

ESTIMATING EXTRA-PAIR PATERNITY IN A HETEROGENOUS SAMPLE

RICHARD MCELREATH

February 26, 2022

The sample comprises mothers with varying numbers of both children and husbands. In principle, each mother and mother-husband pair could have a unique extra-pair paternity (EPP) rate. Estimates of both the average rate and the variation among mothers and mother-husband pairs can be constructed with a multilevel analysis.

Statistical model. Let $y_i \in \{0, 1\}$ be the EPP status of child i . Each child has a corresponding mother with index $\text{MOM}[i]$ and a dyad comprising the mother and the child's social father with index $\text{DYAD}[i]$. Then the probability of y_i is given by:

$$y_i \sim \text{Bernoulli}(p_i)$$
$$\text{logit } p_i = \alpha + z_{\text{MOM}[i]} \sigma + x_{\text{DYAD}[i]} \tau$$

where α is the population average, the vector z contains standardized varying effects for each mother, σ is the standard deviation among mothers, x contains standardized varying effects for each dyad, and τ is the standard deviation among dyads.

The model uses a non-centered parameterization, but is in fact an ordinary multi-level model. The priors are:

$$\begin{aligned}\alpha &\sim \text{Normal}(0, 1.5) \\ z_j &\sim \text{Normal}(0, 1) \\ x_k &\sim \text{Normal}(0, 1) \\ \sigma &\sim \text{Half-Normal}(0, 1) \\ \tau &\sim \text{Half-Normal}(0, 1)\end{aligned}$$

The $\text{Normal}(0, 1.5)$ prior for α establishes a uniform prior on the probability scale and $\text{Half-Normal}(0, 1)$ indicates a normal distribution defined only on positive reals.

DEPARTMENT OF HUMAN BEHAVIOR, ECOLOGY AND CULTURE, MAX PLANCK INSTITUTE FOR EVOLUTIONARY ANTHROPOLOGY, LEIPZIG, GERMANY

E-mail address: richard_mcelreath@eva.mpg.de.

Measurement error. A complication in this analysis is that there is as much as 5% false assignment of paternity. Failing to account for this would artificially inflate our estimate of EPP. We incorporate false-positives into our model by using a mixture likelihood. Given false-positive probability f , the probability of observing a paternity assignment $y_i = 1$ is:

$$\Pr(y_i = 1|p_i) = p_i + f(1 - p_i)$$

Since false negatives are extremely unlikely in our analysis, the probability of observing $y_i = 0$ is:

$$\Pr(y_i = 0|p_i) = (1 - f)(1 - p_i)$$

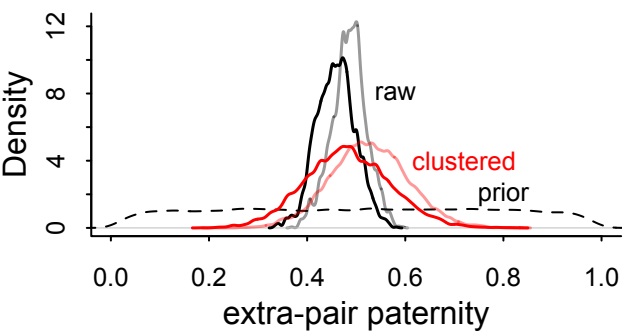
Implementation. We implemented the model in Stan 2.19 (mc-stan.org), a library for Bayesian model construction and fitting. A copy of the script is available at [LINK](#). We drew 8000 samples from 4 chains. All R-hat values were less than 1.01, and there were no divergent transitions or other signs of biased posterior exploration. We show the trace plots for the hyperparameters and a selection of the varying effects on the next page. We also show the same traces as rank histograms (<https://arxiv.org/abs/1903.08008>), which make it much easier to spot problems with chains. Healthy chains, like these, should be relatively uniform.

The marginal posterior distributions for α , σ , and τ are summarized below.

	mean	sd	5.5%	94.5%	n_eff	Rhat
a	-0.07	0.36	-0.63	0.51	6422	1
sigma	0.75	0.55	0.07	1.75	1216	1
tau	2.01	0.54	1.18	2.89	2325	1

On the probability scale, α implies an average posterior EPP rate with mean 0.48 and 95% credible interval from 0.32 to 0.66. Note that there is more variation among dyads (τ) than among mothers (σ). This implies that features of specific pairings explain EPP variation more than features of specific women.

The posterior distribution EPP rate is plotted below in red, against both the prior (dashed) and the naive posterior rate from ignoring clustering by mothers and mother-father dyads (black). We also considered models with no clustering (the black density in the above figure), with clustering only by mother, and clustering only by dyad. All models with clustering produce similar inferences. Code for each, together with WAIC model comparison, can be found in the script ([LINK](#)). For comparison, models which do not include potential false-positives are shown in lighter colors.



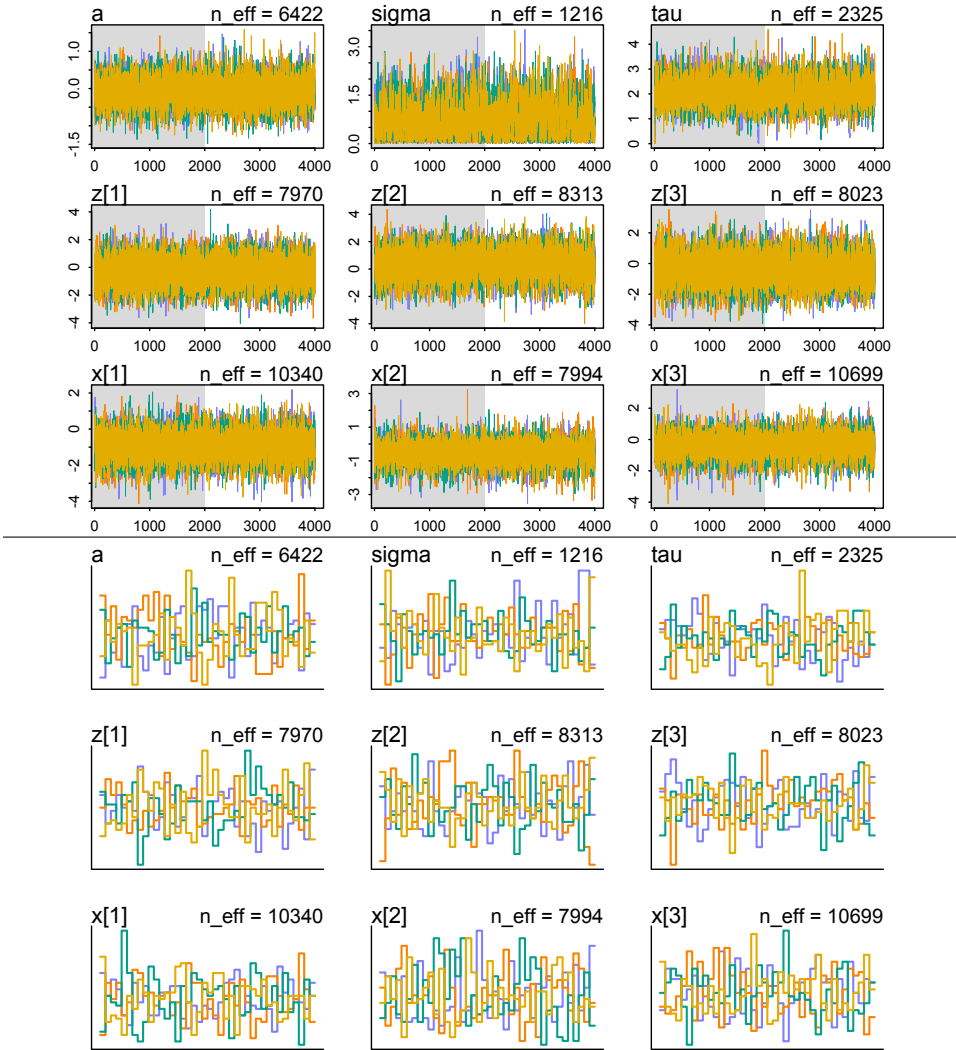


FIGURE 1. Trace plots (top) and rank histograms (bottom) for the model that clusters by mother and mother-husband dyad.