ON THE PROBABLE NUMBER OF JELLY BEANS IN A POUND

BU-252-M D. S. Robson March, 1968

Abstract

If a population of objects is sampled by weight, i.e. if a sample of preassigned weight t is drawn at random then the number N_t of objects in the sample is a random variable. Neglecting excesses over the boundary,

$$P(N_t = n) = F_n(t) - F_{n+1}(t)$$

where $F_n(t)$ is the n-fold convolution of the weight distribution in the population. When t is large relative to the weight of a single object then the Central Limit Theorem gives

$$P(n_1 \leq N_t \leq n_2) \doteq \Phi\left(\frac{t-n_1\mu}{\sigma\sqrt{n_1}}\right) - \Phi\left(\frac{t-(n_2+1)\mu}{\sigma\sqrt{n_2+1}}\right)$$

where μ is the mean weight per object in the population, σ^2 is the variance in weight in the population, and Φ is the standard cumulative normal distribution function.

Biometrics Unit, Department of Plant Breeding and Biometry, Cornell University.

ON THE PROBABLE NUMBER OF JELLY BEANS IN A POUND

BU-252-M D. S. Robson March, 1968

In sampling a population of objects the sample size is sometimes determined by total weight rather than total number. For example, in agronomic experiments a predetermined number of grams of seed may be planted in each experimental plot. We consider here the probability distribution of the number of objects N_t in a random sample of predetermined total weight t.

Let $F_n(t)$ denote the n-fold convolution of the probability distribution of the weights of the individual objects in the population,

$$F_n(t) = P \text{ (total weight of n randomly chosen objects in less than t).}$$

Since the weight of an individual object is non-negative, we then have

$$P(N_t = n) = F_n(t) - F_{n+1}(t) .$$

A small difficulty arises in the case n=0,

$$P(N_t = 0) = 1 - F_1(t)$$

$$= P \text{ (wt. of a single randomly chosen object exceeds t)}$$

and we ignore this boundary problem by assuming that t is very large relative to the weight of a single object, as is the usual case in sampling-by-weight experiments.

This same assumption implies that N_t is stochastically a large number and that we may therefore invoke the Central Limit Theorem to obtain the approximation

$$P(N_t = n) \doteq \Phi\left(\frac{t-n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{t-(n+1)\mu}{\sigma\sqrt{n+1}}\right)$$

where μ denotes the mean weight per object in the population, σ^2 is the variance in weight among individual objects in the population, and Φ is the standard cumulative normal distribution function. Thus, if z_α is defined by $\Phi(z_\alpha) = 1 - \alpha/2$ then

$$P\left\{\frac{t}{\mu} + \frac{z_\alpha}{2}\left(\frac{\sigma}{\mu}\right)^2 - 1 - \frac{\sigma}{\mu}\sqrt{\frac{tz_\alpha}{\mu} + \left(\frac{\sigma z_\alpha}{2\mu}\right)^2} \leq N_t\right.$$

$$\left.\leq \frac{t}{\mu} + \frac{z_\alpha}{2}\left(\frac{\sigma}{\mu}\right)^2 + \frac{\sigma}{\mu}\sqrt{\frac{tz_\alpha}{\mu} + \left(\frac{\sigma z_\alpha}{2\mu}\right)^2}\right\} \doteq 1 - \alpha \quad .$$

Note that another more crude approximation to the distribution of N_t is then given by

$$P(N_t = n) = -\frac{d}{dn}\Phi\left(\frac{t-n\mu}{\sigma\sqrt{n}}\right) \quad .$$