Having a Ball: evaluating scoring streaks and game excitement using in-match trend estimation

Claus Thorn Ekstrøm and Andreas Kryger Jensen
Biostatistics, Institute of Public Health, University of Copenhagen
ekstrom@sund.ku.dk, aeje@sund.ku.dk

16 November, 2020

Abstract

Nu ved jeg godt nok intet om sport, men... !

Keywords: APBRmetrics, Bayesian Statistics, Gaussian Processes, Sports Statistics, Trends

1 Introduction

Sports analytics receive increasing attention within statistics and not just for match prediction or betting but also for game evaluation, in-game and post-game coaching purposes, and for setting strategies and tactics in future matches.

Many popular sports such as football (soccer), basketball, boxing, table tennis, volleyball, American football, and handball involves matches between two teams or players where each team have the possibility of scoring throughout the match. Several research papers seek to predict match results (e.g., Karlis and Ntzoufras (2003);Groll et al. (2019);Gu and Saaty (2019)) or match winners for single matches (Cattelan, Varin, and Firth (2013)) in order to infer the match winner and potentially the winner of a tournament (Ekstrøm et al. (2020); Baboota and Kaur (2018)). While the overall result is highly interesting it conveys very little information about the individual development and trends throughout the match and modeling approaches that allow a finer granularity of the score difference throughout the match is needed.

The trend in score difference between the two teams is a proxy for their underlying strengths. In particular, sustained periods of time where the score difference increases suggests that one team outperforms the other whereas periods where the teams are constantly catching up to each other suggest that the team's strengths in those periods are similar. Modeling the local trend of the score difference will therefore reflect several aspects of the game in particular the team strengths and game dynamics and momentum as they develop through the match.

Figure XX shows the development of the score difference for the final match of the playoffs in the 2020 National Basketball Association (NBA) series between Los Angeles Lakers and Miami Heat. Positive numbers indicate that LA Lakers are leading and the running score difference shows that the Lakers pulled ahead until the third quarter where Miami Heat started to keep up the scoring pace before overtaking the Lakers and reducing the lead.

In this paper we will consider the score difference between two teams as a latent Gaussian process and use the Trend Direction Index (TDI) from Jensen and Ekstrøm (2020b) as a convenient measure to evaluate the local probability of the *monotonicity* of the latent process. The Trend Direction Index uses a Bayesian framework to provide a direct answer to questions such as "What is the probability that the latent process (i.e., that one team is doing better than another) is increasing at a given time-point?" This will allow real-time evaluation of the score difference trend at the current time-point in the game and will provide post-game inference about the "hot" periods of a match where one team out-performed the other. Furthermore we will present the Excitement Trend Index (ETI) as an objective measure of spectator excitement in a given match. The ETI is defined as the expected number of times that the score difference changes monotonicity during a match. If the score difference changes monotonicity othen then that echos a game where both teams frequently score that the whereas a game with a low ETI will represent a one-sided match where one team is doing consistently better than the other over sustained periods of time.

Other authors have considered using continuous processes to model the score difference of matches. Gabel and Redner (2012) shows that NBA basketball score difference is well described by a continuous-time anti-persistent random walk which suggests that a latent Gaussian process might be viable.

- Y. Chen, Dawson, and Müller (2020)
- T. Chen and Fan (2018)

The paper is structured as follows. In the next we introduce trend modeling through latent a Gaussian process and define the Trend Direction Index and Excitement Trend Index that captures the local trends in monotonicity and game excitement, respectively. In Section 3 we apply the our proposed methodology to analyse both the final match of the playoff as well as evaluating the game excitement distribution of the season by considering the ETI from all 1143 matches from the 2019–2020 National Basketball Association (NBA) season. We show how this distribution can be used to assess relative match excitement. We conclude with a discussion in Section 4. Materials to reproduce this manuscript and its analyses can be found at Jensen and Ekstrøm (2020a).

Unused stuff for now: *Philosophical question: What exactly is the sampling model for a single match? We could have a motivating figure here – e.g. a variation of Figure 1 (see its caption).*

2 Methods

Our model is based on the observed score differences D_m in a given match indexed by m. For each match we observe the random variables $\mathcal{D}_m = (t_{m_i}, D_{m_i})_{0<i\leq J_m}$ where $t_{m_1} < t_{m_2} < t_{m_i} < \ldots < t_{m_{J_m}}$ is the ordered time points at which any teams scores, $D_{m_i} = D_m(t_{m_i})$ is the associated difference in scores at time t_{m_i}, and J_m is the total number of goals during the match. We use the convention that D_m is the difference in scores of the home team with respect to the the away team, so that $D_m > 0$ means that the home team is leading. [**??? check that this is correct in the grønthøster**]

We assume that the observed match data are a noisy realization of a latent smooth, random function defined in continuous time evaluated at the random time points where goals occur. Let d_m be the latent functions from which the realizations $\mathcal{D}_m$ are generated. We then propose the following hierarchical model where d_m is a Gaussian process, and the observed data conditional on the scoring times and the values of the latent process at those times are independently normally distributed random variables with a match specific variance:

$$
\begin{aligned}
\boldsymbol{\Theta}_m \mid \boldsymbol{\Psi}_{\boldsymbol{m}}, \mathbf{t}_m &\sim H(\boldsymbol{\Psi}_m) \\
d_m(t) \mid \boldsymbol{\Theta}_m &\sim \mathcal{GP}(\mu_{\boldsymbol{\beta}_m}(t), C_{\boldsymbol{\theta}_m}(s,t)) \\
D_m(t_{m_i}) \mid d_m(t_{m_i}), t_{m_i}, \boldsymbol{\Theta}_m &\overset{iid}{\sim} N\left(d_m(t_{m_i}), \sigma_m^2\right)
\end{aligned}
\tag{1}
$$

where $\boldsymbol{\Theta}_m = (\boldsymbol{\beta}_m, \boldsymbol{\theta}_m, \sigma_m^2)$ is a vector of hyper-parameters governing the dynamics of the latent Gaussian process with a prior distribution H indexed by parameters $\boldsymbol{\Psi}$, and $\mathbf{t}_m = (t_{m_1}, \ldots, t_{m_{J_m}})$ is the vector of time points where goals occurs in the match.

By properties of Gaussian processes (see e.g., Cramer and Leadbetter (1967)) the latent process d_m and its time derivatives (assuming they exist) are distributed as a multivariate Gaussian process. We hence have that

$$
\begin{bmatrix} d_m(s) \\ d'_m(t) \\ d''_m(u) \end{bmatrix} \mid \boldsymbol{\Theta}_m \sim \mathcal{GP}\left(\begin{bmatrix} \mu_{\beta_m}(s) \\ \mu'_{\beta_m}(t) \\ \mu''_{\beta_m}(u) \end{bmatrix}, \begin{bmatrix} C_{\boldsymbol{\theta}_m}(s,s') & \partial_2 C_{\boldsymbol{\theta}_m}(s,t) & \partial_2^2 C_{\boldsymbol{\theta}_m}(s,u) \\ \partial_1 C_{\boldsymbol{\theta}_m}(t,s) & \partial_1\partial_2 C_{\boldsymbol{\theta}_m}(t,t') & \partial_1\partial_2^2 C_{\boldsymbol{\theta}_m}(t,u) \\ \partial_1^2 C_{\boldsymbol{\theta}_m}(u,s) & \partial_1^2\partial_2 C_{\boldsymbol{\theta}_m}(u,t) & \partial_1^2\partial_2^2 C_{\boldsymbol{\theta}_m}(u,u') \end{bmatrix} \right) \tag{2}
$$

where $'$ and $''$ denotes the first and second time derivatives, ∂_j^k is the k'th order partial derivative with respect to the j'th variable

[**??? TALK ABOUR POSTERIOR HERE ???**]

The Excitement Trend Index (ETI) of a particular match, denoted ETI_m, is defined as the expected number of changes in monotonicity of the score differences d_m conditional on the observed data from that match $\mathcal{D}_m$. This is equivalent to value of the expected number of zero-crossings of the posterior distribution of d'_m.

Formally,

$$
\mathrm{ETI}_m(\boldsymbol{\Theta}_m) = \mathrm{E}\left[\#\left\{t \in \mathcal{I}_m : df_m(t) = 0\right\} \mid \mathbf{D}_m, \mathbf{t}_m, \boldsymbol{\Theta}_m\right]
$$

where $\mathcal{I}_m$ is the interval of the time duration of a match i.e., $\mathcal{I}_m = [0; 48]$ minutes without overtime. The ETI is given by the intergal of the local Excitement Trend Index

$$
\mathrm{ETI}_m(\boldsymbol{\Theta}_m) = \int_{\mathcal{I}_m} d\mathrm{ETI}_m(t \mid \boldsymbol{\Theta}_m)\mathrm{d}t
$$

where dETI is the local Excitement Trend Index given by

$$
d\mathrm{ETI}_m(t \mid \boldsymbol{\Theta}_m) = \lambda(t \mid \Theta)\phi\left(\frac{\mu_{df}(t \mid \boldsymbol{\Theta})}{\Sigma_{df}(t,t \mid \boldsymbol{\Theta})^{1/2}}\right)\left(2\phi(\zeta(t \mid \boldsymbol{\Theta})) + \zeta(t \mid \boldsymbol{\Theta})\,\mathrm{Erf}\left(\frac{\zeta(t \mid \boldsymbol{\Theta})}{2^{1/2}}\right)\right)
$$

and $\phi\colon x \mapsto 2^{-1/2}\pi^{-1/2}\exp(-\frac{1}{2}x^2)$ is the standard normal density function, $\mathrm{Erf}\colon x \mapsto 2\pi^{-1/2}\int_0^x \exp(-u^2)\mathrm{d}u$ is the error function, and λ, ω and ζ are functions defined as

$$
\begin{aligned}
\lambda(t \mid \Theta) &= \frac{\Sigma_{d^2f}(t,t \mid \boldsymbol{\Theta})^{1/2}}{\Sigma_{df}(t,t \mid \boldsymbol{\Theta})^{1/2}}\left(1 - \omega(t \mid \boldsymbol{\Theta})^2\right)^{1/2} \\
\omega(t \mid \boldsymbol{\Theta}) &= \frac{\Sigma_{df,d^2f}(t,t \mid \boldsymbol{\Theta})}{\Sigma_{df}(t,t \mid \boldsymbol{\Theta})^{1/2}\Sigma_{d^2f}(t,t \mid \boldsymbol{\Theta})^{1/2}} \\
\zeta(t \mid \boldsymbol{\Theta}) &= \frac{\mu_{df}(t \mid \boldsymbol{\Theta})\Sigma_{d^2f}(t,t \mid \boldsymbol{\Theta})^{1/2}\omega(t)\Sigma_{df}(t,t \mid \boldsymbol{\Theta})^{-1/2} - \mu_{d^2f}(t \mid \boldsymbol{\Theta})}{\Sigma_{d^2f}(t,t \mid \boldsymbol{\Theta})^{1/2}\left(1 - \omega(t \mid \boldsymbol{\Theta})^2\right)^{1/2}}
\end{aligned}
$$

A derivation of this expression can be found in the supplementary material to Jensen and Ekstrøm (2020b).

The posterior distribution of the hyper-parameters given the observed data is then. We define $\widetilde{\mathbf{\Theta}}_m \sim P(\mathbf{\Theta}_m \mid \mathbf{D}_m, \mathbf{\Psi}_m, \mathbf{t}_m)$ hence

$$\widetilde{\mathbf{\Theta}}_m \sim \frac{G(\mathbf{\Theta}_m \mid \mathbf{\Psi}_m, \mathbf{t}_m) \int P(\mathbf{D}_m \mid f(\mathbf{t}_m), \mathbf{\Theta}_m, \mathbf{\Psi}_m, \mathbf{t}_m) dP(f_m(\mathbf{t}_m) \mid \mathbf{\Theta}_m, \mathbf{\Psi}_m, \mathbf{t}_m)}{\iint P(\mathbf{D}_m \mid f_m(\mathbf{t}_m), \mathbf{\Theta}_m, \mathbf{\Psi}_m, \mathbf{t}_m) dP(f_m(\mathbf{t}_m) \mid \mathbf{\Theta}_m, \mathbf{\Psi}_m, \mathbf{t}_m) dG(\mathbf{\Theta}_m \mid \mathbf{\Psi}_m, \mathbf{t}_m)}$$

What we estimate is then the random variable $\widehat{\mathrm{ETI}}_m = \mathrm{ETI}_m(\widetilde{\mathbf{\Theta}}_m)$ which can be summarized by its moments or quantiles.

We need to argue that ETI for $S_a(t_{m_i}) - S_b(t_{m_i})$ is symmetric in a and b so that our choice of "reference group" in D_m is not important. The reason is that we look at both up- and down-crossings at 0 of df_m so the choice of sign in D_m is not relevant.

2.1 Estimation

We have implemented the model described in the previous section in Stan (Carpenter et al. 2017).

Prior mean and covariance:

$$\mu_{\boldsymbol{\beta}_m}(t) = \beta_m, \quad C_{\boldsymbol{\theta}_m}(s,t) = \alpha_m^2 \exp\left(-\frac{(s-t)^2}{2\rho_m^2}\right)$$

with $\alpha_m, \rho_m > 0$. **[??? Note: Infinitely differetiable sample paths. Sample path derivatives are well-defined]**

Hyper-parameters: We used independent priors on $\mathbf{\Theta}_m = (\beta_m, \alpha_m, \rho_m, \sigma_m)$ of the form

$$G(\mathbf{\Theta}_m \mid \mathbf{\Psi}_m, \mathbf{t}_m) = G(\beta_m \mid \Psi_{\beta_m}) G(\alpha_m \mid \Psi_{\alpha_m}) G(\rho_m \mid \Psi_{\rho_m}) G(\sigma_m \mid \Psi_{\sigma_m})$$

where each prior is a heavy-tailed distribution with a moderate variance centered at the marginal maximum likelihood estimates. We used the following distributions

$$\beta_m \sim T\left(\widehat{\beta_m^{\mathrm{ML}}}, 3, 3\right), \quad \alpha_m \sim T^+\left(\widehat{\alpha_m^{\mathrm{ML}}}, 3, 3\right), \quad \rho_m \sim N^+\left(\widehat{\rho_m^{\mathrm{ML}}}, 1\right), \quad \sigma_m \sim T^+\left(\widehat{\sigma_m^{\mathrm{ML}}}, 3, 3\right)$$

where $T^+(\cdot, \cdot, \mathrm{df})$ and $N^+(\cdot, \cdot)$ denotes the location-scale half T- and normal distribution functions with df degrees of freedom. For each match we ran four independent Markov chains for 25,000 iterations each with half of the iterations used for warm-up. Convergence was assessed by trace plots of the MCMC draws and the potential scale reduction factor, $\widehat{R}$, of Gelman and Rubin (1992).

3 Results

We use data from Sports Reference LLC (2020).

General idea: We estimate ETI for all matches in a given season and make a nice plot of the distribution of ETI_m. Then we can rank the matches according to increasing ETI and show the running score difference for e.g., the lowest, median and highest ranked matches. Maybe a large forest plot of ETI_m would look impressive.

Given the posteriors ETI_m we can summarize them by a posterior mean and variance and then do a **meta analysis** where we adjust for game-specific fixed effects such as number of spectators, location, stratify by season and so on.

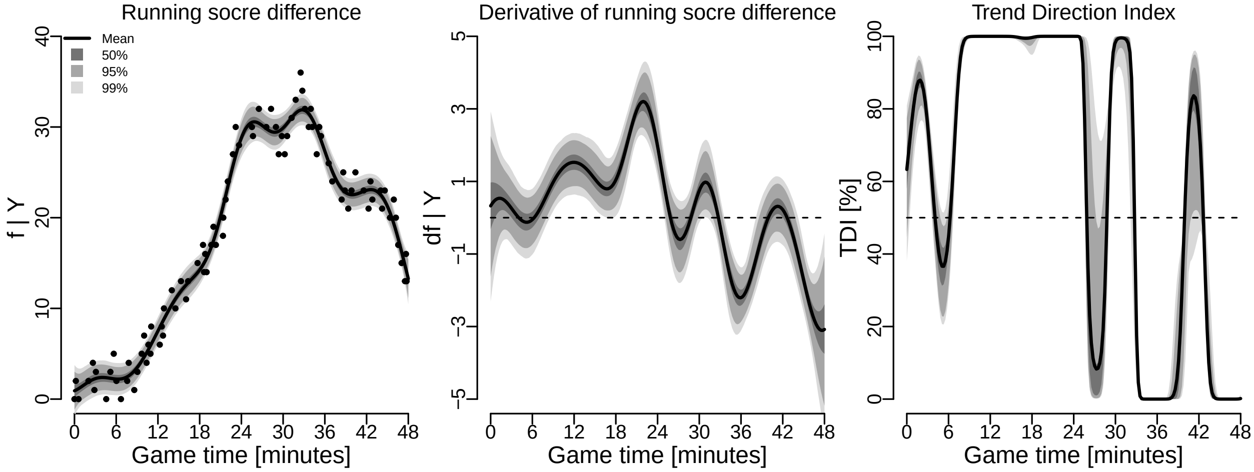

Figure 1: Los Angeles Lakers at Miami Heat. October 11, 2020. **??? Maybe this figure is superfluous in the results sections. We could instead use the first two panels as motivation in the introduction ???**

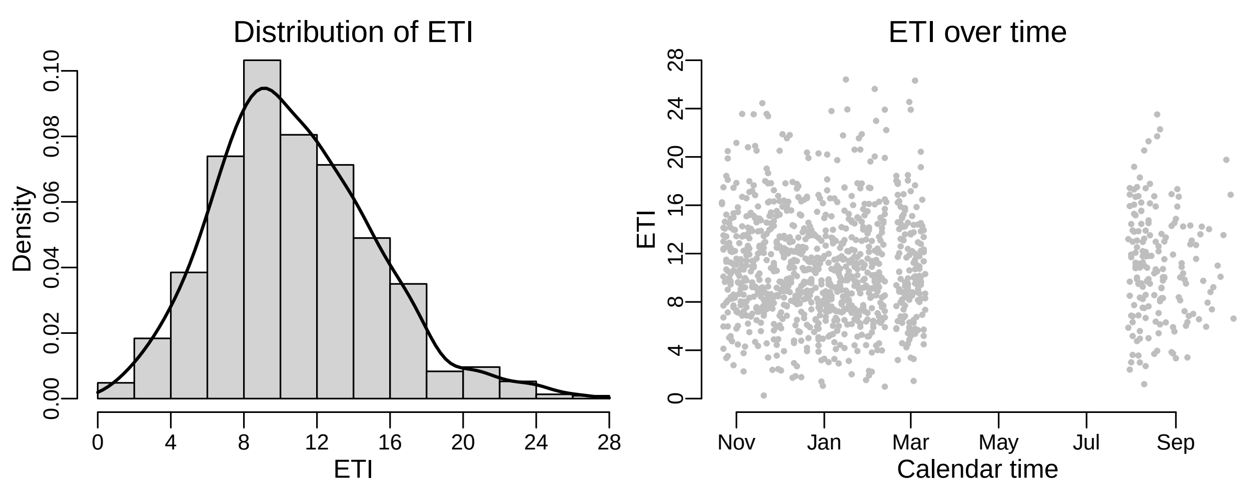

Figure 2: Caption

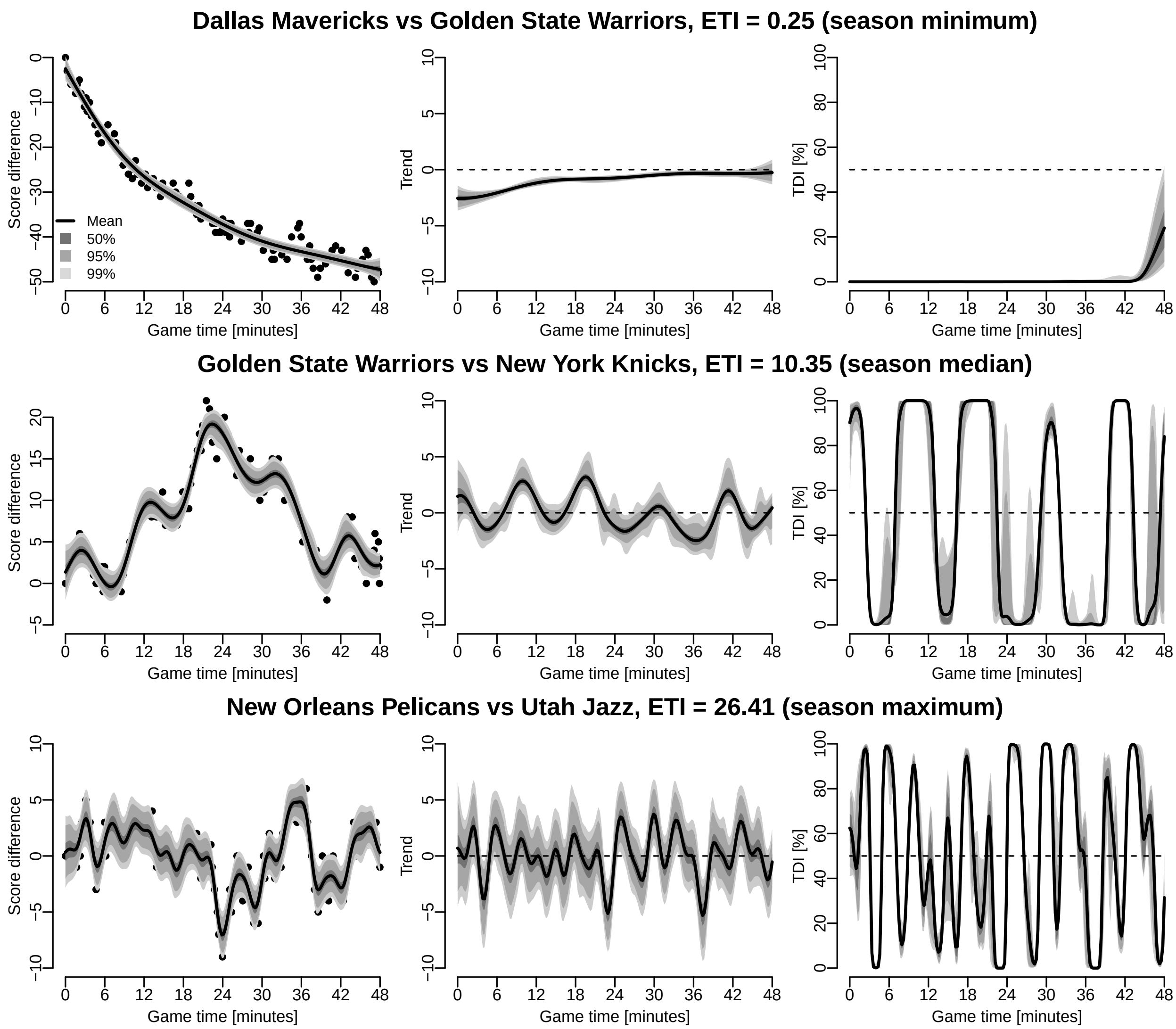

Figure 3: Caption **??? should we make this a nice full page figure and include 25% and 75% ETI games as well ???**

4 Discussion

Some summarization goes on here. We have introduces etc…

We could define a weighted Excitement Trend Index, $\mathrm{ETI}_m^{\mathrm{W}}$, so that changes in monotonicity of the score differences are i) weighted higher towards the end of the game and ii) weighted lower if one team is already far away of the other team. This motivates a weighted ETI of the form

$$\mathrm{ETI}_m^{\mathrm{W}} = \int_{\mathcal{I}_m} w(t, |f_m(t)|) d\mathrm{ETI}_m(t)$$

where w is a bivariate weight function being increasing in its first variable and decreasing in its second variable. Such weight function could be constructed as a product of two kernel functions on $[0; 48] \times \mathbb{R}_{\geq 0}$ with bandwidths based on studies of psychological perception.

A different approach would be to define team-specific excitement index nested with a match. Here we would only look at the **up**-crossings at zero of df_m and we would get two excitement indices for each match $(\mathrm{ETI}_{am}, \mathrm{ETI}_{bm})$. for teams a and b. This would somehow reflect how exciting each team were in match m with respect to chancing the sign of the score differences in their favor.

Acknowledgements

Bibliography

Baboota, Rahul, and Harleen Kaur. 2018. "Predictive Analysis and Modelling Football Results Using Machine Learning Approach for English Premier League." *International Journal of Forecasting* 35 (March). https://doi.org/10.1016/j.ijforecast.2018.01.003.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76 (1).

Cattelan, Manuela, Cristiano Varin, and David Firth. 2013. "Dynamic Bradley–Terry Modelling of Sports Tournaments." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 62 (1): 135–50. https://doi.org/https://doi.org/10.1111/j.1467-9876.2012.01046.x.

Chen, Tao, and Qingliang Fan. 2018. "A Functional Data Approach to Model Score Difference Process in Professional Basketball Games." *Journal of Applied Statistics* 45 (1): 112–27.

Chen, Yaqing, Matthew Dawson, and Hans-Georg Müller. 2020. "Rank Dynamics for Functional Data." *Computational Statistics & Data Analysis*, 106963.

Cramer, Harald, and M. R. Leadbetter. 1967. *Stationary and Related Stochastic Processes – Sample Function Properties and Their Applications.* John Wiley & Sons, Inc.

Ekstrøm, Claus Thorn, Hans Van Eetvelde, Christophe Ley, and Ulf Brefeld. 2020. "Evaluating One-Shot Tournament Predictions." *Journal of Sports Analytics* Preprint (Preprint): 1–10. https://doi.org/10.3233/JSA-200454.

Gabel, Alan, and Sidney Redner. 2012. "Random Walk Picture of Basketball Scoring." *Journal of Quantitative Analysis in Sports* 8 (1).

Gelman, Andrew, and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7 (4): 457–72.

Groll, Andreas, Christophe Ley, Gunther Schauberger, and Hans Van Eetvelde. 2019. "A hybrid random forest to predict soccer matches in international tournaments." *Journal of Quantitative Analysis in Sports* 15: 271–88.

Gu, Wei, and Thomas L. Saaty. 2019. "Predicting the Outcome of a Tennis Tournament: Based on Both Data and Judgments." *Journal of Systems Science and Systems Engineering* 28 (3): 317–43. https://doi.org/10.1007/s11518-018-5395-3.

Jensen, Andreas Kryger, and Claus Thorn Ekstrøm. 2020a. "GitHub Repository for Having a Ball." 2020. https://github.com/aejensen/Having-a-Ball.

———. 2020b. "Quantifying the Trendiness of Trends." *Journal of the Royal Statistical Society: Series C.*

Karlis, Dimitris, and Ioannis Ntzoufras. 2003. "Analysis of Sports Data by Using Bivariate Poisson Models." *Journal of the Royal Statistical Society: Series D (The Statistician)* 52 (3): 381–93. https://doi.org/10.1111/1467-9884.00366.

Sports Reference LLC. 2020. "Basketball Reference." 2020. https://www.basketball-reference.com/.

Appendix

the joint distribution of (f, df, d^2f) conditional on $\mathbf{Y}$, $\mathbf{t}$ and the hyper-parameters $\mathbf{\Theta}$ evaluated at any finite vector $\mathbf{t}^*$ of p time points is

$$\begin{bmatrix} f(\mathbf{t}^*) \\ df(\mathbf{t}^*) \\ d^2f(\mathbf{t}^*) \end{bmatrix} \mid \mathbf{Y}, \mathbf{t}, \mathbf{\Theta} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$$

where $\boldsymbol{\mu} \in \mathbb{R}^{3p}$ is the column vector of posterior expectations and $\mathbf{\Sigma} \in \mathbb{R}^{3p \times 3p}$ is the joint posterior covariance matrix. Partitioning these as

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_f(\mathbf{t}^* \mid \mathbf{\Theta}) \\ \mu_{df}(\mathbf{t}^* \mid \mathbf{\Theta}) \\ \mu_{d^2f}(\mathbf{t}^* \mid \mathbf{\Theta}) \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} \Sigma_f(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) & \Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) & \Sigma_{f,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) \\ \Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta})^T & \Sigma_{df}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) & \Sigma_{df,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) \\ \Sigma_{f,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta})^T & \Sigma_{df,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta})^T & \Sigma_{d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) \end{bmatrix}$$

the individual components are given by

$$\begin{aligned}
\mu_f(\mathbf{t}^* \mid \mathbf{\Theta}) &= \mu_\beta(\mathbf{t}^*) + C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}) \left(C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\
\mu_{df}(\mathbf{t}^* \mid \mathbf{\Theta}) &= d\mu_\beta(\mathbf{t}^*) + \partial_1 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}) \left(C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\
\mu_{d^2f}(\mathbf{t}^* \mid \mathbf{\Theta}) &= d^2\mu_\beta(\mathbf{t}^*) + \partial_1^2 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}) \left(C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\
\Sigma_f(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) &= C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}^*) - C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}) \left(C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right)^{-1} C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}^*) \\
\Sigma_{df}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) &= \partial_1\partial_2 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}) \left(C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right)^{-1} \partial_2 C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}^*) \\
\Sigma_{d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) &= \partial_1^2\partial_2^2 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}^*) - \partial_1^2 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}) \left(C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right)^{-1} \partial_2^2 C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}^*) \\
\Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) &= \partial_2 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}^*) - C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}) \left(C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right)^{-1} \partial_2 C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}^*) \\
\Sigma_{f,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) &= \partial_2^2 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}^*) - C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}) \left(C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right)^{-1} \partial_2^2 C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}^*) \\
\Sigma_{df,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \mathbf{\Theta}) &= \partial_1\partial_2^2 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}) \left(C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right)^{-1} \partial_2^2 C_{\boldsymbol{\theta}}(\mathbf{t}, \mathbf{t}^*)
\end{aligned}$$