ESTIMATION OF PARAMETERS OF PEARSON'S TYPE III POPULATION FROM GENERALISED CENSORED SAMPLES

P. V. K. Iyer and Naunihal Singh

Defence Science Laboratory, Delhi

ABSTRACT

This paper deals with the problem of estimating the parameters of Pearson's Type III population from generalised[1] censored samples. Maximum likelihood method is employed to arrive at the estimates. The ML equations are not readily solvable. This has been facilitated by providing a few graphs and tables. Standard errors (s.e.) of the estimates have been obtained by the method of scoring. A numerical example is discussed to illustrate the practical application of the method.

INTRODUCTION

The problem of estimation of parameters from censored samples has been considered by Hald[2], Cohen[3], Gupta[4], Des Raj[5] and others for various types of distributions. Unlike the problem considered by these authors which involve censoring at two stages only, we are often confronted with situations wherein censoring occurs at multistages owing to the presence of a number of factors beyond control in the observations recorded. To cite an illustration, suppose, we are interested in the study of life of radio tubes or general stores where the life is a random variable. Let N radio tubes be selected randomly from a lot and put to test. Out of this suppose k items fail by the end of time t_k. During this period, if $n_1, n_2 .. n_l$ tubes are either damaged or lost at different intervals, say, $t'_1, t'_2 \ldots t'_l$, it means that no information is available as to when these tubes fail. The only information available is that they have survived upto the time $t'_1, t'_2 \ldots\ldots t'_l$. Similarly we may have to face another situation wherein, life of tyres or likewise other major parts of vehicles under field operations is under study and some of the vehicles are either transferred to other units or damaged due to war operations during the period of experimentation. Such type of data may be treated as a sample censored at various stages. Bartholomew[6] has discussed this type of problem for exponential distribution and Iyer and Singh[1] for normal distribution. Sampford and Taylor[7] have considered a similar problem in randomized block experiments.

In this paper the method of maximum likelihood is employed to arrive at the estimating equations which by successive approximations would yield the estimates of the parameters. To solve ML equations, the sample values are used as first approximations for m and σ. But for α_3 the sample estimate would be biased especially when proportions censored are large. This would necessitate comparatively greater number of iterations for the final solution. To get over this difficulty a graphical procedure to estimate α_3 to a satisfactory degree of accuracy for samples conforming to Type III Distribution has been evolved. Again the likelihood equation for α_3 involves a factor $\frac{\partial}{\partial \alpha_3} \log F(\eta)$ which makes the solution intractable. This has been simplified by providing a table at the appendix for various values of α_3 and η. Standard errors of the estimates have been obtained by the method of scoring.

2. Maximum likelihood solutions

(2·a) *Explanation of notations and symbols.*

(i) Number of items issued N

(ii) Period of observation t_k

(iii) Number of items failed by the time t_k k

(iv) Life of the k items that fail $t_1, t_2, \ldots, t_k$

(v) Number of items missing (censored) during interval t_k before failure n

(vi) Time points at which n_j items are missing t'_j

(vii) Number of failures observed during the interaval t'_j : k_j $j = 1, 2, \ldots, l$

(viii) Number of items remaining upto time t_k N-k-n.

(ix) $n = \sum_{j=1}^{l} n_j$

(x) $\nu_1 = \frac{1}{k} \sum_{j=1}^{k} (t_j - t_k)$, first moment about t_k

(xi) $\theta = \frac{1}{k} \sum_{j=1}^{l} n_j A_j$

(xii) $\Phi = \frac{1}{k} \sum_{j=1}^{l} n_j A_j \eta'_j$

(xiii) $p_T = \int_{-2/a_3}^{\eta_T} f(\eta) d\eta$

(xiv) $p_j = \int_{-2/a_3}^{\eta_j} f(\eta) d\eta \qquad \eta = \frac{t-m}{\sigma}$

(xv) $S = \sum_{j=1}^{k} \left[1 + \frac{a_3}{2} \frac{t_j - m}{\sigma} \right]^{-1}$

where

$$A_j = \frac{f(\eta'_j)}{F(\eta'_j)}, \quad \eta'_j = \frac{t'_j - m}{\sigma}, \quad \eta_T = \frac{t_k - m}{\sigma}$$

$$f(\eta) = c \left[1 + \frac{a_3}{2} \eta \right]^{4/a_3^2 - 1} e^{-\frac{2}{a_3}\eta}$$

$$-\frac{2}{a_3} \leqslant \eta < \infty$$

$$c = \left(\frac{4}{a_3^2} \right)^{4/a_3^2 - \frac{1}{2}} e^{-4/a_3^2} \left[\Gamma(4/a_3^2) \right]^{-1}$$

and $$F(\eta'_j) = \int_{\eta'_j}^{\infty} f(\eta) d\eta$$

(2.b) *Estimating Equations.*

The likelihood function for a sample arising from a situation as discussed in section (1) may be written as

$$L = \text{const.} \prod_{j=1}^{k} f(\eta_j) \prod_{j=1}^{l} [F(\eta_j')]^{N_j} [F(\eta_T)]^{N-k-n} \tag{2.b.1}$$

$$F(\eta_T) = \int_{\eta_T}^{\infty} f(\eta) d\eta$$

Equating to zero the first partial derivatives of the logarithm of the (2.b.1) with respect to m, σ and α_3, we have

$$\frac{2k}{a_3} - \left(\frac{4}{a_3^2} - 1\right)\frac{a_3}{2} S + k\theta + (N-k-n)A_T = 0 \tag{2.b.2}$$

$$\frac{2}{a_3} \sum_{j=1}^{k} \frac{t_j - m}{\sigma} - \frac{4k}{a_3^2} + \left(\frac{4}{a_3^2} - 1\right) S + k\phi + (N-k-n)A_T\eta_T = 0, \tag{2.b.3}$$

$$k\left\{-\frac{8}{a_3^3} \log 4/a_3^2 + \frac{1}{a_3} - \frac{1}{\Gamma(4/a_3^2)} \frac{\partial}{\partial a_3} \Gamma(4/a_3^2)\right\} + \frac{2}{a_3^2} \sum_{j=1}^{k} \frac{t_j - m}{\sigma}$$

$$+ \left(\frac{4}{a_3^2} - 1\right)\frac{1}{a_3}\left(k - S\right) + \sum_{j=1}^{l} \frac{n_j}{F(\eta'_j)} \frac{\partial}{\partial a_3} F(\eta'_j)$$

$$- \frac{8}{a_3^3} \sum_{j=1}^{k} \log\left(1 + \frac{a_3}{2} \frac{t_j - m}{\sigma}\right) + \frac{N-k-n}{F(\eta_T)} \frac{\partial}{\partial a_3} F(\eta_T) = 0 \tag{2.b.4}$$

Substituting the value of S from (2.b.2) in (2.b.3) and (2.b.4), we get

$$\nu_1 = -\sigma\left[\eta_T + \theta + \frac{a_3}{2}\phi + \frac{N-k-n}{k} A_T \left(1 + \frac{a_3}{2}\eta_T\right)\right] \tag{2.b.5}$$

and

$$-\frac{8k}{a_3^3}\left\{\log\frac{4}{a_3^2}-\log\left(\frac{4}{a_3^2}-1\right)+\frac{1}{k}\sum_{j=1}^{k}\log\left(1+\frac{a_3}{2}\frac{t_j-m}{\sigma}\right)\right\}$$

$$+\frac{k}{a_3(1-a_3^2/4)}+\frac{4}{a_3^2}\sum_{j=1}^{k}\frac{t_j-m}{\sigma}$$

$$+\frac{1}{a_3}\left\{k\phi+a_3\sum_{j=1}^{l}\frac{n_j}{F(\eta'_j)}\frac{\partial}{\partial a_3}F(\eta'_j)\right\}$$

$$+\frac{N-k-n}{a_3}\left\{\eta_T A_T+\frac{a_3}{F(\eta_T)}\frac{\partial}{\partial a_3}F(\eta_T)\right\}=0 \qquad \textbf{(2.b.6)}$$

The equation (2.b.5) can be re-written as

$$\hat{m}=\bar{t}+\hat{\sigma}\left\{\theta+\frac{a_3}{2}\phi+\frac{N-k-n}{k}A_T\left(1+\frac{a_3}{2}\eta_T\right)\right\}. \qquad (2.b.7)$$

Putting $\alpha_3 = a_3 + \delta$ in (2.b.6), a_3 being first estimate of a_3 and δ the correction factor, the linear estimate of δ is

$$\delta=\frac{D+A/a_3^2+B/a_3+c(a_3^2/4+1)}{2A/a_3^3+B/a_3^2-ca_3/2} \quad \ldots \quad \ldots \quad \ldots \quad (2.b.8)$$

where

$$A=-8k\left\{\log 4/a_3^2-\log\left(\frac{4}{a_3^2}-1\right)\right.$$

$$\left.+\frac{1}{k}\sum_{j=1}^{k}\log\left(1+\frac{a_3}{2}\frac{t_j-m}{\sigma}\right)\right\}$$

$$B=4\sum_{j=1}^{k}\frac{t_j-m}{\sigma}$$

$$c=k$$

$$D=k\phi+a_3\sum_{j=1}^{l}n_j\frac{\partial}{\partial a_3}\log F(\eta'_j)$$

$$+(N-k-n)\left\{\eta_T A_T+a_3\frac{\partial}{\partial a_3}\log F(\eta_T)\right\}$$

$\hat{m}$, $\hat{\sigma}$ + and $\hat{a}_3$ are the estimates of the parameters.

As indicated in section (1), the main difficulty in solving the above equations is to obtain an approximate and reasonable estimate of α_3. This has been over come by constructing a few graphs (Figs. 1, 2) showing the relation between α_3 and $\lambda = \dfrac{z_1 - z_2}{z_1 - z_3}$ from

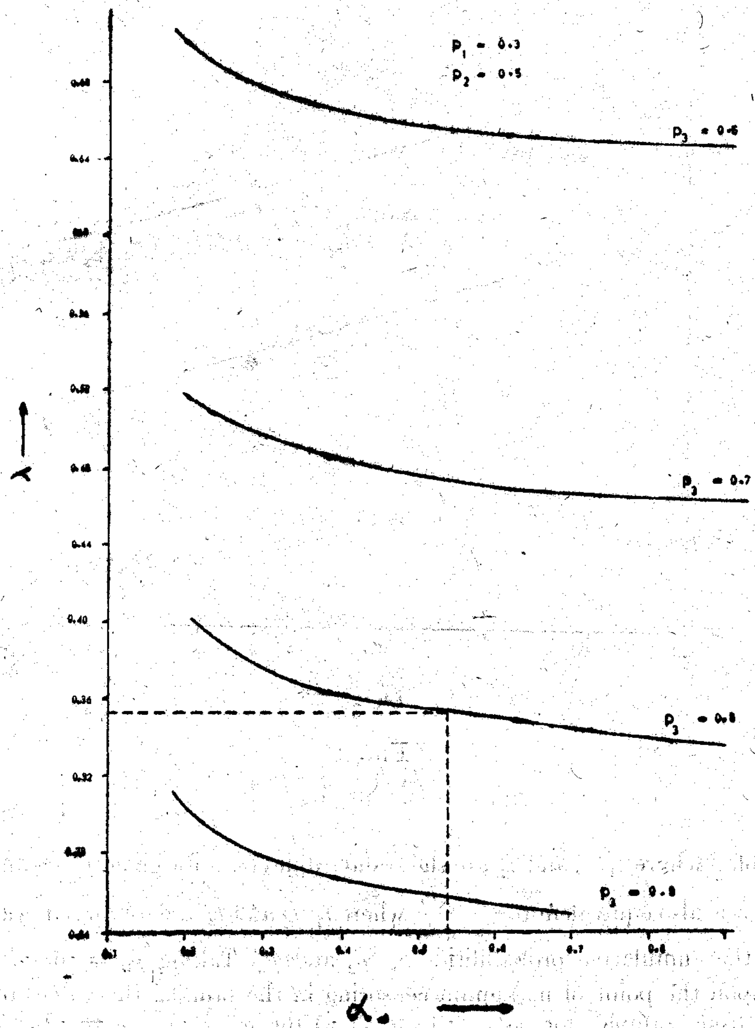

FIG. 1

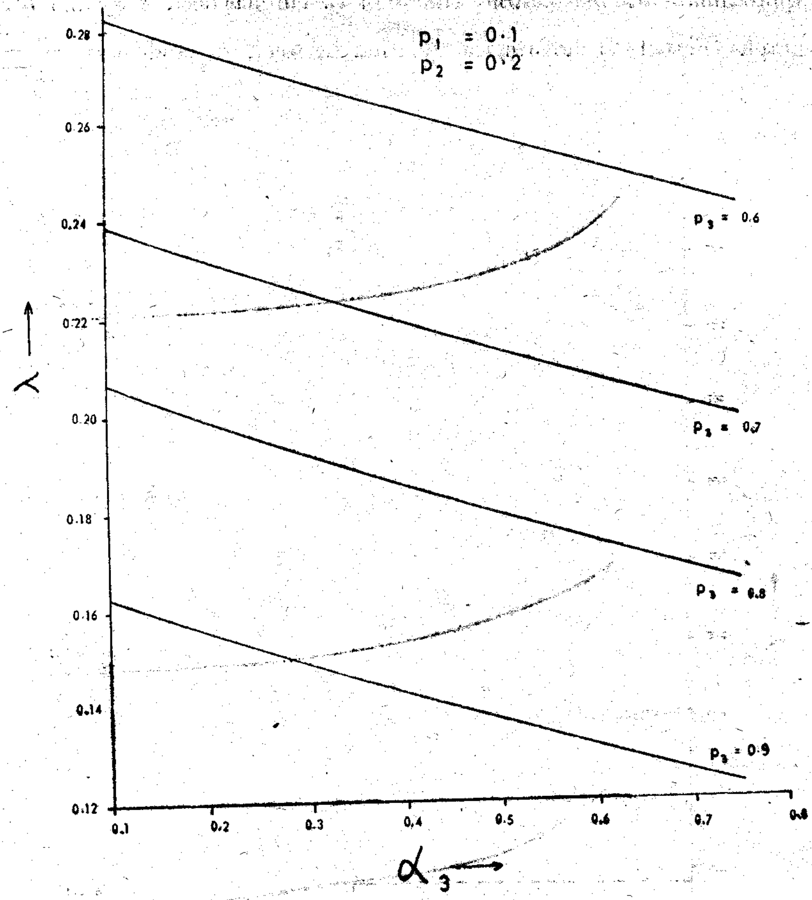

FIG. 2

Salvosa's tables[8] where z_1, z_2 and z_3 are the standard deviates for given p_1, p_2 and p_3. It can be seen that λ is also equivalent to $\frac{t_1 - t_2}{t_1 - t_3}$ where t_1, t_2 and t_3 are observed values corresponding to the cumulative probabilities p_1, p_2 and p_3. Taking $p_3 = \cdot 6$, $\cdot 7$, $\cdot 8$ and $\cdot 9$ depending upon the point of maximum censoring in the sample, the values of λ can be determined from sample for $p_1 = \cdot 3$, $p_2 = \cdot 5$ (or $p_1 = \cdot 1$, $p_2 = \cdot 2$). Using this λ, α_3 can be read from the graph corresponding to the set of values of p_1, p_2 and p_3 actually taken in the calculations. After knowing α_3, σ and m are evaluated from (2.b.5) and (2.b.7) respectively where first estimates of η_j's are obtained from p_j's by using Salvosa's tables and taking $p_j = k_j \Big/ \left(N - \sum_{r=1}^{j} n_r\right)$. δ is then calculated from (2.b.8) to improve the first estimate of α_3 with the aid of the table appended. After this the cycle repeats and the ML estimates of parameters are obtained by successive approximations. The actual procedure is illustrated in the example discussed in section (4).

PRECISION OF ESTIMATES

To obtain asymptotic variances and covariances of the estimates, the symmetric information matrix I $(\hat{m}, \hat{\sigma}, \hat{a}_3)$ is constructed. The elements of this matrix are calculated by scoring process of Fisher[9]. On inversion it will provide the s.e's of the estimates. $\hat{m}$, $\hat{\sigma}$ and $\hat{a}_3$ are taken in order.

$$I_{(\hat{m}, \hat{\sigma}, \hat{a}_3)} = \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{12} & I_{22} & I_{23} \\ I_{13} & I_{23} & I_{33} \end{bmatrix} \qquad (3\cdot1)$$

where

$$I_{11} = \sum_{i=1}^{h} \frac{k}{P_i}\left(\frac{dP_i}{\partial m}\right)^2 + \sum_{j=1}^{l} \frac{n_j}{[F(\eta'_j)]^2}\left(\frac{\partial F(\eta'_j)}{\partial m}\right)^2 + \frac{E(N-k-n)}{[F(\eta_T)]^2}\left(\frac{\partial F(\eta_T)}{\partial m}\right)^2 \quad .. \quad .. \quad .. \quad (3\cdot2)$$

$$I_{22} = \sum_{i=1}^{h} \frac{k}{P_i}\left(\frac{\partial P_i}{\partial \sigma}\right)^2 + \sum_{j=1}^{l} \frac{n_j}{[F(\eta'_j)]^2}\left(\frac{\partial F(\eta'_j)}{\partial \sigma}\right)^2 + \frac{E(N-k-n)}{[F(\eta_T)]^2}\left(\frac{\partial F(\eta_T)}{\partial \sigma}\right)^2 \quad .. \quad .. \quad .. \quad (3\cdot3)$$

$$I_{33} = \sum_{i=1}^{h} \frac{k}{P_i}\left(\frac{\partial P_i}{\partial a_3}\right)^2 + \sum_{j=1}^{l} \frac{n_j}{[F(\eta'_j)]^2}\left(\frac{\partial F(\eta'_j)}{\partial a_3}\right)^2 + \frac{E(N-k-n)}{[F(\eta_T)]^2}\left(\frac{\partial F(\eta_T)}{\partial a_3}\right)^2 \quad .. \quad .. \quad .. \quad (3\cdot4)$$

$$I_{12} = \sum_{i=1}^{h} \frac{k}{P_i}\left(\frac{\partial P_s}{\partial m} \cdot \frac{\partial P_i}{\partial \sigma}\right) + \sum_{j=1}^{l} \frac{n_j}{[F(\eta'_j)]^2}\left(\frac{\partial F(\eta'_j)}{\partial m} \cdot \frac{\partial F(\eta'_j)}{\partial \sigma}\right) + \frac{E(N-k-n)}{[F(\eta_T)]^2}\left(\frac{\partial F(\eta_T)}{\partial m} \cdot \frac{\partial F(\eta_T)}{\partial \sigma}\right) \quad .. \quad .. \quad (3\cdot5)$$

$$I_{13} = \sum_{i=1}^{h} \frac{k}{P_i}\left(\frac{\partial P_i}{\partial m} \cdot \frac{\partial P_i}{\partial a_3}\right) + \sum_{j=1}^{l} \frac{n_j}{[F(\eta'_j)]^2}\left(\frac{\partial F(\eta'_j)}{\partial a_3} \cdot \frac{\partial F(\eta'_j)}{\partial m}\right) + \frac{E(N-k-n)}{[F(\eta_T)]^2}\left(\frac{\partial F(\eta_T)}{\partial m} \cdot \frac{\partial F(\eta_T)}{\partial a_3}\right). \quad .. \quad .. \quad (3\cdot6)$$

$$I_{23}=\sum_{i=1}^{h}\frac{k}{P_i}\left(\frac{\partial P_i}{\partial\sigma}\cdot\frac{\partial P_i}{\partial\sigma}\right)+\sum_{j=1}^{l}\frac{n_j}{[F(\eta'_i)]^2}\left(\frac{\partial F(\eta'_j)}{\partial\sigma}\cdot\frac{\partial F(\eta'_j)}{\partial a_3}\right)$$

$$+\frac{E(N-k-n)}{[F(\eta_T)]^2}\left\{\frac{\partial F(\eta_T)}{\partial\sigma}\cdot\frac{\partial F(\eta_T)}{\partial a_3}\right\} \quad \ldots \quad \ldots \quad \ldots \quad (3\cdot7)$$

where $P_i = \int_{\eta_i}^{\eta_i+1} f(\eta)d\eta,\ \frac{\partial P_i}{\partial m} = \frac{1}{\sigma}\left(f(\eta_i)-f(\eta_i+1)\right),$

$$\frac{\partial P_i}{\partial\sigma}=\frac{1}{\sigma}\left\{\eta_i\, f(\eta_i)-\eta_{i+1}f(\eta_i+1)\right\}$$

$E(n_j) = n$, since n_j's are known. $E(N-k-n)=N(1-p_T)-\sum_{j=1}^{l} n_j\,(p_T-p_j)$ $-\sum_{j=1}^{l} n_j$, h is the number of groups into which the data have been classified. $\frac{\partial P_i}{\partial a_3}$ is noted from the table (appended) as detailed in introduction.

NUMERICAL EXAMPLE

A random sample of size one hundred was taken by using Salvosa's tables with $m=10$, $\sigma=1$ and $a_3=\cdot5$. Let the number of observations n_1 and n_2 not available at $t'_1=8\cdot8169$, $t'_2=9\cdot7674$ be 5 in each of the cases. Taking $t_k = 11\cdot0173$ and $k=80$, the sample yields the following:

* $\bar{t}=9\cdot75204$, $s^2=0\cdot53318$

$\nu_1=-1\cdot26526$ $n=10$

$p_1=\cdot3,\ p_2=\cdot5,\ p_3=\cdot8$

$t_1=9\cdot3868,\ t_2=9\cdot7674,\ t_3=10\cdot4573$ and $\lambda=0\cdot355$,

therefore the graph (Fig. 1) corresponding to $0\cdot8$ gives

$$a_3=0\cdot54$$

and equations (2.b.5) and (2.b.7) give

$\sigma=0\cdot82332,$ $\hat{m}=9\cdot88038.$

With the help of these estimates, we get from (2.b.8)

$\hat{\delta}=0\cdot1600$

therefore corrected $a_3=0\cdot74$

Starting with this set of $\hat{m}$, $\hat{\sigma}$ and a_3, three more sets of approximations are obtained by iteration and shown in Table I. Values of equations (2.b.5), (2.b.7) and (2.b.8) are also given for corresponding sets of approximations which indicate the nature of the convergence of the method.

TABLE I. ITERATIVE ESTIMATION OF PARAMETERS AND VALUES OF EQUATIONS (2.b.5), (2.b.7) AND (2.b.8)

$\hat{a}$	$\hat{m}$	$\hat{a}_3$	δ	(2.b.5)	(2.b.7)	(2.b.8)
82332	9·88038	·540	·1600	·15424	·12773	·01705
74844	9·98445	·700	·2040	·01104	·00827	·01387
75282	9·97709	·740	·2050	·00069	·00184	·00442
75245	9·97964	·745	·2004	·00000	·00011	·00190

* $\bar{t}$ is the sample mean and s^2 the sample variance.

Variances and covariances of the estimates are also determined from (3.1.) and given below:

$\text{Var}(\hat{m}) = 0\cdot009397$, $\text{Var}(\hat{\sigma}) = 0\cdot008905$, $\text{Var}(L_3) = 0\cdot118292$

$\text{Cov}(\hat{m}\hat{\sigma}) = -0\cdot005169$, $\text{Cov}(\hat{m}\hat{a}_3) = 0\cdot007066$, $\text{Cov}(\hat{\sigma}, \hat{a}_3) = -\cdot0\cdot020919$

Acknowledgement—Our sincere thanks are due to Shri P. Samarasimhudu for technical assistance.

REFERENCES

1. IYER, P.V.K., and SINGH, N, *Ind. Soc. Agri. Statist.* **XIV**, (1962).
2. HALD, A., *Skand Aktuairtdsker*, **119**. (1949).
3. COHEN, A. C., *Amer. Statist. Asso.*, **45**, 411 (1950).
4. GUPTA, A. K., *Biometrika*, **39**, 260 (1952).
5. DES RAJ, *Amer. Statist. Asso.*, **48**, 336 (1953).
6. BARTHOLOMEW, D. J., *J. Amer. Statist. Asso.*, **52**, 350 (1957).
7. SAMPFORD, M. R. and TAYLOR. J., *Royal. Stat. Soc.* B **21**, 1, 214 (1959).
8. SALVOSA, L. R., *Ann. Math. Statist. I*, appended, (1930).
9. FISHER, R. A. and YATES, F., "Statistical Tables", XIV, Oliver and Boyd, London, 14, (1949).

Appendix

Table Showing the Values of $\alpha_3 \frac{\partial}{\partial \alpha_3} \log F(\eta)$ for Different Values of α_3 and η

η / α_3	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
−2·00	+0·00294	+0·00602	+0·00908	+0·01187	+0·01393	+0·01461	+0·01295	+0·00812	+0·00170	+0·00000
−1·75	+0·00348	+0·00748	+0·01196	+0·01685	+0·02196	+0·02688	+0·03087	+0·03248	+0·02906	+0·01652
−1·50	+0·00342	+0·00762	+0·01269	+0·01874	+0·02586	+0·03410	+0·04345	+0·05367	+0·06411	+0·07287
−1·25	+0·00245	+0·00571	+0·00991	+0·01525	+0·02192	+0·03019	+0·04038	+0·05289	+0·06817	+0·08679
−1·00	+0·00039	+0·00138	+0·00308	+0·00564	+0·00921	+0·01402	+0·02031	+0·02847	+0·03889	+0·05219
−0·75	−0·00269	−0·00511	−0·00725	−0·00903	−0·01037	−0·01120	−0·01135	−0·01074	−0·00915	−0·00636
−0·50	−0·00643	−0·01296	−0·01959	−0·02632	−0·03314	−0·04002	−0·04696	−0·05391	−0·06084	−0·06770
−0·25	−0·01028	−0·02087	−0·03182	−0·04312	−0·05480	−0·06684	−0·07928	−0·09211	−0·10534	−0·11895
0·00	−0·01348	−0·02738	−0·04158	−0·05622	−0·07129	−0·08680	−0·10275	−0·11918	−0·13603	−0·15334
0·25	−0·01517	−0·03066	−0·04648	−0·06266	−0·07922	−0·09614	−0·11349	−0·13123	−0·14940	−0·16795
0·50	−0·01442	−0·02920	−0·04436	−0·05989	−0·07583	−0·09218	−0·10896	−0·12612	−0·14376	−0·16180
0·75	−0·01028	−0·02137	−0·03321	−0·04577	−0·05906	−0·07296	−0·08757	−0·10278	−0·11862	−0·13500
1·00	−0·00176	−0·00561	−0·01127	−0·01854	−0·02725	−0·03724	−0·04846	−0·06072	−0·07397	−0·08816
1·25	+0·01215	+0·01959	+0·02316	+0·02336	+0·02080	+0·01592	+0·00889	+0·00014	+0·01021	+0·02208
1·50	+0·03257	+0·05572	+0·07148	+0·08126	+0·08623	+0·08714	+0·08486	+0·07969	+0·07201	+0·06250
1·75	+0·06067	+0·10438	+0·13540	+0·15643	+0·16991	+0·17722	+0·17945	+0·17770	+0·17246	+0·16468
2·00	+0·09787	+0·17038	+0·21639	+0·25032	+0·27282	+0·28630	+0·29283	+0·29416	+0·29069	+0·28386