HUNTING MULTIPLE QUANTA BY SELECTIVE LEAST SQUARES

SEPPO MUSTONEN

1. Introduction

In archeological data related to various measures of objects like

– lengths of building blocks of an ancient building,

– worn balance weights used in weighing of commodities,

it may be interesting to search for a reasonable basic unit (length or weight) giving the *observations as integer multiples of the basic unit.* The problem becomes even trickier when there may be *two or more incommensurable basic units* in the same set of observations.

I encountered this problem in 2005 when discussing with Jari Pakkanen who is a specialist in Classical Archealogy and making architectural reconstructions using statistical methods and computer simulations.

For his needs, I immediately implemented the 'cosine quantogram' of D.G.Kendall (1974) [2] as a new Survo program module QUANTA. In this connection, I also came across an idea of using a more straightforward but computationally harder method (SLS) based on selective, conditional least squares.

According to my simulation experiments, SLS seems to work better than Kendall's method especially when several quanta exist.

2. QUANTA program

QUANTA is a Survo [3] program module. The command

```
QUANTA <data>,<variable>,k,L
```

estimates k quanta from the given data set of one variable.

Consider a data set $X_1, X_2, \dots, X_n$ where each observation is an approximate integral multiple of one of positive numbers $q_1, q_2, \dots, q_k$ where typically $k = 1$ or another small integer.

Our task is to estimate the values of quanta $q_1, q_2, \dots, q_k$ on the condition that each of them exceeds a certain minimum value $q_{\min}$.

Kendall proposed using a "cosine quantogram" of the form

$$\phi(q) = \sqrt{(}2/n) \sum_{i=1}^{n} \cos 2\pi\epsilon(i)/q$$

where $0 \le \epsilon(i) < q$ is the remainder when X_i is divided by q. The q-values of highest upward peaks of this function will be considered as candidates for quanta. In Kendall's paper only the case $k = 1$ was studied.

Date: 8 March 2012.

My idea is that the quanta are estimated by a selective, conditional least squares method (SLS) where the sum

$$ss(q_1, \dots, q_k) = \sum_{i=1}^{n} \min[g(X_i, q_1)^2, \dots, g(X_i, q_k)^2]$$

where $g(x, q)$ in the least absolute remainder when x is divided by q, is to be minimized with respect of $q_1, \dots, q_k$ on the condition that each q_i is at least $q_{\min}$. A similar approach was presented by Broadbent [1] (already in 1956) in the case $k = 1$ in a form of ordinary weighted least squares.

When `QUANTA` is activated, the permitted range of quanta is given by a specification

```
RANGE=<lower_limit>(<step>)<upper_limit>
```

and least possible value of a quantum is given by

```
Q_MIN=<q_min>.
```

The default method is SLS and then all combinations of values given by `RANGE` are used as starting values for $q_1, \dots, q_k$ for a minimization process of $ss(q_1, \dots, q_k)$ and it is performed by Powell's conjugate gradient method. By

```
RES=<quantum_number>,<residual>,<coeff>
```

three extra variables are given for saving corresponding information about the results for each case.

3. Experimental studies

The Kendall's and SLS methods will now be compared by simple Monte Carlo experiments using artificial data sets. All these experiments have been carried out by the Survo software [3] and presented here as Survo applications.

3.1. Example 1: One quantum. Analyzing of a dataset consisting of plain multiples of number 3 leads to following results:

```
____________________________________________________________________
DATA X: 9 12 18 21 24 27 33 36 END

RANGE=2(0.2)5 Q_MIN=2
QUANTA X,X,2,CUR+1
Data: X Variable: X  N=8
ss=0
     quantum     # matches
 1   3.000000         8
.....................................................................
METHOD=Kendall RANGE=2(0.0001)5 SCORE_MIN=2
QUANTA X,X,0,CUR+1
Data: X Variable: X  N=8
GPLOT COSQUANT,quant,score / LINE=1 MODE=SVGA Plot the quantogram!
Peaks of Kendall's Cosine Quantogram:
 quantum      score
3.000000   4.000000
____________________________________________________________________
```

Here only numerical results are presented and both methods give the correct result as expected.

When the 'exact' data values $X_j, j = 1, 2, \ldots, 8$ were replaced by noisy values $Y_j = X_j + \epsilon_j$ where errors ϵ_j are independent N(0,0.05) variables, a 1000-fold Monte Carlo experiment gave similar results for both methods. Estimates of q were unbiased with MSE= 0.000005. [1]

3.2. **Example 2: Two quanta.** In the next dataset both multiples of 3 and 7 appear. The SLS method recognizes them exactly as quanta:

```
____________________________________________________________________
DATA X: 9 12 14 18 21 24 27 28 33 35 36 END

RANGE=4(0.2)8 Q_MIN=2
QUANTA X,X,2,CUR+1
Data: X Variable: X  N=11
ss=0
     quantum     # matches
 1   3.000000         8
 2   7.000000         3
____________________________________________________________________
```

while Kendall's method does not succeed as well:

```
____________________________________________________________________
....................................................................
METHOD=Kendall RANGE=2(0.0001)8 SCORE_MIN=1
QUANTA X,X,0,CUR+1
Data: X Variable: X  N=11
GPLOT COSQUANT,quant,score / LINE=1 MODE=SVGA Plot the quantogram!
Peaks of Kendall's Cosine Quantogram:
 quantum      score
2.992800   2.791167
2.351300   1.721541   (3*2.3513=7.0539)
4.540300   1.388202
6.930200   1.233336
____________________________________________________________________
```

The first quantum (2.9928) corresponds to 3 decently but the second quantum 7 is found rather inaccurately (6.9302) as weakest of the four first peaks. A better candidate is the second one (2.3513) since $3 \times 2.3513 = 7.0539$.

Already this simple example indicates that Kendall's method leads to biased results when two or more incommensurable quanta really exist. One has to remember that Kendall obviously concentrated on the single quantum case only.

When a similar Monte Carlo experiment, with 1000 replicates and the same additive error term as in Example 1, is performed, the following results are obtained.

It turns out that in majority of replicates the results will have the same structure as in the previous 'ideal' case, but results of Kendall's method deviate from this basic setup in 7 cases (of 1000) so that the quantum corresponding to the 'true' value $7/3 = 2.333\ldots$ does not appear among the two highest peaks.

For the remaining 993 replicates the following statistics are obtained:

[1] Details of the Monte Carlo technique used in this study are described in Section 4.

```
SLS:
quantum  mean     std.dev.
3        2.99993  0.002311
7        7.00011  0.008037

Kendall:
quantum  mean     std.dev.
3        2.99276  0.002879
7/3      2.35110  0.003660
```

Thus especially the second quantum when calculated according to Kendall's method has a strong bias.

3.3. **Example 2': Two quanta close to each other.** In the next case where the quanta are 1.1 and 1.2 the results for 'pure' data are

```
--------------------------------------------------------------------
DATA X: 5.5 6.0 6.6 7.2 7.7 8.4 8.8 9.6 END

RANGE=1(0.1)2 Q_MIN=1
QUANTA X,X,2,CUR+1
Data: X Variable: X  N=8
ss=1.57772e-030
     quantum     # matches
 1   1.200000        4
 2   1.100000        4
.....................................................................
METHOD=Kendall RANGE=1(0.0001)2 SCORE_MIN=0.5
QUANTA X,X,0,CUR+1
Data: X Variable: X  N=8
GPLOT COSQUANT,quant,score / LINE=1 MODE=SVGA Plot the quantogram!
Peaks of Kendall's Cosine Quantogram:
 quantum      score
1.077000   1.204011
1.412500   0.777207
1.218500   0.564343
--------------------------------------------------------------------
```

Kendall's method gives still weaker results. There appears a spurious peak 1.4125 in the second place and both real quanta get biased estimates 1.077 and 1.2185.

In a Monte Carlo experiment (similar to that of Example 2) it turns out that in only 199 replicates (of 1000) estimates of true quanta 1.1 and 1.2 are obtained as the two strongest peaks by Kendall's method.

Summary of results:

```
SLS: N=1000
quantum  mean     std.dev.
1.1      1.10003    0.00381
1.2      1.19995    0.00381

Kendall: N=199
quantum  mean     std.dev.
1.1      1.07376    0.00329
1.2      1.22575    0.00462
```

3.4. **Example 3: Broadbent** [1]**.** As an example of his weighted least squares method (restricted to one quantum) Broadbent examined certain data of excitation energies of nuclei. The SLS method (unweighted least squares) seems to give same results. As an example, the last dataset of 12 observations is analyzed by `QUANTA`.

```
____________________________________________________________________
*DATA Grant12 A,B,N,M
NX       quantum  Res   coeff
M1.111    12     12.123   123
*
A0.608     1     -0.004    10
*1.283     1     -0.003    21
*1.412     1      0.004    23
*1.663     1      0.010    27
*1.844     1      0.007    30
*2.015     1     -0.006    33
*2.138     1     -0.005    35
*2.268     1      0.002    37
*2.439     1     -0.011    40
*2.513     1      0.002    41
*2.697     1      0.003    44
B2.880     1      0.002    47
*
*RES=quantum,Res,coeff
*RANGE=0.1(0.001)1 Q_MIN=0.05
*QUANTA Grant12,X,1,CUR+1
*Data: Grant12 Variable: X  N=12
*ss=0.000372192
*     quantum     # matches
* 1   0.0612386        12
*
____________________________________________________________________
```

Here the results after the `QUANTA` command has been activated are displayed. Broadbent got the value 0.06124 for the quantum so that the results are identical. Also Kendall's method gives the same result.

4. Appendix

All calculations have been performed as Survo applications. The following extract from a Survo edit field gives a detailed account on Example 2.

The display has been made more legible so that activated commands are shown in red and results of them in green. Text written by the user is black. Normally all text by default is black.

```
____________________________________________________________________
*SAVE EX2 / Two quanta
*
*DATA X: 9 12 14 18 21 24 27 28 33 35 36 END
*
*RANGE=4(0.2)8 Q_MIN=2
*QUANTA X,X,2,CUR+1
*Data: X Variable: X  N=11
*ss=0
*     quantum     # matches
* 1   3.000000         8
* 2   7.000000         3
*....................................................................
*METHOD=Kendall RANGE=2(0.0001)8 SCORE_MIN=1
*QUANTA X,X,0,CUR+1
*Data: X Variable: X  N=11
*GPLOT COSQUANT,quantum,score / LINE=1 MODE=SVGA Plot the quantogram!
*Peaks of Kendall's Cosine Quantogram:
* quantum      score
*2.992800   2.791167
*2.351300   1.721541
*4.540300   1.388202
*6.930200   1.233336
*....................................................................
*
*Creating a data file Q2 with more variables:
*FILE CREATE Q2
*FIELDS:
*1 NA-  8 X           data value
*2 NA-  2 q_index     index of quantum: 1 or 2
*3 NA-  8 Res         residual: Y-coeff*quantum_value
*4 NA-  2 coeff       multiplier of quantum
*5 NA-  8 Y           X+0.05*N.G(0,1,rand(seed))
*END
*
*Copying X values to the data file:
*FILE COPY X TO Q2
*.........................................
*The Monte Carlo experiment is carried out by a Survo macro /EX2.
*/EX2 operates in this case by using a template consisting of
*readily typed text and commands and repeats the set of operations
*until the user interrupts the process.
```

```
*The sucro code has been generated by using Survo in a so-called
*tutorial mode enabling saving of all actions of the user in a sucro
*file.
*
*Do not change lines below (until X). Sucro /EX2 is using them.
*/LMAX          / Enlarging the main Survo window
*FILE DEL K.TXT / Emptying file K.TXT for the results
*/EX2 20001     / Activating sucro /EX2
*seed=20001     / Seed will be incremented in each repetition.
*VAR Y=X+0.05*N.G(0,1,rand(seed)) TO Q2
*..................................................................
*RANGE=2(0.2)8 Q_MIN=2 RES=q_index,Res,coeff
*QUANTA Q2,Y,2,CUR+1
*Data: Q2 Variable: Y  N=11
*ss=0.0443797
*     quantum     # matches
* 1   2.9989061         7
* 2   6.9792463         4
*..................................................................
*RES=q_index,Res,coeff
*METHOD=Kendall RANGE=2(0.0001)8 SCORE_MIN=1
*FILE DEL COSQUANT
X QUANTA Q2,Y,0,CUR+1
*Data: Q2 Variable: Y  N=11
*GPLOT COSQUANT,quant,score / LINE=1 MODE=SVGA Plot the quantogram!
*Peaks of Kendall's Cosine Quantogram:
* quantum      score
*2.9929000   2.7220160
*2.3403000   1.6339178
*4.5412000   1.5761946
*COPY CUR+1,CUR+1 TO K.TXT
*20001 2.9989061 6.9792463 2.9929000 2.3403000 4.5412000
*
*Above sucro /EX2 collects the results and saves them as a new line
*in file K.TXT.
*..................................................................
*Here is the complete listing of sucro /EX2:
*
*TUTSAVE EX2
/
/ def Wseed=W1 Wsls1=W2 Wsls2=W3 WK1=W4 WK2=W5 WK3=W6
/
+ A: {ref set 1}
*{R}{r5}{erase}{print Wseed}{R}
/
*{act}{R}
*{d2}{act}{d4}
*{r2}{next word}{save word Wsls1}{R}
```

```
*{r2}{next word}{save word Wsls2}{R}
/
*{R}
*{d2}{act}{d}{act}{R}
*{d4}{save word WK1}{R}
*{save word WK2}{R}
*{save word WK3}{R}
*{erase}COPY CUR+1,CUR+1 TO K.TXT{R}
*{disp on}{erase}{print Wseed} {print Wsls1} {print Wsls2}
* {print WK1} {print WK2} {print WK3}{home}{u}{act}
*{tempo 2}{5}
*{tempo 0}{disp off}
/
*{Wseed=Wseed+1}{ref jump 1}{goto A}
+ E: {end}
*
*...............................................................
*20 first replicates of the Monte Carlo experiment:
*LOADP K.TXT
*20001 2.9989061 6.9792463 2.9929000 2.3403000 4.5412000
*20002 2.9981507 6.9925593 2.9896000 2.3461000 4.5455000
*20003 6.9965587 2.9993179 2.9917000 2.3487000 4.5400000
*20004 2.9983136 7.0028953 2.9907000 2.3529000 4.5336000
*20005 2.9976509 6.9936417 2.9900000 2.3459000 4.5393000
*20006 3.0008955 6.9957228 2.9945000 2.3523000 4.5338000
*20007 3.0073174 6.9941041 3.0024000 2.3470000 4.5500000
*20008 2.9985545 7.0014290 2.9912000 2.3530000 4.5337000
*20009 3.0022400 7.0003252 2.9957000 2.3515000 4.5385000
*20010 7.0035025 3.0003706 2.9934000 2.3516000 4.5442000
*20011 2.9954907 7.0013877 2.9870000 2.3531000 4.5253000
*20012 2.9977574 7.0050582 2.9900000 2.3548000 4.5316000
*20013 7.0040663 2.9977474 2.9894000 2.3508000 4.5404000
*20014 6.9969261 3.0015846 2.9948000 2.3496000 4.5441000
*20015 6.9973107 3.0016894 2.9951000 2.3515000 4.5448000
*20016 3.0016448 7.0073630 2.9945000 2.3548000 4.5333000
*20017 6.9959618 3.0001869 2.9935000 2.3524000 4.5300000
*20018 2.9994886 7.0122928 2.9924000 2.3553000 4.5321000
*20019 6.9992919 2.9952425 2.9888000 2.3510000 4.5355000
*20020 7.0051272 3.0003758 2.9934000 2.3549000 4.5381000
*. . .
*...............................................................
```

```
*Saving 1000 first replicates in a new Survo data file Q2_1000:
*FILE SAVE K.TXT TO NEW Q2_1000  / FIRST=1 LAST=1000
*FILE EXPAND Q2_1000 / Making more room for variables
*Renaming and creating more variables variables:
*A ready-made template was obtained by a FILE STATUS command.
*FILE UPDATE Q2_1000
* Copied from text file K.TXT
*FIELDS: (active)
*   1 NA_   8 seed      (####)      seed of random number generator
*   2 NA_   8 sls1      (#.######)  first quantum by SLS
*   3 NA_   8 sls2      (#.######)  second quantum by SLS
*   4 NA_   8 K1        (#.######)  first peak by Kendall's method
*   5 NA_   8 K2        (#.######)  second peak by Kendall's method
*   6 NA_   8 K3        (#.######)  third peak by Kendall's method
*   7 NA_   8 SLS1      (#.######)  smaller quantum by SLS
*   8 NA_   8 SLS2      (#.######)  greater quantum by SLS
*   9 NA_   8 K23       (#.######)  smaller quantum by Kendall's method
*  10 NA_   8 K32       (#.######)  greater quantum by Kendall's method
*  11 NA_   1 K_ind     (#####)     =1 if K's method works, =0 otherwise
*END
*
*FILE LOAD +Q2_1000 / IND=ORDER,1,20 VARS=seed,sls1,sls2,K1,K2,K3
*  seed     sls1     sls2       K1       K2       K3
* 20001 2.998906 6.979246 2.992900 2.340300 4.541200
* 20002 2.998151 6.992559 2.989600 2.346100 4.545500
* 20003 6.996559 2.999318 2.991700 2.348700 4.540000
* 20004 2.998314 7.002895 2.990700 2.352900 4.533600
* 20005 2.997651 6.993642 2.990000 2.345900 4.539300
* 20006 3.000896 6.995723 2.994500 2.352300 4.533800
* 20007 3.007317 6.994104 3.002400 2.347000 4.550000
* 20008 2.998555 7.001429 2.991200 2.353000 4.533700
* 20009 3.002240 7.000325 2.995700 2.351500 4.538500
* 20010 7.003502 3.000371 2.993400 2.351600 4.544200
* 20011 2.995491 7.001388 2.987000 2.353100 4.525300
* 20012 2.997757 7.005058 2.990000 2.354800 4.531600
* 20013 7.004066 2.997747 2.989400 2.350800 4.540400
* 20014 6.996926 3.001585 2.994800 2.349600 4.544100
* 20015 6.997311 3.001689 2.995100 2.351500 4.544800
* 20016 3.001645 7.007363 2.994500 2.354800 4.533300
* 20017 6.995962 3.000187 2.993500 2.352400 4.530000
* 20018 2.999489 7.012293 2.992400 2.355300 4.532100
* 20019 6.999292 2.995242 2.988800 2.351000 4.535500
* 20020 7.005127 3.000376 2.993400 2.354900 4.538100
*. . .
*Making a copy Q2_1000S of the data file:
*FILE COPY Q2_1000 TO NEW Q2_1000S
*..................................................................
```

```
*Changing the order of quanta so that
*the SLS quanta are in increasing order,
*checking whether 'true' quanta by Kendall's method are among
*the first two peaks.
*
*VAR SLS1,SLS2,K23,K32,K_ind TO Q2_1000S
*
*SLS1=if(sls1<sls2)then(sls1)else(sls2)
*SLS2=if(sls1<sls2)then(sls2)else(sls1)
*K_ind=if(K2<K3)then(1)else(0)
*K23=if(K2<K3)then(K2)else(K3)
*K32=if(K2<K3)then(K3)else(K2)
*
*FILE LOAD +Q2_1000S / IND=ORDER,1,20 VARS=seed,SLS1,SLS2,K1,K23,K_ind
*  seed     SLS1     SLS2       K1      K23 K_ind
* 20001 2.998906 6.979246 2.992900 2.340300     1
* 20002 2.998151 6.992559 2.989600 2.346100     1
* 20003 2.999318 6.996559 2.991700 2.348700     1
* 20004 2.998314 7.002895 2.990700 2.352900     1
* 20005 2.997651 6.993642 2.990000 2.345900     1
* 20006 3.000896 6.995723 2.994500 2.352300     1
* 20007 3.007317 6.994104 3.002400 2.347000     1
* 20008 2.998555 7.001429 2.991200 2.353000     1
* 20009 3.002240 7.000325 2.995700 2.351500     1
* 20010 3.000371 7.003502 2.993400 2.351600     1
* 20011 2.995491 7.001388 2.987000 2.353100     1
* 20012 2.997757 7.005058 2.990000 2.354800     1
* 20013 2.997747 7.004066 2.989400 2.350800     1
* 20014 3.001585 6.996926 2.994800 2.349600     1
* 20015 3.001689 6.997311 2.995100 2.351500     1
* 20016 3.001645 7.007363 2.994500 2.354800     1
* 20017 3.000187 6.995962 2.993500 2.352400     1
* 20018 2.999489 7.012293 2.992400 2.355300     1
* 20019 2.995242 6.999292 2.988800 2.351000     1
* 20020 3.000376 7.005127 2.993400 2.354900     1
*. . .
*.............................................................
*Checking K_ind values:
*MINSTAT Q2_1000S,CUR+1 / VARS=K_ind
*Basic statistics of data Q2_1000S N=1000
*Variable   mean       stddev           N  minimum    maximum
* K_ind     0.9930000  0.0834144     1000  0.0000000  1.0000000
*
*1000*0.993=993 acceptable cases (Kendall's method)
*.............................................................
```

```
*Summary of results:
*VARS=SLS1,SLS2,K1,K23,K32
*STAT Q2_1000S,CUR+1 / IND=K_ind  RESULTS=0
*Basic statistics: Q2_1000S N=993
*Variable: SLS1     (#.######)
*min=2.99338  in obs.#25
*max=3.007317 in obs.#7
*mean=2.999926 stddev=0.002311 skewness=0.041493 kurtosis=-0.167154
*lower_Q=2.998286 median=2.999924 upper_Q=3.001532
*
*Variable: SLS2     (#.######)
*min=6.978134 in obs.#967
*max=7.025463 in obs.#896
*mean=7.00011  stddev=0.008037 skewness=0.013163 kurtosis=-0.401024
*lower_Q=6.994274 median=7        upper_Q=7.006089
*
*Variable: K1       (#.######)
*min=2.9852   in obs.#345
*max=3.0024   in obs.#7
*mean=2.992756 stddev=0.002879 skewness=0.158769 kurtosis=0.002095
*lower_Q=2.990839 median=2.992679 upper_Q=2.994742
*
*Variable: K23      (#.######)
*min=2.3372   in obs.#311
*max=2.3629   in obs.#638
*mean=2.351096 stddev=0.00366  skewness=-0.388744 kurtosis=0.342191
*lower_Q=2.348815 median=2.351259 upper_Q=2.35374
*
*Variable: K32      (#.######)
*min=4.5191   in obs.#692
*max=6.9345   in obs.#70
*mean=4.60989  stddev=0.401444 skewness=5.586983 kurtosis=29.22313
----------------------------------------------------------------------
```

The current version of this paper can be downloaded from
`http://www.survo.fi/papers/HuntingQuanta2012.pdf`

References

[1] S.R.Broadbent, *Examination of a quantum hypthesis based on a single set of data*, Biometrika 43, 32-44 1956

[2] D.G.Kendall, *Hunting Quanta*, Phil. Trans. R. Soc. London A 276 (1974), 231–266.

[3] `http://www.survo.fi/english`

Department of Mathematics and Statistics, University of Helsinki

E-mail address: `seppo.mustonen@helsinki.fi`