

Appendix: Details of the Model

A.1 Level-1 Model

To recognize the specific offensive and defensive teams involved in a match, I model the shots on goal X taken by team j in a match against team j' as depending on offensive and defensive parameters ϕ_{oj} and $\phi_{aj'}$. The conditional probability of scoring y goals given x shots on goal also depends on offensive and defensive parameters, which I call θ_{oj} and $\theta_{aj'}$:

Level-1: Data Distributions

$$X | \phi_{oj}, \phi_{aj'} \sim \text{Poisson}(\mu), \quad \mu = \phi_{oj} + \phi_{aj'}, \quad \phi_{oj} > 0, \phi_{aj'} > 0.$$

$$Y | \theta_{oj}, \theta_{aj'}, X = x \sim \text{Binomial}(x, \pi), \quad \pi = \Phi(\theta_{oj} + \theta_{aj'}), \quad 0 < \pi < 1.$$

$$Y | \phi_{oj}, \phi_{aj'}, \theta_{oj}, \theta_{aj'} \sim \text{Poisson}(\mu \pi)$$

independent of

$$X - Y | \phi_{oj}, \phi_{aj'}, \theta_{oj}, \theta_{aj'} \sim \text{Poisson}(\mu(1 - \pi)).$$

The function $\Phi(z)$ is the standard Normal CDF ($\Phi(z) = P(Z \leq z)$, where $Z \sim N(0, 1)$).

The final two distributional results are due to the independence assumptions implied by the Poisson and Binomial distributions.

The expected number of shots on goal for team j increases with its offensive shot on goal parameter ϕ_{oj} , as well as with $\phi_{aj'}$, the defensive shot on goal parameter for the defending team j' . So the teams with the most shots on goal have a large ϕ_{oj} and a small $\phi_{aj'}$. The model forces every team to contribute something positive both to their own and to their opponent's mean shots on goal. This ensures that every Poisson distribution will have a positive mean, as is required.

The probability of team j converting a shot on goal against team j' increases with both

θ_{oj} and θ_{dj} , the respective offensive and defensive scoring parameters. So the best teams have a large θ_{oj} and a small (possibly negative) θ_{dj} . The Normal CDF increases monotonically from 0 to 1, so a negative defensive parameter decreases your opponent's probability converting of scoring on a shot on goal.

A.2 Level-2 Model

We expect World Cup soccer teams to be of similar abilities and hence have parameter values that are more similar than different. To model this, we can assume the team parameters are drawn from a unimodal level-2 distribution. Data from every match contributes information about the parameters of this distribution, and this enhances the inference for any particular team or match. For distributional convenience, I assume independent Gamma and Normal densities for the ϕ 's and θ 's, respectively.

Level-2: Team-Parameter Distributions

$$\begin{aligned} \phi_{oj}, \phi_{dj} &\stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, \lambda), & j = 1, \dots, 16. \\ \theta_{oj}, \theta_{dj} &\stackrel{\text{i.i.d.}}{\sim} N(0, \tau^2) \end{aligned}$$

Assuming common level-2 parameters for the offensive and defensive ϕ 's and for the offensive and defensive θ 's, makes the model identifiable. Notice that adding a small constant to every ϕ_{oj} and subtracting it from every ϕ_{dj} would leave the overall goal probabilities unchanged. The same is true if we add and subtract a constant from every θ_{oj}, θ_{dj} pair. Linking the parameters together with these common level-2 distributions forces them all to remain relatively close to their respective level-2 mean values.

A.3 Partitioning Shots On Goal

To model the shot process, imagine a breakdown of shots on goal into shots that are due

entirely to offensive ability, and shots that are due entirely to (lack of) defensive ability:

$$X_o | \phi_{oj} \sim \text{Poisson}(\phi_{oj})$$

independent of

$$X_d | \phi_{dj'} \sim \text{Poisson}(\phi_{dj'})$$

$$X = X_o + X_d \sim \text{Poisson}(\phi_{oj} + \phi_{dj'})$$

This isn't a realistic model for a soccer match because most shots on goal are produced by a combination of offensive execution and defensive breakdowns. I don't believe such a partitioning reflects anything real in terms of numbers of purely offensive and purely defensive shots to expect in a soccer match. However, modeling the problem in this way leads to the same model I assume for the overall shots on goal. So, realistic or not, we can use this hypothetical breakdown to help analyze the offensive and defensive strengths of each team. The partitioned shot totals (X_o 's and X_d 's) are examples of "latent variables": hypothetical variables introduced to break a hard problem into easier sub-problems.

Given ϕ_{oj} and $\phi_{dj'}$, the conditional distribution of the "offensive" goals given the total goals is

$$X_o | X = x \sim \text{Binomial}\left(x, \frac{\phi_{oj}}{\phi_{oj} + \phi_{dj'}}\right), \quad \text{with } X_d = X - X_o.$$

This is the reverse of the result about Poisson and Binomial random variables presented in A.1. If we knew the ϕ 's, we could create a random partition of each shot on goal total into "offensive shots" and "defensive shots" that would be consistent with the model assumptions.

Doing this for each team j in each of its n_j matches yields new estimates for the ϕ 's:

$$n_j \hat{\phi}_{oj} = \sum_{i=1}^{n_j} X_{oi} | \phi_{oj} \sim \text{Poisson}(n_j \phi_{oj}), \quad j = 1, \dots, 16$$

independent of

$$n_j \hat{\phi}_{dj} = \sum_{i=1}^{n_j} X_{di} = \sum_{i=1}^{n_j} (X_i - X_{oi}) | \phi_{dj} \sim \text{Poisson}(n_j \phi_{dj}).$$

Without the partitioned totals, it would be much more difficult to separate one team's ϕ from those of its opponents in the estimation procedure.

Combining the $\hat{\phi}$ distributions with the level-2 distribution for the ϕ 's, we can estimate the level-2 parameters α and λ and make new draws from $\phi_{oj} | \hat{\phi}_{oj}, \alpha, \lambda$ and $\phi_{dj} | \hat{\phi}_{dj}, \alpha, \lambda$. Because the Gamma distribution is conjugate to the Poisson, these posterior distributions are also Gamma:

$$\phi_{oj} | \hat{\phi}_{oj}, \alpha, \lambda \sim \text{Gamma}(\alpha + n_j \hat{\phi}_{oj}, \lambda + n_j), \quad j = 1, \dots, 16$$

independent of

$$\phi_{dj} | \hat{\phi}_{dj}, \alpha, \lambda \sim \text{Gamma}(\alpha + n_j \hat{\phi}_{dj}, \lambda + n_j).$$

The posterior mean of each ϕ can be expressed as a weighted average of the mean α/λ for the level-2 Gamma distribution, and the estimate $\hat{\phi}$, with a larger number of matches n_j leading to more weight on the estimate. For example, for ϕ_{oj} we have

$$E(\phi_{oj} | \hat{\phi}_{oj}, \alpha, \lambda) = \left(\frac{\lambda}{\lambda + n_j} \right) \frac{\alpha}{\lambda} + \left(\frac{n_j}{\lambda + n_j} \right) \hat{\phi}_{oj}.$$

The new ϕ values lead to new partitions of the shot totals, in an iterative procedure called a Markov Chain Monte Carlo (MCMC). After reaching equilibrium, the process yields draws from the posterior distributions of the ϕ 's, conditional only on the shots-on-goal totals, and not on any particular partition.

A.4 Modeling Goal Scoring

I introduce additional latent variables to model the goal-scoring process. For each shot on goal, imagine independent variables

$$Z_{ojk} \sim N(\theta_{oj}, 1/2)$$

$$Z_{aj'k} \sim N(\theta_{aj'}, 1/2)$$

where k indexes the shots on goal by team j against team j' . The k th shot on goal results in a goal if $Z_{ojk} + Z_{aj'k} > 0$, which occurs with probability $\Phi(\theta_{oj} + \theta_{aj'})$. This is true because the sum of these two independent Normal variables has mean $\theta_{oj} + \theta_{aj'}$ and variance $1/2 + 1/2 = 1$. If this method were applied to a basketball or football game, Z_{ojk} and $Z_{aj'k}$ could represent the numbers of points scored by the two teams. Some college football rankings make use of methods like this, because only information about wins and losses may be considered. In the case of goal scoring, Z_{ojk} and $Z_{aj'k}$ are purely hypothetical.

Using a procedure similar to that of 5.1, we can simulate values Z_{ojk} and $Z_{aj'k}$ for each shot on goal, conditional on the θ 's and on the outcome of the shot. If a shot resulted in a goal, then we know

$$S_k = Z_{ojk} + Z_{aj'k} > 0.$$

So we can simulate the sum $S_k = Z_{ojk} + Z_{aj'k}$ from a $N(\theta_{oj} + \theta_{aj'}, 1)$ distribution, constrained to be greater than 0. If the shot were saved, then we would constrain the distribution to be less than 0. There are a variety of ways to generate truncated Normals approximately (e.g., by using an approximation to the inverse-Normal CDF). Damien and Walker (2001) provide an efficient exact method. Given a value for S_k , we can generate the components Z_{ojk} and $Z_{aj'k}$ from their conditional distributions given their sum:

$$Z_{ojk} | S_k = s, \theta_{oj}, \theta_{aj'} \sim N(s/2 + (\theta_{oj} - \theta_{aj'})/2, 1/4)$$

$$Z_{aj'k} | S_k = s, Z_{ojk} = s - Z_{ojk}$$

I repeat this for every shot on goal and find new estimates for each θ by averaging the results for the m_{oj} shots on goal taken and the m_{aj} shots on goal faced by each team j .

$$\hat{\theta}_{oj} = \frac{1}{m_{oj}} \sum_{k=1}^{m_{oj}} Z_{ojk} | \theta_{oj} \sim N(\theta_{oj}, (2m_{oj})^{-1}), \quad j = 1, \dots, 16.$$

$$\hat{\theta}_{aj} = \frac{1}{m_{aj}} \sum_{k=1}^{m_{aj}} Z_{ajk} | \theta_{aj} \sim N(\theta_{aj}, (2m_{aj})^{-1}), \quad j = 1, \dots, 16.$$

Each Z has variance $1/2$, so the average of m has variance $(2m)^{-1}$. These estimates combine with the level-2 distributions for the θ 's to give a standard 2-level Normal hierarchical model. It is straight forward to estimate τ^2 and simulate new values from the posterior distributions $\theta_{oj} | \hat{\theta}_{oj}, \tau$ and $\theta_{aj} | \hat{\theta}_{aj}, \tau$, for $j = 1, \dots, 16$. Given the variance τ^2 of the team parameters, the posterior distributions for the θ 's are Normal with means that are a weighted average of the level-2 mean (0) and the estimate $\hat{\theta}$. The estimate gets more weight with larger numbers of shots on goal.

$$\theta_{oj} | \hat{\theta}_{oj}, \tau^2 \sim N(B_{oj}(0) + (1 - B_{oj})\hat{\theta}_{oj}, (1 - B_{oj})(2m_{oj})^{-1})$$

$$\theta_{aj} | \hat{\theta}_{aj}, \tau^2 \sim N(B_{aj}(0) + (1 - B_{aj})\hat{\theta}_{aj}, (1 - B_{aj})(2m_{aj})^{-1})$$

$$B_{oj} = \frac{(2m_{oj})^{-1}}{(2m_{oj})^{-1} + \tau^2}, \quad B_{aj} = \frac{(2m_{aj})^{-1}}{(2m_{aj})^{-1} + \tau^2}$$

New θ 's facilitate new simulated Z 's, leading to a second MCMC algorithm. After convergence, this algorithm yields draws from the posterior distribution of the θ 's, given only the observed goal and shot-on-goal totals.

I used this algorithm to collect 1600 nearly independent sets of posterior ϕ and θ draws. I modified the algorithm to treat the US as having one offensive team, but two defensive teams, depending on whether Solo or Scurry was in goal. So there are a total of $16 \times 2 = 32$ offensive parameter values and $17 \times 2 = 34$ defensive parameter values generated in each posterior draw.

Combining the parameters for generating shots on goal and for computing goal probabilities, I estimated the expected goals scored and allowed by each team against an average opponent (a hypothetical “reference” team with average ϕ and θ values). Figure 11 displays these expected goal totals for the 16 teams, with goals-for on the horizontal and goals-against on the vertical axis. Individual goal totals would be distributed roughly Poisson with these means.

Figure 11 here

Figure 11: The estimated mean goals scored and allowed by each of the 2007 World Cup teams playing against a hypothetical team with average parameter values. Teams marked below the solid diagonal line would typically win against an average opponent. Teams between the dashed diagonal lines win by between 1 and 2 goals on average against an average opponent. USA2 designates the U.S. team with Briana Scurry in goal, and USA the team with Hope Solo in goal. These two teams were modeled to have the same offensive parameters (hence the same x coordinate) but possibly different defensive parameters.

Conclusions

According to the model, an average team would score goals against Argentina according to a

Poisson process with a mean of about 4.3 (the vertical coordinate in Figure 11). This makes Germany's 11 goals against Argentina seem slightly less incredible, and in the Markov chain simulations, most of those 11 goals were regularly credited to Argentina's poor defense and not to Germany's offense. Nevertheless, Germany stood out as having the largest expected goal differential, outscoring a reference opponent by nearly 2.5 goals on average, topping Brazil (the next best team with an estimated mean of 1.9 goals) due to superior defense.

The model estimates that Germany would allow only 0.2 goals per match on average to the reference team, with probability less than 0.04 of allowing a score on any shot on goal. The next best defensive estimates against the reference team are Brazil, with 0.9 goals allowed on average, and Canada, with probability 0.19 of allowing a score on a particular shot on goal.

Germany's estimated mean goals allowed and defensive scoring probability are both positive despite the fact that Germany did not allow a goal in the 2007 World Cup. One advantage of Bayesian hierarchical models is that they provide a natural way to address extreme outcomes such as Germany's. The model allows for the possibility of a team scoring on Germany in each match, and the fitting procedure conditions on the fact the no team did in estimating the model parameters. The resulting defensive estimates for Germany are outstanding in a way that is consistent with the observed data, but are realistic when considering a future match.

The estimated mean goals allowed for the US team was about 0.4 goals per match larger with Briana Scurry than with Hope Solo. This is the vertical distance between the points marked USA and USA2 in Figure 11. The difference appears large in the context of the other fitted values, but the estimated posterior standard deviation for the difference in mean goals allowed is about 0.9. Generating more than 1600 posterior draws would allow us to

pin down the posterior mean difference more precisely, but the estimated variability of the posterior distribution means that we can not say anything precise about the true parameter values. Scurry allowed 3 goals in an actual match against Brazil, and the fitted values reflect the fact Solo outperformed Scurry in one realization of their matches in the 2007 World Cup. But the posterior distributions also reflect the uncertainty in these estimates due to the relatively small sample sizes, and the inherent variability of the Poisson process. Solo's estimated mean goals allowed was lower than Scurry's in 65% of the posterior draws. But that means there is posterior probability of about 0.35 (with simulation margin of error for 95% confidence of about $1/\sqrt{1600} = 0.025$) that Briana Scurry was the better keeper. If Scurry and Solo could each play a new match against Brazil, the data do not provide strong evidence that Solo would outperform Scurry.