11 Signal detection theory

11.1 Signal detection theory

Signal detection theory (SDT: see D. M. Green & Swets, 1966; MacMillan & Creelman, 2004, for detailed treatments) is a very general, useful, and widely employed method for drawing inferences from data in psychology. It is particularly applicable to two-alternative forced choice experiments, although it can be applied to any situation that can be conceived as a 2 × 2 table of counts.

Table 11.1 gives the basic data and terminology for SDT. There are "signal" trials and "noise" trials, and "yes" responses and "no" responses. When a "yes" response is given for a signal trial, it is called a "hit." When a "yes" response is given for a noise trial, it is called a "false alarm." When a "no" response is given for a signal trial, it is called a "miss." When a "no" response is given for a noise trial, it is called a "correct rejection."

The basic data for an SDT analysis are just the counts of hits, false alarms, misses, and correct rejections. It is common to consider just the hit and false alarm counts which, together with the total number of signal and noise trials, completely describe the data.

Table 11.1 Basic signal detection theory data and terminology.

	Signal trial	Noise trial
Yes response	Hit	False alarm
No response	Miss	Correct rejection

The key assumptions of SDT are shown in Figure 11.1, and involve representation and decision-making assumptions. Representationally, the idea is that signal and noise trials can be represented as values along a uni-dimensional "strength" construct. Both types of trials are assumed to produce strengths that vary according to a Gaussian distribution along this dimension. The signal strengths are assumed to be greater, on average, than the noise strengths, and so the signal strength distribution has a greater mean. In the most common equal-variance form of SDT, both the distributions are assumed to have the same variance. The decision-making assumption of SDT is that yes and no responses are produced by comparing the

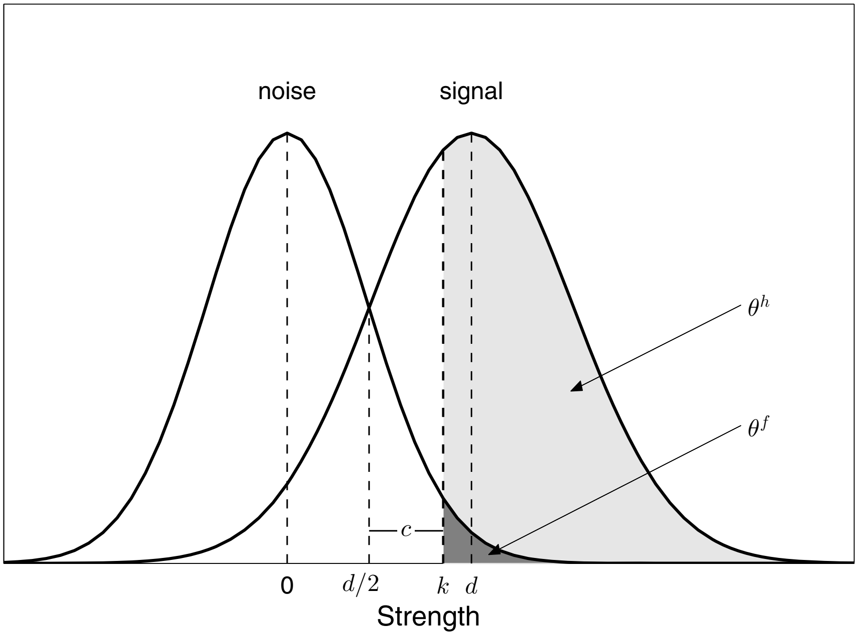


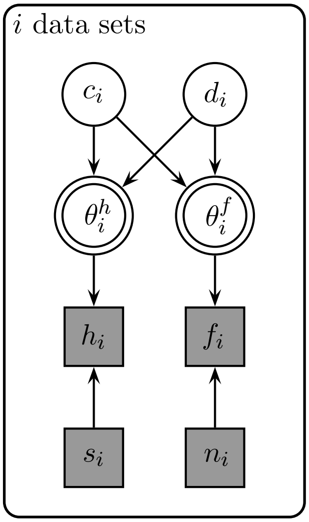
Fig. 11.1 Equal-variance Gaussian signal detection theory framework.

strength of the current trial to a fixed criterion. If the strength exceeds the criterion a "yes" response is made, otherwise a "no" response is made.

Figure 11.1 provides a formal version of the equal-variance SDT model. Since the underlying strength scale has arbitrary units, the variances are fixed to one, and the mean of the noise distribution is set to zero. The mean of the signal distribution is d. This makes d a measure of the *discriminability* of the signal trials from the noise trials, because it corresponds to the distance between the two distributions.

The strength value $d/2$ is special, because it is the criterion value that maximizes the probability of a correct classification when signal and noise trials are equally likely to occur. In this sense, using a criterion of $d/2$ corresponds to unbiased responding. The actual criterion used for responding is denoted k, and the distance between this criterion and the unbiased criterion is denoted c. This makes c a measure of *bias*, because it corresponds to how different the actual criterion is from the unbiased one. Positive values of c correspond to a bias towards saying no, and so to an increase in correct rejections at the expense of an increase in misses. Negative values of c correspond to a bias towards saying yes, and so to an increase in hits at the expense of an increase in false alarms.

The SDT model, with its representation and decision-making assumptions, makes predictions about hit rates and false alarm rates, and so maps naturally onto the counts in Table 11.1. In Figure 11.1, the hit rate, θ^h, is shown as the proportion of the signal distribution above the criterion k. Similarly, the false alarm rate, θ^f, is the proportion of the noise distribution above the criterion k.



$$\begin{aligned}
d_i &\sim \text{Gaussian}(0, \tfrac{1}{2})\\
c_i &\sim \text{Gaussian}(0, 2)\\
\theta_i^h &\leftarrow \Phi(\tfrac{1}{2}d_i - c_i)\\
\theta_i^f &\leftarrow \Phi(-\tfrac{1}{2}d_i - c_i)\\
h_i &\sim \text{Binomial}(\theta_i^h, s_i)\\
f_i &\sim \text{Binomial}(\theta_i^f, n_i)
\end{aligned}$$

Fig. 11.2 Graphical model for signal detection theory.

The usefulness of SDT is that, through this relationship, it is possible to take the sort of data in Table 11.1 and convert the counts of hits and false alarms into psychologically meaningful measures of discriminability and bias. Discriminability is a measure of how easily signal and noise trials can be distinguished. Bias is a measure of how the decision-making criterion being used relates to the optimal criterion.

A graphical model for inferring discriminability and bias from hit and false alarm counts for a number of data sets is shown in Figure 11.2. The hit rates θ_i^h and false alarm rates θ_i^f for the ith data set follow from the geometry of Figure 11.1. They are functions of their associated discriminabilities d_i and biases c_i, using the cumulative standard Gaussian distribution function $\Phi(\cdot)$. The observed counts of hits h_i and false alarms f_i are binomially distributed according to the hits and false alarm rates, and the number of signal trials s_i and noise trials n_i. The priors for discriminability and bias are both Gaussian distributions, constructed to correspond to uniform prior distributions over the hit and false alarm rates.[1]

The script `SDT_1.txt` implements the graphical model in WinBUGS:

```
# Signal Detection Theory
model{
  for (i in 1:k){
  # Observed counts
    h[i] ~ dbin(thetah[i],s[i])
    f[i] ~ dbin(thetaf[i],n[i])
```

[1] The proof relies on the probability integral transform theorem (e.g., Angus, 1994), which says that if X is a continuous real-valued random variable with strictly increasing cumulative distribution function F_X, then $F_X(X)$ is uniformly distributed on $(0,1)$. This means, for example, that if $X \sim \text{Gaussian}(0,1)$ then $\Phi(X) \sim \text{Uniform}(0,1)$.

```
    # Reparameterization Using Equal-Variance Gaussian SDT
    thetah[i] <- phi(d[i]/2-c[i])
    thetaf[i] <- phi(-d[i]/2-c[i])
    # These Priors over Discriminability and Bias Correspond
    # to Uniform Priors over the Hit and False Alarm Rates
    d[i] ~ dnorm(0,0.5)
    c[i] ~ dnorm(0,2)
  }
}
```

The code SDT_1.m or SDT_1.R applies the model to make inferences for three illustrative data sets. Figure 11.3 shows the results produced, plotting the posterior distributions for discriminability, bias, hit rate, and false alarm for each data set.

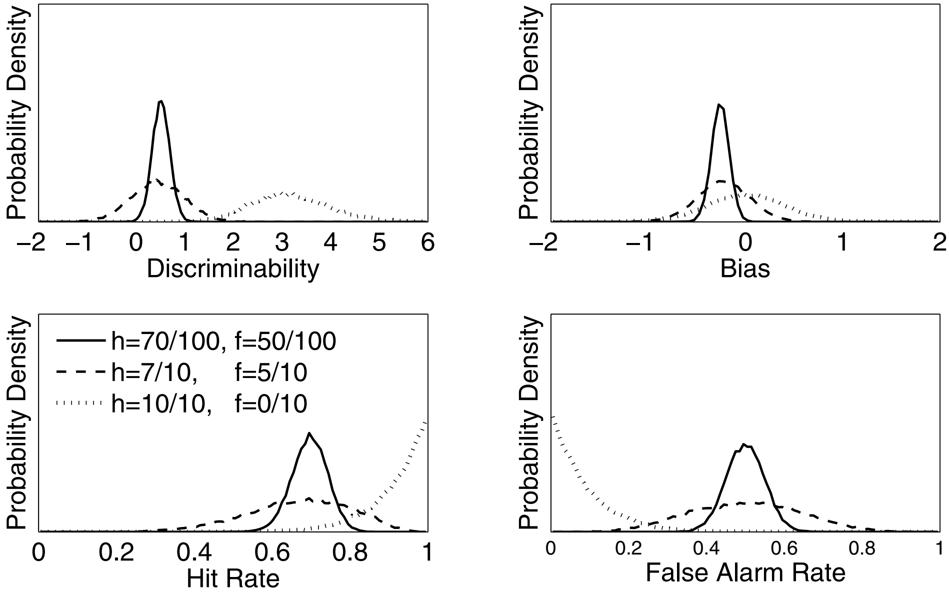


Fig. 11.3 Posterior distributions for discriminability, bias, hit rate, and false alarm rate using three illustrative data sets.

In the first data set, 70 hits and 50 false alarms are observed in 100 target and 100 noise trials. Because of the large number of trials, there is relatively little uncertainty surrounding the hit and false alarm rates, with narrow posteriors centered on 0.7 and 0.5, respectively. Discriminability and bias are also known with some certainty, centered on about 0.5 and −0.25 respectively.

In the second data set, 7 hits and 5 false alarms are observed in 10 target and 10 noise trials. These are the same proportions of hits and false alarms as the first situation, but based on many fewer samples. Accordingly, the posterior distributions have (essentially) the same means, but show much greater uncertainty.

In the third data set, perfect performance is observed, with 10 hits and no false alarms in 10 target and 10 noise trials. The modal hit and false alarm rates are 1.0 and 0.0, but other possibilities have some density. Discriminability is certain to be

Table 11.2 Recognition memory for odors reported by Lehrner et al. (1995).

	Control group		Group I		Group II	
	Old odor	New odor	Old odor	New odor	Old odor	New odor
Old resp.	148	29	150	40	150	51
New resp.	32	151	30	140	40	139

large, although the exact value is not clear. These data provide no information to help estimate bias, and so it retains its prior distribution. This outcome contrasts favorably with traditional frequentist analyses, which have to employ ad hoc edge corrections to avoid both discriminability and bias being undefined when either no hits or no false alarms are observed.

Exercises

Exercise 11.1.1 Do you feel that the priors on discriminability and bias are plausible, a priori? Why or why not? Try out some alternative priors and study the effect that this has on your inference for the data sets discussed above.

Exercise 11.1.2 Lehrner, Kryspin-Exner, and Vetter (1995) report data on the recognition memory for odors of three groups of subjects. Group I had 18 subjects, all with positive HIV antibody tests, and CD-4 counts of 240–700/mm^3. Group II had 19 subjects, all also with positive HIV antibody tests, but with CD-4 counts of 0–170/mm^3. The CD-4 count is a measure of the strength of the immune system, with a normal range being 500–700/mm^3, so Group II subjects had weaker immune systems. Group III had 18 healthy subjects and functioned as a control group. The odor recognition task involved each subject being presented with 10 common household odors to memorize, with a 30-sec. interval between each presentation. After an interval of 15 min., a total of 20 odors were presented to subjects. This test set comprised the 10 previously presented odors, and 10 new odors, presented in a random order. Subjects had to decide whether each odor was “old” or “new.” The signal detection data that resulted are shown in Table 11.2.[2] Analyze these three data sets using signal detection theory to infer the discriminability and bias of the recognition performance for each group. What conclusions do you draw from this analysis? What, if anything, can you infer about individual differences between the subjects in the same groups?

[2] One or two of the counts might be out by one, because these data have been recovered from hit and false alarm rates truncated at two decimal places.

11.2 Hierarchical signal detection theory

We now consider a hierarchical extension of SDT, applied to a different problem where individual subject data are available. This allows us to model possible individual differences using a hierarchical extension of the basic SDT model in the previous case study. The idea is that different subjects have different discriminabilities and biases that are drawn from group-level Gaussian distributions.

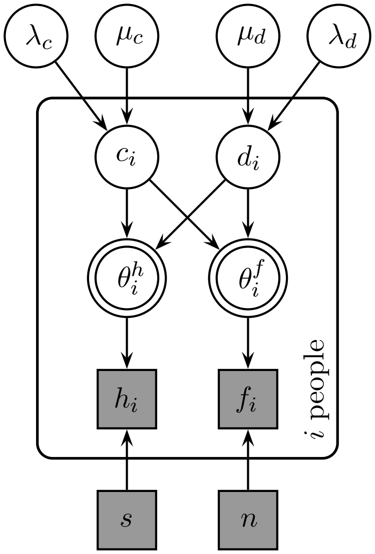
The data come from the empirical evaluation, presented by Heit and Rotello (2005), of a conjecture made by Rips (2001) that inductive and deductive reasoning can be unified within a signal detection theory framework. The conjecture involves considering the strength of an argument as a uni-dimensional construct, but allowing different criteria for induction and deduction. The criterion separates between "weak" and "strong" arguments in the inductive case, and 'invalid" and "valid" arguments in the deductive case, with the deductive criterion being more extreme. Under this conception, deduction is simply a more stringent form of induction. Accordingly, empirical evidence for or against the SDT model has strong implications for the many-threaded contemporary debate over the existence of different kinds of reasoning systems or processes (e.g., Chater & Oaksford, 2000; Heit, 2000; Parsons & Osherson, 2001; Sloman, 1998).

In their study, Heit and Rotello (2005) tested the inductive and deductive judgments of 80 participants on eight arguments. They used a between-subjects design, so that 40 subjects were asked induction questions about the arguments (i.e., whether the conclusion was "plausible"), while the other 40 participants were asked deduction questions (i.e., whether the conclusion was "necessarily true"). These decisions made by participants have a natural characterization in term of the hit and false alarm counts.

A graphical model for analyzing the Heit and Rotello (2005) data is shown in Figure 11.4. It uses SDT to infer the discriminability d_i and bias c_i from hit θ_i^h and false alarm counts θ_i^f and for the ith subject. Individual differences are modeled hierarchically by assuming the individual discriminabilities and biases come from Gaussian group-level distributions, with means and precisions given by μ_d, μ_c, λ_d, and λ_c, respectively.

The script SDT_2.txt implements the graphical model in WinBUGS:

```
# Hierarchical Signal Detection Theory
model{
  for (i in 1:k){
    # Observed counts
    h[i] ~ dbin(thetah[i],s)
    f[i] ~ dbin(thetaf[i],n)
    # Reparameterization Using Equal-Variance Gaussian SDT
    thetah[i] <- phi(d[i]/2-c[i])
    thetaf[i] <- phi(-d[i]/2-c[i])
    # Discriminability and Bias
    c[i] ~ dnorm(muc,lambdac)
    d[i] ~ dnorm(mud,lambdad)
```



$$\begin{aligned}
\mu_d, \mu_c &\sim \text{Gaussian}(0, .001)\\
\lambda_d, \lambda_c &\sim \text{Gamma}(.001, .001)\\
d_i &\sim \text{Gaussian}(\mu_d, \lambda_d)\\
c_i &\sim \text{Gaussian}(\mu_c, \lambda_c)\\
\theta_i^h &\leftarrow \Phi(\tfrac{1}{2}d_i - c_i)\\
\theta_i^f &\leftarrow \Phi(-\tfrac{1}{2}d_i - c_i)\\
h_i &\sim \text{Binomial}(\theta_i^h, s)\\
f_i &\sim \text{Binomial}(\theta_i^f, n)
\end{aligned}$$

Fig. 11.4 Graphical model for hierarchical signal detection theory.

```
  }
  # Priors
  muc ~ dnorm(0,.001)
  mud ~ dnorm(0,.001)
  lambdac ~ dgamma(.001,.001)
  lambdad ~ dgamma(.001,.001)
  sigmac <- 1/sqrt(lambdac)
  sigmad <- 1/sqrt(lambdad)
}
```

The code SDT_2.m or SDT_2.R applies the model to the Heit and Rotello (2005) data. It applies the graphical model separately to the individual data for each experimental condition. Having done this, it produces a display of the joint posterior over the group means for both discriminability and bias, for both experimental conditions, as shown in Figure 11.5.

Exercises

Exercise 11.2.1 Of key interest for testing the Rips (2001) conjecture is how the group-level means for bias and (especially) discriminability differ between the induction and deduction conditions. What conclusion do you draw about the Rips (2001) conjecture base on the current analysis of the Heit and Rotello (2005) data?

Exercise 11.2.2 Heit and Rotello (2005) used standard significance testing methods on their data to reject the null hypothesis that there was no difference between discriminability for induction and deduction conditions. Their analy-

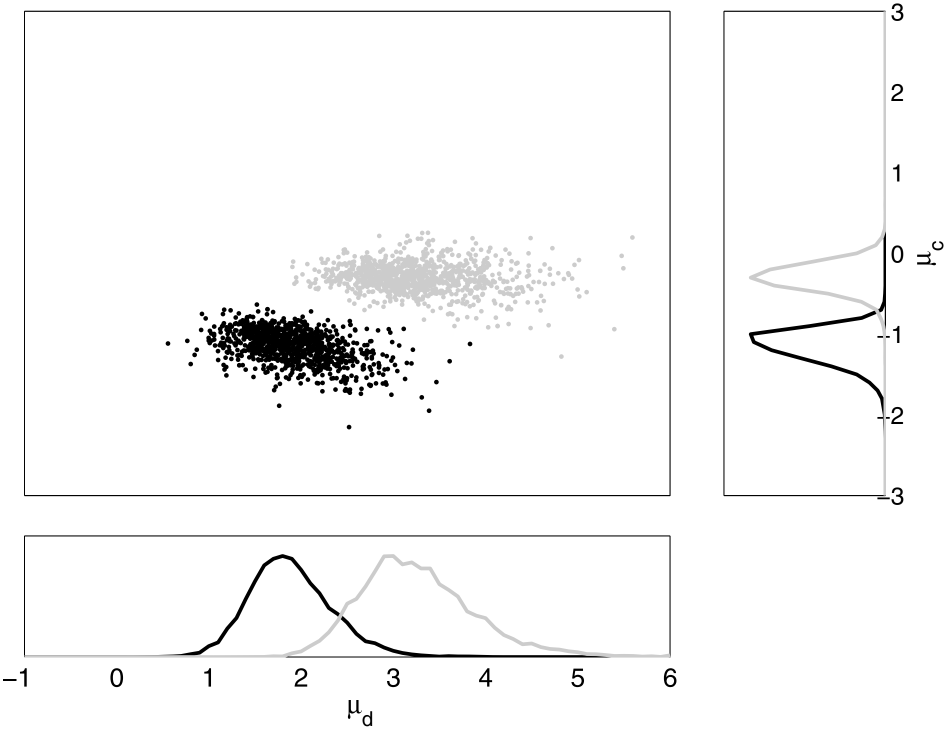


Fig. 11.5 The joint posterior over μ_d and μ_c for the induction (dark) and deduction (light) conditions.

sis involved calculating the mean discriminabilities for each participant, using edge-corrections where perfect performance was observed. These sets of discriminabilities gave means of 0.93 for the deduction condition and 1.68 for the induction condition. By calculating via the t statistic, and so assuming associated Gaussian sampling distributions, and observing that the p-value was less than 0.01, Heit and Rotello (2005) rejected the null hypothesis of equal means. According to Heit and Rotello (2005), this finding of different discriminabilities provided evidence against the criterion-shifting uni-dimensional account offered by SDT. Is this consistent with your conclusions from the Bayesian analysis?

Exercise 11.2.3 Re-run the analysis without discarding burn-in samples. This can be done by setting `nburnin` to 0 in the code `SDT_2.m` or `SDT_2.R`. The result should look something like Figure 11.6. Notice the strange set of samples leading from zero to the main part of the sampled distribution. Explain why these samples exist, and why they suggest burn-in is important in this analysis.

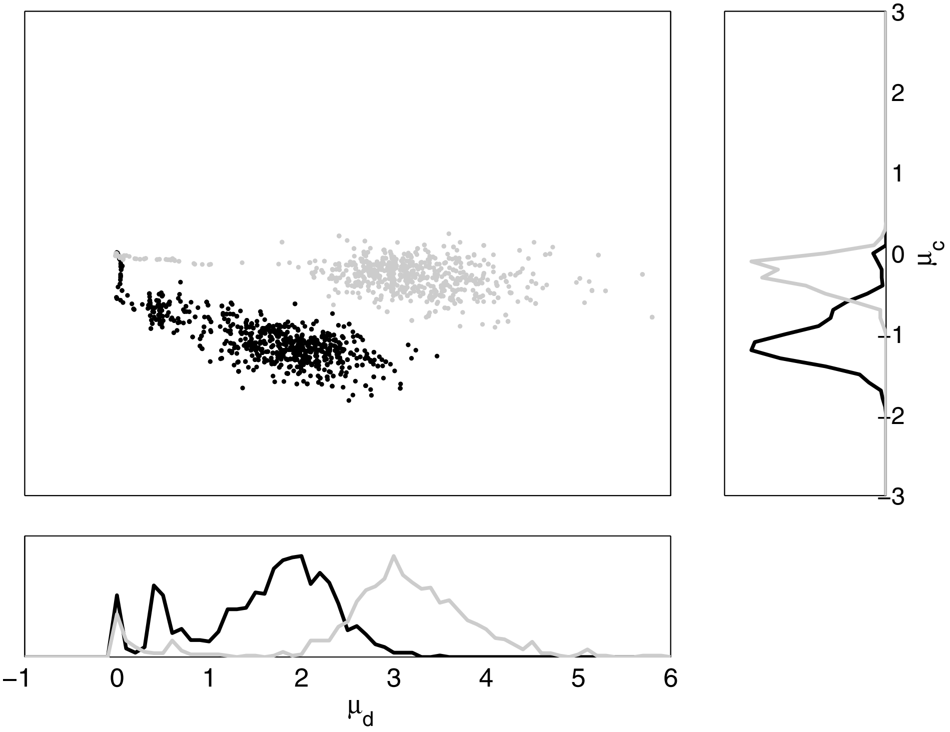


Fig. 11.6 The joint posterior over μ_d and μ_c for the induction (dark) and deduction (light) conditions, without discarding burn-in samples.

11.3 Parameter expansion

WITH DORA MATZKE

Even after the introduction of a burn-in period, there is a sampling issue with the way the chains of the hierarchical variance parameter σ_c behave. As shown in Figure 11.7, the σ_c chains can get stuck near zero, with the samples around 2500 in the induction condition providing an especially good example. This undesirable sampling behavior is fairly common in complicated hierarchical Bayesian models.

The problem is as follows. Suppose that σ_c happens to be estimated near zero. As a result, the individual bias parameters c_i will be pooled toward their population mean μ_c. On the next MCMC iteration, σ_c will be estimated again near zero because it depends on the current values of the c_i parameters. Eventually, the chain of σ_c will break out of the "zero variance trap." However, this may require several iterations and the chain may become trapped again later in sampling.

A good way to enable the sampling process to escape the trap is to use a technique known as parameter expansion (e.g., Gelman, 2004; Gelman & Hill, 2007; Liu & Wu, 1999). This technique involves augmenting the original model with redun-

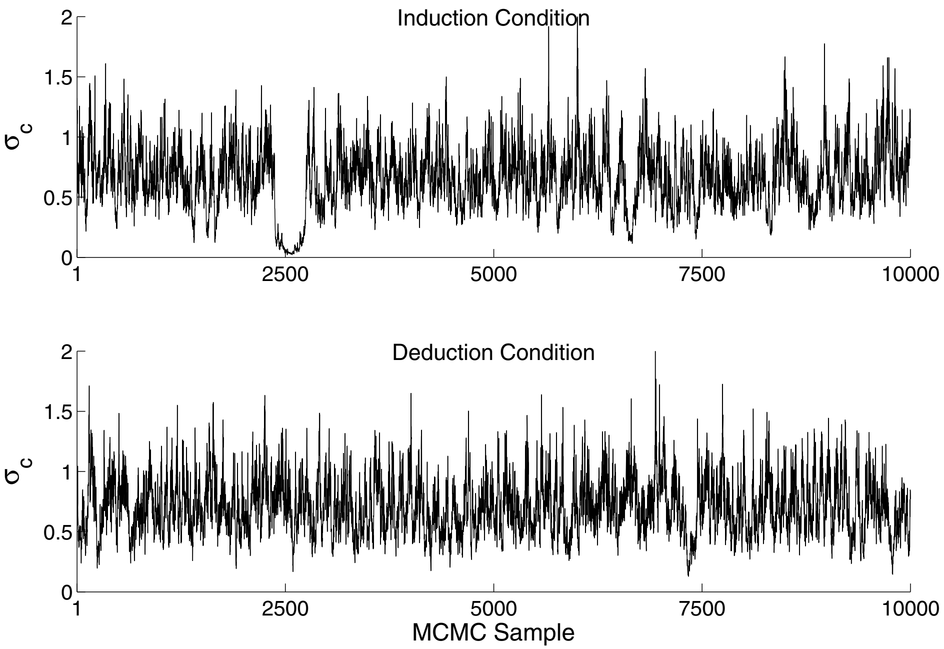


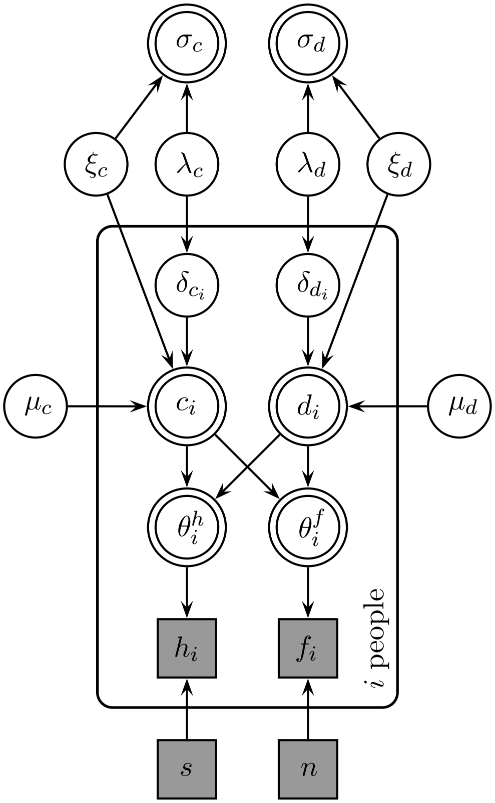
Fig. 11.7 MCMC chains of the σ_c parameter of the hierarchical signal detection model.

dant multiplicative parameters. Specifically, we can extend the hierarchical signal detection model with two multiplicative parameters, say ξ_c and ξ_d. The role of these additional parameters is to rescale the original c_i and d_i parameters and their corresponding standard deviations σ_c and σ_d.

The graphical model that implements this parameter-expanded model is shown in Figure 11.8. Note that the new model is equivalent to the original hierarchical signal detection model, because it simply reparameterizes the original model. In the parameter-expanded model, c_i and σ_c are rescaled by multiplying by ξ_c. The c_i parameter is now expressed in terms of μ_c, ξ_c, and δ_{c_i}, and σ_c is expressed in terms of $|\xi_c|\,\sigma_c^{\text{new}}$. Similarly the d_i and σ_d parameters are rescaled by multiplying them with ξ_d. The d_i parameter is now expressed in terms of μ_d, ξ_d, and δ_{d_i}, and σ_d is expressed in terms of $|\xi_d|\,\sigma_d^{\text{new}}$.

The rationale behind parameter expansion is that updating the ξ_c and ξ_d parameters includes an additional random component in the sampling process. This component causes the samples of $\sigma_c = |\xi_c|\,\sigma_c^{\text{new}}$ and $\sigma_d = |\xi_d|\,\sigma_d^{\text{new}}$ to be less dependent on the previous iteration and prevents the chains getting trapped near zero regardless of how small their previous values were. In order to draw inferences under the original model, however, the parameters from the expanded model must be transformed back to their original scale. For example, inferences about the original σ_c parameter should be based on samples for $|\xi_c|\,\sigma_c^{\text{new}}$ instead of σ_c^{new}.

The script `SDT_3.txt` implements the graphical model in WinBUGS:



$$
\begin{aligned}
\mu_d, \mu_c &\sim \text{Gaussian}(0, .001) \\
\lambda_d, \lambda_c &\sim \text{Gamma}(.1, .1) \\
\xi_d, \xi_c &\sim \text{Beta}(1, 1) \\
\delta_{d_i} &\sim \text{Gaussian}(0, \lambda_d) \\
\delta_{c_i} &\sim \text{Gaussian}(0, \lambda_c) \\
\sigma_d &\leftarrow |\xi_d| \, / \sqrt{\lambda_d} \\
\sigma_c &\leftarrow |\xi_c| \, / \sqrt{\lambda_c} \\
d_i &\leftarrow \mu_d + \xi_d \delta_{d_i} \\
c_i &\leftarrow \mu_c + \xi_c \delta_{c_i} \\
\theta_i^h &\leftarrow \Phi(\tfrac{1}{2} d_i - c_i) \\
\theta_i^f &\leftarrow \Phi(-\tfrac{1}{2} d_i - c_i) \\
h_i &\sim \text{Binomial}(\theta_i^h, s) \\
f_i &\sim \text{Binomial}(\theta_i^f, n)
\end{aligned}
$$

Fig. 11.8 Graphical model for the parameter-expanded hierarchical signal detection theory.

```
# Hierarchical SDT With Parameter Expansion
model{
  for (i in 1:k){
    # Observed counts
    h[i] ~ dbin(thetah[i],s)
    f[i] ~ dbin(thetaf[i],n)
    # Reparameterization Using Equal-Variance Gaussian SDT
    thetah[i] <- phi(d[i]/2-c[i])
    thetaf[i] <- phi(-d[i]/2-c[i])
    # Discriminability and Bias
    c[i] <- muc + xic*deltac[i]
    d[i] <- mud + xid*deltad[i]
    deltac[i] ~ dnorm(0,lambdac)
    deltad[i] ~ dnorm(0,lambdad)
  }
  # Priors
  muc ~ dnorm(0,0.001)
  mud ~ dnorm(0,0.001)
  xic ~ dbeta(1,1)
  xid ~ dbeta(1,1)
  lambdac ~ dgamma(.1,.1)
  lambdad ~ dgamma(.1,.1)
  sigmacnew <- 1/sqrt(lambdac)
  sigmadnew <- 1/sqrt(lambdad)
  sigmac <- abs(xic)*sigmacnew
  sigmad <- abs(xid)*sigmadnew
}
```

The code SDT_3.m or SDT_3.R applies the parameter-expanded model to the Heit and Rotello (2005) data. After you run the code, WinBUGS should display MCMC chains similar to those shown in Figure 11.9. The chains for the variance parameter σ_c now seem to have escaped the zero variance trap.

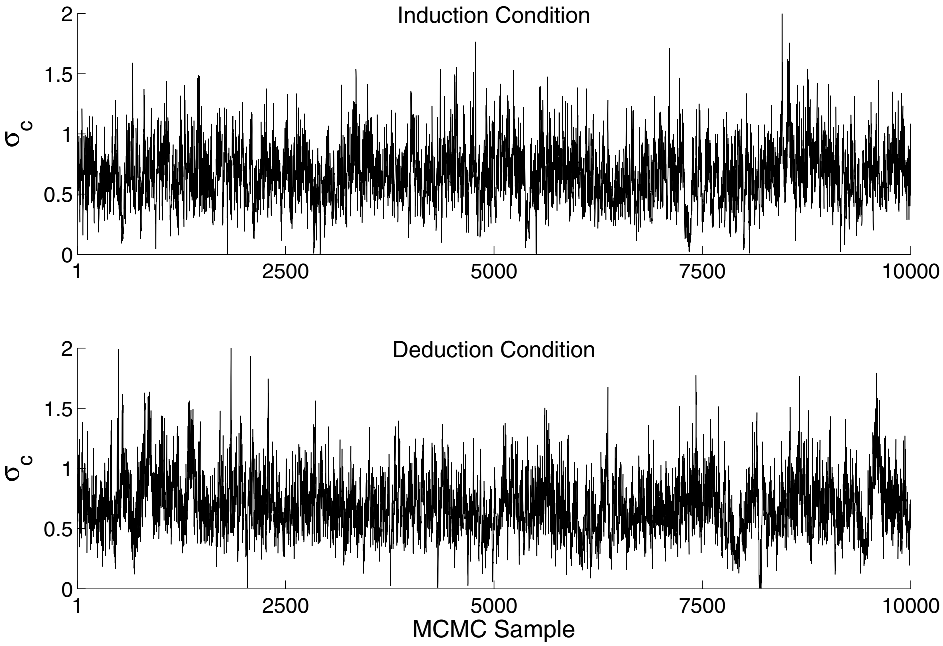


Fig. 11.9 MCMC chains of the σ_c parameter of the parameter-expanded hierarchical signal detection model.

Exercise

Exercise 11.3.1 Experiment with different priors for the unscaled precision, such as dgamma(0.1,0.1), dgamma(0.01,0.01), or dgamma(0.001,0.001), and for the scaling parameters, such as dunif(0,1), dunif(0,2), or dnorm(0,1). How does the prior for the scaled standard deviations change when you change the scaling factor?