Journal of Mathematics and Statistics

Original Research Paper

Formalizing Probability Concepts in a Type Theory

Farida Kachapova

Department of Mathematical Sciences, Auckland University of Technology, New Zealand

Article history
Received: 22-09-2018
Revised: 01-11-2018
Accepted: 10-11-2018

Email: farida.kachapova@aut.ac.nz

Abstract: In this paper we formalize some fundamental concepts of probability theory such as the axiomatic definition of probability space, random variables and their characteristics, in the Calculus of Inductive Constructions, which is a variant of type theory and the foundation for the proof assistant COQ. In a type theory every term and proposition should have a type, so in our formalizations we assign an appropriate type to each object in order to create a framework where further development of formalized probability theory will be possible. Our formalizations are based on mathematical results developed in the COQ standard library; we use mainly the parts with logic and formalized real analysis. In the future we aim to create COQ coding for our formalizations of probability concepts and theorems. In this paper the definitions and some proofs are presented as flag-style derivations while other proofs are more informal.

Keywords: Type Theory, Kolmogorov's Axiomatics, Probability Theory, Calculus of Inductive Constructions, Flag-Style Derivation

1. Introduction

Formalizing mathematics in an axiomatic theory makes mathematical results and applications more reliable and provides assurance of their validity. In recent years several parts of mathematics were formalized in type theories. One of the books on types theories is (Nederpelt and Geuvers, 2014); it describes formal theories from λ-calculus to the calculus of constructions and contains examples of formalizing mathematics in such theories. Type theories underlie proof assistants, so formalizing mathematics in a type theory enables the use of computers for checking the correctness of mathematical proofs. There exist several proof assistants. Here we consider the proof assistant COQ (named after its creator Thierry Coquand) and the formal type theory Calculus of Inductive Constructions (COIC), which is the foundation of COQ. COQ and COIC are described in the book (Bertot and Casteran, 2014) and the COQ website (Coq Development Team). The website contains, among other things, a reference manual and the COQ standard library (which we will further call just the Library). The Library contains formalizations of a significant part of mathematics in COQ, including logic, natural numbers, integers, real numbers, limits, series, Riemann integral, algebraic structures etc. Probability theory has not been developed in COIC except some fragments such as finite probability distributions in (Moreira, 2012). In this paper we formalize the general axiomatic approach to probability developed by Kolmogorov (1950); a more recent presentation of this approach can be found, for example, in (Rosenthal, 2006).

Formalizing mathematical topics in COQ starts with formalizing them in the underlying theory COIC; then computer coding follows. In this paper we formalize in COIC some fundamental probability concepts; coding the results in COQ is a task for the future. The developments in this paper are based on the concepts and theorems that are already formalized in COIC/COQ and presented in the Library. Logical connectives and the constant *False* are defined in COIC/COQ; we denote $\bot$ = *False*. In particular, $\neg P$ denotes $(P \to \bot)$ and we have:

$$False\_ind : \forall P : Prop, \bot \to P. \quad (1)$$

One of the main ideas in type theories is that every object has a type. In this paper we endeavour to assign appropriate types to fundamental probability concepts in order to create a framework where further development of formalized probability theory will be possible.

In Section 2 we axiomatically define a probability space, in Section 3 we give formal definitions of

random variables and related concepts. In Section 4 we formalize concepts related to discrete random variables, including their numerical characteristics. In Section 5 we study random vectors and their numerical characteristics and in Section 6 we look at continuous random variables. All this is done within the theory COIC. For better readability our presentation of results is not strictly formal. In definitions and some proofs we use the flag-style derivation described in (Nederpelt and Geuvers, 2014), while other proofs are more informal. Often we use implicit variables as is done in COIC/COQ and in (Nederpelt and Geuvers, 2014).

2. Probability Space

2.1. Some Facts about Sets

We fix a non-empty universal set U that denotes the sample space (the set of all outcomes):

$$\boxed{\textbf{var } U : Type}$$

The variable ω will be used for objects of type U (outcomes).

The Library introduces subsets of U as objects of type $U \to Prop$, that is predicates on U.

$$\begin{array}{l}\boxed{\textbf{var } U : Type}\\ \quad \text{Definition } P\_set(U) : Type := U \to Prop.\\ \quad \boxed{\textbf{var } A : P\_set(U)}\\ \quad\quad \boxed{\textbf{var } \omega : U}\\ \quad\quad\quad \text{Notation: } \omega \in A \text{ for } A\omega.\\ \quad\quad \text{Definition } A^c : P\_set(U) := \lambda\omega : U.\neg A\omega.\\ \quad \text{Definition } \varnothing : P\_set(U) := \lambda\omega : U.\bot.\\ \quad \text{Definition } \Omega : P\_set(U) := \lambda\omega : U.T.\end{array}$$

Here T denotes $\bot \to \bot$. This flag-style diagram defines the power set $P\_set(U)$ of U, a complement A^c of a set A, the empty set $\varnothing$ and the full set Ω. In a type theory every object should have a type. So when we introduce a variable or a constant we indicate its type immediately after its name (the type is separated from the name by a colon).

The Library has definitions of set operations such as *Union*(B C), *Intersection*(B C) and *Setminus*(B C), which we denote $B \cup C$, $B \cap C$ and $B \setminus C$, respectively.

As usual, $B \subseteq C$ means $\forall \omega$: U, $(\omega \in B \to \omega \in C)$, that is $\forall \omega$: U, $(B\omega \to C\omega)$. The Extensionality axiom from COIC/COQ can be written as:

$$ext : \forall B\ C : P\_set(U), B \subseteq C \wedge C \subseteq B \to B = C. \qquad (2)$$

As in the book (Nederpelt and Geuvers, 2014) we will use the notation $\{x: S \mid Ax\}$ for λx: $S.Ax$, where A is a predicate.

Lemma 2.1

1) $\Omega^c = \varnothing$.
2) $\forall B$: $P\_set(U)$, $(B^c)^c = B$.

Proof

1) In the first part of the following flag-style derivation we use the rule (1) for False elimination and in the second part the Extensionality axiom (2).

$$\begin{array}{l}\boxed{\textbf{var } U : Type}\\ \quad \boxed{\textbf{var } \omega : U}\\ \quad\quad \boxed{\textbf{var } u : \omega \in \Omega^c}\\ \quad\quad\quad u : \Omega^c\omega.\\ \quad\quad\quad u : \neg\Omega\omega.\\ \quad\quad\quad u : T \to \bot.\\ \quad\quad\quad a_1 := \lambda y : \bot.y : \bot \to \bot.\\ \quad\quad\quad a_1 : T.\\ \quad\quad\quad a_2 := ua_1 : \bot.\\ \quad\quad\quad a_2 : \varnothing\omega.\\ \quad\quad\quad a_2 : (\omega \in \varnothing).\\ \quad\quad a_3 := \lambda u : (\omega \in \Omega^c).a_2 : \omega \in \Omega^c \to \omega \in \varnothing.\\ \quad a_4 := \lambda\omega : U.a_3 : \forall\omega : U, \omega \in \Omega^c \to \omega \in \varnothing.\\ \quad a_4 : \Omega^c \subseteq \varnothing.\\ \quad \boxed{\textbf{var } \omega : U}\\ \quad\quad \boxed{\textbf{var } v : \omega \in \varnothing}\\ \quad\quad\quad v : \varnothing\omega.\\ \quad\quad\quad v : \bot.\\ \quad\quad\quad a_5 := False\_ind\,(\omega \in \Omega^c)\,v : \omega \in \Omega^c.\\ \quad\quad a_6 := \lambda v : (\omega \in \varnothing).a_5 : \omega \in \varnothing \to \omega \in \Omega^c.\\ \quad a_7 := \lambda\omega : U.a_6 : \forall\omega : U, \omega \in \varnothing \to \omega \in \Omega^c.\\ \quad a_7 : \varnothing \subseteq \Omega^c.\\ \quad a_8 := conj\,(\Omega^c \subseteq \varnothing)(\varnothing \subseteq \Omega^c)\,a_4a_7 : (\Omega^c \subseteq \varnothing) \wedge (\varnothing \subseteq \Omega^c).\\ \quad a_9 := ext(\Omega^c)\,\varnothing\,a_8 : \Omega^c = \varnothing.\end{array}$$

In the second to last step we used the rule for conjunction introduction.

2) Clearly, this statement does not hold in a constructive theory. So in the second part of the following derivation we use a classical logic.

var $U : Type$

var $B : P\_set(U)$

var $\omega : U$

var $u : \omega \in B$

var $v : \omega \in B^c$

$v : \neg(\omega \in B).$

$v : \omega \in B \to \bot.$

$vu : \bot.$

$a_1 := \lambda v : \omega \in B^c.vu \ : \ \omega \in B^c \to \bot.$

$a_1 : \neg(\omega \in B^c).$

$a_1 : \omega \in (B^c)^c.$

$a_2 := \lambda u : \omega \in B.a_1 \ : \ \omega \in B \to \omega \in (B^c)^c.$

$a_3 := \lambda \omega : U.a_2 \ : \ B \subseteq (B^c)^c.$

var $\omega : U$

var $u : \omega \in (B^c)^c$

$u : \neg\neg(\omega \in B).$

$a_4 := dne(\omega \in B)u : \omega \in B.$

$a_5 := \lambda u : \omega \in (B^c)^c.a_4 \ : \ \omega \in (B^c)^c \to \omega \in B.$

$a_6 := \lambda \omega : U.a_5 \ : \ (B^c)^c \subseteq B.$

$a_7 := conj\,((B^c)^c \subseteq B)(B \subseteq (B^c)^c)\,a_6 a_3 \ :$

$((B^c)^c \subseteq B) \wedge (B \subseteq (B^c)^c).$

$a_8 := ext((B^c)^c)\,B\,a_7 \ : \ (B^c)^c = B.$

In the line with a_4 we are using the classical axiom for elimination of double negation, *dne*: $\forall A$: *Prop*, $\neg\neg A \to A$. In the last line we are using the Extensionality axiom (2). ■

Lemma 2.2

1) $\varnothing^c = \Omega$.
2) $\forall B$: *P_set*(*U*), $B \cup B^c = \Omega$.
3) $\forall B$: *P_set*(*U*), $B \cap B^c = \varnothing$.
4) $\forall B\ C$: *P_set*(*U*), $(B \cup C)^c = B^c \cap C^c$.
5) $\forall B\ C$: *P_set*(*U*), $(B \cap C)^c = B^c \cup C^c$.

Proof

Similarly to the previous lemma, the proofs are derived by using definitions and lemmas from the Ensembles sub-library of the Library. ■

Next we define a countable intersection $\bigcap_{n\in\mathbb{N}} An$ and a countable union $\bigcup_{n\in\mathbb{N}} An$:

var $U : Type$

var $A : nat \to P\_set(U)$

Definition $Inf\_inters(A) : P\_set(U) \ :=$

$(\lambda\omega : U.\forall n : nat, \omega \in An).$

Notation: $\bigcap_{n\in\mathbb{N}} An$ for $Inf\_inters(A)$.

Definition $Inf\_union(A) : P\_set(U) \ :=$

$(\lambda\omega : U.\neg(\forall n : nat, \omega \in (An)^c)).$

Notation: $\bigcup_{n\in\mathbb{N}} An$ for $Inf\_union(A)$.

Thus, $\bigcup_{n\in\mathbb{N}} An$ is defined as $\left(\bigcap_{n\in\mathbb{N}} (An)^c\right)^c$, which means we are using a classical logic.

Lemma 2.3

1) $\left(\bigcap_{n\in\mathbb{N}} An\right)^c = \bigcup_{n\in\mathbb{N}} \left(\lambda n : nat.(An)^c\right)$.
2) $\left(\bigcup_{n\in\mathbb{N}} An\right)^c = \bigcap_{n\in\mathbb{N}} \left(\lambda n : nat.(An)^c\right)$.

Proof

Follows from the definitions. ■

We will use the expression "*An* are disjoint" as an abbreviation for $(\forall i\ j$: *nat*, $i \neq j \to Ai \cap Aj = \varnothing)$.

2.1. Definition of Probability Space

In the Kolmogorov's axiomatics a probability space is defined as a triple (*U*, *F*, *P*), where *U* is a sample space, *F* is a sigma-field of subsets of *U* and *P* is a probability measure on (*U*, *F*). So first we define a sigma-field *F* as a collection of subsets of *U* satisfying three given conditions:

var $U : Type$

var $F : P\_set(U) \to Prop$

Notation: $A \in F$ for FA.

Definition $sigma\_1 : Prop \ := \varnothing \in F.$

Definition $sigma\_2 : Prop \ :=$

$\forall B : P\_set(U), B \in F \to B^c \in F.$

Definition $sigma\_3 : Prop \ :=$

$\forall A : nat \to P\_set(U), (\forall n : nat, An \in F)$

$\to \bigcap_{n\in\mathbb{N}} An \in F.$

Definition $Sigma\_field(F) : Prop \ :=$

$sigma\_1 \wedge sigma\_2 \wedge sigma\_3.$

Theorem 2.4. Properties of a Sigma-Field.

Suppose F is a sigma-field on U. Then the following hold.

1) $\varnothing \in F$.
2) $\forall B\ C$: *P_set*(U), $B \in F \wedge C \in F \to B \cap C \in F$.
3) $\forall B\ C$: *P_set*(U), $B \in F \wedge C \in F \to B \setminus C \in F$.
4) $\forall A$: *nat* $\to$ *P_set*(U), $(\forall n$: *nat*, $An \in F) \to \bigcup_{n \in \mathbb{N}} An \in F$.
5) $\forall B\ C$: *P_set*(U), $B \in F \wedge C \in F \to B \cup C \in F$.

Proof

1) Follows from the *sigma*_1 and *sigma*_2 conditions and Lemma 2.1.1).

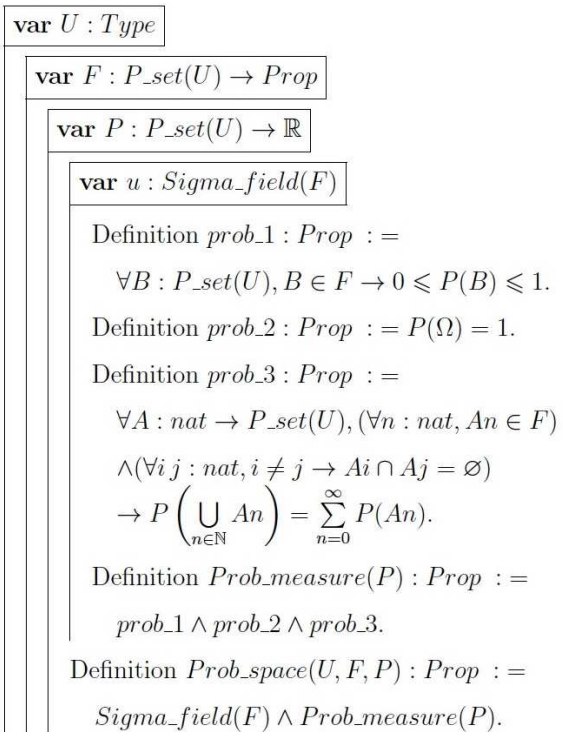
2) Suppose $B \in F \wedge C \in F$. Denote $A0 = B$ and $An = C$ for $n \geq 1$. Then $(\forall n$: *nat*, $An \in F)$ and by the *sigma*_3 condition $\bigcap_{n \in \mathbb{N}} An \in F$, that is $B \cap C \in F$.

3) Follows from the *sigma*_2 condition and part 2), since $B \setminus C = B \cap C^c$.

4) This is proven using the *sigma*_2 and *sigma*_3 conditions, since $\bigcup_{n \in \mathbb{N}} An = \left(\bigcap_{n \in \mathbb{N}} (An)^c \right)^c$.

5) Suppose $B \in F \wedge C \in F$. Then $B^c \in F$ and $C^c \in F$. By part 2) and Lemma 2.2.4), $(B \cup C)^c \in F$, so $((B \cup C)^c)^c \in F$ and by Lemma 2.1.2), $B \cup C \in F$. ■

In the following diagram we define a probability measure on (U, F) and a probability space (U, F, P) following the Kolmogorov's axiomatic definition.



Theorem 2.5. Properties of a Probability Measure.

Suppose (U, F, P) is a probability space. Then the following hold:

1) $P(\varnothing) = 1$.
2) $\forall B\ C$: *P_set*(U); $(B \in F \wedge C \in F \wedge B \cap C = \varnothing)$ $\to P(B \cup C) = P(B) + P(C)$.
3) $\forall B_1, \ldots, B_n$: *P_set*(U), $\left(\bigwedge_{i=1}^{n} B_i \in F \wedge \bigwedge_{i=1}^{n} \bigwedge_{j=1, j \neq i}^{n} (B_i \cap B_j = \varnothing) \right)$ $\to P(B_1 \cup \ldots \cup B_n) = P(B_1) + \ldots + P(B_n)$.
4) $\forall B$: *P_set*(U), $B \in F \to P(B^c) = 1 - P(B)$.
5) $\forall B\ C$: *P_set*(U), $B \in F \wedge C \in F$ $\to P(B \cup C) = P(B) + P(C) - P(B \cap C)$.
6) $\forall A$: *nat* $\to$ *P_set*(U), $(\forall n$: *nat*, $An \in F) \wedge$ $(\forall n$: *nat*, $A(n+1) \subseteq An) \to P\left(\bigcap_{n \in \mathbb{N}} An \right) = \lim_{N \to \infty} P(AN)$.

Proof

1) Denote $A0 = \Omega$ and $An = \varnothing$ for $n \geq 1$. Then $(\forall n$: *nat*, $An \in F)$ and An are disjoint. We have:

$$\Omega = \bigcup_{n \in \mathbb{N}} An$$

and by the *prob*_1 and *prob*_3 conditions:

$$1 = P(\Omega) = \sum_{n=0}^{\infty} P(An) = P(\Omega) + \sum_{n=1}^{\infty} P(\varnothing), \text{ so}$$

$$1 = 1 + \sum_{n=1}^{\infty} P(\varnothing).$$

Since $P(\varnothing) \geq 0$, we get $P(\varnothing) = 0$.

2) Denote $A0 = B$, $A1 = C$ and $An = \varnothing$ for $n > 1$. Then An are disjoint and $(\forall n$: *nat*, $An \in F)$. We have:

$$B \cup C = \bigcup_{n \in \mathbb{N}} An,$$

so by the *prob*_3 condition:

$$P(B \cup C) = \sum_{n=0}^{\infty} P(An) = P(B) + P(C) + 0 = P(B) + P(C),$$

since $P(\varnothing) = 0$ by part 1).

3) This is proven by an external induction on n using part 2).

4) By Lemma 2.2.2), 3), $B \cup B^c = \Omega$ and $B \cap B^c = \varnothing$. By part 2),

$$P(B) + P(B^c) = P(\Omega) = 1, \text{ so } P(B^c) = 1 - P(B).$$

■■

5) $C = (B \cap C) \cup (C \setminus B)$ and $(B \cap C) \cap (C \setminus B) = \varnothing$. By part 2), $P(C) = P(B \cap C) + P(C \setminus B)$ and

$$P(C \setminus B) = P(C) - P(B \cap C). \quad (3)$$

$B \cup C = B \cup (C \setminus B)$ and $B \cap (C \setminus B) = \varnothing$. So by part 2) and (3),

$$P(B \cup C) = P(B) + P(C \setminus B) = P(B) + P(C) - P(B \cap C).$$

6) Denote $B = \lambda n$: $nat.(An)^c$ and $C = \lambda n$: $nat.(Bn \setminus B(n-1))$ with $C0 = B0$. Then Cn are disjoint and

$$\bigcup_{n\in\mathbb{N}} Cn = \bigcup_{n\in\mathbb{N}} Bn = \bigcup_{n\in\mathbb{N}} (An)^c = \left(\bigcap_{n\in\mathbb{N}} An\right)^c,$$

$$\bigcup_{n\in\mathbb{N}} Cn = \left(\bigcap_{n\in\mathbb{N}} An\right)^c. \quad (4)$$

Similarly,

$$\bigcup_{n=0}^{N} Cn = \left(\bigcap_{n=0}^{N} An\right)^c = (AN)^c, \quad (5)$$

since each $A(n+1) \subseteq An$.

By (4), (5), parts 3), 4), and the definition of probability measure,

$$1 - P\left(\bigcap_{n\in\mathbb{N}} An\right) = P\left(\left(\bigcap_{n\in\mathbb{N}} An\right)^c\right) = P\left(\bigcup_{n\in\mathbb{N}} Cn\right) = \sum_{n=0}^{\infty} P(Cn)$$
$$= \lim_{N\to\infty} \sum_{n=0}^{N} P(Cn) = \lim_{N\to\infty} P\left(\bigcup_{n=0}^{N} Cn\right) = \lim_{N\to\infty} P\left((AN)^c\right)$$
$$= 1 - \lim_{N\to\infty} P(AN).$$

So

$$1 - P\left(\bigcap_{n\in\mathbb{N}} An\right) = 1 - \lim_{N\to\infty} P(AN) \text{ and } P\left(\bigcap_{n\in\mathbb{N}} An\right) = \lim_{N\to\infty} P(AN).$$

■

The following diagram defines independence of two events in a probability space.

var $U : Type \mid F : P\_set(U) \to Prop \mid P : P\_set(U) \to \mathbb{R}$

var $u : Prob\_space(U, F, P)$

var $BC : P\_set(U) \mid v : (B \in F \wedge C \in F)$

Definition $Indep\_events(B, C) : Prop :=$

$P(B \cap C) = P(B) \cdot P(C).$

3. Random Variables

As in the rigorous probability theory, we define a random variable X in the following diagram as a real-valued function of outcomes such that each set of the form $\{\omega : U \mid X\omega \le x\}$ is an event (an element of the sigma-field F).

var $U : Type$

var $F : P\_set(U) \to Prop \mid u : Sigma\_field(F)$

var $X : U \to \mathbb{R}$

Definition $Random\_var(X) : Prop :=$

$\forall x : \mathbb{R}, \{\omega : U \mid X\omega \leqslant x\} \in F.$

As was mentioned in subsection 2.1, $\{\omega : U \mid X\omega \le x\}$ is the notation for the predicate $(\lambda\omega : U.\ X\omega \le x)$ of type $U \to Prop$.

Theorem 3.1. Properties of a Random Variable.

Suppose F is a sigma-field on U, X is a random variable and $x \in \mathbb{R}$. Then the following hold:

1) $\{\omega : U \mid X\omega > x\} \in F$.
2) $\{\omega : U \mid X\omega \ge x\} \in F$.
3) $\{\omega : U \mid X\omega < x\} \in F$.
4) $\{\omega : U \mid X\omega = x\} \in F$.

Proof

1) This is proven by applying the *sigma_2* condition, since

$$\{\omega : U \mid X\omega > x\} = \{\omega : U \mid X\omega \le x\}^c.$$

2) Follows from the *sigma_3* condition and part 1), since

$$\{\omega : U \mid X\omega \ge x\} = \bigcap_{n\in\mathbb{N}} \left\{\omega : U \mid X\omega > x - \frac{1}{n+1}\right\}.$$

3) Follows from the *sigma_2* condition and part 2), since

$$\{\omega : U \mid X\omega < x\} = \{\omega : U \mid X\omega \ge x\}^c.$$

4) Follows from part 2) and Theorem 2.4.2) because

$$\{\omega : U \mid X\omega = x\} = \{\omega : U \mid X\omega \le x\} \cap \{\omega : U \mid X\omega \ge x\}.$$ ■

In the following diagram we formally define the (*cumulative*) *distribution function* F_X of a random variable X.

■■

var $U : Type \mid F : P\_set(U) \to Prop \mid P : P\_set(U) \to \mathbb{R}$

var $u : Prob\_space(U, F, P)$

var $X : U \to \mathbb{R} \mid v : Random\_var(X)$

Definition $F_X : \mathbb{R} \to \mathbb{R} := \lambda x : \mathbb{R}.P(\{\omega : U \mid X\omega \leqslant x\})$.

Theorem 3.2. Properties of a Distribution Function.

Suppose (U, F, P) is a probability space and X is *a* random variable. Then the following hold.

1) $\forall x: \mathbb{R}$, $0 \le F_X(x) \le 1$.
2) $\forall x\ y: \mathbb{R}$, $x < y \to F_X(x) \le F_X(y)$, that is the function F_X is non-decreasing.
3) $\lim_{x \to -\infty} F_X(x) = 0$.
4) $\lim_{x \to \infty} F_X(x) = 1$.
5) $\forall x$: $\mathbb{R}$, $\lim_{t \to x+} F_X(t) = F_X(x)$, that is the function F_X is right-continuous.
6) $\forall x: \mathbb{R}$, $P(\{\omega : U \mid X\omega < x\}) = \lim_{t \to x-} F_X(t)$.
7) $\forall x\ y$: $\mathbb{R}$, $x < y \to P(\{\omega : U \mid x < X\omega \le y\}) = F_X(y) - F_X(x)$.
8) $\forall x$: $\mathbb{R}$, $P(\{\omega : U \mid X\omega = x\}) = F_X(x) - \lim_{t \to x-} F_X(t)$.
9) $\forall x$: $\mathbb{R}$, $P(\{\omega : U \mid X\omega > x\}) = 1 - F_X(x)$.

Proof

The theorem is proven using the definitions, Theorem 3.1 and the theory of real numbers developed in the Library. ■

4. Discrete Random Variables

4.1. Definition of a Discrete Random Variable

A random variable X is discrete if its set of values is finite or countable. This is formally expressed in the diagram below:

var $U : Type \mid F : P\_set(U) \to Prop \mid P : P\_set(U) \to \mathbb{R}$

var $u : Prob\_space(U, F, P)$

var $X : U \to \mathbb{R}$

var $g : U \to nat$

var $f : nat \to \mathbb{R}$

Definition $Discr\_var(X, f, g) : Prop :=$

$Random\_var(X) \wedge (\forall \omega : U, X\omega = f(g\omega))$

$\wedge(\forall i\ j : nat, fi = fj \to i = j)$.

Definition $m_X : \mathbb{R} \to \mathbb{R} := \lambda x : \mathbb{R}.P(\{\omega : U \mid X\omega = x\})$.

This definition says: a random variable X is discrete if there exist a function g: $U \to nat$ and an injective function f: $nat \to \mathbb{R}$ such that $X = f \circ g$. In other words, the set of values of X is $\{fi \mid i \in image(g)\}$, which is a finite set or a countable set, depending on the cardinality of the subset $image(g)$ of $\mathbb{N}$. The function m_X is called the *mass function* of the discrete random variable X. Clearly, if x is not a value of X, then

$$\{\omega : U \mid X\omega = x\} = \varnothing \text{ and } m_X(x) = 0.$$

Theorem 4.1. Properties of a Discrete Random Variable.

Suppose *Discr_var* (X, f, g) as in the previous definition. Then the following hold.

1) $\forall x$: $\mathbb{R}$, $0 \le m_X(x) \le 1$.
2) $\sum_{n=0}^{\infty} m_X(fn) = \sum_{n=0}^{\infty} P(\{\omega : U \mid X\omega = fn\}) = 1$.
3) $\forall x$: $\mathbb{R}$, $F_X(x) = \sum_{n=0}^{\infty} P(\{\omega : U \mid X\omega = fn \wedge fn \le x\})$.

Proof

1) Follows from Theorem 3.1.4) and the *prob*_1 condition, since

$$m_X(x) = P(\{\omega : U \mid X\omega = x\}).$$

2) Denote $A = \lambda n$: $nat.\{\omega : U \mid X\omega = fn\}$. Since f is injective, An are disjoint. For any $\omega : U$, $X\omega = fn$ for $n = g\omega$, so we have $\Omega = \bigcup_{n \in \mathbb{N}} An$ and by the *prob*_3 condition:

$$1 = P(\Omega) = \sum_{n=0}^{\infty} P(An) = \sum_{n=0}^{\infty} P(\{\omega : U \mid X\omega = fn\}) = \sum_{n=0}^{\infty} m_X(fn).$$

3) Denote $B = \{\omega : U \mid X\omega \le x\}$. Clearly, $\Omega = \bigcup_{n \in \mathbb{N}} \{\omega : U \mid X\omega = fn\}$ and

$$B = B \cap \Omega = \bigcup_{n \in \mathbb{N}} (B \cap \{\omega : U \mid X\omega = fn\}) = \bigcup_{n \in \mathbb{N}} Cn,$$

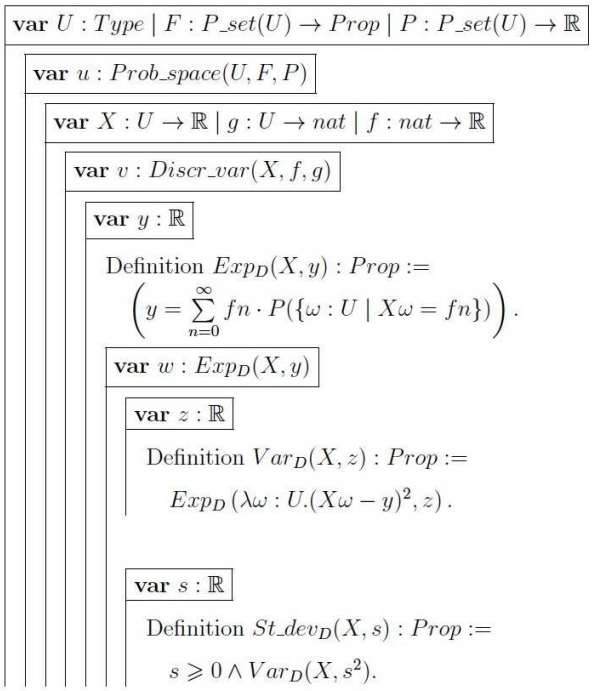
where $Cn = \{\omega : U \mid X\omega = fn \wedge fn \le x\}$. Since Cn are disjoint,

$$F_X(x) = P(B) = \sum_{n=0}^{\infty} P(Cn) = \sum_{n=0}^{\infty} P(\{\omega : U \mid X\omega = fn \wedge fn \le x\}).$$

■

4.2. Numerical Characteristics of Discrete Random Variables

In the following diagram we define the expectation, variance and standard deviation of a discrete random variable X.



The notations $Exp_D(X, y)$, $Var_D(X, z)$ and $St\_dev_D(X, s)$ mean: y is the expectation, z is the variance and s is the standard deviation of the discrete random variable X, respectively.

Theorem 4.2. Expectation Property of a Discrete Variable.

Suppose (U, F, P) is a probability space and $Discr\_var(X, f, g)$. Then

$$\forall h : \mathbb{R} \to \mathbb{R}, [Discr\_var(h \circ X, h \circ f, g) \to$$
$$\forall z : \mathbb{R}, (Exp_D(h \circ X, z) \to$$
$$z = \sum_{n=0}^{\infty} h(fn) \cdot P(\{\omega : U \mid X\omega = fn\}))].$$

Proof

Follows from the definitions. ■

Theorem 4.3. Properties of Expectation.

Suppose (U, F, P) is a probability space and X is a discrete random variable. Then the following hold.

1) $\forall y\ c$: $\mathbb{R}$, $Exp_D(X, y) \to Exp_D(\lambda\omega : U.(c \cdot X\omega), c \cdot y)$.
In ordinary mathematical notations: $E(cX) = cE(X)$.

2) $\forall c$: $\mathbb{R}$, $Exp_D(\lambda\omega : U.c, c)$.
In ordinary mathematical notations: $E(c) = c$.

3) $\forall x\ y$: $\mathbb{R}$, $Exp_D(X, x) \wedge Exp_D(Y, y) \to Exp_D(X + Y, x + y)$.
In ordinary mathematical notations: $E(X+Y) = E(X) + E(Y)$.

4) $\forall y$: $\mathbb{R}$, $Exp_D(X, y) \wedge (\forall \omega : U, X\omega \geq 0) \to y \geq 0$.
In ordinary mathematical notations: $X \geq 0 \to E(X) \geq 0$.

5) $\forall x\ y$: $\mathbb{R}$, $Exp_D(X, x) \wedge Exp_D(Y, y) \wedge (\forall \omega : U, X\omega \leq Y\omega) \to x \leq y$.
In ordinary mathematical notations: $X \leq Y \to E(X) \leq E(Y)$.

Proof

Follows from the definitions and properties of sums and series of real numbers. ■

Theorem 4.4. Properties of Variance.

Suppose (U, F, P) is a probability space and X is a discrete random variable. Then the following hold.

1) $\forall y$: $\mathbb{R}$, $Var_D(X, y) \to y \geq 0$.
In ordinary mathematical notations: $Var(X) \geq 0$.

2) $\forall c$: $\mathbb{R}$, $Var_D(\lambda\omega : U.c, 0)$.
In ordinary mathematical notations: $Var(c) = 0$.

3) $\forall y\ c$: $\mathbb{R}$, $Var_D(X, y) \to Var_D(\lambda\omega : U.(X\omega + c), y)$.
In ordinary mathematical notations: $Var(X + c) = Var(X)$.

4) $\forall y\ c$: $\mathbb{R}$, $Var_D(X, y) \to Var_D(\lambda\omega : U.(c \cdot X\omega), c^2 \cdot y)$.
In ordinary mathematical notations: $Var(cX) = c^2 Var(X)$.

5) $\forall y\ z$: $\mathbb{R}$, $Exp_D(X, y) \wedge Exp_D(\lambda\omega : U.(X\omega)^2, z) \to Var_D(X, z - y^2)$.

In ordinary mathematical notations: $Var(X) = E(X^2) - (E(X)^2)$.

Proof

Follows from the definition of variance and Theorem 4.3. ■

Theorem 4.5. Properties of Standard Deviation.

Suppose (U, F, P) is a probability space and X is a discrete random variable. Then the following hold.

1) $\forall s$: $\mathbb{R}$, $St\_dev_D(X, s) \to s \geq 0$.
In ordinary mathematical notations: $\sigma(X) \geq 0$.

2) $\forall c$: $\mathbb{R}$, $St\_dev_D(\lambda\omega : U.c, 0)$.
In ordinary mathematical notations: $\sigma(c) = 0$.

3) $\forall s\ c$: $\mathbb{R}$, $St\_dev_D(X, s) \to St\_dev_D(\lambda\omega : U.(X\omega + c), s)$.

In ordinary mathematical notations: $\sigma(X + c) = \sigma(X)$.

4) $\forall s\ c$: $\mathbb{R}$, $St\_dev_D(X, s) \to St\_dev_D(\lambda\omega : U.(c \cdot X\omega), |c| \cdot s)$.

In ordinary mathematical notations: $\sigma(cX) = |c|\sigma(X)$.

■■

Proof

Follows from the definition and previous theorem. ■

Theorem 4.6. Markov Inequality.

Suppose (U, F, P) is a probability space and X is a discrete random variable. Then

$$\forall y\ c:\mathbb{R},[c>0\wedge Exp_D(X,y)\wedge(\forall\omega:U,X\omega\geq 0)$$
$$\rightarrow P(\{\omega:U\mid X\omega\geq c\})\leq\frac{y}{c}].$$

Proof

Assume all the hypotheses. Then

$$y=\sum_{n=0}^{\infty} fn\cdot P(\{\omega:U\mid X\omega=fn\}),\ so$$
$$y=\sum_{\substack{n=0\\ fn\geq c}}^{\infty} fn\cdot P(\{\omega:U\mid X\omega=fn\})$$
$$+\sum_{\substack{n=0\\ 0\leq fn<c}}^{\infty} fn\cdot P(\{\omega:U\mid X\omega=fn\}) \quad (6)$$
$$+\sum_{\substack{n=0\\ fn<0}}^{\infty} fn\cdot P(\{\omega:U\mid X\omega=fn\}).$$

We have $(\forall\omega:U,X\omega\geq 0)$, so for any $fn<0$, $\{\omega:U\mid X\omega=fn\}=\varnothing$ and all probabilities in the last sum in (6) equal 0. Clearly,

$$\sum_{\substack{n=0\\ 0\leq fn<c}}^{\infty} fn\cdot P(\{\omega:U\mid X\omega=fn\})\geq 0.$$

Since f is injective by the definition of a discrete random variable, the events $\{\omega:U\mid X\omega=fn\}$ are disjoint and as in Theorem 4.1.3), we have for the first sum in (6):

$$\sum_{\substack{n=0\\ fn\geq c}}^{\infty} fn\cdot P(\{\omega:U\mid X\omega=fn\})\geq c\sum_{\substack{n=0\\ fn\geq c}}^{\infty} P(\{\omega:U\mid X\omega=fn\})$$
$$=c\sum_{n=0}^{\infty} P(\{\omega:U\mid X\omega=fn\wedge fn\geq c\})=cP(\{\omega:U\mid X\omega\geq c\}).$$

Therefore $y\geq cP(\{\omega:U\mid X\omega\geq c\})$ and $P(\{\omega:U\mid X\omega\geq c\})\leq\frac{y}{c}$. ■

Theorem 4.7. Chebychev Inequality.

Suppose (U, F, P) is a probability space and X is a discrete random variable. Then:

$$\forall y\ z\ a:\mathbb{R},[a>0\wedge Exp_D(X,y)\wedge Var_D(X,z)$$
$$\rightarrow P(\{\omega:U\mid\mid X\omega-y\mid\geq a\})\leq\frac{z}{a^2}].$$

Proof

Assume all the hypotheses.

Denote $Y=\lambda\omega:U.(X\omega-y)^2$. Then Y is also a discrete random variable and by the definition of variance, we have Exp_D (Y, z). Applying the Markov inequality (Theorem 4.6) to Y and a^2 we get:

$$P(\{\omega:U\mid Y\omega\geq a^2\})\leq\frac{z}{a^2}. \quad (7)$$

The inequality $Y\omega\geq a^2$ is equivalent to $(X\omega-y)^2\geq a^2$ and to $|X\omega-y|\geq a$, so

$$P(\{\omega:U\mid Y\omega\geq a^2\})=P\{\omega:U\mid\mid X\omega-y\mid\geq a\}$$

and from (7):

$$P(\{\omega:U\mid\mid X\omega-y\mid\geq a\})\leq\frac{z}{a^2}.\ ■$$

5. Random Vectors

5.1. Joint Distribution

In the following diagram we define the joint distribution function F_{XY} of two random variables X and Y; in the same diagram we define the independence of X and Y.

var $U:Type\mid F:P\_set(U)\rightarrow Prop\mid P:P\_set(U)\rightarrow\mathbb{R}$
  var $u:Prob\_space(U,F,P)$
    var $X:U\rightarrow\mathbb{R}\mid v:Random\_var(X)$
      var $Y:U\rightarrow\mathbb{R}\mid w:Random\_var(Y)$
        Definition $F_{XY}:\mathbb{R}\rightarrow\mathbb{R}\rightarrow\mathbb{R}:=$
          $\lambda x\ y:\mathbb{R}.P(\{\omega:U\mid X\omega\leqslant x\wedge Y\omega\leqslant y\})$.
        Definition $Indep\_var(X,Y):Prop:=$
          $\forall x\ y:\mathbb{R},Indep\_events(\{\omega:U\mid X\omega\leqslant x\},\{\omega:U\mid Y\omega\leqslant y\})$.

The joint distribution function of n random variables $X_1, X_2, ..., X_n$ can be defined similarly.

Theorem 5.1. Properties of a Joint Distribution Function.

Suppose (U, F, P) is a probability space, X and Y are random variables. Then the following hold.

1) $\forall x\ y$: $\mathbb{R}$, $0\leq F_{XY}(x,y)\leq 1$.
2) $\forall y$: $\mathbb{R}$, $\lim_{x\rightarrow-\infty} F_{XY}(x,y)=0$.

■■

3) $\forall x\colon \mathbb{R}, \lim_{y\to-\infty} F_{XY}(x, y) = 0.$

4) $\forall y\colon \mathbb{R}, \lim_{x\to\infty} F_{XY}(x, y) = F_Y(y).$

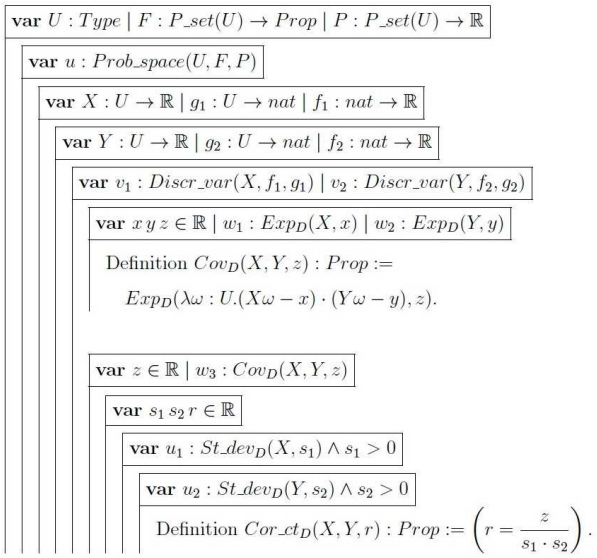
5) $\forall x\colon \mathbb{R}, \lim_{y\to\infty} F_{XY}(x, y) = F_X(x).$

Proof

The theorem is proven using the definitions and the theory of real numbers developed in the Library. ■

5.2. Covariance and Correlation Coefficient

In the following diagram we define the covariance and correlation coefficient of two discrete random variables.



Theorem 5.2. Properties of Covariance

Suppose (U, F, P) is a probability space, X, Y and Z are random variables. Then the following hold.

1) $\forall y\colon \mathbb{R}, Cov_D(X, Y, z) \to Cov_D(Y, X, z)$.
In ordinary mathematical notations:
$Cov(X, Y) = Cov(Y, X)$.

2) $\forall z\colon \mathbb{R}, Cov_D(X, X, z) \leftrightarrow Var_D(X, z)$.
In ordinary mathematical notations: $Cov(X, X) = Var(X)$.

3) $\forall z\ c\colon \mathbb{R}, Cov_D(X, \lambda\omega : U.c, z) \to z = 0$.
In ordinary mathematical notations: $Cov(X, c) = 0$.

4) $\forall z\ c\colon \mathbb{R}, Cov_D(X, Y, z) \to Cov_D(X, \lambda\omega.U.(Y\omega + c), z)$.
In ordinary mathematical notations: $Cov(X, Y + c) = Cov(X, Y)$.

5) $\forall z\ c\colon \mathbb{R}, Cov_D(X, Y, z) \to Cov_D(X, \lambda\omega.\ U.(c \cdot Y\omega), c \cdot z)$.
In ordinary mathematical notations: $Cov(X, cY) = cCov(X, Y)$.

6) $\forall y\ z\colon \mathbb{R}, Cov_D(X, Y, y) \wedge Cov_D(X, Z, z) \to Cov_D(X, Y + Z, y + z)$.
In ordinary mathematical notations: $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$.

7) $\forall x\ y\ z\colon \mathbb{R}, St\_dev_D(X, x) \wedge St\_dev_D(Y, y) \wedge Cov_D(X, Y, z) \to |z| \le x \cdot y$.
In ordinary mathematical notations: $|Cov(X, Y)| \le \sigma(X) \cdot \sigma(Y)$.

8) $\forall x\ y\ z\colon \mathbb{R}, Var_D(X, x) \wedge Var_D(Y, y) \wedge Cov_D(X, Y, z) \to Var_D(X + Y, x + y + 2z)$.
In ordinary mathematical notations: $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$.

9) $Indep\_var(X, Y) \to \forall z\colon \mathbb{R}, (Cov_D(X, Y, z) \to z = 0) \wedge \forall x\ y\colon \mathbb{R}, [Var_D(X, x) \wedge Var_D(Y, y) \to Var_D(X + Y, x + y)]$.

In ordinary mathematical notations: for independent X and Y, $Cov(X, Y) = 0$ and $Var(X + Y) = Var(X) + Var(Y)$.

10) $\forall x\ y\ z\colon \mathbb{R}, Exp_D(X, x) \wedge Exp_D(Y, y) \wedge Exp_D(X \cdot Y, z) \to Cov_D(X, Y, z - x \cdot y)$.

In ordinary mathematical notations: $Cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$.

Proof

Follows from the definitions and previous theorems about variance and standard deviation. ■

Theorem 5.3. Properties of Correlation Coefficient.

Suppose (U, F, P) is a probability space, X and Y are random variables. Then the following hold.

1) $\forall r\colon \mathbb{R}, Cor\_ct_D(X, Y, r) \to Cor\_ct_D(Y, X, r)$.
In ordinary mathematical notations: $\rho(X, Y) = \rho(Y, X)$.

2) $\forall y\colon \mathbb{R}, Var_D(X, y) \wedge y > 0 \to Cor\_ct_D(X, X, 1)$.
In ordinary mathematical notations: $\rho(X, X) = 1$.

3) $\forall r\colon \mathbb{R}, Cor\_ct_D(X, Y, r) \to -1 \le r \le 1$.
In ordinary mathematical notations: $-1 \le \rho(X, Y) \le 1$.

4) $Indep\_var(X, Y) \to \forall r\colon \mathbb{R}, [Cor\_ct_D(X, Y, r) \to r = 0]$.
In ordinary mathematical notations: for independent X and Y, $\rho(X, Y) = 0$.

Proof

Follows from the previous theorem. ■

6. Continuous Random Variables

6.1. Definition of a Continuous Random Variable

A random variable X is continuous if it has a density function. This is formally expressed in the following diagram.

$$\textbf{var } U : Type \mid F : P\_set(U) \to Prop \mid P : P\_set(U) \to \mathbb{R}$$
$$\textbf{var } u : Prob\_space(U, F, P)$$
$$\textbf{var } X : U \to \mathbb{R} \mid v : Random\_var(X)$$
$$\textbf{var } f : \mathbb{R} \to \mathbb{R}$$
$$\text{Definition } Dens\_func(f, X) : Prop := \forall x : \mathbb{R}, \left(fx \geqslant 0 \wedge F_X(x) = \int_{-\infty}^{x} f(t)dt \right).$$
$$\text{Definition } Cont\_var(X, f) : Prop := Dens\_func(f, X).$$

6.2. Numerical Characteristics of Continuous Random Variables

In the following diagram we define the expectation $Exp_C(X)$ of a continuous random variable X.

$$\textbf{var } U : Type \mid F : P\_set(U) \to Prop \mid P : P\_set(U) \to \mathbb{R}$$
$$\textbf{var } u : Prob\_space(U, F, P)$$
$$\textbf{var } X : U \to \mathbb{R} \mid v : Random\_var(X)$$
$$\textbf{var } f : \mathbb{R} \to \mathbb{R} \mid w : Cont\_var(X, f)$$
$$\textbf{var } y : \mathbb{R}$$
$$\text{Definition } Exp_C(X, y) : Prop := \left(y = \int_{-\infty}^{\infty} xf(x)dx. \right)$$

Other numerical characteristics can be given the same definitions as for discrete random variables and vectors. The numerical characteristics of continuous random variables and vectors should have the same properties as in the discrete case, including the Markov and Chebychev inequalities (and excluding Theorem 4.2). But the proofs need more work. A better approach would be to use Lebesgue integral to define an expectation of an arbitrary random variable. However, the definition and theory of Lebesgue integral are not developed in COIC/COQ yet, as far as we know. It is a task for the future. When Lebesgue integral is adequately introduced we will be able to have a universal approach to random variables, instead of considering only discrete and continuous variables and using different techniques in defining their expectations.

7. Conclusion

In this paper we follow the Kolmogorov's axiomatic approach and formalize in the COIC some fundamental concepts of probability theory. First we give a formal definition of a probability space as a triple of a sample space of outcomes, a sigma-field of events and a probability measure on these events. We derive usual basic properties of a probability space. Next we formalize the concepts of a random variable and its distribution function and also the concepts of a discrete random variable and a continuous random variable. We formally introduce the expectation of a random variable, separately for the discrete and continuous cases. We also give formal definitions of other numerical characteristics: variance, standard deviation, covariance and correlation coefficient. We study their properties in detail for the discrete case but not for the continuous case (though most properties should be the same in both cases). We formalize the Markov and Chebychev inequalities.

The definitions and initial proofs are presented in the form of flag-style derivations, as in (Nederpelt and Geuvers, 2014). The rest of the proofs, for readability, are made more informal and brief.

In the future we are planning to use a universal approach and formally define the expectation of an arbitrary random variable (not necessarily discrete or continuous) using a Lebesgue integral, when the Lebesgue integral is formalized in the COQ library. Other possible directions of future research are producing COQ codes for our formalizations and developing other parts of probability theory within COIC.

Acknowledgment

The author thanks the journal's Editor in Chief for his valuable support of this article.

Ethics

This is a mathematical article; no ethical issues can arise after its publication.

References

Bertot, Y. and P. Casteran, 2014. Interactive Theorem Proving and Program Development: Coq'Art: The Calculus of Inductive Constructions. 1st Edn., Springer, ISBN-10: 3662079658, pp: 472.

Coq Development Team. Reference manual. Standard Library. https://coq.inria.fr/.

Kolmogorov, A.N., 1950. Foundations of the Theory of Probability. 2nd Edn., Chelsea Publishing Company, pp: 71.

Moreira, D.A.C.V., 2012. Finite probability distributions in Coq. mei.di.uminho.pt/sites/default/files/dissertacoes//eeum_di_dissertacao_pg16019.pdf

Nederpelt, R. and H. Geuvers, 2014. Type Theory and Formal Proof. 1st Edn., Cambridge University Press, ISBN-10: 110703650X, pp: 436.

Rosenthal, J.S., 2006. A First Look at Rigorous Probability Theory. 2nd Edn., World Scientific, Singapore, ISBN-10: 9812703713, pp: 219.