



Introduction to Bayesian Computation: Rejection Sampling and Importance Sampling

A Beta-Binomial Model for Overdispersion



Tsutakawa et al. (1985) describe the problem of simultaneously estimating the rates of death from stomach cancer for males at risk in the age bracket 45–64 for the largest cities in Missouri. Table 5.1 displays the mortality rates for 20 of these cities, where a cell contains the number n_j at risk and the number of cancer deaths y_j for a given city.

Table 5.1. Cancer mortality data. Each ordered pair represents the number of cancer deaths y_j and the number at risk n_j for an individual city in Missouri.

(0, 1083)	(0, 855)	(2, 3461)	(0, 657)	(1, 1208)	(1, 1025)
(0, 527)	(2, 1668)	(1, 583)	(3, 582)	(0, 917)	(1, 857)
(1, 680)	(1, 917)	(54, 53637)	(0, 874)	(0, 395)	(1, 581)
(3, 588)	(0, 383)				

A Beta-Binomial Model for Overdispersion

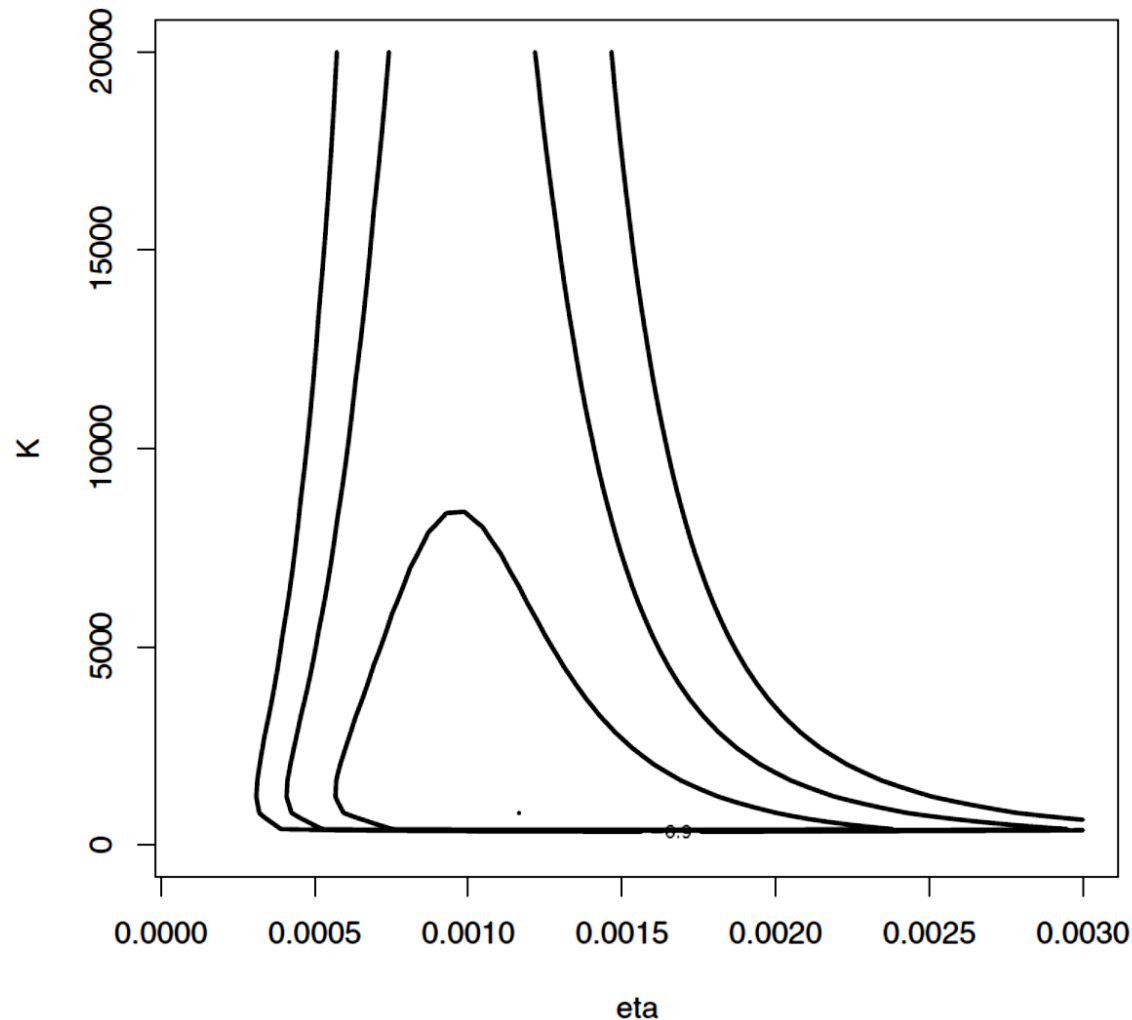


Fig. 5.1. Contour plot of parameters η and K in the beta-binomial model problem.

A Beta-Binomial Model for Overdispersion

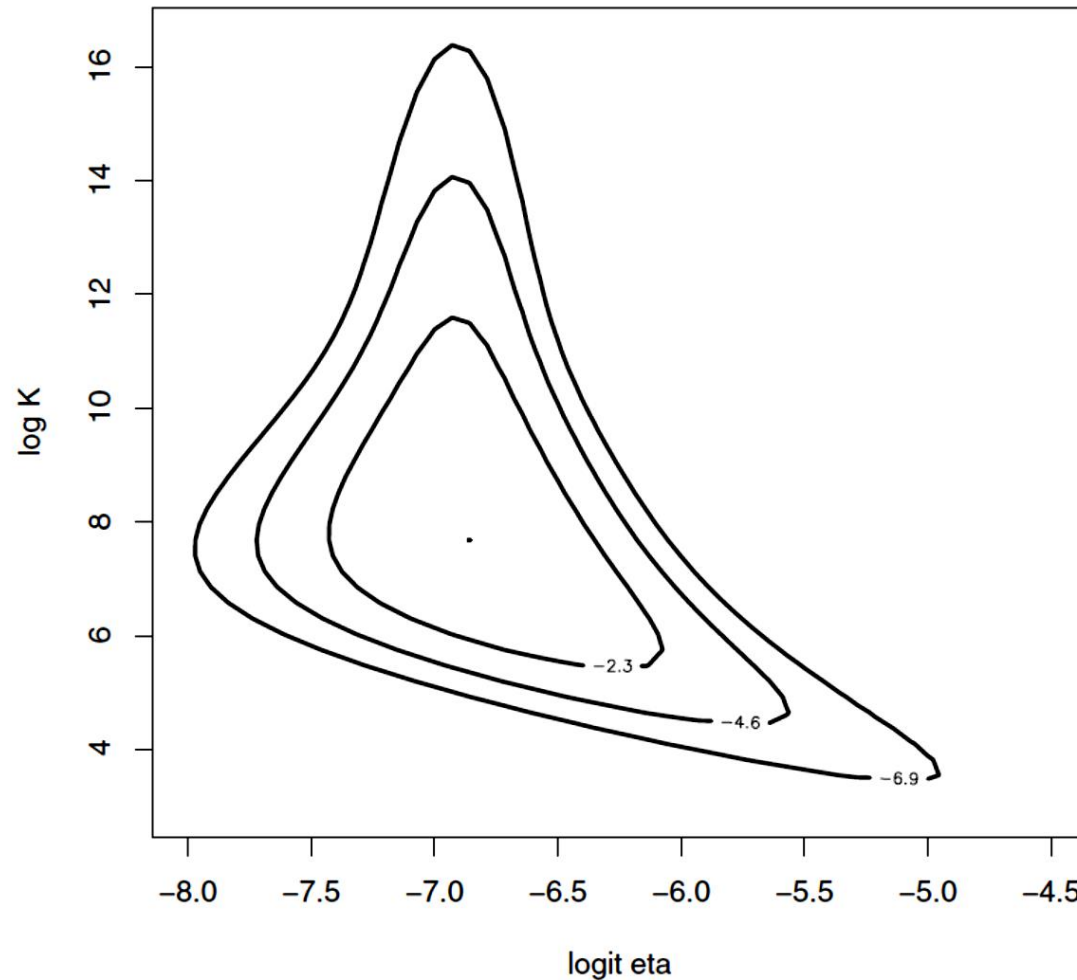


Fig. 5.2. Contour plot of transformed parameters $\text{logit}(\eta)$ and $\log K$ in the beta-binomial model problem.

Rejection Sampling

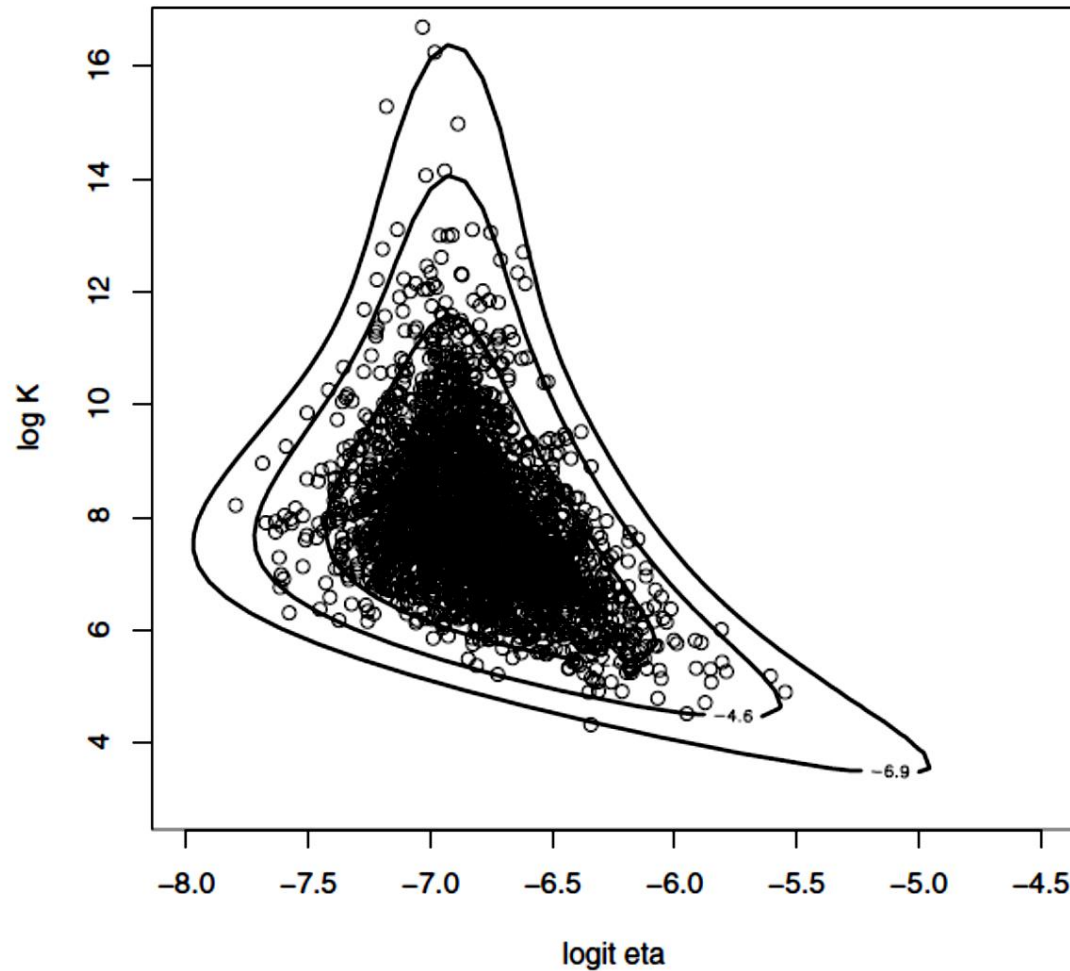


Fig. 5.4. Contour plot of $\text{logit}(\eta)$ and $\log K$ in the beta-binomial model problem together with simulated draws from the rejection algorithm.

Importance Sampling



Let us return to the basic problem of computing an integral in Bayesian inference. In many situations, the normalizing constant of the posterior density $g(\theta|y)$ will be unknown, so the posterior mean of the function $h(\theta)$ will be given by the ratio of integrals

$$E(h(\theta)|y) = \frac{\int h(\theta)g(\theta)f(y|\theta)d\theta}{\int g(\theta)f(y|\theta)d\theta},$$

where $g(\theta)$ is the prior and $f(y|\theta)$ is the likelihood function. If we were able to simulate a sample $\{\theta^j\}$ directly from the posterior density g , then we could approximate this expectation by a Monte Carlo estimate. In the case where we are not able to generate a sample directly from g , suppose instead that we can construct a probability density p that we can simulate and that approximates the posterior density g . We rewrite the posterior mean as

Importance Sampling



$$\begin{aligned} E(h(\theta)|y) &= \frac{\int h(\theta) \frac{g(\theta)f(y|\theta)}{p(\theta)} p(\theta) d\theta}{\int \frac{g(\theta)f(y|\theta)}{p(\theta)} p(\theta) d\theta} \\ &= \frac{\int h(\theta) w(\theta) p(\theta) d\theta}{\int w(\theta) p(\theta) d\theta}, \end{aligned}$$

where $w(\theta) = g(\theta)f(y|\theta)/p(\theta)$ is the weight function. If $\theta^1, \dots, \theta^m$ are a simulated sample from the approximation density p , then the importance sampling estimate of the posterior mean is

$$\bar{h}_{IS} = \frac{\sum_{j=1}^m h(\theta^j) w(\theta^j)}{\sum_{j=1}^m w(\theta^j)}.$$

Importance Sampling



This is called an *importance sampling estimate* because we are sampling values of θ that are important in computing the integrals in the numerator and denominator. The simulation standard error of an importance sampling estimate is estimated by

$$se_{\bar{h}_{IS}} = \frac{\sqrt{\sum_{j=1}^m ((h(\theta^j) - \bar{h}_{IS})w(\theta^j))^2}}{\sum_{j=1}^m w(\theta^j)}.$$

As in rejection sampling, the main issue in designing a good importance sampling estimate is finding a suitable sampling density p . This density should be of a familiar functional form so simulated draws are available. The density should mimic the posterior density g and have relatively flat tails so that the weight function $w(\theta)$ is bounded from above. One can monitor the choice of p by inspecting the values of the simulated weights $w(\theta^j)$. If there are no unusually large weights, then it is likely that the weight function is bounded and the importance sampler is providing a suitable estimate.

Importance Sampling



To illustrate the use of different proposal densities in importance sampling in our example, consider the problem of estimating the posterior mean of a function of $\theta_2 = \log K$ conditional on a value of $\theta_1 = \text{logit}(\eta)$. The posterior density of θ_2 , conditional on θ_1 is given by

$$g_1(\theta_2|\text{data}, \theta_1) \propto \frac{K}{(1+K)^2} \prod_{j=1}^{20} \frac{B(K\eta + y_j, K(1-\eta) + n_j - y_j)}{B(K\eta, K(1-\eta))},$$

where $\eta = \exp(\theta_1)/(1 + \exp(\theta_1))$ and $K = \exp(\theta_2)$. In the following, we write the function `betabinexch.cond` to compute this posterior density conditional on the value $\theta_1 = -6.818793$. This function is written to allow the input of a vector of values of $\theta_2 = \log K$. Also, unlike the other functions in this chapter, the function `betabinexch.cond` returns the value of the density rather than the value of the log density.

Importance Sampling

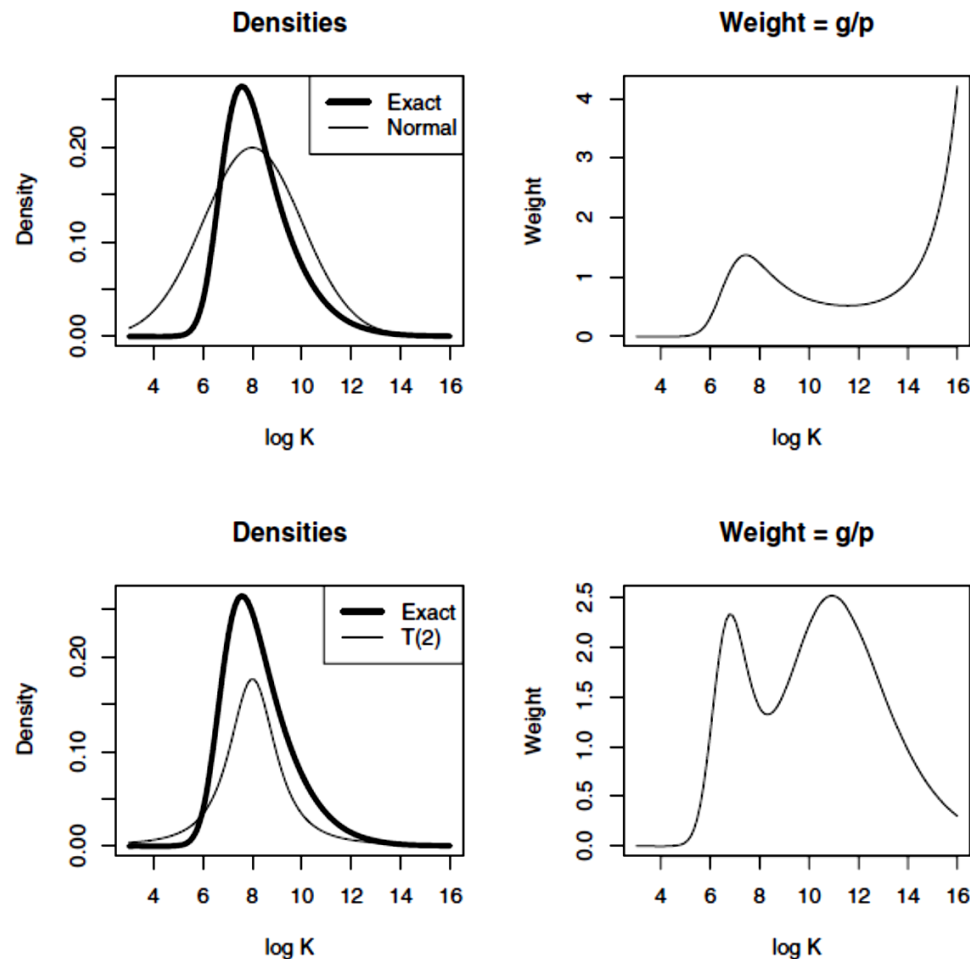


Fig. 5.5. Graph of the posterior density of $\log K$ and weight function using a normal proposal density (top) and a $t(2)$ proposal density (bottom). By using a t proposal density, the weight function appears to be bounded from above.