III-2-2. Nature of binomial distribution

III-2-2-1. Probability distribution of binomial distribution

Statistics is a method for estimation of parent population from data population. Most simple and general method is describing histogram. However, we need summarizing the distribution by several characteristics for statistical analysis. Firstly, we consider description of distribution of population.

Definition of necessary terms

   Expectation value: Sum of products of the obtainable value and possibility of probability event

$$E\big(f(x)\big) = \sum_{i=1}^{n} f(x_i)\, p_i$$

Where

$f(x_i)$: obtainable value by event i

$p(i)$: possibility of event i

$E\big(f(x)\big)$: expectation value

Formula 9

More practically, average value of obtained values by infinitely repeats of probability event

   Average

   Definition 1:   The value obtained by dividing sum of all data by number of data

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

   Definition 2: Expectation value of from obtained data

$$\sum_{i=1}^{n} x_i p_i$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i p_i$$

$\bar{x}$: average of x

Formula 10

   Moment: Value characterizing the distribution of data, calculated as expectation value of $(x - a)^r$

$$\mathrm{E}(x - a)^r = \frac{\sum_{i=1}^{n}(x - a)^r}{n}$$

When r=1, $E(x-a)$ is liner moment

When r=2, $E(x-a)^2$ is quadratic moment

When $a = 0$, $E(x)^r$ is moment around origin

When $a = \mu$, $E(x-\mu)^r$ is moment around mean

μ: mean of parent population

We can say that mean is liner moment around origin and variance is quadratic moment around mean.

SS: Sum of square of the distance of data from average.

$$SS = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$\bar{x}$: average of data

n: number of data

Formula 11

There several possible methods for the estimation of mean of parent population such as simulation by repeated random sampling (MCMC, etc). In those methods, distribution of possible mean value can be directly estimated using the power of computer. We consider expectation value of mean of parent population is average of sample population intuitively and estimate mean of parental population by averaging the sample population. We already accept this as axiomatic truth. However, we have to discuss the adequacy of this expectation method from the purpose of estimation of parental population. The reason for the estimation of mean of parent population is that the value shows the most important character of distribution of parent population, center of distribution.

We learned that binomial distribution express by following equation in former chapter,

$$W(k) = {}_n C_k p^k q^{(n-k)}$$

This means that we can draw the shape of binominal distribution when we know the value of n and p theoretically, because $p + q = 1$. We already know that the shape is unimodal. The position of the peak is most important characteristic of the distribution. The peak is extreme value and the position is obtainable by solution of following derivation equation.

$$\frac{dW(k)}{dk} = 0$$

However, derivation of $C_k p^k q^{(1-k)}$ by k needs several mathematical techniques, and the process of explanation of the solution long and complicated. We will show another method

to obtain the position of the peak without using derivation in following paragraphs at first, then we will show the solution by derivation in later paragraphs.

It may be understandable sensuously.

Figure 6 shows relation of $W(k)$, n and k. when $p = \frac{2}{3}$. This is expressed $B\left(n, \frac{2}{3}\right)$. Taking coin toss game as an analogy, k is times of head and probability of head (A) in a try is $\frac{2}{3}$, and that of tail (B) is $\frac{1}{3}$. W(k) is probability of k times head and $n\text{-}k$ times tail. or probability of k times A and n-k times B. With the increase of n, the plot expands to right side and peak moves to right with the expand of the plot as shown in Fig 6.
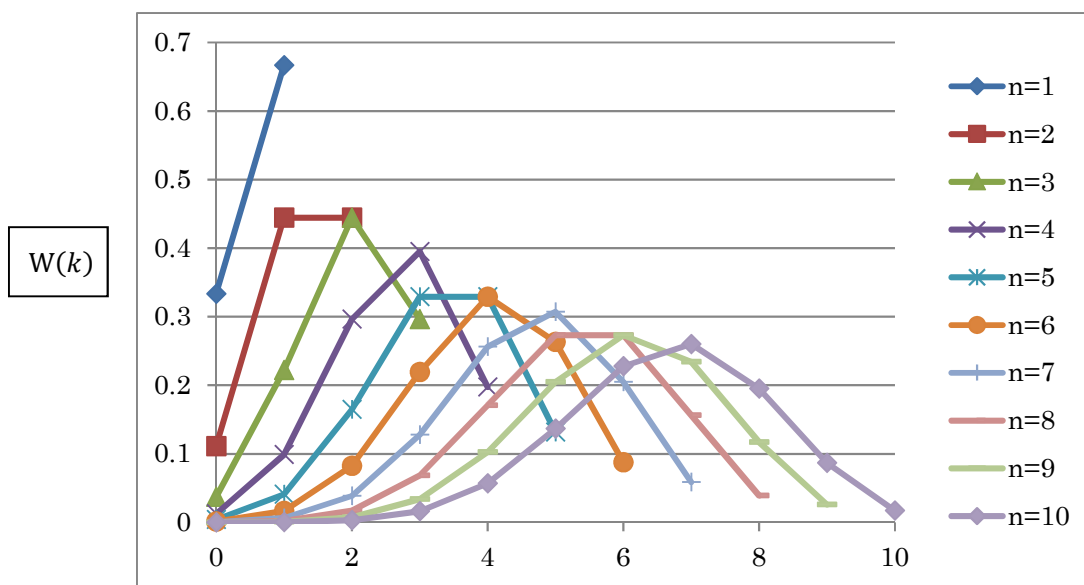

W(k)

Fig.6. $W(k) =_n C_k p^k q^{(1-k)}, B(n, \frac{2}{3})$

In $B(9, \frac{2}{3})$, peak exists at $k = 6$. In $B(6, \frac{2}{3})$ peak exists at $k = 4$, and peak exists at $k = 2$ in $B\left(3, \frac{2}{3}\right)$. In other cases, the positions of peak is not clear, though the position is around k where $\frac{k}{n} = p$. We may consider that the k which gives peak is $k = np$. Simply, probability of A in a try is $\frac{2}{3}$ and we try two times. So maximum probability is obtained at $k = \frac{2}{3} + \frac{2}{3}$, and when we try n times, maximum probability is obtained at np. This is the answer to the question "When you know the probability of A in a try and you try n times, how many times A happen? Answer the most probable times". The answer to this question is expectation value of k.

$$\bar{k} = np. \quad (\bar{k}: \text{expectation value of } k)$$

This is still only our prediction. We need to confirm the prediction.

In above mentioned model, number of repeats in a set is n, and expectation value ($\bar{k}$) can be obtained by summing up products of each value and probability

There are $n+1$ ks; $(0, 1, 2, \cdots n)$.

$$\bar{k} = \sum_{k=0}^{n} W(k) \times k$$

$$= \sum_{k=0}^{n} {}_nC_k p^k q^{(n-k)} \times k$$

$$= \sum_{k=0}^{n} \frac{kn!}{k!\,(n-k)!} p^k q^{(n-k)}$$

$$= \sum_{k=0}^{n} \frac{kn(n-1)!}{k!\,((n-1)-(k-1))!} p \times p^{(k-1)} q^{((n-1)-(k-1))}$$

$$\bar{k} = \sum_{k=0}^{n} \frac{n(n-1)!}{(k-1)!\,((n-1)-(k-1))!} p \times p^{(k-1)} q^{((n-1)-(k-1))}$$

$$\bar{k} = np \sum_{k=1}^{n-1} \frac{(n-1)!}{(k-1)!\,((n-1)-(k-1))!} p^{(k-1)} q^{((n-1)-(k-1))}$$

$\sum_{k=1}^{n-1} \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} p^{(k-1)} q^{((n-1)-(k-1))}$ is a sum of all probability when $n = n-1$.

$$\sum_{k=1}^{n-1} \frac{(n-1)!}{(k-1)!\,((n-1)-(k-1))!} p^{(k-1)} q^{((n-1)-(k-1))} = 1$$

then

$$\bar{k} = np$$

<div align="right">Q.E.D</div>

In a binomial distribution $B(n, p)$, possibility of k times appearance of a phenomenon in n trials ($W(k)$) is as follow

$$W(k) = {}_nC_k p^k q^{(n-k)}$$

$$= \frac{n!}{k!(n-k)!} p^k q^{(n-k)}$$

$$p + q = 1$$

When we take logarithm of both members.

$$\log W(k) = \log(n!) - \log(k!) - \log(n-k)! + k\log(p) + (n-k)\log(q)$$

The equation becomes linear by the transformation and we can discuss each member separately. Here we consider k as continuous valuable (real number, not limited in integer). Generally, k is used expression of integer, so we rewrite k to x. This is a custom and wisdom to prevent misunderstanding.

$$\log W(x) = \log(n!) - \log(x!) - \log(n-x)! + k\log(p) + (n-x)\log(q)$$

Without any prior information, you need to accept following equation.

$$\lim_{x\to\infty} \log x! = \lim_{x\to\infty} \int_1^x \log t \, dt$$

This equation means that $\int_1^x \log t \, dt \fallingdotseq \log x!$, when x is large enough. This equation is simple, and you may accept this equation in an intuitive way.   We need several mathematical techniques for proper proof of this equation and the process of proof is complicated and boring. The readers who do not like such boring story can skip following paragraphs. It makes no problem, though it is useful for readers who want to understand what we are doing to make probability models

Proof of

$$\lim_{x\to\infty} \int_1^x \log t \, dt = \lim_{x\to\infty} \log x!$$

Conclusively, given equation is transformation of following equation. This is our goal.

$$\lim_{x\to\infty} \frac{\int_1^x \log_e t \, dt}{\log_e x!} = 1$$

In the symbol of limitation, there is

$$\int_1^x \log t \, dt$$

The meaning of this equation is area framed by $\log t$, X axis and $X = x$.

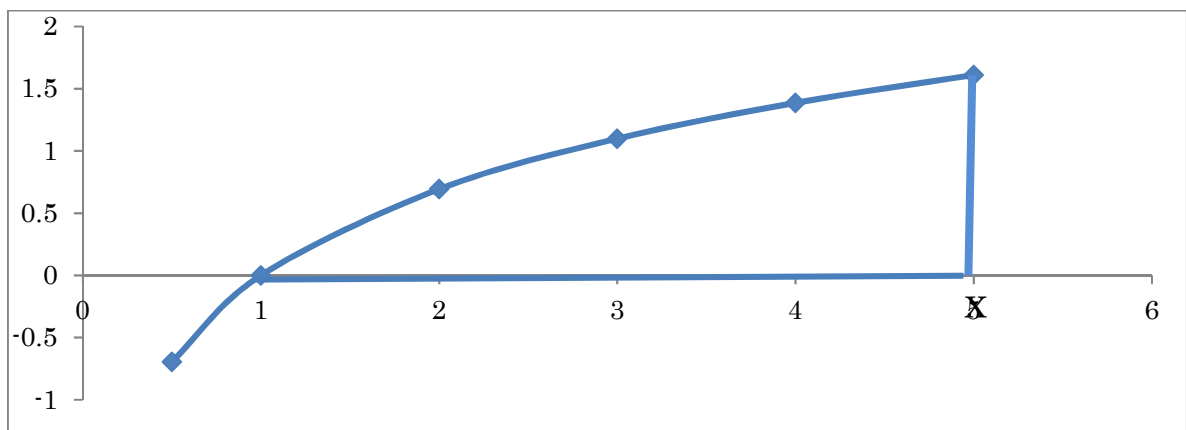$$S_t = \int_1^x \log t \, dt$$

Fig 7.1 Integration range of $\int_1^x \log t \, dt$

We put several rectangles in the figure. Figure 8-1 is overlap of following 2 figures.
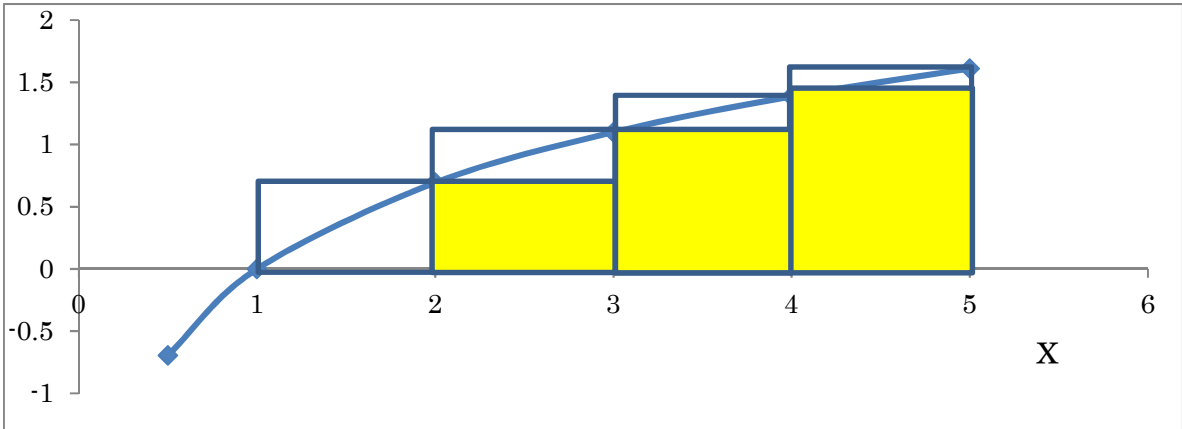
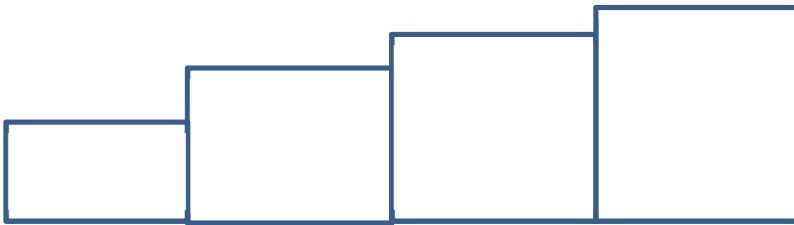Fig.8-1. Illustration diagram of the equation (1)

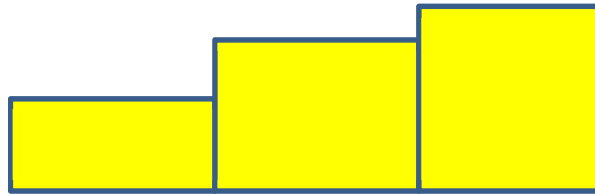Fig. 8-2. Illustration diagram of equation the equation (2)

Fig. 8-3. Illustration diagram of the equation (3)

Width of rectangles are 1. Length of the left rectangle in Fig. 8-2 is $\log 2$. Similarly, the length of other rectangles are $\log 3$, $\log 4$ and $\log 5$. Sum of 4 rectangles is follow.

$$S_l = \log 2 + \log 3 + \log 4 + \log 5 = \log 5!$$

When x increases,

$$S_l = \log x!$$

Similarly. Sum of the area of yellow rectangles is as follow.

$$S_s = \log(x - 1)!$$

When we focus on the size of the areas. Figure 18-1 shows,

$$S_l > S_t > S_s$$

$$\log(x - 1)! < \int_1^x \log t\, dt < \log x!$$

Here,

$$x > 1, \log x! > 0$$

We can divide all members without change of direction of inequality sign

.

$$\frac{\log(x-1)!}{\log\ !} < \frac{\int_1^x \log t\, dt}{\log\ !} < \frac{\log x!}{\log x!}$$

Obviously,

$$\frac{\log x!}{\log x!} = 1$$

Then we consider limit of left member.

$$\frac{\log(x-1)!}{\log x!} = \frac{\log(x-1) + \log(x-2) + \cdots + \log 1}{\log x!}$$

$$= \frac{\log x + \log(x-1) + \log(x-2) + \cdots + \log 1 - \log x}{\log x!}$$

$$= \frac{\log x! - \log x}{\log x!}$$

$$= 1 - \frac{\log x}{\log x!}$$

For the author, $\lim\limits_{x\infty} \frac{\log x}{\log x!} = 0$ is trivial, though sense of triviality is different among

people. It is better to proof

$$\lim_{x\to\infty} \frac{\log x}{\log x!} = 0$$

Proof

$$\log k! > \log k + \log(k-1) + \cdots + \log\left[\frac{k}{2}\right] > \left(\frac{k}{2} - 1\right)\log\left[\frac{k}{2}\right]$$

Meaning of above inequality is as follow

$\left[\frac{k}{2}\right]$: maximum integer number which does not exceed k. When $k = 5$, $\left[\frac{k}{2}\right] = 2$.

When $k = 4$, $\left[\frac{k}{2}\right] = 2$,

$\log k! > \log k + \log(k-1) + \cdots + \log\left[\frac{k}{2}\right]$ means that cantle does not exceed entirety.

$$\text{Log } k! = \log k + \log(k-1)! + \cdots + \log 1$$

$$\log k + \log(k-1) + \cdots + \log\left[\frac{k}{2}\right]$$

All members of right side of inequalities are positive.

So

$$\log k! > \log k + \log(k-1) + \cdots + \log\left[\frac{k}{2}\right]$$

$\log k + \log(k-1) + \cdots + \log\left[\frac{k}{2}\right] > \left(\frac{k}{2}-1\right)\log\left[\frac{k}{2}\right]$ means comparison between sum of the

blue rectangles and the sum of the area of yellow rectangles in figure 9.

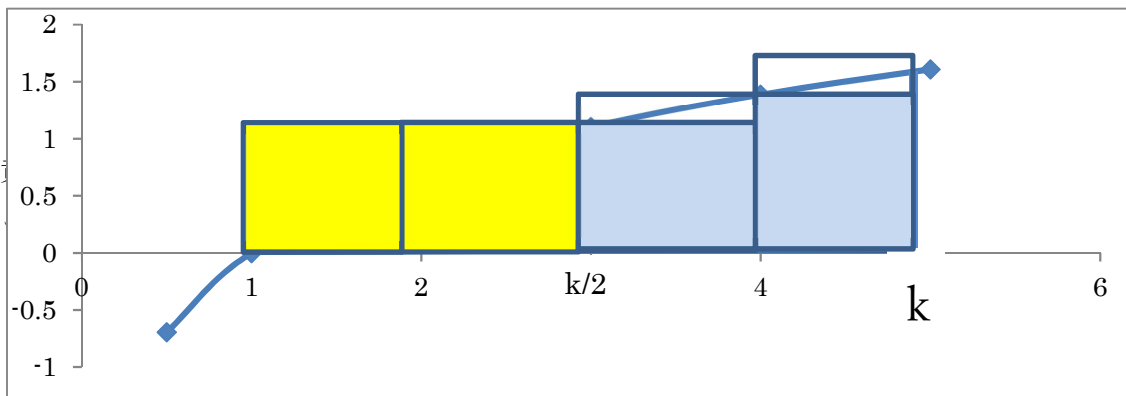Fig. 9. Comparison of magnitude relationship.

$$\log k + \log(k-1) + \cdots + \log\left[\frac{k}{2}\right] > \left(\frac{k}{2}-1\right)\log\left[\frac{k}{2}\right]$$

is proven.

And   also

$$\log k! > \log k + \log(k-1) + \cdots + \log\left[\frac{k}{2}\right] > \left(\frac{k}{2}-1\right)\log\left[\frac{k}{2}\right]$$

is proven.

Conclusively

$$\log k! > \left(\frac{k}{2}-1\right)\log\left[\frac{k}{2}\right]$$

$$\log k! > \left(\frac{k}{2}-1\right)\log\left[\frac{k}{2}\right]$$

we can transform k as x,

$$\log x! > \left(\frac{x}{2}-1\right)\log\left[\frac{x}{2}\right]$$

The relation is the same when $x \to \infty$.

$$\lim_{x\to\infty} \log x! > \lim_{x\to\infty} \left(\frac{x}{2} - 1\right)\log\left[\frac{x}{2}\right]$$

When $x \to \infty$, $\lim_{x\to\infty}\left(\frac{x}{2} - 1\right)\log\left[\frac{x}{2}\right] > 1$

$$\lim_{x\to\infty}\frac{1}{\log x!} < \lim_{x\to\infty}\frac{1}{\left(\frac{x}{2} - 1\right)\log\left[\frac{x}{2}\right]} = 0$$

Here,

$$\log x > 0 \ \ and \ \ \frac{1}{\log x!} > 0$$

$$\frac{\log x}{\log x!} > 0$$

$$0 \le \lim_{x\to\infty}\frac{\log x}{\log x!} \le \lim_{x\to\infty}\frac{\log x}{\left(\frac{x}{2}-1\right)\log\frac{x}{2}} = \lim_{x\to\infty}\frac{\log x}{\left(\frac{x}{2}-1\right)(\log x - \log 2)} = \lim_{x\to\infty}\frac{1}{\left(\frac{x}{2}-1\right)\left(1 - \frac{2}{\log x}\right)} = 0$$

$$0 \le \lim_{x\to\infty}\frac{\log x}{\log x!} \le 0$$

$$\lim_{x\to\infty}\frac{\log x}{\log x!} = 0$$

is proven. The method used in above proof is called false position method.

$$\lim_{x\to\infty}\left(1 - \frac{\log x}{\log x!}\right) = 1$$

$$1 - \frac{\log x}{\log x!} = \frac{\log x! - \log x}{\log x!} = \frac{\log\frac{x!}{x}}{\log x!} = \frac{\log(x-1)!}{\log x!}$$

So

$$\lim_{x\to\infty}\frac{\log(x-1)!}{\log x!} = 1$$

$$1 = \lim_{x\to\infty}\frac{\log(x-1)!}{\log x!} \le \lim_{x\to\infty}\frac{\int_1^x \log t\, dt}{\log x!} \le \lim_{x\to\infty}\frac{\log x!}{\log x!} = 1$$

Using false position method,

$$\lim_{x\to\infty}\frac{\int_1^x \log t\, dt}{\log x!} = 1$$

$$\lim_{x\to\infty}\int_1^x \log t\, dt = \lim_{x\to\infty}\log x!$$

When x is large enough,

$$\log x! \fallingdotseq \int_1^x \log t\, dt$$

We go back to equation 12

$$\log W(x) = \log(n!) - \log(x!) - \log(n-x)! + x\log(p) + (n-x)\log(q)$$

$$\fallingdotseq \log(n!) - \int_1^x \log t\, dt - \int_1^{n-x} \log t\, dt + x\log p + (n-x)\log q$$

In case of $x \to \infty$, we can write

$$\log W(x) = \log(n!) - \int_1^x \log t\, dt - \int_1^{n-x} \log t\, dt + x\log p + (n-x)\log q$$

$$\{\log W(x)\}' = -\log x + \log(n-x) + \log p - \log(1-p)$$

$$= \log \frac{(n-x)p}{x(1-p)}$$

because

$$\frac{dt(t\log t - t)}{dt} = 1 \cdot \log t + t \cdot \frac{1}{t} - 1 = \log t$$

$$\therefore \int \log t\, dt = t\log t - t + C$$

$$\therefore \int_1^x \log t\, dt = x\log x - x - 1\log 1 + 1 = x\log x - x + 1$$

$$\therefore \frac{d\left(\int_1^x \log t\, dt\right)}{dx} = 1 \times \log x + x\frac{1}{x} - 1 = \log x$$

$$\frac{d(\log n!)}{dx} = 0, \frac{d(x\log p)}{dx} = \log p, \frac{d\big((n-x)\log q\big)}{dx} = \log q = \log(1-p)$$

When

$$\frac{(n-x)p}{x(1-p)} = 1,$$

$$\log \frac{(n-x)p}{x(1-p)} = 0$$

$$\frac{(n-x)p}{x(1-p)} = 1$$

$$\textcolor{red}{\because \log 1 = 0}$$

$$(n-x)p = x(1-p)$$

$$np - xp = x - xp$$

$$x = np$$

The $(\log W(x))'$ is monotonically decreasing function and $(\log W(x))'$ is 0 at

$$x = np$$

This result means $\log W(x)$ is local maximum value at $x = np$, and the plot of the

probability $W(x)$ peaks at $x = np$

This conclusion agrees with our empirical prediction

$$\mu = np$$

For readers who know derivation of composite function, the conclusion is obtainable simply.

Rule of derivation of composite function is as follow

$$\frac{df(g(x))}{dx} = \frac{df(g(x))}{dg(x)} \frac{dg(x)}{dx}$$

Here,

$$\log W(x) = \log(n!) - \log(x!) - \log(n-x)! + x\log(p) + (n-x)\log(q)$$

$$W(x) = e^{(\log(n!) - \log(x!) - 1 \quad (n-x)! + x\log(p) + (n-x)\log(q))}$$

$$g(x) = \log(n!) - \log(x!) - \log(n-x)! + x\log(p) + (n-x)\log(q)$$

$$W(g(x)) = e^{g(x)}$$

$$\frac{dW(g(x))}{dg(x)} = e^{g(x)}$$

$$\frac{dg(x)}{dx} = -\log x + \log(n-x) + \log(p) - \log(q)$$

$$\frac{dW(g(x))}{dx} = \frac{dW(g(x))}{dg(x)} \frac{dg(x)}{dx} = e^{g(x)}(-\log x + \log(n-x) + \log(p) - \log(q))$$

$$= e^{g(x)}\left(\log \frac{(n-x)p}{x(1-p)}\right)$$

Calculation of x which gives extreme value

$$\frac{dW(g(x))}{dx} = e^{g(x)}\left(\log \frac{(n-x)p}{x(1-p)}\right) = 0$$

$$e^{g(x)} > 0 \quad (not\ 0)$$

$$\frac{(n-x)p}{x(1-p)} = 1$$

$$(n-x)p = x(1-p)$$

$$np - xp = x - xp$$

$$x = np$$

Next important character is breadth and sharpness of the peak of distribution of parent population. This is the characteristic expressing the average distance of data from average value. Mean of parent population is estimated as a point directly using average of sampled data at first, though we still do not discuss the adequacy of the method. From

this we may consider intuitionally that liner moment around average, $M(x - \bar{x})$, can be used as the characteristics of distribution. The calculation of liner moment is as follow.

$$M(x - \bar{x}) = E(x - \bar{x}) = \frac{\sum_{i=1}^{n}(x - \bar{x})}{n} = 0$$

$$\because \quad \sum_{i=1}^{n}(x - \bar{x}) = 0$$

The result is showing that the character should be expressed as quadratic moment of parent population. Quadratic moment is called variance in classic statistics, which has been developed by Fisher and other pioneers.

Quadratic moment of sample population is as follow.

$$M((x - \bar{x})^2) = \frac{\sum_{i=1}^{n}(x - \bar{x})^2}{n}$$

$$x : \text{data}$$
$$\bar{x} : \text{average of samppled data}$$
$$n : \text{number of sammpled data}$$

We discuss whether we can us quadratic variance sample population as variance of parent population. Not only for theoretical understanding but also for sensory agreement, we estimate mean and quadric moment around mean of parent population from the data randomly obtained from the parent population assuming that parent population completely follows binomial distribution model.

Assuming that there is a probability event in a trial and the probability is p. When the trials are repeated n times in a set, the probability of times of the happening k can be expressed as $p(k)$ and $p(k)$ follows binomial distribution model $B(n, p)$. We can calculate SS in each set. When we sum up infinitely the set, we may expect that we can obtain the characteristics of distribution of parental population. We can get conclusion by actual implementation of this procedure.

One of the simple examples is toss of the coin and we can get 1 dollar when head and we lose 1 dollar when tail.

This example obviously follows binomial distribution model, $B\left(n, \frac{1}{2}\right)$. The mean is 0 and the quadric moment around origin is 1 regardless of n.

We calculate $\frac{SS}{n}$ and $\frac{SS}{n-1}$

In a case of $B\left(1, \frac{1}{2}\right)$

n=1

obtainable data	probability	average	SS	SS/n	SS/(n-1)
-1	1/2	-1	0	0	-
1	1/2	1	0	0	-
Expectation value			0	0	

In a case of $B\left(2,\frac{1}{2}\right)$

Obtainable data	Probability	average	SS	SS/n	SS/(n-1)
$-1 \ -1$	1/4	-1	0	0	0
$-1 \ \ 1$	1/2*	0	2**	1	2
1 \ \ 1	1/4	1	0	0	0
Expectation value			1	1/2***	1 ****

      * : $_2C_1(1/2)*(1/2)$

     ** : $(-1-0)^2+(1-0)^2=2$

   *** : $0\times1/4+2\times1/2+0\times1/4=1$

 **** : $0\times1/4+2\times1/2+0\times1/4$

The author recommends the readers to calculate by themselves for strong understanding.

In a case of $B\left(3,\frac{1}{2}\right)$

Obtainable data	possibility	average	SS	SS/n	SS/(n-1)
$-1 \ -1 \ -1$	1/8	-1	0	0	0
$-1 \ -1 \ \ 1$	3/8	$-1/3$	24/9*	8/9**	12/9***
$-1 \ \ 1 \ \ 1$	3/8	1/3	24/9	8/9	12/9
1 \ \ 1 \ \ 1	1/8	1	0	0	0
Expectation value			2	2/3****	1*****

     * : $(-1-(-1/3))^2+(-1-(-1/3))^2+(1-(-1/3))^2=24/9$

   ** : $(24/9)/3$

  *** : $(24/9)/2$

 **** : $0\times1/8+(8/9)\times(3/8)+(8/9)\times(3/8)+0\times(1/8)$

***** : $0\times1/8+(12/9)\times(3/8)+(12/9)\times(3/8)+0\times(1/8)$

In a case of $B\left(4,\frac{1}{2}\right)$

Obtainable data	probability	average	SS	SS/n	SS/(n-1)

− 1 − 1 − 1 − 1	1/16	− 1	0	0	0
− 1 − 1 − 1 1	4/16	− 1/2	48/16	12/16	16/16
− 1 − 1 1 1	6/16	0	4	1	4/3
− 1 1 1 1	4/16	1/2	48/16	12/16	16/16
1 1 1 1	1/16	1	0	0	0
Expectation value			3	3/4	1

The readers can understand that quadric moment of parent population is calculated as $\frac{SS}{n-1}$.   Rigorous algebraic proof will be given in later part of this chapter. This is explanation for sensory understanding that quadratic moment of sample population can be calculated as $\frac{SS}{n}$ but this is not expectation value of quadratic moment of parent population and quadratic moment of parent population is $\frac{SS}{n-1}$.   Quadratic moment of population is called as variance in statistics and quadratic moment of sample is sample variance and quadratic variance of parent population estimated from sample papulation is called unbiased variance.   When we express unbiased variance as σ^2. The ratio of SS to σ^2, $\left(\frac{SS}{\sigma^2}\right)$, is degree of freedom. In this case degree of freedom is n-1, we have to estimate various variances in statistic and degree of freedom is depending on the required variance and n-1 is not always degree of freedom.

Figures are plot of the $p(k)$ against obtainable averages. Data used is same as the table used for the explanation of unbiased variance.
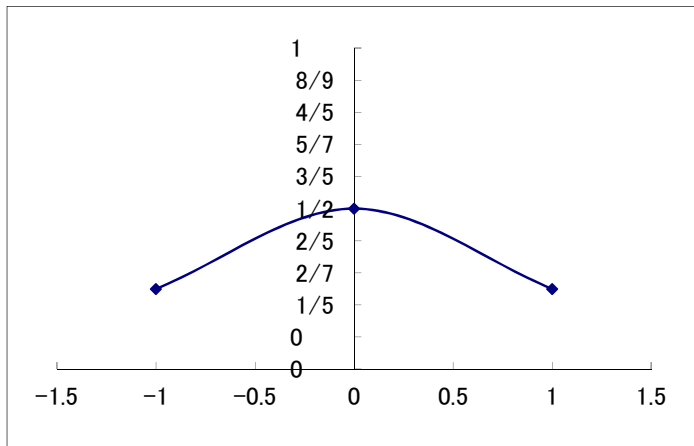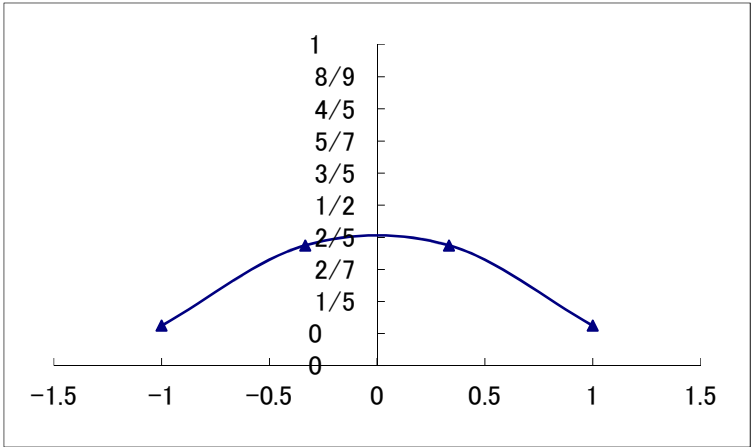

Fig 10. $B\left(2,\frac{1}{2}\right)$

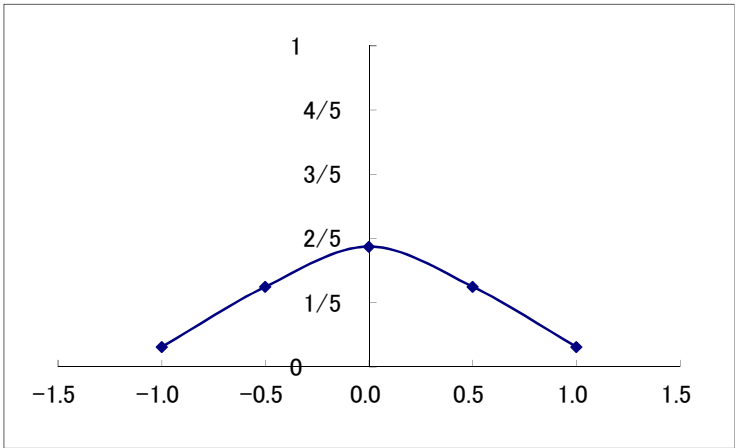Fig.11. $\mathrm{B}\left(3,\frac{1}{2}\right)$

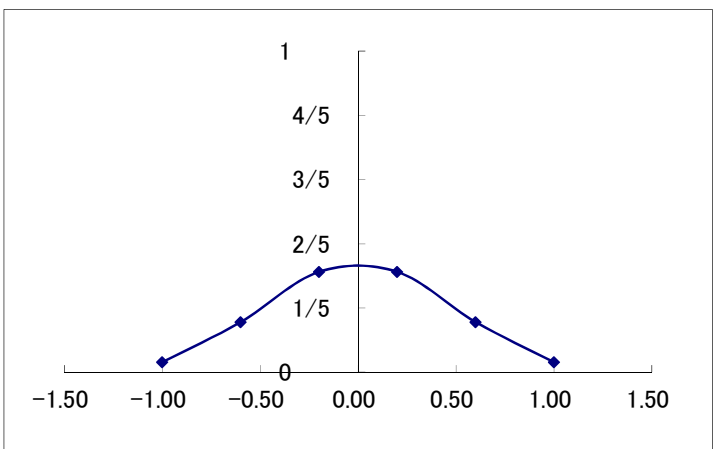Fig.12. $\mathrm{B}\left(4,\frac{1}{2}\right)$

Fig.13. $\mathrm{B}\left(5,\frac{1}{2}\right)$
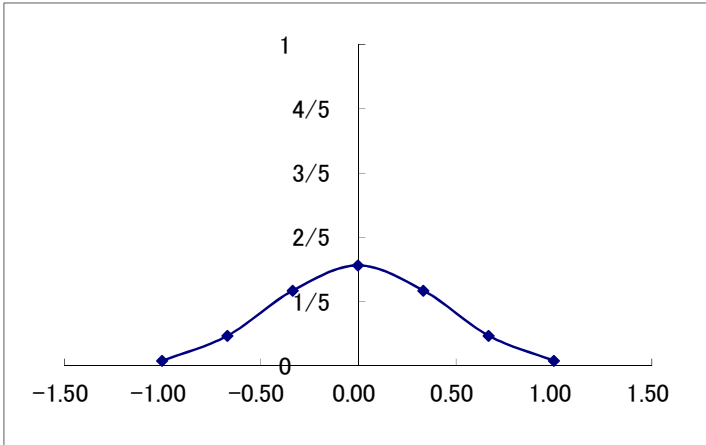
Fig. 14. $B\left(6,\frac{1}{2}\right)$

Values of both side are decreased and peak is sharpened with increase of n.

In case of $B\left(n,\frac{1}{2}\right)$, parent population is originally symmetric. Following is the cases of asymmetric parent population.

Fig.15. Parent population p=1/3

Mean of parental population is $-\frac{1}{3}$, and quadric moment is $\frac{8}{9}$

$B\left(1,\frac{1}{3}\right)$

Obtainable data	probability	average	SS	SS/n	SS/(n-1)
-1	2/3	-1	0	0	-
1	1/3	1	0	0	-
Expectation value			0	0	

$B\left(2, \dfrac{1}{3}\right)$

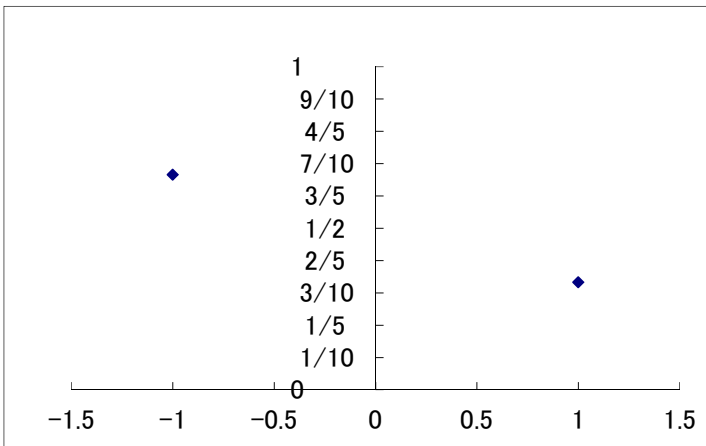| Obtainable data	probability	average	SS	SS/n	SS/(n-1)
$-1-1$	4/9	-1	0	0	0
$-1\quad 1$	4/9*	0	2**	1	2
$1\quad 1$	1/9	1	0	0	0
Expectation value			8/9	4/9***	8/9****

   * : $_2C_1(2/3)*(1/3)$

   ** : $(-1-0)^2 + (1-0)^2 = 2$

   *** : $0\times4/9+1\times4/9+0\times1/9 = 4/9$

  **** : $0\times4/9+2\times4/9+0\times1/9$

$B\left(3, \dfrac{1}{3}\right)$

Obtainable data	probability	average	SS	SS/n	SS/(n-1)
$-1-1-1$	8/27	-1	0	0	0
$-1-1\quad 1$	12/27	$-1/3$	24/9*	8/9**	12/9***
$-1\quad 1\quad 1$	6/27	1/3	24/9	8/9	12/9
$1\quad 1\quad 1$	1/27	1	0	0	0
Expectation value			2	48/81****	8/9*****

   * : $(-1-(-1/3))^2 + (-1-(-1/3))^2 + (1-(-1/3))^2 = 24/9$

   ** : $(24/9)/3$

   *** : $(24/9)/2$

  **** : $0\times8/27 +(8/9)\times(12/27) +(8/9)\times(6/27)+0\times(1/27)$

 ***** : $0\times8/27 +(12/9)\times(12/27) +(12/9)\times(6/27)+0\times(1/27)$

$B\left(4, \dfrac{1}{3}\right)$

Obtainable data	probability	average	SS	SS/n	SS/(n-1)
$-1-1-1-1$	16/81	-1	0	0	0
$-1-1-1\quad 1$	32/81	$-1/2$	48/16	12/16	16/16
$-1-1\quad 1\quad 1$	24/81	0	4	1	4/3
$-1\quad 1\quad 1\quad 1$	8/81	1/2	48/16	12/16	16/16
$1\quad 1\quad 1\quad 1$	1/81	1	0	0	0
Expectation value			3	2/3	8/9

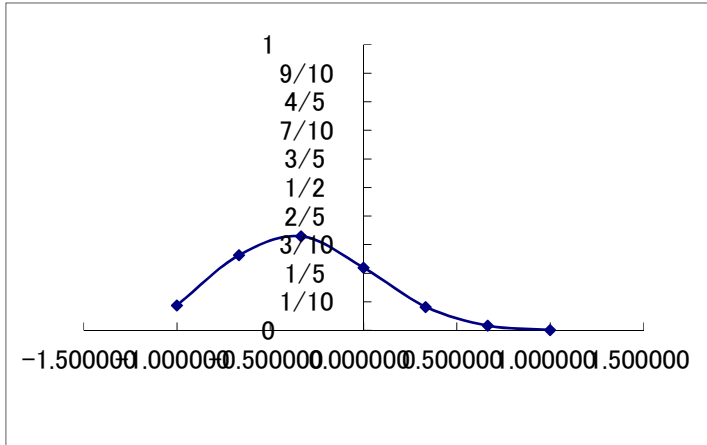Fig. 16. $\mathrm{B}\left(6, \dfrac{1}{3}\right)$

We can understand from the tables and figures that $\dfrac{SS}{n-1}$ is unbiased variance of parent population and the shape of probability of average value became symmetric and sharp with increase of n also in the case of asymmetric parental population. This means that the expectation value of mean concentrates to narrow range around actual mean with increase of n. This square our empirical rule that more sample make the estimation more reliable and named central limit theorem. This means average of sample population close to mean of parent population with increase of sample size. Depending on this theory we can accept use of average of sample population as expectation value of mean of parent population.

In case of binomial distribution, quadratic moment of parent population around mean (μ). This is expectation value the moment can be calculated as follow,

$$\mathrm{E}((k-\mu)^2) = \sum_{k=0}^{n}(k-np)^2\, W_{(k)}$$

$$= \sum_{k=0}^{n}(k^2 - 2knp + n^2p^2)\, W_{(k)}$$

$$= \sum_{k=0}^{n} k^2 W_{(k)} - \sum_{k=0}^{n} 2knp W_{(k)} + \sum_{k=0}^{n} n^2p^2 W_{(k)}$$

$$= \sum_{k=0}^{n} k^2 W_{(k)} - 2np \sum_{k=0}^{n} k W_{(k)} + n^2p^2 \sum_{k=0}^{n} W_{(k)}$$

First member $\sum_{k=0}^{n} k^2 W_{(k)}$ is expectation value of k^2

$$\sum_{k=0}^{n} k^2 W_{(k)} = \mathrm{E}(k^2)$$

$\sum_{k=0}^{n} k W_{(k)}$ in second term in right side is estimation value of mean of parental

population (μ). $\mathrm{in}B(n,p)$

$$\mu = np$$

$$\sum_{k=0}^{n} W_{(k)} = 1$$

$\because$ Sum of all probability is 1

$$\mathrm{E}((k-\mu)^2) = \sum_{k=0}^{n} k^2 W_{(k)} - 2n^2 p^2 + n^2 p^2$$

$$= \mathrm{E}(k^2) - n^2 p^2$$

$$\mathrm{E}\{k(k-1)\} = E(k^2 - k) = E(k^2) - E(k)$$

$$\mathrm{E}\{k(k-1)\} = \sum_{k=0}^{n} k(k-1) \cdot_n C_k p^k q^{n-k}$$

$$= \sum_{k=0}^{n} \frac{k(k-1)n!}{(n-k)!\,k!} \cdot_n C_k p^k q^{n-k}$$

$$\sum_{k=0}^{n} \frac{n(n-1)k(k-1)(n-2)!}{(n-2-(k-2))!\,k!} \cdot_n C_k p^k q^{n-k}$$

$$n(n-1) \sum_{k=0}^{n} \frac{k(k-1)(n-2)!}{(n-2-(k-2))!\,k(k-1)(k-2)!} \cdot_n C_k p^k q^{n-k}$$

$$n(n-1) \sum_{k=0}^{n} \frac{(n-2)!}{((n-2)-(k-2))!\,(k-2)!} \cdot_n C_k p^2 p^{k-2} q^{(n-2)-(k-2)}$$

$$n(n-1)p^2 \sum_{k=0}^{n} \frac{(n-2)!}{((n-2)-(k-2))!\,(k-2)!} \cdot_n C_k p^{k-2} q^{(n-2)-(k-2)}$$

So

$$\mathrm{E}\{k(k-1)\} = n(n-1)p^2$$

Because

$$\sum_{k=0}^{n} \frac{(n-2)!}{((n-2)-(k-2))!\,(k-2)!} \cdot_n C_k p^{k-2} q^{(n-2)-(k-2)} \quad \text{is sum of all probability } B(n-2.p)$$

$$\sum_{k=0}^{n} \frac{(n-2)!}{((n-2)-(k-2))!\,(k-2)!} \cdot_n C_k p^{k-2} q^{(n-2)-(k-2)} = 1$$

$$= n(n-1)p^2$$

Put these transform to $\mathrm{E}\{k(k-1)\} = E(k^2) - E(k)$

$$\mathrm{E}\{k(k-1)\} = E(k^2) - E(k)$$

$$\mathrm{E}(k^2) = n(n-1)p^2 + np = n^2 p^2 - np^2 + np$$

Conclusively,

$$\mathrm{E}((k-\mu)^2) =$$

$$= \mathrm{E}(k^2) - n^2 p^2$$

$$= n^2p^2 - np^2 + np - n^2p^2$$
$$= np(1 - p)$$
$$= npq$$
$$\because p + q = 1$$

As a side note,

$$E((k - \mu)^2) =$$
$$= E(k^2) - n^2p^2$$
$$= E(k^2) - \mu^2$$
$$= E(k^2) - E(k)^2$$
$$E((k - \mu)^2) = V_k$$
$$V_k = E(k^2) - E(k)^2$$
$$V_k : \text{variance of } k$$

Formula 13

This is simplified calculation of variance.

Then we discuss the change of the rage of the estimated mean around proper mean with increase of n using binomial distribution. Concretely, quadratic moment of estimated mean around proper mean.

Mathematically, the relation between $E((M - \mu)^2)$ and n will be discussed , where
   M: calculated average from sample population, μ: proper mean of parent population

Following is an example of $B\left(3.\frac{1}{2}\right)$

Obtainable data	possibility	average(M)	mean(μ)
− 1 − 1 − 1	1/8	− 1	0
− 1 − 1 1	3/8	− 1/3	0
− 1 1 1	3/8	1/3	0
1 1 1	1/8	1	0

Expectation value of $E((M - \mu)^2)$

$$(-1 - 0)^2 \times \frac{1}{8} + \left(-\frac{1}{3} - 0\right)^2 \times \frac{3}{8} + \left(\frac{1}{3} - 0\right)^2 \times \frac{3}{8} + (-1 - 0)^2 \times \frac{1}{8} = \frac{1}{8} + \frac{1}{3\times8} + \frac{1}{3\times8} + \frac{1}{8} = \frac{1}{3}$$

$$B\left(4,\frac{1}{2}\right)$$

Obtainable data	probability	average(M)	mean(μ)
− 1 − 1 − 1 − 1	1/16	− 1	0
− 1 − 1 − 1 1	4/16	− 1/2	0

− 1	− 1	1	1	6/16	0	0		
− 1	1	1	1	4/16	1/2	0		
1	1	1	1	1/16	1	0	0	0

Expectation value of $E((M-\mu)^2)$

$$(-1-0)^2 \times \frac{1}{16} + \left(-\frac{1}{2}-0\right)^2 \times \frac{4}{16} + (0-0)^2 \times \frac{6}{16} + \left(\frac{1}{2}-0\right)^2 \times \frac{4}{16} + (1-0)^2 \times \frac{1}{16}$$

$$= \frac{1}{16} + \frac{1}{16} + 0 + \frac{1}{16} + \frac{1}{16} = \frac{1}{4}$$

In the case of $B\left(n,\frac{1}{2}\right)$, $\sigma^2 = 1$

n=1 $1 = \frac{\sigma^2}{1}$

n=2 $\frac{1}{2} = \frac{\sigma^2}{2}$

n=3 $\frac{1}{3} = \frac{\sigma^2}{3}$

n=4 $\frac{1}{4} = \frac{\sigma^2}{4}$

In the case of $B\left(n,\frac{1}{3}\right)$, $\sigma^2 = \frac{8}{9}$

$B\left(3,\frac{1}{3}\right)$

Obtainable data	probability	average(M)	mean(μ)
− 1 − 1 − 1	8/27	− 1	− 1/3
− 1 − 1 1	12/27	− 1/3	− 1/3
− 1 1 1	6/27	1/3	− 1/3
1 1 1	1 /27	1	− 1/3

Expectation value of $E((M-\mu)^2)$

$$\left(-1-\left(-\frac{1}{3}\right)\right)^2 \times \frac{8}{27} + \left(-\frac{1}{3}-\left(-\frac{1}{3}\right)\right)^2 \times \frac{12}{27} + \left(\frac{1}{3}-\left(-\frac{1}{3}\right)\right)^2 \times \frac{6}{27} + \left(1-\left(-\frac{1}{3}\right)\right)^2 \times \frac{1}{27}$$

$$= \frac{2^2 \times 8}{3^2 \times 27} + \frac{0 \times 12}{27} + \frac{2^2 \times 6}{3^2 \times 27} + \frac{4^2 \times 1}{3^2 \times 27} = \frac{32 + 24 + 16}{3^2 \times 27} = \frac{72}{3^2 \times 27} = \frac{8}{27}$$

$B\left(4,\frac{1}{3}\right)$

Obtainable data   probability   average(M)   mean(μ)

−1	−1	−1	−1	16/81	−1	−1/3
−1	−1	−1	1	32/81	−1/2	−1/3
−1	−1	1	1	24/81	0	−1/3
−1	1	1	1	8/81	1/2	−1/3
1	1	1	1	1/81	1	−1/3

Expectation value of $\mathrm{E}((M - \mu)^2)$

$$\left(-1-\left(-\tfrac{1}{3}\right)\right)^2 \times \tfrac{16}{81} + \left(-\tfrac{1}{2}-\left(-\tfrac{1}{3}\right)\right)^2 \times \tfrac{32}{81} + \left(0-\left(-\tfrac{1}{3}\right)\right)^2 \times \tfrac{24}{81} + \left(\tfrac{1}{2}-\left(-\tfrac{1}{3}\right)\right)^2 \times \tfrac{8}{81} + \left(1-\left(-\tfrac{1}{3}\right)\right)^2 \times \tfrac{1}{81}$$

$$= \frac{2^2 \times 16}{3^2 \times 81} + \frac{32}{2^2 \times 3^2 \times 81} + \frac{24}{3^2 \times 81} + \frac{5^2 \times 8}{2^2 \times 3^2 \times 81} + \frac{4^2 \times 1}{3^2 \times 81}$$

$$= \frac{2^4 \times 16 + 32 + 2^2 \times 24 + 25 \times 8 + 2^2 \times 16}{2^2 \times 3^2 \times 81}$$

$$= \frac{8 \times 81}{2^2 \times 3^2 \times 81} = \frac{8}{4 \times 9}$$

Summary of result of calculation of $\mathrm{B}\left(n, \tfrac{1}{3}\right)$, $\sigma^2 = \tfrac{8}{9}$

n=1 $\dfrac{8}{9} = \dfrac{\sigma^2}{1}$

n=2 $\dfrac{4}{9} = \dfrac{\sigma^2}{2}$

n=3 $\dfrac{8}{27} = \dfrac{\sigma^2}{3}$

n=4 $\dfrac{8}{36} = \dfrac{\sigma^2}{4}$

From these results, we can make prediction that expectation value around mean of parent population, $\mathrm{E}((M - \mu)^2)$, can be calculated as follow

$$\mathrm{E}((M - \mu)^2) = \frac{\sigma^2}{n}$$

Formula 14

III-2-2-2. Algebraical proof.

Now we consider the calculation for estimation value of quadratic moment of average of sample population around proper mean, $\mathrm{E}((M - \mu)^2)$. Here, M is average of sample population, and μ is proper mean of parent population.

We assume the mean of parental population $\mu = 0$, for prevention of unnecessary complicated expression of equation.

We can express the quadratic moment of parent population, σ^2, as follow by

$$\sigma^2 = \lim_{m \to \infty} \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n} \sum_{i=1}^{n} (M_j + e_{ij})^2$$

$$= \lim_{m \to \infty} \frac{1}{m} \sum_{j}^{m} M_j{}^2 + 2 \lim_{m \to \infty} \frac{1}{m} \frac{1}{n} \sum_{j}^{m} \left(M_j \sum_{i=1}^{n} e_{ij} \right) + \lim_{m \to \infty} \frac{1}{m} \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{n} e_{ij}{}^2$$

n: number of trials in a set

m: number of sets

M: average in a set

e_{ij}: deviation or each data from average

First member of the right members, $\lim_{m \to \infty} \frac{1}{m} \sum_{j}^{m} M_j{}^2$, is $E((M - \mu)^2)$, because $\mu = 0$,

Second member $\lim_{m \to \infty} \frac{1}{m} \frac{1}{n} \sum_{j}^{m} (M_j \sum_{i=1}^{n} e_{ij}) = 0$, because e_{ij} is difference from average and

sum of the differences from average is 0

Third member, $\lim_{m \to \infty} \frac{1}{m} \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{n} e_{ij}{}^2$, is expectation value of $\frac{SS}{n}$, because $\sum_{i=1}^{n} e_{ij}{}^2 = SS$

$$\frac{SS}{n - 1} = \sigma^2$$

$$\therefore \lim_{m \to \infty} \frac{1}{m} \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{n} e_{ij}{}^2 = \sigma^2 - \frac{1}{n} \sigma^2$$

$$\sigma^2 = \lim_{m \to \infty} \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n} \sum_{i=1}^{n} (M_j + e_{ij})^2 = E(M^2) + \sigma^2 - \frac{1}{n} \sigma^2$$

$$E(M^2) = \frac{1}{n} \sigma^2$$

Q.E.D

$$\sqrt{E(M^2)} = \frac{\sigma}{\sqrt{n}}$$

Formula 15

$\frac{\sigma}{\sqrt{n}}$ is standard error, the character representing confidential range of estimated mean of parent population. When we need to express the confidential range of estimated mean, we show standard error.

Mathematical summary and conclusion

1. Center and range of distribution characterize the distribution of population.
   Shape of the distribution depends on models

2. Shape of binomial distribution, $B(n, p)$, is determined by n and p.
   When parent distribution follows $B(n, p)$,

$$\mu = np$$
$$\sigma^2 = np(1 - p)$$

3. Center of parental population is mean and we use average of sample population for estimation of mean of parent population from sample population

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\mu = \bar{x}$$

$\bar{x}$: average of sample population

n: number of sampled data (sample size)

μ: estimated mean of parent population

4. Range of the distribution of is expressed by quadratic moment around mean.

$$M(x - \mu)^2: \text{quadratic moment around mean}$$

5. $M(x - \mu)^2$ is estimated as expectation value.

$$M(x - \mu)^2 = E(x - \mu)^2$$

and it can be calculated from sample population as follows

$$M(x - \mu)^2 = E(x - \mu)^2$$

$$= \frac{SS}{n - 1}$$

$$\sigma^2 = \frac{SS}{n - 1}$$

SS: Sum of square

$$SS = \sum_{i=1}^{n} (x - \bar{x})$$

$\bar{x}$: average of sample population

n: number of sampled data (sample size)

σ^2: unbiased variance

σ: standard deviation of parent population, the character showing distribution of parent population

$$s^2 = \frac{SS}{n}$$

s^2: sample variance

s: deviation of sample population

6. $M(\bar{\mu} - \mu)^2$, expectation value of quadratic moment of average of sample population around proper mean of parent population is indicating confidential range of estimated mean from average of sample population.

$$M(\bar{\mu} - \mu)^2 = E(x - \mu)^2 = \frac{\sigma^2}{n}$$

$$\frac{\sigma^2}{n} : \text{square of standard error}$$

III-2-2-3. How to use binomial distribution.

Binomial distribution is a distribution model of discrete type data, we should develop our method for application to continuous type data. Normal distribution model is a development of binomial data for the application and chi square distribution is a model which often be used as the method to detect the difference among discrete type data. However, denominal distribution can be used in some case.

Here, the author illustrates two examples

First example: When we sampled 5 students from a class in school, all students were male.

If the sex ratio is 1:1, $p = \frac{1}{2}$

The possibility of $p(k)s$ are as follows

$$p(5) = \;_5C_5 \times \left(\frac{1}{2}\right)^5 = 1 \times \frac{1}{32} = \frac{1}{32} = 0.03125$$

$$p(4) = \;_5C_4 \times \left(\frac{1}{2}\right)^5 = \frac{5 \cdot 4 \cdot 3 \cdot 2}{4 \cdot 3 \cdot 2 \cdot 1} \times \frac{1}{32} = 5 \times \frac{1}{32} = \frac{5}{32} = 0.15625$$

$$p(3) = \;_5C_3 \times \left(\frac{1}{2}\right)^5 = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} \times \frac{1}{32} = 10 \times \frac{1}{32} = \frac{10}{32} = 0.3125$$

$$p(2) = \;_5C_2 \times \left(\frac{1}{2}\right)^5 = \frac{5 \cdot 4}{2 \cdot 1} \times \frac{1}{32} = 10 \times \frac{1}{32} = \frac{10}{32} = 0.3125$$

$$p(1) = \;_5C_1 \times \left(\frac{1}{2}\right)^5 = \frac{5}{1} \times \frac{1}{32} = 5 \times \frac{1}{32} = \frac{5}{32} = 0.15625$$

$$p(0) = \;_5C_1 \times \left(\frac{1}{2}\right)^5 = 1 \times \frac{1}{32} = \frac{1}{32} = 0.03125$$

The probability of all five students are same sex is 0.03125. The probability to observe this phenomenon is only 3.125%, when the sex ratio in the class is 1;1 and students are randomly sampled.   If we conclude that the sex ratio in the class is not 1:1 or that students are not randomly sampled, the possibility of errancy is only 3.125%. When the sample include a female student, the possibility is 0.15625. It is difficult to make conclusion. Five percent is often used for the threshold of judgement, but in several cases, we apply 1% or 0.1% as threshold in the case we need to keep higher safe level such as possibility of accident and harmful effect.

Second example.

There tow boxes α and β. Many red and white balls are inside of the boxes. We cannot

see the inside of the box. The ratio of red and white balls is 1:2 in box α, and the ratio is 1:1 in box β. We do not know which is α or β. We take a ball from each box. And confirm the color. When we repeat tree times. The results are as shown.

Box	1	2	3
A	Red	White	White
B	Red	Red	Red

We can conclude Box A is box α and Box B is box β, because when we conclude that box B is not α, risk rate is only $\frac{1}{27}$.

(1) When A is α, the possibility of 1red and 2 whites is

$$p(1) = \quad _3C_1\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^2 = 3 \times \frac{2^2}{3^3} = \frac{4}{9}$$

(2) When A is β, the possibility of 1red and 2 whites is

$$p(1) = \quad _3C_1\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2 = 3 \times \frac{1}{2^3} = \frac{3}{8}$$

(3) When B is α, the possibility of 3 red is

$$p(3) = \quad _3C_3\left(\frac{1}{3}\right)^3 = 1 \times \frac{1}{3^3} = \frac{1}{27}$$

(4) When B is β, the possibility of 3 red is

$$p(3) = \quad _3C_3\left(\frac{1}{2}\right)^3 = \frac{1}{2^3} = \frac{1}{8}$$

Possibility of A=α is not denied by (1)

Possibility of A=β is not denied by (2)

Possibility of B=α is denied by (3)

Possibility of B=β is not denied by (4)

We concluded that B is not α, then B must be β. A=α, A=β, B=α and B=β are null hypothesis. we can reach a conclusion when the null hypothesis. In (2), null hypothesis, A=β, is not denied, though we cannot conclude A=β from this result.
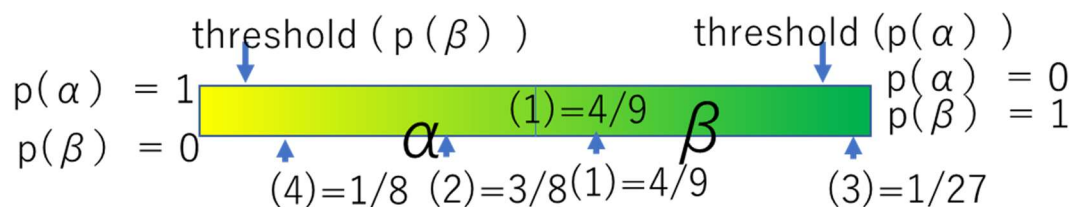

Fig 17. Spectrum of possibilities and positions of result.

Figure 17 is showing the positions of result in spectrum of possibility. Some readers may discover following tricky question. Threshold is arbitrary value. We determine the value considering risks, merits, and situations. In this case the author set the threshold at 0.04. However, if we set the threshold at 0.125, we should accept another conclusion that B=β. This means that we accept two conflicted conclusions (B=β, B=α). The author avoided the conflict by setting threshold at 0.04. Why did the author set threshold at 0.04. There is a reason why the threshold should be 0.04. We repeat similar 4 statistical tests. When we repeat statistical tests plural times. The possibility of false result in crease with increase of repeat times.   When possibility of false is $\frac{1}{8}$ in on trial and we repeat 4 trials, the possibility of appearance of false during 4 trials is $\frac{1}{8} \times 4 = \frac{1}{2}$=0.5. So, we can conclude that the result of (4) is an obtainable result by chance. On the contrary, if the possibility of false is 0.04. the possibility of appearance of false during 4 trials is $0.04 \times 4 = 0.16$. The author considered that this value is small enough.