Global Journal of Mathematical Sciences: Theory and Practical.
ISSN 0974-3200 Volume 5, Number 4 (2013), pp. 237-243

http://www.irphouse.com

Empirical Distribution of the Sample Mean Based on Uniform (0, 1) Random Samples

Yousef M. Emhemmed and Sumaya S. Eljabri

Department of Statistics, Faculty of Science
University of Tripoli, Tripoli, Libya
E-mail: emhemmedy@yahoo.co.uk

Abstract

The Uniform distribution on the interval (0, 1) plays an important role in many statistical applications, such as, its role in simulation procedures when a sequence of random numbers needs to be generated from some parent population. This paper investigates the sampling distribution of the sample mean for random samples drawn from a uniform (0, 1) population. A complete derivation of the probability density function of the sample mean is presented. The normal approximation to the series form of the probability density function of the sample mean is also discussed. To avoid the use of the complicated series form of the probability density function of the sample mean for small sample sizes when the normal approximation is not advisable, the tables of the cumulative distribution function for sample sizes 2, 3, 4 and 5 are constructed. The Minitab statistical software is used throughout this paper.

Keywords: Cumulative Distribution Function (cdf); Asymptotic Distribution.

1. Introduction

The aim is to derive the probability density function of the sample mean of a random sample X_1, X_2, ...,X_n drawn from a uniform distribution on the interval (0, 1). The principal of mathematical induction can be used to derive the general form of the probability density function of the sample mean.

As the sample size increases, the shape of the exact distribution of the sample mean approaches the normal distribution, the approach to normality is seen to be obvious, even for small sample sizes, Mood, Graybill and Boes (1974). This sensitivity to the normal distribution directed our attention to search for an appropriate sample size, at which the approximation to the normal is adequate. The advantage is to avoid the use of the complicated original series form of the sample mean distribution and instead the normal could provide an alternative which could be used accurately. Probability tables of the cdf of the sample mean for small sample sizes are also implemented. The cdf tables provide an alternative way to be used instead of integrating the series form of the exact pdf of the sample mean for small sample sizes.

2. The Exact Distribution of the Sample Mean

This section illustrates the different forms of the probability density function of the sample mean of random samples drawn from the uniform distribution over the interval (0, 1) for different sample sizes. One way to derive this probability density function is to use the indicator function technique together with the convolution formula of the sample mean, Let U be a random variable denoting the sample mean for random sample of size n, then for n=2, the probability density function is given by:

$$f_U(u) = 4u\, I_{(0,\frac{1}{2})}(u) + (4-4u)\, I_{(\frac{1}{2},1)}(u),$$

or equivalently can be written as,

$$f_U(u) = \begin{cases} 4u & ;\ 0 < u < \frac{1}{2}, \\ 4-4u & ;\ \frac{1}{2} < u < 1, \\ 0 & ;\ \text{o.w.} \end{cases}$$

$$= \begin{cases} \frac{2}{(2-1)!}(2u)^{2-1} & ;\ 0 < u < \frac{1}{2}, \\ \frac{2}{(2-1)!}\left[(2u)^{2-1} - \binom{2}{1}(2u-1)^{2-1}\right] & ;\ \frac{1}{2} < u < 1, \\ 0 & ;\ \text{o.w.} \end{cases}$$

The probability density function for n=3, is given by,

$$f_U(u) = \begin{cases} \frac{3}{(3-1)!}(3u)^{3-1} & ;0<u<\frac{1}{3} \\ \frac{3}{(3-1)!}\left[(3u)^{3-1} - \binom{3}{1}(3u-1)^{3-1}\right] & ;\frac{1}{3}<u<\frac{2}{3} \\ \frac{3}{(3-1)!}\left[(3u)^{3-1} - \binom{3}{1}(3u-1)^{3-1} + \binom{3}{2}(3u-2)^{3-1}\right] & ;\frac{2}{3}<u<1 \\ 0 & ;\text{o.w.} \end{cases}$$

Similarly the probability density function for n=4, is shown to be of the form,

$$f_U(u)=\begin{cases} \dfrac{4}{(4-1)!}(4u)^{4-1} & ;\ 0<u<\dfrac{1}{4} \\ \dfrac{4}{(4-1)!}\left\{(4u)^{4-1}-\binom{4}{1}(4u-1)^{4-1}\right\} & ;\ \dfrac{1}{4}<u<\dfrac{1}{2} \\ \dfrac{4}{(4-1)!}\left\{(4u)^{4-1}+\sum\limits_{i=1}^{2}(-1)^i\binom{4}{i}(4u-i)^{4-1}\right\} & ;\ \dfrac{1}{2}<u<\dfrac{3}{4} \\ \dfrac{4}{(4-1)!}\left\{(4u)^{4-1}+\sum\limits_{i=1}^{3}(-1)^i\binom{4}{i}(4u-i)^{4-1}\right\} & ;\ \dfrac{3}{4}<u<1 \\ 0 & ;\ \text{o.w.} \end{cases}$$

According to forms of the probability density function of the sample mean for n= 2, 3 and 4, the general form of the probability density function of any sample size n could then be derived using the principle of the mathematical induction and it takes the following series form,

$$\begin{aligned} f_{\overline{X}_n}(x) &= \frac{n}{(n-1)!}\sum_{k=0}^{n-1}\sum_{r=0}^{k}(-1)^r\binom{n}{r}(nx-r)^{n-1}\, I_{(\frac{k}{n},\frac{k+1}{n})}(x) \\ &= \frac{n}{(n-1)!}\sum_{k=0}^{n-1}\left\{(nx)^{n-1}-\binom{n}{1}(nx-1)^{n-1}+\binom{n}{2}(nx-2)^{n-1}\right. \\ &\qquad \left. -\ldots+(-1)^{n-1}\binom{n}{n-1}[nx-(n-1)]^{n-1}\right\} I_{(\frac{k}{n},\frac{k+1}{n})}(x), \end{aligned}$$

3. Asymptotic Distribution of the Sample Mean:

Since the distribution of a sample mean when sampling from the uniform (0, 1) distribution takes a series form which depends on the sample size n. This distribution presents computational difficulties, especially for large values of n, and hence few, if any, of us would be interested in using it to compute probabilities about the sample mean. One of the objects of this paper is to provide an approximate form of the pdf of the sample mean for large values of n. The central limit theorem allows us to obtain the asymptotic distribution of the sample mean. The random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed as the uniform (0, 1), that is $\mu=1/2$ and $\sigma^2=1/12$ and hence, for large n, the sample mean has a normal distribution with mean 1/2 and variance 1/12n. The exact (dotted curves) and asymptotic probability (solid curves) density functions of the sample mean the sample mean when sampling from the uniform (0, 1) distribution are sketched in Figures (3.1-3.4), and an approach to normality can be observed even for small sample sizes. Furthermore, all the inflection points of both exact and asymptotic curves occur at the same point. The sketched curves may allow for direct visual comparisons and may also provide some sort of evidence for an approach to normality even for small sample sizes. Similar discussion is given in Mood, Graybill and Boes (1974). Although the approach to normality in these figures is evidently clear even for very small sample sizes, goodness of fit tests could be obtained.

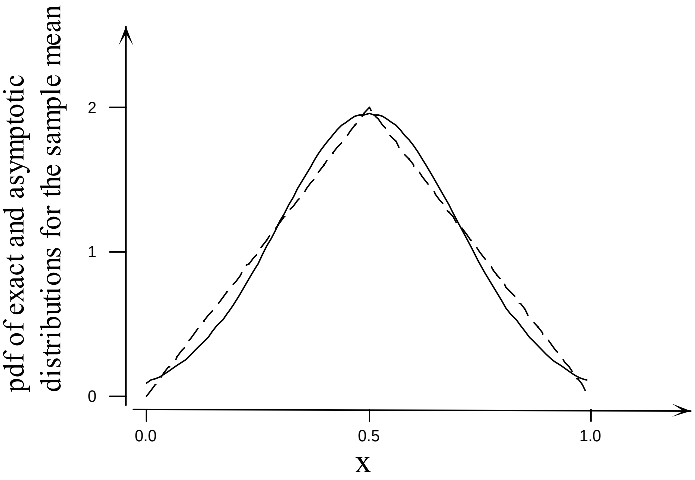

Fig (3.1) The pdf of the exact (dotted line) and asymptotic (solid line) distributions of sample mean with sample size 2.

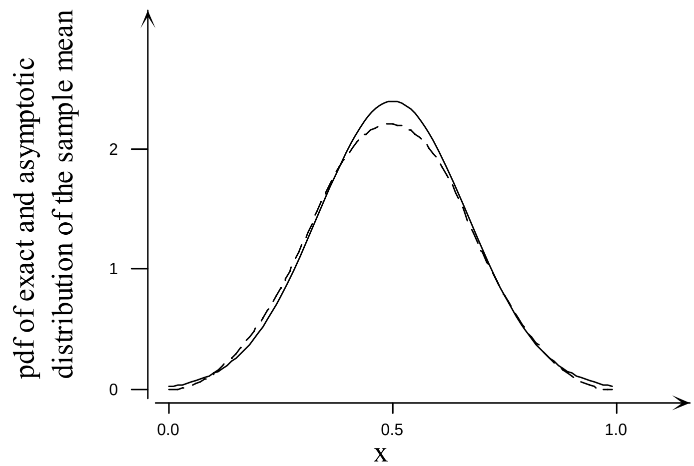

Fig (3.2) The pdf of the exact (dotted line) and asymptotic (solid line) distributions of sample mean with sample size 3.

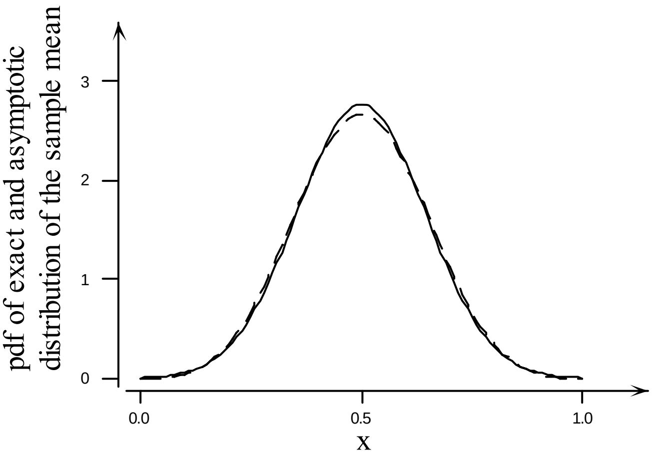

Fig (3.3) The pdf of the exact (dotted line) and asymptotic (solid line) distributions of sample mean with sample size 4.

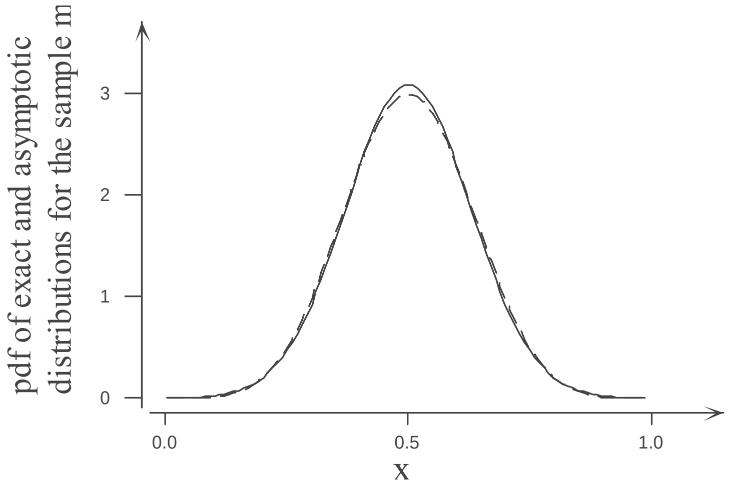

Fig (3.4) The pdf of the exact (dotted line) and asymptotic (solid line) distributions of sample mean with sample size 5.

4. Cumulative Distribution Tables of the Sample Mean:

Many users who seek for accurate results may still prefer to use the exact distribution than to use its normal approximation for small sample sizes. Tables 4.1-4.4 below give the cdf of the exact distribution of the sample mean of random samples of sizes 2, 3, 4 and 5, respectively, drawn from the uniform (0, 1) population. The reason of obtaining these cdf tables is to avoid the tedious work with the exact series form of the pdf of the sample mean and also is to avoid the use of the normal approximation.

Table (4.1) the cdf of the exact distribution of the sample mean of uniform (0, 1) random samples of size 2.

$\bar{x}$	0.00	0.010	0.020	0.030	0.040	0.050	0.06	0.070	0.080	0.090
0.0	0.00	0.000	0.001	0.002	0.003	0.005	0.007	0.01	0.012	0.016
0.1	0.02	0.024	0.029	0.034	0.039	0.045	0.051	0.058	0.065	0.072
0.2	0.08	0.088	0.097	0.106	0.115	0.125	0.135	0.146	0.157	0.168
0.3	0.18	0.192	0.205	0.218	0.231	0.245	0.259	0.274	0.289	0.304
0.4	0.32	0.336	0.353	0.37	0.387	0.405	0.423	0.442	0.461	0.480
0.5	0.50	0.52	0.539	0.558	0.577	0.595	0.613	0.630	0.647	0.664
0.6	0.68	0.696	0.711	0.726	0.741	0.755	0.769	0.782	0.795	0.808
0.7	0.82	0.832	0.843	0.854	0.865	0.875	0.885	0.894	0.903	0.912
0.8	0.92	0.928	0.935	0.942	0.949	0.955	0.961	0.966	0.971	0.976
0.9	0.98	0.984	0.987	0.990	0.993	0.995	0.997	0.998	0.999	0.999

Table (4.2) the cdf of the exact distribution of the sample mean of uniform (0, 1) random samples of size 3.

$\bar{x}$	0.00	0.010	0.020	0.030	0.040	0.050	0.06	0.070	0.080	0.090
0.0	0.00	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.002	0.003
0.1	0.01	0.006	0.008	0.01	0.012	0.015	0.018	0.022	0.026	0.031
0.2	0.04	0.042	0.048	0.055	0.062	0.070	0.079	0.089	0.099	0.110
0.3	0.12	0.134	0.147	0.162	0.177	0.193	0.210	0.227	0.246	0.264
0.4	0.28	0.304	0.325	0.346	0.367	0.389	0.411	0.433	0.455	0.478
0.5	0.50	0.522	0.545	0.567	0.589	0.611	0.633	0.654	0.675	0.696
0.6	0.72	0.736	0.754	0.773	0.790	0.807	0.823	0.838	0.853	0.866
0.7	0.88	0.890	0.901	0.911	0.921	0.931	0.938	0.945	0.952	0.958
0.8	0.97	0.969	0.974	0.978	0.982	0.985	0.988	0.990	0.992	0.994
0.9	0.99	0.997	0.998	0.998	0.999	0.999	0.999	0.999	0.999	1.000

Table (4.3) the cdf of the exact distribution of the sample mean of uniform (0, 1) random samples of size 4.

$\bar{x}$	0.00	0.010	0.020	0.030	0.040	0.050	0.06	0.070	0.080	0.090
0.0	0.00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
0.1	0.00	0.002	0.002	0.003	0.004	0.005	0.008	0.009	0.011	0.014
0.2	0.02	0.021	0.025	0.030	0.035	0.042	0.049	0.057	0.066	0.075
0.3	0.09	0.098	0.111	0.125	0.140	0.156	0.173	0.191	0.210	0.230
0.4	0.25	0.274	0.296	0.320	0.344	0.369	0.395	0.421	0.447	0.473
0.5	0.50	0.527	0.553	0.579	0.605	0.631	0.656	0.680	0.704	0.727
0.6	0.75	0.770	0.790	0.809	0.827	0.844	0.860	0.875	0.889	0.902
0.7	0.91	0.925	0.935	0.943	0.951	0.958	0.965	0.970	0.975	0.979
0.8	0.98	0.986	0.989	0.991	0.993	0.995	0.996	0.997	0.998	0.998
0.9	0.99	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table (4.4) the cdf of the exact distribution of the sample mean of uniform (0, 1) random samples of size 5.

$\bar{X}$	0.00	0.010	0.020	0.030	0.040	0.050	0.06	0.070	0.080	0.090
0.0	0.00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
0.1	0.00	0.003	0.002	0.003	0.004	0.005	0.007	0.009	0.001	0.014
0.2	0.02	0.021	0.025	0.030	0.035	0.042	0.049	0.057	0.066	0.075
0.3	0.09	0.099	0.111	0.125	0.140	0.156	0.173	0.191	0.210	0.230
0.4	0.25	0.274	0.296	0.320	0.344	0.369	0.395	0.421	0.447	0.473
0.5	0.50	0.527	0.553	0.579	0.605	0.631	0.656	0.680	0.704	0.727
0.6	0.75	0.770	0.790	0.810	0.827	0.844	0.860	0.875	0.889	0.902
0.7	0.91	0.925	0.934	0.943	0.951	0.958	0.965	0.970	0.975	0.979
0.8	0.98	0.986	0.989	0.991	0.993	0.995	0.996	0.997	0.998	0.998
0.9	0.99	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

5. References

[1] Glen A., L. Leemis, and Evans D, 2001, "APPL: A Probability programming language", The American Statistician, 55, Number 2, pp.156- 166.

[2] Johnson N. L. and Kotz S. and Balakrishnan N, 1995, "Continuous uni-distributions", Vol. 2, 2nd ed, Upper Saddle River, NJ: Prentice Hall.

[3] Kendall M.G and Stuart A, "The Advanced theory of statistics", Volume 2, Hafner Press.

[4] L'Ecuyer, P, 1994, "Uniform random number generation", Annals of Operations Research, 53:77-120.

[5] Mood A., Graybill F, and Boes D, 1974, "Introduction to the theory of Statistics", 3rd edition, McGraw-Hill book company, New York.

[6] Stephens, M. S, 1986, "Tests for the uniform distribution, Goodness of Fit Techniques", ed. R. B. D'Agostino and M. S. Stephens, 331-366. Marcel Dekker, New York and Basel.

[7] Tausworthe, R. C, 1965, "Random numbers generated by linear recurrence modulo two. Mathematics of Computation", 19:201-209.

[8] Tezuka, S, 1995, "Uniform Random Numbers: Theory and Practice", Kluwer Academic Publishers, Norwell, Massachusetts.