



# PROVA 2 DE PROCESSO ESTOCÁSTICOS

Rafael de Acypreste (200060023) e Rafael Lira (190115858)

Professor Felipe Quintino

# Prova 2 de Processo Estocásticos

## Aplicação ao modelo empírico

Trata-se de um modelo para avaliar as probabilidades de transição entre os estados de precipitação de chuvas.

```
# Importing data
dados <-
  read.delim("dados.txt",
    header = TRUE,
    sep     = ";"
  ) |>
  select(-X) |>
  filter(!is.na(Precipitacao))

# Summary statistics
dados$Precipitacao |> summary()
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.000	0.000	4.173	2.400	131.000

O primeiro passo é discretizar a variável de precipitação, que é feita com a função `cut()` do pacote `base`. Para esse exemplo, a variável será dividida em 3 categorias: sem chuva (precipitação até 0, 1), chuva fraca (precipitação maior que 0, 1 e menor que 10) e chuva forte.

```
# Discretization of the variable
quantiles <- quantile(dados$Precipitacao,
  probs = seq(0.7, 0.9, length.out = 3)
)

breaks <- c(-Inf, 0.001, quantiles, Inf)

state_labels <- factor(
  c(
    "sem chuva",
    "garoa",
    "chuva fraca",
    "chuva moderada",
    "chuva forte"
  ),
  levels = c(
    "sem chuva",
    "garoa",
    "chuva fraca",
    "chuva moderada",
    "chuva forte"
  )
)
```

```
# Discretization
dados <-
  dados |>
  # Discretization
  mutate(rain_status = cut(Precipitacao,
    breaks = breaks,
    labels = state_labels
  ))
```

Para esse exemplo, serão separadas as 10 últimas observações para avaliar as estimativas.

```
dados_teste <- tail(dados, 10)
dados_treinamento <- dados[1:(nrow(dados) - 10), ]
```

Para estimar as transições de estado, é necessário criar uma variável que identifique o estado atual e o estado seguinte. Para isso, é necessário criar uma variável defasada, que pode ser feita com a função `lag()` do pacote `dplyr`. Depois disso, basta avaliar as proporções das transições de estado.

```
# Creating the lagged variable
transicoes_chuva <-
  dados_treinamento |>
  # Lag variable
  mutate(rain_status_lag = lag(rain_status)) |>
  # Exclude the last state
  filter(!is.na(rain_status_lag)) |>
  # Count the transitions
  count(rain_status, rain_status_lag) |>
  # Calculates the estimator
  mutate(
    Prop = round(n / sum(n), digits = 3),
    .by = rain_status
  )
```

A matriz de transição estimada entre os estados sem chuva, garoa, chuva fraca, chuva moderada, chuva forte, nesta ordem, é dada por:

$$P = \begin{pmatrix} 0.822 & 0.041 & 0.051 & 0.045 & 0.04 \\ 0.334 & 0.12 & 0.186 & 0.177 & 0.183 \\ 0.306 & 0.133 & 0.187 & 0.184 & 0.19 \\ 0.306 & 0.121 & 0.171 & 0.193 & 0.21 \\ 0.282 & 0.118 & 0.174 & 0.204 & 0.223 \end{pmatrix} \quad (0.1)$$

Agora, pode-se recuperar a matriz de transição para fazer as estimativas de transição de estado.

```
# Transition matrix
matriz_transicao <-
  transicoes_chuva |>
  select(-n) |>
  pivot_wider(
    names_from = rain_status_lag,
```

```
      values_from = Prop
    ) |>
    column_to_rownames("rain_status") |>
    as.matrix()

ultimo_estado <- dados_treinamento |>
  tail(1) |>
  pull(rain_status)
```

Com a matriz de transição, basta considerar o último estado dos dados (garoa) — consequência da propriedade de Markov — de treinamento para fazer as estimativas de transição de estado.

```
simula_cadeia_markov <- function(n = 10,
                                valor_inicial,
                                matriz_transicao,
                                estados) {

  P <- matriz_transicao
  y <- valor_inicial

  # Simulation of the stochastic process
  for (i in 1:n) {
    # Sample of the next state
    y[i + 1] <- sample(estados, size = 1, prob = P[y[i], ])
  }

  return(y[-1])
}

# Excecution of the function
previsoes <- simula_cadeia_markov(
  valor_inicial = ultimo_estado,
  matriz_transicao = matriz_transicao,
  estados = state_labels,
  n = 10
)
```

E, então, pode-se comparar as previsões com os dados de teste. Para o gráfico, os acertos são indicados pela linha tracejada vermelha.

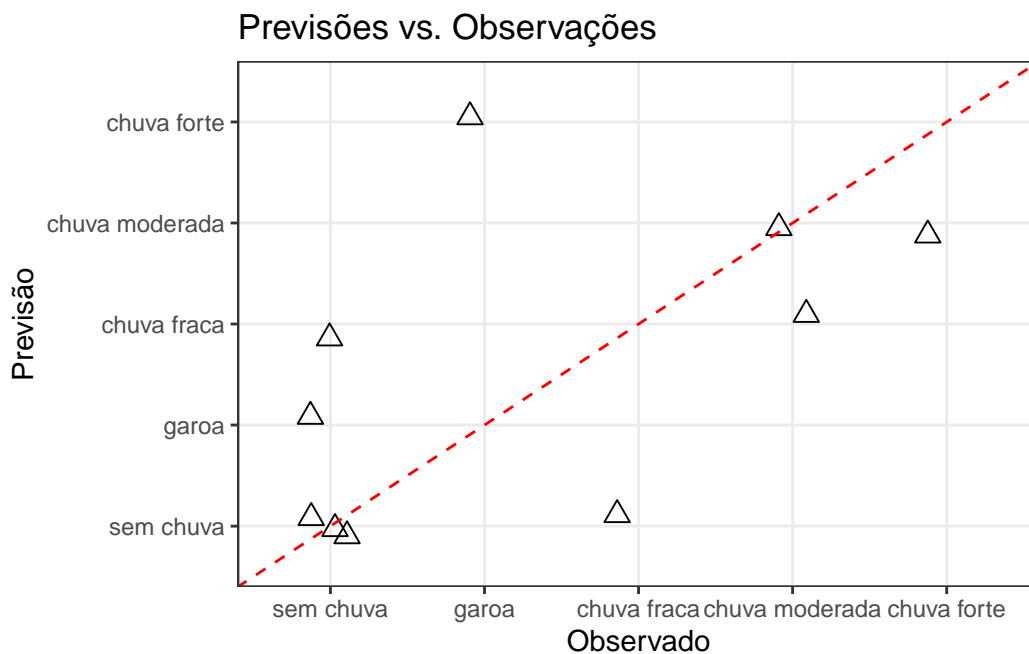
```
comparacao <-
  data.frame(
    observado = dados_teste$rain_status,
    previsao = previsoes
  )

# Imprime a tabela
comparacao
```

	observado	previsao
1	chuva moderada	chuva moderada
2	chuva moderada	chuva fraca
3	chuva forte	chuva moderada
4	sem chuva	garoa

5	sem chuva	chuva fraca
6	sem chuva	sem chuva
7	sem chuva	sem chuva
8	sem chuva	sem chuva
9	chuva fraca	sem chuva
10	garoa	chuva forte

```
# Constroi o gráfico
comparacao |>
  ggplot(aes(x = observado, y = previsao)) +
  geom_jitter(
    size = 3, shape = 2,
    width = 0.15, height = 0.15
  ) +
  geom_abline(
    intercept = 0,
    slope = 1,
    color = "red",
    linetype = "dashed"
  ) +
  theme_bw() +
  labs(
    x = "Observado",
    y = "Previsão",
    title = "Previsões vs. Observações"
  )
)
```





## Questão 2

Espaço para a resposta da questão 2