Tendências em Matemática Aplicada e Computacional, **1**, No. 2 (2000), 333-344.
Sociedade Brasileira de Matemática Aplicada e Computacional.

The Interval Probability: Applications to Discrete Random Variables

M.A. CAMPOS, Centro de Informática, UFPE, Cx.P. 7851, 50732-970 Recife, PE, Brazil.

Abstract. The interval probability is a proposal for solving numerical problems in computing probabilities. This paper presents advances on the calculus of interval probabilities for discrete random variables. Several examples illustrate this new approach.

1. Introduction

Real numbers, $\mathbb{R}$, are usually implemented in computers as floating-point numbers [8] so, most of them can be represented with fixed accuracy. This is a serious problem in all fields where high accuracy calculations are required [7]. Several different representations of real numbers have been proposed, but the most widely used is the floating-point representation [7, 8]. A floating-point system $S = S(b, l, e_{min}, e_{max})$ has a base b, a precision l, and the largest and smallest allowable exponents, e_{max} and e_{min}. S has $2(b-1)b^{l-1}(e_{max} - e_{min} + 1) + 1$ elements, and these are the unique numbers processed by the computer. Unfortunately, the algebraic characteristics of the floating-point system are extremely poor if compared with the ones of the real numbers system. While $\langle \mathbb{R}, +, \cdot \rangle$ is a field in general the addition operation of floating-point numbers is not associative.

Example 1 This example is presented in [1]. The representation of 1/3 with four significant digits is considered. The rounding for the nearest is 0.3333. This value can be substituted by the interval [0.3333,0.3334], using the directed roundings [11]. This interval seems more inaccurate than the real value, but it is more reliable because it shows the present degree of uncertainty. It also shows that 0.3333 is an underestimation of the actual value. The interval representation provides information about the computed value which a single number cannot do.

The examples below show how numerical errors affect the computation of probabilities. Respect to those involving discrete random variables it is well known [14] that (i) $p_k = P(X = k) \geq 0$, $k \in R_X \subseteq \mathbb{R}$, where R_X is the range of the random variable X, and (ii) $\sum_{k \in R_X} p_k = 1$.

Example 2 Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$, be the sample space where the probabilities of the elementary events are $P(\omega_1) = 1/\pi$, $P(\omega_2) = 1/\sqrt{7}$, $P(\omega_3) = 1 - (1/\pi + 1/\sqrt{7})$. If $\pi \approx 3.14$ and $\sqrt{7} \approx 2.64$, then $P(\omega_1) \approx 0.318$, $P(\omega_2) \approx 0.379$, $P(\omega_3) \approx 1 -$

$(0.318 + 0.379) \approx 0.303$. If $\pi \approx 3.1416$ and $\sqrt{7} \approx 2.6458$, then $P(\omega_1) \approx 0.31831$, $P(\omega_2) \approx 0.37796$, $P(\omega_3) \approx 1-(0.31831+0.37796) \approx 0.30373$. The representation of the approximated values for π and $\sqrt{7}$ presents no problem in an usual floating-point system. So the errors in the computations are caused by roundoff and truncation. It is common ground that increasing the precision necessarily does not imply increasing the accuracy. In this example the increasing of the precision will modify the value of $P(\omega_2)$ and $P(\omega_3)$ in the third decimal digit.

Example 3 Suppose that X has a Bernoulli distribution. If $p = 1/31$ and $q = 30/31$ then $\sum_{k=0}^{1} p_k = 1$. If $p = 1/31 \approx 0.032$ and $q = 30/31 \approx 0.968$ then $\sum_{k=0}^{1} p_k = 1.000$ but if $p \approx 0.0322$ and $q \approx 0.9677$ then $\sum_{k \in R_X} p_k \approx 0.9999 < 1$. The important point to be quoted is that $1/31$ is a rational number, therefore either it has a finite representation or it is a periodic fraction, with a period so long that exceeds the precision of the floating-point system.

Example 4 Let X be a binomial random variable with parameters $n = 3$ and $p = 1/3$. Thus $q = 2/3$ and $\sum_{k=0}^{3} p_k = 1$. If $p \approx 0.3333$ and $q \approx 0.6667$ then $p_0 \approx 0.2963$, $p_1 \approx 0.4444$, $p_2 \approx 0.2222$, $p_3 \approx 0.0370$, and $\sum_{k=0}^{3} p_k \approx 0.9999 < 1$.

Example 5 Suppose X follows a Poisson distribution with parameter $\lambda = 2$, and $e = 2.7$ is an approximated value for e. Then, $p_0 \approx 0.137$, $p_1 \approx 0.274$, $p_2 \approx 0.274$, $p_3 \approx 0.183$, $p_4 \approx 0.091$, $p_5 \approx 0.037$, $p_6 \approx 0.012$, and $\sum_{k=0}^{6} P(X = k) \approx 1.007 > 1$.

To solve problems in several fields, including the representation of real numbers in computers, due to its ability to manipulate imprecise data and to control truncation and roundoff errors, during the last years interval mathematics [2, 12, 13] with high accuracy arithmetic [11] has been widely applied.

The interval probability [3, 4] is a proposal for the solution of numerical problems in computing probabilities. Interval mathematics with high accuracy arithmetic controls numerical errors, thus the unique errors that remain when interval probabilities are computed are those inherent to random processes.

This paper presents advances to the calculus of interval probabilities for discrete random variables. It is outlined as follows. Section 2 is a brief remark on the interval probability (more details can be found in [3, 4]). The next section presents discrete random variables defined through intervals. Several examples illustrate the advantage of the proposed approach, and in all of them it is used the BIAS/PROFIL library [9, 10]. This library supports the general definition of computer arithmetic by semimorphism proposed in [11]. The conclusions are in the last section.

2. The Interval Probability

Let $\mathbb{R}$ be the set of real numbers. The set of *intervals*–$\mathbb{IR}$ is $\mathbb{IR} = \{[a_1, a_2] \mid a_1 \leq a_2,\ a_1, a_2 \in \mathbb{R}\}$, where $[a_1, a_2] = \{x \in \mathbb{R} \mid a_1 \leq x \leq a_2\}$. An element in $\mathbb{IR}$ is called an *interval*. The real numbers are denoted by lowercase letters and the elements of $\mathbb{IR}$ are named by uppercase letters and written as $A = [a_1, a_2]$, $X = [x_1, x_2]$, etc. Another notation for an interval is $[\underline{x}, \overline{x}]$. Details about intervals related to equality, order, algebraic structure, topology can be found in [2, 11, 12, 13, 16].

Scientists and engineers solve computational problems involving probabilities by performing approximate calculations with limited precision, i.e., the precision of the floating-point system. So given the probability of the event A the real number $P(A) = p$, either $p \in S$ or $p \notin S$. Therefore in both situations p can be represented by an interval. From the point of view of implementation, p can be the smallest machine representable interval, $[\nabla p, \triangle p]$, that contains it, where ∇ and $\triangle$ are the directed roundings [11]. This means that $[\nabla p, \triangle p]$ is an interval whose endpoints are floating-point numbers such that $\nabla p \leq p \leq \triangle p$.

Campos [3, 4] proposed the interval probability as a probability measure to solve the numerical errors involved in the calculus of real probabilities (all details about the definition of the interval probability can be found in [3, 4]).

Definition 1 *An interval space of probability is a quadruple $\langle \Omega, \mathcal{A}, P, P_v \rangle$ where Ω is the sample space, $\mathcal{A}$ is a σ-field of subsets of Ω, P is the probability function defined on $\mathcal{A}$ and the function $P_v : \mathcal{A} \to I\!I\!R$ has the following properties:*

(*i*) $\forall A \in \mathcal{A}$ *if $P(A)$ exists, then exists $P_v(A) \in I\!I\!R$ such that $P(A) \in P_v(A)$,*

(*ii*) $1 \in P_v(\Omega)$,

(*iii*) $0 \in P_v(\emptyset)$,

(*iv*) *If $A_1, A_2, \ldots$ is a sequence of elements belonging to $\mathcal{A}$ where $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $P_v(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P_v(A_n)$.*

The function P_v is called an interval probability *or an* interval extension for P *or a* validation for P.

The question now is: what additional properties similar to those of the real probability the proposed interval probability has? The following propositions answer this question (the complete proofs are in [4]).

Proposition 1 *If $\{A_n\}_{n \in I\!N}$ is such that $A_1 \subset A_2 \subset \ldots$, $A = \cup_{n=1}^{\infty} A_n = \lim_{n\to\infty} A_n$, then*

$$P_v(A) = \lim_{n\to\infty} P_v(A_n).$$

Proposition 2 *Given $A \in \mathcal{A}$, $P_v(A) = [a_1, a_2]$, and ε a positive real number. A validation for the complementary event, $P_v(A^c)$, is*

$$P_v(A^c) = 1_\varepsilon - P_v(A).$$

So if $p_v = P_v(A) = [p_1, p_2]$, then $q_v = P_v(A^c) = [q_1, q_2] = [1 - p_2 - \varepsilon, 1 - p_1 + \varepsilon]$. The real number ε can be related either with the precision or the error in the computations.

Example 6 If $p_v = [0.625 \cdot 10^{-1}, 0.625 \cdot 10^{-1}]$ then $q_v = [1 - 0.625 \cdot 10^{-1} - \varepsilon, 1 - 0.625 \cdot 10^{-1} + \varepsilon]$.

Proposition 3 *If* $\{A_n\}_{n\in I\!N}$ *is such that* $A_1 \supset A_2 \supset \cdots$, $A = \cap_{n=1}^{\infty} A_n = \lim_{n\to\infty} A_n$ *then*

$$P_v(A) = \lim_{n\to\infty} P_v(A_n) + [-\delta, +\delta].$$

Proposition 4 *Let* $\langle I\!I\!R, \leq\rangle$ *the set of the intervals with the partial order* $\leq$ *proposed by [11]. If* $A \subseteq B$ *then*

$$P_v(A) \leq P_v(B).$$

Proposition 5 *Let* $A, B \in \mathcal{A}$ *be where* $A \subseteq B$, $P_v(A) = [a_1, a_2]$ *and* $P_v(B) = [b_1, b_2]$. *If* $a_1 = a_2 = a$, *then*

$$P_v(B - A) = P_v(B) - P_v(A).$$

Proposition 6 *If* $P_v(A) = [a_1, a_2]$, $P_v(B) = [b_1, b_2]$ *and* $P_v(A \cap B) = P(A \cap B) = [a, a]$, *then*

$$P_v(A \cup B) = P_v(A) + P_v(B) - P_v(A \cap B).$$

Proposition 7 $P_v(\cup_{k=1}^{n} A_k) \leq \sum_{k=1}^{n} P_v(A_k)$.

Proposition 8 *Let* $A_n \in \mathcal{A}$, $\forall n \in I\!N$ *such that* $A_1 \supset A_2 \supset \ldots$ *and* $\lim_{n\to\infty} A_n = \emptyset$. *Then*

$$\lim_{n\to\infty} P_v(A_n) = 0_\varepsilon.$$

3. Interval Probabilities for Discrete Random Variables

In this section interval probabilities are computed for the discrete random variables Bernoulli, Binomial, Poisson, Truncated Poisson, Hypergeometric, Geometric, and Pascal. Examples use the BIAS/PROFIL library [9, 10].

Definition 2 *Given* X *a random variable taking values in a countable set* $R_X = \{x_1, x_2, \cdots\}$, *and* $P(X = k) = p_k, k \in R_X$ *its probability function.* $P_v(X = k)$ *satisfies the following properties:*

(*i*) $P_v(X = k) \in I\!I\!R$,

(*ii*) $1 \in \sum_{k\in R_X} P_v(X = k)$.

$P_v(X = k)$ *is an interval probability function, or an* interval extension for the probability function*, or yet a validation for the real probability function.*

For each one of the random variables defined below the (i) and (ii) conditions must be proved. Condition (i) is straightforward.

3.1. Bernoulli

Definition 3 *Let X be a random variable with Bernoulli distribution, such that $P(X = 1) = p$ and $P(X = 0) = 1 - p = q$. Then the Bernoulli interval probability function is $P_v(X = 1) = p_v = [p_1, p_2]$, and $P_v(X = 0) = q_v = [q_1, q_2]$.*

Lemma 1 *Let p_v and q_v be, then $1 \in [1 - (p_2 - p_1) - \varepsilon, 1 + (p_2 - p_1) + \varepsilon]$.*

Proof. Given $p_v = P_v(X = 1) = [p_1, p_2]$ and $q_v = P_v(X = 0) = [1-p_2-\varepsilon, 1-p_1+\varepsilon]$, then $\sum_{k=0}^{1} P_v(X = k) = [1 - (p_2 - p_1) - \varepsilon, 1 + (p_2 - p_1) + \varepsilon]$. It can be seen that $1 - (p_2 - p_1) - \varepsilon \leq 1$ and $1 + (p_2 - p_1) + \varepsilon \geq 1$. ∎

Example 7 Let $p_v = [0.0322, 0.0323]$ be. Then

$\varepsilon = 10^{-5} \Rightarrow q_v = [0.9676900001, 0.9677900001]$ and
$p_v + q_v = [0.9998900000, 1.0000900001]$.
$\varepsilon = 10^{-6} \Rightarrow q_v = [0.9676990001, 0.9677990001]$ and
$p_v + q_v = [0.9998990000, 1.0000990001]$.

3.2. Binomial

In the binomial distribution, $B(n, p)$, the parameter n is a non-negative integer number. An interval extension for the probability function takes interval values only for p, q, and the arithmetic operations.

Definition 4 *Suppose that $X \sim B(n, p)$. The binomial interval probability function is*

$$P_v(X = k) = \binom{n}{k} p_v^k q_v^{n-k}, k = 0, \cdots, n.$$

Lemma 2 *If $X \sim B(n, p)$ then $1 \in \sum_{k=0}^{n} \binom{n}{k} p_v^k q_v^{n-k}$.*

Proof. Let $p_v = [p_1, p_2]$ and $q_v = [1 - p_2 - \varepsilon, 1 - p_1 + \varepsilon]$, then

$$\begin{aligned} \sum_{k=0}^{n} P_v(X = k) &= \sum_{k=0}^{n} \binom{n}{k} p_v k q_v^{n-k} \\ &= (p_v + q_v)^n \\ &= ([p_1, p_2] + [1 - p_2 - \varepsilon, 1 - p_1 + \varepsilon])^n \\ &= [1 - (p_2 - p_1) - \varepsilon, 1 + (p_2 - p_1) + \varepsilon]^n. \end{aligned}$$

This proof involves details which can be seen in [4]. ∎

Example 8 If X follows a binomial distribution with parameters 3 and 1/3 then $p_v = [0.3330000001, 0.3340000001]$, $q_v = [0.6660000000, 0.6670000000]$, and

$$\begin{aligned} &P_v(X = 0) = [0.2954082960, 0.2967409630], \\ &P_v(X = 1) = [0.4431124440, 0.4457787780], \\ &P_v(X = 2) = [0.2215562220, 0.2232235560], \\ &P_v(X = 3) = [0.0369260371, 0.0372597041], \\ &\textstyle\sum_{k=0}^{3} P_v(X = k) = [0.9970029990, 1.0030030010]. \end{aligned}$$

3.3. Poisson

The method for getting interval probabilities for the Poisson distribution uses specific interval extensions of real functions [4, 5, 12].

Definition 5 *Given X a Poisson random variable, the Poisson interval probability function is*

$$P_v(X = k) = [\underline{\frac{e^{-\lambda}\lambda^k}{k!}}, \overline{\frac{e^{-\lambda}\lambda^k}{k!}}], k = 0, 1, \ldots$$

Lemma 3 $1 \in \sum_{k=0}^{\infty}[\underline{\frac{e^{-\lambda}\lambda^k}{k!}}, \overline{\frac{e^{-\lambda}\lambda^k}{k!}}]$.

Proof.

$$\begin{aligned}
\sum_{k=0}^{\infty} P_v(X = k) &= \sum_{k=0}^{\infty}[\underline{\frac{e^{-\lambda}\lambda^k}{k!}}, \overline{\frac{e^{-\lambda}\lambda^k}{k!}}] \\
&= [\sum_{k=0}^{\infty}\underline{\frac{e^{-\lambda}\lambda^k}{k!}}, \sum_{k=0}^{\infty}\overline{\frac{e^{-\lambda}\lambda^k}{k!}}] \\
&= [1, 1].
\end{aligned}$$

Details of the interval extensions for the functions used in this proof can be seen in [13]. ∎

Example 9 Given X a Poisson random variable with parameter $\lambda = 10$, the objective is to calculate the probability that X be less or equal to 5. So,

$$\begin{aligned}
P_v(X = 0) &= [0.0000454000, 0.0000454001], \\
P_v(X = 1) &= [0.0004539994, 0.0004539995], \\
P_v(X = 2) &= [0.0022699965, 0.0022699966], \\
P_v(X = 3) &= [0.0075666550, 0.0075666551], \\
P_v(X = 4) &= [0.0189166375, 0.0189166376], \\
P_v(X = 5) &= [0.0378332749, 0.0378332750].
\end{aligned}$$

Thus, $P_v(X \leq 5) = [0.0670859628, 0.0670859634]$ and the floating-point result with the usual rounding is $P(X \leq 5) = 0.067$.

3.4. Truncated Poisson

Definition 6 *Given X a Poisson random variable, the Truncated on the right Poisson interval probability function is defined by*

$$P_v(X = j) = \begin{cases} [c_1, c_2][\underline{\frac{e^{-\lambda}\lambda^k}{k!}}, \overline{\frac{e^{-\lambda}\lambda^k}{k!}}], & j \leq k, \\ 0, & j \geq k + 1. \end{cases}$$

Lemma 4 *If X has a truncated Poisson interval probability distribution then $1 \in \sum_{k=0}^{\infty} P_v(X = k)$ if and only if $c_1 \leq 1/\sum_{j=0}^{k}\underline{\frac{e^{-\lambda}\lambda^j}{j!}}$ and $c_2 \geq 1/\sum_{j=0}^{k}\overline{\frac{e^{-\lambda}\lambda^j}{j!}}$.*

Proof.

$$\begin{aligned}
\sum_{k=0}^{\infty} P_v(X=k) &= \sum_{j=0}^{k} P_v(X=j) \\
&= \sum_{j=0}^{k} [c_1, c_2]\left[\underline{\frac{e^{-\lambda}\lambda^j}{j!}}, \overline{\frac{e^{-\lambda}\lambda^j}{j!}}\right] \\
&= \left[\sum_{j=0}^{k} c_1 \underline{\frac{e^{-\lambda}\lambda^j}{j!}}, \sum_{j=0}^{k} c_2 \overline{\frac{e^{-\lambda}\lambda^j}{j!}}\right] \\
&= \left[c_1 \sum_{j=0}^{k} \underline{\frac{e^{-\lambda}\lambda^j}{j!}}, c_2 \sum_{j=0}^{k} \overline{\frac{e^{-\lambda}\lambda^j}{j!}}\right].
\end{aligned}$$

If c_1 and c_2 are real numbers satisfying the above restrictions, the lemma is proved.
∎

Example 10 It is supposed that X is a random variable which follows a truncated Poisson distribution at $k = 4$, with parameter $\lambda = 2$. So,
$[c_1, c_2] = [1.0555794427, 1.0555794428]$, and the interval probabilities are:

$$\begin{aligned}
&P_v(X=0) = [0.1428571429, 0.1428571430], \\
&P_v(X=1) = [0.2857142857, 0.2857142858], \\
&P_v(X=2) = [0.2857142857, 0.2857142858], \\
&P_v(X=3) = [0.1904761905, 0.1904761906], \\
&P_v(X=4) = [0.0952380953, 0.0952380953], \\
&\textstyle\sum_{j=0}^{4} P_v(X=j) = [0.9999999999, 1.0000000003].
\end{aligned}$$

3.5. Hypergeometric

A random variable with non-negative integer parameters N and n has a hypergeometric distribution if

$$P(X=k) = \frac{\binom{n_1}{k}\binom{N-n_1}{n-k}}{\binom{N}{n}}, \; k = 0, 1, \cdots, \min\{n, n_1\}.$$

The example below (Feller [6] on pp. 44) is used to show numerical problems in the computation of probabilities for this random variable.

Example 11 It is supposed that $N = 100$, $n_1 = 2$, and $n = 50$. Then $P(X = 0) = 0.24747\cdots$, $P(X = 1) = 0.5050\cdots$, and $P(X = 2) = 0.24747\cdots$. With four significative digits $\sum_{k=0}^{2} P(X = k) = 0.9998 < 1$.

Different from the former random variables, interval values for this one are needed only at the end of the computed probabilities. This suggests the following definition.

Definition 7 *Let X be a random variable with hypergeometric distribution, the hypergeometric interval probability function is*

$$P_v(X=k)=[\underline{\frac{\binom{n_1}{k}\binom{N-n_1}{n-k}}{\binom{N}{n}}},\overline{\frac{\binom{n_1}{k}\binom{N-n_1}{n-k}}{\binom{N}{n}}}], k=0,1,\cdots,\min\{n,n_1\}.$$

Example 12 Interval results with the data of the Example 11 are:

$$\begin{array}{l} P_v(X=0)=[0.247474747401, 0.247474747501], \\ P_v(X=1)=[0.505050505001, 0.505050505101], \\ P_v(X=2)=[0.247474747401, 0.247474747501], \\ \sum_{k=0}^{2} P_v(X=k)=[0.9999999998, 1.0000000002]. \end{array}$$

To define interval probability distribution for the Geometric and Pascal random variables, it is necessary to propose results to series of intervals.

Definition 8 *Let be the sequence of intervals $\{[x_{n_1}, x_{n_2}]\}$, $n=1,2,\cdots$ and $s_n=\sum_{k=1}^{n}[x_{k_1}, x_{k_2}]$. If exists $\lim_{n\to\infty} s_n$, then*

$$\sum_{n=1}^{\infty}[x_{n_1}, x_{n_2}]=\lim_{n\to\infty} s_n=\lim_{n\to\infty}\sum_{k=1}^{n}[x_{k_1}, x_{k_2}].$$

For real numbers it is well known that

$$\sum_{k=1}^{\infty} ap_k=\lim_{n\to\infty}\sum_{k=1}^{n} ap_k=\lim_{n\to\infty}(a\sum_{k=1}^{n} p_k)=a\lim_{n\to\infty}\sum_{k=1}^{n} p_k=a\sum_{k=1}^{\infty} p_k.$$

The following lemma proves a similar property for intervals, but in the particular situation where the distributive law for intervals is true, i.e, if $X=x$ or $YZ>0$ then $X(Y+Z)=XY+XZ$.

Lemma 5 *Let be the sequence of intervals $\{X_n=[x_{n_1}, x_{n_2}]\}$, $n=1,2,\cdots$, where $X_mX_n>0$ $\forall m,n$. Then*

$$\sum_{n=1}^{\infty}[x,y][x_{n_1}, x_{n_2}]=[x,y]\sum_{n=1}^{\infty}[x_{n_1}, x_{n_2}].$$

Proof.

$$\begin{aligned} \sum_{k=1}^{\infty}[x,y][x_{k_1}, x_{k_2}] &= \sum_{k=1}^{\infty}[xx_{k_1}, yx_{k_2}] \\ &= \lim_{n\to\infty}\sum_{k=1}^{n}[xx_{k_1}, yx_{k_2}] \\ &= \lim_{n\to\infty}\sum_{k=1}^{n}[x,y][x_{k_1}, x_{k_2}] \end{aligned}$$

$$\begin{aligned} &= \lim_{n\to\infty}([x,y]\sum_{k=1}^{n}[x_{k_1},x_{k_2}]) \\ &= [x,y]\lim_{n\to\infty}\sum_{k=1}^{n}[x_{k_1},x_{k_2}] \\ &= [x,y]\sum_{n=1}^{\infty}[x_{n_1},x_{n_2}]. \end{aligned}$$

■

3.6. Geometric

Definition 9 *The* geometric interval probability function *for a random variable with geometric distribution of parameter p is*

$$P_v(X=k)=p_v q_v^{k-1},\ k=1,2,\cdots.$$

Lemma 6 $1\in\sum_{k=1}^{\infty}p_v q_v^{k-1}$.

Proof.

$$\begin{aligned} \sum_{k=1}^{\infty}P_v(X=k) &= \sum_{k=1}^{\infty}[p_1,p_2][q_1,q_2]^{k-1} \\ &= \sum_{k=1}^{\infty}[p_1,p_2][q_1^{k-1},q_2^{k-1}] \\ &= \sum_{k=1}^{\infty}[p_1q_1^{k-1},p_2q_2^{k-1}] \\ &= [\sum_{k=1}^{\infty}p_1q_1^{k-1},\sum_{k=1}^{\infty}p_2q_2^{k-1}] \\ &= [\sum_{k=1}^{\infty}p_1(1-p_2-\varepsilon)^{k-1},\sum_{k=1}^{\infty}p_2(1-p_1+\varepsilon)^{k-1}] \\ &= [p_1\sum_{k=1}^{\infty}(1-p_2-\varepsilon)^{k-1},p_2\sum_{k=1}^{\infty}(1-p_1+\varepsilon)^{k-1}]. \end{aligned}$$

If $-2<\varepsilon-p_1<0$ and $0<\varepsilon+p_2<2$ then $\sum_{k=1}^{\infty}(1-p_2-\varepsilon)^{k-1}=\frac{1}{\varepsilon+p_2}$ and $\sum_{k=1}^{\infty}(1-p_1+\varepsilon)^{k-1}=\frac{1}{p_1-\varepsilon}$. So $\sum_{k=1}^{\infty}P_v(X=k)=[\frac{p_1}{p_2+\varepsilon},\frac{p_2}{p_1-\varepsilon}]$ and the lemma is proved because $\frac{p_1}{p_2+\varepsilon}\leq 1\leq\frac{p_2}{p_1-\varepsilon}$. ■

Example 13 Let be $p_v=[0.3300000001,0.3400000001]$,
so $q_v=[0.6600000000,0.6700000000]$ and the interval probabilities are:

$$
\begin{aligned}
P_v(X=0) &= [0.3300000001, 0.3400000001],\\
P_v(X=1) &= [0.2178000000, 0.2278000000],\\
P_v(X=2) &= [0.1437480000, 0.1526260000],\\
P_v(X=3) &= [0.0948736800, 0.1022594200],\\
P_v(X=4) &= [0.0626166288, 0.0685138114],\\
P_v(X=5) &= [0.0413269751, 0.0459042537],\\
P_v(X\leq 5) &= [0.8903652839, 0.9371034851].
\end{aligned}
$$

3.7. Pascal

The negative distribution with parameters α and p of a given random variable X is $P(X=k)=\binom{\alpha+k-1}{k}p^\alpha q^k$, where k is an integer greater or equal to zero. If α is a positive integer, this distribution is also called the Pascal distribution. The definition below proposes an interval extension for the Pascal distribution random variable.

Definition 10 *If X is a random variable with Pascal distribution then its Pascal interval probability function is*

$$P_v(X=k)=\binom{\alpha+k-1}{k}p_v^\alpha q_v^k,\ \ k=0,1,\cdots.$$

Lemma 7 *If X has an interval Pascal distribution then $1\in\sum_{k=0}^{\infty}\binom{\alpha+k-1}{k}p_v^\alpha q_v^k$.*

Proof. Using the Lemma 6 it remains first to solve $\sum_{k=0}^{\infty}\binom{\alpha+k-1}{k}q_v^k$. In this summation it is supposed that all the real numbers involved are positive. Then it is possible to use Lemma 6 to take p_v^α off.

$$
\begin{aligned}
\sum_{k=0}^{\infty}\binom{\alpha+k-1}{k}q_v^k &= \sum_{k=0}^{\infty}\binom{\alpha+k-1}{k}[q_1,q_2]^k\\
&= \sum_{k=0}^{\infty}\binom{\alpha+k-1}{k}[q_1^k,q_2^k]\\
&= \sum_{k=0}^{\infty}[\binom{\alpha+k-1}{k}q_1^k,\binom{\alpha+k-1}{k}q_2^k]\\
&= [\sum_{k=0}^{\infty}\binom{\alpha+k-1}{k}q_1^k,\sum_{k=0}^{\infty}\binom{\alpha+k-1}{k}q_2^k]\\
&= [\sum_{k=0}^{\infty}\binom{\alpha+k-1}{k}(1-p_2-\varepsilon)^k,\sum_{k=0}^{\infty}\binom{\alpha+k-1}{k}(1-p_1+\varepsilon)^k]\\
&= [(p_2+\varepsilon)^{-\alpha},(p_1-\varepsilon)^{-\alpha}]
\end{aligned}
$$

So, $p_v \sum_{k=0}^{\infty} q_v^k = [(\frac{p_1}{p_2+\varepsilon})^\alpha, (\frac{p_2}{p_1-\varepsilon})^\alpha]$. It is noted that $\frac{p_1}{p_2+\varepsilon} \leq 1, \forall\varepsilon \geq 0 \Rightarrow (\frac{p_1}{p_2+\varepsilon})^\alpha \leq 1$, and $\frac{p_2}{p_1-\varepsilon} \geq 1, \forall\varepsilon \geq 0 \Rightarrow (\frac{p_2}{p_1-\varepsilon})^\alpha \geq 1$. ∎

Example 14 If $p_v = [0.2500000000, 0.2600000001]$, then $q_v = [0.7400000000, 0.7500000000]$ and for $\alpha = 3$ some of the interval probabilities are:

$$
\begin{aligned}
P_v(X = 0) &= [0.0156250000, 0.0175760001],\\
P_v(X = 1) &= [0.0346875000, 0.0395460001],\\
P_v(X = 2) &= [0.0513375000, 0.0593190000],\\
P_v(X \leq 2) &= [0.1016500000, 0.1164410001].
\end{aligned}
$$

Acknowledgements

The author is greatly thankful to CAPES, BRAZIL.

4. Conclusions

Results show that the interval probability can be used for solving numerical problems (which causes many pitfalls in computers [7, 15]) involving random variables. Additionaly, it is possible to control the precision of the realized computations by varying the constant ε.

References

[1] Y. Akyildiz and M.I. Al-Suwaiyel, No Pathologies for Interval Newton's Method, *Interval Computations*, **1** (1993), 60-72.

[2] G. Alefeld and J. Herzberger, "Introduction to Interval Computation", Academic Press, N. Y., 1983.

[3] M.A. Campos, "Uma Extensão Intervalar para a Probabilidade Real", Tese de Doutorado, Departamento de Informática, Universidade Fedral de Pernambuco, 1997.

[4] M.A. Campos, How to Define an Interval Probability, Presented at 1a. Escola Brasileira de Probabilidade (1st. Brazilian School of Probability), IMPA, Rio de Janeiro, 11-16 de agosto de 1997. (submitted to *Markov Processes and Related Fields*)

[5] M.A. Campos, Validating Probabilities for Random Variables with Binomial and Poisson Distributions. Presented at International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics (SCAN 97), September, 12-14, Lyon, France, 1997. (submitted to *Reliable Computing*)

[6] W. Feller, “An Introduction to Probability Theory and Its Applications”, John Wiley & Sons, N. Y., 1968.

[7] G.E. Forsythe, Pitfalls in Computation, or Why a Math Book Isn’t Enough, *Amer. Math Monthly*, (1977), 931-955.

[8] D. Goldberg, What Every Computer Scientist Should Know About Floating-Point Arithmetic, *ACM Computing Surveys*, **23** (1) (1991), 5-48.

[9] O. Knüppel, BIAS - Basic Interval Arithmetic Subroutines, Berich 93.3 des Forschungsschwerpunktes Informations- und Kommunikationstechnik der TU Hamburg-Harburg, TU Hamburg-Harburg, Technische Informatik III, July 1993.

[10] O. Knüppel, PROFIL - Programmers’s Runtime Optimized Fast Interval Library, Berich 93.4 des Forschungsschwerpunktes Informations- und Kommunikationstechnik der TU Hamburg-Harburg, TU Hamburg-Harburg, Technische Informatik III, July 1993.

[11] U.W. Kulisch and W.L. Miranker, “Computer Arithmetic in Theory and Practice”, Academic Press, N. Y., 1981.

[12] R.E. Moore, R.E., “Interval Analysis”, Prentice Hall Inc., Englewood Clifs, N. J., 1966.

[13] R.E. Moore, “Methods and Applications of Interval Analysis”, SIAM, Philadelphia, Pennsylvania, 1979.

[14] P.E. Pfeiffer, “Concepts of Probability Theory”, Second Revised Edition, Dover Publications, Inc, 1978.

[15] I.A. Stegun and M. Abramowitz, Pitfalls in Computations. *J. Soc. Indist. Appl. Math.*, (1956), 207-219.

[16] T. Sunaga, Theory of An Interval Algebra and its Application to Numerical Analysis., *RAAG Memoirs*, **2** (1958), 29-46.