RAPID COMMUNICATIONS

Distribution of Aflatoxin in Pistachios. 5. Sampling and Testing U.S. Pistachios for Aflatoxin

Keywords: *Very large sample testing; sampling variance; experimental distributions*

INTRODUCTION

The clear characterization of aflatoxin infestation in tree or ground nuts requires a knowledge of the distribution of such infestation among the nuts comprising the population under study. In particular, relations between measured sample means and the confidence that can be applied to such results depend critically on the form of aflatoxin distribution (Schatzki, 1995a). Currently, importing countries are proposing new sample testing protocols. It is thus imperative that distributions be established to supply scientific bases for these protocols. In principle, lot distributions can be obtained by measuring aflatoxin level nut by nut. However, because of the rarity of infestation at any level (1 infested nut per 10^4-10^6 nuts is typical), this is not practical. Instead, one takes a sample, or a set of samples, of a size hopefully representative of the lot and attempts to obtain the lot distribution from a statistical analysis of the sample aflatoxin distribution. In a previous publication (Schatzki, 1995a) a method was presented for deriving such lot distributions nonparametrically from sample distributions. In an adjoining publication (Schatzki, 1995b) this method was applied to several sets of data on pistachios, previously published by others. The nature of this distribution in tree nuts, i.e., the difficulty of having a sample be representative of the lot, results in an extremely broad sample distribution. To characterize the lot, one requires a very large number (several hundred) of samples. Consequently, only four such distribution studies were available, and of these only one could be characterized as that of a single lot drawn from two orchards. In the other three cases, which included two sets of processed pistachios, it was necessary to combine results on lots which formed part of a single study or survey but which involved assorted lots that varied considerably in aflatoxin level. In Schatzki (1995b) the point was taken that combining these results into four distinct distributions was acceptable; that is, the similarity of the combined lots outweighed their differences. Indeed, it was found that the resulting four lot distributions showed great similarity and individually were of a form which could be justified on physical grounds. Nevertheless, considerable objection was raised by a number of reviewers on the grounds that true populations were not represented and, to some extent, the question was left in abeyance.

MATERIALS AND METHODS

It was thus thought worthwhile to measure the aflatoxin distribution for a large number of fixed size samples drawn from a single lot and derive the lot distribution therefrom. This lot was derived by starting with a mixed floater and sinker eye reject lot [see Schatzki and Pan (1996), processor A] from the 1995 crop, containing ~6 ng/g aflatoxin. The lot was further sorted, using an image sorter. This sorter is a modified color sorter but uses high-speed digital cameras to obtain grayscale images of the flying nuts. Nuts are sorted on the basis of the detected presence of a shell staining pattern characteristic of early split nuts and thus possible aflatoxin contamination (Pearson, 1996). Two equal size sublots were obtained, one at 0.03 ± 0.01 ng/g aflatoxin and one at 14.6 ± 4.1 ng/g aflatoxin, on the basis of the following analysis. One hundred sixty 500-nut samples, drawn from the aflatoxin-containing sublot, were measured for aflatoxin, using the protocol previously described (Schatzki and Pan, 1996). Another 100 500-nut samples, drawn from the clean sublot, were measured as well. A binning of half-decades was used, as discussed in Schatzki (1995a). Results for total aflatoxin (G_1, B_1, plus B_2; G_2 was almost never seen) are listed in Table 1. For consistency with other workers, results in the table are expressed in terms of supernatant extraction liquid, rather than total extraction liquid, as was done in Schatzki and Pan (1996). As none of the binned sample probabilities P_i significantly exceeded 10% (16/160 or 10/100), the sparse approximation (Schatzki, 1995a) was used to compute the sublot distributions from $p_i = P_i / n$ and $c_i = C_i n$, with $n = 500$ nuts and C_i taken as the geometric mean of the range, where P_i and C_i stand for the sample distribution and p_i and c_i for the lot distribution. The sparse approximation applies when the sample size is small enough so that a sample generally contains at most a single contaminated nut of any detectable aflatoxin level, but large enough that some contaminated samples are obtained. Typically, P_i is between 0.01 and 0.1. These two sublot distributions were then combined to yield the eye reject lot distribution. Figure 1 shows this lot distribution along with the distributions taken from Schatzki (1995b).

RESULTS AND DISCUSSION

Consideration of the five curves of Figure 1 shows striking similarities between the distributions, despite the very different provenances of the five populations. The Sommer and DFA 83−86 data refer to populations of unprocessed pistachios, immediately following harvest. The two DFA 90−91 curves refer to fully processed pistachios, with and without roasting, which have undergone extensive sorting [in Schatzki (1995b) "raw" pistachios were erroneously thought to be unprocessed]. Finally, the eye reject curve shows nuts that were removed by such a sort.

The basis for the dip in the distribution around 8000 ng/g is not clear, although it appears that the total distribution is, in fact, the sum of two contributions, one accounting for the contamination levels below 8000 ng/g, the other for contamination above that. The source of the contamination at low levels is not understood, but at any rate is of little importance in determining

Table 1. Sample Distribution of Sorted Eye Rejects (500-Nut Samples)

	ng/g										
	<0.01	0.01−<0.03	0.03−<0.10	0.1−<0.32	0.32−<1.00	1.00−<3.16	3.16−<10.0	10.0−<31.6	31.6−<100	100−<316	316−
dirty[a]	67	17	18	12	10	12	6	2	9	7	0
clean[b]	86	1	2	8	3	0	0	0	0	0	0

[a] Number of dirty sublot samples among 160 falling into indicated bin. [b] Number of clean sublot samples among 100 falling into indicated bin.

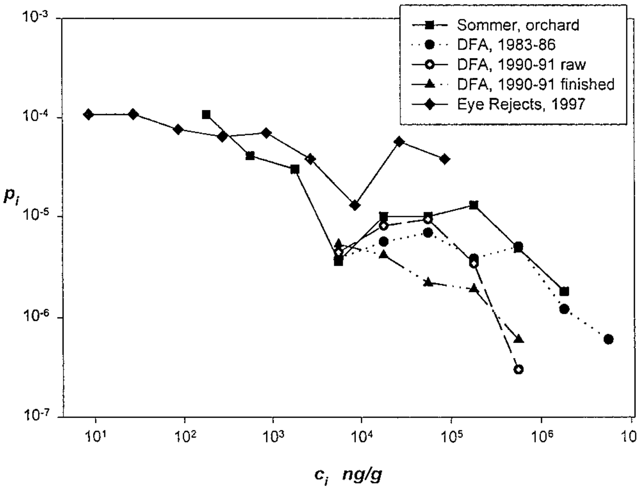

Figure 1. Probability of a single nut having aflatoxin concentration in a half-decade range c_i in U.S. pistachios.

average contamination which is dominated by high c_i values. Contamination at high levels can be understood on physical grounds, however. It is known that preharvest aflatoxin contamination only occurs in nuts of which the hulls were split early on the tree (Sommer et al., 1986), presumably because such hull splitting allows access by fungus with subsequent aflatoxin production. Further, it is known that such splitting occurs during a 6 week period prior to harvest, with a rate during the 2−4 weeks prior to harvest about twice that before and after (Doster and Michailides, 1994). Fungal growth, being a biological process, can be expected to be exponential in time. Hence, the log c axis in Figure 1 represents a time axis in terms of fungal growth and presumably aflatoxin production, and, indeed, a 2-fold maximum is precisely what is observed in aflatoxin concentration. All distributions drop rather precipitously at some higher value of c, $c_{\max}$. The physical basis for the latter may well be the maximum level a nut can sustain or the amount of nutrient available for fungal growth (Schatzki, 1995b). At any rate, such a maximum level is observed not only in tree nuts but in ground nuts as well. The one important difference between the distributions appears to be this value of $c_{\max}$. The value for unsorted U.S. pistachios is a few times 10^6 ng/g, for the fully sorted product ∼560 000 ng/g, and for the reject lot tested here ∼110 000 ng/g (although a low p_i value at ∼300 000 ng/g might have been missed because of the relatively small number of samples, $N = 260$). What is clear is that the assumption made in Schatzki (1995b) is verified: U.S. pistachio populations are similar and, except for possibly removing some high-concentration nuts, sorting techniques do not change the shape of the distribution; that is, such sorting removes nuts of all aflatoxin levels equally and simply lowers the level of the entire distribution.

The reliability of test results is predicted from the variance of the mean concentration of a set of samples. An expression for this quantity may be derived as follows. Suppose first that all aflatoxin-containing nuts are contaminated at the same concentration c with probability p (i.e. $p_i = 0$ for all but one i). The probability of having x contaminated nuts in a sample of n nuts is then given by the Poisson distribution with mean np and var (variance) np. Since the sample concentration C is given by xc/n, the mean of C is $np \times c/n = pc$ and var C is $np \times (c/n)^2 = pc^2/n$. If, instead, contamination occurs at many levels i, $C = \sum C_i$. Since the mean of a sum is the sum of the means and var of a sum is the sum of the var's, mean $C = \sum p_i c_i$ and var $C = \sum p_i c_i^2/n$. If $N > 1$, samples are taken and the average of C, $\langle C \rangle_N$, is computed, then by the central limit theorem, $\langle C \rangle_N$ will be normally distributed with mean = mean C and standard deviation, designated as the standard error (SE) of the estimate of C, computed as

$$\mathrm{SE} = [\mathrm{var}\ C/N]^{0.5} = [(\textstyle\sum p_i c_i^{\,2}/\sum p_i c_i) \times \text{mean } C/nN\ ]^{0.5}$$

after multiplying and dividing by the mean, while the sample mean estimates the lot mean. This expression was derived in Schatzki (1995b), but with use of the sparse approximation. It is now seen to be applicable for all n. To evaluate the standard error, the ratio $\sum p_i c_i^2/\sum p_i c_i$ must be estimated from the lot distribution. Only the terms at high c_i will matter. As noted, the distributions differ mainly in their height, which indicates that the p_i will substantially cancel out. Were p_i constant, ending at $c_{\max}$, a value of $0.76c_{\max}$ would be obtained for this ratio. In the present case actual p_i and c_i values are available for all five distributions. One obtains 2.7×10^6 and 8.6×10^5 ng/g for the two unsorted populations, 3.0×10^5 and 1.6×10^5 ng/g for the processed, sorted (finished and raw) pistachios, and 6.7×10^4 ng/g for the eye rejects measured here. Clearly, a precise value is not available, but a value of 2.4×10^5 ng/g for processed pistachios can be expected to predict the standard error to 25%. The product nN refers to the number of nuts tested, N samples of n nuts each. It does not matter how these nuts are divided into samples, provided $N > 1$.

The resulting equation may be written in terms of the total sample weight (rather than the number of nuts tested) by taking into account that a kilogram contains ∼700 nuts. One obtains SE = [340 × mean (in ng/g)/weight of all samples (in kg)]$^{0.5}$. This makes direct testing difficult. Suppose a buyer will only accept a lot that tests below 4 ng/g. To get 97.5% acceptance will require a lot for which mean + 2SE ≤ 4 ng/g. For the seller there are basically three options: (1) Present a lot with very low mean. A lot of mean = 0.1 ng/g and a 10 kg total sample will result in a standard error of 1.85 ng/g. (2) Accept a high rejection rate, i.e., gamble. For a mean = 2 ng/g and a 10 kg sample the standard error will be 8.3 ng/g, which predicts that 40% of samples will fall above 4 ng/g. (3) Accept a very large sample. If mean = 2 ng/g and total sample = 690 kg, an acceptable standard error of 1 ng/g is obtained. Such a large sample appears impractical. There is a solution available, however. Tests of the image sorter, referred to above, have shown that this sorter is capable of removing all of the aflatoxin-containing pistachios of a (fully commercially sorted) product into a side stream consisting of only 2% of the total product. By obtaining a 690 kg sample, sorting it with the image sorter, testing only the 13.7 kg removed, and multiplying the result by the fraction removed (here 2%), one obtains the mean of the lot to a confidence given by the large 690 kg sample. Only 13.7 kg is destroyed; the remaining 676 kg can be returned to the lot. Moreover, this method is rugged, even if the sorter left 1 ng/g behind in the "clean" 676 kg, the error in the mean would amount to but 1 ng/g. A paper describing tests on this sorter will be published shortly by Pearson and Schatzki. Approaches 1 and 2 have both been used for some time in the pistachio industry. Method 3 is a new approach, now possible because of the availability of the new sorter. It must be pointed out, however, that the results presented here depend, at least in detail, heavily on the form of the aflatoxin distribution. This is known only for U.S. pistachios in which the source of aflatoxin is generally entirely orchard-based. It may not apply to cases when aflatoxin production occurs during storage, leading to different distributions.

ACKNOWLEDGMENT

I thank J. L. Pan for aflatoxin analysis.

LITERATURE CITED

Doster, M. A.; Michailides, T. J. The Development of Early Split Pistachio Nuts and their Contamination by Molds, Aflatoxins and Insects. *California Pistachio Industry Annual Report 1993−1994*; The California Pistachio Commission: Fresno, CA, 1994; pp 68−73.

Pearson, T. C. Machine Vision System for Sorting Stained Pistachio Nuts. *Lebensm. Technol.* **1996**, *29*, 203−209.

Schatzki, T. F. Distribution of Aflatoxin in Pistachios. 1. Lot Distributions. *J. Agric. Food Chem.* **1995a**, *43*, 1561−1565.

Schatzki, T. F. Distribution of Aflatoxin in Pistachios. 2. Distributions in Freshly Harvested Pistachios. *J. Agric. Food Chem.* **1995b**, *43*, 1566−1569.

Schatzki, T. F.; Pan, J. L. Distribution of Aflatoxin in Pistachios. 3. Distribution in Pistachio Process Streams. *J. Agric. Food Chem.* **1996**, *44*, 1076−1084, 2468.

Sommer, N. F.; Buchanan, J. R.; Fortlage, R. J. Relation of Early Splitting and Tattering of Pistachio Nuts to Aflatoxin in the Orchard. *Phytopathology* **1986**, *76*, 692−694.

Received for review October 10, 1997. Revised manuscript received November 4, 1997. Accepted November 7, 1997.

Thomas F. Schatzki[†]
Western Regional Research Center, Agricultural
Research Service, U.S. Department of Agriculture,
Albany, California 94710

JF970872T

[†] E-mail tom@pw.usda.gov