Introduction to RUSH for Statisticians

The statistical model used in RUSH

RUSH is a statistical model for the *score margin process*, which is the difference $M(t) = V(t) - H(t) : 0 \le t \le T$ between the two teams' scores as a function of time. Here T is the length of the game, so that $M(T)$ is the *margin of victory*, a positive number if the visiting team wins and a negative number if they lose. A familiar model assumes that $M(T)$ is normally distributed with mean equal to the difference between the ability parameters of the two teams, possibly including a home field advantage term. One can also build models just based on which team wins using $\text{sign}(M(T))$. RUSH is novel because it utilizes $M(t)$ for $t < T$, and does so in order to deemphasize less informative time periods and scoring plays in the game.

The statistical model used by RUSH to create a probability distribution for the score margin process is derived from point process ideas. We'll first define a model for the score margin process that assumes that the two teams are evenly matched. Suppose that scoring plays occur at times according to a Poisson process with rate λ (ultimately we will model λ as a sum of two parameters called τ's, one for each team, that control how many scoring plays tend to occur in their games). Given that a scoring play occurs, it changes the score margin by some random value $s \in S = \{\pm 2, \pm 3, \pm 4, \pm 6, \pm 7, \pm 8\}$ according to probabilities π_s with $\sum \pi_s = 1$. Positive values correspond to scores for the visiting team and negative values represent scores for the home team, with one team arbitrarily assigned the home label if the game takes place on a neutral field. Since we are assuming the two teams are evenly matched, we have $\pi_s = \pi_{-s}$ for all s. (In RUSH, we lump a touchdown and the following one- or two-point conversion together as a single score, and as a result in college football we have the possibility of a four-point score, which occurs when one team scores a touchdown for six points, and the other team returns their conversion attempt for two points of their own.) The game, then, consists of a Poisson number of scoring plays, and the numbers of points scored on each scoring play are independent and identically distributed.

Now suppose that the two teams are no longer evenly matched, and the difference in their abilities is α, where $\alpha > 0$ means the visiting team is stronger. Suppose that at time t of the game, the score margin (difference between visitor and home scores) is m, let Δt be a small positive length of time, and consider the probability that a scoring play of size s occurs between times t and $t + \Delta t$. If $\alpha = 0$, the Poisson process assumption implies that this probability is approximately $\lambda \times \pi_s \times \Delta t$. For more general α, RUSH assumes that this probability is modified to be proportional to

$$\lambda \times \pi_s \times \Delta t \times \exp\{\alpha f(t, m, s)\}. \quad (1)$$

Here $f(t, m, s)$ is a function that is positive to the extent that we want the scoring play s to indicate that $\alpha > 0$. A reasonable choice is $f(t, m, s) = s$, which treats every point as equally significant, and this leads to a new model for the margin of victory. RUSH, however, uses an f that downweights less important scoring plays. Our f is the change in probability that the visiting team actually wins the game, when the score margin changes at time t from m to $m + s$, under the assumption that the two teams are evenly matched. In notation, where we have defined the score margin process $\{M(t) : 0 \le t \le T\}$ to be the score margin at all times t,

$$f(t, m, s) = Pr_0\{M(T) > 0 | M(t) = m + s\} - Pr_0\{M(T) > 0 | M(t) = m\}, \quad (2)$$

where Pr_0 is the probability measure when the true ability difference $\alpha = 0$ (i.e. the Poisson process model with which we began). f can be nearly one if a scoring play changes the margin from negative to positive late in the game, and it can equal nearly zero if the margin m is, say, 35 at any point in the game, or if m is 9 or more (i.e. if two or more scores are required for the trailing team to catch up) and t is sufficiently close to T.

We then apply usual point process reasoning, letting $\Delta t \to 0$ to obtain the following likelihood function for α (and λ and π) that depends on an observed score margin $\{m(t) : 0 \leq t \leq T\}$ which features n scoring plays s_i that occur at times t_i and result in score margins $m_i = \sum_{k=1}^{i} s_i$:

$$L(\alpha, \lambda, \pi | m(t)) = \lambda^n \left(\prod_{i=1}^{n} \pi_{s_i} \right) \exp \left(\alpha \sum_{i=1}^{n} f(t_i, m_{i-1}, s_i) - \lambda \int_0^T \sum_{\sigma \in S} \pi_\sigma \exp\{\alpha f(u, m(u), \sigma)\} du \right), \tag{3}$$

Note that $f(t, m, s)$ depends on λ and on π but we suppress this dependence in our notation. The integral in the likelihood function must be evaluated numerically, and we use Simpson's rule, and indeed $f(t, m, s)$ must be evaluated approximately. Finally note that this expression reduces to the Poisson process model when $\alpha = 0$.

To combine the results of many games, we assume that the ability difference α_g in the gth game is equal to $\theta_{V_g} - \theta_{H_g} - \eta I_g$, where each team j has an ability parameter θ_j, V_g is the identifier of the visiting team in game g, H_g is the home team, I_g is one if the home field advantage is present and zero if the game occurs at a neutral field, and η is a home field advantage parameter common to all teams (but which is estimated by RUSH). Ratings of individual teams are based on estimates of the θ's, although one should note that they are defined only up to a common additive constant, so we need to make some identifying assumption such as setting the average team to have $\theta = 0$ or the best team to have $\theta = 10$.

To obtain RUSH ratings, we combine the likelihood functions for all the games, assume (usually) that the prior distribution for the θ's is uniform over the whole real line, assume a hierarchical gamma prior for the τ's (recall that $\lambda_g = \tau_{V_g} + \tau_{H_g}$), assume a normal prior for η, and fix values for the π's, since computational shortcuts are possible if we do not continually have to reevaluate everything for new values of the π's. We then obtain a Monte Carlo estimate of the joint distribution of the θ's, τ's, and η using MCMC, and rank the teams by their expected ranks.

Handling overtime

The way overtime has been handled in RUSH has changed a few times: it was not initially obvious how it should be treated since time no longer plays a role and instead an overtime consists of one possession for each team. We now use the following rule. Suppose, for two evenly matched teams, that $1 - \epsilon$ is the probability that a single overtime ends with the game still tied, and that therefore each team wins with probability $\epsilon/2$. If the true ability difference is α, we assume the visiting team wins with probability proportional to $(\epsilon/2) \exp(\alpha/2)$, the home team wins with probability proportional to $\epsilon/2 \exp(-\alpha/2)$, and the game remains tied with probability still proportional to $1-\epsilon$. Note that this approach is similar to that used in defining standard RUSH: we are multiplying the evenly-matched probabilities by terms like $\exp(\alpha f)$, where f is the change in the probability that the visiting team wins. f in the case of overtime can equal $+1/2, -1/2$ and 0 respectively. Generalizing the idea of multiplying by $\exp(\alpha f)$ to time intervals longer than infinitesimal shows that the Bradley-Terry model also uses this approach, so we've shown that RUSH belongs to one class of models that includes a MOV model, and to another class of model that includes Bradley-Terry.