RESEARCHERID THOMSON REUTERS
International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 12, December- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

Some important Truncated Discrete/Continuous Distributions

Dr. J. Purushotham, Asst. Professor (c)
Department of Applied Statistics, Telangana University, Nizamabad - 503322.

Abstract: This paper deals with the truncation of the discrete distributions Poison, Binomial and continuous distributions Exponential, Normal. The truncation is very much useful when some values of a random variable are missing. In such a case we can truncate the random variable beyond such values one side (lower or upper) or two side and can determine the probability mass function (in case of discrete distribution) and probability density function (in case of continuous distribution). Here, we dealt with some important characteristics like mean, variance and moment generating function of the above said distributions. Also, we consider applications of the truncated distributions in the various fields.

Key words: Truncated distributions, Poison, Binomial, Normal distributions.

Introduction

Truncated distributions arise in practical statistics in cases where the ability to record, or even to know about, occurrences is limited to values which lie above or below a given threshold or within a specified range. In other words, it may happen, either by accident or design that information about some of the variable values are missing. In such a case we use truncation. In truncation, the value which falls outside the interval (a,b) is ignored. For example, if the dates of birth of children in a school are examined, these would typically be subject to truncation relative to those of all children in the area given that the school accepts only children in a given age range on a specific date. There would be no information about how many children in the locality had dates of birth before or after the school's cutoff dates if only a direct approach to the school were used to obtain information.

Where sampling is such as to retain knowledge of items that fall outside the required range, without recording the actual values, this is known as censoring, as opposed to the truncation here.

The problem of truncation arise s when a standard statistical model is appropriate for analysis except the values of the random variable failing below or above some value are not measured at all.

Definition: Let X be a random variable with probability mass function (p.m.f) or probability density function (p.d.f). The distribution of X is said to be truncated at the point X=a, if all the values of X ≤ a are discarded. Hence the probability function g(x) of the truncated random variable distribution at X = a is given by

$$g(x) = \frac{f(x)}{P(X>a)}, X > a$$

$$= \frac{f(x)}{\sum_{x>a} f(x)} \ , \ X > a \ \ (for\ discrete\ distribution)$$

$$= \frac{f(x)}{\int_a^\infty f(x)dx} \ , X > a \ \ (for\ continuous\ distribution)$$

If a random variable allowed to assume values on from its lower boundary to specified value b i.e, discarding or removing all possible values of random variable behind X = b is known as right truncation or truncated above and for such truncation the p.m.f or p.d.f is given by

$$g(x) = \frac{f(x)}{P(X<b)} \ , X > b$$

If any random variable is restricted to assume values between X = a and X = b, then the random variable is said to be doubly truncated and for such truncation the p.m.f or p.d.f is given by

$$g(x) = \frac{f(x)}{P(a<X<b)} \ , a < X > b$$

The r[th] moment about origin (non-central moments) for the left truncation distribution is given by

$$\mu_r' = E(X^r) = \sum_a^\infty g(x) \ \ (for\ discrete\ distribution)$$
$$\mu_r' = E(X^r) = \int_a^\infty g(x)dx \ \ (for\ continuous\ distribution)$$

I. Truncated Discrete Distributions

A) Truncated Binomial Distribution: The Binomial distribution is supposed to be the oldest discrete distribution. Binomial distribution is defined by a Bernoulli trial. If a Bernoulli trial is repeated number of times say n with probability of outcome (success) in each trial is constant, then the probability of getting x successes in n trials is given by

$$f(x) = P(X = x) = n_{c_x}\, p^x q^{n-x} \ ; \ \ x = 0,1,2, \dots\dots\dots n; 0 < p < 1$$

Truncation of any distribution may be single (or one sided) or it may be doubly truncated by omitting values on both extremes. A single or one sided binomial distribution is formed if the distribution is truncated at only one end.

Consider, a truncated binomial distribution truncated at X = 0, then its p.m.f is given by

$$g(x) = \frac{f(x)}{P(X>0)} \ , X > o \ i.e, x = 1,2, \dots\dots\dots n$$
$$= \frac{n_{c_x}\, p^x q^{n-x}}{\sum_{x=1}^n f(x)}$$
$$= \frac{n_{c_x}\, p^x q^{n-x}}{\sum_{x=1}^n n_{c_x}\, p^x q^{n-x}}$$
$$= \frac{n_{c_x}\, p^x q^{n-x}}{1-q^n}$$

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 12, December- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

Hence, the truncated p.m.f of Binomial distribution, truncated at X=0 is given by

$$g(x) = \frac{n_{c_x}\, p^x q^{n-x}}{1-q^n} \;;\; x = 1,2, \ldots . n \quad \text{--------} > \ (*)$$

Some times the above is known as positive binomial distribution.
In general, a binomial distribution truncated on the let at X = k is given by

$$g(x) = \frac{f(x)}{P(X>k)}\ , X > k\ i.e, x = k+1, k+2, \ldots \ldots \ldots n$$

$$= \frac{n_{c_x}\, p^x q^{n-x}}{\sum_{x=k}^{n} n_{c_x}\, p^x q^{n-x}} ; x = k+1, k+2, \ldots \ldots . n$$

Consider a binomial distribution in which k observations from the lower end and h observations from the upper end of the distribution are truncated. Then the p.m.f. of the double truncated binomial distribution is given by

$$g(j) = = \frac{n_{c_j}\, p^j q^{n-j}}{\sum_{j=k}^{n-h} n_{c_j}\, p^j q^{n-j}} ;\ k < j < n - h$$

The r^{th} moment about origin (non-central) of truncated binomial distribution are given by dividing the corresponding moment of the complete distribution by $1-q^n$ (by * defined above)

$$\mu_r' = \frac{r^{th}\, moment\ \ of\ binomial\ \ random\ \ variable}{1-q^n}$$

Thus, we get $\mu_1' = mean = \frac{np}{1-q^n}$

$$\mu_2' = \frac{n(n-1)p^2 + np}{1-q^n}$$

and Variance = $\mu_2 = \mu_2' - \mu_1'^2 = \frac{npq\,(1-n)}{1-q^n}$

B) Truncated Poison distribution

Poison distribution was discovered by French mathematician and physicist Simeon Denis Poison (1937). Poison distribution is a limiting case of Binomial distribution under the following conditions.

i) n, the number of trials are very large i.e., $n \to \infty$
ii) p, the probability of success is very small i.e., $p \to 0$
iii) $np = \theta$ (say) is finite.

Under the above conditions,

RESEARCHERID THOMSON REUTERS
International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 12, December- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

$$\lim_{n\to\infty} b(x;n,p) = \lim_{n\to\infty} \frac{n!}{x!(n-x)!}\left(\frac{\theta}{n}\right)^x \left(1-\frac{\theta}{n}\right)^{n-x}$$

$$= \frac{e^{-\theta}\theta^x}{x!} \; ; \qquad x = 0,1,\cdots,\infty$$

Which is the probability function of the Poison distribution with parameter θ.

The right truncated poison distribution is formed if the poison distribution truncated at only right. In this case the probability mass function is given by

$$P(X=k) = \frac{e^{-\theta}\theta^k}{k!\sum_{j=0}^{r}\frac{e^{-\theta}\,\theta^j}{j!}}; \qquad k = 0,1,2,\cdots,r$$

If the first r_1 values 0,1,2,……r_1-1 are omitted then we call it as a left truncated poison distribution and its probability mass function is given by

$$P(X=k) = \frac{e^{-\theta}\theta^x}{k!\left[1-e^{-\theta}\sum_{j=0}^{r_1-1}\frac{\theta^j}{j!}\right]}; \qquad k = r_1, r_1+1,\cdots$$

A very special case of left truncation is the omission of zero, then the corresponding truncated poison distribution is given by

$$P(X=k) = \frac{e^{-\theta}(1-e^{-\theta})\theta^k}{k!} = \frac{(e^{\theta}-1)^{-1}\theta^k}{k!}; \qquad k = 1,2,\cdots$$

This is usually called as positive poisson distribution.
The first and second moment about origin of truncated poison distribution truncated at X=0 is given by

$$\mu_1' = mean = \theta(1-\tfrac{1}{e^{\theta}})^{-1}$$
$$\mu_2' = (\theta^2+\theta)(1-\tfrac{1}{e^{\theta}})^{-1}$$

The variance of truncated poison distribution truncated at X = 0 is

$$\text{Variance} = \mu_2 = \mu_2' - \mu_1'^2 = \theta^2(1-\tfrac{1}{e^{\theta}})^{-1}\left[1-(1-\tfrac{1}{e^{\theta}})^{-1}+\tfrac{1}{\theta}\right]$$

II. Truncated Continuous distributions

A random variable X is said to follow normal distribution with parameters mean (μ) and variance (σ^2) if its probability density function is given by

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 12, December- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

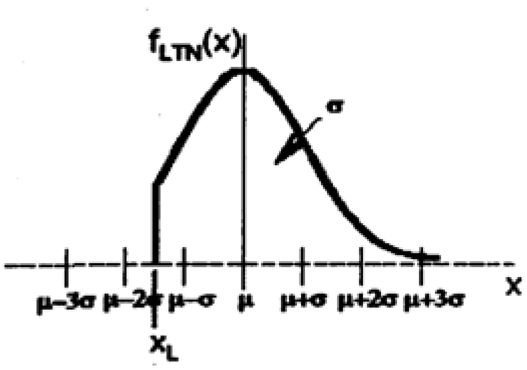
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; \quad -\infty\le x\le\infty;\sigma>0$$

English Mathematician De-Moivre (1733) obtained this continuous distribution as a limiting case of the Binomial distribution and applied it to problems arising in the game of chance. The common use of normal distribution is in model construction, but enumerating the field of application would be lengthy and not really informative and therefore we do not attempt task. The normal distribution is almost used as approximation to theoretical or an unknown distribution.

If the value of x below value X_L cannot be observed due to truncation, then the resulting distribution is left-truncated normal distribution with probability density function and is given by

$$f(x)=\begin{cases}0, & -\infty\le x\le x_L\\ \dfrac{f(x)}{\int_{x_L}^{\infty}f(x)\,dx} & x_L\le x\le\infty\end{cases}$$

Where f(x) is defined above. The left truncated normal distribution graphically shown as below.



If the values of X above some value X_R cannot be observed due to truncation, then the resulting distribution is a right truncated normal distribution with probability density function given by

$$f(x)=\begin{cases}0, & x_R\le x\le\infty\\ \dfrac{f(x)}{\int_{-\infty}^{x_R}f(x)\,dx} & -\infty\le x\le x_R\end{cases}$$

A random variable X has a doubly truncated normal distribution if its probability density function is given by

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 12, December- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

$$f_Y(y)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\left[\frac{1}{\sigma\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\right]^{-1} \qquad a\le x\le b$$

Conclusion: This paper is presented as an overview on truncated discrete and continuous distributions.

References:

[1]. Cohen, A. (1949) "On estimating the mean and standard deviation of truncated normal distribution", Journal of American Statistical Association.
[2]. Shamsur Rahman (2004) " Truncated distributions and their Applications", M.Phil Desertation.
[3]. Cramer G 1975 Mathematical Methods of Statistics (Moscow: Mir) p 631
[4]. Vadzinsky R N 2001 Handbook of Probability Distributions (SanktPetersburg: Science)
[5]. Tokmachev M S, Ryazantsev P P 2010 Simulation of truncated distributions Bulletin of NovSU. No. 55 pp 34–36
[6]. B.C. Arnold, N. Balakrish, H.N. Nagarajua. First Course in order statistics. John Wiely & Sons, New York, 2008.
[7]. A.E. Cohen. Estimation in truncated Poisson distribution when zero and some ones are missing. J. Amer. Stat. Ass., 1960, 55: 342 - 348.
[8]. B.R. Crain. Estimating the parameters of a truncated normal distribution. Applied Mathematics and Computation, 1979, 5: 149 - 158.
[9]. Hassan Okasha, M.M. Mohie-El-Din, Mohamed Mahmoud. On mid-truncated distributions and its applications in order statistic. J. Adv. Res. Stat. Probab., 2011, 3: 47 - 57.