A Hoeffding-Azuma Type Inequality for Random Processes

Mahir Hasanov

Department of Mathematics, Istanbul Beykent University, Türkiye

*Corresponding author: hasanov61@yahoo.com

Abstract. The subject of this paper is a Hoeffding-Azuma type estimation for the difference between an adapted random process and its conditional expectation given a related filtration.

1. Introduction

Hoeffding-Azuma type inequalities have very important applications in probability theory, statistics and different branches of science. In this section we give a brief history of Hoeffding-Azuma type inequalities.

1.1. Classical Hoeffding-Azuma type inequalities. Let \((\Omega, \mathcal{F}, P)\) be a probability triple, where \(\Omega\) is a sample space, \(\mathcal{F}\) is a \(\sigma\)-algebra on \(\Omega\) and \(P\) is a \(\sigma\)-additive probability measure on \(\mathcal{F}\). Let us denote by \(B\) the Borel algebra on \(\mathbb{R}\). Note that \(B = \sigma(\tau_{|x-y|})\), the minimal \(\sigma\)-algebra containing the natural topology \(\tau_{|x-y|}\) on \(\mathbb{R}\).

**Definition 1.1.** A function \(X: \Omega \rightarrow \mathbb{R}\) is called a random variable if \(X\) is \(\mathcal{F}\)-measurable function, i.e. \(X^{-1}(B) \subset \mathcal{F}\).

For definitions from the probability theory, used in this article see ([5], Sections 2.1, 6.1, 12.1 and 12.2).

The classical Hoeffding inequality is about finding upper bounds for the probability that the sum of \(n\) independent random variables exceeds its mean by a positive number \(nt\). The pioneering work was by Hoeffding ([6], Theorem 2) who proved the following theorem.
Theorem 1.1. (Hoeffding Inequality) If \( X_1, X_2, ..., X_n \) are independent random variables and \( a_i \leq X_i \leq b_i \), for \( i = 1, 2, ..., n \). Then,
\[
P(\bar{X} - \mu \geq t) \leq e^{-2nt^2/\sum_{i=1}^{n} (b_i - a_i)^2}
\]
where,
\[
\bar{X} = \frac{X_1 + X_2 + ... + X_n}{n}, \quad \mu = E(\bar{X}).
\]

Besides Hoeffding inequality there are two more classical fundamental results. These are Azuma inequality and Chernoff inequality. Let us give both inequalities in the form expressed by T. Tao ([9], Theorem 2.1.3 and Theorem 2.1.5).

Theorem 1.2. (Chernoff Inequality) Let \( X_1, X_2, ..., X_n \) be independent scalar random variables with \( |X_i| \leq K \) almost surely, with mean \( \mu_i \) and variance \( \sigma_i^2 \). Then for any \( \lambda > 0 \), one has
\[
P(|S_n - \mu| \geq \lambda \sigma) \leq C \max(\exp(-c\lambda^2), \exp(-c\lambda \sigma/K)),
\]
for some constants \( C, c > 0 \), where \( \mu = \sum_{i=1}^{n} \mu_i \), \( \sigma^2 = \sum_{i=1}^{n} \sigma_i^2 \) and \( S_n = X_1 + X_2 + ... + X_n \).

Theorem 1.3. (Azuma Inequality) Let \( X_1, X_2, ..., X_n \) be a sequence of scalar random variables with \( |X_i| \leq 1 \) almost surely. Assume also that we have martingale difference property
\[
E(X_i|X_1, ..., X_{i-1}) = 0
\]
almost surely, for all \( i = 1, ..., n \). Then for any \( \lambda > 0 \) the sum \( S_n = X_1 + X_2 + ... + X_n \) obey the large deviation inequality
\[
P(|S_n| \geq \lambda \sqrt{n}) \leq C \exp(-c\lambda^2)
\]
for some constants \( C, c > 0 \).

1.2. Hoeffding-Azuma type inequalities for Martingale differences. Martingales and Markov chains are known to be widely used areas of Hoeffding-Azuma type inequalities.

Definition 1.2. (see [5], Section 12.1) A sequence of random variables \( Y = \{Y_n : n \geq 0\} \) is martingale with respect to the sequence \( X = \{X_n : n \geq 0\} \) if, for all \( n \geq 0 \)
\[
a) \quad E(|Y_n|) < \infty,
\]
\[
b) \quad E(Y_{n+1}|X_0, X_1, ..., X_n) = Y_n
\]

The Hoeffding inequality for martingale differences is of supreme importance in the theory of martingales (see [5], Section 12.2.)

Theorem 1.4. (Hoeffding inequality for martingale differences) Let \( Y = \{Y_n : n \geq 0\} \) be a martingale, and suppose that there exists a sequence \( K_1, K_2, ..., \) of real numbers such that \( P(|Y_n - Y_{n-1}| \leq K_n) = 1 \) for all \( n \). Then
\[
P(|Y_n - Y_0| \geq x) \leq 2 \exp\left(-\frac{1}{2} x^2 / \sum_{i=1}^{n} K_i^2\right) \quad x > 0.
\]
(1.1)
Inequality (1.1) means that if the martingale differences are bounded then the large deviation of $Y_n$ from its initial value $Y_0$ is small.

There are many new results and their applications in the literature on these issues.

A generalization of Hoeffding inequality for dependent random variables was given in [10]. Optimal bounds for Hoeffding’s inequalities were found in [8]. In [4] it was proved a Hoeffding type inequality to partial sums that are derived from a uniformly ergodic Markov chain. New type of inequalities were introduced in [2] (see also references therein). In [3] some inequalities were obtained for unbounded random variables.

[7] Significantly improved the well-known Bennett-Hoeffding bound for sums of independent random variables by using, instead of the class of all increasing exponential functions, a much larger class of generalized moment functions. The resulting bounds have certain optimality properties.

2. Inequalities for Adopted Random Processes

A random process is a collection of random variables $\{X(t), t \in T\}$. Particularly, a sequence $X_0, X_1, X_2, \ldots, X_n, \ldots$ of random variables defined on the same probability triple $(\Omega, \mathcal{F}, P)$ is a random process.

**Definition 2.1.** Let $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be a sequence of $\sigma$-sub-algebras of $\mathcal{F}$, such that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_n \subseteq \ldots \subseteq \mathcal{F}$. Then $\{\mathcal{F}_n\}_{n=0}^{\infty}$ is called a filtration of $(\Omega, \mathcal{F}, P)$ and a sequence $\{X_n : n \geq 0\}$ of random variables is said to be adapted to the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$ if $X_n$ is $\mathcal{F}_n$ measurable for all $n$.

We denote by $E(X_n|\mathcal{F}_{n-1})$ the condition expectation of $X_n$ given $\mathcal{F}_{n-1}$ for all $n \geq 1$.

As pointed out in the introduction the classical Hoeffding inequality is about finding upper bounds for the probability that the sum of $n$ independent random variables exceeds its mean by $nt$.

Finding upper bounds for the probability that the sum of $n$ terms of an adapted random process exceeds its conditional mean by $x$ is also among the topics of interest.

As far as we know, no previous research has investigated $P\left(\left|\sum_{i=1}^{n} (X_i - E(X_i|\mathcal{F}_{i-1}))\right| \geq x\right)$ for general adopted random processes, which is the main subject of this paper.

The basic result is the following theorem.

**Theorem 2.1.** Let $\{X_n : n \geq 0\}$ be a sequence of random variables which is adapted to a filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$. If $|X_n - E(X_n|\mathcal{F}_{n-1})| \leq C_n$ almost surely, for $n \geq 0$ then

$$P\left(\left|\sum_{i=1}^{n} (X_i - E(X_i|\mathcal{F}_{i-1}))\right| \geq x\right) \leq 2e^{-\frac{x^2}{2\sum_{i=1}^{n} C_i^2}}$$

(2.1)

**Proof.** Let us define the following sequence:
\(Y_0 = X_0;\)
\(Y_1 = X_0 + (X_1 - E(X_1|\mathcal{F}_0));\)
\(\vdots\)
\(Y_n = X_0 + (X_1 - E(X_1|\mathcal{F}_0)) + \ldots + (X_n - E(X_n|\mathcal{F}_{n-1})).\)

We have
\[Y_n = Y_{n-1} + (X_n - E(X_n|\mathcal{F}_{n-1})).\]  \hspace{1cm} (2.2)

Evidently, \(Y_n\) is \(\mathcal{F}_n\) measurable for all \(n\) and \(E\left( (X_n - E(X_n|\mathcal{F}_{n-1})), \mathcal{F}_{n-1} \right) = 0\). Then it follows form (2.2) that \(E(Y_n|\mathcal{F}_{n-1}) = Y_{n-1}\), i.e. \(\{Y_n; n \geq 0\}\) is a martingale sequence. Note that,
\[
\sum_{i=1}^{n} (X_i - E(X_i|\mathcal{F}_{i-1})) = Y_n - Y_0. \hspace{1cm} (2.3)
\]

By (2.3)
\[
P\left( \sum_{i=1}^{n} (X_i - E(X_i|\mathcal{F}_{i-1})) \geq x \right) = P(Y_n - Y_0 \geq x). \hspace{1cm} (2.4)
\]

Let \(\theta > 0\), then
\[
E(e^{\theta(Y_n-Y_0)}) = \int_{\Omega} e^{\theta(Y_n-Y_0)} dP \geq \int_{Y_n-Y_0 \geq x} e^{\theta(Y_n-Y_0)} dP \geq e^{\theta x} \int_{Y_n-Y_0 \geq x} dP.
\]

Hence,
\[
E(e^{\theta(Y_n-Y_0)}) \geq e^{\theta x} \int_{Y_n-Y_0 \geq x} dP = e^{\theta x} P(Y_n - Y_0 \geq x)
\]

and
\[
P(Y_n - Y_0 \geq x) \leq e^{-\theta x} E(e^{\theta(Y_n-Y_0)}). \hspace{1cm} (2.5)
\]

Next, we estimate \(E(e^{\theta(Y_n-Y_0)})\). By the tower property we obtain that
\[
E(e^{\theta(Y_n-Y_0)}) = E \left( E(e^{\theta(Y_n-Y_0)|\mathcal{F}_{n-1}}) \right). \hspace{1cm} (2.6)
\]
\[
E(e^{\theta(Y_n-Y_0)|\mathcal{F}_{n-1}}) = e^{\theta(Y_{n-1}-Y_0)} E(e^{\theta(Y_n-Y_{n-1})|\mathcal{F}_{n-1}}).
\]

We set \(f(y) = e^{\theta y}, |y| \leq C_n\). The function \(f(y) = e^{\theta y}\) is convex, whence it follows that
\[
e^{\theta y} \leq \frac{1}{2}(1 - \frac{y}{C_n}) e^{-\theta C_n} + \frac{1}{2}(1 + \frac{y}{C_n}) e^{\theta C_n} = \left( \frac{1}{2} e^{-\theta C_n} + \frac{1}{2} e^{\theta C_n} \right) + \frac{y}{2 C_n} (e^{\theta C_n} - e^{-\theta C_n}).
\]

Thus,
\[
e^{\theta y} \leq e^{\frac{1}{2} \theta^2 C_n^2} + \frac{y}{2 C_n} (e^{\theta C_n} - e^{-\theta C_n}), \hspace{1cm} |y| \leq C_n. \hspace{1cm} (2.7)
\]

\(Y_n - Y_{n-1} = X_n - E(X_n|\mathcal{F}_{n-1})\) and by the condition of the theorem \(|Y_n - Y_{n-1}| \leq C_n\). Then setting \(y = Y_n - Y_{n-1}\) in (2.7) we obtain
\[
e^{\theta(Y_n-Y_{n-1})} \leq e^{\frac{1}{2} \theta^2 C_n^2} + \frac{Y_n - Y_{n-1}}{2 C_n} (e^{\theta C_n} - e^{-\theta C_n}), \hspace{1cm} |y| \leq C_n. \hspace{1cm} (2.8)
\]

Taking the conditional expectation of (2.8) and using the fact that \(E(Y_n - Y_{n-1}|\mathcal{F}_{n-1}) = 0\) we get
\[ E\left(e^{\theta(Y_n - Y_{n-1})} \mid \mathcal{F}_{n-1}\right) \leq e^{\frac{1}{2}\theta^2 C_n^2}. \]

Therefore,
\[ E\left(e^{\theta(Y_n - Y_0)} \mid \mathcal{F}_{n-1}\right) \leq e^{\theta(Y_{n-1} - Y_0)} e^{\frac{1}{2}\theta^2 C_n^2}. \]

By using this inequality, (2.6) and iterations we get
\[ E\left(e^{\theta(Y_n - Y_0)}\right) = E\left(E\left(e^{\theta(Y_n - Y_0)} \mid \mathcal{F}_{n-1}\right)\right) \leq E\left(e^{\theta(Y_{n-1} - Y_0)}\right) e^{\frac{1}{2}\theta^2 C_n^2} \leq e^{\frac{1}{2}\theta^2 \sum_{i=1}^n C_i^2}. \]

It follows from (2.5) that
\[ P\left(\sum_{i=1}^n (X_i - E(X_i \mid \mathcal{F}_{i-1}) \geq x\right) \leq e^{-\theta x + \frac{1}{2}\theta^2 \sum_{i=1}^n C_i^2}. \]

Finally, minimizing the right side of this inequality in \( \theta \) and replacing the terms under the sum we obtain the needed inequality (2.1).
\[ P\left(\left|\sum_{i=1}^n (X_i - E(X_i \mid \mathcal{F}_{i-1})\right| \geq x\right) \leq 2e^{-\frac{x^2}{2\theta^2 \sum_{i=1}^n C_i^2}}. \]

**Note.** If \( \{X_n : n \geq 0\} \) is a martingale sequence, then \( \sum_{i=1}^n (X_i - E(X_i \mid \mathcal{F}_{i-1})) = X_n - X_0 \) and in this case we get the martingale difference inequality (1.1).

3. Random Processes in the Hilbert Space \( L^2(\Omega, \mathcal{F}, P) \)

Let \( X = \{X_i : i \geq 0\} \) be an adopted random process and \( X_i \in L^2(\Omega, \mathcal{F}_i, P) \) for all \( i \). As can be seen from the following theorem that if \( X_i \in L^2(\Omega, \mathcal{F}_i, P) \) then the conditional expectation \( E(X_i \mid \mathcal{F}_{i-1}) \) is a version of orthogonal projection of \( X_i \) onto the subspace \( L^2(\Omega, \mathcal{F}_{i-1}, P) \).

**Theorem 3.1.** (see [1]) Let \( (X, Z) \) be a bivariate random vector and \( L_Z = \{g(Z) \mid g(Z) \in L_2(\Omega), g \text{ is a Borel function}\} \). Let \( E[|X|^2] < \infty \). Then there exists a Borel function \( g_0 : \mathbb{R} \to \mathbb{R} \) with \( E[(g_0(Z))^2] < \infty \), such that \( E[(X - g_0(Z))^2] = \inf \{E[(X - g(Z))^2] \mid g(Z) \in L_Z\} \). Moreover, \( g_0(Z) = E[X \mid Z] \).

By a using a Hilbert space property we can write
\[ X_i = E(X_i \mid \mathcal{F}_{i-1}) + Y_i, \]  
where \( Y_i \in (L^2(\Omega, \mathcal{F}_{i-1}, P))^\perp \) - the orthogonal complement of the subspace \( L^2(\Omega, \mathcal{F}_{i-1}, P) \). By (3.1) we have \( Y_i \) is \( \mathcal{F}_i \) measurable and
\[ Y_i = X_i - E(X_i \mid \mathcal{F}_{i-1}). \]  

An immediate consequence of (3.2) is

**Corollary 1.** \( E(Y_i \mid \mathcal{F}_{i-1}) = 0 \).

The main conclusion of the above given arguments is given in the following theorem.
Theorem 2. Let $Y_n$ be a version of orthogonal projection of $X_n$ onto the subspace $L^2(\Omega, \mathcal{F}_{n-1}, P)^\perp$, $n \geq 0$.

If $|Y_n| \leq C_n$ almost surely, for $n \geq 0$ then

$$P\left(\left|\sum_{i=1}^{n} Y_i\right| \geq x\right) \leq 2e\left(-\frac{x^2}{\sum_{i=1}^{n} C_i^2}\right).$$

Conflicts of Interest: The author declares that there are no conflicts of interest regarding the publication of this paper.

References


