Variáveis aleatórias discretas

2022-09-20

2. Exemplo 2 dos slides

Lançamento de um dado

Definição da variável A variável aleatória em causa será

X2 - número da face voltada para cima no lançamento de um dado

```
# valores possíveis para o dado 1 a 6, todos com prob 1/6

S2<-data.frame("X2"=1:6,"prob"=1/6)
S2
```

```
##   X2      prob
## 1  1 0.1666667
## 2  2 0.1666667
## 3  3 0.1666667
## 4  4 0.1666667
## 5  5 0.1666667
## 6  6 0.1666667
```

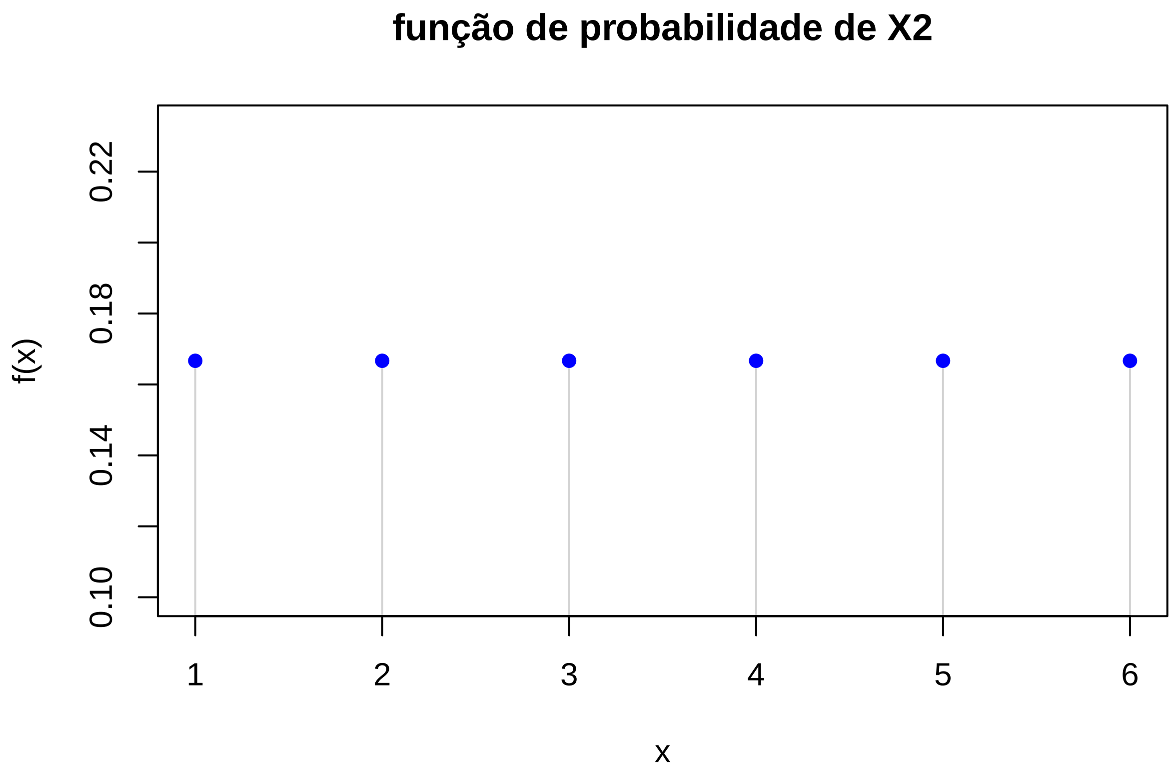
X2 contém os valores da variável, e *prob* a respetiva função de probabilidade.

Representação da função de probabilidade Para obter uma representação gráfica, podemos simplesmente usar a função *plot*

```
plot (x=S2$X2,                    # os valores da variável
      y=S2$prob ,                 # os valores da função de probabilidade
      type="h",                   # linhas verticais entre x e y
      main="função de probabilidade de X2",
      xlab="x",                   # id do eixo x
      ylab="f(x)",   # id do eixo y
      col="light grey"            # cor das linhas
      )

#Adicionar pontos ao gráfico

points(x=S2$X2,
       y=S2$prob,
       pch=16,                    # símbolo usado, 16 é círculo
       col="blue"
       )
```



Cálculo de probabilidade recorrendo à função de probabilidade Para obter a probabilidade de obter mais de 2 pontos no lançamento de um dado, $P[X_2 > 2]$, basta somar a função de probabilidade para os valores da variável que respeitam a condição.

Por exemplo

```
# somar as probabilidades dos valores de x2 (ou seja, em S2$X1) que respeitam a condição ">2"

sum(
  S2[S2$X2>2,]  #condição sobre as linhas (i.e. primeiro argumento)
  $prob)       #valores a somar
```

```
## [1] 0.6666667
```

Alternativamente, podemos usar primeiro uma seleção, o que permite condições mais complexas.

```
quais<-which(S2[,1]>2)       # quais as linhas a somar
sum(S2[quais,2])             # somar a coluna 2 nesses índices
```

```
## [1] 0.6666667
```

Definição da função de distribuição A função de distribuição $F(x)$ (notem a utilização de Maiúscula) para uma variável aleatória discreta é uma função definida em **patamares** (ou função em escada).

Podemos começar por definir os valores dos diferentes patamares, que correspondem às probabilidades acumuladas em cada um dos pontos do suporte de X.

Para isso podemos usar a função genérica **cumsum()**

```
FX2_pontos<-cumsum(S2$prob)
FX2_pontos
```

```
## [1] 0.1666667 0.3333333 0.5000000 0.6666667 0.8333333 1.0000000
```

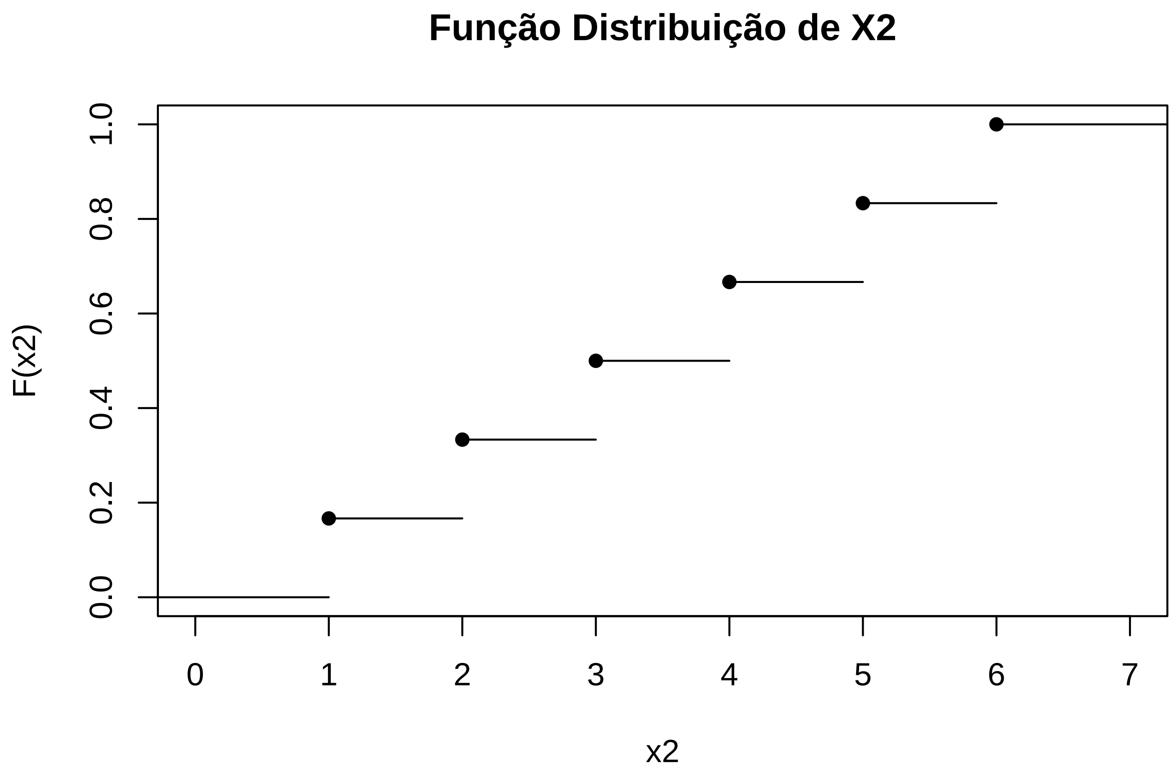
Escrever a função de distribuição

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.167 & 1 \leq x < 2 \\ 0.333 & 2 \leq x < 3 \\ 0.500 & 3 \leq x < 4 \\ 0.667 & 4 \leq x < 5 \\ 0.833 & 5 \leq x < 6 \\ 1 & x \geq 6 \end{cases}$$

Representação gráfica da função de distribuição Para representar graficamente a função em todo o seu domínio vamos recorrer a **stepfun**

```
#Representação gráfica da Função de Distribuição de X

plot.stepfun(
  stepfun(                  # definir a função em patamares
          S2$X2,            # os valores de x a considerar
          c(0,FX2_pontos),  # os patamares a considerar,
                            # ponto adicional inicial 0
          right=FALSE       # intervalos fechados à esquerda
          ),
  verticals=FALSE,          # não colocar traços verticais nos pontos de salto
  pch = 16,                 # tipo de símbolo
  #
  # título e identificação dos eixos
  #
  main="Função Distribuição de X2",
  xlab="x2",
  ylab="F(x2)")
```



Cálculo de probabilidades recorrendo à função de distribuição Recorde-se que

$$P[x_k < X \leq x_t] = \sum_{k+1}^{t} f(x_i) = F(x_t) - F(x_k)$$

Se pretendermos $P[2 < X_2 \leq 4]$ basta calcular $P[X_2 \leq 4] - P[X_2 < 2] = F(4) - F(2) = 0.3333333$

Note: inline code used

```
FX2_pontos[4]-FX2_pontos[2]
```

Se pretendermos $P[2 \leq X_2 \leq 4]$ não podemos fazer da mesma forma diretamente.

Mas, como X_2 é **discreta**, ser ≥ 2 equivale dizer que é > 1, pelo que neste caso podemos calcular $P[X_2 \leq 4] - P[X_2 \leq 1] = F(4) - F(1) = 0.5$

Note: inline code used

```
FX2_pontos[4]-FX2_pontos[1]
```

Valor esperado e variância numa v.a. discreta Seguindo a fórmula

$$E[X] = \sum_{1}^{n} x_i f(x_i)$$

, basta calcular a soma dos produtos da primeira com a segunda colunas definidas:

```
miu_x2<-sum(S2$X2*S2$prob)
miu_x2
```

```
## [1] 3.5
```

Para a variância, podemos aplicar a fórmula simplificada (média dos quadrados menos o quadrado da média) ou a definição formal (média dos quadrados dos desvios face à média).

```
# média dos quadrados menos o quadrado da média
var_x2<- sum((S2$X2^2)*S2$prob)-miu_x2^2
round(var_x2,4)
```

```
## [1] 2.9167
```

```
# média dos quadrados dos desvios face à média
round(sum((S2$X2-miu_x2)^2*S2$prob),4)
```

```
## [1] 2.9167
```

E assim o desvio-padrão será $\sigma_{X_2} = 1.71$