Lista 9 - Introdução à Probabilidade e Estatística

Desigualdades e Teoremas Limites

1 — Um arqueiro aponta a um alvo de 20 cm de raio. Seus disparos atingem o alvo, em média, a 5 cm do centro deste. Assuma que cada disparo é independente de qualquer outro disparo. Limite superiormente a probabilidade do atirador errar o alvo no próximo disparo.

2 — Suponha que X seja uma variável aleatória com média e variância iguais a 20. O que é possível dizer sobre $\mathbb{P}(0 < X < 40)$?

3 — Com sua experiência, um professor sabe que a nota de um estudante na prova final é uma variável aleatória com média 75.

a) Forneça um limite superior para a probabilidade de que a nota de um estudante exceda 85. Suponha, além disso, que o professor saiba que a variância da nota de m estudante é igual a 25.

b) O que se pode dizer sobre a probabilidade de que a nota de um estudante esteja entre 65 e 85?

c) Quantos estudantes teriam que fazer a prova para assegurar, com probabilidade mínima de $0,9$, que a média da turma esteja entre 75 ± 5? Não use o Teorema Central do Limite.

4 — Uma moeda honesta é lançada de forma independente n vezes. Seja S_n o número de caras obtidas nesses n lançamentos. Use a desigualdade de Chebyshev para obter um limitante inferior para a probabilidade de que $\frac{S_n}{n}$ diste de $\frac{1}{2}$ menos do que $0,1$ quando

1. $n = 100$.
2. $n = 10000$.
3. $n = 100000$.

5 — No contexto do problema anterior verifique que

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| < \epsilon\right) = 1$$

para todo $\epsilon > 0$.

6 — Considere uma moeda desonesta com probabilidade p de sair cara. Seja S_n o número de caras obtidas em n lançamentos independentes desta moeda. Escreva um limite semelhante ao problema anterior. Calcule o valor deste limite.

7 — Dez dados honestos são lançados. Encontre, aproximadamente, a probabilidade de que a soma dos números assim obtidos esteja entre 30 e 40.

8 — Suponha que um programa de computador tem $n = 100$ páginas de códigos. Seja X_i o número de erros na $i-$ésima página. Suponha que as variáveis aleatórias X_i tem distribuição Poisson de parâmetro 1 e que são independentes. Seja $Y = \sum_{j=1}^{100} X_j$ o número total de erros. Utilize o Teorema Central do Limite para aproximar $\mathbb{P}[Y < 90]$.

9 — Uma amostra aleatória de n itens é tomada de uma distribuição com media μ e variância σ^2, $0 < \sigma^2$. Utilizando o Teorema Central do Limite, determine o menor número de itens a serem considerados para que a seguinte relação seja satisfeita:

$$\mathbb{P}[|\frac{S_n}{n} - \mu| \leq \frac{\sigma}{4}] \geq 0,99 \ .$$

10 — Uma pessoa possui 100 lâmpadas cujos tempos de vida são exponenciais independentes com média de 5 horas. Se as lâmpadas são usadas uma de cada vez, sendo a lâmpada queimada imediatamente substituída por uma nova, obtenha uma aproximação para a probabilidade de que ainda exista uma lâmpada funcionando após 525 horas.

11 — Mostre que $\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{2}$. (Sugestão: Considere uma sequência de variáveis aleatórias i.i.d com distribuição Poisson e utilize o Teorema Central do Limite.)

12 — Seja $(X_n)_{n\geq 1}$ uma sequência de variáveis aleatórias tal que

1. $\lim_{n\to\infty} \mathbb{E}[X_n] = \alpha$.
2. $\lim_{n\to\infty} \mathbb{V}\mathrm{ar}[X_n] = 0$.

Mostre que, $\forall \epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}[|X_n - \alpha| \geq \epsilon] = 0$.

13 — Um provedor de acesso à Internet está monitorando a duração do tempo de conexões de seus clientes, com o objetivo de dimensionar seus equipamentos. A média é desconhecida, mas o desvio padrão é considerado igual a $\sqrt{50}$ minutos. Uma amostra de 500 conexões resultou em um valor médio observado de 25 minutos. O que dizer da verdadeira média, com confiança $\gamma = 0,92$?

14 — Suponha que X represente a duração da vida de uma peça de equipamento. Admita-se que 100 peças sejam ensaiadas, fornecendo uma duração de vida média de 501, 2 horas. Suponha-se que o desvio padrão seja conhecido e igual a 4 horas. Construa um intervalo de confiança de 95% para a média.

15 — A diretoria de uma escola de segundo grau quer estimar a proporção p de estudantes que conseguiram entender de forma satisfatória as mensagens transmitidas numa exposição de arte. Essa proporção deverá ser estimada com um erro de 5% para um coeficiente de confiança de 90%.

a) Qual é o tamanho de amostra necessário para atender às exigências da diretoria?

b) Que tamanho deverá ter a amostra sabendo que p está entre $0,20$ e $0,60$? E sabendo que $p < 0,20$?

c) Numa amostra de 150 estudantes, 60 apresentaram desempenho satisfatório num teste aplicado na saída da exposição. Qual seria a estimativa intervalar de p nesse caso, para $\gamma = 0,95$?

16 — Uma revista semanal, em artigo sobre a participação das mulheres em curso superior de administração, pretende estimar a proporção p de mulheres neste curso.

a) Quantos estudantes de administração devem ser entrevistados de modo que a proporção p seja estimada com um erro de $0,04$ e uma probabilidade de $0,98$?

b) Se tivermos a informação adicional de que a proporção p é pelo menos 35%, você conseguiria diminuir o tamanho amostral calculado no item anterior? Justifique.

17 — Um estudo da prefeitura indica que 30 das crianças da cidade têm deficit de atenção na escola. Numa amostra de 200 crianças, qual a probabilidade de pelo menos 50 crianças tenham esse problema?

18 — $\star$ Utilize a desigualdade de Chebyshev para mostrar que para toda função continua $f : [0,1] \to \mathbb{R}$,

$$\sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} x^k (1-x)^k \to f(x)$$

uniformemente em $x \in [0,1]$ quando $n \to \infty$.

Respostas dos Exercícios

1 Seja X a distância do ponto atingido ao centro do alvo. Note que $X \geq 0$. Seja Y uma variável aleatória definida como sendo igual a 20 caso $X \geq 20$ e 0 em outro caso. Logo, $X \geq Y$. Tomando esperança temos que

$$\mathbb{E}[X] \geq \mathbb{E}[Y] = 20\mathbb{P}[X \geq 20].$$

Como $\mathbb{E}[X] = 5$ temos que $\mathbb{P}[X \geq 20] \leq \frac{1}{4}$.

Observação: Poderíamos simplesmente ter aplicado a desigualdade de Chebyshev.

2 $\mathbb{P}[0 < X < 40] = \mathbb{P}[-20 < X - 20 < 20] = 1 - \mathbb{P}[|X - 20| \geq 20] \geq 1 - \frac{20}{400} = \frac{19}{20}$.

3 $(a)\mathbb{P}[X \geq 85] \leq \mathbb{E}[X]/85 = 15/17$.
$(b)\mathbb{P}[65 \geq X \leq 85] = 1 - \mathbb{P}[|X - 75| > 10] \geq 1 - \frac{25}{100}$.
$(c)\mathbb{P}[|\sum_{i=1}^{n} X_i/n - 75| > 5] \leq \dfrac{25}{25n}$. Logo $n = 10$.

4 Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes cada uma assumindo os valores 0 e 1 com probabilidade $\frac{1}{2}$. Definiremos que 1 representa cara e 0 representa coroa. Logo $S_n = X_1 + X_2 + \ldots + X_n$ representa o número de caras em n lançamentos. Temos que

$$\mathbb{E}[\frac{S_n}{n}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{2}$$

e, devido a independência, temos que

$$\mathbb{V}\operatorname{ar}(\frac{S_n}{n}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\operatorname{ar}(X_i) = \frac{1}{4n},$$

já que $\mathbb{E}[X_i] = \frac{1}{2}$ e $\mathbb{V}\operatorname{ar}(X_i) = \frac{1}{4}$ para todo $i, i = 1, 2, \ldots, n$. Aplicando a desigualdade de Chebyschev temos que

$$\mathbb{P}(|\frac{S_n}{n} - \frac{1}{2}| \geq 0,1) \leq \frac{1}{4(0,1)^2 n}.$$

Substituindo temos que os limites inferiores fornecidos pela desigualdade de Chebyschev são: $(a)1 - \frac{1}{4}$, $(b)1 - \frac{1}{400}$ e $(c)1 - \frac{1}{4000}$ respectivamente.

5 Analogamente ao problema anterior temos que

$$\mathbb{P}(|\frac{S_n}{n} - \frac{1}{2}| < \epsilon) \geq 1 - \frac{1}{4\epsilon^2 n}.$$

O resultado segue desta desigualdade.

6 Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes cada uma assumindo os valores 0 e 1 com probabilidade p e $1-p$ respectivamente. Por convenção assumiremos que 1 representa cara e 0 representa coroa. Logo $S_n = X_1 + X_2 + \ldots + X_n$ representa o número de caras em n lançamentos. Temos que

$$\mathbb{E}[\frac{S_n}{n}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = p$$

e, devido a independência, temos que

$$\mathbb{V}\operatorname{ar}(\frac{S_n}{n}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\operatorname{ar}(X_i) = \frac{p(1-p)}{n},$$

já que $\mathbb{E}[X_i] = p$ e $\mathbb{V}\operatorname{ar}(X_i) = p(1-p)$ para todo $i, i = 1, 2, \ldots, n$. Aplicando a desigualdade de Chebyschev temos que

$$\mathbb{P}(|\frac{S_n}{n} - p| \geq \epsilon) \leq \frac{p(1-p)}{\epsilon^2 n}$$

para todo $\epsilon > 0$. Isto implica que

$$\lim_{n\to\infty}\mathbb{P}(|\frac{S_n}{n} - p| < \epsilon) = 1.$$

7 Seja X_i o número sorteado pelo i-ésimo dado e seja $S_{10} = \sum_{i=1}^{10} X_i$ a soma dos números sorteados nos lançamentos dos 10 dados. Logo, $\mathbb{P}[30 < S_{10} < 40] = \mathbb{P}[\frac{30-\frac{7}{2}}{\sqrt{10\frac{91}{6}}} < \frac{S_{10}-\frac{7}{2}}{\sqrt{10\frac{91}{6}}} < \frac{40-\frac{7}{2}}{\sqrt{10\frac{91}{6}}}]$. Utilizando a aproximação dada pelo Teorema Central do Limite temos que $\mathbb{P}[30 < S_{10} < 40] \approx \Phi(\frac{40-\frac{7}{2}}{\sqrt{10\frac{91}{6}}}) - \Phi(\frac{30-\frac{7}{2}}{\sqrt{10\frac{91}{6}}})$, onde $\Phi(t)$ é a função de distribuição acumulada de uma variável aleatória com distribuição normal padrão.

8 Note que $\mathbb{E}[Y] = 100$ e que $\mathbb{V}\operatorname{ar}(Y) = 100$. logo, $\mathbb{P}[Y < 90] = \mathbb{P}[\frac{Y-100}{10} < \frac{90-100}{10}]$. Pelo Teorema Central do Limite, $\mathbb{P}[\frac{Y-100}{10} < \frac{90-100}{10}] \approx \Phi(\frac{90-100}{10}) = \Phi(-1) = 0,1587$.

9 Note que $\mathbb{P}[|\frac{S_n}{n} - \mu| \leq \frac{\sigma}{4}] = \mathbb{P}[-\frac{\sigma}{4} \leq \frac{S_n}{n} - \mu \leq \frac{\sigma}{4}] = \mathbb{P}[-\frac{\sqrt{n}}{4} \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq \frac{\sqrt{n}}{4}]$. Pelo Teorema Central do Limite, $\mathbb{P}[-\frac{\sqrt{n}}{4} \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq \frac{\sqrt{n}}{4}] \approx \Phi(\frac{\sqrt{n}}{4}) - \Phi(-\frac{\sqrt{n}}{4}) = 2\Phi(\frac{\sqrt{n}}{4}) - 1$, onde $\Phi(t)$ é a função de distribuição acumulada de uma variável aleatória com distribuição normal padrão. Logo, encontre n tal que $\Phi(\frac{\sqrt{n}}{4}) \geq 0,995$.

10 Análogo ao exercício **9**.

11 Considere uma sequência $(X_n)_{n\geq 1}$ de variáveis aleatórias independentes, identicamente distribuídas, com distribuição Poisson de parâmetro 1. $S_n = \sum_{k=1}^{n} X_k$ tem distribuição Poisson de parâmetro n. Logo temos que

$$\mathbb{P}[S_n \leq n] = \sum_{k=0}^{n} \mathbb{P}[S_n = k] = e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}$$

Como $\mathbb{E}[X_i] = 1$ e $\mathbb{V}\mathrm{ar}(X_i) = 1$ e $\mathbb{P}[S_n \leq n] = \mathbb{P}[\frac{S_n - n}{\sqrt{n}} \leq \frac{n-n}{\sqrt{n}}] = \mathbb{P}[\frac{S_n-n}{\sqrt{n}} \leq 0]$ o Teorema Central do Limite implica que

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \Phi(0) = \frac{1}{2},$$

onde $\Phi(t)$ é a função de distribuição acumulada de uma variável aleatória com distribuição normal padrão.

12 Chebyschev.

13 O intervalo de confiança para a média com variância σ^2 conhecida e coeficiente de confiança γ ou $\gamma 100\%$ é dado por $[\overline{X}_n - a\frac{\sigma}{\sqrt{n}}, \overline{X}_n + a\frac{\sigma}{\sqrt{n}}]$, onde $\overline{X}_n$ é a média amostral e a é tal que $(1-\Phi(a)) = \frac{\alpha}{2}$ com $\gamma = 1-\alpha$. Logo, a é tal que $\Phi(a) = 0,96$. Portanto $a = 1,755, \overline{X}_n = 25, n = 500$ e $\sigma = \sqrt{50}$.

14 Análogo ao exercício **14**.

15 Seja $X_i \sim$ Bernoulli(p) assumindo os valores 0 e 1, onde $X_i = 1$ se o i-ésimo estudante entendeu a mensagem de forma satisfatória e 0 em outro caso. Logo, $\overline{X}_n = \frac{S_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$ representa a proporção dos estudantes que entenderam a mensagem de forma satisfatória e é uma estimativa do valor desconhecido p. A estimativa intervalar para a proporção desconhecida é dada por um intervalo da forma $[\overline{X}_n - \epsilon, \overline{X}_n + \epsilon]$, onde ϵ é a margem de erro. A estimativa intervalar com margem de erro ϵ tem coeficiente de confiança γ se $\gamma = \mathbb{P}[|\overline{X}_n - p| \leq \epsilon]$. Note que

$$\begin{aligned}\gamma &= \mathbb{P}[-\frac{n\epsilon}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{n\epsilon}{\sqrt{np(1-p)}}] \\ &= \mathbb{P}[-\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}] \\ &\approx \Phi(t_{n,\epsilon}) - \Phi(-t_{n,\epsilon}),\end{aligned}$$

onde $\Phi(t)$ é a função de distribuição acumulada de uma variável aleatória com distribuição normal padrão e $t_{n,\epsilon} = \frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}$. A aproximação deve-se ao Teorema Central do Limite. Na prática, toma-se $\gamma = \Phi(t_{n,\epsilon}) - \Phi(-t_{n,\epsilon})$. Pelas propriedades da função Φ segue que $\gamma = 2\Phi(t_{n,\epsilon}) - 1$.

(a) $t_{n,\epsilon} = \frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}$. Logo, $n \leq \frac{t_{n,\epsilon}^2}{4\epsilon^2}$ já que p é desconhecido e portanto limitamos a expressão $p(1-p)$ por $1/4$ que é seu valor máximo no intervalo $[0,1]$. Como $\epsilon = 0,05$ e $\gamma = 0,90$ então $t_{n,\epsilon} = 1,65$. Logo, $n \leq 272,5$. Toma-se $n = 272$.

(b) A função $f(p) = p(1-p)$ tem um máximo absoluto em $I = [0,1]$ em $p = \frac{1}{2}$. Desta forma, se o valor desconhecido de p pertence ao intervalo $(0.2, 0.6)$ limitamos o valor de $f(p)$ por $1/4$ e n deve ser como no problema anterior, $n = 272$.

Se $0 < p < 0,2$ então $f(p) \leq 0.16$. Logo, $n \leq 174,24$. Toma-se $n = 174$.

(c) Na prática substitui-se a proporção desconhecida $\overline{X}_n$ pela proporção amostral $\hat{p}$. Da expressão $\gamma = \mathbb{P}[-\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}]$ temos que $I = [\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$ é um intervalo de confiança para a proporção desconhecida com coeficiente de confiança γ onde z e γ estão relacionados através da equação $\gamma = \Phi(z) - \Phi(-z) = 2\Phi(z) - 1$.

No problema $n = 150, \hat{p} = \frac{60}{150} = 0,40, \gamma = 0,95$ e portanto $z = 1,96$. Logo, $I = [0.3216, 0.4784]$.

16 Idem Exercício **16**.

17 Vamos considerar o caso em que cada criança tem a mesma probabilidade de ter este problema. Definindo

$$X_j = \begin{cases} 1 & \text{se a } j\text{−ésima criança tem esse problema.} \\ 0 & \text{c.c.} \end{cases}$$

temos que $X = X_1 + \cdots + X_{200} \sim$Binomial$(200,\ 0,30)$. A probabilidade a ser calculada é

$$\mathbb{P}[X \geq 50] = \sum_{k=50}^{200} \binom{200}{k} 0,3^k \quad 0,7^{200-k}$$

Vamos aproximar esse valor. Sabemos que $X \sim$Binomial$(200,\ 0,30)$. Logo, $\mathbb{E}[X] = 200\ (0,3) = 60$ e $\mathbb{V}\mathrm{ar}(X) = 200\ (0,3)(0,7) = 42$. Assim sendo, aproximamos a distribuição de X pela distribuição de uma variável aleatória Y com $Y \sim \mathcal{N}(60, 42)$. Logo,

$$\begin{aligned} \mathbb{P}[X \geq 50] &\approx \mathbb{P}[Y \geq 50] \\ &= \mathbb{P}[\frac{Y-60}{\sqrt{42}} \geq \frac{50-60}{\sqrt{42}}] \\ &= 1 - \Phi(-1,42) \\ &= 0,940, \end{aligned}$$

onde $\Phi(t)$ é a função de distribuição acumulada de uma variável aleatória com distribuição normal padrão.

18 Este exercício é opcional. Dica: Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes cada uma assumindo os valores 0 e 1 com probabilidade p e $1-p$ respectivamente. Seja $S_n = X_1 + X_2 + \ldots + X_n$ o número de caras em n lançamentos. Defina o polinômio $r_n(p) = \mathbb{E}[f(\frac{S_n}{n})]$ e estude a expressão $|r_n(p) - f(p)|$.