Chemometrics and Intelligent Laboratory Systems 151 (2016) 103–107

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Sampling plans for beta distributed compositional fractions

K. Govindaraju [a,*], R. Kissling [b]

[a] Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
[b] Fonterra Co-operative Group Limited, Cambridge, New Zealand

ARTICLE INFO

Article history:
Received 24 September 2015
Received in revised form 7 December 2015
Accepted 16 December 2015
Available online 28 December 2015

Keywords:
Beta distribution
Bulk sampling
Compositional fraction
Quality levels
Variables plan

ABSTRACT

Sampling inspection plans for compositional fractions based on the beta distribution are discussed. Design of plans to control the proportion nonconforming levels are covered. It is shown that the traditional plans based on the normal distribution fail to maintain the desired risks. Plans based on the beta distribution are not only suitable for bulk product inspection, but they also achieve considerable economy when composite samples are used in testing.

1. Introduction

Compositional characteristics are the primary quality measures for bulk materials. For example, the percentage protein is a primary quality measure for milk products, and a minimum protein limit of 34% is set for milk powders, see Ref. [1]. Compositional fractions can be modeled using the beta distribution. Published literature on the use of the beta distribution for quality control applications is sparse. Beta distribution is mainly used only as a prior probability distribution for proportion nonconforming in the quality literature. Ref. [2] employed the beta distribution for controlling the mean of the fraction of material having a given characteristic having either single or double sided specification limits. His work was motivated by an application problem involving the percentage of fine product for a refractory cement. Ref. [3] provided a broader computer intensive approach to designing sampling inspection plans for compositional fractions for a given point on the operating characteristic (OC) curve assuming non-normal distributions.

Designing sampling inspection plans for the beta distribution is useful for food quality assurance because many of the food quality measures are compositional fractions. Sampling inspection plan design of Ref. [2] is improved in Section 2. Design of beta distribution based plans for given two points of the OC curve is discussed in later sections. A case study is provided at the end.

* Corresponding author. Tel.: +64 6356 9099; fax: +64 6355 7953.
E-mail address: k.govindaraju@massey.ac.nz (K. Govindaraju).

http://dx.doi.org/10.1016/j.chemolab.2015.12.009
0169-7439/

2. Sampling plans for mean compositional fraction

Let X, the compositional fraction in unit mass g, follows the beta distribution whose density function is given by

$$\mathrm{Beta}(a,b)=\frac{1}{\mathrm{B}(a,b)}x^{a-1}(1-x)^{b-1} 0<x<1, a>0, b>0 \tag{1}$$

where $\mathrm{B}(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function. The beta density is also reparameterized in terms of the expected value $\mathrm{E}(X)=\mu=a/(a+b)$ and *precision* parameter $\theta=a+b$ as $\mathrm{Beta}(\mu\theta,(1\text{-}\mu)\theta)$. For large θ, $\mathrm{Var}(X)=ab/[(a+b)^2(a+b+1)]\simeq ab/\theta^3$.

A probability distribution is termed as a *flexible* distribution when it has at least one *shape* parameter. The shape parameter enables the distribution to take numerous shapes for its density so that a variety of data sets can be well modeled. The beta distribution is a highly flexible distribution because both of its parameters are shape parameters. As a result, both left and right skewed densities can be fitted to the data in addition to symmetric densities. The normal or Gaussian density has a fixed shape and its parameters model just the change in the location and scale of its density. As a result, the normal distribution becomes inflexible to model compositional proportions data which often show a variety of shapes. Even though we cannot directly give a physical meaning to the beta parameters a and b, their sum $(a+b)$ is an inverse measure of variability in the quality characteristic. However the standard beta distribution can be reparameterized in terms of its mean and this particular form is mainly discussed in this paper.

Let the total bulk material amount sampled be G (such as 100 g, 200 ml). G can be expressed as a multiple of the standard or primary unit mass g. Let $m = G/g$ (which need not be an integer). The quantity m is similar to the sample size defined for discrete or non-bulk items. Let the random variable $\hat{\mu}$ be the mean compositional fraction for amount G. Note that $\hat{\mu}$ can be a single measurement based on a well mixed composite, and need not be the arithmetic mean of m measurements of individual test samples.

The distribution of $\hat{\mu}$ was approximated as Beta($m\mu\theta, m(1\text{-}\mu)\theta$) by Ref. [2]. The means and variances of $\hat{\mu}$ and X were matched as follows:

$$\mathrm{E}(\hat{\mu}) = \mathrm{E}(X) = a/(a+b) = \mu, \tag{2}$$

$$\mathrm{Var}(\hat{\mu}) \simeq \mathrm{Var}(X)/m = \mu(1-\mu)/[m(a+b)]. \tag{3}$$

This approximation is very similar to the one given by Ref. [4] for the sum of beta random variables; also see Ref. [5].

For the case of upper specification limit μ_U, the lot acceptance criterion is $\{\hat{\mu} < \kappa\}$, and this inequality is reversed for the lower specification limit μ_L. For given acceptable mean level μ_{AQL}, producer's risk α, rejectable mean level μ_{LQL}, and consumer's risk β, Ref. [2] solved the equations

$$\Pr\left(\hat{\mu} < \kappa | \mu = \mu_{AQL}\right) = 1 - \alpha \tag{4}$$

$$\Pr\left(\hat{\mu} < \kappa | \mu = \mu_{LQL}\right) = \beta \tag{5}$$

to determine the critical limit κ for a given θ. The example given in Ref. [2] sets $\mu_{AQL} = 1\%$, $\alpha = 5\%$, $\mu_{LQL} = 5\%$, $\beta = 5\%$ and $\theta = 300$. The critical value κ was obtained as $\kappa = 0.253$ for an optimum m of 0.535 based on the beta distributions Beta($m\mu_{AQL}\theta, m(1\text{-}\mu_{AQL})\theta$) and Beta($m\mu_{LQL}\theta, m(1\text{-}\mu_{LQL})\theta$); see Fig. 1. It should be noted that this design procedure does not explicitly use the specification limit μ_L or μ_U but the direction of the inequality in the acceptance criterion depends on it.

3. Sampling plans for proportion nonconforming compositional fraction

Variables plans for controlling the proportion nonconforming p are commonly used when compared to sampling plans which control just the mean. The beta distributions shown in Fig. 1 show considerable overlap even when the acceptable and rejectable mean levels are far apart. The proportion nonconforming p obviously depends on the specification limits L and/or U set for the compositional characteristic X.

First consider the case of the lower specification limit L which results in $p = \Pr(X < L|\mu, \theta)$. The sample G gives the estimated mean $\hat{\mu}$ and hence the proportion nonconforming estimate $\hat{p} = \Pr(X < L|\hat{\mu}, \theta)$ for a given θ. Similar to the variables plan based on the normal distribution (see Chapter 10 in Schilling and Neubauer [6]), an acceptance criterion based on the estimated mean and standard deviation of X can be set up. Consider the lot acceptance criterion $\hat{\mu} - k\hat{\sigma} > L$ (or $\hat{\mu} + k\hat{\sigma} < U$ for the upper specification limit) where $\hat{\sigma} \simeq \sqrt{\hat{\mu}(1-\hat{\mu})/\theta}$, the estimated standard deviation of X. It should be noted that $\hat{\sigma}$ is different from the usual sample standard deviation S and depends only on the estimated beta parameter $\hat{\mu}$. The sampling distribution of the acceptance criterion $\hat{\mu} - k\hat{\sigma}$ (or $\hat{\mu} + k\hat{\sigma}$) cannot be easily approximated as a beta distribution (as was done for $\hat{\mu}$) or in any other closed form. Monte Carlo simulation can be employed instead.

It can be noted that the approximation $\hat{\sigma} \simeq \sqrt{\hat{\mu}(1-\hat{\mu})/\theta}$ matches the formula for the standard error of an estimated binomial nonconforming proportion when $n = \theta$ discrete items are tested. For bulk material, testing a single composite sample is sufficient as long as the n equal volume samples are thoroughly mixed. If aliquots of individual samples of equal volumes are physically well mixed, the single measurement of a composite sample is equal to the arithmetic mean of the individual sample values, see Chapter 7 in Ref. [7]. Ref. [8] discussed food safety applications when mixing is imperfect. Unlike the sanitary and safety characteristics covered in Ref. [8], compositional characteristics are easy to homogenize (for example, by blending).

The plan parameters G (or m) and k can be determined for given two-points of the OC curve. Let p_1 be the acceptance quality limit (AQL) for the proportion nonconforming compositional fraction, and p_2 be the rejectable or limiting quality limit (LQL). Let α and β be the producer's and consumer's risks respectively corresponding to p_1 and p_2. The proportion nonconforming p is a function of μ and θ and hence the OC function $P_a(p)$ can be expressed as $P_a(p) = \Pr(\hat{\mu} - k\hat{\sigma} > L|\mu, \theta, k, G)$. The two point design imposes the conditions $P_a(p_1) = 1\text{-}\alpha$ and $P_a(p_2) = \beta$. The amount G or m controls the variability in the estimates $\hat{\mu}$ and $\hat{\sigma}$ while k mainly influences the achieved producer's and consumer's risks.

For a given upper specification limit U and known θ, the mean levels μ_1 and μ_2 corresponding to the desired p_1 and p_1 can be found. For example, the upper specification limit U for water or moisture composition in milk products is usually 5%, see Ref. [1]. For $p_1 = 0.01$ and $p_1 = 0.05$, the corresponding mean levels $\mu_1 = 0.0403$ and $\mu_2 = 0.04312$ can be found using the beta distribution function, see Fig. 2. For fixed $\alpha = 0.05$ and $\beta = 0.10$, the variables plan based on the normal distribution can also be found. R software package *AcceptanceSampling* of Kiermeier [9] gives the normal distribution based known sigma variables plan parameters as $n = 19$ and $k = 1.949$ for $p_1 = 0.01$ and $p_1 = 0.05$. Here n denotes the sample size under the normal model and is distinguished from the

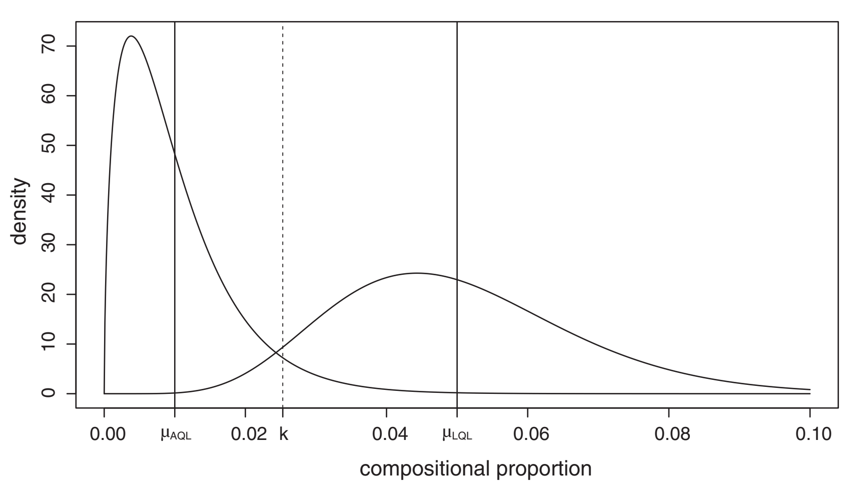

Fig. 1. Design example of Ref. [2].