Bayesian semiparametric analysis for a single item maintenance optimization

Elmira Popova, Paul Damien, Alexander Galenko [1] Tim Hanson[2]

[1]The University of Texas at Austin
Austin, TX, USA

[2]University of Minnesota
Twin Cities, MN, USA

Isaac Newton Institute for Mathematical Sciences Workshop "Construction and Properties of Bayesian Nonparametric Regression Models"

Outline

Main assumptions

- Making optimal (cost efficient) decisions is important to us.
- There are many sources of uncertainties.
- We will optimize the expected value of a predefined random objective function:
- $\min_{x \in \mathbb{X}} E_{\mu}\left[f(x, \xi)\right]$
  - x - set of decision variables (might belong to a constraint set $\mathbb{X}$).
  - ξ - vector of random variables, μ - probability measure (might depend on x).
  - $f(\cdot)$ - cost, utility function, indicator function, etc.
  - $E[\cdot]$ - the expectation operator w.r.t. μ.

Examples

- **Finance:** portfolio optimization problem
  - x - number of shares (or percent of total budget) for each asset in the portfolio
  - μ - the underlying assets' prices
  - $f(\cdot)$ - total return
  - Decision dependent randomness not justified (unless you are Warren Buffet or Bill Gates).
- **Reliability:** maintenance optimization problem
  - x - time to perform maintenance
  - μ - failure times
  - Decision dependent randomness justified - maintenance can change the future failure frequency (faster, slower)

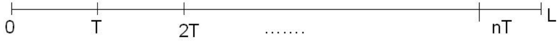
Main objectives

- **What is the impact of the uncertainty model/inference on the objective function?**
- **Does the interaction between decisions and uncertainty exist? Is the *interaction hypothesis* real?**
- Will present two maintenance optimization models:
- First one: assume a certain class of failure distributions and no decision dependent uncertainty. Optimality results are derived.
- The second one - will not make an assumption about the underlying distribution. Will use Bayesian nonparametric approach to estimate it and draw samples from it. The decision dependent randomness will appear.

Policy definition

- *Perform complete overhaul (renewal) of the system every T units of time, if it fails in between, repair it to "as good as old" state.*
- T - integer variable, it belongs to a set of weeks where one can perform the overhaul
- The time between failures of the system is random with a distribution function F.
- Want to find the schedule T, that minimizes the expected total cost over the finite time horizon L.

Graphically,



Barlow and Hunter, 1960, showed that the expected number of failures in $(0, t)$ equals to $E[N(t)] = \int_0^t q(u)du$, where $q(u)$ is the hazard function

Objective function parameters

- $E[N(T)]$ - expected number of failures in $(0, T)$.
- C_{pm} - cost for renewal.
- C_{cm} - cost for minimal repair.
- C_d - cost for downtime due to production loss.
- p - probability that a failed item can cause a production loss.

Objective function

Find T (integer) in $A \subset [0, L]$ such that

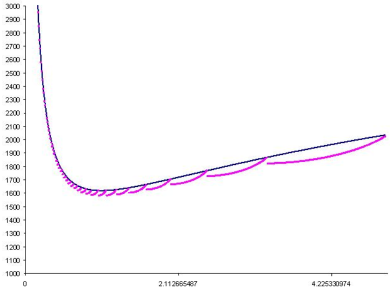
$$\min_{T \in A} z(T)$$

where

$$\begin{aligned} z(T) \quad = \quad & \min_{T \in A} \{ C_{pm} \lceil L/T \rceil + [pC_d + (1-p)C_{cm}] \lfloor L/T \rfloor E[N(T)] \\ & + C_{pm} + [pC_d + (1-p)C_{cm}] \lfloor L/T \rfloor E \left[N(L - T \lfloor \frac{L}{T} \rfloor) \right] \} \end{aligned}$$

Optimization results with IFR failure time distribution

- IFR = Increasing Failure Rate
- The objective function and its continuous counterpart



Properties of the objective function under IFR assumption

- From *A. Galenko, D. Morton, E. Popova, E. Kee, R. Grantom, and A. Sun, "Operational level models and methods for risk informed nuclear asset management", In Proceedings Of The American Nuclear Society International Topical Meeting On Probabilistic Safety Assessment, 2005*:
- $z(T)$ is lower semicontinuous with discontinuities at $L/n, n = 1, 2, 3, \ldots$
- $z(T)$ is increasing and convex in each interval $(L/(n+1), L/n), n = 1, 2, 3, \ldots$
- Its continuous counterpart if quasiconvex.
- The minimum of the continuous counterpart leads to the minimum of $z(T)$.

First model summary

- The main reason to be able to design the algorithm–the increasing failure rate assumption.
- It implies that the system will monotonically age over time.
- It restricts us to a certain class of probability distributions.
- How does the real data look like?

Data description – from STPNOC

- STPNOC - South Texas Project Nuclear Operating Company - nuclear power plant near Houston, TX, USA.
- Time (in days) between failures (right censored) for 5 systems, called Group 1 - Group 5.
- Time frame: 8/1/1988, 12/1/2003. Data are heavily right censored.
- Group 1 : EHC (Electric Hydraulic Control system) Servo valve, 26 (exact, not censored) observations
- Group 2 : EHC accumulator, 20 observations
- Group 3 : EHC isolation valve, 10 observations
- Group 4 : flow control valve (16in), 17 observation
- Group 5: flow control valve (4in), 10 observations

Estimation of the failure rates - current plant practice

- Time between failures is assumed exponential. This makes data gathering easy: at the end of the month one needs to record only the number of failures instead of the time of each failure.
- The prior distribution of the parameter of the exponential is assumed Gamma. Bayesian updates are run at the end of each month.

Classical hazard function shape

- In classical reliability theory - “bathtub” curve:
  - “Burn-in” portion - the probability of failure decreases with age, decreasing hazard
  - “Flat” portion - this is the exponential case, constant hazard
  - “Wear-out” portion - the probability of failure increases with age that leads to “death”, increasing hazard
- The STPNOC risk engineers want to relax these assumptions.

npBayes Lifetime Model

- Accelerated Lifetime Model: $S_x(t) = S_0(e^{x\beta}t)$.
- x is a vector of covariates–in our case dummy variables corresponding to Groups 2, 3, 4, and 5.
- S_x is the reliability function, S_0 is the "baseline" reliability function–Group 1 in our case.
- A mixture of Polya tree (MPT) prior on S_0.
- The MPT prior allows to "center" the Polya tree at a parametric distribution, Weibull in our case.

MPT settings

The Mixture of Polya Trees prior on S_0 is:

$$\begin{aligned} S_0|\theta &\sim PT(c, \rho, G_\theta) \\ \theta &\sim p(\theta) \\ G_\theta(t) &= 1 - \exp(-\theta_2 t^{\theta_1}) \end{aligned}$$

where G_θ is the Weibull distribution, commonly used in reliability settings.

MPT settings

The set of partitions is

$$\Pi_\theta = \{B_\theta(\epsilon) : \epsilon \in \bigcup_{l=1}^{J} \{0,1\}^l\}$$

and the family of positive real numbers is

$$\mathcal{A} = \{\alpha_\epsilon : \epsilon \in \bigcup_{j=1}^{M} \{0,1\}^j\}$$

where $\alpha_{\epsilon_1 \cdots \epsilon_k} = wk^2$ for $w = 1$

- Given Π_θ and $\mathcal{A}$, the Polya tree prior is defined up to level J by the random vectors
- $\mathcal{Y} = \{(Y_{\epsilon 0}, Y_{\epsilon 1}) : \epsilon \in \bigcup_{j=0}^{M-1} \{0,1\}^j\}$ via

$$S_0\{B_\theta(\epsilon_1 \cdots \epsilon_k) | \mathcal{Y}, \theta\} = \prod_{j=1}^{k} Y_{\epsilon_1 \cdots \epsilon_j},$$

for $k = 1, \ldots, M$, where $M = \log_2(n)$ (6 in our case) and

- $S_0(A)$ - baseline measure of any set A.
- $(Y_{\epsilon 0}, Y_{\epsilon 1})$ - independent Dirichlet distributions:

$$(Y_{\epsilon 0}, Y_{\epsilon 1}) \sim \text{Dirichlet}(\alpha_{\epsilon 0}, \alpha_{\epsilon 1}), \quad \epsilon \in \bigcup_{j=0}^{M-1} \{0,1\}^j$$

- Define $p = p(\mathcal{Y}) = (p_1, \ldots, p_{2^J})'$ as

$$p_{j+1} = S_0\{B_\theta(\epsilon_1 \cdots \epsilon_J)|\mathcal{Y}, \theta\} = \prod_{i=1}^{J} Y_{\epsilon_1 \cdots \epsilon_i}$$

where

- $\epsilon_1 \cdots \epsilon_J$ is the base-2 representation of j, $j = 0, \ldots, 2^J - 1$.
- The baseline survival function is,

$$S_0(t|\mathcal{Y}, \theta) = p_N \left[N - 2^J G_\theta(t)\right] + \sum_{j=N+1}^{2^J} p_j,$$

where

- N - the integer part of $2^J G_\theta(t) + 1$
- $g_\theta(\cdot)$ - the density corresponding to G_θ

- The density of $S_0(t|\mathcal{Y},\theta)$ is

$$f_0(t|\mathcal{Y},\theta) = \sum_{j=1}^{2^J} 2^J p_j g_\theta(t) I_{B_\theta(\epsilon_J(j-1))}(t) = 2^J p_N g_\theta(t),$$

- The likelihood for right censored data is given by

$$\mathcal{L}(\mathcal{Y},\theta,\beta) = \prod_{i=1}^{n} f_{x_i}(t_i|\mathcal{Y},\theta,\beta)^{\delta_i} S_{x_i}(t_i|\mathcal{Y},\theta,\beta)^{1-\delta_i}$$

where

- $\mathcal{D}_i$ - i^{th} triple (x_i, t_i, δ_i)
- $T_i \sim S_{x_i}(\cdot)$
- $\delta_i = 0$ equivalent to t_i is a censoring time, $T_i > t_i$, and $\delta_i = 1$ denotes that t_i is a survival time, $T_i = t_i$.

Markov Chain Monte Carlo

- The MCMC algorithm alternately samples the full conditional distributions, $[\beta, \theta | \mathcal{Y}, \mathcal{D}]$ and $[\mathcal{Y} | \beta, \theta, \mathcal{D}]$.
- These sample values $\{(\beta^j, \theta^j, \mathcal{Y}^j)\}_{j=1}^{MC}$ are approximate independent realizations from the target densities.
- Sampling methods taken from Hanson (2006).

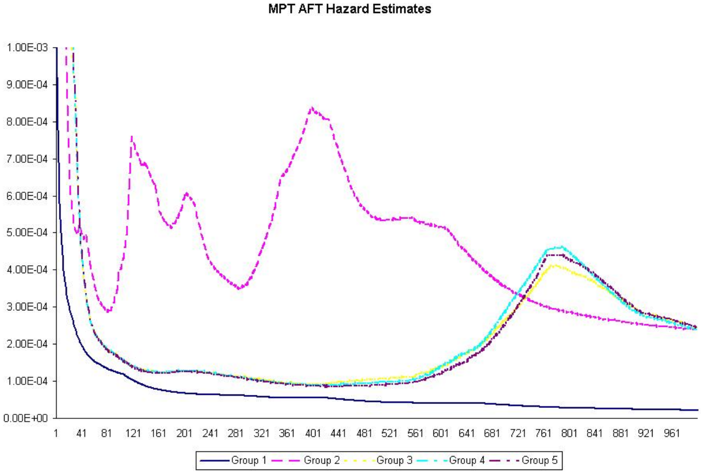
Fitted model

- $S_x(t) = S_0(e^{\beta_j} t), j = 2, 3, 4, 5$
- e^{β_j} is the acceleration factor - the ratio of mean time to failure of Group j to Group 1
- $e^{\beta_{j_1} - \beta_{j_2}}$ - This ratio for Group j_1 relative to j_2

MCMC results

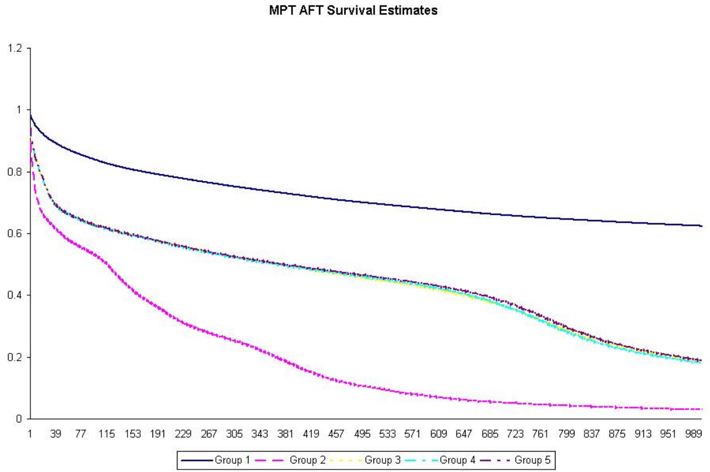
Coef.	Est.	95% CI	Coef.	Est.	95% CI
β_2	3.83	(1.78, 5.11)	β_3	2.61	(0.92, 3.71)
β_4	2.61	(1.00, 3.66)	β_5	2.58	(0.97, 3.62)
$\beta_2 - \beta_3$	1.13	(0.36, 1.97)	$\beta_2 - \beta_4$	1.12	(0.43, 1.96)
$\beta_2 - \beta_5$	1.20	(0.47, 1.95)	$\beta_3 - \beta_4$	-0.01	(-0.42, 0.50)
$\beta_3 - \beta_5$	0.01	(-0.36, 0.52)	$\beta_4 - \beta_5$	0.02	(-0.15, 0.29)

Estimated hazard functions



Predictive survival probabilities

Run the MCMC procedure to estimate the predictive survival probabilities for each interval $(0, i), i = 1, \ldots, 1000$. Time increment is 1 week.



Bayesian regression model

- Historical data from STP for total cost–preventive and corrective
- 13 Explanatory variables–item's characteristics
- Cost data highly skewed to the right
- Bayesian regression model:

$$\log(\text{total cost}_i) = x_{i1}\beta_1 + \cdots + x_{i,13}\beta_{13} + \epsilon_i,$$

where

- $\epsilon_1, \ldots, \epsilon_{5409} | G \overset{iid}{\sim} G$
- $G(t) = w_1 \Phi(t|\mu_1, \tau_1) + w_2 \Phi(t|\mu_2, \tau_2)$
- $\Phi(\cdot|\mu, \tau)$ is the CDF of a $N(0, 1/\tau)$ random variable
- Non-informative priors on $(w_1, \mu_1, \mu_2, \tau_1, \tau_2)$
- Used similar Markov Chain Monte Carlo sampling algorithm to estimate all the parameters

Regression results

Coef.	Est.	95% CI	Coef.	Est.	95% CI
qaqc=1	−0.080	(−0.382, 0.232)	eq=1	−0.248	(−0.374, −0.123)
train=1	0.188	(0.057, 0.322)	arf=1	1.679	(1.308,2.043)
pgi=1	0.071	(−0.015,0.159)	frr=1	0.239	(0.133,0.347)
pgi=2	1.064	(0.907,1.225)	frr=2	0.594	(0.401,0.788)
pgi=3	1.559	(1.201,1.923)	frr=3	1.026	(0.801,1.251)
sc=1	0.080	(−0.237,0.409)	avf=1	−0.249	(−0.400,−0.105)
sc=2	−0.099	(−0.404,0.216)	w_1	0.980	(0.976,0.983)
μ_1	7.628	(7.288,7.966)	τ_1	0.357	(0.343,0.371)
μ_2	0.158	(−0.174,0.489)	τ_2	11.23	(7.83,16.13)

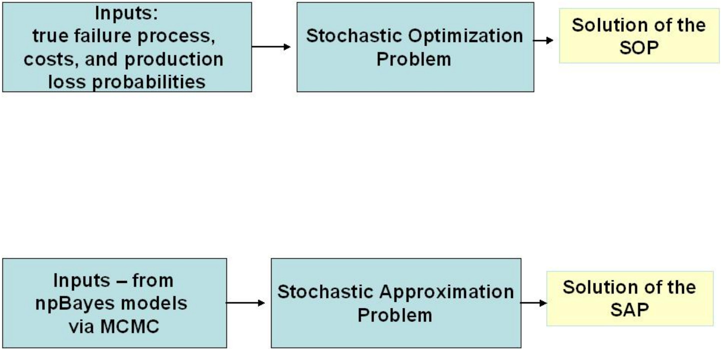
Back to the maintenance optimization problem

- The main goal was to find the optimal T to perform system overhaul.
- Will do this using the Bayesian nonparametric models for failure times and costs.
- Theoretical result - the objective function $z(T)$ is lower semicontinuous with discontinuities at points $L/n, n = 1, 2, \ldots$.
- Given this result and the finite horizon, the minimum is attainable.
- Evaluate the objective function over a grid of points $T = 1, \ldots, L$.

Total cost - stochastic approximation results

- Time scale - weeks, i.e. $i = 1, 2, \ldots$ corresponds to week 1, 2, etc.
- Use the MCMC output from the AFT npBayes model (predictive survival probabilities) to get the expected number of failures in $(0, i)$ as $-\ln S(i)$.
- Obtain the probability of production loss from the STPNOC risk engineers.
- Use the npBayes regression model for the costs to get the cost parameters.

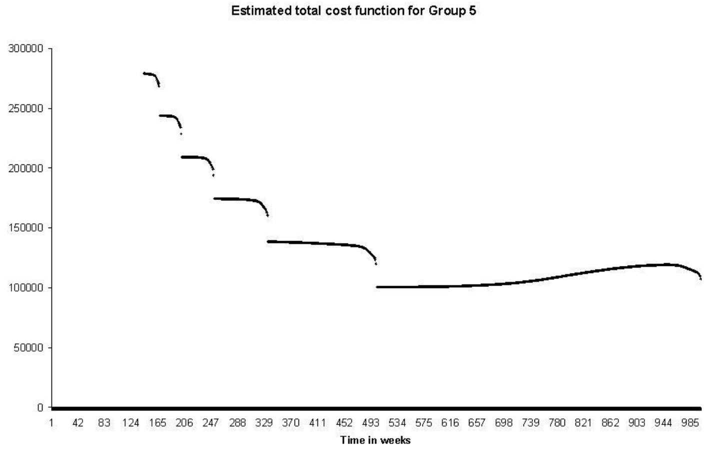
Optimization steps



The minimum cost for each group and the associated time for preventive maintenance T is

Group	Min Cost	T
1	$162,463	1000
2	$281,405	1000
3	$294,209	501
4	$236,987	501
5	$81,630	501

Estimated total cost plot



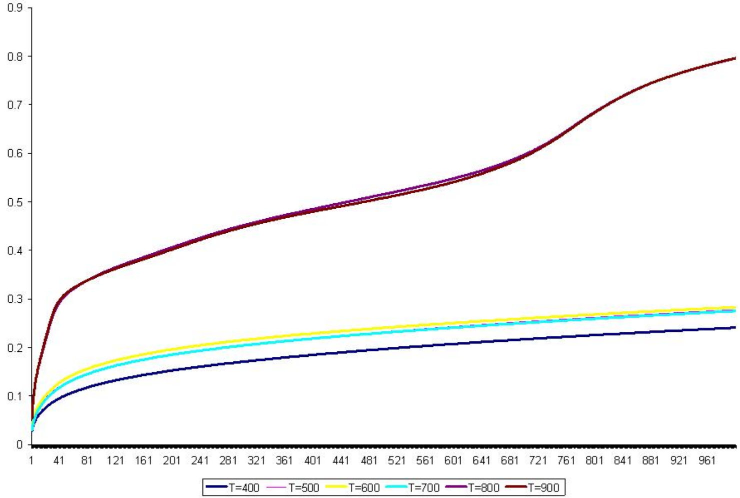
Summary - plots

- The cost function is not monotonic or convex - only grid evaluation will work.
- The hazard functions of Groups 3-5 imply that major overhaul was done in weeks 600-650. Group 1 has a decreasing failure rate. Group 2 is unique: its hazard rate provides evidence that several major overhauls were done to this system.
- “Did these overhauls bring the systems to an as-good-as-new state?”

Simulation Experiments

- Took only the first 400 weeks of data and simulated the next 1000 weeks.
- Repeat this for 500, 600, 700, 800, and 900 weeks.
- At each of these time points we have the future failure behavior of the groups.

Predictive distribution functions for Group 5



- The survival function at censoring times 800 and 900 - different from the functions censored at earlier times.
- Possible conclusion - major overhaul did change the probabilistic behavior of this group.
- An example of a system with *decision dependent* - on the maintenance interventions performed - uncertainties.

Conclusion

- Two Bayesian models - for the failure time distribution, and for the error term in the regression of total costs.
- Some of the hazard plots indicate that major overhauls were performed and their shape is different after that point–an indication of decision dependent change.
- One of the hazards is multi-modal.
- The nonparametric component in the Bayesian model captures important data features: multi-modality and skewness.
- The failure behavior does change with maintenance interventions.

This research has been partially supported by STPNOC grant number 200401743 and the USA National Science Foundation grant number CMMI-0457558.

http://www.me.utexas.edu/ $\sim$ popova