CHAPTER 03 - Log likelihood

3.1 Aims

The aim of this chapter is to get reasonable estimates of the parameters. In the case of my computer programs and the Weibull distribution, I hope to get the estimates as close to my starting values of kappa and rho and to ensure convergence in reasonable computing time.

3.2 Log likelihood (Weibull distribution)

According to Cox and Oakes (1984), the log likelihood for the Weibull distribution is given by

l=d ln k + kd ln p + (k-1)(sum of uncensored ln x_i)-(p^k*sum of x_i^k) (3.1)

Later on, in the absence of censoring (put d=n), we see if the Weibull is an exponential family in the two definitions.

Cox and Oakes (1984), differentiated (3.1) with respect to each parameter to yield the following equations:

U_p = l_p = (kd/p) - (kp^{k-1})(sum of x_i^k) (3.2)

U_k = l_k = (d/k) - (d ln p) + (sum of uncensored ln x_i)
-p^k(sum of x_i^k(ln px_i) (3.3)

According to Cox and Oakes (1984), if k is specified, the MLE ^pk of p can be found explicity by solving U_p=0
as ^p=(d/sum of x_i^k)$^{1/k}$ (3.4)
a result which could be derived from the fact that T^k has an exponential distribution with parameter p^k. Substituting into the equation U_k=0 in (3.3) yields

(d/k) + (sum of uncensored ln x_i) -(d(sum x_i^k ln x_i)/(sum x_i^k))

I will use the above equation to obtain the iterative scheme for kappa, as in section 3.2.1.

3.2.1 Derivation of the iterative scheme for kappa

Rearrange the equation :-

(d/k)=(d*sum x_i^k ln x_i/sum x_i^k)-(sum of uncensored ln x_i)

On the LHS, replace k by k_{n+1}
On the RHS, replace k by k_n

(d/kn+1)=(d*sum x_i^kn ln x_i/sum x_i^kn)-(sum of uncensored ln x_i)

Dividing both sides by d :-

(1/k_{n+1})=(sum x_i^kn ln x_i/sum x_i^kn)-(sum of uncensored ln x_i/d)

-- k_{n+1} = 1/((sum x_i^kn ln x_i/sum x_i^kn)-(sum of uncensored ln x_i/d)) --

---

where k_n and k_{n+1} are the old and new values of k respectively
(sum x_i^kn ln x_i) and (sum x_i^kn) are for all observations (uncensored and censored)
(sum of uncensored ln x_i) are for uncensored observations
d is the no. of failures (uncensored)
x_i is the minimum of T_i and c_i

I use this above equation to obtain a new value of k, k_{n+1}.
I substitute this into (3.4) to obtain p.
Repeat until ABS($k_{n+1}-k_n$) < Tolerance
In my programs (listed in Appendix 02) I use a loop for this purpose.

When I have obtained the estimates of k and p, I used Equations (2.6) to (2.10) in order to calculate the median, lower quartile, upper quartile, midquartile range and interquartile range. These values for each model described in section 1.6 can be compared with that of original k and p, as in Table 2.1.

3.2.2 How the elements of the information matrix are obtained

Cox and Oakes (1984), differentiated (3.2) and (3.3) with respect to each parameter to yield the following equations:

$-I_{pp} = l_{pp} = (-kd/(p^2))-k(k-1)p^{k-2}\text{sum of } x_i^k$ (3.5)

$-I_{pk} = l_{pk} = (d/p)-(p^{k-1}(1+k \ln p)\text{sum of } x_i^k-kp^{k-1}\text{sum of } x_i^k\ln x_i$ (3.6)

$-I_{kk} = l_{kk} = (-d/(k^2))-p^k\text{sum of } x_i^k[\ln(px_i)]^2$ (3.7)

(Where $l_{kk} = d^2l/dk^2$, etc)

Use the new value of k and therefore p to calculate the elements of the information matrix (Equations (3.5), (3.6) and (3.7))

3.3 Asymptotic Theorem for MLE's (From Stats Models, JG, 1990)

Subject to certain technical conditions, if n (in the case of the uncensored model) or d (in the censored models) is large, MLE ^X is approx

(i) Normally distributed
(ii) unbiased, ie E(^X) tends to X
(iii) Efficient, ie V(^X) tends to 1/(I(X))

$SE(\hat{X})=(I(X))^{-0.5}$, ie ^X is approx $N(X,I^{-1})$
So approx. 95% confidence interval for X is

$\hat{X}\pm1.96(I(X))^{-0.5}$

3.3.1 Multiparameter case (From Stats Models, JG, 1990)

If $\underline{X}$ has more than 1 component, $\hat{\underline{X}}$ is approximately multivariate normal with approx. $\hat{\underline{X}} = N(\underline{X},I^{-1})$ where I is the element of the information matrix. I can assume that the Asymptotic Theorem for MLEs and the multiparameter case hold, so that I can calculate standard errors and 95% confidence intervals for the parameters.

3.3.1.1 EXAMPLE

For the Weibull distribution.
$\hat{k}=N(k,I_{kk}^{-1})$ and $\hat{p}=N(p,I_{pp}^{-1})$
The standard errors of ^k and ^p are $(I^{kk})^{-0.5}$ and $(I_{pp})^{-0.5}$ respectively.
The 95% CI of k is $\hat{k}\pm1.96(I_{kk})^{-0.5}$
The 95% CI of p is $\hat{p}\pm1.96(I_{pp})^{-0.5}$

3.4 Examples

I obtained the following results from my programs in examples 1.7.1 to 1.7.4. The examples illustrate the Asymptotic Theorem for MLEs.

CHAPTER 03 - Log likelihood

3.4.1 Uncensored

for 2.3, 1.5

d=75, k=2.56081, p=1.40874
I_{kk} = 1443.31 and I_{pp} = 247.83
SE(^k) = 0.02632 and SE(^p) = 0.06352
With 95% confidence k is between 2.50922 and 2.61240
With 95% confidence p is between 1.28424 and 1.53325

3.4.2 Constant censoring

d=73, k=2.53085, p=1.40673
I_{kk} = 1398.28 and I_{pp} = 236.28
SE(^k) = 0.02674 and SE(^p) = 0.06506
With 95% confidence k is between 2.47843 and 2.58326
With 95% confidence p is between 1.27922 and 1.53424

3.4.3 Realistic case I

d=62, k=2.57465, p=1.41836
I_{kk} = 1145.52 and I_{pp} = 204.30
SE(^k) = 0.02955 and SE(^p) = 0.06996
With 95% confidence k is between 2.51674 and 2.63256
With 95% confidence p is between 1.28123 and 1.55549

3.4.4 Realistic case II

d=39, k=2.82733, p=1.39798
I_{kk} = 2295.47 and I_{pp} = 159.52
SE(^k) = 0.02087 and SE(^p) = 0.07918
With 95% confidence k is between 2.78642 and 2.86824
With 95% confidence p is between 1.24280 and 1.55317

3.5 Exponential Family (From Stats Models, JG, 1990)

We now consider the exponential family of the distributions.
A family of distributions [f(y|theta)] is in the exponential family if we have the form f(y|theta)=exp[((ya(theta)-b(theta))/thi)+c(y,thi)]
where thi is a 'scale parameter' and is 'known' - we concentrate on the way the distribution changes with theta.

EG: Is the Weibull distribution an exponential family? (Assume k is known)

$\ln f(t|p)=\ln k+\ln p+(k-1)\ln p+(k-1)\ln t-p^k t^k$
$\ln f(t|p)=(k-1)\ln t-t^k p^k+(k-1)\ln p+\ln k+\ln p$
We do not have the required form, so no.

NB: Definitions of exponential family vary slightly. Sometimes there is:-
ln f(y|theta)=(g(y)a(theta)-b(theta)/thi)+c(y,thi)
Using the expression for ln f(t|p) and this definition, we find that the Weibull distribution is an exponential family.
$a(p)=-p^k$, $b(p)=-k \ln p$.

Another important distribution that is in the exponential family is the Binomial distribution. There are some probability distributions that are not in the exponential distribution according to the classical definition, but it is according to the alternative definition.

3.5.1 Property of the Exponential Family (Stats Mod., KR, 1990)

We can get mean (and variance) easily. Assume that there is no censoring and we have a canonical parameter. Start with the definition of the exponential family. It is 'well known' that E(dl/d(theta))=0 and $E(d^2l/d(theta)^2) + E[(dl/d(theta))^2] = 0$
We use these well known facts and differentiate the log likelihood with respect to theta.
We have mu=b'(theta) and variance=b''(theta)

EG: Apply to Weibull distribution in 2nd form with b(p)=-k ln p.
Let z=-p, b(z)=-k ln -z. Differentiate this with respect to z.
b'(z)=-k(-1)/(-z)=-k/z. But z=-p, so <u>mu=(-k/-p)=k/p</u>.
NB: If we put k=1, we have an exponential distribution with mean=1/p.

Variance= b''(theta), b'(z)=-k/z, $b''(z)=k/z^2$.
But z=-p, $b''(p)=k/(-p)^2=$<u>kp^{-2}</u>

We can use these results to obtain the coefficient of variation (CV).
Coeff. of variation = sd/mean (SD=square root of variance)
$CV = (k^{0.5}/p)/(k/p) = k^{-0.5}$

We find that the CV>1 when k<1 and CV<1 when k>1.

I can use my estimates of k and p to calculate the mean, variance and coefficient of variation.

I will be using the mean, variance and CV in the accelerated life and proportional hazards models (Chapter 5) for comparison. The interquartile range gives some idea of the variance (or SD), since the interquartile range and variance (or SD) measure the spread.

We now find the coefficient of skewness for the Weibull distribution.
The coefficient of skewness (Pearson's) is given by:-

skewness = 3(mean-median)/sd
=(3/cv)-(3*median/sd).

For the Weibull distribution, the skewness is given by:-

$(3\ sqrt(k))-(3(0.693)^{1/k}/sqrt(k))$

k	p	Mean	Var.	CV	Skew
2.50000	1.50000	1.66667	1.11111	0.63246	3.10478
2.56071	1.40874	1.81773	1.29032	0.62491	3.17594
2.53085	1.40673	1.79910	1.27892	0.62859	3.14107
2.57481	1.41836	1.81534	1.27989	0.62320	3.19232
2.82733	1.39798	2.02244	1.44669	0.59472	3.47716

Table 3.1 - Mean, Variance, Coeff. of variation and Skewness

From Table (3.1) we find that the Weibull distribution is right-skewed. However, the coefficient of skewness varies with the index.

3.6 A test for exponentiality

I now use the likelihood ratio test in order to test for exponentiality. In addition, I assume that the large sample method can be used ie 2 ln (lambda) = 2(l1-l0) follows a chi-square distribution with p1-p0 degrees of freedom, where p0 and p1 are the no. of parameters under H0 and H1 respectively. I substitute (3.4) into (3.1) to eliminate sum(x_i^k) in order to get l=d ln k+ kd ln p+((k-1)*sum of uncensored ln x_i)-d ...(3.8)

Under H0, I put k=1 regardless of p. From (3.8), l_0=d ln p-d

Under H1, I put k<>1 regardless of p.

From (3.8), l_1=d ln k+kd ln p+((k-1)*sum of uncensored ln x_i)-d

l_1-l_0=d(k ln p - ln p)+((k-1)*sum of uncensored ln x_i)

Then I compare 2(l1-l0) with the chi-square distribution on 1 degree of freedom (no parameters under H0 and 1 parameter under H1) and using a significance level of 5%.

I write computer programs to use this test for each of the 4 models described in CHAPTER 1. I calculate the 95% confidence interval of k. If the data is not exponential, this 95% confidence interval should not include 1. But, if the data follows an exponential distribution, this 95% confidence interval MUST include 1.
The program listings and example output are given in APPENDIX 02.

3.6.1 Examples

I obtained the following results from my programs:

	Uncensored	Constant censoring	RC(1)	RC(2)
l_0	-49.30	-48.09	-40.33	-25.93
l_1	- 4.05	- 7.16	-10.62	-12.79
$2(l_1-l_0)$	90.50	81.85	59.42	26.28

Compared with the critical value of chi-square which is 3.841, I find that for each model, it is NOT reasonable to assume exponentiality.

3.7 Computing

Algorithm:-

(01) Simulate, as described in section 1.5. Where appropriate, obtain the censor and times failure times using x_i=min(T_i,c_i). If censoring is present d=d-1 and put indicator variable=1.
(02) Calculate the sum of ln x_i for the uncensored individuals.
(03) Put the present value of k equal to the past value of k.
(04) Calculate the sums of x_i^k and x_i^kln x_i with PRESENT values of k.
(05) Use the iterative scheme for k (section 3.2.1) to calculate the new value of k.
(06) Use this new value of k to obtain p (Equation 3.4)
(07) Use this calculated value of p to obtain sum of $x_i^k[\ln (px_i)]^2$
(08) Is the absolute magnitude of (new value of k)-(present value of k) less than the desired tolerance? If not, then:
(a) Reset the sums in the iteration loop
(b) Goto (3)
(09) Calculate the elements of the information matrix

For the exponentiality test (section 3.6) also include:
(10) Calculate the std. errors and 95% confidence intervals of p and k.
(11) Obtain likelihood equations under H_0 and H_1.
(12) Calculate $2(l_1-l_0)$ and compare it with the critical value of chi-square at the 5% level. If this is greater, it is not reasonable to assume exponentiality, otherwise assume exponentiality.

3.8 Summary

I hope to get the estimates as close to the starting values of kappa and rho and to obtain convergence in reasonable computing time.

The log likelihood function for the Weibull distribution is
$l=d \ln k + kd \ln p + (k-1)(\text{sum of uncensored } \ln x_i)-(p^k*\text{sum of } x_i^k)$

The iterative scheme for kappa is

-- $k_{n+1} = 1/((\text{sum } x_i^kn \ln x_i/\text{sum } x_i^kn)-(\text{sum of uncensored } \ln x_i/d))$ --

where k_n and k_{n+1} are the old and new values of k respectively
(sum $x_i^kn \ln x_i$) and (sum x_i^kn) are for all observations (uncensored and censored)
(sum of uncensored $\ln x_i$) are for uncensored observations
d is the no. of failures (uncensored)
x_i is the minimum of T_i and c_i

The elements of the information matrix are:
$-I_{pp} = l_{pp} = (-kd/(p^2))-k(k-1)p^{k-2}\text{sum of } x_i^k$
$-I_{pk} = l_{pk} = (d/p)-(p^{k-1}(1+k \ln p)\text{sum of } x_i^k-kp^{k-1}\text{sum of } x_i^k\ln x_i$
$-I_{kk} = l_{kk} = (-d/(k^2))-p^k\text{sum of } x_i^k[\ln(px_i)]^2$

The Asymptotic Theorem for MLEs:
Subject to certain technical conditions, if n (in the case of the uncensored model) or d (in the censored models) is large, MLE ^X is approx
(i) Normally distributed
(ii) unbiased, ie E(^X) tends to X
(iii) Efficient, ie V(^X) tends to 1/(I(X))

$SE(\hat{X})=(I(X))^{-0.5}$, ie ^X is approx $N(X,I^{-1})$
So approx. 95% confidence interval for X is
$\hat{X}\pm1.96(I(X))^{-0.5}$

The multiparameter case:
If $\underline{X}$ has more than 1 component, $\hat{\underline{X}}$ is approximately multivariate normal with approx. $\hat{\underline{X}} = N(\underline{X},I^{-1})$ where I is the element of the information matrix. I can assume that the Asymptotic Theorem for MLEs and the multiparameter case hold, so that I can calculate standard errors and 95% confidence intervals for the parameters.

The Weibull distribution is not an exponential family according to the classical definition, but it is according to the alternative definition. The property of an exponential family is that we can find the mean and variance easily. Assuming that we have a canonical parameter, mean=b'(theta) and variance=b''(theta). EG for Weibull: mean=k/p and variance=k/p^2. The coefficient of variation is given by $(b''(\text{theta}))^{0.5}/(b'(\text{theta}))$. In the Weibull distribution we find that

Rep^n ?

CV<1 when k>1 and CV>1 when k<1. If we compare the mean with the median, we find that the Weibull distribution is right-skewed (mean>median). However, the coefficient of skewness varies with the index.

We can use the likelihood ratio test in order to test for exponentiality. Under H0 put k=1 for some p and under H1 put k<>1 for some p. We calculate $2(l_1-l_0)$ (change in scaled deviance) which has a chi-square distribution with p1-p0 degrees of freedom where p0 and p1 are the no. of parameters under H0 and H1 respectively. We compare $2(l_1-l_0)$ with the critical value of chi-square on p1-p0 degrees of freedom. If this change in scaled deviance is greater than this critical distribution, then it is not reasonable to assume exponentiality.

3.9 Later Work

In later work, p will no longer be constant and will be replaced by p_i, where $p_i=\exp(b_0+b_1+...)$
Equation (3.1) will be used and if I substitute p_i into this, I will have the equation and its 1st derivatives in terms of b0,b1, etc.
I will have iterative schemes in b0,b1,... as well as k. My programs will have at least 3 iterative schemes and for each parameter, the absolute difference between the old and new value MUST be less than the desired tolerance. The aims described in 3.1 apply, but instead of kappa and rho, we consider kappa,b_0,b_1,etc. This is hard work and more implicit, but it is more stable than the work I have done in this chapter.