influence of the data and the prior essentially coincide. We shall see examples of Jeffreys priors in future sections.

Finally, we emphasise that if the specific form of vague prior is influential in the analysis, this strongly suggests you have insufficient data to draw a robust conclusion based on the data alone and that you should not be trying to be "non-informative" in the first place.

5.2.4 Location parameters

A location parameter θ is defined as a parameter for which $p(y|\theta)$ is a function of $y - \theta$, and so the distribution of $y - \theta$ is independent of θ. In this case Fisher's information is constant, and so the Jeffreys procedure leads to a uniform prior which will extend over the whole real line and hence be improper. In BUGS we could use `dflat()` to represent this distribution, but tend to use proper distributions with a large variance, such as `dunif(-100,100)` or `dnorm(0,0.0001)`: we recommend the former with appropriately chosen limits, since explicit introduction of these limits reminds us to be wary of their potential influence. We shall see many examples of this use, for example, for regression coefficients, and it is always useful to check that the posterior distribution is well away from the prior limits.

5.2.5 Proportions

The appropriate prior distribution for the parameter θ of a Bernoulli or binomial distribution is one of the oldest problems in statistics, and here we illustrate a number of options. First, both Bayes (1763) and Laplace (1774) suggest using a uniform prior, which is equivalent to Beta$(1, 1)$. A major attraction of this assumption, also known as the Principle of Insufficient Reason, is that it leads to a discrete uniform distribution for the predicted number y of successes in n future trials, so that $p(y) = 1/(n+1)$, $y = 0, 1, ..., n$,* which seems rather a reasonable consequence of "not knowing" the chance of success. On the $\phi = \text{logit}(\theta)$ scale, this corresponds to a standard logistic distribution, represented as `dlogis(0,1)` in BUGS (see code below).

Second, an (improper) uniform prior on ϕ is formally equivalent to the (improper) Beta$(0, 0)$ distribution on the θ scale, i.e., $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$: the code below illustrates the effect of bounding the range for ϕ and hence making these distributions proper. Third, the Jeffreys principle leads to a Beta$(0.5, 0.5)$ distribution, so that $p_J(\theta) = \pi^{-1}\theta^{\frac{1}{2}}(1-\theta)^{\frac{1}{2}}$. Since it is common to use normal prior distributions when working on a logit scale, it is of interest to consider what normal distributions on ϕ lead to a "near-uniform" distribution on θ. Here we consider two possibilities: assuming a prior variance of 2 for ϕ can be shown to give a density for θ that is "flat" at $\theta = 0.5$, while a normal with variance 2.71 gives a close approximation to a standard logistic distribution, as we saw in Example 4.1.1.

```
theta[1]          ~ dunif(0,1)      # uniform on theta
phi[1]            ~ dlogis(0,1)

phi[2]            ~ dunif(-5,5)     # uniform on logit(theta)
logit(theta[2]) <- phi[2]

theta[3]          ~ dbeta(0.5,0.5) # Jeffreys on theta
phi[3]          <- logit(theta[3])

phi[4]            ~ dnorm(0,0.5)    # var=2, flat at theta = 0.5
logit(theta[4]) <- phi[4]

phi[5]            ~ dnorm(0,0.368) # var=2.71, approx. logistic
logit(theta[5]) <- phi[5]
```

We see from Figure 5.1 that the first three options produce apparently very different distributions for θ, although in fact they differ at most by a single implicit success and failure (§5.3.1). The normal prior on the logit scale with variance 2 seems to penalise extreme values of θ, while that with variance 2.71 seems somewhat more reasonable. We conclude that, in situations with very limited information, priors on the logit scale could reasonably be restricted to have variance of around 2.7.

Example 5.2.1. *Surgery (continued): prior sensitivity*
What is the sensitivity to the above prior distributions for the mortality rate in our "Surgery" example (Example 3.3.2)? Suppose in one case we observe 0/10 deaths (Figure 5.2, left panel) and in another, 10/100 deaths (Figure 5.2, right panel). For 0/10 deaths, priors 2 and 3 pull the estimate towards 0, but the sensitivity is much reduced with the greater number of observations.

5.2.6 Counts and rates

For a Poisson distribution with mean θ, the Fisher information is $I(\theta) = 1/\theta$ and so the Jeffreys prior is the improper $p_J(\theta) \propto \theta^{-\frac{1}{2}}$, which can be approximated in BUGS by a `dgamma(0.5, 0.00001)` distribution. The same prior is appropriate if θ is a rate parameter per unit time, so that $Y \sim \text{Poisson}(\theta t)$.

*See Table 3.1 — the posterior predictive distribution for a binomial observation and beta prior is a beta-binomial distribution. With no observed data, $n = y = 0$ in Table 3.1, this posterior predictive distribution becomes the *prior predictive* distribution, which reduces to the discrete uniform for $a = b = 1$.

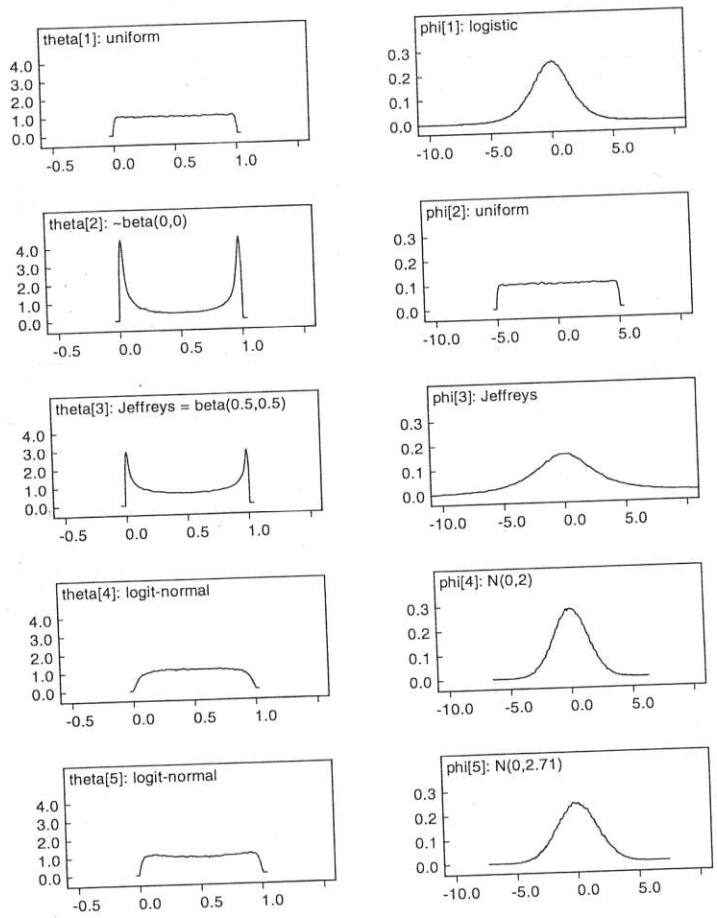

FIGURE 5.1
Empirical distributions (based on 100,000 samples) corresponding to various different priors for a proportion parameter.

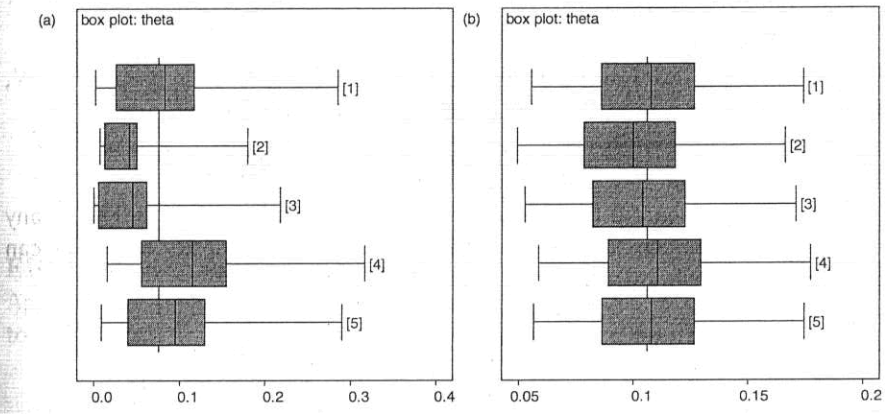

FIGURE 5.2
Box plots comparing posterior distributions arising from the five priors discussed above for mortality rate: (a) 0/10 deaths observed; (b) 10/100 deaths observed.

5.2.7 Scale parameters

Suppose σ is a scale parameter, in the sense that $p(y|\sigma) = \sigma^{-1} f(y/\sigma)$ for some function f, so that the distribution of Y/σ does not depend on σ. Then it can be shown that the Jeffreys prior is $p_J(\sigma) \propto \sigma^{-1}$, which in turn means that $p_J(\sigma^k) \propto \sigma^{-k}$, for any choice of power k. Thus for the normal distribution, parameterised in BUGS in terms of the precision $\tau = 1/\sigma^2$, we would have $p_J(\tau) \propto \tau^{-1}$. This prior could be approximated in BUGS by, say, a `dgamma(0.001,0.001)`, which also can be considered an "inverse-gamma distribution" on the variance σ^2. Alternatively, we note that the Jeffreys prior is equivalent to $p_J(\log \sigma^k) \propto$ const, i.e., an improper uniform prior. Hence it may be preferable to give $\log \sigma^k$ a uniform prior on a suitable range, for example, `log.tau ~ dunif(-10, 10)` for the logarithm of a normal precision. We would usually want the bounds for the uniform distribution to have negligible influence on the conclusions.

We note some potential conflict in our advice on priors for scale parameters: a uniform prior on $\log \sigma$ follows Jeffreys' rule but a uniform on σ is placing a prior on an interpretable scale. There usually would be negligible difference between the two — if there is a noticeable difference, then there is clearly little information in the likelihood about σ and we would recommend a weakly informative prior on the σ scale.

Note that the advice here applies only to scale parameters governing the variance or precision of *observable* quantities. The choice of prior for the variance of *random effects* in a hierarchical model is more problematic — we discuss this in §10.2.3.