

## Application of the Weibull Distribution Function to the Molecular Weight Distribution of Cellulose

A. BROIDO, *Pacific Southwest Forest and Range Experiment Station, Forest Service, U. S. Department of Agriculture, Berkeley, California 94701* and HSIUKANG YOW, *University of California Statewide Air Pollution Research Center, Riverside, California 92502*

### Synopsis

The molecular weight distribution of a linear homologous polymer is usually obtained empirically for any particular sample. Sample-to-sample comparisons are made in terms of the weight- or number-average molecular weights and graphic displays of the distribution curves. Such treatment generally precludes data interpretations in which a distribution can be described in terms of differing proportions of mixed populations. However, a statistical continuous univariate distribution function such as the Weibull can be applied to the molecular weight distributions measured for cellulose nitrate samples. The size distributions of samples degraded under some gentle stress may then be identified as consisting of differing proportions of populations with similar characteristic distributions. Analyzing data in this way should permit interpretation of the results as for breakdown of oligomers in a small-molecule system.

### INTRODUCTION

The polydispersity  $\rho$  of a linear homologous polymer is usually described in terms of the ratio of the weight-average molecular weight,  $\bar{M}_w$  to the number average molecular weight  $\bar{M}_n$ , where

$$\bar{M}_w = \frac{\sum_i n_i w_i^2}{\sum_i n_i w_i} = \sum_i w_i f_i \quad (1)$$

$$\bar{M}_n = \frac{\sum_i n_i w_i}{\sum_i n_i} = \frac{1}{\sum_i \frac{1}{w_i} f_i} \quad (2)$$

where  $n_i$  is the number of molecules of molecular weight  $w_i$ , and

$$f_i = \frac{n_i w_i}{\sum_i n_i w_i} \quad (3)$$

is the molecular weight frequency distribution on a weight basis. If the  $i$  size

classes are ranked in order of increasing molecular weights, the cumulative distribution, i.e., the weight fraction of molecules with molecular weight  $\leq w_j$ , is expressed by

$$F_j = \sum_{i=1}^j f_i \quad (4)$$

On a number basis, the molecular weight frequency distribution is simply

$$h_i = \frac{n_i}{\sum_i n_i} \quad (5)$$

and, with ranked size classes, the number fraction of molecules with molecular weight  $\leq w_j$  is

$$H_j = \sum_{i=1}^j h_i \quad (6)$$

Most often, the frequency distribution is obtained empirically for any particular sample, and sample-to-sample comparisons are made only in terms of  $\bar{M}_w$ ,  $\bar{M}_n$ , and graphic displays of the distribution curves. Frequently, this treatment of the data is adequate—particularly if the weight distribution is clearly characteristic of a single population. On the other hand, such treatment generally precludes data interpretations in which a distribution can be described in terms of differing proportions of mixed populations—especially when these populations have considerable overlap.

For example, consider the case of a monodisperse polymer with degree of polymerization ( $w$ ) of 1000 in which the molecules split in half in a slow process under some gentle stress. The resultant distribution would be bimodal, with the amount of material at  $w = 500$  increasing with time as the amount at  $w = 1000$  decreased. For typical polydispersities, on the other hand, at least the early results would appear simply as a gradual diminution of the average molecular weight, corresponding to a gradual shift to the left of the distribution curve. In general, the results would be indistinguishable from a process involving random breakdown along the length of the molecule, or from a molecular weight decrease by "peeling off" of monomer from one end.

Suppose, however, a characteristic functional relationship can be ascribed to the distribution of polymer sizes in a polymer population. The distribution of the breakdown sizes can then be identified in terms of another such population or as consisting of differing proportions of two or more populations each with its own characteristic distribution. The results can be interpreted then as in the case of the monodisperse system or as with oligomers in a small-molecule system.

One of the newer statistical continuous univariate distribution functions, the Weibull,<sup>1</sup> appeared to have sufficient versatility and analytical tractability for treatment of such problems. This paper describes its application to the molecular weight distributions of samples of cellulose nitrate obtained from filter paper and microcrystalline cellulose. A companion paper<sup>4</sup> discusses its use in interpreting changes in molecular weight distributions during the early pyrolysis of cellulose.

### EXPERIMENTAL

The molecular weight distributions were determined<sup>2,3</sup> by gel permeation chromatography (GPC) on the tetrahydrofuran (THF) solutions of the nitrated samples of filter paper (Whatman #541) and microcrystalline cellulose (Avicel). However, the data processing procedures were modified somewhat. In the earlier determinations, the authors used polystyrene standards to generate calibration curves (logarithm of degree of polymerization  $w$  versus retention volume  $v$ ) for cellulose nitrate in the GPC system. Comparison of the sample average molecular weights thus calculated gave agreement within 10% with viscometric measurements of the same solutions. Since the calibration curves were nearly linear over most of the range of interest, a new computer program was written in which a linear calibration curve was assumed, i.e.,

$$\ln w = sv + k \quad (7)$$

The values for  $s$  and  $k$  were chosen to give the best (least squares) fit to the viscometric measurements for all cellulose nitrate solutions to which the particular calibration curve applied.

For each GPC run, the basic output is a plot of refractometer signal versus retention volume. In the computer-linked data reduction process, the abscissa of the chromatogram is divided into  $\nu$  equal-volume increments and the incremental areas under the curve normalized in terms of the total. Thus,

$$A_i' = A_i/A_T \quad (8)$$

where

$$A_T = \sum_{i=1}^{\nu} A_i \quad (9)$$

is the total area of the chromatogram and  $A_i$  is the area under the chromatogram for  $v_i$ , the  $i$ th volume increment.

For sufficiently small increments,  $w_i$  may be taken as the molecular weight of the polymer molecules at the midpoint of the volume increment and may be obtained from  $v_i$  using eq. (7). With a direct linear relationship between the refractometer signal and the weight concentration of the polymer in  $v_i$ ,

$$f_i \equiv A_i' \quad (10)$$

To a degree of approximation determined by the accuracy of the calibration curve and the size of the increments,

$$h_i = \overline{M}_n(A_i'/w_i) \quad (11)$$

### THE WEIBULL CUMULATIVE DISTRIBUTION FUNCTION

For the present purposes, the Weibull cumulative distribution function may most conveniently be written in the following form:

$$F(x) = 1 - \exp\{-(x - \gamma)/\alpha\}^\beta \quad x \geq \gamma \geq 0, \alpha > 0, \beta > 0 \quad (12)$$

Although this function is finding increasing utility,<sup>5,6</sup> especially for problems involving "failure" (including, e.g., the breakdown of large particles to give a size distribution of smaller particles), it has no theoretical justification.<sup>1</sup> Thus, in

the present problem, there is no *a priori* reason for deciding whether the distribution should be expressed on a weight or number basis, i.e., whether  $F(x)$  should represent the weight fraction or number fraction of molecules of size  $\leq x$ . Neither choice is without its mathematical drawbacks, and it can be readily shown that a Weibull distribution on one basis cannot be converted into a mathematically exact Weibull distribution on the other basis. Empirically, however, it will be shown that, within the limits of experimental variation, reasonably good fits may be obtained on either basis, i.e., if the experimental molecular weight frequency distribution on a weight basis is such that it is reasonably approximated by

$$F = 1 - \exp\{-(w - \gamma)/\alpha\}^\beta \quad (13)$$

another approximation on a number basis may be found such that

$$H = 1 - \exp\{-(w - g)/a\}^b \quad (14)$$

### WEIGHT BASIS

The normalized experimental chromatogram corresponds to a plot of  $dF/d(\ln w)$  versus  $\ln w$ . The theoretical equivalent of the ordinate, obtained by differentiating eq. (13), is

$$f = dF/d(\ln w) = wdF/dw = (\beta/\alpha^\beta)w(w - \gamma)^{\beta-1} \exp\{-(w - \gamma)/\alpha\}^\beta \quad (15)$$

The location or threshold parameter  $\gamma$  corresponds to the  $w$  (more strictly, the  $v$ ) at which the first detectable signal above the solvent base line is seen. There is a single mode at  $w_m$ , defined by

$$[(w_m - \gamma)/\alpha]^\beta + (\gamma/\beta w_m) = 1 \quad (16)$$

such that  $w_m$  is  $\alpha$ , the scale parameter, when the shape parameter  $\beta$  is 1. Further, whenever  $w_m \gg \gamma$ , its value approaches  $\alpha$ , and  $\beta$  is related to the peak height by

$$\beta \simeq ef_m \simeq 2.718 \dots f_m \quad (17)$$

Another useful relationship, apparent from eq. (13), is that, whatever the value of  $\beta$ , when  $w = \gamma + \alpha$ ,

$$F_{\gamma+\alpha} = 1 - e^{-1} = 0.6321 \dots \quad (18)$$

Also, the cumulative Weibull distribution function can be linearized by a double logarithmic transformation; thus,

$$\ln \ln [1/(1 - F)] = \beta \ln (w - \gamma) - \ln \alpha \quad (19)$$

Then, when the correct value of  $\gamma$  is chosen, a plot of the left side of eq. (19) versus  $\ln (w - \gamma)$  gives a straight line with slope  $\beta$ . From this line,  $\alpha$  is obtained readily by eq. (18). To simplify such plots, Weibull probability graph paper is available commercially.

With  $\alpha$ ,  $\beta$ , and  $\gamma$  known, the computation of  $\bar{M}_w$  for the sample is straightforward:

$$\bar{M}_w = \int_{\ln \gamma}^{\infty} w f d(\ln w) = \int_{\gamma}^{\infty} f dw \quad (20)$$

Since  $f = wdF/dw$ ,  $\overline{M}_w$  is simply  $\mu$ , the first moment of  $w$  about the origin. Its analytical expression is

$$\overline{M}_w = \gamma + \alpha \Gamma[(\beta + 1)/\beta] \quad (21)$$

where  $\Gamma[r] = \int_0^\infty x^{r-1} e^{-x} dx$ , the gamma function.

Unfortunately, there is no comparable general analytical expression for

$$\overline{M}_n = 1 / \int_\gamma^\infty (f/w^2) dw \quad (22)$$

Thus,  $\overline{M}_n$  must be evaluated numerically—a tedious task by hand, but a trivial problem for even the smallest computer. Since the polydispersity  $\rho$  is simply the ratio of  $\overline{M}_w$  to  $\overline{M}_n$ , in the general case  $\rho$ , too, must be evaluated numerically.

### NUMBER BASIS

For comparison with a theoretical curve on a number basis, the experimental chromatogram requires an appropriate transformation. For any given weight of polymer, the number of molecules increases as  $w$  decreases. Thus, transformation to a number basis places increasing emphasis on the early (low  $w$ ) portions of the chromatogram, frequently placing greatest emphasis on that portion of the curve with the greatest experimental uncertainty. Nevertheless, from a computational and interpretive standpoint, the transformation is extremely attractive.

In practice, it is perhaps more elegant to transform the theoretical curve to match the experimental chromatogram. On a number basis, the theoretical equivalent of the ordinate is  $(w/\overline{M}_n)h$ , where

$$h = dH/d(\ln w) = wdH/dw = (b/a^b)w(w-g)^{b-1} \exp\{-(w-g)/a\}^b \quad (23)$$

Here,  $g$  is the threshold which, because of the exaggeration of this end of the curve, will usually appear slightly to the left of the threshold on a weight basis (i.e.,  $g < \gamma$ ). The mode is defined by

$$[(w_m - g)/a]^b = 1 + [(1 - 2g/w_m)/b] \quad (24)$$

Regardless of the value of  $b$ , for  $2g > w_m$ ,  $w_m < g + a$ ; for  $2g = w_m$ ,  $w_m = g + a$ ; for  $2g < w_m$ ,  $w_m > g + a$ . As  $b$  increases,  $w_m$  tends very rapidly to  $g + a$ . For  $b \simeq 1$ , the mode is at  $w_m = 2a$ .

Matching the theoretical curve to the experimental chromatogram requires the prior computation of the theoretical value for  $\overline{M}_n$ . Alternatively, a matching theoretical ordinate,  $wh$ , can be used if the total area under the experimental curve is normalized to the experimental value of  $\overline{M}_n$  rather than to unity. Another alternative is to transform the chromatogram to a number basis by dividing each  $A_i$  by the corresponding  $w_i$  and normalizing to

$$\sum_{i=1}^n A_i'/w_i = 1 \quad (25)$$

The comparable theoretical ordinate, then, is simply  $h$ , and the mode on this basis occurs at

$$w_m' = g + a [(b - 1)/b]^{1/b} \quad (26)$$

On a number basis, the relationship between  $a$ ,  $b$ , and  $g$  and  $\bar{M}_w$ ,  $\bar{M}_n$ , and  $\rho$ —as well as common statistical terms such as mean and variance—is straightforward if the method of moments is used:

$$\bar{M}_n = \int_g^{\infty} h dw \quad (27)$$

is the first moment about the origin, i.e., the mean. Its formula is

$$\bar{M}_n = g + a\Gamma[(b + 1)/b] \quad (28)$$

The second moment about the mean, the variance, is given by

$$\sigma^2 = a^2[\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)] \quad (29)$$

Now, since

$$\bar{M}_w = 1/\bar{M}_n \int_g^{\infty} wh dw \quad (30)$$

$\bar{M}_w\bar{M}_n$  is the second moment about the origin. Thus,

$$\bar{M}_w\bar{M}_n = \sigma^2 + \bar{M}_n^2 \quad (31)$$

or

$$\bar{M}_w = \bar{M}_n + \sigma^2/\bar{M}_n \quad (31')$$

Finally,

$$\rho = \bar{M}_w/\bar{M}_n = 1 + \sigma^2/\bar{M}_n^2 \quad (32)$$

### COMPARISON WITH EXPERIMENTAL DATA

The original recorder tracing of a typical chromatogram has a line width of roughly 0.7 mm and a peak height in the range of 20 to 50 mm. Thus, the width of the recording line corresponds to about 3% of the maximum signal. Further, the baseline is not flat, but is drawn between the "start" and "end" of the chromatogram. The selection of these two points, especially the lower terminus where the chromatogram rise is slow, adds additional uncertainty. Finally, the computer program—which involves reading the chromatogram at a limited number of select points and interpolating intermediate values—introduces further uncertainty. For all of these reasons, the criteria we have chosen for a "good fit" between a theoretical curve and the experimental data are agreement within 1% in the cumulative distribution and within 10% in the differential curve.

A typical normalized chromatogram and its normalized cumulative distribution on a weight basis are shown in Figure 1. The corresponding theoretical curves identified as F in Figure 1 were obtained by computer fitting a Weibull curve on a weight basis until agreement along the entire curve fell within the criteria for a "good fit." For the curves identified as H in Figure 1, the experimental data were converted to a number basis and the fitting procedure was repeated. The theoretical curve thus obtained was transposed back to a weight basis for plotting.

Within the criteria of a "good fit," then, Figure 1 demonstrates that Weibull

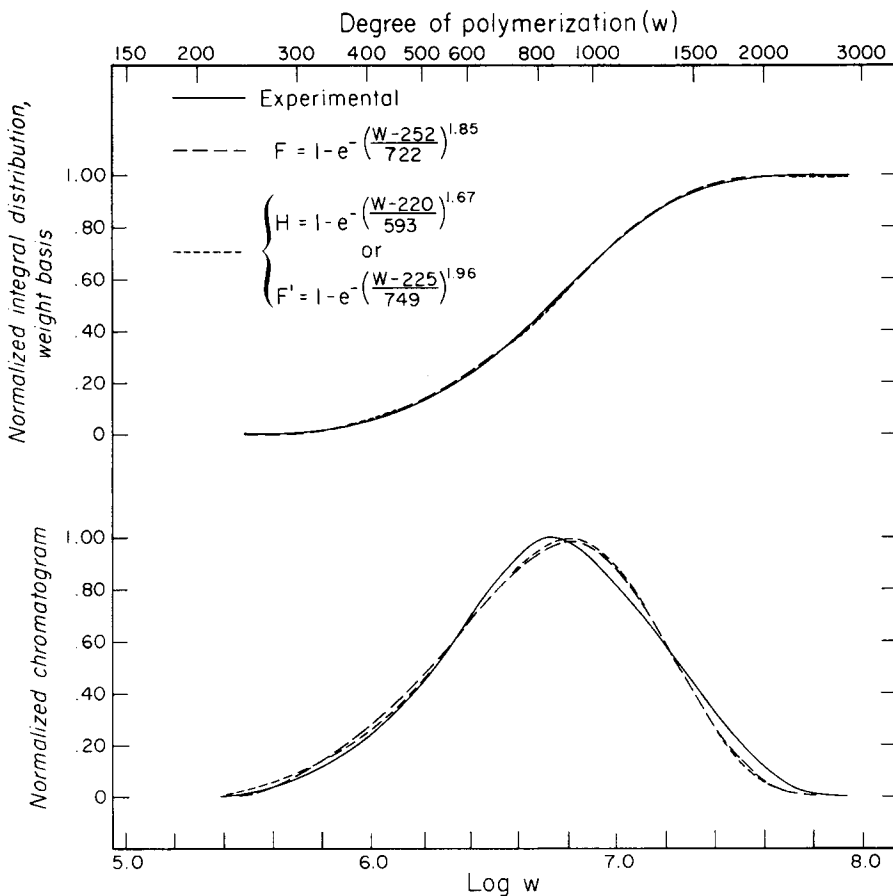


Fig. 1. Normalized experimental chromatogram and cumulative distribution on a weight basis with theoretical curves for Weibull distribution fits on a weight or number basis. Differences between curves for H and F' were too small to distinguish on this scale.

distributions on both a weight and a number basis can be fit to the same experimental distribution. Since mathematically equivalent Weibull distributions on both bases are excluded, one additional comparison was made. The values obtained on transposing the H curve to a weight basis were fed into the computer as a synthetic experimental curve to be fitted by a theoretical Weibull distribution. The resultant equation, labeled F' in Figure 1, agrees with curve H within better than 1% in differential form. However, F' does not meet the criteria for a good fit with the experimental data since its maximum deviation in the cumulative distribution is more than 1.2%.

For both the untreated filter paper and similar paper decrystallized by swelling in liquid ammonia, Weibull parameters (Table I) on both a weight and number basis were found to fit the corresponding experimental curves with agreement like that shown in Figure 1. On both the weight and number basis, the calculated values of  $\bar{M}_w$  and  $\bar{M}_n$  gave excellent agreement with the experimental values (Table II). Since there is more uncertainty in the determination of the experimental mode, agreement between experimental and theoretical values of  $w_m$  is not as good.

TABLE I  
Values for Weibull Parameters Used to Compute  $\bar{M}_w$ ,  $\bar{M}_n$ , and  $w_m$  for Table II

Sample	$\alpha$	$\beta$	$\gamma$	wt. %	$w_m$	$a$	$b$	$g$	no %	$w_m$
Untreated cellulose:										
1	738	1.70	186			564	1.50	152		
2	727	1.70	256			582	1.58	206		
3	722	1.85	252			592	1.67	220		
Decrystallized cellulose:										
1-component	683	1.68	215			544	1.58	168		
monomer	125	1.15	185	5.6	228.4					
trimer	478	2.30	240	16.8	685.2					
tetramer	732	1.70	255	77.6	913.6					
Microcrystalline cellulose:										
1 major	205	1.36	46.5	86	228	132	1.15	23.5	66.7	232
1 minor	34.5	1.65	16.0	14	46	34.5	1.25	16.0	33.3	60
2 major	222	1.11	63.4	76	237	133	1.10	37.0	33.3	244
2 minor	31.2	1.70	8.7	24	37	18.0	1.05	8.4	66.7	35
3 major	220	1.35	35.0	80	237	135	1.10	21.0	50.0	245
3 minor	37.0	1.25	15.0	20	44	25.4	1.14	15.4	50.0	47
3' major	220	1.95	12.0	66	228	119	1.15	60.0	33.3	220
3' minor	68.0	1.28	12.0	34	73	31.0	1.00	13.6	66.7	62

The polydispersities of the microcrystalline cellulose samples are much higher than those of the filter paper samples, and we were totally unsuccessful in our attempts to fit a single-population Weibull distribution curve to any microcrystalline cellulose chromatogram. However, each chromatogram could be described as composed of two overlapping Weibull distributions: a major component (on a weight basis) covering the high molecular weight region and a minor component (generally about 20% on a weight basis) which might be considered to represent further degradation of the major component to lower molecular weights.

The concept may be illustrated by considering microcrystalline cellulose samples 3 and 3'. These two measurements were made on the same nitrated sample of the starting material. For sample 3, the chromatogram was obtained on a freshly prepared solution of the cellulose nitrate. For sample 3', the solution was allowed to stand at room temperature to permit further degradation as indicated by the lower average molecular weight (Table II). The chromatogram for sample 3 was resolved into an 80% fraction with mode near 230 and a 20% fraction of smaller-sized molecules (Table I). The major component of sample 3' also has its mode near 230, but now the major component comprises only 66% of the sample. It is worth noting, but not necessarily significant, that on a



TABLE II  
Experimental Values of  $\bar{M}_w$ ,  $\bar{M}_n$ , and  $w_m$  and  
Theoretical Values Calculated from Parameters in Table I

Sample	$\bar{M}_w$			$\bar{M}_n$			$w_m$		
	Expt.	Eq. F	Eq. H	Expt.	Eq. F	Eq. H	Expt.	Eq. F	Eq. H
Untreated cellulose:									
1	860	844	842	664	665	661	794	867	869
2	916	904	886	749	747	728	828	910	909
3	911	893	891	758	754	749	864	914	916
Decrystallized cellulose:									
1-component composite	832	826	810	663	668	657	723	834	836
		834			660			790	
Microcrystalline cellulose <sup>a</sup> :									
1	207	208	206	116	114	115	240	228	225
2	213	217	198	73	80	73	283	239	243
3	201	199	204	95	95	95	256	236	242
3'	161	162	164	86	87	87	182	206	194

<sup>a</sup> All theoretical curves for microcrystalline cellulose are composite.

number basis all the microcrystalline cellulose samples can be represented as ratios of 2 to 1, 1 to 1, or 1 to 2 of the two components.

The major component of all the microcrystalline cellulose samples had modal values near 230, while the modal values for the single-component Weibull fits to the filter paper samples were very nearly four times as large. Microcrystalline cellulose is produced by the acid hydrolysis of larger cellulose molecules—one of several size-degradation processes which, for cellulose, is characterized by a sharp drop in average size of the material to a “leveling off” degree of polymerization. If, then,  $w_m = 230$  were considered to represent a macromolecular “monomer” at this “leveling off” size, the original filter paper samples would be comprised of tetramers. Then, as the average molecular weight of such samples dropped as the sample was degraded, the degraded chromatograms might be resolvable in terms of appropriate fractions of trimers, dimers, and monomers.

Consider, for example, the decrystallized cellulose sample listed in the tables. A good single-component Weibull distribution curve fit can be found for its chromatogram. However, the average molecular weight of this sample is somewhat lower than those for the untreated cellulose. Correspondingly, the modal value for the Weibull approximation is low. Suppose that during the processing of this sample (during the decrystallization, the nitration, or the solution of the nitrate) partial degradation occurred, with each decomposing tetramer molecule breaking down into a trimer and a monomer. Then, on a weight basis, the yield of monomer should be just one fourth the loss of tetramer while the yield of trimer should be three fourths the tetramer loss. The “composite” values shown in Table II for a distribution synthesized from the components shown in Table I indicate the excellent agreement which may be obtained.

Untreated cellulose sample 1 also has an average molecular weight somewhat lower than the other two untreated samples in Table II. Its chromatogram, too, can be resolved into a major tetramer component with modal degree of polymerization 914 and minor contributions from monomer and trimer components. Since all samples can degrade slightly during processing, this high value for  $w_m$

has been taken as a lower limit for the "true" value of the tetramer. Of course, the degree of polymerization of the monomer then would be about 228, at the low end of the modal values observed for the major component of the microcrystalline cellulose samples. Since microcrystalline cellulose is obtained by degrading larger cellulose molecules until they approach a levelling-off degree of polymerization, a selection at the low end here would seem appropriate.

In conclusion, then, the Weibull distribution function can be applied to the molecular weight distributions of cellulose samples. The use of such a function permits a search for interpretations of experimental results in ways not possible with the usual empirical sample-to-sample comparisons. The application of these techniques to a study of the beginning phases of cellulose pyrolysis is the subject of a companion paper.<sup>4</sup>

Thanks are due Professor Branko Grunbaum, University of Washington, and Professor Forman A. Williams, University of California at San Diego, for many helpful discussions. This work was supported in part by Grants AP00658 from the U.S. Environmental Protection Agency and GP34494 from the National Science Foundation to the University of California Statewide Air Pollution Research Center. This paper does not necessarily reflect the views and policies of these agencies nor does mention of trade names or commercial products constitute endorsement by them or by the U.S. Department of Agriculture.

### References

1. W. Weibull, *J. Appl. Mech.*, **18**, 293 (1951).
2. A. C. Ouano, E. M. Barrall II, A. Broido, and A. C. Javier-Son, *Recent Trends in the Determination of Molecular Weight*, M. Ezrin, Ed., American Chemical Society Advances in Chemistry Series, No. 125, 1973, pp. 187-193.
3. A. Broido, A. C. Javier-Son, A. C. Ouano, and E. M. Barrall II, *J. Appl. Polym. Sci.*, **17**, 3627 (1973).
4. A. Broido and H. Yow, *J. Appl. Polym. Sci.*, **21**, 1677 (1977).
5. Fred H. Steiger, *Chem. Technol.*, **1**, 225 (1971).
6. N. L. Johnson and S. Kotz, *Continuous Univariate Distributions I*, Houghton-Mifflin, Boston, 1970, pp. 250-271.

Received March 25, 1976

Revised May 25, 1976