The Hedging Hyperprior

Stephen D. Simon 1*

Department of Biomedical and Health Informatics,

University of Missouri-Kansas City

Yu Jiang

Division of Epidemiology, Biostatistics, and Environmental Health,

University of Memphis

and

Byron J. Gajewski

Department of Biostatistics,

University of Kansas Medical Center

June 27, 2017

Abstract

Here's a scratch file where I want to park some complex formulas that are not needed (yet) for the regular paper.

In many settings,...

Keywords: Beta-binomial model, Informative prior, Modified power prior

1 Introduction

Stephen Senn is ...

2 A beta-binomial model with an informative prior distribution

Consider a Bayesian ...

3 The hedging hyperprior

The hedging hyperprior ...

4 Alternatives to the hedging hyperprior

There are several ...

Consider the historic prior and the data as two sets of data, D0 and D. If, for example, your informative prior were beta(4, 16), then you would create a pseudo data set with (4-1) successes and (16-1) failures. Apply this pseudo data to a flat prior ($\pi_0(\theta)$), get a posterior distribution the resembles your informative prior.

$\pi(\theta|D_0) \propto L(\theta|D_0)\pi_0(\theta)$

Then use $\pi(\theta|D_0)$ use that as your prior distribution for the new data.

$\pi(\theta|D, D_0) \propto L(\theta|D)\pi(\theta|D_0)$

What Ibrahim et al. (2003) suggested is that you could downweight the historical prior by raising the historical likelihood to the power.

$\pi(\theta|D_0, \tau) \propto L(\theta|D_0)^\tau \pi_0(\theta)$

With $\tau=1$, you get the full strength of the historical prior and with $\tau=0$ you wipe out the historic data and are left with just a flat prior. This formulation does not satisfy the likelihood principle, but you can include a fudge factor ($\int L(\theta|D_0)^\tau \pi_0(\theta)d\theta$) to make things work nicely.

It is fairly easy to show that hedging hyperprior in the beta-binomial model is equivalent to the modified power prior. Notice that

$$L(\theta|D_0)^\tau = \left(\frac{(\alpha+\beta-2)!}{(\alpha-1)!(\beta-1)!} \theta^{\alpha-1}(1-\theta)^{\beta-1} \right)^\tau$$

is proportional to

$$\theta^{(\alpha-1)\tau}(1-\theta)^{(\beta-1)\tau}$$

If you multiply this by a uniform prior for theta and apply the fudge factor, you get a $Beta(1+(\alpha-1)\tau, 1+(\beta-1)\tau)$ distribution for $L(\theta|D_0)$.

The advantage to formulating the problem as the hedging hyperprior is that it makes it easy to see how to use standard Bayesian software like BUGS, JAGS, or STAN. Also, the hedging hyperprior is easier to visualize, allowing you to readily discern its behavior.

You can also use a hierarchical approach. Pretend that you have a multi-center trial with two centers. Consider the Beta(a, b) informative prior as (a+b)-2 pseudo-observations with a-1 successes and b-1 failures from one center of a multi-center trail and the n actual observations with x successes and n-x failures as a second center. Then you can model these two centers using a hierarchical model. The second center will borrow strength from the first center, but the amount of information that it borrows is limited by the center-to-center variation. In other words, you use the prior only to the extent to which the prior and the data agree with one another. When they disagree strongly, the center-to-center variation is large and the posterior distribution for the second center is mostly just based on the likelihood.

$\pi_0 \sim Beta(\alpha_0, \beta_0)$

$\pi \sim Beta(\alpha_0, \beta_0)$

$\alpha_0 \sim Pareto(1)$

$\beta_0 \sim Pareto(1)$

$D_0 = \alpha - 1 \ successes, \ \beta - 1 \ failures.$

$\alpha - 1 \sim Binom(\alpha + \beta - 2, \pi0)$

$x \sim Binom(n, \pi)$

$\pi(\pi0, \pi, \alpha0, \beta0)$

$L(D_0|\alpha, \beta) = \frac{(\alpha+\beta-2)!}{(\alpha-1)!(\beta-1)!} \pi_0^{\alpha-1}(1-\pi_0)^{\beta-1} \frac{\Gamma(\alpha_0)\Gamma(\beta_0)}{\Gamma(\alpha_0+\beta_0)} \pi_0^{\alpha_0-1}(1-\pi_0)^{\beta_0-1} \frac{1}{\alpha_0^2} \frac{1}{\beta_0^2}$

beta-prime distribution: $f(x) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha,\beta)}$

beta distribution: $f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$

logistic distribution: $\dfrac{e^{-\frac{x-\mu}{s}}}{s\left(1+e^{-\frac{x-\mu}{s}}\right)^2}$

$g(x) = log(x); \ g^{-1}(y) = exp(y)$

$\dfrac{e^{y(\alpha-1)}}{(1+e^y)^{\alpha+\beta}B(\alpha,\beta)}e^y$

$\dfrac{e^{y\alpha}}{(1+e^y)^{\alpha+\beta}B(\alpha,\beta)}$

5 Appendix

6 Conclusion

The bibliography ...

7 Bibliography

References

Ibrahim, J. G., M.-H. Chen, and D. Sinha (2003). On optimality properties of the power prior. *Journal of the American Statistical Association 98*(461), 204–213.