Hardy's inequality and its descendants

a probability approach

Klaassen, C.A.J.; Wellner, J.A.

DOI
10.1214/21-EJP711

Publication date
2021

Document Version
Final published version

Published in
Electronic Journal Of Probability

Electronic Journal of Probability

Electron. J. Probab. **26** (2021), article no. 142, 1–34.
ISSN: 1083-6489 https://doi.org/10.1214/21-EJP711

Hardy's inequality and its descendants: a probability approach

Chris A. J. Klaassen* Jon A. Wellner†

Dedicated to the memory of Ron Pyke and Willem R. van Zwet

Abstract

We formulate and prove a generalization of Hardy's inequality [27] in terms of random variables and show that it contains the usual (or familiar) continuous and discrete forms of Hardy's inequality. Next we improve the recent version by Li and Mao [42] of Hardy's inequality with weights for general Borel measures and mixed norms so that it implies the discrete version of Liao [43] and the Hardy inequality with weights of Muckenhoupt [48] as well as the mixed norm versions due to Hardy and Littlewood [29], Bliss [8], and Bradley [14]. An equivalent formulation in terms of random variables is given as well. We also formulate a reverse version of Hardy's inequality, the closely related Copson inequality, a reverse Copson inequality and a Carleman-Pólya-Knopp inequality via random variables. Finally we connect our Copson inequality with counting process martingales and survival analysis, and briefly discuss other applications.

Keywords: reverse Hardy inequality; Copson's inequality; Hardy-Littlewood-Bliss inequality; Muckenhoupt's inequality; Pólya-Knopp inequality; Carleman's inequality; martingales; survival analysis.
MSC2020 subject classifications: 26D15; 60E15.
Submitted to EJP on April 28, 2021, final version accepted on September 30, 2021.

1 Introduction

The classical Hardy inequality is often presented as the following pair of inequalities: the *continuous (or integral form) inequality* says, if $p > 1$ and ψ is a nonnegative p-integrable function on $(0, \infty)$, then

$$\int_0^\infty \left(\frac{1}{x} \int_0^x \psi(y) dy \right)^p dx \leq \left(\frac{p}{p-1} \right)^p \int_0^\infty \psi^p(y)\, dy, \tag{1.1}$$

*University of Amsterdam, The Netherlands. E-mail: c.a.j.klaassen@uva.nl
†University of Washington, United States of America. E-mail: jaw@stat.washington.edu

while the *discrete (or series form) inequality* says, if $p > 1$ and $\{c_n\}_1^\infty$ is a sequence of nonnegative real numbers, then

$$\sum_{n=1}^{\infty}\left(\frac{1}{n}\sum_{k=1}^{n}c_k\right)^p \le \left(\frac{p}{p-1}\right)^p \sum_{k=1}^{\infty}c_k^p. \tag{1.2}$$

For example, see pp. 239–243 of [30], Exercises 3.14 and 3.15 of [55], [41], or [59], Chapter 9.

As Hardy [27] mentions in his Section 5, Landau pointed out that the discrete inequality follows from the integral one by noting that $c_1 \ge c_2 \ge \cdots$ may be assumed, and by choosing an appropriate step function as ψ; see Section 8 of [39].

Our main objective here is to give a unified formulation and proof of the inequalities (1.1) and (1.2) using the notation and language of probability theory. Along the way we will obtain a large family of other corollaries related to weighted Hardy inequalities (as given in [39] and in the book-length treatments [41] and [40]); see Section 2.

There is a vast literature on Hardy's inequality with weights with Muckenhoupt [48], building on [61] and [62], as a milestone. Versions of this inequality are useful in the study of differential equations ([11], [3]); the stability of stochastic processes ([17], [46]); functional inequalities, e.g. Poincaré and log-Sobolev inequalities, ([9], [10], [4], [26], and [2]).

Such versions usually involve two arbitrary Borel measures. A very recent result by Li and Mao [42] is not optimal yet, because it does not contain the discrete version as given by Liao [43]. In Section 3 we shall formulate an improvement of the result by [42] that contains the discrete version by [43] as a special case. Actually our proof of this improvement is based on the discrete result of [43]. An equivalent formulation of our version of Hardy's inequality with weights in terms of random variables will also be given.

Furthermore, we apply our methods from Section 2 to Copson's inequality ([18]) in Section 5 and to the reverse Hardy inequality in Section 4; cf. [53] and [6]. We treat reverse Copson inequalities in the same style in Section 6, and we provide a probabilistic version of the inequalities of Carleman, Pólya, and Knopp in Section 7. In Section 8 we connect our new versions of Copson's inequality formulated in probability terms with counting process martingales arising in survival analysis and reliability theory. The appendix, Section 12, elaborates on survival analysis by briefly explaining connections with the forward (and backward) versions of the Kaplan – Meier estimators appearing in right (and left) censored survival data, including a short description of the analysis of data arising from the question of "when do the baboons come down from the trees". Other applications are presented briefly in Section 11 and a summary of the new inequalities is given in Section 10. Most of the proofs are collected in Section 9.

2 Hardy's inequality

Here is our version of Hardy's inequality that implies both (1.1) and (1.2).

Theorem 2.1. Hardy's inequality
Let X and Y be independent random variables with distribution function F on $(\mathbb{R},\mathcal{B})$, and let ψ be a nonnegative measurable function on $(\mathbb{R},\mathcal{B})$. For $p > 1$

$$E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\le X]} \mid X\right)}{F(X)}\right]^p\right) \le \left(\frac{p}{p-1}\right)^p E\left(\psi^p(Y)\right) \tag{2.1}$$

holds. For continuous distribution functions F this inequality may be rewritten as

$$\int_0^1 \left[\frac{1}{u}\int_0^u \psi_F(v)dv\right]^p du \le \left(\frac{p}{p-1}\right)^p \int_0^1 \psi_F^p(v)dv \tag{2.2}$$

with $\psi_F(v) = \psi(F^{-1}(v))$, $0 < v < 1$, and for such F the constant $(p/(p-1))^p$ is the smallest possible one.

The strength of this inequality (2.1) lies in the fact that it implies both the continuous and the discrete version of Hardy's inequality.

Corollary 2.2.
(i) For any $p > 1$ and nonnegative $\psi \in L_p$, inequality (1.1) holds.
(ii) For any $p > 1$ and nonnegative sequence $\{c_n\}_{n=1}^{\infty} \in \ell_p$, inequality (1.2) holds.

Proof. (i) and (ii) follow from Theorem 2.1 by taking F to be the distribution function corresponding to the uniform probability measure on $[0, K]$ and on $\{1, \ldots, K\}$, respectively, multiplying by K, and taking limits as $K \to \infty$. □

Translating Theorem 2.1 from random variable notation back into analysis yields the following corollary.

Corollary 2.3. *For any $p > 1$, distribution function F on $\mathbb{R}$, and $\psi \in L_p(F)$ we have*

$$\int_{\mathbb{R}} |H_F\psi(x)|^p dF(x) \leq \left(\frac{p}{p-1}\right)^p \int_{\mathbb{R}} |\psi(y)|^p dF(y)$$

where H_F is the F-averaging operator defined for $x \in \mathbb{R}$ and $\psi \in L_p(F)$ by

$$H_F\psi(x) \equiv \frac{\int_{(-\infty,x]} \psi(y)dF(y)}{F(x)} = E\left(\psi(Y) \mid Y \leq x\right). \tag{2.3}$$

Note that H_F generalizes both the discrete and the continuous Hardy averaging operators; see e.g. [39], page 715. Observe that $|H_F\psi| \leq H_F|\psi|$ holds for all measurable ψ with equality if ψ is nonnegative F-a.e. This shows the equivalence of Theorem 2.1 and Corollary 2.3.

Remark 2.4. If we replace $(\mathbf{1}_{[Y \leq X]}, F(X))$ in (2.1) by $(\mathbf{1}_{[Y<X]}, F(X-))$ with the convention $0/0 = 0$, then the inequality does not hold anymore for some distribution functions with jumps. In particular, for X and Y Bernoulli with success probability $P(X = 1) = q$ and with $\psi(0) = 1,\ \psi(1) = 0$ we get

$$E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y<X]} \mid X\right)}{F(X-)}\right]^p\right) = q \tag{2.4}$$

and

$$\left(\frac{p}{p-1}\right)^p E\left(\psi^p(Y)\right) = \left(\frac{p}{p-1}\right)^p (1-q). \tag{2.5}$$

Consequently, inequality (2.1) with $(\mathbf{1}_{[Y \leq X]}, F(X))$ replaced by $(\mathbf{1}_{[Y<X]}, F(X-))$ does not hold here for

$$\frac{1}{1 + \left(1 - \frac{1}{p}\right)^p} < q < 1. \tag{2.6}$$

Remark 2.5. There are distributions for which the constant in (2.1) is not optimal for any $p > 1$. This is the case for all Bernoulli distributions. Let X and Y have a Bernoulli distribution with $P(X = 1) = q = 1 - P(X = 0)$. Then with $\psi(0) = a \geq 0$ and $\psi(1) = b \geq 0$ our Hardy inequality (2.1) becomes

$$(1-q)a^p + q\left((1-q)a + qb\right)^p \leq \left(\tfrac{p}{p-1}\right)^p \left((1-q)a^p + qb^p\right). \tag{2.7}$$

However, by convexity

$$\begin{aligned}(1-q)a^p + q\left((1-q)a+qb\right)^p &\le (1-q)a^p + q\left((1-q)a^p+qb^p\right)\\ &\le (1+q)\left((1-q)a^p+qb^p\right)\end{aligned} \tag{2.8}$$

holds. Consequently, for the Bernoulli distribution with success probability q the optimal constant in our Hardy inequality equals at most $1+q$, for which

$$1+q \le 2 < e = \inf_{p>1}\left(1+\tfrac{1}{p-1}\right)^p \tag{2.9}$$

holds.

Remark 2.6. Since $-X$ has distribution function $P(X \ge -x) = 1 - F_-(-x)$ where $F_-(x) \equiv F(x-)$ denotes the left limit of F at x, Theorem 2.1 immediately implies

$$E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\ge X]} \mid X\right)}{1-F(X-)}\right]^p\right) \le \left(\frac{p}{p-1}\right)^p E\left(\psi^p(Y)\right). \tag{2.10}$$

Note that (2.10) can be rewritten as

$$\int_{\mathbb{R}} |\overline{H}_F\psi(x)|^p dF(x) \le \left(\frac{p}{p-1}\right)^p \int_{\mathbb{R}} |\psi(y)|^p dF(y)$$

where $\overline{H}_F$ is the (right-tail) F-averaging operator defined for $x\in\mathbb{R}$ and $\psi \in L_p(F)$ by

$$\overline{H}_F\psi(x) \equiv \frac{\int_{[x,\infty)}\psi(y)dF(y)}{1-F(x-)} = E\left(\psi(Y) \mid Y \ge x\right) \equiv \Psi(x). \tag{2.11}$$

Thus

$$E\left(\psi(Y)-\psi(x) \mid Y\ge x\right) = \overline{H}_F\psi(x) - \psi(x)$$

is the "mean residual life of $\psi(Y)$" given $[Y \ge x]$. In particular, with $\psi(x) = x$,

$$E\left(Y-x \mid Y \ge x\right) \equiv \Psi(x) - \psi(x)$$

is the "mean residual life function" corresponding to the distribution function F. It turns out that for $\psi(Y) \in L_2(F)$ and F continuous

$$\mathrm{Var}(\psi(Y)) = E\left(\{\psi(Y)-\Psi(Y)\}^2\right),$$

so that the conditional centering operator $I - \overline{H}_F$ is an isometry. For more on this and connections to counting process martingales and survival analysis see [54], [22], and [7]. [60] studies $I-H$ and $I-H^*$ as operators on $L^p(\mathbb{R}^+,\lambda)$ where λ denotes Lebesgue measure.

Remark 2.7. Since the conditional distribution of X given $X \le c$ has distribution function $F(\cdot)/F(c)$ for $c\in\mathbb{R}$ and the same holds for Y, we have the following conditional version of (2.1)

$$\begin{aligned}&E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\le X]} \mid X\right)}{F(X)}\right]^p \middle| X \le c\right)\\ &= E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\le X]} \mid Y\le c, X\right)}{F(X)/F(c)}\right]^p \middle| X \le c\right)\\ &\le \left(\frac{p}{p-1}\right)^p E\left(\psi^p(Y) \mid Y \le c\right),\end{aligned} \tag{2.12}$$

where the inequality stems from (2.1) itself.

Similarly, we have

$$
\begin{aligned}
&E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\geq X]}\mid X\right)}{1-F(X-)}\right]^p \middle| X>c\right)\\
&=E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\geq X]}\mid Y>c, X\right)}{(1-F(X-))/(1-F(c))}\right]^p \middle| X>c\right) \qquad (2.13)\\
&\leq \left(\frac{p}{p-1}\right)^p E\left(\psi^p(Y)\mid Y>c\right).
\end{aligned}
$$

Together (2.12) and (2.13) improve the generalization given in Theorem 3.2 of [56] from continuous distribution functions to arbitrary distributions, namely to

$$
\begin{aligned}
&E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\leq X]}\mid X\right)}{F(X)}\right]^p \mathbf{1}_{[X\leq c]}\right)\\
&\quad + E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\geq X]}\mid X\right)}{1-F(X-)}\right]^p \mathbf{1}_{[X>c]}\right)\\
&=F(c)E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\leq X]}\mid X\right)}{F(X)}\right]^p \middle| X\leq c\right)\\
&\quad +(1-F(c))E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\geq X]}\mid X\right)}{1-F(X-)}\right]^p \middle| X>c\right) \qquad (2.14)\\
&\leq \left(\frac{p}{p-1}\right)^p \left[F(c)E\left(\psi^p(Y)\mid Y\leq c\right)+(1-F(c))E\left(\psi^p(Y)\mid Y>c\right)\right]\\
&=\left(\frac{p}{p-1}\right)^p E\left(\psi^p(Y)\right).
\end{aligned}
$$

Remark 2.8. The Hardy inequality for weighted L_p spaces on $(0,\infty)$, such as Theorem 1.2.1 of [3], also follows from our Hardy inequality for random variables. With $0\leq\varepsilon<(p-1)/p$ and K a large constant, we choose $F(x)=(x/K)^{1-\varepsilon p/(p-1)}\wedge 1,\ x\geq 0$. This results in the inequality

$$
\begin{aligned}
&\left[1-\frac{\varepsilon p}{p-1}\right]^{p+1}\int_0^K\left[\int_0^x \psi(y)y^{-\varepsilon p/(p-1)}dy\right]^p x^{p(\varepsilon-1)}dx \qquad (2.15)\\
&\leq \left[1-\frac{\varepsilon p}{p-1}\right]\left(\frac{p}{p-1}\right)^p\int_0^K \psi^p(y)y^{-\varepsilon p/(p-1)}dy.
\end{aligned}
$$

Taking limits as $K\to\infty$ and writing $\Psi(y)=\psi(y)y^{-\varepsilon p/(p-1)}$ we arrive at

$$
\int_0^\infty\left[\int_0^x\Psi(y)dy\right]^p x^{p(\varepsilon-1)}dx\leq\left[\frac{p-1}{p}-\varepsilon\right]^{-p}\int_0^\infty\Psi^p(y)y^{p\varepsilon}dy, \qquad (2.16)
$$

which is inequality (1.2.1) combined with (1.2.3) of [3]. Note that by choosing $\epsilon=0$ the inequality in the last display reduces to (1.1).

3 Hardy's inequality with weights and mixed norms

To the best of our knowledge the most recent and most general versions of Hardy's inequalities with weights and mixed norms are presented by Liao [43] and Li and Mao [42]. We shall improve the result of [42] so that it contains the discrete version of [43] as a special case. To this end we prove the result of [42] with $(-\infty,x)$ in the inner integral replaced by $(-\infty,x]$, i.e.

Theorem 3.1. Hardy's Inequality with Weights and Mixed Norms
Let $1 < p \le q < \infty$, and suppose that μ and ν are σ-finite Borel measures on $\mathbb{R}$. Then

$$\left[\int_{\mathbb{R}} \left(\int_{(-\infty,x]} \psi d\nu\right)^q d\mu(x)\right]^{1/q} \le k_{q,p}\, B \left[\int_{\mathbb{R}} \psi^p d\nu\right]^{1/p} \tag{3.1}$$

holds for all measurable $\psi : \mathbb{R} \to [0,\infty)$, where $k_{q,p}$ and B are defined by

$$B \equiv \sup_{x\in\mathbb{R}} \mu([x,\infty))^{1/q} \nu((-\infty,x])^{(p-1)/p} \tag{3.2}$$

and, with $\mathrm{Beta}(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$ and $r \equiv (q-p)/p$,

$$k_{q,p} \equiv \left(\frac{r}{\mathrm{Beta}(1/r,(q-1)/r)}\right)^{r/q} \quad \textit{and} \quad k_{p,p} = p(p-1)^{(1-p)/p}. \tag{3.3}$$

Remark 3.2. With the help of Theorem 1.4 of [43] we shall prove our Theorem 3.1 in Section 9. In fact these theorems are equivalent, since our theorem implies his. For nonnegative $a_i, u_i, v_i,\ i = 1,\dots,N$, let μ and ν be measures on $\{1,\dots,N\}$ that have densities u_i and $v_i^{-1/(p-1)}$, respectively, at i with respect to counting measure, and let $\psi(i) = a_i v_i^{1/(p-1)},\ i = 1,\dots,N$. With these choices Theorem 3.1 yields (9.20) and (9.21), and hence Theorem 1.4 of [43].

Remark 3.3. Let C be the smallest constant such that

$$\left[\int_{\mathbb{R}} \left(\int_{(-\infty,x]} \psi d\nu\right)^q d\mu(x)\right]^{1/q} \le C \left[\int_{\mathbb{R}} \psi^p d\nu\right]^{1/p} \tag{3.4}$$

holds in the situation of Theorem 3.1. With $\psi(y) = \mathbf{1}_{[y\le z]}$ this yields

$$\begin{aligned}\nu((-\infty,z])\,[\mu((z,\infty))]^{1/q} &\le \left[\int_{\mathbb{R}} (\nu((-\infty,x\wedge z]))^q\, d\mu(x)\right]^{1/q} \\ &\le C\,[\nu((-\infty,z])]^{1/p}\,,\end{aligned} \tag{3.5}$$

which implies the well known inequality $B \le C$. By Theorem 3.1 we also have $C \le k_{q,p}B$ so $C < \infty$ if and only if $B < \infty$. The constants $k_{q,p}$ first appeared via a (1923) conjecture of Hardy and Littlewood [29] which was later confirmed by Bliss [8]. See Chapter 5 of [40] for a very complete history of these developments and further results.

Theorem 3.1 and Remark 3.3 may be reformulated in terms of random variables as follows.

Theorem 3.4. Probability Version of Hardy's Inequality with Weights and Mixed Norms
Let X and Y be independent random variables with distribution functions F and G respectively, let $1 < p \le q < \infty$, and let U and V be nonnegative measurable functions on $(\mathbb{R},\mathcal{B})$. Furthermore let $C \in [0,\infty]$ be the smallest constant such that

$$\left\{E\left(\left[E\left(\tilde\psi(Y)\mathbf{1}_{[Y\le X]} \mid X\right)\right]^q U(X)\right)\right\}^{1/q} \le C\,\left\{E\left(\tilde\psi^p(Y)V(Y)\right)\right\}^{1/p} \tag{3.6}$$

holds for all nonnegative measurable functions $\tilde\psi$ on $(\mathbb{R},\mathcal{B})$. With

$$B = \sup_{x\in\mathbb{R}} \left[\int_{[x,\infty)} U dF\right]^{1/q} \left[\int_{(-\infty,x]} V^{-1/(p-1)} dG\right]^{(p-1)/p}, \tag{3.7}$$

the string of inequalities

$$B \le C \le k_{q,p} B \tag{3.8}$$

holds, even for $B = \infty$.

Proof. Theorem 3.4 is implied by Theorem 3.1 via the choices $\mu([x,\infty)) = \int_{[x,\infty)} U dF$, $\nu((-\infty, x]) = \int_{(-\infty,x]} V^{-1/(p-1)} dG$ and $\psi = \tilde{\psi} V^{1/(p-1)}$.

With μ a σ-finite measure and $\cup_{i=1}^{\infty} A_i = (0,\infty)$ a partition with $0 < \mu(A_i) < \infty, i = 1, 2, \ldots$, the measure $P(B) = \sum_{i=1}^{\infty} 2^{-i} \mu(B \cap A_i)/\mu(A_i), B \in \mathcal{B}$, is a probability measure dominating μ. Let F and G be the distribution functions of probability measures dominating the measures μ and ν, respectively, from Theorem 3.1. The choices $U(x) = d\mu/dF(x)$ and $V(y) = (d\nu/dG(y))^{1-p}$ show that Theorem 3.4 implies Theorem 3.1. $\square$

Following the arguments of Muckenhoupt [48], in Section 9 we prove the following generalization of his result, which is the special case $q = p$ of our Theorems 3.1 and 3.4.

Theorem 3.5. Probability Version of Muckenhoupt's Inequality
Let X and Y be independent random variables with distribution functions F and G respectively, let $p > 1$, and let U and V be nonnegative measurable functions on $(\mathbb{R}, \mathcal{B})$. Furthermore let $C \in [0,\infty]$ be the smallest constant such that

$$E\left(\left[E\left(\psi(Y)\mathbf{1}_{[Y \le X]} \mid X\right)\right]^p U(X)\right) \le C\, E\left(\psi^p(Y) V(Y)\right) \tag{3.9}$$

holds for all nonnegative measurable functions ψ on $(\mathbb{R}, \mathcal{B})$. With

$$B = \sup_{x \in \mathbb{R}} \int_{[x,\infty)} U dF \left[\int_{(-\infty,x]} V^{-1/(p-1)} dG\right]^{p-1}, \tag{3.10}$$

the string of inequalities

$$B \le C \le \frac{p^p}{(p-1)^{p-1}} B \tag{3.11}$$

holds, even for $B = \infty$.

Remark 3.6. With $U = G^{-p}, V = 1$ and $G = F$ the second inequality in (3.11) does *not* imply our Hardy inequality (2.1). Indeed, for Bernoulli random variables with $P(X = 1) = 1/p = 1 - P(X = 0)$ the factor B equals $1 + (p-1)^{p-1}/p^p$ then and hence the upper bound on C equals $1 + p^p/(p-1)^{p-1}$, which is larger than $(p/(p-1))^p$ for $p \ge p_0 \approx 1.77074$.
However, with $U = G^{-p}, V = 1$ and $G = F$ a continuous distribution function the factor B equals $1/(p-1)$, which shows that (3.11) *does* imply our Hardy inequality (2.1) for this case.

If X is stochastically larger than Y, $Y \preceq X$, and they have no point masses at the same location, then Theorem 3.5 yields an inequality very similar to (2.1). A comparable result is obtained for $X \preceq Y$.

Corollary 3.7. Stochastic ordering
Let X and Y be independent random variables with distribution functions F and G respectively, let $p > 1$, and let ψ be a nonnegative measurable function on $(\mathbb{R}, \mathcal{B})$.
(a) If $P(X = Y) = 0$ and $F(x) \le G(x)$, $x \in \mathbb{R}$, hold, then

$$E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y \le X]} \mid X\right)}{G(X)}\right]^p\right) \le \left(\frac{p}{p-1}\right)^p E\left(\psi^p(Y)\right) \tag{3.12}$$

is valid.

(b) If F is continuous and $F(x) \geq G(x)$, $x \in \mathbb{R}$, holds, then

$$E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\leq X]} \mid X\right)}{F(X)}\right]^p\right) \leq \left(\frac{p}{p-1}\right)^p E\left(\psi^p(Y)\right) \tag{3.13}$$

is valid.

Proof. In case (a) we apply Theorem 3.5 with $U = G^{-p}$ and $V = 1$. Then B from (3.10) equals

$$B = \sup_{r\in\mathbb{R}} G^{p-1}(r) \int_{[r,\infty)} G^{-p} dF. \tag{3.14}$$

If F has no point mass at r, then the stochastic ordering $Y \preceq X$ implies

$$\begin{aligned} &\int_{[r,\infty)} G^{-p} dF = \int_{(r,\infty)} G^{-p} dF \leq \int_{(r,\infty)} G^{-p} dG \\ &= \int_{G^{-1}(u)>r} (G(G^{-1}(u)))^{-p} du \leq \int_{[G(r),1)} (G(G^{-1}(u)))^{-p} du \\ &\leq \int_{[G(r),1)} u^{-p} du = \frac{1}{p-1}\left[G^{1-p}(r) - 1\right] \leq \frac{G^{1-p}(r)}{p-1}. \end{aligned} \tag{3.15}$$

In the first line of the last display and in the second line below we use the characterization $Y \preceq X$ if and only if $Eh(Y) \leq Eh(X)$ for all bounded and non-decreasing functions h; see e.g. [49] Theorem 1.2.8 (ii), page 5, or [57] (1.A.7), page 4.

If F has a point mass at r then G has not and the stochastic ordering $Y \preceq X$ implies

$$\begin{aligned} &\int_{[r,\infty)} G^{-p} dF \leq \int_{[r,\infty)} G^{-p} dG \\ &= \int_{(r,\infty)} G^{-p} dG = \frac{1}{p-1}\left[G^{1-p}(r) - 1\right] \leq \frac{G^{1-p}(r)}{p-1}. \end{aligned} \tag{3.16}$$

Combining (3.14)–(3.16) and (3.9)–(3.11) we arrive at (3.12).
In case (b) we apply Theorem 3.5 with $U = F^{-p}$ and $V = 1$. Then the continuity of F and $G \leq F$ imply that B from (3.10) satisfies

$$\begin{aligned} B &= \sup_{r\in\mathbb{R}} G^{p-1}(r)\left(\int_{[r,\infty)} F^{-p} dF\right) = \sup_{r\in\mathbb{R}} \frac{1}{p-1}\left[F^{1-p}(r) - 1\right] G^{p-1}(r) \\ &\leq \sup_{r\in\mathbb{R}} \frac{1}{p-1}\left[\frac{G(r)}{F(r)}\right]^{p-1} = \frac{1}{p-1} \end{aligned} \tag{3.17}$$

and hence that (3.13) holds. □

4 A reverse Hardy inequality

There are also reversed versions of the classical Hardy inequality: the *continuous (or integral form) inequality* says, if $p > 1$ and ψ is a nonnegative, nonincreasing p-integrable function on $(0,\infty)$, then

$$\int_0^\infty \left(\frac{1}{x}\int_0^x \psi(y)dy\right)^p dx \geq \frac{p}{p-1}\int_0^\infty \psi^p(y)dy, \tag{4.1}$$

while the *discrete (or series form) inequality* says, if $p > 1$ and $\{c_n\}_1^\infty$ is a nonincreasing sequence of nonnegative real numbers, then

$$\sum_{n=1}^\infty \left(\frac{1}{n}\sum_{k=1}^n c_k\right)^p \geq \zeta(p)\sum_{k=1}^\infty c_k^p. \tag{4.2}$$

Here, $\zeta(\cdot)$ is the zeta function. These inequalities have been obtained independently by Renaud [53] and Bennett [6]; see also Lemma 2.1 of [47]. By taking ψ the indicator function of the unit interval we see that (4.1) is sharp and by taking $c_1 = 1, c_2 = c_3 = \cdots = 0$ that (4.2) is sharp.

Here are our random variable versions of (4.1) and (4.2).

Theorem 4.1. Reverse Hardy inequality
Let X and Y be independent random variables both with distribution function F on $(\mathbb{R}, \mathcal{B})$, and let ψ be a nonnegative, nonincreasing function on $(\mathbb{R}, \mathcal{B})$. For $p > 1$ and F absolutely continuous

$$\begin{aligned} E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\le X]} \mid X\right)}{F(X)}\right]^p\right) \quad &\ge \quad \frac{p}{p-1} E\left(\psi^p(Y)\left[1 - F^{p-1}(Y)\right]\right) \\ &\ge \quad E\left(\psi^p(Y)\right) \end{aligned} \tag{4.3}$$

holds with equalities if ψ is constant.

For $p \ge 1$ and F general

$$E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\le X]} \mid X\right)}{F(X)}\right]^p\right) \ge E\left(\psi^p(Y)\right) \tag{4.4}$$

holds with equalities if ψ is constant.

If F is general, but $p \ge 2$ is an integer, then, with $X, Y, X_1, \dots, X_p$ independent and identically distributed and with $X_{(p)} = \max\{X_1, \dots, X_p\}$, we have

$$E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\le X]} \mid X\right)}{F(X)}\right]^p\right) \ge E\left(\psi^p(X_{(p)}) E\left(F^{-p}(Y)\mathbf{1}_{[Y\ge X_{(p)}]} \mid X_{(p)}\right)\right) \tag{4.5}$$

with equality if ψ is constant.

The continuous version (4.1) of the reverse Hardy inequality is contained in (4.3) and the discrete version (4.2) for integer p follows from (4.5).

Corollary 4.2.
(i) For any $p > 1$ and nonnegative, nonincreasing $\psi \in L_p$, inequality (4.1) holds.
(ii) For any integer $p > 1$ and nonnegative, nonincreasing sequence $\{c_n\}_{n=1}^\infty \in \ell_p$, inequality (4.2) holds.

For further developments concerning reverse Hardy type inequalities, see [24].

5 Copson's inequality

Copson [18] presented the following pair of inequalities: the *continuous (or integral form) inequality* says, if $p > 1$ and ψ is a nonnegative p-integrable function on $(0, \infty)$, then

$$\int_0^\infty \left(\int_x^\infty \frac{\psi(y)}{y} dy\right)^p dx \le p^p \int_0^\infty \psi^p(y) dy \tag{5.1}$$

holds, while the *discrete (or series form) inequality* says, if $p > 1$ and a_i and λ_i, $i = 1, 2, \dots$, are nonnegative numbers and $\Lambda_i = \sum_{j=1}^i \lambda_j$, $i = 1, 2, \dots$, is positive, then

$$\sum_{i=1}^\infty \left[\sum_{j=i}^\infty a_j \frac{\lambda_j}{\Lambda_j}\right]^p \lambda_i \le p^p \sum_{j=1}^\infty a_j^p \lambda_j \tag{5.2}$$

holds. We generalize Copson's inequalities as follows.

Theorem 5.1. Copson's inequality
Let X and Y be independent random variables with distribution function F on $(\mathbb{R}, \mathcal{B})$, and let ψ be a nonnegative measurable function on $(\mathbb{R}, \mathcal{B})$. For $p \geq 1$

$$E\left(\left[E\left(\frac{\psi(Y)}{F(Y)}\mathbf{1}_{[Y\geq X]} \mid X\right)\right]^p\right) \leq p^p E\left(\psi^p(Y)\right) \tag{5.3}$$

holds. For absolutely continuous distribution functions F the constant p^p is the smallest possible one.

The strength of this inequality (5.3) lies in the fact that it implies both the continuous and the discrete version of Copson's inequality.

Corollary 5.2.
(i) For any $p \geq 1$ and nonnegative $\psi \in L_p$, inequality (5.1) holds.
(ii) For any $p \geq 1$ and nonnegative sequences $\{a_n\}_{n=1}^{\infty}, \{\lambda_n\}_{n=1}^{\infty} \in \ell_p$ with $\lambda_1 > 0$, inequality (5.2) holds.

Proof. By Tonelli's theorem (Fubini) equality holds in (5.1) and (5.2) for $p = 1$. Let $p > 1$. (i) can be seen by choosing X and Y uniform on $(0, K)$ and taking limits with $K \to \infty$. (ii) needs a longer argument. For $p > 1$ define $\Lambda_i = \sum_{j=1}^{i} \lambda_j$, $p_i = \lambda_i/\Lambda_K$, $i = 1, \ldots, K$, for some natural number K and define the bounded continuous function ψ such that $\psi(i) = a_i$ holds for $i = 1, \ldots, K$. With $F(x) = \sum_{i=1}^{K\wedge\lfloor x\rfloor} p_i$ Theorem 5.1 yields

$$E\left(\left[E\left(\frac{\psi(Y)}{F(Y)}\mathbf{1}_{[Y\geq X]} \mid X\right)\right]^p\right) = \sum_{i=1}^{K}\left[\sum_{j=i}^{K}\frac{a_j}{\Lambda_j/\Lambda_K}p_j\right]^p p_i \tag{5.4}$$
$$= \sum_{i=1}^{K}\left[\sum_{j=i}^{K} a_j\frac{\lambda_j}{\Lambda_j}\right]^p \frac{\lambda_i}{\Lambda_K} \leq p^p \sum_{j=1}^{K} a_j^p\frac{\lambda_j}{\Lambda_K} = p^p E\left(\psi^p(Y)\right).$$

For $K_1 \leq K_2$ this implies

$$\sum_{i=1}^{K_1}\left[\sum_{j=i}^{K_2} a_j\frac{\lambda_j}{\Lambda_j}\right]^p \lambda_i \leq p^p \sum_{j=1}^{K_2} a_j^p \lambda_j. \tag{5.5}$$

Taking limits here for $K_2 \to \infty$ and subsequently $K_1 \to \infty$ we arrive at (5.2). □

Comparison of the left side of (5.3) with the left side of (2.1) and the definition of H_F in (2.3) leads us to define the Copson (or dual) operator H_F^* as follows: for $x \in \mathbb{R}$ and $\psi \in L_p(F)$

$$H_F^*\psi(x) \equiv \int_{[x,\infty)} \frac{\psi(y)}{F(y)} dF(y) = \int_{[x,\infty)} \psi(y) d\Lambda(y) \tag{5.6}$$

where $\Lambda(x) \equiv \int_{[x,\infty)} dF(y)/F(y)$ is the *reverse (or backward)* hazard function corresponding to F. (We will introduce and discuss the forward hazard function $\overline{\Lambda}(x) \equiv \int_{(-\infty,x]} dF(x)/(1 - F(x-))$ in connection with the inequalities of Carleman, Pólya, and Knopp in Section 7.)

As pointed out by Hardy in [28], the discrete Copson inequality is a "reciprocal" or "dual" inequality of the discrete Hardy inequality (1.2), in the sense that one implies the other. But this holds in other senses as well. For a treatment of (1.1) and (5.1) based on the duality of L_p and L_q with $1/p + 1/q = 1$, see [25], section 6.3, especially his Theorem

6.20 and Corollary 6.2.1. In particular when viewed as operators on $L_2(F)$, H_F and H_F^* are adjoint operators: for ψ and χ in $L_2(F)$ we have

$$\begin{aligned} E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\le X]} \mid X\right)}{F(X)}\right]\chi(X)\right) &= E\left(\frac{\psi(Y)\chi(X)}{F(X)}\mathbf{1}_{[Y\le X]}\right) \\ &= E\left(\left[E\left(\frac{\chi(X)}{F(X)}\mathbf{1}_{[Y\le X]} \mid Y\right)\right]\psi(Y)\right). \end{aligned} \tag{5.7}$$

So, H_F and H_F^* have the same norms for $p=2$, and indeed the bounds in (10.1) and (10.2) are the same for $p=2$. Applying Hardy's approach we obtain the equivalence of (2.1) and (5.3).

Theorem 5.3. Equivalence of Hardy's and Copson's inequality
Let X and Y be independent random variables with distribution function F on $(\mathbb{R},\mathcal{B})$. For $p>1$ and all nonnegative measurable functions ψ on $(\mathbb{R},\mathcal{B})$ (2.1) holds if and only if for $p>1$ and all nonnegative measurable functions ψ on $(\mathbb{R},\mathcal{B})$ (5.3) holds.

Although this Theorem 5.3 (formally) renders one of our proofs of Hardy's and Copson's inequality superfluous, we have included both proofs in Section 9 to illustrate the different methods.

Remark 5.4. For $p>1$ there are distributions for which the constant p^p in (5.3) is not optimal. This is the case for all Bernoulli distributions. Let X and Y have a Bernoulli distribution with $P(X=1)=q=1-P(X=0)$. Then with $\psi(0)=a$ and $\psi(1)=b$ the left hand side of our Copson inequality (5.3) equals

$$\begin{aligned} &(1-q)(a+qb)^p + q(qb)^p = (1-q)(1+q)^p\left(\frac{1}{1+q}a+\frac{q}{1+q}b\right)^p + q^{p+1}b^p \\ &\le (1-q)(1+q)^{p-1}\left(a^p+qb^p\right) + q^{p+1}b^p \\ &= (1+q)^{p-1}\left((1-q)a^p+qb^p\right) + q^2\left(q^{p-1}-(1+q)^{p-1}\right)b^p \\ &\le (1+q)^{p-1}\left((1-q)a^p+qb^p\right), \end{aligned} \tag{5.8}$$

where the first inequality follows from Jensen's inequality and the convexity of $x\mapsto x^p$, $x\ge 0$. The right hand side of (5.8) is bounded by

$$2^{p-1}\left((1-q)a^p+qb^p\right) < p^p\left((1-q)a^p+qb^p\right), \tag{5.9}$$

where the strict inequality holds since $p\mapsto p\log p-(p-1)\log 2$ is strictly increasing on $[1,\infty)$ with value 0 at $p=1$ and where the last expression is the upper bound in (5.3).

Remark 5.5. Theorem 5.3 gives a qualitative connection between Hardy's inequality and Copson's inequality (or the "dual Hardy inequality"). The papers by [38], [35], and [36] quantify these connections. These results are strongly related to further work on the connections between the $I-H_F$ and $I-H_F^*$ operators on the one hand, and between the $I-\overline{H}_F$ and $I-\overline{H}_F^*$ operators on the other hand. Also see [12]. Recall that

$$\begin{aligned} H_F\psi(x) &\equiv \frac{\int_{(-\infty,x]}\psi(y)dF(y)}{F(x)}, \quad \overline{H}_F\psi(x) \equiv \frac{\int_{[x,\infty)}\psi(y)dF(y)}{1-F(x-)}, \\ H_F^*\psi(x) &\equiv \int_{[x,\infty)}\frac{\psi(y)}{F(y)}dF(y) \quad \overline{H}_F^*\psi(x) \equiv \int_{(-\infty,x]}\frac{\psi(y)}{1-F(y-)}dF(y) \\ &= -\int_{[x,\infty)}\psi(y)d\Lambda(y), \qquad\qquad = \int_{(-\infty,x]}\psi(y)d\overline{\Lambda}(y), \end{aligned}$$

where

$$\Lambda(x) \equiv \int_{[x,\infty)}\frac{dF(y)}{F(y)}, \qquad \overline{\Lambda}(x) \equiv \int_{(-\infty,x]}\frac{1}{1-F(y-)}dF(y). \tag{5.10}$$

are the *backward* cumulative hazard function and the (*forward*) cumulative hazard functions of survival analysis.

6 A reverse Copson inequality

Reversed versions of the classical Copson inequality are given in Theorems 2 and 4 of Renaud (1986) [53]. His continuous (or integral form) inequality may be rephrased as follows. If $p \geq 1$ holds and ψ is a nonnegative p-integrable function on $(0,\infty)$ such that $x \mapsto \psi(x)/x$ is nonincreasing, then

$$\int_0^\infty \left(\int_x^\infty \frac{\psi(y)}{y} dy \right)^p dx \geq \int_0^\infty \psi^p(y) dy \tag{6.1}$$

holds. His discrete form says: if $p \geq 1$ holds and $a_1/1 \geq a_2/2 \geq \cdots$ are nonnegative numbers, then

$$\sum_{i=1}^\infty \left[\sum_{j=i}^\infty \frac{a_j}{j} \right]^p \geq \sum_{i=1}^\infty a_i^p \tag{6.2}$$

holds.

It seems natural to consider a reverse Copson inequality formulated in terms of random variables. Here is our result in this direction.

Theorem 6.1. Reverse Copson inequality
Let X and Y be independent random variables both with distribution function F on $(\mathbb{R},\mathcal{B})$ and let ψ be a nonnegative p-integrable function on $(\mathbb{R},\mathcal{B})$ with $p \in [1,\infty)$. If the distribution function F is continuous and $x \mapsto \psi(x)/F(x)$ is nonincreasing, then

$$E\left(\left[E\left(\frac{\psi(Y)}{F(Y)} \mathbf{1}_{[Y \geq X]} \mid X \right) \right]^p \right) \geq E\left(\psi^p(Y)\right) \tag{6.3}$$

holds with equality if $\psi = F$ or $p = 1$ holds.

If the distribution function F is continuous, ψ is nonincreasing, and p is an integer, then

$$E\left(\left[E\left(\frac{\psi(Y)}{F(Y)} \mathbf{1}_{[Y \geq X]} \mid X \right) \right]^p \right) \geq p!\, E\left(\psi^p(Y)\right) \tag{6.4}$$

holds with equality if ψ is constant or $p = 1$ holds.

If the distribution function F is arbitrary, ψ is nonincreasing, and p is an integer, then

$$E\left(\left[E\left(\frac{\psi(Y)}{F(Y)} \mathbf{1}_{[Y \geq X]} \mid X \right) \right]^p \right) \geq E\left(\psi^p(Y)\right) \tag{6.5}$$

holds with equality if ψ equals 0, or F is degenerate (i.e. F is concentrated at one point), or $p = 1$ holds.

We conjecture that (6.4), with $p!$ replaced by $\Gamma(p+1)$, and (6.5) hold for all $p \geq 1$, but we have no proof. Note that for F continuous (6.5) with $p \in [1,\infty)$ follows from (6.3). For the situations of the continuous and discrete versions of the original Copson inequality our reverse Copson inequality implies:

Corollary 6.2.
(i) With $p \in [1,\infty)$ and ψ nonnegative p-integrable on $(0,\infty)$ such that $x \mapsto \psi(x)/x$ is nonincreasing (6.1) holds.
(ii) If $p \geq 1$ is an integer and ψ is a nonnegative, nonincreasing, p-integrable function on $(0,\infty)$, then

$$\int_0^\infty \left(\int_x^\infty \frac{\psi(y)}{y} dy \right)^p dx \geq p! \int_0^\infty \psi^p(y) dy \tag{6.6}$$

holds.
(iii) If $p \geq 1$ is an integer and $a_1 \geq a_2 \geq \cdots$ and λ_i, $i = 1, 2, \dots$, are nonnegative numbers and $\Lambda_i = \sum_{j=1}^{i} \lambda_j$, $i = 1, 2, \dots$, is positive, then

$$\sum_{i=1}^{\infty} \left[\sum_{j=i}^{\infty} a_j \frac{\lambda_j}{\Lambda_j} \right]^p \lambda_i \geq \sum_{j=1}^{\infty} a_j^p \lambda_j \tag{6.7}$$

holds.

The proof of this corollary is almost the same as the proof of Corollary 5.2 in Section 5 (but with the inequality signs reversed and the constants changed), and therefore it is omitted.

Remark 6.3. Without continuity of F inequality (6.3) is not generally valid for $p > 1$. Again a counterexample is provided by the Bernoulli distribution. Take $\psi = F$ and $F(x) = (1-q)\mathbf{1}_{[x \geq 0]} + q\mathbf{1}_{[x \geq 1]}$. Now, as a function of the success probability q the left minus the right hand side of (6.3) equals

$$E\left([1 - F(X-)]^p\right) - E\left([F(X)]^p\right) = (1 - 2q) + [q^p - (1-q)^p], \tag{6.8}$$

which takes on both positive and negative values on $[0, 1]$ for $p \neq 2$.

7 The Carleman and Pólya - Knopp inequalities

Another classical pair of inequalities in this family of inequalities are those associated with the names of Pólya and Knopp in the continuous (or integral) case, and Carleman in the discrete case: for a positive function ψ in $L_1(\mathbb{R}^+, \lambda)$,

$$\int_0^\infty \exp\left(\frac{1}{x} \int_0^x \log \psi(y) dy\right) dx \leq e \cdot \int_0^\infty \psi(y) dy \tag{7.1}$$

and, for a sequence of constants $\{c_k\}$,

$$\sum_{k=1}^{\infty} \left(\prod_{j=1}^{k} c_j \right)^{1/k} \leq e \cdot \sum_{j=1}^{\infty} c_j; \tag{7.2}$$

see e.g. [39] section 9, [31], [32], and [50]. By now the reader will anticipate our impulse to reformulate and unify these two inequalities in a more probabilistic vein involving random variables and distribution functions as follows:

Theorem 7.1. *Let ψ be a positive valued function on $\mathbb{R}$ and let X, Y be independent random variables with distribution function F. If $\psi \in L_1(F)$ then*

$$E\left\{ \exp\left(\frac{E\left(1_{[Y \leq X]} \log \psi(Y) | X\right)}{F(X)} \right) \right\} \leq e \cdot E\psi(Y).$$

Corollary 7.2.
(i) For any nonnegative $\psi \in L_1$, inequality (7.1) holds.
(ii) For any positive sequence $\{c_k\} \in \ell_1$ the inequality (7.2) holds

The proof of Corollary 2.2 is applicable to Corollary 7.2 as well.

Kaijser et al. [33] rewrite the classical integral version of the Carleman inequality as follows: replacing $\psi(y)$ in (7.1) by $\psi(y)/y$ yields

$$\int_0^\infty \exp\left(\frac{1}{x} \int_0^x \log \psi(y) dy\right) \frac{dx}{x} \leq \int_0^\infty \psi(x) \frac{dx}{x}. \tag{7.3}$$

This follows by elementary manipulations together with the identity $\int_0^x \log y dy = x(\log x - 1)$. [33] prove (7.1) with strict inequality by proving (7.3) with strict inequality via the following simple convexity argument. By convexity of exp, it follows from Jensen's inequality followed by Fubini's theorem that

$$\begin{aligned}\int_0^\infty \exp\left(\frac{1}{x}\int_0^x \log\psi(y)dy\right)\frac{dx}{x} &\leq \int_0^\infty \frac{1}{x^2}\left\{\int_0^x \psi(y)dy\right\}dx\\ &= \int_0^\infty \psi(y)\left\{\int_y^\infty \frac{1}{x^2}dx\right\}dy = \int_0^\infty \psi(y)\frac{dy}{y}.\end{aligned}$$

Strict inequality follows because equality in Jensen's inequality almost everywhere forces ψ to be constant a.e., but this contradicts finiteness of $\int_0^\infty \psi(y)/y\ dy$.

Now several questions arise: is there a corresponding rewrite of our probabilistic version of the inequalities of Carleman and Pólya – Knopp? The answer is clearly "yes" for continuous distribution functions F. Replacing ψ by ψ/F in (7.1) and arguing as above, but using the identity $\int_{(-\infty,x]} \log F(y)dF(y) = F(x)(\log F(x) - 1)$, yields

$$\int_{\mathbb{R}} \exp\left(\frac{1}{F(x)}\int_{(-\infty,x]} \log\psi(y)dF(y)\right)\frac{dF(x)}{F(x)} < \int_{\mathbb{R}} \psi(y)\frac{dF(y)}{F(y)} = \int_{\mathbb{R}} \psi(y)d(-\Lambda(y))$$

where $\Lambda(x) \equiv \int_{[x,\infty)} dF(y)/F(y)$. This is a "left tail inequality" with motivations from survival analysis.

For the corresponding "right tail inequality" we instead replace ψ by $\psi/(1-F)$. Then reasoning as above yields, for continuous F,

$$\int_{\mathbb{R}} \exp\left(\frac{1}{1-F(x-)}\int_{[x,\infty)} \log\psi(y)dF(y)\right)d\overline{\Lambda}(x) \leq \int_{\mathbb{R}} \psi(y)d\overline{\Lambda}(y)$$

where $\overline{\Lambda}(x) \equiv \int_{(-\infty,x]} dF(y)/(1-F(y-))$.

Note: This notation goes against the classical notation of survival analysis but is in keeping with the current notation of our paper. The usual notation for the "right side" or forward cumulative hazard function is simply $\Lambda(x) = \int_{(-\infty,x]} dF(y)/(1-F(y-))$.

8 Martingale connections and the H operators

In this section we expand on the comments in Sections 2, 5, and 7 concerning martingales, counting processes, and the residual life and dual Hardy operators.

First recall the operators H_F, $\overline{H}_F$, H_F^* and $\overline{H}_F^*$ introduced in Section 5. With I the identity operator and F the *continuous* distribution function of X, Fubini's theorem yields

$$(I - H_F)(I - H_F^*)\psi = \psi, \quad (I - H_F^*)(I - H_F)\psi = \psi - E\psi(X). \tag{8.1}$$

We will also need the classical Hardy operators H and H^* defined by

$$H\psi(x) \equiv \frac{1}{x}\int_0^x \psi(y)dy, \quad \text{and} \quad H^*\psi(x) \equiv \int_x^\infty \frac{\psi(y)}{y}dy.$$

for $\psi \in L_p(\mathbb{R}_+, \lambda)$ where λ denotes Lebesgue measure. Krugliak et al. [37] (see also [38]), showed that

$$(I - H)^{-1}\psi(x) = \psi(x) - H^*\psi(x) = \psi(x) - \int_x^\infty \frac{\psi(y)}{y}dy. \tag{8.2}$$

It is well known (see e.g. [16]) that $I - H$ is an isometry on $L_2(\mathbb{R}_+, \lambda)$.

[54] showed that $R \equiv I - H_F$ is an isometry of $L_2(\mathbb{R}, F)$; see also [7] Appendix A.1, pages 420 – 424. These authors also showed that with $R \equiv I - \overline{H}_F$ and $L \equiv I - \overline{H}_F^*$ we have

$$R \circ L\psi = \psi \ \text{ and } \ L \circ R\psi = \psi - E_F\psi(X)$$

for $\psi \in L_2(F)$. Thus $R^{-1} = L$ on $L_2^0(F) \equiv \{\psi \in L_2(F) : \ E_F\psi(X) = 0\}$, and we see that the analogue of the identity (8.2) becomes

$$\begin{aligned} L\psi(x) &= R^{-1}\psi(x) = (I - \overline{H}_F)^{-1}\psi(x) \\ &= (I - \overline{H}_F^*)\psi(x) = \psi(x) - \int_{(-\infty,x]} \psi(y) d\overline{\Lambda}(y) \end{aligned} \tag{8.3}$$

where $\overline{\Lambda}$ is as defined in (5.10).

To see that this is fundamentally linked to counting process martingales, let X have distribution function F on $\mathbb{R}_+$, and define a one-jump counting process $\{\mathbb{N}(t) : \ t \geq 0\}$ by

$$\mathbb{N}(t) = 1_{[X \leq t]}.$$

This process is (trivially) seen to be nondecreasing in t with probability 1, and hence is a sub-martingale (a process increasing in conditional mean). By the Doob-Meyer decomposition theorem there is an increasing predictable process $\{\mathbb{A}(t) : \ t \geq 0\}$ such that

$$\mathbb{N}(t) = \mathbb{M}(t) + \mathbb{A}(t)$$

where $\{\mathbb{M}(t) : \ t \geq 0\}$ is a mean-0 martingale. In fact for this simple counting process it is well-known that

$$\mathbb{A}(t) = \int_{[0,t]} 1_{[X \geq s]} d\overline{\Lambda}(s)$$

(see e.g. Appendix B of [58], or Chapter 18 of [44]), and hence we see that

$$\mathbb{M}(t) = \mathbb{N}(t) - \int_{[0,t]} 1_{[X \geq s]} d\overline{\Lambda}(s).$$

Comparing this with the identity (8.3) rewritten for a distribution function F on $\mathbb{R}_+$ we see that with $\psi_t(x) = 1_{[x \leq t]}$ and evaluating the resulting identity at $x = X$ we get

$$L\psi_t(X) = 1_{[X \leq t]} - \int_0^t 1_{[X \geq y]} d\overline{\Lambda}(y) = \mathbb{M}(t)$$

where $\overline{\Lambda}(x) \equiv \int_{[0,x]} (1 - F(y-))^{-1} dF(y)$ is the cumulative hazard function corresponding to F on $\mathbb{R}_+$.

But there are still more martingales in this setting which can be represented in terms of the martingale $\mathbb{M}$ by bringing in the residual life operator $R = I - \overline{H}_F$. Consider the increasing family of σ-fields $\{\mathcal{F}_t : \ t \geq 0\}$ given by $\mathcal{F}_t \equiv \sigma\{1_{[X \leq s]} : \ 0 \leq s \leq t\}$. Now let $\psi \in L_2^0(F)$ and consider the process

$$\mathbb{Y}(t) \equiv E\{\psi(X)|\mathcal{F}_t\}, \qquad t \geq 0.$$

Since the σ-fields $\{\mathcal{F}_t\}_{t \geq 0}$ are nested, $\{\mathbb{Y}(t) : \ t \geq 0\}$ is a martingale (and it is often called "Doob's martingale"). Furthermore, it can be represented in terms of the basic martingale $\mathbb{M}$ using the fundamental identity $L \circ R = I$ on $L_2^0(F)$ discussed above: since $\psi = L \circ R\psi$ we see that

$$\mathbb{Y}(t) = E\{\psi|\mathcal{F}_t\} = E\{L \circ R\psi|\mathcal{F}_t\} = \int_{[0,t]} R\psi(s) d\mathbb{M}(s).$$

This set of connections deserves to be explored further. In particular we conjecture that many of the interesting properties of the classical Hardy operator H and the dual Hardy operator H^* established in the series of papers by [37], [38], [12], [35], [13], [36], and [60] will have useful analogues for $\overline{H}_F$ and $\overline{H}_F^*$ in the probability setting for Hardy's inequalities which we have considered here. On the other hand, the martingale connections of the operators L and R perhaps deserve to be better known in the world of classical Hardy type inequalities.

For further explanation of the connections of these processes with right and left censored data problems in survival analysis, see the Appendix, Section 12.

If $X_1, \ldots, X_n$ are i.i.d. with (continuous distribution function) F, then

$$\mathbb{N}_n(t) \equiv \sum_{i=1}^n 1_{[X_i \le t]} = n\mathbb{F}_n(t)$$

is a counting process which is simply the sum of independent counting processes and the sum of the corresponding counting process martingales is again a counting process martingale:

$$\mathbb{M}_n(t) \equiv \sum_{i=1}^n \mathbb{M}_i(t) = \mathbb{N}_n(t) - \int_0^t \mathbb{Y}_n(s)\overline{\Lambda}(s)$$

where $\mathbb{Y}_n(t) \equiv \sum_{i=1}^n 1_{[X_i \ge t]}$ is the number of X_i's "at risk" at time t.

9 Proofs

9.1 Proofs for Section 2

In order to prove our random variable version of Hardy's inequality we need a Lemma. The proof of this Lemma has the same structure as Broadbent's proof of Hardy's inequality (1.2), which is a slightly improved version of Elliot's proof; see [15], [23], and [30], page 240.

Lemma 9.1. *Let a_i and p_i be nonnegative numbers for $i = 1, \ldots, m$, with $p_1 > 0$. For $p > 1$ the inequality*

$$\sum_{n=1}^m \left(\frac{\sum_{i=1}^n a_i p_i}{\sum_{i=1}^n p_i} \right)^p p_n \le \left(\frac{p}{p-1} \right)^p \sum_{n=1}^m a_n^p p_n \tag{9.1}$$

holds.

With $p_i = 1$ this inequality is a finite sum version of the discrete Hardy inequality (1.2). Taking limits as $m \to \infty$ first on the right hand side and subsequently on the left hand side of (9.1) with $p_i = 1$ we obtain the discrete Hardy inequality itself.

Proof of Lemma 9.1. With the notation $P_n = \sum_{i=1}^n p_i$, $A_n = \sum_{i=1}^n a_i p_i$, $B_n = A_n/P_n$, $n = 1, \ldots, m$,
$A_0 = B_0 = P_0 = 0$ we rewrite

$$a_n p_n B_n^{p-1} = (A_n - A_{n-1})\, B_n^{p-1} = (P_n B_n - P_{n-1} B_{n-1})\, B_n^{p-1} \tag{9.2}$$

into

$$P_n B_n^p = a_n p_n B_n^{p-1} + P_{n-1} B_{n-1} B_n^{p-1}. \tag{9.3}$$

By Young's inequality ($uv \le u^p/p + v^{p'}/p'$ with $1/p + 1/p' = 1$ and $p, p', u, v \ge 0$), this implies

$$P_n B_n^p \le a_n p_n B_n^{p-1} + P_{n-1} \left(\frac{1}{p} B_{n-1}^p + \frac{p-1}{p} B_n^p \right) \tag{9.4}$$

and hence

$$\left(P_n - \frac{p-1}{p}P_{n-1}\right)B_n^p \le a_n p_n B_n^{p-1} + \frac{1}{p}P_{n-1}B_{n-1}^p. \tag{9.5}$$

Summing this inequality over n we obtain

$$\sum_{n=1}^m P_n B_n^p - \frac{p-1}{p}\sum_{n=1}^m P_{n-1}B_n^p \le \sum_{n=1}^m a_n p_n B_n^{p-1} + \frac{1}{p}\sum_{n=1}^{m-1} P_n B_n^p, \tag{9.6}$$

which is equivalent to

$$\frac{1}{p}P_m B_m^p + \frac{p-1}{p}\sum_{n=1}^m \left(P_n - P_{n-1}\right)B_n^p \le \sum_{n=1}^m a_n p_n B_n^{p-1}. \tag{9.7}$$

By Hölder's inequality this yields

$$\frac{p-1}{p}\sum_{n=1}^m p_n B_n^p \le \left(\sum_{n=1}^m a_n^p p_n\right)^{1/p}\left(\sum_{n=1}^m B_n^p p_n\right)^{(p-1)/p} \tag{9.8}$$

and hence

$$\left(\sum_{n=1}^m B_n^p p_n\right)^{1/p} \le \frac{p}{p-1}\left(\sum_{n=1}^m a_n^p p_n\right)^{1/p} \tag{9.9}$$

and (9.1). □

Proof of Theorem 2.1. Let $F^{-1}(u) \equiv \inf\{x : \; F(x) \ge u\}$ denote the quantile function corresponding to F. For large N we define $y_{N,i} = F^{-1}(i/N), \; i = 0, \dots, N-1, \; y_{N,N} = \infty$, and we apply Lemma 9.1 with $m = N$ and

$$p_n = \int_{(y_{N,n-1}, y_{N,n}]} dF, \quad a_n = \int_{(y_{N,n-1}, y_{N,n}]} \psi dF/p_n, \quad n = 1, \dots, N. \tag{9.10}$$

By Jensen's inequality we have

$$a_n^p \le \int_{(y_{N,n-1}, y_{N,n}]} \psi^p dF/p_n, \quad n = 1, \dots, N, \tag{9.11}$$

and hence

$$\sum_{n=1}^N a_n^p p_n \le \sum_{n=1}^N \int_{(y_{N,n-1}, y_{N,n}]} \psi^p dF = E\left(\psi^p(Y)\right). \tag{9.12}$$

For any $x \in \mathbb{R}$ there exists an index $n(N,x)$ with $x \in (y_{N,n(N,x)-1}, y_{N,n(N,x)}]$. Consequently we have

$$\sum_{n=1}^N \left(\int_{(-\infty, y_{N,n}]} \psi dF/F(y_{N,n})\right)^p \mathbf{1}_{(y_{N,n-1}, y_{N,n}]}(x) \tag{9.13}$$
$$= \left(\int_{(-\infty, y_{N,n(N,x)}]} \psi dF/F(y_{N,n(N,x)})\right)^p \ge \left(\int_{(-\infty, x]} \psi dF/F(y_{N,n(N,x)})\right)^p$$

and hence by Tonelli's theorem, Fatou's lemma and the right continuity of F

$$\liminf_{N\to\infty} \sum_{n=1}^{N} \left(\frac{\sum_{i=1}^{n} a_i p_i}{\sum_{i=1}^{n} p_i} \right)^p p_n \tag{9.14}$$
$$= \liminf_{N\to\infty} \sum_{n=1}^{N} \int_{\mathbb{R}} \left(\int_{(-\infty, y_{N,n}]} \psi dF / F(y_{N,n}) \right)^p \mathbf{1}_{(y_{N,n-1}, y_{N,n}]}(x) dF(x)$$
$$= \liminf_{N\to\infty} \int_{\mathbb{R}} \sum_{n=1}^{N} \left(\int_{(-\infty, y_{N,n}]} \psi dF / F(y_{N,n}) \right)^p \mathbf{1}_{(y_{N,n-1}, y_{N,n}]}(x) dF(x)$$
$$\geq \int_{\mathbb{R}} \liminf_{N\to\infty} \sum_{n=1}^{N} \left(\int_{(-\infty, y_{N,n}]} \psi dF / F(y_{N,n}) \right)^p \mathbf{1}_{(y_{N,n-1}, y_{N,n}]}(x) dF(x)$$
$$\geq \int_{\mathbb{R}} \liminf_{N\to\infty} \left(\int_{(-\infty, x]} \psi dF / F(y_{N,n(N,x)}) \right)^p dF(x)$$
$$= \int_{\mathbb{R}} \left(\int_{(-\infty, x]} \psi dF / F(x) \right)^p dF(x) = E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y \leq X]} \mid X\right)}{F(X)} \right]^p \right).$$

Combining (9.14), Lemma 9.1 and (9.12) we arrive at a proof of (2.1) from Theorem 2.1.

Let U be uniformly distributed on the unit interval. Since $F^{-1}(U)$ has distribution function F and $F(F^{-1}(u)) = u$ holds for continuous F, inequality (2.1) can be rewritten as (2.2). With $0 < \varepsilon$ small we choose $\psi_F(u) = u^{-(1-\varepsilon)/p}, 0 < u < 1$, and we see that (2.2) is equivalent to the inequality $(p/(p-1+\varepsilon))^p \leq (p/(p-1))^p$. Since $0 < \varepsilon$ may be chosen arbitrarily small, this proves the optimality of the constant in (2.1) and (2.2). □

9.2 Proofs for Section 3

Proof of Theorem 3.1. If B equals infinity, inequality (3.1) is trivial. So, we may assume that B is finite and hence for any $r \in \mathbb{R}$ that $\mu([r,\infty)) = \infty$ implies $\nu((-\infty, r]) = 0$. Define

$$\mathcal{R} = \{r \,:\, \mu([r,\infty)) < \infty,\ r \in \mathbb{R}\}, \quad R_0 = \inf \mathcal{R} \tag{9.15}$$

and choose $R \geq R_0$. If $\mathcal{R} = [R_0, \infty)$ holds, then without loss of generality we may assume that μ is a finite Borel measure and we take $R = R_0$. However, if $\mathcal{R} = (R_0, \infty)$ holds, then we have $\mu([R_0, \infty)) = \infty$ and we take $R > R_0$. Furthermore, define

$$S_0 = \sup\{s \,:\, \mu([s,\infty)) > 0,\ s \in \mathbb{R}\} \tag{9.16}$$

and note that $S_0 = \infty$ might hold. If $S_0 = -\infty$ holds, μ is the null measure and inequality (3.1) is trivial. Let $S \leq S_0$ be such that $M_S = \mu([S,\infty)) > 0$ holds.

We introduce the finite measure $\mu_{R,S}$ that has no mass on $(-\infty, R) \cup (S, \infty)$, equals μ on the interval $[R, S)$ and has mass M_S at the point S. It has total mass $M_{R,S} = \mu([R,\infty))$ and "scaled" distribution function

$$F_{R,S}(x) = \mu([R,x])/M_{R,S} \mathbf{1}_{[x<S]} + \mathbf{1}_{[x \geq S]}, \quad x \in \mathbb{R}, \tag{9.17}$$

with inverse

$$F_{R,S}^{-1}(u) = \inf\{x \,:\, F_{R,S}(x) \geq u\}, \quad u \in [0,1]. \tag{9.18}$$

For $0 < \varepsilon < 1$ we define $\delta = \varepsilon M_S/(M_{R,S} \vee 1)$. With $N = \lceil 1/\delta \rceil$ we choose

$$y_n = F_{R,S}^{-1}(n/N),\ 1 \leq n \leq N-1, \quad y_0 = R, \quad y_N = \infty. \tag{9.19}$$

Note that (y_{n-1}, y_n) might be empty, i.e. $y_{n-1} = y_n$.

In view of $1/N \le \delta = \varepsilon M_S/(M_{R,S} \vee 1) < M_S = \mu_{R,S}(\{S\})$ we have $y_{N-1} = S$ and hence $\mu_{R,S}((y_{N-1},\infty)) = \mu_{R,S}((S,\infty)) = 0$.

By Theorem 1.4 of [43] we have for nonnegative $a_i, u_i, v_i,\ i = 1,\dots,N$,

$$\left\{\sum_{n=1}^{N}\left(\sum_{i=1}^{n} a_i\right)^q u_n\right\}^{1/q} \le k_{q,p}\, B_{\mathrm{d}} \left\{\sum_{i=1}^{N} a_i^p v_i\right\}^{1/p} \tag{9.20}$$

with

$$B_{\mathrm{d}} = \max_{1\le n\le N}\left(\sum_{j=n}^{N} u_j\right)^{1/q}\left(\sum_{i=1}^{n} v_i^{-1/(p-1)}\right)^{(p-1)/p}. \tag{9.21}$$

With $R = y_0 \le y_1 \le \cdots \le y_N = \infty$ as in (9.19) we choose $a_i = \int_{(y_{i-1},y_i]}\psi d\nu$, $u_i = \int_{(y_{i-1},y_i]} d\mu_{R,S},\ v_i = \left(\int_{(y_{i-1},y_i]} d\nu\right)^{1-p}$ with $v_i = 0$ if $\int_{(y_{i-1},y_i]} d\nu = 0,\ i = 2,\dots,N$, and $a_1 = \int_{[R,y_1]}\psi d\nu,\ u_1 = \int_{[R,y_1]} d\mu_{R,S},\ v_1 = \left(\int_{[R,y_1]} d\nu\right)^{1-p}$ with $v_1 = 0$ if $\int_{([R,y_1]} d\nu = 0$. With these choices the left hand side of (9.20) to the power q satisfies

$$\begin{aligned}
\sum_{n=1}^{N}\left(\sum_{i=1}^{n} a_i\right)^q u_n &= \sum_{n=2}^{N}\int_{(y_{n-1},y_n]}\left(\int_{[R,y_n]}\psi d\nu\right)^q d\mu_{R,S} \\
&\quad + \int_{[R,y_1]}\left(\int_{[R,y_1]}\psi d\nu\right)^q d\mu_{R,S} \qquad (9.22)\\
&\ge \sum_{n=2}^{N}\int_{(y_{n-1},y_n]}\left(\int_{[R,x]}\psi d\nu\right)^q d\mu_{R,S}(x) + \int_{[R,y_1]}\left(\int_{[R,x]}\psi d\nu\right)^q d\mu_{R,S}(x)\\
&= \int_{[R,\infty)}\left(\int_{[R,x]}\psi d\nu\right)^q d\mu_{R,S}(x) = \int_{\mathbb{R}}\left(\int_{[R,x]}\psi d\nu\right)^q d\mu_{R,S}(x)\\
&= \int_{\mathbb{R}}\left(\int_{[R,x\wedge S]}\psi d\nu\right)^q d\mu_{R,S}(x) = \int_{\mathbb{R}}\left(\int_{[R,x\wedge S]}\psi d\nu\right)^q d\mu(x).
\end{aligned}$$

Furthermore, by Jensen's inequality (or Hölder) the third factor at the right hand side of (9.20) to the power p satisfies

$$\begin{aligned}
\sum_{i=1}^{N} a_i^p v_i &= \sum_{i=2}^{N}\left(\int_{(y_{i-1},y_i]}\psi d\nu\right)^p v_i + \left(\int_{[R,y_1]}\psi d\nu\right)^p v_1 \qquad (9.23)\\
&\le \sum_{i=2}^{N}\int_{(y_{i-1},y_i]}\psi^p d\nu\left(\int_{(y_{i-1},y_i]} d\nu\right)^{p-1} v_i + \int_{[R,y_1]}\psi^p d\nu\left(\int_{[R,y_1]} d\nu\right)^{p-1} v_1\\
&= \int_{[R,\infty)}\psi^p d\nu \le \int_{\mathbb{R}}\psi^p d\nu,
\end{aligned}$$

where the last expression equals the third factor at the right hand side of (3.1) to the power p. With these choices B_{d} from (9.21) becomes

$$\begin{aligned}
B_{\mathrm{d}} &= B(y_0,\dots,y_N)\\
&= \max\Big\{\left(\mu_{R,S}([R,\infty))^{1/q}\left(\nu(-\infty,y_1]\right)\right)^{(p-1)/p}, \qquad (9.24)\\
&\qquad \max_{2\le n\le N}\left(\mu_{R,S}((y_{n-1},\infty))^{1/q}\left(\nu(-\infty,y_n]\right)\right)^{(p-1)/p}\Big\}.
\end{aligned}$$

For $2 \le n \le N-1$ we have

$$\begin{aligned}&\mu_{R,S}((y_{n-1},\infty)) = \mu_{R,S}([y_n,\infty)) + \mu_{R,S}((y_{n-1},y_n))\\&\le \mu_{R,S}([y_n,\infty))\left[1+\frac{\mu_{R,S}((y_{n-1},y_n))}{M_S}\right] \le \mu_{R,S}([y_n,\infty))\left[1+\frac{M_{R,S}}{NM_S}\right]\\&\le \mu_{R,S}([y_n,\infty))[1+\varepsilon]\end{aligned} \tag{9.25}$$

and analogously we obtain

$$\mu_{R,S}([R,\infty)) \le \mu_{R,S}([y_1,\infty))[1+\varepsilon]. \tag{9.26}$$

This implies that B_{d} from (9.24) becomes [recall $\mu_{R,S}((y_{N-1},\infty)) = 0$]

$$\begin{aligned}&B_{\mathrm{d}} = B(y_0,\ldots,y_N)\\&\le [1+\varepsilon]\max_{1\le n\le N-1}\left(\mu_{R,S}([y_n,\infty))\right)^{1/q}\left(\nu((-\infty,y_n])\right)^{(p-1)/p}\\&\le [1+\varepsilon]\sup_{R\le x\le S}\left(\mu([x,\infty))\right)^{1/q}\left(\nu((-\infty,x])\right)^{(p-1)/p} \le [1+\varepsilon]B,\end{aligned} \tag{9.27}$$

where B is as in (3.2). Since ε may be chosen arbitrarily close to 0, this implies together with (9.17) through (9.23) that inequality (3.1) holds with the left hand side replaced by the right hand side of (9.22) to the power $1/q$.

In the case of $R > R_0$ we have $\mu([R_0,\infty)) = \infty$ and hence $\nu((-\infty,R_0]) = 0$ and monotone convergence shows that the right hand side of (9.22) satisfies

$$\begin{aligned}\lim_{R\downarrow R_0}\int_{\mathbb{R}}\left(\int_{[R,x\wedge S]}\psi d\nu\right)^q d\mu(x) &= \int_{\mathbb{R}}\left(\int_{(R_0,x\wedge S]}\psi d\nu\right)^q d\mu(x)\\&= \int_{\mathbb{R}}\left(\int_{(-\infty,x\wedge S]}\psi d\nu\right)^q d\mu(x).\end{aligned} \tag{9.28}$$

In the case of $R = R_0$ we have $\nu((-\infty,R_0)) = 0$ and hence the right hand side of (9.22) equals

$$\int_{\mathbb{R}}\left(\int_{[R_0,x\wedge S]}\psi d\nu\right)^q d\mu(x) = \int_{\mathbb{R}}\left(\int_{(-\infty,x\wedge S]}\psi d\nu\right)^q d\mu(x). \tag{9.29}$$

In the case of $\mu([S_0,\infty)) = \mu(\{S_0\}) > 0$ we may choose $S = S_0$ and the right hand side of (9.29) equals

$$\int_{\mathbb{R}}\left(\int_{(-\infty,x\wedge S_0]}\psi d\nu\right)^q d\mu(x) = \int_{\mathbb{R}}\left(\int_{(-\infty,x]}\psi d\nu\right)^q d\mu(x). \tag{9.30}$$

In the case of $S_0 = \infty$ or $S_0 < \infty,\ \mu([S_0,\infty)) = 0$ we choose $S < S_0$ and monotone convergence shows that the right hand side of (9.29) satisfies

$$\begin{aligned}&\lim_{S\uparrow S_0}\int_{\mathbb{R}}\left(\int_{(-\infty,x\wedge S]}\psi d\nu\right)^q d\mu(x) = \lim_{S\uparrow S_0}\int_{\mathbb{R}}\left(\int_{(-\infty,x]\cap(-\infty,S]}\psi d\nu\right)^q d\mu(x)\\&= \int_{\mathbb{R}}\left(\int_{(-\infty,x]\cap(-\infty,S_0)}\psi d\nu\right)^q d\mu(x) = \int_{\mathbb{R}}\left(\int_{(-\infty,x]}\psi d\nu\right)^q d\mu(x).\end{aligned} \tag{9.31}$$

Since inequality (3.1) holds with the left hand side replaced by the right hand side of (9.22) to the power $1/q$, the above argument involving (9.28) through (9.31) completes the proof of (3.1) and the theorem. □

For the proof of Theorem 3.5 we need the following Lemma.

Lemma 9.2. *For F and G distribution functions, χ a nonnegative measurable function and $0 < \gamma < 1$ we have*

$$\gamma \int_{(-\infty,x]} \chi(y) \left[\int_{(-\infty,y]} \chi dG \right]^{\gamma-1} dG(y) \leq \left[\int_{(-\infty,x]} \chi dG \right]^{\gamma} \tag{9.32}$$

and

$$\gamma \int_{[y,\infty)} \chi(x) \left[\int_{[x,\infty)} \chi dF \right]^{\gamma-1} dF(x) \leq \left[\int_{[y,\infty)} \chi dF \right]^{\gamma}. \tag{9.33}$$

Proof. By symmetry it suffices to prove (9.32), which with the distribution function $G_x(y) = \int_{(-\infty,y\wedge x]} \chi dG / \int_{(-\infty,x]} \chi dG$ is equivalent to

$$\gamma \int_{-\infty}^{\infty} G_x^{\gamma-1} dG_x \leq 1. \tag{9.34}$$

With the random variable U uniformly distributed on the unit interval the left hand side of this inequality equals and satisfies

$$\gamma E\left(\left[G_x\left(G_x^{-1}(U)\right)\right]^{\gamma-1} \right) \leq \gamma E\left(U^{\gamma-1}\right) = 1. \tag{9.35}$$

□

Proof of Theorem 3.5. The choice $\psi(y) = V^{-1/(p-1)}(y)\mathbf{1}_{[y\leq x]}$ in inequality (3.9) leads to the string of (in)equalities

$$\begin{aligned}
&\left[\int_{(-\infty,x]} V^{-1/(p-1)} dG \right]^p \int_{[x,\infty)} U dF \\
&= E\left(\left[E\left(V^{-1/(p-1)}(Y)\mathbf{1}_{[Y\leq x]} \right) \right]^p U(X)\mathbf{1}_{[X\geq x]} \right) \\
&\leq E\left(\left[E\left(V^{-1/(p-1)}(Y)\mathbf{1}_{[Y\leq x]}\mathbf{1}_{[Y\leq X]} \mid X \right) \right]^p U(X) \right) \\
&\leq C\, E\left(V^{-1/(p-1)}(Y)\mathbf{1}_{[Y\leq x]} \right) = C \int_{(-\infty,x]} V^{-1/(p-1)} dG, \quad x \in \mathbb{R},
\end{aligned} \tag{9.36}$$

which implies the first inequality in (3.11). With

$$h(y) = V^{1/p}(y) \left[\int_{(-\infty,y]} V^{-1/(p-1)} dG \right]^{(p-1)/p^2} \tag{9.37}$$

inequality (9.32) of Lemma 9.2 with $\chi = V^{-1/(p-1)}$ and $\gamma = 1 - 1/p = (p-1)/p$ yields

$$\begin{aligned}
&E\left(h^{-p/(p-1)}(Y)\mathbf{1}_{[Y\leq x]} \right) \\
&= \int_{(-\infty,x]} V^{-1/(p-1)}(y) \left[\int_{(-\infty,y]} V^{-1/(p-1)} dG \right]^{-1/p} dG(y) \\
&\leq \frac{p}{p-1} \left[\int_{(-\infty,x]} V^{-1/(p-1)} dG \right]^{(p-1)/p}.
\end{aligned} \tag{9.38}$$

By Hölder's inequality this implies

$$\begin{aligned}
&E\left(\left[E\left(\psi(Y)\mathbf{1}_{[Y\le X]}\mid X\right)\right]^p U(X)\right)\\
&= E\left(\left[E\left(\psi(Y)h(Y)(h(Y))^{-1}\mathbf{1}_{[Y\le X]}\mid X\right)\right]^p U(X)\right)\\
&\le E\Bigg(E\left(\psi^p(Y)h^p(Y)\mathbf{1}_{[Y\le X]}\mid X\right)\\
&\qquad \left[E\left(h^{-p/(p-1)}(Y)\mathbf{1}_{[Y\le X]}\mid X\right)\right]^{p-1} U(X)\Bigg) \qquad (9.39)\\
&\le \left(\frac{p}{p-1}\right)^{p-1} E\Bigg(E\left(\psi^p(Y)h^p(Y)\mathbf{1}_{[Y\le X]}\mid X\right)\\
&\qquad \left[\int_{(-\infty,X]} V^{-1/(p-1)}dG\right]^{(p-1)^2/p} U(X)\Bigg)\\
&= \left(\frac{p}{p-1}\right)^{p-1} E\Bigg(\psi^p(Y)h^p(Y)E\Bigg(\left[\int_{(-\infty,X]} V^{-1/(p-1)}dG\right]^{(p-1)^2/p}\\
&\qquad U(X)\mathbf{1}_{[Y\le X]}\mid Y\Bigg)\Bigg).
\end{aligned}$$

By the definition of B in (3.10) the right hand side of (9.39) is bounded from above by

$$\begin{aligned}
&\left(\frac{p}{p-1}\right)^{p-1} B^{(p-1)/p} E\Bigg(\psi^p(Y)V(Y)\left[\int_{(-\infty,Y]} V^{-1/(p-1)}dG\right]^{(p-1)/p}\\
&E\left(\left[\int_{[X,\infty)} UdF\right]^{(1/p)-1} U(X)\mathbf{1}_{[Y\le X]}\mid Y\right)\Bigg) \qquad (9.40)\\
&\le \frac{p^p}{(p-1)^{p-1}} B^{(p-1)/p} E\Bigg(\psi^p(Y)V(Y)\left[\int_{(-\infty,Y]} V^{-1/(p-1)}dG\right]^{(p-1)/p}\\
&\qquad \left[\int_{[Y,\infty)} UdF\right]^{1/p}\Bigg),
\end{aligned}$$

where the inequality follows from (9.33) of Lemma 9.2. By the definition of B the last expression is bounded by the right hand side of (3.11), which completes the proof of (3.11). □

9.3 Proofs for Section 4

Proof of Theorem 4.1. Let f be a density of F. The monotonicity of ψ implies

$$\begin{aligned}
\frac{d}{dx}\left[\int_{-\infty}^x \psi(y)dF(y)\right]^p &= p\left[\int_{-\infty}^x \psi(y)dF(y)\right]^{p-1}\psi(x)f(x)\\
&\ge p\psi^p(x)F^{p-1}(x)f(x) \qquad (9.41)
\end{aligned}$$

for Lebesgue almost all $x\in\mathbb{R}$. So we have

$$\left[\int_{-\infty}^x \psi(y)dF(y)\right]^p \ge p\int_{-\infty}^x \psi^p(y)(F(y))^{p-1}dF(y) \qquad (9.42)$$

and hence

$$
\begin{aligned}
&E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\leq X]}\mid X\right)}{F(X)}\right]^p\right)\\
&\geq p\int_{-\infty}^{\infty}\int_{-\infty}^{x}\psi^p(y)F^{p-1}(y)dF(y)F^{-p}(x)dF(x)\\
&= p\int_{-\infty}^{\infty}\int_{y}^{\infty}F^{-p}(x)f(x)dx\psi^p(y)(F(y))^{p-1}dF(y) \qquad (9.43)\\
&= \frac{p}{p-1}\int_{-\infty}^{\infty}\left[F^{1-p}(y)-1\right]\psi^p(y)(F(y))^{p-1}dF(y)\\
&= \frac{p}{p-1}E\left(\psi^p(Y)\left(1-F^{p-1}(Y)\right)\right),
\end{aligned}
$$

which is the first inequality of (4.3). Since ψ^p and $1-F^{p-1}$ are both nonincreasing, $\psi^p(Y)$ and $1-F^{p-1}(Y)$ are nonnegatively correlated and consequently their covariance is nonnegative implying

$$
\begin{aligned}
E\left(\psi^p(Y)\left(1-F^{p-1}(Y)\right)\right) &\geq E\left(\psi^p(Y)\right)E\left(1-F^{p-1}(Y)\right)\\
&= \frac{p-1}{p}E\left(\psi^p(Y)\right). \qquad (9.44)
\end{aligned}
$$

This results in the second inequality of (4.3).

Note that inequality (4.4) and hence the inequality between the left hand side and the right hand side of (4.3) is obvious as ψ is nonincreasing.

Let F be general and p integer. As $X_1,\ldots,X_p$ are independent and identically distributed and $\psi(\cdot)\mathbf{1}_{[\cdot\leq x]}$ is nonincreasing, we have

$$
E\left(\prod_{i=1}^{p}\psi(X_i)\mathbf{1}_{[X_i\leq x]}\right)\geq E\left(\psi^p(X_{(p)})\mathbf{1}_{[X_{(p)}\leq x]}\right) \qquad (9.45)
$$

and hence

$$
E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\leq X]}\mid X\right)}{F(X)}\right]^p\right)\geq E\left(\psi^p(X_{(p)})\mathbf{1}_{[X_{(p)}\leq X]}F^{-p}(X)\right), \qquad (9.46)
$$

which implies (4.5). □

Proof of Corollary 4.2. Let X and Y be uniformly distributed on the interval $(0,K)$. Our reverse Hardy inequality (4.3) becomes

$$
\frac{1}{K}\int_0^K\left[\frac{1}{x}\int_0^x\psi(y)dy\right]^p dx\geq\frac{p}{p-1}\frac{1}{K}\int_0^K\psi^p(y)\left(1-\left(\frac{y}{K}\right)^{p-1}\right)dy, \qquad (9.47)
$$

which for $0<\varepsilon\leq 1$ implies

$$
\int_0^K\left[\frac{1}{x}\int_0^x\psi(y)dy\right]^p dx\geq\frac{p}{p-1}\int_0^{\varepsilon K}\psi^p(y)\left(1-\varepsilon^{p-1}\right)dy, \qquad (9.48)
$$

Taking limits for $K\to\infty$ and subsequently $\varepsilon\downarrow 0$ we arrive at (4.1).

For the second part of the corollary we take X and Y uniformly distributed on $\{1,\ldots,K\}$. In view of $P(X_{(p)}\leq n)=(n/K)^p$ our inequality (4.5) with $\psi(k)=c_k$ becomes

$$
\frac{1}{K}\sum_{n=1}^{K}\left[\frac{\frac{1}{K}\sum_{k=1}^{n}c_k}{n/K}\right]^p\geq\sum_{n=1}^{K}c_n^p\left[\left(\frac{n}{K}\right)^p-\left(\frac{n-1}{K}\right)^p\right]\frac{1}{K}\sum_{k=n}^{K}\left(\frac{k}{K}\right)^{-p}, \qquad (9.49)
$$

which implies

$$\sum_{n=1}^{K}\left[\frac{1}{n}\sum_{k=1}^{n}c_k\right]^p \geq \sum_{n=1}^{K}c_n^p\left[n^p-(n-1)^p\right]\sum_{k=n}^{K_0}\frac{1}{k^p} \tag{9.50}$$

for any integer $K_0 \leq K$ and the corresponding sum vanishing for $n > K_0$. Taking limits as $K \to \infty$ and subsequently $K_0 \to \infty$ we obtain

$$\sum_{n=1}^{\infty}\left[\frac{1}{n}\sum_{k=1}^{n}c_k\right]^p \geq \sum_{n=1}^{\infty}c_n^p\left[n^p-(n-1)^p\right]\sum_{k=n}^{\infty}\frac{1}{k^p}. \tag{9.51}$$

Lemma 2 of [53] shows

$$\left[n^p-(n-1)^p\right]\sum_{k=n}^{\infty}\frac{1}{k^p} \geq \zeta(p) \tag{9.52}$$

for $n \geq 2$. As for $n = 1$ equality holds in (9.52), the proof that for integer p inequality (4.2) can be obtained from our inequality (4.5), is complete. □

9.4 Proofs for Section 5

We will use the following Lemma, which shows the structure of Copson's proof of his Theorem B with sums over infinitely many terms replaced by finite sums; see [18].

Lemma 9.3. *Let a_i and p_i be nonnegative numbers for $i = 1, \dots, m$, with $p_1 > 0$. For $p > 1$ the inequality*

$$\sum_{n=1}^{m}\left(\sum_{i=n}^{m}\frac{a_i p_i}{\sum_{j=1}^{i}p_j}\right)^p p_n \leq p^p\sum_{n=1}^{m}a_n^p p_n \tag{9.53}$$

holds.

Note that part of Theorem B of [18] follows from this inequality by taking limits for $m \to \infty$, first at the right hand side, subsequently within the p-th power at the left hand side, and finally for the first sum at the left hand side.

Proof of Lemma 9.3. With the notation

$$P_n = \sum_{i=1}^{n}p_i, \quad A_n = \sum_{i=n}^{m}\frac{a_i p_i}{P_i}, \quad n = 1, \dots, m, \quad P_0 = A_{m+1} = 0, \tag{9.54}$$

Young's inequality (as in the proof of Lemma 9.1) yields

$$\begin{aligned} &A_n^p p_n - pA_n^{p-1}a_n p_n = A_n^p p_n - pA_n^{p-1}P_n\left(A_n - A_{n+1}\right) \\ &\leq \left(p_n - pP_n\right)A_n^p + P_n\left((p-1)A_n^p + A_{n+1}^p\right) = P_n A_{n+1}^p - P_{n-1}A_n^p \end{aligned} \tag{9.55}$$

for $n = 1, \dots, m$. Summing this inequality over n we obtain

$$\sum_{n=1}^{m}A_n^p p_n - p\sum_{n=1}^{m}a_n A_n^{p-1}p_n \leq 0. \tag{9.56}$$

By Hölder's inequality the second sum in (9.56) is bounded as follows

$$\left(\sum_{n=1}^{m}a_n A_n^{p-1}p_n\right)^p \leq \sum_{n=1}^{m}a_n^p p_n\left(\sum_{n=1}^{m}A_n^p p_n\right)^{p-1}. \tag{9.57}$$

Together with (9.56) this implies

$$\left(\sum_{n=1}^{m}A_n^p p_n\right)^p \leq p^p\left(\sum_{n=1}^{m}a_n A_n^{p-1}p_n\right)^p \leq p^p\sum_{n=1}^{m}a_n^p p_n\left(\sum_{n=1}^{m}A_n^p p_n\right)^{p-1} \tag{9.58}$$

and hence (9.53). □

Proof of Theorem 5.1. As in the proof of Theorem 2.1 we define $y_{N,i} = F^{-1}(i/N),\ i = 0,\dots,N-1,\ y_{N,N} = \infty$, for large N and we apply Lemma 9.3 with $m = N$, but this time we choose

$$p_n = \int_{[y_{N,n-1},y_{N,n})} dF, \quad a_n = \int_{[y_{N,n-1},y_{N,n})} \psi dF/p_n, \quad n = 1,\dots,N. \tag{9.59}$$

By Jensen's inequality we have

$$a_n^p \le \int_{[y_{N,n-1},y_{N,n})} \psi^p dF/p_n, \quad n = 1,\dots,N, \tag{9.60}$$

and hence

$$\sum_{n=1}^N a_n^p p_n \le \sum_{n=1}^N \int_{[y_{N,n-1},y_{N,n})} \psi^p dF = E\left(\psi^p(Y)\right). \tag{9.61}$$

Observe that $F(y_{N,i}-) \le F(y) + 1/N$ holds for $y \in [y_{N,i-1}, y_{N,i})$. Consequently we have

$$\begin{aligned} \sum_{i=n}^N \frac{a_i p_i}{\sum_{j=1}^i p_j} &= \sum_{i=n}^N \int_{[y_{N,i-1},y_{N,i})} \frac{\psi(y)}{F(y_{N,i}-)} dF(y) \\ &\ge \sum_{i=n}^N \int_{[y_{N,i-1},y_{N,i})} \frac{\psi(y)}{F(y)+1/N} dF(y) = \int_{[y_{N,n-1},\infty)} \frac{\psi(y)}{F(y)+1/N} dF(y) \end{aligned} \tag{9.62}$$

and hence by Fatou's lemma

$$\begin{aligned} &\liminf_{N\to\infty} \sum_{n=1}^N \left(\sum_{i=n}^N \frac{a_i p_i}{\sum_{j=1}^i p_j} \right)^p p_n \\ &\ge \liminf_{N\to\infty} \sum_{n=1}^N \int_{[y_{N,n-1},y_{N,n})} \left(\int_{[y_{N,n-1},\infty)} \frac{\psi(y)}{F(y)+1/N} dF(y) \right)^p dF(x) \\ &\ge \liminf_{N\to\infty} \sum_{n=1}^N \int_{[y_{N,n-1},y_{N,n})} \left(\int_{[x,\infty)} \frac{\psi(y)}{F(y)+1/N} dF(y) \right)^p dF(x) \\ &= \liminf_{N\to\infty} \int_{\mathbb{R}} \left(\int_{[x,\infty)} \frac{\psi(y)}{F(y)+1/N} dF(y) \right)^p dF(x) \\ &\ge \int_{\mathbb{R}} \left(\int_{[x,\infty)} \liminf_{N\to\infty} \frac{\psi(y)}{F(y)+1/N} dF(y) \right)^p dF(x) \\ &= E\left(\left[E\left(\frac{\psi(Y)}{F(Y)} \mathbf{1}_{[Y\ge X]} \mid X \right) \right]^p \right). \end{aligned} \tag{9.63}$$

Combining (9.63), Lemma 9.3 and (9.61) we arrive at a proof of Theorem 5.1. □

Proof of Theorem 5.3. Let η be a nonnegative measurable function. Hölder's inequality and subsequently Hardy's inequality (2.1) yield

$$\begin{aligned} &E\left(E\left(\frac{\psi(Y)}{F(Y)} \mathbf{1}_{[Y\ge X]} \mid X \right) \eta(X) \right) = E\left(\psi(Y) \frac{E\left(\eta(X)\mathbf{1}_{[X\le Y]} \mid Y\right)}{F(Y)} \right) \\ &\le \left[E\left(\psi^p(Y)\right)\right]^{1/p} \left[E\left(\left[\frac{E\left(\eta(X)\mathbf{1}_{[X\le Y]} \mid Y\right)}{F(Y)} \right]^{p/(p-1)} \right) \right]^{(p-1)/p} \\ &\le \left[E\left(\psi^p(Y)\right)\right]^{1/p} \left[\left(\frac{p/(p-1)}{p/(p-1)-1} \right)^{p/(p-1)} E\left(\eta^{p/(p-1)}(X) \right) \right]^{(p-1)/p} \\ &= p\left[E\left(\psi^p(Y)\right)\right]^{1/p} \left[E\left(\eta^{p/(p-1)}(X) \right) \right]^{(p-1)/p}. \end{aligned} \tag{9.64}$$

Taking

$$\eta(X) = \left[E\left(\frac{\psi(Y)}{F(Y)} \mathbf{1}_{[Y \geq X]} \mid X \right) \right]^{p-1} \tag{9.65}$$

we obtain Copson's inequality (5.3) from (9.64). Similarly, Hölder's inequality and subsequently Copson's inequality (5.3) yield

$$\begin{aligned} & E\left(\frac{E\left(\psi(Y)\mathbf{1}_{[Y \leq X]} \mid X\right)}{F(X)} \eta(X) \right) = E\left(\psi(Y) E\left(\frac{\eta(X)}{F(X)} \mathbf{1}_{[X \geq Y]} \mid Y \right) \right) \\ & \leq \left[E\left(\psi^p(Y)\right)\right]^{1/p} \left[E\left(\left[E\left(\frac{\eta(X)}{F(X)} \mathbf{1}_{[X \geq Y]} \mid Y \right) \right]^{p/(p-1)} \right) \right]^{(p-1)/p} \\ & \leq \left[E\left(\psi^p(Y)\right)\right]^{1/p} \left[\left(\frac{p}{p-1} \right)^{p/(p-1)} E\left(\eta^{p/(p-1)}(X) \right) \right]^{(p-1)/p} \\ & = \frac{p}{p-1} \left[E\left(\psi^p(Y)\right)\right]^{1/p} \left[E\left(\eta^{p/(p-1)}(X) \right) \right]^{(p-1)/p}. \end{aligned} \tag{9.66}$$

Taking

$$\eta(X) = \left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y \leq X]} \mid X\right)}{F(X)} \right]^{p-1} \tag{9.67}$$

we obtain Hardy's inequality (2.1) from (9.66). □

9.5 Proof for Section 6

Proof of Theorem 6.1. First we prove that for $p \in [1, \infty)$, for arbitrary F and for $x \mapsto \psi(x)/F(x)$ nonincreasing

$$E\left(\left[E\left(\frac{\psi(Y)}{F(Y)} \mathbf{1}_{[Y \geq X]} \mid X \right) \right]^p \right) \geq E\left(\psi^p(Y) \left[\frac{F(Y-)}{F(Y)} \right]^p \right) \tag{9.68}$$

holds. Observe that for continuous F this implies (6.3). To prove (9.68) we follow the line of argument in the proof of Theorem 4 of Renaud [53]. For $x < y$ the monotonicity of ψ/F implies

$$\int_{[x,y)} \frac{\psi}{F} dF \geq \frac{\psi(y)}{F(y)} [F(y-) - F(x-)] \tag{9.69}$$

and hence

$$p \left[\int_{[x,y)} \frac{\psi}{F} \, dF \right]^{p-1} \frac{\psi(y)}{F(y)} \geq p \left[\frac{\psi(y)}{F(y)} \right]^p [F(y-) - F(x-)]^{p-1} \tag{9.70}$$

and

$$\begin{aligned} & \int_{\mathbb{R}} \int_{[x,\infty)} p \left[\int_{[x,y)} \frac{\psi}{F} dF \right]^{p-1} \frac{\psi(y)}{F(y)} \, dF(y) \, dF(x) \\ & \geq \int_{\mathbb{R}} \int_{[x,\infty)} p \left[\frac{\psi(y)}{F(y)} \right]^p [F(y-) - F(x-)]^{p-1} \, dF(y) \, dF(x). \end{aligned} \tag{9.71}$$

In view of $F(F^{-1}(u)-) \le u$ and since $u \le F(y-)$ implies $F^{-1}(u) \le y$, Fubini's theorem shows that the right hand side of (9.71) equals and satisfies

$$\begin{aligned}
&\int_{\mathbb{R}} \left[\frac{\psi(y)}{F(y)}\right]^p \int_{(-\infty,y]} p[F(y-) - F(x-)]^{p-1}\, dF(x)\, dF(y) \\
&= \int_{\mathbb{R}} \left[\frac{\psi(y)}{F(y)}\right]^p \int_0^1 p[F(y-) - F(F^{-1}(u)-)]^{p-1} \mathbf{1}_{[F^{-1}(u)\le y]}\, du\, dF(y) \qquad (9.72)\\
&\ge \int_{\mathbb{R}} \left[\frac{\psi(y)}{F(y)}\right]^p \int_0^1 p[F(y-) - u]^{p-1} \mathbf{1}_{[u\le F(y-)]}\, du\, dF(y) \\
&= \int_{\mathbb{R}} \left[\frac{\psi(y)}{F(y)}\right]^p [F(y-)]^p\, dF(y) = E\left(\psi^p(Y)\left[\frac{F(Y-)}{F(Y)}\right]^p\right).
\end{aligned}$$

Furthermore, for fixed x we define the distribution function $G_x(y) = \int_{[x,y)} (\psi/F)\, dF \,/\, \int_{[x,\infty)} (\psi/F)\, dF$ and we obtain

$$\int_{-\infty}^{\infty} p[G_x(y-)]^{p-1} dG_x(y) = \int_0^1 p[G_x(G_x^{-1}(u)-)]^{p-1} du \le \int_0^1 p u^{p-1} du = 1. \qquad (9.73)$$

This shows that the left hand side of (9.71) is bounded from above by

$$\int_{\mathbb{R}} \left[\int_{[x,\infty)} \psi/F\, dF\right]^p dF(x) = E\left(\left[E\left(\frac{\psi(Y)}{F(Y)}\mathbf{1}_{[Y\ge X]} \mid X\right)\right]^p\right). \qquad (9.74)$$

Combining this with (9.71) and (9.72) we arrive at (9.68) and hence at (6.3).

To prove (6.4) and (6.5) we restrict attention to integer p and let $X, Y, Y_1, \dots, Y_p$ be independent random variables all with distribution function F.

If F is continuous, the monotonicity of ψ implies that

$$\begin{aligned}
&E\left(\left[E\left(\frac{\psi(Y)}{F(Y)}\mathbf{1}_{[Y\ge X]} \mid X\right)\right]^p\right) = E\left(\prod_{i=1}^p \left[E\left(\frac{\psi(Y_i)}{F(Y_i)}\mathbf{1}_{[X\le Y_i]} \mid X\right)\right]\right) \\
&= E\left(E\left(\prod_{i=1}^p \frac{\psi(Y_i)}{F(Y_i)}\mathbf{1}_{[X\le Y_i]} \mid X\right)\right) = E\left(\prod_{i=1}^p \frac{\psi(Y_i)}{F(Y_i)}\mathbf{1}_{[X\le Y_i]}\right) \\
&= p!\, E\left(\left[\prod_{i=1}^p \frac{\psi(Y_i)}{F(Y_i)}\right]\mathbf{1}_{[X\le Y_1\le\cdots\le Y_p]}\right) \ge p!\, E\left(\psi^p(Y_p)\frac{\mathbf{1}_{[X\le Y_1\le\cdots\le Y_p]}}{F(Y_1)\cdots F(Y_p)}\right) \\
&= p!\, E\left(\psi^p(Y_p)\frac{\mathbf{1}_{[Y_1\le Y_2\le\cdots\le Y_p]}}{F(Y_2)\cdots F(Y_p)}\right) = p!\, E\left(\psi^p(Y)\right), \qquad (9.75)
\end{aligned}$$

where equality holds if ψ is constant.

Similarly, if F is arbitrary, we derive

$$\begin{aligned}
&E\left(\left[E\left(\frac{\psi(Y)}{F(Y)}\mathbf{1}_{[Y\ge X]} \mid X\right)\right]^p\right) = E\left(\prod_{i=1}^p \frac{\psi(Y_i)}{F(Y_i)}\mathbf{1}_{[X\le Y_i]}\right) \\
&\ge E\left(\left[\prod_{i=1}^p \frac{\psi(Y_i)}{F(Y_i)}\right]\mathbf{1}_{[X\le Y_1\le\cdots\le Y_p]}\right) \ge E\left(\psi^p(Y_p)\frac{\mathbf{1}_{[X\le Y_1\le\cdots\le Y_p]}}{F(Y_1)\cdots F(Y_p)}\right) \\
&= E\left(\psi^p(Y_p)\frac{\mathbf{1}_{[Y_1\le Y_2\le\cdots\le Y_p]}}{F(Y_2)\cdots F(Y_p)}\right) = E\left(\psi^p(Y)\right). \qquad (9.76)
\end{aligned}$$

One may check that equalities in (9.76) hold if F is degenerate. □

9.6 Proofs for Section 7

Proof of Theorem 7.1. By Hardy's inequality in the probability form (2.1) with ψ replaced by $\psi^{1/p}$ we have

$$E\left(\left[\frac{E(\psi^{1/p}(Y)1_{[Y\le X]}|X)}{F(X)}\right]^p\right) \le \left(\frac{p}{p-1}\right)^p E(\psi(Y)).$$

The left hand side equals

$$\begin{aligned} &E\left([E\left(\exp\{1/p\log\psi(Y)\}|Y\le X, X\right)]^p\right) \\ &\ge E\left([\exp\{1/pE\left(\log\psi(Y)|Y\le X, X\right)\}]^p\right) \\ &= E\left(\exp\left\{\frac{E\left(1_{[Y\le X]}\log\psi(Y)|X\right)}{F(X)}\right\}\right), \end{aligned} \tag{9.77}$$

where the inequality holds in view of Jensen's inequality for conditional expectations and the convexity of $\exp$. The right hand side of (2.9) completes the proof. $\square$

10 Summary

Our sharp inequalities related to Hardy's inequality read as follows.

$$E\left(\psi^p(Y)\right) \le E\left(\left[\frac{E\left(\psi(Y)\mathbf{1}_{[Y\le X]}\mid X\right)}{F(X)}\right]^p\right) \le \left(\frac{p}{p-1}\right)^p E\left(\psi^p(Y)\right), \tag{10.1}$$

where the first inequality holds if F is absolutely continuous and ψ is nonincreasing.

Our sharp inequalities related to Copson's inequality are the following.

$$E\left(\psi^p(Y)\right) \le E\left(\left[E\left(\frac{\psi(Y)}{F(Y)}\mathbf{1}_{[Y\ge X]}\mid X\right)\right]^p\right) \le p^p E\left(\psi^p(Y)\right), \tag{10.2}$$

where the first inequality holds if F is continuous and $x \mapsto \psi(x)/F(x)$ is nonincreasing. Our Hardy inequality with weights and mixed norms is

$$\begin{aligned} &\left\{E\left(\left[E\left(\psi(Y)\mathbf{1}_{[Y\le X]}\mid X\right)\right]^q U(X)\right)\right\}^{1/q} \\ &\le \left(\frac{(q-p)/p}{\mathrm{Beta}(p/(q-p),(q-1)p/(q-p))}\right)^{(q-p)/pq} \\ &\quad \sup_{x\in\mathbb{R}}\left[\int_{[x,\infty)} U dF\right]^{1/q}\left[\int_{(-\infty,x]} V^{-1/(p-1)}dG\right]^{(p-1)/p} \left\{E\left(\psi^p(Y)V(Y)\right)\right\}^{1/p}. \end{aligned} \tag{10.3}$$

Detailed conditions are given in the respective Theorems.

11 Applications and Related Work

We close with a few brief comments concerning applications and related work.

As noted by Diaconis [21], Hardy's inequality (1.2), and especially the weighted version thereof due to Muckenhoupt [48], has been applied by Miclo [46] to obtain useful bounds for the spectral gap for birth-and-death Markov chains. He provides a nice overview of alternative methods and their potential drawbacks. Bobkov and Götze [10] extend the methods of [48] to study optimal constants in log-Sobolev inequalities on $\mathbb{R}$. Because log-Sobolev inequalities are preserved by the formation of products of independent distributions (i.e. tensorization), their results yield log-Sobolev inequalities for product measures. Their results have been refined by Barthe and Roberto [4] who go on in [5] to study modified log-Sobolev inequalities. Saumard and Wellner [56] use the "two-sided" Hardy inequality given by (2.14) to give an alternative proof of Cheeger's inequality. Applications of the Hardy inequality (2.1) with F continuous to semiparametric models for survival analysis were given by Ritov and Wellner [54] and Bickel et al. [7]. As noted in Sections 2, 5, 7, and 8, these results yield martingale connections with the operators $\overline{H}_F$ and $\overline{H}_F^*$.

There has been some related work on Hardy type inequalities with similar unification (of continuous and discrete cases) as an explicit goal: for example, see Kaijser et al. [33] and Evans et al. [24], page 45. Li and Mao [42], pages 257-258, refer to Prokhorov [52]. They all study general measures.

What about related work on formulating probabilistic versions of Hardy type inequalities? We have not found any results in this direction. Despite the many applications of Hardy and Muckenhoupt type inequalities in probability theory over the past 30 years, we are unaware of any explicit mention of these inequalities in terms of random variables. It seems to us that these inequalities should be better known in both the probability and statistics communities, and the probability versions may stimulate both further applications and further theoretical developments. In any case, it seems to be worthwhile to understand when several different formulations can be unified.

In Section 8 we sketched the connection between the operators H_F^* and $\overline{H}_F^*$ appearing in our probabilistic version of Copson's dual inequality and a simple counting process martingale. The key functions $\overline{\Lambda}_F(x)$ and $\Lambda_F(x)$ appearing in those operators (recall (5.10) for the explicit definitions) play an extremely important role in survival analysis and reliability theory. Also note that they do not appear without the probabilistic perspective adopted in our approach. In the Appendix (Section 12) we discuss how these functions arise in connection with left and right censored survival data.

12 Appendix

Here we go further with the discussion concerning the forward and backward hazard functions connected with our random variable versions of the Copson inequalities.

12.1 Censored survival data: from the right and from the left

Suppose that $X_1, \ldots, X_n$ are i.i.d. survival times with d.f. F on $[0, \infty)$. Furthermore, suppose that $Y_1, \ldots, Y_n$ are i.i.d. censoring times (independent of $X_1, \ldots, X_n$) with distribution function G. Unfortunately we do not get to observe the X_i's. Instead, for each individual we observe

$$(Z_i, \delta_i) \equiv (X_i \wedge Y_i, \delta_i) \equiv (X_i \wedge Y_i, \mathbf{1}_{[X_i \le Y_i]}).$$

Nevertheless, our goal is to estimate the cumulative hazard function

$$\overline{\Lambda}_F(t) = \int_{[0,t]} (1 - F(s-))^{-1} dF(s)$$

and the survival function $1 - F$ nonparametrically. Actually, once we have an estimator $\widehat{\overline{\Lambda}}_{F,n}$ of $\overline{\Lambda}_F$, then estimation of $1 - F$ (and hence also F) is immediate since

$$1 - F(t) = \exp(-\overline{\Lambda}_c(t)) \prod_{s \le t} (1 - \Delta\overline{\Lambda}(s)),$$

where $\Delta\overline{\Lambda}(s) \equiv \overline{\Lambda}(s) - \overline{\Lambda}(s-)$ and $\overline{\Lambda}_c(t) \equiv \Lambda(t) - \sum_{s \le t} \Delta\Lambda(s)$. This is the setting of (random, right) – censored survival data, and the (nonparametric) maximum likelihood estimators of $\overline{\Lambda}$ and $1-F$ are the famous Nelson-Aalen estimators $\widehat{\overline{\Lambda}}$ of Λ and Kaplan-Meier estimator $1-\widehat{F}_n$ of $1-F$. This is the random censorship version of right-censored survival data. For treatments of fixed (i.e. deterministic) censoring times, see Pollard [51] and Meier [45].

Before discussing right-censoring further, suppose instead that we observe

$$(W_i,\gamma_i) \equiv (U_i \vee V_i, \mathbf{1}_{[U_i \ge V_i]})$$

where the U_i's are i.i.d. with d.f. F, and the V_i's are i.i.d. G (and independent of the U_i's). The goal again is to estimate the (reverse or backwards) cumulative hazard function $\Lambda_F(t) \equiv \int_{[t,\infty)} dF(s)/F(s)$ and the d.f. F. This is left-censored survival data. Note that Λ_F is the function which arose naturally in the random variable version of Copson's inequality in Section 8. A famous example of left-censored data is the data which arose in a study of the descent times of baboons in the Amboseli Reserve, Kenya. See [63], [64], [19], [20].

In this study the U_i's represent the times when the baboons descended from the trees in the morning while the V_i's represent the times at which the investigators arrived at the study site. If a baboon descended before its observer arrived at the study site, then that baboon's U_i is regarded as being "left – censored". Again the goal is nonparametric estimation of the d.f. of the U_i's.

In this setting, once we have an estimator $\widehat{\Lambda}_{F,n}$ of Λ_F, then estimation of F is immediate since

$$F(t) = \exp\left(-\Lambda_c(t)\right) \prod_{s \ge t} \Delta\Lambda(s)$$

where

$$\Delta\Lambda(s) \equiv \Lambda(s) - \Lambda(s-), \quad \Lambda_c(t) \equiv \Lambda(t) - \sum_{s \ge t} \Delta\Lambda(s).$$

12.2 Nonparametric estimation for right or left censored survival data

First the classical and frequently occurring censoring from the right. To see that $\overline{\Lambda}_F$ and $1-F$ can be estimated nonparametrically from the observed data, consider the following empirical distributions:

$$\begin{aligned}
\mathbb{H}_n^{uc}(t) &= \mathbb{P}_n(\delta \mathbf{1}_{[Z \le t]}) = n^{-1}\sum_{i=1}^n \delta_i \mathbf{1}_{[Z_i \le t]},\\
\mathbb{H}_n^{c}(t) &= \mathbb{P}_n((1-\delta) \mathbf{1}_{[Z \le t]}) = n^{-1}\sum_{i=1}^n (1-\delta_i) \mathbf{1}_{[Z_i \le t]},\\
\mathbb{H}_n(t) &= \mathbb{P}_n \mathbf{1}_{[Z \le t]} = n^{-1}\sum_{i=1}^n \mathbf{1}_{[Z_i \le t]}
\end{aligned}$$

where "uc" stands for "uncensored" observations and "c" stands for "censored" observations. By the strong law of large numbers,

$$\begin{aligned}
\mathbb{H}_n^{uc}(t) &\to_{a.s.} E(\delta \mathbf{1}_{[Z \le t]}) = \int_{[0,t]} (1-G(s-))dF(s) = H^{uc}(t),\\
\mathbb{H}_n^{c}(t) &\to_{a.s.} E((1-\delta) \mathbf{1}_{[Z \le t]}) = \int_{[0,t]} (1-F(s))dG(s) = H^{c}(t),\\
\mathbb{H}_n(t) &\to_{a.s.} P(Z \le t) = 1-(1-F(t))(1-G(t)) = H(t).
\end{aligned}$$

Now note that

$$\overline{\Lambda}_F(t) = \int_{[0,t]} \frac{1}{1-F_-} dF = \int_{[0,t]} \frac{1-G_-}{(1-G_-)(1-F_-)} dF$$
$$= \int_{[0,t]} \frac{1-G_-}{(1-H_-)} dF = \int_{[0,t]} \frac{1}{1-H(s-)} dH^{uc}(s),$$

so we can estimate $\overline{\Lambda}_F$ by

$$\widehat{\overline{\Lambda}}_n(t) \equiv \int_{[0,t]} \frac{1}{1-\mathbb{H}_n(s-)} d\mathbb{H}_n^{uc}(s).$$

Then $1-\widehat{F}_n(t) = \prod_{s\le t}(1-\Delta\widehat{\overline{\Lambda}}_n(s))$ is the Kaplan-Meier [34] estimator of $1-F$.

Now for estimation in the presence of censoring from the left. To see that $\overline{\Lambda}_F$ and Λ_F can be estimated nonparametrically from the observed (left-censored) data, consider the following empirical distributions:

$$\begin{aligned}
\mathbb{H}_n^{uc}(t) &= \mathbb{P}_n(\gamma \mathbf{1}_{[W\le t]}) = n^{-1}\sum_{i=1}^n \gamma_i \mathbf{1}_{[W_i\le t]},\\
\mathbb{H}_n^{c}(t) &= \mathbb{P}_n((1-\gamma) \mathbf{1}_{[W\le t]}) = n^{-1}\sum_{i=1}^n (1-\gamma_i) \mathbf{1}_{[W_i\le t]},\\
\mathbb{K}_n(t) &= \mathbb{P}_n \mathbf{1}_{[W\le t]}.
\end{aligned}$$

Now

$$\begin{aligned}
\mathbb{K}_n^{uc}(t) &\to_{a.s.} E(\gamma \mathbf{1}_{[W\le t]}) = \int_{[0,t]} G(s) dF(s) \equiv K^{uc}(t),\\
\mathbb{K}_n^{c}(t) &\to_{a.s.} E((1-\gamma) \mathbf{1}_{[W\le t]}) = \int_{[0,t]} F(s-) dG(s) \equiv K^{c}(t),\\
\mathbb{K}_n(t) &\to_{a.s.} P(W\le t) = F(t)G(t) = K(t).
\end{aligned}$$

Now note that

$$\Lambda_F(t) = \int_{[s\ge t]} \frac{1}{F(s)} dF(s) = \int_{[s\ge t]} \frac{G(s)}{G(s)F(s)} dF(s) = \int_{[s\ge t]} \frac{G(s)}{K(s)} dF(s) = \int_{[s\ge t]} \frac{1}{K(s)} dK^{uc}(s),$$

so we can estimate the "backwards" Nelson-Aalen hazard function Λ_F by

$$\widehat{\Lambda}_n(t) \equiv \int_{[s\ge t]} \frac{1}{\mathbb{K}_n(s)} d\mathbb{K}_n^{uc}(s).$$

Then $\widehat{F}_n(t) = \prod_{s\ge t}\Delta\widehat{\Lambda}_n(s)$ is the "reverse" or "backwards" Kaplan – Meier estimator of F; see e.g. [64] and [19], [20].

For more on left-censoring, the data in the baboon study, and a plot of the resulting backwards Kaplan-Meier estimator, see Andersen et al. [1], pages 24, 162-165, and 273-274.

References

[1] Per Kragh Andersen, Ørnulf Borgan, Richard D. Gill, and Niels Keiding, *Statistical models based on counting processes*, Springer Series in Statistics, Springer-Verlag, New York, 1993. MR1198884

[2] Dominique Bakry, Ivan Gentil, and Michel Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, Springer, Cham, 2014. MR3155209

[3] Alexander A. Balinsky, W. Desmond Evans, and Roger T. Lewis, *The Analysis and Geometry of Hardy's inequality*, Universitext, Springer, Cham, 2015. MR3408787

[4] F. Barthe and C. Roberto, *Sobolev inequalities for probability measures on the real line*, Studia Math. **159** (2003), 481–497, Dedicated to Professor Aleksander Pełczyński on the occasion of his 70th birthday (Polish). MR2052235

[5] F. Barthe and C. Roberto, *Modified logarithmic Sobolev inequalities on* $\mathbb{R}$, Potential Anal. **29** (2008), 167–193. MR2430612

[6] Grahame Bennett, *Lower bounds for matrices*, Linear Algebra Appl. **82** (1986), 81–98. MR0858964

[7] Peter J. Bickel, Chris A. J. Klaassen, Ya'acov Ritov, and John A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Springer-Verlag, New York, 1998, Reprint of the 1993 original. MR1623559

[8] G. A. Bliss, *An Integral Inequality*, J. London Math. Soc. **5** (1930), no. 1, 40–46. MR1574997

[9] S. G. Bobkov and F. Götze, *Discrete isoperimetric and Poincaré-type inequalities*, Probab. Theory Related Fields **114** (1999), no. 2, 245–277. MR1701522

[10] S. G. Bobkov and F. Götze, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. **163** (1999), no. 1, 1–28. MR1682772

[11] Sergey Bobkov and Friedrich Götze, *Hardy Type Inequalities via Riccati and Sturm-Liouville equations*, Sobolev Spaces in Mathematics. I, Int. Math. Ser. (N. Y.), vol. 8, Springer, New York, 2009, pp. 69–86. MR2508839

[12] Santiago Boza and Javier Soria, *Solution to a conjecture on the norm of the Hardy operator minus the identity*, J. Funct. Anal. **260** (2011), no. 4, 1020–1028. MR2747011

[13] Santiago Boza and Javier Soria, *Averaging operators on decreasing or positive functions: equivalence and optimal bounds*, J. Approx. Theory **237** (2019), 135–152. MR3868629

[14] J. Scott Bradley, *Hardy inequalities with mixed norms*, Canad. Math. Bull. **21** (1978), no. 4, 405–408. MR0523580

[15] T. A. A. Broadbent, *A proof of Hardy's convergence theorem*, J. London Math. Soc. **3** (1928), no. 4, 242–243. MR1574000

[16] Arlen Brown, P. R. Halmos, and A. L. Shields, *Cesàro operators*, Acta Sci. Math. (Szeged) **26** (1965), 125–137. MR0187085

[17] Mu-Fa Chen, *Basic estimates of stability rate for one-dimensional diffusions*, Probability Approximations and Beyond, Lect. Notes Stat., vol. 205, Springer, New York, 2012, pp. 75–99. MR3289377

[18] E. T. Copson, *Note on series of positive terms*, J. London Math. Soc. **2** (1927), no. 1, 9–12. MR1574056

[19] Sándor Csörgő and Lajos Horváth, *Random censorship from the left*, Studia Sci. Math. Hungar. **15** (1980), no. 4, 397–401. MR0688619

[20] Sándor Csörgő and Lajos Horváth, *The baboons come down from the trees quite normally*, Mathematical statistics and applications, Vol. B (Bad Tatzmannsdorf, 1983), Reidel, Dordrecht, 1985, pp. 95–106. MR0851047

[21] Persi Diaconis, *G. H. Hardy and probability???*, Bull. London Math. Soc. **34** (2002), no. 4, 385–402. MR1897417

[22] Bradley Efron and Iain M. Johnstone, *Fisher's information in terms of the hazard rate*, Ann. Statist. **18** (1990), no. 1, 38–62. MR1041385

[23] E. B. Elliott, *A simple exposition of some recently proved facts as to convergency*, J. London Math. Soc. **1** (1926), no. 2, 93–96. MR1574962

[24] W. Desmond Evans, Amiran Gogatishvili, and Bohumír Opic, *The reverse Hardy inequality with measures*, Math. Inequal. Appl. **11** (2008), no. 1, 43–74. MR2376257

[25] Gerald B. Folland, *Real analysis*, second ed., Pure and Applied Mathematics (New York), John Wiley & Sons, Inc., New York, 1999, Modern techniques and their applications, A Wiley-Interscience Publication. MR1681462

[26] N. Gozlan and C. Léonard, *Transport inequalities. A survey*, Markov Process. Related Fields **16** (2010), no. 4, 635–736. MR2895086

[27] G. H. Hardy, *Notes on some points in the integral calculus, (LX). an inequality between integrals*, Messenger of Math. **54** (1925), 150–156.

[28] G. H. Hardy, *Remarks on three recent notes in the Journal*, J. London Math. Soc. **3** (1928), no. 3, 166–169. MR1574154

[29] G. H. Hardy and J. E. Littlewood, *Notes on the Theory of Series (XII): On Certain Inequalities Connected with the Calculus of Variations*, J. London Math. Soc. **5** (1930), no. 1, 34–39. MR1574995

[30] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge, at the University Press, 1952, 1967 reprint of the 1952 2d ed. MR0046395

[31] Maria Johansson, Lars-Erik Persson, and Anna Wedestig, *Carleman's inequality-history, proofs and some new generalizations*, JIPAM. J. Inequal. Pure Appl. Math. **4** (2003), no. 3, Article 53, 19. MR2044402

[32] Sten Kaijser, Ludmila Nikolova, Lars-Erik Persson, and Anna Wedestig, *Hardy-type inequalities via convexity*, Math. Inequal. Appl. **8** (2005), no. 3, 403–417. MR2148234

[33] Sten Kaijser, Lars-Erik Persson, and Anders Öberg, *On Carleman and Knopp's inequalities*, J. Approx. Theory **117** (2002), no. 1, 140–151. MR1920123

[34] E. L. Kaplan and Paul Meier, *Nonparametric estimation from incomplete observations*, J. Amer. Statist. Assoc. **53** (1958), 457–481. MR0093867

[35] V. I. Kolyada, *Optimal relationships between L^p-norms for the Hardy operator and its dual*, Ann. Mat. Pura Appl. (4) **193** (2014), no. 2, 423–430. MR3180926

[36] V. I. Kolyada, *On the optimal relationships between L^P-norms for the Hardy operator and its dual for decreasing functions*, J. Approx. Theory **252** (2020), 105362, 5. MR4051866

[37] Natan Krugljak, Lech Maligranda, and Lars Erik Persson, *On an elementary approach to the fractional Hardy inequality*, Proc. Amer. Math. Soc. **128** (2000), no. 3, 727–734. MR1676324

[38] Natan Kruglyak and Eric Setterqvist, *Sharp estimates for the identity minus Hardy operator on the cone of decreasing functions*, Proc. Amer. Math. Soc. **136** (2008), no. 7, 2505–2513. MR2390520

[39] Alois Kufner, Lech Maligranda, and Lars-Erik Persson, *The prehistory of the Hardy inequality*, Amer. Math. Monthly **113** (2006), no. 8, 715–732. MR2256532

[40] Alois Kufner, Lech Maligranda, and Lars-Erik Persson, *The Hardy inequality*, Vydavatelský Servis, Plzeň, 2007, About its history and some related results. MR2351524

[41] Alois Kufner, Lars-Erik Persson, and Natasha Samko, *Weighted Inequalities of Hardy Type*, second ed., World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2017. MR3676556

[42] Ying Li and Yong-Hua Mao, *The optimal constant in generalized Hardy's inequality*, Math. Inequal. Appl. **23** (2020), no. 1, 257–266. MR4061539

[43] Zhong-Wei Liao, *Discrete Hardy-type inequalities*, Adv. Nonlinear Stud. **15** (2015), no. 4, 805–834. MR3405817

[44] R. S. Liptser and A. N. Shiryayev, *Statistics of random processes. II*, Springer-Verlag, New York-Heidelberg, 1978, Applications, Translated from the Russian by A. B. Aries, Applications of Mathematics, Vol. 6. MR0488267

[45] Paul Meier, *Estimation of a distribution function from incomplete observations*, Perspectives in probability and statistics (papers in honour of M. S. Bartlett on the occasion of his 65th birthday), Academic Press, 1975, pp. 67–87. MR0411040

[46] L. Miclo, *An example of application of discrete Hardy's inequalities*, Markov Process. Related Fields **5** (1999), no. 3, 319–330. MR1710983

[47] M. Milman, *A note on reversed Hardy inequalities and Gehring's lemma*, Comm. Pure Appl. Math. **50** (1997), no. 4, 311–315. MR1438149

[48] Benjamin Muckenhoupt, *Hardy's inequality with weights*, Studia Math. **44** (1972), 31–38, Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, I. MR0311856

[49] Alfred Müller and Dietrich Stoyan, *Comparison Methods for Stochastic Models and Risks*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 2002. MR1889865

[50] Josip Pečarić and Kenneth B. Stolarsky, *Carleman's inequality: history and new generalizations*, Aequationes Math. **61** (2001), no. 1-2, 49–62. MR1820809

[51] David Pollard, *Empirical processes: theory and applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 2, Institute of Mathematical Statistics, Hayward, CA; American Statistical Association, Alexandria, VA, 1990. MR1089429

[52] D. V. Prokhorov, *The Hardy inequality with three measures*, Tr. Mat. Inst. Steklova **255** (2006), no. Funkts. Prostran., Teor. Priblizh., Nelineĭn. Anal., 233–245. MR2302834

[53] P. F. Renaud, *A reversed Hardy inequality*, Bull. Austral. Math. Soc. **34** (1986), no. 2, 225–232. MR0854569

[54] Ya'acov Ritov and Jon A. Wellner, *Censoring, martingales, and the Cox model*, Statistical Inference from Stochastic Processes (Ithaca, NY, 1987), Contemp. Math., vol. 80, Amer. Math. Soc., Providence, RI, 1988, pp. 191–219. MR0999013

[55] Walter Rudin, *Real and Complex Analysis*, McGraw-Hill Book Co., New York-Toronto, Ont.-London, 1966. MR0210528

[56] Adrien Saumard and Jon A. Wellner, *On the isoperimetric constant, covariance inequalities and L_p-Poincaré inequalities in dimension one*, Bernoulli **25** (2019), no. 3, 1794–1815. MR3961231

[57] Moshe Shaked and J. George Shanthikumar, *Stochastic Orders*, Springer Series in Statistics, Springer, New York, 2007. MR2265633

[58] Galen R. Shorack and Jon A. Wellner, *Empirical Processes with Applications to Statistics*, Classics in Applied Mathematics, vol. 59, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009, Reprint of the 1986 original [MR0838963]. MR3396731

[59] J. Michael Steele, *The Cauchy-Schwarz Master Class*, MAA Problem Books Series, Mathematical Association of America, Washington, DC; Cambridge University Press, Cambridge, 2004, An introduction to the art of mathematical inequalities. MR2062704

[60] Michał Strzelecki, *Hardy's operator minus identity and power weights*, J. Funct. Anal. **279** (2020), no. 2, 108532, 34. MR4088501

[61] Giorgio Talenti, *Osservazioni sopra una classe di disuguaglianze*, Rend. Sem. Mat. Fis. Milano **39** (1969), 171–185. MR0280661

[62] Giuseppe Tomaselli, *A class of inequalities*, Boll. Un. Mat. Ital. (4) **2** (1969), 622–631. MR0255751

[63] S. S. Wagner and S. A. Altmann, *What time do the baboons come down from the trees? (an estimation problem)*, Biometrics **29** (1973), 623 – 635.

[64] J.H. Ware and D.L. DeMets, *Reanalysis of some baboon descent data*, Biometrics **32** (1976), 459 – 463.

Acknowledgments. The authors owe thanks to Peter Bickel for pointing out the relevance of Hardy's inequality in the context of information bounds for survival analysis models. Thanks also go to Adrien Saumard for several helpful comments. Finally we thank the AE for suggesting that we should also include Muckenhoupt's inequality in the current study.