Journal of Chromatography A, 1218 (2011) 7841–7849

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Computation of distribution of minimum resolution for log-normal distribution of chromatographic peak heights

Joe M. Davis*

Department of Chemistry and Biochemistry, Southern Illinois University at Carbondale, Carbondale, IL 62901-4409, USA

ARTICLE INFO

Article history:
Received 5 June 2011
Received in revised form 21 August 2011
Accepted 22 August 2011
Available online 3 September 2011

Keywords:
Resolution distribution
Average minimum resolution
Peak-height distribution
Log-normal distribution
Statistical-overlap theory

ABSTRACT

General equations are derived for the distribution of minimum resolution between two chromatographic peaks, when peak heights in a multi-component chromatogram follow a continuous statistical distribution. The derivation draws on published theory by relating the area under the distribution of minimum resolution to the area under the distribution of the ratio of peak heights, which in turn is derived from the peak-height distribution. Two procedures are proposed for the equations' numerical solution. The procedures are applied to the log-normal distribution, which recently was reported to describe the distribution of component concentrations in three complex natural mixtures. For published statistical parameters of these mixtures, the distribution of minimum resolution is similar to that for the commonly assumed exponential distribution of peak heights used in statistical-overlap theory. However, these two distributions of minimum resolution can differ markedly, depending on the scale parameter of the log-normal distribution. Theory for the computation of the distribution of minimum resolution is extended to other cases of interest. With the log-normal distribution of peak heights as an example, the distribution of minimum resolution is computed when small peaks are lost due to noise or detection limits, and when the height of at least one peak is less than an upper limit. The distribution of minimum resolution shifts slightly to lower resolution values in the first case and to markedly larger resolution values in the second one. The theory and numerical procedure are confirmed by Monte Carlo simulation.

1. Introduction

In a recent paper, Enke and Nagels presented evidence that the component concentrations in natural mixtures of metabolytes, light crude oil, and plant extracts follow a log-normal distribution (LND) [1]. By determining parameters of the LND for these mixtures, they were able to estimate the total number of components that are potentially detectable, predict the degree of analytical selectivity and dynamic range needed to detect any additional fraction of undetected components, and investigate the relationship between undetected components, and background levels and chemical noise.

The findings of Enke and Nagels differ from those in several earlier studies, which suggested on theoretical [2,3] and experimental [2,3–7] grounds that component concentrations in complex mixtures follow an exponential distribution. The LND predicts that the component density (i.e., the number of components per unit of concentration) maximizes at an intermediate concentration and approaches zero as the concentration approaches zero. In contrast, the exponential distribution predicts that the component density maximizes as the concentration approaches zero. Because of the possible overlap of signals that one must interpret to evaluate the component density, the low-concentration region can be difficult to measure. Thus, it is possible that previous work based on an exponential distribution is biased.

In general, the distribution of component concentrations determines the distribution of their signal intensities, since concentration and signal are related by the response factor. In chromatography and related methods, the signal intensities are peak heights (or peak areas, which are proportional to peak heights), whose distribution in turn affects the distribution of the minimum resolution needed to separate neighboring peaks. The average of this distribution of minimum resolution is an important metric of statistical-overlap theory (SOT), which relates the number of observed peaks (e.g., maxima) and the number of peaks produced by mixture components. Most calculations in SOT are based on the assumption that peaks heights follow an exponential distribution.

In this paper, general equations are derived for the distribution of minimum resolution, based on a theory of Felinger [8]. Two procedures are proposed for their numerical solution. With these procedures, the distribution of minimum resolution is computed numerically for a LND of peak heights. This distribution then is compared to that for exponentially distributed peak heights, and

* Permanent address: 733 Schloss Street, Wrightsville Beach, NC 28480, USA. Tel.: +1 910 256 4235.
E-mail address: chimicajmd@ec.rr.com

0021-9673/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2011.08.078

the bias in earlier SOT calculations is evaluated. The theory for the distribution of minimum resolution also is expanded to other problems of interest, using the LND as an example. Specifically, the distribution of minimum resolution is computed when small peaks are lost due to noise or detection limits, and when the height of at least one peak is less than an upper limit.

2. Theory

2.1. Review

Consider a multi-component separation in which the heights (or areas) of peaks follow a statistical distribution. Although the heights actually are governed by deterministic attributes (e.g., metabolic processes, biodegradation, etc.), their adherence to a statistical distribution allows them to be treated as a random variable. Ideally, the response factor is the same for all mixture components, such that the distributions of peak height and component concentration are the same. In fact, the response factors of components typically vary, and the two distributions differ.

The theory of the distribution of minimum resolution (or simply resolution distribution) was developed by Felinger [8]. He derived the equation (cited below) relating the ratio of the heights of two Gaussian peaks of equal standard deviation to the minimum resolution required to produce two maxima. He interpreted this relation as a survival function, whose negative derivative is the resolution distribution for peaks with uniformly distributed heights. With a cumulative distribution derived by Feller [9], Felinger then used the inverse probability integral transform (which maps a uniform distribution into a non-uniform one) to derive the resolution distribution for peaks having an exponential distribution of heights. All of Felinger's results were expressed by closed-form expressions. Here, general equations are proposed for the resolution distribution, and numerical approaches are proposed to solve them. However, the principles underlying this work, which are now reviewed, are those of Felinger.

Panel a of Fig. 1 is a graph of a peak-height distribution $g(h)$ vs. the peak height h. In this graph, $g(h)$ is the LND for parameters specified in the figure caption. However, the underlying principles apply to any continuous $g(h)$; the LND is simply shown as an example. The differential area $g(h)dh$ is the probability that a random peak height lies between h and $h+dh$ (the differential areas of other distributions discussed below can be interpreted similarly). The random selection of two peak heights, h_1 and h_2, from this distribution determines the ratio $r=h_2/h_1$, defined so that $0 \le r \le 1$ (i.e., h_2 always is the smaller peak height, unless $r=1$). This ratio itself is a random variable and has its own distribution $f(r)$, which is called the ratio or quotient distribution and is shown in panel b of Fig. 1 (general aspects of the ratio distribution and other examples of it are discussed in Section 2.1 of the Supplementary material). Since the peak heights determining $f(r)$ in panel b are chosen randomly, they can belong to two adjacent peaks. In this case, the ratio distribution $f(r)$ in panel b is connected to the distribution $w(R_s)$ of minimum resolution R_s of these peaks, which is shown in panel c. In summary, the distributions in panels a–c are connected, because the distribution of peak heights in a chromatogram (panel a) causes adjacent pairs of peaks to have different peak-height ratios (panel b), which in turn causes the minimum resolution needed to separate them to vary (panel c). The connection is expressed more rigorously by Felinger's equation relating the peak-height ratio r to the minimum resolution R_s required to separate two Gaussian peaks with equal standard deviations into two maxima [8]

$$r = \left(\sqrt{4R_s^2-1}+2R_s\right)^2 \exp\left(-4R_s\sqrt{4R_s^2-1}\right) \tag{1}$$

Eq. (1) states that two peaks with a peak-height ratio r produce two maxima, when their resolution equals or exceeds R_s. It is graphed in panel d of Fig. 1. The quantitative connection between panels b and c is that the area under $f(r)$ between 0 and r equals the area under $w(R_s)$ between R_s, as determined by Eq. (1) and r, and infinity. This is true, because whatever number of peak pairs in a separation has peak-height ratios less than r, the same number requires a resolution equaling or exceeding R_s for separation. For example, Eq. (1) predicts that $r=0.2$ when $R_s=0.770$ (i.e., two peaks with a 1:5 ratio of heights require a resolution equaling or exceeding 0.770 to produce two maxima), such that the shaded areas between the r values of 0 and 0.2 in panel b, and between the R_s values of 0.770 and infinity in panel c, are equal. By the same argument, the unshaded (white) areas in panels b and c are equal. These are the areas under $f(r)$ between r and 1, and under the resolution distribution between 0.5 (the smallest possible minimum resolution, appropriate to two peaks of equal height) and R_s.

Therefore, one can calculate the resolution distribution by calculating the area under the ratio distribution $f(r)$. The shaded area under $f(r)$ is given by the cumulative distribution function $F(r)$ [10]

$$F(r) = \int_0^r f(z)\,dz \tag{2a}$$

where z is a dummy variable. The unshaded area under $f(r)$ is given by the survival function $F_c(r)$ (or complementary cumulative distribution function) [10]

$$F_c(r) = 1 - F(r) = \int_r^1 f(z)\,dz \tag{2b}$$

Graphs of $F(r)$ and $F_c(r)$ are shown in panel e of Fig. 1, with $F(r)=0$ and $F_c(r)=1$ at $r=0$, and $F(r)=1$ and $F_c(r)=0$ at $r=1$.

2.2. General equations for resolution distribution

From previous arguments, $F_c(r)$ must equal the area under the resolution distribution $w(R_s)$ between 0.5 and the value of R_s paired with r via Eq. (1). Thus

$$F_c(r) = \int_{0.5}^{R_s} w(z)\,dz \tag{3}$$

where z is again a dummy variable. Eq. (3) is an integral equation with a kernel of unity. On its differentiation with respect to R_s, one obtains

$$w(R_s) = \frac{dF_c(r)}{dR_s} \tag{4}$$

Although $F_c(r)$ in Eq. (4) does not depend explicitly on R_s, it has an implicit dependence expressed by Eq. (1), in which R_s varies with r. This dependence is taken into account by applying the chain rule to Eq. (4)

$$w(R_s) = \frac{dF_c(r)}{dr}\frac{dr}{dR_s} \tag{5a}$$

where

$$\frac{dr}{dR_s} = -8\sqrt{4R_s^2-1}\left[\sqrt{4R_s^2-1}+2R_s\right]^2 \exp\left(-4R_s\sqrt{4R_s^2-1}\right) \tag{5b}$$

as derived from Eq. (1) by Felinger [8].

Eq. (5a) shows the resolution distribution $w(R_s)$ is the product of two derivatives. One of them, $dF_c(r)/dr$, varies with the peak-height distribution. The other, dr/dR_s, is always the same. The validity of Eq. (5a) is suggested by its prediction of two resolution distributions derived by Felinger. For a uniform distribution of peak heights, $F_c(r)=1-r$, as shown in Section 2.2 of the Supplementary material. Thus, $dF_c(r)/dr=-1$, and Eq. (5a) reduces to $w(R_s)=-dr/dR_s$, which is the negative of Eq. (5b) and equal to Felinger's result [8]. For