Approximate Bayesian Inference for Case Crossover Models

Alex Stringer, Patrick Brown, Jamie Stafford

2020-04-17

University of Toronto and Centre for Global Health Research, St. Michael's Hospital

Outline

Introduction and motivation

Introduction

A **case crossover** model quantifies the association between mortality/morbidity and short-term exposure to risk factors.

Example: are short-term increases in cigarette smoking associated with increased risk of heart attack?

Example: is using a cell-phone while driving associated with increased risk of crash?

Example: are spikes or drops in daily temperature associated with increased mortality risk in India?

Subject's exposure on day of death is compared to their exposure on one or more previous days on which they did not die.

Higher exposure on day of death provides evidence of a positive association between exposure and mortality risk.

Introduction

A **case crossover** model quantifies the association between mortality/morbidity and short-term exposure to risk factors.

Example: are short-term increases in cigarette smoking associated with increased risk of heart attack?

Example: is using a cell-phone while driving associated with increased risk of crash?

Example: are spikes or drops in daily temperature associated with increased mortality risk in India?

Subject's exposure on day of death is compared to their exposure on one or more previous days on which they did not die.

Higher exposure on day of death provides evidence of a positive association between exposure and mortality risk.

Introduction

A **case crossover** model quantifies the association between mortality/morbidity and short-term exposure to risk factors.

Example: are short-term increases in cigarette smoking associated with increased risk of heart attack?

Example: is using a cell-phone while driving associated with increased risk of crash?

Example: are spikes or drops in daily temperature associated with increased mortality risk in India?

Subject's exposure on day of death is compared to their exposure on one or more previous days on which they did not die.

Higher exposure on day of death provides evidence of a positive association between exposure and mortality risk.

Introduction

A **case crossover** model quantifies the association between mortality/morbidity and short-term exposure to risk factors.

Example: are short-term increases in cigarette smoking associated with increased risk of heart attack?

Example: is using a cell-phone while driving associated with increased risk of crash?

Example: are spikes or drops in daily temperature associated with increased mortality risk in India?

Subject's exposure on day of death is compared to their exposure on one or more previous days on which they did not die.

Higher exposure on day of death provides evidence of a positive association between exposure and mortality risk.

Introduction

A **case crossover** model quantifies the association between mortality/morbidity and short-term exposure to risk factors.

Example: are short-term increases in cigarette smoking associated with increased risk of heart attack?

Example: is using a cell-phone while driving associated with increased risk of crash?

Example: are spikes or drops in daily temperature associated with increased mortality risk in India?

Subject's exposure on day of death is compared to their exposure on one or more previous days on which they did not die.

Higher exposure on day of death provides evidence of a positive association between exposure and mortality risk.

Introduction

A **case crossover** model quantifies the association between mortality/morbidity and short-term exposure to risk factors.

Example: are short-term increases in cigarette smoking associated with increased risk of heart attack?

Example: is using a cell-phone while driving associated with increased risk of crash?

Example: are spikes or drops in daily temperature associated with increased mortality risk in India?

Subject's exposure on day of death is compared to their exposure on one or more previous days on which they did not die.

Higher exposure on day of death provides evidence of a positive association between exposure and mortality risk.

Motivating example: mortality/temperature in India

Indian Million Death Study (MDS): largest comprehensive prospective study of mortality in human history. Data curated by CGHR in Toronto in collaboration with Registrar General of India.

Representative sample of one million housholds in India tracked for 15 years, 2001 – 2016. All deaths recorded by verbal autopsy and classified by two independent physicians, with disputes resolved via moderation process.

Fantastically **accurate** and **comprehensive** source of data on human mortality.

Our question: are extreme **daily maximum temperatures** (very hot or very cold) associated with increased mortality risk due to **stroke**?

Motivating example: mortality/temperature in India

Indian Million Death Study (MDS): largest comprehensive prospective study of mortality in human history. Data curated by CGHR in Toronto in collaboration with Registrar General of India.

Representative sample of one million housholds in India tracked for 15 years, 2001 – 2016. All deaths recorded by verbal autopsy and classified by two independent physicians, with disputes resolved via moderation process.

Fantastically **accurate** and **comprehensive** source of data on human mortality.

Our question: are extreme **daily maximum temperatures** (very hot or very cold) associated with increased mortality risk due to **stroke**?

Motivating example: mortality/temperature in India

Indian Million Death Study (MDS): largest comprehensive prospective study of mortality in human history. Data curated by CGHR in Toronto in collaboration with Registrar General of India.

Representative sample of one million housholds in India tracked for 15 years, 2001 – 2016. All deaths recorded by verbal autopsy and classified by two independent physicians, with disputes resolved via moderation process.

Fantastically **accurate** and **comprehensive** source of data on human mortality.

Our question: are extreme **daily maximum temperatures** (very hot or very cold) associated with increased mortality risk due to **stroke**?

Motivating example: mortality/temperature in India

Indian Million Death Study (MDS): largest comprehensive prospective study of mortality in human history. Data curated by CGHR in Toronto in collaboration with Registrar General of India.

Representative sample of one million housholds in India tracked for 15 years, 2001 – 2016. All deaths recorded by verbal autopsy and classified by two independent physicians, with disputes resolved via moderation process.

Fantastically **accurate** and **comprehensive** source of data on human mortality.

Our question: are extreme **daily maximum temperatures** (very hot or very cold) associated with increased mortality risk due to **stroke**?

Motivating example: mortality/temperature in India

Approach: link $\approx 13,000$ people who died from stroke in the MDS data with existing temperature data from India. Look at the max temperature on each subject's date of death and that on days from 3 – 5 weeks prior. Data already exists, no need to design a study or collect new data for this.

Fu et al. (2018) did this using existing methods:

Motivating example: mortality/temperature in India

Approach: link $\approx 13,000$ people who died from stroke in the MDS data with existing temperature data from India. Look at the max temperature on each subject's date of death and that on days from 3 – 5 weeks prior. Data already exists, no need to design a study or collect new data for this.

Fu et al. (2018) did this using existing methods:

Motivating example: mortality/temperature in India

Approach: link $\approx 13,000$ people who died from stroke in the MDS data with existing temperature data from India. Look at the max temperature on each subject's date of death and that on days from 3 – 5 weeks prior. Data already exists, no need to design a study or collect new data for this.

Fu et al. (2018) did this using existing methods:

Motivating example: mortality/temperature in India

Approach: link $\approx 13,000$ people who died from stroke in the MDS data with existing temperature data from India. Look at the max temperature on each subject's date of death and that on days from 3 – 5 weeks prior. Data already exists, no need to design a study or collect new data for this.

Fu et al. (2018) did this using existing methods:

Motivating example: mortality/temperature in India

Approach: link $\approx 13,000$ people who died from stroke in the MDS data with existing temperature data from India. Look at the max temperature on each subject's date of death and that on days from 3 – 5 weeks prior. Data already exists, no need to design a study or collect new data for this.

Fu et al. (2018) did this using existing methods:

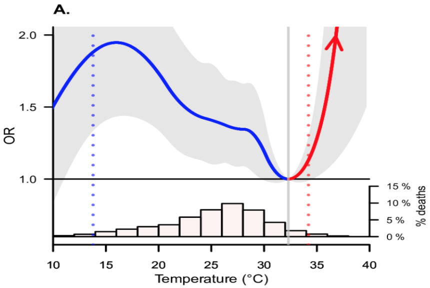

Figure 1: Risk of stroke mortality relative to that at 32 degrees celcius.

Fu et al. (2018) uses **spline** methods and **maximum likelihood** within a modified Cox proportional hazards model. This requires choosing the **number** and **placement** of spline knots:

Fu et al. (2018) uses **spline** methods and **maximum likelihood** within a modified Cox proportional hazards model. This requires choosing the **number** and **placement** of spline knots:

Fu et al. (2018) uses **spline** methods and **maximum likelihood** within a modified Cox proportional hazards model. This requires choosing the **number** and **placement** of spline knots:

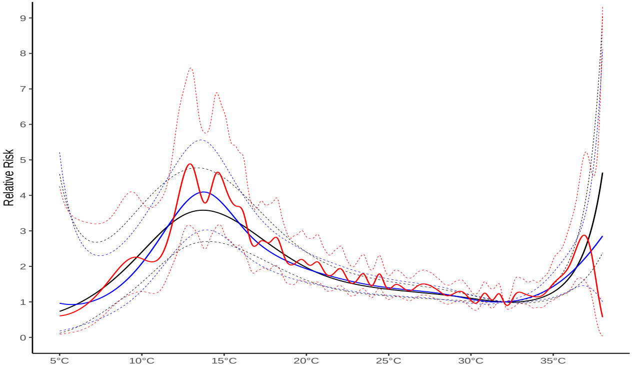

Figure 2: The results are sensitive to the number and placement of the spline knots.

(Note: y-axis differs because they use daily average temperature and we use daily maximum temperature. Pattern is the same.)

Challenges

Challenges:

1. Different **spline knots** will lead to different curves.
2. No **model-based** way to estimate and quantify the uncertainty in the smoothness of the curve.
3. Even if you have a good idea of how to approach 1. and 2., there is no principled way to incorporate this **prior knowledge** about what the curve should look like.

Bayesian inference and **random walk smoothing** address these challenges, but lead to **intractable computations**.

We will develop an approximate Bayesian inference methodology for case crossover models that is **computationally tractable**, provides **model-based** estimation and uncertainty quantification for the smoothness of the curve, and is **not sensitive** to the number and placement of knots ("knot sensitive"... sorry I couldn't help myself).

Challenges

Challenges:

1. Different **spline knots** will lead to different curves.
2. No **model-based** way to estimate and quantify the uncertainty in the smoothness of the curve.
3. Even if you have a good idea of how to approach 1. and 2., there is no principled way to incorporate this **prior knowledge** about what the curve should look like.

Bayesian inference and **random walk smoothing** address these challenges, but lead to **intractable computations**.

We will develop an approximate Bayesian inference methodology for case crossover models that is **computationally tractable**, provides **model-based** estimation and uncertainty quantification for the smoothness of the curve, and is **not sensitive** to the number and placement of knots ("knot sensitive"... sorry I couldn't help myself).

Challenges

Challenges:

1. Different **spline knots** will lead to different curves.
2. No **model-based** way to estimate and quantify the uncertainty in the smoothness of the curve.
3. Even if you have a good idea of how to approach 1. and 2., there is no principled way to incorporate this **prior knowledge** about what the curve should look like.

Bayesian inference and **random walk smoothing** address these challenges, but lead to **intractable computations**.

We will develop an approximate Bayesian inference methodology for case crossover models that is **computationally tractable**, provides **model-based** estimation and uncertainty quantification for the smoothness of the curve, and is **not sensitive** to the number and placement of knots ("knot sensitive"... sorry I couldn't help myself).

Challenges

Challenges:

1. Different **spline knots** will lead to different curves.
2. No **model-based** way to estimate and quantify the uncertainty in the smoothness of the curve.
3. Even if you have a good idea of how to approach 1. and 2., there is no principled way to incorporate this **prior knowledge** about what the curve should look like.

Bayesian inference and **random walk smoothing** address these challenges, but lead to **intractable computations**.

We will develop an approximate Bayesian inference methodology for case crossover models that is **computationally tractable**, provides **model-based** estimation and uncertainty quantification for the smoothness of the curve, and is **not sensitive** to the number and placement of knots ("knot sensitive"... sorry I couldn't help myself).

Challenges

Challenges:

1. Different **spline knots** will lead to different curves.
2. No **model-based** way to estimate and quantify the uncertainty in the smoothness of the curve.
3. Even if you have a good idea of how to approach 1. and 2., there is no principled way to incorporate this **prior knowledge** about what the curve should look like.

Bayesian inference and **random walk smoothing** address these challenges, but lead to **intractable computations**.

We will develop an approximate Bayesian inference methodology for case crossover models that is **computationally tractable**, provides **model-based** estimation and uncertainty quantification for the smoothness of the curve, and is **not sensitive** to the number and placement of knots ("knot sensitive"... sorry I couldn't help myself).

Challenges

Challenges:

1. Different **spline knots** will lead to different curves.
2. No **model-based** way to estimate and quantify the uncertainty in the smoothness of the curve.
3. Even if you have a good idea of how to approach 1. and 2., there is no principled way to incorporate this **prior knowledge** about what the curve should look like.

Bayesian inference and **random walk smoothing** address these challenges, but lead to **intractable computations**.

We will develop an approximate Bayesian inference methodology for case crossover models that is **computationally tractable**, provides **model-based** estimation and uncertainty quantification for the smoothness of the curve, and is **not sensitive** to the number and placement of knots ("knot sensitive"... sorry I couldn't help myself).

Challenges

Challenges:

1. Different **spline knots** will lead to different curves.
2. No **model-based** way to estimate and quantify the uncertainty in the smoothness of the curve.
3. Even if you have a good idea of how to approach 1. and 2., there is no principled way to incorporate this **prior knowledge** about what the curve should look like.

Bayesian inference and **random walk smoothing** address these challenges, but lead to **intractable computations**.

We will develop an approximate Bayesian inference methodology for case crossover models that is **computationally tractable**, provides **model-based** estimation and uncertainty quantification for the smoothness of the curve, and is **not sensitive** to the number and placement of knots ("knot sensitive"... sorry I couldn't help myself).

Challenges

Challenges:

1. Different **spline knots** will lead to different curves.
2. No **model-based** way to estimate and quantify the uncertainty in the smoothness of the curve.
3. Even if you have a good idea of how to approach 1. and 2., there is no principled way to incorporate this **prior knowledge** about what the curve should look like.

Bayesian inference and **random walk smoothing** address these challenges, but lead to **intractable computations**.

We will develop an approximate Bayesian inference methodology for case crossover models that is **computationally tractable**, provides **model-based** estimation and uncertainty quantification for the smoothness of the curve, and is **not sensitive** to the number and placement of knots ("knot sensitive"... sorry I couldn't help myself).

Challenges

Challenges:

1. Different **spline knots** will lead to different curves.
2. No **model-based** way to estimate and quantify the uncertainty in the smoothness of the curve.
3. Even if you have a good idea of how to approach 1. and 2., there is no principled way to incorporate this **prior knowledge** about what the curve should look like.

Bayesian inference and **random walk smoothing** address these challenges, but lead to **intractable computations**.

We will develop an approximate Bayesian inference methodology for case crossover models that is **computationally tractable**, provides **model-based** estimation and uncertainty quantification for the smoothness of the curve, and is **not sensitive** to the number and placement of knots ("knot sensitive"... sorry I couldn't help myself).

Model and Inference Methodology

Case Crossover Models (Stringer et al., 2020a)

Let $Y_i, i \in [n]$ represent the death date of the i^{th} subject, $i = 1, \ldots, n = 13,493$. Let y_i be its realization in the observed data.

Let $S_i = \{c_{i1}, \ldots, c_{iT_i}\}$ be the set of **control days**–days on which the subject *could* have died but didn't–for each subject. Chosen by the analyst as part of the study design.

For $t \in \bigcup_{i=1}^{n} \{y_i\} \cup S_i$ let x_t denote the maximum temperature on day t.

Let $\lambda(x_t)$ represent the probability of dying at temperature x_t.

Assumption: each subject's **baseline risk** of dying due to time-independent risk factors is the same on day y_i and each day $t \in S_i$.

The model is:

$$Y_i|\lambda(\cdot) \overset{ind}{\sim} \text{Multinomial}\left[\lambda(x_{y_i}), \lambda(x_{c_{i1}}), \ldots, \lambda(x_{c_{iT_i}})\right]$$

$$\eta_{it} = \log\lambda(x_t) = u(x_t), t \in \{y_i\} \cup S_i$$

Case Crossover Models (Stringer et al., 2020a)

Let $Y_i, i \in [n]$ represent the death date of the i^{th} subject, $i = 1, \ldots, n = 13,493$. Let y_i be its realization in the observed data.

Let $S_i = \{c_{i1}, \ldots, c_{iT_i}\}$ be the set of **control days**–days on which the subject *could* have died but didn't–for each subject. Chosen by the analyst as part of the study design.

For $t \in \bigcup_{i=1}^{n} \{y_i\} \cup S_i$ let x_t denote the maximum temperature on day t.

Let $\lambda(x_t)$ represent the probability of dying at temperature x_t.

Assumption: each subject's **baseline risk** of dying due to time-independent risk factors is the same on day y_i and each day $t \in S_i$.

The model is:

$$Y_i|\lambda(\cdot) \overset{ind}{\sim} \text{Multinomial}\left[\lambda(x_{y_i}), \lambda(x_{c_{i1}}), \ldots, \lambda(x_{c_{iT_i}})\right]$$

$$\eta_{it} = \log \lambda(x_t) = u(x_t), t \in \{y_i\} \cup S_i$$

Case Crossover Models (Stringer et al., 2020a)

Let $Y_i, i \in [n]$ represent the death date of the i^{th} subject, $i = 1, \ldots, n = 13,493$. Let y_i be its realization in the observed data.

Let $S_i = \{c_{i1}, \ldots, c_{iT_i}\}$ be the set of **control days**–days on which the subject *could* have died but didn't–for each subject. Chosen by the analyst as part of the study design.

For $t \in \bigcup_{i=1}^{n} \{y_i\} \cup S_i$ let x_t denote the maximum temperature on day t.

Let $\lambda(x_t)$ represent the probability of dying at temperature x_t.

Assumption: each subject's **baseline risk** of dying due to time-independent risk factors is the same on day y_i and each day $t \in S_i$.

The model is:

$$Y_i|\lambda(\cdot) \overset{ind}{\sim} \text{Multinomial}\left[\lambda(x_{y_i}), \lambda(x_{c_{i1}}), \ldots, \lambda(x_{c_{iT_i}})\right]$$

$$\eta_{it} = \log \lambda(x_t) = u(x_t), t \in \{y_i\} \cup S_i$$

Case Crossover Models (Stringer et al., 2020a)

Let $Y_i, i \in [n]$ represent the death date of the i^{th} subject, $i = 1, \ldots, n = 13,493$. Let y_i be its realization in the observed data.

Let $S_i = \{c_{i1}, \ldots, c_{iT_i}\}$ be the set of **control days**–days on which the subject *could* have died but didn't–for each subject. Chosen by the analyst as part of the study design.

For $t \in \bigcup_{i=1}^{n} \{y_i\} \cup S_i$ let x_t denote the maximum temperature on day t.

Let $\lambda(x_t)$ represent the probability of dying at temperature x_t.

Assumption: each subject's **baseline risk** of dying due to time-independent risk factors is the same on day y_i and each day $t \in S_i$.

The model is:

$$Y_i | \lambda(\cdot) \overset{ind}{\sim} \text{Multinomial}\left[\lambda(x_{y_i}), \lambda(x_{c_{i1}}), \ldots, \lambda(x_{c_{iT_i}})\right]$$

$$\eta_{it} = \log \lambda(x_t) = u(x_t), t \in \{y_i\} \cup S_i$$

Case Crossover Models (Stringer et al., 2020a)

Let $Y_i, i \in [n]$ represent the death date of the i^{th} subject, $i = 1, \ldots, n = 13,493$. Let y_i be its realization in the observed data.

Let $S_i = \{c_{i1}, \ldots, c_{iT_i}\}$ be the set of **control days**–days on which the subject *could* have died but didn't–for each subject. Chosen by the analyst as part of the study design.

For $t \in \bigcup_{i=1}^{n} \{y_i\} \cup S_i$ let x_t denote the maximum temperature on day t.

Let $\lambda(x_t)$ represent the probability of dying at temperature x_t.

Assumption: each subject's **baseline risk** of dying due to time-independent risk factors is the same on day y_i and each day $t \in S_i$.

The model is:

$$Y_i|\lambda(\cdot) \overset{ind}{\sim} \text{Multinomial}\left[\lambda(x_{y_i}), \lambda(x_{c_{i1}}), \ldots, \lambda(x_{c_{iT_i}})\right]$$

$$\eta_{it} = \log \lambda(x_t) = u(x_t), t \in \{y_i\} \cup S_i$$

Case Crossover Models (Stringer et al., 2020a)

Let $Y_i, i \in [n]$ represent the death date of the i^{th} subject, $i = 1, \ldots, n = 13,493$. Let y_i be its realization in the observed data.

Let $S_i = \{c_{i1}, \ldots, c_{iT_i}\}$ be the set of **control days**–days on which the subject *could* have died but didn't–for each subject. Chosen by the analyst as part of the study design.

For $t \in \bigcup_{i=1}^{n} \{y_i\} \cup S_i$ let x_t denote the maximum temperature on day t.

Let $\lambda(x_t)$ represent the probability of dying at temperature x_t.

Assumption: each subject's **baseline risk** of dying due to time-independent risk factors is the same on day y_i and each day $t \in S_i$.

The model is:

$$Y_i|\lambda(\cdot) \overset{ind}{\sim} \text{Multinomial}\left[\lambda(x_{y_i}), \lambda(x_{c_{i1}}), \ldots, \lambda(x_{c_{iT_i}})\right]$$

$$\eta_{it} = \log \lambda(x_t) = u(x_t), t \in \{y_i\} \cup S_i$$

Smoothing and Bayesian Inference (Stringer et al., 2020a)

The object of inferential interest is the underlying **continuous process** $u(x)$.

Put a **Gaussian process** prior $u(\cdot)|\sigma(\cdot,\cdot) \sim \mathcal{GP}\left[0, \sigma(\cdot,\cdot)\right]$.

Discretize temperature into bins $B_1 = [10, 10.1), \ldots, B_d = [38, 38.1)$. $d = 277$ takes the place of the **number** and the midpoints of B_j take the place of the **placement** of spline knots.

For each $i \in [n], t \in \{y_i\} \cup S_i$ define $j_{i,t} = \{j : x(t) \in B_j\}$.

Define the **piecewise-constant** approximation $U_{j_{i,t}} = u(x_t), i \in [n], t \in \{y_i\} \cup S_i$.

The $\mathcal{GP}$ prior on $u(\cdot)$ is equivalent to a finite-dimensional Gaussian prior on $\boldsymbol{U} = (U_1, \ldots, U_d)|\sigma_u$.

The posterior median and quantiles for $U_j|\boldsymbol{Y}, j \in [d]$ are plotted, giving an estimate and uncertainty quantification for $u(\cdot)$.

The posterior mode and quantiles for $\sigma_u|\boldsymbol{Y}$ provide a model-based estimate and uncertainty quantification for the **smoothness** of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer et al., 2020a)

The object of inferential interest is the underlying **continuous process** $u(x)$.

Put a **Gaussian process** prior $u(\cdot)|\sigma(\cdot,\cdot) \sim \mathcal{GP}\left[0, \sigma(\cdot,\cdot)\right]$.

Discretize temperature into bins $B_1 = [10, 10.1), \ldots, B_d = [38, 38.1)$. $d = 277$ takes the place of the **number** and the midpoints of B_j take the place of the **placement** of spline knots.

For each $i \in [n], t \in \{y_i\} \cup S_i$ define $j_{i,t} = \{j : x(t) \in B_j\}$.

Define the **piecewise-constant** approximation $U_{j_{i,t}} = u(x_t), i \in [n], t \in \{y_i\} \cup S_i$.

The $\mathcal{GP}$ prior on $u(\cdot)$ is equivalent to a finite-dimensional Gaussian prior on $\boldsymbol{U} = (U_1, \ldots, U_d)|\sigma_u$.

The posterior median and quantiles for $U_j|\boldsymbol{Y}, j \in [d]$ are plotted, giving an estimate and uncertainty quantification for $u(\cdot)$.

The posterior mode and quantiles for $\sigma_u|\boldsymbol{Y}$ provide a model-based estimate and uncertainty quantification for the **smoothness** of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer et al., 2020a)

The object of inferential interest is the underlying **continuous process** $u(x)$.

Put a **Gaussian process** prior $u(\cdot)|\sigma(\cdot,\cdot) \sim \mathcal{GP}\left[0, \sigma(\cdot,\cdot)\right]$.

Discretize temperature into bins $B_1 = [10, 10.1), \ldots, B_d = [38, 38.1)$. $d = 277$ takes the place of the **number** and the midpoints of B_j take the place of the **placement** of spline knots.

For each $i \in [n], t \in \{y_i\} \cup S_i$ define $j_{i,t} = \{j : x(t) \in B_j\}$.

Define the **piecewise-constant** approximation $U_{j_{i,t}} = u(x_t), i \in [n], t \in \{y_i\} \cup S_i$.

The $\mathcal{GP}$ prior on $u(\cdot)$ is equivalent to a finite-dimensional Gaussian prior on $\boldsymbol{U} = (U_1, \ldots, U_d)|\sigma_u$.

The posterior median and quantiles for $U_j|\boldsymbol{Y}, j \in [d]$ are plotted, giving an estimate and uncertainty quantification for $u(\cdot)$.

The posterior mode and quantiles for $\sigma_u|\boldsymbol{Y}$ provide a model-based estimate and uncertainty quantification for the **smoothness** of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer et al., 2020a)

The object of inferential interest is the underlying **continuous process** $u(x)$.

Put a **Gaussian process** prior $u(\cdot)|\sigma(\cdot,\cdot) \sim \mathcal{GP}\left[0, \sigma(\cdot,\cdot)\right]$.

Discretize temperature into bins $B_1 = [10, 10.1), \ldots, B_d = [38, 38.1)$. $d = 277$ takes the place of the **number** and the midpoints of B_j take the place of the **placement** of spline knots.

For each $i \in [n], t \in \{y_i\} \cup S_i$ define $j_{i,t} = \{j : x(t) \in B_j\}$.

Define the **piecewise-constant** approximation $U_{j_{i,t}} = u(x_t), i \in [n], t \in \{y_i\} \cup S_i$.

The $\mathcal{GP}$ prior on $u(\cdot)$ is equivalent to a finite-dimensional Gaussian prior on $\boldsymbol{U} = (U_1, \ldots, U_d)|\sigma_u$.

The posterior median and quantiles for $U_j|\boldsymbol{Y}, j \in [d]$ are plotted, giving an estimate and uncertainty quantification for $u(\cdot)$.

The posterior mode and quantiles for $\sigma_u|\boldsymbol{Y}$ provide a model-based estimate and uncertainty quantification for the **smoothness** of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer et al., 2020a)

The object of inferential interest is the underlying **continuous process** $u(x)$.

Put a **Gaussian process** prior $u(\cdot)|\sigma(\cdot,\cdot) \sim \mathcal{GP}\left[0, \sigma(\cdot,\cdot)\right]$.

Discretize temperature into bins $B_1 = [10, 10.1), \ldots, B_d = [38, 38.1)$. $d = 277$ takes the place of the **number** and the midpoints of B_j take the place of the **placement** of spline knots.

For each $i \in [n], t \in \{y_i\} \cup S_i$ define $j_{i,t} = \{j : x(t) \in B_j\}$.

Define the **piecewise-constant** approximation $U_{j_{i,t}} = u(x_t), i \in [n], t \in \{y_i\} \cup S_i$.

The $\mathcal{GP}$ prior on $u(\cdot)$ is equivalent to a finite-dimensional Gaussian prior on $\boldsymbol{U} = (U_1, \ldots, U_d)|\sigma_u$.

The posterior median and quantiles for $U_j|\boldsymbol{Y}, j \in [d]$ are plotted, giving an estimate and uncertainty quantification for $u(\cdot)$.

The posterior mode and quantiles for $\sigma_u|\boldsymbol{Y}$ provide a model-based estimate and uncertainty quantification for the **smoothness** of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer et al., 2020a)

The object of inferential interest is the underlying **continuous process** $u(x)$.

Put a **Gaussian process** prior $u(\cdot)|\sigma(\cdot,\cdot) \sim \mathcal{GP}\left[0, \sigma(\cdot,\cdot)\right]$.

Discretize temperature into bins $B_1 = [10, 10.1), \ldots, B_d = [38, 38.1)$. $d = 277$ takes the place of the **number** and the midpoints of B_j take the place of the **placement** of spline knots.

For each $i \in [n], t \in \{y_i\} \cup S_i$ define $j_{i,t} = \{j : x(t) \in B_j\}$.

Define the **piecewise-constant** approximation $U_{j_{i,t}} = u(x_t), i \in [n], t \in \{y_i\} \cup S_i$.

The $\mathcal{GP}$ prior on $u(\cdot)$ is equivalent to a finite-dimensional Gaussian prior on $\boldsymbol{U} = (U_1, \ldots, U_d)|\sigma_u$.

The posterior median and quantiles for $U_j|\boldsymbol{Y}, j \in [d]$ are plotted, giving an estimate and uncertainty quantification for $u(\cdot)$.

The posterior mode and quantiles for $\sigma_u|\boldsymbol{Y}$ provide a model-based estimate and uncertainty quantification for the **smoothness** of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer et al., 2020a)

The object of inferential interest is the underlying **continuous process** $u(x)$.

Put a **Gaussian process** prior $u(\cdot)|\sigma(\cdot,\cdot) \sim \mathcal{GP}\left[0, \sigma(\cdot,\cdot)\right]$.

Discretize temperature into bins $B_1 = [10, 10.1), \ldots, B_d = [38, 38.1)$. $d = 277$ takes the place of the **number** and the midpoints of B_j take the place of the **placement** of spline knots.

For each $i \in [n], t \in \{y_i\} \cup S_i$ define $j_{i,t} = \{j : x(t) \in B_j\}$.

Define the **piecewise-constant** approximation $U_{j_{i,t}} = u(x_t), i \in [n], t \in \{y_i\} \cup S_i$.

The $\mathcal{GP}$ prior on $u(\cdot)$ is equivalent to a finite-dimensional Gaussian prior on $\boldsymbol{U} = (U_1, \ldots, U_d)|\sigma_u$.

The posterior median and quantiles for $U_j|\boldsymbol{Y}, j \in [d]$ are plotted, giving an estimate and uncertainty quantification for $u(\cdot)$.

The posterior mode and quantiles for $\sigma_u|\boldsymbol{Y}$ provide a model-based estimate and uncertainty quantification for the **smoothness** of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer et al., 2020a)

The object of inferential interest is the underlying **continuous process** $u(x)$.

Put a **Gaussian process** prior $u(\cdot)|\sigma(\cdot,\cdot) \sim \mathcal{GP}\left[0, \sigma(\cdot,\cdot)\right]$.

Discretize temperature into bins $B_1 = [10, 10.1), \ldots, B_d = [38, 38.1)$. $d = 277$ takes the place of the **number** and the midpoints of B_j take the place of the **placement** of spline knots.

For each $i \in [n], t \in \{y_i\} \cup S_i$ define $j_{i,t} = \{j : x(t) \in B_j\}$.

Define the **piecewise-constant** approximation $U_{j_{i,t}} = u(x_t), i \in [n], t \in \{y_i\} \cup S_i$.

The $\mathcal{GP}$ prior on $u(\cdot)$ is equivalent to a finite-dimensional Gaussian prior on $\boldsymbol{U} = (U_1, \ldots, U_d)|\sigma_u$.

The posterior median and quantiles for $U_j|\boldsymbol{Y}, j \in [d]$ are plotted, giving an estimate and uncertainty quantification for $u(\cdot)$.

The posterior mode and quantiles for $\sigma_u|\boldsymbol{Y}$ provide a model-based estimate and uncertainty quantification for the **smoothness** of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer and Brown, 2020)

Because data is only recorded on subjects who died, overall mortality risk in not estimable. Must infer risk **relative** to that at a chosen reference point.

Some care is required in parametrizing the smoothing models in this setting- outside scope for this talk.

Use a constrained **third-order random walk**:

$$-U_{j+2} + 3U_{j+1} - 3U_j + U_{j-1} | \sigma_u^2 \overset{iid}{\sim} \mathrm{N}\left(0, .1^2\sigma_u^2\right), j = 2, \ldots, d-2$$

$$(U_{31}, U_{33} - U_{31}) \overset{ind}{\sim} \mathrm{N}\left[0, 1000 \times \mathrm{diag}(1, .1)\right]$$

$$U_{32} = 0$$

This encodes our **prior belief** that $u(\cdot)$ is **smooth**. How smooth is determined by σ_u, and σ_u is estimated by the data (with uncertainty quantification!)

Changing the bins (spline knots) changes the smoothness, but not the shape, of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer and Brown, 2020)

Because data is only recorded on subjects who died, overall mortality risk in not estimable. Must infer risk **relative** to that at a chosen reference point.

Some care is required in parametrizing the smoothing models in this setting- outside scope for this talk.

Use a constrained **third-order random walk**:

$$
\begin{aligned}
-U_{j+2} + 3U_{j+1} - 3U_j + U_{j-1} | \sigma_u^2 &\overset{iid}{\sim} \mathrm{N}\left(0, .1^2\sigma_u^2\right), j = 2, \ldots, d-2 \\
(U_{31}, U_{33} - U_{31}) &\overset{ind}{\sim} \mathrm{N}\left[0, 1000 \times \mathrm{diag}(1, .1)\right] \\
U_{32} &= 0
\end{aligned}
$$

This encodes our **prior belief** that $u(\cdot)$ is **smooth**. How smooth is determined by σ_u, and σ_u is estimated by the data (with uncertainty quantification!)

Changing the bins (spline knots) changes the smoothness, but not the shape, of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer and Brown, 2020)

Because data is only recorded on subjects who died, overall mortality risk in not estimable. Must infer risk **relative** to that at a chosen reference point.

Some care is required in parametrizing the smoothing models in this setting- outside scope for this talk.

Use a constrained **third-order random walk**:

$$-U_{j+2} + 3U_{j+1} - 3U_j + U_{j-1} | \sigma_u^2 \overset{iid}{\sim} \mathrm{N}\left(0, .1^2\sigma_u^2\right), j = 2, \ldots, d-2$$

$$(U_{31}, U_{33} - U_{31}) \overset{ind}{\sim} \mathrm{N}\left[0, 1000 \times \mathrm{diag}(1, .1)\right]$$

$$U_{32} = 0$$

This encodes our **prior belief** that $u(\cdot)$ is **smooth**. How smooth is determined by σ_u, and σ_u is estimated by the data (with uncertainty quantification!)

Changing the bins (spline knots) changes the smoothness, but not the shape, of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer and Brown, 2020)

Because data is only recorded on subjects who died, overall mortality risk in not estimable. Must infer risk **relative** to that at a chosen reference point.

Some care is required in parametrizing the smoothing models in this setting- outside scope for this talk.

Use a constrained **third-order random walk**:

$$
\begin{aligned}
-U_{j+2} + 3U_{j+1} - 3U_j + U_{j-1} | \sigma_u^2 &\overset{iid}{\sim} \mathrm{N}\left(0, .1^2\sigma_u^2\right), j = 2, \ldots, d-2 \\
(U_{31}, U_{33} - U_{31}) &\overset{ind}{\sim} \mathrm{N}\left[0, 1000 \times \mathrm{diag}(1, .1)\right] \\
U_{32} &= 0
\end{aligned}
$$

This encodes our **prior belief** that $u(\cdot)$ is **smooth**. How smooth is determined by σ_u, and σ_u is estimated by the data (with uncertainty quantification!)

Changing the bins (spline knots) changes the smoothness, but not the shape, of $u(\cdot)$.

Smoothing and Bayesian Inference (Stringer and Brown, 2020)

Because data is only recorded on subjects who died, overall mortality risk in not estimable. Must infer risk **relative** to that at a chosen reference point.

Some care is required in parametrizing the smoothing models in this setting- outside scope for this talk.

Use a constrained **third-order random walk**:

$$-U_{j+2} + 3U_{j+1} - 3U_j + U_{j-1} | \sigma_u^2 \overset{iid}{\sim} \mathrm{N}\left(0, .1^2 \sigma_u^2\right), j = 2, \ldots, d-2$$

$$(U_{31}, U_{33} - U_{31}) \overset{ind}{\sim} \mathrm{N}\left[0, 1000 \times \mathrm{diag}(1, .1)\right]$$

$$U_{32} = 0$$

This encodes our **prior belief** that $u(\cdot)$ is **smooth**. How smooth is determined by σ_u, and σ_u is estimated by the data (with uncertainty quantification!)

Changing the bins (spline knots) changes the smoothness, but not the shape, of $u(\cdot)$.

Approximate Bayesian Inference (Stringer et al., 2020b)

Let $\theta = -2\log\sigma_u$. The objects of inferential interest are the posteriors

$$\pi(U_j|\boldsymbol{Y}) = \int\int \pi(\boldsymbol{U}|\boldsymbol{Y},\theta)\pi(\theta|\boldsymbol{Y})d\boldsymbol{U}_{-j}d\theta, j \in [d]$$

$$\pi(\theta|\boldsymbol{Y}) = \frac{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}}{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}d\theta}$$

Yikes.

Choose a **quadrature grid** and corresponding weights $\left\{\theta^k, \Delta^k : k \in [K]\right\}$. For each $k \in [K]$...

1. Approximate $\pi(\theta^k|\boldsymbol{Y}) \approx \tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})$, a **Laplace approximation** (Tierney and Kadane, 1986),
2. Approximate $\pi(\boldsymbol{U}|\boldsymbol{Y},\theta^k) \approx \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)$, a **Gaussian approximation**. This also gives a univariate Gaussian approximation $\tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k) = \int \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)d\boldsymbol{U}_{-j}$.

Then approximate $\pi(U_j|\boldsymbol{Y}) \approx \sum_{k=1}^{K} \tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k)\tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})\Delta^k$.

Approximate Bayesian Inference (Stringer et al., 2020b)

Let $\theta = -2\log\sigma_u$. The objects of inferential interest are the posteriors

$$\pi(U_j|\boldsymbol{Y}) = \int\int \pi(\boldsymbol{U}|\boldsymbol{Y},\theta)\pi(\theta|\boldsymbol{Y})d\boldsymbol{U}_{-j}d\theta, j \in [d]$$

$$\pi(\theta|\boldsymbol{Y}) = \frac{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}}{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}d\theta}$$

Yikes.

Choose a **quadrature grid** and corresponding weights $\left\{\theta^k, \Delta^k : k \in [K]\right\}$. For each $k \in [K]$...

1. Approximate $\pi(\theta^k|\boldsymbol{Y}) \approx \tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})$, a **Laplace approximation** (Tierney and Kadane, 1986),
2. Approximate $\pi(\boldsymbol{U}|\boldsymbol{Y},\theta^k) \approx \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)$, a **Gaussian approximation**. This also gives a univariate Gaussian approximation $\tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k) = \int \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)d\boldsymbol{U}_{-j}$.

Then approximate $\pi(U_j|\boldsymbol{Y}) \approx \sum_{k=1}^K \tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k)\tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})\Delta^k$.

Approximate Bayesian Inference (Stringer et al., 2020b)

Let $\theta = -2\log\sigma_u$. The objects of inferential interest are the posteriors

$$\pi(U_j|\boldsymbol{Y}) = \int\int \pi(\boldsymbol{U}|\boldsymbol{Y},\theta)\pi(\theta|\boldsymbol{Y})d\boldsymbol{U}_{-j}d\theta, j \in [d]$$

$$\pi(\theta|\boldsymbol{Y}) = \frac{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}}{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}d\theta}$$

Yikes.

Choose a **quadrature grid** and corresponding weights $\left\{\theta^k, \Delta^k : k \in [K]\right\}$. For each $k \in [K]$...

1. Approximate $\pi(\theta^k|\boldsymbol{Y}) \approx \tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})$, a **Laplace approximation** (Tierney and Kadane, 1986),
2. Approximate $\pi(\boldsymbol{U}|\boldsymbol{Y},\theta^k) \approx \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)$, a **Gaussian approximation**. This also gives a univariate Gaussian approximation $\tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k) = \int \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)d\boldsymbol{U}_{-j}$.

Then approximate $\pi(U_j|\boldsymbol{Y}) \approx \sum_{k=1}^K \tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k)\tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})\Delta^k$.

Approximate Bayesian Inference (Stringer et al., 2020b)

Let $\theta = -2\log\sigma_u$. The objects of inferential interest are the posteriors

$$\pi(U_j|\boldsymbol{Y}) = \int\int \pi(\boldsymbol{U}|\boldsymbol{Y},\theta)\pi(\theta|\boldsymbol{Y})d\boldsymbol{U}_{-j}d\theta, j \in [d]$$

$$\pi(\theta|\boldsymbol{Y}) = \frac{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}}{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}d\theta}$$

Yikes.

Choose a **quadrature grid** and corresponding weights $\left\{\theta^k, \Delta^k : k \in [K]\right\}$. For each $k \in [K]$...

1. Approximate $\pi(\theta^k|\boldsymbol{Y}) \approx \tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})$, a **Laplace approximation** (Tierney and Kadane, 1986),
2. Approximate $\pi(\boldsymbol{U}|\boldsymbol{Y},\theta^k) \approx \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)$, a **Gaussian approximation**. This also gives a univariate Gaussian approximation $\tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k) = \int \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)d\boldsymbol{U}_{-j}$.

Then approximate $\pi(U_j|\boldsymbol{Y}) \approx \sum_{k=1}^{K} \tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k)\tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})\Delta^k$.

Approximate Bayesian Inference (Stringer et al., 2020b)

Let $\theta = -2\log\sigma_u$. The objects of inferential interest are the posteriors

$$\pi(U_j|\boldsymbol{Y}) = \int\int \pi(\boldsymbol{U}|\boldsymbol{Y},\theta)\pi(\theta|\boldsymbol{Y})d\boldsymbol{U}_{-j}d\theta, j \in [d]$$

$$\pi(\theta|\boldsymbol{Y}) = \frac{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}}{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}d\theta}$$

Yikes.

Choose a **quadrature grid** and corresponding weights $\left\{\theta^k, \Delta^k : k \in [K]\right\}$. For each $k \in [K]$...

1. Approximate $\pi(\theta^k|\boldsymbol{Y}) \approx \tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})$, a **Laplace approximation** (Tierney and Kadane, 1986),
2. Approximate $\pi(\boldsymbol{U}|\boldsymbol{Y},\theta^k) \approx \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)$, a **Gaussian approximation**. This also gives a univariate Gaussian approximation $\tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k) = \int \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)d\boldsymbol{U}_{-j}$.

Then approximate $\pi(U_j|\boldsymbol{Y}) \approx \sum_{k=1}^{K} \tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k)\tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})\Delta^k$.

Approximate Bayesian Inference (Stringer et al., 2020b)

Let $\theta = -2 \log \sigma_u$. The objects of inferential interest are the posteriors

$$\pi(U_j|\boldsymbol{Y}) = \int \int \pi(\boldsymbol{U}|\boldsymbol{Y}, \theta)\pi(\theta|\boldsymbol{Y}) d\boldsymbol{U}_{-j} d\theta, j \in [d]$$

$$\pi(\theta|\boldsymbol{Y}) = \frac{\int \pi(\boldsymbol{U}, \theta, \boldsymbol{Y}) d\boldsymbol{U}}{\int \pi(\boldsymbol{U}, \theta, \boldsymbol{Y}) d\boldsymbol{U} d\theta}$$

Yikes.

Choose a **quadrature grid** and corresponding weights $\left\{\theta^k, \Delta^k : k \in [K]\right\}$. For each $k \in [K]$...

1. Approximate $\pi(\theta^k|\boldsymbol{Y}) \approx \tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})$, a **Laplace approximation** (Tierney and Kadane, 1986),
2. Approximate $\pi(\boldsymbol{U}|\boldsymbol{Y}, \theta^k) \approx \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y}, \theta^k)$, a **Gaussian approximation**. This also gives a univariate Gaussian approximation $\tilde{\pi}_G(U_j|\boldsymbol{Y}, \theta^k) = \int \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y}, \theta^k) d\boldsymbol{U}_{-j}$.

Then approximate $\pi(U_j|\boldsymbol{Y}) \approx \sum_{k=1}^{K} \tilde{\pi}_G(U_j|\boldsymbol{Y}, \theta^k)\tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})\Delta^k$.

Approximate Bayesian Inference (Stringer et al., 2020b)

Let $\theta = -2\log\sigma_u$. The objects of inferential interest are the posteriors

$$\pi(U_j|\boldsymbol{Y}) = \int\int \pi(\boldsymbol{U}|\boldsymbol{Y},\theta)\pi(\theta|\boldsymbol{Y})d\boldsymbol{U}_{-j}d\theta, j \in [d]$$

$$\pi(\theta|\boldsymbol{Y}) = \frac{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}}{\int \pi(\boldsymbol{U},\theta,\boldsymbol{Y})d\boldsymbol{U}d\theta}$$

Yikes.

Choose a **quadrature grid** and corresponding weights $\left\{\theta^k, \Delta^k : k \in [K]\right\}$. For each $k \in [K]$...

1. Approximate $\pi(\theta^k|\boldsymbol{Y}) \approx \tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})$, a **Laplace approximation** (Tierney and Kadane, 1986),
2. Approximate $\pi(\boldsymbol{U}|\boldsymbol{Y},\theta^k) \approx \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)$, a **Gaussian approximation**. This also gives a univariate Gaussian approximation $\tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k) = \int \tilde{\pi}_G(\boldsymbol{U}|\boldsymbol{Y},\theta^k)d\boldsymbol{U}_{-j}$.

Then approximate $\pi(U_j|\boldsymbol{Y}) \approx \sum_{k=1}^{K} \tilde{\pi}_G(U_j|\boldsymbol{Y},\theta^k)\tilde{\pi}_{LA}(\theta^k|\boldsymbol{Y})\Delta^k$.

Speed and Scalability (Stringer et al., 2020b)

The approximation methodology is **fast** and **scales** to (very) high dimensions.

The **Gaussian approximation** requires repeated high-dimensional optimizations. The objective function is **convex** and has a **sparse Hessian**, and we use **trust region** optimization (Braun, 2014) to do this **fast** and **stable**.

The **Laplace approximation** requires the modes already found for the Gaussian approximations. It requires the determinants of two large sparse matrices for each θ^k, which we compute using the **sparse Cholesky decomposition**.

We follow the **NIM** (**N**ever **I**nvert a **M**atrix) philosophy for fast and memory-efficient computation. For the marginal Gaussian approximations $\tilde{\pi}_G(U_j|\boldsymbol{Y}, \theta^k)$ we require the diagonal of the inverse of a large sparse matrix, which we also get using the sparse Cholesky decompostion (Rue and Martino, 2007).

The computations which need to be done for multiple θ^k are done in **parallel**, and we can (and definitely do) leverage massive cloud computing technology when necessary.

Speed and Scalability (Stringer et al., 2020b)

The approximation methodology is **fast** and **scales** to (very) high dimensions.

The **Gaussian approximation** requires repeated high-dimensional optimizations. The objective function is **convex** and has a **sparse Hessian**, and we use **trust region** optimization (Braun, 2014) to do this **fast** and **stable**.

The **Laplace approximation** requires the modes already found for the Gaussian approximations. It requires the determinants of two large sparse matrices for each θ^k, which we compute using the **sparse Cholesky decomposition**.

We follow the **NIM** (**N**ever **I**nvert a **M**atrix) philosophy for fast and memory-efficient computation. For the marginal Gaussian approximations $\tilde{\pi}_G(U_j|\boldsymbol{Y}, \theta^k)$ we require the diagonal of the inverse of a large sparse matrix, which we also get using the sparse Cholesky decompostion (Rue and Martino, 2007).

The computations which need to be done for multiple θ^k are done in **parallel**, and we can (and definitely do) leverage massive cloud computing technology when necessary.

Speed and Scalability (Stringer et al., 2020b)

The approximation methodology is **fast** and **scales** to (very) high dimensions.

The **Gaussian approximation** requires repeated high-dimensional optimizations. The objective function is **convex** and has a **sparse Hessian**, and we use **trust region** optimization (Braun, 2014) to do this **fast** and **stable**.

The **Laplace approximation** requires the modes already found for the Gaussian approximations. It requires the determinants of two large sparse matrices for each θ^k, which we compute using the **sparse Cholesky decomposition**.

We follow the **NIM** (**N**ever **I**nvert a **M**atrix) philosophy for fast and memory-efficient computation. For the marginal Gaussian approximations $\tilde{\pi}_G(U_j|\boldsymbol{Y}, \theta^k)$ we require the diagonal of the inverse of a large sparse matrix, which we also get using the sparse Cholesky decompostion (Rue and Martino, 2007).

The computations which need to be done for multiple θ^k are done in **parallel**, and we can (and definitely do) leverage massive cloud computing technology when necessary.

Speed and Scalability (Stringer et al., 2020b)

The approximation methodology is **fast** and **scales** to (very) high dimensions.

The **Gaussian approximation** requires repeated high-dimensional optimizations. The objective function is **convex** and has a **sparse Hessian**, and we use **trust region** optimization (Braun, 2014) to do this **fast** and **stable**.

The **Laplace approximation** requires the modes already found for the Gaussian approximations. It requires the determinants of two large sparse matrices for each θ^k, which we compute using the **sparse Cholesky decomposition**.

We follow the **NIM** (**N**ever **I**nvert a **M**atrix) philosophy for fast and memory-efficient computation. For the marginal Gaussian approximations $\tilde{\pi}_G(U_j|\boldsymbol{Y}, \theta^k)$ we require the diagonal of the inverse of a large sparse matrix, which we also get using the sparse Cholesky decompostion (Rue and Martino, 2007).

The computations which need to be done for multiple θ^k are done in **parallel**, and we can (and definitely do) leverage massive cloud computing technology when necessary.

Speed and Scalability (Stringer et al., 2020b)

The approximation methodology is **fast** and **scales** to (very) high dimensions.

The **Gaussian approximation** requires repeated high-dimensional optimizations. The objective function is **convex** and has a **sparse Hessian**, and we use **trust region** optimization (Braun, 2014) to do this **fast** and **stable**.

The **Laplace approximation** requires the modes already found for the Gaussian approximations. It requires the determinants of two large sparse matrices for each θ^k, which we compute using the **sparse Cholesky decomposition**.

We follow the **NIM** (**N**ever **I**nvert a **M**atrix) philosophy for fast and memory-efficient computation. For the marginal Gaussian approximations $\tilde{\pi}_G(U_j|\boldsymbol{Y}, \theta^k)$ we require the diagonal of the inverse of a large sparse matrix, which we also get using the sparse Cholesky decompostion (Rue and Martino, 2007).

The computations which need to be done for multiple θ^k are done in **parallel**, and we can (and definitely do) leverage massive cloud computing technology when necessary.

Speed and Scalability (Stringer et al., 2020b)

The approximation methodology is **fast** and **scales** to (very) high dimensions.

The **Gaussian approximation** requires repeated high-dimensional optimizations. The objective function is **convex** and has a **sparse Hessian**, and we use **trust region** optimization (Braun, 2014) to do this **fast** and **stable**.

The **Laplace approximation** requires the modes already found for the Gaussian approximations. It requires the determinants of two large sparse matrices for each θ^k, which we compute using the **sparse Cholesky decomposition**.

We follow the **NIM** (**N**ever **I**nvert a **M**atrix) philosophy for fast and memory-efficient computation. For the marginal Gaussian approximations $\tilde{\pi}_G(U_j|\boldsymbol{Y}, \theta^k)$ we require the diagonal of the inverse of a large sparse matrix, which we also get using the sparse Cholesky decompostion (Rue and Martino, 2007).

The computations which need to be done for multiple θ^k are done in **parallel**, and we can (and definitely do) leverage massive cloud computing technology when necessary.

Results

Results

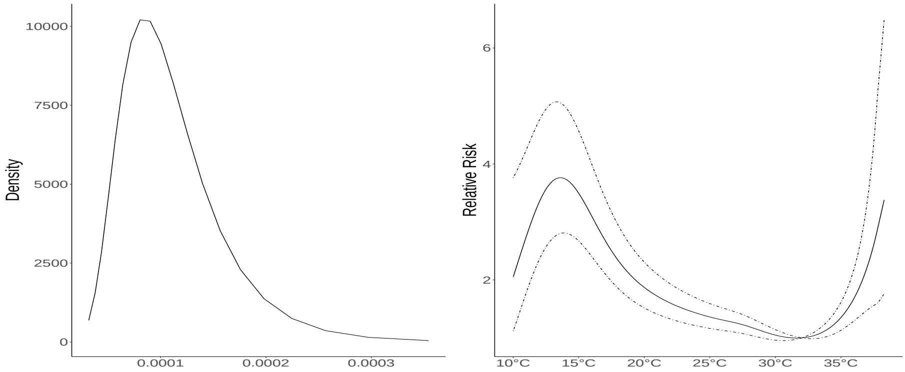

Figure 3: Approximation to the posterior distribution of the smoothing standard deviation (σ_u, left) and posterior median and 95% credible intervals for the temperature effect for the Indian mortality data.

Summary and Extensions

Summary

Introduced an approximate Bayesian inference methodology for case crossover models.

Inferred relative risk of death due to stroke as a function of temperature using data from the Indian Million Death Study.

Bayesian inference provided **model-based estimation and uncertainty quantification** for the smoothness of the temperature effect.

Method is **not sensitive** to the number and placement of bins (spline knots).

Computations are **fast and scalable** due to the use of sparse matrix algebra and stable optimization algorithms.

Summary

Introduced an approximate Bayesian inference methodology for case crossover models.

Inferred relative risk of death due to stroke as a function of temperature using data from the Indian Million Death Study.

Bayesian inference provided **model-based estimation and uncertainty quantification** for the smoothness of the temperature effect.

Method is **not sensitive** to the number and placement of bins (spline knots).

Computations are **fast and scalable** due to the use of sparse matrix algebra and stable optimization algorithms.

Summary

Introduced an approximate Bayesian inference methodology for case crossover models.

Inferred relative risk of death due to stroke as a function of temperature using data from the Indian Million Death Study.

Bayesian inference provided **model-based estimation and uncertainty quantification** for the smoothness of the temperature effect.

Method is **not sensitive** to the number and placement of bins (spline knots).

Computations are **fast and scalable** due to the use of sparse matrix algebra and stable optimization algorithms.

Summary

Introduced an approximate Bayesian inference methodology for case crossover models.

Inferred relative risk of death due to stroke as a function of temperature using data from the Indian Million Death Study.

Bayesian inference provided **model-based estimation and uncertainty quantification** for the smoothness of the temperature effect.

Method is **not sensitive** to the number and placement of bins (spline knots).

Computations are **fast and scalable** due to the use of sparse matrix algebra and stable optimization algorithms.

Summary

Introduced an approximate Bayesian inference methodology for case crossover models.

Inferred relative risk of death due to stroke as a function of temperature using data from the Indian Million Death Study.

Bayesian inference provided **model-based estimation and uncertainty quantification** for the smoothness of the temperature effect.

Method is **not sensitive** to the number and placement of bins (spline knots).

Computations are **fast and scalable** due to the use of sparse matrix algebra and stable optimization algorithms.

Summary

Introduced an approximate Bayesian inference methodology for case crossover models.

Inferred relative risk of death due to stroke as a function of temperature using data from the Indian Million Death Study.

Bayesian inference provided **model-based estimation and uncertainty quantification** for the smoothness of the temperature effect.

Method is **not sensitive** to the number and placement of bins (spline knots).

Computations are **fast and scalable** due to the use of sparse matrix algebra and stable optimization algorithms.

Extensions

The smoothing models and inference methodology were developed in separate projects and each could have its own talk.

They are widely applicable. I have the following examples ready:

1. A Bernoulli GLMM fit to nearly 8 million discharge records from drug treatment centres in the USA, with observations correlated within towns within states,
2. A log-Gaussian Cox process fit to spatially-aggregated mortality counts in England and Wales,
3. Semi-parametric Cox proportional hazards regression fit to a classic dataset on Leukaemia survival times,
4. Simulated semi-parametric Poisson regression where I use stochastic optimization to fit the model to dataset of size up to $n = 1,000,000,000$...

Extensions

The smoothing models and inference methodology were developed in separate projects and each could have its own talk.

They are widely applicable. I have the following examples ready:

1. A Bernoulli GLMM fit to nearly 8 million discharge records from drug treatment centres in the USA, with observations correlated within towns within states,
2. A log-Gaussian Cox process fit to spatially-aggregated mortality counts in England and Wales,
3. Semi-parametric Cox proportional hazards regression fit to a classic dataset on Leukaemia survival times,
4. Simulated semi-parametric Poisson regression where I use stochastic optimization to fit the model to dataset of size up to $n = 1,000,000,000$...

Extensions

The smoothing models and inference methodology were developed in separate projects and each could have its own talk.

They are widely applicable. I have the following examples ready:

1. A Bernoulli GLMM fit to nearly 8 million discharge records from drug treatment centres in the USA, with observations correlated within towns within states,
2. A log-Gaussian Cox process fit to spatially-aggregated mortality counts in England and Wales,
3. Semi-parametric Cox proportional hazards regression fit to a classic dataset on Leukaemia survival times,
4. Simulated semi-parametric Poisson regression where I use stochastic optimization to fit the model to dataset of size up to $n = 1,000,000,000$...

Extensions

The smoothing models and inference methodology were developed in separate projects and each could have its own talk.

They are widely applicable. I have the following examples ready:

1. A Bernoulli GLMM fit to nearly 8 million discharge records from drug treatment centres in the USA, with observations correlated within towns within states,
2. A log-Gaussian Cox process fit to spatially-aggregated mortality counts in England and Wales,
3. Semi-parametric Cox proportional hazards regression fit to a classic dataset on Leukaemia survival times,
4. Simulated semi-parametric Poisson regression where I use stochastic optimization to fit the model to dataset of size up to $n = 1,000,000,000$...

Extensions

The smoothing models and inference methodology were developed in separate projects and each could have its own talk.

They are widely applicable. I have the following examples ready:

1. A Bernoulli GLMM fit to nearly 8 million discharge records from drug treatment centres in the USA, with observations correlated within towns within states,
2. A log-Gaussian Cox process fit to spatially-aggregated mortality counts in England and Wales,
3. Semi-parametric Cox proportional hazards regression fit to a classic dataset on Leukaemia survival times,
4. Simulated semi-parametric Poisson regression where I use stochastic optimization to fit the model to dataset of size up to $n = 1,000,000,000$...

Extensions

The smoothing models and inference methodology were developed in separate projects and each could have its own talk.

They are widely applicable. I have the following examples ready:

1. A Bernoulli GLMM fit to nearly 8 million discharge records from drug treatment centres in the USA, with observations correlated within towns within states,
2. A log-Gaussian Cox process fit to spatially-aggregated mortality counts in England and Wales,
3. Semi-parametric Cox proportional hazards regression fit to a classic dataset on Leukaemia survival times,
4. Simulated semi-parametric Poisson regression where I use stochastic optimization to fit the model to dataset of size up to $n = 1,000,000,000$...

Extensions

The smoothing models and inference methodology were developed in separate projects and each could have its own talk.

They are widely applicable. I have the following examples ready:

1. A Bernoulli GLMM fit to nearly 8 million discharge records from drug treatment centres in the USA, with observations correlated within towns within states,
2. A log-Gaussian Cox process fit to spatially-aggregated mortality counts in England and Wales,
3. Semi-parametric Cox proportional hazards regression fit to a classic dataset on Leukaemia survival times,
4. Simulated semi-parametric Poisson regression where I use stochastic optimization to fit the model to dataset of size up to $n = 1,000,000,000$...

Extensions

...and have the following on deck:

1. A semi-parametric Age-Period-Cohort model for American opioid mortality data, using two-dimensional random walks for age/year interaction effects,
2. A model for estiamating the mass of the Milky Way using astronomical data which is subject to measurement error,
3. Two further examples of the need for stochastic optimization: a Cox-PH model fit to the opioid treatment centre data and a semi-parametric regression fit to astronomical measurements from 1.7 Billion stars,
4. Bayesian neural networks (?)

Extensions

…and have the following on deck:

1. A semi-parametric Age-Period-Cohort model for American opioid mortality data, using two-dimensional random walks for age/year interaction effects,
2. A model for estiamating the mass of the Milky Way using astronomical data which is subject to measurement error,
3. Two further examples of the need for stochastic optimization: a Cox-PH model fit to the opioid treatment centre data and a semi-parametric regression fit to astronomical measurements from 1.7 Billion stars,
4. Bayesian neural networks (?)

Extensions

...and have the following on deck:

1. A semi-parametric Age-Period-Cohort model for American opioid mortality data, using two-dimensional random walks for age/year interaction effects,
2. A model for estiamating the mass of the Milky Way using astronomical data which is subject to measurement error,
3. Two further examples of the need for stochastic optimization: a Cox-PH model fit to the opioid treatment centre data and a semi-parametric regression fit to astronomical measurements from 1.7 Billion stars,
4. Bayesian neural networks (?)

Extensions

...and have the following on deck:

1. A semi-parametric Age-Period-Cohort model for American opioid mortality data, using two-dimensional random walks for age/year interaction effects,
2. A model for estiamating the mass of the Milky Way using astronomical data which is subject to measurement error,
3. Two further examples of the need for stochastic optimization: a Cox-PH model fit to the opioid treatment centre data and a semi-parametric regression fit to astronomical measurements from 1.7 Billion stars,
4. Bayesian neural networks (?)

Extensions

...and have the following on deck:

1. A semi-parametric Age-Period-Cohort model for American opioid mortality data, using two-dimensional random walks for age/year interaction effects,
2. A model for estiamating the mass of the Milky Way using astronomical data which is subject to measurement error,
3. Two further examples of the need for stochastic optimization: a Cox-PH model fit to the opioid treatment centre data and a semi-parametric regression fit to astronomical measurements from 1.7 Billion stars,
4. Bayesian neural networks (?)

References

Braun, M. (2014). trustOptim: An R package for trust region optimization with sparse hessians. *Journal of Statistical Software 60*(4), 1–16.

Fu, S. H., A. Gasparrini, P. S. Rodriguez, and P. Jha (2018). Mortality attributable to hot and cold ambient temperatures in india: a nationally representative case-crossover study. *PLoS Med 15*(7).

Rue, H. and S. Martino (2007). Approximate bayesian inference for hierarchical gaussian markov random field models. *Journal of Statistical Planning and Inference 137*, 3177 – 3192.

Stringer, A. and P. Brown (2020). Proper random walk priors for bayesian smoothing. *In preparation*.

Stringer, A., P. Brown, and J. Stafford (2020a). Approximate bayesian inference for case crossover models. *Under revision for Biometrics*.

Stringer, A., P. Brown, and J. Stafford (2020b). Fast, scalable approximations to marginal posterior distributions in latent gaussian models. *In preparation*.

Tierney, L. and J. B. Kadane (1986). Accurate approximations to posterior moments and marginal densities. *Journal of the American Statistical*

Questions?