INVERSE TRANSFORM METHOD (1D)

(FOR GENERATING NONUNIFORM DISTRIBUTION RANDOM NUMBERS)

SUPPOSE WE WANT TO GENERATE RANDOM NUMBERS ACCORDING TO SOME NONUNIFORM DISTRIBUTION DEFINED BY THE PROBABILITY DENSITY $p(x)$, SUCH THAT

$$\int_{-\infty}^{\infty} p(x)\,dx = 1$$

NOTE: $p(x)$ ALSO HAS TO BE A NON-NEGATIVE FUNCTION

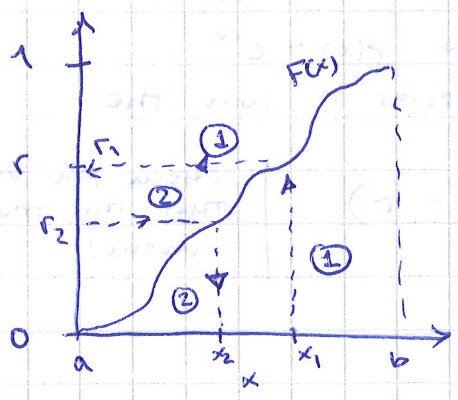
LET US THEN DEFINE THE CUMULATIVE DISTRIBUTION FUNCTION (CDF) OF $p(x)$

$$F(x) \equiv \int_{-\infty}^{x} p(x')\,dx'$$

LET US ALSO DENOTE ITS VALUE AT A GIVEN x AS r, AND CLEARLY $r \in [0,1]$.

NOTE:

- $F(x)$ IS A NON-DECREASING, NON-NEGATIVE FUNCTION.
- SINCE A CDF CAPTURES ALL THE PROPERTIES OF THE UNDERLYING PROBABILITY DENSITY, EACH $p(x)$ HAS ITS OWN, UNIQUE $F(x)$.
- $F(x_0)$ GIVES THE PROBABILITY OF CHOOSING A REAL VALUE x LESS THAN x_0 (ACCORDING TO $p(x)$), THAT IS $F(x_0) = P(x \leq x_0)$.
  [THINK ABOUT THIS FOR A MINUTE].



NOW, LET $p(x)$ BE DEFINED FOR SOME RANGE $[a,b]$. LET'S SAY THE CORRESPONDING CDF IS SHOWN ON THE LEFT.

1. FOR SOME VALUE OF x_1, WE HAVE THE CORRESPONDING VALUE r_1
2. FOR ANY $0 \leq r_2 \leq 1$, THERE IS A UNIQUE CORRESPONDING VALUE x_2 AGAIN ACCORDING TO $p(x)$

SINCE $r = F(x)$, WE CAN OBTAIN THE INVERSE MAPPING (2)

2)

AS $F^{-1}(r) = x.$

THE METHOD IS THUS CLEAR:

1° OBTAIN $F(x) = \int_{-\infty}^{x} p(x')\,dx' = r$

2° CALCULATE ITS INVERSE FUNCTION $F^{-1}(r) = x$

3° GENERATE <u>UNIFORM</u> RANDOM NUMBERS $\{r_i\}$, AND OBTAIN THE NON-UNIFORMLY DISTRIBUTED RANDOM NUMBERS $\{x_i\}$ — ACCORDING TO $p(x)$ — FROM $F^{-1}(r_i) = x_i$

THERE IS A PROPER, RIGOROUS PROOF TO THE METHOD WHICH YOU CAN FIND E.G. IN WIKIPEDIA.

<u>EXAMPLE</u>

$$p(x) = e^{-x},\ x \geq 0 \quad \Big| \quad \text{CHECK: } \int_0^{\infty} p(x)\,dx = -\Big/_{0}^{\infty} e^{-x} = 1 \quad \underline{\text{OK!}}$$

NOW $F(x)$ IS

$$F(x) = \int_0^x e^{-x'}\,dx' = 1 - e^{-x} = r$$

SOLVE x: $\quad x = \underbrace{-\ln(1-r)}_{F^{-1}(r)}$

IN ORDER TO GENERATE RANDOM NUMBERS ACCORDING TO THE EXPONENTIAL DISTRIBUTION $p(x) = e^{-x}$ WE INPUT UNIFORM RANDOM NUMBERS r INTO THE EXPRESSION

$$F^{-1}(r) = -\ln(1-r)$$

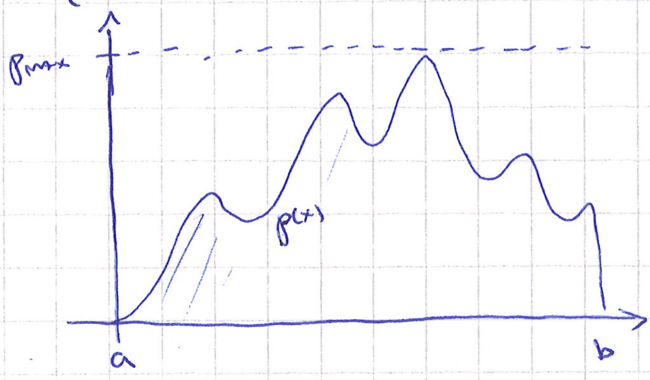
| CHECK FOR YOURSELF THAT THIS TRULY WORKS!

DOWNSIDE TO THE METHOD: WE HAVE TO BE ABLE TO OBTAIN THE ANALYTICAL FORMS OF $F(x)$ AND OF COURSE $F^{-1}(r)$

ALSO NOTE THAT THERE IS A MUCH SIMPLER METHOD, WHICH CAN UNFORTUNATELY BE DRAMATICALLY INEFFICIENT IN SOME CASES:

THE von NEUMANN REJECTION METHOD

(AKA THE HIT-AND-MISS METHOD)



AGAIN, LET'S SUPPOSE OUR $p(x)$ IS DEFINED FOR SOME INTERVAL $[a,b]$ AND ITS MAXIMUM VALUE IS p_{max}.

THE METHOD IN ITS ENTIRETY IS AS FOLLOWS:

1° GENERATE A RANDOM NUMBER $a \leq x \leq b$, CALCULATE $p(x)$

2° GENERATE ANOTHER RANDOM NUMBER $0 \leq r \leq p_{max}$

3° ACCEPT x IF $r \leq p(x)$

RANDOM NUMBERS $\{x_i\}$ GENERATED THIS WAY ARE SAMPLED FROM THE DISTRIBUTION $p(x)$.

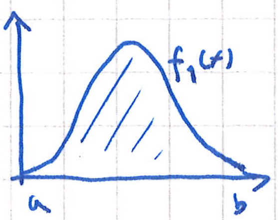
CLEARLY, WHILE THE METHOD IS SUPER SIMPLE, IT CAN LEAD TO VERY INEFFICIENT SAMPLING ESPECIALLY IF THE VARIATIONS IN $p(x)$ ARE STRONG.

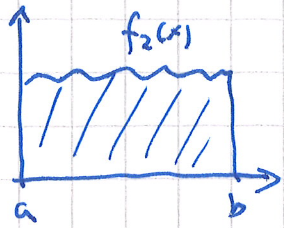
FOR MORE, SEE E.G. GOULD, TOBOCHNIK & CHRISTIAN, SECTION 11.5

Importance Sampling

In Monte Carlo integration the error was $\frac{\sigma}{\sqrt{N}}$.
To make the estimate more accurate, we can either increase N (brute force) or somehow try to make σ smaller.
Importance sampling addresses the latter option.

Consider integrals of these two functions





$\int_a^b f_1(x)\,dx$ etc.

Clearly, in the right-hand side case, MC sampling will result in a smaller variance (and hence, σ).

If we have a function similar to $f_1(x)$ above (i.e., a strongly varying one), what we want is to transform our original integral into something that has less fluctuations.

Let's introduce a positive function $g(x)$ so that it somehow mimics the variations of the original function. $g(x)$ also has the property

$$\int_a^b g(x)\,dx = 1$$

so that it can be taken as a proper probability density.

We then rewrite our original integral as

$$I = \int_a^b \left[\frac{f(x)}{g(x)}\right] g(x)\,dx$$

Note: based on our choice of $g(x)$, the variance of this ratio should be small.

NOW, CONSIDER THE CDF OF $g(x)$

$$G(x) = \int_a^x g(x')\,dx' \quad ; \quad \begin{array}{l} G(b) = 1 \\ G(a) = 0 \end{array}$$

$$\Rightarrow \quad I = \int_a^b \left[\frac{f(x)}{g(x)}\right] dG(x)$$

CHANGE THE INTEGRATION VARIABLE $r = G(x)$

$$\Leftrightarrow \quad x = G^{-1}(r)$$

THEN

$$I = \int_{G(a)=0}^{G(b)=1} \left[\frac{f(x)}{g(x)}\right] dr = \int_0^1 \left[\frac{f(G^{-1}(r))}{g(G^{-1}(r))}\right] dr$$

SO, OUR MONTE CARLO INTEGRATION WOULD BE

$$I_N = \frac{1}{N} \sum_{i=1}^{N} \frac{f(G^{-1}(r_i))}{g(G^{-1}(r_i))}$$

$\uparrow$ UNIFORM RANDOM NUMBER

<u>NOTE:</u> $G^{-1}(r_i)$ MAPS THE UNIFORM RANDOM NUMBER r_i TO A VALUE x_i WHICH FOLLOW THE PROBABILITY DENSITY $\underline{\underline{g(x)}}$!

IN OUR EXAMPLE IN THE LECTURE SLIDES:

$$f(x) = 2x + \sin(8\pi x)\,\mathrm{erfc}(x)/10$$

$$g(x) = 2x$$

$$\Rightarrow \quad G(x) = \int_0^x g(x')\,dx' = x^2$$

SO THAT $G^{-1}(x) = \sqrt{x}$

METROPOLIS METHOD

HOW TO GENERATE POINTS / SAMPLE ACCORDING TO SOME DISTRIBUTION $w(x)$, WHICH IN ADDITION MIGHT NOT BE NORMALIZED?

IN THE CASE OF MULTI-DIMENSIONAL INTEGRALS (THINK: STATISTICAL PHYSICS, LARGE NUMBERS OF PARTICLES) OR IN CASES WHERE SIMPLE IMPORTANCE SAMPLING BASED ON CDFs THERE MAY BE NO SIMPLE SOLUTIONS

→ THE METROPOLIS ALGORITH (METROPOLIS ET AL., 1953)

USEFUL, E.G., FOR CALCULATING AVERAGES OF THE FORM

$$\langle A \rangle = \frac{\int A\, w(\vec{r})\, d\vec{r}}{\int w(\vec{r})\, d\vec{r}} .$$

WHAT THE METHOD DOES: IT CREATES A RANDOM WALK OF POINTS $\{x_i\}$ [FOR SIMPLICITY, WE'LL DO THINGS IN 1D HERE] WHOSE DISTRIBUTION ASYMPTOTICALLY APPROACHES $\sim w(x)$.

THE RANDOM WALK IS DEFINED BY TRANSITION PROBABILITY $T(x_i \to x_j)$, SO THAT $\{x_i\}$ CONVERGES TO $w(x)$.

CONDITION: SUPPOSE WE GOT $\{x_i\} \sim w(x)$. ONCE REACHED, THE TRANSITION PROBABILITIES SHOULD NOT DESTROY THIS EQUILIBRIUM.

BALANCE: THE AVERAGE NUMBER OF ACCEPTED MOVES AWAY FROM THE CURRENT STATE/SITE MUST BE EXACTLY EQUAL TO THE NUMBER OF ACCEPTED MOVES TO THE SITE FROM OTHER SITES.

IN PRACTICE, IT IS CONVENIENT TO IMPOSE A STRONGER CONDITION — SUFFICIENT BUT NOT NECESSARY

DETAILED BALANCE :

$$w(x_i)\, T(x_i \to x_j) = w(x_j)\, T(x_j \to x_i)$$

THAT IS, EVERY MOVE $x_i \to x_j$ IS EXACTLY BALANCED BY THE REVERSE MOVE $x_j \to x_i$.

THE DETAILED BALANCE DOES NOT UNIQUELY DEFINE $T(x_i \to x_j)$, MANY POSSIBLE FORMS.

A SIMPLE CHOICE: $T(x_i \to x_j) = \min\left[1, \frac{w(x_j)}{w(x_i)}\right]$

IN PRACTICE, THE ALGORITHM IS AS FOLLOWS :

0. START FROM x_0
1. CHOOSE TRIAL $x_{TRIAL} = x_i + \Delta x$ ACCORDING TO SOME RANGE, SAY, $[-\delta, \delta]$
2. ACCEPT THE TRIAL WITH PROBABILITY

$$r = \min\left[1, \frac{w(x_j)}{w(x_i)}\right]$$

3. IF ACCEPTED, $x_{i+1} = x_{TRIAL}$
   ELSE $x_{i+1} = x_i$
4. GO TO 1., REPEAT UNTIL DONE.

PRACTICAL THINGS:

- FOR EFFICIENT SAMPLING, IT IS WISE TO CHOOSE x_0 SO THAT $w(x_0)$ IS LARGE / MAXIMUM.
- δ: IF TOO SMALL : TOO MANY ACCEPTED STEPS, INEFFICIENT SAMPLING
  IF TOO LARGE : TOO FEW ACCEPTED STEPS, INEFFICIENT SAMPLING
- A RULE OF THUMB: $\sim 1/3 - 1/2$ OF TRIALS SHOULD BE ACCEPTED FOR EFFICIENT SAMPLING