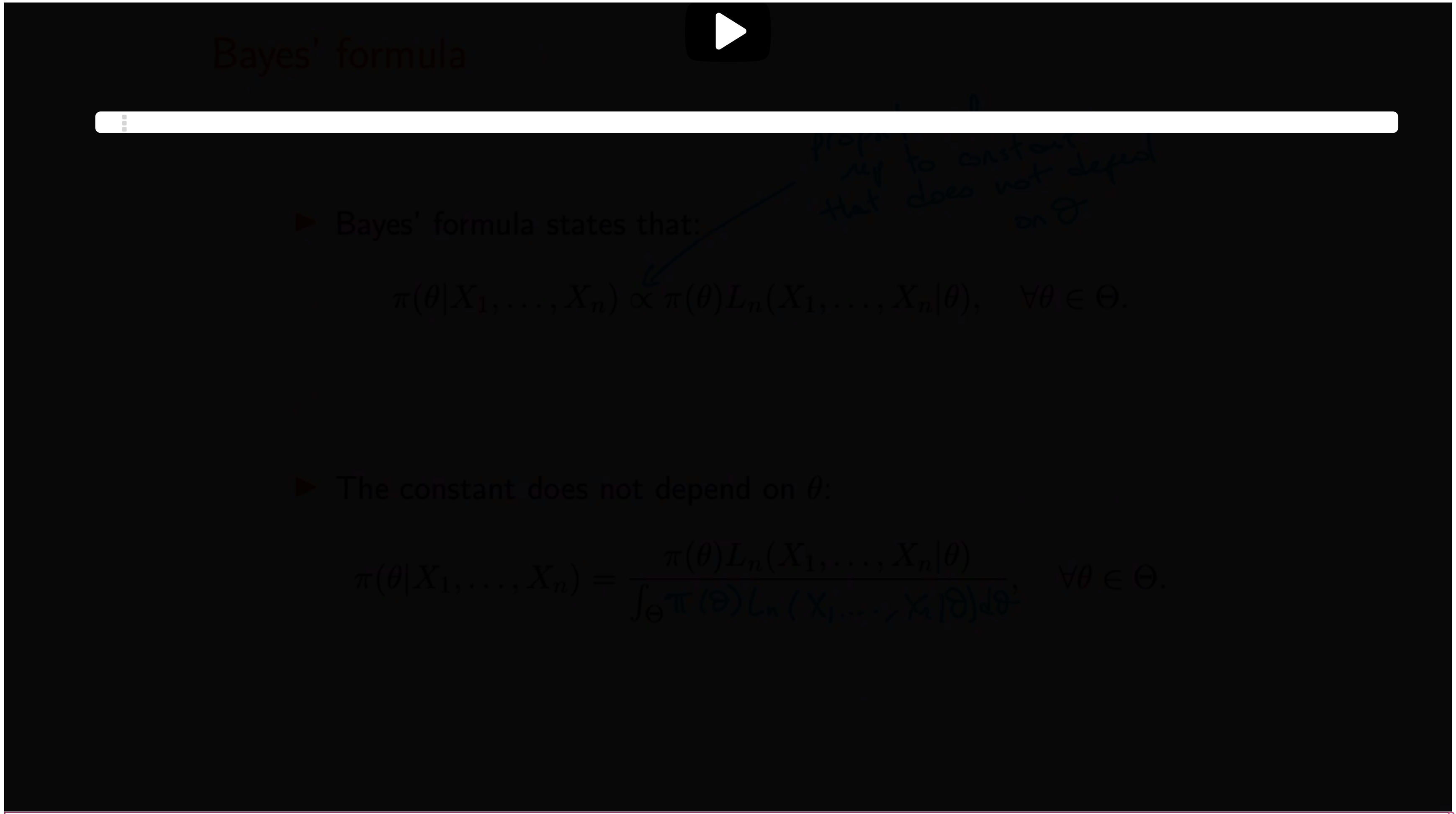
7. The Posterior Distribution, Bayes' Formula

Prior to Posterior



11:09 / 11:09 1.50x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Prior Implications to Posterior: True or False

1/1 point (graded)

Consider the case of a binary parameter $\theta \in \{0, 1\}$. We have the prior distribution $\pi(\theta)$ that satisfies $\pi(0) = p$ and $\pi(1) = 1 - p$, for some $0 \le p \le 1$. Then, we observe $X_1, \ldots, X_n$ with corresponding conditional likelihood function $L(X_1, \ldots, X_n|\theta)$ that is positive for both $\theta = 0$ and $\theta = 1$. Which of the following statements about the posterior distribution $\pi(\theta|X_1, \ldots, X_n)$ is true? (Choose all that apply.)

- [x] If $p = 0$, then the posterior distribution will be identical to the prior distribution.
- [x] If $p = 1$, then the posterior distribution will be identical to the prior distribution.
- [ ] If $p < \frac{1}{2}$, then in the posterior distribution, $\pi(\theta = 0|X_1, \ldots, X_n) \le \pi(\theta = 1|X_1, \ldots, X_n)$ will necessarily be true.
- [ ] If $p > \frac{1}{2}$, then in the posterior distribution, $\pi(\theta = 0|X_1, \ldots, X_n) \ge \pi(\theta = 1|X_1, \ldots, X_n)$ will necessarily be true.

✔

Solution:

In the cases where $p = 0$ or $p = 1$, the prior distribution would have one of $\pi(0)$ or $\pi(1)$ as 1 and the other as zero. As Bayes' rule gives that the posterior distribution $\pi(\theta|X_1, \ldots, X_n)$ is the likelihood function $L(X_1, \ldots, X_n)$ multiplied by the prior distribution, this means that if the prior is 0 for a certain θ, the product will still be zero in the (un-normalized) posterior.

Furthermore, as we've assumed the likelihood to be positive for any θ, the product will be positive in the (un-normalized) posterior. Hence, after normalization, we will get a posterior probability of 1 for the θ that has a probability of 1 in the prior, and a 0 for the other, making the posterior identical to the prior. Thus the choices **"If $p = 0$, then the posterior distribution will be identical to the prior distribution."** and **"If $p = 1$, then the posterior distribution will be identical to the prior distribution."** are correct.

The third and fourth choices are incorrect because it is possible for the likelihood function to skew the posterior probability in the other direction. For example, consider a prior with $\pi(0) = 0.3$, $\pi(1) = 0.7$, as well as likelihoods $L(X_1, \ldots, X_n|\theta = 0) = 0.9$ and $L(X_1, \ldots, X_n|\theta = 0) = 0.1$. Then Bayes' rule gives $\pi(\theta = 0|X_1, \ldots, X_n) = \frac{(0.3)(0.9)}{(0.3)(0.9)+(0.7)(0.1)} = \frac{27}{33}$ and $\pi(\theta = 1|X_1, \ldots, X_n) = \frac{(0.7)(0.1)}{(0.3)(0.9)+(0.7)(0.1)} = \frac{7}{33}$, so we get $\pi(\theta = 0|X_1, \ldots, X_n) > \pi(\theta = 1|X_1, \ldots, X_n)$, contrary to the third statement.

A similar example where the roles of $\theta = 0$ and $\theta = 1$ are reversed would disprove the fourth choice.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Updating Prior (Belief Propagation)

1/1 point (graded)

In this problem, we will explore how to update the belief successively, having observed data. The model is as follows:

- $\theta \in \Theta$, the parameter space; and $\pi(\cdot)$ is the prior distribution of θ.
- We observe i.i.d. (conditional on the parameter) data $X_1, \ldots, X_n$ and calculate the likelihood function $L_n(X_1, \ldots, X_n|\theta)$ (as in the setting of maximum likelihood estimation)
- Write $\phi(X_1, \ldots, X_n)$ as a placeholder function that depends on $X_1, \ldots, X_n$, but not on the parameter θ. (ϕ could stand for different functions in different equations. It's simply a placeholder whenever we want to collect terms that only depend on $X_1, \ldots, X_n$.)

In this context, we add observations one by one, computing the likelihood $L_i(X_1, \ldots, X_i|\theta)$ and posterior $\pi(\theta|X_1, \ldots, X_i)$ after each observation i. Which of the following identities are true? (Choose all that apply.)

- [x] $\pi(\theta|X_1, \ldots, X_n) = \pi(\theta) \cdot L_n(X_1, \ldots, X_n|\theta) \cdot \phi(X_1, \ldots, X_n)$
- [x] $L_n(X_1, \ldots, X_n|\theta) = L_{n-1}(X_1, \ldots, X_{n-1}|\theta) \cdot L_1(X_n|\theta)$
- [x] $\pi(\theta|X_1, \ldots, X_n) = \pi(\theta|X_1, \ldots, X_{n-1}) \cdot L_1(X_n|\theta) \cdot \phi(X_1, \ldots, X_n)$
- [x] $L_n(X_1, \ldots, X_n|\theta) = \frac{\pi(\theta|X_1, \ldots, X_n)}{\pi(\theta)} \phi(X_1, \ldots, X_n)$

Solution:

All the choices are correct. To see this, we proceed as follows. For brevity of notation, let $\alpha_n = \pi(\theta|X_1, \ldots, X_n)$; $\beta_n = L_n(X_1, X_2, \ldots, X_n|\theta)$. We also use X as a compact notation to represent $X_1, X_2, \ldots, X_n$, so that $\phi(X)$ corresponds to $\phi(X_1, \ldots, X_N)$.

- Note that, using Bayes' rule,

$$\begin{aligned}\alpha_n(\theta) &= \pi(\theta|X_1, \ldots, X_n) = \frac{p_n(X_1, \ldots, X_n|\theta)\,\pi(\theta)}{\int_\Theta p_n(X_1, \ldots, X_n|t)\,\pi(t)\;dt} \\ &= \pi(\theta)\,\beta_n(\theta)\,\phi(X),\end{aligned}$$

  where, $\phi(X)$ captures the term in the denominator. This does not depend on θ as we have already integrated over the variable.

- Similarly, using the independence of $X_1, \ldots, X_n$ conditional on θ,

$$\beta_n(\theta) = p_n(X_1, \ldots, X_n|\theta) = p_{n-1}(X_1, \ldots, X_{n-1}|\theta) \cdot p(X_n|\theta) = \beta_{n-1}(\theta) \cdot p(X_n|\theta).$$

- For this part, note that,

$$\begin{aligned}\alpha_n(\theta) &= \pi(\theta|X_1, \ldots, X_n) \\ &= \frac{p_n(X_1, \ldots, X_n|\theta)\,\pi(\theta)}{\widetilde{\phi}(X)} \\ &= \frac{p_{n-1}(X_1, \ldots, X_{n-1}|\theta)\,\pi(\theta)\,p(X_n|\theta)}{\phi(X)} \\ &\propto p(X_n|\theta)\,\underbrace{p_{n-1}(X_1, \ldots, X_{n-1}|\theta)\,\pi(\theta)}_{\propto \alpha_{n-1}(\theta)} \\ &\propto \alpha_{n-1}(\theta)\,p(X_n|\theta).\end{aligned}$$

- This follows by rearranging the first identity and taking the reciprocal of ϕ.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 5 Bayesian statistics:Lecture 17: Introduction to Bayesian Statistics / 7. The Posterior Distribution, Bayes' Formula

Add a Post

All Posts

Updating prior

discussion posted a day ago by Analesdey

Wow! I thought it is straight forward and was I wrong. Its one of the best constructed problems in this course I feel. Had my head scratching.

This post is visible to everyone.

GiulianoCruz

about 4 hours ago

I actually felt it misleading.

Given the notation we have been using and everything we learn on Bayes, I thought $\phi\left(X_1, \ldots, X_n\right)$ would be the integral of the product of likelihood and prior (hence the denominator). It was really not clear to me that we could make up any function of that (say, 1 over the integral). Reading the problem now seems like "ok, I get what you mean". But before seeing the answer it really sounded like the "marginal density of the data" :(

Add a comment

Showing all responses

Preview

Submit