A Note On Mixed Distributions

Rasa Varanka McKean

October 23, 2012

1 Introduction

The author intends to outline and clarify a basic application of mixed distributions. The equations are based on a life insurance publication written more than fifty years ago. By a change in perspective, the same model can be applied to workers compensation insurance for the fitting of probability density curves to a mixture of injury types.[1]

The original life insurance research paper did not consider workers compensation as an application; we can in the following way. Our example will be cash flows and their stopping times for different workers compensation and employers liability injury types. In this model, the cash flow stops or fails when the claim is closed. The WC and EL application is not necessarily based on mortality tables. The claim can close when the employee is healed and returns to work.

It should be noted that although mixed distributions are in use countrywide for workers compensation business, the application described in this paper may or may not be the same as the countrywide model.

The basic equations for the life insurance model are taken from statistical methods in the testing of failure rates. The failures can be due to a variety of causes. As one example, think of a group of cohorts in health insurance, each group of claimants having a certain illness. As another example, think of a population of automobiles, each failing due to a mechanical failure, electrical failure or normal deterioration.

First, we address basic notation.

Consider a mixture of failure sub populations. Denote the number of sub populations by the variable s. There will be s=5 different types of claims in our model and Employers Liability claims also, in a separate 6th sub population. Let r_i be the number of units belonging to the i^{th} sub population. For example, the first sub population contains r_1 units, the second sub population contains r_2 units, the i^{th} sub population contains r_i units, and the last sub population contains r_s units. Given a random sample of n units, the failure of r_1 units is due to cause (1), r_2 units fail due to cause (2), and so on up to r_s.

[1]"Estimation of Parameters of Mixed Exponentially Distributed Failure Time Distributions from Censored Life Test Data" by William Mendenhall and R.J. Hader Source: Biometrika Vol. 45 No. 3/4 (Dec., 1958) pp. 504-520

A random sample of n units is tested up to time $t = T$. Then $\Sigma_{i=1}^{s} r_i = r$ is the total number of units failing before time T and $(n - r)$ units which can't be identified as to sub population survive the test. The data is similar to censored life data.

In visual terms, think of the size of loss distribution as a matrix. Column (1) shows the fatalities, column (2) shows permanent total claims, column (3) all permanent partial, column (4) temporary total, column (5) medical only claims; here an additional column (6) will be included for Employers Liability. The rows of the matrix are the loss limits. The loss limits can start as low as $5,000$ and end as high as $10,000,000$. Note that the subscript i refers to the columns, the sub populations.

The matrix is populated with the number of claims by injury type whose ultimate payout is the size of the loss limit. The cash flow stops or fails when the claim is closed and the loss has reached a limit. We make one assumption to adapt property and casualty insurance to this model, that the claim incurred amounts increase with time. We'll ignore subrogation or other types of reimbursement. In other words, the claims are at their ultimate value. Recall from page 2 that the survival function $G(x)$ accounts for IBNR claims.

Back to notation.

Denote the failure times for the i^{th} sub population by t_{ij}. Then the r_i claims which close in sub population (1) can be ordered as $t_{11},\ t_{12},\ldots,\ t_{1j},\ldots,\ t_{1r_1}$. In other words, there are r_1 claims in sub population (1) and they close in a certain order in time. The r_2 claims in the second sub population can be ordered as $t_{21},\ t_{22},\ldots,\ t_{2j},\ldots,\ t_{2r_2}$. The i^{th} sub population contains r_i claims ordered as $t_{i1},\ t_{i2},\ldots,\ t_{ij},\ldots,\ t_{ir_i}$.

Let the sub populations be mixed in proportions $p_1,\ p_2,\ldots,\ p_s$. The p_i are constants.

Note that the number of different ways the claims can be ordered is: $\dfrac{n!}{r_1!r_2!\ldots r_i!\ldots r_s!(n-r)!}$

To simplify the computation, define a new variable $x_{ij} = t_{ij}/T$. Recall that T is the total allotted time for the experiment and that the t_{ij} are times to failure for each individual cash flow. Necessarily, each t_{ij} is less than the total time T. Then each x_{ij} is less than unity.

Now consider an arbitrary cumulative distribution function $F_i(x)$, the associated density function $f_i(x)$, and the survival function $G_i(x) = 1 - F_i(x)$. This general CDF can be either the exponential, the Weibull or the log normal distributions. Let $F(x) = \Sigma_{i=1}^{s} p_i F_i(x)$ and $G(x) = 1 - F(x)$. In the model below, the survival function $G(x)$ will account for the IBNR claims.

It should be noted that we are accustomed to thinking of $F(1) = 1$ and $G(1) = 0$ for $F(x)$ and $G(x)$ valued at $x = 1$. Here it isn't true because the x_{ij} have a maximum value of unity

As an example, consider the exponential distribution: $F_i(x) = 1 - \exp[-x/\beta_i]$

$$G(1) = 1 - \Sigma_{i=1}^{s} p_i F_i(1) = 1 - \Sigma_{i=1}^{s} p_i + \Sigma_{i=1}^{s} \exp[(-1/\beta_i))] = \Sigma_{i=1}^{s} \exp[(-1/\beta_i)]$$

since $\Sigma_{i=1}^{s} p_i = 1$. Thus $G(1)$ is not equal to zero.

2 The Basic Theory

We consider an arbitrary cumulative distribution function in this section. The calculation of the likelihood function will be clearer without specific detail. Some of the terms in the numerator and denominator of the likelihood function will cancel. The cancellations will be seen more clearly if detail is left out.

In Sections 3, 4, and 5, we consider examples of the mixed exponential, the mixed Weibull, and the mixed log normal distributions. The basic theory holds for an arbitrary CDF.

We need:

1. The probability of the ordered sequences of failure times,
2. The joint probability density functions,
3. The conditional probability of the ordered observations,
4. The likelihood function, and
5. The maximum likelihood estimates of the parameters.

The formula for the probability of the ordered sequences includes the number of possible ordered sequences, the probability of failure for claims in each of the sub populations, and the survival probability at time T for claims still open at the end of the experiment. The probability is evaluated at time $x = t/T$ for $t = T$. The value of x is then $x = 1$.

Given a random sample of n units comprised of i sub populations and total number of claims $r = r_1 + r_2 + \ldots + r_i + \ldots + r_s$, the probability that r_1 units will fail due to cause (1), that r_2 units will fail due to cause (2), that r_i units will fail due to cause (i), and that $(n - r)$ units will survive the test is given by a multinomial distribution.

Denote the above probability by $P(r_1, r_2, \ldots, r_s|n)$ then for $x = 1$ at time T:

$$P(r_1, r_2, \ldots, r_s|n) = \frac{n!}{r_1! r_2! \ldots r_s!(n-r)!} \prod_{i=1}^{s} [p_i F_i(1)]^{r_i} [G(1)]^{(n-r)} \quad (1)$$

At this point, the reader may want to review the joint density functions of order statistics. Good references may be found in the CAS Exam 1 syllabus. Recall that the joint probability density function is equal to the product of the density functions if and only if the random variables are independent.

Now we select the i^{th} sub population conditional on the event that there are r_i claimants in that sub population in order to derive a likelihood equation.

Denote the conditional probability distribution by $P(x_{i1}, x_{i2}, ..., x_{ir_i}|r_i)$ and the conditional probability density function by $p(x_{i1}, x_{i2}, ..., x_{ir_i}|r_i)$.

For the i^{th} sub population the joint density function for the ordered statistics conditional on the probability of r_i claimants in the i^{th} sub population at the end of the experiment is:

$$p(x_{i1}, x_{i2}, ..., x_{ir_i}|r_i) = r_i! \prod_{j=1}^{r_i} f_i(x_{ij})/[F_i(1)]^{r_i}$$

The joint conditional density for all of the s sub populations is the product of the s sub populations:

$$\prod_{i=1}^{s} p(x_{i1}, x_{i2}, ..., x_{ir_i}|r_i) = \prod_{i=1}^{s} r_i! \prod_{j=1}^{r_i} f_i(x_{ij})/[F_i(1)]^{r_i} \tag{2}$$

The likelihood function is the product of equations (1) and (2) above:

$$p(r_1, r_2, \ldots, r_s|n) \prod_{i=1}^{s} p(x_{i1}, x_{i2}, ..., x_{ir_i}|r_i) =$$

$$\frac{n!}{r_1!r_2!\ldots r_s!(n-r)!} \prod_{i=1}^{s} [p_i F_1(1)]^{r_i} [[G(1)]^{(n-r)} \prod_{i=1}^{s} r_i! \prod_{j=1}^{r_i} f_i(x_{ij})/[F_i(1)]^{r_i}$$

Notice that the terms $[F_i(1)]^{r_i}$ in both numerator and denominator cancel. The same is true for the product of the $r_i!$.

We are left with the likelihood and the log likelihood functions:

$$L = \frac{n!}{(n-r)!} \prod_{i=1}^{s} p_i^{r_i} \prod_{j=1}^{r_i} f_i(x_{ij}) [G(1)]^{(n-r)} \tag{3}$$

$$\ln L = \ln \frac{n!}{(n-r)!} + \Sigma_{i=1}^{s} p_i^{r_i} + \Sigma_{j=1}^{r_i} \ln f_i(x_{ij}) + (n-r) \ln[G(1)] \tag{4}$$

The p_i will be redefined here to clarify their relationship in the curve fitting process. It's important to note that there is no effect on the model or the calculations. As we will see in the maximum likelihood examples below, redefining the p_i clarifies that the most weight in the tail is given to the serious injury types. For instance, at some point in the actuarial data, the proportion p_2 of permanent total claims dominates the other injury type weights. Stay tuned...

Thus, we define a new functional form for the proportions p_i. Define $p_i(x) = p_i G_i(x)/G(x)$ for $G(x)$ the survival function as defined in the Introduction above. The proportions of the injury type curves now depend on the survival function. Define $k_i = p_i(1) = p_i G_i(1)/G(1)$.

At this point, maximum likelihood estimates will be computed for specific examples. The MLEs depend on the number of parameters in the distribution and we continue with specific forms of the equations.

3 The Mixed Exponential Distribution

Take the case of the mixed exponential distribution with $s = 2$ sub populations.

The number of distinct sequences of claims is $\dfrac{n!}{r_1!r_2!(n-r)!}$

The density functions are $f_1(x_{1j}) = (1/\beta_1)exp[-(x_{1j}/\beta_1)]$ and $f_2(x_{2j}) = (1/\beta_2)exp[-(x_{2j}/\beta_2)]$

The CDFs are $F_1(x_{1j}) = 1 - exp[-(x_{1j}/\beta_1)]$ and $F_2(x_{2j}) = 1 - exp[-(x_{2j}/\beta_2)]$

The survival functions are $G_1(x_{1j}) = exp[-(x_{1j}/\beta_1)]$ and $G_2(x_{2j}) = exp[-(x_{2j}/\beta_2)]$

Given a random sample of n units comprised of two sub populations and total number of claims $r = r_1 + r_2$, the probability that r_1 units will fail due to cause (1), that r_2 units will fail due to cause (2), and that $(n - r)$ units will survive the test is given by a multinomial distribution.

Denote the above conditional probability by $P(r_1, r_2|n)$ then for $x = 1$ at time T:

$$P(r_1, r_2|n) = \frac{n!}{r_1!r_2!(n-r)!}[p_1F_1(1)]^{r_1}[p_2F_2(1)]^{r_2}[G(1)]^{(n-r)} \tag{5}$$

The joint distribution in this example before conditioning is given by the product of the $f_1(x_{1j})$ and the $f_2(x_{2j})$ for the two sub populations $i = 1, 2$ and for all j.

For the 1^{st} and 2^{nd} sub populations the respective joint conditional density functions are:

$$p(x_{11}, x_{12}, ..., x_{1r_1}|r_1) = r_1!\prod_{j=1}^{r_1} f_1(x_{1j})/[F_1(1)]^{r_1} \tag{6}$$

$$p(x_{21}, x_{22}, ..., x_{2r_2}|r_2) = r_2!\prod_{j=1}^{r_2} f_2(x_{2j})/[F_2(1)]^{r_2} \tag{7}$$

Taking the product of the above three equations (5), (6) and (7) yields the likelihood function and the log likelihood function:

$$L = p(r_1, r_2|n)\prod_{i=1}^{2} p(x_{i1}, x_{i2}, ..., x_{ir_i}|r_i) = \frac{n!}{(n-r)!}\prod_{i=1}^{s} p_i^{r_i}\prod_{j=1}^{r_1} f_1(x_{1j})\prod_{j=1}^{r_2} f_2(x_{2j})[G(1)]^{(n-r)} \tag{8}$$

$$\ln L = \ln\frac{n!}{(n-r)!} + \Sigma_{i=1}^{s} r_i \ln p_i + \Sigma_{j=1}^{r_1}\ln f_1(x_{1j}) + \Sigma_{j=1}^{r_2}\ln f_2(x_{2j}) + (n-r)\ln[G(1)] \tag{9}$$

In order to derive maximum likelihood parameters, start by taking the partial derivative of the log likelihood function with respect to the first parameter.

$$\frac{\partial \ln L}{\partial \beta_1} = \frac{\partial \ln \dfrac{n!}{(n-r)!}}{\partial \beta_1} + \Sigma_{i=1}^{s} \frac{\partial r_i \ln p_i}{\partial \beta_1} + \Sigma_{j=1}^{r_1} \frac{\partial \ln f_1(x_{1j})}{\partial \beta_1} + \Sigma_{j=1}^{r_1} \frac{\partial \ln f_2(x_{2j})}{\partial \beta_1} + \frac{\partial (n-r) \ln[G(1)]}{\partial \beta_1} \tag{10}$$

Note that the first term in equation (9), the factorial, is a constant. The derivative of a constant is zero and the first term in equation (10) will disappear. The same is true for the second term since the p_i are constants. The derivative of the function $f_2(x_{2j})$ with respect to β_1 will also disappear since it is a function of β_2 but not β_1.

The following terms in the derivative of the log likelihood function with respect to β_1 remain:

$$\frac{\partial \ln L}{\partial \beta_1} = \Sigma_{j=1}^{r_1} \frac{\partial \ln f_1(x_{1j})}{\partial \beta_1} + (n-r) \frac{\partial \ln[G(1)]}{\partial \beta_1} \tag{11}$$

Consider the first term in equation (11):

$$f_1(x_{1j}) = \frac{1}{\beta_1} \exp[-(\frac{x_{ij}}{\beta_1})]$$

$$\ln f_1(x_{1j}) = -\ln \beta_1 - \frac{x_{1j}}{\beta_1}$$

$$\frac{\partial \ln f_1(x_{1j})}{\partial \beta_1} = -\frac{1}{\beta_1} + \frac{x_{1j}}{\beta_1^2}$$

$$\Sigma_{j=1}^{r_1} \frac{\partial \ln f_1(x_{1j})}{\partial \beta_1} = \Sigma_{j=1}^{r_1} - (\frac{1}{\beta_1}) + \Sigma_{j=1}^{r_1} (\frac{x_{1j}}{\beta_1^2}) = -\frac{r_1}{\beta_1} + (-\frac{r_1}{\beta_1^2}) \Sigma_{j=1}^{r_1} (\frac{x_{1j}}{r_1})$$

$$\Sigma_{j=1}^{r_1} \frac{\partial \ln f_1(x_{1j})}{\partial \beta_1} = -\frac{r_1}{\beta_1} + (\frac{r_1}{\beta_1^2}) \bar{x_1} \tag{12}$$

where $\bar{x_1}$ is the average of the r_1 values x_{1j}.

Consider the second term in the derivative of the log likelihood equation:

$$(n-r) \frac{\partial \ln[G(1)]}{\partial \beta_1} = (n-r) \frac{\partial \ln[1 - \Sigma_{i=1}^{2} p_i F_i(1)]}{\partial \beta_1} = \frac{(n-r)(p_1/\beta_1^2) \exp[-\frac{1}{\beta_1}]}{(p_1 \exp[-\frac{1}{\beta_1}] + p_2 \exp[-\frac{1}{\beta_2}])} \tag{13}$$

Substituting these results, equations (12) and (13) back into equation (11), we have so far:

$$\frac{\partial \ln L}{\partial \beta_1} = -\frac{r_1}{\beta_1} + (\frac{r_1}{\beta_1^2})\bar{x_1} + \frac{(n-r)(p_1/\beta_1^2)\exp[-\frac{1}{\beta_1}]}{(p_1 \exp[-\frac{1}{\beta_1}] + p_2 \exp[-\frac{1}{\beta_2}])}$$

To see the results more clearly, define the variable:

$$k_1 = \frac{(p_1)\exp[-\frac{1}{\beta_1}]}{(p_1 \exp[-\frac{1}{\beta_1}] + p_2 \exp[-\frac{1}{\beta_2}])} \tag{14}$$

Then we have the result:

$$\frac{\partial \ln L}{\partial \beta_1} = \frac{k(n-r)}{\beta_1^2} - \frac{r_1}{\beta_1} + \frac{r_1 \bar{x_1}}{\beta_1^2} \tag{15}$$

We can compute the derivative with respect to β_2 in a similar way:

$$\frac{\partial \ln L}{\partial \beta_2} = -\frac{r_2}{\beta_2} + (\frac{r_2}{\beta_2^2})\bar{x_2} + \frac{(n-r)(p_2/\beta_2^2)\exp[-\frac{1}{\beta_2}]}{(p_1 \exp[-\frac{1}{\beta_1}] + p_2 \exp[-\frac{1}{\beta_2}])}$$

And since

$$k_2 = (1-k_1) = 1 - \frac{p_1 \exp[-1/\beta_1]}{(p_1 \exp[-1/\beta_1] + p_2 \exp[-1/\beta_2])} = \frac{p_2 \exp[-1/\beta_2]}{(p_1 \exp[-1/\beta_1] + p_2 \exp[-1/\beta_2])} \tag{16}$$

we can then write:

$$\frac{\partial \ln L}{\partial \beta_2} = \frac{(1-k)(n-r)}{\beta_2^2} - \frac{r_2}{\beta_2} + \frac{r_2 \bar{x_2}}{\beta_2^2} \tag{17}$$

There remains one more derivative to take before setting the above derivatives in equations (15) and (17) equal to zero and solving for the optimal parameters. Note that there is a constraint in the system that the proportional amounts p_i add to unity when summed.

Recall from the introduction that the p_i will be redefined:

$p_i(x) = p_i G_i(x)/G(x)$, that $p_i(1) = p_i G_i(1)/G(1)$, and that $1 = \Sigma_{i=1}^{s} p_i$.

Recall also that for the exponential distribution: $F_i(x) = 1 - \exp[-x/\beta_i]$

Then, for the case of two sub populations where $s = 2$:

$$G(1) = 1 - \Sigma_{i=1}^{2} p_i F_i(1) = 1 - \Sigma_{i=1}^{2} p_i [1 - \exp[-x/\beta_i]] = 1 - \Sigma_{i=1}^{2} p_i + \Sigma_{i=1}^{2} p_i \exp[-\frac{1}{\beta_i}] = \Sigma_{i=1}^{2} p_i \exp[-\frac{1}{\beta_i}] \tag{18}$$

$$k_i = p_i(1) = \frac{p_i G_i(1)}{G(1)} = \frac{p_i[1 - F_i(1)]}{G(1)} = \frac{p_i \exp -[\frac{1}{\beta_i}]}{\Sigma_{i=1}^2 p_i \exp[-\frac{1}{\beta_i}]} \tag{19}$$

Before calculating the maximum likelihood equation in its entirety, firstly consider the term in the log likelihood equation (9) that involves $G(1)$. Utilizing equations (18) and (19):

$$\frac{\partial \ln G(1)}{\partial p_i} = \frac{1}{1 - F(1)} \frac{\partial[1 - p_1 F_1(1) - p_2 F_2(1)]}{\partial p_1} = \frac{1}{1 - F(1)} \frac{\partial[1 - p_1 F_1(1) - (1 - p_1) F_2(1)]}{\partial p_1}$$

$$\frac{\partial \ln G(1)}{\partial p_1} = \frac{[-F_1(1) + F_2(1)]}{1 - F(1)} = \frac{\exp(-\frac{1}{\beta_1}) - \exp(-\frac{1}{\beta_2})}{p_1 \exp(-\frac{1}{\beta_1}) + p_2 \exp(-\frac{1}{\beta_2})} = \frac{k_1}{p_1} - \frac{k_2}{p_2} \tag{20}$$

Now we'll compute the entire equation for $\frac{\partial \ln L}{\partial p_1}$ to reflect the constraint in the system of two sub populations.

$$\frac{\partial \ln L}{\partial p_1} = (n - r)\frac{\partial \ln G(1)}{\partial p_1} + \frac{\partial r_1 \ln p_1}{\partial p_1} + \frac{\partial r_2 \ln(1 - p_1)}{\partial p_1} = (n - r)[\frac{k_1}{p_1} - \frac{k_2}{p_2}] + \frac{r_1}{p_1} - \frac{r_2}{p_2} \tag{21}$$

Gathering the terms together from equations (15), (17), and (21), we can state that the system of maximum likelihood equations for two sub populations is the following. Each of these equations will be set to zero to derive the optimal set of parameters with a constraint:

$$\frac{\partial \ln L}{\partial \beta_1} = \frac{k_1(n-r)}{\beta_1^2} - \frac{r_1}{\beta_1} + \frac{r_1 \bar{x_1}}{\beta_1^2} = 0$$

$$\frac{\partial \ln L}{\partial \beta_2} = \frac{(1-k_1)(n-r)}{\beta_2^2} - \frac{r_2}{\beta_2} + \frac{r_2 \bar{x_2}}{\beta_2^2} = 0$$

$$\frac{\partial \ln L}{\partial p_1} = (n - r)[\frac{k_1}{p_1} - \frac{k_2}{p_2}] + \frac{r_1}{p_1} - \frac{r_2}{p_2} = 0$$

At this point, let's review which of the quantities are known and which are unknown.

The quantity n is the total number of cohorts in the population. The quantity r is the total number of known claims at the end of the experiment $t = T$. The quantities r_1 and r_2 are the number of known claims in the first and second sub populations respectively. The quantities $\bar{x}_i$ are the average values of the claims in each sub population. These are the known quantities.

The unknown quantities are the optimal β_i and the optimal p_i. At this point, we have three equations in three unknowns.

For the case of an arbitrary number of sub populations below, the seriously interested reader can work out similar equations, following the same steps and techniques as above.

Here we state the equations for an arbitrary number of sub populations:

$$L = \frac{n!}{(n-r)!}[G(1)]^{(n-r)} \prod_{i=1}^{s} p_i^{r_i} \prod_{i=1}^{s} \prod_{j=1}^{r_i} f_i(x_{ij}) \tag{22}$$

$$\ln L = \ln \frac{n!}{(n-r)!} + (n-r)\ln[G(1)] + \Sigma_{i=1}^{s} r_i \ln p_i + \Sigma_{i=1}^{s} \Sigma_{j=1}^{r_i} \ln f_i(x_{ij}) \tag{23}$$

$$\frac{\partial \ln L}{\partial \beta_i} = (n-r)\frac{\partial \ln[G(1)]}{\partial \beta_i} + \Sigma_{j=1}^{r_i} \frac{\partial \ln f_i(x_{ij})}{\partial \beta_i} = \frac{k_i(n-r)}{\beta_i^2} - \frac{r_i}{\beta_i} + \frac{r_i \bar{x}_i}{\beta_i^2} = 0 \tag{24}$$

$$\frac{\partial \ln L}{\partial p_i} = (n-r)\frac{\partial \ln G(1)}{\partial p_i} + \frac{\partial r_i \ln p_i}{\partial p_i} + \frac{\partial r_s \ln(1-\Sigma_{i=1}^{s-1} p_i)}{\partial p_i} = (n-r)[\frac{k_i}{p_i} - \frac{k_s}{p_s}] + \frac{r_i}{p_i} - \frac{r_s}{p_s} = 0 \tag{25}$$

Note that the equation for $\frac{\partial \ln L}{\partial p_i}$ holds for $i = 1, 2, \ldots, (s-1)$. The partial derivative with respect to p_s has been eliminated by the constraint $\Sigma_{i=1}^{s} p_i = 1$.

4 The Mixed Weibull Distribution

The probability density function, cumulative distribution function, and survival function for the Weibull distribution differs slightly in the exponential. Each of the functions is shown below.

The Weibull density functions are $f_i(x_{ij}) = (c_i/\beta_i)(x_{ij}/\beta_i)^{(c_i-1)} exp[-(x_{1j}/\beta_1)^{c_i}]$

The Weibull CDFs are $F_i(x_{ij}) = 1 - \exp[-(x_{ij}/\beta_i)^{c_i}]$

The Weibull survival functions are $G_i(x_{ij}) = \exp[-(\frac{x_{ij}}{\beta_i})^{c_i}]$ and $G(1) = \Sigma_{i=1}^{s} p_i \exp[-(1/\beta_i)^{c_i}]$

The likelihood and log likelihood functions are the same as before but the exact form of the density and survival functions will differ:

$$L = p(r_1, r_2, \ldots, r_i|n) \prod_{i=1}^{s} p(x_{i1}, x_{i2}, ..., x_{ir_i}|r_i) = \frac{n!}{(n-r)!} \prod_{i=1}^{s} p_i^{r_i} \prod_{i=1}^{s} \prod_{j=1}^{r_i} f_i(x_{ij})[G(1)]^{(n-r)}$$

$$\ln L = \ln \frac{n!}{(n-r)!} + \Sigma_{i=1}^{s} r_i \ln p_i + \Sigma_{i=1}^{s} \Sigma_{j=1}^{r_i} \ln f_i(x_{ij}) + (n-r)\ln[G(1)]$$

The derivatives with respect to β_i and p_i for both the Weibull and the exponential are similar and will seem familiar to the reader. However, as we will see, the derivative with respect to c_i is very different for the Weibull than for the exponential. In practical terms, the implementation will be more difficult.

Firstly, we'll take the derivative with respect to β_i:

$$\frac{\partial \ln L}{\partial \beta_i} = (n-r)\frac{\partial \ln[G(1)]}{\partial \beta_i} + \Sigma_{j=1}^{r_i}\frac{\partial \ln f_i(x_{ij})}{\partial \beta_i}$$

$$= \frac{(n-r)}{1-F(1)}\frac{\partial \Sigma_{i=1}^{s} p_i \exp[-(\frac{1}{\beta_i})^{c_i}]}{\partial \beta_i} + \Sigma_{j=1}^{r_i}\frac{\partial}{\partial \beta_i}[\ln c_i - \ln \beta_i + (c_i-1)(\ln x_{ij} - \ln \beta_i) - (\frac{x_{ij}}{\beta_i})^{c_i}]$$

$$= (n-r)\frac{p_i(\frac{c_i}{\beta_i})(\frac{1}{\beta_i})^{c_i}\exp[-(\frac{1}{\beta_i})^{c_i}]}{\Sigma_{i=1}^{s} p_i \exp[-(\frac{1}{\beta_i}^{c_i})]} + \Sigma_{j=1}^{r_i}[-\frac{1}{\beta_i} - (\frac{c_i-1}{\beta_i}) + (\frac{c_i}{\beta_i})(\frac{x_{ij}}{\beta_i})^{c_i}]$$

As before, define

$$k_i = \frac{p_i \exp[-(\frac{1}{\beta_i})^{c_i}]}{\Sigma_{i=1}^{s} p_i \exp[-(\frac{1}{\beta_i}^{c_i})]}$$

Then:

$$\frac{\partial \ln L}{\partial \beta_i} = \frac{c_i k_i (n-r)}{\beta_i^{c_i+1}} - \frac{c_i r_i}{\beta_i} + (\frac{c_i}{\beta_i^{c_i+1}})\Sigma_{j=1}^{r_i} x_{ij}^{c_i} \qquad (26)$$

Secondly, we'll take the derivative with respect to c_i:

$$\frac{\partial \ln L}{\partial c_i} = (n-r)\frac{\partial \ln[G(1)]}{\partial c_i} + \Sigma_{j=1}^{r_i}\frac{\partial \ln f_i(x_{ij})}{\partial c_i}$$

$$= \frac{(n-r)}{1-F(1)}\frac{\partial \ln \Sigma_{i=1}^{s} p_i \exp[-(\frac{1}{\beta_i})^{c_i}]}{\partial c_i} + \Sigma_{j=1}^{r_i}\frac{\partial}{\partial c_i}[\ln c_i - \ln \beta_i + (c_i-1)(\ln x_{ij} - \ln \beta_i) - (\frac{x_{ij}}{\beta_i})^{c_i}]$$

$$= (n-r)(\frac{1}{\beta_i})^{c_i}(\ln \beta_i)\frac{p_i \exp[-(\frac{1}{\beta_i})^{c_i}]}{\Sigma_{i=1}^{s} p_i \exp[-(\frac{1}{\beta_i}^{c_i})]} + \Sigma_{j=1}^{r_i}[\frac{1}{c_i} + (\ln x_{ij} - \ln \beta_i) - (\frac{x_{ij}}{\beta_i})^{c_i}\ln(\frac{x_{ij}}{\beta_i})]$$

$$= (n-r)(k_i)(\frac{1}{\beta_i})^{c_i}(\ln \beta_i) + \frac{r_i}{c_i} - r_i \ln \beta_i + \Sigma_{j=1}^{r_i}[\ln x_{ij} - (\frac{x_{ij}}{\beta_i})^{c_i}\ln(\frac{x_{ij}}{\beta_i})] \qquad (27)$$

The derivative of the Weibull with respect to p_i is similar to the exponential since the density function is not a function of p_i.

$$\frac{\partial \ln L}{\partial p_i} = (n-r)\frac{\partial \ln G(1)}{\partial p_i} + \frac{\partial r_i \ln p_i}{\partial p_i} + \frac{\partial r_s \ln(1-\Sigma_{i=1}^{s-1} p_i)}{\partial p_i}$$
$$= (n-r)[\frac{k_i}{p_i} - \frac{k_s}{p_s}] + \frac{r_i}{p_i} - \frac{r_s}{p_s} \quad (28)$$

Here we summarize the maximum likelihood equations for the mixed Weibull distribution, referring back to equations (26), (27), and (28):

$$\frac{\partial \ln L}{\partial \beta_i} = \frac{c_i k_i (n-r)}{p_i \beta_i^{c_i+1}} - \frac{c_i r_i}{\beta_i} + (\frac{c_i}{\beta_i^{c_i+1}})\Sigma_{j=1}^{r_i} x_{ij}^{c_i}$$

$$\frac{\partial \ln L}{\partial c_i} = (n-r)(k_i)(\frac{1}{\beta_i})^{c_i}(\ln \beta_i) + \frac{r_i}{c_i} - r_i \ln \beta_i + \Sigma_{j=1}^{r_i}[\ln x_{ij} + (\frac{x_{ij}}{\beta_i})^{c_i} \ln(\frac{x_{ij}}{\beta_i})]$$

$$\frac{\partial \ln L}{\partial p_i} = (n-r)[\frac{k_i}{p_i} - \frac{k_s}{p_s}] + \frac{r_i}{p_i} - \frac{r_s}{p_s}$$

Again, note that the equation for $\frac{\partial \ln L}{\partial p_i}$ holds for $i = 1, 2, \ldots, (s-1)$.

5 The Mixed Log Normal Distribution

Each of the probability density functions, cumulative distribution functions, and survival functions for both the normal and the log normal distributions will be shown below for easy reference.

The normal density function is given in the standard notation:

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \exp[-t^2/2]$$

By a change of variable, replacing t with $w = (t-\mu)/\sigma$ and replacing dt with $dw = \frac{1}{\sigma}dt$, the normal density function is:

$$\phi(\frac{t-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{(t-\mu)^2}{2\sigma^2}](\frac{1}{dt})d(\frac{t-\mu}{\sigma}) = \frac{1}{\sigma}\phi(t)$$

By a different change of variable, replacing t with $w = (\ln t - \mu)/\sigma$ and dt with $dw = \frac{1}{t\sigma}dt$, the log normal density function is:

$$\phi(\frac{\ln t - \mu}{\sigma}) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{(\ln t - \mu)^2}{2\sigma^2}](\frac{1}{t\sigma dt})d(\frac{(\ln t - \mu)}{\sigma}) = \frac{1}{t\sigma}\phi(t) \tag{29}$$

Now consider the normal distribution function, which is the integral of the density function:

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp^{[-t^2/2]}\, dt$$

Continuing in our notation with the change of variables above:

$$\Phi_i(\frac{x_{ij} - \mu_i}{\sigma_i}) = \int_{-\infty}^{(x_{ij}-\mu_i)/\sigma_i} \phi(t)dt = \frac{1}{\sigma_i} \int_{-\infty}^{w_{ij}} \phi(t)dt$$

$$\Phi_i(\frac{\ln x_{ij} - \mu_i}{\sigma_i}) = \int_{-\infty}^{(\ln x_{ij}-\mu_i)/\sigma_i} \phi(t)dt = \frac{1}{\sigma_i} \int_{-\infty}^{w_{ij}} \frac{1}{t}\phi(t)dt \tag{30}$$

Notice in the first equality of each of the two equations immediately above, that the integrand is the same as that for the normal distribution. What has changed is the upper limit of integration. In practical terms, to calculate the value of the log normal distribution function for a given value of x_{ij}, compute the value of $w_{ij} = (\ln x_{ij} - \mu_i)/\sigma_i$ and then look up w_{ij} in a table for the normal distribution. No additional tables are necessary for the log normal distribution function.

The survival function for the normal distribution is known as the Q-function in engineering textbooks. The survival function for the log normal distribution is expressed similarly:

$$Q(x) = \int_{x}^{\infty} \phi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp^{[-t^2/2]}\, dt = 1 - \Phi(x)$$

$$Q(\frac{x_{ij} - \mu_i}{\sigma_i}) = \int_{(x_{ij}-\mu_i)/\sigma_i}^{\infty} \phi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{(x_{ij}-\mu_i)/\sigma_i}^{\infty} \exp^{[-t^2/2]}\, dt = 1 - \Phi_i(\frac{x_{ij} - \mu_i}{\sigma_i})$$

$$Q(\frac{\ln x_{ij} - \mu_i}{\sigma_i}) = \int_{(\ln x_{ij}-\mu_i)/\sigma_i}^{\infty} \phi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{(\ln x_{ij}-\mu_i)/\sigma_i}^{\infty} \exp^{[-t^2/2]}\, dt = 1 - \Phi_i(\frac{\ln x_{ij} - \mu_i}{\sigma_i}) \tag{31}$$

We'll continue the derivation focused only on the log normal distribution. The likelihood and log likelihood functions are the same as before but the exact form of the density and survival functions will differ for the log normal.

Recall that equations (3) and (4) give us the likelihood and log likelihood functions for the general mixed distribution. In the standard notation for the log normal density and distribution functions, the analogous equations are now:

$$L = \frac{n!}{(n-r)!} \prod_{i=1}^{s} p_i^{r_i} \prod_{i=1}^{s} \prod_{j=1}^{r_i} \phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i})[Q(\frac{-\mu_i}{\sigma_i})]^{(n-r)} \tag{32}$$

$$\ln L = \ln \frac{n!}{(n-r)!} + \Sigma_{i=1}^{s} r_i \ln p_i + \Sigma_{i=1}^{s} \Sigma_{j=1}^{r_i} \ln \phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i}) + (n-r) \ln[Q(\frac{-\mu_i}{\sigma_i})] \tag{33}$$

Where, as before: $\Phi(x) = \Sigma_{i=1}^{s} p_i \Phi_i(x)$ and $Q(x) = 1 - \Phi(x)$.

The partial derivatives are taken with respect to the variables μ_i and σ_i. As before, the derivatives of the first two terms in $\ln L$ vanish when the partials are taken. The first two terms in $\ln L$ contain factorials and the variables p_i but not the variables μ_i and σ_i.

$$\frac{\partial \ln L}{\partial \mu_i} = \Sigma_{j=1}^{r_i} \frac{\partial \ln \phi_i(\frac{\ln x_{ij} - \mu_i}{\sigma_i})}{\partial \mu_i} + (n-r) \frac{\partial \ln[Q(\frac{-\mu_i}{\sigma_i})]}{\partial \mu_i} \tag{34}$$

Consider the partial with respect to μ_i of the density function in the first summation. By the chain rule:

$$\frac{\partial \ln \phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i})}{\partial \mu_i} = \frac{1}{\phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i})} \frac{\partial \phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i})}{\partial \mu_i} = \frac{1}{\phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i})} [\frac{1}{\sqrt{2\pi}}][\frac{2(\ln x_{ij} - \mu_i)}{2\sigma_i^2}] \exp[\frac{-(\ln x_{ij} - \mu_i)^2}{2\sigma_i^2}]$$

$$\frac{\partial \ln \phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i})}{\partial \mu_i} = \frac{(\ln x_{ij} - \mu_i)}{\sigma_i^2} \tag{35}$$

since the term $\phi(\frac{(\ln x_{ij} - \mu_i)}{\sigma_i})$ cancels from both the numerator and the denominator.

Similarly, for the partial with respect to σ_i:

$$\frac{\partial \ln \phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i})}{\partial \sigma_i} = \frac{1}{\phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i})} \frac{\partial \phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i})}{\partial \sigma_i} = \frac{1}{\phi(\frac{\ln x_{ij} - \mu_i}{\sigma_i})} [\frac{1}{\sqrt{2\pi}}][\frac{(\ln x_{ij} - \mu_i)^2}{\sigma_i^3}] \exp[\frac{-(\ln x_{ij} - \mu_i)^2}{2\sigma_i^2}]$$

$$\frac{\partial \ln \phi(\dfrac{\ln x_{ij} - \mu_i}{\sigma_i})}{\partial \sigma_i} = \frac{(\ln x_{ij} - \mu_i)^2}{\sigma_i^3} \tag{36}$$

Before proceeding, recall the operation of differentiation under the integral sign. Don't feel bad about looking it up in Wikipedia if you don't remember the formula.

For the function $F(x)$, with the proper conditions of continuity and differentiability allowing us to interchange a derivative and an integral, we have from the fundamental theorem of calculus:

$$F(x) = \int_{a(x)}^{b(x)} f(x,t)dt$$

$$\frac{\partial F(x)}{\partial dx} = f(x, b(x))\frac{\partial b(x)}{\partial x} - f(x, a(x))\frac{\partial a(x)}{\partial x} + \int_{a(x)}^{b(x)} \frac{\partial f(x,t)}{\partial x} dt \tag{37}$$

Now consider the last term in equation (34), the term with the survival function. We'll see that keeping the integrand as a function of only the variable t is a definite advantage here. If the integrand is not a function of μ_i or σ_i then differentiation under the integral sign will be particularly easy since the partial with respect to the integrand will vanish.

$$Q(\frac{-\mu_i}{\sigma_i}) = 1 - \Phi(\frac{-\mu_i}{\sigma_i}) = 1 - \Sigma_{i=1}^{s} p_i \int_{-\infty}^{(-\mu_i/\sigma_i)} \phi(t)dt$$

For clarity, let's first compute:

$$\frac{\partial Q(\dfrac{-\mu_i}{\sigma_i})}{\partial \mu_i} = -\Sigma_{i=1}^{s} p_i \frac{\partial}{\partial \mu_i} \int_{-\infty}^{(-\mu_i/\sigma_i)} \phi(t)dt \tag{38}$$

Then, insert (38) into:

$$\frac{\partial \ln Q(\dfrac{-\mu_i}{\sigma_i})}{\partial \mu_i} = \frac{1}{Q(\dfrac{-\mu_i}{\sigma_i})} \frac{\partial Q(\dfrac{-\mu_i}{\sigma_i})}{\partial \mu_i} = \Sigma_{i=1}^{s} \frac{-p_i}{Q(\dfrac{-\mu_i}{\sigma_i})} \frac{\partial}{\partial \mu_i} \int_{-\infty}^{(-\mu_i/\sigma_i)} \phi(t)dt$$

The integrand and the lower limit of integration are not functions of μ_i. By equation (37), the differentiation reduces to that of the upper limit of integration:

$$\frac{\partial \ln Q(\frac{-\mu_i}{\sigma_i})}{\partial \mu_i} = \Sigma_{i=1}^{s} \frac{-p_i}{Q(\frac{-\mu_i}{\sigma_i})} \frac{\partial(-\mu_i/\sigma_i)}{\partial \mu_i} = \Sigma_{i=1}^{s} \frac{(p_i/\sigma_i)}{Q(\frac{-\mu_i}{\sigma_i})} \tag{39}$$

Similarly, taking the partial with respect to σ_i, yields:

$$\frac{\partial \ln Q(\frac{-\mu_i}{\sigma_i})}{\partial \sigma_i} = \Sigma_{i=1}^{s} \frac{-p_i}{Q(\frac{-\mu_i}{\sigma_i})} \frac{\partial(-\mu_i/\sigma_i)}{\partial \sigma_i} = \Sigma_{i=1}^{s} \frac{(-p_i/\sigma_i^2)}{Q(\frac{-\mu_i}{\sigma_i})} \tag{40}$$

Gathering the terms in equations (35) and (39), we have from equation (34), the derivative of the log likelihood function with respect to μ_i:

$$\begin{aligned} \frac{\partial \ln L}{\partial \mu_i} &= \Sigma_{j=1}^{r_i} \frac{\partial \ln \phi_i(\frac{\ln x_{ij} - \mu_i}{\sigma_i})}{\partial \mu_i} + (n-r) \frac{\partial \ln[Q(\frac{-\mu_i}{\sigma_i})]}{\partial \mu_i} \\ &= \Sigma_{j=1}^{r_i} \frac{1}{\sigma_i} (\frac{\ln x_{ij} - \mu_i}{\sigma_i}) + (n-r) \Sigma_{i=1}^{s} \frac{(p_i/q_i)}{Q(\frac{-\mu_i}{\sigma_i})} \end{aligned} \tag{41}$$

Gathering the terms in equations (36) and (40), we have the derivative of the log likelihood function with respect to σ_i:

$$\begin{aligned} \frac{\partial \ln L}{\partial \sigma_i} &= \Sigma_{j=1}^{r_i} \frac{\partial \ln \phi_i(\frac{\ln x_{ij} - \mu_i}{\sigma_i})}{\partial \sigma_i} + (n-r) \frac{\partial \ln[Q(\frac{-\mu_i}{\sigma_i})]}{\partial \sigma_i} \\ &= \Sigma_{j=1}^{r_i} \frac{1}{\sigma_i} (\frac{\ln x_{ij} - \mu_i}{\sigma_i})^2 + (n-r) \Sigma_{i=1}^{s} \frac{(-p_i/\sigma_i^2)}{Q(\frac{-\mu_i}{\sigma_i})} \end{aligned} \tag{42}$$

The derivative of the Log Normal with respect to p_i is similar to the exponential and Weibull functions since the density function is not a function of p_i.

$$\frac{\partial \ln L}{\partial p_i} = (n-r) \frac{\partial \ln Q(\frac{-\mu_i}{\sigma_i})}{\partial p_i} + \frac{\partial r_i \ln p_i}{\partial p_i} + \frac{\partial r_s \ln(1 - \Sigma_{i=1}^{s-1} p_i)}{\partial p_i}$$

$$= -(n-r)\Sigma_{i=1}^{s} \frac{1}{Q(\frac{-\mu_i}{\sigma_i})} [\frac{\partial}{\partial p_i} p_i \int_{-\infty}^{(-\mu_i/\sigma_i)} \phi_i(t)dt] + \frac{r_i}{p_i} - \frac{r_s}{p_s}$$

$$= -(n-r)\Sigma_{i=1}^{s} \frac{1}{Q(\frac{-\mu_i}{\sigma_i})} [\int_{-\infty}^{(-\mu_i/\sigma_i)} \phi_i(t)dt] + \frac{r_i}{p_i} - \frac{r_s}{p_s}$$

$$= -(n-r)[\Sigma_{i=1}^{s} \frac{\Phi_i(\frac{-\mu_i}{\sigma_i})}{Q(\frac{-\mu_i}{\sigma_i})}] + \frac{r_i}{p_i} - \frac{r_s}{p_s}$$

$$= -(n-r)[\Sigma_{i=1}^{s} \frac{\Phi_i(\frac{-\mu_i}{\sigma_i})}{1 - \Sigma_{i=1}^{s} p_i \Phi_i(\frac{-\mu_i}{\sigma_i})}] + \frac{r_i}{p_i} - \frac{r_s}{p_s} \tag{43}$$

The maximum likelihood equations (41), (42), and (43) for the mixed log normal distribution are a challenge. The integral in the formula of the distribution has no closed form solution. This integral appears in both the numerator and denominator of the summation in equation (43). The values $\Phi(x)$ can be approximated very accurately for asymptotic (large) values of x. However, equation (43) could involve several approximations at each step of the MLE iteration. Thus, the algorithm could be lengthy. Professional optimization software is highly advisable.

6 Summary

A model of mixed distributions pertinent to workers compensation insurance is adapted from life insurance. Maximum likelihood equations for the mixed exponential, mixed Weibull, and mixed log normal distributions are derived for the fitting of a mixture of probability density curves by injury type. Implementation of the model requires optimization software.