On Fisher Information of Some Functions of Phase Type Variates

Pavithra Celeste R and T.G.Deepak

Dept. of Mathematics

Indian Institute of Space Science and Technology

Thiruvananthapuram, Kerala, India

deepak@iist.ac.in

February 13-15, 2019

Phase type distributions

Distributions of absorption times in a Markov Process with $p < \infty$ transient states (the phases) and one absorbing state.

Phase type distributions

Distributions of absorption times in a Markov Process with $p < \infty$ transient states (the phases) and one absorbing state.

Phase type distributions were first considered by Neuts.

Phase type distributions

Distributions of absorption times in a Markov Process with $p < \infty$ transient states (the phases) and one absorbing state.

Phase type distributions were first considered by Neuts.

M.F. Neuts. *Probability distributions of Phase type*. In Liber Amicorum Prof. Emeritus H. Florin, Pages 173-206, 1975.

Phase type distributions

Distributions of absorption times in a Markov Process with $p < \infty$ transient states (the phases) and one absorbing state.

Phase type distributions were first considered by Neuts.

M.F. Neuts. *Probability distributions of Phase type*. In Liber Amicorum Prof. Emeritus H. Florin, Pages 173-206, 1975.

M.F. Neuts. *Matrix Geometric Solutions in Stochastic Models*, John Hopkins University press, 1981.

Motivations

Dense, in the metric of weak convergence of distributions, in all distributions on $[0, \infty)$.

Motivations

Dense, in the metric of weak convergence of distributions, in all distributions on $[0, \infty)$.

PH distributions act as the computational vehicle for many applied probability models since they constitute a very versatile class of distributions defined on the non negative real line that lead to models which are algorithmically tractable.

Motivations

Dense, in the metric of weak convergence of distributions, in all distributions on $[0, \infty)$.

PH distributions act as the computational vehicle for many applied probability models since they constitute a very versatile class of distributions defined on the non negative real line that lead to models which are algorithmically tractable.

Their formulation allow us to retain the Markov structure of Stochastic Models while being act as a reasonable approximation to a general distribution.

Continuous Phase type distributions

Let $\{X_t : t \geq 0\}$ be a CTMC on the finite state space $E = \{1, 2, ..., p, p+1\}$, where the states 1,2,....,p are transient (i.e given that we start in any one of these states, there is a non-zero probability that we will never return to it) and the state p+1 is absorbing.

Continuous Phase type distributions

Let $\{X_t : t \geq 0\}$ be a CTMC on the finite state space $E = \{1, 2, ..., p, p+1\}$, where the states 1,2,....,p are transient (i.e given that we start in any one of these states, there is a non-zero probability that we will never return to it) and the state p+1 is absorbing. Then $\{X_t : t \geq 0\}$ has an intensity matrix of the form

$$\Lambda = \begin{bmatrix} S & s^0 \\ \mathbf{0} & 0 \end{bmatrix}$$

where S is a $p \times p$ dimensional matrix (satisfying $S_{ii} < 0$ and$S_{ij} \geq 0$, for $i \neq j$), and t is a p-dimensional column vector satisfying $Se + s^0 = 0$.

Let $\beta_i = P\{X_0 = i\}$.

Then $(\beta_1, \beta_2, ..., \beta_p, \beta_{p+1})$ is called the initial probability vector of $\{X_t : t \geq 0\}$.

Let $\beta_i = P\{X_0 = i\}$.

Then $(\beta_1, \beta_2, ..., \beta_p, \beta_{p+1})$ is called the initial probability vector of $\{X_t : t \geq 0\}$.

Let $\tau = \inf \{t : X_t = p + 1\}$.

Let $\beta_i = P\{X_0 = i\}$.

Then $(\beta_1, \beta_2, ..., \beta_p, \beta_{p+1})$ is called the initial probability vector of $\{X_t : t \geq 0\}$.

Let $\tau = \inf \{t : X_t = p + 1\}$.

Then $\tau \sim PH(\beta, S)$ where $\beta = (\beta_1, \beta_2, ..., \beta_p)$.

Let $\beta_i = P\{X_0 = i\}$.

Then $(\beta_1, \beta_2, ..., \beta_p, \beta_{p+1})$ is called the initial probability vector of $\{X_t : t \geq 0\}$.

Let $\tau = \inf \{t : X_t = p + 1\}$.

Then $\tau \sim PH(\beta, S)$ where $\beta = (\beta_1, \beta_2, ..., \beta_p)$.

- $P\{\tau \leq x\} = 1 - \beta exp(Sx)\mathbf{e}$

Let $\beta_i = P\{X_0 = i\}$.

Then $(\beta_1, \beta_2, ..., \beta_p, \beta_{p+1})$ is called the initial probability vector of $\{X_t : t \geq 0\}$.

Let $\tau = \inf \{t : X_t = p + 1\}$.

Then $\tau \sim PH(\beta, S)$ where $\beta = (\beta_1, \beta_2, ..., \beta_p)$.

- $P\{\tau \leq x\} = 1 - \beta exp(Sx)\mathbf{e}$
- pdf of τ is

  $f(x) = \beta exp(Sx)s^0$

EM algorithm for parameter estimation of *PH*

S. Asmussen,O. Nerman and M. Olsson. Fitting phase-type distributions via the EM algorithm. Scand. J. Statist. *1996,* 23, *419-441*

EM algorithm for parameter estimation of PH

S. Asmussen,O. Nerman and M. Olsson. Fitting phase-type distributions via the EM algorithm. Scand. J. Statist. *1996,* 23, *419-441*

- Consider $y_1, y_2, \ldots, y_M$ be realization of i.i.d random variables from $PH_p(\beta, S)$

EM algorithm for parameter estimation of *PH*

S. Asmussen,O. Nerman and M. Olsson. Fitting phase-type distributions via the EM algorithm. Scand. J. Statist. *1996, 23, 419-441*

- Consider $y_1, y_2, \ldots, y_M$ be realization of i.i.d random variables from $PH_p(\beta, S)$
- Let $x = (x_1, x_2, \ldots, x_M)$ be the complete data, $\theta = (\beta, S, s^0)$, where $s^0 = -S\mathbf{e}$, be the parameter set and $f(.)$, the *pdf* of the PH variate

EM algorithm for parameter estimation of *PH*

S. Asmussen,O. Nerman and M. Olsson. Fitting phase-type distributions via the EM algorithm. Scand. J. Statist. *1996, 23, 419-441*

- Consider $y_1, y_2, \ldots, y_M$ be realization of i.i.d random variables from $PH_p(\beta, S)$
- Let $x = (x_1, x_2, \ldots, x_M)$ be the complete data, $\theta = (\beta, S, s^0)$, where $s^0 = -S\mathbf{e}$, be the parameter set and $f(.)$, the *pdf* of the PH variate
- Likelihood function is,

$$L_f(\theta; x) = \prod_{i=1}^{p} {\beta_i}^{B_i} \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} S_{ij}^{N_{ij}} e^{-S_{ij} Z_i} \prod_{l=1}^{p} {s_l^0}^{N_l} e^{-s_l^0 Z_l}$$

B_i, N_i, N_{ij} and Z_i are the sufficient statistics.

- B_i, the number of trajectories that start in phase i, $i = 1, 2, \cdots, p$.
- N_i, the number of trajectories for which absorption occurs from phase i, $i = 1, 2, \cdots, p$.
- N_{ij}, the number of transitions that occur from phase i to phase j, $1 \leq i,\ j \leq p,\ i \neq j$.
- Z_i, the total sojourn time in phase i for all the M trajectories combined, for $i = 1, 2, \cdots, p$.

The maximum likelihood estimators are,

$$\hat{S}_{ij} = \frac{N_{ij}}{Z_i}, \quad \hat{s}_i^0 = \frac{N_i}{Z_i} \text{ and } \hat{\beta}_i = \frac{B_i}{M}$$

The maximum likelihood estimators are,
$\hat{S}_{ij} = \dfrac{N_{ij}}{Z_i}, \quad \hat{s}_i^0 = \dfrac{N_i}{Z_i}$ and $\hat{\beta}_i = \dfrac{B_i}{M}$
Let θ_0 be any initial value of parameters and l_f be the log-likelihood function.

The maximum likelihood estimators are,

$\hat{S}_{ij} = \dfrac{N_{ij}}{Z_i}, \quad \hat{s}_i^0 = \dfrac{N_i}{Z_i}$ and $\hat{\beta}_i = \dfrac{B_i}{M}$

Let θ_0 be any initial value of parameters and l_f be the log-likelihood function.

EM Algorithm works as follows

1. E-Step: Calculate $h : \theta \rightarrow \mathbb{E}_{\theta_0}(l_f(\theta, x)|Y = y\ )$

The maximum likelihood estimators are,

$\hat{S}_{ij} = \dfrac{N_{ij}}{Z_i}, \quad \hat{s}_i^0 = \dfrac{N_i}{Z_i}$ and $\hat{\beta}_i = \dfrac{B_i}{M}$

Let θ_0 be any initial value of parameters and l_f be the log-likelihood function.

EM Algorithm works as follows

1. E-Step: Calculate $h : \theta \rightarrow \mathbb{E}_{\theta_0}(l_f(\theta, x) | Y = y\ )$
2. M-Step: $\theta_0 = argmax_\theta h(\theta)$

The maximum likelihood estimators are,
$\hat{S}_{ij} = \dfrac{N_{ij}}{Z_i}$, $\hat{s}_i^0 = \dfrac{N_i}{Z_i}$ and $\hat{\beta}_i = \dfrac{B_i}{M}$
Let θ_0 be any initial value of parameters and l_f be the log-likelihood function.

EM Algorithm works as follows

1. E-Step: Calculate $h : \theta \to \mathbb{E}_{\theta_0}(l_f(\theta, x)|Y = y)$
2. M-Step: $\theta_0 = argmax_\theta h(\theta)$
3. Go to E-Step

Estimates of Sufficient Statistics

Let $M(y,\beta,S)=\int\limits_{o}^{y} e^{S(y-u)}s^0\beta e^{Su}du$. Given a sample value y from $PH_p(\beta,S)$, we have the following estimates (conditional expectations given y) of the sufficient statistics:

$$
\begin{aligned}
\hat{B}_i(y,\beta,S) &= \frac{\beta_i \mathbf{e}_i^{\mathsf{T}} e^{Sy} s^0}{\beta e^{Sy} s^0} \\
\hat{Z}_i(y,\beta,S) &= \frac{M_{ii}(y,\beta,S)}{\beta e^{Sy} s^0} \\
\hat{N}_i(y,\beta,S) &= \frac{s_i^0 \beta e^{Sy} \mathbf{e}_i}{\beta e^{Sy} s^0} \\
\hat{N}_{ij}(y,\beta,S) &= \frac{S_{ij} M_{ji}(y,\beta,S)}{\beta e^{Sy} s^0}
\end{aligned}
$$

1. FSPH Class

V. Ramaswamy, and N. C. Viswanath. "Phase Type Distributions with Finite Support". *Stochastic Models*, 30:576-597, 2014.

Some Functions of PH Variates

1. FSPH Class

V. Ramaswamy, and N. C. Viswanath. "Phase Type Distributions with Finite Support". *Stochastic Models*, 30:576-597, 2014.

Definition

A random variable X has distribution $FSPH_1(\beta, S)$ over the interval $(a, b), a < b$ if it has the same distribution as the random variable $a + Ymod(b - a)$, where $Y \sim PH(\beta, S)$.

Some Functions of PH Variates

1. FSPH Class

V. Ramaswamy, and N. C. Viswanath. "Phase Type Distributions with Finite Support". *Stochastic Models*, 30:576-597, 2014.

Definition

A random variable X has distribution $FSPH_1(\beta, S)$ over the interval $(a, b), a < b$ if it has the same distribution as the random variable $a + Y mod(b - a)$, where $Y \sim PH(\beta, S)$.

Let $\mathbf{s}^0 = -S\mathbf{e}$

- $f(x) = \beta e^{S(x-a)}\{(I - e^{S(b-a)})^{-1}\mathbf{s}^0, a < x < b$

Some Functions of PH Variates

1. FSPH Class

V. Ramaswamy, and N. C. Viswanath. "Phase Type Distributions with Finite Support". *Stochastic Models*, 30:576-597, 2014.

Definition

A random variable X has distribution $FSPH_1(\beta, S)$ over the interval $(a, b), a < b$ if it has the same distribution as the random variable $a + Y mod(b - a)$, where $Y \sim PH(\beta, S)$.

Let $\mathbf{s}^0 = -S\mathbf{e}$

- $f(x) = \beta e^{S(x-a)}\{(I - e^{S(b-a)})^{-1}\mathbf{s}^0, a < x < b$
- Dense in the set of all distributions with support (a, b).

2. LogPH Variate

A. Ghosh, R Jana, V Ramaswami, J Rowland, N. K. Shankaranarayanan.

Modeling and characterization of large-scale Wi-Fi traffic in public hot-spots.

In: Proc. IEEE INFOCOM 2011, Sharighai, China, 2921-2929

Definition

The LogPH distribution, $LogPH(\beta, S)$, is defined as the distribution of the random variable $Y = e^X$ where X has a PH distribution with parameters (β, S)

2. LogPH Variate

A. Ghosh, R Jana, V Ramaswami, J Rowland, N. K. Shankaranarayanan.

Modeling and characterization of large-scale Wi-Fi traffic in public hot-spots.

In: Proc. IEEE INFOCOM 2011, Sharighai, China, 2921-2929

Definition

The LogPH distribution, $LogPH(\beta, S)$, is defined as the distribution of the random variable $Y = e^X$ where X has a PH distribution with parameters (β, S)

Its density function is

$$f_Y(y) = \frac{1}{y}\beta e^{S \log y} s^0, \quad y \geq 1$$

2. LogPH Variate

A. Ghosh, R Jana, V Ramaswami, J Rowland, N. K. Shankaranarayanan.

Modeling and characterization of large-scale Wi-Fi traffic in public hot-spots.

In: Proc. IEEE INFOCOM 2011, Sharighai, China, 2921-2929

Definition

The LogPH distribution, $LogPH(\beta, S)$, is defined as the distribution of the random variable $Y = e^X$ where X has a PH distribution with parameters (β, S)

Its density function is

$$f_Y(y) = \frac{1}{y}\beta e^{S \log y} s^0, \quad y \geq 1$$

Dense in the set of all distribution functions defined on $[1, \infty)$.

2. LogPH Variate

A. Ghosh, R Jana, V Ramaswami, J Rowland, N. K. Shankaranarayanan. Modeling and characterization of large-scale Wi-Fi traffic in public hot-spots. *In: Proc. IEEE INFOCOM* 2011, Sharighai, China, 2921-2929

Definition

The LogPH distribution, $LogPH(\beta, S)$, is defined as the distribution of the random variable $Y = e^X$ where X has a PH distribution with parameters (β, S)

Its density function is

$$f_Y(y) = \frac{1}{y}\beta e^{S \log y} s^0, \quad y \geq 1$$

Dense in the set of all distribution functions defined on $[1, \infty)$.

Has many successful uses in the context of queuing and reliability and is used to model insurance data.

Motivation

- Both FSPH and logPH random variates are functions of PH variates.
- Both are dense and have many applications
- The analysis and estimation of functions of PH variates are also important.

Cases Considered

We consider three types of functions.

- The function $g(X)$, of the PH variate X, is differentiable for all $X = x$ and either the derivative at x is strictly positive or negative

  Example: LogPH Variate

Cases Considered

We consider three types of functions.

- The function $g(X)$, of the PH variate X, is differentiable for all $X = x$ and either the derivative at x is strictly positive or negative
  Example: LogPH Variate
- The derivative of g is continuous and non zero for all but finite number of values of x and hence for every real number y, there exists $n = n(y)$ inverses .
  Example: $|X - k|$, where $k \in \mathbb{R}^+$

Cases Considered

We consider three types of functions.

- The function $g(X)$, of the PH variate X, is differentiable for all $X = x$ and either the derivative at x is strictly positive or negative
  Example: LogPH Variate
- The derivative of g is continuous and non zero for all but finite number of values of x and hence for every real number y, there exists $n = n(y)$ inverses .
  Example: $|X - k|$, where $k \in \mathbb{R}^+$
- $Y = g(X)$ and g is invertible only in a finite interval and at each point y the function is having countable number of inverses
  Example: FSPH variate, $\sin(X)$

Case 1

- Let $X \sim PH_p(\beta, S)$

Case 1

- Let $X \sim PH_p(\beta, S)$
- Let the function $Y = g(X)$ be differentiable for all $X = x$ and the derivative g' at x is either strictly positive or negative.

Case 1

- Let $X \sim PH_p(\beta, S)$
- Let the function $Y = g(X)$ be differentiable for all $X = x$ and the derivative g' at x is either strictly positive or negative.
- g will be invertible

Case 1

- Let $X \sim PH_p(\beta, S)$
- Let the function $Y = g(X)$ be differentiable for all $X = x$ and the derivative g' at x is either strictly positive or negative.
- g will be invertible
- n observations $y_1, y_2, \dots, y_n$ from $Y \Rightarrow g^{-1}(y_1), g^{-1}(y_2), \dots, g^{-1}(y_n)$, will be observations from $g^{-1}(Y) = X$, a PH variate

Case 1

- Let $X \sim PH_p(\beta, S)$
- Let the function $Y = g(X)$ be differentiable for all $X = x$ and the derivative g' at x is either strictly positive or negative.
- g will be invertible
- n observations $y_1, y_2, \dots, y_n$ from $Y \Rightarrow g^{-1}(y_1), g^{-1}(y_2), \dots, g^{-1}(y_n)$, will be observations from $g^{-1}(Y) = X$, a PH variate
- Estimate the parameters of Y using that of X

Case 2

Let $Y = g(X), X \sim PH_p(\beta, S)$ where, the derivative of g is continuous and non zero for all but finite number of values of x . Then for every y

1. there exist a positive integer $n = n(y)$ and real numbers $x_1, x_2, \ldots, x_n$ such that, $g[x_k] = y, \;\; g'[x_k] \neq 0, \;\; k = 1, 2, \ldots, n(y).$

Case 2

Let $Y = g(X), X \sim PH_p(\beta, S)$ where, the derivative of g is continuous and non zero for all but finite number of values of x . Then for every y

1. there exist a positive integer $n = n(y)$ and real numbers $x_1, x_2, \ldots, x_n$ such that, $g[x_k] = y, \;\; g'[x_k] \neq 0, \;\; k = 1, 2, \ldots, n(y).$
2. there does not exist any x such that $g(x) = y$ or $g^{'}(x) = 0$ in which case we write $n(y) = 0$

Case 2

Let $Y = g(X), X \sim PH_p(\beta, S)$ where, the derivative of g is continuous and non zero for all but finite number of values of x . Then for every y

1. there exist a positive integer $n = n(y)$ and real numbers $x_1, x_2, \dots, x_n$ such that, $g[x_k] = y, \;\; g'[x_k] \neq 0, \;\; k = 1, 2, \dots, n(y)$.
2. there does not exist any x such that $g(x) = y$ or $g^{'}(x) = 0$ in which case we write $n(y) = 0$

pdf of Y is,

$$h(y) = \begin{cases} \sum_{k=1}^{n} \beta e^{Sx_k(y)} s^0 |g^{'}(x_k(y)|^{-1} & n > 0 \\ 0 & n = 0 \end{cases}$$

Given a sample $y_1, y_2, \ldots, y_M$ from Y, we get sample points $x_{11}(y_1), \ldots, x_{1n_1}(y_1), \ldots, x_{M1}(y_M), \ldots, x_{Mn_M}(y_M)$ from X such that $g(x_{il}(y_i)) = y_i, \quad i = 1, 2, \ldots, M \quad l = 1, 2, \ldots, n_i$ where $n_i = n(y_i)$.

Given a sample $y_1, y_2, \ldots, y_M$ from Y, we get sample points $x_{11}(y_1), \ldots, x_{1n_1}(y_1), \ldots, x_{M1}(y_M), \ldots, x_{Mn_M}(y_M)$ from X such that $g(x_{il}(y_i)) = y_i, \quad i = 1, 2, \ldots, M \quad l = 1, 2, \ldots, n_i$ where $n_i = n(y_i)$.

$$\hat{B}_i(y, \beta, S) = P(X(0) = i | g(X) = y) = \sum_{k=1}^{n_y} |g'(x_k(y))|^{-1} \frac{\beta_i \mathbf{e}_i^T e^{Sx_k(y)} s^0}{h(y)}$$

$$\hat{Z}_i(y, \beta, S) = \sum_{k=1}^{n_y} |g'(x_k(y))|^{-1} \frac{M_{ii}(x_k(y), \beta, S)}{h(y)}$$

$$\hat{N}_i(y, \beta, S) = \sum_{k=1}^{n_y} |g'(x_k(y))|^{-1} \frac{s_i^0 \beta e^{Sx_k(y)} \mathbf{e}_i}{h(y)}$$

$$\hat{N}_{ij}(y, \beta, S) = \sum_{k=1}^{n_y} |g'(x_k(y))|^{-1} \frac{S_{ij} M_{ji}(x_k(y), \beta, S)}{h(y)}$$

Case 3

- $Y = g(X), \quad X \sim PH_p(\beta, S)$

Case 3

- $Y = g(X), \quad X \sim PH_p(\beta, S)$
- g is invertible in an interval of length k

Case 3

- $Y = g(X), \quad X \sim PH_p(\beta, S)$
- g is invertible in an interval of length k
- at each point y, g has a countable number of inverses, $h(y) + l(n)$ where $l(n) = kn - m$, $h(y)$ is the inverse in the interval of length k

Case 3

- $Y = g(X), \quad X \sim PH_p(\beta, S)$
- g is invertible in an interval of length k
- at each point y, g has a countable number of inverses, $h(y) + l(n)$ where $l(n) = kn - m$, $h(y)$ is the inverse in the interval of length k
- *pdf* of Y is

$$\begin{aligned} f(y) &= \sum_{n=0}^{\infty} \beta e^{S(h(y)+kn-m)} s^0 |h^{'}(y)| \\ &= \beta e^{S(h(y)-m)} (I - e^{Sk})^{-1} s^0 |h^{'}(y)| \end{aligned}$$

Define

$$
\begin{aligned}
g(n|y) &= P(X = h(y) + l(n) | Y = y) \\
&= \frac{\beta e^{S(h(y)-m)} e^{knS} s^0}{f(y)} |h'(y)|
\end{aligned}
$$

Define

$$\begin{aligned} g(n|y) &= P(X = h(y) + l(n) | Y = y) \\ &= \frac{\beta e^{S(h(y)-m)} e^{knS} s^0}{f(y)} |h'(y)| \end{aligned}$$

Put $a(y) = e^{S(h(y)-m)}(I - e^{Sk})^{-1} s^0$

$b(y) = \beta e^{S(h(y)-m)}(I - e^{Sk})^{-1}$

and $c(y) = \dfrac{|h'(y)|}{\beta a(y)}$

Define

$$
\begin{aligned}
g(n|y) &= P(X = h(y) + l(n) | Y = y) \\
&= \frac{\beta e^{S(h(y)-m)} e^{knS} s^0}{f(y)} |h^{'}(y)|
\end{aligned}
$$

Put $a(y) = e^{S(h(y)-m)}(I - e^{Sk})^{-1} s^0$

$b(y) = \beta e^{S(h(y)-m)}(I - e^{Sk})^{-1}$

and $c(y) = \dfrac{|h^{'}(y)|}{\beta a(y)}$

$$
\begin{aligned}
\text{then, } \hat{B}_i(y, \beta, S) &= c(y)\beta_i a_i(y) \\
\hat{Z}_i(y, \beta, S) &= c(y)M^*_{ii}(y, \beta, S) \\
\hat{N}_i(y, \beta, S) &= c(y)s^0_i b_i(y) \\
\hat{N}_{ij}(y, \beta, S) &= c(y)S_{ij}M^*_{ji}(y, \beta, S)
\end{aligned}
$$

where $M^*(y, \beta, S) = \sum_{n=0}^{\infty} M(h(y) + kn - m, \beta, S)$

Example

Let $X \sim PH_p(\beta, S)$ and $Y = \sin(X)$.

Example

Let $X \sim PH_p(\beta, S)$ and $Y = \sin(X)$. Then the pdf of Y is

$$f(y) = \begin{cases} \frac{1}{\sqrt{1-y^2}} \left[\beta(I - e^{2\pi S})^{-1} \left(e^{\pi S - sin^{-1}(y)S} + e^{S\sin^{-1}(y)}\right)\right] s^0 & 0 < y < 1 \\ \frac{1}{\sqrt{1-y^2}} \beta \left(I - e^{2\pi S}\right)^{-1} e^{\pi S} \left[e^{(\pi + \sin^{-1}(y))S} + e^{-\sin^{-1}(y)S}\right] s^0 & -1 < y < 0 \\ 0 & otherwise \end{cases}$$

Define, $E_i = [\mathbf{0}\ \mathbf{0}\ \ldots\ \mathbf{0}\ \mathbf{I}\ \mathbf{0}\ \ldots\ \mathbf{0}]$

and

$$C = \begin{bmatrix} S & s^0\beta & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & S & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & S \end{bmatrix}.$$

Estimates of the sufficient statistics are

$$\begin{aligned}
\hat{B}_i(y,\beta,S) &= \frac{(1-y^2)^{-1/2}}{f(y)}\beta_i e_i^{\mathsf{T}}(I-e^{2\pi S})^{-1}\Big[\Big(e^{S\sin^{-1}(y)}+\\
&\quad e^{\pi S - S\sin^{-1}(y)}\Big)\mathbf{1}_{0<y<1} + e^{\pi S}\Big(e^{\pi S+\sin^{-1}(y)S}+\\
&\quad e^{-\sin^{-1}(y)S}\Big)\mathbf{1}_{-1<y<0}\Big]s^0\\
\hat{N}_i(y,\beta,S) &= \frac{(1-y^2)^{-1/2}}{f(y)}\beta(I-e^{2\pi S})^{-1}\Big[\Big(e^{S\sin^{-1}(y)}+\\
&\quad e^{\pi S - S\sin^{-1}(y)}\Big)\mathbf{1}_{0<y<1} + e^{\pi S}\Big(e^{\pi S+\sin^{-1}(y)S}+\\
&\quad e^{-\sin^{-1}(y)S}\Big)\mathbf{1}_{-1<y<0}\Big]e_i s_i^0
\end{aligned}$$

$$
\begin{aligned}
\hat{Z}_i(y,\beta,S) &= \frac{(1-y^2)^{-1/2}}{f(y)}\Big\{\Big[E_1(I-e^{2\pi C})^{-1}e^{\sin^{-1}(y)C}E_2^{\mathsf{T}}+ \\
&E_1(I-e^{2\pi C})^{-1}e^{(\pi-\sin^{-1}(y))C}E_2^{\mathsf{T}}\Big]_{ii}\mathbf{1}_{0<y<1}+ \\
&\Big[E_1(I-e^{2\pi C})^{-1}e^{(\pi-\sin^{-1}(y))C}E_2^{\mathsf{T}}+ \\
&E_1(I-e^{2\pi C})^{-1}e^{(2\pi+\sin^{-1}(y))C}E_2^{\mathsf{T}}\Big]_{ii}\mathbf{1}_{-1<y<0}\Big\} \\
\hat{N}_{ij}(y,\beta,S) &= S_{ij}\frac{(1-y^2)^{-1/2}}{f(y)}\Big\{\Big[E_1(I-e^{2\pi C})^{-1}e^{\sin^{-1}(y)C}E_2^{\mathsf{T}}+ \\
&E_1(I-e^{2\pi C})^{-1}e^{(\pi-\sin^{-1}(y))C}E_2^{\mathsf{T}}\Big]_{ji}\mathbf{1}_{0<y<1}+ \\
&\Big[E_1(I-e^{2\pi C})^{-1}e^{(\pi-\sin^{-1}(y))C}E_2^{\mathsf{T}}+ \\
&E_1(I-e^{2\pi C})^{-1}e^{(2\pi+\sin^{-1}(y))C}E_2^{\mathsf{T}}\Big]_{ji}\mathbf{1}_{-1<y<0}\Big\}
\end{aligned}
$$

Fisher Information

- Fisher Information Matrix(FIM) describes the amount of information that the data provide about unknown parameters

Fisher Information

- Fisher Information Matrix(FIM) describes the amount of information that the data provide about unknown parameters
- Measures the overall sensitivity of the log-likelihood functions to changes of parameters

Fisher Information

- Fisher Information Matrix(FIM) describes the amount of information that the data provide about unknown parameters
- Measures the overall sensitivity of the log-likelihood functions to changes of parameters
- Used for the testing of hypothesis and in the construction of confidence regions for the unknown parameters

Fisher Information

- Let L be the likelihood function and θ be the parameter vector. The $U = \dfrac{\partial L}{\partial \theta}$ is called the *score statistic*.

Fisher Information

- Let L be the likelihood function and θ be the parameter vector. The $U = \dfrac{\partial L}{\partial \theta}$ is called the *score statistic*.
- Score statistic is used for inference about parameter values in generalized linear models.

Fisher Information

- Let L be the likelihood function and θ be the parameter vector. The $U = \dfrac{\partial L}{\partial \theta}$ is called the *score statistic*.
- Score statistic is used for inference about parameter values in generalized linear models.
- We have, $\mathbb{E}(U) = 0$.

Fisher Information

- Let L be the likelihood function and θ be the parameter vector. The $U = \dfrac{\partial L}{\partial \theta}$ is called the *score statistic*.
- Score statistic is used for inference about parameter values in generalized linear models.
- We have, $\mathbb{E}(U) = 0$.
- Variance of U, *i.e.,* $\mathbb{E}(U^2)$ is called the Fisher Information.

Fisher Information

- Let L be the likelihood function and θ be the parameter vector. The $U = \frac{\partial L}{\partial \theta}$ is called the *score statistic*.
- Score statistic is used for inference about parameter values in generalized linear models.
- We have, $\mathbb{E}(U) = 0$.
- Variance of U, *i.e.*, $\mathbb{E}(U^2)$ is called the Fisher Information.
- The inverse of the Fisher information matrix gives the covariance matrix for the estimates of the parameters.

Fisher Information

$$\textit{The expected FIM} = E\left[\frac{\partial L(\theta;X)}{\partial \theta}\frac{\partial L(\theta;X)}{\partial \theta^T}\right] = -E\left[\frac{\partial^2 L(\theta;X)}{\partial \theta\ \partial \theta^T}\right]$$

Fisher Information

The expected FIM $= E\left[\frac{\partial L(\theta;X)}{\partial\theta}\frac{\partial L(\theta;X)}{\partial\theta^T}\right] = -E\left[\frac{\partial^2 L(\theta;X)}{\partial\theta\ \partial\theta^T}\right]$

D. Oakes. Direct calculation of the information matrix via the EM. J. R. Statis. Soc.: Series B (Statistical Methodology)*. 1999,* 61*(2), 479-482.*

$$\frac{\partial^2 L(\theta; y)}{\partial\theta^2} = \left\{\frac{\partial^2 Q(\hat{\theta}/\theta)}{\partial\hat{\theta}^2} + \frac{\partial^2 Q(\hat{\theta}/\theta)}{\partial\theta\partial\hat{\theta}}\right\}_{\hat{\theta}=\theta}$$

$$\text{where } Q(\hat{\theta}/\theta) = E_\theta\left(l_f(\hat{\theta}; x)/y\right),$$

M. Bladt, L. J. Esparza and B. F. Nielsen. Fisher information and statistical inference for phase-type distributions. J. Appl. Prob. Spec. *2011,* 48A, *277-293.*

M. Bladt, L. J. Esparza and B. F. Nielsen. Fisher information and statistical inference for phase-type distributions. J. Appl. Prob. Spec. *2011,* 48A, *277-293.*

$$\text{Let, } \theta = (\beta_1, \beta_2, \cdots \beta_{p-1}, s_1^0, S_{12}, \cdots S_{1p}, S_{21}, s_2^0, S_{23} \cdots S_{2p}, \cdots S_{p1}, S_{p2} \cdots S_{p,p-1}, s_p^0)$$

be the parameter vector of order $p - 1 + p^2$.

M. Bladt, L. J. Esparza and B. F. Nielsen. Fisher information and statistical inference for phase-type distributions. J. Appl. Prob. Spec. *2011,* 48A, *277-293.*

$$\text{Let, } \theta = (\beta_1, \beta_2, \cdots \beta_{p-1}, s_1^0, S_{12}, \cdots S_{1p}, S_{21}, s_2^0, S_{23} \cdots S_{2p}, \cdots S_{p1}, S_{p2} \cdots S_{p,p-1}, s_p^0)$$

be the parameter vector of order $p - 1 + p^2$.

Take $\beta_p = 1 - \sum_{j=1}^{p-1} \beta_j$ and $S_{ii} = -\sum_{\substack{j=1 \\ j \neq i}}^{p} S_{ij} - s_i^0$.

M. Bladt, L. J. Esparza and B. F. Nielsen. Fisher information and statistical inference for phase-type distributions. J. Appl. Prob. Spec. *2011,* 48A, *277-293.*

$$\text{Let, } \theta = (\beta_1, \beta_2, \cdots \beta_{p-1}, s_1^0, S_{12}, \cdots S_{1p}, S_{21}, s_2^0, S_{23} \cdots S_{2p}, \cdots S_{p1}, S_{p2} \cdots S_{p,p-1}, s_p^0)$$

be the parameter vector of order $p - 1 + p^2$.

Take $\beta_p = 1 - \sum_{j=1}^{p-1} \beta_j$ and $S_{ii} = -\sum_{\substack{j=1 \\ j \neq i}}^{p} S_{ij} - s_i^0$.

$$Q(\hat{\theta}/\theta) = \sum_{i=1}^{p} \log(\hat{\beta}_i) \sum_{k=1}^{n} \hat{B}_i^k + \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \sum_{k=1}^{n} \log(\hat{S}_{ij}) \hat{N}_{ij}^k - \sum_{i=1}^{p} \sum_{j=1 j \neq i}^{p} \sum_{k=1}^{n} \hat{S}_{ij} \hat{Z}_i^k + \sum_{i=1}^{p} \sum_{k=1}^{n} \log(\hat{t}_i) \hat{N}_i^k - \sum_{i=1}^{p} \sum_{k=1}^{n} \hat{s}_i^0 \hat{Z}_i^k$$

$$
\begin{aligned}
Q(\hat{\theta}/\theta) &= \sum_{i=1}^{p-1} \log(\hat{\beta}_i) U_i \beta_i + \log\left(1 - \sum_{i=1}^{p-1} \hat{\beta}_i\right) U_p (1 - \sum_{i=1}^{p-1} \beta_i) + \\
&\quad \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \log(\hat{S}_{ij}) V_{ij} S_{ij} + \sum_{i=1}^{p} T_i V_{ii} + \sum_{i=1}^{p} \log(\hat{s_i^0}) W_i s_i^0
\end{aligned}
$$

$$
\begin{aligned}
U_i &= \sum_{k=1}^{M} \frac{e_i^{\mathsf{T}} e^{S y_k} s^0}{f(y_k)} \\
W_i &= \sum_{k=1}^{M} \frac{\beta e^{S y_k} e_i}{f(y_k)}
\end{aligned}
$$

$T_i = -\sum_{\substack{j=1 \\ j \neq i}}^{p} \hat{S}_{ij} - \hat{s_i^0}$ $\quad V_{ij} = \sum_{k=1}^{M} (1/f(y_k))(e_j^{\mathsf{T}} M(y_k, \beta, S) e_i)$.

For $i, j = 1, 2, \cdots p - 1$, the $(i, j)^{th}$ element is

$$\frac{\partial U_i}{\partial \beta_j} - \frac{\partial U_p}{\partial \beta_j};$$

For $i, j = 1, 2, \cdots p-1$, the $(i,j)^{th}$ element is

$$\frac{\partial U_i}{\partial \beta_j} - \frac{\partial U_p}{\partial \beta_j};$$

for $m = 1, 2, \cdots p-1$ and $i, j = 1, 2 \cdots p$, the $(ip-1+j, m)^{th}$ element is

$$\frac{\partial U_m}{\partial S_{ij}} - \frac{\partial U_p}{\partial S_{ij}} \text{ if } i \neq j, \ \frac{\partial U_m}{\partial s_i^0} - \frac{\partial U_p}{\partial s_i^0} \text{ if } i = j;$$

For $i, j = 1, 2, \cdots p-1$, the $(i,j)^{th}$ element is

$$\frac{\partial U_i}{\partial \beta_j} - \frac{\partial U_p}{\partial \beta_j};$$

for $m = 1, 2, \cdots p-1$ and $i, j = 1, 2 \cdots p$, the $(ip-1+j, m)^{th}$ element is

$$\frac{\partial U_m}{\partial S_{ij}} - \frac{\partial U_p}{\partial S_{ij}} \text{ if } i \neq j, \ \frac{\partial U_m}{\partial s_i^0} - \frac{\partial U_p}{\partial s_i^0} \text{ if } i = j;$$

the $(m, ip-1+j)^{th}$ element is given by

$$\frac{\partial V_{ij}}{\partial \beta_m} - \frac{\partial V_{ii}}{\partial \beta_m} \text{ if } i \neq j, \ \frac{\partial W_i}{\partial \beta m} - \frac{\partial V_{ii}}{\partial \beta_m} \text{ if } i = j;$$

for $i, j, m, n = 1, 2 \cdots p$, the $(ip - 1 + j, mp - 1 + n)^{th}$ element is

$$\frac{\partial V_{ij}}{\partial S_{mn}} - \frac{\partial V_{ii}}{\partial S_{mn}} \text{ if } i \neq j, m \neq n, \qquad \frac{\partial V_{ij}}{\partial s^0_m} - \frac{\partial V_{ii}}{\partial s^0_m} \text{ if } i \neq j, m = n,$$

$$\frac{\partial W_i}{\partial S_{mn}} - \frac{\partial V_{ii}}{\partial S_{mn}} \text{ if } i = j, m \neq n, \qquad \frac{\partial W_i}{\partial s^0_m} - \frac{\partial V_{ii}}{\partial s^0_m} \text{ if } i = j, m = n.$$

Computation

For the computation of the above derivatives, put,

$$R_i(u) \quad = \quad \beta e^{Su} e_i$$

and

$$Q_i(u) \quad = \quad e_i^T e^{Su} s^0.$$

Computation

For the computation of the above derivatives, put,

$$R_i(u) \quad = \quad \beta e^{Su} e_i$$

and

$$Q_i(u) \quad = \quad e_i^T e^{Su} s^0.$$

Then we have

$$U_i = \sum_{k=1}^{M} \frac{Q_i(y_k)}{f(y_k)}, \qquad W_i = \sum_{k=1}^{M} \frac{R_i(y_k)}{f(y_k)}, \text{ and}$$

Computation

For the computation of the above derivatives, put,

$$R_i(u) \quad = \quad \beta e^{Su} e_i$$

and

$$Q_i(u) \quad = \quad e_i^T e^{Su} s^0.$$

Then we have

$$U_i = \sum_{k=1}^{M} \frac{Q_i(y_k)}{f(y_k)}, \qquad W_i = \sum_{k=1}^{M} \frac{R_i(y_k)}{f(y_k)}, \text{ and}$$

We need to compute $\frac{\partial Q_i}{\partial \theta_m}, \frac{\partial R_i}{\partial \theta_m}, \frac{\partial f}{\partial \theta_m}, \frac{\partial e^{Su}}{\partial \theta_m}$

Computation of M

We have,

$$M(y, \beta, S) = \int_{o}^{y} e^{S(y-u)} s^{0} \beta e^{Su} du.$$

Computation of M

We have,

$M(y, \beta, S) = \int\limits_{o}^{y} e^{S(y-u)} s^0 \beta e^{Su} du.$

Now using the properties of integration of matrices,

$M(y, \beta, S) = E_1 e^{Cy} E_2^{\mathsf{T}}$, where $E_i = [\mathbf{0}\ \mathbf{0}\ \ldots\ \mathbf{0}\ \boldsymbol{I}\ \mathbf{0}\ \ldots\ \mathbf{0}]$ and

$$C = \begin{bmatrix} S & s^0\beta & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & S & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & S \end{bmatrix}.$$

Computation of M

We have,

$M(y, \beta, S) = \int\limits_{o}^{y} e^{S(y-u)} s^0 \beta e^{Su} du.$

Now using the properties of integration of matrices,

$M(y, \beta, S) = E_1 e^{Cy} E_2^{\mathsf{T}}$, where $E_i = [\mathbf{0}\ \mathbf{0}\ \ldots\ \mathbf{0}\ \mathbf{I}\ \mathbf{0}\ \ldots\ \mathbf{0}]$ and

$$C = \begin{bmatrix} S & s^0\beta & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & S & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & S \end{bmatrix}.$$

For the computation of $\frac{\partial V_{ij}}{\partial \theta_m}$, we need $\dfrac{\partial e^{Cy}}{\partial \theta_m}$

By uniformization,

$$\frac{\partial e^{Su}}{\partial \theta_m} = \sum_{r=1}^{\infty} b_r \frac{\partial K^r}{\partial \theta_m} + \frac{\partial c}{\partial \theta_m} u e^{Su} (K - I)$$

where, $c = Max\{-S_{ii} : 1 \leq i \leq p\}$ and $K = \frac{1}{c} S + I$. For $r \geq 1$,

$$\frac{\partial K^r}{\partial \theta_m} = \sum_{l=0}^{r-1} \left[K^l \frac{\partial K}{\partial \theta_m} K^{r-1-l} \right]$$

and

$$\frac{\partial K}{\partial \theta_m} = \frac{1}{c} \frac{\partial S}{\partial \theta_m} - \frac{1}{c^2} \frac{\partial c}{\partial \theta_m} S$$

By uniformization,

$$\frac{\partial e^{Su}}{\partial \theta_m} = \sum_{r=1}^{\infty} b_r \frac{\partial K^r}{\partial \theta_m} + \frac{\partial c}{\partial \theta_m} u e^{Su} (K - I)$$

where, $c = Max\{-S_{ii} : 1 \leq i \leq p\}$ and $K = \frac{1}{c}S + I$. For $r \geq 1$,

$$\frac{\partial K^r}{\partial \theta_m} = \sum_{l=0}^{r-1} \left[K^l \frac{\partial K}{\partial \theta_m} K^{r-1-l} \right]$$

and

$$\frac{\partial K}{\partial \theta_m} = \frac{1}{c} \frac{\partial S}{\partial \theta_m} - \frac{1}{c^2} \frac{\partial c}{\partial \theta_m} S$$

Assume that the maximum of the diagonal of $-S$ is appeared in row l

$$\frac{\partial c}{\partial S_{ij}} = \begin{cases} 0 & \text{if} \quad i \neq l,\ j \neq i \\ 1 & \text{if} \quad i = l,\ j \neq i \end{cases}$$

and

$$\frac{\partial c}{\partial s_i^0} = \begin{cases} 0 & \text{if} \quad i \neq l \\ 1 & \text{if} \quad i = l \end{cases}$$

$$\frac{\partial c}{\partial S_{ij}} = \begin{cases} 0 & if \quad i \neq l,\ j \neq i \\ 1 & if \quad i = l,\ j \neq i \end{cases}$$

and

$$\frac{\partial c}{\partial s_i^0} = \begin{cases} 0 & if \quad i \neq l \\ 1 & if \quad i = l \end{cases}$$

For $i \neq j$

$$\left(\frac{\partial S}{\partial S_{ij}}\right)_{(r,s)} = \begin{cases} 0 & if \quad i \neq r \\ -1 & if \quad i = r,\ j \neq s \\ 1 & if \quad i = r,\ j = s \end{cases}$$

and $\left(\frac{\partial S}{\partial s_i^0}\right)_{(i,j)} = \begin{cases} -1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$

$$
\begin{aligned}
\frac{\partial f(y)}{\partial \beta_m} &= (e_m^T - e_p^T) e^{Sy} s^0 \\
\frac{\partial f(y)}{\partial S_{ij}} &= \beta \left(\frac{\partial e^{Sy}}{\partial S_{ij}} s^0 + e^{Sy} \frac{\partial s^0}{\partial S_{ij}} \right) \\
\frac{\partial f(y)}{\partial s_m^0} &= \beta e^{Sy} e_m + \beta \frac{\partial e^{Sy}}{\partial s_m^0} s^0
\end{aligned}
$$

Case 1

In this case, the FIM of $Y = g(X)$ will be same as the FIM of $g^{-1}(Y)$, which is a *PH* variate.

Case 1

In this case, the FIM of $Y = g(X)$ will be same as the FIM of $g^{-1}(Y)$, which is a *PH* variate.

If $y_1, y_2, \ldots, y_M$ be a sample from Y, then take the inverse of each observation and compute the *FIM* of the PH variate using the sample $g^{-1}(y_1), g^{-1}(y_2), \ldots, g^{-1}(y_M)$.

Case 2

Let $y_1, y_2, \ldots, y_M$ be a sample from Y. Then,

$$
\begin{aligned}
U_i &= \sum_{l=1}^{M} \sum_{k=1}^{n_{y_l}} |g'(x_k(y_l))|^{-1} \frac{e_i^T e^{S x_k(y_l)} s^0}{h(y_l)} \\
W_i &= \sum_{l=1}^{M} \sum_{k=1}^{n_{y_l}} |g'(x_k(y_l))|^{-1} \frac{\beta e^{S x_k(y_l)} e_i}{h(y_l)} \\
T_i &= -\sum_{\substack{j=1 \\ j \neq i}}^{p} \hat{S}_{ij} - \hat{s}_i^0 \\
\text{and } V_{ij} &= \sum_{l=1}^{M} \sum_{k=1}^{n_{y_l}} \frac{|g'(x_k(y_l))|^{-1}}{h(y_l)} e_j^T M(x_k(y_l), \beta, S) e_i.
\end{aligned}
$$

Case 3

We have, $M^*(y, \beta, S) = \sum_{n=0}^{\infty} M(h(y) + kn - m, \beta, S)$

Case 3

We have, $M^*(y, \beta, S) = \sum_{n=0}^{\infty} M(h(y) + kn - m, \beta, S)$
Which can be computed as
$M^*(y, \beta, S) = E_1 e^{C(h(y)-m)}(I - e^{Ck})^{-1} E_2^T$.

Case 3

We have, $M^*(y,\beta,S) = \sum_{n=0}^{\infty} M(h(y)+kn-m,\beta,S)$

Which can be computed as

$M^*(y,\beta,S) = E_1 e^{C(h(y)-m)}(I-e^{Ck})^{-1}E_2^T$.

Also, $M^*(y,\beta,S) =$

$(I-e^{Sk})^{-1}M(h(y)-m,\beta,S)+(I-e^{Sk})^{-1}M(k,\beta,S)(I-e^{Sk})^{-1}e^{S(h(y)-m})$

$$
\begin{aligned}
U_i &= \sum_{l=1}^{n} \frac{|h^{'}(y)|e_i^{\mathsf{T}} e^{S(h(y_l)-m)}(I - e^{Sk})^{-1} s^0}{f(y_l)} \\
W_i &= \sum_{l=1}^{n} \frac{|h^{'}(y)|\beta e^{S(h(y)-m)}(I - e^{Sk})^{-1} e_i}{f(y_l)} \\
T_i &= -\sum_{j=1,\, j\neq i}^{p} \hat{S}_{ij} - \hat{s}_i^0 \\
V_{ij} &= \sum_{l=1}^{n} \frac{|h^{'}(y)|e_j^{\mathsf{T}} M^*(y_l, \beta, S) e_i}{f(y_l)}
\end{aligned}
$$

Put

$$R_i(u) = \beta e^{Su}(I - e^{Sk})^{-1} e_i$$

Put

$$R_i(u) = \beta e^{Su}(I - e^{Sk})^{-1} e_i$$

and

$$Q_i(u) = e_i^T e^{Su}(I - e^{Sk})^{-1} s^0$$

Put

$$R_i(u) = \beta e^{Su}(I - e^{Sk})^{-1} e_i$$

and

$$Q_i(u) = e_i^T e^{Su}(I - e^{Sk})^{-1} s^0$$

Hence we get,

$$\begin{aligned} U_i &= \sum_{l=1}^{n} \frac{|h^{'}(y)|}{f(y_l)} Q_i(h(y_l) - m) \\ W_i &= \sum_{l=1}^{n} \frac{|h^{'}(y)|}{f(y_l)} R_i(h(y_l) - m) \end{aligned}$$

Put

$$R_i(u) = \beta e^{Su}(I - e^{Sk})^{-1} e_i$$

and

$$Q_i(u) = e_i^T e^{Su}(I - e^{Sk})^{-1} s^0$$

Hence we get,

$$U_i = \sum_{l=1}^{n} \frac{|h'(y)|}{f(y_l)} Q_i(h(y_l) - m)$$

$$W_i = \sum_{l=1}^{n} \frac{|h'(y)|}{f(y_l)} R_i(h(y_l) - m)$$

and

$$V_{ij} = \sum_{l=1}^{n} \frac{|h'(y)|}{f(y_l)} E_1 e^{C(h(y)-m)}(I - e^{kC})^{-1} E_2^T e_i$$

For $q \in \{1, 2, \ldots, p-1+p^2\}$,

$$\frac{\partial U_i}{\partial \theta_q} = \sum_{l=1}^{n} \frac{|h'(y)|}{(f(y_l))^2} \left[f(y_l) \frac{\partial Q_i(h(y_l)-m)}{\partial \theta_q} - (Q_i h(y_l) - m) \frac{\partial f(y_l)}{\partial \theta_q} \right]$$

$$\frac{\partial W_i}{\partial \theta_q} = \sum_{l=1}^{n} \frac{|h'(y)|}{(f(y_l))^2} \left[f(y_l) \frac{\partial R_i(h(y_l)-m)}{\partial \theta_q} - (R_i h(y_l) - m) \frac{\partial f(y_l)}{\partial \theta_q} \right]$$

$$\frac{\partial V_{ij}}{\partial \theta_q} = \sum_{l=1}^{n} \frac{|h'(y)|}{(f(y_l))^2} \left\{ f(y_l) E_1 \left[\frac{\partial}{\partial \theta_q} (e^{C(h(y)-m)})(I - e^{kC})^{-1} + e^{C(h(y)-m)} \frac{\partial}{\partial \theta_q} (I - e^{kC})^{-1} \right] E_2^{\mathsf{T}} e_i + \frac{\partial f(y_l)}{\partial \theta_q} E_1 e^{C(h(y)-m)} (I - e^{kC})^{-1} E_2^{\mathsf{T}} e_i \right\}$$

Differentiating the *pdf*, we get,

$$\frac{\partial f}{\partial \beta_q} = |h^{'}(y)| \left[e_q^{\mathsf{T}} e^{S(h(y)-m)} (I - e^{Sk})^{-1} s^0 - e_p^{\mathsf{T}} e^{S(h(y)-m)} (I - e^{Sk})^{-1} s^0 \right]$$

$$\frac{\partial f}{\partial S_{ij}} = |h^{'}(y)| \left[\beta \frac{\partial}{\partial S_{ij}} (e^{S(h(y)-m)}) (I - e^{Sk})^{-1} s^0 + \beta e^{S(h(y)-m)} \frac{\partial}{\partial S_{ij}} (I - e^{Sk})^{-1} s^0 \right]$$

and $$\frac{\partial f}{\partial s_q^0} = |h^{'}(y)| \left[\beta e^{S(h(y)-m)} (I - e^{Sk})^{-1} e_q + \beta \frac{\partial}{\partial s_q^0} (e^{S(h(y)-m)}) (I - e^{Sk})^{-1} s^0 + \beta e^{S(h(y)-m)} \frac{\partial}{\partial s_q^0} (I - e^{Sk})^{-1} s^0 \right].$$

Thus for evaluating the above derivatives we need the terms $\frac{\partial}{\partial S_{ij}}(I - e^{Sk})^{-1}$ and $\frac{\partial}{\partial s_q^0}(I - e^{Sk})^{-1}$.

Thus for evaluating the above derivatives we need the terms $\frac{\partial}{\partial S_{ij}}(I - e^{Sk})^{-1}$ and $\frac{\partial}{\partial s_q^0}(I - e^{Sk})^{-1}$. By uniformization,

$$\begin{aligned}
\frac{\partial}{\partial \theta_q}(I - e^{Sk})^{-1} &= \sum_{n=1}^{\infty} \frac{\partial e^{nkS}}{\partial \theta_q} \\
&= \sum_{n=1}^{\infty}\sum_{r=1}^{\infty} b_{r,nk} \frac{\partial K^r}{\partial \theta_q} + \sum_{n=1}^{\infty} \frac{\partial c}{\partial \theta_q} nke^{nkS}(K - I).
\end{aligned}$$

Thus for evaluating the above derivatives we need the terms $\frac{\partial}{\partial S_{ij}}(I - e^{Sk})^{-1}$ and $\frac{\partial}{\partial s_q^0}(I - e^{Sk})^{-1}$. By uniformization,

$$\begin{aligned}\frac{\partial}{\partial \theta_q}(I - e^{Sk})^{-1} &= \sum_{n=1}^{\infty} \frac{\partial e^{nkS}}{\partial \theta_q} \\ &= \sum_{n=1}^{\infty}\sum_{r=1}^{\infty} b_{r,nk}\frac{\partial K^r}{\partial \theta_q} + \sum_{n=1}^{\infty}\frac{\partial c}{\partial \theta_q} nke^{nkS}(K - I).\end{aligned}$$

Hence,

$$\frac{\partial}{\partial \theta_q}(I - e^{Sk})^{-1} = \sum_{n=1}^{\infty}\sum_{r=1}^{\infty} b_{r,kn}\frac{\partial K^r}{\partial \theta_q} + \frac{\partial c}{\partial \theta_q} e^{Sk}(I - e^{Sk})^{-2}(K - I).$$

FSPH

10000 observations are taken from the beta $B(0.5, 5)$ distribution.

FSPH

10000 observations are taken from the beta $B(0.5, 5)$ distribution. Fit this data with a $FSPH_1(\alpha, T)$ variate of order 2 and the approximate ML estimates are

$$\alpha = \begin{bmatrix} 0.5531 & 0.4469 \end{bmatrix}$$

FSPH

10000 observations are taken from the beta $B(0.5, 5)$ distribution. Fit this data with a $FSPH_1(\alpha, T)$ variate of order 2 and the approximate ML estimates are

$$\alpha = \begin{bmatrix} 0.5531 & 0.4469 \end{bmatrix}$$

and

$$T = \begin{bmatrix} -9.6181 & 1.0762 \\ 12.6514 & -66.7070 \end{bmatrix}.$$

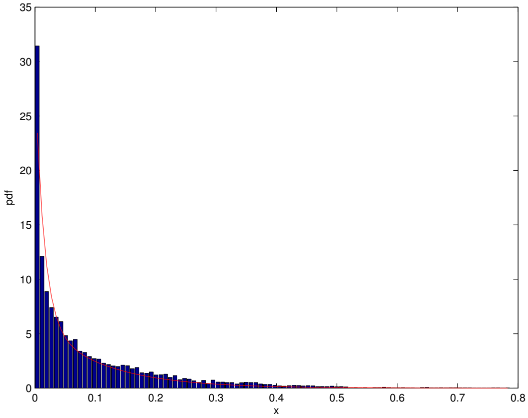

Figure : $B(0.5, 5)$ fitted with $FSPH_1(\alpha, T)$ of order 2.

Table : Correlations for $B(0.5, 5)$ fitted with $FSPH_1$.

Parameter	$\hat{\alpha}_1$	$\hat{t}_1$	$\hat{T}_{12}$	$\hat{T}_{21}$	$\hat{t}_2$
$\hat{\alpha}_1$	1.0000	-0.1879	0.3353	0.5468	-0.1594
$\hat{t}_1$	-0.1879	1.0000	0.9344	-0.9101	-0.5402
$\hat{T}_{12}$	o.3353	0.9344	1.0000	0.3193	0.1711
$\hat{T}_{21}$	0.5468	-0.9101	0.3193	1.0000	0.9228
$\hat{t}_2$	-0.1594	-0.5402	0.1711	0.9228	1.0000


Thank You!