

MAST2012

Dataanalyse
Jørn Vatn/Januar 2024

Repetisjon

Vedlikehold defineres som en kombinasjon av alle tekniske og administrative aktiviteter, inkludert ledelsesaktiviteter som har til hensikt å opprettholde eller gjenvinne en tilstand som gjør en enhet i stand til å utføre en krevd funksjon. Vi skiller mellom:

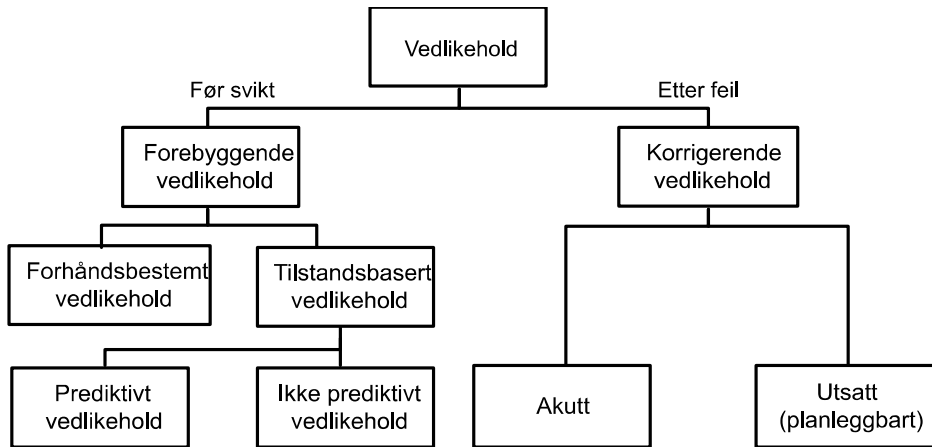
- ▶ *Forebyggende vedlikehold:* Vedlikehold som utføres for å vurdere og/eller minske degraderingen og redusere sannsynligheten for svikt i en enhet
- ▶ *Korrigerende vedlikehold:* Vedlikehold som utføres etter at en feil er funnet, og som har som formål å gjenopprette en enhet til en tilstand der den kan oppfylle krevd funksjon

Forebyggende vedlikehold

For forebyggende vedlikehold skiller vi mellom:

- ▶ *Forhåndsbestemt vedlikehold*: Forebyggende vedlikehold som utføres i henhold til etablerte tidsintervaller eller antall bruksenheter, men uten forutgående tilstandsundersøkelser
- ▶ *Tilstandsbasert vedlikehold*: Forebyggende vedlikehold som omfatter vurdering av fysisk tilstand, analyse og mulige påfølgende vedlikeholdstiltak, hvor vi igjen skiller mellom:
 - ▶ *Prediktivt vedlikehold*: Tilstandsbasert vedlikehold som utføres etter en prognose utledet av gjentatt analyse eller kjente egenskaper og evaluering av de vesentlige parameterne for degradering av enheten
 - ▶ *Ikke prediktivt vedlikehold*: Tilstandsbasert vedlikehold hvor det ikke finnes en prognose for utvikling av degradering

Vedlikehold



Prediktivt vedlikehold

Tilstandsbasert vedlikehold som utføres etter en prognose utledet av gjentatt analyse eller kjente egenskaper og evaluering av de vesentlige parameterne for degradering av enheten. Vi har tre viktige elementer:

1. Tidligdeteksjon (anomaly detection). På et tidlig tidspunkt avdekke avvik fra normalsituasjonen. Utfordringen er å skille støy i måledata fra underliggende trender.
2. Diagose. Avdekke type hvilke deler av systemet som har defekter, og hvilke feilmekanismer som er til stede.
3. Prognose. Etablere modeller som beskriver feilutvikling, og tid til en kritisk tilstand vil oppstå.

Dataanalyse

- ▶ For å bestemme riktig tidspunkt for vedlikehold trenger vi matematiske modeller
- ▶ Slike modeller inneholder modellparametere slik som MTTF, P-F intervall, aldringsparameter i Weibullfordelingen osv
- ▶ Målet med dataanalysen er å fastsette tallverdier for modellparametrene



Eksempel: Periodisk utskifting

- ▶ Anta at en enhet har økende sviktintensitet, dvs badekarskurven øker i høyre del
- ▶ Vi antar at tid-til-svikt for slike enheter er Weibullfordelt med aldringsparameter α og midlere tid til svikt $MTTF = \Gamma(1 + 1/\alpha)/\lambda$
- ▶ Sviktintensiteten er gitt ved

$$z(t) = \alpha \lambda^\alpha t^{\alpha-1} \quad (1)$$

- ▶ Dersom vi erstatter enheten med en enhet som er så god som ny med vedlikeholdsintervall τ kan det vises at effektiv feilrate er gitt ved

$$\lambda_E(\tau) = \left(\frac{\Gamma(1 + 1/\alpha)}{MTTF} \right)^\alpha \tau^{\alpha-1} \quad (2)$$

- ▶ Effektiv feilrate er forventet antall svikt per tidsenhet som funksjon av vedlikeholdsintervallet

Kostnadsmodell

- ▶ Anta nå at kostnaden for å utføre FV-aksjonen (bytte) er c_{FV}
- ▶ og kostnaden ved en ikke-planlagt svikt er c_U
- ▶ Da blir forventede kostnader per tidsenhet:

$$C(\tau) = c_{FV}/\tau + \lambda_E(\tau)c_U = c_{FV}/\tau + \lambda_E(\tau)(c_{KV} + c_{EP})$$

- ▶ hvor c_{KV} er reparasjonskostnadene, og c_{EP} er forventet (expected) produksjonstap ved en svikt

Optimalt intervall

- ▶ Ved å derivere kostnadsfunksjonen mht τ , og sette den deriverte lik 0 finner vi at optimalt intervall er gitt ved:

$$\tau^* = \frac{\text{MTTF}}{\Gamma(1 + 1/\alpha)} \left(\frac{c_{\text{FV}}}{c_{\text{U}}(\alpha - 1)} \right)^{1/\alpha}$$

- ▶ For å finne tallsvar trenger vi tallverdier for MTTF, α , c_{FV} og c_{U} .

Hvilke data har vi?

- ▶ Anta at vi betrakter et lager i en pumpe som har en tendens til å ryke
- ▶ Vi antar videre at vi har samlet inn tid-til-svikt data for slike pumper
- ▶ Vi har to typer observasjoner, ordinere tid-til-svikt data hvor vi har målt tiden fra lageret er satt inn til det svikter, målt f eks i driftstimer
- ▶ Så er det noen datapunkter som ikke representerer svikt, men de kan representere at vi har byttet ut lageret forebyggende, eller at vi har solgt pumpen før lageret havarerte. Slike datapunkter betegnes sensurerte levetider, eller sensurerte tid-til-svikt tider.
- ▶ Observasjonene betegnes $t_1, t_2, \dots, t_n$, hvor vi setter et merke (*) på de sensurerte levetidene

Kaplan Meier plott

- ▶ For grafisk å visualisere observasjonene kan vi lage et Kaplan Meier plott
- ▶ Kaplan Meier plottet viser overlevelsessannsynligheten, dvs $R(t)$
- ▶ Plottet kan nyttiggjøre seg både ordinære tid-til-svikt tider, og sensurerte tid-til-svikt tider
- ▶ Det er hensiktsmessig å sortere observasjonene, og vi lar $t_{(1)}, t_{(2)}, \dots, t_{(n)}$,
- ▶ Vi vil nå lage et plott for observasjonene hvor vi lar x-aksen representere tiden, og y-aksen overlevelsessannsynligheten

Kaplan Meier plott, forts

- ▶ La $n_{(i)}$ være antall komponenter “under risiko”, dvs som fortsatt lever, ved tid $t_{(i)}$ og $s_{(i)}$ være antall svikt ved tid $t_{(i)}$
- ▶ I regelen er $s_{(i)} = 1$, men kan være større enn 1, dersom vi observerer like tid-til-svikt tider (med målenøyaktigheten vi benytter)
- ▶ Dersom $t_{(i)}$ er en sensurert observasjon, er $s_{(i)} = 0$
- ▶ Like før $t_{(i)}$ er det $n_{(i)}$ enheter som fungerer, og like etter er det $n_{(i)} - s_{(i)}$ som fungerer. Den “lokale” sannsynligheten for “å overleve” er da gitt ved:

$$\frac{n_{(i)} - s_{(i)}}{n_{(i)}}$$

- ▶ For å overleve fra 0 til tid t må vi ha hatt “suksess” hver gang, dvs vi kan estimere den “totale” overlevelsessannsynligheten ved:

$$\hat{R}(t) = \prod_{t_{(i)} < t} \frac{n_{(i)} - s_{(i)}}{n_{(i)}}$$

Kaplan Meier plott, eksempel

- ▶ Anta vi har følgende observasjoner:
- ▶ 8,9,7,6,12,18,14,18*,6,9,11, 24,30* og 28*
- ▶ I oppsettet vi bruker så lar vi $I_{(i)}$ være en indikatorvariabel som er lik 1 dersom det er en ordinær observasjon, og lik 0 dersom det er en sensurert observasjon.

Kaplan Meier plott, eksempel

t_i	I_i	n_i	s_i	$(n_i - s_i)/n_i$	$R(t_i)$
6	1	14	2	12/14=0.86	0.86
6	1	13	0	13/13=1	0.86
7	1	12	1	11/12=0.92	0.79
8	1	11	1	10/11=0.91	0.71
9	1	10	2	8/10=0.8	0.57
9	1	9	0	9/9=1	0.57
11	1	8	1	7/8=0.88	0.5
12	1	7	1	6/7=0.86	0.43
14	1	6	1	5/6=0.83	0.36
18	1	5	1	4/5=0.8	0.29
18	0	4	0	4/4=1	0.29
24	1	3	1	2/3=0.67	0.19
28	0	3	0	3/3=1	0.19
30	0	3	0	3/3=1	0.19

Introduksjon til parameterestimering

- ▶ Med estimering mener vi å fastsette verdien på parametere i en stokastisk modell
 - ▶ En estimator er en formel for å estimere en parameter, og er en funksjon av observasjonene
 - ▶ Et estimat er en tallverdi, når vi setter inn verdiene på observasjonene i estimatoren
- ▶ Vi betrakter to situasjoner:
 - ▶ Konstant sviktintensitet
 - ▶ Økende sviktintensitet - Weibullfordeling

MLE = Maximum likelihood estimation

- ▶ La $T_1, T_2, \dots, T_n$ være levetider og sensurerte levetider
- ▶ Store bokstaver benyttes når vi betrakter disse som stokastiske variabler, og små bokstaver benyttes når vi setter inn tallverdiene
- ▶ La $\mathbf{t}$ være en vektor av sammensatt av alle observerte levetider og sensurerte levetider
- ▶ La U være mengden av usensurerte levetider, og la C være mengden av sensurerte levetider
- ▶ La θ være en vektor av ukjente parametere som vi ønsker å estimere
- ▶ Rimelighetsfunksjonen er nå gitt ved:

$$L(\theta; \mathbf{t}) = \prod_{i \in U} f(t_i; \theta) \prod_{i \in C} R(t_i; \theta)$$

MLE = Maximum likelihood estimation, forts

$$L(\theta; \mathbf{t}) = \prod_{i \in U} f(t_i; \theta) \prod_{i \in C} R(t_i; \theta)$$

- ▶ $L(\theta; \mathbf{t})$ representerer sannsynligheten for at vi har observert det vi har gjort i lys av dataene (observasjonene)
- ▶ ML estimatorene er nå gitt ved verdien av som maksimerer $L(\theta; \mathbf{t})$
- ▶ Vi benytter symbolet $\hat{\theta}$ for å angi løsningen av dette optimaliseringsproblemet
- ▶ For å forenkle analysen, er det ofte lettere å maksimere $l(\theta; \mathbf{t}) = \ln L(\theta; \mathbf{t})$, hvor $\ln x$ er den naturlige logaritmen til x

MLE prinsippet for eksponensialfordelingen

- ▶ Anta at vi har observert levetider for en eller flere komponenter
- ▶ La $T_1, T_2, \dots, T_n$ være levetider og sensurerte levetider
- ▶ La x være antall levetider, og $n - x$ være antall sensurerte levetider
- ▶ For eksponensialfordelingen har vi ($\theta = \lambda$):
 - ▶ $f(t) = \lambda e^{-\lambda t}$
 - ▶ $R(t) = e^{-\lambda t}$

MLE prinsippet for eksponensialfordelingen

- ▶ For eksponensialfordelingen har vi ($\theta = \lambda$):
 - ▶ $f(t) = \lambda e^{-\lambda t}$ og $R(t) = e^{-\lambda t}$
- ▶ Setter så inn for eksponensialfordelingen i det generelle uttrykket:
 $L(\theta; \mathbf{t}) = \prod_{i \in U} f(t_i; \theta) \prod_{i \in C} R(t_i; \theta)$, og får: $L(\lambda; \mathbf{t}) = \prod_{i \in U} \lambda e^{-\lambda t_i} \prod_{i \in C} e^{-\lambda t_i}$
- ▶ Dersom vi nå tar logaritmen får vi:

$$l(\lambda; \mathbf{t}) = \ln L(\lambda; \mathbf{t}) = x \ln \lambda - \lambda \sum_{i \in U} t_i - \lambda \sum_{i \in C} t_i$$

- ▶ Siden vi har x svikttider, og $n - x$ sensurerte tider får vi:

$$l(\lambda; \mathbf{t}) = x \ln \lambda - \lambda \sum_{i=1}^n t_i$$

- ▶ Deriverer mht λ , og setter lik 0:

$$\frac{dl(\lambda; \mathbf{t})}{d\lambda} = \frac{x}{\lambda} - \sum_{i=1}^n t_i = 0 \implies \hat{\lambda} = \frac{x}{\sum_{i=1}^n t_i}$$

MLE prinsippet for Weibul

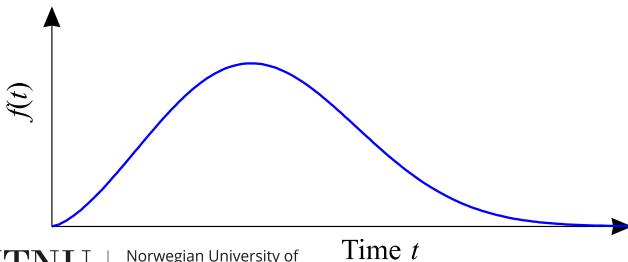
- ▶ Anta at vi har observert levetider for en eller flere komponenter
- ▶ La $T_1, T_2, \dots, T_n$ være levetider og sensurerte levetider
- ▶ La $I_1, I_2, \dots, I_n$ være indikatorer slik at $I_i = 1$, dersom tid nummer i er en levetid, og 0 hvis det er en sensurert levetid
- ▶ Rimelighetsfunksjonen, $L(\alpha, \lambda; \mathbf{t})$, er nå gitt ved produktet av sannsynlighetstetthetene i punktene $t_1, t_2, \dots, t_n$ for levetidene multiplisert med produktet av overlevelsessannsynlighetene i punktene $t_1, t_2, \dots, t_n$ for de sensurerte levetidene
- ▶ Log-rimelighetsfunksjon blir:

$$l(\alpha, \lambda; \mathbf{t}) = \ln L(\alpha, \lambda; \mathbf{t}) = \sum_{i=1}^n I_i [\ln \alpha + \alpha \ln \lambda + (\alpha - 1) \ln t_i] - \sum_{i=1}^n (\lambda t_i)^\alpha$$

▶ Åpne Excel-fil - MLE fanen

Prediktivt vedlikehold og data

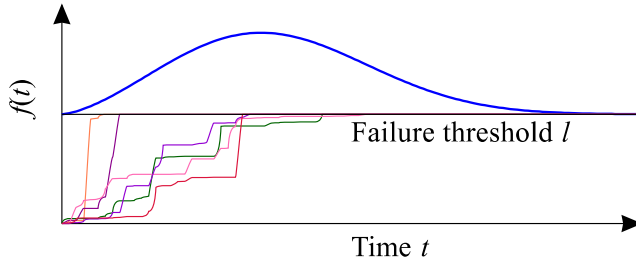
- ▶ I klassisk forutbestemt vedlikehold er det kun alder, driftstid, kjørte kilometer som er utgangspunkt for tidspunkt for vedlikehold, f eks bytte av slidedeler
- ▶ Vi har da kun sannsynlighetstettheten, $f(t)$ som grunnlag for å bestemme intervaller
- ▶ MLE-tilnærmingen gir oss modellparameterene for å angi, og derved sviktintensiteten $z(t)$ og tilhørende effektiv feilrate $\lambda_E(\tau)$:



Prediktivt vedlikehold og data,forts

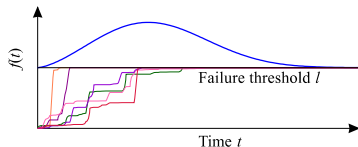
- ▶ I prediktivt vedlikehold er idéen å benytte tilstand som utgangspunkt for å lage prediksjoner for feilutvikling, og tidspunkt for svikt
- ▶ Nedenfor har vi indikert ulike forløp for helseindikatoren
- ▶ Det er viktig å merke seg at forløpet for helseindikatoren, dvs tilstandsindikatoren er stokastisk (tilfeldig)
- ▶ Det betyr også at tid-til-svikt blir tilfeldig
- ▶ Men i motsetning til situasjonen uten tilstandsinforamsjon, så kan vi nå benytte helseindikatoren til å oppdatere sannsynlighetsfordelingen for tid-til-svikt etter hvert som vi får data om helseindikatoren, og usikkerheten i tid-til-svikt reduseres betydelig

Prediktivt vedlikehold og data,forts



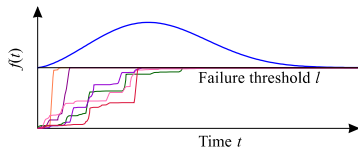
- Merk at forløpet for helseindikatoren, dvs tilstandsindikatoren er stokastisk (tilfeldig)

Prediktivt vedlikehold og data,forts



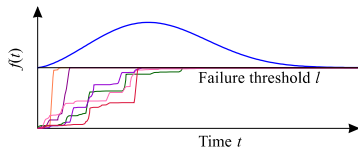
- ▶ La nå $Y(t)$ betegne tilstanden til enheten ved tid t , dvs verdien til helseindikatoren
- ▶ Siden forløpet til helseindikatoren er stokastisk (tilfeldig), vil også endringen i helseindikatoren i løpet av tiden Δt være tilfeldig.
- ▶ La $\Delta Y(t) = Y(t + \Delta t) - y(t)$ være endringen i verdien til helseindikatoren i løpet av tiden Δt
- ▶ Anta nå som en forenkling at $\Delta Y(t) = \Delta Y$ ikke avhenger av hverken t eller verdien helseindikatoren har ved tid t

Prediktivt vedlikehold og data,forts



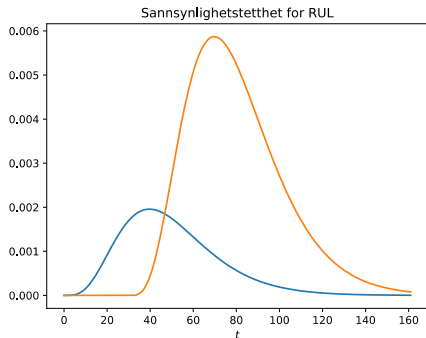
- ▶ Dersom vi har samlet inn data for helseindikatoren for en eller flere enheter over tid, hvor disse enhetene kan betraktes å operere under omtrent de samme operasjonelle betingelser, kan vi beregne inkrementene ΔY_i hvor vi velger en passe tidsenhet for Δt
- ▶ Typisk dersom MTTF er i størrelsesorden ett år, kan vi velge $\Delta t =$ ett døgn, dvs vi måler og loggfører endringer i helseindikatoren dag for dag
- ▶ Dataene vi nå får, dvs ΔY_i -verdiene kan nå benyttes til å estimere modellparameterene til fordelingen for *inkrementene*

Prediktivt vedlikehold og data,forts



- ▶ Når vi nå har estimert modellparameterne i fordelingen til inkrementene, kan vi også beregne sannsynlighetsfordelingen for tid-til-svikt fortløpende etter hvert som vi får informasjon om verdien på helseindikatoren, dvs tilstandsindikatoren
- ▶ Matematisk kan dette være litt utfordrende, og vi vil som regel trenge noen numeriske beregningsrutiner
- ▶ Studenter som ønsker å fortsette med en mastergrad, vil få muligheten til å lære seg slike beregningsmetoder, dvs hvordan man kan predikere RUL med tilhørende usikkerhet

Prediktivt vedlikehold og data, usikkerhet i RUL



- ▶ Den blå kurven viser RUL (Remaining Useful Lifetime) slik vi kan beskrive den ved tid $t = 0$, dvs like etter at vi har installert en ny enhet
- ▶ Den oransje kurven viser RUL ved et senere tidspunkt, og når vi har fått informasjon om tilstanden til enheten

Thank you for your attention

