Chebyshev’s inequality for Banach-space-valued random elements

Ling Zhou and Ze-Chun Hu*
Nanjing University

Abstract In this paper, we obtain a new generalization of Chebyshev’s inequality for random elements taking values in a separate Banach space.

Keywords Chebyshev’s inequality, Banach space

1 Introduction

Chebyshev’s inequality states that for a random variable $X$ with mean $E(X)$ and variance $\text{Var}(X)$, and any $\varepsilon > 0$,

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

This inequality plays an important role in probability theory and statistics. Several generalization for random vectors have been made. A natural one is as follows.

Suppose that $X$ is an $n$-dimensional random vector, then for $\varepsilon > 0$,

$$P\{\|X - E(X)\| \geq \varepsilon\} \leq \frac{\text{Var}(X)}{\varepsilon^2},$$

where $\| \cdot \|$ denotes the Euclidean norm in $\mathbb{R}^n$ and $\text{Var}(X) = E[\|X - E(X)\|^2]$. This result can be seen in Laha and Rohatgi (1979), P. 446-451. Other generalizations are given in Marshall and Olkin (1960), Godwin (1955) and Mallows (1956).

Grenander (1963) proved a Chebyshev’s inequality for Hilbert-space-valued random elements as follows: if $X$ is a random element taking values in a Hilbert space $H$ with $E(\|X\|^2) < \infty$, then for $\varepsilon > 0$,

$$P\{\|X\| \geq \varepsilon\} \leq \frac{E(\|X\|^2)}{\varepsilon^2}.$$

*Corresponding author: Department of Mathematics, Nanjing University, Nanjing 210093, PR China
E-mail address: huzc@nju.edu.cn
Chen (2007) proved the following new generalization of Chebyshev’s inequality for random vectors.

**Theorem 1.1** Suppose that $X$ is an $n$-dimensional random vector with positive definite covariance matrix $\Sigma$. Then, for any $\varepsilon > 0$,

$$P\{(X - E(X))^T \Sigma^{-1} (X - E(X)) \geq \varepsilon\} \leq \frac{n}{\varepsilon},$$

where the superscript “$T$” denotes the transpose of a matrix.

Rao (2010) extended Theorem 1.1 to random elements taking values in a separable Hilbert space as follows.

**Theorem 1.2** Suppose that $X$ is a random element taking values in a separable Hilbert space $H$ with expectation zero, positive definite covariance operator $S$, and probability distribution $\mu$ such that $\int_H ||x||^2 \mu(dx) < \infty$. Then, for every $\varepsilon > 0$,

$$P\{(SX, X) > \varepsilon\} \leq \frac{\left[\int_H ||x||^2 \mu(dx)\right]^2}{\varepsilon}$$

and

$$P\{(S^{-1}X, X) > \varepsilon\} \leq \frac{\left[||S^{-1}|| \int_H ||x||^2 \mu(dx)\right]^2}{\varepsilon},$$

where covariance operator $S$ is the Hermitian operator determined uniquely by the quadratic form $(Sy, y) = \int_H (x, y)^2 \mu(dx)$.

In this paper, we will extend Theorem 1.2 to random elements taking values in a separable Banach space.

## 2 Main result

Suppose that $B$ is a real separable Banach space, $\| \cdot \|$ is the norm, and $\mathcal{B}(B)$ is the Borel $\sigma$-algebra. $B^*$ is the dual space of $B$, i.e. $B^*$ is the family of all bounded linear functionals on $B$. It’s well known that $B^*$ is also a Banach space with the operator norm $\| \cdot \|^*$ defined by

$$\|f\|^* = \sup_{\|x\| \leq 1} |f(x)|, \forall f \in B^*.$$

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A mapping $X : (\Omega, \mathcal{F}) \to (B, \mathcal{B}(B))$ is called measurable if for any $A \in \mathcal{B}(B)$, we have $X^{-1}(A) \in \mathcal{F}$, where $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$. We call such $X$ random element taking values in $B$. Let $\mu$ be the probability distribution of $X$, i.e. $\mu$ is the probability measure on $(B, \mathcal{B}(B))$ defined by

$$\mu(A) = P(X^{-1}(A)), \forall A \in \mathcal{B}(B).$$
**Definition 2.1** Suppose that a mapping \( X : (\Omega, \mathcal{F}) \to (B, \mathcal{B}(B)) \) is measurable. \( X \) is called a step function if it can be expressed by
\[
X = \sum_{i=1}^{\infty} x_i I_{A_i},
\]
where \( \forall i \in \mathbb{N}, x_i \in B, A_i \in \mathcal{F} \) and \( A_i \cap A_j = \emptyset, \forall i \neq j \).

By the separability of \( B \), we can easily get

**Lemma 2.2** Suppose that \( X : (\Omega, \mathcal{F}) \to (B, \mathcal{B}(B)) \) is measurable. Then there exists a sequence \( \{X_n, n \geq 1\} \) of step functions such that
\[
\lim_{n \to \infty} \sup_{\omega \in \Omega} \|X_n(\omega) - X(\omega)\| = 0
\]
and \( \forall \omega \in \Omega, \forall n \geq 1, \|X_n(\omega)\| \leq 2\|X(\omega)\| \).

To state our main result, we should define the covariance operator of random element \( X \) taking values in the Banach space \( B \). For any \( x \in B, f \in B^* \), define \((x, f) = (f, x) = f(x)\).

**Theorem 2.3** Suppose that \( X \) is a random element taking values in \( B \) with probability distribution \( \mu \). If \( \int_B \|x\|^2 \mu(dx) < \infty \), then the quadratic form
\[
(Sf, g) = \int_B (x, f)(x, g) \mu(dx), \forall f, g \in B^*, \tag{2.1}
\]
uniquely determines one bounded linear operator \( S : B^* \to B \).

**Proof. Existence:** Firstly, we prove that \( \forall f \in B^*, \exists Sf \in B \) such that
\[
(Sf, g) = \int_B f(x)g(x) \mu(dx), \forall g \in B^*. \tag{2.2}
\]
By Lemma 2.2, there exists a sequence \( \{X_n, n \geq 1\} \) of step functions such that \( X_n \) converges to \( X \) uniformly and \( \|X_n(\omega)\| \leq 2\|X(\omega)\|, \forall \omega \in \Omega, \forall n \geq 1 \). Let \( \mu_n \) be the probability distribution of \( X_n \). Then for any \( n \geq 1 \), we have
\[
\int_B \|x\|^2 \mu_n(dx) = \int_\Omega \|X_n\|^2 dP \leq 4 \int_\Omega \|X\|^2 dP = 4 \int_B \|x\|^2 \mu(dx) < \infty.
\]
Let \( X_1 \) have the expression that \( X_1 = \sum_{i=1}^{\infty} x_i I_{A_i} \) as in Definition 2.1. For \( f \in B^* \), define
\[
S_1f = \sum_{i=1}^{\infty} f(x_i) P(A_i)x_i. \tag{2.3}
\]
Since
\[
\sum_{i=1}^{\infty} \| f(x_i) P(A_i) x_i \| = \int_{\Omega} \| f(X_1) X_1 \| dP \leq \| f \| \int_{\Omega} \| X_1 \|^2 dP = \| f \| \int_{B} \| x \|^2 \mu_1(dx) < \infty, \tag{2.4}
\]

$S_1f$ is well defined and we have
\[
S_1f = \int_{\Omega} f(X_1) X_1 dP. \tag{2.5}
\]
By (2.3) and (2.4), we get that for any $g \in B^*$,
\[
(S_1f, g) = \sum_{i=1}^{\infty} f(x_i) P(A_i) g(x_i) = \int_{\Omega} f(X_1) g(X_1) dP = \int_{B} f(x) g(x) \mu_1(dx). \tag{2.6}
\]
For $n = 2, 3, \ldots$, define $S_n f$ from $X_n$ similar to $S_1 f$. In particular, we have
\[
S_n f = \int_{\Omega} f(X_n) X_n dP, \tag{2.7}
\]
and for any $g \in B^*$,
\[
(S_n f, g) = \int_{\Omega} f(X_n) g(X_n) dP = \int_{B} f(x) g(x) \mu_n(dx). \tag{2.8}
\]
For any $n, m = 1, 2, \ldots$, by (2.5) and (2.6), we have
\[
\| S_n f - S_m f \| = \left\| \int_{\Omega} (f(X_n) X_n - f(X_m) X_m) dP \right\| \leq \int_{\Omega} \| f(X_n) X_n - f(X_m) X_m \| dP = \int_{\Omega} \| f(X_n) (X_n - X_m) + X_m (f(X_n) - f(X_m)) \| dP \leq \| f \| \int_{\Omega} (\| X_n \| + \| X_m \|) (\| X_n - X_m \|) dP. \tag{2.8}
\]
By the fact that $\| X_n(\omega) \| \leq 2 \| X(\omega) \|, \forall n \geq 1, \forall \omega \in \Omega$, we have
\[
\int_{\Omega} (\| X_n \| + \| X_m \|) (\| X_n - X_m \|) dP \leq 16 \int_{\Omega} \| X \|^2 dP = 16 \int_{B} \| x \|^2 \mu(dx) < \infty.
\]
Then it follows from (2.8) and the dominated convergence theorem that
\[ \|S_n f - S_m f\| \to 0 \text{ as } n, m \to \infty, \]
which implies that \{S_n f, n \geq 1\} is a Cauchy sequence in \( B \). Thus there exists a unique element denoted by \( Sf \) such that \( S_n f \) converges to \( Sf \) in \( B \). Furthermore, by (2.7) and integral transformation, we have
\[
(Sf, g) = \int_{\Omega} f(X)g(X)dP = \int_B f(x)g(x)\mu(dx), \quad \forall g \in B^*.
\]

Secondly, we prove that \( S \) is linear. Suppose that \( f_1, f_2 \in B^* \) and \( a, b \in \mathbb{R} \). Then for any \( g \in B^* \),
\[
(S(af_1 + bf_2), g) = \int_B (x, af_1 + bf_2)(x, g)\mu(dx)
\]
\[
= a \int_B (x, f_1)(x, g)\mu(dx) + b \int_B (x, f_2)(x, g)\mu(dx)
\]
\[
= (aSf_1, g) + (bSf_2, g)
\]
which implies that \( S(af_1 + bf_2) = aSf_1 + bSf_2 \).

Thirdly, we prove that \( S \) is bounded. By (2.6), for any \( n \geq 2 \), we have
\[
\|S_n f\| \leq \|f\|^* \int_{\Omega} \|X_n\|^2dP. \tag{2.9}
\]
Notice that \( S_n f \) converges to \( Sf \) and \( X_n \) converges uniformly to \( X \). Then letting \( n \to \infty \) in (2.9), we get
\[
\|Sf\| \leq \|f\|^* \int_{\Omega} \|X\|^2dP = \|f\|^* \int_B \|x\|^2\mu(dx),
\]
which implies that \( S \) is a bounded operator from \( B^* \) to \( B \).

**Uniqueness:** Suppose that \( S' : B^* \to B \) is another bounded linear operator satisfying that
\[
(S'f, g) = \int_B (x, f)(x, g)\mu(dx), \forall f, g \in B^*.
\]
Then for any \( f \in B^* \), we have
\[
(S'f - Sf, g) = 0, \quad \forall g \in B^*,
\]
which implies that \( S'f - Sf = 0 \). Thus \( S' = S \).

Suppose that \( P^* \) is a probability measure on \( (B^*, B(B^*)) \). Since \( S \) is a bounded linear operator from \( B^* \) to \( B \), we can check that \( f \mapsto (Sf, f) \) is a nonnegative continuous functional on \( B^* \), and thus it is measurable with respect to \( B(B^*) \). For any \( \varepsilon > 0 \), define
\[
D_\varepsilon = \{ f \in B^* : (Sf, f) \geq \varepsilon \}.
\]
Then

\[ P^*(D_\varepsilon) \leq \frac{1}{\varepsilon} \int_{D_\varepsilon} (Sf, f) P^*(df) \]

\[ \leq \frac{1}{\varepsilon} \int_{B^*} (Sf, f) P^*(df) \]

\[ = \frac{1}{\varepsilon} \int_{B^*} \left( \int_B f^2(x) \mu(dx) \right) P^*(df) \]

\[ \leq \frac{1}{\varepsilon} \int_{B^*} \left( \int_B (\|f\|^*)^2 \|x\|^2 \mu(dx) \right) P^*(df) \]

\[ = \frac{1}{\varepsilon} \int_{B^*} (\|f\|^*)^2 P^*(df) \left( \int_B \|x\|^2 \mu(dx) \right). \]

We have known that \( S \) is nonnegative definite, i.e. for any \( f \in B^* \), \((Sf, f) \geq 0\). Furthermore, if \( S \) is positive definite in the sense that \((Sf, f) = 0\) implies that \( f = 0 \), then \( S \) is invertible. For any \( y \in B \), we have

\[ (SS^{-1}y, S^{-1}y) = \int_B (x, S^{-1}y)^2 \mu(dx) \geq 0, \]

i.e.

\[ (y, S^{-1}y) = \int_B (x, S^{-1}y)^2 \mu(dx) \geq 0. \]

Define

\[ D'_\varepsilon = \{ y \in B : (S^{-1}y, y) \geq \varepsilon \}. \]

Then

\[ P\{X \in D'_\varepsilon\} = \int_{D'_\varepsilon} \mu(dy) \]

\[ \leq \frac{1}{\varepsilon} \int_{D'_\varepsilon} (y, S^{-1}y) \mu(dy) \]

\[ \leq \frac{1}{\varepsilon} \int_B (y, S^{-1}y) \mu(dy) \]

\[ = \frac{1}{\varepsilon} \int_B \left( \int_B (x, S^{-1}y)^2 \mu(dx) \right) \mu(dy) \]

\[ \leq \frac{1}{\varepsilon} \int_B \left( \int_B \|x\|^2 (\|S^{-1}y\|^*)^2 \mu(dx) \right) \mu(dy) \]

\[ \leq \frac{1}{\varepsilon} \int_B \left( \int_B \|x\|^2 \|S^{-1}\|^2 \mu(dx) \right) \|y\|^2 \mu(dy) \]

\[ = \frac{1}{\varepsilon} \|S^{-1}\|^2 \left[ \int_B \|x\|^2 \mu(dx) \right]^2, \]

where \( \|S^{-1}\| \) is the operator norm of \( S^{-1} : B \to B^* \). Hence we have the following result.
Theorem 2.4 Suppose that $X$ is a random element taking values in $\mathcal{B}$ with probability distribution $\mu$ satisfying that $\int_{\mathcal{B}} \|x\|^2 \mu(dx) < \infty$, $P^*$ is a probability measure on $(\mathcal{B}^*, \mathcal{B}(\mathcal{B}^*))$, $S : \mathcal{B}^* \to \mathcal{B}$ is the bounded linear operator defined by Theorem 2.3. Then for any $\varepsilon > 0$, we have

$$P^* \{ f \in \mathcal{B}^* : (Sf, f) \geq \varepsilon \} \leq \frac{1}{\varepsilon} \int_{\mathcal{B}^*} (\|f\|)^2 P^*(df) \left( \int_{\mathcal{B}} \|x\|^2 \mu(dx) \right)$$

and

$$P\{(S^{-1}X, X) \geq \varepsilon \} \leq \frac{1}{\varepsilon} \|S^{-1}\|^2 \left[ \int_{\mathcal{B}} \|x\|^2 \mu(dx) \right]^2.$$

3 Remarks

In this section, we show that Theorem 2.4 extends Theorem 1.2. Let $H$ be a separable Hilbert space with inner product $(\cdot, \cdot)_H$ and $H^*$ be the dual space of $H$ with norm $\| \cdot \|^*$. Let $X$ be a random element taking values in $H$ with probability distribution $\mu$ satisfying $\int_{H} \|x\|^2 \mu(dx) < \infty$. Let $S_H$ be the covariance operator defined in Theorem 1.2, and $S$ be the bounded linear operator from $H^*$ to $H$ defined by Theorem 2.3.

By Riesz representation theorem, there exists an isometry $T : H \to H^*$. In fact, for any $x \in H$, $Tx(y) = (y, x)_H$, $\forall y \in H$, and $\|Tx\|^* = \|x\|$. As in Section 2, for any $x \in H, f \in H^*$, we define $(x, f) = (f, x) = f(x)$. Then we have $ST = S_H$. In fact, for any $y \in H$, we have $STy \in H$ and

$$(STy, y)_H = Ty(STy) = (STy, Ty) = \int_H ((Ty(x))^2 \mu(dx) = \int_H (x, y)_H^2 \mu(dx) = (S_Hy, y)_H.$$ 

Define one probability measure $P^*$ on $(H^*, \mathcal{B}(H^*))$ by

$$P^*(A) = \mu(T^{-1}(A)), \forall A \in \mathcal{B}(H^*).$$

Then

$$P^*(D_\varepsilon) = \mu(T^{-1}(D_\varepsilon)) = P(X^{-1}(T^{-1}(D_\varepsilon))) = P\{\omega \in \Omega : TX(\omega) \in D_\varepsilon \} = P\{\omega \in \Omega : (STX(\omega), TX(\omega)) \geq \varepsilon \}. \quad (3.1)$$

By the fact that $\|Tx\|^* = \|x\|, \forall x \in H$, and integral transformation, we have

$$\int_{H} \|x\|^2 \mu(dx) = \int_{H} (\|Tx\|^*)^2 \mu(dx) = \int_{H^*} (\|f\|^*)^2 P^*(df). \quad (3.2)$$
By Theorem 2.4, (3.1) and (3.2), we obtain
\[
P\{\omega \in \Omega : (STX(\omega), TX(\omega)) \geq \varepsilon \} \leq \frac{1}{\varepsilon} \int_{H^*} (\|f\|^2) P^*(df) \left( \int_H \|x\|^2 \mu(dx) \right)
= \frac{1}{\varepsilon} \left( \int_H \|x\|^2 \mu(dx) \right)^2.
\] (3.3)

By \(ST = S_H\) and the definition of \(T\), we have
\[(STX(\omega), TX(\omega)) = (S_H X, X)_H.\] (3.4)

It follows from (3.3) and (3.4) that
\[
P\{(S_H X, X)_H \geq \varepsilon \} \leq \frac{1}{\varepsilon} \left( \int_H \|x\|^2 \mu(dx) \right)^2.
\] (3.5)

On the other hand, if \(S_H\) is positive definite, then \(S\) is positive definite by \(ST = S_H\) and the isometry of \(T\). Note that \(S^{-1} X \in H^*\) and
\[(S^{-1} X, X) = S^{-1} X(X) = T^{-1} S^{-1} X(X) = (X, T^{-1} S^{-1} X)_H = (S^{-1}_H X, X)_H.\]
Then we have
\[
P\{X \in D'_{\varepsilon}\} = P\{(S^{-1} X, X) \geq \varepsilon\} = P\{(S^{-1}_H X, X)_H \geq \varepsilon\}.
\] (3.6)

By \(S^{-1}_H = T^{-1} S^{-1}\) and the isometry of \(T\), we have \(\|S^{-1}\| = \|S^{-1}_H\|\). By Theorem 2.4 and (3.6), we obtain that
\[
P\{(S^{-1}_H X, X) \geq \varepsilon\} \leq \frac{1}{\varepsilon} \|S^{-1}_H\|^2 \left( \int_H \|x\|^2 \mu(dx) \right)^2.
\] (3.7)

Inequalities (3.5) and (3.7) are just those two ones in Theorem 1.2.

Acknowledgments

Research supported by NNSFC (Grant No. 10801072) and the Fundamental Research Funds for the Central Universities.

References

[1] Chen, X., 2007. A new generalization of Chebyshev inequality for random vectors. arXiv: 0707.0805v1[math.ST] 5 Jul 2007.

[2] Godwin, H.J., 1955. On generalizations of Chebyshev inequality. Journal of the American Statistical Association 50, 923-945.
[3] Grenander, U., 1963. Probabilities on Algebraic Structures. Wiley, New York, Almqvist and Wiksell, Stockholm.

[4] Laha, R.G., Rohatgi, V.K., 1979. Probability Theory. Wiley, New York.

[5] Mallows, C.L., 1956. Generalizations of Chebyshev inequalities. Journal of Royal Statistical Society, Series B 18, 139-171.

[6] Marshall, A., Olkin, I., 1960. Multivariate Chebyshev inequalities. Annals of Mathematical Statistics 31, 1001-1014.

[7] Rao, B.L.S.P., 2010. Chebyshev’s inequality for Hilbert-space-valued random elements. Statistics and Probability Letters 80, 1039-1042.