Estimating the Parameters of the Distribution of a Mixture of Two Poisson Populations

I. Eisenberger
Communications Systems Research Section

If a random variable is such that, with probability p, it comes from a Poisson distributed population with parameter γ_1 and, with probability $(1 - p)$, it comes from a Poisson-distributed population with parameter γ_2, its density function is given by

$$g(x) = \frac{p\gamma_1^x e^{-\gamma_1} + (1-p)\gamma_2^x e^{-\gamma_2}}{x!}, \qquad x = 0,1,2,\cdots$$

The problem of estimating p, γ_1, and γ_2 is considered with respect to a DSIF application involving certain types of equipment for which the density function of time to failure obeys the exponential law.

I. Introduction

The general purpose of this article is to consider the problem of estimating the parameters of the distribution of a mixture of two Poisson populations. In particular, the method used in estimating the parameters will be applied to a practical problem arising in DSIF subsystems reliability.

Suppose one has M pieces of equipment, or devices, and the following are known:

(1) The distribution of the failure time of m_1 of these devices obeys the exponential law, $\lambda_1 \exp(-\lambda_1 t)$.

(2) The distribution of the failure time for the remaining $m_2 = M - m_1$ devices obeys the exponential law $\lambda_2 \exp(-\lambda_2 t)$.

(3) Each device can be repaired after failure and put back into service with no change in its operating characteristics.

The parameters m_1, λ_1 and λ_2 are unknown and it is desired to estimate these values from the failure information gained by operating the devices simultaneously for a given length of time, T.

It will be seen that by a suitable choice of a statistical model the problem of estimating m_1, λ_1 and λ_2 is transformed into that of estimating the parameters of the distribution of a mixture of two Poisson populations.

II. Distribution of a Mixture of Two Poisson Populations

If a random variable is such that, with probability p, it comes from a population which is Poisson distributed with parameter γ_1, and with probability $1 - p$ it comes from a population which is also Poisson distributed but with parameter γ_2, then its probability density function is given by

$$g(x) = \frac{p\gamma_1^x e^{-\gamma_1} + (1-p)\gamma_2^x e^{-\gamma_2}}{x!}, \qquad x = 0,1,2,3,\cdots$$

Here is a simple statistical model which makes use of the well-known fact that the number of failures in time T for a device with an exponential failure law is Poisson distributed with parameter λT. Let us assume that each device was chosen randomly and independently from a large reservoir of devices such that $100p\%$ of them have a failure rate λ_1 and the remainder have a failure rate λ_2. Under this assumption, the number of failures in time T for each device can be considered as a set of M independent observations taken from a population which is distributed as a mixture of two Poissons with parameters p, $\gamma_1 = \lambda_1 T$ and $\gamma_2 = \lambda_2 T$.

It should be observed, first, that $p \neq m_1/M$. In fact, m_1 is itself a random variable with a binomial distribution and, if it were known, m_1/M would be the maximum-likelihood estimate of p in the event that our statistical model described the actual physical situation and we were interested in the value of p for its own sake. As it is, m_1 can be estimated as the nearest integer to $\hat{p}M$.

It should also be noted that the cases of interest in this analysis are those for which, for a given T, the difference between λ_1 and λ_2 is not too great. For example, suppose $\lambda_1 = 0.01$, $\lambda_2 = 0.05$, $T = 500$, $M = 20$ and $m_1 = 14$. Then, after time T, we would have a group of 14 observations clustered about a mean value of 5 with a standard deviation of $\sqrt{5}$, and a group of 6 observations clustered about a mean value of 25 with a standard deviation of 5, and it would be highly unlikely, under the given statistical circumstances, that we would assign a given observation to the wrong group. Thus by knowing m_1 with virtual certainty, we would take m_1/M as an estimate of p (if that is what is desired) $N_1/14T$ as an estimate of λ_1 and $N_2/6T$ as an estimate of λ_2, where N_1 is the total number of failures in time T for the first group of 14 devices and N_2 is the total number of failures in time T for the devices in the second group.

It is only when it is impossible to classify the observations into two such groups with reasonable certainty that our statistical model comes into play. On the other hand, if λ_1 and λ_2 do not differ sufficiently for given T, then what is likely to happen, as we shall see, is that the solutions to the likelihood equations will converge to $\hat{p} = 1$ or 0, $\hat{\lambda}_1 = \hat{\lambda}_2 = N/MT$, where N is the total number of failures for all devices in time T. However, since the difference between the means of the component populations of the mixture is proportional to T, while the standard deviations are proportional to $\sqrt{T}$, then, for T sufficiently large more meaningful estimates can be obtained.

III. Maximum-Likelihood Estimators of p, λ_1 and λ_2

The method of maximum-likelihood (M.L.) will be used to estimate p, γ_1 and γ_2. Then M.L. estimates of λ_1 and λ_2 are given by $\hat{\lambda}_1 = \hat{\gamma}_1/T$, $\hat{\lambda}_2 = \hat{\gamma}_2/T$.

Due to the assumed independence of the observations, the likelihood function can be written as

$$L = \prod_{i=1}^{M} \frac{p\gamma_1^{k_i} e^{-\gamma_1} + (1-p)\gamma_2^{k_i} e^{-\gamma_2}}{k_i!}$$

where k_i is the number of failures in time T for the ith device. Then one has

$$\ell n\, L = \sum_{i=1}^{M} \left\{ -\ell n\, k_i + \ell n \left[p\gamma_1^{k_i} e^{-\gamma_1} + (1-p)\gamma_2^{k_i} e^{-\gamma_2} \right] \right\}$$

$$\frac{\partial \ell n L}{\partial p} = \sum_{i=1}^{M} \frac{A_1(k_i)}{D(k_i)}$$

$$\frac{\partial \ell n L}{\partial \gamma_1} = \sum_{i=1}^{M} \frac{A_2(k_i)}{D(k_i)}$$

$$\frac{\partial \ell n L}{\partial \gamma_2} = \sum_{i=1}^{M} \frac{A_3(k_i)}{D(k_i)}$$

where

$$A_1(k_i) = \gamma_1^{k_i} e^{-\gamma_1} - \gamma_2^{k_i} e^{-\gamma_2}$$

$$A_2(k_i) = pe^{-\gamma_1}\left(k_i\gamma_1^{k_i-1} - \gamma_1^{k_i}\right)$$

$$A_3(k_i) = (1-p)e^{-\gamma_2}\left(k_i\gamma_2^{k_i-1} - \gamma_2^{k_i}\right)$$

$$D(k_i) = k_i!\, g(k_i)$$

Setting the three partial derivatives equal to zero and solving for p, γ_1 and γ_2 provides the M.L. estimates of the parameters.

IV. Estimating p, λ_1 and λ_2 using Real Data

By using random numbers, 17 sets of sample values were generated such that each set can be considered as being taken from a population with density function $g(x)$, for various conditions and parameters. The method used for solving the likelihood equations was as follows.

Let

$$F_1 = \frac{\partial \ell nL}{\partial p},\ F_2 = \frac{\partial \ell nL}{\partial \gamma_1},\ F_3 = \frac{\partial \ell nL}{\partial \gamma_2}$$

Then, using the initial values $p^{(1)} = p$, $\gamma_1^{(1)} = \gamma_1$ and $\gamma_2^{(1)} = \gamma_2$, the following system of linear equations were solved in order to obtain the increments Δp, $\Delta\gamma_1$ and $\Delta\gamma_2$.

$$\frac{\partial F_i}{\partial p}\Delta p + \frac{\Delta F_i}{\partial \gamma_1}\Delta\gamma_1 + \frac{\partial F_i}{\partial \gamma_2}\Delta\gamma_2 = -F_i\,, \qquad i = 1,2,3$$

The above procedure was then repeated with $p^{(2)} = p^{(1)} + k\Delta p \gamma_1^{(2)} = \gamma_1^{(1)} + k\Delta\gamma_1$ and $\gamma_2^{(2)} = \gamma_2^{(1)} + k\Delta\gamma_2$, where k is a proportionality factor usually, but not always, taken as one. This iterative procedure was continued until the F_i were all zero or near zero. The final values of p, λ_1 and λ_2 were taken as the estimates. The results are shown in Table 1. It can be seen that for cases 1, 2, 6 and 12, $\hat{p}$ converged to one or zero and $\hat{\lambda}_1 = \hat{\lambda}_2$. However, for each case more satisfactory results were obtained when T was increased.

The large sample variances of the estimators were also computed and given in the last three columns of Table 1.

The author is grateful to Dr. Gary Lorden of Caltech for suggesting the statistical model used in the above analysis.

Table 1. Maximum-likelihood estimates of the parameters of the distribution of a mixture of two Poisson populations and the large sample variances of the estimators

Case	p	λ_1	λ_2	M	T	Maximum likelihood			Large sample		
						$\hat{p}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	M var. $(\hat{\lambda}_1)$	M var. $(\hat{\lambda}_2)$	M var. $(\hat{p})$
1	0.7	.01	.0125	20	1000	1	.0106	.0106	3.048×10^{-5}	1.527×10^{-4}	8.7234
2	0.7	.01	.0125	40	1000	1	.0109	.0109			
3	0.7	.01	.0125	40	1500	.8366	.0103	.0144	5.285×10^{-6}	2.573×10^{-5}	1.4534
4	0.7	.01	.0125	20	2000	.8811	.0100	.0169	4.552×10^{-6}	2.163×10^{-5}	.6787
5	0.7	.01	.0125	40	2000	.7380	.00953	.0140			
6	0.7	.01	.016	20	500	1	.0110	.0110	1.540×10^{-5}	8.209×10^{-5}	.7323
7	0.7	.01	.016	20	1000	.8089	.00982	.0184	3.114×10^{-6}	1.590×10^{-5}	.1327
8	0.7	.01	.016	20	1500	.7699	.0116	.0182	1.338×10^{-6}	6.569×10^{-6}	.0544
9	0.7	.01	.025	20	500	.7109	.0107	.0249	2.738×10^{-6}	1.858×10^{-5}	.0233
10	0.7	.01	.025	40	500	.7616	.0109	.0258			
11	0.7	.01	.03	20	500	.6156	.0114	.0302	2.027×10^{-6}	1.557×10^{-5}	.0142
12	0.3	.01	.0125	40	1500	0	.0117	.0117	2.603×10^{-5}	6.385×10^{-6}	1.6169
13	0.3	.01	.0125	40	2000	.6131	.0129	.0138	2.116×10^{-5}	5.491×10^{-6}	.7431
14	0.3	.01	.016	20	1500	.2344	.00944	.0157	4.846×10^{-6}	2.040×10^{-6}	.0589
15	0.3	.01	.016	40	1500	.1878	.00950	.0155			
16	0.3	.01	.025	20	500	.3400	.0110	.0225	9.192×10^{-6}	6.396×10^{-6}	.0254
17	0.3	.01	.025	40	500	.3767	.0108	.0227			