The relation between the closeness of random variables and their distributions

by Péter Major
Mathematical Institute of the Hungarian Academy of Sciences

Abstract: If two random variables are close to each other, then the same relation holds for their distributions. This statement cannot be reversed. Nevertheless, some non-trivial results can be proved in this in the opposite direction if the question is formulated in the right way. Given two probability measures on a general metric space let us try to construct two random variables with these distributions which are close to each other. In this problem the distribution of the random variables are prescribed, but we have the freedom to "couple them", to define their joint distribution at our taste.

We show that the questions how close two probability measures are to each other and how close random variables can be constructed with such distributions are closely related. Such problems will be discussed here. In particular, we introduce the notion of the Prochorov metric, the quantile transform and discuss their most important properties. It is worth mentioning that these results are strongly related to a classical result in combinatorics, the König–Hall theorem.

Introduction

If two random variables are close to each other, then the same relation holds for their distributions. The converse statement does naturally not hold. For instance these two random variables can be independent. But an interesting and non-trivial answer can be given to the following question. Let two (close) probability measures μ and ν be given on the Borel σ-algebra of a metric space (X, ρ). Is it possible to construct a probability space $(\Omega, \mathcal{A}, P)$ and two random variables ξ and η on it in such a way that the distribution of ξ is μ, the distribution of η is ν, and the random variables ξ and η are close to each other? Naturally, in a detailed discussion we must tell explicitly what we mean by closeness of probability measures and random variables.

In the study of this problem the following approach is natural: Put $(\Omega, \mathcal{A}, P) = (X \times X, \mathcal{A} \times \mathcal{A}, P)$, where $\times$ denotes direct product, $\mathcal{A}$ is the σ-algebra induced by the metric ρ, or more precisely by the topology generated by it, and P is an appropriate probability measure on the space $(X, \mathcal{A})$. Furthermore, let us define the random variables $\xi(x_1, x_2) = x_1$ and $\eta(x_1, x_2) = x_2$, $(x_1, x_2) \in X \times X$, on this space. After the introduction of these objects the problem of constructing random variables with prescribed distributions close to each other can be reformulated in the following way: Let us construct a probability measure P on the space $(X \times X, \mathcal{A} \times \mathcal{A})$ whose marginal distributions are the measures μ and ν, and which is essentially concentrated close to the diagonal $\{(x, x),\ x \in X\}$. In the investigation of this problem a classical combinatorial result, the König–Hall theorem (sometimes called the marriage problem), or more precisely a continuous version of this result is very useful. We formulate these results, and also write down their proofs in an Appendix.

König–Hall theorem, (marriage problem). *Let us consider n boys and n girls such that some boys and girls know each other. (All acquaintances are mutual.) We want to put these boys and girls into pairs (make married couples) in such a way that all persons put into a pair know each other. This is possible if and only if an arbitrary group of the girls knows at least as many boys as the number of this group.*

In a more formal way: Let us consider a bipartitated graph consisting of two sets $Y = \{y_1, \dots, y_n\}$ and $Z = \{z_1, \dots, z_n\}$ and a map $Y \times Z \to \{0, 1\}$. We can interpret the relation $d(y, z) = 1$, $y \in Y$, $z \in Z$ so that the points y and z are connected, while $d(y, z) = 0$ means that they are not connected. For all sets $A \subset Y$ let us define the set $B(A) \subset Z$ which contains the points which are connected to one of the points of A, i.e. let

$$B(A) = \{z\colon z \in Z, \text{ and there exists such an } y \in A, \text{ for which } d(y, z) = 1\}.$$

There exists a factorization of this bipartitated graph, i.e. we can divide the points of the sets Y and Z into pairs $(y_j, z_{\pi(j)})$, $y_j \in Y$, $z_{\pi(j)} \in Z$, $j = 1, 2, \dots, n$, in such a way that $d(y_j, z_{\pi(j)}) = 1$ for all $j = 1, 2, \dots, n$, and $\pi(j)$, $j = 1, \dots, n$, is an appropriate permutation of the set $\{1, \dots, n\}$ if and only if $|B(A)| \ge |A|$ for all sets $|A| \subset Y$, where $|C|$ denotes the cardinality of a set C.

The continuous version of the König–Hall theorem. *Let r depots with stocks $u_1, u_2, \dots, u_r$ and s plants with claims $v_1 \dots, v_s$ be given such that $\sum\limits_{j=1}^{r} u_j = \sum\limits_{k=1}^{s} v_k$. Let certain depots and plants be connected by a route. We can satisfy all claims of the plants by transporting the stocks of the depots on these routes if and only if the joint demand of an arbitrary group of plants is not greater than the joint stock of the depots which are connected by a route with one of these plants.*

In a more formal way: Let us consider a bipartitated graph consisting of two sets $Y = \{y_1, \dots, y_r\}$ and sets $Z = \{z_1, \dots, z_s\}$ and a map $d\colon Y \times Z \to \{0, 1\}$. We connect two points y and z, $y \in Y$, $z \in Z$, if $d(y, z) = 1$, and we do not connect them if $d(y, z) = 0$. Furthermore, let two weight function $u(y)$, $u(y) \ge 0$, $y \in Y$ and $v(z)$, $v(z) \ge 0$, $z \in Z$ be given such that $\sum\limits_{y \in Y} u(y) = \sum\limits_{z \in Z} v(z)$. For all sets $A \subset Y$ let us define the set $B(A) \subset Z$ by the formula

$$B(A) = \{z\colon z \in Z, \text{and there exists such an } y \in A, \text{ for which } d(y, z) = 1\}.$$

There exists a "transport function" $w(y, z) \ge 0$ with the properties

i.) $\sum\limits_{z\colon\, d(y,z)=1} w(y, z) = u(y)$ *for all* $y \in Y$,
and $\sum\limits_{y\colon\, d(y,z)=1} w(y, z) = v(z)$ *for all* $z \in Z$.

ii.) The inequality $w(y, z) > 0$ holds only if $d(y, z) = 1$,

if and only if the relation $\sum\limits_{z \in B(A)} v(z) \ge \sum\limits_{y \in A} u(y)$ holds for all sets $A \subset Y$.

Problems

0.) Let us show that the conditions in the König–Hall theorem and in its continuous version are symmetric for the sets Y and Z, i.e. these conditions remain valid, if we replace in them (and in the definition of the set $B(A)$) the sets Y and Z.

1.) Let two probability measures μ and ν be given on the σ-algebra determined by the topology generated by the metric of a separable metric space (X, ρ). Let $B^\alpha = \{x\colon \rho(x, B) < \alpha\}$ denote the open neighborhood of the set $B \subset X$ of radius α. Let us assume that the measures μ and ν satisfy the condition $\mu(B) \le \nu(B^\alpha) + \beta$ for all closed sets $B \subset X$ with some numbers $\alpha > 0$ and $\beta > 0$. Then for all $\varepsilon > 0$ a probability space $(\Omega, \mathcal{A}, P)$ can be constructed, together with two random variables ξ and η on this probability space with values in the space (X, ρ) whose distributions are μ and ν respectively, and which satisfy the relation

$$P(\rho(\xi, \eta) > \alpha + \varepsilon) \le \beta + \varepsilon \tag{a}$$

Conversely, if two random variables ξ and η with distributions μ and ν respectively satisfy relation (a), then $\mu(B) \le \nu(B^{\alpha+\varepsilon}) + \beta + \varepsilon$ for all closed sets B.

If X is not only a separable, but also a complete separable metric space, then under the above conditions relation (a) also holds with $\varepsilon = 0$. (In the proof of this statement we may apply the result by which a uniformly compact sequence of probability measures on a complete metric space has a subsequence which is convergent with respect to the weak convergence of probability measures.)

2.) Let a separable metric space (X, ρ) be given together with the Borel σ-algebra $\mathcal{A}$ induced by the natural topology of this space. Let $\mathcal{M}$ denote the space of probability measures on the space $(X, \mathcal{A})$, and let us introduce the following function $d(\cdot, \cdot)$ on the pairs of probability measures on the space (X, ρ): If $\mu \in \mathcal{M}$ and $\nu \in \mathcal{M}$, then

$$d(\mu, \nu) = \inf\{\alpha\colon \mu(B) \le \nu(B^\alpha) + \alpha \text{ for all closed sets } B \subset X\},$$

where B^α has the same meaning as in the previous problem. Let us show that $d(\cdot, \cdot)$ is a metric on the space $\mathcal{M}$ which metrizes the weak convergence in this space, i.e. the relation $\mu_n \Rightarrow \mu$, as $n \to \infty$, holds for a sequence of measures $\mu_n \in \mathcal{M}$, $n = 1, 2, \dots$, and $\mu \in \mathcal{M}$, where $\Rightarrow$ denotes weak convergence, if and only if $d(\mu_n, \mu) \to 0$. The space $(\mathcal{M}, d)$ is a separable metric space, and if (X, ρ) is a complete separable metric space, then the same property holds for $(\mathcal{M}, d)$.

Remark: The property whether a metric generating a topology is complete or non-complete is not topologically invariant, i.e. it is possible that a complete and a non-complete metric generate the same topology on a space. There is a classical result by which the space of probability measures on a complete metric space can be endowed with a metric which induces weak convergence, and with which the space of probability measures is a complete metric space. By the above remark this result does not imply automatically that also the metric introduced in problem 2 has this property.

We shall also prove the following *Statement A:*

Statement A: *Let us apply the notations of problem 2. If $\mu_n \in \mathcal{M}$, $n = 1, 2, \dots$, is a sequence of probability measures on a metric space (X, ρ), and this sequence of measures together with a probability measure $\mu \in \mathcal{M}$ satisfy the relation $\mu_n \Rightarrow \mu$, then there exists a probability space $(\Omega, \mathcal{A}, P)$ and a sequence of random variables ξ_n, $n = 1, 2, \dots$, together with a random variable ξ on this space in such a way that the distribution of ξ_n is μ_n, $n = 1, 2, \dots$, the distribution of ξ is μ, and $\xi_n \to \xi$ with probability one.*

The proof of *Statement A* is based on the observation that in an appropriate construction it can be achieved that the sets depending on the index n where the distance between the random variables ξ_n and ξ is relatively large almost overlap each other. This can be achieved if the joint distribution of the random variables ξ_n is appropriately chosen. As a consequence, the almost sure convergence in *Statement A* is less useful than it may seem at first sight. The deep results of the probability theory containing "with probability one ... " statements depend on the joint distribution of the random variables. On the other hand, *Statement A* allows the changement of the joint distribution of the random variables.

The proof of *Statement A* is simpler in complete separable spaces, and in this case the probability space where the convergent sequence of random variables is constructed can be chosen in a very special way. To prove *Statement A* in this special case it is useful first to solve the following problem.

3.) Let (X, ρ) be a complete metric space, μ a probability measure on the Borel σ-algebra of this space. Let us consider the special probability space $(\Omega, \mathcal{A}, P)$ for which $\Omega = [0, 1]$, $\mathcal{A}$ is the Borel σ-algebra on the interval $[0, 1]$, and P is the Lebesgue measure on the Borel σ-algebra of the interval $[0, 1]$. On this probability space a random variable can be constructed with values on the space (X, ρ) whose distribution is μ.

4.) Let us prove *Statement A* in the case when (X, ρ) is a complete separable metric space. Show that in this case the probability space $(\Omega, \mathcal{A}, P)$ where the convergent sequence of random variables is constructed can be chosen in the following special way: $\Omega = [0, 1]$, $\mathcal{A}$ is the Borel σ-algebra on the interval $[0, 1]$, and P is the Lebesgue measure on the Borel σ-algebra of the interval $[0, 1]$.

5.) *Statement A* also holds for convergent sequence of probability measures on an arbitrary separable (not necessarily complete) metric space. (In this space the sequence of convergent random variables has to be constructed on an appropriate (generally very large) probability space $(\Omega, \mathcal{A}, P)$.)

6.) Let ξ_n, $n = 1, 2, \dots$, and ξ be (X, ρ) valued random variables, where (X, ρ) is a separable metric space. Let us assume that $\xi_n \Rightarrow \xi$, where $\Rightarrow$ denotes stochastic convergence. The distributions μ_n of the random variables ξ_n and the distribution μ of the random variable ξ satisfy the relation $\mu_n \Rightarrow \mu$, where $\Rightarrow$ denotes weak convergence of probability measures.

We have investigated the question that given two probability measures μ and ν on a metric space (X, ρ) how a pair of μ distributed ξ and ν distributed η random variables can be constructed which are close to each other. We have seen that informally saying this question leads to the to the following "transport problem": How can a system of points with mass distribution μ be transported with relatively few movements to a system of points with mass distribution ν? If the metric space (X, ρ) where we are working is the real line with the usual metric, then because of the simple structure of this space the "transport problem" we have to handle becomes considerably simpler. In this case if some "natural evaluation of the transport cost" is considered (see the subsequent problem 9, where such a problem is formulated in an explicit form) it is useful to exclude the following possibility: There exist such pairs of numbers $x_1 < x_2$ and $x_3 < x_4$ for which the point x_1 is transported to the point x_4 and the x_2 to the point x_3. In this case the transports $x_1 \to x_3$ and $x_2 \to x_4$ are more economic. The following construction, called the quantile transform, excludes the possibility of such non-economic transports.

To define the quantile transform we recall the following fact often used also in mathematical statistics. If ξ is a random variable with distribution F on the real line, then under some slight restrictions the random variable $\eta = F(\xi)$ is a random variable with uniform distribution on the interval $[0, 1]$. Conversely, if η is a random variable with uniform distribution on the interval $[0, 1]$, then $\xi = F^{-1}(\eta)$, where $F^{-1}(x)$ is the inverse of the distribution function $F(x)$ is an F distributed random variable. In the next problem we formulate the above result in a more precise and slightly more general form.

7.) Let ξ be a random variable with distribution function $F(x) = P(\xi < x)$ and η a random variable uniformly distributed in the interval $[0, 1]$. Let us define the generalized inverse of the (not necessarily strictly) monotone increasing function $F(x)$ by the formula $F^{-1}(x) = \sup\{u \colon F(u) < x\}$. Then $\bar{\xi} = F^{-1}(\eta)$ is an F distributed random variable. Conversely, let ε be a random variable with uniform distribution in the interval $[0, 1]$ which is independent of the random variable ξ. Then $\bar{\eta} = \tilde{F}(\xi, \varepsilon) = F(\xi) + \varepsilon[F(\xi + 0) - F(\xi)]$, where $F(x + 0) = \lim\limits_{h>0, h\to 0} F(x + h)$, is uniformly distributed in the interval $[0, 1]$.

8.) Let F and G be two distribution functions. If ζ is a random variable with uniform distribution on the interval $[0, 1]$, then the random variables $\bar{\xi} = F^{-1}(\zeta)$ and $\bar{\eta} = G^{-1}(\zeta)$, where the inverse functions F^{-1} are G^{-1} are defined in the same way as in the previous problems, have distributions F and G. If $\bar{\bar{\xi}}$ is a random variable with distribution F, and ε is an on the interval $[0, 1]$ uniformly distributed random variable independent of the random variable $\bar{\bar{\xi}}$, then the random variable $\bar{\bar{\eta}} = G^{-1}(\tilde{F}(\bar{\bar{\xi}}, \varepsilon))$ has distribution G. The distributions of the random vectors $(\bar{\xi}, \bar{\eta})$ and $(\bar{\bar{\xi}}, \bar{\bar{\eta}})$ agree.

In the problems of probability theory we sometimes have to construct two random variables ξ and η who have prescribed distributions μ and ν. Actually in the problems of probability theory generally it is enough to give the joint distribution of the random vector (ξ, η). Two constructions which determine random vectors with the same distribution can be considered equivalent. Hence the construction of both random vectors $(\bar{\xi},\bar{\eta})$ and $(\bar{\bar{\xi}},\bar{\bar{\eta}})$ given in the previous problem are called the quantile transform in the literature. In the next problem we formulate an optimality property of the quantile transform.

9.) Let ξ and η be two random variables on the real line with distribution functions $F(x)$ and $G(x)$, which satisfy the conditions $E|\xi|<\infty$, $E|\eta|<\infty$. Beside this, let us consider a convex function $\Phi(x)$ on the real line. Then

$$E\Phi(\xi-\eta)\ge\int_0^1\Phi\left(F^{-1}(x)-G^{-1}(x)\right)\,dx>-\infty,$$

where $F^{-1}(x)$ and $G^{-1}(x)$ are the inverse functions defined in problem 7. If the random vector is defined by means of the quantile transform, then the two sides of the above inequality are equal.

9a.) Let ξ and η be two random variables on the real line with distribution functions $F(x)$ and $G(x)$, which satisfy the conditions $E|\xi|<\infty$, $E|\eta|<\infty$. Then $E|\xi-\eta|\ge\int_{-\infty}^{\infty}|F(x)-G(x)|\,dx$, and the inequality can be replaced by identity if the pair (ξ,η) is constructed by means of quantile transform.

Finally we formulate some problems different of the previous ones which may be useful in some investigations. Their proofs apply some non-trivial facts from measure theory like the existence of conditional distribution, the Banach decomposition theorem or the Radon–Nikodym theorem which implies the latter result.

10.) Let three complete separable metric spaces (X_i,ρ_i), $i=1,2,3$ be given, and let $\mathcal{A}_i$, $i=1,2,3$, denote the σ-algebra induced by the topology of these spaces. Let μ be a probability measure on the space $(X_1\times X_2,\mathcal{A}_1\times\mathcal{A}_2)$ and ν a probability measure on the space $(X_2\times X_3,\mathcal{A}_2\times\mathcal{A}_3)$. Let us assume that the projections of the measures μ and ν to the space X_2 agree. Then there exists such a probability measure P on the space $(X_1\times X_2\times X_3,\mathcal{A}_1\times\mathcal{A}_2\times A_3)$ whose projection to the space $X_1\times X_2$ is μ, and to the space $X_2\times X_3$ is ν.

Let us remark that if (X,ρ) is a complete separable space, then the direct product of infinitely many copies of this space $X\times X\times\cdots$ can also be considered as a complete metric space with an appropriate metric $\bar{\rho}$. Indeed, we may assume that the metric in the space (X,ρ) is such that $\rho(x,\bar{x})\le1$ for all points $x\in X$ and $\bar{x}\in X$, by introducing for instance the new metric $\rho'(x,\bar{x})=\min(\rho(x,\bar{x}),1)$ if it is necessary. Then the metric

$$\bar{\rho}\left((x_1,x_2,\dots),(\bar{x}_1,\bar{x}_2,\dots)\right)=\sum_{k=1}^{\infty}\frac{1}{2^k}\rho(x_k,\bar{x}_k)$$

can be defined on the product space $X\times X\times\cdots$. The space $(X\times X\times\cdots,\bar{\rho})$ is a separable complete metric space with this metric.

It follows from the results of problem 10 that if μ and ν are two probability measures on a complete separable metric space (X,ρ) and we want to construct two random variables ξ and η with distributions μ and ν respectively which are close to each other then we can apply the following approach. We introduce some auxiliary random variables ζ and $\bar{\zeta}$ with values in the space (X,ρ) in such a way that the members of the pairs (ξ,ζ) and $\left(\eta,\bar{\zeta}\right)$ are close to each other, ξ is a μ, η is a ν distributed random variable, and the distributions of ζ and $\bar{\zeta}$ agree. Furthermore, by the remark made after problem 10 this approach can be applied also in the case if we want to approximate a series of random variables $\xi_1,\xi_2,\dots$ with a prescribed distribution with another sequence of random variables $\eta_1,\eta_2,\dots$ of possibly different distribution.

The statement of the previous paragraph can be slightly strengthened. Let us have a μ distributed random variable ξ on a sufficiently rich probability space $(\Omega,\mathcal{A},P)$ and a probability measure ν on a product space $X\times X$, where (X,ρ) is a complete separable metric space. Let us further assume that the projection of the probability measure ν to the first coordinate of the product space $X\times X$ is the distribution μ of the random variable ξ. Then a random variable η can be constructed on the probability space $(\Omega,\mathcal{A},P)$ for which the random vector (ξ,η) is ν distributed. This means that if we want to construct a coupled pair (ξ,η) with prescribed joint distribution on a sufficiently rich probability space, then we may demand that the random variable ξ (with the right distribution) be fixed at the start. An analogous statement also holds if we consider two sequences of random variables $\xi_1,\xi_2,\dots$ and $\eta_1,\eta_2,\dots$ instead of the random variables ξ and η.

We shall prove the above statement in the next problem 11. We remark that a probability space is "sufficiently rich" in the sense we need it if there exists an in the interval $[0,1]$ uniformly distributed random variable in this space which is independent of the random variable ξ or the random sequence $\xi_1,\xi_2,\dots$ which we want to couple with an appropriate random variable or random sequence.

11.) let ν be a probability measure on a product space $X\times X$, where (X,ρ) is a complete separable metric space. Let μ be the projection of the measure ν to the first coordinate of the space $X\times X$. Let ξ be a μ distributed random variable on a probability space $(\Omega,\mathcal{A},P)$, and let us assume that there exists a random variable χ on this probability space $(\Omega,\mathcal{A},P)$ with uniform distribution on the interval $[0,1]$ which is independent of ξ. Then such a random variable η can be constructed for which the distribution of the random pair (ξ,η) is ν.

12.) If ξ and η are two random variables taking their values on a measurable space $(X,\mathcal{A})$, the distribution of ξ is μ and the distribution of η is ν, then $P(\xi\neq\eta)\ge \operatorname{Var}(\mu,\nu)$, where $\operatorname{Var}(\mu,\nu)=\sup\limits_{A\in\mathcal{A}}|\mu(A)-\nu(A)|$, is the variational distance between the measures μ and ν. For arbitrary probability measures μ and ν there exist such μ and ν distributed random variables ξ and η, for which the two sides of the above inequality are equal.

Finally, we give a concise proof of two statements with the help of the above results. The first of them is the functional central limit theorem which is also called the invariance principle in the literature. The second statement enables us to deduce some useful consequences of the functional central limit theorem. Before its formulation let us recall the following form of the central limit theorem.

Central limit theorem. *Let $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$, with some positive integers n_k, be a triangular array, i.e. assume that the random variables $\xi_{k,j}$, $1\le j\le n_k$, are independent for a fixed number k. Assume that $E\xi_{k,j}=0$, $E\xi_{k,j}^2=\sigma_{k,j}^2<\infty$, $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}\sigma_{k,j}^2=1$. If beside this the triangular array $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$, satisfies the Lindeberg condition, i.e.*

$$\lim_{k\to\infty} E\xi_{k,j}^2 I(|\xi_{k,j}|>\varepsilon)=0 \quad \text{for all numbers } \varepsilon>0,$$

where $I(A)$ denotes the indicator function of a set A, then the sums $S_k=\sum\limits_{j=1}^{n_k}\xi_{k,j}$, $k=1,2,\dots$, converge in distribution to the standard normal distribution as $k\to\infty$.

The functional central limit theorem, formulated below, says that under the conditions of the central limit theorem also the following sharper statement holds.

Functional central limit theorem. *Let $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$, be a triangular array such that $E\xi_{k,j}=0$, $E\xi_{k,j}^2=\sigma_{k,j}^2<\infty$, $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}\sigma_{k,j}^2=1$. Let us also assume that the triangular array $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$, satisfies the Lindeberg condition. Introduce the partial sums $S_{k,l}=\sum\limits_{j=1}^{l}\xi_{k,j}$, $k=1,2,\dots$, $1\le l\le n_k$, and define the random broken line $S_k(t)$, $0\le t\le 1$, $k=1,2,\dots$, in the following way. Put $u_{k,l}=\sum\limits_{j=1}^{l}\sigma_{k,j}^2$, $\bar u_{k,l}=\dfrac{u_{k,l}}{u_{k,n_k}}$, let $S_k(\bar u_{k,l})=S_{k,l}$, $S_k(0)=0$, $k=1,2,\dots$, $1\le l\le n_k$, and let the function $S_k(t)$ be linear in the intervals $[0,\bar u_{k,1}]$ and $[\bar u_{k,l-1},\bar u_{k,l}]$, $k=1,2,\dots$, $2\le l\le n_k$. Then the random broken lines $S_k(t)$, $0\le t\le 1$, can be considered as random variables taking values in the Banach space $C([0,1])$ of continuous functions endowed with the supremum norm. The functional central limit theorem states that the distributions of the random variables $S_k(t)$ (in the space $C([0,1])$) converge in distribution to the so-called Wiener measure, i.e. to the distribution of a Wiener process $W(t)$, $0\le t\le 1$, as $k\to\infty$.*

13.) Prove (with the help of the results in this work, the central limit theorem, inequalities for the maximum of sums of independent random variables and some basic facts about Wiener processes) the functional central limit theorem.

The coupling results of this work can be useful also in the proof of the following statement which may help for instance to understand why the functional central limit theorem is a useful result of probability theory.

14.) Let μ_n, $n=1,2,\dots$, be a sequence of probability measures on a separable (not necessarily complete) metric space (X,ρ) which converges weakly to a probability measure μ_0 on this space. Let $\mathcal{F}$ be a measurable map from this separable metric space (X, ρ) to some other separable metric space (Y, ρ_1) such that the map $\mathcal{F}$ is continuous with probability one with respect to the limit measure μ_0. Let us consider the probability measures $\mathcal{F}\mu_n$, defined as $\mathcal{F}\mu_n(B) = \mu\{x\colon \mathcal{F}(x) \subset B\}$ if $B \subset Y$ is a Borel-measurable set, $n = 0, 1, \dots,$ on the space (Y, ρ_1). Prove with the help of the result of Problem 5 that the probability measures $\mathcal{F}\mu_n$ converge weakly the probability measure $\mathcal{F}\mu_0$ as $\to \infty$.

Remarks

The result of problem 1 was originally proved by V. Strassen. I learned about it and its close relation to the "transport problem" and the König–Hall theorem, an important method of combinatorics, from the works of R. M. Dudley. Let us remark that this result gives a non-trivial answer to the question how closely two random variables or two stochastic processes with prescribed distributions can be put to each other. Nevertheless this result does not give a real help in most coupling problems. The reason for this deficiency is that generally it is not simpler to check formula (a) for *all* closed sets than to construct a coupling in an explicit way.

The result of problem 2 was proven by Yu. V. Prochorov, and the metric introduced there is called the Prochorov metric in the literature. The result of problem 4 (and the result of problem 3 which serves as basis for its solution) belongs to A. V. Skorochod. The construction supplying the solution of problem 5 belongs to R. M. Dudley. The generalization of problem 4 given in problem 5 is non-trivial. One could try to deduce the result of problem 5 to problem 4 by embedding a separable metric space to a complete separable metric space. Such an embedding is always possible, but it may happen that the embedded space is a non-measurable subset of the larger space. Hence one cannot get a solution for problem 5 in such a simple way.

The quantile transform is a well-known method in probability theory, and it is frequently used in certain investigations. It is hard to relate its introduction to a definite person. It is worth mentioning that the proof of the optimality property of the quantile transform formulated in problem 9 is based on its reformulation for a "transport problem". It is the simple structure of the real line which enables us to prove such an explicit result in this case.

The results of problems 10, 11 and 12 are more or less well-known for mathematicians working in this subject. However, I had the impression that the answer to the question how big freedom we have to construct random variables or random sequences with prescribed distribution is not sufficiently well understood. Hence I thought it may be useful to discuss some results which may help to study such questions.

The functional central limit theorem is a classical result of probability theory. Its usual proof is based on the investigation of the convergence of probability measures. The functional central limit theorem together with the result formulated in problem 14 explain why all "natural" functions of partial sums of independent random variables have a limit distribution which is independent of the distribution of the summands and can be expressed as the distribution of a functional of a Wiener process.

Solutions

0.) Let us define for all sets $B \subset Z$ the set

$$A(B) = \{y\colon y \in Y,\ d(y,z) = 1 \text{ for some } z \in B\}.$$

We have to show that if the conditions of the König–Hall theorems are satisfied, then $|A(B)| \ge |B|$. This statement is equivalent to the relation $|Y \setminus A(B)| \le |Z \setminus B|$. It follows from the conditions of the problem that $|B(Y \setminus A(B))| \ge |Y \setminus A(B)|$. Hence, it is enough to show that $B(Y \setminus A(B)) \subset Z \setminus B$. This relation holds, since if $y \in B(Y \setminus A(B))$, i.e. there exists an $z \notin A(B)$ such that $d(y,z) = 1$, then by the definition of the set $A(B)$ $y \notin B$, i.e. $B(Y \setminus A(B)) \subset Z \setminus B$.

The analogous statement in the case of the continuous version of the König–Hall theorem can be proved similarly. In this case the statement we have to prove is equivalent to the inequality $\sum\limits_{y \in Y \setminus A(B)} u(y) \le \sum\limits_{z \in Z \setminus B} v(z)$ because of the identity $\sum\limits_{y \in Y} u(y) = \sum\limits_{z \in Z} v(z)$. This inequality follows from the relation $B(Y \setminus A(B)) \subset Z \setminus B$.

1.) Let us define the probability space $(\Omega, \mathcal{A}, P)$ with the following choice: $\Omega = X \times X$, $\mathcal{A}$ is the σ algebra generated by the topology of the space $X \times X$, and P is an appropriate probability measure on the measure space $(\Omega, \mathcal{A})$ we still have to define. Put $\xi(x_1, x_2) = x_1$ and $\eta(x_1, x_2) = x_2$. We solve the problem if we can construct a probability measure P on the space $(\Omega, \mathcal{A})$ which satisfies the following relations:

a.) $P(A \times X) = \mu(A)$, $P(X \times A) = \nu(A)$ for all measurable sets $A \subset X$.

b.) $P\left(\{(x_1, x_2)\colon \rho(x_1, x_2) > \alpha + \varepsilon\}\right) \le \beta + \varepsilon$.

We shall construct such a probability measure P with the help of the continuous version of the König–Hall theorem.

First we define a bipartitated graph with an appropriate weight function. To do this we introduce some notations. Let $G(x, \alpha)$ denote the ball with center x and radius α in the metric space (X, ρ). Let $x_1, x_2, \dots,$ be an everywhere dense sequence on the space X, and let us fix a number $\varepsilon > 0$. As $\bigcup\limits_{n=1}^{\infty} G\left(x, \frac{\varepsilon}{5}\right) = X$, there exists such a number $N = N(\varepsilon)$ for which the set $W_N = \bigcup\limits_{n=1}^{N(\varepsilon)} G\left(x, \frac{\varepsilon}{5}\right)$ satisfies the relations $\mu(W_N) > 1 - \frac{\varepsilon}{2}$ and $\nu(W_N) > 1 - \frac{\varepsilon}{2}$. Let us define the sets $V_k = G\left(x_k, \frac{\varepsilon}{5}\right) \setminus \bigcup\limits_{j=1}^{k-1} G\left(x_j, \frac{\varepsilon}{5}\right)$, $k = 1, \dots, N$ and $V_{N+1} = X \setminus W_N$. Then V_k, $k = 1, \dots, N+1$ is a partition of the space X, $d(V_k) \le \frac{\varepsilon}{5}$, if $1 \le k \le N$, where $d(A)$ denotes the diameter of set $A \subset X$. Further, $\mu(V_{N+1}) < \frac{\varepsilon}{2}$ and $\nu(V_{N+1}) < \frac{\varepsilon}{2}$. We shall call the point x_k the center of the set V_k, $1 \le k \le N$.

We define the following bipartitated graph $(Y, Z, d(\cdot,\cdot))$: $Y = \{y_1, \dots, y_{N+1}\} = \{V_1, \dots, V_{N+1}\}$, $Z = \{z_1, \dots, z_{N+1}\} = \{V_1, \dots, V_{N+1}\}$, $d(y_j, z_k) = 1$, if $\rho(x_j, x_k) \le$

$\alpha + \frac{\varepsilon}{2}$, $1 \le j, k \le N$, and also $d(y_{N+1}, z_k) = 1$ and $d(y_j, z_{N+1}) = 1$ for all indices $1 \le j, k \le N+1$. In all other cases we define $d(y, z) = 0$. (The relation $d(y, z) = 1$ means that the points y and z are connected.) We also introduce the following weight functions $u(\cdot)$ and $v(\cdot)$ on the sets Y and Z: $u(y_j) = \mu(V_j)$, $v(z_j) = \nu(V_j)$, $j = 1, \dots, N$, $u(y_{N+1}) = \mu(V_{n+1}) + \beta$, $v(z_{N+1}) = \nu(V_{n+1}) + \beta$. We claim that this bipartitated graph and weight function satisfy the conditions of the continuous version of the König–Hall theorem.

The desired inequality obviously holds for such sets $A \subset Y$, for which $y_{N+1} \in A$. Indeed, in this case $B(A) = Z$, since y_{N+1} is connected to all points of the set Z. If $y_{N+1} \notin A$ then put $D_1 = \bigcup\limits_{V_j \in A} V_j$ and $D_2 = \bigcup\limits_{V_j \in B(A)} V_j$. In this case $\sum\limits_{y \in A} u(y) = \mu(D_1)$, and $\sum\limits_{z \in B(A)} v(z) = \nu(D_2) + \beta$, since $y_{N+1} \notin A$ and $z_{N+1} \in B(A)$. Hence, it is enough to show that $\bar{D}_1^\alpha \subset D_2$, where $\bar{D}_1$ denotes the closure of the set D_1. But if $x \in \bar{D}_1$, then there exists such a set $y_j = V_j \in A$ whose center x_j satisfies the inequality $\rho(x, x_j) \le \frac{\varepsilon}{5}$, thus $G(x, \alpha) \subset G\left(x_j, \alpha + \frac{\varepsilon}{5}\right)$. We claim that $G(x, \alpha) \subset G\left(x_j, \alpha + \frac{\varepsilon}{5}\right) \subset D_2$, with this point x, hence $\bar{D}_1^\alpha \subset D_2$ as we claimed. Indeed, if this statement did not hold, then there would exist a point $v \in G\left(x_j, \alpha + \frac{\varepsilon}{5}\right)$, such that the element V_k of the partition $\{V_1, \dots, V_{N+1}\}$ for which $v \in V_k$ and its center x_k have the following properties: The points V_k and V_j are not connected in the bipartitated graph we have defined, hence $1 \le k \le N$, and $\rho(x_j, x_k) \ge \alpha + \frac{\varepsilon}{2}$. But this is not possible, since $d(V_k) \le \frac{\varepsilon}{5}$ and thus $\rho(x_j, x_k) \le \rho(x_j, v) + \frac{\varepsilon}{5} \le \alpha + \frac{2\varepsilon}{5}$. We have shown the continuous version of the König–Hall theorem can be applied to this system.

Let $\bar{w}(y, z)$, $y \in Y$, $z \in Z$ be a "transport function" satisfying the continuous version of the König–Hall theorem in the above system, and let us define the following function $w_1(y_j, z_k)$ with its help:

$$w_1(y_j, z_k) = \bar{w}(y_j, z_k) \quad \text{if } 1 \le j, k \le N, \\ \text{and } w_1(y_j, z_k) = 0, \quad \text{if } j = N+1 \text{ or } k = N+1.$$

This function $w_1(y, z)$ satisfies the following properties:

i.) $w_1(y_j, z_k) \ge 0$, and $w_1(y_j, z_k) > 0$ only if $1 \le j, k \le N$ and $\rho(x_j, x_k) < \alpha + \frac{\varepsilon}{2}$

ii.) $\sum\limits_{z \in Z} w_1(y_j, z) \le u(y_j) = \mu(V_j)$, $\sum\limits_{y \in Y} w_1(y, z_k) \le v(z_k) = \nu(V_k)$.

iii.) $\sum\limits_{y \in Y, z \in Z} w_1(y, z) \ge 1 - \beta - \varepsilon$, because $\sum\limits_{y \in Y, z \in Z} w_1(y, z) \ge \sum\limits_{y \in Y, z \in Z} w(y, z) - u(y_{N+1}) - v(z_{N+1}) \ge 1 + \beta - 2(\beta + \frac{\varepsilon}{2})$.

By the properties of the function $w_1(y, z)$ there is a function $w_2(y_j, z_k) \ge 0$, $1 \le j, k \le N+1$, such that the function $w(y, z) = w_1(y, z) + w_2(y, z)$ satisfies the properties

$$\sum_{z \in Z} w(y_j, z) = u(y_j) = \mu(V_j), \quad \sum_{y \in Y} w(y, z_k) = v(z_k) = \nu(V_k)$$

We shall define a probability measure P which satisfies properties a.) and b.) with the help of the function $w(y,z)$. Put

$$P(C\times D)=\frac{\mu(C)}{\mu(W_j)}\frac{\nu(C)}{\nu(W_k)}w(y_j,z_k)\quad \text{if } C\subset W_j,\ D\subset W_k$$

with some indices $1\le j,k\le N+1$. In the general case let us define

$$P(C\times D)=\sum_{j=1}^{N+1}\sum_{k=1}^{N+1}P\left((C\cap V_j)\times(D\cap V_k)\right).$$

After this definition the measure P can be extended from the above rectangular sets to the whole σ algebra $\mathcal{A}$ in a unique way. We claim that this measure P satisfies both properties a.) and b.). Indeed, for all sets $A\subset V_j$, $1\le j\le N+1$

$$P(A\times X)=\sum_{k=1}^{N+1}P(A\times V_k)=\frac{\mu(A)}{\mu(V_j)}\sum_{k=1}^{N+1}w(y_j,z_k)=\mu(A),$$

and this implies that $P(X\times A)=\nu(A)$ for all measurable sets $A\subset X$. The second part of statement a.) can be proved similarly. On the other hand,

$$\begin{aligned}P((x_1,x_2)\colon \rho(x_1,x_2)\le\alpha+\varepsilon)&\ge\sum_{(j,k)\colon\ \rho(x_j,x_k)\le\alpha+\frac{\varepsilon}{2}}P(V_j\times V_k)\\&=\sum_{(j,k)\colon\ \rho(x_j,x_k)\le\alpha+\frac{\varepsilon}{2}}w(V_j\times V_k)\\&\ge\sum_{y_j\in Y,\ z_k\in Z}w_1(V_j\times V_k)\ge 1-\beta-\varepsilon\end{aligned}$$

by the property iii.), and this statement is equivalent to condition b.)

Conversely, if property (a) holds, then for arbitrary closed set B $\{\xi\subset B,\eta\notin B^{\alpha+\varepsilon}\}\subset\{(\xi,\eta)\colon \rho(\xi,\eta)>\alpha+\varepsilon\}$, hence $P(\xi\in B,\eta\notin B^{\alpha+\varepsilon})\le\beta+\varepsilon$. As $\{\xi\in B\}\subset\{\xi\in B,\eta\notin B^{\alpha+\varepsilon}\}\cup\{\eta\in B^{\alpha+\varepsilon}\}$, this implies that $\mu(B)\le\beta+\varepsilon+\nu(B^{\alpha+\varepsilon})$, and this is what we have to prove.

To prove the last statement of problem 1 let us make the following observation. If X is a complete metric space, and P_n, $n=1,2,\dots$, is a sequence of probability measures on the space $X\times X$ whose marginal distributions are two measures μ and ν which do not depend on the index n, then this sequence of probability measures has a convergent subsequence if the weak convergence of probability measures are considered.

To prove this statement let us observe that for all numbers $\varepsilon>0$ there exists such a compact set $K\subset X$ for which $\mu(K)\ge 1-\frac{\varepsilon}{2}$, and $\nu(K)\ge 1-\frac{\varepsilon}{2}$. Hence the compact set $K\times K\subset X\times X$ and a sequence of probability measures P_n on the space $X\times X$

whose marginal distributions are μ and ν satisfy the inequality $P_n(K\times K)\ge 1-\varepsilon$ for all indices n. By some classical results in probability theory this implies that the sequence of the probability measures P_n is tight, hence it has a weakly convergent subsequence.

Let P_n, $n=1,2,\dots$, be a sequence of probability measures on the space $X\times X$ with marginal distributions μ and ν which satisfy property (a) with numbers $\varepsilon_n=\frac{1}{n}$. Let P_{n_k}, $k=1,2,\dots$, be a convergent subsequence of this sequence, and let P denote its limit. The marginal distributions of the measure P are μ and ν. We claim that a random vector with distribution P satisfies property (a) also with the number $\varepsilon=0$. Indeed, the sets of the form $\{(x_1,x_2)\colon (x_1,x_2)\in X\times X, \rho(x_1,x_2)>\alpha+\varepsilon\}$ are open. Hence

$$\beta\ge\limsup_{k\to\infty}P_{n_k}\left(\{(x_1,x_2)\colon \rho(x_1,x_2)>\alpha+\varepsilon\}\right)\ge P\left(\{(x_1,x_2)\colon \rho(x_1,x_2)>\alpha+\varepsilon\}\right)$$

for all numbers $\varepsilon>0$. This implies that

$$P\left(\{(x_1,x_2)\colon \rho(x_1,x_2)>\alpha\}\right)=\lim_{\varepsilon\to0}P\left(\{(x_1,x_2)\colon \rho(x_1,x_2)>\alpha+\varepsilon\}\right)\le\beta.$$

Problem 1 is proved.

2.) Let us first show that the function $d(\cdot,\cdot)$ is a metric. i.) $d(\mu,\mu)=0$. On the other hand, we show that in the case $d(\mu,\nu)=0$ $\mu=\nu$. Indeed, if $d(\mu,\nu)=0$, then $\mu(F)\le\nu(F)$ for all closed sets $F\subset X$, since $F=\bigcap\limits_{\varepsilon\to0}F^\varepsilon$, and $\mu(F)\le\liminf\limits_{\varepsilon\to0}(\nu(F^\varepsilon)+\varepsilon)=\nu(F)$. We show that also the identity $\mu(F)=\nu(F)$ holds for closed sets F. Indeed, as $\mu(G)\ge\nu(G)$ for all open sets G, hence $\mu(F)=\lim\limits_{\varepsilon\to0}\mu(F^\varepsilon)\ge\lim\limits_{\varepsilon\to0}\nu(F^\varepsilon)=\nu(F)$. This implies that $\mu(A)=\nu(A)$ for all closed or open sets A. On the other hand, a probability measure is uniquely determined by its values on the open sets, hence $\mu=\nu$. ii.) $d(\mu,\nu)=d(\nu,\mu)$. Let $d(\mu,\nu)<\alpha$ with some number $\alpha>0$. We show that in this case $d(\nu,\mu)\le\alpha$. Let $F\subset X$ be an arbitrary closed set, and let us define the set $F'=X\setminus F^\alpha$. We claim that $(F')^\alpha\subset X\setminus F$. Indeed, if $y\in(F')^\alpha$ then $d(y,X\setminus F^\alpha)<\alpha$, that is there exists such a point $z\in X$, for which $d(z,F)\ge\alpha$, and $d(z,y)<\alpha$. This implies that $y\notin F$, hence $(F')^\alpha\subset X\setminus F$. With the help of this relation we get that $1-\mu(F^\alpha)=\mu(F')\le\nu((F')^\alpha)+\alpha\le\nu(X\setminus F)+\alpha=1-\nu(F)+\alpha$, i.e. $\nu(F)\le\mu(F^\alpha)+\alpha$. Hence $d(\nu,\mu)\le\alpha$. In such a way we have shown that $d(\nu,\mu)\le d(\mu,\nu)$. Because of symmetry reasons $d(\mu,\nu)=d(\nu,\mu)$. iii.) $d(\mu_1,\mu_3)\le d(\mu_1,\mu_2)+d(\mu_2,\mu_3)$. If $d(\mu_1,\mu_2)=\alpha$, $d(\mu_2,\mu_3)=\beta$, then $\mu_1(F)\le\mu_2(F^{\alpha+\varepsilon})+\alpha+\varepsilon$, $\mu_2(F^{\alpha+\varepsilon})\le\mu_3((F^{\alpha+\varepsilon})^{\beta+\varepsilon})+\beta+\varepsilon$ for all numbers $\varepsilon>0$ and closed sets F. As $(F^{\alpha+\varepsilon})^{\beta+\varepsilon})\subset F^{\alpha+\beta+2\varepsilon}$ this implies that $\mu_1(F)\le\mu_3(F^{\alpha+\beta+2\varepsilon})+\alpha+\beta+2\varepsilon$. Since this relation holds for all closed sets F and numbers $\varepsilon>0$ it implies relation iii.).

We show that $d(\mu_n,\mu)\to0$ if and only if $\mu_n\Rightarrow\mu$. If $d(\mu_n,\mu)\to0$, then $\limsup\limits_{n\to\infty}\mu_n(F)\le\mu(F^\varepsilon)+\varepsilon$ for all closed sets F and numbers $\varepsilon>0$. As $F=\bigcap\limits_{\varepsilon\to0}F^\varepsilon$, $\lim\limits_{\varepsilon\to0}(\mu(F^\varepsilon)+\varepsilon)=\mu(F)$. This relation implies the inequality $\limsup\limits_{n\to\infty}\mu_n(F)\le\mu(F)$ for all closed sets F, and this is equivalent to the statement $\mu_n\Rightarrow\mu$.

To prove the other direction of the statement first we show that for all $\varepsilon > 0$ there exists an integer $N = N(\varepsilon) > 0$ and a partition $V_1, \dots, V_N, V_{N+1}$ of the space X which satisfies the following conditions: $\mu(\partial V_k) = 0$, $k = 1, \dots, N+1$, $\bar{d}(V_k) < \varepsilon^2$, $k = 1, \dots, N$, and $\mu(V_{N+1}) \le \frac{\varepsilon}{2}$, where ∂A denotes the boundary of the set A and $\bar{d}(A)$ its diameter. Indeed, let $x_1, x_2, \dots$ be an everywhere dense set in the space X. For all x_k let us choose a ball $G(x_k, \delta_k)$ in the space X of center x_k and radius δ_k with some δ_k such that $\frac{\varepsilon^2}{2} < \delta_k < \varepsilon^2$ for which $\mu\left(\partial G(x_k, \delta_k)\right) = 0$. The union of these balls covers the whole space X. Let us choose such a number $N = N(\varepsilon)$ for which $\mu\left(\bigcup\limits_{k=1}^{N} G(x_k, \delta_k)\right) \le \frac{\varepsilon}{2}$. Let $V_{N+1} = X \setminus \left(\bigcup\limits_{k=1}^{N} G(x_k, \delta_k)\right)$ and $V_k = G(x_k, \delta_k) \setminus \left(\bigcup\limits_{j=1}^{k=1} G(x_j, \delta_j)\right)$, $k = 1, \dots, N$. These sets satisfy the conditions we have imposed.

For all closed sets $F \subset X$ let us define the set $B(F) = \bigcup\limits_{k\colon V_k \cap F \ne \emptyset} V_k$. Let us observe that $F \subset B(F) \subset F^\varepsilon \cup V_{N+1}$. Further, $\lim\limits_{n\to\infty} \sup\limits_{F \text{ closed set}} |\mu_n(B(F) - \mu(B(F)| = 0$, because $\lim\limits_{n\to\infty} \mu_n(V_k) = \mu(V_k)$ for all $k = 1, \dots, N+1$, there are finitely many sets of the form $B(F)$, and all of them are the union of finitely many sets V_k. Hence there exists a threshold index $n = n(\varepsilon)$ independent of the closed set F such that

$$\mu(F^\varepsilon) - \mu_n(F) \ge \mu(B(F)) - \mu(V_{N+1}) - \mu_n(B(F)) \ge -\varepsilon$$

for all closed sets F. This implies that $d(\mu_n, \mu) \le \varepsilon$, if $n \ge n_0(\varepsilon)$. Hence $\lim\limits_{n\to\infty} d(\mu_n, \mu) = 0$ as we claimed.

We show that the metric space $(\mathcal{M}, d)$ is separable. Let $x_1, x_2, \dots,$ be an everywhere dense set in the space X, let $\mathcal{M}_0$ be the set of those discrete probability measures which are concentrated in the finite subsets of the set $\{x_1, x_2, \dots\}$ and the measure of all points is a rational number. This is a countable set, hence it is enough to show that $\mathcal{M}_0$ is an everywhere dense subset of the set $\mathcal{M}$. This can be seen for instance by a slight modification of the previous argument. This shows that for an arbitrary measure $\mu \in \mathcal{M}$ and number $\varepsilon > 0$ we can choose a partition $V_1, \dots, V_N$ of the set X and a number $\eta > 0$ for which the following statement holds: If a probability measure $\nu \in \mathcal{M}$ satisfies the condition $|\mu(V_k) - \nu(V_k)| \le \eta$ for all numbers $k = 1, \dots, N$, then the measure ν also satisfies the relation $d(\mu, \nu) < \varepsilon$. Since for all $\mu \in \mathcal{M}$ $\mathcal{M}_0$ contains such a measure ν, hence $\mathcal{M}_0$ is an everywhere dense subset of the set $\mathcal{M}$.

We show that if X is a complete separable space, then the same property holds for the space $(\mathcal{M}, d)$. It is enough to show that if

$$\lim_{n\to\infty} \sup_{m\colon m \ge n} d(\mu_n, \mu_m) = 0,$$

then the series μ_n, $n = 1, 2, \dots$ is relatively compact, i.e. it has a weakly convergent subsequence. To prove this statement it is enough to show that for all numbers $\varepsilon > 0$ there exists a compact set $K \subset X$, for which $\mu_n(K) \ge 1 - \varepsilon$ for all indices n.

The statement from which we want to deduce the completeness of the space $(\mathcal{M}, d)$ can be slightly weakened. It is enough to show that for arbitrary $\varepsilon > 0$ there exists a compact set $K = K(\varepsilon)$ whose neighbourhood of radius ε $K^\varepsilon = \{x\colon \rho(x, K) < \varepsilon\}$ satisfies the inequality $\mu_n(K^\varepsilon) \ge 1 - \varepsilon$ for all indices $n = 1, 2, \dots$. Indeed, let us consider such sets $K(\varepsilon 2^{-m})$ with $\varepsilon 2^{-m}$ for all $m = 1, 2, \dots$, and let us define the set $K = \bigcap\limits_{m=1}^{\infty} (K(\varepsilon 2^{-m})^{\varepsilon 2^{-m}}$. Then $\mu_n(K) \ge 1 - \sum\limits_{m=1}^{\infty} \varepsilon 2^{-m} = 1 - \varepsilon$. Hence it is enough to prove that this set K is relatively compact, (that is, its closure is compact). To show this let us recall the result by which a subset A of a separable complete metric space is relatively compact if and only if this set A has a finite δ-net for all $\delta > 0$, i.e. there exists a finite subset of the space X such that for arbitrary $x \in A$ the distance of x from one of the points of this finite set is less than δ. This property holds for the above set K, since for all numbers $\delta > 0$ there exists an integer m such that $\delta > \varepsilon 2^{-m}$. Further, the set $K(\varepsilon 2^{-m})^{\varepsilon 2^{-m}}$ has a finite δ-net for this m, and this is also a finite δ-net for the original set K.

This weakened statement can be proved in the following way: For a fixed number $\varepsilon > 0$ let us choose an index $n_0 = n_0(\varepsilon)$ such that the relation $d(\mu_{n_0}, \mu_n) \le \frac{\varepsilon}{4}$ holds for all $n \ge n_0$, and let $K_0 \subset X$ be a compact set such that $\mu_{n_0}(K_0) \ge 1 - \frac{\varepsilon}{4}$. Then $\mu_n(K_0^{\varepsilon/4}) \ge \mu_{n_0}(K_0) - \frac{\varepsilon}{4} \ge 1 - \frac{\varepsilon}{2}$ for all numbers $n \ge n_0$. Let us choose such a compact set K_1, for which $\mu_n(K_1) \ge 1 - \frac{\varepsilon}{2}$ for all numbers $n \le n_0$. Then the set $K = K_1 \cup K_0$ is compact, and $\mu_n(K^\varepsilon) \ge 1 - \varepsilon$ for all numbers $n = 1, 2, \dots$. Indeed, $\mu_n(K^\varepsilon) \ge \mu_n(K_0^{\varepsilon/4}) \ge 1 - \varepsilon$, if $n \ge n_0$, and $\mu_n(K^\varepsilon) \ge \mu_n(K_1) \ge 1 - \varepsilon$, if $n \le n_0$. The statements of problem 2 are proved.

3.) Let us observe that the space X has a partition $\mathcal{X}_1 = \{A_1, A_2, \dots\}$ such that $\bar{d}(A_j) \le 1$, and $\mu(\partial A_j) = 0$ for all indices $j = 1, 2, \dots$. Here $\bar{d}(A)$ denotes the diameter of the set A and ∂A its boundary. (This statement follows from the argument presented at the start of problem 2, when it was shown that the relation $\mu_n \Rightarrow \mu$ implies that $d(\mu_1, \mu_2) \to 0$.) By splitting further the elements of this partition we get a more and more refined sequence of partitions $\mathcal{X}_1 \supset \mathcal{X}_2 \supset \cdots \mathcal{X}_k \supset \cdots$ of the space X such that the elements of these partitions satisfy the relations $\bar{d}(A_{j_1,\dots,j_k}) \le 2^{-k}$ and $\mu(\partial A_{j_1,\dots,j_k}) = 0$. ($A_{j_1,\dots,j_k} \in \mathcal{X}_k$.) Let us define a similar sequence of more and more refined partitions $\mathcal{Y}_1 \supset \mathcal{Y}_2 \supset \cdots \mathcal{Y}_k \supset \cdots$ of the intervals $(0, 1]$ (endowed with the Lebesgue measure) in the following way:

Put $\mathcal{Y}_1 = \{B_1, \dots, B_k, \dots\}$, $B_k = (b_{k-1}, b_k]$, $b_k = \sum\limits_{j=1}^{k} \mu(A_j)$, $k = 1, 2, \dots$, and if the sets $B_{j_1,\dots,j_k} = (b_{j_1,\dots,j_{k-1}}, b_{j_1,\dots,j_k}]$, of the partition $\mathcal{Y}_k$ are already defined, and they are defined so that the indentity $b_{j_1,\dots,j_k} - b_{j_1,\dots,j_{k-1}} = \mu(A_{j_1,\dots,j_k})$ holds, then split all intervals $B_{j_1,\dots,j_k}$ into subsequent non-overlapping intervals

$$B_{j_1,\dots,j_k,s} = (b_{j_1,\dots,j_k,s-1}, b_{j_1,\dots,j_k,s}],$$

of length $\mu(A_{j_1,\dots,j_k,s})$. These smaller intervals will be the elements of the partition $\mathcal{Y}_{k+1}$ of the interval $[0, 1]$. After the definition of these partitions let us fix for all integers $k \ge 1$ and elements $A_{j_1,\dots,j_k}$ of the partition $\mathcal{X}_k$ a point $x_{j_1,\dots,j_k} \in A_{j_1,\dots,j_k}$.

Now for all $k=1,2,\dots$ we define a random variable $\xi_k(y)$ on the probability space $((0,1],\mathcal{B},\lambda)$ with values on the space X by means of the following formula: Put $\xi_k(y)=x_{j_1,\dots,j_k}$ if $y\in B_{j_1,\dots,j_k}$. We claim that the limit $\xi(y)=\lim\limits_{k\to\infty}\xi_k(y)$ exists for all $y\in(0,1]$, and it is a μ distributed random variable.

The above limit really exists, because for all $y\in(0,1]$ a uniquely determined sequence of decreasing intervals $B_{j_1,\dots,j_k}$, $k=1,2,\dots$, exists in such a way that $y\in B_{j_1,\dots,j_k}$. This implies that the sequence of points $\xi_k(y)\in A_{j_1,\dots,j_k}$ is a Cauchy, hence also a convergent sequence.

The union of the σ-algebras $\mathcal{X}_k$, $k=1,2,\dots$, generates the Borel σ-algebra in the space (X,ρ). Hence to prove that the above defined random variable ξ has distribution μ it is enough to show that $P\left(\xi\in A_{j_1,\dots,j_k}\right)=\mu\left(A_{j_1,\dots,j_k}\right)$ for all numbers $k=1,2,\dots$ and sets $A_{j_1,\dots,j_k}$. It follows from the construction of the random variable ξ that if $y\in B_{j_1,\dots,j_k}$ then $\xi(y)\in\bar A_{j_1,\dots,j_k}$, where $\bar A$ denotes the closure of the set A. First we show that it follows from this fact and the relation

$$P\left(\xi(y)\in\partial A_{j_1,\dots,j_k}\right)=0\quad\text{for all numbers } k=1,2,\dots \text{ and indices } j_1,\dots,j_k, \quad (*)$$

to be proven later that ξ is μ distributed. Indeed, these relations imply that

$$\begin{aligned}P\left(\xi(y)\in A_{j_1,\dots,j_k}\right)&\ge P\left(\xi(y)\in\bar A_{j_1,\dots,j_k}\right)-P\left(\xi(y)\in\partial A_{j_1,\dots,j_k}\right)\\&\ge\lambda\left(B_{j_1,\dots,j_k}\right)=\mu\left(A_{j_1,\dots,j_k}\right),\end{aligned}$$

and summing up these inequalities we get that

$$1=\sum_{j_1,\dots,j_k}P\left(\xi(y)\in A_{j_1,\dots,j_k}\right)\ge\sum_{j_1,\dots,j_k}\mu\left(A_{j_1,\dots,j_k}\right)=1.$$

As a consequence, the above inequalities are actually identities, and the distribution of ξ is μ.

To prove formula $(*)$ let us observe that since $\mu(\partial A_{j_1,\dots,j_k})=0$ for all numbers $\varepsilon>0$ there exists a number $\delta=\delta(\varepsilon)>0$ such that the neighbourhood $(\partial A_{j_1,\dots,j_k})^\delta$ of radius δ of the set $\partial A_{j_1,\dots,j_k}$ satisfies the inequality $\mu\left((\partial A_{j_1,\dots,j_k})^\delta\right)<\varepsilon$. Let us choose an integer s so large that the diameters of the elements of the partition $\mathcal{X}_k$ satisfy the inequality $\bar d(A_{j'_1,\dots,j'_s})<\frac{\delta}{2}$ for all sets $A_{j'_1,\dots,j'_s}$. Since $\rho(\xi_s(y),\xi(y))\le\max\limits_{j'_1,\dots,j'_s}\bar d(A_{j'_1,\dots,j'_s})\le\frac{\delta}{2}$, hence in the case $\xi(y)\in\partial A_{j_1,\dots,j_k}$ $\rho\left(\xi_s(y),\partial A_{j_1,\dots,j_k}\right)<\frac{\delta}{2}$. This implies that if $\xi(y)\in\partial A_{j_1,\dots,j_k}$, then the indices $j'_1,\dots,j'_s$ for which $y\in B_{j'_1,\dots,j'_s}$ are such that $\xi_s(y)\in A_{j'_1,\dots,j'_s}$, $\rho\left(A_{j'_1,\dots,j'_s},\partial A_{j_1,\dots,j_k}\right)<\frac{\delta}{2}$, hence $A_{j'_1,\dots,j'_s}\subset(\partial A_{j_1,\dots,j_k})^\delta$. This implies that

$$\begin{aligned}&\{y\colon \xi(y)\in\partial A_{j_1,\dots,j_k}\}\\&\qquad\subset\left\{y\colon \xi_s(y)\in A_{j'_1,\dots,j'_s}\subset(\partial A_{j_1,\dots,j_k})^\delta \text{ with some indices } j'_1,\dots,j'_s\right\},\end{aligned}$$

and

$$\lambda\{y\colon \xi(y)\in\partial A_{j_1,\dots,j_k}\}\le \sum_{(j'_1,\dots,j'_s)\colon\, A_{j'_1,\dots,j'_s}\subset(\partial A_{j_1,\dots,j_k})^\delta}\lambda\left(B_{j'_1,\dots,j'_s}\right)$$

$$=\sum_{(j'_1,\dots,j'_s)\colon\, A_{j'_1,\dots,j'_s}\subset(\partial A_{j_1,\dots,j_k})^\delta}\mu\left(A_{j'_1,\dots,j'_s}\right)\le\mu\left((\partial A_{j_1,\dots,j_k})^\delta\right)\le\varepsilon.$$

Since this statement holds for all $\varepsilon>0$, this implies relation $(*)$ and the statement of problem 3.

4.) We construct the random variables ξ_n with distribution μ_n, $n=1,2,\dots$, and the random variable ξ with distribution μ simultaneously by means of the construction of problem 3. The only novelty is that we apply the same partitions $\mathcal{X}_1\subset\mathcal{X}_2\subset\cdots$ in the construction of all random variables ξ_n, $n=1,2,\dots$, and ξ, and demand that the boundaries of the elements $A_{j_1,\dots,j_k}\in\mathcal{X}_k$ of the partitions satisfy the relation $\mu(\partial A_{j_1,\dots,j_k})=0$ and $\mu_n(\partial A_{j_1,\dots,j_k})=0$ for all $k=1,2,\dots$, $A_{j_1,\dots,j_k}\in\mathcal{X}_k$ and $n=1,2,\dots$. We remark that it causes no extra difficulty to find partitions $\mathcal{X}_k$ with such elements whose boundary have the above property. Indeed, we can for instance introduce a new probability measure $\bar\mu=\dfrac{\mu}{2}+\sum\limits_{n=1}^{\infty}2^{-n-1}\mu_n$ and construct a sequence of more and more refined partitions of the space X in the same way as it was done in the construction of problem 3 with the only difference that we replace the measure μ by the measure $\bar\mu$. Then the boundaries of the partitions satisfy the desired requirement for all measures μ and μ_n, $n=1,2,\dots$. We claim that if $\mu_n\Rightarrow\mu$, then $\xi_n\to\xi$ with probability one for the random variables constructed in the above way.

Let us observe that if $\mu_n\Rightarrow\mu$, then $\lim\limits_{n\to\infty}\mu_n(A_{j_1,\dots,j_k})=\mu(A_{j_1,\dots,j_k})$ for all $j_1,\dots,j_k$, since the boundaries of these sets have zero μ measure. This relation and the structure of our construction imply that $b_{j_1,\dots,j_k}(n)\to b_{j_1,\dots,j_k}$ for all $k=1,2,\dots$ and indices $j_1,\dots,j_k$ as $n\to\infty$, where $b_{j_1,\dots,j_k}(n)$ and $b_{j_1,\dots,j_k}$ denote the right-hand side end-points of the intervals $B_{j_1,\dots,j_k}(n)$ and $B_{j_1,\dots,j_k}$ which appear in the partitions $\mathcal{Y}_k(n)$ and $\mathcal{Y}_k$ of the intervals $(0,1]$ introduced during the definition of the random variables ξ_n and ξ.

For a fixed small number $\varepsilon>0$ let us choose a number $k=k(\varepsilon)$ for which $2^{-k}<\varepsilon$. Then the diameter of the elements $A_{j_1,\dots,j_k}$ of the partition $\mathcal{X}_k$ are less than $2^{-k}<\varepsilon$. Let us choose a finite set D_k of k-tuples $\{j_1,\dots,j_k\}$ for which $\sum\limits_{(j_1,\dots,j_k)\colon\,(j_1,\dots,j_k)\in D_k}\mu(A_{j_1,\dots,j_k})>1-\dfrac{\varepsilon}{2}$. There exists an index $n_0=n_0(\varepsilon)$ such that the set

$$B(\varepsilon)=B(k,\varepsilon,n_0)=\bigcup_{(j_1,\dots,j_k)\in D_k}\bigcup_{n\ge n_0}\left(B_{j_1,\dots,j_k}\triangle B_{j_1,\dots,j_k}(n)\right)$$

satisfies the inequality $\lambda(B(\varepsilon))<\dfrac{\varepsilon}{2}$, where $A\triangle B=(A\setminus B)\cup(B\setminus A)$ denotes the symmetric difference of two sets A and B. The above relations imply that the set

$B_1(\varepsilon) = \bigcup_{(j_1,\dots,j_k)\in D_k} (B_{j_1,\dots,j_k} \setminus B(\varepsilon))$ satisfy the relation $\lambda(B_1(\varepsilon)) > 1-\varepsilon$. Define the set $B_2(\varepsilon) = B_1(\varepsilon) \cap Y_0 \cap \bigcap_{n=1}^{\infty} Y_{0,n}$, where $Y_{0,n} = (0,1) \setminus \bigcup_{k=1}^{\infty} \bigcup_{j_1,\dots,j_k} \{b_{j_1,\dots,j_k}(n)\}$, $n=1,2,\dots$, and $Y_0 = (0,1) \setminus \bigcup_{k=1}^{\infty} \bigcup_{j_1,\dots,j_k} \{b_{j_1,\dots,j_k}(n)\}$. The inequality $\lambda(B_2(\varepsilon)) > 1-\varepsilon$ also holds, since the set $B_1(\varepsilon) \setminus B_2(\varepsilon)$ is countable. Furthermore, if $n > n_0$ and $y \in B_2(\varepsilon)$, then $\rho(\xi_n(y), \xi(y)) \le \varepsilon$, since in this case $\xi_n(y)$ and $\xi(y)$ are in the closure of the same set $A_{j_1,\dots,j_k}$. This implies that

$$\lambda\left(\limsup_{n\to\infty} |\xi_n - \xi| \ge \varepsilon\right) < \varepsilon.$$

Since this relation holds for all $\varepsilon > 0$, it implies the statement of problem 3.

5.) Put $\bar X = X \times [0,1]$, and let us define the measurable space $(\Omega, \mathcal{A})$ as $\Omega = \bar X \times X \times \cdots \times X \times \cdots$, and $\mathcal{A}$ is the product of the σ-algebras on the coordinate spaces $\bar X$ and X. In the point $\omega = (x, u, x_1, x_2, \dots) \in \Omega$ let us define the random variables ξ and ξ_n by the formulas $\xi(\omega) = x$, $\xi_n(\omega) = x_n$, $n = 1, 2, \dots$. The measure P will be defined with the help of the measure $\bar\mu = \mu \times \lambda$ introduced on the space $\bar X$ and appropriately defined conditional distributions $Q_n((x,u), A)$, $n = 1, 2, \dots$, where $(x,u) \in \bar X = X \times [0,1]$, and $A \subset X$ is a measurable subset of the space X. (The function $Q_n((x,u), A)$ is called a conditional distribution function if $Q_n((x,u), \cdot)$ is a probability measure on the space $(X, \mathcal{A})$ for all points $(x,u) \in \bar X$, and $Q_n(\cdot, A)$ is a measurable function for all sets $A \in \mathcal{A}$.) The measure P will be defined in the following way: The distribution of the first coordinate $(x,u) \in \bar X$ is $\bar\mu$, the coordinates $x_n \in X$, $n = 1, 2, \dots$, are conditionally independent for fixed (x,u), and the conditional distribution of the coordinate x_n under this condition is $Q_n((x,u), \cdot)$. In a formal way, put

$$P(A \times A_1 \times \cdots \times A_n) = \int_{(x,u)\in A} Q_1((x,u), A_1) \dots Q_n((x,u), A_n))\, d\mu(x)\, du$$
$$\text{for all measurable sets } A \subset \bar X,\ A_j \subset X,\ j = 1, \dots, n.$$

(By some non-trivial results of the measure theory the above formula and its extension really defines a probability measure P. It is worth mentioning that this fact follows from such a result (the Tulcea–Ionescu theorem), which does not demand some nice topological properties of the space. This is the reason why this method can be applied in non-complete metric spaces which may not have good nice topological properties.)

For all numbers $k = 1, 2, \dots$, let us define such a partition $\mathcal{A}_k = \{A_{1,k}, A_{2,k}, \dots\}$ of X for which $\bar d(A_{j,k}) < \frac{1}{k}$, and $\mu(\partial A_{j,k}) = 0$, $j = 1, 2, \dots$, where $\bar d(A)$ denotes the diameter and ∂A the boundary of the set A. Let us define for all $k = 1, 2, \dots$ an index $m(k)$ such that $\mu\left(\bigcup_{j\ge m(k)} A_{j,k}\right) \le \frac{1}{k^2}$ and $\mu(A_{j,k}) > 0$ for all $1 \le j \le m(k)$.

(To satisfy this latter conditions we can re-index the sets $A_{j,k}$ if this is necessary.) After this let us consider a series of numbers $1 = n_1 < n_2 < n_3 < \cdots$ such that

$$|\mu_n(A_{j,k}) - \mu(A_{j,k})| < \frac{\mu(A_{j,k})}{k^2 m(k)}, \quad \text{for all numbers } 1 \le j < m(k) \quad \text{and } n_k \le n < n_{k+1}.$$

This is possible, since $\lim\limits_{n\to\infty} \mu_n(A_{j,k}) = \mu(A_{j,k})$ for all numbers k and j because of the relation $\mu(\partial A_{j,k}) = 0$.
Let us introduce the numbers

$$\lambda_{j,k,n} = \frac{\min\left(\mu(A_{j,k}), \mu_n(A_{j,k})\right)}{\mu(A_{j,k})}, \quad n_k \le n < n_{k+1}, \quad j < m(k).$$

Clearly, $1 - \frac{1}{k^2} \le \lambda_{j,k,n} \le 1$. Let us define the conditional distributions $Q_n((x,u),\cdot)$ first only to some pairs $(x,u) \in \bar X$ and n which satisfy certain conditions. Put

$$Q_n((x,u),C) = \frac{\mu_n(C \cap A_{j,k})}{\mu_n(A_{j,k})}, \quad \text{if } n_k \le n < n_{k+1},\ x \in A_{j,k},\ 1 \le j < m(k), \quad \text{and } 0 \le u \le \lambda_{j,k,n} \quad \text{for all sets } C \in \mathcal{A}.$$

On the domain where we have not defined the conditional distributions $Q_n(\cdot,\cdot)$ yet we want to do this in such a way which guarantees that the projection of the measure P to the n-th coordinate is μ_n. To do this let us introduce the numbers

$$P_n = \left(1 - \sum_{j=1}^{m(k)-1} \min\left[\mu_n(A_{j,k}), \mu(A_{j,k})\right]\right), \quad n_k \le n < n_{k+1}.$$

(This number is the μ measure of that part of the set $\bar X = X \times [0,1]$ where we have still not defined the conditional distribution $Q_n(\cdot,\cdot)$). Then let us also define the following probability measures $\bar\mu_n$ on the space X:

$$\bar\mu_n(C) = \frac{1}{P_n}\left[\mu_n(C) - \sum_{j=1}^{m(k)-1} \min\left[\mu_n(A_{j,k}), \mu(A_{j,k})\right] \frac{\mu_n(C \cap A_{j,k})}{\mu_n(A_{j,k})}\right], \quad C \in \mathcal{A},$$

and put

$$Q_n((x,u),C) = \bar\mu_n(C), \quad \text{if } C \in \mathcal{A}, \quad n_k \le n < n_{k+1}, \quad \text{and } x \in A_{j,k},\ j \ge m(k) \text{ or } x \in A_{j,k}, j < m(k) \text{ and } \lambda_{j,k,n} < u \le 1.$$

We claim that

$$\int Q_n((x,u),C)\mu(\,dx)\,du = \mu_n(C), \quad C \in \mathcal{A}. \tag{+}$$

Indeed, if $n_k \le n < n_{k+1}$, then $1 \le j < m(k)$ and in the case $C \in \mathcal{A}$

$$\begin{aligned}\int Q_n((x,u), C\cap A_{j,k})\mu(\,dx)\,du &= \mu(A_{j,k})\lambda_{j,k,n}\frac{\mu_n(C\cap A_{j,k})}{\mu_n(A_{j,k})} + \mu_n(C\cap A_{j,k}) \\ &- \min\left(\mu(A_{j,k}),\mu(A_{j,k})\right)\frac{\mu_n(C\cap A_{j,k})}{\mu_n(A_{j,k})} = \mu_n(C\cap A_{j,k}).\end{aligned}$$

On the other hand, put $B_k = X\setminus\left(\bigcup\limits_{j=1}^{m(k)-1} A_{j,k}\right)$. Then the system of sets $\{A_{j,k}, 1\le j < m(k), B_k\}$ supplies a partition of the space X. In particular, $A_{j,k}\cap B_k=\emptyset$ for all $1\le j<m(k)$. Observe that

$$\int Q_n((x,u), C\cap B_k)\mu(\,dx)\,du = \mu_n(C\cap B_k).$$

By summing up these identities we get relation (+), which means that the random variables ξ_n have the prescribed distribution μ_n for all numbers $n=1,2,\dots$.
On the other hand,

$$P\left(\sup_{n_k\le n<n_{k+1}}\rho(\xi_n,\xi)\ge\frac1k\,\middle|\,\bar\xi=(x,u)\right)=0$$

if $(x,u)\notin X_1(k)\subset\bar X$, where $\bar\xi(x,u,x_1,x_2,\dots)=(x,u)$, and

$$X_1(k)=B_k\times[0,1]\cup\bigcup_{j=1}^{m(k)-1}\left\{(x,u)\colon x\in A_{j,k},\ \inf_{n_k\le n<n_{k+1}}\lambda_{j,k,n}\le u\le 1\right\}.$$

Hence

$$\begin{aligned}P\left(\sup_{n_k\le n<n_k}\rho(\xi_n,\xi)>\frac1k\right)&\le\mu\times\lambda(X_1(k))\\ &=\sum_{j=1}^{m(k)}\left(1-\min_{n_k\le n<n_{k+1}}\lambda_{j,k,n}\right)\mu(A_{j,k})+\mu(B_k)\le\frac{2}{k^2}.\end{aligned}$$

Since

$$\sum_{k=1}^\infty P\left(\sup_{n_k\le n<n_k}\rho(\xi_n,\xi)>\frac1k\right)\le\sum_{k=1}^\infty\frac{2}{k^2}<\infty$$

it follows from the Borel–Cantelli lemma that $\xi_n\to\xi$ with probability one if $n\to\infty$.

6.) Let F be an arbitrary closed set. As $F=\bigcap\limits_{\delta\to0}F^\delta=F$, where F^δ denotes the open neighbourhood of radius δ of the set F, hence for all $\varepsilon>0$ there exists a $\delta=\delta(\varepsilon)>0$ such that $\mu(F^\delta)<\mu(F)+\varepsilon$. As $\xi_n\Rightarrow\xi$ stochastically, hence for all $\delta>0$ and

$\varepsilon > 0$ there exists such an index $n_0 = n_0(\varepsilon, \delta)$ for which $P(\rho(\xi, \xi_n) > \delta) < \varepsilon$. Furthermore, $\{\omega\colon \xi(\omega) \notin F^\delta\} \subset \{\omega\colon \xi_n(\omega) \notin F\} \cup \{\omega\colon \rho(\xi_n(\omega), \xi(\omega)) > \delta\}$, hence $1 - \mu(F^\delta) \le 1 - \mu_n(F) + \varepsilon$, and $\mu(F) + \varepsilon \ge \mu(F^\delta) \ge \mu_n(F) - \varepsilon$, if $n \ge n_0$. This implies that $\limsup\limits_{n\to\infty} \mu_n(F) \le \mu(F) + 2\varepsilon$. Since this inequality holds for all $\varepsilon > 0$, this implies that $\limsup\limits_{n\to\infty} \mu_n(F) \le \mu(F)$ for all closed sets F, i.e. $\mu_n \Rightarrow \mu$.

7.) To prove that $\xi = F^{-1}(\eta)$ is F distributed it is enough to show that

$$\{\omega\colon \eta(\omega) < F(x)\} \subset \{\omega\colon F^{-1}(\eta)(\omega) < x\} \subset \{\omega\colon \eta(\omega) \le F(x)\}.$$

The middle term is contained in the right-hand side term. Indeed, if $F^{-1}(\eta)(\omega) < x$ then there exists such a number $h > 0$ for which $F^{-1}(\eta)(\omega) = \sup\{u\colon F(u) < \eta(\omega)\} < x - h$. But this implies that $\eta(\omega) \le F(x)$. Indeed, if the relation $F(x) < \eta(\omega)$ held, then the number x would be in the set of numbers u whose supremum defines the quantity $F^{-1}(\eta)$, and this contradicts to the inequality $F^{-1}(\eta)(\omega) < x - h$.

To prove the left-hand side of this relation observe that if $\eta(\omega) < F(x)$, then $\eta(\omega) = F(x) - h$, and $F^{-1}(\eta(\omega)) = F^{-1}(F(x) - h)) = \sup\{v\colon F(v) < F(x) - h\}$ with an appropriate number $h > 0$. On the other hand, $\sup\{v\colon F(v) < F(x) - h\} < x$, because $F(x)$ is a function continuous from the left, and as a consequence $F(v) < F(x) - h$ implies that there exists such a number $\delta = \delta(h) > 0$, for which $v < x - \delta$, that is $F^{-1}(\eta(\omega)) = \sup\{v\colon F(v) < F(x) - h\} \le x - \delta < x$. Hence also the left-hand side part of the above relation holds.

To prove that $\tilde{F}(\xi, \varepsilon)$ is uniformly distributed in the interval $[0, 1]$, let us define the quantity $z(x) = \sup\{y\colon F(y) < x\}$ for all real numbers x. As the distribution function $F(x)$ is continuous from the left, hence $F(z) \ge x$. Let us consider the cases $F(z) = x$ and $F(z) < x$ separately.

If $F(z) = x$, then $F(u + 0) < x$ for all numbers $u < z$, and $\{\omega\colon \tilde{F}(\xi(\omega), \varepsilon(\omega)) < x\} = \{\omega\colon \xi(\omega) < z\}$. Hence $P(\tilde{F}(\xi, \varepsilon) < x) = P(\xi < z) = F(z) = x$.

If $F(z) < x$, then $F(z + 0) \ge x$, and

$$\begin{aligned}&\{\omega\colon \tilde{F}(\xi(\omega, \varepsilon(\omega)) < x\} \\ &\qquad = \{\omega\colon \xi(\omega) < z\} \cup \{\omega\colon \xi(\omega) = z, F(z) + \varepsilon(\omega)[F(z + 0) - F(z)] < x\}.\end{aligned}$$

Hence

$$\begin{aligned}P\left(\tilde{F}(\xi, \varepsilon) < x\right) &= P(\xi < z) + P(\xi(\omega) = z)\, P(\varepsilon(\omega)[F(z + 0) - F(z)] < x - F(z)) \\ &= F(z) + [F(z + 0) - F(z)]\frac{x - F(z)}{F(z + 0) - F(z)} = x.\end{aligned}$$

This implies the second statement of problem 7.

8.) The results of problem 7 imply that the distribution of the random variable $\bar{\xi}$ is F, and the distribution of $\bar{\eta}$ is G. Furthermore $\tilde{F}(\bar{\bar{\xi}}, \varepsilon)$ is uniformly and $\bar{\bar{\eta}}$ is G

distributed random variable. To show that the distribution of the vectors $(\bar\xi,\bar\eta)$ and $(\bar{\bar\xi},\bar{\bar\eta}$ agree, it is enough to show that $F^{-1}(\tilde F(\bar{\bar\xi},\varepsilon))=\bar{\bar\xi}$ with probability one. Indeed, this means that both random vectors can be represented (with the exception of a set of measure zero) as the transformation of an appropriate in the interval $[0,1]$ uniformly distributed random variable, (both coordinates of the vectors are obtained as the transform of the same random variable) and the representation of both vectors applies the same transformation.

The statement from which we can deduce the statement of problem 8 can be even weakened a little bit. It is enough to show that $P(F^{-1}(\tilde F(\bar{\bar\xi},\varepsilon))\le\bar{\bar\xi})=1$, since if one random variables is larger than another random variable with the same distribution functions with probability one, then these two random variables equal with probability one.

If $\bar{\bar\xi}(\omega)=x$, then

$$F^{-1}(\tilde F(\bar{\bar\xi}(\omega),\varepsilon(\omega)))=\sup\{u\colon F(u)<\tilde F(x,\varepsilon(\omega))\},$$

and since $F(v)\ge\tilde F(x,\varepsilon(\omega))$ if $v>x$, this implies that $F^{-1}(\tilde F(\bar{\bar\xi}(\omega),\varepsilon(\omega)))\le x=\bar{\bar\xi}(\omega)$ in this case. Problem 8 is solved.

9.) Let us remark that if ξ and η are defined by the formula $\xi=F^{-1}(\zeta)$ and $\eta=G^{-1}(\zeta)$, where ζ is a random variable with uniform distribution in the interval $[0,1]$, that is these random variables are defined by means of the quantile transform, then the two sides of the inequality investigated in this problem are equal. Let us first prove this inequality in the special case when the distributions of both random variables ξ and η are concentrated in a finite subset $X=\{x_1,\dots,x_n\}$, $x_1<x_2<\cdots<x_n$ of the real line. Put $p_j=P(\xi=x_j)$, $q_k=P(\eta=x_k)$, and let us consider the joint distribution of the random variables ξ and η $r(x_j,y_k)=P(\xi=x_j,\eta=x_k)$, $1\le j,k\le n$. Let us introduce the quantities

$$\begin{aligned}r(x_j,x_k,y_j,y_k)&=\min\left(r(x_j,y_k),r(x_k,y_j)\right),\\ r&=\max_{\substack{\{x_j,x_k,y_j,y_k\}\in X\times X\times X\times X\\ x_j<x_k,\ y_j<y_k}} r(x_j,x_k,y_j,y_k)\end{aligned}$$

Let us observe that if the vector (ξ,η) is constructed by means of the quantile transform, then $r=0$ for the above defined number r. Furthermore, the condition $r=0$ and the distribution F of the random variable ξ together the distribution G of the random variable η also determine the joint distribution of the vector (ξ,η), i.e. the property $r=0$ characterizes the quantile transform.

Indeed, we may assume without violating the generality that the probability space $(\Omega,\mathcal{A},P)$, where the random vector (ξ,η) is defined has the following structure: $\Omega=X\times X$, $\mathcal{A}$ is the discrete σ-algebra on $X\times X$, and the random variables are defined as $\xi(x_j,x_k)=x_j$ $\eta(x_j,x_k)=x_k$, $1\le j,k\le n$. In this space the measure P with the property $r=0$ we are looking for can be represented as the solution of the following “transport problem”: Let us consider the bipartitated graph consisting of the pairs (x_j, x_k), $1 \le j, k \le n$, attach the weights $p(x_j) = p_j = P(\xi = x_j)$ to the first coordinates and the weights $q(x_k) = q_k = P(\eta = x_k)$, $1 \le j, k \le n$ to the second coordinates. Then let us consider such a transports where weights p_j are sent from the first coordinates x_j, and weights q_k arrive to the second coordinates Let $r(j,k)$ denote the mass transported from the point x_j to x_k. Then $P(\xi = x_j, \eta = x_k) = r(j,k)$ denotes a joint distribution with the prescribed marginal distributions, and all joint distributions with such marginal distributions can be presented in such a way. The condition $r = 0$ means that the transport of the weights p_j to weights q_k is done in the following way: We transport the mass p_1 from the point x_1 first to the place x_1, then if some mass is still left to the point x_2, then to the point x_3, e.t.c. After this we transport the mass p_2 from the point x_2 coordinate to the smallest point where some empty space is still left, then to the next smallest space, e.t.c. Then we fill the smallest places where not all masses are sent to from the points x_3, x_4, e.t.c. This means in particular, that the condition $r = 0$ and the marginal distributions of the random variables ξ and η determine the joint distribution of the random vector (ξ, η).

We claim that

$$\min_{\substack{\xi \text{ is } F \text{ distributed} \\ \eta \text{ is } G \text{ distributed}}} E\Phi(\xi - \eta) = E\Phi(\bar\xi - \bar\eta), \tag{b}$$

where $(\bar\xi, \bar\eta)$ is a random vector with marginal distributions F and G such that $r = 0$.

Indeed, if the distribution of the random vector (ξ, η) is such that $r \neq 0$, then there exists a quadruple (x_j, x_k, y_j, y_k), $x_j < x_k$ and $y_j < y_k$ such that

$$\tilde r = r(x_j, x_k, y_j, y_k) = \min\left(r(x_j, y_k), r(x_k, y_j)\right) > 0.$$

We show that in this case a new joint distributions $\tilde r(x_j, x_k)$, $1 \le j, k \le n$, can be introduced in such a way that the marginal distribution of a random vector $(\tilde\xi, \tilde\eta)$ with this distribution has marginal distributions F and G, and $E\Phi(\xi - \eta) \le E\Phi(\tilde\xi - \tilde\eta)$. Furthermore, if Φ is a strictly convex function then there is a strict inequality in the last relation.

We shall define this probabilities in the following way.

$$\begin{aligned}
\tilde r(x_j, y_j) &= r(x_j, y_j) + \tilde r \\
\tilde r(x_k, y_k) &= r(x_k, y_k) + \tilde r \\
\tilde r(x_j, y_k) &= r(x_j, y_k) - \tilde r \\
\tilde r(x_k, y_j) &= r(x_k, y_j) - \tilde r \\
\tilde r(x, y) &= r(x, y) \quad \text{otherwise.}
\end{aligned}$$

Then the marginal distributions of the random vector $(\tilde\xi, \tilde\eta)$ are the prescribed ones, and

$$E\Phi(\tilde\xi - \tilde\eta) - E\Phi(\xi - \eta) = \tilde r \left(\Phi(x_j - y_j) + \Phi(x_k - y_k) - \Phi(x_j - y_k) - \Phi(x_k - y_j)\right).$$

This expression is non-negative. Moreover, it is strictly positive, if the function $\Phi(\cdot)$ is strictly convex. Indeed,

$$x_j - y_k < \frac{x_j - y_j}{x_k - y_k} < x_k - y_j,$$

hence the convexity of the function Φ implies that

$$\Phi(x_j - y_j) + \Phi(x_k - y_k) \le \Phi(x_j - y_k) + \Phi(x_k - y_j),$$

(here we also exploit that $(x_j - y_j) + (x_k - y_k) = (x_j - y_k) + (x_k - y_j)$,) and there is a strict inequality if Φ is a strictly convex function.

We show that the above relations imply formula (b). The proof of this statement is simpler in the case when $\Phi(\cdot)$ is a strictly convex function. Indeed, the function $E\Phi(\xi-\eta)$ takes its minimum somewhere if such pairs of vectors (ξ,η) are considered which are prescribed at the left-hand of formula (b). By the results of the previous paragraph this minimum is taken in the case $r = 0$. If the function $\Phi(\cdot)$ is convex, but not necessarily strictly monotone, then for all $\varepsilon > 0$ let us define the function $\Phi_\varepsilon(x) = \Phi(x) + \varepsilon x^2$. Then $\Phi_e(\cdot)$ is a strictly convex function, and we get by letting $\varepsilon \to 0$ the relation

$$\Phi(\xi - \eta) = \lim_{\varepsilon \to \varepsilon} \Phi_\varepsilon(\xi - \eta) \ge \lim_{\varepsilon \to \varepsilon} \Phi_\varepsilon(\bar\xi - \bar\eta) = \Phi(\bar\xi - \bar\eta).$$

This implies the statement of problem 9 in the case when the distribution of the random variables ξ and η are concentrated in finitely many points.

If F and G are two distribution functions concentrated to a finite interval, then it is useful to approximate these distributions by the discrete distributions $F_n(x) = F\left(\frac{[nx]}{n}\right)$ and $G_n(x) = G\left(\frac{[nx]}{n}\right)$. More precisely, it is worthwhile to approximate the two dimensional distribution $H(x,y)$ with marginal distributions $F(x)$ and $G(y)$ by the distribution $H_n(x,y) = H\left(\frac{[nx]}{n}, \frac{[ny]}{n}\right)$, where $[u]$ denotes the integer part of the number u. By applying the result of problem 9 in the already proven case and letting $n \to \infty$ we get the desired results in this case.

The general case can be reduced to the previous case by means of an appropriate limiting procedure. Let us truncate our random variables at the level $\pm u$. We do this truncation in such a way that if the vector (ξ,η) takes its value outside the square $[-u,u] \times [-u,u]$, then the truncated version of this vector has its value in the origin. Then carrying out the limit procedure $u \to \infty$ we can get the result of problem 9 in the general case. We make some comments which may be useful when carrying out this limit procedure. The convexity of the function Φ implies that $\Phi(x) \ge Ax + B$ for all real numbers x with appropriate constants A and B, hence $E\Phi(\xi - \eta) \ge -|A|E(|\xi| + |\eta| - |B| > -\infty$ under the conditions of problem 9. Let us remark that by adding an appropriate linear function to $\Phi(x)$ we can reduce the problem to the case when $\Phi(x) \ge 0$ for all real numbers x, and $\Phi(0) = 0$.

We may also assume that $E\Phi(\xi-\eta)<\infty$, since the statement of problem 9 is otherwise trivial. These remarks may simplify the limit procedure. At the left-hand side of the inequality we can apply the monotone convergence theorem, and at the right-hand side the lemma Fatou. We omit the details.

9a.) As $f(x)=|x|$ is a convex function, the result of problem 9 can be applied in this case. To finish the proof it is enough to show the identity

$$\int_0^1 \left|F^{-1}(x)-G^{-1}(x)\right|\,dx=\int_{-\infty}^{\infty}|F(x)-G(x)|\,dx.$$

This relation can be seen for instance by considering the domain

$$D=\{(x,y)\colon -\infty<x<\infty,\ \min(F(x),G(x))<y\le\max(F(x),G(x))\}\subset R^2,$$

and calculating the area of this domain first by integrating with respect to the variable y to get the left-hand side and then to integrate first with respect to the variable x to get the right-hand side of this identity. (Observe, that the intersection of the domain D with the horizontal line $y=u$ is an interval whose end-points are $\min(F^{-1}(u),G^{-1}(u))$ and $\max(F^{-1}(u),G^{-1}(u))$.)

10.) Let γ be the projection of the measure μ or (what is equivalent because of the conditions of the problem) of the measure ν to the space X_2. Furthermore, let $P(x,A)$ be the conditional distribution of the measurable sets of the form $A\times X_2$ with respect to prescribed value $x\in X_2$ on the space $(X_1\times X_2,\mathcal{A}_1\times\mathcal{A}_2,\mu)$, and $Q(x,C)$ the conditional distribution of the measurable sets of the form $X\times C$ with respect to prescribed values $x\in X_2$ on the space $(X_2\times X_3,\mathcal{A}_2\times \mathcal{A}_3,\nu)$. That is, let $\mu(A\times B)=\int_B P(x,A)\gamma(\,dx)$ for all measurable sets $A\in\mathcal{A}$, and $\nu(B\times C)=\int_B Q(x,C)\gamma(\,dx)$ for all sets $C\in\mathcal{A}_3$. (Such conditional distributions exist on complete separable metric spaces.) Let us define the measure of measurable sets of the form $A\times B\times C$ in the space $X_1\times X_2\times X_3$ by means of the formula $P(A\times B\times C)=\int_B P(x,A)Q(x,C)\gamma(\,dx)$. This measure can be extended to the whole space az $(X_1\times X_2\times X_3,\mathcal{A}_1\times\mathcal{A}_2\times\mathcal{A}_3)$, and it has the required properties.

11.) Let $P(x,A)$, where $x\in X$, A is a measurable subset of the space X, the conditional distribution of the second coordinate of the space $X\times X$ endowed with the measure ν with respect to the first coordinate, that is let $P(x,\cdot)$ be a probability measure on the measurable sets of the space X, $P(\cdot,A)$ a measurable function for all measurable sets $A\subset X$, and $\nu(A\times B)=\int_A P(x,B)\mu(\,dx)$ for all measurable sets $A\subset X$ and $B\subset X$. By a classical result of the probability theory (measure theory) such a conditional distribution exists.

Let us consider the probability space $(Q,\mathcal{B},\lambda)$, where $Q=[0,1]$, $\mathcal{B}$ is the Borel σ-algebra on the interval $[0,1]$, and λ is the Lebesgue measure on the σ-algebra $\mathcal{B}$ of the interval $[0,1]$. We claim that a set of random variables $\zeta(x,u)$, $u\in Q=[0,1]$, $x\in X$ indexed by a parameter $x\in X$ can be constructed on the probability space $(Q,\mathcal{B},\lambda)$ in such a way that $\zeta(x,\cdot)$ is a random variable with distribution $P(x,\cdot)$ for all points $x\in X$, where $P(x,A)$ is the conditional distribution defined in the previous paragraph, and $\zeta(\cdot,\cdot)$ is a measurable function on the product space $X\times[0,1]$ with the natural product σ-algebra. Actually we shall prove only a slightly weaker statement. We only claim the distribution of the random variable $\zeta(x,\cdot)$ agrees with the formerly given measure $P(x,\cdot)$ for almost all points x with respect to the measure μ. But this is not a real restriction. If we replace the measures $P(x,\cdot)$ by some other probability measures on a set of the points x with μ measure zero we get an equivalent version of the conditional distribution $P(x,A)$.

The above statement can be proved by a natural adaptation of the construction given in the solution of problem 3. In that problem we have described a possible construction of a random variable on the probability space $(Q,\mathcal{B},\lambda)$ with prescribed distribution on a separable complete metric space (X,ρ). We want to show that by applying this method we can construct a measurable function $\zeta(x,u)$, $x\in X$, $u\in[0,1]$ on the product space $X\times[0,1]$ in such a way that for a fixed $x\in X$ $\zeta(x,\cdot)$ is a $\mu=P(x,\cdot)$ distributed random variable on the probability space $([0,1],\mathcal{B},\lambda)$, where $\mathcal{B}$ denotes the Borel σ-algebra and λ the Lebesgue measure on the unit interval $[0,1]$. We construct such a function $\zeta(x,u)$ by applying the method of construction described in the solution of problem 3 with the following modification. We choose such partitions $\mathcal{X}_k$ whose elements $A_{j_1,\dots,j_k}$ has such boundaries for which the condition $\bar\mu(\partial A_{j_1,\dots,j_k})=0$ is satisfied, where the measure $\bar\mu$ is the projection of the measure ν to the second coordinate of the space $X\times X$. Then the properties of the conditional distributions imply that $P(x,(\partial A_{j_1,\dots,j_k}))=0$ for all partitions $\mathcal{X}_k$ and all (countably many) sets $A_{j_1,\dots,j_k}$ in these partitions except a measurable set of $x\in X$ with μ measure zero. Let us apply the construction described in problem 3 for all such points x for which the boundaries of all elements in the partitions have zero $P(x,\cdot)$ probabilities. For the sake of a complete construction let us define $\zeta(x,u)=x_0$, on the exceptional set of points x, where $x_0\in X$ is an arbitrary fixed point. Let us observe that by applying this construction we get the function $\zeta(x,u)$ as the limit of discrete valued measurable functions. Then the limit of these functions is also measurable. The random variables $\zeta(x,u)$ constructed in such a way have the desired properties.

Define $\eta(\omega)=\zeta(\xi(\omega),\chi(\omega))$ with the help of the above function $\zeta(x,u)$ and the independent random variables $\xi(\omega)$ and $\chi\omega)$ appearing in the formulation of problem 11. (The idea behind this definition is that $P(\zeta(\xi(\omega),\chi(\omega))\in B|\xi(\omega)=x)=P(\zeta(x,\chi(\omega)\in B)=P(x,B)$.) We claim that this random variable η satisfies the statement of the problem, that is $P(\xi(\omega)\in A,\eta(\omega)\in B)=\nu(A\times B)$ for all measurable sets $A\subset X$ and $B\subset X$. Indeed, since the distribution of the random vector $(\chi(\omega),\xi(\omega))$ is $dv\,\mu(dx)$ on the space $[0,1]\times X$, hence we get that

$$\begin{aligned}
P(\xi(\omega)\in A,\eta(\omega)\in B)&=E\left(I(\xi(\omega)\in A)I(\zeta(\xi(\omega),\chi(\omega))\in B)\right)\\
&=\int\left(\int I\left(x\in A\right)I(\zeta(x,v)\in B)\,dv\right)\mu(\,dx)\\
&=\int I(x\in A)P(x,B)\mu(\,dx)
\end{aligned}$$

$$= \int_A P(x,B)\mu(\,dx) = \nu(A\times B),$$

where $I(C)$ is the indicator function of the set C. In such a way we have solved problem 11.

12.) The measures μ and ν have a decomposition of the following form: $\mu=\gamma+\mu_1$, $\nu=\gamma+\nu_1$, where the measure γ is the "joint part" of the measures μ and ν, the measures μ_1 and ν_1 are singular to each other. This means that there exists such a set $C\in A$, for which the measure μ_1 is concentrated to the set C and the measure ν_1 is concentrated to the set $X\setminus C$, that is $\mu_1(X\setminus C)=0$, and $\nu_1(C)=0$. To see that such a decomposition exists let us consider a measure dominating both measures μ and ν (we can choose for instance the measure $\frac{\mu+\nu}{2}$ as this dominating measure), let $f(x)$ be the density function of the measure μ and $g(x)$ the density function of the measure ν with respect to this dominating measure. Let us define the measure γ as the measure with density function $\min(f(x),g(x))$, the measure μ_1 as the measure with density function $f(x)-\min(f(x),g(x))$ the measure ν_1 as the measure with density function $g(x)-\min(f(x),g(x))$ with respect to the dominating measure. Finally, define the set C by the formula $C=\{x\colon f(x)\ge g(x)\}$. This decomposition satisfies the requested properties. Beside this, $\mu(A)-\nu(A)\le\mu(C)-\nu(C)$ for all sets $A\in\mathcal{A}$. Let us observe that $\mu(C)-\nu(C)=\mu(C)-\gamma(C)=\mu_1(C)=\mu_1(X)=1-\gamma(X)$, which implies that $\sup\limits_{A\in\mathcal{A}}(\mu(A)-\nu(A))=1-\gamma(X)$. The quantities $\nu(A)-\mu(A)$ can be estimated similarly. As a consequence, we get that $\operatorname{Var}(\mu,\nu)=\sup\limits_{A\in\mathcal{A}}|\mu(A)-\nu(A)|=1-\gamma(X)$. If ξ and η are μ respectively ν distributed random variables, then

$$P(\xi\ne\eta)\ge P(\{\xi\in C\}\cap\{\eta\notin C\})\ge\mu(C)-\nu(C)=\operatorname{Var}(\mu,\nu).$$

To make such a construction where the two sides of the above inequality are equal let us define the following probability space $(\Omega,\mathcal{B},P)$. Put $\Omega=X\cup(X\times X)$, and let $\mathcal{B}$ be the natural σ-algebra on the space Ω whose restriction to the set X is $\mathcal{A}$, and to the set $X\times X$ $\mathcal{A}\times\mathcal{A}$. Let us define the measure P in the following way. Let the restriction of the measure μ P to the set X γ, and to the set $X\times X$ $D\cdot\mu_1\times\nu_1$, where $D^{-1}=\mu_1(X)=\nu_1(X)$, that is D is the natural norming factor. Let us define the random variables ξ and η in the following way: If $\omega=x\in X$, than $\xi(\omega)=\eta(\omega)=x$, if $\omega=(x_1,x_2)\in X\times X$, then $\xi(\omega)=x_1$, $\eta(\omega)=x_2$. With such a definition the random variable ξ is μ and the random variable η is ν distributed random variable. Furthermore, $P(\xi=\eta)=\gamma(X)=1-\operatorname{Var}(\mu,\nu)$, since the restriction of the measure P to the set $X\times X$ is concentrated on the set $C\times(X\setminus C)$, where $\xi(\omega)\ne\eta(\omega)$. We have solved problem 12.

Remark: It had some technical reasons why the random variables which yield identity in the inequality of problem 12 were constructed on the probability space $\Omega=X\cup(X\times X)$ and not on the space $\Omega_1=X\times X$, which might have been a more natural choice. But if we had worked in the probability space Ω_1, then the measure γ should have been concentrated to the diagonal $D = \{(x,x)\colon x \in X\}$. On the other hand, there are cases when this diagonal D is a non-measurable subset of the space $\Omega_1 = X \times X$, and this causes some problems. We wanted to avoid this difficulty and make a construction which also works in such cases.

13.) It is enough to prove that under the conditions of the central limit theorem a triangular array $\tilde\xi_{k,j}$, $k = 1,2\dots,$ $1 \le j \le n_k$, and Wiener processes $W_k(t)$, $0 \le t \le 1$, $k = 1,2,\dots$, can be constructed in such a way that the distribution of the random variables $\tilde\xi_{k,j}$ and $\xi_{k,j}$ agree, and the random broken line $\tilde S_{k,j}(\cdot)$ defined with the help of the random variables $\tilde\xi_{k,j}$ in the same way as the random broken line $S_k(t)$ with the help of the random variables $\xi_{k,j}$ satisfy, together with the constructed Wiener processes $W_k(t)$, the relation $\sup\limits_{0\le t\le 1} |\tilde S_k(t) - W_k(t)| \Rightarrow 0$ as $k \to \infty$, where $\Rightarrow$ denotes stochastic convergence.

Let us observe that the Lindeberg condition implies that $\lim\limits_{k\to\infty} \sup\limits_{1\le j\le n_k} \sigma^2_{k,j} = 0$, where $\sigma^2_{k,j} = E\xi^2_{k,j}$. Let us consider the sequences $u_{k,l} = \sum\limits_{j=1}^{l} \sigma^2_{k,j}$, fix some $\varepsilon > 0$, of the form $\varepsilon = \dfrac{1}{M}$, where M is a positive integer, and define for all $k = 1,2,\dots$, the subsequence $0 = m_{k,0} < m_{k,1} < \cdots < m_{k,s}$, $m_{k,r} = m_{k,r}(\varepsilon)$, $1 \le r \le s$, $s = s(k,\varepsilon)$, of the sequence $0,1,2,\dots,n_k$ in the following way: Put $m_{k,0} = 0$, and if $m_{k,r}$ is already defined, then the number $m_{k,r+1}$ is defined as the number $L > m_{k,r}$ for which $u_{k,L} - u_{k,m_{k,r}} \ge \varepsilon$, $u_{k,L-1} - u_{,mk,r} < \varepsilon$ if $u_{k,n_k} - u_{k,m_{k,r}} \ge \varepsilon$, and $m_{k,r+1} = n_k$ and $r+1 = s(k,\varepsilon)$ is otherwise. Put $\bar u_{k,j} = \dfrac{u_{k,j}}{u_{k,n_k}}$, $1 \le j \le n_k$. Let us observe that $\lim\limits_{k\to\infty} \sup\limits_{1\le r\le s(k,\varepsilon)} \dfrac{|\bar u_{k,m_r} - \bar u_{k,m_{r-1}} - \varepsilon|}{\varepsilon} = 0$, and the random variables $T_{k,j} = \dfrac{S_{k,m_{k,r_j}} - S_{k,m_{k,r_{j-1}}}}{\sqrt{\bar u_{k,m_{k,j}} - \bar u_{k,m_{k,j-1}}}}$ satisfy the central limit theorem with arbitrary indices j for which $1 \le r_j \le s(k,\varepsilon)$, and $S_{k,r} = \sum\limits_{j=1}^{r} \xi_{k,j}$. Let us also observe that $s(k,\varepsilon) \le \varepsilon^{-1}$.

The above results together with the coupling results of this paper (e.g.problem 2 or Statement A can be applied) imply that such pairs of independent random variables $(\tilde T_{k,j}, X_{k,j})$, $1 \le j \le s(k,\varepsilon)$ can be constructed for all $k = 1,2,\dots$ for which the random variables $X_{k,j}$ have standard normal distribution, the distributions of the random variables $T_{k,j}$ and $\tilde T_{k,j}$ agree, and $\sup\limits_{1\le j\le s(k,\varepsilon)} \left|\tilde T_{k,j} - X_{k,j}\right| \Rightarrow 0$, where $\Rightarrow$ denotes convergence in probability. Let $Z_{k,l,\varepsilon} = \sum\limits_{j=1}^{l} \sqrt{\bar u_{k,m_{k,j}} - \bar u_{k,m_{k,j-1}}}\tilde T_{k,j}$, $Y_{k,l,\varepsilon} = \sum\limits_{j=1}^{l} \sqrt{\bar u_{k,m_{k,j}} - \bar u_{k,m_{k,j-1}}} X_{k,j}$, $1 \le l \le s(k,\varepsilon)$, and put $Z_{k,0,\varepsilon} = 0$ and $Y_{k,0,\varepsilon} = 0$, Then the relation $\sup\limits_{1\le l\le s(k,\varepsilon)} |Z_{k,l,\varepsilon} - Y_{k,l,\varepsilon}| \Rightarrow 0$ also holds. Since this relation holds for all numbers ε of the form $\varepsilon=\dfrac{1}{M}$, the relation

$$\sup_{1\le l\le s(k,\varepsilon_k)}|Z_{k,l,\varepsilon_k}-Y_{k,l,\varepsilon_k}|\Rightarrow 0,\quad \text{and } \varepsilon_k\to 0 \text{ if } k\to\infty \tag{1}$$

also holds with an appropriate sequence ε_k, $k=1,2,\dots$. Because of the distribution of the sequences $(Z_{k,j,\varepsilon_k},\ 1\le j\le s(k,\varepsilon_k))$, $(Y_{k,j,\varepsilon_k},\ 1\le j\le s(k,\varepsilon_k)$ and the result of Problem 10 a triangular array $\tilde S_{k,j}$ and a sequence of Wiener processes $W_k(t)$ can be constructed in such a way that the distribution of the sequences $\tilde\xi_{k,j}$ and $\xi_{k,j}$, $1\le j\le n_k$, agree for a fixed $k=1,2,\dots$, and beside this the partial sums $\tilde S_{k,l}=\sum\limits_{j=1}^{l}\tilde\xi_{k,j}$ satisfy the identity $\tilde S_{k,m_{k,l}}=Z_{k,l,\varepsilon_k}$, $W_k(\bar u_{k,m_{k,l}})=Y_{k,l,\varepsilon_k}$, $1\le l\le s(k,\varepsilon_k)$. (Actually we construct directly the partial sums $\tilde S_{k,j}$ and not the random variables $\xi_{k,j}$. We do this under the condition that their values are prescribed for certain indices and their distributions must agree with the distribution of the corresponding partial sums of the random variables $\xi_{k,j}$. The Wiener processes $W_k(\cdot)$ can be constructed in a simpler way by using its Markov property and some basic facts about Wiener processes.)

We claim that the random broken lines $\tilde S_k(t)$ determined by the above constructed triangular array $\tilde\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$ together with the Wiener processes $W_k(t)$ constructed together with it satisfy the relation $\sup\limits_{0\le t\le 1}|\tilde S_k(t)-W_k(t)|\Rightarrow 0$. It follows from the construction and relation (1) that

$$\sup_{1\le r\le s(k,\varepsilon_k)}\left|\tilde S_k(\bar u_{k,m_{k,r}})-W_k(\bar u_{k,m_{k,r}})\right|\Rightarrow 0,$$

and beside this $\lim\limits_{k\to\infty}\sup\limits_{1\le r\le s_k}\left(\bar u_{i,m_{k,r}}-\bar u_{k,m_{k,r-1}}\right)=0$. Hence to prove the functional central limit theorem it is enough to prove that

$$P\left(\sup_{0\le r<s(k,\varepsilon_k)}\ \sup_{\bar u_{k,m_{k,r}}\le u\le\bar u_{k,m_{k,r+1}}}\left|W_k(u)-W_k(\bar u_{k,m_{k,r}})\right|>\varepsilon\right)$$
$$\le\sum_{0\le r<s(k,\varepsilon_k)}P\left(\sup_{\bar u_{k,m_{k,r}}\le u\le\bar u_{k,m_{k,r+1}}}\left|W_k(u)-W_k(\bar u_{k,m_{k,r}}\right|>\varepsilon\right)\to 0 \tag{2a}$$

$$P\left(\sup_{0\le r<s(k,\varepsilon_k)}\ \sup_{\bar u_{k,m_{k,r}}\le u\le\bar u_{k,m_{k,r+1}}}\left|\tilde S_k(u)-\tilde S_k(\bar u_{k,m_{k,r}})\right|>\varepsilon\right)$$
$$\le\sum_{0\le r<s(k,\varepsilon_k)}P\left(\sup_{\bar u_{k,m_{k,r}}\le u\le\bar u_{k,m_{k,r+1}}}\left|\tilde S_k(u)-\tilde S_k(\bar u_{k,m_{k,r}})\right|>\varepsilon\right)\to 0 \tag{2b}$$

as $k\to\infty$ for all $\varepsilon>0$. Relation (2.2b) can be rewritten because of the structure of the random broken lines $\tilde S_k(t)$ as

$$\sum_{0\le r<s(k,\varepsilon_k)}P\left(\sup_{m_{k,r}+1\le p\le m_{k,r+1}}\left|\sum_{j=m_{k,r}+1}^{p}\tilde\xi_{k,j}\right|>\varepsilon\right)\to 0\quad\text{for all } \varepsilon>0. \tag{2c}$$

The proof of relation (2a) is relatively simple, since the probabilities at its right-hand side can be calculated explicitly. Indeed,

$$
\begin{aligned}
&P\left(\sup_{\bar u_{k,m_{k,r}}\le u\le \bar u_{k,m_{k,r+1}}}\left|W_k(u)-W_k(\bar u_{k,m_{k,r}}\right|>\varepsilon\right)\\
&\qquad =P\left(\sup_{0\le u\le 1}|W_k(u)|>\frac{\varepsilon}{\sqrt{\bar u_{k,m_{k,r+1}}-\bar u_{k,m_{k,r}}}}\right)\\
&\qquad \le 4\left(1-\Phi\left(\frac{\varepsilon}{\sqrt{\bar u_{k,m_{k,r+1}}-\bar u_{k,m_{k,r}}}}\right)\right),
\end{aligned}
$$

where $\Phi(\cdot)$ denotes the standard normal distribution function. Because of the relation $\lim\limits_{k\to\infty}\sup\limits_{1\le r<s(k,\varepsilon_k)}(\bar u_{k,m_{k,r+1}}-\bar u_{k,m_{k,r}})=0$ we have

$$
\lim_{k\to\infty}\frac{\left(1-\Phi\left(\frac{\varepsilon}{\sqrt{\bar u_{k,m_{k,r+1}}-\bar u_{k,m_{k,r}}}}\right)\right)}{\bar u_{k,m_{k,r+1}}-\bar u_{k,m_{k,r}}}=0.
$$

The above relations together with the identity $\sum\limits_{r=1}^{s(k,\varepsilon_k)}(\bar u_{k,m_{k,r+1}}-\bar u_{k,m_{k,r}})=1$ imply formula (2a).

The proof of formula (2c) needed to complete the proof of the functional central limit theorem is more difficult. We describe one possible proof. Let us first observe that because of the Lindeberg condition there exists some sequence δ_k of positive numbers, $k=1,2,\dots$, such that $\lim\limits_{k\to\infty}\delta_k=0$ and $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}E\tilde\xi_{k,j}^2I(|\tilde\xi_{k,j}|>\delta_k)=0$. Let us fix such a sequence δ_k, and let us choose the decomposition $\tilde\xi_{k,j}=\bar\xi_{k,j}+\bar{\bar\xi}_{k,j}$ with $\bar\xi_{k,j}=\tilde\xi_{k,j}I(|\tilde\xi_{k,j}|>\delta_k)-E\tilde\xi_{k,j}I(|\tilde\xi_{k,j}|>\delta_k)$ and $\bar{\bar\xi}_{k,j}=\tilde\xi_{k,j}I(|\tilde\xi_{k,j}|\le\delta_k)-E\tilde\xi_{k,j}I(|\tilde\xi_{k,j}|\le\delta_k)$. Then $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}E\bar\xi_{k,j}^2=0$, $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}E\bar{\bar\xi}_{k,j}^2=1$, $E\bar{\bar\xi}_{k,j}^4\le\delta_k^2E\bar{\bar\xi}_{k,j}^2$, and to prove relation (2c) it is enough to show that

$$
\sum_{0\le r<s(k,\varepsilon_k)}P\left(\sup_{m_{k,r}+1\le p\le m_{k,r+1}}\left|\sum_{j=m_{k,r}+1}^{p}\bar\xi_{k,j}\right|>\varepsilon\right)\to 0\quad\text{for all }\varepsilon>0,\tag{3a}
$$

$$
\sum_{0\le r<s(k,\varepsilon_k)}P\left(\sup_{m_{k,r}+1\le p\le m_{k,r+1}}\left|\sum_{j=m_{k,r}+1}^{p}\bar{\bar\xi}_{k,j}\right|>\varepsilon\right)\to 0\quad\text{for all }\varepsilon>0.\tag{3b}
$$

We can write $P\left(\sup\limits_{m_{k,r}+1\le p\le m_{k,r+1}}\left|\sum\limits_{j=m_{k,r}+1}^{p}\bar\xi_{k,j}\right|>\varepsilon\right)\le\frac{\sum\limits_{j=m_{k,r}+1}^{m_{k,r+1}}E\bar\xi_{k,j}^2}{\varepsilon^2}$ by the Kolmogorov inequality. We get relation (3.1) by summing up this relation for $0\le r\le$

$s(k,\varepsilon_k)$ and applying the relation $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k} E\bar\xi_{k,j}^2=0$.

The proof of relation (3b) demands a more intricate argument. Let us observe that the following version of the Kolmogorov inequality holds.

$$P\left(\sup_{m_{k,r}+1\le p\le m_{k,r+1}}\left|\sum_{j=m_{k,r}+1}^{p}\bar{\bar\xi}_{k,j}\right|>\varepsilon\right)\le\frac{E\left(\sum\limits_{j=m_{k,r}+1}^{m_{k,r+1}}\bar{\bar\xi}_{k,j}\right)^4}{\varepsilon^4}.$$

Indeed, the sequence of the random variables $\left(\sum\limits_{j=m_{k,r}+1}^{p}\bar{\bar\xi}_{k,j}\right)^4$, $p=m_{k,r}+1,\dots,$ $m_{k,r+1}$, is a semimartingale. (This follows for instance from the fact that the convex function of a martingale is a semimartingale.) Then the above inequality can be proved similarly to the proof of the Kolmogorov inequality with the help of this fact. We can prove formula (3b) with the help of a good estimate on $E\left(\sum\limits_{j=m_{k,r}+1}^{m_{k,r+1}}\bar{\bar\xi}_{k,j}\right)^4$. Indeed, we have

$$\begin{aligned}E\left(\sum_{j=m_{k,r}+1}^{m_{k,r+1}}\bar{\bar\xi}_{k,j}\right)^4&\le\sum_{j=m_{k,r}+1}^{m_{k,r+1}}E\bar{\bar\xi}_{k,j}^4+3\left(\sum_{j=m_{k,r}+1}^{m_{k,r+1}}E\bar{\bar\xi}_{k,j}^2\right)^2\\&\le\left(\delta_k^2+3\sup_{0\le r<s(,\varepsilon_k)}\left(u_{k,m_{k,r+1}}-u_{k,m_{k,r}}\right)\right)\left(\sum_{j=m_{k,r}+1}^{m_{k,r+1}}E\bar{\bar\xi}_{k,j}^2\right).\end{aligned}$$

(Here we have exploited that because of the relation $E\bar{\bar\xi}_{k,j}=0$ and the independence of the random variables $\bar{\bar\xi}_{k,j}$ by carrying out the multiplications in the expression $E\left(\sum\limits_{j=m_{k,r}+1}^{m_{k,r+1}}\bar{\bar\xi}_{k,j}\right)^4$ we get a sum whose summands are the expected value of products of the random variables $\xi_{k,j}$. Those summands which contain a term with power one equal zero.) The above relations imply that

$$\begin{aligned}&\sum_{0\le r<s(k,\varepsilon_k)}P\left(\sup_{m_{k,r}+1\le p\le m_{k,r+1}}\left|\sum_{j=m_{k,r}+1}^{p}\bar{\bar\xi}_{k,j}\right|>\varepsilon\right)\\&\qquad\le\frac{\left(\delta_k^2+3\sup\limits_{0\le r<s(,\varepsilon_k)}\left(u_{k,m_{k,r+1}}-u_{k,m_{k,r}}\right)\right)\left(\sum\limits_{j=1}^{n_k}E\bar{\bar\xi}_{k,j}^2\right)}{\varepsilon^4}.\end{aligned}$$

Formula (3b) follows from this inequality, since $\lim\limits_{k\to\infty}E\sum\limits_{j=1}^{n_k}\bar{\bar\xi}_{k,j}^2=1$, $\lim\limits_{k\to\infty}\delta_k=0$, and $\lim\limits_{k\to\infty}\sup\limits_{0\le r<s(k,\varepsilon_k)}\left(u_{k,m_{k,r+1}}-u_{k,m_{k,r}}\right)=0$.

14.) Let us consider a sequence of random variables ξ_n with distribution μ_n, $n = 0, 1, 2, \dots$, on the space (X, ρ) such that the random variables ξ_n converge with probability one to the random variable ξ_0 as $n \to \infty$. This is possible by the result of Problem 5. Then also the $\mathcal{F}\mu_n$ distributed random variables $\eta_n = \mathcal{F}\xi_n$ converge to the $\mathcal{F}\mu_0$ distributed random variable $\eta_0 = \mathcal{F}\xi_0$ with probability one. This implies the statement of Problem 14.

Appendix

The proof of the König–Hall theorem. The necessity of the condition is obvious. We prove sufficiency by induction with respect to the size of the set Y. If $|Y| = 1$, then the statement is obvious. Let us assume that we know the sufficiency of the condition for $|Y| = k < n$, and let us prove it for $|Y| = n$. We distinguish two cases:

a.) There exists a set $A \subset Y$ such that $0 < |A| = k < n$, and $|B(A)| = |A|$.

b.) For all sets $A \subset Y$ such that $0 < |A| < n$ $|B(A)| > |A|$.

In case a.) we claim that the conditions of the theorem hold for both pairs of sets $\bar{Y} = A$, $\bar{Z} = B(A)$ and $\bar{Y} = Y \setminus A$, $\bar{Z} = Z \setminus B(A)$ and the restriction of the function $d(y, z)$ to these sets $\bar{Y} \times \bar{Z}$. Then the inductive hypothesis implies the sufficiency of the condition in case a.) In case $\bar{Y} = A$, $\bar{Z} = B(A)$ this statement is obvious. If $\bar{Y} = Y \setminus A$ and $\bar{Z} = Z \setminus B(A)$ let us consider the set $C \subset Y \setminus A$. Put $\bar{C} = C \cup A$. Then $|C| = |\bar{C}| - k$, $|B(\bar{C})| \ge |\bar{C}|$, and $B(C) \cap \bar{Z} \supset B(\bar{C}) \setminus B(A)$. Hence $|B(C) \cap \bar{Z}| \ge |B(\bar{C})| - k \ge |\bar{C}| - k$, and $|B(C) \cap \bar{Z}| \ge |C|$, as we have claimed.

In case b.) let us consider a point $z_j \in Z$, for which $d(y_1, z_j) = 1$. Let us pair the point y_1 with the point z_j. To complete the proof of the theorem it is enough to show that the sets $\bar{Y} = Y \setminus \{y_1\}$ and $\bar{Z} = Z \setminus \{z_j\}$ satisfy the conditions of the theorem in case b.). But this is obvious, since in case b.) $|B(A) \cap \bar{Z}| \ge |B(A)| - 1 \ge |A|$ for all sets $A \subset Z$.

The proof of the continuous version of the König–Hall theorem. The necessity of the condition is obvious also in this case. We prove the sufficiency of the conditions by means of the original König–Hall theorem.

Let us first consider the special case when there is an integer N such that $u(y_j) = \dfrac{k_j}{N}$ for all $y_j \in Y$ and $v(z_l) = \dfrac{p_l}{N}$ for all $z_l \in Z$, where k_j and p_l are integers. In this case let us define the following bipartitated graph: $\bar{Y} = \{(y_j, m(j)), y_j \in Y,\ 1 \le m(j) \le k_j\}$, $\bar{Z} = \{(z_l, n(l)), z_l \in Z,\ 1 \le n(l) \le p_l\}$, $\bar{d}((y_j, m(j)), (z_l, n(l))) = d(y_j, z_l)$. Then the bipartitated graph $(\bar{Y}, \bar{Z}, \bar{d}(\cdot, \cdot))$ satisfies the conditions of the König–Hall theorem. Indeed, it is enough to check the conditions of the König–Hall theorem for those sets $A \subset \bar{Y}$ for which in the case $\bar{y} = (y_j, m(j)) \in A$ with some points $(y_j, m(j)) \in \bar{Y}$ also the relation $(y_j, k) \in A$ holds for all points (y_j, k), $1 \le k \le k_j$. On the other hand, this condition agrees with the relation $\sum\limits_{z \in B(A)} v(z) \ge \sum\limits_{y \in A} u(y)$ for all $A \in Y$.

Hence there exists such a pairing of the elements of the sets $\bar{Y}$ and $\bar{Z}$, in which the points $\bar{y}$ and $\bar{z}$ in a pair satisfy the relation $\bar{d}(\bar{y},\bar{z}) = 1$. Put $\bar{w}(\bar{y},\bar{z}) = 1$, if $\bar{y}$ and $\bar{z}$ are in a pair, and $\bar{w}(\bar{y},\bar{z}) = 0$ otherwise. Then the function $w(y_j, z_l) = \frac{1}{N} \sum_{\substack{\bar{y}=(y_j,m(j)),1\le m(j)\le k_j \\ \bar{z}=(z_l,n(l)),1\le n_l\le p_l}} \bar{w}(\bar{y},\bar{z})$ satisfies the statement of the theorem in this special case.

The general case can be reduced to the already proved situation with an appropriate approximation. We can assume without violating the generality that $\sum_{y\in Y} u(y) = \sum_{z\in Z} v(z) = 1$. For all $N = 1, 2, \dots$ let us define the following system approximating the original one. Put $\bar{Y} = \bar{Y}_N = Y \cup \{r+1\}$, $\bar{Z} = \bar{Z}_N = Z \cup \{s+1\}$, $\bar{d}(y,z) = \bar{d}_N(y,z) = d(y,z)$, if $y \in Y$, $z \in Z$, and $\bar{d}(y_{r+1}, z) = \bar{d}_N(y_{r+1}, z) = \bar{d}(y, z_{s+1}) = \bar{d}_N(y, z_{s+1}) = 1$, (this means that the points y_{r+1} and z_{s+1} are connected to all points of the other set), $\bar{u}_N(y) = \frac{[Nu(y)]}{N}$, if $y \in Y$, $\bar{v}_N(z) = \frac{[Nv(z)]}{N}$, if $z \in Z$, where $[u]$ denotes the integer part of the number u, beside this $\bar{u}_N(y_{r+1}) = \sum_{y\in Y} (u(y) - \bar{u}_N(y)) = 1 - \sum_{y\in Y} \bar{u}_N(y)$, and $\bar{v}_N(y_{s+1}) = \sum_{z\in Z} (v(z) - \bar{v}_N(z)) = 1 - \sum_{z\in Z} \bar{v}_N(z)$. Let us observe that this new system also satisfies the conditions of the theorem. Indeed, if $A \subset Y_N$, then the set of points $B_N(A)$ of $\bar{Z}_N$ connected to a set $A \subset Y_N$ is $\bar{B}_N(A) = B(A) \cup \{z_{s+1}\}$, and $\sum_{y\in A} \bar{u}_N(y) \le \sum_{y\in A} u(y) \le \sum_{z\in B(A)} v(z) \le \sum_{z\in B(A)} \bar{v}_N(z) + \bar{v}_N(z_{s+1}) \le \sum_{z\in \bar{B}(A)} \bar{v}_N(z)$. If $y_{r+1} \in A$, then $\bar{B}_N(A) = \bar{Z}_N$, and $\sum_{y\in A} \bar{u}_N(y) \le 1 = \sum_{z\in \bar{Z}_N} \bar{v}_N(z)$. Hence the already proved part of the Theorem can be applied in this case.

Let us consider for all $N = 1, 2, \dots$ a "transport function" $\bar{w}_N(y,z)$, $y \in \bar{Y}$, $z \in \bar{Z}$ satisfying the theorem. As $0 \le \bar{w}_N(y,z) \le 1$ for all points $y \in \bar{Y}$, $z \in \bar{Z}$, and numbers $N = 1, 2, \dots$ a subsequence $N_k \to \infty$ of the integers can be chosen in such a way that the limit $w(y,z) = \lim_{k\to\infty} w_{N_k}(y,z)$ exists for all points $y \in \bar{Y}$ and $z \in \bar{Z}$. This function $w(y,z)$ satisfies the theorem.