Far East J.Theo.Stat. 23(2)(2007),133-164

THE BETA DISTRIBUTION, MOMENT METHOD, KARL PEARSON AND R.A. FISHER

K.O. BOWMAN[1] and L.R. SHENTON[2]

[1]Computational Sciences and Engineering Division
Oak Ridge National Laboratory, P.O.Box 2008, 4500N, MS-6191
Oak Ridge, TN 37831-6191, U.S.A.
e-mail: bowmanko@ornl.gov
[2]Department of Statistics, University of Georgia
Athens, Georgia 30602, U.S.A.

October 19, 2007

*Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC-05-00OR22725 with the U.S. Department of Energy.

Abstract

Simulation studies provide four moment approximating distributions to each of the four parameters of a beta distribution (Pearson Type I). Two of the parameters refer to origin and scale, two to shape (skewness and kurtosis). Type I random number generator is checked out, and the stability of moments of random samples of size n over cycles; particular attention is paid to shape parameter moments. In Type I region of validity (referred to skewness and kurtosis), moment methods become unstable in the neighborhood of Type III (χ^2) line, and ultimately abort. Thus extremely large variances and large higher moments arise. We probe the cause of this phenomenon. Simulation studies are turned to since alternative power series methods are forbiddingly complicated. However, use is made of the delta method to provide asymptotic variances of the estimators, and asymptotic variances of percentage points of the basic distribution. An account of work on the subject by K. Pearson, some of it a century ago, is given. In particular an important paper by Pearson and Filon provides some estimates of probable errors of moment parameter estimators such as the basic distribution parameters, the mode, the skewness and others. The heated controversy between Pearson and Fisher is considered.

2000 Mathematics Subject Classification: 62E20.

Keywords: Beta density; Delta method; Kurtosis variance; Moments; Monte Carlo simulation; Percentage points; Skewness variance; Variance of percentage points.

1 Introduction

Estimating the four parameters of the density

$$B(x;a,b;\,p,q) = c(x-a)^{p-1}(b-x)^{q-1} \qquad (a<x<b;\ p,q>0)$$

is perhaps the most complicated of all three main Pearson types which include I or Beta, IV, and VI. The regions of validity are shown in Figure 1, along with density types.

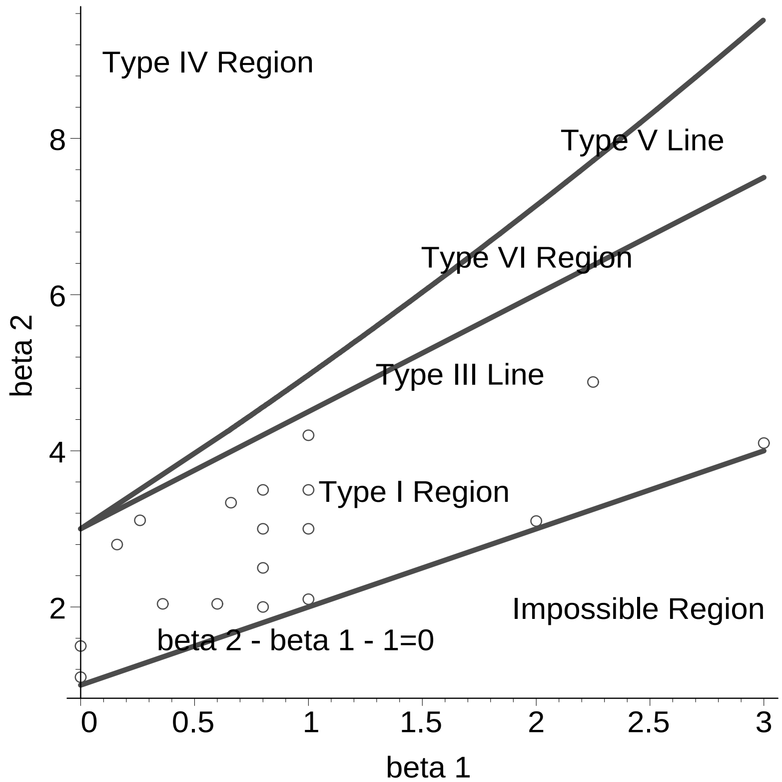

Figure 1: Pearson type distributions over β plane

Type III line is $2\beta_2 - 3\beta_1 - 6 = 0$. Circles represent the cases studied.

We consider moment estimators $(a^*, b^*;\ p^*, q^*)$ for the four parameters $(a, b;\ p, q)$. In particular

$$p^* = \frac{r^*}{2}\left\{1 - \frac{(r^*+2)\sqrt{b_1}}{\sqrt{D^*}}\right\}, \qquad (\sqrt{b_1} = m_3/m_2^{3/2}, b_2 = m_4/m_2^2, m_s = \sum(x_i-\bar{x})^s/n)$$

$$q^* = \frac{r^*}{2}\left\{1 + \frac{(r^*+2)\sqrt{b_1}}{\sqrt{D^*}}\right\},$$

$$b^* - a^* = \sqrt{m_2}\sqrt{D^*}/2,$$

$$a^* = m_1' - (b^* - a^*)p^*/(r^*),$$

$$r^* = 6(b_2 - b_1 - 1)/(6 + 3b_1 - 2b_2)$$

$$D^* = (r^* + 2)^2 b_1 + 16(r^* + 1).$$

(When asterisks are excluded, estimators take their population values; for example $\sqrt{b_1} \to \sqrt{\beta_1}$, $p^* \to p$, etc.)

It will be seen that the estimators all involve the second, third, and forth central moments m_2, m_3, m_4 along with the mean m_1'. Thus to study the moments $\mu_s(t^*)$, $(s = 1, 2, 3, 4)$, t^* being one of the estimating parameters, in the form

$$\mu_s(t^*) \sim T_0 + T_1/n + T_2/n^2 + \cdots,$$

n being the sample size, would involve four dimensional Taylor series in the deviates

$$\xi_s = m_s' - \mu_s' \qquad (s = 1, 2, 3, 4)$$

where here we revert to non-central moments. The reason for this is that using expectations, through independence

$$E\exp\{\alpha_1\xi_1 + \alpha_2\xi_2 + \cdots\} = \left[E\left\{\exp\left(\frac{d_1(x - \mu_1')}{n} + \frac{d_2(x^2 - \mu_2')}{n} + \cdots\right)\right\}\right]^n,$$

whereas this breaks down for

$$E\{\exp[\alpha_1(m_1 - Em_1) + \alpha_2(m_2 - Em_2) + \cdots]\}.$$

Similarly, although p^* and q^* are functions of $\sqrt{b_1}$, b_2 only, there is no ready answer to the problem of expanding

$$E\{\exp[\alpha_1(\sqrt{b_1} - E\sqrt{b_1}) + \alpha_2(b_2 - Eb_2)]\}.$$

We point out that the problem is much simpler when the end-points a, b are known and taken to be $a = 0$, $b = 1$. Here, p and q have simple moment estimators based on the mean and variance, and Taylor series developments are readily set up (Bowman and Shenton, 1992). The maximum likelihood approach is given by Lau and Lau (1991).

In summary, so far, it is clear that a mathematical analysis using Taylor series in descending power of n, is out of reach even allowing for approaches assisted by mathematical language packages (Derive, Maxima, Maple, etc). The reader may glance at first order terms in the moments of the parameters given in the appendix.

The aim being to throw some light on the response of **the moment estimators of the parameters to the many types of samples thrown up in random sampling**

from specified basic Beta forms, we are compelled to resort, in the main to simulation studies.

There is a historical component in our study. About, a century ago Karl Pearson (1857-1936) in 1894 and 1895, introduced his system of curves in connection with the mathematical theory of evolution. At the time the normal distribution had no rivals, so that faced with skewness in natural measurements (for examples (1894) (i) measurements of 998 specimens (pawns) from penultimate to hindmost tooth on the carapace, (ii) variation in the number of Mullerian glands in the fore-legs of 2000 swine), Pearson resorted to a mixture of two normal components to allow for skewness contamination. When this approach failed he considered many components as a whole producing Bell-shaped data histograms conceptualized as stemming from a simple differential equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{(x+a)}{Ax^2+Bx+c}. \qquad (y \text{ being a density})$$

In the solutions, a discriminant arises defining the main types of curves and transition types.

Three years later, Pearson and Filon (1898) produced a long paper "On the probable errors of frequency constants and on the influence of random selection on Variation and Correlation." The term probable error is no longer in use, but refers to a deviate on the normal distribution corresponding to 25%; including the negative deviate, 50% is entailed. In terms of normality then it may be interpreted as 0.6745σ. This study of Pearson and Filon is quite difficult to understand; not surprisingly it has for the most part been "ignored" or overlooked. It appears to be deriving first order terms in maximum likelihood estimation somewhat earlier than the studies of R. A. Fisher (1890-1962). Fisher and Pearson had a difference of opinion in the approach to estimation, in particular relating to moments and maximum likelihood (m.l.) in the case of the Beta distribution (referred to as an example in the sequel). There was a quibble about numerical accuracy (8 or so decimal digits were used in the computations), and both parties were insecure when asymptotics were involved; here asymptotic efficiency related to the notion that m.l. estimators had less variance than other estimators. However this problem was only lightly touched upon. To gain a flavor of the conflict, Pearson (1936) remarks

"What is the practical value of showing that the distribution of a statistical parameter may be **approximately** represented by a normal curve when $n > 40,000$ when you are seeking what happens when $n = 400$?" He continues, asking about a statement such as (in our notation)

$$E(\text{standard deviation}) \sim \sigma_0 + \sigma_1/\mathrm{n} + \sigma_2/\mathrm{n}^2 + \cdots \qquad (\mathrm{n} \to \infty)$$

If truncated to two terms, how large must n be for acceptance? With this background, we consider

(i) validity of the random number generator for Beta distributions.

(ii) For various sample sizes ($n = 100-5000$) the cycle length needed to produce consistent (not in the statistical sense) estimates of the first four moments of the 4 parameters a, b, p, q.

(iii) Having satisfied (ii), we tabulate several cases of moments for the moment estimators, including two of special interest since they spring from papers of Pearson-Filon (1898) and Pearson (1902). Let it be noted that Pearson (1936) refers to the Koshal-Fisher paper, which in fact does not exist. It happened partly through correspondence that Pearson's attention became focused on Koshal's (1933) paper in JRSS, the subject being the improvement of estimation by moments by the method of maximum likelihood. This apparently acted as a lighting rod to Pearson who discovered that the study owed a great deal to Fisher. Thus Pearson referred to the Koshal paper as the Koshal-Fisher paper.

(iv) Checks on the simulated values of the moments of $c^* = 6 + 3b_1 - 2b_2$, $\sqrt{\beta_1}$, and β_2, using Taylor series expansions and Padè approximation.

(v) Dominant asymptotics for moments using the delta method.

2 Notation and Formulas with Comments

First of all, it is convenient at this point to briefly define the fundamental entities used in the study.

Type I density:

(1) End points 0,1:

$$B(x; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1} \qquad (0 \le x \le 1; p, q > 0)$$

(2) End points $a, b; b > a$, and shape parameters p, q:

$$B(x; a, b, p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(x-a)^{p-1}(b-x)^{q-1}}{(b-a)^{p+q-1}} \qquad (a \le x \le b; p, q > 0)$$

Moments for the general case: (Johnson and Kotz, 1970)

Mean:

$$\mu_1' = a + (b-a)p/(p+q)$$

Variance σ^2:

$$\mu_2 = (b-a)^2 pq/[(p+q)^2(p+q+1)]$$

Skewness: $\sqrt{\beta}_1$;

$$\frac{\mu_3}{\mu_2^{3/2}} = \frac{2(q-p)\sqrt{p^{-1}+q^{-1}+(pq)^{-1}}}{p+q+2} = \alpha_3$$

Kurtosis β_2;

$$\frac{\mu_4}{\mu_2^2} = \frac{3(p+q+1)[2(p+q)^2 + pq(p+q-6)]}{pq(p+q+2)(p+q+3)} = \alpha_4$$

Parameters:

$$p, q = \frac{r}{2}\left\{1 \mp (r+2)\frac{\sqrt{\beta_1}}{\sqrt{D}}\right\}$$

$$b - a = \sqrt{\mu_2}\sqrt{D}$$

$$a = \mu_1' - (b-a)p/(p+q)$$

where

$$r = \frac{6(\beta_2 - \beta_1 - 1)}{6 + 3\beta_1 - 2\beta_2}, \qquad D = (r+2)^2\beta_1 + 16(r+1).$$

Region of validity:

$$\begin{cases} \beta_2 - \beta_1 - 1 > 0 & (\text{Boundary L}_1), \\ 2\beta_2 - 3\beta_1 - 6 < 0 & (\text{Boundary L}_2). \end{cases}$$

Approximations (used in the text)

Symmetry: $\sqrt{\beta_1} = 0$

$$p = q = r/2 = 3(\beta_2 - 1)/(6 - 2\beta_2) \qquad (1 < \beta_2 < 3).$$

Asymmetry:

(1) Neighborhood of L_1

$$p, q \sim \frac{r}{2}\left\{1 \mp \frac{2\sqrt{\beta_1}}{\sqrt{4\beta_1 + 16}}\right\}.$$

(2) Neighborhood of L_2 and $\sqrt{\beta_1} > 0$.

$$\begin{cases} p \sim 4r(r+1)/[\beta_1(r+2)^2] & (r \to \infty), \\ q \sim r. \end{cases}$$

Comments:

(1) In the vicinity of L_1, the distributions are U-shaped, the skewness depending on $p - q$.

(2) In the region "midway" between L_1 and L_2 the distributions are J or inverted J-shaped (interchanging p and q if $\sqrt{\beta_1} < 0$).

(3) In the vicinity of L_2, the distributions are bell shaped (single mode).

Comments on the distribution

The Type I Pearson distribution in the general case involves four parameters (2 terminals, 2 indices) and is valid over a 2-dimensional infinite space. Infinite varieties of structure are therefore possible. Proximity of the skewness and kurtosis to transition regions (Type III and Type VI), is an important driving force in the approximation procedure. For example with $\sqrt{\beta_1} = 1$ (Table 1), q increases as the line L_2 is approached.

Table 1. Response of Parameters p and q to nearness to L_2 Line

$\beta_1 = 1$

β_2	4.00	4.20	4.30	4.40	4.45	4.49	4.50
p	1.82	2.41	2.81	3.32	3.63	3.92	4.00
q	10.18	19.59	31.69	58.68	143.37	743.08	∞

Series and Continued Fractions (c.f.)

We include this subject since rational functions are frequently better algorithms to turn to in the face of divergence tendencies.

If $f(x) = c_1 x + c_2 x^2 + \cdots$, then the c.f. is

$$\frac{a_1 x}{1+}\ \frac{a_2 x}{1+}\ \cdots = \frac{a_1}{1/x+}\ \frac{a_2}{1+}\ \frac{a_3}{1/x+}\ \frac{a_4}{1+}\cdots$$

where

$$a_{2s} = \psi_{s+1}\phi_{s-1}/\psi_s\phi_s, \quad a_{2s+1} = -\phi_{s+1}\psi_s/\psi_{s+1}\phi_s \quad (s = 1, 2, \cdots)$$

and $a_1 = \phi_1$, $\phi_0 = 1$, $\psi_1 = 1$, $\phi_s = |c_1, c_3, \cdots, c_{2s-1}|$, $\psi_s = |c_2, c_4, \cdots, c_{2s-2}|$, here diagonal notation is used for a determinant.

Stieltjes (1918, 425-429) showed that for a power series

$$\frac{c_0}{n} - \frac{c_1}{n^2} + \frac{c_2}{n^3} - \frac{c_3}{n^4} + \cdots$$

there is the c.f.

$$\frac{a_1}{n+}\ \frac{b_1/a_1}{1+}\ \frac{a_2/b_1}{n+}\ \frac{b_2/a_2}{1+}\ \frac{a_3/b_2}{n+}\ \frac{b_3/a_3}{1+}\ \cdots \qquad \text{if valid}$$

where $a_s = A_s/A_{s-1}$ and $b_s = B_s/B_{s-1}$, and in diagonal notation

$$A_s = |c_0, c_2, \cdots, c_{2s}|, \qquad B_s = |c_1, c_3, \cdots, c_{2s+1}|,$$

$$A_0 = B_0 = 1 \qquad\qquad (A_1 = c_0, B_1 = c_1).$$

For example, $A_2 = c_0 c_2 - c_1^2$.

If the first coefficient in the series is changed to $c_0 + \mu$, then the modified determinants become

$$\begin{cases} A_s^* = A_s + \mu C_{s-1} & \text{if } C_s = |c_2, c_4, \cdots, c_{2s}|, \\ B_s^* = B_s & (C_0 = 1). \end{cases}$$

An example of this is the c.f. for $E(2b_2 - 3b_1 - 6)$ in contrast to that for $E(2b_2 - 3b_1)$. Also see for example, Bowman and Shenton (1989), Wall (1948), Jones and Thron (1990) for further properties.

3 Random Number Generator for the Beta Distribution

3.1 Basics

First of all we carry out a check on the random number (r.n.) generator package (G05FEF, NAG Fortran Library Routine). Runs of over 10^6 r.n. are shown (Table 2).

Table 2. Validity of r.n. Generator for Type I Distribution

N		β_1	β_2	p	q
10^6	G	0	1.1	0.0789	0.0789
	E	0.0008	1.1007		
2.5×10^6	G	0.0	1.5	0.5000	0.5000
	E	0.0973	1.4720		
10^6	G	0.36	2.04	0.4680	0.8920
	E	0.4128	2.0421		
10^6	G	1.0	2.1	0.0345	0.0905
	E	0.9991	2.0979		
5×10^6	G	1.0	4.2	2.4075	19.5925
	E	1.0007	4.2035		
10^6	G	2.0	3.1	0.0218	0.0816
	E	2.0002	3.1000		
10^6	G	3.0	4.1	0.0152	0.0730
	E	3.0064	4.1058		

Comment: G = given population value, E = r.n.generator results. On the whole quite satisfactory. However the case $\beta_1 = 0$, $\beta_2 = 1.5$ compared to $\beta_1 = 0$, $\beta_2 = 1.1$ is rather surprising. Mean and Variance are taken to be zero and unity respectively.

It is to be expected that when complicated functions are simulated, approximants will deteriorate somewhat; for example, p^*, q^*, a^*, and b^* are all anything but simple structures in terms of the basic statistics, m_1', m_2, $\sqrt{b_1}$ and b_2.

In view of our interest in the historical use of the Type I distribution, we give (Table 3) simulations of two data sets. (i) A set given by Pearson (1895); we direct attention to the results for the estimators b^* and q^*. (ii) A set given by Pearson and Filon (1898).

Table 3. Two Examples

(a) $\mu_1' = 3.9832, \mu_2 = 5.5599, \beta_1 = 0.662858, \beta_2 = 3.3362804$

$a = 0.277847, b = 16.931730, p = 1.69751993, q = 5.93198896,$ Pearson (1936)

n		$\mu_1'^*$	σ^*	$\sqrt{\beta_1^*}$	β_2^*	cycles
500	a^*	0.2949	0.2641	-0.9195	5.2222	9998
	b^*	17.8270	14.7234	53.7288	3870.0172	
	p^*	1.7022	0.4160	1.7652	9.6632	
	q^*	6.7555	12.7718	54.8813	3930.9467	
750	a^*	0.2898	0.2156	-0.6796	3.9952	7500
	b^*	17.2982	3.8956	5.6047	86.1983	
	p^*	1.7005	0.3304	1.2400	6.0619	
	q^*	6.2270	3.0012	6.4776	104.1592	
1000	a^*	0.2871	0.1858	-0.5937	4.0990	5000
	b^*	17.1748	3.0288	5.0712	90.9073	
	p^*	1.6993	0.2811	1.1501	6.6421	
	q^*	6.1628	2.3231	7.1879	148.9588	
2000	a^*	0.2870	0.1279	-0.3177	3.1289	5000
	b^*	16.9878	1.7760	1.0588	5.3589	
	p^*	1.6918	0.1881	0.6239	3.7720	
	q^*	5.9915	1.2741	1.2977	6.3947	

(b) $\sqrt{\beta_1} = 0.5091, \beta_2 = 3.1108, c = 6 + 3\beta_1 - 2\beta_2 = 0.5559$

$a = -0.8182, b = 17.2264, p = 4.7837, q = 15.2014$, Pearson and Filon (1898)

n		$\mu_1'^*$	σ^*	$\sqrt{\beta_1}^*$	β_2^*	cycles	25%	75%
1000	a^*	-0.8060	0.5384	-1.1788	5.2433	4865	-1.0882	-0.4140
	b^*	27.0080	2.7646	196.1832	5419.4640			
	p^*	4.9345	1.8952	2.0742	9.7576		3.5918	5.6915
	q^*	35.7296	440.6835	57.2115	3609.2558			
	c^*	0.6023	0.2477	-0.1803	2.5314		0.4285	0.7859
	$\sqrt{\beta_1}^*$	0.5049	.0728	0.1047	2.9502		0.4547	0.5536
	β_2^*	3.0892	0.1970	0.2878	2.7678		2.9453	3.2215
2000	a^*	-0.8236	0.4017	-1.0447	4.8757	4978	-1.0418	-0.5339
	b^*	20.7770	52.7646	46.5267	2565.8943			
	p^*	4.9131	1.3896	1.8888	8.9112		3.9250	5.4913
	q^*	22.3520	104.8210	45.5275	2475.9034			
	c^*	0.5752	0.1907	-0.2607	2.8574		0.4483	0.7119
	$\sqrt{\beta_1}^*$	0.5078	0.0525	0.1127	2.9738		0.4717	0.5428
	β_2^*	3.1033	0.1491	0.2954	2.8512		2.9955	3.2023

Comment: In both cases, the higher moment parameters $\sqrt{\beta_1}$ and β_2 are extremely fragile. However the mean and standard deviation are fairly reliable. Out of 5000 cycles, 135 samples had c values outside the Type I region for $n = 1000$, and 22 samples for $n = 2000$.

(c) n^{-1} terms of Covariances

	a^*	b^*	p^*	q^*	$\sigma(n = 2000)$
a^*	307.0093	-2824.3282	-915.6832	-4949.4605	0.3918
b^*	-0.8032	40271.7541	9650.1311	65369.7048	4.4873
p^*	-0.9735	0.8958	2881.8598	16617.3212	1.2004
q^*	-0.8601	0.9919	0.9425	107856.1934	7.3436

Values in the lower half triangle are correlations.

3.2 Stability and Closeness of $c = 6 + 3\beta_1 - 2\beta_2$ to Zero

Simulations present problems because the Beta distribution responds differently to (a) $0 < p < 1$, $0 < q < 1$. (b) $0 < p < 1$, $q > 1$ or $0 < q < 1$, $p > 1$, and (c) $p > 1$, $q > 1$, when seeking its inversion. In addition sampling from the Beta in the neighborhood of the gamma line $2\beta_2 = 3\beta_1 + 6$, is dangerously near to chaos. For in this case the moment parameter $r^* \to \infty$ responding to the denominator $6 + 3\beta_1 - 2\beta_2$.

Two examples of simulating the moments of q^* (Table 4) bring out clearly a contrast in results due to the different values of $c = 6+3\beta_1-2\beta_2$. As c (or c^*) becomes smaller sampling is likely to pick up values of $c^* > 0$ but small (if $c^* < 0$, the run is excluded), so that if $\sqrt{b_1} > 0$ then q^* is r^* approximately, and r^* has denominator c^* (see the approximation in section [2]).

Table 4. Variability Over Cycles

(a) $\beta_1 = 0.8, \beta_2 = 3.0, a = 0, b = 1, p = 0.7601, q = 2.2319, c = 2.4, n = 50$

Cycle	$E(a^*)$	$\sigma(a^*)$	$\sqrt{\beta}_1(a^*)$	$\beta_2(a^*)$
2000	0.00024	0.0087	-0.2538	3.0064
4000	0.00026	0.0086	-0.2606	3.0827
6000	0.00029	0.0086	-0.2447	3.0686
8000	0.00033	0.0086	-0.2574	3.0675
10000	0.00028	0.0086	-0.2545	3.0704

(b) $\sqrt{\beta}_1 = 1.0, \beta_2 = 4.2, a = 0, b = 1, p = 2.4075, q = 19.5925, c = 0.6, n = 1000$

Cycle	$E(q^*)$	$\sigma(q^*)$	$\sqrt{\beta}_1(q^*)$	$\beta_2(q^*)$
913	174.3	3932	30.1	907
1826	116.5	2831	40.6	1693
2734	87.0	2314	497	2534
3632	74.6	2010	57.1	3352
4542	74.4	1866	52.4	3628

Cycle	$E(p^*)$	$\sigma(p^*)$	$\sqrt{\beta}_1(p^*)$	$\beta_2(p^*)$
913	2.2503	0.4965	0.7804	3.4104
1826	2.2512	0.5060	0.9131	3.9460
2734	2.2456	0.5099	0.8872	3.8601
3632	2.2394	0.5048	0.8921	3.9828
4542	2.2396	0.5058	0.9081	3.9791

Out of 5000 cycles, 458 samples gave c values outside the Type I region.

Karl Pearson did observe that one or other of p^*, q^* could be excessively large (in the context of estimation of parameters), but suggested, rather vaguely that this characteristic might not have much effect on the "shape" of the distribution. We shall return to this point later (section [6]).

Note that there is a non-negligible region for which c^* is small (or negative) in which case q^* will be large. Contrast this aspect with Type III sampling for which the "region" degenerates to a line, and c^* is no longer involved in estimation.

3.3 Checks on the Simulations using $c^* = 6 + 3b_1 - 2b_2$

Internal checks on simulations, such as repeated cycle consistency, are not completely reliable. An alternative is to develop Taylor series, provided the fundamental deviates involved, in expectation can be evaluated. It would appear to be obvious to attempt Taylor series for p^*, and q^* in terms of $b_1 - Eb_1$ and $b_2 - Eb_2$. However the joint moments of the deviates are excessively complicated since they involve moment parameters as ratios. The estimators a^* and b^* in terms of 4-dimensional Taylor series are also not attractive.

So we turn to c^* which we studied (Bowman and Shenton, 1975) in Taylor series developments, purely from interest in the basic algorithm used. The report considers four moments of $\sqrt{b_1}$, b_2, and c^*, mainly from Pearson distributions. At that time, the series involved terms to n^{-8} for $\sqrt{b_1}$, and terms to n^{-6} for b_2 and c^*. Using a criteria for usage by direct summation, "safe sample sizes" are included. For the Pearson system the study is confined to the Type I region including the Type III line; asymptotics tend to become fragile tools in the Type IV and Type VI regions. If the sample size involved in any particular case is within the "safe" value, direct summation is acceptable; otherwise resort to rational fractions (ratios of polynomials in n, the sample size) in the form

$$\frac{nq_0}{n+} \frac{p_1}{1+} \frac{q_1}{n+} \frac{p_2}{1+} \frac{q_2}{n+} \cdots$$

for the mean, with obvious modifications for higher moments (for example, drop the initial n component for the variance). Examples are shown in Table 5.

See the notation section for remarks on continued fractions.

Table 5. Moments of c^* by Simulation and Taylor Series

(a) $n = 100, \sqrt{\beta_1} = 0.6, \beta_2 = 2.04, p = 0.4680, q = 0.8920, c = 3$

	$E(c^*)$	$\sigma(c^*)$	$\sqrt{\beta_1}(c^*)$	$\beta_2(c^*)$	cycle
Simulation	3.1714	0.2498	-0.1506	3.4075	10000
Series	3.0109	0.2488	-0.0329	3.3691	

(b) $n = 100, \sqrt{\beta_1} = 0, \beta_2 = 1.5, p = q = 0.5, c = 3$

	$E(c^*)$	$\sigma(c^*)$	$\sqrt{\beta_1}(c^*)$	$\beta_2(c^*)$	cycle
Simulation	3.0968	0.1551	-0.3089	3.2932	25000
Series	3.0102	0.1632	-0.3582	3.2856	

(For this case Robert Byers (CDC) used the SPlus package over 5100 cycles for $n = 200$ and found $E(c^* = 3.005$, against our series and Padé value 3.0050498. Dr. Byers also gave $\sigma(c^*) = 0.0611$ compared to our approximation 0.0559 using linear interpolation in the 1975 tabulation.

(c) $n = 2000, \beta_1 = 0.259182, \beta_2 = 3.11082, p = 4.7837, q = 15.2014, c = 0.5559$

	$E(c^*)$	$E(b_1)$	$E(b_2)$	cycle
Simulation	0.5752	0.26058	3.1033	4978
Series	0.5689	0.26058	3.1064	

Remarks: On the whole one can have confidence in the random number generator and its use in simulation studies. One does note the surprising discrepancy in the skewness estimates, in case (a). It could be related to the fact that the sampled distribution in U-shaped. As against this explanation case (b) is noticed - here however we have a symmetrical U-shape. Nonetheless case (c) is somewhat disturbing.

(d) $n = 1000, \beta_1 = 1.0, \beta_2 = 4.2, p = 2.4075, q = 19.5925, c = 0.6$

	$E(c^*)$	$\sigma(c^*)$	$\sqrt{\beta_1}(c^*)$	$\beta_2(c^*)$	cycle
Simulation	0.7626	0.3418	-0.1096	2.4092	4542
Series (Padè)	0.6568	0.4889	-1.3965	{ 7.6830 7.6060	

Remarks: The Padè approximants are based on Taylor series for the moments to order n^{-6} modified as rational fractions to reduce apparent divergence (Bowman and Shenton, 1975). The β-point is near the Type III line; note that in this case the value of c on the L_1 line ($\beta_1 = 1$) is 5, so that $0.6/5 = 0.12$. Thus this case is likely to have large variances for b^*, and q^*. From Table 8 for $\beta_1 = 1$, $\beta_2 = 4.25$, the n^{-1} variances of p^*, and q^* are 882 and 1158487 respectively.

(e) Padè Approximants, $n = 1000, \beta_1 = 1.0, \beta_2 = 4.2$

Uses terms to	$\mu_1'(c^*)$	$\mu_2(c^*)$	$\mu_3(c^*)$	$\mu_4(c^*)$
1	0.664181	0.240475	-0.169440	-0.303402
2	0.656842	0.238955	-0.162744	0.466210
3	0.656843	0.239008	-0.163177	0.438915
4	0.656843	0.239009	-0.163195	0.434521

4 Simulation Moments for a^*, b^*, p^*, q^*, c^*

A selection of sampled populations (Table 5) gives some idea of the response of simulated moments of the estimators to points (β_1, β_2) in the β-plane. As we have indicated previously, the response has three patterns;

(a) β-points in the neighborhood of $L_1 \equiv \beta_2 - \beta_1 - 1 = 0$ (J-shaped)

(b) β-points in the "middle" region (U-shaped)

(c) β-points in the neighborhood of $L_2 = 6 + 3\beta_1 - 2\beta_2 = 0$.

For practical purposes (c) is the most important, putting aside the question of sample size. Here a chaos zone exists when sampling can produce a cluster of large values of r^*, stemming from its denominator. This danger zone is difficult to define considering the extent of the β-plane and the included Type I region. If we limit this region to $0 < \sqrt{\beta_1} < 2$ or so, then one could compare the value of c^* observed to c^{**} on the L_1 line. As an alternative consider the kurtosis on line L_1 and line L_2, where the data point is (b_1, b_2) and belongs to Type III region. On line L_1, $\beta_2^{(1)} = 1 + b_1$, and on L_2, $\beta_2^{(2)} = 3 + 3b_1/2$. For the data point the kurtosis $\beta_2^{(d)} = b_2$. Hence consider the critical distance ratio

$$c_r(b_1, b_2) = \frac{\beta_2^{(2)} - \beta_2^{(d)}}{\beta_2^{(2)} - \beta_2^{(1)}} = \frac{3 + 3b_1/2 - b2}{2 + b_1/2}$$

If $c_r < 0$ for example, then we should consider fitting Type III distribution.

We now turn to dominant ($n \to \infty$) asymptotics for the moments of the estimators.

The Table 6. gives four moments of the estimators, sample size, and cycle size. The mean of a^* should be near to -2.18; that of b^* close to 10.75. Results of special interest are noted. Evidently, since we have $\sqrt{\beta_1} > 0$, problems are to be expected in some cases with b^* and q^*, and especially with small samples. In the present content, small is about 100, whereas stability in some cases may only be reached for $n = 5000$. Farther comments appear in a footnote to the Table 6.

Table 6. An Example: Pearson(1984). 7 Sample Sizes

(a)

$a = -2.18, b = 10.75, p = 3.6155, q = 7.5663, \mu_1 = 2, \sigma^2 = 3, \sqrt{\beta}_1 = 0.4, \beta_2 = 2.8, c = 0.88$

n		$\mu_1'^*$	σ^*	$\sqrt{\beta}_1^*$	β_2^*	cycles	c^* outside
100	a^*	-2.0511	1.1763	-5.2580	89.0642	47450	2550
	b^*	23.7758	1673.5923	208.4123	44466.239		
	p^*	3.7501	5.1713	33.8736	2242.8162		
	q^*	36.5366	3464.6444	210.2097	45105.492		
250	a^*	-2.1673	0.7571	-1.7229	8.8806	19674	326
	b^*	14.8278	128.6360	98.2245	11405.113		
	p^*	3.8332	2.2639	3.9879	32.5877		
	q^*	16.9468	288.7022	93.7617	10536.316		
500	a^*	-2.1927	0.5652	-1.6876	9.6754	9971	29
	b^*	12.0045	22.5196	60.0872	4377.1758		
	p^*	3.7931	1.6330	4.2114	40.9884		
	q^*	10.5141	61.4841	69.2238	5627.6438		
750	a^*	-2.1896	0.4485	-1.3564	7.6771	7498	2
	b^*	11.2220	3.4939	6.4907	77.6990		
	p^*	3.7356	1.2046	3.0770	24.5414		
	q^*	8.5632	6.6409	9.3961	140.0576		
1000	a^*	-2.1860	0.3815	-1.1824	7.1602	5000	0
	b^*	11.0612	2.5392	5.1087	64.5928		
	p^*	3.7003	0.9846	2.5556	19.1353		
	q^*	8.2145	4.4700	7.9262	129.3426		
3000	a^*	-2.1778	0.2074	-0.5267	3.4796	3000	0
	b^*	10.8220	1.0774	0.9645	4.7998		
	p^*	3.6332	0.4895	0.8770	4.4860		
	q^*	7.7149	1.6126	1.2657	6.1432		
5000	a^*	-2.1766	0.1558	-0.3626	3.3530	3000	0
	b^*	10.7804	0.7806	0.6108	3.6961		
	p^*	3.6205	0.3597	0.5915	3.6974		
	q^*	7.6336	1.1515	0.8136	4.3059		

Note the largeness of the skewness and kurtosis of b^* and q^* when $n = 750$, and how this largeness is exacerbated and spreads to the mean and standard deviation as n decreases.

(b) n^{-1} terms of Covariances

	a^*	b^*	p^*	q^*	$\sigma(n = 5000)$
a^*	127.4565	-414.7359	-276.1989	-702.2247	0.1597
b^*	-0.6716	2991.5589	1142.1251	4240.9823	0.7735
p^*	-0.9519	0.8125	660.5074	1866.2431	0.3635
q^*	-0.7821	0.9749	0.9130	6325.3879	1.1248

Values in the lower half triangle are correlations.

5 First Order Covariances

One possibility to better understand the structure of Type I estimation is to consider the large sample covariances, since exact covariances are out of the question. For in the case of moment estimators, we are dealing with functions of the mean, variance, third and forth central moments of a sample. The so-called delta method (see Appendix) is a very useful tool, and is quite straightforward for implementation on a calculator.

As an example, consider $y = m_3/m_2^{3/2}$. Then

$$\delta y = \frac{\delta m_3}{\mu_2^{3/2}} - \frac{3}{2}\frac{\mu_3 \delta m_2}{\mu_2^{5/2}},$$

so that to order n^{-1},

$$Var_1 y = \frac{Var_1(m_3)}{\mu_2^3} - \frac{3\mu_3}{\mu_2^4} Cov_1(m_2, m_3) + \frac{9}{4}\frac{\mu_3^2}{\mu_2^5} Var_1(m_2)$$

Similarly for $\delta(m_1', \sqrt{b_1})$ for example.

For given values of the basic moments, covariances (Table 6) are set up (values of n are to be inserted). These should alert the user to trouble spots, by scrutinizing the means and standard deviations in particular, and also the coefficient of variation.

In Table 7 we give a selection of first-order variances of p^* and q^*. Recall that these two estimators are solely dependent on the skewness and kurtosis. These should be useful in gaining some idea of distributional properties, and deciding whether to use a Type III model in preference to a Type I.

Table 7. Coefficients of $Var_1(p^*)$ and $Var_1(q^*)$

β_2	$\beta_1 = 0.5$		$\beta_1 = 1.0$		$\beta_1 = 2.0$		$\beta_1 = 3.0$	
	p^*	q^*	p^*	q^*	p^*	q^*	p^*	q^*
1.00	0.0665	0.1398						
1.75	0.1046	0.4410						
2.00	0.4313	2.0125						
2.50	3.8394	26.9379	0.2401	2.0424				
3.00	44.5818	822.949	1.7378	21.7378				
3.50			12.0893	394.763	0.1158	1.9139		
4.00			158.582	30929	0.6089	14.5656		
4.25			882.259	1158487	1.2545	41.1081	0.0208	0.5025
5.00					13.1950	2467.50	0.2965	10.8214
5.50					101.588	166178	0.9996	64.8096
6.00							3.4124	593.731
7.00							75.7705	579413

5.1 Checks on the Simulations using Approximate Percentage Points

We simulate percentage points for samples of 3000 and 5000 of the Pearson example in Table 6. The percentage points are also approximated by the Bowman-Shenton algorithm (1979a, 1979b), an algorithm using least squares and rational fractions in terms of the skewness and kurtosis. The algorithm is applicable to cases in which $\sqrt{\beta_1} < 2$; the restriction on β_2 is mild. A computer program has been written by Davis and Stephens (1983). Comparisons are shown in Table 8.

Table 8. Percentage Point Comparisons

$a = -2.18, b = 10.75, p = 3.6155, q = 7.5663, \mu_1 = 2, \sigma^2 = 3, \sqrt{\beta}_1 = 0.4, \beta_2 = 2.8, c = 0.88$

n	Parameter		0.01	0.05	0.10	0.90	0.95	0.99
3000	a^*	MC	-2.72	-2.54	-2.46	-1.94	-1.87	-1.78
		BS	-2.74	-2.55	-2.45	-1.93	-1.87	-1.77
	b^*	MC	8.93	9.31	9.55	12.17	12.80	14.02
		BS	8.97	9.36	9.60	12.23	12.80	14.07
	p^*	MC	2.75	2.95	3.05	4.27	4.52	5.03
		BS	2.77	2.96	3.07	4.28	4.53	5.08
	q^*	MC	5.07	5.61	5.91	9.72	10.60	12.50
		BS	5.18	5.65	5.96	9.80	10.72	12.84
5000	a^*	MC	-2.59	-2.45	-2.38	-1.99	-1.94	-1.85
		BS	-2.58	-2.45	-2.38	-1.98	-1.94	-1.85
	b^*	MC	9.27	9.63	9.81	11.78	12.15	12.96
		BS	9.29	9.64	9.85	11.81	12.18	12.95
	p^*	MC	2.92	3.08	3.18	4.08	4.26	4.63
		BS	2.92	3.09	3.19	4.09	4.26	4.62
	q^*	MC	5.52	5.97	6.24	9.07	9.69	10.84
		BS	5.55	6.02	6.29	9.14	9.73	10.99

The agreement between simulation approximations and algorithmic values are most gratifying. It must not be overlooked that in practice asymptotic results are valid for sufficiently large samples only - for moderate sample sizes they may be quite misleading. However, in the present case an alternative is lacking.

6 Variation of Percentage Points and the Pearson System

Let the αth percentage point of a distribution be

$$P_\alpha^* = m_1' + \sqrt{m_2}\pi_\alpha^*(\sqrt{b_1}, b_2)$$

where $\pi_\alpha^*(\sqrt{b_1}, b_2)$ is the standard αth level, and $(P_\alpha^* - m_1')/\sqrt{m_2} = \pi_\alpha^*(\sqrt{b_1}, b_2)$. π_α^* may be approximated satisfactorily by using the Bowman Shenton rational fraction approach. In fact

$$\pi(\sqrt{\beta_1}, \beta_2) = \pi_1(\sqrt{\beta_1}, \beta_2)/\pi_2(\sqrt{\beta_1}, \beta_2)$$

where

$$\pi_i(\sqrt{\beta_1}, \beta_2) = \sum\sum\nolimits_{0 \le r+s \le 3} a_{r,s}^{(i)}(\sqrt{\beta_1})^r {\beta_2}^s. \qquad (i = 1, 2)$$

The method is described in Bowman and Shenton (1979a, 1979b).

Now using incremental calculus

$$\begin{aligned}\delta\pi_\alpha &= \delta m_1' + \frac{\delta m_2}{2\sqrt{\mu_2}}\pi_\alpha^*(\sqrt{\beta_1}, \beta_2) + \sqrt{\mu_2}\left(\frac{\delta\pi^*\delta b_1}{\delta\beta_1} + \frac{\delta\pi^*\delta b_2}{\delta\beta_2}\right) \\ &= \delta m_1' + M_2\delta m_2 + P_1\delta b_1 + P_2\delta b_2,\end{aligned}$$

where

$$\begin{aligned}M_2 &= \pi_\alpha^*(\sqrt{\beta_1}, \beta_2)/(2\sqrt{\mu_2}), \\ P_1 &= \frac{\partial\pi_\alpha^*}{\partial\beta_1}\sqrt{\mu_2}, \\ P_2 &= \frac{\partial\pi_\alpha^*}{\partial\beta_2}\sqrt{\mu_2}.\end{aligned}$$

Examples are given in Table 9 relating to cases considered in the test. A set of percentage points are given along with the corresponding variance (to order n^{-1}) and standard deviation (to order $n^{-1/2}$). As was expected, variances for the levels considered are small compared to the extremes possible for variances of a^*, b^*, and p^*, q^*.

Note that the n^{-1} variances are not affected by a change of the mean.

Table 9. Percentage Points of Pearson Distributions and Its Variances

Moments of Pearson Distributions							
μ_1'	3.5010	2.0000	1.0000	1.0000	1.0000	1.0000	1.0000
μ_2	2.8251	3.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sqrt{\beta_1}$	0.5091	0.4000	1.0000	1.0000	1.0000	1.0000	1.0000
β_2	3.1108	2.8000	3.0000	3.5000	4.0000	4.2500	4.5000
Percentage Points							
1%	0.3856	-1.2408	0.0004	-0.2127	-0.4222	-0.5112	-0.5893
2.5%	0.7049	-0.9286	0.0023	-0.1928	-0.3466	-0.4041	-0.4514
5%	1.0314	-0.5951	0.0060	-0.1579	-0.2576	-0.2908	-0.3164
10%	1.4443	-0.1641	0.0247	-0.0720	-0.1132	-0.1230	-0.1286
25%	2.2644	0.7071	0.1561	0.1923	0.2344	0.2523	0.2680
50%	3.3421	1.8599	0.6529	0.7519	0.8045	0.8221	0.8361
75%	4.5700	3.1494	1.6114	1.5825	1.5658	1.5597	1.5546
90%	5.7711	4.3601	2.5916	2.4646	2.3889	2.3616	2.3390
95%	6.5212	5.0817	3.0856	2.9978	2.9285	2.9007	2.8765
97.5%	7.1822	5.6949	3.4133	3.4405	3.4148	3.3988	3.3831
99%	7.9525	6.3692	3.6769	3.9066	3.9916	4.0114	4.0231
n^{-1} Coefficients of Variance							
1%	8.7865	7.9536	0.2930	1.1515	2.7940	4.4060	7.1927
2.5%	4.7128	4.6058	0.2646	0.7193	1.2830	1.7805	2.5655
5%	3.3811	3.5588	0.2751	0.3837	0.6229	0.7956	1.0057
10%	3.0700	3.3627	0.1580	0.2439	0.4470	0.5582	0.6759
25%	3.2428	3.6249	0.2957	0.6584	0.8789	1.0002	1.1777
50%	3.4356	3.8643	1.7760	1.4380	1.2540	1.2061	1.1891
75%	5.1815	5.4390	3.7398	2.6390	2.2312	2.1458	2.1249
90%	8.5319	8.1419	5.0990	4.7657	4.5569	4.5776	4.7450
95%	12.3071	10.8480	4.5667	6.1326	6.7892	7.0296	7.3179
97.5%	18.7113	15.3129	3.6311	7.7314	10.5567	11.6385	12.5791
99%	36.4997	26.6484	3.9527	11.5963	20.4312	24.8478	29.2626
$n^{-1/2}$ Coefficients of σ							
1%	2.9642	2.8202	0.5413	1.0731	1.6715	2.0990	2.6819
2.5%	2.1709	2.1461	0.5144	0.8481	1.1327	1.3344	1.6017
5%	1.8388	1.8865	0.5245	0.6194	0.7892	0.8920	1.0029
10%	1.7522	1.8338	0.3975	0.4938	0.6686	0.7472	0.8221
25%	1.8008	1.9039	0.5438	0.8114	0.9375	1.0001	1.0852
50%	1.8535	1.9658	1.3327	1.1992	1.1198	1.0982	1.0905
75%	2.2763	2.3322	1.9339	1.6245	1.4937	1.4649	1.4577
90%	2.9209	2.8534	2.2581	2.1830	2.1347	2.1395	2.1783
95%	3.5081	3.2936	2.1370	2.4764	2.6056	2.6513	2.7052
97.5%	4.3257	3.9132	1.9055	2.7805	3.2491	3.4115	3.5467
99%	6.0415	5.1622	1.9881	3.4053	4.5201	4.9848	5.4095

7 Conclusion

The basic Koshal paper (1933) undoubtedly owed much of its mathematical expertise to Fisher and to a less extent to D.J. Finney. Nonetheless Koshal must have been a worthy worker to be the recipient of the attentions of these two statistical and mathematical leaders.

The papers must be reviewed in the light of the prevailing scientific environment around the turn of the century, and a few decades later. The narrow view would focus on the failure of the normal distribution to produce a good model for natural phenomenal data, leading to a more general model such as the Pearson system and differential series models (Gram-Charlier, etc.). Estimation procedures surfaced focusing on the discovery of the best approach; maximum likelihood against moment methods and efficient estimators based on smallness of asymptotic variance. The asymptotic assessment in practice slowly trickled down to subjective vagueness.

The broader view would note the emphasis on applied mathematics; Fisher took an astronomy course doubtless highlighting aspects of the theory of errors. whereas Pearson was influenced by Galton and theories of hereditary and evolution. Numerical analysis was merely a trivial subject left to self-learning techniques. A first text on the subject appeared in 1924 (Calculus of Observations, Whittaker and Robinson). Electronic computers as against mechanical were some 3-4 decades in the future. As for the asymptotic aspect, it seems likely that both Pearson and Fisher were unaware of the epoch making lectures of Émile Borel at the Sorbonne early in the present century. Note in passing that recent work by us on the statistical aspects of asymptotics (power series for moments) show marked divergence especially for complicated structures such as those associated with Type I sampling.

Some reflections on the Koshal-Turner (1930), Koshal (1933), Fisher (1937), Pearson (1895), Pearson-Filon (1898) follow.

(i) Fisher and Pearson had open access to two prestigious journals (*Biometrika, Annals of Eugenics*).

(ii) The case of a Type I model being the center of the controversy was pure serendipity. A more difficult model of 4 parameters would have been hard to find.

(iii) The rivals were apparently on fairly safe ground asymptotically with a sample of $n = 1000$ - further similar material from Koshal-Turner was available but not used. However 'safeness' considerations only apply if we concentrate on a criterion of 'goodness of fit', ignoring (for the data concerned) the chaotic approximate distributions of two of the estimators (b^*, and q^*).

(iv) With regard to (iii), Pearson-Filon made a valiant attempt to quantify, in modern terminology, percentage points for four parameters, along with the range, mode, mean, standard deviation, mean to mode distance, skewness, and modal frequency. Our study considers simulated covariances, and asymptotic covariances.

(v) We have attempted to define a gray area (of infinite extent) in the β-plane for which chaotic distributions may arise for some of the parameters (if $\sqrt{b_1} > 0$, particularly the parameter b for the range, and q for an index.) This area is bounded by Type III line and a nearly parallel line in the Type I region. This region will be deeper near the line $\beta_1 = b_1$ and tail off sharply to the left and slowly to the right. Another approach to this unsafe region, is to consider the n^{-1} variances of p^* and q^* and make a judgment relating them to the corresponding mean (in other words, consider the coefficient of variation).

(vi) The problem of the approximate distribution of percentage points awaits attention.

In the single sentence, it seems Karl Pearson was amongst scientists to set up calculus of morphological majors relating to plants and animals.

Appendix: First Order Covariances for the Estimators

A Background: The Delta Method

Since it is quite out of the question to simulate the full moments of the four estimators $(a^*, b^*;\ p^*, q^*)$ with their ten covariances over a representative region, we can partially overcome this problem by considering dominant asymptotics; we have to assume that bias is not important.

We are using moment estimators, and functions of moments. Thus the so-called delta approach turns out to be answer to asymptotic covariances; delta, here referring to an increment as in the calculus. An excellent account of the procedure is given in Kendall and Stuart (1963, pp228-245).

B Fundamental Covariances

There are three essential results for moments in general.

$$\begin{cases} Var_1(m_s) = \mu_{2s} - \mu_s^2 + s^2\mu_2\mu_{s-1}^2 - 2s\mu_{s-1}\mu_{s+1} \qquad (n \to \infty), \\ Cov_1(m_s, m_t) = \mu_{s+t} - \mu_s\mu_t + st\mu_2\mu_{s-1}\mu_{t-1} - s\mu_{s-1}\mu_{t+1} - t\mu_{s+1}\mu_{t-1}, \\ Cov_1(m_1', m_s) = \mu_{s+1} - s\mu_2\mu_{s-1}. \end{cases} \tag{1}$$

To derive $Var_1(b_2)$ say consider

$$y = m_4/m_2^2.$$

Then

$$\delta y = \delta m_4/\mu_2^2 - 2\mu_4\delta m_2/\mu_2^3.$$

Square and use (A1).

$$Var_1(b_1) = \beta_1(4\beta_4 - 24\beta_2 + 36 + 9\beta_1\beta_2 - 12\beta_3 + 35\beta_1), \tag{2}$$

$$Var_1(b_2) = \beta_6 - 4\beta_2\beta_4 + 4\beta_2^3 - \beta_2^2 + 16\beta_1\beta_2 - 8\beta_3 + 16\beta_1. \tag{3}$$

$$Cov_1(b_1, b_2) = 2\beta_5 - 3\beta_4\beta_1 - 4\beta_3\beta_2 + 6\beta_2^2\beta_1 + 3\beta_1\beta_2 - 6\beta_3 + 12\beta_1^2 + 24\beta_1, \tag{4}$$

$$\begin{aligned} Var_1(6 + 3b_1 - 2b_2) &= 4\beta_6 - 16\beta_2\beta_4 + 16\beta_2^3 + 72\beta_1\beta_4 - 24\beta_5 && (5)\\ &- 72\beta_1\beta_2^2 + 48\beta_2\beta_3 + 81\beta_1^2\beta_2 - 108\beta_1\beta_3 && (6)\\ &- 4\beta_2^2 - 188\beta_1\beta_2 + 40\beta_3 + 171\beta_1^2 + 100\beta_1. && (7) \end{aligned}$$

Here $\beta_3 = \mu_3\mu_5/\sigma^8$, $\beta_4 = \mu_6/\sigma^6$, $\beta_5 = \mu_3\mu_7/\sigma^{10}$, $\beta_6 = \mu_8/\sigma^8$.

C Higher moments of the Pearson System

Define $\nu_s = \mu_s/\sigma^s$. Then

$$\nu_{s+1} = \frac{s}{\Delta_s}\left[(\beta_2 + 3)(\sqrt{\beta_1})\nu_s + (4\beta_2 - 3\beta_1)\nu_{s-1}\right], \qquad (s = 1, 2, \cdots; \nu_0 = 1, \nu_1 = 0)$$

where

$$\Delta_s = 6(\beta_2 - \beta_1 - 1) - s(2\beta_2 - 3\beta_1 - 6).$$

All moments exist for the type I distribution. "Below" Type III line the highest moment is μ_s, where $s = [x] + 1$, and $x = 6(\beta_2 - \beta_1 - 1)/(2\beta_2 - 3\beta_1 - 6)$. ($[x]$ = integer part of x.)

The formula is due to Bowman and Shenton (1973).

D Further Formulas Needed

$$Cov_1(m_1', m_2) = \mu_3 \qquad \text{(subscript means the } \mathrm{n}^{-1} \text{ coefficient)}$$

$$Var_1(m_1') = \mu_2$$

$$Cov_1(m_1', b_1) = \sigma(2\beta_2 - 3\beta_1 - 6)\sqrt{\beta_1}$$

$$Cov_1(m_1', b_2) = \sigma(\mu_5/\sigma^5 - 4\sqrt{\beta_1} - 2\beta_2\sqrt{\beta_1})$$

$$Cov_1(m_2, b_1) = \sigma^2(2\sqrt{\beta_1}\mu_5/\sigma^5 - 5\beta_1 - 3\beta_1\beta_2)$$

$$Cov_1(m_2, b_2) = \sigma^2(\mu_6/\sigma^6 + \beta_2 - 4\beta_1 - 2\beta_2^2)$$

E Formulas for the Deltas of the Estimators

E.1 Parameter r^*

$$r^* = \frac{6(b_2 - b_1 - 1)}{c^*} = \frac{3(2b_2 - 2b_1 - 2)}{c^*} = \frac{3(2b_2 - 3b_1 - 6 + b_1 + 4)}{c^*} = -3 + \frac{3(b_1 + 4)}{c^*}$$

$$\begin{aligned}
\delta r^* &= \frac{3\delta b_1}{c} - \frac{3(\beta_1 + 4)(3\delta b_1 - 2\delta b_2)}{c^2} \\
&= \frac{3c\delta b_1 - 3(\beta_1 + 4)(3\delta b_1 - 2\delta b_2)}{c^2} \\
&= \frac{6(\beta_1 + 4)\delta b_2 + \delta b_1(18 + 9\beta_1 - 6\beta_2 - 9\beta_1 - 36)}{c^2} \\
&= \frac{6[(\beta_1 + 4)\delta b_2 - (3 + \beta_2)\delta b_1]}{c^2}.
\end{aligned}$$

Hence

$$Var_1(r^*) = \frac{36}{c^4}[(\beta_1 + 4)^2 Var_1(b_2) - 2(\beta_1 + 4)(3 + \beta_2)Cov_1(b_1, b_2) + (3 + \beta_2)^2 Var_1(b_1)].$$

E.2 The Estimator p^*

We have

$$\begin{aligned}
p &= \frac{r}{2}\left\{1 - \frac{(r + 2)\sqrt{\beta_1}}{\sqrt{(r + 2)^2\beta_1 + 16(r + 1)}}\right\} \\
&= \frac{r}{2} - \frac{r(r + 2)\sqrt{\beta_1}}{2\sqrt{D}} \qquad (D = (r + 2)^2\beta_1 + 16(r + 1))
\end{aligned}$$

Hence,

$$p^* = \frac{r^*}{2} - \frac{r^*(r^* + 2)\sqrt{b_1}}{2\sqrt{D^*}}$$

and

$$\delta p^* = \frac{\delta r^*}{2} - \frac{(r + 1)\sqrt{\beta_1}}{\sqrt{D}}\delta r^* - \frac{(r + 2)r\delta b_1}{4\sqrt{\beta_1}\sqrt{D}} + \frac{r(r + 2)\sqrt{\beta_1}}{4D^{3/2}}\delta D^*$$

where

$$\delta D^* = 2(r + 2)\beta_1\delta r^* + (r + 2)^2\delta b_1 + 16\delta r^*$$

This finally

$$\begin{aligned}
\delta p^* &= \frac{\delta r^*}{2} - \frac{(r + 1)\sqrt{\beta_1}}{\sqrt{D}}\delta r^* - \frac{(r + 2)r\delta b_1}{4\sqrt{\beta_1}\sqrt{D}} + \frac{r(r + 2)\sqrt{\beta_1}}{4D^{3/2}}[2(r + 2)\beta_1\delta r^* + (r + 2)^2\delta b_1 + 16\delta r^*] \\
&= A\delta r^* + B\delta b_1
\end{aligned}$$

say, where

$$
\begin{aligned}
A &= -\frac{(r+1)\sqrt{\beta_1}}{\sqrt{D}} + \frac{1}{2} + \frac{r(r+2)}{4D^{3/2}}\sqrt{\beta_1}[(2r+4)\beta_1 + 16] \\
&= \frac{1}{2} - \frac{(r+1)\sqrt{\beta_1}}{\sqrt{D}} + \frac{r\sqrt{\beta_1}}{2D^{3/2}}[(r+2)^2\beta_1 + 8(r+2)] \\
&= \frac{1}{2} - \frac{(r+1)\sqrt{\beta_1}}{\sqrt{D}} + \frac{r\sqrt{\beta_1}}{2D^{3/2}}(D - 8r) \\
&= \frac{1}{2} - \frac{(r+1)\sqrt{\beta_1}}{\sqrt{D}} + \frac{r\sqrt{\beta_1}}{2\sqrt{D}} - \frac{4r^2\sqrt{\beta_1}}{D^{3/2}} \\
&= \frac{1}{2} - \frac{\sqrt{\beta_1}}{2\sqrt{D}}(r+2) - \frac{4r^2\sqrt{\beta_1}}{D^{3/2}}
\end{aligned}
$$

$$
\begin{aligned}
B &= -\frac{(r+2)r}{4\sqrt{\beta_1}\sqrt{D}} + \frac{r(r+2)\sqrt{\beta_1}(r+2)^2}{4D^{3/2}} \\
&= -\frac{(r+2)r}{4\sqrt{\beta_1}\sqrt{D}} + \frac{r(r+2)}{4\sqrt{\beta_1}D^{3/2}}(D - 16r - 16] \\
&= -\frac{4r(r+1)(r+2)}{\sqrt{\beta_1}D^{3/2}}
\end{aligned}
$$

$$
\begin{aligned}
\delta p^* &= \left[\frac{1}{2} - \frac{(r+2)\sqrt{\beta_1}}{2\sqrt{D}} - \frac{4r^2\sqrt{\beta_1}}{D^{3/2}}\right]\frac{6}{c^2}[(4+\beta_1)\delta b_2 - (3+\beta_2)\delta b_1] - \frac{4r(r+1)(r+2)}{\sqrt{\beta_1}D^{3/2}}\delta b_1 \\
&= I\delta b_1 + J\delta b_2
\end{aligned}
$$

Hence the n^{-1} term in $Var(p^*)$ follows by squaring and using (1) - (3).

E.3 Variance of q^* and $Cov_1(p^*, q^*)$

$Var_1(q^*)$ follows from δp^* by replacing $\sqrt{\beta_1}$ by $-\sqrt{\beta_1}$. With this we set up the n^{-1} covariance for p^*, q^*.

E.4 Formula for δa^*

$$
a^* = m_1' - \frac{p^*}{r^*} \cdot \frac{\sqrt{m_2 D^*}}{2}
$$

$$
\begin{aligned}
\delta a^* &= \delta m_1' - \frac{\sqrt{\mu_2 D}}{2r}\delta p^* - \frac{p\sqrt{D}}{4r\sqrt{\mu_2}}\delta m_2 + \frac{p\sqrt{\mu_2 D}}{2r^2}\delta r^* - \frac{p\sqrt{\mu_2}}{4r\sqrt{D}}\delta D^* \\
&= \delta m_1' - \frac{p\sqrt{D}}{4r\sqrt{\mu_2}}\delta m_2 - \frac{\sqrt{\mu_2 D}}{2r}(I\delta b_1 + J\delta b_2) + \frac{p\sqrt{\mu_2 D}}{2r^2}(R_1\delta b_1 + R_2\delta b_2)
\end{aligned}
$$

$$-\frac{p\sqrt{\mu_2}}{4r\sqrt{D}}(D_1\delta b_1 + D_2\delta b_2)$$

$$= \quad \delta m_1' - \frac{p\sqrt{D}}{4r\sqrt{\mu_2}}\delta m_2 + A_1\delta b_1 + A_2\delta b_2$$

E.5 Formula for b^*

$$b^* = a^* + \sqrt{m_2 D^*}/2$$

$$\delta b^* = \delta a^* + \frac{\sqrt{D}}{4\sqrt{\mu_2}}\delta m_2 + \frac{\sqrt{\mu_2}}{4\sqrt{D}}\delta D^*$$

$$\delta b^* = \delta m_1' + B_1\delta b_1 + B_2\delta b_2 + B_3\delta m_2$$

References

Bowman, K.O. and Shenton, L.R. (1973). Note on the distribution of $\sqrt{\beta_1}$ in sampling from Pearson distributions, *Biometrika*, 60, 155-167.

Bowman, K.O. and Shenton, L.R. (1975). Tables of moments of the skewness and kurtosis statistics in non-normal sampling. Tech. Report UCCND-CSD-8, Union Carbide Corporation, Oak Ridge, Tennessee.

Bowman, K.O. and Shenton, L.R. (1979a). Approximate percentage points for Pearson distributions, *Biometrika*, 66, 147-151.

Bowman, K.O. and Shenton, L.R. (1979b). Further approximate Pearson percentage points and Cornish-Fisher, *Comm.Stat.Part B, Simulation and Computation*, B(8), 231-244.

Bowman, K.O. and Shenton, L.R. (1989). *Continued Fractions in Statistical Applications*, Marcel Dekker, Inc., New York.

Bowman, K.O. and Shenton, L.R. (1992). Parameter estimation for the beta distribution, *Journal of Statistical Computation and Simulation*, 43, 217-228.

Davis, C. and Stephens, M. (1983). AS192 approximate percentage points using Pearson curves, *Applied Statistics*, 32(3), 322-327.

Fisher, R.A. (1937). Professor Karl Pearson and method of moments, *Annals of Eugenics*, VII, Part IB, 303-318.

Johnson, N.L. and Kotz, S. (1970). *Continuous Univariate Distributions, Vol. 2*, Houghton Mifflin Company, Boston.

Jones, W. and Thron, W. (1990). *Continued Fractions: Analytic theory and Applications*, Addison-Wesly Publishing Co. London.

Kendall, M.G. and Stuart, A. (1963). *Advanced Theory of Statistics, Vol. 1. Distribution Theory*, Hafner Publishing Company, New York.

Koshal, R. (1933). Application of the method of maximum likelihood to the improvement of curves fitted by the method of moments, *JRSS*, 96, 303-313.

Koshal, R. and Turner, A. (1930). Studies in the sampling of cotton for the determination of fiber properties, *Journal of the Textile Institute Transaction*, XXI, 325-370.

Lau, H. and Lau, A. (1991). Effective procedures for estimating beta distribution's parameters and their confidence intervals, *Journal of Statistical Computation and Simulation*, 38, 139-150.

Pearson, K. (1894). Contributions to the mathematical theory of evolution, *Phil.Trans. A*, 185, 71-110.

Pearson, K. (1895). Contributions to the mathematical theory of evolution, ii. skew variation in homogeneous material, *Phil.Trans. A*, 186, 343-414.

Pearson, K. (1895). On the mathematical theory of errors of judgment with special reference to the personal equation. *Phil.Trans.A*, 198, 235-299.

Pearson, K. (1936). Methods of moments and method of maximum likelihood, *Biometrika*, 28, 34-59.

Pearson, K. and Filon, L. (1898). Mathematical contributions to the theory of evolution, iv. on the probable errors of frequency constants and on the influence of random selection on variation and correlation, *Phil.Trans.A.*, 191, 229-311.

Stieltjes, T.J. (1918). *Recherches sur les fractions continues, Tome 2*, VII, Euvres Groningen.

Wall, H. (1948). *Analytic Theory of Continued Fractions*, Chelsea Publishing Company, Bronx, N.Y.

Whittaker, E.T. and Robinson, J. (1924). *Calculus of Observations: Treatise on Numerical Mathematics*, Blackie and Sons, Limited, London.