Weibullfordelingen

Kjetil L. Nielsen

Innhold

1 Teori

Weibullfordelingen er en kontinuerlig sannsynlighetsfordeling som er kanskje best kjent for sine bruksområder i industrien, blant annet for å beregne levetiden til industrielle komponenter [1, 2]. Den brukes også, blant annet, innenfor væranalyse for å beregne vindhastigheter og bølgehøyde [3]. Weibullfordelingen har to paramtere: β, som også kalles formfaktoren, og η, som også kalles skaleringsfaktor. Eksponensialfordelingen er et spesialtilfelle av Weibullfordelingen med $\beta = 1$ og $\eta = 1/\alpha$. Dersom en stokastisk variabel T kan beskrives ved Weibullfordelingen, skriver vi $T \sim \mathcal{W}(\beta, \eta)$.

1.1 Tetthetsfunksjon og fordelingsfunksjon

Tetthetsfunksjonen for Weibullfordelingen er gitt ved

Tetthetsfunksjon for Weibullfordelingen

$$f(t) = \frac{\beta}{\eta^\beta} t^{\beta-1} e^{-(t/\eta)^\beta} \quad t \geq 0, \tag{1}$$

For $t < 0$ definerer vi $f(t) = 0$.

Vi finner fordelingsfunksjonen ved å integrere:

$$F(t) = P(T \leq t) = \int_{-\infty}^{t} f(u)du = \int_{0}^{t} \frac{\beta}{\eta^\beta} u^{\beta-1} e^{-(u/\eta)^\beta} du$$

Ved å gjøre substitusjonen $w = -(u/\eta)^\beta$ får vi $dw = -\frac{\beta}{\eta^\beta} u^{\beta-1} du$ og

$$\int_{0}^{t} \frac{\beta}{\eta^\beta} u^{\beta-1} e^{-(u/\eta)^\beta} du = \int_{w} -e^w dw = \left[-e^w\right]_w = \left[-e^{-(u/\eta)^\beta}\right]_0^t$$

Setter vi inn grensene, ender vi opp med følgende fordelingsfunsjon:

Fordelingsfunksjon for Weibullfordelingen

$$F(t) = 1 - e^{-(t/\eta)^\beta} \quad t \geq 0 \tag{2}$$

Eksempel 1

Anta at bølgehøyden, B, rundt en oljeplatform (målt i meter) kan beskrives ved en Weibullfordeling med parametere $\beta = 1.5$ og $\eta = 2.5$. Finn sannsynligheten for at en tilfeldig bølge vil ha en høyde på

a) mer enn 2.0 m.

b) mellom 1.0 og 3.0 m.

Løsning: Vi har at $B \sim \mathcal{W}(1.5, 2.5)$.

a)

$$P(B > 2.0) = 1 - P(B \leq 2.0) = 1 - F(2.0) = 1 - (1 - e^{-(2.0/2.5)^{1.5}}) \approx \underline{\underline{0.4889}}$$

b)

$$\begin{aligned} P(1.0 < B < 3.0) &= P(B < 3.0) - P(B \leq 1.0) \overset{kont.}{=} P(B \leq 3.0) - P(B \leq 1.0) \\ &= F(3.0) - F(1.0) = 1 - e^{-(3.0/2.5)^{1.5}} - (1 - e^{-(1.0/2.5)^{1.5}}) \\ &\approx \underline{\underline{0.5079}} \end{aligned}$$

1.2 Parameterene i Weibullfordelingen

Figur 1 og 2 viser hvordan formparameteren, β, og skalaparameteren, η, påvirker tetthetsfunksjonen til Weibullfordelingen. Som navnet tilsier, har β en sterk påvirkning på selve karakteristikken til formen av kurven. For $0 < \beta < 1$ vil kurven avta monotont mot null, samt ha en asymptote ved $t = 0$. For $\beta > 1$ har vi $f(0) = 0$ og kurven har en mer bjelleaktig form med et toppunkt. Når $\beta = 1$ har vi i praksis eksponensiellfordelingen (se figur 1). Kurven starter da i $f(0) = 1$.

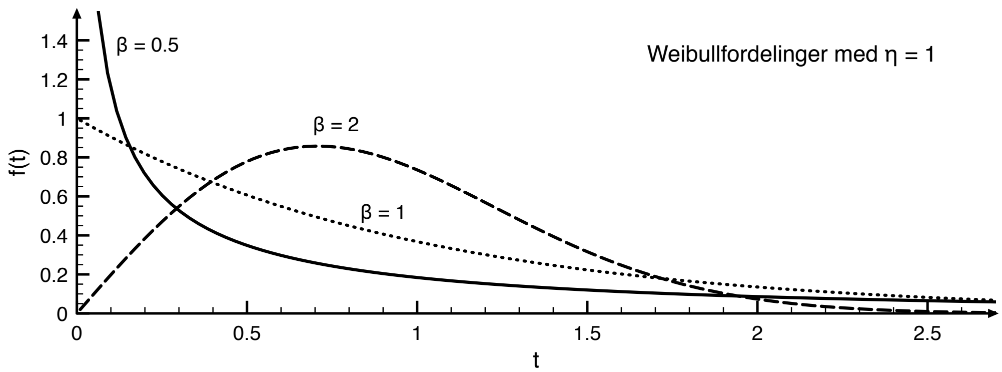

Fig. 1: Tetthetsfunksjonen til Weibullfordelingen for ulike verdier av β.

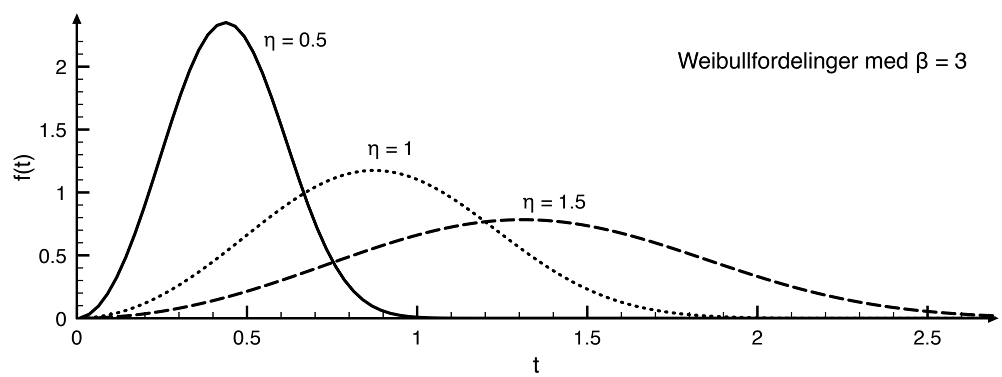

Fig. 2: Tetthetsfunksjonen til Weibullfordelingen for ulike verdier av η.

Skaleringsfaktoren, η, påvirker kurven ved en utstrekking langs t-aksen. Ved større verdier for η, blir kurven bredere og lavere (se figur 2). I levetidsanalyse for komponenter, kalles η ofte den karakteristiske levetiden [1]. Dersom vi setter $t = \eta$, får vi

$$F(\eta) = 1 - e^{-(\eta/\eta)^\beta} = 1 - e^{-1} \approx 0.6321$$

Med andre ord, etter en tid $t = \eta$ har ca. 63% av komponentene sviktet.

Når man ser på kurven til tetthetsfunksjonen etterhvert som β øker i figur 1 i tillegg til å se på hvordan kurven ser ut når $\beta = 3$ (figur 2), så ser det ut som at kurven blir mer og mer lik tetthetsfunksjonen til en normalfordeling. Fortsetter vi å øke verdien for β vil likheten med normalfordelingen etterhvert minke da kurven går fra å være høyretung, alstå at kurven har en lengre hale mot venstre, (se kurven $\beta = 2$ i figur 1) til å bli venstretung (hale mot høyre). Når $2.6 < \beta < 3.7$ vil kurven være tilnærmet symmetrisk og Weibullfordelingen er tilnærmet lik normalfordelingen [4].

1.3 Forventning og varians

La T være en stokastisk variabel gitt ved $T \sim \mathcal{W}(\beta, \eta)$. Vi finner forventingsverdien til T ved å ta utgangspunkt i definisjonen av forventingsverdi for en kontinuerlig variabel:

$$\mu = \int_{-\infty}^{\infty} t \cdot f(t)dt = \int_{0}^{\infty} t \cdot \frac{\beta}{\eta^\beta} t^{\beta-1} e^{-(t/\eta)^\beta} dt = \int_{0}^{\infty} \frac{\beta}{\eta^\beta} t^\beta e^{-(t/\eta)^\beta} dt$$

Dersom vi gjør substitusjonen $u = (t/\eta)^\beta$, får vi

$$\mu = \int_{0}^{\infty} \eta \; u^{1/\beta} e^{-u} du$$

Vi kan gjøre dette uttrykket mer kompakt ved å bruke gammafunksjonen som for $z \geq 0$ er definert ved

$$\Gamma(z) = \int_{0}^{\infty} t^{(z-1)} e^{-t} dt$$

Dersom vi setter $z = 1 + \frac{1}{\beta}$, får vi

Forventingsverdi for Weibullmodell

$$\mu = \eta \cdot \Gamma(1 + \frac{1}{\beta}) \tag{3}$$

Dersom z er et heltall, vil gammafunksjonen ta formen

$$\Gamma(z) = (z-1)! \quad z = 1, 2, 3, \ldots$$

noe som gjør utregningene mye enklere. Dersom z ikke er et heltall, finner vi verdien av gammafunksjonen fra tabeller (se formelheftet).

Vi finner også variansen til T utifra definisjonen:

$$\sigma^2 = \int_{-\infty}^{\infty} (t-\mu)^2 f(t)dt$$

Dersom vi bruker at $\text{Var}(T) = E(T^2) - \mu^2$, får vi

$$\sigma^2 = \int_{-\infty}^{\infty} t^2 f(t)dt - \mu^2 = \int_0^{\infty} t^2 \frac{\beta}{\eta^\beta} t^{\beta-1} e^{-(t/\eta)^\beta} dt - \left(\eta \cdot \Gamma(1+\frac{1}{\beta})\right)^2$$

Vi gjør substitusjonen $u = (t/\eta)^\beta$ og får

$$\begin{aligned}\sigma^2 &= \int_0^{\infty} \eta^2 u^{2/\beta} e^{-u} du - \left(\eta \cdot \Gamma(1+\frac{1}{\beta})\right)^2 \\ &= \eta^2 \left[\int_0^{\infty} u^{2/\beta} e^{-u} dt - \Gamma(1+\frac{1}{\beta})^2\right]\end{aligned}$$

Dersom vi bruker gammafunksjonen med $z = 1 + \frac{2}{\beta}$, får vi

Varians for Weibullmodell

$$\sigma^2 = \eta^2 \left(\Gamma(1+\frac{2}{\beta}) - \Gamma(1+\frac{1}{\beta})^2\right) \tag{4}$$

Eksempel 2

Anta at vi har to stokastiske variable $T_1 \sim \mathcal{W}(1, \eta)$ og $T_2 \sim \exp(\alpha)$. Siden T_2 følger eksponensialfordelingen, har vi $E(T_2) = 1/\alpha$ og $\text{Var}(T_2) = 1/\alpha^2$. Ta utgangspunkt i $E(T_1)$ og $\text{Var}(T_1)$, og vis at $E(T_1) = E(T_2)$ og $\text{Var}(T_1) = \text{Var}(T_2)$.

Løsning: Vi fant tidligere at

$$\begin{aligned}\mu &= \eta \cdot \Gamma(1+\frac{1}{\beta}) \\ \sigma^2 &= \eta^2 \left(\Gamma(1+\frac{2}{\beta}) - \Gamma(1+\frac{1}{\beta})^2\right)\end{aligned}$$

Siden $T_1 \sim \mathcal{W}(\beta = 1, \eta)$, setter vi $\beta = 1$ og får

$$E(T_1) = \eta \cdot \Gamma(1+\frac{1}{1}) = \eta \cdot \Gamma(2) = \eta \cdot (2-1)! = \eta$$

Dersom vi setter $\eta = 1/\alpha$, får vi at $E(T_1) = E(T_2)$.

Vi setter så $\beta = 1$ for variansen og får

$$\begin{aligned}\mathrm{Var}(T_1) &= \eta^2 \left(\Gamma(1 + \frac{2}{1}) - \Gamma(1 + \frac{1}{1})^2 \right) \\ &= \eta^2 \left(\Gamma(3) - \Gamma(2)^2 \right) = \eta^2 \left((3-1)! - [(2-1)!]^2 \right) \\ &= \eta^2 (2-1) = \eta^2\end{aligned}$$

Igjen, dersom vi setter $\eta = 1/\alpha$, får vi at $\mathrm{Var}(T_1) = \mathrm{Var}(T_2)$. Dette er ikke et overraskende resultat siden vi innledningsvis poengterte at eksponensialfordelingen er et spesialtilfelle av Weibullfordelingen med $\eta = 1/\alpha$.

1.4 Utvidet Weibullfordeling

Det finnes også en utvidet Weibullfordeling som inkluderer en tredje parameter γ. Vi skriver da $T \sim \mathcal{W}(\beta, \eta, \gamma)$. Tetthetsfunksjonen for den utvidede Weibullfordelingen er:

Tetthetsfunksjon for utvidet Weibullfordeling

$$f(t) = \frac{\beta}{\eta^\beta}(t-\gamma)^{\beta-1} e^{-((t-\gamma)/\eta)^\beta} \quad t \geq \gamma \tag{5}$$

Legg merke til at tetthetsfunksjonen nå starter på $t = \gamma$. Dersom vi velger $\gamma = 0$ ender vi opp med den tradisjonelle Weibullfordelingen. Siden γ angir hvor fordelingen starter, kaller vi γ for lokasjonsparameteren.

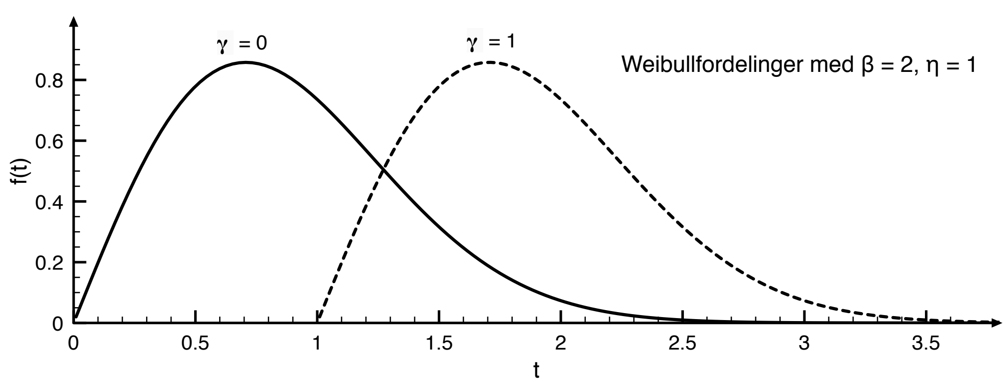

Fig. 3: Tetthetsfunksjonen til Weibullfordelingen for ulike verdier av γ.

1.5 Weibull-plott

Weibull-plott er en metode for å undersøke om et datasett følger en toparameters Weibull-fordeling ($\gamma = 0$). Vi kan også bruke metoden til å estimere parameterene β og η. Metoden går ut på å skrive om fordelingsfunksjonen $F(t)$ slik at vi ender opp med en lineær funksjon på formen $y = mx + b$. Har vi et dataset, kan vi dermed gjøre en lineær regresjon for å sjekke hvor godt datasettet passer til denne rette linjen. Dersom datasettet passer til den rette linjen, kan vi konkludere med at fenomentet vi undersøker følger en Weibullfordeling.

Vi starter med fordelingsfunksjonen:

$$
\begin{aligned}
F(t) &= 1 - e^{-(t/\eta)^\beta} \\
1 - F(t) &= e^{-(t/\eta)^\beta}
\end{aligned}
$$

Vi tar så logaritmen på begge sider og får

$$
\begin{aligned}
\ln\left(1 - F(t)\right) &= -(t/\eta)^\beta \\
\ln \frac{1}{1 - F(t)} &= (t/\eta)^\beta
\end{aligned}
$$

Vi tar logaritmen på begge sider enda en gang og får

$$\ln\ln \frac{1}{1 - F(t)} = \beta \ln(t) - \beta \ln(\eta)$$

Dersom vi skriver $y = \ln\ln \frac{1}{1-F(t)}$, $x = \ln(t)$ og $b = -\beta\ln(\eta)$, får vi

$$y = \beta x + b$$

som er likningen for en rett linje med β som stigningstall. Dersom vi plotter $\ln\ln \frac{1}{1-F(t)}$ mot $\ln(t)$ for et dataset og ser at grafen blir en tilnærmet rett linje, kan vi konkludere med at fenomentet følger en Weibullfordeling.

Det er likevel en liten hake. Det er lett å regne ut $\ln(t)$, men for å regne ut $\ln\ln \frac{1}{1-F(t)}$, trenger vi $F(t)$. For å estimere $F(t)$ fra et dataset, kan vi bruke Bernards approksimasjon [5]. Anta at vi ser på n industrielle komponenter. La t_i være tiden det tok før komponent nr. i sviktet, hvor vi har sortert komponentene slik at $i = 1$ er den første komponenten som sviktet, $i = 2$ den andre komponenten som sviktet osv. Bernards approksimasjon sier da at

$$\widehat{F(t_i)} = \frac{i - 0.3}{n + 0.4}$$

1.5.1 Godhetsmål for lineær regresjon

Ved å utføre en lineær regresjon, kan man finne en rett linke som "passer best" til de dataene vi har. Vi kan dermed bruke den estimerte regresjonslinjen til å estimere en verdi for parameterene i Weibullfordelingen. Dette høres problemfritt ut, men for å avgjøre om dataene følger en Weibullfordeling, må vi først avgjøre om dataene faktisk er lineære (etter omskrivingen beskrevet i forrige avsnitt). For å avgjøre dette, kan vi bruke en såkalt "goodness-of-fit"-test. Vi skal ikke gå inn på dette temaet her, men det kan vises at korrelasjonskoeffisienten kvadrert, R^2, kan brukes som en "goodness-of-fit"-mål for vår regresjonsmodell [6]. Koeffisienten, R^2, vil være et tall mellom 0 og 1 der $R^2 \approx 1$ betyr at dataene passer veldig godt til vår estimerte regresjonslinje. Dersom $R^2 > 0.65$ sier vi at dataene passer "godt nok" til regresjonslinjen og vi konkluderer med at Weibullfordelingen er en god sannsynlighetsmodell for å beskrive det observerte fenomenet.

Eksempel 3

Vi observerte 9 industrielle komponenter og målte tiden det tok før hver av dem sviktet. Vi målte følgende tider (målt i timer):

$$1652.5,\ 2023.5,\ 2361.7,\ 2893.5,\ 3053.4,\ 3155.8,\ 3939.6,\ 4139.8,\ 4526.4$$

Vi bruker Bernards approksimasjon for å beregne $F(t)$ for hvert datapunkt. I tillegg regner vi ut $\ln\ln\frac{1}{1-\widehat{F(t_i)}}$ og $\ln(t_i)$:

Komp.	t_i	$\widehat{F(t_i)}$	$\ln t_i$	$\ln\ln\frac{1}{1-\widehat{F(t_i)}}$
1	1652.5	0.0745	7.4100	-2.5585
2	2023.5	0.1809	7.6126	-1.6117
3	2361.7	0.2872	7.7671	-1.0831
4	2893.5	0.3936	7.9702	-0.6927
5	3053.4	0.5000	8.0240	-0.3665
6	3155.8	0.6064	8.0570	-0.0670
7	3939.6	0.7128	8.2788	0.2212
8	4139.8	0.8191	8.3284	0.5364
9	4526.4	0.9255	8.4177	0.9543

Vi plotter $\ln\ln\frac{1}{1-\widehat{F(t_i)}}$ mot $\ln(t_i)$ og gjør en lineær regresjon ved hjelp av minste-kradratersmetode i tillegg til å regne ut korrelasjonskoeffisienten, R. Resultatene kan sees i figur 4. Minste-kvadraters-metode gir oss $R = 0.990647$ (som gir $R^2 = 0.98139$). Det betyr at regresjonslinjen passer veldig godt til dataene, og vi kan konkludere at Weibull fordelingen er en passende modell for å beskrive fenomenet dataene kommer fra. Stigningstallet til den estimerte regresjonslinjen, er 3.2354, som da også blir et estimat for β.

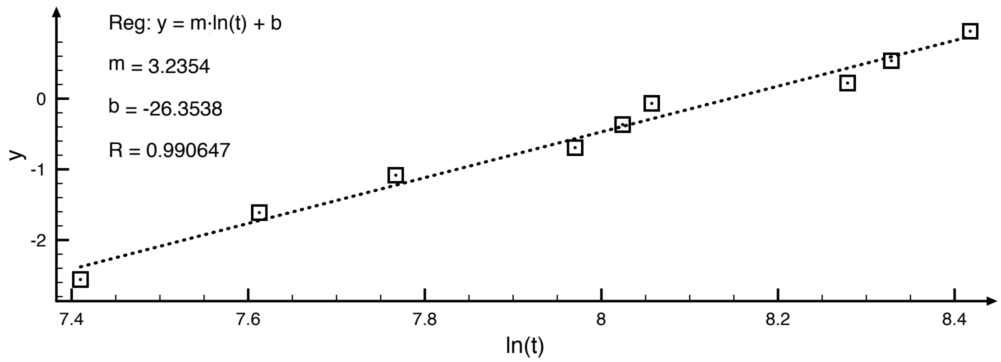

Fig. 4: Lineær regresjon for $y = \ln\ln\frac{1}{1-\widehat{F(t_i)}}$ og $\ln(t)$.

Regresjonen gav også $b = -26.3538$. Vi fant tidligere ut at $b = -\beta\ln(\eta)$. En omskriving av denne, gir oss $\eta = e^{-b/\beta} \approx e^{-(-26.3538)/3.2354} \approx 3447.4$. For å oppsummere:

$$\hat{\beta} = 3.2354$$
$$\hat{\eta} = 3447.4$$

2 Oppgaver

1. Anta at vi har en stokastisk variabel $X \sim \mathcal{W}(2,5)$. Finn sannsynligheten for at en tilfeldig måling av X er

   a) er mindre enn 2.

   b) minst 1.

   c) mellom 3 og 4.

2. Anta at bølgehøyden (målt i meter) for et havområde som vurderes for en oljeplatform har en Weibullfordeling med $\beta = 2.5$ og $\eta = 11$.

   a) Finn sannsynligheten for at bølgehøyden vil være høyere enn 14 m.

   b) Finn forventet bølgehøyde for dette havområdet.

3. Vi ser på 7 komponenter og måler tiden det tar før de svikter. Tidene ble målt til (i timer):

$$66,\ 112,\ 188,\ 242,\ 542,\ 789,\ 817$$

a) Bruk måledataene til å undersøke om Weibullfordelingen er en passende fordeling for å beskrive levetiden til komponentene.

b) Estimer parameterene i denne Weibullfordelingen.

c) Bruk Weibullmodellen til å estimere forventet levealder til en tilfeldig komponent.

4. Vi lar 100 komponenter stå i 200 timer. Etter 200 timer har 8 komponenter sviktet. Vi markerer tiden det tok før de sviktet (målt i timer):

$$18.6, 57.0, 69.9, 127.6, 149.9, 163.3, 180.9, 199.7$$

a) Bruk måledataene til å undersøke om eksponensialfordelingen er en passende fordeling for å beskrive levetiden til komponentene.

b) Estimer parameteren, α, i denne eksponensialfordelingen.

c) Beregn sannsynligheten for at en tilfeldig komponent lever i mindre enn 200 timer.

d) I vårt forsøk sviktet 8 av 100 komponenter i løpet av 200 timer. Beregn sannsynligheten for dette skulle skje (anta at komponentene svikter uavhengig av hverandre).

Se neste side for fasit.

3 Fasit

1. a) 0.148
   b) 0.961
   c) 0.170
2. a) 0.16
   b) 9.76 m
3. a) Ja, siden $R^2 = 0.95$
   b) $\hat{\beta} = 1.0744$, $\hat{\eta} = 440.45$
   c) ca. 428 timer.
4. a) Ja, siden $R^2 = 0.977 \approx 1$ og $\hat{\beta} = 1.00073197 \approx 1$.
   b) $\hat{\alpha} \approx 3.489 \cdot 10^{-3}$
   c) 0.0674
   d) 0.129

Referanser

[1] Alf Harbitz (2001) *Statistikk og sannsynlighetsregning*, Fagbokforlaget.

[2] Richard A. Johnson (1994) *Miller & Freund's Probability & Statistics For Engineers*, Prentice-Hall Inc. Englewood Cliffs, New Jersay 07632

[3] David M. Levine, Patricia P. Ramsey & Robert K. Smidt (2001) *Applied Statistics for Engineers and Scientists*, Prentice-Hall Inc. Uppper Saddle River, New Jersey 07458.

[4] `http://reliawiki.org/index.php/Distributions_Used_in_Accelerated_Testing`

[5] WARWICK MANUFACTURING GROUP (2007) *The use of Weibull in defect data analysis*, University of Warwick.

[6] Gunnar G. Løvås (2013) *Statistikk for universiteter og høgskoler* Universitetsforlaget AS