THE GEGENBAUER DISTRIBUTION REVISITED : SOME RECURRENCE RELATIONS FOR MOMENTS, CUMULANTS, ETC., ESTIMATION OF PARAMETERS AND ITS GOODNESS OF FIT

By

M. BORAH

Science College Kohima, Nagaland

(Received : January, 1982)

SUMMARY

The Gegenbauer distribution is further investigated by working out recursive relationship of the probabilities for computing individual probabilitihs and recurrence reburrhnce relations for moments, cumulants. Estimation of the parameters by the method of cumulants is studied. The distribution is fitted to biological data for an empirical comparison. The resulting fit is found to be good in comparison to others.

INTRODUCTION

Plunkett and Jain [5] obtained the Gegenbauer distribution by mixing the Hermite distribution with the Gamma distribution. Further they have studied some of its limiting cases and its goodness of fit by using (1) Student's historic data on haemocytometer counts of yeast cells and (II) Accidents to 647 women working on H.E. Shells during 5 weeks.

In this paper the Gegenbauer distribution is further investigated. Recursive relationship of the probabilities have been suggested for computing the individual probabilities. Further the recurrence relations for moments, cumulants, etc. are also investigated. Estimation of the parameters by the method of cumulants is also studied. The distribution is fitted to one sets of biological data for an empirical comparison. It has been observed that the resulting fit is quite good in comparison to others, Gegenbauer Polynomial of order n may be defined as

$$G_n^\lambda(\alpha,\beta)=\sum_{j=1}^{[\frac{n}{2}]}\frac{n!}{j!}\beta^j\frac{\Gamma(n+\lambda-j)\,\alpha^{n-2j}}{\Gamma(n-2j+1)\,\Gamma(\lambda)} \quad \ldots(1)$$

($[k]$ denotes the integer part of k), where α and β are positive constants such that $\alpha+\beta<1$, and $\lambda>o$. Plunkett and Jain [5] have obtained Gegenbauer distribution with probability generating function (pgf)

$$g(s)=(1-a-\beta)^{-\lambda}(1-\alpha s-\beta s^2)^{-\lambda} \quad \ldots(2)$$

RECURRENCE RELATIONS FOR POLYNOMIALS

Differentiating the gf of Gegenbauer polynomial Plunkett and Jain have obtained the recurrence relation viz.

$$G_r^\lambda(\alpha,\beta)=\lambda\,\alpha\,G_{r-1}^{\lambda+1}(\alpha,\beta)+2\,\lambda\,\beta\,(r-1)\,G_{r-1}^{\lambda+1}(\alpha,\beta),\ r\geqslant 2 \quad \ldots(3)$$

Again equating the coefficient of s^r on both sides of the identity

$$(1-\alpha s-\beta s^2)^{-(\lambda+1)}=(1-\alpha s-\beta s^2)^{-\lambda}/(1-\alpha s-\beta s^2),$$

they have obtained another recurrence relation as

$$G_r^\lambda(\alpha,\beta)=G_r^{\lambda+1}(\alpha,\beta)-r\alpha\,G_{r-1}^{\lambda+1}(\alpha,\beta)-\beta r(r-1)\,G_{r-1}^{\lambda+1}(\alpha,\beta),\ r\geqslant 2 \quad \ldots(4)$$

Equation (3) and (4) yield the following interesting recurrence relation among the polynomials

$$G_r^\lambda(\alpha,\beta)=\alpha\,(r+\lambda)\,G_r^\lambda(\alpha,\beta)+r\beta(2\lambda+r-1)\,G_{r-1}^\lambda(\alpha,\beta),\quad r\geqslant 2 \quad \ldots(5)$$

where

$$G_0^\lambda(\alpha,\beta)=1,$$

$$G_1^\lambda(\alpha,\beta)=\alpha\lambda\,G_0^\lambda(\alpha,\beta).$$

It can conveniently be used for the computation of the polynomials of higher order by putting $r=2,3,4,\ldots$ etc. in (5).

PROBABILITIES AND RECURRENCE RELATIONS

$$P_n=\frac{\partial^n g(s)}{\partial s^n}\Big/ n!\ \Big|_{s=0} \quad \ldots(6)$$

we have $P_0=(1-\alpha-\beta)^\lambda$

$P_1 = \alpha\lambda\,(1-\alpha-\beta)^\lambda$

$P_n = P_0\, G_n^\lambda(\alpha,\beta)\,/\,n!$, for any positive integral values of n. Obviously we may obtain the recurrence relation for the probabilities of the Gegenbauer distribution by using the relations (5) and (6) as

$$P_{r+1}=\frac{1}{r+1}\,[\,\alpha\,(\lambda+r)\,P_r+\beta\,(2\lambda+r-1)\,P_{r-1}\,],\ r\geqslant 1 \quad ...(7)$$

This relation is often found very useful for the computation of individual probabilities for the fitting of the distribution.

RECURRENCE RELATIONS FOR MOMENTS

The moment generating function (mgf) of the distribution is given as

$$M(s)=(1-\alpha-\beta)^\lambda\,(1-\alpha e^s-\beta e^{2s})^{-\lambda} \quad ...(8)$$

Differentiating (8) *w.r.t.* α and β and equating the coefficients $\frac{s^r}{r!}$ from both sides, we have

$$\mu'_{r+1}=\alpha\frac{\partial\mu'_r}{\partial\alpha}+2\beta\,\frac{\partial\mu'_r}{\partial\beta}+\frac{\lambda(\alpha+2\beta)}{(1-\alpha-\beta)}\mu'_r \quad ...(9)$$

and

$$\mu_{r+1}=\alpha\frac{\partial\mu_r}{\partial\alpha}+2\beta\,\frac{\partial\mu_r}{\partial\beta}+\frac{r(\alpha+4\beta-\alpha\beta)}{(1-\alpha-\beta)^2}\mu_r \quad ...(10)$$

where $\mu'_r=E(X^r)$, $\mu_r=E[\{X-E(X)\}^r]$, the rth order raw and central moments respectively. It is also noted that the moments recurrence relations may conveniently be written as

$$\mu'_{r+1}=\lambda\sum_{j=0}^{r}\binom{r}{j}\left(\frac{\alpha+2^{j+1}\beta}{1-\alpha-\beta}\right)\mu'_{r-j}+\sum_{j=0}^{r-1}\binom{r}{j+1}\left(\frac{\alpha+2^{j+1}\beta}{1-\alpha-\beta}\right)\mu'_{r-j} \quad ...(11)$$

and

$$\mu_{r+1}=\lambda\sum_{j=0}^{r}\binom{r}{j}\left[\frac{(1-\alpha-\beta)(\alpha+2^{j+1}\beta)+(\alpha+2\beta)(\alpha+2^{j}\beta)}{(1-\alpha-\beta)^2}\right]\mu_{r-j}$$

$$+\sum_{j=0}^{r-1}\binom{r}{j+1}\left(\frac{\alpha+2^{j+1}\beta}{1-\alpha-\beta}\right)\mu_{r-j}-\lambda\left(\frac{\alpha+2\beta}{1-\alpha-\beta}\right)\mu_r \quad ...(12)$$

which are independent of any differential form.

5. Recurrence Relations for Cumulants

Differentiating the cumulant *gf* of the distribution w.r.t. α and β and equating the coefficients of $\frac{s^r}{r!}$ from both sides, we have

$$K_{r+1}=\alpha\frac{\partial K_r}{\partial\alpha}+2\beta\frac{\partial K_r}{\partial\beta},\ r>1 \quad ...(13)$$

where K_r denotes the rth order cumutant of the distribution. Similarly again differentiating the *cgf* of the distribution w.r.t. s and equating the coefficient of $\frac{s^r}{r!}$ from both sides, we have another recurrence relation for the cumulants as

$$K_{r+1}=\lambda\frac{(\alpha+2^{r+1}\beta)}{1-\alpha-\beta}+\sum_{j=0}^{r-1}\binom{r}{j+1}\left(\frac{\alpha+2^{j+1}\beta}{1-\alpha-\beta}\right)K_{r-j},\ r\geqslant1 \quad ...(14)$$

Cumulants of higher order may be obtained by putting $r=1, 2, 3, 4...$ etc. in the equation (13), where $K_1=\frac{\lambda(\alpha+2\beta)}{1-\alpha-\beta}$. Thus we have

$$K_2=\frac{\lambda(\alpha+4\beta)}{1-\alpha-\beta}+\left(\frac{\alpha+2\beta}{1-\alpha-\beta}\right)K_1$$

$$K_3=\lambda\left(\frac{\alpha+8\beta}{1-\alpha-\beta}\right)+2\left(\frac{\alpha+2\beta}{1-\alpha-\beta}\right)K_2+\left(\frac{\alpha+4\beta}{1-\alpha-\beta}\right)K_1 \quad ...(15)$$

6. Recurrence Relations for Factorial Moments

If $\mu_{(r)}$ denotes the rth order factorial moment of the distribution, the recurrence relation may by obtained as

$$\mu_{(r+1)}=\left(\frac{\alpha+2\beta}{1-\alpha-\beta}\right)(r+\lambda)\mu_{(r)}$$

$$+\left(\frac{r\beta}{1-\alpha-\beta}\right)(2\lambda+r-1)\mu_{(r-1)},\ r\geqslant1 \quad ...(16)$$

Now, putting $r=1, 2, 3. ...$in (14), we have

$$\mu_{(2)}=\lambda(\lambda+1)\left(\frac{\alpha+2\beta}{1-\alpha-\beta}\right)^2+\left(\frac{2\lambda\beta}{1-\alpha-\beta}\right).$$

$$\mu_{(3)}=\lambda(\lambda+1)(\lambda+2)\left(\frac{\alpha+2\beta}{1-\alpha-\beta}\right)^3+\frac{6\lambda(\lambda+1)\beta(\alpha+2\beta)}{(1-\alpha-\beta)^2}$$

where

$$\mu_{(1)}=\frac{\lambda(\alpha+2\beta)}{1-\alpha-\beta}.$$

7. Goodness of Fit

The distribution was fitted to Student's historic data on haemocytometer counts of yeast cells for which two-parameter Hermite (Kemp and Kemp [2]) and Negative binomiai (Bliss, [1]) and three parameter Neyman Type A (Neyman, [4]) distributions respectively have been fitted. Since obtaining maximum likelihood estimates is very cumbersome, adhoc methods were used to estimate the parameters. The parameters α, β and λ of Gegenbauer distribution can be estimated in the following methods.

(*a*) Estimation from first three sample moments.

$$\hat{\lambda}=\frac{3K_1(K_2-K_1)\pm K_1\sqrt{(9(K_1-K_2)^2-4K_1A)}}{2A}$$

$$\hat{\alpha}=\frac{2\hat{\lambda}(2K_1-K_2)+2K_1^2}{D} \qquad \ldots(17)$$

$$\hat{\beta}=\frac{\hat{\lambda}(K_2-K_1)+K_1^2}{D}$$

where

$$A=K_3+2K_1-3K_2$$

$$D=2\hat{\lambda}^2+3\hat{\lambda}K_1-K_1^2-\hat{\lambda}K_2$$

(*b*) Estimation from first two sample moments and ratio of first two frequencies in the sample.

Let us take

$$\theta=\frac{f_1}{f_0}=\frac{P_1}{P_0}=\alpha\lambda,$$

Hence λ can be calculated in terms of mean, variance and θ as.

$$\hat{\lambda}=\frac{-B\pm\sqrt{(B^2-4AC)}}{2A}, \qquad \ldots(18)$$

where

$$A=2\{\theta-(2K_1-K_2)\}$$

$$B=(3K_1-K_2)\,\theta-2K_1^2$$

$$C=\theta K_1^2$$

Then α and β can be calculated from the last two equations of (17).

TABLE 1

Haemocytometer Counts of Yeast Cells

No. of Yeast Cells per square	*Observed frequencies*	*Gegenbauer (Method of three moments)*	*Gegenbauer (Plunkett and Jain 1975)*	*Negative Binomial (Bliss 1953)*	*Hermite (Kemp and Kemp 1965)*	*Neyman Type A (Neyman, 1939)*
0	213	214.15	215.11	214.27	213.12	214.8
1	128	123.00	121.00	122.65	122.91	121.3
2	37	44.88	45.44	45.00	46.71	45.7
3	18	13.36	13.85	13.42	13.31	13.7
4	3	3.55	3.62	3.54	3.16	3.6
5	1	0.86	0.81	0.86	0.64	0.8
6	0	0.20	0.17	0.26	0.15	0.1
Total	400	400.00	400.00	400.00	400.00	400.0
χ^2		2.8342	3.3151	3.3196	3.8825	3.4466
		$\hat{\alpha}$=.1980200	$\hat{\alpha}$=.4500459409			
		$\hat{\beta}$=.00410075	$\hat{\beta}$=.05880742304			
		$\hat{\lambda}$=2.89828	$\hat{\lambda}$=1.249822047			

It is noted that two sets of estimates (α, β, λ) are obtained corresponding to the values of λ. The chosen sets are those which satisfy the condition $(1-\alpha-\beta)<1$, $\lambda>0$ and give the smaller χ^2 value to the goodness of fit.

When mean and variance were large, use was made of the method of the first three moments. When the first two frequencies were large in comparison with remaining, the method involving the first two moments and the ratio of the first two frequencies was used in estimation. For the sake of reference, the fit of Gegenbauer as given by Plunkett and Jain [5], Negative Binomial by Bliss [1], Hermite by Kemp and Kemp [2] and Neyman Type A by Neyman [4] are also given alongside in Table I.

ACKNOWLEDGEMENT

The author is very much grateful to Professor J. Medhi, Head of the Department of Statistics for his guidance and supervision.

REFERENCES

(1) Bliss, C.I. (1953) : Fitting the negative binomial distribution to biological data. *Biometrics*, **9**, 176-196.

(2) Kemp, C.D. and Kemp, A.W. (1965) : Some properties of the Hermite distribution. *Biometrika*, **52**, 381-394.

(3) Lahiri, M. (1971) : On a generalization of Hermite polynomials. *Proc. Amer. Math. Soc.* **27**, 117-121.

(4) Neyman, J. (1939) : On a class of 'Contagious' distributions applicable in entomology and bacterio logy, *Ann. Math. Statist.* **10**, 35-57.

(5) Plunkett, I.G and Jain G.C. (1975) : Three generalized negative binomial distributions. *Biom. Z.* **17**, 286—302.