Computational Complexity of the Fisher Information

Ali Eshragh

(Joint work with Nigel Bean and Joshua Ross)

School of Mathematical and Physical Sciences & CARMA
The University of Newcastle, Australia

INRIA, Paris
October 6, 2014

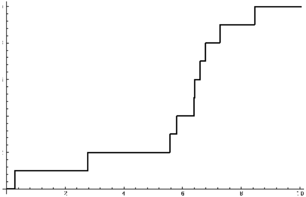
Motivation

- **Epidemiology**

Motivation

- **Epidemiology**

- A **Growing** Population



Definition and Notation

- Let X_t denote the **population size** at time t.

Definition and Notation

- Let X_t denote the **population size** at time t.
- $\{X_t : t \in \mathrm{R}_0^+\}$ is a **stochastic process** .

Definition and Notation

- Let X_t denote the **population size** at time t.
- $\{X_t : t \in \mathrm{R}_0^+\}$ is a **stochastic process** .
- Suppose $\{X_t : t \in \mathrm{R}_0^+\}$ is a **simple birth process (SBP)** with the **birth rate** λ. Moreover, $X_0 \overset{a.s.}{=} x_0$.

Definition and Notation

- Let X_t denote the **population size** at time t.
- $\{X_t : t \in \mathrm{R}_0^+\}$ is a **stochastic process** .
- Suppose $\{X_t : t \in \mathrm{R}_0^+\}$ is a **simple birth process (SBP)** with the **birth rate** λ. Moreover, $X_0 \overset{a.s.}{=} x_0$.
- It is **Markovian**, that is

$$\Pr(X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n, \ldots, , X_{t_1} = x_1) = \Pr(X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n) ,$$

for all possible values of n and $t_1, \ldots, t_{n+1}$.

Definition and Notation

- Let X_t denote the **population size** at time t.
- $\{X_t : t \in \mathrm{R}_0^+\}$ is a **stochastic process** .
- Suppose $\{X_t : t \in \mathrm{R}_0^+\}$ is a **simple birth process (SBP)** with the **birth rate** λ. Moreover, $X_0 \overset{a.s.}{=} x_0$.
- It is **Markovian**, that is

  $$\Pr(X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n, \ldots, , X_{t_1} = x_1) \;=\; \Pr(X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n)\,,$$

  for all possible values of n and $t_1, \ldots, t_{n+1}$.
- The **transition probability** is equal to

  $$\Pr(X_{s+t} = j | X_s = i) \;=\; \binom{j-1}{i-1} e^{-\lambda t i}(1 - e^{-\lambda t})^{j-i}\,.$$

Likelihood Function

- **Estimating** the unknown parameter λ through **maximum likelihood** method.

Likelihood Function

- **Estimating** the unknown parameter λ through **maximum likelihood** method.
- Take the **observations** $X_{t_1}, \ldots, X_{t_n}$ at observation times $0 < t_1 \leq \ldots \leq t_n \leq \tau$, respectively.

Likelihood Function

- **Estimating** the unknown parameter λ through **maximum likelihood** method.
- Take the **observations** $X_{t_1}, \ldots, X_{t_n}$ at observation times $0 < t_1 \leq \ldots \leq t_n \leq \tau$, respectively.
- Construct the **likelihood function**

$$\mathcal{L}(x_1, \ldots, x_n; \lambda) = \Pr(X_{t_1} = x_1, \ldots, X_{t_n} = x_n | \lambda)$$

Likelihood Function

- **Estimating** the unknown parameter λ through **maximum likelihood** method.
- Take the **observations** $X_{t_1}, \ldots, X_{t_n}$ at observation times $0 < t_1 \leq \ldots \leq t_n \leq \tau$, respectively.
- Construct the **likelihood function**

$$\mathcal{L}(x_1, \ldots, x_n; \lambda) = \Pr(X_{t_1} = x_1, \ldots, X_{t_n} = x_n | \lambda)$$

$$= \prod_{i=2}^{n} \Pr(X_{t_i} = x_i | X_{t_{i-1}} = x_{i-1}, \ldots, X_{t_1} = x_1) \Pr(X_{t_1} = x_1)$$

Likelihood Function

- **Estimating** the unknown parameter λ through **maximum likelihood** method.
- Take the **observations** $X_{t_1}, \ldots, X_{t_n}$ at observation times $0 < t_1 \leq \ldots \leq t_n \leq \tau$, respectively.
- Construct the **likelihood function**

$$\begin{aligned}\mathcal{L}(x_1, \ldots, x_n; \lambda) &= \Pr(X_{t_1} = x_1, \ldots, X_{t_n} = x_n | \lambda) \\ &= \prod_{i=2}^{n} \Pr(X_{t_i} = x_i | X_{t_{i-1}} = x_{i-1}, \ldots, X_{t_1} = x_1) \Pr(X_{t_1} = x_1) \\ &= \prod_{i=2}^{n} \Pr(X_{t_i} = x_i | X_{t_{i-1}} = x_{i-1}) \Pr(X_{t_1} = x_1)\end{aligned}$$

Likelihood Function

- **Estimating** the unknown parameter λ through **maximum likelihood** method.
- Take the **observations** $X_{t_1}, \ldots, X_{t_n}$ at observation times $0 < t_1 \leq \ldots \leq t_n \leq \tau$, respectively.
- Construct the **likelihood function**

$$\begin{aligned}\mathcal{L}(x_1, \ldots, x_n; \lambda) &= \Pr(X_{t_1} = x_1, \ldots, X_{t_n} = x_n | \lambda) \\ &= \prod_{i=2}^{n} \Pr(X_{t_i} = x_i | X_{t_{i-1}} = x_{i-1}, \ldots, X_{t_1} = x_1) \Pr(X_{t_1} = x_1) \\ &= \prod_{i=2}^{n} \Pr(X_{t_i} = x_i | X_{t_{i-1}} = x_{i-1}) \Pr(X_{t_1} = x_1) \\ &= \prod_{i=1}^{n} \binom{x_i - 1}{x_{i-1} - 1} e^{-\lambda(t_i - t_{i-1}) x_{i-1}} (1 - e^{-\lambda(t_i - t_{i-1})})^{x_i - x_{i-1}} .\end{aligned}$$

Observation Times

- **When** should we take the observations $X_{t_1}, \ldots, X_{t_n}$?

Observation Times

- **When** should we take the observations $X_{t_1}, \ldots, X_{t_n}$?
- Presumably, a good choice is finding observation times $t_1, \ldots, t_n$ such that the **expected volume of information** obtained from these observations to estimate the unknown parameter λ is **maximized**.

Observation Times

- **When** should we take the observations $X_{t_1}, \dots, X_{t_n}$?
- Presumably, a good choice is finding observation times $t_1, \dots, t_n$ such that the **expected volume of information** obtained from these observations to estimate the unknown parameter λ is **maximized**.
- A good tool to measure the expected volume of information gained from a set of observations is the **Fisher Information**.

Observation Times

- **When** should we take the observations $X_{t_1}, \ldots, X_{t_n}$?
- Presumably, a good choice is finding observation times $t_1, \ldots, t_n$ such that the **expected volume of information** obtained from these observations to estimate the unknown parameter λ is **maximized**.
- A good tool to measure the expected volume of information gained from a set of observations is the **Fisher Information**.
- It can be shown that

$$\mathcal{FI}_{(X_{t_1},\ldots,X_{t_n})}(\lambda) \quad = \quad E_{\mathcal{L}}\left[\left(\frac{d}{d\lambda}\ln(\mathcal{L}(X_{t_1},\ldots,X_{t_n};\lambda))\right)^2\right].$$

Observation Times

- **When** should we take the observations $X_{t_1}, \ldots, X_{t_n}$?
- Presumably, a good choice is finding observation times $t_1, \ldots, t_n$ such that the **expected volume of information** obtained from these observations to estimate the unknown parameter λ is **maximized**.
- A good tool to measure the expected volume of information gained from a set of observations is the **Fisher Information**.
- It can be shown that

$$\mathcal{FI}_{(X_{t_1},\ldots,X_{t_n})}(\lambda) = E_{\mathcal{L}}\left[\left(\frac{d}{d\lambda}\ln(\mathcal{L}(X_{t_1},\ldots,X_{t_n};\lambda))\right)^2\right].$$

- Hence, $(t_1^*, \ldots, t_n^*) \in \operatorname{argmax}\{\mathcal{FI}_{(X_{t_1},\ldots,X_{t_n})}(\lambda)\}$.

Fisher Information and Optimal Observation Times

Proposition (Becker and Kersting, 1983)

The **Fisher information** *for a SBP with the parameter* λ, *the initial value of* x_0 *and the observation times of* $(t_1, \ldots, t_n)$ *is as follows:*

$$\mathcal{FI}_{(x_{t_1},\cdots,x_{t_n})}(\lambda) = x_0 \sum_{i=1}^{n} \frac{(t_i - t_{i-1})^2}{e^{-\lambda t_{i-1}} - e^{-\lambda t_i}} .$$

Fisher Information and Optimal Observation Times

Proposition (Becker and Kersting, 1983)

The **Fisher information** *for a SBP with the parameter* λ, *the initial value of* x_0 *and the observation times of* $(t_1, \ldots, t_n)$ *is as follows:*

$$\mathcal{FI}_{(X_{t_1},\cdots,X_{t_n})}(\lambda) \quad = \quad x_0 \sum_{i=1}^{n} \frac{(t_i - t_{i-1})^2}{e^{-\lambda t_{i-1}} - e^{-\lambda t_i}} \,.$$

Optimal Observation Times (Becker and Kersting, 1983)

$$\mathbf{t_i^* \approx \frac{3}{\lambda} \log\left(1 + \frac{i}{n}(e^{\frac{\lambda\tau}{3}} - 1)\right) \quad for\ i = 1, \ldots, n}$$

Definition and Notation

- Suppose that at each observation time, we can count the population, **partially**.

Definition and Notation

- Suppose that at each observation time, we can count the population, **partially**.
- At each observation time, each individual can be counted **independently** with probability **p**.

Definition and Notation

- Suppose that at each observation time, we can count the population, **partially**.
- At each observation time, each individual can be counted **independently** with probability **p**.
- $\mathbf{Y_t}$ is the number of individuals observed at at time t.

Definition and Notation

- Suppose that at each observation time, we can count the population, **partially**.
- At each observation time, each individual can be counted **independently** with probability $\mathbf{p}$.
- $\mathbf{Y_t}$ is the number of individuals observed at at time t.
- $(Y_t|X_t = x) \sim \texttt{Binomial}(\mathbf{x}, \mathbf{p})$.

Definition and Notation

- Suppose that at each observation time, we can count the population, **partially**.
- At each observation time, each individual can be counted **independently** with probability **p**.
- $\mathbf{Y_t}$ is the number of individuals observed at at time t.
- $(Y_t|X_t = x) \sim \mathtt{Binomial}(\mathbf{x}, \mathbf{p})$.
- We call the stochastic process $\{Y_t : t \in \mathrm{R}_0^+\}$ the **partially-observable simple birth process (POSBP)** with parameters (λ, p).

Definition and Notation

- Suppose that at each observation time, we can count the population, **partially**.
- At each observation time, each individual can be counted **independently** with probability **p**.
- $\mathbf{Y_t}$ is the number of individuals observed at at time t.
- $(Y_t|X_t = x) \sim \texttt{Binomial}(\mathbf{x}, \mathbf{p})$.
- We call the stochastic process $\{Y_t : t \in \mathrm{R}_0^+\}$ the **partially-observable simple birth process (POSBP)** with parameters (λ, p).
- $\text{POSBP}(\lambda, 1) \equiv \text{SBP}(\lambda)$.

Markovian or non-Markovian?

Theorem (Bean, Elliott, Eshragh and Ross; 2014)

The POSBP $\{Y_t : t \in \mathrm{R}_0^+\}$ *with parameters* (λ, p) *is* **not Markovian**.

Markovian or non-Markovian?

Theorem (Bean, Elliott, Eshragh and Ross; 2014)

The POSBP $\{Y_t : t \in \mathbb{R}_0^+\}$ with parameters (λ, p) is **not Markovian**.

- However,

$$\Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n} | X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n})$$

$$= \prod_{i=1}^{n} \Pr(Y_{t_i} = y_{t_i} | X_{t_i} = x_{t_i}) .$$

Likelihood Function

- The likelihood function:

$$\mathcal{L}(y_{t_1}, \ldots, y_{t_n}; \lambda) = \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n})$$

Likelihood Function

- The likelihood function:

$$
\begin{aligned}
\mathcal{L}(y_{t_1}, \ldots, y_{t_n}; \lambda) &= \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n}) \\
= \sum_{x_{t_1}, \ldots, x_{t_n}} & \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n} | X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
& \Pr(X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n})
\end{aligned}
$$

Likelihood Function

- The likelihood function:

$$
\begin{aligned}
\mathcal{L}(y_{t_1}, \ldots, y_{t_n}; \lambda) &= \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n} | X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
&\qquad \Pr(X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \Pr(Y_{t_i} = y_{t_i} | X_{t_i} = x_{t_i}) \Pr(X_{t_i} = x_{t_i} | X_{t_{i-1}} = x_{t_{i-1}})
\end{aligned}
$$

Likelihood Function

- The likelihood function:

$$
\begin{aligned}
\mathcal{L}(y_{t_1}, \ldots, y_{t_n}; \lambda) &= \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n} | X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
&\qquad \Pr(X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \Pr(Y_{t_i} = y_{t_i} | X_{t_i} = x_{t_i}) \Pr(X_{t_i} = x_{t_i} | X_{t_{i-1}} = x_{t_{i-1}}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n}
\end{aligned}
$$

Likelihood Function

- The likelihood function:

$$
\begin{aligned}
\mathcal{L}(y_{t_1}, \ldots, y_{t_n}; \lambda) &= \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n} | X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
&\qquad \Pr(X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \Pr(Y_{t_i} = y_{t_i} | X_{t_i} = x_{t_i}) \Pr(X_{t_i} = x_{t_i} | X_{t_{i-1}} = x_{t_{i-1}}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \binom{x_{t_i}}{y_{t_i}} p^{y_i} q^{x_{t_i} - y_{t_i}}
\end{aligned}
$$

Likelihood Function

- The likelihood function:

$$
\begin{aligned}
\mathcal{L}(y_{t_1},\ldots,y_{t_n};\lambda) &= \Pr(Y_{t_1}=y_{t_1},\ldots,Y_{t_n}=y_{t_n}) \\
&= \sum_{x_{t_1},\ldots,x_{t_n}} \Pr(Y_{t_1}=y_{t_1},\ldots,Y_{t_n}=y_{t_n}|X_{t_1}=x_{t_1},\ldots,X_{t_n}=x_{t_n}) \\
&\qquad \Pr(X_{t_1}=x_{t_1},\ldots,X_{t_n}=x_{t_n}) \\
&= \sum_{x_{t_1},\ldots,x_{t_n}} \prod_{i=1}^{n} \Pr(Y_{t_i}=y_{t_i}|X_{t_i}=x_{t_i}) \Pr(X_{t_i}=x_{t_i}|X_{t_{i-1}}=x_{t_{i-1}}) \\
&= \sum_{x_{t_1},\ldots,x_{t_n}} \prod_{i=1}^{n} \binom{x_{t_i}}{y_{t_i}} p^{y_i} q^{x_{t_i}-y_{t_i}} \binom{x_{t_i}-1}{x_{t_{i-1}}-1} \upsilon_{i-1,i}^{x_{t_{i-1}}} (1-\upsilon_{i-1,i})^{x_{t_i}-x_{t_{i-1}}},
\end{aligned}
$$

Likelihood Function

- The likelihood function:

$$\begin{aligned}
\mathcal{L}(y_{t_1}, \ldots, y_{t_n}; \lambda) &= \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n} | X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
&\qquad \Pr(X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \Pr(Y_{t_i} = y_{t_i} | X_{t_i} = x_{t_i}) \Pr(X_{t_i} = x_{t_i} | X_{t_{i-1}} = x_{t_{i-1}}) \\
&= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \binom{x_{t_i}}{y_{t_i}} p^{y_i} q^{x_{t_i} - y_{t_i}} \binom{x_{t_i} - 1}{x_{t_{i-1}} - 1} \upsilon_{i-1,i}^{x_{t_{i-1}}} (1 - \upsilon_{i-1,i})^{x_{t_i} - x_{t_{i-1}}},
\end{aligned}$$

where $q := 1 - p$ and $\upsilon_{i-1,i} := e^{-\lambda(t_i - t_{i-1})}$.

Fisher Information

- The Fisher Information:

$$\mathcal{FI}_{(Y_{t_1},\ldots,Y_{t_n})}(\lambda) = E_{\mathcal{L}}\left[\left(\frac{d\log(\mathcal{L})}{d\lambda}\right)^2\right]$$

Fisher Information

- The Fisher Information:

$$\begin{aligned}\mathcal{FI}_{(Y_{t_1},\ldots,Y_{t_n})}(\lambda) &= E_{\mathcal{L}}\left[\left(\frac{d\log(\mathcal{L})}{d\lambda}\right)^2\right] \\ &= \sum_{y_{t_1},\ldots,y_{t_n}}\left(\frac{d\log(\mathcal{L})}{d\lambda}\right)^2\mathcal{L}\end{aligned}$$

Fisher Information

- The Fisher Information:

$$
\begin{aligned}
\mathcal{FI}_{(Y_{t_1},\ldots,Y_{t_n})}(\lambda) &= E_{\mathcal{L}}\left[\left(\frac{d\log(\mathcal{L})}{d\lambda}\right)^2\right] \\
&= \sum_{y_{t_1},\ldots,y_{t_n}} \left(\frac{d\log(\mathcal{L})}{d\lambda}\right)^2 \mathcal{L} \\
&= \sum_{y_{t_1},\ldots,y_{t_n}} \left(\frac{\frac{d\mathcal{L}}{d\lambda}}{\mathcal{L}}\right)^2 \mathcal{L}
\end{aligned}
$$

Fisher Information

- The Fisher Information:

$$\begin{aligned}
\mathcal{FI}_{(Y_{t_1},\ldots,Y_{t_n})}(\lambda) &= E_{\mathcal{L}}\left[\left(\frac{d\log(\mathcal{L})}{d\lambda}\right)^2\right] \\
&= \sum_{y_{t_1},\ldots,y_{t_n}} \left(\frac{d\log(\mathcal{L})}{d\lambda}\right)^2 \mathcal{L} \\
&= \sum_{y_{t_1},\ldots,y_{t_n}} \left(\frac{\frac{d\mathcal{L}}{d\lambda}}{\mathcal{L}}\right)^2 \mathcal{L} \\
&= \sum_{y_{t_1},\ldots,y_{t_n}} \frac{\left(\frac{d\mathcal{L}(y_{t_1},\ldots,y_{t_n};\lambda)}{d\lambda}\right)^2}{\mathcal{L}(y_{t_1},\ldots,y_{t_n};\lambda)}.
\end{aligned}$$

Theoretical Result

Proposition (Bean, Eshragh and Ross; 2014)

For a POSBP with n observations and time horizon τ, the FI is an **increasing** *function of t_n. Hence, the* **optimal observation time** *for the last observation, that is t_n^*, is equal to $\boldsymbol{\tau}$.*

Theoretical Result

Proposition (Bean, Eshragh and Ross; 2014)

For a POSBP with n observations and time horizon τ, the FI is an **increasing** *function of t_n. Hence, the* **optimal observation time** *for the last observation, that is t_n^*, is equal to $\boldsymbol{\tau}$.*

Proposition (Bean, Eshragh and Ross; 2014)

If $t_1^*, \ldots, t_n^*$ are optimal observation times for a POSBP with parameters (λ, p) and time-horizon τ, then $\frac{\mathbf{t_1^*}}{\boldsymbol{\tau}}, \ldots, \frac{\mathbf{t_n^*}}{\boldsymbol{\tau}}$ are **optimal observation times** *for a POSBP with parameters $(\boldsymbol{\lambda\tau}, p)$ and time-horizon $\mathbf{1}$.*

Truncated Summation

- The Fisher Information:

$$\mathcal{FI}_{(Y_{t_1},\ldots,Y_{t_n})}(\lambda) \quad = \quad \sum_{y_{t_1},\ldots,y_{t_n}} \frac{\left(\frac{d\mathcal{L}(y_{t_1},\ldots,y_{t_n};\,\lambda)}{d\lambda}\right)^2}{\mathcal{L}(y_{t_1},\ldots,y_{t_n};\,\lambda)}.$$

Truncated Summation

- The Fisher Information:

$$\mathcal{FI}_{(Y_{t_1},\ldots,Y_{t_n})}(\lambda) \quad = \quad \sum_{y_{t_1},\ldots,y_{t_n}} \frac{\left(\frac{d\mathcal{L}(y_{t_1},\ldots,y_{t_n};\,\lambda)}{d\lambda}\right)^2}{\mathcal{L}(y_{t_1},\ldots,y_{t_n};\,\lambda)}\,.$$

- Here, the likelihood function $\mathcal{L}(y_{t_1},\ldots,y_{t_n};\,\lambda)$ is equal to

$$\sum_{x_{t_1},\ldots,x_{t_n}} \prod_{i=1}^{n} \binom{x_{t_i}}{y_{t_i}} p^{y_i}(1-p)^{x_{t_i}-y_{t_i}} \binom{x_{t_i}-1}{x_{t_{i-1}}-1} \upsilon_{i-1,i}^{x_{t_{i-1}}} (1-\upsilon_{i-1,i})^{x_{t_i}-x_{t_{i-1}}}\,,$$

where $\upsilon_{i-1,i} := e^{-\lambda(t_i-t_{i-1})}$.

Truncated Summation

- The Fisher Information:

$$\mathcal{FI}_{(Y_{t_1},\ldots,Y_{t_n})}(\lambda) = \sum_{y_{t_1},\ldots,y_{t_n}} \frac{\left(\frac{d\mathcal{L}(y_{t_1},\ldots,y_{t_n};\lambda)}{d\lambda}\right)^2}{\mathcal{L}(y_{t_1},\ldots,y_{t_n};\lambda)}.$$

- Here, the likelihood function $\mathcal{L}(y_{t_1},\ldots,y_{t_n};\lambda)$ is equal to

$$\sum_{x_{t_1},\ldots,x_{t_n}} \prod_{i=1}^{n} \binom{x_{t_i}}{y_{t_i}} p^{y_i}(1-p)^{x_{t_i}-y_{t_i}} \binom{x_{t_i}-1}{x_{t_{i-1}}-1} \upsilon_{i-1,i}^{x_{t_{i-1}}} (1-\upsilon_{i-1,i})^{x_{t_i}-x_{t_{i-1}}},$$

where $\upsilon_{i-1,i} := e^{-\lambda(t_i-t_{i-1})}$.

- By exploiting **Chebyshev's inequality**, we have

$$\begin{aligned}\Pr\left(E[Z]-12\sqrt{Var(Z)} \le Z \le E[Z]+12\sqrt{Var(Z)}\right) &\ge 1-\frac{1}{12^2} \\ &= 99.3\%.\end{aligned}$$

Conditional Expectations

- Motivating from Chebyshev's inequality:

$$0 \;\leq\; y_{t_i} \;\leq\; E[Y_{t_i}] + 12\sqrt{Var(Y_{t_i})}$$

$$\max\{1, y_{t_1}, \ldots, y_{t_n}\} \;\leq\; x_{t_n} \;\leq\; E[X_{t_n}|Y_{t_n} = y_{t_n}] + 12\sqrt{Var(X_{t_n}|Y_{t_n} = y_{t_n})}$$

Conditional Expectations

- Motivating from Chebyshev's inequality:

$$0 \leq y_{t_i} \leq E[Y_{t_i}] + 12\sqrt{Var(Y_{t_i})}$$

$$\max\{1, y_{t_1}, \ldots, y_{t_n}\} \leq x_{t_n} \leq E[X_{t_n} | Y_{t_n} = y_{t_n}] + 12\sqrt{Var(X_{t_n} | Y_{t_n} = y_{t_n})}$$

Lemma (Eshragh, Bean and Ross; 2014)

If $\{X_t\}$ is a **SBP** *with parameter λ and $\{Y_t\}$ is the corresponding* **POSBP** *with parameters (λ, p), then we have*

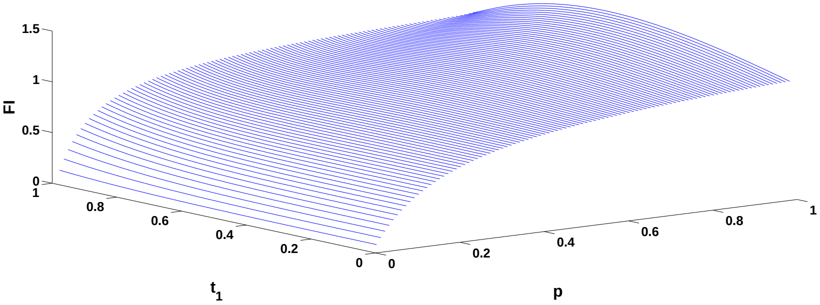
$$E[Y_t] = pe^{\lambda t}, \quad Var(Y_t) = p(pe^{2\lambda t} + (1-2p)e^{\lambda t})$$

$$E[X_t | Y_t = y_t] = \frac{y_t e^{\lambda t} + (1-p)(e^{\lambda t} - 1)}{pe^{\lambda t} + 1 - p}$$

$$Var(X_t | Y_t = y_t) = \frac{(y_t + 1)(1-p)e^{\lambda t}(e^{\lambda t} - 1)}{(pe^{\lambda t} + 1 - p)^2}.$$

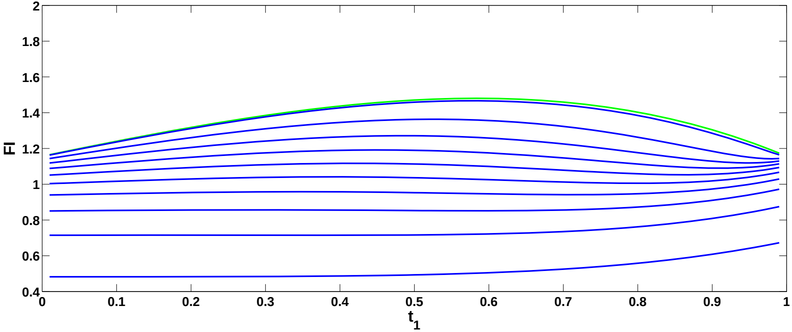
Results for $\lambda = 2$, $n = 2$ and $t_2^* = \tau = 1$

- Fisher Information vs. t_1 and p



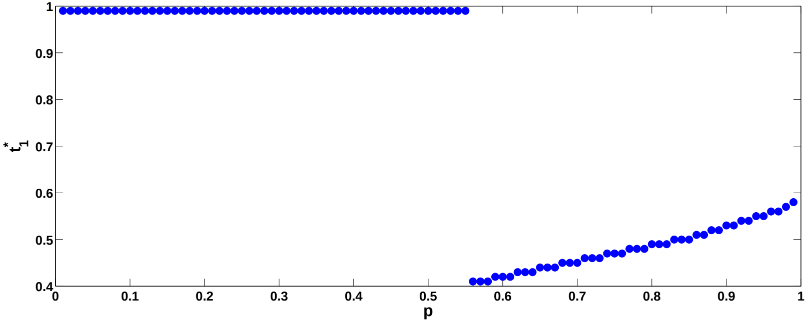
Results for $\lambda = 2$, $n = 2$ and $t_2^* = \tau = 1$

- The Fisher Information vs. t_1



Results for $\lambda = 2$, $n = 2$ and $t_2^* = \tau = 1$

- Optimal observation time t_1^* vs. p



The Chain Rule

- The likelihood function

$$\mathcal{L}(y_{t_1}, y_{t_2}|\lambda) = \Pr(Y_{t_2} = y_{t_2} | Y_{t_1} = y_{t_1}, \lambda) \Pr(Y_{t_1} = y_{t_1}|\lambda) \,.$$

The Chain Rule

- The likelihood function

$$\mathcal{L}(y_{t_1}, y_{t_2}|\lambda) = \Pr(Y_{t_2} = y_{t_2}|Y_{t_1} = y_{t_1}, \lambda)\Pr(Y_{t_1} = y_{t_1}|\lambda).$$

- Accordingly,

$$\log\left(\mathcal{L}(y_{t_1}, y_{t_2}|\lambda)\right) = \log\left(\Pr(Y_{t_2} = y_{t_2}|Y_{t_1} = y_{t_1}, \lambda)\right) + \log\left(\Pr(Y_{t_1} = y_{t_1}|\lambda)\right).$$

The Chain Rule

- The likelihood function

$$\mathcal{L}(y_{t_1}, y_{t_2}|\lambda) = \Pr(Y_{t_2} = y_{t_2}|Y_{t_1} = y_{t_1}, \lambda)\Pr(Y_{t_1} = y_{t_1}|\lambda).$$

- Accordingly,

$$\begin{aligned}\log\left(\mathcal{L}(y_{t_1}, y_{t_2}|\lambda)\right) = {} & \log\left(\Pr(Y_{t_2} = y_{t_2}|Y_{t_1} = y_{t_1}, \lambda)\right) \\ & + \log\left(\Pr(Y_{t_1} = y_{t_1}|\lambda)\right).\end{aligned}$$

- The Fisher Information:

$$\mathcal{FI}_{(Y_{t_1}, Y_{t_2})}(\lambda) = \mathcal{FI}_{(Y_{t_2}|Y_{t_1})}(\lambda) + \mathcal{FI}_{(Y_{t_1})}(\lambda).$$

Two-Parameter Geometric Distribution

Definition

A discrete random variable V has the "**Two-Parameter Geometric**" distribution with parameters $\alpha \in [0, 1)$ and $\beta \in (0, 1)$, denoted by **TPG**(α, β), if its **probability mass function** (**p.m.f.**) is

$$P_V(v) = \begin{cases} \alpha & \text{for } v = 0 \\ (1-\alpha)\beta(1-\beta)^{v-1} & \text{for } v = 1, 2, \dots . \end{cases}$$

Three-Parameter Negative Binomial Distribution

Definition

Suppose $V_1, \ldots, V_r$ are **i.i.d.** random variables with common $\mathrm{TPG}(\alpha, \beta)$ distribution. If $\mathbf{W} := \sum_{\mathbf{i}=1}^{\mathbf{r}} \mathbf{V_i}$, then W has "**Three-Parameter Negative Binomial**" distribution with parameters $\mathbf{r}$, α and β, denoted by $\mathbf{TPNB}(\mathbf{r}, \alpha, \beta)$.

Three-Parameter Negative Binomial Distribution

Definition

Suppose $V_1, \dots, V_r$ are **i.i.d.** random variables with common $\mathrm{TPG}(\alpha, \beta)$ distribution. If $\mathbf{W := \sum_{i=1}^{r} V_i}$, then W has "**Three-Parameter Negative Binomial**" distribution with parameters $\mathbf{r}$, α and β, denoted by $\mathbf{TPNB(r, \alpha, \beta)}$.

Proposition (Bean, Eshragh and Ross; 2014)

If W follows the $TPNB(r, \alpha, \beta)$ distribution, then its **p.m.f.** *is*

$$P_W(w) = \begin{cases} \alpha^r & \text{for } w = 0 \\ \displaystyle\sum_{\xi=1}^{\min\{r,w\}} \binom{w-1}{\xi-1} \beta^{\xi}(1-\beta)^{w-\xi} \binom{r}{\xi} (1-\alpha)^{\xi} \alpha^{r-\xi} & \text{for } w \geq 1 \end{cases}$$

The Distribution of Y_t

Theorem (Bean, Eshragh and Ross; 2014)

Consider the **POSBP** $\{Y_t,\ t \geq 0\}$ *with* **parameters** (λ, p) *and the* **initial population size** $x_0 \geq 1$. *For any real value* $t > 0$, *the random variable* Y_t *follows the* $\mathbf{TPNB(x_0, (1-p)\beta_t, \beta_t)}$ *distribution where*

$$\beta_t := \frac{e^{-\lambda t}}{p + (1-p)e^{-\lambda t}} .$$

The Fisher Information for a Single Observation

Proposition (Bean, Eshragh and Ross; 2014)

Consider the **POSBP** $\{Y_t,\ t \geq 0\}$ *with* **parameters** (λ, p)*. The Fisher Information of a single observation* Y_{t_1} *for parameter* λ *is equal to*

$$\mathcal{FI}_{\mathbf{Y}_1}(\lambda) = \frac{pt_1^2\left(p+(1-p)(1-e^{-\lambda t_1})e^{-\lambda t_1}\right)}{(1-e^{-\lambda t_1})(p+(1-p)e^{-\lambda t_1})^2}.$$

The Distribution of $(Y2|Y1 = y_{t_1})$

Theorem (Bean, Eshragh and Ross; 2014)

Consider the **POSBP** $\{Y_t,\ t \geq 0\}$ *with* **parameters** (λ, p). *Then*

$$\mathbf{W} \stackrel{\mathbf{d}}{=} (\mathbf{Y_{t_2}}|\mathbf{Y_{t_1}} = \mathbf{y_{t_1}}) + \mathbf{V}$$

where $(Y_{t_2}|Y_{t_1} = y_{t_1})$ *and* V *are mutually independent and*

$$W \sim TPNB(y_{t_1} + 1, (1-p)\beta^{\circ}, \beta^{\circ})$$

and

$$V \sim TPG((1-p)\beta_{t2-t1}, \beta_{t2-t1}).$$

Bounds for the General Form of the Fisher Information

Theorem

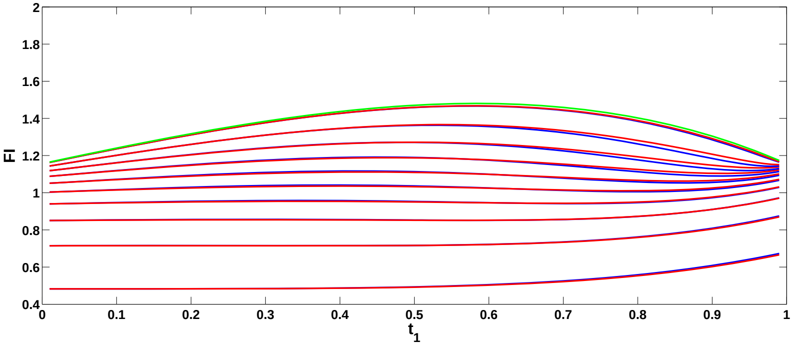
If $Z_1, \ldots, Z_n$ are independent random variables from distributions with common unknown parameter γ and $\mathbf{g} : \mathbb{R}^{\mathbf{n}} \rightarrow \mathbb{R}$ is a real-value function, then

$$\mathcal{FI}_{\mathbf{g}(\mathbf{Z}_1,\ldots,\mathbf{Z}_n)}(\gamma) \leq \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathcal{FI}_{\mathbf{Z}_\mathbf{i}}(\gamma) .$$

Furthermore, equality occurs if and only if g is a sufficient estimator for γ.

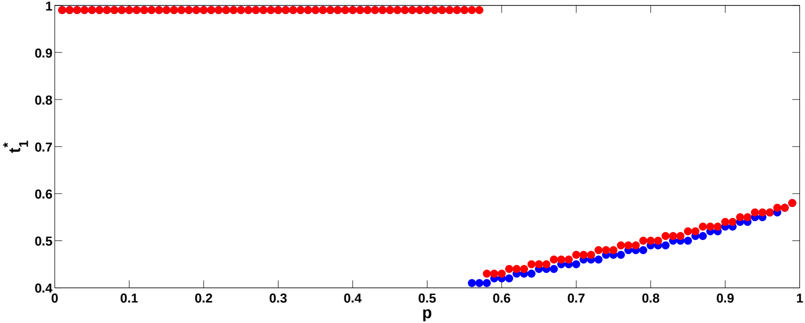
Results for $\lambda = 2$, $n = 2$ and $t_2^* = \tau = 1$

- The Fisher Information (blue) and its Approximation (red) vs. t_1



Results for $\lambda = 2$, $n = 2$ and $t_2^* = \tau = 1$

- Optimal observation time t_1^* vs. p



Bounds for the Fisher Information

- By exploiting the last two theorems, we found a **lower** and an **upper** bounds for the Fisher Information.

Bounds for the Fisher Information

- By exploiting the last two theorems, we found a **lower** and an **upper** bounds for the Fisher Information.

Theorem (Bean, Eshragh and Ross; 2014)

The approximation function for the Fisher Information **lies within** *the lower and upper bounds found for the Fisher Information.*

Bounds for the Fisher Information

- By exploiting the last two theorems, we found a **lower** and an **upper** bounds for the Fisher Information.

Theorem (Bean, Eshragh and Ross; 2014)

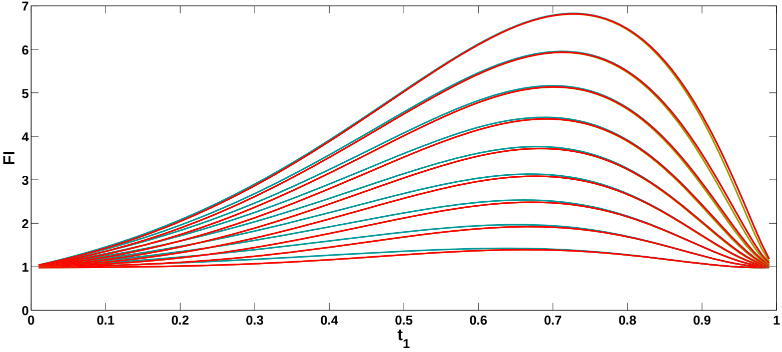
The approximation function for the Fisher Information ***lies within*** *the lower and upper bounds found for the Fisher Information.*

Theorem (Bean, Eshragh and Ross; 2014)

The lower and upper bounds for the Fisher Information ***approach together*** *as λ tends to infinity.*

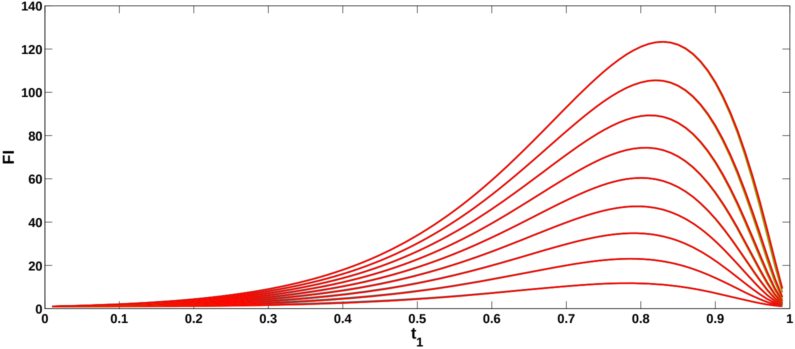
Results for $\lambda = 6$, $n = 2$ and $t_2^* = 1$

- Lower (brown) and Upper (green) Bounds for The Fisher Information and its Approximation (red) vs. t_1



Results for $\lambda = 10$, $n = 2$ and $t_2^* = 1$

- Lower (brown) and Upper (green) Bounds for The Fisher Information and its Approximation (red) vs. t_1



Further Developments

- Developing analogous approximation for **higher values** of n .

Further Developments

- Developing analogous approximation for **higher values** of n.
- Investigating the quality of the **approximation**

$$\mathcal{FI}^{x_0}(\lambda) \approx x_0 \mathcal{FI}^{1}(\lambda)$$

for $x_0 > 1$.

Further Developments

- Developing analogous approximation for **higher values** of n .
- Investigating the quality of the **approximation**

$$\mathcal{FI}^{\mathbf{x}_0}(\lambda) \approx \mathbf{x}_0 \mathcal{FI}^1(\lambda)$$

  for $x_0 > 1$.
- Finding the Fisher Information to estimate parameter **p** along with λ, both together.

End

Thank you $\cdots$ Questions?