Probabilidade II
Aula 12

Junho de 2009

Mônica Barros, D.Sc.

Conteúdo

- Desigualdade de Markov
- Desigualdade de Chebyshev
- Desigualdade de Jensen
- Lei Fraca dos Grandes Números

Introdução

- A variância de uma variável aleatória mede a dispersão em torno da média.
- A desigualdade de Chebyshev nos fornece uma maneira de entender como a variância mede a dispersão em torno da média, e nos permite encontrar limites superiores e inferiores para certas probabilidades.

Introdução

- Estes limites não são necessariamente próximos dos valores reais das probabilidades, e são usados principalmente em discussões teóricas, e não como aproximações.
- A desigualdade de Markov é um resultado teórico importante que, entre outras coisas, nos ajuda a demonstrar a desigualdade de Chebyshev.

Introdução

- Estas duas desigualdades nos fornecem limites superiores e inferiores para probabilidades quando apenas a média (Markov) é conhecida ou quando só a média e a variância são dadas (Chebyshev).
- Em geral, estes limites para as probabilidades não são eficientes do ponto de vista computacional, mas também os requisitos necessários para o seu cálculo são mínimos, pois precisamos apenas conhecer no máximo a média e a variância.

Desigualdade de Markov

- Teorema
- Seja u(X) uma função não negativa da variável aleatória X.
- Se E [u(X)] existe, então para qualquer constante positiva <u>c</u> temos:

$$\Pr\left(u(X) \geq c\right) \leq \frac{E\left(u\left(X\right)\right)}{c}$$

Desigualdade de Markov

- Demonstração – caso contínuo
- Seja A o conjunto{x : u(x) ≥ c}. Então o valor esperado de u(X) pode ser escrito como:

$$E\left(u\left(X\right)\right) = \int_{-\infty}^{\infty} u(x) f(x) dx = \int_A u(x) f(x) dx + \int_{A^*} u(x) f(x) dx$$

- onde A* é o complemento de A , ou seja: A* = { X : u(X) < c}

Desigualdade de Markov

- Cada uma das integrais do lado direito acima é não negativa, e então o lado esquerdo da equação é maior ou igual a cada uma destas integrais. Em particular:

$$E\left(u\left(X\right)\right) \geq \int_A u(x) f(x) dx$$

- Mas, pela definição do conjunto A , se X ∈ A então u (X) ≥ c. Assim:

$$E\left(u\left(X\right)\right) \geq \int_A u(x) f(x) dx \geq c \int_A f(x) dx = c.\Pr(X \in A) = c.\Pr\left(u(X) \geq c\right)$$

Desigualdade de Markov

- Isto é:
  - E [u (X)] ≥ c. Pr {u (X) ≥ c }
  - Pr{ u (X) ≥ c } ≤ E [u (X)] /c

- Note que, na demonstração da desigualdade de Markov a única hipótese é que E [u(X)] existe e é uma função não negativa.

Desigualdade de Markov

- Exemplo 1
- Seja Y uma v.a. tal que $E(Y^4) \leq 100$. Use esta informação para encontrar um limite superior para Pr(Y ≥ 5).
- Solução
  - Seja $W = Y^4$. Então W é uma v.a. não negativa cuja média é, no máximo, 100. Note que se Y ≥ 5 então $W = Y^4 \geq 625$. Pela desigualdade de Markov:

$$\Pr(Y \geq 5) = \Pr(W \geq 625) \leq \frac{E(W)}{625} = \frac{100}{625}$$

Desigualdade de Chebyshev

- A desigualdade de Chebyshev pode ser encarada como um corolário da desigualdade de Markov.

- Seja X uma v.a. qualquer com média μ e variância σ^2, ambas finitas. Então:

$$\Pr\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Desigualdade de Chebyshev

- Demonstração
- Como $u(X) = (X-\mu)^2$ é uma função não-negativa, podemos aplicar a desigualdade de Markov e obter:

$$\Pr(u(X) \geq k) = \Pr\left((X-\mu)^2 \geq k^2\right) \leq \frac{E(u(X))}{k^2} = \frac{E\left((X-\mu)^2\right)}{k^2} = \frac{\sigma^2}{k^2}$$

- Mas, os eventos:

$$(X-\mu)^2 \geq k^2 \quad \text{e} \quad |X-\mu| \geq k$$

- São equivalentes, o que prova a desigualdade de Chebyshev.

Desigualdade de Chebyshev

- Analogamente podemos escrever:

$$\Pr\left(\left|X-\mu\right|\le k\right)\ge 1-\frac{\sigma^2}{k^2}$$

- A desigualdade de Chebyshev pode ser escrita de maneira mais conveniente em termos da distância (em unidades do desvio padrão) em relação à média, ou seja:

Desigualdade de Chebyshev

- A probabilidade de X estar a uma distância "grande" (maior que k desvios padrões) da sua média é pequena (menor que $1/k^2$).

$$\Pr\left(\left|X-\mu\right|\ge k.\sigma\right)\le \frac{1}{k^2}$$

- A probabilidade de X estar a uma distância "pequena" (menor que k desvios padrões) da sua média é grande (maior que $1- 1/k^2$).

$$\Pr\left(\left|X-\mu\right|\le k.\sigma\right)\ge 1-\frac{1}{k^2}$$

Desigualdade de Chebyshev

- Os limites fornecidos pelas desigualdades de Markov e Chebyshev são, como já dissemos, “grandes demais” para uso como aproximações numéricas.

- Suas maiores virtude são as de funcionar sob condições muito pouco restritivas, exigindo apenas o conhecimento da média (Markov) ou da média e variância (Chebyshev).

Desigualdade de Chebyshev

- Exemplo 2
- Comparação de uma probabilidade exata com o limite dado por Chebyshev.
- Seja X uma v.a. Unif($-\sqrt{3}$, $+\sqrt{3}$).
- Verifique que E(X) = 0 e VAR(X) =1
- Calcule exatamente:

$$\Pr\left\{\left|X\right|\ge\frac{3}{2}\right\}=1-\Pr\left\{\left|X\right|\le\frac{3}{2}\right\}=1-\int_{-3/2}^{+3/2}\frac{1}{2\sqrt{3}}dx=1-\frac{3}{2\sqrt{3}}=0.1340$$

Desigualdade de Chebyshev

- Por Chebyshev, k =3/2 e o limite superior para esta probabilidade torna-se (vide slide 14):

$$\Pr\left(\left|X-0\right| \geq \frac{3}{2}.(1)\right) \leq \frac{1}{(3/2)^2} = \frac{4}{9} = 0.4444$$

- Ou seja, o limite dado por Chebyshev é mais de 3 vezes maior que o valor real neste caso.

Desigualdade de Chebyshev

- Exemplo 3
- Neste caso a distribuição da v.a. é desconhecida.
- O número de TVs produzido por uma fábrica num intervalo de uma semana é uma v.a. com média 500.
- 1) Encontre um limite para a probabilidade do número de TVs produzido esta semana exceder 1000.

Desigualdade de Chebyshev

- Neste caso usamos a desigualdade de Markov, pois só conhecemos a média. Note que a variável de interesse (número de TVs produzidas é intrinsecamente não negativa, justificando o uso de Markov).

$$\Pr(X \geq 1000) \leq \frac{E(X)}{1000} = \frac{500}{1000} = \frac{1}{2}$$

Desigualdade de Chebyshev

- 2) Suponha agora que conhecemos também a variância do número de TVs produzidas na semana, que é igual a 100.
- O que se pode dizer sobre a probabilidade do número de TVs produzidas esta semana estar entre 400 e 600?
- Pela desigualdade de Chebyshev , e usando μ = 500, σ^2 = 100 temos:

$$\Pr\{\left|X-500\right| \leq 100\} = \Pr\{400 \leq X \leq 600\} \geq 1 - \frac{100}{(100)^2} = 1 - \frac{1}{100} = \frac{99}{100}$$

Desigualdade de Chebyshev

- Exemplo 4
  - Seja X uma variável Expo(1). Use a desigualdade de Markov para provar que Pr(X >c) ≤ 1/c onde c é um número positivo qualquer.
- Solução
  - Note que a desigualdade de Markov pode ser usada diretamente em X, que é uma variável não negativa.
  - Também, E(X) = 1. Por Markov:
  - Pr(X > c) ≤ E(X)/c = 1/c

Desigualdade de Chebyshev

- A extensão deste exemplo para uma variável Exponencial qualquer é trivial.
- Se X é Exponencial(λ) então:

$$\Pr(X > c) \leq \frac{E(X)}{c} = \frac{1}{\lambda . c}$$

- Para qualquer constante positiva c.
- Note que isso é realmente verdadeiro, pois:

$$\Pr(X > c) = 1 - F(c) = 1 - \{1 - e^{-\lambda c}\} = e^{-\lambda c} = \frac{1}{e^{+\lambda c}} = \frac{1}{1 + (\lambda c) + \frac{(\lambda c)^2}{2!} + \frac{(\lambda c)^3}{3!} + \ldots} \leq \frac{1}{\lambda c}$$

Desigualdade de Chebyshev

- Exemplo 5
- Aplicação a uma sequência de Bernoullis.
- Sejam $X_1, X_2, \ldots, X_n$ iid Bernoulli(p), de tal forma que $Y = X_1 + X_2 + \ldots + X_n$ é Binomial (n, p).
- Pela desigualdade de Chebyshev:

$$\Pr\{|Y - np| \geq k\} \leq \frac{VAR(Y)}{k^2} = \frac{npq}{k^2}$$

Desigualdade de Chebyshev

- Mas, a função n.p.q = n.p.(1-p) tem um máximo em p=1/2 (faça o gráfico) e então podemos garantir que:

$$\Pr\{|Y - np| \geq k\} \leq = \frac{n}{4k^2}$$

- Considere agora a proporção de sucessos nas n repetições, isto é: $\hat{p} = \frac{Y}{n}$

Desigualdade de Chebyshev

- Pelas propriedades das funções lineares de v.a., pode-se provar que:

$$E(\hat{p}) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = \frac{np}{n}$$

$$VAR(\hat{p}) = VAR\left(\frac{Y}{n}\right) = \frac{1}{n^2}VAR(Y) = \frac{npq}{n^2} = \frac{pq}{n}$$

- Aplicando Chebyshev a p^ segue que:

Desigualdade de Chebyshev

$$\Pr\{|\hat{p} - p| \geq k\} \leq \frac{VAR(\hat{p})}{k^2} = \frac{pq}{nk^2} \leq \frac{1}{4nk^2}$$

- Esta última expressão tem implicações importantes em amostragem.
- Por exemplo, se n > 1000, a probabilidade de p^ diferir do valor verdadeiro de p por mais de 0.1 é, no máximo, 0.025 para qualquer valor de p.

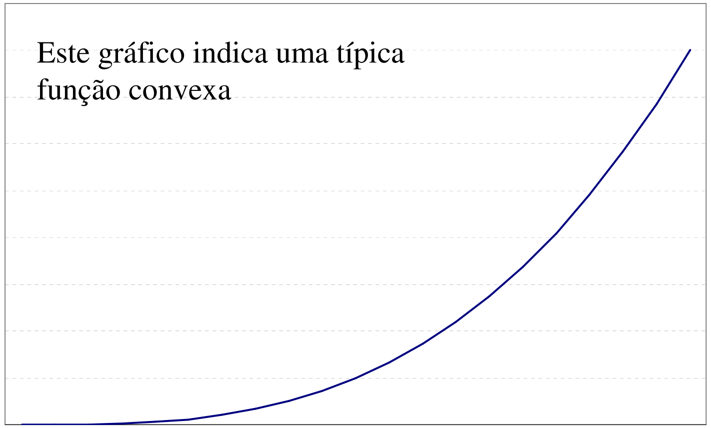
Desigualdade de Chebyshev

- Exemplo 6
- Com base na expressão do slide anterior, calcule n (tamanho da amostra) necessário para que a probabilidade da diferença entre o p real e o estimado ser maior que 3% seja menor ou igual a 5%.
- Solução
- Aqui k = 0.03 e $1/(4nk^2) = 5\%$
- Logo: $1/4n(3/100)^2 = 10000/36n = 5/100$
- $n = 10^6/180 = 5555$ aproximadamente

Desigualdade de Jensen

- Antes de enunciar esta desigualdade, é preciso lembrar o que são funções convexas.
- Uma função f(x) é convexa se:
- $f\{t.x + (1-t).y\} \leq t.f(x) + (1-t).f(y)$ para $0 \leq t \leq 1$ e x e y num intervalo [a,b].
- Se acima a desigualdade é estrita dizemos que f é estritamente convexa.

Desigualdade de Jensen



Desigualdade de Jensen

- A desigualdade de Jensen pode ser escrita como:

$$E\{f(X)\} \geq f\{E(X)\}$$

alternativamente

$$f\{E(X)\} \leq E\{f(X)\} \quad \text{se f é uma função convexa}$$

- A igualdade ocorre se f não é estritamente convexa ou se X tem uma distribuição degenerada (i.e, X tem toda a probabilidade num único ponto).

Desigualdade de Jensen

- A demonstração será omitida.
- A desigualdade de Jensen será importante em Inferência Estatística para provar alguns resultados, por exemplo, o Teorema de Rao-Blackwell.

- Uma aplicação interessante segue.

Desigualdade de Jensen

- Exemplo 7
- Use a desigualdade de Jensen para provar que a variância de uma v.a. é sempre não negativa.
- Solução
- A função $f(x) = x^2$ é convexa. Assim, pela desigualdade de Jensen:
- $E(X^2) \geq \{ E(X)\}^2$ □ $E(X^2) - \{ E(X)\}^2 \geq 0$ □ $VAR(X) \geq 0$

Lei Fraca dos Grandes Números

- A média μ de uma distribuição pode ser encarada como a média empírica de X no longo prazo após um número muito grande de repetições.

- Sejam $X_1, X_2, ..., X_n$ iid com média μ e variância σ^2, ambas finitas.
- Mostraremos que a média amostral "convergirá" para μ no sentido indicado a seguir.

Lei Fraca dos Grandes Números

- Note que a soma $S_n = X_1 + X_2 + ... + X_n$ é uma v.a. com média $n.\mu$ e variância $n.\sigma^2$ e assim ambas a média e a variância de S_n crescem à medida que tomamos mais termos na soma.

- A média e a variância da média amostral são:

$$E(\overline{X}_n) = E\left(\frac{S_n}{n}\right) = \frac{n.\mu}{n} = \mu \qquad VAR(\overline{X}_n) = VAR\left(\frac{S_n}{n}\right) = \frac{n.\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Lei Fraca dos Grandes Números

- Definição – Convergência em Probabilidade
- Dizemos que uma sequência de v.a. $Y_1, Y_2, ..., Y_n$ converge em probabilidade para outra v.a. Y quando n tende a infinito se:

$$\lim \Pr(|Y_n - Y| \geq \varepsilon) = 0 \quad \text{para todo } \varepsilon > 0$$
equivalentemente:
$$\lim \Pr(|Y_n - Y| < \varepsilon) = 1 \quad \text{para todo } \varepsilon > 0$$

Onde os limites se referem a n tendendo a infinito

Lei Fraca dos Grandes Números

- Em outras palavras:
- Y_n converge em probabilidade para Y se n tende a infinito se, e somente se, Y_n está arbitrariamente perto de Y com uma probabilidade tão grande quanto necessário para n suficientemente grande.

- Em muitos casos (como na "lei fraca" a seguir), a convergência em probabilidade será para uma constante.

ENCE
Escola Nacional de Ciências Estatísticas
IBGE

Lei Fraca dos Grandes Números

- Teorema (Lei Fraca dos Grande Números)
- Sejam $X_1, X_2, ..., X_n$ iid com média μ e variância σ^2, ambas finitas.
- Então, para qualquer n ≥ 1:

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

- Converge em probabilidade para $\mu = E(X_i)$ quando n tende a infinito.

ENCE
Escola Nacional de Ciências Estatísticas
IBGE

Lei Fraca dos Grandes Números

- Demonstração
  - Segue direto da desigualdade de Chebyshev e da definição de convergência em probabilidade.

$$\Pr\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \leq \frac{VAR\left(\overline{X}_n\right)}{\varepsilon^2} = \frac{\sigma^2}{n.\varepsilon^2} \to 0 \text{ quando n} \to \infty$$

ENCE
Escola Nacional de Ciências Estatísticas
IBGE

Lei Fraca dos Grandes Números

- Exemplo 8
  - Sejam $X_1, X_2, ..., X_n$ iid Unif(0,1). Seja $X_{(n)}$ o máximo de $X_1, X_2, ..., X_n$. Mostre que $X_{(n)}$ converge em probabilidade para 1 quando n tende a infinito.
  - Solução
  - Note que $\Pr(X_{(n)} > 1) = 0$
  - Também: $\Pr(X_{(n)} \leq 1 - \varepsilon) = \{\Pr(X_i \leq 1 - \varepsilon)\}^n =$ $= (1 - \varepsilon)^n$ para $0 < \varepsilon < 1$ e então o limite desta probabilidade quando $n \to \infty$ é zero.

ENCE
Escola Nacional de Ciências Estatísticas
IBGE

Lei Fraca dos Grandes Números

- Exemplo 9 – Simulação de Monte Carlo
- Uma aplicação muito útil da lei fraca dos grandes números está no cálculo de integrais que não podem ser resolvidas numericamente.
- Isso pode ser feito através de simulação de Monte Carlo, como mostrado a seguir.

Lei Fraca dos Grandes Números

- Suponha que queremos calcular:

$$I(f) = \int_0^1 f(x)dx$$

- E esta integral não tem uma solução analítica, ou não pode ser avaliada através de tabelas ou algum método "fácil". O que fazer?

Lei Fraca dos Grandes Números

- Uma solução bastante comum é apelar para métodos de simulação.
- A idéia neste caso é:
  - Gere um conjunto de variáveis iid Unif(0,1) X_1, X_2, ..., X_n
  - Calcule uma aproximação para a integral baseada nos n valores simulados, a saber:

$$\hat{I}(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$$

Lei Fraca dos Grandes Números

- Pela lei fraca dos grandes números, este valor, se n é grande, deve estar próximo de sua média, que é E{f(X)}, que é simplesmente:

$$E\{f(X)\} = \int_0^1 f(x)(1)dx = I(f)$$

- Este esquema pode ser alterado para, por exemplo, modificar o intervalo de integração.

Lei Fraca dos Grandes Números

- Exemplo 10 – aplicação do exemplo 9
- Suponha que desejamos avaliar:

$$I(f) = \int_0^1 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \Phi(1) - 0.5$$

- Neste caso já sabemos a resposta, pela tabela da N(0,1), que é 0.3413.

Lei Fraca dos Grandes Números

- Como resolver isso por Monte Carlo?
- Geramos n observações iid da Unif(0,1) e calculamos:

$$\hat{I}(f) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left\{\frac{-X_i^2}{2}\right\}$$

- Faça algumas simulações no Excel.
- Numa 1a. simulação, gerei n = 1000 variáveis Unif(0,1) e encontrei 0.3427 como o valor aproximado da integral (erro percentual de 0.4%)

Lei Fraca dos Grandes Números

- Numa segunda simulação, com 10 mil números gerados, o valor aproximado da integral foi 0.3407 (erro percentual de 0.2%).

- Numa terceira simulação, gerei 50 mil números e obtive 0.3415 como valor aproximado da integral, um erro percentual de 0.1%.

Exemplo – para casa

- Seja X uma variável aleatória discreta com valores maiores ou iguais a zero. Seja G(s) = E(s^x) a função geradora de probabilidades de X, e suponha que G(s) é finita para todo s. Seja <u>u</u> um número positivo qualquer.

- Usando o mesmo tipo de argumentos que na demonstração da desigualdade de Chebyshev prove que:

$$\Pr(X \le u) \le \frac{G(s)}{s^u}, 0 \le s \le 1$$

Exemplo – para casa

- Seja X ~ Poisson(λ). Use as desigualdades de Chebyshev e Markov para mostrar que:

- 1) $\Pr\left(X \le \frac{\lambda}{2}\right) \le \frac{4}{\lambda}$

- 2) $\Pr(X \ge 2\lambda) \le \frac{1}{\lambda}$

Exemplo – para casa

- Uma variação do método de Monte Carlo
- Suponha que desejamos avaliar:

$$I(f)=\int_a^b f(x)dx$$

- Seja g(x) uma densidade em [a, b].
- Gere X_1, X_2, ..., X_n da densidade g(x) e estime I através de:

$$\hat{I}(f)=\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)}{g(X_i)}$$

Exemplo – para casa

- a) Use este exemplo para definir um procedimento para avaliar uma integral definida no intervalo [-1,+1]
- b) Você consegue usar este esquema para avaliar:

$$I(f)=\int_0^\infty x^2 e^{-x}dx$$

- c) E para avaliar a mesma integral que acima mas no intervalo (0,2)?