UDC 551.501.777:519.241.7:551.577.35

Note on the Sample Size to Achieve Normality for Estimators for the Gamma Distribution

L. R. SHENTON—*Office of Computer Activities, The University of Georgia, Athens, Ga.*

K. O. BOWMAN—*Oak Ridge National Laboratories, Oak Ridge, Tenn.*

ABSTRACT—The two-parameter gamma distribution is determined by its shape parameter γ and scale parameter, β. Maximum likelihood estimators of these $\hat{\gamma}/\gamma$, $\hat{\beta}/\beta$ are distributed independent of β and tend to be normally distributed as the sample size n increases. Values of n are given (for selected values of γ) to ensure near normality.

Mooley (1973) raises some interesting questions on when a variate can be described as being normally distributed. He considers the size of the sample n {the sample being assumed to consist of independent identically distributed gamma variates with density $f(x)=(x/\beta)^{\gamma-1}[\exp(-x/\beta)]/[\beta\Gamma(\gamma)]$ for $\gamma>0$, $\beta>0$} such that the maximum likelihood estimators $\hat{\beta}$ and $\hat{\gamma}$ of β and γ respectively shall have an approximate normal distribution. There are several approaches to this, one being to use the fact that, as n becomes large, $\hat{\beta}$ and $\hat{\gamma}$ tend to become normal. Near normality can be described in terms of the skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) moment ratios, and we might say that approximate normality is achieved when $-\epsilon_1<\sqrt{\beta_1}<\epsilon_1$ and $-\epsilon_2<\beta_2-3<\epsilon_2$ where ϵ_1 and ϵ_2 are small and positive. Actually, $\sqrt{\beta_1}$ and β_2 may be functionally related; thus limiting β_2 to an interval automatically controls the range of $\sqrt{\beta_1}$. Here,

TABLE 1.—*Sample size to achieve approximate normality for the distribution of $\hat{\gamma}$, $\hat{\beta}$*

$\beta_2(\hat{\gamma})$	γ	n	$\sqrt{\beta_1(\hat{\gamma})}$	$\sqrt{\beta_1(\hat{\beta})}$	$\beta_2(\hat{\beta})$
3.450	0.1	87	0.484	0.789	4.002
4.000	.3	54	.712	.704	3.825
4.000	0.5	60	.714	.582	3.564
4.000	1.0	66	.719	.466	3.361
4.000	1.5	68	.716	.432	3.309
4.000	2.0	69	.718	.401	3.263
4.000	3.0	70	.720	.376	3.225
4.000	5.0	71	.721	.359	3.201
4.000	10.0	71	.721	.349	3.185
3.334	0.1	115	.417	.684	3.750
3.750	.3	69	.618	.620	3.640
3.750	.5	77	.618	.512	3.437
3.750	1.0	85	.623	.410	3.279
3.750	1.5	87	.622	.380	3.240
3.750	2.0	89	.624	.354	3.204
3.750	3.0	90	.625	.331	3.175
3.750	5.0	90	.626	.317	3.156
3.750	10.0	91	.626	.307	3.144
3.219	0.1	172	.338	.557	3.500
3.500	.3	99	.506	.516	3.441
3.500	.5	110	.508	.427	3.303
3.500	1.0	116	.503	.375	3.237
3.500	1.5	125	.509	.316	3.166
3.500	2.0	128	.511	.294	3.141
3.500	3.0	130	.512	.275	3.121
3.500	3.5	130	.513	.271	3.116
3.500	4.0	130	.513	.268	3.112
3.500	5.0	131	.513	.263	3.108
3.500	10.0	131	.513	.255	3.099
3.108	0.1	339	.239	.395	3.250
3.250	.3	187	.361	.373	3.231
3.250	.5	207	.363	.310	3.160
3.250	1.0	220	.357	.270	3.123
3.250	1.5	240	.361	.228	3.086
3.250	2.0	245	.363	.212	3.073
3.250	3.0	249	.363	.198	3.063
3.250	5.0	250	.364	.190	3.056
3.250	10.0	251	.364	.184	3.052
3.043	0.1	842	.150	.250	3.100
3.100	.3	454	.228	.239	3.095
3.100	.5	506	.229	.198	3.065
3.100	1.0	534	.226	.173	3.050
3.100	1.5	582	.229	.146	3.035
3.100	2.0	596	.230	.135	3.030
3.100	3.0	605	.230	.127	3.026
3.100	5.0	609	.231	.121	3.023
3.100	10.0	611	.231	.118	3.021
3.004	0.1	8275	.048	.080	3.010
3.010	.3	4368	.073	.077	3.010
3.010	.5	4880	.073	.064	3.007
3.010	1.0	5230	.072	.055	3.005
3.010	1.5	5720	.073	.046	3.004
3.010	2.0	5863	.073	.043	3.003
3.010	5.0	5992	.073	.039	3.002
3.010	10.0	6006	.073	.037	3.002

we set ϵ_2 such that $\beta_2 \leq 3.5$, 3.25, 3.1, and 3.01 and find how large n must be to ensure these results with respect to the larger of the kurtosis for the distribution of $\hat{\beta}$ and $\hat{\gamma}$. The assessments in table 1 have been evaluated from the asymptotic moments of $\hat{\beta}$ and $\hat{\gamma}$ to order n^{-6} given in Bowman and Shenton (1970) and Shenton and Bowman (1972); in our notation, $\beta=\alpha$, $\gamma=\rho$.

Note that, in our tabulation, $B_2(\hat{\beta})<B_2(\hat{\gamma})$ except when $0<\hat{\gamma}<0.3$. Moreover, accepting a sample size of $n=75$ to induce normality (Mooley gives 75 yr as the needed record to make monthly rainfall gamma parameter estimators nearly normal) is equivalent to accepting $\max\{\beta_2(\hat{\beta}), \beta_2(\hat{\gamma})\}$ in the region of 3.75 to 4.00. We point out that the upper 5-percent and 1-percent usual normal deviates 1.6449 and 2.326 would become 1.621 and 2.472, respectively, if we use a Pearson type VII curve with $\beta_1=0$ and $\beta_2=4.0$ instead of the normal; similarly, for type VII with $\beta_1=0$ and $\beta_2=3.5$, the upper 5-percent and 1-percent points are 1.633 and 2.416, respectively. Thus Pearson's type VII with $\beta_2=4.0$ at the 5-percent level is midway between that for the normal and type VII with $\beta_2=4.0$; but the 1 percent for type VII with $\beta_2=3.5$ is nearer to type VII with $\beta_2=4.0$ than to the normal.

ACKNOWLEDGMENTS

This work was supported partially by the Southeastern Forest Experiment Station under Department of Defense Contract No. IPR19-2-8006 and was sponsored by the U.S. Atomic Energy Commission under contract with the Union Carbide Corporation.

REFERENCES

Bowman, Kimiko O., and Shenton, L. R., "Small Sample Properties of Estimators for the Gamma Distribution," *Report* CTC-28, Union Carbide Corporation, Oak Ridge, Tenn., **1970**, 160 pp.

Mooley, Diwakar, "An Estimate of the Distribution and Stability Period of the Parameters of the Gamma Probability Model Applied to Monthly Rainfall Over Southeast Asia During the Summer Monsoon," *Monthly Weather Review*, Vol. 101, No. 12, Dec. **1973**, pp. 884–890.

Shenton, L. R., and Bowman, Kimiko O., "Further Remarks on Maximum Likelihood Estimators for the Gamma Distribution," *Technometrics*, Vol. 14, No. 3, Aug. **1972**, pp. 725–733.

[*Received September 25, 1972*]