On bivariate telescopic distributions of probability and their application

Petr Volf

Institute of Information Theory and Automation, Czech Academy of Sciences, Praha, Czech Republic.

Contributing authors: volf@utia.cas.cz;

Abstract

This paper describes an interesting class of discrete probability distributions that arise from the discretization of continuous distributions. They are sometimes also called telescopic distributions, a term inspired by the concept of "telescoping series" in mathematics. After some introductory remarks, we will focus primarily on deriving two-dimensional telescopic distributions, i.e., discretized two-dimensional continuous probability distributions, which are typically constructed using a suitable copula. We will then present examples using both simulated and real data to demonstrate the advantages and potential applications of these models. Among other things, we will derive and apply the "telescopic" form of the bivariate negative binomial distribution.

Keywords: Discretization of continuous distribution, telescopic distribution, bivariate distribution, copula, bivariate negative binomial model.

1 Introduction and preliminaries

This study deals with an interesting class of discrete distributions of probability, which are constructed from continuous distributions by their discretization. They are sometimes also referred to as telescopic distributions, their name is derived most likely from the term "telescoping series" (cf. *https://en.wikipedia.org/wiki/Telescoping_series*). In fact, the term telescopic probability distribution is (so far) used rather rarely; probably one of the most basic articles on these distributions is by Roknabadi et al. (2009). The definition offered there is as follows:

Definition

We say that a discrete non-negative random variable X has a telescopic distribution if its probabilities can be written in the form

$$P(X = K) = q^{z(K;\theta)} - q^{z(K+1;\theta)}, \tag{1}$$

where $K = 0, 1, 2, ...$, $q \in (0, 1)$, function $z(x; \theta)$ is for each θ (which is the distribution parameter) non-decreasing in x, $z(0; \theta) = 0$, $\lim_{x\to\infty} z(x; \theta) = \infty$.

This formulation seems to be unnecessarily complicated. It suffices to consider some continuous non-negative random variable Y with distribution function $F(y)$, and then define the discrete probabilities of X as

$$P(X = K) = F(K + 1) - F(K) = \bar{F}(K) - \bar{F}(K + 1), \tag{2}$$

where $\bar{F}(y) = 1 - F(y)$ is the corresponding survival function. And of course $\lim_{x\to\infty} \bar{F}(x) = 0$, so that $P(X \geq K) = \bar{F}(K)$.

In fact, the variable Y does not need to be restricted to the positive semi-axis, it is sufficient to set $\bar{F}(0) = 1$ in (2). Eventually, the probabilities should be normalized if for the original variable $P(Y < 0) > 0$, to keep $\sum_{K=0}^{\infty} P(X = K) = 1$.

This actually brings us to a concept that is already more common, namely the discretization of continuous probability distributions. Essentially, such discretization may also be applied to continuous distributions defined on the whole real axis, e.g., as in Roy (2003), where the discretized normal distribution is used to estimate the achieved tolerances in reliability analysis problems.

In the following section, we will first comment on some properties of these distributions and recall that some standard discrete distributions also have a "telescopic" form (2). We will then propose a procedure for constructing two-dimensional discrete probability distributions in a similar manner, i.e., by discretizing bivariate continuous distributions. This approach not only extends and enriches the possibilities of modelling two-dimensional discrete data, but also simplifies their processing. This will be demonstrated using examples with both simulated and real data.

2 Remarks on properties of telescopic distributions

In Roknabadi et al. (2009) the main focus is on a certain analogy between the properties of related continuous and discretized distributions, with an emphasis on models applied in the field of reliability analysis. Our perspective here is broader and more general. Before extending the idea of discretization to bivariate distributions, in the following remarks we will show that some standardly used discrete distributions satisfy the above definition (2), i.e., they may be expressed as discretized continuous distributions.

1. The term "telescopic" (∼extensible, adjustable, e.g., a type of telescope or ladder) obviously refers to the property of partial sums in a sequence of probabilities (1)

or (2), namely that

$$\sum_{k=M}^{N} P(X=k) = q^{z(M;\theta)} - q^{z(N+1;\theta)} = \bar{F}(M) - \bar{F}(N+1).$$

Let us mention here another interesting property; it is quite easy to show that

$$E(X) = \sum_{K=0}^{\infty} K \cdot (\bar{F}(K) - \bar{F}(K+1)) = \sum_{K=1}^{\infty} \bar{F}(K) \leq \int_0^{\infty} \bar{F}(y)dy.$$

Recall that the last expression equals $E(Y)$ provided it is finite.

2. As already mentioned, the idea of discretizing a continuous distribution as in (2) is not new. Let us imagine a continuous variable Y observed in such a way that only the counts of values in given intervals are known. Then an analogy to the relation (2) could be applied and the relative frequencies computed. They estimate the distribution function of Y at the boundaries of intervals, which is then compared with the model distribution function. A traditional example is the estimation of parameters of the Weibull distribution, which is actually a classic estimation method from the "pre-computer" era.
3. Thus, by applying (2) a broad class of discrete probability distributions is obtained. They are essentially quite simple (just as simple as the continuous distribution from which the discrete distribution is derived). This greatly enriches our choice of parametric models for discrete data. And these models have one advantage: they are in fact specified by their distribution function, whereas the standard discrete distributions usually have their distribution functions given by partial sums of probabilities, which is not very comfortable. On the other hand, it is well known that also some standard discrete distributions can be expressed in "telescopic" form (2). We shall recall these expressions for several of the most frequently used types.
4. It is easy to show that the geometric distribution can be obtained by (2) from the exponential distribution:

   $P(X=K) = \exp(-\lambda K) - \exp(-\lambda(K+1)) = \exp(-\lambda K)(1-\exp(-\lambda))$.

   It follows that this defines a geometric distribution with probability of success $p = 1-\exp(-\lambda)$, $1-p = \exp(-\lambda)$.
5. The Poisson distribution can also be expressed in this "telescopic" way: Consider a homogeneous Poisson process with parameter λ, i.e., the series of endpoints of successively arranged independent exponential random variables with this parameter. Then the variable X equal to the number of these points in the interval $(0,1)$ has the Poisson distribution, again with the same λ. At the same time, the variable T_K equal to the length of partial sum of K of these exponential variables has the gamma distribution $Gamma(K,\lambda)$, K is the shape parameter, λ intensity ($1/\lambda$ is the scale). Its density equals $f_K(z) = \lambda^K z^{K-1} e^{-\lambda z}/\Gamma(K)$, distribution function $F_K(z) = \gamma(K,\lambda z)/\Gamma(K)$. Here $\Gamma(K)$ is the complete gamma function, $\gamma(a,b) = \int_0^b t^{a-1}e^{-t}dt$ is the incomplete (lower) gamma function.

Hence the Poisson random variable X from above fulfils the following condition:

$$P(X = K) = P(T_K < 1) - P(T_{K+1} < 1) = \frac{\gamma(K, \lambda)}{\Gamma(K)} - \frac{\gamma(K+1, \lambda)}{\Gamma(K+1)}. \qquad (3)$$

The function $S(t; \lambda) = \gamma(t, \lambda)/\Gamma(t)$ is called the regularized incomplete gamma function. Simultaneously, for each $\lambda > 0$, as a function of $t > 0$, it is non-increasing, with $\lim_{t\to 0+} S(t) = 1$ and $\lim_{t\to\infty} S(t) = 0$, Therefore, it could be a survival function of some continuous random variable. Thus, (3) corresponds to the definition of the telescopic type distribution.

6. It follows that also the negative binomial distribution is of this type, because it can be obtained from the Poisson distribution by a mixing, in the following way: let us assume that the Poisson parameter λ is random with a gamma distribution. Specifically, let the parameter λ of the Poisson variable X have a priori $Gamma(r, p/(1-p))$ distribution (denote its density by $g(\lambda; r, p)$), then the resulting random variable (denote it by Z) has a negative binomial distribution with parameters r and p, and it holds that:

$$P(Z = K; r, p) = \int_0^\infty P(X = K; \lambda) g(\lambda; r, p) d\lambda =$$

$$= \int_0^\infty S(K; \lambda) g(\lambda; r, p) d\lambda - \int_0^\infty S(K+1; \lambda) g(\lambda; r, p) d\lambda =$$
$$= I_{1-p}(K, r) - I_{1-p}(K+1, r),$$

where $I_x(a, b) = B(x; a, b)/B(a, b)$ is a (lower) regularized incomplete beta function. Here $B(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt$ is an incomplete (lower) beta function, $B(a, b)$ is a complete beta function.

In such a way the distribution functions of Poisson or negative binomial distributions may be expressed in a compact form. The advantage of such a characterization will be shown in the next section dealing with bivariate distributions. It will enable rather easy construction of different variants of bivariate distributions, with the aid of suitable copulas.

7. Similarly, it is well known that the binomial distribution with parameters N, p can also be expressed in such a form, as its distribution function fulfils $F(k) = P(X \leq k) = I_{1-p}(N-k, k+1)$, for $k = 0, 1, ..., N$. All these relations can be found in various general sources, for instance in Wikipedia.

3 Bivariate telescopic distributions

It can be said that the advantage of discretized continuous models becomes much more apparent when working with bivariate discrete data. Constructing bivariate versions of classical discrete distributions is usually quite complicated. It is easy to extend Bernoulli's '0–1' distribution or to construct the bivariate geometric distribution from it, as in Basu and Dhar (1995), Marshall and Olkin (1985), but for other discrete distributions this could be a problem. For example, a version of the bivariate Poisson distribution described below is quite popular.

Let X_1, X_2, X_3 be mutually independent random variables with Poisson distribution with parameters $\lambda_1, \lambda_2, \lambda_3$, then $Y = X_1 + X_3$ and $Z = X_2 + X_3$ form a bivariate Poisson distribution. Their covariance is λ_3 (i.e., positive and given directly by the parameter values, so this may be a limitation). Marginal distributions are Poisson again, with parameters $\lambda_1 + \lambda_3$ and $\lambda_2 + \lambda_3$; their joint probability function has no longer a simple form. This and other kinds of bivariate discrete distributions are presented, for instance, in Kocherlakota and Kocherlakota (1992) or Johnson et al. (1997). A general procedures of constructing multivariate discrete distributions are described also in Ng et al. (2010), with an example of negative binomial distribution. In these models, the formulas for the joint probability are, as a rule, even more complicated.

On the other hand, the method based on the discretization of continuous distributions is relatively simple. The first step is to choose a suitable bivariate continuous distribution. This is not a problem when using appropriate marginal continuous distributions and the chosen copula. This approach has, of course, already been proposed, essentially again attempting to model a continuous distribution based on discretized data, similarly to what is shown for the univariate Weibull distribution in Remark 2. Such an approach is presented, for example, in Barbiero (2022). In the present study we are actually looking at the problem in the opposite way: we assume discrete data and try to find a suitable and simple discrete model for them, as well as a fairly universal method for such modelling.

Recall that each copula $C(u, v, \theta)$ represents a bivariate uniform distribution function of random variables U, V on $[0, 1]^2$, θ being the corresponding parameter. If $\bar{F}_S(s), \bar{F}_T(t)$ are the marginal survival function of two continuous variables S, T, then the copula defines the joint survival function as

$$\bar{F}_2(s, t) = P(S > s, T > t) = C(\bar{F}_S(s), \bar{F}_T(t), \theta). \tag{4}$$

Naturally, different copulas lead to different associated models, even with different correlations (some copulas allow for the correlation in only a sub-interval of $[-1, 1]$), the parameter θ has to be estimated. But then the discretization providing the joint distribution of some discrete random variables X, Y is straightforward:

$$P(X = K, Y = L) = \bar{F}_2(K, L) - \bar{F}_2(K, L+1) - \bar{F}_2(K+1, L) + \bar{F}_2(K+1, L+1). \tag{5}$$

Thus, in this simple way a large set of sufficiently flexible two-dimensional discrete distributions can be obtained. From the last remarks of the previous section it follows that this method also applies to several standard discrete distributions.

The choice of a suitable copula is based on comparing the fit of the models, in fact there is no objective rule for the best copula selection. One of key points could be the choice of a model that captures the correlation adequately. Notice also that, in general, the marginal distributions of S and T may be of different types.

However, one should be aware that the construction of a two-dimensional distributions from two marginal distributions connected by a copula may not capture the dependence sufficiently, as the interdependence of the two variables can be much more complex. Sometimes another problem may arise concerning the properties of distributions constructed with the aid of copula. For instance, in univariate cases,

both exponential and geometric distributions are "memoryless", which means that $P(X > t|X > s) = P(X > t - s)$ for $t > s$. While the bivariate geometric distribution proposed by Basu and Dhar (1995) still retains this property, bivariate exponential distributions constructed via copula do not keep it. Hence the same concerns the bivariate geometric distributions obtained from them by discretization (5). To illustrate this difference, both versions of bivariate geometric distribution will be compared in the following simulated examples.

4 Application

In this section, the construction of two-dimensional telescopic distributions in accordance with (5) will be shown on relatively simple examples. The data will be available in the form of paired random sample of size N from a couple of discrete random variables X, Y. In the first step, our effort will be to find suitable marginal distributions for both X and Y. In addition to several discretized continuous models, we will use also several standard discrete distributions (Poisson, geometric, negative binomial). The parameters of chosen distributions will be estimated with the aid of the maximum likelihood method. Then, the model goodness-of-fit to the data will be checked both graphically and with the Pearson's chi-squared test. The data will usually be limited to a few values or classes, for example $0, 1, 2, ..., M - 1$ and $\geq M$. In fact, even the correct use of the chi-squared test requires aggregation of classes with low probability. Then, also taking into account estimated correlation $\rho(X, Y)$, an appropriate copula will be selected and the corresponding bivariate discrete distribution constructed in accordance with (4) and (5). In doing so, we use the parameters of the marginal distributions already estimated, so that in this step we only need to estimate the copula parameter θ. And again, the goodness-of-fit will be assessed with the chi-squared test. It means that the observed frequencies O_j will be compared with the expected frequencies, E_j, predicted by the model. The criterion will therefore use the quantity

$$C^2 = \sum_j (O_j - E_j)^2 / E_j,$$

where the summation is over all distinct classes. The value of C^2 will be compared to the 95% quantile of the chi-squared distribution with degrees of freedom equal to $(N_c - 1 - N_p)$, where N_c is the number of classes and N_p the number of estimated model parameters. Thus, the basic requirement for the success is a good agreement for both the bivariate model and the two marginal models.

Simultaneous estimation of all parameters of the bivariate models has been attempted, too. The results were practically identical, with minor deviations caused by the method of randomized search for optimal parameters values (optimal in the sense of maximum likelihood). The advantage of two-step method is its much smaller computational complexity, as each step solves only a relatively simple MLE task, while the numerical method used for joint estimation takes much more time and may have problems with global optimum finding.

The equivalence of both approaches, sequential and joint, could follow from the "two-level" structure of distribution constructed with copula. The first level is the

copula itself defining a joint distribution of two random variables U, V on $[0,1]^2$, the second level corresponds to their transformation to required variables X, Y. In fact, even the random generation of data from the bivariate distribution X, Y joined by a copula proceeds standardly according to these two steps: First, a representation of U, V is generated from their distribution given by the copula, then both U and V are transformed, separately, to X and Y. However, in the case of discrete distribution of X, Y with a finite set of values, a direct method of generation is easier (it avoids the generation from copula): First, all probabilities $P(X = K, Y = L)$ are computed, again according to (4) and (5), then a representation of X, Y is generated from them. Naturally, both approaches lead to equivalent results.

The likelihood function for given data and parameters has the form

$$L = \prod_{i=0}^{M-1} \prod_{j=0}^{M-1} P(X = i, Y = j)^{n(i,j)} \cdot \prod_{i=0}^{M-1} P(X = i, Y \geq M)^{n(i,M)} \cdot$$

$$\cdot \prod_{j=0}^{M-1} P(X \geq M, Y = j)^{n(M,j)} \cdot P(X \geq M, Y \geq M)^{n(M,M)},$$

where $n(i,j)$ are observed frequencies of values (classes) (i,j), and probabilities follow from (4) and (5). However, as the computation of derivatives of the log-likelihood function is rather complicated, optimal value of θ is obtained as a rule via random search. That is also why the confidence intervals of parameters are not evaluated directly (their computation needs the second derivatives of the log-likelihood). In simulated examples each experiment is repeated many times, thus the empirical distribution of parameters estimates is obtained and the empirical versions of confidence intervals can be derived from it.

4.1 Examples with simulated data

The aim of this section is to verify the proposed model fitting method using artificial examples. Specifically, the model construction will be demonstrated using the example of a bivariate geometric distribution. It will be derived according to (5) by discretizing the bivariate exponential distribution, where this bivariate exponential distribution will be constructed according to (4) using the Morgenstern copula. The Morgenstern copula has the following form:

$$C(u, v, \theta_M) = uv(1 + \theta_M(1 - u)(1 - v)).$$

Its parameter θ_M lies in $[-1, 1]$, correlation $\rho(U, V)$ equals $\theta_M/3$, while Kendal's $\tau(U, V)$ equals $2\theta_M/9$.

The parameters were specified as follows: random variables X and Y had the geometric distributions with parameters $p_x = 0.4, p_y = 0.5$, respectively, the parameter of copula was $\theta_M = 0.8$. Sample of N=500 items was randomly generated, values of both variables were aggregated to $0, 1, 2, 3, 4+$, where the last class contained values 4 and larger. Then the joint distribution of (X, Y) had 25 classes. Since the model had

	p_x	p_y	θ_M	C^2_M	θ_G	C^2_G
mean	0.3998	0.5009	0.7969	20.2996	1.2548	21.2098
median	0.3999	0.5005	0.8030	20.1259	1.2530	21.4291
5% quantile	0.3761	0.4744	0.5750	11.7869	1.1740	12.2730
95% quantile	0.4253	0.5289	0.9760	32.4143	1.3515	34.0081
st. deviation	0.0149	0.0166	0.1269	6.3943	0.0535	6.7023

Table 1 Sample characteristics (from 1000 replications) of estimated p_x, p_y, copula parameters θ_M, θ_G in models with Morgenstern and Gumbel copula, and characteristics of values of chi-squared criterion C^2 for both models.

3 parameters, the chi-squared criterion for the goodness-of-fit test had 21 degrees of freedom, critical 95% quantile of corresponding chi-squared distribution was 32.670.

Data generation and parameters estimation were repeated 1000 times, so that for each parameter estimate a sample of 1000 values was obtained. During the data analysis process, models with two different copulas were compared. Namely, in addition to the Morgenstern copula, i.e., the copula used in data generation, the Gumbel copula was also used for data explanation:

$$C(u, v, \theta_G) = exp(-[(-\ln(u))^{\theta_G} + (-\ln(v)).^{\theta_G}]^{1/\theta_G}).$$

Its parameter $\theta_G \geq 1$ allows for any non-negative correlation between U and V, its Kendall's $\tau(U, V)$ equals $1-1/\theta_G$. Table 1 presents the sample characteristics of results (i.e. computed from 1000 replications of the experiment).

In the second set of experiments, the data were generated from the classical bivariate geometric distribution model of Basu and Dhar (1995). This distribution is fully defined by the bivariate Bernoulli distribution. Our choice of parameters was $p_{00} = 0.35, p_{11} = 0.25, p_{10} = 0.15, p_{01} = 0.25$, which again led to $p_x = p_{10} + p_{11} = 0.4, p_y = p_{01} + p_{11} = 0.5$. In order to analyze possible differences between this model and models obtained via copulas, these data were fitted by distributions constructed according to (4) and (5). Again, models with both the Morgenstern and Gumbel copulas were considered. As above, each generated sample had 500 items, values were aggregated to $0, 1, 2, 3, 4+$, the sampling was repeated 1000 times. The results are displayed in Table 2.

	p_x	p_y	θ_M	C^2_M	θ_G	C^2_G
mean	0.4002	0.4998	0.6260	36.7120	1.2210	33.0834
median	0.3996	0.4997	0.6405	35.6052	1.2210	32.2361
5% quantile	0.3760	0.4734	0.3945	20.6897	1.1330	18.1480
95% quantile	0.4267	0.5265	0.8322	57.0888	1.3160	51.6502
st. deviation	0.0153	0.0165	0.1290	11.0587	0.0563	10.1499

Table 2 The same characteristics as in Table 1, now in the case of data generated from the Basu-Dhar model.

Table 1 shows that the estimates of the parameters of the marginal distributions p_x and p_y are, as expected, quite accurate. The same can be said about the estimate of the

	Y=0	1	2	3	4+	sum
X=0	15	15	4	2	1	37
1	17	18	9	3	5	52
2	4	16	12	6	5	43
3	2	6	5	2	4	19
4+	2	4	5	0	4	15
sum	40	59	35	13	19	166

Table 3 Observed numbers of accidents.

parameter θ_M, which, however, exhibits quite a large variation. A criterion value C_M^2 of around 20 is quite acceptable. Of interest are the last two columns, which show that even the Gumbel copula is capable of fitting the data, naturally with a different value of the theta parameter. Note also that in both cases the values of theta correspond roughly to the same Kendal's τ. This result seems to support the conjecture that the copulas performances are quite similar. Simultaneously, however, Table 2 shows that the applicability of copula-based models is rather limited. The example indicates that the structure of relationship between X and Y in the Basu-Dhar model is of a different kind. Both copulas yield roughly equally imprecise results; the criterion values of best models are generally greater than the critical test value. And again, the results (precision of fit) for both copulas are comparable, though, naturally, the optimal theta parameters differ.

4.2 Real data example

The data of this example have been analyzed many times, e.g., in Edwards and Gurland (1961) or Kopociňska (1999), first appearing in Farmer and Chambers (1939). The data display the numbers of accidents of 166 London bus drivers, comparing the accident rates in two subsequent years. Already Farmer and Chambers have noticed a kind of stability in the number of accidents for individual drivers: *Drivers who sustain multiple accidents in the first year of observation have a higher mean accident rate in subsequent years than drivers who have no accidents in their first year.* This was confirmed by the observed positive correlation between the number of accidents in the first year, X, and the second year, Y.

The modelling of these data was mostly done using the bivariate negative binomial distribution, but it will be shown that these solutions are not optimal. The negative binomial distribution can be constructed from the Poisson distribution by mixing (see Remark 6 in the introductory section), or also by the convolution of r independent geometric distributions. For instance, in Kopociňska (1999) the bivariate negative binomial distribution was obtained by the convolution of bivariate geometric distributions, so that both X and Y had the negative binomial distribution with the same parameter r. Then, naturally, it is easy to find better marginal negative binomial distributions. But it turns out that neither these, especially regarding the Y component, are optimal. It will be shown that there exists another solution using the bivariate distribution of "telescopic" form (5) leading to a better fit. And that such a solution is also simpler.

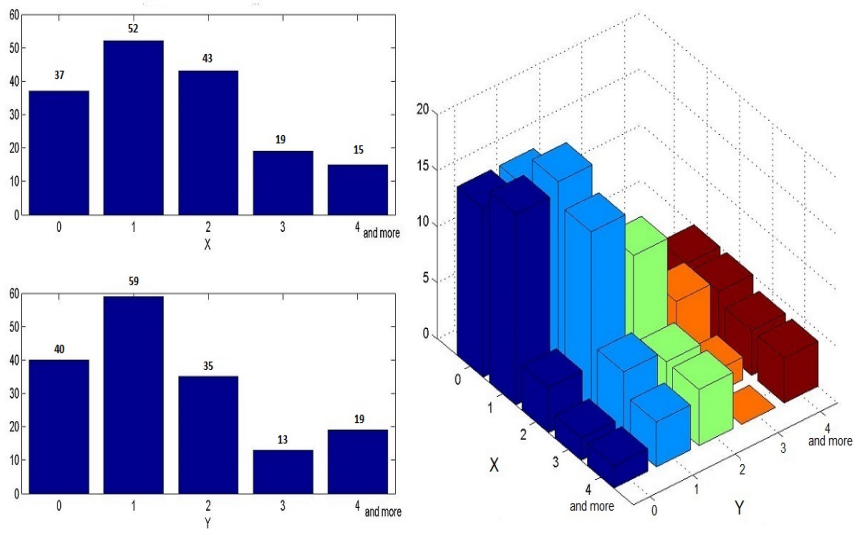

Fig. 1 Left: graphs of observed frequencies of X and Y. Right: bivariate graph of values (X, Y).

	Y=0	1	2	3	4+	sum
X=0	17.865	11.668	4.051	1.745	1.571	36.900
1	13.069	21.335	9.845	4.644	4.470	53.363
2	6.367	14.699	8.808	4.717	5.100	39.691
3	2.699	7.172	5.060	3.034	3.748	21.713
4+	1.418	4.109	3.294	2.214	3.298	14.333
sum	41.418	58.983	31.058	16.354	18.187	166.000

Table 4 Numbers of accidents predicted by the model.

The data, displayed in Figure 1 and Table 3, were available already aggregated, with 5 classes for both X and Y, namely $\{0, 1, 2, 3, 4 \text{ and more}\}$, and contained a total of $N = 166$ pairs of (X_i, Y_i). Hence, all the characteristics reported here are computed from such incomplete data. Figure 1 shows the observed frequencies, both for X and Y separately as well as together. During the search for suitable marginal distributions, it turned out that for the X component, both the negative binomial distribution (with optimal parameters $r_x = 9.994, p_x = 0.863$, yielding the goodness-of-fit criterion value $C^2 = 0.6787$), and the discretized Weibull distribution (with parameters of scale $\alpha_x = 0.2514$ and shape $\beta_x = 1.6422, C^2 = 0.7078$) were acceptable. The chi-squared test had 5-1-2=2 degrees of freedom, its 95% quantile was 5.9915. Good agreement is also shown by the two plots in upper plots of Figure 2. For the Y component, however, these distributions were not quite optimal. This is shown in Figure 2, bottom left, where the data are compared with the negative binomial model (with maximum likelihood parameter estimates $r_y = 4.521$, $p_y = 0.745$, the test value

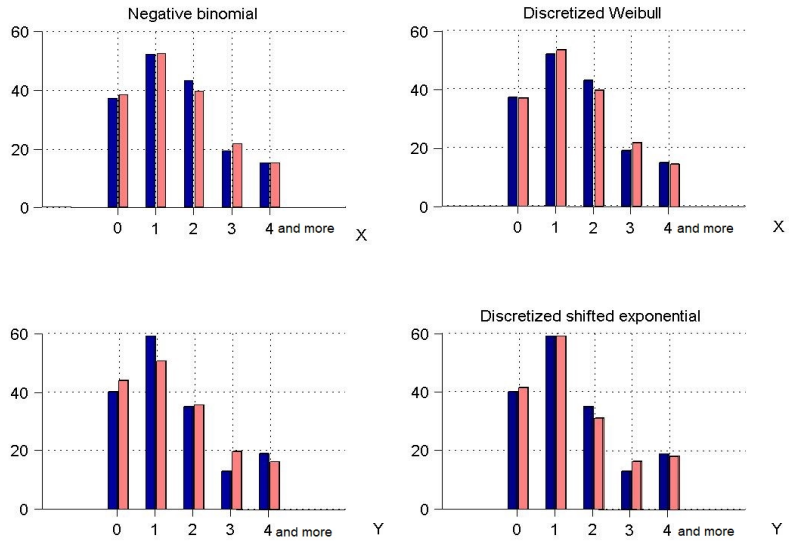

Fig. 2 Plot of the agreement between the frequencies observed (blue, dark) and predicted by the model (pink, light). Left for negative binomial models, top right for the discretized Weibull model used for the variable X, bottom for the shifted exponential model, i.e., used for Y

was $C^2 = 4.524$, p-value of the test 0.1042). Neither discretized Weibull distribution fitted well, the data required a distribution with longer right tail. For example, a discretized log-normal distribution turned out to be a good choice for variable Y ($\mu = 0.5055, \sigma = 0.7172, C^2 = 0.8677$), but the log-normal fit for the component X was rather poor (with $C^2 = 5.0722$). Even from the graphs in Figure 1 it is seen that the distributions of X and Y are quite different.

Continuing the search and preferring simple models, we have finally revealed that the component Y could be modelled well also via a discretized exponential distribution shifted from zero. Its fit to the data is shown in the bottom right plot of Figure 2. Optimal parameters were estimated as $\lambda = 0.6414$, shift $\delta = 0.5528$ (theoretically, this distribution is then defined only for $y \geq \delta$, it was therefore necessary to set $\bar{F}(0) = 1$). The goodness of fit criterion was $C^2 = 1.2729$. And again, no shifted exponential model fitted to component X sufficiently.

Then, a suitable copula had to be selected for the construction of the joint distribution. Since the correlation of X and Y was positive (estimated as $r(X, Y) = 0.3180$), the Gumbel copula was chosen. We used the Weibull distribution for the X variable, the shifted exponential distribution for Y, and the joint discrete distribution of telescopic type was then constructed again according to (5). The maximum likelihood estimate of the copula parameter was $\theta = 1.335$. The goodness-of-fit criterion resulted in $C^2 = 10.8127$, under 5x5-1-5=19 degrees of freedom, the corresponding 95% quantile was 30.1437, p-value of the test 0.930. Predicted numbers of accidents are displayed in Table 4. Just for comparison, theoretical correlation in the Gumbel copula with $\theta = 1.335$ equals $\rho(U, V) = 0.3227$.

4.3 Use of bivariate negative binomial distribution

In the Remark 6 of Section 2 the telescopic form of negative binomial distribution has been derived. Specifically, it was shown that its probabilities can be expressed via the regularized incomplete (lower) beta function:

$$P(X = K) = I_{1-p}(K, r) - I_{1-p}(K + 1, r), \text{ where } I_q(a, b) = B(q; a, b)/B(a, b).$$

The function $I_q(t, b)$, as a function of $t \geq 0$, has the properties of survival function of some continuous random variable, that is why a version of bivariate negative binomial distribution can be constructed using (4) and (5). Parameters p_x, r_x, p_y, r_y have been estimated above, both marginal distributions were then joined with the Gumbel copula. The maximum likelihood estimate of the copula parameter was $\theta = 1.346$, the goodness-of-fit criterion $C^2 = 13.317$, p-value of the test 0.822. These results were slightly worse than the results achieved above, probably due to poorer fit of the negative binomial distribution to component Y.

For comparison, when checking the results in Kopociňska (1999), it was revealed that $C^2 = 14.970$, the model there had one less parameter (it had a common parameter r in the bivariate negative binomial distribution). This bivariate model could be therefore regarded as quite acceptable, too, its disadvantage lied in the complicated form of its bivariate distribution probabilities. Moreover, none of marginal models was satisfying, namely the fit for Y was rejected by the chi-squared test.

This example is also interesting from another perspective. Let us calculate, directly from the observed data, the transition matrix from X to Y. We obtain this from the matrix in Table 3 by dividing each row by its sum. Let us now assume that this represents the transition matrix of a homogeneous Markov chain. After several iterations, the stationary distribution is reached with probabilities 0.2641, 0.3563, 0.2009, 0.0703, 0.1085 for states $0, 1, 2, 3, 4+$. These probabilities thus represent the stationary distribution of the number of accidents for a typical driver during one year.

5 Conclusion

In this paper a class of discrete probability distributions, called the telescopic distributions, was recalled, and the related method of discretization of continuous distributions presented. The main attention was then focussed on the analogous construction of bivariate discrete distributions. The method of model construction and its parameters estimation was described and illustrated with the aid of simulated examples. Certain limitations of such models flexibility have been discussed, too. Finally, the applicability of this approach has been demonstrated on a real data example. An interesting question concerns the interpretation of such models. The original idea is represented, for example, by the Weibull distribution example in Remark 2 of Section 2, but in most applications the challenge lies rather in finding a suitable and computationally simple quantitative model; the interpretation or practical motivation for a specific type of model is often somewhat unclear. This also concerns the real data example presented here.

References

Barbiero, A. (2022). Discrete analogues of continuous bivariate probability distributions. *Annals of Operations Research*, 312. 23-43. DOI: 10.1007/s10479-019-03388-8

Basu, A.P. and Dhar, S. (1995). Bivariate geometric distribution. *Journal of Applied Statistical Science*, 2. 33-44.

Edwards, C.B. and Gurland, J. (1961). A class of distributions applicable to accidents. *J. Amer. Statist. Assoc.*, 56. 503-517. DOI: 10.1080/01621459.1961.10480641

Farmer, E. and Chambers, E. G. (1939). A study of accident proneness among motor drivers. Industrial Health Research Board Report No. 84, London.

Johnson, N., Kotz, S., and Balakrishnan, N. (1997). Discrete Multivariate Distributions. Wiley, New York. ISBN: 978-0-471-12844-1

Kocherlakota, S. and Kocherlakota, K., 1992. Bivariate Discrete Distributions, Marcel Dekker, New York. DOI: 10.1201/9781315138480

Kopociňska, I. (1999). Bivariate negative binomial distribution of the Marshall–Olkin type. *Applicationes Mathematicae*, 25, 4. 457-461.

Marshall, A.W. and Olkin, I. (1985). A family of bivariate distributions generated by the bivariate Bernoulli distribution. *J. Amer. Statist. Assoc.*, 80. 332-338. DOI: 10.1080/01621459.1985.10478116

Ng, C.M., Ong, S.H., Srivastava, H.M. (2010). A class of bivariate negative binomial distributions with different index parameters in the marginals. *Appl. Mathematics and Computation*, 217. 3069-3087. DOI: 10.1016/j.amc.2010.08.040

Roknabadi, A.H.R., Borzadaran, G.R.M, and Khorashadizadeh, M. (2009). Some Aspects of Discrete Hazard Rate Function in Telescopic Families. *Economic Quality Control*, 24, 1. 35-42. DOI: 10.1515/EQC.2009.35

Roy, D. (2003). The Discrete Normal Distribution. *Comm. Statist. Theory Methods*, 32. 1871-1883. DOI: 10.1081/STA-120023256

Wikipedia, Telescoping series, retrieved from
$https://en.wikipedia.org/wiki/Telescoping_series$ (12.2025).