This Maple worksheet accompanies the paper:

Di Nardo E. (2010) *A new approach to Sheppard's corrections.* In press Mathematical Methods in Statistics. (http://arxiv.org/abs/1004.4989)

A new approach to Sheppard's corrections

$$a_n = \sum_{j=0}^{n} \binom{n}{j} \left(2^{1-j} - 1\right) B_j \, h^j \, \tilde{a}_{n-j}$$

Sheppard's corrections

E. Di Nardo*
elvira.dinardo@unibas.it
http://www.unibas.it/utenti/dinardo/home.html;
Tel: +39 0971205890, Fax: +39 0971205896

G. Guarino**
giuseppe.guarino@rete.basilicata.it

* Dipartimento di Matematica e Informatica, Università degli Studi della Basilicata,
Viale dell'Ateneo Lucano n.10, 85100 Potenza, Italy

**Medical School, Università del Sacro Cuore (Rome branch),
Largo Agostino Gemelli n.8, 00168 Roma, Italy

Introduction

***Abstract*:** in the real world, continuous variables are observed and recorded in finite precision through a rounding or coarsening operation, i.e. a grouping rule. A compromise between the desire to know and the cost of knowing is then a necessary consequence.
Attention has been paid in the literature to the computation of moments when data are grouped into classes. The moments computed by means of the resulting grouped frequency distribution are looked upon as a first approximation to the moments of the parent distribution, but they suffer from the error committed in grouping. A good correction procedure is given by Sheppard's corrections that are nowadays still employed. Sheppard's corrections are usually referred to continuos parent distribution. But grouping includes also censoring or splitting data into categories during collection or publication,

and so it does not only involve continuous variables.
A very simple closed-form formula for Sheppard's corrections has been recovered by the classical umbral calculus (see [5]) as well as a more general closed-form formula for discrete parent distributions (see [2]). No attention was paid in the literature to multivariate generalizations of Sheppard's corrections, probably due to the complexity of the resulting formulae (see [1]). Via the umbral calculus, the generalization to the multivariate case turns to be straightforward.
All these new formulae are particularly suited to be implemented in MAPLE. The theoretical background of these formulae can be found in Di Nardo E. (2010) (see [3])

Application Areas/Subject: combinatorics & algebraic methods in statistics.
Keywords: raw moment, grouped moment, Sheppard's correction, umbral calculus.
See Also: background on umbral calculus in [4]

Initialization

> *restart*

raw2grp

Suppose $X = (X_1, X_2, .., X_j)$ a multivariate random vector.
The raw multivariate moment of X of order $t_1, .., t_j$ is denoted by $r_{t_1, .., t_j}$.
The moments calculated from the grouped frequencies are denoted by $g_{t_1, .., t_j}$.
Assume $h_1, .., h_j$ are not-zero width window for each component and $m_1, .., m_j$ the numbers of consecutive values grouped in a frequency class of width h_k.
The procedure ***raw2grp*** gives raw moments $r_{t_1, .., t_j}$ in terms of grouped moments $g_{t_1, .., t_j}$, by using formula (31) of the paper [3].

In particular, set the variable t = 0 when Sheppard's corrections are required for continuous parent distribution.

Note:
that sequence f1 in the procedure refers to formula (14) and sequence f2 refers to formula (17).

```
> raw2grp := proc(V, t)
      local n, M, eF1, eF2;
      n := nops(V);
      M := expand(mul((mu[i] + h[i]*f1[i] + `if`(t = 0, 0, h[i]*f2[i]/m[i]))^V[i], i = 1 .. n));
      M := add(x*g[seq(degree(x, mu[i]), i = 1 .. n)], x = M);
      M := eval(M, [seq(mu[i] = 1, i = 1 .. n)]);
      eF1 := seq(seq(f1[i]^j = (2^(1 - j) - 1)*bernoulli(j), j = 1 .. max(op(V))), i = 1 .. n);
      if t = 0 then
```

```
    eF2 := NULL
  else
    eF2 := seq( seq( f2[i]^j = `if`( irem(j, 2) = 1, 0, (1/2)^j / (j + 1) ), j = 1 ..max(op(V)) ), i = 1 ..n )
  end if;
  eval(M, [eF1, eF2, g0$n = 1])
end proc:
```

>

Examples

continuous parent distributions

```
> raw2grp([2, 2], 0);
```

$$g_{2,2} - \frac{1}{12} h_2^2 g_{2,0} - \frac{1}{12} h_1^2 g_{0,2} + \frac{1}{144} h_1^2 h_2^2 \qquad \textbf{(3.1.1)}$$

discrete parent distributions

```
> raw2grp([2, 2], 1);
```

$$\frac{1}{12} \frac{h_1^2 g_{0,2}}{m_1^2} + g_{2,2} - \frac{1}{144} \frac{h_1^2 h_2^2}{m_1^2} - \frac{1}{12} h_1^2 g_{0,2} + \frac{1}{144} h_1^2 h_2^2 - \frac{1}{12} h_2^2 g_{2,0} \qquad \textbf{(3.1.2)}$$

$$- \frac{1}{144} \frac{h_1^2 h_2^2}{m_2^2} + \frac{1}{12} \frac{h_2^2 g_{2,0}}{m_2^2} + \frac{1}{144} \frac{h_1^2 h_2^2}{m_1^2 m_2^2}$$

>

grp2raw

Suppose $X = (X_1, X_2, .., X_j)$ a multivariate random vector.
The raw multivariate moment of X of order $t_1, .., t_j$ is denoted by $r_{t_1, .., t_j}$.
The moments calculated from the grouped frequencies are denoted by $g_{t_1, .., t_j}$.
Assume $h_1, .., h_j$ are not-zero width window for each component and $m_1, .., m_j$ the number of consecutive values grouped in a frequency class of width h_k.
The procedure ***grp2raw*** gives grouped moments $g_{t_1, .., t_j}$ in terms of raw moments $r_{t_1, .., t_j}$, by using formula (32) of the paper [3].

In particular, set the variable t = 0 when Sheppard's corrections are required for continuous parent distribution.

Note:
that sequence f1 in the procedure refers to formula (14) and sequence f2 refers to formula (17).

```
> grp2raw := proc(V, t)
    local n, M, eF1, eF2;
```

$n := nops(V);$

$M := expand\left(mul\left(\left(\mu_i + h_i \cdot \mathrm{f1}_i + \grave{i}f\hat{}\left(t = 0,\ 0,\ \frac{h_i \cdot \mathrm{f1}_i}{m_i}\right)\right)^{V_i},\ i = 1..n\right)\right);$

$M := add\left(\mathrm{x} \cdot \mathrm{r}_{seq\left(degree\left(x, \mu_i\right),\ i = 1..n\right)},\ x = M\right);$

$M := eval\left(M,\ \left[seq\left(\mu_i = 1,\ i = 1..n\right)\right]\right);$

if $t = 0$ **then**

$eF1 := NULL$

else $eF1 := seq\left(seq\left(f1_i^j = \left(2^{1-j} - 1\right) \cdot \mathrm{bernoulli}(j),\ j = 1..\max(op(V))\right),\ i = 1..n\right);$

end if;

$eF2 := seq\left(seq\left(f2_i^j = \grave{i}f\hat{}\left(irem(j, 2) = 1,\ 0,\ \frac{\left(\frac{1}{2}\right)^j}{j+1}\right),\ j = 1..\max(op(V))\right),\ i = 1..n\right);$

$eval\left(M,\ \left[eF1, eF2, r_{0\$n} = 1\right]\right)$

end proc:

>

Examples

continuous parent distributions

> $grp2raw([2, 2], 0);$

$$r_{2,2} + \frac{1}{12} h_2^2 r_{2,0} + \frac{1}{12} h_1^2 r_{0,2} + \frac{1}{144} h_1^2 h_2^2 \qquad \textbf{(4.1.1)}$$

discrete parent distributions

> $grp2raw([2, 2], 1);$

$$-\frac{1}{144}\frac{h_1^2 h_2^2}{m_1^2} - \frac{1}{12}\frac{h_1^2 r_{0,2}}{m_1^2} - \frac{1}{12}\frac{h_2^2 r_{2,0}}{m_2^2} + \frac{1}{144} h_1^2 h_2^2 + \frac{1}{12} h_2^2 r_{2,0} + \frac{1}{12} h_1^2 r_{0,2} \qquad \textbf{(4.1.2)}$$

$$-\frac{1}{144}\frac{h_1^2 h_2^2}{m_2^2} + r_{2,2} + \frac{1}{144}\frac{h_1^2 h_2^2}{m_1^2 m_2^2}$$

Tests

The procedure ***raw2grp*** gives raw moments $r_{t_1,..,t_j}$ in terms of grouped moments $g_{t_1,..,t_j}$.

If the output is evaluated using $\boldsymbol{g_{t_1,..,t_j}} = \boldsymbol{grp2raw}([t_1,..,t_j])$ you obtain the raw moment again.

continuous parent distributions

> $r2g := raw2grp([2, 2], 0);$

$$r2g := g_{2,2} - \frac{1}{12} h_2^2 g_{2,0} - \frac{1}{12} h_1^2 g_{0,2} + \frac{1}{144} h_1^2 h_2^2 \qquad \textbf{(5.1)}$$

> $expand\left(eval\left(r2g,\ \left[g_{2,2} = grp2raw([2, 2], 0),\right.\right.\right.$

$$g_{2,0} = grp2raw([2,0],0),\quad g_{0,2} = grp2raw([0,2],0)]));$$

$$r_{2,2} \tag{5.2}$$

discrete parent distributions

> $r2g := raw2grp([2,2],1);$

$$r2g := g_{2,2} + \frac{1}{12}\frac{h_2^2 g_{2,0}}{m_2^2} - \frac{1}{144}\frac{h_1^2 h_2^2}{m_1^2} - \frac{1}{12}h_1^2 g_{0,2} + \frac{1}{144}h_1^2 h_2^2 - \frac{1}{12}h_2^2 g_{2,0} \tag{5.3}$$

$$- \frac{1}{144}\frac{h_1^2 h_2^2}{m_2^2} + \frac{1}{12}\frac{h_1^2 g_{0,2}}{m_1^2} + \frac{1}{144}\frac{h_1^2 h_2^2}{m_1^2 m_2^2}$$

> $expand(eval(r2g, [g_{2,2} = grp2raw([2,2],1), g_{2,0} = grp2raw([2,0],1), g_{0,2} = grp2raw([0,2],1)]));$

$$r_{2,2} \tag{5.4}$$

>

The procedure ***grp2grp*** gives grouped moments $g_{t_1,..,t_j}$ in terms of raw moments $r_{t_1,..,t_j}$.
If the output is evaluated using $\boldsymbol{r_{t_1,..,t_j} = raw2grp([t_1,..,t_j])}$ you obtain the grouped moments again.

continuous parent distributions

> $g2r := grp2raw([2,2],0);$

$$g2r := r_{2,2} + \frac{1}{12}h_2^2 r_{2,0} + \frac{1}{12}h_1^2 r_{0,2} + \frac{1}{144}h_1^2 h_2^2 \tag{5.5}$$

> $expand(eval(g2r, [r_{2,2} = raw2grp([2,2],0), r_{2,0} = raw2grp([2,0],0), r_{0,2} = raw2grp([0,2],0)]));$

$$g_{2,2} \tag{5.6}$$

discrete parent distributions

> $g2r := grp2raw([2,2],1);$

$$g2r := -\frac{1}{144}\frac{h_1^2 h_2^2}{m_1^2} + \frac{1}{144}h_1^2 h_2^2 + \frac{1}{12}h_2^2 r_{2,0} + \frac{1}{12}h_1^2 r_{0,2} - \frac{1}{144}\frac{h_1^2 h_2^2}{m_2^2} + r_{2,2} \tag{5.7}$$

$$- \frac{1}{12}\frac{h_1^2 r_{0,2}}{m_1^2} + \frac{1}{144}\frac{h_1^2 h_2^2}{m_1^2 m_2^2} - \frac{1}{12}\frac{h_2^2 r_{2,0}}{m_2^2}$$

> $expand(eval(g2r, [r_{2,2} = raw2grp([2,2],1), r_{2,0} = raw2grp([2,0],1), r_{0,2} = raw2grp([0,2],1)]));$

(5.8)

$g_{2,\,2}$ **(5.8)**

>

Conclusions

We have shown how the corrections of moments resulting from grouping into classes may be summarized in few closed-form formulae.

Once more, this algorithm shows how the classical umbral calculus should be taken into account for managing sequence of numbers related to random variables, since many calculations are reduced. For example, the reader interested in recovering corrections for cumulants and factorial moments, by using the classical umbral calculus, can refer to [4].

References

[1] Baten, W.D. (1931) Correction for the Moments of a Frequency Distribution in Two Variables. Ann. Math. Stat 2, No. 3, 309-319.

[2] Craig, C.C. (1936) Sheppard's corrections for a discrete variable. Ann.Math. Stat 7, No. 2, 55-61.

[3] Di Nardo E. (2010) A new approach to Sheppard's corrections. Math. Meth. Stat. in press. (http://arxiv.org/abs/1004.4989)

[4] Di Nardo, E., Senato, D. (2006) An umbral setting for cumulants and factorial moments. European J. Combin. 27, No. 3, 394–413. (http://www.arxiv.org/abs/math/0412052)

[5] Di Nardo, E., Guarino, G., Senato, D. (2008) A unifying framework for k-statistics, polykays and their multivariate generalizations. Bernoulli. 14, No. 2, 440–468. (http://www.unibas.it/utenti/dinardo/BEJ6163290408.pdf)