Single Rating Model

For the case of a single rating outcome (case strength), we separately model each group of participants as follows: Given a set of ratings R indexed by $i = 1 \ldots N$, and an $N \times P$ design matrix X (with columns corresponding to regressors $p = 1 \ldots P$), we assume for a given observation i corresponding to subject s and case c:

$$
\begin{aligned}
R_i &\sim \left[\mathcal{N}(\theta_i, \sigma^2)\right]_0^{100} && (1)\\
\theta_i &= X_{i\cdot} \cdot \beta_{s(i)c(i)\cdot} && (2)\\
\beta_{scp} &= \mu_p + \gamma_{cp} + \epsilon_{sp}\tau_{cp} && (3)\\
\gamma_{cp} &\sim \mathcal{T}_\nu(0, \eta_p^2) && (4)\\
\epsilon_{sp} &\sim \mathcal{T}_{\nu'}(0, 1) && (5)\\
\mu_p &\sim \mathcal{N}(50, 50) && (6)\\
\eta_p &\sim \mathrm{Ca}^+(0, 50) && (7)\\
\tau_{cp} &\sim \mathrm{Ca}^+(0, 50) && (8)\\
\sigma &\sim \mathrm{Ca}^+(0, 5) && (9)\\
\nu, \nu' &\sim \mathcal{N}^+(0, 100) && (10)
\end{aligned}
$$

That is:

- ratings are generated from a normal distribution censored to lie in the range $[0, 100]$ (1)
- linear predictors of ratings are weighted sums of subject-, case-, and regressor-specific effects (2)
- ϵ effects for each subject are drawn from regressor-specific distribution (5)
- γ effects for each case are themselves drawn from a regressor-specific distribution of effects (4)
- effects at the case and single-subject level are modeled as robuts/fat-tailed, with Student-t distributions (5,4)
- variances (η^2, τ^2, σ^2) are modeled using weakly informative half-Cauchy priors (Gelman 2006), while degrees of freedom (ν, ν') are modeled using weak half-Normal priors

This approach allows for flexible fitting (including estimates of variance) at the regressor, case, and individual levels, while at the same time leveraging partial pooling to share statistical strength across these levels (Gelman and Hill 2006).

Multiple Rating Model

For the case in which subjects provide multiple ratings (punishment, case strength, etc.) for a given scenario, we model the resulting vector of ratings, R_r, $r =$

$1 \dots N_r$, similarly:

$$
\begin{aligned}
R_i &\sim \left[\mathcal{N}(\theta_i, \sigma^2_{r(i)})\right]_0^{100} && (11)\\
\theta_i &= X_{i\,\cdot} \cdot \beta_{s(i)c(i)\,\cdot\,r(i)} && (12)\\
\beta_{scpr} &= \mu_{p,r} + \gamma_{cpr} + \epsilon_{spr} && (13)\\
\gamma_{cp\,\cdot} &\sim \mathcal{T}_\nu(0, \Sigma_p) && (14)\\
\epsilon_{spr} &\sim \mathcal{T}_{\nu'}(0, \tau^2 pr) && (15)\\
\Sigma_p &= L_p \mathrm{diag}(\eta_p) L_p^\top && (16)\\
\Omega_p = L_p L_p^\top &\sim \mathrm{LKJ}(1) && (17)\\
\mu_{pr} &\sim \mathcal{N}(50, 50) && (18)\\
\eta_{pr} &\sim \mathrm{Ca}^+(0, 50) && (19)\\
\tau_{pr} &\sim \mathrm{Ca}^+(0, 50) && (20)\\
\sigma_r &\sim \mathrm{Ca}^+(0, 5) && (21)\\
\nu, \nu' &\sim \mathcal{N}^+(0, 100) && (22)
\end{aligned}
$$

Here, we have used a "long" or "melted" representation of R in which each index i corresponds to a single observation of a single rating scale $r(i)$. This allows us to more easily handle missing data in the model fitting procedure (see below). The model is almost equivalent to concatenating N_r versions of the first model, one for each rating, aside from two key differences: First (14) and (17) involve a multivariate t-distribution on the population effects specific to each case (γ). That is, we allow for covariance among the ratings for each effect at the population level, where the magnitudes of the variances are again controlled by η and the correlations $\Omega = LL^\top$ are modeled according to the LKJ distribution (Lewandowski, Kurowicka, and Joe 2009) through their Cholesky factorization (17).[1] Second, in order to more accurately estimate variances in the presence of missing data, we have restricted this model to a single value of τ across all cases (for each outcome and regressor) (13).

Gelman, Andrew. 2006. "Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)." *Bayesian Analysis* 1 (3). International Society for Bayesian Analysis: 515–34.

Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge university press.

Lewandowski, Daniel, Dorota Kurowicka, and Harry Joe. 2009. "Generating Random Correlation Matrices Based on Vines and Extended Onion Method." *J. Multivar. Anal.* 100 (9): 1989–2001.

[1]Implemented as `L ~ lkj_corr_chol(1)` in Stan.