Agri. Res. J. Kerala, 1978 16 (]) 24—27

PROPERTIES OF THE SAMPLE VARIANCE IN SAMPLES FROM A GENERALISED AUTOREGRESSIVE SCHEME

E. J. THOMAS

College of Agriculture, Vellayani, Kerala

The distribution of sample variance $s^2 = \Sigma (x_i - \bar{x})^2 / (n-1)$, where the x_i's are randomly and independently drawn from a normal population with mean a and variance S^2, is,

$$P(s^2) = \frac{\left(\frac{n-1}{2S^2}\right)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} . e^{\frac{-(n-1)s^2}{2S^2}} . (s^2)^{\frac{n-1}{2}-1}$$

(Rietz, 1937J. The moments of s^2 are as follows. Mean $(M_1') = S^2$, variance $(M_2) = (\frac{2}{n-1}). S^4$, $M_3 = \frac{8}{(n-1)^2} . S^6$, $M_4 = \frac{12(n+3)}{(n-1)^3} . S^8$. Thus

$\beta_1 = 8/(n-1)$ and $\beta_2 = 3 + \frac{12}{n-1}$. ('Student', 1908)

In this paper, x_i's are assumed to follow a generalised autoregressive scheme,

$$x_t + B_1.x_{t-1} + B_2.x_{t-2} + ... + B_p.x_{t-p} = e_t + C_1.e_{t-1} + C_2.e_{t-2} .... + C_n.e_{t-n},$$

where the e_t's follow a normal distribution with mean 0 and variance S^2. Further, e_t's are assumed to be independent. Under this scheme, the properties of s^2 are investigated. The algebra involved in the computation of the third and fourth moments tend to be laborious. The methods are utilised to find the first four moments and also β_1 and β_2, for the simpler model, $x_t = e_t + a. e_{t-1}$.

Materials and Methods

The variable x_t follows the generalised autoregressive scheme,

$x_t + B_1.x_{t-1} + B_2.x_{t-2} + ... + B_p.x_{t-p} = e_t + C_1.e_{t-1} + C_2.e_{t-2} + ... + C_n.e_{t-n}$. This can be written in the form,

$x_t = e_t + a_1.e_{t-1} + a_2.e_{t-2} + ...$ where,

$$\frac{1 + C_1.u + C_2.u^2 + ... + C_n.u^n}{1 + B_1.u + B_2.u^2 + ... + B_p.u^p} = 1 + a_1.u + a_2.u^2 + ...$$

Since the e, 's are assumed to be normally and independently disiributed with mean 0 and variance S^2, it follows that, $E(x_t) = 0$.

$E(x_t^2) = (1 + a_1^2 + a_2^2 + \ldots). S^2 = A_0 . S^2$ (say) and

$E(x_t.x_{t+k}) = (a_k + a_1.a_{k+1} + a_2.a_{k+2} + \ldots) S^2 = A_k.S^2$, where

$A_k = a_k + a_1.a_{k+1} + a_2.a_{k+2} + \ldots\ldots$

To obtain the four moments of s^2, the usual procedure of finding expectations of the first four powers of s^2 is adopted. This involves expansions of expressions of the from $(\Sigma x_t^2)^p . C \Sigma x_t)^{2q}$, where p and q are zero or positive integers such that $p < 4$, $q < 4$, and $p + q = 4$.

$$(\Sigma x_t^2)^2 = \Sigma x_t^4 + 2 \sum_{r=1}^{n-1} \sum_{t=1}^{n-r} x_t^2.x_{t+r}^2$$

$$(\Sigma x_t^2)^3 = \Sigma x_t^6 + 3 \sum_{r=1}^{n-1} \sum_{t=1}^{n-r} \sum_{42} x_t^4.x_{t+r}^2 + 6 \sum_{r=1}^{n-2} \sum_{s=1}^{n-r-} \sum_{t=1}^{n-r-s} x_t^2.x_{t+r}^2.x_{t+r+s}^2$$

where $\sum_{42} x_t^4 x_{t+r}^2 = x_t^4 x_{t+r}^2 + x_t^2 x_{t+r}^4$

The other expressions, being still more lengthy, are not listed here,

Expectations of terms iike $x_t^{2k_1} . x_{t+r}^{k_2} . x_{t+r+s}^{k_3} . x_{t+r+s+w}^{k_4} . x_{t+r+s+w+v}^{k_5}$. If. $x_{t \ldots s\, w+v\, u}$, where $k_2, k_3, \ldots\ldots k_7$ can take values from 0 to 7 and k_1 takes values from 0 to 4, such that $2.k_1 + k_2 + k_3 + k_4 + k_5 + k_6 + k_7 < 8$, are required for the first four moments of s^2. The following results are easily derived.

$E(x_t) = 0$, $E(x_t^2) = A_0.S^2$, $E(x_t x_{t+r}) = A_r.S^2$, $E(x_t^4) = 3.A_0^2.S^4$

$E(x_t^2.x_{t+r}^2) = (A_0^2 + 2A_r^2) S^4$, $\sum_{31} E(x_t^3 . x_{t+t}) = 6.A_0.A_r.S^4$

$E(x_t.x_{t+r}.x_{t+r+s}.x_{t+r+s+w}) = (A_r.A_w + A_s.A_{r+s+w} + A_{r+s}.A_{s+w}) . S^4$

$E(x_t^6) = 15.A_0^3.S^6$, $\sum_{51} E(x_t^5.x_{t+r}) = 30.A_0^2.A.S^6$

$\sum_{42} E(x_t^4.x_{t+r}^2) = (6.A_0^3 + 24.A_0.A_r^2) S^6$, $E(x_t^3.x_{t+r}^3) = (9.A_0^2.A_t + 6.A_r^3).S^6$

Expections of other expressions are still more lengthy and hence are not reproduned here.

Results and discussion

Using the methods described, it is possible to derive the moments of s^2, where $s^2 = \Sigma (x_i - \bar{x})^2/(n-1) = (\Sigma x_i^2)/(n-1) - (\Sigma x_i)^2/n(n-1)$

Thus, $$M_1' = \left(A_0 - \frac{2}{n(n-1)} \sum_{r=1}^{n-1} (n-r) A_r \right) S^2$$

$$M_2 = \frac{S^4}{n^2(n-1)^2} \Big[2n^2(n-1).A_0^2 + 4n(n-2) \sum_{r=1}^{n-1} (n-r).A_r^2 + 4n(n-4) \sum_{r=1}^{n-1} (n-r) A_0 A_r + 8\Big[\sum_{r=1}^{n-1} (n-r).A_r\Big]^2 - 4n \sum_{r=1}^{n-2} \sum_{s=1}^{n-r-1} (n-r-s) \Big\{ A_0 (A_r + A_s + A_{r+s}) + 2(A_r.A_s + A_r.A_{r+s} + A_s.A_{r+s}) \Big\} \Big]$$

Expressions for M_3 and M_4 in the general case happen to be very lengthy and hence are not given in this paper.

For a simple particular case it is easy to work out the moments. Hence the model, $x_t = e_t + a.e_{t-1}$ is taken to illustrate the method. In this case, $A_0 = 1 + a^2$, $A_1 = a$ and $A_p = 0$ for p greater than 1. This leads to the following results.

$$M_1' = S^2 [1 - (2a/n) + a^2]$$

$$M_2 = \frac{S^4}{n^2(n-1)^2} \Big[2n^2(n-1) - 8n(n-1).a + 8(n^3 - 2n^2 + n + 1).a^2 - 8n(n-1).a^3 + 2n^2(n-1).a^6 \Big]$$

$$M_3 = \frac{S^6}{n^3(n-1)^3} \Big[8n^3(n-1) - 48n^2(n-1).a + 24n(3n^3 - 7n^2 + 4n + 4).a^2 - 32(5n^3 - 6n^2 - 3n - 2).a^3 + 24n(3n^3 - 7n^2 + 4n + 4).a^4 - 48n^2(n-1).a^5 + 8n^3(n-1).a^6 \Big]$$

$$M_4 = \frac{S^8}{n^4(n-1)^4} \Big[12n^4(n-1)(n+3) - 16n^2(n-1)(6n^2 + 18n - 3681).a + 96n(n^5 + 2n^4 + 3n^3 - 40n^2 + 106n - 90).a^2 + 48n(8n^4 - 58n^3 - 3457n^2 + 3329n + 320).a^3 + 24(9n^6 + 14n^5 + n^4 - 252n^3 + 684n^2 - 280n + 40).a^4 + 48n(8n^4 - 58n^3 - 3457n^2 + 3329n + 320).a^5 + 96n(n^5 + 2n^4 + 3n^3 - 40n^2 + 106n - 90).a^6 - 8n^2(n-1)(12n^2 + 36n - 7362).a^7 + 12n^4(n-1)(n+3).a^8 \Big]$$

When a = 0, these reduce to $M_1' = S^2$, $M_2 = \left(\frac{2}{n-1}\right) S^4$,

$$M_3 = \frac{8}{(n-1)^2} S^6$$

$M_4 = \frac{12(n+2)}{(n-1)^3} S^8$, agreeing with the results for the case where x_t's are randomly and indepently distributed.

When $a \neq 0$ also, it can be seen that β_1 tends to 0 and β_2 tends to 3 as n tends to infinity. It is difficult to get the exact distribution of s^2 in this case. But the distribution can be approximated by a normal distribution with mean $(1+a^2). S^2$ and variance $(2/a) . (a^4 + 8n^2 + 2).S^4$. It is obvious that s^2 is a biased estimate of s^2 for $a \neq 0$

Summary

An attempt is made to study the properties of the sample variance in samples from a generalised autoregressive scheme, by computing the moments. Eventhough the expressions are lengthy, the moments can be derived. For the simple particular case $x_t = e_t + a.e_{t-1}$, the moments have been worked out. The exact distribution has not been derived, but the large sample approximation is found to be normal with mean $S^2 (1 + a^2)$ and variance $(2/n) (a^4 + 8a^2 + 2).S^4$ When $a \neq 0$, s^2 is a biased of S^2.

സംഗ്രഹം

സാമാന്യ വൻകൃത സ്വയംസമാശ്രയണ പദ്ധതിയിൽ നിന്നുള്ള സാമ്പിൾ വ്യതിയാനത്തിന്റെ ഗുണധർമ്മങ്ങളെപ്പറ്റി പഠിക്കുവാൻ ഇതിന്റെ ആഘൂർണ്ണങ്ങളുടെ പരികലനം വഴി ഒരു ശ്രമം നടത്തിയിരിക്കുന്നു. വ്യംജകങ്ങൾ വളരെ ദൈർഘ്യമുള്ളവയാണെന്നു വരികിലും ആഘൂർണ്ണങ്ങളെ വ്യൽപാദിപ്പിക്കുവാൻ സാദ്ധ്യമാണ്. $x_t = e_t + a. e_{t-1}$ എന്ന ലഘു വിശേഷസ്ഥിതിയിൽ ആഘൂർണ്ണങ്ങളെ പരികലനം ചെയ്തിട്ടുണ്ട്. കൃത്യമായ വിതരണ ഫലനങ്ങൾ വ്യൽപാദിപ്പിക്കപ്പെട്ടിട്ടില്ലെങ്കിലും വിതരണത്തിന്റെ ബൃഹത് സാമ്പിൾ ഏകദേശനം, $S^2 (1 + a^2)$fflocu^ojo $\frac{2}{n}(a^4 + 8a^2 + 2)S^4$ വ്യതയാനവുമായുള്ള നോർമൽ വിതരണമാണെന്നു കാണുന്നു. $a \neq 0$ ആയിരിക്കുമ്പോൾ s^2 എന്നത് S^2 ന്റെ അഭിനതിയുള്ള ആകലനം ആകുന്നു.

REFERENCES

Rietz. H. L. 1937. some topics in sampling theory. *Bull. Amer. Math. Soc. April.* 1937.

'Student'. 1908. Probable error of the mean. Biometrika. 6.

(M. S. Received 27-4-1977)