CDF/MEMO/STATISTICS/PUBLIC/6820

A Guide to the Pearson Type IV Distribution

Joel Heinrich—University of Pennsylvania

December 21, 2004

1 Introduction

The Pearson Type IV distribution was employed recently to (empirically) characterize the shape of a pull distribution in [1]. Unfortunately, the information necessary to utilize Pearson Type IV must be gleaned from a relatively small number of references that are individually incomplete, and perhaps not readily available to many physicists. It was therefore deemed useful to summarize the collected information in a brief CDF note, which also includes one or two possibly new results. The goal is to present enough information to enable others to use this distribution in their analyses.

Pearson Type IV is used to fit observed distributions obtained from data or Monte Carlo simulations. The Pearson types (I–XII) [2, 3, 4] were intended to provide distributions to approximate all unimodal possibilities: distributions that are modeled well by Pearson Type IV (asymmetric with extensive tails) are not modeled well by the standard "textbook" distributions. One often requires a fit to an empirically obtained distribution to estimate quantities like P-values, to generate toy Monte Carlo (simulating an observed spectrum), or simply to characterize an observed distribution with a small number of parameters. Pearson Type IV seems especially well suited to model pull distributions, which are often asymmetric with non Gaussian tails.

2 The Pearson Type IV p.d.f.

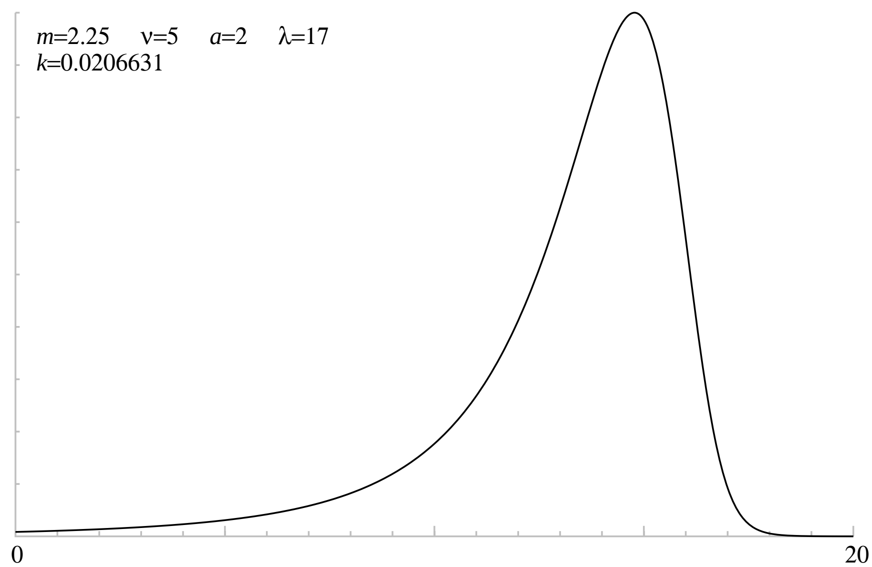
The Pearson Type IV probability density function is given by

$$f(x)dx = k\left[1+\left(\frac{x-\lambda}{a}\right)^2\right]^{-m}\exp\left[-\nu\tan^{-1}\left(\frac{x-\lambda}{a}\right)\right]dx \qquad (m > 1/2)$$

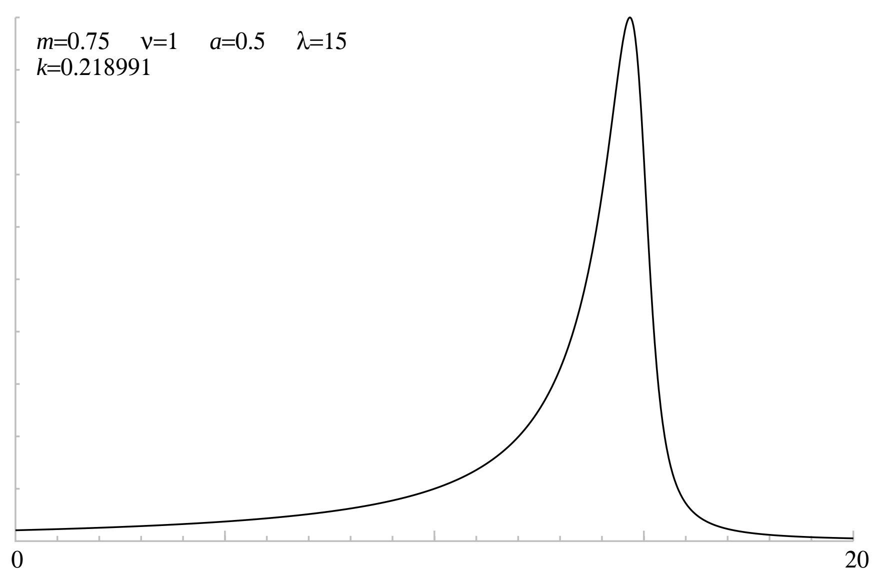
where m, ν, a, and λ are real-valued parameters, and $-\infty < x < \infty$ (k is a normalization constant that depends on m, ν, and a). (The symbols "m", "ν", and "a" are adopted from [2, 5], and "λ" is taken from [6], since [2, 5] set $\lambda = 0$.) Since the p.d.f. is invariant under the simultaneous change ($a \to -a$, $\nu \to -\nu$), we also specify the convention $a > 0$. When $m \le 1/2$, the p.d.f. is not normalizable.

Pearson Type IV is essentially an asymmetric version of Student's t distribution (when $\nu = 0$, it *is* a Student's t). In particular, for small values of the parameter m, the tails are much longer than those of a Gaussian. For $m = 1$, Pearson Type IV becomes an asymmetric version of the Breit-Wigner (Cauchy) distribution.

We show two plots of the Pearson Type IV p.d.f. below:



The case $m = 2.25$ shows a tail that is obviously non Gaussian.



The case $m = 0.75$ shows a very extensive tail.

3 Mode and Inflection Points

This topic is covered in [5] for the general Pearson family of distributions. The derivative of the Pearson IV p.d.f. with respect to x is given by

$$\frac{1}{f(x)}\frac{df(x)}{dx} = -2m\frac{x-\lambda+\frac{a\nu}{2m}}{a^2+(x-\lambda)^2}$$

which means that there is always a single mode $\mathcal{M}$ given by

$$\mathcal{M} = \lambda - \frac{a\nu}{2m}$$

where the derivative is zero. The second derivative is given by

$$\frac{1}{f(x)}\frac{d^2f(x)}{dx^2} = \frac{2m(2m+1)(x-\lambda+\frac{a\nu}{2m})^2 - \frac{a^2}{2m}(4m^2+\nu^2)}{\left[a^2+(x-\lambda)^2\right]^2}$$

which is zero at exactly two inflection points $\mathcal{I}_+$ and $\mathcal{I}_-$ given by

$$\mathcal{I}_\pm = \mathcal{M} \pm \frac{a}{2m}\sqrt{\frac{4m^2+\nu^2}{2m+1}}$$

The inflection points are thus always equidistant from the mode. The value of the second derivative at the mode is

$$\left.\frac{d^2f(x)}{dx^2}\right|_{x=\mathcal{M}} = \frac{-(2m)^3}{a^2(4m^2+\nu^2)}f(\mathcal{M})$$

which is always negative, proving that the mode is located at the maximum of the p.d.f., not a minimum.

4 Moments

This section is based on [2, 5]. The mean $\langle x \rangle$ of the p.d.f. is

$$\langle x \rangle = \lambda - \frac{a\nu}{2(m-1)} \qquad (m > 1)$$

(The moments can be calculated without knowing k.) For $\nu \neq 0$ and $m \leq 1$, we have $\langle x \rangle = \pm\infty$, depending on the sign of ν: $\langle x \rangle = -\infty$ when $\nu > 0$.

The variance μ_2 is given by

$$\mu_2 = \langle (x - \langle x \rangle)^2 \rangle = \frac{a^2}{r^2(r-1)}(r^2 + \nu^2) \qquad (m > 3/2)$$

where we have followed [2, 5] in using r as an abbreviation for $2(m-1)$. For $m \leq 3/2$, the variance is infinite.

Similarly, the third and fourth[1] moments are

$$\mu_3 = -\frac{4a^3\nu(r^2+\nu^2)}{r^3(r-1)(r-2)} \qquad (m > 2)$$

$$\mu_4 = \frac{3a^4(r^2+\nu^2)[(r+6)(r^2+\nu^2) - 8r^2]}{r^4(r-1)(r-2)(r-3)} \qquad (m > 5/2)$$

These expressions are often written in the alternative[2] form

$$\sqrt{\beta_1} \equiv \frac{\mu_3}{\mu_2^{3/2}} = \frac{-4\nu}{r-2}\sqrt{\frac{r-1}{r^2+\nu^2}}$$

$$\beta_2 \equiv \frac{\mu_4}{\mu_2^2} = \frac{3(r-1)[(r+6)(r^2+\nu^2) - 8r^2]}{(r-2)(r-3)(r^2+\nu^2)}$$

The recurrence relation

$$\mu_n = \frac{a(n-1)}{r^2[r-(n-1)]}\left[-2\nu r\mu_{n-1} + a(r^2+\nu^2)\mu_{n-2}\right] \qquad (n \geq 2)$$

connects the moments—by definition, $\mu_0 = 1$ and $\mu_1 = 0$.

[1] Unfortunately, the expression for μ_4 is incorrect in [5]; it's given correctly in [2].

[2] As we define $\sqrt{\beta_1}$ to have the same sign as μ_3, $\sqrt{\beta_1}$ may be either positive or negative.

5 Normalization

Unfortunately, after an excellent start, [5] gives no information about how to calculate the normalization factor k. Refs. [2, 6, 7] all give

$$k = \frac{2^{2m-2}|\Gamma(m+i\nu/2)|^2}{\pi a\Gamma(2m-1)} = \frac{\Gamma(m)}{\sqrt{\pi}a\Gamma(m-1/2)}\left|\frac{\Gamma(m+i\nu/2)}{\Gamma(m)}\right|^2$$

The complex gamma function is available via CERNLIB `CGAMMA` [8].

Instead of calling `CGAMMA`, one can use the fact that

$$\left|\frac{\Gamma(x+iy)}{\Gamma(x)}\right|^2 = \frac{1}{F(-iy, iy; x; 1)}$$

where F is the hypergeometric function [9], sometimes written ${}_2F_1$, which for $|z| \leq 1$ can be calculated via the series

$$F(a,b;c;z) = 1 + \frac{ab}{c}\frac{z}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{z^2}{2!} + \frac{a(a+1)(a+2)b(b+1)(b+2)}{c(c+1)(c+2)}\frac{z^3}{3!} + \cdots$$

While for $|z| = 1$ this series converges absolutely when $\Re(c-a-b) > 0$, the nth term is of order $n^{-(1+c-a-b)}$ (for large n), so convergence is slow (when $|z| = 1$) unless $\Re(c-a-b) \gg 1$. That is, the series for $F(-iy, iy; x; 1)$ converges rapidly only when $x \gg 1$.

From the relation $\Gamma(z+1) = z\Gamma(z)$, it is trivial to show that

$$\begin{aligned}\left|\frac{\Gamma(x+iy)}{\Gamma(x)}\right|^2 &= \left[1+\left(\frac{y}{x}\right)^2\right]^{-1}\left|\frac{\Gamma(x+1+iy)}{\Gamma(x+1)}\right|^2 \\ &= \left[1+\left(\frac{y}{x}\right)^2\right]^{-1}\left[1+\left(\frac{y}{x+1}\right)^2\right]^{-1}\left|\frac{\Gamma(x+2+iy)}{\Gamma(x+2)}\right|^2 = \cdots\end{aligned}$$

So a workable strategy for small x is to calculate $F(-iy, iy; x+n; 1)$ via the series, for some n chosen to be sufficiently large, and work down to $n = 0$ using these relations.

Reference [7] simply suggests using

$$\left|\frac{\Gamma(x+iy)}{\Gamma(x)}\right|^2 = \prod_{n=0}^{\infty}\left[1+\left(\frac{y}{x+n}\right)^2\right]^{-1}$$

but for large y, this scheme, in practice, is too CPU-intensive even when only moderate precision is required.

5.1 example C code

In fact, it takes more space to describe the calculation of k than is necessary for its implementation, so we present C code that does the job:

```
#include<math.h>
#include<float.h>
#include<assert.h>

double gammar2(double x,double y) {
  /* returns abs(gamma(x+iy)/gamma(x))^2 */
  const double y2=y*y, xmin = (2*y2>10.0) ? 2*y2 : 10.0;
  double r=1, s=1, p=1, f=0;
  while(x<xmin) {
    const double t = y/x++;
    r *= 1 + t*t;
  }
  while (p > s*DBL_EPSILON) {
    p *= y2 + f*f;
    p /= x++ * ++f;
    s += p;
  }
  return 1.0/(r*s);
}

double type4norm(double m,double nu,double a) {
  /* returns k */
  assert(m>0.5);
  return 0.5*M_2_SQRTPI*gammar2(m,0.5*nu)*exp(lgamma(m)-lgamma(m-0.5))/a;
}
```

When testing, the following special cases [9] are useful:

$$F(-iy, iy; \tfrac{1}{2}; 1) = \cosh(\pi y) \qquad\qquad F(-iy, iy; 1; 1) = \frac{\sinh(\pi y)}{\pi y}$$

6 The Cumulative Distribution

The cumulative distribution is defined as

$$P(x) = \int_{-\infty}^{x} f(t)dt$$

The calculation of this quantity in the general case (real-valued m) is poorly covered by the existing literature. Ref. [10] gives a continued fraction expansion. No one seems to have noticed that $P(x)$ can be expressed in terms of the hypergeometric function, which we will now demonstrate:

The cumulative distribution is

$$P(x) = ka \int_{-\infty}^{\frac{x-\lambda}{a}} (1+t^2)^{-m} e^{-\nu \tan^{-1} t} dt$$

Using the substitution $t = \tan\theta$ one obtains

$$P(x) = ka \int_{-\pi/2}^{\tan^{-1}\frac{x-\lambda}{a}} \cos^{2m-2}\theta e^{-\nu\theta} d\theta = kae^{\nu\pi/2} \int_{0}^{\frac{\pi}{2}+\tan^{-1}\frac{x-\lambda}{a}} \sin^{2m-2}\phi e^{-\nu\phi} d\phi$$

where $\phi = \theta + \pi/2$. A more familiar form for the integral

$$I = \int_0^y e^{-\nu\phi} \sin^r \phi \, d\phi$$

is achieved through the substitution $w = 1 - e^{2i\phi} = -2ie^{i\phi}\sin\phi$, which yields

$$I = (-2i)^{-r-1} \int_0^z w^r (1-w)^{(i\nu - r - 2)/2} dw$$

where z is given by

$$z = 1 - e^{2iy} = -2ie^{iy}\sin y = \frac{2}{1 - i\frac{x-\lambda}{a}}$$

The integral w.r.t. w is an incomplete Beta function [11], which is related to the hypergeometric function by

$$\int_0^z w^{\alpha-1}(1-w)^{\beta-1} dw = \frac{z^\alpha}{\alpha} F(\alpha, 1-\beta; \alpha+1; z) = \frac{z^\alpha (1-z)^\beta}{\alpha} F(1, \alpha+\beta; \alpha+1; z)$$

So we have

$$I = \frac{e^{-\nu y} \sin^{r+1} y}{r+1} e^{iy} F\left(1, \frac{r+2+i\nu}{2}; r+2; -2ie^{iy}\sin y\right)$$

which yields

$$P(x) = \frac{ka}{2m-1}\left[1 + \left(\frac{x-\lambda}{a}\right)^2\right]^{-m} \exp\left[-\nu \tan^{-1}\left(\frac{x-\lambda}{a}\right)\right] \times$$

$$\left(i - \frac{x-\lambda}{a}\right) F\left(1, m + \frac{i\nu}{2}; 2m; \frac{2}{1 - i\frac{x-\lambda}{a}}\right)$$

or more compactly,

$$\frac{P(x)}{f(x)} = \frac{a}{2m-1}\left(i - \frac{x-\lambda}{a}\right) F\left(1, m + \frac{i\nu}{2}; 2m; \frac{2}{1 - i\frac{x-\lambda}{a}}\right)$$

This is the main result. The corresponding series will be absolutely convergent when $x < \lambda - a\sqrt{3}$. For $x > \lambda + a\sqrt{3}$ one can use the identity

$$P(m, \nu, a, \lambda; x) = 1 - P(m, -\nu, a, -\lambda; -x)$$

To handle the case $|x - \lambda| < a\sqrt{3}$ one can apply a "linear transformation" (e.q. 15.3.7 in [9]) to produce

$$P(x) = \frac{1}{1 - e^{-(\nu + 2im)\pi}} -$$

$$\frac{iaf(x)}{i\nu - 2m + 2}\left[1 + \left(\frac{x-\lambda}{a}\right)^2\right] F\left(1, 2-2m; 2-m+\frac{i\nu}{2}; \frac{1 + i\frac{x-\lambda}{a}}{2}\right)$$

There are many schemes for calculating the hypergeometric function other than direct summation of the power series. One interesting example is provided in [12], which can also serve as a brief introduction to the hypergeometric function itself.

7 Generation

An elegant method for generating Pearson IV random deviates when $m > 1$ is given as an exercise in Luc Devroye's "Non-Uniform Random Variate Generation" [13]. Those interested in the method should read the text, which is available online; here we only present our implementation (in C), which, along with the code that calculates k, constitutes our solution to the exercise.

```
#include <math.h>
#include <assert.h>
double ranu(void);  /* uniform 0 to 1 */
double type4norm(double m,double nu,double a);

double rpears4(double m,double nu,double a,double lam) {
  /* returns random Pearson IV deviate */
  const double k=type4norm(m,nu,1.0), b=2*m-2, M=atan(-nu/b);
  const double cosM=b/sqrt(b*b+nu*nu), r=b*log(cosM)-nu*M, rc=exp(-r)/k;
  double x,z;
  assert(m>1);
  do {
    int s=0;
    z = 0;
    if( (x=4*ranu()) > 2 ) {
      x -= 2;
      s = 1;
    }
    if (x > 1) x = 1 - (z=log(x-1)) ;
    x = (s) ? M + rc*x : M - rc*x;
  } while (fabs(x) >= M_PI_2 || z + log(ranu()) > b*log(cos(x)) - nu*x - r);
  return a*tan(x) + lam;
}
```

8 Fitting to Pearson IV

When Karl Pearson developed his family of distributions, the maximum likelihood method was unknown. Pearson used the method of moments, which is not really adequate in many cases, but may be used to provide starting values to a maximum likelihood fitter.

The method of moments for Pearson IV is covered in [2, 5]: One simply computes the moments $\langle x \rangle$, μ_2, $\sqrt{\beta_1}$ and β_2 from the data, and finds the Pearson IV parameters that would give these moments. The following equations achieve the desired result:

$$r = 2(m-1) = \frac{6(\beta_2 - \beta_1 - 1)}{2\beta_2 - 3\beta_1 - 6} \qquad \nu = -\frac{r(r-2)\sqrt{\beta_1}}{\sqrt{16(r-1) - \beta_1(r-2)^2}}$$

$$a = \frac{\sqrt{\mu_2[16(r-1) - \beta_1(r-2)^2]}}{4} \qquad \lambda = \langle x \rangle - \frac{(r-2)\sqrt{\beta_1}\sqrt{\mu_2}}{4}$$

The preferred method, maximum likelihood fitting, requires minimizing the negative log likelihood given by

$$-\ln L = m \sum_{i=1}^{N} \ln \left[1 + \left(\frac{x_i - \lambda}{a} \right)^2 \right] + \nu \sum_{i=1}^{N} \tan^{-1} \left(\frac{x_i - \lambda}{a} \right) - N \ln k$$

where there are N observed data points x_i. This must be done numerically. If analytic derivatives with respect to the parameters are desired, the only non elementary ones are

$$\frac{\partial \ln k}{\partial \nu} = -\Im\psi(m + i\nu/2) \qquad \frac{\partial \ln k}{\partial m} = 2[\ln 2 + \Re\psi(m + i\nu/2) - \psi(2m - 1)]$$

which can be evaluated using the `CPSIPG` function from CERNLIB [14].

References

[1] J. Heinrich, 2nd Top Mass Workshop, Nov 11–12 2003, Lake Geneva WI. www-cdf.fnal.gov/internal/WebTalks/Archive/0311/031112_top_mass_workshop_nov12morning/07_031112_top_mass_workshop_nov12morning_David_AmbroseJoelHeinrich_1_tm3.pdf

[2] Karl Pearson, "Contributions to the Mathematical Theory of Evolution.—II. Skew Variation in Homogeneous Material", Philosophical Transactions of the Royal Society of London A, **186**, (1895), page 343.

[3] Karl Pearson, "Mathematical Contributions to the Theory of Evolution.—X. Supplement to a Memoir on Skew Variation", Philosophical Transactions of the Royal Society of London A, **197**, (1901), page 443.

[4] Karl Pearson, "Mathematical Contributions to the Theory of Evolution.—XIX. Second Supplement to a Memoir on Skew Variation", Philosophical Transactions of the Royal Society of London A, **216**, (1916), page 429.

[5] Alan Stuart and J. Keith Ord, "Kendall's Advanced Theory of Statistics", Volume 1 6th ed., (Oxford University Press, Oxford, 1994), §6.2, §6.8, §6.11.

[6] Harold Jeffreys, "Theory of Probability", 3rd ed., (Oxford University Press, Oxford, 1961), §2.3, page 75.

[7] Yuichi Nagahara, "The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters", Statistics & Probability Letters, **43**, (1999), page 251.

[8] CERN Program Library, C305, CGAMMA, wwwasdoc.web.cern.ch/wwwasdoc/shortwrupsdir/c305/top.html

[9] M. Abramowitz and I.A. Stegun, editors, "Handbook of Mathematical Functions", (United States Department of Commerce, National Bureau of Standards, Washington, D.C. 1964; and Dover Publications, New York, 1968), chapter 15.

[10] L.R. Shenton and J.A. Carpenter, "The Mills Ratio and the Probability Integral for a Pearson Type IV Distribution", Biometrika, **52**, (1964), page 119.

[11] M. Abramowitz and I.A. Stegun, editors, ibid., §6.6, page 263.

[12] William H. Press, et al., "Numerical Recipes", 2nd edition, (Cambridge University Press, Cambridge, 1992), §5.14 and §6.12
`lib-www.lanl.gov/numerical/bookcpdf/c5-14.pdf`
`lib-www.lanl.gov/numerical/bookcpdf/c6-12.pdf`

[13] Luc Devroye, "Non-Uniform Random Variate Generation", (Springer-Verlag, New York, 1986), Chapter VII, §2.7, Exercise 1, page 308.
`cgm.cs.mcgill.ca/~luc/rnbookindex.html`

[14] CERN Program Library, C317, CPSIPG,
`wwwasdoc.web.cern.ch/wwwasdoc/shortwrupsdir/c317/top.html`