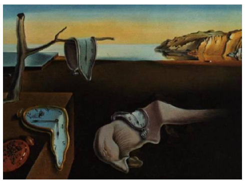

Journal of Statistical Software
September 2009, Volume 31, Code Snippet 2. http://www.jstatsoft.org/

A SAS/IML Macro for Computing Percentage Points of Pearson Distributions

Wei Pan
University of Cincinnati

Abstract

The Pearson distribution family provides approximations to a wide variety of observed distributions using the first four moments or the first three moments with a left or right boundary. Curve fitting utilizing Pearson distributions has been extensively applied in many fields. However, in practice, it is quite unwieldy to obtain percentage points of Pearson distributions when consulting the massive tables of Pearson and Hartley (1972) or using the out-of-date computer programs (Amos and Daniel 1971; Bouver and Bargmann 1974; Davis and Stephens 1983). The present study compiled a convenient computer program for computing the percentage points using the contemporary SAS/IML (SAS Institute Inc. 2008) macro language.

Keywords: Pearson distribution, Pearson curve, percentage point, curve fitting, SAS/IML.

1. Introduction

Pearson distributions (Pearson 1895) provide approximations to a wide variety of frequent distributions of empirical data using first four moments, or first three moments and left or right boundary. Curve fitting utilizing Pearson distributions has been extensively applied to statistical methodology as well as many practical fields, such as applied physics (Winterbon 1983), crystallography (Hall, Veeraraghavan, Rubin, and Winchell 1977), oceanography (Delignon, Garello, and Hillion 1997), water resources (US Water Resources Council 1967), and so on.

To find Pearson percentage points, researchers normally rely on Pearson and Hartley (1972)'s tables which give approximate percentage points in terms of skewness $\sqrt{\beta_1}$ and kurtosis β_2. However, the 28-page tables are unwieldy and often require second difference interpolation for both $\sqrt{\beta_1}$ and β_2. To deal with the cumbersome calculations, some computer programs were created for computing the percentage points (Amos and Daniel 1971; Bouver and Bargmann

1974; Davis and Stephens 1983). Unfortunately, the existing programs are over 25 years old and, therefore, not compatible with modern statistical software. Hence, it is imperative to have an updated and efficient computer program for computing percentage points of Pearson distributions, which constituted the purpose of the study.

2. Numerical method

The Pearson percentage points were approximated by Bowman and Shenton (1979) using the rational fraction approximation, i.e., the 19-point formula:

$$P(\sqrt{\beta_1}, \beta_2) = \pi_1(\sqrt{\beta_1}, \beta_2)/\pi_2(\sqrt{\beta_1}, \beta_2), \tag{1}$$

where for $i = 1, 2,$

$$\pi_i(\sqrt{\beta_1}, \beta_2) = \sum_{0 \le r+s \le 3}\sum a_{r,s}^{(i)}(\sqrt{\beta_1})^r \beta_2^s,$$

with $a_{0,0}^{(2)} = 1$ and all other coefficients $a_{r,s}^{(i)}$, called "A" array, are displayed in Davis and Stephens (1983) (Table 1, p. 324). For a particular percentile of α = 1.0%, 2.5%, 5.0%, 10.0%, 25.0%, 50.0%, 75.0%, 90.0%, 95.0%, 97.5%, or 99.0%, 19 points are chosen from the "A" array to be plugged into the formula. The numerical error was assessed as less than 1.0% of the true value (Bowman and Shenton 1979; Davis and Stephens 1983).

The approximation (1) gives a standardized percentage point y_α for a standardized Pearson variable Y. A transformation is needed to obtain an unstandardized percentage point from a standardized percentage point. For a random variable X of interest, its unstandardized percentage point can be calculated from $x_\alpha = \mu + \sigma \times y_\alpha$, where μ is the mean of X and σ is the standard deviation of X.

If the first four moments are known, the skewness $\sqrt{\beta_1}$ and kurtosis β_2 can be calculated as follows:

$$\sqrt{\beta_1} = \mu_3/\sigma^3 \text{ and } \beta_2 = \mu_4/\sigma^4.$$

Note that $\sigma = \mu_2$. If only the first three moments and left boundary are available, the value of β_2 can be determined by using Müller and Vahl (1976)'s algorithm. For the right boundary, the value of β_2 can be determined for $-X$ using the same algorithm.

3. SAS/IML macro program

Based on the numerical method presented in Section 2, a SAS/IML (SAS Institute Inc. 2008) macro program was written as follows (see the SAS/IML source file for the code):

```
%macroPearson(m =, sd =, rb1 =, b2 =, bndry =, type =);
```

where

`m` = mean μ;

`sd` = standard deviation σ;

`rb1` = skewness $\sqrt{\beta_1}$;

`b2` = kurtosis β_2;

`bndry` = left or right boundary;

$$\texttt{type} = \begin{cases} 1 & \text{if the first four moments are used;} \\ 2 & \text{if the first three moments and left boundary are used;} \\ 3 & \text{if the first three moments and right boundary are used.} \end{cases}$$

When running the macro program, if the first four moments are used (i.e., `type = 1`), simply input a "`.`" for `bndry`; if the first three moments and left or right boundary are used (i.e., `type = 2` or `type = 3`), simply input a "`.`" for `b2`.

The program also has user-friendly features to provide instant failure messages as clues for users to adjust parameter inputs. The failure messages are:

Message 0: Successful parameter inputs;

Message 1: σ is negative;

Message 2: `type` is not 1, 2, or 3;

Message 3: The absolute value of $\sqrt{\beta_1}$ is larger than 2.0;

Message 4: μ is impossible with the value entered for boundary;

Message 5: β_2 cannot be computed for the first 3 moments and left (or right) boundary;

Message 6: β_2 is out of the range, that is (Davis and Stephens 1983):

$$(\beta_2 < 1.5 \times |\sqrt{\beta_1}| + 1.5 \text{ or } \beta_2 > 0.2 \times |\sqrt{\beta_1}| + 10.8) \text{ if } |\sqrt{\beta_1}| \le 1;$$
$$(\beta_2 < 3.9 \times |\sqrt{\beta_1}| - 0.9 \text{ or } \beta_2 > 4.8 \times |\sqrt{\beta_1}| + 6.2) \text{ if } |\sqrt{\beta_1}| > 1.$$

4. Evaluation of the program

To evaluate the efficiency of the SAS/IML (SAS Institute Inc. 2008) macro program, some Pearson curve percentage points approximated by the program were compared with the corresponding ones in Pearson and Hartley (1972)'s tables (Table 32, p. 276), taking $\mu = 0$, $\sigma = 1$, $\sqrt{\beta_1} = 1.3$, and $\beta_2 = 4.2$ as an example. As can be seen in Table 1, the differences between the Pearson percentage points computed from the SAS/IML macro and those from Pearson and Hartley (1972)'s tables are all less than 0.0025 in term of the absolute value.

In addition, for comparing unstandardized Pearson curve percentage points, we took an example that was used in Pearson and Hartley (1972) (p. 79). That is, for $\mu = 0.08333$, $\sigma = 0.05$, $\sqrt{\beta_1} = 1.619$, and $\beta_2 = 6.7905$, the unstandardized Pearson curve percentage point for the

Percentile	SAS/IML Macro	Pearson and Hartley table	Difference
1.0%	−0.9789	−0.9786	−0.0003
2.5%	−0.9738	−0.9754	0.0016
5.0%	−0.9650	−0.9668	0.0018
10.0%	−0.9383	−0.9395	0.0012
25.0%	−0.7959	−0.7934	−0.0025
50.0%	−0.3412	−0.3411	−0.0001
75.0%	0.5042	0.5057	−0.0015
90.0%	1.4987	1.4996	−0.0009
95.0%	2.1122	2.1111	0.0011
97.5%	2.6092	2.6082	0.0010
99.0%	3.1176	3.1171	0.0005

Table 1: Approximated Pearson percentage points for `type = 1`: $\mu = 0$, $\sigma = 1$, $\sqrt{\beta_1} = 1.3$, and $\beta_2 = 4.2$.

percentile of 99% was 0.2531 from the SAS/IML macro program; whereas the Pearson and Hartley (1972)'s tables had 0.2535. The difference between them was less than 0.0004.

The examples above are for the case that the first four moments are known (i.e., `type = 1`). In order to verify that the SAS/IML macro is also effective for the case that the first three moments and left or right boundary are known (i.e., `type = 2` or `type = 3`), two more examples are taken, one for `type = 2`: $\mu = 0$, $\sigma = 1$, $\sqrt{\beta_1} = 1$, and the left boundary $= -1$; and the other for `type = 3`: $\mu = 0$, $\sigma = 1$, $\sqrt{\beta_1} = 0.5$, and the right boundary $= 2.932345$. The SAS/IML macro returned $\beta_2 = 3$ for `type = 2` and $\beta_2 = 2.4$ for `type = 3`. Table 2 and Table 3 show the the Pearson percentage points computed from the SAS/IML macro and those from Pearson and Hartley (1972)'s tables for `type = 2` and `type = 3`, respectively. From the tables, we can see that the differences between the Pearson percentage points computed from the SAS/IML macro and those from Pearson and Hartley (1972)'s tables are less than 0.0030 and 0.0044, respectively, in term of the absolute value.

Percentile	SAS/IML Macro	Pearson and Hartley table	Difference
1.0%	−0.9996	−0.9998	0.0002
2.5%	−0.9977	−0.9985	0.0008
5.0%	−0.9940	−0.9938	−0.0002
10.0%	−0.9753	−0.9753	0.0000
25.0%	−0.8439	−0.8437	−0.0002
50.0%	−0.3471	−0.3472	0.0001
75.0%	0.6114	0.6114	0.0000
90.0%	1.5916	1.5946	−0.0030
95.0%	2.0856	2.0859	−0.0003
97.5%	2.4133	2.4130	0.0003
99.0%	2.6769	2.6767	0.0002

Table 2: Approximated Pearson percentage points for `type = 2`: $\mu = 0$, $\sigma = 1$, $\sqrt{\beta_1} = 1$, and the left boundary $= -1$; the estimated $\beta_2 = 3$.

Percentile	SAS/IML Macro	Pearson and Hartley table	Difference
1.0%	−1.5008	−1.5006	−0.0002
2.5%	−1.4509	−1.4473	−0.0036
5.0%	−1.3652	−1.3688	0.0036
10.0%	−1.2235	−1.2262	0.0027
25.0%	−0.8277	−0.8282	0.0005
50.0%	−0.1414	−0.1415	0.0001
75.0%	0.7074	0.7070	0.0004
90.0%	1.4577	1.4587	−0.0010
95.0%	1.8481	1.8480	0.0001
97.5%	2.1369	2.1325	0.0044
99.0%	2.3974	2.3962	0.0012

Table 3: Approximated Pearson percentage points for `type = 3`: $\mu = 0$, $\sigma = 1$, $\sqrt{\beta_1} = 0.5$, and the right boundary = 2.932345; the estimated $\beta_2 = 2.4$.

5. Concluding remarks

The new program was written using the SAS/IML (SAS Institute Inc. 2008) macro language that can help researchers obtain Pearson percentage points much more efficiently than computing interpolation with Pearson and Hartley (1972)'s tables. It is worthy to note that this SAS/IML macro as well as other existing computer programs and Pearson and Hartley (1972)'s tables provide a means of finding percentage points of Pearson distributions other than percentiles or probability values of the distributions. In further study, it would be desirable to develop a computer program to obtain probability values of Pearson distributions to meet the needs of significance tests using Pearson distributions.

References

Amos DE, Daniel SL (1971). "Tables of Percentage Points of Standardized Pearson Distributions." *Research Report SC-RR-71 0348*, Sanida Laboratories, Albuquerque, NM.

Bouver H, Bargmann RE (1974). "Tables of the Standardized Percentage Points of the Pearson System of Curves in Terms of β_1 and β_2." *Technical Report No. 107*, Department of Statistics and Computer Science, University of Georgia, Georgia, GA.

Bowman KO, Shenton LR (1979). "Approximate Percentage Points for Pearson Distributions." *Biometrika*, **66**(1), 147–151.

Davis CS, Stephens MA (1983). "Approximate Percentage Points Using Pearson Curves." *Applied Statistics*, **32**(3), 322–327.

Delignon Y, Garello R, Hillion A (1997). "Statistical Modelling of Ocean SAR Images." *IEE Proceedings - Radar, Sonar and Navigation*, **144**(6), 348–354.

Hall Jr MM, Veeraraghavan VG, Rubin H, Winchell PG (1977). "The Approximation of Symmetric X-ray Peaks by Pearson Type VII Distributions." *Journal of Applied Crystallography*, **10**(1), 66–68.

Müller PH, Vahl H (1976). "Pearson's System of Frequency Curves Whose Left Boundary and First Three Moments Are Known." *Biometrika*, **63**(1), 191–194.

Pearson ES, Hartley HO (1972). *Biometrika Tables for Statisticians*, volume II. Cambridge University Press, New York.

Pearson K (1895). "Contributions to the Mathematical Theory of Evolution. II. Skew Variations in Homogeneous Material." *Philosophical Transactions of the Royal Society of London, Series A*, **186**, 343–414.

SAS Institute Inc (2008). *SAS/IML 9.2 User's Guide*. Cary, NC. URL http://www.sas.com/.

US Water Resources Council (1967). "A Uniform Technique for Determining Flood Flow Frequencies." *Bulletin No. 15*, US Water Resources Council, Washington, DC.

Winterbon KB (1983). "Pearson Distributions for Ion Ranges." *Applied Physics Letters*, **42**(2), 205–206.

Affiliation:

Wei Pan
P.O. Box 210049
University of Cincinnati
Cincinnati, OH 45221-0049, United States of America
Telephone: +1/513/556-2610
Fax: +1/513/556-3535
E-mail: wei.pan@uc.edu

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 31, Code Snippet 2 *Submitted:* 2009-02-17
September 2009 *Accepted:* 2009-07-07