CHAPTER 02 - Distributions of failure time

2.1 Introduction

In this and later chapters I am going to deal with continuous distributions, particularly Weibull. Why?

(1) Time is assumed to be a continuous random variable and I will be modelling time in my programs.

(2) Easy to integrate, hence easy to simulate. However, it is not always true that continuous distributions are easy to integrate (eg Normal, Gamma). Numerical methods should be used for those distributions that are difficult to integrate (eg Simpson's Rule), but this is beyond the scope of this project!

However, I may use a discrete distribution that relates the no. of successes/failures out of a given no. of trials, but only Binomial. Why? The Binomial distribution is an exponential family (see section 3.5) and has a logit link. This may be valuable for logistic regression and I will be describing its use in Medical Statistics (Appendix 05).

2.2 Survivor functions

"Consider a homogeneous population of individuals each having a 'failure time'. That is, we deal with a single positive random variable T. In particular, an origin and scale for measuring time are assumed to be clearly defined. We examine the general specification of the distribution of T and then consider various special distributions that are useful. We write this as

$$\mathcal{F}_T(t)=pr(T>=t) \quad ...(2.1)$$

for the survivor function of T, omitting the suffix T when the random variable is clear from the context." (CO, 1984)

It is equal to 1-F(t), where F(t) is the CDF of any probability distribution. Since the CDF is between 0 and 1, the survivor function is also between 0 and 1. For continuous distributions (eg Weibull), we differentiate this result with respect to t to get the PDF

$$f_T(t)=-\mathcal{F}'_T(t) \quad ...(2.2)$$

such that $\mathcal{F}_T(t)$=Integral of $f_T(t)$ between t and infinity

EG For the Weibull distribution

$$F_T(t)=1-exp(-(pt)^k)$$

$$\mathcal{F}_T(t)=1-(1-(exp-(pt)^k)$$

$$=exp(-(pt)^k)$$

Differentiating with respect to t gives

$$f_T(t)=kp(pt)^{k-1}exp(-(pt)^k)$$

2.3 Hazard functions and integrated hazard

"$\mathcal{F}T(.)$ and $f_T(.)$ provide two mathematically equivalent ways of specifying the distribution of a continuous positive random variable and there are of course many equivalent functions. One with special value in the present context is the hazard function, or age-specific failure rate, defined by

$$h_T(t)=pr(t<=T<t+delta\ t<=T)/delta \quad ...(2.3)$$

By the definition of conditional probability, we have

$$h(t)=f(t)/\mathcal{F}(t) \quad ...(2.4)$$ (Omitting the suffix T)" (CO, 1984)

EG For the Weibull distribution

$f(t)=kp(pt)^{k-1}exp(-(pt)^k)$

$\mathcal{F}(t)=exp(-(pt)^k)$

$h(t)=kp(pt)^{k-1}exp(-(pt)^k)/(exp(-(pt)^k))$

$\quad =kp(pt)^{k-1}$

If the distributions are continuous, then in accordance with Cox and Oakes (1984) we can use (2.2) to substitute into (2.4). This gives the following result for h(t):

$h(t)=-\mathcal{F}'(t)/\mathcal{F}(t)$

$\quad =(-d/dt)\ ln\ \mathcal{F}(t)$

"So that, because $\mathcal{F}(0)=1$,
$\mathcal{F}(t)$=exp[-Integral of h between 0 and t]=exp[-H(t)] (2.5)
say, where H(.) is called the integrated hazard." (CO, 1984)

EG For Weibull $H(t)=(pt)^k$

The PDF f(t) can be related to the survivor function, hazard function and/or integrated hazard :-

$f(t)=-\mathcal{F}'(t)$ OR

$f(t)=h(t)\mathcal{F}(t)$ OR

$f(t)=h(t)exp[-H(t)]$

2.4 Computing

I write Fortran programs to demonstrate the behaviour of the survivor, hazard and density functions as well as the integrated hazard, with increasing time (this includes censoring, where appropriate).
However, I used INTER-CHART to plot the graphs. These programs can be extended as a basis for the accelerated life and proportional hazards model (see Chapter 5).

2.5 Properties of probability distributions

I look at the proporties of probability distributions and their behaviour of survivor functions, hazard functions, PDFs and integrated hazards with respect to time. I will also find the median, etc of these distributions.

2.5.1 The Weibull distribution

The Weibull distribution has PDF $f(t)=kp(pt)^{k-1}exp(-(pt)^k)$, survivor function $\mathcal{F}(t)=exp(-(pt)^k)$, hazard function $h(t)=kp(pt)^{k-1}$ and integrated hazard $H(t)=(pt)^k$.

Their rates of change with respect to time are:

(df/dt) $=\mathcal{F}h' + \mathcal{F}'h$ (Product rule)

$\quad =\mathcal{F}h' - fh$ (using 2.2)

$\quad =(exp(-(pt)^k)k(k-1)p^k t^{k-1}$

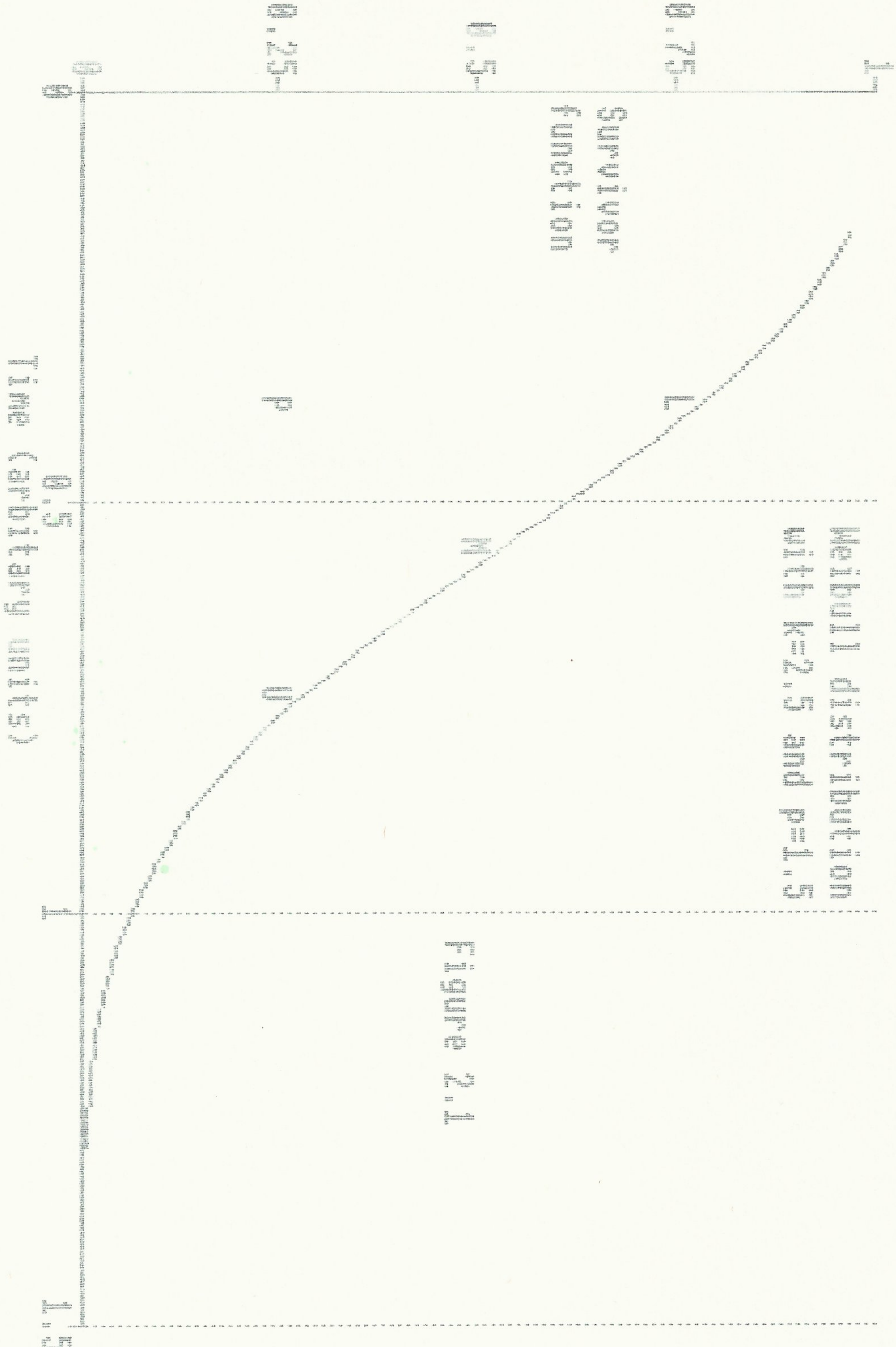
$\quad\quad -kp(pt)^{k-1}exp(-(pt)^k)kp(pt)^{k-1}$

$\quad =(k(k-1)p^k t^{k-1}-[(kp(pt)^{k-1})^2])exp(-(pt)^k)$

$(d\mathcal{F}/dt)\ =-kp(pt)^{k-1}exp(-(pt)^k)$

$(dh/dt)\ =k(k-1)p^k t^{k-1}$

$(dH/dt)\ =kp^k t^{k-1}$



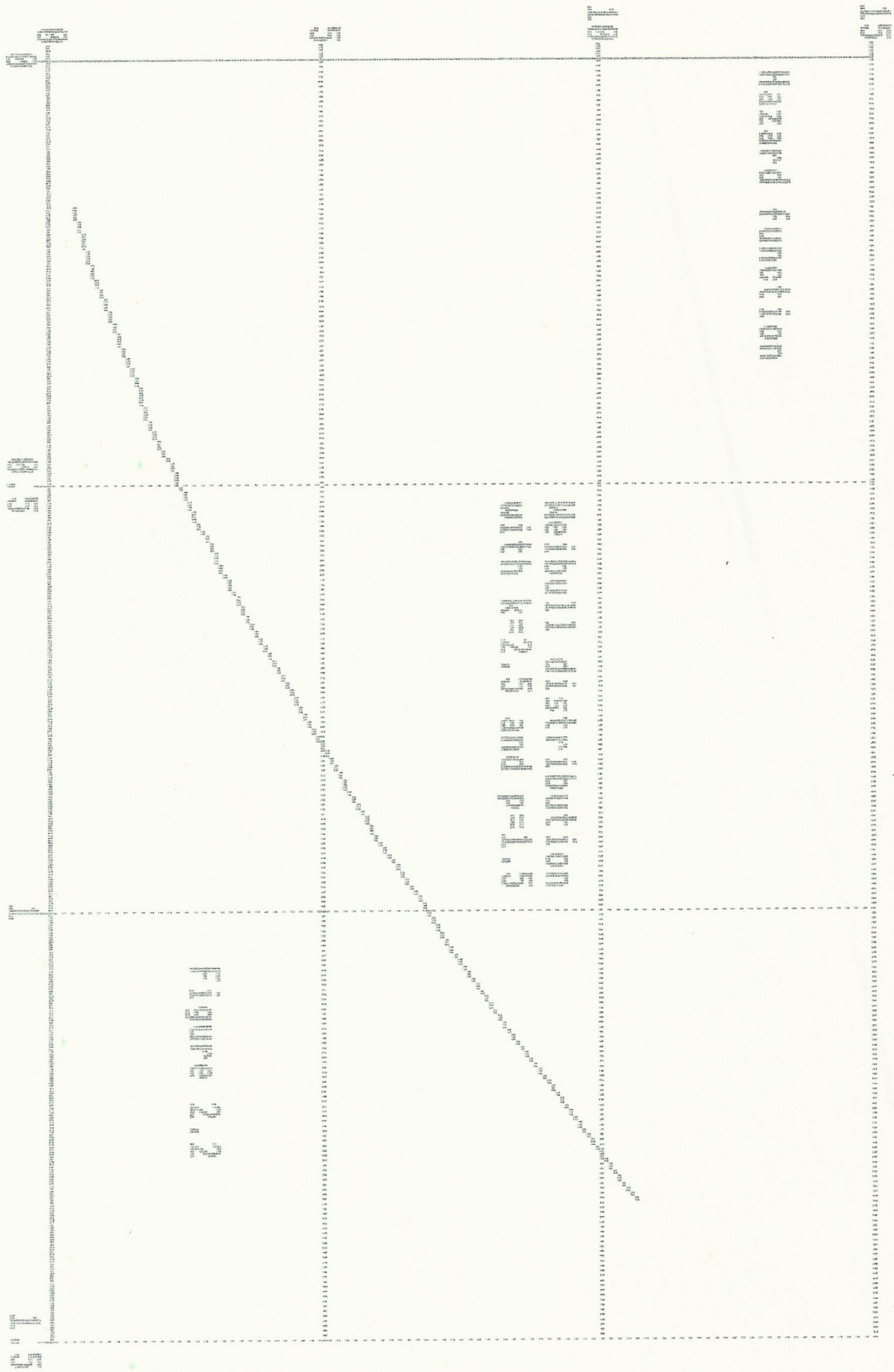

Hazard function
Weibull distribution
with k=2.5 and p=1.5
15
10
5
0
0
0.5
1
1.5
Time (arbitary units)

Figure 2.2

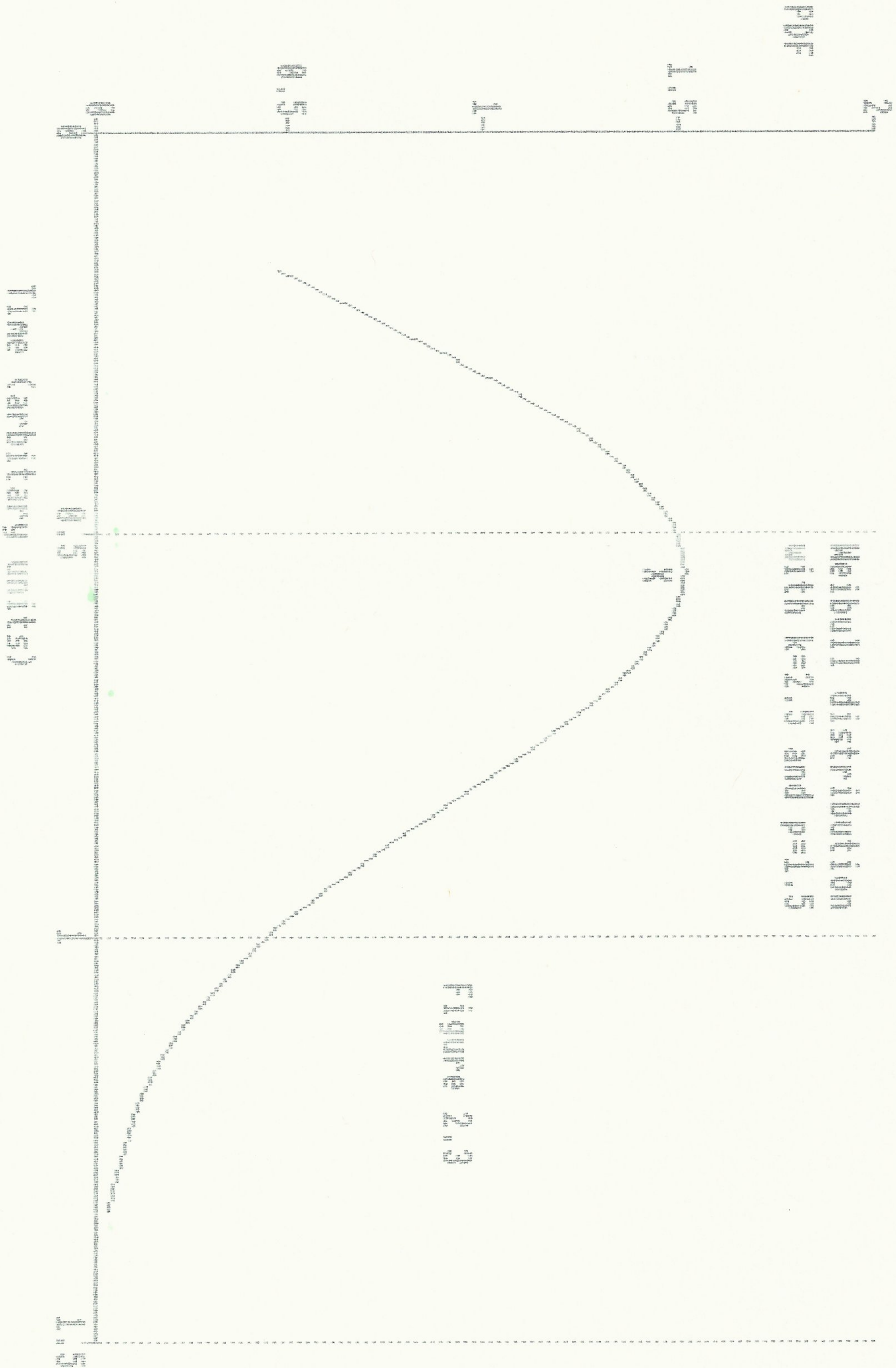

Weibull distribution
with k=2.5 and p=1.5
PDF
2
1.5
1
0.5
0
X
0
0.5
1
1.5
Time (arbitary units)

Figure 2.3

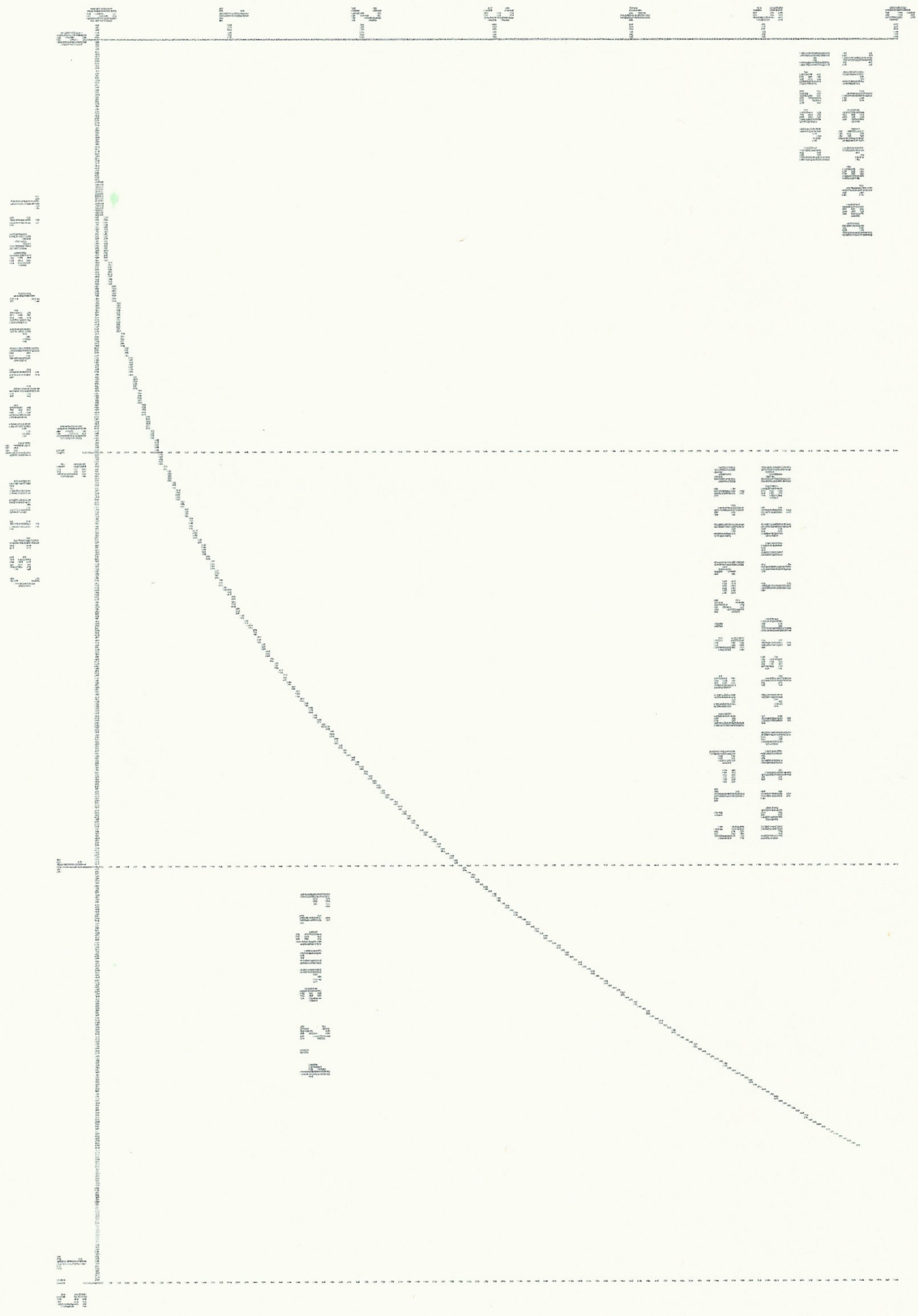

Integrated
Hazard
Weibull distribution
with k=2.5 and p=1.5
Time (arbitary units)
0
1
2
3
4
5
6
0
0.5
1
1.5

Figure 2.4

2.5.1.1 Survivor function

If k=1, we have the exponential distribution. Plotting a graph of ln *F* against t^k, gives a straight line with gradient $-(p)^k$ (Assume that k is known). When t=0, survivor function=1 and when t=infinity, survivor function=0. It is possible to plot a graph of the survivor function against time. Such a graph is shown in Figure 2.1 for the uncensored case, with k=2.5 and p=1.5. The gradient of this graph is -f(t). The reason for this minus sign is that the survivor function DECREASES as time increases. When f(t) is a maximum the rate of change of the survivor function is at its greatest and when f(t) is a minimum the rate of change of the survivor function is at its minimum.

The survivor function has lots of special features that can easily be deduced from the graph and/or mathematically (see section 2.5.1.5). EG:
The time whose survivor function is 0.50 is the median.(Point M)
The time whose survivor function is 0.25 is the upper quartile.(Point U)
The time whose survivor function is 0.75 is the lower quartile.(Point L)
On the graph, define a point P, where y=0.25 and x=lower quartile time, such that lines LP and PU are formed.
The difference between the times whose survivor functions are 0.25 and 0.75 respectively is called the INTERQUARTILE RANGE. Graphically, this represents the length of the line PU.
The average of the two times whose survivor functions are 0.25 and 0.75 respectively is called the MID-QUARTILE RANGE.
Graphically, this represents the area of the triangle LPU.

A typical example of this graph is radioactive decay with a constant half-life (in the case of the exponential distribution). When k<1 the rate of change of the survivor function with respect to time (ie d*F*/dt) for the Weibull distribution is less than that of the exponential distribution and when k>1 the rate of change of the survivor function with respect to time (ie d*F*/dt) for the Weibull distribution is greater than that of the exponential distribution. Survivor functions are useful in many real models, eg we may want to model the chance of survival when a man is age x years old.

2.5.1.2 Hazard function

If k=1, we have the exponential distribution. Plotting a graph of ln h against $(t)^{k-1}$ gives a straight line with gradient (k-1) and intercept ln k + ln p + (k-1) ln p. It is possible to plot a graph of the hazard function against time. Such a graph is shown in Figure 2.2 for the uncensored case, with k=2.5 and p=1.5. The result should be an increasing curve when k>2, a decreasing curve when k<2 or a straight line when k=1,2. When k=1 or k=2, the rate of change of the hazard function with respect to time (ie dh/dt) for the Weibull distribution is constant. When k=1, the gradient is 0 and when k=2, the gradient is (2*p*p). When k<3, the rate of change of the hazard function with respect to time (ie dh/dt) for the Weibull distribution decreases with increasing time and when k>=3 the rate of change of the hazard function with respect to time (ie dh/dt) increases with increasing time.

2.5.1.3 PDFs

A graph of PDF against time is shown in Figure 2.3. Putting (df/dt)=0 gives the maximum PDF and the maximum rate of change of the survivor function (ignore sign). Point X on the graph represents this maximum rate of change of the survivor function.

2.5.1.4 Integrated Hazard

When k=1, we have the exponential distribution. Plotting a graph of ln H against ln t gives a straight line with gradient k and intercept . A graph of integrated hazard against time is shown in Figure 2.4. When k<1, the integrated hazard decreases with increasing time, k=1 the integrated hazard increases linearly with increasing time and when k>1 the integrated hazard increases non-linearly with increasing time. The rate of change of integrated hazard with respect to time (ie dH/dt) is always positive. When k>2, the rate of change of integrated hazard with respect to time (ie dH/dt) increases with increasing time, decreases otherwise.

2.5.1.5 Statistics

We can find the median, lower quartile, upper quartile, interquartile range and mid-quartile range of the Weibull distribution.

To find the median, put CDF=0.5, ie survivor function=0.5.

$0.5=\exp(-(pt)^k)$

$\ln 0.5=-(p^k)t^k$ (Taking logs on both sides)

$t^k(p^k)=0.693$ (Changing signs)

$t_m=[0.693]^{1/k}/p$ (Equation 2.6)

where t_m is the median time.

To find the lower quartile, put CDF=0.25, ie survivor function=0.75.

$0.75=\exp(-(pt)^k)$

$\ln 0.75=-(p^k)t^k$

$t^k(p^k)=0.$

$t_{lq}=[0.288]^{1/k}/p$ (Equation 2.7)

where t_{lq} is the lower-quartile time.

To find the upper quartile, put CDF=0.75, ie survivor function=0.25.

$0.25=\exp(-(pt)^k)$

$\ln 0.25=-(p^k)t^k$

$t^k(p^k)=0.$

$t_{uq}=[1.386]^{1/k}/p$ (Equation 2.8)

where t_{uq} is the upper-quartile time.

To find the interquartile range, subtract (2.7) from (2.8)

$t_{uq}-t_{lq}=[1.386-0.288]^{1/k}/p=[1.098]^{1/k}/p$ ...(2.9)

To find the mid-quartile range, take the mean of (2.7) and (2.8)

$(t_{uq}+t_{lq})/2=0.5((1.386+0.288)^{1/k}/p=[0.837]^{1/k}/p$ ...(2.10)

An example of these statistics is given in Table 2.1 (From Viewsheet). I can use these statistics in order to compare how well my estimations are (Details on estimation are in Chapter 03).

Index	Rate	Median	Lower-Q	Upper-Q	IQ Range	Mid-Q
2.50000	1.50000	0.57576	0.40502	0.75972	0.35470	0.58237
2.56081	1.40874	0.61519	0.43639	0.80642	0.37004	0.62141
2.53085	1.40673	0.61503	0.43450	0.80880	0.37429	0.62165
2.57465	1.41836	0.61149	0.43456	0.80041	0.36584	0.61749
2.82733	1.39798	0.62835	0.46039	0.80292	0.34253	0.63165

Table 2.1 Examples of the statistics

I may use these equations in later work (especially for accelerated life and proportional hazard models).

2.6 Summary

For any continuous probability distribution, the PDF f(t) can be related to the survivor function, hazard function and/or integrated hazard :-

$$f(t)=-F'(t) \quad \text{OR}$$
$$f(t)=h(t)F(t) \quad \text{OR}$$
$$f(t)=h(t)\exp[-H(t)]$$

The reason for the minus sign in the first equation is that the survivor function DECREASES as time increases.

When k=1 in the Weibull distribution, we get the exponential distribution. The survivor function of any probability distribution is 1 when t=0 and tends to 0, when t tends to infinity. Its rate of change with respect to time is always negative and is a maximum when the PDF is a maximum. It has lots of special features that can easily be deduced graphically and/or mathematically, eg we can find the median, lower quartile, upper quartile, interquartile range, mid-quartile range, etc.

When we have an exponential distribution, the survivor function can be used in modelling radioactive decay. The half-life is defined as the time whose survivor function is 0.5.
If k<1, the survivor function of the Weibull distribution is less steeper than that of the exponential distribution.
If k>1, the survivor function of the Weibull distribution is steeper than that of the exponential distribution.

The hazard function of any probability distribution is 0 when t=0 and tends to infinity, when t tends to infinity.
If the hazard function is constant, the distribution is exponential.
If the hazard function has a constant non-zero slope when k=2, the distribution is Weibull.

The integrated hazard of any probability distribution is 0 when t=0 and tends to infinity, when t tends to infinity.
If the integrated hazard has a constant slope, the distribution is exponential, Weibull otherwise.

2.7 Later Work

Chapter 2 has important uses in later work :

PDFs are relevant for likelihood functions, MLEs, exponential family considerations, accelerated life and proportional hazard models.

Survivor functions, hazard functions and integrated hazards are relevant for accelerated life and proportional hazard models.

I will be doing the following in later work :

(1) use F77 programs to demonstrate that the product limit estimator gives some estimate of the survivor function and use the K-S test.

(2) use F77 programs to demonstrate the accelerated life and proportional hazard models for various statistical techinques, eg product limit estimator.

In addition, I will be considering exponential family and their properties in Chapter 3.