Results relating to a conjecture on the validity of a certain hypothesis test

Haben Michael[1]

[1]University of Massachusetts, Amherst, MA (hmichael@math.umass.edu)

SUMMARY: We consider the classical problem of forming a confidence interval for the mean of a distribution with bounded support based on an iid sample. Gaffke (2005) proposed such an interval and established that it is asymptotically valid. The interval's remarkably tight bounds on the mean led Gaffke to offer two conjectures: 1) The CI has guaranteed coverage for all sample sizes (Gaffke's weaker conjecture), and 2) Further, even after relaxing the assumption of identical distributions, the CI remains valid simultaneously for all the underlying means (Gaffke's stronger conjecture). Dormant in the literature for many years, Gaffke's conjectures have received renewed attention in recent years (Waudby-Smith and Ramdas, 2024; Phan et al., 2021; Learned-Miller and Thomas, 2019). We do not offer a proof of the conjecture but instead offer a collection of related results, including properties of the test statistic and special cases of the conjecture.

Let $n \geq 1$ pairs (a_i, b_i) be given such that $0 \leq a_i \leq 1$, $b_i \geq 1$, and $a_i < b_i$, $i = 1, \ldots, n$. Let $X_1, \ldots, X_n$, be mutually independent random variables with mean 1, with X_i supported on $\{a_i, b_i\}$, $i = 1, \ldots, n$, implying

$$X_i = \begin{cases} a_i & \text{w.p.} \quad \frac{b_i - 1}{b_i - a_i} \\ b_i & \text{w.p.} \quad \frac{1 - a_i}{b_i - a_i}. \end{cases}$$

Define the function $K : \mathbb{R}^{(\mathbb{N})} \to \mathbb{R}$ as follows.

Definition 1. *For $Z_0, \ldots, Z_n$ independent standard exponential random variables,*

$$K(x_1, \ldots, x_n) = \mathbb{P}(\sum_{i=1}^{n} (x_i - 1) Z_i < Z_0).$$

Gaffke's strong conjecture (Gaffke, 2005) is that for any $n \in \mathbb{N}$ and $\alpha \in [0, 1]$,

$$\mathbb{P}(K(X_1, \ldots, X_n) \leq \alpha) \leq \alpha. \tag{1}$$

Gaffke (2005) relates this conjecture to the validity of a certain one-sided test for the mean of independent, not necessarily identically distributed random variables with a uniform lower bound. See Section 6.

1 Properties of K

Theorem 2 lists properties following directly from Definition 1.

Theorem 2. *For reals $x_1, \ldots, x_n$,*

1. $0 \le K(x_1, \ldots, x_n) \le 1$.
2. $K(x_1, \ldots, x_n) = 1$ *if* $\max_i x_i \le 1$, *and* $K(x_1, \ldots, x_n) \to \infty$ *if* $\max_i x_i \to \infty$ *and* $\min_i x_i$ *is bounded below.*
3. $K(x_1, \ldots, x_n)$ *is symmetric in* $x_1, \ldots, x_n$.
4. $K(x_1, \ldots, x_n)$ *is non-increasing in each* x_i, *and strictly decreasing if* $x_i > 1$ *for some* i.
5. $K(x_1, \ldots, x_n, y) \underset{\le}{\overset{\ge}{=}} K(x_1, \ldots, x_n)$ *as* $y \underset{\ge}{\overset{\le}{=}} 1$.
6. $K(y_1^{(t)}, \ldots, y_n^{(t)})$ *is non-increasing in* $t \ge 0$, *where* $y_i^{(t)}$ *is given by the transformation* $x_i \mapsto y_i^{(t)} = 1 + t(x_i - 1)$.

Theorem 3. *For nonnegative reals $x_1, \ldots, x_n$,*

1. $\prod_{i: x_i \ge 1} \frac{1}{x_i} \le K(x_1, \ldots, x_n) \le \prod_{i=1}^n \frac{1}{x_i}$.
2. *If* $x_i \ge 1, i = 1, \ldots, n$, *then* $K(x_1, \ldots, x_n) = \frac{1}{\prod_{i=1}^n x_i}$.

Proof. Let $Z_0, \ldots, Z_n$ be independent standard exponentials, so their common cdf is $q \mapsto \max(0, 1 - e^{-q})$.

$$\begin{aligned}
K(x_1, \ldots, x_n) &= \mathbb{P}(\sum_{i=1}^n (x_i - 1) Z_i < Z_0) \\
&= \mathbb{E}(\mathbb{P}(\sum_{i=1}^n (x_i - 1) Z_i < Z_0 \Big| \sum_{i=1}^n (x_i - 1) Z_i)) \\
&= \mathbb{E}(\exp(-\sum_{i=1}^n (x_i - 1) Z_i); \sum_{i=1}^n (x_i - 1) Z_i \ge 0) + \mathbb{P}(\sum_{i=1}^n (x_i - 1) Z_i < 0). \quad (2)
\end{aligned}$$

If $x_i \ge 1, i = 1, \ldots, n$, then $\sum_{i=1}^n (x_i - 1) Z_i > 0$ almost surely, and (2) is

$$\begin{aligned}
&= \mathbb{E}(\exp(-\sum_{i=1}^n (x_i - 1) Z_i)) \\
&= \prod_{i=1}^n \mathbb{E}(\exp(-(x_i - 1) Z_i)) \\
&= \frac{1}{\prod_{i=1}^n x_i}.
\end{aligned}$$

Otherwise, (2) is

$$\le \mathbb{E}(\exp(-\sum_{i=1}^n (x_i-1)Z_i); \sum_{i=1}^n (x_i-1)Z_i \ge 0) + \mathbb{E}(\exp(-\sum_{i=1}^n (x_i-1)Z_i); \sum_{i=1}^n (x_i-1)Z_i < 0)$$

$$= \mathbb{E}(\exp(-\sum_{i=1}^n (x_i-1)Z_i))$$
$$= \prod_{i=1}^n \mathbb{E}(\exp(-(x_i-1)Z_i))$$
$$= \frac{1}{\prod_{i=1}^n x_i}.$$

□

Theorem 5 gives an explicit formula for K.

Definition 4.

$$k_0 : \mathbb{R} \to (0,1], \quad k_0(x) = \begin{cases} 1, & x \le 1 \\ 1/x, & x > 1. \end{cases}$$

For $x \ge 0$, $k_0(x) = 1 \wedge (1/x)$.

Theorem 5. *For distinct $x_1, \ldots, x_n$,*

$$K(x_1, \ldots, x_n) = \sum_{i=1}^n k_0(x_i) \prod_{j=1, j\neq i}^n \frac{x_i - 1}{x_i - x_j}. \tag{3}$$

Proof. Fix $x_1, \ldots, x_n$. The characteristic function of $\sum_{i=1}^n (x_i-1)Z_i - Z_0$ is

$$\phi(t) = \frac{1}{1+it} \prod_{i=1}^n \frac{1}{1-(x_i-1)it}.$$

By a Fourier inversion formula (Gil-Pelaez, 1951),

$$K(x_1, \ldots, x_n) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\mathrm{Im}(\phi(t))}{t} dt. \tag{4}$$

Let $y_0 = -1$ and $y_i = x_i - 1, i = 1, \ldots, n$, so $\phi(t) = \prod_{i=0}^n (1 - iy_it)^{-1}$ and $\mathrm{Im}\frac{1}{1-iy_it} = \frac{y_it}{1+y_i^2t^2}$. Let $\phi(t)$ have partial fraction decomposition

$$\phi(t) = \prod_{i=0}^n \frac{1}{1-iy_it} = \sum_{i=0}^n \frac{c_i}{1-iy_it}, \quad c_i = \frac{1}{\prod_{j:j\neq i}(1-y_j/y_i)}.$$

The integral in (4) is

$$\begin{aligned}\int_0^\infty \frac{\mathrm{Im}(\phi(t))}{t}dt &= \int_0^\infty \sum_{i=0}^n \frac{c_i y_i t}{1+y_i^2 t} \\ &= \sum_{i=0}^n c_i \arctan(y_i t)\Big|_0^\infty \\ &= \frac{\pi}{2}\sum_{i=0}^n \frac{\mathrm{sgn}(y_i)}{\prod_{j:j\neq i}(1-y_j/y_i)}. \end{aligned} \tag{5}$$

Substituting (5) into (4),

$$\begin{aligned} K(x_1,\ldots,x_n) &= \frac{1}{2}(1-\sum_{i=0}^n \frac{\mathrm{sgn}(y_i)y_i^n}{\prod_{j:j\neq i}(y_i-y_j)}) \qquad (6)\\ &= \frac{1}{2}(1-\frac{(-1)^{n+1}}{\prod_{j=1}^n(-x_j)} - \sum_{i=1}^n \frac{\mathrm{sgn}(x_i-1)(x_i-1)^n}{x_i\prod_{j:j\neq i}(x_i-x_j)}). \end{aligned}$$

The identity

$$\sum_{i=1}^n \frac{(x_i-1)^{n-1}}{\prod_{j:j\neq i}(x_i-x_j)} = 1$$

follows by interpreting the lhs as the Lagrange interpolating polynomial of $x \mapsto 1$ at nodes $x_1,\ldots,x_n$, evaluated at $x=1$, allowing (6) to be written as

$$K(x_1,\ldots,x_n) = \frac{1}{2}\sum_{i=1}^n \frac{(x_i-1)^{n-1}}{\prod_{j:j\neq i}(x_i-x_j)}(1+\frac{1}{\prod_{j=1}^n x_j} - \frac{|x_i-1|}{x_i}). \tag{7}$$

A similar identity,

$$\sum_{i=1}^n (1-x_i)^{n-1}\prod_{j:j\neq i}\frac{-x_j}{x_i-x_j} = 1$$

follows by evaluating at 0 the Lagrange interpolating polynomial for $x \mapsto (1-x)^{n-1}$ at nodes $x_1,\ldots,x_n$, implying

$$\sum_{i=1}^n \frac{(x_i-1)^{n-1}}{\prod_{j:j\neq i}(x_i-x_j)}\frac{1}{\prod_{j=1}^n x_j} = \sum_{i=1}^n \frac{(x_i-1)^{n-1}}{\prod_{j:j\neq i}(x_i-x_j)}\frac{1}{x_i}.$$

Therefore (7) is

$$K(x_1,\ldots,x_n) = \sum_{i=1}^n \frac{(x_i-1)^{n-1}}{\prod_{j:j\neq i}(x_i-x_j)}\frac{1}{2}(1+\frac{1}{x_i}-\frac{|x_i-1|}{x_i}),$$

where

$$\frac{1}{2}(1+\frac{1}{x_i}-\frac{|x_i-1|}{x_i}) = \begin{cases} 1, & x \leq 1 \\ 1/x, & x > 1. \end{cases}$$

$\square$

1.1 Recursion formula and consequences

Theorem 6. *Let $f_0 : \mathbb{R} \to \mathbb{R}$ be given. Then a function $f : \mathbb{R}^{(\mathbb{N})} \to \mathbb{R}$ satisfies for all distinct $x_1, \ldots, x_n$,*

$$f(x_1, \ldots, x_n) = \begin{cases} \frac{x_1-1}{x_1-x_n} f(x_1, \ldots, x_{n-1}) + \frac{x_n-1}{x_n-x_1} f(x_2, \ldots, x_n), & n > 1 \\ f_0(x_1), & n = 1 \end{cases} \tag{8}$$

if and only if

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_0(x_i) \prod_{j:j\neq i}^{n} \frac{x_i - 1}{x_i - x_j}. \tag{9}$$

Then also, for $n \geq 3$,

$$f(\boldsymbol{x}_{-3}) = \frac{x_1 - x_3}{x_1 - x_2} f(\boldsymbol{x}_{-2}) + \frac{x_2 - x_3}{x_2 - x_1} f(\boldsymbol{x}_{-1}),$$

that is,

$$(x_2 - x_1) f(\boldsymbol{x}_{-3}) + (x_1 - x_3) f(\boldsymbol{x}_{-2}) + (x_3 - x_1) f(\boldsymbol{x}_{-1}) = 0. \tag{10}$$

Proof. Suppose first that $f(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_0(x_i) \prod_{j:j\neq i}^{n} \frac{x_i-1}{x_i-x_j}$. Then,

$$\begin{aligned}
&\frac{x_1 - x_3}{x_1 - x_2} f(\boldsymbol{x}_{-2}) + \frac{x_2 - x_3}{x_2 - x_1} f(\boldsymbol{x}_{-1}) \\
&= \frac{x_1 - x_3}{x_1 - x_2} \sum_{i\neq 2} f_0(x_i) \prod_{j:j\neq i, j\neq 2} \frac{x_i - 1}{x_i - x_j} + \frac{x_2 - x_3}{x_2 - x_1} \sum_{i\neq 1} f_0(x_i) \prod_{j:j\neq i, j\neq 1} \frac{x_i - 1}{x_i - x_j} \\
&= \frac{x_1 - x_3}{x_1 - x_2} f_0(x_1) \prod_{j:j\neq 1, j\neq 2} \frac{x_1 - 1}{x_1 - x_j} + \frac{x_2 - x_3}{x_2 - x_1} f_0(x_2) \prod_{j:j\neq 2, j\neq 1} \frac{x_2 - 1}{x_2 - x_j} \\
&+ \frac{1}{x_1 - x_2} \sum_{j\neq 1, j\neq 2} f_0(x_i)((x_1 - x_2) \prod_{j:j\neq i, j\neq 2} \frac{x_i - 1}{x_i - x_j} + (x_3 - x_2) \prod_{j:j\neq i, j\neq 1} \frac{x_i - 1}{x_i - x_j}) \\
&= (x_1 - x_3) f_0(x_1) \prod_{j\neq 1} \frac{x_1 - 1}{x_1 - x_j} + (x_2 - x_3) f_0(x_2) \prod_{j\neq 2} \frac{x_2 - 1}{x_2 - x_j} \\
&+ \frac{1}{x_1 - x_2} \sum_{i\neq 1, i\neq 2} f_0(x_i)((x_1 - x_3)(x_i - x_2) + (x_3 - x_2)(x_i - x_1)) \prod_{j:j\neq i} \frac{x_i - 1}{x_i - x_j} \\
&= f_0(x_1) \prod_{j\neq 1, j\neq 3} \frac{x_1 - 1}{x_1 - x_j} + f_0(x_2) \prod_{j\neq 2, j\neq 3} \frac{x_2 - 1}{x_2 - x_j} + \sum_{i\neq 1, i\neq 2, i\neq 3} f_0(x_i) \prod_{j:j\neq i j\neq 3} \frac{x_i - 1}{x_i - x_j} \\
&= \sum_{i\neq 3} f_0(x_i) \prod_{j:j\neq i, j\neq 3} \frac{x_i - 1}{x_i - x_j} \\
&= f(_{-3}).
\end{aligned}$$

Therefore f satisfies (10).

The value of the function

$$(x_1, \ldots, x_n) \mapsto \sum_{i=1}^{n} f_0(x_i) \prod_{j:j\neq i} \frac{x_i - 1}{x_i - x_j}$$

is unaffected by the presence of any x_i that equals 1. Taking $x_3 = 1$ in the above derivation shows that (9) satisfies the recursion (8). It satisfies the base case, as well, so it must equal f. □

Theorem 6 establishes properties of K using the sum-product formula (3), independently of the probabilistic definition of K in Definition 1. Kaminsky et al. (1984) established the recursion formula (8) for K as defined in Definition 1. By the preceding theorem, the result of Kaminsky et al. (1984) implies the conclusion of Theorem 5. We don't know if it was previously known that this recursion formula implies the sum-product formula (9). In the other direction, Theorem 5 gives an alternative proof of the result of Kaminsky et al. (1984) for standard exponentials, and strengthens it by (10).

Definition 1 shows that f is symmetric. Similarly, the formula (9), which is symmetric in x_i, also shows that the recursive definition (8) is well-defined, i.e., doesn't depend on the x_i, x_j chosen.

Definition 7.

$$p(x, y) : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto \frac{y - 1}{y - x}, \tag{11}$$

for $x \neq y$.

Theorem 8. *For reals* $x \geq 0, y \geq 0$, x *and* y *on opposite sides of* 1,

$$p(x, y) \leq k_0(x).$$

When $x < 1, y > 1$, *equality occurs as* $y \to \infty$. *When* $x > 1, y < 1$, *equality occurs as* $y \to 0$.

Proof. If $x \leq 1$, the rhs is 1 while the lhs is a probability ≤ 1. If $x > 1$, then $y < 1$ and $p(x, y) = 1 - (x - 1)/(x - y)$ is decreasing in y, and maximized at $y = 0$, when both sides of the inequality are $1/x$. □

Theorem 9. *For distinct non-negative reals* $x, x_1, \ldots, x_n$,

$$k_0(x)K(x_1, \ldots, x_n) \leq K(x, x_1, \ldots, x_n).$$

Proof. If $x \leq 1$, monotonicity of K implies

$$k_0(x)K(x_1, \ldots, x_n) \leq K(x_1, \ldots, x_n) \leq K(x, x_1, \ldots, x_n).$$

Suppose $b = x > 1$. By induction on n. The base case requires $1/bK(x_1) \leq K(b, x_1)$. If $x_1 \leq 1$, $1/bK(x_1) = 1/b = K(b) \leq K(x_1, b)$, while if $x_1 \geq 1$, $1/bK(x_1) = 1/(bx_1) = K(b, x_1)$. Supposing the result holds for $n - 1$, then

$$\begin{aligned}
\frac{1}{b}K(x_1, \ldots, x_n) &= \frac{1}{b}(p(x_n, x_1)K(x_1, \ldots, x_{n-1}) + p(x_1, x_n)K(x_2, \ldots, x_n)) \\
&\leq p(x_n, x_1)K(b, x_1, \ldots, x_{n-1}) + p(x_1, x_n)K(b, x_2, \ldots, x_n) \\
&= K(b, x_1, \ldots, x_n).
\end{aligned}$$

□

Theorem 10. *Given* $\boldsymbol{a} \in [0,1]^r, b_1 \geq 1, b_2 \geq 1$,

$$k_0(b_2)K(\boldsymbol{a}, b_1) \leq K(\boldsymbol{a}, b_1 b_2).$$

Proof. The conclusion is

$$\begin{aligned}
&0 \leq K(\boldsymbol{a}, b_1 b_2) - \frac{1}{b_2} K(\boldsymbol{a}, b_1) \\
&0 \leq 1 + \prod_{i=1}^{r} p(a_i, b_1 b_2)(\frac{1}{b_1 b_2} - 1) - \frac{1}{b_2}(\prod_{i=1}^{r} p(a_i, b_1)(\frac{1}{b_1} - 1) + 1) \\
&0 \leq (b_2 - 1)(1 - \prod_{i=1}^{r} p(a_i, b_1 b_2)) - \frac{b_1 - 1}{b_1}(\prod_{i=1}^{r} p(a_i, b_1 b_2)) - \prod_{i=1}^{r} p(a_i, b_1)) \\
&0 \leq (b_2 - 1)(1 - g(b_1 b_2)) - \frac{b_1 - 1}{b_1}(g(b_1 b_2) - g(b_1)), \qquad (12)
\end{aligned}$$

where g is the nonnegative increasing function given by $x \mapsto \prod_{i=1}^{r} \frac{x-1}{x-a_i}$. We bound the two terms on the rhs of (12) using the log-derivative $g'(x)/g(x) = \frac{1}{x-1} \sum_{i=1}^{r} \frac{1-a_i}{x-a_i}$. For the first term,

$$\begin{aligned}
1 - g(b_1 b_2) &= \int_{b_1 b_2}^{\infty} g'(x) dx \\
&= \int_{b_1 b_2}^{\infty} \frac{1}{x-1} \sum_{i=1}^{r} \frac{1-a_i}{x-a_i} g(x) dx \\
&\geq g(b_1 b_2) \int_{b_1 b_2}^{\infty} \frac{1}{x-1} \sum_{i=1}^{r} \frac{1-a_i}{x-a_i} dx \\
&\geq g(b_1 b_2) \int_{b_1 b_2}^{\infty} \sum_{i=1}^{r} \frac{1-a_i}{(x-a_i)^2} dx \\
&= g(b_1 b_2) \sum_{i=1}^{s} \frac{1-a_i}{b_1 b_2 - a_i}.
\end{aligned}$$

For the second term,

$$\begin{aligned}
g(b_1 b_2) - g(b_1) &= \int_{b_1}^{b_1 b_2} \frac{1}{x-1} \sum_{i=1}^{r} \frac{1-a_i}{x-a_i} g(x) dx \\
&\leq \int_{b_1}^{b_1 b_2} \frac{1}{(x-1)^2} \sum_{i=1}^{r} (1-a_i) g(b_1 b_2) dx \\
&= g(b_1 b_2) \sum_{i=1}^{r} (1-a_i) \frac{b_1(b_2-1)}{(b_1-1)(b_1 b_2 - 1)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&(b_2-1)(1-g(b_1b_2)) - \frac{b_1-1}{b_1}(g(b_1b_2)-g(b_1)) \\
&\geq g(b_1b_2)\sum_{i=1}^{s}\frac{1-a_i}{b_1b_2-a_i} - g(b_1b_2)\sum_{i=1}^{r}(1-a_i)\frac{b_1(b_2-1)}{(b_1-1)(b_1b_2-1)} \\
&= (b_2-1)g(b_1b_2)(\sum_{i=1}^{r}\frac{1-a_i}{b_1b_2-1} - \sum_{i=1}^{r}\frac{1-a_i}{b_1b_2-a_i}) \\
&\geq 0,
\end{aligned}$$

since $b_1b_2-1 \leq b_1b_2-a_i, i=1,\ldots,r$. □

The corresponding result for $a' \in [0,1]$, $k_0(a')K(a,\boldsymbol{b}) \leq K(aa',\boldsymbol{b})$, follows from the monotonicity property (Theorem 2) as in Theorem 9 . Theorem 10 in conjunction with Theorem 23 below strengthens Theorem 9, in the special case where there is a single component x_i of $\boldsymbol{x}$ that is ≥ 1.

Definition 11. *A "leaf node" is a vector lying in* $[0,1]^{(\mathbb{N})}$ *or in* $[1,\infty)^{(\mathbb{N})}$.

If x_i and x_j lie on opposite sides of the number 1 then $p(x_i,x_j)=(x_j-1)/(x_j-x_i)$ and $p(x_j,x_i)=(x_i-1)/(x_i-x_j)$ are both ≥ 0 and (8) is a convex combination. By iterating the recursion, $K(x_1,\ldots,x_n)$ can be expressed as a convex combination of the values of K on leaf nodes, namely, 1 and the reciprocal of products of b-values.

Theorem 12. *Suppose* $f:\mathbb{R}^{(\mathbb{N})}\to\mathbb{R}$ *satisfies the recursion* (8)*. Then there are non-negative reals* $w_1,\ldots,w_{m+n}$*,* $\sum_{i=1}^{m+n}w_i=1$*, such that*

$$f(a_1,\ldots,a_m,b_1,\ldots,b_n)=\sum_{i=1}^{m}w_if(a_1,\ldots,a_i)+\sum_{i=1}^{n}w_{m+i}f(b_1,\ldots,b_i), \tag{13}$$

where an empty sum is taken to be 0*. The weights* w_i *depend on* a_i *and* b_i *only through the function* p *given in* (11).

Proof. By induction on $k=m+n$. If $k=1, m=0$, or $n=0$, then take $w_1=1$. Suppose that $\min(m,n)>0$, and suppose the conclusion (13) holds for any m,n with $m+n=k$, and let $(\boldsymbol{a}, \boldsymbol{b}) = (a_1, \ldots, a_m, b_1, \ldots, b_n)$, be given with $m + n = k + 1$. Then,

$$\begin{aligned}
f(\boldsymbol{a}, \boldsymbol{b}) &= p(a_m, b_n) f(a/a_{m-1}, b) + p(b_n, a_m) f(a, b/b_{n-1}) \\
&= p(a_m, b_n)(\sum_{i=1}^{m-1} w'_i f(a_1, \ldots, a_i) + \sum_{i=1}^{n} w'_{m-1+i} f(b_1, \ldots, b_i)) \\
&+ p(b_n, a_m)(\sum_{i=1}^{m} w''_i f(a_1, \ldots, a_i) + \sum_{i=1}^{n-1} w''_{m+i} f(b_1, \ldots, b_i)) \\
&= \sum_{i=1}^{m-1} f(a_1, \ldots, a_i)(p(a_m, b_n) w'_i + p(b_n, a_m) w''_i) + p(b_n, a_m) w''_m f(a_1, \ldots, a_m) \\
&+ \sum_{i=1}^{n-1} f(b_1, \ldots, b_i)(p(a_m, b_n) w'_{m-1+i} + p(b_n, a_m) w''_{m+i}) + p(a_m, b_n) w'_{m-1+n} f(b_1, \ldots, b_n),
\end{aligned}$$

where w'_i and w''_i are non-negative, $\sum_{i=1}^{m+n-1} w'_i = 1$, and $\sum_{i=1}^{m+n-1} w''_i = 1$. Then

$$w_i = \begin{cases} p(a_m, b_n) w'_i + p(b_n, a_m) w''_i, & 1 \leq i \leq m-1 \\ p(b_n, a_m) w''_m, & i = m \\ p(a_m, b_n) w'_{i-1} + p(b_n, a_m) w''_i, & m+1 \leq i \leq m+n-1 \\ p(a_m, b_n) w'_{m-1+n}, & i = m+n \end{cases}$$

are nonnegative and sum to 1. □

When f in Theorem 12 is taken to be K, the conclusion is that there are $w_1, \ldots, w_n$, $\sum_{i=1}^{n} w_i \leq 1$, $w_i \geq 0$, so that

$$K(a_1, \ldots, a_m, b_1, \ldots, b_n) = 1 - \sum_{i=1}^{n} w_i + \sum_{i=1}^{n} w_i/(b_1 \cdot \ldots \cdot b_i).$$

The effect on f of scaling all arguments to or from 1 uniformly is determined by the effect on leaf vectors.

Theorem 13. *Let f be defined as in Theorem 6, and for $\alpha \geq 0$ let $h_\alpha : \mathbb{R} \to \mathbb{R} : x \mapsto 1 + \alpha(x-1)$. Then $f(h_\alpha(x_1), \ldots, h_\alpha(x_n))$ is monotonic in α for any $(x_1, \ldots, x_n) \in \mathbb{R}^n$ iff it is for those $(x_1, \ldots, x_n) \in \mathbb{R}^n$ with $x_i \leq 1$, all i, and those with $x_i \geq 1$, all i.*

Proof. By Theorem 12 f may be expressed as a convex combination of its values at vectors with all components ≤ 1 or all components ≥ 1, with weights that depend on $x_1, \ldots, x_n$ only through $p(\cdot, \cdot)$. However, $p(h_\alpha(x), h_\alpha(y)) = p(x, y)$. □

1.2 K leave-one-outs and c

Multiplying the lhs of the recursion formula (8) by $1 = p(x_i, x_j) + p(x_j, x_i)$ gives

$$(\frac{x_i - 1}{x_i - x_j} + \frac{x_j - 1}{x_j - x_i}) K(\boldsymbol{x}) = \frac{x_i - 1}{x_i - x_j} K(\boldsymbol{x}_{-j}) + \frac{x_j - 1}{x_j - x_i} K(\boldsymbol{x}_{-i}),$$

which may be re-arranged as

$$\frac{K(\boldsymbol{x}_{-j}) - K(\boldsymbol{x})}{x_j - 1} = \frac{K(\boldsymbol{x}_{-i}) - K(\boldsymbol{x})}{x_i - 1}.$$

Therefore the ratio

$$\frac{K(\boldsymbol{x}_{-i}) - K(\boldsymbol{x})}{x_i - 1}$$

doesn't depend on the omitted component x_i, only on $x_1, \ldots, x_n$.

Definition 14. *For reals $x_1, \ldots, x_n$, distinct from 1,*

$$c(\boldsymbol{x}) = \frac{K(\boldsymbol{x}_{-i}) - K(\boldsymbol{x})}{x_i - 1}, 1 \leq i \leq n.$$

Combined with Theorem 2 and 16 , K is seen to be Lipschitz in each component locally at 1 with Lipschitz constant $c(\boldsymbol{x})$:

$$|K(\boldsymbol{x}) - K(x_1, \ldots, 1, \ldots, x_n)| = c(\boldsymbol{x})|x_i - 1|$$

Theorem 15.

$$c(\boldsymbol{x}) = \sum_{i=1}^{n} c_0(x_i) \prod_{j:j \neq i} p(x_j, x_i) \tag{14}$$

where

$$c_0 : \mathbb{R} \to [0, 1], \quad c_0(x) = \begin{cases} 0, & 0 < x < 1 \\ \frac{1}{x}, & x > 1 \\ \textit{undefined}, & \textit{otherwise.} \end{cases}$$

Proof. If $0 < x < 1$, $c(x) = (k(1) - 1)/(x - 1) = 0 = c_0(x)$, and if $x > 1$, $c(x) = (k(1) - 1/x)/(x - 1) = 1/x = c_0(x)$. By Theorem 6 the formula (14) follows on establishing the recursion (8) for c,

$$p(x_1, x_n)c(\boldsymbol{x}_{-1}) + p(x_n, x_1)c(\boldsymbol{x}_{-n}) = c(\boldsymbol{x}). \tag{15}$$

Suppose $n = 2$. If $x_1 < 1$ and $x_2 < 1$, both sides of (15) are 0. If $x_1 > 1$ and $x_2 > 1$, (15) becomes

$$\frac{x_2 - 1}{x_2 - x_1} \frac{1}{x_2} + \frac{x_1 - 1}{x_1 - x_2} \frac{1}{x_2} = \frac{1}{x_1 x_2},$$

which also holds. If $x_1 < 1$ and $x_2 > 1$, (15) becomes

$$p(x_1, x_2)\frac{1}{x_2} = \frac{1}{x_2 - 1}(1 - \frac{p(x_1, x_2)}{x_2} - p(x_2, x_1)),$$

which also holds. When $n \geq 3$,

$$
\begin{aligned}
&p(x_1, x_n)c(\boldsymbol{x}_{-1}) + p(x_n, x_1)c(\boldsymbol{x}_{-n}) \\
&= p(x_1, x_n)\frac{K(\boldsymbol{x}_{-1,-2}) - K(\boldsymbol{x}_{-1})}{x_2 - 1} + p(x_n, x_1)\frac{K(\boldsymbol{x}_{-n,-2}) - K(\boldsymbol{x}_{-n})}{x_2 - 1} \\
&= \frac{1}{x_2 - 1}(K(\boldsymbol{x}_{-2}) - K(\boldsymbol{x})) \\
&= c(\boldsymbol{x}),
\end{aligned}
$$

using the recursion (8) for K. □

The process used to define c can be iterated. The function c satisfies the recursion (8), so that as with K the ratio

$$d(\boldsymbol{x}) = \frac{c(\boldsymbol{x}_{-i}) - c(\boldsymbol{x})}{x_i - 1}$$

must be independent of i. By the same method of proof as in Theorem 15, d satisfies the formula

$$d(\boldsymbol{x}) = \sum_{i=1}^{n} d_0(x_i) \prod_{j:j \neq i} p(x_j, x_i)$$

where

$$d_0 : \mathbb{R} \to [0,1], \quad d_0(x) = \begin{cases} 0, & 0 < x < 1 \\ \frac{1}{x(x-1)}, & x > 1 \\ \text{undefined}, & \text{otherwise.} \end{cases}$$

Therefore, as the leave-one-out

$$K(\boldsymbol{x}_{-i}) = (x_i - 1)c(\boldsymbol{x})$$

may be expessed as a linear function of x_i with coefficients independent of i, the "leave-two-out"

$$K(\boldsymbol{x}_{-i,-j}) = (x_i + x_j - 1)(c(\boldsymbol{x}) - d(\boldsymbol{x})) - c(\boldsymbol{x}) + x_i x_j d(\boldsymbol{x})$$

may be expressed as a quadratic function of x_i, x_j with coefficients independent of i, j, and so on.

Properties following from the formula (14):

Theorem 16. *For distinct non-negative reals $x_1, \ldots, x_n$, different from 1,*

1. $c(\boldsymbol{x}) = 0$ *if* $\max_i x_i \leq 1$ *and* $c(\boldsymbol{x}) = 1/\prod_i x_i$ *if* $\min_i x_i \geq 1$.

2. $0 \leq c(\boldsymbol{x}) \leq 1$

Proof. The formula (14) implies that $c(\boldsymbol{x}) = 0$ i $x_i \leq 1$, all i. Since k_0 and c_0 agree on $(1, \infty)$, comparison of formulas (3) and (14) shows that K and c agree when $x_i > 1$, all i. Theorem 12 then implies that c may be written as a convex combination

$$c(\boldsymbol{a}, \boldsymbol{b}) = \sum_{i=1}^{n} w_i / (b_1 \cdot \ldots \cdot b_i),$$

which must then lie in $[0, 1]$. □

Theorem 17. *For distinct non-negative reals $x_1 < \ldots < x_n$, $K(\boldsymbol{x}_{-i})$ is non-decreasing in i.*

Proof.

$$K(\boldsymbol{x}_{-i}) = K(\boldsymbol{x}) + c(\boldsymbol{x})(x_i - 1)$$

and by Theorem 16 $c(\boldsymbol{x}) \geq 0$. □

Theorem 18. *Let $x_1, \ldots, x_n$, be distinct non-negative reals different from 1, and for $\alpha \geq 0$ let h_α be the homothety transformation given in Theorem 13. Then*

$$c(h_\alpha(\boldsymbol{x})) - c(\boldsymbol{x}) = K(h_\alpha(\boldsymbol{x})) - K(\boldsymbol{x}).$$

Proof. Let f be defined as the difference $c - k$. Then for $\boldsymbol{x}$ as in the theorem statement,

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} f_0(x_i) \prod_{j:j \neq i} p(x_j, x_i),$$

where

$$f_0 : \mathbb{R} \to \mathbb{R}, \quad f_0(x) = \begin{cases} 0, & 0 < x < 1 \\ 1, & x > 1 \\ \text{undefined}, & \text{otherwise.} \end{cases}$$

However, both p and f_0 are invariant under h_α. □

Theorem 19. *For distinct non-negative reals $x_1, \ldots, x_n$,*

$$\min_i K(\boldsymbol{x}_{-i})^{\frac{n}{n-1}} \leq K(\boldsymbol{x}) \leq \max_i K(\boldsymbol{x}_{-i}).$$

The first inequality is tight iff $x_i \leq 1$, all i, or all x_i are equal. The second inequality is tight iff $x_i \leq 1$, all i.

Proof. The second inquality is

$$K(\boldsymbol{x}) \leq K(\boldsymbol{x}) + c(\boldsymbol{x})(\max_i x_i - 1).$$

If any x_i is ≥ 1 then the rhs is $\geq K(\boldsymbol{x})$, and the inequality gap vanishes iff $\max_i x_i \downarrow 1$. If $x_i \leq 1$ for all i then both sides are $= 1$.

The first inequality is

$$K(\boldsymbol{x}) + (\min_i x_i - 1)c(\boldsymbol{x}) \le K(\boldsymbol{x})^{\frac{n}{n-1}},$$

or

$$(1 - \min_i x_i)c(\boldsymbol{x}) \ge K(\boldsymbol{x})(1 - K(\boldsymbol{x})^{-\frac{1}{n}}).$$

If $\min_i x_i \le 1$, the lhs is non-negative and the rhs is non-positive, with equality iff $x_i \le 1$, all i. If $x_i \ge 1$ for all i then the first inequality is

$$(\frac{\prod_{i=1}^n x_i}{\max_i x_i})^{\frac{n}{n-1}} \le \prod_{i=1}^n x_i$$

or $\prod_{i=1}^n x_i \le (\max_i x_i)^n$, which holds with equality iff all x_i are equal. □

Theorem 20. *For distinct non-negative reals $x_1, \dots, x_m$, and distinct non-negative reals $y_1, \dots, y_n$, all different from 1, if*

$$\max_i K(\boldsymbol{x}_{-i}) \le \min_i K(\boldsymbol{y}_{-i})$$

then

$$K(\boldsymbol{x}) \le K(\boldsymbol{y}).$$

Proof. Since $c(\boldsymbol{x}) \ge 0, c(\boldsymbol{y}) \ge 0$,

$$K(\boldsymbol{x}) \le K(\boldsymbol{x}) + c(\boldsymbol{x})(\max_i x_i - 1) \le K(\boldsymbol{y}) + c(\boldsymbol{y})(\min_i y_i - 1) \le K(\boldsymbol{y})$$

unless $x_i < 1$, all i, or $y_i > 1$, all i, or both. If $x_i < 1$, all i, then $1 = \max_i K(\boldsymbol{x}_{-i}) \le \min K(\boldsymbol{y}_{-i})$ implies $y_i \le 1$, all i, and $K(\boldsymbol{y}) = 1 \ge K(\boldsymbol{x})$. Suppose $\min_i y_i > 1$. Then $K(\boldsymbol{y}) = c(\boldsymbol{y}) = 1/\prod_i y_i$, and

$$\max_i K(\boldsymbol{x}_{-i}) \le K(\boldsymbol{y}) + c(\boldsymbol{y})(\min_i y_i - 1) = K(\boldsymbol{y}) \min_i y_i = \frac{\min_i y_i}{\prod_i y_i}.$$

By Theorem 19

$$K(\boldsymbol{x}) \le \max_i K(\boldsymbol{x}_{-i}) \le (\max_i K(\boldsymbol{x}_{-i}))^{\frac{n}{n-1}} \le (\frac{\min_i y_i}{\prod_i y_i})^{\frac{n}{n-1}},$$

and the rightmost term is $\le K(\boldsymbol{y}) = 1/\prod_i y_i$ iff $(\min_i y_i)^n \le \prod_i y_i$. □

Theorem 21. *For distinct non-negative reals $x_1, \dots, x_n$, different from 1,*

$$K(\boldsymbol{x}) = 1 - \sum_{i=1}^n (x_i - 1)c(x_1, \dots, x_i).$$

Proof. The $n = 1$ case follows from the definitions of k_0 and c_0. The $n \geq 2$ case is

$$\begin{aligned}
&1 - K(\boldsymbol{x}) \\
&= 1 - K(x_1) + K(x_1) - K(x_1, x_2) + \ldots + K(x_1, \ldots, x_{n-1}) - K(x_1, \ldots, x_n) \\
&= K(x_1/x_1) - K(x_1) + K((x_1, x_2)/x_2) - K(x_1, x_2) + \ldots + K(\boldsymbol{x}_{-n}) - K(\boldsymbol{x}) \\
&= (x_1 - 1)c(x_1) + (x_2 - 1)c(x_1, x_2) + \ldots + (x_n - 1)c(x_1, \ldots, x_n).
\end{aligned}$$

□

Lemma 22. *1. Let $\boldsymbol{a} = (a_1, \ldots, a_r)$ with $a_i \in [0, 1]$ for $1 \leq i \leq r, r \geq 1$. Let $b_1 > 1, b_2 > 1$. Then*

$$K(\boldsymbol{a}, b_1, b_2) \leq K(\boldsymbol{a}, b_1 b_2)$$

with equality iff $\boldsymbol{a} = \mathbf{1}$.

2. Let $\boldsymbol{b} = (b_1, \ldots, b_s)$ with $b_i \geq 1$ for $1 \leq i \leq s, s \geq 1$. Let a_1 and a_2 lie in $[0, 1)$. Then

$$K(a_1 a_2, \boldsymbol{b}) \leq K(a_1, a_2, \boldsymbol{b})$$

with equality iff $\boldsymbol{b} = \mathbf{1}$.

Proof. 1. Let $\lambda_1 = b_1 - 1, \lambda_2 = b_2 - 1$, and $\lambda_{12} = \lambda_1 + \lambda_2 + \lambda_1 \lambda_2 = b_1 b_2 - 1$. Let $\Delta = \sum_{i=1}^{r} (1 - a_i) Z_i$, where $Z_i, i = 0, \ldots, r$, are independent standard exponentials. Then $\Delta \geq 0$ a.s. Let

$$\begin{aligned}
F_{\lambda_1 W_1 + \lambda_2 W_2} &: q \mapsto \max(0, 1 - \frac{\lambda_1 e^{-q/\lambda_1} - \lambda_2 e^{-q/\lambda_2}}{\lambda_1 - \lambda_2}) \text{ and} \\
F_{\lambda_{12} W_1} &: q \mapsto \max(0, 1 - e^{-q/\lambda_{12}})
\end{aligned}$$

denote the cdfs of $\lambda_1 W_1 + \lambda_2 W_2$ and $\lambda_{12} W_1$, where W_1 and W_2 are independent standard exponentials.

By Definition 1 ,

$$K(a_1, a_r, b_1, b_2) - K(\boldsymbol{a}, b_1, b_2) = \mathbb{E}(F_{\lambda_1 W_1 + \lambda_2 W_2}(Z_0 + \Delta) - F_{\lambda_{12} W_1}(Z_0 + \Delta).)$$

The conclusion of the lemma follows on showing

$$F_{\lambda_1 W_1 + \lambda_2 W_2}(Z_0 + \delta) - F_{\lambda_{12} W_1}(Z_0 + \delta) \geq 0$$

for all non-random $\delta \geq 0$, with equality iff $\delta = 0$. Define h by

$$\begin{aligned}
h(\delta) &= \mathbb{E}(F_{\lambda_1 W_1 + \lambda_2 W_2}(Z_0 + \delta) - F_{\lambda_{12} W_1}(Z_0 + \delta)) \\
&= \int_0^\infty (F_{\lambda_1 W_1 + \lambda_2 W_2}(z + \delta) - F_{\lambda_{12} W_1}(z + \delta)) e^{-z} dz \\
&= \frac{\lambda_{12}}{\lambda_{12} + 1} e^{-\delta/\lambda_{12}} - \frac{1}{\lambda_1 - \lambda_2} (\frac{\lambda_1^2}{\lambda_1 + 1} e^{-\delta/\lambda_1} - \frac{\lambda_2^2}{\lambda_2 + 1} e^{-\delta/\lambda_2}).
\end{aligned}$$

By Rolle's theorem and induction, the sum of 3 exponential $h(\delta)$ has at most 2 roots. The function h has a double root at 0, while $h''(0) = 1/\lambda_{12} > 0$. Therefore h increases from 0 in a neighborhood of 0. Moreover $\lambda_{12} > \max(\lambda_1, \lambda_2)$ so that $e^{-\delta/\lambda_{12}}$, which has a positive coefficient, is the dominant term. As a result, $h(\delta) \downarrow 0$ as $\delta \to \infty$, implying $\delta = 0$ is the only root on $[0, \infty)$.

2. Let $\lambda_1 = 1-a_1, \lambda_2 = 1-a_2$, and $\lambda_{12} = 1-a_1a_2 = \lambda_1+\lambda_2-\lambda_1\lambda_2$. Let $\Delta = \sum_{i=1}^r (1-a_i)Z_i$, where $Z_i, i = 0, \ldots, r$, are independent standard exponentials. Then $\Delta \geq 0$ a.s. Similarly to the first part,

$$K(a_1, a_2, \boldsymbol{b}) - K(a_1a_2, \boldsymbol{b}) = \mathbb{E}(F_{\lambda_{12}W_1}(\Delta - Z_0) - F_{\lambda_1 W_1+\lambda_2 W_2}(\Delta - Z_0)),$$

and the conclusion of the lemma follows on showing

$$\mathbb{E}(F_{\lambda_{12}W_1}(\Delta - Z_0) - F_{\lambda_1 W_1+\lambda_2 W_2}(\Delta - Z_0)) \geq 0$$

for all non-random $\delta \geq 0$, with equality iff $\delta = 0$. Define h by

$$\begin{aligned} h(\delta) &= \mathbb{E}(F_{\lambda_1 W_1+\lambda_2 W_2}(Z_0+\delta) - F_{\lambda_{12}W_1}(Z_0+\delta)) \\ &= \frac{\lambda_{12}}{\lambda_{12}-1}e^{-\delta/\lambda_{12}} - \frac{1}{\lambda_1-\lambda_2}\left(\frac{\lambda_1^2}{\lambda_1-1}e^{-\delta/\lambda_1} - \frac{\lambda_2^2}{\lambda_2-1}e^{-\delta/\lambda_2}\right). \end{aligned}$$

The remainder of the proof is the same as in the last part: h has at most 2 roots, $h(0) = h'(0) = 0$, $h''(0) > 0$, and $h(\delta) \downarrow 0$ as $\delta \to \infty$, implying $h(\delta) > 0$ for $\delta > 0$.

□

Theorem 23. *Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}_+^n$. Let $A = \{i : 0 \leq x_i < 1\}$ and $B = \{i : x_i \geq 1\}$. Suppose $\{i,j\} \subset A$ or $\{i,j\} \subset B$, $x_i \neq 1, x_j \neq 1$. Then*

$$K(\boldsymbol{x}_{-i,-j}, x_i x_j) \leq K(\boldsymbol{x}),$$

with equality iff $\boldsymbol{x}_{-i,-j} = \mathbf{1}$.

Proof. The proof is by induction on $\min(|A|,|B|)$, with the cases $\min(|A|,|B|) = 1$ and $\min(|A|,|B|) = 2$ treated separately.

1. $\min(|A|,|B|) = 0$. Then $\boldsymbol{x} = \boldsymbol{x}_A$ or $\boldsymbol{x} = \boldsymbol{x}_B$. If $\boldsymbol{x} = \boldsymbol{x}_A$, then $K(\boldsymbol{x}) = 1 = K(\boldsymbol{x}_{-i,-j}, x_i x_j)$. If $\boldsymbol{x} = \boldsymbol{x}_B$, then $K(\boldsymbol{x}) = 1/\prod_i x_i = K(\boldsymbol{x}_{-i,-j}, x_i x_j)$.

2. $\min(|A|,|B|) = 1$. By induction on $\max(|A|,|B|)$. If $\max(|A|,|B|) = 1$ there is nothing to do.

Suppose that $\max(|A|,|B|) = 2$. Then either $\boldsymbol{x}_A = a$ and $\boldsymbol{x}_B = (b_1, b_2)$, or $\boldsymbol{x}_A = (a_1, a_2)$ and $\boldsymbol{x}_B = b$. In the first case,

$$K(a, b_1, b_2) - K(a, b_1b_2) = p(a, b_1)K(b_1, b_2) + p(b_1, a)K(a_1, b_2) - p(a, b_1b_2)K(b_1b_2) - p(b_1b_2, a)$$

$$= -b_1(b_2-1)\frac{(b_2-1)(1-a)}{b_2-a} + (b_1b_2-1)\frac{b_1(b_2-1)(1-a)}{b_1b_2-a}.$$

If $a = 1$, the above line is 0. Otherwise, dividing through by $1-a$, the last line becomes $p(a, b_1b_2) - p(a, b_2) > 0$, as $t \mapsto p(a,t)$ is strictly increasing when $a \in [0,1)$.

Next supposing $\boldsymbol{x}_A = (a_1, a_2)$ and $\boldsymbol{x}_B = b$,

$$K(a_1, a_2, b) - K(a_1 a_2, b) = p(a_1, b)(p(a_2, b)/b + p(b, a_2)) + p(b, a_1) - p(a_1 a_2, b)/b - p(b, a_1 a_2)$$

$$= \frac{b-1}{b - a_1 a_2} - \frac{b-1}{b-a_1}\frac{b-1}{b-a_2}.$$

If $b = 1$ the above line is 0. Otherwise, dividing through by $b - 1$, the last line becomes $1 - (a_1 + (1 - a_1)a_2)$ which is > 0 for $\{a_1, a_2\} \subset [0, 1)$.

Suppose by way of induction that the conclusion of the theorem holds for some $m \geq 2$, $m = \max(|A|, |B|)$ and $\min(|A|, |B|) = 1$. We consider the case that $|A| = m+1$ and $|B| = 1$, with the case $|A| = 1, |B| = m + 1$ being analogous. Then $\{i, j\} \subset A$, and there is $u \in A$ distinct from i and j. Let $B = \{v\}$. By the inductive hypothesis and the $\min(|A|, |B|) = 0$ case given above,

$$\begin{aligned} K(\boldsymbol{x}) &= K(\boldsymbol{x}_A, \boldsymbol{x}_B) = p(x_u, x_v)K(\boldsymbol{x}_{A/u}, x_v) + p(x_v, x_u)K(\boldsymbol{x}_A) \\ &\geq p(x_u, x_v)K(\boldsymbol{x}_{A/\{u,i,j\}}, x_i x_j, x_v) + p(x_v, x_u)K(\boldsymbol{x}_{A/\{i,j\}}, x_i x_j) \\ &= K(\boldsymbol{x}_{A/\{i,j\}}, x_i x_j, x_v). \end{aligned}$$

Equality holds above iff $K(\boldsymbol{x}_{A/u}, x_v) = K(\boldsymbol{x}_{A/\{u,i,j\}}, x_i x_j, x_v)$ and $K(\boldsymbol{x}_A) = K(\boldsymbol{x}_{A/\{i,j\}}, x_i x_j)$, which holds iff $\boldsymbol{x}_{A/\{i,j\}}$ are all $= 1$ and $\boldsymbol{x}_B = x_v = 1$, by the inductive hypothesis.

3. $\min(|A|, |B|) = 2$. This case is given by Lemma 22 .

Finally, suppose by way of induction that the conclusion of the theorem holds for some $m \geq 2$, and suppose $\min(|A|, |B|) = m + 1$. We give the case that $\{i, j\} \subset A$, with the case that $\{i, j\} \subset B$ being analogous. Let $v \in B$, and let $u \in A$ be distinct from i and j. By the inductive hypothesis,

$$\begin{aligned} K(\boldsymbol{x}) &= K(\boldsymbol{x}_A, \boldsymbol{x}_B) = p(x_u, x_v)K(\boldsymbol{x}_{A/u}, \boldsymbol{x}_B) + p(x_v, x_u)K(\boldsymbol{x}_A, \boldsymbol{x}_{B/v}) \\ &\geq p(x_u, x_v)K(\boldsymbol{x}_{A/\{u,i,j\}}, x_i x_j, \boldsymbol{x}_B) + p(x_v, x_u)K(\boldsymbol{x}_{A/\{i,j\}}, x_i x_j, \boldsymbol{x}_{B/v}) \\ &= K(\boldsymbol{x}_{A/\{i,j\}}, x_i x_j, \boldsymbol{x}_B). \end{aligned}$$

Equality holds iff $K(\boldsymbol{x}_{A/u}, \boldsymbol{x}_B) = K(\boldsymbol{x}_{A/\{u,i,j\}}, x_i x_j, \boldsymbol{x}_B)$ and $K(\boldsymbol{x}_A, \boldsymbol{x}_{B/v}) = K(\boldsymbol{x}_{A/\{i,j\}}, x_i x_j, \boldsymbol{x}_{B/v})$, which holds iff $\boldsymbol{x}_{A/\{i,j\}}$ and $\boldsymbol{x}_B$ are all $= 1$, by the inductive hypothesis. $\square$

2 Interpretations of K

2.1 Stochastic process

2.1.1 Martingale

Recursion (8) invites a martingale formulation of K. Fix $r \geq 0, s \geq 0, \boldsymbol{a} \in [0, 1]^r, \boldsymbol{b} \in [0, \infty)^s$. Define discrete-time integer-valued processes $I(t)$ on $0, \ldots, r$, and $J(t)$ on $0, \ldots, s$ as follows.

For $t = 0$, $I(0) = J(0) = 0$. For $t > 0$,

$$(I(t+1), J(t+1)) = \begin{cases} (I(t), J(t)) & \text{w.p. } 1 \text{ if } I(t) = r \text{ or } J(t) = s \\ (I(t)+1, J(t)) & \text{w.p. } p(a_{I(t)}, b_{J(t)}) \text{ if } I(t) < r \text{ and } J(t) < s \\ (I(t), J(t)+1) & \text{w.p. } p(b_{J(t)}, a_{I(t)}) \text{ if } I(t) < r \text{ and } J(t) < s. \end{cases}$$

Then $K((\boldsymbol{a})^r_{I(t)}, (\boldsymbol{b})^s_{J(t)})$ is a bounded discrete-time martingale. When $I(t) = r$ or $J(t) = s$, $K((\boldsymbol{a})^r_{I(t+1)}, (\boldsymbol{b})^s_{J(t+1)}) = K((\boldsymbol{a})^r_{I(t)}, (\boldsymbol{b})^s_{J(t)})$ deterministically, and when both $I(t) < r$ and $J(t) < s$,

$$\begin{aligned} &\mathbb{E}(K((\boldsymbol{a})^r_{I(t+1)}, (\boldsymbol{b})^s_{J(t+1)}) | K((\boldsymbol{a})^r_{I(t)}, (\boldsymbol{b})^s_{J(t)})) \\ &= p(a_{I(t)}, b_{J(t)}) K((\boldsymbol{a})^r_{I(t+1)}, (\boldsymbol{b})^s_{J(t)}) + p(b_{J(t)}, a_{I(t)}) K((\boldsymbol{a})^r_{I(t)}, (\boldsymbol{b})^s_{J(t+1)}) \\ &= K((\boldsymbol{a})^r_{I(t)}, (\boldsymbol{b})^s_{J(t)}). \end{aligned}$$

Define the bounded stopping time T by the event $I(t) = r$ or $J(t) = s$. Then

$$K(\boldsymbol{a}, \boldsymbol{b}) = \mathbb{E} K((\boldsymbol{a})^r_{I(T)}, (\boldsymbol{b})^s_{J(T)}).$$

2.1.2 Markov chain and gambling interpretation

Let $a_1, a_2, \ldots$ be a sequence in $[0, 1]$ and $b_1, b_2, \ldots$ a sequence in $[1, \infty)$. Define a discrete-time process R_t on $\mathbb{Z}$ with increments ± 1 as follows. At $t = 0$, $R_0 = 0$. If at time t, R has taken U up steps and D down steps then

$$R_{t+1} = \begin{cases} R_t + 1 & \text{w.p. } p(b_{U+1}, a_{D+1}) \\ R_t - 1 & \text{w.p. } p(a_{D+1}, b_{U+1}). \end{cases}$$

Since $U + D = t$ and $U - D = r_t$, $U = (R_t + t)/2$ and $D = (t - R_t)/2$, and so R_t is a time- and space-inhomogeneous Markov chain. Given natural numbers r and s, define the "payoff" of the sample path as follows. Let $\boldsymbol{a} = (a_i)^r_{i=1}$ and $\boldsymbol{b} = (b_i)^s_{i=1}$. Let T_U and T_D denote the hitting times for $R_t = -t + 2s$ and $R_t = t - 2r$. When $T_U = T_D \wedge T_U$, $R_{T_D \wedge T_U}$ has made s up moves and $T_U - s$ down moves, and the payoff is defined to be $K(a_{D+1}, \ldots, a_r) = 1$. When $T_D = T_D \wedge T_U$, $R_{T_D \wedge T_U}$ has made r down moves and $T_D - r$ up moves, and the payoff is defined to be $K(b_{U+1}, \ldots, b_s) = (\prod_{i=U+1}^s b_i)^{-1}$. Then $K(\boldsymbol{a}, \boldsymbol{b})$ is the expected payoff.

Player A begins with chips $a_1, \ldots, a_r$ and player B with chips $b_1, \ldots, b_s$. The object of the game is to bankrupt the other player. In each of a sequence of rounds the two players put up a chip and a coin is flipped with probability $p(b_U, a_D) = \frac{1-a_D}{b_U - a_D}$ of landing heads, where D and U represent the number of rounds A and B, respectively, have previously won. The value of a chip staked in a round is therefore related to its relative distance from 1. If the coin does land heads, A wins the current round, and B loses his chip; otherwise B wins and A loses his chip. Hitting $R_t = -t + 2s$ represents bankruptcy for B, $R_t = t - 2r$ for A. At bankruptcy, the other player cashes in his remaining chips, and K is the valuation, e.g., $K((\boldsymbol{b})^s_{i=U+1}) = (\prod_{i=U+1}^s b_i)^{-1}$ if B wins. $K(\boldsymbol{a}, \boldsymbol{b})$ represents the expected payoff for the winner. See Fig. 1.

The symmetry of K says that strategy doesn't matter. A player can play chips in any order without affecting the expected payoff.

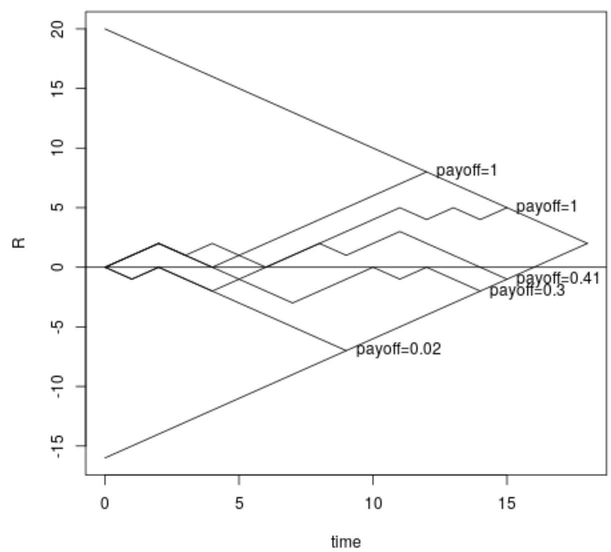

Figure 1: Interpretation of $K(\boldsymbol{a}, \boldsymbol{b})$ as the expected payoff in a game. Five games are depicted, where player A begins with the 8 chips $\boldsymbol{a} = (0.59, 0.01, 0.29, 0.28, 0.81, 0.26, 0.72, 0.91)$ and B with 10 chips $\boldsymbol{b} = (1.05, 1.84, 3.68, 1.29, 1.4, 1.43, 1.34, 2.05, 1.06, 1.13)$.

2.2 Interpolating polynomials

The formula (3) may be interpreted as the value of a polynomial interpolant and, relatedly, divided differences.

Given $f : \mathbb{R} \to \mathbb{R}$ and distinct reals $x_1, \ldots, x_n$, let $L[f; x_1, \ldots, x_n] : \mathbb{R} \to \mathbb{R}$ denote the unique polynomial of degree $n-1$ passing through $(x_1, f(x_1)), \ldots, (x_n, f(x_n))$. By comparison of the Lagrange form of this polynomial to (3),

$$K(x_1, \ldots, x_n) = -\frac{L[t \mapsto k_0(t)(t-1)^n; x_1, \ldots, x_n](1)}{\prod_{i=1}^n (1 - x_i)}, \tag{16}$$

That is, $K(x_1, \ldots, x_n)$ is the interpolating polynomial of $t \mapsto k_0(t)(t-1)^n$ at nodes $x_1, \ldots, x_n$, evaluated at 1, after dividing by $-\prod_{i=1}^n (1 - x_i)$.

Next, let $[x_1, \ldots, x_n; f]$ denote the divided difference of order $n-1$ of f at nodes $x_1, \ldots, x_n$, that is, the leading coefficient of the Newton form of the unique polynomial of degree $n-1$ through $(x_1, f(x_1)), \ldots, (x_n, f(x_n))$ (de Boor, 2005). Then

$$K(x_1, \ldots, x_n) = [x_1, \ldots, x_n; t \mapsto k_0(t)(1-t)^{n-1}]. \tag{17}$$

The interpretation of K using interpolating polynomials opens up many results, including an integral representation, the derivative of K with respect to a node, and confluent formulas (de Boor, 2005). See also the proof of Theorem 5 where the relation to Lagrange polynomials is applied. However, many of the results for interpolating polynomials do not carry over. The reason is that in the interpretations (16) and (17) the function to be interpolated, $t \mapsto k_0(t)(t-1)^n$ or $t \mapsto k_0(t)(1-t)^{n-1}$, depends on the number n of nodes used for interpolation.

As a consequence results for K resemble but still differ from the corresponding results for interpolating polynomials. For example, compare the explicit formula for $K(x_1, \ldots, x_n)$ given by (3),

$$\sum_{i=1}^{n} k_0(x_i) \prod_{j:j\neq i} \frac{x_i - 1}{x_i - x_j}.$$

with the formula for the Lagrange interpolant through $(x_i, k_0(x_i)), i = 1, \ldots, n$,

$$\sum_{i=1}^{n} k_0(x_i) \prod_{j:j\neq i} \frac{x_j - 1}{x_j - x_i}.$$

Similarly, given distinct i, j, $1 \leq i \leq n, 1 \leq j \leq n$, compare the recursion formula (8) for K,

$$K(\boldsymbol{x}) = \frac{x_i - 1}{x_i - x_j} K(\boldsymbol{x}_{-j}) + \frac{x_j - 1}{x_j - x_i} K(\boldsymbol{x}_{-i})$$

with the corresponding formula given by Neville's algorithm for computing the value at 1 of the polynomial interpolating $(x_k, k_0(x_k)), k = 1, \ldots, n$,

$$K(\boldsymbol{x}) = \frac{x_i - 1}{x_i - x_j} K(\boldsymbol{x}_{-i}) + \frac{x_j - 1}{x_j - x_i} K(\boldsymbol{x}_{-j})$$

or the divided differences recursion formula,

$$[\boldsymbol{x}; K] = \frac{1}{x_i - x_j} [\boldsymbol{x}_{-j}; K] + \frac{1}{x_j - x_i} [\boldsymbol{x}_{-i}; K].$$

3 Extensions

Let Θ_0 denote the set of all finite products of nonnegative two-point distributions with mean 1, and let Θ_1 denote the set of all finite products of nonnegative two-point distributions at least one of which has mean strictly greater than 1.

We first consider generalizations of K via Theorem 6 allowing the values on leaf nodes to be arbitrary. The resulting functions need not be symmetric so in the following definition we use one example of a choice function on the indices of the function arguments.

Definition 24. *Given a function $f : [0,1]^{(\mathbb{N})} \cup [1,\infty)^{(\mathbb{N})} \to \mathbb{R}$. extend its domain to $\mathbb{R}_+^{(\mathbb{N})}$ by recursion 8. That is, for $\boldsymbol{x} \in [0,\infty)^{(\mathbb{N})}$, $|\boldsymbol{x}| \geq 2$, $\boldsymbol{x} \notin [0,1]^{(\mathbb{N})} \cup [1,\infty)^{(\mathbb{N})}$, let $i_0 = \min\{i : x_i \leq 1\}, j_0 = \min\{j : x_j \geq 1\}$, and*

$$f(\boldsymbol{x}) = p(x_{i_0}, x_{j_0}) f(\boldsymbol{x}_{-i_0}) + p(x_{j_0}, x_{i_0}) f(\boldsymbol{x}_{-j_0}).$$

Let $\mathbb{K}$ denote the class of all functions so defined, as the values on the leaf nodes vary.

The function K is admissible in $\mathbb{K}$ in the following sense.

Theorem 25. *Given $f \in \mathbb{K}$,* *1. If $f(\boldsymbol{x}) < K(\boldsymbol{x})$ for a leaf node $\boldsymbol{x}$, then for some $\alpha \in [0,1]$ there exists $\theta_0 \in \Theta_0$ such that $\mathbb{P}_{\theta_0}(f(\boldsymbol{x}) \leq \alpha) > \mathbb{P}_{\theta_0}(K(\boldsymbol{x}) \leq \alpha)$.*

2. If $f(\boldsymbol{x}) > K(\boldsymbol{x})$ for a leaf node $\boldsymbol{x}$, then for some $\alpha \in [0,1]$ there exists $\theta_1 \in \Theta_1$ such that $\mathbb{P}_{\theta_1}(f(\boldsymbol{x}) \leq \alpha) < \mathbb{P}_{\theta_1}(K(\boldsymbol{x}) \leq \alpha)$.

Proof. 1. (a) Suppose there is $\boldsymbol{a} \in [0,1]^n$ such that $f(\boldsymbol{a}) < K(\boldsymbol{a}) = 1$. Let $\boldsymbol{X} \sim TP(\boldsymbol{a}, b\mathbf{1}, \mathbf{1})$, with b to be specified. For $\alpha \in [f(\boldsymbol{a}), 1)$, $\mathbb{P}(f(\boldsymbol{X}) \leq \alpha) \geq \prod_{i=1}^n \mathbb{P}(X_i = a_i) = \prod_{i=1}^n p(a_i, b, 1)$ and $\prod_{i=1}^n p(a_i, b, 1) \to 1$ as $b \to \infty$. So for large enough b, $\mathbb{P}(f(\boldsymbol{X}) \leq \alpha) > \alpha$.

(b) Suppose there is $\boldsymbol{b} \in [1,\infty)^n$ such that $f(\boldsymbol{b}) < K(\boldsymbol{b}) = 1/\prod_{i=1}^n b_i$. Let $\boldsymbol{X} \sim TP(\mathbf{0}, \boldsymbol{b}, \mathbf{1})$. For $\alpha \in [f(\boldsymbol{b}), 1/\prod_{i=1}^n b_i)$, $\mathbb{P}(f(\boldsymbol{X}) \leq \alpha) \geq \prod_{i=1}^n p(b_i, 0, 1) = 1/\prod_{i=1}^n b_i$, where $1/\prod_{i=1}^n b_i$ is $> \alpha = f(\boldsymbol{b})$.

2. (a) Suppose there is $\boldsymbol{a} \in [0,1]^n$ such that $f(\boldsymbol{a}) > K(\boldsymbol{a}) = 1$. Then $\mathbb{P}(f(\boldsymbol{X}) \leq 1) \leq 1 - \prod_{i=1}^n \mathbb{P}(X_i = a_i) < 1 = \mathbb{P}(K(\boldsymbol{X}) \leq 1)$ holds for any non-degenerate $X_i \in \Theta_0 \cup \Theta_1$, $i = 1, \ldots, n$. Therefore 2. holds with $\alpha = 1$.

(b) Suppose there is $\boldsymbol{b} \in [1,\infty)]^n$ such that $f(\boldsymbol{b}) > K(\boldsymbol{b}) = 1/\prod_{i=1}^n b_i$. Let $\boldsymbol{X} \sim TP(\mathbf{0}, \boldsymbol{b}, \boldsymbol{\mu})$ with $\boldsymbol{\mu} \in (1,\infty)^n$ variable. Then as $\boldsymbol{\mu} \uparrow \boldsymbol{b}$ element-wise, $\mathbb{P}(f(\boldsymbol{X}) \leq 1/\prod_{i=1}^n b_i) < 1 - \prod_{i=1}^n \mathbb{P}(X_i = b_i) \to 0$ whereas $\mathbb{P}(K(\boldsymbol{X}) \leq 1/\prod_{i=1}^n b_i) = \prod_{i=1}^n \mathbb{P}(X_i = b_i) \to 1$. Therefore 2. holds with $\alpha = 1/\prod_{i=1}^n b_i$.

□

We next focus on a sub-class of symmetric functions, namely, those elements of $\mathbb{K}$ that, like K, are multiplicative on leaf nodes.

Theorem 26. *As in Theorem 6, given a function $f_0 : [0,\infty) \to \mathbb{R}$, define*

$$f : [0,\infty)^{(\mathbb{N})} \to \mathbb{R} : (x_1, \ldots, x_n) \to \sum_{i=1}^n f_0(x_i) \prod_{j:j\neq i} \frac{x_i - 1}{x_i - x_j}. \tag{18}$$

Then $f(\boldsymbol{x}) = \prod_{i=1}^n f(x_i)$ for all $\boldsymbol{x} \in [0,1]^n \cup [1,\infty)^n$ and $n \geq 1$, if and only if for all pairs of real numbers x, y,

$$f(x)f(y) = \frac{(y-1)f(y) - (x-1)f(x)}{y - x}. \tag{19}$$

Then either $f \equiv 0$ or there exists a pair of reals (q_1, q_2) such that

$$f_0(x) = \begin{cases} \frac{1}{q_1 + (1-q_1)x} & x \in [0,1) \\ \frac{1}{q_2 + (1-q_2)x} & x \geq 1. \end{cases} \tag{20}$$

Proof. The necessity of the functional equation follows from (8) applied to (18), since $f(x,y) = p(x,y)f(y) + p(y,x)f(x)$ is the Newton quotient (19). Conversely, suppose by way of induction the result holds for some $n \geq 1$, and let $\boldsymbol{x} = (x_1, \ldots, x_{n+1})$ be a leaf node.

$$
\begin{aligned}
f(x_1, \ldots, x_{n+1}) &= p(x_1, x_{n+1}) f(\boldsymbol{x}_{-1}) + p(x_{n+1}, x_1) f(\boldsymbol{x}_{-(n+1)}) \\
&= p(x_1, x_{n+1}) \prod_{i=2}^{n+1} f(x_i) + p(x_{n+1}, x_1) \prod_{i=1}^{n} f(x_i) \\
&= (p(x_1, x_{n+1}) f(x_{n+1}) + p(x_{n+1}, x_1) f(x_1)) \prod_{i=2}^{n} f(x_i) \\
&= f(x_1, x_{n+1}) \prod_{i=2}^{n} f(x_i) \\
&= \prod_{i=1}^{n+1} f(x_i).
\end{aligned}
$$

To find the solutions to the functional equation (19) we show more generally that if a function $f : D \to \mathbb{R}$ satisfies

$$
f(x)f(y) = \frac{(y-1)f(y) - (x-1)f(x)}{y-x}, \qquad x \neq y \tag{21}
$$

on a domain $D \subset \mathbb{R}$ with at least 3 points including a point $y \in D$ satisfying $y \neq 1, f(y) \neq 0, f(y) \neq 1$, then for some $q \in \mathbb{R}$,

$$
f(x) = \frac{1}{q + (1-q)x}.
$$

Equation (20) follows by taking $D = (0, 1)$ and $D = [1, \infty)$. Suppose f isn't the 0 function on D. Let $x, y \in D$ be distinct, $(y-1)f(y) \neq 0$, and $y(f(y) - 1 + (1 - f(y))x) \neq 0$, which can be found under our assumptions. Then the functional equation (21) may be re-written as

$$
\begin{aligned}
f(x) &= \frac{(y-1)f(y)}{(yf(y)-1) + x(1-f(y))} = \frac{1}{c_1 + c_2 x} \\
c_1 &= \frac{yf(y)-1}{(y-1)f(y)}, c_2 = \frac{1-f(y)}{(y-1)f(y)}.
\end{aligned}
$$

Substituting this expression for f into the functional equation, for $x \neq y$, yields

$$
\frac{1}{c_1 + c_2 x} \frac{1}{c_1 + c_2 y} = \frac{(c_1 + c_2)(y-x)}{(c_1 + c_2 x)(c_1 + c_2 y)},
$$

implying $c_1 + c_2 = 1$. $\square$

Definition 27. *For a pair of reals q_1, q_2,*

$$
K_{q_1, q_2} : [0, \infty)^{(\mathbb{N})} \to \mathbb{R} : (x_1, \ldots, x_n) \to \sum_{i=1}^{n} f_0(x_i) \prod_{j : j \neq i} \frac{x_i - 1}{x_i - x_j}.
$$

with f_0 as in (20).

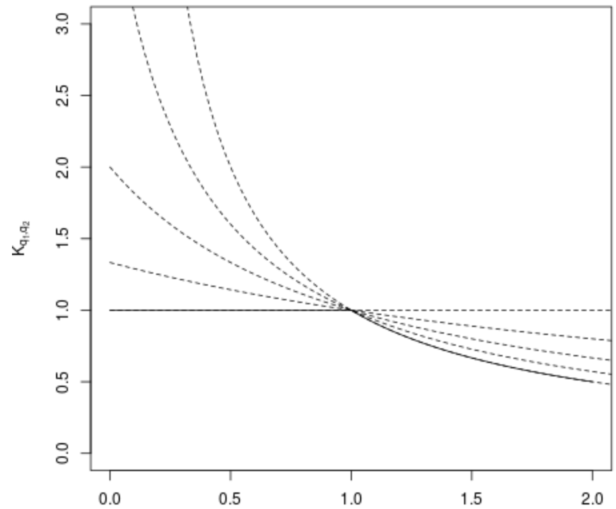

Figure 2: The functions K_{q_1,q_2} for some values $(q_1, q_2) \in [0,1]^2$, and the lower envelope given by $K = K_{1,0}$.

Several of the basic properties of K carry over to K_{q_1,q_2} when (q_1, q_2) belongs to the unit square.

Theorem 28. *Suppose* $(q_1, q_2) \in [0,1]^2$. *Then*

1. $0 \le k(x_1, \dots, x_n) \le 1$.
2. $K(x_1, \dots, x_n)$ *is symmetric in* $x_1, \dots, x_n$.
3. $K(x_1, \dots, x_n)$ *is non-increasing in each* x_i*, and strictly decreasing if* $x_i > 1$ *for some* i.
4. $K(x_1, \dots, x_n, y) \gtreqqless K(x_1, \dots, x_n)$ *according as* $y \lesseqqgtr 1$.

The function $K = K_{1,0}$ corresponds to $(q_1, q_2) = (1, 0)$ and is a lower envelope for $\{K_{q_1,q_2} : (q_1, q_2) \in [0,1]^2\}$. See Fig. 2. The likelihood ratio statistic (38) of Wang and Zhao (2003) can be situated in this family, which clarifies the result of Gaffke (2005) that $K \le R$. The likelihood ratio R is obtained by minimizing over the diagonal in $[0,1] \times [0,1]$ whereas K is obtained by minimizing over the unit square.

Corollary 29 (Gaffke (2005, Theorem 2.1))**.** *For all* $\boldsymbol{x} \in [0,\infty)^{(\mathbb{N})}$, $K(\boldsymbol{x}) \le R(\boldsymbol{x})$.

Proof. For any $(q_1, q_2) \in [0,1]^2$

$$K_{1,0}(\boldsymbol{x}) \le K_{q_1,q_2}(\boldsymbol{x})$$

holds whenever $\boldsymbol{x} = x$ is a nonnegative real number, and so by the multiplicative property the inequality also holds whenever $\boldsymbol{x}$ is any leaf node, and so by Theorem 12 the inequality also holds for any $\boldsymbol{x} \in \mathbb{R}_+^{(\mathbb{N})}$. Therefore, for any $\boldsymbol{x} \in \mathbb{R}_+^{(\mathbb{N})}$,

$$
\begin{aligned}
R &= \inf_{q\in[0,1]} \prod_{i=1}^{n} \frac{1}{q + (1-q)x_i} \\
&= \inf_{q\in[0,1]} K_{q,q}(\boldsymbol{x}) \\
&\geq \inf_{(q_1,q_2)\in[0,1]^2} K_{q_1,q_2}(\boldsymbol{x}) \\
&= K_{1,0}(\boldsymbol{x}) = K(\boldsymbol{x}).
\end{aligned}
$$

□

Theorem 30. *1. If $q_1 \notin [0,1]$ or $q_2 \notin [0,1]$, there is $\alpha \in [0,1]$ and $\mathbb{P}_0 \in \Theta_0$ such that for any n, $|\boldsymbol{X}| = n$, $\mathbb{P}_0(K_{q_1,q_2}(\boldsymbol{X}) \leq \alpha) > \alpha$.*

2. If $q_1 \in [0,1]$ and $q_2 \in [0,1]$, then for any $\alpha \in (0,1)$,

$$
\sup_{\mathbb{P}_1\in\Theta_1} \left(\mathbb{P}_1(K(\boldsymbol{X}) \leq \alpha) - \mathbb{P}_1(K_{q_1,q_2}(\boldsymbol{X}) \leq \alpha)\right) \geq 0. \tag{22}
$$

3. If $n = 1$, there is $\alpha \in [0,1]$ such that the inequality is strict.

4. If $n \geq 2$, the inequality (22) is strict.

Proof. 1. If $q_1 \notin [0,1]$ then there is some $x \in [0,1)$ such that $K_{q_1,q_2}(x) < K(x)$. Then with $X \sim TP(x,b,1)$, $\mathbb{P}(K_{q_1,q_2}(X) \leq \alpha) \geq p(x,b) = \frac{b-1}{b-x}$ is $> \alpha$ iff $b > \frac{1-\alpha x}{1-\alpha}$. If $q_2 \notin [0,1]$ then there is some $x \in [1,\infty)$ such that $K_{q_1,q_2}(x) < K(x)$. Then with $X \sim TP(0,x,1)$, $\alpha = \max(K_{q_1,q_2}(x), 0)$, $\mathbb{P}(K_{q_1,q_2}(X) \leq \alpha) \geq p(x,0) = 1/x$ which is $> \alpha$. Since $K_{q_1,q_2}(\boldsymbol{x},1) = K_{q_1,q_2}(\boldsymbol{x})$ for all $(q_1,q_2) \in [0,1]^2$ and $\boldsymbol{x} \in \mathbb{R}_+^{(\mathbb{N})}$, the $n = 1$ case can be extended to $n > 1$ by taking $X_i, i = 2, \ldots, n$, to be point masses at 1.

2. If $(q_1,q_2) \in [0,1]^2$, then for any $\boldsymbol{x} \in \mathbb{R}_+^{(\mathbb{N})}$, $K_{1,0}(\boldsymbol{x}) \leq K_{q_1,q_2}(\boldsymbol{x})$, so $\mathbb{P}(K_{1,0}(\boldsymbol{X}) \leq \alpha) \geq \mathbb{P}(K_{q_1,q_2}(\boldsymbol{X}) \leq \alpha)$ for any $\mathbb{P}$.

3. As above, if $q_1 \notin [0,1]$ or $q_2 \notin [0,1]$ there is $x \in [0,1)$ or $x \in [1,\infty)$, respectively, such that $K_{q_1,q_2}(x) < K(x)$, and so this result follows from Theorem 25.

4. Suppose first that $q_1 \in (0,1]$, so that $K(x) < K_{q_1,0}(x)$ for all $x \in (0,1)$. For $a \in (0,1), b \in (1,\infty), \alpha \in (0,1), \mu > 1$, consider the inequalities

$$
K(a,b) \leq \alpha < K_{q_1,0}(a,b) \tag{23}
$$

which may be re-written as

$$
1 \leq \frac{\alpha - p(a,b)/b}{1 - p(a,b)} < K_{q_1,0}(a,b).
$$

The middle term is α at $b = 1$ and $\to \infty$ as $b \to \infty$, and so by continuity there is some $b = b^* \in [1, \infty)$ for which (23) holds. Then with X_1, X_2 both $\sim TP(a, b^*, \mu)$ for any $0 \le a \le 1 < \mu < b^*$,

$$
\begin{aligned}
\mathbb{P}(K(X_1, X_2) \le \alpha) &= \mathbb{P}((X_1, X_2) = (a, b^*)) + \mathbb{P}((X_1, X_2) = (b^*, a)) + \mathbb{P}((X_1, X_2) = (b^*, b^*)) \\
&> \mathbb{P}((X_1, X_2) = (b^*, b^*)) \\
&= \mathbb{P}(K_{q_1,0}(X_1, X_2) \le \alpha).
\end{aligned}
$$

Since $K_{q_1,0}(x) \ge K_{q_1,q_2}(x)$ for all $q_2 \in [0, 1]$ and $x \ge 0$, $\mathbb{P}(K_{q_1,q_2}(X) \le \alpha) \le \mathbb{P}(K_{q_1,0}(X) \le \alpha) \le \mathbb{P}(K(X) \le \alpha)$ with $\mathbb{P} \in \Theta_1$.

Next suppose that $q_2 \in [0, 1)$, so that $K(x) < K_{1,q_2}(x)$ for all $x > 1$. Let $b = 1/\alpha$, $a < 1 < \mu \le b$, and $X \sim TP(a, b, \mu)$. Then $\mathbb{P}(K(X) \le \alpha) = \mathbb{P}(X = b) = p(b, a, \mu) > 0$ whereas $\mathbb{P}(K_{1,q_2}(X) \le \alpha) = 0$, since $K_{1,q_2}(b) > K(b) = \alpha$ and $K_{1,q_2}(a) \ge K_{1,q_2}(b)$ by anti-monotonicity (Theorem 28). Since $K_{1,q_2}(x) \le K_{q_1,q_2}(x)$ for all $q_1 \in [0, 1]$ and $x \ge 0$, $\mathbb{P}(K_{q_1,q_2}(X) \le \alpha) \le \mathbb{P}(K_{1,q_2}(X) \le \alpha) \le \mathbb{P}(K(X) \le \alpha)$ with $\mathbb{P} \in \Theta_1$.

Finally, the $n = 1$ and $n = 2$ cases given here can be extended to greater values of n by taking additional X_i to be point masses, as mentioned above.

□

4 Asymptotic results

Theorem 31. *Given $X_i \sim TP(a_i, b_i, 1)$, $i = 1, 2, \ldots$, assume as $n \to \infty$,*

1. $\sum_{i=1}^{\infty} ((1 - a_i)(b_i - 1)(2 - a_i - b_i)^2)/i^2 < \infty$
2. $\frac{1}{n} \sum_{i=1}^{n} (1 - a_i)(b_i - 1) \not\to 0$
3. $\frac{\max_{1 \le i \le n} (1-a_i)(b_i-1)}{\sum_{i=1}^{n} (1-a_i)(b_i-1)} \to 0.$

Then as $n \to \infty$, $K(X_1, \ldots, X_n)$ converges in distribution to a standard uniform random variable.

Proof. We make only superficial changes to the proof of Gaffke (2005), which applied to the case of identically distributed X_i.

By Kolmogorov's SLLN, $\sum_{i=1}^{\infty} \mathrm{Var}((X_i - 1)^2)/i^2 < \infty$ implies

$$\frac{1}{n} \sum_{i=1}^{n} ((X_i - 1)^2 - \mathbb{E}((X_i - 1)^2)) \to 0 \text{ a.s.}$$

Since $\mathrm{Var}((X_i - 1)^2) = (1 - a_i)(b_i - 1)(2 - a_i - b_i)^2$, Assumption 1 provides the required hypothesis.

We next show that $\frac{1}{n}\max_{1\le i\le n}(X_i-1)^2 \to 0$ a.s. If $(X_i-1)^2$ is bounded on a sample path, the result follows, otherwise let $n_k, k=1,2,\ldots,$ denote the increasing sequence of record times. Then

$$\begin{aligned}
&\frac{1}{n_k}(X_{n_k}-1)^2 + (\frac{1}{n_k} - \frac{1}{n_k-1}\sum_{i=1}^{n_k-1}((X_i-1)^2 - \mathbb{E}(X_i-1)^2) \\
&= \frac{1}{n_k}\sum_{i=1}^{n_k}((X_i-1)^2 - \mathbb{E}(X_i-1)^2) - \frac{1}{n_k-1}\sum_{i=1}^{n_k-1}((X_i-1)^2 - \mathbb{E}(X_i-1)^2) \\
&\to 0 \text{ a.s.}
\end{aligned}$$

implies $\frac{1}{n_k}(X_{n_k}-1)^2 \to 0$ a.s. Next,

$$\frac{\max_{1\le i\le n}(X_i-1)^2}{\sum_{i=1}^n (X_i-1)^2} = \frac{\max_{1\le i\le n}(X_i-1)^2}{\sum_{i=1}^n \mathbb{E}(X_i-1)^2}\frac{\sum_{i=1}^n \mathbb{E}(X_i-1)^2}{\sum_{i=1}^n (X_i-1)^2}$$

where Assumption 2 and the SLLN from the first step imply the second factor in the last line converges to 1 a.s. Another application implies

$$\frac{\max_{1\le i\le n}(X_i-1)^2}{\sum_{i=1}^n \mathbb{E}(X_i-1)^2} = \frac{\frac{1}{n}\max_{1\le i\le n}(X_i-1)^2}{\frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i-1)^2} \to 0 \text{ a.s.}$$

Therefore,

$$\frac{\max_{1\le i\le n}(X_i-1)^2}{\sum_{i=1}^n (X_i-1)^2} \to 0 \text{ a.s.}$$

Let $Z_1, Z_2, \ldots$ be independent standard exponential random variables. Along almost any sample path $x_1, x_2, \ldots,$ of $X_1, X_2, \ldots,$ the assumptions of the Lindeberg-Feller CLT for the sequence $(x_1-1)Z_1, (x_2-1)Z_2, \ldots,$ are satisfied, so that

$$\frac{\sum_{i=1}^n (x_i-1)Z_i - \sum_{i=1}^n (x_i-1)}{(\sum_{i=1}^n (x_i-1)^2)^{1/2}}$$

converges in distribution to a standard normal random variable. Using Definition 1,

$$|K(x_1,\ldots,x_n) - \Phi(-\frac{\sum_{i=1}^n (x_i-1)}{(\sum_{i=1}^n (x_i-1)^2)^{1/2}})| \to 0.$$

Finally, Assumptions 2, implying $\sum \sigma_i^2 \to \infty$, and 3 are the Lindeberg-Feller conditions for the sequence $X_1-1, X_2-1, \ldots,$ so

$$\frac{\sum_{i=1}^n (X_i-1)}{(\sum_{i=1}^n (X_i-1)^2)^{1/2}}$$

converges in distribution to a standard normal random variable. Since its negation must also then converge to a standard normal,

$$\Phi(-\frac{\sum_{i=1}^n (X_i-1)}{(\sum_{i=1}^n (X_i-1)^2)^{1/2}})$$

converges in distribution to a standard uniform random variable, and so also must $K(X_1,\ldots,X_n)$. $\square$

5 Special cases of Gaffke's strong conjecture and reductions

Theorem 32 (Gaffke (2005, Lemma 4.1))**.** *Suppose $n = 1$ or $n = 2$. Then* (1) *holds.*

Proof. We exhibit the cdf of K directly and show that it is stochastically larger than a standard uniform cdf. Suppose $n = 1$. Then $K(\boldsymbol{X}) = K(X_1) = 1$ w.p. $p(a_1, b_1)$ and $K(\boldsymbol{X}) = 1/b_1$ w.p. $p(b_1, a_1)$, so

$$\mathbb{P}(K(\boldsymbol{x}) \leq \alpha) = \begin{cases} 0, & \alpha \in [0, 1/b_1) \\ p(b_1, a_1), & \alpha \in [1/b_1, 1) \\ 1, & \alpha = 1. \end{cases}$$

Since $p(b_1, a_1) \leq 1/b_1$, $\mathbb{P}(K(\boldsymbol{x}) \leq \alpha) \leq \alpha$ for all α.

Suppose $n = 2$. The state space of $\boldsymbol{X} = (X_1, X_2)$ consists of $(b_1, b_2), (a_1, b_2), (b_1, a_2), (a_1, a_2)$. We assume that $K(a_1, b_2) \leq K(b_1, a_2)$, with the case of the opposite inequality being analogous. The cdf of $K(X_1, X_2)$ is then

$$\mathbb{P}(K(X_1, X_2) \leq \alpha) = \begin{cases} 0, & \alpha \in [0, 1/(b_1 b_2)) \\ p(b_1, a_1)p(b_2, a_2), & \alpha \in [1/(b_1 b_2), K(a_1, b_2)) \\ p(b_2, a_2), & \alpha \in [K(a_1, b_2), K(b_1, a_2)) \\ p(b_2, a_2) + p(b_1, a_1)p(a_2, b_2), & \alpha \in [K(b_1, a_2), 1) \\ 1 & \alpha = 1. \end{cases}$$

We apply the monotonicity properties of p, namely $p(a, b)$ is increasing in each argument and $p(b, a)$ is decreasing in each argument, when $a \in [0, 1]$ and $b \geq 1$. This monotonicity implies $p(b_1, a_1)p(b_2, a_2) \leq 1/(b_1 b_2)$, so that the cdf lies below α through the second interval of constancy given above. In the third interval the cdf is $p(b_2, a_2) \leq 1/b_2 \leq p(a_1, b_2)/b_2 + p(b_2, a_1) = K(a_1, b_2)$, the second inequality following since the rhs is a convex combination of $1/b_2$ and 1.

Consider finally the fourth interval, where the cdf is $p(b_2, a_2) + p(b_1, a_1)p(a_2, b_2)$ and $\alpha \geq K(b_1, a_2) = p(b_1, a_2) + p(a_2, b_1)/b_1$. Suppose first $b_1 \leq b_2$. Then $p(b_2, a_2) + p(b_1, a_1)p(a_2, b_2) \leq p(b_2, a_2) + p(a_2, b_2)/b_1$. Since $b_1 \leq b_2$ implies $p(a_2, b_2) \geq p(a_2, b_1)$, comparing the two convex combinations, $p(b_2, a_2) + p(a_2, b_2)/b_1 \leq p(b_1, a_2) + p(a_2, b_1)/b_1 = K(b_1, a_2)$. Suppose next that $b_1 > b_2$, so that $p(a_1, b_1) > p(a_1, b_2)$. Then rewriting the cdf value, $p(b_2, a_2) + p(b_1, a_1)p(a_2, b_2) = p(b_1, a_1) + p(b_2, a_2)p(a_1, b_1) \leq p(b_1, a_1) + p(a_1, b_1)/b_2 \leq p(b_2, a_1) + p(a_1, b_2)/b_2 = K(a_1, b_2) \leq K(b_1, a_2)$. Therefore, $\mathbb{P}(K(\boldsymbol{x}) \leq \alpha) \leq \alpha$ for all α. □

Theorem 33 (Gaffke (2005, Lemma 2.2))**.** *Let $\boldsymbol{a} = a\mathbb{1}, a \in [0, 1]$ and $\boldsymbol{b} = b\mathbb{1}, b \geq 1$, $a \neq b$. Then* (1) *holds.*

Proof. Gaffke (2005) proves this result by showing that another statistic, denoted there W, is stochastically larger than a standard uniform, and that K is in turn stochastically larger than W. A direct proof is below.

Let $p = p(a,b) = (b-1)/(b-a)$. Assume without loss that $\alpha = K(a\mathbb{1}_r, b\mathbb{1}_s)$ where $r \geq 0$, $s \geq 0$, and $r + s = n$.

The vectors $\{a,b\}^n$ making up the probability space are ordered by the number of occurrences of a, or equivalently b, among their components: if $r_1 > r_2$, i.e., $s_1 < s_2$, then by monotonicity

$$K(a\mathbb{1}_{r_1}, b\mathbb{1}_{s_1}) \leq K(a\mathbb{1}_{r_2}, b\mathbb{1}_{s_2}).$$

Therefore

$$\begin{aligned}\mathbb{P}(K(\boldsymbol{x}) \leq \alpha) &= \mathbb{P}\big(K(\boldsymbol{x}) \leq K(a\mathbb{1}_r, b\mathbb{1}_s)\big) \\ &= \sum_{i=0}^{r} \mathbb{P}(a\mathbb{1}_i, b\mathbb{1}_{s+r-i}) \\ &= \sum_{i=0}^{r} \binom{r+s}{i} p^i (1-p)^{r+s-i}.\end{aligned}$$

Using the martingale interpretation of K given in Section 2.1.1,

$$K(a\mathbb{1}_r, b\mathbb{1}_s) = \sum_{i=0}^{r-1} \binom{s+i}{i} p^i (1-p)^s + \sum_{i=0}^{s-1} \binom{r+i}{i} p^r (1-p)^i b^{i-s}.$$

By Theorem 8, we have $1 - p \leq 1/b$, so $K(a\mathbb{1}_r, b\mathbb{1}_s)$ is lower bounded by

$$\begin{aligned}&\sum_{i=0}^{r-1} \binom{s+i}{i} p^i (1-p)^s + \sum_{i=0}^{s-1} \binom{r+i}{i} p^r (1-p)^i b^{i-s} \\ &\geq \sum_{i=0}^{r-1} \binom{s+i}{i} p^i (1-p)^s + \sum_{i=0}^{s-1} \binom{r+i}{i} p^r (1-p)^i (1-p)^s \\ &= \sum_{i=0}^{r-1} \binom{s+i}{i} p^i (1-p)^s + p^r (1-p)^s \binom{r+s}{s-1}.\end{aligned}$$

By our choice of α we have $K(a\mathbb{1}_r, b\mathbb{1}_s) \in \{K(\boldsymbol{x}) \leq \alpha\}$, so it suffices to establish the binomial inequality

$$\sum_{i=0}^{r} \binom{r+s}{i} p^i (1-p)^{r+s-i} \;\leq\; \sum_{i=0}^{r-1} \binom{s+i}{i} p^i (1-p)^s + p^r (1-p)^s \binom{r+s}{s-1}. \tag{24}$$

We use a probabilistic interpretation. Consider $r + s$ independent Bernoulli trials with common success probability p. The left-hand side is

$$\begin{aligned}\sum_{i=0}^{r} \binom{r+s}{i} p^i (1-p)^{r+s-i} &= \mathbb{P}(\geq s \text{ failures in } r+s \text{ trials}) \\ &= \mathbb{P}(\geq s \text{ failures in } r+s-1 \text{ trials}) \\ &\quad + (1-p)\,\mathbb{P}(s-1 \text{ failures in the first } r+s-1 \text{ trials}).\end{aligned}$$

For $s \geq 1$, the second term is

$$(1-p)\,\mathbb{P}(s-1 \text{ failures in the first } r+s-1 \text{ trials}) = p^r(1-p)^s\binom{r+s-1}{s-1}$$
$$\leq p^r(1-p)^s\binom{r+s}{s-1}.$$

The first term, by a union bound, is

$$\mathbb{P}(\geq s \text{ failures in } r+s-1 \text{ trials}) = \mathbb{P}\Big(\bigcup_{j=s}^{r+s-1}\{s \text{ failures and } j-s \text{ successes in the first } j \text{ trials}\}\Big)$$

$$\leq \sum_{j=s}^{r+s-1}\mathbb{P}(s \text{ failures and } j-s \text{ successes in the first } j \text{ trials})$$

$$= \sum_{j=s}^{r+s-1}\binom{j}{s}(1-p)^s p^{j-s}$$
$$= (1-p)^s\sum_{i=0}^{r-1}\binom{s+i}{s}p^i$$
$$= (1-p)^s\sum_{i=0}^{r-1}\binom{s+i}{i}p^i.$$

Combining the two bounds yields (24). □

Definition 34. *Let $\boldsymbol{a} \in [0,1]^n$, $\boldsymbol{b} \in [0,\infty)^n$, $n \geq 1$ be given. Let $\boldsymbol{X} = (X_1,\ldots,X_n)$ consist of independent random variables with mean 1, with X_i supported on $\{a_i, b_i\}$, $i = 1,\ldots,n$. For $I \in \{0,1\}^{m\times n}$ with distinct rows $I_1,\ldots,I_m$, define the $m \times n$ matrix $\boldsymbol{x}$ through*

$$x_{ij} = x(I_{ij}) = \begin{cases} a_j, & \text{if} \quad I_{ij} = 0 \\ b_j, & \text{if} \quad I_{ij} = 1. \end{cases}$$

with rows $\boldsymbol{x}_i = \boldsymbol{x}_{I_i} = \boldsymbol{x}(I_i) \in \prod_{j=1}^n\{a_j, b_j\}$

$$(\boldsymbol{x}(I_i))_j = \begin{cases} a_j & \text{if} \quad I_{ij} = 0 \\ b_j & \text{if} \quad I_{ij} = 1. \end{cases}$$

Theorem 35. *Gaffke's strong conjecture* (1) *is equivalent to the assertion that, for any $n \geq 1$ and the 2^{2^n} sets of distinct index sets $I_1,\ldots,I_m$, $1 \leq m \leq 2^n$,*

$$\sum_{i=1}^m \mathbb{P}((X_1,\ldots,X_n) = \boldsymbol{x}_{I_i}) \leq \max_{1\leq i\leq m} K(\boldsymbol{x}_{I_i}).$$

Proof. If the first form of the conjecture (1) is true, taking $\alpha = \max_{1\le i\le m} K(\boldsymbol{x}_{I_i})$,

$$\begin{aligned}
\sum_{i=1}^{m} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}_{I_i}) &\le \sum_{I:K(\boldsymbol{x}_I)\le \max_{1\le i\le m} K(\boldsymbol{x}_{I_i})} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}_{I_i}) \\
&= \sum_{I:K(x_I)\le\alpha} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}_{I_i}) \\
&= \mathbb{P}(K(\boldsymbol{X}) \le \alpha) \\
&\le \alpha = \max_{1\le i\le m} K(\boldsymbol{x}_{I_i}).
\end{aligned}$$

Conversely, letting $\{I_1, \ldots, I_m\} = \{I : K(\boldsymbol{x}_I) \le \alpha\}$,

$$\begin{aligned}
\mathbb{P}(K(\boldsymbol{X}) \le \alpha) &= \sum_{I:K(\boldsymbol{x}_I)\le\alpha} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}_I) \\
&= \sum_{i=1}^{m} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}_{I_i}) \le \alpha.
\end{aligned}$$

$\square$

In this form, a schematic depiction of the conjecture is, for all $\boldsymbol{a} \in [0,1]^n$, $\boldsymbol{b} \in [1,\infty)^n$, $I \in \{0,1\}^{m\times n}$,

$$\begin{array}{rcccl}
 & \mathbb{P}(x_{11}) \cdot \mathbb{P}(x_{12}) \cdots \cdot \mathbb{P}(x_{1,n}) & & & K(x_{11}, x_{12}, \ldots, x_{1,n}) \\
+ & \mathbb{P}(x_{21}) \cdot \mathbb{P}(x_{22}) \cdots \cdot \mathbb{P}(x_{2,n}) & \le & \bigvee & K(x_{21}, x_{22}, \ldots, x_{2,n}) \\
 & \vdots & & & \vdots \\
+ & \mathbb{P}(x_{m,1}) \cdot \mathbb{P}(x_{m,2}) \cdots \cdot \mathbb{P}(x_{m,n}) & & \bigvee & K(x_{m,1}, x_{m,2}, \ldots, x_{m,n})
\end{array} \tag{25}$$

5.1 Induction

For fixed j, $x_{ij} = x(I_{ij})$ is either a_j or b_j, depending on whether $I_{ij} = 0$ or $= 1$. If the column only contains one of these two values, i.e., $1 \le j \le n$, $x_{ij} = a_j$ for all i, or $x_{ij} = b_j$ for all $i, 1 \le i \le m$, then factoring out this common value, say $\mathbb{P}(x_{1,j})$, reduces to the $n-1$ case,

$$\begin{aligned}
\sum_{i=1}^{m} \mathbb{P}(\boldsymbol{x}_i) &= \mathbb{P}(x_{1,j}) \sum_{i=1}^{m} \mathbb{P}(\boldsymbol{x}_{i,-j}) \\
&\le \mathbb{P}(x_{1,j}) \max_{1\le i\le m} K(\boldsymbol{x}_{i,-j})
\end{aligned}$$

and an application of Theorems 8 and 9 completes the inductive step,

$$\begin{aligned}
&\le k_0(x_{1,j}) \max_{1\le i\le m} K(\boldsymbol{x}_{i,-j}) \\
&\le \max_{1\le i\le m} K(\boldsymbol{x}_i).
\end{aligned}$$

Suppose therefore that for each $j, 1 \leq j \leq n$, there are i, i', such that $x_{i,j} = a_j, x_{i',j} = b_j$. In terms of the matrix I (Definition 34), the condition is that $0 < \sum_{i=1}^m I_{ij} < m, 1 \leq j \leq n$. Then for any $j, 1 \leq j \leq n$, by the inductive hypothesis,

$$\begin{aligned}\sum_{i=1}^m \mathbb{P}(\boldsymbol{x}_i) &= \mathbb{P}(a_j) \sum_{i:x_{ij}=a_j} \mathbb{P}(x_{i,-j}) + \mathbb{P}(b_j) \sum_{i:x_{ij}=b_j} \mathbb{P}(x_{i,-j}) \\ &\leq \mathbb{P}(a_j) \max_{i:x_{ij}=a_j} K(\boldsymbol{x}_{i,-j}) + \mathbb{P}(b_j) \max_{i:x_{ij}=b_j} K(\boldsymbol{x}_{i,-j}).\end{aligned}$$

A candidate for an inductive step in a proof of (25) is

$$\min_{1\leq j\leq n} (\mathbb{P}(a_j) \max_{i:x_{ij}=a_j} K(\boldsymbol{x}_{i,-j}) + \mathbb{P}(b_j) \max_{i:x_{ij}=b_j} K(\boldsymbol{x}_{i,-j})) \leq \max_{1\leq i\leq m} K(\boldsymbol{x}_i). \tag{26}$$

Lemma 36. *Let $\boldsymbol{a} \in [0,1]^n$, $\boldsymbol{b} \in [1,\infty)^n$, $I \in \{0,1\}^{m\times n}$ be given. Suppose $x_1 = a_1$ and $y_1 = b_1$, and that Theorem 35 holds for $n-1$. Then $\mathbb{P}(\boldsymbol{x}) + \mathbb{P}(\boldsymbol{y}) > \max(K(\boldsymbol{x}), K(\boldsymbol{y}))$ iff*

1. *$K(\boldsymbol{x}) < K(\boldsymbol{y})$ and $c(\boldsymbol{x}) > c(\boldsymbol{y})$ and $|\frac{K(\boldsymbol{x})-K(\boldsymbol{y})}{c(\boldsymbol{x})-c(\boldsymbol{y})}| < b_1 - 1$, or*

2. *$K(\boldsymbol{x}) > K(\boldsymbol{y})$ and $c(\boldsymbol{x}) > c(\boldsymbol{y})$ and $|\frac{K(\boldsymbol{x})-K(\boldsymbol{y})}{c(\boldsymbol{x})-c(\boldsymbol{y})}| < 1 - a_1$.*

Proof. By the inductive hypothesis,

$$\begin{aligned}\mathbb{P}(\boldsymbol{x}) + \mathbb{P}(\boldsymbol{y}) &\leq \mathbb{P}(a_1)K(\boldsymbol{x}_{-1}) + \mathbb{P}(b_1)K(\boldsymbol{y}_{-1}) \\ &= \mathbb{P}(a_1)(K(\boldsymbol{x}) + (a_1-1)c(\boldsymbol{x})) + \mathbb{P}(b_1)(K(\boldsymbol{y}) + (b_1-1)c(\boldsymbol{y})).\end{aligned} \tag{27}$$

Expression (27) strictly exceeding $K(\boldsymbol{x})$, i.e.,

$$\begin{aligned}0 &< (\mathbb{P}(a_1)-1)K(\boldsymbol{x}) + (1-\mathbb{P}(a_1))K(\boldsymbol{y}) + \frac{b_1-1}{b_1-a_1}(a_1-1)c(\boldsymbol{x}) + \frac{1-a_1}{b_1-a_1}(b_1-1)c(\boldsymbol{y}) \\ &= \frac{1-a_1}{b_1-a_1}(K(\boldsymbol{y})-K(\boldsymbol{x})) + \frac{(1-a_1)(b_1-1)}{b_1-a_1}(c(\boldsymbol{y})-c(\boldsymbol{x})),\end{aligned}$$

is equivalent to

$$K(\boldsymbol{x}) - K(\boldsymbol{y}) < -(b_1-1)(c(\boldsymbol{x}) - c(\boldsymbol{y})). \tag{28}$$

Similarly, assuming (27) strictly exceeds $K(\boldsymbol{y})$ means

$$K(\boldsymbol{x}) - K(\boldsymbol{y}) > (1-a_1)(c(\boldsymbol{x}) - c(\boldsymbol{y})). \tag{29}$$

We consider by cases the implications for the ratio

$$|\frac{K(\boldsymbol{x})-K(\boldsymbol{y})}{c(\boldsymbol{x})-c(\boldsymbol{y})}|$$

of the possible signs of $K(\boldsymbol{x}) - K(\boldsymbol{y})$ and $c(\boldsymbol{x}) - c(\boldsymbol{y})$.

1. $K(\boldsymbol{x}) < K(\boldsymbol{y})$ and $c(\boldsymbol{x}) < c(\boldsymbol{y})$. Then (28) cannot hold.

2. $K(\boldsymbol{x}) < K(\boldsymbol{y})$ and $c(\boldsymbol{x}) > c(\boldsymbol{y})$. Then (28) implies the ratio is $< b_1 - 1$ while (29) implies the ratio is $> (-(1 - a_1))$, which is vacuous.

3. $K(\boldsymbol{x}) > K(\boldsymbol{y})$ and $c(\boldsymbol{x}) < c(\boldsymbol{y})$. Then (29) cannot hold

4. $K(\boldsymbol{x}) > K(\boldsymbol{y})$ and $c(\boldsymbol{x}) > c(\boldsymbol{y})$. Then (28) implies the ratio is $> -(b_1 - 1)$, which is vacuous, while (29) implies the ratio is $< (1 - a_1)$.

$\square$

Theorem 37. *Let $\boldsymbol{a} \in [0,1]^n$, $\boldsymbol{b} \in [1,\infty)^n$, $I \in \{0,1\}^{m\times n}$ be given, and suppose that* (25) *holds for $n-1$. Then if*

1. *for some $j, 1 \le j \le n$, $\sum_{i=1}^m I_{ij} \in \{0, m\}$*

the inequality (26) *holds. Suppose $0 < \sum_{i=1}^m I_{ij} < m, 1 \le j \le n$, and let $i_a(j) = \arg\max_{i:I_{ij}=0} K(\boldsymbol{x}_{i,-j})$ and $i_b(j) = \arg\max_{i:I_{ij}=1} K(\boldsymbol{x}_{i,-j})$ denote the indices of the two types of column maxima. The the following are sufficient conditions for* (26)*:*

2. *for any $j, 1 \le j \le n$, $\sum_{i:x_{ij}=a_j} \mathbb{P}(\boldsymbol{x}_{i,-j}) \ge \sum_{i:x_{ij}=b_j} \mathbb{P}(\boldsymbol{x}_{i,-j})$*

3. *for any $j, 1 \le j \le n$, $K(\boldsymbol{x}_{i_a(j)}) \ge K(\boldsymbol{x}_{i_b(j)})$*

4. *for any $j, 1 \le j \le n$, $c(\boldsymbol{x}_{i_a(j)}) \ge c(\boldsymbol{x}_{i_b(j)})$*

5. *for all $i, 1 \le i \le m$, there is $j, 1 \le j \le n$ such that $I_{ij} = 0$ and $I_{ij'} = 1, j' \ne j$.*

Proof. 1. The sufficiency of $\sum_{i=1}^m I_{ij} \in \{0, m\}$ is given in the discussion preceding the theorem.

2. If for some j, $\sum_{i:x_{ij}=a_j} \mathbb{P}(\boldsymbol{x}_{i,-j}) \ge \sum_{i:x_{ij}=b_j} \mathbb{P}(\boldsymbol{x}_{i,-j})$ then since $\mathbb{P}(a_j) + \mathbb{P}(b_j) = 1$, $\mathbb{P}(a_j) \sum_{i:x_{ij}=a_j} \mathbb{P}(\boldsymbol{x}_{i,-j}) + \mathbb{P}(b_j) \sum_{i:x_{ij}=b_j} \mathbb{P}(\boldsymbol{x}_{i,-j}) \le \mathbb{P}(a_j) \sum_{i:x_{ij}=a_j} \mathbb{P}(\boldsymbol{x}_{i,-j})$. By the inductive hypothesis and Theorem 9, the rhs is $\le \mathbb{P}(a_j) \max_{i:x_{ij}=a_j} K(\boldsymbol{x}_{i,-j}) \le \max_{i:x_{ij}=a_j} K(a_j, \boldsymbol{x}_{i,-j}) = \max_{i:x_{ij}=a_j} K(\boldsymbol{x}_i) \le \max_{1\le i\le m} K(\boldsymbol{x}_i)$.

3. If for some j, $\max_{i:x_{ij}=a_j} K(\boldsymbol{x}_{i,-j}) \ge \max_{i:x_{ij}=b_j} K(\boldsymbol{x}_{i,-j})$ then again by convexity

$$\mathbb{P}(a_j) \max_{i:x_{ij}=a_j} K(\boldsymbol{x}_{i,-j}) + \mathbb{P}(b_j) \max_{i:x_{ij}=b_j} K(\boldsymbol{x}_{i,-j}) \le \max_{i:x_{ij}=a_j} K(\boldsymbol{x}_{i,-j})$$

and by monotonicity $\max_{i:x_{ij}=a_j} K(\boldsymbol{x}_{i,-j}) \le \max_{i:x_{ij}=a_j} K(\boldsymbol{x}_{i,-j}, a_j) = \max_{i:x_{ij}=a_j} K(\boldsymbol{x}_i) \le \max_{1\le i\le m} K(\boldsymbol{x}_i)$.

4. Given by Lemma 36.

5. Let $i_0 = \arg\max_{1\le i\le m} c(\boldsymbol{x}_i)$ and let j_0 be the index corresponding to i_0 given by the condition. Then $i_a(j_0) = i_0$, and so $c(\boldsymbol{x}_{i_a(j_0)}) = c(\boldsymbol{x}_{i_0}) \ge c(\boldsymbol{x}_{i_b(j_0)})$, and sufficient condition (4) applies.

$\square$

Corollary 38. *Suppose $m = 1$ or $m = 2$. Then* (25) *holds.*

Proof. When $m = 1$ the claim is that for any $\boldsymbol{x} \in \prod_{i=1}^{n}\{a_i, b_i\}$,

$$\mathbb{P}(\boldsymbol{x}) \leq K(\boldsymbol{x}). \tag{30}$$

When $n = 1$, Theorem 32 implies (30). Assume (30) holds for $n-1$, and let $\boldsymbol{x} \in \prod_{i=1}^{n}\{a_i, b_i\}$ be given. Then by Theorems 8 and 9

$$\mathbb{P}(\boldsymbol{x}) \leq \mathbb{P}(x_1)K(\boldsymbol{x}_{-1}) \leq K(\boldsymbol{x}).$$

When $m = 2$ the claim is that for any $n \geq 1$ and $I_1 \in \{0,1\}^n$, $I_2 \in \{0,1\}^n$,

$$\prod_{j=1}^{n} p(x(I_{1j}), x(\bar{I}_{1j})) + \prod_{j=1}^{n} p(x(I_{2j}), x(\bar{I}_{2j})) \leq \max(K(\boldsymbol{x}(I_1)), K(\boldsymbol{x}(I_2))). \tag{31}$$

Again the $n = 1$ case is given by Theorem 32, so assume (31) holds for $n - 1$, and let $I_1 \in \{0,1\}^n$ and $I_2 \in \{0,1\}^n$ be given. If for some $i, 1 \leq i \leq n$, $x_i = y_i$ then $\mathbb{P}(\boldsymbol{x}) + \mathbb{P}(\boldsymbol{y}) = \mathbb{P}(x_i)(\mathbb{P}(\boldsymbol{x}_{-i}) + \mathbb{P}(\boldsymbol{y}_{-i})) \leq \mathbb{P}(x_i)\max(K(\boldsymbol{x}_{-i}), K(\boldsymbol{y}_{-i})) \leq \max(K(\boldsymbol{x}), K(\boldsymbol{y}))$, again by Theorems 8 and 9. Otherwise assume $x_i \neq y_i, 1 \leq i \leq n$. Then if $\boldsymbol{x}$ is a leaf node so is $\boldsymbol{y}$, and one of them is then $(a_1, \ldots, a_n)$, in which case the rhs of (31) is 1 and the inequality is satisfied. Therefore assume without loss of generality that $(x_1, x_2) = (a_1, b_2), (y_1, y_2) = (b_1, a_2)$, $a_1 \neq b_1, a_2 \neq b_2$. In this case, Theorem 37 (5) establishes the inductive step. $\square$

A drawback of working with the second form of Gaffke's conjecture given in Theorem 35 is that it obscures which special cases are useful to establish. For example, in the condition given in Theorem 37 that for some j, $\sum_{i:x_{ij}=a_j} \mathbb{P}(\boldsymbol{x}_{i,-j}) \geq \sum_{i:x_{ij}=b_j} \mathbb{P}(\boldsymbol{x}_{i,-j})$, strict inequality cannot hold for any subset of the probability space of the form $\mathbb{P}(K(\boldsymbol{x}) \leq \alpha, \alpha \in (0,1)$. See Section 5.2.

5.2 Cylinder sets on the probability space

Given a vector $\boldsymbol{x} = \prod_{i=1}^{n}(a_i, b_i)$ with $x_i = a_i$ for some index i, the monotonicity of K implies that changing x_i to b_i cannot increase $K(\boldsymbol{x})$. Therefore the closure of $lhs(\boldsymbol{a}, \boldsymbol{b}, I)$ under these shifts can only tighten (25) , increasing $lhs(\boldsymbol{a}, \boldsymbol{b}, I)$ without affecting $rhs(\boldsymbol{a}, \boldsymbol{b}, I)$. Since any subset $\{\boldsymbol{x} : K(\boldsymbol{x}) \leq \alpha\}$ of the probability space must be closed in this sense, there is no loss in generality in assuming $lhs(\boldsymbol{a}, \boldsymbol{b}, I)$ is closed.

Theorem 39. *Given $I \in \{0,1\}^{m \times n}$, suppose whenever $I_{ij} = 0$ for some i, j, there exists an i' such that $I_{i'j} = 1$ and $I_{i'j'} = I_{ij}$ for $j' \neq j$. Then*

$$\sum_{i=1}^{m}\prod_{j=1}^{n} p(x(I_{ij}), x(\bar{I}_{ij})) = \mathbb{P}(\bigcup_{i=1}^{m}\bigcap_{j:I_{ij}=1}\{X_j = b_j\}). \tag{32}$$

Proof. The lhs of (32) is the probability of the event

$$\bigcup_{i=1}^{m}((\bigcap_{j:I_{ij}=a_j}\{X_j = a_j\})\bigcap(\bigcap_{j:I_{ij}=b_j}\{X_j = b_j\})), \tag{33}$$

and since $((\bigcap_{j:I_{ij}=a_j}\{X_j = a_j\})\bigcap(\bigcap_{j:I_{ij}=b_j}\{X_j = b_j\})) \subset \bigcap_{j:I_{ij}=b_j}\{X_j = b_j\}$, the lhs is $\leq$ the rhs in (32) . For the converse direction, fix i in $1,\ldots,m$, and let $B = \{j : I_{ij} = 1\}$ and $A = \{1,\ldots,n\}/B = \{j : I_{ij} = 0\}$. The contribution of row i to the union event in the rhs of (32) is

$$\begin{aligned}
\bigcap_{j:I_{ij}=1}\{X_j = b_j\} &= (\bigcap_{j\in B}\{X_j = b_j\}\bigcap(\bigcap_{k\in A}\{X_k = b_k\}\bigcup\{X_k = a_k\}) \\
&= \bigcap_{k\in A}((\bigcap_{j\in B}\{X_j = b_j\})\bigcap(\{X_k = b_k\}\bigcup\{X_k = a_k\})) \\
&= \bigcup_{\boldsymbol{x}\in\prod_{k\in A}\{a_k,b_k\}}((\bigcap_{j\in B}\{X_j = b_j\})\bigcap(\bigcap_{x_k\in\boldsymbol{x}}\{X_k = x_k\})),
\end{aligned}$$

where the outer intersection in the last line is taken over the $2^{|A|}$ possible tuples $\boldsymbol{x}$ formed by choosing a_k or b_k for each $k \in A$. By the assumed closure property, each event $(\bigcap_{j\in B}\{X_j = b_j\})\bigcap(\bigcap_{x_k\in\boldsymbol{x}}\{X_k = x_k\}))$ in the union belongs to (33). $\square$

Then (25) may be replaced with loss of generality, by

$$\mathbb{P}(\bigcup_{i=1}^{m}\bigcap_{j\in J_i}\{X_j = b_j\}) \leq \max_{i=1}^{m} K((b_j)_{j\in J_i},(a_j)_{j\notin J_i})$$

where $J_1,\ldots,J_m$ are any distinct subsets of $\{1,\ldots,n\}$.

Lemma 40. *Given* $\boldsymbol{a} \in [0,1]^n$ *and* $\boldsymbol{b} \in [1,\infty)^n$, *let* $B = \prod_{i=1}^n b_i$. *Then*

$$\prod_{i=1}^{n}\frac{B-1}{B-a_i} \leq 1 - \prod_{i=1}^{n}\frac{1-a_i}{b_i-a_i}.$$

In the notation of (11)

$$\prod_{i=1}^{n} p(a_i,B) \leq 1 - \prod_{i=1}^{n} p(b_i,a_i).$$

Proof. Suppose $b > 1, \beta > 1$, and $a \in [0,1)$. Since $\frac{d^2}{db^2}(b-a_1)(\beta/b-a_2) = -\frac{2\beta a_1}{b^3} < 0$, $b \mapsto (b-a_1)(\beta/b-a_2)$ is concave, implying $b \mapsto \frac{1-a_1}{b-a_1}\frac{1-a_2}{\beta/b-a_2}$ is convex on $[1,\beta]$. Therefore

$$\frac{1-a_1}{b-a_1}\frac{1-a_2}{\beta/b-a_2} \leq \max(\frac{1-a_2}{\beta-a_2},\frac{1-a_1}{\beta-a_1}).$$

Suppose by way of induction that $a_{n+1} \in [0,1)$ and $b_{n+1} > 0$ are given and that

$$\prod_{i=1}^{n}\frac{1-a_i}{b_i-a_i} \leq \max_{1\leq i\leq n}\frac{1-a_i}{\prod_{j=1}^n b_j - a_i}.$$

An application of the inductive hypothesis along with the base case with $\beta = \prod_{j=1}^{n+1} b_j$ yields

$$\begin{aligned}\prod_{i=1}^{n+1} \frac{1-a_i}{b_i - a_i} &\leq \max_{1\leq i\leq n} \frac{1-a_i}{\prod_{j=1}^{n} b_j - a_i} \frac{1-a_{n+1}}{b_{n+1} - a_{n+1}} \\ &= \max_{1\leq i\leq n} \frac{1-a_i}{\beta/b_{n+1} - a_i} \frac{1-a_{n+1}}{b_{n+1}-a_{n+1}} \\ &\leq \max_{1\leq i\leq n} \max\left(\frac{1-a_i}{\beta - a_i}, \frac{1-a_{n+1}}{\beta - a_{n+1}}\right) \\ &= \max_{1\leq i\leq n+1} \frac{1-a_i}{\beta - a_i}.\end{aligned}$$

Therefore,

$$1 - \prod_{i=1}^{n} \frac{1-a_i}{b_i - a_i} \geq 1 - \max_{1\leq i\leq n} \frac{1-a_i}{\prod_{j=1}^{n} b_j - a_i} = \min_{1\leq i\leq n} \frac{\prod_{j=1}^{n} b_j - 1}{\prod_{j=1}^{n} b_j - a_i} \geq \prod_{i=1}^{n} \frac{\prod_{j=1}^{n} b_j - 1}{\prod_{j=1}^{n} b_j - a_i}.$$

□

Theorem 41. *Suppose* $\boldsymbol{a} \in [0,1]^n$, $\boldsymbol{b} \in [1,\infty)^n$, *and* $J_1, \ldots, J_m$, *are disjoint subsets of* $\{1, \ldots, n\}$. *Then* (25) *holds.*

Proof. Let i be fixed, $1 \leq i \leq m$, and let $J = \cup_{j=1}^{m} J_j$. Among the terms maximized over on the rhs of (25) there is, corresponding to J_i, $K((b_j)_{j\in J_i}, (a_j)_{j\notin J_i})$. For $j, 1 \leq j \leq m$, denote $B_j = \prod_{k\in J_j} b_k$. By Theorem 23 and (8)

$$\begin{aligned}K((b_j)_{j\in J_i}, (a_j)_{j\notin J_i}) &\geq K(B_i, (a_j)_{j\notin J_i}) \\ &= \frac{1}{B_i} \prod_{j\in J/J_i} p(a_j, B_i) + 1 - \prod_{j\in J/J_i} p(a_j, B_i). \qquad (34)\end{aligned}$$

The last expression is a convex combination of 1 and the smaller $1/B_i$. The lhs of (25) may also be written as a convex combination by conditioning on whether $\bigcap_{k\in J_j}\{X_k = b_k\}$ occurs for at least one j different from i,

$$\prod_{j:j\neq i}(1 - \mathbb{P}(\bigcap_{k\in J_j}\{X_k = b_k\}))\mathbb{P}(\bigcap_{j\in J_i}\{X_j = b_j\}) + 1 - \prod_{j:j\neq i}(1 - \mathbb{P}(\bigcap_{k\in J_j}\{X_k = b_k\})).$$

The lhs is a convex combination of 1 and $\mathbb{P}(\bigcap_{j\in J_i}\{X_j = b_j\})$ and by Theorem 8 the latter is $\leq \prod_{j\in J_i} 1/b_j = 1/B_i$. Therefore, if the lhs of (25) were strictly larger than the right hand side, and therefore also larger than (34), necessarily

$$\begin{aligned}\prod_{j:j\neq i}(1 - \mathbb{P}(\bigcap_{k\in J_j}\{X_k = b_k\})) &< \prod_{j\in J/J_i} p(a_j, B_i), \text{ or} \\ \prod_{j:j\neq i}(1 - \prod_{k\in J_j} p(b_k, a_k)) &< \prod_{j:j\neq i}\prod_{k\in J_j} p(a_k, B_i).\end{aligned}$$

Applying Lemma 40,

$$\prod_{j:j\neq i}\prod_{k\in J_j} p(a_k, B_j) < \prod_{j:j\neq i}\prod_{k\in J_j} p(a_k, B_i).$$

Therefore, for some j different from i, $\prod_{k\in J_j} p(a_k, B_j) < \prod_{k\in J_j} p(a_k, B_i)$, implying that for some j different from i, $B_j < B_i$, i.e.,

$$B_i > \min_{j:j\neq i} B_j.$$

If this inequality held for each i we would reach a contradiction as some $B_i, 1 \leq i \leq m$, must be minimal. Therefore, the lhs of (25) does not exceed at least one of the terms on the rhs of (25). $\square$

Theorem 42. *Suppose $\boldsymbol{a} \in [0,1]^n$, $\boldsymbol{b} \in [1,\infty)^n$, and J_1 and J_2 are subsets of $\{1,\ldots,n\}$. Then* (25) *holds.*

Proof. The lhs of (25) is

$$\begin{aligned}
&\mathbb{P}(\bigcap_{j\in J_1}\{X_j = b_j\}\bigcup\bigcap_{j\in J_2}\{X_j = b_j\}) \\
&= \mathbb{P}(\bigcap_{j\in J_1}\{X_j = b_j\}) + \mathbb{P}(\bigcap_{j\in J_2}\{X_j = b_j\}) - \mathbb{P}(\bigcap_{j\in J_1}\{X_j = b_j\}\bigcap\bigcap_{j\in J_2}\{X_j = b_j\}) \\
&= \mathbb{P}(\bigcap_{j\in J_1\cap J_2}\{X_j = b_j\})(\mathbb{P}(\bigcap_{j\in J_2/J_1}\{X_j = b_j\}) + (1-\mathbb{P}(\bigcap_{j\in J_2/J_1}\{X_j = b_j\}))\mathbb{P}(\bigcap_{j\in J_1/J_2}\{X_j = b_j\})) \\
&= \prod_{j\in J_1\cap J_2} p(b_j, a_j)(\prod_{j\in J_2/J_1} p(b_j, a_j) + (1-\prod_{j\in J_2/J_1} p(b_j, a_j))\prod_{j\in J_1/J_2} p(b_j, a_j)).
\end{aligned}$$

The rhs of (25) is

$$\max(K((a_j)_{j\in J_2/J_1}, (b_j)_{j\in J_1}), K((a_j)_{j\in J_1/J_2}, (b_j)_{j\in J_2})).$$

If (25) does not hold, then in particular the lhs would strictly exceed $K((a_j)_{j\in J_2/J_1}, (b_j)_{j\in J_1})$. Then by Theorem 23, denoting $B_1 = \prod_{j\in J_1} b_j$,

$$\prod_{j\in J_2/J_1} p(b_j, a_j) + (1-\prod_{j\in J_2/J_1} p(b_j, a_j))\prod_{j\in J_1/J_2} p(b_j, a_j) > \frac{K((a_j)_{j\in J_2/J_1}, B_1)}{\prod_{j\in J_1\cap J_2} p(b_j, a_j)}.$$

By repeated application of Theorem 10,

$$\begin{aligned}
\prod_{j\in J_1\cap J_2} p(b_j, a_j)K(\prod_{j\in J_1/J_2} b_j, (a_j)_{j\in J_2/J_1}) &\leq \prod_{j\in J_1\cap J_2} k_0(b_j)K(\prod_{j\in J_1/J_2} b_j, (a_j)_{j\in J_2/J_1}) \\
&\leq K(B_1, (a_j)_{j\in J_2/J_1}).
\end{aligned}$$

Therefore a violation of (25) would imply

$$\begin{aligned}
&\prod_{j\in J_2/J_1} p(b_j,a_j) + (1-\prod_{j\in J_2/J_1} p(b_j,a_j))\prod_{j\in J_1/J_2} p(b_j,a_j)\\
&> K(\prod_{j\in J_1/J_2} b_j,(a_j)_{j\in J_2/J_1})\\
&= 1-\prod_{j\in J_2/J_1} p(a_j,\prod_{k\in J_1/J_2} b_k) + \frac{1}{\prod_{k\in J_1/J_2} b_k}\prod_{j\in J_2/J_1} p(a_j,\prod_{k\in J_1/J_2} b_k).
\end{aligned}$$

The remainder is as in the proof of Theorem 41. Comparing terms of the two convex combinations, the above inequality implies

$$\prod_{j\in J_2/J_1} p(b_j,a_j) > \frac{1}{\prod_{k\in J_1/J_2} b_k}\prod_{j\in J_2/J_1} p(a_j,\prod_{k\in J_1/J_2} b_k)$$

and

$$1-\prod_{j\in J_2/J_1} p(b_j,a_j) < \prod_{j\in J_2/J_1} p(a_j,\prod_{k\in J_1/J_2} b_k).$$

The second inequality, by Lemma 40, implies $\prod_{k\in J_2/J_1} b_k < \prod_{k\in J_1/J_2} b_k$, which in turn implies $\prod_{k\in J_2} b_k < \prod_{k\in J_1} b_k$. Repeating the argument with the rhs term $K((a_j)_{j\in J_1/J_2},(b_j)_{j\in J_2})$ rather than $K((a_j)_{j\in J_2/J_1},(b_j)_{j\in J_1})$, or just from symmetry considerations, a violation of (25) would lead also to $\prod_{k\in J_1} b_k < \prod_{k\in J_2} b_k$, a contradiction. □

The proofs of Theorems 41 and 42 show that, when $lhs(\boldsymbol{a},\boldsymbol{b},I)$ has the special structure assumed by those theorems, the inequality (25) may be tightened by replacing each $K((a)_{j\notin J_i},(b)_{j\in J_i})$ on the rhs with $K((a)_{j\notin J_i},\prod_{j\in J_i} b_j)$.

6 Background

Suppose $X_1,\dots,X_n$, are independent random variables with known lower bounds $l_1,\dots,l_n$. Given $\mu_0 > \max_i l_i$, consider the problem of testing

$$H_0 : \max_i \mathbb{E}X_i \le \mu_0 \text{ vs. } H_a : \max_i \mathbb{E}X_i > \mu_0. \tag{35}$$

Shifting the data by $\min l_i$, and then scaling by $\max l_i - \min l_i$, linearity of expectation implies that (35) is the same as testing, based on nonnegative, independent $X_1,\dots,X_n$,

$$H_0 : \max_i \mathbb{E}X_i \le 1 \text{ vs. } H_a : \max_i \mathbb{E}X_i > 1. \tag{36}$$

Gaffke (2005) proposes to use $K(X_1,\dots,X_n)$ as a test statistic for (36), rejecting H_0 iff $K(X_1,\dots,X_n)\le\alpha$. This test controls the false positive rate at level α iff, for all nonnegative, independent $X_1,\dots,X_n$, with means in $[0,1]$,

$$\mathbb{P}(K(X_1,\dots,X_n)\le\alpha) < \alpha. \tag{37}$$

Since K is non-increasing in each argument (Theorem 2), another shift of the data reduces the problem to showing (37) holds for all non-negative, independent $X_1, \ldots, X_n$, with mean 1. As a final reduction, Gaffke (2005), citing Theorem 2.4 and Lemma 2.5 of Gaffke (2004), notes that it is enough to consider nonnegative, independent $X_1, \ldots, X_n$, with mean 1, with each X_i supported on two points. This reduction also follows directly from Theorem 1 of Dubins (1962). . In this form, Gaffke's strong conjecture (37) asserts that K provides a valid test statistic for the hypothesis testing problem (35). A weaker conjecture of Gaffke (2005) is that (37) holds under the further requirement that the X_i be identically distributed.

Another test for (35) is the likelihood ratio test of Wang and Zhao (2003). The test statistic is given by

$$R : \mathbb{R}_+^{(\mathbb{N})} \to \mathbb{R} : (x_1, \ldots, x_n) \mapsto \inf_{q \in [0,1]} \prod_{i=1}^{n} \frac{1}{q + (1-q)x_i}, \tag{38}$$

and the decision rule at level α is to reject the null based on a sample $X_1, \ldots, X_n$, when $R(X_1, \ldots, X_n) \leq \alpha$. Gaffke (2005) establishes that $K(x_1, \ldots, x_n) \leq R(x_1, \ldots, x_n)$ (see also Section 3), so that establishing (37) for K implies that it also holds for R, and so the validity of the likelihood ratio test follows from that of the test based on K.

For motivation underlying the form of the statistic K, as well as equivalent formulations using uniformly distributed random variables, see Gaffke (2005).

References

de Boor, C. (2005). Divided differences. *Surveys in Approximation Theory 1*, 46–69.

Dubins, L. E. (1962). On extreme points of convex sets. *Journal of Mathematical Analysis and Applications 5*(2), 237–244.

Gaffke, N. (2004). Nonparametric one-sided testing for the mean and related extremum problems. *Mathematical Methods of Statistics 13*(4), 369–391.

Gaffke, N. (2005). Three test statistics for a nonparametric one-sided hypothesis on the mean of a nonnegative variable. *Mathematical Methods of Statistics 14*(4), 451–467.

Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika 38*(3-4), 481–482.

Kaminsky, K., E. Luks, and P. Nelson (1984). Strategy, nontransitive dominance and the exponential distribution. *Australian Journal of Statistics 26*(2), 111–118.

Learned-Miller, E. and P. S. Thomas (2019). A new confidence interval for the mean of a bounded random variable. *arXiv preprint arXiv:1905.06208*.

Phan, M., P. Thomas, and E. Learned-Miller (2021). Towards practical mean bounds for small samples. In *International Conference on Machine Learning*, pp. 8567–8576.

Wang, W. and L. H. Zhao (2003). Nonparametric tests for the mean of a non-negative population. *Journal of Statistical Planning and Inference 110*(1-2), 75–96.

Waudby-Smith, I. and A. Ramdas (2024). Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology 86*(1), 1–27.