Bài 5 - Thống kê Bayes tính toán (Computational Bayesian Statistics)

Thống kê máy tính và ứng dụng (CLC)

Vũ Quốc Hoàng (vqhoang@fit.hcmus.edu.vn)
FIT - HCMUS

Ngày 14 tháng 3 năm 2022

Nội dung

Nội dung

Công thức Bayes

Trên không gian mẫu Ω, các biến cố $\{E_1, E_2, ..., E_n\}$ được gọi là một họ đầy đủ nếu

- $E_i \cap E_j = \varnothing$ khi $i \neq j$ (mutually exclusive, loại trừ, "không trùng"),
- $\Omega = \bigcup_{i=1}^{n} E_i$ (exhaust all possibilities, đầy đủ, "không sót").

Công thức Bayes (Bayes' rule)

$$\underbrace{P(E_i|D)}_{\text{posterior}} = \frac{\overbrace{P(D|E_i)}^{\text{likelihood}}\overbrace{P(E_i)}^{\text{prior}}}{\underbrace{P(D)}_{\text{evidence}}} = \frac{P(D|E_i)P(E_i)}{\sum_{j=1}^{n} P(D|E_j)P(E_j)} \propto P(D|E_i)P(E_i).$$

Công thức Bayes hướng dẫn cách "cập nhật xác suất" hay "phân bổ lại niềm tin" (reallocation of **credibility**).

Suy diễn Bayes

Các **dữ liệu** (data) được sinh ra từ một **mô hình xác suất** (probabilistic model). Mô hình được xác định bởi các **tham số** (parameter). Từ các dữ liệu quan sát, **suy diễn** (inference) về mô hình qua tham số. **Suy diễn Bayes** (Bayesian inference)

1. Xác định dữ liệu, mô hình và hàm hợp lý của tham số,
2. "Chọn" phân phối tiên nghiệm phù hợp cho tham số,
3. "Tính" phân phối hậu nghiệm của tham số theo công thức Bayes

$$\underbrace{p(\theta|D)}_{\text{posterior}} = \frac{p(D|\theta)p(\theta)}{p(D)} = \frac{\overbrace{L(\theta|D)}^{\text{likelihood}}\overbrace{p(\theta)}^{\text{prior}}}{\underbrace{p(D)}_{\text{evidence}}} \propto L(\theta|D)p(\theta),$$

4. Kiểm tra phân phối hậu nghiệm có sinh dữ liệu phù hợp ("posterior predictive check"). Nếu không, thử mô hình và/hoặc dữ liệu khác,
5. "Dùng" phân phối hậu nghiệm để đưa ra các kết luận.

Suy diễn Bayes - Ví dụ 1

Bài toán. Một nhà máy sản xuất bóng với 4 loại kích thước 1, 2, 3, 4. Đặt 3 quả bóng loại kích thước 2. Nhận được 3 quả bóng với các kích thước: 1.77, 2.23, 2.70. Hỏi nhà máy có sản xuất đúng loại đã đặt?

Suy diễn. Ta mô hình kích thước bóng là biến ngẫu nhiên $X \sim \mathcal{N}(\mu, \sigma^2)$ với tham số μ có thể nhận một trong 4 giá trị

$$\begin{cases} E_1 : \mu = \mu_1 = 1.0, \\ E_2 : \mu = \mu_2 = 2.0, \\ E_3 : \mu = \mu_3 = 3.0, \\ E_4 : \mu = \mu_4 = 4.0. \end{cases}$$

Ta có

$$p(X = x|E_i) = f_{\mathcal{N}(\mu_i,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu_i)^2}{2\sigma^2}}, i = 1, ..., 4.$$

Suy diễn Bayes - Ví dụ 1 (tt)

Ta nhận được X_1, X_2, X_3 độc lập và cùng phân phối với X. Cụ thể, ta có dữ liệu

$$D = (X_1 = 1.77) \cap (X_2 = 2.23) \cap (X_3 = 2.70).$$

Hàm hợp lý của tham số μ trên dữ liệu D theo mô hình đã chọn là

$$\begin{aligned} L(\mu_i|D) &= p(D|E_i) = p\left((X_1 = 1.77) \cap (X_2 = 2.23) \cap (X_3 = 2.70)|E_i\right) \\ &= p(X_1 = 1.77|E_i)p(X_2 = 2.23|E_i)p(X_3 = 2.70|E_i) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^3 e^{\frac{-\left((1.77-\mu_i)^2+(2.23-\mu_i)^2+(2.70-\mu_i)^2\right)}{2\sigma^2}} \\ &\propto e^{\frac{-\left((1.77-\mu_i)^2+(2.23-\mu_i)^2+(2.70-\mu_i)^2\right)}{2\sigma^2}}, i = 1, ..., 4. \end{aligned}$$

"Giả sử" phân phối tiên nghiệm là

$$p(E_i) = \frac{1}{4} \propto 1, i = 1, ..., 4.$$

Suy diễn Bayes - Ví dụ 1 (tt)

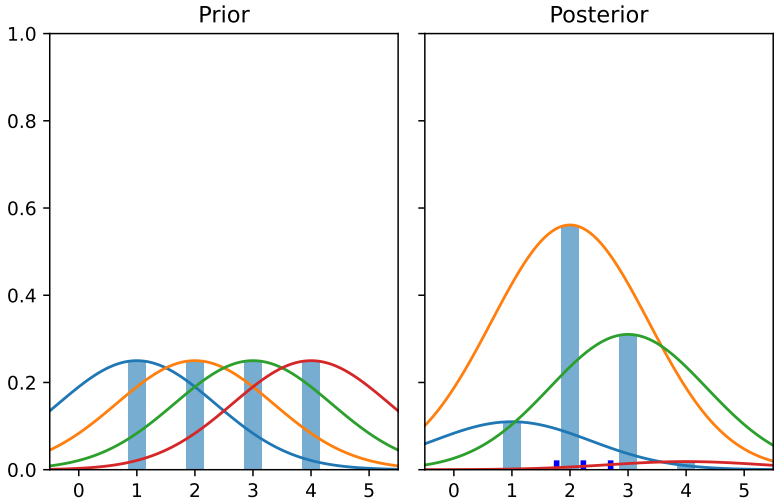
Ta tính phân phối hậu nghiệm theo công thức Bayes

$$p(E_i|D) = \frac{p(D|E_i)p(E_i)}{\sum_{j=1}^{4} p(E_j|D)p(E_j)} = \frac{e^{\frac{-\left((1.77-\mu_i)^2+(2.23-\mu_i)^2+(2.70-\mu_i)^2\right)}{2\sigma^2}}}{\sum_{j=1}^{4} e^{\frac{-\left((1.77-\mu_j)^2+(2.23-\mu_j)^2+(2.70-\mu_j)^2\right)}{2\sigma^2}}}, i = 1, ..., 4.$$

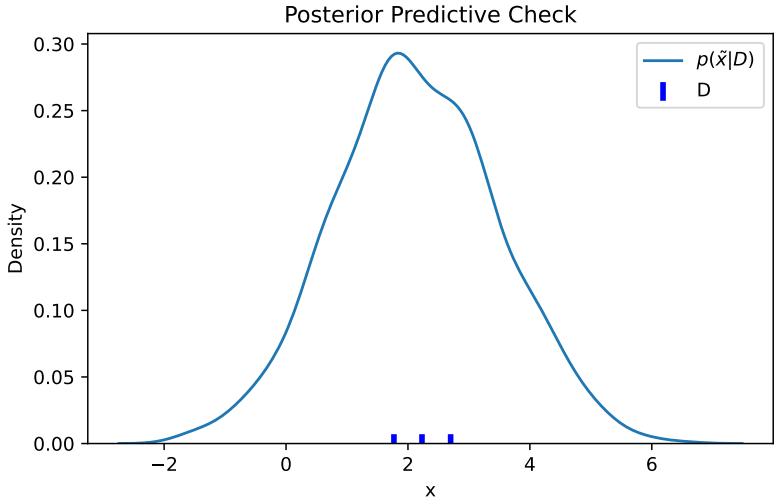
Với một số giá trị của σ^2, tính toán ta có

	$\sigma^2 = 1$	$\sigma^2 = 1.35$	$\sigma^2 = 2$
$P(\mu = 1\|D)$	7%	11%	16%
$P(\mu = 2\|D)$	64%	56%	47%
$P(\mu = 3\|D)$	29%	31%	32%
$P(\mu = 4\|D)$	1%	2%	5%

Suy diễn Bayes - Ví dụ 1 (tt)



Suy diễn Bayes - Ví dụ 1 (tt)



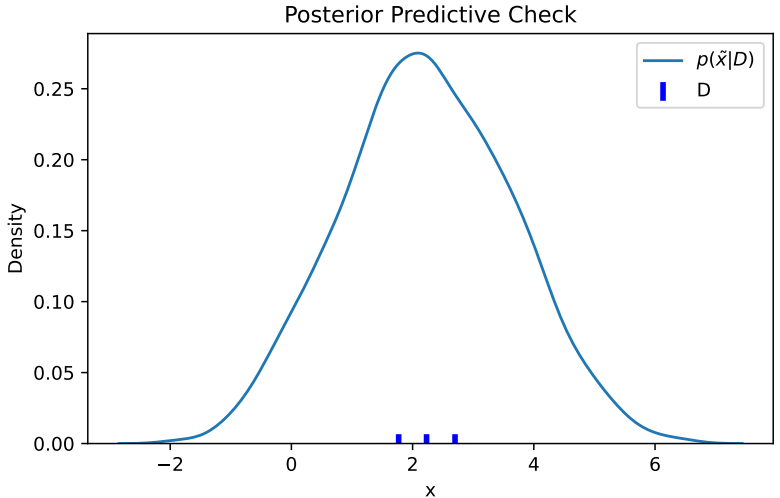
Suy diễn Bayes - Ví dụ 1 - PyMC3

```
import pymc3 as pm
import theano

mu = np.array([1.0, 2.0, 3.0, 4.0])
prior = 1/4 * np.ones(len(mu))
x = np.array([1.77, 2.23, 2.70])
sigma2 = 1.35

with pm.Model() as model:
    mu_index_var = pm.Categorical("mu_index", p=prior)
    mu_var = theano.shared(mu)[mu_index_var]
    x_var = pm.Normal("x", mu=mu_var, tau=1/sigma2,
       observed=x)
    trace = pm.sample(1000, return_inferencedata=False)
```

Suy diễn Bayes - Ví dụ 1 - PyMC3 (tt)

```
mu_index_posterior = trace["mu_index"]
values, counts = np.unique(mu_index_posterior,
   return_counts=True)
print(mu[values])
# [1. 2. 3. 4.]
print(counts/np.sum(counts))
# [0.1225 0.5425 0.3135 0.0215]

x_post_sample = pm.sample_posterior_predictive(trace, 1000,
    model)["x"].flatten()
sns.kdeplot(x_post_sample)
plt.scatter(x, np.zeros(len(x)))
```

Suy diễn Bayes - Ví dụ 1 - PyMC3 (tt)

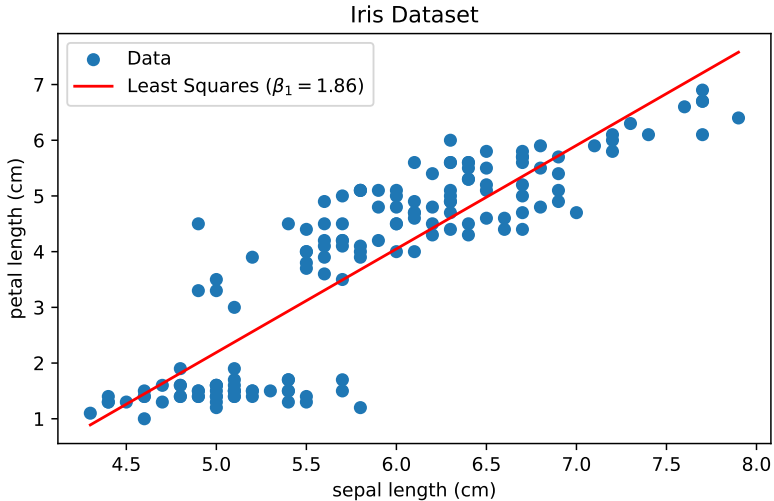


Suy diễn Bayes - Ví dụ 2

Iris Dataset

- Wikipedia: `https://en.wikipedia.org/wiki/Iris_flower_data_set`
- UCI: `https://archive.ics.uci.edu/ml/datasets/iris`
- Scikit-learn: `https://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html`

Suy diễn Bayes - Ví dụ 2 (tt)



Suy diễn Bayes - Ví dụ 2 (tt)

Bài toán. Từ dữ liệu Iris, xác định sự phụ thuộc của petal-length vào sepal-length.

Suy diễn. "Từ gợi ý của dữ liệu", ta mô hình sự phụ thuộc của petal-length (y) vào sepal-length (x) như sau

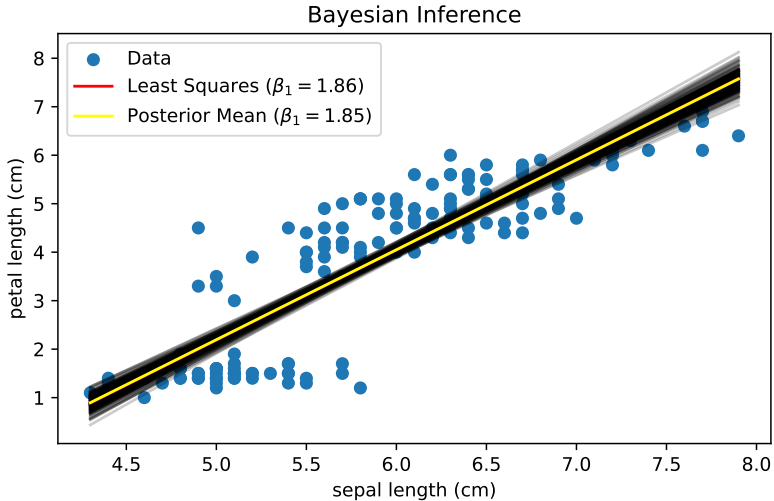
$$\begin{aligned}
\hat{y} &= \beta_0 + \beta_1 x, \\
y &\sim \mathcal{N}(\hat{y}, \sigma), \\
\beta_0 &\sim \mathcal{N}(0, 10^2), \\
\beta_1 &\sim \mathcal{N}(0, 10^2), \\
\sigma &\sim \mathcal{U}(0, 1000).
\end{aligned}$$

Dùng PyMC3, "tính" phân phối hậu nghiệm, ta có các kết quả sau.

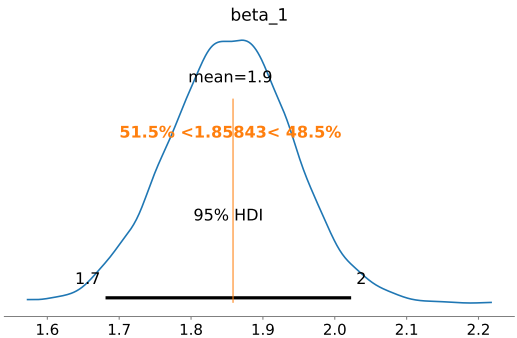
Suy diễn Bayes - Ví dụ 2 (tt)

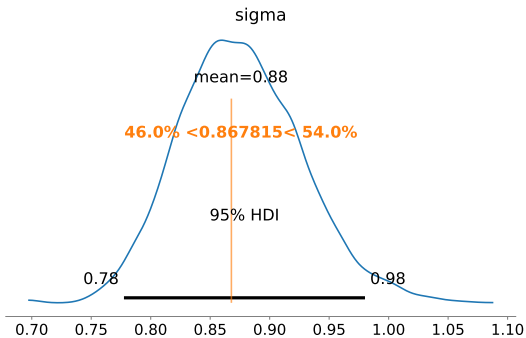
```
with pm.Model() as model:
    beta_0 = pm.Normal("beta_0", mu=0, sigma=10)
    beta_1 = pm.Normal("beta_1", mu=0, sigma=10)
    sigma = pm.Uniform("sigma", lower=0, upper=1000)
    y_hat = beta_0 + beta_1*x
    y_var = pm.Normal("y", mu=y_hat, sigma=sigma,
                      observed=y)
    trace = pm.sample(5000, return_inferencedata=False)
```

Suy diễn Bayes - Ví dụ 2 (tt)

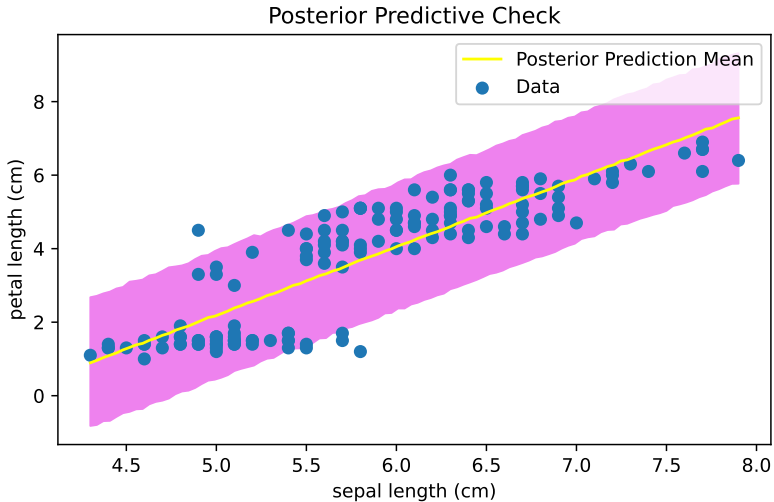


Suy diễn Bayes - Ví dụ 2 (tt)





Suy diễn Bayes - Ví dụ 2 (tt)



Nội dung

Mô hình nhị thức

Biến ngẫu nhiên rời rạc Y được gọi là kết quả của một **phép thử Bernoulli** (Bernoulli trial) hay có **phân phối Bernoulli** (Bernoulli distribution) với tham số p ($0 \leq p \leq 1$), kí hiệu $Y \sim \text{Bernoulli}(p)$, nếu

$$\begin{cases} P(Y = 1) = p, \\ P(Y = 0) = 1 - p. \end{cases}$$

($Y = 1$) thường được gọi là **"thành công"** (success), ($Y = 0$) là **"thất bại"** (failure), p là xác suất thành công.

Dãy biến ngẫu nhiên $Y_1, Y_2, \ldots$ được gọi là một dãy phép thử Bernoulli hay một **quá trình Bernoulli** (Bernoulli process) nếu $Y_1, Y_2, \ldots$ độc lập và cùng phân phối Bernoulli(p).

Mô hình nhị thức (tt)

Bài toán. Cho dữ liệu của dãy phép thử Bernoulli $D = \{Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n\}$, "tìm" p.

Suy diễn. Xem tham số $\theta = p$ là biến ngẫu nhiên liên tục, nhận giá trị trong $[0, 1]$. Với $Y \sim \text{Bernoulli}(\theta)$ có giá trị quan sát $y \in \{0, 1\}$, hàm hợp lý của θ theo y là

$$p(y|\theta) = \theta^y(1-\theta)^{1-y} = \begin{cases} \theta & y = 1, \\ 1-\theta & y = 0. \end{cases}$$

Do đó, với dữ liệu $D = \{y_i\}_{i=1}^n$, hàm hợp lý của θ theo D là

$$p(D|\theta) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} = \theta^{\sum_{i=1}^n y_i}(1-\theta)^{\sum_{i=1}^n (1-y_i)} = \theta^z(1-\theta)^{n-z},$$

với $z = \sum_{i=1}^n y_i$ là số lần thành công trong n lần quan sát của dữ liệu D.

Mô hình nhị thức (tt)

"Chọn" phân phối tiên nghiệm cho θ là **phân phối Beta** (Beta distribution) với tham số a, b $(a > 0, b > 0)$, $\theta \sim \text{Beta}(a, b)$, ta có

$$p(\theta; a, b) \propto \theta^{a-1}(1-\theta)^{b-1}, 0 \leq \theta \leq 1.$$

(Xem thêm phân phối Beta `https://en.wikipedia.org/wiki/Beta_distribution`.)

Từ công thức Bayes, θ có phân phối hậu nghiệm sau khi quan sát dữ liệu D là

$$p(\theta|D) \propto p(D|\theta)p(\theta; a, b) \propto \theta^{z}(1-\theta)^{n-z}\theta^{a-1}(1-\theta)^{b-1} \propto \theta^{z+a-1}(1-\theta)^{n-z+b-1}.$$

Nhận xét, $(\theta|D) \sim \text{Beta}(a+z, b+n-z)$.
Do đó, phân phối Beta được gọi là **phân phối tiên nghiệm liên hợp** (conjugate prior distribution) của phân phối Bernoulli.

Mô hình nhị thức (tt)

Lưu ý, các **"siêu tham số"** (hyperparameter) a, b của phân phối tiên nghiệm và $\kappa = a + b$ thường được gọi là **pseudo-count** vì

- Trước khi quan sát dữ liệu D, $\theta \sim \text{Beta}(a, b)$, ta tin vào xác suất thành công θ như thể ta đã thấy a lần thành công và b lần thất bại (trong tổng số $\kappa = a + b$ lần),
- Dữ liệu D cho thấy z lần thành công và $n - z$ lần thất bại (trong tổng số n lần),
- Sau khi quan sát dữ liệu D, $(\theta|D) \sim \text{Beta}(a + z, b + n - z)$, ta tin vào xác suất thành công như thể ta đã thấy $a + z$ lần thành công và $b + n - z$ lần thất bại (trong tổng số $\kappa + n$ lần).

Mô hình nhị thức (tt)

Như vậy, có thể nói phân phối hậu nghiệm là "tổng hợp" hay "thỏa hiệp" giữa phân phối tiên nghiệm và hàm hợp lý trên dữ liệu. Chẳng hạn

- Phân phối tiên nghiệm $\theta \sim \text{Beta}(a, b)$ có kỳ vọng

$$E(\theta) = \frac{a}{a+b} = \frac{a}{\kappa},$$

- Hàm hợp lý từ dữ liệu $L(\theta|D) = p(D|\theta) = \theta^z(1-\theta)^{n-z}$ đạt cực đại tại

$$\hat{\theta}_{\text{MLE}} = \frac{z}{n},$$

- Phân phối hậu nghiệm $(\theta|D) \sim \text{Beta}(a+z, b+n-z)$ có kỳ vọng

$$\begin{aligned} E(\theta|D) &= \frac{a+z}{a+z+b+n-z} = \frac{a+z}{\kappa+n} = \frac{\kappa}{\kappa+n}\frac{a}{\kappa} + \frac{n}{\kappa+n}\frac{z}{n} \\ &= \frac{\kappa}{\kappa+n}E(\theta) + \frac{n}{\kappa+n}\hat{\theta}_{\text{MLE}}. \end{aligned}$$

Mô hình nhị thức - Ví dụ

- Dữ liệu: $z = 1, n = 10$ (quan sát thấy 1 lần thành công trong 10 lần), hàm hợp lý

$$p(D|\theta) = \theta^z(1-\theta)^{n-z} = \theta^1(1-\theta)^9$$

đạt cực đại tại

$$\hat{\theta}_{\text{MLE}} = \frac{z}{n} = \frac{1}{10} = 0.1$$

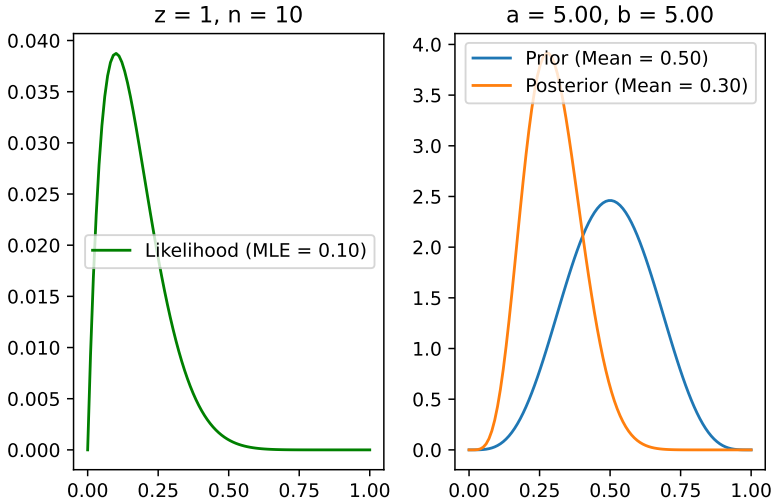
- Phân phối tiên nghiệm: $a = 5, b = 5, \kappa = a + b = 10$ (như thể đã quan sát thấy 5 lần thành công trong 10 lần), $\theta \sim \text{Beta}(a, b) = \text{Beta}(5, 5)$ có

$$E(\theta) = \frac{a}{\kappa} = \frac{5}{10} = 0.5.$$

- Phân phối hậu nghiệm: $a + z = 6, \kappa + n = 20$ (như thể đã quan sát thấy 6 lần thành công trong 20 lần), $(\theta|D) \sim \text{Beta}(a + z, b + n - z) = \text{Beta}(6, 14)$ có

$$E(\theta|D) = \frac{a+z}{\kappa+n} = \frac{6}{20} = 0.3 = \frac{\kappa}{\kappa+n}E(\theta) + \frac{n}{\kappa+n}\hat{\theta}_{\text{MLE}} = 0.5E(\theta) + 0.5\hat{\theta}_{\text{MLE}}.$$

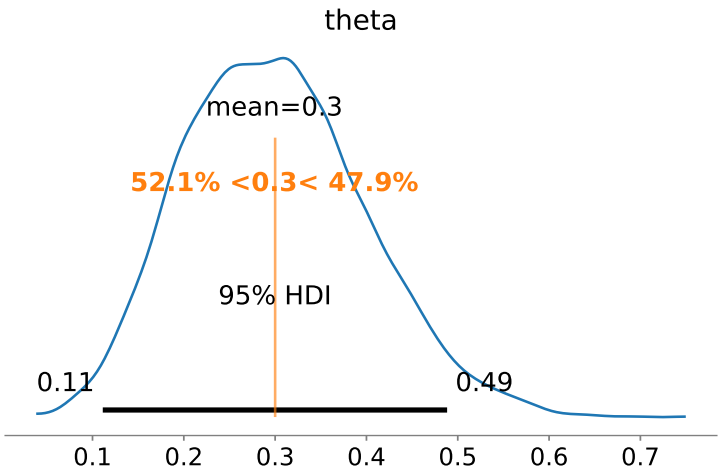
Mô hình nhị thức - Ví dụ (tt)



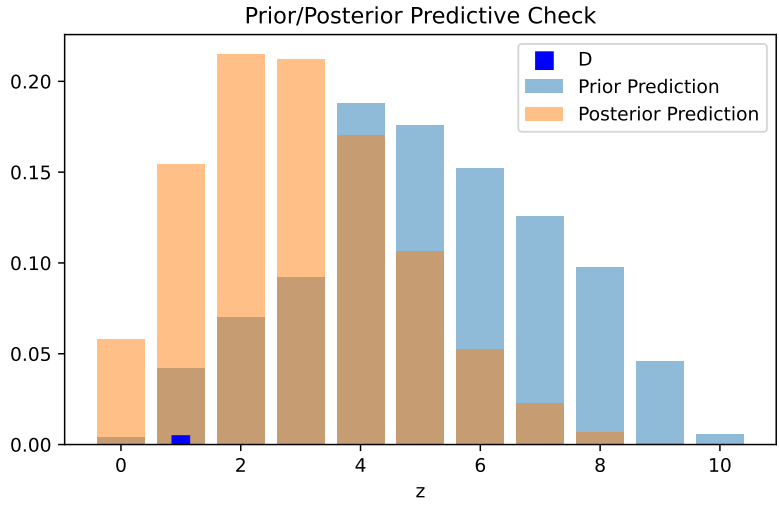
Mô hình nhị thức - Ví dụ - PyMC3

```
with pm.Model() as model:
    theta = pm.Beta("theta", alpha=a, beta=b)
    z_var = pm.Binomial("z", p=theta, n=n, observed=z)
    trace = pm.sample(5000, return_inferencedata=False)
    prior_pred = pm.sample_prior_predictive()
    posterior_pred = pm.sample_posterior_predictive(trace)

    az.plot_posterior(trace, var_names=["theta"],
            hdi_prob=0.95, ref_val=posterior_dist.mean())
```

Mô hình nhị thức - Ví dụ - PyMC3 (tt)



Mô hình nhị thức - Ví dụ - PyMC3 (tt)



Tài liệu tham khảo

Chapter 2, 6. John K. Kruschke. *Doing Bayesian Data Analysis – A Tutorial with R, JAGS, and Stan*. Elsevier, 2015.