Integrated quantile functions: properties and applications

Alexander A. Gushchin[a,b,c,*]**, Dmitriy A. Borzykh**[b]

[a]*Steklov Mathematical Institute, Gubkina 8, 119991 Moscow, Russia*
[b]*National Research University Higher School of Economics,*
Myasnitskaya 20, 101000 Moscow, Russia
[c]*Lomonosov Moscow State University,*
Leninskie gory, 119991 Moscow, Russia

gushchin@mi.ras.ru (A.A. Gushchin), dborzykh@hse.ru (D.A. Borzykh)

Abstract   In this paper we provide a systematic exposition of basic properties of integrated
distribution and quantile functions. We define these transforms in such a way that they charac-
terize any probability distribution on the real line and are Fenchel conjugates of each other. We
show that uniform integrability, weak convergence and tightness admit a convenient character-
ization in terms of integrated quantile functions. As an application we demonstrate how some
basic results of the theory of comparison of binary statistical experiments can be deduced using
integrated quantile functions. Finally, we extend the area of application of the Chacon–Walsh
construction in the Skorokhod embedding problem.

Keywords   Quantile functions, integrated quantile functions, integrated distribution
functions, convex stochastic order, binary experiments, Chacon–Walsh construction
2010 MSC   60E05, 62E15, 60B10, 26A48

1   Introduction

Integrated distribution and quantile functions or simple transformations of them play
an important role in probability theory, mathematical statistics, and their applications

such as insurance, finance, economics etc. They frequently appear in the literature, often under different names. Moreover, in many occasions they are defined under additional assumption of integrability of a random variable or at least integrability of the positive or the negative part of a variable. Let us point out only few references. For a random variable X, let F_X be the distribution function of X and q_X any quantile function of X. Examples of integrated distribution functions or their simple modifications are:

- The function

$$\Psi_X(x) := \int_{-\infty}^{x} F_X(t)\,dt, \quad x \in \mathbb{R}, \tag{1}$$

  considered in [11].

- The integrated survival function

$$H_X(x) := \int_{x}^{+\infty} \left(1 - F_X(t)\right) dt, \quad x \in \mathbb{R}, \tag{2}$$

  of X, also called the stop-loss transform, see, e.g., [19, 20].

- The potential function

$$U_X(x) := -\mathsf{E}|x - X|, \quad x \in \mathbb{R}, \tag{3}$$

  of X, see, e.g., [4, 6].

These transforms characterize the distribution of X only if the expectations $\mathsf{E}X^-$, or $\mathsf{E}X^+$, or $\mathsf{E}|X|$ respectively are finite; otherwise, the transforms equal $+\infty$ or $-\infty$ identically and do not allow to identify the distribution of X.

The examples of integrated quantile functions or their simple transformations are:

- The absolute Lorenz curve

$$\mathrm{AL}_X(u) := \int_{0}^{u} q_X(s)\,ds, \quad u \in [0,\,1], \tag{4}$$

  see, e.g., [21] and the references therein.

- The Hardy–Littlewood maximal function

$$\mathrm{HL}_X(u) := \frac{1}{1-u} \int_{u}^{1} q_X(s)\,ds, \quad u \in [0,\,1), \tag{5}$$

  of X, see [14].

- The Conditional Value at Risk

$$\mathrm{CV@R}_X(u) := \frac{1}{u} \int_{0}^{u} q_X(s)\,ds, \quad u \in (0,\,1], \tag{6}$$

  see, e.g., [22, 23], also called the Average Value at Risk [11], or the expected shortfall, or the expected tail loss.

Again, these transforms characterize the distribution of X if either $\mathsf{E}[X^-] < \infty$ or $\mathsf{E}[X^+] < \infty$, otherwise, they are equal to $+\infty$ or $-\infty$ identically.

The main goal of this paper is a systematic exposition of basic properties of integrated distribution and quantile functions. In particular, we define the integrated distribution and quantile functions for any random variable X in such a way that each one of these functions determines uniquely the distribution of X. Further, we show that such important notions of probability theory as uniform integrability, weak convergence and tightness can be characterized in terms of integrated quantile functions (see Section 3). In Section 4 we show how some basic results of the theory of comparison of binary statistical experiments can be deduced using our results in previous two sections. Finally, in Section 5 we extend the area of application of the Chacon–Walsh construction in the Skorokhod embedding problem with the help of integrated quantile functions.

One of the key points of our approach is that we define integrated distribution and quantile functions as Fenchel conjugates of each other. This is due to the fact that their derivatives, distribution functions and quantile functions, are generalized inverses (see, e. g., [8, 11]). This convex duality result can be found in [21] and [11, Lemma A.26], and constitutes implicitly one of two main results in [22, 23].

Let us note that we consider only univariate distributions in this paper. However, it is reasonable to mention a possible generalization to the multidimensional case based on ideas from optimal transport. The integrated quantile function of a random variable X, as it is defined in our paper, is a convex function whose gradient pushes forward the uniform distribution on $(0, 1)$ into the distribution of X; moreover, the integrated distribution function is the Fenchel transform of the integrated quantile function and its gradient pushes forward the distribution of X into the uniform distribution on $(0, 1)$ if the distribution of X is continuous. It the multidimensional case the existence of such functions follows from the McCann theorem [18]. Namely, if μ is the distribution on $\mathbb{R}^d$, then there exists a (unique up to an additive constant) convex function V whose gradient pushes forward the uniform distribution on the unit cube (or, say, the unit ball) in $\mathbb{R}^d$ into μ. Additionally, if μ vanishes on Borel subsets of Hausdorff dimension $d - 1$, then the Fenchel transform V^* of V pushes forward μ to the corresponding uniform distribution. We refer to [3, 5, 9] and [13] for recent advances in this area.

It is more convenient for us to speak about random variables rather than distributions. However, if a probability space is not specified, the symbols P and E for probability and expectation enter into consideration only via distributions of random variables and may refer to different probability spaces. This allows us to replace occasionally random variables by their distributions in the notation.

For the reader's convenience, we recall some terminology and elementary facts concerning convex functions of one real variable. A convex function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is proper if its effective domain

$$\operatorname{dom} f := \left\{ x \in \mathbb{R} : f(x) < +\infty \right\}$$

is not empty. The subdifferential $\partial f(x)$ of f at a point x is defined by

$$\partial f(x) = \left\{ u \in \mathbb{R} : f(y) \geq f(x) + u(y - x) \text{ for every } y \in \mathbb{R} \right\}.$$

If f is a proper convex function and x is an interior point of dom f, then $\partial f(x) = [f'_-(x), f'_+(x)]$, where $f'_-(x)$ and $f'_+(x)$ are the left and right derivatives of f at x respectively. The conjugate of f, or the Fenchel transform, is the function f^* on $\mathbb{R}$ defined by

$$f^*(u) = \sup_{x \in \mathbb{R}} [xu - f(x)].$$

The conjugate function is lower semicontinuous and convex. The Fenchel–Moreau theorem says that if f is a proper lower semicontinuous convex function, then f is the conjugate of f^*, i.e.

$$f(x) = \sup_{u \in \mathbb{R}} [xu - f^*(u)], \quad x \in \mathbb{R};$$

moreover, for $x, u \in \mathbb{R}$,

$$u \in \partial f(x) \quad \Longleftrightarrow \quad x \in \partial f^*(u) \quad \Longleftrightarrow \quad f(x) + f^*(u) = xu.$$

2   Integrated distribution and quantile functions: definitions and main properties

2.1   Definition and properties of integrated distribution functions

The distribution function F_X of a random variable X given on a probability space $(\Omega, \mathscr{F}, \mathsf{P})$ is defined by $F_X(x) = \mathsf{P}(X \le x)$, $x \in \mathbb{R}$. Since F_X is bounded, for any choice of $x_0 \in \mathbb{R}$, the integral $\int_{x_0}^{x} F_X(t)\,dt$ is defined and finite for all $x \in \mathbb{R}$.[2] In contrast to this case, the function Ψ_X in (1) corresponding to the choice $x_0 = -\infty$, takes value $+\infty$ identically if $\mathsf{E}[X^-] = \infty$.

Definition 1. The *integrated distribution function* of a random variable X is defined by

$$\mathsf{J}_X(x) := \int_0^x F_X(t)\,dt, \quad x \in \mathbb{R}.$$

Theorem 1. *An integrated distribution function J_X has the following properties*:

 (i) $\mathsf{J}_X(0) = 0$.

 (ii) J_X *is convex, increasing and finite everywhere on* $\mathbb{R}$.

 (iii) *for $a < b$,*
$$\mathsf{J}_X(b) - \mathsf{J}_X(a) = \mathsf{E}\big[(b - X)^+ - (X - a)^-\big], \tag{7}$$
   in particular, for $x \in \mathbb{R}$,
$$\mathsf{J}_X(x) = \mathsf{E}\big[(x - X)^+ - X^-\big] = \mathsf{E}\big[(X - x)^+ - X^+ + x\big]. \tag{8}$$

 (iv) $\lim\limits_{x \to -\infty} \mathsf{J}_X(x) = -\mathsf{E}[X^-]$ *and* $\lim\limits_{x \to +\infty} (x - \mathsf{J}_X(x)) = \mathsf{E}[X^+]$.

---

[2]Throughout the paper, if $b < a$, by convention $\int_a^b f(x)\,dx := -\int_b^a f(x)\,dx$.

(v) $\lim_{x \to -\infty} \frac{J_X(x)}{x} = 0$ *and* $\lim_{x \to +\infty} \frac{J_X(x)}{x} = 1$.

(vi) *The subdifferential of* J_X *satisfies*

$$\partial J_X(x) = \left[F_X(x-0), \ F_X(x) \right], \quad x \in \mathbb{R}, \tag{9}$$

in particular, $(J_X)'_-(x) = F_X(x-0)$ *and* $(J_X)'_+(x) = F_X(x)$.

(vii) $J_{-X}(x) = x + J_X(-x)$ *for all* $x \in \mathbb{R}$.

It is clear from (vi), that the integrated distribution function uniquely determines the distribution.

Proof. It is evident that (i) holds and J_X is finite and increasing. For $a < b$, we have

$$F_X(a)(b-a) \le J_X(b) - J_X(a) = \int_a^b F_X(t)\, dt \le F_X(b-0)(b-a). \tag{10}$$

It follows that, for any $x, \ y \in \mathbb{R}$,

$$J_X(y) \ge J_X(x) + p(y - x),$$

if $p \in [F_X(x-0), \ F_X(x)]$. Now the convexity of J_X follows, which, in turn, implies (vi).

Next, by Fubini's theorem, for $a < b$,

$$\int_a^b F_X(t)\, dt = \int_a^b \mathsf{E}[\mathbb{1}_{\{X \le t\}}]\, dt = \mathsf{E}\left[\int_a^b \mathbb{1}_{\{X \le t\}}\, dt \right]$$
$$= \mathsf{E}\left[(b - X)^+ - (X - a)^- \right].$$

Thus, we have proved (7). The second equality in (8) is trivial, and the first one follows from (7) if we put $a = 0$ or $b = 0$ depending on the sign of x.

Let us prove (iv). The function $(x - X)^+ - X^-$ is increasing in x, hence $\mathsf{E}[(x - X)^+ - X^-] \to -\mathsf{E}[X^-]$ as $x \to -\infty$ by the monotone convergence theorem. This proves the first equality in (iv). Similarly, $x - (x - X)^+ + X^-$ is increasing in x, hence $\mathsf{E}[x - (x - X)^+ + X^-] \to \mathsf{E}[X^+]$ as $x \to +\infty$ by the monotone convergence theorem.
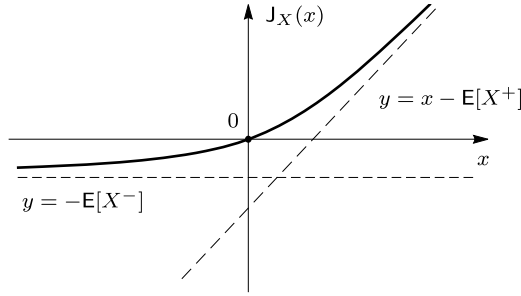
Finally, (v) and (vii) follow from (10) and (8) respectively. $\qquad\square$

Corollary 1. *If X is an integrable random variable, then, for any $x \in \mathbb{R}$,*

$$\Psi_X(x) = J_X(x) + \mathsf{E}[X^-],$$
$$H_X(x) = J_X(x) + \mathsf{E}[X^+] - x,$$
$$U_X(x) = x - \mathsf{E}|X| - 2J_X(x),$$

where Ψ_X, H_X, and U_X are defined in (1)–(3), *in particular,*

$$\Psi_X(x) + H_X(x) = -U_X(x).$$

Fig. 1. A typical graph of an integrated distribution function if the expectations $\mathsf{E}[X^-]$ and $\mathsf{E}[X^+]$ are finite

Theorem 2. *If $J : \mathbb{R} \to \mathbb{R}$, $J(0) = 0$, is a convex function satisfying*

$$\lim_{x \to -\infty} \frac{J(x)}{x} = 0 \quad and \quad \lim_{x \to +\infty} \frac{J(x)}{x} = 1,$$

then there exists on some probability space a random variable X for which $\mathsf{J}_X = J$.

Proof. Since J is convex and finite everywhere on the line, it has the right-hand derivative at each point, and $J(x) = \int_0^x J'_+(t)\, dt$. Moreover, similarly to the proof of (v) in Theorem 1, $\lim_{x \to -\infty} J'_+(x) = \lim_{x \to -\infty} \frac{J(x)}{x} = 0$ and $\lim_{x \to +\infty} J'_+(x) = \lim_{x \to +\infty} \frac{J(x)}{x} = 1$. Put $F(x) := J'_+(x)$. Due to convexity of J, F is an increasing and right-continuous function. So we can conclude, that F is the distribution function of some random variable X and $\mathsf{J}_X = J$. $\quad\square$

2.2 Definition and properties of integrated quantile functions

We call every function $q_X \colon (0,\ 1) \to \mathbb{R}$ satisfying

$$F_X\big(q_X(u) - 0\big) \le u \le F_X\big(q_X(u)\big), \quad u \in (0,\ 1),$$

a *quantile function* of a random variable X. The functions q_X^L and q_X^R defined by

$$q_X^L(u) := \inf\{x \in \mathbb{R} \colon F_X(x) \ge u\},$$
$$q_X^R(u) := \inf\{x \in \mathbb{R} \colon F_X(x) > u\},$$

are called the *lower (left)* and *upper (right)* quantile functions of X. Of course, the lower and upper quantile functions of X are quantile functions of X, and we always have

$$q_X^L(u) \le q_X(u) \le q_X^R(u), \quad u \in (0,\ 1),$$

for any quantile function q_X.

It follows directly from the definitions that, for any $x \in \mathbb{R}$ and $u \in (0,\ 1)$,

$$q_X^L(u) \le x \quad \Leftrightarrow \quad u \le F_X(x), \tag{11}$$
$$q_X^R(u) \ge x \quad \Leftrightarrow \quad u \ge F_X(x - 0). \tag{12}$$

See, e. g., [8, 11] for more information on quantile functions (generalized inverses).

Definition 2. The Fenchel transform of the integrated distribution function of a random variable X

$$\mathsf{K}_X(u) = \sup_{x \in \mathbb{R}} \{xu - \mathsf{J}_X(x)\}, \quad u \in \mathbb{R}, \tag{13}$$

is called the *integrated quantile function* of X.

This definition is motivated by the fact mentioned in the introduction, that a function whose derivative is a quantile function must coincide with the Fenchel transform of J_X up to an additive constant. The next theorem clarifies this point.

Theorem 3. *An integrated quantile function K_X has the following properties* :

(i) *The function K_X is convex and lower semicontinuous. It takes finite values on $(0, 1)$ and equals $+\infty$ outside $[0, 1]$.*

(ii) *The Fenchel transform of K_X is J_X, i. e. for any $x \in \mathbb{R}$,*

$$\mathsf{J}_X(x) = \sup_{u \in \mathbb{R}} \{xu - \mathsf{K}_X(u)\}. \tag{14}$$

(iii) $\min_{u \in \mathbb{R}} \mathsf{K}_X(u) = 0$, $\{u \in \mathbb{R} : \mathsf{K}_X(u) = 0\} = [F_X(0 - 0), \ F_X(0)]$.

(iv) *for every $u \in [0, 1]$,*

$$\mathsf{K}_X(u) = \int_{u_0}^{u} q_X(s) \, ds, \tag{15}$$

where u_0 is any zero of K_X.

(v) $\mathsf{K}_X(0) = \mathsf{E}[X^-]$ *and* $\mathsf{K}_X(1) = \mathsf{E}[X^+]$.

(vi) *The subdifferential of K_X satisfies*

$$\partial \mathsf{K}_X(u) = \left[q_X^L(u), \ q_X^R(u)\right], \quad u \in (0, 1), \tag{16}$$

in particular, $(\mathsf{K}_X)'_-(u) = q_X^L(u)$ and $(\mathsf{K}_X)'_+(x) = q_X^R(u)$.

(vii) $\mathsf{K}_{-X}(u) = \mathsf{K}_X(1 - u)$ *for all $u \in [0, 1]$.*

It is clear from (ii) and the similar remark after Theorem 1 that the integrated quantile function uniquely determines the distribution.

Proof. Since J_X is a proper convex continuous function, it follows from the definition of K_X and the Fenchel–Moreau theorem that K_X is convex and lower semicontinuous, (14) holds, and for all $x, \ u \in \mathbb{R}$

$$u \in \partial \mathsf{J}_X(x) \quad \Leftrightarrow \quad x \in \partial \mathsf{K}_X(u) \quad \Leftrightarrow \quad F_X(x - 0) \leq u \leq F_X(x), \tag{17}$$

where the last equivalence follows from (9). In particular,

$$\partial \mathsf{K}_X(u) = \varnothing, \quad \text{if } u \notin [0, 1],$$

and, for $u \in (0, 1)$,

$$x \in \partial \mathsf{K}_X(u) \quad \Leftrightarrow \quad q_X^L(u) \leq x \leq q_X^R(u),$$

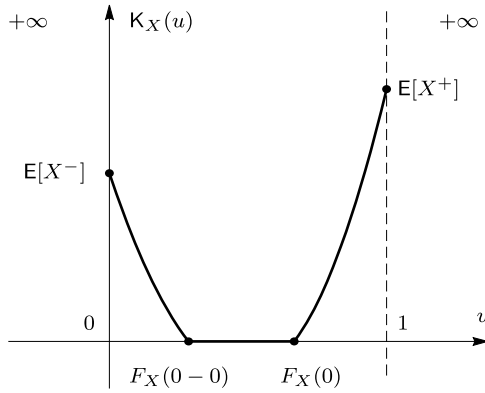due to (11) and (12). Thus, we have proved (i), (ii) and (vi).

Fig. 2. A typical graph of an integrated quantile function

Putting $x = 0$ in (14) and (17), we get $\inf_{u \in \mathbb{R}} \mathsf{K}_X(u) = 0$ and this infimum is attained at u if and only if $u \in [F_X(0 - 0), \ F_X(0)]$. This constitutes assertion (iii). Now (iv) follows from preceding statements.

Statement (v) follows from the definition of K_X and Theorem 1 (iv):

$$\mathsf{K}_X(0) = - \inf_{x \in \mathbb{R}} \mathsf{J}_X(x) = - \lim_{x \to -\infty} \mathsf{J}_X(x) = \mathsf{E}[X^-],$$
$$\mathsf{K}_X(1) = \sup_{x \in \mathbb{R}} (x - \mathsf{J}_X(x)) = \lim_{x \to +\infty} (x - \mathsf{J}_X(x)) = \mathsf{E}[X^+].$$

Finally, (vii) follows from the definition of K_X and Theorem 1 (vii). $\qquad\square$

Corollary 2. *For any random variable X,*

$$\mathsf{K}_X(F_X(x)) = xF_X(x) - \mathsf{J}_X(x), \quad x \in \mathbb{R}.$$

Proof. Put $u := F_X(x)$ and $g(y) := \mathsf{J}_X(y) - yu, \ y \in \mathbb{R}$. According to (9), $\partial g(y) = [F_X(y - 0) - u, \ F_X(y) - u]$, in particular, $0 \in \partial g(y)$ if $y = x$. This means that the function g attains its minimum at x and, hence, we have $\mathsf{K}_X(u) = \sup_{y \in \mathbb{R}} \{yu - \mathsf{J}_X(y)\} = xu - \mathsf{J}_X(x)$. $\qquad\square$

Theorem 4. *If a convex lower semicontinuous function $K : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ satisfies*

$$(0, \ 1) \subseteq \operatorname{dom} K \subseteq [0, \ 1]$$

and there is $u_0 \in [0, \ 1]$ such that $K(u_0) = 0$, then there exists on some probability space a random variable X for which $\mathsf{K}_X = K$.

Proof. Under our assumptions

$$K(u) = \int_{u_0}^{u} q(s)\,ds, \quad u \in [0, \ 1],$$

where $q(u) = K'_-(u), \ u \in (0, \ 1)$, is increasing and left continuous. Let us define a probability space $(\Omega, \ \mathscr{F}, \mathsf{P})$ as follows: $\Omega = (0, \ 1)$, $\mathscr{F}$ is the Borel σ-field and P is

the Lebesgue measure. Put $X(\omega) := q(\omega)$. Now if $G(x) := \inf\{u \in (0,\,1): q(u) > x\}$, then it is easy to verify that $q(u) \le x \Leftrightarrow G(x) \ge u$, cf. (11). It follows that G is the distribution function of X and, hence, $q = q_X^L$ on $(0,\,1)$. This means that the left-hand derivative of K and K_X coincide on $(0,\,1)$. In addition, their minimums over this interval are equal to zero. Therefore, $K = K_X$ on $(0, 1)$ and, hence, everywhere on $\mathbb{R}$. $\qquad\square$

Remark 1. An alternative way to prove Theorem 4 is to introduce the Fenchel transform J of K and to show that J satisfies the assumptions of Theorem 2. However, our proof yields not only a characterization statement of Theorem 4 but also an explicit representation of a random variable with a given integrated quantile function. Of course, this representation (namely, of a random variable with given distribution as its quantile function with respect to the Lebesgue measure on $(0,\,1)$) is well known.

It is convenient to introduce *shifted integrated quantile functions*:

$$K_X^{[0]}(u) := K_X(u) - K_X(0), \quad u \in [0,\,1], \qquad \text{if } K_X(0) = \mathsf{E}[X^-] < \infty,$$
$$K_X^{[1]}(u) := K_X(u) - K_X(1), \quad u \in [0,\,1], \qquad \text{if } K_X(1) = \mathsf{E}[X^+] < \infty.$$

Now we can express the functions defined in (4)–(6) in terms of shifted integrated quantile functions. If $\mathsf{E}[X^-] < \infty$, then the absolute Lorenz curve coincides with $K_X^{[0]}$:

$$\mathrm{AL}_X(u) = K_X^{[0]}(u), \quad u \in [0,\,1].$$

Since Ψ_X is obtained from J_X by adding a constant $\mathsf{E}[X^-] = K_X(0)$ by Corollary 1, the absolute Lorenz curve is the Fenchel transform of Ψ_X:

$$\mathrm{AL}_X(u) = \sup_{x \in \mathbb{R}}\{xu - \Psi_X(x)\}, \quad u \in [0,\,1].$$

The Conditional Value at Risk satisfies

$$\mathrm{CV@R}_X(u) = \frac{1}{u}K_X^{[0]}(u) = \sup_{x \in \mathbb{R}}\{x - \Psi_X(x)/u\}, \quad u \in (0,\,1].$$

The Hardy–Littlewood maximal function satisfies

$$\mathrm{HL}_X(u) = \frac{1}{u-1}K_X^{[1]}(u) = \frac{1}{u-1}K_{-X}^{[0]}(1-u) = -\,\mathrm{CV@R}_{-X}(1-u), \quad u \in [0,\,1).$$

2.3 Convex orders

Let us recall the definitions of convex orders in the univariate case.

For an arbitrary function $\psi \colon \mathbb{R} \to \mathbb{R}_+$, define C_ψ as the space of all continuous functions $f \colon \mathbb{R} \to \mathbb{R}$ such that

$$\sup_{x \in \mathbb{R}} \frac{|f(x)|}{1 + \psi(x)} < \infty.$$

Let X and Y be random variables. We say that

- *X is less than Y in convex order ($X \leq_{cx} Y$) if $E|X| < \infty$, $E|Y| < \infty$, and $E[\varphi(X)] \leq E[\varphi(Y)]$ for any convex function $\varphi \in C_{|x|}$;*

- *X is less than Y in increasing convex order ($X \leq_{icx} Y$) if $E[X^+] < \infty$, $E[Y^+] < \infty$, and $E[\varphi(X)] \leq E[\varphi(Y)]$ for any increasing convex function $\varphi \in C_{x^+}$;*

- *X is less than Y in decreasing convex order ($X \leq_{decx} Y$) if $E[X^-] < \infty$, $E[Y^-] < \infty$, and $E[\varphi(X)] \leq E[\varphi(Y)]$ for any decreasing convex function $\varphi \in C_{x^-}$.*

It is trivial that $X \leq_{icx} Y$ if and only if $-X \leq_{decx} -Y$. Also it is easy to see that $X \leq_{cx} Y$ if and only if $X \leq_{icx} Y$ and $X \leq_{decx} Y$.

The following theorem is well known. We provide its proof which reduces to the duality between integrated distribution and quantile functions.

Theorem 5. *Let X and Y be random variables.*

(i) *If $E|X| < \infty$, $E|Y| < \infty$, then the following statements are equivalent:*

    (a) $X \leq_{cx} Y$;

    (b) $K_X^{[1]}(u) \geq K_Y^{[1]}(u)$ *for all $u \in [0, 1]$ and $K_X^{[1]}(0) = K_Y^{[1]}(0)$;*

    (c) $K_X^{[0]}(u) \geq K_Y^{[0]}(u)$ *for all $u \in [0, 1]$ and $K_X^{[0]}(1) = K_Y^{[0]}(1)$;*

    (d) $K_X^{[1]}(u) \geq K_Y^{[1]}(u)$ *and $K_X^{[0]}(u) \geq K_Y^{[0]}(u)$ for all $u \in [0, 1]$.*

(ii) *$X \leq_{icx} Y$ if and only if $E[X^+] < \infty$, $E[Y^+] < \infty$, and $K_X^{[1]}(u) \geq K_Y^{[1]}(u)$ for all $u \in [0, 1]$;*

(iii) *$X \leq_{decx} Y$ if and only if $E[X^-] < \infty$, $E[Y^-] < \infty$, and $K_X^{[0]}(u) \geq K_Y^{[0]}(u)$ for all $u \in [0, 1]$.*

Proof. First, let us prove (ii). It is well known (see, e. g., [24]) that $X \leq_{icx} Y$ if and only if $E[X^+] < \infty$, $E[Y^+] < \infty$, and $E[(X - x)^+] \leq E[(Y - x)^+]$ for all $x \in \mathbb{R}$. Taking (8) into account, the last condition can be rewritten as

$$J_X(x) + E[X^+] \leq J_Y(x) + E[Y^+], \quad x \in \mathbb{R},$$

which in turn, is equivalent to

$$K_X(u) - E[X^+] \geq K_Y(u) - E[Y^+], \quad u \in [0, 1],$$

by the definition of integrated quantile function. The claim follows.

Now, (iii) follows from (ii) and the first part of the remark before Theorem 5. Now, the second part of this remark shows equivalence (a) ⇔ (d) in (i).

Next, the equalities in (b) and (c) are both equivalent to $E[X] = E[Y]$. On the other hand, the inequalities in (d) reduce to $-E[X] \geq -E[Y]$ and $E[X] \geq E[Y]$ for $u = 0$ and $u = 1$ respectively. It follows that (d) implies (b) and (c). Finally, it is straightforward to check that (b) and (c) are equivalent and, hence, imply (d). □

Further properties of convex orders see, e.g., in [20, 24].
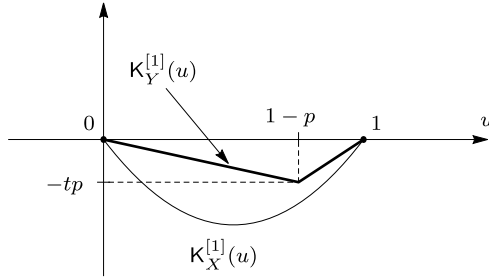
Fig. 3. Graphs of shifted integrated quantile functions $\mathsf{K}_X^{[1]}$ and $\mathsf{K}_Y^{[1]}$ in Example 1

2.4 Examples

In this subsection we demonstrate how the developed techniques can be used to derive two elementary well-known inequalities, see [10, p. 152]. This approach allows us to find the distributions at which the corresponding extrema are attained. So the inequalities obtained in this way are sharp.

Example 1. Let X be a random variable with zero mean and finite variance $\mathsf{D}(X) = \sigma^2$. It is required to find a *sharp* upper bound for the probability $\mathsf{P}(X \geq t)$, where t is a fixed positive number.

We solve a converse problem. Namely, let $p := \mathsf{P}(X \geq t)$ be fixed. Our purpose is to find a sharp lower bound for variances $\mathsf{D}(X) = \mathsf{E}[X^2]$ over all random variables X such that $\mathsf{E}[X] = 0$, and $\mathsf{P}(X \geq t) = p$.
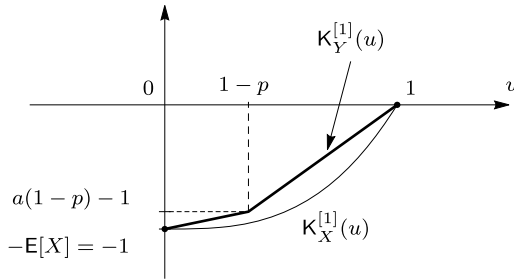
The above class of distributions has a minimal element with respect to the convex order. Indeed, let Y be a discrete random variable with $\mathsf{P}(Y = t) = p$ and $\mathsf{P}(Y = -\frac{tp}{1-p}) = 1 - p$. It is clear that $\mathsf{E}[Y] = 0$ and $\mathsf{P}(Y \geq t) = p$. If X is another random variable with these properties, then $\mathsf{K}_X^{[1]}(u) \leq \mathsf{K}_Y^{[1]}(u)$ for all $u \in [0, 1]$. Indeed, $\mathsf{K}_X^{[1]}(0) = \mathsf{K}_Y^{[1]}(0) = 0$ and the graph of $\mathsf{K}_Y^{[1]}$ consists of two straight segments, see Fig. 3. Since $\mathsf{P}(X \geq t) = p$, $q_X^R(u) \geq t$ for $u \in [1 - p, 1]$. In particular,

$$\mathsf{K}_X^{[1]}(1 - p) = -\int_{1-p}^1 q_X^R(s)\,ds \leq -pt = \mathsf{K}_Y^{[1]}(1 - p).$$

Due to convexity of integrated quantile functions, this implies $\mathsf{K}_X^{[1]}(u) \leq \mathsf{K}_Y^{[1]}(u)$ for all $u \in [0, 1]$, see Fig. 3. Hence, $X \geq_{cx} Y$ by Theorem 5 (i). Therefore, $\mathsf{E}[f(X)] \geq \mathsf{E}[f(Y)]$ for any convex function f. In particular, $\sigma^2 = \mathsf{E}[X^2] \geq \mathsf{E}[Y^2] = \frac{t^2 p}{1-p}$. Resolving this inequality with respect to $p = \mathsf{P}(X \geq t)$, we obtain the required upper bound

$$\mathsf{P}(X \geq t) \leq \frac{\sigma^2}{\sigma^2 + t^2}. \tag{18}$$

To show that the estimate in (18) is sharp it is enough to put $p = \frac{\sigma^2}{\sigma^2 + t^2}$ in the definition of a random variable Y and to check that $\mathsf{E}[Y^2] = \sigma^2$ and, for $X = Y$, the equality holds in (18).

Fig. 4. Graphs of shifted integrated quantile functions $K_X^{[1]}$ and $K_Y^{[1]}$ in Example 2

Example 2. Let X be a strictly positive random variable, i.e. $F_X(0) = 0$, such that $E[X] = 1$ and $E[X^2] = b$. It is required to find a *sharp* lower bound for the probability $P(X > a)$, where $a \in (0, 1)$ is fixed.

We will proceed in the similar way as in the previous example. Namely, let $p := P(X > a)$ be fixed. Our purpose is to find a sharp lower bound for the second moment $E[X^2]$ over all random variables X such that $F_X(0) = 0$, $E[X] = 1$ and $P(X > a) = p$.

The above class of distributions has a minimal element with respect to the convex order. Indeed, let Y be a random variable such that $P(Y = a) = 1 - p$ and $P(Y = \frac{1 - a(1-p)}{p}) = p$. It is obvious that $F_Y(0) = 0$, $E[Y] = 1$, and $P(Y > a) = p$. If X is another random variable with these properties, then $K_X^{[1]}(u) \le K_Y^{[1]}(u)$ for all $u \in [0, 1]$. Indeed, $K_X^{[1]}(0) = K_Y^{[1]}(0) = -1$ and the graph of $K_Y^{[1]}$ consists of two straight segments, see Fig. 4. Since $P(X \le a) = 1 - p$, $q_X^L(u) \le a$ for $u \in [0, 1-p]$. In particular,

$$K_X^{[1]}(1 - p) = \int_0^{1-p} q_X^L(s)\, ds - E[X] \le a(1 - p) - 1 = K_Y^{[1]}(1 - p).$$

Due to convexity of integrated quantile functions, this implies $K_X^{[1]}(u) \le K_Y^{[1]}(u)$ for all $u \in [0, 1]$, see Fig. 4. Hence, $X \ge_{cx} Y$ by Theorem 5 (i). Therefore, $E[f(X)] \ge E[f(Y)]$ for any convex function f. In particular, $b = E[X^2] \ge E[Y^2] = a^2(1-p) + \frac{(1-a(1-p))^2}{p^2} p$. Resolving this inequality with respect to $p = P(X > a)$, we obtain the required lower bound

$$P(X > a) \ge \frac{(1 - a)^2}{b - a(2 - a)}. \tag{19}$$

The sharpness of the estimate in (19) follows if we put $p = \frac{(1-a)^2}{b-a(2-a)}$ in the definition of a random variable Y and verify that $E[Y^2] = b$ and, for $X = Y$, the equality holds in (19). Remark that replacing the right-hand side in (19) by a smaller quantity $\frac{(1-a)^2}{b}$, we arrive at the inequality (7.6) in [10, p. 152].

3 Uniform integrability and weak convergence

3.1 Tightness and uniform integrability

In this subsection we study conditions for tightness and uniform integrability of a family of random variables in terms of integrated quantile function. It is a natural question because both tightness and uniform integrability are characterized in terms of one-dimensional distributions of these variables.

Theorem 6. *Let (X_α) be a family of random variables. Then the following statements are equivalent*:

  (i) *The family of distributions $\{\mathrm{Law}(X_\alpha)\}$ is tight.*

  (ii) *For every $u,\ v \in (0,\ 1)$, $\sup_\alpha |\mathsf{K}_{X_\alpha}(u) - \mathsf{K}_{X_\alpha}(v)| < \infty$.*

  (iii) *The family of functions $\{\mathsf{K}_{X_\alpha}\}$ is pointwise bounded on $(0,\ 1)$.*

  (iv) *The family of functions $\{\mathsf{K}_{X_\alpha}\}$ is equicontinuous on every $[a,\ b] \subset (0,\ 1)$.*

Proof. (i) $\Rightarrow$ (iv) Let $[a,\ b] \subset (0,\ 1)$. The tightness condition implies that there is $C > 0$ such that $F_{X_\alpha}(-C) < a$ and $F_{X_\alpha}(C) \geq b$ for all α. Hence, $-C < q^L_{X_\alpha}(u) \leq C$ for all α and $u \in [a,\ b]$. Thus, the functions K_{X_α} are even uniformly Lipschitz continuous on $[a,\ b]$.

(ii) $\Rightarrow$ (iii) Let $c \in (0,\ 1)$. Take $a \in (0,\ c)$ and $b \in (c,\ 1)$. By the assumption, there is $L > 0$ such that $\mathsf{K}_{X_\alpha}(a) - \mathsf{K}_{X_\alpha}(c) \leq L$ and $\mathsf{K}_{X_\alpha}(b) - \mathsf{K}_{X_\alpha}(c) \leq L$ for all α. Let $u_{0,\alpha}$ be a point, where $\mathsf{K}_{X_\alpha}(u_{0,\alpha}) = 0$. If α is such that $u_{0,\alpha} < c$, then, by the three chord inequality,

$$\frac{\mathsf{K}_{X_\alpha}(c)}{c - u_{0,\alpha}} \leq \frac{\mathsf{K}_{X_\alpha}(b) - \mathsf{K}_{X_\alpha}(c)}{b - c},$$

therefore, $\mathsf{K}_{X_\alpha}(c) \leq cL/(b-c)$. Similarly, if α is such that $u_{0,\alpha} > c$, then $\mathsf{K}_{X_\alpha}(c) \leq (1-c)L/(c-a)$.

(ii) $\Rightarrow$ (i) Let $\varepsilon > 0$. By the assumption, there is $L > 0$ such that $\mathsf{K}_{X_\alpha}(\varepsilon) - \mathsf{K}_{X_\alpha}(\varepsilon/2) > -L$ and $\mathsf{K}_{X_\alpha}(1 - \varepsilon/2) - \mathsf{K}_{X_\alpha}(1 - \varepsilon) \leq L$ for all α. The first inequality yields

$$-L < \mathsf{K}_{X_\alpha}(\varepsilon) - \mathsf{K}_{X_\alpha}(\varepsilon/2) = \int_{\varepsilon/2}^{\varepsilon} q^L_{X_\alpha}(s)\,ds \leq \frac{\varepsilon}{2} q^L_{X_\alpha}(\varepsilon),$$

which shows that $-\frac{2L}{\varepsilon} < q^L_{X_\alpha}(\varepsilon)$ and, hence, $F_{X_\alpha}(-\frac{2L}{\varepsilon}) < \varepsilon$. Similarly, from the second inequality, one gets $F_{X_\alpha}(\frac{2L}{\varepsilon}) \geq 1 - \varepsilon$ for all α. This proves the tightness of the laws of X_α.

Since implications (iv) $\Rightarrow$ (ii) and (iii) $\Rightarrow$ (ii) are obvious, the claim follows. $\square$

Theorem 7. *Let $\{X_\alpha\}$ be a family of random variables. Then the following statements are equivalent*:

  (i) *The family $\{X_\alpha\}$ is uniformly integrable.*

  (ii) *The family of integrated quantile functions $\{\mathsf{K}_{X_\alpha}\}$ is equicontinuous on $[0,\ 1]$.*

  (iii) *The family of integrated quantile functions $\{\mathsf{K}_{X_\alpha}\}$ is relatively compact in the space $C[0,\ 1]$ of continuous functions with supremum norm.*

Proof. Let us consider the probability space $(\Omega, \mathscr{F}, \mathsf{P})$ as in the proof of Theorem 4 and define random variables $Y_\alpha(\omega) = q^L_{X_\alpha}(\omega)$. Then $X_\alpha \stackrel{d}{=} Y_\alpha$ and it is enough to study the uniform integrability of the family $\{Y_\alpha\}$. Without loss of generality, we suppose that $X_\alpha = Y_\alpha$.

Let us recall that a family $\{X_\alpha\}$ is uniformly integrable if and only if $\mathsf{E}|X_\alpha|$ are bounded and $\mathsf{E}[|X_\alpha|\mathbb{1}_A]$ are uniformly continuous, i. e. $\sup_\alpha \mathsf{E}[|X_\alpha|\mathbb{1}_A] \to 0$ as $\mathsf{P}(A) \to 0$; moreover, the boundedness of $\mathsf{E}|X_\alpha|$ is a consequence of the uniform continuity if the measure P has no atomic part, in particular, in our case. On the other hand, by the Arzela–Ascoli theorem a set in $C[0, 1]$ is relatively compact if and only if it is uniformly bounded and equicontinuous.

We shall check that the uniform boundedness and the equicontinuity of $\{\mathsf{K}_{X_\alpha}\}$ are equivalent to uniform boundedness of $\mathsf{E}|X_\alpha|$ and the uniform continuity of $\mathsf{E}[|X_\alpha|\mathbb{1}_A]$, respectively. In view of the above this is sufficient for the proof of the theorem.

By the properties of integrated quantile functions,

$$\sup_{u\in[0,\,1]} \mathsf{K}_{X_\alpha}(u) = \max\left(\mathsf{E}\left[X_\alpha^-\right], \mathsf{E}\left[X_\alpha^+\right]\right) \leq \mathsf{E}|X_\alpha| \leq 2 \sup_{u\in[0,\,1]} \mathsf{K}_{X_\alpha}(u).$$

Hence, $\sup_\alpha \mathsf{E}|X_\alpha| < \infty$ if and only if the family $\{\mathsf{K}_{X_\alpha}\}$ is uniformly bounded.

For a fixed $\varepsilon > 0$, let $\delta > 0$ be such that $\sup_\alpha \mathsf{E}[|X_\alpha|\mathbb{1}_A] < \varepsilon$ for any Borel set $A \subseteq (0, 1)$ with $\mathsf{P}(A) < \delta$. Let $u_1, u_2 \in [0, 1]$ satisfy $0 < u_2 - u_1 < \delta$. Then, for any α,

$$\left|\mathsf{K}_{X_\alpha}(u_2) - \mathsf{K}_{X_\alpha}(u_1)\right| = \left|\int_{u_1}^{u_2} X_\alpha(\omega)\,d\omega\right| \leq \int_{u_1}^{u_2} \left|X_\alpha(\omega)\right|d\omega < \varepsilon.$$

Conversely, fix $\varepsilon > 0$ and let $\delta > 0$ be such that $|\mathsf{K}_{X_\alpha}(u_2) - \mathsf{K}_{X_\alpha}(u_1)| < \varepsilon$ for all α if $|u_2 - u_1| < \delta$. Since $X_\alpha(\omega)$ is increasing in ω, the following inequality holds for any Borel subset $A \subseteq (0, 1)$:

$$\int_{(0,\,\mathsf{P}(A)]} X_\alpha(\omega)\,d\omega \leq \int_A X_\alpha(\omega)\,d\omega \leq \int_{[1-\mathsf{P}(A),\,1)} X_\alpha(\omega)\,d\omega. \qquad (20)$$

Therefore, if $\mathsf{P}(A) < \delta$ then

$$\begin{aligned}
\mathsf{E}\left[|X_\alpha|\mathbb{1}_A\right] &= \int_{A\cap\{X_\alpha<0\}} -X_\alpha(\omega)\,d\omega + \int_{A\cap\{X_\alpha>0\}} X_\alpha(\omega)\,d\omega \\
&\leq \int_{(0,\,\mathsf{P}(A\cap\{X_\alpha<0\})]} -X_\alpha(\omega)\,d\omega + \int_{[1-\mathsf{P}(A\cap\{X_\alpha>0\}),\,1)} X_\alpha(\omega)\,d\omega \\
&\leq \max_{u\in(0,\mathsf{P}(A))}\left(\mathsf{K}_{X_\alpha}(0) - \mathsf{K}_{X_\alpha}(u)\right) + \max_{u\in[1-\mathsf{P}(A),1)}\left(\mathsf{K}_{X_\alpha}(1) - \mathsf{K}_{X_\alpha}(u)\right) \\
&< 2\varepsilon. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square
\end{aligned}$$

The following criterion of uniform integrability is proved in [17].

Theorem 8 (Leskelä and Vihola). *A family $\{X_\alpha\}$ of integrable random variables is uniformly integrable if and only if there is an integrable random variable X such that $|X_\alpha| \leq_{icx} X$ for all α.*

Proof. Without loss of generality, we may assume that all X_α are nonnegative. To simplify notation, let $K_\alpha(u) := \mathsf{K}^{[1]}_{X_\alpha}(u)$, $u \in [0, 1]$. Then K_α are increasing continuous convex functions with $K_\alpha(1) = 0$. According to Theorems 4, 5 (ii) and 7, it is enough to prove that the family $\{K_\alpha\}$ is equicontinuous if and only if there is an increasing continuous convex function $K(u)$, $u \in [0, 1]$, with $K(1) = 0$ such that

$$K_\alpha(u) \geq K(u) \quad \text{for all } u \in [0, 1] \text{ for all } \alpha.$$

The sufficiency is evident. Indeed, if $0 \leq u_1 \leq u_2 \leq 1$, then, for all α,

$$0 \leq K_\alpha(u_2) - K_\alpha(u_1) \leq K_\alpha(1) - K_\alpha\big(1 - (u_2 - u_1)\big) = -K_\alpha\big(1 - (u_2 - u_1)\big)$$
$$\leq -K\big(1 - (u_2 - u_1)\big) = K(1) - K\big(1 - (u_2 - u_1)\big),$$

and the equicontinuity follows from the continuity of K.

Let us define K as the lower semicontinuous convex envelope of $\inf_\alpha K_\alpha$. To prove the necessity, it is enough to show that $K(1) = 0$ if the family $\{K_\alpha\}$ is equicontinuous. Fix $\varepsilon > 0$ and let $\delta > 0$ be such that $|K_\alpha(u_2) - K_\alpha(u_1)| < \varepsilon$ for all α if $|u_2 - u_1| \leq \delta$. In particular, $K_\alpha(1 - \delta) > -\varepsilon$. Since K_α is convex, we have $K_\alpha(u) > -\frac{\varepsilon}{\delta}(1 - u)$ for all $u \in [0, 1 - \delta]$ and for all α. Moreover, since K_α is increasing, $K_\alpha(u) \geq -\varepsilon$ for all $u \in [1 - \delta, 1]$ and for all α. Combining, we get

$$\inf_\alpha K_\alpha(u) \geq \min\left(-\frac{\varepsilon}{\delta}(1 - u), -\varepsilon\right) \geq -\frac{\varepsilon}{\delta} + \frac{\varepsilon(1 - \delta)}{\delta}u,$$

for all $u \in [0, 1]$. It follows that $K(u) \geq -\frac{\varepsilon}{\delta} + \frac{\varepsilon(1-\delta)}{\delta}u$ for all $u \in [0, 1]$, in particular, $K(1 - \delta) > -2\varepsilon$. The claim follows. $\qquad\square$

3.2   Weak convergence

In this subsection (X_n) is a sequence of random variables.

Theorem 9. *The following statements are equivalent*:

(i) *The sequence (X_n) weakly converges.*

(ii) *There is a sequence (c_n) of numbers such that, for every $u \in (0, 1)$, the sequence $(\mathsf{K}_{X_n}(u) - c_n)$ converges to a finite limit.*

(iii) *The sequence (K_{X_n}) converges uniformly on every $[\alpha, \beta] \subseteq (0, 1)$.*

Moreover, in this case if X is a weak limit of (X_n), then $\mathsf{K}_X(u) = \lim_{n \to \infty} \mathsf{K}_{X_n}(u)$ for all $u \in (0, 1)$.

Remark 2. If $\mathsf{E}[X_n^-] < \infty$ (resp. $\mathsf{E}[X_n^+] < \infty$) for all n, then the pointwise convergence of $\mathsf{K}^{[0]}_{X_n}$ (resp. $\mathsf{K}^{[1]}_{X_n}$) on $(0, 1)$ is sufficient (use Theorem 9, (ii) $\Rightarrow$ (i)) but not necessary for the weak convergence of X_n.

Theorem 10. *Let (X_n) weakly converge and $\mathsf{E}|X_n| < \infty$ (resp. $\mathsf{E}[X_n^-] < \infty$, resp. $\mathsf{E}[X_n^+] < \infty$). Then the following statements are equivalent*:

(i) *The sequence $(|X_n|)$ (resp. (X_n^-), resp. (X_n^+)) is uniformly integrable.*

(ii) *The sequence of functions* (K_{X_n}) *converges pointwise on* $[0, 1]$ *(resp.* $[0, 1)$*, resp.* $(0, 1]$*) to a continuous function with finite values.*

(iii) *The sequence* (K_{X_n}) *converges uniformly to a finite-valued function on* $[0, 1]$ *(resp. on every* $[0, \beta] \subseteq [0, 1)$*, resp. on every* $[\alpha, 1] \subseteq (0, 1]$*).*

Remark 3. In contrast to Remark 2, a combination of the weak convergence of X_n and the uniform integrability of X_n^- (resp. X_n^+) can be expressed in terms of the shifted integrated quantile functions $\mathsf{K}_{X_n}^{[0]}$ (resp. $\mathsf{K}_{X_n}^{[1]}$). For instance, let a sequence (X_n) weakly converge to X and the sequence (X_n^+) is uniformly integrable. Then the pointwise limit of $\mathsf{K}_{X_n}^{[1]}(u)$ satisfies

$$\lim_{n\to\infty} \mathsf{K}_{X_n}^{[1]}(u) = \lim_{n\to\infty} \mathsf{K}_{X_n}(u) - \lim_{n\to\infty} \mathsf{E}\big[X_n^+\big]$$
$$= \mathsf{K}_X(u) - \mathsf{E}\big[X^+\big] = \mathsf{K}_X^{[1]}(u), \quad u \in (0, 1], \tag{21}$$

and is continuous on $(0, 1]$. Conversely, if the functions $\mathsf{K}_{X_n}^{[1]}$ converge pointwise to a continuous limit on $(0, 1]$, then X_n weakly converges, say, to X (use Theorem 9, (ii) $\Rightarrow$ (i)). In particular, for any $u \in (0, 1)$,

$$\lim_{n\to\infty} \mathsf{K}_{X_n}^{[1]}(u) = \lim_{n\to\infty} \mathsf{K}_{X_n}(u) - \lim_{n\to\infty} \mathsf{E}\big[X_n^+\big] = \mathsf{K}_X(u) - \lim_{n\to\infty} \mathsf{E}\big[X_n^+\big].$$

Continuity of the limiting function in the left-hand side of the above formula at $u = 1$ implies $\lim_{n\to\infty} \mathsf{E}[X_n^+] = \lim_{u\uparrow 1} \mathsf{K}_X(u) = \mathsf{E}[X^+]$.

Proof of Theorems 9 and 10. First, let us suppose that (X_n) weakly converges to X. It is well known that then $q_{X_n}^L(u) \to q_X^L(u)$ as $n \to \infty$ for every continuity point u of q_X^L. Put $u_{n,0} := F_{X_n}(0)$ and $u_0 := F_X(0)$.

Assume for the moment that X_n are uniformly bounded. Then, for any $u \in [0, 1]$,

$$\mathsf{K}_{X_n}(u) = \int_0^u q_{X_n}(s)\, ds - \int_0^{u_{n,0}} q_{X_n}(s)\, ds = \int_0^u q_{X_n}(s)\, ds + \int_0^1 \big(q_{X_n}(s)\big)^-\, ds$$
$$\to \int_0^u q_X(s)\, ds + \int_0^1 \big(q_X(s)\big)^-\, ds = \mathsf{K}_X(u)$$

by the dominated convergence theorem. Moreover, by Theorem 7 the sequence (K_{X_n}) is relatively compact in $C[0, 1]$. Combined with pointwise convergence, this shows that (K_{X_n}) converges to K_X uniformly on $[0, 1]$.

If no assumptions on X_n are imposed, let us introduce the function $g_C(x) := \max(\min(x, C), -C), C > 0$, and define random variables

$$Y_n := g_C(X_n), \qquad Y := g_C(X).$$

Then (Y_n) weakly converges to Y. Hence, $\mathsf{K}_{Y_n} \to \mathsf{K}_Y$ uniformly on $[0, 1]$ as it has just been proved. However, $\mathsf{K}_{Y_n} = \mathsf{K}_{X_n}$ on $[F_{X_n}(-C), F_{X_n}(C)] \ni u_{n,0}$ and $\mathsf{K}_Y = \mathsf{K}_X$ on $[F_X(-C), F_X(C)] \ni u_0$. Given $[\alpha, \beta] \subseteq (0, 1)$, choose $C > 0$ so that $[\alpha, \beta] \subseteq [F_X(-C), F_X(C)]$ and $[\alpha, \beta] \subseteq [F_{X_n}(-C), F_{X_n}(C)]$ for all n, which is possible

by tightness. Therefore, $\mathsf{K}_{X_n}(u) \to \mathsf{K}_X(u)$ uniformly in u on $[\alpha, \beta] \subseteq (0, 1)$. In particular, (K_{X_n}) converges pointwise to K_X on $(0, 1)$.

To complete the proof of Theorem 9 it remains to prove implication (ii) $\Rightarrow$ (i). Let $u, v \in (0, 1)$. By the assumption, the sequence $\mathsf{K}_{X_n}(u) - \mathsf{K}_{X_n}(v)$ converges to a finite limit and, hence, is bounded. By Theorem 6, the laws of X_n are tight. Let (X_{n_k}) be a weakly convergent subsequence. It follows from what has been proved that the integrated quantile function $K(u)$ of its limit coincides with $\lim_{k \to \infty} \mathsf{K}_{X_{n_k}}(u)$ for $u \in (0, 1)$. Therefore, for all $u \in (0, 1)$,

$$\lim_{n \to \infty} \left(\mathsf{K}_{X_n}(u) - c_n \right) = \lim_{k \to \infty} \left(\mathsf{K}_{X_{n_k}}(u) - c_{n_k} \right) = K(u) - \lim_{k \to \infty} c_{n_k}.$$

This implies that c_{n_k} converges to a finite limit and that $K(u)$ is obtained from $\lim_{n \to \infty} (\mathsf{K}_{X_n}(u) - c_n)$ by adding a constant. Since K is an integrated quantile function, this constant is determined uniquely. Thus, K is the same for all weakly convergent subsequences, which means that (X_n) weakly converges.

It is enough to prove Theorem 10 in one of three cases, for example, in the case $\mathsf{E}[X^-] < \infty$. Assume that (X_n^-) is uniformly integrable. Then $\mathsf{E}[X_n^-] \to \mathsf{E}[X^-]$, where X is a weak limit of (X_n). In other words, $\mathsf{K}_{X_n}(0) \to \mathsf{K}_X(0)$. Thus, we have (ii). Moreover, the sequence $(\mathsf{K}_{X_n^-})$ is equicontinuous. It follows that (K_{X_n}) converges uniformly on every segment $[0, \beta] \subseteq (0, 1)$. Implication (iii) $\Rightarrow$ (ii) is trivial. If (ii) holds, then $\lim_{n \to \infty} \mathsf{K}_{X_n}(u)$ is a continuous function in $u \in [0, 1)$. On the other hand, this limit is $\mathsf{K}_X(u)$ for $u \in (0, 1)$. Hence, $\mathsf{E}[X^-] = \mathsf{K}_X(0) = \lim_{n \to \infty} \mathsf{E}[X_n^-]$, and the sequence (X_n^-) is uniformly integrable. □

4 Applications to binary statistical models

The theory of statistical experiments deals with the problem of comparing the information in different experiments. The foundation of the theory of experiments was laid by Blackwell [1, 2], who first studied a notion of being more informative for experiments. Since it is difficult to give an explicit definition of statistical information, the theory of statistical experiments evaluates the performance of an experiment in terms of the set of available risk functions, in general, for arbitrary decision spaces and loss functions. For the theory of statistical experiments we refer to [15, 25], and especially to [27, 28], where the reader can find unexplained results and additional information.

In this paper we consider only binary statistical experiments, or dichotomies, $\mathbb{E} = (\Omega, \mathscr{F}, \mathsf{P}, \mathsf{P}')$. It is known that for binary models, it is enough to deal with testing problems, i. e. with tests as decision rules and with the probabilities of errors of the first and the second kinds of a test.

Let us introduce some notation. Q is any probability measure dominating P and P', $z := d\mathsf{P}/d\mathsf{Q}$ and $z' := d\mathsf{P}'/d\mathsf{Q}$ are the corresponding Radon–Nikodým derivatives. E, E', and E_{Q} are the expectations with respect to P, P' and Q respectively. Note that $\mathsf{P}(z = 0) = 0$ and $Z := z'/z$, where $0/0 = 0$ by convention, is the Radon–Nikodým derivative of the P-absolutely continuous part of P' with respect to P.

For an experiment $\mathbb{E} = (\Omega, \mathscr{F}, \mathsf{P}, \mathsf{P}')$, denote by $\Phi(\mathbb{E})$ the set of all test functions φ in $\mathbb{E}$, i.e. measurable mappings from $(\Omega, \mathscr{F})$ to $[0, 1]$. It is convenient for us to interpret $\varphi(\omega)$ as the probability to accept the null hypothesis P and to reject the
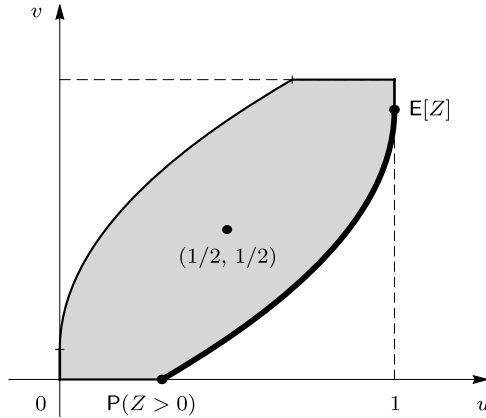
Fig. 5. The shaded area represents the set $\mathfrak{N}(\mathbb{E})$. The thick curve corresponds to admissible, or Neyman–Pearson tests φ^* with the following property: if $\alpha(\varphi) \leq \alpha(\varphi^*)$ and $\beta(\varphi) \leq \beta(\varphi^*)$ for some $\varphi \in \Phi(\mathbb{E})$, then both inequalities are equalities. See [27, Chapter 2] for more details. The thick curve together with the horizontal segment $[0, \ \mathsf{P}(Z > 0)] \times \{0\}$ is the graph of the function $\mathsf{r}_{\mathbb{E}}(1 - u)$

alternative P' if ω is observed. Then $\alpha(\varphi) := \mathsf{E}[1 - \varphi]$ and $\beta(\varphi) := \mathsf{E}'[\varphi]$ are the probabilities of errors of the first and the second kind respectively of a test φ.

Denote

$$\mathfrak{N}(\mathbb{E}) := \{(\mathsf{E}[\varphi], \ \mathsf{E}'[\varphi]) : \varphi \in \Phi(\mathbb{E})\} = \{(1 - \alpha(\varphi), \ \beta(\varphi)) : \varphi \in \Phi(\mathbb{E})\}.$$

It is well known that $\mathfrak{N}(\mathbb{E})$ is a convex and closed subset of $[0, 1] \times [0, 1]$, contains $(0, 0)$, and is symmetric with respect to the point $(1/2, 1/2)$, see, e.g., [16, p. 62]. In Fig. 5 we present a set $\mathfrak{N}(\mathbb{E})$ of generic form. Introduce also the *risk function*

$$\mathsf{r}_{\mathbb{E}}(u) := \inf \{\beta(\varphi) : \varphi \in \Phi(\mathbb{E}), \ \alpha(\varphi) = u\}, \quad u \in [0, 1], \tag{22}$$

that is the smallest probability of the second kind error if the probability of the first kind error is u. It follows that the set $\mathfrak{N}(\mathbb{E})$ and the risk function $\mathsf{r}_{\mathbb{E}}$ are connected by

$$\mathsf{r}_{\mathbb{E}}(u) = \inf \{v : (1 - u, v) \in \mathfrak{N}(\mathbb{E})\} \tag{23}$$

and

$$\mathfrak{N}(\mathbb{E}) = \{(u, \ v) \in [0, 1] \times [0, 1] : \mathsf{r}_{\mathbb{E}}(1 - u) \leq v \leq 1 - \mathsf{r}_{\mathbb{E}}(u)\}. \tag{24}$$

In particular, $\mathsf{r}_{\mathbb{E}}$ is a continuous convex decreasing function taking values in $[0, 1]$ and $\mathsf{r}_{\mathbb{E}}(1) = 0$. Therefore, by Theorem 4, $\mathsf{r}_{\mathbb{E}}(u)$ coincides on $[0, 1]$ with an integrated quantile function corresponding to some distribution. The following result determines this distribution and explains why it is natural to use integrated quantile functions for binary models.

Proposition 1. *For all* $u \in [0, 1]$, $\mathsf{r}_{\mathbb{E}}(u) = \mathsf{K}_{-Z}(u)$, *where* K_{-Z} *is the integrated quantile function corresponding to the distribution of the negative likelihood ratio* $-Z = -z'/z$ *under the null hypothesis.*

Proof. Let $\varphi_0 \in \Phi(\mathbb{E})$ and $x \in \mathbb{R}_+$. Then the straight line with the slope x and passing through the point $(\mathsf{E}[\varphi_0], \mathsf{E}'[\varphi_0])$ lies below the graph of $K(u) := \mathsf{r}_{\mathbb{E}}(1-u)$ on $[0, 1]$ if and only if, for every $\varphi \in \Phi(\mathbb{E})$,

$$\mathsf{E}'[\varphi] \geq \mathsf{E}'[\varphi_0] + x\big(\mathsf{E}[\varphi] - \mathsf{E}[\varphi_0]\big).$$

Passing to a dominating measure Q, the above inequality can be rewritten as

$$\mathsf{E}_{\mathsf{Q}}\big[(z' - xz)(\varphi - \varphi_0)\big] \geq 0.$$

This holds for every $\varphi \in \Phi(\mathbb{E})$ if and only if

$$\varphi_0 = \mathbb{1}_{\{z' < xz\}} + \varphi_0 \mathbb{1}_{\{z'=xz\}} \quad \mathsf{Q}\text{-a.s.} \tag{25}$$

Let $u \in (0, 1)$ and take any $x \in [q_Z^L(u), q_Z^R(u)]$. Then $u \in [F_Z(x-0), F_Z(x)]$, so there is $\gamma \in [0, 1]$ such that $u = (1-\gamma)F_Z(x-0) + \gamma F_Z(x)$. Finally, put $\varphi_0 := \mathbb{1}_{\{z' < xz\}} + \gamma \mathbb{1}_{\{z'=xz\}}$. Since $Z = z'/z$ P-a.s., we get $\mathsf{E}[\varphi_0] = u$ and, obviously, φ_0 satisfies (25). This means that $x \in \partial K(u)$. Conversely, let $u \in (0, 1)$ and $x \in \partial K(u)$. Take any $\varphi_0 \in \Phi(\mathbb{E})$ such that $\mathsf{E}[\varphi_0] = u$ and $\beta(\varphi_0) = K(u)$. Then φ_0 satisfies (25), which implies $\mathsf{E}[\varphi_0] \in [F_Z(x-0), F_Z(x)]$. Hence, $x \in [q_Z^L(u), q_Z^R(u)]$.

It is clear that $K(0) = \mathsf{K}_Z(0) = 0$. Now taking into account that K and K_Z are convex functions, K_Z is continuous on $[0, 1]$, and $\partial K(u) = \partial \mathsf{K}_Z(u)$ for $u \in (0, 1)$, it remains to prove that $K(1) \leq \mathsf{K}_Z(1) = \mathsf{E}[Z]$. This is easy: take $\varphi_0 := \mathbb{1}_{\{z>0\}}$, then $\mathsf{E}[\varphi_0] = 1$ and $\mathsf{E}'[\varphi_0] = \mathsf{P}'(z > 0) = \mathsf{E}_{\mathsf{Q}}[z' \mathbb{1}_{\{z>0\}}] = \mathsf{E}[Z]$. Finally, $\mathsf{r}_{\mathbb{E}}(u) = K(1-u) = \mathsf{K}_Z(1-u) = \mathsf{K}_{-Z}(u)$. $\square$

Remark 4. A usual way to prove that the set $\mathfrak{N}(\mathbb{E})$ is closed is based on weak compactness of test functions, see, e.g., [16]. The reader may readily verify that the closedness of $\mathfrak{N}(\mathbb{E})$ follows directly from the above proof.

Let us also introduce the *minimum Bayes risk function* (the *error function*)

$$\mathsf{b}_{\mathbb{E}}(\pi) := \inf_{\varphi \in \Phi(\mathbb{E})} \big((1-\pi)\alpha(\varphi) + \pi\beta(\varphi)\big), \quad \pi \in [0, 1].$$

It can be expressed in terms of risk function $\mathsf{r}_{\mathbb{E}}$ and vice versa. Indeed, for any $\pi \in (0, 1)$,

$$\mathsf{b}_{\mathbb{E}}(\pi) = \inf_{u \in [0, 1]} \big((1-\pi)u + \pi \mathsf{r}_{\mathbb{E}}(u)\big)$$

$$= -\pi \sup_{u \in [0, 1]} \left(-\frac{1-\pi}{\pi}u - \mathsf{r}_{\mathbb{E}}(u)\right) = -\pi \sup_{u \in [0, 1]} \left(-\frac{1-\pi}{\pi}u - \mathsf{K}_{-Z}(u)\right) =$$

$$= -\pi \mathsf{J}_{-Z}\left(-\frac{1-\pi}{\pi}\right) = 1 - \pi - \pi \mathsf{J}_Z\left(\frac{1-\pi}{\pi}\right). \tag{26}$$

In particular, it follows from Theorem 1 that

$$\lim_{\pi \downarrow 0} \frac{\mathsf{b}_{\mathbb{E}}(\pi)}{\pi} = \lim_{x \to +\infty} \big(x - \mathsf{J}_Z(x)\big) = \mathsf{E}[Z],$$

$$\lim_{\pi \uparrow 1} \frac{\mathsf{b}_{\mathbb{E}}(\pi)}{1 - \pi} = 1 - \lim_{x \downarrow 0} \frac{\mathsf{J}_Z(x)}{x} = \mathsf{P}(Z > 0),$$

see [27, Lemma 14.6] and [28, p. 607].

Conversely, using Definition 2 and (26), we get, for $u \in [0, 1]$,

$$
r_{\mathbb{E}}(u) = \mathsf{K}_{-Z}(u) = \sup_{x \in \mathbb{R}} \big(xu - \mathsf{J}_{-Z}(x) \big) = \sup_{x < 0} \big(xu - \mathsf{J}_{-Z}(x) \big)
$$

$$
= \sup_{\pi \in (0, 1)} \left(-\frac{1 - \pi}{\pi} u - \mathsf{J}_{-Z}\left(-\frac{1 - \pi}{\pi} \right) \right) = \sup_{\pi \in (0, 1)} \frac{1}{\pi} \big(b_{\mathbb{E}}(\pi) - (1 - \pi)u \big),
\tag{27}
$$

see [28, p. 590]. Here we have used that $\mathsf{J}_{-Z}(x) = x$ for $x \geq 0$.

Finally, let us introduce one more characteristic of binary models, namely the distribution of the 'likelihood ratio'

$$
\mu_{\mathbb{E}}(A) := \mathsf{P}(Z \in A), \quad A \in \mathscr{B}(\mathbb{R}_+).
$$

Now let us present some basic notions and results from the theory of comparison of dichotomies. All these facts are well known, see e. g. [27, Chapter 3] and [28, Chapter 10]. Our aim is to show how they can be deduced with the help of the results in Sections 2 and 3.

Definition 3. Let $\mathbb{E} = (\Omega, \mathscr{F}, \mathsf{P}, \mathsf{P}')$ and $\widetilde{\mathbb{E}} = (\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathsf{P}}, \widetilde{\mathsf{P}}')$ be two binary experiments. $\mathbb{E}$ is said to be *more informative* than $\widetilde{\mathbb{E}}$, denoted by $\mathbb{E} \succeq \widetilde{\mathbb{E}}$ or $\widetilde{\mathbb{E}} \preceq \mathbb{E}$, if $\mathfrak{N}(\mathbb{E}) \supseteq \mathfrak{N}(\widetilde{\mathbb{E}})$. $\mathbb{E}$ and $\widetilde{\mathbb{E}}$ are called *equivalent* ($\mathbb{E} \sim \widetilde{\mathbb{E}}$) if $\mathbb{E} \succeq \widetilde{\mathbb{E}}$ and $\mathbb{E} \preceq \widetilde{\mathbb{E}}$. The *type* of an experiment is the totality of all experiments which are equivalent to the given experiment.

Proposition 2. *Let $\mathbb{E}$ and $\widetilde{\mathbb{E}}$ be binary experiments. The following statements are equivalent*:

   (i) $\mathbb{E} \succeq \widetilde{\mathbb{E}}$;

   (ii) $r_{\mathbb{E}} \leq r_{\widetilde{\mathbb{E}}}$;

   (iii) $b_{\mathbb{E}} \leq b_{\widetilde{\mathbb{E}}}$;

   (iv) $\mu_{\widetilde{\mathbb{E}}} \preceq_{decx} \mu_{\mathbb{E}}$.

Proof. (i) $\Leftrightarrow$ (ii) follows from (23) and (24). (ii) $\Leftrightarrow$ (iii) is a consequence of (26) and (27). Finally, (ii) $\Leftrightarrow$ (iv) follows from Proposition 1 and Theorem 5. □

Corollary 3. *Let $\mathbb{E}$ and $\widetilde{\mathbb{E}}$ be binary experiments. The following statements are equivalent*:

   (i) $\mathbb{E} \sim \widetilde{\mathbb{E}}$;

   (ii) $r_{\mathbb{E}} = r_{\widetilde{\mathbb{E}}}$;

   (iii) $b_{\mathbb{E}} = b_{\widetilde{\mathbb{E}}}$;

   (iv) $\mu_{\mathbb{E}} = \mu_{\widetilde{\mathbb{E}}}$.

Proposition 3.   (i) *The mapping $\mathbb{E} \rightsquigarrow r_{\mathbb{E}}$ is onto the set of all convex continuous decreasing functions $r \colon [0, 1] \to [0, 1]$ such that $r(1) = 0$.*

(ii) *The mapping* $\mathbb{E} \rightsquigarrow b_{\mathbb{E}}$ *is onto the set of all concave functions* $b \colon [0, 1] \to [0, 1]$ *such that* $b(\pi) \leq \pi \wedge (1 - \pi)$.

(iii) *The mapping* $\mathbb{E} \rightsquigarrow \mu_{\mathbb{E}}$ *is onto the set of all probability measures* μ *on* $(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$ *such that* $\int x \, \mu(dx) \leq 1$.

Proof. (i) Let $r(u)$, $u \in [0, 1]$, be a convex continuous decreasing function with $r(1) = 0$ and $r(0) \leq 1$. Let $\Omega = [0, 1]$ and $\mathscr{F}$ be the Borel σ-field. Define P on $(\Omega, \mathscr{F})$ as the Lebesgue measure and P' as the measure with the distribution function

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ r(1 - x), & \text{if } 0 \leq x < 1, \\ 1, & \text{if } x \geq 1. \end{cases} \tag{28}$$

Then $Z(u) = F'_-(u)$ P-a.s. As in the proof of Theorem 4, it follows that $r(1 - u) = K_Z(u)$. Proposition 1 allows us to conclude that $r_{\mathbb{E}} = r$.

(iii) First, it is evident that $\mu_{\mathbb{E}}$ is a probability measure on $(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$ such that $\int x \, \mu_{\mathbb{E}}(dx) \leq 1$ for any dichotomy $\mathbb{E}$. Now, let μ be a probability measure on $(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$ such that $\int x \, \mu(dx) \leq 1$. Put $\Omega = [0, +\infty]$ and let $\mathscr{F}$ be the Borel σ-field. Define P as the probability measure which coincides with μ on Borel subsets of $\mathbb{R}_+$. Finally, define P' by

$$\mathsf{P}'(B \cap \mathbb{R}_+) := \int_{B \cap \mathbb{R}_+} x \, \mu(dx), \qquad \mathsf{P}'(\{+\infty\}) := 1 - \int_{\mathbb{R}_+} x \, \mu(dx).$$

If $\mathbb{E}$ is defined as $\mathbb{E} = (\Omega, \mathscr{F}, \mathsf{P}, \mathsf{P}')$, it is clear that $\mu_{\mathbb{E}} = \mu$.

(ii) First, it follows from the definition of the error function that $0 \leq b_{\mathbb{E}}(\pi) \leq \pi \wedge (1 - \pi)$, $\pi \in [0, 1]$, and that $b_{\mathbb{E}}$ is concave. If b is a function with these properties, then define $J(x) := x - (1 + x)b(\frac{1}{1+x})$, $x \geq 0$, cf. (26); put also $J(x) = 0$ for $x < 0$. Using concavity of b, it is easy to check that J is convex on $\mathbb{R}_+$. Since $b(0) = 0$, we have $\lim_{x \to +\infty} \frac{J(x)}{x} = 1$. The inequalities $0 \leq b(\pi) \leq 1 - \pi$ imply that $0 \leq J(x) \leq x$ for all $x \geq 0$. In particular, J is convex on $\mathbb{R}$ and, by Theorem 2, J is the integrated distribution function of some nonnegative random variable Z. Finally, the inequality $b(\pi) \leq \pi$ implies that $J(x) \geq x - 1$, which means that $\mathsf{E}[Z] \leq 1$ by Theorem 1. Hence, b is the error function of an experiment $\mathbb{E}$ such that $\mu_{\mathbb{E}} = \mathrm{Law}(Z)$. $\qquad\square$

Let us note that the proofs of (i) and (iii) give more than it is stated. Starting with a function r or a measure μ from corresponding classes, we construct an experiment such that its risk function (resp., the distribution of the likelihood ratio) coincides with r (resp. μ). Now, if we start in (i) with the risk function $r_{\mathbb{E}}$ of an experiment $\mathbb{E}$, we obtain a new experiment, say, $\varkappa(\mathbb{E})$, equivalent to $\mathbb{E}$. Moreover, experiments $\mathbb{E}_1$ and $\mathbb{E}_2$ are equivalent if and only if $\varkappa(\mathbb{E}_1) = \varkappa(\mathbb{E}_2)$. In other words, the rule $\mathbb{E} \rightsquigarrow \varkappa(\mathbb{E})$ is a representation of binary experiments. Another representation is given in the proof of (iii).

Definition 4. Let $\mathbb{E} = (\Omega, \mathscr{F}, \mathsf{P}, \mathsf{P}')$ and $\widetilde{\mathbb{E}} = (\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathsf{P}}, \widetilde{\mathsf{P}}')$ be two binary experiments. $\mathbb{E}$ is called ε-*deficient* with respect to $\widetilde{\mathbb{E}}$ if for any $\widetilde{\varphi} \in \Phi(\widetilde{\mathbb{E}})$ there is $\varphi \in \Phi(\mathbb{E})$ such that $\alpha(\varphi) \leq \alpha(\widetilde{\varphi}) + \varepsilon/2$ and $\beta(\varphi) \leq \beta(\widetilde{\varphi}) + \varepsilon/2$. The number

$$\delta_2(\mathbb{E}, \widetilde{\mathbb{E}}) := \inf \{\varepsilon \geq 0 \colon \mathbb{E} \text{ is } \varepsilon\text{-deficient with respect to } \widetilde{\mathbb{E}}\}$$

is called the (asymmetric) *deficiency* of $\mathbb{E}$ with respect to $\widetilde{\mathbb{E}}$. Define also the (symmetric) *deficiency*

$$\Delta_2(\mathbb{E}, \widetilde{\mathbb{E}}) := \max\left(\delta_2(\mathbb{E}, \widetilde{\mathbb{E}}), \delta_2(\widetilde{\mathbb{E}}, \mathbb{E})\right)$$

between $\mathbb{E}$ and $\widetilde{\mathbb{E}}$.

It is easy to check that $\mathbb{E} \succeq \widetilde{\mathbb{E}}$ if and only if $\delta_2(\mathbb{E}, \widetilde{\mathbb{E}}) = 0$. Hence, $\mathbb{E} \sim \widetilde{\mathbb{E}}$ if and only if $\Delta_2(\mathbb{E}, \widetilde{\mathbb{E}}) = 0$. It is also easy to check that δ_2 and Δ_2 satisfy the triangle inequality and, hence, Δ_2 is a metric on the space of types of experiments. We shall see after the next proposition that this metric space is a compact space.

Proposition 4. *Let $\mathbb{E}$ and $\widetilde{\mathbb{E}}$ be binary experiments. The following statements are equivalent*:

(i) *$\mathbb{E}$ is ε-deficient with respect to $\widetilde{\mathbb{E}}$.*

(ii) *For all $u \in [0, 1 - \frac{\varepsilon}{2}]$,*

$$r_{\mathbb{E}}\left(u + \frac{\varepsilon}{2}\right) \leq r_{\widetilde{\mathbb{E}}}(u) + \frac{\varepsilon}{2}. \tag{29}$$

(iii) *For all $\pi \in (0, 1)$,*

$$b_{\mathbb{E}}(\pi) \leq b_{\widetilde{\mathbb{E}}}(\pi) + \frac{\varepsilon}{2}. \tag{30}$$

Proof. (i) $\Leftrightarrow$ (ii) follows immediately from Definition 4, so our goal is to prove (ii) $\Leftrightarrow$ (iii) using dual relations (13) and (14). A direct proof of (i) $\Leftrightarrow$ (iii) can be found in [27].

To simplify the notation, put $\mathsf{K} := \mathsf{K}_Z$, $\mathsf{J} := \mathsf{J}_Z$, while the corresponding functions in the experiment $\widetilde{\mathbb{E}}$ are denoted by $\widetilde{\mathsf{K}}$ and $\widetilde{\mathsf{J}}$. Since $r_{\mathbb{E}}(u) = \mathsf{K}(1 - u)$ (29) is equivalent to

$$\mathsf{K}(u) \leq \widetilde{\mathsf{K}}\left(u + \frac{\varepsilon}{2}\right) + \frac{\varepsilon}{2} \quad \text{for all } u \in \left[0, 1 - \frac{\varepsilon}{2}\right]. \tag{31}$$

In turn, if follows from (26) that (30) is equivalent to

$$\mathsf{J}(x) \geq \widetilde{\mathsf{J}}(x) - \frac{\varepsilon}{2}(1 + x) \quad \text{for all } x > 0. \tag{32}$$

Since $\mathsf{J}(x) = 0$ for $x \leq 0$, we have $\mathsf{K}(u) = \sup_{x \geq 0}\{xu - \mathsf{J}(x)\}$ and similarly for $\widetilde{\mathsf{K}}$. Thus, it follows from (32) that, for $u \geq 0$,

$$\mathsf{K}(u) \leq \sup_{x \geq 0}\left\{xu + \frac{\varepsilon}{2}(1 + x) - \widetilde{\mathsf{J}}(x)\right\} = \widetilde{\mathsf{K}}\left(u + \frac{\varepsilon}{2}\right) + \frac{\varepsilon}{2}.$$

Conversely, let (31) hold true, and let $x > 0$ be such that $F_{\widetilde{Z}}(x - 0) \geq \frac{\varepsilon}{2}$, where $\widetilde{Z}$ is the Radon–Nikodým derivative of the $\widetilde{\mathsf{P}}$-absolutely continuous part of $\widetilde{\mathsf{P}}'$ with respect to $\widetilde{\mathsf{P}}$. Then

$$\mathsf{J}(x) \geq \sup_{u \in [0, 1]}\left\{xu - \mathsf{K}(u)\right\} \geq \sup_{u \in [0, 1 - \frac{\varepsilon}{2}]}\left\{xu - \mathsf{K}(u)\right\}$$

$$\geq \sup_{u\in[0,\,1-\frac{\varepsilon}{2}]}\left\{xu - \frac{\varepsilon}{2} - \widetilde{\mathsf{K}}\left(u + \frac{\varepsilon}{2}\right)\right\} = \sup_{u\in[\frac{\varepsilon}{2},\,1]}\left\{xu - \frac{(1+x)\varepsilon}{2} - \widetilde{\mathsf{K}}(u)\right\}$$

$$= -\frac{\varepsilon}{2}(x+1) + \widetilde{\mathsf{J}}(x),$$

where the last equality follows from the fact that the supremum in (14) is attained at $u \in [F_{\widetilde{\mathsf{Z}}}(x-0),\ F_{\widetilde{\mathsf{Z}}}(x)]$, cf. (17). It remains to note that if x is such that $F_{\widetilde{\mathsf{Z}}}(x-0) < \frac{\varepsilon}{2}$, then $\widetilde{\mathsf{J}}(x) \leq \frac{\varepsilon x}{2}$ and (32) is obviously true. $\qquad\square$

As a consequence, we obtain the following expressions for $\delta_2(\mathbb{E},\ \widetilde{\mathbb{E}})$ and $\Delta_2(\mathbb{E},\ \widetilde{\mathbb{E}})$, see ([27], [28, p. 604]):

Corollary 4. *Let $\mathbb{E}$ and $\widetilde{\mathbb{E}}$ be binary experiments. Then*

$$\delta_2(\mathbb{E},\ \widetilde{\mathbb{E}}) = \frac{1}{2}\sup_{\pi\in[0,\,1]}\left\{\mathsf{b}_{\mathbb{E}}(\pi) - \mathsf{b}_{\widetilde{\mathbb{E}}}(\pi)\right\},$$

$$\Delta_2(\mathbb{E},\ \widetilde{\mathbb{E}}) = \frac{1}{2}\sup_{\pi\in[0,\,1]}\left|\mathsf{b}_{\mathbb{E}}(\pi) - \mathsf{b}_{\widetilde{\mathbb{E}}}(\pi)\right| = \frac{1}{2}L(F,\ \widetilde{F}),$$

where $L(\cdot,\ \cdot)$ is the Lévy distance between distribution functions, F is defined as in (28) with $\mathsf{r} = \mathsf{r}_{\mathbb{E}}$, and $\widetilde{F}$ is defined similarly with $\mathsf{r} = \mathsf{r}_{\widetilde{\mathbb{E}}}$.

The subset of concave functions b on $[0,\ 1]$ satisfying $0 \leq \mathsf{b}(\pi) \leq \pi \wedge (1-\pi)$ is clearly closed with respect to uniform convergence and is equicontinuous. By the Arzela–Ascoli theorem, this subset is a compact in the space $C[0,\ 1]$ with sup-norm. Therefore, the space of types of experiments is a compact metric space with Δ_2-metric.

Definition 5. Let $\mathbb{E} = (\Omega, \mathscr{F}, \mathsf{P}, \mathsf{P}')$ and $\mathbb{E}^n = (\Omega^n, \mathscr{F}^n, \mathsf{P}^n, \mathsf{P}'^n)$, $n \geq 1$, be binary experiments. We say that $\mathbb{E}^n$ *weakly converges* to $\mathbb{E}$ if $\Delta_2(\mathbb{E}^n, \mathbb{E}) \to 0$ as $n \to \infty$.

Proposition 5. *Let $\mathbb{E}$ and $\mathbb{E}^n$, $n \geq 1$, be binary experiments. The following statements are equivalent*:

  (i) $\Delta_2(\mathbb{E}^n,\ \mathbb{E}) \to 0$.

  (ii) $\mathsf{r}_{\mathbb{E}^n}$ *converges to* $\mathsf{r}_{\mathbb{E}}$ *pointwise on* $(0,\ 1]$.

  (ii′) $\mathsf{r}_{\mathbb{E}^n}$ *converges uniformly to* $\mathsf{r}_{\mathbb{E}}$ *on any* $[a,\ 1] \subset (0,\ 1]$.

  (iii) $\mathsf{b}_{\mathbb{E}^n}$ *converges uniformly to* $\mathsf{r}_{\mathbb{E}}$ *on* $[0,\ 1]$.

  (iv) $\mu_{\mathbb{E}^n}$ *weakly converges to* $\mu_{\mathbb{E}}$.

Proof. The equivalences (i) $\Leftrightarrow$ (ii) and (i) $\Leftrightarrow$ (iii) follow from Corollary 4, and the equivalence of (ii), (ii′), and (iv) is a consequence of Theorem 10 and Proposition 1. However, we prefer to give a direct proof of the equivalence (i) $\Leftrightarrow$ (ii) without using the Lévy distance.

Assume (i). By (29),

$$\mathsf{r}_{\mathbb{E}^n}(u) + \frac{\Delta_2(\mathbb{E}^n,\ \mathbb{E})}{2} \geq \mathsf{r}_{\mathbb{E}}\left(u + \frac{\Delta_2(\mathbb{E}^n,\ \mathbb{E})}{2}\right), \qquad 0 \leq u \leq 1 - \frac{\Delta_2(\mathbb{E}^n,\ \mathbb{E})}{2},$$

$$r_{\mathbb{E}}\left(u - \frac{\Delta_2(\mathbb{E}^n, \mathbb{E})}{2}\right) + \frac{\Delta_2(\mathbb{E}^n, \mathbb{E})}{2} \geq r_{\mathbb{E}^n}(u), \qquad \frac{\Delta_2(\mathbb{E}^n, \mathbb{E})}{2} \leq u \leq 1.$$

Passing to the limit as $n \to \infty$, we get

$$\liminf_{n \to \infty} r_{\mathbb{E}^n}(u) \geq r_{\mathbb{E}}(u) \quad \text{for } 0 \leq u < 1,$$

$$r_{\mathbb{E}}(u) \geq \limsup_{n \to \infty} r_{\mathbb{E}^n}(u) \quad \text{for } 0 < u \leq 1.$$

Combining these inequalities, we obtain $\lim_{n \to \infty} r_{\mathbb{E}^n}(u) = r_{\mathbb{E}}(u)$ for $0 < u < 1$. Since risk functions vanish at 1, the convergence holds for $u = 1$ as well.

Now the converse implication (ii) $\Rightarrow$ (i) is proved by standard compactness arguments. $\qquad\square$

5   Chacon–Walsh revisited

The Skorokhod embedding problem was posed and solved by Skorokhod [26] in the following form: given a centered distribution μ with finite second moment, find a stopping time T such that $\mathsf{E}[T] < \infty$ and $\mathrm{Law}(B_T) = \mu$, where $B = (B_t)_{t \geq 0}$, $B_0 = 0$, is a standard Brownian motion. Chacon and Walsh [4] suggest to construct T as the limit of an increasing sequence of stopping times T_n, each being the first exit time (after the previous one) of B from a compact interval. This construction has a simple graphical interpretation in terms of the potential functions of B_{T_n} (we recall that potential functions are defined in (3)).

Cox [6] extends the Chacon–Walsh construction to a more general case. He considers a Brownian motion $B = (B_t)_{t \geq 0}$ with a given integrable starting distribution μ_0 for B_0 and a general integrable target distribution μ. A solution T (such that $\mathrm{Law}(B_T) = \mu$) must be found in the class of *minimal* stopping times.

It is easy to observe that the Chacon–Walsh construction has a graphical interpretation in terms of integrated quantile functions as well; moreover, in our opinion, the picture is more simple. We give alternative proofs of the result in [4] and of some results in [6]. Moreover, we construct a minimal stopping time in some special case where μ_0 and μ may be non-integrable.

Let us recall the definition of the balayage. For a probability measure μ on $\mathbb{R}$ and an interval $I = (a, b)$, $-\infty < a < b < +\infty$, the balayage μ_I of μ on I is defined as the measure which coincides with μ outside $[a, b]$, vanishes on (a, b), and such that
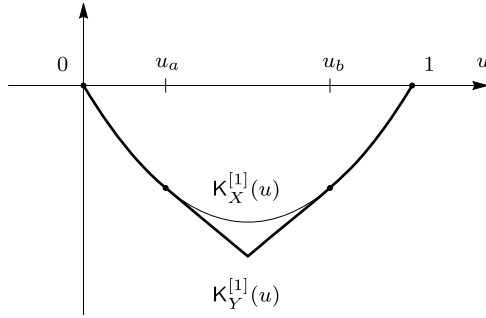
$$\mu_I(\{a\}) = \int_{[a, b]} \frac{b - x}{b - a} \mu(dx), \qquad \mu_I(\{b\}) = \int_{[a, b]} \frac{x - a}{b - a} \mu(dx). \qquad (33)$$

Since

$$\int_{[a, b]} \mu_I(dx) = \int_{[a, b]} \mu(dx) \quad \text{and} \quad \int_{[a, b]} x\, \mu_I(dx) = \int_{[a, b]} x\, \mu(dx), \qquad (34)$$

the balayage μ_I is a probability measure and has the same mean as μ (if defined). It follows that, if $B = (B_t)_{t \geq 0}$ is a continuous local martingale with $\langle B, B \rangle_\infty = \infty$ a.s. (e. g. a Brownian motion), μ is the distribution of B_S, where S is a stopping time, and the stopping time T is defined by

Fig. 6. Graphs of shifted integrated quantile functions $\mathsf{K}_X^{[1]}$ and $\mathsf{K}_Y^{[1]}$: the distribution of Y is the balayage of the distribution of X

$$T := \inf\{t \geq S\colon B_t \notin I\}, \tag{35}$$

then $T < +\infty$ a. s. and the distribution of B_T is the balayage μ_I.

Let X and Y be random variables with the distributions μ and μ_I respectively. It is clear that

$$q_Y^L(u) = \begin{cases} q_X^L(u), & \text{if } 0 < u \leq F_X(a-0) \text{ or } F_X(b) < u < 1, \\ a, & \text{if } F_X(a-0) < u \leq F_X(a-0) + \mu_I(\{a\}), \\ b, & \text{if } F_X(a-0) + \mu_I(\{a\}) < u \leq F_X(b). \end{cases}$$

Moreover, the second equality in (34) can be rewritten as

$$\mathsf{K}_X\big(F_X(b)\big) - \mathsf{K}_X\big(F_X(a-0)\big) = \mathsf{K}_X\big(F_Y(b)\big) - \mathsf{K}_X\big(F_Y(a-0)\big).$$

This allows us to describe how to obtain the integrated quantile function of Y: pass the tangent lines with the slopes a and b to the graph of K_X, replace the curve on this graph between points where the graph meets the lines by the corresponding segments of these lines. If the point of intersection of these lines lies below the horizontal axis, then shift the resulting graph vertically upwards so that this point will come on the horizontal axis.

If $\mathsf{E}[X^+] < \infty$ (resp. $\mathsf{E}[X^-] < \infty$), then $\mathsf{E}[Y^+] < \infty$ (resp. $\mathsf{E}[Y^-] < \infty$), and the last step is not needed if we deal with shifted integrated quantile functions $\mathsf{K}_X^{[1]}$ and $\mathsf{K}_Y^{[1]}$ (resp. $\mathsf{K}_X^{[0]}$ and $\mathsf{K}_Y^{[0]}$), see Fig. 6. We state this fact in the following lemma only in the case where $\mathsf{E}[X^+] < \infty$. Its proof is immediate from the previous paragraph.

Lemma 1. *Let μ be the distribution of a random variable X with $\mathsf{E}[X^+] < \infty$, Law$(Y) = \mu_I$, where $I = (a, b)$ is a finite interval. Put $u_a := F_X(a-0)$ and $u_b := F_X(b)$. Then $\mathsf{E}[Y^+] < \infty$ and*

$$\mathsf{K}_Y^{[1]}(u) = \begin{cases} \mathsf{K}_X^{[1]}(u), & \text{if } u \notin (u_a, u_b), \\ \big(a(u - u_a) + \mathsf{K}_X^{[1]}(u_a)\big) \vee \big(b(u - u_b) + \mathsf{K}_X^{[1]}(u_b)\big), & \text{if } u \in (u_a, u_b). \end{cases}$$

In particular, $\mathsf{K}_Y^{[1]}(u) \leq \mathsf{K}_X^{[1]}(u)$ for all $u \in [0, 1]$.

The next lemma is a key tool in our future construction.

Lemma 2. *Let X and Y be random variables such that $\mathsf{E}[X^+] < \infty$, $\mathsf{E}[Y^+] < \infty$, and $\mathsf{K}_Y^{[1]}(u) \leq \mathsf{K}_X^{[1]}(u)$ for all $u \in [0, 1]$. Fix $v \in (0, 1)$. Then there is a random variable Z such that $\mathsf{K}_Y^{[1]}(u) \leq \mathsf{K}_Z^{[1]}(u) \leq \mathsf{K}_X^{[1]}(u)$ for all $u \in [0, 1]$, $\mathsf{K}_Z^{[1]}(v) = \mathsf{K}_Y^{[1]}(v)$, and the distribution of Z is a balayage of the distribution of X.*

Proof. Without loss of generality, we may assume that $\mathsf{K}_X^{[1]}(v) > \mathsf{K}_Y^{[1]}(v)$. Let us consider the following equation:

$$xv - \mathsf{J}_X(x) - \mathsf{E}[X^+] = \mathsf{K}_Y^{[1]}(v). \tag{36}$$

The maximum of the left-hand side over x equals $\mathsf{K}_X(v) - \mathsf{E}[X^+] = \mathsf{K}_X^{[1]}(v)$ and is greater than the right-hand side. Moreover, it is attained at $x \in [q_X^L(v), q_X^R(v)]$. Further, applying Theorem 1 (iv)–(v), we get

$$\lim_{x \to +\infty} \left(xv - \mathsf{J}_X(x)\right) = \lim_{x \to +\infty} \left(x - \mathsf{J}_X(x)\right) + \lim_{x \to +\infty} (v - 1)x = -\infty$$

and

$$\lim_{x \to -\infty} \left(xv - \mathsf{J}_X(x)\right) = \lim_{x \to -\infty} x\left(v - \frac{\mathsf{J}_X(x)}{x}\right) = -\infty.$$

Since the left-hand side of (36) is a concave function in x, the equation (36) has two solutions $a < q_X^L(v)$ and $b > q_X^R(v)$, i.e. $F_X(a) < v < F_X(b - 0)$.

Using Corollary 2, rewrite equation (36) in the form

$$\mathsf{K}_X^{[1]}\left(F_X(x)\right) = \mathsf{K}_Y^{[1]}(v) + x\left(F_X(x) - v\right).$$

This equality for $x = a$ (resp. $x = b$) says that the straight line with the slope a (resp. b) and passing through the point $(v, \mathsf{K}_Y^{[1]}(v))$ meets the curve $\mathsf{K}_X^{[1]}$ at the point where the first coordinate is $F_X(a)$ (resp. $F_X(b)$). Due to (17), these straight lines are tangent lines to the curve $\mathsf{K}_X^{[1]}$. Comparing with Lemma 1, we obtain that a random variable Z such that its distribution is the balayage of the distribution of X on $I = (a, b)$ satisfies all the requirements. $\square$

From now on, we assume that there is a probability space with filtration $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geq 0}, \mathsf{P})$ and an $(\mathscr{F}_t, \mathsf{P})$-Brownian motion $B = (B_t)_{t \geq 0}$ with an arbitrary initial distribution. For $c > 0$, let

$$H_c = \inf\{t \geq 0 \colon B_t \geq c\}.$$

The next lemma is inspired by Theorem 5 in [7].

Lemma 3. *Let S be a stopping time and T defined by (35) with $I = (a, b)$. If $\mathsf{E}[B_S^+] < \infty$ and $c\mathsf{P}(S \geq H_c) \leq \mathsf{E}[B_S \mathbb{1}_{\{S \geq H_c\}}]$, then $\mathsf{E}[B_T^+] < \infty$ and $c\mathsf{P}(T \geq H_c) \leq \mathsf{E}[B_T \mathbb{1}_{\{T \geq H_c\}}]$.*

Proof. By the strong Markov property, in view of boundedness of the random variables under the conditional expectations below,

$$\mathsf{E}[B_T - B_S | \mathscr{F}_S] = 0 \quad \text{and} \quad \mathsf{E}[B_T - B_{(S \vee H_c) \wedge T} | \mathscr{F}_{(S \vee H_c) \wedge T}] = 0.$$

Since $\{S \geq H_c\} \in \mathscr{F}_S$ and $\{S < H_c \leq T\} \in \mathscr{F}_{(S \vee H_c) \wedge T}$, we get

$$
\begin{aligned}
c\mathsf{P}(T \geq H_c) &= c\mathsf{P}(S \geq H_c) + c\mathsf{P}(T \geq H_c > S) \\
&\leq \mathsf{E}[B_S \mathbb{1}_{\{S \geq H_c\}}] + \mathsf{E}[B_{(S \vee H_c) \wedge T} \mathbb{1}_{\{S < H_c \leq T\}}] \\
&= \mathsf{E}[B_T \mathbb{1}_{\{S \geq H_c\}}] + \mathsf{E}[B_T \mathbb{1}_{\{S < H_c \leq T\}}] = \mathsf{E}[B_T \mathbb{1}_{\{T \geq H_c\}}]. \qquad \square
\end{aligned}
$$

Let us also recall that T is a *minimal* stopping time if any stopping time $R \leq T$ with $\mathrm{Law}(B_R) = \mathrm{Law}(B_T)$ satisfies $R = T$ a. s.

Theorem 11. *Let μ_0 and μ be distributions on $\mathbb{R}$ such that $\int_{\mathbb{R}} x^+ \, \mu(dx) < \infty$ and*

$$
\int_{\mathbb{R}} (x - y)^+ \, \mu_0(dx) \leq \int_{\mathbb{R}} (x - y)^+ \, \mu(dx) \quad \text{for all } y \in \mathbb{R}.
$$

Let B be a Brownian motion with the initial distribution $\mathrm{Law}(B_0) = \mu_0$. Then there is an increasing sequence of stopping times $0 = T_0 \leq T_1 \leq \cdots \leq T_n \leq \ldots$ such that $T := \lim_{n \to \infty} T_n$ is a minimal a. s. finite stopping time, the distribution of B_{T_n} is a balayage of the distribution of $B_{T_{n-1}}$ for each $n = 1, 2, \ldots$, and $\mathrm{Law}(B_T) = \mu$.

Proof. Put $X_0 := B_{T_0}$ and let $\mathrm{Law}(Y) = \mu$. Then $\mathsf{K}_Y^{[1]}(u) \leq \mathsf{K}_{X_0}^{[1]}(u)$ for all $u \in [0, 1]$. Take an arbitrary sequence $\{v_n\}$ of distinct points in $(0, 1)$ such that $\{v_n : n = 1, 2, \ldots\}$ is dense in $[0, 1]$. Recursively define X_n as Z in Lemma 2 applied to $X = X_{n-1}$, Y, and $v = v_n$. Then we obtain a sequence $\{X_n\}$ of random variables such that

$$
\mathsf{K}_Y^{[1]}(u) \leq \mathsf{K}_{X_n}^{[1]}(u) \leq \mathsf{K}_{X_{n-1}}^{[1]}(u) \leq \mathsf{K}_{X_0}^{[1]}(u), \quad u \in (0, 1],
$$

and $\mathsf{K}_Y^{[1]}(v_n) = \mathsf{K}_{X_n}^{[1]}(v_n)$, which implies $\mathsf{K}_Y^{[1]}(v_n) = \mathsf{K}_{X_m}^{[1]}(v_n)$ for all n and $m \geq n$. Then $\lim_{n \to \infty} \mathsf{K}_{X_n}^{[1]}(u)$ exists, is finite for all $u \in (0, 1)$, and coincides with $\mathsf{K}_Y^{[1]}(u)$ on the set $\{v_n : n = 1, 2, \ldots\}$. Being a convex function in u, this limit coincides with $\mathsf{K}_Y^{[1]}(u)$ everywhere on $(0, 1)$. It follows from Remark 3 that X_n weakly converges to Y and the sequence $\{X_n^+\}$ is uniformly integrable.

Moreover, the construction in Lemma 2 provides an interval (a, b) denoted by I_n such that the distribution of X_n is the balayage of the distribution of X_{n-1} on I_n. Now recursively define

$$
T_n := \inf\{t \geq T_{n-1} : B_t \notin I_n\}.
$$

Then B_{T_n} has the same distribution as X_n. Since

$$
\mathsf{E}[B_0 \mathbb{1}_{\{0 \geq H_c\}}] = \mathsf{E}[B_0 \mathbb{1}_{\{B_0 \geq c\}}] \geq c\mathsf{P}(B_0 \geq c),
$$

we conclude from Lemma 3 that, for any $c > 0$ and n,

$$
c\mathsf{P}(T_n \geq H_c) \leq \mathsf{E}[B_{T_n} \mathbb{1}_{\{T_n \geq H_c\}}] \leq \mathsf{E}[B_{T_n}^+] \leq \mathsf{E}[Y^+].
$$

If $\mathsf{P}(T = \infty) = \delta > 0$, then the limit of the expression on the left in the last inequality is greater than or is equal to $c\delta$, which is greater than the right-hand side if c is large enough. This contradiction proves that $T < \infty$ a.s. This implies that B_{T_n} converges a.s. to B_T and, hence, $\mathrm{Law}(B_T) = \mu$.

It remains to prove that T is a minimal stopping time. According to Theorem 4.1 in [12], it is enough to find a one-to-one function G such that $G(B)^T$ is a closed submartingale.

Let $g(x)$, $x \in \mathbb{R}$, be a continuously differentiable function with the following properties: $g \equiv 1$ on $[0, +\infty)$ and is strictly positive and increasing on $(-\infty, 0]$, $\int_{-\infty}^0 g(x)\,dx < \infty$, and $g'(x) \leq 1$ for all x. Put $G(y) := \int_0^y g(x)\,dx$, then, in particular, G is strictly increasing and bounded from below, and $G(y) = y$ for $y \geq 0$. By Itô's formula,

$$G(B_t) = G(B_0) + \int_0^t g(B_s)\,dB_s + \frac{1}{2}\int_0^t g'(s)\,ds =: G(B_0) + M_t + A_t,$$

where $G(B_0)$ is an integrable random variable, M is a local martingale and $[M, M]_t = \int_0^t g^2(B_s)\,ds \leq t$. Hence, by the Burkholder–Davis–Gundy inequality, $\mathsf{E}\sup_{s \leq t}|M_s|$ is integrable; in particular, M is a martingale. Finally, A is an increasing process and $A_t \leq t/2$. Therefore, $G(B)$ is a submartingale and, hence, so are the stopped processes $G(B)^T$ and $G(B)^{T_n}$. Note that, by construction, the process $(B - B_0)^{T_n}$ is bounded (by the sum of the lengths of I_k, $k \leq n$) for a fixed n. Hence, $G(B_{t \wedge T_n}) \leq B_{t \wedge T_n}^+ \leq B_0^+ + \sup_{s \leq T_n}|B_s - B_0|$. We conclude that the submartingale $G(B)^{T_n}$ is uniformly integrable, hence, $G(B_{t \wedge T_n}) \leq_{icx} G(B_{T_n})$ for any n and t. On the other hand, $B_{T_n}^+ \leq_{icx} B_T^+$. Combining, we get $[G(B_{t \wedge T_n})]^+ \leq_{icx} B_T^+$ for any n and t. We can pass to the limit as $n \to \infty$ in this inequality, which shows that the family $[G(B_{t \wedge T})]^+$, $t \in \mathbb{R}_+$, is uniformly integrable. The claim follows. $\qquad\square$

Remark 5. It has been already mentioned in the proof of Theorem 5 that the assumptions on μ_0 and μ in Theorem 11 are equivalent to $\mu_0 \leq_{icx} \mu$.

Remark 6. Let μ_0 and μ satisfy the assumptions of Theorem 11. As a by-product, we have obtained the following classical characterization of increasing convex order: there exist random variables X_0 and X defined on the same probability space such that $\mathrm{Law}(X_0) = \mu_0$, $\mathrm{Law}(X) = \mu$ and

$$\mathsf{E}[X|X_0] \geq X_0.$$

Indeed, take $X_0 = B_0$ and $X = B_T$ and use the uniform integrability of $(B_{T_n}^+)_{n \geq 1}$ to obtain the desired inequality.

Acknowledgments

References

[1] Blackwell, D.: Comparison of experiments. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, pp. 93–102. University of California Press, Berkeley and Los Angeles (1951). MR0046002

[2] Blackwell, D.: Equivalent comparisons of experiments. Ann. Math. Statistics **24**, 265–272 (1953). MR0056251

[3] Carlier, G., Chernozhukov, V., Galichon, A.: Vector quantile regression: An optimal transport approach. Ann. Statist. **44**(3), 1165–1192 (2016). MR3485957. https://doi.org/10.1214/15-AOS1401

[4] Chacon, R.V., Walsh, J.B.: One-dimensional potential embedding. In: Séminaire de Probabilités, X. Lecture Notes in Math., vol. 511, pp. 19–23. Springer (1976). MR0445598

[5] Chernozhukov, V., Galichon, A., Hallin, M., Henry, M.: Monge–Kantorovich depth, quantiles, ranks and signs. Ann. Statist. **45**(1), 223–256 (2017). MR3611491. https://doi.org/10.1214/16-AOS1450

[6] Cox, A.M.G.: Extending Chacon-Walsh: minimality and generalised starting distributions. In: Séminaire de Probabilités XLI. Lecture Notes in Math., vol. 1934, pp. 233–264. Springer (2008). https://doi.org/10.1007/978-3-540-77913-1_12

[7] Cox, A.M.G., Hobson, D.G.: Skorokhod embeddings, minimality and non-centred target distributions. Probab. Theory Related Fields **135**(3), 395–414 (2006). https://doi.org/10.1007/s00440-005-0467-y

[8] Embrechts, P., Hofert, M.: A note on generalized inverses. Math. Methods Oper. Res. **77**(3), 423–432 (2013). https://doi.org/10.1007/s00186-013-0436-7

[9] Faugeras, O., Rüschendorf, L.: Markov morphisms: a combined copula and mass transportation approach to multivariate quantiles. Mathematica Applicanda **45**(1), 21–63 (2017)

[10] Feller, W.: An Introduction to Probability Theory and Its Applications. Vol. II. Second edition. John Wiley & Sons, Inc., New York-London-Sydney (1971). MR0270403

[11] Föllmer, H., Schied, A.: Stochastic Finance: An Introduction in Discrete Time, 4th edn. Walter de Gruyter & Co., Berlin (2016)

[12] Gushchin, A.A., Urusov, M.A.: Processes that can be embedded in a geometric Brownian motion. Theory of Probability & Its Applications **60**(2), 246–262 (2016). https://doi.org/10.1137/S0040585X97T987594

[13] Hallin, M.: On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$. ECARES Working Paper 2017-34 (2017)

[14] Hardy, G.H., Littlewood, J.E.: A maximal theorem with function-theoretic applications. Acta Math. **54**(1), 81–116 (1930). MR1555303. https://doi.org/10.1007/BF02547518

[15] Le Cam, L.: Asymptotic Methods in Statistical Decision Theory. Springer (1986). MR0856411. https://doi.org/10.1007/978-1-4612-4946-7

[16] Lehmann, E.L., Romano, J.P.: Testing Statistical Hypotheses, 3rd edn. Springer (2005). MR2135927

[17] Leskelä, L., Vihola, M.: Stochastic order characterization of uniform integrability and tightness. Stat. Probab. Lett. **83**(1), 382–389 (2013). https://doi.org/10.1016/j.spl.2012.09.023

[18] McCann, R.J.: Existence and uniqueness of monotone measure-preserving maps. Duke Math. J. **80**(2), 309–323 (1995). https://doi.org/10.1215/S0012-7094-95-08013-2

[19] Müller, A.: Orderings of risks: A comparative study via stop-loss transforms. Insurance: Mathematics and Economics **17**(3), 215–222 (1996). https://doi.org/10.1016/0167-6687(96)90002-5

[20] Müller, A., Stoyan, D.: Comparison Methods for Stochastic Models and Risks. John Wiley & Sons, Ltd., Chichester (2002). MR1889865

[21] Ogryczak, W., Ruszczyński, A.: Dual stochastic dominance and related mean-risk models. SIAM J. Optim. **13**(1), 60–78 (2002). https://doi.org/10.1137/S1052623400375075

[22] Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. Journal of Risk **2**, 21–42 (2000)

[23] Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. Journal of Banking & Finance **26**(7), 1443–1471 (2002)

[24] Rüschendorf, L.: Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios. Springer (2013). https://doi.org/10.1007/978-3-642-33590-7

[25] Shiryaev, A.N., Spokoiny, V.G.: Statistical Experiments and Decisions: Asymptotic Theory. World Scientific Publishing Co., Inc., River Edge, NJ (2000). https://doi.org/10.1142/9789812779243

[26] Skorokhod, A.V.: Studies in the Theory of Random Processes. Addison-Wesley Publishing Co., Inc., Reading, Mass. (1965). MR0185620

[27] Strasser, H.: Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory. Walter de Gruyter & Co., Berlin (1985). https://doi.org/10.1515/9783110850826

[28] Torgersen, E.: Comparison of Statistical Experiments. Cambridge University Press, Cambridge (1991). https://doi.org/10.1017/CBO9780511666353