The Annals of Statistics
1982, Vol. 10, No. 2, 479–484

MIXTURES OF EXPONENTIAL DISTRIBUTIONS[1]

BY NICHOLAS P. JEWELL

Princeton University

Arbitrary nonparametric mixtures of exponential and Weibull (fixed shape) distributions are considered as possible models for a lifetime distribution. A characterization of such distributions is given by the well-known characterization of Laplace transforms. The maximum likelihood estimate of the mixing distribution is investigated and found to be supported on a finite number of points. It is shown to be unique and weakly convergent to the true mixing measure with probability one. A practical algorithm for computing the maximum likelihood estimate is described. Its performance is briefly discussed and some illustrative examples given.

1. Introduction. Parametric models for lifetime distributions have been extensively used in the study of data arising from times to failure of "units" (e.g. patients, machine components, etc.) under observation. We refer to Kalbfleisch and Prentice (1980) for a description of these models. Recently there has been much interest in models where the distribution function of the random time-to-failure is a *mixture* of a family of lifetime distributions. For finite mixture models we refer the reader to Mendenhall and Hader (1958), Rider (1961), Kao (1959), Cheng and Fu (1980). In engineering applications, an infinite mixture has been found to be appropriate and useful in describing shock models and random variations in hazard rates due to fluctuations in manufacturing tolerances (stochastic hazard rates); see Harris and Singpurwalla (1968), Doyle, Hansen and McNolty (1980), and Hill, Saunders and Laud (1980) for specific details. Recently, in studies of mortality in demography, infinite mixture models have been suggested for populations whose members differ in their endowment for longevity. For this and other applications in demography, see, for example, Vaupel, Manton and Stallard (1979).

Mixture models have also received a great deal of attention in statistics. Such models involve interesting estimation problems—see, for example, Blum and Susarla (1977), and Simar (1976). Hill, Saunders and Laud (1980) consider maximum likelihood estimation of mixing probabilities p_k, $k \geq 1$, when the data have a density of the form $g = \Sigma\, p_k f_k$ for known densities f_k. Laird (1978) obtains some general results about the existence of maximum likelihood estimates of arbitrary mixing measures. Here we provide a direct proof of the existence of a unique maximum likelihood (ML) estimate of an arbitrary measure in the case where the model is a mixture of exponential distributions. This direct approach yields more detailed properties of the estimate including weak convergence to the true population mixing measure. Our proof is easily extended to the situation where observations are arbitrarily right-censored. The methods can also be applied to mixtures of most one-parameter families of exponential family distributions.

The mixture model is also appropriate in the context of empirical Bayes estimation; see Robbins (1964, 1980). In this situation, the mixing measure represents an unknown prior distribution function which we wish to estimate using a set of observations.

Finally the mixture model approach provides a smooth estimate of the survival distribution function which may be preferable to the standard nonparametric step function estimates. It also has the exploratory value of providing useful information in determining possible non-homogeneous subpopulations.

Received November 1980; revised August 1981.

[1] Supported in part by a grant from the National Science Foundation.

AMS 1970 *subject classifications.* Primary 62G05, secondary 62N05.

Key words and phrases. Mixtures, survival distributions, Weibull distributions, exponential distributions, maximum likelihood, consistency, EM algorithm.

2. Preliminaries. We shall consider a lifetime distribution $F(t)$ having a density $f(t)$. The hazard rate of the distribution is given by $\lambda(t) = f(t)/\{1 - F(t)\}$ for t such that $F(t) < 1$. In this paper we shall be concerned with mixtures of the exponential distribution. The distribution function is then given by $F(t) = \int_0^\infty (1 - e^{-\lambda t})\, dM(\lambda)$, where M is any cumulative distribution function (cdf) supported on $(0, \infty)$ with $M(+\infty) = 1$.

Standard results on the Laplace transform now reveal the sort of distributions that are generated when you mix exponentials. Since mixtures of Weibull distributions with fixed shape parameter p are just powers of mixtures of exponentials, we can state a general result concerning mixture of Weibull distributions.

THEOREM. *Let F be a* cdf *on* $[0, \infty)$ *with* $F(0) = 0$, $F(+\infty) = 1$. *Then F arises as a mixture of Weibull distributions with fixed shape parameter $p = \gamma^{-1}$ if and only if $1 - F(x^\gamma)$ is a completely monotone function of x on* $(0, \infty)$.

PROOF. Feller (1971, page 439), together with the remark before the theorem, immediately provides the result.

There is a corollary to this result which has apparently not been noticed before.

COROLLARY. *The Weibull distribution with shape parameter p is a mixture of Weibull distributions with fixed shape parameter q so long as $p < q$.*

PROOF. Using the theorem, we need only check that $1 - F_p(x^{1/q})$ is completely monotone when $q > p$, where F_p is the Weibull distribution with shape parameter p, i.e. $F_p(x) = 1 - \exp(-\lambda x^p)$. Hence $1 - F_p(x^{1/q}) = \exp(-\lambda x^r)$ where $r = p/q$. It remains to show that the function $\exp(-\lambda x^r)$ is a completely monotone function of x when $r < 1$. This is easy.

COMMENTS.

1. In particular, the corollary tells us that any Weibull distribution with shape parameter less than 1 arises as a mixture of exponentials. Also the exponential distribution itself arises as a mixture of Weibull distributions with fixed shape parameter p, so long as $p > 1$.

2. As we shall see in Section 3, the Weibull distribution with shape parameter p does *not* arise as a non-trivial mixture of Weibull distributions with shape parameter p.

3. The mixing distribution M which yields a given F can be found using the inverse Laplace transform via the Fourier-Mellin integral. For example, the Weibull distribution with fixed shape parameter $p = ½$ and scale parameter μ arises as a mixture of exponentials where the mixing measure M is given by

$$dM(\lambda) = \frac{\mu e^{-\mu^2/4\lambda}}{2\sqrt{\pi\lambda^3}}\, d\lambda,$$

i.e. the stable distribution of order ½. For the details of similar calculations, see Doyle, Hansen and McNolty (1980).

4. We have the identity

$$1 - F(x^\gamma) = \exp\left\{-\int_0^{x^\lambda} \lambda(t)\, dt\right\}.$$

Feller (1971, page 441) shows that if ϕ is completely monotone and χ is a positive function with a completely monotone derivative then $\phi(\chi)$ is completely monotone. Now if λ is completely monotone, $\int_0^x \lambda(t)\, dt$ is a positive function with a completely monotone derivative. Hence $1 - F(x)$ is completely monotone. Similarly, since for $\lambda < 1$, x^γ is positive and has a completely monotone derivative, $1 - F(x)$ completely monotone implies $1 - F(x^\gamma)$ is completely monotone. Thus a sufficient condition for a survival distribution to be a mixture of exponentials is the complete monotonicity of the hazard function. It is easily seen that this condition is not necessary.

We remark also that it is easy to derive equations linking the first m derivatives of $\lambda(t)$ in terms of the first $(m+1)$ moments of the probability measure $e^{-\lambda x^p}\, dM(\lambda)/\kappa$ where $\kappa = \int_0^\infty e^{-\lambda x^p}\, dM(\lambda)$.

3. Maximum likelihood estimation of the mixing measure. In this section, we consider the existence and properties of the ML estimator of the mixing measure M given a sample of n independent observations from the mixture distribution $F(t) = \int_0^\infty (1 - e^{-\lambda t})\, dM(\lambda)$. Our remarks in Section 2 show that our results immediately extend to mixtures of Weibull distributions with fixed and known shape parameter. Our approach to ML estimation of the mixing measure heavily depends on work of Simar (1976) on ML estimation of a compound Poisson process.

Before we can make any sense of estimation of the mixing measure we must consider the problem of identifiability. Identifiability of the mixing measure of exponential distributions follows from Section 4 of Teicher (1961). So we can now turn to considering ML estimation of M. Suppose we have n independent observations, $t_1, \cdots, t_n$, arising from the mixture distribution $F(t) = \int_0^\infty (1 - e^{-\lambda t})\, dM(\lambda)$. The log-likelihood function can be written as

$$\Phi = \textstyle\sum_{j=1}^n \log\left\{ \int_0^\infty \lambda e^{-\lambda t_j}\, dM(\lambda) \right\} = \sum_{j=1}^n \log \beta_j = \Phi(\beta_1, \cdots, \beta_j) \tag{1}$$

where $\beta_j = \int_0^\infty \lambda e^{-\lambda t_j}\, dM(\lambda)$, $j = 1, \cdots, n$. We wish to determine the cdf M (if any) which maximizes the function Φ.

Let $\mathscr{F}$ be the set of positive measures on $[0, \infty)$ with total mass of at most 1 and $\mathscr{F}_0$ those measures in $\mathscr{F}$ with no mass at 0. Equations (1) define a map ϕ from $\mathscr{F}$ to the n-tuple $(\beta_1, \cdots, \beta_n)$ in $\mathbb{R}^n$. If $A \subseteq \mathbb{R}^n$ is the image of $\mathscr{F}$ under ϕ, it is easy to verify that A is compact and convex. Also Φ is strictly concave on A so that it attains its maximum value at a unique point of A which we shall denote by $(\hat\beta_1, \cdots, \hat\beta_n)$. Let $\hat M$ be a measure of mass of at most 1 with $\phi(\hat M) = (\hat\beta_1, \cdots, \hat\beta_n)$. Suppose the mass of $\hat M$ is $1 - \varepsilon$ where $\varepsilon > 0$. Then $\Phi(\hat M + \varepsilon\delta(1)) > \Phi(\hat\beta_1, \cdots, \hat\beta_n)$ where $\delta(1)$ represents a unit mass at 1. Thus all measures $\hat M$ in $\mathscr{F}$ with $\phi(\hat M) = (\hat\beta_1, \cdots, \hat\beta_n)$ have total mass 1. Note that this also shows that no $\hat M$ in $\mathscr{F}$ with $\phi(\hat M) = (\hat\beta_1, \cdots, \hat\beta_n)$ has any positive mass at 0. For if there existed such an $\hat M$ with mass $\alpha (> 0)$ at 0, then $\phi(\hat M - \alpha\delta(0)) = \phi(\hat M) = (\hat\beta_1, \cdots, \hat\beta_n)$ which, by the above, shows that $\hat M - \alpha\delta(0)$ has total mass 1 which contradicts the fact that $\hat M$ has total mass 1. We have now established the existence of an ML estimate of M which is in $\mathscr{F}_0$.

Let $G \in \mathscr{F}$ and $\phi(G) = (g_1, \cdots, g_n) \in A$. Consider the function h of ε defined on $[0, 1]$ by $h(\varepsilon) = \Phi(\phi\{(1-\varepsilon)\hat M + \varepsilon G\}) = \sum_{j=1}^n \log\{(1-\varepsilon)\hat\beta_j + \varepsilon g_j\}$. Now h is a strictly concave function of ε with a maximum at $\varepsilon = 0$. Hence, differentiation with respect to ε leads to

$$\textstyle\sum_{j=1}^n \frac{g_j - \hat\beta_j}{\hat\beta_j} \le 0. \tag{2}$$

With $G = \delta(\lambda)$, this gives

$$\frac{1}{n} \textstyle\sum_{j=1}^n \frac{\lambda e^{-\lambda t_j}}{\hat\beta_j} \le 1, \tag{3}$$

and this is true for all $\lambda \ge 0$.

Integrating both sides of (3) with respect to the measure $\hat M$ shows that equality holds in (3) everywhere on the support of $\hat M$. Polya and Szego (1925) show that a non-identically vanishing exponential polynomial $\sum_{j=1}^n p_j(y)\exp(k_j y)$, where p_j is a real ordinary polynomial of degree n_j, admits at most $\sum_{j=1}^n (n_j + 1) - 1$ zeros, counting multiplicities. Hence the support of $\hat M$ is finite and contains at most n distinct points.

We also can deduce information concerning the location of the points of support of $\hat M$ from the fact that equality holds in (3) everywhere on the support of $\hat M$. If $\lambda < t_{(n)}^{-1}$ then the derivative of $1/n \sum_{j=1}^n \lambda \exp(-\lambda t_j)/\hat\beta_j$ is positive and since the function is zero when $\lambda = 0$

we know that the points of support of $\hat{M}$ lie to the right of $t_{(n)}^{-1}$. Similarly we can deduce that the points of support of $\hat{M}$ lie to the left of $t_{(1)}^{-1}$.

We have thus shown that any ML estimate of M has finite support, containing at most n distinct points, each of these points lying in the closed interval $[t_{(n)}^{-1}, t_{(1)}^{-1}]$.

The point $(\hat{\beta}_1, \cdots, \hat{\beta}_n)$ in A is unique and it uniquely determines the number r and the points $\{\lambda_1, \cdots, \lambda_r\}$ in the support of any measure M with $\phi(M) = (\hat{\beta}_1, \cdots, \hat{\beta}_n)$. It remains to determine the mass p_m of $\hat{M}$ at λ_m for $m = 1, \cdots, r$. We have $\hat{\beta}_j = \sum_{m=1}^{r} p_m \lambda_m \exp(-\lambda_m t_j)$, $j = 1, \cdots, n$. We know there is one solution $(p_1, \cdots, p_r)$ to this set of equations. Suppose there is another solution $q_1, \cdots, q_r$. Then

$$\textstyle\sum_{m=1}^{r} p_m \lambda_m \exp(-\lambda_m t_j) = 0, \quad j = 1, \cdots, n, \quad \text{where } P_m = p_m - q_m.$$

Thus the function $\sum_{m=1}^{n} P_m \lambda_m \exp(-\lambda_m t)$ admits at least n distinct zeros. By the result in Polya and Szego given above, we know that $n \le r - 1$. But $r \le n$. This contradiction shows that $p_1, \cdots, p_r$ are uniquely determined and thus the ML estimate is unique.

REMARK. We may adapt the above argument for the case where k of the n observations are right-censored. The ML estimate $\hat{M}$ exists and is supported on a finite number, r, of distinct points. In this case r is bounded above by the integral part of $n - k/2$. However $\hat{M}$ may now have support at the origin and we no longer have the condition on the points lying in $[t_{(n)}^{-1}, t_{(1)}^{-1}]$. When $k = 0, 1, n - 1, n$, it is again easy to show that $\hat{M}$ is unique. However it does not appear to be unique for other values of k for certain combinations of the values $\lambda_1, \cdots, \lambda_r$ and the observations, although the number of points r and the points $\lambda_1, \cdots, \lambda_r$ are still uniquely determined.

We now prove a theorem describing the convergence of the unique ML estimate, denoted by M_n, derived from n independent observations.

THEOREM. *As $n \to \infty$, the sequence $\{M_n\}$ of* ML *estimates converges weakly with probability one to the true mixing distribution M_0.*

PROOF. The sequence $\{M_n\}$ has a subsequence which converges weakly to a positive measure M on $(0, \infty)$ with total mass of at most 1. Let F_n be the empirical distribution function associated with $t_1, \cdots, t_n$. Then, with probability one, F_n converges weakly to F as $n \to \infty$. Now, equation (2) above yields

$$n^{-1} \textstyle\sum_{j=1}^{n} \{f_{M_0}(t_j)/f_{M_n}(t_j)\} \le 1,$$

where for any measure $G \in \mathscr{F}_0$, $f_G(t) = \int_0^\infty \lambda e^{-\lambda t} dG(\lambda)$. This can be rewritten as

$$\int_0^\infty [f_{M_0}(t)/f_{M_n}(t)]\, dF_n(t) = \int_0^\infty h_n(t)\, dF_n(t) \le 1,$$

where $h_n(t) = f_{M_0}(t)/f_{M_n}(t)$. Thus, for a fixed $\kappa > 0$,

$$\int_{1/\kappa}^{\kappa} h_n(t)\, dF_n(t) \le 1. \tag{4}$$

Now there exist constants $a, b > 0$ such that the measure M_n has mass of at least $\delta > 0$ on $[a, b]$ for large n. Otherwise (4) is contradicted. Thus, for n large enough, $f_{M_n}(t) \ge \delta \min(ae^{-at}, be^{-bt})$, and so $h_n(t) \le (C/t)\{\min(ae^{-at}, be^{-bt})\}^{-1}$ for some constant $C > 0$. We thus have $h_n(t)$ bounded uniformly in n by a constant and so, applying the dominated convergence theorem together with standard techniques to (4), we obtain $\int_{1/\kappa}^{\kappa} \{f_{M_0}(t)/f_M(t)\}\, dF(t) \le 1$. Since the choice of κ was arbitrary we have $E\{f_{M_0}(t)/f_M(t)\} \le 1$ by the monotone convergence theorem.

Consider the function ψ on $\mathscr{F}_0$ given by

$$\psi(G) = E\left[\log\left\{\frac{f_G(t)}{f_{M_0}(t)}\right\}\right]$$

where the expectation is taken over the random variable t. By Jensen's inequality we have

$$\psi(G) \leq \log\left\{E\left[\frac{f_G(t)}{f_{M_0}(t)}\right]\right\} = 0$$

with equality if and only if $G = M_0$. Thus if we suppose $M \neq M_0$, we can consider

$$\lim_{\varepsilon\to 0} \frac{\psi((1-\varepsilon)M + \varepsilon M_0) - \psi(M)}{\varepsilon} > \psi(M_0) - \psi(M) > 0.$$

But this limit can be evaluated and is equal to $E\{f_{M_0}(t)/f_M(t)\} - 1$ which gives a contradiction. Thus $M = M_0$. Therefore, with probability one, any convergent subsequence of $\{M_n\}$ converges to M_0 and the result is established.

4. Computation of M_n. We know that M_n is a step function with a finite number, r, of steps. There are a variety of routines for maximizing the likelihood of a finite mixture of exponentials when the mixture is a step function with a *known* finite number of steps. Hasselblad (1969) suggested the following iterative scheme to evaluate $p_1, \cdots, p_r$ and $\lambda_1, \cdots, \lambda_r$, which are the masses and points of mass of M_n respectively, for fixed r:

at the kth *step, we have* $p_j^{(k)}, \lambda_j^{(k)}, j = 1, \cdots, r$;

put $g^{(k)}(t) = \sum_{j=1}^{r} p_j^{(k)}\lambda_j^{(k)} \exp(-\lambda_j^{(k)}t)$

and $f_j^{(k)}(t) = \lambda_j^{(k)} \exp(-\lambda_j^{(k)}t), j = 1, \cdots, r$;

then $p_j^{(k+1)} = p_j^{(k)}[\sum_{i=1}^{n} \{f_j^{(k)}(t_i)/g(t_i)\}]/n$,

and $\lambda_j^{(k+1)} = \sum_{i=1}^{n}\{f_j^{(k)}(t_i)/g^{(k)}(t_i)\}/\sum_{i=1}^{n}\{t_i f_j^{(k)}(t_i)/g^{(k)}(t_i)\}, j = 1, \cdots, r$.

This is, in fact, a special case of the EM algorithm for computation of ML estimates and Dempster, Laird and Rubin (1977) show that the iterative scheme converges to the ML estimate. Since $r(\leq n)$ is unknown in our case, one approach is to start with $r = 1$, maximize the likelihood and increase r as long as the log of the likelihood keeps increasing. To use the above algorithm, we need initial values for $p_1, \cdots, p_r, \lambda_1, \cdots, \lambda_r$. In practice, choice of initial estimates is relatively unimportant and equal values of the p_j's together with equally spaced λ_j's lying within the interval $[t_{(n)}^{-1}, t_{(1)}^{-1}]$ works well.

Only minor adjustments to the above algorithm are necessary to handle mixtures of Weibull distributions with fixed shape parameter p. Note that p must be chosen in advance since the family of mixtures of Weibull distributions with shape parameter $p > 0$ is not identifiable, by the Corollary in Section 2. Also, with a more refined use of the EM algorithm, we can use the above approach when some of the data are censored.

The algorithm described above to estimate M_n suffers the limitations and drawbacks which occur in estimating the parameters of a finite mixture of distributions. These are discussed in Hasselblad (1969). In summary, the variances of the ML estimates may be very large if the different scale parameters occurring in the mixture are not sufficiently distinct. Typically, more than 100 iterations are needed for convergence to 3 decimal places and the sample size needs to be large in order to detect "close" subpopulations. (Fortunately, in engineering and demographic applications, large sample sizes are often available.)

To illustrate, we indicate the results of three examples. First, 100 observations were generated from two exponential distributions with $\lambda_1 = 1, \lambda_2 = 15$ and mixing probabilities, .2 and .8 respectively. In this case, the estimate M_n yielded $\hat{\lambda}_1 = .863$ and $\hat{\lambda}_2 = 23.6$ with $\hat{p}_1 = .242$ and $\hat{p}_2 = .758$. Second observations were generated from a mixture of three exponentials with respective parameters $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 5, p_1 = .5, p_2 = .4$ and $p_3 = .1$. With sample size $n = 100$, M_n still had only one point of support. With $n = 200$, it had two points of support. With $n = 1000$, the estimates were $\hat{\lambda}_1 = 1.04, \hat{\lambda}_2 = 2.15, \hat{\lambda}_3 = 13.6, \hat{p}_1 = .500, \hat{p}_2 = .290$ and $\hat{p}_3 = .211$. Thirdly, 10,000 observations were generated from a mixture of three exponentials with $\lambda_1 = 1, \lambda_2 = 9, \lambda_3 = 15, p_1 = .2, p_2 = .1$ and $p_3 = .7$. The estimate M_n yielded $\hat{\lambda}_1 = 1.03, \hat{\lambda}_2 = 8.50, \hat{\lambda}_3 = 14.7, \hat{p}_1 = .195, \hat{p}_2 = .303$ and $\hat{p}_3 = .502$.

Thus the convergence proved in the previous section can be very slow. Care also is necessary in choosing which family of distributions to mix over—refer to the corollary in Section 2. For example, n observations were generated from the Weibull distribution with $\lambda = 1$ and shape parameter 0.5. Under the assumption of a mixture of exponentials (which is correct), M_n was computed. The table below gives some of the results. The true measure is given in Comment 3 of Section 2.

n		M_n		
10	$\hat{\lambda}_1 = 4.40$,	$\hat{\lambda}_2 = 0.646$,	$\hat{p}_1 = .507$,	$\hat{p}_2 = .493$
50	$\hat{\lambda}_1 = 2.74$,	$\hat{\lambda}_2 = 0.291$,	$\hat{p}_1 = .578$,	$\hat{p}_2 = .422$
100	$\hat{\lambda}_1 = 2190$,	$\hat{\lambda}_2 = 4.07$,	$\hat{\lambda}_3 = 2.63$,	$\hat{\lambda}_4 = .312$
	$\hat{p}_1 = .059$,	$\hat{p}_2 = .237$,	$\hat{p}_3 = .299$,	$\hat{p}_4 = .406$

Acknowledgements. The author would like to thank G. S. Watson who introduced him to mixtures of exponential distributions and also the referees for suggestions which improved the presentation of this paper.

REFERENCES

BLUM, J. R., and SUSARLA, V. (1977). Estimation of a mixing distribution function. *Ann. Probability* **5** 200–209.

CHENG, S. W., and FU, J. C. (1980). Estimation of mixed Weibull parameters in life testing. Florida State Univ. Statist. Rept. M564.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM Algorithm. *J. Roy. Statist. Soc. B* **39** 1–38 (with discussion).

DOYLE, J., HANSEN, E. and MCNOLTY, F. (1980). Properties of the mixed exponential failure process. *Technometrics* **22** 555–565.

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

HARRIS, C. M. and SINGPURWALLA, N. (1968). Life distributions derived from stochastic hazard functions. *IEEE Trans. on Reliability* **R-17** 70–79.

HASSELBLAD, V. (1969). Estimation of finite mixtures of distributions from the exponential family. *J. Amer. Statist. Assoc.* **64** 1459–1471.

HILL, D. L., SAUNDERS, R. and LAUD, P. W. (1980). Maximum likelihood estimation for mixtures. *Canad. J. Statist.* **8** 87–93.

KALBFLEISCH, J. D. and PRENTICE, R. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

KAO, J. H. K. (1959). A graphical estimation of mixed Weibull parameters in life-testing of electron tubes. *Technometrics* **1** 389–407.

LAIRD, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *J. Amer. Statist. Assoc.* **73** 805–811.

MENDENHALL, W. and HADER, J. R. (1958). Estimation of parameters of mixed exponentially failure time distributions from censored life test data. *Biometrika* **45** 504–520.

POLYA, G. and SZEGO, G. (1925). *Aufgaben und Lehrsatze aus der Analysis*, 2. Springer, Berlin.

RIDER, P. R. (1961). The method of moments applied to a mixture of two exponential distributions. *Ann. Math. Statist.* **32** 143–147.

ROBBINS, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

ROBBINS, H. (1980). Estimation and prediction for mixtures of the exponential distribution. *Proc. Nat. Acad. Sciences* **77** 2382–2383.

SIMAR, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4** 1200–1209.

TEICHER, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244–248.

VAUPEL, J. W., MANTON, K. G. and STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16** 439–454.

GROUP IN BIOSTATISTICS
DEPARTMENT OF BIOMEDICAL AND
ENVIRONMENTAL HEALTH SCIENCES
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CA 94720