Cross Validated

Bayesian posterior: is multiplying likelihood by prior (rather than simulation) an acceptable approach?

Asked 2 years, 9 months ago Active 2 years, 9 months ago Viewed 2k times

2

1

Ken Rice has a helpful introductory set of slides available online called 'Bayesian Statistics (a very brief introduction)'.

http://faculty.washington.edu/kenrice/BayesIntroClassEpi515kmr2016.pdf

On slide 23 he gives this formulation, which comes directly from Bayes theorem:

Posterior ∝ Likelihood × Prior

However, within a section on 'when priors don't matter (much)', on slide 33 he describes a method whereby you multiply the likelihood function by the prior to get the posterior. But he describes this as "semi-Bayesian". (On slide 35, I think he's referring to the same thing when he mentions an "approximate Bayes" approach, and describing "full Bayes" as better.)

My question is: in what sense is taking a prior expressed as a functional form and multiplying it by a likelihood function only semi-Bayesian?

Is it just that the (normal) likelihood he presents is only an approximation to the real likelihood function? Or is it because the multiplication he presents is only approximate? Or is there something more fundamentally 'semi' about this type of Bayesianism?

More generally, the focus of texts on Bayesian inference seems to be on simulation (especially MCMC) approaches. Is this because it is 'wrong' to get your posterior distribution from multiplying a prior distribution by the likelihood function generated by some new data? Or is it because the analytical route is not often available to you?

bayesian

asked Feb 17 '17 at 16:53

jamse
23 3

1 I don't have an answer to the "why does the author call this semi-Bayesian" question, but no there is nothing wrong with multiplying prior and likelihood, and yes the reason why simulation is emphasized is because models which allow you to do the computation analytically are not always rich enough to specify very meaningful priors, etc. – Chris Haug Feb 17 '17 at 17:37

The posterior is not the prior times the likelihood. It's proportional to it. What this implies, for example, is that maximizing the posterior is the same as maximizing the prior times the likelihood. But if you actually want to compute the posterior, you need to normalize, and this normalization step is often intractable. – Qiaochu Yuan Feb 17 '17 at 18:57

I think the "semi-Bayesian" accounts to his way of calculating confidence intervals. As others have mentioned, $posterior \propto prior \times likelihood$ is as Bayesian as it gets. – bayerj Feb 17 '17 at 20:13

2 It is approximate Bayes because the likelihood is replaced with a Normal approximation, not because of the missing constant. – Xi'an Feb 18 '17 at 16:49

1 Answer

6

Bayes theorem is

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

so posterior is *proportional* to likelihood times prior. For it to be equal we need to multiply the right-hand side of the equation by a normalizing constant, so that it integrates to unity, what makes posterior a proper probability distribution.

Constant does not change anything about finding maximum of the function, since each possible output of the function is multiplied by the same constant, so if you are only interested in point estimate (*maximum a posteriori* estimate), then you can ignore the normalizing constant. However if you want to obtain proper posterior distribution, then it is needed and we often use MCMC to find it and solve the equation.

See also the Why Normalizing Factor is Required in Bayes Theorem? thread.

Edit

But, as noticed by *Xi'an*, what the slides that you refer to actually say is that the author by "semi-Bayesian" approach means using normal distribution as likelihood function and normal priors:

$$\begin{aligned}
\text{Prior: } \beta &\sim N(\mu_0, \sigma_0^2) \\
\text{Likelihood: } &\text{approx } N(\hat{\beta}, \widehat{\text{StdErr}}[\hat{\beta}]^2) \\
\text{Posterior: } \beta &\sim N\left(\mu_0 w + \hat{\beta}(1-w), \frac{1}{1/\sigma_0^2 + 1/\widehat{\text{StdErr}}[\hat{\beta}]^2}\right), \\
\text{where } & w = \frac{1/\sigma_0^2}{1/\sigma_0^2 + 1/\widehat{\text{StdErr}}[\hat{\beta}]^2}
\end{aligned}$$

This makes computation very easy since using conjugate priors, but it may not be the best approximation for all cases (recall that normal distribution is continuous, symmetric, and reaches from $-\infty$ to ∞ -- this is not true for many different kinds of data!).

edited Apr 13 '17 at 12:44
Community ♦
1

answered Feb 17 '17 at 19:55
Tim ♦
71.3k 14 152 253

2 +1. I'd just like to add the importance of why we actually care about normalizing, which is buried in the last answer to the question you linked: there's not necessarily an easy way to know the possible range of values of a posterior, especially if the posterior was computed by simulation. So without normalizing, you don't actually know how big two posterior likelihoods are relative to each other, and can't use the resulting values for anything other than to say "this one is bigger than that one". – shadowtalker Feb 17 '17 at 20:15

Thanks for the answer Tim. Apologies if this is wrong form - I'm new around here, so not sure on etiquette around questioners commenting on answers. Do you think it would be worth adding a sentence or two on @ssdecontrol's reasoning to this answer? – jamse Feb 20 '17 at 19:15

Also, up in the comments to the question @Xi'an suggests another reason this might be considered approximate Bayes: "because the likelihood is replaced with a Normal approximation". Do you see any truth in that, and if so, would it be worth noting it as a second element of semi-ness in the answer? – jamse Feb 20 '17 at 19:15

@jamse we ask each other to update our answers all the time, it's appropriate as long as it's polite (and you're definitely polite enough!). Note that you also have the option to suggest an edit to the poster by clicking the "edit" button below the answer. – shadowtalker Feb 20 '17 at 19:31