Phase Type Distributions
2022-10-10
BFN/bfn

Consider the setting of Sections 3.4.2 and 3.7 The Markov chain $\{X_n; n \geq 0\}$ with one step transition probabilities

$$\boldsymbol{P} = \left\| \begin{array}{cc} \boldsymbol{Q} & \boldsymbol{R} \\ \boldsymbol{0} & \boldsymbol{I} \end{array} \right\|.$$

Where $\boldsymbol{Q}$ is $r \times r$, $\boldsymbol{I} - \boldsymbol{Q}$ is non singular, such that all r states are transient.

Define $Y = \min_{n \geq 0}\{X_n \geq r\}$ to be the time of absorption in one of the absorbing states.

Let $\alpha_i = \mathbb{P}\{X_0 = i\}, \quad i = 0, 1, \ldots, r, \boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_{r-1})$.

Then

$$\begin{aligned} \mathbb{P}\{Y > y\} &= \mathbb{P}\{X_y < r\} = \boldsymbol{\alpha}\boldsymbol{Q}^y\boldsymbol{e} \\ \mathbb{P}\{Y = y\} &= \boldsymbol{\alpha}\boldsymbol{Q}^{y-1}\boldsymbol{R}\boldsymbol{e} \end{aligned}$$

The first expression gives the probability of the Markov chain being in one of the transient states at time y (not yet absorbed)

The second expression gives the probability of absorption in (exactly) time y

If our focus is primarily on the absorption time,

$$\boldsymbol{P} = \left\| \begin{array}{cc} \boldsymbol{Q} & \boldsymbol{r} \\ \boldsymbol{0} & 1 \end{array} \right\|.$$

with $\boldsymbol{r} = \boldsymbol{R}\boldsymbol{e} = \boldsymbol{e} - \boldsymbol{Q}\boldsymbol{e}$

(Discrete) phase type distribution

we reparameterise to get

$$\boldsymbol{P} = \left\| \begin{array}{cc} \boldsymbol{S} & \boldsymbol{s} \\ \boldsymbol{0} & 1 \end{array} \right\|.$$

with initial probability distribution $(\boldsymbol{\alpha}, \alpha_r), \quad \alpha_r = 1 - \boldsymbol{\alpha}\boldsymbol{e}$. We frequently assume $\alpha_r = 0$ $(\boldsymbol{\alpha}\boldsymbol{e} = 1)$.

We write $Y \sim \mathrm{PH}(\boldsymbol{\alpha}, \boldsymbol{S})$: a representation

$$\begin{aligned}
\mathbb{P}\{Y = y\} &= \boldsymbol{\alpha}\boldsymbol{S}^{y-1}\boldsymbol{s}, \quad \text{Probability mass function (density)} \\
\mathbb{P}\{Y > y\} &= \boldsymbol{\alpha}\boldsymbol{S}^{y}\boldsymbol{e}, \quad \text{Survival function} \\
\mathbb{P}\{Y \le y\} &= 1 - \boldsymbol{\alpha}\boldsymbol{S}^{y}\boldsymbol{e}, \quad \text{(Cumulative) distribution function}
\end{aligned}$$

From Chapter 3:

$$\mathbb{E}(Y) = \boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{e}$$

Higher order moments

Probability generating function

$$\begin{aligned}
\phi(\theta) &= \mathbb{E}\left(\theta^{Y}\right) = \sum_{y=0}^{\infty} \theta^{y}\mathbb{P}\{Y = y\} = \alpha_r + \sum_{y=1}^{\infty} \theta^{y}\boldsymbol{\alpha}\boldsymbol{S}^{y-1}\boldsymbol{s} = \alpha_r + \boldsymbol{\alpha}\left[\sum_{y=1}^{\infty} \theta^{y}\boldsymbol{S}^{y-1}\right]\boldsymbol{s} \\
&= \alpha_r + \theta\boldsymbol{\alpha}\left[\sum_{y=1}^{\infty}(\theta\boldsymbol{S})^{y-1}\right]\boldsymbol{s} = \alpha_r + \theta\boldsymbol{\alpha}\left[\sum_{i=0}^{\infty}(\theta\boldsymbol{S})^{i}\right]\boldsymbol{s} = \alpha_r + \theta\boldsymbol{\alpha}(\boldsymbol{I} - \theta\boldsymbol{S})^{-1}\boldsymbol{s}
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}\phi(\theta)}{\mathrm{d}\theta} &= \boldsymbol{\alpha}\left[\sum_{y=1}^{\infty} y\theta^{y-1}\boldsymbol{S}^{y-1}\right]\boldsymbol{s} = \boldsymbol{\alpha}\left[\sum_{y=1}^{\infty}\left(\sum_{k=1}^{y} 1\right)(\theta\boldsymbol{S})^{y-1}\right]\boldsymbol{s} = \boldsymbol{\alpha}\left[\sum_{y=1}^{\infty}\sum_{k=1}^{y}(\theta\boldsymbol{S})^{y-1}\right]\boldsymbol{s} \\
&= \boldsymbol{\alpha}\left[\sum_{k=1}^{\infty}\sum_{y=k}^{\infty}(\theta\boldsymbol{S})^{y-1}\right]\boldsymbol{s} = \boldsymbol{\alpha}\left[\sum_{k=1}^{\infty}\sum_{y=k}^{\infty}(\theta\boldsymbol{S})^{k-1}(\theta\boldsymbol{S})^{y-k}\right]\boldsymbol{s} \\
&= \boldsymbol{\alpha}\left[\sum_{k=1}^{\infty}(\theta\boldsymbol{S})^{k-1}\sum_{y=k}^{\infty}(\theta\boldsymbol{S})^{y-k}\right]\boldsymbol{s} = \boldsymbol{\alpha}\left[\sum_{\ell=0}^{\infty}(\theta\boldsymbol{S})^{\ell}\sum_{z=0}^{\infty}(\theta\boldsymbol{S})^{z}\right]\boldsymbol{s} \\
&= \boldsymbol{\alpha}(\boldsymbol{I} - \theta\boldsymbol{S})^{-2}\boldsymbol{s}
\end{aligned}$$

For $\theta = 1$ we get

$$\mathbb{E}(Y) = \boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-2}\boldsymbol{s} = \boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{e}$$

using

$$\boldsymbol{S}\boldsymbol{e} + \boldsymbol{s} = \boldsymbol{e} \Leftrightarrow \boldsymbol{s} = (\boldsymbol{I} - \boldsymbol{S})\boldsymbol{e}$$

Taking further derivatives we get

$$\begin{aligned}
\frac{\mathrm{d}\phi(\theta)}{\mathrm{d}\theta} &= \boldsymbol{\alpha}\boldsymbol{S}^{k-1}(\boldsymbol{I}-\theta\boldsymbol{S})^{-k-1}\boldsymbol{s} \\
\mathbb{E}\left(\prod_{i=0}^{k-1}(X-i)\right) &= \boldsymbol{\alpha}\boldsymbol{S}^{k-1}(\boldsymbol{I}-\boldsymbol{S})^{-k}\mathbf{1}
\end{aligned}$$

Continuous time phase type distributions

Consider a Markov jump process $\{J(t); t \geq 0\}$ with generator

$$\boldsymbol{A} = \left\| \begin{array}{cc} \boldsymbol{S} & \boldsymbol{s} \\ \mathbf{0} & 0 \end{array} \right\|$$

$$\begin{aligned}
\mathbb{P}\{X > x\} &= \boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{e} \\
\mathbb{P}\{x \leq X \leq x+h\} &= \boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{s}h \cong f_X(x)h \\
f_X(x) &= \boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{s} \\
\mathbb{P}\{X \leq x\} &= 1-\boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{e} \\
L_X(\theta) &= \mathbb{E}\left(e^{-\theta X}\right) = \alpha_r + \int_0^\infty e^{-\theta x}\boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{s}\mathrm{d}x = \alpha_r + \boldsymbol{\alpha}\left[\int_0^\infty e^{-\theta x} e^{\boldsymbol{S}x}\mathrm{d}x\right]\boldsymbol{s} \\
&= \alpha_r + \boldsymbol{\alpha}\int_0^\infty e^{-(\theta\boldsymbol{I}-\boldsymbol{S})x}\mathrm{d}x\boldsymbol{s} = \alpha_r + \boldsymbol{\alpha}(\theta\boldsymbol{I}-\boldsymbol{S})^{-1}\boldsymbol{s} \\
\frac{\mathrm{d}^n L_X(\theta)}{\mathrm{d}\theta^n} &= (-1)^n n!\boldsymbol{\alpha}(\theta\boldsymbol{I}-\boldsymbol{S})^{-n-1}\boldsymbol{s} \\
\mathbb{E}\left(X^n\right) &= n!\boldsymbol{\alpha}(-\boldsymbol{S})^{-n}\boldsymbol{e}, \quad \text{using} \\
\boldsymbol{S}\boldsymbol{e}+\boldsymbol{s} &= 0 \Leftrightarrow \boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}
\end{aligned}$$

Simplest (nearly trivial) example

$$\begin{aligned}
\boldsymbol{A} &= \left\| \begin{array}{cc} -\lambda & \lambda \\ 0 & 0 \end{array} \right\| \\
(\boldsymbol{\alpha}, \boldsymbol{S}) &= ((1), ||-\lambda||)
\end{aligned}$$

$X \sim \mathrm{PH}\left((1), ||-\lambda||\right)$ or $X \sim \exp\left(\lambda\right)$

Probabilistic derivation of mean

W_j time spent in j before absorption

$$
\begin{aligned}
\mathbb{E}(W_j|J(0)=i) &= \mathbb{E}\left(\int_0^X \mathbf{1}\{J(t)=j\}|J(0)=i)\mathrm{d}t\right) = \mathbb{E}\left(\int_0^\infty \mathbf{1}\{J(t)=j, X>t\}|J(0)=i)\mathrm{d}t\right) \\
&= \int_0^\infty P_{ij}(t)\mathrm{d}t = \int_0^\infty \left[e^{\boldsymbol{S}t}\right]_{i,j} \mathrm{d}t \\
\boldsymbol{U} &= \int_0^\infty e^{\boldsymbol{S}t}\mathrm{d}t = (-\boldsymbol{S})^{-1} \\
\mathbb{E}(X) &= \boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{e} = \boldsymbol{\alpha}(-\boldsymbol{S})^{-1}\boldsymbol{e}
\end{aligned}
$$

Operations with phase type distributions

Phase type distributions are closed under a number of operations. The results can be proven probabilistically (as well as analytically)

Sums of independent PH variables

Suppose $X_i \sim \mathrm{PH}(\boldsymbol{\alpha}_i, \boldsymbol{S}_i)$ independent (discrete or continuous)

The distribution of $X = X_1 + X_2$

First assume both exponential (geometric)

$$
\boldsymbol{A} = \left\| \begin{array}{cc|c} -\lambda_1 & \lambda_1 & 0 \\ 0 & -\lambda_2 & \lambda_2 \\ \hline 0 & 0 & 0 \end{array} \right\|
$$

Two life stages.

Generally $X_1 \sim \mathrm{PH}(\boldsymbol{\alpha}, \boldsymbol{S})$ with $\{J_1(t); t \geq 0\}$ and, $X_2 \sim \mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{T})$ with $\{J_2(t); t \geq 0\}$. Define

$$J(t) = \begin{cases} J_1(t) & t < X_1 \\ J_2(t - X_1) & X_1 \leq t \end{cases}$$

State changes between "X_1" states: $S_{ij}(\mathrm{d}t)$
State changes between "X_2" states: $T_{ij}(\mathrm{d}t)$
State changes from "X_1" states to "X_2" states: $s_i(\mathrm{d}t)\beta_j$

$$\boldsymbol{A} = \left\| \begin{array}{cc|c} \boldsymbol{S}_1 & \boldsymbol{s}_1\boldsymbol{\beta} & 0 \\ 0 & \boldsymbol{S}_2 & \boldsymbol{s}_2 \\ \hline \mathbf{0} & \mathbf{0} & 0 \end{array} \right\|$$

So $X \sim \mathrm{PH}(\boldsymbol{\gamma}, \boldsymbol{L})$ with

$$(\boldsymbol{\gamma}, \boldsymbol{L}) = \left((\boldsymbol{\alpha}, \alpha_{r_1}\boldsymbol{\beta}), \left\| \begin{array}{cc} \boldsymbol{S} & \boldsymbol{s}_1\boldsymbol{\beta} \\ 0 & \boldsymbol{T} \end{array} \right\| \right)$$

Alternative analytic proof via generating function/Laplace transform. By induction the result holds for finite sums.

Example Erlang distributions

$$(\boldsymbol{\alpha}, \boldsymbol{S}) = \left((1, 0, 0, \ldots, 0), \left\| \begin{array}{ccccccc} -\lambda & \lambda & 0 & \ldots & 0 & 0 \\ 0 & -\lambda & \lambda & \ldots & 0 & 0 \\ 0 & 0 & -\lambda & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots\ddots\ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -\lambda & \lambda \\ 0 & 0 & 0 & \ldots & 0 & -\lambda \end{array} \right\| \right)$$

Example generalized Erlang distributions

$$(\boldsymbol{\alpha}, \boldsymbol{S}) = \left((1,0,0,\dots,0), \left|\left| \begin{array}{cccccc} -\lambda_1 & \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & \dots & 0 & 0 \\ 0 & 0 & -\lambda_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots\ddots\ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_{n-1} & \lambda_{n-1} \\ 0 & 0 & 0 & \dots & 0 & -\lambda_n \end{array} \right|\right| \right)$$

Mixture of (independent) PH variables

As before $X_i \sim \mathrm{PH}(\boldsymbol{\alpha}_i, \boldsymbol{S}_i)$ $X = IX_1 + (1-I)X_2$ with I indicator $\mathbb{E}(I) = p$ independent of X_i.

$X \sim \mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{T})$

$$\begin{aligned} (\boldsymbol{\beta}, \boldsymbol{T}) &= \left((p\boldsymbol{\alpha}_1, (1-p)\boldsymbol{\alpha}_2)), \left|\left| \begin{array}{cc} \boldsymbol{S}_1 & \boldsymbol{0} \\ 0 & \boldsymbol{S}_2 \end{array} \right|\right| \right) \\ f_X(x) &= p_1\boldsymbol{\alpha}_1 e^{\boldsymbol{S}_1 x}\boldsymbol{s}_1 + p_2\boldsymbol{\alpha}_2 e^{\boldsymbol{S}_2 x}\boldsymbol{s}_2 \end{aligned}$$

By induction the result holds for finite mixtures

Example hyper exponential distributions

$$(\boldsymbol{\alpha}, \boldsymbol{S}) = \left((p_1, p_2, p_3, \dots, p_n), \left|\left| \begin{array}{cccccc} -\lambda_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -\lambda_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & -\lambda_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots\ddots\ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_{n-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & -\lambda_n \end{array} \right|\right| \right)$$

Order statistics of independent PH variables

Consider first $X = \min(X_1, X_2)$, with $X_i \sim \exp(\lambda_i)$ (similar/same argument for$X_i \sim \mathrm{geo}(p_i)$)

Similarity with Markov jump process (continuous time Markov chain) - race between two exponentials $X \sim \exp(\lambda_1 + \lambda_2)$

Minimum of two discrete PH distributions

We need to simultaneously keep track of the state in both chains.

Denote the states of X_1 as $\{1,2\}$ and that of X_2 as $\{a,b\}$.

$$\boldsymbol{S}_1 = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \left\| \begin{array}{cc} p_{11} & p_{12} \\ p_{21} & p_{22} \end{array} \right\| \end{array} \quad \boldsymbol{S}_2 = \begin{array}{c} \\ a \\ b \end{array} \begin{array}{c} \begin{array}{cc} a & b \end{array} \\ \left\| \begin{array}{cc} q_{11} & q_{12} \\ q_{21} & q_{22} \end{array} \right\| \end{array}$$

$$\boldsymbol{T} = \begin{array}{c} \\ 1,a \\ 1,b \\ 2,a \\ 2,b \end{array} \begin{array}{c} \begin{array}{cccc} 1,a & 1,b & 2,a & 2,b \end{array} \\ \left\| \begin{array}{cccc} p_{11}q_{11} & p_{11}q_{12} & p_{12}q_{11} & p_{12}q_{12} \\ p_{11}q_{21} & p_{11}q_{22} & p_{12}q_{21} & p_{12}q_{22} \\ p_{21}q_{11} & p_{21}q_{12} & p_{22}q_{11} & p_{22}q_{12} \\ p_{21}q_{21} & p_{21}q_{22} & p_{22}q_{21} & p_{22}q_{22} \end{array} \right\| \end{array} = \left\| \begin{array}{cc} p_{11}\boldsymbol{S}_2 & p_{12}\boldsymbol{S}_2 \\ p_{21}\boldsymbol{S}_2 & p_{22}\boldsymbol{S}_2 \end{array} \right\| = \boldsymbol{S}_1 \otimes \boldsymbol{S}_2, \quad \text{(Kronecker product)}$$

In general for matrices $\boldsymbol{A}$ $(n \times n)$ and $\boldsymbol{B}$.

$$\boldsymbol{A} \otimes \boldsymbol{B} = = \left\| \begin{array}{cccc} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \dots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \dots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & \ddots\ddots\ddots & \vdots \\ a_{n1}\boldsymbol{B} & a_{n2}\boldsymbol{B} & \dots & a_{nn}\boldsymbol{B} \end{array} \right\|$$

In summary $X = \min(X_1, X_2)$, $X_i \sim \mathrm{PH}(\boldsymbol{\alpha}_i, \boldsymbol{S}_i)$, then $X \sim \mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{T})$ with

$$\begin{aligned} \boldsymbol{\beta} &= \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 \\ \boldsymbol{T} &= \boldsymbol{S}_1 \otimes \boldsymbol{S}_2, \quad \text{(discrete)} \\ \boldsymbol{T} &= \boldsymbol{S}_1 \oplus \boldsymbol{S}_2 = \boldsymbol{S}_1 \otimes \boldsymbol{I} + \boldsymbol{I} \otimes S_2, \quad \text{(continuous)} \end{aligned}$$

$X = \max(X_1, X_2)$, $X_i \sim \mathrm{PH}(\boldsymbol{\alpha}_i, \boldsymbol{S}_i)$, then $X \sim \mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{T})$ with

$$\begin{aligned} \boldsymbol{\beta} &= (\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2, \alpha_{2,r_2}\boldsymbol{\alpha}_1, \alpha_{1,r_1}\boldsymbol{\alpha}_2) \\ \boldsymbol{T} &= \left\| \begin{array}{ccc} \boldsymbol{S}_1 \otimes \boldsymbol{S}_2 & \boldsymbol{S}_1 \otimes \boldsymbol{s}_2 & \boldsymbol{s}_1 \otimes \boldsymbol{S}_2 \\ \boldsymbol{0} & \boldsymbol{S}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_2 \end{array} \right\|, \quad \text{(discrete)} \\ \boldsymbol{T} &= \left\| \begin{array}{ccc} \boldsymbol{S}_1 \oplus \boldsymbol{S}_2 & \boldsymbol{I} \otimes \boldsymbol{s}_2 & \boldsymbol{s}_1 \otimes \boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{S}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_2 \end{array} \right\|, \quad \text{(continuous)} \end{aligned}$$

same principle but more involved for general order statistics - see Bladt and Nielsen

Random sums

$X_i \sim \mathrm{PH}(\boldsymbol{\alpha}, \boldsymbol{S})$, $N \sim \mathrm{PH}(\boldsymbol{\gamma}, \boldsymbol{K})$ N discrete. $X = \sum_{i=1}^{N} X_i$.

Again we need a state space that is the product space of the generic space of the X_i's and N.

We see $X \sim \mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{T})$ with

$$\begin{aligned} \boldsymbol{\beta} &= \boldsymbol{\gamma} \otimes \boldsymbol{\alpha} \\ \boldsymbol{T} &= \boldsymbol{S} \otimes \boldsymbol{I} + \boldsymbol{s}\boldsymbol{\alpha} \otimes \boldsymbol{K} \end{aligned}$$

Example $X_i \sim \exp(\lambda), N \sim \mathrm{geo}(p)$ to get $\boldsymbol{T} = -\lambda \cdot 1 + \lambda \cdot (1-p) = -\lambda p$, so X is exponential.