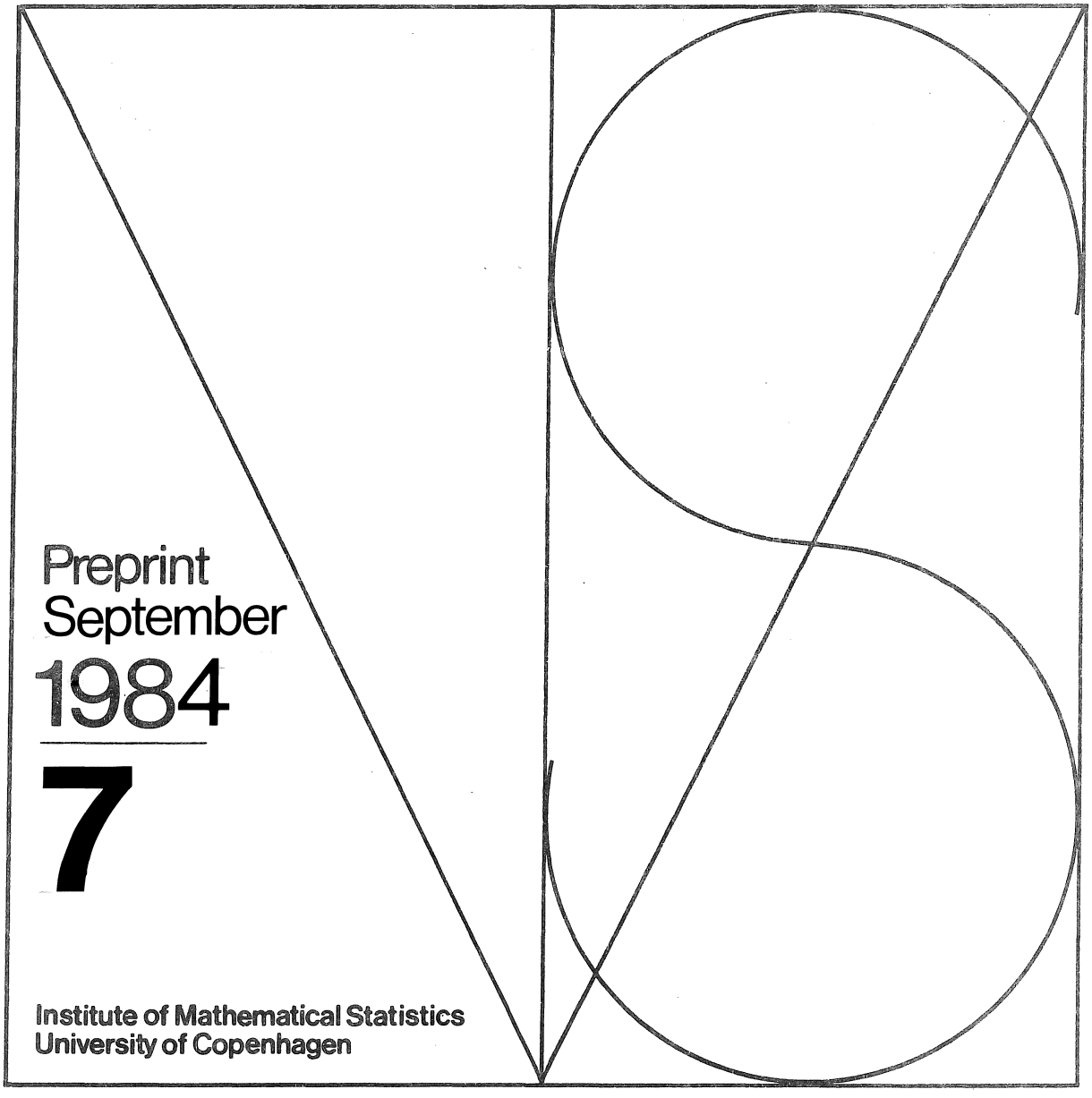
Philip Hougaard

Frailty Models Derived from the Stable Distributions



Philip Hougaard

FRAILTY MODELS DERIVED FROM THE STABLE DISTRIBUTIONS

Preprint 1984 No. 7

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

September 1984

Abstract A new three-parameter family of distributions on the positive numbers is proposed. It includes as special cases the stable distributions on the positive numbers, the gamma distributions, the degenerate distributions and the inverse Gaussian distributions. Only the latter correspond to interior points in the parameter space. The family is characterized by a simple form of the Laplace transform, from which moments, convolutions, infinite divisibility and other properties can be derived. Also there are results about mixtures of certain distributions over the positive stable distributions. The family is a natural exponential family in one of the three parameters. The distributions in the family are natural candidates for application as frailty distribution for use in life table methods for heterogenous populations. Also it is discussed, what properties such distributions preferably should have. As an example the survival after myocardial infarction is considered.

Key words: positive stable distributions; inverse Gaussian distributions; gamma distributions; heterogeneity; life tables; mixtures.

1. INTRODUCTION

The aim of this paper is to derive a family of distributions, which are natural candidates for application as frailty distribution in life tables for heterogenous populations. In Section 2 we discuss what properties such a family of frailty distributions should have.

In the rest of the paper we consider a specific family of distributions, with three parameters α, δ and θ. The family generalizes the inverse Gaussian distributions, which correspond to $\alpha = 1/2$. It is derived from the stable distributions on the positive numbers $(\theta = 0)$. The family is naturally completed to contain the gamma distributions $(\alpha = 0)$. Trivially it contains the degenerate distributions $(\alpha = 1)$. The family has a number of nice properties generalizing those of the inverse Gaussian family. However the family is only an exponential family in one of the three parameters. Another disadvantage is that the density is more complicated than the inverse Gaussian density. In Section 3 we give a short review of the positive stable distributions and in Section 4 we derive the family and examine its properties. In Section 5 the formulae relevant for the application of the new family in frailty models are explicitly stated. In Section 6 we consider the survival after myocardial infarction as an example.

2. LIFE TABLE METHODS FOR HETEROGENOUS POPULATIONS

The idea to suggest and examine the family came from a consideration of life table methods for heterogenous populations. Ordinary life table methods implicitly assume that the population under study is homogenous, i.e. all individuals have the same risk. Vaupel, Manton and Stallard (1979) suggested a model for heterogenous populations, where the population was considered as a mixture of individuals, each having a survival distribution. To each individual there corresponds a quantity, the so-called frailty, denoted Z, describing the relative risk for the individual. The hazard in age t, for an individual with frailty Z is assumed to have the form $Z\mu(t)$, where $\mu(t)$ describes the age dependence. Vaupel et al (1979) suggested using a gamma distribution for Z, because the family of gamma distributions is closed under the selection induced by mortality, that is if the distribution of Z at birth is gamma, also the distribution among survivors in any given age, that is the conditional distribution given survival until that age, is gamma with the original shape parameter, but a different scale parameter. Hougaard (1984) showed that all natural exponential families on the positive numbers shared the same property of being closed under mortality selection. His main example was the inverse Gaussian distributions, because the formulae were rather simple in that case. A family of frailty distributions should preferably be a natural exponential family, because of the property of closedness. Also models created by starting observation in different ages are consistent with each other. If the family does not have this property we can generate a family with the property, introducing one more parameter. The family which we suggest is derived in this way, starting with the positive stable distributions.

Hougaard (1984) discussed a model with cause-specific frailties. Suppose mortality is divided into k causes and let the cause-specific frailties be

$Z_1,\ldots,Z_k$ and assume that the hazard of death of cause i for a given individual conditional on the frailty is $Z_i\mu_i(t)$, where $\mu_i(t)$ describes the age effect. Then the total hazard is $\Sigma_i Z_i\mu_i(t)$. Whether analyzing total mortality or cause-specific mortality we must remember that causes are always made up of a number of subcauses. Suppose that we are analyzing total mortality, but that it would have been more relevant if we had analyzed it with k causes as above. Under the assumption that the cause-specific hazards are proportional, that is there exist $\mu(t)$ and $c_1,\ldots,c_k \in (0,\infty)$, such that $\mu_i(t) = c_i\mu(t)$, the total hazard is $\mu(t)\Sigma c_i Z_i$, also if the cause-specific frailties are dependent. That means that the model for total mortality is a frailty model, with frailty $\Sigma c_i Z_i$. Therefore when we analyze total mortality we should preferably use a frailty distribution, which naturally allows for a description as $\Sigma c_i Z_i$. All distributions allow a number of such descriptions with dependent Z_i's. If the frailty distribution is infinitely divisible there are also such descriptions with independent Z_i's for all combinations of c_i. Probably it will then also include some relevant descriptions with dependent Z_i's. Having an infinitely divisible frailty distribution is therefore not in conflict with the fact that causes can be split into subcauses. As shown in Lemma 3 below the distributions $P(\alpha,\delta,\theta)$ are all infinitely divisible.

The assumption of proportionality between the cause-specific hazards is often unrealistic. However, this is not only a problem for models allowing for heterogeneity. It is the same with ordinary parametric models. For example if the cause-specific hazards are Weibull hazards with different shape parameters, the total hazard cannot be a Weibull hazard.

If there is a large number of possible causes with the same cause-specific hazard and independent identically distributed frailties, we might approximate the distribution of ΣZ_i. If the Z_i's have second moments the distribution

of the sum is approximately normal. Because the frailties are positive we might prefer to approximate with a normal distribution truncated at 0. Probably it only gives little difference at birth, but in older ages it will give a larger difference. The truncated normal distributions are a two parameter exponential family and can be handled using the results from Hougaard (1984). If the second moments do not exist, we might instead approximate by a stable distribution, most naturally by a positive stable distribution. It is an advantage if the frailty distribution is stable, because it then has a domain of attraction, meaning that it can approximate sums of variables from a large class of distributions.

Also it is preferable if the frailty distribution satisfies simple mixture results. If the mixture result can be described as a product result, like (4.4) below for the positive stable distributions, it might describe different risk factors, which act multiplicatively. Hougaard (1984) mentioned a similar product result for the gamma distribution. However such results should not be taken too literally, because there might be many different products giving the same distribution. For example $P(1/2, \delta, 0)$ can be written as a product using (4.4), but because it is the reciprocal of a gamma distribution with shape 1/2, it can also be described as the product of the reciprocal of a gamma with larger shape parameter and the reciprocal of a beta distribution. These products are rather unsymmetrical. It would be better if there was a description as a product of i.i.d. variables, that is if $\log Z$ were infinitely divisible. The positive stable distributions, the gamma distributions and trivially the degenerate distributions are infinitely divisible in the logarithm. However the suggested family does not in general have this property.

In short infinite divisibility of Z allow the causes to be groups of subcauses and infinite divisibility of $\log Z$ allows for a multiplicative

effect of several unobserved risk factors.

Probably the frailties for different causes or subcauses are dependent, thus creating dependence between the competing risks. For a discussion of such dependence, see Prentice et al (1978). The dependence and interaction between causes and the individual differences, which we have attempted to describe by the frailty are all difficult problems. Analyzing these detailed would require a number of extensive studies, including some with data not presently obtainable. At the moment it is impossible to make a "correct" model. That we suggest the distribution to be infinitely divisible, does not mean that we believe, that the subcauses are independent or have proportional hazards, but only that the least we can require is that the method is relevant under simple assumptions.

Infinite divisibility of Z is preserved under the selection induced by mortality. This is, however, not the case for infinite divisibility of the logarithm and for stability. The extreme logical consequence of this, assuming that frailty only describes genetical properties of an individual, is to assume that the distribution is stable at the time of conception, but because observation starts at birth or later, we do not find a stable distribution, but one from the family defined in Section 4. In practice frailty is not solely genetical or even constant through life. It describes risk factors not taken into account by the model and might therefore include smoking habits or present diseases. If it is possible to measure important risk factors, we should of course measure them. But we have to be aware of the possibility, that there are risk factors, which we have not thought about. Also having data on cause of death makes more detailed conclusions possible.

In conclusion about the applicability of these models for life tables for heterogenous populations, I think that the models are not correct, but they include more realistic concepts than ordinary life table models.

3. STABLE DISTRIBUTIONS

A distribution is said to be stable (Feller, 1971, p. 169) if the sum of independent random variables from the distribution suitably normalized follows the same distribution. Suppose $X_1, X_2, \ldots$ are i.i.d. The distribution is then stable if for each n, there exists constants c_n, γ_n such that $X_1 + \ldots + X_n \stackrel{d}{=} c_n X_1 + \gamma_n$, where $\stackrel{d}{=}$ means having the same distribution. In this paper we will only consider the strictly stable distributions, which allow γ_n to be 0. The norming constants have the form $c_n = n^{1/\alpha}$, for some $\alpha \in (0,2]$, called the characteristic exponent. The only stable distributions with finite variance are the normal distributions ($\alpha = 2$) and the degenerate distributions, which have $\alpha = 1$, considered as strictly stable distributions. Usually the degenerate distributions are not stable by definition, because they are uninteresting exceptions for many theorems and formulae. In this paper we will, however, consider them as stable because they are naturally included. The stable distributions on the positive numbers has $\alpha \in (0,1]$ and apart from scale factors they are given by the following form of Laplace transform

$$L(s) = E \exp(-sX) = \exp(-s^{\alpha}). \tag{3.1}$$

This distribution will be denoted $P(\alpha,\alpha,0)$. For $\alpha < 1$, the density is, cf. Feller (1971, p. 583), using $\gamma = -\alpha$

$$f_{\alpha}(x) = \frac{-1}{\pi x} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha+1)}{k!} (-x^{-\alpha})^k \sin(\alpha k \pi) \tag{3.2}$$

These densities are bell-shaped, see Gawronski (1984).

4. DERIVATION AND PROPERTIES OF THE FAMILY

Let X follow the standard stable distribution $P(\alpha,\alpha,0)$ given by the Laplace transform (3.1), where $\alpha \in (0,1]$. The distribution $P(\alpha,\delta,0)$ for $\delta > 0$ is then defined as the distribution of $(\delta/\alpha)^{1/\alpha} X$. This distribution has Laplace transform $L(s) = \exp(-\delta s^{\alpha}/\alpha)$ and the density for $\alpha < 1$ is $f_{\alpha}\{x(\alpha/\delta)^{1/\alpha}\}(\alpha/\delta)^{1/\alpha}$, where f_{α} is defined in (3.2). It follows from the Laplace transform, that for $\theta \geq 0$ and $\alpha \in (0,1)$

$$f_{\alpha}\{x(\alpha/\delta)^{1/\alpha}\}(\alpha/\delta)^{1/\alpha} \exp(-\theta x) \exp(\delta\theta^{\alpha}/\alpha) \qquad (4.1)$$

is a probability density. The distribution is denoted $P(\alpha,\delta,\theta)$. For $\alpha = 1$, we can generalize in the same way, using the relevant one point measures instead of Lebesgue measure. It is easier, however, to define $P(1,\delta,\theta)$ as the degenerate distribution in δ. Thus it does not depend on θ. The operation of including the parameter θ above is actually a standard operation, which because of the special parametrization seems more complicated than necessary. Letting β be the scale parameter, the density has the form

$$\frac{f_{\alpha}(x/\beta) \exp(-\theta x)}{\beta L_{\alpha}(\beta\theta)}$$

The parametrization is chosen to give simpler formulae later on.

Lemma 1. Basic properties of $P(\alpha,\delta,\theta)$

a. The density for $\alpha < 1$ is

$$f(x;\alpha,\delta,\theta) = -\exp(-\theta x + \delta\theta^{\alpha}/\alpha)(\pi x)^{-1}$$

$$\sum_{k=1}^{\infty} \frac{\Gamma(k\alpha+1)}{k!} (-\delta x^{-\alpha}/\alpha)^{k} \sin(\alpha k\pi)$$

b. The Laplace transform is

$$L(s) = \exp\left[-\frac{\delta}{\alpha}\{(\theta+s)^{\alpha} - \theta^{\alpha}\}\right] \qquad (4.2)$$

c. For fixed α,δ the family is an exponential family with natural observation x and natural parameter $-\theta$.

d. For $\theta = 0$ and $\alpha < 1$ the γth moment exists iff $\gamma < \alpha$, in which case it is

$$(\delta/\alpha)^{\gamma/\alpha}\,\Gamma(1-\gamma/\alpha)/\Gamma(1-\gamma)$$

For $\theta > 0$ or $\alpha = 1$ all moments exist and $EX = \delta\theta^{\alpha-1}$

The kth cumulant, $k > 1$, is

$$\kappa_k = \delta(1-\alpha)(2-\alpha)\cdot\ldots\cdot(k-1-\alpha)\theta^{\alpha-k}$$

e. If X follows $P(\alpha,\delta,\theta)$ and $c > 0$, the distribution of cX follows $P(\alpha,c^{\alpha}\delta,\theta/c)$.

Proof a. Combine (3.2) and (4.1). b. The integral can be found by means of the Laplace transform for $P(\alpha,\delta,0)$. c. Immediate from the density in a. d. For $\theta = 0$ and $\alpha < 1$ the result is well known, cf. Feller(1971, p.578) and Shanbhag and Sreehari (1977). The result for $\theta > 0$ is found by differentiation of the Laplace transform at $s = 0$. e. Follows from the Laplace transform.

The parameter space above is $0 < \alpha \leq 1$, $\delta > 0$, $\theta \geq 0$. To some extent it is possible to include the boundary. The value $\delta = 0$ corresponds to a distribution degenerate in 0, which is not interesting. Fixing $\alpha < 1$ and δ the limit $\theta \to \infty$ yields degenerate distributions in 0. By letting δ be a function of θ, we can also find other degenerate distributions for $\theta \to \infty$. Most interesting, however, is that the limits for $\alpha \to 0$ are the gamma distributions. Therefore we will define $P(0,\delta,\theta)$ to be the gamma distribution with shape parameter δ and scale parameter $1/\theta$, that is the distribution with density $\theta^{\delta}x^{\delta-1}\exp(-\theta x)/\Gamma(\delta)$. In this case we need $\theta > 0$. Therefore, the parameter space, $(0,1]\times(0,\infty)\times[0,\infty) \cup \{0\}\times(0,\infty)\times(0,\infty)$, say Ω, is neither open nor closed and it is not a cartesian product. For later reference we note that the Laplace transform for $\alpha = 0$ is

$$L(s) = \{\theta/(\theta+s)\}^{\delta} \tag{4.3}$$

Part c., d. and e. of Lemma 1 generalize to the full parameter set.

Lemma 2 Limits in distributions

a. The family P is continuous in distribution, that is if a sequence of points $(\alpha_n,\delta_n,\theta_n)$ from Ω converges to a point (α,δ,θ) in Ω the distributions $P(\alpha_n,\delta_n,\theta_n)$ converges weakly to $P(\alpha,\delta,\theta)$

b. For fixed α and $\mu > 0$, $P(\alpha,\mu\theta^{1-\alpha},\theta)$ converges weakly to μ for $\theta \to \infty$.

Proof a. For the weak convergence we show pointwise convergence of the Laplace transform. The only step which is nontrivial is $\alpha \to 0$, which we prove for $\theta_n = \theta$, $\delta_n = \delta$. It follows from l'Hopitals rule that

$$L_n(s) \to \exp\left[-\delta d/d\alpha\{(\theta+s)^{\alpha}-\theta^{\alpha}\}\ /d/d\alpha(\alpha)\right]\Big|_{\alpha=0}$$

$$= \exp[-\delta\{\ln(\theta+s) - \ln\theta\}] = \{\theta/(\theta+s)\}^{\delta}.$$

b. The mean and variance of $P(\alpha,\mu\theta^{1-\alpha},\theta)$ is μ and $(1-\alpha)\mu/\theta$ respectively, from which the result trivially follows.

Lemma 3 Convolutions and infinite divisibility.

a. Let $X_1,\ldots,X_n$ be independent, X_i following $P(\alpha,\delta_i,\theta)$ The distribution of $X_1+\ldots+X_n$ is then $P(\alpha,\Sigma\delta_i,\theta)$.

b. $P(\alpha,\delta,\theta)$ is infinitely divisible. Actually for any n it is the sum of n i.i.d. $P(\alpha,\delta/n,\theta)$ variables.

Proof a. The Laplace transform of the sum is the product of Laplace transforms. The result follows because δ enters linearly in the exponent in the Laplace transform, cf. (4.2) and (4.3) . b. Immediate from a.

Actually we can show a stronger result. The distributions are generalized gamma convolutions, (g.g.c.) see Thorin (1977), a property which implies infinite divisibility. For a distribution to be a g.g.c. the Laplace-Stieltjes

transform

$$g(s) = E \exp(sX), \mathrm{Re}(s) \leq 0$$

has to be of the form

$$g(s) = \exp\{as - \int_0^\infty \log(1 - s/y)dU(y)\} ,$$

where $a \geq 0$ and $U(y)$ is nondecreasing, and satisfies three regularity conditions, which will be mentioned in the proof below. Thorin (1977) mentions that the positive stable distribution $P(\alpha,\delta,0)$ for $\alpha \in (0,1)$ is a g.g.c. with $U(y) = \sin(\pi\alpha)(by)^\alpha/\pi$ for some $b > 0$. We will show that all distributions $P(\alpha,\delta,\theta)$ are g.g.c. by means of the following general lemma.

Lemma 4 Generalized gamma convolutions.
If a distribution is a generalized gamma convolution, the same applies for any $\theta \geq 0$ for the distribution given by the density $\exp(-\theta x)/L(\theta)$ with respect to the measure given by the original distribution, with Laplace transform $L(\theta)$.

Proof The Laplace-Stieltjes transform of the distribution is $g_\theta(s) = g(s-\theta)/g(-\theta)$. Using the known form of g yields

$$g_\theta(s) = \exp[as - \int_0^\infty \{\log \frac{1+(\theta-s)/y}{1+\theta/y}\} dU(y)$$

$$= \exp\{as - \int_\theta^\infty \log(1 - s/u)dU(u-\theta)\},$$

which is of the g.g.c. form with

$$\tilde{U}(u) = \begin{cases} 0 & 0 < u \leq \theta \\ U(u-\theta) & u > \theta \end{cases}$$

Thus we only need to show, that $\tilde{U}$ satisfies the three regularity conditions.

1. $\tilde{U}(0) = 0$, clearly

2. $\int_0^1 |\log u| \, d\tilde{U}(u) < \infty$, follows because it is smaller than $\int_0^1 |\log y| dU(y)$, which is finite by assumption.

3. $\int_1^\infty u^{-1} d\tilde{U}(u) < \infty$. The integral equals

$$\int_{\max(1,\theta)}^{\infty} u^{-1} dU(u-\theta) = \int_{\max(0,1-\theta)}^{\infty} (y+\theta)^{-1} \, dU(y)$$

$$\leq \theta^{-1} \int_0^1 dU(y) + \int_1^\infty y^{-1} \, dU(y)$$

$$= \theta^{-1} \, U(1) + \int_1^\infty y^{-1} \, dU(y),$$

which is finite by assumption.

Corollary 1 $P(\alpha,\delta,\theta)$ is a generalized gamma convolution.

Proof For $\alpha \in (0,1)$ it follows from the result of Thorin (1977) and the lemma. For $\alpha = 1$ it is trivially a g.g.c., U the 0-measure and $a = \delta$. Also for $\alpha = 0$ the result is trivially true, $a = 0$ and U concentrated in one point.

Bondesson (1981) extends the generalized gamma convolutions by defining T_γ, $\gamma > 0$, as the distributions allowing a description as

$$\frac{g'(s)}{g(s)} = a + \int_{[0,\infty)} (y-s)^{-\gamma} \, dU(y) \, ,$$

again with U satisfying some regularity conditions. The g.g.c!s then corresponds to $\gamma = 1$. For $\gamma_1 < \gamma_2$ we have $T_{\gamma_1} \subset T_{\gamma_2}$ with strict inclusion. For $\gamma \in (0,1]$ T_γ is exactly the distributions obtained as weak limits of finite convolutions of distributions $P(1-\gamma,\delta,\theta)$.

Also there are mixture and product results for the positive stable distributions. First we describe a product result. If X and Y are independent, X follows $P(\alpha,\delta_X,0)$ and Y follows $P(\gamma,\delta_Y,0)$, then $Z = XY^{1/\alpha}$ follows

$$P(\alpha\gamma, \delta_X^{\gamma}\delta_Y\alpha^{1-\gamma}, 0) \tag{4.4}$$

This result is in Feller (1971,p.176). The value of δ_Z can be calculated by conditioning on Y in the Laplace transform. The same result can be reinterpreted as a mixture result. If Z conditional on Y is $P(\alpha, \delta_X Y, 0)$ and Y is $P(\gamma, \delta_Y, 0)$, then Z is $P(\alpha\gamma, \delta_X^{\gamma}\delta_Y\alpha^{1-\gamma}, 0)$. Our parametrization is suggested because it gives simple mean values for $\theta > 0$, simple convolution results and unifies with the gamma distribution, but inter alia with the loss of a simple value of δ_Z. However if X and Y are standard stable distributions as defined in (3.1) also Z will be a standard stable distribution.

Another product result is in Williams (1977), who showed that for $n = 1,2,\ldots$ the distribution $P(1/n, 1, 0)$ is obtained as $1/(X_1 \cdot X_2 \cdot \ldots \cdot X_{n-1})$, where $X_1, \ldots, X_{n-1}$ are independent, $X_i \sim P(0, i/n, 1)$. For $n = 2$ this is the well known result that $P(1/2, \delta, 0)$ is the reciprocal of a gamma distribution with shape parameter 1/2.

As shown in Bondesson (1978) the gamma distributions are infinitely divisible in the logarithm, that is if X follows $P(0, \delta, \theta)$, $\log X$ has an infinitely divisible distribution. From this and the product result of Williams (1977) it trivally follows that for $1/\alpha$ integer the distributions $P(\alpha, \delta, 0)$ are infinitely divisible in the logarithm. Bondesson (1981) showed that the logarithm of $P(\alpha, \delta, 0)$ is an extended generalized gamma convolution, that is the difference between two g.g.c.'s if and only if $1/\alpha$ is an integer; being an e.g.g.c. implying infinite divisibility. However $P(\alpha, \delta, 0)$ is infinitely divisible in the logarithm for all α and δ, as shown by Shanbhag and Sreehari (1977). It is trivial that the distributions $P(1, \delta, \theta)$ are infinitely divisible in the logarithm. For $\theta > 0$ the logarithm of $P(1/2, \delta, \theta)$ is, however, not infinitely divisible, because the tails are decreasing too fast,

cf. Ruegg (1970) and Bondesson (1981). This shows that infinite divisibility of the logarithm is not preserved under the selection induced by mortality. In conclusion we have infinite divisibility in the logarithm on the boundary of the parameter set, but at least for some interior points that property is not present.

There are also nice mixture results for other distributions, inter alia the Weibull and Gompertz distributions. But first we prove a more general lemma, formulated for the kind of mixtures considered in life table methods for heterogenous populations.

Lemma 5 Mixtures over $P(\alpha,\delta,\theta)$

Suppose that T conditional on X has a distribution with integrated hazard $X\,M(t)$, that is $P(T>t|X) = \exp\{-X\,M(t)\}$, where $M(t)$ is a known nondecreasing function. If X has the distribution $P(\alpha,\delta,\theta)$, $\alpha \in (0,1]$, the marginal distribution of T has integrated hazard

$$\frac{\delta}{\alpha}[\{\theta + M(t)\}^{\alpha} - \theta^{\alpha}]$$

Proof $P(T > t) = E\,P(T > t|X)$, which can be found from the Laplace transform, see also Hougaard (1984).

Corollary 2 Mixtures of Weibulls and Gompertz' over the positive stable distributions.

a. If the distribution of T given X is Weibull $(\gamma,\beta X)$, that is with hazard $X\gamma\beta t^{\gamma-1}$ and X follows $P(\alpha,\delta,0)$, $\alpha \in (0,1]$ the distribution of T is Weibull $(\alpha\gamma,\delta\beta^{\alpha}/\alpha)$.

b. If the distribution of T given X is Gompertz $(\gamma,\beta X)$, that is with hazard $X\gamma\beta\exp(\gamma t)$ and X follows $P(\alpha,\delta,0)$, $\alpha \in (0,1]$ the distribution of T is Gompertz $(\alpha\gamma,\delta\beta^{\alpha}/\alpha)$.

Jewell (1982) showed that a Weibull (p,r) is a mixture over s of Weibulls (q,s) if and only if $p < q$. The corollary specifies the mixing distribution. Hougaard (1984) found the mixing distribution for $p = q/2$, corresponding to $\alpha = 1/2$. A reformulation of a. as a product result is that if $T \sim$ Weibull (γ,β) and $X \sim P(\alpha,\delta,0)$ independent of T, the product $T X^{-1/\gamma}$ is Weibull$(\alpha\gamma,\delta\beta^{\alpha}/\alpha)$. The Gompertz distribution is often parametrized with hazard $a b^{t}$. In this parametrization the result of the marginal distribution is $\tilde{a}\tilde{b}^{t}$, with $\tilde{b} = b^{\alpha}$.

For $\alpha = 1/2$ we have the inverse Gaussian family, with an unusual parametrization. The density is

$$f(x) = \delta\, \pi^{-1/2} \exp(2\delta\theta^{1/2}) x^{-3/2} \exp(-\theta x - \delta^2/x),$$

such that it is an exponential family in the parameters δ^2 and θ.

Also the gamma distributions are an exponential family with two parameters, δ and θ.

Because the density is complicated, maximum likelihood estimation is difficult. For fixed α and δ the family is exponential and therefore maximum likelihood coincides with the method of moments, for the natural observations, that is $\hat{\theta}$ is the solution to $\delta\theta^{\alpha-1} = \bar{X}$, which has a unique solution except for $\alpha = 1$. This equation will also be one of the likelihood equations in the general estimation problem. The method of moments gives an explicit solution. Let s^2 and k be the empirical variance and third cumulant and define $R = s^4/(k\bar{X})$. The moment estimator is then given by

$$\alpha = 2 - 1/(1-R),\ \theta = (1-\alpha)\bar{X}/s^2, \delta = \bar{X}\theta^{1-\alpha},$$

which has a solution for $0 \le R \le 1/2$, where R is defined as 0 for $s^2 = 0$. Another possibility for estimation is to use the Laplace transform, analogous to Heathcote (1977). A paper on estimating the index α of a general stable distribution is DuMouchel (1983), which also gives references

to other papers on that subject. In the life table situation X is not observed, thus requiring a completely different procedure. An example is mentioned in the next section.

5. RESULTS FOR FRAILTY DISTRIBUTIONS FROM THE FAMILY

This section includes formulae for application of the distributions $P(\alpha,\delta,\theta)$ as frailty distribution. That is we assume the hazard for a person with frailty Z is $Z\mu(t)$ and that Z is distributed according to $P(\alpha,\delta,\theta)$. Let $M(t) = \int_0^t \mu(u)du$. Then the distribution of Z among the survivors in age t is $P(\alpha,\delta,\theta+M(t))$, cf. Hougaard (1984).

The parameter space is strange. For $\alpha \in (0,1)$, the condition is $\delta > 0$, $\theta \geq 0$, but for $\alpha = 0$, which gives the gamma distributions, we need $\delta > 0$, $\theta > 0$ and for $\alpha = 1$, the degenerate distributions, we need $\delta > 0$, whereas the distribution does not depend on θ, and we can therefore restrict θ in any way we like, for example $\theta \in \mathbb{R}$, $\theta \geq 0$ or $\theta = 0$. That means that the parameter space is neither open nor closed and it is not a Cartesian product. Some of the interesting hypothesis are on the boundary, for example the hypothesis of stability ($\theta = 0$), the hypothesis of gamma distribution or equivalently constant coefficient of variation ($\alpha = 0$), see below, and the hypothesis of homogeneity ($\alpha = 1$). In the latter case, there are two more problems. Firstly because θ is then a nuisance parameter which is present only under the alternative, and secondly because for large θ's there are distributions arbitrarily close to the degenerate distributions, cf. Lemma 2 b. For $\alpha < 1$ the distributions have density with respect to Lebesgue measure, but for $\alpha = 1$ the distributions are degenerate. Because we do not observe the variable Z itself, but only the survival time, which is a mixture over Z, it is not a problem in estimation that the distributions do not have densities with respect to a common measure.

The relation between the integrated population hazard $\Lambda(t)$ and the integrated individual hazard $M(t)$ is

$$\Lambda(t) = \begin{cases} \delta \ln\{1 + M(t)/\theta\} & \alpha = 0 \\ \delta\,[\,\{\theta + M(t)\}^{\alpha} - \theta^{\alpha}]/\alpha & 0 < \alpha \leq 1, \end{cases}$$

which can be inverted to

$$M(t) = \begin{cases} \theta\,[\exp\{\Lambda(t)/\delta\} - 1] & \alpha = 0 \\ \theta\,[\,\{1 + \alpha\Lambda(t)\delta^{-1}\theta^{-\alpha}\}^{1/\alpha} - 1] & 0 < \alpha \leq 1 \end{cases}$$

For the population hazard the two cases unifies to

$$\lambda(t) = \delta\{\theta + M(t)\}^{\alpha-1}\mu(t)$$

At age t the squared coefficient of variation is $(1-\alpha)\{\theta + M(t)\}^{-\alpha}/\delta$, expressing that for $\alpha = 0$ it is constant, for $0 < \alpha < 1$ it is decreasing converging to 0 and for $\alpha = 1$ it is constant, equal to 0. The gamma distributions are the only distributions for which a non-zero coefficient of variation is constant under mortality selection, as follows from Morris (1982).

With completely unspecified $\mu(t)$ the parameters α,δ and θ are not identifiable, a problem which is discussed in Heckman and Singer (1982) and Hougaard (1984). We will comment on the estimation problem with parametrized $\mu(t)$, assuming a Weibull distribution for each individual. That is suppose $\mu(t) = \gamma t^{\gamma-1}$. Assuming Z following $P(\alpha,\delta,\theta)$ implies that the population hazard is $\lambda(t) = \delta\{\theta + t^{\gamma}\}^{\alpha-1}\,\gamma t^{\gamma-1}$. We could include a multiplicative parameter c, such that $\mu(t) = c\,\gamma\,t^{\gamma-1}$, but as $P(\alpha,\delta,\theta)$ includes a scale parameter, the model is not more general. If $\theta = 0$ the model collapses to a twodimensional model, only $\alpha\gamma$ and δ/α can be estimated. That is testing the hypothesis of stability $(\theta=0)$ involves a nuisance parameter, which is present only under the alternative. This happens also for other distributions than the Weibull, cf. Lemma 5. Apart from the problems of the boundary of the parameter space, this gives a completely ordinary parametric model for the hazard and the four parameters α,δ,θ and γ can be estimated and analyzed

in the usual way, but the estimates have to be derived using an iterative procedure.

6. AN EXAMPLE

As an illustration we will consider the survival after myocardial infarction. This disease is the death of a part of the heart muscle, caused by insufficient oxygen supply. In the first days after the infarction mortality is very high and heart arhytmies and other complications are frequent. After the infarction the mortality decreases. Usually the acute phase with high mortality (about 6 weeks) is interpreted as the result of a process, where the organism adjusts to the reduced heart volume. An alternative or supplementary explanation of the decrease in mortality is heterogeneity. Probably there are large differences between the patients in how well their heart is able to satisfy the needs of the organism. The differences are only partly observable by the localization of the infarction and enzyme values, which correlates to the size of the infarction.

We will suggest a model, that assumes that heterogeneity is the only explanation of the decrease in mortality, that is the individual hazard is assumed constant and the frailty distribution is assumed to be $P(\alpha,\delta,\theta)$. Thus the population hazard is $\delta(\theta+t)^{\alpha-1}$. This model is applied to a data set of 1140 admissions during 1977-79 to Glostrup Hospital, Denmark. This data material is analyzed in detail elsewhere, using a Cox-model with time-dependent covariates, but not taking any heterogeneity into account, see Madsen et al (1983) and Hougaard and Madsen (1983). In this material the survival time is measured in days only, and as 102 patients die the first day, we analyze it as grouped data. The patients were followed up after one year, except for three emigrants, who were last seen at the hospital on day 18. This makes our model multinomial with censored data. The cell probabilities are nonlinear functions of the parameters, the probability of death day t, which covers time $t-1$ until t is

$$P_t = \begin{cases} \{1 + (t-1)/\theta\}^{-\delta} - (1+t/\theta)^{-\delta} & \alpha = 0 \\ \exp[-\delta\{(\theta+t-1)^{\alpha}-\theta^{\alpha}\}/\alpha] - \exp[-\delta\{(\theta+t)^{\alpha}-\theta^{\alpha}\}/\alpha] & 0 < \alpha \leq 1 \end{cases}$$

The likelihood function can be maximized in the usual way; we have iterated by means of Newtons algorithm using the second derivative of the log likelihood function.

Table 1 gives the estimates and corresponding standard errors for the full model and under the hypotheses $\alpha = 0$ and $\theta = 0$. Also the logarithm of the maximized likelihood function is reported. For the model where the daily death probabilities are freely varying this function is -2245.89, such that the likelihood ratio test statistic for goodness-of-fit is 357.5 with 356 degrees of freedom. However, as most of the cells are empty, the χ^2- distribution is not a good approximation. After grouping the survival times, the goodness-of-fit is not as good, with fewer deaths than expected in the period 30 - 120 days after the infarction.

Heart failure is known to be an important complication, implying a poor prognosis. We have tried to analyze the patients with and without this complication separately. For these two groups the results are described in Table 2. For patients without heart failure the estimate is on the boundary, suggesting a stable distribution. Thus the calculation of standard errors is meaningless. A partial explanation of this might be the method of determining the presence of heart failure. The determination is based on an X-ray taken after admission. In some cases the patient has died before the X-ray examination, thus invalidating the use of detected heart failure as a risk factor the first days, because some of the deaths in the group without heart failure are wrongly classified. The estimate is, however, still on the boundary, if we analyze the data conditioning on survival status after a couple of days.

Under the hypothesis $\alpha = 1/2$, the estimate is also on the boundary $\theta = 0$, because the mortality decreases too fast to be explained by an inverse Gaussian frailty. The fit is rather bad, for example for all patients the logarithm of the likelihood function is -2550.08.

For all patients analyzed together and for the patients with heart failure, the hypotheses $\alpha = 0$ and $\theta = 0$ are both rejected, the fit for $\theta = 0$ being worse than for $\alpha = 0$. For the patients without heart failure the fit for $\alpha = 0$ is much worse than for $\theta = 0$.

Measured by the coefficient of variation the estimates imply a heterogeneity of 5.4 for all patients, 3.7 for patients with heart failure and ∞ for patients without. Compared to patients with heart failure the mortality in the other group is practically zero. Splitting the patients thus works as excluding a group of almost zero mortality from the heart failure patients, thus explaining the decrease in coefficient of variation from 5.4 to 3.7. The infinite value for patients without heart failure is possible because this group only has small influence on the total mortality.

It is not possible to conclude from this simple analysis whether one or the other explanation (or both explanations) of the mortality decrease is correct. The aim of this section is only to suggest the explanation based on heterogeneity and to demonstrate that it is practically possible to analyze models, where the survival distributions are explicitly formulated after a consideration of heterogeneity.

Table 1

Estimates (and standard errors)

	α	δ	θ	log L
Full model	0.116(0.038)	0.040(0.006)	0.07(0.05)	-2424.64
$\alpha = 0$		0.0598(0.0037)	0.28(0.06)	-2428.67
$\theta = 0$	0.241(0.013)	0.0254(0.0013)		-2431.67

Table 2

Estimates (and standard errors) for subgroups.

Patients with heart failure

	α	δ	θ	log L
Full model	0.114(0.044)	0.083(0.013)	0.14(0.09)	-1862.07
$\alpha = 0$		0.122(0.008)	0.41(0.09)	-1865.02
$\theta = 0$	0.263(0.015)	0.0481(0.0027)		-1870.28

Patients without heart failure

	α	δ	θ	log L
Full model	0.281	0.0053	0	-400.90
$\alpha = 0$		0.0136(0.0022)	0.28(0.16)	-408.62
$\theta = 0$	0.281(0.046)	0.0053(0.0009)		-400.90

ACKNOWLEDGEMENT

The author is supported by the Danish Natural Science Research Council.

The comments from Søren Johansen are greatly appreciated.

REFERENCES

1. Bondesson, L. (1978). On infinite divisibility of powers of a gamma variable. Scand. Actuarial J., 31 - 40.

2. Bondesson, L. (1981). Classes of infinitely divisible distributions and densities. Z Wahrscheinlichkeitstheorie verw. Gebiete 57, 39 - 71.

3. DuMouchel, W.H. (1983). Estimating the stable index α in order to measure tail thickness: A critique. Ann. Statist. 11, 1019 - 1031.

4. Feller, W. (1971). An Introduction to Probability Theory and Its Applications. Volume II, Second Edition. Wiley.

5. Gawronski, W. (1984). On the bell-shape of stable densities. Ann. Prob. 12, 230 - 242.

6. Heathcote, C.R. (1977). The integrated squared error estimation of parameters. Biometrika 64, 255 - 64.

7. Heckmann, J.J. & Singer, B. (1982). The identification problem in econometric models for duration data. Advances in Econometrics, 39 - 77. (ed. Werner Hildenbrand), Cambridge University Press.

8. Hougaard, P. (1984). Life table methods for heterogenous populations: Distributions describing the heterogeneity, Biometrika 71, 75 - 83.

9. Hougaard, P. and Madsen, E.B. (1983). Dynamic evaluation of short-term prognosis after myocardial infarction. Research Report 83/3, Statistical Research Unit, Copenhagen, Denmark. To appear in Statistics in Medicine 4, 1985.

10. Jewell, N.P. (1982). Mixtures of exponential distributions. Ann.Statist. 10, 479 - 484.

11. Madsen, E.B., Hougaard, P. and Gilpin, E. (1983). Dynamic evaluation of prognosis from time-dependent variables in acute myocardial infarction. The American J. of Cardiology, 51, 1579 - 1583.

12. Morris, C.N. (1982). Natural exponential families with quadratic variance functions. Ann. Statist. 10, 65 - 80.

13. Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Flournoy, N., Farewell, V.T. & Breslow, N.E. (1978). The analysis of failure times in the presence of competing risks. Biometrics 34, 541 - 554.

14. Ruegg, A.F. (1970). A characterization of certain infinitely divisible laws. Ann. Math. Statist. 41, 1354 - 1356.

15. Shanbhag, D.N. and Sreehari, M. (1977). On certain self-decomposable distributions. Z. Wahrscheinlichkeitstheorie verw. Gebiete 38, 217 - 222.

16. Thorin, O. (1977). On the infinite divisibility of the Pareto distribution. Scand. Actuarial J. 31 - 40.

17. Vaupel, J.W., Manton, K.G. & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16, 439 - 454.

18. Williams, E.J.(1977). Some representations of stable random variables as products. Biometrika 64, 167 - 169.

PREPRINTS 1983

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Jacobsen, Martin: Two Operational Characterizations of Cooptional Times.

No. 2 Hald, Anders: Nicholas Bernoulli's Theorem.

No. 3 Jensen, Ernst Lykke and Rootzén, Holger: A Note on De Moivre's Limit Theorems: Easy Proofs.

No. 4 Asmussen, Søren: Conjugate Distributions and Variance Reduction in Ruin Probability Simulation.

No. 5 Rootzén, Holger: Central Limit Theory for Martingales via Random Change of Time.

No. 6 Rootzén, Holger: Extreme Value Theory for Moving Average Processes.

No. 7 Jacobsen, Martin: Birth Times, Death Times and Time Substitutions in Markov Chains.

No. 8 Hougaard, Philip: Convex Functions in Exponential Families.

PREPRINTS 1984

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Rootzén, Holger and Sternby, Jan: Consistency in Least Squares Estimation: A Bayesian Approach.

No. 2 Hougaard, Philip: Parameter Transformations in Multiparameter Nonlinear Regression Models.

No. 3 Jacobsen, Martin: Cooptional Times and Invariant Measures for Transient Markov Chains.

No. 4 Rootzén, Holger: Attainable Rates of Convergence of Maxima.

No. 5 Asmussen, Søren and Thorisson, Hermann: Boundary Problems and Large Deviation Results for Queue Length Processes.

No. 6 Björnsson, Ottó J.: Notes on Right-(Left-)Continuous Functions.

No. 7 Hougaard, Philip: Frailty Models Derived from the Stable Distributions.