Metrika (2006) 64: 131–138
DOI 10.1007/s00184-006-0039-x

ORIGINAL ARTICLE

**Antonia Castaño-Martínez**
**Fernando López-Blázquez**

# Heuristic approximation to Cramér-von Mises type statistics

Received: 1 April 2004 / Accepted: 9 May 2005 / Published online: 21 February 2006


**Abstract** We derive a formal expansion for a distribution in terms of another distribution. As a particular case we get the formal Edgeworth expansion. The heuristic procedure that we present is used to obtain approximations for distribution functions of the Cramér-von Mises and Watson goodness-of-fit statistics. Finally we compare our results with some obtained in the literature.

**Keywords** Approximation to distributions · Cramér-von Mises statistic · Watson statistic · Orthogonal polynomials

## 1 Introduction

The representation of a distribution in terms of another is widely used as a technique for obtaining approximations of distribution functions. One of the most popular representations is the Edgeworth expansion to approximate a distribution in terms of its cumulants and the normal distribution. General applications of the Edgeworth expansion have been discussed by Wallace (1958).

In this work, we present a heuristic procedure to obtain approximations of a distribution function $F_1$ in terms of another, $F_2$, which is known. As a particular case we get the formal Gram-Charlier type A expansion which is arranged to obtain the Edgeworth expansion for the normalized sum of $n$ iid random variables.

A. Castaño-Martínez (✉)
Universidad de Cádiz Departamento de Estadística e Investigación Operativa
Polígono Río San Pedro, 11510, Puerto Real (Cádiz), Spain
E-mail: antonia.castano@uca.es

F. López-Blázquez
Universidad de Sevilla, Departamento de Estadística e Investigación Operativa
Tarfia, s/n, 41012, Sevilla, Spain
E-mail: lopez@us.es

The main purpose of this work is to give heuristic approximations for the small-sample distribution functions of the Cramér-von Mises and Watson goodness-of-fit statistics from their asymptotic distributions.

The statistics of Watson and Cramér-von Mises are used to test the null hypothesis that $n$ observations come from a continuous cumulative distribution function $F(x)$. The exact sampling distributions under the null hypothesis of these statistics for any sample size $n$, are not known. For most goodness-of-fit purposes, the percentage points in the upper tail are required. It is clear that the upper tail of the distribution is far too difficult to calculate exactly, and attempts have been made to approximate the distribution by well known systems of curves.

The paper is structured as follows. In the following section we present a heuristic procedure to approximate distributions. In section 3 we apply step by step this procedure to the Cramér-von Mises distribution. We show some results and compare with some given in the literature. Similarly, in section 4 we give corresponding results for the Watson distribution.

## 2 Heuristic approximation

We consider two distribution functions $F_1$, $F_2$, and a system of orthogonal functions with respect to the distribution $F_2$, denoted by $\{\psi_k\}_{k\geq 0}$, with $\psi_0 \equiv 1$.

If we consider the indicator function of a random variable $X$ with distribution function $F_1$, in a formal way we have

$$I_{(-\infty,x]}(X) = F_2(x) + \sum_{k=1}^{\infty} \alpha_k(x)\, \psi_k(X) \tag{1}$$

with

$$\alpha_k(x) = \frac{\int_{-\infty}^{x} \psi_k(y)\, dF_2(y)}{\|\psi_k\|_2^2}, \tag{2}$$

where the norm, $\|\cdot\|_2$ is given by:

$$\|\psi_k\|_2^2 = \int_{-\infty}^{\infty} \psi_k^2(x)\, dF_2(x).$$

So, taking expectations (with respect to $F_1$) on both sides of (1), we have

$$F_1(x) = F_2(x) + \sum_{k=1}^{\infty} \alpha_k(x)\, E[\psi_k(X)]. \tag{3}$$

Considering truncated expansions in the right hand side of (3), we get approximations to the distribution function $F_1$, in terms of $F_2$, i.e.

$$F_1(x) \simeq F_2(x) + \sum_{k=1}^{j} \alpha_k(x)\, E[\psi_k(X)]. \tag{4}$$

It appears formidable to develop rigorously this heuristic method in any generality. As a possible justification of this procedure we will show that it is possible to obtain formal Edgeworth expansions following the above method.

Let $X_1, \ldots, X_n$ a random sample from one distribution $F$ which has cumulants $k_1, k_2, k_3, \ldots$. Consider $Z_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma$ where $\mu = k_1$ and $\sigma^2 = k_2$. The expansion of the indicator function $I_{(-\infty,x]}(y)$ in terms of the Hermite polynomials is

$$I_{(-\infty,x]}(y) = \Phi(x) - \phi(x) \sum_{k=1}^{\infty} \frac{H_{k-1}(x)}{k!} H_k(y), \tag{5}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote respectively the distribution and density function of a standard normal variable, and $H_k$ is the Hermite polynomial of order $k$. These polynomials satisfy the three terms recurrent formula

$$\begin{aligned} H_k(x) &= x H_{k-1}(x) - (k-1) H_{k-2}(x), \quad k \geq 1 \\ H_{-1}(x) &= 0, \qquad H_0(x) = 1, \end{aligned}$$

see Chihara (1978). Substituting $y$ by $Z_n$, in (5) we obtain

$$I_{(-\infty,x]}(Z_n) = \Phi(x) - \phi(x) \sum_{k=1}^{\infty} \frac{H_{k-1}(x)}{k!} H_k(Z_n). \tag{6}$$

Taking expected values with respect to $F$ in (6) and truncating the series we get

$$\begin{aligned} P(Z_n \leq x) \simeq \Phi(x) - \phi(x) \Bigg[ & \frac{k_3}{3!\sqrt{n}} H_2(x) + \frac{k_4}{4!n} H_3(x) \\ & + \frac{k_5}{5!n^{3/2}} H_4(x) + \frac{1}{6!} \left( \frac{k_6}{n^2} + \frac{10k_3^2}{n} \right) H_5(x) \Bigg] + \cdots . \end{aligned} \tag{7}$$

The right hand side of (7) is the formal Gram-Charlier Type A expansion for the distribution function of the normalized sum of n iid random variables having a common cdf $F$ (Cramér (1946)), and after reordering (7) in ascending powers of $1/\sqrt{n}$ we obtain the Edgeworth expansion for the distribution of the standardized sample mean, given by

$$\begin{aligned} P(Z_n \leq x) \simeq \Phi(x) \\ -\phi(x) \left[ \frac{k_3 H_2(x)}{3!\sqrt{n}} + \left( \frac{k_4 H_3(x)}{4!} + \frac{10k_3^2 H_5(x)}{6!} \right) \frac{1}{n} \right] + \cdots . \end{aligned} \tag{8}$$

## 3 Approximation to the Crámer-von Mises distribution

Consider the Cramér-von Mises statistic

$$W_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \, dF(x),$$

where $F_n(x)$ is the empirical distribution function of a random sample of size $n$ from a population with distribution function $F$ assumed to be a continuous function on the whole line. We refer to Csörgő and Faraway (1996) for the historical exposition of this goodness-of-fit statistic along with the main mathematical results and a review of the relevant literature with many corrections. Here we list only those properties of this statistic and of its asymptotic distribution, denoted $W^2_\infty$, which are necessary to get the approximation given in (4).

– Anderson and Darling (1952) obtained one expression for the distribution of $W^2_\infty$ from the solution of a certain differential equation:

$$F_{W^2_\infty}(x) = \frac{1}{\pi^{3/2}x^{1/2}} \sum_{k=0}^{\infty} \frac{\Gamma(k+1/2)}{k!} (4k+1)^{1/2} \times \exp\left\{-\frac{(4k+1)^2}{16x}\right\} K_{1/4}\left\{\frac{(4k+1)^2}{16x}\right\}, x > 0, \tag{9}$$

where $K_\nu(\cdot)$, $\nu > -1/2$, is the modified Bessel function of the second kind.

– The cumulants of $W^2_\infty$, say $k_s$, verify the following relation, see Pearson and Stephens (1962):

$$k_s = 2^{s-1} \frac{(s-1)!}{\pi^{2s}} \sum_{j=1}^{\infty} \frac{1}{j^{2s}} \tag{10}$$

– Therefore the moments of all orders exist and the first eight moments of $W^2_\infty$ obtained from (10) are:

$$\begin{aligned} &m_1 = 0.16666666666,\ m_2 = 0.05,\ &&m_3 = 0.02420634921,\\ &m_4 = 0.01667989418,\ m_5 = 0.01496212121,\ &&m_6 = 0.01651200982,\\ &m_7 = 0.02160988831,\ m_8 = 0.03269556345. \end{aligned}$$

– The first four moments of $W^2_n$ obtained from the central moments given in Pearson and Stephens (1962):

$$\begin{aligned} m_{n,1} &= \frac{1}{6},\\ m_{n,2} &= \frac{1}{20} - \frac{1}{60n},\\ m_{n,3} &= \frac{61}{2520} - \frac{37}{1512n} + \frac{1}{126n^2},\\ m_{n,4} &= \frac{1261}{75600} - \frac{3833}{113400n} + \frac{2071}{75600n^2} - \frac{1}{120n^3}. \end{aligned}$$

We can determine a system of orthogonal polynomials with respect to the density of asymptotic Cramér-von Mises statistic, which is obtained from (9), using the

method given in Chihara (1978), page 17, Exercise 3.1. So, the first four orthogonal polynomials obtained from the first eight moments of $W_\infty^2$, are

$$\begin{aligned}
\psi_1(x) &= x - 0.1666666666,\\
\psi_2(x) &= x^2 - 0.7142857143x + 0.06904761905,\\
\psi_3(x) &= x^3 - 1.660541586x^2 + 0.6170535139x - 0.04402152221,\\
\psi_4(x) &= x^4 - 3.009453709x^3 + 2.542564319x^2 - 0.6517911511x +\\
&\quad +0.03767163579.
\end{aligned}$$

We also need the norm of the above polynomials:

$$\begin{aligned}
\|\psi_1\|_2^2 &= 0.022222222222,\\
\|\psi_2\|_2^2 &= 0.002842025699,\\
\|\psi_3\|_2^2 &= 0.0008935723022,\\
\|\psi_4\|_2^2 &= 0.0005206325960.
\end{aligned}$$

Let $E_{n,k} = E\left[\psi_k\left(W_n^2\right)\right]$, then from the expressions given for the polynomials and for the moments of $W_n^2$, we have:

$$\begin{aligned}
E_{n,1} &= 0,\\
E_{n,2} &= -0.016666666666\frac{1}{n},\\
E_{n,3} &= 0.003204793630\frac{1}{n} + 0.007936507937\frac{1}{n^2},\\
E_{n,4} &= -0.002532738271\frac{1}{n} + 0.003509626646\frac{1}{n^2} - 0.008333333333\frac{1}{n^3}.
\end{aligned}$$

If we denote $F_{W_n^2}$ as the distribution function of the Cramér-von Mises statistic we have the following approximation in terms of its asymptotic distribution:

$$F_{W_n^2}(x) \simeq F_{W_\infty^2}(x) + \sum_{k=1}^{4} \alpha_k(x)\, E_{n,k}, \tag{11}$$

with

$$\alpha_k(x) = \frac{\int_{-\infty}^{x} \psi_k(y)\, dF_{W_\infty^2}(y)}{\|\psi_k\|_2^2}. \tag{12}$$

We note that the coefficients (12) are easily calculated on a computer, so we have the necessary elements to get the approximation (11).

In Table 1 we show some comparisons with the approximation given by Csörgő and Faraway (1996) and the values considered as exact obtained by Csörgő and Faraway (1996), a part of which is from Knott (1974). As we can notice we have obtained results slightly more accurate than those given by Csörgő and Faraway (1996). This is remarkable since their results rest on a fully rigorous mathematical approximation, firmly developed by Götze (1979).

**Table 1** Approximations to the Cramér-von Mises distribution

| $n$ | $x$ | $F_{W_n^2}(x)$ | Approx. (11) | Approx. Csörgő and Faraway (1996) |
|---|---|---|---|---|
| 2 | 0.48901 | 0.975 | 0.9769954760 | 0.9681751969 |
| 2 | 0.55058 | 0.99 | 0.9881409934 | 0.9808432316 |
| 3 | 0.8224 | 0.999 | 0.9981970834 | 0.9978556645 |
| 3 | 0.6398 | 0.99 | 0.9910413971 | 0.9884969825 |
| 6 | 0.69443 | 0.99 | 0.9904452092 | 0.9885935585 |
| 7 | 0.34397 | 0.9 | 0.9002770924 | 0.8998157898 |
| 8 | 0.7072 | 0.99 | 0.9903136860 | 0.9898779312 |
| 10 | 0.3450 | 0.9 | 0.9001813058 | 0.8999070681 |

## 4 Approximation to the Watson distribution

The Watson statistic is a modification of the Cramér-von Mises statistic, given by

$$U_n^2 = n \int_{-\infty}^{\infty} \left( F_n(x) - F(x) - \int_{-\infty}^{\infty} (F_n(x) - F(x))\, dF(x) \right)^2 dF(x),$$

where $F$ is continuous as before. We can observe that $U_n^2$ has the form of a variance while the Cramér-von Mises statistic has the form of a second moment about the origin; in this sense the modification corresponds to a correction for the mean. This makes $U_n^2$ rotationally invariant, when it is adapted for testing goodness of fit on the unit circunference of a circle. Again we refer to Csörgő and Faraway (1996) for references and further discussion, and cite here only results that we need for $U_n^2$ and its asymptotic form, $U_\infty^2$:

– The distribution function of $U_\infty^2$, $F_{U_\infty^2}$, has the following expression, see Watson (1961):

$$F_{U_\infty^2}(u) = 1 - \sum_{k=1}^{\infty} (-1)^{k-1} 2e^{-2k^2\pi^2 u}, \quad u > 0.$$

Hence the density function of $U_\infty^2$, $\phi$, is

$$\phi(u) = \sum_{k=1}^{\infty} (-1)^{k-1} 4k^2\pi^2 e^{-2k^2\pi^2 u}, \quad u > 0. \tag{13}$$

– The moment generating function of $U_\infty^2$, also given by Watson (1961) is

$$E\left(e^{\theta U_\infty^2}\right) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{2}{1 - \left(\theta/2k^2\pi^2\right)},$$

from which we can obtain an expression for the $j$-th moment, $m_j$ of $U_\infty^2$:

$$m_j = \frac{j!}{2^{j-1}\pi^{2j}} \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i^{2j}}. \tag{14}$$

– The first four moments of $U_n^2$ obtained from the central moments given in Stephens (1963) are

$$
\begin{aligned}
m_{n,1} &= \frac{1}{12}, \\
m_{n,2} &= \frac{7}{720} - \frac{1}{360n}, \\
m_{n,3} &= \frac{31}{20160} - \frac{41}{30240n} + \frac{1}{2520n^2}, \\
m_{n,4} &= \frac{127}{403200} - \frac{103}{181440n} + \frac{127}{302400n^2} - \frac{1}{8400n^3}.
\end{aligned}
$$

Similarly as for the Cramér-von Mises statistic, we can determine a system of orthogonal polynomials with respect to the density (13). Following similar arguments as in section 3, we have the expression for the first four orthogonal polynomials:

$$
\begin{aligned}
\psi_1(x) &= x - 0.0833333333, \\
\psi_2(x) &= x^2 - 0.2619047619x + 0.01210317460, \\
\psi_3(x) &= x^3 - 0.5389952153x^2 + 0.07390350877x - 0.002456092884, \\
\psi_4(x) &= x^4 - 0.9157528834x^3 + 0.2531110208x^2 - 0.02404995526x + \\
&\quad +0.0006365329436,
\end{aligned}
$$

with respective norms

$$
\begin{aligned}
\|\psi_1\|_2^2 &= 0.00277777777777, \\
\|\psi_2\|_2^2 &= 0.00002991937516, \\
\|\psi_3\|_2^2 &= 0.7139680969 \cdot 10^{-6}, \\
\|\psi_4\|_2^2 &= 0.3003525491 \cdot 10^{-7}.
\end{aligned}
$$

Letting $E_{n,k} = E\left[\psi_k\left(U_n^2\right)\right]$, from the expressions given for the polynomials and for the moments of $U_n^2$, we obtain:

$$
\begin{aligned}
E_{n,1} &= 0, \\
E_{n,2} &= -0.0027777777777\frac{1}{n}, \\
E_{n,3} &= 0.0001413888256\frac{1}{n} + 0.0003968253968\frac{1}{n^2}, \\
E_{n,4} &= -0.00002917077354\frac{1}{n} + 0.00005657954363\frac{1}{n^2} - \\
&\quad -0.0001190476190\frac{1}{n^3}.
\end{aligned}
$$

So if $F_{U_n^2}$ is the distribution function of the Watson statistic, we get:

$$
F_{U_n^2}(x) \simeq F_{U_\infty^2}(x) + \sum_{k=1}^{4} \alpha_k(x) E_{n,k}, \tag{15}
$$

**Table 2** Approximations to the Watson distribution

| $n$ | $x$ | $F_{U_n^2}(x)$ | Approx. (15) | Approx. Csörgő and Faraway (1996) |
|---|---|---|---|---|
| 6 | 0.2087 | 0.975 | 0.9745433154 | 0.9742627229 |
| 6 | 0.2450 | 0.99 | 0.9900560198 | 0.9894690391 |
| 8 | 0.2121 | 0.975 | 0.9748605892 | 0.9745985778 |
| 8 | 0.2513 | 0.99 | 0.9900534859 | 0.9897855525 |
| 9 | 0.2532 | 0.99 | 0.9900191083 | 0.9898242904 |
| 10 | 0.2548 | 0.99 | 0.9900143789 | 0.9898700734 |
| 10 | 0.1164 | 0.80 | 0.8001599380 | 0.8001907288 |

where

$$\alpha_k(x) = \frac{\int_{-\infty}^{x} \psi_k(y)\,\phi(y)\,dy}{\|\psi_k\|_2^2}. \tag{16}$$

In Table 2 we compare our approximation with those given by Csörgő and Faraway (1996) and the values considered as exact obtained from simulation. As we can notice our heuristic approximation seems to produce results slightly closer to the exact values than those given by Csörgő and Faraway (1996). Again, this is remarkable since their approximation has a sound mathematical basis, rooted in Götze (1979).

**Acknowledgements** The authors are thankful to the referees for their suggestions and helpful comments. This research was partially supported by grants FQM-331, FQM-270, MTM 2004-0909 and BMF2002-04525-C02-02.

## References

Anderson TW, Darling DA (1952) Asymptotic theory of certain goodness of fit criteria based on stochastic processes. Ann Math Stat 23:193–212

Chihara T (1978) An introduction to orthogonal polynomials. Gordon-Breach, New York

Cramér H (1946) Mathematical methods of statistics. Princeton University Press, Princeton

Csörgő S, Faraway JJ (1996) The exact and asymptotic distributions of Cramér-von Mises Statistics. J R Stat Soc B 58:221–234

Götze F (1979) Asymptotic expansions for bivariate von Mises functionals. Zeitschrift Wahrsch Ver Geb 50:333–355

Knott M (1974) The distribution of the Cramér-von Mises statistic for small sample sizes. JR Stat Soc B 36:430–438

Pearson ES, Stephens MA (1962) The goodness-of-fit tests based on $W_n^2$ and $U_n^2$. Biometrika 49:397–402

Stephens MA (1963) The distribution of the goodness-of-fit statistic $U_n^2$ : I. Biometrika 50: 303–313

Wallace DL (1958) Asymptotic approximations to distributions. Ann Math Stat 29:635–654

Watson GS (1961) Goodness-of-fit tests on a circle Biometrika 48:109–114