Statistical Models For Data In Compact Groups

K. Gerald van den Boogaart
Ernst-Moritz-Arndt Universität Greifswald,
Institut for mathematics and computer sciences
boogaart @ uni-greifswald.de

Samples with group scale

Geology is concerned with many different scales. Beside the classic scale hierarchy from binary over categorical, ordinal, interval, natural to real scale and multivariate vector scale, geological statistic is concerned with special equivalance classes of groups: rotations $SO(3)$ discribing plate movements, the sphere [6] $S^2 = SO(3)/$"rotations around z" discribing directions, the simplex [1] discribing compositions, crystallographic orientations $SO(3)/G$ [4], orientations of planes and axis $S^2/\{-1\}$ e.g. of schist or folding, configurations of objects and orders of sequences given by $S_n/$"equivalence of layers". Here I would like to give a general view to statistics of observations from these compact groups G and their derived spaces like (right-) quotients of groups G/R with $\forall \bar{g}_1, \bar{g}_1 \in G/R : \bar{g}_1 = \bar{g}_2 :\Leftrightarrow \exists r \in R : g_1 = g_2 r$.

Group scale differs from real scale by some important aspects:

- Groups have only one operation, which I will write as a multiplicative operation.
- There is essentially only one $\mathbb{R}$ and only one $\mathbb{R}^d$, but there are many different groups.
- Real statistics rely on $0, 1, \| \|^2, <$ and are concerned with mean, variance, cdf and ranks – things quite meaningless on groups –. With groups we have subgroups and symmetry.

Theoretical background for compact groups: Haar-measure, representations

On a compact group G the canonical measure replacing the Lebesgue-measure is the Haar measure μ_G defined as the only measure invariant under the operation of the group[7] (i.e. $\forall A \forall g \in G : \mu_G(A) = \mu_G(gA)$) and unit mass $\mu_G(G) = 1$. $L^2(G) := L^2(\mu_G)$ of G is spanned by the matrix elements of the representations $T_l(g) = (T_l^{mn}(g))_{mn}, l = 0, \ldots, \infty$ [9] of G. The T_l are group homomorphisms.

An important problem of the statistics of groups is the consideration of symmetry and thus to define for any subgroups $L, R \lhd G$ an orthogonal basis of $L^2(LGR) = \{f \in L^2(G) : f(lgr) = f(g) \forall l \in L \forall r \in R\}$. Such basis can be constructed by fixing orthogonal rectangular matrices A_l with $\mathbf{im} A_l = \langle T_l(g) : g \in L \rangle^\perp$, and B_l with $\mathbf{im} B_l = \langle T_l(g) : g \in R \rangle^\perp$ and define[4, cf.]:

$$T_l^{LR}(g) := A_l^t T_l(g) B_l$$

Characteristic functions, convolution and moments

A replacement for characteristic function of a distribution P on G is a mapping from the frequency domain $\mathbb{N}_0$ to matrices of varying frame size given by:

$$f_P^*(l) := \int T_l(g) dP(g) = \int T_l(g) f_P(g) d\mu_G(g)$$

In general $f_P^*(i)$ is a function in $L^\infty(\mathbb{N}_0)$ and the characteristic function of a convolution

$$(P_1 * P_2)(A) := \int_G \int_G 1_A(g_1 g_2) dP_1(g_1) dP_2(g_2)$$

is given by $f^*_{P_1 * P_2}(l) = f^*_{P_1}(l) f^*_{P_2}(l)$. A concept of linear moments in the sense of R^d does not exist for groups. The $\mu_l := \int T_l(g) dP(g) = f^*_P(l)$ can instead be seen as non centered or harmonic moments of P. The μ_l are not elements of the group, however linear with respect to group operations: $\mu_l^{\sigma X \tau} = T_l(\sigma) \mu_l^X T_l(\tau)$

Symmetry, kernels and distances

Let us call a function left symmetric with respect to $S \subset G$ when $\forall s \in S \forall g \in G : f(sg) = f(g)$, right symmetric, when $\forall s \in S \forall g \in G : f(gs) = f(g)$, double symmetric, when $\forall s \in S \forall t \in S \forall g \in G : f(sgt) = f(g)$ and conjugationally symmetric, when $\forall s \in S \forall g \in G : f(sgs^{-1}) = f(g)$.

Some statistical procedures rely on kernels or covariance functions, which are positive semidefinite functions $c(g_1, g_2)$ invariant under operation of the group: $c(sg_1t, sg_2t) = c(g_1, g_2)$ implying $c(g_1, g_2) = k(g_1^{-1} g_2)$ with a conjugationally symmetric $k(g) = \sum_l \alpha_l \mathbf{tr} T_l(g)$. Positive definiteness is equivalent to $\forall l : \alpha_l > 0$. For some statistical procedures (kernel density estimation, location parameters) we need such sequence $(\alpha_l)_l$. For a given sequence $(\alpha_l)_l$ a distance, which is useful even after symmetrization, of two group elements is defined by:

$$d_\alpha(g_1, g_2) := \sqrt{\sum_l \alpha_l^2 \|T_l(g_1) - T_l(g_2)\|^2}$$

Measures of location and spread, symmetric location

In real scale the first and second moment are measures of location and spread. Here we can consider a single μ_l as a measure of location as well of spread, which is well known for the special case of the spherical Fisher distribution[6], however $\mu_l \notin G$. A location parameter $g \in G$ with $\sum_l \alpha_l^2 \|T_l(g) - \hat{\mu}_l\|^2 \to \min$ can be considered as a parameter of location, especially, when we consider that this g is also given by:

$$g_\alpha := \operatorname*{argmin}_{g_0} E[d_\alpha(g_0, g)], \quad \hat{g}_\alpha := \operatorname*{argmin}_{g \in G} \sum_i d_\alpha(g, g_i)^2$$

Correspondingly a metric variance spread could be defined by:

$$\mathbf{var}_\alpha(g) := E[d_\alpha(g_\alpha, g)^2], \quad \hat{\mathbf{var}}_\alpha(g) := \frac{1}{n-1} \sum_i d_\alpha(\hat{g}_\alpha, g_i)^2$$

g_α is unique up to the symmetry acutally present in the distribution or in the data. Thus e.g. for the uniform distribution every $g \in G$ is with equal right the location. When we assume some symmetry in the distribution, we can remove all distance that can be induced by that symmetry by replacing the distance d_α by:

$$d_\alpha^{LR}(g_1, g_2) := \sqrt{\sum_l \alpha_l^2 \|T_l^{LR}(g_1) - T_l^{LR}(g_2)\|^2}$$

For trivial α this can get a trivial distance $d_\alpha^{LR}(g_1, g_2) \equiv 0$. However for $(\alpha_l)_l, \alpha_l > 0 \forall l$ this distance not degenerated and based on them we can define a symmetric location and spread parameters by:

$$g_\alpha^{LR} := \operatorname*{argmin}_{g_0 \in L \backslash G / R} E[d_\alpha^{LR}(g_0, g)], \quad \hat{g}_\alpha^{LR} := \operatorname*{argmin}_{g \in L \backslash G / R} \sum_i d_\alpha^{LR}(g, g_i)^2$$

$$\mathbf{var}_\alpha^{LR}(g) := E[d_\alpha^{LR}(g_\alpha^{LR}, g)^2], \quad \hat{\mathbf{var}}_\alpha^{LR}(g) := \frac{1}{n-1} \sum_i d_\alpha^{LR}(\hat{g}_\alpha^{LR}, g_i)^2$$

All these estimators are strongly consistently up to the symmetry actually present in the data.

To have an interpretation of the α-variance it is important to get an idea of the meaning of the d_α on the specific group.

Empirical distribution, V,U-statistics

Instead of empirical cdfs we use the symmetric empirical distribution:

$$\hat{P}(A) = \frac{1}{n}\sum_{i=1}^{n} \int\int 1_A(\sigma g_i \tau) d\mu_L(\sigma) d\mu_R(\tau)$$

$\hat{P}(A)$ estimates $P(A)$ strongly consistent for any A with $\mu(A) > 0$. This type of convergence replaces convergence in every continuity point of the distribution function. For any functional of the distribution given by a k-positional kernel Ψ:

$$\beta = \int \dots \int \Psi(g_1, \dots, g_k) dP(g_1) \dots dP(g_k)$$

we can define a consistent V-estimate[8] on the group by:

$$\hat{\beta}_V = \int \dots \int \Psi(g_1, \dots, g_k) d\hat{P}(g_1) \dots d\hat{P}(g_k) = \frac{1}{n^k} \sum_{i_1=1}^{n} \dots \sum_{i_k=1}^{n} \Psi(g_{i_1}, \dots, g_{i_k})$$

or a consistent and unbiased U-estimate[8] on the group by:

$$\hat{\beta}_U = \frac{1}{\binom{n}{k} k!} \sum_{i_1 \neq \dots \neq i_k} \Psi(g_{i_1}, \dots, g_{i_k})$$

The difference to the classical V- and U- estimators is only the notation, which changed uses measures instead of the no longer defined distribution functions.

Symmetric kernel density estimation

In general a kernel density estimation [5, cf.] on G can be given by the convolution of a kernel on G with the discrete distribution given by the data g_i. When we assume f to be symmetric with respect to $L \lhd G$ from left and $R \lhd G$ from right we might give a correspondingly symmetric kernel density estimate by:

$$\hat{f}(\bar{g}) = \frac{1}{n}\sum_{i=1}^{n} \int\int\int k(s g_i^{-1} l g r) d\mu_S(s) d\mu_L(l) d\mu_R(r) = \frac{d(\mu_L * K * \hat{P} * \mu_R)(\bar{g})}{d\mu_{G/S}(\bar{g})}$$

With $\alpha_0 = 1$ $\hat{f}$ is an unbiased estimator for a smoothed version of the density: $E[\hat{f}] = \frac{d(K*P)(g)}{d\mu_G(g)}$ with $P = \mu_L * P * \mu_R$ due to assumed symmetry of P, $\mu_L * K = K * \mu_L$, and $P * K = K * P$ due to the conjugational symmetry of K. Other aspects of kernel density estimation, like prediction error, consistency [5, cf.] and confidence bounds [4, cf.] are similar to the special cases for Stiefel manifolds or crystallographic orientations discussed elsewhere.

Distributions and models

A general class of exponential families for our, scale containing the uniform, the matrix Langevin [5, cf.], von Mises-Fisher-matrix [6] and Beran's [3] general distribution families on the sphere as special cases and used for crystallographic orientations in [4], is given as follows

Fix a group G, left and right symmetry groups L, R (eventually $L = R = \{1\}$) and a maximum degree of series expansion D. Then for every choice of $\Theta = (\Theta_l)_l$, $\Theta_l \in R^{MN} \ni <$

$T_l^{LR}(g) >$ the following is an exponential family distribution density, which is symmetric with respect to L from left and with respect to R from right:

$$\frac{dP_{LGRD\Theta}(g)}{d\mu_G(g)} = A(\Theta)\exp\left(\sum_{l\leq D}\mathbf{tr}\Theta_l^t T_l^{LR}(g)\right), \quad A(\Theta) = \text{normalization}$$

The results of [4] on crystallographic exponential families can be generalized: the normalization constant always exists due to the compactness of the space. The family can alternatively be parameterized by all moments μ_l with l showing up in the sum. The maximum likelihood estimator factorizes over the corresponding empirical moments. The distribution is the one with maximum μ_G-entropy among all distributions with that given moments.

Regression in group scale can be done by generalized linear regression models with natural link based on a linear expansion of Θ- parameter of $P_{LGRD\Theta}$ or the Langevin distribution in terms of real and categorical regressors and in terms of $T_l(g)$ of group scale regressors and in terms of $A_l T_l^{LR}(\bar{g}) B_l^t$ of regressors $\bar{g} \in L\backslash G/R$ [4].

References

[1] Aichison, J (1986) *The Statistical Analysis of Compositional Data*, Monographs on Statistics and Applied Probability, Chapman & Hall Ltd, London

[2] Bauer (2002) *Wahrscheinlichkeitstheorie*, 5.Auflage, Walter deGruyter, Berlin und New York

[3] Beran, R. (1979) Exponential families for directional data, *The Ann. Stat.* **7**, pp. 1162-1178

[4] Boogaart, (2002) *Statistics of individual crystallographic orientation measurements*, Industriemathematik und Angewante Mathematik, Shaker Verlag, Aachen

[5] Chikuse, Y. (2003) *Statistics on Special Mannifolds*, Lecture Notes in Statistics 174, Springer, New York

[6] Mardia, K.V., Jupp,P.E. (2000) *Directional statistics*, John Wiley&Sons, New York

[7] Nachbin, L. (1972) *The Haar Integral*, Krieger Publ. Co, New York

[8] Shao, J. (1998) *Mathematical Statistics*, Springer, New York

[9] Vilenkin, N.J., Klimyk, A.U. (1991) *Representation theory of Lie Groups and Special Functions*, Volume 1, Kluwer Academic Publisher

RÉSUMÉ

A general statistics for compact groups allows to analyze many different geological data in a uniform way. On one hand statistical analysis at compact group scale differs in questions, methodology and computational problems substantially from other scales. On the other hand based on representation theory most classical ideas can be transported in a general way into group scale. Only some examples are given here. While the classical aspects of normal distribution and linearity get meaningless, new canonical concepts of symmetry and uniformity arise. The concept of moments separates from the concept of parameters of location and scale. While the moments naturally express symmetry, we need to invent new descriptions of location for symmetric distributions.