

BETA REPARAMETRIZATION

LS

CONTENTS

1. Current Model: focus on likelihood and prior for P	1
2. Re-parametrise with a Beta and integrate out P	3
2.1. Reparametrization with ν as prior sample size	4
2.2. Reparametrization with mean and variance	4
2.3. Integrating P out	4
3. Allowing for different values on the diagonal	6

1. CURRENT MODEL: FOCUS ON LIKELIHOOD AND PRIOR FOR P

Our current likelihood is:

$$\begin{aligned}
 (1) \quad p(y|z, P, K) &= \prod_{i=1}^{N-1} \prod_{j=i+1}^N p(y_{ij}|z, P, K) \\
 (2) \quad &= \prod_{i=1}^{N-1} \prod_{j=i+1}^N \binom{n_{ij}}{y_{ij}} p_{z_i, z_j}^{y_{ij}} (1 - p_{z_i, z_j})^{n_{ij} - y_{ij}}
 \end{aligned}$$

The prior probability distribution on the P matrix entries p_{pq} is:

$$\begin{aligned}
 (3) \quad p(p_{pq} | \mu_k, \sigma^2, \beta_{max}) &= \\
 (4) \quad &\begin{cases} \text{TruncNorm}(\mu_k, \sigma^2) \mathbb{I}(0.5, \beta_{max}) & \text{for } k = q - p > 0 \\ 0.5 & \text{with } p = 1 & \text{for } k = q - p = 0 \\ 1 - p_{pq} & \text{with } p = 1 & \text{for } k = q - p < 0 \end{cases}
 \end{aligned}$$

where k is the index of the level set $L_{(k)} \subset P$, the collection of the entries satisfying the condition $p - q = k$. In other words, they correspond to the entries on the k -th diagonal above the main diagonal within matrix P . Additionally, we set the probabilities within the main diagonal to be 0.5, reflecting an equal chance of preference if two items belong to the same block. The maximum attainable probability is determined by the user and denoted as β_{max} , where $\beta_{max} < 1$. This constraint is imposed to prevent extreme probabilities, such as values near 0 or 1.

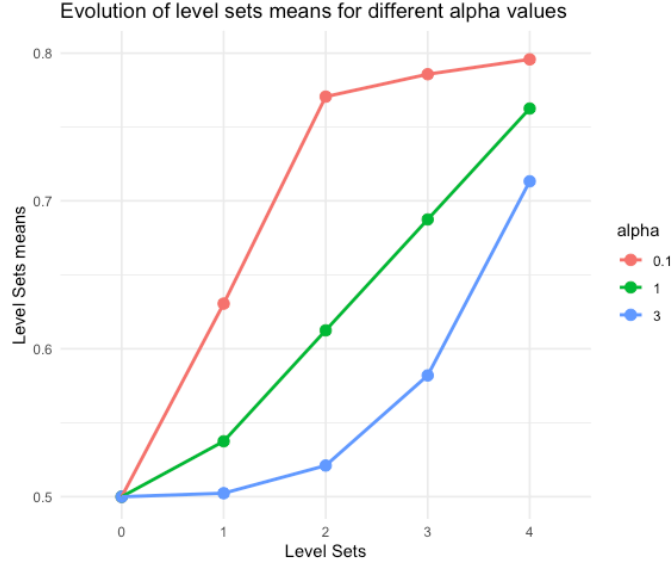


FIGURE 1

To enforce the constraint $0.5 < \mu_{(1)} < \mu_{(2)} < \dots < \mu_{(K-1)} < \beta_{\max}$, we consider a power-law function $\mu_{\alpha,k} = 0.5 + g(k)^\alpha \mathbb{I}(0.5, \beta_{\max})$. This power law is monotonically increasing in $g(k)$, and it is flexible enough to account for different increasing rates by modulating accordingly the α values. Indeed, we introduce the α hyperparameter, upon which, jointly with k , $\mu_{\alpha,k}$ depends as follows:

$$(5) \quad \mu_{\alpha,k} = f(\alpha, k) = \left(\frac{(\beta_{\max} - 0.5)^{(1/\alpha)}}{K} \times k \right)^\alpha + 0.5 \quad \text{for } k = 0, \dots, K$$

The equation above may look rather abstract, but the interpretation for α is quite straightforward. The parameter α regulates the rate of increase of the means $\mu_{\alpha,k}$ across the level sets, from $k = 1$ to $k = K$. We have an interpretation for α magnitude:

- For α values exceeding 1, a convex power-law function emerges, engendering a steady but accelerating increase in the level set truncations toward β_{\max} .
- When α is between 0 and 1, the power-law function becomes concave, promptly pushing values toward β_{\max} . This reflects a pronounced bias toward higher probabilities.
- Notably, as α is equal to 1, the power-law function becomes linear, leading to a constant increment in the $\mu_{\alpha,k}$ parameter.

2. RE-PARAMETRIZE WITH A BETA AND INTEGRATE OUT P

We first redefine the proportional likelihood over the blocks, meaning that instead of considering the victories/games between players, we count the victories/games occurring between blocks. Therefore, we introduce two new quantities:

$$(6) \quad \bar{y}_{pq} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N y_{ij} \cdot \mathbb{I}(z_i = p, z_j = q)$$

$$(7) \quad \bar{n}_{pq} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N n_{ij} \cdot \mathbb{I}(z_i = p, z_j = q)$$

where

$$\mathbb{I}(z_i = p, z_j = q) = \begin{cases} 1 & \text{if } z_i = p, z_j = q \\ 0 & \text{otherwise} \end{cases}$$

The likelihood expressed over the blocks will look like:

$$p(y \mid z, P) \propto \prod_{p=1}^{K-1} \prod_{q=p+1}^K p_{pq}^{\bar{y}_{pq}} \cdot (1 - p_{pq})^{\bar{n}_{pq} - \bar{y}_{pq}}$$

where \bar{n}_{pq} and \bar{y}_{pq} are respectively the number of games played between block p and block q and the number of victories of block p vs block q .

We can replace the truncated normals over the level sets with beta distribution so that $p_{pq} \sim \text{TrunBeta}(\alpha_k, \beta_k) \mathbb{I}_{p_{pq}}(0.5, \beta_{max})$. We will have the product of $K - 1$ truncated beta priors.

$$(8) \quad p(P \mid z) \propto \prod_{k=1}^{K-1} \left(p_{pq}^{\alpha_k - 1} \cdot (1 - p_{pq})^{\beta_k - 1} \right)^{K-k} \mathbb{I}(p_{pq} \in [0.5, \beta_{max}])$$

$$(9) \quad = \prod_{p=1}^{K-1} \prod_{q=p+1}^K \left(p_{pq}^{\alpha_k - 1} \cdot (1 - p_{pq})^{\beta_k - 1} \right) \mathbb{I}(p_{pq} \in [0.5, \beta_{max}])$$

To induce the same constraint as above, we could reparametrise the Beta by expressing α_k and β_k parameters in terms of the mean $\mu_{\alpha,k}$ and the confidence in the mean prior ν .

We have that:

$$(10) \quad \begin{aligned} \mathbb{E}[p_{pq} \mid z] &= \frac{\alpha_k}{\alpha_k + \beta_k} \cdot \underbrace{\frac{1}{F(\beta_{max}; \alpha_k, \beta_k) - F(0.5; \alpha_k, \beta_k)}}_{Z_0} \\ &\approx \frac{\alpha_k}{\alpha_k + \beta_k} \quad \text{for } \sigma^2 < 0.001 \end{aligned}$$

2.1. Riparametrization with ν as prior sample size. In terms of estimating p_{pq} , $\mu_{\alpha,k}$ represents our prior guess at the true value of p_{pq} and $\nu = \alpha_k + \beta_k$ represents our confidence in this guess, expressed on the same scale as the sample size. Therefore we can approximate their prior beliefs about p_{pq} with a truncated beta distribution having parameters

$$\begin{aligned} \alpha_k &= \mu_{\alpha,k} \cdot Z_0 \cdot \nu \\ \beta_k &= \nu \cdot (1 - \mu_{\alpha,k} \cdot Z_0) \end{aligned} \quad (11)$$

where Z_0 is the normalising constant of the prior truncated distribution that we express in (19), and α_k, β_k have been both obtained by solving the system of equations $\mathbb{E}(p_{pq}) = \mu_{\alpha,k}$ and $\alpha_k + \beta_k = \nu$ and $p_{pq} \sim \text{TruncBeta}(\alpha_k, \beta_k)$.

2.2. Riparametrisation with mean and variance. If we were working with a non-truncated beta we may reparametrise the Beta(α_k, β_k)

$$(12) \quad \alpha_k = \mu_{\alpha,k} \left(\frac{\mu_{\alpha,k}(1 - \mu_{\alpha,k})}{\sigma^2} - 1 \right), \text{ if } \sigma^2 < \mu_{\alpha,k}(1 - \mu_{\alpha,k})$$

$$(13) \quad \beta_k = (1 - \mu)\nu = (1 - \mu_{\alpha,k}) \left(\frac{\mu_{\alpha,k}(1 - \mu_{\alpha,k})}{\sigma^2} - 1 \right), \text{ if } \sigma^2 < \mu_{\alpha,k}(1 - \mu_{\alpha,k}).$$

However, we are working with a truncated beta, and the parameters correspondence is not that straightforward.

2.3. Integrating P out. The first step is to show that the beta-binomial conjugacy is preserved under truncation of the prior.

$$(14)$$

$$p(P \mid y, z) \propto p(y \mid z, P) \cdot p(P \mid z)$$

$$(15)$$

$$\propto \prod_{p=1}^{K-1} \prod_{q=p+1}^K p_{pq}^{\bar{y}_{pq}} \cdot (1 - p_{pq})^{\bar{n}_{pq} - \bar{y}_{pq}} \cdot \prod_{p=1}^{K-1} \prod_{q=p+1}^K \left(p_{pq}^{\alpha_k - 1} \cdot (1 - p_{pq})^{\beta_k - 1} \right) \mathbb{I}(p_{pq} \in [0.5, \beta_{max}])$$

$$(16)$$

$$= \prod_{p=1}^{K-1} \prod_{q=p+1}^K p_{pq}^{\bar{y}_{pq} + \alpha_k - 1} \cdot (1 - p_{pq})^{\bar{n}_{pq} - \bar{y}_{pq} + \beta_k - 1} \mathbb{I}(p_{pq} \in [0.5, \beta_{max}])$$

which is the kernel of a truncated beta distributions $\text{TruncBeta}(\bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}) \mathbb{I}(p_{pq} \in [0.5, \beta_{max}])$, where $\bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}$ are:

$$(17) \quad \bar{\alpha}_{pq,k} = \bar{y}_{pq} + \alpha_k$$

$$(18) \quad \bar{\beta}_{pq,k} = \bar{n}_{pq} - \bar{y}_{pq} + \beta_k$$

and the normalising constant is

$$(19) \quad F(\beta_{max}; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}) - F(0.5; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k})$$

where $F(x; \alpha, \beta)$ is the cumulative distribution function of a Beta(α, β) distribution evaluated in x .

The expected value of such a distribution is:

$$\begin{aligned}
 \mathbb{E}(p_{pq} \mid y, z) &= \frac{\bar{\alpha}_{pq,k}}{\bar{\alpha}_{pq,k} + \bar{\beta}_{pq,k}} \cdot \frac{1}{F(\beta_{max}; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}) - F(0.5; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k})} \\
 &= \frac{\bar{\alpha}_{pq,k}}{\bar{\alpha}_{pq,k} + \bar{\beta}_{pq,k}} \cdot \frac{1}{Z} = \frac{\bar{y}_{pq} + \alpha_k}{\bar{n}_{pq} + \alpha_k + \beta_k} \frac{1}{Z} \\
 &= \left[\frac{\bar{n}_{pq}}{\bar{n}_{pq} + \alpha_k + \beta_k} \cdot \frac{\bar{y}_{pq}}{\bar{n}_{pq}} + \frac{\alpha_k + \beta_k}{\bar{n}_{pq} + \alpha_k + \beta_k} \cdot \frac{\alpha_k}{\alpha_k + \beta_k} \right] \frac{1}{Z} \\
 (20) \quad &= \frac{\bar{n}_{pq}}{\bar{n}_{pq} + \alpha_k + \beta_k} \cdot \underbrace{\frac{\bar{y}_{pq}}{\bar{n}_{pq}} \frac{1}{Z}}_{\text{rescaled sample mean}} + \frac{\alpha_k + \beta_k}{\bar{n}_{pq} + \alpha_k + \beta_k} \cdot \underbrace{\frac{\alpha_k}{\alpha_k + \beta_k} \frac{1}{Z}}_{\text{rescaled prior mean}}
 \end{aligned}$$

Now, we marginalise out the P parameter, by integrating it out from the joint distribution as follows:

$$(21) \quad p(y \mid z) = \int_0^1 p(y, P \mid z) dP = \int_0^1 p(y \mid z, P) \cdot p(P \mid z) dP$$

$$(22) \quad \propto \int_{0.5}^{\beta_{max}} \prod_{p=1}^{K-1} \prod_{q=p+1}^K p_{pq}^{\bar{y}_{pq} + \alpha_k - 1} \cdot (1 - p_{pq})^{\bar{n}_{pq} - \bar{y}_{pq} + \beta_k - 1} dp_{pq}$$

In the integral above, we recognise the product of $K \cdot (K - 1)$ truncated beta distributions $\text{TruncBeta}(\bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}) \mathbb{I}(p_{pq} \in [0.5, \beta_{max}])$, where $\bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}$ are as in (17) and (18).

Employing a well-known simplification strategy, we divide the expression within the integral by the truncated beta's normalising constant already mentioned in (19)

Now, we multiply and divide (22) by the normalising constant:

$$\begin{aligned}
 &\prod_{p=1}^{K-1} \prod_{q=p+1}^K F(\beta_{max}; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}) - F(0.5; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}) \cdot \\
 (23) \quad &\int_{0.5}^{\beta_{max}} \prod_{p=1}^{K-1} \prod_{q=p+1}^K \frac{1}{F(\beta_{max}; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}) - F(0.5; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k})} p_{pq}^{\bar{\alpha}_{pq,k}} \cdot (1 - p_{pq})^{\bar{\beta}_{pq,k}}
 \end{aligned}$$

The expression within the integral now is just a product of density, and therefore it equals one. We are left with:

$$(24) \quad p(y \mid z) \propto \prod_{p=1}^{K-1} \prod_{q=p+1}^K F(\beta_{max}; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}) - F(0.5; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k})$$

where $\bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}$ are as in (17) and (18).

Also, to estimate P is possible, since $p(p_{pq} | y, z) \sim \text{TruncBeta}(\bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}) \mathbb{I}(p_{pq} \in [0.5, \beta_{\max}])$ and therefore

$$\begin{aligned}
 \mathbb{E}(p_{pq}) &= \frac{\bar{\alpha}_{pq,k}}{\bar{\alpha}_{pq,k} + \bar{\beta}_{pq,k}} \cdot \frac{1}{F(\beta_{\max}; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k}) - F(0.5; \bar{\alpha}_{pq,k}, \bar{\beta}_{pq,k})} \\
 &= \frac{\bar{\alpha}_{pq,k}}{\bar{\alpha}_{pq,k} + \bar{\beta}_{pq,k}} \cdot \frac{1}{Z} = \frac{\bar{y}_{pq} + \alpha_k}{\bar{n}_{pq} + \alpha_k + \beta_k} \frac{1}{Z} \\
 &= \left[\frac{\bar{n}_{pq}}{\bar{n}_{pq} + \alpha_k + \beta_k} \cdot \frac{\bar{y}_{pq}}{\bar{n}_{pq}} + \frac{\alpha_k + \beta_k}{\bar{n}_{pq} + \alpha_k + \beta_k} \cdot \frac{\alpha_k}{\alpha_k + \beta_k} \right] \frac{1}{Z} \\
 (25) \quad &= \frac{\bar{n}_{pq}}{\bar{n}_{pq} + \alpha_k + \beta_k} \cdot \underbrace{\frac{\bar{y}_{pq}}{\bar{n}_{pq}} \frac{1}{Z}}_{\text{rescaled sample mean}} + \frac{\alpha_k + \beta_k}{\bar{n}_{pq} + \alpha_k + \beta_k} \cdot \underbrace{\frac{\alpha_k}{\alpha_k + \beta_k} \frac{1}{Z}}_{\text{rescaled prior mean}}
 \end{aligned}$$

3. ALLOWING FOR DIFFERENT VALUES ON THE DIAGONAL

Here we want to relax the assumption that the entries on the diagonal are fixed to 0.5, since this might be too restricting and could significantly lower the goodness of fit of the model. Instead, we model the level set 0, that is, the main diagonal, as a truncated normal centred around 0.5. The truncation spans the whole restricted probability space from $1 - \beta_{\max}$ to β_{\max} . The prior probability distribution on the P matrix entries p_{pq} is:

$$\begin{aligned}
 (26) \quad & p(p_{pq} | \mu_k, \sigma^2, \beta_{\max}) = \\
 (27) \quad & \begin{cases} \text{TruncNorm}(\mu_k, \sigma^2) \mathbb{I}(p_{pq} \in [0.5, \beta_{\max}]) & \text{for } k = q - p > 0 \\ \text{TruncNorm}(0.5, \sigma^2) \mathbb{I}(1 - \beta_{\max}, \beta_{\max}) & \text{for } k = q - p = 0 \\ 1 - p_{pq} \text{ with } p = 1 & \text{for } k = q - p < 0 \end{cases}
 \end{aligned}$$

[to be continued]