Variável aleatória discreta - Distribuição de probabilidades - Distribuição acumulada de probabilidades

Quando se pensa em um jogo, como aconteceu historicamente no início da Teoria das Probabilidades, além de ser fundamental conhecer todos os resultados do jogo (que são representados no espaço amostra) e respectivas probabilidades, temos a definição de uma característica, na qual podemos apostar. Por exemplo, ao se lançar um dado, a característica pode ser o número da face superior; nesse caso, temos as seguintes possibilidades: face 1, face 2, face 3, face 4, face 5 e face 6 para apostar. Essa característica, *número da face superior do dado* é denominada **variável aleatória**.

As **variáveis aleatórias discretas (v.a.d.)** são aquelas provenientes de contagem. Por exemplo, no lançamento de três moedas, *o número de caras obtidas* é uma variável aleatória discreta e *a soma dos pontos obtidos no lançamento de dois dados* é outro exemplo de variável aleatória discreta.

As **variáveis aleatórias contínuas (v.a.c.)** são aquelas provenientes de medidas. Por exemplo, *altura dos alunos de uma escola* e *volume líquido em garrafas de 750 ml* são variáveis aleatórias contínuas. Elas ficarão para serem estudadas posteriormente.

Uma variável aleatória discreta é, portanto, uma **função** que a cada elemento do espaço amostra associa um número real e deve ser representada por uma letra maiúscula. No exemplo citado em que a v.a.d. é o *número de caras obtidas em três lances de uma moeda*, vamos representá-la por *X*. O espaço amostra associado ao experimento, como já estudamos, é dado por:

$$\Omega = \{(c,c,c)\,,(c,c,k)\,,(c,k,c)\,,(c,k,k)\,,(k,c,c)\,,(k,c,k)\,,(k,k,c)\,,(k,k,k)\}$$

Esse espaço amostra é o domínio da função X ($D(X) = \Omega$), o contradomínio é o conjunto dos números reais ($CD(X) = \mathbb{R}$) e as imagens de *X* (*número de caras),* para cada elemento do domínio, são:

$X(c,c,c) = 3$ $\quad X(c,c,k) = 2$ $\quad X(c,k,c) = 2$ $\quad X(c,k,k) = 1$

$X(k,c,c) = 2$ $\quad X(k,c,k) = 1$ $\quad X(k,k,c) = 1$ $\quad X(k,k,k) = 0$

Assim, $Im(X) = \{0,1,2,3\}$, ou, podemos dizer que os valores da v.a.d. *X* são: 0, 1, 2 e 3.

Quando se fala em jogo, o interesse é pelas probabilidades de ganho. Assim, é importante saber os valores em que podemos apostar, que no caso do exemplo, são os valores de *X*, ou seja, 0, 1, 2 e 3, mas, também, precisamos conhecer as probabilidades de ocorrência de cada valor, isto é, as chances de ganho ao se apostar em cada valor. Essa necessidade é modelada por outra função, denominada **função distribuição de probabilidades da variável aleatória *X*** e representada por p_X (note que a letra é minúscula, diferente da função probabilidade, já estudada, que é representada pela letra *P* maiúscula).

A **função distribuição de probabilidades** de uma v.a.d. é uma função que tem como domínio a imagem da v.a. e como contradomínio, um número real entre zero e um, porque representa probabilidade. Essa função é definida como a probabilidade da v.a. assumir cada valor, probabilidade essa determinada a partir das probabilidades obtidas no espaço amostra. No caso, como nada foi dito em relação à moeda, deve-se considerar que ela seja honesta. Assim, teremos que o espaço amostra é equiprovável e, então, $P(c,c,c) = P(c,c,k) = P(c,k,c) = P(c,k,k) = P(k,c,c) = P(k,c,k) == P(k,k,c) = P(k,k,k) = 1/8$

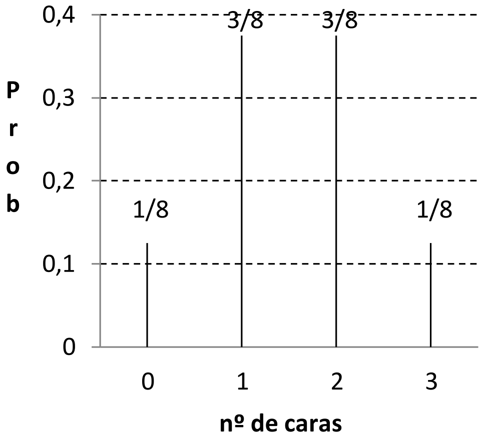
No exemplo em estudo, a $Im(X) = \{0,1,2,3\}$ é o domínio da função p_X e as imagens são dadas por:

$$p_X(0) = P(X = 0) = P(k, k, k) = 1/8$$
$$p_X(1) = P(X = 1) = P[(c, k, k) \cup (k, c, k) \cup (k, k, c)] = 1/8 + 1/8 + 1/8 = 3/8$$
$$p_X(2) = P(X = 2) = P[(c, c, k) \cup (c, k, c) \cup (k, c, c)] = 1/8 + 1/8 + 1/8 = 3/8$$
$$p_X(3) = P(X = 3) = P(c, c, c) = 1/8$$

Essa forma de construir a função distribuição de probabilidades da v.a.d. pode ser simplificada, quando a representamos na forma de uma **tabela**, denominada **distribuição de probabilidades da v.a.d.**, como apresentada a seguir.

X	0	1	2	3
P(X)	1/8	3/8	3/8	1/8

Como as probabilidades esgotam o espaço amostra, a **soma das probabilidades deve totalizar um**.



Além da tabela, podemos representar a distribuição de probabilidades de uma v.a.d. por meio de um **gráfico**, que deverá ter como representação gráfica, um segmento de reta. O uso de *softwares,* nem sempre permite que se faça um segmento, mas, apenas, colunas estreitas, que podem ser aceitas, quando se utiliza *softwares*.

A distribuição de probabilidades é a modelagem matemática (modelo teórico) da distribuição de frequências, que se encontra quando fazemos a parte experimental.

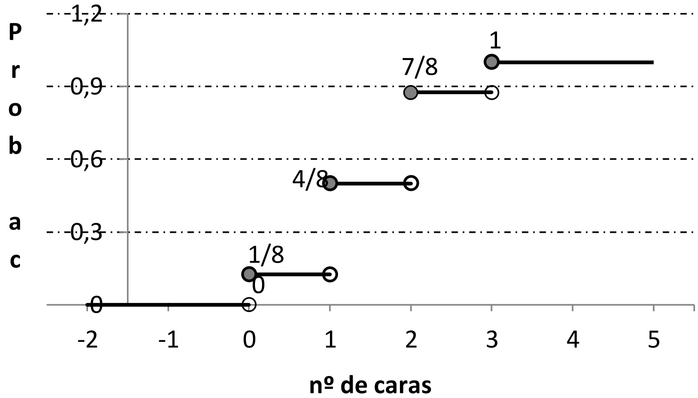
Assim como na prática, muitas vezes precisamos de informações acumuladas. Também na modelagem podemos falar em **função distribuição acumulada de probabilidades** de uma variável aleatória e correspondentemente, em distribuição de probabilidades acumulada dessa variável.

A função distribuição acumulada de probabilidades (ou função distribuição de probabilidades acumulada, ou, simplesmente, função repartição) de uma variável aleatória X é definida no conjunto dos números reais e com valores no intervalo real de 0 a 1, por: $\boldsymbol{F(t) = P(X \leq t)}$, $\forall\, \boldsymbol{t} \in \mathbb{R}$**.**

No caso do exemplo, em que a variável aleatória discreta é definida por *X: número de caras obtidas em três lances de uma moeda*, a função distribuição de probabilidades acumulada de X pode ser representada por fórmula, ou gráfico (***distribuição acumulada de probabilidades – ou, distribuição de probabilidades acumulada - função repartição - da v.a.d. X***):

$F : \mathbb{R} \rightarrow [0\,,1] \subset \mathbb{R}$, tal que:

$$F(t) = P(X \leq t) = \begin{cases} 0 & se \quad t < 0 \\ \dfrac{1}{8} & se \quad 0 \leq t < 1 \\ \dfrac{4}{8} & se \quad 1 \leq t < 2 \\ \dfrac{7}{8} & se \quad 2 \leq t < 3 \\ 1 & se \quad t \geq 3 \end{cases}$$



Observe que a distribuição de probabilidades da v.a.d. X pode ser obtida diretamente da função repartição, observando que a v.a.d. X assume valores, com probabilidade não nula, exatamente nos pontos em que a função repartição dá "um salto" e, ainda, a probabilidade, em cada ponto é igual ao "tamanho do salto". Assim, no exemplo, temos que $P(X=2)=F(2)-F(1)=\frac{7}{8}-\frac{4}{8}=\frac{3}{8}$.

As probabilidades de uma v.a. assumir valores pertencentes a um intervalo real podem ser determinadas valendo-se da função distribuição acumulada de probabilidades. Retomando a definição da função acumulada, $\boldsymbol{F(t)=P(X\leq t)}$, $\forall\, \boldsymbol{t} \in \mathbb{R}$**,** temos os seguintes casos a considerar:

$\boldsymbol{P(a<X\leq b)=F(b)-F(a)}$ **-** note que $\boldsymbol{F(b)}$ já inclui $\boldsymbol{P(X=b)}$, como queremos e quando tiramos $\boldsymbol{F(a)}$, já tiramos $\boldsymbol{P(X=a)}$ que não queremos;

$\boldsymbol{P(a\leq X\leq b)=F(b)-F(a)+P(X=a)}$ **-** note que $\boldsymbol{F(b)}$ já inclui $\boldsymbol{P(X=b)}$, como queremos, quando tiramos $\boldsymbol{F(a)}$, tiramos $\boldsymbol{P(X=a)}$ que queremos e, portanto, precisamos somar $\boldsymbol{P(X=a)}$;

$\boldsymbol{P(a<}X<b\boldsymbol{)=F(b)-F(a)-P(X=b)}$ **-** note que $\boldsymbol{F(b)}$ inclui $\boldsymbol{P(X=b)}$, como não queremos, precisamos tirar $\boldsymbol{P(X=b)}$**,** quando tiramos $\boldsymbol{F(a)}$, tiramos $\boldsymbol{P(X=a)}$ que não queremos;

$\boldsymbol{P(a\leq X<}b\boldsymbol{)=F(b)-F(a)-P(X=b)+P(X=a)}$ **-** note que $\boldsymbol{F(b)}$ inclui $\boldsymbol{P(X=b)}$, como não queremos, precisamos tirar $\boldsymbol{P(X=b)}$**,** quando tiramos $\boldsymbol{F(a)}$, tiramos $\boldsymbol{P(X=a)}$ que queremos e, portanto, precisamos somar $\boldsymbol{P(X=a)}$**.**

Outro exemplo:

No lançamento de dois dados, determine a distribuição de probabilidades, por tabela e gráfico, e a distribuição acumulada de probabilidades, por fórmula e gráfico da v.a.d.: Y: máximo entre os dois pontos obtidos. Determine, usando a função acumulada: (a) $P(2<Y\leq 6)$*; (b)* $P(2\leq Y\leq 4)$*; (c)* $P(Y>2{,}3)$*.*

Para que se possa obter a distribuição de probabilidades de qualquer v.a.d., inicialmente, é preciso determinar o espaço amostra e respectivas probabilidades.

O espaço amostra associado ao experimento do lançamento de dois dados é:

$$\Omega=\{(1{,}1),(1{,}2),(1{,}3),(1{,}4),(1{,}5),(1{,}6),(2{,}1),(2{,}2),(2{,}3),(2{,}4),(2{,}5),(2{,}6),$$
$$(3{,}1),(3{,}2),(3{,}3),(3{,}4),(3{,}5),(3{,}6),(4{,}1),(4{,}2),(4{,}3),(4{,}4),(4{,}5),(4{,}6),$$
$$(5{,}1),(5{,}2),(5{,}3),(5{,}4),(5{,}5),(5{,}6),(6{,}1),(6{,}2),(6{,}3),(6{,}4),(6{,}5),(6{,}6)\}$$

Como nada foi dito a respeito dos dados, devemos considerá-los honestos, o que equivale a dizer que o espaço amostra é equiprovável. Assim, como temos 36 elementos no espaço, cada elemento terá probabilidade igual a 1/36. Equivalentemente, podemos calcular a probabilidade de cada elemento do espaço amostra da seguinte forma: $P(1{,}1)=P(1_{1º}\cap 1_{2º})=P(1_{1º})\times P(1_{2º})=\frac{1}{6}\times\frac{1}{6}=\frac{1}{36}$, já que o número no segundo dado independe do número no primeiro dado (eventos independentes; probabilidade da intersecção é produto das probabilidades). Os demais se obtêm analogamente.

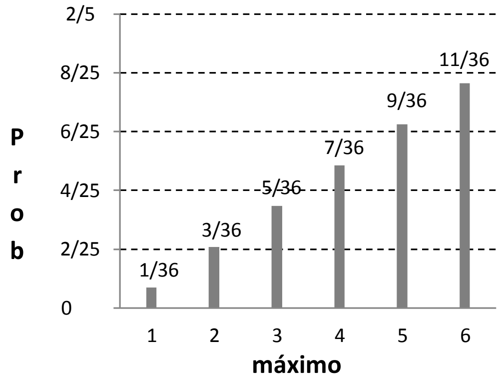
Como a variável Y é definida como *o máximo entre os dois pontos obtidos*, temos que o máximo pode ser: 1, 2, 3, 4, 5, e 6. Note que o *máximo* de $(2{,}3)$ é 3 e o *máximo* de $(4{,}4)$ é 4. Observe que precisamos definir "*o máximo entre os dois pontos*" e não "*o maior dos dois pontos*", porque o par $(4{,}4)$ não tem maior, apenas máximo.

A fim de determinarmos a distribuição de probabilidades de *Y*, precisamos determinar as probabilidades de cada valor que a variável *Y* pode assumir. Assim, $p_Y(1) = P(Y = 1) = P(1,1) = {}^{1}/_{36}$, $p_Y(2) = P(Y = 2) = P(\{(1,2) \cup (2,1) \cup (2,2)\}) = P(1,2) + P(2,1) + P(2,2) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36}$. Analogamente obtemos as demais probabilidades para os outros possíveis valores da v.a.d. *Y*.

Com as probabilidades obtidas, podemos fazer a representação, por tabela, da distribuição de probabilidades da v.a.d. *Y*, apresentada a seguir.

Y	1	2	3	4	5	6
P(Y)	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

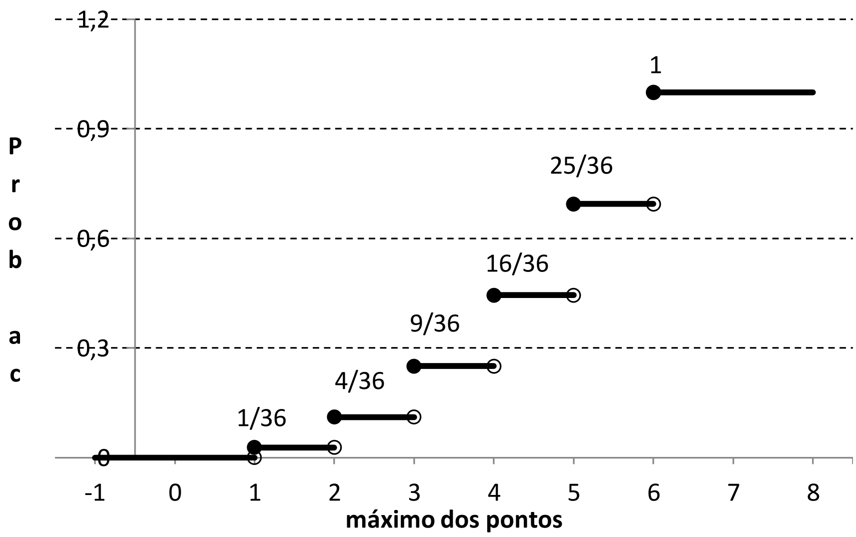
A representação gráfica da distribuição de probabilidades de Y é:



A função distribuição de probabilidades acumulada, por meio de uma fórmula, será, portanto:

$$F : \mathbb{R} \to [0\,,1] \subset \mathbb{R}, \text{ tal que: } \quad F(t) = P(X \le t) = \begin{cases} 0 & se \quad t < 1 \\ {}^{1}/_{36} & se \quad 1 \le t < 2 \\ {}^{4}/_{36} & se \quad 2 \le t < 3 \\ {}^{9}/_{36} & se \quad 3 \le t < 4 \\ {}^{16}/_{36} & se \quad 4 \le t < 5 \\ {}^{25}/_{36} & se \quad 5 \le t < 6 \\ 1 & se \quad t \ge 6 \end{cases}$$

Assim, a representação gráfica da distribuição de probabilidades acumulada da v.a.d. *Y* é dada por:



(a) $P(2 < Y \leq 6) = F(6) - F(2) = 1 - \frac{4}{36} = \frac{32}{36}$

(b) $P(2 \leq Y \leq 4) = F(4) - F(2) + P(Y = 2) = \frac{16}{36} - \frac{4}{36} + [F(2) - F(1)] = \frac{12}{36} + \left[\frac{4}{36} - \frac{1}{36}\right] = \frac{12}{36} + \frac{3}{36} = \frac{15}{36}$

(c) $P(Y > 2{,}3) = 1 - P(Y \leq 2{,}3) = 1 - F(2{,}3) = 1 - \frac{4}{36} = \frac{32}{36}$

A seguir, propomos alguns exercícios para resolver:

***Exercício 1**. Em um jogo, paga-se R$5,00 na compra de uma ficha que permite a participação. O jogador tem o direito a lançar um dado, que se considera honesto. Se o jogador obtém face correspondente a um número múltiplo de 3, tem o direito a lançar o dado novamente e, nesse caso, se sair face correspondente a um número primo, ganha o equivalente ao quádruplo do total de pontos obtidos nas duas jogadas, em reais. Determine a distribuição de probabilidades da v.a.d. "G: ganho do jogador em uma única participação", por tabela e gráfico, assim como da distribuição de probabilidades acumulada, por fórmula e gráfico.*

***Exercício 2.** No lançamento de dois dados, determine a distribuição de probabilidades, por tabela e gráfico, e a distribuição acumulada de probabilidades, por fórmula e gráfico da v.a.d.: X: soma dos pontos obtidos nos dois lançamentos. Determine usando a função acumulada:*

(a) $P(4 \leq X \leq 9)$; *(b)* $P(3 < X < 11)$; *(c)* $P(3{,}8 \leq X \leq 9{,}2)$; *(d)* $P(2 \leq X < 10)$; *(e)* $P(X < 8)$.

***Exercício 3**. Uma urna tem fichas numeradas, indistintas a menos do número. Ela tem três fichas com o número 2, cinco fichas com o número 3 e uma ficha com o número 4. Duas fichas são sorteadas, sem reposição, dessa urna. Determine a distribuição de probabilidades, por tabela e gráfico, e a distribuição acumulada de probabilidades, por fórmula e gráfico, da v.a.d. Z: soma dos pontos obtidos nas duas fichas selecionadas. Utilize a função acumulada para determinar:*

(a) $P(Z < 6{,}3)$; *(b)* $P(4 < Z < 7)$; *(c)* $P(4 \leq Z < 6)$; *(d)* $P(Z > 5)$; *(e)* $P(2{,}3 \leq Z \leq 6{,}4)$.

***Exercício 4**. No lançamento de três moedas em que a primeira é honesta, a segunda tem probabilidade de cara igual ao dobro da de coroa e a terceira é tal que a probabilidade de cara é 1/3 da de coroa, determine a distribuição de probabilidades, por tabela e a distribuição de probabilidades, por fórmula, da v.a.d. W: número de caras obtidas nos três lançamentos.*

***Exercício 5**. Três tetraedros regulares, com faces numeradas: 1, 2, 3 e 4 são lançados simultaneamente. O primeiro deles, a ser lançado, tem probabilidade de cada face diretamente proporcional ao número da face, o segundo é honesto e o terceiro é tal que a probabilidade de cada face par é igual a 0,2 e a probabilidade da face 3 é o quíntuplo da probabilidade da face 1. Determine a distribuição de probabilidades, por tabela, da v.a.d. S: soma dos pontos obtidos nos lançamentos dos três tetraedros (um tetraedro regular é uma pirâmide com quatro faces constituídas por triângulos equiláteros).*