Estatística - Listas de Exercícios

Jorge Augusto Salgado Salhani

Agosto, 2022

1 Lista 2 - Probabilidades

1.1 Considere W a variável aleatória definida como o número de caras menos o número de coroas com três jogadas de uma moeda (o experimento é, assim, "lançar três vezes sucessivas uma moeda").

(a) Considerando a moeda honesta, liste os elementos do espaço amostral para três lançamentos da moeda e, para cada ponto amostral, atribua o valor w de W.

Sendo três lançamentos de moedas com cara (C), com probabilidade p_c de ocorrência, e coroa (K), com probabilidade $p_k = 1 - p_c$ de ocorrência, o espaço amostral, a variável $w \in W$ dada por $w = \#C + \#K$ e a probabilidade de ocorrência são

Ω	(CCC)	(CCK)	(CKC)	(CKK)	(KCC)	(KCK)	(KKC)	(KKK)
W	3	1	1	-1	1	-1	-1	-3
P	p_c^3	$p_c^2 p_k$	$p_c^2 p_k$	$p_k^2 p_c$	$p_c^2 p_k$	$p_k^2 p_c$	$p_k^2 p_c$	p_k^3

Logo $\{(W, P)\} = \{(3, p_c^3), (1, 3p_c^2 p_k), (-1, 3p_k^2 p_c), (-3, p_k^3)\}$. Para moeda honesta, $p_c = p_k = 1/2$ então $\{(W, P)\} = \{(3, 1/2^3), (1, 3(1/2^3)), (-1, 3(1/2^3)), (-3, 1/2^3)\}$

(b) Se a moeda não fosse honesta, o espaço amostral seria diferente e os valores atribuídos seriam diferentes?

Não. Os elementos do espaço amostral e, junto com eles, os valores que a variável aleatória $w \in W$ apresentam independe da probabilidade de ocorrerem.

(c) Considerando a moeda honesta, calcule $E(W)$ e $V(W)$

Sendo válidas as expressões para o valor esperado $E[W]$ e variância $V[W]$

$$E[W] = \sum_{i=1}^{N} w_i P(w_i) = \mu$$

$$V[W] = \begin{cases} \sum_{i=1}^{N} P(w_i)[w_i - \mu]^2 & \\ E[(W-\mu)^2] & = E[W^2 - 2W\mu + \mu^2] = E[W^2] - 2E[W]\mu + \mu^2 \\ & = E[W^2] - (E[W])^2 \end{cases}$$

Usando do conjunto $\{W, P\}$ do item **(a)**, temos que

$$E[W] = 3p_c^3 + 3p_c^2 p_k + (-3)p_k^2 p_c + (-3)p_k^3 = 0$$

onde assumimos que $p_c = p_k = 1/2$ para moeda honesta. Para o cálculo da variância, também temos que

$$E[W^2] = 9p_c^3 + 3p_c^2 p_k + 3p_k^2 p_c + 9p_k^3 = 3$$

Logo, $V[W] = E[W^2] - E[W]^2 = 3$.

(d) Assumindo que a moeda seja viciada, de modo que cara seja duas vezes mais provável de ocorrer do que coroa, determine a distribuição de probabilidade da variável aleatória W, sua função de distribuição acumulada, sua esperança e sua variância.

Reutilizando do resultado anterior, $\{(W, P)\} = \{(3, p_c^3), (1, 3p_c^2 p_k), (-1, 3p_k^2 p_c), (-3, p_k^3)\}$, teremos agora que $p_c = 2p_k \implies p_c = 2(1 - p_c) = 2 - 2p_c \implies p_c = 2/3$. Assim, temos que

$$\begin{aligned} \{(W, P)\} = \{&(3, (2/3)^3), \\ &(1, 3(2/3)^2(1/3)) \\ &(-1, 3(1/3)^2(2/3)) \\ &(-3, (1/3)^3)\} \end{aligned}$$

Sendo a distribuição de probabilidade acumulada dada por

$$F_W(w) = P(W \le w)$$
$$F_W(w) = \sum_{w_i \le w} P(W = w_i)$$

$$\begin{aligned} \{(W, F)\} = \{&(-\infty \le w < -3), 0 \\ &(-3 \le w < -1), (1/3)^3 + \sum F_0, \\ &(-1 \le w < 1, 3(1/3)^2(2/3)) \sum F_1, \\ &(1 \le w < 3), 3(2/3)^2(1/3) + \sum F_2, \\ &(3 \le w < +\infty), (2/3)^3 + \sum F_3\} \end{aligned}$$

A esperança é dada por

$$E[W] = 3(2/3)^3 + 3(2/3)^2(1/3) + (-3)(1/3)^2(2/3) + (-3)(1/3)^3 = 1$$

Para a variância, precisamos da quantidade

$$E[W^2] = 9(2/3)^3 + 3(2/3)^2(1/3) + 3(1/3)^2(2/3) + 9(1/3)^3 = 3.5556$$

Logo $V[W] = 2.5556$.

(e) Sobre o mesmo experimento, consideremos a variável aleatória Y definida como sendo 1 se o segundo lançamento for cara e 0 se não. Qual a distribuição de probabilidade de Y? E a função de distribuição acumulada? Calcular também a esperança e variância.

Como a ocorrência de C no segundo lançamento ocorre em 3 de 6 possibilidade de modo aleatório, temos que $\{W, P\} = \{(1, p_c), (0, (1 - p_c))\}$ distribuição de probabilidade. Também $\{(W, F)\} = \{(-\infty \leq w < 0, 0), (0 \leq w < 1, (1 - p_c)), (1 \leq w < +\infty, 1)\}$.

Para o valor esperado, temos $E[W] = p_c$, $E[W^2] = p_c$ e $V[W] = p_c - p_c^2 = p_c(1 - p_c)$.

1.2 Experimento: três cartas são retiradas, sucessivamente, de um baralho sem reposição.

(a) Determine a distribuição de probabilidade f_X para a variável aleatória X, definida como o "número de espadas". Calcule também a esperança $E[X]$ e a variância $V[X]$

Para este caso, temos um experimento dentro de uma população finita com N elementos (total de cartas) contendo K elementos bem sucedidos (total de cartas de espada), das quais retiramos sucessivamente n elementos (número de cartas escolhidas ao acaso) e obtemos k sucessos observados (número de cartas de espadas escolhidas).

Assim sendo, considerando que existam $\binom{K}{k}$ ocorrências de sucesso, $\binom{N-K}{n-k}$ ocorrências de não-sucesso e $\binom{N}{n}$ ocorrências em totalidade, vale a distribuição hipergeométrica dada por

$$p_X(k) = P(X = k) = \frac{\#\text{sucessos} \times (1 - \#\text{sucessos})}{\#\text{ocorrências}} = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Retiramos $n = 3$ cartas. Portanto o número de espadas (k) apresenta valores de $k = 0, 1, 2, 3$ de um espaço fixo de $N = 52$ cartas das quais $K = 12$ são de espadas. Logo

k	0	1	2	3
$f_{X=k}$	0.4135	0.4359	0.1376	0.0129

Além disso, temos que $E[X] = 0.75$ e $V[X] = 0.5404$

(b) Determine a distribuição de probabilidade f_Y para a variável Y definida como o "número de outros". Calcule também $E[Y]$ e $V[Y]$.

O resultado é idêntico ao anterior, visto que o número de ouros ocorre em mesma proporção que o número de espadas, com mesma chance de ocorrência

(c) Determine a distribuição de probabilidade para a variável Z, definida como o "número de espadas mais o número de ouros". Calcule também $E[Z]$ e $V[Z]$.

Retiramos $n = 3$ cartas. Sabemos que temos a ocorrência de, no máximo 3 cartas de ouro ou 3 cartas de espada. Logo $k = 0, 1, 2, 3$, tal como o item anterior. Temos fixo um espaço de $N = 52$ cartas, porém é considerado "sucesso" a escolha de cartas de ouro OU espada, fazendo com que $K = 26$ (ouros OU espadas). Logo

k	0	1	2	3
$f_{y=k}$	0.1176	0.3824	0.3824	0.1176

E também, $E[Y] = 1.5$ e $V[Y] = 0.7206$.

(d) Verifique se o evento $X = k$ (com $k = 0, 1, 2$) é independente do evento $Y = l$ (com $l = 0, 1, 2$).

Para que sejam independentes, deve valer que $P(X \cap Y) = P(X)P(Y)$. Sendo $P(X \cap Y) = P(X)$, então os eventos não são independentes.

1.3 Seja X a variável aleatória com a seguinte distribuição de probabilidade

x	-3	6	9
$f(x)$	1/6	1/2	1/3

(a) Calcule $E[(2X + 1)^2]$, utilizando $E[X]$ e $E[X^2]$

Expandindo a expressão inicial, temos $E[(2X + 1)^2] = E[4X^2 + 4X + 1] = 4E[X^2] + 4E[X] + 1$.

Temos que $E[X] = 5.5$ e $E[X^2] = 46.5$. Logo, $E[(2X + 1)^2] = 209$.

(b) Sejam $X_1 \sim X$ e $X_2 \sim X$. Calcule a distribuição de probabilidade de $Y = X_1 + X_2$ se X_1 e X_2 são independentes. Também $E[Y]$ e $V[Y]$.

Ao combinarmos as variáveis X_1 e X_2 temos

	-3	6	9
-3	-6	3	6
6	3	12	15
9	6	15	18

Como as variáveis são independentes, a probabilidade de cada célula $P(X_1 \cap X_2) = P(X_1)P(X_2)$. Logo, a distribuição de probabilidade é dada por

y	-6	3	6	12	15	18
$P(Y = y)$	1/36	(2)1/32	(2)1/18	1/4	(2)1/6	1/9

Com ela, temos que $E[Y] = 10.6875$ e $V[Y] = 77.0519$.

(c) Calcule a distribuição de probabilidade de $Z = 2X$. Também $E[Z]$ e $V[Z]$.

Para a variável Z temos

z	-6	12	18
$P(Z = z)$	1/6	1/2	1/3

Temos que $E[Z] = 11$ e $V[Z] = 65$.

1.4 Sendo X uma variável aleatória seguindo uma distribuição Uniforme Discreta, com valores no conjunto $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Determine:

(a) $P(X \geq 7)$

Sendo uma distribuição uniforme, sabemos que todos os eventos apresentam probabilidade $P(X = x) = 1/10$. Logo

$$P(X \geq 7) = \sum_{i=7}^{10} P(X = x_i) = 4(1/10) = 0.4$$

(b) $P(3 < X \leq 7)$

Tal como o item anterior, temos que $P(3 < X \leq 7) = \sum_{i=4}^{7}(1/10) = 4(1/10) = 0.4$.

(c) $P(X \leq 7 | X \geq 6)$

Aqui, sabemos que $P(A \cap B) = P(A|B)P(B) \implies P(A|B) = P(A \cap B)/P(B)$.

Em particular, temos $P(X \leq 7) = 7(1/10)$, $P(X \geq 6) = 5(1/10)$, $P(X \leq 7 \cap X \geq 6) = P(X = 6 \cup X = 7) = 2(1/10)$.

Logo $P(X \leq 7 | X \geq 6) = 2/5 = 0.4$.

(d) $E[X]$ e $V[X]$

Para a distribuição, temos $E[X] = 5.5$ e $V[X] = 8.25$.

(e) Os eventos $X \leq 7$ e $X \leq 4$ são independentes?

Para que sejam independentes, $P(A \cap B) = P(A)P(B)$. Como para a intersecção, $P(X \leq 7 \cap X \leq 4) = P(X \leq 4)$, os eventos não são independentes.

1.5 Determine o valor de $c \in \mathbf{R}$, de modo que a função a seguir represente uma distribuição de probabilidade da variável aleatória X.

$$f(x) = c\binom{5}{x}\binom{7}{4-x}, \text{para } x = 0, 1, 2.$$

Como $X = \{0, 1, 2\}$ e a distribuição de probabilidade deve ser tal que $\sum_i P(x_i) = 1$, temos que

$$c\binom{5}{0}\binom{7}{4} + c\binom{5}{1}\binom{7}{3} + c\binom{5}{2}\binom{7}{2} = 1$$

$$c(420) = 1 \implies c = 1/420 = 0.0024$$

(a) Suponha $X_1 \sim X$ e $X_2 \sim X$ independentes. Calcular a distribuição de probabilidade $Y = X_1 + X_2$. Calcular também $E[Y]$ e $V[Y]$ e a distribuição de probabilidade de $Z = 2X$.

Numericamente a variável $x \sim X$ apresenta distribuição

x	0	1	2
$f(x)$	0.0840	0.420	0.5040

Para duas variáveis com mesma distribuição $\sim X$ temos

	0	1	2
0	0	1	2
1	1	2	3
2	2	3	4

Cuja distribuição de probabilidade é dada por

x	0	1	2	3	4
$f(x)$	0.0071	0.0706	0.2611	0.2117	0.2540

Portanto $E[Y] = 2.2438$ e $V[Y] = 1.0647$.

A distribuição de probabilidade $Z = 2X$ é idêntica à de X.

1.6 Suponha que um conjunto de 116 celulares contenham 6 com defeitos e 110 que funcionem normalmente. Se X é o número de celulares defeituosos em uma amostra de 18 celulares escolhidos aleatoriamente do conjunto, determine $P(X > 3)$.

Sabemos que no conjunto de $N = 116$ existe um número finito de $K = 6$ itens defeituosos. Ao selecionarmos $n = 18$ itens (supondo que não haja reposição), a probabilidade que um número k de eventos (itens defeituosos) ocorram apresenta distribuição hipergeométrica dada por

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Como $P(X > 3) = 1 - P(X \leq 3)$ e $P(X \leq 3) = \sum_{i=0}^{3} P(X = x_i)$, temos que $P(X > 3) = 0.0052$.

1.7 Demonstre as seguintes propriedades do valor esperado de uma variável aleatória X

(a) Se $X = C$, onde C é uma constante, então $E[X] = C$.

Sendo válido que $P(X = C) = 1$, pois X assume apenas um valor único igual a C, temos que

$$E[X] = \sum_i P(X = x_i)x_i = \sum_i 1C = C.$$

(b) Se C é uma constante, então $E[CX] = CE[X]$.

Partindo da definição, temos que

$$E[CX] = \sum_i P(X = x_i)[Cx_i] = C \sum_i P(X = x_i)x_i = CE[X]$$

1.8 Demonstre as seguintes propriedades da variância de uma variável aleatória X

(a) Se C for uma constante, então $V[X+C] = V[X]$

Como podemos escrever $V[X] = E[X^2] - E^2[X]$, temos que

$$V[X+C] = E[(X+C)^2] - E^2[(X+C)] = E[X^2 + 2XC + C^2] - \left[\sum_i P(X = x_i)(x_i + C)\right]^2$$
$$= E[X^2] + 2CE[X] + C^2 - \left[\sum_i P(X = x_i)x_i + C\right]^2 = E[X^2] + 2CE[X] + C^2 - \Big[E[X] + C\Big]^2$$
$$= E[X^2] + 2CE[X] + C^2 - E^2[X] - 2E[X]C - C^2 = E[X^2] - E^2[X] = V[X]$$

Logo $V[X+C] = V[X]$, como gostaríamos.

(b) Se C for uma constante, então $V[CX] = C^2V[X]$

Como no item anterior, temos que

$$V[CX] = E[(CX)^2] - E^2[CX] = E[C^2X^2] - \left[\sum_i P(X = x_i)(Cx_i)\right]^2$$
$$= C^2E[X^2] - \left[C\sum_i P(X = x_i)x_i\right]^2 = C^2E[x^2] - C^2E[X] = C^2[E[X^2] - E[X]] = C^2V[X]$$

Logo $V[CX] = C^2V[X]$, como gostaríamos.

1.9 Um vendedor de cachorro quente trabalha na porta do estádio do Morumbi em dias de jogo. Ele pode deixar preparado 5, 6 ou 7 dúzias de sanduíches, que custam a ele R\$5 a dúzia. Sabe-se que a procura por cachorro quente (X), no seu ponto, é uma variável aleatória com a função de probabilidade a seguir. Sabe-se que cada dúzia de sanduíche é vendida a R\$ 12 e os sanduíches não vendidos vão para um canil que cpaga R\$2 pela dúzia.

X	4	5	6	7
p_i	0.2	0.3	0.3	0.2

(a) Qual é o número N de dúzias de sanduíches que devem ser preparados de modo a maximizar o lucro médio do vendedor?

(b) Se o número de dúzias preparadas é fixado em N (o resultado de **(a)**), qual é a distribuição de probabilidade da variável aleatória Y definida como "lucro obtido no dia"? Calcular $E[Y]$ e $V[Y]$.

(c) Nas condições do item **(b)**, e assumindo que o que acontece cada dia é independente do que aconteceu no dia anterior, qual a distribuição de probabilidade da variável aleatória Z, definida como "lucro médio obtido em dois dias seguidos"? E se os dias não fossem consecutivos? Calcular $E[Z]$ e $V[Z]$.

1.10 O número de pedidos de reparo que uma construtora recebe por mês é uma variável aleatória. Em média, são recebidos 7.5 pedidos por mês. Determine a probabilidade de que em um mês qualquer, a construtora receba:

(a) Exatamente 2 pedidos de reparo.

Como lidamos com uma taxa média de pedidos/mês fixa, o processo pode ser modelado como uma distribuição de Poisson tal que $\lambda = 7.5$.

Sabemos que

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Dessa forma, temos que para a ocorrência de exatamente $k = 2$ pedidos, $P(X = 2) = 0.0156$

(b) No máximo 2 pedidos de reparo.

Utilizando a mesma distribuição anterior, para que ocorram no máximo 2 pedidos, temos $P(X \leq 2) = \sum_{i=0}^{2} P(X = x_i)$. Assim $P(X \leq 2) = 0.0203$.

(c) No mínimo 8 pedidos de reparo.

Análogo aos itens anteriores, temos que $P(X \geq 8) = \sum_{i=8} P(X = x_i) = 1 - \sum_{i=0}^{7} P(X = x_i)$. Logo $P(X \geq 8) = 0.4754$.

Imagine agora que o período analisado é de 10 meses. A média de pedidos é 90 por ano. Calcular a probabilidade de que, num período qualquer, a construtora receba:

(a) Exatamente 20 pedidos de reparo.

(b) No máximo 20 pedidos de reparo.

(c) No mínimo 80 pedidos de reparo.