ISSN 0973-8975

J.Mech.Cont. & Math. Sci., Vol.-8, No.-2, January (2014) Pages 1210-1216

A CIRCLE THEOREM IN THE SAMUELSON DOMAIN

By

[1]Mihir B. Banerjee, [2]J.R. Gupta, [3]R.G. Shandil and [4]Jyotirmoy Mukhopadhyay

[1]C-36/45, Sector-K, Aligunj, Lucknow, India.

[2, 3] Department of Mathematics, H.P. University, Shimla, India.

[4] Institute of Mechaninics of Continua and Mathematical Sciences, Kolkota, India.

Abstract

In the subject matter of mathematical statistics, let the domain of mathematical activity that draws its inspiration from and nurtures the lead provided by the seminal paper of the American Economist and Nobel Prize (1970) winner P.A. Samuelson entitled, "How Deviant can you be?" and published in the Journal of the American Statistical Association in 1968, on the maximum and the minimum deviations, from the mean (denoted presently by m and m' *respectively) in a set of n observations with given mean* μ *and standard deviation* σ*, be henceforth defined as the Samuelson Domain. The present communication is in the Samuelson Domain. A circle theorem in the* $m\sigma-$*plane is rigorously established and exhibited step by step for the sheer delight of its simplicity and elegance. A crude first approximation yields a result that is inferior to Samuelson's but a more precise investigation of the consequences of the circle theorem shows that Samuelson's famous work on the existence of bounds, for a set of n real numbers, in terms of* μ, σ *and n can be improved upon provided n exceeds a critical value.*

Keywords and phrases : Samuelson Domain, mean, standard deviation, maximum and the minimum deviations, critical value

বিমূর্ত সার (Bengali version of the Abstract)

গাণিতিক পরিসংখ্যানের বিষয়বস্তুতে গাণিতিক ক্রিয়ার ক্ষেত্র প্রসঙ্গে অনুপ্রেরণা এবং পরর্বতী উন্নয়নের পথ প্রর্দশকের অবদান মার্কিন অর্থনীতিবিদ এবং নোবেল পুরস্কার বিজয়ী (1970) পি-এ-স্যামুইলসন (P.A. Samuelson) যুগান্তকারী পত্র "How Deviant can you be?" শিরনামে আমেরিকান স্ট্যাটিস্টিক্যাল এ্যসোসিয়েশনের জার্নালে 1968 সালের প্রকাশিত পত্র , মধ্যক থেকে বৃহত্তম ও

ক্ষুদ্রতম বিচ্যুতি যখন প্রদত্ত মধ্যক μ *এবং সমক পার্থক্য* σ *সহ* n *- সংখ্যক পর্যবেক্ষনের সেটের মধ্যক (বর্তমানে যথাক্রমে* m *এবং* m' *হিসাবে চিহ্নিত করা হয়েছে) । এরপর থেকে ইহাকে স্যামুইলসনের ক্ষেত্র হিসাবে সংজ্ঞাত করা হবে । বর্তমান পত্রটি সকল অর্থে স্যামুইলসন ক্ষত্রের অর্ন্তগত ।* $m\sigma$ *– সমতলে একটি বৃত্ত - উপপাদ্যকে বিস্তৃত ভাবে প্রতিষ্ঠিত করা হয়েছে এবং ইহার সারল্য এবং সৌন্দর্য্যের বিশুদ্ধ আনন্দের জন্য ধাপে ধাপে প্রর্দশন করা হয়েছে । অপরিপক্ক প্রথম আসন্নমানে (A crude first approximation) একটি ফলাফল প্রদান করে , যাহা স্যামুইলসনের মান থেকে নিম্নতর কিন্তু বৃত্ত-উপপাদ্যের পরিনতির আরও অধিক যথাযথ অনুসন্ধান দেখায় যে* n *- বাস্তব সংখ্যার সেটের জন্য সীমাবদ্ধতার অস্তিত্বের উপর স্যামুইলসনের বিখ্যাত কাজটিকে* m, σ *-এর* n *- পদের সাহায্যে উন্নীত করা যায় যদি* n *ক্রান্তিক মানকে অতিক্রম করে ।*

1. MATHEMATICAL ANALYSIS

Let the discrete random statistical variate x take n values $x_1, x_2, x_3, \cdots, x_n$ such that $a = x_{\min} = x_1, x_2, x_3, \cdots x_n = x_{\max} = b$. We define the mean μ, the standard deviation σ and the range r, of the variate x as

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

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \,, \tag{2}$$

$$= \sqrt{\frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - n\mu^2\right]}$$

and $$r = x_{\max} - x_{\min} = b - a \,. \tag{3}$$

The two well-known theorems are then valid in the above context.

Samuelson's (1968) Theorem on the Existence of Bounds in terms of μ, σ and n:

The minimum of the above set of real numbers is bounded below and the maximum is bounded above in terms of their mean μ, standard deviation σ and the number of values n.

Samuelson's theorem implies that

$$\mu - \sigma\sqrt{n-1} \le x_i \le \mu + \sigma\sqrt{n-1} \qquad \text{for all} \quad i = 1,2,3,\cdots,n \tag{4}$$

or equivalently

$$(\mu - x_i)^2 \le (n-1)\,\sigma^2 \qquad \textit{for all} \quad i = 1,2,3,\cdots,n\,. \tag{5}$$

We therefore have

$$m^2 = \max_{(i)}(\mu - x_i)^2 \le (n-1)\,\sigma^2\,. \tag{6}$$

Banerjee-Shandil (1995) Semi-Circle Theorem

The point (μ,σ) must of necessity lie within the semi-circle in the upper half of the $\mu\sigma-$plane which has the range of the variate x as diameter.

Banerjee-Shandil Semi-Circle Theorem implies that

$$\left(\mu - \frac{a+b}{2}\right)^2 + \sigma^2 \le \left(\frac{b-a}{2}\right)^2. \tag{7}$$

Although ultimate priority seems hard to pin down there is no denying the fact that the present geometrical version of the statement of inequality (7) was first published, in the context of statistical analysis, in the year 1995, in the Journal of the Bharat Ganita Parishad, by Banerjee and Shandil.

We now make use of the above two theorems to establish the main result of the present paper.

Theorem on Circle of Freedom for the point (μ,σ):

The point (m,σ) in the $m\sigma-$plane must necessarily lie within the circle whose centre is at $\left(0,\frac{\sqrt{n-1}}{2}r\right)$ and radius equal to $\frac{\sqrt{n-1}}{2}r$.

Proof. Consider two mutually exclusive cases namely (a) $\mu \ge \frac{a+b}{2}$ (b) $\mu < \frac{a+b}{2}$.

When case (a) is valid, we have

$$\frac{a+b}{2} \le \mu \le b, \text{ since } a \le \mu \le b.$$

Further, since $m = \max\limits_{(i)}(\mu - x_i)$, it follows that

$$|m| = \left|\max_{(i)}(\mu - x_i)\right| = \max_{(i)} |(\mu - x_i)|$$

$$= |\mu - a|, \text{ since } \mu \ge \frac{a+b}{2},$$

$$= \left|\mu - \frac{a+b}{2} + \frac{a+b}{2} - a\right|,$$

$$\le \left|\mu - \frac{a+b}{2}\right| + \left|\frac{a+b}{2} - a\right|,$$

$$\le \left|\mu - \frac{a+b}{2}\right| + \frac{b-a}{2}, \text{ since } b \ge a \tag{8}$$

$$\le \left|\mu - \frac{a+b}{2}\right| + R, \text{ where } R = \frac{r}{2}. \tag{9}$$

Inequality (9) implies that

$$\left(|m| - R\right)^2 \le \left(\mu - \frac{a+b}{2}\right)^2, \tag{10}$$

which upon using Banerjee-Shandil semi-circle theorem yields

$$m^2 + R^2 - 2|m|R \le R^2 - \sigma^2. \tag{11}$$

When case (b) is valid, we have

$$a \le \mu < \frac{a+b}{2}, \text{ since } a \le \mu \le b.$$

Further, since $m = \max\limits_{(i)}(\mu - x_i)$, it follows that

$$|m| = \left|\max_{(i)}(\mu - x_i)\right| = \max_{(i)} |(\mu - x_i)|$$

$$= |\mu - b|, \text{ since } \mu < \frac{a+b}{2},$$

$$= \left| \mu - \frac{a+b}{2} + \frac{a+b}{2} - b \right|,$$

$$\leq \left| \mu - \frac{a+b}{2} \right| + \left| \frac{a+b}{2} - b \right|,$$

$$\leq \left| \mu - \frac{a+b}{2} \right| + \frac{b-a}{2}, \text{ since } b \geq a$$

$$\leq \left| \mu - \frac{a+b}{2} \right| + R, \text{ where } R = \frac{r}{2}$$

which is same as (9) and in this case also we obtain (11).

Now using Samuelson's theorem on bounds, we derive from inequality (11) that

$$m^2 - 2\sqrt{n-1}\,\sigma R + \sigma^2 \leq 0. \tag{12}$$

Let $R\sqrt{n-1} = s$, inequality (12) gives

$$m^2 - 2\sigma s + \sigma^2 \leq 0. \tag{13}$$

Adding s^2 to both sides of inequality (13), we get

$$m^2 + s^2 - 2\sigma s + \sigma^2 \leq s^2, \tag{14}$$

that is,

$$m^2 + (\sigma - s)^2 \leq s^2, \tag{15}$$

and hence the theorem.

Theorem on conditional superiority of the present bound on m^2 over that of Samuelson's:

If $n \geq \frac{2s}{\sigma}$

then

$$m^2 \leq 2s\sigma - \sigma^2 \leq (n-1)\sigma^2.$$

Proof. Since $n \ge \frac{2s}{\sigma}$, we have

$$n\sigma^2 \ge 2s\sigma,$$

that is,

$$2s\sigma - \sigma^2 \le n\,\sigma^2 - \sigma^2. \qquad (16)$$

Combining inequalities (13) and (16), we derive

$$m^2 \le 2s\sigma - \sigma^2 \le n\,\sigma^2 - \sigma^2 \le (n-1)\,\sigma^2, \qquad (17)$$

and hence the theorem.

An Example of the illustrative case:

Let $D = \{x_1, \cdots, x_n\}$ he any distribution

$$a = x_{min} \le x_1, \cdots, x_n \le b = x_{max}.$$

Let

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \ \text{ and } \ \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2,$$

$$R = \frac{x_{max} - x_{min}}{2} \ \text{ and } \ s = R\sqrt{n-1}.$$

Suppose, in the $\mu\sigma -$plane, the point (μ, σ) lies inside the triangle ABC

where $A = (a,0)$, $B = (b,0)$ and $C = \left(\frac{a+b}{2}, R\right)$

then

$$\mu - \sigma \ge a \quad \text{and} \quad \mu + \sigma \le b. \qquad (18)$$

Let k be any positive integer.

Let D' be the distribution $\left\{x_1, \ldots\ldots, x_n, \frac{\mu-\sigma, \cdots, \mu-\sigma}{(k \ terms)}, \frac{\mu+\sigma, \cdots, \mu+\sigma}{(k \ terms)}\right\}$.

Let $x', \mu', \sigma', R', s'$ correspond to D'. Then

(i) $n' = n + 2k$,

(ii) $\mu' = \frac{1}{n+2k}\left[\sum_{i=1}^{n} x_i + k(\mu-\sigma) + k(\mu+\sigma)\right] = \frac{1}{n+2k}\left[n\mu + 2k\mu\right] = \mu$

(iii) $\sigma'^2 = \frac{1}{n+2k}\left[\sum_{i=1}^{n} (x_i-\mu)^2 + 2k\sigma^2\right] = \frac{1}{n+2k}\left[n\sigma^2 + 2k\sigma^2\right] = \sigma^2$

(iv) $R' = R$ by (18),

(v) $s' = \sqrt{n'-1}\,.R' = \sqrt{n+2k-1}\,.R$.

Now

$$\frac{n'}{s'} = \frac{(n+2k)}{R\sqrt{n+2k-1}} > \frac{1}{R}\sqrt{n+2k} \quad i.e. \quad \frac{n'}{s'} \to \infty \quad as \quad k \to \infty .$$

It follows that for sufficiently large value of k

$$\frac{n'}{s'} > \frac{2}{\sigma} \qquad \text{i.e.} \qquad n' > \frac{2s'}{\sigma'} .$$

It is well known that corresponding to every point (μ, σ) inside the triangle ABC there is a distribution of the type D. Thus there is a large class of distributions satisfying

$$n > \frac{2s}{\sigma} .$$

REFERENCES

1) Samuelson,P.A., How deviant can you be?, J. American Statistical Association, **63**,1522-1525,1968.

2) Banerjee, M.B. and Shandil, R.G., A theorem on mean and standard deviation of a statistical variate, Ganita,**46**, 21-23,1995.