ENCE – CÁLCULO DE PROBABILIDADE II

Semestre 2009.01 – Profa. Monica Barros

Lista de exercícios 5

Problema 1

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 um número positivo qualquer.

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

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

Solução

Pela desigualdade de Markov, se u(X) é uma função não negativa, então:

Pr{ u (X) ≥ c } ≤ E [u (X)] /c

Seja X uma variável aleatória discreta com valores maiores ou iguais a zero e G(s) a sua função geradora de probabilidades. Então s^X é não negativa para 0 ≤ s ≤ 1 e podemos usar Markov para escrever:

$$\Pr\{s^X \ge s^u\} \le \frac{E(s^X)}{s^u} \quad \text{onde } c > 0$$

Mas, podemos escrever:

$$\Pr\{s^X \ge s^u\} = \Pr\{\ln(s^X) \ge \ln(s^u)\} = \Pr\{X \ln(s) \ge u.\ln(s)\} \le \frac{E(s^X)}{s^u}$$

Como 0 ≤ s ≤ 1, ln(s) < 0, o que leva à inversão do sinal na desigualdade acima.

$$\Pr\{X \ln(s) \ge u.\ln(s)\} = \Pr\{X \le u\} \le \frac{E(s^X)}{s^u} = \frac{G(s)}{s^u}$$

Problema 2

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

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

$$\Pr(X \geq 2\lambda) \leq \frac{1}{\lambda}$$

Solução (na verdade são tentativas, não consegui chegar às respostas corretas)

Se X é Poisson(λ) então E(X) = VAR(X) = λ e E(X^2) = VAR(X) + λ^2

Por Markov, como X é uma v.a. não negativa:

$$\Pr\{X \geq c\} \leq \frac{E(X)}{c} = \frac{\lambda}{c}$$

Por Chebyshev:

$$\Pr(|X-\mu| \leq k) \geq 1 - \frac{\sigma^2}{k^2} \quad \text{e} \quad \Pr(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Substituindo a média e a variância da Poisson aqui temos:

$$\Pr(|X-\lambda| \leq k) \geq 1 - \frac{\lambda^2}{k^2} \Leftrightarrow \Pr\{-k \leq X-\lambda \leq +k\} \geq 1 - \frac{\lambda^2}{k^2} \Leftrightarrow \Pr\{\lambda - k \leq X \leq \lambda + k\} \geq 1 - \frac{\lambda^2}{k^2}$$

$$\Pr(|X-\lambda| \geq k) \leq \frac{\lambda^2}{k^2} \Leftrightarrow \Pr\{X-\lambda \leq -k\} + \Pr\{X-\lambda \geq +k\} \leq \frac{\lambda^2}{k^2}$$
$$\Leftrightarrow \Pr\{X \leq \lambda - k\} + \Pr\{X \geq \lambda + k\} \leq \frac{\lambda^2}{k^2}$$

No item a) queremos provar que: $\Pr\left(X \leq \frac{\lambda}{2}\right) \leq \frac{4}{\lambda}$

Use Chebyshev com $\lambda^2/k^2 = 4/\lambda$ e então $k^2 = \lambda^3/4$

Substituindo acima:

$$\Pr\left\{X \leq \lambda - \frac{\lambda\sqrt{\lambda}}{2}\right\} + \Pr\left\{X \geq \lambda + \frac{\lambda\sqrt{\lambda}}{2}\right\} \leq \frac{\lambda^2}{k^2} = \frac{4}{\lambda}$$

Para $\lambda > 4$, que é o que fornece resultados interessantes acima, a primeira probabilidade refere-se a X ser menor que um número negativo, e é zero. Logo:

$$\Pr\left\{X \geq \lambda + \frac{\lambda\sqrt{\lambda}}{2}\right\} \leq \frac{4}{\lambda}$$

$$\Pr\left\{X \geq \frac{2\lambda + \lambda\sqrt{\lambda}}{2}\right\} \leq \frac{4}{\lambda}$$

Usando Markov com $\lambda/c = 4/\lambda$ leva a: $c = \lambda^2/4$ .Então:

$$\Pr\left\{X \geq \frac{\lambda^2}{4}\right\} \leq \frac{4}{\lambda} \Rightarrow \Pr\left\{X < \frac{\lambda^2}{4}\right\} \geq 1 - \frac{4}{\lambda}$$

No item b) faça $\lambda^2/k^2 = 1/\lambda$

Substituindo acima leva a:

$$\Pr\left(\left|X - \lambda\right| \geq \lambda^{3/2}\right) \leq \frac{\lambda^2}{k^2} = \frac{1}{\lambda} \Rightarrow \Pr\left\{X - \lambda \geq \lambda^{3/2}\right\} + \Pr\left\{X - \lambda \leq -\lambda^{3/2}\right\} \leq \frac{1}{\lambda}$$

Se $\lambda \geq 1$, a segunda probabilidade do lado direito é a de X ser menor que a de um número negativo, e então:

$$\Pr\left\{X - \lambda \geq \lambda^{3/2}\right\} \leq \frac{1}{\lambda} \Rightarrow \Pr\left\{X \geq \lambda + \lambda\sqrt{\lambda}\right\} \leq \frac{1}{\lambda} \Rightarrow \Pr\left\{X \geq \lambda\left(1 + \sqrt{\lambda}\right)\right\} \leq \frac{1}{\lambda}$$

Por Markov, fazendo $\lambda/c = 1/\lambda$ leva a: $c = \lambda^2$ .Então:

$$\Pr\left\{X \geq \lambda^2\right\} \leq \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

Problema 3

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 $X_1, X_2, \ldots, X_n$ da densidade g(x) e estime I através de:

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

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)?

Solução

a) A forma mais simples de resolver o problema talvez seja definindo g(x) como a densidade Uniforme no intervalo [-1,+1].

$$g(x) = \frac{1}{2} \quad \text{para -1} \leq \text{x} \leq +1$$

$$\hat{I}(f) = \frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i)}{1/2} = \frac{2}{n}\sum_{i=1}^{n} f(X_i)$$

b) Neste caso sabemos que o resultado da integral é a função Gama com argumento 3, ou seja, 2! = 2.

Precisamos gerar uma amostra de uma densidade definida em (0, + ∞). Por exemplo, poderíamos gerar n v.a. da densidade Expo(1), que são geradas facilmente a partir de variáveis Unif(0,1). Se X_1, X_2, ..., X_n são iid Expo(1) então podem ser construídas a partir de U_1, U_2, ..., U_n iid Unif(0,1) apenas fazendo a transformação: $X_i = -\ln(U_i)$.

A integral aproximada por Monte Carlo torna-se então:

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

Por exemplo, execute o programa Matlab monte_carlo_lista_5.m que está em anexo ao material do curso.

Numa simulação com 10 mil observações, encontrei como resultado da integral por Monte Carlo o valor 1.9822 (o valor real é 2), o que significa um erro percentual de 0.89%.

Numa outra simulação para a função Gama(5) (quais as modificações necessárias acima?) e usando 20 mil observações, o valor real é 4! = 24 e a integral simulada por Monte Carlo foi 25.8727, o que corresponde a um erro percentual de 7.80%.

Para esta mesma integral, fiz uma outra simulação, agora gerando 50 mil observações e encontrei 24.0077 como o resultado da simulação de Monte Carlo, um erro percentual de apenas 0.03%.

c) Agora a simulação é bastante fácil, pois o intervalo da integral é finito, e a amostra gerada será da densidade g(x), uma Uniforme(0,2).

Podemos generalizar esta idéia: se o interesse é avaliar a integral da função Gama no intervalo (0,t) basta empregar v.a. Uniformes neste mesmo intervalo.

No nosso problema específico, a integral aproximada por Monte Carlo será:

$$\hat{I}(f) = \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)}{1/2} = \frac{2}{n}\sum_{i=1}^{n} f(X_i) = \frac{2}{n}\sum_{i=1}^{n} X_i^2.\exp(-X_i)$$

Onde os X_i são iid Unif(0,2). Como criá-los a partir de v.a. Unif(0,1)? É fácil. Basta fazer $X_i = 2.U_i$ onde U_i é Unif(0,1).

Abaixo estão alguns resultados de simulações feitas para resolver a integral em c). O programa Matlab para fazer estas simulações está em anexo, é o Matlab monte_carlo_lista_5b.m

Número de v.a. simuladas	Valor da integral
10,000	0.6441
50,000	0.6465
100,000	0.6473
1,000,000	0.6467
10,000,000	0.6467

Problema 4

Numa certa empresa de computação, o salário anual médio dos funcionários com menos de 5 anos de experiência é R$ 36000, com desvio padrão de R$ 1800. Toma-se uma amostra de 36 funcionários com menos de 5 anos de experiência. Qual a probabilidade do salário médio na amostra exceder R$ 36500?

Solução

É uma aplicação típica to teorema central do limite. A distribuição dos salários individuais é desconhecida, mas desejamos saber algo sobre a distribuição do salário médio numa amostra de tamanho razoavelmente grande, e assim podemos invocar o TCL.

$E(X_i) = 36000$, $VAR(X_i) = (1800)^2$

$$E(\overline{X}) = 36000$$

$$VAR(\overline{X}) = \frac{(1800)^2}{36} \Rightarrow dp(\overline{X}) = \frac{(1800)}{6} = 300$$

$$\Pr\{\overline{X} > 36500\} = \Pr\left\{\frac{\overline{X} - 36000}{300} > \frac{36500 - 36000}{300}\right\} \approx 1 - \Phi\left(\frac{36500 - 36000}{300}\right) = 1 - \Phi\left(\frac{5}{3}\right) =$$
$$= 1 - \Phi(1.6667) = 1 - 0.9522 = 0.0478$$

Problema 5

Um elevador pode transportar até 10 passageiros. O peso de cada passageiro é uma variável aleatória com média 70 kg e desvio padrão 10 kg. Normas de segurança estabelecem que o peso máximo transportado no elevador não pode exceder 715 kg mais de 5% do tempo.

a) Qual a probabilidade de que o peso total dos passageiros exceda 715 kg?

b) Encontre o percentil 95% da distribuição do peso total dos passageiros do elevador. As normas de segurança estão sendo obedecidas?

Solução

A solução é, novamente, pelo TCL.

Seja T o peso total. Então T é uma v.a. com média 700 kg e variância $10(10)^2$.

$$\Pr\{T > 715\} = \Pr\left\{\frac{T - 700}{10\sqrt{10}} > \frac{715 - 700}{10\sqrt{10}}\right\} \approx 1 - \Phi\left(\frac{15}{10\sqrt{10}}\right) = 1 - \Phi(0.4743) = 1 - 0.6824 = 0.3176$$

A distribuição do peso total é aproximadamente N(700, 1000) e seu percentil 95% é 752.01 kg. Isso significa que 5% "do tempo" o peso total ultrapassa 752 kg, o que é claramente fora das normas de segurança do elevador. No item anterior nós já tínhamos uma indicação disso, pois a probabilidade de ultrapassar 715 kg (iesto é, estar fora das normas de segurança) era cerca de 32%, bem acima do valor desejado de 5%.

Problema 6

Sejam X_1 , X_2 , ... , X_n iid $N(0, \sigma^2)$.

a) Qual a média e variância de X_i^2 ?

b) Como aproximar Pr{ $X_1^2 + X_2^2 + ...+ X_n^2 \leq x$) em termos de $\Phi(.)$?

Solução

$\frac{X_i^2}{\sigma^2}$ tem densidade qui-quadrado com 1 grau de liberdade para i =1,2,...n

Logo, sua média e variância são:

$$E\left(\frac{X_i^2}{\sigma^2}\right)=1 \Rightarrow \mathrm{E}\left(X_i^2\right)=\sigma^2 \text{ e } \mathrm{VAR}\left(\frac{X_i^2}{\sigma^2}\right)=2 \Rightarrow \quad \mathrm{VAR}\left(X_i^2\right)=2\sigma^4$$

Então, a soma dos X_i^2 ´s tem média e variância:

$$E\left(\sum_{i=1}^{n} X_i^2\right)=n\sigma^2$$

$$VAR\left(\sum_{i=1}^{n} X_i^2\right)=2n\sigma^4$$

Pelo Teorema Central do Limite, se n for grande podemos normalizar esta soma de tal forma que ela seja aproximadamente N(0,1).

Seja T = $X_1^2 + X_2^2 + ...+ X_n^2$ então:

$$Z=\frac{T-n\sigma^2}{\sqrt{2n\sigma^4}}=\frac{\sum_{i=1}^{n} X_i^2-n\sigma^2}{\sqrt{2n\sigma^4}} \approx N(0,1)$$

$\Pr(T \leq x)=\Pr\left(\frac{T-n\sigma^2}{\sqrt{2n\sigma^4}} \leq \frac{x-n\sigma^2}{\sqrt{2n\sigma^4}}\right) \approx \Phi\left(\frac{x-n\sigma^2}{\sqrt{2n\sigma^4}}\right)$ é a aproximação desejada.

Problema 7

Sejam X_1 , X_2 , ... , X_n iid N(0, 1).

a) Calcule Pr{ $X_1^2 + X_2^2 + ...+ X_{100}^2 \leq$ 120)

b) Calcule Pr{ 80 $\leq X_1^2 + X_2^2 + ...+ X_{100}^2 \leq$ 120)

c) Ache c tal que Pr{ $X_1^2 + X_2^2 + ...+ X_{100}^2 \leq$ 100 + c) = 0.95

Solução

Note que os X_i^2 são iid Qui-Quadrado com 1 grau de liberdade e portanto sua soma é também Qui-Quadrado.

a) $\Pr\left\{X_1^2 + X_2^2 + \ldots X_{100}^2 \le 120\right\} = \Pr\left\{\chi_{100}^2 \le 120\right\} = \Pr\left\{\frac{\chi_{100}^2 - 100}{\sqrt{100}} \le \frac{120-100}{\sqrt{100}}\right\} \approx \Phi\left(\frac{20}{10}\right) = \Phi(2) =$

$= 0.9773$

b) $\Pr\left\{80 \le X_1^2 + X_2^2 + \ldots X_{100}^2 \le 120\right\} = \Pr\left\{80 \le \chi_{100}^2 \le 120\right\} = \Pr\left\{\frac{80-100}{\sqrt{100}} \le \frac{\chi_{100}^2 - 100}{\sqrt{100}} \le \frac{120-100}{\sqrt{100}}\right\}$

$\approx \Phi\left(\frac{20}{10}\right) - \Phi\left(\frac{-20}{10}\right) = \Phi(2) - \Phi(-2) = 0.9773 - 0.0228 = 0.9545$

c) $\Pr\left\{X_1^2 + X_2^2 + \ldots X_{100}^2 \le 100 + c\right\} = \Pr\left\{\chi_{100}^2 \le 100 + c\right\} = \Pr\left\{\frac{\chi_{100}^2 - 100}{\sqrt{100}} \le \frac{100 + c - 100}{\sqrt{100}}\right\}$

$\approx \Phi\left(\frac{c}{10}\right) = 0.95 \Rightarrow \frac{c}{10} = 1.645 \Rightarrow c = 16.45$

Problema 8

Seja f(x) = $1/x^2$ onde x > 1. Considere uma amostra aleatória de tamanho 72 desta densidade e calcule aproximadamente a probabilidade de que mais do que 50 ítens na amostra sejam menores que 3.

Solução

O problema não está corretamente especificado. A média e variância dos X's não são finitas, não podemos aplicar o TCL.

Problema 9

O número de acidentes de trânsito numa estrada é uma variável aleatória N com distribuição Poisson com média 100. Use a aproximação Normal para encontrar k tal que:

Pr(100 - k < N < 100 + k) ≈ 0.90

amostra sejam menores que 3.

Solução

E(N) = VAR(N) = 100 e N pode ser encarado como a soma de 100 v.a. Poisson com parâmetro 1, e assim são válidas as condições do TCL.

Pelo TCL:

$$\frac{N-100}{\sqrt{100}}=\frac{N-100}{10} \quad \text{é aprox. N(0,1)}$$

$$\Pr\{100-k\le N\le 100+k\}=\Pr\left\{\frac{100-k-100}{10}\le\frac{N-100}{10}\le\frac{100+k-100}{10}\right\}=$$

$$=\Pr\left\{\frac{-k}{10}\le\frac{N-100}{10}\le\frac{+k}{10}\right\}\approx\Phi\left(\frac{+k}{10}\right)-\Phi\left(\frac{-k}{10}\right)=\Phi\left(\frac{+k}{10}\right)-\left\{1-\Phi\left(\frac{+k}{10}\right)\right\}=2\Phi\left(\frac{+k}{10}\right)-1$$

Logo:

$$2\Phi\left(\frac{k}{10}\right)-1=0.90\Rightarrow\Phi\left(\frac{k}{10}\right)=0.95\Rightarrow\frac{k}{10}=1.645\Rightarrow k=16.45$$

Problema 10

Você joga R$ 5 em cada repetição independente de um jogo no qual a probabilidade de ganhar é 50%. Use o Teorema Central do Limite para aproximar a probabilidade de que, após 50 jogadas, você terá perdido mais de R$75.

amostra sejam menores que 3.

Solução

Seja X_i = +5 se você ganha a i-ésima jogada e X_i = -5 se você perde a i-ésima jogada. As probabilidades de ganhar e perder são iguais, p = q = ½. Então $E(X_i) = 0$ e $E(X_i^2) = 25(1/2) + (25)(1/2) = 25 = VAR(X_i)$.

Seja $Y = X_1 + X_2 + ... + X_{50}$ o total de dinheiro ganho após as 50 jogadas

Então E(Y) = 0 e VAR(Y) = 50(25)

Pelo TCL:

$$\Pr\{Y<-75\}=\Pr\left\{\frac{Y-0}{\sqrt{50(25)}}<\frac{-75-0}{\sqrt{50(25)}}\right\}\approx\Phi\left(\frac{-75}{25\sqrt{2}}\right)=\Phi\left(\frac{-3}{\sqrt{2}}\right)=\Phi(-2.1213)=0.0169$$

Problema 11

Seja Z_n uma variável aleatória Poisson(n). Mostre que a distribuição limite de $Y_n=\dfrac{Z_n-n}{\sqrt{n}}$ é N(0,1).

Solução

Este problema é apenas uma aplicação direta do TCL, como o problema 9. Se Z_n é uma variável aleatória Poisson(n), ela pode ser pensada como uma soma de n variáveis Poisson(1) e assim o TCL se aplica à variável padronizada Y_n definida acima.