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

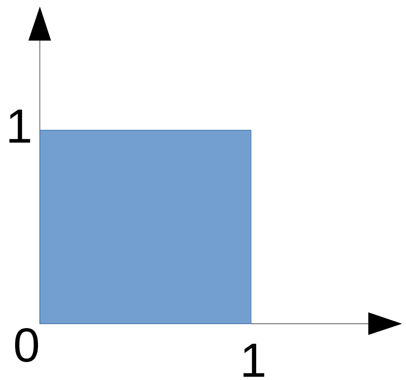
$$E[X]=E[E[X|Y]]$$

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

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

- 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

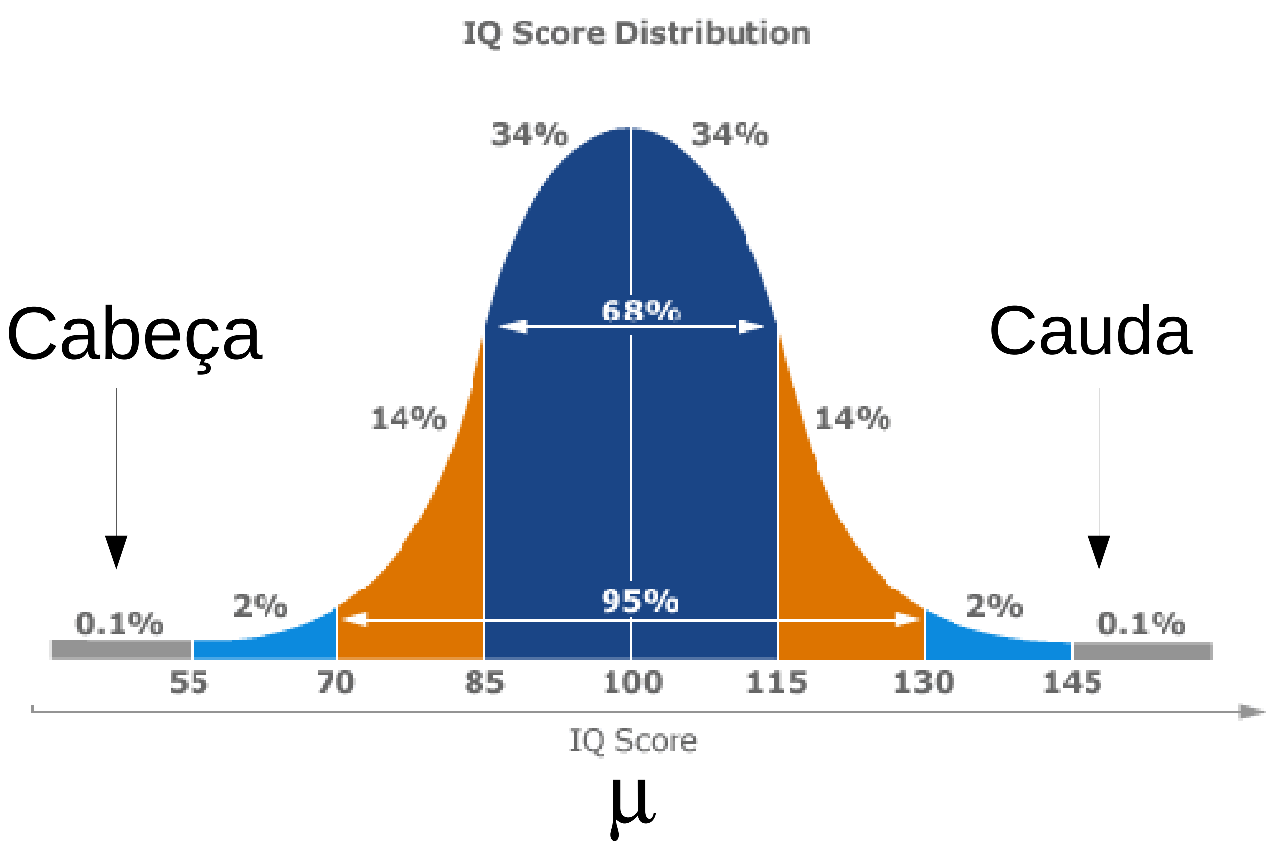
$P[A] \leq U_A$ ← U_A é um limitante superior

$P[A] \geq L_A$ ← L_A é um limitante inferior

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

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



- 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

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

$$N=\sum_i I(X_i) \longleftarrow$$

v.a. indicadora de número primo
(vale 1 quando argumento é primo)

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

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

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

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

Desigualdade de Markov

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

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

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

- 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]

    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

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

- 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

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

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

$\longleftarrow$ prob de *X* estar *k* desvios padrão longe da média

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

Caso Interessante

- Se $k=\sqrt{2}$, então temos:

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

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

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

- μ = 50*(2/5) = 20 , σ^2 = 50*(2/5)*(3/5) = 12
- {N >= 40} = {N – μ >= 20}
- kσ = 20 → k = 5/sqrt(3)

$$P[N \geq 40] \leq P[|N-\mu| \geq 20] \leq \frac{1}{(5/\sqrt{3})^2} = \frac{3}{25} = 0.12$$

- Resultado melhor que por Markov!

Desigualdade de Chernoff

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

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

- 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

$$P[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu}$$ ← prob da cauda (depois da média)

$$P[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \right)^{\mu}$$ ← prob da cabeça (antes da média)

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

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

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

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

- Resultado bem melhor que por Chebyshev!

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, μ = 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 > μ + λ] < 1/n

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

- (1+δ)μ = μ + λ → δ = λ/μ

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

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

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
- *A*(*n*) ocorre *with high probability* (whp) quando

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

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

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

  - 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!