

The concept of exchangeability in ensemble forecasting

J. Bröcker and H. Kantz

Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Strasse 38, 01187 Dresden, Germany

Received: 24 August 2010 – Revised: 1 December 2010 – Accepted: 12 December 2010 – Published: 5 January 2011

Abstract. A set of random variables is exchangeable if its joint distribution function is invariant under permutation of the arguments. The concept of exchangeability is discussed, with a view towards potential application in evaluating ensemble forecasts. It is argued that the paradigm of ensembles being an independent draw from an underlying distribution function is probably too narrow; allowing ensemble members to be merely exchangeable might be a more versatile model. The question is discussed whether established methods of ensemble evaluation need alteration under this model, with reliability being given particular attention. It turns out that the standard methodology of rank histograms can still be applied. As a first application of the exchangeability concept, it is shown that the method of minimum spanning trees to evaluate the reliability of high dimensional ensembles is mathematically sound.

1 Introduction

A widely employed means to convey probabilistic forecast information are ensembles. Ensembles are particularly popular for forecasting dynamical processes. In this situation, the ensemble members are generated by running a model of the dynamics into the future. Heterogeneity of the ensemble members is commonly achieved by slightly different initial conditions as well as heterogeneity of the model itself. In effect, the ensemble members are considered “equally likely” scenarios of the future evolution of the process under concern (see for example Toth et al., 2003)¹.



Correspondence to: J. Bröcker
(broecker@pks.mpg.de)

¹This interpretation does not apply to situations in which the ensemble members are obviously distinguishable, for example if they are generated using significantly different models (Poor Man’s Ensemble). Such ensembles are not considered here.

Opinions differ slightly as to how precisely ensembles are to be interpreted, but a common view is the following, often referred to as *Monte-Carlo* interpretation (Stephenson and Doblas Reyes, 2000). An *ensemble* is a collection X_1, \dots, X_K of random variables, drawn independently from a common distribution function F_θ , the *forecast distribution*. The forecast distribution depends on a parameter θ , which can be considered as the (random) state of the forecasting scheme. In other words, F_θ is the distribution of the ensemble members conditioned on the current state of the forecasting scheme being θ . The details of the forecasting scheme are of no interest in the present study; we can imagine it as some information processing device providing either a forecast distribution or an ensemble. In fact, often the forecast distribution F_θ is but a mental construct and not operationally available. The *verification* is yet another random variable X_0 which is independent of X_1, \dots, X_K , conditionally on θ .

The forecasting scheme is called *reliable* if the distribution of X_0 , conditionally on θ , is given by F_θ . The Monte-Carlo interpretation implies that F_θ is reliable if and only if conditionally on θ , $X_0 \dots X_K$ are independent random variables with common distribution function F_θ . Less formally stated, the verification behaves like just another ensemble member. A consequence of reliability is that the rank of X_0 among all ensemble members assumes the values $1, \dots, K + 1$ with equal probability (namely $1/(K + 1)$), either conditionally or unconditionally on θ . (The rank of X_0 is defined as the number of indices n so that $X_n \leq X_0$; the concept of rank obviously makes sense only if the X_k are scalars.) This fact is used to build tests for reliability. More specifically, reliability implies that for $l = 1, \dots, K + 1$

$$\mathbb{P}(\text{rank}(X_0) = l | \theta) = \frac{1}{K + 1}, \quad (1)$$

Let $T_i, i = 1, \dots, L$ sets covering the range of the parameter θ . Then a consequence of Eq. (1) is

$$\mathbb{P}(\text{rank}(X_0) = l | \theta \in T_i) = \frac{1}{K + 1}. \quad (2)$$

This means that conditioned on $\theta \in T_i$, the histogram of $\text{rank}(X_0)$ should be flat. This can be statistically tested, see for example Bröcker (2008). (A technical difficulty, ignored here, is that θ is typically not available operationally.) The relation (2) is only necessary for reliability, not sufficient; see Bröcker (2008); Hamill (2001) for a discussion.

There are two issues with this rank based reliability analysis, both of which lead us to study *exchangeable* ensembles. Random variables are called exchangeable if their distribution function is symmetric. In formulae, the random variables X_1, \dots, X_K are exchangeable if

$$M(x_1, \dots, x_K) = M(x_{\pi(1)} \dots x_{\pi(K)}),$$

where M is the joint distribution function of X_1, \dots, X_K and π is an arbitrary permutation of K elements (a definition for non-scalar random variables will be given in Sect. 4.) The first issue with the rank approach is that it obviously cannot be used in higher dimensions. Hansen and Smith (2004) suggested an alternative approach based on the minimum spanning tree. Effectively, the high dimensional ensemble is condensed into an ensemble with real valued members. The ensemble members are however not independent; yet they are exchangeable, due to the fact that the minimum spanning tree is a symmetric function. This is looked at in detail in Sect. 2. The second issue is that in the Monte-Carlo interpretation, even though the ensemble members are independent conditionally on θ , they are generally *not* independent conditionally on $\theta \in T_i$. More precisely, we have that

$$\mathbb{P}(\{X_1 < x_1, \dots, X_K < x_K\}|\theta) = F_\theta(x_1) \cdot \dots \cdot F_\theta(x_K),$$

is a product, while $\mathbb{P}(\{X_1 < x_1, \dots, X_K < x_K\}|\theta \in T_i)$ is not. Equation (5) below gives the exact expression for this probability, from which we gather that it is still symmetric. That is, the ensemble remains exchangeable under this conditioning. If the ensemble was only exchangeable in the first place, conditioning would not destroy that property. This is looked at in detail in Sect. 3.

These two issues raise the question as to whether the Monte-Carlo interpretation should be replaced by the weaker assumption that the ensemble members are exchangeable random variables. It is demonstrated (mainly in Sect. 4) that in doing so, not much will be lost. Most importantly, the property (1) of the rank distribution is still valid (for a reliable ensemble), whence the rank based reliability analysis can be applied to exchangeable ensembles. In particular, the analysis shows that the minimum spanning tree approach is mathematically sound.

2 Evaluation of high dimensional ensemble forecasts

If the ensemble members are elements of a high dimensional vector space V , the rank is no longer defined, and the approach to testing reliability as outlined in the Introduction ceases to apply. One possible approach is to choose

a function $\phi : V \rightarrow \mathbb{R}$ and check the reliability of the ensemble $\{\phi(X_1) \dots \phi(X_K)\}$ with respect to the verification $\phi(X_0)$. Obviously, this is but a necessary test for the original problem, since by applying the function ϕ , we might have neglected various degrees of freedom of the problem.

Another approach was proposed by Hansen and Smith (2004). Let $\phi : V^K \rightarrow \mathbb{R}$ a function which is symmetric, that is, if π is any permutation of the numbers $\{1, \dots, K\}$, then $\phi(v_1, \dots, v_K) = \phi(v_{\pi(1)}, \dots, v_{\pi(K)})$. In Hansen and Smith (2004), $\phi(v_1, \dots, v_K)$ is the length of the *minimum spanning tree*, that is the shortest graph connecting the points v_1, \dots, v_K . Using ϕ , a new verification ξ_0 and a new ensemble ξ_1, \dots, ξ_K with values in \mathbb{R} are generated as follows:

$$\xi_k := \phi_k(X_0, \dots, X_K) \quad k = 0 \dots K, \quad (3)$$

where

$$\phi_k(X_0, \dots, X_K) := \phi(X_0, \dots, X_{k-1}, X_{k+1}, \dots, X_K). \quad (4)$$

In other words, ξ_k is the length of the minimum tree spanning $X_0 \dots X_K$ with X_k left out. The problem now is that the random variables $\xi_0 \dots \xi_K$ are not any longer independent, even if the original variables $X_0 \dots X_K$ are. Indeed, we would expect them to be very highly dependent, since any two of them have all but one argument in common. It is therefore not clear if the rank of ξ_0 within ξ_1, \dots, ξ_K has uniform distribution. What rescues the approach though is the fact that, under the hypothesis of reliability, the variables $\xi_0 \dots \xi_K$ are exchangeable. As will be shown in Sect. 4, the rank histogram can then still be expected to be flat. To see that the variables $\xi_0 \dots \xi_K$ are exchangeable, note that it follows readily from the definitions that permuting the $X_0 \dots X_K$ has the same effect as applying the same perturbation to the $\xi_0 \dots \xi_K$. (The symmetry of the function ϕ is crucial here.) Since the distribution of the $X_0 \dots X_K$ is invariant under permutations, the same holds thus for $\xi_0 \dots \xi_K$, which are therefore exchangeable random variables. Note also that we are not using the independence of the X_k , but only that they are exchangeable themselves. Thus for the ξ_k to be exchangeable it is sufficient that the X_k are exchangeable. In other words, building the minimum spanning trees, exchangeability of ensemble members is preserved, although not independence.

3 Forecast stratification

As mentioned in the Introduction, even though the ensemble members are independent conditionally on θ , they are generally *not* independent conditionally on $\theta \in T$, where T is some subset of the range of θ . Indeed, suppose the X_1, \dots, X_K are independent given θ , and let G be the distribution of θ . Then we have the relation

$$\begin{aligned} \mathbb{P}(\{X_1 < x_1, \dots, X_K < x_K\}|\theta \in T) \\ = \frac{1}{c} \int_T F_\theta(x_1) \cdot \dots \cdot F_\theta(x_K) dG(\theta), \quad (5) \end{aligned}$$

with

$$c = \int_T dG(\theta).$$

In general, the expression (5) cannot be written as a product with the individual factors being one-dimensional distribution functions. In any event though, the right hand side of (5) is symmetric, that is, invariant with respect to permutations of the x_i . Thus, conditionally on θ , the ensemble is exchangeable. Exchangeability though is a convex property, that is, it is preserved when averaging over distribution functions. More precisely, let the ensemble X_1, \dots, X_K be a draw from a symmetric distribution function $M_\theta(x_1, \dots, x_K)$, where as before θ is itself random. If again G is some distribution function for θ , then the average

$$\bar{M}(x_1, \dots, x_K) = \int M_\theta(x_1, \dots, x_K) dG(\theta), \quad (6)$$

is again a symmetric distribution function. We can therefore conclude that any averaging over θ does not destroy the exchangeability of the ensemble. What we learn from these considerations is that the classical Monte-Carlo interpretation gives independent ensemble members only if we condition on the full model state θ , while any other conditioning results in merely exchangeable random variables. As a consequence, we might want to study exchangeable random variables, and we might want to drop the Monte-Carlo interpretation and assume merely exchangeability of the ensemble members in the first place. The next section demonstrates that not much is lost in doing so.

4 Exchangeable random variables

Let X_1, \dots, X_K be a finite series of real valued random variables. As mentioned in the Introduction, the random variables are said to be exchangeable if their joint distribution function is symmetric. If the X_i have values in \mathbb{R}^d , they are defined as exchangeable if for any family of sets $\{A_k \subset \mathbb{R}^d; k = 1 \dots K\}$ and any permutation π of the elements $\{1, \dots, K\}$, it holds that

$$\mathbb{P}\{X_k \in A_k, k = 1, \dots, K\} = \mathbb{P}\{X_{\pi(k)} \in A_k, k = 1, \dots, K\}, \quad (7)$$

A sub-selection of exchangeable random variables is exchangeable. Similarly, an infinite series X_1, X_2, \dots of random variables is defined to be exchangeable if the first k members $\{X_1, \dots, X_k\}$ are exchangeable for any $k \in \mathbb{N}$.

A few properties of exchangeable random variables are discussed. Exchangeable random variables are identically distributed. More generally, infinite series of exchangeable random variables have shift invariant distributions and therefore form stationary processes. To see this, note that

$$\begin{aligned} \mathbb{P}(\{X_k \in A_k, k = 1, \dots, K\} \cup \{X_{K+1} \in \mathbb{R}^d\}) \\ = \mathbb{P}(\{X_{k+1} \in A_k, k = 1, \dots, K\} \cup \{X_1 \in \mathbb{R}^d\}) \end{aligned}$$

by cyclic permutation of the X_1, \dots, X_{K+1} and using exchangeability. The left and right hand sides are, respectively, $\mathbb{P}\{X_k \in A_k, k = 1, \dots, K\}$ and $\mathbb{P}\{X_{k+1} \in A_k, k = 1, \dots, K\}$, proving the shift invariance of the distribution.

Next we will look at the distribution of ranks of exchangeable random variables, which is relevant for reliability tests, and also for the general interpretation of ensembles. Let again π be a permutation of the numbers $\{1, \dots, K\}$. For any such permutation, we have the event $A_\pi := \{X_{\pi(1)} < \dots < X_{\pi(K)}\}$, that is, the event that the $X_{\pi(k)}, k = 1, \dots, K$ are ordered. Obviously, the A_π are disjoint for different permutations π . Due to exchangeability, the A_π 's all have the same probability. This fact supports the intuitive notion that the order in which ensemble members come is irrelevant and does not carry any information. Furthermore, if the joint distribution function of the X_k 's is continuous, then $X_k = X_j$ for $k \neq j$ happens with probability zero only, and therefore $\mathbb{P}(\cup A_\pi) = 1$, where the union runs over all permutations π of the numbers $\{1, \dots, K\}$. Since there are $K!$ such permutations, it must be that $\mathbb{P}(A_\pi) = 1/K!$, that is, each one of possible orderings of the X_k 's occurs with probability $1/K!$. There are exactly $(K-1)!$ permutations π with $\pi(k) = l$ for arbitrary but fixed $k, l \leq K$. Hence, the probability that X_k has rank l is equal to $\frac{(K-1)!}{K!} = \frac{1}{K}$. In other words

$$\mathbb{P}(\text{rank}(X_k) = l) = \frac{1}{K}. \quad (8)$$

This result can be derived as a consequence of the following statement, which is of interest on its own and easy to prove (see Lehmann, 1959). Write $X_{[k]}$ for the k -th order statistics, that is, the random variable at position k when X_1, \dots, X_K are put in ascending order. Then for any function $f(x_1, \dots, x_K)$ it holds that

$$\begin{aligned} \mathbb{E}(f(X_1, \dots, X_K) | X_{[k]} = x_k, k = 1, \dots, K) \\ = \frac{1}{K!} \sum_{\pi} f(x_{\pi(1)} \dots x_{\pi(K)}), \quad (9) \end{aligned}$$

where the sum ranges over all permutations of K elements. A noteworthy fact is that the right hand side is independent of the actual distribution of the X_k . To prove relation (8) using (9), fix some m and l and apply (9) to

$$f(x_1, \dots, x_K) = \begin{cases} 1 & \text{if rank}(x_m) = l \\ 0 & \text{otherwise} \end{cases}$$

With this choice, the left hand side of relation (9) becomes equal to $\mathbb{P}(\text{rank}(X_m) = l | X_{[k]} = x_k, k = 1, \dots, K)$. On the right hand side of relation (9), the sum just counts the number of permutations π with $\pi(m) = l$, of which there are exactly $(K-1)!$, as noted before. Thus the right hand side of (9) gives $1/K$; we have established

$$\mathbb{P}(\text{rank}(X_m) = l | X_{[k]} = x_k, k = 1, \dots, K) = 1/K.$$

This implies (8).

5 De Finetti's theorem

For an infinite series of exchangeable random variables, the following remarkable fact (known as de Finetti's theorem, see Kingman, 1978; Hewitt and Savage, 1955) has been demonstrated. There exists a random variable θ with some distribution G so that the representation

$$\mathbb{P}\{X_1 < x_1, \dots, X_K < x_K\} = \int F_\theta(x_1) \cdot \dots \cdot F_\theta(x_K) dG(\theta),$$

holds for all finite K , with $F_\theta(x) := \mathbb{P}(X_k < x | \theta)$. In particular, an infinite series of exchangeable random variables is always independent conditionally on θ . This theorem is false though for only finite series of exchangeable random variables.

Of course, we would like to view an infinite number of variables as an idealisation of a sufficiently large but finite number of variables. In the present context, the statement of de Finetti implies that if we start with an exchangeable ensemble, then in the limit of infinitely many members, the ensemble can in fact be considered as independently drawn from some underlying random distribution function. Does that mean that for sufficiently large ensembles, we are back in the classic Monte-Carlo interpretation of ensembles? And how large is "sufficiently large"? Any attempt to answer this question will have to incorporate details of the ensemble generation process.

In order to appreciate the difficulties involved, consider the following example, which shows that "just large" is not enough. An urn contains N balls, either red or blue, in proportion θ , which is itself considered random. Drawing either with or without replacement results in an exchangeable sequence of red and blue balls. Furthermore, if θ is given, drawing *with* replacement yields an independent series of red and blue balls, so de Finetti's theorem is true here even for only finite number of draws. This is not the case for drawing *without* replacement. The colour of the very last ball, for example, is known after $N - 1$ draws; given θ , the draws are *not* independent. If only a few balls are drawn though from a very large urn, then drawing with and without replacement essentially amounts to the same. Consequently, the draws would be essentially independent in both cases. Thus, in order that drawing without replacement gives an independent series, the essential bit is not that the number of draws is large, but that the number of draws is *small* compared to the number of balls left in the urn. As de Finetti's theorem assumes an infinite series of draws, the urn has to contain an infinite amount of balls in the first place. Given such a generous urn, it does not matter if we draw with or without replacement, the series is independent in either case, and de Finetti's theorem becomes true.

We learn from this example that de Finetti's theorem is *not* an idealisation of a situation with many draws from an urn, but rather of a situation with comparably *few* draws from an urn containing a very large number of balls. The implications

this has for the application of de Finetti's theorem to ensembles given typical procedures of ensemble generation will be subject to future research.

6 Conclusion

Starting from the standard Monte-Carlo interpretation of ensembles, it was shown that the ensemble members retain their independence only if their distribution conditional on the latent variable is considered. Otherwise, the independence is destroyed, and they are merely exchangeable. This leads us to study exchangeable random variables. A central result is that under exchangeability, any permutation of the ensemble members is equally likely, which formalises the intuitive understanding that the order in which ensemble members come is irrelevant. In addition, this implies that a given ensemble member assumes any rank with equal probability. This fact is at the basis of standard reliability tests such as Talagrand histograms.

Furthermore, we discussed de Finetti's theorem, which essentially asserts that an infinitely large ensemble with exchangeable members can be considered a Monte-Carlo ensemble. The usefulness of this result though for merely finite ensembles is questionable.

To demonstrate the applicability of the concept, the method of minimum spanning trees, suggested by Hansen and Smith (2004) to evaluate the reliability of large dimensional ensembles, was revisited. It was shown that the method amounts to forming a "new ensemble", which is not independent but still exchangeable. We therefore concluded that the corresponding histograms can be expected to be flat.

Acknowledgements. The authors acknowledge fruitful discussions with the members of the time series analysis group at the Max-Planck Institute, in particular Stefan Siebert.

The service charges for this open access publication have been covered by the Max Planck Society.

Edited by: O. Talagrand

Reviewed by: two anonymous referees

References

- Bröcker, J.: On reliability analysis of multi-categorical forecasts, *Nonlin. Processes Geophys.*, 15, 661–673, doi:10.5194/npg-15-661-2008, 2008.
- Hamill, T. M.: Interpretation of rank histograms for verifying ensemble forecasts, *Mon. Weather Rev.*, 129(3), 550–560, 2001.
- Hansen, J. A. and Smith, L. A.: Extending the limits of forecast verification with the minimum spanning tree, *Mon. Weather Rev.*, 132(6), 1522–1528, 2004.
- Hewitt, E. and Savage, L. J.: Symmetric measures on cartesian products, *T. Am. Math. Soc.*, 80(2), 470–501, 1955.
- Kingman, J. F. C.: Uses of exchangeability, *The Annals of Probability*, 6(2), 183–197, 1978.

Lehmann, E. L.: Testing statistical hypotheses, John Wiley & Sons, Inc., 1959.

Stephenson, D. B. and Doblas Reyes, F. J.: Statistical methods for interpreting monte carlo ensemble forecasts, *Tellus, series A*, 52(3), 300–322, 2000.

Toth, Z., Talagrand, O., Candille, G., and Zhu, Y.: Probability and ensemble forecasts, in: *Forecast Verification*, edited by: Jolliffe, I. T. and Stephenson, D. B., chapter 7, pages 137–163, John Wiley & Sons, Ltd., Chichester, 2003.