Analyzing shape, accuracy, and precison of shooting results with the shotGroups package

Daniel Wollschläger*

January 10, 2014

1 Introduction

The `shotGroups` package provides functions to read in, plot, statistically describe, analyze, and compare shot group data with respect to group shape, precision, and accuracy. This includes graphical methods, descriptive statistics, and inference tests using standard, but also nonparametric and robust statistical methods. The data can be imported from files produced by OnTarget PC and OnTarget TDS,[1] or from custom data files in text format with a similar structure.

Use `help(package="shotGroups")` for a list of all functions and links to the detailed help pages with information on options, usage and output.

Note that this vignette currently does not include the diagrams produced by each analysis function.

2 Performing an analysis

Analyzing shot groups usually takes the following steps:

- Read in data (section 2.1)
- Perform either a comprehensive numerical as well as graphical analysis of a group's shape, location (accuracy), and spread (precision) with `analyzeGroup()` (section 2.2) ...
- ... or analyze these aspects of a group separately with `groupShape()` (section 2.3), `groupSpread()` (section 2.4), `groupLocation()` (section 2.5) and a number of utility functions to calculate individual statistical parameters (section 2.7)
- Numerically and visually compare different groups in terms of their shape, location (accuracy), and spread (precision) with `compareGroups()` (section 2.6)

A good description of statistical techniques for shot group analysis can be found at `http://ballistipedia.com/`.

2.1 Reading in data

To import data into R, it should be saved as a text file with the following format:

- The file should have one row for each shot, and one column for each coordinate or any other variable.

*Email: `dwoll@kuci.org`

[1] `http://www.ontargetshooting.com/`

- Columns should be separated by tabs or other whitespace. This type of text file can be exported from OnTarget PC/TDS, or from a spreadsheet application like Excel or Calc.
- The file needs a header in the first line giving the variable names, and should contain at least the (x, y)-coordinates of points of impact, either with variable names `Point.X`, `Point.Y` or just `X`, `Y`.
- For several analysis functions, the following additional variables are useful: `Group` (group number), `Distance` (distance to target), and `Aim.X`, `Aim.Y` (point of aim).
- Note that R is case sensitive, so the aforementioned variable names need to match exactly.
- If you have output files from OnTarget PC/TDS, you can read multiple files with `readDataOT1()` (for OnTarget PC v1.1*), or with `readDataOT2()` (for OnTarget PC v2.1* and OnTarget TDS v3.7*).
- If you have other whitespace or tab-delimited text files with the structure outlined above, you can read multiple files with `readDataMisc()`.
- If your data is saved in some other text file format, consult the help for `read.table()`.

```
> library(shotGroups, verbose=FALSE)       # load shotGroups package
> ## read text files and save to data frame
> ## not run, use data frame provided in package instead
> DFgroups <- readDataMisc(fPath="c:/path/to/files",
+                          fNames=c("series1.dat", "series2.dat"))
```

Note that by default, OnTarget places the origin of the coordinate system in the upper left corner, This can be taken into account by using the function `analyzeGroup()` (section 2.2) which performs a comprehensive analysis of a group's shape, precision and accuracy. When analyzing these aspects separately with `groupShape()` (section 2.3), `groupSpread()` (section 2.4), and `groupLocation()` (section 2.5), the scatterplots will be vertically mirror-reversed.

2.2 Performing a comprehensive analysis

`analyzeGroup()`: This function is a convencience wrapper for the functions presented in sections 2.3, 2.4, and 2.5. It analyzes a group's shape, precision, and accuracy in one go, and collects the results.

```
> library(shotGroups, verbose=FALSE)       # load shotGroups package
> analyzeGroup(DFtalon, conversion='m2mm')
>
> ## output not shown, see following sections for results
```

2.3 Analyzing group shape

`groupShape()`: Assess (multivariate) normality, identify outliers and get a sense for the shape of the bivariate distribution.

Reported statistical parameters and tests:

- Correlation matrix including a robust estimate
- Outlier identification using robust Mahalanobis distances and adjusted quantiles from the χ^2-distribution
- Shapiro-Wilk normality tests for the distribution of x- and y-coordinates

- Energy test for bivariate normality of (x, y)-coordinates

Plots:

- Combined plot for multivariate outlier identification using robust Mahalanobis distances and adjusted quantiles from the χ^2-distribution
- χ^2 Q-Q-plot for eyeballing multivariate normality of (x, y)-coordinates
- Heatmap of a nonparametric 2D-kernel density estimate for the (x, y)-coordinates together with group center and error ellipse based on a robust estimate for the covariance matrix
- Q-Q-plots for eyeballing normality of x- and y-coordinates
- Histogram for x- and y-coordinates including a fitted normal distribution as well as a non-parametric kernel density estimate

```
> library(shotGroups, verbose=FALSE)       # load shotGroups package
> groupShape(DFtalon, bandW=0.25, outlier='mcd')

$corXY
        X       Y
X  1.0000 -0.2931
Y -0.2931  1.0000

$corXYrob
        X       Y
X 1.00000 0.09767
Y 0.09767 1.00000

$Outliers
 [1]  22  24  25  26  28  31  32  33  35  39  81  82  83  85 158

$ShapiroX

        Shapiro-Wilk normality test

data:  X
W = 0.9471, p-value = 3.105e-06


$ShapiroY

        Shapiro-Wilk normality test

data:  Y
W = 0.9552, p-value = 1.769e-05


$multNorm

        Energy test of multivariate normality: estimated parameters

data:  x, sample size 180, dimension 2, replicates 999
E-statistic = 3.74, p-value < 2.2e-16
```

2.4 Analyzing group spread – precision

`groupSpread()`: Assess precision using empirical and parametric spread measures with confidence intervals. Where possible, also use robust estimates based on the MCD method for estimating of the covariance matrix.

Reported statistical parameters and tests:

- Standard deviations of x- and y coordinates including a robust estimate together with parametric and bootstrap confidence intervals (in original measurement units and MOA)
- Covariance matrix of (x, y)-coordinates including a robust estimate
- Empirical mean and median radius as well as estimated Rayleigh precision parameter σ, estimated Rayleigh radial standard deviation RSD, and estimated Rayleigh mean radius MR together with parametric and bootstrap confidence intervals for σ, RSD, and MR (in original measurement units and MOA)
- Maximum pairwise distance (center-to-center, a.k.a. maximum spread, in original measurement units and MOA)
- Width and height of bounding box with diagonal and figure of merit as well as of the (oriented) minimum bounding box (in original measurement units and MOA)
- Radius for the minimum enclosing circle (in original measurement units and MOA)
- Length of semi-major and semi-minor axis of the confidence ellipse including these measures for the confidence ellipse based on a robust estimate for the covariance matrix (in original measurement units and MOA)
- Aspect ratio and flattening of the (robust) confidence ellipse
- Estimates for the circular error probable CEP (Rayleigh, Grubbs-Patnaik, modified RAND-234, in original measurement units and MOA)

Plots:

- Scatterplot of the (x, y)-coordinates together with group center, circle with average distance to center, and confidence ellipse – the latter also based on a robust estimate for the covariance matrix
- Scatterplot of the (x, y)-coordinates together with the bounding box, minimum bounding box, minimum enclosing circle, and maximum group spread
- Histogram for distances to group center including a Rayleigh fit and a nonparametric kernel density estimate

```
> library(shotGroups, verbose=FALSE)       # load shotGroups package
> groupSpread(DFtalon, CEPtype=c("Rayleigh", "Grubbs", "RAND"), level=0.95,
+             sigmaType='Rayleigh', dstTarget=10, conversion='m2mm')

$sdXY
          X      Y
unit 2.2746 2.7308
MOA  0.7819 0.9388

$sdXci
      sdX (  sdX ) sdX basic ( sdX basic ) sdX BCa ( sdX BCa )
```

```
unit 2.0614 2.5374      1.9289      2.6129    2.0052    2.7573
MOA  0.7087 0.8723      0.6631      0.8983    0.6893    0.9479

$sdYci
      sdY ( sdY ) sdY basic ( sdY basic ) sdY BCa ( sdY BCa )
unit 2.4749 3.046      2.4280       3.051    2.4719     3.104
MOA  0.8508 1.047      0.8347       1.049    0.8498     1.067

$sdXYrob
          X      Y
unit 2.0556 2.3094
MOA  0.7066 0.7939

$covXY
       X      Y
X  5.174 -1.820
Y -1.820  7.457

$covXYrob
       X      Y
X 4.2253 0.5141
Y 0.5141 5.3332

$distToCtr
      mean median  sigma    RSD    MR
unit 2.949 2.6696 2.5078 1.6430 3.143
MOA  1.014 0.9178 0.8621 0.5648 1.081

$sigmaCI
     sigma ( sigma ) sigma basic ( sigma basic ) sigma BCa (
unit  2.3433  2.7136        2.2121        2.7994      2.2697
MOA   0.8056  0.9329        0.7605        0.9624      0.7803
     sigma BCa )
unit      2.8872
MOA       0.9925

$RSDci
      RSD (  RSD ) RSD basic ( RSD basic ) RSD BCa ( RSD BCa )
unit 1.5352 1.7778      1.4493      1.8340    1.4870    1.8915
MOA  0.5278 0.6112      0.4982      0.6305    0.5112    0.6502

$MRci
      MR (  MR ) MR basic ( MR basic ) MR BCa ( MR BCa )
unit 2.937 3.401     2.7725      3.508   2.8447    3.619
MOA  1.010 1.169     0.9531      1.206   0.9779    1.244

$maxPairDist
  unit    MOA
16.819  5.782

$groupRect
     width height    FoM  diag
unit 14.05 13.840 13.945 19.72
MOA   4.83  4.758  4.794  6.78
```

```
$groupRectMin
     width height    FoM   diag
unit 15.18 12.517 13.851 19.679
MOA   5.22  4.303  4.762  6.765

$minCircleRad
 unit   MOA
8.409 2.891

$confEll
     semi-major semi-minor
unit      7.181      5.039
MOA       2.469      1.732

$confEllRob
     semi-major semi-minor
unit      5.764      5.035
MOA       1.981      1.731

$confEllShape
aspectRatio  flattening       trace         det
     1.4253      0.2984     12.6311     35.2687

$confEllShapeRob
aspectRatio  flattening       trace         det
     1.1448      0.1265      9.6123     22.6821

$CEP
$CEP$Rayleigh
       50%   90%   95%
unit 2.953 5.689 6.582
MOA  1.015 1.956 2.263

$CEP$Grubbs
        50%   90%   95%
unit 2.8913 5.570 6.445
MOA  0.9939 1.915 2.216

$CEP$RAND
       50%   90%   95%
unit 2.891 5.570 6.445
MOA  0.994 1.915 2.216
```

2.5 Analyzing group location – accuracy

`groupLocation()`: Assess accuracy of a group using empirical and parametric measures.
Reported statistical parameters and tests:

- (x, y)-offset of group center relative to point of aim with the same measure for a robust estimation of the group center
- Distance from group center to point of aim with the same measure for a robust estimation of the group center (in original measurement units and MOA)

- Hotelling's T^2-test result for equality of the true group center with point of aim
- Parametric and bootstrap confidence intervals for the true center's x- and y-coordinate

Plots:

- Scatterplot of the (x, y)-coordinates together with group center as well as a robust estimation of group center. If requested, bullet holes are drawn to scale on a target background (currently, only a very limited number of target types is implemented).

```
> library(shotGroups, verbose=FALSE)       # load shotGroups package
> groupLocation(DFtalon, dstTarget=10, conversion='m2cm',
+               level=0.95, plots=2, target='BDS25m', caliber=5.56)

$ctr
      X       Y
 0.8947 -0.3432

$ctrRob
     X      Y
0.4391 0.3890

$distPOA
  unit    MOA
0.9583 3.2943

$distPOArob
  unit    MOA
0.5867 2.0168

$Hotelling
Analysis of Variance Table

             Df Hotelling-Lawley approx F num Df den Df  Pr(>F)
(Intercept)   1            0.156     13.9      2    178 2.5e-06
Residuals   179

$ctrXci
         x (     x )
t      0.5602 1.22927
basic  0.5724 1.24622
BCa   -0.7375 0.04347

$ctrYci
         y (     y )
t     -0.7448 0.05849
basic -0.7351 0.04053
BCa   -0.7375 0.04347
```

2.6 Comparing groups

`compareGroups()`: Compare two or more groups with regards to their precision and accuracy using empirical measures and statistical tests.

`compareGroups()` requires an additional factor `Series` that identifies shot groups. OnTarget PC/TDS' `Group` variable identifies groups just within one file, `Series` should number groups also across different original files. When you read in data with `readDataOT1()`, `Series` is added automatically (same for `readDataOT2()` and `readDataMisc()`). Otherwise, you can just copy variable `Groups` to `Series` in a data frame called `shots` with `shots$Series <- shots$Group` .

Reported statistical parameters and tests:

- Group centers relative to the respective point of aim
- Distances from group centers to their respective point of aim (in original measurement units and MOA)
- MANOVA result from testing equality of group centers relative to their respective point of aim
- Group correlation matrices for the (x, y)-coordinates
- Group standard deviations for the x- and y-coordinates including parametric 95%-confidence intervals (in original measurement units and MOA)
- Average distances from points to their respective group center (in original measurement units and MOA)
- Maximum pairwise distance between points for each group (center-to-center, a.k.a. maximum spread, in original measurement units and MOA)
- Figure of merit FoM and diagonal of the minimum bounding box for each group (in original measurement units and MOA)
- Radius of the minimum enclosing circle for each group (in original measurement units and MOA)
- Rayleigh estimate for the 50% circular error probable (CEP) in each group (in original measurement units and MOA)
- Ansari-Bradley-test results from testing equality of group variances for x- and y-coordinates – when two groups are compared. With more than two groups, the Fligner-Killeen-test is used
- Wilcoxon-Rank-Sum-test result from testing equality of average point distances to their respective group center – when two groups are compared. With more than two groups, the Kruskal-Wallis-test is used

Plots:

- Scatterplot showing all groups as well as their respective center and 50%-confidence ellipse
- Scatterplot showing all groups as well as their respective minimum bounding box, minimum enclosing circle, and maximum group spread
- Scatterplot showing all groups as well as their respective minimum enclosing circle and circle with average distance to center

```
> library(shotGroups, verbose=FALSE)       # load shotGroups package
> DFsub <- subset(DFtalon, Series %in% 1:3)
> compareGroups(DFsub, conversion='m2mm')
```

```
$ctr
  Series1 Series2 Series3
X  0.3475   3.856 -0.7985
Y -0.1910  -2.913 -1.6140

$distPOA
     Series1 Series2 Series3
unit  0.3965   4.833   1.801
MOA   0.1363   1.661   0.619

$MANOVA
Analysis of Variance Table

            Df Wilks approx F num Df den Df  Pr(>F)
(Intercept)  1 0.676     13.4      2     56 1.7e-05
Series       2 0.504     11.4      4    112 7.9e-08
Residuals   57

$corXY
$corXY$Series1
        X       Y
X  1.0000 -0.4632
Y -0.4632  1.0000

$corXY$Series2
        X       Y
X  1.0000 -0.2143
Y -0.2143  1.0000

$corXY$Series3
        X       Y
X  1.0000 -0.5081
Y -0.5081  1.0000


$sdXY
  Series1 Series2 Series3
X  0.9404   3.354   1.755
Y  1.4912   4.220   1.656

$sdXYmoa
  Series1 Series2 Series3
X  0.3233   1.153  0.6033
Y  0.5126   1.451  0.5691

$sdXYci
      Series1 Series2 Series3
sdX (  0.7151   2.551   1.335
sdX )  1.3735   4.899   2.563
sdY (  1.1340   3.209   1.259
sdY )  2.1780   6.164   2.418

$sdXYciMOA
      Series1 Series2 Series3
```

```
sdX (  0.2458  0.8768  0.4588
sdX )  0.4722  1.6840  0.8811
sdY (  0.3898  1.1033  0.4328
sdY )  0.7487  2.1190  0.8313

$meanDistToCtr
     Series1 Series2 Series3
unit 1.253   4.825   2.096
MOA  0.4306  1.659   0.7206

$maxPairDist
     Series1 Series2 Series3
unit   7.642   15.65   8.077
MOA    2.627    5.38   2.777

$bbFoM
     Series1 Series2 Series3
unit   5.241  12.170   5.757
MOA    1.802   4.184   1.979

$bbDiag
     Series1 Series2 Series3
unit   7.712   17.22   8.732
MOA    2.651    5.92   3.002

$minCircleRad
     Series1 Series2 Series3
unit   3.821   7.825   4.039
MOA    1.314   2.690   1.388

$CEPray
     Series1 Series2 Series3
unit  1.4395   4.402  1.9700
MOA   0.4949   1.513  0.6772

$FlignerX

        Approximative Fligner-Killeen Test

data:  X by Series (1, 2, 3)
chi-squared = 18.23, p-value < 2.2e-16


$FlignerY

        Approximative Fligner-Killeen Test

data:  Y by Series (1, 2, 3)
chi-squared = 21.42, p-value < 2.2e-16


$Kruskal

        Approximative Kruskal-Wallis Test
```

```
data:  dstCtr by Series (1, 2, 3)
chi-squared = 15.34, p-value = 0.002
```

2.7 Utility functions

The `shotGroups` package also provides a number of utility functions that can be used separately. These functions accept either a data frame or a matrix with (x, y)-coordinates. For more information on additional options, see the respective help page.

- `getBoundingBox()`, `getMinBBox()`: Calculates the vertices of the bounding box / of the minimum-area (possibly oriented) bounding box
- `getMinCircle()`: Calculates center and radius of the minimum enclosing circle
- `getCEP()`: Calculates the Rayleigh, Grubbs-Patnaik, and modified RAND R-234 estimate for the circular error probable
- `getConfEll()`: Calculates the 2D-confidence ellipse under the assumption of bivariate normality. Also includes the ellipse based on a robust estimate for the covariance matrix of the (x, y)-coordinates
- `getDistToCtr()`: Calculates the distances of a set of points to their center
- `getMaxPairDist()`: Calculates the maximum of pairwise distances between points
- `getRayParam()`: Estimates the radial precision parameter σ of the Rayleigh distribution together with the Rayleigh radial standard deviation RSD and the Rayleigh mean MR, including parametric confidence intervals
- `getMOA()`, `fromMOA()`: Converts object size to MOA or vice versa.

```
> library(shotGroups, verbose=FALSE)       # load shotGroups package
> getBoundingBox(DFtalon)                  # axis-aligned bounding box

$pts
  xleft ybottom  xright    ytop
  -4.43   -4.37    9.62    9.47

$width
[1] 14.05

$height
[1] 13.84

$FoM
[1] 13.95

$diag
[1] 19.72

> getMinBBox(DFtalon)                      # minimum-area bounding box

$pts
          X      Y
[1,] -2.447 11.428
```

```
[2,] -5.161 -3.512
[3,]  7.155 -5.750
[4,]  9.869  9.190

$width
[1] 15.18

$height
[1] 12.52

$FoM
[1] 13.85

$diag
[1] 19.68

$angle
   Y
79.7

> getMinCircle(DFtalon)                     # minimum covering circle

$ctr
[1] 2.940 3.015

$rad
[1] 8.409

> getCEP(DFtalon, type=c("Rayleigh", "Grubbs"))    # circular error probable

$Rayleigh
       50%   90%   95%
unit 2.953 5.689 6.582
MOA  1.015 1.956 2.263

$Grubbs
        50%   90%   95%
unit 2.8913 5.570 6.445
MOA  0.9939 1.915 2.216

$ellShape
aspectRatio  flattening
     1.4253      0.2984

$ctr
     X      Y
0.8947 0.3432

> getConfEll(DFtalon)                       # confidence ellipse

$ctr
     X      Y
0.8947 0.3432
```

```
$ctrRob
      X       Y
 0.4481 -0.3938

$cov
      X     Y
X 5.174 1.820
Y 1.820 7.457

$covRob
        X       Y
X  4.3125 -0.4921
Y -0.4921  5.2850

$size
     semi-major semi-minor
unit      3.432     2.4081
MOA       1.180     0.8278

$sizeRob
     semi-major semi-minor
unit     2.7643     2.3907
MOA      0.9503     0.8219

$shape
aspectRatio  flattening       trace         det
     1.4253      0.2984     12.6311     35.2687

$shapeRob
aspectRatio  flattening       trace         det
     1.1562      0.1351      9.5975     22.5494

$magFac
[1] 1.18

> getMaxPairDist(DFtalon)                 # maximum pairwise distance

$d
[1] 16.82

$idx
[1]  39 169

> getRayParam(DFtalon)                    # Rayleigh parameter estimates

$sigma
  sigma sigCIlo sigCIup
  2.508   2.343   2.714

$RSD
    RSD RSDciLo RSDciUp
  1.643   1.535   1.778

$MR
```

```
    MR MRciLo MRciUp
 3.143  2.937  3.401

> getMOA(c(1, 2, 10),   dst=100, conversion='m2cm')  # convert to MOA

[1] 0.3438 0.6875 3.4377

> fromMOA(c(0.5, 1, 2), dst=100, conversion='m2cm')  # convert from MOA

[1] 1.454 2.909 5.818
```

3 Included data sets

The `shotGroups` package includes a number of empirical data sets with shooting results:

- `DF300BLK`: One group of shooting a Noveske AR-15 rifle in 300BLK at 100yd (20 observations)[2]
- `DFcciHV`: Two groups of shooting a PWS T3 rifle in .22LR at 100yd (40 observations, see footnote 2)
- `DFcm`: Several groups of shooting a 9x19mm pistol at 25m (487 observations)
- `DFtalon`: Several groups of shooting a Talon SS air rifle at 10m (180 observations)[3]
- `DFsavage`: Several groups of shooting a Savage 12 FT/R rifle in .308 Win at distances from 100 to 300m (180 observations, see footnote 3)
- `DFscar17`: One group of shooting an FN SCAR 17 rifle in .308 Win at 100yd (10 observations, see footnote 2)

4 TODO

- Make diagrams show up in this vignette
- Provide the option to render data in units SMOA (exactly 1" at 100 yards) or mils, instead of MOA
- Provide the option to specify bullet diameter, and then render the points in all group plots using that size, with alpha-blending so that overlaps can be distinguished (as in `groupLocation(..., plots="2")` see section 2.5)
- Include units somewhere on plots – either axis labels or title
- Human-readable output from functions `compareGroups()`, `groupLocation()`, `groupShape()`, `groupSpread()` by using `print()` methods
- Rayleigh parameter estimates from possibly elliptical distributions – see DiDonato (2007): "Computation of the Circular Error Probable (CEP) and Confidence Intervals in Bombing Tests"
- Repeat simulations from McMillan & McMillan (2008) to estimate CEP correction factors for elliptical distributions for different confidence levels

[2]Thanks: David Bookstaber `http://ballistipedia.com/`

[3]Thanks: Charles & Paul McMillan `http://statshooting.com/`

- Function which accepts extreme order statistics (maximum spread) to allow inference for Rayleigh parameters σ, RSD, MR
- Find and fit platykurtic distribution type that allows for more near misses than normal distribution
- Add formulas for calculated statistics and literature references to this vignette
- Allow dates be associated with group data to track accuracy and precision performance over time