Aula 9

Aula passada

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

Aula de hoje

- Gerando amostras complicadas
- Variância amostral
- Simulação

Gerando Amostras Complicadas

- Até agora, assumimos conhecimento da função de probabilidade, *f*(*i*)
    - ou ao menos de sua constante de normalização
- **Problema:** em muitos casos, não temos esta função

- Máquina caça níquel no cassino
- Você chega com 20 reais
- Cada rodada custa 1 real, e possivelmente dá recompensa
- Quantas rodadas de jogo até você perder tudo (inclusive o que ganhou)?
- Sim, você vai perder tudo se jogar o suficiente!

Amostras no Cassino

- Seja *T* v.a. que denota o número de rodadas até você perder tudo quando você chega com 20 reais
  - queremos calcular E[T]
- **Problema:** não temos função de probabilidade de *T*, ou seja $f(i) = P[T = i]$, para $i = 20,21,\ldots$.
  - f(20) = P[T = 20] = perder todas as vezes consecutivas
  - f(40) = P[T = 40] = ?

- Seja *R* v.a. que denota o retorno da máquina (possivelmente zero), com função de probabilidade *g*, ou seja g(j) = P[R = j]
- Seja *S* o valor total remanescente
- A cada rodada, valor total diminui de um e aumenta de *R*
- Rodadas são independentes

Amostras no Cassino

- Seja m o valor inicial (no caso, 20 reais)
- Podemos definir a soma iterativamente

$$S_k = m + \sum_{j=1}^{k} (R_j - 1)$$, onde R_j é a recompensa na rodada j

- E definir o valor para nossa v.a. *T*

$$T = min\{ k | m + \sum_{j=1}^{k} (R_j - 1) \leq 0 \}$$

  - primeira rodada onde a soma é menor ou igual a zero
- Como estimar E[T]?

Simular as rodadas!

  - sabemos gerar amostras para *R*, pois *g* é dada

Algoritmo

- Estimador de E[T] usando a média amostral

  T = 0;

  para i = 1, …, n

      S = m; t = 0;

      enquanto (S > 0)

          Gerar amostra *r* com distribuição *g*

          S = S + r – 1; t = t + 1;

      T = T + t;

  retorna T/n
- Convergência depende da variância de T
  - necessário para definir *n*, número de amostras
- Quanto vale $Var[T] = \sigma_T^2$?

Variância do Estimador

- Sabemos que estimador via média amostral tem variância

$$\hat{\mu}_T^n = 1/n \sum_{i=1}^{n} T_i \qquad Var[\hat{\mu}_T^n] = \frac{\sigma_T^2}{n}$$

- Não sabemos σ_T^2, mesmo se soubermos σ_R^2
- Podemos estimar σ_T^2 - *Monte Carlo to the rescue!*
- Definição

$$\sigma_T^2 = Var[T] = E[(T - \mu_T)^2]$$

- Estimador (ingênuo, veremos)

$$\hat{\sigma}_{T,n}^2 = 1/n \sum_{i=1}^{n} (T_i - \mu_T)^2$$

Mas não temos μ_T, mas podemos usar seu estimador!

Estimando a Variância

- Estimador (ingênuo, veremos)

$$\hat{\sigma}^2_{T,n} = 1/n \sum_{i=1}^{n} (T_i - \hat{\mu}^n_T)^2$$

Usando o estimador para valor esperado

- Qualidade do estimador: se for sem viés, seu valor esperado é igual ao valor sendo estimado para todo *n*, no caso σ_T^2

$$E[\hat{\sigma}^2_{T,n}] = E[1/n \sum_{i=1}^{n} (T_i - \hat{\mu}^n_T)^2] = 1/n \sum_{i=1}^{n} E[(T_i - \hat{\mu}^n_T)^2]$$

$$= 1/n \sum_{i=1}^{n} E[(T_i^2 - 2 T_i \hat{\mu}^n_T) + (\hat{\mu}^n_T)^2] = \ldots = \frac{n-1}{n} \sigma^2_T$$

- Este estimador é *enviesado*! Mas viés é de fácil correção
  - fator multiplicativo de n/(n-1) corrige o viés

Variância Amostral

- Estimador para variância sem viés (*sample variance*)

$$\hat{s}^2_{T,n} = \frac{n}{n-1}\hat{\sigma}^2_{T,n} = \frac{1}{n-1}\sum_{i=1}^{n}\left(T_i - \hat{\mu}^n_T\right)^2$$

- Que pode ser calculado mais facilmente como

$$M_1 = \sum_{i=1}^{n} T_i \qquad M_2 = \sum_{i=1}^{n} \left(T_i\right)^2$$

$$\hat{s}^2_{T,n} = \frac{M_2 - M_1^2/n}{n-1} \qquad \hat{\mu}^n_T = \frac{M_1}{n}$$

- Variância amostral de *T* pode ser usada para calcular variância do estimador μ_T^n, e com isto definir *n*

Gerando Amostras Complicadas

- Primeiro lance na sinuca
- Ver vídeo: https://www.youtube.com/watch?v=iR8BClwp5fE
- Aleatoriedade: posição da bola branca, local de contato, ângulo do taco, velocidade do taco

- Todo o resto é física Newtoniana (determinístico)
- Quantas bolas são encaçapadas? Valor esperado?

Simular o sistema
(até bolas pararem)

Simulação de Sistemas

- **Simulador:** gerador sofisticado de amostras que são difíceis de gerar!
  - simular comportamento de sistema complicado a partir de seus componentes básicos
- Método de Monte Carlo é a teoria que sustenta simulação
  - Geradores de variáveis aleatórias, estimadores baseados em média amostral e variância amostral
- Técnica amplamente utilizada para avaliar sistemas com eventos aleatórios
  - lógica depende do simulador depende do domínio, não é muito fácil generalizar