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 $t$.

- $\{X_t : t \in \mathbb{R}_0^+\}$ is a stochastic process.
Definition and Notation

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

- $\{X_t : t \in \mathbb{R}_0^+\}$ is a stochastic process.

- Suppose $\{X_t : t \in \mathbb{R}_0^+\}$ is a simple birth process (SBP) with the birth rate $\lambda$. Moreover, $X_0 \overset{a.s.}{=} x_0$. 

Definition and Notation

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

- $\{X_t : t \in \mathbb{R}_0^+\}$ is a **stochastic process**.
- Suppose $\{X_t : t \in \mathbb{R}_0^+\}$ is a **simple birth process (SBP)** with the **birth rate** $\lambda$. 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 $t$.

- $\{X_t : t \in \mathbb{R}_0^+\}$ is a **stochastic process**.

- Suppose $\{X_t : t \in \mathbb{R}_0^+\}$ is a **simple birth process (SBP)** with the **birth rate** $\lambda$. 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

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

- **Estimating** the unknown parameter $\lambda$ through **maximum likelihood** method.

The likelihood function is constructed as:

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

$$= \prod_{i=2}^{n} \left( x_{i-1} - x_i \right) e^{-\lambda (t_{i-1} - t_i)} x_i e^{-\lambda (t_{i-1} - t_i)} (1 - e^{-\lambda (t_{i-1} - t_i)}) x_i - x_{i-1}.$$
Likelihood Function

- **Estimating** the unknown parameter $\lambda$ 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 $\lambda$ 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 $\lambda$ 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**

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

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

- **Estimating** the unknown parameter $\lambda$ 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**

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

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

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

- **Estimating** the unknown parameter $\lambda$ 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**

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

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

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

\[
= \prod_{i=1}^{n} \left( \frac{x_i - 1}{x_{i-1} - 1} \right) e^{-\lambda(t_i - t_{i-1})} x_{i-1} (1 - e^{-\lambda(t_i - t_{i-1})}) x_i - x_{i-1} .
\]
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 $\lambda$ is **maximized**.
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 $\lambda$ is **maximized**.

- A good tool to measure the expected volume of information gained from a set of observations is the **Fisher Information**.
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 $\lambda$ 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

$$\mathcal{F}\mathcal{I}(X_{t_1}, \ldots, X_{t_n})(\lambda) = E[\left(\frac{d}{d\lambda} \ln(\mathcal{L}(X_{t_1}, \ldots, X_{t_n}; \lambda))\right)^2].$$
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 $\lambda$ 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, $(t_1^*, \ldots, t_n^*) \in \arg\max \{\mathcal{FI}(X_{t_1}, \ldots, X_{t_n})(\lambda)\}.$
Proposition (Becker and Kersting, 1983)

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

$$
\mathcal{F}\mathcal{I}(x_{t_1}, \ldots, x_{t_n})(\lambda) = x_0 \sum_{i=1}^{n} \frac{(t_i - t_{i-1})^2}{e^{-\lambda t_{i-1}} - e^{-\lambda t_i}}.
$$
Fisher Information and Optimal Observation Times

**Proposition (Becker and Kersting, 1983)**

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

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

**Optimal Observation Times (Becker and Kersting, 1983)**

$$t_i^* \approx \frac{3}{\lambda} \log \left( 1 + \frac{i}{n} \left( e^{\frac{\lambda \tau}{3}} - 1 \right) \right) \quad \text{for } i = 1, \ldots, n$$
Suppose that at each observation time, we can count the population, \textit{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 $p$. 

\[
\text{Suppose that at each observation time, we can count the population, } p \text{.} \\
\text{At each observation time, each individual can be counted } \text{independently } p. 
\]
Definition and Notation

- Suppose that at each observation time, we can count the population, \textit{partially}.

- At each observation time, each individual can be counted \textit{independently} with probability \( p \).

- \( Y_t \) is the number of individuals observed at at time \( t \).
Definition and Notation

- Suppose that at each observation time, we can count the population, \textit{partially}.

- At each observation time, each individual can be counted \textit{independently} with probability $p$.

- $Y_t$ is the number of individuals observed at time $t$.

- $(Y_t|X_t = x) \sim \text{Binomial}(x, p)$. 
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 $p$.

- $Y_t$ is the number of individuals observed at time $t$.

- $(Y_t | X_t = x) \sim \text{Binomial}(x, p)$.

- We call the stochastic process $\{Y_t : t \in \mathbb{R}_0^+\}$ the *partially-observable simple birth process (POSBP)* with parameters $(\lambda, p)$. 
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 $p$.

- $Y_t$ is the number of individuals observed at time $t$.

- $(Y_t | X_t = x) \sim \text{Binomial}(x, p)$.

- We call the stochastic process $\{Y_t : t \in \mathbb{R}_0^+\}$ the **partially-observable simple birth process (POSBP)** with parameters $(\lambda, p)$.

- $\text{POSBP}(\lambda, 1) \equiv \text{SBP}(\lambda)$. 
Markovian or non-Markovian?

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

The POSBP \( \{ Y_t : t \in \mathbb{R}_0^+ \} \) with parameters \((\lambda, p)\) 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 $(\lambda, p)$ 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:

\[
\mathcal{L}(y_t_1, \ldots, y_t_n; \lambda) = \Pr(Y_t_1 = y_t_1, \ldots, Y_t_n = y_t_n)
= \sum_{x_{t_1}, \ldots, x_{t_n}} \Pr(Y_t_1 = y_t_1, \ldots, Y_t_n = y_t_n | X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \Pr(X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n})
\]

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

The likelihood function:

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

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

\[
= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \Pr(Y_{t_i} = y_{t_i} | X_{t_i} = x_{t_i}) \Pr(X_{t_i} = x_{t_i} | X_{t_{i-1}} = x_{t_{i-1}})
\]
The likelihood function:

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

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

\[
= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \Pr(Y_{t_i} = y_{t_i} | X_{t_i} = x_{t_i}) \Pr(X_{t_i} = x_{t_i} | X_{t_{i-1}} = x_{t_{i-1}})
\]

\[
= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \left( p_{y_{t_i}}^y q_{x_{t_i}}^{y_{t_i}} \right)
\]

where \( q_{x_{t_i}} := 1 - p_{y_{t_i}} \) and \( \nu_{i-1,i} := e^{-\lambda (t_{i} - t_{i-1})} \).
The likelihood function:

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

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

\[ \times \Pr(X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \]

\[ = \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \Pr(Y_{t_i} = y_{t_i} | X_{t_i} = x_{t_i}) \Pr(X_{t_i} = x_{t_i} | X_{t_{i-1}} = x_{t_{i-1}}) \]

\[ = \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \left( \frac{x_{t_i}}{y_{t_i}} \right)^{y_{t_i}} q^{x_{t_i} - y_{t_i}} \]
The likelihood function:

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

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

\[
= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \Pr(Y_{t_i} = y_{t_i} | X_{t_i} = x_{t_i}) \Pr(X_{t_i} = x_{t_i} | X_{t_{i-1}} = x_{t_{i-1}})
\]

\[
= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \binom{x_{t_i}}{y_{t_i}} p^{y_{t_i}} q^{x_{t_i} - y_{t_i}} \left( \frac{x_{t_i} - 1}{x_{t_{i-1}} - 1} \right)^{x_{t_{i-1}} - 1} (1 - \nu_{i-1,i})^{x_{t_i} - x_{t_{i-1}}},
\]
The likelihood function:

\[
\mathcal{L}(y_{t_1}, \ldots, y_{t_n}; \lambda) = \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n}) \\
= \sum_{x_{t_1}, \ldots, x_{t_n}} \Pr(Y_{t_1} = y_{t_1}, \ldots, Y_{t_n} = y_{t_n} | X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
\times \Pr(X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}) \\
= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \Pr(Y_{t_i} = y_{t_i} | X_{t_i} = x_{t_i}) \Pr(X_{t_i} = x_{t_i} | X_{t_{i-1}} = x_{t_{i-1}}) \\
= \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \left( \begin{array}{c} x_{t_i} \\ y_{t_i} \end{array} \right)^{p_{y_{t_i}} - y_{t_i}} \left( \begin{array}{c} x_{t_i} - 1 \\ x_{t_{i-1}} - 1 \end{array} \right)^{x_{t_{i-1}} - 1} (1 - v_{i-1,i})^{x_{t_i} - x_{t_{i-1}}} ,
\]

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

- The Fisher Information:

\[ \mathcal{FI}(\gamma_1, \ldots, \gamma_n)(\lambda) = \mathbb{E}_\mathcal{L} \left[ \left( \frac{d \log(\mathcal{L})}{d\lambda} \right)^2 \right] \]
The Fisher Information:

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

\[ = \sum_{\gamma_{t_1}, \ldots, \gamma_{t_n}} \left( \frac{d \log(\mathcal{L})}{d \lambda} \right)^2 \mathcal{L} \]
The Fisher Information:

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

\[ = \sum_{\gamma_{t_1}, \ldots, \gamma_{t_n}} \left( \frac{d \log(\mathcal{L})}{d\lambda} \right)^2 \mathcal{L} \]

\[ = \sum_{\gamma_{t_1}, \ldots, \gamma_{t_n}} \left( \frac{d \mathcal{L}}{d\lambda} \right)^2 \mathcal{L} \]
Fisher Information

The Fisher Information:

\[ F \mathcal{I}(y_{t_1}, \ldots, y_{t_n})(\lambda) = \mathbb{E}_\mathcal{L} \left[ \left( \frac{d \log(\mathcal{L})}{d\lambda} \right)^2 \right] \]

\[ = \sum_{y_{t_1}, \ldots, y_{t_n}} \left( \frac{d \log(\mathcal{L})}{d\lambda} \right)^2 \mathcal{L} \]

\[ = \sum_{y_{t_1}, \ldots, y_{t_n}} \left( \frac{d \mathcal{L}}{d\lambda} \right)^2 \mathcal{L} \]

\[ = \sum_{y_{t_1}, \ldots, y_{t_n}} \left( \frac{d \mathcal{L}(y_{t_1}, \ldots, y_{t_n}; \lambda)}{d\lambda} \right)^2 \frac{\mathcal{L}(y_{t_1}, \ldots, y_{t_n}; \lambda)}{\mathcal{L}(y_{t_1}, \ldots, y_{t_n}; \lambda)}. \]
Proposition (Bean, Eshragh and Ross; 2014)

For a POSBP with \( n \) observations and time horizon \( \tau \), 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 \( \tau \).
Theoretical Result

Proposition (Bean, Eshragh and Ross; 2014)

For a POSBP with \( n \) observations and time horizon \( \tau \), 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 \( \tau \).

Proposition (Bean, Eshragh and Ross; 2014)

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

- The Fisher Information:

\[
\mathcal{F}(y_{t_1}, \ldots, y_{t_n})(\lambda) = \sum_{y_{t_1}, \ldots, y_{t_n}} \frac{(dL(y_{t_1}, \ldots, y_{t_n}; \lambda))^2}{L(y_{t_1}, \ldots, y_{t_n}; \lambda)}.
\]
Truncated Summation

The Fisher Information:

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

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

\[
\sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^n \left( \frac{x_{t_i}}{y_{t_i}} \right)^p (1-p)^{x_{t_i}-y_{t_i}} \left( \frac{x_{t_i} - 1}{x_{t_i+1} - 1} \right)^{x_{t_i+1}-i} (1 - \nu_{i-1,i})^{x_{t_i}-x_{t_i+1}},
\]

where \( \nu_{i-1,i} := e^{-\lambda(t_i-t_{i-1})} \).
Truncated Summation

- The Fisher Information:

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

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

\[ \sum_{x_{t_1}, \ldots, x_{t_n}} \prod_{i=1}^{n} \binom{x_{t_i}}{y_{t_i}} p^{y_{t_i}}(1 - p)^{x_{t_i} - y_{t_i}} \left( \frac{x_{t_i} - 1}{x_{t_i - 1} - 1} \right)^w_{i-1,i} (1 - \nu_{i-1,i})^{x_{t_i} - x_{t_i - 1}}, \]

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

- By exploiting Chebyshev’s inequality, we have

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

Motivating from Chebyshev’s inequality:

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

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

Motivating from Chebyshev’s inequality:

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

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

Lemma (Eshragh, Bean and Ross; 2014)

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

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

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

\[ \text{Var}(X_t \mid Y_t = y_t) = \frac{(y_t + 1)(1 - p)e^{\lambda t}(e^{\lambda t} - 1)}{(pe^{\lambda t} + 1 - p)^2}. \]
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,

\[ \log (\mathcal{L}(y_{t_1}, y_{t_2}|\lambda)) = \log (\Pr(Y_{t_2} = y_{t_2}|Y_{t_1} = y_{t_1}, \lambda)) \]
\[ + \log (\Pr(Y_{t_1} = y_{t_1}|\lambda)). \]
The Chain Rule

- The likelihood function

\[ \mathcal{L}(y_{t1}, y_{t2} | \lambda) = \Pr(Y_{t2} = y_{t2} | Y_{t1} = y_{t1}, \lambda) \Pr(Y_{t1} = y_{t1} | \lambda). \]

- Accordingly,

\[ \log (\mathcal{L}(y_{t1}, y_{t2} | \lambda)) = \log (\Pr(Y_{t2} = y_{t2} | Y_{t1} = y_{t1}, \lambda)) + \log (\Pr(Y_{t1} = y_{t1} | \lambda)). \]

- The Fisher Information:

\[ \mathcal{FI}(y_{t1}, y_{t2})(\lambda) = \mathcal{FI}(y_{t2} | y_{t1})(\lambda) + \mathcal{FI}(y_{t1})(\lambda). \]
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 $\text{TPG}(\alpha, \beta)$, if its probability mass function (p.m.f.) is

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

**Definition**

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

Proposition (Bean, Eshragh and Ross; 2014) If $W$ follows the $TPNB(r, \alpha, \beta)$ distribution, then its p.m.f. is

$$P_W(w) = \begin{cases} \alpha^r & \text{for } w = 0 \\ \frac{\sum_{\xi=1}^{\min\{r, w\}} (w-1)_{\xi-1} \beta^\xi (1-\beta)^{w-\xi} (r^\xi \alpha^{r-\xi}) (1-\alpha)^{\xi}}{\alpha^r} & \text{for } w \geq 1 \end{cases}$$
Three-Parameter Negative Binomial Distribution

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

Proposition (Bean, Eshragh and Ross; 2014)

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

$$
P_W(w) = \begin{cases} 
\alpha^r & \text{for } w = 0 \\
\min \{ r, w \} & \\
\sum_{\xi=1}^{\min \{ r, w \}} \left( \frac{w - 1}{\xi - 1} \right) \beta^{\xi} (1 - \beta)^{w - \xi} \binom{r}{\xi} (1 - \alpha)^{\xi} \alpha^{r - \xi} & \text{for } w \geq 1
\end{cases}
$$
The Distribution of $Y_t$

Theorem (Bean, Eshragh and Ross; 2014)

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

$$\beta_t := \frac{e^{-\lambda t}}{p + (1 - p)e^{-\lambda t}}.$$
Proposition (Bean, Eshragh and Ross; 2014)

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

\[
\mathcal{FI}_{Y_1}(\lambda) = \frac{pt_1^2 (p + (1 - p)(1 - e^{-\lambda t_1})e^{-\lambda t_1})}{(1 - e^{-\lambda t_1})(p + (1 - p)e^{-\lambda t_1})^2}.
\]
The Distribution of \((Y_{t_2}\mid Y_{t_1} = y_{t_1})\)

Theorem (Bean, Eshragh and Ross; 2014)

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

\[ W \overset{d}{=} (Y_{t_2}\mid Y_{t_1} = y_{t_1}) + V \]

where \((Y_{t_2}\mid Y_{t_1} = y_{t_1})\) and \(V\) are mutually independent and

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

and

\[ V \sim TPG((1 - p)\beta_{t_2 - t_1}, \beta_{t_2 - t_1}). \]
Theorem

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

$$\mathcal{FI}_g(Z_1, \ldots, Z_n)(\gamma) \leq \sum_{i=1}^{n} \mathcal{FI}_{Z_i}(\gamma).$$

Furthermore, equality occurs if and only if $g$ is a sufficient estimator for $\gamma$. 
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. $p$
By exploiting the last two theorems, we found a lower and an upper 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.
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 $\lambda$ 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

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

for $x_0 > 1$.

- Finding the Fisher Information to estimate parameter $\theta$ along with $\lambda$, both together.
End

Thank you ··· Questions?