Aula 8

Aula passada

- Gerando amostras de v.a. discretas
- Gerando Geométrica
- Método da transformada inversa
- Gerando Binomial
- Gerando permutações

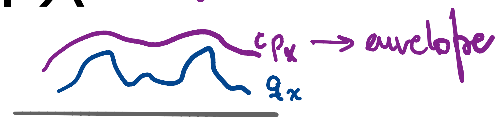
Aula de hoje

- Método da rejeição (*rejection sampling*)
- Exemplos
- *Importance Sampling*
- Exemplos
- Generalização

Rejection Sampling

- Técnica fundamental para geração de amostras aleatórias
  - usada em Markov Chain Monte Carlo (veremos depois)
- **Ideia:** usar distribuição conhecida para gerar amostras de outra distribuição, mais complicada
- Sejam *X* e *Y* duas v.a. com distribuições p_x e q_x , definidas no mesmo suporte → Se $p_x > 0$, então $p_y > 0$
  - $p_x = P[X = x]$, $q_x = P[Y = \cancel{x}\, y]$
- Assuma que $q_x <= c\, p_x$ para alguma constante *c* e todo *x*
  - ou seja, $c\, p_x$ é uma *função envelope* para q_x

  $c p_x$ está por cima de q_x



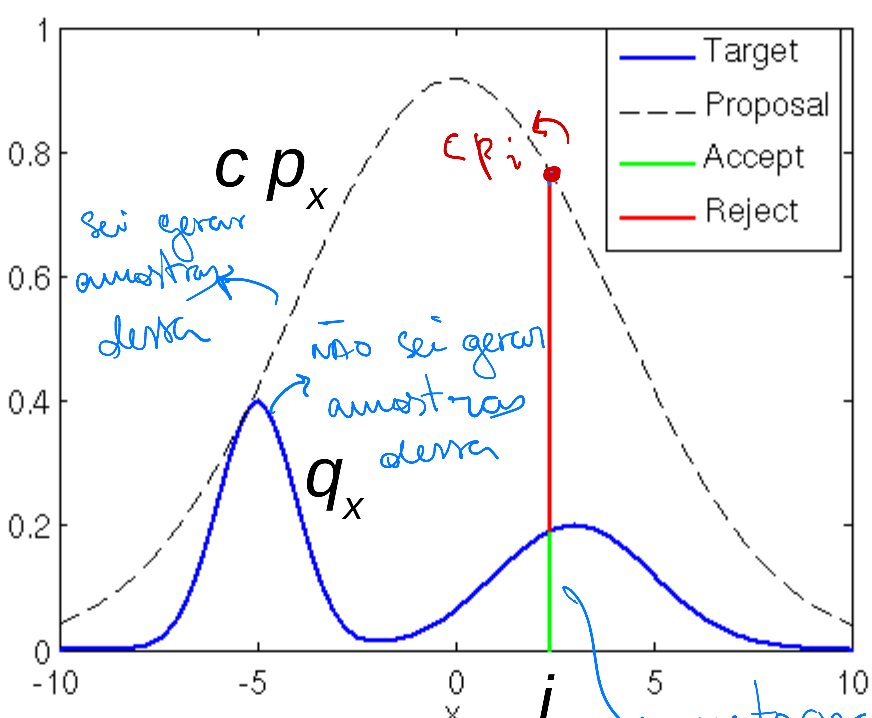
- Supor que sabemos gerar amostras para v.a. *X*
  - ex. método da transformada inversa
- Como usar isto para gerar amostras para v.a. *Y*?

Rejection Sampling

- Algoritmo (proposto por von Neumann) → um dos pais do métodos de MC
  1) Gerar valor para i a partir de p_x $\quad i \sim p_x$
  2) Gerar u uniforme(0, $c\ p_i$) contínua, usando o i gerado
  3) Se $u < q_i$, retorna i, caso contrário vai para 1)

  $P[Y=i]$ ↳ rejeita a amostra

- Graficamente



↳ retorna i com probabilidade $p_x - q_x$

- Algoritmo pode *rejeitar* amostra gerada de X várias vezes
  - *rejection sampling*
- Rejeita com probabilidade proporcional a diferença de p_x e q_x

Rejection Sampling Funciona

- Mostrar que algoritmo funciona
  - ou seja, algoritmo gera amostra *i* com probabilidade q_i
- Considere um valor i e o evento *aceitar*

  P[X=i, *aceitar*] = P[X=i] P[*aceitar* | X= i] = $p_i\, q_i/(c\, p_i) = q_i/c$
- Probabilidade de aceitar (por prob. total)

$$P[aceitar]=\sum_i P[X=i, aceitar]=\sum_j \frac{q_j}{c}=\frac{1}{c}$$

- Ao final de cada rodada, algoritmo aceita com prob. 1/c
- Temos então

$$\frac{q_i}{c}$$

$$P[\underbrace{X=i}_{\text{prob. do algoritmo retornar } i}|aceitar]=\frac{P[X=i, aceitar]}{P[aceitar]}=q_i=P[Y=i]$$

$$\frac{1}{c}$$

Figueiredo 2018

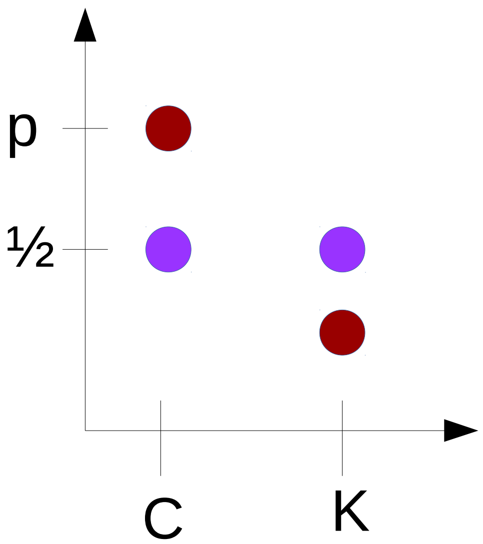
Rodadas de *Rejection Sampling*

- A cada rodada, algoritmo aceita com prob. 1/c $= P[\text{aceitar}]$
- Número de rodadas até aceitar é aleatório. Distribuição?
  - Geométrica com parâmetro 1/c $= p$
  - valor esperado = c $= 1/p$ Na média, precisamos de c rodadas para gerar uma amostra
  - complexidade de caso médio para cada amostra
- Escolha do valor para c é muito importante
  - menor valor tal que $q_x \le c\, p_x$ para todo x quero um envelope que mais se aproxima de q_x
- Valor depende da "distância" entre q_x e p_x
  - se muito diferentes, pode demandar c muito grande

Técnica funciona também com v.a. contínua

Exemplo

- Dado moeda enviesada, cara com prob p > ½
- Como gerar moeda sem viés?



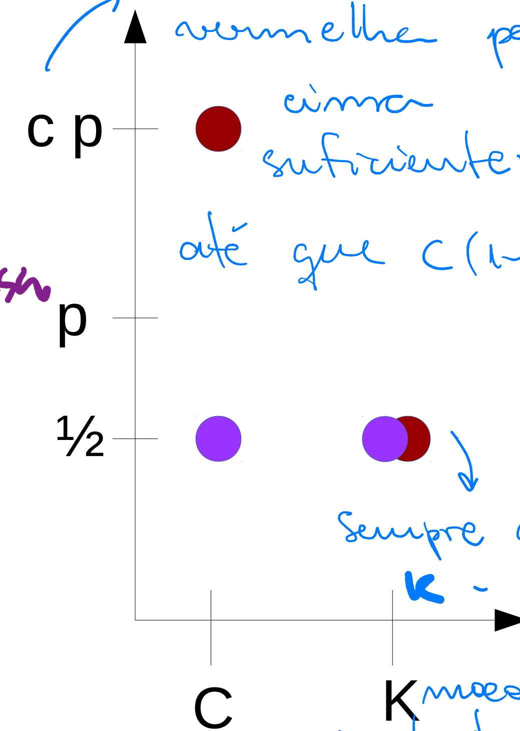
● Moeda enviesada

● Moeda honesta

quero gerar esta

- Encontrar constante c tal que $q_x \le c\, p_x$
- $c(1-p) = ½ \rightarrow c = 1/(2(1-p))$

Empurro a vermelha para cima suficientemente até que $c(1-p) = \frac{1}{2}$



Sempre aceita K - pois a moeda original estava enviesada para sair C

1) Gerar valor para i = {C, K} a partir de p_x *(moeda enviesada)*

2) Gerar u uniforme(0, $c\, p_i$) contínua, usando o i gerado

3) Se $u < q_i$, retorna i, caso contrário vai para 1)

Figueiredo 2018

Exemplo 2

- Gerar amostras da v.a. *Y* contínua com densidade $f_Y(x) = 20x(1-x)^3$, $0<x<1$
- Usando método da rejeição. Que distribuição proponente?
  - uniforme[0,1], $g_X(x) = 1$, $0<x<1$
- Determinar c tal que $f(x) <= c\, g(x)$ para $0<x<1$
- **Ideia:** encontrar máximo de f(x)/g(x) → g(x) é cte, então basta derivar f(x)
  - derivar, igualar a zero, resolver para x, encontrar valor
  - máximo em x = ¼, c = 135/64

$f(1/4) = 135/64$

1) Gerar valor para *x* a partir de *g(x)* (uniforme[0,1])
2) Gerar *u* uniforme(0, *c g(x)*) contínua, usando o *x* gerado
3) Se *u* < *f*(x), retorna *x*, caso contrário vai para 1)

Cenário Problemático

- Algoritmo de Monte Carlo é estimador universal
  - média amostral converge para valor esperado
- **Problema:** variância do estimador pode ser muito alta!
  - muitas, muitas amostras serão necessárias
- Exemplo $G_N = N \cdot M_n$

$$G_N = \sum_{i=1}^{N} g(i) \qquad M_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i) \quad , X_i \sim \text{Unif}(1,N)$$

- $g(i)$ é pequena (ou zero) para muitos valores de i, e grande para poucos valores
- Teremos que gerar muitas amostras para “acertar” os valores importantes de $g(i)$
- Ideias para atacar o problema?

Importance Sampling

→ Técnica de variance reduction

Devemos tentar gerar os elementos das regiões que possuem mais massa

- Amostrar com probabilidade diferente da original
  - uniforme no exemplo anterior
- Amostrar com maior probabilidade região mais importante para o problema em questão → Pode criar um viés para as regiões de grande massa
  - *importance sampling*
- Compensar pelo aumento desta probabilidade
- Usar método de Monte Carlo em problema reformulado
  - com novas funções de distribuição que compensem o viés
- **Objetivo:** reduzir a variância do estimador

  Quero gerar mais amostras de regiões importantes sem gerar viés

  - isto nem sempre ocorre, pois se exagerar demais vai ter que compensar com mais amostras!
  - técnica tem que ser usada com cuidado

Importance Sampling

o problema da uniforme é dar a mesma massa para TODAS AS PARCELAS

- Seja X uma v.a. uniforme(1,N)

v.a. uniforme

quero estimar

$$E[g(X)] = \sum_{i=1}^{N} P[X=i]\, g(i) = \frac{1}{N}\sum_{i=1}^{N} g(i) = \frac{G_N}{N}$$

- Seja Y v.a. com distribuição dada por h(i) > 0 para todo i

↳ escolho h de forma que eu saiba gerar suas amostras

$$E_h\left[\frac{g(Y)}{h(Y)}\right] = \sum_{i=1}^{N} \frac{g(i)}{\cancel{h(i)}}\cancel{h(i)} = \sum_{i=1}^{N} g(i) = G_N$$

Estou interessado nesta distribuição

Boa notícia! Podemos escolher h(i)

- Podemos estimar G_N estimando o valor esperado através da média amostral

  $y \sim h$, média amostral de $\frac{g(y)}{h(y)}$

  - Monte Carlo aplicado a outro valor esperado
- Usar h para reduzir variância!

Importance Sampling

- Seja Y_i uma sequência iid de v.a. com distribuiição $h(i)$

$$M_n = \frac{1}{n}\sum_{i=1}^{n}\frac{g(Y_i)}{h(Y_i)}$$

Como podemos escolher h(i), podemos usar o método da transformada inversa para gerar $y_i \sim h(i)$

- Média e variância do novo estimador

$$E_h[M_n] = E_h\left[\frac{1}{n}\sum_{i=1}^{n}\frac{g(Y_i)}{h(Y_i)}\right] = \frac{1}{n}\sum_{i=1}^{n}\overbrace{E_h\left[\frac{g(Y_i)}{h(Y_i)}\right]}^{G_N} = G_N$$

$$Var_h[M_n] = \frac{\sigma_{g/h}^2}{n}$$

Variância da v.a. g(Y)/h(Y)

- M_n converge para $E_h[\, g(Y) \,/\, h(Y) \,] = G_N$
- Variância do estimador depende da variância de g(Y)/h(Y)

Variância da Nova v.a.

- Temos g(Y)/h(Y), onde *Y* possui distribuição *h*, com $h_i > 0$ para todo *i*

independe de h ← G_N^2

$$\sigma^2_{g/h} = Var_h\left[\frac{g(Y)}{h(Y)}\right] = E_h\left[\left(\frac{g(Y)}{h(Y)}\right)^2\right] - E_h\left[\frac{g(Y)}{h(Y)}\right]^2$$

$\sigma_x^2 = \mathbb{E}[x^2] - \mathbb{E}^2[x]$

$\rightarrow \sigma^2_{g/h} = \sum_{i=1}^{N} g(i)^2/h(i) - G_N^2$

$= G_N$

- Segundo momento

$$E_h\left[\left(\frac{g(Y)}{h(Y)}\right)^2\right] = \sum_{i=1}^{N}\left(\frac{g(i)}{h(i)}\right)^2 h(i) = \sum_{i=1}^{N}\frac{g(i)^2}{h(i)}$$

- E_h[g(Y)/h(Y)] não depende de *h*, mas segundo momento depende

O objetivo é encontrar um h que coloque massa onde g(i) é grande

  - variância de M_n depende apenas do segundo momento

Figueiredo 2018

Exemplo

assumption:

- Dado N, calcular $G_N = \sum_{i=1}^{N} g(i)$ onde $g(i) = i \log i$
- Seja Y_i seq iid de v.a. com distribuição $h(i) > 0$ para todo i
- **Opção 1**: $h(i) = 1/N$ para todo i, ou seja, h(i) é uniforme(1,N)
  - o que temos feito até agora
- Segundo momento?

$$\sum_{i=1}^{N} \frac{g(i)^2}{\underbrace{h(i)}_{= 1/N}} = N \sum_{i=1}^{N} g(i)^2 = N \sum_{i=1}^{N} i^2 \log^2 i$$

- Como reduzir segundo momento?
- **Ideia:** escolher h(i) proporcionalmente a g(i)
  - maiores valores tem maior probabilidade (com cuidado)

Exemplo

→ constante

- **Opção 2**: h(i) = i / K_2 , ou seja linearmente proporcional a *i*
  - onde K_2 = 1+2+...+N = N(N+1)/2 → K_2 deve ser tal que $\sum h(i) = 1$
- Segundo momento? $i^2 \log^2 i$

$$\sum_{i=1}^{N} \frac{g(i)^2}{h(i)} = K_2 \sum_{i=1}^{N} \frac{g(i)^2}{i} = K_2 \sum_{i=1}^{N} i \log^2 i$$

$h(i) = 1/K_2$

→ dou mais importância a valores mais altos

- **Opção 3**: h(i) = i^3 / K_3 , ou seja cúbica em *i*
  - onde K_3 = $1^3+2^3+...+N^3$ = $N^2(1 + N)^2/4$ → A constante agora é proporcional a N^4 !!

$$\sum_{i=1}^{N} \frac{g(i)^2}{h(i)} = K_3 \sum_{i=1}^{N} \frac{g(i)^2}{i^3} = K_3 \sum_{i=1}^{N} \frac{\log^2 i}{i}$$

→ Escondemos a sujeira aqui

Exemplo

- Qual é a melhor opção?
- A que tiver menor variância (ou menor segundo momento)
- Com N=1000, vamos calcular!

Opção 1

$$E_h\left[\left(\frac{g(Y)}{h(Y)}\right)^2\right]=1.44\,x\,10^{13}$$

Opção 2

$$E_h\left[\left(\frac{g(Y)}{h(Y)}\right)^2\right]=1.03\,x\,10^{13}$$

Opção 3

$$E_h\left[\left(\frac{g(Y)}{h(Y)}\right)^2\right]=2.75\,x\,10^{13}$$

- Melhor estimador é a **opção 2** (menor variância)
  - significa menos amostras para um mesmo erro
  - ou menos erro para um mesmo número de amostras
- Qual seria a melhor h(i) possível? → Queremos encontrar h(i) que induza a menor variância possível

O Melhor $h(i)$ Possível

- h(i) que induz variância zero é o melhor possível!

$$\sigma^2_{g/h} = E_h\left[\left(\frac{g(Y)}{h(Y)} - \mu_{g/h}\right)^2\right] = \sum_{i=1}^{N} \underbrace{\left(\frac{g(i)}{h(i)} - \mu_{g/h}\right)^2}_{\text{Forço que seja zero}} h(i)$$

Definição do valor esperado

$E[\, g(Y)/h(Y)\,] = G_N$

↳ constante, que não conhecemos!

- Se $\dfrac{g(i)}{h(i)} = \mu_{g/h}$ para todo *i*, então variância é zero!
- A princípio, podemos escolher qualquer *h*(*i*). Qual o problema então?
- Precisamos saber $\mu_{g/h}$ que é justamente o que queremos estimar! (Estamos usando o que queremos estimar)
- **Ideia**: tentar aproximar esta relação com heurísticas

uma delas é fazer h(i) proporcional a g(i)

Lembrando que o valor esperado também é um somatório

Generalização

- Supor que queremos calcular um determinado valor esperado, onde X tem distribuição dada por *f*

$$\mu_f = E_f[g(X)] = \sum_i g(i) f(i)$$

- Podemos aplicar *Importance Sampling* neste problema
  - amostrar mais regiões mais importantes para *g*
- Seja *h* outra distribuição para v.a. X, tal que f(i) > 0 → h(i) > 0

um somatório de uma f ξ vezes uma probabilidade é um valor esperado em cima

$$\mu_f = \sum_i \frac{g(i) f(i)}{h(i)} h(i) = E_h\left[\frac{g(X) f(X)}{h(X)}\right]$$

- Temos um outro valor esperado E_h que pode ser estimado
  - *X* em E_h tem distribuição *h*

1) Gero amostras $X \sim h(x)$

2) Aplico $\frac{g(x) f(x)}{h(x)}$

→ Agora a h casa com o produto g.f

- Podemos reduzir variância escolhendo *h*

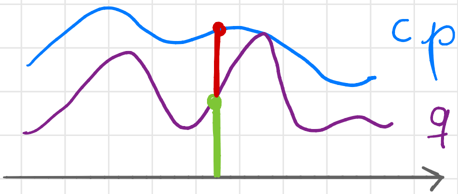
→ casando h(i) com g(i).f(i)

Generalização 2

- Tudo vale para o caso de v.a. contínua
  - trocar distribuição por densidade, somatório por integral
- Muitas aplicações no caso contínuo
  - *Monte Carlo Ray Tracing* não funciona sem *Importance Sampling*
  - Espaço de integração é muito grande para uniforme dar bons resultados
  - Muitas heurísticas são usadas nesta aplicação
- **Outro uso:** estimar eventos de baixa probabilidade
  - E[I(evento A)] = p_A , onde A é evento de interesse
  - Mesma ideia, reduzir variância do estimador, aumentar amostragem do evento de interesse

Importance Sampling

Técnica para gerar amostras de uma distribuição qualquer



$c\,p(x)$ → envelope para $q(x)$:

$q(x)$

$$q(x) \leq c \cdot p(x) \,, \quad c \geq 1$$

1. Gerar uma amostra i de p_x → $i \sim p_x$
2. Gerar uma amostra uniforme em $[0, c \cdot p(i)]$

$$u \sim \mathcal{U}\,[0, c \cdot p(i)]$$

3. Se $\begin{cases} u \leq q(i), & \text{retorna } i \\ u > q(i), & \text{repete o processo} \end{cases}$

Uma rodada é aceita com prob. $p = \frac{1}{c}$. Assim, na média, precisamos de $\frac{1}{p} = c$ rodadas para gerar uma amostra.

Importance Sampling

Técnica para gerar amostras de uma distribuição reduzindo-se a variância do estimador

Queremos gerar amostras de $g(x)$ mas podemos utilizar amostras de uma distribuição conhecida $h(x)$ tal que $h(x) \geq 0$ se $g(x) \geq 0$

Truque: modificar o valor esperado:

$$\mathbb{E}[g(x)] = \mathbb{E}\left[\frac{g(x)}{h(x)} \cdot h(x)\right] = \mathbb{E}_h\left[\frac{g(x)}{h(x)}\right]$$

Se $y_i \sim h(x) \longrightarrow \mathbb{E}[g(x)] = \mathbb{E}_h\left[\frac{g(x)}{h(x)}\right] \sim \frac{1}{n}\sum_{i=0}^{n}\frac{g(y_i)}{h(y_i)}$