An Evolutionary Algorithm for Integer Programming*

Günter Rudolph

Informatik Centrum Dortmund e.V., Joseph-von-Fraunhofer-Str. 20
D-44227 Dortmund

Abstract. The mutation distribution of evolutionary algorithms usually is oriented at the type of the search space. Typical examples are binomial distributions for binary strings in genetic algorithms or normal distributions for real valued vectors in evolution strategies and evolutionary programming. This paper is devoted to the construction of a mutation distribution for unbounded integer search spaces. The principle of maximum entropy is used to select a specific distribution from numerous potential candidates. The resulting evolutionary algorithm is tested for five nonlinear integer problems.

1 Introduction

Evolutionary algorithms (EAs) represent a class of stochastic optimization algorithms in which principles of organic evolution are regarded as rules in optimization. They are often applied to real parameter optimization problems [2] when specialized techniques are not available or standard methods fail to give satisfactory answers due to multimodality, nondifferentiability or discontinuities of the problem under consideration. Here, we focus on using EAs in integer programming problems of the type

$$\max\{f(x) \,:\, x \in M \subseteq \mathbb{Z}^n\} \tag{1}$$

where $\mathbb{Z}$ denotes the set of integers. Note that the feasible region M is not required to be bounded. Consequently, the encoding of the integer search space with fixed length binary strings as used in standard genetic algorithms (GA) [7, 6] is not feasible. The approach to use an evolution strategy (ES) [13, 14] by embedding the search space $\mathbb{Z}^n$ into $\mathbb{R}^n$ and truncating real values to integers has, however, also its deficiency: As evolution strategies usually operate on real valued spaces they include features to locate optimal points with arbitrary accuracy. In integer spaces these features are not necessary because the smallest distance in ℓ_1-norm between two different points is 1. Therefore, as soon as the step sizes drop below 1 the search will stagnate. Thus, EAs for integer programming should operate directly on integer spaces.
Early approaches in this direction can be found in [4] and [10]. They proposed random search methods on integer spaces in the spirit of a $(1+1)$–ES:

*This work was done in the BMFT project 'EVOALG' under grant 01 IB 403 A.

Choose $x^{(0)} \in M$ and set $t = 0$
repeat
    $y^{(t+1)} = x^{(t)} + z^{(t)}$
    if $y^{(t+1)} \in M$ and $f(y^{(t+1)}) > f(x^{(t)})$ then $x^{(t+1)} = y^{(t+1)}$
    else $x^{(t+1)} = x^{(t)}$
    increment t
until stopping criterion fulfilled

Here, $z^{(t)}$ denotes the random vector used at step t. Gall [4] and Kelahan/Gaddy [10] used a (continuous) bilateral power distribution with density

$$p(x) = \frac{1}{2\,k\,a^{1/k}}\,|x|^{1/k-1} \cdot 1_{[-a,a]}(x) \quad , k \in \mathbb{N},$$

to generate random vector z via $z_i = a \cdot (1 - 2\,u)^k$ with $u \sim U[0,1]$ for each vector component and by truncating these values to integers. The factor a is used to shrink the support of the distribution during the search by a geometrical schedule. Since the support is bounded and its range tends to zero as time increases, the above algorithm is at best locally convergent. Moreover, the truncation of random variables values drawn from continuous mutation distributions to integer values might complicate theoretical investigations. Therefore, the usage of mutation distributions with support $\mathbb{Z}^n$ seems most natural for integer problems (1).
The remainder of the paper is devoted to the construction and test of an appropriate mutation distribution. As there are several candidates for such a mutation distribution we pose some requirements for the desired distribution before using the concept of maximum entropy to guide our final choice. The resulting evolutionary algorithm, which is oriented at a (μ, λ)-ES, is tested for five nonlinear integer problems.

2 Construction of Mutation Distribution

Assume that the current position of the algorithm in the search space is the point $x \in \mathbb{Z}^n$. Since the algorithm does not have prior knowledge of the response surface there is absolute uncertainty about the next step to be taken. In each coordinate direction one might ask the questions: Should we go left or right ? How many steps should be taken in the chosen direction ? The first question has a simple solution: Since we do not have prior knowledge which direction will offer improvement, we should move to the left or right with the same probability. The second question is more difficult to answer: Since we do not know which step size will be successful, we could draw a step size $s \in \mathbb{N}_0$ at random. But which distribution should be used for random variable s ? There are several candidates, for example Poisson, Logseries and geometrical distributions [9]. To select a specific distribution with given mean, say $\theta > 0$, we may use the concept of *maximum entropy*: A distribution with maximum entropy is the one which is spread out as uniformly as possible without contradicting the given information. It agrees with what is known, but expresses maximum uncertainty with respect to all other matters [8]. The usage of this concept to select a distribution usually leads to the task to solve a nonlinear, constrained optimization problem analytically. This will be done in subsection 2.1. The symmetrization of the chosen distribution and its extension to the multivariate case will be considered in subsection 2.2. Finally, we discuss possibilities to control the average step size during the search in subsection 2.3.

2.1 Maximum Entropy

DEFINITION 1
Let p_k with $k \in \mathbb{N}_0$ denote the density of a discrete random variable. Then

$$H(p) = -\sum_{k=0}^{\infty} p_k \log p_k \tag{2}$$

is called the *entropy* of the distribution, provided that series (2) converges†. □

PROPOSITION 1
Let X be a discrete random variable with support $\mathbb{N}_0$ and mean $\mathrm{E}[X] = \theta > 0$. If the distribution of X is requested to possess maximum entropy, then X must have geometrical distribution with density

$$P\{X = k\} = \frac{1}{\theta+1}\left(1 - \frac{1}{\theta+1}\right)^k \quad , \; k \in \mathbb{N}_0 \; . \tag{3}$$

PROOF : The optimal values for p_k can be obtained after partial differentation of the Lagrangian

$$L(p, a, b) = -\sum_{k=0}^{\infty} p_k \log p_k + a\left(\sum_{k=0}^{\infty} p_k - 1\right) + b\left(\sum_{k=0}^{\infty} k \cdot p_k - \theta\right) .$$

The details are omitted due to lack of space. □

2.2 Symmetrization and Extension to the Multivariate Case

A discrete distribution is symmetric with respect to 0, if $p_k = p_{-k}$ for all k being elements of the support. How this can be achieved ?

PROPOSITION 2 ([12, p. 436–437])
Let X be a discrete random variable with support $\mathbb{N}_0$ and Y a random variable with support $\mathbb{Z} \backslash \mathbb{N}$ and $Y \stackrel{d}{=} -X$. If X and Y are stochastically independent, then $Z \stackrel{d}{=} X + Y$ possesses a symmetric distribution with respect to 0. □

Here, X possesses geometrical distribution with $P\{X = k\} = p \cdot (1-p)^k$, so that for Y holds $P\{Y = k\} = p \cdot (1-p)^{-k}$. The convolution of both distributions leads to the distribution of Z. We distinguish two cases:

†Convention: $0 \cdot \log 0 = 0$.

I. Let $k \geq 0$. Then $P\{Z = k\} = P\{X + Y = k\} =$

$$\begin{aligned}
\sum_{j=0}^{\infty} P\{X = j + k\} \cdot P\{Y = -j\} &= \sum_{j=0}^{\infty} p \cdot (1-p)^{j+k} \cdot p \cdot (1-p)^j = \\
p^2 \, (1-p)^k \sum_{j=0}^{\infty} \left[(1-p)^2\right]^j &= \frac{p}{2-p} \, (1-p)^k \quad .
\end{aligned}$$

II. Let $k < 0$. Then $P\{Z = k\} = P\{X + Y = k\} =$

$$\begin{aligned}
\sum_{j=0}^{\infty} P\{X = j\} \cdot P\{Y = -j + k\} &= \sum_{j=0}^{\infty} p \cdot (1-p)^{j} \cdot p \cdot (1-p)^{j-k} = \\
p^2 \, (1-p)^{-k} \sum_{j=0}^{\infty} \left[(1-p)^2\right]^j &= \frac{p}{2-p} \, (1-p)^{-k} \quad .
\end{aligned}$$

Thus, the probability function of Z is

$$P\{Z = k\} = \frac{p}{2-p} \, (1-p)^{|k|} \;\; , \; k \in \mathbb{Z} \tag{4}$$

with $\mathrm{E}[Z] = 0$ and $\mathrm{Var}[Z] = 2\,(1-p)\,/\,p^2$. Next, we consider the multivariate case. Two questions are of special interest: Does the extension to the multivariate case remain symmetric ? What is the expectation and variance of the step size in n-dimensional space ?

Definition 2 ([3])
A discrete multivariate distribution with support $\mathbb{Z}^n$ is called ℓ_1-symmetric if the probability to generate a specific point $k \in \mathbb{Z}^n$ is of the form

$$P\{X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n\} = g(\|k\|_1) \; ,$$

where $k = (k_1, \ldots, k_n)'$ with $k_i \in \mathbb{Z}$ and $\|k\|_1 = \sum_{i=1}^{n} |k_i|$ denotes the ℓ_1-norm. □

We shall generate the mutation vector Z by n independent random variables Z_i with probability function (4). Let $k \in \mathbb{Z}^n$ be any point of length $\|k\|_1 = s$, s fixed. Then $P\{Z_1 = k_1, Z_2 = k_2, \ldots, Z_n = k_n\} =$

$$\begin{aligned}
\prod_{i=1}^{n} P\{Z_i = k_i\} &= \left(\frac{p}{2-p}\right)^n \prod_{i=1}^{n} (1-p)^{|k_i|} = \\
\left(\frac{p}{2-p}\right)^n (1-p)^{\sum_{i=1}^{n} |k_i|} &= \left(\frac{p}{2-p}\right)^n (1-p)^{\|k\|_1} \; .
\end{aligned}$$

That means that each point with length s is sampled with the same probability. Therefore, the multivariate distribution is ℓ_1-symmetric. To determine the expectation and variance of the step size in n-dimensional space, we require the expectation of random variable $|Z_1|$. Since $P\{|Z_1| = j\} = P\{Z_1 = j\} + P\{Z_1 = -j\}$ for $j \in \mathbb{N}$ and $P\{|Z_1| = 0\} = P\{Z_1 = 0\}$ we obtain

$$P\{|Z_1| = j\} = \begin{cases} \dfrac{p}{2-p} & \text{, if } j = 0 \\[2ex] \dfrac{2\,p}{2-p}\,(1-p)^j & \text{, if } j \in \mathbb{N} \end{cases} .$$

Straightforward calculations lead to

$$\mathrm{E}[\,|Z_1|\,] = \frac{2\,(1-p)}{p\,(2-p)} \quad \text{and} \quad \mathrm{V}[\,|Z_1|\,] = \frac{2\,(1-p)}{p^2}\left[1 - \frac{2\,(1-p)}{(2-p)^2}\right] .$$

As the random variables $|Z_i|$ are stochastically independent we obtain for vector Z

$$\mathrm{E}[\,\|Z\|_1\,] = n \cdot \mathrm{E}[\,|Z_1|\,] \quad \text{and} \quad \mathrm{V}[\,\|Z\|_1\,] = n \cdot \mathrm{V}[\,|Z_1|\,] \;. \tag{5}$$

It should be noted that (4) could have been derived from the ansatz

$$P\{Z = k\} = \frac{q^{|k|}}{\sum\limits_{j=-\infty}^{\infty} q^{|j|}} = \frac{q^{|k|}}{1 + 2\sum\limits_{j=1}^{\infty} q^j} = \frac{(1-q)\cdot q^{|k|}}{1+q}$$

with $q = 1 - p$. Another approach with

$$P\{Z = k\} = \frac{q^{k^2}}{\sum\limits_{j=-\infty}^{\infty} q^{j^2}} = \frac{q^{k^2}}{1 + 2\sum\limits_{j=1}^{\infty} q^{j^2}} = \frac{q^{k^2}}{\vartheta_3(q,0)} \;, \tag{6}$$

where $\vartheta_3(q,x)$ denotes the third Theta function [1, entry 16.27.3], gives exactly the distribution with maximum entropy under the constraints $P\{Z = k\} = P\{Z = -k\}$ for all $k \in \mathbb{Z}$ and $\mathrm{V}[Z] = \sigma^2$ (see Appendix). There are, however, three problems:

1. The multivariate version of (6) is symmetric with respect to ℓ_2-norm.
2. The control of (6) with parameter σ would require the approximate determination of the zeros of a highly nonlinear equation (see Appendix), as soon as parameter σ is altered.
3. There exists neither a closed form of $\vartheta_3(q,0)$ nor of its partial sums, so that the generation of the associated random numbers must be expensive and inaccurate.

While the first point is a matter of taste, the last two points are prohibitive for the usage of distribution (6) in an optimization algorithm. The last two problems do not occur with distribution (4). Firstly, the random variable Z

can be generated by the difference of two geometricly distributed independent random variables (both with parameter p) and a geometric random variable G is obtained as follows: Let U be uniformly distributed over $[0,1) \subset \mathbb{R}$. Then

$$G = \left\lfloor \frac{\log(1-U)}{\log(1-p)} \right\rfloor$$

is geometricly distributed with parameter p. Secondly, the distribution could be controlled by the mean step size (5), so that one obtains

$$S = n \cdot \frac{2\,(1-p)}{p\,(2-p)} \quad \Leftrightarrow \quad p = 1 - \frac{S/n}{(1+(S/n)^2)^{1/2}+1}\ , \tag{7}$$

where $S = \mathrm{E}[\, \|Z\|_1 \,]$.

2.3 Parameter Control

As soon as an probabilistic optimization algorithm approaches the optimum, the step size of the algorithm must decrease to balance the probability to generate a new successful point. There are several ways to control the parameters of the mutation distribution. A simple idea is to decrease the step size s by a deterministic schedule, say $s_t = s_0/t$ or $s_t = \beta^t \cdot s_0$ with $\beta \in (0,1)$. This is sufficient for problems with only one local (= global) optimum. But for problems with more than one local optimum such a schedule would force the algorithm to approach the closest local optimum. Therefore, it might be useful to offer the chance to increase the step size, too. Evolution strategies employ the following technique [2]: An *individual* consists of a vector $x \in \mathbb{Z}^n$ and a mutation control parameter $s \in \mathbb{R}_+$. Both x and s are regarded as *genes* changeable by genetic operators. First, the mean step size s is *mutated* by multiplication with a lognormally distributed random variable: $s' = s \cdot \exp(N)$, where N is a normally distributed random variable with zero mean and variance $\tau^2 = 1/n$. Thus, the mean step size is decreased or increased by a factor with the same probability and it is likely that a better step size will also produce a better point. Since a mean step size below 1 is not useful for integer problems, the mutated mean step size is set to 1 if the value is less than 1.
Finally, vector x is mutated by adding the difference of two independent geometricly distributed random numbers to each vector component. Both geometric random variables have parameter p depending on the new step size s' via (7).

3 Computational Results

3.1 Sketch of the Algorithm

The evolutionary algorithm to be developed here is basically a (μ, λ)-ES [2] (outlined below). Initially, vector x of each individual is drawn uniformly from the starting area $\mathcal{M} \in \mathbb{Z}^n$, which need not contain the global optimum. The initial value of s is chosen proportional to the nth root of the volume of $\mathcal{M}$. Recombination of two individuals is performed as follows: The step size parameters of the parents are averaged and the new vector x is generated by choosing the vector component of the first or second individual with the same probability. Infeasible individuals are sampled anew.

```
initialize μ individuals
calculate the fitness of the individuals
repeat
    do λ times:
        select two individuals at random
        perform recombination to obtain an offspring
        mutate the offspring
        calculate the fitness of the offspring
    od
    select μ best offspring
until termination criterion fulfilled
```

3.2 Test problems

Problem 2.20 (unconstrained) of [14] with $x \in \mathbb{Z}^{30}$: $f_1(x) = -\|x\|_1$ with known solution: $x_i = 0$ with $f(x) = 0$. Starting area $\mathcal{M} = [-1000, 1000]^{30} \subset \mathbb{Z}^{30}$. Initial mean step size $s_0 = 1000/3$.
Problem 1.1 of [14] with $x \in \mathbb{Z}^{30}$: $f_2(x) = -x'\,x$ with known solution: $x_i = 0$ with $f(x) = 0$. Starting area $\mathcal{M} = [-1000, 1000]^{30} \subset \mathbb{Z}^{30}$. Initial mean step size $s_0 = 1000/3$.
Problem 8 of [5] with $x \in \mathbb{Z}^5$:

$$\mathrm{f_3(x) = (15\ 27\ 36\ 18\ 12)\,x - x'} \begin{pmatrix} 35 & -20 & -10 & 32 & -10 \\ -20 & 40 & -6 & -31 & 32 \\ -10 & -6 & 11 & -6 & -10 \\ 32 & -31 & -6 & 38 & -20 \\ -10 & 32 & -10 & -20 & 31 \end{pmatrix} \mathrm{x}$$

Best solutions known: $x = (0\ 11\ 22\ 16\ 6)'$ and $x = (0\ 12\ 23\ 17\ 6)'$ with $f_3(x) = 737$. Starting area $\mathcal{M} = [0, 100]^5 \subset \mathbb{Z}^5$. Initial mean step size $s_0 = 50/3$.
Derived from problem 9 of [5] with $x_i \geq 0$.

$$\mathrm{f_4(x)} = -\sum_{\mathrm{i=1}}^{10} \mathrm{x_i} \cdot \left[\log\left(\frac{\mathrm{x_i + i}}{0.1 + \sum_{\mathrm{i=1}}^{10} \mathrm{x_i}}\right) - \mathrm{d_i}\right] - 6 \cdot (\mathrm{A^2 + B^2 + C^2}) \,,$$

with $d = (6.089, 17.164, 34.054, 5.914, 24.721, 14.986, 24.1, 10.708, 26.662, 22.179)$, $A = x_1 + 2\,x_2 + 2\,x_3 + x_6 + x_{10} - 2$, $B = x_4 + x_5 + x_6 + x_7 - 1$ and $C = x_3 + x_7 + x_8 + x_9 + x_{10} - 1$. Best solution known: $x = (3\ 0\ 0\ 3\ 0\ 0\ 0\ 3\ 0\ 0)'$ with $f(x) \approx 150.533$. Starting area $\mathcal{M} = [50, 150]^{10} \subset \mathbb{Z}^{10}$. Initial mean step size $s_0 = 50/3$.
Problem 2 of [11] with $x_i \geq 0$: $f_5(x) = \prod_{i=1}^{15} [1 - (1 - r_i)^{x_i}]$ with constraints $c'\,x \leq 400$ and $w'\,x \leq 414$ (see [11] for details).
Best solution known: $x = (3\ 4\ 6\ 4\ 3\ 2\ 4\ 5\ 4\ 2\ 3\ 4\ 5\ 4\ 5)'$ with $f_5(x) \approx 0.945613$. Starting area $\mathcal{M} = [0, 6]^{15} \subset \mathbb{Z}^{15}$. Initial step size $s_0 = 2$.

3.3 Results

The evolutionary algorithm was tested with $\mu = 30$ and $\lambda = 100$. The test statistics over 1000 runs for each problem are summarized in tables 1 and 2 below.

	min	max	mean	std.dev.	skew
f_1	106	1519	147.0	96.4	6.68
f_2	115	159	135.6	6.7	0.25
f_3	30	198	107.7	30.5	-0.17
f_4	38	769	94.3	85.9	3.48
f_5	16	49434	582.6	2842.5	9.90

Table 1: Statistics of the first hitting time distribution.

Table 1 shows the minimum, maximum, mean, standard deviation and skewness of the sample of first hitting times (of the global maximum) obtained from 1000 independent runs. Problems f_1 and f_2 only have one local (= global) maximum. Surprisingly, the distribution of the first hitting times is skewed to the right significantly for problem f_1, while the tails of the distribution for problem f_2 are balanced. But as can be seen from table 2 containing the percentiles of the ordered sample, at least 90% of all runs solved problem f_1 in not more than 140 generations. The reasons for these different characteristics are unknown at the moment.

	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.97	.99
f_1	118	120	123	125	126	128	131	134	140	276	416	624
f_2	128	130	132	133	135	137	139	141	144	147	149	152
f_3	65	81	94	102	110	118	125	133	145	155	161	173
f_4	46	49	53	57	62	70	82	108	194	260	331	455
f_5	24	26	28	31	34	38	44	84	487	2245	5655	13038

Table 2: Percentiles of the first hitting time distribution

Problems f_3, f_4 and f_5 possess more than one local maxima. Seemingly, problems f_3 and f_4 do not cause difficulties, maybe due to the low dimensionality of the problems. The results for problem f_5 reveal that this problem is solvable for 80% of the runs in less than 100 generations, but there must be local maxima on isolated peaks preventing the population to generate better solutions by recombination, so that mutation is the only chance to move from the top of a locally optimal peak to the global one.

4 Conclusions

The principle of maximum entropy is a useful guide to construct a mutation distribution for evolutionary algorithms to be applied to integer problems. This concept can and should be used to construct mutation distributions for arbitrary search spaces, because special *a priori* knowledge of the problem type

may be formulated as constraints of the maximum entropy problem, so that this knowledge (and only this) will be incorporated into the search distribution.
The evolutionary algorithm developed here demonstrated the usefulness of this approach. It was able to locate the global optima of five nonlinear integer problems relatively fast for at least 80 percent of 1000 independent runs per problem. Clearly, a test bed of only five problems may not be regarded as a basis to judge the power of the algorithm, but these first results and the 'simplicity' of the mutation distribution are encouraging for further theoretical research.

References

[1] M. Abramowitz and I.A. Stegun (Eds.). *Handbook of Mathematical Functions.* Dover Publications, New York, 1965.

[2] T. Bäck and H.-P. Schwefel. An overview of evolutionary algorithms for parameter optimization. *Evolutionary Computation*, 1(1):1–23, 1993.

[3] K.-T. Fang, S. Kotz, and K.-W. Ng. *Symmetric Multivariate and Related Distributions.* Chapman and Hall, London and New York, 1990.

[4] D.A. Gall. A practical multifactor optimization criterion. In A. Lavi and T.P. Vogl, editors, *Recent Advances in Optimization Techniques*, pages 369–386. Wiley, New York, 1966.

[5] A. Glankwahmdee, J.S. Liebman, and G.L. Hogg. Unconstrained discrete nonlinear programming. *Engineering Optimization*, 4:95–107, 1979.

[6] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison Wesley, Reading (MA), 1989.

[7] J.H. Holland. *Adaptation in natural and artificial systems.* The University of Michigan Press, Ann Arbor, 1975.

[8] E.T. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, 4(3):227–241, 1968.

[9] N.L. Johnson and S. Kotz. *Distributions in Statistics: Discrete Distributions.* Houghton Mifflin, Boston, 1969.

[10] R.C. Kelahan and J.L. Gaddy. Application of the adaptive random search to discrete and mixed integer optimization. *International Journal for Numerical Methods in Engineering*, 12:289–298, 1978.

[11] R. Luus. Optimization of system reliability by a new nonlinear integer programming procedure. *IEEE Transactions on Reliability*, 24(1):14–16, 1975.

[12] P.H. Müller (Ed.). *Wahrscheinlichkeitsrechnung und mathematische Statistik: Lexikon der Stochastik.* Akademie Verlag, Berlin, 5th edition, 1991.

[13] I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Frommann–Holzboog Verlag, Stuttgart, 1973.

[14] H.-P. Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie.* Birkhäuser, Basel, 1977.

Appendix: A Family of Maximum Entropy Distributions

The question addressed here is: Which discrete distribution with support $\mathbb{Z}$ is symmetric with respect to 0, has a variance $\sigma^2 > 0$ or a mean deviation s and possesses maximum entropy ? The answer requires the solution of the following nonlinear optimization problem: $-\sum_{k=-\infty}^{\infty} p_k \log p_k \rightarrow \max$ subject to

$$\begin{aligned} p_k &= p_{-k} \quad \forall k \in \mathbb{Z} , & (8)\\ \sum_{k=-\infty}^{\infty} p_k &= 1 , & (9)\\ \sum_{k=-\infty}^{\infty} |k|^m \, p_k &= \sigma^2 & (10)\\ p_k &\geq 0 \quad \forall k \in \mathbb{Z} . & (11) \end{aligned}$$

We may neglect condition (11), if the solution of the surrogate problem fulfills these inequalities. We may therefore differentiate the Lagrangian

$$L(p,a,b) = -\sum_{k=-\infty}^{\infty} p_k \log p_k + a \cdot \left(\sum_{k=-\infty}^{\infty} p_k - 1 \right) + b \cdot \left(\sum_{k=-\infty}^{\infty} |k|^m \, p_k - \sigma^2 \right)$$

to obtain the necessary condition (9) and (10) and $-1 - \log p_k + a + b\,|k|^m = 0$ or alternatively

$$p_k = e^{a-1} \cdot (e^b)^{|k|^m} \quad \forall k \in \mathbb{Z} . \tag{12}$$

Exploitation of the symmetry condition (8) and substitution of (12) in (9) leads to

$$\sum_{k=-\infty}^{\infty} p_k = p_0 + 2 \cdot \sum_{k=1}^{\infty} p_k = e^{a-1} \cdot \left(1 + 2 \cdot \sum_{k=1}^{\infty} (e^b)^{|k|^m} \right) = 1 . \tag{13}$$

Let $q = e^b < 1$ so that $S(q) := \sum_{k=1}^{\infty} q^{|k|^m} < \infty$ and $q \cdot S'(q) = \sum_{k=1}^{\infty} |k|^m \, q^{|k|^m}$. Then condition (13) becomes

$$e^{1-a} = 1 + 2 \cdot S(q) = \begin{cases} (1+q)/(1-q) & \text{, if } m = 1 \\ \vartheta_3(q,0) & \text{, if } m = 2 \end{cases} , \tag{14}$$

where $\vartheta_3(q,z)$ denotes the third Theta function [1]. Substitution of (12) in (10) yields

$$\sum_{k=-\infty}^{\infty} |k|^m \, p_k = 2 \cdot \sum_{k=1}^{\infty} |k|^m \, p_k = 2\, e^{a-1} \cdot \sum_{k=1}^{\infty} |k|^m \, q^{|k|^m} = 2\, e^{a-1} \cdot q \cdot S'(q) = \sigma^2 \ ,$$

so that with substitution of (14) for $m = 2$ one obtains

$$\frac{2\, q\, S'(q)}{\vartheta_3(q,0)} = \sigma^2 , \tag{15}$$

while $m = 1$ gives (7) with $q = 1 - p$ and $s/n = \sigma^2$. The value of q in (15) for given σ can be determined numerically only. Substitution of (14) in (12) gives (6) for $m = 2$ and (4) for $m = 1$.