

On the Continuity Property of the Probability Function and Its Applications

Subhash Bagui ¹ and K. L. Mehra ²

¹Department of Mathematics and Statistics
The University of West Florida
Pensacola, FL 32514 USA

²Department of Mathematical and Statistical Sciences
University of Alberta
Edmonton, AB, T6G 2G1, Canada

This article is distributed under the Creative Commons by-nc-nd Attribution License.
Copyright © 2021 Hikari Ltd.

Abstract

In this paper, we study the ‘continuity’ property of probability functions and explore thoroughly its well-known “equivalence” to the axiom of ‘countable additivity’. Students in early university courses in Probability and Statistics are often quite fearful of dealing with topics involving ‘convergence’ of random variables - and their distributions - where the ‘continuity’ property plays a major role, namely, those involving notions of almost sure (a.s.) convergence, convergence in probability (pr.) or convergence in distribution (d) etc.. We illustrate this via simple well-structured proofs and multiple informative examples. A lack of adequate emphasis on the ‘continuity’ property in early courses creates a gap in students’ understanding of the role of convergence of distributions in large sample statistical analysis. This article aims to fill up this gap. The article is directed towards students and teachers of introductory courses in Probability and Statistics.

Keywords: Continuity property of Probability, \liminf , \limsup , Almost Sure convergence, In-probability convergence, In-distribution convergence, Murphy's law

1. Introduction

Probability as a concept emanates from our intuitive understanding of random phenomena and the occurrence or non-occurrence of events under their incidence, and so does all its associated terminology. As a function, it assigns a number to each possible event under the phenomenon, a number that indicates the extent of its likelihood to occur or not occur. To add clarity and precision to the preceding assertion, let E denote the random phenomenon or experiment at hand and $\Omega = \Omega_{\mathcal{E}}$ the associated sample space, namely, the set of all possible individual outcomes of E – termed its sample points. If Ω is countable, i.e., either finite or denumerable, then all subsets of Ω , including the empty set ϕ and the whole space Ω , are possible events. If, however, Ω is uncountable (infinite), only the measurable subsets, i.e., those to which probability numbers are assignable, including the null event ϕ and the sure event Ω – are events (see the definition of σ -algebra of events below).

In general, clearly though, if $A, B, C \dots$ etc. are events, so should be their unions, intersections, and complementary subsets $A \cup B$, $A \cap B$ and $A^c = \Omega - A$, $B^c = \Omega - B$ etc., respectively. If a collection of subsets \mathbf{Q} is closed under finite unions, intersections, and complementation, it is called a 'field' or 'algebra'. If it is also closed additionally under all countable – finite or denumerable – set operations, it is designated as a σ -field or σ -algebra of events. Thus, for example, for any denumerable sequence $\{A_n : n \geq 1\}$ of events in a σ -

algebra \mathbf{Q} , $\bigcup_{n=1}^{\infty} A_n$ and $\bigcap_{n=1}^{\infty} A_n$ would also be events belonging to it; so too would be the events

$$\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \text{ or } \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Example 1. If the real line R is the sample space, the class C of open, closed, or half-open – half-closed intervals (a, b) , $[a, b]$, $(a, b]$, $[a, b)$, where $-\infty \leq a < b \leq +\infty$ with a, b infinite only on the open side, can be events. The minimal σ -field containing C is known as Borel field of events. It contains the minimal field containing C . The Borel field can be generated by adding to it all sets obtained through countable set operations, including complementation, on elements of C .

Definition 1.1. [3]. For any denumerable sequence of events $\{A_n : n \geq 1\}$ in Ω , whether convergent or not, the limiting events

$$(1.1) \quad \begin{cases} \liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \lim_n \bigcap_{k=n}^{\infty} A_k \text{ and} \\ \limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \lim_n \bigcup_{k=n}^{\infty} A_k \end{cases}$$

always exist. Since $\bigcap_{k=n}^{\infty} A_k \subset A_n \subset \bigcup_{k=n}^{\infty} A_k$ for all n , with $(\bigcap_{k=n}^{\infty} A_k) \uparrow$ and $(\bigcup_{k=n}^{\infty} A_k) \downarrow$, taking limits as $n \rightarrow \infty$ on both sides of $(\bigcap_{k=n}^{\infty} A_k) \subset (\bigcup_{k=n}^{\infty} A_k)$, it follows directly from (1.1) that

$$(1.2) \quad \liminf_n A_n = \lim_n \bigcap_{k=n}^{\infty} A_k \subset \lim_n \bigcup_{k=n}^{\infty} A_k = \limsup_n A_n .$$

Definition 1.2. If $\liminf_n A_n = \limsup_n A_n$, then we say that events $A_n \rightarrow A$, as $n \rightarrow \infty$.

Remark 1.1. (i) If $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$, then $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n = A$;

(ii) If $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$, then $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n = A$.

Remark 1.2. Since $(\bigcup_{k=n}^{\infty} A_k)$ stands for the event that at least one of the A_k 's, $k \geq n$ does occur, the event $\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ denotes that the event $(\bigcup_{k=n}^{\infty} A_k)$ holds true at each stage n and also thereafter at each $(n+1)$ th, $(n+2)$ th, \dots stage for all $n \geq 1$. In other words, $\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ stands for the event that " A_n 's occur infinitely often".

Probability Space and Probability Function

Definition: To every random phenomenon or experiments E, there corresponds a probability space consisting of the triplet (Ω, \mathbf{Q}, P) with $\Omega = \Omega_{\xi}$, the corresponding sample space, \mathbf{Q} a suitable σ -field of events and P a non-negative set function defined on \mathbf{Q} and satisfying the following axioms:

- (1.3) A1. $P(\Omega) = 1$;
 A2. P is finitely additive; that is, for any finite sequence of disjoint events

$$\{A_i, 1 \leq i \leq n\} \quad P\left(\bigcup_{i=1}^n A_i\right) = P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i);$$

- A3. The probability function P is continuous; that is, for any sequence of events $\{A_n; n \geq 1\}$ if $A_n \rightarrow A$, as $n \rightarrow \infty$, $P(A_n) \rightarrow P(A)$, as $n \rightarrow \infty$.

It immediately follows that: (i) $P(\phi) = 0$ (since by Axiom A2 of finite additivity, $P(\Omega) + P(\phi) = P(\Omega)$ which yields $P(\phi) = 0$); (ii) if $A \subset B$, then $P(A) \leq P(B)$ (since again by Axiom A2, $P(B) = P(BA) + P(BA^c) = P(A) + P(BA^c) \geq P(A)$ clearly); and (iii) that P is sub-

additive: that is, $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$ and $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ (since $\bigcup_{i=1}^n A_i = \bigcup_{k=1}^n B_k = \sum_{k=1}^n B_k$,

where the events $B_1 = A_1$ and $B_k = A_1^c A_2^c \cdots A_{k-1}^c A_k$ for $k \geq 2$ are all disjoint with each B_k a subset of A_k the preceding part (ii), along with Axioms A2 of 'finite additivity' and A3 of 'continuity' imply, respectively, the two sub-additivity assertions (iii) as stated.)

It so happens that in undergraduate courses in Probability or Statistics, Probability Space is commonly

introduced with the following alternative but "equivalent" set of Axioms, with Axioms A2 and A3 replaced

by a single Axiom of "*Countable Additivity*." This, indeed, makes the axiomatic description of probability

space (Ω, \mathbf{Q}, P) a little simpler but at some cost, which we shall be elaborating on below:

Alternative Definition: To every random phenomenon or experiment ε , there corresponds a probability space (Ω, \mathbf{Q}, P) with $\Omega = \Omega_\varepsilon$ and \mathbf{Q} as defined in (1.3) above and P a non-negative set function defined on \mathbf{Q} , now satisfying axioms:

$$(1.4) \quad \text{B1. } P(\Omega) = 1$$

$$\text{B2. } P \text{ is countably additive; that is, for any finite or denumerable sequence } \{A_n; n \geq 1\} \text{ of disjoint events in } \Omega, \quad P(\bigcup_{i \geq 1} A_i) = P(\sum_{i \geq 1} A_i) = \sum_{i \geq 1} P(A_i).$$

As mentioned above, the axiomatic definition (1.4) of probability space (Ω, \mathbf{Q}, P) is a little simpler based on the '*countable additivity*' axiom relative to the one in (1.3), but it is at the cost of relegating the important and crucial '*continuity*' property of the Probability function P to the background, as only an implication of - and playing second fiddle to - the property of '*countable additivity*'. It is our opinion, from a pedagogical standpoint, that students in undergraduate classes do need absolutely to thoroughly grasp the '*continuity*' property' of P . The property has a connection to and facilitates a better understanding of various types of stochastic convergences. It also plays an important implicit role in the study and development of large sample inferential methodology. We shall illustrate this viewpoint with simple examples in Section 4 below. In Section 2 and 3, we give elaborate proofs of the "equivalence" of "*countable additivity*" and various "*continuity*" properties and their connection to the three basic notions of stochastic convergences. Section 5 includes some concluding remarks.

2. Countable Additivity and the Continuity Property

We shall demonstrate below that the '*countable additivity*' Axiom B2 in (1.4) implies the '*continuity*' Axiom A3 in (1.3). Conversely, on the other hand, the continuity Axiom A3 needs the '*finite additivity*' Axiom A2 to imply the '*countable additivity*' Axiom B2 (see Lemma 2.3

below). First, however, in Lemma 2.1 we shall show that, in the presence of 'finite additivity' A2, the 'continuity' property in A3 can be equivalently replaced with either of the "continuity from below" or "continuity from above" properties C1 or C2, respectively, as defined below:

- (2.1) C1. 'Continuity from below': If a sequence of events $A_n \uparrow A$, as $n \rightarrow \infty$ then $P(A_n) \uparrow P(A)$, as $n \rightarrow \infty$.
- C2. 'Continuity from above': If a sequence of events $A_n \downarrow A$, as $n \rightarrow \infty$ then $P(A_n) \downarrow P(A)$, as $n \rightarrow \infty$.

Lemma 2.1. (i) The 'continuity' Axiom A3 implies each of the 'continuity from below' and the 'continuity from above' properties C1 and C2, respectively, in (2.1); and (ii) In the presence of 'finite additivity' Axiom A2, the 'continuity' properties C1 and C2 are equivalent and imply the 'continuity' Axiom A3.

In other words, in the presence of 'finite additivity' Axiom A2, each of the 'continuity' properties C1 from 'below' and C2 from 'above' are - individually, on their own - equivalent to the 'continuity' Axiom A3.

Proof. Part (i) of the lemma follows directly from the 'continuity' Axiom A3 by its definition, since in the 'continuity' properties C1 and C2, the required condition $A_n \rightarrow A$, as $n \rightarrow \infty$, is in there *ab-initio*. In part (ii), the assertion that, in the presence of Axiom A2 of 'finite additivity', the 'continuity' properties C1 and C2 are equivalent follows by their complementary duality: To see that $C1 \Rightarrow C2$, suppose that C1 holds; then if $A_n \downarrow A$, as $n \rightarrow \infty$, which is equivalent to $A_n^c \uparrow A^c$ (by Axiom 2 of 'finite additivity') and, therefore, implies $P(A_n^c) \uparrow P(A^c)$. This last assertion implies $P(A_n) = 1 - P(A_n^c) \downarrow P(A) = 1 - P(A^c)$, as $n \rightarrow \infty$. Thus, $C1 \Rightarrow C2$. Conversely, similar reverse reasoning would imply that $C2 \Rightarrow C1$. This establishes the stated equivalence of 'continuity' properties C1 and C2.

It remains to demonstrate that C1 and C2 ($C1 \Leftrightarrow C2$) imply the 'continuity' Axiom 3: To accomplish this, suppose that C1 and C2 hold, using which we shall first establish the following set of standard inequalities in (2.2) below, from which the deduction $C1, C2 \Rightarrow A3$ would be immediate:

Let $\{A_n: n \geq 1\}$ be a sequence of events, then the following inequalities hold:

$$(2.2) \quad P(\liminf A_n) \leq \liminf P(A_n) \leq \limsup P(A_n) \leq P(\limsup A_n).$$

The middle inequality being obvious, we need to demonstrate the left-most and the right-most ones: Since from (1.1), $\bigcap_{k=n}^{\infty} A_k \uparrow \bigcup_{n=1}^{\infty} (\bigcap_{k=n}^{\infty} A_k) = \liminf A_n$, as $n \rightarrow \infty$, it follows by the 'continuity from below' property C1 that

$$(2.3) \quad P\left(\bigcap_{k=n}^{\infty} A_k\right) \uparrow P(\liminf_n A_n).$$

Also, since $\bigcap_{k=n}^{\infty} A_k \subset A_n$, we have

$$(2.3a) \quad P\left(\bigcap_{k=n}^{\infty} A_k\right) \leq P(A_n) \quad \text{for all } n,$$

so that taking \liminf_n on both sides of (2.3a), in view of (2.3), the left-most inequality

$$(2.4) \quad P(\liminf_n A_n) \leq \liminf_n P(A_n)$$

of (2.2) follows forthwith. To obtain the right-most inequality in (2.2), note from (1.1) that

$$\bigcup_{k=n}^{\infty} A_k \downarrow \bigcap_{n=1}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k\right) = \limsup_n A_n, \text{ as } n \rightarrow \infty, \text{ so that by property C2}$$

$$(2.5) \quad P\left(\bigcup_{k=n}^{\infty} A_k\right) \downarrow P(\limsup_n A_n) \text{ as } n \rightarrow \infty.$$

Again since $A_n \subset \bigcup_{k=n}^{\infty} A_k$, we have

$$(2.5a) \quad P(A_n) \leq P\left(\bigcup_{k=n}^{\infty} A_k\right) \quad \text{for all } n,$$

so that taking \limsup_n on both sides of (2.5a), it follows by the same reasoning as above for the left-most inequality in (2.2) that, in view of (2.5), the stated right-most inequality holds, viz.,

$$(2.6) \quad \limsup_n P(A_n) \leq P(\limsup_n A_n);$$

this establishes the inequalities (2.2). Now we deduce the 'continuity' property A3 from the inequalities (2.2):

Suppose $A_n \rightarrow A$, as $n \rightarrow \infty$; i.e., $\liminf_n A_n = \limsup_n A_n = A$. From inequalities (2.2), it immediately follows that $\lim_n P(A_n) = P(A)$. This establishes the continuity Axiom A3. The proof of part (ii) of the lemma is complete.

Also, as a *concluding comment*, it is worth pointing out that in the above proof, we could have easily deduced the inequality (2.6) from the one in (2.4), which is its complementary dual inequality. To see this note that the inequality (2.4) should hold for the complementary sequence $\{A_n^c : n \geq 1\}$ also, so that we have

$$(2.7) \quad P(\liminf_n A_n^c) \leq \liminf_n P(A_n^c).$$

Now note that whereas the LHS of (2.7) equals

$$(2.7a) \quad P[\liminf_n (\Omega - A_n)] = P(\Omega - \limsup_n A_n) = 1 - P(\limsup_n A_n),$$

the RHS of (2.7) equals

$$(2.7b) \quad \liminf_n [1 - P(A_n)] = 1 - \limsup_n P(A_n).$$

From (2.7), (2.7a) and (2.7b), we thus obtain

$$1 - P(\limsup_n A_n) \leq 1 - \limsup_n P(A_n),$$

or equivalently that

$$\limsup_n P(A_n) \leq P(\limsup_n A_n),$$

which is the inequality (2.6). This completes the comment. \square

We now state a simple lemma.

Lemma 2.2. [3]. If $\{A_n : n \geq 1\}$ is a sequence of disjoint events in \mathbf{Q} , then both

$$(2.8) \quad \liminf_n A_n = \phi \text{ and } \limsup_n A_n = \phi.$$

Proof. Since A_i 's are mutually disjoint, it clearly follows that $\bigcap_{i=n}^{\infty} A_i = \phi$ for all $n \geq 1$.

Therefore,

$$(2.9) \quad \liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i = \bigcup_{n=1}^{\infty} (\phi) = \phi.$$

This establishes the first equality in (2.8). To prove the second, set $B_n = \bigcup_{i=n}^{\infty} A_i$ and note that $\{B_n : n \geq 1\}$ is a decreasing sequence of events. Let $\omega \in B_1$, then $\omega \in A_i$ for some $i = n_0 \geq 1$, which implies that $\omega \notin B_{n_0+j} = \bigcup_{i=n_0+j}^{\infty} A_i$ for all $j \geq 1$, since A_i 's are mutually disjoint (or equivalently that $\omega \notin A_{n_0+j}$ for any $j \geq 1$). So every element of $B_1 \notin B_{n_0+j}$ for all $j \geq 1$, for some $n_0 = 1, 2, \dots$ which implies that

$$(2.10) \quad \limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i = \bigcap_{n=1}^{\infty} B_n = \phi.$$

The proof is complete. \square

We now, as mentioned above, state and prove the equivalence of Axioms A2 and A3 - respectively, of 'finite additivity' and 'continuity' - in (1.3) together to the Axiom B2 of 'countable additivity' in (1.4).

Lemma 2.3. *The 'finite additivity' Axiom A2 and the 'continuity' Axiom A3 in (1.3) together are equivalent to the single 'countable additivity' Axiom B2 in the alternative but the equivalent definition (1.4) of probability space (Ω, \mathbf{Q}, P) .*

Proof. We first prove $B2 \Rightarrow A2$ and $A3$, for which first note that the 'countable additivity' Axiom B2 includes the 'finite additivity' Axiom A2 by its very definition, so that we simply need to show that $B2 \Rightarrow A3$: For this, consider any nondecreasing sequence of events $\{A_n : n \geq 1\}$ such that $A_n \uparrow A$, as $n \rightarrow \infty$, and set $B_1 = A_1$ and $B_k = A_{k-1}^c A_k$ for $k \geq 2$. It is easy

to see that B_k 's are all disjoint and satisfy $\bigcup_{k=1}^n B_k = \bigcup_{k=1}^n A_k$ for all $n \geq 1$ (and so also

$\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k$). Then using the 'countable' additivity Axiom B2, we obtain

$$(2.11) \quad \lim_n P(A_n) = \lim_n P\left(\bigcup_{k=1}^n A_k\right) = \lim_n P\left(\sum_{k=1}^n B_k\right) = \lim_n \sum_{k=1}^n P(B_k) = \sum_{k=1}^{\infty} P(B_k) \\ = P\left(\bigcup_{k=1}^{\infty} B_k\right) = P\left(\bigcup_{k=1}^{\infty} A_k\right) = P(A),$$

so that (2.11) establishes the 'continuity' property C1 which, being equivalent in the presence of 'finite additivity' to the 'continuity' Axiom A3 by Lemma 2.1, establishes A3. The proof of $B2 \Rightarrow A2, A3$ is complete.

To prove the converse that $A2$ and $A3 \Rightarrow B2$, let $\{A_n : n \geq 1\}$ be a denumerable sequence of disjoint events. Then using the 'finite additivity' Axiom A2, we have

$$(2.12) \quad P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\bigcup_{k=1}^n A_k\right) + P\left(\bigcup_{k=n+1}^{\infty} A_k\right) = \sum_{k=1}^n P(A_k) + P(B_{n+1}),$$

where $B_{n+1} = \bigcup_{k=n+1}^{\infty} A_k$, as defined in Lemma 2.2. Since by Lemma 2.2, $P(B_{n+1}) \downarrow 0$, as $n \rightarrow \infty$, taking limits as, $n \rightarrow \infty$, on the RHS of (2.12), we obtain for the disjoint sequence of events

$\{A_n : n \geq 1\}$ that $P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$. This, together with Axiom A2, establishes the

'countable additivity'

Axiom B2. The proof of Lemma 2.3 is complete. \square

3. Stochastic Convergences and the "Continuity" Property

In this section, we shall describe the role and connection of 'continuity' property to the three basic concepts of stochastic convergences.

3.1. Convergence in Distribution and the 'Continuity' Property

The "Continuity" property of Probability P as a set function $P : (\Omega, \mathcal{Q}) \rightarrow [0,1]$ reads as follows: For any sequence of events $\{A_n : n \geq 1\}$ and an event $A \subset \mathcal{Q}$, $\lim_{n \rightarrow \infty} A_n = A$ implies that

$\lim_{n \rightarrow \infty} P(A_n) = P(A)$. The reverse question one may ask: Given that a certain type of sequences

of events $\{A_n : n \geq 1\}$ and their corresponding limiting events $A \subset Q$, belonging to a suitable sub-class of events $Q' \subset Q$, satisfy $P(A_n) \rightarrow P(A)$, as $n \rightarrow \infty$, what can we conclude about the sequences $\{A_n\}$ of events, as $n \rightarrow \infty$? To make this question amenable to a meaningful answer, let $\{X_n : n \geq 1\}$ be a sequence of real r.v. 's and X a real r.v. defined on a probability space (Ω, Q, P) , and set $A_n = [\omega : X_n(\omega) \in B] = [X_n^{-1}(B)]$, and $A = [\omega : X(\omega) \in B] = [X^{-1}(B)]$ for any Borel set $B \in \mathcal{B} = \text{Borel } \sigma\text{-field} \subset \mathbb{R} = \text{real line}$. The induced probability functions $P_n(\cdot), P(\cdot)$, defined respectively by $P_n(B) = P(X_n^{-1}(B)), P(B) = P[X^{-1}(B)]$ on $(\mathbb{R}, \mathcal{B})$, are referred to as their respective probability distributions.

Suppose that, as $n \rightarrow \infty$

$$(3.1) \quad P_n(B) = P[X_n^{-1}(B)] = P(A_n) \rightarrow P(A) = P[X^{-1}(B)] = P(B),$$

for all Borel sets B belonging to a certain sub-class $\mathcal{B}' \subset \mathcal{B}$. Under what sets of conditions on the sub-class \mathcal{B}' for which convergence (2.13) holds, can we conclude any convergence property for the sequence $\{A_n\} = \{X_n^{-1}(B)\}$, or for that matter, for the sequence $\{X_n\}$ of r.v. 's? A partial answer lies in the definition of distributional convergence of the sequence $\{X_n\}$ to X , as $n \rightarrow \infty$:

(3.2) **Definition 3.1.** We say that $\{X_n\}$ converges in distribution to r.v. a X (notationally $X_n \xrightarrow{d} X$), as $n \rightarrow \infty$, if and only if the sub-class $\mathcal{B}' (\subset \mathcal{B})$ of Borel sets B , for which the convergence (3.1) holds, consists of all sets that have P -null boundary $\delta(B)$; that is, the subclass $\mathcal{B}' = \{B : P(\delta(B)) = 0\}$, where $\delta(B) = \bar{B} - B^0$, \bar{B} and B^0 being, respectively, the closure and interior of Borel set B .

The elements of \mathcal{B}' are also referred to as P -continuity sets. In fact, in the preceding Definition 3.1, it is enough to assume that the distribution functions (d.f. 's)

$$(3.3) \quad F_n(x) = P(X_n \leq x) \rightarrow F(x) = P(X \leq x),$$

as $n \rightarrow \infty$, at all continuity points $x \in C(F) = \{x : x \text{ a continuity point of } F\} \subset \{x : -\infty < x < \infty\}$ (see Lemma 3.1 below).

Definition 3.1 of distributional convergence, as we have seen above, provides a partial converse and a connection to the 'continuity' property of probability function P . \square

Now we shall state and prove a very useful and informative 'equivalence' lemma [1]:

Lemma 3.1. We say that $X_n \xrightarrow{d} X$, as $n \rightarrow \infty$, if and only if any of one the following four assertions hold:

- (1) $P_n(B) \rightarrow P(B)$, as $n \rightarrow \infty$, \forall Borel sets $B \in C'(P) = \{B: P(\delta(B))=0\}$ (Definition 3.1);
- (2) $F_n(x) \rightarrow F(x)$, as $n \rightarrow \infty$, $\forall x \in C(P) = \{x: x \text{ a continuity point of } F\}$ (Assertion 3.3);
- (3) For every closed Borel set $H \in \mathcal{B}$, we have

$$\limsup_n P_n(H) \leq P(H);$$
- (4) For every open Borel set $G \in \mathcal{B}$, we have

$$\liminf_n P_n(G) \geq P(G).$$

Proof. First note that the equivalence of assertions (3) and (4) follows immediately through complementation argument: Since if (3) holds and G is an open set, then $\liminf_n P_n(G^c) \leq P(G^c)$, G^c being closed, or equivalently that $1 - \liminf_n P_n(G) \leq 1 - P(G)$ which yields $\liminf_n P_n(G) \geq P(G)$. This establishes (3) \Rightarrow (4). Similarly, the reverse argument would yield (4) \Rightarrow (3). This establishes equivalence of (3) and (4). To see that (3) and/or (4) \Rightarrow (1) is also straight forward: note that if (3) and/or (4) are true, then for any $B \in C'(P) \subset \mathcal{B}$ we have

$$(3.4) \quad P(B^0) \leq \liminf_n P_n(B) \leq \limsup_n P_n(B) \leq P_n(\bar{B});$$

since $P(\delta(B)) = P(\bar{B}) - P(B^0) = 0$, so that $P(\bar{B}) = P(B^0) = P(B)$, it follows at once from (3.4) that $\lim_n P_n(B) = P(B)$ for all P -null Borel sets $B \in C'(P)$. This establishes (3) and/or (4) \Rightarrow (1). To prove their equivalence, we need to prove the convergence (1) \Rightarrow (3) and/or (4); we shall show (1) \Rightarrow (3): Let H be a closed Borel set; then for any sequence of positive constants $\{\eta_k\} \downarrow 0$, as $k \rightarrow \infty$, we can always select a sequence $\{H_k\}$ of closed sets $H_k = \{x: |x - H| \leq \delta_k\}$ with P -null boundaries $\delta(H_k) = 0$ for all $k = 0, 1, \dots$ (This selection is feasible, since there can be at most only a countable number of k 's with $P(\delta(H_k)) > 0$.) Now since $H \subset H_k \forall k$ and they are all P -continuity sets, if (1) holds we have

$$(3.5) \quad \limsup_n P_n(H) \leq \lim_n P(H_k) = P(H_k) \quad \forall k = 1, 2, \dots,$$

so that, in view of $H_k \downarrow H$, as $k \rightarrow \infty$, letting $k \rightarrow \infty$ in (3.5) leads to assertion (3). This proves the equivalence of (3), (4) and (1).

It remains to establish the equivalence of (1) and (2): For this, note that (1) \Rightarrow (2) follows just by their definitions. This is so since $x \in C(F)$ means that $(-\infty, x]$ is a P -continuity set, so that if (1) holds then $F_n(x) \rightarrow F(x)$ as $n \rightarrow \infty$, for all $x \in C(F)$. This establishes (1) \Rightarrow (2). To prove converse that (2) \Rightarrow (1), which is a tough nut to crack, we need to go through two elaborate steps: First let U denote the class of intervals $\{(a, b]: -\infty \leq a < b < \infty$ with P -null boundaries}. Then U is closed under finite intersections and consists of only P -

continuity sets, so that if (2) holds, it at once follows that $P_n(B) \rightarrow P(B)$ for all $B \in \mathcal{U}$. Furthermore, if B_1, B_2, \dots, B_m belong to \mathcal{U} , then by the inclusion-exclusion formula

$$\begin{aligned}
 (3.6) \quad P_n\left(\bigcup_{k=1}^m B_k\right) &= \sum_{i=1}^m P_n(B_i) - \sum_{i<j} P_n(B_i B_j) + \sum_{i<j<k} P_n(B_i B_j B_k) - \dots \\
 &\rightarrow \sum_{i=1}^m P(B_i) - \sum_{i<j} P(B_i B_j) + \sum_{i<j<k} P(B_i B_j B_k) - \dots \\
 &= P\left(\bigcup_{k=1}^m B_k\right).
 \end{aligned}$$

Now if G be any Borel set, then for some countable sequence $\{B_n\}$ of elements of \mathcal{U} , $\bigcup_n B_n$

, so that given $\varepsilon > 0$, we can choose an m such that $P\left(\bigcup_{k=1}^m B_k\right) \geq P(G) - \varepsilon$. In view of this, it at once follows from (3.6) that

$$\begin{aligned}
 (3.7) \quad P(G) - \varepsilon \leq P\left(\bigcup_{k=1}^m B_k\right) &= \lim_n P_n\left(\bigcup_{k=1}^m B_k\right) \\
 &\leq \liminf_n P_n\left(\bigcup_{k=1}^\infty B_k\right) = \liminf_n P_n(G).
 \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, letting $\varepsilon \downarrow 0$ in (3.7) yields the assertion (4). Since (4) is equivalent to (1) as proved above, we have established the converse $(2) \Rightarrow (1)$, so that $(1) \Leftrightarrow (2)$ is proved. The proof complete.

□

3.2. Convergences Almost Sure (a.s.) and in Probability (pr.) vis-à-vis the "Continuity" Property

We now describe how the preceding two convergence notions are based on and connected to the 'continuity' property of P .

Almost Sure Convergence $X_n \xrightarrow{a.s.} X$, as $n \rightarrow \infty$.

Let $\{X_n : n \geq 1\}$ be a sequence of r.v. 's and X a r.v. defined on (Ω, \mathcal{Q}, P) as in sub-section 3.1 above.

(3.8) **Definition 3.2.** We say that the sequence of r.v. 's $\{X_n\}$ converges to the r.v. X almost surely (a.s.) (symbolically $X_n \xrightarrow{a.s.} X$), as $n \rightarrow \infty$, if and only if $P\{\omega : X_n(\omega) \rightarrow X(\omega)\} = 1$;

or equivalently,

(3.8a) $X_n \xrightarrow{a.s.} X$, as $n \rightarrow \infty$, if and only if for each given $\varepsilon > 0$, however small, at all points $\omega \in \Omega$, except possibly at some ω 's constituting a P -null set, $|X_n(\omega) - X(\omega)| < \varepsilon$ for all $n \geq n(\varepsilon, \omega)$ for some sufficiently large $n = n(\varepsilon, \omega)$.

In view of (3.8a), we can re-express (3.8) Definition 3.2 also as:

$$(3.9) \quad X_n \xrightarrow{a.s.} X, \text{ as } n \rightarrow \infty, \text{ if and only if } P(X_n \rightarrow X) = \\ P\left[\bigcap_{\varepsilon>0} \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} \{|X_i - X| < \varepsilon\}\right] = P\left[\bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} \{|X_i - X| < \frac{1}{k}\}\right] = P\left[\bigcap_{k=1}^{\infty} \liminf_n B_n^c\left(\frac{1}{k}\right)\right] = 1$$

where we have set $B_n(\varepsilon) = \{|X_n - X| \geq \varepsilon\}$ in (3.9), the second equality in (3.9) indicating that the event $\{X_n \rightarrow X\}$ is measurable. Indeed, the equation (3.9) can also be equivalently expressed in terms of their complementary counterparts (obtained through complementation) as in (3.9a) below:

$$(3.9a) \quad X_n \xrightarrow{a.s.} X, \text{ as } n \rightarrow \infty, \text{ if and only if } P(X_n \not\rightarrow X) = \\ P\left[\bigcup_{\varepsilon>0} \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} \{|X_i - X| \geq \varepsilon\}\right] = P\left[\bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} \{|X_i - X| \geq \frac{1}{k}\}\right] = P\left[\bigcup_{k=1}^{\infty} \limsup_n B_n\left(\frac{1}{k}\right)\right] = 0.$$

Upon examining (3.9a) (or (3.9)), especially the first and the last equations, and noting that $\left[\bigcup_{i=n}^{\infty} \{|X_n - X| \geq \varepsilon\}\right] \downarrow \left(\left[\bigcap_{i=n}^{\infty} \{|X_n - X| < \varepsilon\}\right] \uparrow\right)$, as $n \rightarrow \infty$, it at once follows that assertion

(3.9a) (or (3.9)), is equivalent to the assertion:

$$(3.10) \quad X_n \xrightarrow{a.s.} X, \text{ as } n \rightarrow \infty, \text{ if and only if, for each } \varepsilon > 0, \\ P[\limsup_n \{|X_n - X| \geq \varepsilon\}] = P[\lim_n \bigcup_{i=n}^{\infty} \{|X_n - X| \geq \varepsilon\}] = 0,$$

or equivalently,

$$(3.10a) \quad X_n \xrightarrow{a.s.} X, \text{ as } n \rightarrow \infty, \text{ if and only if, for each } \varepsilon > 0, \\ P[\liminf_n \{|X_n - X| < \varepsilon\}] = P[\lim_n \bigcap_{i=n}^{\infty} \{|X_n - X| < \varepsilon\}] = 1.$$

Finally, since the sequence $\bigcup_{i=n}^{\infty} \{|X_n - X| \geq \varepsilon\}$ of events in criterion (3.10) decreases to an empty event ϕ ,

as $n \rightarrow \infty$, it follows at once that the assertion $P[\limsup_n \{|X_n - X| \geq \varepsilon\}] = 0$ in (3.10) would hold, if and only if, for each $\varepsilon > 0$

$$(3.11) \quad P\left(\bigcup_{i=n}^{\infty} \{|X_n - X| \geq \varepsilon\}\right) \downarrow 0, \text{ as } n \rightarrow \infty.$$

The criterion (3.10) for $X_n \xrightarrow{a.s.} X$, as $n \rightarrow \infty$, therefore, become equivalent to an easily verifiable one given by (3.11).

Here notice that the criterion (3.11), - and, therefore, the preceding ones in (3.10) or (3.10a) for $X_n \xrightarrow{a.s.} X$, as $n \rightarrow \infty$ - all utilize and are based on the "continuity" properties, including those from 'above' or 'below', in their formulations.

Alternative Representations of a.s. Convergence Criteria:

By definition (3.8) of Almost Sure (a.s.) convergence, $X_n \xrightarrow{a.s.} X$, as $n \rightarrow \infty$, if and only if

(3.12) the set $C = \{\omega : X_n(\omega) \rightarrow X(\omega)\}$ on which $X_n \rightarrow X$, as $n \rightarrow \infty$, has $P(C) = 1$; or equivalently,

(3.12a) the set $C^c = \{\omega : X_n(\omega) \not\rightarrow X(\omega)\}$, namely, the set on which $X_n \not\rightarrow X$, as $n \rightarrow \infty$, has $P(C^c) = 0$

Also, for the set C in (3.12), note from (3.8a) that

(3.13) $\omega \in C$, if and only if, for each $\varepsilon > 0$, there exists an $n(\varepsilon, \omega)$ such that $|X_k(\omega) - X(\omega)| < \varepsilon$ for all sufficiently large $k \geq n(\varepsilon, \omega)$;

or equivalently, that

(3.13a) $\omega \in C^c$, if and only if for each $\varepsilon > 0$, there is for every integer n , no matter how large, at least $k \geq n$ for which $|X_k(\omega) - X(\omega)| \geq \varepsilon$ holds; that is, if and only if, $\omega \in$ some event $A_n(\varepsilon) = \bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}$, $n = 1, 2, 3, \dots$

Criterion 1⁰: $X_n \xrightarrow{a.s.} X$ iff for each $\varepsilon > 0$, $P[\limsup_n \{|X_n - X| \geq \varepsilon\}] = 0$.

Note from (2.20a) that since $A_n(\varepsilon)$ is a non-increasing sequence, as $n \rightarrow \infty$, with

$$(3.14) \quad A_n(\varepsilon) \downarrow \left[\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\} \right] = \left[\limsup_n \{|X_n - X| \geq \varepsilon\} \right],$$

It follows from (3.13a), in view of (3.14), that

$$(3.15) \quad \omega \in C^c \text{ if and only if } \omega \in \left[\limsup_n \{|X_n - X| \geq \varepsilon\} \right].$$

In view of (3.15), therefore, we can rewrite definition (3.12a) as: $X_n \xrightarrow{a.s.} X$ iff the Criterion 1⁰ above holds. □

Criterion 2⁰: $X_n \xrightarrow{a.s.} X$ iff for each $\varepsilon > 0$, $P[\bigcup_n \{|X_n - X| \geq \varepsilon\}] = 0$.

Note from (3.14) that $\left[\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\} \right] = \left[\limsup_n \{|X_n - X| \geq \varepsilon\} \right]$ stands for the event that at each stage $n = 1, 2, \dots$ there is at least one $k \geq n$ for which the event $\{|X_k - X| \geq \varepsilon\}$ does occur. This is equivalent to saying that the last event occurs infinitely often (i.o.). Thus,

the event $[\limsup_n \{|X_n - X| \geq \varepsilon\}]$ is the same as the event $[\{|X_n - X| \geq \varepsilon, \text{ i.o.}\}]$. Hence, we can rewrite (3.12a) as: $X_n \xrightarrow{\text{a.s.}} X$ iff $P[\{|X_n - X| \geq \varepsilon, \text{ i.o.}\}] = 0$. \square

Criterion 3⁰: $X_n \xrightarrow{\text{a.s.}} X$ iff for each $\varepsilon > 0$, $P[\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}] \downarrow 0$, as $n \rightarrow \infty$.

For this, we shall show that the present criterion is equivalent to the one in Criterion 1⁰: To see this, note that on account of convergence (3.14) and the 'continuity' property, we already have

$$(3.16) \quad P[A_n(\varepsilon)] = P[\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}] \downarrow P[\limsup_n \{|X_n - X| \geq \varepsilon\}], \text{ as } n \rightarrow \infty.$$

So if we assume that $P[\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}] \downarrow 0$, as $n \rightarrow \infty$, then clearly the limit in (3.23), namely, $P[\limsup_n \{|X_n - X| \geq \varepsilon\}]$, must also be equal to zero. Conversely, if we assume

$P[\limsup_n \{|X_n - X| \geq \varepsilon\}] = 0$, then $P[\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}]$ must $\downarrow 0$ by (3.16). This establishes the desired equivalence. We can, therefore, write (3.12a) also as: $X_n \xrightarrow{\text{a.s.}} X$ as $n \rightarrow \infty$ iff $P[\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}] \downarrow 0$, as $n \rightarrow \infty$. The present criterion is the easiest to demonstrate among all others in practice.

Direct Proof of Criterion 3⁰.

Note from (3.13) that if $\omega \in C$, then given any $\varepsilon > 0$, there exists an $n(\varepsilon, \omega)$ sufficiently large such that $|X_n(\omega) - X(\omega)| < \varepsilon$ for all $n \geq n(\varepsilon, \omega)$, and so $\omega \notin A_n(\varepsilon)$ for all such sufficiently large n . Thus $[A_n(\varepsilon) \cap C] \downarrow \phi$, as $n \rightarrow \infty$, at all these $\omega \in C$. Similarly, if $\omega \in C^c$, it must belong in the limit, as $n \rightarrow \infty$, to the event $[\limsup_n \{|X_n - X| \geq \varepsilon\}]$; thus $[A_n(\varepsilon) \cap C^c] \downarrow [\limsup_n \{|X_n - X| \geq \varepsilon\} \cap C^c] = C^c$ by (3.15). Hence, as $n \rightarrow \infty$,

$$(3.17) \quad P[A_n(\varepsilon)] = P[A_n(\varepsilon) \cap C] + P[A_n(\varepsilon) \cap C^c] \rightarrow P(\phi) + P(C^c) = 0.$$

By (3.17), therefore, $X_n \xrightarrow{\text{a.s.}} X$ implies that $P[\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}] = P[A_n(\varepsilon)] \downarrow 0$, as $n \rightarrow \infty$, for all $\varepsilon > 0$. Conversely, suppose that $P[\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}] = P[A_n(\varepsilon)] \downarrow 0$, as $n \rightarrow \infty$; or equivalently that as $n \rightarrow \infty$,

$$(3.18) \quad P[\bigcap_{k=n}^{\infty} \{|X_k - X| < \varepsilon\}] = P[A_n^c(\varepsilon)] \rightarrow 1.$$

But by the 'continuity' property (from below), as $n \rightarrow \infty$,

$$(3.19) \quad P\left[\bigcap_{k=n}^{\infty} \{|X_k - X| < \varepsilon\}\right] \uparrow P\left[\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{|X_k - X| < \varepsilon\}\right] \\ = P[\liminf_n \{|X_n - X| < \varepsilon\}].$$

It follows, therefore, from (3.18) and (3.19) that $P[\liminf_n \{|X_n - X| < \varepsilon\}] = 1$, so that one can at once conclude from (3.19) that $X_n \xrightarrow{\text{a.s.}} X$ as $n \rightarrow \infty$. The proof is complete. \square

The Convergence in Probability: $X_n \xrightarrow{p} X$, as $n \rightarrow \infty$,

Definition 3.3. We say $X_n \xrightarrow{p} X$, as $n \rightarrow \infty$, if and only if for each $\varepsilon > 0$,

$$(3.20) \quad P[|X_n - X| \geq \varepsilon] = P[B_n(\varepsilon)] \rightarrow 0 \text{ as } n \rightarrow \infty,$$

where we set $B_n(\varepsilon) = \{|X_n - X| \geq \varepsilon\}$.

By analogy with the "almost sure" convergence discussion above, note that the definition (2.18) again is based on and utilizes directly the continuity Axiom A3, with the set $B_n(\varepsilon) \rightarrow$ the null empty set ϕ , as $n \rightarrow \infty$.

3.3. Relationships between Almost Sure, In-Probability, and Distributional Convergences

In this sub-section, we present two lemmas [4] that elaborate on the relationships among the three basic notions of stochastic convergence of r.v.'s considered above.

Let $\{X_n\}$ be a sequence of real r.v. 's and X a real r.v. defined on $(\Omega, \mathcal{Q}, P) \rightarrow (\mathbb{R}, \mathcal{B})$, as in the preceding subsections 3.1 and 3.2.

Lemma 3.2. (i) If $X_n \xrightarrow{\text{a.s.}} X$, as $n \rightarrow \infty$, then $X_n \xrightarrow{\text{pr}} X$, as $n \rightarrow \infty$; (ii) Conversely, if $X_n \xrightarrow{\text{pr}} X$, as $n \rightarrow \infty$, then there is a subsequence $\{X_{n_k}\} \subset \{X_n\}$ such that as $k \rightarrow \infty$, $n_k \rightarrow \infty$ and $X_{n_k} \xrightarrow{\text{a.s.}} X$.

Proof. Part (i) of the lemma follows at once from the definitions of these convergences and the Criterion 3⁰ (or (3.11)) of subsection 3.2: $X_n \xrightarrow{\text{a.s.}} X$, as $n \rightarrow \infty$, then clearly by virtue of the obvious inequality (3.12) below and Criterion 3⁰, we obtain

$$(3.21) \quad P\{|X_n - X| \geq \varepsilon\} \leq P\left[\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}\right] \rightarrow 0,$$

as $n \rightarrow \infty$; that is, $X_n \xrightarrow{\text{pr}} X$, as $n \rightarrow \infty$. This proves part (i). To prove the converse as in part (ii), suppose $X_n \xrightarrow{\text{pr}} X$, as $n \rightarrow \infty$. Then, by its very definition, for each integer $k \geq 1$ there exists an $n = n(k)$ such that for all $n \geq n(k)$,

$$P\{|X_n - X| \geq \frac{1}{2^k}\} < \frac{1}{2^k}.$$

Now to select an appropriate subsequence $\{X_{n_k}\}$, define $n_1 = n(1)$, $n_2 = \max\{(n_1 + 1), n(2)\}$, \dots , $n_k = \max\{(n_{k-1} + 1), n(k)\}$, $k = 1, 2, \dots$, and set $X_k^* = X_{n_k}$; and note that $A_k = [|X_k^* - X| \geq \frac{1}{2^k}]$, then $P(A_k) \leq \frac{1}{2^k}$ and

$$(3.22) \quad P\left[\bigcup_{k=n}^{\infty} \left\{ |X_k^* - X| \geq \frac{1}{2^k} \right\}\right] \leq \sum_{k=n}^{\infty} P(A_k) \leq \sum_{k=n}^{\infty} \frac{1}{2^k} = \frac{1}{2^{n-1}} \rightarrow 0,$$

as $n \rightarrow \infty$. From (3.22) it follows that $X_n^* \xrightarrow{\text{a.s.}} X$, as $n \rightarrow \infty$; or equivalently that $X_{n_k} = X_k^* \xrightarrow{\text{a.s.}} X$, as $k \rightarrow \infty$. In other words, the sequence $\{X_n\}$ converges almost surely, as $n \rightarrow \infty$, through the subsequence $\{n_k\} \subset n$, as $k \rightarrow \infty$. The proof of part (ii) complete. \square

Lemma 3.3. (i) If $X_n \xrightarrow{\text{pr}} X$, then $X_n \xrightarrow{d} X$, as $n \rightarrow \infty$; and conversely (ii) if $X_n \xrightarrow{d} c$, a constant, then $X_n \xrightarrow{\text{pr}} c$, as $n \rightarrow \infty$.

Proof. We first prove the direct part (i): First note that for any $x' < x$,

$$(3.23) \quad \begin{aligned} P(X \leq x') &= P(X_n \leq x, X \leq x') + P(X_n > x, X \leq x') \\ &\leq P(X_n \leq x) + P(X_n - X > x - x') \leq P(X_n \leq x) + P(|X_n - X| > x - x') \end{aligned}$$

or equivalently, that

$$(3.23a) \quad F(x') \leq F_n(x) + P(|X_n - X| > x - x'),$$

for $x' < x$. Now suppose that $X_n \xrightarrow{\text{pr}} X$, as $n \rightarrow \infty$ then $P(|X_n - X| > x - x') \rightarrow 0$, so that taking \liminf_n on the RHS of (3.23a), we obtain

$$(3.24) \quad F(x') \leq \liminf_n F_n(x) \text{ for all } x' < x.$$

Similarly, interchanging the roles of X_n with X and those of $x' < x$ with $x < x''$ in (3.23) – (3.23a), we obtain $F_n(x) \leq F(x'') + P(|X_n - X| > x'' - x)$. Taking \limsup_n on both sides of preceding inequality, we obtain

$$(3.25) \quad \limsup_n F_n(x) \leq F(x'') \text{ for all } x < x''.$$

From (3.24) and (3.25), we have

$$(3.25a) \quad F(x') \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x'') \text{ for all } x' < x < x''.$$

Now if x is a continuity point of F , i.e., $x \in C(F)$, then letting $x' \uparrow x$ and $x'' \downarrow x$, (3.25a) at once yields

$$(3.26) \quad \lim_n F_n(x) = F(x) \text{ for all } x \in C(F);$$

that is, $X_n \xrightarrow{d} X$, as $n \rightarrow \infty$. This proves part (i). To prove the converse as in part (ii), first note that if $X_n \xrightarrow{d} c$, a constant, then the limiting d.f. F is: $F(x) = 0$ for each $x < c$ and $F(x) = 1$ for $x \geq c$. Therefore, for each $\varepsilon > 0$,

$$(3.27) \quad \begin{aligned} P(|X_n - c| \geq \varepsilon) &\leq P(X_n - c \geq \varepsilon) + P(X_n - c \leq -\varepsilon) \\ &= P(X_n \geq \varepsilon + c) + P(X_n \leq c - \varepsilon) \\ &= 1 - P(X_n < c + \varepsilon) + F_n(c - \varepsilon) \rightarrow 1 - F(c + \varepsilon) + F(c - \varepsilon) = 1 - 1 + 0 = 0, \end{aligned}$$

as $n \rightarrow \infty$, the convergence and the last equality in (3.27) following since $(c - \varepsilon)$ and $(c + \varepsilon)$ are both continuity points of limiting F with $F(c + \varepsilon) = 1$ and $F(c - \varepsilon) = 0$, respectively. Part (ii) of the lemma is complete. \square

4. Examples and Applications

In early university undergraduate probability courses, the continuity property of probability functions is generally not introduced and discussed explicitly. If this concept is grasped correctly and thoroughly in early courses, it would greatly facilitate a better grasp of more complex probability concepts later.

4.1. Application to distribution functions

Examples 1 [6]. Consider a real-valued random variable X with a distribution function $F_X(x)$. It is known

that $F_X(x)$ has the following properties:

(DF1) If $x_1 < x_2$, then $F_X(x_1) \leq F_X(x_2)$, i.e., $F_X(x)$ is a nondecreasing function

(DF2) $\lim_{n \rightarrow \infty} F_X(x) = 1$; $\lim_{n \rightarrow -\infty} F_X(x) = 0$

(DF3) $\lim_{h \rightarrow 0^+} F_X(x+h) = F_X(x)$, i.e., $F_X(x)$ is right continuous

(DF4) $\lim_{h \rightarrow 0^+} F_X(x-h) = F_X(x) - P(X = x)$.

(DF5) If F_X is also left continuous, then $P(X = x) = 0$; i.e., $x \in C(F)$.

For proving these properties, as stated above, the continuity property is usually not invoked in junior-level classes. Here we shall employ this property to prove (DF2) to (DF4); (DF1) and (DF5) are evident.

To show (DF2), let $\{a_n\}$ be an increasing sequence of real numbers such that $a_n \uparrow \infty$, as

$n \rightarrow \infty$, and set $A_n = \{X \leq a_n\}$. Clearly, $A_n \uparrow$ with $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n = \{X < \infty\}$. It follows that

$$(4.1) \quad \lim_{x \rightarrow \infty} F_X(x) = \lim_{n \rightarrow \infty} F_X(a_n) = \lim_{n \rightarrow \infty} P(X \leq a_n) = \lim_{n \rightarrow \infty} P(A_n)$$

$$= P(\lim_{n \rightarrow \infty} A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(X < \infty) = 1.$$

The equation (4.1) establishes the first part of DF2; the second part follows similarly.

To show (DF3), let $\{a_n\}$ be a decreasing sequence of real numbers so that $a_n \downarrow 0$, as $n \rightarrow \infty$

and set $A_n = \{X \leq x + a_n\}$. Clearly, $A_n \downarrow$ with $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n = \{X \leq x\}$. It follows that

$$(4.2) \quad \begin{aligned} \lim_{h \rightarrow 0^+} F_X(x+h) &= \lim_{n \rightarrow \infty} F_X(x+a_n) = \lim_{n \rightarrow \infty} P(X \leq x+a_n) = \lim_{n \rightarrow \infty} P(A_n) \\ &= P(\lim_{n \rightarrow \infty} A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(X \leq x) = F_X(x), \end{aligned}$$

which establishes (DF3).

To show (DF4), let $a_n \downarrow 0$, $B_n = \{X \leq x - a_n\}$ and $C_n = \{x - a_n < X \leq x\}$. Thus, the sequence

of events $\{B_n\}$ is increasing with $\lim_{n \rightarrow \infty} B_n = \bigcup_{n=1}^{\infty} B_n = \{X < x\}$, and the sequence of events $\{C_n\}$

is decreasing with

$\lim_{n \rightarrow \infty} C_n = \bigcap_{n=1}^{\infty} C_n = \{X = x\}$. Now

$$(4.3) \quad \begin{aligned} F_X(x) &= P(X \leq x) = P(X \leq x - a_n) + P(x - a_n < X \leq x) \\ &= P(B_n) + P(C_n), \end{aligned}$$

so that taking limits on both sides (4.3), as $n \rightarrow \infty$, we have

$$(4.4) \quad \begin{aligned} F_X(x) &= \lim_{n \rightarrow \infty} P(B_n) + \lim_{n \rightarrow \infty} P(C_n) = P(\lim_{n \rightarrow \infty} B_n) + P(\lim_{n \rightarrow \infty} C_n) \\ &= P(X < x) + P(X = x) = F_X(x-) + P(X = x). \end{aligned}$$

Rearranging terms in (4.4), we have $F_X(x-) = F_X(x) - P(X = x)$. This establishes (DF4).

The proof of (DF5) is obvious. \square

Example 2 [7]. Let X be a r.v. such that $\text{Var}(X) = 0$. Then X is a constant equal to $E(X)$ with Probability one.

To prove this, we will use the continuity property of P . By Chebyshev's inequality, for any $n \geq 1$,

$$(4.5) \quad P\{|X - E(X)| > \frac{1}{n}\} \leq n^2 V(X) = 0.$$

Let $B_n = \{|X - E(X)| > \frac{1}{n}\}$. Then the sequence of events $\{B_n\}$ is increasing \uparrow with

$\lim_{n \rightarrow \infty} B_n = \bigcup_{n=1}^{\infty} B_n = \{X \neq E(X)\}$ and, by (4.5), $P(B_n) = 0 \quad \forall \quad n = 1, 2, \dots$. Thus,

$$(4.6) \quad P(X \neq E(X)) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = P(\lim_{n \rightarrow \infty} B_n) = \lim_{n \rightarrow \infty} P(B_n) = 0.$$

The equation (4.6) implies that $P(X = E(X)) = 1$. The proof is complete. \square

Example 3 [7]. If $F_X(x)$ is continuous, then $P(X = x) = 0 \quad \forall \quad x$. To prove this, we shall again use the continuity property of Probability P :

Let $A_n = \{x - \frac{1}{n} < X \leq x\}$. Clearly, $\{A_n\}$ is a decreasing sequence of events with $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n = \{X = x\}$. Now

$$\begin{aligned}
 (4.7) \quad P(X = x) &= P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n) \\
 &= \lim_{n \rightarrow \infty} P\left(x - \frac{1}{n} < X \leq x\right) = \lim_{n \rightarrow \infty} [F_X(x) - F_X(x - \frac{1}{n})] \\
 &= F_X(x) - F_X(x-) = 0, \text{ since } F_X \text{ is continuous at } x.
 \end{aligned}$$

The proof is complete. \square

4.2. Applications to almost sure (a.s.) convergence

By the equation (3.11), we say that $X_n \xrightarrow{a.s.} X$ iff $\lim_{n \rightarrow \infty} P\left[\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}\right] = 0$. In the following examples, we illustrate *a.s.* convergence with the help of continuity property of Probability P .

Example 4. Let X_1, X_2, \dots, X_n be i.i.d., distributed as (\square) X with density $f_{\theta}(x) = \frac{1}{\theta}$, $0 \leq x \leq \theta$. Suppose $X_{(n)} = \max(X_1, X_2, \dots, X_n)$. We shall show that $X_{(n)} \xrightarrow{a.s.} \theta$, as $n \rightarrow \infty$:

Let $\varepsilon > 0$ and set $A_n = \{\omega : |X_{(n)}(\omega) - \theta| \geq \varepsilon\}$. Note that $X_{(n)} \leq X_{(n+1)} \leq \theta$, $|X_{(n+1)} - \theta| \leq |X_{(n)} - \theta|$ so that $A_{n+1} \subset A_n$ for all n . It follows that

$$\begin{aligned}
 P(\limsup_n A_n) &= \lim_{n \rightarrow \infty} P\left[\bigcup_{k=n}^{\infty} A_k\right] = \lim_{n \rightarrow \infty} P\left[\bigcup_{k=n}^{\infty} \{|X_{(k)} - \theta| \geq \varepsilon\}\right] \\
 &= \lim_{n \rightarrow \infty} P\{|X_{(n)} - \theta| \geq \varepsilon\} = \lim_{n \rightarrow \infty} P(X_{(n)} - \theta \leq -\varepsilon), \text{ (since } \{X_{(n)} - \theta > \varepsilon\} = \phi) \\
 &= \lim_{n \rightarrow \infty} P(X_{(n)} \leq \theta - \varepsilon) = \lim_{n \rightarrow \infty} P(X_1 \leq \theta - \varepsilon, X_2 \leq \theta - \varepsilon, \dots, X_n \leq \theta - \varepsilon) \\
 &= \lim_{n \rightarrow \infty} [P(X_1 \leq \theta - \varepsilon)]^n,
 \end{aligned}$$

since X_i 's are i.i.d. From above, therefore, we have

$$\begin{aligned}
 (4.8) \quad \lim_{n \rightarrow \infty} P\left[\bigcup_{k=n}^{\infty} \{|X_{(k)} - \theta| \geq \varepsilon\}\right] &= \lim_{n \rightarrow \infty} \left(\int_0^{\theta - \varepsilon} \frac{1}{\theta} dx\right)^n \\
 &= \lim_{n \rightarrow \infty} \left(1 - \frac{\varepsilon}{\theta}\right)^n = 0 \text{ as } 0 < \varepsilon < \theta.
 \end{aligned}$$

The result $X_{(n)} \xrightarrow{a.s.} \theta$, as $n \rightarrow \infty$, follows by criterion 3⁰ (or criterion 1⁰). \square

Example 5. Let X_1, X_2, \dots, X_n be i.i.d. \square as X , which has the density $f_\theta(x) = e^{-(x-\theta)}$, $x \geq \theta$. Suppose we set $X_{(1):n} = \min(X_1, X_2, \dots, X_n)$. We shall show that $X_{(1):n} \xrightarrow{a.s.} \theta$, as $n \rightarrow \infty$:

Let $\varepsilon > 0$ and set $A_n = \{\omega : |X_{(1):n}(\omega) - \theta| \geq \varepsilon\}$. Note that $X_{(1):n} \geq X_{(1):n+1} \geq \theta$ and $|X_{(1):n+1} - \theta| \leq |X_{(1):n} - \theta|$, so that $A_{n+1} \subset A_n$ for all n . Now,

$$\begin{aligned} P(\limsup_n A_n) &= \lim_{n \rightarrow \infty} P\left[\bigcup_{k=n}^{\infty} A_k\right] = \lim_{n \rightarrow \infty} P\left[\bigcup_{k=n}^{\infty} \{|X_{(1):k} - \theta| \geq \varepsilon\}\right] \\ &= \lim_{n \rightarrow \infty} P\{|X_{(1):n} - \theta| \geq \varepsilon\} = \lim_{n \rightarrow \infty} P(X_{(1):n} - \theta \geq \varepsilon), \text{ (since } \{X_{(1):n} - \theta \leq -\varepsilon\} = \phi \text{)} \\ &= \lim_{n \rightarrow \infty} P(X_{(1):n} \geq \theta + \varepsilon) = \lim_{n \rightarrow \infty} P(X_1 \geq \theta + \varepsilon, X_2 \geq \theta + \varepsilon, \dots, X_n \geq \theta + \varepsilon) \\ &= \lim_{n \rightarrow \infty} [P(X_1 \geq \theta + \varepsilon)]^n, \end{aligned}$$

since X_i 's are i.i.d. Thus, from above, we have

$$(4.9) \quad \begin{aligned} \lim_{n \rightarrow \infty} P\left[\bigcup_{k=n}^{\infty} \{|X_{(1):k} - \theta| \geq \varepsilon\}\right] &= \lim_{n \rightarrow \infty} \left(\int_{\theta+\varepsilon}^{\infty} e^{-(x-\theta)} dx\right)^n = \lim_{n \rightarrow \infty} \left(\int_{\varepsilon}^{\infty} e^{-x} dx\right)^n \\ &= \lim_{n \rightarrow \infty} (e^{-n\varepsilon}) = 0, \text{ as } \varepsilon > 0. \end{aligned}$$

The result $X_{(1):n} \xrightarrow{a.s.} \theta$, as $n \rightarrow \infty$, again follows by the criterion 3⁰ (or criterion 1⁰).

\square

Example 6 [7]. Let $\{X_i : i \geq 1\}$ be a sequence of i.i.d. Bernoulli r.v.'s with $E(X_i) = p$.

Suppose $S_n = \sum_{i=1}^n X_i$. We shall show that $\frac{S_n}{n} = \bar{X}_n \xrightarrow{a.s.} E(X_i) = p$, as $n \rightarrow \infty$. First, note that $S_n \square$ Binomial (n, p) . For the Binomial distribution, it can be verified that

$$(4.10) \quad P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq 2e^{-(n\varepsilon^2/4)}.$$

(The proof of (4.10) is given in the Appendix.) Let $A_n = \left\{\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right\}$. Then,

$$(4.11) \quad \begin{aligned} P(\limsup_n A_n) &= \lim_{n \rightarrow \infty} P\left[\bigcup_{m=n}^{\infty} A_m\right] = \lim_{n \rightarrow \infty} P\left[\bigcup_{m=n}^{\infty} \left\{\left|\frac{S_m}{m} - p\right| \geq \varepsilon\right\}\right] \\ &\leq \lim_{n \rightarrow \infty} \sum_{m=n}^{\infty} P\left(\left|\frac{S_m}{m} - p\right| \geq \varepsilon\right) \end{aligned}$$

by Bool's inequality (sub-additivity), so that from (4.11) and inequality (4.10) above, we have

$$(4.11a) \quad \begin{aligned} P(\limsup_n \left\{\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right\}) &= \lim_{n \rightarrow \infty} P\left[\bigcup_{m=n}^{\infty} \left\{\left|\frac{S_m}{m} - p\right| \geq \varepsilon\right\}\right] \\ &\leq 2 \lim_{n \rightarrow \infty} \sum_{m=n}^{\infty} e^{-m\varepsilon^2/4} \end{aligned}$$

$$= 2(1 - e^{-\varepsilon^2/4})^{-1} \lim_{n \rightarrow \infty} (e^{-n\varepsilon^2/4}) = 0.$$

From (4.11a), in view of Criterion 1⁰, it follows that $[\frac{S_n}{n} - p] \xrightarrow{a.s.} 0$, or equivalently, $\frac{S_n}{n} \xrightarrow{a.s.} p$, as $n \rightarrow \infty$. Roughly speaking, this says that not only does the Probability of finding $\frac{S_n}{n}$ far from p vanishes but that the Probability that any of the subsequent (S_n/n) 's are far from p also vanishes, as $n \rightarrow \infty$. Results of this type are called strong laws of large numbers.

4.3. Applications to Miscellaneous Problems

In this section, we give applications of the 'continuity' property in two other interesting contexts, different from the above.

Example 7 (Eventual extinction) [5]. Consider a population whose members are capable of producing offsprings of the same kind. Let X_0 denote the size of the initial generation and regard its offsprings as constituting the first generation. Denote the number in the first generation by X_1 and, correspondingly, by X_n the number in the n th generation. Assume further that (i) $0 < P(X_1 = 0) < 1$; and also that (ii) $0 < P(X_{n+1} = 0 | X_n > 0) < 1$ holds for all n .

Now since the event $\{X_n = 0\}$ implies (obviously) the event $\{X_{n+1} = 0\}$, it follows precisely on account of assumptions (i) and (ii) that the strict inequality $P(X_n = 0) < P(X_{n+1} = 0)$ holds for all n . To see this clearly, note in view of the obvious equation $P(X_{n+1} = 0 | X_n = 0) = 1$ that

$$\begin{aligned} (4.12) \quad P(X_{n+1} = 0) &= P(X_{n+1} = 0, X_n = 0) + P(X_{n+1} = 0, X_n > 0) \\ &= P(X_n = 0)P(X_{n+1} = 0 | X_n = 0) + \\ &P(X_n > 0)P(X_{n+1} = 0 | X_n > 0) \\ &= P(X_n = 0) + P(X_n > 0)P(X_{n+1} = 0 | X_n > 0) > P(X_n = 0), \end{aligned}$$

the last inequality in (4.12) following since both $P(X_{n+1} = 0 | X_n > 0)$ and $P(X_n > 0)$ are strictly positive: The first probability $P(X_{n+1} = 0 | X_n > 0)$ is positive for all n by assumption (ii); and the second $P(X_n > 0)$, while positive for $n = 1$ by assumption (i), remains positive for $n = 2, 3, \dots$ by successive holding of (4.12) inductively for each $n = 1, 2, 3, \dots$ (i.e., by induction on n using (4.12)).

Thus, $P(X_n = 0)$ is strictly increasing \uparrow , as $n \rightarrow \infty$, so that $\lim_{n \rightarrow \infty} P(X_n = 0)$ does exist. What does it mean? To answer this question, we use the continuity property of the probability function P as follows:

$$(4.12a) \quad \lim_{n \rightarrow \infty} P(X_n = 0) = P[\lim_{n \rightarrow \infty} \{X_n = 0\}] \\ = P\left(\bigcup_{n=1}^{\infty} \{X_n = 0\}\right) = P(\text{population eventually dies out}) > 0;$$

That is, limiting probability of $\{X_n = 0\}$, namely, the probability of eventual extinction, is positive, which conclusion agrees with the common belief that if the offspring number hits zero in any particular generation, the population would eventually become extinct. \square

Example 8 (Murphy's Law) [7]. Murphy's law says that "anything that can go wrong, will go wrong". We will illustrate this via a coin-tossing problem. Suppose a fair coin is tossed repeatedly. Let s denote any fixed sequence of heads and tails of length r . Thus, among 2^r equally likely possibilities, one of them is s . With Probability one, the sequence s will eventually occur in r consecutive tosses of the coin.

Consider a fair die with 2^r faces, where each face corresponds to one of the 2^r outcomes of tossing the coin r times, and one of them is face s . Now roll the die repeatedly.

Let A_k be the event that the face s appears for the first time on the k th roll. There are 2^{rk} distinct outcomes of k rolls, and they are equally likely as it is a fair die. The number of favorable cases constituting the event A_k are $(2^r - 1)^{k-1} \cdot 1 = (2^r - 1)^{k-1}$, with 2^{rk} as the total number of cases in k rolls. Thus, by the classical definition of Probability,

$$(4.13) \quad P(A_k) = \frac{(2^r - 1)^{k-1}}{2^{rk}}.$$

Since $A_k \cap A_j = \phi$ for $k \neq j$, we have by Axiom (A3) that

$$(4.14) \quad P\left(\bigcup_{k=1}^m A_k\right) = \sum_{k=1}^m P(A_k) = \sum_{k=1}^m \frac{(2^r - 1)^{k-1}}{2^{rk}} = \frac{1}{2^r} \sum_{k=1}^m \left(1 - \frac{1}{2^r}\right)^{k-1} \\ = \frac{1}{2^r} \left[\frac{1 - \left(1 - \frac{1}{2^r}\right)^m}{1 - \left(1 - \frac{1}{2^r}\right)} \right] = 1 - \left(1 - \frac{1}{2^r}\right)^m,$$

which is the Probability that face s appears at all m rolls of the die.

Now consider n tosses of the coin, and let $m = \lfloor \frac{n}{r} \rfloor$ (where $\lfloor x \rfloor$ is the integer part of x). Thus, the n tosses can be divided into m sequences of length r with a remainder $n - mr$. Let B_n be the event that none of these m sequences is s , and let C_n be the event that the sequence s does not occur anywhere in the n tosses. Then

$$(4.15) \quad C_n \subseteq B_n = \left(\bigcup_{k=1}^m A_k\right)^c,$$

because rolling the die m times and tossing the coin mr times yield the same sample space of equally likely outcomes. By the monotonicity and complementation rule of Probability, (4.14) and (4.15) yield

$$(4.16) \quad P(C_n) \leq P(B_n) = \left(1 - \frac{1}{2^r}\right)^m \rightarrow 0 \text{ as } n \rightarrow \infty, \text{ or equivalently, as } m \rightarrow \infty \text{ (since } mr \leq n < (m+1)r \text{).}$$

Now the event that s eventually occurs is $\lim_{n \rightarrow \infty} P(C_n^c)$. Thus, by the 'continuity' property of Probability, complementation rule and (4.16), we obtain

$$(4.17) \quad P(\lim_{n \rightarrow \infty} C_n^c) = \lim_{n \rightarrow \infty} P(C_n^c) = 1 - \lim_{n \rightarrow \infty} P(C_n) = 1.$$

This proves the assertion. \square

5. Concluding Remarks

In Section 2, we have described in detail the sense in which the 'continuity' Axiom A3 in definition (1.3) of Probability Space (Ω, \mathcal{Q}, P) is equivalent to that of 'countable additivity' Axiom B2 in its alternative but the equivalent definition (1.4).

In Section 3, we have discussed its connection to and the role it plays in defining with nuance the three basic notions of stochastic convergence of random quantities (that is, of random variables, vectors, or functions): Convergence Almost Sure, Convergence in Probability, and Convergence in Distribution. This fundamental 'Continuity' property is frequently not discussed with adequate emphasis in early university courses in Probability and Statistics. From a pedagogical standpoint, however, it appears vital that it be done so at an early introductory stage. In our opinion, this would be of considerable help in laying down sound initial foundations to start with and later for a firmer grasp of advanced probabilistic concepts. Since the main objective of this paper was to simply focus and highlight where the 'Continuity' property comes into play explicitly, we limited ourselves to the discussion only the above three basic notions of stochastic convergences. In other notions of convergences of random quantities, for example, convergence of moments, convergence in L_p or other types of metric convergences or of appropriate transforms etc., the 'Continuity' property plays only an implicit role if at all.

The paper should be of value to students and teachers of introductory university courses in Probability or Statistics and may serve as a useful teaching reference.

References

- [1] Billingsley, P., *Convergence of Probability Measures*, John Wiley & Sons Inc., New York, 1968. <https://doi.org/10.1002/9780470316962>
- [2] Durrett, R., *Elementary Probability for Applications*, Cambridge University Press, NY, 2009. <https://doi.org/10.1017/cbo9780511841521>
- [3] Heathcote, C.R., *Probability: Elements of the Mathematical Theory*, Dover Publications, NY, 2000.
- [4] Loeve, M., *Probability Theory I*, 4th Edition, Springer-Verlag, NY, 1977. <https://doi.org/10.1007/978-1-4757-6288-4>

[5] Ross, S., *Stochastic Processes*, Wiley, NY, 1983.

[6] Sevrerini, A., *Elements of Distribution Theory*, Cambridge University Press, NY, 2005.

[7] Stirzaker, D., *Elementary Probability*, 2nd Edition, Cambridge University Press, NY, 2007. <https://doi.org/10.1017/cbo9780511755309>

Appendix

(A.1). Definition [2]: Suppose $\{A_n : n \geq 1\}$ is a sequence of events defined on a probability space (Ω, \mathcal{Q}, P) .

$$(i) \text{ If } A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots, \text{ then } \lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$$

$$(ii) \text{ If } A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots, \text{ then } \lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$$

(A.2). DeMorgan's Law [5]:

$$(i) \left(\bigcup_{i=1}^k A_i \right)^c = \bigcap_{i=1}^k A_i^c; \quad \left(\bigcap_{i=1}^{\infty} A_i \right)^c = \bigcup_{i=1}^{\infty} A_i^c$$

$$(ii) \left(\bigcap_{i=1}^k A_i \right)^c = \bigcup_{i=1}^k A_i^c; \quad \left(\bigcup_{i=1}^{\infty} A_i \right)^c = \bigcap_{i=1}^{\infty} A_i^c.$$

(A.3). A Useful Identity on $\bigcup_{i=1}^{\infty} A_i$ [5]:

Let $B_1 = A_1$ and $B_i = A_1^c A_2^c \dots A_{i-1}^c A_i$ for $i \geq 2$, then B_i 's are disjoint and $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i = \sum_{i=1}^{\infty} B_i$.

(A.4). Proof of the statement (4.10):

Let $\{X_i : i \geq 1\}$ be a sequence of i.i.d. Bernoulli r.v.'s with $E(X_i) = p$, $q = 1 - p$. Set

$S_n = \sum_{i=1}^n X_i$. Then, $S_n \sqcup$ Binomial (n, p) and $E(S_n) = np$. Therefore, the mgf of S_n , the binomial mgf, is

$$(A.5) \quad M_{S_n}(t) = E(e^{tS_n}) = \prod_{i=1}^n E(e^{tX_i}) = (q + pe^t)^n.$$

Below, we will show that

$$(A.6) \quad P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leq 2e^{-n\varepsilon^2/4}.$$

By Chernoff's bound, note that for each $\varepsilon > 0$,

$$(A.7) \quad P(S_n - np > n\varepsilon) = P[S_n > n(p + \varepsilon)]$$

$$\leq \frac{E(e^{tS_n})}{e^{n(p+\varepsilon)t}} = \frac{M_{S_n}(t)}{e^{n(p+\varepsilon)t}}, \forall t > 0.$$

Thus, the following also holds true (since (A.7) holds for $\forall t > 0$),

$$(A.8) \quad P(S_n - np > n\varepsilon) \leq \min_{t>0} \frac{M_{S_n}(t)}{e^{n(p+\varepsilon)t}}.$$

Below, we simplify the RHS of (A.7) as follows: By (A.5), the RHS of (A.7) can be written as,

$$(A.9) \quad \begin{aligned} \min_{t>0} \frac{M_{S_n}(t)}{e^{n(p+\varepsilon)t}} &= \min_{t>0} \frac{(q + pe^t)^n}{e^{n(p+\varepsilon)t}} = \min_{t>0} (qe^{-pt} + pe^t e^{-pt})^n e^{-n\varepsilon t} \\ &= \min_{t>0} (qe^{-pt} + pe^{qt})^n e^{-n\varepsilon t} \\ &\leq \min_{t>0} (qe^{p^2 t^2} + pe^{q^2 t^2})^n e^{-n\varepsilon t} \\ &\leq \min_{t>0} (qe^{t^2} + pe^{t^2})^n e^{-n\varepsilon t} = \min_{t>0} [e^{nt^2 - n\varepsilon t}] = e^{-n\varepsilon^2/4}, \end{aligned}$$

the first inequality in (A.9) above holding since $e^x \leq x + e^{x^2}$ and the last equality since the minimum occurs at $t = \frac{\varepsilon}{2}$.

By (A.8) and (A.9), we conclude that

$$(A.10) \quad P(S_n - np > n\varepsilon) \leq e^{-n\varepsilon^2/4}.$$

Likewise, we can also show that

$$(A.11) \quad P(S_n - np < -n\varepsilon) \leq e^{-n\varepsilon^2/4}.$$

Finally, by combining (A.10) and (A.11), we get

$$P(|\frac{S_n}{n} - p| > \varepsilon) = P(S_n - np > n\varepsilon) + P(S_n - np < -n\varepsilon) \leq 2 e^{-n\varepsilon^2/4}.$$

The proof is complete. \square

Received: June 29, 2021; Published: July 21, 2021