Introdução à modelagem unidimensional

Elias Teixeira Krainski
eliaskr@ufpr.br

Setembro 2017

Exemplo de motivação

Dias chuvosos em Tokyo

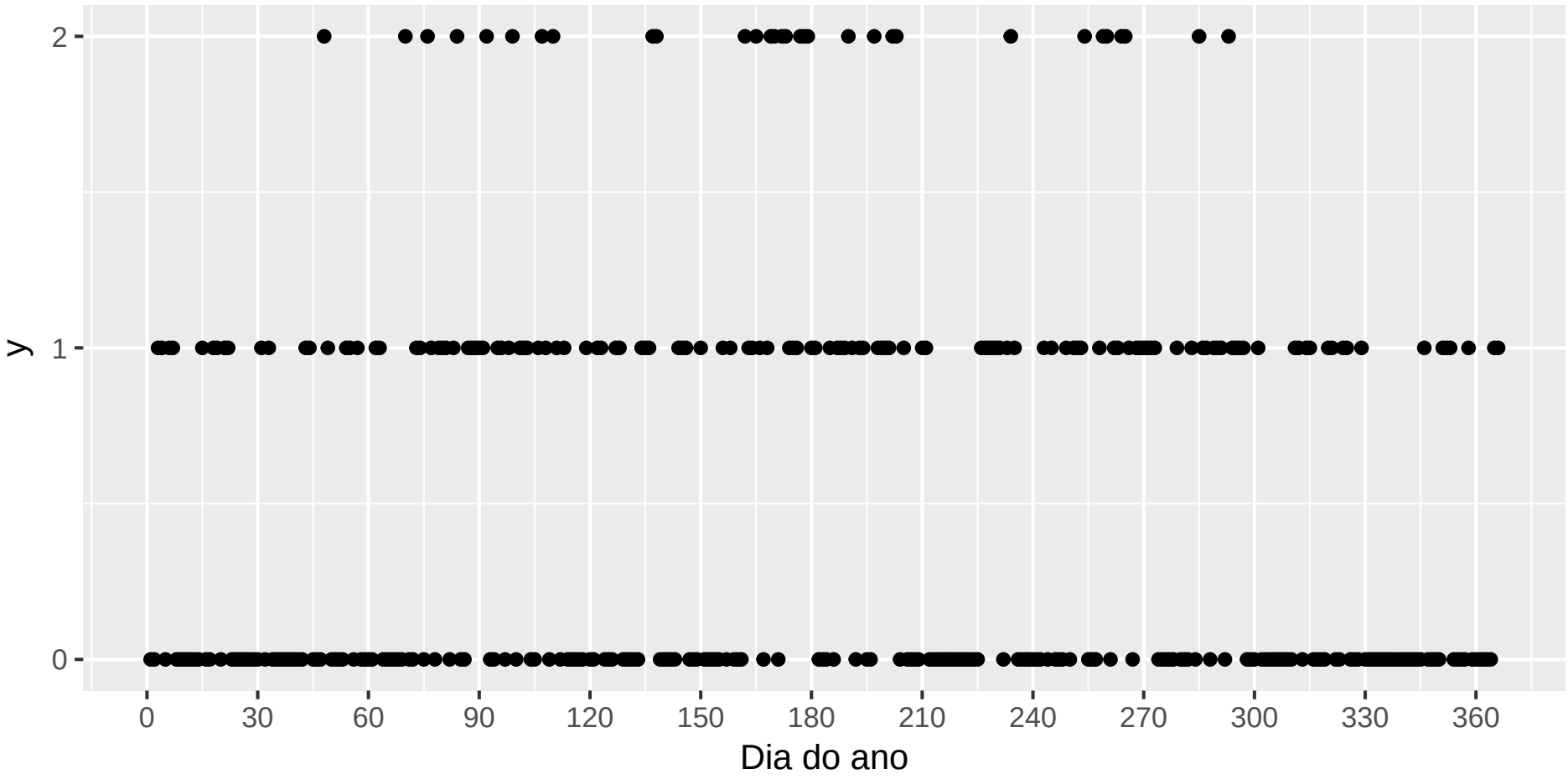
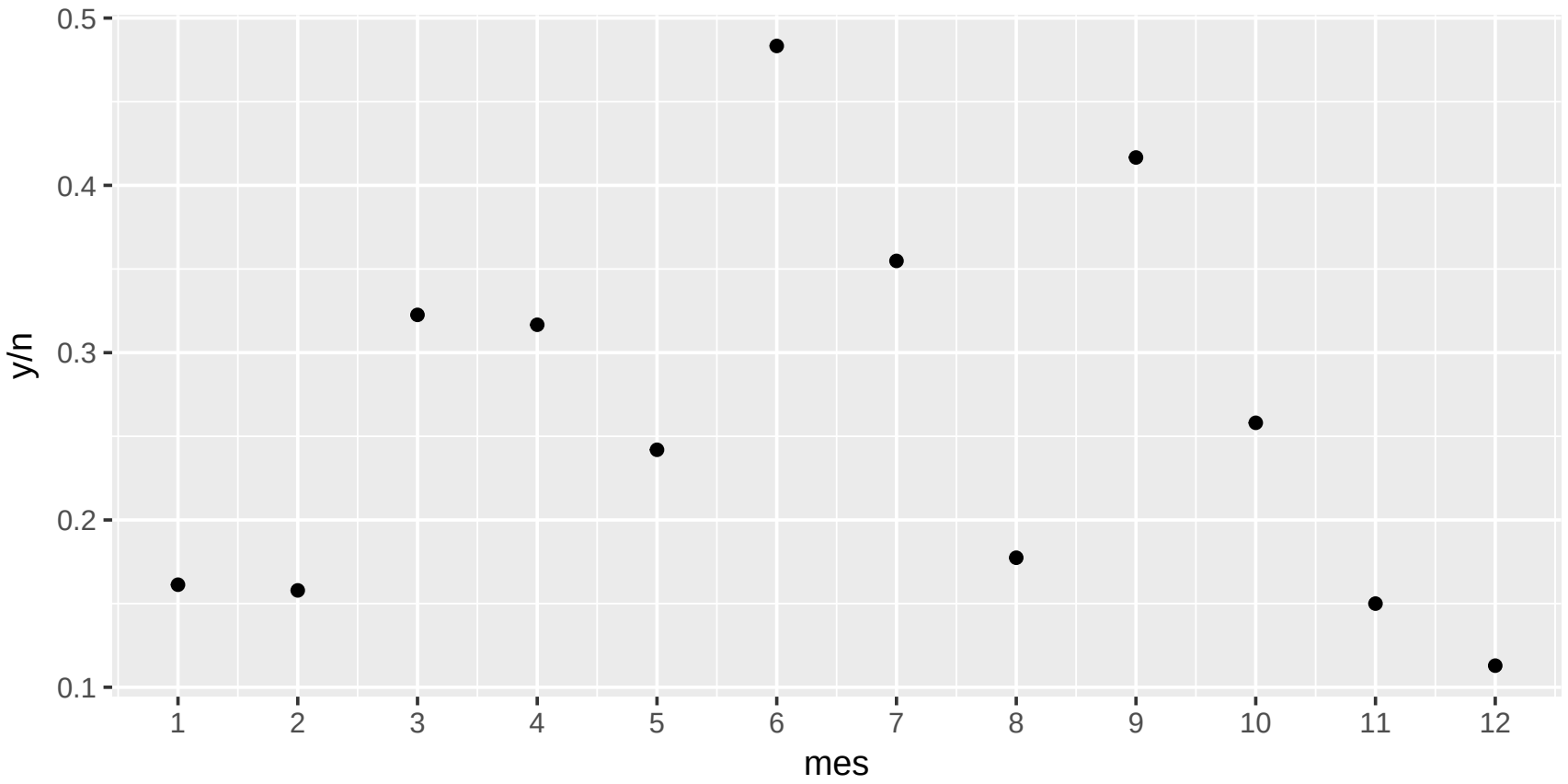


Figura 1: Choveu ou não em cada dia do ano durante dois anos.

Descritiva

Proporção de dias chuvosos por mês

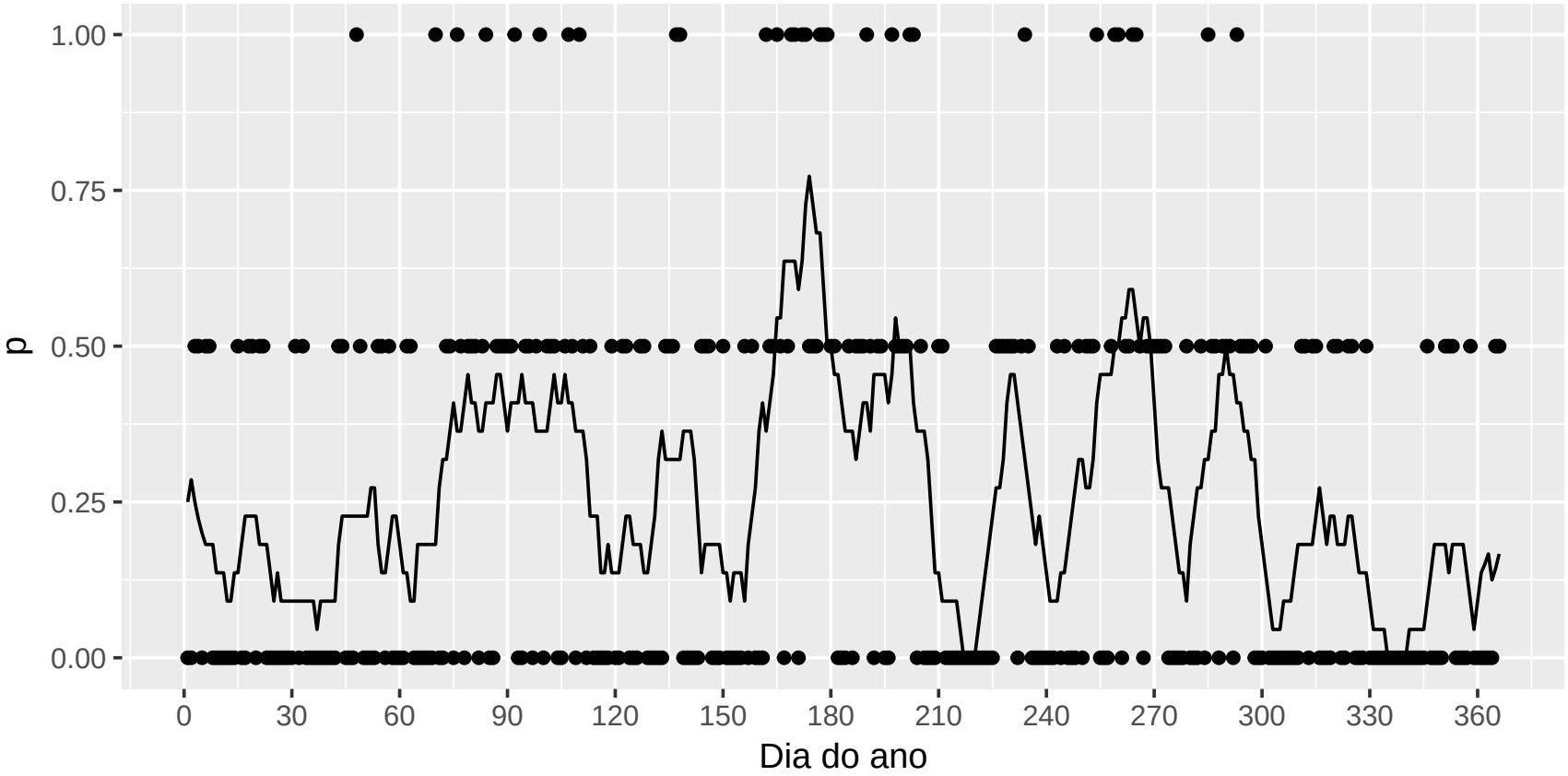


- proporção de dias chuvosos não muda tão abruptamente
- porém, muito da variabilidade foi suprimida.

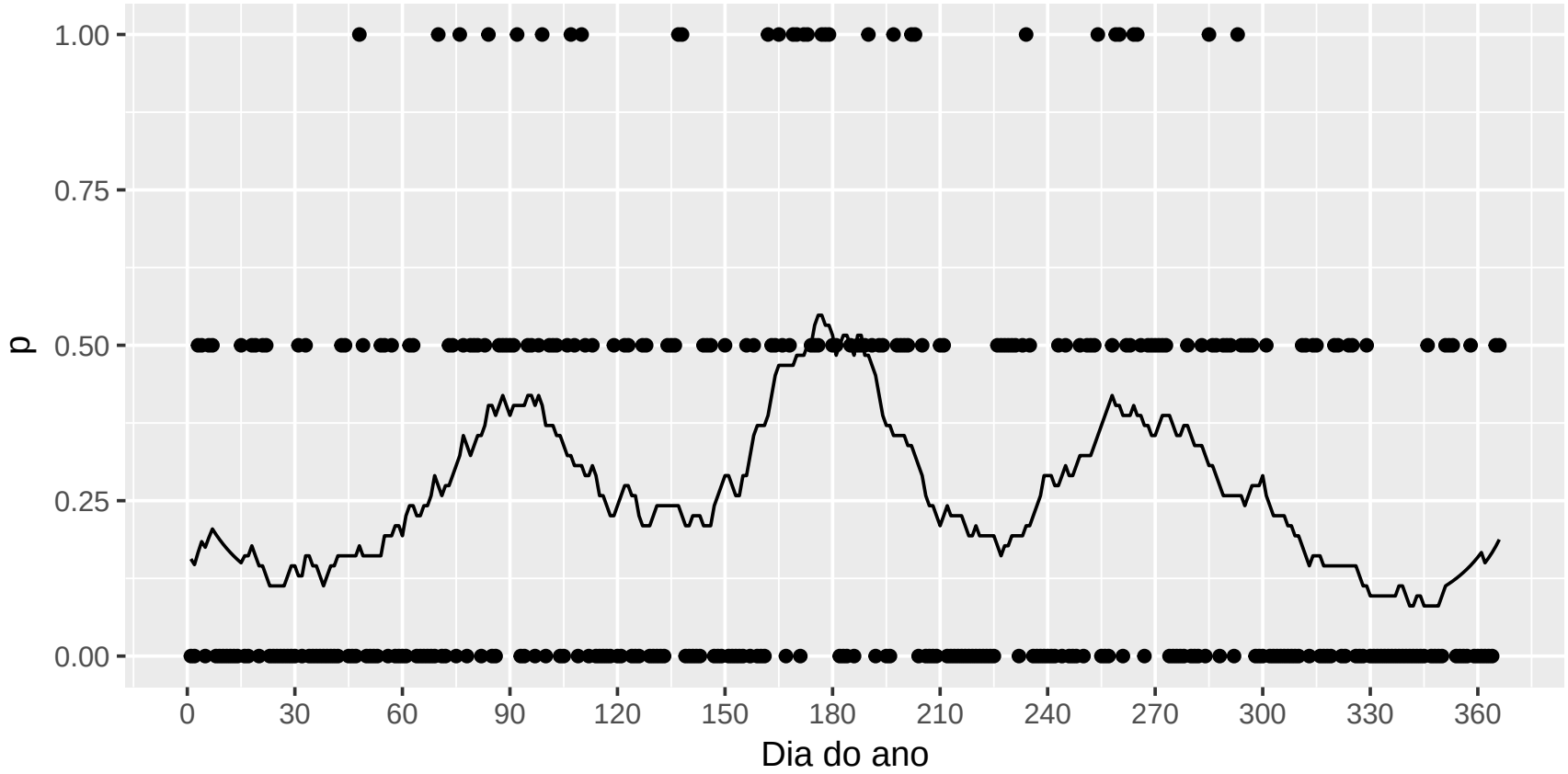
Janela deslizante

- cada dia: média dos dados nos dias mais próximos
- proximidade: à uma distância menor que k dias
- -> janelas deslizantes de amplitude $2 * k$ ao longo do ano

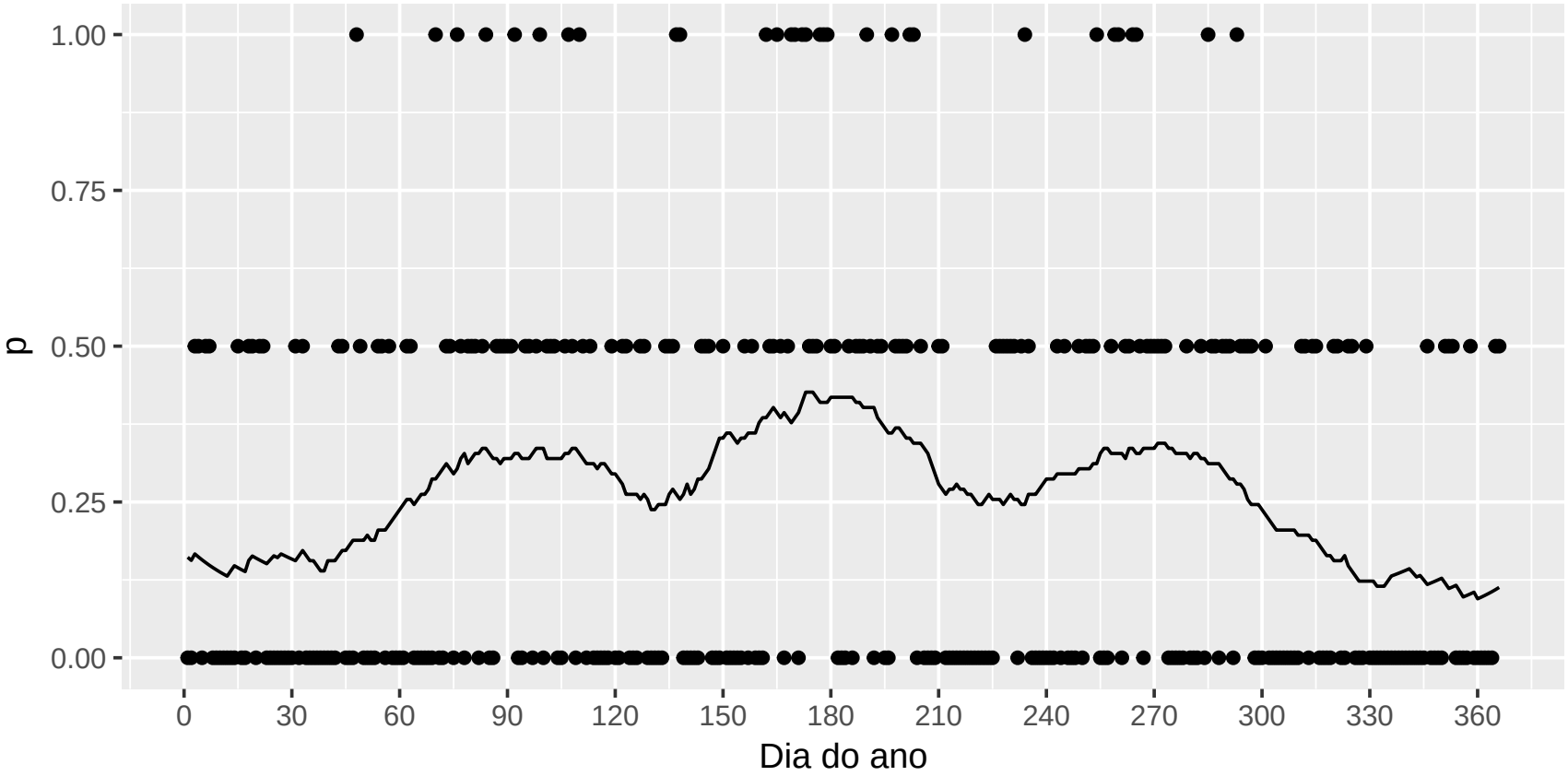
Janela deslizante, k=5



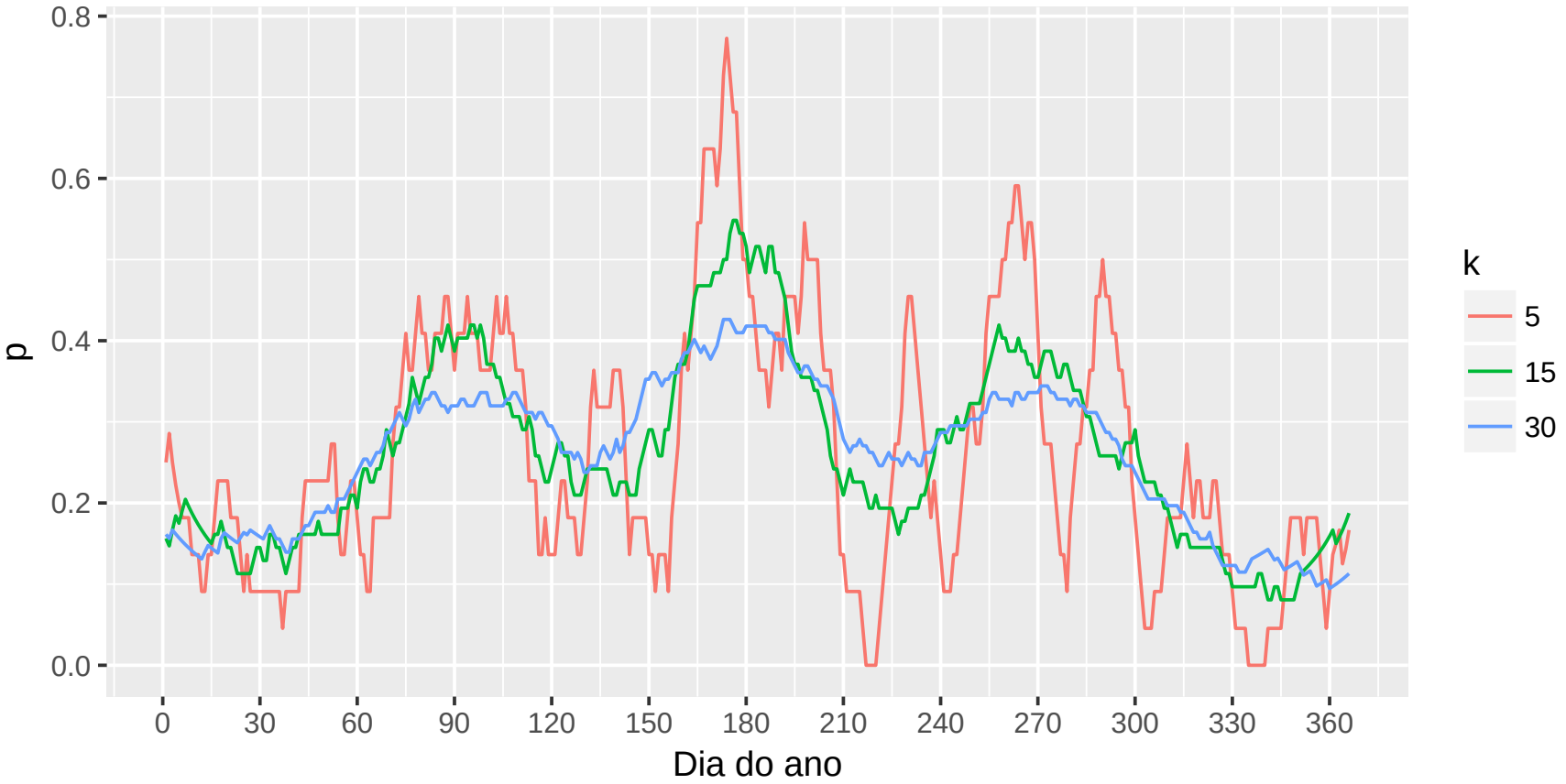
Janela deslizante, k=15



Janela deslizante, k=30



Janela deslizante, k=5,15,30



Modelagem

Modelagem

- **Savage**: “Devemos construir modelos tão grandes quanto elefantes”

Modelagem

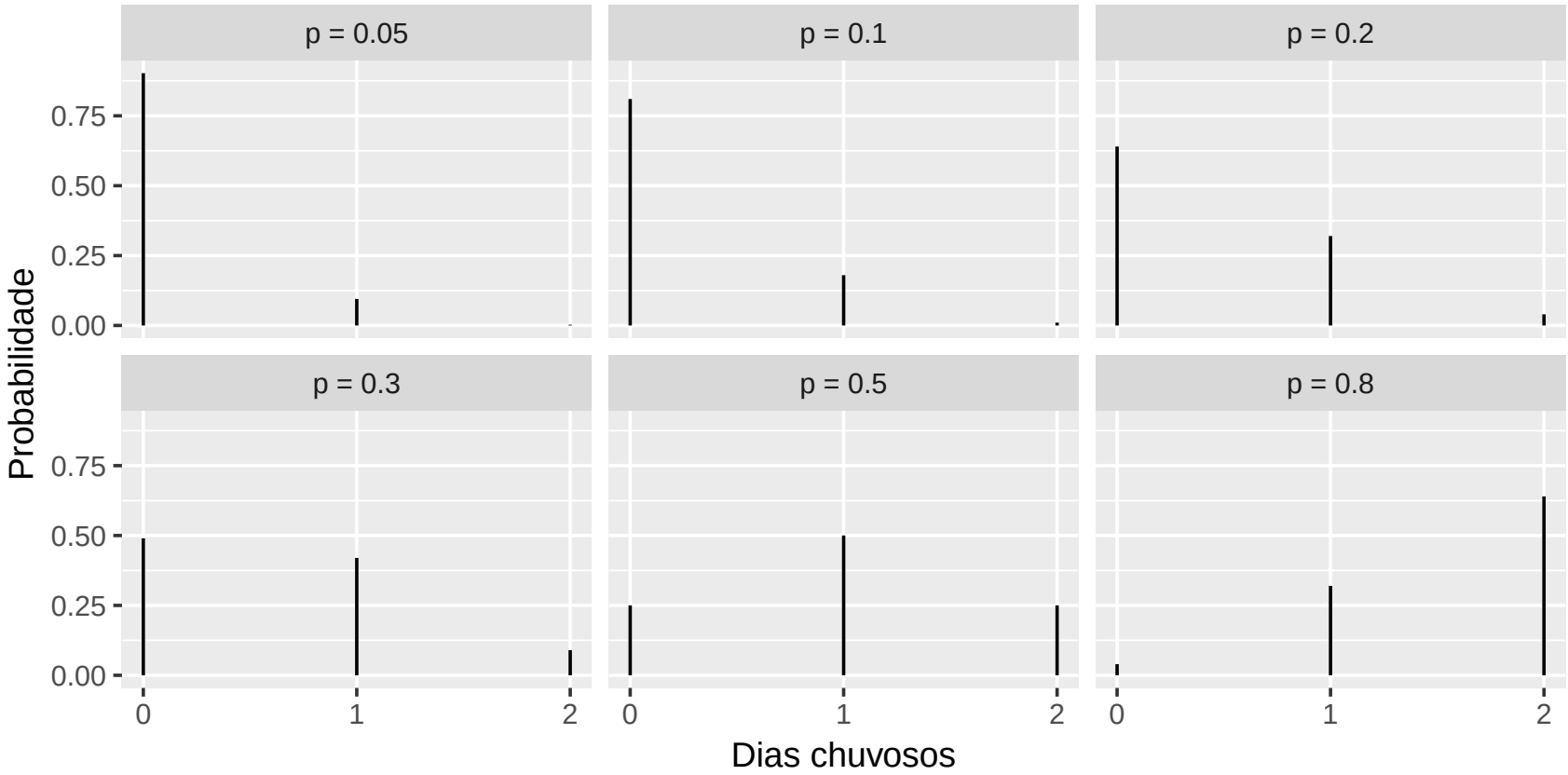
- **Savage**: “Devemos construir modelos tão grandes quanto elefantes”
- **von Neumann**: “Com quatro parâmetros eu posso estimar um elefante, e com cinco eu posso fazê-lo mexer sua tromba”

Variabilidade amostral

- pode ou não chover um dia mesmo que a probabidade de chuva nesse dia não tenha alterado
- variabilidade nos dados mesmo o parâmetro não se alterando
- assumir distribuição probabilística: **P(Dados | Parâmetros)**

Variabilidade amostral

- pode ou não chover um dia mesmo que a probabidade de chuva nesse dia não tenha alterado
- variabilidade nos dados mesmo o parâmetro não se alterando
- assumir distribuição probabilística: **P(Dados | Parâmetros)**



Abordagem

- assumir distribuição probabilística aos dados
  - condicional à variáveis explicativas e parâmetros:
    P(chuva | dia do ano, parâmetros)

Abordagem

- assumir distribuição probabilística aos dados
  - condicional à variáveis explicativas e parâmetros:
    P(chuva | dia do ano, parâmetros)
- **Observáveis**:
  - *Resposta*: chuva/não chuva
  - *Covariável*: dia do ano
- **não observáveis**:
  - parâmetros

Abordagem

- assumir distribuição probabilística aos dados
  - condicional à variáveis explicativas e parâmetros:
    P(chuva | dia do ano, parâmetros)
- **Observáveis**:
  - *Resposta*: chuva/não chuva
  - *Covariável*: dia do ano
- **não observáveis**:
  - parâmetros

- após observar os dados, a incerteza é sobre os parâmetros
- incerteza no modelo

Ocorrência de chuva em Tokyo

Opções iniciais de modelagem: GLM

- modelos lineares generalizados, *Generalized Linear Models* - GLM
- probabilidade modelada: *logito* de η

$$p_i = \frac{1}{1 + \mathrm{e}^{-\eta_i}}$$

- η_i: alguma função do dia do ano/tempo t
- probabilidade, p_i, de chuva em função do tempo, t
- qual função de tempo, t?

Polinômios em t

- polinômio de ordem m

$$\eta_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + ... + \beta_m t_i^m$$

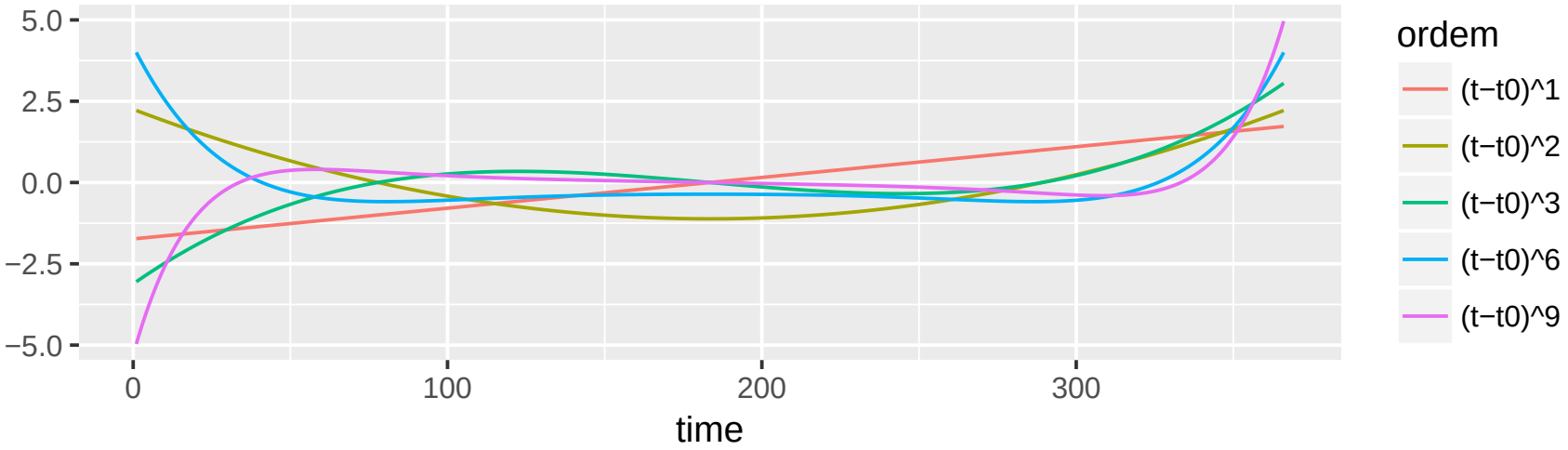
estimar os parâmetros β_j, $j = 1, ..., m$

Polinômios em t

- polinômio de ordem m

$$\eta_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + ... + \beta_m t_i^m$$

estimar os parâmetros β_j, $j = 1, ..., m$

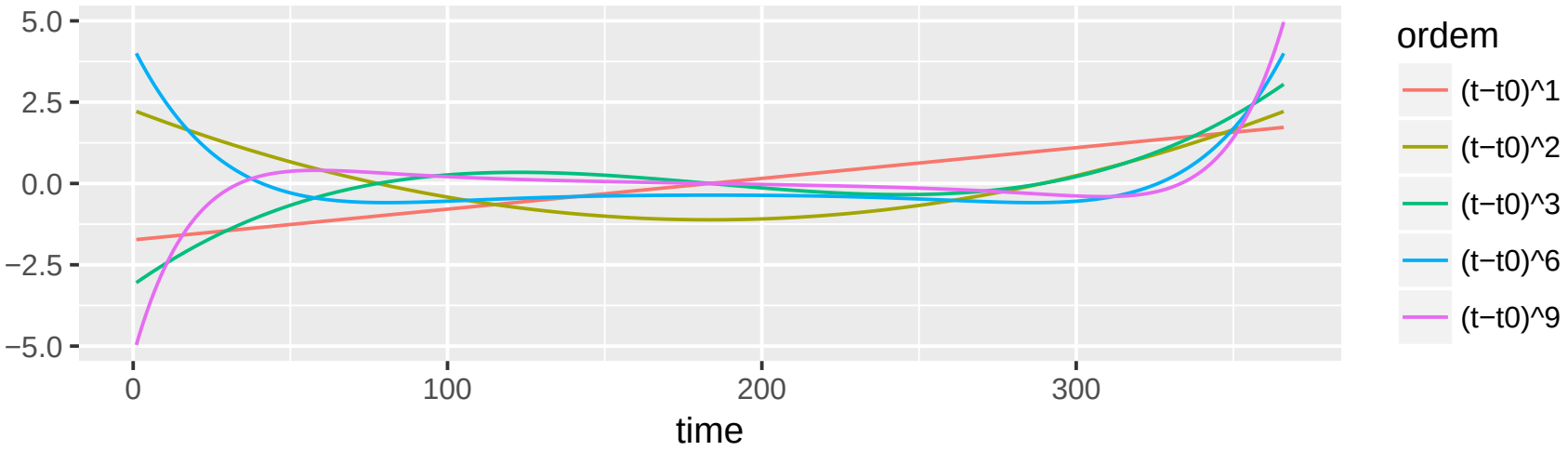


Polinômios em t

- polinômio de ordem m

$$\eta_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + ... + \beta_m t_i^m$$

estimar os parâmetros β_j, $j = 1, ..., m$



- polinômios são flexíveis, mas carecem de interpretabilidade
- estabilidade computacional: polinômios ortogonais

Resultado considerando polinômios

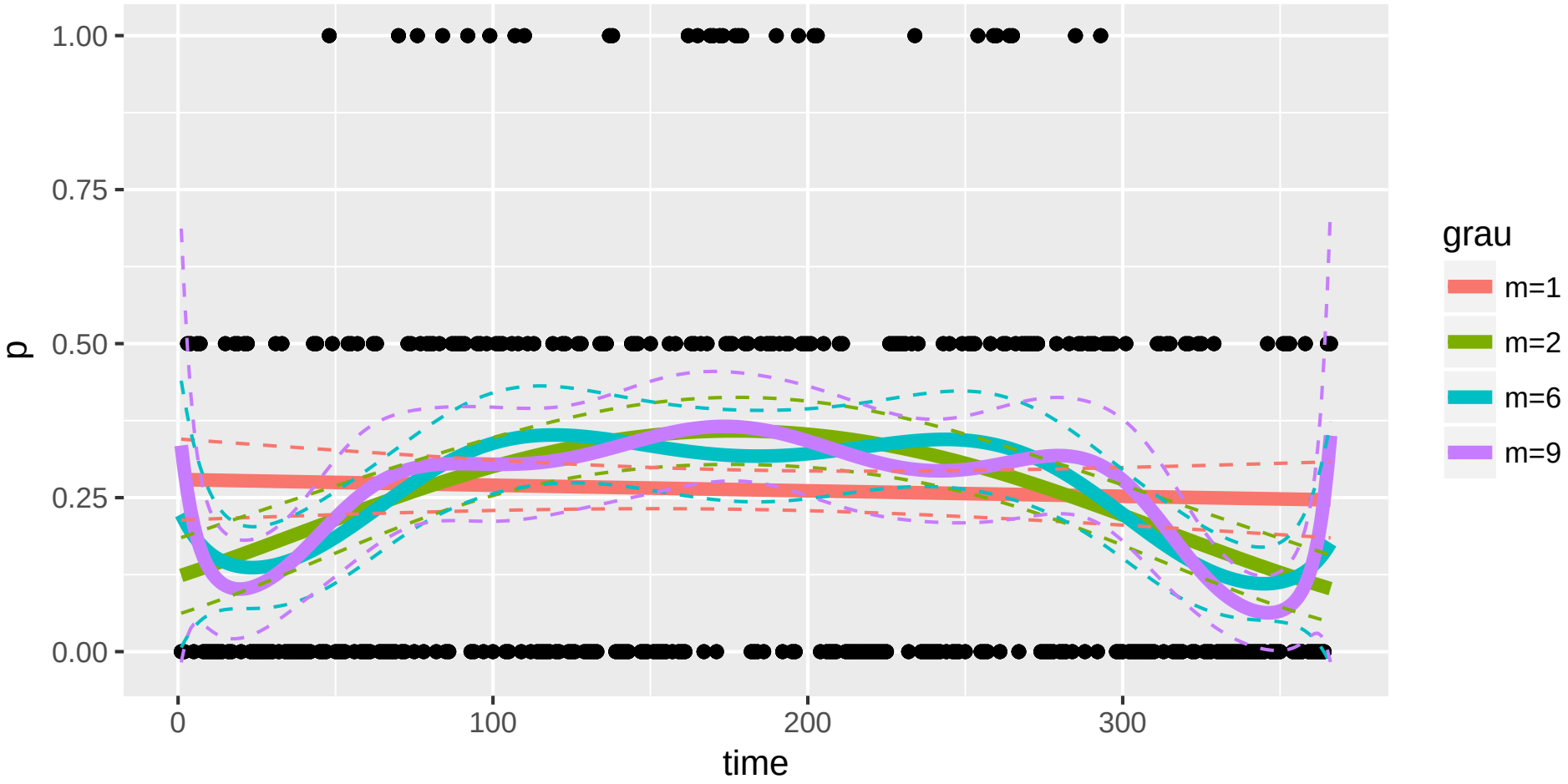


Figura 3: Curvas de predição (+ incerteza), para diferentes graus.

Funções bases

- representar/subdividir o espaço da variável

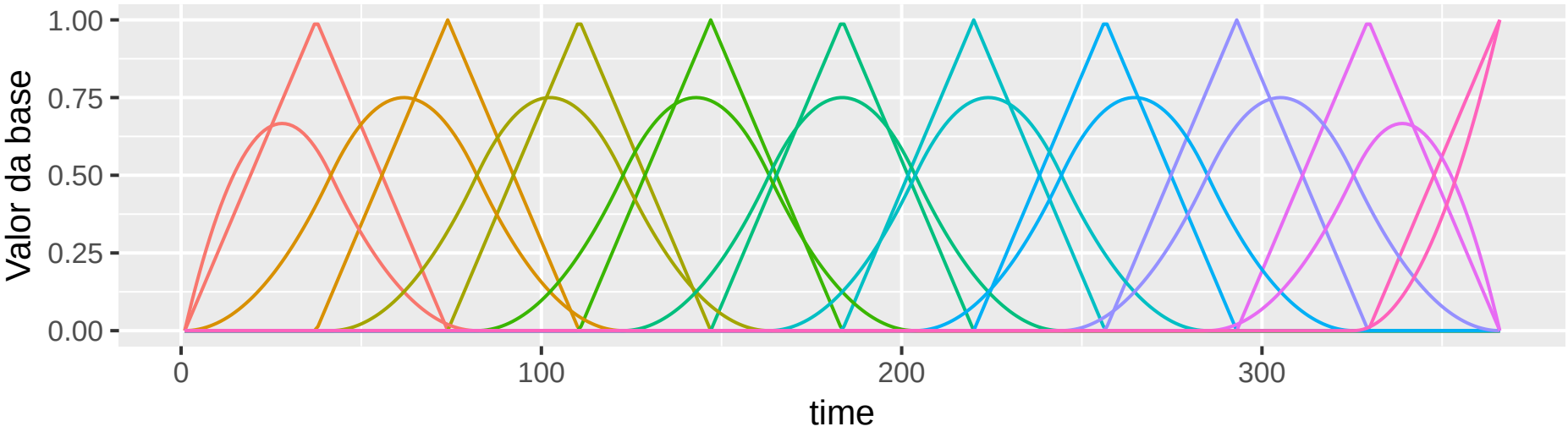


Figura 4: B-splines de primeiro e segundo grau, com 8 graus de liberdade.

Funções bases

- representar/subdividir o espaço da variável

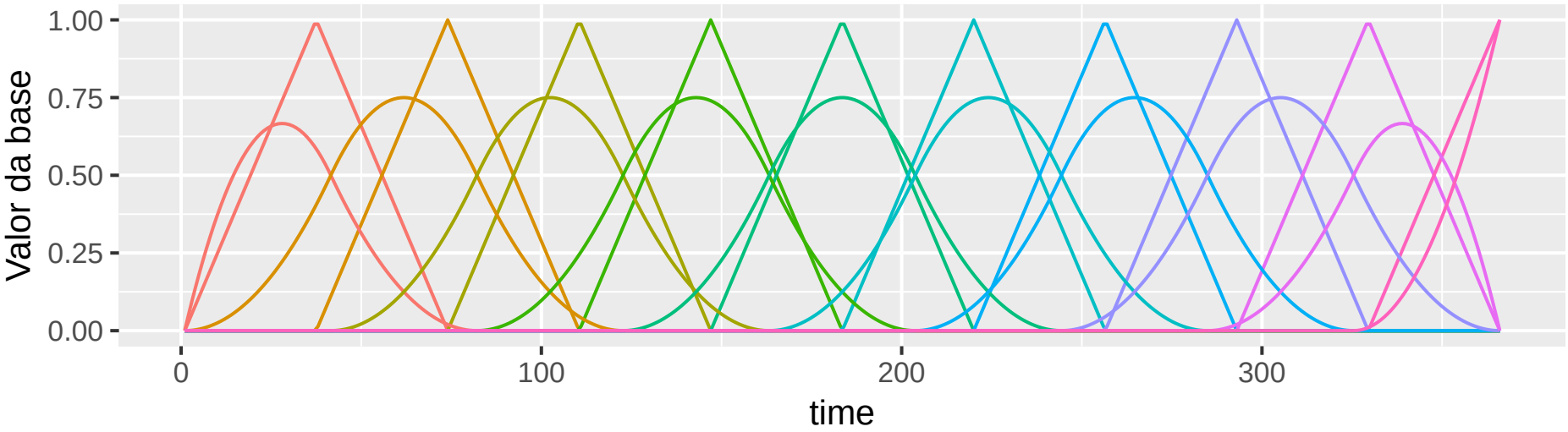


Figura 4: B-splines de primeiro e segundo grau, com 8 graus de liberdade.

- suporte compacto
  - cada função base representa uma parte
  - valores não nulos em parte da variável
  - coeficientes de regressão: ativação naquela parte

Usando B-splines de grau 2

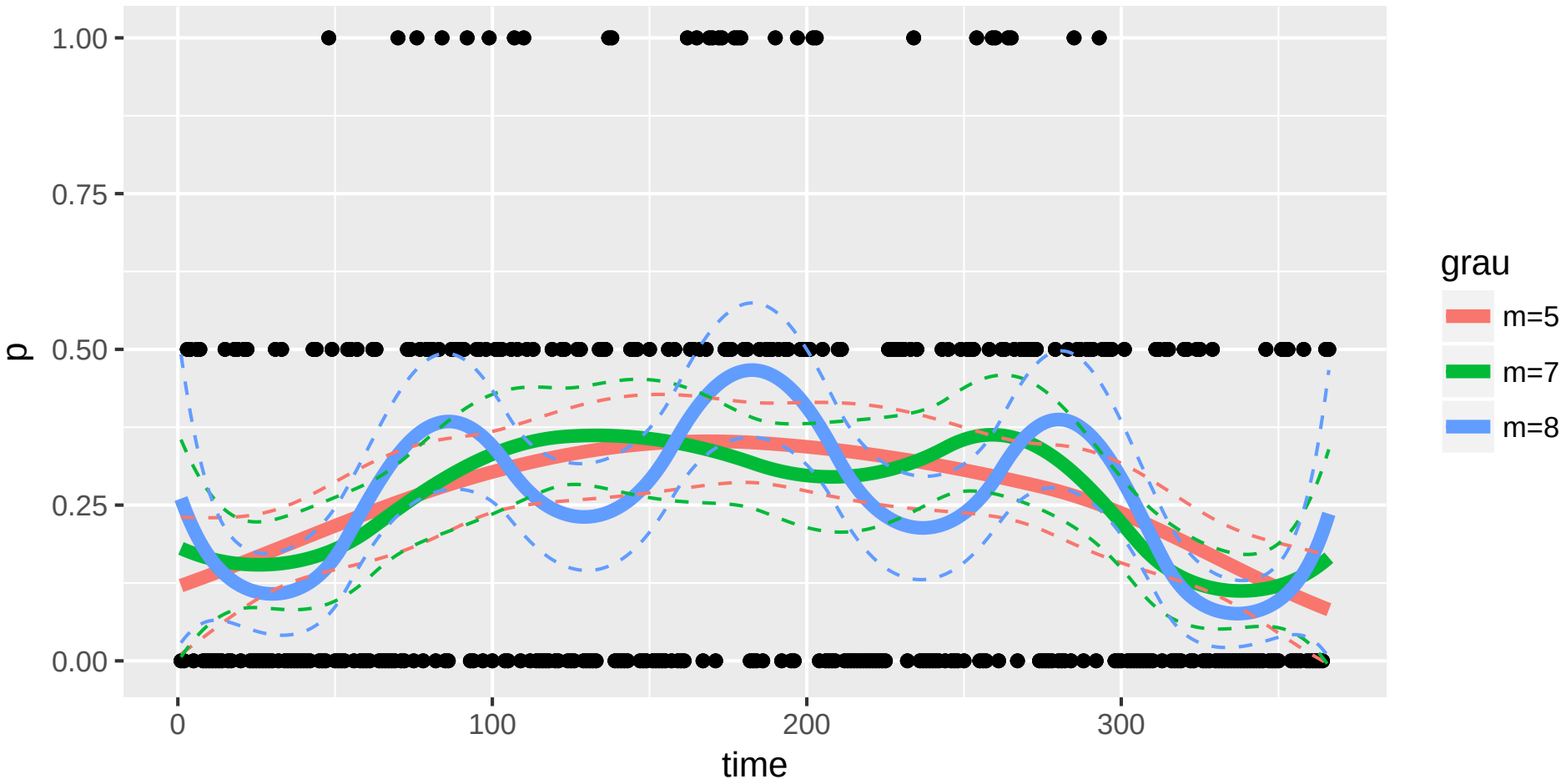


Figura 5: Curvas de predição (e bandas de incerteza), para diferentes graus de liberdade.

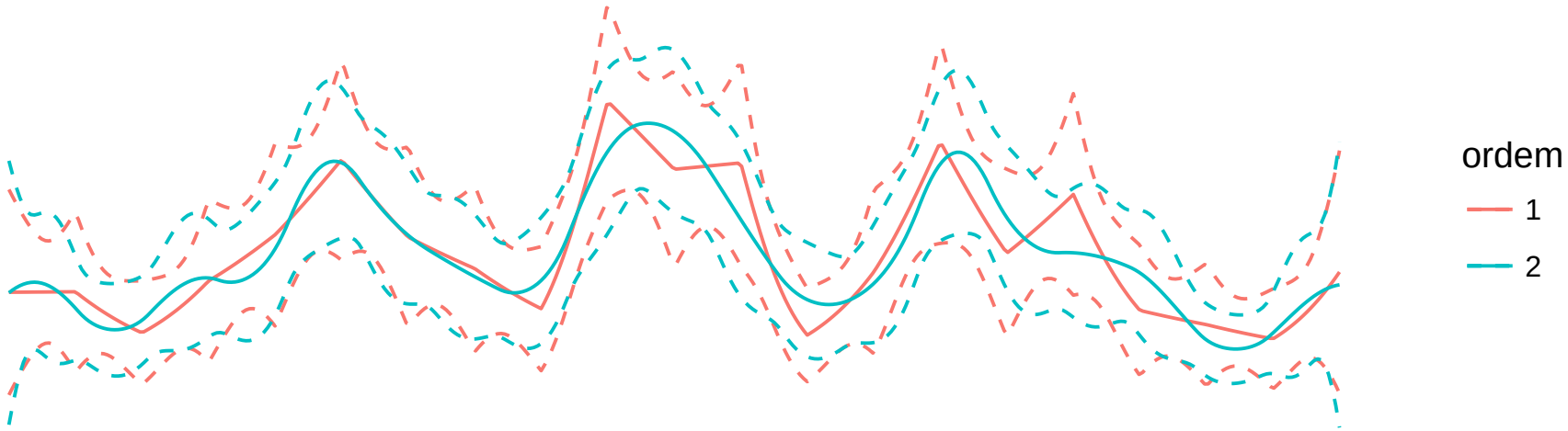
splines

- facilidade maior de interpretação que polinômios
- muitos graus de liberdade: *ajuste excessivo* aos dados
- no limite volta ao caso de análise de cada dia separadamente

Bases e coeficientes

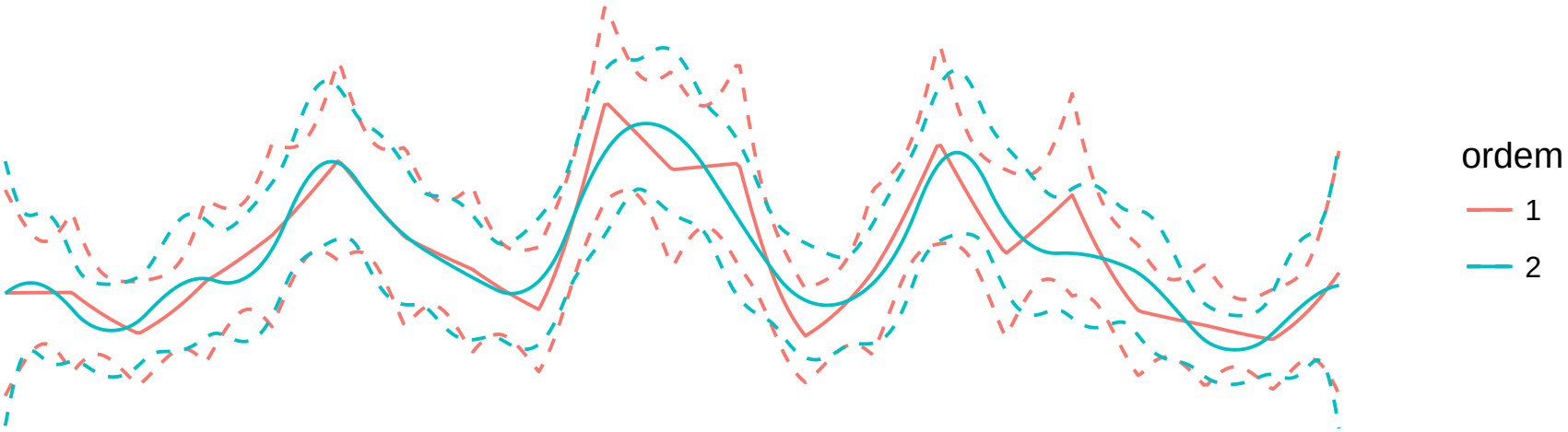
Exemplo: 20 *B-splines* de ordens 1 e 2

- Resultado:

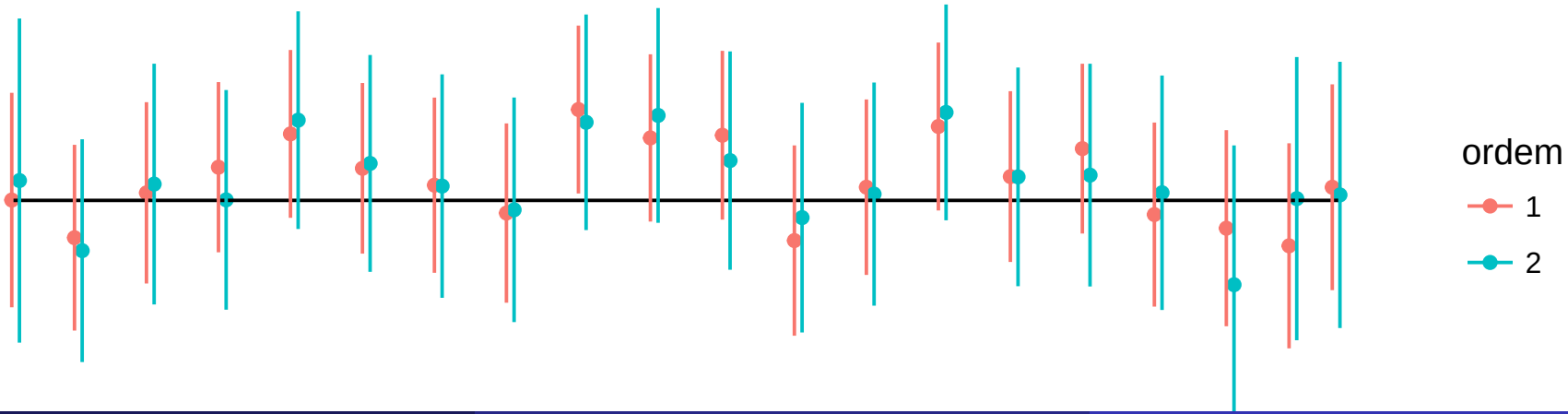


Exemplo: 20 *B-splines* de ordens 1 e 2

- Resultado:

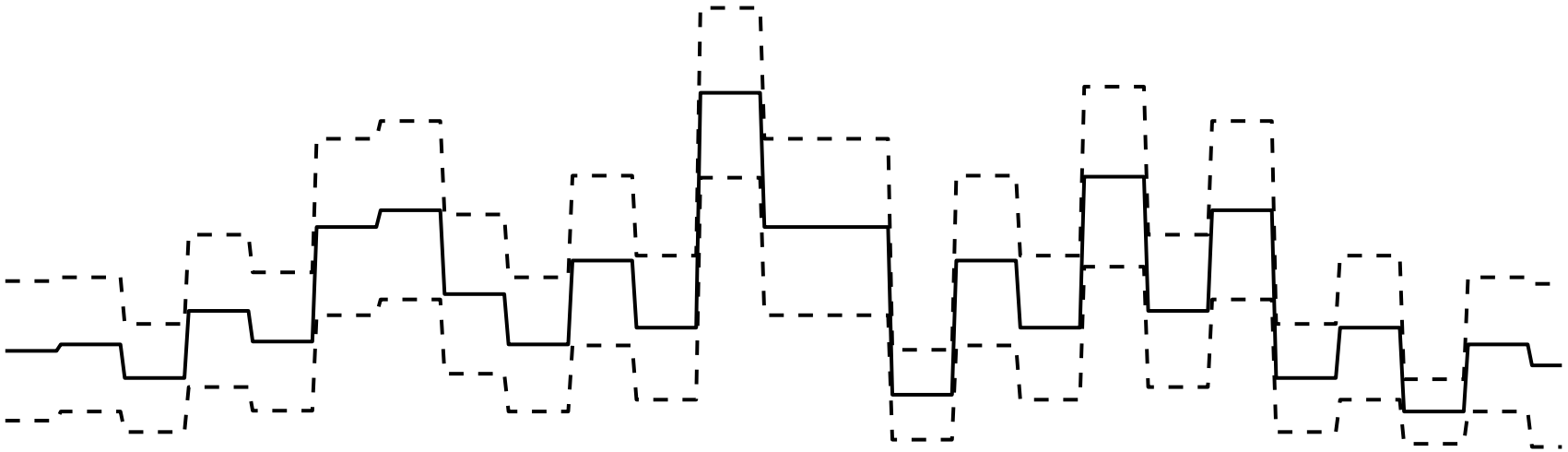


- O que ocorre com os coeficientes:



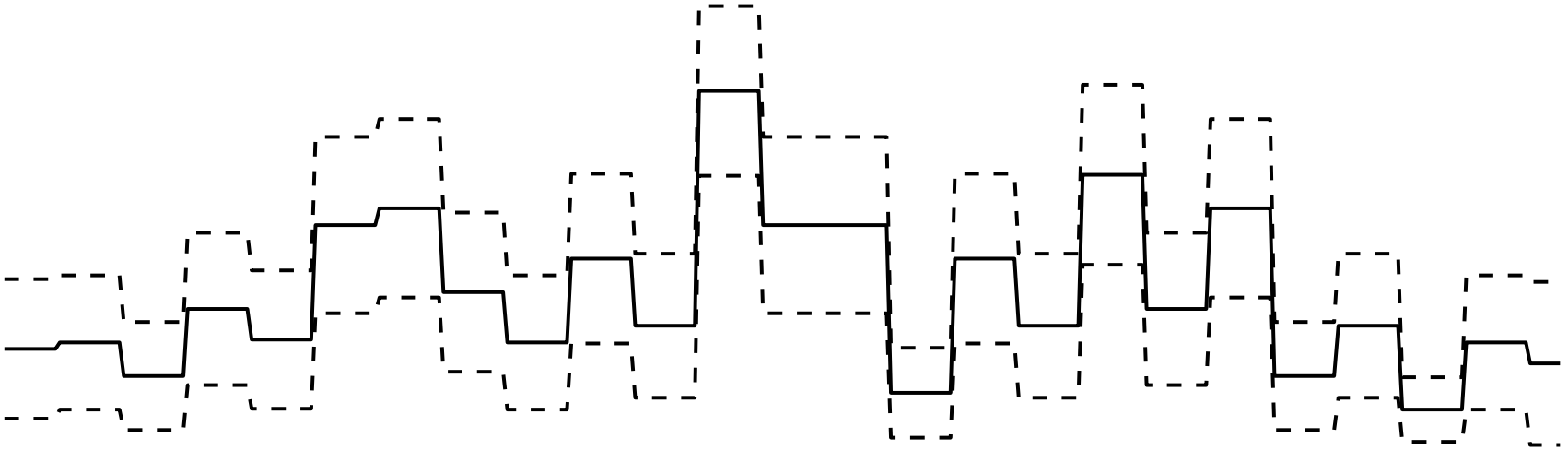
Tempo discretizado (15 dias) como fator

- Resultado:

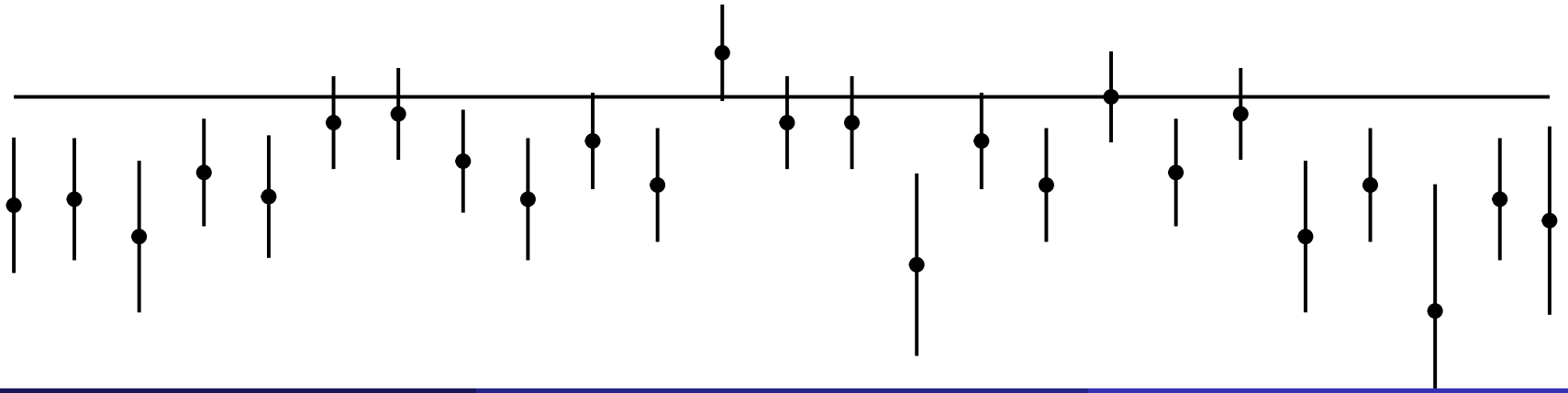


Tempo discretizado (15 dias) como fator

- Resultado:



- O que ocorre com os coeficientes:



Ideia

Ideia: Modelar os coeficientes

- Seja $\underline{\eta}$ escrito como

$$\underline{\eta} = \mathbf{Z}\underline{b}$$

- $\mathbf{Z}$: matriz cujas linhas somam 1
- $\underline{b}$ vetor de coeficientes

Ideia: Modelar os coeficientes

- Seja $\underline{\eta}$ escrito como

$$\underline{\eta} = \mathbf{Z}\underline{b}$$

- $\mathbf{Z}$: matriz cujas linhas somam 1
- $\underline{b}$ vetor de coeficientes
- considerar algum modelo para suavizar $\underline{b}$
- **suavizar**: b_i similar a b_{i+1}

Ideia: Modelar os coeficientes

- Seja $\underline{\eta}$ escrito como

$$\underline{\eta} = \mathbf{Z}\underline{b}$$

- $\mathbf{Z}$: matriz cujas linhas somam 1
- $\underline{b}$ vetor de coeficientes
- considerar algum modelo para suavizar $\underline{b}$
- **suavizar**: b_i similar a b_{i+1}
- Exemplo: tempo discretizado

em k períodos:

$$\mathbf{Z} = \begin{pmatrix} 1 & & & \\ \vdots & & & \\ 1 & & & \\ & 1 & & \\ & \vdots & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & \vdots \\ & & & 1 \end{pmatrix}_{n \times k}$$

Ideia: Modelar os coeficientes

- Seja $\underline{\eta}$ escrito como

$$\underline{\eta} = \mathbf{Z}\underline{b}$$

- $\mathbf{Z}$: matriz cujas linhas somam 1
- $\underline{b}$ vetor de coeficientes
- considerar algum modelo para suavizar $\underline{b}$
- **suavizar**: b_i similar a b_{i+1}
- Exemplo: tempo discretizado

em k períodos:

$$\mathbf{Z} = \begin{pmatrix} 1 & & & \\ \vdots & & & \\ 1 & & & \\ & 1 & & \\ & \vdots & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & \vdots \\ & & & 1 \end{pmatrix}_{n \times k}$$

- se $k = n$ temos $\mathbf{Z} = \mathbf{I}_{n \times n}$

Modelo de passeio aleatório

Modelo para diferenças sucessivas

- modelagem das **diferenças sucessivas**
  - Seja b_1, .., b_n variáveis aleatórias
  - Seja as diferenças sucessivas: $b_i - b_{i+1}$
  - Considere $b_i - b_{i+1} \sim N(0, \sigma^2)$

Modelo para diferenças sucessivas

- modelagem das **diferenças sucessivas**
  - Seja b_1, .., b_n variáveis aleatórias
  - Seja as diferenças sucessivas: $b_i - b_{i+1}$
  - Considere $b_i - b_{i+1} \sim N(0, \sigma^2)$

- σ^2 é o parâmetro desse modelo
- quanto menor σ^2, menor a variação em $b_i - b_{i+1}$
- sem variação $= 0 \leftarrow \sigma^2 \rightarrow \infty =$ sem suavização

Distribuição conjunta

- distribuição conjunta do vetor $\underline{b} = b_1, ..., b_n$

$$
\begin{aligned}
p(\underline{b}) &= \prod_{i=1}^{n-1} p(b_i | b_{i+1}) \\
&= \prod_{i=1}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(b_i - b_{i+1})^2}{2\sigma^2}}
\end{aligned}
$$

Distribuição conjunta

- distribuição conjunta do vetor $\underline{b} = b_1, ..., b_n$

$$
\begin{aligned}
p(\underline{b}) &= \prod_{i=1}^{n-1} p(b_i|b_{i+1}) \\
&= \prod_{i=1}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(b_i - b_{i+1})^2}{2\sigma^2}} \\
&= (2\pi\sigma^2)^{(n-1)/2} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n-1}(b_i - b_{i+1})^2}
\end{aligned}
$$

Distribuição conjunta (cont.)

- Notar que

$$\begin{aligned}\sum_{i=1}^{n-1}(b_i - b_{i+1})^2 &= \sum_{i=1}^{n-1}(b_i^2 - 2b_i b_{i+1} + b_{i+1}^2) \\ &= \sum_{i=1}^{n-1} b_i^2 - 2\sum_{i=1}^{n-1} b_i b_{i+1} + \sum_{i=2}^{n} b_i^2\end{aligned}$$

Distribuição conjunta (cont.)

- Notar que

$$\begin{aligned}\sum_{i=1}^{n-1}(b_i - b_{i+1})^2 &= \sum_{i=1}^{n-1}(b_i^2 - 2b_i b_{i+1} + b_{i+1}^2) \\ &= \sum_{i=1}^{n-1} b_i^2 - 2\sum_{i=1}^{n-1} b_i b_{i+1} + \sum_{i=2}^{n} b_i^2 = \underline{b}^T \mathbf{R} \underline{b}\end{aligned}$$

onde

$$\mathbf{R} = \begin{pmatrix} 1 & -1 & & & & & & & \\ -1 & 2 & -1 & & & & & & \\ & -1 & 2 & -1 & & & & & \\ & & & \ddots & & & & & \\ & & & & -1 & 2 & -1 & & \\ & & & & & -1 & 2 & -1 & \\ & & & & & & & 1 & \end{pmatrix}$$

Distribuição conjunta (cont.)

- Notar que

$$\begin{aligned}\sum_{i=1}^{n-1}(b_i - b_{i+1})^2 &= \sum_{i=1}^{n-1}(b_i^2 - 2b_i b_{i+1} + b_{i+1}^2) \\ &= \sum_{i=1}^{n-1} b_i^2 - 2\sum_{i=1}^{n-1} b_i b_{i+1} + \sum_{i=2}^{n} b_i^2 = \underline{b}^T \mathbf{R} \underline{b}\end{aligned}$$

onde

$$\mathbf{R} = \begin{pmatrix} 1 & -1 & & & & & & \\ -1 & 2 & -1 & & & & & \\ & -1 & 2 & -1 & & & & \\ & & & \ddots & & & & \\ & & & & -1 & 2 & -1 & \\ & & & & & -1 & 2 & -1 \\ & & & & & & & 1 \end{pmatrix}$$

- $p(\underline{b}) = (2\pi\sigma^2)^{(n-1)/2} \mathrm{e}^{-\frac{1}{2\sigma^2}\underline{b}^T \mathbf{R} \underline{b}}$

Distribuição de b para o exemplo de Tóquio

- é razoável considerar b_1 similar a b_n
- passeio aleatório cíclico
- multiplicando a distribuição conjunta anterior por

$$(2\pi\sigma^2)^{1/2}\mathrm{e}^{-\frac{(b_n-b_1)^2}{2\sigma^2}}$$

Distribuição de b para o exemplo de Tóquio

- é razoável considerar b_1 similar a b_n
- passeio aleatório cíclico
- multiplicando a distribuição conjunta anterior por

$$(2\pi\sigma^2)^{1/2}\mathrm{e}^{-\frac{(b_n-b_1)^2}{2\sigma^2}}$$

temos

$$p(\underline{b}_c) = (2\pi\sigma^2)^{n/2}\mathrm{e}^{-\frac{1}{2\sigma^2}\underline{b}_c^T\mathbf{R}_c\underline{b}_c}$$

onde

$$\mathbf{R}_c = \begin{pmatrix} 2 & -1 & & & & & & -1 \\ -1 & 2 & -1 & & & & & \\ & -1 & 2 & -1 & & & & \\ & & & \ddots & & & & \\ & & & & -1 & 2 & -1 & \\ & & & & & -1 & 2 & -1 \\ -1 & & & & & & -1 & 2 \end{pmatrix}$$