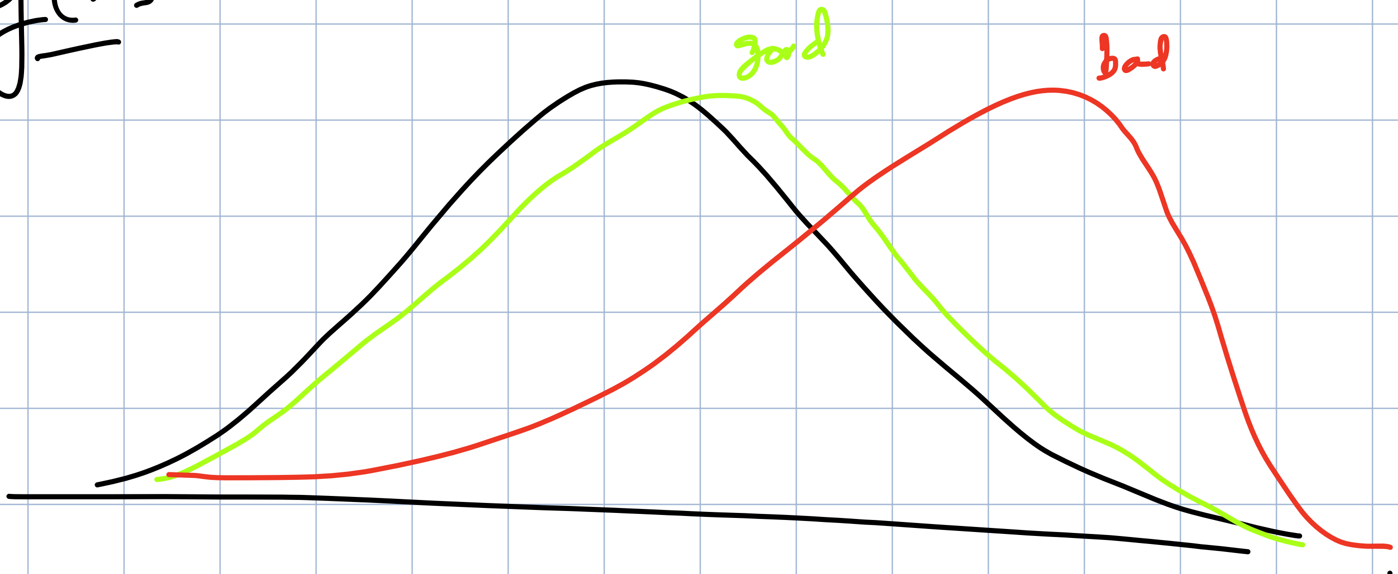
$c = \mu$

$$\boxed{Var = 0} !$$

Why? Knowing $g(x)$ already require known μ which is what we want to estimate

Thus even 1 sample is enough.

$g(x)$



Exercise:

$$S = x_1 + x_2 + \dots x_n \qquad x_i \in R$$

What is the "best" sampling strategy to estimate the mean.

what if

$$\vec{S} = \vec{x}_1 + \vec{x}_2 + \vec{x}_3 + \dots + \vec{x}_n \qquad \vec{x}_i \in R^d$$

Importance Sampling does not let ~~the~~ us sample from $f(x)$. Just lets us estimate $E_f(\cdot)$ using $g(x)$

Note: there are cases when using $f(x)$ for estimation of certain quantities is bad. Can you give one such example?

$w(x)$ & $f(x)$ don't co-operate

$\exists x$ s.t $|w(x)|$ is high but $f(x)$ is low.

Exercise: Discuss when softmax can be approximated via randomly uniform sampling of tokens.

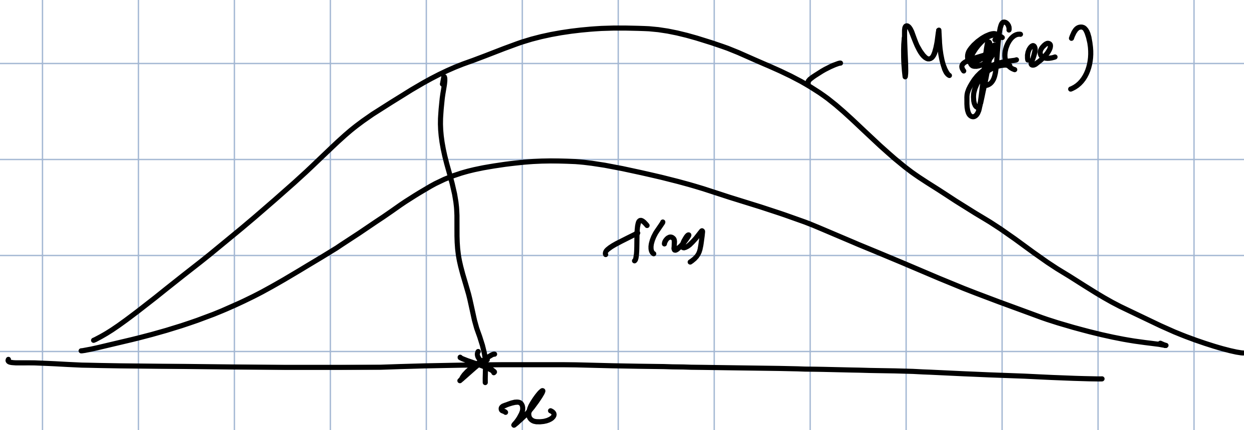
Rejection Sampling

How to Sample from $f(x)$ if there is no direct way

Say we can sample from $g(x)$

let M be such that $Mg(x) \geq f(x)$

the



$x \sim g(x)$

Accept with $P = \dfrac{f(x)}{Mg(x)}$ otherwise reject and resample.

$$p(x) \cancel{\propto} g(x) \cdot \frac{f(x)}{M g(x)}$$

$$\propto \frac{f(x)}{M}$$

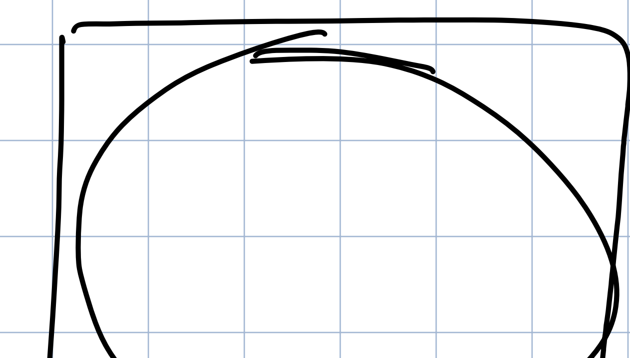
How do choose good g

- M: small
- $f \sim g$ g should be closer to f

So that rejection rate is small

Exercise:

$d = 2$



Sample a point uniformly from.

Circle using other Uniform [0,1] rand variable.

$$\text{rejection rate} = \frac{\text{Area}(\text{Sq}) - \text{Area}(\text{Circle})}{\text{Area}(\text{Sq})}$$

$$= 1 - \frac{\text{Area}(\text{Circle})}{\text{Area}(\text{Sq})}$$

$d = 2$?

As $d \uparrow$?

Can same approach be used?