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

12 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 increased attention within statistics and not just for match prediction or betting but also for game evaluation, coaching purposes, and for setting strategies and tactics in future matches.

Overall result of individual matches

We introduce the Excitement Trend Index (ETI) as an objective measure of spectator excitement in a given match.

- Excitement defineret som skift af hvem, der er i føring
- Vurdering af om et hold trender lige pt.
- Identificere hvilket "hot periods" et hold har i løbet af en kamp til efterfølgende evaluering

Reference to Quantifying the Trendiness of Trends Jensen and Ekstrøm (2020b).

We need to reference and discuss the following papers

- Chen, Dawson, and Müller (2020)
- Chen and Fan (2018)
- Gabel and Redner (2012)

Vi kan overveje, om vi ikke udelukkende burde fokusere på Eddy i dette manus og foreslå dette som et objektivt excitement index i stedet for at gå ind i en diskussion om TDI også. Jeg tænker også i forhold til at dette værk måske ikke behøver være så voluminøst.

Materials to reproduce this manuscript can be found at Jensen and Ekstrøm (2020a).

Philosophical question: What exactly is the sampling model for a single match?

The paper is structured as follows. In the next we introduce the Gaussian process that can be .. and derive the TDI and ETI.

In Section XX we apply the ... of every match from the 2019-2020 NBA basketball season

2 Methods

Let m index a given match between teams a and b and let $D_m(t_{mi}) = S_a(t_{mi}) - S_b(t_{mi})$ be the difference in scores at times $t_{m1} < t_{m2} < t_{mi} < \ldots < t_{mJ_m}$ being the ordered event times when a score by either team a or b occurs during match m.

We use the model

$$\begin{aligned}(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m, \sigma_m^2) \mid \boldsymbol{\Psi}_m, \mathbf{t}_m &\sim G(\boldsymbol{\Theta}_m \mid \boldsymbol{\Psi}_m, \mathbf{t}_m) \\ f_m(t) \mid \boldsymbol{\Theta}_m &\sim \mathcal{GP}(\mu_{\boldsymbol{\beta}_m}(t), C_{\boldsymbol{\theta}_m}(t, t')) \\ D_m(t_{mi}) \mid f_m(t_{mi}), t_{mi}, \boldsymbol{\Theta}_m &\overset{iid}{\sim} N\left(f_m(t_{mi}), \sigma_m^2\right)\end{aligned} \tag{1}$$

where f_m models a latent trajectory of score differences. We also have that

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

where $d^k\mu_{\boldsymbol{\beta}}$ is the k'th derivative of $\mu_{\boldsymbol{\beta}}$ and ∂_j^k denotes the k'th order partial derivative with respect to the j'th variable

We then define the Excitement Trend Index (ETI) as the number of zero-crossings of df_m conditional on the observed score differences. 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:

$$\mu_{\boldsymbol{\beta}_m}(t) = \beta_m, \quad C_{\boldsymbol{\theta}_m}(t, t') = \alpha_m^2 \exp\left(-\frac{(t-t')^2}{2\rho_m^2}\right)$$

with $\alpha_m, \rho_m > 0$.

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.

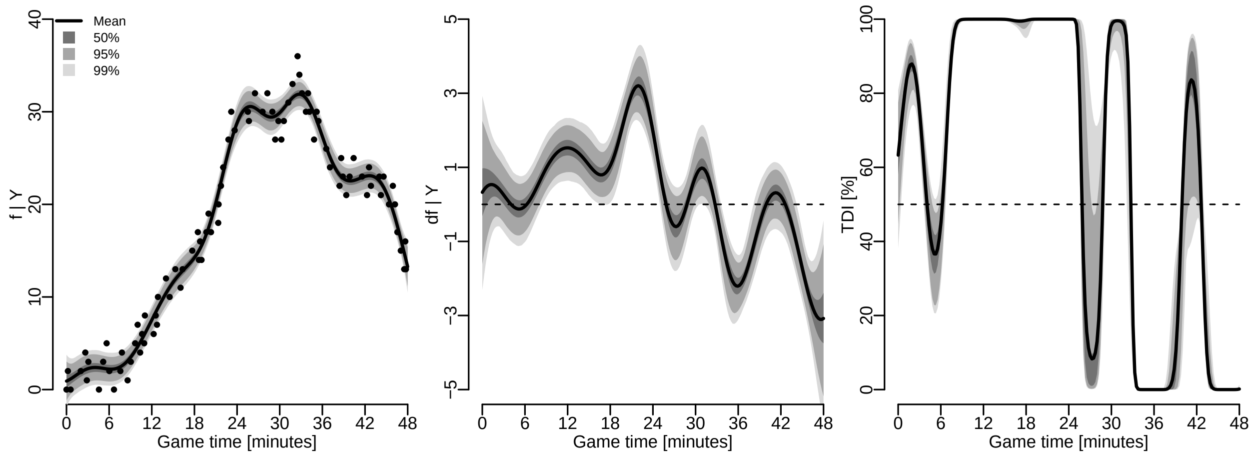

Figure 1: Los Angeles Lakers at Miami Heat. October 11, 2020

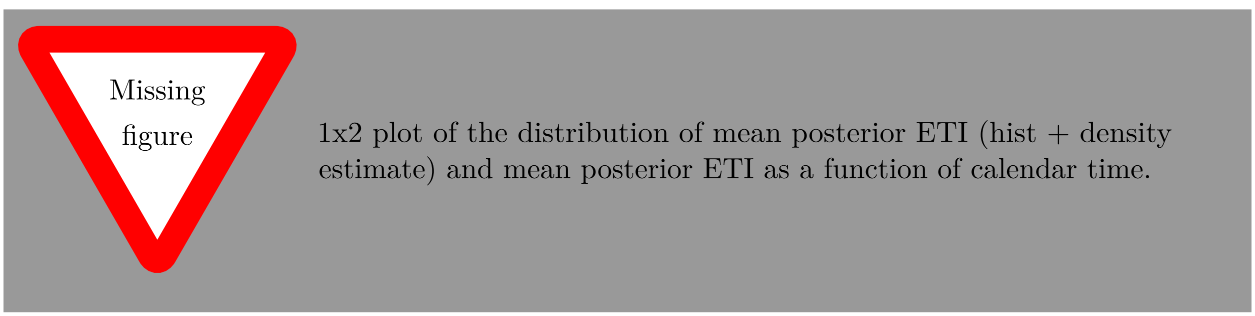

Figure 2: Caption

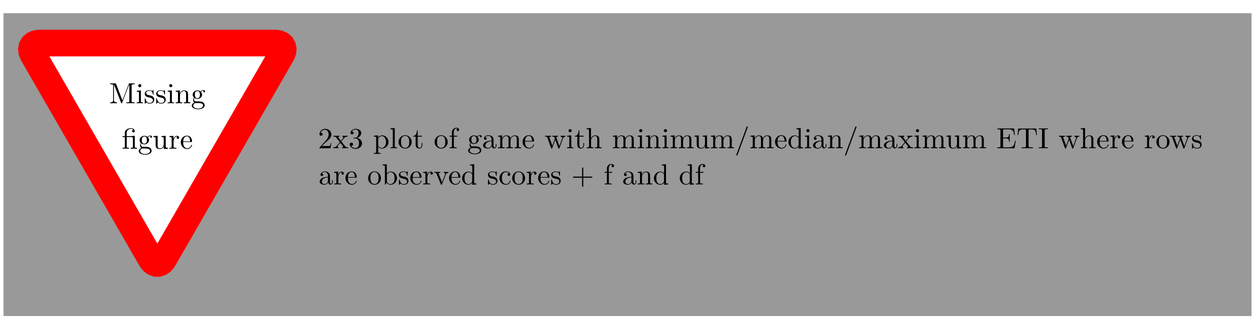

Figure 3: Caption

4 Discussion

We could define a weighted Excitement Trend Index, ETI_m^W, so that zero crossings of the derivative of score differences are weighted higher towards the end of the game as in

$$\text{ETI}_m^W = \int_{\mathcal{I}_m} d\text{ETI}_m(t) w(t) \text{d}t$$

where w is an increasing weight function.

Maybe changes in monotonicity of the score differences are not so important if one team is already far ahead of the other team. This would motivate a weighted ETI of the form

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

where $w(\cdot, \cdot)$ is a decreasing function in its second variable.

Another 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 $(\text{ETI}_{am}, \text{ETI}_{bm})$. 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

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).

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.

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.

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*.

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

Appendix (say what?)

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}$$