Bayesian hierarchical modelling of MMA judges decisions

HJA Meulenbelt

March 4, 2025

Ordinal regression

Ordinal variables are variables whose value exists on an arbitrary scale where only the relative ordering between different values is significant. Using *K+1* ordered cutpoints we can partition the latent continous space $X = \mathbb{R}$ into K intervals from which we can compute K ordinal probabilities. Because of their fixed values, the bounding cut points, $c_0 = -\infty$ and $c_K = \infty$, are sometimes ignored in practice such that the cut points are defined as only the *K-1* interior boundaries.

Induced Dirichlet prior

To avoid the non-identifiability of the interior cut points from propagating to the posterior distribution we have to fix the cut points around some anchor point, ϕ. This allows us to transform the K probabilities $p_1, \ldots, p_K$ to a new variable that encodes the normalization, $S = \sum_{k=1}^{K} p_k = 1$ and the *K-1* cut points *S*, $c_1, \ldots, c_{K-1}$.

More formally if we are using a latent logistic probability density function with

$$\sigma(x) = \frac{1}{1+e^{-x}}$$
$$\lambda(x) = \sigma^{-1}(x) = \log \frac{x}{1-x}$$

and

$$\rho(x) = \frac{d\sigma}{dx}(x) = \sigma(x) \cdot (1 - \sigma(x))$$

then we can write the map from cut points to probabilities as

$$
\begin{aligned}
p_k &= \sigma(\phi - c_{k-1}) - \sigma(\phi - c_k) \\
\sum_{k=1}^{K} p_k &= S
\end{aligned}
$$

with the inverse map given by

$$
\begin{aligned}
S &= 1 \\
c_1 &= \phi - \lambda(S - p_1) \\
&... \\
c_k &= \phi - \lambda(\sigma(\phi - c_{k-1}) - p_k)
\end{aligned}
$$

The Jacobian matrix for this transformation is given by a left column of ones

$$J_{k,1} = \frac{\partial p_k}{\partial S} = 1$$

a diagonal band for $k > 1$

$$J_{k,k} \frac{\partial p_k}{\partial c_{k+1}} = -\rho(\phi - c_k)$$

and an off-diagonal band

$$J_{k,k+1} \frac{\partial p_k}{\partial c_{k+1}} = \rho(\phi - c_k)$$

and zeros elsewhere.

The pushforward probability density function for the induced prior model is then given by

$$\phi(c, S | \alpha, \phi) = Dirichlet(p(c, \phi) | \alpha) \cdot \delta(S - 1) \cdot |J(c, \phi)|$$

where Dirichlet is the Dirichlet probability density function and $\delta(S-1)$ is a Dirac delta function encoding the summation constraint. It is straightforward to marginalize out the auxiliary variable S to give a probability density function that defines our desired cut point prior model

$$\phi(c | \alpha, \phi) = Dirichlet(p(c, \phi | \alpha) \cdot |J(c, \phi)|$$

This prior model is implemented in Stan as follows:

```
functions {
  real induced_dirichlet_lpdf(vector c, vector alpha, real phi) {
    int K = num_elements(c) + 1;
    vector[K - 1] sigma = inv_logit(phi - c);
    vector[K] p;
    matrix[K, K] J = rep_matrix(0, K, K);

    // induced ordinal probabilities
    p[1] = 1 - sigma[1];

    for (k in 2:(K - 1)) {
      p[k] = sigma[k - 1] - sigma[k];
    }

    p[K] = sigma[K - 1];

    // baseline column of Jacobian
    for (k in 1:K) {
      J[k, 1] = 1;
    }

    // diagonal entries of Jacobian
    for (k in 2:K) {
      real rho = sigma[k - 1] * (1 - sigma[k - 1]);
      J[k, k] = - rho;
      J[k - 1, k] = rho;
    }

    return dirichlet_lpdf(p | alpha) + log_determinant(J);
  }
}
```

Decision-making of MMA judges

Theory

Ties are allowed but are actively discouraged in the rules. The rules state a *10-10* is not to be used "as an excuse by a judge that cannot assess the differences in the round". Any discernible difference (perhaps just one extra effective strike) should remove the *10-10* option within a round.

By far, the most common score is a *10-9.* This score can reflect an extremely close round or a round of marginal domination.

Winning a round by a large margin will result in a *10-8.* Judges should always give a *10-8* when one fighter "has dominated the action of the round, had duration of the domination and also impacted their opponent with either effective strikes or effective grappling manoeuvres that have diminished the abilities of their opponent". A *10-8* should be considered when a fighter shows either dominance (e.g. when the opponent continually has to defend) or impact (e.g. by visibly hurting their opponent, causing a lack of control or ability).

Finally, a *10-7* score can be given when a fighter completely overwhelms their opponent, so the judge considers the fight should be stopped. A fighter must display immense dominance and impact to warrant this score.

Practice

Judge *Sal d'Amato* is one the best-known judges in MMA. He judged *2960* rounds of *1031* fights as follows:

Score	Frequency
10-9	2804
10-8	141
10-7	2

The frequency of the scores, c, are modelled using a multinomial distribution

$$\begin{aligned} c &\sim MultiNominial(\theta) \\ \theta &\sim Dirichlet(\alpha_k) \\ \alpha_k &\sim Normal(0,1) \end{aligned}$$

An important feature of the Dirichlet distribution is that the underlying categories are correlated with each other. The Dirichlet distribution is constrained to $[0, 1]$ and the sum of all its components is equal to 1. The higher the values of one α_k, the lower the possible values of the other α_k by necessity. Figure 1 shows that α_{10-9} is strongly correlated with α_{10-8} and α_{10-7}. High values for α_{10-9} appear with low values for α_{10-8} and α_{10-7}. Because the distributions of α_{10-8} and α_{10-7} are both naturally low, their scatterplot is all clustered at low values for both variables.

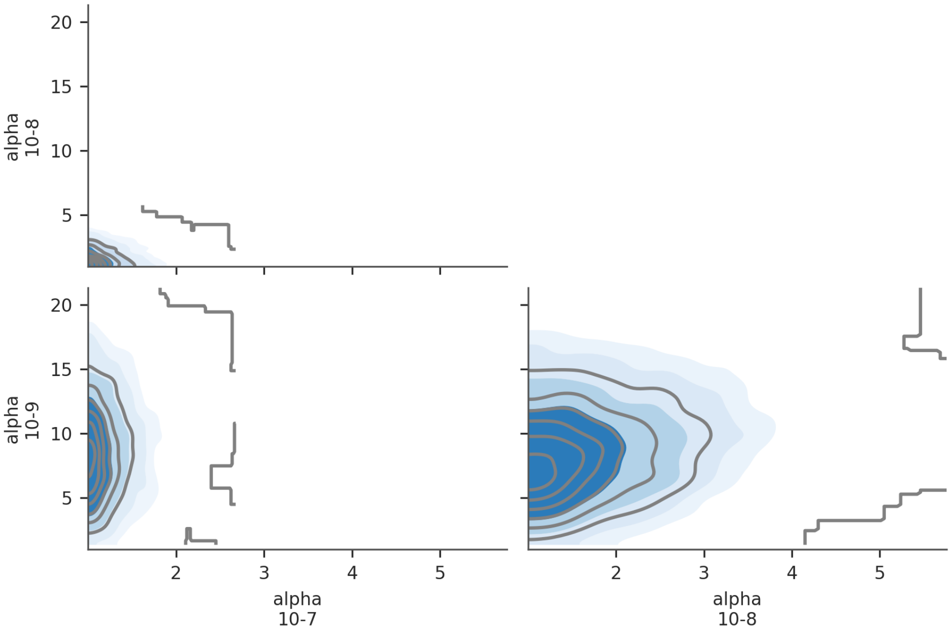

Figure 1: KDE pair plot

For *d'Amato* we retieve the following values for α_k

Score	α_k
10-9	8.87
10-8	1.86
10-7	1.22

As an alternative to a KDE plot, we can visualize this relationship with a ternary plot.

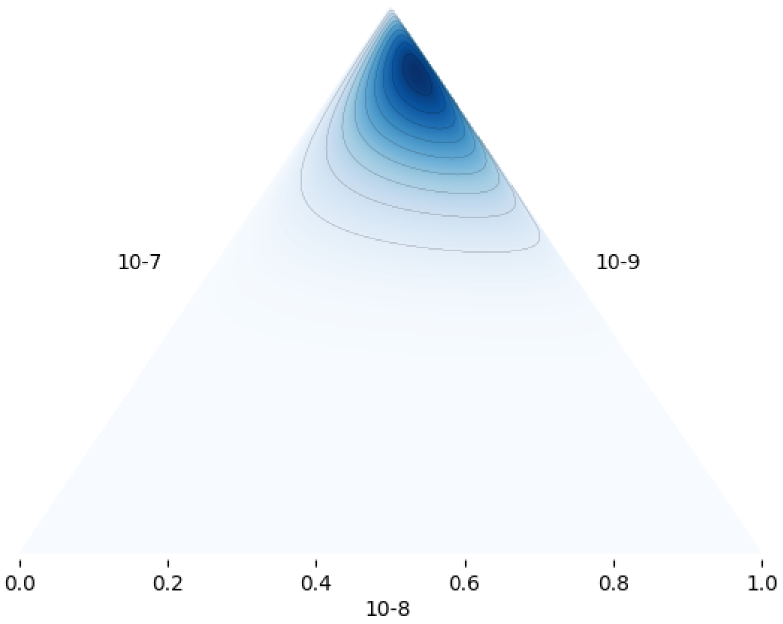

Figure 2: Ternary plot

Clearly, the probability of a score is not uniformly distributed. This domain knowledge needs to be translated into informative priors that are used as input parameters for the *InducedDirichlet*-distribution.

Methodology

Let $y_{r,j} \in \{7-10, 8-10, 9-10, 10-10, 10-9, 10-8, 10-7\}$ denote the score given by a judge j in round r. There are $n = 1, ..., N$ observations of judges' scores and, for brevity, we will refer to these as y_n.

The scores are modelled using an ordered-logit regression with mean λ_n and the cutoffs of each category denoted by c = *(*c_1*, ...,* c_6*)*.

In addition, there are J judges and a judge base his/her score on k = *1, ...,* K predictors. Each predictor k is multiplied with the difference between the two fighters, $X_{k,n}$ to obtain the latent effect.

Each judge has an individual set of parameters β_j = *(*$\beta_{j,1}$*, ...,* $\beta_{j,K}$*)* representing the value they attribute to each action. Correlation between the judge's preferences is introduced as this is extremely likely to exist.

Model

The following hierarchical model is constructed

$$
\begin{aligned}
y_n &\sim OrderedLogit(\lambda_n, c)\\
\lambda_n &= \beta_{j_n} X_n\\
\beta_j &= Normal(\mu, \Sigma)\\
\Sigma &= Diag(\sigma)\Omega Diag(\sigma)\\
\mu &\sim Normal(0, 5)\\
\sigma &\sim HalfNormal(0, 2.5)\\
\Omega &\sim LKJ(2)\\
c &\sim InducedDirichlet(1, 0)
\end{aligned}
$$

The $LKJ(\eta)$ distribution provides a prior over correlation matrices. When $\eta = 1$, the density is uniform over all matrices. When $0 \leq \eta \leq 1$, stronger (both positive and negative) correlations are favoured and when $\eta > 1$ weaker correlations are favoured. In the case of perfect correlation $\eta = 0$ and in the case of no correlation $\eta \to \infty$.

Non-centered parameterization

The multivariate normal density and LKJ prior on correlation matrices both require their matrix parameters to be factored. It is advised to parameterize the model directly in terms of Cholesky factors of correlation matrices using the multivariate version of the non-centered parameterization This parameterization removes the dependency between β, μ and σ.

We introduce z that has the same dimensions as β but the elements of z are independently and identically distributed standard normal

$$
\begin{aligned}
z &\sim Normal(0, 1)\\
\beta &= \mu + Lz
\end{aligned}
$$

where $LL^{\intercal} = \Sigma$ is the Cholesky factor of Σ.