Efficient Online Estimation and Remaining Useful Life Prediction Based on the Inverse Gaussian Process

XU, Ancha

Zhejiang Gongshang University

Joint work with Jingyang Wang (NUS), Yincai Tang (ECNU), Piao Chen (ZJU-UIUC)

Degradation: examples

- Degradation: changes of key performance characteristic over time
- Performance characteristic: capacity, light intensity, wear level, crack









Degradation: more examples

• Performance characteristic: gear vibration, measurement accuracy, printhead ink migration, corrosion





Cross-Section of Inkjet Printhead





Degradation and failure time



Degradation and failure time



Remaining useful life (RUL)



RUL demo



RUL demo



RUL demo



RUL demo



Degradation: applications



Degradation: more applications



Preventive maintenance

Predict warranty costs

Burn-in test

Degradation modelling: a stochastic process

- The degradation level stochastically increases over time.
- Degradation of different units differs.



Monotonic degradation data – examples



Migration data of printheads

Capacitance loss of capacitors

Battery capacity degradation

- Existing stochastic processes: the gamma process and the inverse Gaussian (IG) process
- Physical interpretation: limiting compound Poisson process

The Inverse Gaussian (IG) process

A stationary IG process {𝒱(t), t ≥ 0} is a stochastic process which satisfies the following properties: i) 𝒱(0) = 0 with probability 1; ii) {𝒱(t), t ≥ 0} has independent increments; iii) The increment ΔY_{ts} = 𝒱(t) – 𝒱(s) follows the IG distribution 𝒯(Δt/α, βΔt²) with Δt = t - s > 0. The probability density function (PDF) of an IG distribution 𝒯(a, b), a > 0, b > 0 is

$$\mathcal{E}_{\mathcal{IG}}(u;a,b) = \left(\frac{b}{2\pi u^3}\right)^{1/2} \exp\left[-\frac{b(u-a)^2}{2a^2u}\right], \ u>0.$$

- Non-stationary IG process: introduce the time-scale transformation $\Lambda_{\lambda}(t)$ to the stationary IG process, which is a monotone increasing function of t, and λ is the unknown parameter to be estimated.
- Most existing research is on modelling and offline estimation based on the IG process, very few on online inference and RUL prediction

Problem setting

- Consider *n* systems and the degradation of each system follows a nonstationary IG process.
- Let 0 = t₀ < t₁ < t₂ < ··· < t_m < ··· be the inspection time points, where the degradation levels of all the systems are measured. Let y_{i,j} be the degradation level of the *i*th system at time t_j and Y⁽ⁱ⁾_{0:m} = (y_{i,0},..., y_{i,m}) be the collected degradation observations for the *i*th system up to the time point t_m.
- The degradation data from all the systems up to t_m are denoted as $\mathbf{Y}_{0:m} = (\mathbf{Y}_{0:m}^{(1)}, \dots, \mathbf{Y}_{0:m}^{(n)}).$

Two objectives

- Assume the current time point is t_m . The first task is to estimate $\theta = (\alpha, \beta, \lambda)$ based on $\mathbf{Y}_{0:m}$. The RUL of the system at time t_m is then defined as $\mathcal{X}_m = \inf\{x : \mathcal{Y}(x + t_m) \ge \omega | y_m < \omega\}$, where ω is the failure threshold.
- At the next inspection time point t_{m+1} , the new observations $(y_{1,m+1}, \ldots, y_{n,m+1})$ from the *n* systems become available. Our next objective is to efficiently obtain $\hat{\theta}^{(m+1)}$ by using the new observations, the previous estimates $\hat{\theta}^{(m)}$ and possibly only a few summary statistics based on the historical data **Y**_{0:m}. Afterwards, the *in-situ* RUL prediction can also be performed.

The idea - a composite framework

• If λ is known, the updates of $\hat{\alpha}$ and $\hat{\beta}$ can be derived from their MLEs based on $\mathbf{Y}_{0:m}$

$$\hat{\alpha}^{(m)} = \frac{n\Lambda_{\lambda}(t_m)}{\sum_{i=1}^n y_{i,m}}, \quad \hat{\beta}^{(m)} = \frac{nm}{\sum_{i=1}^n \sum_{j=1}^m \frac{\Delta\Lambda_j^2}{\Delta y_{i,j}} - \frac{n^2\Lambda_{\lambda}^2(t_m)}{\sum_{i=1}^n y_{i,m}}}.$$

• When the new degradation measurements $\mathbf{y}_{m+1} = (y_{1,m+1}, \dots, y_{n,m+1})$ are collected, the update of $\hat{\alpha}^{(m+1)}$ is straightforward

$$\hat{\alpha}^{(m+1)} = \frac{n\Lambda_{\lambda}(t_{m+1})}{\sum_{i=1}^{n} y_{i,m+1}}$$

which does not need any information from $\mathbf{Y}_{0:m}$. In terms of $\hat{\beta}^{(m+1)}$, by decomposing the denominator, we have the recursive formula

$$\hat{\beta}^{(m+1)} = \frac{n(m+1)}{\frac{nm}{\hat{\beta}^{(m)}} + \left[\hat{\alpha}^{(m)}\right]^2 \sum_{i=1}^n y_{i,m} - \left[\hat{\alpha}^{(m+1)}\right]^2 \sum_{i=1}^n y_{i,m+1} + \sum_{i=1}^n \frac{\Delta \Lambda_{m+1}^2}{\Delta y_{i,m+1}}}$$

The idea - a composite framework

- Composite procedure: $\hat{\lambda}^{(m)} \to (\hat{\alpha}^{(m)}, \hat{\beta}^{(m)}) \to \hat{\lambda}^{(m+1)} \to (\hat{\alpha}^{(m+1)}, \hat{\beta}^{(m+1)}) \to \dots$
- How to update $\hat{\lambda}^{(m+1)}$ based on $(\hat{\alpha}^{(m)}, \hat{\beta}^{(m)})$?
 - The profile likelihood does not permit an efficient recursion
- Is there any theoretical guarantee on the composite procedures?
 - Convergence and asymptotic normality may not be easy to establish

Update $\hat{\lambda}$

• One-step estimator: given a preliminary estimator $\hat{ heta}$, the one-step estimator $\hat{ heta}$ is

$$\hat{\theta} = \tilde{\theta} + [I(\tilde{\theta})]^{-1}\dot{L}(\tilde{\theta}),$$

where $I(\cdot)$ is the Fisher information and $\dot{L}(\cdot)$ is the score function.

- If $\tilde{\theta}$ is \sqrt{n} -consistent and the function $\theta \mapsto \dot{L}(\theta)$ satisfies certain differentiability conditions, the one-step estimator $\hat{\theta}$ is \sqrt{n} -consistent and asymptotically efficient.
- \bullet The one-step estimator $\hat{\lambda}^{(m+1)}$ can be derived as

$$\hat{\lambda}^{(m+1)} = \hat{\lambda}^{(m)} + V_{m+1}(\hat{\lambda}^{(m)}) \frac{1}{n} \dot{\mathcal{L}}(\alpha, \beta, \hat{\lambda}^{(m)} | \mathbf{Y}_{0:m+1}),$$

where $V_{m+1}(\lambda)$ is the inverse of Fisher information contributed by $\mathbf{Y}_{0:m+1}$, and $L(\cdot)$ is the likelihood function.

• However, $V_{m+1}(\hat{\lambda}^{(m)})$ cannot be efficiently updated from $V_m(\hat{\lambda}^{(m-1)})$ as I_j 's have to be recalculated for different estimators of λ . Recall that the estimates $\hat{\alpha}^{(m)}$ and $\hat{\beta}^{(m)}$ will be used in $\hat{\lambda}^{(m)}$. Therefore, an approximation to $V_{m+1}(\hat{\lambda}^{(m)})$ at each step can be $\tilde{V}_{m+1} = \left[\sum_{j=1}^{m+1} I_j(\lambda^{(j-1)}|\hat{\alpha}^{(j-1)}, \hat{\beta}^{(j-1)})\right]^{-1}$ and it is easy to see that \tilde{V}_m is recursive because

$$\widetilde{V}_{m+1}^{-1} = \widetilde{V}_m^{-1} + I_{m+1}(\hat{\lambda}^{(m)}|\hat{\alpha}^{(m)},\hat{\beta}^{(m)}).$$

• The recursion for λ can be approximated as

$$\hat{\lambda}^{(m+1)} = \hat{\lambda}^{(m)} + \frac{1}{n} \widetilde{V}_{m+1} \dot{L}(\alpha, \beta, \hat{\lambda}_n^{(m)} | \mathbf{Y}_{0:m+1}),$$

The algorithm

- After collecting the degradation values measured at least three times, the offline estimation procedure is implemented to obtain the initial estimates of the parameters, denoted as $\hat{\alpha}^{(3)}$, $\hat{\beta}^{(3)}$ and $\hat{\lambda}^{(3)}$. Compute $\tilde{V}_3 = \left[\sum_{j=1}^3 I_j(\lambda^{(3)}|\hat{\alpha}^{(3)},\hat{\beta}^{(3)})\right]^{-1}$.
- **2** After the *m*th iteration, $m \ge 3$, when new observations \mathbf{y}_{m+1} is collected, the estimates of α and β are updated by $\hat{\lambda}^{(m)}$. Denote the updated estimates as $\hat{\alpha}^{(m+1)}$ and $\hat{\beta}^{(m+1)}$, respectively.
- **③** Update \widetilde{V}_{m+1} . Then substitute $\hat{\alpha}^{(m+1)}$ and $\hat{\beta}^{(m+1)}$ to obtain $\hat{\lambda}^{(m+1)}$.
- Seperat Steps 3 and 4 until no new observations are collected.

Asymptotic results

Theorem

For every $m \geq 3$, we have that $(\hat{\alpha}^{(m)}, \beta^{(m)}, \hat{\lambda}^{(m)})$ converges to $(\alpha_0, \beta_0, \lambda_0)$ in probability when $n \to \infty$. Furthermore, the estimator sequence $\sqrt{n}\{(\hat{\alpha}^{(m)}, \beta^{(m)}, \hat{\lambda}^{(m)}) - (\alpha_0, \beta_0, \lambda_0)\}$ converges in distribution to a 3-dimensional normal random vector with mean zero and covariance matrix Σ_m , where Σ_m can be recursively updated.

RUL prediction

- Recall the RUL at t_m is defined as $\mathcal{X}_m = \inf\{x : \mathcal{Y}(x + t_m) \ge \omega | y_m < \omega\}$.
- The CDF of \mathcal{X}_m can be readily derived by the equivalence of the two events $\{\mathcal{X}_m < x\}$ and $\{\mathcal{Y}(x + t_m) \ge \omega\}$

$$\begin{aligned} \mathcal{F}_{\mathcal{X}_m}\left(x|y_m\right) &= \mathbb{P}\{\mathcal{Y}(x+t_m) \geq \omega\} = \mathbb{P}\{\mathcal{Y}(x+t_m) - y_m \geq \omega - y_m\} \\ &= \Phi\left(\frac{\sqrt{\beta}\left[\Delta\Lambda_x - \alpha(\omega - y_m)\right]}{\sqrt{\omega - y_m}}\right) - \exp\left(2\alpha\beta\Delta\Lambda_x\right)\Phi\left(-\frac{\sqrt{\beta}\left[\Delta\Lambda_x + \alpha(\omega - y_m)\right]}{\sqrt{\omega - y_m}}\right). \end{aligned}$$

• The estimates $(\hat{\alpha}^{(m)}, \hat{\beta}^{(m)}, \hat{\lambda}^{(m)})$ can then be used to sequentially update the CDF $F_{\mathcal{X}_m}(\cdot)$.

• Other reliability characteristics can also be obtained.

IG random effects model

- The degradation data from different systems can exhibit heterogeneities because of variability of raw materials, fluctuations in the production processes and different operating environments.
- To account for the heterogeneities, the random-effect model has been extensively used in degradation modelling by letting one of the model parameters vary across different systems.
- We let the drift parameter α be a normal random variable, i.e., $\alpha \sim N(\mu, \sigma^2)$, by assuming that $\mu \gg \sigma$ so that the possibility of a negative α is neglectable.
- The unknown parameters are now $\boldsymbol{\theta} = (\beta, \mu, \sigma, \lambda)$.

Update β, μ, σ given λ

- Major difficulty: in presence of random effects, the MLEs of β , μ , σ do not have closed forms, which pose difficulties in developing recursions
- Idea: estimate the missing parameters $\alpha_1, \ldots, \alpha_n$ and then use them to estimate μ and σ .
- Given λ and observed degradation increments $\Delta \mathbf{y}_1, \ldots, \Delta \mathbf{y}_m$, the ML estimators of β and α_i 's are respectively

$$\hat{\beta}^{(m)} = \frac{nm}{\sum_{i=1}^{n} \phi_{i,m}}, \ \hat{\alpha}_i^{(m)} = \frac{\Lambda_{\lambda}(t_m)}{y_{i,m}}, \ i = 1, \dots, n,$$

where $\phi_{i,m} = \sum_{j=1}^{m} \frac{\Delta \Lambda_j^2}{\Delta y_{i,j}} - \frac{\Lambda_{\lambda}^2(t_m)}{y_{i,m}}$.

• Bias correction: because of the excess parameters and the reduced sample size of a single system, these estimators can be highly biased.

Bias correction and recursion of $\hat{\beta}$

- Observe that a unbiased estimator for $1/\beta$ is $T_m = \sum_{i=1}^n \phi_{im}/[n(m-1)]$
- Taylor's expansion gives $\mathbb{E}\left[1/T_m\right] \approx \beta + \mathsf{Var}(T_m)\beta^3$
- An approximate estimator of the variance is

$$\widehat{\operatorname{Var}(T_m)} = \frac{\frac{1}{n-1}\sum_{i=1}^n (\phi_{im} - \overline{\phi}_m)^2}{n(m-1)^2}.$$

 $\bullet\,$ The closed-form estimator of β with bias correction can be derived as

$$\tilde{\beta}^{(m)} = \frac{n(m-1)}{\sum_{i=1}^{n} \phi_{im}} - \frac{\sum_{i=1}^{n} (\phi_{im} - \bar{\phi}_m)^2}{n(n-1)(m-1)^2} \left(\frac{nm}{\sum_{i=1}^{n} \phi_{i,m}}\right)^3$$

• It can be shown $\phi_{i,m}$ is recursive

$$\phi_{i,m+1} = \sum_{j=1}^{m+1} \frac{\Delta \Lambda_j^2}{\Delta y_{i,j}} - \frac{\Lambda_\lambda^2(t_{m+1})}{y_{i,m+1}} = \phi_{i,m} + \frac{y_{i,m+1}\Delta y_{i,m+1}}{y_{i,m}} \left[\frac{\Lambda_\lambda(t_{m+1})}{y_{i,m+1}} - \frac{\Delta \Lambda_{m+1}}{\Delta y_{i,m+1}}\right]^2$$

Bias correction and recursion of $\hat{\alpha}_i$, $\hat{\mu}$ and $\hat{\sigma}$

Note that

$$\mathbb{E}\left[\hat{\alpha}_{i}^{(m)}\right] = \mathbb{E}\left[\mathbb{E}\left[\hat{\alpha}_{i}^{(m)}|\alpha_{i}\right]\right] = \mathbb{E}\left[\alpha_{i} + \frac{1}{\beta\Lambda_{\lambda}(t_{m})}\right] = \mu + \frac{1}{\beta\Lambda_{\lambda}(t_{m})}$$

which implies the bias is $1/(\beta \Lambda_{\lambda}(t_m))$.

 \bullet A closed-form estimator of α with bias correction is

$$\tilde{\alpha}_{i}^{(m)} = \hat{\alpha}_{i}^{(m)} - \frac{1}{\tilde{\beta}^{(m)}\Lambda_{\lambda}(t_{m})} = \frac{\Lambda_{\lambda}(t_{m})}{y_{i,m}} - \frac{1}{\tilde{\beta}^{(m)}\Lambda_{\lambda}(t_{m})}, \ i = 1, \dots, n.$$

• Afterwards,

$$\tilde{\mu}^{(m)} = \frac{1}{n} \sum_{i=1}^{n} \tilde{\alpha}_i^{(m)}.$$

$$\tilde{\sigma}^{(m)} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} \left(\tilde{\alpha}_{i}^{(m)} - \tilde{\mu}^{(m)}\right)^{2} - \frac{2}{\left[\tilde{\beta}^{(m)}\right]^{2}\left[\Lambda_{\lambda}(t_{m})\right]^{4}} - \frac{\tilde{\mu}^{(m)}}{\tilde{\beta}^{(m)}\left[\Lambda_{\lambda}(t_{m})\right]^{3}}}.$$

Update λ given β, μ, σ

• One-step estimator

$$ilde{\lambda}^{(m+1)} = ilde{\lambda}^{(m)} + rac{1}{n} \widetilde{RV}_{m+1} RS_{m+1} (\Delta \mathbf{y}_{m+1}, ilde{\lambda}^{(m)}),$$

where $\widetilde{RV}_{m+1} = \left[\sum_{j=1}^{m+1} Rl_j(\tilde{\lambda}^{(j-1)} | \tilde{\alpha}^{(j-1)}, \tilde{\beta}^{(j-1)})\right]^{-1}$ is the approximation of the inverse of the Fisher information and $RS_j(\Delta \mathbf{y}_j, \lambda)$ is the derivative of the log-likelihood based on the *j*th degradation increments with respect to λ .

• It is easy to verify that \widetilde{RV}_m can be sequentially updated as

$$\widetilde{RV}_{m+1}^{-1} = \widetilde{RV}_m^{-1} + RI_{m+1}(\widetilde{\lambda}^{(m)}|\widetilde{\alpha}^{(m)}, \widetilde{\beta}^{(m)})$$

and $RS_{m+1}(\Delta \mathbf{y}_{m+1}, \tilde{\lambda}^{(m)})$ only depends on $\Delta \mathbf{y}_{m+1}$.

The algorithm

- Use the the first three degradation measurements to implement the offline estimation. Obtain $\left(\tilde{\beta}^{(3)}, \tilde{\alpha}_{1}^{(3)}, \dots, \tilde{\alpha}_{n}^{(3)}, \tilde{\lambda}^{(3)}\right)$ and then obtain $\tilde{\mu}^{(3)}$ and $\tilde{\sigma}^{(3)}$ using the pseudo sample $\tilde{\alpha}^{(3)}$. Compute $\widetilde{RV}_{3} = \left[\sum_{j=1}^{3} Rl_{j}(\tilde{\lambda}^{(3)}|\tilde{\alpha}^{(3)}, \tilde{\beta}^{(3)})\right]^{-1}$.
- After the *m*th iteration, $m \ge 3$, when new observations \mathbf{y}_{m+1} is collected, first update $\phi_{i,m+1}$. Then $\tilde{\beta}^{(m+1)}$, $\tilde{\alpha}^{(m+1)}$, $\tilde{\mu}^{(m+1)}$ and $\tilde{\sigma}^{(m+1)}$ can be iteratively updated by $\lambda^{(m)}$.
- **③** Update \widetilde{RV}_{m+1} . Then substitute $\tilde{\beta}^{(m+1)}$ and $\tilde{\alpha}^{(m+1)}$ to obtain $\lambda^{(m+1)}$.
- Repeat Steps 3 and 4 until no new observations are collected.

RUL prediction

Theorem

Assume that the observed degradation measurements are $0 < y_1 < \cdots < y_m < \omega$ at current time t_m which follows an IG random-effect process with $\alpha \sim N(\mu, \sigma^2)$. The CDF of the RUL \mathcal{X}_m is

$$F_{\mathcal{X}_m}(x|y_m) = \Phi\left(\frac{-\kappa_1\mu_m + \kappa_2}{\sqrt{1 + \kappa_1^2\tau_m}}\right) - \exp\left(\kappa_3\mu_m + \frac{\kappa_3^2\tau_m}{2}\right) \Phi\left(\frac{-\kappa_1\mu_m - \kappa_2 - \kappa_1\kappa_3\tau_m}{\sqrt{1 + \kappa_1^2\tau_m}}\right),$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution, $\mu_m = \frac{\beta \Lambda_\lambda(t_m) + \mu \sigma^{-2}}{\beta y_m + \sigma^{-2}}$, $\tau_m = (\beta y_m + \sigma^{-2})^{-1}$, $K_1 = \sqrt{\beta(\omega - y_m)}$, $K_2 = \frac{\sqrt{\beta} \Delta \Lambda_x}{\sqrt{(\omega - y_m)}}$ and $K_3 = 2\beta \Delta \Lambda_x$.

Setting

- $\alpha = 3$, $\beta = 10$ and a power transformation $\Lambda_{\lambda}(t) = t^2$ with $\lambda = 2$.
- n = 15 and m = 100.
- Random effects: $\alpha \sim N(3, 0.8^2)$, i.e., $\mu = 3$ and $\sigma = 0.8$.

Simulation

No random effects



Simulation

Random effects







6 Time

Δ

8 10

2

10

Simulation

Random effects



Time

10

10

Example

Lithium-ion battery capacity degradation



Example

Online estimation - no random effects



Hundred cycles

Online estimation - random effects



No random effects vs Random effects



Figure: The AIC values and *p*-values for IG models with and without random effects.

Example

Test $\alpha \sim N(\mu, \sigma)$



Hundred cycles

Example

RUL prediction



Figure: RUL prediction of Lithium battery #6 based on the two IG models.

Conclusion

- The IG process is one of the most important degradation processes in degradation modelling.
- Online estimation and RUL prediction based on the IG process have not been well studied in the literature.
- We have proposed for the first time the efficient online estimation methods considering the IG process with and without random effects.
- Compared with the filtering methods commonly used for the Weiner process, our methods are computationally efficient and do not have the impoverishment problems.
- It is possible to be extended to the gamma process.

Related work on degradation

- Liangliang Zhuang, Ancha Xu, and Xiaolin Wang, A prognostic driven predictive maintenance framework based on Bayesian deep learning, Reliability Engineering & System Safety. Vol. 234, 109181, 2023.
- Shirong Zhou, Ancha Xu, Yincai Tang and Lijuan Shen, Fast Bayesian inference of reparameterized gamma process with random effects, IEEE Transactions on Reliability. Vol 73(1), 399-412, 2024.
- Ancha Xu, Binbing Wang, Di Zhu, Jihong Pang and Xinze Lian, Bayesian reliability assessment of permanent magnet brake under small sample size, **IEEE Transactions on Reliability**. 2024. doi: 10.1109/TR.2024.3381072
- Ancha Xu, Jinyang Wang, Yincai Tang and Piao Chen, Efficient online estimation and remaining useful life prediction based on the inverse Gaussian process, **Naval Research Logistics**, Minor revision.

Related work on degradation

- Liangliang Zhuang, Ancha Xu, Guanqi Fang and Yincai Tang, Multivariate inverse Gaussian process with common effects, **Journal of Quality Technology**. Major revision.
- Liangliang Zhuang, Ancha Xu, Yijun Wang and Yincai Tang, Remaining useful life prediction for two-phase degradation model based on reparameterized inverse Gaussian process, **European Journal of Operational Research**. Major revision.
- Ancha Xu, Yijun Wang and Yincai Tang, Remaining useful life prediction for gamma degradation processes: a recursive Bayesian approach, IEEE Transactions on Reliability. Major revision.

Thank You!

Q&A