The Probability Distribution of the Sum of Several Dice approach to a Normal Distribution

Reshma R. Chavan
Assistant Professor, Saraswati College of Engineering, Kharghar, India,
c.reshma2601@gmail.com

Abstract: The normal distributions are a very important *class* of statistical distributions. All normal distributions are symmetric and have bell-shaped density curves with a single peak. The probability distribution of the sum of several dice approach to a normal distribution as the number of dice is increased. In this paper the probability distribution of the sum of dice (for *k >2)* is derived using the method of moment generating function.

Key words : Moment Generating Function, Normal Distribution, Central Limit Theorem

INTRODUCTION

In this paper the method of moment generating function (mgf) is used to derive the probability distribution of the sum of k>2 dice. Finally, the probability distribution of sum of dice approach to normal distribution as the number of dice is increased is showed by using graphs.

METHODOLOGY

The Moment Generating Function:

The Moment Generating Function (mgf) of a random variable X is defined as the expected value or weighted average of the function

$M_X(t) = E(e^{tX})$ for $t \in \mathbb{R}$ ⋯ (1)

If X is a discrete random variable the mgf is $M_X(t) = \sum e^{tx} p(x)$ for all x where $p(x)$ is probability mass function of X. If X is a continuous random variable the mgf is $M_X(t) = \int_{-\infty}^{+\infty} e^{tx} p(x)$ where $p(x)$ is probability density function of X. Suppose X a discrete random variable has the probability distribution:

$f(x_j) = P(X = x_j) = p_j \quad j = 1,2,\dots,k$.

Then the mgf is

$M_X(t) = p_1 e^{tx_1} + p_2 e^{tx_2} + \dots + p_k e^{tx_k}$.

In other words, if the random variable X has the mgf: $M_X(t) = p_1 e^{tx_1} + p_2 e^{tx_2} + \dots + p_k e^{tx_k}$, then its probability is $f(x_j) = P(X = x_j) = p_j \quad j = 1,2,\dots,k$ ⋯ (2)

The Probability Distribution of the Sum of *k* Dice[1-2]:

Consider the experiment of rolling k fair dice, and let X_i represents the number that comes up when *i-th* fair die is rolled, $i = 1,2,\dots,k$. Then the probability distribution of each X_i is given by:

$$f(x) = \begin{cases} \frac{1}{6} & x = 1,2,\dots,6 \\ 0 & otherwise \end{cases}$$

and its moment generating function *(mgf)* is:

$$M_X(t) = E(e^{tX}) = \frac{1}{6}(e^t + e^{2t} + e^{3t} + \dots + e^{6t})$$

Since the random variables $X_1, X_2, \dots, X_6$ are independent, the mgf of the sum S is:

$$M_S(t) = E(e^{tX}) = E\left[e^{t(X_1+X_2+\cdots+X_6)}\right]$$

$$= E[e^{tX_1} \cdot e^{tX_2} \cdots e^{tX_6}] = \prod_{i=1}^{k} E(e^{tX_i})$$

$$= \prod_{i=1}^{k} \left[\frac{1}{6}(e^t + e^{2t} + e^{3t} + \dots + e^{6t})\right]$$

$$= \frac{1}{6^k}(e^t + e^{2t} + e^{3t} + \dots + e^{6t})^k$$

$$M_S(t) = \frac{1}{6^k}(e^t + e^{2t} + e^{3t} + \dots + e^{6t})^k \qquad \cdots (3)$$

We can now use result (3). To obtain the probability distribution of X_k , where $k = 2,3,4$ and 5.

Sum of 2 fair dice (for *k* =2)

$$M_x(t) = \frac{1}{6^2}(e^t + e^{2t} + e^{3t} + \cdots + e^{6t})^2$$

$$= \frac{1}{36}(e^{2t} + 2e^{3t} + 3e^{4t} + 4e^{5t} + 5e^{6t} + 6e^{7t} + 5e^{8t} + 4e^{9t} + 3e^{10t} + 2e^{11t} + e^{12t}) \qquad \cdots (4)$$

From result (2), the probability distribution of sum of 2 fair dice is,

Table 1:

x	2	3	4	5	6	7	8	9	10	11	12
f(x)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Similarly, the mgf $M_x(t)$ of the sum of $k = 3,4,5,6$ dice is obtained from equation (3):

Sum of 3 fair dice (for *k* =3)

$$M_x(t) = \frac{1}{6^3}(e^t + e^{2t} + e^{3t} + \cdots + e^{6t})^3$$

$$= \frac{1}{216}(e^{3t} + 3e^{4t} + 6e^{5t} + 10e^{6t} + 15e^{7t} + 21e^{8t} + 25e^{9t} + 27e^{10t} + 27e^{11t} + 25e^{12t} + 21e^{13t} + 15e^{14t} + 10e^{15t} + 6e^{16t} + 3e^{17t} + e^{18t}) \qquad \cdots (5)$$

Sum of 4 fair dice (for *k* =4)

$$M_x(t) = \frac{1}{6^4}(e^t + e^{2t} + e^{3t} + \cdots + e^{6t})^4$$

$$= \frac{1}{1296}(e^{4t} + 4e^{5t} + 10e^{6t} + 20e^{7t} + 35e^{8t} + 56e^{9t} + 80e^{10t} + 104e^{11t} + 125e^{12t} + 140e^{13t} + 146e^{14t} + 140e^{15t} + 125e^{16t} + 104e^{17t} + 80e^{18t} +$$

ISSN 2229-5518

$56e^{19t} + 35e^{20t} + 20e^{21t} + 10e^{22t} + 4e^{23t} + e^{24t}) \quad \cdots (6)$

Sum of 5 fair dice (for *k* =5)

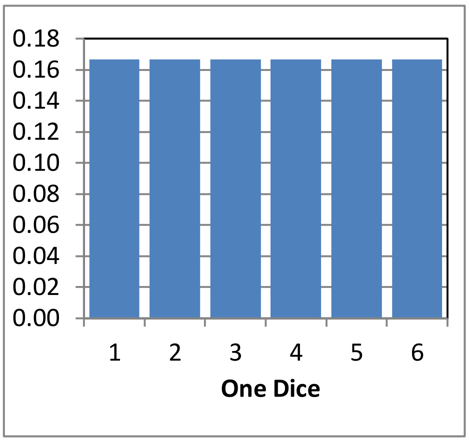
$$M_x(t) = \frac{1}{6^5}(e^t + e^{2t} + e^{3t} + \cdots + e^{6t})^5$$

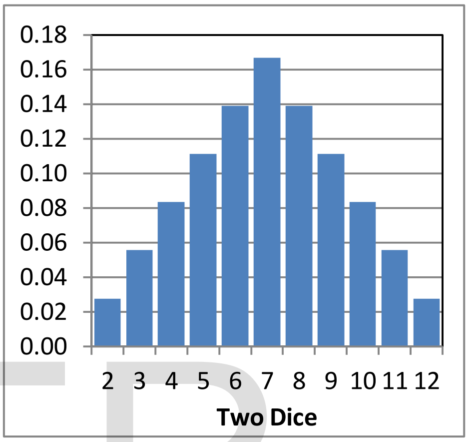
$$= \frac{1}{7776}(e^{5t} + 5e^{6t} + 15e^{7t} + 35e^{8t} + 70e^{9t} + 126e^{10t} + 205e^{11t} + 305e^{12t} + 420e^{13t} + 540e^{14t} + 651e^{15t} + 735e^{16t} + 780e^{17t} + 780e^{18t} + 735e^{19t} + 651e^{20t} + 540e^{21t} + 420e^{22t} + 305e^{23t} + 205e^{24t} + 126e^{25t} + 70e^{26t} + 35e^{27t} + 15e^{28t} + 5e^{29t} + e^{30t}) \quad \cdots (6)$$

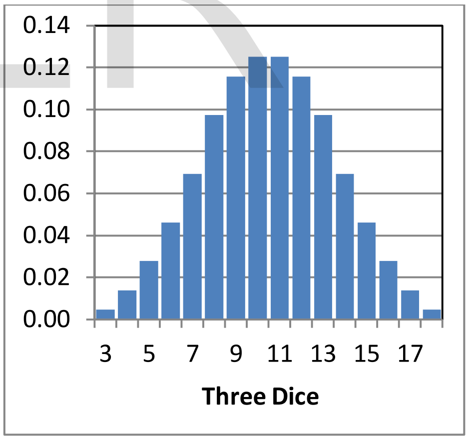
The probability distributions of the sum *S* for *k* = 3, 4, and 5 are easily obtained from the above expressions for mgf, and result (2). Table 2 shows the probability distributions of the sum of *k* dice for *k* = 3, 4, and 5 .

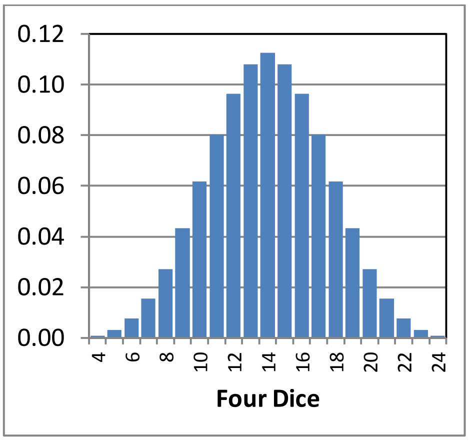
Table 2:

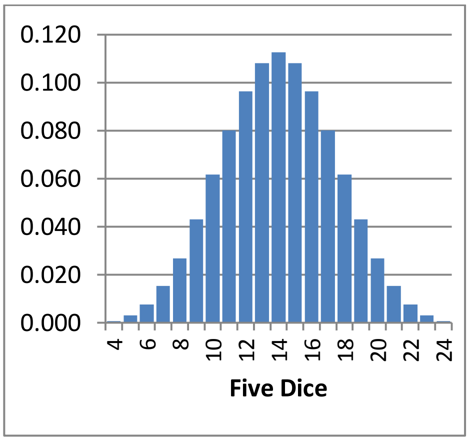
k = 3		*k =4*		*k =5*	
x	*f(x)*	*x*	*f(x)*	*x*	*f(x)*
3	1/216	4	1/1296	5	1/7776
4	3/216	5	4/1296	6	5/7776
5	6/216	6	10/1296	7	15/7776
6	10/216	7	20/1296	8	35/7776
7	15/216	8	35/1296	9	70/7776
8	21/216	9	56/1296	10	126/7776
9	25/216	10	80/1296	11	205/7776
10	27/216	11	104/1296	12	305/7776
11	27/216	12	125/1296	13	420/7776
12	25/216	13	140/1296	14	540/7776
13	21/216	14	146/1296	15	651/7776
14	15/216	15	140/1296	16	735/7776
15	10/216	16	125/1296	17	780/7776
16	6/216	17	104/1296	18	780/7776
17	3/216	18	80/1296	19	735/7776
18	1/216	19	56/1296	20	651/7776
		20	35/1296	21	540/7776
		21	20/1296	22	420/7776
		22	10/1296	23	205/7776
		23	4/1296	24	205/7776
		24	1/1296	25	126/7776
				26	70/7776
				27	35/7776
				28	15/7776
				29	5/7776
				30	1/7776

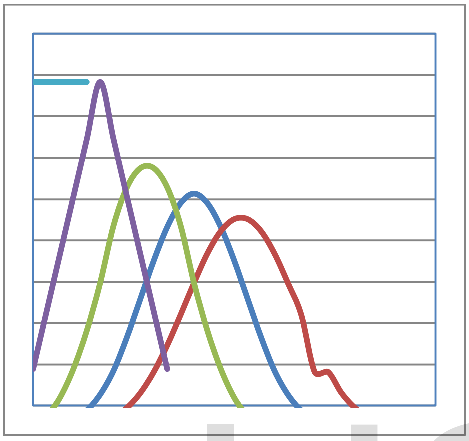












Notice that there are a few important changes to these graphs as the numbers of dice increase. First, the central and most frequent sum of each graph moves to the right (getting larger) as more dice are added. The curve gets closer and closer to the common bell shape of a normal distribution. In fact, the Central Limit Theorem provides some insight into why the sum of a bunch of random dice must always approximate this normal distribution. This is part of what makes that bell curve so common. In probability theory, the **central limit theorem** (**CLT**) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution[2]. That is, suppose that a sample is obtained containing a large number of observations, each observation being randomly generated in a way that does not depend on the values of the other observations, and that the arithmetic average of the observed values is computed. If this procedure is performed many times, the central limit theorem says that the computed values of the average will be distributed according to the normal distribution (commonly known as a "bell curve"). The CLT is responsible for this remarkable result.

References

[1] The Probability Distribution of the Sum of Several Dice: Slot Applications (Ashok K Singh, Rohan J. Dalpatadu and Anthony F. Lucas), *UNLV Gaming Research & Review Journal* (2013)

[2] Advanced Engineering Mathematics by Erwin Kreyszig 9th edition, Advanced Engineering Mathematics by Jain and Iyengar

[3] http://en.wikipedia.org/wiki/Probability_distribution