Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

AULA 6 - INFERÊNCIA BAYESIANA

por **Cibele Russo**

ICMC/USP - São Carlos SP

PROGRAMA

- O paradigma Bayesiano.
- Os diferentes tipos de prioris.
- Distribuições conjugadas.
- Estimação Bayesiana.
- Densidade preditiva.
- Computação Bayesiana.
- Exemplos com PyMC3.

Referências e leituras recomendadas:

1. Migon, H. S., Gamerman, D. and Louzada, F. (2014). Statistical Inference: An Integrated Approach, Second Edition, CRC Press.
2. Caffo, B. (2016). Statistical Inference for Data Science. Leanpub. Disponível em https://leanpub.com/LittleInferenceBook
3. https://pt.wikipedia.org/wiki/Infer%C3%AAncia_bayesiana
4. Pacote PyMC https://pypi.org/project/pymc/

O PARADIGMA BAYESIANO

Seja uma amostra aleatória $X_1, \ldots, X_n$ que vem de um modelo $p(x|\theta), \theta \in \Theta$ e sejam $x_1, \ldots, x_n$ os dados observados.

Sob o **paradigma clássico ou frequentista**, θ é um **parâmetro fixo e desconhecido**.

Sob o **paradigma Bayesiano**, consideramos modelos probabilísticos para **representar a incerteza** a respeito de θ.

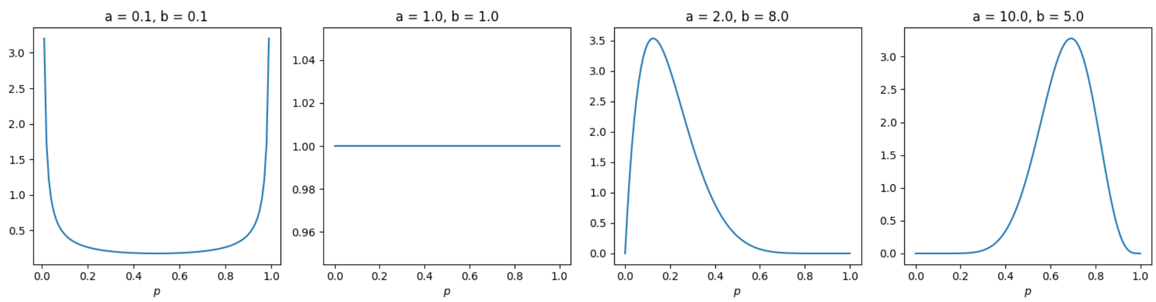
Se temos informação **a priori** sobre o parâmetro θ, antes de observar os dados, por que não usá-la?

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

EXEMPLO:

Na aplicação de dados bancários que vimos na Aula 1, suponha que antes de observar os dados, exista um conhecimento prévio de que a proporção p de inadimplentes esteja em torno de 20%. Como incorporar esse conhecimento prévio? Uma possibilidade seria considerar uma distribuição a priori beta para a proporção p.

[1]:
```python
# Mais sobre a distribuição beta: https://pt.wikipedia.org/wiki/
↪Distribui%C3%A7%C3%A3o_beta

import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
from scipy.stats import beta

# Conjuntos de parâmetros da distribuição a priori para a proporção de␣
↪inadimplentes
a_b_params = ((0.1, 0.1), (1, 1), (2, 8), (10, 5))
p = np.linspace(0, 1, 100)

# Plota as densidades da beta para cada conjunto de parâmetros
plt.figure(figsize=(15,4))
for i, (a, b) in enumerate(a_b_params):
    plt.subplot(1, len(a_b_params), i+1)

    prior = beta(a, b)

    plt.plot(p, prior.pdf(p))
    plt.xlabel(r'$p$')
    plt.title("a = {:.1f}, b = {:.1f}".format(a, b))
    plt.tight_layout()
```



Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

TEOREMA DE BAYES

Ver, por exemplo: https://pt.wikipedia.org/wiki/Teorema_de_Bayes

Sejam

- $p(\theta)$ a **distribuição a priori** para θ.
- $l(\theta, x) = p(x|\theta)$ a **verossimilhança** de θ.
- $p(\theta|x)$ a **distribuição a posteriori** de θ.

O Teorema de Bayes ilustra o aumento de informação com a introdução do conhecimento a priori

$$p(\theta|x) = \frac{p(x,\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\displaystyle\int p(\theta, x)d\theta}.$$

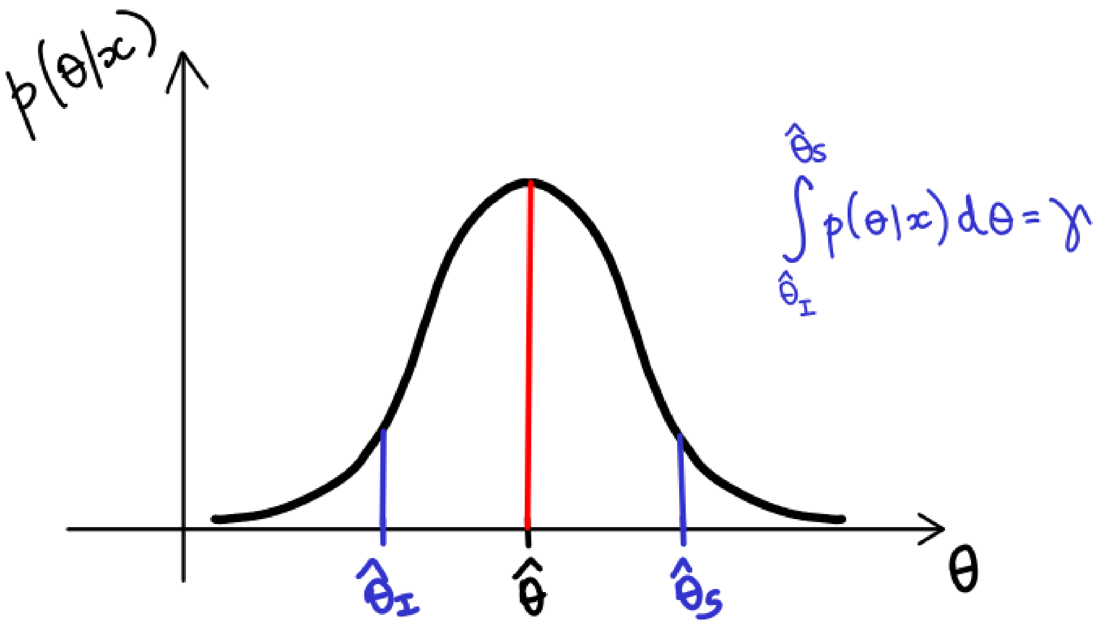
Em outras palavras:

distribuição a posteriori $\propto$ verossimilhança $\times$ distribuição a priori

Para um valor fixo de x, as duas fontes de informação para θ são

- a função $l(\theta; x) = p(x|\theta)$ fornece a plausibilidade ou verossimilhança de cada um dos possíveis valores de θ e
- $p(\theta)$ é chamada distribuição a priori de θ,

que, combinadas, levam à distribuição a posteriori de θ, $p(\theta|x)$.



Assim, a forma usual do teorema de Bayes é $p(\theta|x) \propto l(\theta; x)p(\theta)$.

O termo omitido $p(x)$ é apenas uma constante normalizadora e não depende de θ.

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

Para x fixo, a **verossimilhança** fornece a **plausibilidade** de cada um dos possíveis valores de θ.

Já a **distribuição a priori** $p(\theta)$ incorpora o **conhecimento do pesquisador**.

Essas duas quantidades combinadas são levadas à **distribuição a posteriori** de θ.

A **distribuição a posteriori** de θ dados $x_1, \ldots, x_n$ observados é dada por:

$$p(\theta|x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n|\theta)p(\theta)}{\int_\Theta p(x_1, \ldots, x_n|\theta)p(\theta)d\theta}$$

- $p(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta) = L(\theta|x_1, \ldots, x_n)$ é a **função de verossimilhança** de θ.
- O denominador
$$\int_\Theta p(x_1, \ldots, x_n|\theta)p(\theta)d\theta = C(x_1, \ldots, x_n).$$

É comum escrever que

$p(\theta|x_1, \ldots, x_n) = \dfrac{L(\theta|x_1, \ldots, x_n)p(\theta)}{C(x_1, \ldots, x_n)} \propto L(\theta|x_1, \ldots, x_n)p(\theta),$

DISTRIBUIÇÃO PREDITIVA

A constante normalizadora da posteriori pode ser facilmente recuperada pois $p(\theta|x) = kp(x|\theta)p(\theta)$ onde

$k^{-1} = \int p(x|\theta)p(\theta)d\theta = E_\theta[p(X|\theta)] = p(x)$

chamada **distribuição preditiva** (ou marginal) de X.

Esta é a distribuição esperada para a observação x dado θ. Assim,

- Antes de observar X podemos checar a adequação da priori fazendo predições via $p(x)$.
- Se X observado recebia pouca probabilidade preditiva então o modelo deve ser questionado, revisado, ou existe observação aberrante.

Se, após observar $X = x$, estamos interessados na previsão de uma quantidade Y, também relacionada com θ, e descrita probabilisticamente por $p(y|\theta)$ então

$$p(y|x) = \int p(y, \theta|x)d\theta = \int p(y|\theta, x)p(\theta|x)d\theta$$

Os conceitos de priori e posteriori são relativos àquela observação que está sendo considerada no momento. Assim, $p(\theta|x)$ é a posteriori de θ em relação a X (que já foi observado) mas é a priori de θ em relação a Y (que não foi observado ainda).

Após observar $Y = y$ uma nova posteriori (relativa a $X = x$ e $Y = y$) é obtida aplicando-se novamente o teorema de Bayes.

Exemplo (Gamerman e Migon, 1993)

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

Um médico "desconfia" que um paciente pode ter uma doença. Baseado na sua experiência, no seu conhecimento sobre esta doença e nas informações dadas pelo paciente, **ele assume que a probabilidade do paciente ter a doença é 0.7**. A quantidade de interesse, desconhecida, é definida como

$$\theta = \begin{cases} 1, & \text{se o paciente tem a doença,} \\ 0, & \text{se o paciente não tem a doença.} \end{cases}$$

Para aumentar sua quantidade de informação sobre a doença o médico aplica um teste X relacionado com θ através da distribuição

$P(X = 1|\theta = 0) = 0.40$ e

$P(X = 1|\theta = 1) = 0.95$

e o resultado do teste foi positivo (ou seja, observou-se $X = 1$).

É bem intuitivo que a probabilidade de doença deve ter aumentado após este resultado e a questão aqui é quantificar este aumento. Usando o teorema de Bayes, segue que

$P(\theta = 1|X = 1) \propto l(\theta = 1; X = 1)p(\theta = 1) = (0.95)(0.7) = 0.665$

$P(\theta = 0|X = 1) \propto l(\theta = 0; X = 1)p(\theta = 0) = (0.40)(0.3) = 0.120.$

A constante normalizadora é tal que $P(\theta = 0|X = 1) + P(\theta = 1|X = 1) = 1$, i.e.,

$k(0.665) + k(0.120) = 1$ e $k = 1/0.785$.

Portanto, a distribuição a posteriori de θ é

$P(\theta = 1|X = 1) = 0.665/0.785 = 0.847$

$P(\theta = 0|X = 1) = 0.120/0.785 = 0.153.$

O aumento na probabilidade de doença não foi muito grande porque a verossimilhança $l(\theta = 0; X = 1)$ também era grande (o modelo atribuia uma plausibilidade grande para $\theta = 0$ mesmo quando $X = 1$).

Agora o médico aplica outro teste Y cujo resultado está relacionado a θ através da seguinte distribuição

$P(Y = 1|\theta = 0) = 0.04$ e

$P(Y = 1|\theta = 1) = 0.99.$

Mas antes de observar o resultado deste teste é interessante obter sua distribuição preditiva.

Como θ é uma quantidade discreta segue que

$p(y|x) = \sum_\theta p(y|\theta)p(\theta|x)$

e note que $p(\theta|x)$ é a priori em relação a Y.

Assim,

$P(Y = 1|X = 1) = P(Y = 1|\theta = 0)P(\theta = 0|X = 1) + P(Y = 1|\theta = 1)P(\theta = 1|X = 1)$

$= (0.04)(0.153) + (0.99)(0.847) = 0.845$

$P(Y = 0|X = 1) = 1 - P(Y = 1|X = 1) = 0.155.$

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

O resultado deste teste foi negativo ($Y = 0$).

Neste caso, é também intuitivo que a probabilidade de doença deve ter diminuido e esta redução será quantificada por uma nova aplicação do teorema de Bayes,

$$P(\theta = 1|X = 1, Y = 0) \propto l(\theta = 1; Y = 0)P(\theta = 1|X = 1) \propto (0.01)(0.847) = 0.0085$$

$$P(\theta = 0|X = 1, Y = 0) \propto l(\theta = 0; Y = 0)P(\theta = 0|X = 1) \propto (0.96)(0.153) = 0.1469.$$

A constante normalizadora é $1/(0.0085 + 0.1469) = 1/0.1554$ e assim a distribuição a posteriori de θ é

$$P(\theta = 1|X = 1, Y = 0) = 0.0085/0.1554 = 0.055$$

$$P(\theta = 0|X = 1, Y = 0) = 0.1469/0.1554 = 0.945.$$

Verifique como a probabilidade de doença se alterou ao longo do experimento

$$P(\theta = 1) = \begin{cases} 0.7 & \text{antes dos testes} \\ 0.847 & \text{após o teste X} \\ 0.055 & \text{após X e Y} \end{cases}$$

Note também que o valor observado de Y recebia pouca probabilidade preditiva. Isto pode levar o médico a repensar o modelo, i.e.,

(i) Será que $P(\theta = 1) = 0.7$ é uma priori adequada?

(ii) Será que as distribuições amostrais de X e Y estão corretas? O teste X é tão inexpressivo e Y é realmente tão poderoso?

OS DIFERENTES TIPOS DE PRIORIS

Destacamos as distribuições a priori

- **Priori não-informativa**
  - Uniforme
  - Priori vaga (às vezes imprópria)
- **Priori informativa**
  - Conhecimento do pesquisador dá informação sobre os parâmetros
- **Priori conjugada**
  - Priori e posteriori tem a mesma distribuição, a menos dos parâmetros (em geral facilita os cálculos)

DISTRIBUIÇÕES CONJUGADAS

Leituras recomendadas:

- Notas Ricardo Ehlers e Paulo Justiniano: `http://www.leg.ufpr.br/~paulojus/CE227/ce227/node1.html`
- Conjugate Prior Explained, with examples & proofs: `https://towardsdatascience.com/conjugate-prior-explained-75957dc80bfb`

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

Vantagem de usar distribuições a priori conjugadas: principalmente **ganho de custo computacional.**

	Priori	**Núcleo da Verossimilhança**	**Posteriori**
θ proporção	Beta	Bernoulli	Beta
θ média	Normal	Normal	Normal
taxa de falha	Gama	Poisson	Gama
	Dirichlet	Multinomial	Dirichlet

EXEMPLO DE PRIORI CONJUGADA BETA-BERNOULLI

Ver https://towardsdatascience.com/conjugate-prior-explained-75957dc80bfb

No exemplo do banco, se considerarmos que

- $X = \begin{cases} 1, & \text{se o cliente é classificado como inadimplente,} \\ 0, & \text{caso contrário.} \end{cases}$
- $X \sim Bernoulli(p)$
- Verossimilhança:

Para n suficientemente grande, pelo TLC sabemos que a distribuição amostral de $\bar{X}$ se aproxima da normal (será usada para graficar a verossimilhança)

$$\bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Além disso, $Y = \sum_{i=1}^{n} X_i \sim binomial(n, p)$.

- Priori: $p \sim beta(2, 8)$
- Posteriori: $p|k \sim beta(k + a, n - k + b)$

onde k é o número de sucessos observados na amostra.

[2]:
```python
import pandas as pd

# Indique o seu diretório se necessário
#pkgdir = '/hdd/MBA/ECD/Data'
#dados = pd.read_csv(f'{pkgdir}/dados_banco.csv', index_col=0)

# Dados banco - Leitura dos dados
dados = pd.read_csv('https://raw.githubusercontent.com/cibelerusso/
↪Estatistica-Ciencia-Dados/main/Data/dados_banco.csv', index_col=0)

dados.head()
```

[2]:
```
        Sexo  Idade  Empresa  Salario  Saldo_cc  Saldo_poupança  \
Cliente
75928      M     32  Privada  5719.00    933.79             0.0
```

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

```
52921        F      28    Privada  5064.00     628.37            0.0
8387         F      24   Autônomo  4739.00     889.18            0.0
54522        M      30    Pública  5215.00    1141.47            0.0
45397        M      30   Autônomo  5215.56     520.70            0.0

         Saldo_investimento  Devedor_cartao  Inadimplente
Cliente
75928                   0.0         6023.68             0
52921                   0.0         1578.24             0
8387                    0.0         2578.70             0
54522                   0.0         4348.96             0
45397                   0.0         1516.78             1
```

[3]:
```
# Vamos trabalhar com uma amostra

import random

a = 10
b = 5

amostra = dados.sample(n=100, replace=False, random_state=10)

n = len(amostra)
k = amostra['Inadimplente'].sum()
posteriori = beta(a + k, n - k + b)

k/n
```

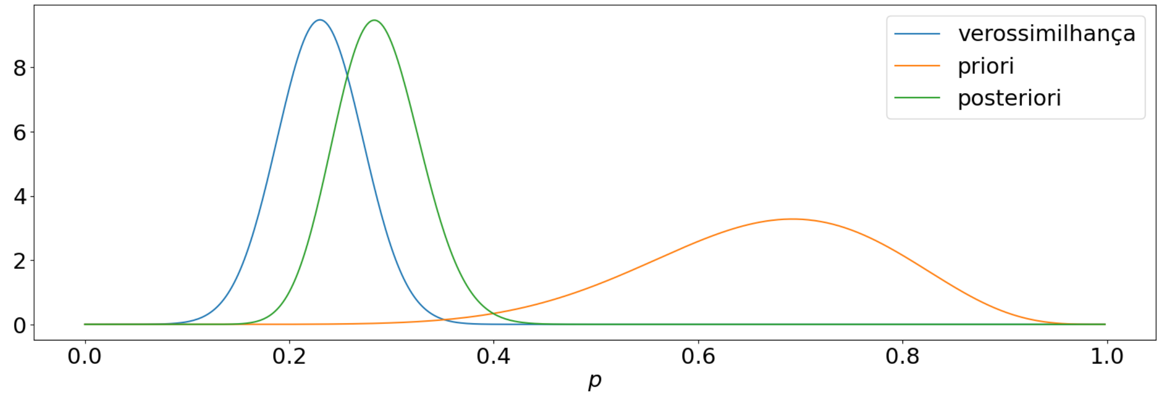
[3]:
```
0.23
```

[4]:
```
from scipy.stats import norm

# Eixo x entre 0 e 1 de .002 em .002.
x_axis = np.arange(0, 1, 0.002)

# Plota as densidades da beta para cada conjunto de parâmetros
plt.figure(figsize=(20,6))

prior = beta(a, b)

p_chapeu = amostra['Inadimplente'].mean()
dp = np.sqrt(p_chapeu*(1-p_chapeu)/n)
```

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

```python
media = p_chapeu
dp = np.sqrt(media*(1-media)/n)

plt.s = 0
plt.rcParams.update({'font.size': 22})

plt.plot(x_axis, norm.pdf(x_axis, media, dp), label='verossimilhança')
plt.plot(x_axis, prior.pdf(x_axis), label='priori')
plt.plot(x_axis, posteriori.pdf(x_axis), label='posteriori')
plt.xlabel(r'$p$')
plt.legend()
```

[4]: <matplotlib.legend.Legend at 0x7d990dd4f2e0>



[5]:
```python
# Estimador bayesiano EAP (Esperança a posteriori)
print('Média: %.2f' % posteriori.mean())



# E para calcular um intervalo de credibilidade, decidimos uma probabilidade
# Por exemplo 95% para a credibilidade
# Uma maneira seria definir que 2,5% de cada cauda como os limites do intervalo␣
↪(chamado intervalo simétrico)
# Este método é válido quando a posteriori se aproxima de uma distribuição␣
↪simétrica, pois nesse caso tende a gerar o intervalo com menor amplitude
# A seguir, apresentamos outra solução com um intervalo de credibilidade de menor␣
↪amplitude.

LI = posteriori.ppf(.025)
LS = posteriori.ppf(.975)
print("Intervalo com 95% de credibilidade: ({:.3f}, {:.3f})".format(LI,LS))
```

```
Média: 0.29
```

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

```
Intervalo com 95% de credibilidade: (0.208, 0.373)
```

ESTIMAÇÃO BAYESIANA

Leituras recomendadas:

- Notas Ricardo Ehlers e Paulo Justiniano: `http://www.leg.ufpr.br/~paulojus/CE227/ce227/node1.html`
- Migon, H. S., Gamerman, D. and Louzada, F. (2014). Statistical Inference: An Integrated Approach, Second Edition, CRC Press.

INTRODUÇÃO À TEORIA DA DECISÃO

Um problema de decisão fica completamente especificado pela descrição dos seguintes espaços:

(i) Espaço do parâmetro ou estados da natureza, Θ.

(ii) Espaço dos resultados possíveis de um experimento, Ω.

(iii) Espaço de possíveis ações, A.

Uma regra de decisão δ é uma função definida em Ω que assume valores em A, i.e. $\delta : \Omega \to A$. A cada decisão δ e a cada possível valor do parâmetro θ podemos associar uma **perda** $L(\delta, \theta)$ assumindo valores positivos. Definimos assim uma **função de perda**.

Definição: O **risco de uma regra de decisão**, denotado por $R(\delta)$, é a **perda esperada a posteriori**, i.e.

$$R(\delta) = E_{\theta|\mathbf{x}}[L(\delta, \theta)].$$

Uma regra de decisão δ^* é **ótima** se tem risco mínimo, i.e. $R(\delta^*) < R(\delta), \, \forall \delta$. Esta regra será denominada **regra de Bayes** e seu risco, **risco de Bayes**.

Exemplo:

Um laboratório farmaceutico deve decidir pelo lançamento ou não de uma nova droga no mercado. O laboratório só lançará a droga se achar que ela é eficiente mas isto é exatamente o que é desconhecido.

Podemos associar um parâmetro θ aos estados da natureza:

- droga é eficiente ($\theta = 1$),
- droga não é eficiente ($\theta = 0$)

e as possíveis ações como

- lança a droga ($\delta = 1$),
- não lança a droga ($\delta = 0$).

Suponha que foi possível construir a seguinte tabela de perdas levando em conta a eficiência da droga,

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

	eficiente ($\theta = 1$)	**não eficiente ($\theta = 0$)**
lança ($\delta = 1$)	-500	600
não lança ($\delta = 0$)	1500	100

Vale notar que estas perdas traduzem uma **avaliação subjetiva** em relação à gravidade dos erros cometidos.

Suponha agora que a incerteza sobre os estados da natureza é descrita por

$P(\theta = 1) = \pi, 0 < \pi < 1$

avaliada na distribuição atualizada de θ (seja a priori ou a posteriori).

Note que, para δ fixo, $L(\delta, \theta)$ é uma variável aleatória discreta assumindo apenas dois valores com probabilidades π e $1 - \pi$.

Assim, usando a definição de risco obtemos que

$R(\delta = 0) = E(L(0, \theta)) = \pi 1500 + (1 - \pi)100 = 1400\pi + 100$

$R(\delta = 1) = E(L(1, \theta)) = \pi(-500) + (1 - \pi)600 = -1100\pi + 600$

Uma questão que se coloca aqui é, para que valores de π a regra de Bayes será de lançar a droga.

Não é difícil verificar que as duas ações levarão ao mesmo risco, i.e.

$R(\delta = 0) = R(\delta = 1)$ se somente se $\pi = 0.20$.

Além disso, para $\pi < 0.20$ temos que $R(\delta = 0) < R(\delta = 1)$ e a regra de Bayes consiste em **não lançar a droga** enquanto que $\pi > 0.20$ implica em $R(\delta = 0) > R(\delta = 1)$ e a regra de Bayes deve ser de **lançar a droga**.

ESTIMADORES DE BAYES

Considere uma amostra aleatória $X_1, \ldots, X_n$, tomada de uma distribuição com função de (densidade) de probabilidade $p(x|\theta)$, onde o valor do parâmetro θ é desconhecido.

Em um problema de inferência como este o valor de θ deve ser estimado a partir dos valores observados na amostra.

Se $\theta \in \Theta$ então é razoável que os possíveis valores de um estimador $\widehat{\theta}(\mathbf{x})$ também devam pertencer ao espaço Θ.

Além disso, um bom estimador é aquele para o qual, com alta probabilidade, apresenta o seguinte erro

$$\widehat{\theta}(\mathbf{x}) - \theta$$

próximo de zero.

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

FUNÇÕES DE PERDA

Para cada possível valor de θ e cada possível estimativa $a \in \Theta$ vamos associar **uma perda** $L(a,\theta)$ de modo que quanto maior a distância entre a e θ maior o valor da perda. A perda esperada a posteriori é dada por

$$E[L(a,\theta)|\mathbf{x}] = \int L(a,\theta)p(\theta|\mathbf{x})d\theta$$

O **estimador de Bayes** será aquele que minimiza a perda esperada.

Função de perda quadrática

$$L(a,\theta) = (a-\theta)^2$$

O estimador de Bayes para θ será a **média de sua distribuição atualizada** (EAP: expected a posteriori).

Função de perda absoluta (introduz punições que crescem linearmente com o erro de estimação)

$$L(a,\theta) = |a-\theta|$$

O estimador de Bayes para θ é a **mediana de sua distribuição atualizada**.

Função de perda 0-1 (associam uma perda fixa a um erro cometido, não importando sua magnitude)

$$L(a,\theta) = \begin{cases} 1 & \text{se} \quad |a-\theta| > \epsilon \\ 0 & \text{se} \quad |a-\theta| < \epsilon \end{cases}$$

para todo $\epsilon > 0$.

Neste caso, o estimador de Bayes é a **moda da distribuição atualizada de** θ (MAP: maximum a posteriori).

A moda da posteriori de θ também é chamado de estimador de máxima verossimilhança generalizado (EMVG) e é o mais fácil de ser obtido dentre os estimadores vistos até agora.

Exemplo:

Suponha que queremos estimar a proporção θ de itens defeituosos em um grande lote. Para isto será tomada uma amostra aleatória $X_1, \ldots, X_n$ de uma distribuição de Bernoulli com parâmetro θ. Usando uma priori conjugada beta(a, b), após observar a amostra a distribuição a posteriori é beta($a+k, b+n-k$) onde $k = \sum_{i=1}^{n} x_i$.

A média desta distribuição beta é dada por $(a+k)/(a+b+n)$ e portanto o estimador de Bayes de θ usando perda quadrática é

$\delta(\mathbf{x}) = \dfrac{a + \sum_{i=1}^{n} X_i}{a+b+n}$.

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

ESTIMAÇÃO POR INTERVALOS

Pode-se desenvolver a estimação por intervalos por meio do **intervalo de credibilidade (ou intervalo de confiança Bayesiano)** baseado no distribuição a posteriori.

Definição:

C é um **intervalo de credibilidade** de 100(1-α)%, ou nível de credibilidade (ou confiança) $1-\alpha$, para θ se $P(\theta \in C) \geq 1-\alpha$.

Obs: Quanto menor for o tamanho do intervalo, mais concentrada é a distribuição do parâmetro, ou seja o tamanho do intervalo informa sobre a dispersão de θ.

O **intervalo com o menor comprimento possível** é obtido tomando-se os valores de θ com maior densidade a posteriori, e esta idéia é expressa matematicamente na definição abaixo.

Definição:

Um intervalo de credibilidade C de 100(1-α)% para θ é de **máxima densidade a posteriori (MDP)** se $C = \{\theta \in \Theta : p(\theta|\mathbf{x}) \geq k(\alpha)\}$ onde $k(\alpha)$ é a maior constante tal que $P(\theta \in C) \geq 1-\alpha$.

Usando esta definição, todos os pontos dentro do intervalo MDP terão densidade maior do que qualquer ponto fora do intervalo.

Um problema com os intervalos MDP é que eles não são invariantes a transformações 1 a 1, a não ser para transformações lineares. O mesmo problema ocorre com intervalos de comprimento mínimo na inferência clássica.

COMPUTAÇÃO BAYESIANA

Leituras recomendadas:

- Métodos de Monte Carlo: `https://pt.wikipedia.org/wiki/M%C3%A9todo_de_Monte_Carlo`
- Notas Ricardo Ehlers e Paulo Justiniano: `http://www.leg.ufpr.br/~paulojus/CE227/ce227/node1.html`
- Salvatier, John; Wiecki, Thomas V.; Fonnesbeck, Christopher. Probabilistic programming in Python using PyMC3. PeerJ Computer Science, v. 2, p. e55, 2016. Disponível em `https://peerj.com/articles/cs-55/.Acessoem15/05/2024`.

A obtenção de informações a partir da distribuição a posteriori dos parâmetros pode envolver a avaliação de probabilidades ou esperanças, que exigem métodos computacionais baseados em simulações, como

- Método de Monte Carlo simples
- Monte Carlo com função de importância,
- Algoritmo de Metropolis-Hastings,
- Amostrador de Gibbs,

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

- Método do Bootstrap Bayesiano,
- Monte Carlo via cadeias de Markov (MCMC),
- Monte Carlo Hamiltoniano (HMC),
- No-U-Turn Sampler (NUTS).

HMC e NUTS aproveitam as informações de gradiente da probabilidade e alcançam uma convergência muito mais rápida do que os métodos de amostragem tradicionais, especialmente para modelos mais complexos.

O pacote PyMC3 do Python usam a programação probabilística para executar os métodos de HMC como o NUTS. Esse tipo de programação permite a especificação flexível e ajuste de modelos estatísticos bayesianos com sintaxe intuitiva e legível, embora poderosa, que é próxima da sintaxe natural que os estatísticos usam para descrever modelos.

EXEMPLOS

[6]:
```
!pip install theano
!pip install arviz==0.15.1
```

EXEMPLO BETA-BERNOULLI: CLIENTES DO BANCO

[9]:
```
with pm.Model() as model:
    p = pm.Beta("p", 2, 8) #priori
    obs = pm.distributions.discrete.Bernoulli("obs", p,␣
 ↪observed=amostra['Inadimplente'])

    idata = pm.sample(2000, tune=1500, return_inferencedata=True)
```

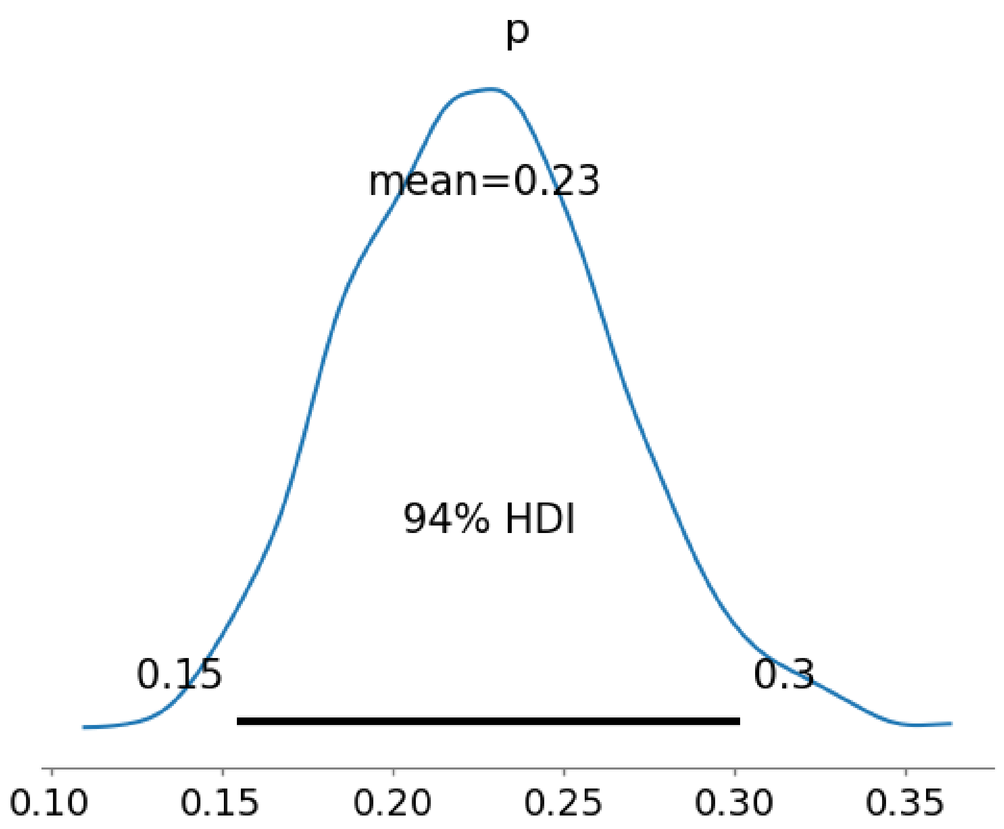
```
<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>
```

[10]:
```
az.summary(idata)
```

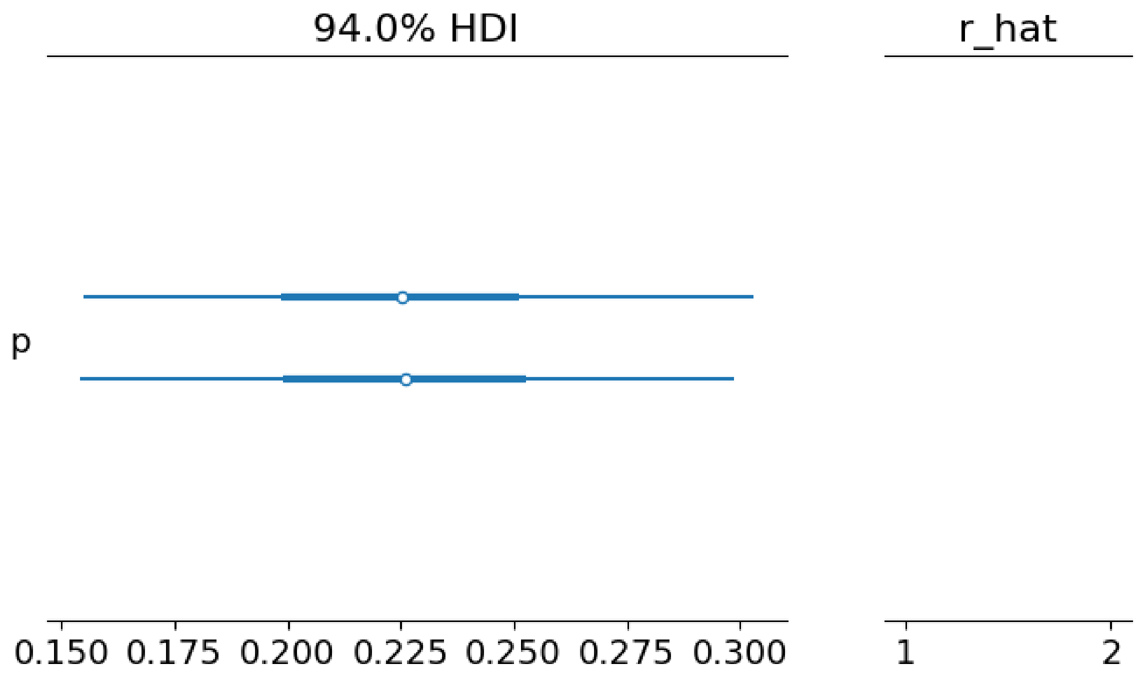
[10]:
```
    mean     sd  hdi_3%  hdi_97%  mcse_mean  mcse_sd  ess_bulk  ess_tail  \
p  0.227  0.039   0.155    0.302      0.001    0.001    1737.0    2710.0

   r_hat
p    1.0
```

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

```
[11]: az.plot_posterior(idata);
```




```
[12]: az.plot_forest(idata, r_hat=True);
```



Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

EXEMPLO: NORMAL COM DADOS SIMULADOS

Fonte: https://docs.pymc.io/pymc-examples/examples/pymc3_howto/api_quickstart.html

[13]:
```
with pm.Model() as model:
    mu = pm.Normal("mu", mu=0, sigma=1) # priori
    obs = pm.Normal("obs", mu=mu, sigma=1, observed=np.random.randn(100)) #␣
↪verossimilhança
```

[14]:
```
model.basic_RVs
```

[14]: [mu, obs]

[15]:
```
model.free_RVs
```

[15]: [mu]

[16]:
```
model.observed_RVs
```

[16]: [obs]

VARIÁVEIS ALEATÓRIAS NÃO OBSERVÁVEIS

[17]:
```
with pm.Model():
    x = pm.Normal("x", mu=0, sigma=1)
```

VARIÁVEIS ALEATÓRIAS OBSERVÁVEIS

[18]:
```
with pm.Model():
    obs = pm.Normal("x", mu=0, sigma=1, observed=np.random.randn(100))
```

INFERÊNCIA

AMOSTRAGEM

[19]:
```
with pm.Model() as model:
    mu = pm.Normal("mu", mu=0, sigma=1)
    obs = pm.Normal("obs", mu=mu, sigma=1, observed=np.random.randn(100))

    idata = pm.sample(2000, tune=1500, return_inferencedata=True)
```

```
<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>
```

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

<IPython.core.display.HTML object>

[20]:
```
idata.posterior.dims
```

[20]:
```
Frozen({'chain': 2, 'draw': 2000})
```

AMOSTRAGEM COM 6 CADEIAS EM PARALELO

[21]:
```
with pm.Model() as model:
    mu = pm.Normal("mu", mu=0, sigma=1)
    obs = pm.Normal("obs", mu=mu, sigma=1, observed=np.random.randn(100))

    idata = pm.sample(cores=4, chains=6, return_inferencedata=True)
```

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

[22]:
```
idata.posterior["mu"].shape
```

[22]:
```
(6, 1000)
```

PODEMOS INCLUIR PASSOS COM OUTROS MÉTODOS

[23]:
```
with pm.Model() as model:
    mu = pm.Normal("mu", mu=0, sigma=1)
    obs = pm.Normal("obs", mu=mu, sigma=1, observed=np.random.randn(100))

    step = pm.Metropolis()
    trace = pm.sample(1000, step=step)
```

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

[24]:
```
with pm.Model() as model:
    mu = pm.Normal("mu", mu=0, sigma=1)
    sd = pm.HalfNormal("sd", sigma=1)
    obs = pm.Normal("obs", mu=mu, sigma=sd, observed=np.random.randn(100)) #verossimilhança

    step1 = pm.Metropolis(vars=[mu])
```

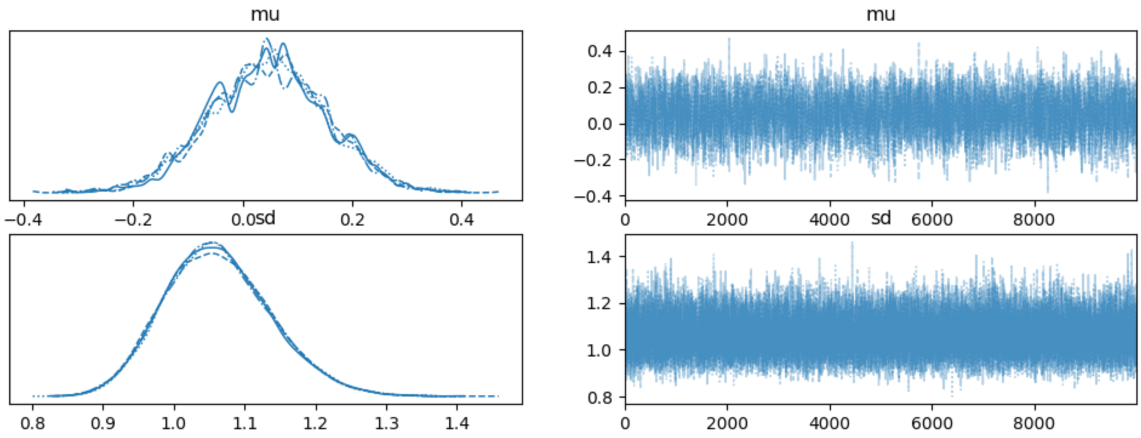
Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.

```
    step2 = pm.Slice(vars=[sd])
    idata = pm.sample(10000, step=[step1, step2], cores=4,␣
↪return_inferencedata=True)
```

```
<IPython.core.display.HTML object>

<IPython.core.display.HTML object>
```

ANÁLISE DE RESULTADOS

[25]:
```
az.plot_trace(idata);
```



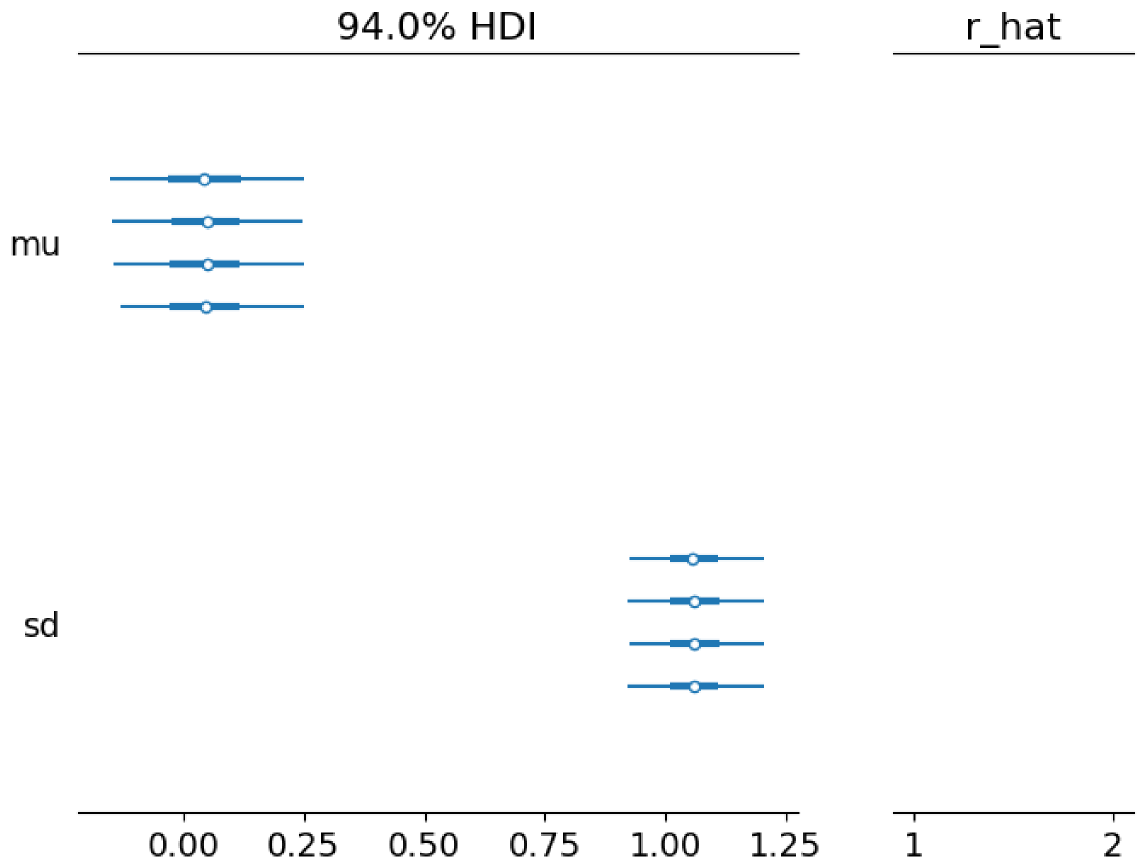
ESTATÍSTICA DE GELMAN-RUBIN (R CHAPÉU)

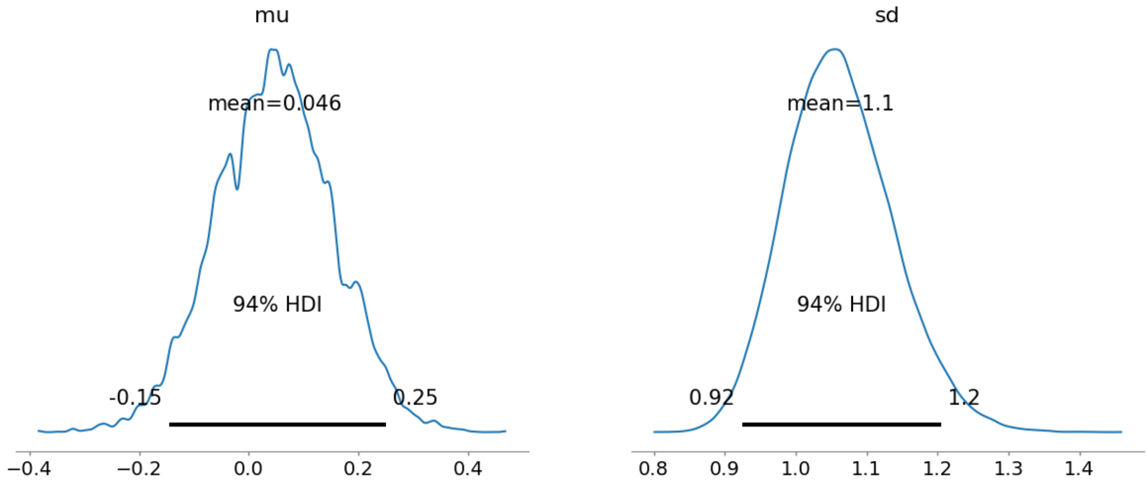
[26]:
```
az.summary(idata)
```

[26]:
```
     mean     sd  hdi_3%  hdi_97%  mcse_mean  mcse_sd  ess_bulk  ess_tail  \
mu  0.046  0.106  -0.146    0.250      0.001    0.001    5906.0    6753.0
sd  1.063  0.075   0.924    1.205      0.000    0.000   38451.0   29012.0

    r_hat
mu    1.0
sd    1.0
```

[27]:
```
az.plot_forest(idata, r_hat=True);


az.plot_posterior(idata);
```

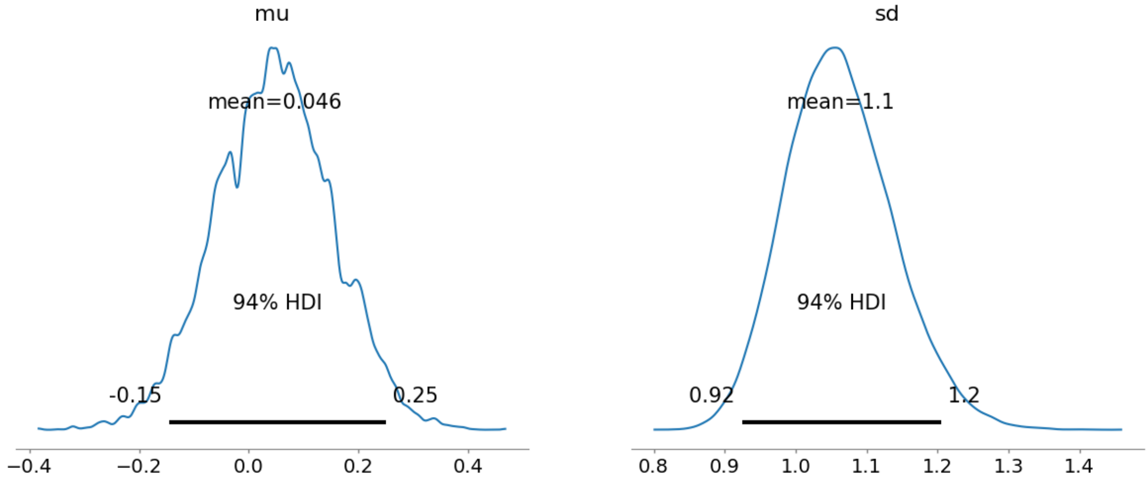
Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.





[28]:
```
az.plot_posterior(idata);
```

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.



[28]:

Copyright © 2024. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização.