

Avaliação 2

1 Penalised Complexity Prior

Considere $X \sim \text{Bernoulli}(\theta)$ como o modelo complexo com o parâmetro de complexidade θ e $X \sim \text{Bernoulli}(1/2)$ como o modelo de base.

1. Encontre a medida de divergência KL entre o modelo Bernoulli de base e o modelo Bernoulli complexo;
2. Seguindo os princípios de construção da PCP (veja Daniel Simpson et. al., 2017), encontre a PCP para o parâmetro θ ;
3. Plote a PCP para diferentes valores da taxa de penalização e faça uma discussão sobre a interpretação da taxa de penalização;
4. Estudo de simulação: assumindo $X_i \sim \text{Bernoulli}(0.35)$, $i = 1, \dots, 300$, plote a distribuição a posteriori de θ para diferentes valores da taxa de penalização. O estudo mostra que o Princípio da Navalha de Occam é satisfeito?

1.1 Medida KL

A função massa de probabilidade do modelo complexo é dada por

$$p(X = k \mid \theta) = \theta^k (1 - \theta)^{1-k} \quad (1)$$

e do modelo base por

$$q(X = k \mid \theta_0 = 1/2) = \theta_0^k (1 - \theta_0)^{1-k} \quad (2)$$

A medida de divergência de Kullback-Leibler, para distribuições discretas, é definida por

$$\text{KL}(p||q) = \sum_{k \in K} p(k) \log \frac{p(k)}{q(k)} \quad (3)$$

sendo neste caso $K = \{0, 1\}$. Temos então:

$$\begin{aligned} \text{KL}(p||q) &= \sum_{k=0}^1 p(k \mid \theta) \log \frac{p(k \mid \theta)}{q(k \mid \theta_0)} \\ &= \theta \log \left(\frac{\theta}{\theta_0} \right) + (1 - \theta) \log \left(\frac{1 - \theta}{1 - \theta_0} \right) \end{aligned} \quad (4)$$

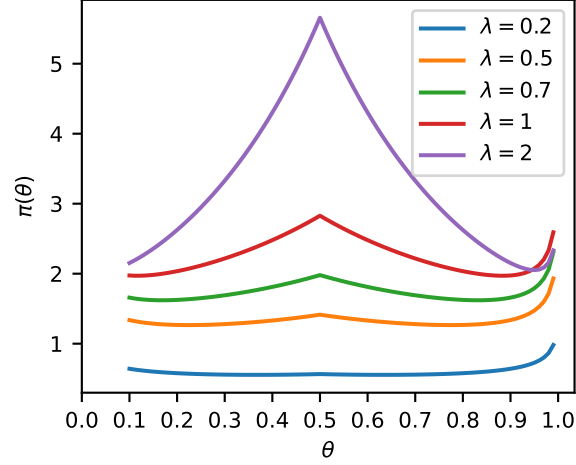


Figura 1: Densidade da PCP para diferentes valores da taxa de penalização λ

1.2 Construção da PCP

Primeiramente, definimos a distância entre as distribuições por

$$d(\theta) = \sqrt{2 \text{KL}(p||q)}. \quad (5)$$

Para a construção da PCP, assumimos que $d(\theta)$ segue uma distribuição exponencial com parâmetro λ :

$$\pi(d(\theta)) = \lambda e^{-\lambda d(\theta)}, \quad d(\theta) > 0, \quad (6)$$

onde λ é a taxa de penalização do modelo complexo. Então, aplicamos uma mudança de variáveis $d \rightarrow \theta$ para obtermos a densidade da PCP:

$$\pi(\theta) = \pi_d(d(\theta)) \left| \frac{\partial}{\partial \theta} d(\theta) \right|. \quad (7)$$

A densidade da PCP fica então:

$$\pi(\theta) = \lambda e^{-\lambda d(\theta)} \left| \frac{1}{d(\theta)} \left[\log \left(\frac{\theta}{\theta_0} \right) - \log \left(\frac{1-\theta}{1-\theta_0} \right) \right] \right|. \quad (8)$$

1.3 Gráficos da PCP

Na fig. 1, plotamos a densidade da PCP para alguns valores da taxa de penalização λ . Observe que, quanto maior λ , maior peso é dado ao modelo base (maior peso para $\theta = 1/2$), ou seja, mais o modelo complexo tende a ser penalizado.

1.4 Estudo de simulação

No estudo de simulação, assumimos $X_i \sim \text{Binomial}(\theta = 0.35)$, $i = 1, \dots, 300$.

Seja $n = 300$ o tamanho da amostra e n_1 o número de sucessos, ou seja $n_1 = \sum_{i=1}^n \mathbb{I}(x_i)$, onde $\mathbb{I}(\cdot)$ é a função indicadora. A distribuição a posteriori para θ fica então:

$$\begin{aligned}\pi(\theta \mid \mathbf{x}) &\propto \mathcal{L}(\theta \mid \mathbf{x})\pi(\theta) \\ &= \theta^{n_1}(1 - \theta)^{n - n_1}\pi(\theta),\end{aligned}\tag{9}$$

$$\tag{10}$$

com $\pi(\theta)$ dada pela eq. (8).

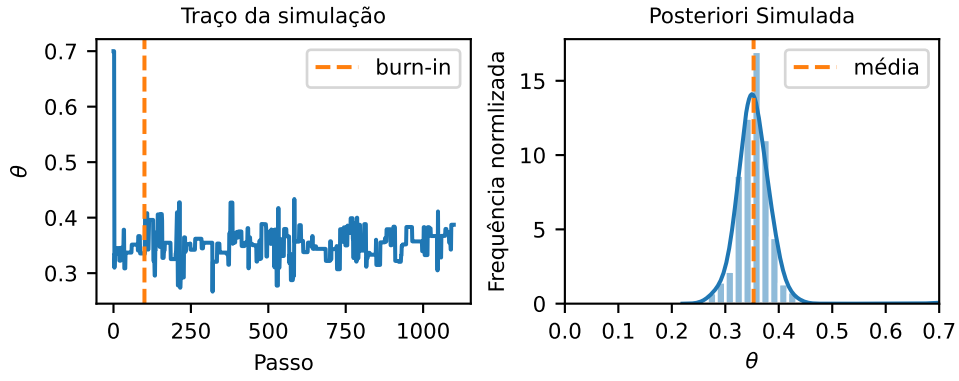
Não conhecemos, na literatura, distribuição com esta densidade. Então, para simularmos valores desta posteriori, recorreremos ao método MCMC de Metropolis com Passeio Aleatório. Adotamos a seguinte estratégia de simulação:

- Fixamos $n_1 = 0.35 \times 300 = 105$;
- Distribuição para a geração dos valores candidatos: $\mathcal{N}(\cdot, s^2)$, $s^2 = 0.2^2$;
- Inicialização: $\theta = 0.7$;
- Total de passos/amostras simuladas: 1.100;
- Descarte das 100 primeiras amostras iniciais (*burn-in*);
- Taxa de penalização: $\lambda = [1, 10, 50, 100]$.

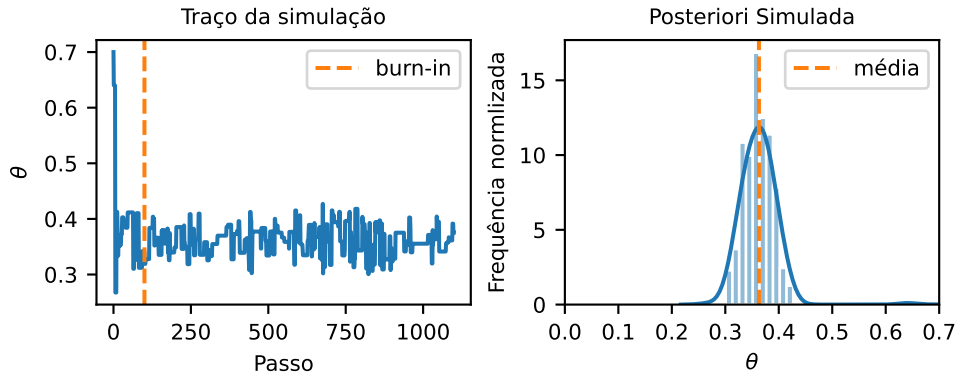
As densidades a posteriori simuladas para θ , bem como o gráfico de traço da simulação, são mostrados na fig. 2. Podemos observar que:

- Com $\lambda = 1$ e $\lambda = 10$, praticamente não houve influência da priori na inferência, ficando a média da posteriori em torno de $\theta = 0.35$.
- Já com $\lambda = 50$ e $\lambda = 100$, a PCP de fato penalizou o modelo complexo ao "puxar" a posteriori em direção ao modelo base, isto é, a média a posteriori ficou mais próxima de $\theta_0 = 1/2$.

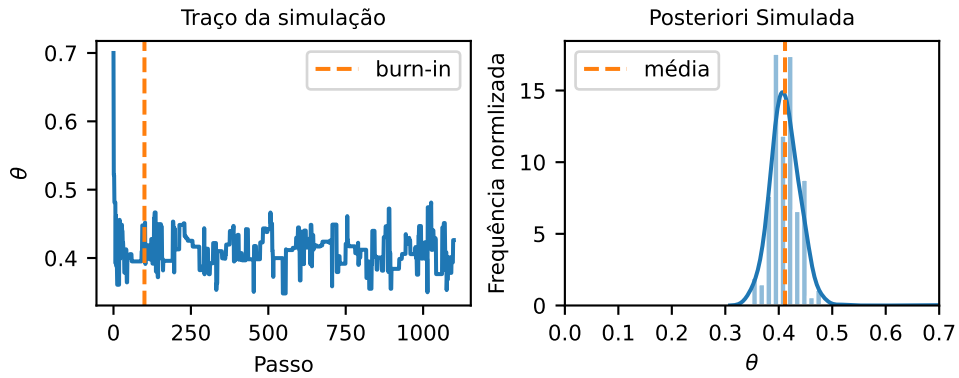
Concluimos então que o modelo com PCP de fato satisfaz o Princípio da Navalha de Occam ao priorizar o modelo mais simples. Porém, o estudo de simulação mostrou que esta priorização depende do ajuste da taxa de penalização. Na prática, este ajuste pode não ser fácil.



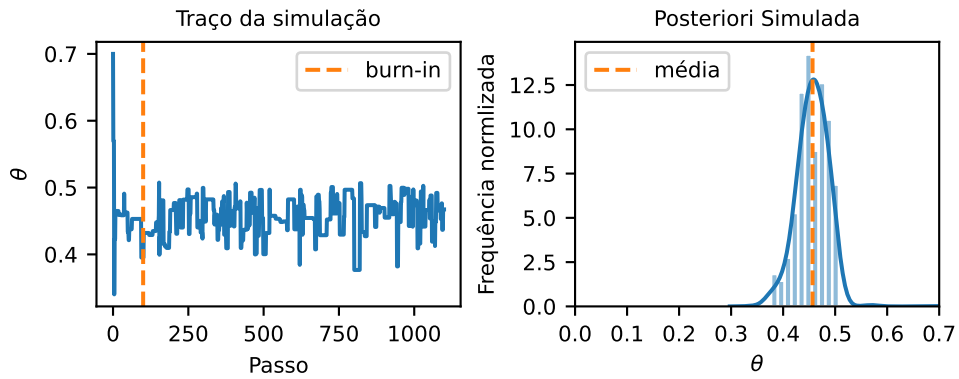
(a) Simulação com $\lambda = 1$



(b) Simulação com $\lambda = 10$



(c) Simulação com $\lambda = 50$



(d) Simulação com $\lambda = 100$

Figura 2: Traço da simulação MCMC e densidade a posteriori $\pi(\theta \mid \mathbf{x})$ para diferentes valores da taxa de penalização λ .

2 Modelo Hierárquico e Power Prior

Suponha que estamos interessados em estimar a média da concentração química θ de um processo industrial. Seja $X \mid \theta \sim \mathcal{N}(\theta, \sigma^2 = 1)$, uma medida observada recentemente de θ . Anteriormente à observação recente, foram observadas três medidas de concentração: $X_i \mid \theta_i \sim \mathcal{N}(\theta_i, \sigma^2 = 1)$, $i = 1, 2, 3$, tais que $\theta_i \mid \theta \sim \mathcal{N}(\theta, \tau^2 = 1)$. Considere uma a priori constante para θ .

1. Determine a distribuição a posteriori $\pi(\theta \mid X_1, X_2, X_3)$;
2. Determine a distribuição a posteriori $\pi(\theta \mid X)$ com uma priori para θ baseada nas informações passadas (*Power Prior*). Por favor, eu gostaria de ver alguma discussão a respeito dos casos extremos $a_0 = 0$ e $a_0 = 1$.
3. Estudo de simulação: Através de um estudo de simulação, para $\theta \in \mathbb{R}^+$ e $n = 50$, plote a probabilidade de cobertura de θ considerando intervalos de credibilidade 95% e valores de a_0 . O que você pensa sobre a sensibilidade das probabilidades de cobertura de θ com respeito a a_0 ?

2.1 Posteriori $\pi(\theta \mid x_1, x_2, x_3)$

Este exercício é semelhante ao discutido em aula e na Lista 6 (Modelo Hierárquico), com exceção de que, no modelo da Lista 6, consideramos no terceiro estágio do modelo hierárquico uma priori $\theta \sim \mathcal{N}(0, w^2)$ ao invés de constante. Porém, conforme mostrado no livro do Gelman (*Bayesian Data Analysis*, 3ª ed.), seção 5.4, pg. 117, eq. 5.20, a distribuição a posteriori neste caso (com priori constante) é semelhante àquela com priori $\mathcal{N}(0, w^2)$ quando $w \rightarrow \infty$:

$$\pi(\theta \mid \mathbf{x}) \sim \mathcal{N}(\tilde{\mu}, \tilde{w}^2), \quad (11)$$

$$\tilde{w}^2 = \left(\sum_{j=1}^J \frac{1}{\frac{\sigma^2}{n_j} + \tau^2} + \frac{1}{w^2} \right)^{-1}, \quad (12)$$

$$\tilde{\mu} = \tilde{w}^2 \left(\sum_{j=1}^J \frac{\bar{x}_j}{\frac{\sigma^2}{n_j} + \tau^2} \right). \quad (13)$$

No caso concreto desta questão, $\sigma^2 = \tau^2 = 1$, $n_j = 1$ e $\bar{x}_j = x_j$, para $j = 1, 2, 3$. Então, as equações se reduzem para:

$$\pi(\theta \mid x_1, x_2, x_3) \sim \mathcal{N}(\tilde{\mu}, \tilde{w}^2), \quad (14)$$

$$\tilde{\mu} = \frac{x_1 + x_2 + x_3}{3} \quad (15)$$

$$\tilde{w}^2 = \frac{2}{3} \quad (16)$$

2.2 Posteriori $\pi(\theta|x)$

A *Power Prior* é definida como

$$\pi(\theta \mid \mathbf{x}_0) \propto [\mathcal{L}(\theta_0 \mid \mathbf{x}_0)]^{a_0} \pi(\theta) \quad (17)$$

onde $\mathcal{L}(\theta_0 \mid \mathbf{x}_0)$ é a função de verossimilhança dos dados históricos \mathbf{x}_0 e a_0 é o hiperparâmetro que controla a influência dos dados históricos.

Na Lista 5, vimos que, quando utilizamos o modelo normal e consideremos $\pi(\theta) \propto 1$, a posteriori é normal:

$$\pi(\theta \mid \mathbf{x}, \mathbf{x}_0) \sim \mathcal{N}(\theta \mid \mu_*, \sigma_*^2) \quad (18)$$

com

$$\begin{aligned} \mu_* &= \frac{\frac{n}{\sigma^2} \bar{x} + a_0 \frac{n_0}{\sigma_0^2} \bar{x}_0}{\frac{n}{\sigma^2} + a_0 \frac{n_0}{\sigma_0^2}} \\ \sigma_*^2 &= \frac{1}{\frac{n}{\sigma^2} + a_0 \frac{n_0}{\sigma_0^2}} \end{aligned} \quad (19)$$

Em nosso caso concreto, temos que $n_0 = 3$, $\bar{x}_0 = \tilde{\mu}$ (definido na seção anterior) $n = 1$, $\bar{x} = x$ e $\sigma^2 = 1$. Falta calcularmos σ_0^2 . Utilizando a lei da variância total:

$$\sigma_0^2 = \text{Var}(X_i \mid \theta) = \mathbb{E}[\text{Var}(X_i \mid \theta_i) \mid \theta] + \text{Var}(\mathbb{E}[X_i \mid \theta_i] \mid \theta) \quad (20)$$

Mas:

$$\text{Var}(X_i \mid \theta_i) = \sigma^2 = 1 \quad (21)$$

$$\mathbb{E}[X_i \mid \theta_i] = \theta_i \quad (22)$$

$$\text{Var}[\theta_i \mid \theta] = \tau^2 = 1, \quad (23)$$

e então

$$\begin{aligned} \sigma_0^2 &= \mathbb{E}[\sigma^2 \mid \theta] + \text{Var}(\theta_i \mid \theta) \\ &= \sigma^2 + \tau^2 \\ &= 2. \end{aligned} \quad (24)$$

Finalmente, temos:

$$\begin{aligned} \mu_* &= \frac{x + \frac{3}{2} a_0 \bar{x}_0}{1 + \frac{3}{2} a_0} \\ \sigma_*^2 &= \frac{1}{1 + \frac{3}{2} a_0} \end{aligned} \quad (25)$$

Vemos então que a posteriori é uma média ponderada do dado atual e dados históricos. No caso extremo em que $a_0 = 0$, estamos ignorando os dados passados e fazendo a inferência somente com o dado atual. Já no caso extremo em que $a_0 = 1$, a ponderação dos dados históricos é feita com peso 3/5 e peso 2/5 para o dado atual.

2.3 Estudo de simulação

A simulação foi realizada seguindo os seguintes passos:

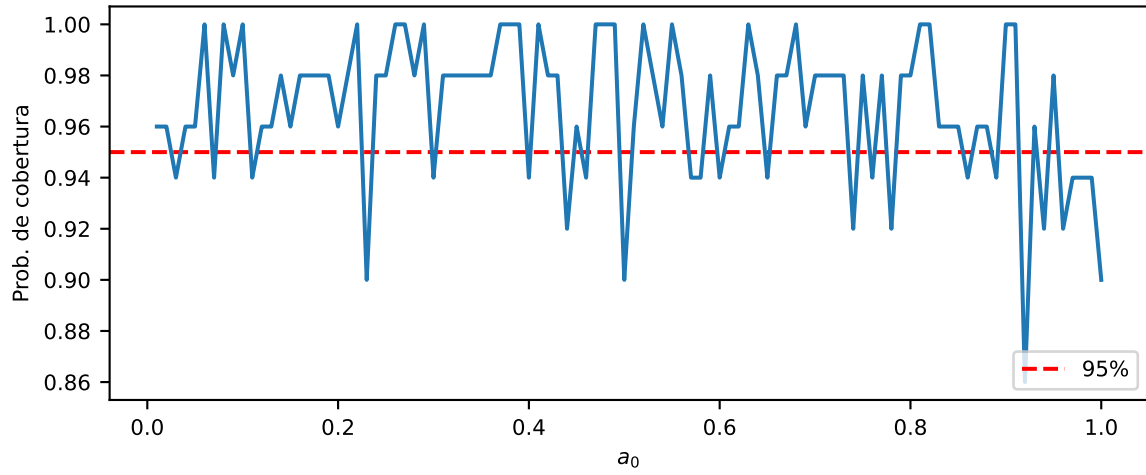
1. Fixamos $\theta = 5$ (escolhido pelo “martelo” da natureza);
2. Para cada a_0 no intervalo $(0.01, 0.02, \dots, 1)$, fazemos:
 - (a) Geramos os dados históricos x_1, x_2, x_3 com o modelo hierárquico:
 - i. Primeiro, geramos θ_i à partir de $\mathcal{N}(\theta, \tau^2 = 1)$
 - ii. Depois, geramos x_i à partir de $\mathcal{N}(\theta_i, \sigma^2 = 1)$;
 - (b) Geramos o dado atual x à partir de $\mathcal{N}(\theta, \sigma^2 = 1)$;
 - (c) Calculamos a posteriori $\pi(\theta|x)$ à partir da *Power Prior* e do dado atual;
 - (d) Calculamos o intervalo de credibilidade 95% e verificamos se este cobre o θ ;
3. Repetimos os passos (a) a (d) $n = 50$ vezes e verificamos com que frequência os intervalos de credibilidade cobrem θ . Esta frequência é então uma estimativa da probabilidade de cobertura.

A fig. 3 apresenta os resultados da simulação. Com $n = 50$, observamos alta variabilidade nos resultados em função da semente de simulação utilizada. Aumentamos então o número de intervalos simulados para $n = 10000$, e então verificamos que os resultados concordavam para sementes diferentes.

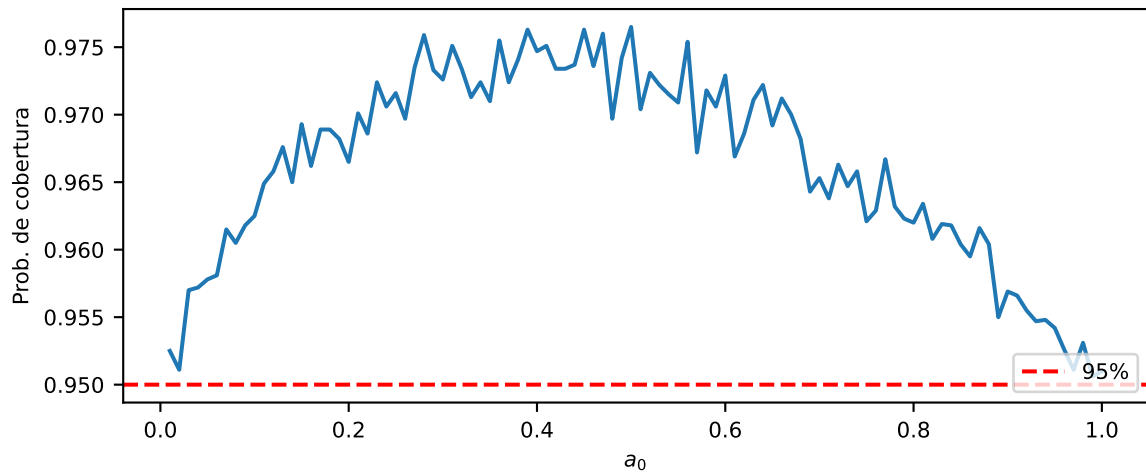
Observa-se na fig. 3 que o modelo apresentou probabilidade de cobertura acima do esperado, sobretudo quando a_0 está em torno de 0.5. Este resultado foi semelhante à simulação realizada na Lista 5. Para identificar as causas desta sobre cobertura, uma análise mais aprofundada seria necessária.

3 Implementações

Os questões 1 e 2 foram implementadas em Python. O código encontra-se no repositório <https://www.github.com/cleitonmoya/ib>.



(a) $n = 50$ intervalos



(b) $n = 10000$ intervalos

Figura 3: Simulação da probabilidade de cobertura para intervalos de credibilidade 95% com (a) $n = 50$ intervalos; (b) $n = 10000$ intervalos. De modo semelhante ao resultado da Lista 5 (Power Prior), observamos uma sobrecobertura (probabilidade acima do esperado).