# Filtros Bayesianos

## Princípios de Inferência Estatística

Maicon C. Germano

IBB UNESP

24 de maio de 2023

# Sumário

# O que é inferência Bayesiana?

- A inferência Bayesiana é um método estatístico que leva em conta informações prévias sobre o que está sendo analisado.
- Ela combina essas informações com os novos dados que estão sendo analisados para chegar a uma conclusão mais precisa e confiável.
- O parâmetro populacional $\theta$ é tratado como uma quantidade aleatória, o que leva a abordagens substancialmente diferentes à modelagem e inferência em comparação à inferência clássica.

# O que é inferência Bayesiana?

- A inferência Bayesiana é baseada em $f(\theta \mid y)$ ao invés de se basear em $f(y \mid \theta)$, ou seja, na probabilidade do parâmetro condicional aos dados obtidos.
- Para utilizar a inferência Bayesiana, é necessário especificar uma distribuição a priori de probabilidades $f(\theta)$ que representa as crenças sobre a distribuição de $\theta$ antes de se considerar qualquer informação proveniente dos dados.
- A noção de distribuição a priori para o parâmetro $\theta$ está no cerne do pensamento bayesiano.

# Inferência Bayesiana vs. Clássica

- Existem diferentes maneiras de analisar dados e chegar a conclusões estatísticas. Duas dessas maneiras são a inferência Bayesiana e a inferência clássica.

- A inferência Bayesiana é um método que permite que se leve em conta informações prévias sobre o que está sendo analisado, como experiências anteriores ou conhecimentos sobre o assunto. Ela combina essas informações com os novos dados que estão sendo analisados, para chegar a uma conclusão mais precisa e confiável.

# Inferência Bayesiana vs. Clássica

- Já a inferência clássica é uma maneira mais tradicional de analisar dados, sem levar em conta informações prévias. Nesse método, são usados estimadores de máxima verossimilhança, que buscam encontrar a estimativa mais provável dos parâmetros do modelo que está sendo analisado.

- A principal diferença entre esses dois métodos é que a inferência Bayesiana leva em conta informações prévias, enquanto a inferência clássica não o faz. Isso pode fazer com que a inferência Bayesiana seja mais precisa e confiável em algumas situações.

# Distribuição a priori

- Ao tentar estimar um parâmetro $\theta$, é comum ter algum conhecimento ou crença prévia sobre o valor de $\theta$ antes de considerar os dados.
- A distribuição a priori representa a crença prévia sobre o valor de $\theta$, antes de observar os dados.
- A informação da distribuição a priori é combinada com a informação dos dados para obter a distribuição a posteriori, que reflete a crença atualizada sobre o valor de $\theta$.

# Distribuição a Priori

- A distribuição a priori pode ser baseada em conhecimento prévio, experiências anteriores, especialistas ou modelos teóricos.
- A distribuição a priori pode influenciar as inferências feitas a partir dos dados, pois pode levar a diferentes distribuições a posteriori e conclusões diferentes.
- A escolha da distribuição a priori pode ser subjetiva e pode levar a diferentes inferências entre diferentes indivíduos ou grupos.

Em sua forma básica, o Teorema de Bayes é simplesmente um resultado de probabilidade condicional:

**Teorema de Bayes**

### Teorema (1)

Se A e B são dois eventos com $P(A) > 0$ então:

$$P(B \mid A) = \frac{P(A|B)P(B)}{P(A)}$$

**Exemplo 1.** Um procedimento de testes de diagnóstico para HIV é aplicado a uma população de alto risco; acredita-se que 10% desta população é positiva para o HIV. O teste de diagnóstico é positivo para 90% das pessoas que de fato são HIV-positivas, e negativo para 85% das pessoas que não são HIV-positivas. Qual a probabilidade de resultados falso-positivo e falso-negativo?

**Notação:**

$A$: a pessoa é HIV-positiva, $B$: o resultado do teste é positivo. Temos dados que $P(A) = 0,1$, $P(B \mid A) = 0,9$ e $P(\bar{B} \mid \bar{A}) = 0,85$. Então:

$$P(\text{falso positivo}) = P(\bar{A} \mid B) = \frac{P(B \mid \bar{A})P(\bar{A})}{P(B)}$$

$$= \frac{0,15 \times 0,9}{(0,15 \times 0,9) + (0,9 \times 0,1)} = 0,6$$

# Continuação do exemplo 1

De forma similar,

$$
\begin{aligned}
P(\text{falso negativo}) &= P(A \mid \bar{B}) \\
&= \frac{P(\bar{B} \mid A)P(A)}{P(\bar{B})} \\
&= \frac{0,1 \times 0,1}{(0,1 \times 0,1) + (0,85 \times 0,9)} = 0,0129
\end{aligned}
$$

### Intervalos de credibilidade

- A ideia do intervalo de credibilidade é criar um análogo ao intervalo de confiança da estatística frequentista.
- Estimativas pontuais não fornecem uma medida de precisão, portanto, é preferível informar um intervalo no resultado das análises.
- Na estatística frequentista, isso causa problemas, pois os parâmetros não são considerados aleatórios, tornando impossível dar um intervalo com a interpretação de que há uma certa probabilidade de o parâmetro estar dentro do intervalo.
- Em vez disso, os intervalos de confiança devem ser interpretados como se a amostragem fosse repetida inúmeras vezes, haveria uma probabilidade especificada de que o intervalo contivesse o parâmetro - ou seja, o intervalo é aleatório e não o parâmetro.

# Intervalos de credibilidade

Na abordagem bayesiana, os parâmetros são tratados como aleatórios. Assim, uma região $C_\alpha(y)$ é uma região de credibilidade $100(1-\alpha)\%$ para $\theta$ se

$$\int_{C_\alpha(y)} f(\theta \mid y)\mathrm{d}\theta = 1 - \alpha. \tag{1}$$

Ou seja, existe uma probabilidade de $1-\alpha$, com base na distribuição posteriori, que $\theta$ esteja contido em $C_\alpha(y)$.

# Intervalos de credibilidade

- Uma dificuldade com o intervalo de credibilidade (que também ocorre com intervalos de confiança), é que eles não são unicamente definidos.
- Qualquer região com probabilidade $1 - \alpha$ é um intervalo válido.
- Deseja-se um intervalo que contenha apenas os valores mais plausíveis do parâmetro, por isso é habitual impor uma restrição adicional, que a largura do intervalo seja tão pequena quanto possível.
- Isso equivale a um intervalo (ou região) da forma: $C_\alpha(y) = \{\theta : f(\theta \mid y) \geq \gamma\}$

Em que $\gamma$ é escolhido para garantir que:

$$\int_{C_\alpha(y)} f(\theta \mid y)\mathrm{d}\theta = 1 - \alpha$$

Tais regiões são chamadas de regiões com mais alta densidade posteriori (HPD). Esses intervalos são encontrados numericamente, embora existam valores tabulados para algumas distribuições posteriores univariadas. É importante lembrar que a escolha do valor de $\alpha$ pode afetar a largura do intervalo, resultando em um "perde/ganha": um valor pequeno resulta em um intervalo largo, enquanto um valor grande resulta em um intervalo com baixa probabilidade de conter o parâmetro.

**Exemplo 2.** (Média normal.) Sejam $Y_1, \ldots, Y_n$ variáveis independentes de uma distribuição $N\left(\theta, \sigma^2\right)$, ( $\sigma^2$ conhecido) com uma priori para $\theta$ da forma $\theta \sim N\left(b, d^2\right)$.

**Resolução:**
Com essas informações, obtém-se a posteriori:

$$\theta \mid y \sim N\left(\frac{\frac{b}{d^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{d^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{d^2} + \frac{n}{\sigma^2}}\right) \tag{2}$$

# Como obter a posteriori

A essência da abordagem bayesiana é tratar o parâmetro desconhecido $\theta$ como uma variável aleatória, especificar uma distribuição a priori para $\theta$ que represente as convicções sobre $\theta$ antes de ver os dados, usar o Teorema de Bayes para atualizar as convicções a priori na forma de probabilidades a posteriori e fazer inferências apropriadas. Portanto há quatro passos característicos da abordagem bayesiana:

1. especificação da verossimilhança do modelo $L(\theta \mid y) \equiv f(y \mid \theta)$;
2. determinação da priori $f(\theta)$;
3. cálculo da distribuição posteriori $f(\theta \mid y)$,obtida pelo Teorema de Bayes;
4. extrair inferências da distribuição posteriori.

# Como obter a posteriori

O Teorema da Bayes expresso em termos de variáveis aleatórias, com densidades denotadas genericamente por $f(.)$, tem a forma:

$$f(\theta \mid y) = \frac{f(\theta)f(y \mid \theta)}{f(y)} = \frac{f(\theta)f(y \mid \theta)}{\int f(\theta)f(y \mid \theta)d\theta} \tag{3}$$

No caso contínuo $f$ é a função de densidade de probabilidade como usual, mas no caso discreto, $f$ é a função de massa de probabilidade de $Y$ $[P(Y = y)]$. De forma similar, $\theta$ pode ser discreto ou contínuo mas, caso seja discreto, $\int f(\theta)f(y \mid \theta)d\theta$ deve ser interpretado como $\sum_j f(\theta_j) f(y \mid \theta_j)$.
Pode-se então dizer que:

$$posteriori = \frac{priori \times verossimilhana}{marginal} \tag{4}$$

# Como obter a posteriori

Note que o denominador no Teorema de Bayes é uma função apenas de $y$, resultante de uma integração em $\theta$ (ou seja, $\theta$ foi "integrado fora"). Desta forma, uma outra maneira de escrever o Teorema de Bayes é:

$$f(\theta \mid y) \propto f(\theta)f(y \mid \theta) \tag{5}$$

ou, em palavras, "a posteriori é proporcional à priori vezes a verossimilhança".

**Voltando ao nosso exemplo:** (Média normal.) Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes com uma distribuição $N\left(\theta, \sigma^2\right)$, em que $\sigma^2$ é conhecido.
Então,

$$f\left(y_i \mid \theta\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{\left(y_i - \theta\right)^2}{2\sigma^2}\right\} \tag{6}$$

e a verossimilhança fica

$$L(\theta \mid y) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^{n}\left(y_i - \theta\right)^2}{2\sigma^2}\right\} \tag{7}$$

$$\propto \exp\left\{-\frac{\sum_{i=1}^{n}\left(y_i - \theta\right)^2}{2\sigma^2}\right\}$$

# Como obter a posteriori

Após algum algebrismo é possível mostrar que esta expressão pode ser escrita na forma

$$L(\theta \mid y) \propto \exp\left\{-\frac{(\theta-\bar{y})^2}{2\left(\sigma^2/n\right)}\right\} \tag{8}$$

Agora, suponha que as convicções a priori sobre $\theta$ podem elas mesmas serem representadas por uma distribuição normal $\theta \sim \mathrm{N}\left(b, d^2\right)$.

Novamente, esta escolha visa obter uma análise matemática simples, mas deve apenas ser usada se tal escolha é de fato uma boa aproximação à crença a priori sobre $\theta$. Então, pelo Teorema de Bayes,

$$\begin{aligned} f(\theta \mid y) &\propto \exp\left\{-\frac{(\theta-b)^2}{2d^2}\right\}\exp\left\{-\frac{(\theta-\bar{y})^2}{2\left(\sigma^2/n\right)}\right\} \\ &= \exp\left\{-\frac{1}{2}\left[\frac{(\theta-b)^2}{d^2}+\frac{(\theta-\bar{y})^2}{\sigma^2/n}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{d^2}+\frac{1}{\sigma^2/n}\right)\theta^2-2\theta\left(\frac{b}{d^2}+\frac{\bar{y}}{\sigma^2/n}\right)\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{d^2}+\frac{1}{\sigma^2/n}\right)\left[\left(\theta-\frac{\frac{b}{d^2}+\frac{\bar{y}}{\sigma^2/n}}{\frac{1}{d^2}+\frac{1}{\sigma^2/n}}\right)^2\right]\right\} \end{aligned}$$

# Como obter a posteriori

Portanto,

$$\theta \mid y \sim \mathrm{N}\left(\frac{\frac{b}{d^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{d^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{d^2}+\frac{n}{\sigma^2}}\right) \tag{9}$$

que é igual a equação 2 do nosso exemplo.

# Continuação do exemplo

Agora, como a distribuição Normal é unimodal e simétrica, a região HPD de $100(1-\alpha)\%$ para $\theta$ é:

$$\left(\frac{\frac{b}{d^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{d^2}+\frac{n}{\sigma^2}}\right) \pm z_{\alpha/2}\left(\frac{1}{\frac{1}{d^2}+\frac{n}{\sigma^2}}\right)^{\frac{1}{2}}$$

em que $z_{\alpha/2}$ é o ponto com porcentagem desejada da distribuição normal padrão $N(0,1)$. Note, além do mais que à medida que $d \to \infty$ o intervalo se torna:

$$\bar{y} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

que é precisamente o intervalo de confiança para $100(1-\alpha)\%$ de $\theta$ obtido em inferência clássica. Neste caso especial, o intervalo de credibilidadea e o intervalo de confiança são idênticos, embora as suas interpretações sejam bastante diferentes.

ESTIMATIVA DE ESTADO E DE PARÂMETROS EM MODELAGEM DO TRATAMENTO DE TUMOR PROSTÁTICO VIA RADIOTERAPIA E HORMONIOTERAPIA

Prof. Dr. José Mir Justino da Costa
Prof. Dr. Jeremias da Silva Leão

# Resumo do trabalho

- Neste trabalho, foi apresentada uma adaptação de um modelo matemático existente na literatura para tratar o câncer de próstata utilizando radioterapia e hormonioterapia em conjunto.
- O modelo adaptado é composto por um sistema de equações diferenciais acopladas.
- Como o trabalho está relacionado a Problemas Inversos, foi utilizada uma análise inversa com um filtro Bayesiano.
- Estima as variáveis de estado e parâmetros do modelo.
- A qualidade do ajuste foi avaliada usando o erro quadrático médio (EQM).

# Objetivo geral

Apresentar um modelo matemático da literatura, baseado em Equações Diferenciais Ordinárias (EDO) e modificá-lo para incorporar tratamento de câncer de próstata via radioterapia e hormonioterapia, bem como usar filtro Bayesiano para estimativa combinada de estados e parâmetros.

# Geração de medidas

Para a geração das medidas que foram utilizadas neste trabalho adotamos o seguinte:

$$\eta_{medido} = \eta_{simulado} + \varepsilon$$

Em que, $\eta_{medido}$, resulta da soma de $\eta_{simulado}$, que é a solução do problema direto e de $\varepsilon$. Aqui consideramos $\varepsilon$ como uma variável aleatória tendo distribuição normal, com média igual a zero e variância conhecida e igual a 5% do valor máximo de cada população medida. Vale ressaltar que foi gerado medidas apenas para a população de células normais e tumorais.

# Solução do Problema Inverso

- Os problemas inversos formam um conjunto de problemas matemáticos onde se objetiva determinar a causa de um fenômeno particular (solução do problema), observando-se o efeito por ele produzido (dados).

- Neste trabalho foi utilizado o filtro de partículas proposto por (LIU; WEST, 2001) para estimar simultaneamente os estados do processo (população celular) e os parâmetros do modelo matemático modificado.

Em um problema de estimativa de estado consideramos o seguinte modelo de evolução:

$$x_k = f\left(x_{k-1}, v_{k-1}\right) \tag{10}$$

onde,

- $\mathbf{x}$ é o vetor de estado que contém todas as variáveis que serão estimadas dinamicamente
- o subscrito $k = 1, 2, ...$, representa o instante de tempo $t$
- $f$ é geralmente uma função não linear das variáveis de estado $\mathbf{x}$ e do vetor de incertezas $\mathbf{v}$

# Filtro de Partículas

Sendo $z_k \in \boldsymbol{R}^n$ as medidas observadas e $\eta \in R^n$ as incertezas associadas a essas medidas, podemos considerar o modelo de observação ou de medidas como:

$$z_k = h(x_k, \eta_k) \tag{11}$$

O problema de estimação de estado tem por objetivo obter informações sobre $x_k$ baseado nas equações (10 - 11), e as seguintes suposições são feitas:

(i) A sequência $x_k$ para $k = 1, 2, \ldots$ é um processo Markoviano de primeira ordem, ou seja

$$\pi\left(x_k \mid x_0, x_1, \ldots, x_{k-1}\right) = \pi\left(x_k \mid x_{k-1}\right)$$

(ii) A sequência $\mathrm{z_k}$ para $k = 1, 2, \ldots$ é um processo Markoviano com respeito a história de $\mathrm{x_k}$, isto é:

$$\pi\left(z_k \mid z_0, z_1, \ldots, z_{k-1}\right) = \pi\left(z_k \mid z_{k-1}\right)$$

(iii) A sequência $x_k$ depende das observações passadas através da sua própria história, ou seja:

$$\pi\left(x_k \mid x_{k-1}, z_{1:k-1}\right) = \pi\left(x_k \mid x_{k-1}\right)$$

onde a expressão $\pi(a \mid b)$ representa a densidade de probabilidade de $a$ dado $b$.

Outras suposições também são feitas, tais como (KAIPIO; SOMERSALO, 2006):

(a) Os vetores de ruídos $\boldsymbol{v}_i$ e $v_j$ para $i \neq j$, bem como $\eta_i$ e $\eta_j$ são mutuamente independentes e mutuamente independentes do estado inicial $\mathrm{x}_0$.

(b) Os vetores $v_i$ e $\eta_i$ são mutuamente independente para todos os valores de $i \neq j$.

Sendo $\pi\left(x_0 \mid z_0\right) = \pi\left(x_0\right)$ uma informação a priori no instante inicial $t_0$ em um problema de filtragem o objetivo é obter uma aproximação da distribuição a posteriori $\pi\left(x_k \mid z_{1:k}\right)$. Esta posteriori é obtida em dois passos: previsão e atualização.

Como neste trabalho estamos interessados na estimativa combinada de estados e parâmetros, usoou-se o filtro de (LIU; WEST, 2001) que é uma generalização do filtro ASIR. Neste filtro a inferência é feita sobre

a densidade a posteiori conjunta $\pi\left(x_k, \theta \mid z_{1:k}\right)$. Sendo $\theta$ o vetor de parâmetros e $\boldsymbol{x}_k$ as variáveis de estado. (LIU; WEST, 2001) assume que para um vetor de parâmetros estáticos $\theta$ a aproximação da densidade a posteriori $\pi\left(\theta \mid z_k\right)$ é feita por densidade suavizada via kernel

$$\pi\left(\theta \mid z_k\right) \approx \sum_{i=1}^{N} w_k^i N\left(\theta \mid m_k^i, h^2 V_k\right) \tag{12}$$

# Filtro de Partículas

Dessa forma temos que a distribuição artificial do parâmetros $\boldsymbol{\theta}$ é representada pela mistura ponderada por $\boldsymbol{w}_k^i$ de distribuições normais com média $\boldsymbol{m}_{\boldsymbol{k}}^i$ e variância $h^2 \boldsymbol{V}_k$ ((LIU; WEST, 2001),(COSTA, 2015)).

A sugestão ainda de (LIU; WEST, 2001) é a de que a constante $h$, parâmetro de suavização, seja escolhida como uma função suave e decrescente com relação a $N$. Outras considerações feitas foram que:

$$\boldsymbol{m}_k^i = a\boldsymbol{\theta}_k^i + (1-a)\overline{\boldsymbol{\theta}}_k. \tag{13}$$

$$a = \sqrt{(1-h^2)}. \tag{14}$$

$$\overline{\boldsymbol{\theta}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\theta}_k^i. \tag{15}$$

$$\boldsymbol{V}_k = \frac{1}{N}\sum_{i=1}^{N} \left(\boldsymbol{\theta}_k^i - \overline{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}_k^i - \overline{\boldsymbol{\theta}}\right)^T, \tag{16}$$

onde $T$ denota o transposto da matriz, $a$ está relacionado a um fator de desconto $\delta$ de acordo com (LIU; WEST, 2001) da seguinte forma :

$$a = \frac{3\delta - 1}{2\delta} \tag{17}$$

# Resultados

# Conclusão

A utilização do filtro de partículas para estimativa combinada de variáveis de estado e parâmetros funcionou de forma bastante satisfatória. Até mesmo para a variável, população de células imunológicas a qual não tínhamos medida, o filtro conseguiu boa estimativa quando utilizadas 5000 partículas. Com relação aos parâmetros também foram obtidos, em geral, um bom ajuste quando comparados aos valores exatos. Dada a natureza não linear do modelo utilizado alguns parâmetros precisaram de quantidade de partículas maiores do que para outros parâmetros.