Signal detection theory and vestibular perception: III. Estimating unbiased fit parameters for psychometric functions

Experimental Brain Research

Shomesh E. Chaudhuri, Daniel M. Merfeld

S. E. Chaudhuri, D. M. Merfeld
Jenks Vestibular Physiology Laboratory,
Massachusetts Eye and Ear Infirmary,
Boston, MA, USA
e-mail: shomesh_chaudhuri@meei.harvard.edu

D. M. Merfeld (corresponding author)
Otology and Laryngology,
Harvard Medical School,
Boston, MA, USA
e-mail: dan_merfeld@meei.harvard.edu
Telephone: (617) 573-5595
Fax: (617) 573-5596

Online Resource 1

As detailed in the main body, the methods and notation used to subtract the order n^{-1} asymptotic bias term from the maximum likelihood estimate were first described by McCullagh (1987, 1989). An alternative approach was later developed by Firth (1993) who removed the order n^{-1} bias by introduction of an appropriate shift in the score function. Firth's approach is especially useful for models where there is a finite probability that the estimate is on a boundary of the parameter space and/or where the bias-corrected estimator is undefined (Kosmidis and Firth, 2010). Neither concern seems relevant for our application, which is one reason why we chose to focus on the bias-correction approach investigated and reported in the main body.

To the best of our knowledge, the modified score approach has not been tried for bias-corrected psychometric function fits in the 20 years since it was introduced. Since there is no fundamental reason that this approach could not be applied to our probit model we show summary calculations and findings below. We also provide simulation results for direct bias reduction on μ and σ.

To implement this approach we can calculate the bias vector by modifying the score function (U) which is the gradient of the log-likelihood function (Firth 1993). The log-likelihood function for our binary response model is:

$$l(\theta_1, \theta_2; \mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \left[\log \left(y_i \psi(x_i; \theta_1, \theta_2) + (1 - y_i)\left(1 - \psi(x_i; \theta_1, \theta_2)\right) \right) \right]$$

where $\mathbf{x}$ is the stimulus vector, $\mathbf{y}$ is the binary response vector and (θ_1, θ_2) parameterizes the psychometric function ψ.

The maximum likelihood estimate is the solution to $U = \nabla l = 0$. Similarly, the bias-reduced maximum likelihood estimate is the solution $U + A = 0$ where A is a correction function based either on the expected or observed information matrix (Firth 1993). During simulations we observed insignificantly small differences between using the observed or expected information matrices to calculate A (Fig. S1), and so we proceed using the expected information matrix as it can be directly related to the BRGLM estimation technique and is simpler to calculate.

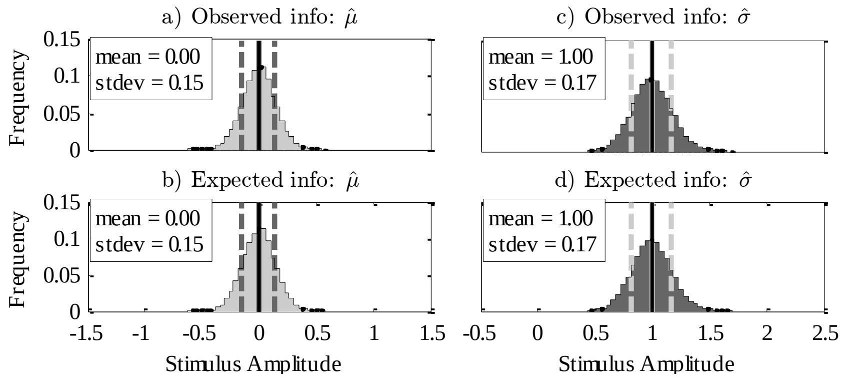

Fig. S1 Bias reduction using the observed or expected information matrices with $(\theta_1, \theta_2) = (b_1, b_2)$ to calculate A for a 3D/1U staircase and $n = 100$. Panels a) and c) show the histograms for $\hat{\mu}$ and $\hat{\sigma}$ using the observed information matrix. Panels b) and d) show the histograms for $\hat{\mu}$ and $\hat{\sigma}$ using the expected information matrix. The solid black line is the actual parameter value, the solid gray line is the mean of the parameter estimates, and the dashed gray lines indicate one standard deviation either side of the mean

Using summation notation, A is defined as:

$$A_r = \kappa^{u,v}\left(\kappa_{r,u,v} + \kappa_{r,uv}\right)/2$$

where $\kappa_{r,u,v} = n^{-1}\mathrm{E}[U_r U_u U_v]$, $\kappa_{r,uv} = n^{-1}\mathrm{E}[U_r U_{uv}]$, and $\kappa^{u,v}$ denotes the inverse of expected information matrix $\kappa_{u,v}$ (Firth 1993).

Letting $c_i = (2y_i - 1)/\left(y_i\psi_i + (1 - y_i)(1 - \psi_i)\right)$ we find that the score vector has components:

$$U_{\theta_1} = \sum_{i=1}^{n} c_i \frac{\partial \psi_i}{\partial \theta_1} \quad , \quad U_{\theta_2} = \sum_{i=1}^{n} c_i \frac{\partial \psi_i}{\partial \theta_2}$$

so the observed information matrix is:

$$I = -\begin{bmatrix} \sum_{i=1}^{n} c_i \frac{\partial^2 \psi_i}{\partial {\theta_1}^2} - \left(c_i \frac{\partial \psi_i}{\partial \theta_1}\right)^2 & \sum_{i=1}^{n} c_i \frac{\partial^2 \psi_i}{\partial \theta_1 \partial \theta_2} - (c_i)^2 \frac{\partial \psi_i}{\partial \theta_1}\frac{\partial \psi_i}{\partial \theta_2} \\ \sum_{i=1}^{n} c_i \frac{\partial^2 \psi_i}{\partial \theta_1 \partial \theta_2} - (c_i)^2 \frac{\partial \psi_i}{\partial \theta_1}\frac{\partial \psi_i}{\partial \theta_2} & \sum_{i=1}^{n} c_i \frac{\partial^2 \psi_i}{\partial {\theta_2}^2} - \left(c_i \frac{\partial \psi_i}{\partial \theta_2}\right)^2 \end{bmatrix}$$

and letting $g_i = \left(\psi_i(1 - \psi_i)\right)^{-1}$ the expected information matrix is:

$$\mathrm{E}[I] = \begin{bmatrix} \sum_{i=1}^{n} \left(\frac{\partial\psi_i}{\partial\theta_1}\right)^2 g_i & \sum_{i=1}^{n} \left(\frac{\partial\psi_i}{\partial\theta_1}\right)\left(\frac{\partial\psi_i}{\partial\theta_2}\right) g_i \\ \sum_{i=1}^{n} \left(\frac{\partial\psi_i}{\partial\theta_1}\right)\left(\frac{\partial\psi_i}{\partial\theta_2}\right) g_i & \sum_{i=1}^{n} \left(\frac{\partial\psi_i}{\partial\theta_2}\right)^2 g_i \end{bmatrix}$$

The remaining quantities required to calculate A can be found below. Once those have been calculated we can then use numeric methods to solve $U + A = 0$ to find the bias-reduced maximum likelihood estimates (Firth 1993).

Letting $h_i = (\psi_i)^{-2} - (1 - \psi_i)^{-2}$, the remaining quantities required to calculate A are the $\kappa_{r,u,v} = n^{-1}\mathrm{E}[U_r U_u U_v]$ terms:

$$\kappa_{\theta_1,\theta_1,\theta_1} = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{\partial\psi_i}{\partial\theta_1}\right)^3 h_i$$

$$\kappa_{\theta_1,\theta_1,\theta_2} = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{\partial\psi_i}{\partial\theta_1}\right)^2 \left(\frac{\partial\psi_i}{\partial\theta_2}\right) h_i$$

$$\kappa_{\theta_1,\theta_2,\theta_2} = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{\partial\psi_i}{\partial\theta_1}\right) \left(\frac{\partial\psi_i}{\partial\theta_2}\right)^2 h_i$$

$$\kappa_{\theta_2,\theta_2,\theta_2} = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{\partial\psi_i}{\partial\theta_2}\right)^3 h_i$$

and the $\kappa_{r,uv} = n^{-1}\mathrm{E}[U_r U_{uv}]$ terms:

$$\kappa_{\theta_1,\theta_1\theta_1} = \frac{1}{n}\sum_{i=1}^{n} \left[\left(\frac{\partial\psi_i}{\partial\theta_1}\right)\left(\frac{\partial^2\psi_i}{\partial{\theta_1}^2}\right) g_i - \left(\frac{\partial\psi_i}{\partial\theta_1}\right)^3 h_i\right]$$

$$\kappa_{\theta_1,\theta_1\theta_2} = \frac{1}{n}\sum_{i=1}^{n} \left[\left(\frac{\partial\psi_i}{\partial\theta_1}\right)\left(\frac{\partial^2\psi_i}{\partial\theta_1\partial\theta_2}\right) g_i - \left(\frac{\partial\psi_i}{\partial\theta_1}\right)^2 \left(\frac{\partial\psi_i}{\partial\theta_2}\right) h_i\right]$$

$$\kappa_{\theta_1,\theta_2\theta_2} = \frac{1}{n}\sum_{i=1}^{n} \left[\left(\frac{\partial\psi_i}{\partial\theta_1}\right)\left(\frac{\partial^2\psi_i}{\partial{\theta_2}^2}\right) g_i - \left(\frac{\partial\psi_i}{\partial\theta_1}\right)\left(\frac{\partial\psi_i}{\partial\theta_2}\right)^2 h_i\right]$$

$$\kappa_{\theta_2,\theta_1\theta_1} = \frac{1}{n}\sum_{i=1}^{n} \left[\left(\frac{\partial^2\psi_i}{\partial{\theta_1}^2}\right)\left(\frac{\partial\psi_i}{\partial\theta_2}\right) g_i - \left(\frac{\partial\psi_i}{\partial\theta_1}\right)^2 \left(\frac{\partial\psi_i}{\partial\theta_2}\right) h_i\right]$$

$$\kappa_{\theta_2,\theta_1\theta_2} = \frac{1}{n}\sum_{i=1}^{n} \left[\left(\frac{\partial\psi_i}{\partial\theta_2}\right)\left(\frac{\partial^2\psi_i}{\partial\theta_1\partial\theta_2}\right) g_i - \left(\frac{\partial\psi_i}{\partial\theta_1}\right)\left(\frac{\partial\psi_i}{\partial\theta_2}\right)^2 h_i\right]$$

$$\kappa_{\theta_2,\theta_2\theta_2} = \frac{1}{n}\sum_{i=1}^{n} \left[\left(\frac{\partial\psi_i}{\partial\theta_2}\right)\left(\frac{\partial^2\psi_i}{\partial{\theta_2}^2}\right) g_i - \left(\frac{\partial\psi_i}{\partial\theta_2}\right)^3 h_i\right]$$

The modified score technique for implementing bias-reduction was used to estimate $\hat{\mu}$ and $\hat{\sigma}$ for a 3D/1U staircase with n =100, μ =0, σ =1 and an initial stimulus of 2. Both the $(\mathrm{b}_1, \mathrm{b}_2)$ parameterization of ψ and the (μ, σ) parameterization of ψ were used to calculate A. Table S1 lists the means and standard deviations on $\hat{\mu}$ and $\hat{\sigma}$ using the modified score technique with these parameterizations. For convenience, the GLM and BRGLM estimates are also listed.

Table S1 $\hat{\mu}$ and $\hat{\sigma}$ means ± (standard deviations) for the GLM, BRGLM and modified score techniques

	GLM	**BRGLM**	**Modified score $(\mathbf{b_1}, \mathbf{b_2})$**	**Modified score $(\boldsymbol{\mu}, \boldsymbol{\sigma})$**
$\hat{\boldsymbol{\mu}}$	0.00±(0.14)	0.00±(0.14)	0.00±(0.14)	0.00±(0.14)
$\hat{\boldsymbol{\sigma}}$	0.97±(0.17)	1.00±(0.18)	1.00±(0.18)	0.97±(0.17)

Category I Bias (<10% of stdev) Category II Bias (10-25% of stdev) Category III Bias (>25% of stdev)

As expected the BRGLM technique and modified score technique using the (b_1, b_2) parameterization produce the same results. Furthermore, to our surprise, bias-reduction directly on μ and σ failed to remove the bias on the $\hat{\sigma}$ estimates. This, alongside results from the paper, suggests that, for this application, the nonlinear transformation from $(\hat{b}_1, \hat{b}_2)$ to $(\hat{\mu}, \hat{\sigma})$ is an important step in producing the unbiased $\hat{\sigma}$ estimates. Similar to Tables 1 and 4 in the paper, the generality of these results for psychometric function fits were demonstrated (not shown) by using multiple adaptive sampling procedures, termination criteria and initial stimuli.