

5. Parametric Regression Model

The Accelerated Failure Time (AFT) Model

Denote by $S_1(t)$ and $S_2(t)$ the survival functions of two populations. The **AFT model** says that there is a constant $c > 0$ such that

$$S_1(t) = S_2(ct) \text{ for all } t > 0$$

It implies that the aging rate of population 1 is c times as much as that of population 2.

let μ_i be the mean survival time for population i and let φ_i be the population quantiles such that $S_i(\varphi_i) = \theta$ for some $\theta \in (0,1)$. Then

$$\mu_2 = \int_0^{\infty} S_2(t) dt = c \int_0^{\infty} S_2(cu) du = c \int_0^{\infty} S_1(u) du = c\mu_1$$

$$S_2(\varphi_2) = \theta = S_1(\varphi_1) = S_2(c\varphi_1)$$

Assume that $S_2(t)$ is a strictly decreasing function. Then we have

$$\varphi_2 = c\varphi_1$$

This simple argument shows that under the accelerated failure time model, the expected survival time, median survival time of population 2 all are c times as much as those of population 1.

Life Regression

Suppose we have a sample of size n from a target population. For subject i ($i = 1, 2, \dots, n$), we have observed values of covariates $z_{i1}, z_{i2}, \dots, z_{ip}$ and possibly censored survival time T_i . The procedure *Proc Lifereg* in SAS fits models to data specified by the following equations:

$$\log(T_i) = \beta_0 + \beta_1 z_{i1} + \dots + \beta_p z_{ip} + \sigma \varepsilon_i$$

Where β_0, \dots, β_p are the regression coefficients of interest, σ is the scale parameter and ε_i are the random disturbance terms, usually assumed to be independent and identically distributed with some density function $f(\varepsilon)$.

Consider the interpretation of β_k ($k = 1, \dots, p$): Holding other covariate values fixed, let us increase covariate z_k by one unit from z_k to $z_k + 1$ and denote by T_1 and T_2 the corresponding survival times for the two populations with covariate values z_k and $z_k + 1$ (with other covariate values fixed). Then T_1 and T_2 can be expressed as

$$T_1 = e^{\beta_0 + \beta_1 z_1 + \dots + \beta_k z_k + \dots + \beta_p z_p} e^{\sigma \varepsilon_1} = c_1 e^{\sigma \varepsilon_1}$$

$$T_2 = e^{\beta_0 + \beta_1 z_1 + \dots + \beta_k (z_k + 1) + \dots + \beta_p z_p} e^{\sigma \varepsilon_2} = c_2 e^{\sigma \varepsilon_2}$$

Where, $c_2 = c_1 * e^{\beta_k}$

The corresponding survival functions are

$$S_1(t) = P[T_1 \geq t] = P[c_1 e^{\sigma \varepsilon_1} \geq t] = P[e^{\sigma \varepsilon_1} \geq c_1^{-1} t]$$

$$S_2(t) = P[T_2 \geq t] = P[c_2 e^{\sigma \varepsilon_2} \geq t] = P[e^{\sigma \varepsilon_2} \geq c_2^{-1} t]$$

Since ε_1 and ε_2 have the same distribution and $c_2 = c_1 * e^{\beta_k}$

$$S_2(e^{\beta_k} t) = P[e^{\sigma \varepsilon_2} \geq c_2^{-1} e^{\beta_k} t] = P[e^{\sigma \varepsilon_2} \geq c_1^{-1} e^{-\beta_k} e^{\beta_k} t] = P[e^{\sigma \varepsilon_1} \geq t] = S_1(t)$$

So, we have accelerated failure time model between populations 1 (covariate value = z_k) and 2 (covariate value = $z_k + 1$) with $c = e^{\beta_k}$.

If we increase the covariate value of z_k by one unit while holding other covariate values unchanged, the corresponding average survival time μ_2 and μ_1 will be related by $\mu_2 = e^{\beta_k} \mu_1$. If β_k is small, then

$$\frac{\mu_2 - \mu_1}{\mu_1} = e^{\beta_k} - 1 \approx \beta_k$$

Similarly,

$$\frac{\varphi_2 - \varphi_1}{\varphi_1} = e^{\beta_k} - 1 \approx \beta_k$$

Therefore, when β_k is small, it can interpreted as the percentage increase if $\beta_k > 0$ or percentage decrease if $\beta_k < 0$ in the average survival time and/or median survival time when we increase the covariate value of z_k by one unit.

Some Popular AFT models

$$\log(T) = \beta_0 + \beta_1 z_1 + \cdots + \beta_p z_p + \sigma \varepsilon$$

some popular distributions:

Distribution of ε	Distribution of T	Syntax in Proc Lifereg
extreme values (2 par.)	Weibull	dist = weibull
extreme values (1 par.)	exponential	dist = exponential
log-gamma	gamma	dist = gamma
logistic	log-logistic	dist = llogistic
normal	log-normal	dist = lnormal

In *Proc Lifereg* of SAS, all models are named for the distribution of T rather than the distribution of ε . Although these above models fitted by *Proc Lifereg* all are AFT models (so the regression coefficients have a unified interpretation), different distributions assume different shapes for the hazard function.

The exponential Model

For the model:

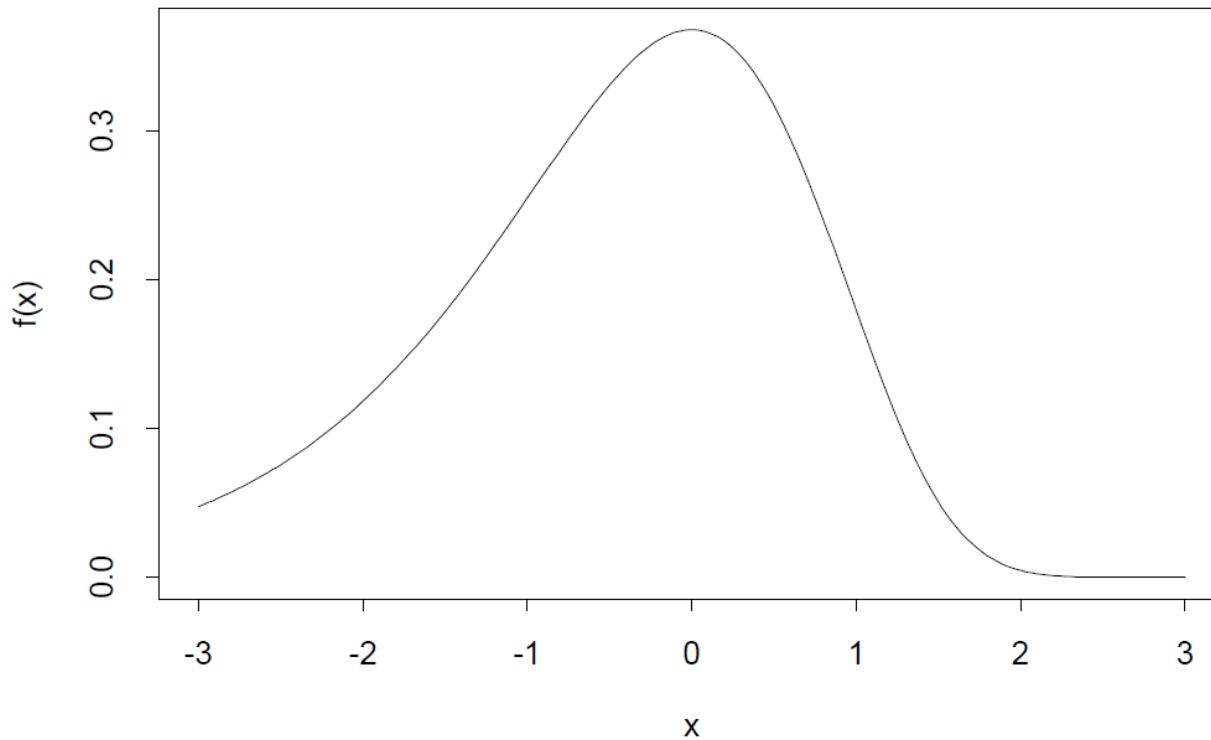
$$\log(T) = \beta_0 + \beta_1 z_1 + \cdots + \beta_p z_p + \sigma \varepsilon,$$

if we assume that $\sigma = 1$ and ε has a standard *extreme value* distribution ($f(\varepsilon) = e^{\varepsilon - e^{\varepsilon}}$), then T follows an exponential distribution:

$$T \sim \exp(e^{-\beta_0 - \beta_1 z_1 - \cdots - \beta_p z_p})$$

Note: e^{ε} has the standard exponential distribution with constant hazard 1.

The density function of the standard extreme value distribution



Notice that distribution of T at any covariate vector z is exponential with constant hazard (independent of t), i.e.

$$\lambda(t|z) = e^{-\beta_0 - \beta_1 z_1 - \dots - \beta_p z_p}$$

So automatically, we get *proportional hazards models*. For a given set of covariates (z_1, z_2, \dots, z_p) , the corresponding survival function is

$$S(t|z) = e^{-\lambda(t|z)}$$

Let $\beta_j^* = -\beta_j$. Then equivalently

$$\lambda(t|z) = e^{\beta_0^* + \beta_1^* z_1 + \dots + \beta_p^* z_p}$$

Therefore, if we increase the value of covariate z_k ($k = 1, \dots, p$) by one unit from z_k to $z_k + 1$ while holding other covariate values fixed, then the ratio of the corresponding hazards is equal to

$$\frac{\lambda(t|z_k + 1)}{\lambda(t|z_k)} = e^{\beta_k^*}$$

Thus $e^{\beta_k^*}$ can be interpreted as the hazard ratio corresponding to one unit increase in the covariate z_k , or equivalently, β_k^* can be interpreted as the increase in **log**-hazard as the value of covariate z_k increases by one unit (while other covariate values being held constant).

Note: Another SAS procedure *Proc Phreg* fits a proportional hazards model to the data and outputs the regression coefficient estimates in **log**-hazard form (i.e., in β_k^*). If an exponential model fits the data well, the regression coefficient estimates in outputs from *Proc Lifereg* using `dist=exponential` and *Proc Phreg* should be just opposite to each other.

Example

<http://www.mcw.edu/biostatistics/Faculty/Faculty/JohnPKleinPhD/SurvivalAnalysisBook/DataSetsBothEditions.htm> (BMT for HD and NHL data, from section 1.10):

Data on 43 bone marrow transplant patients were collected. Patients had either Hodgkin's disease (*hodgkins*=1) or Non-Hodgkin's Lymphoma (*hodgkins*=0), and were given either an allogeneic (Allo, *allo*=1) transplant (from a HLA match sibling donor) or autogeneic (Auto, *allo*=0) transplant (their own marrow was cleansed and returned to them after a high dose of chemotherapy). Other covariates are Karnofsky score (a subjective measure of how well the patient is doing, ranging from 0-100) and waiting time (in months) from diagnosis to transplant. It is of substantial interest to see the difference in leukemia-free survival (in days) between those patients given an Allo or Auto transplant, after adjusting for patients' disease status, Karnofsky score and waiting time. (SAS code)

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits	Chi-Square	Pr > ChiSq	
Intercept	1	0.6834	0.6408	-0.5726 1.9394	1.14	0.2862	
allo	1	0.1333	0.4607	-0.7697 1.0363	0.08	0.7723	
hodgkins	1	-1.3185	0.5275	-2.3524 -0.2845	6.25	0.0124	
kscopy	1	0.0758	0.0094	0.0574 0.0942	64.90	<.0001	
wtime	1	0.0093	0.0072	-0.0049 0.0235	1.66	0.1975	
Scale	0	1.0000	0.0000	1.0000 1.0000			
Weibull Shape	0	1.0000	0.0000	1.0000 1.0000			

The Weibull Model

For the model:

$$\log(T) = \beta_0 + \beta_1 z_1 + \cdots + \beta_p z_p + \sigma \varepsilon,$$

if we assume ε has a standard *extreme value* distribution ($f(\varepsilon) = e^{\varepsilon - e^\varepsilon}$), then T follows an Weibull distribution with survival function of:

$$S(t|z) = \exp \left\{ - \left[t e^{-z^T \beta} \right]^{1/\sigma} \right\}.$$

Note: The only difference between the Weibull model and the exponential model is that the scale parameter σ is estimated rather than being set to be one.

Equivalently, in terms of log hazard function:

$$\log[\lambda(t|z)] = \left(\frac{1}{\sigma} - 1 \right) \log(t) - \log(\sigma) - z^T (\beta/\sigma)$$

Let $\alpha = 1/\sigma$, $\beta_0^* = -\log(\sigma) - \beta_0/\sigma$, and $\beta_j^* = -\beta_j/\sigma$ for $j = 1, \dots, p$. Then

$$\log[\lambda(t|z)] = (\alpha - 1) \log(t) + \beta_0^* + \beta_1^* z_1 + \cdots + \beta_p^* z_p$$

Thus we also get a proportional hazards model and the coefficient β_k^* ($k = 1, \dots, p$). also has the interpretation that it is the increase in **log**-hazard when the value of covariate z_k increases by one unit while other covariate values being held unchanged. The function

$$\lambda_0(t) = t^{\alpha-1} e^{\beta_0^*} = \alpha t^{\alpha-1} e^{-\beta_0/\sigma} = \alpha t^{\alpha-1} e^{-\alpha \beta_0}$$

is the baseline hazard (*i.e.*, when $z = 0$).

Note: If the Weibull model is a reasonable model for your data and you use Proc Lifereg and Proc Phreg to fit the data, then the regression coefficient estimates not only have opposite signs (except possibly for the intercept) but also have different magnitude (depending on whether $\sigma > 1$ or $\sigma < 1$), since $\beta_k^* = -\beta_k/\sigma$ for $k = 1, \dots, p$.

However, testing $H_0: \beta_k^* = 0$ is equivalent to testing $H_0: \beta_k = 0$

Example: BMT data revisited (SAS code)

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	0.4258	0.8463	-1.2329	2.0845	0.25	0.6148
allo	1	0.2080	0.5659	-0.9012	1.3172	0.14	0.7132
hodgkins	1	-1.3746	0.6465	-2.6417	-0.1075	4.52	0.0335
kscore	1	0.0793	0.0121	0.0557	0.1029	43.22	<.0001
wtime	1	0.0104	0.0094	-0.0081	0.0289	1.22	0.2692
Scale	1	1.2733	0.2044	0.9297	1.7440		
Weibull Shape	1	0.7854	0.1260	0.5734	1.0757		

Output

Question: How to tell whether exponential model or Weibull model fits data better?

The log-normal Model

For the model:

$$\log(T) = \beta_0 + \beta_1 z_1 + \cdots + \beta_p z_p + \sigma \varepsilon,$$

if we assume ε has a standard *normal* distribution ($\varepsilon \sim N(0,1)$), then T follows a log-normal distribution with hazard function of:

$$\log[\lambda(t|z)] = \log \left[\lambda_0 \left(t e^{-z^T \beta} \right) \right] - z^T \beta.$$

Where,

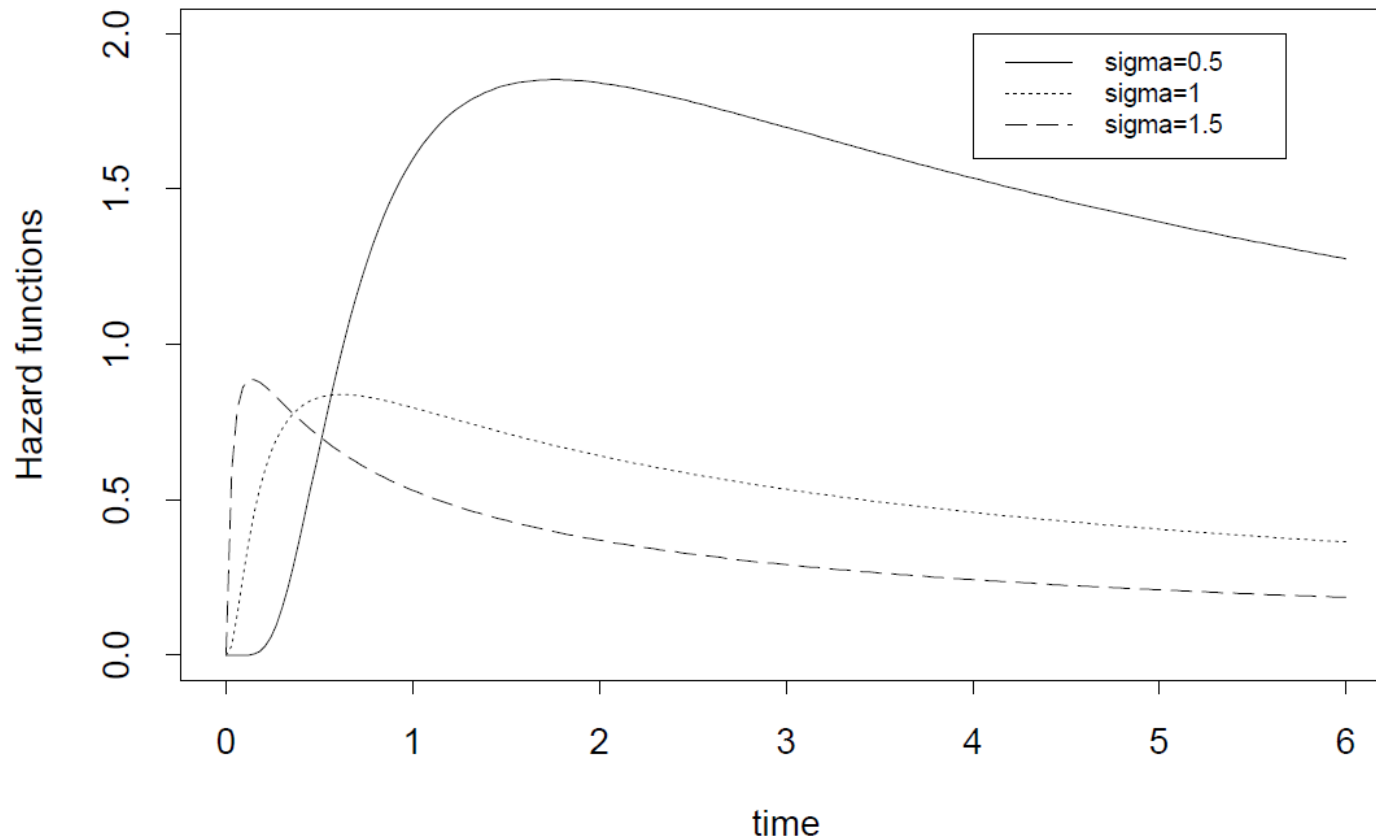
$$\lambda_0(t) = \frac{\phi \left(\frac{\log(t)}{\sigma} \right)}{\left[1 - \Phi \left(\frac{\log(t)}{\sigma} \right) \right] \sigma t},$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{and} \quad \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

Note:

1. Obviously this is no longer a proportional hazards model.
2. The function $\lambda_0(t)$ is not the baseline hazard function. If such a function is desired, it can be obtained from the equation above by setting $z = 0$.

Some typical patterns that the hazard function $\lambda_0(t)$



The inverted U-shaped of the log-normal hazard is often appropriate for repeated events such as a residential move (*i.e.*, the interest is the time to next move). Immediately after a move, the hazard for another move is likely to be low, then increases with time, and eventually begins to decline since people tend to not move as they get older.

The survival function $S(t|z)$ at any covariate value z can be expressed as

$$\Phi^{-1}[S(t|z)] = \beta_0^* + \beta_1^* z_1 + \cdots + \beta_p^* z_p - \alpha \log(t)$$

Or equivalently,

$$S(t|z) = \Phi[\beta_0^* + \beta_1^* z_1 + \cdots + \beta_p^* z_p - \alpha \log(t)]$$

Where $\alpha = 1/\sigma$, and $\beta_j^* = \beta_j/\sigma$ for $j = 0, 1, \dots, p$.

This is a *probit regression model* with intercept depending on t .

Note: The survival function indicates that the coefficients β_j^* can be estimated using *Proc Logistic* or *Proc Genmod* by specifying probit link function. Specifically, pick a time point of interest, say t_0 . Then dichotomize each subject based on his/her survival status at t_0 (in this case $\alpha \log(t_0)$ is absorbed into the intercept). Of course, there are some limitations using this approach. First, there should not be censoring prior to time t_0 . Second, the scale parameter σ is not estimable. Third, we will lose efficiency since we did not use all the information on the exact timing of the events. Since normal distribution and the logistic distribution we will introduce soon behave similarly to each other, the parameters β_j^* here have similar interpretation as the parameters in log-logistic model.

Example: BMT data revisited ([SAS code](#))

Example: BMT data revisited ([SAS code](#))

Output

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	0.5064	1.1563	-1.7600	2.7727	0.19	0.6614
allo	1	0.3520	0.5902	-0.8048	1.5088	0.36	0.5509
hodgkins	1	-1.3137	0.6898	-2.6656	0.0382	3.63	0.0568
kscore	1	0.0663	0.0143	0.0383	0.0944	21.54	<.0001
wtime	1	0.0142	0.0115	-0.0083	0.0367	1.53	0.2160
Scale	1	1.6296	0.2441	1.2149	2.1857		

The results from this model are quite similar to the results from other models.

The Log-Logistic Model

For the model:

$$\log(T) = \beta_0 + \beta_1 z_1 + \cdots + \beta_p z_p + \sigma \varepsilon,$$

if we assume ε has a standard logistic distribution

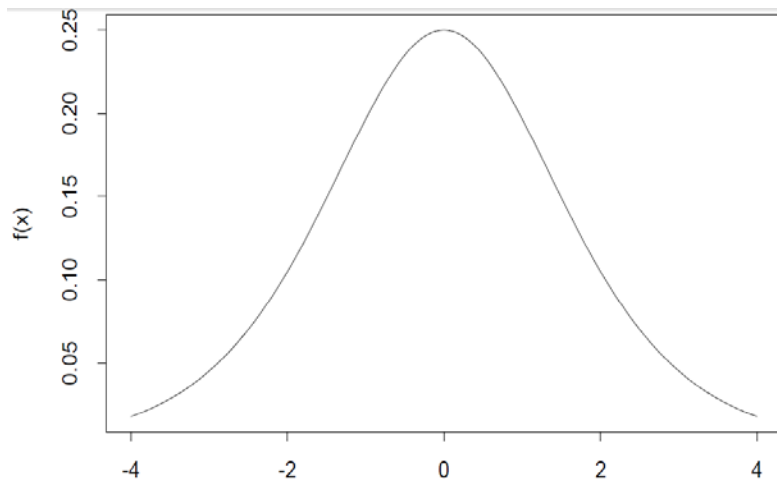
$$f(\varepsilon) = \frac{e^\varepsilon}{(1+e^\varepsilon)^2},$$

then T follows an log-logistic distribution with hazard function of:

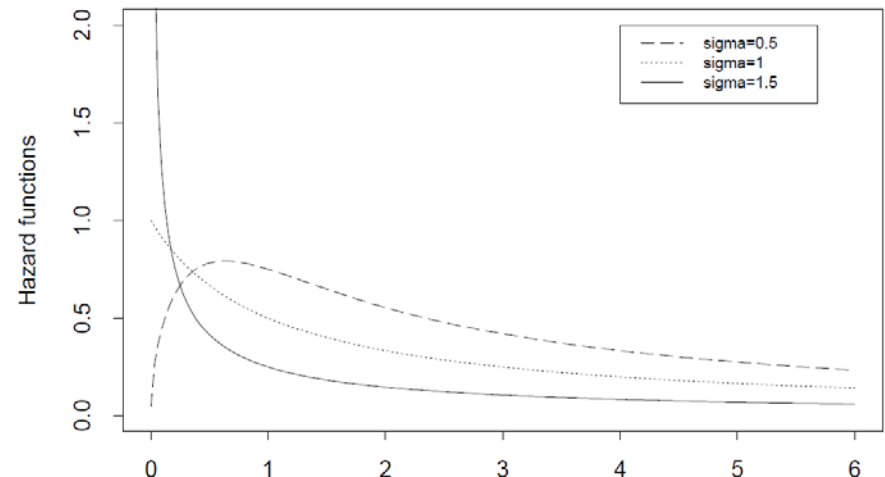
$$\lambda(t|z) = \frac{\alpha t^{\alpha-1} e^{-z^T \beta / \sigma}}{1 + t^\alpha e^{-z^T \beta / \sigma}}$$

Where $\alpha = 1/\sigma$.

Density function of standard logistic distribution



Hazard functions for the log-logistic distribution



The random variable T has a very simple survival function at covariate value z .

$$S(t|z) = \frac{1}{1 + (te^{-z^T \beta})^{1/\sigma}}$$

Some simple algebra then shows that

$$\log \left[\frac{S(t|z)}{1 - S(t|z)} \right] = \beta_0^* + \beta_1^* z_1 + \dots + \beta_p^* z_p + \alpha \log(t)$$

where $\beta_j^* = -\beta_j/\sigma$ for $j = 0, 1, \dots, p$.

This is nothing but a *logistic regression model* with the intercept depending on t . Since $S(t|z)$ is the probability of surviving to time t for any given time t , the ratio $\frac{S(t|z)}{1-S(t|z)}$ is often called the *odds* of surviving to time t . Therefore, with one unit increase in z_k while other covariates being held fixed, the *odds ratio* is given by

$$\frac{S(t|z_k + 1)/(1 - S(t|z_k + 1))}{S(t|z_k)/(1 - S(t|z_k))} = e^{\beta_k^*}$$

which is a constant over time. Therefore, we have a *proportional odds models*. Hence β_k^* can be interpreted as the log odds ratio (for surviving) with one unit increase in z_k and $-\beta_k^*$ is the log odds ratio of dying before time t with one unit increase in z_k . At the times when the event of failure is rare (such as the early phase of a study), $-\beta_k^*$ can also be approximately interpreted as the log relative risk of dying. The log-logistic model is the *only* one that is both an AFT model and a proportional odds model.

Obviously, ε has the following cumulative distribution function

$$F(\varepsilon) = \frac{e^\varepsilon}{1 + e^\varepsilon}, \varepsilon \in (-\infty, \infty)$$

whose inverse function

$$\text{logit}(\pi) = \log\left[\frac{\pi}{1-\pi}\right], \quad \pi \in (0,1),$$

is often called the logit function.

Note: As in the case of log-normal distribution, the regression coefficients β_k^* can be estimated using *Proc logistic* or *Proc Genmod*. See the log-normal model for the procedure and the limitations of such approach.

Example: BMT data revisited (**SAS code**)

Analysis of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	0.5715	1.0620	-1.5101	2.6530	0.29	0.5905
allo	1	0.3419	0.5764	-0.7878	1.4715	0.35	0.5531
hodgkins	1	-1.4494	0.6693	-2.7613	-0.1376	4.69	0.0303
kscore	1	0.0669	0.0137	0.0402	0.0937	24.04	<.0001
wtime	1	0.0131	0.0108	-0.0081	0.0343	1.46	0.2264
Scale	1	0.9382	0.1567	0.6763	1.3016		

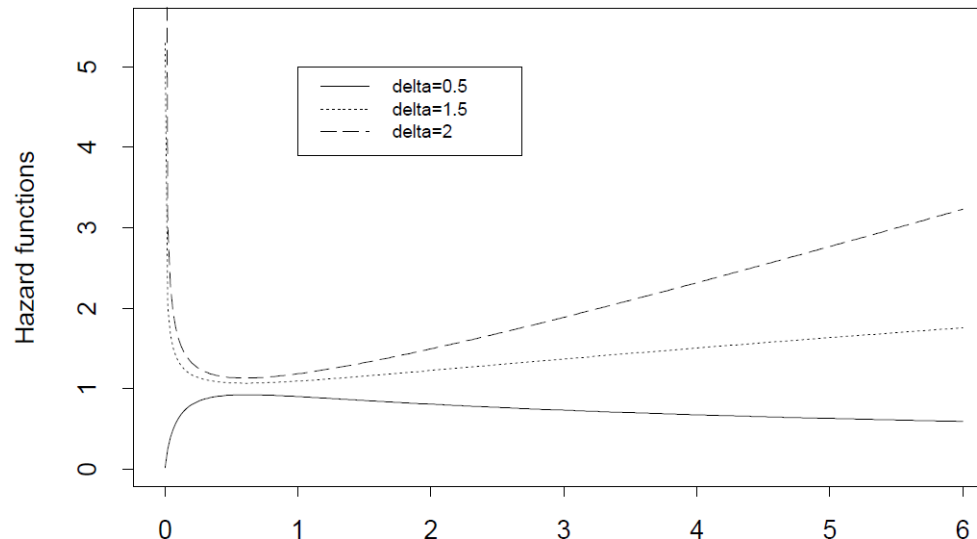
Output

The Gamma Model

The procedure *Proc Lifereg* in SAS actually fits a generalized gamma model (not a standard gamma model) to the data by assuming $T_0 = e^\varepsilon$ to have the following density function

$$f(t) = \frac{|\delta| \left(\frac{t^\delta}{\delta^2} \right)^{\frac{1}{\delta^2}} \exp \left(-\frac{t^\delta}{\delta^2} \right)}{t \Gamma \left(\frac{1}{\delta^2} \right)}$$

where δ is called shape with label Gamma shape parameter in the output of *Proc Lifereg* when we specify dist=gamma. The hazard function of this gamma distribution does not have a closed form and is presented in graph below for some δ 's. **Note:** They are **not** the hazard functions of the standard gamma distribution.



Clearly from this plot, the hazard function is an inverted U-shaped function of time when $\delta < 1$ and takes a U-shape when $\delta > 1$. This feature makes gamma model very appropriate to model the survival times, especially for human. In practice, the hazard function is determined jointly by the scale parameter σ and the shape parameter δ and we need to examine the resulting hazard function case by case.

For a given set of covariates (z_1, z_2, \dots, z_p) , let $c = e^{\beta_0 + \beta_1 z_1 + \dots + \beta_p z_p} = e^{z^T \beta}$. Then $\log(T) = z^T \beta + \sigma \varepsilon$ implies $T = e^{z^T \beta} * [e^\varepsilon]^\sigma$. Hence the survival function for this population is

$$S(t|z) = P[T \geq t] = P[cT_0^\sigma \geq t] = P\left[T_0 \geq (c^{-1}t)^{\frac{1}{\sigma}}\right] = P[T_0 \geq b(t)]$$

$$= \int_{b(t)}^{\infty} \frac{|\delta| \left(\frac{x^\delta}{\delta^2}\right)^{\frac{1}{\delta^2}} \exp\left(-\frac{x^\delta}{\delta^2}\right)}{x \Gamma\left(\frac{1}{\delta^2}\right)} dx = \begin{cases} \int_{\alpha b(t)^\delta}^{\infty} \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy & \text{if } \delta > 0 \\ \int_0^{\alpha b(t)^\delta} \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy & \text{if } \delta < 0 \end{cases}$$

where $\alpha = \frac{1}{\delta^2}$. This final integration can be calculated using built-in functions in any popular software.

Example: BMT data revisited (SAS code)

Output

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	0.4299	1.1407	-1.8057	2.6656	0.14	0.7062
allo	1	0.3734	0.5870	-0.7772	1.5240	0.40	0.5247
hodgkins	1	-1.3098	0.6784	-2.6394	0.0199	3.73	0.0535
kscore	1	0.0696	0.0206	0.0292	0.1100	11.38	0.0007
wtime	1	0.0133	0.0120	-0.0102	0.0367	1.23	0.2678
Scale	1	1.5771	0.3805	0.9829	2.5307		
Shape	1	0.2047	1.0278	-1.8098	2.2192		

Some special cases:

- $\delta = 1 \Rightarrow T|z$ has the Weibull distribution.
- $\delta = 0 \Rightarrow T|z$ has the log-normal distribution. We need the following approximation in order to show this:

$$\Gamma(x) \approx \sqrt{2\pi} x^{x-\frac{1}{2}} e^{-x} \text{ as } x \rightarrow \infty$$

- $\delta = \sigma \Rightarrow T|z$ has the standard gamma distribution with the following density:

$f(t|z) = \frac{t^{K-1} e^{-\frac{t}{\gamma}}}{\gamma^K \Gamma(K)}$, where $K = \frac{1}{\delta^2}$ is the shape parameter and $\gamma = \delta^2 e^{z^T \beta}$ is the scale parameter in the standard gamma distribution.

Goodness-of-fit Using Likelihood Ratio Test

1. generalized gamma $(\delta, \sigma) \supset$ standard gamma $(\delta = \sigma) \supset$ exponential distribution $(\delta = \sigma = 1)$.
2. generalized gamma $(\delta, \sigma) \supset$ Weibull $(\delta = 1, \sigma) \supset$ exponential distribution $(\delta = \sigma = 1)$.
3. generalized gamma $(\delta, \sigma) \supset$ log-normal $(\delta = 0, \sigma)$.

We can use the above nested models to conduct likelihood ratio test for the bone marrow transplant data. Below are Maximum likelihood values for different models

Model	Maximum Likelihood
Gamma	-60.91
Log-logistic	-61.15
Log-normal	-60.93
Weibull	-61.21
Exponential	-62.49

Assuming the Gamma model is a reasonable model for the data, the LRT indicates that the log-normal model is equally good and the Weibull model is also acceptable. Since the log-normal model and the Weibull model have the same number of parameters, we might want to take the log-normal model as the final model based on the larger maximum log-likelihood value.

$\delta = 0 \Rightarrow T|z$ has the log-normal distribution

$$T = e^{z^T \beta} (T_0)^\sigma, \text{ so } T_0 = \left(T e^{-z^T \beta} \right)^{\frac{1}{\sigma}},$$

$$\frac{dT_0}{dT} = \frac{1}{\sigma} \left(T e^{-z^T \beta} \right)^{\frac{1}{\sigma}-1} e^{-z^T \beta}, T_0^\delta = \left(T e^{-z^T \beta} \right)^{\frac{\delta}{\sigma}}$$

Density of T_0 :

$$f(t_0) = \frac{|\delta| \left(\frac{t_0^\delta}{\delta^2} \right)^{\frac{1}{\delta^2}} \exp \left(-\frac{t_0^\delta}{\delta^2} \right)}{t_0 \Gamma \left(\frac{1}{\delta^2} \right)}$$

$$f(t) = \frac{|\delta| \left(\frac{(te^{-z^T \beta})^\delta}{\delta^2} \right)^{\frac{1}{\delta^2}} \exp \left(-\frac{(te^{-z^T \beta})^\delta}{\delta^2} \right)}{(te^{-z^T \beta})^{\frac{1}{\sigma}} \Gamma \left(\frac{1}{\delta^2} \right)} \frac{1}{\sigma} (te^{-z^T \beta})^{\frac{1}{\sigma}-1} e^{-z^T \beta}$$

$$= \frac{|\delta|}{\sigma t} \frac{1}{\Gamma \left(\frac{1}{\delta^2} \right)} \left((te^{-z^T \beta})^{\frac{\delta}{\sigma}} \frac{1}{\delta^2} \right)^{\frac{1}{\delta^2}} \exp \left(-\frac{(te^{-z^T \beta})^\delta}{\delta^2} \right)$$

$$= \frac{|\delta|}{\sigma t} \frac{1}{\Gamma \left(\frac{1}{\delta^2} \right)} e^{\frac{\delta}{\sigma} \frac{(\ln(t) - \ln(e^{z^T \beta})) + \ln \left(\frac{1}{\delta^2} \right)}{\delta^2}} e^{-\frac{\delta}{\sigma} \frac{(\ln(t) - \ln(e^{z^T \beta}))}{\delta^2}}$$

$$= \frac{|\delta|}{\sigma t} \frac{1}{\Gamma \left(\frac{1}{\delta^2} \right)} e^{\frac{\delta \frac{\ln(t) - \mu}{\sigma} + \ln \left(\frac{1}{\delta^2} \right) - e^{\frac{\ln(t) - \mu}{\sigma}}}{\delta^2}}$$

$$\left(\mu = \ln \left(e^{z^T \beta} \right) \right)$$

As $\delta \rightarrow 0$, we have the following approximation:

$$\Gamma\left(\frac{1}{\delta^2}\right) \approx \sqrt{2\pi} \left(\frac{1}{\delta^2}\right)^{\frac{1}{\delta^2}-\frac{1}{2}} e^{-\frac{1}{\delta^2}}$$

$$\Gamma(x) \approx \sqrt{2\pi} x^{x-\frac{1}{2}} e^{-x} \text{ as } x \rightarrow \infty$$

$$e^{\delta \frac{\ln(t)-\mu}{\sigma}} \approx 1 + \delta \frac{\ln(t)-\mu}{\sigma} + \delta^2 \frac{[\ln(t)-\mu]^2}{2\sigma^2}$$

Therefore:

$$\begin{aligned} f(t) &= \frac{|\delta|}{\sigma t} \frac{1}{\Gamma\left(\frac{1}{\delta^2}\right)} e^{\frac{\delta \frac{\ln(t)-\mu}{\sigma} + \ln\left(\frac{1}{\delta^2}\right) - e^{\delta \frac{\ln(t)-\mu}{\sigma}}}{\delta^2}} \\ &\approx \frac{|\delta|}{\sigma t} \frac{1}{\sqrt{2\pi} \left(\frac{1}{\delta^2}\right)^{\frac{1}{\delta^2}-\frac{1}{2}} e^{-\frac{1}{\delta^2}}} e^{\frac{\delta \frac{\ln(t)-\mu}{\sigma} + \ln\left(\frac{1}{\delta^2}\right) - \left(1 + \delta \frac{\ln(t)-\mu}{\sigma} + \delta^2 \frac{[\ln(t)-\mu]^2}{2\sigma^2}\right)}{\delta^2}} \\ &= \frac{1}{\sqrt{2\pi}\sigma t} e^{\frac{[\ln(t)-\mu]^2}{2\sigma^2}} \longleftarrow \text{Log-normal} \end{aligned}$$