Fisher Information for a Partially-Observable Simple Birth Process

Ali Eshragh
(Joint work with Nigel Bean and Joshua Ross)

School of Mathematical Sciences
The University of Adelaide, Adelaide, 5005 Australia

Australia and New Zealand Applied Probability Workshop
Brisbane, Australia
8-11 July, 2013

Definition and Notation

- Let $\{X_t : t \in \mathrm{R}_0^+\}$ denote a **simple birth process (SBP)** with parameter λ. Moreover, $X_0 \overset{a.s.}{=} x_0$.

Definition and Notation

- Let $\{X_t : t \in \mathrm{R}_0^+\}$ denote a **simple birth process (SBP)** with parameter λ. Moreover, $X_0 \overset{a.s.}{=} x_0$.

- It is **Markovian** with infinitesimal conditions

$$\Pr(X_{t+h} = j | X_t = i) = \begin{cases} \lambda ih + \mathcal{O}(h) & \text{for } j = i + 1 \\ 1 - \lambda ih + \mathcal{O}(h) & \text{for } j = i \\ \mathcal{O}(h) & \text{otherwise} \end{cases} .$$

Definition and Notation

- Let $\{X_t : t \in \mathrm{R}_0^+\}$ denote a **simple birth process (SBP)** with parameter λ. Moreover, $X_0 \overset{a.s.}{=} x_0$.
- It is **Markovian** with infinitesimal conditions

$$\Pr(X_{t+h} = j | X_t = i) = \begin{cases} \lambda i h + \mathcal{O}(h) & \text{for } j = i+1 \\ 1 - \lambda i h + \mathcal{O}(h) & \text{for } j = i \\ \mathcal{O}(h) & \text{otherwise} \end{cases} .$$

- **Transition probability** $\Pr(X_{s+t} = j | X_s = i) \; := \; p_{ij}(t)$:

$$p_{ij}(t) \;=\; \binom{j-1}{i-1} e^{-\lambda t i}(1 - e^{-\lambda t})^{j-i} .$$

The Fisher Information

- **Estimating** the unknown parameter λ .

The Fisher Information

- **Estimating** the unknown parameter λ.
- Take the **sample** $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ at sampling times $0 < t_1 \leq t_2 \leq \ldots \leq t_n$, respectively.

The Fisher Information

- **Estimating** the unknown parameter λ.
- Take the **sample** $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ at sampling times $0 < t_1 \leq t_2 \leq \ldots \leq t_n$, respectively.
- Finding the **volume of information** obtained from the sample to estimate the unknown parameter λ.

The Fisher Information

- **Estimating** the unknown parameter λ.
- Take the **sample** $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ at sampling times $0 < t_1 \leq t_2 \leq \ldots \leq t_n$, respectively.
- Finding the **volume of information** obtained from the sample to estimate the unknown parameter λ.
- A good tool to measure the volume of information gained from a sample is the **Fisher Information**.

The Fisher Information

- **Estimating** the unknown parameter λ.
- Take the **sample** $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ at sampling times $0 < t_1 \leq t_2 \leq \ldots \leq t_n$, respectively.
- Finding the **volume of information** obtained from the sample to estimate the unknown parameter λ.
- A good tool to measure the volume of information gained from a sample is the **Fisher Information**.
- It can be shown that

$$\mathcal{FI}_{(X_{t_1}, X_{t_2}, \cdots, X_{t_n})}(\lambda) = E_{\mathcal{L}}\left[\left(\frac{d}{d\lambda}\ln(\mathcal{L}(X_{t_1}, X_{t_2}, \ldots, X_{t_n}; \lambda))\right)^2\right].$$

The Fisher Information for the Simple Birth Process

Proposition (Becker and Kersting, 1983)

The **Fisher Information** *for the simple birth process with the parameter* λ*, the initial value of* x_0 *and the sampling times of* $(t_1, t_2, \ldots, t_n)$ *is as follows:*

$$\mathcal{FI}_{(x_{t_1}, x_{t_2}, \cdots, x_{t_n})}(\lambda) = x_0 \sum_{i=1}^{n} \frac{(t_i - t_{i-1})^2}{e^{-\lambda t_{i-1}} - e^{-\lambda t_i}} .$$

Definition and Notation

- Suppose that at each sampling time t_i, we can observe the population, **partially**.

Definition and Notation

- Suppose that at each sampling time t_i, we can observe the population, **partially**.
- $\mathbf{Y_t}$ is the sample observed at time t.

Definition and Notation

- Suppose that at each sampling time t_i, we can observe the population, **partially**.
- $\mathbf{Y_t}$ is the sample observed at time t.
- $(Y_t | X_t = x) \sim \texttt{Binomial}(\mathbf{x}, \mathbf{p})$.

Definition and Notation

- Suppose that at each sampling time t_i, we can observe the population, **partially**.
- $\mathbf{Y_t}$ is the sample observed 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).

Definition and Notation

- Suppose that at each sampling time t_i, we can observe the population, **partially**.
- $\mathbf{Y_t}$ is the sample observed 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; 2013)

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; 2013)

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

- However,

$$\Pr(Y_{t_1} = y_{t_1}, Y_{t_2} = y_{t_2}, \ldots, Y_{t_n} = y_{t_n} | X_{t_1} = x_{t_1}, X_{t_2} = x_{t_2}, \ldots, X_{t_n} = x_{t_n})$$

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

The Fisher Information for the POSBP

- The Fisher Information:

$$\mathcal{FI}_{(Y_{t_1},Y_{t_2},\cdots,Y_{t_n})}(\lambda) = \sum_{y_{t_1},y_{t_2},\ldots,y_{t_n}} \frac{\left(\frac{d\mathcal{L}(y_{t_1},y_{t_2},\ldots,y_{t_n};\,\lambda)}{d\lambda}\right)^2}{\mathcal{L}(y_{t_1},y_{t_2},\ldots,y_{t_n};\,\lambda)}.$$

The Fisher Information for the POSBP

- The Fisher Information:

$$\mathcal{FI}_{(Y_{t_1},Y_{t_2},\cdots,Y_{t_n})}(\lambda) = \sum_{y_{t_1},y_{t_2},\ldots,y_{t_n}} \frac{\left(\frac{d\mathcal{L}(y_{t_1},y_{t_2},\ldots,y_{t_n};\,\lambda)}{d\lambda}\right)^2}{\mathcal{L}(y_{t_1},y_{t_2},\ldots,y_{t_n};\,\lambda)}.$$

- Here, the likelihood function $\mathcal{L}(y_{t_1},y_{t_2},\ldots,y_{t_n};\,\lambda)$ is equal to

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

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

The Fisher Information for the POSBP

- The Fisher Information:

$$\mathcal{FI}_{(Y_{t_1},Y_{t_2},\cdots,Y_{t_n})}(\lambda) = \sum_{y_{t_1},y_{t_2},\ldots,y_{t_n}} \frac{\left(\frac{d\mathcal{L}(y_{t_1},y_{t_2},\ldots,y_{t_n};\,\lambda)}{d\lambda}\right)^2}{\mathcal{L}(y_{t_1},y_{t_2},\ldots,y_{t_n};\,\lambda)}.$$

- Here, the likelihood function $\mathcal{L}(y_{t_1},y_{t_2},\ldots,y_{t_n};\,\lambda)$ is equal to

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

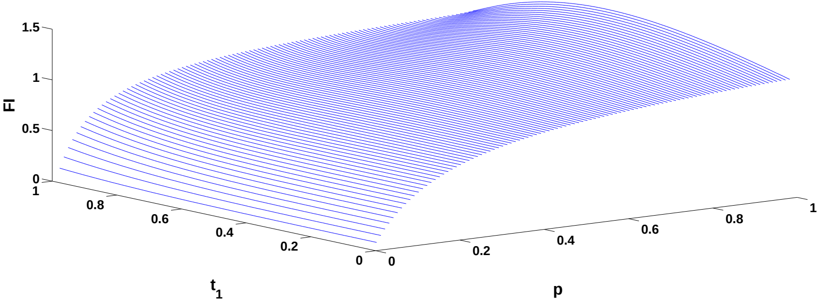
where $\upsilon_{i-1,i} := e^{-\lambda(t_i-t_{i-1})}$.

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

$$\Pr\left(E[Z]-12\sqrt{Var(Z)} \leq Z \leq E[Z]+12\sqrt{Var(Z)}\right) \geq 1-\frac{1}{12^2} = 99.3\%.$$

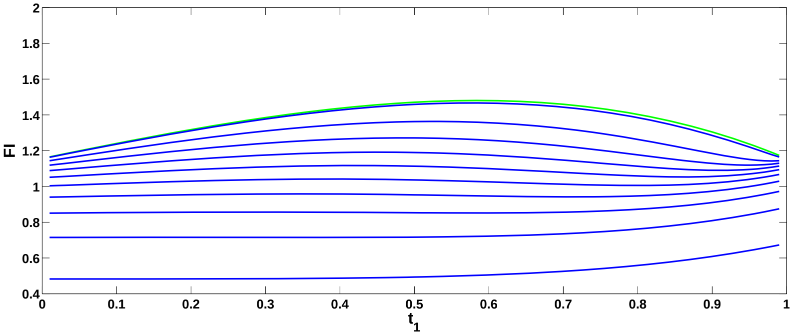
Results for $x_0 = 1$, $\lambda = 2$, $n = 2$ and $t_2 = 1$

- The Fisher Information vs. t_1 and p



Results for $x_0 = 1$, $\lambda = 2$, $n = 2$ and $t_2 = 1$

- The Fisher Information vs. t_1



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,

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

- The Fisher Information:

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

Delayed Geometric Distribution

Definition

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

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

Delayed Geometric Distribution

Definition

A discrete random variable V has the "**Delayed Geometric**" distribution with parameters $\alpha \in [0,1)$ and $\beta \in (0,1)$, denoted by $\mathbf{DG}(\alpha,\beta)$, if its **probability mass function** (**p.m.f.**) is

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

Remark

The $DG(\beta,\beta)$ and $DG(0,\beta)$ distributions reduce, respectively, to the **Geometric distribution-failure model** *and* **-success model** *both with parameter β.*

Delayed Negative Binomial Distribution

Definition

Suppose $V_1, \cdots, V_r$ are **i.i.d.** random variables with common $DG(\alpha, \beta)$ distribution. If $\mathbf{W} := \sum_{i=1}^{r} \mathbf{V}_i$, then W has "**Delayed Negative Binomial**" distribution with parameters $\mathbf{r}$, α and β, denoted by $\mathbf{DNB}(\mathbf{r}, \alpha, \beta)$.

Delayed Negative Binomial Distribution

Definition

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

Proposition (Bean, Eshragh and Ross; 2013)

If W follows the $DNB(r, \alpha, \beta)$ 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; 2013)

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* $\mathbf{DNB(x_0, (1-p)\beta_t, \beta_t)}$ *distribution where*

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

The Distribution of Y_t

Theorem (Bean, Eshragh and Ross; 2013)

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* $\mathbf{DNB}(\mathbf{x_0}, (\mathbf{1}-\mathbf{p})\boldsymbol{\beta}_\mathbf{t}, \boldsymbol{\beta}_\mathbf{t})$ *distribution where*

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

Corollary (Bean, Eshragh and Ross; 2013)

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

The Fisher Information for a Single Observation

Proposition (Bean, Eshragh and Ross; 2013)

Consider the **POSBP** $\{Y_t,\ t \geq 0\}$ *with* **parameters** (λ, p) *and the* **initial population size** $x_0 = 1$. *The Fisher Information of a single observation* Y_{t_1} *for parameter* λ *is equal to*

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

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

Theorem (Bean, Eshragh and Ross; 2013)

Consider the **POSBP** $\{Y_t,\ t \geq 0\}$ *with* **parameters** (λ, p) *and the* **initial population size** $x_0 = 1$. *Then*

$$\mathbf{W_1 \stackrel{d}{=} (Y_{t_2}|Y_{t_1} = y_{t_1}) + V_1}$$

where $(Y_{t_2}|Y_{t_1} = y_{t_1})$ *and* V_1 *are mutually independent and* $W_1 \sim DNB(y_{t_1} + 1, (1-p)\beta^\circ, \beta^\circ)$ *and* $V_1 \sim DG((1-p)\beta_{t2-t1}, \beta_{t2-t1})$.

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

Theorem (Bean, Eshragh and Ross; 2013)

Consider the **POSBP** $\{Y_t,\ t \geq 0\}$ *with* **parameters** (λ, p) *and the* **initial population size** $x_0 = 1$. *Then*

$$\mathbf{W_1} \overset{\mathbf{d}}{=} (\mathbf{Y_{t_2}} | \mathbf{Y_{t_1}} = \mathbf{y_{t_1}}) + \mathbf{V_1}$$

where $(Y_{t_2}|Y_{t_1} = y_{t_1})$ *and* V_1 *are mutually independent and* $W_1 \sim DNB(y_{t_1} + 1, (1-p)\beta^\circ, \beta^\circ)$ *and* $V_1 \sim DG((1-p)\beta_{t2-t1}, \beta_{t2-t1})$. *Moreover,*

$$(\mathbf{Y_{t_2}} | \mathbf{Y_{t_1}} = \mathbf{y_{t_1}}) \overset{\mathbf{d}}{=} \mathbf{W_2} + \mathbf{V_2}$$

where $W_2 \sim DNB(y_{t_1}, (1-p)\beta^\circ, \beta^\circ)$ *and* $V_2 \sim DG\left(\left(pe^{\lambda(t_2-t_1)} + 1 - p\right)\beta^\circ, \beta^\circ\right)$ *are two independent random variables.*

Bounds for the General Form of the Fisher Information

Theorem

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

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

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

Bounds for the General Form of the Fisher Information

Theorem

If $Z_1, \cdots, Z_n$ 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

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

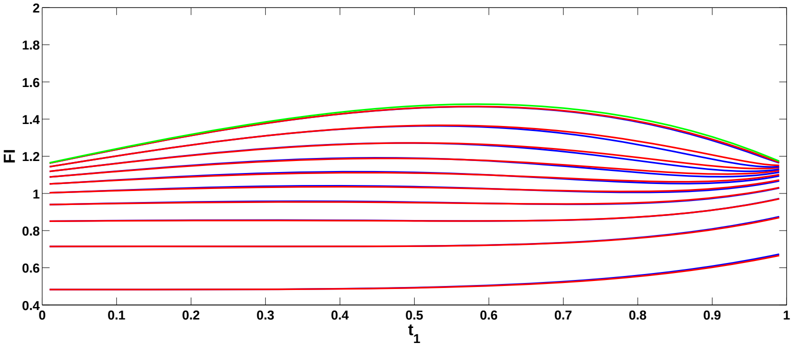
Furthermore, equality occurs if and only if g is a sufficient estimator for γ.

- Also, the **Carmer-Rao lower bound** implies that

$$\mathcal{FI}_{\mathbf{g}(\mathbf{Z}_1, \cdots, \mathbf{Z}_n)}(\gamma) \geq \frac{\left(\frac{\partial \mathrm{E}\left[\mathbf{g}(\mathbf{Z}_1, \cdots, \mathbf{Z}_n)\right]}{\partial \gamma}\right)^2}{\mathrm{Var}\left(\mathbf{g}(\mathbf{Z}_1, \cdots, \mathbf{Z}_n)\right)}.$$

Results for $x_0 = 1$, $\lambda = 2$, $n = 2$ and $t_2 = 1$

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



Bounds for the Fisher Information

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

Bounds for the Fisher Information

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

Theorem (Bean, Eshragh and Ross; 2013)

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 Fisher Information.

Theorem (Bean, Eshragh and Ross; 2013)

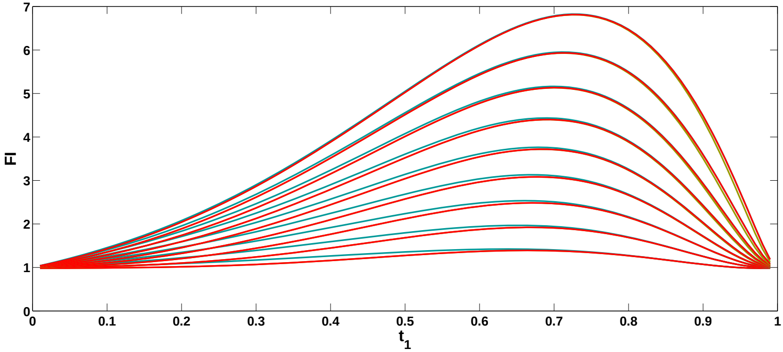
The approximation function for the Fisher Information **lies within** *the lower and upper bounds found for the Fisher Information.*

Theorem (Bean, Eshragh and Ross; 2013)

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

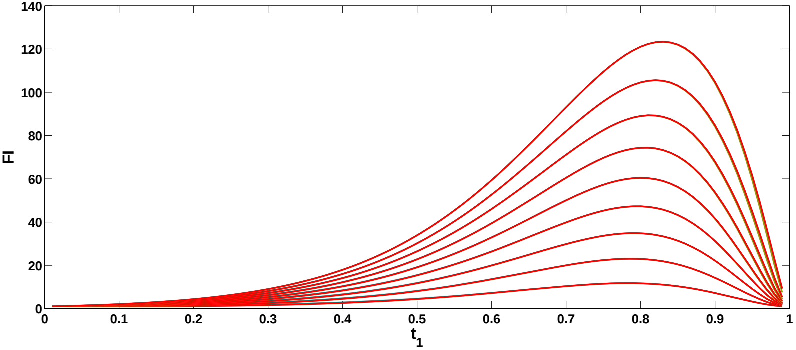
Results for $x_0 = 1$, $\lambda = 6$, $n = 2$ and $t_2 = 1$

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



Results for $x_0 = 1$, $\lambda = 10$, $n = 2$ 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.
- Finding the Fisher Information to estimate parameter **p** along with λ, both together.

End

Thank you $\cdots$ Questions?