

# Weibullanalys

Jan Enger  
Matematisk statistik  
KTH

Weibull-fördelningen är en mycket viktig fördelning inom tillförlitlighetsanalysen. Den används ofta för att modellera mekaniska komponenters livslängder. Weibullfördelningen uppkommer bl.a. som en gränsfördelning till minimum av ett stort ("oändligt") antal oberoende stokastiska variabler. I praktiska sammanhang kan ofta en livslängd uppfattas som ett sådant minimum. Uttrycket "en kedja är inte starkare än sin svagaste länk" återspeglar ett sådant resonemang.

**Definition** En stokastisk variabel  $X$  är weibullfördelad med skalparameter  $a$  och formparameter  $c$  om dess överlevnadsfunktion ges av

$$R(t) = e^{-(t/a)^c} \text{ för } t \geq 0$$

Täthetsfunktionen ges av  $f(t) = \frac{ct^{c-1}}{a^c} e^{-(t/a)^c}$  och felintensiteten är således  $\lambda(t) = \frac{ct^{c-1}}{a^c}$ . Weibullfördelningen är alltså IFR om  $c \geq 1$  och DFR om  $c \leq 1$ . Exponentialfördelningen är en weibullfördelning med formparameter  $c = 1$ .

Om man sätter  $b = 1/a^c$  ser man att överlevnadsfunktionen kan skrivas

$$R(t) = e^{-bt^c} \text{ för } t \geq 0$$

Om nu form- och skalparametrar  $a$  respektive  $c$  är okända måste dessa skattas. Vi skall här se på två sätt att skatta parametererna, varav det ena sättet är grafiskt.

## Maximum-likelihoodskattning

Vi antar att det föreligger ett censurerat Typ II stickprov på livslängder av komponenter med weibullfördelade livslängder, dvs vi antar att  $x_{(1)}, x_{(2)}, \dots, x_{(r)}$  är de ordnade livslängderna för de komponenter som brustit vid en undersökning av  $n$  komponenter, ordnade efter storlek. Undersökningen avbryts alltså då  $r$  komponenter gått sönder. Vi antar provningen sker utan återläggning.

Med  $b = (1/a)^c$  har vi alltså att överlevnadsfunktionen är  $R(t) = \exp(-bt^c)$  och att täthetsfunktionen är  $f(t) = bct^{c-1} \exp(-bt^c)$

Likelihoodfunktionen skall vara "troligheten" att få det stickprov man faktiskt erhållit. "Troligheten" för en observerad livslängd  $x_i$  är täthetsfunktionen värde i denna punkt,  $f(x_i)$ . "Troligheten" för en komponent som inte brustit vid tidpunkt  $x_{(r)}$  är sannolikheten att livslängden just är större än  $x_{(r)}$ , dvs  $R(x_{(r)})$ . Eftersom observationerna är oberoende blir likelihoodfunktionen produkten av dessa "troligheter"

$$L(b, c; x_{(1)}, x_{(2)}, \dots, x_{(r)}) = \left( \prod_{i=1}^r (bcx_{(i)}^{c-1} \exp(-bx_{(i)}^c)) \right) (\exp(-bx_{(r)}^c))^{n-r} = \\ b^r c^r \exp(-b \sum_{i=1}^r x_{(i)}^c - b(n-r)x_{(r)}^c) \prod_{i=1}^r x_{(i)}^{c-1}$$

Vi skall söka de värden på  $b$  och  $c$  som maximerar  $L(b, c; x_{(1)}, x_{(2)}, \dots, x_{(r)})$ , och det är då enklare att i stället betrakta logaritmen av  $L$  som maximeras för samma värden.

$$\ln(L) = r \ln(b) + r \ln(c) + (c-1) \sum_{i=1}^r \ln(x_{(i)}) - b \sum_{i=1}^r x_{(i)}^c - b(n-r)x_{(r)}^c$$

Derivera med avseende på  $b$  och  $c$ , och sätter derivatorna lika med 0.

$$\frac{\delta \ln(L)}{\delta b} = \frac{r}{b} - \sum_{i=1}^r x_{(i)}^c - (n-r)x_{(r)}^c = 0 \quad (1)$$

$$\frac{\delta \ln(L)}{\delta c} = \frac{r}{c} + \sum_{i=1}^r \ln(x_{(i)}) - b \sum_{i=1}^r x_{(i)}^c \ln(x_{(i)}) - b(n-r)x_{(r)}^c \ln(x_{(r)}) = 0 \quad (2)$$

Ur den första löser vi lätt ut  $b$ .

$$b = \frac{r}{\sum_{i=1}^r x_{(i)}^c + (n-r)x_{(r)}^c} \quad (3)$$

Vi sätter in detta värde i (2) och erhåller

$$\frac{\delta \ln(L)}{\delta c} = \frac{r}{c} + \sum_{i=1}^r \ln(x_{(i)}) - \frac{r \{ \sum_{i=1}^r x_{(i)}^c \ln(x_{(i)}) + (n-r)x_{(r)}^c \ln(x_{(r)}) \}}{\sum_{i=1}^r x_{(i)}^c + (n-r)x_{(r)}^c} = 0 \quad (4)$$

Denna sista ekvation kan vi inte lösa analytiskt utan får beräknas numeriskt för att på så sätt erhålla  $c$ . Detta ger ML-skattningarna av  $b$  och  $c$ . ML-skattningen av  $a$  fås sedan ur relationen mellan  $a$ ,  $b$  och  $c$ ,  $a = 1/b^{1/c}$ .

Om man har ett fullständigt censurerat stickprov ser man lätt att ML-skattningarna fås om man i ekvationerna ovan sätter  $r = n$ , man provar ju tills alla  $n$  komponenter brustit. Formeluttrycken blir något enklare eftersom  $n - r = 0$  i detta fall, och i ekvationerna (3) och (4) försvinner termerna med faktorn  $n - r$ .

**Ann:** Analysen ovan är gjord under förutsättningen att alla komponenter startas samtidigt och att således censureringen sker vid tidpunkten  $x_{(r)}$ . Så behöver inte vara fallet. Om de startas vid olika tidpunkter så är de censurerade

tiderna inte lika. Genom en enkel modifiering av härledningen ovan kan man se att ekvationerna (3) och (4) ersätts av

$$b = \frac{r}{\sum_{i=1}^n x_i^c} \quad (5)$$

och

$$\frac{\delta \ln(L)}{\delta c} = \frac{r}{c} + \sum_{i; x_i \text{ ocensurerad}} \ln(x_i) - \frac{r \left\{ \sum_{i=1}^n x_i^c \ln(x_i) \right\}}{\sum_{i=1}^n x_i^c} = 0 \quad (6)$$

där  $r$  är antalet ocensurerade tider och  $x_i$  är  $i$ :te komponentens livslängd, censurerad eller ocensurerad.

## Weibullpapper

Man kan på ett enkelt sätt använda s.k. Weibullpapper för att skatta parametrarna  $a$  och  $c$ .

Idén med fördelningpapper är att skala om axlarna i ett diagram så att fördelningsfunktionen för fördelningsfamiljen ifråga blir en rät linje. Det betyder att om man plottar data mot lämpliga skattningar av fördelningsfunktionen skall punkterna anpassa sig till en rät linje. Gör de inte det, kan det tyda på att data inte kommer från den förmodade fördelningsfamiljen.

Låt oss se på hur weibullpapper är konstruerade. Fördelningsfunktionen är ju

$$y = F(x) = 1 - e^{-(x/a)^c} \text{ för } x \geq 0 \quad (7)$$

dvs

$$1 - F(x) = e^{-(x/a)^c} \text{ för } x \geq 0 \quad (8)$$

Logaritmera (8) två gånger. Då erhåller vi

$$\ln(-\ln(1 - F(x))) = c \ln(x) - c \ln(a) \quad (9)$$

Med  $v = \ln(-\ln(1 - y)) = \ln(-\ln(1 - F(x)))$  och  $u = \ln(x)$  kan denna ekvation skrivas

$$v = cu - c \ln(a) \quad (10)$$

vilken är en rät linje i  $(u, v)$ -planet. Axlarna kan sedan skalas i de ursprungliga variablerna  $x$  och  $y$ , dvs  $x$ -axeln har logartimisk skala och  $y$ -axeln dubbellogaritmisk. Lutningen av linjen i  $(u, v)$ -skalan är lika med formparametern  $c$ .

Om nu det föreligger  $n$  observationer från en weibullfördelning plottar vi dessa  $x$ -observationer mot lämpliga värden på  $y$ -axeln. Vi antar först att data inte är censurerade. Den punkt, som det  $i$ :te minsta värdet  $x_{(i)}$  skall plottas mot, borde vara fördelningsfunktionens värde  $F(x_{(i)})$ , sannolikheten att få en

observation mindre eller lika med  $x_{(i)}$ . Men vi känner ju inte  $F$ , utan måste plotta mot en lämplig skattning istället. En sådan skulle kunna vara  $\frac{i}{n}$  som är den empiriska fördelningsfunktionens värde i observationen.  $x_{(i)}$  plottas då mot relativa antalet observationer som är mindre eller lika med  $x_{(i)}$ . Detta förfarande kan dock medföra vissa anomalier. Den största observationen  $x_{(n)}$  skall då plottas mot 1, som kommer att ligga oändligt högt upp på  $v$ -axeln. Bättre är att plotta  $x_{(i)}$  mot medianrangen. Medianrangen kan fås i bifogade tabell. Gå in på aktuellt  $n$ , antal observationer, och plotta  $x_{(i)}$  mot det värde  $u_i^{(n)}$  som ges av det  $i$ :te värdet i kolumnen. Med mycket god approximation är den  $i$ :e medianrangen lika med  $\frac{j - 0.3}{n + 0.4}$ . I nästa avsnitt finns en förklaring till medianranger.

Efter att ha plottat värdena, anpassa en rät linje till data. Sedan är det dags att skatta parametrarna. Enklast är att skatta formparametern  $a$ . Om  $x = a$  är ju fördelningsfunktionens värde  $1 - e^{-1} \approx 0.6321$  oberoende av vad  $c$  är. Vi ser helt enkelt efter vid vilket  $x$ -värde den anpassade kurvan skär linjen  $y = 1 - e^{-1} \approx 0.63$ . Den finns angiven som hjälplinje.

Den anpassade linjen har som sagt lutning  $c$  i de logaritmerade koordinaterna,  $(u, v)$ . Parallellflytta nu linjen så att den går genom origo  $(0,0)$  i de logaritmerade variablerna  $(u, v)$ , det betyder punkten  $(1, 1 - e^{-1})$  i  $(x, y)$ -skalan. Denna punkt är inringad i bifogade weibullpapper. Den parallellflyttade linjen är  $v = cu$ . Om vi i denna linje sätter  $u = -1$  fås  $v = -c$ . Linjen  $u = -1$  finns som hjälplinje och linjärt skalad d.v.s. i  $v$ -skalan. På denna kan man sedan läsa av  $c$ .

Antag nu att vi har censurerade data precis som när vi betraktade ML-skattning. Det är lätt att skaffa sig en weibull-plott även i detta fall. De observerade livslängderna är ju de första  $r$  i en tänkt, men ofullbordad, serie av  $n$  stycken livslängder. Plotta alltså de observerade livslängderna mot de  $r$  första medianrangerna i serien av ranger som motsvarar  $n$ . Beräkna sedan skattningarna av  $a$  och  $c$  på samma sätt ovan.

## Medianranger

Vi antar att  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  är ett ordnat stickprov från en stokastisk variabel med fördelningsfunktion  $F$ . I ett  $F$ -fördelningspapper skall vi plotta observationen  $x_{(i)}$  mot fördelningsfunktionens värde i denna punkt, dvs egentligen mot  $F(x_{(i)})$ . Men  $F$  är okänd, så vi får uppskatta detta värde på lämpligt sätt. Vi konstaterar att  $P(F(X_{(i)}) \leq y) = P(X_{(i)} \leq F^{-1}(y))$  som är sannolikheten att minst  $i$  stycken av observationerna är mindre eller lika med  $F^{-1}(y)$ . Men sannolikheten att en given observation är mindre än  $F^{-1}(y)$ ,  $X \leq F^{-1}(y)$  är  $F(F^{-1}(y)) = y$  och  $Z =$ antalet observationer som är mindre än eller lika med  $F^{-1}(y)$  är då  $\text{Bin}(n, y)$ . Därför får vi att

$$G(y) = P(F(X_{(i)}) \leq y) = P(Z \geq i) = \sum_{k=i}^n \binom{n}{k} y^k (1-y)^{n-k}$$

där  $G(y)$  alltså är fördelningsfunktionen till  $F(X_{(i)})$ . Täthetsfunktionen ges av  $g(y) = G'(y)$  som man kan visa (lämnas som övning) blir

$$g(y) = \frac{n!}{(i-1)!(n-i)!} y^{i-1} (1-y)^{n-i}$$

vilken är täthetsfunktionen till en så kallad beta-fördelning med parametrar  $i$  och  $n-i+1$ ,  $beta(i, n-i+1)$ .

Som uppskattning  $F(X_{(i)})$  skulle man då kunna ta väntevärdet av denna beta-fördelning

$$\begin{aligned} E(F(X_{(i)})) &= \int_0^1 yg(y)dy = \int_0^1 y \frac{n!}{(i-1)!(n-i)!} y^{i-1} (1-y)^{n-i} dy \\ &= \frac{i}{n+1} \int_0^1 \frac{(n+1)!}{(i)!(n-i)!} y^i (1-y)^{n-i} dy = \frac{i}{n+1} \end{aligned}$$

eftersom den sista integranden är täthetsfunktion till  $beta(i+1, n-i+1)$ . Dessa väntevärden kallas medelranger.

Ett annat sätt är att skatta  $F(X_{(i)})$  med medianen i fördelningen d.v.s. med det värde  $u_i^{(n)}$  som är lösningen till  $G(u_i^{(n)}) = 0.5$  d.v.s.

$$\int_0^{u_i^{(n)}} \frac{n!}{(i-1)!(n-i)!} y^{i-1} (1-y)^{n-i} dy = 0.5$$

Det är medianrangerna. Det finns undersökningar som tyder på att medianrangerna ligger närmare det korrekta värdet och har mindre spridning än medelrangerna. Därför föredrar vi medianranger.

Man kan visa att medianrangerna  $u_i^{(n)}$  kan approximeras med värdena  $\frac{i-0.3}{n+0.4}$ , den s.k. Bernhards formel. Det absoluta felet är mindre än 0.00125, varför denna approximation räcker för praktiska bruk.

## Exempel

**Exempel. 1** I ett försök sattes 10 komponenter med weibullfördelade livslängder under provning tills alla komponenterna brustit. De observationerna livslängderna blev (100-tal timmar)

0.28 0.47 0.68 0.84 1.26 1.61 1.7 1.74 2.58 5.03.

Detta är ett icke-censurerat försök och för att beräkna ML-skattningen av formparametern  $c$  sätter vi  $r = n = 10$  i (4). Med hjälp av MATLAB och m-filen *fzero* beräknades ML-skattningen till  $\hat{c} = 1.33$ . Sättes detta värde in

i (3) erhålles skattningen  $\hat{b} = 0.4650$ . Därefter beräknar vi skattningen av formparametern  $a$ ,  $\hat{a} = 1/\hat{b}^{1/\hat{c}} = 1.77$

I weibullplotten plottar vi data mot medianrangerna  $u_1^{(10)}, u_2^{(10)}, \dots, u_{10}^{(10)}$ . Till data anpassar vi en rät linje. För att skatta  $a$  ser vi efter var denna linje skär hjälplinjen  $y = 1 - e^{-1} = 0.63$ . Skärningen tycks ske i  $x = 1.8$ . Detta värde är skattningen  $a^*$ .

Parallellflytta linjen så att den går genom origo i  $(u, v)$ -planet, dvs genom den punkt som är inringad i weibullpappret. Denna linje skär den lodräta hjälplinjen ( $u = 0$  i  $(u, v)$ -planet) i  $v = 1.3$ . Vi får således skattningen av formparametern  $c^* = 1.3$ . Dessa skattningar är i god överensstämmelse med ML-skattningarna. En viss osäkerhet råder dock i att anpassa en rät linje till data.

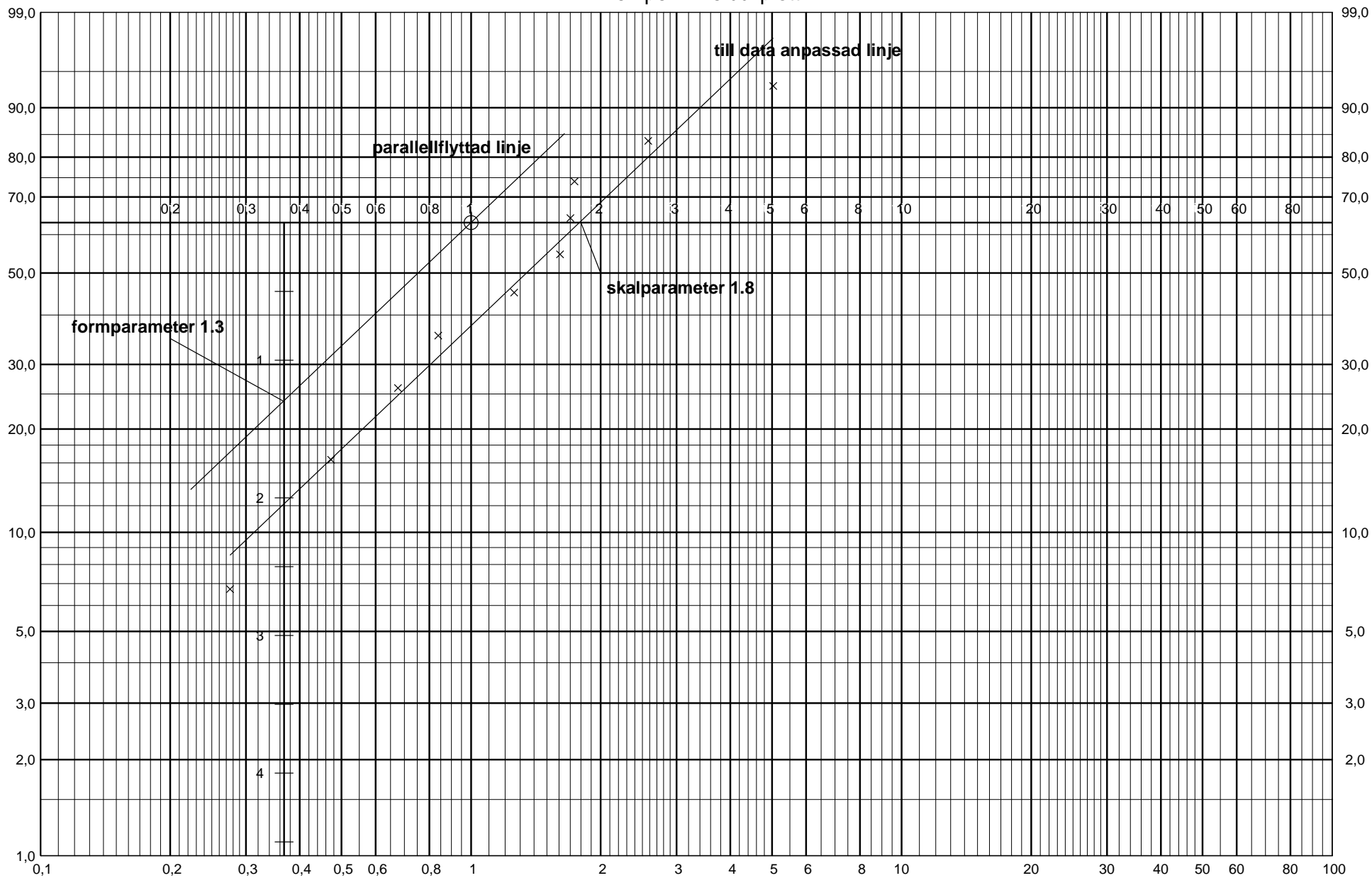
**Exempel. 2** I ett annat försök provades 20 komponenter, och man avsåg att hålla på tills 8 brustit. Livslängdstiderna för dessa 8 komponenter blev

0.96, 1.18, 1.19, 1.32, 1.34, 1.70, 1.80, 1.85

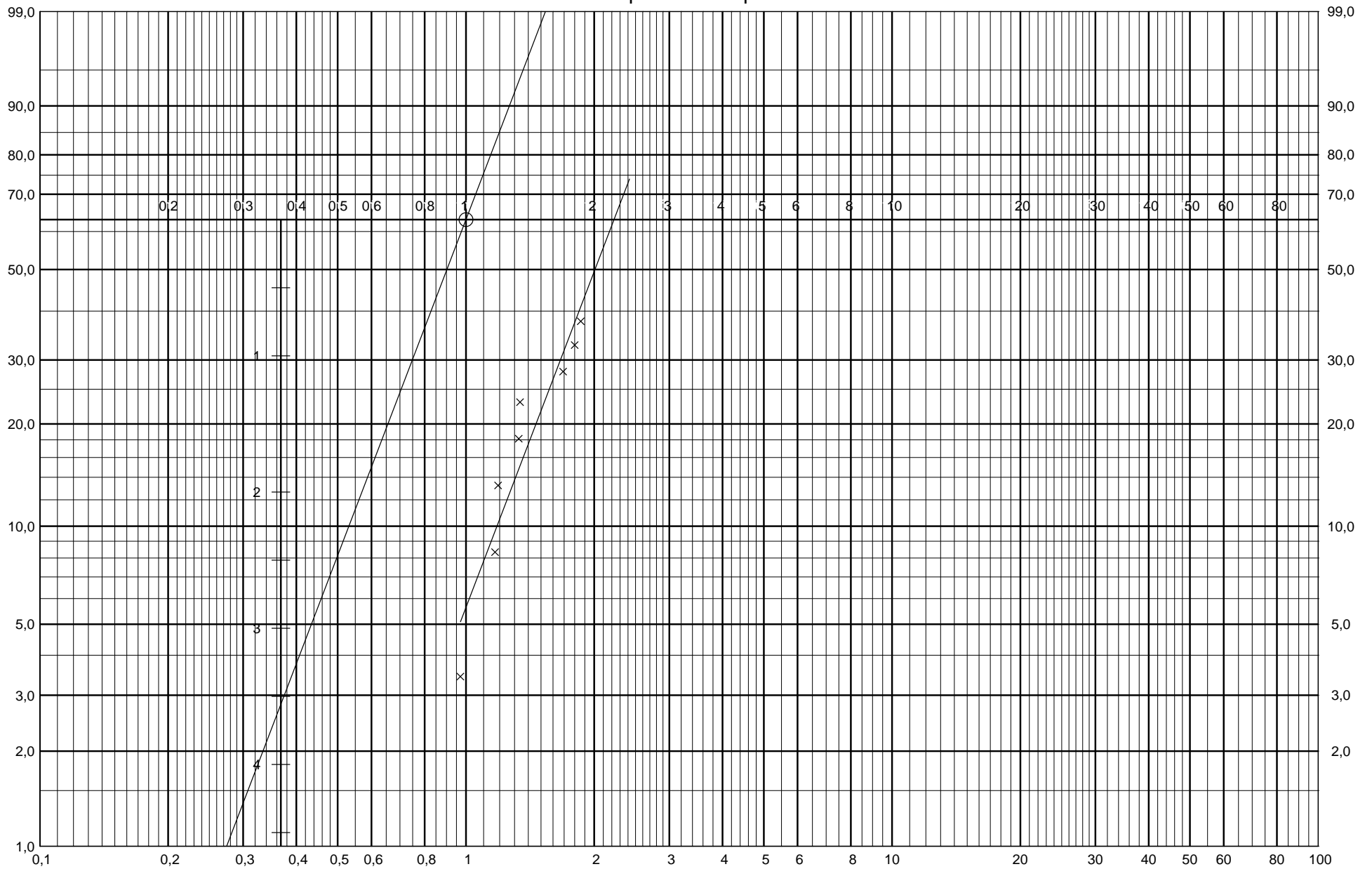
Vi har en censurerad provning med  $n=20$  och  $r=8$ . Vi får lösa (4) numeriskt och erhåller  $\hat{c} = 3.8137$ . På samma sätt som i föregående exempel erhåller vi skattningen av  $b$  från (3),  $\hat{b} = 0.04817$ . Det ger  $\hat{a} = 2.2151$ .

Weibullplotten får vi på samma sätt som i exempel 1. Vi skall dock i detta exempel plotta de observerade livslängderna ovan mot medianrangerna  $u_1^{(20)}, u_2^{(20)}, \dots, u_8^{(20)}$ . Från weibullplotten ser vi att skattningarna ges av  $a^* \approx 2.2$  och  $c^* \approx 3.6$ .

# Exempel 1 weibullplott

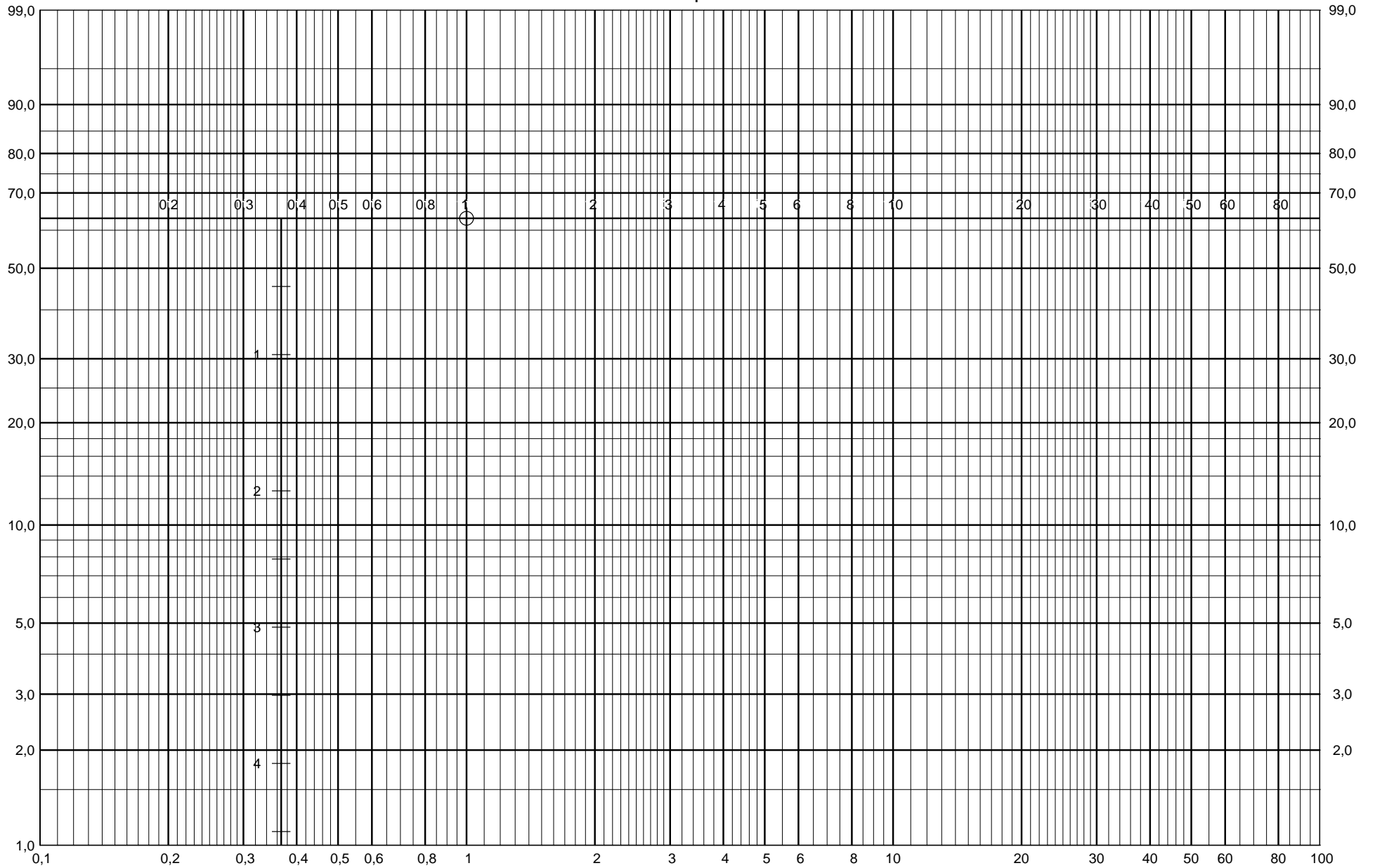


Exempel 2 weibullplott





# Weibullplott





Stickprovsstorlek  $n$

$i$	21	22	23	24	25	30	25	40	45	50
1	3.25	3.10	2.97	2.85	2.73	2.28	1.96	1.72	1.53	1.38
2	7.86	7.51	7.19	6.90	6.62	5.53	4.75	4.16	3.70	3.33
3	12.53	11.97	11.46	10.99	10.55	8.81	7.57	6.63	5.90	5.31
4	17.21	16.44	15.73	15.09	14.49	12.10	10.39	9.10	8.10	7.29
5	21.89	20.91	20.01	19.19	18.43	15.40	13.22	11.58	10.30	9.28
6	26.57	25.38	24.30	23.30	22.38	18.69	16.05	14.06	12.51	11.26
7	31.26	29.86	28.58	27.41	26.32	21.99	18.87	16.54	14.71	13.25
8	35.94	34.33	32.86	31.51	30.27	25.28	21.70	19.01	16.92	15.24
9	40.63	38.81	37.15	35.62	34.22	28.58	24.53	21.49	19.12	17.22
10	45.31	43.29	41.43	39.73	38.16	31.87	27.36	23.97	21.33	19.21
11	50.00	47.76	45.72	43.84	42.11	35.17	30.19	26.45	23.53	21.19
12	54.69	52.24	50.00	47.95	46.05	38.46	33.02	28.93	25.74	23.18
13	59.37	56.71	54.28	52.05	50.00	41.76	35.85	31.41	27.94	25.17
14	64.06	61.19	58.57	56.16	53.95	45.06	38.68	33.89	30.15	27.15
15	68.74	65.67	62.85	60.27	57.89	48.35	41.51	36.37	32.35	29.14
16	73.43	70.14	67.14	64.38	61.84	51.65	44.34	38.84	34.56	31.13
17	78.11	74.62	71.42	68.49	65.78	54.94	47.17	41.32	36.77	33.11
18	82.79	79.09	75.70	72.59	69.73	58.24	50.00	43.80	38.97	35.10
19	87.47	83.56	79.99	76.70	73.68	61.54	52.83	46.28	41.18	37.09
20	92.14	88.03	84.27	80.81	77.62	64.83	55.66	48.76	43.38	39.07
21	96.75	92.49	88.54	84.91	81.57	68.13	58.49	51.24	45.59	41.06
22		96.90	92.81	89.01	85.51	71.42	61.32	53.72	47.79	43.05
23			97.03	93.10	89.45	74.72	64.15	56.20	50.00	45.03
24				97.15	93.38	78.01	66.98	58.68	52.21	47.02
25					97.27	81.31	69.81	61.16	54.41	49.01
26						84.60	72.64	63.63	56.62	50.99
27						87.90	75.47	66.11	58.82	52.98
28						91.19	78.30	68.59	61.03	54.97
29						94.47	81.13	71.07	63.23	56.95
30						97.72	83.95	73.55	65.44	58.94
31							86.78	76.03	67.65	60.93
32							89.61	78.51	69.85	62.91
33							92.43	80.99	72.06	64.90
34							95.25	83.46	74.26	66.89
35							98.04	85.94	76.47	68.87
36								88.42	78.67	70.86
37								90.90	80.88	72.85
38								93.37	83.08	74.83
39								95.84	85.29	76.82
40								98.28	87.49	78.81
41									89.70	80.79
42									91.90	82.78
43									94.10	84.76
44									96.30	86.75
45									98.47	88.74
46										90.72
47										92.71
48										94.69
49										96.67
50										98.62

För alla praktiska behov är approximationen  $u_i^{(n)} = \frac{i - 0.3}{n + 0.4}$  tillräcklig. Det absoluta felet är mindre än 0.00125, dvs 0.125 %.