Aula 4

Aula passada

- Função de distribuição
- Bernoulli
- Sequência de v.a.
- Binomial, Geométrica, Zeta
- Valor esperado
- Variância
- Distribuição conjunta
- Independência de v.a.

Aula de hoje

- Valor esperado condicional
- Espaço amostral contínuo, função densidade
- Limitantes para probabilidade
- Desigualdades de Markov, Chebyshev, Chernoff
- *with high probability*

Esperança Condicional

- Seja *X* e *Y* duas v.a. definidas sobre um mesmo espaço amostral

$$f_{X|Y}(i,j)=P[X=i|Y=j]$$

Distribuição condicional de *X* dado *Y*

- Valor esperado condicional
  - valor esperado restrito a um subconjunto do espaço amostral (definido pelo valor de outra v.a.)

$$E[X|Y=j]=\sum_{i \in O_X} i\, f_{X|Y}(i,j)$$

  - segue definição de valor esperado (nada de novo)!

Esperança Condicional

- Muito útil para calcular valores esperados
  - uso similar a prob conditional para calcular marginal
- Seja E[X | Y= j] uma função da v.a. Y
  - repare que E[X|Y] é uma v.a. ao considerar j aleatório
- Temos que

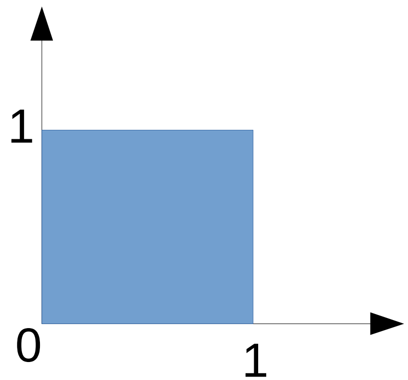
$$E[X] = E[E[X|Y]] \longleftarrow$$

propriedade da torre da esperança

- Exemplo: S = $N_1 + N_2 + \ldots + N_K$
  - parcela N_i~Bin(n,p), iid, e K~Geo(q)
  - número de parcelas é aleatório
- E[S] = ?

Espaço Amostral Contínuo

- Até agora, espaço amostral era discreto (contável)
- Espaço amostral contínuo (não contável)
  - ex. números reais, pontos do quadrado [0,1]x[0,1], etc



Problema: Como dar probabilidade a cada ponto do espaço amostral?

Solução: Dar probabilidade a “subconjuntos” do espaço amostral

- pedaços contíguos do espaço amostral tem uma “densidade” de probabilidade
- eventos representam tais pedaços

Função de Densidade

- Seja *A* um evento qualquer de *S*
- Dizemos que f(x) é uma função densidade sse

$$P[A]=\int_A f(x)\,dx$$

Área da função densidade dentro do espaço definido por *A*

- Mesma restrições que antes

$$0\le\int_A f(x)\,dx\le 1 \qquad \int_S f(x)\,dx=1$$

- Exemplo: *S* = [*a*, *b*], para *a*, *b* constantes

$$f(x)=\frac{1}{b-a},\, x\in[a,b] \qquad P[A]=\int_{a'}^{b'} f(x)\,dx=\frac{b'-a'}{b-a}$$

- *A* = [*a'*, *b'*] com *a'*>= *a*, *b'*<=*b*

Segue Tudo

- Todos os conceitos são equivalentes
  - independência, exclusão mútua, prob. totais, etc
  - trocar somatório por integral
- Definir v.a. para *S* contínuo

$X : S \rightarrow \mathbb{R}$ ← v.a. contínua, com imagem não contável

- Todos os conceitos sobre v.a. são equivalentes
  - função de distribuição (chamada de densidade)
  - valor esperado, variância, propriedades
  - distribuição conjunta, etc
  - trocar somatório por integral nas definições, usando CDF quando necessário

Limitantes para Probabilidade

- Calcular probabilidade de um evento pode ser difícil
  - analiticamente intratável
  - computacionalmente intratável
- Calcular um limitante inferior ou superior para probabilidade pode ser mais fácil

$$P[A] \leq U_A \longleftarrow U_A \text{ é um limitante superior}$$

$$P[A] \geq L_A \longleftarrow L_A \text{ é um limitante inferior}$$

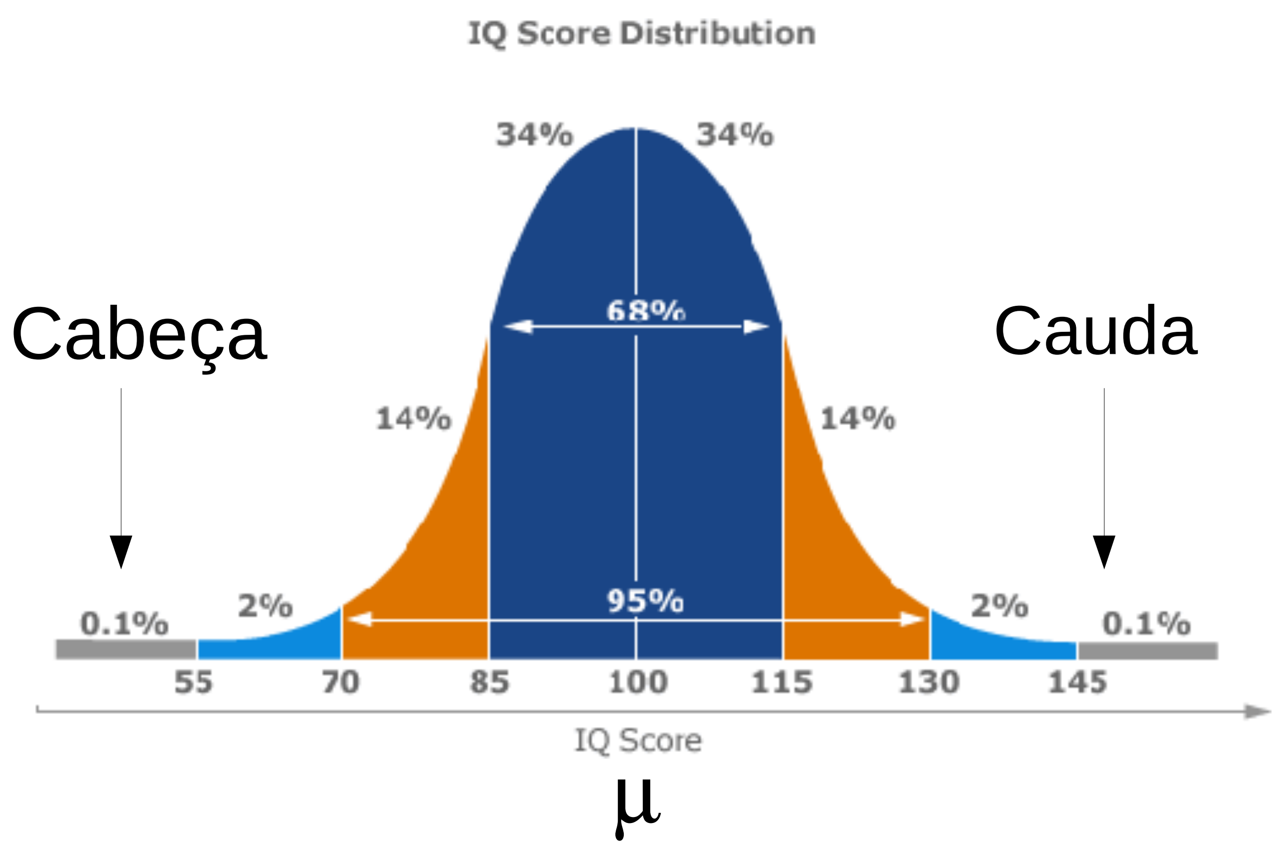
↳ variável aleatória

- Em geral, estamos interessados na probabilidade da cauda ou cabeça da distribuição

↳ Queremos saber dos eventos raros, que não acontecem com tanta frequência

Cauda e Cabeça

- Seja *X* uma v.a com μ = E[X]
- Cauda: valores de X bem maiores que μ
- Cabeça: valores de X bem menores que μ



prob. do evento ser k vezes maior do que a média

- Exemplos
- P[X > *k*μ] : prob. da cauda, *k* > 1
- P[X < *k*μ] : prob. da cabeça, *k* < 1
- Probabilidade de eventos extremos, mais raros

Figueiredo 2018

Exemplo

- Jogar um dado honesto de 10 faces 50 vezes
- N = número de vezes que o resultado foi um primo
- X_i = resultado do dado na i-ésima rodada $\in [1, 10]$

$$\underbrace{N}_{\text{Binom.}}=\sum_i \overbrace{I}^{\text{Bernoulli}}(X_i)$$ ← v.a. indicadora de número primo (vale 1 quando argumento é primo)

Espaço Amostral: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10} ↳ Eventos de interesse

$$P[N \geq 40] = ?$$

- Qual é a distribuição de N?
- $P[X_i = 1] = 2/5 = 4/10$
- N~Bin(50, 2/5)

Soma de Indicadoras é uma binomial

Como os eventos são mutuamente exclusivos, podemos somá-los

$$P[N \geq 40] = \sum_{i=40}^{50} P[N=i] = \sum_{i=40}^{50} \underbrace{\binom{50}{i}\left(\frac{2}{5}\right)^i \left(\frac{3}{5}\right)^{50-i}}_{P[N=i,p]}$$

- difícil calcular coeficientes de Newton
- somatório poderia ser muito longo

Figueiredo 2018

Desigualdade de Markov

- Importante limitante superior para probabilidade de um evento Relativamente fácil de usar
  - relação entre valor esperado e probabilidade
- Para qualquer v.a. *X* não negativa e constante a>0, temos: evento

$$\underbrace{P[X \geq a]}_{\text{evento}} \leq \frac{E[X]}{\underbrace{a}_{U_L}}$$

Só faz sentido para *a* > E[*X*], *a* está na cauda da distrib.

Se $a < \mathbb{E}[x]$, $\mathbb{E}[x]/a > 1$, então a desigualdade não nos ajuda muito

- Prova
  - I(X >= a) : v.a. indicadora do evento X >= a
  - Então aI(X >=a) <= X
  - Aplicando esperança dos dois lados, temos E[aI(X >=a)] <= E[X] $= a \cdot \mathbb{E}[I(X \geq a)] = 1 \cdot P[X \geq a] + 0 \cdot P[X < a]$

= P[X >= a] <= E[X] / a

Voltando ao Exemplo

- N~Bin(50, 2/5) $$P[N \geq 40] = \sum_{i=40}^{50} \binom{50}{i} \left(\frac{2}{5}\right)^{i} \left(\frac{3}{5}\right)^{50-i}$$

- Podemos aplicar desigualdade de Markov

$$P[X \geq a] \leq \frac{E[X]}{a}$$

Para Binom(n, p), $\mathbb{E}[X] = np$

- E[N] = 50*2/5 = 20

$$P[N \geq 40] \leq \frac{20}{40} = \frac{1}{2}$$

← Chance de ver 40 ou mais primos é menor do que 1/2

Não sabemos qual é a probabilidade, mas sabemos que é menor que 1/2

Desigualdade de Chebyshev

- Outro importante limitante superior para probabilidade de um evento
  - relação entre valor esperado, variância e probabilidade
- Mais precisa que desigualdade de Markov
  - Markov foi aluno de Chebyshev (russos)
- Para qualquer v.a. *X* com valor esperado μ e variância σ^2, e qualquer $k>0$, temos

Probabilidade da V.A. estar k desvios padrões fora da média

$$P[\underbrace{|X-\mu|}_{\text{o quanto estou longe da média}} \geq k\sigma] \leq \frac{1}{k^2}$$

← prob de *X* estar *k* desvios padrão longe da média

- Prova
  - $Y = (X - \mu)^2 \quad a = (k\sigma)^2$
  - Aplicar desigualdade de Markov usando *Y* e *a*

Caso Interessante

- Se $k=\sqrt{2}$, então temos:

Qualquer V.A. concentra sua massa em 1.41 desvios padrões da média

$$P[|X-\mu|\geq 1.41\,\sigma]\leq \underbrace{\frac{1}{2}}_{UL}$$

- Probabilidade de X estar fora do intervalo [μ – 1.41σ , μ + 1.41σ] é menor do que ½
  - vale para qualquer distribuição da v.a.

Voltando ao Exemplo

- N~Bin(50, 2/5) $$P[N \ge 40] = \sum_{i=40}^{50} \binom{50}{i} \left(\frac{2}{5}\right)^{i} \left(\frac{3}{5}\right)^{50-i}$$
- Podemos aplicar desigualdade de Chebyshev

$$P[|X-\mu| \ge k\sigma] \le \frac{1}{k^2}$$

n × p   n p (1 - p)

- μ = 50*(2/5) = 20 , σ^2 = 50*(2/5)*(3/5) = 12

$\sigma = 2\sqrt{3}$

- {N >= 40} = {N – μ >= 20}

10

- kσ = 20 → k = ~~5~~ 10/sqrt(3)

$$P[N \ge 40] \le P[|N-\mu| \ge 20] \le \frac{1}{(\cancel{5}\,10/\sqrt{3})^2} = \frac{3}{\cancel{25}\,100} = 0.\cancel{12}\,03$$

$P[N - 20 \geqslant 20] \leq P[|N-20| \geqslant 20]$

- Resultado melhor que por Markov!

Desigualdade de Chernoff

- Limitante superior para para soma de v.a. independentes

$$X = Y_1 + \ldots + Y_n$$

Não se aplica p/ qualquer v.a.

- Resultado muito importante e muito usado
  - perfeito para Binomial (soma de Bernoulli)
  - muitas variações das desigualdades (mais fáceis de usar)
- Seja Y_i ~Bern(p), μ = E[X] = np e qualquer $\delta > 0$, temos

fator 1+δ após a média

→ preciso somente de μ, mas estou limitado à condição anterior

$$P[X \geq \overbrace{(1+\delta)}\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

← prob da cauda (depois da média)

cauda

$$P[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$$

← prob da cabeça (antes da média)

cabeça

Voltando ao Exemplo

- N~Bin(50, 2/5) $$P[N \geq 40] = \sum_{i=40}^{50} \binom{50}{i} \left(\frac{2}{5}\right)^{i} \left(\frac{3}{5}\right)^{50-i}$$
- Podemos aplicar desigualdade de Chernoff

$$P[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

Consigo um limitante melhor pois Chernoff foi feito para soma de V.A. independentes

- μ = 50*(2/5) = 20
- $(1+\delta)\mu$ = 40 → δ = 1

$$P[N \geq (1+1)20] \leq \left(\frac{e^{1}}{(1+1)^{1+1}}\right)^{20} = \frac{e^{20}}{2^{40}} = 0.00044$$

- Resultado bem melhor que por Chebyshev!

Figueiredo 2018

Desigualdade de Chernoff

- Considere Y_i resultado de moeda honesta é cara, iid

$$X = Y_1 + \ldots + Y_n$$

- X = número de caras em *n* jogadas, $\mu = n/2$
- Onde está quase toda a “massa” da distribuição?
  - prob. da cauda vai a zero com *n*
- Ou seja, qual o valor de λ tal que $P[X > \mu + \lambda] < 1/n$

$$P[X \geq (1+\delta)\mu] \leq e^{-\delta^2 \mu/3}$$ ← variação da desigualdade de Chernoff

- $(1+\delta)\mu = \mu + \lambda \;\rightarrow\; \delta = \lambda/\mu$

$$P[X \geq \mu + \lambda] \leq e^{-\left(\frac{\lambda}{\mu}\right)^2 \mu/3} = e^{-\frac{2\lambda^2}{3n}} = 1/n$$

$$\lambda = \sqrt{3/2\, n \ln n}$$

With High Probability (whp)

- Seja *n* um parâmetro de um modelo probabilístico
  - ex. número de rodadas de um dado, vértices no grafo G(n,p)
- Seja *A*(*n*) um evento no respectivo espaço amostral → Quando determino n, tenho um espaço-amostral
- *A*(*n*) ocorre *with high probability* (whp) quando

Se mudo n, mudo o espaço-amostral

$$P[A(n)] \geq 1 - \frac{1}{n^{\alpha}}$$ ← para algum α > 1 constante

- Repare que $\lim_{n \to \infty} P[A(n)] = 1$ ← convergência em probabilidade

Evento tende a ocorrer sempre

- Do exemplo anterior temos:

$$X \leq \frac{n}{2} + \sqrt{3/2\, n \ln n} = \frac{n}{2} + \theta\left(\sqrt{n \ln n}\right)$$, w.h.p.

  - ou seja, número de caras será praticamente sempre menor do que a média + sqrt(*n* log *n*), ao jogar *n* vezes

Voltando ao Exemplo

- Se $\lambda = \sqrt{3/2\, n \ln n}$
- Então $P[X \geq \mu + \lambda] \leq 1/n$
- Exemplo: *n*=1000 (lançar moeda 1000 vezes)
  - μ = 500
  - $\lambda = \sqrt{3/2\, n \ln n} = \sqrt{1500 \ln 1000} = 101.8$
- Então

$$P[X \geq 500 + 102] = P[X \geq 602] \leq 0.001$$

- Ou seja, observar 602 caras ou mais é bastante raro, ao jogar uma moeda honesta 1000 vezes
- Podemos apostar com bastante segurança!

Desigualdade de Markov: $P[X \geq a] \leq \underbrace{\frac{\mathbb{E}[X]}{a}}_{\text{Limitante superior da prob.}}$

Só faz sentido se $a \geq \mathbb{E}[X]$
(prob. da cauda)

Desigualdade de Chebyshev: $P[|X-\mu| \geq k \cdot \sigma] \leq \frac{1}{k^2}$

Prob. do evento acontecer a k desvios padrões da média

Desigualdade de Chernoff: Apenas para $Y = X_1 + X_2 + \ldots + X_N$
onde X_i são independentes

$$\underbrace{P[X \geq (1+\delta) \cdot \mu]}_{\text{Cauda}} \leq \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$$

$$\underbrace{P[X \leq (1-\delta)\mu]}_{\text{Cabeça}} \leq \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu}$$