The Annals of Applied Probability
1992, Vol. 2, No. 4, 951–972

MODELLING HETEROGENEITY IN SURVIVAL ANALYSIS BY THE COMPOUND POISSON DISTRIBUTION

BY ODD O. AALEN

University of Oslo

When making probabilistic models for survival times, one should consider the fact that individuals are heterogeneous. The observed changes in population intensities (or hazard rates) over time are a mixed result of two influences: on the one hand, the actual changes in the individual hazards, and, on the other hand, the selection due to high-risk individuals leaving the risk group early. I will consider the common multiplicative model for heterogeneity, but with the new feature that the random proportionality factor has a compound Poisson distribution. This distribution is studied in some detail. It is pointed out how its application to the survival situation extends a model of Hougaard, inheriting several nice properties. One important feature of the model is that it yields a subgroup of zero susceptibility, which "survives forever." This is a relevant model in medicine and demography. Two examples are given where the model is fitted to data concerning marriage rates and fertility.

1. Introduction. In order to properly interpret results of survival analysis, one has to consider the fact that individual risks differ in, possibly, unknown ways. This heterogeneity may be difficult to assess, but is nevertheless of great importance. It may distort observed survival curves and intensities (hazard rates). This has been discussed by a number of authors, for instance, Manton, Stallard and Vaupel (1981), Vaupel and Yashin (1985), Hougaard (1984, 1986a, b), Aalen (1988) and Vaupel (1990).

When intensities are estimated, for instance by incidence rates, one may be interested in how they change as a function of time. Quite often they are seen to be rising at the start, reaching a maximum and then declining. Hence the intensity has a unimodal shape with a finite mode (which should be clearly distinguished from the distribution as such being unimodal). This, for instance, is typical of death rates for cancer patients, meaning that the longer the patient lives, beyond a certain time, the more improved are his or her chances. It is also a well-known phenomenon of divorce rates when the time scale is the duration of marriage. The maximal rate of divorce which occurs after a few years is often (falsely) interpreted to mean that (most) marriages are going through a crisis and then improving [Aaberge, Kravdal and Wennemo (1989)]. In the examples analyzed in the present paper, concerning marriage and fertility rates, the same unimodal intensity is observed; see Figures 7 and 8 below.

Received January 1991; revised December 1991.

AMS 1980 *subject classifications*. Primary 60E05; secondary 62M05.

Key words and phrases. Heterogeneity, frailty, survival analysis, compound Poisson distribution.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Applied Probability*. www.jstor.org

It is likely that unimodal intensities are often a result of selection and do not reflect an underlying development on the individual level. The population intensity starts to decline simply because the high-risk individuals have already died or been divorced, and so forth. The intensity of a given individual might well continue to rise.

An additional feature which is often seen is that the total integral under the intensity appears to be finite; that is, the distribution is defective. In practical terms this means that some individuals have zero susceptibility; they will "survive forever." For instance, some patients survive their cancer, some people never marry, some marriages are not prone to be dissolved, and so on. In medicine there are several examples of diseases primarily attacking people with a particular susceptibility, for instance of a genetic kind, other people having virtually zero susceptibility of getting the disease. Another example is fertility. Some couples are unable to conceive children, so that the distribution of times to first child birth for a population of couples will be defective. Analyzing unemployment, one is also faced with the fact that some people may be completely unable to get a job.

The aim of the present paper is to study a simple probabilistic model for heterogeneity which incorporates a nonsusceptible group in a natural fashion. The model is an extension of one studied by Hougaard (1986a, b). The compound Poisson distribution plays a prominent role in this extension, being used here as a mixing distribution. The extension was first suggested by the present author [Aalen (1988)] but was only briefly discussed there. Hence the object of the present paper is to study further the choice of the compound Poisson distribution as a mixing distribution in survival models. Incidentally, the use of the terminology "compound Poisson distribution" follows Feller (1971), who gives some results on this class of distributions. [It should be noted that the term "compound" is used somewhat differently by different authors; see, for instance, Cox and Oakes (1984) for another use.]

The starting point will be the following simple and much applied model for heterogeneity: One considers the time to occurrence of a particular event. The intensity (hazard rate) of an individual is given as the product of an individual specific quantity Z and a basic intensity $\lambda(t)$:

$$\text{Individual intensity} = Z\lambda(t). \tag{1}$$

Clearly, Z may be considered as a random variable over the population of individuals. It will be called a mixing variable and it is the distribution of Z which is termed the mixing distribution. What may be observed in a population is not the individual intensity, but the net result for a number of individuals with differing values of Z.

The use of the compound Poisson distribution for Z is not only mathematically convenient, but might also be seen as natural in a more substantial sense. The distribution arises as a sum of a random number of independent gamma variables, where the number of terms in the sum is Poisson distributed. This might be viewed as a kind of shock model, where the vulnerability of the subject has been shaped by a random number of shocks, each of random size.

[A related point of view is presented by Becker and Rittgen (1990).] This is not unreasonable as a formal model, but one should obviously not interpret it too literally in any given practical case. After all, the statistical model fitting presented in the examples below is a quite superficial undertaking.

It should be mentioned that there is a certain connection with the so-called "mover–stayer" models [see, e.g., Sampson (1990)]. The model of the present paper may perhaps also be termed a mover–stayer model (the stayers being those with $Z = 0$), but the mathematical form is more natural and leads to a simpler form of the population intensity than previous mover–stayer models. These have mainly assumed that one part of the population has zero susceptibility, while the other part has a fixed positive susceptibility.

Below, the extension of Hougaard's model will be described and some properties of the compound Poisson distribution will be studied and illustrated graphically. Several aspects of model (1) with this mixing distribution will be studied. Finally, two practical examples will be analyzed, one concerning marriage incidence and the other concerning fertility.

2. Extending Hougaard's model.

2.1. *The Laplace transform of the mixing variable.* For model (1) the population survival function is given as follows:

$$S(t) = E\{e^{-Z\Lambda(t)}\},$$

where $\Lambda(t) = \int_0^t \lambda(u)\,du$. Let $L(s)$ denote the Laplace transform of Z. The above formula may be rewritten as

$$S(t) = L(\Lambda(t)). \tag{2}$$

This connection with the Laplace transform was pointed out by Philip Hougaard and has been very efficiently exploited by him in the papers mentioned above. It follows that, when seeking distributions for the mixing variable Z, it is natural to try those which have explicit Laplace transforms. I will here consider the following extension of a parametric family of Laplace transforms suggested by Hougaard (1986a):

$$L(s) = \exp\left\{\frac{\alpha}{(1-\alpha)\delta}\left[1 - \left(1 + \frac{\delta\gamma}{\alpha}s\right)^{1-\alpha}\right]\right\}, \qquad \alpha, \delta \geq 0, \gamma > 0. \tag{3}$$

For certain values of the parameters, the analytic expression above is not immediately well defined, but should be defined by continuity. If α or δ equals 0, the distribution of Z is degenerate at γ. These cases correspond to no heterogeneity being present. The remaining case to be defined by continuity is the following important one:

$$L(s) = \left\{\frac{1}{1 + \delta\gamma s}\right\}^{1/\delta} \quad \text{when } \alpha = 1,$$

which is the Laplace transform of a gamma distribution.

The parameters of the mixing distribution defined by (3) are interpreted as follows: First, γ is the expectation of Z and is seen to be a simple scale parameter. Next, δ is the squared coefficient of variation, this being a natural parameter since the spread of the distribution of Z is important in determining the degree of heterogeneity. Finally, α is a parameter dividing the class of distributions in two major subcategories: For $\alpha < 1$ the family of distributions is the one discussed in detail by Hougaard (1986a), who derived it from the stable distributions. Note that a different parameterization is used here, in the belief that these parameters have a more natural interpretation in the context of survival analysis. The extension to $\alpha > 1$ was suggested by Aalen (1988) and shown to yield compound Poisson distributions (generated by gamma variables). The two subcategories are separated by the family of gamma distributions ($\alpha = 1$). The general class of distributions considered here has also been studied by Bar-Lev and Enis (1986) and Jørgensen (1987), although in an entirely different context.

2.2. *Compound Poisson mixing distribution.* In this section it will be explained how the subfamily $\alpha > 1$ of (3) arises as a class of compound Poisson distributions generated by gamma variables. Such a distribution may be written as follows [Feller (1971)]:

$$Z = \begin{cases} X_1 + X_2 + \cdots + X_N, & \text{if } N > 0, \\ 0, & \text{if } N = 0, \end{cases} \tag{4}$$

where N is Poisson distributed with expectation ρ, while $X_1, X_2, \ldots$ are independent and gamma distributed with scale parameter ν and shape parameter η. The Laplace transforms of the gamma and Poisson distributions are given by $L_X(s) = \{\nu/(\nu + s)\}^\eta$ and $L_N(s) = \exp(-\rho + \rho e^{-s})$, respectively. The following standard derivation can now be applied:

$$\begin{aligned} L(s) &= E\{e^{-sZ}\} = E\{e^{-s(X_1 + \cdots + X_N)}\} \\ &= E\{L_X(s)^N\} = L_N(-\ln(L_X(s))). \end{aligned} \tag{5}$$

Inserting the previous expressions gives the following Laplace transform of Z:

$$L(s) = \exp\left\{-\rho + \rho\left(\frac{\nu}{\nu + s}\right)^\eta\right\}. \tag{6}$$

This is the same Laplace transform as (3), but with a different parameterization. The connection between the two parameter sets is derived as follows. By differentiation of (6) one finds the first and second moments. Recalling that γ is the expectation, one writes $\gamma = EZ = \rho\eta/\nu$. Further, δ is the squared coefficient of variation; hence $\delta = \operatorname{Var} Z/EZ^2 = (\eta + 1)/(\eta\rho)$. Finally, one sees that $\alpha = \eta + 1$. Solving the equations gives

$$\rho = \frac{\alpha}{\delta(\alpha - 1)}, \qquad \nu = \frac{\alpha}{\delta\gamma}, \qquad \eta = \alpha - 1.$$

Inserting this into the Laplace transform (6) brings it into the form (3), which will be used in the rest of the paper.

From the definition (4) one may immediately deduce the following. The distribution consists of two parts: a positive probability of being equal to 0 and a continuous density on the positive real line. The discrete part is

$$P(Z = 0) = \exp\{-\rho\} = \exp\left\{-\frac{\alpha}{\delta(\alpha - 1)}\right\}, \tag{7}$$

which increases with α and δ. Conditioning with respect to N and using the fact that the sum of the X's is gamma distributed, the density of the continuous part may be written immediately and put in the following form:

$$f(z;\alpha,\delta,\gamma) = \exp\left\{-\frac{\alpha}{\delta}\left(\frac{z}{\gamma} + \frac{1}{\alpha - 1}\right)\right\} \times \frac{1}{z}\sum_{k=1}^{\infty} \frac{(\alpha/\delta)^{k\alpha}(z/\gamma)^{k(\alpha-1)}}{k!\Gamma(k(\alpha - 1))(\alpha - 1)^k}. \tag{8}$$

The positive probability at $z = 0$ corresponds to an assumption of nonsusceptibility for a part of the population.

It may be of interest to consider separately the continuous part of the distribution. The expectation is given as

$$E(Z|Z > 0) = \frac{\gamma}{P(Z > 0)} = \gamma\Big/\left(1 - \exp\left\{-\frac{\alpha}{\delta(\alpha - 1)}\right\}\right).$$

Similarly, the squared coefficient of variation for the continuous part may be derived as

$$(\delta + 1)P(Z > 0) - 1 = (\delta + 1)\left(1 - \exp\left\{-\frac{\alpha}{\delta(\alpha - 1)}\right\}\right) - 1.$$

As might be expected, this quantity is always smaller than δ, the squared coefficient of variation for the whole distribution.

2.3. *Connection between the classes of distributions.* The subclass $\alpha < 1$ of (3) is the family suggested by Hougaard (1986a), which was proved by him to consist of absolutely continuous nonnegative distributions with unimodal densities. Hence the atom at $z = 0$ disappears when α goes below 1. Hougaard gives an expression for the density valid for $\alpha < 1$ and an interesting question is how this is related to the density (8) of the absolutely continuous part of the compound Poisson distributions. Considering (8) for $\alpha < 1$, one meets with the difficulty that the argument of the gamma function becomes negative. To circumvent this, one may use the following standard formula [see, e.g.,

Abramowitz and Stegun (1972), page 256]:

$$\Gamma(y)\Gamma(1-y) = \frac{\pi}{\sin(\pi y)}.$$

(This can be considered formally true even at integer y's where both sides equal ∞.) Applying this yields

$$\frac{1}{\Gamma(k(\alpha-1))} = \frac{1}{\pi}\Gamma(1-k(\alpha-1))\sin(\pi k(\alpha-1)).$$

Inserting this into (8) yields

$$f(z;\alpha,\delta,\gamma) = \exp\left\{-\frac{\alpha}{\delta}\left(\frac{z}{\gamma}+\frac{1}{\alpha-1}\right)\right\} \times \frac{1}{\pi z}\sum_{k=1}^{\infty}\frac{(\alpha/\delta)^{k\alpha}(z/\gamma)^{k(\alpha-1)}\Gamma(1-k(\alpha-1))\sin(\pi k(\alpha-1))}{k!(\alpha-1)^k}.$$

By reparameterization this is the same as the density given by Hougaard (1986a). Hougaard's parameters, denoted with subscript H, are given as follows:

$$\alpha_H = 1-\alpha, \qquad \delta_H = \gamma^{1-\alpha}(\alpha/\delta)^{\alpha}, \qquad \theta_H = \frac{\alpha}{\delta\gamma}.$$

Hence the density given by (8) is valid for all $\alpha > 0$ when the gamma function is extended to negative arguments. Its integral over the positive half-line equals 1 when $\alpha \le 1$, and the integral is less than 1 when $\alpha > 1$ with the rest of the probability, given by (7), being placed at 0.

2.4. *Asymptotic limits.* A number of asymptotic results exist for the present distributions, and the most important ones will be mentioned briefly.

ASYMPTOTIC NORMALITY. Consider the standardized variable

$$Y = \frac{Z-\gamma}{\gamma\sqrt{\delta}},$$

which has expectation 0 and standard deviation 1. By taking the limit of the Laplace transform, it is deduced that the distribution of Y converges to a standard normal distribution when $\delta \downarrow 0$; see also Jørgensen (1987).

ASYMPTOTIC POISSON DISTRIBUTION. When $\alpha \to \infty$ the distribution of $Z/(\delta\gamma)$ converges to a Poisson distribution with expectation $1/\delta$. This follows by applying a limit argument to the Laplace transform.

ASYMPTOTIC STABLE DISTRIBUTION. For $\alpha < 1$ the class of stable distributions appears as a limit when δ and γ goes jointly to ∞ in an appropriate fashion; see Hougaard (1986a) for details.

3. Shape of the compound Poisson distributions. When using the compound Poisson distribution, it is important to be acquainted with its properties, for instance the shape of the density of the absolutely continuous part. There does not appear to be much information on this in the literature, and therefore some discussion and several illustrative figures are given here.

Of the three parameters, γ is merely a scale parameter, while the other two, α and δ, determine the shape of the distribution. When $\alpha \leq 1$ it has been proven by Hougaard (1986a) that the distributions of Z are unimodal. For $\alpha > 1$ the positive probability at 0, combined with an absolutely continuous density, implies that the distribution is not unimodal [see also Bar-Lev and

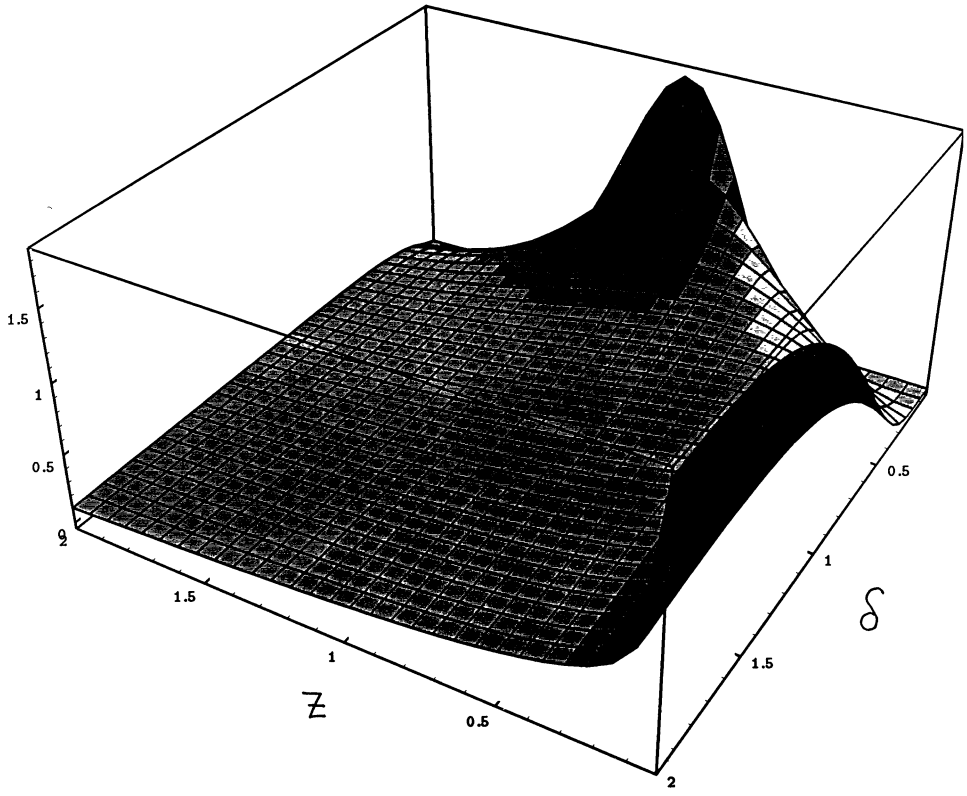


FIG. 1. *Densities of the continuous part of the compound Poisson distribution for* $\alpha = 1.5$. *By varying the parameter* δ *the family of densities is shown as a surface, where a particular density is obtained by cutting the surface parallel to the axis marked z. In this case the densities appear unimodal; but see the text for qualifications.* [*Technically, the figure shows* $f(z; 1.5, \delta, 1)$ *for* $0.01 \leq z \leq 2$ *and* $0.05 \leq \delta \leq 2$ *with grid size* 0.069×0.067.]

Enis (1986)]. One may, however, ask whether the density part $f(z; \alpha, \delta, \gamma)$ is unimodal. This would seem a desirable feature of the model, unless groups with clearly distinct susceptibilities were involved. An impression of the shapes of the distributions is most easily given in a graphical manner. Since $f(z; \alpha, \delta, \gamma)$, given by (8), is not defined at $z = 0$, I will first discuss the limit of this expression, for $\alpha > 1$, when z approaches 0.

The limit of the density part (8) when z approaches 0 depends on the value of α. The first factor of $f(z; \alpha, \delta, \gamma)$ (the exponential function) converges to a finite positive limit, and so poses no problem. The rest of the right-hand side of (8) can be written as follows:

$$\sum_{k=1}^{\infty} \frac{(\alpha/\delta)^{k\alpha} z^{k(\alpha-1)-1}}{k!\Gamma(k(\alpha-1))(\alpha-1)^k \gamma^{k(\alpha-1)}}. \tag{9}$$

Depending on the value of α, the exponent of z in the above sum will be positive from a certain k on. The sum of terms from this k will converge to 0 when $z \downarrow 0$. This follows from the Lebesgue convergence theorem due to the

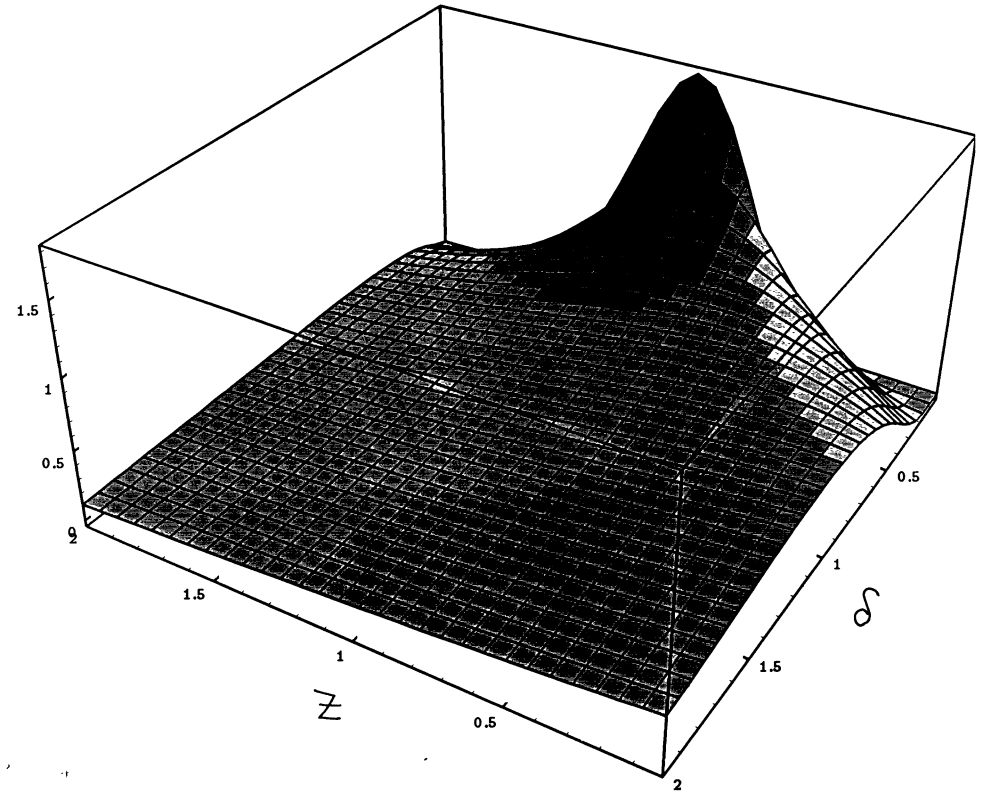


FIG. 2. *Densities of the continuous part of the compound Poisson distribution for* $\alpha = 2$; *see Figure* 1 *for details.*

fact that (9) converges for all positive z, and contains only nonnegative summands each decreasing in z from k on. Three cases must be distinguished:

1. *The case* $1 < \alpha < 2$. The first few terms of (9) will have negative powers of z, with the dominant term being proportional to $z^{\alpha-2}$. Hence $f(z; \alpha, \delta, \gamma)$ goes to ∞ when z goes to 0.
2. *The case* $\alpha = 2$. All terms of (9) go to 0, except the first which determines the limit. Hence the limit of $f(z; \alpha, \delta, \gamma)$ when z goes to 0 equals $(2/\delta)^2 \exp(-2/\delta)/\gamma$.
3. *The case* $\alpha > 2$. All terms of (9) contain positive powers of z and hence $f(z; \alpha, \delta, \gamma)$ goes to 0 when z goes to 0.

A number of figures have been made to illustrate the possible shapes of the density part. Without any loss of generality the expectation γ is put equal to 1. For some values of α, three-dimensional plots of (8) are then presented with z and δ as independent variables; that is, $f(z, \alpha_0, \delta, 1)$ is plotted as a function in two variables for specific values of α_0. (The plots are made by the Mathematica program [Wolfram (1988)].) The densities for various values of δ will be the

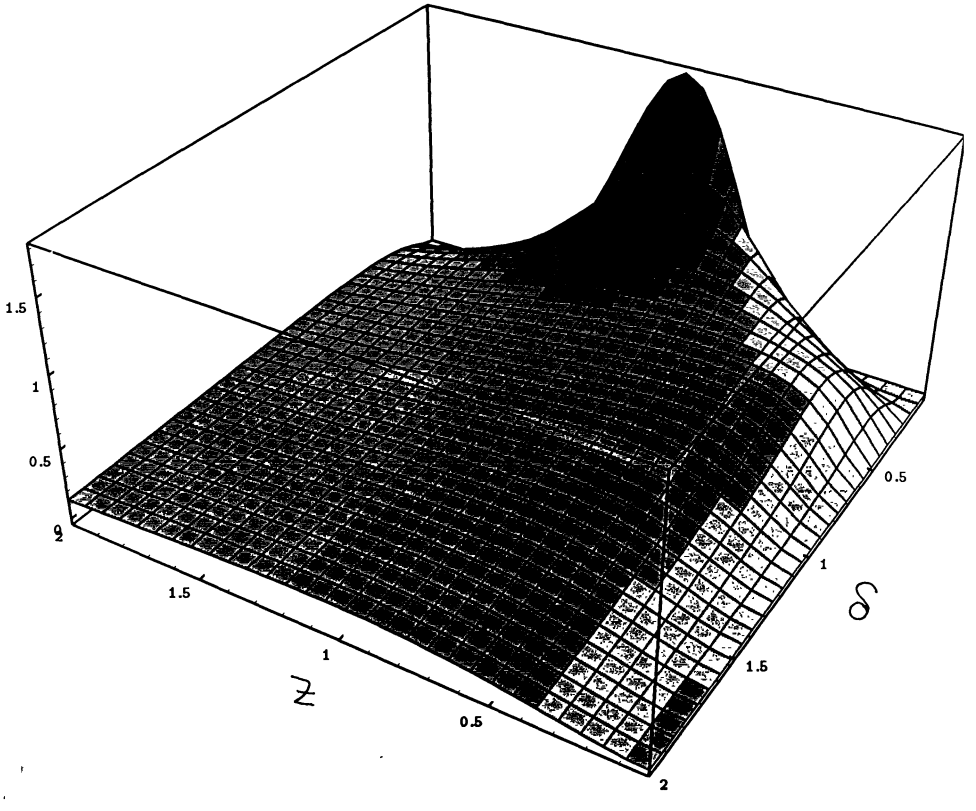


FIG. 3. *Densities of the continuous part of the compound Poisson distribution for* $\alpha = 4$; *see Figure* 1 *for details.*

cuts through the resulting surface parallel to the z-axis. The advantage of this way of plotting is that it gives a good impression of how the densities vary with the parameter δ. The three categories enumerated above will be discussed separately.

For the case $1 < \alpha < 2$, I have chosen $\alpha_0 = 1.5$ as a representative value. The corresponding plot is presented in Figure 1. Considering cuts through the surface parallel to the axis marked z, one sees that the densities appear to be unimodal, decreasing monotonically from a mode at 0 when δ is large, and apparently having a mode larger than 0 when δ is small. From considerations above one knows that the value at $z = 0$ is in all cases ∞. Hence the impression of unimodality cannot be correct for small values of δ. One-dimensional plots with finer detail show that the densities are "nearly" unimodal also for small δ, except that, when z approaches 0, the density curve reaches a minimum and then goes off towards ∞. This latter part, however, generally constitutes a very small part of the density when considering its integral. Hence, for practical purposes, the densities can be considered to be nearly unimodal even for small

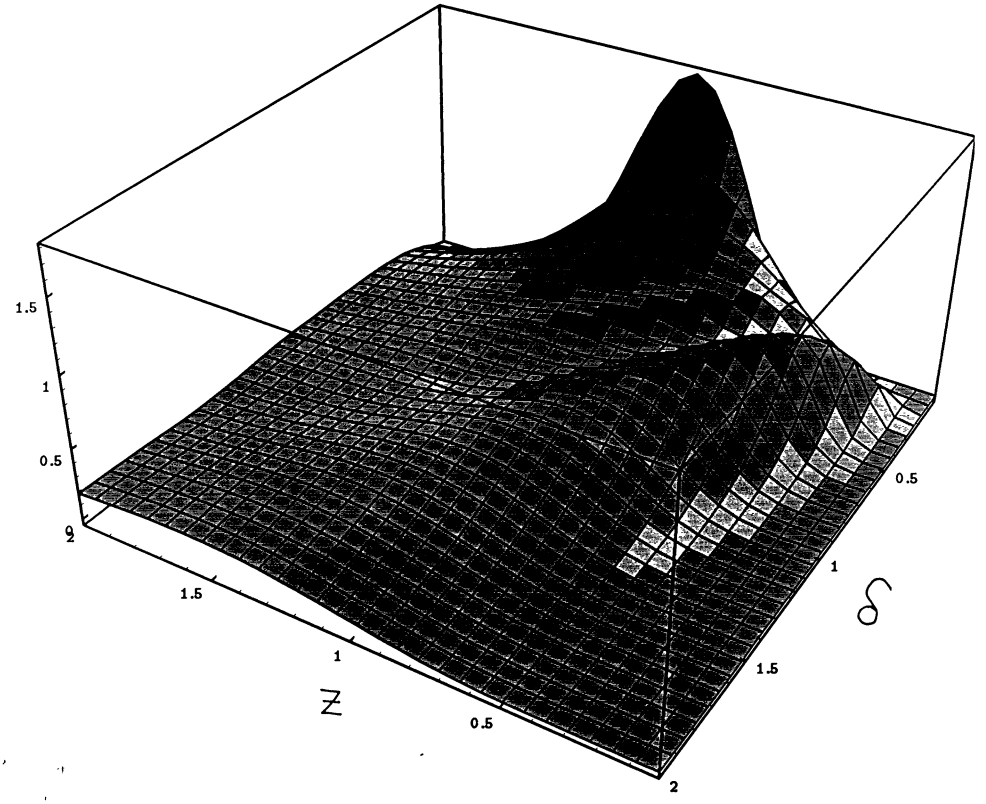


FIG. 4. *Densities of the continuous part of the compound Poisson distribution for* $\alpha = 10$; *see Figure* 1 *for details.*

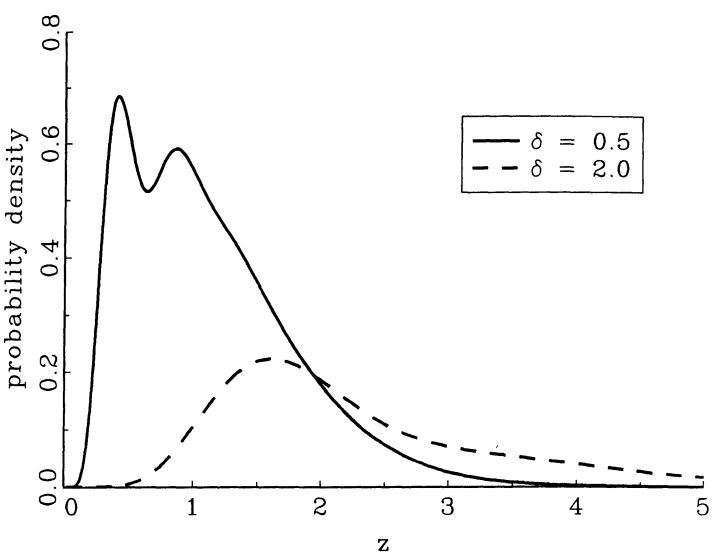


FIG. 5. *Density of the continuous part of the compound Poisson distribution for* $\alpha = 10$, $\gamma = 1$ *and two values of* δ.

values of δ. This seems to hold for most α between 1 and 2. In a sense the divergence to ∞ when z approaches 0 can be viewed as a remainder of the atom at 0.

The case $\alpha = 2$ is illustrated in Figure 2. Apparently the densities are unimodal, monotonically decreasing from a finite value at 0 when δ is large, but having a positive mode for small δ.

The case $\alpha > 2$ is more complicated. Three-dimensional plots are shown for α_0 equals 4 and 10; see Figures 3 and 4. In the first case the distributions appear to be unimodal with a positive mode. In the second case there is a "valley" in the figure, implying cases with two modes. This is also illustrated in Figure 5. Drawing plots for larger values of α reveals several modes arising. This is not surprising since the distributions converge towards Poisson distributions when α increases. Hence one would expect multiple modes to arise, eventually converging into the discrete atoms of the Poisson distribution.

What is the practical implication of these multiple modes when the heterogeneity model is applied? Preliminary experience with statistical fitting of the heterogeneity model have given values of α between 1 and 5, for which the densities would be unimodal, or nearly so. So large values of α are perhaps not very common in practice. But if they do occur, then the resultant multiple modes would mean that the population should consist of several subgroups with quite distinct risk levels. This may occasionally be true, but it should be documented by other information too, and not only by the fit of a model.

From all three-dimensional figures it is apparent how the normal distribution seems to arise for values of δ very close to 0. This is in accordance with the asymptotic theory mentioned earlier.

4. Application to heterogeneity in survival analysis.

4.1. *Population survival function and intensity.* When used in conjunction with model (1), the particular class of mixing distributions considered here produces nice and useful formulas. Combining (2) and (3) gives the following population survival function:

$$(10)\quad S(t) = \exp\left\{ \frac{\alpha}{(1-\alpha)\delta}\left[1 - \left(1 + \frac{\delta\gamma}{\alpha}\Lambda(t)\right)^{1-\alpha}\right]\right\} \quad \text{if } \alpha \neq 1,\ \alpha > 0,$$

$$(11)\quad S(t) = \left\{ \frac{1}{1 + \delta\gamma\Lambda(t)} \right\}^{1/\delta} \quad \text{if } \alpha = 1,$$

with the corresponding population intensity

$$(12)\qquad \mu(t) = -\frac{d}{dt}\ln(S(t)) = \frac{\gamma\lambda(t)}{\{1 + \alpha^{-1}\delta\gamma\Lambda(t)\}^{\alpha}} \quad \text{for } \alpha > 0.$$

Note the limiting case $\alpha = 0$, yielding $\mu(t) = \gamma\lambda(t)$, which should be expected when no heterogeneity is present. One reason for being interested in the particular class of mixing distributions considered here is the simple form of (12).

It should be noted that $\mu(t)$ has a finite integral over $(0, \infty)$ when $\alpha > 1$. This means that the survival distribution is defective, corresponding to the fact that some individuals have zero intensity of "dying." These individuals are, of course, precisely those with $Z = 0$, and it is the positive probability of this event [see formula (7)] that produces the "infinite" lifetimes. In fact, this kind of situation has previously been modelled by means of the so-called "mover–stayer" model [see, e.g., Sampson (1990)]. The present model is mathematically more attractive, and has the advantage that it combines the zero susceptibility for some individuals with a continuous variation in susceptibility for other individuals.

Considering more generally the methodology of survival analysis, it is interesting that (12) for $\alpha > 1$ produces a large class of intensities corresponding to defective survival distributions. The ordinary parametric models of survival analysis do not have this property. Such defective distributions are of relevance in many contexts; see Section 1 for some further discussion. The finite integral of the population intensity for $\alpha > 1$ means that eventually the population intensity $\mu(t)$ must decline to 0 however much $\lambda(t)$ increases. This is reasonable since eventually the survivors will be dominated by those of zero susceptibility.

In order to inspect more closely what effect mixing may have on the intensity, one may consider the case of an increasing Weibull basic intensity, that is, $\lambda(t) = at^k$ with $k > 0$. It is easily seen that the population intensity $\mu(t)$ will eventually decrease towards 0 whenever $\alpha > k/(k+1)$, continue to increase indefinitely when α is less than this value and approach a finite limit in the case of equality. In the first case the population intensity will increase

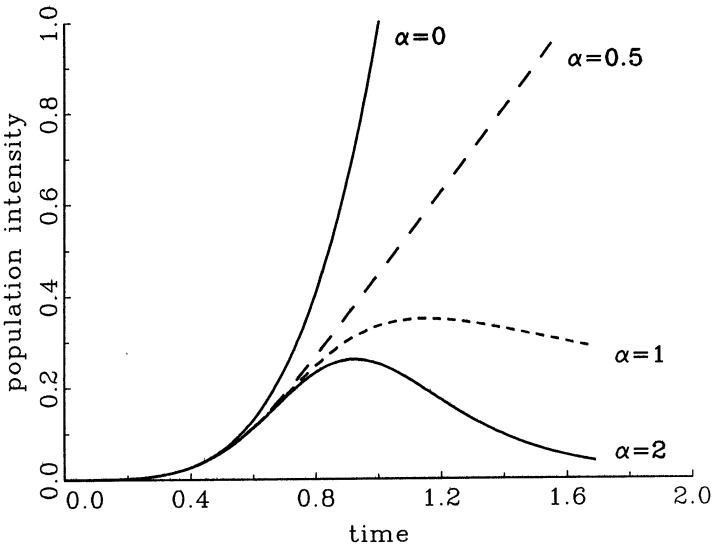


FIG. 6. *The effect of heterogeneity on the intensity. Population intensities, $\mu(t)$, are shown for $\delta = 10$, $\gamma = 1$ and various values of α, with basic intensity $\lambda(t) = t^4$. The case $\alpha = 0$ corresponds to no heterogeneity.*

up to a maximal point before decreasing; hence this is a unimodal intensity—see Section 1. An illustration of some population intensities for various values of α and a Weibull basic intensity with $k = 4$ is given in Figure 6. In this figure the curve for $\alpha = 0$ corresponds to no heterogeneity; in the other curves it is apparent how heterogeneity "depresses" the intensity, with α equal to 1 and 2 corresponding to unimodal shapes.

The occurrence of basic intensities of the Weibull type may sometimes be justified from its property of being an extreme value distribution [Leadbetter, Lindgren and Rootzen (1983)]. It is, for instance, well known how the extreme value character of the Weibull distribution makes it reasonable as a model for the basic intensity of cancer incidence. This is related to the so-called Armitage–Doll model; see also Aalen (1988).

Another important particular case is that of a Gompertz basic intensity, that is, $\lambda(t) = ae^{bt}$. It may be seen that $\mu(t)$ eventually decreases to 0 whenever $\alpha > 1$, approaches a constant when $\alpha = 1$ and increases indefinitely when $\alpha < 1$.

4.2. *Results for surviving individuals.* Consider the individuals surviving at a time t. Due to selection their mixing distribution will have changed compared to what it was at time 0, since the individuals with the smallest Z's will have had the greatest chances of surviving. An important advantage of Hougaard's family is that the mixing distribution of survivors will still belong to the same class, but with new parameters. Since the class considered here is

just a parametric extension of Hougaard's family, the same might be expected to hold for it. Let T be a random variable with distribution given by the survival function $S(t)$. The Laplace transform of the mixing distribution for survivors at time t is given by

$$L_t(s) = E(e^{-sZ}|T > t) = \frac{L(s + \Lambda(t))}{L(\Lambda(t))}$$

[Hougaard (1984)]. Substituting (3) and rearranging yields the Laplace transform $L(s)$ anew, but with changed parameters. The new parameters are as follows:

$$(13) \quad \alpha_t = \alpha, \qquad \gamma_t = \gamma\left(1 + \frac{\delta\gamma}{\alpha}\Lambda(t)\right)^{-\alpha}, \qquad \delta_t = \delta\left(1 + \frac{\delta\gamma}{\alpha}\Lambda(t)\right)^{\alpha-1}.$$

The fact that α_t is independent of t means that the mixing distributions stay within the particular subclass defined by this parameter when "mortality" selection acts. In particular, if the mixing distribution is compound Poisson at the start, it remains so throughout. One may, furthermore, note that the expectation γ_t decreases with t, as should be expected. On the other hand, the squared coefficient of variation decreases for Hougaard's family ($\alpha < 1$), is constant for the gamma mixing distributions, but increases for the compound Poisson mixing distributions. Finally, it may be deduced from (13) that $\gamma_t^{1-1/\alpha}\delta_t$ is in all cases a constant independent of t.

For $\alpha > 1$, the probability of zero susceptibility for the survivors at time t is given by

$$(14) \qquad P(Z = 0|T > t) = \exp\left\{-\frac{\alpha}{(\alpha - 1)\delta}\left(1 + \frac{\delta\gamma}{\alpha}\Lambda(t)\right)^{1-\alpha}\right\},$$

which, as expected, increases with t.

An interesting limiting case may be mentioned: Let $\alpha > 1$ and consider the susceptible individuals (that is, $Z > 0$) only. It may be interesting to consider the properties of those susceptible individuals who survive until a large time t. In other words, one wants to find the limiting distribution of the mixing variable Z for such individuals. The answer, which is easily derived from the Laplace transform, may be formally given as follows: Let t increase towards ∞ and assume that $\Lambda(t)$ then also goes to ∞. In that case the limiting distribution of $Z\Lambda(t)$ among the susceptible survivors at time t turns out to be a gamma distribution with expectation and variance both equal to $\alpha - 1$. Hence, when excluding those of zero risk, the eventual behaviour of the compound Poisson model is the same as that of a gamma model.

4.3. *Multivariate modelling.* When individuals in a study belong to families or other groups where there may be similarities in risk, then this association can be modelled within the present framework. For simplicity, it will be assumed that a group consists of only two individuals, A and B, and that conditional on a group-specific quantity Z the lifetimes of the individuals are

independent with intensities given as

$$Z\lambda_A(t) \quad \text{and} \quad Z\lambda_B(t),$$

respectively. Note that individuals from the same group have the same value of the mixing variable. Let T_A and T_B be the survival times of A and B, respectively. The joint population survival function is given as

$$\begin{aligned} P(T_A > t_A, T_B > t_B) &= E\big[P(T_A > t_A, T_B > t_B|Z)\big] \\ &= E\big[\exp\big(-Z(\Lambda_A(t_A) + \Lambda_B(t_B))\big)\big]. \end{aligned}$$

Introducing the Laplace transform yields

$$P(T_A > t_A, T_B > t_B) = L(\Lambda_A(t_A) + \Lambda_B(t_B)),$$

which is the counterpart of (2). Once again, this elegant expression for the survival function is one of the main attractive aspects of the model.

This model has been studied by several authors, for instance, Clayton (1978), Hougaard (1986b) and Oakes (1989), with various mixing distributions. The object here is to point out that the compound Poisson mixing distribution is another such candidate. Its use implies that if one member of a group is nonsusceptible, then the other member is too.

5. More general compound Poisson distributions. So far it has been supposed that the compound Poisson distribution is generated as a random sum of gamma variables. Of course, the X-variables of (4) may also have other nonnegative distributions. Applying (5) with a general Laplace transform $L_X(s)$ gives

$$L(s) = \exp\big(-\rho + \rho L_X(s)\big).$$

Combining with (2) produces the population survival function

$$S(t) = \exp\big(-\rho + \rho L_X(\Lambda(t))\big)$$

and the corresponding population intensity

$$\mu(t) = -\rho\lambda(t)L'_X(\Lambda(t)).$$

There exist a number of nonnegative distributions with explicit Laplace transforms that could be candidates for $L_X(s)$; see, for instance, Feller (1971), Chapter 13. One example is the stable distributions with Laplace transform $\exp(-s^\beta)$ for β between 0 and 1. These distributions do not have finite expectations, however, and this carries over to $L(s)$. Another example from Feller's book (Section 13.3) is the Laplace transform

$$\left[s + 1 - \sqrt{(s+1)^2 - 1}\right]^r$$

valid for all $r > 0$. This corresponds to a probability density involving Bessel functions.

It seems that the case of gamma-distributed X-variables, as used in this paper, would be the simplest one. Whether other possibilities may also be of practical importance must be judged on the basis of further research.

6. Example: Modelling incidence of marriage. Marriage is an example of an event that does not happen, eventually, to everybody. A certain percentage never marry and models of marriage incidence must be able to account for this. Borgan and Ramlau-Hansen (1985) presented rates of first marriage for women born in Denmark in 1940, and I will use this as an example.

It will be assumed that the basic intensity is of Weibull type, with $\lambda(t) = a(t - 15)^k$ for $t \geq 15$, where t is age measured in years. The model implies that all women have an increasing intensity of getting married from age 15 on. The fact that the observed incidence peaks around age 23 and becomes rather low after age 30 (see Figure 7) is therefore interpreted as a selection phenomenon due to heterogeneity, meaning that those most prone to marriage will marry quite early, and that those who remain will have less tendency to marry. The aim is to describe this variation by means of model (1), with Z having a compound Poisson distribution.

It is obvious that the model is a rather primitive one. It is not likely that the individual intensity increases in the same manner for all women, and certainly this increase should stop at a certain age. The latter feature could, of course, be incorporated, but it is not done here, for the sake of simplicity, and,

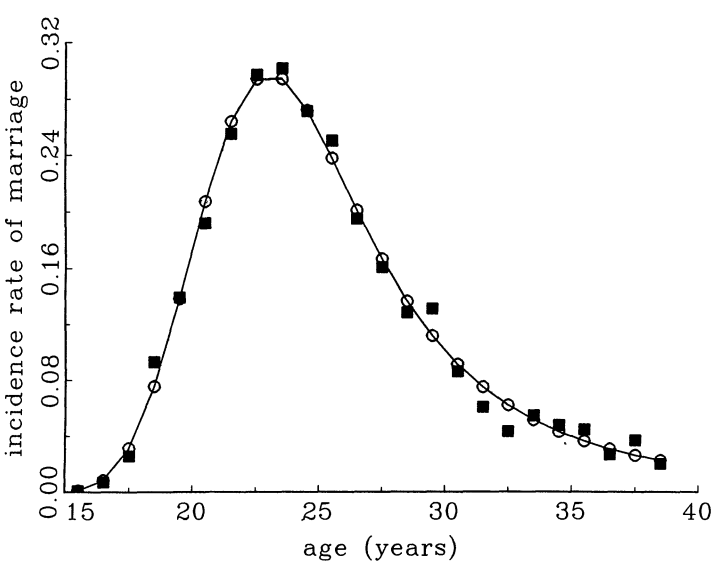


FIG. 7. *Incidence rate of marriage per year. Solid squares are observed rates; open circles are those estimated by the model.*

anyway, the results are only going to be used for ages up to 39 years. With all the simplification, though, it will turn out that the model fits quite nicely to the data, and so one has a parametric model that describes the situation well, and which may be useful for practical purposes. One important feature of the model is that it naturally incorporates the defective character of the distribution of time to marriage, due to those who never marry. I also believe the procedure of including heterogeneity in the modelling process gives insight, even though one should certainly not trust the model too much. Of course, a good fit does not prove its validity; clearly a number of quite different models would have fitted quite as well [Aalen (1988) and Heckman and Walker (1990)].

In spite of the above objections, the fact that a simple heterogeneity model fits well to the data, as it will turn out to do, shows that heterogeneity is certainly a phenomenon that must be taken seriously in the interpretation of marriage rates. For instance, there is no reason to assume that the falling rates after age 23 are mainly due to individual women meeting with greater difficulties in finding a marriageable partner. Certainly, this will eventually be true, but it is compounded with a, probably strong, heterogeneity effect.

One further practical purpose should be mentioned in discussing the model. In demography a lot of attention is being paid to graduation of incidence rates [Hoem (1976)]. The parametric models studied here should be quite useful for analytic graduation, yielding simple, and often well-fitting formulas. These may be of use, for instance, for comparing different populations.

Having discussed the model and its potential usefulness, I will now turn to the estimation of the parameters for the mentioned population of Danish women born in 1940. The aim is to fit our model to the occurrence/exposure rates estimated by Borgan and Ramlau-Hansen (1985). The raw data are not available in their paper, but I have been supplied with these from one of the authors [Borgan (personal communication)]. The risk set at age 15 is 32,534 women.

From the data one can estimate, for each age i, the number R_i at risk at the beginning of the one-year age interval. This corresponds to the number of unmarried ones adjusted for censoring due to death. One also knows the number of marriages, N_i, during the interval. The likelihood function is given as follows:

$$L(\delta, \alpha, a, k) = \prod_{i=16}^{i=39} \{S(i)/S(i-1)\}^{R_i - N_i} \{1 - (S(i)/S(i-1))\}^{N_i},$$

where $S(t)$ is given by (10), with

$$\Lambda(t) = \int_{15}^{t} a(u-15)^k \, du = \frac{a}{k+1}(t-15)^{k+1}.$$

It should be noted that the model is overparameterized; it is not possible to identify γ separately from a. Hence in the estimation γ is put equal to 1, thus centering the mixing distribution at 1. The maximum likelihood estimates for the Danish women are presented in Table 1.

TABLE 1
Maximum likelihood estimates for the marriage data

Parameters	δ	α	a	k
Estimates	0.78	1.72	2.38×10^{-3}	2.79
s.e.	0.02	0.04	0.11×10^{-3}	0.03

The fit of the model is judged by comparing the occurrence/exposure rates (or observed incidence rates), as computed by Borgan and Ramlau-Hansen (1985), with expected incidence rates based on the model. (Note that the term "incidence rate" is used here in a somewhat different sense than by Borgan and Ramlau-Hansen.) The observed rates are computed as N_i/R_i, while the expected rates are computed as $-\ln(S(i)/S(i-1))$, where $S(t)$ is defined as above with estimates substituted for the parameters. The result is shown in Figure 7; it is seen that the overall fit is good. Nevertheless, there is a clearly significant deviation when measured by chi-square, this being not so surprising in view of the large number of individuals in the data set.

The estimated value of α is 1.72 with a very small standard error. Hence one is clearly within the compound Poisson domain of the model. The probability of never marrying may be estimated from (7) by inserting estimates for the parameters, the result being 4.7%. This may be compared to the probability of not being married before age 39, which is the upper age limit of the data set. Computing this either from the occurrence/exposure rates or from the estimated model yields in both cases the same result, namely 5.7%.

One might consider the group composed of individuals who are unmarried at their twenty-fifth birthday, say, and ask for their mixing distribution. From the theory it follows that this is still a compound Poisson distribution with the parameter α unchanged and with the other parameters given by (13). Inserting estimates gives $\gamma_{25} = 0.17$ and $\delta_{25} = 1.62$, showing that the expectation of Z has decreased considerably from its original value of 1, while the squared coefficient of variation has doubled. Furthermore, the conditional probability of never marrying (that is, $Z = 0$) given that the woman has turned 25 is estimated as 23.0%.

7. Example: Modelling fertility.

7.1. *Calculations of fertility.* It is well known that between 5% and 10% of all couples are unable to conceive children. If couples have tried unsuccessfully to conceive for, say, a year, they may start to get worried about whether they are infertile, and it might be of interest to compute the conditional probability of their being able to conceive a child in, say, one more year. The model of the present paper, with $\alpha > 1$ might be useful in this context, since it contains a flexible range of mixing distributions. Of course, $P(Z = 0)$ will be the probability of infertility, and the variation of Z over the positive line expresses varying

fecundabilities among the fertile ones (by fecundability is meant the ability to conceive).

The formulas of Section 4.2 are of relevance here. Equation (14) gives the probability of infertility for those who have not conceived up to a certain time. The probability that they will not conceive in a further period of given length follows from (10) with the new parameters of (13) inserted. To get parameter estimates, one would need to have data on times required to conceive for couples who have decided to have a child and who do not use any contraception. Needless to say, these kinds of data, which shall be valid for a given modern society, may be hard to come by, although data for certain special populations exist in the literature [Heckman and Walker, (1990) and Tietze (1968)]. Due to the limited availability of such data, a data set from a somewhat different context shall be used for illustration in the next section.

7.2. *Analyzing fertility after stillbirth.* After a stillbirth it appears that most women will tend to conceive a new child as soon as possible. Hence analyzing the time to next birth from a heterogeneity point of view might give some information on the natural variation in fecundability. In this example I will use data from the Norwegian medical birth registry. Of course, all the women have proven to be fertile in the sense of having conceived a child, although it was stillborn. Hence one cannot derive any estimate of the general population infertility from these data. Also, the fecundabilities of women having had a stillbirth may be different from those of other women. Nevertheless, it might be of some interest to fit the model of the present paper. I will use the same model as in the marriage example of Section 6. Certainly, all the qualifications made in that example are valid here too.

The data set consists of all Norwegian women who had their first birth during 1967 to 1971, who were at the time of this birth married and below 25 years of age, and for whom the child was stillborn. Here the time until birth of a second child will be studied. There is a total of 451 women who have been followed for a period varying from 10 to 15 years. In intervals varying from one month at the beginning to two years at the end of the 15-year period, it has been registered how many women give birth during the interval, and how many are censored.

The data have first been analyzed by the actuarial method. The observed incidence rates (occurrence/exposure rates) of a second birth are shown in Figure 8, and clearly exhibit a strongly unimodal shape. The corresponding actuarial survival curve is shown in Figure 9.

The model to be fitted is the same as in the marriage example, except that the cumulative basic intensity is given by

$$\Lambda(t) = \frac{a}{k+1}(t - 0.75)^{k+1}$$

for $t \geq 0.75$, where t denotes years since stillbirth. The latter equation implies that the Weibull model is only valid from 0.75 years, that is 9 months,

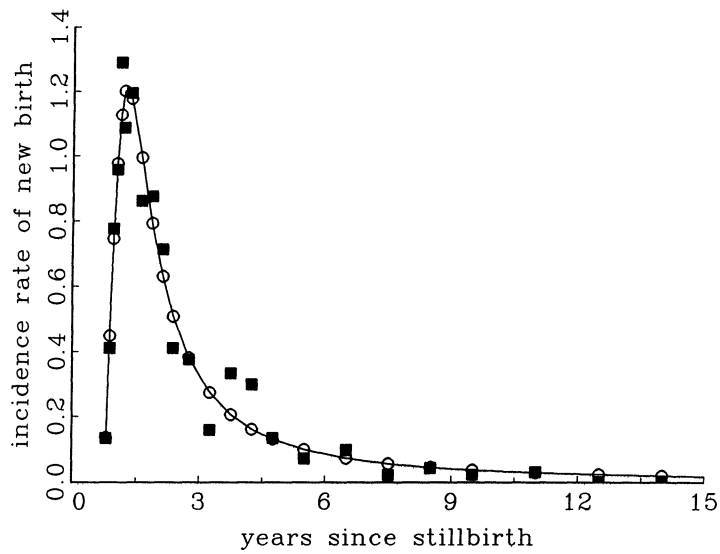


FIG. 8. *Incidence rate per year of new birth for women having had a stillbirth. Solid squares are observed rates; open circles are those estimated by the model.*

FIG. 9. *Proportion having not yet experienced a second birth among women with a stillborn first child. Solid squares represent the actuarial survival curve; open circles connected with lines (the circles are largely hidden behind the squares) give the curve estimated by the model.*

TABLE 2
Maximum likelihood estimates for the birth data

Parameters	δ	α	a	k
Estimates	1.55	1.31	0.43	1.15
s.e.	0.34	0.12	0.10	0.20

reflecting the basic fact the interval between births will in almost all cases exceed 9 months. The maximum likelihood estimates are given in Table 2.

The estimated value of α is 1.31 which is within the domain of the compound Poisson distribution. The probability of never conceiving a second child may be estimated from (7) by inserting estimates for the parameters, the result being 6.3%.

The survival curve $S(t)$ with estimates substituted for the parameters is presented in Figure 9, where it may be compared to the actuarial survival curve. Expected incidence rates based on the fitted model have been computed as in Section 6 and are presented in Figure 8, where they may be compared to the observed incidence rates. Apparently, the fit is quite good.

Acknowledgments. I am grateful to Ørnulf Borgan for providing me with the complete data used in the marriage example. The data have been extracted from official Danish statistics. I am also grateful to the Norwegian Medical Birth Registry for providing the data on child births, and to Berit Sandstad for help with organizing and analyzing these data.

REFERENCES

AABERGE, R., KRAVDAL, Ø. and WENNEMO, T. (1989). Unobserved heterogeneity in models of marriage dissolution. Discussion Paper 42, Central Bureau of Statist., Norway.

AALEN, O. O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine* **7** 1121–1137.

ABRAMOWITZ, M. and STEGUN, I. A. (1972). *Handbook of Mathematical Functions*. Dover, New York.

BAR-LEV, S. K. and ENIS, P. (1986). Reproducibility and natural exponential families with power variance functions. *Ann. Statist.* **14** 1507–1522.

BECKER, N. and RITTGEN, W. (1990). Some mathematical properties of cumulative damage models and their application in cancer epidemiology. *Biometrical J.* **32** 3–15.

BORGAN, Ø. and RAMLAU-HANSEN, H. (1985). Demographic incidence rates and estimates of intensities with incomplete information. *Ann. Statist.* **13** 564–582.

CLAYTON, D. (1978). A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65** 141–151.

COX, D. R. and OAKES, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

HECKMAN, J. J. and WALKER, J. R. (1990). Estimating fecundability from data on waiting times to first conception. *J. Amer. Statist. Assoc.* **85** 283–294.

HOEM, J. M. (1976). The statistical theory of demographic rates. A review of current developments (with discussion). *Scand. J. Statist.* **3** 169–185.

HOUGAARD, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* **71** 75–83.

HOUGAARD, P. (1986a). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73** 387–396.

HOUGAARD, P. (1986b). A class of multivariate failure time distributions. *Biometrika* **73** 671–678.

JØRGENSEN, B. (1987). Exponential dispersion models. *J. Roy. Statist. Soc. Ser. B* **49** 127–162.

LEADBETTER, M. R., LINDGREN, G. and ROOTZEN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin.

MANTON, K. G., STALLARD, E. and VAUPEL, J. W. (1981). Methods for comparing the mortality experience of heterogeneous populations. *Demography* **18** 389–410.

OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.* **84** 487–493.

SAMPSON, M. (1990). A Markov chain model for unskilled workers and the highly mobile. *J. Amer. Statist. Assoc.* **85** 177–180.

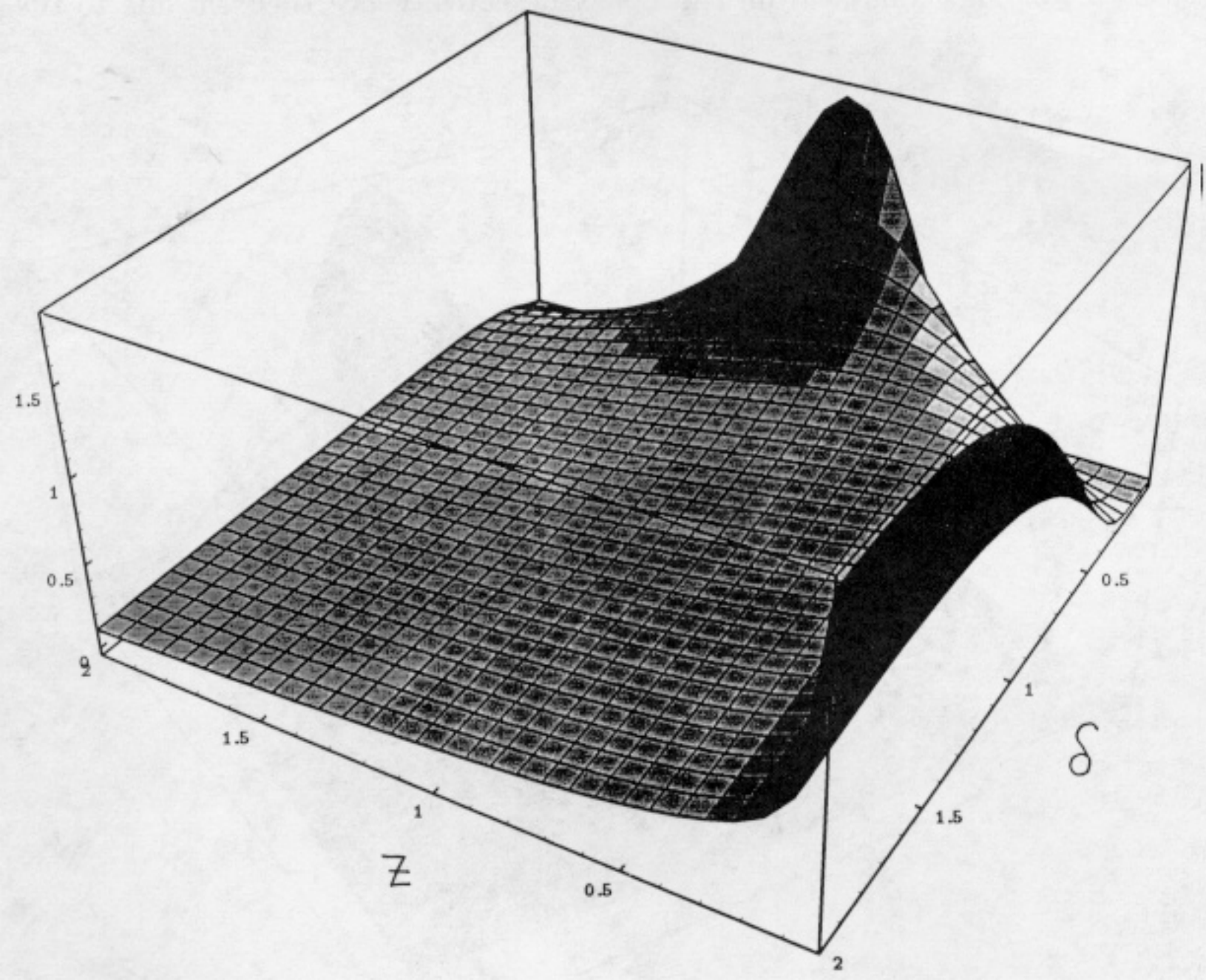
TIETZE, C. (1968). Fertility after discontinuation of intrauterine and oral contraception. *Internat. J. Fertility* **13** 385–389.

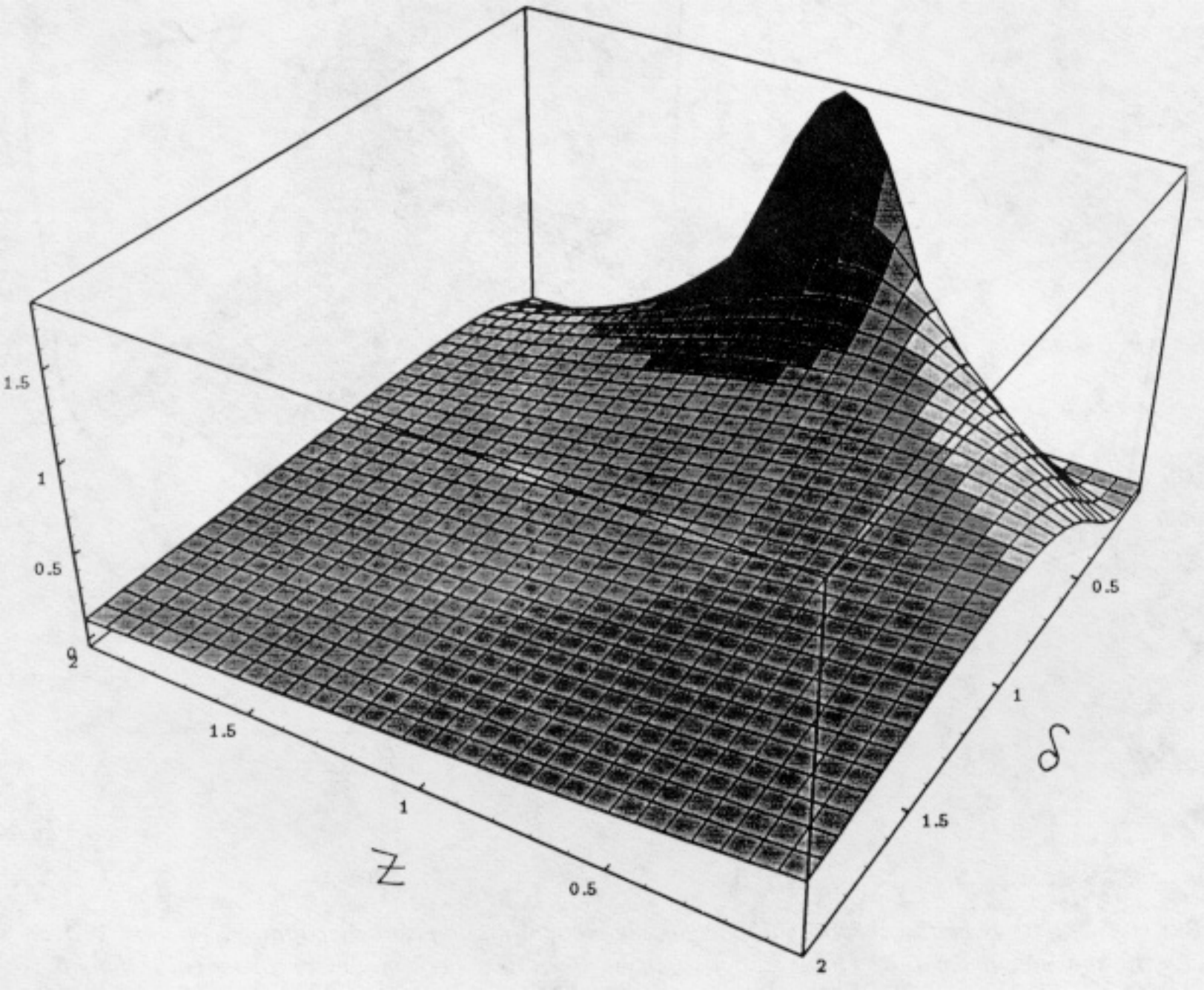
VAUPEL, J. W. (1990). Relatives' risk: Frailty models of life history data. *Theoret. Population Biol.* **37** 220–234.

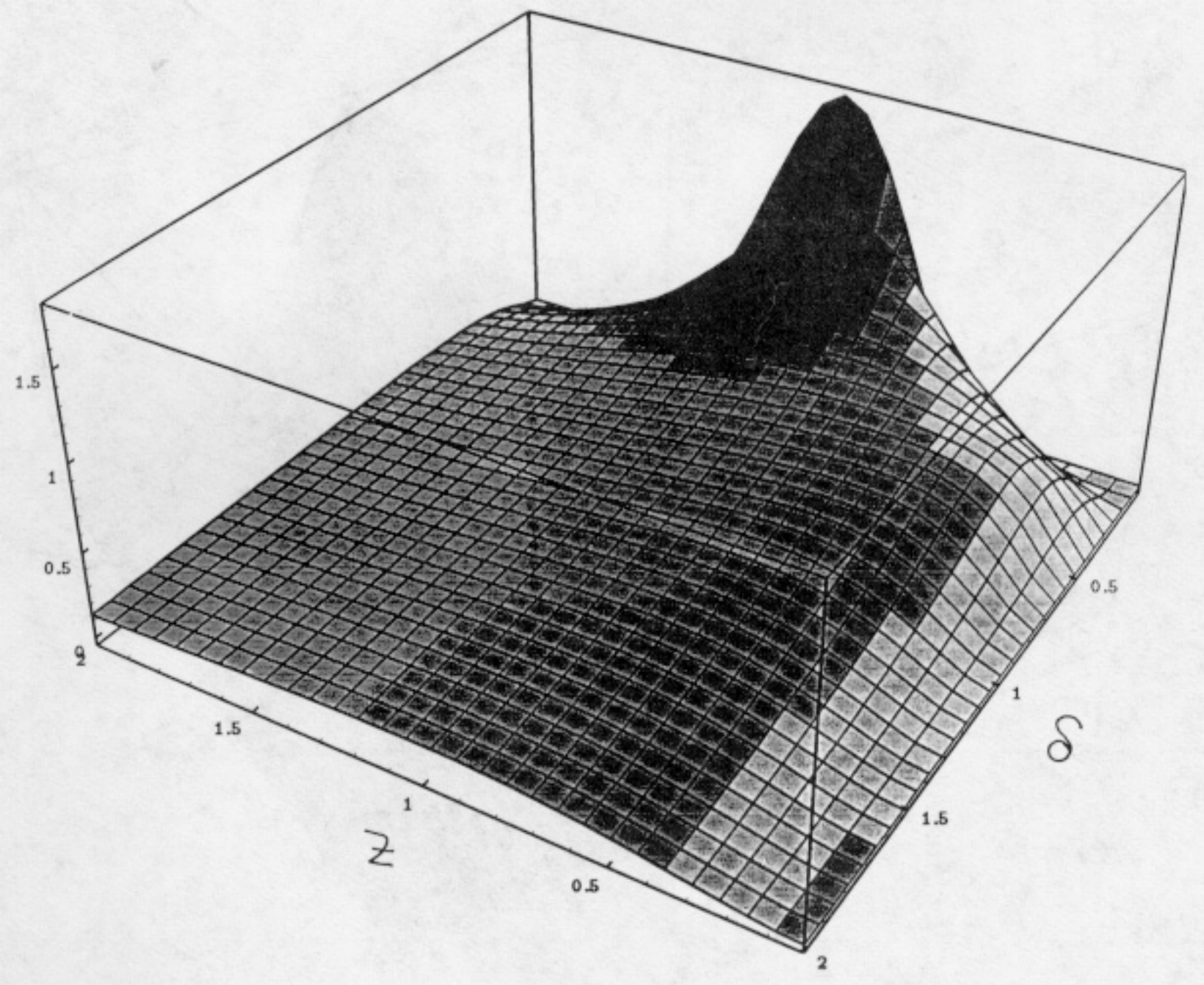
VAUPEL, J. W. and YASHIN, A. I. (1985). The deviant dynamics of death in heterogeneous populations. In *Sociological Methodology* (E. F. Borgatta, ed.) 179–211. Jossey-Bass, San Fransisco.

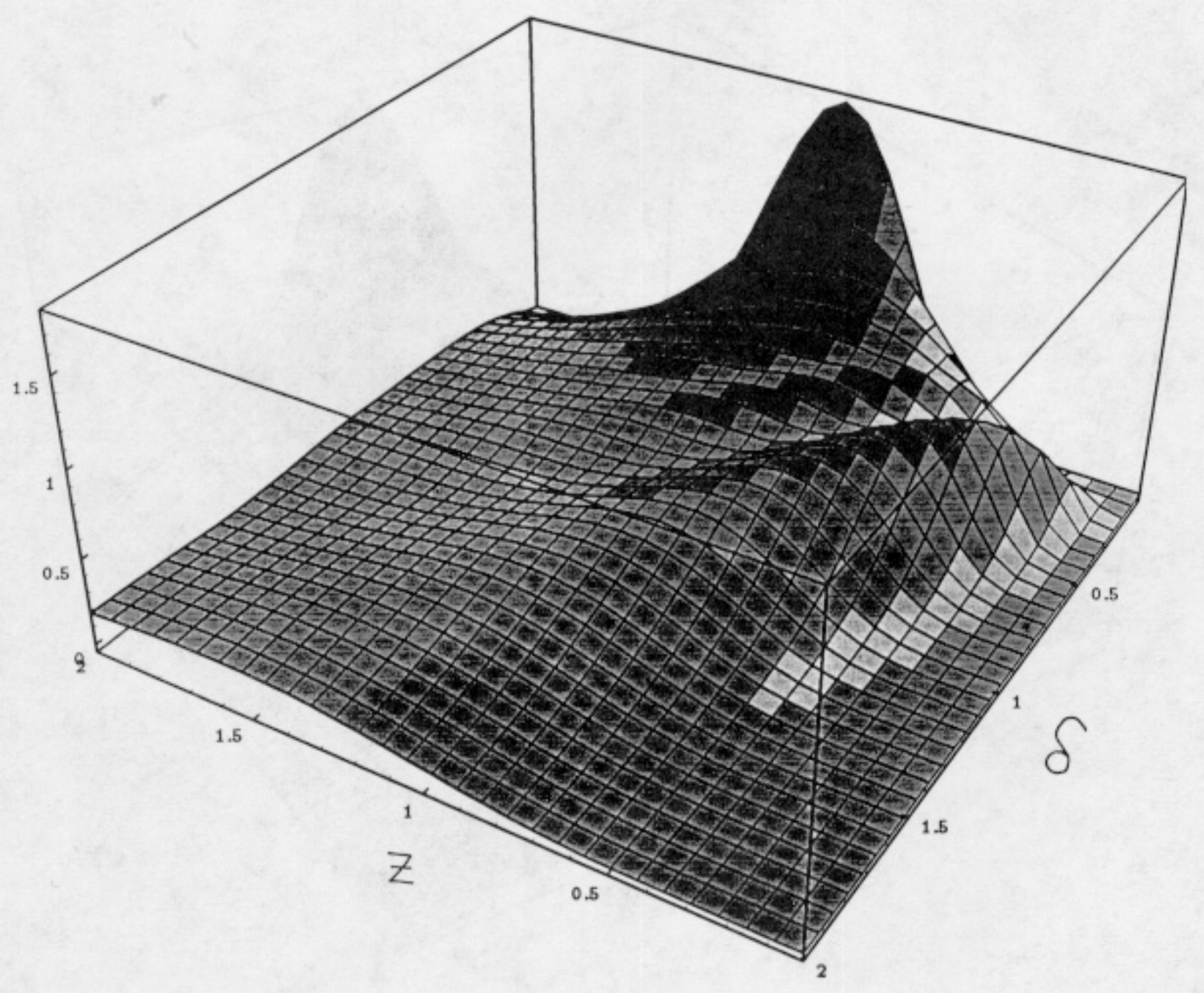
WOLFRAM, S. (1988). *Mathematica: A System for Doing Mathematics by Computer*. Addison-Wesley, Reading, Mass.

SECTION OF MEDICAL STATISTICS
UNIVERSITY OF OSLO
P.O. BOX 1122 BLINDERN
N-0317 OSLO
NORWAY


1.5
1
0.5
0
1.5
1
0.5
Z
0.5
1
1.5
2
δ


1.5
1
0.5
0
2
1.5
1
0.5
Z
0.5
1
1.5
2
δ


1.5
1
0.5
0
2
1.5
1
0.5
z
0.5
1
1.5
2
δ


1.5
1
0.5
0
1.5
1
0.5
Z
0.5
1
1.5
2
δ