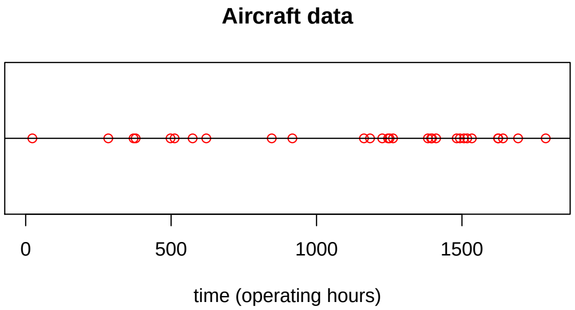
Extended geometric process and its applications to reliability

Laurent Bordes & Sophie Mercier

Université de Pau et des Pays de l'Adour

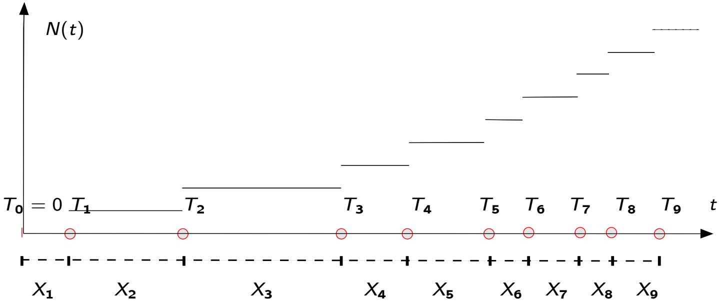
AMMSI 2014 - Toulouse

Recurrent event times data



$n = 29$ successive failure times (operating hours) of an air-conditioning equipment of a Boeing 720 aircraft (Proschan, 1963). Data are available in Lindsey (2004). Are the inter-arrival times stochastically decreasing or increasing?

Notations and some models



- T_i time of the i-th failure/repair and $X_i = T_i - T_{i-1}$ is the i-th inter-arrival time ($i \geq 1$);
- Modeling the distribution of the counting process $t \mapsto N(t)$: Non Homogeneous Poisson Process (ABAO), Renewal Process (AGAN) and many other *in between* situations (BP, Virtual Age, etc.).
  Aim: discuss the Geometric Process approach.

Outline

Outline

Geometric Process (GP) and its generalization

Lam (1988) introduced generalized the Renewal Process (RP).
$\boldsymbol{Y} = (Y_n)_{n\geq 1}$ is a RP and $a > 0$ a real number. We set:

$$X_n = a^{n-1} Y_n \quad \text{for} \quad n \geq 1,$$

and call $\boldsymbol{X} = (X_n)_{n\geq 1}$ a Geometric Process. Then the distribution of $\boldsymbol{X}$ (or $\boldsymbol{N} = (N(t))_{t\geq 0}$) depends on (a, F) where F is the unknown cdf of the underlying RP $\boldsymbol{Y} = (Y_n)_{n\geq 1}$ (it is the cdf of X_1). Without a parametric assumption on F the GP is a semiparametric model.

Generalized GP: We consider here that the link function between X_n and Y_n is not necessarily geometric, we assume that there exists a sequence $(b_n)_{n\geq 1}$ such that $X_n = a^{b_n} Y_n$ for $n \geq 1$. We mainly assume that $(b_n)_{n\geq 1}$ is known but we discuss the case where $b_n = g(n; \theta)$ where θ is an additional euclidean parameter.

Outline

Semiparametric estimation method

Assume that the first n failure times are observed. Following Lam (1992), setting $Z_k = \log X_k = b_k\beta + \mu + \varepsilon_k$ for $k = 1, \ldots, n$ where

$$\beta = \log a, \quad \mu = \mathbb{E}[\log Y_k] \quad \text{and} \quad \varepsilon_k = \log Y_k - \mu$$

the unknown parameter β is estimated by solving

$$(\hat{\mu}_n, \hat{\beta}_n) = \arg\min_{\mu,\beta} \sum_{k=1}^{n} (Z_k - \beta b_k + \mu)^2.$$

Finally we have $\hat{a}_n = \exp(\hat{\beta}_n)$. The cdf F is naturally estimated by the empirical cdf of pseudo-observations $\tilde{Y}_k$ of Y_k defined by $\tilde{Y}_k = \hat{a}_n^{-b_k} X_k$, then we have

$$\hat{F}_n(t) = \frac{1}{n} \sum_{k=1}^{n} 1_{\{\tilde{Y}_k \leq t\}}.$$

Some asymptotic results (1/3)

Assume that $\mathbb{E}[Z_1^2] < +\infty$, we have $\hat{a}_n = \exp(\hat{\beta}_n)$.

Strong Law of Large Numbers. SLLN for weighted sums of iid random variables (Cuzick, 1995; Bai et al. 2000): $\alpha_n(\hat{\beta}_n - \beta) \xrightarrow{a.s.} 0$ where $\alpha_n^2 = n^{-1}\sum_{k=1}^n b_k^2 - \left(n^{-1}\sum_{k=1}^n b_k\right)^2$. Since $(\alpha_n)_{n\geq 1}$ is non decreasing $\hat{\beta}_n \xrightarrow{a.s.} \beta$).

Law of Iterated Logarithm. LIL for weighted sums of iid random variables (Bai et al. 1997) leads to

$$\limsup_{n\to+\infty} \frac{\sqrt{n}\alpha_n^2}{b_n\sqrt{\log n}}|\hat{\beta}_n - \beta| \leq 2\sqrt{2}\sigma \quad \text{a.s.}$$

where $\sigma^2 = \text{var}(Z_1)$.

Some asymptotic results (2/3)

Central Limit Theorem. The CLT is obtained by combining the previous results with the Lindeberg-Feller theorem. If in addition to $\mathbb{E}[Z_1^2] < +\infty$ we have $\sqrt{n}\alpha_n/b_n \to +\infty$, then $\sqrt{n}\alpha_n(\hat{a}_n - a) \xrightarrow{d} \mathcal{N}(0, a^2\sigma^2)$, and

$$\hat{\sigma}_n^2 = \frac{1}{n-2}\sum_{k=1}^{n}\left(Z_k - \hat{\beta}_n b_k + \hat{\mu}_n\right)^2$$

is a consistent unbiased estimator of σ^2.

Uniform Strong Consistency. Assume that Z_1 has a bounded df g, a bounded second order moment and that

$$\limsup_{n\to+\infty} \frac{b_n^2\sqrt{\log n}}{\sqrt{n}\alpha_n^2} = 0,$$

then $\|\hat{F}_n - F\|_\infty \xrightarrow{a.s.} 0$.

Some asymptotic results (3/3)

Remarks.

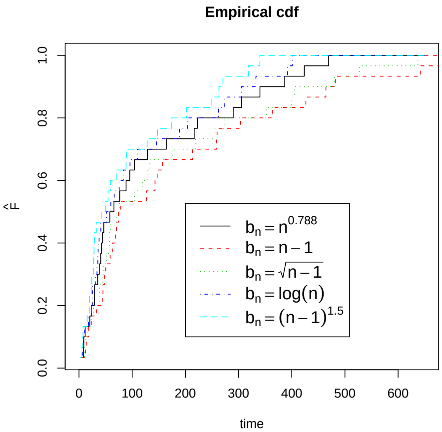
1. Conditions on $(b_n)_{n\geq 1}$ are satisfied whenever $b_n = (n-1)^\alpha$ or $b_n = (\log n)^\alpha$ with $\alpha > 0$.
2. If $b_n = (n-1)^\alpha$ the convergence rate in the CLT is $n^{\alpha+1/2}$. We retrieve the Lam et al. (2004) result.
3. The usual $\sqrt{n}$ CLT convergence rate is obtained for $b_n = \log n$.

Outline

Aircraft data

Estimates of a for various b_n

b_n	$(n-1)^{0.788}$	$\log n$	$\sqrt{n-1}$	$n-1$	$(n-1)^{3/2}$
$\hat{a}$	0.900	0.620	0.740	0.952	0.992
95% CI for a	[0.798,1.003]	[0.275,0.966]	[0.489,0.991]	[0.901,1.003]	[0.982,1.001]



More general model

If $X_k = a^{g(k;\theta)} Y_k$ for $k \geq 1$ then we estimate μ, β and θ by minimizing

$$c(\mu, \beta, \theta) = \sum_{k=1}^{n} (Z_k - \beta g(k;\theta) - \mu)^2.$$

Parameters μ and β can be expressed as functions of θ, indeed:

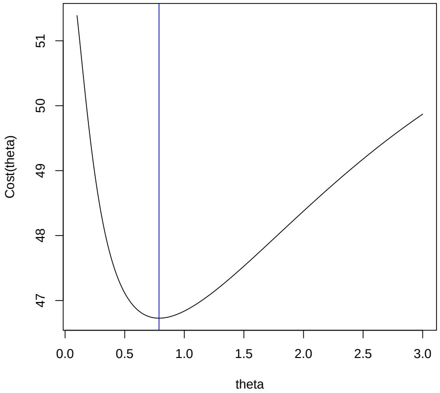
$$\begin{aligned}
\mu_n(\theta) &= \frac{\left(\sum_{k=1}^n g(k;\theta)\right)\left(\sum_{k=1}^n y_k g(k;\theta)\right) - \left(\sum_{k=1}^n y_k\right)\left(\sum_{k=1}^n g^2(k;\theta)\right)}{\left(\sum_{k=1}^n g(k;\theta)\right)^2 - n\left(\sum_{k=1}^n g^2(k;\theta)\right)}, \\
\beta_n(\theta) &= \frac{\left(\sum_{k=1}^n g(k;\theta)\right)\left(\sum_{k=1}^n y_k\right) - n\left(\sum_{k=1}^n y_k g(k;\theta)\right)}{\left(\sum_{k=1}^n g(k;\theta)\right)^2 - n\left(\sum_{k=1}^n g^2(k;\theta)\right)},
\end{aligned}$$

hence $\hat{\theta}_n = \arg\min_\theta \mathcal{C}_n(\theta)$ where

$$\mathcal{C}_n(\theta) = \sum_{k=1}^{n} (Z_k - \beta_n(\theta) g(k;\theta) - \mu_n(\theta))^2.$$

More general model for aircraft data

We minimize $\theta \mapsto \mathcal{C}_n(\theta)$ for $b_n = (n-1)^\theta$.



Results: we obtain $b_n = (n-1)^{0.788}$ and we do not reject $H_0 : a = 1$ at the 5% significance level.

Outline

Mean number of failures on $[0, t]$: n(t)

Approximating $n(t)$: For $c > 0$ and $t \geq 0$ define $\tau^c = \inf\{n \geq 1; X_n < c\}$ and $n^c(t) = \mathbb{E}\left[\sum_{n=1}^{\tau^c - 1} 1_{\{T_n \leq t\}}\right]$ then $n^c(t) \leq n(t) = \mathbb{E}[N(t)]$. By the monotone convergence theorem $n^c(t) \to n(t)$.
We set $u_n^c(t) = \mathbb{P}(T_n \leq t, X_1 \geq c, \ldots, X_n \geq c)$ and we have

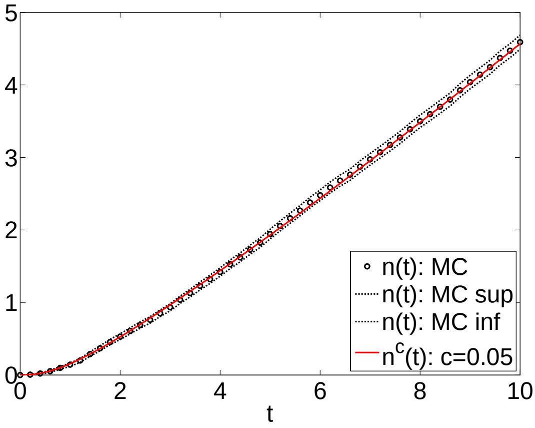
$$n^c(t) = \sum_{n=1}^{\lfloor t/c \rfloor} u_n^c(t).$$

$(u_n^c(t))_{n \geq 1}$ may be computed recursively using

$$\begin{aligned} u_1^c(t) &= (F(t) - F(c))^+, \\ u_{n+1}^c(t) &= \frac{1}{a^{b_{n+1}}} \int_0^{(t-c)^+} u_n^c(u) f\left(\frac{t-u}{a^{b_{n+1}}}\right) du. \end{aligned}$$

Numerical example

Assumptions: $a = 0.8$, $Y_1 \sim \Gamma(2.5, 1)$, $b_n = (\log n)^{0.7}$.



A replacement policy

Assumptions: $a < 1$, $X_i < s \Rightarrow$ replacement (instantaneous) at some cost c_R, at failure time replacement (instantaneous) at cost c_F. We call $C(s)$ the asymptotic unitary cost per unit time.

- Setting $C([0,t])$ the cumulated cost on $[0,t]$ we have:

$$C(s) = \lim_{t\to+\infty} \frac{C([0,t])}{t} \quad \text{a.s.}$$

- We have

$$C(s) = \frac{c_R + c_F\mathbb{E}[\tau^s - 1]}{\mathbb{E}[T_{\tau^s}]},$$

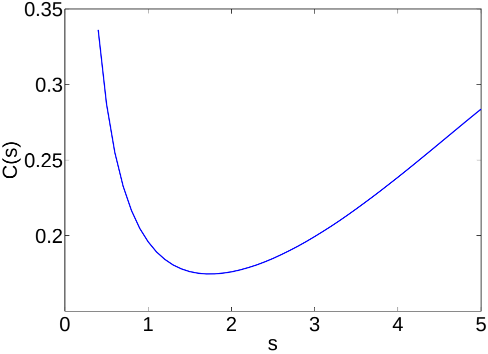
with

$$\mathbb{E}[\tau^s - 1] = \sum_{k=1}^{+\infty} v_k^s \quad \text{and} \quad \mathbb{E}[T_{\tau^s}] = \mathbb{E}[Y_1]\left(1 + \sum_{k=1}^{+\infty} a^{b_{k+1}} v_k^s\right)$$

and $v_k^s = \prod_{i=1}^k \bar{F}(s/a^{b_i})$ with $\bar{F} = 1 - F$.

Numerical example

Assumptions: $a = 0.8$, $Y_1 \sim \Gamma(2.5, 1)$, $b_n = (\log n)^{0.7}$, $c_R = 1$ and $c_F = 0.5$.



The cost function reaches its minimum at $s^{opt} = \arg\min_{s>0} C(s) \approx 1.70$.

Concluding remarks

- Extension of the classical Geometric Process $\Rightarrow$ allows easy interpretation.
- New statistical results but some important questions $\Rightarrow$
  - weak convergence result for $(\hat{a}_n, \hat{F}_n)$ (or convergence results $(\hat{a}_n, \hat{\theta}_n, \hat{F}_n)$);
  - easy to introduce covariates in a;
  - formal test to choose between several sequences $(b_n)_{n\geq 1}$.
- Reliability $\Rightarrow$
  - alternative replacement policy could be more appropriate;
  - find an upper bound for the the pseudo-renewal function.