Estatística - Listas de Exercícios

Jorge Augusto Salgado Salhani

Novembro, 2022

1 Lista 5 - Estimação Pontual e Intervalo de Confiança

1. Seja X uma variável aleatória com distribuição uniforme *Unif(0, a)*. Note que $E[X] = a/2$, mas a é, em princípio, desconhecido. É planejado o seguinte experimento: são extraídas duas instâncias X_1 e X_2 de X, independentes. Qual a probabilidade de $a > X_1 + X_2$?

Dado $X_1 < a$, para satisfazer a desigualdade, X_2 é restrito a $X_2 < a - X_1$. Sendo $f =$*Unif(0, a)*, temos que

$$P(a > X_1 + X_2) = P(X_1 < a)P(X_2 < a - X_1) = \int_0^a f dx_1 \int_0^{a-X_1} f dx_2$$

Como $f = 1/a$ para $0 \leq x \leq a$ (para sua integral ser 1 em todo o espaço e ser definida como distribuição de probabilidade), temos que

$$\begin{aligned} P(a > X_1 + X_2) &= (1/a)^2 \int_0^a dx_1 \int_0^{a-X_1} dx_2 = (1/a)^2 \int_0^a (a - X_1) dx_1 \\ &= (1/a)^2[(a)^2 - (1/2)(a)^2] = 1/2 \end{aligned}$$

Logo $P(a < X_1 + X_2) = 1/2$.

2. Seja X uma variável aleatória com distribuição de Bernoulli, com probabilidade p de tomar o valor 1 e $1 - p$ de tomar o valor 0. Note que $E[X] = p$, mas p é desconhecido, em princípio. São extraídas n instâncias $X_1, X_2, ..., X_n$, independentes. Sendo $\overline{X} = \sum_k X_k/n$, qual a probabilidade de $\overline{X} \leq 0.9p$?

Sendo os eventos variáveis de Bernoulli com valor 0 e 1, a distribuição de $X_1 + X_2 + ... + X_n = \sum_k X_k$ segue distribuição binomial. Assim, sendo $\overline{X} = \sum_k X_k/n$, temos que $n\overline{X} \sim$ *Binom(n, np)*. Logo

$$P(\overline{X} \leq 0.9p) = P(n\overline{X} \leq np0.9) = \sum_{i=0}^{\lfloor np0.9 \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$$

3. Seja X uma variável aleatória com distribuição normal *N(a,1)*. Note que $E[X] = a$, mas a é desconhecido, em princípio. São extraídas duas observações X_1, X_2 de X, independentes. Qual a probabilidade de $a < min(X_1, X_2)$?

Como a função *min*(X_1, X_2) não cria vínculo entre as variáveis, basta que $X_1 < a$ e $X_2 < a$. Assim

$$P(a < \min(X_1, X_2)) = P(a < X_1)P(a < X_2).$$

Como *N(a,1)* é simétrico em relação a a, então $P(a < X_1) = 1/2 = P(a < X_2)$. Logo $P(a < min(X_1, X_2)) = 1/4$.

3. Seja X uma variável aleatória com distribuição normal $N(\mu, 1)$, cuja média μ é desconhecida. Será extraída uma amostra de 4 elementos e com elas serão calculadas $A = \left[\sum_i^4 X_i/4\right] - 1$ e $B = \left[\sum_i^4 X_i/4\right] + 1$.

A e B são variáveis aleatórias? A e B têm distribuição normal? Qual a média e desvio padrão de A e B? A e B são independentes?

Como A e B são definidas por variáveis aleatórias, A e B são também aleatórias.

Sabendo que se

$$X \sim N(\mu, \sigma^2) \implies Y = X + c \sim N(\mu + c, \sigma^2)$$

ou seja, soma de valores a uma variável aleatória de distribuição normal mantém a distribuição normal, mas com média deslocada, podemos usar esse fato para afirmar que A e B possuem também distribuição normal.

Além disso, sabendo que

$$X \sim N(\mu, \sigma^2) \implies Y = \sum_i^n X_i \sim N(n\mu, n\sigma^2) \tag{1}$$

Como $\sum_i^4 X_i = 4(A + 1)$ e $\sum_i^4 X_i = 4(B - 1)$, então $4(A + 1) \sim N(4\mu, 4\sigma^2)$ e $4(B - 1) \sim N(4\mu, 4\sigma^2)$.

Sabendo agora que

$$X \sim N(\mu, \sigma^2) \implies Y = cX \sim N(c\mu, c^2\sigma^2)$$

então, se $Z_A = 4(A+1)$, $A = (1/4)Z_A - 1$ e portanto $A \sim N(\mu - 1, (1/4)\sigma^2)$. Análogo, temos se $Z_B = 4(B-1)$, $B = (1/4)Z_B + 1$ e portanto $B \sim N(\mu + 1, (1/4)\sigma^2)$.

Por fim, como $B = A + 2$, A e B têm vínculo, e portanto, não são independentes.

5. Sejam A e B variáveis aleatórias do exercício anterior. Qual a probabilidade de $\mu \notin [A, B]$?

Como $B > A$, pois $B = A + 2$ então segue que

$$P(\mu \notin [A, B]) = 1 - P(A \leq \mu \leq B)$$

Como vale que

$$X \sim N(\mu, \sigma^2) \implies Y = (X - \mu)/\sigma \sim N(0, 1)$$

Então se $A \sim N(\mu - 1, 1/4)$, sendo $\mu_A = \mu - 1$ e $\sigma_A^2 = 1/4$, então temos que $Y_A = [A - (\mu - 1)]/(1/2)$ é tal que $Y_A \sim N(0, 1)$. Em termos de μ, podemos escrever a relação $-Y_A/2 + A + 1 = \mu$.

Logo, rearranjando a desigualdade para $\mu \notin [A, B]$, temos que

$$\begin{aligned} P(\mu \notin [A, B]) &= 1 - P(A \leq \mu \leq A + 2) \\ &= 1 - P(A \leq A + 1 - Y_A/2 \leq A + 2) \\ &= 1 - P(-1 \leq -Y_A/2 \leq 1) = 1 - P(2 \geq Y_A \geq -2) \end{aligned}$$

Com $Y_A \sim N(0, 1)$, por fim

$$P(\mu \notin [A, B]) = 1 - \int_{-2}^{2} N(0, 1) dx$$

6. Será extraída uma amostra de n elementos independentes $X_1, X_2, ...X_n$ da variável $X \sim N(1, \sigma^2)$, da qual a variância σ^2 é desconhecida. A parti dessa amostra será computada a estatística (o estimador)

$$Y^2 = \frac{1}{n-1} \sum_i (X_i - 1)^2$$

Qual a probabilidade de $Y^2/2 \leq \sigma^2 \leq 2Y^2$?

Responder para n pequeno com distribuição χ^2 e para n grande, usando a aproximação do teorema central do limite. Comparar com $n = 20$.

Sabemos que

$$X \sim N(0, 1) \implies Q = \sum_i^k X_i^2 \sim \chi^2(k)$$

dessa forma, podemos tomar $Z_i = (X_i - 1)/\sigma$ de modo que $X_i \sim N(1, \sigma^2))$, então $Z_i \sim N(0, 1)$. Logo, também podemos reescrever que $(X_i - 1) = Z_i\sigma$.

Assim, a estatística desejada pode ser reescrita como $Y^2(n-1) = \sum_i (X_i - 1)^2 = \sum_i (Z_i\sigma)^2 = \sigma^2 \sum_i Z_i$, fazendo com que $Y^2(n-1)/\sigma^2 = Q \sim \chi^2$. Ou também, em termos de σ^2, temos $\sigma^2 = Y^2(n-1)/Q$.

Para o intervalo desejado, temos que

$$\begin{aligned} P(Y^2/2 \leq \sigma^2 \leq 2Y^2) &= P(Y^2/2 \leq Y^2(n-1)/Q \leq 2Y^2) \\ &= P(1/[2(n-1)] \leq 1/Q \leq 2/(n-1)) \\ &= P(2(n-1) \geq Q \geq (n-1)/2) \end{aligned}$$

Logo

$$P(Y^2/2 \leq \sigma^2 \leq 2Y^2) = \int_{(n-1)/2}^{2(n-1)} \chi^2 dx$$

Caso n grande, é válida a aproximação de chi^2 para uma distribuição $N(\mu, \sigma^2)$ tais que $\mu = n$ e $\sigma^2 = 2n$. Assim

$$P(Y^2/2 \leq \sigma^2 \leq 2Y^2) = \int_{(n-1)/2}^{2(n-1)} N(n, 2n) dx$$

7. Será extraída uma amostra de n elementos independentes, $X_1, X_2, ..., X_n$ da variável $X \sim N(\mu, \sigma^2)$, da qual os parâmetros $\theta = (\mu, \sigma^2)$ são desconhecidos. São computadas as estatísticas $\overline{X}$ e S^2

$$\overline{X} = \frac{1}{n} \sum_i X_i \qquad S^2 = \frac{1}{n-1} \sum_i (X_i - \overline{X})^2$$

Qual a probabilidade de $\overline{X} - S/10 \leq \mu \leq \overline{X} + S/10$?

Responder para n pequeno usando a distribuição t de Student e para n grande usando aproximação pela distribuição normal.

Sabendo que, se $\overline{X}$, S^2 definidos acima,

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

com t_{n-1} distribuição t de Student com n graus de liberdade, dessa forma, temos que

$$-TS/\sqrt{n} + \overline{X} = \mu$$

E o intervalo desejado pode ser reescrito na forma

$$\begin{aligned} P(\overline{X} - S/10 \leq \mu \leq \overline{X} + S/10) &= P(\overline{X} - S/10 \leq -TS/\sqrt{n} + \overline{X} \leq \overline{X} + S/10) \\ &= P(-S/10 \leq -TS/\sqrt{n} \leq +S/10) \\ &= P(-\sqrt{n}/10 \leq -T \leq \sqrt{n}/10) \\ &= P(\sqrt{n}/10 \geq T \geq -\sqrt{n}/10) \end{aligned}$$

Logo

$$P(\overline{X} - S/10 \leq \mu \leq \overline{X} + S/10) = \int_{-\sqrt{n}/10}^{\sqrt{n}/10} t_{n-1} dx$$

Para n grande e aproximação da distribuição t para uma normal, temos válido que

$$E[t_k] = 0 \qquad Var[t_k] = \frac{k}{k-2}$$

Logo, para n grande a distribuição tende a $N(0, k/k-2)$, com $k = n-1$. Assim

$$P(\overline{X} - S/10 \leq \mu \leq \overline{X} + S/10) = \int_{-\sqrt{n}/10}^{\sqrt{n}/10} N(0, n/(n-3))dx$$

8. Seja X uma variável aleatória com distribuição de Bernoulli, com probabilidade p de tomar valor 1 e $1-p$ de tomar valor 0. Note que $E[X] = p$, mas p é desconhecido, em princípio. São extraídas n observações $X_1, X_2, ..., X_n$ de X, independentes. Construa um intervalo de confiança de 95% para p a partir da estatística $f = \overline{X}$ (frequência amostral).

Sendo $\overline{X} = \sum_k X_k/n$ e $\sum_k X_k \sim Binom(n,p)$, então temos que $n\overline{X} = \sum_k X_k \sim Binom(n, np)$

Para que o intervalo tenha 95% de confiança, $1 - \alpha = 0.95 \implies \alpha = 0.05$ e $\alpha/2 = 0.025$. Sendo E extremo do intervalo de confiança, sabemos que

$$E = (z_{\alpha/2})\frac{s}{\sqrt{n}}$$

Com $z_{\alpha/2}$ percentil da distribuição relativa a $\alpha/2$.

Como é válido que, para a distribuição binomial, $Var[X] = \sigma^2 = p(1-p)$, então $s = \sqrt{p(1-p)}$. O limite E é dado por

$$E = \sum_{i=0}^{np} \binom{n}{i} p^i (1-p)^{n-i} \frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

Em abordagem "conservadora", ou seja, considerando um intervalo maior aproximado por uma distribuição normal (n grande), podemos assumir válido que

$$E = z_{\alpha/2}\frac{1}{2\sqrt{n}}$$

Dessa forma

$$E = \int_{-\infty}^{0.0025} N(0,1)\frac{1}{2\sqrt{n}}dx$$

representa os extremos (esquerda, para $\alpha/2$, e direita para $1-\alpha/2$) do intervalo de confiança.