Depth in Multivariate Statistics

Ignacio Cascos

Department of Statistics
Universidad Carlos III de Madrid

Madrid, February 2019

Outline

A little taste EDA

Main ingredient depth functions

- How to cook them

Full menu applications

Side dishes extensions

In-depth recipe control charts

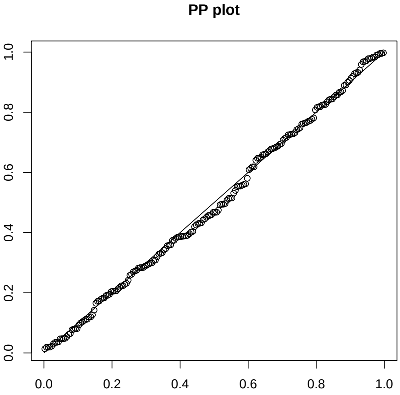
Outline

Depth functions

- A depth function quantifies how central a point $x \in \mathbb{R}^d$ is wrt a multivariate probability distribution (or a data cloud).
- The more central points are assessed high depths, while peripheral ones assume low values of depth.
- Each notion of depth induces a center-outward ordering in a multivariate dataset.
- Many univariate data-analytic techniques can be extended to the multivariate setting by using depth functions.

PPplot & DDplot

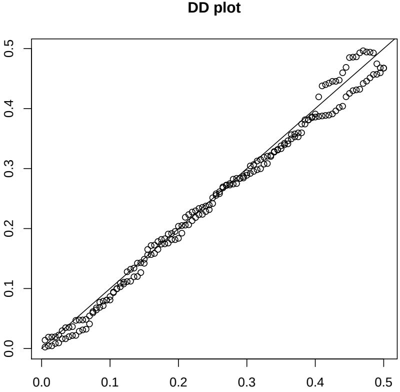
Liu, Parelius, and Singh, AOS, 1999



$$(F_n(x_i), G_m(x_i))$$

PPplot & DDplot

Liu, Parelius, and Singh, AOS, 1999



$$(\mathrm{D}(x_i; F_n), \mathrm{D}(x_i; G_m))$$

$$\mathrm{D}(x; F) = \min\{F(x), 1 - F(x)\}$$

DDplot, Multivariate goodness-of-fit

Liu, Parelius, and Singh, AOS, 1999

A data set and a population distribution

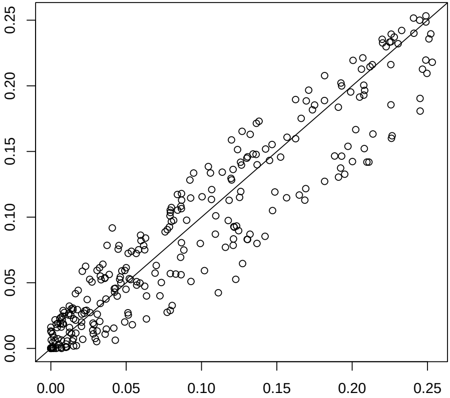
Plot the empirical depth of every point of the data set versus its population depth.

Two data sets

Take every point from each of the two data sets and plot its empirical depth wrt the first data set versus the empirical depth wrt the second.

DDplot, Multivariate goodness-of-fit

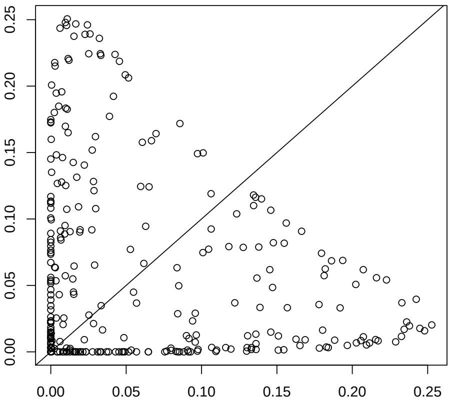
Liu, Parelius, and Singh, AOS, 1999



Two data sets drawn from the same distribution.

DDplot, Multivariate goodness-of-fit

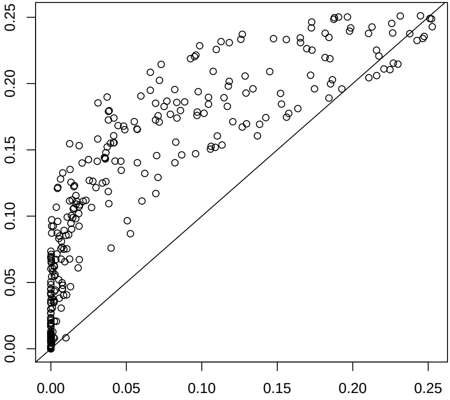
Liu, Parelius, and Singh, AOS, 1999



Location shift.

DDplot, Multivariate goodness-of-fit

Liu, Parelius, and Singh, AOS, 1999

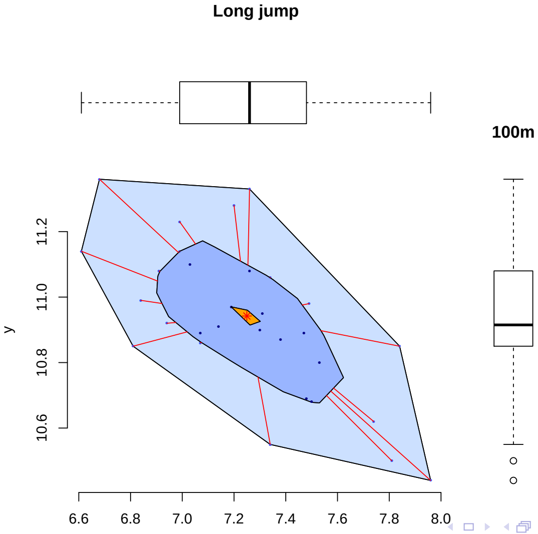


Shift in scale.

Boxplot & Bagplot

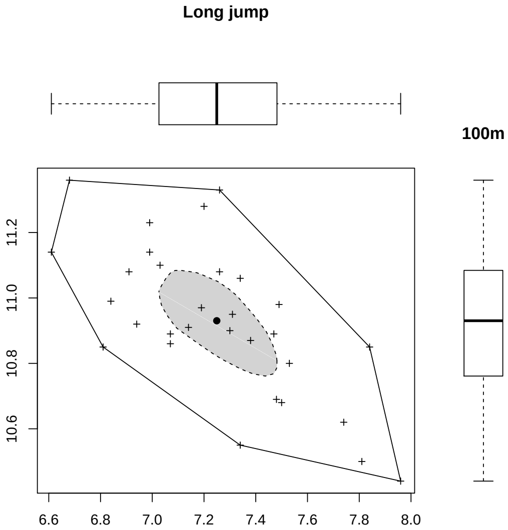
Rousseeuw, Ruts, and Tukey, AS, 1999, `aplpack`

Decathlon @ Athens 2004: long jump & 100m

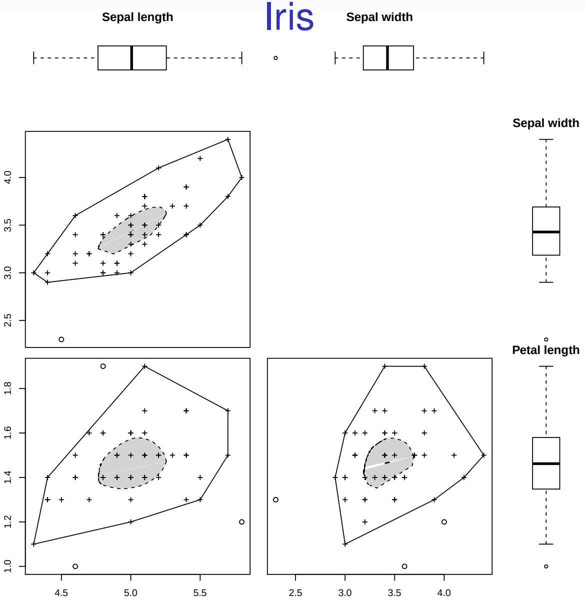


BEXplot

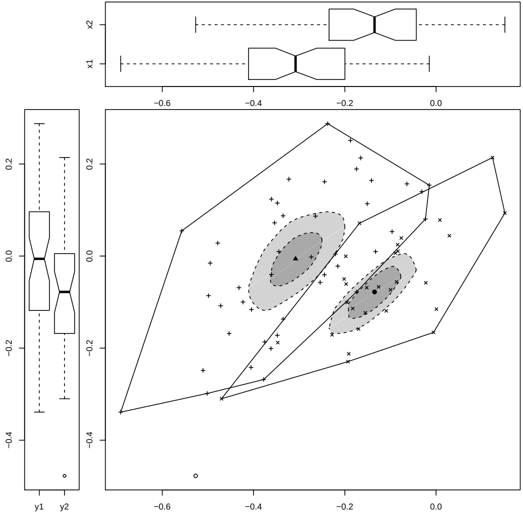
Cascos and Ochoa, in preparation



Iris


Sepal length
Sepal width
Sepal width
Petal length
4.0
3.5
3.0
2.5
1.8
1.6
1.4
1.2
1.0
4.5
5.0
5.5
2.5
3.0
3.5
4.0

Hemophilia



Outline

Depth function

Liu, AOS, 1990, Zuo and Serfling, AOS, 2000, Dyckerhoff, ASA, 2004

A depth function, $\mathrm{D}(x;P)$ (or $\mathrm{D}(x)$ shortly), satisfies:

- D1 Affine invariance. $\mathrm{D}(Ax+b;P_{AX+b}) = \mathrm{D}(x;P_X)$ for every nonsingular $A \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$;
- D2 Vanishes at infinity. $\mathrm{D}(x;P) \longrightarrow 0$ if $\|x\| \to \infty$;
- D3 Upper semicontinuity. Level sets are closed ;
- D4 Monotonicity relative to deepest point. Decreasing in rays from center ;
- D4' Quasiconcavity. Level sets are convex .

The level sets are or depth-trimmed regions are

$$\mathrm{D}^\alpha(P) = \{x \in \mathbb{R}^d : \mathrm{D}(x;P) \geq \alpha\}.$$

Conversely $\mathrm{D}(x;P) = \sup\{\alpha : x \in \mathrm{D}^\alpha(P)\}$.

Depth function

Liu, AOS, 1990, Zuo and Serfling, AOS, 2000, Dyckerhoff, ASA, 2004

A depth function, $\mathrm{D}(x;P)$ (or $\mathrm{D}(x)$ shortly), satisfies:

- D1 Affine invariance. $\mathrm{D}(Ax+b;P_{AX+b}) = \mathrm{D}(x;P_X)$ for every nonsingular $A \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$;
- D2 Vanishes at infinity. $\mathrm{D}(x;P) \longrightarrow 0$ if $\|x\| \to \infty$;
- D3 Upper semicontinuity. Level sets are closed ;
- D4 Monotonicity relative to deepest point. Decreasing in rays from center ;
- D4' Quasiconcavity. Level sets are convex .

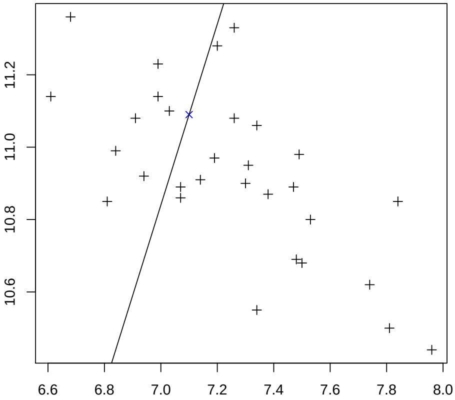
The level sets are or depth-trimmed regions are

$$\mathrm{D}^{\alpha}(P) = \{x \in \mathbb{R}^d : \mathrm{D}(x;P) \geq \alpha\}.$$

Conversely $\mathrm{D}(x;P) = \sup\{\alpha : x \in \mathrm{D}^{\alpha}(P)\}$.

Halfspace depth (Tukey, 1975)

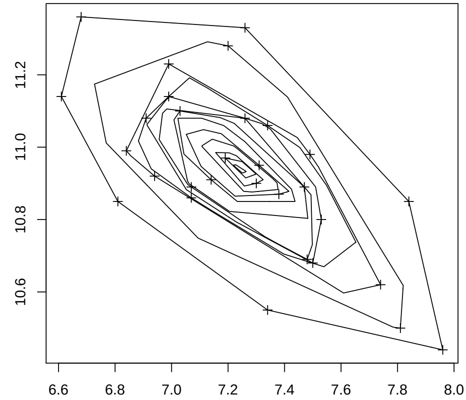
Decathlon @ Athens 2004: long jump vs. 100m



$$\mathrm{HD}_n(x) = n^{-1} \min_{u \in \mathbb{R}^d} \#\{i : \langle X_i, u \rangle \geq \langle x, u \rangle\}$$

Halfspace depth (Tukey, 1975)

Decathlon @ Athens 2004: long jump vs. 100m



Halfspace depth

Tukey, 1975, Rousseeuw and Ruts, Metrika, 1999

Population halfspace depth

$$\mathrm{HD}(x; P) = \inf\{P(H) : x \in H \text{ closed halfspace}\};$$

- Univariate $\mathrm{HD}(x; P_X) = \min\{F_X(x), 1 - F_X(x)\}$;
- Satisfies Properties D1–D4 and D4'.

Halfspace trimming

$$\mathrm{HD}^\alpha(P) = \bigcap\{H : H \text{ closed halfspace } P(H) > 1 - \alpha\};$$

- Univariate $\mathrm{HD}^\alpha(P_X) = [F_X^{-1}(\alpha), F_X^{-1}(1 - \alpha)]$.

Halfspace depth

Tukey, 1975, Rousseeuw and Ruts, Metrika, 1999

Population halfspace depth

$$\mathrm{HD}(x;P) = \inf\{P(H) : x \in H \text{ closed halfspace}\} \, ;$$

- Univariate $\mathrm{HD}(x;P_X) = \min\{F_X(x), 1 - F_X(x)\}$;
- Satisfies Properties D1–D4 and D4'.

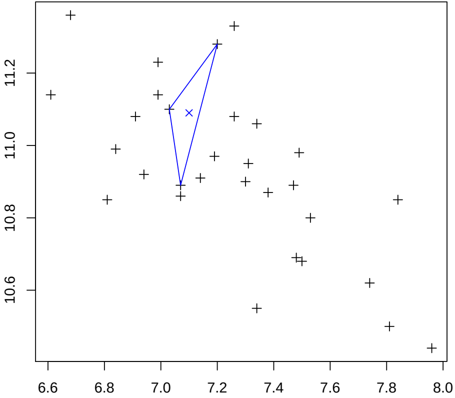
Halfspace trimming

$$\mathrm{HD}^{\alpha}(P) = \bigcap\{H : H \text{ closed halfspace } P(H) > 1 - \alpha\} \, ;$$

- Univariate $\mathrm{HD}^{\alpha}(P_X) = [F_X^{-1}(\alpha), F_X^{-1}(1 - \alpha)]$.

Simplicial depth (Liu, AOS, 1990)

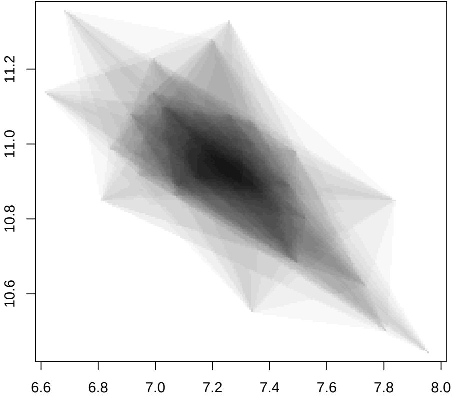
Decathlon @ Athens 2004: long jump vs. 100m



$$\mathrm{SD}_n(x) = \binom{n}{d+1}^{-1} \sum_{1 \le i_1 < i_2 < \ldots < i_{d+1} \le n} \mathbb{I}(x \in \mathrm{co}\{X_{i_1}, X_{i_2}, \ldots, X_{i_{d+1}}\})$$

Simplicial depth (Liu, AOS, 1990)

Decathlon @ Athens 2004: long jump vs. 100m



Simplicial depth

Liu, AOS, 1990

Population simplicial depth

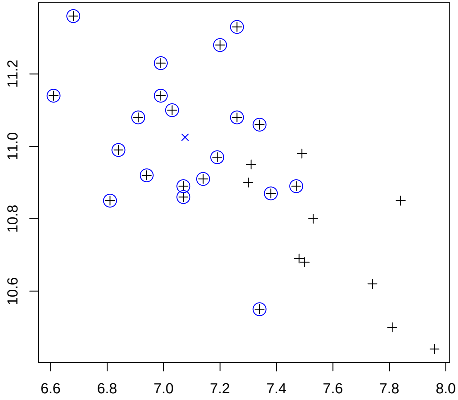
$$\mathrm{SD}(x;P) = \Pr(x \in \mathrm{co}\{X_1, X_2, \ldots, X_{d+1}\})$$

where $X_1, \ldots, X_{d+1}$ are independent with distribution P.

- Univariate $\mathrm{SD}(x;P) = 2F(x)(1 - F(x))$;
- Satisfies D1–D3. Also D4 on absolutely continuous and angularly symmetric distributions.

Zoid trimming (Koshevoy and Mosler, AOS, 1997)

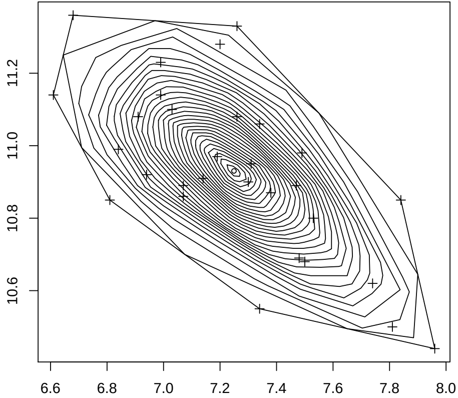
Decathlon @ Athens 2004: long jump vs. 100m



$$\mathrm{ZD}_n^{k/n} = \mathrm{co}\left\{ \frac{1}{k}\sum_{i=1}^{k} X_{\pi(i)} \; : \; \pi \text{ permutation of } \{1,\ldots,n\} \right\}$$

Zonoid trimming (Koshevoy and Mosler, AOS, 1997)

Decathlon @ Athens 2004: long jump vs. 100m



Zoond trimming

Koshevoy and Mosler, AOS, 1997

Population zonoid trimmed regions

$$\mathrm{ZD}^{\alpha}(P) = \left\{ \int \boldsymbol{x} \mathrm{d}Q(x) : \ Q \text{ probability distribution } Q \leq \alpha^{-1} P \right\}$$

- Univariate

$$\mathrm{ZD}^{\alpha}(P_X) = \left[\frac{1}{\alpha} \int_0^{\alpha} F_X^{-1}(t) \mathrm{d}t \, , \, \frac{1}{\alpha} \int_{1-\alpha}^{1} F_X^{-1}(t) \mathrm{d}t \right]$$

- Properties D1–D4 and D4’, $\mathrm{ZD}^1(P) = \{\mathbf{E}X\}$.
- Zonoid depth $\mathrm{ZD}(x; P) = \sup\{\alpha : \ x \in \mathrm{ZD}^{\alpha}(P)\}$.

How to cook them

- R package `depth` can be used to compute the halfspace and simplicial depths and to obtain the contour-plots of their level sets as well as the deepest points.
- R package `ddalpha` contains depth-based classification algorithms and also routines to compute several depth functions, in particular the zonoid depth in dimension $d \geq 2$.

Outline

Applications of data depth

- Multivariate EDA (Rousseeuw, Ruts, and Tukey, AS, 1999; Liu, Parelius, and Singh, AOS, 1999)
- Clustering and classification (Ghosh and Chaudhuri, Bernoulli, 2005; Li, Cuesta-Albertos, and Liu, JASA, 2012)
- Outlier detection (Dang and Serfling, JSPI, 2010)
- Inference procedures including control charts (Liu, JASA, 1995; Cascos and López-Díaz, AMM, 2018)
- Imputation of missing data (Mozharovskyi, Josse, and Husson, JASA, 2019)
- Risk measurement (Cascos and Molchanov, FS, 2007)

Outline

Parameter depth

Mizera, AOS, 2002; Mizera and Müller, JASA, 2004

- Nonfit: Element from a parameter space that is not a suitable for a parameter wrt a sample.
- Depth: Fraction of points to be deleted from a sample in order to make our candidate a *nonfit*.

Notions of parameter depth

- Halfspace depth: a nonfit is a point out of the convex hull.
- Regression depth: (Rousseeuw and Hubert, JASA, 1999) a nonfit is a line that does not contain any data point and whose residuals change sign, at most, once.
- Location-scale depth: Mizera and Müller, JASA, 2004.
- Scatter matrices: (Chen, Gao, and Ren, AOS, 2018; Paindaveine and Van Bever, AOS, 2018)

Parameter depth

Mizera, AOS, 2002; Mizera and Müller, JASA, 2004

- Nonfit: Element from a parameter space that is not a suitable for a parameter wrt a sample.
- Depth: Fraction of points to be deleted from a sample in order to make our candidate a *nonfit*.

Notions of parameter depth

- Halfspace depth: a nonfit is a point out of the convex hull.
- Regression depth: (Rousseeuw and Hubert, JASA, 1999) a nonfit is a line that does not contain any data point and whose residuals change sign, at most, once.
- Location-scale depth: Mizera and Müller, JASA, 2004.
- Scatter matrices: (Chen, Gao, and Ren, AOS, 2018; Paindaveine and Van Bever, AOS, 2018)

Parameter depth induced by a probability functional

Cascos and López-Díaz, JMVA, 2012

Consider any functional of a probability $T : \mathbb{P} \mapsto \mathbb{R}^k$

- Trimmed regions

$$\mathrm{D}_T^{\alpha}(P) = \{T(Q) : Q \text{ prob. distribution}, \ Q \leq \alpha^{-1} P\}, \text{ for } 0 < \alpha \leq 1$$

- Depth function

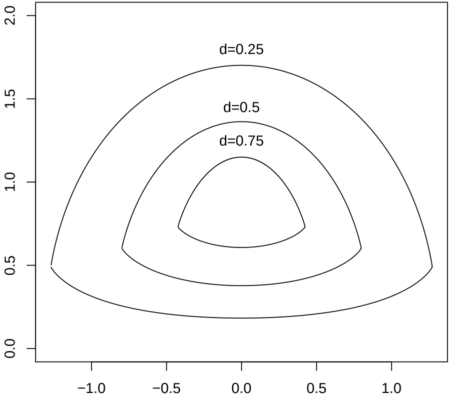
$$\mathrm{D}_T(x; P) = \sup\{0 < \alpha \leq 1 : x \in \mathrm{D}_T^{\alpha}(P)\}$$

- Location-scale depth. If P is univariate distribution and $T(P) = (\mu(P), \sigma(P))$

$$\mathrm{LSD}((m, s); P_X) = \mathrm{ZD}((m, s^2 + m^2); P_{X,X^2})$$

Location-scale depth

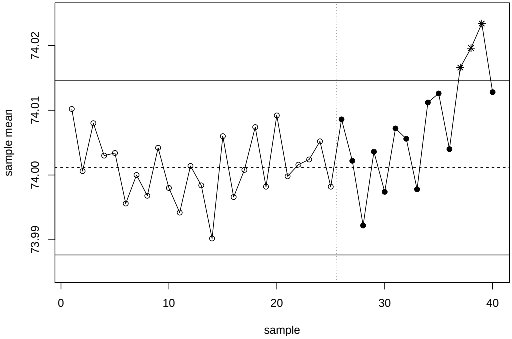
Location-scale regions of levels 0.75, 0.5, and 0.25 wrt a standard normal distribution



Outline

Quality Control

$\overline{X}$-chart

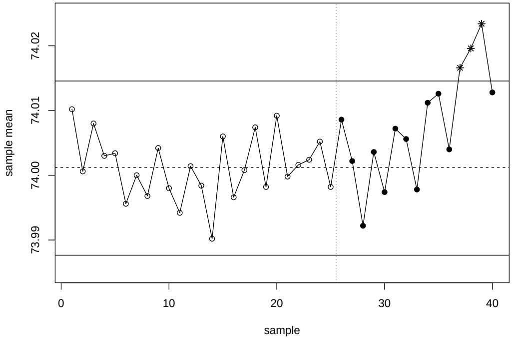


Depth-based multivariate rank

$$r_P(x) = P(\{y : \mathrm{D}(y; P) \leq \mathrm{D}(x; P)\})$$

Quality Control

$\overline{X}$-chart



Depth-based multivariate rank

$$r_P(x) = P(\{y :\ \mathrm{D}(y;P) \le \mathrm{D}(x;P)\})$$

Quality Control

Liu, JASA, 1995

Multivariate Control Charts

- Control Chart for r instead of X-Chart.

$$r_{\hat{P}_m}(X_i) = \#\{Y_j : \mathrm{D}_m(Y) \leq \mathrm{D}_m(X_i)\}/m\,.$$

- Control Chart for Q (average rank) instead of $\overline{X}$-Chart.

$$Q(\hat{P}_m; X_1, \ldots, X_k) = \frac{1}{k}\sum_{i=1}^{n} r_{\hat{P}_m}(X_i)\,.$$

- Control Chart for S (accumulated rank) instead of CUSUM Chart.

$$S(\hat{P}_m; X_1, \ldots, X_n) = \sum_{i=1}^{n}\left(r_{\hat{P}_m}(X_i) - \frac{1}{2}\right)\,.$$

Quality Control

Liu, JASA, 1995

Multivariate Control Charts

- Control Chart for r instead of X-Chart.

$$r_{\hat{P}_m}(X_i) = \#\{Y_j : \mathrm{D}_m(Y) \leq \mathrm{D}_m(X_i)\}/m\,.$$

- Control Chart for Q (average rank) instead of $\overline{X}$-Chart.

$$Q(\hat{P}_m; X_1, \ldots, X_k) = \frac{1}{k}\sum_{i=1}^{n} r_{\hat{P}_m}(X_i)\,.$$

- Control Chart for S (accumulated rank) instead of CUSUM Chart.

$$S(\hat{P}_m; X_1, \ldots, X_n) = \sum_{i=1}^{n}\left(r_{\hat{P}_m}(X_i) - \frac{1}{2}\right)\,.$$

Quality Control

Liu, JASA, 1995

Multivariate Control Charts

- Control Chart for r instead of X-Chart.

$$r_{\hat{P}_m}(X_i) = \#\{Y_j : \mathrm{D}_m(Y) \leq \mathrm{D}_m(X_i)\}/m\,.$$

- Control Chart for Q (average rank) instead of $\overline{X}$-Chart.

$$Q(\hat{P}_m; X_1, \ldots, X_k) = \frac{1}{k}\sum_{i=1}^{n} r_{\hat{P}_m}(X_i)\,.$$

- Control Chart for S (accumulated rank) instead of CUSUM Chart.

$$S(\hat{P}_m; X_1, \ldots, X_n) = \sum_{i=1}^{n}\left(r_{\hat{P}_m}(X_i) - \frac{1}{2}\right)\,.$$

Nonparametric control charting based on depth

- Control charts for rational samples of size n.
- For each sample statistic $\hat{\theta} = T(\hat{P}_n)$ is monitored.

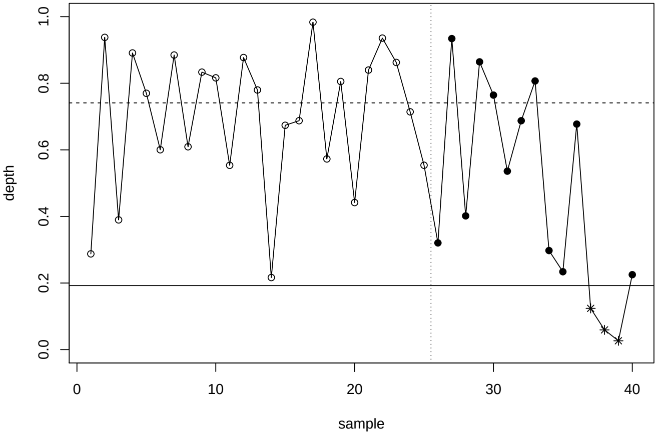
Ph. I.1 Build the historical dataset out of the $m = k \times n$ observations coming from k samples of size n.

Ph. I.2 Resample from the historical dataset taking samples of size k in order to estimate the unique CL.

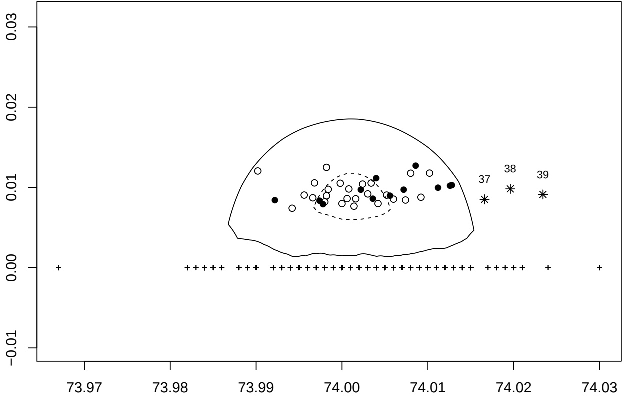
Ph. I.3 For each trial sample check $\mathrm{D}_m(\hat{\theta}) \geq \mathrm{CL}$. If not, delete the corresponding trial sample and return to step Ph.I.1.

Ph. II For each on-going sample, evaluate $\mathrm{D}_m(\hat{\theta})$ and rise an alarm if it lies below CL.

Location-scale control chart



Location-scale control chart



Conclusions

- Data depth refers to the centrality of a point wrt a data cloud (or probability distribution).
- Parameter depth refers to how well does an element of a parameter space fit a probability distribution as its parameter.
- Applications of data depths and parameter depths have been briefly introduced.

Conclusions

- Data depth refers to the centrality of a point wrt a data cloud (or probability distribution).
- Parameter depth refers to how well does an element of a parameter space fit a probability distribution as its parameter.
- Applications of data depths and parameter depths have been briefly introduced.

Conclusions

- Data depth refers to the centrality of a point wrt a data cloud (or probability distribution).
- Parameter depth refers to how well does an element of a parameter space fit a probability distribution as its parameter.
- Applications of data depths and parameter depths have been briefly introduced.

Acknowledgements

Thank you!!