New Maximum-Likelihood Method for Reduction of Distribution-Coefficient Data

Simon L. Marshall
Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

The sampling distributions for the activity coefficient γ and distribution ratio (or K-value) in binary fluid mixtures are derived from the assumptions that the pressure errors are negligible, and that the vapor and liquid composition measurements from which they are derived are normally distributed with known mean and variance. The distribution of K is shown to be not normal, with deviations from normality becoming particularly pronounced at the extremes of the composition range. The dispersion of this distribution is considerably greater than that predicted by the classic error-propagation formula. Deviation from normality invalidates the most important assumption implicit in the use of the least-squares criterion as the basis for parameter estimation from activity-coefficient data. This has important implications for the estimation of Henry's law constants for solutes in very dilute solutions, as well as for Raoult's law-based infinite-dilution activity coefficients. The use of the exact distributions of γ and K as the basis for a more rigorous maximum-likelihood method for parameter estimation in excess Gibbs energy models based on either Henry's law or Raoult's law standard states is demonstrated.

Introduction

The principle of maximum likelihood (ML) has found extensive application to parameter-estimation problems arising in chemical engineering—in particular, to the reduction of vapor−liquid equilibrium (VLE) data and the cognate problem of testing the thermodynamic consistency of such measurements. The starting point in all calculations based on this principle is the calculation of a joint probability density function for the measurements, in which the mean value is assumed to be predicted by the functional relationship postulated to exist between the measurements; the parameters are then adjusted, by some numerical optimization technique, to maximize the value of the likelihood function (or more usually, its logarithm). The goal in VLE data reduction is to describe the thermodynamic properties of fluid mixtures in terms of empirical or semiempirical representations of the excess molar Gibbs energy as a function of the composition. The parameters in such model equations need to be estimated to optimize the agreement between the measured and calculated phase compositions at equilibrium. Application of the ML principle to this parameter-estimation problem can proceed in two ways.

The most common procedure involves determination of the activity coefficients γ_1, γ_2 from the measured phase compositions, and fit of these quantities to the composition dependence predicted by the postulated representation of the excess Gibbs energy in the liquid phase. For this purpose, the dependent variables can be set equal to the activity coefficients themselves (Fabries and Renon, 1975; Gmehling et al., 1977; Van Ness et al., 1978), the excess molar Gibbs energy $g^E/RT = x_1 \ln \gamma_1 + x_1 \ln \gamma_1$ (Van Ness et al., 1973; Byers et al., 1973; Abbott and Van Ness, 1975), or $\ln(\gamma_1/\gamma_2)$ and the pressure p (Kemény et al., 1982). Kemény et al. (1982) considered a number of other possible choices for the dependent variables. Maximization of the log-likelihood function in this approach is equivalent to weighted nonlinear least-squares (WNLS). This involves minimization of a quadratic form in the residuals (Aitken, 1935), in which the weight given to each possible product of residuals at a particular design point is an element of the reciprocal variance−covariance matrix for the responses. If there is only one response, or if correlations between the responses are neglected, this quadratic form reduces to a weighted sum of squared residuals. The basic problem that arises in all WNLS is that of estimating the variance–covariance matrix. In the absence of replicate measurements, the covariances are estimated by means of an error-propagation formula based on a Taylor-series expansion, application of which to activity-coefficient data was first described by Ulrichson and Stevenson (1972). WNLS treatments can be classified according to a hierarchy of approximations made in the treatment of the covariances. The most rigorous treatments allow for both heteroscedasticity and correlated responses, and require estimates of the covariance matrix at each design point. Less rigorous treatments assume the covariance matrix to be the same for all data; the simplest approximation (the diagonal uniform weighting approximation) assumes that the weighting matrices are all proportional to the identity matrix. As discussed by Bard (1974) and Bates and Watts (1988, Chap. 4), application of ML theory in situations where the covariances are unknown requires minimization of the determinant of the moment matrix of residuals (Box and Draper, 1965).

The alternative approach, which has been adopted much less widely, is to fit the calculated pressure-composition curves directly to the data, without prior calculation of the activity coefficients at each experimental composition. Here, the ML method is equivalent to minimization of a weighted sum of squared residuals for all the data, subject to vectors of functional constraints containing the activity coefficients derived from the excess Gibbs energy model. Such generalized nonlinear least squares (GNLS) calculations can be carried out by use of Gauss-Newton algorithms developed by Britt and Luecke (1973) and Jefferys (1980) (BLJ) for a single vector of functional constraints, and by Anderson et al. (1978) for two vectors of constraints. These "errors-in-variables" or "random-regressor" approaches rely on fewer assumptions concerning the covariances of the variables, but suffer from the disadvantage of being complex and requiring good initial parameter estimates for convergence.

A previous article compared the performance of this GNLS approach to binary VLE data reduction with the multiresponse WNLS fit of activity coefficients and minimization of the moment-matrix determinant (which is equivalent to performing WNLS with unknown weights). The calculations described therein illustrated the important influence of the weighting matrix in WNLS, and drew attention to the lack of a completely satisfactory way of estimating this matrix, particularly at the extremes of the composition ranges, where errors in composition data can be expected to exert a pronounced effect on extrapolated (infinite-dilution) activity-coefficient values and the integral test of thermodynamic consistency. The latter conclusion suggested that it might be profitable to determine the statistical distribution of the activity coefficients estimated from VLE data, and to use this distribution to set up an alternative ML treatment of VLE-derived activity coefficients that does not make use of a least-squares criterion and is in principle unaffected by the lack of knowledge of the covariances between the responses.

An overview of the present article is as follows. In the next section, the two fundamental results relating the statistical distribution of the product and quotient of two random variables to their joint probability density are introduced. In the third section of the article these two results are used to derive the sampling distribution of K and γ for a solute in the Henry's law limit (where the activity of the other component —the solvent—is negligibly different from unity), and where errors in the pressure can be neglected. In the fourth section, the joint distribution is derived for the K values of both components, and used as the basis for the maximum-likelihood fit of the activity coefficients to a commonly used empirical representation of excess Gibbs energy.

Product and Quotient of Two Random Variables

The distribution of the product and quotient of two random variables is a topic that is not usually covered in elementary treatises on statistics, but the specific problem of determining the distribution of the quotient of two normally distributed random variables has been considered by several authors (Geary, 1930; Fieller, 1932; Hinkley, 1969; Shanmugalingam, 1982), mainly in connection with problems in biology and agriculture. The best way of deriving the general expressions for these distributions is by determining the appropriate cumulative probability distributions and differentiating them with respect to the variable. Details of this procedure are given by Mood et al. (1974, pp. 187–188). The results can be stated thus: If the joint probability distribution of two random variables U and V is $f_{UV}(u,v)$ (where U, V can assume all positive and negative values), the probability densities of the product $P = UV$ and quotient $Q = U/V$ are

$$f_P(p) = \int_{-\infty}^{\infty} \frac{1}{|u|} f_{UV}\left(u, \frac{p}{u}\right)\, du = \int_{-\infty}^{\infty} \frac{1}{|v|} f_{UV}\left(\frac{p}{v}, v\right)\, dv \quad (1)$$

and

$$f_Q(q) = \int_{-\infty}^{\infty} |u| f_{UV}(u, qu)\, du. \quad (2)$$

Henry's Law Limit: Distribution of *K* Value

The purpose of this section is to establish the distribution of the quantity

$$K_1 \equiv \gamma_1 \frac{p_1^0}{p} = \frac{Y_1}{X_1}, \quad (3)$$

which describes the vapor–liquid partitioning of component 1 of a binary mixture under conditions where vapor-phase nonideality and Poynting corrections can be ignored, and the liquid-phase mole fraction X_1 is small enough that the activity of the solvent (component 2) is negligibly different from unity. In terms of Eq. 2, this means formally identifying U with Y_1, V with X_1, and Q with K_1, where the former two variables are independently distributed, with densities $g(X_1)$ and $h(Y_1)$ defined within the square $\{(X_1,Y_1) | 0 \le X_1 \le 1, 0 \le Y_1 \le 1\}$. This restriction, and the fact that K can be either less than or greater than unity, makes some modifications of Eq. 2 necessary. As shown in Appendix A, the resulting distribution possesses the general form

$$f_K(K) = \begin{cases} \int_0^1 X g(X) h(KX)\, dX & K \le 1 \\ (1/K^2) \int_0^1 Y h(Y) g(Y/K)\, dY & K > 1. \end{cases} \quad (4)$$

To proceed further, it is necessary to specify the individual distribution functions g and h appearing in the preceding formula. Since the mole fractions contain random measurement errors, it is natural to choose normal distributions for each. But the use of such distributions is, strictly speaking, fundamentally incompatible with the restriction of X and Y to values between 0 and 1. This incompatibility can, however, be removed by introducing a truncated normal distribution, which is of similar analytical form, but normalized to the finite interval [0,1]. For example, assuming that $g(X)$ possesses a mean X_0 and standard deviation σ_X, the truncated normal distribution is

$$g(X) = C_X e^{-((X-X_0)/\sigma_X\sqrt{2})^2}, \tag{5}$$

where the normalization constant is given by

$$C_X^{-1} = \int_0^1 e^{-((X-X_0)/\sigma_X\sqrt{2})^2} dX = \frac{\sigma_X\sqrt{2\pi}}{2}\left[\operatorname{erf}\left(\frac{1-X_0}{\sigma_X\sqrt{2}}\right) + \operatorname{erf}\left(\frac{X_0}{\sigma_X\sqrt{2}}\right)\right],$$

and

$$\operatorname{erf}(x) \equiv \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2} du = 1 - \operatorname{erfc}(x)$$

is the standard error function. If X_0 is many (at least 6) standard deviations from 0 or 1, the error functions approach unity and the usual form of the normal distribution (for which $C_X = 1/\sigma_X\sqrt{2\pi}$) is recovered. Similarly

$$h(Y) = C_Y e^{-((Y-Y_0)/\sigma_Y\sqrt{2})^2}, \tag{6}$$

where

$$C_Y^{-1} = \int_0^1 e^{-((Y-Y_0)/\sigma_Y\sqrt{2})^2} dY = \frac{\sigma_Y\sqrt{2\pi}}{2}\left[\operatorname{erf}\left(\frac{1-Y_0}{\sigma_Y\sqrt{2}}\right) + \operatorname{erf}\left(\frac{Y_0}{\sigma_Y\sqrt{2}}\right)\right],$$

As shown in Appendix B, the density function can be written in the form

$$\begin{aligned} f_K(K) = C_X C_Y\Bigg[& 2\sigma_X^2\left\{G\left(A_x^2+1, 2A_xB_x, B_x^2; \frac{1-X_0}{\sigma_x\sqrt{2}}\right) - G\left(A_x^2+1, -2A_xB_x, B_x^2; \frac{X_0}{\sigma_x\sqrt{2}}\right)\right\} \\ & + \sigma_X\sqrt{2}\left\{F\left(A_x^2+1, 2A_xB_x, B_x^2; \frac{1-X_0}{\sigma_x\sqrt{2}}\right) + F\left(A_x^2+1, -2A_xB_x, B_x^2; \frac{X_0}{\sigma_x\sqrt{2}}\right)\right\}\Bigg], \end{aligned} \tag{7}$$

where $A_x \equiv K\sigma_X/\sigma_Y$ and $B_x \equiv (KX_0 - Y_0)/\sigma_Y\sqrt{2}$, and the functions F and G are defined by

$$F(a,b,c;x) = \int_0^x e^{-(ax^2+bx+c)} dx \quad \text{and} \quad G(a,b,c;x) = \int_0^x x e^{-(ax^2+bx+c)} dx \tag{8}$$

From the perspective of data analysis, the quantities of most interest are the mean and standard deviation of this distribution. The mean is most directly obtained from the integral formulas for $f_K(K)$ by applying the definition of an expectation value

$$\langle K\rangle \equiv \int_0^\infty K f_K(K)\ dK = \int_0^1 K\left[\int_0^1 X_g(X)h(KX)\ dX\right]dK + \int_1^\infty K\left[(1/K^2)\int_0^1 Yh(Y)g(Y/K)\ dY\right]dK \tag{9}$$

or, interchanging the order of integration

$$\langle K\rangle = \int_0^1 Xg(X)\left[\int_0^1 Kh(KX)\ dK\right]dX + \int_0^1 Yh(Y)\left[\int_1^\infty (1/K)g(Y/K)\ dK\right]dY \tag{10}$$

When $g(X)$ and $h(Y)$ are replaced by the truncated normal distributions, the second integral over K is

$$\int_1^\infty \frac{1}{K} C_Y e^{-((Y/K-Y_0)/\sigma_Y\sqrt{2})^2} dK = \int_0^1 \frac{1}{v} C_Y e^{-((Yv-Y_0)/\sigma_Y\sqrt{2})^2} dv \tag{11}$$

which clearly does not exist. Thus, the mean of the distribution of K values cannot be rigorously defined, and neither can the standard deviation σ_K, in view of the definition

$$\sigma_K^2 = \int_0^\infty [K - \langle K\rangle]^2 f_K(K)\ dK. \tag{12}$$

The counterpart of Eq. 4 for the distribution of the quotient of two random variables distributed over $(-\infty,\infty)$ is

$$f_K(K) = \int_{-\infty}^\infty |X| g(X)h(KX)\ dX, \tag{13}$$

and if g and h are normal distributions, the expectation value

$$\int_0^\infty K f_K(K)\ dK = \frac{1}{2\pi\sigma_X\sigma_Y}\int_{-\infty}^\infty |X| e^{-((X-X_0)/\sigma_X\sqrt{2})^2} \times \left\{\int_0^\infty K e^{-((KX-Y_0)/\sigma_Y\sqrt{2})^2} dK\right\} dX \tag{14}$$

is also infinite, since

$$\int_0^{\infty} e^{-((KX-Y_0)/\sigma_Y\sqrt{2})^2} dK = \frac{Y_0\sigma_Y}{X^2}\sqrt{\frac{\pi}{2}}\left[1+\operatorname{erf}\left(\frac{Y_0}{\sigma_Y\sqrt{2}}\right)\right] + \left(\frac{\sigma_Y}{X}\right)^2 e^{-(Y_0/\sigma_Y\sqrt{2})^2}. \quad (15)$$

The ratio distribution assumes a particularly simple form when both means are equal to zero, Thus, following Geary (1930)

$$f_K(K) = \frac{1}{2\pi\sigma_X\sigma_Y}\int_{-\infty}^{\infty} |X| e^{-(X/\sigma_X\sqrt{2})^2} e^{-(KX/\sigma_Y\sqrt{2})^2} dX = \frac{\sigma_X}{\pi\sigma_Y}\frac{1}{1+\left(\dfrac{K\sigma_X}{\sigma_Y}\right)^2} \quad (16)$$

which is known as the Cauchy distribution; that the mean and variance of this distribution are undefined can be seen more clearly than in the general case where the means are unequal to zero. The divergence of the mean and standard deviation can be understood as a consequence of the finite probability that X is zero and Y/X is not defined. But in many cases of interest, this probability is not significant. Thus, although a divergent "Cauchy component" of $f_K(K)$ is always present, it becomes vanishingly small if the mean of the X distribution is many standard deviations different from zero. Under these circumstances the lower limit of the integral defining $f_K(K)$ can be replaced, with negligible error, by a positive number X_L, and the distribution becomes

$$f_K(K) \approx \frac{1}{2\pi\sigma_X\sigma_Y}\int_{X_L}^{\infty} X e^{-(X/\sigma_X\sqrt{2})^2} e^{-(KX/\sigma_Y\sqrt{2})^2} dX$$

$$= \frac{\sigma_X}{\pi\sigma_Y}\int_{X_L}^{\infty} u e^{-u^2-(A_x u+B_x)^2} du + \frac{X_0}{\pi\sigma_Y\sqrt{2}}\int_{X_L}^{\infty} e^{-u^2-(A_x u+B_x)^2} du \quad (17)$$

where the required integrals can be expressed in terms of the results

$$\int_x^{\infty} e^{-(au^2+bu+c)} du = \frac{1}{2}\sqrt{\frac{\pi}{a}}\, e^{b^2/4a-c}\operatorname{erfc}\left(\frac{ax+b}{2\sqrt{a}}\right) \quad (18)$$

and

$$\int_x^{\infty} u e^{-(au^2+bu+c)} du = \frac{e^{-(ax^2+bx+c)}}{2a} - \frac{b}{4a}\sqrt{\frac{\pi}{a}}\, e^{b^2/4a-c}\operatorname{erfc}\left(\frac{ax+b}{2\sqrt{a}}\right) \quad (19)$$

Application of the classic error-propagation formula to K gives

$$\delta K = K[|\delta X|/X + |\delta Y|/Y] \quad (20)$$

Table 1. Vapor–Liquid Equilibrium in the Bromine(1)–Carbon Tetrachloride(2) System at 298 K

p_1^*/torr	X_1^*	Y_1^*	K^{**}	$\delta K^{\dagger}$
1.52	0.00420	0.00200	0.47619	0.13039
1.60	0.00420	0.00211	0.50238	0.13101
2.39	0.00599	0.00314	0.52421	0.09222
4.27	0.01020	0.00562	0.55098	0.05442
5.43	0.01300	0.00714	0.54923	0.04269
9.57	0.02360	0.01259	0.53347	0.02345
9.83	0.02380	0.01293	0.54328	0.02329
10.27	0.02500	0.01351	0.54040	0.02216

*Measurements by Lewis and Storch (1917).

**$K^0 = Y_1/X_1$.

†$\delta K^0 = K^0\left[\dfrac{\delta X}{X} + \dfrac{\delta Y}{Y}\right]$.

from which it can be concluded that the contribution of errors in X to the errors in the partition ratio will become especially significant when the values of X,Y are comparable to their uncertainties $\delta X, \delta Y$. It is of interest to examine this conclusion in the light of the previous analysis, with the identifications $\delta X \equiv \sigma_X$ and $\delta Y \equiv \sigma_Y$. For this purpose, use can be made of the classic measurements by Lewis and Storch (1917) of vapor–liquid equilibrium in the bromine(1)–carbon tetrachloride(2) system, for bromine mole fractions that are sufficiently small that Henry's law clearly applies. The original data consisted of partial pressures of bromine above the solutions, but these can be readily converted into vapor-phase mole fractions by dividing by atmospheric pressure. The data are reproduced in Table 1, together with the ideal values $K^0 = Y^0/X^0$ and the uncertainties calculated from the error-propagation formula. The K-dependence of the distribution function can be plotted by generating values of K according to $K = \alpha Y^0/X^0$, where α is a dimensionless scale factor that varies between 0 and 2, and Y^0, X^0 are taken from the first, third, and fifth lines of Table 1. These plots are shown in Figure 1, for which $\sigma_X = 0.0001$ and $\sigma_Y = 0.0005$.

The width at half-height of the first curve is $\delta\alpha \approx 0.3$, corresponding to $\delta K = 0.3/K^0 = 0.6$. This is more than twice as large as the error estimate given in the last column of Table 1. It therefore appears that the error-propagation formula produces overly optimistic estimates of errors in K. The curve corresponding to the highest concentration in Table 1 is not easily shown on the same chart: it has a width of about 0.1 with a maximum value of about 20.

As shown by Lewis and Storch (1917), the dependence of bromine partial pressure on liquid-phase mole fraction is very well represented by a straight-line plot with zero intercept. Although this confirms the validity of Henry's law, random measurement errors in the mole fractions result in slopes (Henry's law constants) varying between 0.47619 and 0.55098. For each pair of data Y^0, X^0 a distribution curve can be drawn, with a maximum very close to the ideal value Y^0/X^0, but since these maxima (or most probable values) are in general different, the best one can hope for is a compromise value that will maximize the joint probability—the likelihood function—for the whole data set. Such a compromise is exactly analogous to that implied by the minimization of squared residuals in conventional data reduction: the only difference is that in the present situation, it is not necessary to assume

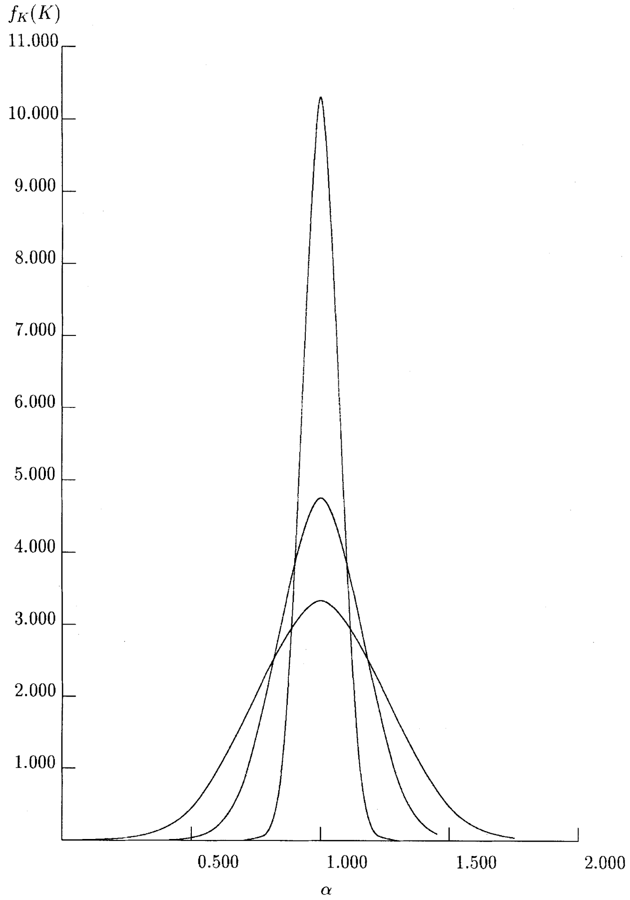

Figure 1. Distribution function calculated for $X_0 = 0.00394$, Y_0 (curve 1), 0.003 (curve 2), and 0.005 (curve 3); $\sigma_X = 0.0002$, $\sigma_Y = 0.0005$.

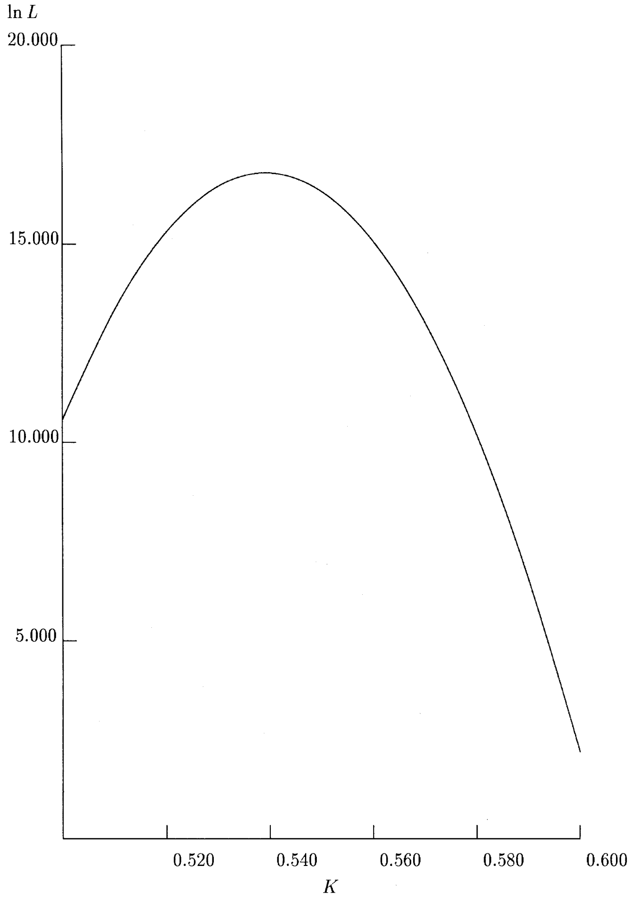

Figure 2. Log-likelihood function calculated for the data of Table 1, as a function of *K*.

that the variables are normally distributed. But since, as shown by Figure 1, the distribution curves become narrower and taller as X_0 increases (or, more precisely, as X_0/σ_X increases), the compromise value is biased toward those corresponding to the smallest relative errors in X_0. This effect is also exactly analogous to weighted least-squares fitting, if the weights are set equal to the reciprocal variances of the data; the difference is that for the normal distributions implied in the least-squares approaches, the width of the distribution curve at each datum can be more simply related to the variance. With these observations in mind, the likelihood function can be formed by calculating the product of f_K for all data and an assumed value of K: the result is the curve shown in Figure 2, with a maximum at 0.53927_5, which is closer to the higher-concentration data. The excellent agreement of this ML estimate with the value of 0.539 given by Lewis and Storch (1917) provides indirect confirmation that the assignment $\sigma_X = 0.0001$ and $\sigma_Y = 0.0005$ was reasonable.

Extension of the preceding analysis to include the effect of errors in the pressure on the distribution of the solute activity coefficient is possible if the vapor-phase fugacity coefficients and Poynting corrections are neglected. Thus, assuming that K and the scaled pressure $p^* \equiv p/p_1^0$ are distributed independently with densities given by Eq. 4 and

$$f_{P^*}(p^*) = C_{P^*}\exp\left[-\left(\frac{p^* - p^{0*}}{\sigma_{p^*}\sqrt{2}}\right)^2\right], \tag{21}$$

where the normalization constant

$$C_{P^*}^{-1} = \sigma_{p^*}\sqrt{\frac{\pi}{2}}\left[1+\mathrm{erf}\left(\frac{p^{0*}}{\sigma_{p^*}\sqrt{2}}\right)\right]$$

is negligibly different from $\sigma_{p^*}\sqrt{2\pi}$ under most conditions, the joint probability density for the two variables is $f_{P^*}(p^*)f_K(K)$ and the distribution of the activity coefficient follows by application of the product formula, Eq. 1

$$f_\Gamma(\gamma) = \int_0^\infty \frac{1}{p^*} f_{P^*}(p^*) f_K\left(\frac{\gamma}{p^*}\right)\, dp^* \tag{22}$$

Evaluation of this integral by numerical quadratures is straightforward, since the density function for p^* becomes vanishingly small outside a narrow range, and in the limit as $\sigma_{p^*} \to 0$ becomes like a Dirac delta function. The use of a scaled pressure variable in this equation is appropriate if the Henry's law constant or the vapor pressure of the pure solute is known. This is clearly not the case in the correlation of gas-solubility data, which are used to *determine* Henry's law constants; there, the Henry's law constant is best included as an adjustable parameter in the activity-coefficient model applied to the liquid (as in the earlier treatment of the Br_2-CCl_4 system). But a more important problem that arises is that it is no longer reasonable to neglect the fugacity coefficients and Poynting corrections. These quantities give rise to a complicated nonlinear dependence of the activity coefficient on pressure and composition, for which the distribution function cannot be analytically determined.

Another problem to which such an approach is applicable is the analysis of isopiestic equilibration experiments, which are an important source of thermodynamic data for concentrated electrolyte solutions. If the molalities of unknown and reference solutions are denoted by m and m_r, respectively, application of the isofugacity condition to the common solvent results in the equation (Rard and Platford, 1991)

$$\nu m \phi = \nu_r m_r \phi_r \tag{23}$$

where ϕ, ϕ_r and ν, ν_r are the molal osmotic coefficients and ionic dissociation numbers of unknown and reference electrolytes, respectively. This equation can be rewritten as

$$\phi = \phi_r K \tag{24}$$

where $K = \nu_r m/(\nu m_r)$. Since ν_r/ν is just a constant, the distribution in the values of the osmotic coefficient can be related to the individual distributions of ϕ_r and K by

$$f_\Phi(\phi) = \int_0^\infty \frac{1}{\phi_r} f_{\Phi_r}(\phi_r) f_K\left(\frac{\phi}{\phi_r}\right) d\phi_r \tag{25}$$

The distribution of K is given by equations of very similar form to those described earlier, the most important difference being in the normalization of the molality distribution functions to $[0,\infty)$ rather than $[0,1]$ (which, as observed earlier for the pressure, is unlikely to have any significant practical effect). Knowledge of the uncertainty in the osmotic coefficient is clearly of importance in the fit of such data to analytical expressions (such as the Pitzer equations). In isopiestic experiments, the reference osmotic coefficients are generally obtained either by interpolation from tabulated, smoothed values, or preferably by use of an analytical formula. If the latter were arrived at by a least-squares calculation, the calculated values of ϕ_r could be assigned a standard deviation equal to the standard deviation of the residuals of the fit. The application of the ratio-distribution theory to the analysis of isopiestic data would evidently require not only the characterization of the precision of the experimental molalities [for example, along the lines described by Hefter et al. (1998)], but also a careful weighted nonlinear least-squares analysis of the selected reference data. Both these points require specialized considerations that are beyond the scope of the present discussion.

Raoult's Law Limit: Joint Distribution of K_1 and $K2$

In this section attention is devoted to the more general situation in which the activities of both components are different from unity, and the deviations from ideality are to be characterized in terms of activity coefficients based on a Raoult's law standard state. In this case it is necessary to admit the possibility that the density functions for the distribution ratios of each component might overlap, so that they would no longer be independently distributed. Such a phenomenon is particularly likely at the extremes of the composition range. Thus, from Eq. 20 it is evident that as the mole fraction of the one component tends to unity, the errors in the distribution coefficient of that component will be much smaller than those in the distribution coefficient of the other component. But it is also true from the preceding analysis that the dispersion in the density function for the K-value of the minor component becomes very large, resulting in the possibility of substantial overlap between the two distributions. To examine this possibility in more detail it is therefore desirable to consider the joint distribution function of $K_1 \equiv Y_1/X_1$ and $K_2 \equiv (1-Y_1)/(1-X_1)$. If values of Y_1 and X_1 are each described by truncated normal distributions with respective means and standard deviations X_1^0, σ_X, and Y_1^0, σ_Y, the joint probability distribution is

$$f_{XY}(X_1,Y_1) = C_X e^{-1/2((X_1-X_1^0)/\sigma_X)^2} C_Y e^{-1/2((Y_1-Y_1^0)/\sigma_Y)^2}. \tag{26}$$

The distribution function for the K-values can be derived from the preceding two equations by application of the transformation formula for multivariate distribution functions (Anderson, 1958, pp. 10–11)

$$f_{K_1K_2}(K_1,K_2) = f_{XY}(X_1,Y_1)\left|\frac{\partial(X_1,Y_1)}{\partial(K_1,K_2)}\right|. \tag{27}$$

As shown in Appendix C, the result of applying this formula is

$$f_{K_1K_2}(K_1,K_2) = f_{XY}(X_1,Y_1)\begin{cases} \dfrac{(1-K_2)(K_1-1)}{(K_1-K_2)^3} & (K_1,K_2)\in S_1 \\ \dfrac{(1-K_1)(K_2-1)}{(K_2-K_1)^3} & (K_1,K_2)\in S_2, \end{cases} \tag{28}$$

where X_1,Y_1 are to be obtained from the inverse of the transformation, and permissible values of K_1,K_2 lie in the semiinfinite strips $S_1 = \{(K_1,K_2)|K_1 \geq 1, 0 \leq K_2 \leq 1\}$, and $S_2 = \{(K_1,K_2)|K_2 \geq 1, 0 \leq K_1 \leq 1\}$. The lack of resemblance between this equation and that derived in the preceding section for a single solute is a bit perplexing, but this disappears when the marginal distributions are calculated. As shown in Appendix D, these are

$$f_{K_1}(K_1) = C_X C_Y \int_0^{1/\max(1,K_1)} \exp\left[-\left(\frac{X_1-X_1^0}{\sigma_X\sqrt{2}}\right)^2\right] \times \exp\left[-\left(\frac{K_1(X_1-X_1^0)+K_1X_1^0-Y_1^0}{\sigma_Y\sqrt{2}}\right)^2\right] X_1 dX_1 \tag{29}$$

Table 2. Vapor–Liquid Equilibrium in the Bromine(1)–Carbon Tetrachloride(2) System at 273 K

p*/torr	X_1**	Y_1†	γ††	γ‡
69.56	0.9194	0.901524	1.004529	2.510712
70.07	0.9120	0.895390	1.013164	2.460722
70.86	0.8567	0.856054	1.042807	2.102787
71.18	0.8523	0.850660	1.046290	2.126151
67.94	0.6628	0.752281	1.135674	1.474477
63.11	0.4938	0.678815	1.277699	1.182968
64.19	0.4900	0.680324	1.312554	1.188635
60.95	0.3859	0.600984	1.397952	1.169946
60.70	0.3822	0.601318	1.406476	1.157201
53.84	0.2482	0.501486	1.602110	1.054682
54.18	0.2453	0.497970	1.619849	1.064721
45.13	0.1178	0.318192	1.795310	1.030390
39.55	0.0545	0.180531	1.929443	1.012649
37.71	0.0479	0.168656	1.955473	0.972739
36.24	0.0220	0.078918	1.914580	1.008298
34.41	0.00957	0.036617	1.939049	0.988783
32.66	0.00479	0.017915	1.798666	0.951975
34.02	0.001582	0.005879	1.861889	1.000697
33.97	0.000936	0.003533	1.888146	1.000937
33.89	0.000762	0.002773	1.816783	0.999284
33.89	0.000762	0.002773	1.816783	0.999284

*Sum of partial pressures measured by Barthel and Dodé (1954).
**Liquid-phase mole fraction measured by Barthel and Dodé (1954).
†Vapor-phase mole fractions calculated from partial pressures.
††Activity coefficient $\gamma_1 = Y_1 p/(X_1 p_1^0)$, $p_1^0 = 67.9$ torr.
‡Activity coefficient $\gamma_2 = Y_2 p/(X_2 p_2^0)$, $p_2^0 = 33.85$ torr.

and

$$f_{K_2}(K_2) = C_X C_Y \int_0^{1/\max(K_2,1)} X_2 \exp\left[-\left(\frac{X_2 - X_2^0}{\sigma_X\sqrt{2}}\right)^2\right] \times \exp\left[-\left(\frac{K_2(X_2 - X_2^0) + K_2 X_2^0 - Y_2^0}{\sigma_Y\sqrt{2}}\right)^2\right] dX_2 \quad (30)$$

which can in turn be expressed in terms of the same functions used in Eq. 7; the difference is in the upper limits of the integrals.

ML fit of Raoult's law–based activity coefficient data requires use of the joint density function. This can be illustrated by application to further vapor–liquid equilibrium measurements by Barthel and Dodé (1954) on the Br_2–CCl_4 system. These authors extended the measurements by Lewis and Storch (1917) to a wider concentration range. The data are listed, in slightly modified form, in Table 2; they are of particular interest in the present context, because they contain measurements for several very low bromine mole fractions, at which statistical dispersion effects can be expected to be significant. Barthel and Dodé (1954) represented the concentration-dependence of the bromine activity coefficient by the equation

$$\ln\gamma_1 = \alpha(1 - X_1)^2 + \beta(1 - X_1)^3 \quad (31)$$

and found numerical values $\alpha = 1.197$ and $\beta = -0.493$; they did not explicitly consider the carbon tetrachloride activity coefficient. In terms of the more commonly used Redlich-Kister functions, for which

$$\ln\gamma_1 = (A + 3B)(1 - X_1)^2 - 4B(1 - X_1)^3 \quad (32)$$

$$\ln\gamma_2 = (A - 3B)(1 - X_1)^2 + 4B(1 - X_1)^3 \quad (33)$$

the corresponding values of the constants A and B easily can be shown to be $A = \alpha + 3\beta/4 = 0.82725$ and $B = -\beta/4 = 0.12325$. The usual data-reduction method of plotting g^E/RTX_1X_2 against X_1 (Smith and Van Ness, 1987, pp. 350–355) gives very poor results, especially toward the extremes of the concentration range, where small errors in the activity coefficients are greatly magnified by dividing by the mole fractions. For example, at the point with $X_1 = 0.00479$, $g^E/RTX_1X_2 = -9.68$, which is vastly different from the other values. A more satisfactory estimate of these parameters can be obtained from a simultaneous fit of both activity coefficients to the preceding equations, by minimization of the moment-matrix determinant, which requires the least restrictive assumptions on the form of the covariance matrix. Application of the conjugate-gradient minimization method gives $A = 0.9075870$, $B = 0.2699164$ after three iterations and a residual standard deviation of 0.0480, which is a considerable improvement on the 0.1132 obtained with $A = 0.82725$ and $B = 0.12325$. The discrepancy between the two estimates also illustrates the importance of using both activity coefficients for parameter estimation, rather than just one.

The log-likelihood function consists of the logarithm of the joint probability density, evaluated at each pair of data (X_1^0, Y_1^0) with $K_1 = \gamma_1 p_1^0/p$ and $K_2 = \gamma_2 p_2^0/p$, and summed over the entire data set. Consider, for example, the contribution of the first point, $p = 69.56$ torr $= 9{,}274$ Pa, $X_1^0 = 0.9194$, $Y_1^0 = 0.901524$. Assuming $\sigma_X = 0.0001$ and $\sigma_Y = 0.0005$, the error functions in the normalization constants are essentially equal to 1, so that $C_X = 1/\sigma_X\sqrt{2\pi} = 0.3989423E+04$ and $C_Y = 1/\sigma_Y\sqrt{2\pi} = 0.7978846E+03$. At ($A = 0.9075914$, $B = 0.2699164$) (which is the first point required in the numerical approximation of the gradient with respect to the parameters), the Redlich-Kister functions give

$$\gamma_1 = 1.0106474 \Rightarrow K_1 = \gamma_1 p_1^0/p = 0.9865290$$

$$\gamma_2 = 2.5137721 \Rightarrow K_2 = \gamma_2 p_2^0/p = 1.2232775$$

These numbers are to be compared with the experimental partition ratios, which are

$$K_1^0 = Y_1^0/X_1^0 = 0.980557$$

$$K_2^0 = \left(1 - Y_1^0\right)/\left(1 - X_1^0\right) = 1.221787$$

The agreement between these pairs of K-values is apparently quite good—to within about 1%—but the effect of the discrepancy between them is greatly magnified by the small value of σ_X. Thus, the values of X and Y corresponding to these calculated K-values are found from the inverse transformation to be $X = 0.9431001$, $Y = 0.9303957$ (see Appendix C), so the arguments of the exponentials are the negative squares of

$$\frac{X_1 - X_1^0}{\sigma_X\sqrt{2}} = \frac{0.9431001 - 0.9194000}{0.0001 \times 1.4142136} = 167.585298$$

$$\frac{Y_1 - Y_1^0}{\sigma_Y\sqrt{2}} = \frac{0.9303957 - 0.9015240}{0.0005 \times 1.4142136} = 40.8307210$$

These give extremely small numbers upon exponentiation. Finally, since the preexponential factor is

$$C_X C_Y = \frac{(1 - K_1)(K_2 - 1)}{(K_2 - K_1)^3} = 0.7214927E + 06$$

the contribution to $-\ln L$ is

$$(167.585298)^2 + (40.8307210)^2 - \ln(0.7214927E + 06) = 0.2973849E + 05.$$

Proceeding similarly for all the other points and adding the results leads to the value $-\ln L = 0.8724649E + 08$.

Conjugate-gradient minimization of $-\ln L$ can fail if the components of the gradient with respect to the parameters are sufficiently large to give rise to extrapolations far outside the physically reasonable regions of parameter space. This problem can be overcome by limiting the values of the scalar parameter used in the line searches along the conjugate directions. A simpler alternative, which was adopted here, involves dividing $-\ln L$ by a constant scaling factor. (Such scaling is, however, unlikely to succeed if the minimum being sought is poorly defined as a result of excessively imprecise data, the existence of multiple minima, of the presence of redundant parameters in the model.) When the objective function was scaled in this way (by a factor of $1E + 09$) the conjugate gradient search converged rapidly: from the initial estimate generated by minimization of the determinant criterion, it led in three iterations to the point (0.9347971, 0.3419962), at which $-\ln L = 0.7468119E + 05$. Since the determinant criterion is based on the assumption that the errors in the data are described by a bivariate normal distribution with unknown covariance matrix, the discrepancy between the initial and final parameter estimates illustrates the importance of deviations from normality.

The activity coefficients are shown in Figure 3, together with the curves fitted by the determinant method and the ML method. The curve fitted by the determinant method passes approximately through the middle of the cluster of experimental points at the lowest values of X_1, while the ML curve misses nearly all of these points. This seems reasonable; in the determinant method, no attempt is made to limit the influence of the dilute-solution data by assignment of multiplicative weights. A striking consequence of this is the difference in the extrapolated activity coefficients at infinite dilution, particularly for γ_2: 3.6 as compared with about 3.2. This corresponds to a difference of about 0.3 kJ mol^{-1} in the estimate of the excess chemical potential of component 2 at infinite dilution. Finally, in the fit of nonlinear models by *any* method, it is necessary to admit the possibility that multiple local minima could exist in parameter space. This seems to be precluded by the simple analytical form of the function considered here, but is likely to arise with multiparameter functions containing terms that almost vanish for some values of the variables.

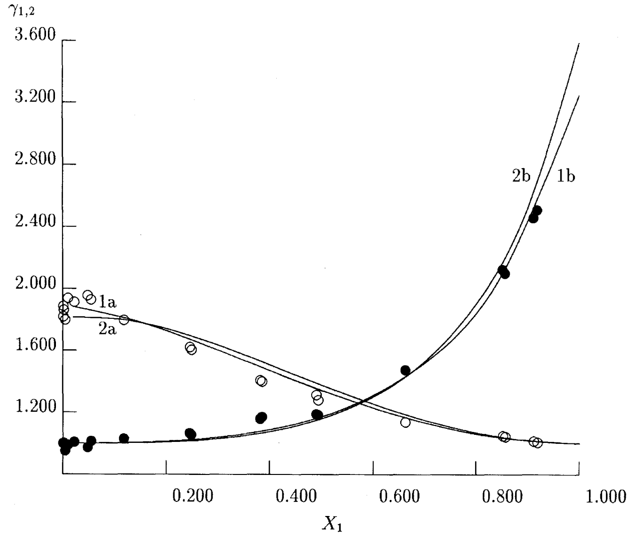

Figure 3. Fit of activity coefficients from Table 2.

Open circles: γ_1; filled circles: γ_2. Curve 1a: determinant fit of γ_1; curve 1b: determinant fit of γ_2; curve 2a: ML fit of γ_1; curve 2b: ML fit of γ_2.

Discussion

The calculations presented in the preceding section demonstrated that the exact distribution function of the K-values can be used as an alternative basis for ML parameter estimation. The purpose of this section is to consider various aspects of the relationship between this new approach and conventional parameter estimation methods.

In least-squares methods, it can be analytically proved that the contours of the likelihood function for a two-parameter model in the vicinity of the minimum are ellipses, with aspect ratios that can be related to the standard errors in the parameter estimates. In contrast, the analytical form of the likelihood contours in the present method is less easily established, but the curve shown in Figure 2 for the one-dimensional model seems qualitatively consistent with what one would obtain from a least-squares calculation. But considering that the likelihood contour shapes reflect the combined effects of the precision of the data, the number of points, and the sensitivity of the model to the parameter values and the variables, they seem to be of limited value in illustrating the distinction between the present approach and conventional WNLS or GNLS. Of far more practical importance in this regard is the manner in which the influence of the dilute-solution VLE data on parameter estimates is taken into account. When the determinant criterion is used, no differential weighting is applied to the data; in WNLS, deemphasis of the dilute-solution data is achieved through assumed values of the variances of the activity-coefficient data, while in the present approach the same effect is achieved more rigorously by specifying the precision of the individual vapor and liquid compositions. The effects of these distinctions on the extrapolation behavior of the fitted model are quite striking, as illustrated in Figure 3.

Since the present approach involves no assumption that the errors in the independent variable (liquid-phase composition) can be neglected, it invites comparison with the GNLS treatment described by Anderson et al. (1978), and the multiequation generalization of the BLJ method described in a previous article. Although the present analysis follows a closely similar argument and is founded on statistical assumptions that are essentially equivalent, it cannot be regarded as an errors-in-variables model *in sensu stricto*, because the mean values used in setting up the log-likelihood function are fixed equal to the experimental data. In this regard it parallels more closely the VLE data-reduction scheme proposed by Kemény et al. (1982), which is based on the "effective variance" approximation. The conversion of the theory presented in this article into an errors-in-variables method for the analysis of ratio data would require the addition of a scheme for updating the regressors at each iteration. It is not immediately obvious how this could be done.

One limitation already identified is the inability to include the effects of errors in pressure on the distribution of activity coefficients, due to the nonlinearities associated with fugacity coefficients and Poynting corrections. These effects are much better treated by direct GNLS analysis of VLE data. Considering that in low-pressure VLE data these effects are often small and largely cancel each other (Smith and Van Ness, 1987, p. 347), this limitation is unlikely to be particularly serious in many practical situations. In any event, the ML method should provide good initial parameter estimates for GNLS calculations. The two approaches are therefore best regarded as complementary rather than as mutually exclusive alternatives. Parameter estimation involving the least-squares analysis of activity coefficients has the advantage of being able to reveal systematic trends indicative of incompatibility between the excess Gibbs energy model and the data. The same observation applies *a fortiori* to the present theory: by eliminating the need for the generally unknown covariance matrix at each point, the amount of useful information that can be extracted from the composition dependence of activity coefficients has been considerably increased—particularly for data containing appreciable scatter. A particularly important attribute of the present ML method is that it provides a theoretically valid basis for extrapolation of partition ratios to very low concentrations, where methods based on correlation of g^E/RTX_1X_2 can fail badly.

The possible application to analysis of gas-solubility data and isopiestic equilibration experiments has already been mentioned. A further practical problem where such extrapolations are of importance is in the vapor−liquid partitioning of electrolytes from aqueous solutions at high temperatures (Marshall et al., 2000; Palmer et al., 1999a,b). This has been recognized for many years as an important cause of corrosion in the steam cycles of fossil-fuel and nuclear power stations. Traces of electrolyte impurities that are completely dissociated under ambient conditions become strongly associated as neutral molecules or ion pairs at high temperatures (as the dielectric constant of the water solvent decreases), and therefore susceptible to volatilization. Measurements of such vapor−liquid transfer are inherently difficult, and rely on the determination of extremely small concentrations—measurements of the equilibrium partitioning quotients contain considerable scatter. It is usually required to extract from such data an estimate of the inherent volatility of the electrolyte (in effect, a molal Henry's law constant) by adopting a suitable activity-coefficient model for the aqueous phase and extrapolating to infinite dilution. The theory presented here permits such extrapolations to be made much more rigorously than would be possible with the conventional unweighted least-squares techniques that are typically applied to such data.

Conclusions

The problem of fitting experimental partition ratio data to theoretical models has been solved by a ML parameter estimation procedure, based on the exact distribution function for the ratio of two random variables described by truncated normal distributions. For the case of a single solute obeying Henry's law, this distribution function can be expressed in terms of error-function integrals. The most important advantage of the proposed procedure in comparison with conventional least-squares fitting procedures is that knowledge of the mean and variance is not required: these are undefined, as is also true for the well-known case of the ratio of two normally distributed random variables. The joint distribution function for the vapor−liquid partition ratios of each component in a binary mixture can be derived by application of the change-of-variable rule for multivariate distribution functions. ML fits of activity coefficients based on this joint distribution function allow experimental uncertainties to be more rigorously taken into account in the estimation of infinite-dilution activity coefficients. Possible applications to the analysis of isopiestic experiments and modeling of VLE in electrolytes at high temperatures have been identified.

Acknowledgments

Funding for the research was provided by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences, U.S. Department of Energy under Contract DEAC05-00OR22725, Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC. I express my sincere thanks to Dr. J. G. Blencoe for many helpful discussions, and to Dr. M. R. Ally for useful suggestions.

Literature Cited

Abbott, M. M., and H. C. Van Ness, "Vapor-Liquid Equilibrium: Part III. Data Reduction with Precise Expressions for g^E," *AIChE J.*, **21**, 65 (1975).

Aitken, A. C., "On Least Squares and Combinations of Observations," *Proc. Roy. Soc. Edinburgh*, **55**, 42 (1935).

Anderson, T. F., D. S. Abrams, and E. A. Grens II, "Evaluation of Parameters for Nonlinear Thermodynamic Models," *AIChE J.*, **24**, 20 (1978).

Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, Wiley, New York (1958).

Bard, Y., *Nonlinear Parameter Estimation*, Academic Press, New York (1974).

Barthel, C., and M. Dodé, "Etude du Système Brome-Tétrachlorure de Carbon au Moyen de Radiobrome 82," *Bull. Soc. Chim. Fr.*, 1312 (1954).

Bates, D. M., and D. G. Watts, *Nonlinear Regression Analysis and its Applications*, Wiley, New York (1988).

Box, G. E. P., and N. R. Draper, "The Bayesian Estimation of Common Parameters from Several Responses," *Biometrika*, **52**, 355 (1965).

Britt, H. I., and R. H. Luecke, "The Estimation of Parameters in Nonlinear, Implicit Models," *Technometrics*, **15**, 233 (1973).

Byers, S. M., R. E. Gibbs, and H. C. Van Ness, "Vapor-Liquid Equilibrium: Part II. Correlations from p-x Data for 15 Systems," *AIChE J.*, **19**, 245 (1973).

Fabries, J., and H. Renon, "Methods of Evaluation and Reduction of Vapor-Liquid Equilibrium Data of Binary Mixtures," *AIChE J.*, **21**, 735 (1975).

Fieller, E. C., "The Distribution of the Index in a Normal Bivariate Population," *Biometrika*, **24**, 428 (1932).

Geary, R. C., "The Frequency Distribution of the Quotient of Two Normal Variates," *J. Roy. Stat. Soc.*, **93**, 442 (1930).

Gmehling, J., U. Onken, W. Arlt, and U. Weidlich, *Vapor-Liquid Equilibrium Data Collection*, Vol. 1, Part 1, DECHEMA, Frankfurt am Main, Germany (1977).

Hefter, G., P. M. May, S. L. Marshall, J. Cornish, and I. Kron, "Improved Apparatus and Procedures for Isopiestic Measurements at Elevated Temperatures," *Rev. Sci. Instrum.*, **68**, 2558 (1998).

Hinkley, D. V., "On the Ratio of Two Correlated Normal Random Variables," *Biometrika*, **56**(3), 635 (1969).

Jefferys, W. H., "On the Method of Least Squares," *Astrophys. J.*, **85**, 177 (1980).

Kemény, S., J. Manczinger, S. Skjold-Jørgensen, and M. Tóth, "Reduction of Thermodynamic Data by Means of the Multiresponse Maximum Likelihood Principle," *AIChE J.*, **28**, 20 (1982).

Lewis, G. N., and H. Storch, "The Potential of the Bromine Electrode; the Free Energy of Dilution of Hydrogen Bromide; the Distribution of Bromine Between Several Phases," *J. Amer. Chem. Soc.*, **39**, 2544 (1917).

Marshall, S. L., D. A. Palmer, J. M. Simonson, and D. B. Joyce, *Volatility of Aqueous Acetic Acid, Formic Acid, and Sodium Acetate*, EPRI Rep. TR-113089, Electric Power Research Institute, Palo Alto, CA (2000).

Mood, A. M., F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*, McGraw-Hill, New York (1974).

Palmer, D. A., J. M. Simonson, R. W. Carter, J. P. Jensen, D. B. Joyce, and S. L. Marshall, *Volatility of Aqueous Sodium Hydroxide, Bisulfate and Sulfate*, EPRI Rep. TR-105801, Electric Power Research Institute, Palo Alto, CA (1999a).

Palmer, D. A., J. M. Simonson, R. W. Carter, J. P. Jensen, and S. L. Marshall, *Vapor-Liquid Partitioning of Sulfuric Acid and Ammonium Sulfate*, EPRI Rep. TR-112359, Electric Power Research Institute, Palo Alto, CA (1999b).

Rard, J. A., and R. F. Platford, "Experimental Methods: Isopiestic," *Activity Coefficients in Electrolyte Solutions*, 2nd ed., K. S. Pitzer, ed., CRC Press, Boca Raton, FL, p. 209 (1991).

Shanmugalingam, S., "On the Analysis of the Ratio of Two Correlated Normal Variables," *Statistician*, **31**, 251 (1982).

Smith, J. M., and H. C. Van Ness, *Introduction to Chemical Engineering Thermodynamics*, 4th ed., Wiley, New York (1987).

Ulrichson, D. L., and F. D. Stevenson, "Effects of Experimental Errors on Thermodynamic Consistency and on Representation of Vapor-Liquid Equilibrium Data," *Ind. Eng. Chem. Fundam.*, **11**, 287 (1972).

Van Ness, H. C., S. M. Byers, and R. E. Gibbs, "Vapor-Liquid Equilibrium: Part I. An Appraisal of Data Reduction Methods," *AIChE J.*, **19**, 238 (1973).

Van Ness, H. C., F. Pedersen, and P. Rasmussen, "Vapor-Liquid Equilibrium: Part V. Data Reduction by Maximum Likelihood," *AIChE J.*, **21**, 1055 (1978).

Appendix A: Derivation of Equation 4

The required distribution is derived by considering the probability $\text{Prob}(Y/X \le K)$, where the subscript indicating component 1 has been suppressed for clarity. Since this is the same as $\text{Prob}(X/Y \ge 1/K) = 1 - \text{Prob}(X/Y \le 1/K)$, K can be assumed without loss of generality to be less than 1. Then it follows that

$$\text{Prob}(Y/X \le K) = \int_{X=0}^{1} g(X)\left[\int_{Y=0}^{KX} h(Y)dY\right]dX \tag{A1}$$

the region over which the integral is to be evaluated is the triangle with vertices at (0,0), $(1,K)$, and (1,0). The corresponding probability density is obtained from this expression as the limit of

$$\frac{\text{Prob}(Y/X \le K + \delta K) - \text{Prob}(Y/X \le K)}{\delta K} = \frac{\text{Prob}(K \le Y/X \le K + \delta K)}{\delta K} \tag{A2}$$

as δK tends to 0. Thus

$$\begin{aligned}\text{Prob}(K \le Y/X \le K + \delta K) &= \int_{X=0}^{1} g(X)\left[\int_{Y=KX}^{(K+\delta K)X} h(Y)\ dY\right]dX \\ &\approx \int_{X=0}^{1} g(X)h(KX)X\ \delta K\, dX\end{aligned} \tag{A3}$$

which, after substitution into the previous equation, gives the probability density

$$f_K(K) = \int_0^1 Xg(X)h(KX)\ dX \tag{A4}$$

The form of the distribution for $K > 1$ can be obtained from the cumulative probability $\text{Prob}(X/Y \le 1/K)$, which by analogous reasoning is

$$\text{Prob}(X/Y \le 1/K) = \int_0^1 h(Y)\left[\int_0^{Y/K} g(X)\ dX\right]\ dY \tag{A5}$$

$$\text{Prob}(X/Y \le 1/(K + \delta K)) = \int_0^1 h(Y)\left[\int_0^{Y/(K+\delta K)} g(X)\ dX\right]\ dY \tag{A6}$$

The increment in cumulative probability corresponding to Eq. 6 is

$$\begin{aligned}\text{Prob}(1/(K+\delta K) \le X/Y \le K) &= \int_0^1 \int_{Y/K}^{Y/(K+\delta K)} h(Y)g(X)\ dX\ dY \\ &\approx \int_0^1 h(Y)g(Y/K)Y\left[\frac{1}{K+\delta K} - \frac{1}{K}\right]dY \\ &\approx -\int_0^1 Yh(Y)g(Y/K)\frac{\delta K}{K^2}\ dY.\end{aligned} \tag{A7}$$

Forming the difference quotient and taking the limit as $\delta K \to 0$, the required probability density is found to be

$$\frac{d\ \text{Prob}(X/Y \ge K)}{dK} = -\frac{d\ \text{Prob}(X/Y \le K)}{dK} = \frac{1}{K^2}\int_0^1 Yh(Y)g(Y/K)\ dY \tag{A8}$$

When $K = 1$, the two forms are equal, since

$$\int_0^1 Yh(Y)g(Y)\ dY = \int_0^1 Xh(X)g(X)\ dX. \tag{A9}$$

Therefore, combining both cases,

$$f_K(K) = \begin{cases} \int_0^1 Xg(X)h(KX)\ dX & K \le 1 \\ (1/K^2)\int_0^1 Yh(Y)g(Y/K)\ dY & K > 1. \end{cases} \tag{A10}$$

Appendix B: Evaluation of the Distribution Function of *K*

To evaluate

$$f_K(K) = C_X C_Y \int_0^1 X e^{-((X-X_0)/\sigma_X\sqrt{2})^2} e^{-((KX-Y_0)/\sigma_Y\sqrt{2})^2} dX, \tag{B1}$$

apply the substitution $u = (X - X_0)/\sigma_X\sqrt{2}$. This results in

$$\begin{aligned} \sqrt{2}C_X C_Y \sigma_X &\int_{X_0/\sigma_X\sqrt{2}}^{(1-X_0)/\sigma_X\sqrt{2}} \left(\sigma_X\sqrt{2}u + X_0\right) e^{-u^2-(A_x u + B_x)^2} du \\ &= C_X C_Y \Bigg[2\sigma_X^2 \int_{X_0/\sigma_X\sqrt{2}}^{(1-X_0)/\sigma_X\sqrt{2}} u e^{-u^2-(A_x u + B_x)^2} du \\ &\quad + \sigma_X\sqrt{2}u \int_{X_0/\sigma_X\sqrt{2}}^{(1-X_0)/\sigma_X\sqrt{2}} e^{-u^2-(A_x u + B_x)^2} du \Bigg] \end{aligned} \tag{B2}$$

where $A_x \equiv K\sigma_X/\sigma_Y$ and $B_x \equiv (KX_0 - Y_0)/\sigma_Y\sqrt{2}$. Each integral can be further subdivided:

$$\begin{aligned} \int_{X_0/\sigma_X\sqrt{2}}^{(1-X_0)/\sigma_X\sqrt{2}} u e^{-u^2-(A_x u + B_x)^2} du &= \int_0^{(1-X_0)/\sigma_X\sqrt{2}} u e^{-u^2-(A_x u + B_x)^2} du \\ &\quad - \int_0^{X_0/\sigma_X\sqrt{2}} u e^{-u^2-(A_x u - B_x)^2} du; \end{aligned} \tag{B3}$$

$$\begin{aligned} &\int_{X_0/\sigma_X\sqrt{2}}^{(1-X_0)/\sigma_X\sqrt{2}} e^{-u^2-(A_x u + B_x)^2} du \\ &= \int_0^{(1-X_0)/\sigma_X\sqrt{2}} e^{-u^2-(A_x u + B_x)^2} du + \int_0^{X_0/\sigma_X\sqrt{2}} e^{-u^2-(A_x u - B_x)^2} du. \end{aligned} \tag{B4}$$

The integrals required here are of the form

$$F(a,b,c;x) = \int_0^x e^{-(ax^2+bx+c)}dx \text{ and } G(a,b,c;x) = \int_0^x x e^{-(ax^2+bx+c)}dx, \tag{B5}$$

where $a \equiv A_x^2 + 1$, $b \equiv \pm 2A_x B_x$, and $c \equiv B_x^2$. Although neither of them is expressible in terms of elementary functions, they can both easily be reduced to error function integrals. Thus, by completeing the square in the quadratic argument of the exponential and applying elementary transformations,

$$F(a,b,c;x) = \frac{1}{2}\sqrt{\frac{\pi}{a}}\, e^{b^2/4a-c}\left[\operatorname{erf}\left(\frac{b}{2\sqrt{a}}\right) + \operatorname{erf}\left(\frac{ax+b}{2\sqrt{a}}\right)\right], \tag{B6}$$

and since

$$\int_0^x x e^{-(ax^2+bx+c)}dx = \frac{1}{2a}\left[\int_0^x (2ax+b)e^{-(ax^2+bx+c)}dx - b\int_0^x e^{-(ax^2+bx+c)}dx\right], \tag{B7}$$

it follows that

$$G(a,b,c;x) = \frac{e^{-c} - e^{-(ax^2+bx+c)}}{2a} - \frac{b}{4a}\sqrt{\frac{\pi}{a}}\, e^{b^2/4a-c}\left[\operatorname{erf}\left(\frac{b}{2\sqrt{a}}\right) + \operatorname{erf}\left(\frac{ax+b}{2\sqrt{a}}\right)\right]. \tag{B8}$$

Appendix C: Application of the Transformation Formula

Application of the transformation formula requires the elements of the appropriate Jacobian matrix. Since

$$\frac{\partial K_1}{\partial X_1} = -\frac{K_1}{X_1} \quad \frac{\partial K_1}{\partial Y_1} = \frac{K_1}{Y_1} \quad \frac{\partial K_2}{\partial X_1} = \frac{K_2}{1-X_1} \quad \frac{\partial K_2}{\partial Y_1} = -\frac{K_2}{1-Y_1} \tag{C1}$$

and

$$\left|\frac{\partial(K_1,K_2)}{\partial(X_1,Y_1)}\right| = \begin{vmatrix} \dfrac{\partial K_1}{\partial X_1} & \dfrac{\partial K_1}{\partial Y_1} \\ \dfrac{\partial K_2}{\partial X_1} & \dfrac{\partial K_2}{\partial Y_1} \end{vmatrix} = K_1 K_2\left[\frac{1}{X_1(1-Y_1)} - \frac{1}{Y_1(1-X_1)}\right], \tag{C2}$$

the Jacobian determinant appearing in Eq. 41 is

$$\left|\frac{\partial(K_1,K_2)}{\partial(X_1,Y_1)}\right| = \left|\frac{\partial(X_1,Y_1)}{\partial(K_1,K_2)}\right|^{-1} = \left\{K_1 K_2\left[\frac{1}{X_1(1-Y_1)} - \frac{1}{Y_1(1-X_1)}\right]\right\}^{-1}. \tag{C3}$$

From the inverse of the transformation, expressed by formulas giving X_1,Y_1 in terms of K_1,K_2, namely

$$X_1 = \frac{1-K_2}{K_1-K_2} \qquad Y_1 = K_1\frac{1-K_2}{K_1-K_2}, \tag{C4}$$

it is clear that not all values of K_1, K_2 will result in physically acceptable values of X_1, Y_1: obviously X_1 cannot be negative or greater than unity. The nonnegativity condition on X_1 is satisfied either if $K_2 < 1$ and $K_1 > K_2$, or if $K_2 > 1$ and $K_2 > K_1$. In the first case, the smallest value $X_1 = 0$ corresponds to $K_2 = 1$, and the largest value $X_1 = 1$ is obtained when $K_2 = 0$ and $K_1 = 1$. Proceeding similarly for the other case, it therefore follows that permissible values of K_1, K_2 lie in the semiinfinite strips $S_1 = \{(K_1,K_2)|K_1 \geq 1, 0 \leq K_2 \leq 1\}$, and $S_2 = \{(K_1,K_2)|K_2 \geq 1, 0 \leq K_1 \leq 1\}$. With these conditions established, the determinant in the distribution function can be rewritten entirely in terms of K_2, K_1 by substitution of Eqs. C4 into Eq. C3. After simplification

$$\left\{K_1K_2\left[\frac{1}{X_1(1-Y_1)} - \frac{1}{Y_1(1-X_1)}\right]\right\}^{-1} = \begin{cases} \dfrac{(1-K_2)(K_1-1)}{(K_1-K_2)^3} & (K_1,K_2)\in S_1 \\ \dfrac{(1-K_1)(K_2-1)}{(K_2-K_1)^3} & (K_1,K_2)\in S_2. \end{cases} \tag{C5}$$

The required joint distribution function of K_1 and K_2 is therefore

$$f_{K_1K_2}(K_1,K_2) = f_{XY}(X_1,Y_1)\begin{cases} \dfrac{(1-K_2)(K_1-1)}{(K_1-K_2)^3} & (K_1,K_2)\in S_1 \\ \dfrac{(1-K_1)(K_2-1)}{(K_2-K_1)^3} & (K_1,K_2)\in S_2, \end{cases} \tag{C6}$$

where X_1, Y_1 are to be obtained from Eq. C4.

Appendix D: Marginal Distribution Functions of K_1 and K_2

The marginal distribution of K_1 is obtained by holding K_1 constant and integrating over all permissible values of K_2. In view of the preceding discussion, this integral possesses two different forms. For $(K_1,K_2)\in S_2$ (that is, $K_1 < 1$)

$$\begin{aligned} f_{K_1}(K_1) &= \int_1^\infty f_{XY}(X_1,Y_1)\frac{(1-K_1)(K_2-1)}{(K_2-K_1)^3}\,dK_2 \\ &= \int_1^\infty C_X \exp\left(-\frac{1}{2\sigma_X^2}\left[\frac{1-K_2}{K_1-K_2} - X_1^0\right]^2\right) C_Y \exp\left(-\frac{1}{2\sigma_Y^2}\left[K_1\frac{1-K_2}{K_1-K_2} - Y_1^0\right]^2\right) \\ &\qquad \times \frac{(1-K_1)(K_2-1)}{(K_2-K_1)^3}\,dK_2 \end{aligned} \tag{D1}$$

while for $(K_1, K_2)\in S_1$ (that is, $K_1 > 1$)

$$\begin{aligned} f_{K_1}(K_1) &= \int_0^1 f_{XY}(X_1,Y_1)\frac{(1-K_2)(K_1-1)}{(K_1-K_2)^3}\,dK_2 \\ &= \int_0^1 C_X \exp\left(-\frac{1}{2\sigma_X^2}\left[\frac{1-K_2}{K_1-K_2} - X_1^0\right]^2\right) C_Y \exp\left(-\frac{1}{2\sigma_Y^2}\left[K_1\frac{1-K_2}{K_1-K_2} - Y_1^0\right]^2\right) \\ &\qquad \times \frac{(1-K_2)(K_1-1)}{(K_1-K_2)^3}\,dK_2 \end{aligned} \tag{D2}$$

Proceeding similarly, the marginal distribution of K_2 is

$$\begin{aligned} f_{K_2}(K_2) &= \int_1^\infty f_{XY}(X_1,Y_1)\frac{(1-K_2)(K_1-1)}{(K_1-K_2)^3}\,dK_1 \\ &= \int_1^\infty C_X \exp\left(-\frac{1}{2\sigma_X^2}\left[\frac{1-K_2}{K_1-K_2} - X_1^0\right]^2\right) C_Y \exp\left(-\frac{1}{2\sigma_Y^2}\left[K_1\frac{1-K_2}{K_1-K_2} - Y_1^0\right]^2\right) \\ &\qquad \times \frac{(1-K_2)(K_1-1)}{(K_1-K_2)^3}\,dK_1 \end{aligned} \tag{D3}$$

for $(K_1,K_2)\in S_1$ (that is, $K_2 < 1$), and

$$\begin{aligned} f_{K_2}(K_2) &= \int_0^1 f_{XY}(X_1,Y_1)\frac{(1-K_1)(K_2-1)}{(K_2-K_1)^3}\,dK_1 \\ &= \int_0^1 C_X \exp\left(-\frac{1}{2\sigma_X^2}\left[\frac{1-K_2}{K_1-K_2} - X_1^0\right]^2\right) C_Y \exp\left(-\frac{1}{2\sigma_Y^2}\left[K_1\frac{1-K_2}{K_1-K_2} - Y_1^0\right]^2\right) \\ &\qquad \times \frac{(1-K_1)(K_2-1)}{(K_2-K_1)^3}\,dK_1 \end{aligned} \tag{D4}$$

for $(K_1,K_2)\in S_2$ (that is, $K_2 > 1$). These integrals can be transformed as follows, by change of variable. Thus, in Eq. D1, substitute

$$X_2 \equiv 1 - X_1 = \frac{1-K_1}{K_2-K_1} \qquad dX_2 = -\frac{X_2^2}{1-K_1}\,dK_2 \tag{D5}$$

and observe that

$$\frac{(1-K_1)(K_2-1)}{(K_2-K_1)^3}dK_2=\frac{K_2-1}{K_2-K_1}\left(\frac{1-K_1}{K_2-K_1}\right)^2\frac{dK_2}{1-K_1}$$
$$=-X_1dX_2=X_1dX_1. \tag{D6}$$

Then, since $K_2=1\Rightarrow X_1=0$ and $K_2=\infty\Rightarrow X_1=1$, the integral transforms to

$$f_{K_1}(K_1)=C_XC_Y\int_0^1\exp\left[-\left(\frac{X_1-X_1^0}{\sigma_X\sqrt{2}}\right)^2\right]$$
$$\times\exp\left[-\left(\frac{K_1(X_1-X_1^0)+K_1X_1^0-Y_1^0}{\sigma_Y\sqrt{2}}\right)^2\right]X_1\ dX_1, \tag{D7}$$

for $K_1<1$. In Eq. D2, the substitution

$$X_2\equiv 1-X_1=\frac{K_1-1}{K_1-K_2}\quad dX_2=\frac{X_2^2}{K_1-1}dK_2 \tag{D8}$$

similarly leads to

$$\frac{(1-K_2)(K_1-1)}{(K_1-K_2)^3}dK_2=X_1dX_2=-X_1dX_1, \tag{D9}$$

Since $K_2=0\Rightarrow X_1=1/K_1$ and $K_2=1\Rightarrow X_1=0$, the integral transforms to

$$f_{K_1}(K_1)=C_XC_Y\int_0^{1/K_1}\exp\left[-\left(\frac{X_1-X_1^0}{\sigma_X\sqrt{2}}\right)^2\right]$$
$$\times\exp\left[-\left(\frac{K_1(X_1-X_1^0)+K_1X_1^0-Y_1^0}{\sigma_Y\sqrt{2}}\right)^2\right]X_1\ dX_1, \tag{D10}$$

for $K_1>1$. These two results can be combined into one formula by writing

$$f_{K_1}(K_1)=C_XC_Y\int_0^{1/\max(1,K_1)}\exp\left[-\left(\frac{X_1-X_1^0}{\sigma_X\sqrt{2}}\right)^2\right]$$
$$\times\exp\left[-\left(\frac{K_1(X_1-X_1^0)+K_1X_1^0-Y_1^0}{\sigma_Y\sqrt{2}}\right)^2\right]X_1\ dX_1. \tag{D11}$$

For $K_2<1$, the substitution of

$$X_1=\frac{1-K_2}{K_1-K_2}\qquad dX_1=-\frac{X_1^2}{1-K_2}dK_1 \tag{D12}$$

into Eq. D3 results in

$$\frac{(1-K_2)(K_1-1)}{(K_1-K_2)^3}dK_1=\frac{K_1-1}{K_1-K_2}\left(\frac{1-K_2}{K_1-K_2}\right)^2\frac{dK_1}{1-K_2}$$
$$=-X_2dX_1. \tag{D13}$$

The extra factor of K_1 in the argument of the second exponential can be disposed of by noting that

$$K_1\frac{1-K_2}{K_1-K_2}=(1-K_2)\left(1+\frac{K_2}{K_1-K_2}\right), \tag{D14}$$

from which follows, after simple algebra,

$$K_1\frac{1-K_2}{K_1-K_2}-Y_1^0=K_2(X_1-X_1^0)+Y_2^0-K_2X_2^0$$
$$=-K_2(X_2-X_2^0)+Y_2^0-K_2X_2^0. \tag{D15}$$

Since $K_1=1\Rightarrow X_1=1$ and $K_1=\infty\Rightarrow X_1=0$, the transformed integral is

$$f_{K_2}(K_2)=C_XC_Y\int_0^1X_2\exp\left[-\left(\frac{X_1-X_1^0}{\sigma_X\sqrt{2}}\right)^2\right]$$
$$\times\exp\left[-\left(\frac{K_2(X_2-X_2^0)+K_2X_2^0-Y_2^0}{\sigma_Y\sqrt{2}}\right)^2\right]dX_1. \tag{D16}$$

Finally, for $K_2>1$, the substitution of

$$X_1=\frac{K_2-1}{K_2-K_1}\qquad dX_1=\frac{X_1^2}{K_2-1}dK_1 \tag{D17}$$

into Eq. D4 results in

$$\frac{(1-K_1)(K_2-1)}{(K_2-K_1)^3}dK_1=X_2dX_1, \tag{D18}$$

and since $K_1=0\Rightarrow X_1=1-1/K_2$ and $K_1=1\Rightarrow X_1=1$, the transformed integral is

$$f_{K_2}(K_2)=C_XC_Y\int_{1-1/K_2}^1X_2\exp\left[-\left(\frac{X_1-X_1^0}{\sigma_X\sqrt{2}}\right)^2\right]$$
$$\times\exp\left[-\left(\frac{K_2(X_2-X_2^0)+K_2X_2^0-Y_2^0}{\sigma_Y\sqrt{2}}\right)^2\right]dX_1. \tag{D19}$$

Combining Eqs. D16 and D17

$$f_{K_2}(K_2)=C_XC_Y\int_{1-1/\max(K_2,1)}^1X_2\exp\left[-\left(\frac{X_1-X_1^0}{\sigma_X\sqrt{2}}\right)^2\right]$$
$$\times\exp\left[-\left(\frac{K_2(X_2-X_2^0)+K_2X_2^0-Y_2^0}{\sigma_Y\sqrt{2}}\right)^2\right]dX_1 \tag{D20}$$

or, with the trivial change in variable from X_1 to X_2

$$f_{K_2}(K_2) = C_X C_Y \int_0^{1/\max(K_2,1)} X_2 \exp\left[-\left(\frac{X_2 - X_2^0}{\sigma_X\sqrt{2}}\right)^2\right] \times \exp\left[-\left(\frac{K_2(X_2 - X_2^0) + K_2 X_2^0 - Y_2^0}{\sigma_Y\sqrt{2}}\right)^2\right] dX_2 \quad \text{(D21)}$$

Comparison of Eqs. D11 and D21 shows that the marginal density functions of K_1 and K_2 are not only the same form, but are also closely similar in form to the expression derived in Appendix B for the case of a single solute in the Henry's law limit. This is indeed an attractive situation, since it means that the same computer codes can be applied to both the Henry's law and Raoult's law cases.

Manuscript received Sept. 13, 2002, and revision received Mar. 19, 2003.