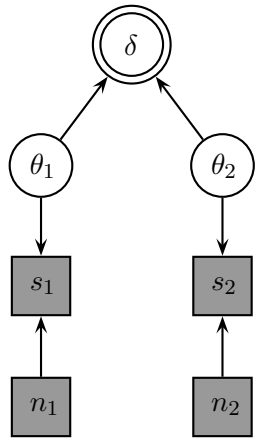
9 Comparing binomial rates

9.1 Equality of proportions

In their article "After the promise: the STD consequences of adolescent virginity pledges," Brückner and Bearman (2005) analyzed a series of interviews conducted as part of the National Longitudinal Study of Adolescent Health. The focus of the article was on the sexual behavior of adolescents, aged 18–24, who have made a virginity pledge. This is a public or written pledge to remain a virgin until marriage.

Consider the hypothesis that the sexual behavior of pledgers is not very different from that of non-pledgers, except for the fact that pledgers are less likely to use condoms when they first have sex. The Brückner and Bearman (2005) study presents relevant data. Those adolescents who reported using a condom for their first sexual encounter numbered 424 out of 777 ($\approx 54.6\%$) for the pledgers, and 5416 out of 9072 ($\approx 59.7\%$) for non-pledgers. To what extent does a statistical analysis support the assertion that pledgers are less likely than non-pledgers to use a condom?



$$
\begin{aligned}
s_1 &\sim \text{Binomial}(\theta_1, n_1) \\
s_2 &\sim \text{Binomial}(\theta_2, n_2) \\
\theta_1 &\sim \text{Beta}(1, 1) \\
\theta_2 &\sim \text{Beta}(1, 1) \\
\delta &\leftarrow \theta_1 - \theta_2
\end{aligned}
$$

Fig. 9.1 Bayesian graphical model for the comparison of two proportions.

The Bayesian model selection approach for these data is simple and general. We assume that the number of condom users among the pledgers (i.e., $s_1 = 424$ out of $n_1 = 777$) and among the non-pledgers (i.e, $s_2 = 5416$ out of $n_2 = 9072$) are governed by binomial rates θ_1 and θ_2, respectively, which are given uniform priors. The difference between the two rate parameters is $\delta = \theta_1 - \theta_2$. This leads to the

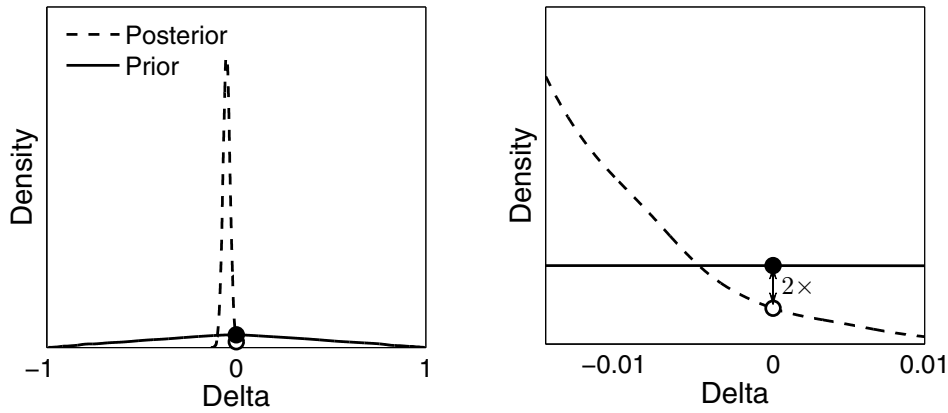


Fig. 9.2 Prior and posterior distributions of the rate difference δ for the pledger data. The left panel shows the distributions across their entire range, and the right panel zooms in on the area that is relevant for the Savage–Dickey test. The markers show the densities at $\delta = 0$ needed to estimate the Bayes factor.

same graphical model we originally considered in Section 3.2, as is shown again in Figure 9.1.

The null hypothesis states that the rates θ_1 and θ_2 are equal, and hence $\mathcal{H}_0 : \delta = 0$. The unrestricted alternative hypothesis states that the rates are free to vary, $\mathcal{H}_1 : \delta \neq 0$. Thus, applying the Savage–Dickey method for finding the Bayes factor requires the prior and posterior distributions for δ.

The script `Pledgers_1.txt` implements the graphical model in WinBUGS:

```
# Pledgers
model{
  # Rates and Difference
  theta1 ~ dbeta(1,1)
  theta2 ~ dbeta(1,1)
  delta <- theta1-theta2
  # Data
  s1 ~ dbin(theta1,n1)
  s2 ~ dbin(theta2,n2)
  # Prior Sampling
  theta1prior ~ dbeta(1,1)
  theta2prior ~ dbeta(1,1)
  deltaprior <- theta1prior-theta2prior
}
```

The code `Pledgers_1.m` or `Pledgers_1.R` calls WinBUGS to draw samples from the posterior and prior of the rate difference δ. The left panel of Figure 9.2 shows the distributions on their entire range, and the right panel zooms in on the relevant region around $\delta = 0$. The critical point $\delta = 0$ is supported about twice as much under the prior as it is under the posterior, and so the Bayes factor is approximately 2 in favor of the alternative hypothesis.

A reasonable interpretation of this Bayes factor is that the data do not provide much evidence in favor of one hypothesis over the other. This seems more conservative than the interpretation that may be drawn from Bayesian parameter estimation, since the Bayesian 95% credible interval for the posterior of δ is approximately $(-0.09, -0.01)$ and does not include 0. The reason for the discrepancy is that the Bayesian hypothesis test punishes $\mathcal{H}_1$ for assigning prior mass to values of δ that yield very low likelihoods (see Berger & Delampady, 1987, for a discussion).

Exercises

Exercise 9.1.1 Because the rate parameters θ_1 and θ_2 both have a uniform prior distribution, the prior distribution for the difference parameter δ can be found analytically as a triangular distribution. What are the advantages of using this result, rather than relying on computational sampling? What are the disadvantages?

Exercise 9.1.2 In the current analysis, we put independent priors on θ_1 and θ_2. Do you think this is plausible? How would you change the model to take into account the possible dependence? How would this affect the outcome of the Bayesian test?

Exercise 9.1.3 This example corresponds to a rare case in which the Bayes factor is available analytically. $BF_{01} = p\left(D \mid \mathcal{H}_0\right) / p\left(D \mid \mathcal{H}_1\right)$ is given by

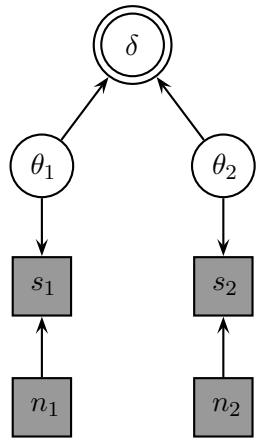
$$BF_{01} = \frac{\binom{n_1}{s_1}\binom{n_2}{s_2}}{\binom{n_1+n_2}{s_1+s_2}} \frac{(n_1+1)(n_2+1)}{n_1+n_2+1}.$$

Calculate the Bayes factor analytically, and compare it to the result obtained using the Savage–Dickey method.

Exercise 9.1.4 For the pledger data, a frequentist test for equality of proportions indicates that $p \approx 0.006$. This tells us that when $\mathcal{H}_0$ is true (i.e., the proportions of condom users are equal in the two groups), then the probability is about 0.006 that we would encounter a result at least as extreme as the one that was in fact observed. What conclusions would you draw based on this information? Discuss the usefulness of the Bayes factor and the p-value in answering the scientific question of whether pledgers are less likely than non-pledgers to use a condom.

9.2 Order-restricted equality of proportions

Whether pledgers are less likely than non-pledgers to use condoms has a natural interpretation as involving the order-restriction that the rate is lower for the pledgers than for the non-pledgers. This corresponds to a different alternative hypothesis $\mathcal{H}_2 : \delta < 0$.



$$
\begin{aligned}
s_1 &\sim \text{Binomial}(\theta_1, n_1) \\
s_2 &\sim \text{Binomial}(\theta_2, n_2) \\
(\theta_1, \theta_2) &\sim \text{Uniform}(0,1)^2,\ \theta_2 > \theta_1 \\
\delta &\leftarrow \theta_1 - \theta_2
\end{aligned}
$$

Fig. 9.3 Bayesian graphical model for the order-restricted comparison of two proportions.

The graphical model for this order-restricted analysis is shown in Figure 9.3, and involves changing the priors on the rates. Since the order-restriction is that $\delta < 0$, the requirement is that $\theta_2 > \theta_1$. This is the same sort of restriction involved in the two-country quiz example in Section 6.4, where it was addressed by using `theta2 ~ dunif(0,1)` and `theta1 ~ dunif(0,theta2)`. While this approximate approach worked satisfactorily for inferring posterior distributions of parameters, an exact approach is needed in the current context of model comparison. This is because the approximate method does not generate a uniform prior distribution of θ_1 and θ_2 over the region of the joint parameter space satisfying the order-restriction, and model selection, unlike parameter estimation, is likely to be sensitive to this mismatch between available information and the implementation of the model.

To make this clear, Figure 9.4 shows samples from the joint prior distribution of (θ_1, θ_2) under the approximate method in the left panel. It is visually clear that the approximate method places too much density in the bottom-left corner of (θ_1, θ_2), because each θ_1 is drawn uniformly, and then θ_1 is constrained to be less than θ_2. The exact distribution is shown in the middle panel, and gives equal density to the valid region in which $\theta_2 > \theta_1$. The impact of the approximation on the prior for δ is shown in the right panel of Figure 9.4, and involves too much prior density being given to values near zero.

The script `Pledgers_2.txt` implements the graphical model in WinBUGS:

```
# Pledgers, Order Constrained Rates
model{
  # Order Constrained Rates
  thetap[1:2] ~ dmnorm(mu[],TI[,])
  theta1 <- phi(cos(angle)*thetap[1]-sin(angle)*abs(thetap[2]))
  theta2 <- phi(sin(angle)*thetap[1]+cos(angle)*abs(thetap[2]))
  # Data
  s1 ~ dbin(theta1,n1)
  s2 ~ dbin(theta2,n2)
```

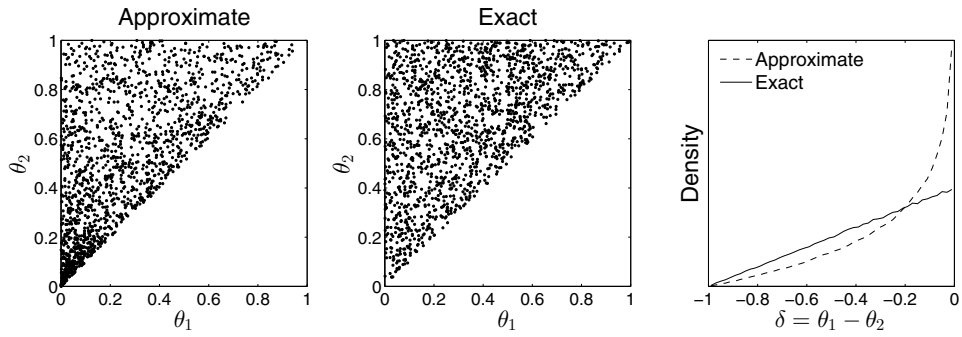


Fig. 9.4 Samples from approximate (left panel) and exact (middle panel) approaches to implementing the order-constrained prior on (θ_1, θ_2) in which $\theta_2 > \theta_1$. The right panel shows the implied prior distributions on $\delta = \theta_1 - \theta_2$ arising from both methods.

```
  # Difference
  delta <- theta1-theta2
  # Prior Sampling
  thetapprior[1:2] ~ dmnorm(mu[],TI[,])
  theta1prior <- phi(cos(angle)*thetapprior[1]-sin(angle)*abs(thetapprior[2]))
  theta2prior <- phi(sin(angle)*thetapprior[1]+cos(angle)*abs(thetapprior[2]))
  deltaprior  <- theta1prior-theta2prior
  # Constants
  angle <- 45*3.1416/180
  TI[1,1] <- 1
  TI[1,2] <- 0
  TI[2,1] <- 0
  TI[2,2] <- 1
  mu[1] <- 0
  mu[2] <- 0
}
```

An exact method is used for the order-constrained prior, involving jointly drawing samples from a bivariate standard Gaussian, rotating these samples by 45 degrees, and then transforming the rotated samples into rates that lie in the unit square.

The code `Pledgers_2.m` or `Pledgers_2.R` draws samples from the posterior and prior of the rate difference δ under the order-constrained prior. The left panel of Figure 9.5 shows the distributions on their entire range, and the right panel zooms in on the relevant region around $\delta = 0$. It is clear that order-restriction has the effect of doubling the density of the prior, but has little effect on the posterior. This means that the critical point $\delta = 0$ is now supported about four times as much under the prior as it is under the posterior, and so the Bayes factor is approximately 4 in favor of the alternative hypothesis.

Exercise

Exercise 9.2.1 Consider an order-restricted test of $\mathcal{H}_0 : \delta = 0$ versus $\mathcal{H}_3 : \delta > 0$. What do you think the result will be? Check your intuition by implementing the appropriate graphical model, and estimating the Bayes factor.

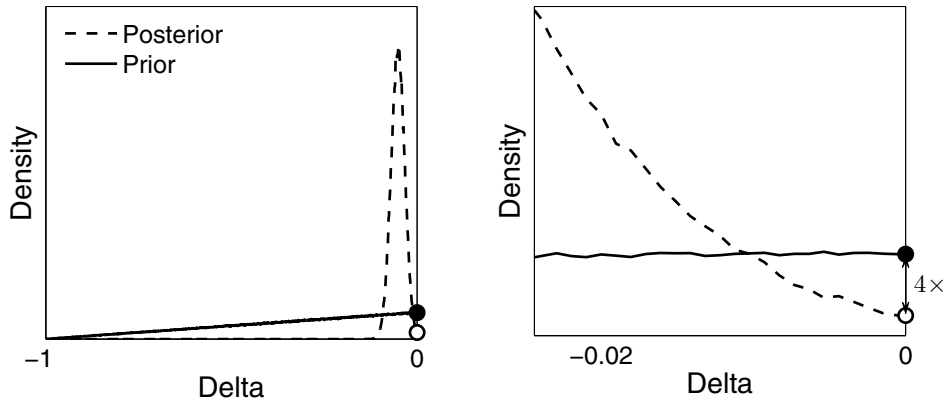


Fig. 9.5 Prior and posterior distributions of the rate difference δ for the pledger data, under the order-restricted analysis. The left panel shows the distributions across their entire range, and the right panel zooms in on the area that is relevant for the Savage–Dickey test. The markers show the densities at $\delta = 0$ needed to estimate the Bayes factor.

Fig. 9.6 Example pair of similar pictures used in Experiment 3 from Zeelenberg et al. (2002).

9.3 Comparing within-subject proportions

In their article "Priming in implicit memory tasks: Prior study causes enhanced discriminability, not only bias," Zeelenberg, Wagenmakers, and Raaijmakers (2002) reported three experiments in two-alternative forced-choice perceptual identification. In the test phase of each experiment, a stimulus (e.g., a picture of a clothespeg) is briefly presented and masked. Immediately after the mask the subject is confronted with two choice options, examples of which are shown in Figure 9.6. One is the target (e.g., a picture of the clothespeg) and a similar foil alternative (e.g., a picture of a stapler). The subject's goal is to identify the target.

Prior to the test phase, the Zeelenberg et al. (2002) experiments featured a study phase, in which subjects studied a subset of the choice alternatives that would also be presented in the later test phase. Two conditions were critical: the "study-

neither" condition, in which neither choice alternative was studied, and the "study-both" condition, in which both choice alternatives were studied. In the first two experiments reported by Zeelenberg et al. (2002), subjects choose the target stimulus more often in the study-both condition than in the study-neither condition. This *both-primed benefit* suggests that prior study leads to enhanced discriminability, not just a bias to prefer the studied alternative (e.g., Ratcliff & McKoon, 1997).

Here we focus on statistical inference for Experiment 3 from Zeelenberg et al. (2002). In the study phase of this experiment, all 74 subjects were presented with 21 pairs of similar pictures, as in the example in Figure 9.6. In the test phase, all participants had to identify briefly presented target pictures among a set of two alternatives. The test phase was composed of 42 pairs of similar pictures, 21 of which had been presented in the study phase.

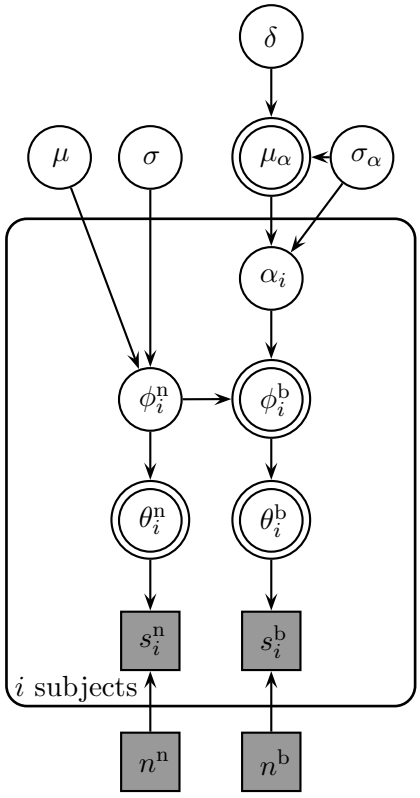
In order to assess the evidence in favor of the both-primed benefit, the authors carried out a standard analysis and computed a one-sample t-test:

> Mean percentage of correctly identified pictures was calculated for each participant. When neither the target nor the foil had been studied, 71.5% of the pictures were correctly identified. When both the target and the foil had been studied, 74.7% of the pictures were correctly identified. The difference between the study-both and study-neither conditions was significant, $t(73) = 2.19$, $p < .05$.

Our Bayes factor test of the both-primed benefit is shown in Figure 9.7. The model assumes that the number of correct choices made by the ith subject is binomially distributed with accuracy rate parameters θ_i^b and θ_i^n in the study-both and study-neither conditions, respectively. We allow for individual differences between the accuracy of subjects, and individual differences in the impact of the both-primed benefit.

Individual differences in the study-neither condition are modeled as having a Gaussian distribution, with the rates themselves given by a probit transformation. This is conceptually the same idea as the logit transformation approach in Section 6.6, with probit and logit transformations being quantitatively slightly different. The probit transform is shown in Figure 9.8, and is the inverse cumulative distribution function of the standard Gaussian distribution. This means, for example, that a probit rate of $\phi = 0$ maps on to a rate of $\theta = \Phi(0) = 0.5$, and a probit rate of $\phi = 1.96$ maps on to a rate of $\theta = \Phi(1.96) = 0.975$. The standard Gaussian distribution on the probit scale, $\phi \sim \text{Gaussian}(0, 1)$, corresponds to a standard uniform distribution on the rate scale, $\theta \sim \text{Uniform}(0, 1)$. Formally, the ith subject has $\phi_i^n \sim \text{Gaussian}(\mu, 1/\sigma^2)$ and their accuracy rate is then given by $\theta_i^n = \Phi(\phi_i^n)$, transforming a real number into a probability.

The model in Figure 9.7 uses the probit scale to implement the potential both-primed benefit. The benefit for the ith subject is drawn from a Gaussian distribution with mean μ_α and and standard deviation σ_α. This effect additively changes the study-neither into the study-both accuracy on the probit scale, so that $\phi_i^b = \phi_i^n + \alpha_i$. The model incorporates a parameter δ that quantifies the critical effect size, $\delta = \mu_\alpha / \sigma_\alpha$. Effect size is a dimensionless quantity, and this makes it relatively easy



$$
\begin{aligned}
\mu &\sim \text{Gaussian}(0,1)_{\mathcal{I}(0,\infty)} \\
\sigma &\sim \text{Uniform}(0,10) \\
\delta &\sim \text{Gaussian}(0,1)_{\mathcal{I}(0,\infty)} \\
\sigma_\alpha &\sim \text{Uniform}(0,10) \\
\mu_\alpha &\leftarrow \delta\sigma_\alpha \\
\alpha_i &\sim \text{Gaussian}(\mu_\alpha, 1/\sigma_\alpha^2) \\
\phi_i^{\mathrm{n}} &\sim \text{Gaussian}(\mu, 1/\sigma^2) \\
\phi_i^{\mathrm{b}} &\leftarrow \phi_i^{\mathrm{n}} + \alpha_i \\
\theta_i^{\mathrm{n}} &\leftarrow \Phi(\phi_i^{\mathrm{n}}) \\
\theta_i^{\mathrm{b}} &\leftarrow \Phi(\phi_i^{\mathrm{b}}) \\
s_i^{\mathrm{n}} &\sim \text{Binomial}(\theta_i^{\mathrm{n}}, n^{\mathrm{n}}) \\
s_i^{\mathrm{b}} &\sim \text{Binomial}(\theta_i^{\mathrm{b}}, n^{\mathrm{b}})
\end{aligned}
$$

Fig. 9.7 Graphical model for the analysis of the Zeelenberg et al. (2002) data.

to define a principled prior, such as the Cauchy distribution (i.e., a t distribution with one degree of freedom) and the standard Gaussian distribution (e.g., Gönen, Johnson, Lu, & Westfall, 2005; Rouder et al., 2009). The latter prior is known as the "unit information prior," as it carries as much information as a single observation (Kass & Wasserman, 1995). Our model uses the standard Gaussian distribution prior.

For the parameters that are not the focus of the statistical test (i.e., μ_ϕ, σ_ϕ, and σ_α) relatively uninformative priors are used. The prior for the group mean of the study-neither condition, μ_ϕ, is a standard Normal truncated to be greater than zero, which on the rate scale corresponds to a uniform distribution from 0.5 to 1. For σ_ϕ and σ_α, the model uses uniform priors over a large enough range from 0 to 10.

With this account of the data-generating model in place, we can now turn to hypothesis testing. The null hypothesis states that there is no both-primed benefit, and hence the effect size is zero: $\mathcal{H}_0 : \delta = 0$. The alternative, order-restricted hypothesis states that there is a both-primed benefit, and hence $\mathcal{H}_1 : \delta > 0$.

The script `Zeelenberg.txt` implements the graphical model in WinBUGS:

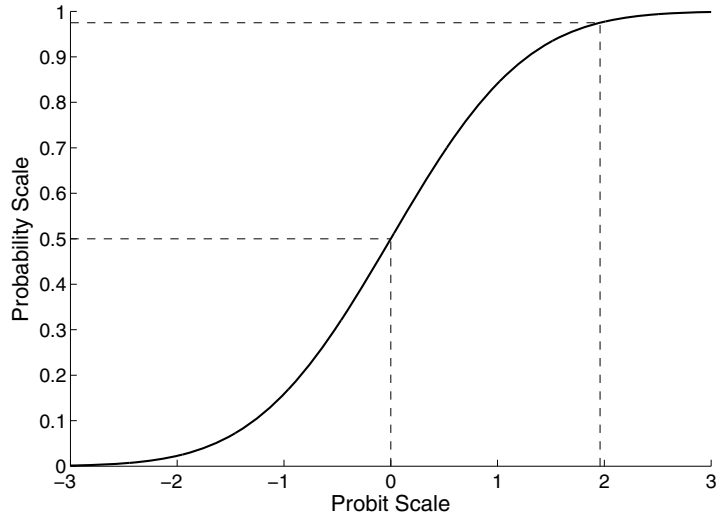


Fig. 9.8 The probit transformation.

```
# Zeelenberg
model{
  for (i in 1:ns){
    # Data
    sb[i] ~ dbin(thetab[i],nb)
    sn[i] ~ dbin(thetan[i],nn)
    # Probit Transformation
    thetab[i] <- phi(phib[i])
    thetan[i] <- phi(phin[i])
    # Individual Parameters
    phin[i] ~ dnorm(mu,lambda)
    alpha[i] ~ dnorm(mualpha,lambdaalpha)
    phib[i] <- phin[i]+alpha[i]
  }
  # Priors
  mu ~ dnorm(0,1)I(0,)
  sigma ~ dunif(0,10)
  lambda <- pow(sigma,-2)
  # Priming Effect
  sigmaalpha ~ dunif(0,10)
  lambdaalpha <- pow(sigmaalpha,-2)
  delta ~ dnorm(0,1)I(0,)
  mualpha <- delta*sigmaalpha
  # Sampling from Prior Distribution for Delta
  deltaprior ~ dnorm(0,1)I(0,)
}
```

The code `Zeelenberg.m` or `Zeelenberg.R` draws samples from the posterior and the prior for the effect size δ, and Figure 9.9 shows the results of applying the Savage–Dickey method. The critical effect size of $\delta = 0$ is supported about 4 times as much under the prior as it is under the posterior. That is, the data have decreased

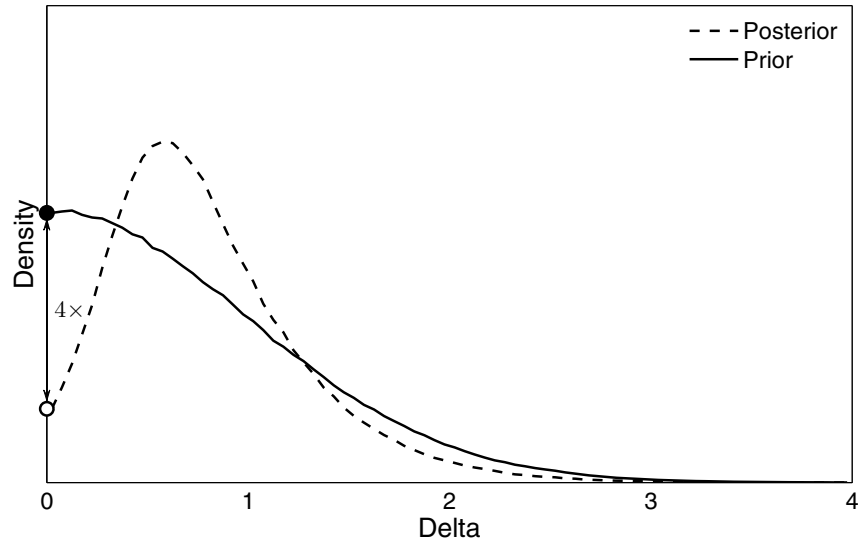


Fig. 9.9 Prior and posterior distributions of the effect size δ for the Zeelenberg et al. (2002) data. The markers show the densities at $\delta = 0$ needed to estimate the Bayes factor.

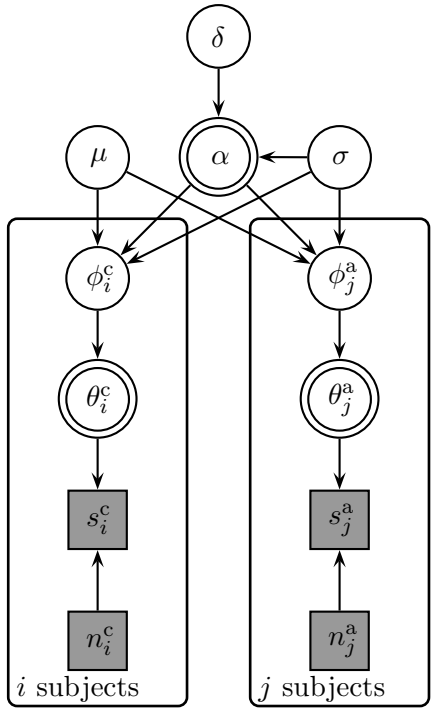
the support for $\delta = 0$ by a factor of 4. Thus, the Bayes factor is about 4 in favor of the alternative hypothesis, and a reasonable conclusion might be that the data weakly support the assertion that there is a both-primed benefit.

Exercise

Exercise 9.3.1 The Zeelenberg data can also be analyzed using the Bayesian t-test discussed in Chapter 8. Think of a few reasons why this might not be such a good idea. Then, despite your reservations, apply the Bayesian t-test anyway. How do the results differ? Why?

9.4 Comparing between-subject proportions

In their article "How specific are executive functioning deficits in Attention Deficit Hyperactivity Disorder and autism?", Geurts, Verté, Oosterlaan, Roeyers, and Sergeant (2004) studied the performance of children with ADHD and autism on a range of cognitive tasks. Here we focus on a small subset of the data and consider the question whether children who develop typically (i.e., "normal controls") outperform children with ADHD on the Wisconsin Card Sorting Test (WCST). The WCST requires people to learn, by trial and error, to sort cards according to an implicit rule. The complication is that, over the course of the experiment, the sorting rule sometimes changes. This means that in order to avoid too many mistakes,



$$
\begin{aligned}
\delta &\sim \text{Gaussian}(0, 1) \\
\mu &\sim \text{Gaussian}(0, 1) \\
\sigma &\sim \text{Uniform}(0, 10) \\
\alpha &\leftarrow \delta\sigma \\
\phi_i^{\mathrm{c}} &\sim \text{Gaussian}(\mu + \alpha/2, 1/\sigma^2) \\
\phi_j^{\mathrm{a}} &\sim \text{Gaussian}(\mu - \alpha/2, 1/\sigma^2) \\
\theta_i^{\mathrm{c}} &\leftarrow \Phi(\phi_i^{\mathrm{c}}) \\
\theta_j^{\mathrm{a}} &\leftarrow \Phi(\phi_j^{\mathrm{a}}) \\
s_i^{\mathrm{c}} &\sim \text{Binomial}(\theta_i^{\mathrm{c}}, n_i^{\mathrm{c}}) \\
s_j^{\mathrm{a}} &\sim \text{Binomial}(\theta_j^{\mathrm{a}}, n_j^{\mathrm{a}})
\end{aligned}
$$

Fig. 9.10 Graphical model for the analysis of the Geurts et al. (2004) data.

people have to suppress the tendency to perseverate, and discover and adopt the new rule. Because of these task demands, performance on the WCST is thought to quantify cognitive flexibility or set shifting ability.

The experiment of interest contains data from 26 normal controls and 52 children with ADHD. Each child performed the WCST, and the measure of interest is the number of correctly sorted cards relative to the total number of sorting opportunities. The WCST provides a maximum of 128 cards to sort, but, depending on a child's performance, this number could also be lower. Overall, the group of normal controls sorted the cards correctly on 65.4% of the cases, and the group of ADHD children sorted the cards correctly on 66.9% of the cases. The null hypothesis states that the probability of sorting cards correctly does not differ between the normal controls and the ADHD children. A between-subjects frequentist t-test on the proportion of correctly sorted cards does not allow one to reject the null hypothesis, $t(40.2) = 0.37$, $p = 0.72$. But this statistic does not quantify the evidence in favor of the null hypothesis.

The key difference between this example and the Zeelenberg example in Section 9.3 is that the experimental design is now between-subjects. Figure 9.10 shows the graphical model for this design. There are now separate plates for the control and ADHD groups. Within these groups the ith or jth child, respectively, has a success rate of θ_i^c and θ_j^a. These account for the observed data, which are the number of correctly sorted cards s_i^c and s_j^a out of the number of attempts n_i^c and n_j^a.

Individual variation is again modeled by a Gaussian distribution on the probit scale. The means of the control and ADHD groups are $\mu+\alpha/2$ and $\mu-\alpha/2$, so that they differ by α, and both Gaussians have the same standard deviation σ. Thus, on the probit scale, $\phi_i^c \sim \text{Gaussian}\left(\mu+\alpha/2, 1/\sigma^2\right)$ and $\phi_i^a \sim \text{Gaussian}\left(\mu-\alpha/2, 1/\sigma^2\right)$, with the transformations $\theta_i^c = \Phi\left(\phi_i^c\right)$ and $\theta_j^a = \Phi\left(\phi_j^a\right)$.

As before, the parameters that are not the focus of the statistical test are given relatively uninformative priors. Also as before, a parameter $\delta = \alpha/\sigma$ is used to quantify effect size, and is given a "unit information" standard normal prior. The null hypothesis states that normal controls and ADHD children perform the same on the WCST, and hence the effect size is zero: $\mathcal{H}_0 : \delta = 0$. The unrestricted alternative hypothesis states that there is a difference in performance, and hence $\mathcal{H}_1 : \delta \neq 0$.

The script `Geurts.txt` implements the graphical model in WinBUGS:

```
# Geurts
model{
  for (i in 1:nsc){
    kc[i] ~ dbin(thetac[i],nc[i])
    thetac[i] <- phi(phic[i])
    phic[i] ~ dnorm(muc,lambda)
  }
  for (j in 1:nsa){
    ka[j] ~ dbin(thetaa[j],na[j])
    thetaa[j] <- phi(phia[j])
    phia[j] ~ dnorm(mua,lambda)
  }
  muc <- mu+alpha/2
  mua <- mu-alpha/2
  # Priors
  mu ~ dnorm(0,1)
  sigma ~ dunif(0,10)
  alpha <- delta*sigma
  lambda <- pow(sigma,-2)
  delta ~ dnorm(0,1)
  # Sampling from Prior Distribution for Delta
  deltaprior ~ dnorm(0,1)
}
```

The code `Geurts.m` or `Geurts.R` draws samples from the posterior and the prior for the effect size δ. The results are shown in Figure 9.11. The ADHD children performed slightly better than the normal controls, and this is reflected in a posterior distribution for δ. This posterior is slightly asymmetric around zero, and assigns more mass to negative than to positive values of δ. The Bayesian 95% credible interval for δ is approximately $(-0.54, 0.42)$.

Figure 9.11 shows that the data have made the value $\delta = 0$ about four times more likely than it was before. Thus, the data support the claim, with a Bayes factor of about 4, that normal controls and ADHD children perform equally well on the WCST, when compared to the claim that these groups perform differently.

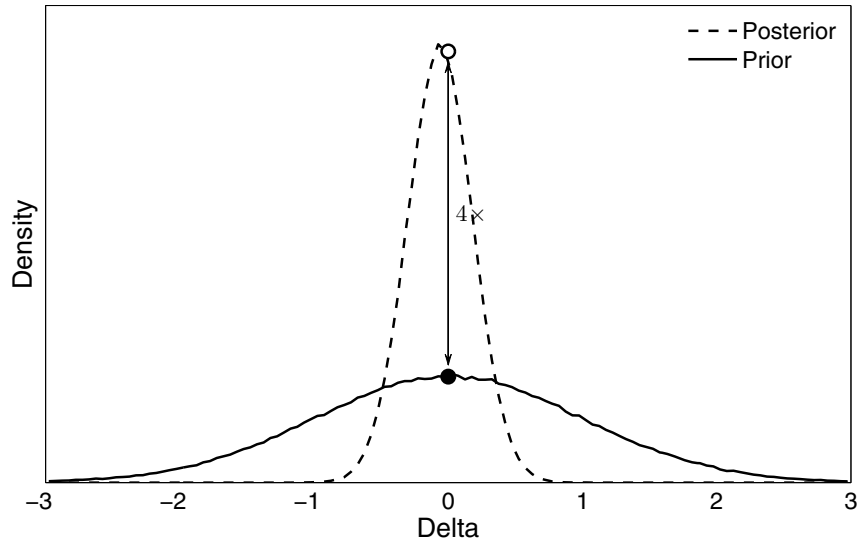


Fig. 9.11 Prior and posterior distributions of the effect size δ for the Geurts et al. (2004) data. The markers show the densities at $\delta = 0$ needed to estimate the Bayes factor.

Exercises

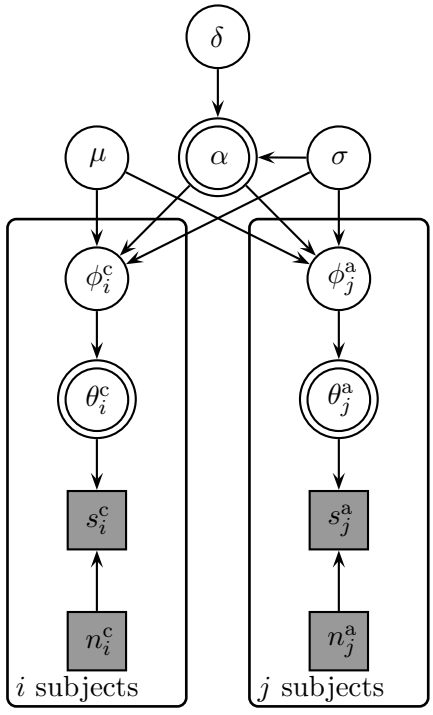
Exercise 9.4.1 A between-subjects frequentist t-test on the proportion of correctly sorted cards does not allow one to reject the null hypothesis, $t\,(40.2) = 0.37$, $p = 0.72$. In what way does the Bayesian approach improve upon the frequentist inference?

Exercise 9.4.2 In what way is the model of the data in Figure 9.10 superior to the statistical model assumed by the t-test?

9.5 Order-restricted between-subjects comparison

A natural modification of the alternative hypothesis for the Geurts et al. (2004) data is to impose the order-restriction corresponding to the assumption that normal controls perform better than ADHD children. This alternative hypothesis is $\mathcal{H}_2 : \delta > 0$. This hypothesis seems reasonable to entertain because of its a priori plausibility, but the data suggest that, if anything, the reverse is true. What can we expect when we test $\mathcal{H}_0 : \delta = 0$ versus $\mathcal{H}_2 : \delta > 0$?

First, note that the posterior for δ is not far from being symmetric around zero. If it were completely symmetric, the height of both the prior and the posterior is multiplied by 2, so that their ratio stays the same. Second, the posterior for δ is not quite symmetric around zero, and assigns slightly more mass to values that are inconsistent with $\mathcal{H}_2$. This will slightly increase the support for $\mathcal{H}_0$ over $\mathcal{H}_2$. These



$$
\begin{aligned}
\delta &\sim \text{Gaussian}(0,1)_{\mathcal{I}(0,\infty)} \\
\mu &\sim \text{Gaussian}(0,1) \\
\sigma &\sim \text{Uniform}(0,10) \\
\alpha &\leftarrow \delta\sigma \\
\phi_i^{\mathrm{c}} &\sim \text{Gaussian}(\mu+\alpha/2, 1/\sigma^2) \\
\phi_j^{\mathrm{a}} &\sim \text{Gaussian}(\mu-\alpha/2, 1/\sigma^2) \\
\theta_i^{\mathrm{c}} &\leftarrow \Phi(\phi_i^{\mathrm{c}}) \\
\theta_j^{\mathrm{a}} &\leftarrow \Phi(\phi_j^{\mathrm{a}}) \\
s_i^{\mathrm{c}} &\sim \text{Binomial}(\theta_i^{\mathrm{c}}, n_i^{\mathrm{c}}) \\
s_j^{\mathrm{a}} &\sim \text{Binomial}(\theta_j^{\mathrm{a}}, n_j^{\mathrm{a}})
\end{aligned}
$$

Fig. 9.12 Graphical model for the analysis of the Geurts et al. (2004) data.

two considerations lead us to expect that the evidence in favor of $\mathcal{H}_0$ over $\mathcal{H}_2$ will be slightly larger than that of $\mathcal{H}_0$ over $\mathcal{H}_1$. First, note that the posterior for δ is not far from being symmetric around zero. If it were completely symmetric, the height of both the prior and the posterior is multiplied by 2, so that their ratio stays the same. Second, the posterior for δ is not quite symmetric around zero, and assigns slightly more mass to values that are inconsistent with $\mathcal{H}_2$. This will slightly increase the support for $\mathcal{H}_0$ over $\mathcal{H}_2$. These two considerations lead us to expect that the evidence in favor of $\mathcal{H}_0$ over $\mathcal{H}_2$ will be slightly larger than that of $\mathcal{H}_0$ over $\mathcal{H}_1$.

Figure 9.12 shows the modified graphical model for the order-restricted analysis, which simply restricts the prior on δ to positive values.

The script `GeurtsOrderRestricted.txt` implements the graphical model in WinBUGS:

```
# Geurts, Order Restricted
model{
  for (i in 1:nsc){
    kc[i] ~ dbin(thetac[i],nc[i])
    thetac[i] <- phi(phic[i])
    phic[i] ~ dnorm(muc,lambda)
  }
  for (j in 1:nsa){
    ka[j] ~ dbin(thetaa[j],na[j])
```

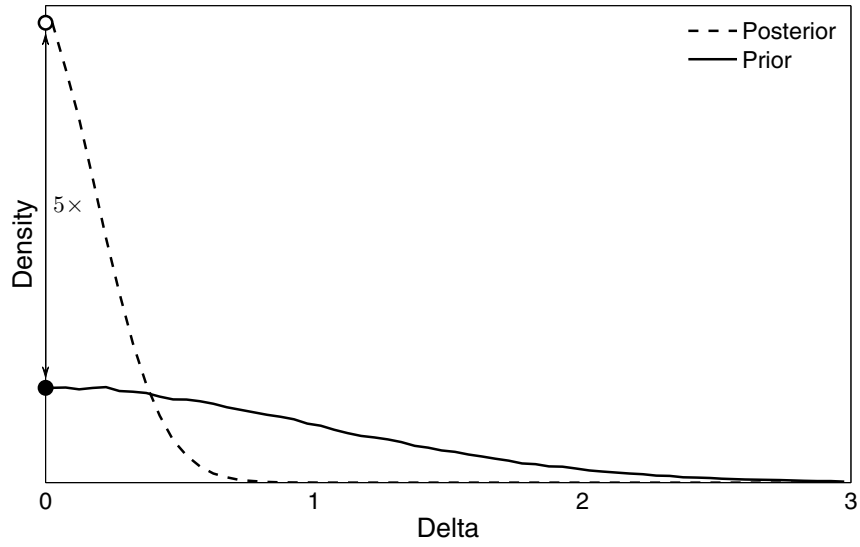


Fig. 9.13 Prior and posterior distributions of the effect size δ for the Geurts et al. (2004) data. The markers show the densities at $\delta = 0$ needed to estimate the Bayes factor.

```
    thetaa[j] <- phi(phia[j])
    phia[j] ~ dnorm(mua,lambda)
  }
  muc <- mu+alpha/2
  mua <- mu-alpha/2
  # Priors
  mu ~ dnorm(0,1)
  sigma ~ dunif(0,10)
  alpha <- delta*sigma
  lambda <- pow(sigma,-2)
  delta ~ dnorm(0,1)I(0,)
  # Sampling from Prior Distribution for Delta
  deltaprior ~ dnorm(0,1)
}
```

The code `GeurtsOrderRestricted.m` or `GeurtsOrderRestricted.R` draws samples from the posterior and the prior for δ. Figure 9.13 shows the results, confirming the expectation that the order-restriction slightly increases the evidence in favor of $\mathcal{H}_0$. The Bayes factor in favor of the null hypothesis is now about 5. Thus, the data support the assertion that normal controls and children with ADHD perform similarly on the WCST, although the evidence is not overwhelming.

Exercises

Exercise 9.5.1 For the order-restricted comparison of $\mathcal{H}_0 : \delta = 0$ versus $\mathcal{H}_2 : \delta > 0$, what is the maximum support in favor of $\mathcal{H}_0$ that could possibly be obtained, given the present number of subjects, and given that the average rate of correct card sorts is 65%?

Exercise 9.5.2 What is the maximum support for the earlier unrestricted test of $\mathcal{H}_0 : \delta = 0$ versus $\mathcal{H}_1 : \delta \neq 0$?