Sequential Bayesian Approach and Prediction

Aleksandar Dogandžić

2017-03-11

Contents

READING: [Gelman et al. 2014, §1.3, 2.1, 2.5].

Sequential Bayesian Approach

SUPPOSE that we have observed x_1 and x_2, where x_1 comes first.[1] We wish to infer about θ. Conditioning on $X_1 = x_1$ yields

$$f_{X_2,\Theta \mid X_1}(x_2, \theta \mid x_1) = f_{X_2 \mid \Theta, X_1}(x_2 \mid \theta, x_1)\, f_{\Theta \mid X_1}(\theta \mid x_1) \tag{1}$$

where

- $f_{X_2 \mid X_1, \Theta}(x_2 \mid x_1, \theta)$ is the new, updated likelihood function of θ based on x_2 and
- $f_{\Theta \mid X_1}(\theta \mid x_1)$ is the new, updated prior for θ.

Now, (1) implies

$$f_{\Theta \mid X_1, X_2}(\theta \mid x_1, x_2) \propto f_{X_2 \mid \Theta, X_1}(x_2 \mid \theta, x_1)\, f_{\Theta \mid X_1}(\theta \mid x_1) \tag{2}$$

which is simply the Bayes' rule with an extra conditioning on X_1.[2]

[1] For example, the subscript denotes a time index.

Put the quantities we know on the right and the quantities we do not know on the left.

[2] We obtain the ordinary Bayes' rule by removing $\mid X_1$ from all terms in (2).

Prediction

SUPPOSE that we have observed $X_1 = x_1$ and wish to predict X_2. For this purpose, we use the *predictive distribution*[3] $f_{X_2 \mid X_1}(x_2 \mid x_1)$. We derive this pdf as follows. Recall (1):

$$f_{X_2,\Theta \mid X_1}(x_2, \theta \mid x_1) = f_{X_2 \mid \Theta, X_1}(x_2 \mid \theta, x_1)\, f_{\Theta \mid X_1}(\theta \mid x_1). \tag{3}$$

[3] say a probability density function (pdf), for simplicity of exposition

Now, marginalize the joint pdf of X_2 and Θ in (3) with respect to the unknown parameter Θ, i.e., integrate Θ out:

$$\begin{aligned} f_{X_2|X_1}(x_2 \mid x_1) &= \int f_{X_2,\Theta|X_1}(x_2, \theta \mid x_1)\,\mathrm{d}\theta \\ &= \int f_{X_2|\Theta,X_1}(x_2 \mid \theta, x_1)\, f_{\Theta|X_1}(\theta \mid x_1)\,\mathrm{d}\theta. \end{aligned} \tag{4}$$

Conditionally independent observations

Consider conditionally independent X_1 and X_2 given $\Theta = \theta$, i.e.,

$$f_{X_1,X_2|\Theta}(x_1, x_2 \mid \theta) = f_{X_1|\Theta}(x_1 \mid \theta)\, f_{X_2|\Theta}(x_2 \mid \theta) \tag{5}$$

or, equivalently,

$$f_{X_2|X_1,\Theta}(x_2 \mid x_1, \theta) = f_{X_2|\Theta}(x_2 \mid \theta). \tag{6}$$

Now, (4) simplifies to

$$f_{X_2|X_1}(x_2 \mid x_1) = \int f_{X_2|\Theta}(x_2 \mid \theta)\, f_{\Theta|X_1}(\theta \mid x_1)\,\mathrm{d}\theta. \tag{7}$$

Suppose that X_1 and X_2 are independent given $\Theta = \theta$, coming from

$$\{X_i \mid \Theta = \theta\}_{i=1}^{2} \sim \mathrm{Bin}(N_i, \theta)$$

i.e., the joint likelihood for measurements x_1 and x_2 is

$$\begin{aligned} p_{X_1,X_2|\Theta}(x_1, x_2 \mid \theta) &= p_{X_1|\Theta}(x_1 \mid \theta)\, p_{X_2|\Theta}(x_2 \mid \theta) \\ &= \binom{N_1}{x_1} \theta^{x_1}(1-\theta)^{N_1-x_1} \binom{N_2}{x_2} \theta^{x_2}(1-\theta)^{N_2-x_2}\, \mathbb{1}_{(0,1)}(\theta). \end{aligned}$$

As we have seen earlier in handout `introBayes`, the class of conjugate prior pdfs for Θ under this data model is

$$\begin{aligned} f_\Theta(\theta) &= \mathrm{Beta}(\theta \mid \alpha, \beta) \\ &\propto \theta^{\alpha-1}(1-\theta)^{\beta-1}\, \mathbb{1}_{(0,1)}(\theta). \end{aligned}$$

✻ The posterior pdf of θ is

$$\begin{aligned} f_{\Theta|X_1,X_2}(\theta \mid x_1, x_2) &\propto p_{X_1,X_2|\Theta}(x_1, x_2 \mid \theta)\, f_\Theta(\theta) \\ &\propto \theta^{x_1+x_2+\alpha-1}(1-\theta)^{N_1+N_2-x_1-x_2+\beta-1}\, \mathbb{1}_{(0,1)}(\theta) \end{aligned}$$

which is the kernel of the $\mathrm{Beta}(x_1 + x_2 + \alpha, \beta + N_1 - x_1 + N_2 - x_2)$ pdf, see the table of distributions. Hence,

$$f_{\Theta|X_1,X_2}(\theta \mid x_1, x_2) = \mathrm{Beta}(\theta \mid x_1 + x_2 + \alpha, \beta + N_1 - x_1 + N_2 - x_2).$$

⁎ THE posterior pdf $f_{\Theta|X_1}(\theta\,|\,x_1)$ given only $X_1 = x_1$. Now,

$$\begin{aligned} f_{\Theta|X_1}(\theta\,|\,x_1) &\propto p_{X_1|\Theta}(x_1\,|\,\theta)\, f_\Theta(\theta) \\ &\propto \theta^{x_1+\alpha-1}(1-\theta)^{N_1-x_1+\beta-1}\, \mathbb{1}_{(0,1)}(\theta) \end{aligned}$$

which is the kernel of the $\mathrm{Beta}(x_1+\alpha, \beta+N_1-x_1)$ pdf; therefore,

$$f_{\Theta|X_1}(\theta\,|\,x_1) = \mathrm{Beta}(\theta\,|\,x_1+\alpha, \beta+N_1-x_1).$$

⁎ PREDICTING X_2 after observing $X_1 = x_1$. Since X_1 and X_2 are independent given $\Theta = \theta$, we apply (7):

$$\begin{aligned} p_{X_2|X_1}(x_2\,|\,x_1) &= \int_0^1 p_{X_2|\Theta}(x_2\,|\,\theta)\, f_{\Theta|X_1}(\theta\,|\,x_1)\, \mathrm{d}\theta \\ &= \binom{N_2}{x_2} \frac{\Gamma(\alpha+\beta+N_1)}{\Gamma(x_1+\alpha)\Gamma(\beta+N_1-x_1)} \\ &\quad \cdot \underbrace{\int_0^1 \theta^{x_1+x_2+\alpha-1}(1-\theta)^{\beta+N_1-x_1+N_2-x_2-1}\, \mathrm{d}\theta}_{c} \\ &= \binom{N_2}{x_2} \frac{\Gamma(\alpha+\beta+N_1)}{\Gamma(x_1+\alpha)\Gamma(\beta+N_1-x_1)} \frac{\Gamma(x_1+x_2+\alpha)\Gamma(\beta+N_1-x_1+N_2-x_2)}{\Gamma(\alpha+\beta+N_1+N_2)} \end{aligned}$$

$c = \Gamma(x_1+x_2+\alpha)\Gamma(\beta+N_1-x_1+N_2-x_2) / \Gamma(\alpha+\beta+N_1+N_2)$

which is the predictive probability mass function (pmf) of X_2 given $X_1 = x_1$.

⁎ COMMENTS:

- Here, we have used the fact that $\mathrm{Beta}(\alpha, \beta)$ pdf of a random variable Θ has the following form (see the distribution table):

$$f_\Theta(\theta) = \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{\text{normalizing constant}} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

which implies

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

DC Level Estimation in AWGN with Known Variance: Predictive Distribution

SUPPOSE that we have collected N independent, identically distributed (i.i.d.) observations $(X[n])_{n=0}^{N-1}$ given θ and that they follow

$$\{X[n]\,|\,\Theta = \theta\} \sim \mathcal{N}(\theta, \sigma^2).$$

We wish to predict the next observation, denoted by $X_\star$, which is independent of $(X[n])_{n=0}^{N-1}$ given $\Theta = \theta$, and follows

$$\{X_\star \mid \Theta = \theta\} \sim \mathcal{N}(\theta, \sigma^2). \tag{8}$$

Then, $X_\star$ and

$$\overline{X} = \frac{1}{N}\sum_{n=0}^{N-1} X[n]$$

are also independent given $\Theta = \theta$; consequently,

$$f_{X_\star,\Theta|\overline{X}}(x_\star, \theta \mid \overline{x}) = f_{X_\star|\Theta}(x_\star \mid \theta)\, f_{\Theta|\overline{X}}(\theta \mid \overline{x}). \tag{9}$$

Based on (9), we have

$$\mathrm{E}_{X_\star,\Theta|\overline{X}}(\cdot \mid \overline{x}) = \mathrm{E}_{\Theta|\overline{X}}\big[\mathrm{E}_{X_\star|\Theta}(\cdot \mid \Theta) \mid \overline{x}\big]. \tag{10}$$

Recall that $\overline{x}$ is a sufficient statistic for θ based on $\boldsymbol{x} = (x[n])_{n=0}^{N-1}$; then, by sufficiency,

$$f_{\Theta|\boldsymbol{X}}(\theta \mid \boldsymbol{x}) = f_{\Theta|\overline{X}}(\theta \mid \overline{x}).$$

Our predictive distribution is

$$\begin{aligned} f_{X_\star|\boldsymbol{X}}(x_\star \mid \boldsymbol{x}) &= f_{X_\star|\overline{X}}(x_\star \mid \overline{x}) \\ &= \int \underbrace{f_{X_\star|\Theta}(x_\star \mid \vartheta)\, f_{\Theta|\overline{X}}(\vartheta \mid \overline{x})}_{f_{X_\star,\Theta|\overline{X}}(x_\star, \vartheta \mid \overline{x})}\, \mathrm{d}\vartheta. \end{aligned} \tag{11}$$

Focus on (9):

$$f_{X_\star,\Theta|\overline{X}}(x_\star, \theta \mid \overline{x}) = \underbrace{f_{X_\star|\Theta}(x_\star \mid \theta)}_{\mathcal{N}(x_\star \mid \theta, \sigma^2)}\ \underbrace{f_{\Theta|\overline{X}}(\theta \mid \overline{x})}_{\mathcal{N}(\theta \mid \mu_N, \tau_N^2)}$$

see handout `introBayes` where we derived

$$f_{\Theta|\overline{X}}(\theta \mid \overline{x}) = \mathcal{N}\big(\theta \mid \mu_N(\overline{x}), \tau_N^2\big)$$

with

$$\begin{aligned} \mu_N(\overline{x}) &= \frac{(N/\sigma^2)\overline{x} + (1/\tau_0^2)\mu_0}{N/\sigma^2 + 1/\tau_0^2} \\ \frac{1}{\tau_N^2} &= \frac{N}{\sigma^2} + \frac{1}{\tau_0^2}. \end{aligned}$$

and, therefore,

$$f_{X_\star,\Theta|\overline{X}}(x_\star, \theta \mid \overline{x}) \propto \exp\Big[-\frac{1}{2\sigma^2}(x_\star - \theta)^2\Big] \exp\Big\{-\frac{1}{2\tau_N^2}[\theta - \mu_N(\overline{x})]^2\Big\}$$

which is the kernel of a bivariate Gaussian pdf. Hence, $f_{X_\star,\Theta|\overline{X}}(x_\star, \theta \mid \overline{x})$ is a bivariate Gaussian pdf.

☞ WE wish to find the predictive pdf $f_{X_\star|\overline{X}}(x_\star \mid \overline{x})$.

Integrating θ out[4] in (11) is easy, see handout `revprob`. Since we know that the predictive pdf $f_{X_\star|\overline{X}}(x_\star \mid \overline{x})$ must be Gaussian, we just need to find its mean:

[4] i.e., marginalizing with respect to θ

$$\begin{aligned} \mathrm{E}_{X_\star|\overline{X}}(X_\star \mid \overline{x}) &= \mathrm{E}_{X_\star,\Theta|\overline{X}}(X_\star \mid \overline{x}) \\ &\overset{\text{iter. exp.}}{=} \underbrace{\mathrm{E}_{\Theta|\overline{X}}\big[\mathrm{E}_{X_\star|\Theta}(X_\star \mid \Theta) \mid \overline{x}\big]}_{\text{see (10)}} \\ &\overset{\text{see (8)}}{=} \mathrm{E}_{\Theta|\overline{X}}(\Theta \mid \overline{x}) \\ &= \mu_N(\overline{x}) \end{aligned}$$

and variance

use the law of conditional variances based on (10)

$$
\begin{aligned}
\operatorname{var}_{X_\star \mid \bar{X}}(X_\star \mid \bar{x}) &= \mathrm{E}_{\Theta \mid \bar{X}}\big[\underbrace{\operatorname{var}_{X_\star \mid \Theta}(X_\star \mid \Theta)}_{\sigma^2,\text{ see (8)}} \mid \bar{x}\big] \\
&\quad + \operatorname{var}_{\Theta \mid \bar{X}}\big[\underbrace{\mathrm{E}_{X_\star \mid \Theta}(X_\star \mid \Theta)}_{\Theta,\text{ see (8)}} \mid \bar{x}\big] \\
&= \sigma^2 + \tau_N^2.
\end{aligned}
$$

Hence,

$$
f_{X_\star \mid \bar{X}}(x_\star \mid \bar{x}) = \mathcal{N}\big(x_\star \mid \mu_N(\bar{x}), \sigma^2 + \tau_N^2\big).
$$

Acronyms

i.i.d. independent, identically distributed. 3

pdf probability density function. 1–4

pmf probability mass function. 3

References

Gelman, A., J. B. Carlin, H. S. Stern, David B. Dunson, Aki Vehtari, and D. B. Rubin (2014). *Bayesian Data Analysis*. 3rd ed. Boca Raton, FL: Taylor & Francis (cit. on p. 1).