Best predictors in logarithmic distance between positive random variables

H. GZYL

Abstract

The metric properties of the set in which random variables take their values lead to relevant probabilistic concepts. For example, the mean of a random variable is a best predictor in that it minimizes the L_2 distance between a point and a random variable. Similarly, the median is the same concept but when the distance is measured by the L_1 norm.

Also, a *geodesic distance* can be defined on the cone of strictly positive vectors in $\mathbb{R}^n$ in such a way that, the minimizer of the distance between a point and a collection of points is their geometric mean.

That geodesic distance induces a distance on the class of strictly positive random variables, which in turn leads to an interesting notions of conditional expectation (or best predictors) and their estimators. It also leads to different versions of the Law of Large Numbers and the Central Limit Theorem. For example, the lognormal variables appear as the analogue of the Gaussian variables for version of the Central Limit Theorem in the logarithmic distance.

Mathematics Subject Classification 2010: 60B99, 60B12, 60A99.

Keywords: Prediction in logarithmic distance, Law of large numbers in logarithmic distance, Central Limit Theorem in logarithmic distance, Logarithmic geometry for positive random variables.

1. INTRODUCTION AND PRELIMINARIES

The study of random variables and processes taking values in spaces with geometries other than Euclidean is not new. Consider the textbooks by Kunita and Watanabe [5] or by Hsu [3] to mention just two. In this line of work, the distance between points in the base manifold is replaced by a geodesic distance derived from a Riemannian metric placed. Such distance is inherited by random variables taking values in the manifold.

It should not then be surprising that the notion of best predictor of a random variable by variables of a given class, should depend on the metric of the manifold. In this note we shall consider the manifold to be $\mathcal{M} = (0, \infty)^n$, which is an open set in $\mathbb{R}^n$, which is also a commutative group with respect to component wise multiplication. We postpone the study of the geometry of this group to the appendix. Here we mention that what we do is the commutative version of a more elaborate geometry in the space of symmetric matrices. The reader can check with Lang [6] in which a relation of this geometry to Bruhat-Tits spaces is explained, or in Lawson and Lim [7] or Mohaker [9]

and references therein, where the geometric mean property in the class of symmetric matrices is established. More recently Resigny et al. [1] and Schwartzman [10] used the same geometric setting to study the role of such geometry in a large variety of applications. The applications of the geometric ideas in these references concern the non-commutative case, but the simplest commutative case and its potential usefulness for positive random variables seems not to have been explored.

As mentioned in the abstract, it is the purpose of this note to explore the possible usefulness of measuring distances between positive numbers by a logarithmic distance resulting from an interesting group invariant metric.

In the appendix we establish that the distance between any two points $x_i, x_2 \in \mathcal{M}$ is given by

$$d(x_1, x_2)^2 = \sum_{i=1}^n (\ln x_1(i) - \ln x_2(i))^2. \quad (1.1)$$

This makes $\mathcal{M}$ a Tits-Bruhat space in which the distance satisfies a semi-parallelogram law. This is contained in Theorem 8.1. We shall use this property to establish the uniqueness of conditional expectations. And the group structure in $\mathcal{M}$ will be inherited in a curious way by the conditional expectations (or by the best predictors) in the logarithmic distance (8.1).

Starting from the logarithmic distance on $\mathcal{M}$, and from the fact that it satisfies the semi-parallelogram law, we come to the main objective of the paper, which is to consider the notion of best predictor (conditional expectations) in that distance. These matters will be taken up in Sections 2 and 3, where we shall introduce the notion of ℓ -expected value and ℓ -conditional expectation, which will denote the best predictors in the logarithmic distance (hence the ℓ -prefix). We examine there some of the basic properties of these constructs.

In Section 4 we present the two most basic estimators, namely, that of the ℓ -mean and that of the ℓ -variance, and explain how the law of large numbers and the central limit theorem for these estimators relates to the standard law of large numbers and the central limit theorems.

In section 5 we examine how the notion of martingale related to the ℓ -conditional expectation relates to the standard notion of martingale. We shall do it in discrete time, but the extension to continuous time is quite direct. In Section 6 we examine Markowitz portfolio theory when the distance between (gross) returns is the logarithmic distance.

As said, we leave the study of the geometry on $\mathcal{M}$ to the appendix. There we explain how the logarithmic distance between strictly positive vectors is actually a geodesic distance in that manifold. For that we shall present some results from Lang's [6], but in a simpler, commutative setup. The basic idea behind our constructions has been very much studied in geometry. The idea is to put a group action upon $\mathcal{M}$ and construct a scalar product on the tangent bundle TM that is invariant under the group action. That scalar product determines the logarithmic (geodesic) distance on $\mathcal{M}$ that interests us. That appendix can be read independently of the previous sections dealing with probabilistic aspects.

2. BEST PREDICTORS IN LOGARITHMIC DISTANCE

Our set up here consists of a probability space $(\Omega, \mathcal{F}, P)$ and we shall be concerned with the cone $\mathcal{C}$ of P -almost everywhere (a.e. for short) finite and strictly positive ($\mathcal{M}$ -valued) random variables. As usual, we identify variables that are P -a.e. equal. Since the operations among vectors are component wise, to reduce to the case $n = 1$ only takes a simple notational change. To shorten the description of the random variables used in the statements coming up below, let us introduce the following notations. For $p > 1$ (we shall be concerned with $p = 1, 2$ only) define:

$$L_p = \{X \in \mathcal{F} \mid E[|X_i|^p] < \infty, \ i = 1, \dots, n\}$$

$$Ln_p = \{X \in \mathcal{C} \mid \ln X \in L_p\}, \quad LLn_p = L_p \cap Ln_p.$$

Let X_1 and X_2 be two strictly positive random variables in Ln_2 . The (logarithmic) distance between them is defined to be

$$d_\ell(X_1, X_2)^2 \equiv E \left[\sum_{i=1}^n (\ln X_1(i) - \ln X_2(i))^2 \right] \quad (2.1)$$

Since we are identifying variables that are a.e equal, $d_\ell(X_1, X_2)$ is a distance on $\mathcal{C}$. Similarly to $m = E[X]$ being the constant that minimizes the Euclidean (squared) distance to X , we have

PROPOSITION 2.1. With the notations introduced above, let $X \in Ln_2$. The vector m_ℓ that minimizes the logarithmic distance to X is given by

$$m_\ell(X) = \exp(E[\ln X]).$$

The proof of the first assertion is computational, and the second results from an

application of Jensen's inequality. When there is no risk of confusion, we shall write $m_\ell(X) = m_\ell$. Keep in mind that the operations are componentwise, and that $m_\ell(X)_j = \exp(E[\ln X_j])$ for $j = 1, \dots, n$. If $X \in LLn_1$, we also have $m_\ell \leq E[X]$.

COMMENT. To define $m_\ell(X)$ it suffices that $X \in Ln_1$, but if we want it to be a best predictor in the logarithmic distance, we require $X \in Ln_2$.

And the analogue of the notions of covariance and centering are contained in the following definition.

DEFINITION 2.1. Let now $X, Y \in Ln_2$. We define the logarithmic covariance matrix of the non-negative random variables X and Y by

$$Cov_\ell(X, Y) \equiv E[(\ln X - \ln m_\ell(X))(\ln Y - \ln m_\ell(Y))^t] = Cov(\ln X, \ln Y).$$

Let Σ be the matrix with components $E[(\ln X_i - \ln m_\ell(X_i))(\ln Y_j - \ln m_\ell(Y_j))]$. If the matrix Σ is invertible, we define the “centered” (in logarithmic distance) version of X by

$$X^c \equiv \exp\left(\Sigma^{-1/2}(\ln X - \ln m_\ell(X))\right)$$

The need for the exponentiation is clear: First we have to “undo” the taking of the logarithms and second, the argument of the exponential function is a vector in $\mathbb{R}^n$ which yields a positive vector after exponentiation. It takes a simple computation to verify that

$$m_\ell(X^c) = \mathbf{1}, \quad \Sigma_\ell(X^c) = \mathbb{I}.$$

A variation on the previous theme consists of predicting a variable Y by a variable X in logarithmic distance. The extension of the previous result is contained in the following statement.

PROPOSITION 2.2. Let Y and X be in Ln_2 . Then the $\sigma(X)$ –measurable random variable that minimizes the logarithmic distance (2.1) to Y is given by

$$E_\ell[Y|X] = \exp(E[\ln Y | X]).$$

And we also have $E_\ell[Y|X] \leq E[Y | X]$.

The proof of Proposition 2.2 follows the same pattern as the standard proof. Just notice that $\phi(X) = \exp(E[\ln Y | X])$ is a bounded, $\sigma(X)$ –measurable random variable, such that $\ln \phi(X) = E[\ln Y | X]$ minimizes the Euclidean square distance to $\ln Y$.

COMMENT. As above, we remark that $E_\ell[Y|X]$ can be defined for $Y \in Ln_1$, but if we want it to be a best predictor in logarithmic distance, we need $Y \in Ln_2$.

Note that the last inequality mentioned in the statement does not mean that one of the estimators is better than the other in any sense. They are minimizers in different metrics. Also, since linear combinations in an exponent are transported as scaling and powers, we have the following analogue to linear prediction for positive random variables.

PROPOSITION 2.3. Let Y and X be positive real variables with square integrable logarithms. The values of $a > 0$ and $b \in \mathbb{R}$ that make $Y^\# \equiv aX^b$ the best predictor of Y in the logarithmic metric, are given by

$$\begin{cases} a = \exp(E[(\ln Y)] - bE[(\ln X)]) \\ b = \frac{1}{D} (E[\ln X \ln Y] - E[\ln X]E[\ln Y]) \\ D = E[(\ln X)^2] - (E[\ln X])^2 = \sigma^2(\ln X). \end{cases}$$

The proof follows the standard computation starting from the definition of $d(Y, aX^b)_\ell$. Certainly the result is natural as the linear structure of $\mathbb{R}$ is transferred multiplicatively onto $(0, \infty)$ by the exponential mapping. Also, the extension to random variables taking values in higher dimensional $\mathcal{M}$ is direct, but notationally more cumbersome.

A simple computation leads to

$$m_\ell(Y^\#) = E_\ell[Y^\#] = e^{E[\ln Y]}, \quad \sigma_\ell(Y^\#) = b^2 \sigma^2(\ln X).$$

3. LOGARITHMIC CONDITIONAL EXPECTATION AND SOME OF ITS PROPERTIES

Here we extend the semi-parallelogram property mentioned in Theorem (8.1) to strictly positive random variables.

LEMMA 3.1. All random variables mentioned are supposed to be in Ln_2 . Let X_1 and X_2 be as mentioned. Then there exists $Z \in Ln_2$ such that for any Y we have

$$d(X_1, X_2)_\ell^2 + 4d(Z, Y)_\ell^2 \leq 2d(Y, X_1)_\ell^2 + 2d(Y, X_2)_\ell^2.$$

To prove this, use the second comment after Theorem (8.1) at every $\omega \in \Omega$ to obtain the pointwise version of the semi-parallelogram property, and then integrate with respect to P . Clearly $Z = (X_1 X_2)^{1/2} \in LLn_2$. Below we apply this to obtain the uniqueness of the extension of the standard notion of conditional expectation.

THEOREM 3.1. Let $\mathcal{G} \subset \mathcal{F}$ be a σ -algebra, and let Y be non-negative with square integrable logarithm. Then, the unique -up to a set of P measure 0-, positive $X^* \in \mathcal{G}$ that makes $d(Y, X)_\ell^2$ minimum over $\{X \in \mathcal{G}, X > 0, E[(\ln X)^2] < \infty\}$, is given by $X^* = \exp(E[\ln Y | \mathcal{G}])$. To be consistent with the notations introduced above, we shall write $X^* = E_\ell[Y | \mathcal{G}]$.

PROOF. The existence follows the same pattern of proof as the propositions in the previous section, that is $E[\ln Y | \mathcal{G}]$ minimizes the ordinary square distance to $\ln Y$, and it is the unique (up to sets of P measure 0). We shall use the semi-parallelogram property to verify the uniqueness. For that, let X some other possible minimizer of the logarithmic distance. Now set $Z = \sqrt{XX^*}$ (keep in mind the second comment after Theorem (8.1)), and observe that according to the semi-parallelogram property

$$d(X^*, X)_\ell^2 + 4d(Y, Z)_\ell^2 \leq 2d(Y, X^*)_\ell^2 + 2d(Y, X)_\ell^2.$$

Since by definition, $d(Y, Z)_\ell^2$ is larger than any of the two distances in the right hand side of the inequality, it follows that necessarily $d(X^*, X)_\ell^2 = 0$.

Notice that in the group structure on $\mathcal{M}$ described in the Appendix, the analogue to the multiplication by scalars is replaced by the exponentiation. This is used to verify the analogue of the standard definition of conditional expectation in logarithmic distance. For the heuristics see the comment in the Appendix.

THEOREM 3.2. Let Y be a $\mathcal{M}$ valued random variable such that $Y \in Ln_1$. Then $E_\ell[Y | \mathcal{G}] = \exp(E[\ln Y | \mathcal{G}])$ is the unique $\mathcal{G}$ -measurable, $\mathcal{M}$ -valued random variable, such that for any bounded $\mathcal{G}$ -measurable real valued H the following holds:

$$E_\ell[Y^H] = E_\ell[E_\ell[Y | \mathcal{G}]^H].$$

PROOF. It follows the standard pattern. Notice that $Y^H \in Ln_1$, therefore

$$E_\ell[Y^H] = e^{E[H \ln Y]} = e^{E[HE[\ln Y | \mathcal{G}]]} = e^{E[\ln(\exp(E[\ln Y | \mathcal{G}]))^H]} = E_\ell[E_\ell[Y | \mathcal{G}]^H].$$

The intermediate steps consist of an application of the standard conditional expectation rules.

Let us now verify some standard and non standard properties of the notion of conditional expectation introduced above. Keep in mind that the arithmetic operations with positive vectors are componentwise.

THEOREM 3.3. Let $Y \in LLn_2$ and let $\mathcal{H} \subset \mathcal{G}$ be two sub- σ -algebras of $\mathcal{F}$. Then, up to a set of measure 0, the following hold:

$$1) E_\ell[Y | \{\emptyset, \Omega\}] = E_\ell[Y].$$

$$2) E_\ell[E_\ell[Y|\mathcal{G}]|\mathcal{H}] = E_\ell[Y|\mathcal{H}].$$

3) Let $Y_1, \dots, Y_k$ be in LLn_2 , and $w_i \in \mathbb{R}$. The analogue of the linearity property of the standard conditional expectation is the following multiplicative property:

$$E_\ell[\prod_{i=1}^k Y_i^{w_i} | \mathcal{G}] = \prod_{i=1}^k \left(E_\ell[Y|\mathcal{G}] \right)^{w_i}$$

4) If Y is independent of $\mathcal{G}$ in the standard sense, then $E_\ell[Y|\mathcal{G}] = E_\ell[Y]$.

PROOF. The first assertion is simple consequence of the definition. To verify the second we start from the definition and carry on:

$$E_\ell[E_\ell[Y|\mathcal{G}]|\mathcal{H}] = \exp \left(E[\ln \exp E[\ln Y|\mathcal{G}|\mathcal{H}]] \right) = \exp \left(E[E[\ln Y|\mathcal{G}|\mathcal{H}]] \right),$$

and now apply the standard tower property of conditional expectations to finish the proof of the assertion.

It is in the proof of (3) where the logarithmic distance plays a curious role. The proof of the assertion is a simple computation starting from the definition:

$$E_\ell[\prod_{i=1}^k Y_i^{w_i} | \mathcal{G}] = \exp \left(E \left[\sum w_i \ln Y_i | \mathcal{G} \right] \right) = \prod_{i=1}^k \left(E_\ell[Y|\mathcal{G}] \right)^{w_i}.$$

The fourth property is also simple to establish using the definition and the standard notion of independence.

4. ESTIMATORS AND LIMIT THEOREMS

In this section we shall consider the case $n = 1$. The notation is a bit simpler in this case. That is, we shall forget about the symbols in boldface for a while.

Making use of Proposition (8.1) the following definition is clear:

DEFINITION 4.1. Let $X_1, \dots, X_K$ be positive random variables. We define their empirical logarithmic mean by

$$\hat{m}_{\ell,K}(X) = \left(\prod_{j=1}^K X_j \right)^{1/K}.$$

And a the standard law of large numbers becomes:

THEOREM 4.1. Let $X_j, j \geq 1$ be a collection of i.i.d. positive random variables defined on $(\Omega, \mathcal{F}, P)$ having finite logarithmic variance σ_ℓ^2 and mean m_ℓ . Then

$\hat{m}_{\ell,K}(X)$ is an unbiased estimator of the logarithmic mean $m_\ell(X)$ and

$$\hat{m}_{\ell,K}(X) = \left(\prod_{j=1}^K X_j \right)^{1/K} \rightarrow m_\ell(X)$$

almost surely w.r.t. P as $K \rightarrow \infty$.

The proof is clear. Since

$$\hat{m}_{\ell(X),K} = \exp \left(\frac{1}{K} \sum_{j=1}^K \ln X_j \right),$$

we can invoke the strong law of large numbers, see Borkhar [2] or Jacod and Protter [4], plus the continuity of the exponential function to obtain our assertion. That $\hat{m}_{\ell,K}(X)$ has logarithmic mean $m_\ell(X)$ is clear.

In analogy with the standard notion of empirical variance, we can introduce

DEFINITION 4.2. With the notations introduced above and under the assumptions in Theorem 4.1, the empirical estimator of the logarithmic variance is defined by

$$\hat{\sigma}_{\ell,K}^2(X) = \frac{1}{K-1} \sum_{j=1}^K (\ln X_j - \ln \hat{m}_\ell(X))^2.$$

And, as in basic statistics, we have

THEOREM 4.2. With the notations introduced above, and under the assumptions of Theorem (4.1), $\hat{\sigma}_\ell^2(X)$ is an unbiased estimator of the logarithmic variance and

$$\hat{\sigma}_{\ell,K}^2(X) \rightarrow \sigma_\ell^2(X)$$

almost surely w.r.t. P as $K \rightarrow \infty$.

But perhaps more interesting is the following version of the central limit theorem. It brings to the fore the role of lognormal variables as the analogue to the Gaussian random variables in the class of positive variables.

THEOREM 4.3. Suppose that $X_j, j \geq 1$ are a collection of i.i.d. random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with logarithmic mean $m_\ell = E[\ln X_j]$ and $E[(\ln X_j)^2] < \infty$. Then

$$\left(\prod_{j=1}^K \frac{X_j}{m_\ell} \right)^{1/\sqrt{K}} \rightarrow e^X$$

in law as $K \rightarrow \infty$, where $X \sim N(0, \sigma_\ell^2)$.

PROOF. Observe that

$$\left(\prod_{j=1}^K \frac{X_j}{m_\ell}\right)^{1/\sqrt{K}} = \exp\left(\frac{1}{\sqrt{K}} \sum_{j=1}^K (\ln X_j - \ln m_\ell)\right).$$

From the standard proof of the central limit theorem we know that $\frac{1}{\sqrt{K}} \sum_{j=1}^K (\ln X_j - \ln m_\ell)$ converges in probability to an $N(0, \sigma_\ell^2)$ random variable and therefore, since the exponential function is continuous, the same convergence holds for $\left(\prod_{j=1}^K \frac{X_j}{m_\ell}\right)^{1/\sqrt{K}}$. Thus concludes the proof of our assertion.

5. ℓ -MARTINGALES IN DISCRETE TIME

As there is a notion of ℓ -conditional expectation, there must be a corresponding notion of ℓ -martingale. In this section we examine some of its very simple properties. As usual, the basic setup consists of the probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\{\mathcal{F}_n, n \geq 0\}$.

THEOREM 5.1. Let the $\mathcal{M}$ -valued process $\{X_n; n \geq 0\}$ be such that $X_n \in \mathcal{F}_n$ and the $\xi_n = \ln X_n$ are integrable. Then X_n is an ℓ -martingale (resp. sub-martingale, super-martingale) if and only if $\{\xi_n\}$ is an ordinary martingale (resp. sub-martingale, super-martingale).

Also, if X_n is an ℓ -martingale, it is an ordinary sub-martingale.

PROOF. For $n \geq 0$ and $k \geq 1$

$$E_\ell[X_{n+k} | \mathcal{F}_n] = e^{E[\xi_{n+k} | \mathcal{F}_n]}$$

from which the assertion of the theorem drops out. For the second assertion note that

$$E_\ell[X_{n+k} | \mathcal{F}_n] = X_n = e^{\xi_n} = e^{E[\xi_{n+k} | \mathcal{F}_n]} \leq E[e^{\xi_{n+k}} | \mathcal{F}_n] = E[X_{n+k} | \mathcal{F}_n]$$

The middle step drops out from Jensen's inequality.

The corresponding version of the Doob decomposition theorem, say for sub-martingales, goes as follows.

THEOREM 5.2. With the notations introduced above, let $\{X_n\}$ be an $\mathcal{M}$ -valued ℓ -sub-martingale. Then there exist an $\mathcal{M}$ -valued ℓ -martingale $\{Y_n\}$ and an increasing $\mathcal{M}$ -valued process A_n , such that $X_n = Y_n A_n$.

PROOF. Just apply the Doob decomposition theorem to $\xi_n = \ln X_n$ and use

$$X_n = e^{\xi_n}.$$

6. LOGARITHMIC GEOMETRY AND PORTFOLIO THEORY

Let us introduce a slight change of notation to conform with the notation in financial modeling. By the generic R we shall denote the (gross) return of any asset of portfolio, which means the quotient of its current value divided by its initial value.

To begin with, we proved in the Appendix, see (8.3), that the curve $R_1^w R_2^{1-w}$ is a geodesic in the logarithmic distance between the points R_1 and R_2 . That curve can be thought of as a weighted geometric mean of R_1 and R_2 . This remark leads to variation on the theme of “return” of a portfolio. In our setup, a generic portfolio w , characterized by the weights $w_1, \dots, w_K$, of assets with gross returns $R_1, \dots, R_K$, has a weighted return given by the geometric mean: $\prod_{i=1}^K R_i^{w_i}$. To push the geodesic interpretation a bit further, that geometric mean can be thought of as a sequence of geodesic walks joining say R_1 to R_K . Anyway, note that the logarithm of the ℓ -mean m_ℓ given by,

$$\ln m_\ell = \sum_{i=1}^K w_i E[\ln R_i] \quad (6.1)$$

is clearly the logarithmic rate of growth of the portfolio w . Recall as well that the square of the logarithmic distance of m_ℓ to $\prod_{i=1}^K R_i^{w_i}$ is given by

$$d(\prod_{i=1}^K R_i^{w_i}, m_\ell)^2 = \text{Var}(\sum_{i=1}^K w_i \ln R_i) = (w, \Sigma w) \quad (6.2)$$

where Σ is the covariance matrix of the logarithmic returns. Imitating Markowitz’s portfolio theory, we assign to any portfolio w its logarithmic mean $m_\ell(w)$ and its logarithmic variance $\sigma_\ell^2(w)$. According to Markowitz’s proposal a portfolio is optimal when it minimizes the variance for a given expected value of its (rate of) return. The following result contains the analogue of the classical Markowitz result, but for the case of the logarithmic results.

PROPOSITION 6.1. With the notations introduced above, the weights $w_i^*, \dots, w_K^*$ that make the logarithmic variance, $\sigma_\ell^2(w) = d(\prod_{i=1}^K R_i^{w_i}, m_\ell)^2$ minimal subject to the constraints $\sum w_i = 1$ and $m_\ell(w) = e^\mu$, are the same as the weights that minimize $\text{Var}(\sum_{i=1}^K w_i \ln R_i)$ subject to $E[\sum_{i=1}^K w_i \ln R_i] = \mu$ and $\sum w_i = 1$.

This is a standard quadratic optimization problem, whose solution is simple starting from (6.2). We refer the interested reader to Luenberger [8] or to Shiryayev [11] for

more details about the classical Markowitz portfolio optimization theory.

7. CONCLUDING COMMENTS

In this note we proposed an alternative metric in the set of positive vectors, such that when distance between random variables is measured in this metric, the standard notions of best predictors, their estimation, and the classical convergence results, acquire a different but intuitively related form.

Also, as a simple application to finance, when assets are characterized by their gross returns (which by definition are positive random variables), the concept of return of a portfolio becomes a weighted geometric average, and the standard portfolio choice methodology appears in a slightly different guise. Readers familiar with the basics of the methodology will find it clear that the analogue of the efficient frontier, market portfolio, market line and CAPM have a counterpart within the formalism developed above, but this is not the place to pursue the matters.

8. APPENDIX: THE LOGARITHMIC DISTANCE BETWEEN POSITIVE VECTORS

We shall think of the vectors in $\mathbb{R}^n$ as functions $\xi : \{1, \dots, n\} \rightarrow \mathbb{R}$, and all standard arithmetical operations either as component wise operations among vectors or point wise operations among functions. Let us denote by $\mathcal{M} = \{x \in \mathbb{R}^n \mid x(i) > 0, i = 1, \dots, n\}$ the set of all positive vectors. $\mathcal{M}$ is an open set in $\mathbb{R}^n$ which is trivially a manifold over $\mathbb{R}^n$, having $\mathbb{R}^n$ itself as tangent space at each point. We shall use the standard notation TM_x to stress this point.

COMMENT. As a collateral detail we mention that $\mathcal{M}$ is a vector space, in which the (commutative) group operation is given by the componentwise multiplication, and the standard multiplication by scalars is given by $(a, x) \in (\mathbb{R}, \mathcal{M}) \rightarrow x^a$. This detail helps to intuitively understand Theorem 3.2.

Here $\mathcal{M}$ plays the role that the positive definite matrices play in the works by Lang, Lawson-Lim and Mohaker mentioned in the Introduction. The role of the group of invertible matrices in those references is to be played here by $G = \{g \in \mathbb{R}^n \mid g(i) \neq 0, i = 1, \dots, n\}$, which clearly is an Abelian group respect to the standard product, in which the identity, denoted by e , is the vector with all components equal to 1. We shall make use the action $G : \mathcal{M} \rightarrow \mathcal{M}$ of G on $\mathcal{M}$ defined by $\tau_g(x) = g^{-1}xg^{-1}$. This action is clearly transitive on $\mathcal{M}$, and can be defined in the obvious way as an action on $\mathbb{R}^n$.

The transitivity of the action allows us to transport the scalar product on TM_e to any TM_x as follows. The scalar product between ξ and η at TM_e is defined to be the standard Euclidean product $(\xi, \eta) = \sum \xi_i \eta_i$, where we shall switch between $\xi(i)$ and ξ_i as typographical convenience dictates. Since $x = \tau_g(e)$ with $g = x^{-1/2}$. We define the scalar product transported to TM_x by

$$(\xi, \eta)_x \equiv (x^{-1}\xi, x^{-1}\eta) = (x^{-2}\xi, \eta).$$

This scalar product allows us to define the length of a differentiable curve as follows:

Let $x(t)$ be a differentiable curve in $\mathcal{M}$, its length is given by

$$\int_0^1 \sqrt{(\dot{x}, \dot{x})_x} dt.$$

With this definition, the distance between $x_1, x_2 \in \mathcal{M}$ is defined by the expected

$$d(x_1, x_2) = \inf \left\{ \int_0^1 \sqrt{(\dot{x}, \dot{x})_x} dt \mid x(t) \text{ differentiable such that } x_1 = x(0) \ x_2 = x(1) \right\} \quad (8.1)$$

It takes an application of the Euler-Lagrange formula to see that the equation of the geodesics in this metric is

$$\ddot{x}(t) = x^{-1}\dot{x}^2, \quad x(0) = x_1, \quad x(1) = x_2, \quad (8.2)$$

the solution to which is

$$x(t) = x_1 e^{t \ln(x_2/x_1)} = x_2 x_1^{(1-t)}. \quad (8.3)$$

This allows us to compute the distance between x_1 and x_2 as

$$d(x_1, x_2)^2 = \sum_{i=1}^n (\ln x_1(i) - \ln x_2(i))^2. \quad (8.4)$$

Similarly, the solution to (8.2) subject to $x(0) = x$, and $\dot{x}(0) = \xi$ is the (exponential) mapping $xe^{t\xi}$. With this notations we recall some results (in this simpler setup) from Chapter 5 of Lang (1995) under

THEOREM 8.1. With the notations introduced above we have:

- 1) The exponential mapping is metric preserving through the origin.
- 2) The derivative of the exponential mapping is measure preserving, that is, $\exp'(\xi)v = ve^\xi$ as a mapping $TM_x \rightarrow TM_{x \exp \xi}$, satisfies

$$(v, v) = (\exp'(\xi)v, \exp'(\xi)v)_{\exp(\xi)}$$

3) With the metric given by (1.1), $\mathcal{M}$ is a Bruhat-Tits space, that is, it is a complete metric space in which the semi-parallelogram law holds. This means that, given any $x_1, x_2 \in \mathcal{M}$, there exists a unique $z \in \mathcal{M}$ such that for any $y \in \mathcal{M}$ the following holds

$$d(x_1, x_2)^2 + 4d(z, y)^2 \leq 2d(y, x_1)^2 + 2d(y, x_2)^2.$$

COMMENT.

1) The action τ_g defined a few paragraphs above coincides with parallel transport along geodesics.

2) The proofs take some space but are systematic and computational. In our case, commutativity makes things considerably simpler. The completeness of $\mathcal{M}$ is transferred from $\mathbb{R}^n$ via the exponential mapping.

3) The point z mentioned in item (3) is given by $z = \sqrt{x_1 x_2}$. Actually, a simple calculation provides the proof of the following slightly more general statement.

LEMMA 8.1. Let $x_1, \dots, x_K$ be K points in $\mathcal{M}$. The point $\bar{x}_\ell$ that minimizes the sum of logarithmic distances (1.1) to the given points is given by their geometric mean, that is

$$\bar{x}_\ell = \left(\prod_{j=1}^K x_j \right)^{1/K}$$

ACKNOWLEDGEMENTS

I wish to thank the reviewers/editors for their comments to improve the presentation.

REFERENCES

- [1] Arsigny, V., Fillard, P., Pennec, X. and Ayach, N. (2007). *Geometric Means in a Novel Vector Space Structure on Symmetric positive definite matrices*, SIAM J. Matrix Theory, **29**, 328-347.
- [2] Borkhar, V. *Probability Theory*, Springer, New York, (1995).
- [3] Hsu, E.P. *Stochastic Analysis on Manifolds*, Amer. Math. Soc., Providence, (2002).
- [4] Jacod, J. and Protter, P. *Probability Essentials*, Springer, New York, (2000).
- [5] Kunita, H. and Watanabe, S. *Stochastic Differential Equations and Diffusion Processes*, North Holland Pub. Co, Amsterdam, (1989).
- [6] Lang, S. *Math talks for undergraduates*, Springer, New York, (1999).
- [7] Lawson, J.D. and Lim, Y. (2001). *The Geometric mean, matrices, metrics and more*, Amer. Math., Monthly, **108**, 797-812.
- [8] Luenberger, D.G. *Investment Science*, Princeton Univ. Press, Princeton, (1980).
- [9] Mohaker, M. *A differential geometric approach to the geometric mean of symmetric positive definite matrices*, SIAM Jour. Matrix Analysis & Applic., **26**, 735-747.
- [10] Schwartzman, A. (2015). *Lognormal distribution and geometric averages of positive definite matrices*, Int. Stat. Rev., **84**, 456-486.
- [11] Shiryaev, A.N, *Essentials of Stochastic Finance. Facts, Models, Theory*, World Scient. Pubs. (1999).

Henryk Gzyl
Centro de Finanzas IESA, Caracas, (Venezuela)
email: henryk.gzyl@iesa.edu.ve