Technical Note Accompanying "Group Decision-Making on an Optimal Stopping Problem"

Michael D. Lee
Department of Cognitive Sciences
University of California, Irvine

Michael J. Paradowski
Command and Control Division
Defence Science and Technology Organisation

Abstract

This note provides formal details for the hierarchical signal detection theory used in "Group Decision-Making on an Optimal Stopping Problem".

A Hierarchical Signal Detection Model

The hierarchical signal detection model involves three levels. At the bottom data-level are the observed counts of hits, false alarms, misses and correct rejections for every decision-maker. At the middle signal-detection level is a distribution over discriminability and bias parameters, derived from distributions over hit and false alarm rates. At the top-most group level is distribution over the means and variances for both the discriminability and bias dimensions of an uncorrelated bi-variate Gaussian group distribution.

Group Level

Formally, we denote the mean and precision (i.e., the reciprocal of the variance) of the group-level Gaussian for d' as $\mu^{d'}$ and $\lambda^{d'}$, and similarly for c we have μ^c and λ^c. To perform Bayesian inference, we need to make prior assumptions about the distributions of these means and precisions. The standard approach is to use what are known as conjugate priors (see, for example Bernardo & Smith, 1994, pp. 436–442), which simplifies the computations involves.

In this case , conjugacy requires that the priors for the means of d' and c are Gaussians, themselves having means $\phi^{d'}$ and ϕ^c and precisions $\gamma^{d'}$ and γ^c, giving

$$\begin{aligned} \mu^{d'} \mid \phi^{d'}, \gamma^{d'} &\sim \text{Gaussian}\big(\phi^{d'}, \gamma^{d'}\big) \\ \mu^{c} \mid \phi^{c}, \gamma^{c} &\sim \text{Gaussian}\big(\phi^{c}, \gamma^{c}\big). \end{aligned}$$

Conjugate priors for the precisions of d' and c are Gamma distributions, with shapes $\xi^{d'}$

and ξ^c and scales of $\omega^{d'}$ and ω^c, so that

$$\begin{aligned}\lambda^{d'} \mid \xi^{d'}, \omega^{d'} &\sim \text{Gamma}\big(\xi^{d'}, \omega^{d'}\big) \\ \lambda^{c} \mid \xi^{c}, \omega^{c} &\sim \text{Gamma}\big(\xi^{c}, \omega^{c}\big).\end{aligned}$$

Throughout all the simulations reported, the values $\phi^{d'} = 0$, $\phi^{d'} = 0$, $\gamma^{d'} = .033$, $\gamma^c = .067$, $\xi^{d'} = 3/2$ and $\xi^c = 3/2$, $\omega^{d'} = .033$ and $\omega^c = .067$ were used. These values were chosen to provide proper approximations to the theoretically justified improper 'non-informative' prior distributions $\mu^{d'} \propto 1$, $\mu^c \propto 1$, $\lambda^{d'} \propto \sqrt{\lambda^{d'}}$ and $\lambda^c \propto \sqrt{\lambda^c}$, which would correspond to choosing $\gamma^{d'} = \gamma^c = \omega^{d'} = \omega^c = 0$, while still allowing for reasonably efficient sampling. As a check for the robustness of all modeling conclusions to these prior assumptions the values of $\gamma^{d'}$, γ^c, $\omega^{d'}$, and ω^c were all halved and doubled, and observed to have negligible quantitative effects on the results.

Signal Detection Level

A novel feature of our hierarchical model is that we allow prior information to be introduced directly into the sampling of the discriminability and bias values at the signal detection level. This is very naturally done in the hit and false alarm parameterization. Formally, again using conjugacy, the prior for hit θ_h and false-alarm θ_f rates are Beta distributions, with 'prior samples' for the ith subject α_i^h, α_i^f, α_i^m, and α_i^r, so that

$$\begin{aligned}\theta_i^h &\sim \text{Beta}\big(\alpha_i^h, \alpha_i^m\big) \\ \theta_i^f &\sim \text{Beta}\big(\alpha_i^f, \alpha_i^r\big).\end{aligned}$$

Throughout all the analyses reported, we assume a independent uniform prior distributions over the hit and false alarm rates, by using $\alpha_i^h = \alpha_i^m = \alpha_i^f = \alpha_i^r = 1$. This corresponds to assuming simply that participants know they must make both accept and reject decisions (see Jaynes, 2003, pp. 385).

Posterior Sampling

The basic approach is having observed the data, which are counts for the ith subject, $D_i = (h_i, f_i, m_i, r_i)$, to use a standard Bayesian method known as a Gibbs sampler (e.g., Gelman, Carlin, Stern, & Rubin, 2004; Gilks, Richardson, & Spiegelhalter, 1996; Mackay, 2003) to make inferences about the discriminability and bias distributions of each decision-maker, and then use these distributions over collections of decision-makers to make inference about the mean and variance of their Gaussian distribution.

The Gibbs sampler operates by sweeping repeatedly through the various parameters of the model, sampling each conditionally dependent on the current values of the other. In our sampling scheme, we uses an efficient form of collapsed Gibbs sampling (Chen, Shao, & Ibrahim, 2000) that also facilitates the use of prior information at the signal detection level, by first sampling hit and false alarm rates for each subject

$$\begin{aligned}\theta_i^h \mid D_i &\sim \text{Beta}\big(\alpha_i^h + h_i, \alpha_i^m + m_i\big) \\ \theta_i^f \mid D_i &\sim \text{Beta}\big(\alpha_i^f + f_i, \alpha_i^r + r_i\big),\end{aligned}$$

and converting them to d'_i and c_i values

$$
\begin{aligned}
d'_i &= \Phi^{-1}\left(\theta^h_i\right) - \Phi^{-1}\left(\theta^f_i\right) \\
c &= -\frac{1}{2}\left[\Phi^{-1}\left(\theta^h_i\right) + \Phi^{-1}\left(\theta^f_i\right)\right],
\end{aligned}
$$

where $\Phi(x) = 1/\sqrt{2\pi}\int_0^x \exp\left(-t^2/2\right) \mathrm{d}t$. We then update the means and precision at the group level, according to

$$
\begin{aligned}
\mu^{d'} \mid \lambda^{d'}, \phi^{d'}_k, \gamma^{d'}_k &\sim \text{Gaussian}\Big(\frac{\phi^{d'}_k \gamma^{d'}_k + \lambda^{d'} \sum_{i=1}^m d'_i}{\gamma^{d'} + m\lambda^{d'}}, \gamma^{d'}_k + m\lambda^{d'}\Big) \\
\mu^{c} \mid \lambda^{c}, \phi^{c}_k, \gamma^{c}_k &\sim \text{Gaussian}\Big(\frac{\phi^{c}_k \gamma^{c}_k + \lambda^{c} \sum_{i=1}^m c_i}{\gamma^{c}_k + m\lambda^{c}}, \gamma^{c} + m\lambda^{c}\Big) \\
\lambda^{d'} \mid \mu^{d'}, \xi^{d'}, \omega^{d'} &\sim \text{Gamma}\Big(\xi^{d'} + \frac{1}{2}m, \omega^{d'} + \frac{1}{2}\sum_{i=1}^m (d'_i - \mu^{d'})^2\Big) \\
\lambda^{c} \mid \mu^{c}, \xi^{c}, \omega^{c} &\sim \text{Gamma}\Big(\xi^{c} + \frac{1}{2}m, \omega^{c} + \frac{1}{2}\sum_{i=1}^m (c_i - \mu^{c})^2\Big).
\end{aligned}
$$

In each of the analyses reported, we conducted 10^4 sweeps. After an initial 'burnin' period, set to be 10^2, these samples of the parameter values come from the full joint posterior distribution. From this distribution, all of the standard quantities and distributions of interest—such as expectations for the means and variances at the group level, or marginal distributions over any of the parameters—can be calculated automatically.

Bayes Factors

Besides inferences involving parameter values, the Bayesian approach allows for model or hypothesis testing using a form of likelihood ratios known as Bayes Factors (e.g., Kass & Raftery, 1995). These ratios are essentially likelihoods that integrate out (average over) all of the free parameters in competing models, and so automatically take into account both the goodness-of-fit and complexity of the models (Pitt, Myung, & Zhang, 2002).

The Bayes Factors we report compare a 'same' model M_s that assumes no difference between two sets of decision making data D_1 and D_2, with a 'different' model M_d that does assume change, and so requires separate parameterizations for both data sets. This means that M_s has parameters $\boldsymbol{\theta_s} = (\mu^{d'}, \mu^c, \lambda^{d'}, \lambda^c)$ while M_d has parameters $\boldsymbol{\theta_d} = (\mu^{d'}_1, \mu^c_1, \lambda^{d'}_1, \lambda^c_1, mu^{d'}_2, \mu^c_2, \lambda^{d'}_2, \lambda^c_2)$.

The Bayes Factor, BF, is the ratio

$$
\mathrm{BF} = \frac{p(D \mid M_s)}{p(D \mid M_d)} = \frac{\int p(D \mid \boldsymbol{\theta_s}, M_s) p(\boldsymbol{\theta_s} \mid M_s)\ \mathrm{d}\boldsymbol{\theta_s}}{\int p(D \mid \boldsymbol{\theta_d}, M_d) p(\boldsymbol{\theta_d} \mid M_d)\ \mathrm{d}\boldsymbol{\theta_d}},
$$

and can be computed by measuring the likelihood function for a set of N samples from the posterior distributions, $p(\boldsymbol{\theta_s} \mid D)$ and $p(\boldsymbol{\theta_d} \mid D)$, as provided by the Gibbs sampler. This likelihood is

$$
L_s = \prod_{i=1}^m \frac{\sqrt{\lambda^{d'}\lambda^c}}{2\pi} \exp\left\{-\frac{1}{2}\left[\lambda^{d'}\left(d'_i - \mu^{d'}\right)^2 + \lambda^c\left(c_i - \mu^c\right)^2\right]\right\}
$$

for the same model, where there are m decision makers in the data set D , and

$$L_d = \prod_{j=1}^{2} \prod_{i=1}^{m} \frac{\sqrt{\lambda_j^{d'} \lambda_j^{c}}}{2\pi} \exp\left\{ -\frac{1}{2} \left[\lambda_j^{d'} \left(d'_{ij} - \mu^{d'_j} \right)^2 + \lambda_j^{c} \left(c_{ij} - \mu_j^{c} \right)^2 \right] \right\}$$

for the different model, where there are m decision makers in both the data set D_1 and D_2, and d'_{ij} and c_{ij} now denote the discriminability and bias for the ith member of the jth set.

The marginal probabilities required to calculate the Bayes Factor are then approximated by the harmonic means of the set of likelihood values sampled from the posterior (see Raftery, 1996). Using an extension of their notation, the approximation is

$$\mathrm{BF} \approx \frac{\left\langle 1 / \left\| \{L_s\} \right\|_{\mathrm{post}} \right\rangle^{-1}}{\left\langle 1 / \left\| \{L_d\} \right\|_{\mathrm{post}} \right\rangle^{-1}},$$

where $\langle \cdot \rangle$ denotes the expectation.

References

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory*. New York: Wiley.

Chen, M. H., Shao, Q. M., & Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian computation*. New York: Spinger-Verlag.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (Second ed.). London: Chapman and Hall.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London: Chapman and Hall.

Jaynes, E. T. (2003). *Probability theory: The logic of science* (G. L. Bretthorst, Ed.). New York: Cambridge University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

Mackay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge: Cambirdge University Press.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*(3), 472–491.

Raftery, A. E. (1996). Hypothesis testing and model selection. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 163–187). Boca Raton (FL): Chapman & Hall/CRC.