Math 170S

Lecture Notes Section 6.8 [*][†]

Bayesian estimate

Instructor: Swee Hong Chan

NOTE: The notes is a summary for materials discussed in the class and is not supposed to substitute the textbook. Please send me an email if you find typos.

---

1 Conditional probability: Review

Let A and B be two events. The **conditional probability** of A *given that B has already happened* is

$$P[A \mid B] := \frac{P(A \cap B)}{P[B]}.$$

2 Problem: coin toss

The friendly instructor has three coins in his drawer, with success probability 0.2, 0.5, and 0.8 respectively. He picks one coin uniformly at random, and flips the coin twice. Compute the probability that both tosses are head.

Answer. Let Θ be the success probability of the chosen coin. Let X_1 and X_2 be the result of the first and second coin toss, respectively.

Then X_1 and X_2 are Bernoulli random variables with success probability Θ, and

$$P[\Theta = 0.2] = \frac{1}{3}; \qquad P[\Theta = 0.5] = \frac{1}{3}; \qquad P[\Theta = 0.8] = \frac{1}{3}.$$

Note that

$$P[X_1, X_2 = 1 \mid \Theta = 0.2] = (0.2)^2;$$

$$P[X_1, X_2 = 1 \mid \Theta = 0.5] = (0.5)^2;$$

$$P[X_1, X_2 = 1 \mid \Theta = 0.8] = (0.8)^2.$$

So we have (BT)

$$
\begin{aligned}
P[X_1, X_2 = 1] &= P[X_1, X_2 = 1 \mid \Theta = 0.2]\, P[\Theta = 0.2] + \\
&\quad P[X_1, X_2 = 1 \mid \Theta = 0.5]\, P[\Theta = 0.5] + \\
&\quad P[X_1, X_2 = 1 \mid \Theta = 0.8]\, P[\Theta = 0.8] \\
&= (0.2)^2 \left(\frac{1}{3}\right) + (0.5)^2 \left(\frac{1}{3}\right) + (0.8)^2 \left(\frac{1}{3}\right) \\
&= 0.31,
\end{aligned}
$$

as desired. $\qquad\square$

3 Bayes theorem: Review

Theorem 1. *For any event A and B,*

$$P[A \mid B] = \frac{P[B \mid A]P[A]}{P[B]}.$$

$\square$

4 Problem: Coin toss again

Compute the probability

$$P[\Theta = 0.2 \mid X_1, X_2 = 1]; \qquad P[\Theta = 0.5 \mid X_1, X_2 = 1];$$

$$P[\Theta = 0.8 \mid X_1, X_2 = 1].$$

Answer: By Bayes theorem (BT)

$$P[\Theta = 0.2 \mid X_1, X_2 = 1] = \frac{P[X_1, X_2 = 1 \mid \Theta = 0.2]P[\Theta = 0.2]}{P[X_1, X_2 = 1]}$$

$$= \frac{(0.2)^2(1/3)}{0.31} \approx 0.044;$$

$$P[\Theta = 0.5 \mid X_1, X_2 = 1] = \frac{P[X_1, X_2 = 1 \mid \Theta = 0.5]P[\Theta = 0.5]}{P[X_1, X_2 = 1]}$$

$$= \frac{(0.5)^2(1/3)}{0.31} \approx 0.268;$$

$$P[\Theta = 0.8 \mid X_1, X_2 = 1] = \frac{P[X_1, X_2 = 1 \mid \Theta = 0.8]P[\Theta = 0.8]}{P[X_1, X_2 = 1]}$$

$$= \frac{(0.8)^2(1/3)}{0.31} \approx 0.688.$$

5 Bayesian inference: COVID-19

Consider the forecasting models for COVID-19.

- On March 2020, one model suggested that the total fatalities in US might reach up to 240,000 people.

- However, on April 2020, as more data is collected, the estimates (by using the same model) was revised to up to 85,000 fatalities.

This is one instance where we revise our estimate to give the best prediction, based on the most up-to-date information. This revision is done by Bayesian inference.

6 Bayesian inference (discrete)

1. Let X be a random variable with distribution f_θ for some parameter Θ.

2. Θ is a **discrete** random variable on Ω with an unknown pmf π.

3. **Problem:** Estimate the unknown pmf π.

4. **Input:**

   - Sample values $x_1, \ldots, x_n$ from n experiments.

   - A **prior pmf** π_{prior} which we think is the best estimate for π **before** we run the experiments.

5. **Output:** A **posterior pmf** π_{post} that we think is the best estimate for π **after** we observe the experiments.

6. **Method:**

(a) Compute the quantity

$$K \; := \; \sum_{\theta \in \Omega} f_\theta(x_1) \ldots f_\theta(x_n) \pi_{prior}[\Theta = \theta].$$

(b) Compute the posterior pmf π_{post} by

$$\pi_{post}[\Theta = \theta] \; = \; \frac{f_\theta(x_1) \ldots f_\theta(x_n) \, \pi_{prior}[\Theta = \theta]}{K}.$$

(c) That is it, you are done!

7   Example: coin toss again

The friendly instructor has three coins, with success probability 0.2, 0.5, and 0.8. He picks one coin at random following an unknown pmf. In the language of Bayesian inference,

- X is a Bernoulli random variable with success probability Θ;

- Θ is randomly picked from the set $\{0.2, 0.5, 0.8\}$ following some unknown pmf π.

- Since we have no information regarding π, our best guess would be π_{prior} is the uniform distribution,

$$\pi_{prior}[\Theta = 0.2] = \pi_{prior}[\Theta = 0.5] = \pi_{prior}[\Theta = 0.8] = \frac{1}{3}.$$

Now the chosen coin is flipped twice, and both outcomes

are equal to head, so $x_1 = x_2 = 1$.

After observing these experiments, we update our prediction on π by the given method and compute K: (BT)

$$K = \sum_{\theta \in \Omega} f_\theta(x_1) \ldots f_\theta(x_n) \pi_{prior}[\Theta = \theta]$$

$$= f_{0.2}(1) f_{0.2}(1) \pi_{prior}[\Theta = 0.2] \; + \; f_{0.5}(1) f_{0.5}(1) \pi_{prior}[\Theta = 0.5]$$

$$= (0.2)(0.2)(1/3) \; + \; (0.5)(0.5)(1/3) \; + \; (0.8)(0.8)(1/3)$$

$$= 0.31.$$

The posterior pmf is then given by

$$\pi_{post}[\Theta = 0.2] = \frac{f_\theta(x_1) \ldots f_\theta(x_n)\, \pi_{prior}[\Theta = \theta]}{K}$$

$$= \frac{f_{0.2}(1) f_{0.2}(1)\, \pi_{prior}[\Theta = 0.2]}{K}$$

$$= \frac{(0.2)(0.2)(1/3)}{0.31} \approx 0.044;$$

$$\pi_{post}[\Theta = 0.5] = \frac{f_{0.5}(x_1) f_{0.5}(x_2)\, \pi_{prior}[\Theta = 0.5]}{K}$$

$$= \frac{(0.5)(0.5)(1/3)}{0.31} \approx 0.268;$$

$$\pi_{post}[\Theta = 0.8] = \frac{f_{0.8}(1) f_{0.8}(1)\, \pi_{prior}[\Theta = 0.8]}{K}$$

$$= \frac{(0.8)(0.8)(1/3)}{0.31} \approx 0.688;$$

This is our posterior pmf.

8 Bayesian inference (continuous)

Bayesian inference for the continuous case works the same way with discrete case, except that sum in the formula for K is replaced with integrals.

$$K := \int_{-\infty}^{\infty} f_\theta(x_1) \ldots f_\theta(x_n) \pi_{prior}[\theta] \, d\theta.$$

Remark 2. Some of our notations here are different from the textbook:

- The random variable X with distribution f_θ in our notes is the random variable Y with distribution $g(\cdot \mid \theta)$ in the textbook;

- The prior pdf π_{prior} in our notes is $h(\theta)$ in the textbook;

- The posterior pdf π_{post} in our notes is $k(\theta)$ in the textbook;

- The quantity K in our notes is the quantity $k_1(y)$ in the textbook.

9 Problem: Binomial and beta

Let X be binomial distribution with parameters n and θ,

$$f_\theta(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad x = 0, 1, \ldots, n.$$

Suppose that π_{prior} is the beta pdf with parameter α, β,

$$\pi_{prior}[\theta] = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \qquad 0 < \theta < 1,$$

where Γ is the gamma function. Suppose that we have performed one experiment with outcome equal to x. Compute π_{post}.

Answer: We first compute K, (BT)

$$K = \int_{-\infty}^{\infty} f_\theta(x_1) \dots f_\theta(x_n) \pi_{prior}[\theta] \, d\theta$$

$$= \int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \, d\theta$$

$$= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \theta^x (1-\theta)^{n-x} \theta^{\alpha-1}(1-\theta)^{\beta-1} \, d\theta$$

$$= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \, d\theta$$

$$= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(n+\alpha+\beta)}$$

$$\int_0^1 \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)} \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \, d\theta.$$

The term inside the integral is exactly the pdf of the beta random variable with parameter $\alpha + x$ and $n + \beta - \alpha$, so this integral is equal to 1, and

$$K = \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(n+\alpha+\beta)}.$$

Therefore, the posterior distribution is

$$
\begin{aligned}
\pi_{post}[\theta] &= \frac{f_\theta(x_1)\ldots f_\theta(x_n)\pi_{prior}[\theta]}{K} \\
&= \frac{1}{K}\binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&= \frac{1}{K}\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^x(1-\theta)^{n-x}\,\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1},
\end{aligned}
$$

which is the pdf of the beta random variable with parameter $\alpha + x$ and $n + \beta - x$. □

Second solution: The posterior distribution is given by

$$
\begin{aligned}
\pi_{post}[\theta] &= \frac{f_\theta(x_1)\dots f_\theta(x_n)\pi_{prior}[\theta]}{K} \\
&= \frac{1}{K}\binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&= \frac{1}{K}\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \\
&= C\,\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1},
\end{aligned}
$$

where $C := \frac{1}{K}\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$. so π_{post} is equal to a **constant multiple** of the pdf beta$(x+\alpha, n+\beta-x)$.

Key observation:

- π_{post} is a **constant multiple** of the pdf beta($x+\alpha$, $n + \beta - x$);

- π_{post} is a **pdf**;

- Thus π_{post} is the pdf beta($x + \alpha$, $n + \beta - x$).

Hence we conclude that

$$\pi_{post}[\theta] = \frac{\Gamma(n + \alpha + \beta)}{\Gamma(\alpha + x)\Gamma(n + \beta - x)}\theta^{x+\alpha-1}(1 - \theta)^{n-x+\beta-1}.$$

$\square$

10 Bayes estimator

In the scenario of Bayes inference, the estimate for the unknown parameter Θ is not a *fixed number*, but a *random variable.* However, there are situations in real life where we are asked to give a fixed number $\widehat{\theta}$ as our estimate. Then **Bayesian estimator** $\widehat{\theta}$ would depend on the penalty for errors created by incorrect guesses:

1. The loss function is $(\theta - \widehat{\theta})^2$, the **square of the error**. Then best guess $\widehat{\theta}$ would be the mean of the posterior pdf.

2. The loss function is $|\theta - \widehat{\theta}|$, the **absolute value of the error**. Then best guess $\widehat{\theta}$ would be the median of the posterior pdf.

Remark 3. The quantity $\widehat{\theta}$ in our notes is written as $w(y)$ in the textbook.

11 Bayesian estimator: example

Let X be the binomial random variable with parameters n and θ. Let π_{prior} be the beta pdf with parameters α and β. Suppose that we have one sample with value x. Compute the Bayesian estimator $\widehat{\theta}$ that minimizes

- square of the error;

- absolute value error, if $\alpha + x = n + \beta - x = 1$.

Answer: We have previously calculated that π_{post} is the pdf for beta($\alpha + x$, $n + \beta - x$). The Bayesian estimator for the mean square error is then the mean of beta($\alpha + x$, $n + \beta - x$),

$$\widehat{\theta} = \frac{(\alpha + x)}{(\alpha + x) + (n + \beta - x)}.$$

The Bayesian estimator for the absolute value error is the median of beta($\alpha + x$, $n + \beta - x$) = beta(1,1), and is thus equal to $\frac{1}{2}$[1]. $\square$

---

[1]The median of beta(α, β) is equal to $\frac{1}{2}$ if $\alpha = \beta$.