Article

Asymptotic Results in Broken Stick Models: The Approach via Lorenz Curves

Gheorghiță Zbăganu

"Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, 050711 Bucharest, Romania; gheorghitazbaganu@yahoo.com

check for updates

Abstract: A stick of length 1 is broken at random into n smaller sticks. How much inequality does this procedure produce? What happens if, instead of breaking a stick, we break a square? What happens asymptotically? Which is the most egalitarian distribution of the smaller sticks (or rectangles)? Usually, when studying inequality, one uses a Lorenz curve. The more egalitarian a distribution, the closer the Lorenz curve is to the first diagonal of $[0, 1]^2$. This is why in the first section we study the space of Lorenz curves. What is the limit of a convergent sequence of Lorenz curves? We try to answer these questions, firstly, in the deterministic case and based on the results obtained there in the stochastic one.

1. Introduction: Defining the Problem

Some time ago we were interested in the emergence and evolution of inequality if a whole is divided at random into n pieces. How much inequality is there in that division? The whole may be the surface, the population, the total GDP of the world, the number of species of insects or birds, etc. What means "at random", and how can we model this phenomenon?

This is the so-called "broken stick model". It was introduced by MacArthur [1] in 1957, and it has been extensively studied since then. A probabilistic approach can be found in Cohen [2]. Since "at random" usually means "uniformly distributed", these papers were about uniform spacings. Many books and papers deal with them, e.g., Aly [3], Bansal et al. [4], Beirlant [5], Barton and David [6], Csörgö [7], Cucala [8], Durbin [9], Le Cam [10], Pyke [11,12], Rao [13], Stephens [14,15], Tung [16], Wilks [17].

The inequality produced by these models was also considered via Lorenz curves, generalized Lorenz curves, and Gini coefficients (see for instance Arnold [18,19], Sarabia et al. [20] or Stephens [14,15]). The convergence of the empirical Lorenz curves to a theoretical one was studied by Goldie in 1977 (see Reference [21]). The order between Lorenz curves is a particular stochastic ordering (see Marshall and Olkin [22] and Stoyan [23]).

Our study was prompted by the strange fact that if we replace "at random = uniform distribution" with "at random = any absolutely continuous distribution on [0, 1]", the inequality produced in the broken stick model seems to increase. Simulations with different distributions made us conjecture that the Lorenz curves of the spacings always have a limit, and that limit is below the graph of $L(x) = x + (1 - x) \ln(1 - x)$, meaning that all the other distributions of spacings dominate the spacings produced by the uniform distribution in the Lorenz order.

We prove that in a particular case.

In order that this paper be as self-contained as possible, we study in the first three sections the case when the spacings are produced by a deterministic algorithm, not by a random one. The main results are Proposition 14 and its analog in terms of spacings, Proposition 20, which say that some mixtures increase the inequality.

The last two sections focus on the random case. The main result is Proposition 27 which implies the fact that if the density of the random variables used to break the stick is a step function, then the limit of the corresponding Lorenz curves does exist, and it lies below the Lorenz curve of the exponential distribution. The last section deals with the broken rectangle: the limit of the Lorenz curves does exist and one can compute its Gini coefficient even if we do not have an analytic formula for the Lorenz curve.

We have collected in the Appendix A several results which may be well known, but to which we were not able to find appropriate references.

2. Generalities: The Lorenz and Pre-Lorenz Curve of a Probability Distribution on the Non-Negative Half-Line

Notation. *Let $X \geq 0$ be a non-negative random variable. We will denote by the same letter F or F_X both its distribution and its distribution function. There will be no danger of confusion, since if B is a Borel set, then F(B) will mean $P(X \in B)$, while if $x \in \mathbb{R}$, F(x) will actually mean $F((-\infty, x]) = P(X \leq x)$. Moreover, as $F_X(x) = 0$, for all $x < 0$ we may as well think that $F : [0, \infty) \rightarrow \mathbb{R}$.*

We will denote by $F^{-1} : [0, 1) \rightarrow \mathbb{R}$ its **superior quantile** defined by $F^{-1}(p) = inf\{t : F(t) > p\}$.

Notice that we define the quantiles only on the interval $[0, 1)$, since $F^{-1}(1)$ may be ∞ if the distribution does not have a compact support.

The expectation EX depends only on F. In terms of distribution, it will be denoted by $e(F)$. Since X is non-negative, it is more or less obvious that:

$$e(F) = \int x dF(x) = \int_0^1 F^{-1}(y) dy \tag{1}$$

whereby $\int_0^1 F^{-1}(y) dy$, we understand the Riemann integral $\lim_{\varepsilon \downarrow 0} \int_0^{1-\varepsilon} F^{-1}(y) dy$. If X is integrable, then $e(F) < \infty$, otherwise $e(F) = \infty$.

Let $\Lambda_F : [0, 1) \rightarrow [0, \infty)$ be defined by:

$$\Lambda_F(p) = \int_0^p F^{-1}(y) dy \tag{2}$$

As F^{-1} is non-decreasing, $\Lambda_F(p) \leq p F^{-1}(p)$, thus the definition of Λ_F makes sense. We call Λ_F the **pre-Lorenz** curve of the distribution F.

The **Lorenz curve** is defined by:

$$L_F(p) = \frac{\Lambda_F(p)}{\Lambda_F(1-0)}, \ 0 \leq p < 1 \tag{3}$$

and makes sense only for distribution with finite expectation. Here and in the sequel $\Lambda_F(1-0) = \lim_{p \uparrow 1} \Lambda_F(p)$. The limit does exist due to the monotonicity reasons.

The pre-Lorenz curve makes sense for any probability distribution on $[0, \infty)$, even if its expectation is infinite.

The pre-Lorenz curve was introduced by Shorrocks [24] in 1983 under the name of "Generalized Lorenz Curve", and it was studied by many authors under this name. We think that "pre-Lorenz" is a more appropriate name, since one computes the Lorenz curve after computing its pre-Lorenz one and dividing by the expectation of the distribution.

Example 1. *Let $F = \begin{pmatrix} a & b \\ p & q \end{pmatrix} = p\delta_a + q\delta_b$ with $0 \leq a < b$.*

Here and in the sequel, δ_a means the Dirac measure at a defined by $\delta_a(B) = 1_B(a) = \begin{cases} 1 & \text{if } a \in B \\ 0 & \text{elsewhere} \end{cases}.$

Then $F^{-1}(t) = a1_{(0,\,p)}(t) + b1_{[p,1)}(t)$, therefore $\Lambda_F(t) = at1_{(0,p)}(t) + (ap + b(t - p))1_{[p,1)}(t)$, $\Lambda_F(1 - 0) = ap + bq,$

$$L_F(t) = \begin{cases} \frac{at}{ap+bq} & \text{if } t < p \\ \frac{ap+b(t-p)}{ap+bq} & \text{if } p \leq t < 1 \end{cases} \tag{4}$$

Example 2. *Let $F(x) = (1 - \frac{1}{x^\alpha})1_{[1,\infty)}(x)$ be the Pareto distribution of parameter $\alpha > 0$. Then $F^{-1}(p) = (1 - p)^{-\frac{1}{\alpha}}$ and*

$$\Lambda_F(p) = \begin{cases} \frac{\alpha}{1-\alpha}(1 - (1-p)^{-\frac{1-\alpha}{\alpha}}) & \text{if } \quad 0 < \alpha < 1 \\ -\ln(1-p) & \text{if } \quad \alpha = 1 \\ \frac{\alpha}{\alpha-1}(1 - (1-p)^{\frac{\alpha-1}{\alpha}}) & \text{if } \quad \alpha > 1 \end{cases}, \quad \Lambda_F(1-0) = \begin{cases} \infty & \text{if } \alpha \leq 1 \\ \frac{\alpha}{\alpha-1} & \text{if } \alpha > 1 \end{cases} \tag{5}$$

$$L_F(t) = \begin{cases} 0 & \text{if } \alpha \leq 1 \\ 1 - (1-p)^{\frac{\alpha-1}{\alpha}} & \text{if } \alpha > 1 \end{cases} \tag{6}$$

Unlike the mapping $F \mapsto L_F$, the mapping $F \mapsto \Lambda_F$ is a one-one operator.

Proposition 1. *Let P be the set of all the probability distributions on $([0, \infty), B([0, \infty)))$. Let $F \in P$. Then $\Lambda_F(0) = 0, \Lambda_F(1 - 0) = \Lambda_F(1) = e(F)$. Moreover, Λ_F is convex and non-decreasing, and its right derivative is F^{-1}.*

Proof. Obvious. Lemma A1 says that any convex function defined on $[a, b]$ is continuous on (a, b), and it has both right and left derivatives. Its right derivative is $(\Lambda_F)'_r(p) = \lim\limits_{\delta \downarrow 0} \frac{\int_p^{p+\delta} F^{-1}(y)dy}{\delta} = F^{-1}(p)$ for all $p \in [0, 1)$ if F^{-1} stands for the superior quantile which is right continuous. □

Let now C be the set of all convex functions $L : [0,1) \to [0, \infty)$ which are convex and non-decreasing. The following result concerns the unicity.

Proposition 2. *The mapping $F \to \Lambda_F$ from P to C is onto and one–one.*

Proof. Let F, G be two distributions from P such that $\Lambda_F = \Lambda_G$. Taking the right derivatives, it follows that $F^{-1} = G^{-1}$. As F, G are right continuous, it follows that $F = G$.

This is a consequence of the following elementary fact: Let $F, G : \mathbb{R} \to \mathbb{R}$ be non-decreasing functions. Let $\alpha : (F(-\infty), F(\infty)) \to \mathbb{R}$, $\beta : (G(-\infty), G(\infty)) \to \mathbb{R}$ be two functions having the property that $F(\alpha(p) - 0) \leq p \leq F(\alpha(p) + 0)$, $G(\beta(p) - 0) \leq p \leq G(\beta(p) + 0)$. If $\alpha = \beta$, then $F(x) = G(x)$ for any continuity point of F or G. Indeed, suppose, to make a choice, that $F(x) < G(x)$. Let p such that $F(x) < p < G(x)$. Thus $F(x) < F(\alpha(p) + 0)$, $G(x) > G(\alpha(p) - 0) \Rightarrow x \leq \alpha(p)$, $x \geq \alpha(p) \Rightarrow x = \alpha(p)$ and $F(x) < F(x + 0)$, $G(x) > G(x - 0)$. Thus, x must be a discontinuity point both for F and G. This set is at most countable.

On the other hand, if $\Lambda \in C$ then, its right derivative $\lambda = \Lambda'$ is right continuous and non-decreasing. According to Lemma A4, there exists a distribution function F such that $\lambda = F^{-1}$ hence $\Lambda_F = \Lambda$. □

Remark 1. *If F has finite expectation, we can see that $L_F(1 - 0) = 1$. Indeed, $L_F(p) = \frac{\Lambda_F(p)}{\Lambda_F(1)} = \frac{\Lambda_F(p)}{e(F)}$. If a function $L : [0, 1] \to \mathbb{R}$ is convex, non-decreasing, $L(0) = 0$, $L(1) = 1$ but $L(1 - 0) < 1$, we call it a* ***defective Lorenz curve.***

If F_n and F are distributions on the real line, their weak convergence will be denoted by "$F_n \Rightarrow F$". Thus $F_n \Rightarrow F$ means $\int f dF_n \to \int f dF$ for any continuous and bounded $f : [0, \infty) \to \mathbb{R}$.
The connection between the Λ- transform and the weak convergence is given by the following.

Proposition 3. *Let F_n, $F \in \mathbf{P}$ be probability distributions on $[0, \infty)$. Then $F_n \Rightarrow F$ if and only if $\Lambda_{F_n}(p) \to \Lambda_F(p)$ for any $p \in (0, 1)$. Moreover, if the limit Λ of the sequence $(\Lambda_{F_n})_n$ does exist and it is finite, then the weak limit of the sequence $(F_n)_n$ does exist, too, and $\Lambda = \Lambda_{\lim F_n}$.*

Proof. According to Lemma A5 from the Appendix A, $F_n \Rightarrow F$ if and only if $F_n^{-1}(p) \to F^{-1}(p)$ for $p \in [0, 1]$, with the possible exception of a countable set, N.

If $p \notin N$, then for any $\varepsilon > 0$ there exists n_ε such that $n \geq n_\varepsilon \Rightarrow F_n^{-1}(p) \leq F^{-1}(p) + \varepsilon$. As F_n^{-1} are non-decreasing and non-negative, for any $0 \leq F_n^{-1}(t) \leq F^{-1}(p) + \varepsilon$ for all $t \leq p$, $t \notin N$. Therefore, we may apply Lebesgue theorem of dominated convergence to infer that $\int_0^p F_n^{-1}(t)dt \to \int_0^p F^{-1}(t)dt$ or, which is the same thing, that $\Lambda_{F_n}(p) \to \Lambda_F(p)$.

Conversely, if $\Lambda_{F_n}(p) \to \Lambda_F(p)$, then, according to Lemma A3, the right derivatives converge too: $(\Lambda_{F_n})'(p) \to (\Lambda_F)'(p)$ at all points p where $(\Lambda_F)_r'(p) = (\Lambda_F)_l'(p)$. But the right derivatives are $F_n^{-1}(p)$ and $F^{-1}(p)$: it follows that $F_n^{-1} \to F^{-1}$ with the possible exception of a countable set. Apply again Lemma A5 (iv) to derive that F_n converges to F with the possible exception of a countable set. But this means precisely that $F_n \Rightarrow F$.

If we only know that $\Lambda_{F_n}(p) \to \Lambda(p)$ for all $0 < p < 1$ and Λ is finite, then its right derivative, Λ', is the superior quantile of some probability distribution on $[0, \infty)$, and the proof goes in the same way. □

The Lorenz curve does not have the same good properties: it is possible that $F_n \Rightarrow F$ but $L_{F_n}(p)$ does not converge to L_F.

Example 3. *Let $F_n = \begin{pmatrix} a & na \\ 1 - \frac{1}{n} & \frac{1}{n} \end{pmatrix}$, $a > 0$. Obviously $F_n \Rightarrow \delta_a$ but $L_{F_n}(t) =$*

$$\begin{cases} \frac{at}{2a - \frac{1}{n}} & \text{if } t < 1 - \frac{1}{n} \\ \frac{a(1 - \frac{1}{n}) + na(t - 1 + \frac{1}{n})}{2a - \frac{1}{n}} & \text{if } 1 - \frac{1}{n} \leq t < 1 \end{cases}$$ *converges to $\frac{t}{2}$ and **not** to $L_{\delta_a}(t) = t$.*

Example 4. *The same, modified a bit: $F_n = \begin{pmatrix} 1 & a_n + 1 \\ 1 - \frac{1}{n} & \frac{1}{n} \end{pmatrix}$. Then $F_n \Rightarrow \delta_1$, $\Lambda_{F_n}(p) \to p \; \forall p \in [0, 1)$ but $L_{F_n}(p) = \frac{p}{1 + \frac{a_n}{n}} \; \forall p \in [0, 1 - \frac{1}{n})$ may have no limit at all if the sequence $\frac{a_n}{n}$ has no limit.*
The reason is the loss of mass to the infinity: if $X_n \geq 0$ and $X_n \to X$ then $EX \leq \lim_{n \to \infty} EX_n$. This happens due to Fatou's Lemma.
If there is no loss of mass we can state.

Proposition 4. *Let $\mathbf{P_1}$ be the set of the probability distributions from $\mathbf{P}$ which have finite expectation. Suppose that F_n, $F \in \mathbf{P_1}$, $F_n \Rightarrow F$ and $e(F_n) \to e(F) > 0$. Then, $L_{F_n}(p) \to L_F(p)$ for all $0 \leq p \leq 1$.*
Conversely, if $L_n : [0, 1] \to [0, 1]$ are Lorenz curves, $L_n \to L$ and L is continuous at $p = 1$, then there exist probability distributions F_n and F from $\mathbf{P_1}$ such that $e(F_n) = e(F) = 1$ and $F_n \Rightarrow F$.
Or, reformulated in terms of random variables:
Let $(X_n)_{n \geq 0}$ be random variables from L^1. Let $X \geq 0$ be another random variable such that $EX = 1$. Then $\frac{X_n}{EX_n} \xrightarrow{D} X$ $\Rightarrow L_{X_n} \to L_X$. Conversely, if $L_{X_n} \to L$ and L is continuous at $p = 1$, then there exists a non-negative random variable on the probability space $(\Omega, K, P) = ((0, 1), B((0, 1)), \text{Lebesgue})$ such that $\frac{X_n}{EX_n} \xrightarrow{L^1} X$.

Proof. From Proposition 3, we know that $F_n \Rightarrow F \Leftrightarrow \Lambda_{F_n}(p) \to \Lambda_F(p)$. On the other hand, $e(F_n) = \Lambda_{F_n}(1 - 0)$ and $e(F_n) = \Lambda_F(1 - 0)$. Thus,

$F_n \Rightarrow F \Rightarrow \Lambda_{F_n}(p) \to \Lambda_F(p), \Lambda_{F_n}(1-0) \to \Lambda_F(1-0) \Rightarrow \frac{\Lambda_{F_n}(p)}{\Lambda_{F_n}(1-0)} \to \frac{\Lambda_F(p)}{\Lambda_F(1-0)} \Rightarrow L_{F_n}(p) \to L_F(p)$ for all $0 \le p \le 1$. Conversely, if $L_{F_n}(p) \to L_F(p)$, we apply Lemma A3 from the Appendix A, and the right derivatives converge in L^1. The right derivatives $\lambda_n = (L'_n)r$ and $\lambda = (L')r$ are right continuous and non-decreasing, thus there exist distribution functions F_n and F such that $F_n^{-1} = \lambda_n$ and $F^{-1} = \lambda$ (Lemma A4(iv)). Obviously, $e(F_n) = e(F) = \int_0^1 \lambda_n(x)dx = \int_0^1 \lambda(x)dx = 1$ and, according to Lemma A5, we see that $F_n \Rightarrow F$. □

The situation changes if $e(F_n)$ does not converge to $e(F)$.

If the sequence $(e(F_n))_n$ has no limit at all, then Example 4 points out that the sequence $L_{F_n}(p) = \frac{\Lambda_{F_n}(p)}{e(F_n)}$ has no limit, too, (since $\Lambda_{F_n}(p) \to \Lambda_F(p)$ for any $0 < p < 1$).

If $e(F_n) \to ke(F)$, we know that $k \ge 1$ (Fatou's Lemma). Then the limit is a defective Lorenz curve: $\lim_n L_{F_n}(p) = \lim_n \frac{\Lambda_{F_n}(p)}{e(F_n)} = \frac{\Lambda_F(p)}{ke(F)} = \frac{1}{k}L_F(p)$. Thus, we can add to the proposition 4 the following fact.

Proposition 5. *Suppose that $F_n, F \in P_1$, $F_n \Rightarrow F$ and $e(F_n) \to ke(F) > 0$. Then, $L_{F_n}(p) \to \frac{1}{k}L_F(p)$ for all $0 \le p < 1$. If the limit of a sequence of Lorenz curves is defective, then the convergence does not ensure the convergence of expected values. As a limit case, if $L_{F_n}(p) \to 0$ for $0 \le p < 1$ and $0 < e(F_n) < M$ for some $M > 0$ then, $F_n \Rightarrow \delta_0$.*

Proof. Only the last claim needs a proof. Remark that $\Lambda_{\delta_0}(p) = 0$ thus if $L_{F_n}(p) \to 0$ then $\Lambda_{F_n} \to \Lambda_{\delta_0}$. In this case Proposition 3 says that $F_n \Rightarrow \delta_0$. If the the sequence $(e(F_n))_n$ is not bounded, the claim is not true. For instance, if $F_n = \begin{pmatrix} 1 & n^2 \\ 1-\frac{1}{n} & \frac{1}{n} \end{pmatrix}$ then $L_{F_n}(p) \to 0$ but $F_n \Rightarrow \delta_1$ and not to δ_0. □

3. The Empirical Lorenz and Pre-Lorenz Curves of a Sequence of Non-Negative Real Numbers

Here, we study the deterministic case: if the sequence of points which generates the smaller sticks is an usual one. We generalize a bit, giving up the restriction that the points lie between 0 and 1.

Let $a = (a_n)_{n \ge 1}$ be a sequence of non-negative reals. We attach to it the sequence of probability distributions.

$$F_n = F_n(a) = \frac{\sum_{k=1}^n \delta_{a_k}}{n}, \quad F_n^* = F_n^*(a) = \frac{\sum_{k=1}^n \delta_{\frac{n}{s_n}a_k}}{n}, \quad s_n = a_1 + \cdots + a_n \tag{7}$$

Next, we denote by $\Lambda_n = \Lambda_n(a)$ and $L_n = L_n(a)$ the pre-Lorenz and the Lorenz curves of F_n. Thus:

$$\Lambda_n = \Lambda_n(a) = \Lambda_{F_n}(a), \quad L_n = L_n(a) = L_{F_n}(a) \tag{8}$$

We want to know when these curves have a limit. First, a remark which is useful for calculus:

Proposition 6. *Let $a = (a_n)_{n \ge 1}$ be a sequence of non-negative reals. Let $n \ge 2$ be fixed. Let also $O_n(a) = (a_{1:n}, a_{2:n}, \ldots, a_{n,n})$ be the sorted (ascending) values of $(a_1, a_2 \ldots, a_n)$. Denote by $S_{k:n}$ the sum*

$$S_{k:n} = a_{1:n} + a_{2:n} + \cdots + a_{k:n}, \text{ for } 1 \le k \le n \tag{9}$$

Then,

(i)　Λ_n *is the polygonal line which joins the points* $(0,0)$, $(\frac{1}{n}, \frac{S_{1,n}}{n})$, $(\frac{2}{n}, \frac{S_{2,n}}{n})$, $\ldots$, $(\frac{n}{n}, \frac{S_{n,n}}{n})$, *and L_n is the polygonal line which joins the points* $(0,0)$, $(\frac{1}{n}, \frac{S_{1,n}}{S_{n,n}})$, $(\frac{2}{n}, \frac{S_{2,n}}{S_{n,n}})$, $\ldots$, $(\frac{n}{n}, \frac{S_{n,n}}{S_{n,n}})$. *Or, formally, for any $1 \le \cdots k \le n$ we have:*

$$\Lambda_n(\frac{k-1+\varepsilon}{n}) = \frac{(1-\varepsilon)S_{k-1,n} + \varepsilon S_{k,n}}{n}, L_n(\frac{k-1+\varepsilon}{n}) = \frac{(1-\varepsilon)S_{k-1,n} + \varepsilon S_{k,n}}{S_{n,n}} \tag{10}$$

(ii) $L_{F_n}(\boldsymbol{a}) = \Lambda_{F_n^*}(\boldsymbol{a})$.

(iii) $\Lambda_n(\alpha\boldsymbol{a}) = \alpha\Lambda_n(\boldsymbol{a})$, $L_n(\alpha\boldsymbol{a}) = L_n(\boldsymbol{a})$ for any $\alpha > 0$. The Lorenz curve is invariant to homotheties.

(iv)   if $\sigma \in S_n$ is a permutation, then $\Lambda_n(a_{(1)}, a_{(2)}, \ldots, a_{(n)}) = \Lambda_n(\boldsymbol{a})$. Both Λ_n and L_n are invariant with respect to permutation of the first n terms of $\boldsymbol{a}$.

Proof. The distribution function of F_n is $F_n(x) = \frac{\sum_{k=1}^n 1_{[a_k, \infty)}(x)}{n}$ a.s.o. The assertions are elementary and left to the reader. $\square$

Example 5. *If z_j are some points in plane, we denote by* $\boldsymbol{Poly}(z_1 \to z_2 \to \ldots \to z_k) = \boldsymbol{Poly}(z_j \, ; \, 1 \leq j \leq k)$ *the polygonal line which joins the points $z_1, z_2, \ldots, z_k$ from plane (in this order). Then:*

(i) $\boldsymbol{a} = (1, 0, 1, 0, 1, 0, \ldots)$. *Then $\boldsymbol{O_n}(\boldsymbol{a}) = (0, \ldots, 0, 1, \ldots, 1)$ where, if $n = 2k$, we have k zeros and k ones, and if $n = 2k+1$, we have k zeros and $k+1$ ones. If $n = 2k$, then Λ_n is the polygonal line which joins the points* $(0, 0)$, $(\frac{1}{n}, 0), \ldots, (\frac{k}{n}, 0), (\frac{k+1}{n}, \frac{1}{n}), \ldots, (\frac{n}{n}, \frac{k}{n})$. *Many points are on the same line, thus it is easier to write $\Lambda_{2k} = \boldsymbol{Poly}((0,0) \to (\frac{1}{2}, 0) \to (1, \frac{1}{2}))$ and $\Lambda_{2k+1} = \boldsymbol{Poly}((0,0) \to (\frac{k}{2k+1}, 0) \to (1, \frac{k+1}{2k+1}))$. As $S_{n:n} = k$ if $n = 2k$ and $S_{n:n} = k+1$ if $n = 2k+1$, it follows that $L_{2k} = \boldsymbol{Poly}((0,0) \to (\frac{1}{2}, 0) \to (1, 1))$ and $\Lambda_{2k+1} = \boldsymbol{Poly}((0,0) \to (\frac{k}{2k+1}, 0) \to (1, 1))$.*

(ii)   *More general, if $\boldsymbol{a}$ is a periodic sequence: $\boldsymbol{a} = (x_1, \ldots, x_k, x_1, \ldots, x_k, \ldots)$, then the describing of Λ_n and L_n become very annoying if we do not know the ordering of the numbers x_j. But if we assume that $x_1 \leq \ldots \leq x_k$, the describing can be done as follows: suppose that $n = km + r$ with $0 \leq r < k$, then $\boldsymbol{O_n}(\boldsymbol{a}) = (x_1, \ldots, x_1, x_2, \ldots, x_2, \ldots, x_r, \ldots, x_r, x_{r+1}, \ldots, x_{r+1}, \ldots, x_k, \ldots, x_k)$, where x_j occur $m + 1$ times if $j \leq r$, and it occurs m times if $j > r$. Thus $S_{n:n} = mS_{k:k} + S_{r:k}$ and*

$$\Lambda_n = \boldsymbol{Poly}((0,0) \to (\tfrac{m+1}{n}, \tfrac{m+1}{n} x_1) \to (\tfrac{2(m+1)}{n}, \tfrac{m+1}{n}(x_1 +$$

$$x_2)) \to \ldots \to (\tfrac{r(m+1)}{n}, \tfrac{m+1}{n}(x_1 + \ldots + x_r)) \to (\tfrac{r(m+1)+m}{n}, \tfrac{m+1}{n}(x_1 + x_2 + \ldots + x_r) +$$

$$\tfrac{m}{n} x_{r+1}) \to (\tfrac{r(m+1)+m(k-r)}{n}, \tfrac{m+1}{n}(x_1 + x_2 + \ldots + x_r) + \tfrac{m}{n}(x_{r+1} + \ldots + x_k)) \to (1, \tfrac{mS_{k:k}+S_{r:k}}{n})), \text{ and}$$

$$L_n = \boldsymbol{Poly}((0,0) \to (\tfrac{m+1}{n}, \tfrac{m+1}{S_{n:n}} x_1) \to (\tfrac{2(m+1)}{n}, \tfrac{m+1}{S_{n:n}}(x_1 + x_2)) \to \ldots \to (\tfrac{r(m+1)}{n}, \tfrac{m+1}{S_{n:n}}(x_1 + x_2 + \ldots +$$

$$x_r)) \to (\tfrac{r(m+1)+m}{n}, \tfrac{m+1}{S_{n:n}}(x_1 + x_2 + \ldots + x_r) + \tfrac{m}{n} x_{r+1}) \to (\tfrac{r(m+1)+m(k-r)}{n}, \tfrac{m+1}{S_{n:n}}(x_1 + x_2 + \ldots + x_r) +$$

$$\tfrac{m}{S_{n:n}}(x_{r+1} + \ldots + x_k)) \to (1, 1))$$

(iii)   *Suppose that $a_n = \alpha^n$. If $\alpha > 0$, the sequence is increasing, if $\alpha = 0$ it is constant, and if $\alpha < 0$ it is decreasing. In the first two cases, no sort is necessary hence $\Lambda_n = \boldsymbol{Poly}((0,0) \to \{(\frac{k}{n}, \frac{s_k}{n}); 1 \leq k \leq n\})$, $L_n = \boldsymbol{Poly}((0,0) \to \{(\frac{k}{n}, \frac{s_k}{s_n}); 1 \leq k \leq n\})$ where $s_n = 1^\alpha + 2^\alpha + \ldots + n^\alpha$ and $s_0 = 0$. If $\alpha = -\beta < 0$ then $\boldsymbol{O_n}(\boldsymbol{a}) = (\frac{1}{n^\beta}, \frac{1}{(n-1)^\beta}, \ldots, \frac{1}{2^\beta}, 1)$, therefore $\Lambda_n = \boldsymbol{Poly}((0,0) \to \{(\frac{k}{n}, \frac{s_k - s_{n-k}}{n}); 1 \leq k \leq n\})$, $L_n = \boldsymbol{Poly}((0,0) \to \{(\frac{k}{n}, \frac{s_k - s_{n-k}}{s_n}); 1 \leq k \leq n\})$*

(iv) $\boldsymbol{a} = (1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, \ldots (2^4 \text{ times}), 0, 0, \ldots (2^5 \text{ times}), \ldots)$ *Let $N = 1 + 2 + \ldots + 2^{n-1} = 2^n - 1$. If $n = 2m$ is even, $\boldsymbol{O_n}(\boldsymbol{a}) = (0, 0, \ldots (\frac{2}{3}(2^n - 1) \text{ times}), 1, 1, \ldots, (\frac{1}{3}(2^n - 1) \text{ times}))$ and if $n = 2m + 1$ is odd, $\boldsymbol{O_n}(\boldsymbol{a}) = (0, 0 \ldots (\frac{1}{3}(2^n - 1) \text{ times}), 1, 1, \ldots, (\frac{2}{3}(2^n - 1) \text{ times}))$. In the first case, $F_n = \begin{pmatrix} 0 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}$, and in the second one it is $F_n = \begin{pmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$. For other types of n, we obtain intermediate values. The pre-Lorenz curves move between $(p - \frac{2}{3})_+$ and $(p - \frac{1}{3})_+$, and the Lorenz ones move between $3(p - \frac{2}{3})_+$ and $2(p - \frac{1}{3})_+$. None of them have a limit.*

The first problem is: for what kind of sequences does the sequence $\Lambda_n(\boldsymbol{a})$ have a finite limit.

This is the easy one.

Proposition 7. *Let* $\boldsymbol{a} = (a_n)_{n \geq 1}$ *be a sequence of non-negative integers. The following assertions are equivalent:*

(i) *The sequence* $(F_n(\boldsymbol{a}))_{n \geq 1}$ *has a weak limit,* $F = F(\boldsymbol{a})$;
(ii) *The sequence* $(\Lambda_n(\boldsymbol{a}))_{n \geq 1}$ *has a finite limit,* $\Lambda_F(\boldsymbol{a})$;
(iii) *The sequence* $(f(a_n))_{n \geq 1}$ *has finite Cesaro limit for any continuous bounded* $f : [0, \infty) \to \mathbb{R}$ *and this limit is equal to* $\int f dF$.

Proof. The equivalence (i) $\Leftrightarrow$ (ii) is actually Proposition 3. According to the well-known Portmanteau proposition (see, for instance, References [25,26]), one knows that $F_n \Rightarrow F$ if and only if $\int f dF_n \to \int f dF$ for any $f : \mathbb{R} \to \mathbb{R}$ bounded continuous. In our case, $\int f dF_n(\boldsymbol{a}) = \frac{f(a_1) + f(a_2) + \dots + f(a_n)}{n}$. If this limit exists for any *continuous and bounded* f, then the sequence of probability measures $(F_n)_n$ should have a limit weak limit too.

In order to detect the limit, we have to study the empirical distribution functions $F_n(x) = \frac{|\{k \leq n : a_k \leq x\}|}{n}$. The hard way is to compute $\Lambda_n(\boldsymbol{a})$. If we want to find the limit of $L_n(\boldsymbol{a})$, the fact that $F_n \Rightarrow F$ does not help very much unless we can prove that $e(F_n)$ converges to $e(F)$, too. □

Sometimes the following result may help.

Lemma 1. *Let* $f_n : [a, b] \to [m, M]$ *be convex and non-decreasing. Suppose that for any* $x \in (a, b)$ *there exists a sequence* $(x_n)_n$ *such that* $x_n \to x$ *and the sequence* $(f_n(x_n))_n$ *is convergent.*
Then, the limit $f(x) := \lim_n f_n(x)$ *does exist for any* $x \in (a, b)$, *and it is convex and non-decreasing. Moreover,* $f(x) = \lim_n f_n(x)$.

Proof. Let $a < x < b$ and $f(x) = \lim_{n \to \infty} f_n(x_n)$. Write $f(x) - f_n(x) = f(x) - f_n(x_n) + f_n(x_n) - f_n(x)$. Let $\varepsilon > 0$ be small enough in order that $a < x - \varepsilon < x + \varepsilon < b$. Let n_ε be such that $n \geq n \Rightarrow x - \varepsilon < x_n < x + \varepsilon$. For $n \geq n_\varepsilon$, we have $f_n'{}_l(x - \varepsilon) \leq \frac{f_n(x) - f_n(x_n)}{x - x_n} \leq (f_n)'{}_r(x + \varepsilon)$ hence $|f_n(x) - f_n(x_n)| \leq (f_n)'{}_r(x + \varepsilon)|x - x_n|$. By chord inequality, $(f_n)'{}_r(x + \varepsilon) \leq \frac{f_n(b) - f_n(x + \varepsilon)}{b - (x + \varepsilon)}$ and the last quantity is smaller than $\frac{M - m}{b - (x + \varepsilon)}$. Thus, we obtained the evaluation $\left| f(x) - f_n(x) \right| \leq |f(x) - f_n(x_n)| + \frac{M - m}{b - (x + \varepsilon)} |x - x_n| \to 0$ if $n \to \infty$. □

A first positive answer for the existence of the limit is:

Proposition 8. *Let* $\boldsymbol{a} \geq 0$. *Suppose that* $a_n \to a$. *If* $a > 0$, *then* $F_n(\boldsymbol{a}) \Rightarrow \delta_a$, $F_n^*(\boldsymbol{a}) \Rightarrow \delta_0$ *and* $L_n(\boldsymbol{a})(t) \to t$, $0 \leq t \leq 1$. *If* $= 0$, *then* $F_n(\boldsymbol{a}) \Rightarrow \delta_0$. *If, moreover, the series* $\sum_{k=1}^\infty a_k$ *is convergent then* $F_n^*(\boldsymbol{a}) \Rightarrow \delta_0$ *and* $L_n(\boldsymbol{a})(t) \to 0$.

Proof. The first claim is obvious: if f is continuous and bounded, then $\lim \frac{\sum_{k=1}^n f(a_k)}{n} = \lim f(a_n) = f(a)$ for $a \geq 0$. To prove that $F_n^*(\boldsymbol{a}) \Rightarrow \delta_1$ write $a_n = a + \varepsilon_n$ with $\varepsilon_n \to 0$. Let $T_n = \varepsilon_1 + \dots + \varepsilon_n$, $S_n = a_1 + \dots + a_n = na + T_n$. Notice that $\frac{na_k}{S_n} = \frac{na + n\varepsilon_k}{na + T_n} = \frac{1 + \frac{\varepsilon_k}{n}}{1 + \frac{T_n}{na}}$ converge to 1 as $n \to \infty$ for any $k \leq n$. Let $f : [0, \infty) \to [0, \infty)$ be uniformly continuous and bounded. Let $\varepsilon > 0$ be arbitrary small and let $\delta = \delta(\varepsilon)$ be such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$.

Let n_ε be such that $n > n_\varepsilon \Rightarrow \left| \frac{na_k}{S_n} - 1 \right| < \delta$. Notice that $\lim_{n \to \infty} \frac{\sum_{k=1}^n f(\frac{na_k}{S_n})}{n} = \lim_{n \to \infty} \frac{\sum_{k=n_\varepsilon+1}^n f(\frac{na_k}{S_n})}{n}$ hence

$$f(1) - \varepsilon = \lim_{n \to \infty} \frac{(n - n_\varepsilon)(f(1) - \varepsilon)}{n} \leq \lim_{n \to \infty} \frac{\sum_{k=n_\varepsilon+1}^n f(\frac{na_k}{S_n})}{n} \leq \lim_{n \to \infty} \frac{(n - n_\varepsilon)(f(1) + \varepsilon)}{n} = f(1) + \varepsilon.$$

As ε is arbitrary, $\lim_{n \to \infty} \frac{\sum_{k=1}^n f(\frac{na_k}{S_n})}{n} = f(1) = \int f d\delta_1$. As $e(\delta_1) = e(F_n^*) = 1$, the Lorenz curves $L_n(\boldsymbol{a})$ converge to the Lorenz curve of δ_1, namely, to $L(t) = t$.

If $a = 0$, we compute the Lorenz curve, i.e., the polygonal line given by the points $\left(\frac{k}{n}, \frac{b_1+\ldots+b_k}{S_n}\right)$, where $b_k = a_{k:n}$. Let $c_n = b_{n-k+1}$ and $_k = c_1 + \cdots + c_k$. Note that the sequence $(c_n)_n$ is non-increasing, $S_n = \Sigma_n = c_1 + \cdots + c_n, b_1 + \cdots + b_k = \Sigma_n - \Sigma_{n-k}$. If we write $\frac{b_1+\ldots+b_{[nt]}}{S_n} = \frac{\Sigma_n - \Sigma_{[nt]}}{\Sigma_n} = 1 - \frac{\Sigma_{[nt]}}{\Sigma_n}$ and remark that $0 < t < 1 \Rightarrow lim\Sigma_n = lim\, \Sigma_{[nt]} = \sum_{k=1}^{\infty} a_k$, it becomes obvious that $\lim_{n\to\infty} \frac{b_1+\ldots+b_{[nt]}}{S_n} = 0$. $\Leftrightarrow L_n(a) \to 0$.

By Proposition 6 (ii), we know that $L_{F_n}(a) = \Lambda_{F_n^*}(a)$. As $\Lambda_{F_n^*}$ converges to 0, it results that $F_n^* \Rightarrow \delta_0$. $\square$

Remark 2. *If $\sum_{k=1}^{\infty} a_k = \infty$, $a_n \to 0$ we can say nothing about the limit of $L_n(a)$. For instance, if $a = \left(\frac{1}{n^\alpha}\right)_{n\geq 1} 0 < \alpha < 1$ is the sequence from Example 5(iii) below, then $a_n \to 0$ but the sequence $L_n(a)(p)$ has the limit$L(p) = 1 - (1-p)^{1-\alpha}$. Therefore, the limit of $F_n^*(a)$ is the probability distribution with the distribution function $F^*(x) = \left(1 - \left(\frac{1-\alpha}{x}\right)^{\frac{1}{\alpha}}\right)_+$ which is a Pareto-type one. Thus, it is very possible that $F_n(a)$ and $F_n(b)$ have the same limit, but $F_n^*(a)$ and $F_n^*(b)$ have a different limit—if any. Conversely, $F_n^*(a)$ and $F_n^*(\lambda a)$ have the same limit, but $F_n(a)$ and $F_n(\lambda a)$ have different limits. Many other examples will follow.*

Remark 3. *The order does matter. Precisely, if we know that $F_n(a) \Rightarrow F$ and we modify the terms of a according to some permutation $\sigma : \{1, 2, \ldots\} \to \{1, 2, \ldots\}$, then we can say nothing about the limit of $F_n(a \circ \sigma)$. Suppose, for instance, that $a = (0,1,0,1,0,1, \ldots .)$ and that σ is chosen in such a way that $a \circ \sigma = (0,1,0,0,1,1,0,0,0,0,1,1,1,1,0(23\ times),1(23\ times),0(24\ times),1(24\ times),\ldots\ldots)$. The reader can check that $(F_n(a \circ \sigma))_n$ has no limit. This is a main difference between the Cesaro convergence and the usual convergence.*
We study now what happens if we modify the sequence a. The question is: how sensitive is L with respect to that change, and, in general if $F_n(a)$ has a limit (or $F_n^*(a)$ has a limit) what can we say about the limits of $F_n(\varphi(a))$ or $F_n^*(\varphi(a))$ if $\varphi : [0, \infty) \to [0, \infty)$?
A partial answer is:

Proposition 9. *(i) Let a be a non-decreasing and non-bounded sequence of non-negative reals. Let b be a bounded sequence of non-negative reals. Suppose that $(F_n^*(a))_n$ is weakly convergent to some probability distribution F and that $\frac{S_n^2}{n^2 a_n} \to \infty$ as $n \to \infty$, where $S_n = a_1 + a_2 + \cdots + a_n$.*
Then $F_n^*(a + b) \Rightarrow F$, too, and $L_n(a + b) \to L_F$.
(ii) If $F_n(a) \Rightarrow F$, then $F_n(\varphi(a)) \Rightarrow F \circ \phi^{-1}$.
(iii) If $L_n(a) \to L$, then $L_n(\alpha a + \beta)(t) \to \frac{L(t)+\frac{\beta}{\alpha\mu}t}{1+\frac{\beta}{\alpha\mu}t}$, where $\mu = \lim_n \frac{S_n}{n} > 0, \alpha > 0, \beta > 0, t \in (0,1)$.

Proof. (i). Let $T_n = b_1 + b_2 + \cdots + b_n$. The assumption is that $\int f dF_n^*(a) \to \int f dF$ for any $f : \mathbb{R} \to \mathbb{R}$ bounded and uniformly continuous. According to the definition of F_n^*, that is the same with:

$$\frac{\sum_{j=1}^{n} f\left(\frac{n}{S_n} a_j\right)}{n} \to \int f dF \tag{11}$$

The claim is that:

$$\frac{\sum_{j=1}^{n} f\left(\frac{n}{S_n+T_n}(a_j + b_j)\right)}{n} \to \int f dF \tag{12}$$

Let $\varepsilon > 0$ be arbitrary and let $\delta = \delta(\varepsilon)$ such that $|x - x'| \leq \delta \Rightarrow |f(x) - f(x')| \leq \varepsilon$. We prove that there exists a positive integer n_0 such that $n \geq n_0 \Rightarrow \left|\frac{n}{S_n} a_j - \frac{n}{S_n+T_n}(a_j + b_j)\right| < \delta$ for any $1 \leq j \leq n$. Indeed, let $M > 0$ such that $b_j \leq M$ for any j. Since a is non-decreasing, we have $\left|\frac{na_j}{S_n} - \frac{n(a_j+b_j)}{S_n+T_n}\right| = \frac{n}{S_n}\frac{|T_n a_j - S_n b_j|}{S_n+T_n} \leq \frac{n}{S_n}\frac{T_n a_j + S_n b_j}{S_n+T_n} \leq \frac{n}{S_n}\frac{nMa_j + S_n M}{S_n+T_n} \leq \frac{nM}{S_n}\frac{na_n + S_n}{S_n+T_n} \leq M\left(\frac{n^2 a_n}{S_n^2} + \frac{n}{S_n}\right)$.

According to our assumptions $\frac{S_n^2}{n^2 a_n} \to \infty \Rightarrow \frac{n^2 a_n}{S_n^2} \to 0$. On the other hand, $\frac{n}{S_n} \to 0$, since a_n is non-decreasing and non-bounded. It follows that there exists n_0 such that $n \geq n_0 \Rightarrow \left| \frac{na_j}{S_n} - \frac{n(a_j + b_j)}{S_n + T_n} \right|$

$< \delta$ for $1 \leq j \leq n \Rightarrow \left| f(\frac{na_j}{S_n}) - f(\frac{n(a_j + b_j)}{S_n + T_n}) \right| < \varepsilon$. Therefore, $\frac{\left| \sum_{j=1}^{n} f(\frac{n}{S_n + T_n}(a_j + b_j)) - \sum_{j=1}^{n} f(\frac{na_j}{S_n}) \right|}{n} < \varepsilon$ for any $n \geq n_0$. As ε is arbitrary, the claim (12) immediately follows.

(ii) is obvious.

(iii) The assumption is that the sequence of polygonal lines given by the points $\left(\frac{k}{n}, \frac{a_{1:n} + \dots + a_{k:n}}{S_n} \right)_{0 \leq k \leq n}$ converges to some Lorenz curve L. We want to see what happens with the polygonal lines given by the points $\left(\frac{k}{n}, \frac{\alpha(a_{1:n} + \dots + a_{k:n}) + k\beta}{\alpha S_n + n\beta} \right)_{0 \leq k \leq n} = \left(\frac{k}{n}, \frac{\frac{a_{1:n} + \dots + a_{k:n}}{S_n} + \frac{k\beta}{\alpha S_n}}{1 + \frac{n\beta}{\alpha S_n}} \right)_{0 \leq k \leq n}$. If $k = [nt]$ with $0 < t < 1$, then

$n \to \infty \Rightarrow \frac{a_{1:n} + \dots + a_{k:n}}{S_n} \to L(t)$, $k \frac{\beta}{\alpha S_n} \to \frac{\beta t}{\alpha \mu}$, $\frac{n\beta}{\alpha S_n} \to \frac{\beta}{\alpha \mu}$ if $\mu < \infty$. If $\mu = \infty$, the limit remains $L(t)$. $\square$

Remark 4. *If $\mu = 0$, the result fails to be true. However, one may check that if $\sum_{k=1}^{\infty} a_k < \infty$, then $F_n(\alpha a + \beta) \Rightarrow \delta_1$.*

Remark 5. *One may see that if $\beta \geq 0$, the Lorenz curve increases. This is a particular case of a theorem of Arnold [27] which says that if $X \geq 0$ a random variable, $EX > 0$ then the Lorenz curve of $f(X)$ is above the Lorenz curve of X if and only if f is non-decreasing and $\frac{f(x)}{x}$ is non-increasing. In our case $f(x) = \alpha x + \beta$.*

Combining. *Now we study what happens if we combine two sequences a and b. Precisely, if we know the limits for $L_n(a)$ and $L_n(b)$, what can we say about the limit of this combination.*
By combination we mean the sequence $a \& b = (a_1, b_1, a_2, b_2, \dots.)$.
A problem that arises is to describe the limit of $F_n^*(a \& b)$ if we know the limits of $F_n^*(a)$ and $F_n^*(b)$.

We cannot expect the result to be some mixture of the two limit distributions as the following example shows:

Proposition 10. *Let $a = (1, 2, 3, 4, \dots..)$ and $b = (\alpha, 2\alpha, 3\alpha, \dots.)$. Then $F_n^*(a) = F_n^*(b) \Rightarrow Uniform(0, 2)$. However $F_n^*(a \& b) \Rightarrow \frac{1}{2}\left(\textbf{Uniform}(0, \frac{4}{1+\alpha}) + \textbf{Uniform}(0, \frac{4\alpha}{1+\alpha}) \right)$.*

Proof. $F_n^*(a) = F_n^*(b) = \frac{\sum_{k=1}^{n} \delta_{\frac{2k}{n+1}}}{n}$. If f is bounded and continuous, then $\int f dF_n^*(a) = \frac{\sum_{k=1}^{n} f(\frac{2k}{n+1})}{n} \to \int_0^1 f(2x) dx = \frac{1}{2} \int_0^2 f(x) dx = \int f d\mu$ where $\mu = Uniform(0, 2)$.

In order to compute $\int f dF_n^*(a \& b)$, we have to consider two cases: if n is odd or even. If $n = 2m$ then:

$$\int f dF_n^*(a \& b) = \frac{\sum_{j=1}^{m} f(\frac{4j}{(1+\alpha)(m+1)}) + \sum_{j=1}^{m} f(\frac{4\alpha j}{(1+\alpha)(m+1)})}{2m} =$$
$$= \frac{m+1}{2m} \left(\frac{1}{m+1} \sum_{j=1}^{m} f(\frac{4}{1+\alpha} \cdot \frac{j}{m+1}) \right)$$
$$+ \frac{2}{2m+1} \sum_{j=1}^{m} f(\frac{4\alpha}{1+\alpha} \cdot \frac{j}{m+1}))$$

and the last sum converges to:

$$\frac{1}{2}\left(\int_0^1 f(\frac{4}{1+\alpha}x) dx + \int_0^1 f(\frac{4\alpha}{1+\alpha}x) dx \right) = \frac{1}{2} \int f d(\textbf{Uniform}(0, \frac{4}{1+\alpha}) + \textbf{Uniform}(0, \frac{4\alpha}{1+\alpha}))$$

If $n = 2m + 1$, the computations are almost the same.

Notice that if $\alpha = 1$ (hence $a = b$), then $F_n^*(a \& b) \Rightarrow \textbf{Uniform}(0, 2)$. $\square$

There is no general result concerning the limit of $L_n(a \& b)$. However, we can state some partial ones.

Proposition 11. *Let a, b be two non-decreasing sequences of positive numbers and let $S_n = a_1 + a_2 + \cdots + a_n$. Suppose that $a\&b$ is also non-decreasing, $\frac{S_{n+1}}{S_n} \to 1$ as $n \to \infty$ and that the weak limit $\lim_n F_n^*(a) = F$ does exist and $e(F) = 1$.*
Then, both $L_n(b)$ and $L_n(a\&b)$ converge to the Lorenz curve of F, L_F.
Or, otherwise written, $\lim\limits_{n\to\infty} F_n^*(b) = \lim\limits_{n\to\infty} F_n^*(a\&b) = F$.

Proof. The assumption is that the points $\left(\frac{[np]}{n}, \frac{S_{[np]}}{S_n}\right) \to (p, L_F(p)$ as $n \to \infty$ for any $0 < p < 1$. Let then $T_n = b_1 + b_2 + \cdots + b_n$. Notice that $S_n \leq T_n \leq S_{n+1}$. It means that $\frac{S_{[np]}}{S_{n+1}} \leq \frac{T_{[np]}}{S_n} \leq \frac{S_{[np]+1}}{S_n}$. We supposed that $\frac{S_{n+1}}{S_n} \to 1$ as $n \to \infty$. It follows that $\lim\limits_{n\to\infty} \frac{S_{[np]+1}}{S_n} = \lim\limits_{n\to\infty} \frac{S_{[np]+1}}{S_{[np]}} \lim\limits_{n\to\infty} \frac{S_{[np]}}{S_n} = 1 \cdot L_F(p)$ and $\lim\limits_{n\to\infty} \frac{S_{[np]}}{S_{n+1}} = \lim\limits_{n\to\infty} \frac{S_{[np]}}{S_n} \lim\limits_{n\to\infty} \frac{S_n}{S_{n+1}} = L_F(p) \cdot 1$ hence $\lim\limits_{n\to\infty} \left(\frac{[np]}{n}, \frac{T_{[np]}}{T_n}\right) = (p, L_F(p))$, or, which is the same thing, that $L_n(b) \to L_F$. □

Let $c = a\&b$ and let $U_n = c_1 + \cdots + c_n$. If $n = 2m$ is even, then $U_n = S_n + T_n$ an if $n = 2m - 1$ is odd, then $U_n = S_m + T_{m-1}$. On the other hand:

$$\frac{U_k}{U_n} = \begin{cases} \frac{S_j + T_j}{S_m + T_m} & if\ k = 2j, n = 2m \\ \frac{S_j + T_{j-1}}{S_m + T_m} & if\ k = 2j - 1, n = 2m \\ \frac{S_j + T_j}{S_m + T_{m-1}} & if\ k = 2j, n = 2m - 1 \\ \frac{S_j + T_{j-1}}{S_m + T_{m-1}} & if\ k = 2j - 1, n = 2m - 1 \end{cases} \tag{13}$$

Suppose that $n \to \infty$ and $k \to \infty$ in such a way that $\frac{k}{n} \to p$ for some $0 < p < 1$. Then $\frac{[k/2]}{[n/2]} \to p$ too. In other words, $\frac{j}{m} \to p$, where j and m are those from (13). In all the four cases $\frac{U_k}{U_n} \to L_F(p)$.
For instance:

$$\left|\frac{S_j + T_j}{S_m + T_m} - L_F(p)\right| = \left|\frac{S_m}{S_m + T_m}\left(\frac{S_j}{S_m} - L_F(p)\right) + \frac{T_m}{S_m + T_m}\left(\frac{T_j}{T_m} - L_F(p)\right)\right| \leq max\left(\left|\frac{S_j}{S_m} - L_F(p)\right|, \left|\frac{T_j}{T_m} - L_F(p)\right|\right)$$

As a byproduct, we may notice:

Proposition 12. *Let a, b be two non-decreasing sequences of positive numbers. Suppose that $a\&b$ is also non-decreasing, the weak limits coincide: $\lim\limits_{n\to\infty} F_n^*(a) = \lim\limits_{n\to\infty} F_n^*(b) = F$ and $e(F) = 1$. Then $\lim\limits_{n\to\infty} F_n^*(a\&b) = F$. As a particular case, $\lim\limits_{n\to\infty} F_n^*(a\&a) = F$.*

Proof. It is the same. We do not need the hypothesis that $\frac{S_{n+1}}{S_n} \to 1$. □

Mixture. *Mixture is a generalization of combining. If b and c are two sequences and $a = b\& c$, then b and c have the same weight in a. It is possible that the weights can be different. Then, we have a **mixture.***

Definition 1. *Let N be the set of positive integers. Let $B = \{i_1 < i_2 < i_3 < \cdots\}, C = \{j_1 < j_2 < j_3 < \cdots\} \subseteq N$ be such that $B \cup C = N$ and B and C are disjointed. Let b and c be two sequences of reals. Let $\beta_n = |B \cap \{1, \ldots, n\}|$ β, $\gamma_n = |C \cap \{1, \ldots, n\}|\gamma$. Suppose that $\frac{\beta_n}{n} \to p \in (0, 1)$ and $\frac{\gamma_n}{n} \to q \in (0, 1) = q = 1 - p$. Define $a_n = b_k$ if $n \in B$, $n = i_k$ and $a_n = c_k$ if $n \in C$, $n = j_k$. Then a is called a (B, C)—mixture of b and c. Let $S_n(b) = b_1 + \cdots + b_n$ and $S_n(c) = c_1 + \cdots + c_n$. Then it is obvious that $S_n(a) = S_{\beta_n}(b) + S_{\gamma_n}(c)$. If, moreover the limits $\lim \frac{S_{\beta_n}(b)}{S_n(a)} = \pi_b, \lim \frac{S_{\gamma_n}(c)}{S_n(a)} = \pi_c$ do exist, then we call the sequence "a good(B, C)—mixture of b and c". One may notice that, for instance, if $B = \{1, 3, 5, \ldots\}$, $C = \{2, 4, 6\}$, then $a = b\& c$. Thus, combining is indeed a particular case of mixture.*

The generalization is obvious. Split the positive integers in sets $B_1, \ldots, B_k$. Consider k sequences $\mathbf{b}_1, \ldots, \mathbf{b}_k$. Suppose that all the sets $B_j = \{i_{j.1}, \ i_{j.2}, \ldots\}$ are infinite. Let $\beta_{j,n} = |B_j \cap \{1, 2, \ldots, n\}|$ and $p_j = \lim_n \frac{\beta_{j,n}}{n}$. Define the mixture by $a_n = b_{j,k}$ if $n \in B_j$, $n = i_{j,k}$. Then $S_n(\mathbf{a}) = S_{\beta_{1,n}}(\mathbf{b}_1) + S_{\beta_{2,n}}(\mathbf{b}_2) + \ldots + S_{\beta_{k,n}}(\mathbf{b}_k)$.

Good Mixtures. *The mixture* $(B_1, \ldots, B_k)$ *of* $(\mathbf{b}_1, \ldots, \mathbf{b}_k)$ *is called a* **good mixture** *if all the limits* $\pi_j = \lim_n \frac{S_{\beta_{j,n}}(\mathbf{b}_j)}{S_n(\mathbf{a})}$ *do exist. Remark that always* $\pi_1 + \pi_2 + \cdots + \pi_k = 1$.

Proposition 13. *(i) Let* $\mathbf{a}$ *be a* (B, C) *–mixture of* $\mathbf{b}$ *and* $\mathbf{c}$*. Suppose that* $F_n(\mathbf{b}) \Rightarrow H_b$ *and* $F_n(\mathbf{c}) \Rightarrow H_c$*. Then* $F_n(\mathbf{a}) \Rightarrow pH_b + qH_c$*.*
(ii) Suppose that $\mathbf{a}$ *is a* **good** (B, C) *mixture of* $\mathbf{b}$ *and* $\mathbf{c}$ *and* $F_n^*(\mathbf{b}) \Rightarrow H_b^*$ *and* $F_n^*(\mathbf{c}) \Rightarrow H_c^*$*. Then* $F_n^*(\mathbf{a}) \Rightarrow pH_b^* \circ (h_{\pi_b/p})^{-1} + qH_c^* \circ (h_{\pi_c/q})^{-1}$ *where* $h_\lambda(x) = \lambda x$ *is the homothety. Here* $H \circ f^{-1}$ *means the image of* H *given by* f*, defined by* $H \circ f^{-1}(B) = \mu(f^{-1}(B))$ *for all Borel sets* B*.*
(iii) Generalization.
If B *is a* $(B_1, B_2, \ldots, B_k))$ *mixture of* $\mathbf{b}_1, \ldots, \mathbf{b}_k$ *and* $F_n(\mathbf{b}_j) \Rightarrow H_j$ *then* $F_n(\mathbf{a}) \Rightarrow \sum_{k=1}^n p_j H_j$*. If the mixture is good and* $F_n^*(\mathbf{b}_j) \Rightarrow H_{b_j}^*$ *then* $F_n^*(\mathbf{a}) \Rightarrow \sum_{j=1}^k p_j H_{b_j}^* \circ (h_{\pi_j/p_j})^{-1}$*.*

Proof.　(i) Let $f : [0, \infty) \to [0, \infty)$ be continuous and bounded. We know that as $n \to \infty$, $\frac{\sum_{k=1}^n f(b_k)}{n} \to \int f dH_b$ and $\frac{\sum_{k=1}^n f(c_k)}{n} \to \int f dH_c$. We also know that $\frac{\beta_n}{n} \to p \in (0, 1)$ and $\frac{\gamma_n}{n} \to q$. Then $\frac{\sum_{k=1}^n f(a_k)}{n} = \frac{\beta_n}{n} \frac{\sum_{k=1}^{\alpha_n} f(b_k)}{\beta_n} + \frac{\gamma_n}{n} \frac{\sum_{k=1}^{\beta_n} f(c_k)}{\gamma_n} \to p \int f dH_b + q \int f dH_c = \int f d(pH_b + qH_c)$.

(ii) Recall that $F_n^*(\mathbf{a}) = \frac{\sum_{k=1}^n \delta_{\frac{n}{S_n} a_k}}{n}$ We know by hypothesis that $\frac{\sum_{k=1}^n f(\frac{n b_k}{S_n(b)})}{n} \to \int f dH_b^*$ and $\frac{\sum_{k=1}^n f(\frac{n b_k}{S_n(c)})}{n} \to \int f dH_c^*$ for any uniform continuous and bounded $f : [0, \infty) \to [0, \infty)$. We want to prove that $\frac{\sum_{k=1}^n f(\frac{n a_k}{S_n(a)})}{n} \to p \int f \circ h_{\pi_b/p} dH_b^* + q \int f \circ h_{\pi_b/q} dH_c^*$ for any bounded continuous f. It is known (see, for instance, Reference [26] p. 371) that it is enough to consider only continuous functions with compact support, provided that the guessed limit is indeed a probability. This is our case. Thus, we write:

$$\frac{\sum_{k=1}^n f(\frac{n a_k}{S_n(a)})}{n} = \frac{\beta_n}{n} \frac{\sum_{k=1}^{\alpha_n} f(\frac{n S_{\beta_n}(b)}{S_n(a)\beta_n} \cdot \frac{\beta_n b_k}{S_{\beta_n}(b)})}{\beta_n} + \frac{\gamma_n}{n} \frac{\sum_{k=1}^{\beta_n} f((\frac{n S_{\gamma_n}(c)}{S_n(a)\beta_n} \cdot \frac{\gamma_n b_k}{S_{\gamma_n}(c)}))}{\gamma_n}$$

Notice that when $n \to \infty$, $\frac{n S_{\beta_n}(b)}{S_n(a)\beta_n} = \frac{n}{\beta_n} \frac{S_{\beta_n}(b)}{S_n(a)} \to \frac{\pi_b}{p}$ and $\frac{n S_{\gamma_n}(b)}{S_n(a)\gamma_n} \to \frac{\pi_c}{q}$. Taking into account the fact that f is uniformly continuous and its support is included in some interval $[0, M]$, the reader can check that:

$$\lim_n \frac{\sum_{k=1}^{\alpha_n} f(\frac{n S_{\beta_n}(b)}{S_n(a)\beta_n} \cdot \frac{\beta_n b_k}{S_{\beta_n}(b)})}{\beta_n} = \lim_n \frac{\sum_{k=1}^{\alpha_n} f(\frac{\pi_b}{p} \cdot \frac{\beta_n b_k}{S_{\beta_n}(b)})}{\beta_n}$$

and in the same way that the same way $\lim_n \frac{\sum_{k=1}^{\beta_n} f((\frac{n S_{\gamma_n}(c)}{S_n \gamma_n} \cdot \frac{\gamma_n b_k}{S_{\gamma_n}(c)}))}{\gamma_n} = \lim_n \frac{\sum_{k=1}^{\beta_n} f((\frac{\pi_c}{q} \cdot \frac{\gamma_n b_k}{S_{\gamma_n}(c)}))}{\gamma_n}$. But, according to our hypothesis $\lim_n \frac{\sum_{k=1}^{\alpha_n} f(\frac{\pi_b}{p} \cdot \frac{\beta_n b_k}{S_{\beta_n}(b)})}{\beta_n} = \int f(\frac{\pi_b}{p} x) dH_b^*(x)$ and $\lim_n \frac{\sum_{k=1}^{\beta_n} f((\frac{n S_{\gamma_n}(c)}{S_n \beta_n} \cdot \frac{\gamma_n b_k}{S_{\gamma_n}(c)}))}{\gamma_n} = \int (\frac{\pi_c}{q} x) dH_c^*(x)$.

The result is that

$$\lim_n \frac{\sum_{k=1}^n f(\frac{n a_k}{S_n})}{n} = p \int f(\frac{\pi_b}{p} x) dH_b^*(x) + q \int f(\frac{\pi_c}{q} x) dH_c^*(x) = \int f d(pH_b^* \circ (h_{\frac{\pi_b}{p}})^{-1} + qH_c^* \circ (h_{\frac{\pi_c}{q}})^{-1})$$

(iii) Same proof. □

Corollary 1. *Assume the same conditions as in Proposition 13. If $H_{b_j} = H$ for al j, then $F_n(\mathbf{a}) \Rightarrow H$ and if $H^*_{b_j} = H^*$ for all j, then $F^*_n(\mathbf{a}) \Rightarrow \sum_{j=1}^k p_j H^* \circ (h_{\pi_j/p_j})^{-1}$.*

Proof. Obvious. □

Lorenz Domination. *Let F and G be two probability distributions on $[0, \infty)$ and let L_F, L_G be their Lorenz curves. If $L_F \geq L_G$, we say that F is Lorenz dominated by G or, to quote Barry Arnold [28], that "G exhibits more inequality than F". Usually one denotes this domination by $F \leq_L G$. It is known that $F \leq_L G \Leftrightarrow \int u dF \leq \int u dG$ for all convex functions $u : [0, \infty) \to: [0, \infty)$ for which the integrals are finite. See Reference [29].*

It is interesting that a mixture of the type described in Corollary 1 always increases the inequality.

Proposition 14. *Let F be a probability distribution on: $[0, \infty)$. Let $(p_j)_{1 \leq j \leq k}$, and $(\pi_j)_{1 \leq j \leq k}$ be convex combinations with positive coefficients. Let $\lambda_j = \frac{\pi_j}{p_j}$. Then $F \leq_L \sum_{j=1}^k p_j F \circ (h_{\lambda_j})^{-1}$.*

Proof. Let u be convex. Then, $\int f d(\sum_{j=1}^k p_j F \circ (h_{\lambda_j})^{-1}) = \int \sum_{j=1}^k p_j f(\lambda_j x) dF(x) \geq \int f(\sum_{p_j}^k p_j \lambda_j x) dF(x) = \int f dF$. □

Example 6. $a_n = n^\alpha$.
The sequence $(\Lambda_n)_n$. If $\alpha > 0$ $a_n \to \infty$ and the sequence is increasing. $(F_n(\mathbf{a}))_n$ has no limit (unless we consider the abstract distribution δ). Thus $\Lambda_n(\mathbf{a})(p) \to \infty$. If $\alpha = 0$, $a_n = 1 \Rightarrow F_n = \delta_1$. If $\alpha < 0$,

$$a_n \to 0 \Rightarrow F_n = \delta_0 \text{ hence the result is } \lim_n \Lambda_n(p) = \begin{cases} 0 & if \quad \alpha < 0 \\ p & if \quad \alpha = 0 \\ \infty & if \quad \alpha > 0 \end{cases}.$$

The sequence $(L_n)_n$: If $\alpha > 0$ the sequence is increasing. The extreme points of the polygonal line L_n are $(\frac{k}{n}, \frac{s_k}{s_n}); 1 \leq k \leq n$. Here $s_k = 1^\alpha + 2^\alpha + \cdots + k^\alpha$. Notice that $\frac{k^{\alpha+1}+\alpha}{\alpha+1} < s_k < \frac{(k+1)^{\alpha+1}-1}{\alpha+1}$ hence $\frac{k^{\alpha+1}+\alpha}{(n+1)^{\alpha+1}-1} < \frac{s_k}{s_n} < \frac{(k+1)^{\alpha+1}-1}{n^{\alpha+1}+\alpha}$. Let $0 < p < 1$ and $(k_n)_n$ be a sequence such that $\frac{k_n}{n} \to p$. Then $\frac{(\frac{k_n}{n})^{\alpha+1}+\frac{\alpha}{n^{\alpha+1}}}{(1+\frac{1}{n})^{\alpha+1}-\frac{1}{n^{\alpha+1}}} < \frac{s_{k_n}}{s_n} < \frac{(\frac{k_n+1}{n})^{\alpha+1}-\frac{1}{n^{\alpha+1}}}{1^{\alpha+1}+\frac{\alpha}{n^{\alpha+1}}}$. Passing to limit as $n \to \infty$ we obtain $\frac{s_{k_n}}{s_n} \to p^{\alpha+1}$. Thus, $L_n(\frac{k_n}{n}) \to p^{\alpha+1}$. By Lemma 1, we get that $L_n(p) \to p^{\alpha+1}$. Similar considerations may be applied in the other cases; the result is that

$$\lim_n L_n(p) = \begin{cases} 0 & if \quad \alpha \leq -1 \\ 1 - (1-p)^{\alpha+1} & if \quad -1 < \alpha \leq 0 \\ p^{\alpha+1} & if \quad \alpha > 0 \end{cases}. \text{ Notice that if } \alpha \leq -1, \text{ the limit is a degenerate Lorenz curve.}$$

Example 7. $a = (1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, \ldots (2^4 \text{ times}), 0, 0, \ldots (2^5 \text{ times}), \ldots)$.
Let $N = 1 + 2 + \ldots + 2^{n-1} = 2^n - 1$. If $n = 2m$ is even, $O_n(a) = (0, 0 \ldots (\frac{2}{3}(2^n - 1) times), 1, 1, \ldots (\frac{1}{3}(2^n - 1) times))$ and if $n = 2m + 1$ is odd, $O_n(a) = (0, 0 \ldots (\frac{1}{3}(2^n - 1) times), 1, 1, \ldots (\frac{2}{3}(2^n - 1) times))$. In the first case, $F_n = \begin{pmatrix} 0 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}$ and in the second one it is $F_n = \begin{pmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$. For other types of n we obtain intermediate values. The pre-Lorenz curves move between $(p - \frac{2}{3})_+$ and $(p - \frac{1}{3})_+$, and the Lorenz ones move between $3(p - \frac{2}{3})_+$ and $2(p - \frac{1}{3})_+$. The limits do not exist.

Example 8. *A generic example of injective sequences a with the property that $F_n(a) \Rightarrow U(0, 1)$. Let $(A_k)_k$ be a sequence of positive numbers. Suppose that $A_k > 1$ is increasing and $\lim_k A_k = \infty$. Let $\sigma(0) = 0$, $\sigma(k) = [A_1] + [A_2] + \cdots + [A_k]$ for $k \geq 1$. ([A]means the integer part of A). Suppose, moreover, that*

$\frac{A_{k+1}}{A_k} \to 1$ *and that the ratio $\frac{A_m}{A_k}$ is **not a** rational number. Any positive integer n can be written uniquely as $n = \sigma(k) + j$ for some $k = k(n)$ and $j = j(n) \in \{0, \dots, [A_{k+1}] - 1\}$. For this n, define $a_n = \frac{j+1}{A_{k+1}}$. Then, all the elements of **a** are different, $F_n(a) \Rightarrow U(0,1)$ and $L_n(a)(p) \to p^2 = L_{U(0,1)}(p)$.*

Proof. The first assertion is obvious: if $a_m = a_n$ for some $m \neq n$, then $\frac{i}{A_k} = \frac{j}{A_m}$ for some i, j, k, m positive integers, contradicting the fact that $\frac{A_m}{A_k}$ is not a rational number. For the second one, we have to check that $\frac{f(a_1) + f(a_2) + \dots + f(a_n)}{n} \to \int_0^1 f(x)dx$ for any $f: \mathbb{R} \to \mathbb{R}$ bounded and continuous. (Proposition 7). Let us denote by U_A the distribution $\frac{\sum_{i=1}^{[A]} \delta_{\frac{i}{A}}}{[A]}$. Then $\int f dU_A = \frac{\sum_{i=1}^{[A]} f(\frac{i}{A})}{[A]}$. It is easy to see that $\lim\limits_{A \to \infty} \int f dU_A = \int_0^1 f(x)dx$ for any bounded continuous f (even more general, for every Riemann integrable f on $[0,1]$).

Let now $k \geq 1$, $j < [A_{k+1}]$ and $n = \sigma(k) + j$.

Then, $\frac{f(a_1) + f(a_2) + \dots + f(a_n)}{n} = \frac{[A_1] \int f dU_{A_1} + [A_2] \int f dU_{A_2} + \dots + [A_k] \int f dU_{A_k} + f(a_{\sigma(k)+1}) + \dots + f(a_{\sigma(k)+j})}{n}$. Suppose that $f \geq 0$. Then, as $\sigma(k) \leq n \leq \sigma(k+1)$, we have the evaluation

$$\frac{\sum_{i=1}^{k} [A_i] \int f dU_{A_i}}{\sigma(k+1)} \leq \frac{\sum_{j=1}^{n} f(a_j)}{n} \leq \frac{\sum_{i=1}^{k+1} [A_i] \int f dU_{A_i}}{\sigma(k)}$$

Apply Stolz-Cesaro lemma: as $k \to \infty$, the right-hand term has the same limit as $\frac{[A_{k+2}] \int f dU_{A_{k+2}}}{[A_{k+1}]}$, and the left hand one has the same limit as $\frac{[A_{k+1}] \int f dU_{A_{k+1}}}{[A_{k+2}]}$. But both limits are the same, namely, $\int_0^1 f(x)dx$. This ends the proof if $f \geq 0$. If not, write $f = f_+ - f_-$ and that is all. Thus, $F_n(\mathbf{a}) \Rightarrow U(0,1) \Leftrightarrow \Lambda_n(p) \to \frac{p^2}{2}$. To prove the last assertion, we have to check that the expectations $e(F_n(\mathbf{a}))$ converge to the expectation of $U(0,1)$, i.e., converge to $\frac{1}{2}$. The reasoning is the same as before. The expectation of U_A is $e_A := \frac{[A+1]}{2A}$ hence $\frac{\sum_{i=1}^{k}[A_i]e_{A_i}}{\sigma(k+1)} \leq \frac{\sum_{j=1}^{n} a_j}{n} \leq \frac{\sum_{i=1}^{k+1}[A_i]e_{A_i}}{\sigma(k)}$ converges to $\frac{1}{2}$ by the same argument of Stolz-Cesaro. □

4. The Empirical Lorenz and Pre-Lorenz Curves in the Broken Stick Models

Here, we suppose that the stick is not broken at random but using a deterministic sequence of points. What happens when n is great? These sticks are called **spacings** or **lags**. We shall be interested in the existence of limit for **spacings** of **a**.

Definition 2. *Let $\mathbf{a} = (a_n)_n$ be a sequence of reals. For any $n \geq 2$ we sort the first n terms of $\mathbf{a}$ as $a_{1:n} \leq a_{2:n} \leq \dots \leq a_{n,n}$. For $1 \leq j \leq n-1$ let $b_{j,n} = a_{(j+1):n} - a_{j:n}$. These are the **spacings** of $\mathbf{a}$ at level n.*

The vector $D_n(\mathbf{a}) = (b_{1,n}, \dots, b_{n-1,n})_{1 \leq j \leq n-1}$ is the vector of spacings after n trials.

The probability measure $G_n = G_n(a) = \sum_{j=1}^{n-1} \frac{\delta_{b_{j,n}}}{n-1}$ is the empirical distribution of spacings after n trials.

Remark 6. *Let $m_n(\mathbf{a}) = a_{1:n} = min(a_1, \dots, a_n)$ and $M_n(\mathbf{a}) = a_{n:n} = max(a_1, \dots, a_n)$. We may think that the interval $(m_n(\mathbf{a}), M_n(\mathbf{a}))$ is a stick which is broken by points $a_{j:n}$, $2 \leq j \leq n-1$. The $n-1$ segments obtained after this operation are the spacings. If two points coincide, the segment will be a degenerated one. From the definition, we see that the Lorenz curve of spacings is invariant with respect to scalings.*

Proposition 15. $L_{G_n(a)} = L_{G_n(\alpha a + \beta)}$ *for any $\alpha > 0$, $\beta \in \mathbb{R}$.*

Proof. The spacings which correspond to $(\alpha, a_n + \beta)_n$ are $\alpha b_{j,n} = \alpha(a_{j+1:n} - a_{j:n})$. $\square$

Proposition 16. *If the sequence $\boldsymbol{a}$ is bounded, then $G_n \Rightarrow \delta_0$ and $e(G_n) \to 0$ as $n \to \infty$.*

Proof. The second assertion is obvious: $e(G_n) = \sum_{j=1}^{n-1} \frac{b_{j,n}}{n-1} = \frac{M_n - m_n}{n-1} \to 0$ as $n \to \infty$. To check the first one, let $f \in C_b(\mathbb{R})$. Then $\int f dG_{n+1}(a) = \frac{f(b_1)+f(b_2)+...f(b_n)}{n}$ with $b_j = a_{(j+1)(n+1)} - a_{j:(n+1)}$. Suppose that $m \le a_n \le M$ for some $m < M$ and for every n. Then $\frac{b_1 + b_2 + ... b_n}{n} \le \frac{M-m}{n} \to 0$ as $n \to \infty$. Let $\varepsilon > 0$ be fixed and let $\delta = \delta(\varepsilon)$ such that $|x| < \delta \Rightarrow |f(x) - f(0)| < \varepsilon$. Notice that $k_n \left| \left\{ j \le n : b_j \ge \delta \right\} \right|$ is always smaller that $\frac{(M-m)}{\delta}$ hence $\frac{k_n}{n} \to 0$.

Let $A = \inf f$, $B = \sup f$. Then $\frac{f(b_1)+f(b_2)+...f(b_n)}{n} = \frac{\sum_{j:b_j<\delta} f(b_j) + \sum_{j:b_j\ge\delta} f(b_j)}{n} \le \frac{(n-k_n)(f(0)+\varepsilon)+Bk_n}{n}$. And, similarly, $\frac{f(b_1)+f(b_2)+...f(b_n)}{n} \ge \frac{(n-k_n)(f(0)-\varepsilon)-Ak_n}{n}$. Thus, $\frac{(n-k_n)(f(0)-\varepsilon)-Ak_n}{n} \le \int f dG_{n+1}(a) \le \frac{(n-k_n)(f(0)+\varepsilon)-Bk_n}{n}$. Passing to limit as $n \to \infty$ we get $f(0) - \varepsilon \le \liminf_{n\to\infty} \int f dG_{n+1}(a) \le \limsup_{n\to\infty} \int f dG_{n+1}(a) \le f(0) + \varepsilon$. It means that $G_n \Rightarrow \delta_0$ as claimed. $\square$

Remark 7. *That is why the distributions G_n are not very informative if we want to study the asymptotic behavior of the broken stick. We normalize them in such a way that their expectation be equal to 1, replacing them with $G_n^* = \frac{1}{n-1} \sum_{j=1}^{n-1} \delta_{\frac{(n-1)b_{j,n}}{M_n-m_n}}$. Now, $e(G_n^*) = 1$ for any n. The Lorenz curve remains the same, but we may hope that the weak limit of G_n^* does exist and it is not trivial. Moreover, $\Lambda_{G_n^*} = L_{G_n}$.*

Definition 3. *$G_n^*(\boldsymbol{a}))$ is called the normalized empirical distribution of the spacings generated by $\boldsymbol{a}$.*

Example 9. $\boldsymbol{a} = (1,0,1,0,1,0,....)$; as $O_n(\boldsymbol{a}) = (0,...,0,1,...,1)$ the spacings are $(0,...,0,1,0,...0)$ $(n-2$ times 0 and once a $1)$. Thus, $\Lambda_{G_n} \to 0$ as $n \to \infty$. As $G_n^* = \begin{pmatrix} 0 & n-1 \\ \frac{n-2}{n-1} & \frac{1}{n-1} \end{pmatrix} \Rightarrow \delta_0$, $\Lambda_{G_{n^*}} = L_{G_n} \to 0$ too as $n \to \infty$. Recall that $F_n(\boldsymbol{a}) \Rightarrow \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

Example 10. $\boldsymbol{a}$ *is a periodic sequence:* $\boldsymbol{a} = (x_1,...,x_k, x_1,...,x_k,...)$ *with* $x_1 \le x_2 \le \cdots \le x_k$;
As $O_n(\boldsymbol{a}) = (x_1,...,x_1,x_2,...,x_2,...,x_r,...,x_r,x_{r+1},...,x_{r+1},...,x_k,...,x_k)$ *we see that the spacings are* $(0,...,0,x_2 - x_1,0,...,0,x_3 - x_2,0,...,0,...,x_k - x_{k+1},0,...0)$ *($n-k$ times 0 and once the differences $x_{j+1} - x_j$. If $n = km$, then:*
$$G_n^* = \begin{pmatrix} 0 & (n-1)\frac{x_2-x_1}{x_k-x_1} & (n-1)\frac{x_3-x_2}{x_k-x_1} & \cdots & (n-1)\frac{x_k-x_{k-1}}{x_k-x_1} \\ \frac{n-k}{n-1} & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \end{pmatrix}. \text{ Thus, } \Lambda_{G_{n^*}} = L_{G_n} \to 0.$$

Example 11. $a_n = n^\alpha$. *If $\alpha > 1$, the spacings are $(2^\alpha - 1, 3^\alpha - 2^\alpha,...,n^\alpha - (n-1)^\alpha)$ hence,*
$$G_n^* = \begin{pmatrix} (n-1)\frac{2^\alpha-1}{n^\alpha-1} & (n-1)\frac{3^\alpha-2^\alpha}{n^\alpha-1} & (n-1)\frac{4^\alpha-3^\alpha}{n^\alpha-1} & \cdots & (n-1)\frac{n^\alpha-(n-1)^\alpha}{n^\alpha-1} \\ \frac{1}{n-1} & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \end{pmatrix}. \text{ Then, } \Lambda_{G_{n^*}} \text{ is the}$$
polygonal line given by the points $(\frac{k}{n-1}, \frac{k^\alpha-1}{n^\alpha-1})_k$. If $k_n = [np]$, these points converge to (p, p^α). According to Lemma 1, $\Lambda_{G_{n^*}}(p) = L_{G_n}(p) \to p^\alpha$ as $p \to \infty$.
If $\alpha = 1$, the spacings are $(1,1,...,1)$. $G_n = \delta_1$, $\Lambda_{G_n}(p) = L_{G_n}(p) = p$.
If $\alpha \in (0,1)$, the spacings $(2^\alpha - 1, 3^\alpha - 2^\alpha,...,n^\alpha - (n-1)^\alpha)$. As they are decreasing:
$$G_n^* = \begin{pmatrix} (n-1)\frac{n^\alpha-(n-1)^\alpha}{n^\alpha-1} & (n-1)\frac{(n-1)^\alpha-(n-2)^\alpha}{n^\alpha-1} & (n-1)\frac{(n-2)^\alpha-(n-3)^\alpha}{n^\alpha-1} & \cdots & (n-1)\frac{2^\alpha-1}{n^\alpha-1} \\ \frac{1}{n-1} & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \end{pmatrix}.$$
Then, $\Lambda_{G_{n^*}}$ is the polygonal line given by the points $(\frac{k}{n-1}, \frac{n^\alpha-(n-k)^\alpha}{n^\alpha-1})_k$. If $k_n = [np]$, these points converge to $(p, 1 - (1-p)^\alpha)$. According to Lemma 1, $\Lambda_{G_{n^*}}(p) = L_{G_n}(p) \to 1 - (1-p)^\alpha$ as $p \to \infty$.
If $\alpha = 0$, The spacings are $(0,0,...,0)$ and $G_n = \delta_0$. The Lorenz curve makes no sense.
If $\alpha < 0$ the sequence $\boldsymbol{a}$ is convergent, then $L_{G_n} \to 0$ according to Proposition 17 below.

$a = (1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, \ldots (2^4$ *times*$), 0, 0, \ldots (2^5$ *times*$), \ldots)$. *The spacings are the same as in Example 1.* $\Lambda_{G_n{}^*} = L_{G_n} \to 0$ *but* $F_n(a)$ *has no limit.*

A first question is: if we know that the sequence $(F_n^*(\mathbf{a}))_n$ has a limit, does it result that the sequence $G_n^*(\mathbf{a})$ has a limit? First, we should notice that there is no connection between the limit of F_n^* and the limit of G_n^*.

Here is an extension of Proposition 16: sometimes not only $G_n(\mathbf{a}) \Rightarrow \delta_0$, but $G_n^*(\mathbf{a}) \Rightarrow \delta_0$, too. In addition, it is also possible that $F_n(\mathbf{a})$ has no limit even if $G_n(\mathbf{a})$ has one.

Proposition 17. *Let* $\mathbf{a} = (a_n)_n$ *be bounded. Suppose that it has only a discrete set of limit points. Then* $G_n^*(a) \Rightarrow \delta_0$, $L_{G_n}(a) \to 0$. *It is possible that* $L_{F_n}(a)$ *and* $\Lambda_{F_n}(a)$ *have no limit.*

Proof. A discrete set is at most countable. Let A $= \{a_n : n \geq 1\}$. By hypothesis, the set of limit points B $=$ Closure(A)\A is at most countable hence Closure (A) is at most countable, too. Let a $=\inf(\mathbf{a})$ and b $=\sup(\mathbf{a})$.

The set D $=$ (a, b)\Closure(A) is open, hence it can be written as D $= \bigcup_{j \in J} I_j$, where $I_j = (\alpha_j, \beta_j)$ are open intervals with endpoints in the set A. Let $\mu_j = \beta_j - \alpha_j$ and $\boldsymbol{\mu} = (\mu_n)_n$. As A, B are null sets, $\sum_{j \in J} \mu_j = b - a$, hence the series is convergent. According to Proposition 8, $L_n(\boldsymbol{\mu}) \to 0$ as $n \to \infty$. If we relabel the sequence $\mathbf{a}$ in such a way that the sequence $(\mu_n)_n$ is non-increasing, we see that $L_{G_n}(\mathbf{a}) = L_n(\boldsymbol{\mu}) \to 0$. Otherwise written, $\lim_{n \to \infty} L_{G_n{}^*(a)}(p) = 0$. □

It is easy to find cases when L_{G_n} has a limit but L_{F_n} does not have one, it is enough to consider Example 11 from above: it has the same spacings as the one from Example 9, thus $L_{G_n} \to 0$ but L_{F_n} has no limit.

Remark 8. *A consequence of Proposition 17 is that if the sequence* a *is bounded, a condition in order that* G_n^* *have a limit different than* δ_0, *is that* $\lambda(Closure\{a_n : n \geq 1\}) > 0$. *The sequence should be dense in a set of positive Lebesgue measure. Here* λ *is the Lebesgue measure. For instance, if* $a_n = \sin n$, *then* $(Cl\{a_n : n \geq 1\}) = [-1, 1]$. *(For this very sequence we are not able to decide if* $G_n^*(a)$ *has a limit. The computer says it does not—if you can believe it).*
The converse holds too: it is possible that $F_n^*(a)$ *has a limit but* $G_n^*(a)$ *does not.*

Proposition 18. *Let* $\varepsilon = (\varepsilon_n)_n \in [0, 1]$ *and* $a = (1, 1 + \varepsilon_1, 2, 2 + \varepsilon_2, 3, 3 + \varepsilon_3, \ldots)$.
Then $F_n^* \Rightarrow \boldsymbol{Uniform}(0, 2)$, *but it is possible that* $G_n^*(a)$ *has no limit.*

Proof. Let $= (1, 2, 3, \ldots), c = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots)$.
Then $\Lambda_{F_n^*}(b)(p) = L_{F_n}(\mathbf{b})(p) \to p^2 = \Lambda_{\text{Uniform}(0,2)}(p)$.
According to proposition 10, the fact that $\Lambda_{F_n{}^*}(a) \to \Lambda_{\text{Uniform}(0,2)}$ implies that $F_n^* \Rightarrow \text{Uniform}(0, 2)$. The sequence $\mathbf{c} = \mathbf{b} + \boldsymbol{\varepsilon}$ has the same property: $F_n^*(\mathbf{c})$ converges to **Uniform** (0,2), too: this is a consequence of Proposition 10. Notice that $\mathbf{a} = \mathbf{b}\&\mathbf{c}$ $\mathbf{c}$. As $\mathbf{a}$ is non-decreasing too, Proposition 11 says that $F_n^*(\mathbf{a}) \Rightarrow \textbf{Uniform}(0, 2)$.

The spacings are easy to compute:
If $n = 2m$ is even then $D_n(\mathbf{a}) = (\varepsilon_1, 1 - \varepsilon_1, \varepsilon_2, 1 - \varepsilon_2, \ldots, \varepsilon_{m-1}, 1 - \varepsilon_{m-1}, \varepsilon_m)$ and
If $n = 2m + 1$, $D_n(\mathbf{a}) = (\varepsilon_1, 1 - \varepsilon_1, \varepsilon_2, 1 - \varepsilon_2, \ldots, \varepsilon_m, 1 - \varepsilon_m)$.
Their sum S_n is equally to $\left[\frac{n}{2}\right]$ if n is odd or to $\frac{n}{2} - 1 + \varepsilon_{\frac{n}{2}}$ if n is even.

If, for example, $n = 2m$ then $G_n^*(\mathbf{a}) = \frac{\sum_{k=1}^{2m} (\delta_{2\varepsilon_k} + \delta_{2-2\varepsilon_k})}{2m}$ and the limit of this sequence has no reasons to exist. □

Remark 9. *Sometimes it is not very difficult to compute $G_n^*(a)$. If $\varepsilon_n \in \{\alpha, \beta\}$, where $0 < \alpha < \beta <$*
$1 - \beta < 1 - \alpha < 1$ and $n = 2m + 1$, then $G_n^* = \begin{pmatrix} 2\alpha & 2\beta & 2(1-\beta) & 2(1-\alpha) \\ \frac{i_m}{m} & \frac{j_m}{m} & \frac{j_m}{m} & \frac{i_m}{m} \end{pmatrix}$ (if $n = 2m$ is slightly
different). Here, we denoted $i_m = \left| \{ j \le m : \varepsilon_j = \alpha \} \right|$ and $j_n = m - N(\alpha)$. Thus, G_n^* is a distribution of the form
$G_n^* = \begin{pmatrix} 2\alpha & 2\beta & 2(1-\beta) & 2(1-\alpha) \\ p_n & q_n & q_n & p_n \end{pmatrix}$, where $p_n + q_n = \frac{1}{2}$. If the frequencies $(p_n)_n$ have a limit as $n \to \infty$,
then G_n^* has a limit, too, therefore $L_{G_n^*}$ has a limit too. But that is by no means the rule.

Remark 10. *If $G = \begin{pmatrix} 2\alpha & 2\beta & 2(1-\beta) & 2(1-\alpha) \\ p & q & q & p \end{pmatrix}$ with $a < b < \frac{1}{2}$ and $p + q = \frac{1}{2}$, then L_G is the*
polygonal line given by the points $(0,0)$, $(p, 2ap)$, $(\frac{1}{2}, 2ap + 2bq)$, $(1 - p, 2(ap + q))$, $(1,1)$. Its Gini coefficient
(i.e., $Gini(G) = 1 - 2\int_0^1 L_G(x)dx$) is equal to $\frac{1}{2} - b - 4p(b - a) - 4p^2(b - a)$ lies between $\frac{1}{2} - b$ and $\frac{1}{2} - a$.
We think that the distributions G_n given by Remark 9 have always the property that $Gini(G_n^*) \le \frac{1}{2}$. For instance,
if $\varepsilon_n = \alpha < \frac{1}{2}$ is constant, then $Gini(G_n^*) = \frac{1-2\alpha}{2}$.

Combining

Proposition 19. *Suppose that the sequence a has the property that $\lim G_n^*(a) = G$ and $\lim L_n(a) = L$.*
Then, $\lim G_n^*(a \& a) = \frac{1}{2}(\delta_0 + G)$ and $\lim L_n(a \& a)(p) = L((2p - 1)_+)$.

Proof. Let $n = 2m + 1$.

Then $O_n(a \& a) = (a_{1:m+1}, a_{1:m+1}, a_{2:m+1}, a_{2:m+1}, \dots, a_{m+1:m+1}, a_{m:m}, a_{m:m+1})$ and the spacings
are $D_n(a \& a) = (0, d_1, 0, d_2, \dots, 0, d_m)$, where $d_j = a_{j+1, m+1} - a_{j.m+1}$. Then $G_n^*(a) = \dfrac{m\delta_0 + \sum_{j=1}^m \delta_{\frac{2md_j}{\Delta_m}}}{2m}$
where $\Delta_m = a_{m+1, m+1} - a_{1, m+1}$. □

Mixture. *We can say something about a special case of mixture.*

Proposition 20. *Let $T_0 < T_1 < \cdots < T_k$ and $I_j = [T_{j-1}, T_j)$. Let a be a sequence valued into $[T_0, T_k) = \cup_{j=1}^k I_j$*
having the property that $\inf a = T_0$ and $\sup a = T_k$. For $n \ge 2$ let $n_j = \left| \{ i \le n : a_i \in I_j \} \right|$. Suppose that the
limits $p_j = \lim\limits_{n \to \infty} \frac{n_j}{n}$ do exist and $p_j > 0$ for all $1 \le j \le k$.
Let $A_j = \{ n : a_n \in I_j \} = \{ i_{j,1} < i_{j,2} \dots \}$ and let the sequences b_j be defined by $b_{j,n} = a_{i_{j,n}}$. Suppose that
$G_n^*(b_j) \Rightarrow H_j$, where H_j are some probability distributions on $[0, \infty)$ $[0, \infty)$ and $e(H_j) = 1$.
For any fixed $n \ge 2$, let $d_{i,n} = a_{i+1:n+1} - a_{i:n+1}$ be the n spacings corresponding to $(a_1, \dots, a_{n+1})$.
Suppose, finally, that there are no gaps: $\inf \{ a_n : n \in A_j \} = T_{j-1}$ and $\sup \{ a_n : n \in A_j \} = T_{j-1}$. Then
(i) $G_n^*(a) \Rightarrow \sum_{j=1}^n p_j H_j \circ (h_{\pi_j/p_j})^{-1}$ with $\pi_j = \frac{T_j - T_{j-1}}{T_k - T_0}$.
(ii) The Lorenz curves $L_n(a)$ converge to the Lorenz curve of $\sum_{j=1}^n p_j H_j \circ (h_{\pi_j/p_j})^{-1}$.
(iii) If all the probability distributions H_j coincide with the same H then the Lorenz curves $L_n(a)$ converge to the
Lorenz curve of $\sum_{j=1}^n p_j H \circ (h_{\pi_j/p_j})^{-1}$ which is always below L_H. In words, the mixture of sequences of the
same type always exhibits more inequality.

Proof. (i) Let $f : [0, \infty) \to R$ be continuous and with compact support. We have to compute the limit
of $\dfrac{\sum_{i=1}^n f(\frac{nd_{i,n}}{a_{n+1:n+1} - a_{1:n+1}})}{n}$. As $\inf a = T_0$, $\sup a = T_k$, this limit is the same with the limit of $\dfrac{\sum_{i=1}^n f(\frac{nd_{i,n}}{T})}{n}$ for
$T = T_k - T_0$— if it exists.

There are two kinds of spacings: spacings $d_{i,n}$ for which both $a_{i+1,n+1}$ and $a_{i,n+1}$ are in the same
interval I_j. Call them **spacings of type j** and denote by $B_{j,n}$ their set. If n is great enough, $\left| B_{j,n} \right| = n_j - 1$.

The other kind of spacings are those $d_{i,n}$ for which $a_{i,n+1} \in I_j$ but $a_{i+1,n+1} \in I_{j+1}$ for some
$1 \le j \le k - 1$. The number of spacings of the second kind is finite—actually, if n is great enough,

it must be equal to $k - 1$, due to the fact of our hypothesis that all p_j are positive. That is why we can neglect them in computing the limit. Then:

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} f(\frac{nd_{i,n}}{T})}{n} = \lim_{n\to\infty} \frac{\sum_{j=1}^{k} \sum_{i\in B_{j,n}} f(\frac{nd_{i,n}}{T})}{n} = \sum_{j=1}^{k} \lim_{n\to\infty} \frac{n_j}{n} \frac{\sum_{i\in B_{j,n}} f(\frac{n(T_j-T_{j-1})}{n_jT} \cdot \frac{n_j d_{i,n}}{T_j-T_{j-1}})}{n_j}. \text{ Notice that } \frac{n_j}{n} \to p_j,$$

$$\frac{n(T_j-T_{j-1})}{n_jT} \to \frac{\pi_j}{p_j} \text{ and } \frac{\sum_{i\in B_{j,n}} f(\frac{n_j d_{i,n}}{T_j-T_{j-1}})}{n_j} \to \int f dH_j$$

As f is uniformly continuous:

$$\frac{\sum_{i\in B_{j,n}} f(\frac{n(T_j-T_{j-1})}{n_jT} \cdot \frac{n_j d_{i,n}}{T_j-T_{j-1}})}{n_j} \to \int f(\frac{\pi_j}{p_j}x)dH_j(x).$$

.

Therefore, $\lim_{n\to\infty} \frac{\sum_{i=1}^{n} (\frac{nd_{t,n}}{T})}{n} = \sum_{j=1}^{k} \int p_j f(\frac{\pi_j}{p_j}x)dH_j(x)$ as claimed.

(ii) Notice that $e(\sum_{j=1}^{n} p_j H_j \circ (h_{\pi_j/p_j})^{-1}) = 1$ (since $e(H_j) = 1$ for all j), but in that case $\Lambda = L$ and apply Proposition 3.

(iii) If f is convex, then:

$$\int f d(\sum_{j=1}^{k} p_j H \circ h_{\pi_j/p_j})^{-1} = \int \sum_{j=1}^{k} p_j f(\frac{\pi_j}{p_j}x)dH(x) \geq \int f(\sum_{j=1}^{k} p_j \frac{\pi_j}{p_j}x)dH(x) = \int f dH.$$

□

Remark 11. *Here we supposed that there are no gaps: the spacings tend to 0. What happens if we accept some gaps?*

Proposition 21. Finite number of gaps. *Let $\boldsymbol{a}$ be a bounded sequence, $\alpha = \inf a$, $\beta = \sup a$. Suppose that $Cl\{a_n : n \geq 1\} = I_1 \cup I_2 \cup \ldots \cup I_k$ where $I_j = [\alpha_j \beta_j]$ with $\alpha = \alpha_1 < \beta_1 < \alpha_2 < \beta_2 < \cdots < \alpha_k < \beta_k = \beta$.*
Let $n_j = |i \leq n : a_i \in I_j|$. Suppose that the limits $p_j = \lim_{n\to\infty} \frac{n_j}{n}$ do exist and $p_j > 0$ for all $1 \leq j \leq k$. Let $A_j = \{n : a_n \in I_j\} = \{i_{j,1} < i_{j,2} < \cdots\}$ and let the sequences $\boldsymbol{b}_j$ be defined by $b_{j,n} = a_{i_{j,n}}$. Suppose that $G_n^*(\boldsymbol{b}_j) \Rightarrow H_j$, where H_j are some probability distributions on $[0, \infty)$. For any fixed $n \geq 2$, let $d_{i,n} = a_{i+1:n+1} - a_{i:n+1}$ be the n spacings corresponding to $(a_1, \ldots, a_{n+1})$.
Let $\pi_j = \frac{\beta_j - \alpha_j}{\beta - \alpha}$ Then $G_n^*(\boldsymbol{a}) \Rightarrow \sum_{j=1}^{n} p_j H_j \circ (h_{\pi_j/p_j})^{-1}$.

The Lorenz curves $L_n(\boldsymbol{a})$ converge to $k\, L_{\sum_{j=1}^{k} p_j H_j \circ (h_{\pi_j/p_j})^{-1}}$ where $k = \sum_{j=1}^{k} \frac{\beta_j - \alpha_j}{\beta - \alpha} < 1$. The limit is a degenerate Lorenz curve.

Proof. The same proof, the difference is that now $\sum_{j=1}^{k} \pi_j < 1$. Notice that $e(\sum_{j=1}^{n} p_j H_j \circ (h_{\pi_j/p_j})^{-1}) = \sum_{j=1}^{k} \pi_j$. Apply Proposition 5. □

Conclusion: Gaps or no gaps, the Lorenz curve of a mixture of sequences of the same type is below the mother Lorenz curve.

5. The Empirical Lorenz Curves in the Stochastic Broken Stick Models

Now suppose that a stick of length 1 is broken at random into n smaller sticks by some i.i.d. random variables $X_1, \ldots, X_{n+1}$. Empirical evidence shows that the empiric distribution function of normed spacings always has a limit.

We shall prove that in some situations.

Precisely, let $X = (X_n)_n \geq 0$ be i.i.d. random variables having a common distribution F. Suppose that they are integrable and $0 < \mu = EX_j < \infty$. Let $a = \text{essinf } X_j$, $b = \text{esssup } X_j$, and $D_{1,n} = X_{1:n} - a$, $D_{2,n} = X_{2:n} - X_{1:n}, \ldots, D_{n,n} = X_{n:n} - X_{n-1:n}$. If $b < \infty$, we could also add the $(n+1)$-th spacing, namely, $D_{n+1,n} = b - X_{n:n}$.

The Glivenko theorem says that the empirical distribution $F_n(X) = \sum_{j=1}^{n} \frac{\delta_{X_j}}{n}$ converges weakly to F almost surely. More than that, the distribution functions $F_n x(x) = \frac{|\{j \leq n : X_j \leq x\}|}{n}$ converge almost surely uniformly to $F(x)$.

We prove that even the normed empirical functions defined by $F_n^*(X) = \frac{\sum_{k=1}^{n} \delta_{\frac{n}{S_n} X_k}}{n}$ converge almost surely to the distribution of $\frac{X_k}{\mu}$:

Lemma 2. *Let F_n and F be distribution functions and $(\alpha_n)_n$ be a sequence of non-negative numbers. Suppose that $F_n \Rightarrow F$ and $\alpha_n \to \alpha$. Then $F_n(\alpha_n x) \to F(\alpha x)$ at all continuity points of F.*

Proof. Let $\varepsilon > 0$ and n_ε such that $n > n_\varepsilon \Rightarrow \alpha - \varepsilon < \alpha_n < \alpha + \varepsilon$. Then $F_n((\alpha - \varepsilon)x) \leq F_n(\alpha_n x) \leq F_n((\alpha + \varepsilon)x)$. Suppose that F is continuous at αx, $(\alpha - \varepsilon)x$ and $(\alpha + \varepsilon)x$ (this is no problem, the discontinuity set of points for F is at most countable). Then $F((\alpha - \varepsilon)x) \leq \liminf_n F_n(\alpha_n x) \leq \liminf_n F_n(\alpha_n x) \leq F_n((\alpha + \varepsilon)x)$. As ε is arbitrary small, the result immediately follows. □

Remark 12. *Another version of Glivenko's theorem is in terms of **order statistics**. Note that $F_n(x) = \frac{k}{n}$ if and only if $X_{k:n} \leq x < X_{k+1:n}$. Let $p \in (0,1)$ and let $(k_n)_n$ be a sequence of positive integers with the property that $\frac{k_n}{n} \to p$ (for instance $k_n = [np]$). Then $F_n(X_{k_n:n}) = \frac{k_n}{n} \to p \Rightarrow X_{k_n:n} \to F^{-1}(p)$ almost surely for all p for which $F^{-1}(x)$ is unique. As a special case, $X_{[np]} \to F^{-1}(p)$ a.s. for all $p \in (0,1)$ if F is bijective.*

Proposition 22. *Let $X = (X_n)_n$ be i.i.d. non-negative F-distributed random variables. Let $F_n^*(X) = \frac{\sum_{k=1}^{n} \delta_{\frac{n}{S_n} X_k}}{n}$. Then $F_n^*(X) \Rightarrow F \circ (h_{\frac{1}{\mu}})^{-1}$ almost surely, where $\mu = EX_j > 0$. In other words, they converge to the distribution of $\frac{X_j}{\mu}$. Moreover, the empirical Lorenz curves $L_n(X)$ converge almost surely to the Lorenz curve of F, denoted by L_F.*

Proof. According to the strong law of large numbers, the sequence $\alpha_n = \frac{S_n}{n}$ converges almost surely to μ. Let $F_n(x)$ be the empirical distribution function of **X**. The distribution function of $\frac{\sum_{k=1}^{n} \delta_{\frac{n}{S_n} X_k}}{n}$ is $F_n(\alpha_n x)$. We know that $F_n(x) \to F(x)$ almost surely at all continuity points of F and $\alpha_n \to \mu$ almost surely. Then, by Lemma 2 we see that $F_n(\alpha_n x) \to F(\mu x)$ almost surely at all discontinuity points of $G(x) = F(\mu x)$. The distributions $F_n^*(X)$ were constructed in such a way that $e(F_n^*(X)) = 1$. By Proposition 4, their Lorenz curves converge to the Lorenz curve of F, *denoted by L_F*. □

The problem of convergence of G_n^* to some limit distribution is more delicate. We believe that the almost sure weak limit does always exist, but we are not able to prove it. Besides, we can prove some weaker convergence results.

Definition 4. *Let F be a distribution on real line. We say that F has the property (D) if the weak limit of $G_n^*(X) = \frac{\sum_{k=2}^{n} \delta_{\frac{n}{X_n^* - X_{*,n}} D_{k,n}}}{n-1}$ does exist in probability for any sequence X of i.i.d. F-distributed random variables. It means that the empirical distribution function $G_n^*(X)(x) = \frac{|\{k \leq n : \frac{n}{X_n^* - X_{*,n}} D_{k,n} \leq x\}|}{n-1}$—which are random variables—converge in probability to some distribution function $G(x)$ at all continuity points of G. In that case we say that F is the mother distribution and that G is the born distribution. Call the quantities $\frac{D_{k,n}}{X_n^* - X_{*,n}}$ the normed spacings.*

Proposition 23. *If F has the property* (**D**) *and* $h(x) = ax + b$, *then* $F \circ h^{-1}$ *has the same property. Moreover, F and* $F \circ h^{-1}$ *bear the same G.*

Proof. Obvious. If $\Upsilon_j = \alpha X_j + \beta$, then the normed spacings are the same for **X** and Υ. □

The first positive result is that all discrete probability distributions have the property (**D**). All of them bear δ_0.

Proposition 24. *Suppose* $F = \sum_{k=1}^{\infty} p_k \delta_{a_k}$ *is discrete. Then* $G_n^*(\mathbf{X}) \Rightarrow \delta_0$ *in probability.*
Thus, any discrete distribution on the real line has property (D): *it gives birth to the Dirac distribution* δ_0.

Proof. Let $\mathbf{X} = (X_n)_{n \geq 1}$ be a sequence of *F-distributed i.i.d.* random variables, let $\varepsilon > 0$ be arbitrary small, and let k such that $p_1 + \cdots + p_k > 1 - \varepsilon$. For any $n \geq 1$ let $n_j = \left| \left\{ j \leq n : X_n = a_j \right\} \right|$. By the law of large numbers, $\frac{n_j}{n} \to p_j$. At least $n_1 - 1 + n_2 - 1 + \cdots + n_k - 1$ spacings will be equal to 0. Thus, $G_n^*(\{0\}) \geq \frac{\sum_{j=1}^{k} n_j - k}{n} \to \sum_{j=1}^{k} p_j > 1 - \varepsilon$. It means that $\lim_{n \to \infty} G_n^*(\{0\}) = 1 \lim_{n \to \infty} G_n^*(\{0\}) = 1$. That obviously implies $G_n^*(\mathbf{X}) \Rightarrow \delta_0$ in probability. □

Remark 13. *All the random variables produced by computer using pseudo random numbers are discrete. Using the computer to guess a limit can be misleading many times.*
The second positive result is that F = Uniform (0,1) has the property (**D**)—*it gives birth to the standard exponential distribution, denoted by Exp* (1). *According to Proposition 25 below this will hold for any F = Uniform*(a, b). *We shall use the following result:*

Lemma 3. *Let* $\mathbf{X} = (X_n)_n$ *and* $\Upsilon = (Y_n)_n$ *be two sequences of random variables. Suppose that* X_n *and* Y_n *have the same distributions. If* $X_n \to c$ *(a.s.) for some constant* $c \in \mathbb{R}$, *then* $Y_n \to c$ *(in probability).*

Proof. $P(|Y_n - c| > \varepsilon) = P(|X_n - c| > \varepsilon) \to 0$. □

Remark 14. *Unfortunately, it is not true that if we know that* $X_n \to 0$ *(a.s.) and* $Y_n \sim X_n \Rightarrow Y_n \to 0$ *a.s. A simple counterexample on* $(0, 1)$ *is if:*
$$\Upsilon = (1_{(0,\frac{1}{2})}, 1_{(\frac{1}{2},1)}, 1_{(0,\frac{1}{3})}, 1_{(\frac{1}{3},\frac{2}{3})}, 1_{(\frac{2}{3},1)}, 1_{(0,\frac{1}{4})}, 1_{(\frac{1}{4},\frac{2}{4})}, 1_{(\frac{2}{4},\frac{3}{4})}, 1_{(\frac{3}{4},1)}, 1_{(0,\frac{1}{5})}, \ldots) \text{ and}$$
$$\mathbf{X} = (1_{(0,\frac{1}{2})}, 1_{(0,\frac{1}{2})}, 1_{(0,\frac{1}{3})}, 1_{(0,\frac{1}{3})}, 1_{(0,\frac{1}{3})}, 1_{(0,\frac{1}{4})}, 1_{(0,\frac{1}{4})}, 1_{(0,\frac{1}{4})}, 1_{(0,\frac{1}{4})}, 1_{(0,\frac{1}{5})}, \ldots).$$
And it is not true if the limit is not a constant.

Proposition 25. *If F = Uniform*(a, b), *then F has the property* (**D**). *It gives birth to the standard exponential distribution Exp* (1).
As a consequence, the sequence of Lorenz Curves $L_n(\mathbf{X})$ *converge in probability to* $L(t) = t + (1 - t) \ln(1 - t)$.

Proof. Let (Ω, K, P) be a probability space. According to Proposition 23, we can put $a = 0$, $b = 1$. Let $\mathbf{X} = (X_n)_{n \geq 1}$ be a sequence of i.i.d. random variables uniformly distributed, let $O_n(\mathbf{X}) = (X_{1:n}, X_{2:n}, \ldots, X_{n:n})$, and $\mathbf{D}_{n+1}(\mathbf{X}) = (D_{j,n+1})_{0 \leq j \leq n}$ be the $n + 1$ spacings $D_{j,n+1} = X_{j+1,n} - X_{j,n}$, with the convention that $X_{n+1,n} = 1$.

It is well known that $\mathbf{D}_n(\mathbf{X}) \sim$Uniform (P_{n+1}), where $P_n = \left\{ x \in \mathbb{R}_+^n : x_1 + \cdots + x_n = 1 \right\}$ is the unitary simplex—see any book of order statistics, for instance, References [10–12,17].

Let, on the other hand, $\xi = (\xi_n)_{n \geq 1}$ be a sequence of i.i.d. random variable distributed as Exp (1). Let $S_n = \xi_1 + \cdots + \xi_n$. It is also known (see References [10–12,17]) that $(\frac{\xi_1}{S_n}, \frac{\xi_2}{S_n}, \ldots, \frac{\xi_n}{S_n})$ is also uniformly distributed on P_n. Thus, $\mathbf{D}_n(\mathbf{X}) =^D (\frac{\xi_1}{S_n}, \frac{\xi_2}{S_n}, \ldots, \frac{\xi_n}{S_n})$ for any $n \geq 2$. Therefore, $n\mathbf{D}_n(\mathbf{X})$ has the same distribution as $(\frac{n\xi_1}{S_n}, \frac{n\xi_2}{S_n}, \ldots, \frac{n\xi_n}{S_n})$. Consequently, the Lorenz curves of $G_n^*(\mathbf{X}) = \frac{\sum_{j=1}^n \delta_{nD_{j,n}}}{n}$ and

$F_n^*(\xi)\dfrac{\sum_{j=1}^{n}\delta_{n\xi_j}}{S_n}$ have the same distribution. According to Glivenko's theorem, the empirical distribution

functions $F_{n+1}(\xi)(x)=\dfrac{\sum_{j=1}^{n+1}1_{\{\xi_j\le x\}}}{n+1}$ converge almost surely uniformly to the distribution function of the Exp distribution, namely, to $(x)=1-e^{-x}$. Otherwise written, the empirical distributions

$F_n=\dfrac{\sum_{j=1}^{n}\delta_{\xi_j}}{n+1}$ converge almost surely to $F\mathrm{Exp}\,(1)$ Let $\Omega_1=\left\{\omega\in\Omega:\dfrac{\sum_{j=1}^{n}\delta_{\xi_j(\omega)}}{n}\Rightarrow F\right\}$ and $\Omega_2=$

$\left\{\omega\in\Omega:\dfrac{\sum_{j=1}^{n}\xi_j(\omega)}{n}\to 1\right\}$. According to Glivenko's theorem and to the strong law of large numbers,

$P(\Omega_1)=P(\Omega_2)=1$. Thus, if $\omega\in\Omega_1\cap\Omega_2$ the sequence of empirical distributions $F_n^*(\xi(\omega))=$

$\dfrac{\sum_{j=1}^{n}\delta_{\frac{n\xi_j(\omega)}{\xi_1(\omega)+\xi_2(\omega)+\dots+\xi_n(\omega)}}}{n}$ weakly converge to Exp (1). All these distribution have the same expectation,

namely, 1, hence for almost all $\omega\in\Omega$, the limit of the Lorenz curves $L_{F_n^*(\xi(\omega))}(t)$ is the same, namely, the Lorenz curve of Exp (1) which is $L(t)=t+(1-t)\ln(1-t)$. The curves $L_{G_n^*(\mathbf{X})}(t)$ have the same distribution with $L_{F_n^*(\xi(\omega))}(t)$; therefore, they have the same limit in probability—namely, $L(t)$. $\square$

Remark 15. *The result has been known for some time and can be proved in several ways. For instance, Stephens [14,15] gives an alternate proof of the fact that the limit of the Lorenz curves in the broken stick model is $L(t)=t+(1-t)\ln(1-t)$.*

Proposition 26. *The exponential distribution has the property (**D**), too: it gives birth to δ_0.*

Proof. Let $\mathbf{X}=(X_n)_{n\ge 1}$ be iid standard exponentially distributed random variables. The density is $\pi(x)=e^{-x}1_{(0,\infty)}(x)$. The reader can check that the density of the order statistics $O_n(\mathbf{X})$ is $p(x_1,\dots,x_n)=n!e^{-(x_1+x_2+\dots+x_n)}1_{\{x_1<x_2<\dots<x_n\}}$ and the density of $D_n(\mathbf{X})$ is $q(x_1,\dots,x_n)=n!e^{-(nx_1+(n-1)x_2+\dots+x_n)}1_{[0,\infty)^n}(\mathbf{x})$.

Thus, the distribution of $D_n(\mathbf{X})$ is $\prod_{j=1}^{n}\mathrm{Exp}(n+1-j)$, otherwise written $D_n(\mathbf{X})\overset{D}{=}(\frac{\xi_1}{n},\frac{\xi_2}{n-1},\dots,\frac{\xi_n}{1})$, where $(\xi_j)_{1\le j\le n}$ are standard independent exponentially distributed.

The normalized empirical distribution is $G_n^*(\mathbf{X})=\dfrac{\sum_{j=1}^{n}\delta_{\frac{nD_{j,n}}{X_n^*}}}{n}$, (since $\sum_{j=1}^{n}D_{j,n}=X_n^*=$

$\max(X_1,\dots,X_n)$) and its empirical tail is $\overline{G}_n(x)=\dfrac{\sum_{j=1}^{n}1_{\{\frac{nD_{j,n}}{X_n^*}>x\}}}{n}$. We shall prove that $\overline{G}_n(x)\to 0$ in probability for any $x>0$ and that will mean that $G_n^*(\mathbf{X})\Rightarrow\delta_0$.

The idea is to prove that $\overline{G}_n(x)\to 0$ in L^1. As these random variables are positive, the only thing we have to do is to check that $E\overline{G}_n(x)\to 0$. But $E\overline{G}_n(x)=\frac{1}{n}\sum_{j=1}^{n}P(\frac{nD_{j,n}}{X_n^*}>x)=\frac{1}{n}\sum_{j=1}^{n}P(\frac{n\xi_j}{jX_n^*}>x)=\frac{1}{n}\sum_{j=1}^{n}P(\xi_j>\frac{jx}{n}X_n^*)$.

As the random variables ξ_j are i.i.d. and the function $\varphi(x)=\max(x_1,\dots,x_n)$ is symmetrical, the probabilities $P(\xi_j>\frac{jx}{n}X_n^*)$ coincide with $P(\xi_1>\frac{jx}{n}X_n^*)$.

Let $\psi(t)=P(\xi_1>tX_n^*)$. Then we have $E\overline{G}_n(x)=\frac{1}{n}\sum_{j=1}^{n}P(\xi_1>\frac{jx}{n}X_n^*)=\frac{1}{n}\sum_{j=1}^{n}\psi(\frac{jx}{n})$.

Notice that ψ is decreasing therefore the sum $\frac{1}{n}\sum_{j=1}^{n}\psi(\frac{jx}{n})$ is smaller than $\int_0^1\psi(tx)dt$. It follows that $x>0\Rightarrow E\overline{G}_n(x)\le\int_0^1 P(\xi_1>txX_n^*)dt$. But the sequence $(X_n^*)_n$ is non-decreasing and tends to infinity almost surely. It follows that $P(\xi_1>txX_n^*)\downarrow 0\Rightarrow E\overline{G}_n(x)\to 0$ as $n\to\infty\Rightarrow\overline{G}_n(x)\to 0$ in $L^1\Rightarrow\overline{G}_n(x)\to 0$ in probability. $\square$

Mixtures. *The stochastic equivalents of Propositions 20 and 21.*

Proposition 27. *$f:[a,b)\to(0,\infty)$ be a positive density and let F be its corresponding distribution. Next, let $a=T_0<T_1<\dots<T_k=b$ be a division of $[a,b)$ and $I_j=\left[T_{j-1},T_j\right),1\le j\le k$. Let F_j be the probability*

distributions on I_j given by the densities $f_j = \frac{f1_{[T_{j-1}, T_j)}}{F(T_j) - F(T_{j-1})}$. *Suppose that all the probability distributions F_j have the property (**D**) and that F_j bears H_j. Then F has the property (**D**).*

(i) *$G_n^*(\boldsymbol{X})$ converges weakly in probability to $\sum_{j=1}^n p_j H_j \circ (h_{\pi_j/p_j})^{-1}$ with $\pi_j = \frac{\lambda_j}{b-a}$, $\lambda_j = T_j - T_{j-1}$ and $p_i = F(T_j) - F(T_{j-1})$.*

(ii) *The Lorenz curves $L_n(\boldsymbol{X})$ converge to the Lorenz curve of $\sum_{j=1}^n p_j H_j \circ (h_{\pi_j/p_j})^{-1}$. If all the probability distributions H_j coincide with the same H, then the Lorenz curves $L_n(\boldsymbol{a})$ converge to the Lorenz curve of $\sum_{j=1}^n p_j H \circ (h_{\pi_j/p_j})^{-1}$ which is always below L_H. In other words, the mixture of sequences of the same type always exhibits more inequality.*

Proof. Let $\boldsymbol{X} = (X_n)_n$ be a sequence of i.i.d F-distributed random variables. Let $A_j = \{n : X_n \in I_j\} = \{i_{j,1} < i_{j,2} < i_{j,3}; \ldots\}$ and $n_j = |A_j \cap \{1, 2, \ldots, n\}|$. From the law of large numbers, $\frac{n_j}{n} \to p_j F(T_j) - F(T_{j-1}) > 0$ almost surely. Define $\Upsilon_{j,m} = X_n$ if $X_n \in I_j$ and $n = i_{j,m}$.

The sequences $\boldsymbol{\Upsilon}_j = (\Upsilon_{j,m})_m$ are again i.i.d but their distribution is F_j. We know that F_j bears H_j. Meaning that $\frac{\sum_{i=1}^{n_j} g(\frac{n_j}{\lambda_j} \Delta_{i,n_j}(j))}{n_j} \xrightarrow{P} \int g dH_j$ for any g continuous with compact support. Here $\Delta_{i,n}(j)$ are the spacings of $\boldsymbol{\Upsilon}_j$ and $n_j \to \infty$. We want to find the limit in probability for $\frac{\sum_{i=1}^n g(\frac{n}{b-a} D_{i,n})}{n}$, where $D_{i,n} = X_{i+1:n} - X_{i:n}$, $1 \leq i \leq n-1$ are the $n-1$ spacings of $\boldsymbol{X}$.

Let $B_{j,n} = \{i \leq n : i, i+1 \in A_{j,n}\}$. Then $|B_{j,n}| = n_j - 1$ and $i \in B_{j,n} \Rightarrow D_{i,n} = \Delta_{m,n_j}(j)$ for some $m < n_j$. All but k spacings are of one of these k types. Then, as $\frac{1}{b-a} = \frac{\pi_j}{\lambda_j}$, we can write $lim \frac{\sum_{i=1}^n g(\frac{n}{b-a} D_{i,n})}{n} = lim \sum_{j=1}^k \frac{n_j}{n} \frac{\sum_{i=1}^{n_j-1} g(\frac{n\pi_j}{n_j} \frac{n_j \Delta_{i,n_j}(j)}{\lambda_j})}{n_j}$.

When $n \to \infty$, $n_j \to \infty$ almost surely and $\frac{n\pi_j}{n_j} \to \frac{\pi_j}{p_j}$ almost surely. Then, as g is uniformly continuous and has a compact support, $\frac{\sum_{i=1}^{n_j-1} g(\frac{n\pi_j}{n_j} \frac{n_j \Delta_{i,n_j}(j)}{\lambda_j})}{n_j} \xrightarrow{P} \int g(\frac{\pi_j}{p_j} x) dH_j(x)$. The remaining claims have the same proof as in Proposition 21. $\square$

Corollary 2. *Let $a = T_0 < T_1 < \cdots < T_k = b$, If F has the density $f = \sum_{j=1}^k \alpha_j 1_{[T_j, T_{j+1})}$ then F has the property (D). It bears the distribution $H = \sum_{j=1}^k \alpha_j \lambda_j exp((b-a)\alpha_j)$ which dominate always Exp (1) in Lorenz order. Here, $\lambda_j = T_j - T_{j-1}$. Therefore, the Lorenz curves $L_n(\boldsymbol{X})$ converge in probability to the Lorenz curve of H.*

Proof. Let $\lambda_j = T_j - T_{j-1}$. If f is a density, then $\sum_{j=0}^k \alpha_j \lambda_j$ must be equal to 1. Suppose that $\alpha_j > 0$ for all $1 \leq j \leq n$. Apply Proposition 27. The densities are $f_1 = \sum_{j=0}^k \frac{1}{\lambda_j} 1_{[T_j, T_{j+1})}$ and $p_j = \alpha_j \lambda_j$. All H_j are the same, by Proposition 25, namely, $H = \exp(1)$. $\square$

Next, $\frac{\pi_j}{p_j} = \frac{1}{(b-a)\alpha_j}$ hence $\sum_{j=1}^n p_j H \circ (h_{\pi_j/p_j})^{-1} = \sum_{j=1}^n \alpha_j \lambda_j \exp((b-a)\alpha_j)$.

Remark 16. *What happens if we have gaps? Then we apply Proposition 21. The weak limit remains the same, but the Lorenz curves converge to a defective one. Precisely:*

Corollary 3. *Suppose the mother distribution has the density $F = \sum_{j=1}^k \alpha_j 1_{[s_j, t_j)}$, where $s_1 < t_1 \leq s_2 < t_2 \leq \cdots \leq s_k < t_k$. Let $\lambda_j = t_j - s_j$, $L = t_k - s_1$, $q = \frac{\lambda_1 + \ldots + \lambda_k}{L}$. Then, F has the property (D)and bears the same $H = \sum_{j=1}^k \alpha_j \lambda_j exp((b-a)\alpha_j)$, but the Lorenz curves $L_n(\boldsymbol{X})$ converge in probability to qL_H which can be defective if $t_j < s_{j-1}$ for some j.*

Remark 17. *One may see that the order of the intervals I_j and of the possible gaps $[t_j, s_{j-1})$ is not important. What matters is their lengths. For instance the probability distributions F_1, F_2 having the densities $f_1 =$*

$\alpha 1_{[1,3)} + \beta 1_{[4,5)}$ and $f_2 = \beta 1_{[0,1)} + \alpha 1_{[2,4)}$ with $\alpha, \beta > 0$ and $2\alpha + \beta > 0$ bear the same probability distribution $H = 2\alpha exp(4\alpha) + \beta exp(4\beta)$. However, the empirical Lorenz curves $L_n(X)$ converge to the defective curve $L^* = \frac{3}{4}L_H$. Here $\lambda_1 = 2$, $\lambda_2 = 1$, $\pi_1 = \frac{1}{2}$, $\pi_2 = \frac{1}{4}$.

The Broken Rectangle. *Suppose that the rectangle $R = [a, a + L] \times [b, b + l)$ is broken by a sequence of points $z = (x_n, y_n)_{n \geq 1}$ which belong to R as follows: for any $n \geq 1$ we sort increasingly the components $(x_j)_{1 \leq j \leq n}$ and $(y_j)_{1 \leq j \leq n}$ obtaining $(x_{j:n})_{1 \leq j \leq n}$ and $(y_{j:n})_{1 \leq j \leq n}$. Then add the endpoints $x_{0:n} = a, x_{n+1:n} = a + L$, $y_{0:n} = b$, $y_{n+1:n} = b + l$ and construct the rectangle $D_{i.j} = [x_{i:n}, x_{i+1:n}) \times [y_{j:n}, y_{j+1:n})$ for $0 \leq i, j \leq n$. Its area is $\sigma_{i,j} = (x_{i+1:n} - x_{i:n})(y_{i+1:n} - y_{i:n})$. Then consider the normalized empirical distributions $G_n^*(z) = \frac{\sum_{0 \leq i,j \leq n} \delta_{\frac{(n+1)^2}{Ll}\sigma_{i,j}}}{(n+1)^2}$ (all of them have the expectation equal to 1).*
Do they have a weak limit?

Proposition 28. *Let $X = (X_n)_{n \geq 1}$ and $Y = (Y_n)_{n \geq 1}$ be two independent sequences of i.i.d random variables. Let F_X be the distribution of X_n and F_Y be the distribution of Y_n. and $Z = (X_n, Y_n)_{\boldsymbol{n} \geq \boldsymbol{1}}$ 1. Suppose both F_X and F_Y have the property (**D**), namely F_X gives birth to H_X and F_Y gives birth to H_Y. Let ξ_X and ξ_Y be two random variables distributed as H_X and H_Y and H be the distribution of $\xi_X \xi_Y$. Let $D_n(X) = (X_{i+1:n+1} - X_{i:n+1})_{1 \leq i \leq n}(U_1, \ldots U_n)$, $D_n(Y) = (Y_{j+1:n+1} - Y_{j:n+1})_{1 \leq j \leq n}(V_1, \ldots V_n)$. Thus, the assumption is that $G_n^*(X) \Rightarrow H_X$ X and $G_n^*(Y) \Rightarrow H_Y$ in probability. Finally, let $G_n^*(Z) = \frac{\sum_{1 \leq i,j \leq n} \delta_{\frac{n^2 U_i V_j}{(M_{n+1}(X) - m_{n+1}(X))(M_{n+1}(Y) - m_{n+1}(Y))}}}{n^2}$ be the empirical distribution of the normalized spacings produced by Z. Here $M_n(X) = \max(X_1, \ldots, X_n)$, $m_n(X) = \min(X_1, \ldots, X_n)$ a.s.o. Then*

(i) $G_n(Z) \Rightarrow H$ *in probability and the Lorenz curves of $(U_i V_j)_{1 \leq i,j \leq n}$ tend to $L(p) = E\xi_X E\xi_Y L_{\xi_X \xi_Y}(p)$.*

(ii)   *If $E\xi_X = E\xi_Y = 1$, the limit is the Lorenz curve $L(p) = L_{\xi_X \xi_Y}(p)$.*

(iii) *As a particular case, if $F_X = F_Y = Uniform(0, 1)$, then ξ_X and ξ_Y are exponentially distributed hence H has the tail $\overline{H}(x) = \int_0^\infty e^{-(y + \frac{x}{y})} dy$. The empirical Lorenz curves converge in probability to the Lorenz curve of H.*

(iv) $L_{\xi_X \xi_Y} \leq min(L_{H_X}, L_{H_Y})$ *hence always $L \leq min(L_{H_X}, L_{H_Y})$.*

      In the particular case (iii), $L(p) \leq p + (1-p)\ln(1-p)$.

Proof.   (i) We know that $G_n^*(X) \Rightarrow H_X$ and $G_n^*(Y) \Rightarrow H_Y$ in probability.   It means that $\int g dG_n^*(X) \xrightarrow{P} \int g dH_X$ and $\int g dG_n^*(Y) \xrightarrow{P} \int g dH_Y$ for any bounded continuous g. We claim that $G_n^*(X) \otimes G_n^*(Y) \Rightarrow H_X \otimes H_Y$ in probability. The easy way is to use the equivalence "$X_n \xrightarrow{P} X$ if and only if from any subsequence $(X_{k_n})_n$ one may extract a sub-sub-sequence $(X_{k_{\sigma(n)}})_n$ such that $X_{k_{\sigma(n)}} \xrightarrow{a.s.} X$".

In our case, from any subsequence $(k_1 < k_2 < \cdots)$ we can extract a sub-sub-sequence $G_{k_{\sigma(n)}}^*(X) \Rightarrow H_X, G_{k_{\sigma(n)}}^*(Y) \Rightarrow H_Y$ almost surely. In that case, $G_{k_{\sigma(n)}}^*(X) \otimes G_{k_{\sigma(n)}}^*(Y) \Rightarrow H_X \otimes H_Y$ almost surely, because one knows that if F_n, G_n, F, G are probability distributions, then $F_n \Rightarrow F$, $G_n \Rightarrow G$ implies $F_n \otimes G_n \Rightarrow F \otimes G$ (prove follows immediately using characteristic functions).

In our case:
$$G_n(X) = \frac{\sum_{1 \leq i \leq n} \delta_{\frac{nU_i}{(M_{n+1}(X) - m_{n+1}(X))}}}{n},$$

$$G_n(Y) \frac{\sum_{1 \leq j \leq n} \delta_{\frac{nV_i}{(M_{n+1}(Y) - m_{n+1}(Y))}}}{n}, G_n(X) \otimes G_n(Y) = \frac{\sum_{1 \leq i,j \leq n} \delta_{(\frac{nU_i}{M_{n+1}(X) - m_{n+1}(X)}, \frac{nV_j}{M_{n+1}(Y) - m_{n+1}(Y)})}}{n^2}. \text{ It follows that}$$

$\frac{\sum_{1 \leq i,j \leq n} g(\frac{nU_i}{M_{n+1}(X) - m_{n+1}(X)}, \frac{nV_j}{M_{n+1}(Y) - m_{n+1}(Y)})}{n^2} \xrightarrow{P} \int g dH_X \otimes H_Y$ for any $g : \mathcal{R}^2 \to \mathcal{R}$ bounded continuous.

As a particular case, we set $g(x, y) = h(xy)$ with $h : \mathcal{R} \to \mathcal{R}$ bounded continuous and the result is that $\frac{\sum_{1 \leq i,j \leq n} h(\frac{n^2 U_i V_j}{(M_{n+1}(X) - m_{n+1}(X))(M_{n+1}(Y) - m_{n+1}(Y))})}{n^2} \xrightarrow{P} \int h(xy) dH_X \otimes H_Y(x, y) = Eh(\xi_X \xi_Y)$.

The remaining (i) is a consequence of Proposition 5.

(iii) If $F_X = F_\Upsilon = $ Uniform $(0,1)$, then $H_X = H_\Upsilon = Exp\ (1)$. If ξ_X, ξ_Υ are independent and exponentially distributed, then $P(\xi_X \xi_\Upsilon > x) = E(P(\xi_\Upsilon > x\ /\ \xi_X | \xi_X)) = E(e^{-\frac{x}{\xi_X}}) = \int_0^\infty e^{-(y+\frac{x}{y})} dy$.

(iv) If X, Y are two non-negative independent random variables, $EX = a > 0$, $E\Upsilon = b > 0$, then $L_{X\Upsilon} \leq \min(L_X, L_\Upsilon)$. Indeed, it is enough to consider $a = b = 1$. If $u : [0, \infty) \to \mathcal{R}$ is convex, then $E_u(X\Upsilon) = E(E_u(X\Upsilon)|\Upsilon) \geq E_u(E(X\Upsilon)|\Upsilon)$ (Jensen's inequality) $= E_u(\Upsilon E(X|\Upsilon)) = E_u(\Upsilon EX) = E_u(\Upsilon)$.

Also, in the same way, $E_u(X\Upsilon) \geq E_u(X)$. □

Remark 18. *Otherwise written, the stick broken at random exhibits always smaller inequality than the rectangle broken at random.*

Generalization is easy. If, instead of a square we have the hypercube $[0,1]^k$, then the empirical Lorenz curves tend, in the uniform case, to the Lorenz curve of $\xi_1, \ldots, \xi_k$ where $(\xi_j)_{1 \leq j \leq k}$ are i.i.d. standard exponentially distributed random variables.

Remark 19. *Gini index. It is interesting that we can compute the Gini index for the broken rectangle, even if we do not know its distribution analytically. It is known (see [3,27,28]) that the Gini index can be computed as $Gini(L_X) = 1 - \frac{Emin(X,Y)}{EX}$, where Y is an independent copy of X.*

If $X \sim Exp\ (1)$, then $\min(X,Y) \sim Exp\ (2)$, hence it is obvious that in the uniform broken stick model the limit of empirical Gini is $\frac{1}{2}$.

In the broken square model, the limit distribution of spacings has the tail $\overline{H}(x) = \int_0^\infty e^{-(y+\frac{x}{y})} dy$. If X has the distribution H and Y is an independent copy of X, then the tail of $\min(X,Y)$ is $\overline{H}^2(x)$. As $E\xi = 1$, we get $Gini(L_X) = 1 - Emin(X,Y) = 1 - \int_0^\infty (\int_0^\infty e^{-(y+\frac{x}{y})} dy)^2 dx = 1 - \int_0^\infty \int_0^\infty \int_0^\infty e^{-(y+\frac{x}{y})} e^{-(z+\frac{x}{z})} dz\,dy\,dx = 1 - \int_0^\infty \int_0^\infty \int_0^\infty e^{-(y+z+x(\frac{1}{y}+\frac{1}{z}))} dx\,dz\,dy = 1 - \int_0^\infty \int_0^\infty \frac{yze^{-(y+z)}}{y+z} dz\,dy = 1 - \int_0^\infty y \int_y^\infty \frac{(u-y)e^{-u}}{u} du\,dy = 1 - \int_0^\infty (y(e^{-y} - y \int_y^\infty \frac{e^{-u}}{u} du)) dy = \int_0^\infty \int_y^\infty y^2 \frac{e^{-u}}{u} du\,dy = \int_0^\infty (\int_0^u y^2 \frac{e^{-u}}{u} dy) du = \int_0^\infty \frac{e^{-u}}{u} (\int_0^u y^2 dy) du = \int_0^\infty \frac{u^2 e^{-u}}{3} du = \frac{2}{3}.$

Therefore, in the broken square model the limit for the Gini index is $\frac{2}{3}$. It is empirically confirmed by the computer.

6. Conclusions and Open Problems

We have proved that in many cases, the empirical distributions of the normed spacings generated in the process of breaking the stick, denoted by $G_n^*(\mathbf{X})$, have a limit and that the limit of their associated Lorenz curves $L_n(\mathbf{X})$ lies below the Lorenz curve of the exponential distribution. Thus, they are more inegalitarian that the exponential distribution. As a byproduct, the limit of their Gini coefficient is always greater than $\frac{1}{2}$. As a principle, no random division of a whole to many individuals according to the principle of broken stick—be it land, wealth or income—can produce a Gini index less than 0.5.

We were not able to fix some problems such as:

1. Prove or disprove that if $\mathbf{X} = (X_n)$ are i.i.d., then $G_n^*(\mathbf{X})$ always has a weak limit ALMOST SURELY;
2. Prove or disprove that if $\text{essinf } X_n = -\infty$ or $\text{ess sup } X_n = \infty$, then $G_n^*(\mathbf{X}) \Rightarrow \delta_0$ ALMOST SURELY.
3. Let $a_n = \{n\alpha\}$, where $\alpha \notin \mathbf{Q}$. Is it possible to find a formula for weak $\lim_{n\to\infty} G_n^*((a_n)_n)$?
4. Let $\mathbf{a} = (0, 1, 0, \frac{1}{2}, 1, 0, \frac{1}{3}, \frac{2}{3}, 1, 0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1, 0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1, 0, \frac{1}{6}, \frac{2}{6}, \ldots\ldots)$. Find $\lim_n F_n^*(\mathbf{a}), \lim_n G_n^*(\mathbf{a})$;
5. The same question for a $= (0, 1, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \frac{5}{6}, \frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \frac{1}{9}, \frac{2}{9}, \ldots)$;
6. Rearrangement. Empirical evidence suggests that if a mother distribution F has density f and g is another density having the property that $\lambda(\{f > t\}) = \lambda(\{g > t\})$ for every $t > 0$, then G is also a mother distribution and bears the same H as F. For instance, $f_1(x) = 2x1_{(0,1)}(x)$, $f_2(x) = 2(1-x)1_{(0,1)}(x)$ and $f_3(x) = 4\min(x, 1-x)$, bear the same H. Or, in terms of random variables: Let $U = (U_n)_n$ and $V = (V_n)_n$ be two independent sequences of i.i.d. uniform $(0,1)$,

distributed random variables. Then, $W_1 = \max(U, V)$, $W_2 = \min(U, V)$, $W_3 = \frac{1}{2}(U + V)$ have the same limit distribution of the spacings. Prove or disprove that.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

We prove some (maybe well-known) facts about convex functions, quantiles, and for other technicalities for which we were not able to find appropriate references and also for the sake of self-containment.

Convexity

Lemma A1. *Let $f : (a, b) \to \mathbb{R}$ be convex. Then*

(i)   *its right derivative $f'_r(x) = \lim\limits_{\delta \downarrow 0} \frac{f(x+\delta)-f(x)}{\delta}$ is non-decreasing and right continuous.*

(ii)   *its left derivative $f'_l(x) = \lim\limits_{\delta \downarrow 0} \frac{f(x)-f(x-\delta)}{\delta}$ is non-decreasing and left continuous.*

(iii)   *moreover, $f'_r(x) = f'_l(x + 0)$ for any $x \in (a, b)$.*

(iv)   *the two derivatives coincide with the possible exception of a (at most) countable set.*

Proof. The convexity of f is equivalent with the chord inequality. For any $a < x_1 < x_2 < x_3 < b$ one has $\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_1)}{x_3-x_1} \leq \frac{f(x_3)-f(x_2)}{x_3-x_2}$ with the obvious consequence:

$$a < x_1 < x_2 < b \Rightarrow f'_r(x_1) \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq f'_l(x_2) \tag{A1}$$

These two derivatives are non-decreasing, therefore all the lateral limits $f'_r(x + 0)$, $f'_r(x - 0)$, $f'_l(x + 0)$ and $f'_l(x - 0)$ do exist. Keeping x_1 fixed in $(A1)$ and letting $x_2 \downarrow x_1$ one gets $f'_r(x_1) \leq f'_l(x + 0)$; if one keeps x_2 fixed and lets $x_1 \uparrow x_2$, one gets $f'_r(x_2 - 0) \leq f'_l(x_2)$. Thus,

$$f'_r(x - 0) \leq f'_l(x) \leq f'_t(x) \leq f'_l(x + 0) \text{ for any } a < x < b \tag{A2}$$

In order to prove the continuities, we remark that for any $a < x_1 < x_2 < b$, there exist y_1, y_2 such that:

$$a < x_1 < y_1 < y_2 < x_2 < b \text{ and } \frac{f(y_1) - f(x_1)}{y_1 - x_1} \leq \frac{f(y_2) - f(y_1)}{y_2 - y_1} \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \tag{A3}$$

Indeed, if $\frac{f(y)-f(x_1)}{y-x_1} = \frac{f(x_2)-f(x_1)}{x_2-x_1}$ for some $\in (x_1, x_2)$, then f is affine on (x_1, x_2), hence any y_1, y_2 satisfy (A3). If not, choose $y_2 \in (x_1, x_2)$ arbitrarily such that $\frac{f(y_2)-f(x_1)}{y_2-x_1} < \frac{f(x_2)-f(x_1)}{x_2-x_1}$. As f is continuous, there must exist $y_1 \in (x_1, y_2)$ such that $\frac{f(y_2)-f(y_1)}{y_2-y_1} < \frac{f(x_2)-f(x_1)}{x_2-x_1}$. Then, by chord inequality:

$$f'_r(x_1) \leq \frac{f(y_1) - f(x_1)}{y_1 - x_1} \leq \frac{f(y_2) - f(y_1)}{y_2 - y_1} \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq f'_l(x_2) \tag{A4}$$

If in (A4) we let $y_2 \downarrow x_1$, then $y_1 \downarrow x_1$ too, and we get $f'_r(x_1) \leq f'_r(x_1 + 0) \leq \frac{f(x_2)-f(x_1)}{x_2-x_1} \leq f'_l(x_2)$. If we let now $x_2 \downarrow x_1$, we find that $f'_r(x_1 + 0) \leq f'_r(x_1) \leq f'_l(x_1 + 0)$. Thus, f'_r is right continuous and, moreover, the inequality $f'_r(x + 0) \leq f'_l(x + 0)$ may occur only if $f'_r(x) = f'_l(x + 0)$. This equality obviously implies the fact that f'_r is right continuous: if we replace any function with its limits at right (provided that they do exist) we get a right-continuous function. The fact that f'_l is left continuous can

be proved in a similar way. Finally, both f'_l and f'_r are non- decreasing, hence they are continuous with the possible exception of some at most countable set. The equality $f'_r(x) = f'_l(x+0)$ implies that f'_r and f'_l coincide at continuity points. □

Lemma A2. *Let* $f : [a,b) \rightarrow \mathbb{R}$ *be convex and continuous at a. Let* $\lambda = f'_r$ *be the right derivative. Then* λ *is Riemann integrable on any interval* $[a.x]$ *with* $a < x < b$ *and, moreover:*

$$f(x) = f(a) + \int_a^x \lambda(t)dt \tag{A5}$$

Proof. Let $a < x_1 < x_2 < \cdots < x_n = x$ be a division of $[a,x]$. As $f(x) - f(a) = \sum_{k=1}^n \frac{f(x_k)-f(x_{k-1})}{x_k-x_{k-1}}(x_k - x_{k-1})$ we have, according to (A1) the evaluation:

$$\sum_{k=1}^n f'_r(x_{k-1})(x_k - x_{k-1}) \leq f(x) - f(a) \leq \sum_{k=1}^n f'_l(x_k)(x_k - x_{k-1})$$

Both f'_r and f'_l are bounded on $[a,x]$ and continuous with the possible exception of a countable set. Therefore they are Riemann integrable and, passing to the limit when the norm of the division tends to 0 we get $\int_a^x f'_r(t)dt \leq f(x) - f(a) \leq \int_a^x f'_l(t)dt$. The claimed equality results from the fact that $f'_r \geq f'_l$. □

The convex functions have a property that it is not true in general: the convergence of functions implies the almost everywhere convergence of the derivatives. Precisely

Lemma A3. *Let* $f_n, f : [a,b] \rightarrow \mathbb{R}$ *be convex and continuous. Suppose that* $f_n \rightarrow f$. *Then*

(i) $(f'_n)_r(x) \rightarrow f'_r(x)$ *at any point x for which* $f'_l(x) = f'_r(x)$.

(ii)   *If, moreover,* f_n *and* f *are non-decreasing, then the convergence of* $(f'_n)_r$ *to* f'_r *holds in* $L^1(a,b)$, *too.*

Proof.

(i)   Let $x \in (a, b)$. Let $\delta > 0$ be such that $a < x - \delta < x + \delta < b$, then, according to (A1) $\frac{f(x)-f(x-\delta)}{\delta} \leq f'_l(x) \leq f'_r(x) \leq \frac{f(x+\delta)-f(x)}{\delta}$. It follows that: $\frac{f_n(x)-f_n(x-\delta)}{\delta} - \frac{f(x+\delta)-f(x)}{\delta} \leq (f_n)'_r(x) - f'_r(x) \leq \frac{f_n(x+\delta)-f_n(x)}{\delta} - \frac{f(x)-f(x-\delta)}{\delta}$. Letting $n \rightarrow \infty$, we get: $\frac{f(x)-f(x-\delta)}{\delta} - \frac{f(x+\delta)-f(x)}{\delta} \leq \liminf_n((fn)'_r(x) - f'_r(x)) \leq \limsup_n\left((fn)'_r(x) - f'_r(x)\right) \leq \frac{f(x+\delta)-f(x)}{\delta} - \frac{f(x)-f(x-\delta)}{\delta}$ which further implies:

$$\limsup_n \left|(fn)'_r(x) - f'_r(x)\right| \leq \frac{f(x + \delta) - f(x)}{\delta} - \frac{f(x) - f(x - \delta)}{\delta} \tag{A6}$$

Now let (A6) assume that $\delta \downarrow 0$. It follows that

$$\limsup_n \left|(fn)'_r(x) - f'_r(x)\right| \leq f'_r(x) - f'_l(x) \tag{A7}$$

(ii)   If we take into account Lemma A1 (iv), it follows that $(fn)'_r(x) \rightarrow f'_r(x)$ for $x \in [a,b]$ with the possible exception of a countable set. As any countable set is a null set for the Lebesgue measure, we can write that $(fn)'_r \rightarrow f'_r$ almost everywhere (a.e.). From (A5) we have that

$\int_a^b (f_n')_r(x)dx = f_n(b) - f_n(a)$, $\int_a^b (f')_r(x)dx = f(b) - f(a)$. To prove the convergence in L^1, notice that $|x| = x_+ = \frac{x+|x|}{2} \Rightarrow |x| = 2x_+ - x$. We write

$$\int_a^b \left| (f')_r(x) - (f_n')_r(x) \right| dx = 2 \int_a^b \left((f')_r(x) - (f_n')_r(x) \right)_+ dx - \int_a^b \left((f')_r(x) - (f_n')_r(x) \right) dx$$

The second integral is equal to $(f(b) - f_n(b)) - (f(b) - f_n(b))$ tends to 0 as $n \to \infty$ as $f_n \to f$. The second one tends to 0 too because of the domination $\left((f')_r - (f_n')_r \right)_+ \leq (f')_r \in L^1(a, b)$. Thus, we prove the convergence in L^1. □

Now we know that the right derivative of a convex function $\Lambda : [0, 1) \to \mathcal{R}_+$ with the property that $\Lambda(0) = 0$ is a mapping which is non-negative, right continuous, non-decreasing and, moreover, Λ has the representation.

$$\Lambda(x) = \int_0^x \lambda(t)dt \text{ for all } x \in (0, 1) \tag{A8}$$

We claim that λ is the superior quantile of some distribution function $F : [0, \infty) \to [0, 1]$.

Quantiles.

Definition 5. *Let $F : \mathbb{R} \to \mathbb{R}$ be non-decreasing and not a constant. Let $m = F(-\infty)$, $M = F(\infty)$. A quantile of F is any function $Q : (m, M) \to \mathcal{R}$ with the property that $Q(p) = x \Rightarrow F(x - 0) \leq p \leq F(x + 0)$.*
The mappings $Q^+(p) = \sup F^{-1}((-\infty, p])$, $(Q^-(p) = \inf F^{-1}([p, \infty)))$ are called the superior (inferior) quantiles of F. Sometimes one denotes the superior quantile by F^{-1}.

Lemma A4. *(i). If F is one to one, then the quantile is unique. Moreover, if Q_1 and Q_2 are two quantiles of F, then they coincide with the possible exception of a countable set.*
*(ii). $Q(p - 0) < Q(p + 0)$ if and only if the level set $I_F(p) = \{x : F(x) = p\}$ is a non-degenerate interval. Otherwise written, the continuity points of Q are those points for which either the level set $I_F(p)$ is void or it is a singleton. As a consequence, the unique quantile of a one to one non-decreasing function F is **continuous**.*
(iii). The superior quantile is always right—continuous. Moreover, the only quantile which is right continuous is the superior one.
(iv). If $\lambda : (m, M) \to \mathbb{R}$ is non-decreasing and right continuous, then there exists $F : \mathbb{R} \to [m, M]$ which is non-decreasing and right—continuous such that λ is the superior quantile of F. As a particular case, if $\lambda : (0, 1) \to \mathbb{R}$ is non-decreasing and right continuous then there exists a distribution function F such that $\lambda = F^{-1}$.

Proof. (i) Suppose that Q_1 and Q_2 are two quantiles of F and that there exists some p such that $Q_1(p) = x_1$ and $Q_2(p) = x_2$. Thus $F(x_1 - 0) \leq p \leq F(x_1 + 0)$, $F(x_2 - 0) \leq p \leq F(x_2 + 0)$. Suppose $x_1 < x_2$. It follows that $F(x_1 - 0) \leq p \leq F(x_1 + 0) \leq F(x_2 - 0) \leq p \leq F(x_2 + 0) \Rightarrow F(x_1 + 0) = F(x_2 - 0) = p \Rightarrow (x_1, x_2) \subseteq I_F(p)$. Thus $Q_1 = Q_2$. But the set $\{p \in (0, 1) : I_F(p)$ contains an open interval$\}$ is always at most countable.

(ii) Suppose that $I_F(p)$ contains the interval (x_1, x_2) and $x_1 < x_2$. We claim that $Q(p - 0) \leq x_1$ and $Q(p + 0) \geq x_2$. Indeed, let ε, δ be small enough such that $x_1 < x_1 + \varepsilon < x_2 - \delta < x_2$. Let $p' < p$ and $x' = Q(p')$. Then $x' \leq x_1 + \varepsilon$. Indeed, we know that $F(x' - 0) \leq p' < F(x' + 0)$. If $x' > x_1 + \varepsilon$, then $F(x' - 0) \geq F(x_1 + \varepsilon - 0)$ and that is absurd $p' \geq F(x' - 0) \geq p$. Thus $p' < p \Rightarrow Q(p') \leq x_1 + \varepsilon$ for any $\varepsilon > 0$. It means that $Q(p' - 0) \leq x_1$. In the same way one proves chat $p' > p \Rightarrow Q(p') \geq x_2 - \delta$ for any $\delta > 0$ hence $(p + 0) \geq x_2$.

On the other hand, let us suppose that $p \in (m, M)$ has the property that $I(p)$ has at most one point. Then there exists $x \in \mathcal{R}$ such that $x' < x \Rightarrow F(x' + 0) < p$ and $x'' > x \Rightarrow F(x'' - 0) > p$. We claim that $Q(p - 0) = Q(p + 0) = x$. Indeed, suppose that $Q(p - 0) = x' < x$ or that $Q(p + 0) = x'' > x$. Suppose that we are in the first situation. Let $p_n < p$ be a sequence such that $p_n \to p$.

Let $x_n = Q(p_n) \Rightarrow F(x_n - 0) \le p_n \le F(x_n + 0)$. As $x_n \le x'$ we see that $F(x_n - 0) \le p_n \le F(x' - 0) < p$ for all n. This is absurd, since $p_n \to p$. We are forced to accept that $x' = x$. In the same way one checks that $x'' = x$.

(iii) The superior quantile is $Q(p) = \sup\{x : F(x) \le p\}$. If $\{x : F(x) = p\}$ is at most a singleton, Q is continuous at p and we have nothing to prove. If not, then $Q(p) = \sup\{x : F(x) = p\}$ hence $Q(p) = x \Rightarrow F(x - 0) = F(x) = p$. Let $p_n > p$ be a sequence such that $p_n \to p$ and let $(p_n) = x_n$. Let $x' = \lim x_n$.

If $x' > x$, then $F(x) < F(x' - 0)$ therefore $p = F(x - 0) < F(x' - 0) \le F(x_n - 0) \le p_n$. Passing to the limit we find the absurdity $p = F(x) < F(x' - 0) \le p$.

Now suppose that Q is a quantile for F and it is right continuous. The discontinuities of Q are in those points p where the level set $I_F(p)$ contains non-degenerate intervals. This set is at most countable. For those points, let $\alpha(p) = \inf I_F(p) < \beta(p) = \sup I_F(p)$. Then we claim that $Q(p + 0) = \beta(p)$. Indeed, it is clear that $Q(p + 0) \ge \beta(p)$. If, ad absurdum, $Q(p + 0) > \beta(p)$, then there exists $x^* > \beta(p)$ such that $Q(p') \ge x^*$ for any $p' > p$. Then $p = F(\beta(p)) < F(x^* - 0) \le F(Q(p') - 0) \le p$ for all $p' > p$. This is an absurdity: if $p < p' < F(x^* - 0)$ then p' cannot be greater than $F(x^* - 0)$. It follows that $Q(p + 0) = \beta(p)$. If Q is right continuous, then $Q(p) = Q(p + 0) = \beta(p)$. But this is precisely the superior quantile, F^{-1}.

(iv) Define $F(x) = \begin{cases} m & \text{if } x \le \lambda(m + 0) \\ \sup\{p : \lambda(p) \le x\} & \text{if } m > \lambda(p) \\ M & \text{if } \lambda(M - 0) \le x \end{cases}$. Then F is right continuous. It will be

enough to prove that λ is a quantile for F, since the only quantile which is right continuous is the superior one. We have to check that $\lambda(p) = x \Rightarrow F(x - 0) \le p \le F(x)$. The level set $\{q : \lambda(q) = x\}$ is an interval with endpoints p_1 and p_2. Then $F(x) = p_2 \ge p$, $F(x - 0) = p_1 \le p$ hence $F(x - 0) \le p \le F(x)$. It follows that λ is the superior quantile for F. $\square$

The connection between the weak convergence of the probability measures on the real line and quantiles is given by following result.

Lemma A5. *Let $F_n, F : \mathcal{R} \to [0, 1]$ be distribution functions. Then $F_n(x) \to F(x)$ for all x which are continuity points of F if and only if $F_n^{-1}(p) \to F^{-1}(p)$ for $p \in (0, 1)$ with the possible exception of a countable set.*

Proof. Let $\Gamma = \{x : F_n(x) \to F(x)\}$. The complement of Γ is at most countable hence, Γ is a dense subset of $\mathbb{R}$. For x in Γ we have the equivalences

$$F(x) \le p \Leftrightarrow \forall k \exists\, m = m(k) \text{ such that } F_{m+n}(x) \le p + \frac{1}{k} \,\forall n$$

$$F(x) \ge p \Leftrightarrow \forall k \exists\, m = m(k) \text{ such that } F_{m+n}(x) \ge p - \frac{1}{k} \,\forall n$$

Here, m, n, k are positive integers
Or

$$\cap_{k=1}^{\infty} \cup_{m=1}^{\infty} \cap_{n=1}^{\infty} \left\{ F_{m+n} \le p + \frac{1}{k} \right\} \cap \Gamma = \cap_{k=1}^{\infty} \liminf_k \left\{ F_n \le p + \frac{1}{k} \right\} \cap \Gamma \tag{A9}$$

$$\{F \ge p\} \cap \Gamma = \cup_{k=1}^{\infty} \cap_{m=1}^{\infty} \cap_{n=1}^{\infty} \left\{ F_{m+n} \ge p - \frac{1}{k} \right\} \cap \Gamma = \cap_{k=1}^{\infty} \liminf_k \left\{ F_n \ge p - \frac{1}{k} \right\} \cap \Gamma \tag{A10}$$

As Γ is dense, $\sup(\{F \le p\} \cap \Gamma) = \sup(\{F \le p\})$ and $\inf(\{F \le p\} \cap \Gamma) = \inf(\{F \le p\})$.

Let $Q^+ = \sup\{F \le p\}$, $Q_n^+ = \sup\{F_n \le p\}$, $Q^- = \inf\{F \ge p\}$, and $Q_n^- = \inf\{F_n \ge p\}$, be the superior (inferior) quantiles of F and F_n. Taking the supremum in (A9) and infimum in (A10) we get $Q^+(p) = \lim\limits_{k \to \infty} (\liminf\limits_{n} Q_n{}^+)(p + \frac{1}{k}) \le (\liminf\limits_{n} Q_n{}^+)(p + 0)$.

$$Q^-(p) = \lim_{k \to \infty} (\limsup_{n} Q_n{}^-)(p - \frac{1}{k}) \ge (\limsup_{n} Q_n{}^-)(p - 0) \tag{A11}$$

Since all the four functions from (A11)—namely, Q^+, Q^-, $(\liminf\limits_{n} Q_n{}^+)$, $(\limsup\limits_{n} Q_n{}^-)$, are nondecreasing, they are all continuous with the possible exception of a set which is at most countable.

Moreover, excepting another at most countable set, $Q^+(p) = Q^-(p)$ and $Q_n^+(p) = Q_n^-(p)$. Put all these exception sets in a set N which is at most countable. It follows that if $p \notin N$ we have the inequalities: $(\liminf\limits_{n} Q_n{}^+)(p) = (\liminf\limits_{n} Q_n{}^+)(p + 0) \ge Q^+(p)$, $(\limsup\limits_{n} Q_n{}^-)(p) = (\limsup\limits_{n} Q_n{}^-)(p - 0) \le Q^-(p)$ which further implies $(\liminf\limits_{n} Q_n{}^+)(p) \ge Q^+(p) = Q^-(p) \ge (\limsup\limits_{n} Q_n{}^-)(p)$ otherwise written $F_n{}^{-1}(p) \to F^{-1}(p)$ for all $p \in (0,1)/(M \cup N)$.

Conversely, suppose that $F_n{}^{-1}(p) \to F^{-1}(p)$ for $p \in \Gamma$ with $\Gamma \subseteq (0,1)$ dense. According to Lemma A4(iv) F_n is the superior quantile of $F_n{}^{-1}$ and F is the superior quantile of F^{-1}. The proof goes in the same way as before. □

Remark A1. *Actually, one can be more precise:* $F_n{}^{-1}(p) \to F^{-1}(p)$ *for all p which are continuity points for* F^{-1} *(see Proposition 5 [30] (p. 250) or, for more general cases, [7,21,31]).*

References

1. MacArthur, R.H. On the relative abundance of bird species. *Proc. Natl. Acad. Sci. USA* **1957**, *43*, 293–295. [PubMed]
2. Cohen, J.E. Alternate derivation of a species-abundance relation. *Am. Nat.* **1968**, *102*, 165–172. [CrossRef]
3. Aly, E.A.A.; Beirlant, J.; Horvath, L. Strong and weak approximation of k-spacings processes. *Z. Wahrschein. Gebiete* **1984**, *66*, 461–484.
4. Bansal, P.; Arora, S.; Mahajan, K.K. Estimates of Inequality Indices Based on Simple Random, Ranked Set, and Systematic Sampling. *ISRN Probab. Stat.* **2013**. [CrossRef]
5. Beirlant, J. Strong approximations of the empirical and quantile processes of uniform spacings. *Limit Theoremes Probab. Stat.* **1984**, *1*, 77–89.
6. Barton, D.E.; David, F.N. Tests for randomness of points on a line. *Biometrika* **1956**, *43*, 104–112.
7. Csörgö, M.; Yu, H. Weak Approximations for Empirical Lorenz Curves and Their Goldie Inverses of Stationary Observations. *Adv. Appl. Probab.* **1999**, *31*, 698–719.
8. Cucala, L. Le Cam spacings theorem in dimension two. *arXiv* **2005**, arXiv:Math/0507367.
9. Durbin, J. Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. *Biometrika* **1975**, *62*, 5–22.
10. Le Cam, L. Un theoreme sur la division d'une intervalle par des points pres au hazard. *Publ. Inst. Statist. Univ. Paris* **1958**, *7*, 7–16.
11. Pyke, R. Spacings. *J. Roy. Statist. Sot. Ser. B* **1965**, *27*, 395–449.
12. Pyke, R. Spacings revisited. *Sixth Berkeley Symp. Math. Statist. Probab.* **1972**, *1*, 417–427.
13. Rao, J.S.; Sethuraman, J. Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. *Ann. Stat.* **1975**, *3*, 299–313. [CrossRef]
14. Stephens, M.A. Tests for the Uniform Distribution. In *Goodness-of-Fit Techniques*; D'Agostino, R., Stephens, M.A., Eds.; Marcel Dekker: New York, NY, USA, 1986.
15. Stephens, C.J. Towards Understanding the Lorenz Curve Using the Uniform distribution. In Proceedings of the Gini-Lorenz Conference, Siena, Italy, 23–26 May 2005; University of Siena: Siena, Italy, 2005.
16. Tung, D.T.; Jamalamadaka, S.R. U statistics based on spacings. *J. Stat. Plann. Infer.* **2012**, *142*, 673–684.
17. Wilks, S.S. *Mathematical Statistics*; John Wiley & Sons: Hoboken, NJ, USA, 1962.

18. Arnold, B.C. Majorization: Here, there and everywhere. *Stat. Sci.* **2007**, *22*, 407–413.

19. Arnold, B.C. *Majorization and the Lorenz Order: A Brief Introduction*; Springer: New York, NY, USA, 1987.

20. Sarabia, J.M.; Castillo, E.; Slottje, D.J. An ordered family of Lorenz curves. *J. Econom.* **1999**, *91*, 43–60.

21. Goldie, C.M. Convergence theorems for empirical Lorenz curves and their inverses. *Adv. Appl. Probab.* **1977**, *9*, 765–791.

22. Marshall, A.W.; Olkin, I.; Arnold, B.C. *Inequalities: Theory of Majorization and Its Applications*, 2nd ed.; Springer: New York, NY, USA, 2010.

23. Stoyan, D. *Qualitative Eigenschaften und Abschatzungen stochastischer Modelle*; Academie Verlag: Berlin, Germany, 1977.

24. Shorrocks, A.F. Ranking Income Distributions. *Economica* **1983**, *50*, 3–17. [CrossRef]

25. Billingsley, P. *Probability and Measure*; John Wiley & Sons: Hoboken, NJ, USA, 1995.

26. Billingsley, P. *Convergence of Probability Measures*; John Wiley & Sons: Hoboken, NJ, USA, 2013.

27. Arnold, B.C. Lorenz ordering of exponential order statistics. *Stat. Probab. Lett.* **1984**, *11*, 485–490.

28. Arnold, B.C. The Lorenz Curve: Evergreen after 100 years. In *Advances on Income Inequality and Concentration Measures*; Betti, G., Lemmi, A., Eds.; Routledge Frontiers of Political Economy; Routledge: London, UK, 2008; pp. 34–46.

29. Shaked, M.; Shanthikumar, J. *Stochastic Orders*; Springer: New York, NY, USA, 2007.

30. Fristedt, B.E.; Gray, L.F. *A Modern Approach to Probability Theory*; Springer Science & Business Media: New York, NY, USA, 2013.

31. Ahidar-Coutrix, A.; Berthet, P. Convergence of Multivariate Quantile Surfaces. Available online: https://hal.archives-ouvertes.fr/hal-01754841/document (accessed on 10 March 2020).