An existence theorem for bounds on the expectation of a random variable. Its opportunities for utility theories. V. 2

Alexander Harin

Modern University for the Humanities

4. October 2015

An existence theorem for bounds on the expectation of a random variable. Its opportunities for utility theories. V. 2

An existence theorem is proven for the case of a discrete random variable that can take on only a finite set of possible values. If the random variable takes on values in a finite interval and there is a lower non-zero bound on the modulus of (at least one) its central moment, then non-zero bounds on its expectation exist near the borders of the interval. The revealed bounds can be considered as "forbidden zones" for the expectation. They can be useful, e.g., in utility and prospect theories.

Alexander Harin
aaharin@gmail.com
Modern University for the Humanities

Contents

1. Introduction

Bounds on functions of random variables are considered in a number of works. At that, information of moments of random variables is used quite often. Bounds for probabilities and expectations of convex functions of discrete random variables with finite support are considered in Prékopa (1990). Inequalities on expectations of functions are considered in Prékopa (1992). The inequalities are based on the knowledge of moments of discrete random variables. A class of lower bounds on the expectation of a convex function using the first two moments of the random variable with a bounded support is considered in Dokov and Morton (2005).

Bounds on the exponential moments of $min(y, X)$ and $X1\{X < y\}$ using the first two moments of a random variable X are considered in Pinelis (2011).

Information of moments of a random variable can be used also for bounds on its expectation. This is done in the present article.

These bounds (or bounding inequalities) on the expectation of a random variable are expressed in terms of its minimal non-zero central moments (if such minimal non-zero central moments exist), in particular in terms of its minimal dispersion. The expectation and dispersion are sufficiently widespread in the probability theory to draw a conclusion of possible usefulness of these bounds.

A random variable, in itself, can play a part of the identical function and can be considered as a formal example of possible use of these bounds on the expectation of a random variable in the scope of the above bounds on functions. Linear functions of this variable can be also considered for such a use without essential modifications.

The dispersion is a common measure of a scattering. The scattering can be caused by noise and/or uncertainty, measurement errors, etc. So, one can suppose that the theorem can be used in practice in researches of the influence of a scattering of experimental data on the expectations of these data near the borders of intervals.

Sketches of versions of the theorem have at least partially explained some problems of utility and prospect theories, including the underweighting of high and the overweighting of low probabilities, risk aversion, the "four-fold pattern" paradox, etc. (see, e.g., Harin 2012), and have been used in the analysis of Prelec's probability weighting function at the probabilities $p \sim 1$ (see Steingrimsson and Luce, 2007, Aczél and Luce, 2007 and Harin, 2014).

Due to the convenience of abbreviations and to the history of creation and development of the topic of this article, the term "bound" is often referred to here as the term "restriction," especially in mathematical expressions.

2. Preliminary notes

In the present article, the first and simplest case of a discrete random variable with finite support is considered. Other cases may be considered later.

Let us consider a discrete random variable X such that there is a probability space $(\Omega, Æ, P)$ and $X : \Omega \to R$. Let us suppose that

$$X = \{x_k\} : k = 1, 2, \ldots, K : 2 \le K < \infty$$

and

$$a \le x_k \le b : 0 < (b - a) < \infty$$

and the probability mass function is

$$f_X(x) = P(X = x) \equiv P(\{\omega \in \Omega : X(\omega) = x\}) \ .$$

Let us consider further the expectation of X

$$E(X) \equiv \sum_{k=1}^{K} x_k f_X(x_k) \equiv \mu \ ,$$

its central moments

$$E(X - \mu)^n = \sum_{k=1}^{K} (x_k - \mu)^n f_X(x_k)$$

and possible interrelationship between the expectation and moments.

3. Maximality

Let us search for the probability mass function $f_X(x)$ such that a central moment of X attains the maximal possible absolute value.

It is intuitively evident that the maximal possible absolute value of a central moment is obtained for the probability mass function, which is concentrated at the borders of the interval. Nevertheless, for the sake of mathematical rigor, this statement must be proven.

For the sake of simplicity, in this section, the probability mass function $f_X(x)$ will be used in a simplified form as $f(x) \equiv f_X(x)$.

3.1. Pairs

In the scope of this section, let us analyze the realizations x_k of the random variable X relative to μ.

Let us consider two possible realizations (points) x_a and x_b of the random variable X and the corresponding probabilities

$$f(x_a) \equiv f_X(x_a) \quad \text{and} \quad f(x_b) \equiv f_X(x_b).$$

For the purposes of this article, let us introduce a term "pair."

Sometimes, one may need to mark objects associated with pairs. Let us mark them by an additional subscript. To not confuse with the abbreviation of the term "probability," let us choose a subscript "C" ("couple").

Definition 3.1. Pair. Two realizations (points) x_a and x_b of the discrete random variable X, satisfying

$$a \le x_a \le \mu \le x_b \le b \ ,$$

will be called a "**pair**" (or a "couple")

$$X_{Pair} \equiv X_{Couple} \equiv X_C \equiv (x_a, x_b) \equiv (x_{C.a}, x_{C.b})$$

relative to μ if the balance

$$(\mu - x_a) f(x_a) = (x_b - \mu) f(x_b)$$

is true, in other words, if $\mu \equiv E(X)$ is the expectation of x_a and x_b as well. At that, if X may be considered as a set, then a pair may be considered as a subset X_C of the set X, having the same expectation μ as X.

Note, if $x_a = x_b$ then the balance can be also considered as true, though formally.

The sum of the probabilities $f(x_a)$ and $f(x_b)$ is assumed to be non-zero and (for the convenience of abbreviations, to not numerously use the long punctilious definition of the probability) can be named as the weight of the pair (couple) $w_{Pair} \equiv w_{Couple} \equiv w_C$ or simply w

$$w \equiv w_C \equiv f(x_a) + f(x_b) \equiv P(X = x_a) + P(X = x_b) > 0 \ .$$

The central moment $E_C(X_C\text{-}\mu)^n \equiv E_{Couple}(X_{Couple}\text{-}\mu)^n$ of this pair (couple) is

$$E_C(X_C - \mu)^n \equiv (x_a - \mu)^n f(x_a) + (x_b - \mu)^n f(x_b).$$

Its absolute value is limited by the sum of the absolute values of its components

$$\begin{aligned} | E_C(X_C - \mu)^n | &\le | (x_a - \mu)^n f(x_a) | + | (x_b - \mu)^n f(x_b) | = \\ &= (\mu - x_a)^n f(x_a) + (x_b - \mu)^n f(x_b) \end{aligned} .$$

3.2. Limiting function

Let us define a bounding function for a central moment of the pair. To not confuse the abbreviation of this function with the point b, this function will be named the limiting function L.

From the expressions of the balance and weight of the pair (couple)

$$f(x_b) = \frac{\mu - x_a}{x_b - \mu} f(x_A) = w_C - f(x_a)$$

and

$$f(x_a) + f(x_b) = \frac{\mu - x_a + x_b - \mu}{x_b - \mu} f(x_a) = \\ = \frac{x_b - x_a}{x_b - \mu} f(x_a) = w_C \quad ,$$

one may replace $f(x_a)$ and $f(x_b)$ by functions of only x_a, μ, x_b and w_C

$$f(x_a) = \frac{x_b - \mu}{x_b - x_a} w_C \quad \text{and} \quad f(x_b) = \frac{\mu - x_a}{x_b - x_a} w_C$$

and obtain

$$| E_C (X - \mu)^n | \le (\mu - x_a)^n f(x_a) + (x_b - \mu)^n f(x_b) = \\ = (\mu - x_a)^n \frac{x_b - \mu}{x_b - x_a} w_C + (x_b - \mu)^n \frac{\mu - x_a}{x_b - x_a} w_C \equiv \\ \equiv (\mu - x_a)^n \frac{x_b - \mu}{x_b - x_a} w + (x_b - \mu)^n \frac{\mu - x_a}{x_b - x_a} w \quad .$$

Definition 3.2. Limiting function. For the purposes of this article, one may define a **limiting function** $L_C(x_a, \mu, x_b, n, w_C)$ or, abbreviated, $L(x_a, \mu, x_b, n, w)$ or simply L_C or L for a central moment of a pair (couple). This function depends only on x_a, μ, x_b, n, w_C

$$L_{Couple}(x_a, \mu, x_b, n, w_{Couple}) \equiv L_C(x_a, \mu, x_b, n, w_C) \equiv L(x_a, \mu, x_b, n, w) \equiv \\ \equiv (\mu - x_a)^n \frac{x_b - \mu}{x_b - x_a} w_C + (x_b - \mu)^n \frac{\mu - x_a}{x_b - x_a} w_C \quad .$$

Note, here x_a and x_b are variables, but μ, n, and w_C are parameters.

The absolute value of a central moment, say $|E_C(X_C-\mu)^n|$, of the pair (couple) is, by definition, limited (bounded) by this limiting function

$$| E_C (X_C - \mu)^n | \le L_C(x_a, \mu, x_b, n, w_C) \quad .$$

3.3. Search for the maximum. Derivatives

Let us find the maximum of the limiting function $L_C(x_a, \mu, x_b, n, w_C)$ for x_a and x_b.

3.3.1. Differentiation with respect to x_a

Let us differentiate $L(x_a, \mu, x_b, n, w)$ with respect to x_a

$$\frac{\partial L(x_a,\mu,x_b,n,w)}{\partial x_a} =$$

$$= \frac{\partial\left((\mu - x_a)^n \frac{x_b-\mu}{x_b-x_a} w + (x_b-\mu)^n \frac{\mu - x_a}{x_b - x_a} w \right)}{\partial x_a} =$$

$$= \{[-n(x_b-x_a)+(\mu-x_a)](\mu-x_a)^{n-1} + $$

$$+[-(x_b-x_a)+(\mu-x_a)](x_b-\mu)^{n-1}\}\frac{x_b-\mu}{(x_b-x_a)^2}w =$$

$$= \{[(\mu-x_a)-n(x_b-x_a)](\mu-x_a)^{n-1}-(x_b-\mu)^n\}\frac{x_b-\mu}{(x_b-x_a)^2}w$$

If $n \geq 1$ and $(\mu\text{-}x_a)<(x_b\text{-}x_a)$, that is, if $x_b>\mu$ and $x_b\text{-}x_a>0$, then

$$(\mu - x_a) - n(x_b - x_A) < 0$$

and

$$\frac{\partial L(x_a,\mu,x_b,n,w)}{\partial x_a} < 0 \, .$$

So, at $n \geq 1$, for $\mu < x_b \leq b$ (and, as can easily be seen, for $a \leq x_a < \mu$) the first derivative with respect to x_a is strictly less than zero. That is, we have

$$L(a,\mu,x_b,n,w) > L(x_a,\mu,x_b,n,w)$$

for $a \leq x_a < \mu < x_b \leq b$ or for $[a, b]$ except for the specific point μ.

If $(\mu\text{-}x_a)=(x_b\text{-}x_a)$, that is, if $x_b=\mu$, then from

$$(\mu - x_a)f(x_a) = (x_b - \mu)f(x_b) \ ,$$

we obtain

$$(\mu - x_a) = (\mu - \mu)\frac{f(x_b)}{f(x_a)} = 0$$

or $x_a = \mu$.

To include the specific point μ into the ranges of variation of the arguments x_a and x_b of the inequality

$$L(a,\mu,x_b,n,w) > L(x_a,\mu,x_b,n,w) \ ,$$

let us estimate the derivative $\partial L(x_a, \mu, x_b, n, w)/\partial x_a$ for both $x_a \rightarrow \mu$ and $x_b \rightarrow \mu$. One may impose some natural conditions of non-zero values of probabilities: $f(x_a)>0$ and $f(x_b)>0$.

Let, say, μ-x_a be the basic term. Then

$$(x_b - \mu) = \frac{f(x_a)}{f(x_b)}(\mu - x_a)$$

and

$$x_b - x_a = (x_b - \mu) + (\mu - x_a) =$$
$$= \left(\frac{f(x_a)}{f(x_b)} + 1\right)(\mu - x_a) = \frac{w}{f(x_b)}(\mu - x_a) \ .$$

If $x_a \rightarrow \mu$ then the derivative

$$\{[(\mu - x_a) - n(x_b - x_a)](\mu - x_a)^{n-1} - (x_b - \mu)^n\}\frac{x_b - \mu}{(x_b - x_a)^2} w =$$
$$= \left\{\left[1 - n\frac{w}{f(x_b)}\right] - \left(\frac{f(x_a)}{f(x_b)}\right)^n\right\}(\mu - x_a)^n \frac{f(x_a)}{f(x_b)}\left(\frac{f(x_b)}{w}\right)^2 w(\mu - x_a)^{-1} = .$$
$$= \left\{\left[1 - n\frac{w}{f(x_b)}\right] - \left(\frac{f(x_a)}{f(x_b)}\right)^n\right\}\frac{f(x_a)f(x_b)}{w}(\mu - x_a)^{n-1} \xrightarrow[n>1;\ x_a \to \mu]{} 0$$

So (at $n>1$, if μ-x_a tends to 0, then the derivative)

$$\frac{\partial L(x_a,\mu,x_b,n,w)}{\partial x_a} \xrightarrow[n>1;\ x_a \to \mu]{} 0 \ .$$

Therefore, for $a \le x_a \le \mu \le x_b \le b$, the derivative $\partial L(x_a, \mu, x_b, n, w)/\partial x_a \le 0$.

Let us include the point μ into the ranges of variation of the arguments x_a and x_b of the inequality $L(a, \mu, x_b, n, w)>L(x_a, \mu, x_b, n, w)$. Let us consider an intermediate point, say $x_a=(a+\mu)/2$.

If, for $a \le x_a \le \mu \le x_b \le b$, the derivative $\partial L(x_a, \mu, x_b, n, w)/\partial x_a \le 0$, then, for $a \le x_a \le \mu \le x_b \le b$, the function $L(x_a, \mu, x_b, n, w) \ge L(\mu, \mu, x_b, n, w)=L(\mu, \mu, \mu, n, w)$ (and $L((a+\mu)/2, \mu, x_b, n, w) \ge L(\mu, \mu, \mu, n, w)$).

If, for $a \le x_a<\mu<x_b \le b$, the derivative $\partial L(x_a, \mu, x_b, n, w)/\partial x_a<0$ then, for $a<x_a<\mu<x_b \le b$, the function $L(a, \mu, x_b, n, w)>L(x_a, \mu, x_b, n, w)$ and $L(a, \mu, x_b, n, w)>L((a+\mu)/2, \mu, x_b, n, w)$.

Therefore,

$$L(a,\mu,x_b,n,w) > L\left(\frac{a+\mu}{2},\mu,x_b,n,w\right) \ge L(\mu,\mu,\mu,n,w)$$

or

$$L_C(a,\mu,x_b,n,w) > L_C(\mu,\mu,\mu,n,w) \ .$$

We have included the specific point μ into the ranges of variation of arguments of the inequality $L(a, \mu, x_b, n, w)>L(x_a, \mu, x_b, n, w)$ and the inequality is true for $a \le x_a \le \mu \le x_b \le b$.

So, at $n>1$, the limiting function $L_C(x_a, \mu, x_b, n, w_C)$ has the maximum

$$Max(L_C(x_a,\mu,x_b,n,w_C)) = L_C(a,\mu,x_b,n,w_C) \ .$$

for x_a for the total interval $[a, b]$.

3.3.2. Differentiation with respect to x_b

Let us differentiate $L(x_a, \mu, x_b, n, w)$ with respect to x_b

$$\frac{\partial L(x_a,\mu,x_b,n,w)}{\partial x_b}=$$

$$=\frac{\partial\left((\mu-x_a)^n \dfrac{x_b-\mu}{x_b-x_a}w+(x_b-\mu)^n \dfrac{\mu-x_a}{x_b-x_a}w \right)}{\partial x_b}=$$

$$=\{[(x_b-x_a)-(x_b-\mu)](\mu-x_a)^{n-1}+$$

$$+[n(x_b-x_a)-(x_b-\mu)](x_b-\mu)^{n-1}\}\frac{\mu-x_a}{(x_b-x_a)^2}w=$$

$$=\{(\mu-x_a)^n+[n(x_b-x_a)-(x_b-\mu)](x_b-\mu)^{n-1}\}\frac{\mu-x_a}{(x_b-x_a)^2}w$$

At $n\geq 1$, if $(x_b-x_a)>(x_b-\mu)$, that is, if $x_a<\mu$, then

$$n(x_b-x_a)-(x_b-\mu)>0$$

and (if $x_b-x_a>0$)

$$\frac{\partial L(x_a,\mu,x_b,n,w)}{\partial x_b}>0\,.$$

If $(x_b-x_a)=(x_b-\mu)$, that is, if $x_a=\mu$, then $x_b=\mu$ (see above).

So, at $n\geq 1$, for $a\leq x_a<\mu<x_b<b$ the first derivative with respect to x_b is strictly greater than zero. That is, we have

$$L(x_a,\mu,x_b,n,w)<L(x_a,\mu,b,n,w)\ .$$

for $a\leq x_a<\mu<x_b\leq b$ or for $[a, b]$ except for the specific point μ.

To include the specific point μ into the ranges of variation of the arguments x_a and x_b, let us estimate the derivative $\partial L(x_a, \mu, x_b, n, w)/\partial x_b$ for both $x_b\rightarrow\mu$ and $x_a\rightarrow\mu$ under the same natural conditions of non-zero values of probabilities: $f(x_a)>0$ and $f(x_b)>0$.

Let, say, $x_b-\mu$ be the basic term. Then

$$(\mu-x_a)=\frac{f(x_b)}{f(x_a)}(x_b-\mu)$$

and

$$x_b-x_a=\left(1+\frac{f(x_b)}{f(x_a)}\right)(x_b-\mu)=\frac{w}{f(x_a)}(x_b-\mu)\ .$$

If $x_b \to \mu$, then the derivative

$$\{(\mu - x_a)^n + [n(x_b - x_a) - (x_b - \mu)](x_b - \mu)^{n-1}\}\frac{\mu - x_a}{(x_b - x_a)^2} w =$$

$$= \left\{\left(\frac{f(x_b)}{f(x_a)}\right)^n + \left[n\frac{w}{f(x_a)} - 1\right]\right\}(x_b - \mu)^{n-1}\frac{f(x_b)}{f(x_a)}\left(\frac{f(x_a)}{w}\right)^2 w = \quad .$$

$$= \left\{\left(\frac{f(x_b)}{f(x_a)}\right)^n + \left[n\frac{w}{f(x_a)} - 1\right]\right\}\frac{f(x_b)f(x_a)}{w}(x_b - \mu)^{n-1} \xrightarrow[n>1;\ x_b \to \mu]{} 0$$

So (for $n>1$, if x_b (and x_a) tend to μ, then)

$$\frac{\partial L(x_a, x_b, x_b, n, w)}{\partial x_b} \xrightarrow[n>1;\ x_b \to x_b]{} 0 \ .$$

Therefore, for $a \le x_a \le \mu \le x_b \le b$, the derivative $\partial L(x_a, \mu, x_b, n, w)/\partial x_b \ge 0$.

Let us include the specific point μ into the ranges of variation of the arguments x_a and x_b of the inequality $L(x_a, \mu, b, n, w) > L(x_a, \mu, x_b, n, w)$. Let us consider an intermediate point, say $x_b = (\mu+b)/2$.

If, for $a \le x_a \le \mu \le x_b \le b$, the derivative $\partial L(x_a, \mu, x_b, n, w)/\partial x_b \ge 0$ then, for $a \le x_a \le \mu \le x_b \le b$, the function $L(x_a, \mu, \mu, n, w) = L(\mu, \mu, \mu, n, w) \le L(x_a, \mu, x_b, n, w)$ (and $L(\mu, \mu, \mu, n, w) \le L((x_a, \mu, (\mu+b)/2, n, w))$.

If, for $a \le x_a < \mu < x_b \le b$, the derivative $\partial L(x_a, \mu, x_b, n, w)/\partial x_b > 0$ then, for $a \le x_a < \mu < x_b < b$, the function $L(x_a, \mu, x_b, n, w) < L(x_a, \mu, b, n, w)$ and $L((a+\mu)/2, \mu, x_b, n, w) < L(x_a, \mu, b, n, w)$.

Therefore,

$$L(\mu, \mu, \mu, n, w) \le L\left(x_a, \mu, \frac{a+\mu}{2}, n, w\right) < L(x_a, \mu, b, n, w)$$

or

$$L_C(\mu, \mu, \mu, n, w) < L_C(x_a, \mu, b, n, w) \ .$$

We have included the specific point μ into the ranges of variation of arguments of the inequality $L(x_a, \mu, x_b, n, w) < L(x_a, \mu, b, n, w)$ and the inequality is true for $a \le x_a \le \mu \le x_b \le b$.

So, at $n>1$, the limiting function $L_C(x_a, \mu, x_b, n, w_C)$ has the maximum

$$Max(L_C(x_a, \mu, x_b, n, w_C)) = L_C(x_a, \mu, b, n, w_C) \ .$$

for x_b for the total interval $[a, b]$.

3.3.3. The maximum

So, at $n>1$, for $a \le x_a \le \mu \le x_b \le b$, the limiting function

$$L_C(x_a, \mu, x_b, n, w) = (\mu - x_a)^n \frac{x_b - \mu}{x_b - x_a} w + (x_b - \mu)^n \frac{\mu - x_a}{x_b - x_a} w$$

attains its maximum at the borders $x_a = a$ and $x_b = b$ of the interval $[a, b]$

$$Max(L_C(x_a, \mu, x_b, n, w_C)) = L_C(a, \mu, b, n, w_C) = \\ = (\mu - a)^n \frac{b - \mu}{b - a} w_C + (b - \mu)^n \frac{\mu - a}{b - a} w_C .$$

So, at $n>1$, the absolute value $|E_{Couple}(X-\mu)^n| \equiv |E_C(X-\mu)^n|$ of a central moment of the pair (couple) (x_a, x_b) is limited by the maximal limiting function L_C, that is concentrated at the borders $x_a=a$ and $x_b=b$ of the interval $[a, b]$

$$| E_C(X - \mu)^n | \le L_C(a, \mu, b, n, w_C) = \\ = (\mu - a)^n \frac{b - \mu}{b - a} w_C + (b - \mu)^n \frac{\mu - a}{b - a} w_C .$$

3.4. Representation by pairs. Succession of situations

3.4.1. Preliminary considerations

Let us analyze whether the total probability (weight)

$$W_K \equiv \sum_{k=1}^{K} f(x_k) \equiv P(\Omega),$$

and central moments

$$E(X-\mu)^n = \sum_{k=1}^{K} (x_k - \mu)^n f(x_k).$$

of the variable X of Section 2 can be exactly represented by those of a set of pairs.

The final goal of this section is to exactly represent the modulus of any central moment of the variable X of Section 2 by a sum of moduli of central moments of a set of pairs of the same variable and to estimate this sum by the limiting functions.

The discrete random variable X can be treated as a set of points $\{x_k\}$. The probability mass function f of Section 2 can be also treated as a set of values $\{f(x_k)\}$ associated with $\{x_k\}$. A pair (x_a, x_b) defined in this section is a subset of the set $\{x_k\}$. If there are $K.C : K.C \geq 1$ pairs then, if there is a need, one can denote the $k.C^{th}$ pair (couple), such that $k.C \in [1, K.C]$, as $\{x_{k.C.a}, x_{k.C.b}\}$. The weight of this pair can be denoted as $w_{k.C}$. (The multiple notation, e.g. $x_{k.C.a}$, is used to avoid numerous three-storey and even four-storey indices in the text).

In this subsection we should distinguish between objects, characteristics, etc., which are associated with pairs, and objects, characteristics, etc., which are (still) not associated with pairs. To do this, let us denote objects, characteristics, etc., which are associated with pairs, as objects of pairs, pairs' characteristics, etc. Let us also denote objects, characteristics, etc., which are (still) not associated with pairs, as **original** ones.

Linearity of sums

Let us mention the linearity of sums of weights and moments.

The total weight

$$W_K = \sum_{k=1}^{K} f(x_k),$$

and moments

$$E(X - x_0)^n = \sum_{k=1}^{K} (x_k - x_0)^n f(x_k).$$

of X depend linearly on the values $f(x_k)$. The sum is their linear function as well. Therefore:

1) the total weight of a sum equals the sum of the (constituent) weights

and

2) the moment of a sum equals the sum of the moments.

The sum of the central moments of pairs is limited by the sum of the maximal limiting functions (those are linear functions of $f(x_k)$ as well) of these pairs. One can see, indeed, that if for $k.C^{th}$ pair

$$| E_{k.C}(X_{k.C} - \mu)^n | \le L_{k.C}(a, \mu, b, n, w_{k.C}) \ ,$$

then for $K.C$ pairs

$$\sum_{k.C=1}^{K.C} | E_{k.C}(X_{k.C} - \mu)^n | \le \sum_{k.C=1}^{K.C} L_{k.C}(a, \mu, b, n, w_{k.C}) \ .$$

3.4.2. Situations

Let us divide the points x_k into three groups:

1) $x_{k.a}<\mu$,
2) $x_{k.\mu}=\mu$ (zero central moment(s)),
3) $x_{k.b}>\mu$.

Let us introduce the numbers $K.a$, $K.\mu$ and $K.b$, such that $k.a \leq K.a$, $k.\mu \leq K.\mu$, $k.b \leq K.b$ and

$$K.a + K.\mu + K.b = K .$$

Owing to $x_{k.\mu} - \mu \equiv 0$, an arbitrary non-zero central moment depends only on $K.a$ and $K.b$. Let us consider in turn situations with various numbers

$$K.ab \equiv K.a + K.b .$$

from $K.ab = 0$ to the general situation.

Situations $K.ab = 0$ and $K.ab = 1$

Evidently (in more detail see Harin 2015), Situations $K.ab = 0$ and $K.ab=1$ (if they exist) do not contribute to the non-zero central moments.

Further, as a rule, we shall not consider the cases those do not contribute to the non-zero central moments, namely $x_k : f(x_k)=0$ and $x_k = \mu$.

Situation $K.ab=2$

Here, the only possible case, which contributes to the non-zero central moments, is the case $K.a=1$ and $K.b=1$.

If $K.a=1$ and $K.b=1$, then we have the balance

$$(\mu - x_{1.a}) f(x_{1.a}) = (x_{1.b} - v) f(x_{1.b}) \ .$$

Therefore, the original points $x_{1.a}$ and $x_{1.b}$ are the required pair of the previous subsections.

Evidently, the total weight and moments of the pair are equal to those of the original points.

So, the original total weight and moments of Situation $K.ab=2$ can be exactly represented by the total weight and moments of a pair of the previous subsections.

Remark 3.3

Let us further, for definiteness, enumerate the points $x_{k.a}$ and $x_{k.b}$, for example, from those furthest from μ, to those closest to μ.

Divided sets

Let us define “divided” or “exactly divided” sets.

Definition 3.4. Let us suppose given an initial set of points $\{x_k\}$ and the initial set of values $\{f(x_k)\}$ associated with $\{x_k\}$ as in Section 2.

A divided or exactly **divided set of points** $\{x_k\}$ (with respect to the **initial set of points**) is defined as the same initial set of points $\{x_k\}$ such that at least one value $f(x_k)$ (associated with a point x_k) is divided into, at least, two parts $f_1(x_k)$ and $f_2(x_k)$ satisfying the equality

$$f(x_k) = f_1(x_k) + f_2(x_k) \ .$$

A divided or exactly **divided set of values** (with respect to the **initial set of values**) is the set of values associated with the divided set of points.

The notation of a divided value may be more complex, e.g.

$$f(x_k) \equiv f_{1(k)}(x_k) + f_{2(k)}(x_k)$$

or, more generally,

$$f(x_k) \equiv \sum_{d(k)=1}^{D(k)} f_{d(k)}(x_k) \quad : \ 2 \leq D(k) < \infty \ .$$

More generally, every value $f(x_k)$ (that will be either divided or not divided) of the initial set of values $\{f(x_k)\}$ may be written via the values $f_{d(k)}(x_k)$ of the exactly divided set $\{f_{d(k)}(x_k)\}$, by definition, as

$$f(x_k) \equiv \sum_{d(k)=1}^{D(k)} f_{d(k)}(x_k) \quad : \ 1 \leq D(k) < \infty \ .$$

Note, the divided set of points and the initial set of points are the same sets. The divided set of values and the set of initial values differ from each other. Because of these properties, there is a reason to distinguish between divided and initial sets of points with the help of their associated sets of values.

Note, that a divided set of points can serve as the new initial set of points for a subsequent division.

Evidently, the total weight and moments of the divided set of points are equal to those of the initial set of points.

Let us consider the total weight and moments of a divided set of points.

By the definition, the total weight of a divided value $f_{d(k)}(x_k)$ is equal to the initial value $f(x_k)$

$$f(x_k) = \sum_{d(k)=1}^{D(k)} f_{d(k)}(x_k)$$

for every initial value $f(x_k)$. Therefore, the total weight of the divided set is equal to that of the initial set.

Both the divided values $f_{d(k)}(x_k)$ and the initial value $f(x_k)$ are associated with the same point x_k. Therefore and by the definition, the sum of moments of every divided point is equal to the moment of the initial point

$$\sum_{d(k)=1}^{D(k)} (x_k - x_0)^n f_{d(k)}(x_k) = (x_k - x_0)^n \sum_{d(k)=1}^{D(k)} f_{d(k)}(x_k) = (x_k - x_0)^n f(x_k) \ .$$

Therefore, the total moment of the whole divided set is equal to that of the whole initial set.

One can see, indeed, that, by definition, the total weight W_D of the divided set of points is

$$W_D \equiv \sum_{k=1}^{K} \sum_{d(k)=1}^{D(k)} f_{d(k)}(x_k) \equiv \sum_{k=1}^{K} f(x_k) \equiv W_K$$

and the total moment $E_D(X-x_0)^n$ of the divided set of points is

$$\begin{aligned} & E_D(X - x_0)^n \equiv \sum_{k=1}^{K} \sum_{d(k)=1}^{D(k)} (x_k - x_0)^n f_{d(k)}(x_k) \equiv \\ & \equiv \sum_{k=1}^{K} (x_k - x_0)^n \sum_{d(k)=1}^{D(k)} f_{d(k)}(x_k) \equiv \sum_{k=1}^{K} (x_k - x_0)^n f(x_k) \equiv \ . \\ & \equiv E(X - x_0)^n \end{aligned}$$

So, we have specified the properties of the divided sets: the total weight and moments of a divided set of points are equal to the total weight and moments of the initial set of points.

Situation $K.ab = 3$

Here, there are only two possible cases those can contribute to the non-zero central moments: the case of $K.a=2$ and $K.b=1$, or the case of $K.a=1$ and $K.b=2$.

Let us consider the case of $K.a=2$ and $K.b=1$.

Let us make the first step of the representation of the total weight and central moments of the original set of points by the total weight and central moments of a set of pairs.

The value $f(x_{1.b})$ can be exactly divided into two parts $f_1(x_{1.b})$ and $f_2(x_{1.b})$ satisfying the balance

$$(\mu - x_{1.a}) f(x_{1.a}) = (x_{1.b} - \mu) f_1(x_{1.b})$$

and the equality of divided sets

$$f_2(x_{1.b}) = f(x_{1.b}) - f_1(x_{1.b}) \ .$$

Here, the points $x_{1.a}$ and $x_{1.b}$ are the initial set of points. The divided points are the same points. The values $f(x_{1.a})$ and $f(x_{1.b})$ are the initial set of values. The values $f(x_{1.a})$, $f_1(x_{1.b})$ and $f_2(x_{1.b})$ are the divided set of values.

Due to the properties of the divided sets, the total weight and moments of the divided set of points are equal to those of the initial set of points.

The first portion of the original set of points is the set $x_{1.a}$ and $x_{1.b}$ of the divided original set with the associated values $f(x_{1.a})$ and $f_1(x_{1.b})$. Since the balance

$$(\mu - x_{1.a}) f(x_{1.a}) = (x_{1.b} - \mu) f_1(x_{1.b})$$

is true, the two points $x_{1.a}$ and $x_{1.b}$ of the divided set with the associated values $f(x_{1.a})$ and $f_1(x_{1.b})$ are the required pair of the previous subsections. Therefore, the total weight and moments of the pair are equal to those of the first portion of the divided original set of points.

So, the first step of the representation has been done. The total weight and moments of the pair as of the first portion of the set of the pairs are equal to those of the first portion of the divided original set of points.

This can be seen in more detail for the central moments

$$\begin{aligned}
&E(X-\mu)^n = (x_{1.a}-\mu)^n f(x_{1.a}) + (x_{2.a}-\mu)^n f(x_{2.a}) + (x_{1.b}-\mu)^n f(x_{1.b}) = \\
&= (x_{1.a}-\mu)^n f(x_{1.a}) + (x_{1.b}-\mu)^n f_1(x_{1.b}) + \\
&+ (x_{2.a}-\mu)^n f(x_{2.a}) + (x_{1.b}-\mu)^n f_2(x_{1.b}) = \\
&= E_{1.C}(X-\mu)^n + \\
&+ (x_{2.a}-\mu)^n f(x_{2.a}) + (x_{1.b}-\mu)^n f_2(x_{1.b})
\end{aligned} \ .$$

As a result of the first step, the number of unpaired values is diminished by one and we come to Situation $K.ab_{Diminished}=K.ab-1=2$.

Let us make the second step of the representation.

The initial balance

$$(\mu - x_{1.a}) f(x_{1.a}) + (\mu - x_{2.a}) f(x_{2.a}) = (x_{1.b} - \mu) f(x_{1.b})$$

remains

$$(\mu - x_{1.a}) f(x_{1.a}) + (\mu - x_{2.a}) f(x_{2.a}) =$$
$$= (x_{1.b} - \mu) f_1(x_{1.b}) + (x_{1.b} - \mu) f_2(x_{1.b})$$

and, subtracting the balance of the pair, we come to Situation *K.ab=2* for $f(x_{2.a})$ and $f_2(x_{1.b})$

$$(\mu - x_{2.a}) f(x_{2.a}) = (x_{1.b} - \mu) f_2(x_{1.b}) \ .$$

As has been proven above, the total weight and moments of Situation *K.ab=2* can be exactly represented by the total weight and moments of a pair of the previous subsections. So, the second step is the final one.

This can be seen in more detail for the central moments

$$E(X-\mu)^n = E_{1.C}(X-\mu)^n + (x_{2.a} - \mu)^n f(x_{2.a}) + (x_{1.b} - \mu)^n f_2(x_{1.b}) = \ .$$
$$= E_{1.C}(X-\mu)^n + E_{2.C}(X-\mu)^n \ .$$

So, the final step of the representation has been done. The total weight and moments of the final portion of the set of the pairs of points are equal to those of the final portion of the divided original set of points.

So, Situation *K.ab=3*, at *K.a=2* and *K.b=1*, can be represented by the sum of the first step and the final step.

So, the total weight and moments of the divided original set of points are equal to those of the initial original set of points. For every step, the total weight and moments of the portion of the set of the pairs are equal to those of the portion of the divided original set of points. Both the total weight and moments depend linearly on the values of the members of the sets. Therefore, the total weight and moments of the sum of the portions are equal to the sum of the constituent weights and moments correspondingly. Therefore, for whole Situation *K.ab=3*, at *K.a=2* and *K.b=1*, the total weight and moments of the set of the pairs are equal to those of the original set of points.

If *K.a=1* and *K.b=2*, then the consideration is analogous to the preceding one.

So, the total weight and moments of Situation *K.ab=3* can be exactly represented by the total weight and moments of a set of pairs of the previous subsections.

General Situation *K.ab*. Suppose general Situation $K.ab \geq 4$, $K.a \geq 1$ and $K.b \geq 1$ (the case of $K.a=0$ and $K.b \geq 1$ and the case of $K.b=0$ and $K.a \geq 1$ cannot exist or do not contribute to the non-zero central moments (in more detail see Harin, 2015)).

Let us consider $f(x_{1.a})$ and $f(x_{1.b})$. There are only two possible variants:
less possible but more easy Variant 1 (equality)

$$(\mu - x_{1.a}) f(x_{1.a}) = (x_{1.b} - \mu) f(x_{1.b}) \ .$$

and more possible but less easy Variant 2 (inequality)

$$(\mu - x_{1.a}) f(x_{1.a}) \neq (x_{1.b} - \mu) f(x_{1.b})$$

Let us make the first step of the representation of the total weight and moments of the original set of points by the total weight and moments of the set of the pairs. This first step may be implemented in one of the two forms depending on whether Variant 1 (equality) or Variant 2 (inequality) takes place.

Variant 1 (equality). Due to

$$(\mu - x_{1.a}) f(x_{1.a}) = (x_{1.b} - \mu) f(x_{1.b})$$

the two points $x_{1.a}$ and $x_{1.b}$ are the required pair of the previous subsections. Therefore, the total weight and moments of this first pair are the same as those of this first portion of the original set.

As a result of this first step within the scope of Variant 1 (equality), the number of unpaired values is diminished by two and from Situation *K.ab* we come to Situation $K.ab_{Diminished}=K.ab-2$. Here, the number $K.ab_{Diminished}=K.ab-2$ is composed of $2, \ldots, K.a$ and $2, \ldots, K.b$.

Let us make the first step of the representation within the scope of Variant 2 (inequality).

Variant 2 (inequality). If

$$(\mu - x_{1.a}) f(x_{1.a}) \neq (x_{1.b} - \mu) f(x_{1.b}) \ ,$$

then there are only two possible cases as well:

$$(\mu - x_{1.a}) f(x_{1.a}) < (x_{1.b} - \mu) f(x_{1.b})$$

and

$$(\mu - x_{1.a}) f(x_{1.a}) > (x_{1.a} - \mu) f(x_{1.a}) \ .$$

Suppose, for example, that

$$(\mu - x_{1.A}) f(x_{1.A}) < (x_{1.B} - \mu) f(x_{1.B}) \ .$$

Then one should divide the value $f(x_{1.b})$ into two parts $f_1(x_{1.b})$ and $f_2(x_{1.b})$ satisfying the balance

$$(\mu - x_{1.A}) f(x_{1.A}) = (x_{1.B} - \mu) f_1(x_{1.B})$$

and the equality of divided sets

$$f_2(x_{1.B}) = f(x_{1.B}) - f_1(x_{1.B}) \ .$$

Here, the points $x_{1.a}$ and $x_{1.b}$ are the initial set of points. The divided points are the same points. The values $f(x_{1.a})$ and $f(x_{1.b})$ are the initial set of values. The values $f(x_{1.a})$, $f_1(x_{1.b})$ and $f_2(x_{1.b})$ are the divided set of values.

Due to the properties of the divided sets, the total weight and moments of the divided set of points are equal to those of the initial set of points.

The first portion of the original set of points is the subset $(x_{1.a}, x_{1.b})$ of the divided original set with the associated values $f(x_{1.a})$ and $f_1(x_{1.b})$. Since the balance

$$(\mu - x_{1.a}) f(x_{1.a}) = (x_{1.b} - \mu) f_1(x_{1.b})$$

is true, two points $x_{1.a}$ and $x_{1.b}$ of the divided set with the values $f(x_{1.a})$ and $f_1(x_{1.b})$ are the required pair of the previous subsections. Therefore, the total weight and moments of the pair are equal to those of the first portion of the original set of points.

So, within the scope of Variant 2 (inequality), the first step of the representation has been done. The total weight and moments of the pair as of the first portion of the set of the pairs are equal to those of the first portion of the divided original set of points.

As a result of this first step within the scope of Variant 2 (inequality), the number of unpaired values is diminished by one (taking into account the part $f_2(x_{1.b})$ of the value $f(x_{1.b})$) and we come to Situation $K.ab_{Diminished}=K.ab-1$. Note, that the number $K.ab$ is composed of $1, \ldots, K.a$ and $1, \ldots, K.b$. And here, the number $K.ab_{Diminished}=K.ab-1$ is composed of $2, \ldots, K.a$ and $2, \ldots, K.b$ plus one.

If

$$(\mu - x_{1.a}) f(x_{1.a}) > (x_{1.a} - \mu) f(x_{1.a}) \ ,$$

then the consideration is analogous to the preceding one.

So, we have considered the first step of diminishing the number $K.ab$ for general Situation $K.ab\geq 4$ within the scopes of both parallel variants. It diminishes $K.ab$ by one or two.

Evidently, such a step may be a general intermediate one.

A general intermediate step

Evidently, a general intermediate step of the representation of the total weight and moments of the intermediate portion of the original set of points by the total weight and moments of the intermediate portion of the set of pairs is similar to the above first step.

For illustrativeness, an example of this general intermediate step may be written via formulae for the total weight and central moments.

Let us suppose a general intermediate situation such that there are already $g.C.ab : 1 \le g.C.ab < K.C.ab,$ pairs (couples) which represent the total weight and moments of some original points and there are still $K.a - g.a + 1$ of $x_{k.a}$ points and $K.b - g.b + 1$ of $x_{k.b}$ points, the total weight and moments of which are still not represented by those of pairs.

The total weight for the general intermediate situation before this general intermediate step can be represented as

$$W_{K.ab} = \sum_{k.C.ab=1}^{g.C.ab} w_{k.C.ab} + \sum_{k.a=g.a}^{K.a} f(x_{k.a}) + \sum_{k.b=g.b}^{K.b} f(x_{k.b}) \ .$$

The central moments for this general intermediate situation before the general intermediate step can be represented as

$$E(X-\mu)^n =$$
$$= \sum_{k.C.ab=1}^{g.C.ab} E_{k.C.ab}(X-\mu)^n + \sum_{k.a=g.a}^{K.a} (x_{k.a}-\mu)^n f(x_{k.a}) + \sum_{k.b=g.b}^{K.b} (x_{k.b}-\mu)^n f(x_{k.b}) \ .$$

Let us illustrate the general intermediate step for this general intermediate situation.

Variant 1 (equality). The general intermediate step can be seen in more detail for the total weights

$$W_{K.ab} = \sum_{k.C.ab=1}^{g.C.ab} w_{k.C.ab} + f(x_{g.a}) + f(x_{g.b}) + \sum_{k.a=g.a+1}^{K.a} f(x_{k.a}) + \sum_{k.b=g.b+1}^{K.b} f(x_{k.b}) =$$
$$= \sum_{k.C.ab=1}^{g.C.ab+1} w_{k.C.ab} + \sum_{k.a=g.a+1}^{K.a} f(x_{k.a}) + \sum_{k.b=g.b+1}^{K.b} f(x_{k.b}) \ .$$

The general intermediate step can be seen in more detail for the central moments

$$E(X-\mu)^n =$$
$$= \sum_{k.C.ab=1}^{g.C.ab} E_{k.C.ab}(X-\mu)^n + (x_{gk.a}-\mu)^n f(x_{g.a}) + (x_{g.b}-\mu)^n f(x_{g.b}) +$$
$$+ \sum_{k.a=g.a+1}^{K.a} (x_{k.a}-\mu)^n f(x_{k.a}) + \sum_{k.b=g.b+1}^{K.b} (x_{k.b}-\mu)^n f(x_{k.b}) =$$
$$= \sum_{k.C.ab=1}^{g.C.ab+1} E_{k.C.ab}(X-\mu)^n + \sum_{k.a=g.a+1}^{K.a} (x_{k.a}-\mu)^n f(x_{k.a}) + \sum_{k.b=g.b+1}^{K.b} (x_{k.b}-\mu)^n f(x_{k.b}) \ .$$

Variant 2 (inequality). The general intermediate step can be seen in more detail for the total weights

$$\begin{aligned} W_{K.ab} &= \sum_{k.C.ab=1}^{g.C.ab} w_{k.C.ab} + f(x_{g.a}) + f_{1(g)}(x_{g.b}) + f_{2(g)}(x_{g.b}) + \\ &+ \sum_{k.a=g.a+1}^{K.a} f(x_{k.a}) + \sum_{k.b=g.b+1}^{K.b} f(x_{k.b}) = \\ &= \sum_{k.C.ab=1}^{g.C.ab+1} w_{k.C.ab} + \sum_{k.a=g.a+1}^{K.a} f(x_{k.a}) + f_{2(g)}(x_{g.b}) + \sum_{k.b=g.b+1}^{K.b} f(x_{k.b}) \end{aligned} .$$

The general intermediate step can be seen in more detail for the central moments

$$\begin{aligned} E(X-\mu)^n &== \sum_{k.C.ab=1}^{g.C.ab} E_{k.C.ab}(X-\mu)^n + \\ &+ (x_{gk.a}-\mu)^n f(x_{g.a}) + (x_{g.b}-\mu)^n f_{1(g)}(x_{g.b}) + (x_{g.b}-\mu)^n f_{2(g)}(x_{g.b}) + \\ &+ \sum_{k.a=g.a+1}^{K.a} (x_{k.a}-\mu)^n f(x_{k.a}) + \sum_{k.b=g.b+1}^{K.b} (x_{k.b}-\mu)^n f(x_{k.b}) = \\ &= \sum_{k.C.ab=1}^{g.C.ab+1} E_{k.C.ab}(X-\mu)^n + \sum_{k.a=g.a+1}^{K.a} (x_{k.a}-\mu)^n f(x_{k.a}) + \\ &+ (x_{g.b}-\mu)^n f_{2(g)}(x_{g.b}) + \sum_{k.b=g.b+1}^{K.b} (x_{k.b}-\mu)^n f(x_{k.b}) \end{aligned} .$$

So, we have considered the general step of diminishing the number *K.ab* for general Situation *K.ab≥4* within the scopes of both parallel variants. It diminishes *K.ab* by one or two.

Evidently, this general step may be repeated as many times as needed to reach final Situations $K.ab_{Diminished}=3$ or $K.ab_{Diminished}=2$.

For every step, if the original set of points is divided, then the total weight and moments of the divided original set of points are equal to those of the initial original set of points. For every step, the total weight and moments of the portion of the set of the pairs are equal to those of the portion of the (divided) original set of points. Both the total weight and moments depend linearly on the values of the members of the sets. Therefore, the total weight and moments of the sum of the portions are equal to the sum of the constituent weights and moments correspondingly. Therefore, for whole general Situation *K.ab*, the total weight and moments of the set of the pairs are equal to those of the original set of points.

So, in general Situation *K.ab : K.ab≥4*, at *K.a≥1* and *K.b≥1*, the total weight and moments of a discrete random variable X of Section 2 may be exactly represented by the total weight and moments of the pairs of this section.

So, we have proven for the total weight

$$\sum_{k.a=1}^{K.a} f(x_{k.a}) + \sum_{k.b=1}^{K.b} f(x_{k.b}) = \sum_{k.C.ab=1}^{K.C.ab} w_{k.C.ab}$$

We have proven for the central moments

$$\sum_{k.a=1}^{K.a} (x_{k.a}-\mu)^n f(x_{k.a}) + \sum_{k.b=1}^{K.b} (x_{k.b}-\mu)^n f(x_{k.b}) = \sum_{k.C.ab=1}^{K.C.ab} E_{k.C.ab}(X-\mu)^n$$

and

$$| E(X-\mu)^n | \le \sum_{k.C.ab=1}^{K.C.ab} | E_{k.C.ab}(X-\mu)^n | \ .$$

3.5. General limitations

3.5.1. Weights

Let us consider the weights (probabilities) of groups of realizations (points) x_k of X, of groups of pairs and general limitations on them.

Remembering

$$K.a + K.\mu + K.b = K$$

of the preceding subsection, the total weights of these groups may be denoted as W_a, W_μ and W_b such that

$$W_a \equiv \sum_{k.a \le K.a} f(x_{k.a}) \ , \ W_\mu \equiv \sum_{k.\mu \le K.\mu} f(x_{k.\mu}) \ , \ W_b \equiv \sum_{k.b \le K.b} f(x_{k.b})$$

and the sum of the weights (probabilities) is

$$W_a + W_\mu + W_b = W_K \equiv P(\Omega) = 1 \ .$$

Let us denote the total weight of the total set of all the pairs (couples) as $W_{Couple} \equiv W_C$, the weight of the set of the formal pairs $\{\mu_{k.C.\mu}, \mu_{k+1.C.\mu}\}$ as $W_{C.\mu}$ and the total weight of the set of the pairs $\{x_{k.C.a}, x_{k.C.b}\}$ as $W_{C.ab}$. By this definition, the weight of, e.g., $k.C.ab^{\text{th}}$ pair (couple) $(x_{k.C.a}, x_{k.C.b})$, is denoted as $w_{k.C.ab}$ and

$$\sum_{k.C=1}^{K.C} w_{k.C} \equiv W_C \ , \quad \sum_{k.C.\mu \le K.C.\mu} w_{k.C.\mu} \equiv W_{C.\mu} \ , \quad \sum_{k.C.ab=1}^{K.C.ab} w_{k.C.ab} \equiv W_{C.ab} \ ,$$

and we have

$$W_C = W_{C.\mu} + W_{C.ab} \ .$$

Evidently,

$$W_{C.\mu} = W_\mu \ .$$

Due to the preceding subsection,

$$W_{C.ab} = W_a + W_b \ .$$

Therefore,

$$W_C = W_{C.\mu} + W_{C.ab} = W_K \equiv P(\Omega) = 1 \ .$$

3.5.2. The general limiting function

In the preceding subsection we have proven

$$|E(X-\mu)^n| \le \sum_{k.C.ab=1}^{K.C.ab} |E_{k.C.ab}(X-\mu)^n| \ .$$

The maximal limiting functions $L_{k.C.ab}(a, \mu, b, n, w_{k.C.ab})$, satisfying

$$|E_{k.C.ab}(X-\mu)^n| \le L_{k.C.ab}(a,\mu,b,n,w_{k.C.ab}) \ ,$$

allow estimating the central moments $E(X\text{-}\mu)^n$ of the random variable X

$$|E(X-\mu)^n| \le \sum_{k.C.ab=1}^{K.C.ab} |E_{k.C.ab}(X-\mu)^n| \le$$
$$\le \sum_{k.C.ab=1}^{K.C.ab} L_{k.C.ab}(a,\mu,b,n,w_{k.C.ab}) \ .$$

This estimate can be easily simplified. From

$$L_{k.C.ab}(a,\mu,b,n,w_{k.C.ab}) =$$
$$= (\mu-a)^n \frac{b-\mu}{b-a} w_{k.C.ab} + (b-\mu)^n \frac{\mu-a}{b-a} w_{k.C.ab} =$$
$$= \left[(\mu-a)^n \frac{b-\mu}{b-a} + (b-\mu)^n \frac{\mu-a}{b-a} \right] w_{k.C.ab}$$

there follows

$$\sum_{k.C.ab=1}^{K.C.ab} L_{k.C.AB}(a,\mu,b,n,w_{k.C.ab}) =$$
$$= \sum_{k.C.ab=1}^{K.C.ab} \left[(\mu-a)^n \frac{b-\mu}{b-a} + (b-\mu)^n \frac{\mu-a}{B-a} \right] w_{k.C.ab} = \ .$$
$$= \left[(\mu-a)^n \frac{b-\mu}{b-a} + (b-\mu)^n \frac{\mu-a}{b-a} \right] \sum_{k.C.ab=1}^{K.C.ab} w_{k.C.ab}$$

By definition,

$$\sum_{k.C.ab=1}^{K.C.ab} w_{k.C.ab} \equiv W_{C.ab} \ .$$

Due to the preceding subsubsection, $W_C = W_{C.\mu} + W_{C.ab} = 1$. Therefore, we have

$$\sum_{k.C.ab=1}^{K.C.ab} w_{k.C.ab} \le 1 \ .$$

and

$$\sum_{k.C.ab=1}^{K.C.ab} L_{k.C.ab}(a,\mu,b,n,w_{k.C.ab}) \le (\mu-a)^n \frac{b-\mu}{b-a} + (b-\mu)^n \frac{\mu-a}{b-a}$$

and

$$\begin{aligned} &| E(X-\mu)^n | \le \sum_{k.C.ab=1}^{K.C.ab} L_{k.C.ab}(a,\mu,b,n,w_{k.C.ab}) \le \\ &\le (\mu-a)^n \frac{b-\mu}{b-a} + (b-\mu)^n \frac{\mu-a}{b-a} \end{aligned} \ .$$

So, we have considered a discrete random variable X with finite support. X takes on values in a finite interval $[a, b]$. We have proven that the maximal possible modulus of a central moment of this variable is attained for the probability mass function which is concentrated at the borders of the interval. We have also obtain an estimate of this maximal possible modulus of a central moment of X

$$| E(X-\mu)^n | \le (\mu-a)^n \frac{b-\mu}{b-a} + (b-\mu)^n \frac{\mu-a}{b-a} \qquad (1)$$

4. Theorem

4.1. Preliminary considerations

Remark 4.1. Simplification. Let us simplify the inequality for a central moment (1)

$$|E(X-\mu)^n| \le (\mu-a)^n \frac{b-\mu}{b-a} + (b-\mu)^n \frac{\mu-a}{b-a} =$$
$$= [(\mu-a)^{n-1} + (b-\mu)^{n-1}] \frac{(\mu-a)(b-\mu)}{b-a} = \qquad .$$
$$= \left[\left(\frac{\mu-a}{b-a}\right)^{n-1} + \left(\frac{b-\mu}{b-a}\right)^{n-1}\right](b-a)^{n-1} \frac{(\mu-a)(b-\mu)}{b-a}$$

Keeping in mind $a \le \mu \le b$ we have

$$0 \le \frac{\mu-a}{b-a} \le 1 \quad \text{and} \quad 0 \le \frac{\mu-a}{b-a} \le 1 \ .$$

For $n \ge 2$ we have

$$\left(\frac{\mu-a}{b-a}\right)^{n-1} + \left(\frac{b-\mu}{b-a}\right)^{n-1} =$$
$$= \frac{\mu-a}{b-a}\left(\frac{\mu-a}{b-a}\right)^{n-2} + \frac{b-\mu}{b-a}\left(\frac{b-\mu}{b-a}\right)^{n-2} \le .$$
$$\le \frac{\mu-a}{b-a} + \frac{b-\mu}{b-a} = \frac{b-a}{b-a} \equiv 1$$

So,

$$\left[\left(\frac{\mu-a}{b-a}\right)^{n-1} + \left(\frac{b-\mu}{b-a}\right)^{n-1}\right](b-a)^{n-1} \frac{(\mu-a)(b-\mu)}{b-a} \le$$
$$\le (b-a)^{n-1} \frac{(\mu-a)(b-\mu)}{b-a} = \qquad .$$
$$= (b-a)^{n-2}(\mu-a)(b-\mu)$$

So,

$$|E(X-\mu)^n| \le (b-a)^{n-2}(\mu-a)(b-\mu) \ . \qquad (2)$$

Let us define two terms for the purposes of this article:

Definition 4.2. Bound (restriction) on the expectation.

A "**non-zero bound** (**restriction) on the expectation**

$$restriction_{Expectation} \equiv r_{Expect} \equiv r > 0$$

signifies the impossibility for the expectation to be located closer to a border of the interval than some non-zero distance $r_{Expect} > 0$.

In other words, a non-zero bound designates the existence of a non-zero distance from a border of the interval. Within this distance, it is impossible for the expectation to be located.

This bound may be denoted also as a "forbidden zone" for the expectation near a border of the interval.

The "bound" for one border and the "bound" for another border constitute the "bounds" for the borders.

The value of a non-zero bound (or the width of a non-zero "forbidden zone") signifies the minimal possible distance between the expectation and a border of the interval. For brevity, the term "the value of a bound" may be shortened to "the bound."

Definition 4.3. A non-zero bound on a central moment.

At the beginning, let us define a "non-zero bound on the dispersion $\sigma^2_{Min.2} \equiv \sigma^2_{Min}$" to be the minimal value of the dispersion $\sigma^2 \equiv E(X-\mu)^2$ satisfying

$$E(X-\mu)^2 \geq \sigma^2_{Min.2} \equiv \sigma^2_{Min} > 0 \ .$$

Let us define analogously a general "**non-zero bound on a central moment** $|\boldsymbol{\sigma^n_{Min.n}}|$" to be the minimal absolute value of a central moment $E(X-\mu)^n$ satisfying

$$|E(X-\mu)^n| \geq |\sigma^n_{Min.n}| > 0 \ .$$

4.2. Theorem and notes

4.2.1. Theorem

Theorem. Existence theorem. Suppose, a discrete random variable X with finite support takes on values in an interval $[a, b] : 0<(b\text{-}a)<\infty$. If there is a non-zero lower bound $|\sigma^n{}_{Min.n}| > 0$ on the modulus of, at least one, its central moment $|E(X\text{-}\mu)^n| \geq |\sigma^n{}_{Min.n}| : 2\leq n<\infty$, then the non-zero bounds (restrictions) $restriction_{Expectation} \equiv r_{Expect} > 0$ on the expectation exist near the borders of the interval $[a, b]$, such that

$$a < (a + r_{Expect}) \leq \mu \equiv E(X) \leq (b - r_{Expect}) < b \ . \tag{3}$$

Proof. From a composition of the conditions of the theorem and (2) we have

$$0 < |\sigma^n{}_{Min.n}| \leq |E(X-\mu)^n| \leq (b-a)^{n-2}(\mu - a)(b-\mu) \ .$$

This composition can be simplified by denoting

$$\alpha \equiv \mu - a$$

or

$$\beta \equiv b - \mu$$

One may rewrite the composition as

$$0 < |\sigma^n{}_{Min.n}| \leq (b-a)^{n-2}(\mu - a)(b-\mu) =$$
$$(b-a)^{n-2}\alpha(b - a - (\mu - a)) =$$
$$(b-a)^{n-2}((b-a) - \alpha)\alpha$$

or

$$0 < |\sigma^n{}_{Min.n}| \leq |E(X-\mu)^n| \leq (b-a)^{n-2}(\mu - a)(b-\mu) =$$
$$(b-a)^{n-2}(\mu - a)\beta = (b-a)^{n-2}(\mu - b + b - a)\beta =$$
$$(b-a)^{n-2}(-\beta + (b-a))\beta$$

and

$$0 < \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-2}} \leq (b-a)\alpha - \alpha^2$$

or as the exactly analogous and equivalent expression

$$0 < \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-2}} \leq (b-a)\beta - \beta^2 \ .$$

Let us choose further, e.g., β from the two above choices (to not confuse α with a). So, we have the inequality

$$\beta^2 - (b-a)\beta + \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-2}} \leq 0 \ . \tag{4}$$

For the equation

$$\beta^2 - (b-a)\beta + \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-2}} = 0 \ , \tag{5}$$

its roots are

$$\beta_{1,2} = \frac{b-a}{2} \pm \sqrt{\left(\frac{b-a}{2}\right)^2 - \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-2}}} \ . \tag{6}$$

Let us analyze these roots.

Let us consider the function

$$\Phi \equiv \beta^2 - (b-a)\beta + \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-2}} \ .$$

Its derivatives are

$$\frac{\partial\Phi}{\partial\beta} = 2\beta - (b-a) \quad \text{and} \quad \frac{\partial^2\Phi}{\partial\beta^2} = 2 > 0 \ .$$

The first derivative is equal to zero and the function has its minimum at

$$\beta_0 = \frac{b-a}{2} \ .$$

The point β_0 is located between the points of the roots

$$\beta_2 = \frac{b-a}{2} - \sqrt{\left(\frac{b-a}{2}\right)^2 - \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-2}}}$$

and

$$\beta_1 = \frac{b-a}{2} + \sqrt{\left(\frac{b-a}{2}\right)^2 - \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-2}}}$$

(where $\beta_2 \leq \beta_1$) of the equation (5). The function Φ is equal to zero at the roots. Therefore, the values of the function are less than zero when β is located between the roots β_2 and β_1.

Therefore, the inequality (4) is true at

$$\beta_2 \le \beta \le \beta_1$$

and, remembering $\beta \equiv b - \mu$ or $\mu \equiv b - \beta$, the inequality (4) is true at

$$b - \beta_1 \le \mu \le b - \beta_2 \ .$$

Expression (6) for the roots of equation (5) is symmetric with respect to $(b-a)/2$. In particular, $\beta_1 + \beta_2 = (b-a)$. Therefore, $\beta_1 = (b-a) - \beta_2$ and $b - \beta_1 = b - (b-a) + \beta_2 = a + \beta_2$. Therefore, one can write

$$a + \beta_2 \le \mu \le b - \beta_2 \ .$$

So, one may determine the non-zero bound (restriction) $restriction_{Expectation} \equiv r_{Expect} > 0$ on the expectation as

$$r \equiv r_{Expect} \equiv \beta_2 = \frac{b-a}{2} - \sqrt{\left(\frac{b-a}{2}\right)^2 - \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-2}}}$$

or

$$r_{Expect} = \frac{b-a}{2}\left(1 - \sqrt{1 - 4\left(\frac{|\sigma_{Min.n}|}{b-a}\right)^n}\right)$$

or

$$r \equiv r_{Expect} = \frac{b-a}{2} - \sqrt{\left(\frac{b-a}{2}\right)^2 - \sigma^2{}_{Min.n}\left(\frac{|\sigma_{Min.n}|}{b-a}\right)^{n-2}} \ . \qquad (7)$$

Under the conditions of (1), $0<(b-a)<\infty$, and $|E(X-\mu)^n| \ge |\sigma^n{}_{Min.n}| > 0$ and $2 \le n < \infty$, the bounds r_{Expect} on the expectation of the random variable are non-zero.

Therefore, under these conditions, we have proven the theorem in a form of the bounding inequality (3)

$$a < (a + r_{Expect}) \le E(X) \le (b - r_{Expect}) < b \ .$$

4.2.2. Lower bounds on central moments

If there are non-zero lower bounds on the moduli of more than one central moment, then non-zero bounds on the expectation exist for every non-zero lower bound on the modulus of the central moment. Evidently, the maximal of these bounds on the expectation can be used as the tightest ones.

4.2.3. Countable and continuous random variables

The theorem uses only three conditions of:

1) the estimate (1) for the maximal possible modulus of a central moment;
2) finite interval;
3) finite power of a central moment.

Therefore, the theorem will be true for the case of a non-zero lower bound on the modulus of, at least one, finite power central moment $|E(X-\mu)^n|$: $2\leq n<\infty$, of a countable or continuous random variable, which takes on values in a finite interval, as soon as the estimate (1) for the maximal possible modulus of a central moment will be proven for this variable.

4.2.4. Dispersion ($n = 2$)

For the most important case of $n = 2$ and the dispersion $|\sigma^n{}_{Min.n}| = \sigma^2{}_{Min.2} = \sigma^2{}_{Min}$, denoting the half of the length of the interval $[a, b]$ as

$$h \equiv h_{Half} \equiv \frac{b-a}{2} \ ,$$

one can laconically rewrite the inequality (4) as

$$\beta^2 - 2h\beta + \sigma^2{}_{Min} < 0$$

and the roots of the equation $\beta^2 - 2h\beta + \sigma^2{}_{Min} = 0$ as

$$\beta_{1,2} = h \pm \sqrt{h^2 - \sigma^2{}_{Min}} \qquad ,$$

or, denoting the bounds on the expectation $r \equiv r_{Expect} \equiv \beta_2$,

$$r = h - \sqrt{h^2 - \sigma^2{}_{Min}} \ . \qquad (8)$$

The maximal possible dispersion is not more than $((b-a)/2)^2$. So, denoting the maximal possible standard deviation as

$$\sigma_{Max} = \frac{b-a}{2} \ ,$$

we have

$$r_{Expect} = \sigma_{Max} - \sqrt{\sigma^2{}_{Max} - \sigma^2{}_{Min}}$$

or, e.g.,

$$r_{Expect} = \sigma_{Max}\left(1 - \sqrt{1 - \frac{\sigma^2{}_{Min}}{\sigma^2{}_{Max}}}\right) \ .$$

4.2.5. Infinitesimal case

For the case of $\sigma_{Min.n} \rightarrow 0$ one can easily obtain for (3) from either (7) or (2)

$$a < \left(a + \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-1}} \right) \le E(X) \le \left(b - \frac{|\sigma^n{}_{Min.n}|}{(b-a)^{n-1}} \right) < b$$

or (denoting for compactness $\mu \equiv E(X)$)

$$a < a + |\sigma^n{}_{Min.n}| \left(\frac{|\sigma^n{}_{Min.n}|}{b-a} \right)^{n-1} \le \mu \le b - |\sigma^n{}_{Min.n}| \left(\frac{|\sigma^n{}_{Min.n}|}{b-a} \right)^{n-1} < b \qquad (9)$$

For $n = 2$ and $\sigma_{Min.n} = \sigma_{Min.2} = \sigma_{Min}$ one can rewrite (10) as

$$a < \left(a + \frac{\sigma^2{}_{Min}}{b-a} \right) \le E(X) \le \left(b - \frac{\sigma^2{}_{Min}}{b-a} \right) < b \ . \qquad (10)$$

or

$$a < \left(a + \sigma_{Min} \frac{\sigma_{Min}}{b-a} \right) \le E(X) \le \left(b - \sigma_{Min} \frac{\sigma_{Min}}{b-a} \right) < b \ . \qquad (11)$$

5. Opportunities of the theorem for decision, utility and prospect theories

The dispersion is a common measure of a scattering. The scattering can be caused by noise and/or uncertainty, measurement errors, etc.

The dispersion of a random variable can model the consequence of real scattering. More rigorously, the non-zero dispersion signifies that the minimal dispersion of the random variable is bounded from below by a non-zero value. In other words, this signifies “a non-zero bound on the dispersion.”

So, the theorem can be used in researches of the influence of the scatter of experimental data on their expectations near the borders of intervals.

Noise and uncertainty are widespread phenomena in economics, in particular in decision, utility and prospect theories (see, e.g., Schoemaker and Hershey, 1992, Butler and Loomes, 2007). The essential feature of problems of these theories is their intense manifestation near the borders of the scale of probability (see, e.g., Tversky and Wakker, 1995).

Sketches of versions of the above existence theorem have at least partially explained the problems, including underweighting of high and the overweighting of low probabilities, risk aversion, the "four-fold pattern" paradox, etc. (see, e.g., Harin 2012). So, the theorem can be used also in decision, utility and prospect theories, especially in researches of Prelec’s weighting function.

Conclusions

Suppose a discrete random variable $X = \{x_k\} : k=1, 2, \ldots, K : 2 \le K < \infty$, takes on values in an interval $[a, b] : 0 < (b\text{-}a) < \infty$. Suppose there is a non-zero lower bound on the modulus of (at least one) its central moment $|E(X\text{-}E(X))^n| : 2 \le n < \infty$ (this bound is denoted as $|\sigma^n{}_{Min.n}|$, so, $|E(X\text{-}E(X))^n| \ge |\sigma^n{}_{Min.n}| > 0$).

Under these conditions, the existence theorem is proven for non-zero bounds (restrictions) $restriction_{Expectation} \equiv r_{Expect} \equiv r > 0$ on the expectation $E(X)$ near the borders of the interval. The theorem is proven in the form of bounding inequality (3)

$$a < (a + r_{Expect}) \le E(X) \le (b - r_{Expect}) < b \ .$$

In other words, under the above conditions, the non-zero "forbidden zones" (those widths are equal to the non-zero bounds r_{Expect}) for the expectation are proven to exist near the borders a and b of the interval $[a, b]$.

In this inequality (3) formula (7) for the bounds r_{Expect} on the expectation is

$$r_{Expect} = \frac{b-a}{2} - \sqrt{\left(\frac{b-a}{2}\right)^2 - \sigma^2{}_{Min.n}\left(\frac{|\sigma_{Min.n}|}{b-a}\right)^{n-2}} \ .$$

The above bounding inequality (3) and formula (7) for the bounds on the expectation are the two main results of the present article.

The general formula (7) for the bounds on the expectation can be rewritten for the most important case of $n=2$ in the particular laconic form of (8)

$$r = h - \sqrt{h^2 - \sigma^2{}_{Min}}$$

for the minimum $|\sigma^n{}_{Min.n}| = \sigma^2{}_{Min.2} \equiv \sigma^2{}_{Min} > 0$ of the dispersion σ^2, denoting the half of the interval $h_{Half} \equiv h \equiv (b\text{-}a)/2$ and $r \equiv r_{Expect}$.

The general inequality (3) and particular inequality (10) can be rewritten for $n=2$ and $\sigma_{Min} \to 0$ in the particular forms of (11) or (12)

$$a < \left(a + \sigma_{Min}\frac{\sigma_{Min}}{b-a}\right) \le E(X) \le \left(b - \sigma_{Min}\frac{\sigma_{Min}}{b-a}\right) < b \ .$$

The theorem can be used in utility and prospect theories, especially in researches of Prelec's weighting function.

References

Aczél, J., and R. D. Luce. 2007. “A behavioral condition for Prelec’s weighting function on the positive line without assuming $W(1)=1$,” *Journal of Mathematical Psychology* **51** (2007), pp. 126–129.

Dokov, S. P., D.P. Morton. 2005. Second-Order Lower Bounds on the Expectation of a Convex Function. *Mathematics of Operations Research* **30**(3) 662–677

Harin, A. 2012. “Data dispersion in economics (II) – Inevitability and Consequences of Restrictions”, *Review of Economics & Finance* **2** (2012), no. 4: 24–36.

Harin, A. 2014. “The random--lottery incentive system. Can $p \sim 1$ experiments deductions be correct?” 16th conference on the Foundations of Utility and Risk, **16** (2014), 1-30.

Harin, A. 2015. “An existence theorem for bounds (restrictions) on the expectation of a random variable. Its opportunities for decision, utility and prospect theories,” *MPRA Paper, Item ID:* 66692, (2015).

Kahneman, D., and R. Thaler. 2006. “Anomalies: Utility Maximization and Experienced Utility,” *Journal of Economic Perspectives* **20** (2006), no. 1, 221–234.

Pinelis, I. 2011. “Exact lower bounds on the exponential moments of truncated random variables.” *Journal of Applied Probability* **48**(2) 547-560.

Prékopa, A. 1990. “The discrete moment problem and linear programming,” *Discrete Applied Mathematics* **27**(3) 235–254.

Prékopa, A. 1992. “Inequalities on Expectations Based on the Knowledge of Multivariate Moments,” *Lecture Notes-Monograph Series* **22** *Stochastic Inequalities* 309-331.

Schoemaker, P., and J. Hershey. 1992. “Utility measurement: Signal, noise, and bias,” *Organizational Behavior and Human Decision Processes* **52** (1992) no. 3, 397–424.

Steingrimsson, R., and R. D. Luce. 2007. “Empirical evaluation of a model of global psychophysical judgments: IV. Forms for the weighting function,” *Journal of Mathematical Psychology* **51** (2007), pp. 29–44.

Tversky, A., and P. Wakker. 1995. “Risk attitudes and decision weights,” *Econometrica* **63** (1995), 1255–1280.