Airo International Research Journal

Volume XVI, ISSN: 2320-3714
April, 2018
Impact Factor 0.75 to 3.19

UGC Approval Number 63012

A STUDY ON THE THEORY OF PROBABILITY DISTRIBUTIONS

Dr. Veena Taneja
Department of Commerce
Dr. Ashwini Kumar Nagpal
Professor in mathematics
Seth G. L. Bihani S. D. P. G. College, Sri Ganganagar (Rajasthan)

ABSTRACT

In probability theory and statistics, a probability distribution is a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment. In more technical terms, the probability distribution is a description of a random phenomenon in terms of the probabilities of events. For instance, if the random variable X is used to denote the outcome of a coin toss ("the experiment"), then the probability distribution of X would take the value 0.5 for X = heads, and 0.5 for X = tails (assuming the coin is fair). Examples of random phenomena can include the results of an experiment or survey.

KEYWORDS: Probability, Distribution, Sample

INTRODUCTION

A probability distribution is specified in terms of an underlying sample space, which is the set of all possible outcomes of the random phenomenon being observed. The sample space may be the set of real numbers or a set of vectors, or it may be a list of non-numerical values; for example, the sample space of a coin flip would be {heads, tails}.

Probability distributions are generally divided into two classes. A discrete probability distribution (applicable to the scenarios where the set of possible outcomes is discrete, such as a coin toss or a roll of dice) can be encoded by a discrete list of the probabilities of the outcomes, known as a probability mass function. On the other hand, a continuous probability distribution (applicable to the scenarios where the set of possible outcomes can take on values in a continuous range (e.g. real numbers), such as the temperature on a given day) is typically described by probability density functions (with the probability of any individual outcome actually being 0). The normal distribution is a commonly encountered continuous probability distribution. More complex experiments, such as those involving stochastic processes defined in continuous time, may demand the use of more general probability measures.

A probability distribution whose sample space is one-dimensional (for example real numbers, list of labels, ordered labels or binary) is called univariate, while a distribution whose sample space is a vector space of dimension 2 or more is called multivariate.

A univariate distribution gives the probabilities of a single random variable taking on various alternative values; a multivariate distribution (a joint probability distribution) gives the probabilities of a random vector – a list of two or more random variables – taking on various combinations of values. Important and commonly encountered univariate probability distributions include the binomial distribution, the hypergeometric distribution, and the normal distribution. The multivariate normal distribution is a commonly encountered multivariate distribution.

To define probability distributions for the simplest cases, it is necessary to distinguish between discrete and continuous random variables.

In the discrete case, it is sufficient to specify a probability mass function {\displaystyle p} assigning a probability to each possible outcome: for example, when throwing a fair dice, each of the six values 1 to 6 has the probability 1/6. The probability of an event is then defined to be the sum of the probabilities of the outcomes that satisfy the event; for example, the probability of the event "the dice rolls an even value" is

$$p(2)+p(4)+p(6)=1/6+1/6+1/6=1/2.$$

THEORY OF PROBABILITY DISTRIBUTIONS

{\displaystyle p(2)+p(4)+p(6)=1/6+1/6+1/6=1/2.}

In contrast, when a random variable takes values from a continuum then typically, any individual outcome has probability zero and only events that include infinitely many outcomes, such as intervals, can have positive probability. For example, the probability that a given object weighs exactly 500 g is zero, because the probability of measuring exactly 500 g tends to zero as the accuracy of our measuring instruments increases. Nevertheless, in quality control one might demand that the probability of a "500 g" package containing between 490 g and 510 g should be no less than 98%, and this demand is less sensitive to the accuracy of measurement instruments.

Continuous probability distributions can be described in several ways. The probability density function describes the infinitesimal probability of any given value, and the probability that the outcome lies in a given interval can be computed by integrating the probability density function over that interval. On the other hand, the cumulative distribution function describes the probability that the random variable is no larger than a given value; the probability that the outcome lies in a given interval can be computed by taking the difference between the values of the cumulative distribution function at the endpoints of the interval. The cumulative distribution function is the anti-derivative of the probability density function provided that the latter function exists.

The concept of the probability distribution and the random variables which they describe underlies the mathematical discipline of probability theory, and the science of statistics. There is spread or variability in almost any value that can be measured in a population (e.g. height of people, durability of a metal, sales growth, traffic flow, etc.); almost all measurements are made with some intrinsic error; in physics many processes are described probabilistically, from the kinetic properties of gases to the quantum mechanical description of fundamental particles. For these and many other reasons, simple numbers are often inadequate for describing a quantity, while probability distributions are often more appropriate.

Several more specific example of an application:

- The cache language models and other statistical language models used in natural language processing to assign probabilities to the occurrence of particular words and word sequences do so by means of probability distributions.

- Probabilistic load flow in power-flow study explains the uncertainties of input variables as probability distribution and provide the power flow calculation also in term of probability distribution
- Formally, if X is a continuous random variable, then it has a probability density function $f(x)$, and therefore its probability of falling into a given interval, say [a, b] is given by the integral

$$\mathrm{P}[a \leq X \leq b] = \int_a^b f(x)\, dx$$

In particular, the probability for X to take any single value a (that is $a \leq X \leq a$) is zero, because an integral with coinciding upper and lower limits is always equal to zero.

The definition states that a continuous probability distribution must possess a density, or equivalently, its cumulative distribution function be absolutely continuous. This requirement is stronger than simple continuity of the cumulative distribution function, and there is neither a special class of distributions, singular distributions, which are neither continuous nor discrete nor a mixture of those. An example is given by the Cantor distribution. Such singular distributions however are never encountered in practice.

Note on terminology: some authors use the term "continuous distribution" to denote the distribution with continuous cumulative distribution function. Thus, their definition includes both the (absolutely) continuous and singular distributions.

DISCUSSION

By one convention, a probability distribution {\displaystyle \,\mu } is called continuous if its cumulative distribution function {\displaystyle F(x)=\mu (-\infty ,x]} is continuous and, therefore, the probability measure of singletons {\displaystyle \mu \{x\}\,=\,0}. {\displaystyle \,x}.

Another convention reserves the term continuous probability distribution for absolutely continuous distributions. These distributions can be characterized by a probability density function: a non-negative Lebesgue integrable function {\displaystyle \,f} defined on the real numbers such that

$$F(x) = \mu(-\infty, x] = \int_{-\infty}^{x} f(t)\, dt.$$

When the quality characteristic under consideration is an attribute, the sampling plan that is used to make a decision on the disposition of the lots of manufactured products is known as lot-by-lot sampling by attributes. When the quality characteristic is measurable on a continuous scale, the associated sampling plan is known as lot-by-lot sampling by variables. When production is continuous, the formation of lots for lot-by-lot inspection may be impracticable or somewhat artificial. Inspection procedures which have been developed for such situations are termed as continuous sampling plans. Special purpose plans are the sampling inspection procedures defined under the conditions that (a) the production is steady so that the results of past, current and future lots are broadly indicative of a continuing process, (b) the lots are submitted substantially in the order of their production; and (c) inspection is by attributes with quality defined in terms of a fraction nonconforming.

A sampling plan is usually specified by one or more parameters such as sample size(s) and acceptance number(s), with which it is operated for making a decision on the lot. A sampling scheme or system is designated by a set of specified switching rules besides the sample size(s) and acceptance number(s). The discriminatory power of a sampling plan or a scheme or a system is revealed by its operating characteristic (OC) curve. Average outgoing quality limit, average sample number and average total inspection are other measures of assessing the performance of sampling plans. The determination of the parameters of a sampling plan or a scheme subject to certain conditions providing protection to the producer as well as to the consumer is called designing a sampling plan or scheme. It is essential to ensure that the producer is protected from the rejection of the submitted lots which according to his production process are satisfactory ones and the consumer is safeguarded from receiving lots which consist of poor quality items.

Hence, sampling plans should be generally derived with the objective of providing protection to the producer and the consumer.

A common approach to design an acceptance sampling plan or scheme is to require that the corresponding OC curve should pass through two designated points, namely, acceptable quality level and limiting quality level associated with the producer's risk and the consumer's risk, respectively. These points would be considered to select the curve in accordance with a desired degree of discrimination.

Other criteria such as average outgoing quality limit, average sample number and average total inspection are also followed in designing a plan or a scheme.

This research has focused on the designing aspects of lot-by-lot sampling by attributes and special purpose plans by attributes as well as variables for a prescribed set of requirements providing protection to the producer as well as the consumer. A detailed study on the evaluation of repetitive group sampling inspection plans, skiplot sampling plans with single sampling plan by attributes and by variables as the reference plans, skip-lot sampling plans with double and repetitive group sampling plans by attributes as the reference plans and tightened - normal - tightened sampling schemes is made.

The plans under this system are generally viewed as an extension and application of continuous sampling plan (CSP) to lots. Such types of plans are used when the quality of the submitted product is good as demonstrated by the producer's quality history. A specific skip-lot plan under its operation uses a given lot inspection plan by attributes called the reference sampling plan. Single, double and repetitive group sampling plans can be used as the reference plans under skip-lot sampling procedure.

CONCLUSION

Probability theory is the foundation for statistical inference. A probability distribution is a device for indicating the values that a random variable may have. There are two categories of random variables. These are discrete random variables and continuous random variables. A continuous variable can assume any value within a specified interval of values assumed by the variable. In a general case, with a large number of class intervals, the frequency polygon begins to resemble a smooth curve. A continuous probability distribution is a probability density function. The area under the smooth curve is equal to 1 and the frequency of occurrence of values between any two points equals the total area under the curve between the two points and the x-axis.

REFERENCES

- American National Standard Institute/American Society for Quality Control (ANSI/ASQC) Standard A2. Terms, Symbols and Definitions for Acceptance Sampling, American Society for Quality Control, Milwaukee, Wisconsin, US.
- Army Chemical Center. Master Sampling Plans for Single, Duplicate, Double and Multiple Sampling, Manual No. 2, United States Department of Army, Chemical Corps Engineering Agency, US.
- Balamurali, S., and Chi-Hyuck, J. (2006). Repetitive Group Sampling Procedure for Variables Inspection, Journal of Applied Statistics, 33(10), pp. 327 - 338.
- Balamurali, S., Heckon, P., Chi-Hyuck, J., Kwang-Jae, K., and Jaewook, L. (2015).
- Designing of Variables Repetitive Group Sampling Plan Involving Minimum Average Sample Number, Communications in Statistics - Simulation and Computation, 34, pp. 799 - 809.
- Brugger (2015), R. M. A Simplification for Skip-lot Procedure Formulation, Journal of Quality Technology, 7(4), pp. 165 - 167.
- Burgess (2014). A Graphical Method of Determining a Single Sampling Plan, Industrial Quality Control, 4(6), pp. 25 - 27.
- Calvin, T.W (2013). TNT Zero Acceptance Number Sampling, American Society for Quality Control Annual Technical Conference Transactions, Philadelphia, PA, pp. 35 - 39.
- Cameron, J. M (2014). Tables for Constructing and for Computing the Operating Characteristics of Single Sampling Plans, Industrial Quality Control, 9(1), pp. 37 - 39.
- Carr, W. E (2015). Sampling Plan Adjustment for Inspection Error and Skip-lot Plans, Journal of Quality Technology, 14(3), pp. 105 - 116.
- Case, K. E., and Keats, J. B (2012). On the Selection of a Prior Distribution in Bayesian Acceptance Sampling, Journal of Quality Technology, 14(1), pp.10 - 18.
- Chakraborty, T. K (2014). Single Sampling LTPD Plans as a Stochastic Programming Problem, Sankhya - Series B, 54, pp. 242 - 248.
- Cox, D. C (2013). Skip-Lot Sampling Plans, Quality, 21(3), pp. 105 - 116.
- Derman, C., Littauer, S., and Solomon, H (2012). Tightened Multi-Level Continuous Sampling Plans, Annals of Mathematical Statistics, 28(2), pp. 395 - 404.