DISTRIBUTIONAL TRANSFORMATIONS WITHOUT ORTHOGONALITY RELATIONS

CHRISTIAN DÖBLER

Abstract. Distributional transformations characterized by equations relating expectations of test functions weighted by a given biasing function on the original distribution to expectations of the test function’s higher derivatives with respect to the transformed distribution play a great role in Stein’s method and were, in great generality, first considered by Goldstein and Reinert. We prove two abstract existence and uniqueness results for such distributional transformations, generalizing their $X - P$ bias transformation. On the one hand, we show how one can abandon previously necessary orthogonality relations by subtracting an explicitly known polynomial depending on the test function from the test function itself. On the other hand, we prove that for a given nonnegative integer $m$ it is possible to obtain the expectation of the $m$-th derivative of the test function with respect to the transformed distribution in the defining equation, even though the biasing function may have $k < m$ sign changes, if these two numbers have the same parity. We also explain how these findings may be applied to improve recent results by Pike and Ren on convergence rates for certain random sums to the Laplace distribution.

1. Introduction

Distributional transformations play a great role in Stein’s method in connection with certain coupling constructions, which are often an essential tool for bounding the quantities arising from Stein’s equation. Important and well studied examples are given by the well known size bias transformation (see e.g. [GR96], [AGK13] or [AG10]) and the zero bias transformation, which was introduced in [GR97] for mean zero random variables with finite and strictly positive variance. For an introduction to Stein’s method for normal approximation we refer to the book [CGS11], which includes an extensive discussion of the use of various coupling constructions in Stein’s method. For a general introduction to Stein’s method we refer to the book [BC05].

Technische Universität München, Fakultät für Mathematik, Lehrstuhl für Mathematische Physik, D-85740 München, Germany.
christian.doebler@tum.de

Keywords: Stein’s method, distributional transformations, zero bias transformation, size bias transformation, equilibrium distribution, random sums.
Recall that for a nonnegative random variable $X$ with finite and positive mean $\mu$ one says that a random variable $X^*$ has the $X$-size biased distribution if the identity $E[Xf(X)] = \mu E[f(X^*)]$ holds for all bounded and measurable functions $f$ on $[0, \infty)$. Existence of the $X$-size biased distribution is easily seen by just letting the distribution of $X^*$ have Radon-Nikodym derivative $x \mapsto x/\mu$ with respect to the distribution of $X$. In contrast, if $X$ is a given real-valued random variable with variance $\sigma^2 \in (0, \infty)$ and if $E[X] = 0$, then a random variable $X^*$ is said to have the $X$-zero biased distribution if $E[Xf(X)] = \sigma^2 E[f'(X^*)]$ for all Lipschitz continuous functions $f$ on $\mathbb{R}$. It was shown in [GR97] that, for a given real-valued random variable $X$, the $X$-zero biased distribution exists uniquely if and only if $0 < E[X^2] < \infty$ and $E[X] = 0$ and that the distribution of $X^*$ is always absolutely continuous with respect to Lebesgue measure. In [GR05], given a real-valued random variable $X$, an integer $m \geq 0$ and a function $P$ on $\mathbb{R}$, Goldstein and Reinert addressed the general problem of when a random variable $X(P)$ and a constant $\alpha$ exist such that

\begin{equation}
E[P(X)F(X)] = \alpha E[F^{(m)}(X(P))]
\end{equation}

holds for a sufficiently large class of functions $F$ on $\mathbb{R}$. Their most general result in this direction, Theorem 2.1 of [GR05], guarantees existence and uniqueness of the distribution for such a random variable $X(P)$, if $P$ has exactly $m$ sign changes on $\mathbb{R}$ and if there exists an $\alpha > 0$ such that the orthogonality relations $E[X^kP(X)] = m!\alpha \delta_{k,m}$ hold for $k = 0, 1, \ldots, m$.

The main purpose of the present paper is to generalize Theorem 2.1 of [GR05] in two respects: Firstly, in Theorem 2.1 below, we make sure that one can do without the orthogonality relations by replacing the term $F(X)$ on the left hand side of (1) by $F(X) - L_F(X)$, where $L_F$ is an explicit polynomial of degree at most $m - 1$, which depends on $F$ and the sign change points of $P$. We further show that the distribution of $X(P)$ is always absolutely continuous with respect to Lebesgue measure, if $m \geq 1$. Secondly, we consider the case that the number $k$ of sign changes of the function $P$ is strictly smaller than the order $m$ of the derivative we would like to have on the right hand side of (1). Our general existence and uniqueness result, Theorem 3.3 which is in fact a generalization of Theorem 2.1 makes sure that the desired distributional transformation exists, if we additionally assume that $k$ and $m$ have the same parity. We also hint at how Theorem 3.3 (or, in fact, Proposition 3.1) can be applied to improve results from [PR12] on convergence rates of certain random sums to the Laplace distribution. However, this paper is mainly intended to extend the abstract theory of distributional transformations and we wish to present applications to distributional convergence and concrete error bounds by Stein’s method elsewhere.

The paper is structured as follows: In Section 2 we state and prove Theorem 2.1 using
a probabilistic approach and discuss properties of this distributional transformation. In Section 3 we prove Theorem 3.3 on the case of fewer than $m$ sign changes and finally, in Section 4 we give an analytical proof of the existence part in Theorem 2.1 which invokes the Riesz representation theorem and which was our original proof of this result. Since we prefer to use the symbol $P$ for probability measures, from now on we denote the biasing function by $B$ instead.

2. Biasing functions with $m$ sign changes

Let $B : \mathbb{R} \rightarrow \mathbb{R}$ be a measurable function, which will henceforth be called a biasing function. We say $B$ has no sign changes, if it is either nonnegative or nonpositive on $\mathbb{R}$. For $m \in \{1, 2, \ldots \}$ we say that $B$ has $m$ sign changes occurring at the points $x_1 < x_2 < \ldots < x_m$, if the following holds: Letting $J_1 := (-\infty, x_1], J_2 := (x_1, x_2], \ldots, J_m := (x_{m-1}, x_m] \text{ and } J_{m+1} := (x_m, \infty)$ we have for each integer $1 \leq k \leq m$ and for all $x \in J_k, y \in J_{k+1}$ that $B(x) \cdot B(y) \leq 0$ and there is an $x \in J_k \cup J_{k+1}$ such that $B(x) \neq 0$. As was already noted in [GR05], the points where the sign changes occur may not be unique if there are non-trivial subintervals of $\mathbb{R}$, where $B$ is identically equal to zero. Note that this definition is slightly more general than the one in [GR05] in that they would additionally demand the existence of an $x \in J_k \cup J_{k+1}$ with $B(x) \neq 0$ for each $k = 1, 2, \ldots, m + 1$. This generalization also implies that a given $B$ may be considered to have both, $m$ and $m'$ sign changes, where $m \neq m'$, if there are non-trivial subintervals of $\mathbb{R}$, where $B$ is identically equal to zero. Throughout, for $m \in \{1, 2, \ldots \}$, we denote by $\mathcal{F}^m$ the class of all functions $F \in C^{m-1}(\mathbb{R})$ such that $F^{(m-1)}$ is still Lipschitz-continuous. For $m = 0$, we denote by $\mathcal{F}^0$ the class of all bounded and Borel-measurable functions on $\mathbb{R}$. Further, we adopt the standard conventions that empty sums are set equal to zero and empty products are set equal to one.

For a function $F$ on $\mathbb{R}$ and $m \geq 0$ real numbers $x_1 < x_2 < \ldots < x_m$ we define the polynomial $L_F := L_{F,x_1,\ldots,x_m}$ of degree at most $m - 1$ by $L_F := 0$ if $m = 0$ and for $m \geq 1$ we define $L_F$ to be the interpolation polynomial corresponding to the function $F$ and the nodes $x_1, \ldots, x_m$, i.e.

\begin{equation}
L_F(x) := \sum_{k=1}^{m} F(x_k) \prod_{j=1 \atop j \neq k}^{m} \frac{x - x_j}{x_k - x_j}.
\end{equation}

The main purpose of this section is to give a proof of the following theorem, which is a generalization of Theorem 2.1 in [GR05].

**Theorem 2.1.** Let $m$ be a nonnegative integer and let $B : \mathbb{R} \rightarrow \mathbb{R}$ be a measurable biasing function having $m \geq 0$ sign changes at the points $x_1 < x_2 < \ldots < x_m$, and
suppose that $B$ is nonnegative on $J_{m+1} = (x_m, \infty)$ (on $\mathbb{R}$, if $m = 0$). Assume further that $X$ is a real-valued random variable on some probability space $(\Omega, \mathcal{A}, P)$ such that $E|X^j B(X)| < \infty$ for $j = 0, 1, \ldots, m$ and

$$\alpha := \frac{1}{m!} E \left[ B(X)(X - x_m)(X - x_{m-1}) \cdots (X - x_1) \right] \neq 0.$$ 

Then, $\alpha$ is necessarily positive and there exists a unique distribution for a random variable $X^{(B)}$ such that for all $F \in \mathcal{F}^m$ we have

\begin{equation}
\alpha E \left[ F^{(m)}(X^{(B)}) \right] = E \left[ B(X)(F(X) - L_F(X)) \right],
\end{equation}

with $L_F$ as defined in (2). Furthermore, if $m \geq 1$, then the distribution of $X^{(B)}$ is absolutely continuous with respect to Lebesgue measure.

Remark 2.2. (a) If $X$ and $B$ additionally satisfy the orthogonality conditions $E[X^j B(X)] = 0$ for all $j = 0, 1, \ldots, m - 1$, then the distribution $\mathcal{L}(X^{(B)})$ of $X^{(B)}$ reduces to the $X - B$ biased distribution from [GR05] as is easily seen by writing the polynomial $L_F$ in terms of the monomials $1, X, \ldots, X^{m-1}$. Also, in this case for the same reason we have $\alpha = (m!)^{-1} E[X^m B(X)]$. So it is justified to call the distribution of $X^{(B)}$ the generalized $X - B$ biased distribution.

(b) Note that if, according to our definition of sign changes, $B$ has both, $m$ and $m'$ sign changes for $m \neq m'$, then we see from (3) that these two points of view lead to different distributions for $X^{(B)}$. Also, if we may consider $B$ to have sign changes at $x_1 < \ldots < x_m$ as well as at $y_1 < \ldots < y_m$, then the resulting $\alpha$’s and, again, the distributions of $X^{(B)}$’s are different, in general, which is in contrast to the theory from [GR05], where such ambiguities are ruled out by their orthogonality assumptions on $X$ with respect to $B$. Thus, one should actually denote the variable $X^{(B)}$ by $X^{(B; x_1, \ldots, x_m)}$ to prevent these ambiguities. We will, however, not do so but rather assume that it is understood or mention how many sign changes at what exact points the function $B$ is supposed to have.

(c) For the existence part of Theorem 2.1 we give two different proofs: An analytical proof, which uses the Riesz representation theorem, and a probabilistic proof, which relies on an explicit construction of the random variable $X^{(B)}$. Remarkably, the same construction of $X^{(B)}$ as in [GR05] is still valid in this more general setting. However, we were not able to generalize the proof of Theorem 2.1 in [GR05] to a proof of our Theorem 2.1.

(d) In the case $m = 1$, one can easily show that the function $p$ given by

$$p(t) = \frac{1}{\alpha} E \left[ B(X) \left(1_{\{x_1 \leq t \leq X\}} - 1_{\{X \leq t < x_1\}} \right) \right]$$


is a probability density function on \( \mathbb{R} \), whose associated distribution satisfies the requirements for the generalized \( X - B \) biased distribution, thus yielding a direct proof of existence and absolute continuity in this case.

(e) Note that if \( F \in \mathcal{F}^m \), then one can easily show by induction on \( k = 0, 1, \ldots, m \) that there exist finite constants \( c_k > 0 \) such that

\[
|F^{(m-k)}(x)| \leq c_k (1 + |x|^k)
\]

for each \( x \in \mathbb{R} \). Hence, if \( X \) satisfies the conditions from Theorem 2.1, then \( E\left[B(X)F(X)\right] \) exists for each \( F \in \mathcal{F}^m \).

(f) The assumption \( E|X^jB(X)| < \infty \) for \( j = 0, 1, \ldots, m \) is easily seen to be equivalent to \( E|B(X)| < \infty \) and \( E|X^mB(X)| < \infty \).

Proof of uniqueness in Theorem 2.1. The argument for uniqueness is the same as in [GR05] and is only included for reasons of completeness. Let \( \mu \) and \( \nu \) both be probability measures on \( (\mathbb{R}, \mathcal{B}(\mathbb{R})) \) such that random variables \( V \sim \mu \) and \( W \sim \nu \) satisfy the conditions on \( X^{(B)} \). For a function \( f \in C_c(\mathbb{R}) \), the class of continuous functions with compact support, consider the function \( F := I^m f := I^m(f) \) on \( \mathbb{R} \). Here, \( If(x) := \int_0^x f(t)dt \) and \( I^m \) is the \( m \)-th iterate of \( I \). Then, \( F^{(m)} = f \) and, since \( \|f\|_\infty < \infty \), it follows from Remark 2.2 (e) that \( E|F(X)B(X)| < \infty \) and by (3) we have that

\[
\int_\mathbb{R} f(x)d\mu(x) = E[F^{(m)}(V)] = \frac{1}{\alpha}E\left[B(X)\left(F(X) - L_F(X)\right)\right] = E[F^{(m)}(W)]
\]

= \( \int_\mathbb{R} f(x)d\nu(x) \).

Since the class \( C_c(\mathbb{R}) \) is separating probability measures, this implies that \( \mu = \nu \). \( \square \)

Probabilistic existence proof. We first argue for a fixed random variable \( X \) by induction on \( m \geq 0 \) that \( \alpha \) is positive for each biasing function \( B \) satisfying the assumptions of Theorem 2.1. If \( m = 0 \), then it follows that \( B \) is nonnegative on \( \mathbb{R} \) and, hence, also \( \alpha = E[B(X)] \geq 0 \). Since we have assumed that \( \alpha \neq 0 \), it follows that \( \alpha > 0 \). Now, suppose that \( m \geq 1 \) and assume the claim for all biasing functions having \( m - 1 \) sign changes. Define \( \tilde{B}(x) := B(x) \between x_{m-1} \). Then \( \tilde{B} \) changes signs at the points \( x_1 < x_2 < \ldots < x_{m-1} \) but no more at \( x_m \) and these exhaust the sign change points of \( \tilde{B} \). Thus, \( \tilde{B} \) has \( m - 1 \) sign changes and \( \tilde{B}(x) \geq 0 \) for \( x \geq x_{m-1} \). Hence, from the induction hypothesis it follows that
\[ \beta := E\left[ \tilde{B}(X) \prod_{j=1}^{m-1} (X - x_j) \right] > 0, \]

whenever we know that \( \beta \neq 0 \). But, since

\[ \beta = E\left[ \tilde{B}(X) \prod_{j=1}^{m-1} (X - x_j) \right] = E\left[ B(X) \prod_{j=1}^{m} (X - x_j) \right] = \alpha \neq 0 \]

by assumption, we know that \( \alpha = \beta > 0 \). This completes the proof that \( \alpha \) is positive. Now, we give the explicit construction of the random variable \( X^{(B)} \) from [GR05].

Let \( Y, U_1, U_2, \ldots, U_m \) be independent random variables such that \( U_j \) has the density \( p_j(u) := ju^{j-1}1_{(0,1)}(u) \) (\( 1 \leq j \leq m \)) and \( Y \) has distribution \( \nu \) given by

\[ (4) \quad d\nu(y) := \frac{1}{\alpha m!} \prod_{i=1}^{m} (y - x_i) B(y) d\mu(y), \]

where \( \mu \) is the distribution of \( X \). Note that \( \prod_{i=1}^{m} (y - x_i) B(y) \) is nonnegative on \( \mathbb{R} \) by the assumption on \( B \) and that this and the definition of \( \alpha \), which is always positive, imply that \( \nu \) is indeed a probability measure and, hence, such a \( Y \) exists.

Now, we define the random variable

\[ (5) \quad X^{(B)} := x_m + \sum_{k=1}^{m} \left( \prod_{i=k}^{m} U_i \right) (x_{k-1} - x_k), \]

where \( x_0 := Y \). We claim that \( X^{(B)} \) satisfies (3). This claim will be proved by induction on \( m = 0, 1, \ldots \). If \( m = 0 \), then the claim reduces to

\[ E[B(X) F(X)] = \alpha E[F(Y)], \]

where \( \alpha = E[B(X)] \). But this immediately follows from the definition of \( \nu = \mathcal{L}(Y) \).

Now, suppose that \( m \geq 1 \) and that the claim is proved for \( m-1 \). Again, we consider the function \( \tilde{B}(x) := (x - x_m) B(x) \) with exactly \( m-1 \) sign changes occurring at the points \( x_1 < \ldots < x_{m-1} \) and such that \( \tilde{B}(x) \geq 0 \) for \( x \geq x_{m-1} \). Furthermore, we have

\[ X^{(B)} = (1 - U_m)x_m + U_m \left( x_{m-1} + \sum_{k=1}^{m-1} \left( \prod_{i=k}^{m-1} U_i \right) (x_{k-1} - x_k) \right) \]

and since
\[
\prod_{i=1}^{m} (y - x_i)B(y) = \prod_{i=1}^{m-1} (y - x_i)\tilde{B}(y)
\]

and

\[
\alpha m! = \mathbb{E}\left[ B(X) \prod_{i=1}^{m} (X - x_i) \right] = \mathbb{E}\left[ \tilde{B}(X) \prod_{i=1}^{m-1} (X - x_i) \right]
\]

we conclude from the induction hypothesis that

\[
X^{(\tilde{B})} := x_{m-1} + \sum_{k=1}^{m-1} \left( \prod_{i=k}^{m-1} U_i \right) (x_{k-1} - x_k)
\]

with the same \( Y = x_0 \) satisfies the assumptions on a random variable with the generalized \( X - \tilde{B} \) biased distribution. Now, let \( F \in \mathcal{F}^m \) be given and define \( G := F' \). Furthermore, for \( t \in [0, 1] \) we let \( G_t(x) := G(x_m + t(x - x_m)) \), noting that \( G_t^{(k)}(x) = G_t^{(k)}(x_m + t(x - x_m)) t^k \) for \( k = 0, 1, \ldots, m - 1 \) and that each \( G_t \in \mathcal{F}^{m-1} \). Since \( X^{(B)} = x_m + U_m (X^{(\tilde{B})} - x_m) \) and \( X^{(\tilde{B})}, U_m \) are independent, we have from Fubini’s theorem that

\[
\alpha \mathbb{E}\left[ F^{(m)}(X^{(B)}) \right] = \alpha \mathbb{E}\left[ (m-1)! (x_m + U_m (X^{(\tilde{B})} - x_m)) \right] = \alpha \int_0^1 \mathbb{E}\left[ G^{(m-1)}(x_m + t(X^{(\tilde{B})} - x_m)) \right] m t^{m-1} dt = \alpha m \int_0^1 \mathbb{E}\left[ G_t^{(m-1)}(X^{(\tilde{B})}) \right] dt.
\]

Noting that

\[
\alpha m = \frac{\alpha m!}{(m - 1)!} = \frac{1}{(m - 1)!} \mathbb{E}\left[ \tilde{B}(X) \prod_{j=1}^{m-1} (X - x_j) \right],
\]

we can thus conclude from the induction hypothesis that
\[ \alpha E \left[ F(m) (X(B)) \right] = \int_0^1 E \left[ \tilde{B}(X)(G_t(X) - L_{G_t}(X)) \right] dt \]
\[ = \int_0^1 E \left[ \tilde{B}(X) \left( G_t(X) - \sum_{k=1}^{m-1} G_t(x_k) \prod_{j=1 \atop j \neq k}^{m-1} \frac{X - x_j}{x_k - x_j} \right) \right] dt \]
\[ = E \left[ B(X) \int_0^1 (X - x_m) \left( G_t(X) - \sum_{k=1}^{m-1} G_t(x_k) \prod_{j=1 \atop j \neq k}^{m-1} \frac{X - x_j}{x_k - x_j} \right) dt \right]. \tag{7} \]

Now, for each real \( z \neq x_m \)
\[ \int_0^1 G_t(z) dt = \frac{1}{z - x_m} \int_0^1 F'(x_m + t(z - x_m))(z - x_m) dt \]
\[ = \frac{1}{z - x_m} \int_x^z F'(s) ds = \frac{F(z) - F(x_m)}{z - x_m}, \tag{8} \]

implying that
\[ (x - x_m) \int_0^1 G_t(x) dt = F(x) - F(x_m) \tag{9} \]

for each \( x \in \mathbb{R} \).

From (8) and (9) we conclude for \( x \in \mathbb{R} \) that
\[ (x - x_m) \int_0^1 \left( G_t(x) - \sum_{k=1}^{m-1} G_t(x_k) \prod_{j=1 \atop j \neq k}^{m-1} \frac{x - x_j}{x_k - x_j} \right) dt \]
\[ = F(x) - F(x_m) - \sum_{k=1}^{m-1} \left( F(x_k) - F(x_m) \right) \frac{x - x_m}{x_k - x_m} \prod_{j=1 \atop j \neq k}^{m-1} \frac{x - x_j}{x_k - x_j} \]
\[ = F(x) - F(x_m) - \sum_{k=1}^{m-1} F(x_k) \prod_{j=1 \atop j \neq k}^{m-1} \frac{x - x_j}{x_k - x_j} \]
\[ + F(x_m) \sum_{k=1 \atop j \neq k}^{m-1} \frac{x - x_j}{x_k - x_j}. \]
Now we notice that for each \( x \in \mathbb{R} \) we have

\[
\sum_{k=1}^{m-1} \prod_{j=1 \atop j \neq k}^m \frac{x - x_j}{x_k - x_j} = 1 - \prod_{j=1}^{m-1} \frac{x - x_j}{x_m - x_j},
\]

which is clear from the Lagrange form of the interpolation polynomial corresponding to the constant function 1 and the nodes \( x_1, \ldots, x_m \). Using this, we obtain that

\[
(x - x_m) \int_0^1 \left( G_t(x) - \sum_{k=1}^{m-1} G_t(x_k) \prod_{j=1 \atop j \neq k}^m \frac{x - x_j}{x_k - x_j} \right) dt
\]

\[
= F(x) - \sum_{k=1}^m F(x_k) \prod_{j=1 \atop j \neq k}^m \frac{x - x_j}{x_k - x_j}
\]

\[
= F(x) - L F(x).
\]

Plugging this into (7) we see that

\[
\alpha E \left[ F^{(m)}(X^{(B)}) \right] = E \left[ B(X) (F(X) - L F(X)) \right],
\]

as claimed. \( \square \)

**Proof of absolute continuity if \( m \geq 1 \).** To prove this claim, we use the explicit construction of \( X^{(B)} \) given in (5). Thus, we have that

\[
X^{(B)} = x_m + \sum_{k=1}^m \left( \prod_{i=k}^m U_i \right) (x_{k-1} - x_k)
\]

\[
= U_1 \left( \prod_{i=2}^m U_i \right) (Y - x_1) + \left( x_m + \sum_{k=2}^m \left( \prod_{i=k}^m U_i \right) (x_{k-1} - x_k) \right)
\]

\[
= : U_1 f(Y, U_2, \ldots, U_m) + g(U_2, \ldots, U_m).
\]

Let \( N \in \mathcal{B}(\mathbb{R}) \) be a given set such that \( \lambda(N) = 0 \), where \( \lambda \) denotes Lebesgue measure on the line. Then,

\[
P \left( X^{(B)} \in N \right) = E \left[ P \left( X^{(B)} \in N \mid Y, U_2, \ldots, U_m \right) \right].
\]
Note that \( f(Y, U_2, \ldots, U_m) = \left( \prod_{i=2}^{m} U_i \right) (Y - x_1) \neq 0 \) \( P \)-almost surely, since \( P(Y = x_j) = 0 \) for \( 1 \leq j \leq m \) by the definition of \( \nu = \mathcal{L}(Y) \) in (4) and also \( P(U_i = 0) = 0 \) for each \( i = 1, 2, \ldots, m \). Thus, by independence of \( U_1 \) and \( (Y, U_2, \ldots, U_m) \) we have for each choice of \( y \in \mathbb{R} \setminus \{x_1, \ldots, x_m\} \) and \( u_2, \ldots, u_m \in (0, 1) \) that

\[
P(X^{(B)} \in N \mid Y = y, U_2 = u_2, \ldots, U_m = u_m) \\
= P(f(y, u_2, \ldots, u_m)U_1 + g(u_2, \ldots, u_m) \in N) \\
= P\left(U_1 \in \frac{N - g(u_2, \ldots, u_m)}{f(y, u_2, \ldots, u_m)} \right) \\
= P\left(U_1 \in \tilde{N}(y, u_2, \ldots, u_m) \right),
\]

(11)

where

\[
\tilde{N}(y, u_2, \ldots, u_m) = \frac{N - g(u_2, \ldots, u_m)}{f(y, u_2, \ldots, u_m)} = \left\{ \frac{x - g(u_2, \ldots, u_m)}{f(y, u_2, \ldots, u_m)} \mid x \in N \right\}.
\]

By the properties of Lebesgue measure it follows that

\[
\lambda(\tilde{N}(y, u_2, \ldots, u_m)) = \frac{1}{|f(y, u_2, \ldots, u_m)|} \lambda(N) = 0,
\]

so that we conclude that \( P(U_1 \in \tilde{N}(y, u_2, \ldots, u_m)) = 0 \), because \( U_1 \) has an absolutely continuous distribution. Hence, by (11) also

\[
P(X^{(B)} \in N \mid Y, U_2, \ldots, U_m) = 0 \text{ \( P \)-almost surely. Thus, from (10) we infer that} \]

\[P(X^{(B)} \in N) = 0. \text{ Hence, the distribution of } X^{(B)} \text{ is absolutely continuous with respect to } \lambda.
\]

Remark 2.3. With the notation of the above existence proof, one can easily deduce from the identity

\[
E[f(X^{(B)})] = \int_0^1 E[f(x_m + t(X^{(B)}(\hat{B}) - x_m))] mt^{m-1} dt,
\]

valid for bounded and measurable \( f \), and an easy change of variable that for \( m \geq 2 \) a density \( p \) of \( X^{(B)} \) is given by

\[
p(x) = m \int_0^1 \tilde{p}\left(x_m + \frac{x - x_m}{t}\right)t^{m-2} dt,
\]
where \( \tilde{p} \) is a density of \( X^{(B)} \). This observation may be used to derive density formulas iteratively, beginning with the case \( m = 1 \), see Remark 2.2 (d). It also gives rise to an inductive proof of absolute continuity of the distribution of \( X^{(B)} \).

For the zero-bias and the size-bias transformations it is known that if the distribution of the random variable \( X \) is a mixture of the distributions of certain variables \( X_s, s \in S \), then also the biased distribution of \( X \) is a mixture of the biased distributions of the \( X_s \) (see \[Gol10\] for the zero-bias case and \[AGK13\] for the size-bias case). This property easily generalizes to our situation. Thus, let \( X \) be independent of the family \( (X_s)_{s \in S} \), where \( I \sim \gamma \) and \( X_s \) has distribution \( \mu_s \) for each \( s \in S \). Then, \( X := X_I \) has distribution \( \gamma K \). For each \( s \in S \) let \( \alpha_s := E[B(X_s)(X_s - x_1) \cdots (X_s - x_m)] \) and assume that

\[
E|B(X)| = \int_S E|B(X_s)|d\gamma(s) < \infty \quad \text{and} \quad E|X^m B(X)| = \int_S E|X^m B(X_s)|d\gamma(s) < \infty. 
\]

From (12) and Remark 2.2 (f) we conclude that

\[
0 < \alpha := E[B(X)(X - x_1) \cdots (X - x_m)] = \int_S \alpha_s d\gamma(s) < \infty. 
\]

Further, for each \( s \in S \) let \( X^{(B)}_s \) have the generalized \( X_s - B \) biased distribution. Let \( J \) be independent of the family \( (X^{(B)}_s)_{s \in S} \) having distribution \( P(J \in A) := \int_A \frac{\alpha_s}{\alpha} d\gamma(s), A \in S. \)

**Proposition 2.4.** Under the above assumptions the variable \( X^{(B)} := X^{(B)}_J \) has the generalized \( X - B \) biased distribution.

*Proof.* The easy proof is quite standard: For \( F \in \mathcal{F}^m \) we have by Fubini’s theorem

\[
E[B(X)(F(X) - L_F(X))] = \int_S E[B(X_s)(F(X_s) - L_F(X_s))]d\gamma(s) \\
= \int_S \alpha_s E[F^{(m)}(X^{(B)}_s)]d\gamma(s) = \alpha \int_S \frac{\alpha_s}{\alpha} E[F^{(m)}(X^{(B)}_s)]d\gamma(s) \\
= \alpha \int_S E[F^{(m)}(X^{(B)}_s)]P(J \in ds) = \alpha E[F^{(m)}(X^{(B)}_j)].
\]
It is actually not strictly necessary to assume that $X_s$ satisfies the assumptions of Theorem 2.1 for each $s \in S$. In fact, from (12) and Remark 2.2 (f) it follows that $X_s^{(B)}$ exists for $\gamma$-a.e. $s \in S$. Letting the Radon-Nikodym derivative of the distribution of $J$ with respect to $\gamma$ be zero for all $s$ in the exceptional null set, the proof goes through as before.

3. Generalization to fewer than $m$ sign changes

Although Theorem 2.1 is already quite general, in practice it might happen that one would like the order $m$ of the derivative on the right hand side of (3) to be larger than the number, say $k$, of sign changes of the function $B$ on the left hand side of (3). For example, if $X$ is a nonnegative random variable with finite and non-zero expectation, then $X^e$ is said to have the equilibrium distribution with respect to $X$, if

$$E[f(X) - f(0)] = E[X]E[f'(X^e)]$$

holds for all Lipschitz-continuous functions $f$. Couplings with this distributional transformation were successfully used for exponential approximation by Stein’s method in [PR11b] and [PR11a]. Thus, it appears as if in (13) we would have $m = 1$ but $k = 0$, since $B = 1$. But, as it turns out, this distributional transformations is nevertheless covered by Theorem 2.1 by letting $B(x) := \text{sign}(x)$, for example. Then, as a function on $\mathbb{R}$, $B$ has exactly one sign change at $x_1 = 0$ and Theorem 2.1 may be invoked. Since $X$ was assumed nonnegative, this is not quite reflected in equation (13). However, there are cases of distributional transformations, which are used in practice and which are not quite covered by Theorem 2.1. For example, in their analysis of the rate of convergence for the distributional convergence of certain random sums of mean zero random variables to the Laplace distribution, in [PR12] the authors make sure that for each real valued random variable $X$ such that $P(X < 0) = P(X > 0) = 1/2$ and with $0 < E[X^2] < \infty$, there exists a unique distribution for a random variable $X^{(L)}$ such that

$$E[f(X) - f(0)] = \frac{1}{2}E[X^2]E[f''(X^{(L)})]$$

holds for all continuously differentiable functions $f$ with a Lipschitz derivative. As they used Theorem 2.1 of [GR05] with the distributional transformation given by $B(x) = \text{sign}(x)$ twice in a row, they had to make sure that the orthogonality assumptions of that theorem were satisfied. This is where the assumption $P(X < 0) = P(X > 0) = 1/2$ comes from. Invoking Theorem 2.1 instead, we are able to
prove the following statement, which immediately generalizes \([14]\) to the class of all mean zero \(X\) with finite second moment.

**Proposition 3.1.** Let \(X\) be a real-valued random variable such that \(0 < E[X^2] < \infty\). Then, there exists a unique distribution for a random variable \(\hat{X}\) such that

\[
E[ f(X) - f(0) - f'(0)X] = \frac{1}{2} E[X^2] E[f''(\hat{X})]
\]

holds for all continuously differentiable functions \(f\) with a Lipschitz derivative. Further, the distribution of \(\hat{X}\) is always absolutely continuous with respect to Lebesgue measure.

**Remark 3.2.** Using the transformation from Proposition 3.1, one could easily generalize the results from \([PR12]\) to random sums with general mean zero summands and even to summands with small, non-zero means.

**Proof of Proposition 3.1.** Uniqueness can be seen in a similar way as in the proof of Theorem 2.1. The existence proof is very similar to the proof of Theorem 3.4 in \([PR12]\): Let \(X\) and \(f\) be as in the statement of Proposition 3.1. Define the function \(B\) on \(\mathbb{R}\) by

\[
B(x) := \text{sign}(x) := \begin{cases} 
-1 & , x < 0 \\
0 & , x = 0 \\
1 & , x = 1 
\end{cases}
\]

having exactly one sign change at \(x_1 = 0\). Thus, since \(\alpha := E[B(X)(X - 0)] = E[|X|] \in (0, \infty)\), by Theorem 2.1, there exists a random variable \(\tilde{X}\) such that

\[
E[|X|] E[g'(\tilde{X})] = E[\text{sign}(X)(g(X) - g(0))]
\]

holds for all Lipschitz functions \(g\) on \(\mathbb{R}\). Now, since \(\frac{d}{dx}(\frac{x^2}{2}\text{sign}(x)) = |x|\), we have

\[
\beta := E[\tilde{X}\text{sign}(\tilde{X})] = E|\tilde{X}| = \frac{1}{\alpha} E[\text{sign}(X)\frac{1}{2}X^2\text{sign}(X)] = \frac{1}{2\alpha} E[X^2] \in (0, \infty).
\]

Thus, again by Theorem 2.1 there exists a random variable \(\hat{X}\) having the \(\hat{X} - B\) biased distribution. This means that
\[ \beta E[h'(\hat{X})] = E[\text{sign}(\hat{X})(h(\hat{X}) - h(0))] = E[\text{sign}(\hat{X})h(\hat{X})] - h(0)E[\text{sign}(\hat{X})] \]

(19) \[ = E[\text{sign}(\hat{X})h(\hat{X})] - h(0)\frac{1}{\alpha}E[X] \]

holds for all Lipschitz functions \( h \). Since \( X \) has finite second moment, one can easily see that (17) also holds for absolutely continuous functions \( g \) such that \( |g'(x)| \) is \( O(x) \) as \( |x| \rightarrow \infty \). In particular this holds for \( g(x) := \text{sign}(x)(f(x) - f(0)) \) with \( g'(x) = \text{sign}(x)f'(x) \) for \( x \neq 0 \). Thus, from (17), (18) and (19) we conclude that

\[ E[f(X) - f(0)] = E[\text{sign}(X)g(X)] = \alpha E[g'(\hat{X})] = \alpha E[\text{sign}(\hat{X})f'(\hat{X})] \]

(20) \[ = \frac{1}{2}E[X^2] E[f''(\hat{X})] + f'(0)E[X] , \]

proving (15). Absolute continuity of \( \mathcal{L}(\hat{X}) \) follows immediately from Theorem 2.1. \( \square \)

Next, we will use the result of Proposition 3.1 to give a generalization of Theorem 2.1 to cases, where the number \( k \) of sign changes of \( B \) may be smaller than the order \( m \) of the derivative we would like to have in the defining identity for the biased distribution. However, we will have to assume that \( k \equiv m \mod 2 \), i.e. that \( k \) and \( m \) have the same parity. In what follows, for nonnegative integers \( n, j \) we denote by \( (n)_j \) the falling factorial, i.e. \( (n)_0 := 1 \) and \( (n)_j := n(n-1)\ldots(n-j+1) \) if \( j \geq 1 \).

**Theorem 3.3.** Let \( k \leq m \) be nonnegative integers with the same parity and let \( B \) be a measurable function on \( \mathbb{R} \) having \( k \) sign changes at the points \( x_1 < x_2 < \ldots < x_k \) such that \( B(x) \geq 0 \) for all \( x \geq x_k \), if \( k \geq 1 \) and for all \( x \in \mathbb{R} \), if \( k = 0 \). Further, let \( X \) be a real-valued random variable such that \( E[|B(X)X^2|] < \infty \) for all \( 0 \leq j \leq m \) and such that

(21) \[ \alpha := \frac{1}{k!} E\left[B(X) \prod_{j=1}^{k} (X - x_j) \right] \neq 0 . \]

If \( k = 0 \), assume further that the generalized \( X - B \) biased distribution from Theorem 2.1 is not the Dirac measure at 0. Then, there exists a unique distribution for a random variable \( X^{(B,m)} \) such that

(22) \[ E\left[B(X)(F(X) - R_F(X) - L_F(X))\right] = \beta E\left[F^{(m)}(X^{(B,m)})\right] \]
holds for each $F \in \mathcal{F}^m$, where, with

$$a_i^{(j)} := \sum_{l=1}^{k} x_l^{k+j-i-1} \prod_{r \neq l} (x_l - x_r) = \sum_{\sum_{j=1}^{k} \alpha_j = j-i} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k} \quad (j \geq i \geq 0),$$

we define the polynomial $R_F$ by

$$R_F(x) := R_{F;x_1,\ldots,x_k}(x) := \prod_{j=1}^{k} (x - x_j) \sum_{i=0}^{m-k-1} \left( \sum_{j=i}^{m-1} \frac{F(k+j)(0)a_i^{(j)}}{(k+j)!} \right) x^i,$$

if $k \geq 1$ and by

$$R_F(x) := \sum_{j=0}^{m-1} \frac{F(j)(0)}{j!} x^j,$$

if $k = 0$. Then, $R_F$ is equal to zero, whenever $k = m$ and has degree at most $m-1$, if $k < m$. Furthermore, $L_F$ still denotes the interpolation polynomial for $F$ corresponding to the nodes $x_1,\ldots,x_k$ given by (2) but with $m$ replaced by $k$. Additionally, $\beta$ is always positive and is given by

$$\beta := \frac{1}{m!} E\left[B(X) \left( X^m - \sum_{l=1}^{k} x_l^m \prod_{r \neq l} \frac{X - x_r}{x_l - x_r} \right) \right]$$

if $k \geq 1$ and by $\beta = (m!)^{-1} E[B(X)X^m]$, if $k = 0$. Also, the distribution of $X^{(B,m)}$ is always absolutely continuous with respect to Lebesgue measure unless $k = m = 0$.

**Proof.** From Theorem 2.1 we know that $\alpha > 0$. Let $F \in \mathcal{F}^m$ be given. By the assumptions on $X$ one can conclude again from Theorem 2.1 that $E[B(X)(F(X) - L_F(X))]$ exists and that there is a random variable $Y$ having the generalized $X - B$ biased distribution, so that

$$E\left[B(X)(F(X) - L_F(X)) \right] = \alpha E\left[F^{(k)}(Y) \right].$$

From our assumption in the case $k = 0$ and from Theorem 2.1 for $k \geq 1$, we know that $Y$ is not almost surely equal to zero. Thus, if $m \geq k + 2$, by Proposition 3.1 we know that there is a random variable $Y_1$ satisfying
\[
E[F^{(k)}(Y)] = F^{(k)}(0) + F^{(k+1)}(0)E[Y]
+ E[F^{(k)}(Y) - F^{(k)}(0) - F^{(k+1)}(0)Y]
= F^{(k)}(0) + F^{(k+1)}(0)E[Y] + \beta_1 E[F^{(k+2)}(Y_1)],
\]
(28)

where \(\beta_1 = \frac{1}{2} E[Y^2]\). Now, if \(m \geq k + 4\), then again by Proposition 3.1 we can find a random variable \(Y_2\) such that

\[
E[F^{(k+2)}(Y_1)] = F^{(k+2)}(0) + F^{(k+3)}(0)E[Y_1]
+ E[F^{(k+2)}(Y_1) - F^{(k+2)}(0) - F^{(k+3)}(0)Y_1]
= F^{(k+2)}(0) + \frac{F^{(k+3)}(0)}{3! \beta_1} E[Y^3] + \beta_2 E[F^{(k+4)}(Y_2)],
\]
(29)

since \(E[Y_1] = \frac{1}{6 \beta_1} E[Y^3] = \frac{1}{3! \beta_1} E[Y^3]\) and with

\[
\beta_2 = \frac{1}{2} E[Y_2^2] = \frac{1}{2} \frac{1}{12 \beta_1} E[Y^4] = \frac{1}{4! \beta_1} E[Y^4].
\]

Rearranging (28) and (29) we find that

\[
E[F^{(k)}(Y)] = F^{(k)}(0) + F^{(k+1)}(0)E[Y] + \frac{F^{(k+2)}(0)}{2} E[Y^2] + \frac{F^{(k+3)}(0)}{3!} E[Y^3]
+ \frac{1}{4!} E[Y^4] E[F^{(k+4)}(Y_2)].
\]
(30)

Inductively, for \(l = 1, \ldots, \frac{m-k}{2}\) we find that there exists \(Y_l\) such that, with \(Y_0 := Y\) we have

\[
E[F^{(k+2l-2)}(Y_{l-1})] = F^{(k+2l-2)}(0) + F^{(k+2l-1)}(0)E[Y_{l-1}]
+ E[F^{(k+2l-2)}(Y_{l-1}) - F^{(k+2l-2)}(0) - F^{(k+2l-1)}(0)Y_{l-1}]
= F^{(k+2l-2)}(0) + \frac{F^{(k+2l-1)}(0)}{(2l-1)! \beta_{l-1}} E[Y^{2l-1}] + \beta_l E[F^{(k+2l)}(Y_l)],
\]
(31)

where

\[
\beta_l = \frac{1}{(2l)! \beta_1 \cdots \beta_{l-1}} E[Y^{2l}].
\]

Again by induction we find the following analog of (30):
(32) \[ E[F^{(k)}(Y)] = \sum_{j=0}^{m-k-1} \frac{F^{(k+j)}(0)}{j!} E[Y^j] + \frac{1}{(m-k)!} E[Y^{m-k}] E[F^{(m)}(Y^{m-k})] \]

Now note that for \( j = 0, 1, \ldots, m - k \) with the function \( F_j(x) := \frac{x^{k+j}}{(k+j)_k} \) we have from (27) that

(33) \[ E[Y^j] = E[F_j^{(k)}(Y)] = \frac{1}{\alpha} E[B(X)(F_j(X) - L_{F_j}(X))] \].

Clearly, \( Q_j(x) := F_j(x) - L_{F_j}(x) \) is a polynomial of degree \( k + j \) having the zeroes \( x_1 < \ldots < x_k \). Thus, there exists a polynomial \( q_j \) of degree \( j \) such that

\[ Q_j(x) = q_j(x) \prod_{l=1}^{k} (x - x_l) \]

Now, first suppose that \( k = 0 \). Then, we have \( F_j(x) = Q_j(x) = q_j(x) = x^j \). Thus, from (32) we can conclude that

\[ \alpha E[F(Y)] = \sum_{j=0}^{m-k-1} \frac{F^{(j)}(0)}{j!} E[B(X)X^j] \]

(34) \[ + \frac{1}{m!} E[B(X)X^m] E[F^{(m)}(Y^m)] \].

Letting \( X^{(B,m)} := Y^m \) the claim follows in the case \( k = 0 \) from (27) and (34). From now on, we will assume that \( k \geq 1 \). In order to find \( q_j \) in this case, we write

\[ x^{k+j} - L_{x^{k+j}} = x^{k+j} - \sum_{l=1}^{k} \prod_{r \neq l}^{k} \frac{x - x_r}{x_l - x_r} \]

(35) \[ = \prod_{r=1}^{k} (x - x_r) \sum_{i=0}^{j+k-1} x_i \prod_{l=1}^{k} \frac{x^{k+j-1-i}}{(x_l - x_r)} \]

(32) \[ E[F^{(k)}(Y)] = \sum_{j=0}^{m-k-1} \frac{F^{(k+j)}(0)}{j!} E[Y^j] + \frac{1}{(m-k)!} E[Y^{m-k}] E[F^{(m)}(Y^{m-k})] \]
the last identity because the left hand side is a polynomial of degree $j + k$ and, hence, the right hand side must also be. Thus, as a neat by-product we have proved that

\begin{equation}
\sum_{l=1}^{k} \frac{x_l^n}{\prod_{r\neq l}(x_l - x_r)} = 0 \quad \text{for all } n \leq k - 2.
\end{equation}

From (35) we conclude that $q_j$ is given by

\begin{equation}
q_j(x) = \frac{1}{(k + j)_k} \sum_{i=0}^{j} \left( \sum_{l=1}^{k} \frac{x_l^{k+j-1-i}}{\prod_{r\neq l}(x_l - x_r)} \right) x^i = \frac{1}{(k + j)_k} \sum_{i=0}^{j} a_i^{(j)} x^i.
\end{equation}

Hence, from (33) and (37) we find for $j = 0, 1, \ldots, m - k$ that

\begin{equation}
E[Y^j] = \frac{1}{\alpha(k + j)_k} \sum_{i=0}^{j} a_i^{(j)} E\left[B(X)X^i \prod_{l=1}^{k}(X - x_l)\right]
\end{equation}

Plugging this into (32) we arrive at

\begin{align*}
E[F^{(k)}(Y)] &= \sum_{j=0}^{m-k-1} \frac{F^{(k+j)}(0)}{j!} \frac{1}{\alpha(k + j)_k} \sum_{i=0}^{j} a_i^{(j)} E\left[B(X)X^i \prod_{l=1}^{k}(X - x_l)\right] \\
&\quad + \frac{1}{\alpha(m - k)!(m)_k} \sum_{i=0}^{m-k} a_i^{(m-k)} E\left[B(X)X^i \prod_{l=1}^{k}(X - x_l)\right] E\left[F^{(m)}(Y_{m-k})\right] \\
&= \frac{1}{\alpha} \sum_{i=0}^{m-k-1} E\left[B(X)X^i \prod_{l=1}^{k}(X - x_l)\right] \sum_{j=i}^{m-k-1} \frac{F^{(k+j)}(0)a_i^{(j)}}{(k + j)!} \\
&\quad + \frac{1}{\alpha m!} \sum_{i=0}^{m-k} a_i^{(m-k)} E\left[B(X)X^i \prod_{l=1}^{k}(X - x_l)\right] E\left[F^{(m)}(Y_{m-k})\right] \\
&= \frac{1}{\alpha} E\left[B(X)R_F(X)\right] \\
&\quad + \frac{1}{\alpha m!} \sum_{i=0}^{m-k} a_i^{(m-k)} E\left[B(X)X^i \prod_{l=1}^{k}(X - x_l)\right] E\left[F^{(m)}(Y_{m-k})\right]
\end{align*}

Now, from reading (35) backwards (with $m = k + j$) we obtain
\[ \sum_{i=0}^{m-k} a_i^{(m-k)} \prod_{j=1}^{k} (x - x_j) x^i = \sum_{i=0}^{m-k} \sum_{l=1}^{k} \frac{x_l^{m-1-i}}{\prod_{r \neq l} (x_l - x_r)} \prod_{j=1}^{k} (x - x_j) x^i \]

\[ = x^m - Lx^m = x^m - \sum_{l=1}^{k} x_l^m \prod_{r \neq l} \frac{x - x_r}{x_l - x_r}. \]

Thus, from (39) and (40) we see that

\[ E[F^{(k)}(Y)] = \frac{1}{\alpha} E[B(X)R_F(X)] + \frac{\beta}{\alpha} E[F^{(m)}(Y_{m-k})]. \]

Letting \( X^{(B,m)} := Y_{m-k} \) now follows from (27) and (41).

To see that \( \beta > 0 \), note that we know from our assumption in the case \( k = 0 \) and from Theorem 2.1 in the case \( k \geq 1 \) that \( Y \) cannot almost surely be equal to zero. Thus, the even moments of \( Y \) are also non-zero. Since we know from (32) that \( \beta = \frac{\alpha}{(m-k)!} E[Y^{m-k}] \) with \( \alpha > 0 \) and as \( m - k \) is even, it follows that also \( \beta > 0 \).

Knowing that \( \beta \) is necessarily positive, uniqueness of the distribution for \( X^{(B,m)} \) can be proved as for \( X^{(B)} \) in the proof of Theorem 2.1. Absolute continuity of \( \mathcal{L}(X^{(B,m)}) \) in the case that not both, \( m \) and \( k \) are equal to zero, now follows from Theorem 2.1 and Proposition 3.1. It remains to show the alternative representation for the numbers \( a_i^{(j)} \) in (23). This is given by Lemma 3.4.

**Lemma 3.4.** For \( k \geq 1 \) let \( x_1, \ldots, x_k \) be distinct real (or complex) numbers. Then, for each nonnegative integer \( n \) we have the identity

\[ \sum_{l=1}^{k} x_l^n \prod_{r \neq l} (x_l - x_r) = \sum_{(\alpha_1, \ldots, \alpha_k) \in \mathbb{N}^k_0} \prod_{j=1}^{k} x_\alpha_j. \]

**Proof.** We prove the claim by induction on \( k \), simultaneously for all \( n \geq 0 \). If \( k = 1 \), then it is clearly true. Now assume that \( k \geq 1 \) and that \( x_1, \ldots, x_k, x_{k+1} \) are distinct numbers. Then, we can write

\[ \sum_{l=1}^{k+1} x_l^n \prod_{r \neq l} (x_l - x_r) = \sum_{l=1}^{k} x_l^n \prod_{r \neq l} (x_l - x_r) + x_{k+1}^n \sum_{l=1}^{k+1} \frac{1}{\prod_{r \neq l} (x_l - x_r)} =: S_1 + S_2 \]

Noting that
\[ x_l^n - x_{k+1}^n = \left( x_l - x_{k+1} \right) \sum_{i=0}^{n-1} x_i^i x_{k+1}^{n-1-i} \]

we conclude from the induction hypothesis that

\[
S_1 = \sum_{l=1}^{k} \frac{(x_l - x_{k+1})}{\prod_{r=1}^{k+1}(x_l - x_r)} \sum_{i=0}^{n-1} x_i^i x_{k+1}^{n-1-i} = \sum_{i=0}^{n-1} x_k^{i+1} \sum_{l=1}^{k} \frac{x_l^i}{\prod_{r=1, r \neq l}^{k+1}(x_l - x_r)}
\]

\[
= \sum_{i=0}^{n-1} x_k^{i+1} \sum_{(\beta_1, \ldots, \beta_k) \in \mathbb{N}_0^k} \frac{x_{\beta_1}^1 x_2^{\beta_2} \cdots x_k^{\beta_k}}{\prod_{j=1}^{k} \beta_j = i - k + 1}
\]

Thus, it only remains to show that \( S_2 = 0 \). But this follows from (36), completing the proof. \( \square \)

**Remark 3.5.** (a) We may call the distribution of \( X^{(B,m)} \) the \( X - (B, m) \) biased distribution. Note, however, that, as for \( X^{(B)} \), the distribution of \( X^{(B,m)} \) is sensitive to the number \( k \) and the choice of the sign change points \( x_1 < \ldots < x_k \), if these are ambiguous (see Remark 2.2 (b)).

(b) It is easy to see that an analog of Proposition 2.4 also exists for the \( X - (B, m) \) biased distribution.

(c) One can see from examples that the condition that \( k \) and \( m \) have the same parity cannot be abandoned without substitution. In fact, if \( X \) has support equal to \( \mathbb{R} \), then one cannot find a random variable \( X^e \) such that (13) is satisfied for all Lipschitz \( f \), because it is easy to see that the corresponding distribution would need to have a density proportional to \( q(t) := 1_{[0, \infty)}(t)P(X > t) - 1_{(-\infty,0)}(t)P(X \leq t) \), which is negative for \( t < 0 \). Note that contrarily, if there is an \( x_1 \in \mathbb{R} \) such that \( X \geq x_1 \) almost surely (and \( E[X] > x_1 \)), then letting \( B(x) := \text{sign}(x - x_1) \) having one sign change at \( x_1 \), by Theorem 2.1 we find a random variable \( X^{(B)} \) such that \( E[f(X) - f(x_1)] = \alpha E[f'(X^{(B)})] \) with \( \alpha = E[X - x_1] \).

In view of Remark 3.5, (c) it would be nice to know, if for each real random variable \( X \) with \( E|X| < \infty \) we can always find another random variable \( Y \) and constants \( \beta > 0 \) and \( c_f, f \) Lipschitz on \( \mathbb{R} \), such that

\[
E[f(X) - c_f] = \beta E[f'(Y)]
\]
holds for each Lipschitz function $f$. By Remark 3.5 (c) this is true for all $X$, which are almost surely bounded below. Thus, only those $X$ with support equal to $\mathbb{R}$ must be considered to find a counterexample. Note that such a counterexample would imply that the condition that $k$ and $m$ in Theorem 3.3 have the same parity is also necessary, in general.

4. Analytical proof of the existence part in Theorem 2.1

**Lemma 4.1.** Let $F : \mathbb{R} \to \mathbb{R}$ be $m$-times differentiable for an integer $m \geq 0$ such that $f := F^{(m)} > 0$ on $\mathbb{R}$. Then, $F$ has at most $m$ zeroes.

*Proof.* We prove the claim by induction on $m$. Since $F^{(0)} = f > 0$ has no zeroes if $m = 0$, the assertion easily follows in this case. Now, let $m \geq 1$ and assume that the claim is true for $(m - 1)$-times differentiable functions. Suppose that $f$ has $m + 1$ distinct zeroes $y_1 < \ldots < y_m < y_{m+1}$. Then, by Rolle’s theorem there exist points $z_k \in (y_k, y_{k+1})$ such that $F'(z_k) = 0$ for $k = 1, \ldots, m$. Since the points $z_1, \ldots, z_m$ are necessarily pairwise distinct, this contradicts the induction hypothesis applied to $F'$.

**Lemma 4.2.** Let $m \geq 0$ be an integer and let $G \in C^m(\mathbb{R})$ be a function such that $f(x) := G^{(m)}(x) \geq \varepsilon$ for all $x \in \mathbb{R}$, where $\varepsilon > 0$. Then, for each fixed real number $a$ and all $x \geq a$ we have that

$$G(x) \geq \sum_{j=0}^{m-1} G^{(j)}(a) \frac{(x-a)^j}{j!} + \frac{\varepsilon}{m!} (x-a)^m.$$

Hence, for each polynomial $Q$ of degree at most $m-1$ it follows that

$$\lim_{x \to \infty} (G(x) + Q(x)) = +\infty.$$ 

*Proof.* The second assertion follows easily from the first one. We prove the first claim by induction on $m \geq 0$. If $m = 0$, then $G^{(m)}(x) = f(x) \geq \varepsilon$ for each $x \in \mathbb{R}$, which is the claim for $m = 0$. Now, assume that $m \geq 1$ and that the claim holds for $m - 1$. Then, from the fundamental theorem of calculus and the induction hypothesis we conclude that for all $x \geq a$
\[ G(x) = G(a) + \int_a^x G'(t)dt \]

\[ \geq G(a) + \int_a^x \left( \sum_{j=0}^{m-2} G^{(j+1)}(a) \frac{(t-a)^j}{j!} + \frac{\varepsilon}{(m-1)!}(t-a)^{m-1} \right) dt \]

\[ = G(a) + \sum_{j=0}^{m-2} G^{(j+1)}(a) \frac{(x-a)^{j+1}}{(j+1)!} + \frac{\varepsilon}{m!}(x-a)^m \]

\[ = \sum_{j=0}^{m-1} G^{(j)}(a) \frac{(x-a)^j}{j!} + \frac{\varepsilon}{m!}(x-a)^m. \]

\[ \square \]

Recall that for real numbers \( x_1 < x_2 < \ldots < x_m \) we let \( J_1 := (-\infty, x_1], J_k := (x_{k-1}, x_k] \) for \( 2 \leq k \leq m \) and \( J_{m+1} := (x_m, \infty) \).

**Lemma 4.3.** Let \( f : \mathbb{R} \to \mathbb{R} \) be a nonnegative continuous function and let \( x_1 < x_2 < \ldots < x_m \) be real numbers. Then, there is a unique function \( F \in C^m(\mathbb{R}) \) such that \( F^{(m)} = f \) and \( (-1)^{m+1-k}F(x) \geq 0 \) for all \( x \in J_k \) and each \( 1 \leq k \leq m+1 \).

**Proof.** We first prove the easier uniqueness claim. Let \( F_1 \) and \( F_2 \) be two such functions. Since \( F_1^{(m)} - F_2^{(m)} = f - f = 0 \) identically, we know that \( Q := F_1 - F_2 \) is a polynomial with degree at most \( m-1 \). By continuity we have \( Q(x_k) = F_1(x_k) - F_2(x_k) = 0 - 0 = 0 \) for each \( k = 1, 2, \ldots, m \). This implies that \( Q \) must be the zero polynomial, i.e. \( F_1 = F_2 \).

Now we turn to the existence of \( F \). We first assume that there is an \( \varepsilon > 0 \) such that \( f(x) \geq \varepsilon \) for each \( x \in \mathbb{R} \). We define the function \( G := I_m^m f \), where, for a real number \( a \), we let \( I_a^m f(x) := (I_a f)(x) := \int_a^x f(t)dt \) and \( I_a^m \) is the \( m \)-th iterate of the operator \( I_a \). Then, \( G \in C^m(\mathbb{R}) \) and \( G^{(m)} = f \). Furthermore, we have \( G(x_k) = 0 \) and one can easily see by induction on \( m \) that \( G(x) > 0 \) for all \( x \in J_{m+1} \). Since, in general, \( G(x_k) \neq 0 \) for \( 1 \leq k \leq m-1 \), we let \( L := L_{G;x_1,\ldots,x_m} \) be the unique (interpolation) polynomial of degree \( \leq m-1 \) such that \( L(x_k) = G(x_k) \) for \( 1 \leq k \leq m \) and define \( F := G - L \). Of course, it holds that \( F^{(m)} = G^{(m)} = f \). Further, by construction we have \( F(x_k) = G(x_k) - L(x_k) = 0 \) for all \( k = 1, \ldots, m \). By Lemma 4.1 we conclude that \( F \) has exactly the zeroes \( x_1, \ldots, x_m \). In particular, either \( F(x) > 0 \) for each \( x \in J_{m+1} \) or \( F(x) < 0 \) for each \( x \in J_{m+1} \). The second alternative being impossible by Lemma 4.2 and the intermediate value theorem we conclude that \( F \) is strictly positive on \( J_{m+1} \). Next, we make sure that \( F \) really changes signs at the points \( x_1, \ldots, x_m \). Since \( F(x_k) = 0 \) it is enough to show that \( F'(x_k) \neq 0 \) for each
Now, we only assume that $f \geq 0$ is nonnegative and for each $n \in \mathbb{N}$ we let $f_n := f + 1/n$. Then, for each $n \geq 1$, $f_n \geq n^{-1}$ satisfies the assumptions of the case just treated. Additionally, the sequence $(f_n)_{n \in \mathbb{N}}$ converges uniformly to $f$. This implies that $I_a f_n$ converges to $I_a f$ uniformly on compact intervals (for each $a \in \mathbb{R}$), yielding that also $G_n := I_{x_m}^n f_n$ converges to $G := I_{x_m}^n f$ uniformly on compacts. By the specific Lagrange form of the interpolation polynomial, one can easily see that also $L_{G_n;x_1,\ldots,x_m}$ converges pointwise to $L_{G;x_1,\ldots,x_m}$ as $n \to \infty$. Thus, letting $F_n := G_n - L_{G_n;x_1,\ldots,x_m}$ and $F := G - L_{G;x_1,\ldots,x_m}$ we know from the first case that $(-1)^{n+1-k} F_n(x) \geq 0$ for all $x \in J_k$ and each $1 \leq k \leq m+1$ and since $F_n(x) \xrightarrow{n \to \infty} F(x)$ for each $x \in \mathbb{R}$, the same applies to $F$.

**Lemma 4.4.** Let $B$ be a measurable biasing function on $\mathbb{R}$, having $m \in \mathbb{N}$ sign changes occurring at the points $x_1, < \ldots < x_m$ as above. Then, for each nonnegative, continuous function $f$ on $\mathbb{R}$, there exists a unique function $F$ on $\mathbb{R}$ such that $F^{(m)} = f$ and $F(x) \cdot B(x) \geq 0$ for all $x \in \mathbb{R}$. Furthermore, letting $G_f := G_{f;x_1,\ldots,x_m} := I_{x_m}^n f$ and denoting by $L_{G_f}$ the interpolation polynomial of degree at most $m-1$ corresponding to the function $G_f$ and to the nodes $x_1,\ldots,x_m$, we have that $F = G_f - L_{G_f}$.

**Proof.** This follows immediately from Lemma 4.3 and its proof.

**Analytical proof of Theorem 2.1.** From the first lines of the probabilistic existence proof, which are independent of the remainder of that proof, we already know that $\alpha = (m!)^{-1} E[B(X)(X - x_1) \cdot \ldots \cdot (X - x_m)] > 0$. Further, we concentrate on the non-trivial case that $m \geq 1$. We define the operator $T : C_c(\mathbb{R}) \to \mathbb{R}$ by

\[
\tag{43} T f := \frac{1}{\alpha m!} E \left[ B(X)(G_f(X) - L_{G_f}(X)) \right],
\]

with $G_f$ and $L_{G_f}$ as in the statement of Lemma 4.4. Since $\|f\|_{\infty} < \infty$ for $f \in C_c(\mathbb{R})$, $T$ is well-defined by the assumptions on $X$ and $B$. It is also easy to see that $T$ is linear. In order to invoke the Riesz representation theorem, we aim at showing that $T$ is also positive. Thus, let $f \in C_c(\mathbb{R})$ be nonnegative. By Lemma 4.4 we know that $F(x) \cdot B(x) \geq 0$ for all $x \in \mathbb{R}$, where $F(x) = G_f(x) - L_{G_f}(x)$. This immediately implies that $T f \geq 0$ and, hence, $T$ is a positive, linear operator on $C_c(\mathbb{R})$. By the Riesz representation theorem there exists a unique (positive) Radon measure $\nu$ on $(\mathbb{R}, B(\mathbb{R}))$ such that
\( T f = \int_{\mathbb{R}} f(x) d\nu(x) \) for each \( f \in C_c(\mathbb{R}) \).

In order to show that \( \nu \) is in fact a probability measure, we choose nonnegative functions \( f_n \in C_c(\mathbb{R}) \), \( n \geq 1 \), such that \( f_n \not \rightarrow 1 \) pointwise. Since the functions \( f_n \) are uniformly bounded (by 1), one can show similarly as in the proof of Lemma 4.3 that the \( G_{f_n} \) converge to \( G_1 \) pointwise as \( n \rightarrow \infty \) and one can show inductively that \( |G_{f_n}| \leq |G_1| \), where \( G_1(x) = \sum_{k=0}^{m} \frac{(x-x_k)^k}{k!} \). Thus, since \( B(X)G_1(X) \) is integrable by the assumptions of Theorem 2.1, we conclude from the dominated convergence theorem that

\[
\lim_{n \to \infty} T f_n = \lim_{n \to \infty} \frac{1}{\alpha m!} E \left[ B(X)(G_{f_n}(X) - L_{G_{f_n}}(X)) \right] = \frac{1}{\alpha m!} E \left[ B(X)(G_1(X) - L_{G_1}(X)) \right].
\]

Note that by construction \( Q(x) := G_1(x) - L_{G_1}(x) \) is a polynomial of degree \( m \) such that \( Q(x_k) = 0 \) for \( k = 1, \ldots, m \). Hence, there exists \( c \neq 0 \) such that \( Q(x) = c \prod_{k=1}^{m} (x-x_k) \). Since \( c = Q^{(m)} = G_1^{(m)} = 1 \), we conclude from (45) that

\[
\lim_{n \to \infty} T f_n = \frac{1}{\alpha m!} E \left[ B(X) \prod_{k=1}^{m} (X-x_k) \right] = \frac{\alpha m!}{\alpha m!} = 1.
\]

On the other hand, by the monotone convergence theorem and (44) we have

\[
\lim_{n \to \infty} T f_n = \lim_{n \to \infty} \int_{\mathbb{R}} f_n(x) d\nu(x) = \int_{\mathbb{R}} 1 d\nu = \nu(\mathbb{R}).
\]

From (46) and (47) we conclude that \( \nu \) is indeed a probability measure. Thus, we can choose a random variable \( X^{(B)} \) on some probability space with distribution \( \nu \). In order to show that \( X^{(B)} \) satisfies (3), we let \( F \in \mathcal{F}^m \) be given. Then, since \( F^{(m-1)} \) is Lipschitz, we know that \( f := F^{(m)} \) exists almost everywhere and is bounded. Let \( (f_n)_{n \geq 1} \) be a sequence in \( C_c(\mathbb{R}) \) converging to \( f \) pointwise such that \( \|f_n\|_\infty \leq \|f\|_\infty \) for all \( n \geq 1 \). Such a sequence can be constructed by convolution with suitable mollifiers with compact support, for example. Then, by an argument similar to that leading to (45), one can see, using (43), (44) and the dominated convergence theorem twice, that
\[ \alpha E\left[F^{(m)}(X^{(B)})\right] = \alpha E\left[f(X^{(B)})\right] = \alpha \lim_{n \to \infty} E\left[f_n(X^{(B)})\right] = \lim_{n \to \infty} E\left[B(X)(G_{f_n}(X) - L_{G_{f_n}}(X))\right] = E\left[B(X)(G_f(X) - L_{G_f}(X))\right]. \tag{48} \]

Now, it is easily seen by successive differentiation that
\[ F = G_f + T_{m-1,x_m}F, \]
where \( T_{m-1,x_m}F \) is the Taylor polynomial of order \( m - 1 \) around \( x_m \) corresponding to \( F \).

Since the interpolation polynomial of degree \( \leq m - 1 \) corresponding to \( T_{m-1,x_m}F \) is still \( T_{m-1,x_m}F \), this implies that
\[ L_{G_f} = L_{G_f} + L_{T_{m-1,x_m}F} = L_{G_f} - T_{m-1,x_m}F \]
and, hence,
\[ F - L_{G_f} = G_f + T_{m-1,x_m}F - \left(L_{G_f} - T_{m-1,x_m}F\right) = G_f - L_{G_f}. \tag{49} \]

From (48) and (49) it finally folllows that
\[ \alpha E\left[F^{(m)}(X^{(B)})\right] = E\left[B(X)(F(X) - L_{G_f}(X))\right], \]
which was to be proved. \( \square \)

References

[AG10] R. Arratia and L. Goldstein. Size bias, sampling, the waiting time paradox, and infinite divisibility: when is the increment independent? arXiv:1007.3910, 2010.

[AGK13] R. Arratia, L. Goldstein, and F. Kochman. Size bias for one and all. arXiv:1308.2729, 2013.

[BC05] A. D. Barbour and Louis H. Y. Chen, editors. An introduction to Stein’s method, volume 4 of Lecture Notes Series. Institute for Mathematical Sciences. National University of Singapore. Singapore University Press, Singapore, 2005. Lectures from the Meeting on Stein’s Method and Applications: a Program in Honor of Charles Stein held at the National University of Singapore, Singapore, July 28–August 31, 2003.

[CGS11] Louis H. Y. Chen, Larry Goldstein, and Qi-Man Shao. Normal approximation by Stein’s method. Probability and its Applications (New York). Springer, Heidelberg, 2011.

[Gol10] Larry Goldstein. Bounds on the constant in the mean central limit theorem. Ann. Probab., 38(4):1672–1689, 2010.

[GR96] Larry Goldstein and Yosef Rinott. Multivariate normal approximations by Stein’s method and size bias couplings. J. Appl. Probab., 33(1):1–17, 1996.

[GR97] Larry Goldstein and Gesine Reinert. Stein’s method and the zero bias transformation with application to simple random sampling. Ann. Appl. Probab., 7(4):935–952, 1997.

[GR05] Larry Goldstein and Gesine Reinert. Distributional transformations, orthogonal polynomials, and Stein characterizations. J. Theoret. Probab., 18(1):237–260, 2005.
[PR11a] Erol Peköz and Adrian Röllin. Exponential approximation for the nearly critical Galton-Watson process and occupation times of Markov chains. *Electron. J. Probab.*, 16:no. 51, 1381–1393, 2011.

[PR11b] Erol A. Peköz and Adrian Röllin. New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.*, 39(2):587–608, 2011.

[PR12] John Pike and Haining Ren. Stein’s method and the Laplace distribution. arXiv:1210.5775, 2012.