The Trendiness of Trends

Andreas Kryger Jensen
(joint with and energized by Prof. Ekstrøm on Business Class)

March 13th, 2019 (the best day of the year)

Biostatistics, Institute of Public Health
University of Copenhagen

The problem

Changes in trends

A statement often seen in the news is that at this very moment we see a significant change in the trend of something.

Recent examples from the Danish public news:

- Proportion of smokers
- Proportion of injuries from fireworks on New Year's Eve
- Proportion of children being baptized
- Average price of a one-family house

The questions

It is trendy to talk about changes in trends.

But...

- What is a trend?
- What is a change in a trend?
- How „trendy” is a trend?
- Can we quantify and estimate the trendiness of a trend?

Politiken, January 02, 2019

Første gang i 20 år: Flere danskere ryger

Prisen på cigaretter skal op. Sådan siger fagfolk på baggrund af årets undersøgelse af rygevaner, der viser, at andelen af rygere er steget markant på to år.

For første gang i to årtier er der en signifikant stigning i andelen af danskere, der ryger. Det viser den årlige undersøgelse af rygevaner blandt 5.017 danskere, som Sundhedsstyrelsen og tre organisationer offentliggør torsdag.

Five definitions

Definition 1

Definition 1
Reality evolves in continuous time $t \in \mathcal{T} \subset \mathbb{R}$.

Definition 2

There exists a random, latent function $f = \{f(t) : t \in \mathcal{T}\}$ governing the time evolution of some outcome. We can observe random variables sampled from f at discrete time points with additive measurement noise

$$Y_i = f(t_i) + \varepsilon_i, \quad t_i \in \mathcal{T}, \quad \mathbb{E}[\varepsilon_i \mid t_i] = 0$$

We are interested in understanding the dynamics of f conditional on observed data $(Y_i, t_i)_{i=1}^n$.

Definition 3
The trend of f is its instantaneous slope given by

$$df(t) = \left(\frac{\mathrm{d}f(s)}{\mathrm{d}s}\right)(t)$$

We say that f exhibits a positive trend at t if $df(t) > 0$ and a negative trend at t if $df(t) < 0$.

Definition 4
A change in the trend of f is a point where df changes sign (i.e., a positive trend becomes a negative trend or vice versa).

Definition 4
A change in the trend of f is a point where df changes sign (i.e., a positive trend becomes a negative trend or vice versa).

A change in trend never occurs at any given time point! It is a gradual and continuous event. The correct question is:

„What is the probability that the trend is changing at time t given what we have observed?"

Definition 5

We define a local Trend Direction Index as the conditional probability

$$\mathrm{TDI}(t, \delta) = P(df(t + \delta) > 0 \mid \mathcal{F}_t)$$

where $\mathcal{F}_t$ is the sigma algebra of all available information up until time t.

Definition 5

Definition 5
We define a local Trend Direction Index as the conditional probability

$$\mathrm{TDI}(t, \delta) = P(df(t+\delta) > 0 \mid \mathcal{F}_t)$$

where $\mathcal{F}_t$ is the sigma algebra of all available information up until time t.

- The sign of δ determines if we are estimating ($\delta \leq 0$) or forecasting ($\delta > 0$).
- A similar index can be defined for $df(t+\delta) < 0$ but that is just $1 - \mathrm{TDI}(t, \delta)$ hence redundant.
- Examples in the public news are often with $t = \text{now}$ and $\delta = 0$ (making change-point analyses impossible).

Functional Data Approach

Latent Gaussian Process model

Gaussian Process

A random function $\{f(t) : t \in \mathcal{T}\}$ is a Gaussian Process if and only if $(f(t_1), \ldots, f(t_n))$ has a multivariate normal distribution for every $(t_1, \ldots, t_n) \subset \mathcal{T}$ with $n < \infty$. $f \sim \mathcal{GP}(\mu(\cdot), C(\cdot, \cdot))$.

Latent Gaussian Process model

Gaussian Process
A random function $\{f(t) : t \in \mathcal{T}\}$ is a Gaussian Process if and only if $(f(t_1), \ldots, f(t_n))$ has a multivariate normal distribution for every $(t_1, \ldots, t_n) \subset \mathcal{T}$ with $n < \infty$. $f \sim \mathcal{GP}(\mu(\cdot), C(\cdot, \cdot))$.

We assume that the observed data is generated by the following hierarchical model:

$$f \mid m, \theta \sim \mathcal{GP}(m, C_\theta(\cdot, \cdot))$$

$$Y_i \mid f(t_i), \Theta, t_i \overset{iid}{\sim} N(f(t_i), \sigma^2), \quad \Theta = (m, \theta, \sigma^2)$$

where $C_\theta \colon \mathcal{T} \times \mathcal{T} \mapsto \mathbb{R}$ is a parametric covariance function.

Why a latent $\mathcal{GP}$?

The conditional joint distribution of (f, df) is a bivariate Gaussian Process:

$$\begin{bmatrix} f(s) \\ df(t) \end{bmatrix} \mid m, \theta \sim \mathcal{GP}\left(\begin{bmatrix} m \\ 0 \end{bmatrix}, \begin{bmatrix} C_\theta(s, s') & \partial_2 C_\theta(s, t) \\ \partial_1 C_\theta(t, s) & \partial_1 \partial_2 C_\theta(t, t') \end{bmatrix} \right)$$

Implication: We know the conditional distribution of (f, df) at any finite set of time point as well as their joint distribution with the observed data.

Conditional joint distribution

Let $\mathbf{Y}$ be the vector of random variables observed at time points $\mathbf{t}$ and $\mathbf{t}^*$ any finite subset of $\mathcal{T}$ not necessarily equal to $\mathbf{t}$.

The conditional joint distribution of everything is then

$$\begin{bmatrix} f(\mathbf{t}^*) \\ df(\mathbf{t}^*) \\ \mathbf{Y} \end{bmatrix} \mid \mathbf{t}, \Theta \sim N\left(\begin{bmatrix} m \\ 0 \\ m \end{bmatrix}, \begin{bmatrix} C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \\ C_\theta(\mathbf{t}, \mathbf{t}^*) & \partial_2 C_\theta(\mathbf{t}, \mathbf{t}^*) & C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \end{bmatrix} \right)$$

Conditional joint posterior distribution

Directly the joint conditional posterior distribution of the latent functions is

$$\begin{bmatrix} f(\mathbf{t}^*) \\ df(\mathbf{t}^*) \end{bmatrix} \mid \mathbf{Y}, \mathbf{t}, \Theta \sim N\left(\begin{bmatrix} \mu_f(\mathbf{t}^*) \\ \mu_{df}(\mathbf{t}^*) \end{bmatrix}, \begin{bmatrix} \Sigma_{f,f}(\mathbf{t}^*, \mathbf{t}^*) & \Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^*) \\ \Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^*)^T & \Sigma_{df,df}(\mathbf{t}^*, \mathbf{t}^*) \end{bmatrix} \right)$$

$$\begin{aligned}
\mu_f(\mathbf{t}^*) &= m + C_\theta(\mathbf{t}^*, \mathbf{t}) \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right]^{-1} (\mathbf{Y} - m) \\
\mu_{df}(\mathbf{t}^*) &= \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right]^{-1} \mathbf{Y} \\
\Sigma_{f,f}(\mathbf{t}^*, \mathbf{t}^*) &= C_\theta(\mathbf{t}^*, \mathbf{t}^*) - C_\theta(\mathbf{t}^*, \mathbf{t}) \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right]^{-1} C_\theta(\mathbf{t}^*, \mathbf{t})^T \\
\Sigma_{df,df}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_1 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right]^{-1} \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t})^T \\
\Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - C_\theta(\mathbf{t}^*, \mathbf{t}) \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I\right]^{-1} \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t})^T
\end{aligned}$$

Example

Let's see what this looks like on $\mathcal{T} = [0; 1]$ in the noise-free case ($\sigma^2 = 0$) with known parameters.

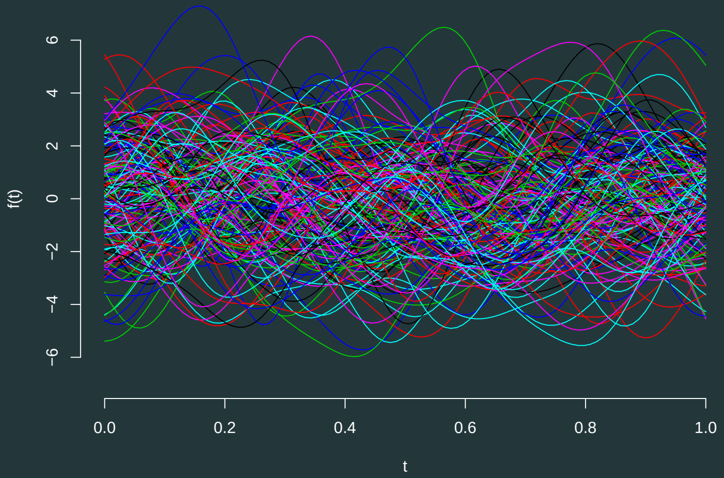
We use the exponential squared covariance function

$$C_\theta(s, t) = \alpha^2 \exp\left(-\frac{(s-t)^2}{2\rho^2}\right), \quad \theta = (\alpha, \rho)$$

where α is the standard deviation of the latent function and ρ is its *length-scale*. We set $\alpha = 2$ and $\rho = 0.1$.

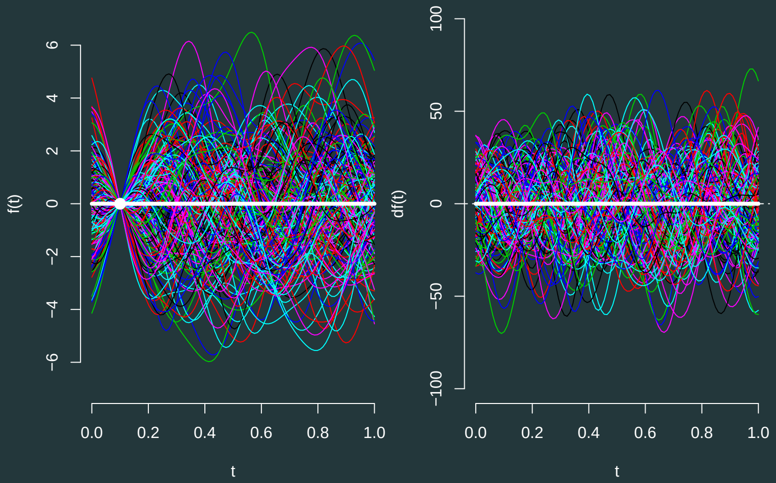
Example - prior distribution of f

What the world looks like without observed data



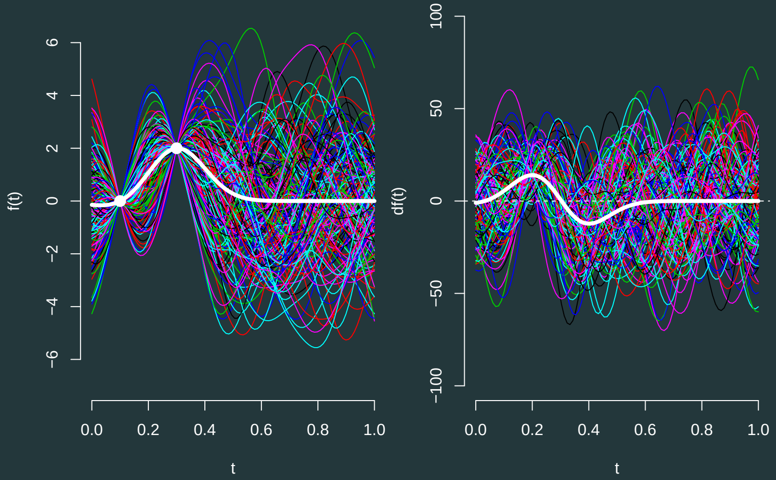
Example - posterior distribution of (f, df)

Posterior distributions conditional on 1 observation



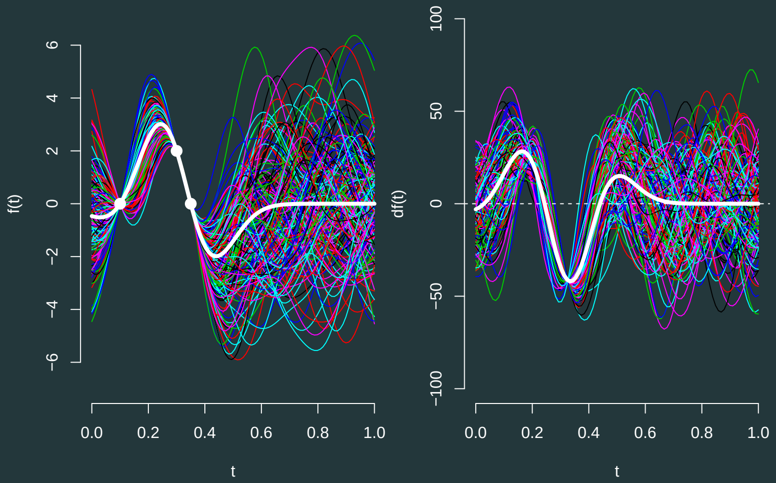
Example - posterior distribution of (f, df)

Posterior distributions conditional on 2 observations



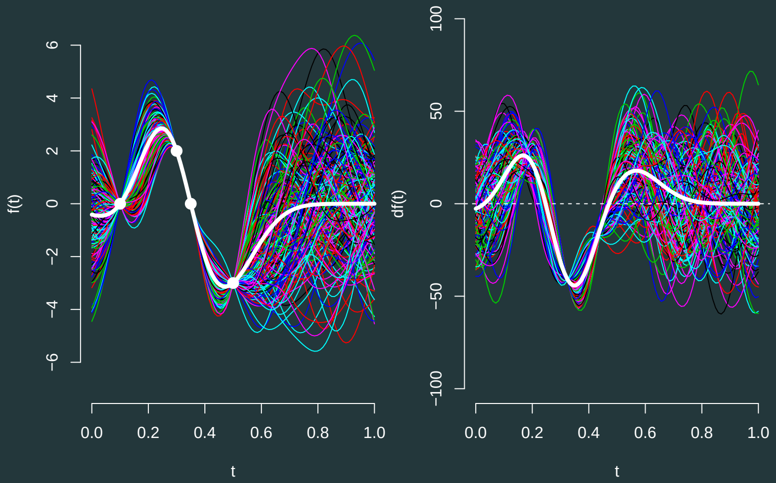
Example - posterior distribution of (f, df)

Posterior distributions conditional on 3 observations



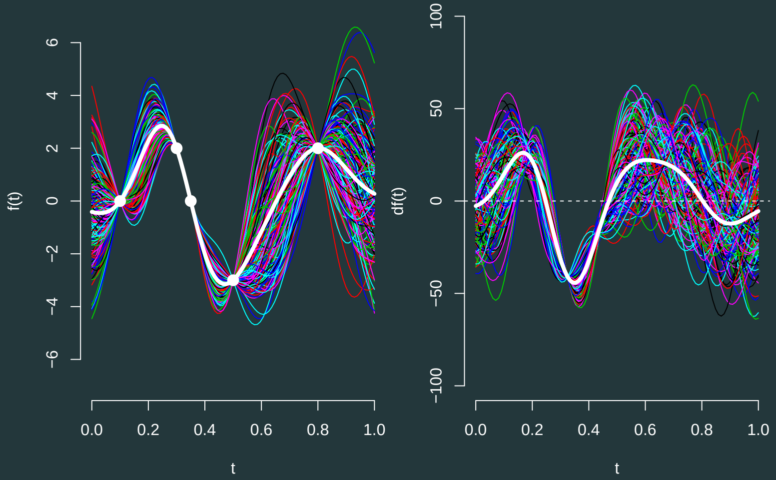
Example - posterior distribution of (f, df)

Posterior distributions conditional on 4 observations



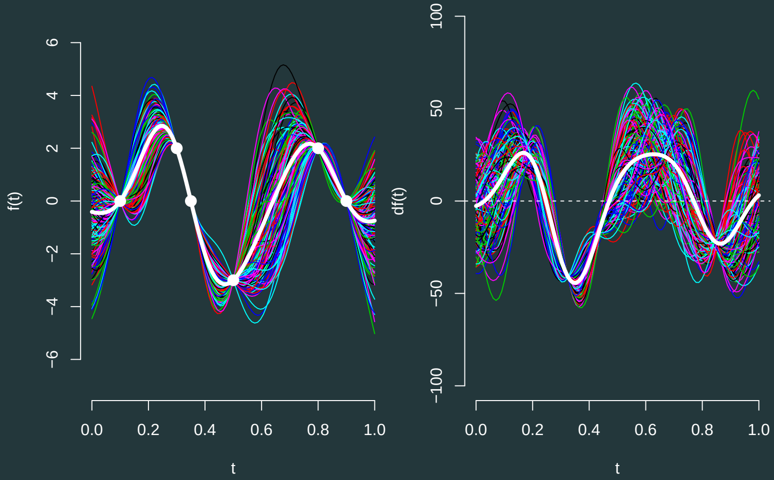
Example - posterior distribution of (f, df)

Posterior distributions conditional on 5 observations



Example - posterior distribution of (f, df)

Posterior distributions conditional on 6 observations



Trend Direction Index

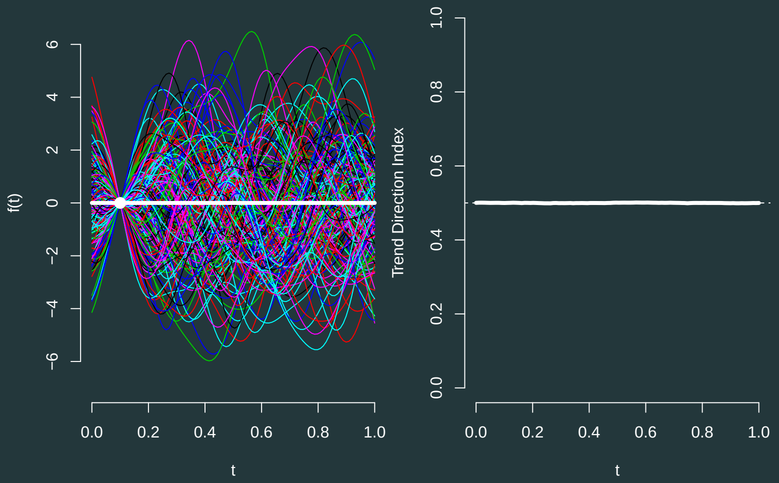
Recall the definition of the Trend Direction Index

$$\mathrm{TDI}(t, \delta) = P(df(t+\delta) > 0 \mid \mathcal{F}_t)$$

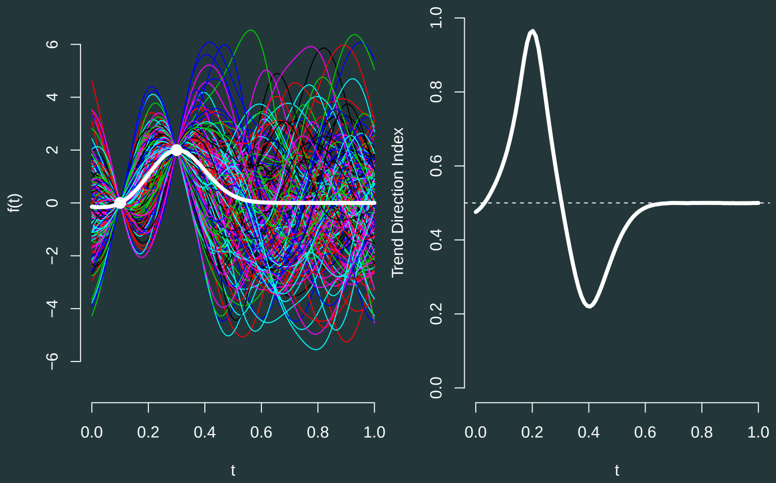
Letting $\mathcal{F}_t$ be generated by $\{\mathbf{Y}, \mathbf{t}\}$, we may express the Trend Direction Index as

$$\begin{aligned}\mathrm{TDI}(t, \delta \mid \Theta) &= P\left(df(t+\delta) > 0 \mid \mathbf{Y}, \mathbf{t}, \Theta\right) \\ &= \int_0^\infty N\left(u, \mu_{df}(t+\delta), \Sigma_{df,df}(t+\delta, t+\delta)^{1/2}\right) \mathrm{d}u\end{aligned}$$

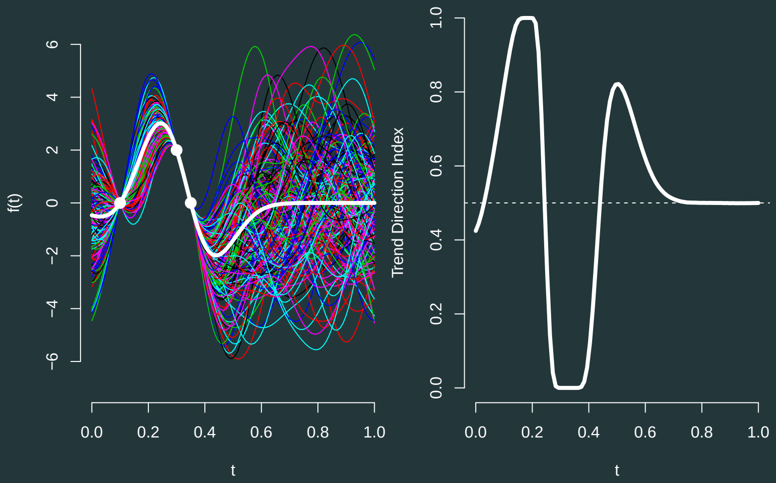
Example - posterior distribution of f and TDI



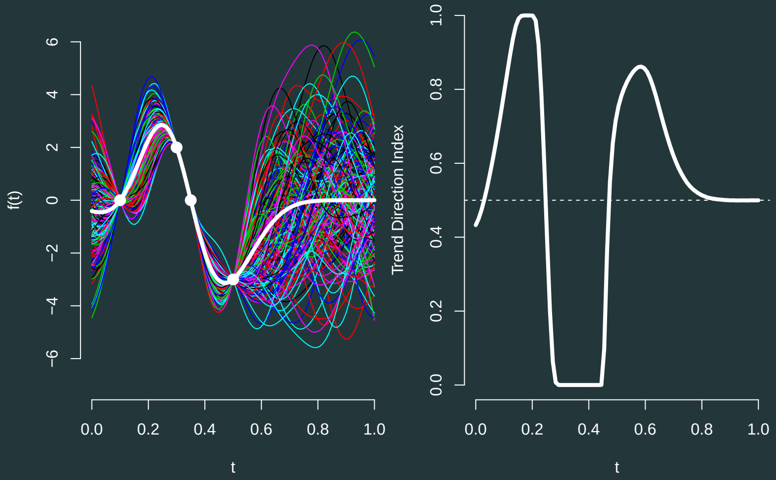
Example - posterior distribution of f and TDI



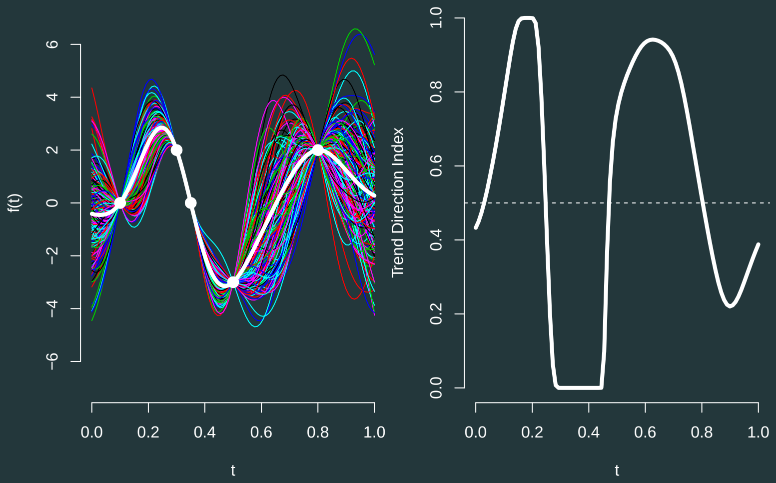
Example - posterior distribution of f and TDI



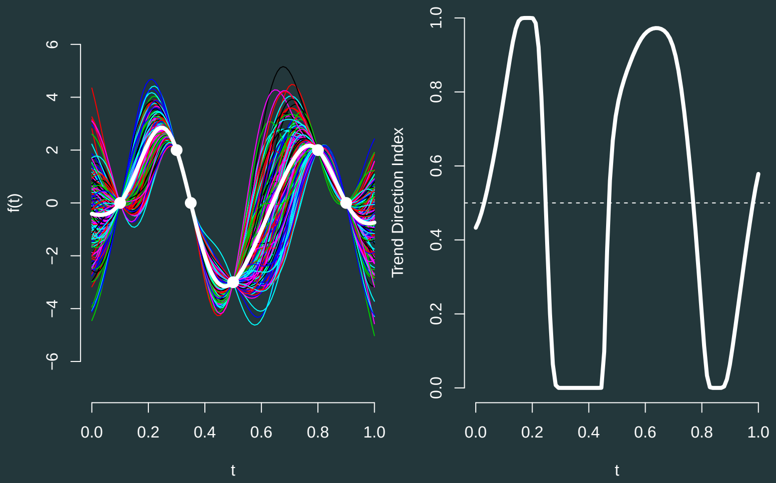
Example - posterior distribution of f and TDI



Example - posterior distribution of f and TDI



Example - posterior distribution of f and TDI



Estimation

Two approaches

- **Empirical Bayes estimation**
  - Maximize the marginal likelihood and plug-in
  - $\widehat{\Theta} = \arg\sup_{\Theta} \log L(\mathbf{Y} \mid \mathbf{t}, \Theta)$ and $\mathrm{TDI}(t, \delta \mid \widehat{\Theta})$
  - ☺ Fast and easy to implement
  - ☹ Can only be used in a very simple models where the marginal likelihood is analytically available
  - ☹ Does not account for the uncertainty in Θ

Two approaches

- **Empirical Bayes estimation**
  - Maximize the marginal likelihood and plug-in
  - $\widehat{\Theta} = \arg\sup_{\Theta} \log L(\mathbf{Y} \mid \mathbf{t}, \Theta)$ and $\mathrm{TDI}(t, \delta \mid \widehat{\Theta})$
  - ☺ Fast and easy to implement
  - ☹ Can only be used in a very simple models where the marginal likelihood is analytically available
  - ☹ Does not account for the uncertainty in Θ

- **Fully Bayesian estimation**
  - Markov chain Monte Carlo estimation with priors on Θ
  - ☺ Easily extendable to more complex models
  - ☺ Accounts for all the uncertainty in the model. $\mathrm{TDI}(t, \delta \mid \Theta)$ is now a distribution over probabilities
  - Requires special implementions or... **Stan**!

Fully Bayesian estimation in Stan

```
functions {
}

data {
}

parameters {
}

transformed parameters {
}

model {
}

generated quantities {
}
```

https://mc-stan.org/

The Stan program 1/2

```
functions{
  matrix pred_rng(real[] xPred, vector y, real[] x, real m,
                  real alpha, real rho, real sigma) {
    //...
  }
}

data {
  int<lower = 1> n;
  real x[n];
  vector[n] y;
  int<lower = 1> nPred;
  real xPred[nPred];
}

parameters {
  real mu;
  real<lower = 0> alpha;
  real<lower = 0> rho;
  real<lower = 0> sigma;
}
```

The Stan program 2/2

```
model {
  matrix[n, n] K = cov_exp_quad(x, alpha, rho);
  for (i in 1:n) {
    K[i, i] = K[i, i] + square(sigma);
  }

  mu ~ normal(26.835, 3);
  alpha ~ normal(0, 3);
  rho ~ inv_gamma(3.548762, 10.221723); // <- what is this?
  sigma ~ normal(0, 3);

  y ~ multi_normal_cholesky(rep_vector(mu, n), cholesky_decompose(K));
}

generated quantities {
  matrix[nPred, 4] pred;
  pred = pred_rng(xPred, y, x, mu, alpha, rho, sigma);
}
```

Application

From the Danish Health Authority[1]

Danskernes rygevaner 2018

Oprettet 3. januar 2019

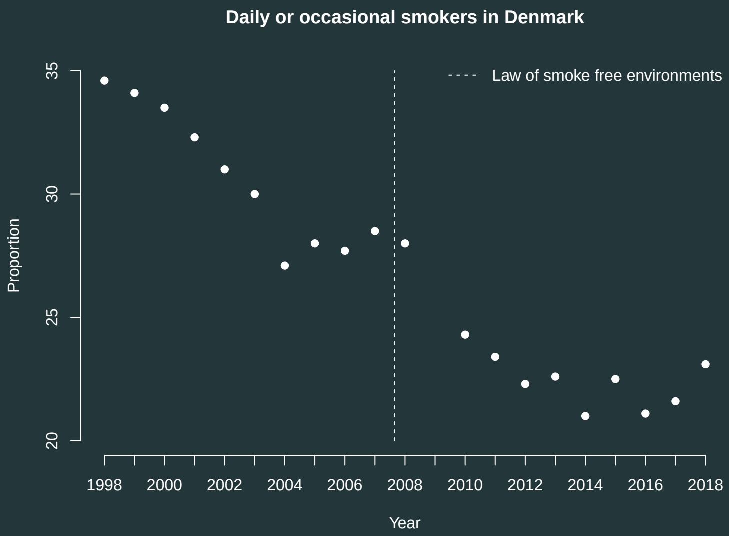
Hjerteforeningen, Sundhedsstyrelsen, Kræftens Bekæmpelse og Lungeforeningen har undersøgt danskernes rygevaner. Undersøgelsen er baseret på et repræsentativt udsnit af danskerne i forhold til alder, køn, religion og uddannelse. I alt har 5.017 danskere deltaget i undersøgelsen. Dataindsamlingen er foretaget af TNS Gallup via deres internetpanel i november 2018.

*** Data fra undersøgelsen i 2009 er ikke medtaget, da borgere med lavt uddannelsesniveau var underrepræsenteret i forhold til tidligere undersøgelser.*

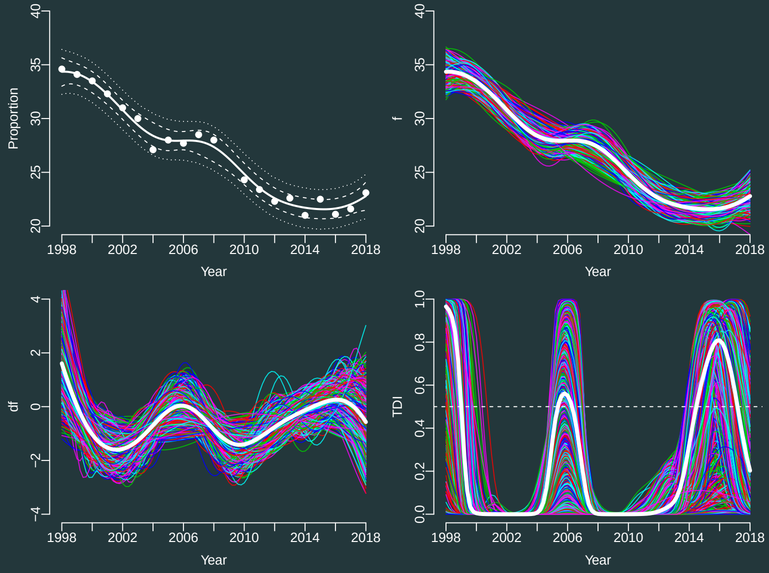
I denne undersøgelse er studiepopulationen imidlertid så stor, at stort set alle forskellene bliver statistisk signifikant på et 95% signifikansniveau.

[1] Nøgletal - Danskernes rygevaner 2018, www.sst.dk/da/udgivelser/2019/danskernes-rygevaner-2018

Proportion of smokers in Denmark



Model fit



Proportion
Year
f
df
TDI
1998
2002
2006
2010
2014
2018
20
25
30
35
40
−4
−2
0
2
4
0.0
0.2
0.4
0.6
0.8
1.0

Extensions

Stationarity and latent dynamics

A stationary covariance function $C_\theta(s,t) = C_\theta(|s-t|)$ is often used out of convenience. In the previous model we used the exponential squared covariance function. It can be shown that $\partial_2 C_\theta(t,t) = 0$ for all stationary covariance functions hence $f(t) \perp df(t)$.

Stationarity and latent dynamics

A stationary covariance function $C_\theta(s,t) = C_\theta(|s-t|)$ is often used out of convenience. In the previous model we used the exponential squared covariance function. It can be shown that $\partial_2 C_\theta(t,t) = 0$ for all stationary covariance functions hence $f(t) \perp\!\!\!\perp df(t)$.

By construction $df(t) \mid f(t), \Theta = \alpha(t) + \beta(t) f(t) + \epsilon(t)$ and therefore $\beta(t) = 0$ imposing a strong assumption on the latent dynamics.

To obtain $\beta(t) \neq 0$ we must enable non-stationarity. This can be achived flexibly by adding a functional latent class to the model.

Functional latent class model extension

We add a time-varying latent class as follows

$$
\begin{aligned}
w(t) \mid \phi &\sim S^{\infty}_{K-1}(\phi) \\
G(t) \mid w(t), \phi &\sim \text{Multinomial}(w(t)) \\
f(t) \mid G(t) = k, \theta_k, \mathbf{w}, \phi &\sim \mathcal{GP}(m, C_{\theta_k}(\cdot, \cdot))
\end{aligned}
$$

where S^{∞}_{K-1} is a distribution on the functional $K-1$ dimensional simplex

$$
\left\{ w_k(t), k = 1, \ldots, K, t \in \mathcal{T}, w_k(t) > 0, \sum_{k=1}^{K} w_k(t) = 1 \right\}
$$

Functional latent class model extension

Marginalizing out $G(t)$ we obtain

$$\begin{bmatrix} f(s) \\ df(t) \end{bmatrix} \mid m, \{w_k, \theta_k\}_{k=1}^K, \phi \sim \mathcal{GP}\left(\begin{bmatrix} m \\ 0 \end{bmatrix}, \begin{bmatrix} C_{11}(s,s') & C_{12}(s,t) \\ C_{21}(t,s) & C_{22}(t,t') \end{bmatrix} \right)$$

where

$$\begin{aligned} C_{11}(s,s') &= \sum_{k=1}^K w_k(s) C_{\theta_k}(s,s') w_k(s') \\ C_{12}(s,t) &= \partial_2 C_{11}(s,t) \\ &= \sum_{k=1}^K w_k(s) \partial_2 C_{\theta_k}(s,t) w_k(t) + \sum_{k=1}^K w_k(s) C_{\theta_k}(s,t) dw_k(t) \end{aligned}$$

and $C_{12}(t,t)$ is now generally non-zero and therefore conditionally $df(t) \not\perp\!\!\!\perp f(t)$.

Applying the non-stationary model

We use $K = 2$ hence $\mathbf{w}(t) = (w_1(t), 1 - w_1(t))$.

We model $w_1(t)$ by a transformed polynomial B-spline expansion with $P = 15$

$$w_1(t) = \operatorname{logit}^{-1}\left(\phi_0 + \sum_{p=1}^{P} \phi_p B(t)\right)$$

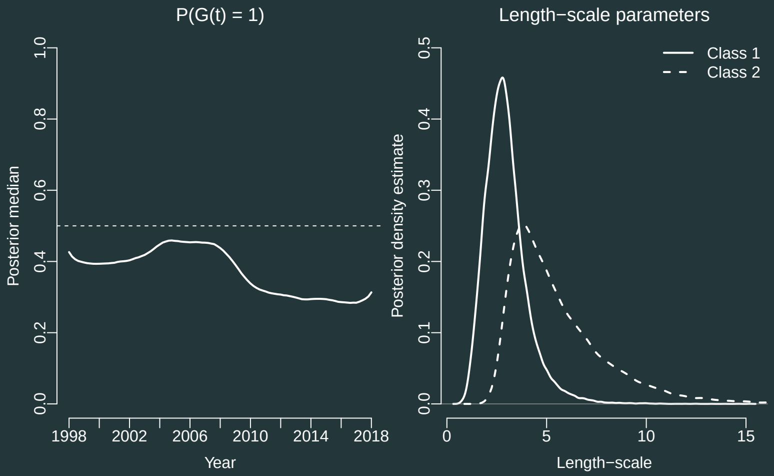
and regularize by an autogressive prior

$$\phi_1 \sim N(0,1), \quad \phi_p \sim N(\phi_{p-1}, \tau), \quad \tau \sim N(0,1)$$

Its derivative then has the closed form expression

$$dw_1(t) = \left(\sum_{p=1}^{P} \phi_p \frac{\mathrm{d}}{\mathrm{dt}} B_p(t)\right) \mathrm{d}\operatorname{logit}^{-1}(t)\left(\phi_0 + \sum_{p=1}^{P} \phi_p B_p(t)\right)$$

Latent class fit



A useful link

https://github.com/aejensen/TrendinessOfTrends

- Stan implementations
- Data and code to reproduce the figures
- Work-in-progress preprint of the manuscript
- These slides

Thank you!