Lecture 1 How random number generators (don't) work.

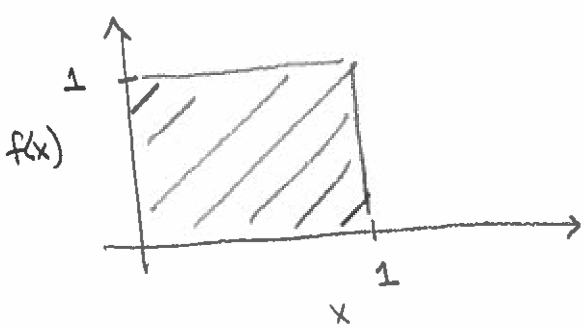
Plan

1. Review syllabus
2. Quantile functions
3. Box Muller
4. Rejection sampling
5. Importance sampling.

Quantile functions: we can generate a lot of random variables using Unif(0,1) random variables.

- Running assumption that we can generate as many Unif(0,1) r.v.s as we want.

Unif(0,1)



pdf $f(x) = 1, \quad 0<x<1$
"probability density function"

cdf $F(x) = \mathbb{P}\{U \le x\}$

$$= \int_0^x f(u)\,du$$

$$= \int_0^x du$$

$$= x, \quad 0<x<1$$

"cumulative distribution function"

- **Exercise**: given just ONE Unif(0,1) random variable, we can generate INFINITELY MANY ind. Unif(0,1) r.v.s

- Hint: use binary notation.
- Caveat: This is NOT how computers work.
- The cdf of a random variable X is $F(x) = \mathbb{P}\{X \le x\}$.
- The quantile function is $Q = F^{-1}$ if $x \mapsto F(x)$ is strictly increasing, continuous
- More generally, $Q(p) = \inf\{y \in \mathbb{R} : F(y) \ge p\}$

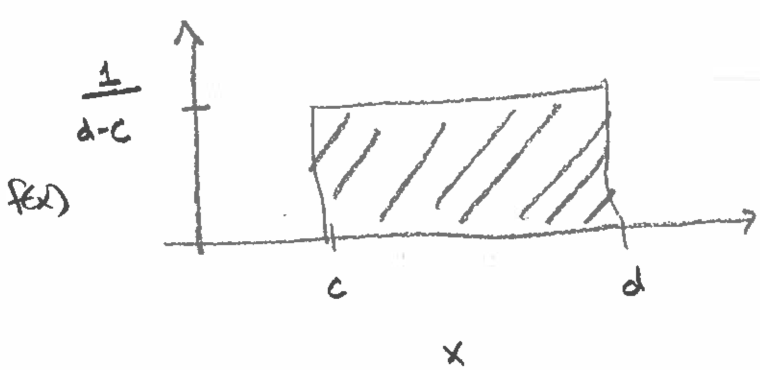
Proposition (Random number generation). If $U \sim \mathrm{Unif}(0,1)$, then $Q(U)$ has cdf F.

Pf 1. First, check the Galois inequalities $Q(p) \le x$ if and only if $p \le F(x)$ $\forall\, x, p \in \mathbb{R}$.

2. Next, $\mathbb{P}\{Q(U) \le x\} = \mathbb{P}\{U \le F(x)\} = F(x)$. $\square$

Examples:

1. $\mathrm{Unif}(c,d)$



$$F(x) = \frac{x-c}{d-c}, \quad c < x < d$$

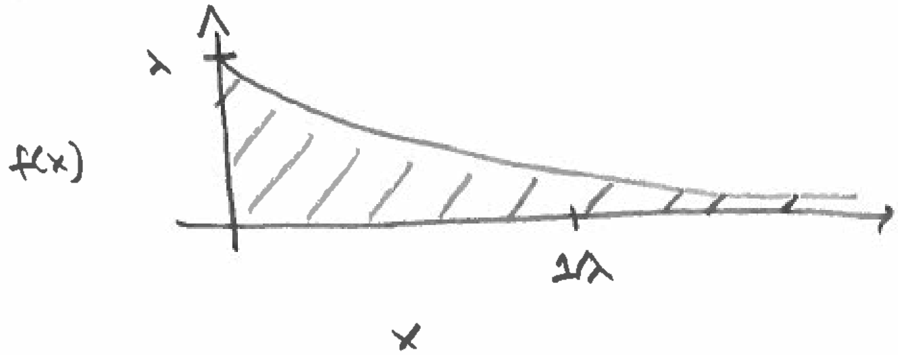
$$Q(p) = c + (d-c)\,p, \quad 0 < p < 1$$

$$\Rightarrow \boxed{c + (d-c)\,U \sim \mathrm{Unif}(c,d)}$$

2. $\exp(\lambda)$

$$F(x) = 1 - e^{-\lambda x}, \quad x > 0$$

$$Q(p) = \frac{-\log(1-p)}{\lambda}$$



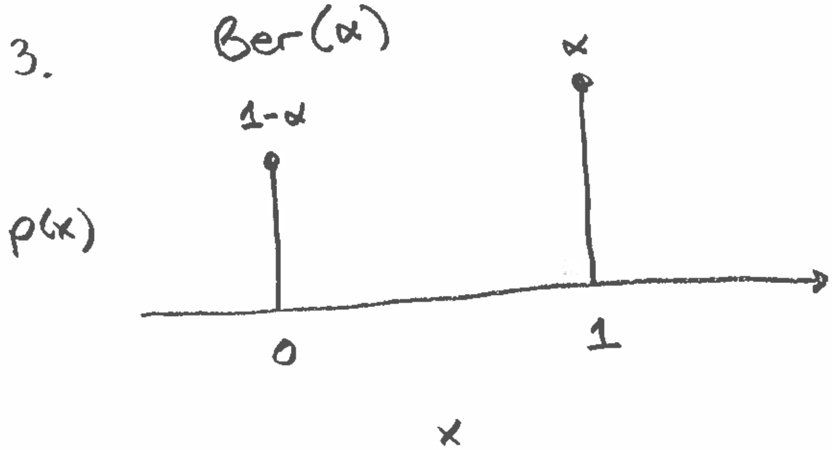
$$\Rightarrow \quad \frac{-\log(1-u)}{\lambda} \sim \exp(\lambda)$$

Using $1-u \sim \text{Unif}(0,1)$, there is a simpler formula $\boxed{\frac{-\log u}{\lambda} \sim \exp(\lambda)}$

3. $\text{Ber}(\alpha)$

$$F(x) = \begin{cases} 0, & x < 0 \\ 1-\alpha, & 0 < x < 1 \\ 1, & x > 1 \end{cases}$$

$$Q(p) = \mathbb{1}\{p \geq 1-\alpha\}$$



$$\Rightarrow \quad \mathbb{1}\{u \geq 1-\alpha\} \sim \text{Ber}(\alpha) \quad \text{OR} \quad \boxed{\mathbb{1}\{u \leq \alpha\} \sim \text{Ber}(\alpha)}$$

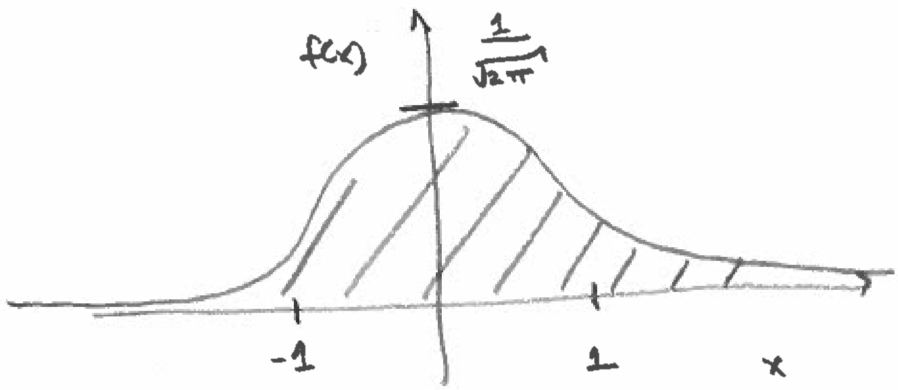
<u>Q</u>: How do we generate a vector of independent r.v.s $\vec{X} = (X_1, \dots, X_N)$ with $X_i \sim \text{Ber}(\alpha_i)$?

A: Generate $\vec{u} = (u_1, \dots, u_N)$, $\vec{\alpha} = (\alpha_1, \dots, \alpha_N)$, $\vec{X} = \mathbb{1}\{\vec{u} \leq \vec{\alpha}\}$

Q: we've got $\text{Unif}(c,d)$, $\exp(\lambda)$, $\text{Ber}(\alpha)$. what else can we do?

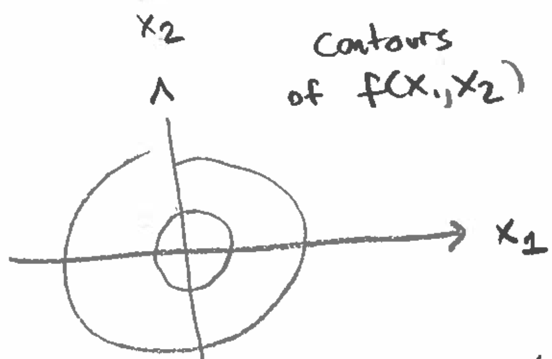
Box Muller: How can we generate $X \sim \mathcal{N}(0,1)$?

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$



- **Idea**: Generate TWO indep. Gaussians $X_1, X_2 \sim \mathcal{N}(0,1)$

$$f(x_1,x_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2+x_2^2)\right)$$



1. Change of variables $\begin{cases} x_1 = r\cos\theta \\ x_2 = r\sin\theta \end{cases} \Rightarrow |J| = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$

2. change of density $f(x_1,x_2)\,dx_1 dx_2 = \frac{1}{2\pi} \exp\left(-\frac{1}{2}r^2\right) r \, d\theta\, dr.$
$= f(\theta, r)\, d\theta\, dr.$

3. Change of variables $S = \frac{1}{2}r^2 \Rightarrow |J| = |r| = r.$

4. Change of density $f(\theta,r)\,d\theta\,dr = \frac{1}{2\pi}\exp(-s)\,d\theta\,ds$
$= f(\theta,s)\,d\theta\,ds.$

5. pdf $f(\theta,s) = \frac{1}{2\pi}\exp(-s) \Rightarrow \theta, S$ indep. r.v.'s with $\theta \sim \text{Unif}(0,2\pi)$, $S \sim \exp(1)$

$$\Rightarrow \boxed{\begin{cases} X_1 = \sqrt{-2\log U_1}\,\cos(2\pi U_2) \sim \mathcal{N}(0,1) \\ X_2 = \sqrt{-2\log U_2}\,\sin(2\pi U_2) \sim \mathcal{N}(0,1) \end{cases}}$$

and X_1, X_2 are indep.

<u>Q</u>: How can we generate a $\mathcal{N}(\vec{M}, \Sigma)$ random vector?

<u>A</u>: Generate independent $Z_1, \dots, Z_N \sim \mathcal{N}(0,1)$ and set $\vec{X} = \vec{M} + \Sigma^{1/2} \vec{Z}$.

- <u>Summary</u>: we can generate lots of r.v.'s using Unif(0,1) r.v.s
- <u>Caveat</u>: Your computer builds on these foundations and uses even <u>faster</u> algorithms, e.g. "Ziggurat" (Marsaglia & Tsang, 1984).

<u>Rejection sampling</u>: Want to sample from pdf f, only know how to sample from pdf g.

a) Sample $X \sim g$, compute $r = \frac{f(x)}{g(x)M}$ (≤ 1).

b) Sample $U \sim \text{Unif}(0,1)$. If $U \leq r$, accept X. Otherwise, try again.

<u>Q</u>: What is the best M?

$\underline{A}$: $M = \sup_{x \in \mathbb{R}} \frac{f(x)}{g(x)}$

<u>Proposition (Rejection works):</u> Random variable X from rejection sampling has pdf f, waiting time to success is $\text{Geo}\left(\frac{1}{M}\right)$.

<u>Pf</u> 1. Calculate $\mathbb{P}(X \in dx, \text{success}) = \mathbb{P}(X \in dx)\mathbb{P}(\text{success} \mid X = x)$

$$= g(x) \frac{f(x)}{g(x) M} = \frac{f(x)}{M}$$

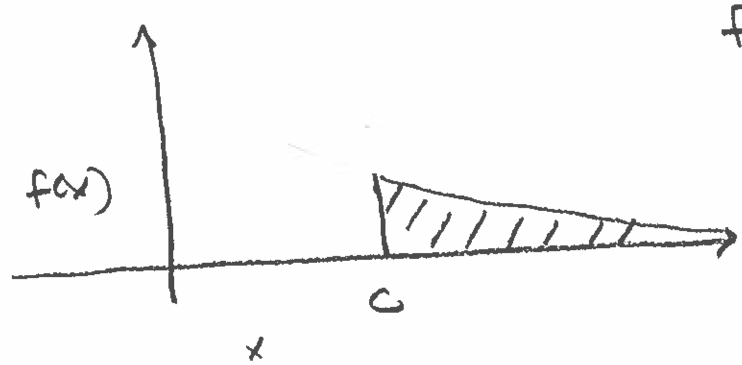
2. Integrate $\mathbb{P}(\text{success}) = \int \mathbb{P}(X \in dx, \text{success})\, dx = \frac{1}{M}$

3. Apply Bayes rule $\mathbb{P}(X \in dx \mid \text{success}) = \frac{\mathbb{P}(X \in dx, \text{success})}{\mathbb{P}(\text{success})}$

$$= \frac{f(x)/M}{1/M} = f(x).$$ $\square$

<u>Ex</u>: We want to sample $Z \mid Z > C$ where $Z \sim \mathcal{N}(0,1)$

$$f(x) = \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)}{\mathbb{P}(Z \geq C)}, \quad x > C$$



<u>Idea</u>: Let's try sampling $X \sim \mathcal{N}(c, 1)$

$$g(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-c)^2\right)$$

$$\frac{f(x)}{g(x)M} = \underbrace{\frac{\exp\left(-xc + \frac{1}{2}c^2\right)}{P(Z \geq c)M}}_{\text{Need to make this } \leq 1}, \quad x > c$$

$$\boxed{\text{Sample } X \sim \mathcal{N}(c,1), \text{ compute } r = \exp\left(-c(x-c)\right)\mathbb{1}\{x \geq c\}, \text{ accept w/ prob } r}$$

<u>Exercise</u>: What is the success probability for generating a truncated Gaussian?

<u>Hint</u>: It's $\frac{1}{M} \geq \frac{1}{\sqrt{2\pi}} \frac{c}{c^2+1}$. Prove it using "Mills' ratio".

<u>Q</u>: Is this better than generating a bunch of Gaussians $Z_1, \dots, Z_N \sim \mathcal{N}(0,1)$ and selecting the ones larger than c?

<u>Importance sampling</u>: want to sample from pdf f, only know how to sample from pdf g.

a) Sample $X_1, \dots, X_N \sim g$.

b) Set $w_i = \frac{f(x_i)}{g(x_i)M}$ for $1 \leq i \leq N$.

c) Approximate any average $\mu_h = \int f(x)h(x)\,dx$ as

$$\hat{h}_{imp} = \frac{w_1 h(x_1) + \dots + w_N h(x_N)}{w_1 + \dots + w_N}.$$

Q : what is the best M?

A : whatever makes the computation of w_i simplest

Proposition (Importance sampling works): Random variable $w_i h(X_i)$ from importance sampling has expected value $\frac{1}{M}\mu_h$ (hence $\mathbb{E}\, w_i = \frac{1}{M}$).

Pf $\mathbb{E}\left[w_i h(X_i)\right] = \int g(x) \frac{f(x)}{g(x) M} h(x)\,dx = \frac{1}{M}\mu_h$. □

Ex We want to sample $Z \mid Z > c$ where $Z \sim \mathcal{N}(0,1)$

Let's try sampling $X_i \sim \mathcal{N}(c, 1)$ for $1 \le i \le N$.

$$w_i = \frac{f(x_i)}{g(x_i) M} = \underbrace{\frac{\exp\left(-x_i c + \frac{1}{2}c^2\right)}{\mathbb{P}(Z \ge c)\, M}}_{\text{Need to make this simple.}}, \quad x_i > c$$

$$\boxed{\text{Sample } X_1, \dots, X_N \sim \mathcal{N}(c, 1), \text{ compute } w_i = e^{-cX_i} \mathbb{1}\{X_i > c\} \text{ (or maybe } w_i = e^{-c(X_i - c)} \mathbb{1}\{X_i > c\}\text{), use } (w_i, X_i)_{1 \le i \le N} \text{ for computations}}$$

Proposition (Importance beats rejection): For a fn h, let

$$\hat{h}_{rej} = \frac{\mathbb{1}\{U_1 \le r_1\} h(X_1) + \dots + \mathbb{1}\{U_N \le r_N\} h(X_N)}{\mathbb{1}\{U_1 \le r_1\} + \dots + \mathbb{1}\{U_N \le r_N\}}$$

be the estimate of $\mu_h = \int f(x) h(x)$ from rejection sampling.

Let $\hat{h}_{imp} = \dfrac{w_1 h(x_1) + \cdots + w_N h(x_N)}{w_1 + \cdots + w_N}$

be the estimate of μ_h from importance sampling. Then as $N \to \infty$

$$\sqrt{N}\left(\hat{h}_{rej} - \mu_h\right) \to \mathcal{N}(0, V_{rej}) \qquad \text{"} \hat{h}_{rej} = \mu_h \pm \frac{1}{\sqrt{N}} V_{rej}^{1/2} \text{"}$$

$$\sqrt{N}\left(\hat{h}_{imp} - \mu_h\right) \to \mathcal{N}(0, V_{imp}) \qquad \text{"} \hat{h}_{imp} = \mu_h \pm \frac{1}{\sqrt{N}} V_{imp}^{1/2} \text{"}$$

where

$$V_{imp} = \int \frac{f^2(x)}{g(x)} |h(x) - \mu_h|^2 \, dx \leq V_{rej} = M \int f(x) |h(x) - \mu_h|^2 \, dx.$$

True because $M \geq \dfrac{f(x)}{g(x)} \quad \forall \, x \in \mathbb{R}$

Pf Exercise using the "delta method". $\square$