# Geometric processes and its extensions

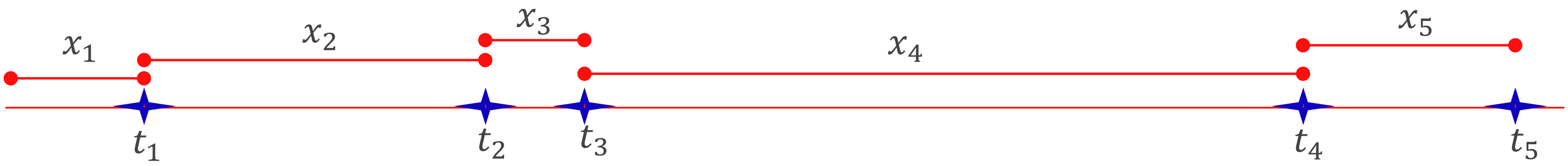
Prof Shaomin Wu (s.m.wu@kent.ac.uk)
Kent Business School
University of Kent

# Outline

- Background
- Literature review
- New models and their performance
- Future research

# Recurrent event data analysis

- Analysis of recurrent events data:



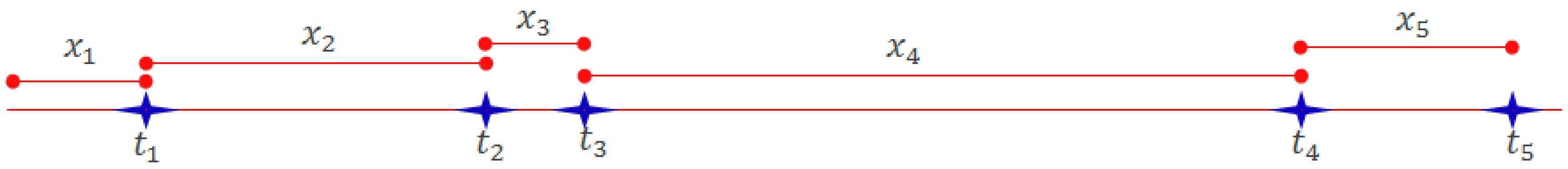

- Applications
  - modelling of insurance claims/warranty claims: times between claims
  - modelling of the outbreak of an epidemic disease: the number of cases
  - modelling of time between failures of technical systems (software or hardware)

Challenges

1. Common difficulties in reliability engineering:
   - Lacks of failures, technical systems are normally reliable and do not have many failure data
   - Censoring:  when the observation period ends, not all units have failed - some are survivors
2. Recurrent event data analysis is also widely studied in the healthcare sector, in which
   - the effect of covariates is the focus and
   - they have more data available for modelling

# Recurrent event data analysis

- Recurrent events:




- Notations:

$$S_n = \sum_{i=1}^{n} X_i$$

$$Y_t = \sum_{n=1}^{\infty} \chi\{S_n \le t\} = \sup\{n: S_n \le t\}$$

$$m(t) = E[Y_t]$$

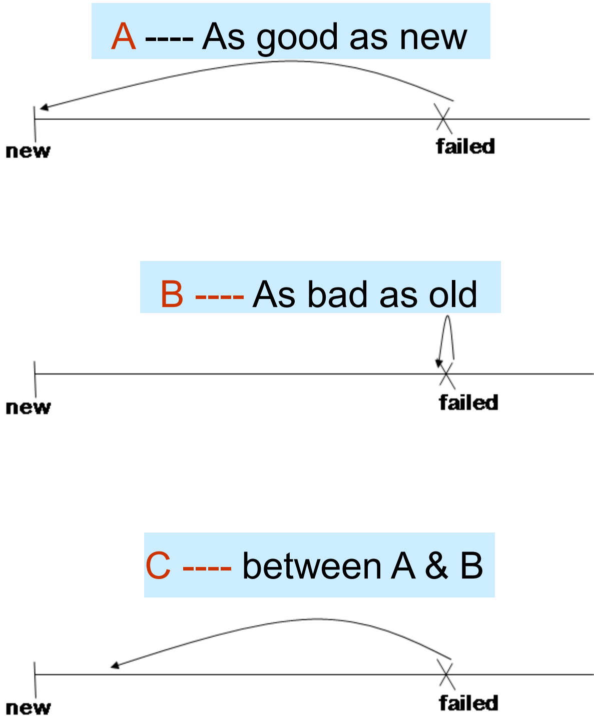
Questions in reliability mathematics:

1. What are the distributions of the gap times, i.e., the distributions of $X_i$?
2. How many events occurred within a given time, i.e., How can we estimate $m(t)$?

# Literature review

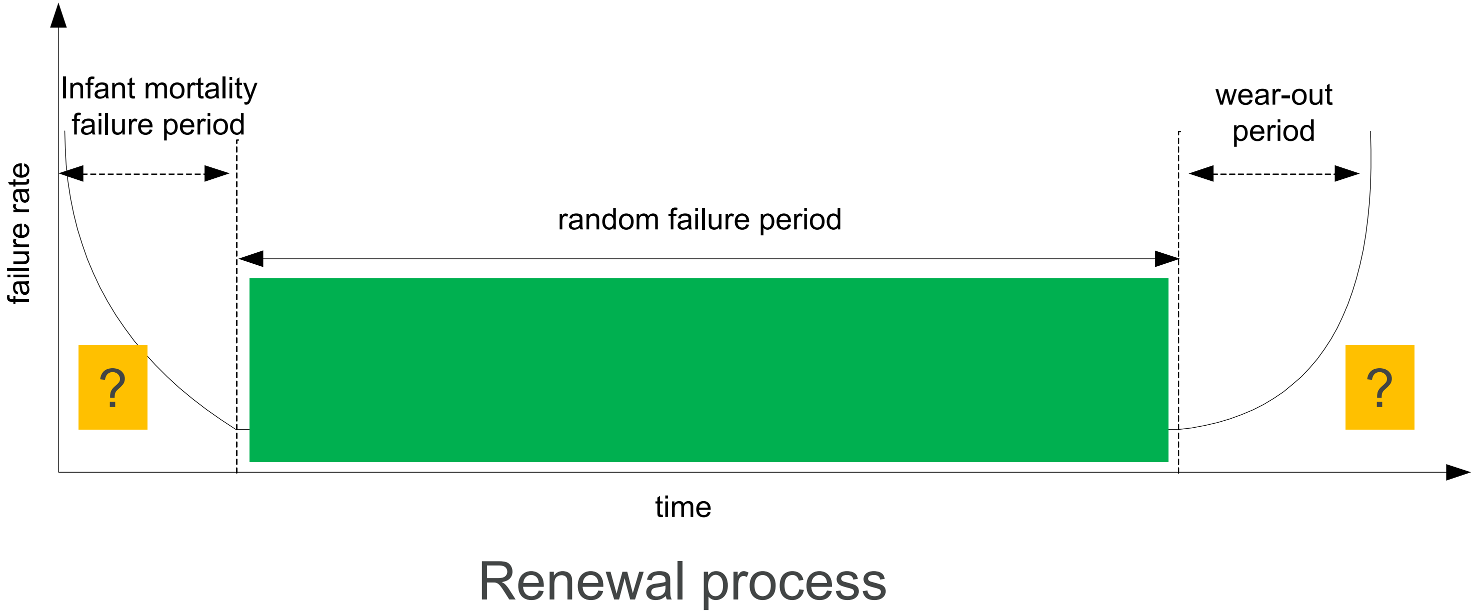
# Existing modelling methods

- Parametric methods, for example
  - Renewal process (RP): replacement, repaired as good as new
  - Nonhomogeneous Poisson process (NHPP): minimal repair
  - Virtual age models
  - Geometric process
- Non-parametric methods



# Renewal process

- Given a sequence of non-negative random variables $\{X_k, k = 1,2, \dots\}$, if they are Independent and identically distributed, then $\{X_k, k = 1,2, \cdots\}$ is called a **renewal process** (RP).



# Geometric process

- Given a sequence of non-negative random variables $\{X_k, k = 1,2, \dots\}$, if they are independent and the cdf of $X_k$ is given by $\boldsymbol{F(a^{k-1}x)}$ for $k = 1,2, \dots$, where $a$ is a positive constant, then $\{X_k, k = 1,2, \cdots\}$ is called a **geometric process** (GP) (Lam, 1988).

- Remarks
  - If $a > 1$, then $\{X_k, k = 1, 2, \cdots\}$ is stochastically decreasing.
  - If $a < 1$, then $\{X_k, k = 1, 2, \cdots\}$ is stochastically increasing.
  - If $a = 1$, then $\{X_k, k = 1, 2, \cdots\}$ is a renewal process (RP).
  - If $\{X_k, k = 1,2, \dots\}$ is a GP and $X_1$ follows the Weibull distribution, then the shape parameter of $X_k$ for $k = 2,3, \dots$ remains the same as that of $X_1$.

# Literature review

- The GP has been extensively studied since its introduction in 1988 (Lam, 1988), mainly due to its elegance and convenience in deriving mathematical properties in applications

**Geometric Process (GP)**

- $X_k \sim F(a_k x)$, $1 = a_1 \le \cdots \le a_k \ldots$ (Finkelstein, 1993)
- $X_k = a^{k-1} W_k$, $W_k$ are iid (Wang, et. al. 1996)
- $X_k \sim F(k^{-a} x)$, (Braun, et. al. 2005)
- $X_k \sim F((\alpha a^{k-1} + \beta\, b^{k-1})x)$(Wu, et. al., 2006)
- $\{a_i^{n-M_i} Z_n, M_i \le n < M_{i+1}\}$ forms a renewal process (Chan, et. al., 2006)
- $X_k \sim F\left(a^{-b_k} x\right)$, (Bordes, et. al. 2013)

**Geometric Process (GP)**

↓

$X_k \sim F(g(k)x)$

$g(k)$ is a function of $k$

# Motivation for the doubly geometric process (DGP)

The process $\{X_k, k = 1,2, \dots\}$ with $X_k \sim F(g(k)x)$ has the following restrictive implications

- **Invariance of the CV** (coefficient of variation): Given a GP $\{X_1, X_2, \dots\}$, then

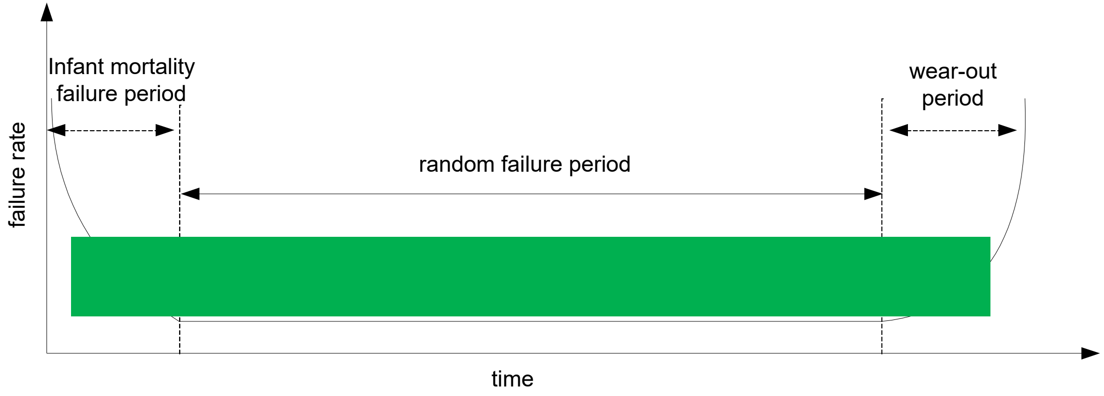
$$CV_k = \frac{\sqrt{V(X_k)}}{E(X_k)} = \frac{\sqrt{E(X_1^2) - \left(E(X_1)\right)^2}}{E(X_1)}$$

- **Invariance of the shape parameter**: Suppose $X_1 \sim 1 - \exp\{-\left(\frac{x}{\theta_1}\right)^{\theta_2}\}$, then

$$F(g(k)x) = 1 - \exp\{-\left(\frac{x}{\theta_1 \left(g(k)\right)^{-1}}\right)^{\theta_2}\}$$

# Literature review

- **Monotonicity of the GP**: from the remarks, the GP $\{X_k, k = 1,2, \ldots\}$ change monotonously, either increasing or decreasing



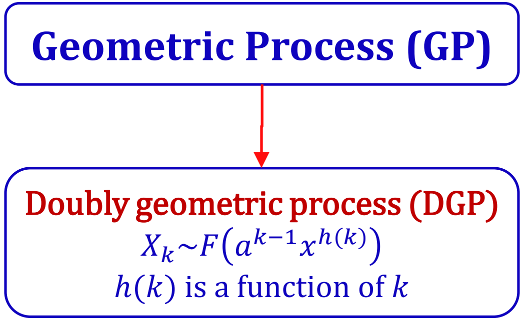

**Geometric Process (GP)**

- $X_k \sim F((\alpha a^{k-1} + \beta\, b^{k-1})x)$(Wu, et. al., 2006)
- $\{a_i^{n-M_i} Z_n, M_i \le n < M_{i+1}\}$ forms a renewal process (Chan, et. al., 2006)

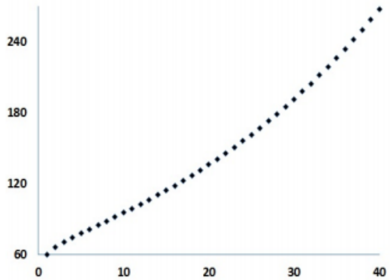
# Doubly geometric processes and double ratio geometric processes

# Doubly geometric process

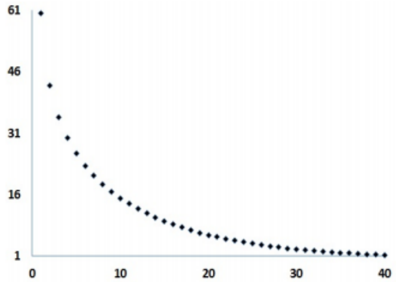
- Given a sequence of non-negative random variables $\{X_k, k = 1,2, \dots\}$, if they are independent and the cdf of $X_k$ is given by $\boldsymbol{F(a^{k-1}x^{h(k)})}$ for $k = 1,2, \dots$, where $a$ is a positive constant, then $\{X_k, k = 1,2, \cdots\}$ is called a **doubly geometric process** (GP) (Wu, 2018).



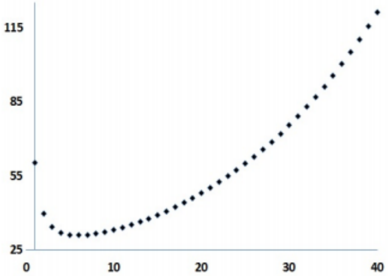
# Examples



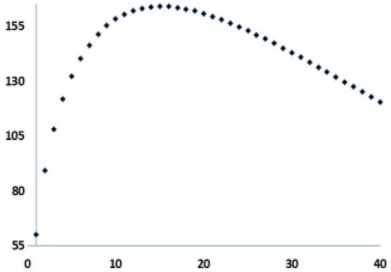

(a) $a = 0.97, b = -0.05, \theta_1 = 40$ and $\theta_2 = 0.6$.




(b) $a = 1.1, b = 0.2, \theta_1 = 40$ and $\theta_2 = 0.6$.




(c) $a = 0.92, b = 0.4, \theta_1 = 40$ and $\theta_2 = 0.6$.



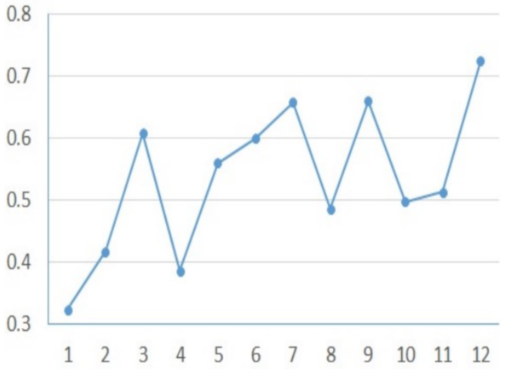

(d) $a = 1.02, b = -0.3, \theta_1 = 40$ and $\theta_2 = 0.6$.

Figure 1: DGPs with different parameter settings: the $Y$-axes are $\mathbb{E}[X_k]$ and the $X$ axes are $k$, $F(x) = 1 - e^{-(\frac{x}{\theta_1})^{\theta_2}}$, $h(k) = (1 + \log(k))^b$.

# Data analysis

Table 1: Time between warranty claims of 22 identical items (unit: day).

| Months \ Shipments | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | CV |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 10 | 8 | 13 | 7 | 8 | 16 | 9 | 6 | 7 | 15 | 11 | 9 | 13 | 7 | 9 | 6 | 13 | 10 | 9 | 5 | 0.323 |
| 2 | 7 | 4 | 8 | 6 | 9 | 6 | 1 | 8 | 8 | 9 | 11 | 10 | 10 | 9 | 7 | 8 | 1 | 3 | 9 | 12 | 0.417 |
| 3 | 11 | 7 | 15 | 3 | 4 | 3 | 3 | 13 | 9 | 13 | 6 | 4 | 3 | 5 | 5 | 6 | 3 | 2 | 8 | 5 | 0.607 |
| 4 | 8 | 3 | 12 | 6 | 7 | 6 | 11 | 9 | 9 | 7 | 10 | 7 | 8 | 11 | 6 | 5 | 8 | 5 | 6 | 17 | 0.385 |
| 5 | 4 | 3 | 4 | 2 | 8 | 6 | 7 | 15 | 7 | 9 | 10 | 5 | 2 | 6 | 4 | 14 | 3 | 7 | 10 | 13 | 0.559 |
| 6 | 11 | 8 | 5 | 10 | 4 | 5 | 7 | 8 | 1 | 6 | 11 | 1 | 3 | 4 | 3 | 9 | 4 | 5 | 16 | 13 | 0.599 |
| 7 | 7 | 7 | 22 | 3 | 5 | 14 | 12 | 5 | 4 | 7 | 9 | 4 | 4 | 6 | 17 | 4 | 13 | 3 | 6 | 5 | 0.658 |
| 8 | 11 | 8 | 4 | 5 | 4 | 12 | 6 | 10 | 3 | 4 | 8 | 3 | 5 | 12 | 9 | 10 | 3 | 11 | 4 | 4 | 0.486 |
| 8 | 4 | 3 | 16 | 7 | 1 | 8 | 3 | 6 | 1 | 5 | 6 | 4 | 4 | 12 | 5 | 2 | 4 | 5 | 5 | 6 | 0.660 |
| 10 | 2 | 5 | 9 | 4 | 3 | 10 | 11 | 8 | 1 | 12 | 8 | 6 | 10 | 7 | 2 | 3 | 9 | 10 | 6 | 9 | 0.497 |
| 11 | 5 | 4 | 8 | 4 | 7 | 12 | 1 | 9 | 5 | 8 | 4 | 7 | 3 | 2 | 3 | 5 | 13 | 8 | 7 | 6 | 0.513 |
| 12 | 4 | 5 | 2 | 6 | 1 | 7 | 6 | 10 | 4 | 3 | 12 | 2 | 2 | 17 | 4 | 13 | 6 | 1 | 9 | 5 | 0.724 |




- The AICc of DGP is 630.090 and that of GP is 630.242

# Reliability data: Time between failure data

Table 2: The datasets, including TBF(Time between failures).

| No. | Dataset | $N_0$ | References |
|---|---|---|---|
| 1 | Hydraulic system (LHD3) | 25 | Kumar and Klefsjö (1992) |
| 2 | Propulsion diesel engine failure data | 71 | Ascher and Feingold (1984) |

**Reliability data**

**The least squares method**

**Table 3** Comparison of the performance of the GP and the DGP based on the least squares method

| *No.* | *Parameters of the DGP* | | | *Parameters of the GP* | | $RMSE_{DGP}$ | $RMSE_{GP}$ |
|---|---|---|---|---|---|---|---|
| | $\hat{a}$ | $\hat{b}$ | $\hat{\mu}$ | $\hat{a}$ | $\hat{\mu}$ | | |
| 1 | 0.944 (0.0559) | 0.499 (0.174) | 531.406 (109.390) | 1.0382 (0.0315) | 209.841 (67.652) | *111.729* | 144.431 |
| 2 | 0.909 (0.0607) | 0.488 (0.280) | 147.624 (62.664) | 0.972 (0.0181) | 56.702 (20.486) | *65.670* | 69.810 |

**The maximum likelihood method**

**Table 4** Comparison of the performance of the GP and the DGP based on the maximum likelihood method

| *No.* | *Estimated parameters of the DGP* | | | | *Estimated parameters of the GP* | | | $AICc_{DGP}$ | $AICc_{GP}$ | $AICc_{PL}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{a}$ | $\hat{b}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{a}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | | | |
| 1 | 0.884 (0.0938) | 0.638 (0.352) | 449.165 (337.92) | 0.789 (0.227) | 1.0147 (0.0230) | 168.807 (58.139) | 1.0287 (0.159) | *301.376* | 304.182 | 311.851 |
| 2 | 0.899 (0.0714) | 0.502 (0.349) | 147.636 (103.569) | 0.964 (0.281) | 0.983 (0.0151) | 73.070 (19.461) | 1.295 (0.182) | *318.030* | 319.445 | 323.094 |

# GP-like models that can describe non-monotonic trends



- There are GP-like models that can describe non-monotonic trends

**Geometric Process (GP)**

- $X_k \sim F\big((\alpha a^{k-1} +$
  - 4 parameters
- $\{a_i^{n-M_i} Z_n, M_i \leq n < M_{i+1}\}$ forms a renewal process (Chan, et. al., 2006)
  - Flexible, but many parameters
- $X_k \sim F\big(a^{-b_k} x^{h(k)}\big)$, (Wu, 2018)
  - 2 parameters, the shape parameters changes over $k$

- **Invariance of the shape parameter**: Suppose $X_1 \sim 1 - \exp\{-\left(\frac{x}{\theta_1}\right)^{\theta_2}\}$, then

$$F(g(k)x) = 1 - \exp\{-\left(\frac{x}{\theta_1 (g(k))^{-1}}\right)^{\theta_2}\}$$

# Requirements of a new GP-like model

University of Kent

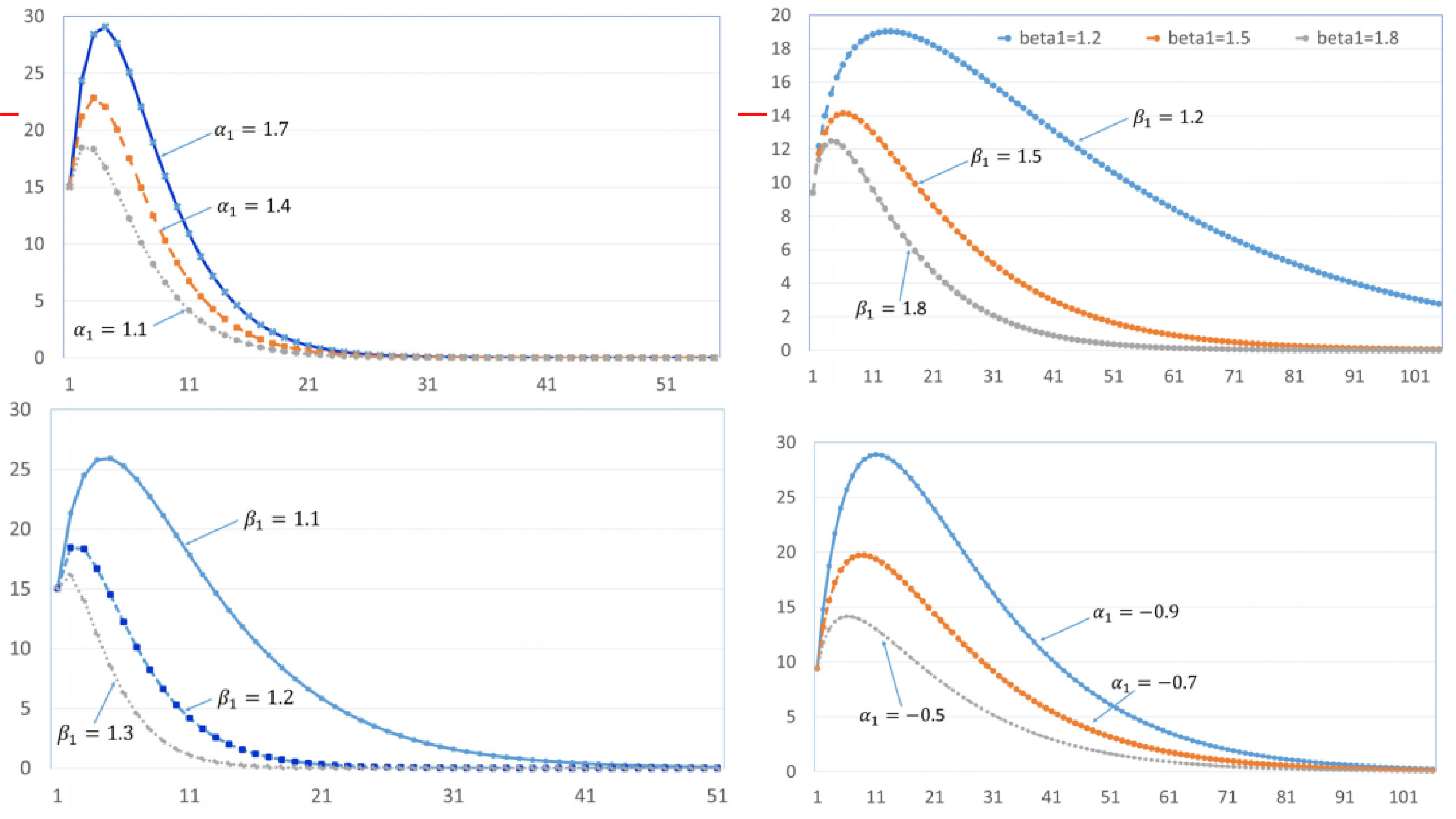
- We require a model satisfying three conditions
  - A parsimonious model: with fewer parameters
  - Be able to describe non-monotonic trends
  - If $X_1$ follows the Weibull distribution, the distributions of $X_k$ (for $k = 2,3, \ldots$) have the same shape parameter as that of $X_1$

- DRGP: Given a sequence of non-negative random variables $\{Z_k^D, k = 1,2, \ldots\}$, if they are independent and the cdf of $Z_k^D$ is given by $F_k^D(t) = 1 - \exp\{-\int_0^t b_k h(a_k u) du\}$ for $k = 1,2, \ldots$, where $a_k$ and $b_k$ are positive parameters (or ratios) and $a_1 = b_1 = 1$. $\{Z_k^D, k = 1,2, \ldots\}$ is called the **double-ratio geometric process** (DRGP) (Wu, 2021)

# An equivalent definition

- Lemma: Given a sequence of non-negative random variables $\{Z_k^D, k = 1,2,\dots\}$, if they are independent and the cdf of $a_k^{-1} Z_k^D$ is given by $F_k^D(t) = 1 - \left(1 - F_1^D(a_k t)\right)^{\frac{b_k}{a_k}}$ for $k = 1,2,\dots$, where $a_k$ and $b_k$ are positive parameters (or ratios), and $a_1 = b_1 = 1$. Then $\{Z_k^D, k = 1,2,\dots\}$ is a double-ratio geometric process.

- [Monotonicity] Suppose $h(t)$ is a monotonously increasing function in $t$, $\{Z_k^D, k = 1,2,\dots\}$ is a DRGP, then
- If both $a_k$ and $b_k$ are increasing in $k$, then the DRGP is stochastically decreasing;
- If both $a_k$ and $b_k$ are decreasing in $k$, then the DRGP is stochastically increasing; and
- If both $a_k$ (or $b_k$) is increasing in $k$ and $b_k$ (or $a_k$) is decreasing, then the DRGP may not be stochastically monotonic.



α1 = 1.7
α1 = 1.4
α1 = 1.1
beta1=1.2
beta1=1.5
beta1=1.8
β1 = 1.2
β1 = 1.5
β1 = 1.8
β1 = 1.1
β1 = 1.2
β1 = 1.3
α1 = −0.9
α1 = −0.7
α1 = −0.5

# Datasets for testing

| No. | Dataset | $n$ | Description | Reference |
|---|---|---|---|---|
| 1 | LHD3 | 25 | failures of a load-haul-dump (LHD) machine deployed at Kiruna mine, Sweden | Kumar and Klefsjö (1992) |
| 2 | LHD11 | 28 | failures of a load-haul-dump (LHD) machine deployed at Kiruna mine, Sweden | Kumar and Klefsjö (1992) |
| 3 | Calvert Cliffs | 23 | diesel generator failure data from power plant "Calvert Cliffs" | Kvam et al. (2002) |
| 4 | Pump D | 30 | reliability data collected from a main pump (A) at an oil refinery | Percy and Alkali (2007) |

Test for non-monotonic trend.

| No. | p-value from the Ljung-Box test | p-value from $V_1$ | p-value from $V_4$ |
|---|---|---|---|
| 1 | 0.886 | 0.00866 | 0.0216 |
| 2 | 0.899 | 0.0159 | 0.0114 |
| 3 | 0.509 | 0.0194 | 0.0276 |
| 4 | 0.107 | 0.00874 | 0.00180 |

Table 5: AICc values of the models

| No. | EGP | DGP | TGP | DRGP-I | DRGP-II | DRGP-III | DRGP-IV |
|---|---|---|---|---|---|---|---|
| 1 | 310.198 | 301.345 | 309.431 | 307.040 | 305.281 | 300.228 | 300.228 |
| 2 | 328.790 | 323.711 | 324.103 | 325.802 | 324.983 | 321.312 | 321.312 |
| 3 | 224.545 | 221.461 | 222.055 | 221.490 | 222.299 | 219.632 | 219.632 |
| 4 | 297.200 | 300.939 | 298.127 | 306.622 | 308.924 | 295.031 | 295.031 |

Table 6: Parameters

| No. | DRGP-III ($\hat{a}_k = k^{\hat{\alpha}_1}$ and $\hat{b}_k = \hat{\beta}_1^{k-1}$) | | | | DRGP-IV ($\hat{a}_k = \hat{\beta}_2^{k-1}$ and $\hat{b}_k = k^{\hat{\alpha}_2}$) | | | |
|---|---|---|---|---|---|---|---|---|
| | $\widehat{\alpha}_1$ | $\widehat{\beta_1}$ | $\widehat{\theta}_1$ | $\widehat{\theta}_2$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_2$ | $\widehat{\theta}_1$ | $\widehat{\theta}_2$ |
| 1 | 6.942 | 0.848 | 1.231 | 553.510 | 1.603 | 0.490 | 1.231 | 553.510 |
| | (5.502) | (0.0621) | (0.200) | (308.283) | (0.671) | (0.287) | (0.200) | (340.652) |
| 2 | 14.233 | 0.891 | 1.085 | 347.046 | 1.205 | 0.254 | 1.085 | 347.046 |
| | (27.230) | (0.0555) | (0.166) | (251.771) | (0.617) | (0.660) | (0.165) | (251.777) |
| 3 | -29.898 | 0.855 | 0.960 | 92.937 | 1.192 | 50.442 | 0.960 | 92.937 |
| | (144.543) | (0.0861) | (0.182) | (106.597) | (0.935) | (920.240) | (0.176) | (106.516) |
| 4 | -8.996 | 1.281 | 1.297 | 4.629 | -2.673 | 2.304 | 1.297 | 4.629 |
| | (4.827) | (0.0787) | (0.185) | (2.985) | (0.689) | (1.0352) | (0.186) | (2.985) |

# Contributions and further research

- **Contributions**
  - the DGP can model a stochastic process with varying CV
  - the DGP can model recurrent event processes where $F_k(x)$'s have different shape parameters over $k$'s, which can be done by neither the GP-like models nor other repair models such as reduction of age models
  - the DGP and DRGP can model not only monotonously increasing or decreasing stochastic processes, but also processes with complicated failure intensity functions such as the bathtub shaped curves and the upside-down bathtub shaped curves
- **Further research**
  - Both the DGP and DRGP have the limitation that they are parametric models, a Bayesian nonparametric method is under development

# References


- Braun WJ, Li W and Zhao YQ (2005). Properties of the geometric and related processes. *Naval Research Logistics* 52(7):607–616.
- Chan JS, Yu PL, Lam Y and Ho AP (2006). Modelling sars data using threshold geometric process. *Statistics in Medicine* 25 (11):1826–1839.
- Finkelstein M (1993). A scale model of general repair. *Microelectronics Reliability* 33(1):41–44.
- Kvam, P.H., Singh, H. and Whitaker, L.R., (2002). Estimating distributions with increasing failure rate in an imperfect repair model. *Lifetime Data Analysis*, 8(1), pp.53-67.
- Kumar U and Klefsjö B (1992). Reliability analysis of hydraulic systems of LHD machines using the power law process model. *Reliability Engineering & System Safety* 35(3):217–224.
- Percy, D. F., & Alkali, B. M. (2007). Scheduling preventive maintenance for oil pumps using generalized proportional intensities models. *International Transactions in Operational Research*, 14(6), 547-563.
- Wang H and Pham H (1996). A quasi renewal process and its applications in imperfect maintenance. *International Journal of Systems Science* 27(10):1055–1062.
- Wu S and Clements-Croome D (2006). A novel repair model for imperfect maintenance. *IMA Journal of Management Mathematics* 17(3):235–243.
- Wu, S. (2017). Doubly geometric processes and applications. *Journal of the Operational Research Society*, https://doi.org/10.1057/s41274-017-0217-4
- Wu, S., (2021). The double ratio geometric process for the analysis of recurrent events. *Naval Research Logistics (NRL)* **https://doi.org/10.1002/nav.22021**.

University of Kent

I would like to thank the organisers of t*he 3rd International Conference on System Reliability and Safety Engineering (SRSE 2021)* for their invitation