Moments of Order Statistics From the Inverse Gaussian Distribution

William W. S. Chen, Internal Revenue Service

Order statistics often play a central role in optimal statistical inference procedures. In current literature, there are many papers with tabulations of expected values, variances, and covariances of order statistics with samples drawn from the Gaussian Distribution. However, none deals with the similar problem when a sample is drawn from the Inverse Gaussian (IG) Distribution. In both theory and application, the IG Distribution has received more attention in recent years.

M.Y. Chan, A.C. Cohen, and B.J. Whitten (1983) gave a three-parameter IG distribution with zero mean and unit variance and tabulated the cumulative probability function as a function of the standardized variate, z, and of the shape parameter, k. In this study, we adopt their standardized IG cumulative distribution function, and with the same selected shape parameter 0.0(0.1)2.5, we tabulate the mean, variance, and covariance of order statistics up to sample size 25. Using these tabulated values, we can establish the best linear unbiased estimates of the location and scale parameters. Finally, the variances and covariances of these parameters are tabulated.

■ Introduction

The computation of theoretical means, variances, covariances, and correlation of order statistics in small samples from a Gaussian Distribution dates back as early as Jones (1948) and Godwin (1949). Later, Teichroew (1956) extended sample sizes to 20 and gave 10-decimal-place accuracy. Yamauti (1972) provided 8-decimal-place tables of product moments for sample sizes of up to 30. G.L. Tietjen, D.K. Kahaner, and R.J. Beckman (1977) have computed variances and covariances of the normal order statistics for sample sizes up to 50. Most recently, R.S. Parrish (1992) applied a Gauss-Legendre Quadrature technique and gave 25 decimal places of the means of normal order statistics for sample sizes up to 500. He also constructed a table of variances and covariances of normal order statistics of 25 decimal places for sample sizes up to 20, 15 decimal places for sample sizes up to 40, and 10 decimal places for sample sizes up to 50. From the results listed above, we have no doubt that the Gaussian Distribution is the most commonly used distribution, and these moments or product moments are needed.

Recently, the Inverse Gaussian Distribution has received increased attention. At first, this distribution had been used to answer questions relating to the physical phenomenon of Brownian motion. Hadwiger (1940) applied the Inverse Gaussian density function as the solution to a functional equation in his study of a reproduction function associated with a biological population. Later, this curve was referred to as the Hadwiger fertility curve by Hoem and Berge (1975). Tweedie noted the inverse relationship between its cumulant generating functions and those of the Gaussian Distribution and coined the name "Inverse Gaussian Distribution." This distribution has also been studied extensively by numerous investigators. A detailed list is in reference [2]. A noteworthy investigator is V. Seshadri (1993) (see reference [14] covering both theoretical and applied work in this area of study). In reference [2] (1983), the authors compare the I.G. distribution with the most commonly used lifespan distributions, such as the exponential, the lognormal,the gamma, and the Weibull distribution. It has been suggested that the I.G. distribution can be a useful alternative. They derived a standardized probability density function with zero mean, unit variance, and shape parameter, k. The cumulative distribution function as a function of standardized variate and k as the shape parameter were then tabulated. That approach is the foundation of this study, producing the means, variances, and covariances of the order statistics of this important distribution. However, only 5 decimal places of accuracy are presented. We would expect that the table will be widely used throughout the industry for quality control. By making use of the means, variances, and covariances of order statistics tabulated in the tables, we can also compute the coefficients for the best linear unbiased estimators of the location and scale parameters of Inverse Gaussian Distribution and variances and covariances of these estimators in cases of complete or type II right censored samples.

Basic Formulas

M.Y. Chan, A.C. Cohen, and B.J.Whitten (1983) gave a three-parameter version of the I.G. distribution with parameters μ, σ, γ, where γ is the origin, $\mu + \gamma$ is the mean, and σ is the standard deviation. They then derived a standardized distribution with mean zero, unit variance, and shape parameter k, where k is the third standard moment. Thus, $z=(x-\mu-\gamma)/\sigma$, and the probability density function becomes

$$f(z;0,1,k) = \frac{1}{\sqrt{2\pi}} \left\{ \frac{3}{3+kz} \right\}^{3/2} \exp\left\{ -\frac{3z^2}{2(3+kz)} \right\} \tag{2.1}$$

where $-3/k < z < \infty$, zero elsewhere. From (2.1), they can derive the cumulative distribution function of the standardized I.G.distribution. It subsequently follows that

$$F(z;0,1,k) = \Phi\left\{ \frac{z}{\sqrt{1+kz/3}} \right\} + e^{18/k^2} \Phi\left\{ \frac{-(z+6/k)}{\sqrt{1+kz/3}} \right\} \tag{2.2}$$

where $\Phi(\,)$ is the cumulative standardized normal distribution. As we can see from (2.2), they have expressed the cumulative distribution function of the standardized I.G. distribution as the sum of two cumulative standardized normal components. It is not difficult to show from basic analysis that as $k \to 0$, the cumulative distribution function of (2.2) will approach the standardized normal distribution. Using the density function of (2.1) and cumulative distribution function of (2.2), we can start to construct our single and product moments of the Inverse Gaussian Distribution. The general fundamental definitions of these moments are given in reference [1]. For example, our definition of expected values of order statistics i for given sample sizes n is formula (5.2.1) on page 108, the definition of variance is formula(5.2.2) on page 108, and the definition of covariance is formula (5.2.4) on page 108. Clearly, theoretical integration of these formulas without the aid of a computer is intractable. Thus, a numerical method is necessary. Three different numerical methods have been applied. The first method used power series approximation, the second method applied Gauss-Legendre quadrature technique by using 96 lattice points, while the last approach is the same as a previous one but increased to 512 lattice points. The mathematical analysis in error bounds using this method has been discussed in detail in reference[12], section 2.1 or 2.2. It has been determined that the maximum error will not exceed order of 10^{-30}. Hence, only the final approximated results are presented in our tables.

Numerical Computation

All computations were performed in double precision by Fortran 77 on the HP 9000/755 model computer system. The single integrals of equations were evaluated by using the 512-point Gaussian quadrature formula over different lengths of intervals. Except in the normal case, i.e., shape parameter k=0, the lower bound of the integral is easy to determine by computing -3/k. For example, for the shape parameter k=0.1, we can use -30 as our lower bound. If the shape parameter yields 1.5, then we use -2.0 as our lower bound. All other cases can easily be determined. However, there are no rules for us to determine the upper limits for a given shape parameter, k. By tabulating (2.2), as in the work by the three authors (1983), we determine z values such that the cumulative functions are equal to 1. As expected, when the shape parameter increases, the distribution has gradually turned out to be a very long, thin, upper tail. For example,when k=2.5, the upper limit could reach 55, covering the whole distribution range. However, we must be very careful about the range of double integral due to the fact that it is a dependent variable. In other words, the lower limit of x_j is x_i while the lower limit of x_i is -3/k, depending on k. The upper limit of x_j will be dependent upon tabulation and k. If we ignore the dependency of the two variables, it could cause some negative association as i is very close to j. This is not the case in our selected shape parameters. When this problem was studied, we used the power-series approach first. This only led to two-digit accuracy when compared with the mean values of three/four digits in the covariance approximation. Hence, we switched to the 96 Gauss-Legendre quadrature formula. Improvement was dramatic. In normal cases, the means or covariances coincided with R.S. Parrish's published results. When sample sizes grow large, there is a huge combinatoric integer value to be multiplied by the result of the double integration. Usually, the 96 points formula can lead to 13-digit accuracy, but if, say, seven-digit integer numbers are multiplied, then only six-digit accuracy is achieved. The second cause of inaccuracy is the increasing shape parameter as the distribution gradually evolves to be a long, thin, upper tail. However, 96 lattice points are evenly distributed over the integral domain, causing more area or volume loss in the lower tails than in the upper tails. It is clear that we need more points in the lower tails to compensate for this loss. One of the best ways to make up for this defect is by increasing the number of lattice points. We are pleased with this final approach. During the computation of the mean, variance, and covariance, we have written one short calling program with eight subfunctions or subroutines. In the main program, we only read in the preassigned exponential values of n,i,j and pass these values to subfunctions to compute each possible combination of the power of probability density function or cumulative functions. All the computations use double precision formats to ensure that at least the first ten digits are accurate. With this accuracy of the tabulated mean, variances, and covariances as our basis, the best linear unbiased estimation of location and scale parameters for both completed or type II right censored samples could reach the desired accuracies.

■ Verification

We used at least fifteen different identities or recurrence relations to check the table values and found that they all matched. Some of the identities related to product moments or second-order moments about the origin. Since they are not given in the table, we have not listed them below. Equation (4.1) gives the easiest and quickest check of the mean values. It simply says that for a given sample size, in our case the sum of the mean values of the order statistics, the value is always zero. As we glance at the table of mean values, we can see that this statement is correct. Cole (1951) developed a recurrence relation of expected values of (4.2), and we randomly selected four different shape parameters and sample sizes to verify the table mean values. Equation (4.3) is the most powerful identity to use to check the interrelation between the mean values and covariances.

Similarly, using five randomly selected shape parameters, sample sizes, and randomly selected index i and j, we can easily see that the N. Balakrishnan (1989) identity has been matched. It is fun to use these formulas and to check our tabulated values. We encourage readers to try some. Listed below are the identities used:

$$\sum_{i=1}^{n} \mu_{i:n} = n\mu_{1:1} , \qquad (4.1)$$

For $1 \le i \le n-1$

$$i\mu_{i+1:n} + (n-i)\mu_{i:n} = n\mu_{i:n-1} , \qquad (4.2)$$

For $2 \le i < j \le n$ (4.3)

$$(i-1)\sigma_{i,j:n} + (j-i)\sigma_{i-1,j:n} + (n-j+1)\sigma_{i-1,j-1:n}$$
$$= n\left(\sigma_{i-1,j-1:n-1} + (\mu_{i-1:n-1} - \mu_{i-1:n})(\mu_{j-1:n-1} - \mu_{j:n})\right) ,$$

Equation Used	Skewed Parameter	Sample Sizes	Index Used	Both sides of Equation equal
(4.2)	1.2	15	i = 8	-1.21918
(4.2)	1.5	15	i = 8	-1.75030
(4.2)	2.0	20	i = 10	-4.92265
(4.2)	2.5	25	i = 20	15.33447
(4.3)	0.2	19	i=5 j=6	1.46562
(4.3)	0.6	20	i=7 j=8	1.17180
(4.3)	1.2	10	i=5 j=7	0.87064
(4.3)	1.5	15	i=10 j=13	1.45909
(4.3)	2.1	20	i=8 j=14	0.52172

■ Acknowledgments

The author appreciates the assistance of J. Harte of IRS who helped to edit the manuscript.

■ References

Arnold, B.C.; Balakrishnan, N.; and Nagaraja, H.N. (1992): *A First Course in Order Statistics*, John Wiley and Sons, Inc.

Chan, M.Y.; Cohen, A.C.; and Whitten, B.J. (1983): "The standardized inverse gaussian distribution table of the cumulative probability function," *Communication In Statistics, Simulation, and Computation*, pp. 423-442.

Chen, W.W.S. (1982): "Simulation on probability points for testing of Lognormal or Weibull Distribution with a small sample," *Journal of Statistical Computation and Simulation*, Vol. 15, pp. 201-210.

Chhikara, R.S. and Folks, J.L. (1974): "Estimation of the Inverse Gaussian Distribution Function," *Journal of the American Statistical Association*, 69, pp. 250-54.

Chhikara, R.S. and Folks, J.L. (1977): "The Inverse Gaussian Distribution as a Lifetime Model," *Technometrics*, Vol. 19, No. 4, pp. 461-468.

Cole, R.H. (1951): "Relations Between Moments of Order Statistics," *Annals of Mathematical Statistics*, 22, pp. 308-310.

Godwin, H.J. (1949): "Some Low Moments of Order Statistics," *Annals of Mathematical Statistics*, 20, pp. 279-285.

Harter, H.L. (1961): "Expected Values of Normal Order Statistics," *Biometrika*, 48, pp. 151-165.

Jones, H.L. (1948): "Exact Lower Moments of Order Statistics in Small Samples from a Normal Distribution," *Annals of Mathematical Statistics*, 19, pp. 270-273.

Parrish, R.S (1992): "Computing Expected Values of Normal Order Statistics," *Communication In Statistics, Simulation and Computation*, pp. 57-70.

Parrish, R.S. (1992): "Computing Variances and Covariances of Normal Order Statistics," *Communication In Statistics, Simulation, and Computation*, pp.71-101.

Seshadri, V. (1993): *The Inverse Gaussian Distribution, A Case Study in Exponential Families*, Clarendon Press, Oxford Science Publications.

Stroud, A.H. and D. Secrest (1969): *Gaussian Integration Formulas*, Prentice-Hall, Englewood Cliffs, NJ.

Teichroew, D. (1956): "Table of Expected Values of Order Statistics and Products of Order Statistics for Samples of Size Twenty and Less from the Normal Distribution," *Annals of Mathematical Statistics*, 27, pp. 410-426.

Tietjen, G.L.; D.K. Kahaner; and R.J. Beckman (1977): "Variances and Covariances of the Normal Order Statistics for Sample Sizes 2 to 50," Selected Tables in *Mathematical Statistics V*, pp. 1-73.

Tweedie, M.C.K. (1957): "Statistical Properties of the Inverse Gaussian Distribution II," *Annals of Mathematical Statistics*, 28, pp. 695-705.

Expected Values of Inverse Gaussian Order Statistics
Shape Parameter k

n	i	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
1	1	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
2	1	-.56419	-.56399	-.56341	-.56244	-.56110	-.55940	-.55735	-.55498
2	2	.56419	.56399	.56341	.56244	.56110	.55940	.55735	.55498
3	1	-.84628	-.84140	-.83595	-.82996	-.82346	-.81649	-.80908	-.80129
3	2	.00000	-.00918	-.01833	-.02741	-.03638	-.04522	-.05389	-.06238
3	3	.84628	.85058	.85428	.85736	.85984	.86171	.86298	.86367
4	1	-1.02938	-1.01985	-1.00969	-.99894	-.98764	-.97585	-.96363	-.95105
4	2	-.29701	-.30604	-.31472	-.32302	-.33093	-.33840	-.34544	-.35202
4	3	.29701	.28767	.27806	.26821	.25816	.24797	.23765	.22727
4	4	1.02938	1.03822	1.04635	1.05375	1.06040	1.06629	1.07142	1.07580
5	1	-1.16296	-1.14928	-1.13493	-1.11998	-1.10450	-1.08856	-1.07223	-1.05558
5	2	-.49502	-.50214	-.50873	-.51474	-.52017	-.52501	-.52926	-.53291
5	3	.00000	-.01188	-.02371	-.03545	-.04706	-.05849	-.06971	-.08068
5	4	.49502	.48737	.47924	.47065	.46165	.45227	.44256	.43256
5	5	1.16296	1.17593	1.18813	1.19952	1.21008	1.21979	1.22863	1.23661

Variances and covariances of Inverse Gaussian Order
Statistics Shape Parameter k

n	i	j	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
1	1	1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1	1	.68169	.65372	.62628	.59947	.57335	.54801	.52348	.49981
2	2	2	.68169	.71010	.73886	.76785	.79698	.82614	.85524	.88417
3	1	1	.55947	.52555	.49293	.46169	.43187	.40353	.37668	.35132
3	2	2	.44867	.44833	.44732	.44564	.44332	.44038	.43686	.43279
3	3	3	.55947	.59459	.63082	.66802	.70608	.74485	.78421	.82402
4	1	1	.49172	.45581	.42172	.38949	.35915	.33069	.30411	.27936
4	2	2	.36046	.35261	.34431	.33562	.32660	.31729	.30776	.29807
4	3	3	.36046	.36781	.37463	.38088	.38653	.39156	.39596	.39972
4	4	4	.49172	.52936	.56864	.60946	.65170	.69520	.73984	.78545
5	1	1	.44753	.41089	.37641	.34412	.31401	.28607	.26023	.23642
5	2	2	.31152	.30050	.28928	.27794	.26654	.25513	.24379	.23257
5	3	3	.28683	.28655	.28571	.28431	.28238	.27994	.27701	.27362
5	4	4	.31152	.32228	.33273	.34281	.35245	.36163	.37030	.37842
5	5	5	.44753	.48630	.52711	.56988	.61447	.66078	.70864	.75791
2	1	2	.31831	.31809	.31743	.31634	.31483	.31293	.31064	.30801
3	1	2	.27566	.26776	.25963	.25132	.24289	.23439	.22586	.21734
3	1	3	.16487	.16471	.16423	.16343	.16234	.16096	.15930	.15740
3	2	3	.27566	.28330	.29061	.29757	.30413	.31027	.31597	.32119
4	1	2	.24559	.23463	.22370	.21287	.20219	.19171	.18147	.17152
4	1	3	.15801	.15417	.15010	.14584	.14143	.13689	.13225	.12756
4	1	4	.10468	.10456	.10420	.10361	.10279	.10176	.10053	.09911
4	2	3	.23594	.23572	.23503	.23390	.23234	.23037	.22800	.22527
4	2	4	.15801	.16160	.16492	.16793	.17064	.17301	.17505	.17675
4	3	4	.24559	.25653	.26739	.27811	.28863	.29890	.30887	.31850
5	1	2	.22433	.21186	.19963	.18770	.17612	.16494	.15418	.14389
5	1	3	.14815	.14247	.13670	.13087	.12503	.11921	.11344	.10776
5	1	4	.10577	.10344	.10092	.09823	.09541	.09247	.08944	.08634
5	1	5	.07422	.07412	.07384	.07337	.07272	.07191	.07094	.06984
5	2	3	.20844	.20474	.20070	.19634	.19168	.18677	.18164	.17634
5	2	4	.14994	.14978	.14929	.14848	.14736	.14594	.14424	.14229
5	2	5	.10577	.10790	.10981	.11148	.11290	.11407	.11499	.11565
5	3	4	.20844	.21175	.21466	.21716	.21923	.22087	.22207	.22285
5	3	5	.14815	.15369	.15907	.16425	.16919	.17387	.17826	.18235
5	4	5	.22433	.23699	.24977	.26260	.27543	.28818	.30080	.31323