Cohen, R.D. (1991) "Why Do Random Samples Represent Populations So Accurately?," Journal of Chemical Education, Vol. 60, pp. 902-903.

Why Do Random Samples Represent Populations So Accurately?

Ruben D. Cohen
Department of Mechanical Engineering & Materials Science, Rice University, Houston, Texas 77251

Selecting samples from large *populations* (sets or collections of numbers, measurements, or observations related to actual or conceptual situations[1]) is sometimes the only way to obtain certain statistical characteristics of the population without having to deal with huge amounts of available information. It is generally accepted that if the sample is large and taken at random (selected without prejudice), then it quite accurately represents the statistics of the population, such as distribution, probability, mean, standard deviation, etc. This work provides a simple and easy way to follow proof for this.

For simplicity, our example concerns variously sized clusters in dispersion. Although the choice of examples to carry out the proof is arbitrary, we choose this for its application in many fields—chemical engineering, chemistry, and physics.

Consider a dispersion containing a population of N clusters, each composed of a number of primary particles i, where i = 1, 2, 3,....The figure illustrates a hypothetical situation involving a dispersion of N = 11 variously sized clusters. We should emphasize, however, that in practical situations, N is extremely large, usually at least 10^8.

If we denote N_i to be the number of clusters of size i in the actual dispersion, the probability P_i of finding a cluster of size i in that population is simply given by

$$P_i = \frac{N_i}{N} \qquad (i = 1, 2, 3,...) \quad (1)$$

Consequently, P_i would be the *actual* size distribution probability of the clusters in the dispersion.

However, due to the very large numbers involved, it would be a formidable task to examine all N clusters in order to evaluate P_i accurately. It is preferable to obtain a random, relatively large sample (but significantly smaller than the population) and measure its size distribution, hoping that it comprises the statistical properties of the actual distribution.

Assume now that the extracted sample has a total of n clusters such that

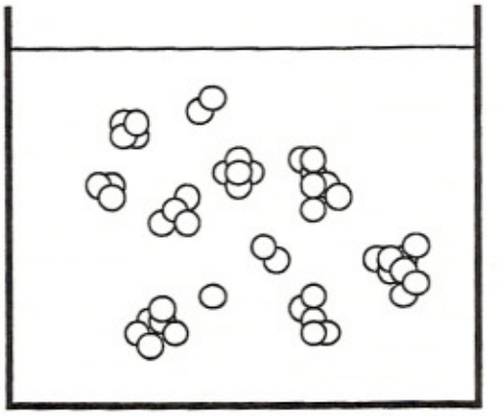
$$N >> n >> 1 \quad (2)$$

That is, the sample is large, but significantly smaller than the population, to make counting easier.

Recall that the probability of having n_i clusters of size i in a sample with a total of n aggregates (taken randomly from a population whose size distribution probability is exactly equal to P_i) is simply given by the binomial distribution.[1] Then the occurrence of *any* size distribution n_i within the sample can be denoted by a finite probability $\tilde{p}\ (n_i, n)$ such that

$$\tilde{p}\ (n_i, n) = \frac{n!}{(n - n_i)!\ n_i!}\ [1 - P_i]^{n - n_i}\ P_i^{n_i} \quad (3)$$

Since many real cases involve sampling without replacement, the hypergeometric distribution is more suitable. However, the hypergeometric distribution approaches the binomial distribution if the sample size, though large, is still much smaller than the actual population,[2] such as the case considered here.



Schematic diagram of a suspension with N = 11 variously sized clusters.

Now we search for the size distibution that would *most probably* be observed within the sample. This can be carried out easily by maximizing the function $\ln \tilde{p}\ (n_i, n)$ in eq 3 with respect to n_i, that is,

$$\frac{\partial}{\partial n_i} \ln \tilde{p}\ (n_i, n) = 0 \qquad (i = 1, 2, 3,...) \quad (4)$$

Applying the above to the right side of eq 3 gives

$$\frac{\partial}{\partial n_i} [\ln n! - \ln (n - n_i)! - \ln n_i! + (n - n_i) \ln (1 - P_i) + n_i \ln P_i\ ] = 0 \quad (5)$$

If we assume a sample large enough so that

$$n >> n_i >> 1 \quad (6)$$

then Stirling's approximation for $\ln n!$, $\ln (n - n_i)!$, and $\ln n_i!$ can be used.[2] These are given by

$$\ln n! = n \ln n - n \quad (7)$$

$$\ln (n - n_i)! = (n - n_i) \ln (n - n_i) - (n - n_i) \quad (8)$$

and

$$\ln n_i! = n_i \ln n_i - n_i \quad (9)$$

using only the first two terms of the series.

Substituting eqs 7–9 into eq 5 and carrying out the partial differentiation yields

$$\ln\left[\frac{n - n_i}{n_i}\right] - \ln\left[\frac{1 - P_i}{P_i}\right] = 0 \qquad (i = 1, 2, 3,...) \quad (10)$$

[1] Miller, I.; Freund, J. E. *Probability and Statistics for Engineers;* Prentice-Hall: New Jersey, 1965.
[2] Kreyszig, E. *Advanced Engineering Mathematics,* 3rd ed.;Wiley: New York, 1972.

which is satisfied if

$$\frac{n_i}{n} = P_i \qquad (i = 1, 2, 3, \ldots) \quad (11)$$

Equation 11 agrees with eq 1, obviously proving that if a sample is relatively large and selected at random (by virtue of eqs 6 and 3, respectively), then its statistical properties will *most probably* be similar to those of the population.

"Kinetics of Pentaamminenitritocobalt(III) to Pentaamminenitrocobalt(III) Linkage Isomerization Revisited"—A Correction

W. Gregory Jackson
University College (NSW), ADFA, Canberra, Australia 2600

The title reaction and its reverse have received considerable scrutiny from the point of view of mechanistic detail, especially in the last decade (*1–7*). In the March issue of *this Journal* last year aspects of this work were again presented (*8*), without reference to any of this recent literature. The detectable reversibility of the nitrito–nitro isomerization reaction for the Cl^- salt in the solid state was confirmed, and aside from identifying a propagated misprint in the earlier literature (*9,10*), the only aspect that appears to be new is the suggestion that the nitrito to nitro isomerization is significantly reversible in aqueous solution, and this is certainly incorrect.

$$[(NH_3)_5Co\text{–}ONO]^{2+} \underset{\leftarrow}{\overset{K_{ON}}{\rightleftharpoons}} [(NH_3)_5Co\text{–}NO_2]^{2+}$$

Thus the authors imply that that the kinetics of isomerization in solution can be studied commencing with either the nitrito or nitro isomer, although more conveniently using the nitrito form because the equilibrium favors the nitro form (K_{ON} = 3.4, in the range 20–47 °C (*8*)).

The nitrito-to-nitro isomerization can proceed by a spontaneous or a much faster base catalyzed pathway, but the equilibrium position is of course independent of the mechanistic route. We previously (*2, 3*) determined that there is no detectable nitrito complex under equilibrium conditions by either route, implying $K_{ON} > 99$. A particularly sensitive analytical method is to take the equilibrium mixture of isomers and treat it with HCl, which leaves the nitro isomer untouched but converts the nitrito form to the aqua complex via N–O cleavage. (The nitro form reacts rapidly, also by N–O cleavage, only in a very strongly acidic medium such as neat CF_3SO_3H (*6*).) The complexes $[Co(NH_3)_5NO_2]^{2+}$ and $[Co(NH_3)_5OH_2]^{3+}$ are then readily separated by ion-exchange chromatography and determined individually (spectrophotometrically, for example). For the equilibrium mixture in aqueous solution, no residual nitrito complex (1%) was detected by this method (*2, 3*). If the result reported by Phillips et al. (*8*) were correct, then there would be 23% nitrito form at equilibrium in aqueous solution. Given the substantial differences in the electronic spectra of the nitro and nitrito linkage isomers (*2*), this would be easily detected, particularly around 324 nm where the absorbance due to the nitrito complex is a minimum (ε 162 $M^{-1}\,cm^{-1}$) and that due to the nitro complex is a maximum (ε 1720 $M^{-1}\,cm^{-1}$). A convincing demonstration is provided by the absorbance–time plot at 324 nm commencing with the pure (*2, 3*) nitro complex. This gives a horizontal trace over several hours at 25 °C (Cary 210 instrument, [Co] = 5.26×10^{-4} M, absorbance = 0.905±0.005), whereas the results of Phillips et al. (*8*) would require the absorbance diminish to $100[(0.77 \times 1720) + 0.23 \times 162)]/1720 = 79\%$ of the original at this wavelength.

We conclude that the nitrito-to-nitro isomerization in aqueous solution (as in many other solvents (*4*)) proceeds to completion with no detectable reversibility. Phillips et al. may have been led to their wrong result through the use of two-component spectrophotometric analysis with incorrect molar absorptivies for one or both of the two isomers. Some molar absorptivities were reported (*8*), but the agreement with literature values for the nitrito isomer was not good. If one isomer is an impurity in the other then anomalies do not appear (*11*) in the two-component spectrophotometric analysis (nor do they appear in the kinetics), and therein could lie the difficulty. Alternatively, the nitrito complex, as synthesized (*8*), contained traces of HCl that rapidly generated some aqua complex in unbuffered solution, and this was not distinguished from the spectrally similar nitrito species.

The position of the nitrito–nitro equilibrium is clearly different for solution and solid state. Furthermore, in the solid state the equilibrium should depend on the counterion. Phillips et al., using IR analysis for the Cl^- salts in KBr discs, report K_{ON} ca 4.0 (20% nitrito, 80% nitro complex) in the range 40–60 °C (8) when starting with the nitrito isomer, and a similar value (4.8) when commencing with the nitro complex. However, using our ion-exchange analytical method described above, we find only 9% residual nitrito complex after equilibration of the pure Cl^- salt (K_{ON} = 10, 65 °C), a significantly different result. Furthermore, isomerization is almost complete commencing with the perchlorate salt (2% residual nitrito isomer, K_{ON} ca. 50, 65 °C). The difference in equilibrium for the two salts is not surprising since the lattices are different, and it is suggested that our result for the chloride salt differs to the published value (8) because under pressure KBr interferes with the lattice of the chloride salts. The question is also raised whether the observed solid state equilibrium position should be the same when approached from either side if the nitrito and nitro salts are not isomorphous, i.e., they have different lattices.

Literature Cited

1. Jackson, W. G.; Lawrance, G. A.; Lay, P. A.; Sargeson, A. M. *J. Chem. Soc. Chem. Commun.* **1982**, *70*.
2. Jackson, W. G.; Lawrance, G. A.; Lay, P. A.; Sargeson, A. M. *Inorg. Chem.* **1980**, *19*, 904–910.
3. Jackson, W. G.; Lawrance, G. A.; Lay, P. A.; Sargeson, A. M. *J. Chem. Educ.* **1981**, *58*, 734–738.
4. Jackson, W. G.; Lawrance, G. A.; Lay, P. A.; Sargeson, A. M. *Aust. J. Chem.* **1982**, *35*, 1561–1580.
5. Jackson, W. G.; Randall, M. L.; Sargeson, A. M.; Marty, W. *Inorg. Chem.* **1983**, *22*, 1013.
6. Jackson, W. G. *Inorg. Chem.* **1987**, *26*, 3857–3859
7. Jackson, W. G. *Inorg. Chim. Acta.* **1988**, *149*, 101–104
8. Phillips, W. M.; Choi, S.; Larrabee, J. A. *J. Chem. Educ.* **1990**, *67*, 267–269.
9. Beattie, I. R.; Satchell, P. *N. Trans. Faraday Soc.* **1956**, *52*, 1590–1593.
10. Basolo, F.; Hammaker, G. S. *Inorg. Chem.* **1962**, *1*, 1–5.
11. Jackson, W. G.; Begbie, C. M. *Inorg. Chem.* **1983**, *22*, 1190–1197.