Polymer Analysis

Chapter 2.   Error Analysis/Statistical Descriptions of Data.

All Polymer Properties are Disperse:

Polymeric materials are subject to dispersion in all analytic properties. For example, the melting point in a low molecular weight organic or inorganic is a fixed value constant which might display some variability over 1 or 2 degrees. The melting point for high-density polyethylene (HDPE), for example can vary from about 110°C to about 160°C depending on processing and generally displays a broad dispersion over about 10 to 20°C. Spectroscopic analysis of polymers using techniques such as infra red adsorption (IR) and nuclear magnetic resonance (NMR) rely mostly on local chemical groups, so can display fairly sharp absorption bands.   In these spectroscopic techniques, dispersion is shown by the existence of peak splittings or the presence of a number of chemical species in small amounts. Additionally, many absorption bands in polymers are difficult to describe analytically and pertain to various acoustic modes associated with the conformation of long chain structures. These are typically broad bands due to wide dispersion in these long chain conformations. X-ray diffraction from polymer crystallites generally displays broad diffraction peaks associated with small crystals and a high degree of disorder within crystallites. Polymers also display dispersion in structural and chemical orientation which should be viewed with statistics. Dispersions in mechanical properties are always seen in polymers. Statistics are also used to describe the dispersion of chain size, molecular weight, and topological arrangement of tacticity. All descriptions of polymer chain size are based on statistics. It is critical to realized at the onset of any analysis of a polymeric system that every physical property will be described by a distribution.

Error Analysis in Analytic Methods:

In the physical sciences each analytic measurement must be associated with an assessment of the confidence which should be associated with the analytic description. A value for some property of a material is of no use unless some estimate of the expected error and distribution is provided.   For example, for a commercial sample of HDPE a technician might report the melting point of a blown film as 135°C. You might be involved in using this material in a packaging application where the material will be subjected to shipping at a maximum temperature of 95°C. *From the reported melting point and operating temperature would you feel confident that this material would meet the specs of the packaging company?* The correct answer is that the measured value is of no use without a description of the statistical distribution in the samples as well as a statistical description of the distribution of melting points, i.e. onset, peak and maximum melting point.

The technician should be required to measure at least 10 samples (preferably more) and calculate a standard deviation for the reported value. Additionally, similar determinations of the onset of melting and the maximum melting point would be needed to determine if this material meets the specifications of the packager.

There are a number of useful texts which describe the correct handling of experimental data. The most commonly cited reference is "Data Reduction and Error Analysis for the Physical Sciences" by P. R. Bevington, 1969 McGraw Hill. This section of the course will summarize some of the major points of Bevington with an emphasis on applications in polymeric systems.

Types of Error:

Statisticians categorize three types of error. This framework may be useful in considering a specific experiment such as determination of the d-spacing from an XRD peak.

Illegitimate Error: These errors involve an operator error, e. g. you have placed the wrong sample in the diffractometer and are determining the d-spacing for the wrong sample.   There is no

statistical description for these errors unless you want to consider sociology. Your main protection against illegitimate errors is to always consider, when faced with extremely unexpected results, that the results involve a human error. Always look carefully at completely unexpected results.

Systemic Error: There errors are also not subject to a statistical analysis. They result from faulty calibration of the instrument or other problems which result in a constant shift of the data. In XRD a systematic error might involve confusion of 2θ with θ leading to roughly a doubling of the d-spacings. Systemic errors can sometimes be corrected after the fact if one is careful. It is important to keep a good record of the analysis that was performed partly for the purpose of correction of systemic errors of this type.

Random Error: Random errors are the main area which statistics can deal with (i.e. standard deviation, mean). Random errors have two sources. 1) Random variability in the samples themselves. 2) Limited precision of the analytic equipment. These two sources can be distinguished if more precise equipment is available or a standard sample is available.

In describing the error involved in an analytic measurement all types of error should be considered.

Accuracy of a measurement pertains to how close a measurement is to the actual value. Precision refers to the reproducibility of the measurement whether or not it is close to the actual value. If systematic error is present a measurement might be extremely precise but completely inaccurate! Also, if the equipment is well calibrated a measurement could be extremely accurate but not particularly precise. Mostly the importance of these terms is in communication of your confidence in a particular value and what the possible sources of error are.

Statistical Analysis of Measurements:

Generally a given measurement will be conducted several times in order to determine the statistical distribution for the measurement. The values most commonly determined are the mean, μ, and the standard deviation, σ:

$$\mu = \frac{1}{N}\sum_{1}^{N} x_i$$

$$\sigma^2 = \frac{1}{N-1}\sum_{1}^{N}(x_i - \mu)^2$$

N is the number of measurements made, the value from each measurement is x_i for measurement i. The square of the standard deviation is called the variance.

Distribution Functions (see P. C. Heimenz, Polymer Chemistry, 1984, Marcel Dekker, pp. 34):

If a measurement is not single valued then it is common to fit the number distribution of values to a distribution function. The simplest distribution functions will involve two parameters, the mean, μ and the standard deviation, σ. More complicated continuous distribution functions will involve higher moments of the distribution. The k'th moment of a distribution is given by:

$$\text{k' th moment} = \sum_{i} f_i (x_i - x_s)^k$$

where f_i is the fraction of all measurements which have the value x_i, i.e. N_i/N, x_s is basis for the moment and k is the order of the moment. For example, the mean, μ, is the first moment (k = 1) about the origin ($x_s = 0$). The variance, σ^2, is the second moment (k = 2) about the mean ($x_s = \mu$).

The molecular weight distributions commonly used in polymer science can be described in terms of the more broadly used statistical description of moments.

The number average molecular weight, M_n is the same as the first moment about the origin or the mean. The weight average molecular weight (mass average), M_w, is given by:

$$M_w = \frac{\sum\limits_i N_i M_i^2}{\sum\limits_i N_i M_i} = \frac{\sum\limits_i f_i M_i^2}{\sum\limits_i f_i M_i}$$

or the ratio of the second to the first moment about the origin.

The polydispersity index, M_w/M_n, is the ratio of the second moment to the square of the first moment about the origin.

Given M_w and M_n the standard deviation of a distribution of molecular weights can be determined:

$$\sigma = M_n \left(\frac{M_w}{M_n} - 1 \right)^{\frac{1}{2}}$$

Several specific distribution functions are mentioned below. A distribution is usually considered unimodal if one of these continuous distribution functions describes the distribution of measured values. If the data is best described by several of these functions it is termed bimodal, trimodal etc. A bimodal distribution in lamellar thicknesses in polyethylene might be generated if crystallization occurred in two distinct steps such as primary crystallization and secondary epitaxial spherulitic decoration for instance. This distribution might be described by two Gaussian functions, so a total of 4 parameters, 2 means and 2 standard deviations.

Binomial Distribution:

Consider a tactic polymer such as polypropylene. As discussed in class, on passing along the main chain of the polymer there is a choice in handedness for the substitutent methyl group which could be considered similar to a coin toss, i.e. if the substitutent occurs with the same handedness as the previous mer unit this would be similar to a coin tossed heads (R for tacticity) and if it occurs with the opposite handedness (S for tacticity) this would be similar to a coin tossed tails. If the tacticity of the polymer is determined at random (atactic) then there is equal probability for R and S handedness. The binomial distribution describes such a situation where the "probability of success" is given by p, (consider R a success, here p = 0.5 for atactic polymers). The probability of observing "n" R handedness mer units in a chain of N mer units when the probability of seeing an R handiness mer unit is p is given by:

$$P_{\text{Binomial}}(N, n, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

the mean, μ, and standard deviation, σ, for the binomial distribution are given by:

$$\mu = Np$$

$$\sigma^2 = Np(1-p)$$

$P_{Binomial}$ could be used to plot a distribution function of tacticities for example. Note that there is a finite probability for a completely R polymer (0! =1).

Poisson Distribution:

When N is large and the mean, μ, is constant with sample size, the binomial distribution simplifies to the Poisson Distribution:

$$P_{Poisson}(n,\mu) = \frac{\mu^n}{n!} e^{-\mu}$$

where the standard deviation is given by:

$$\sigma = \sqrt{\mu}$$

The Poisson Distribution is used to describe small samples of large populations, so is appropriate in analytic techniques which involve counting of events, such as XRD and light counters, and mass spectrometers which measure events.   The standard deviation in such a counting measurement (counting of events) is the square root of the number of counts. For example, an integrated diffraction peak might generate 10,000 counts on a proportional detector. The error in this value is ± 100.   If an analytic instrument does not report "counts" or "events" then the Poission distribution value for the standard deviation can not be used directly.

Gaussian Distribution:

For very large samples, N, with a finite probability of success, p, a smooth distribution is usually observed. Such a distribution can be described by a Gaussian function:

$$P_{Gaussian}(n,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{n-\mu}{\sigma}\right)^2\right)$$

The Gaussian distribution is the basis of the "bell-shaped curve". It is also used to describe random walks in diffusion as well as the path of a polymer coil under theta-conditions. Integration of the Gaussian distribution as a weighting factor for r^2, the square of the chains end-to-end distance yields the mean square end-to-end distance for a "Gaussian" chain, Nl^2, where l is the step size and N is the number of randomly arranged steps in the non-interacting chain. For the Gaussian chain $\mu = 0$ and $\sigma^2 = nl^2$.

Other Distributions:

There are many other distribution functions commonly used in polymer science such as the Lorentzian Distribution (see hand out) and the Maxwellian Distribution (see Polymer Materials Science by J. Schultz for instance).   The definition of these distributions will rely on an understanding of the moments of a continuous distribution described above.

Covariance:

In many analytic experiments several parameters are determined from a single measurement through the use of a fit to experimental data. For instance, a light scattering curve can be used to determine the molecular weight (first moment about the origin) and second viral coefficient, A_2, through the Zimm plot (figure 2.8 in our text). Often the two parameters which are measured are not completely independent in terms of the fit to the data. Under such conditions it is necessary to determine the covariance of the two parameters. The covariance reflects the degree to which two parameters effect each other. If more than two parameters are unknown, then a covariance matrix can be constructed to determine the extent two which any two parameters are associated. For two parameters the covariance is defined by:

$$\sigma_{uv}^2 = \left\langle \left(u - \langle u \rangle \right) \left(\left(v - \langle v \rangle \right) \right) \right\rangle$$

where <> indicates a mean. Covariance is included for completeness here. Determination of the covariance is generally rare in the literature.

Propagation of Errors:

In many analytic techniques a value is measured, an error is determined, and this value is used to calculate a parameter of interest. For example, in the determination of the modulus of a sample the extension of the sample is measured with some error and is normalized against the original length to determine the engineering strain. The force applied to the sample is measured with an experimental error and is normalized by the cross sectional area to determine the stress. These two parameters are plotted (stress versus strain) and a curve fit is used to determine the modulus at low strain. In order to determine the standard deviation in the modulus the experimentally determined errors in length, force and area must be propagated to the stress and strain and the error in these parameters must be propagated to a linear fit of the data points. Propagation of errors is a rudimentary tool necessary to perform polymer analysis.

Consider a measurement of the absorption, A, and absorption coefficient, α, using a single wavelength of light which passes through a sample and a photomultiplier tube which reports counts. The relevant equation is the Beer-Lambert Law for linear absorption (pp. 40 and 54 in Cambell):

$$A = \ln\left(\frac{I_0}{I}\right) = \ln\left(\frac{1}{T}\right) = \alpha c l$$

where c is the concentration of absorbing species and l is the sample thickness. T is called the transmission. For a solid sample c = 1. Two measurements are necessary, I_0 with no sample and I with a sample of thickness l. If the two measurements are I_0 = 100,000 counts , and I = 20,000 counts, the respective standard deviation is given by a Poission Distribution function as the square root of the number of events,

I_0 = 100,000 ± 320 counts; I = 20,000 ± 140 counts.

To propagate the error in I_0 and I to 1/T the general error propagation rule in differential form can be used:

$$\sigma_x^2 = \sigma_u^2\left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2\left(\frac{\partial x}{\partial v}\right)^2 + 2\sigma_{uv}^2\left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right)$$

Here it is safe to assume that I and I_0 are uncorrelated (unrelated) so the covariance, $\sigma_{uv} = 0$. The two standard deviations are given above. For $x = 1/T$, $\delta x/\delta I_0 = 1/I$, and $\delta x/\delta I = -I_0/I^2$. The above equation yields:

$$\sigma_{1/T}^2 = \frac{\sigma_{I_0}^2}{I^2} + \frac{I_0^2\sigma_I^2}{I^4}$$

or

$$\frac{\sigma_{1/T}^2}{(1/T)^2} = \frac{\sigma_{I_0}^2}{I_0^2} + \frac{\sigma_I^2}{I^2}$$

this follows the general rule for ratios given by Bevington (see handout). The propagated value and error for $1/T$ is, **$1/T = 5 \pm 0.04$**.

The absorption, A, is the natural log of $1/T$ and the propagated error is given by:

$$\sigma_A^2 = \sigma_{1/T}^2\left(\frac{\partial A}{\partial(1/T)}\right)^2 = \frac{\sigma_{1/T}^2}{(1/T)^2}$$

So, **A = 1.61 $\pm$ 0.01**.

The sample thickness, l, is 0.1 ± 0.02 cm as calculated from a series of 10 measurements using the mean and number of samples equation given above. The absorption coefficient, α, is calculated from $\alpha = A/cl$, where c is 1 for a solid sample. Replacing the variables in the equation above for the error in $1/T$, α = **16 $\pm$ 3**. *Notice the reduction in the number of significant figures associated with the large error.*

It should also be noted that often the largest source of error is related to factors which are of minor significance to the measurement, e.g. here error in the sample thickness dominates the error in the absorption coefficient.

Purpose of Error Analysis:

Although it is assumed that you can determine the error value for any quantity you measure in science, the most important part of error analysis involves interpretation of the significance of the analysis in view of the error and in view of logic and the reasonableness of the results. Error analysis is a critical factor in both demonstrating the *scientific reasonableness* of a result, as well as forming a basis of a *scientific critique* of work performed for you by a technician or outside lab. Error analysis is always at the heart of a *scientific argument*. A measured value has no meaning without an analysis of the associated error. Such an analysis can be quantitative, as above, or qualitative, based on your scientific judgment. The value in either case is based on the reasonableness and logic of the approach. In the remainder of this course special emphasis will be given to qualitative and quantitative assessments of the error in the analytic techniques covered.

Least-Squares Fits:

In order to determine analytic parameters it is often necessary to perform a curve fit to a raw data set. For example the determination of modulus from a stress-strain measurement requires a linear fit of the type $\sigma = E\varepsilon$ where s is the stress, $\sigma = F/A$, and e is the strain, $\varepsilon = \Delta l/l$, and E is the Youngs modulus. The error in the stress measurement could be propagated from the estimated error in the Force and area measurements. These errors can be propagated to the modulus in a linear fit through a least squares minimization of χ^2. χ^2 is a measure of the difference between the actual data points and the projected points associated with the fit parameter. It bears resemblance to the variance,

$$\chi^2 = \sum_1^N \frac{\left(y_i - f(x_i)\right)^2}{\sigma_i^2}$$

The propagated uncertainty in the coefficients for the least squares fit can be obtained in a computer program by calculation of the second derivative of χ^2 with respect to variation in the parameter:

$$\sigma_{a_j}^2 = \frac{2}{\partial^2 \chi^2 / \partial^2 a_j^2}$$

A full discussion of least-squares fitting routines and propagation of error is given in Bevington. For linear functions a relatively simple algorithm for propagation of error is given in the handout.