encyclopedia

MDPI

Entry

Count Random Variables

Sandra Mendonça [1,2], António Alberto Oliveira [3], Dinis Pestana [2,4,5,*] and Maria Luísa Rocha [2,6,7]

1 Departamento de Matemática—FCEE, Universidade da Madeira, Campus Universitário da Penteada, 9020-105 Funchal, Portugal; sandram@staff.uma.pt
2 Centro de Estatística e Aplicações, Universidade de Lisboa (CEAUL), Campo Grande, 1749-016 Lisboa, Portugal; maria.ls.rocha@uac.pt
3 Instituto Superior de Administração e Finanças (ISAF), Luanda, Angola; a.a.oliveira@hotmail.com
4 Faculdade de Ciências da Universidade de Lisboa (DEIO), Campo Grande, 1749-016 Lisboa, Portugal
5 Instituto de Investigação Científica Bento da Rocha Cabral, Calçada Bento da Rocha Cabral 14, 1250-012 Lisboa, Portugal
6 Faculdade de Economia e Gestão, Universidade dos Açores, Rua da Mãe de Deus, 9500-321 Ponta Delgada, Portugal
7 CEEAplA—Centro de Estudos de Economia Aplicada do Atlântico, Universidade dos Açores, Rua da Mãe de Deus, 9500-321 Ponta Delgada, Portugal
* Correspondence: ddpestana@ciencias.ulisboa.pt

Definition: The observation of randomness patterns serves as guidance for the craft of probabilistic modelling. The most used count models—Binomial, Poisson, Negative Binomial—are the discrete Morris' natural exponential families whose variance is at most quadratic on the mean, and the solutions of Katz–Panjer recurrence relation, aside from being members of the generalised power series and hypergeometric distribution families, and this accounts for their many advantageous characteristics. Some other basic count models are also described, as well as models with less obvious but useful randomness patterns in connection with maximum entropy characterisations, such as Zipf and Good models. Simple tools, such as truncation, thinning, or parameter randomisation, are straightforward ways of constructing other count models.

Keywords: discrete models; count random variables; Panjer's family; hierarchical models

check for updates

Citation: Mendonça, S.; Oliveira, A.A.; Pestana, D.; Rocha, M.L. Count Random Variables. *Encyclopedia* **2024**, *4*, 1367–1380. https://doi.org/10.3390/encyclopedia4030089

Academic Editor: David Carfì

Received: 8 August 2024
Revised: 18 September 2024
Accepted: 20 September 2024
Published: 23 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

For any $\mathcal{S} = \{x_k\}_{k\in\mathbb{K}}$, with $\mathbb{K} \subseteq \mathbb{N}_0 = \{0, 1, \dots\}$, and for any sequence $\{p_k\}_{k\in\mathbb{K}}$ such that $p_k \geq 0$ for any $k \in \mathbb{K}$ and $\sum_{x_k\in\mathcal{S}} p_k = 1$, $X = \begin{cases} x_k, & x_k \in \mathcal{S} \\ p_k = \mathbb{P}[X = x_k] \end{cases}$ is a discrete lattice random variable with support $\mathcal{S}$ and probability mass function $\{p_k\}_{k\in\mathbb{K}}$. If $x_k = k \in \mathbb{N}_0$, X is a count random variable.

In most cases, the probability mass function $\{p_k\}_{k\in\mathbb{K}}$ is not interesting, since it is difficult to deal with and there is no clear interpretation of the pattern of randomness it describes. The craft of probabilistic modelling (Gani (1986) [1]) uses a diversity of criteria to describe and select models, namely, those arising from randomness patterns (such as counts in Bernoulli trials, sampling with or without replacement, and random draws from urns). Another source of the rationale description of count models are characterisation theorems based on structural properties (e.g., a power series distribution with mean = variance, or maximum Shannon entropy with prescribed arithmetic and/or geometric mean). Recurrence relationships (for instance, $p_{k+1} = (a + \frac{b}{k+1})\, p_k$, $k = \nu, \nu+1, \dots$) or mathematical properties (for instance, the variance being at most a quadratic function of the expectation) also define interesting families of discrete random variables. On the other hand, asymptotic properties such as arithmetic properties, namely, infinite divisibility, discrete self-decomposability, and stability, serve as guidance in model choice.

Section 2 describes the discrete uniform random variables, modelling equiprobability patterns resulting from the principle of insufficient reason, of which the Bernoulli random

variable with parameter $\frac{1}{2}$ is the simplest example. Section 3 is a detailed overview of count models—Binomial, Negative Binomial—arising from the observation of random patterns in Bernoulli trials (including the Poisson random variable, as a limit of Binomial random variables under a mean stability restriction, and the Hypergeometric sampling without replacement model, herein in the context of conditioning on the sum of two independent random variables). In Section 4, the recurrence relation holding for the probability mass function of Binomial, Poisson, and Negative Binomial random variables investigated by Katz [2] and by Panjer [3] is extended to describe Hess et al. [4]'s family of basic count distributions. Section 5 briefly discusses alternative organisations of count models, namely, via Power Series distributions or Kemp's [5] generalised hypergeometric probability distributions. Section 6 constrasts the equilibrium pattern of Zipf's [6] law with the equiprobability modelled by the discrete uniform random variables. Sections 7 and 8 discuss ways of transforming random models, respectively, by randomising parameters and via the discretisation of countinuous random variables. Section 9 briefly discusses the role of characterisations in the craft of modelling count data. Further issues are briefly described in Section 10.

2. Bernoulli Random Variables, Principle of Insufficient Reason and Discrete Uniform Random Variables

Let $p \in [0,1]$ be the probability of the event $\mathcal{E}$ occurring as the outcome of an experiment. Either $\mathcal{E}$ (sometimes referred to as success) occurs once, or its complementary event $\overline{\mathcal{E}}$ (referred to as failure) occurs, so the number of occurrences of $\mathcal{E}$ in a single trial is either 1, with probability p, or 0, with probability $1-p$. We shall use the notation

$$B = \begin{cases} 0 & 1 \\ 1-p & p \end{cases},$$

where the first line indicates the support of the count random variable B, and the second line the probabilities of the support points.

The above random variable is called "Bernoulli", with parameter p, in honour of the genial probabilist Jacques Bernoulli, author of the fascinating *Ars Conjectandi* [7], published posthumously in 1713 by his nephew Niklaus Bernoulli (also a probabilist).

We shall use the notation $X \sim Bernoulli(p)$; if $p = \frac{1}{2}$, meaning that $\mathcal{E}$ and $\overline{\mathcal{E}}$ are equiprobable, it should be assumed that there is insufficient reason to assign different probabilities to $\mathcal{E}$ and $\overline{\mathcal{E}}$.

The principle of insufficient reason (renamed principle of indifference by Maynard Keynes [8]) was used in the foundation texts of Jacques Bernoulli [7] and Laplace [9] for assigning epistemic probabilities to equiprobable events. The natural extension of $X \sim Bernoulli(\frac{1}{2})$ is the equiprobable count model

$$Y = \begin{cases} k, & k = 0, \ldots, n \\ p_k = \frac{1}{n+1} \end{cases},$$

named Discrete Uniform with parameter n, that we denote as $Y \sim DiscreteUniform(n)$.

3. Count Random Variables in Bernoulli Trials

Let $\{T_1, T_2, \ldots\}$ be a sequence of independent random experiments (trials) whose outcomes are either $\mathcal{E}$, with probability $p \in (0,1)$, or $\overline{\mathcal{E}}$, with probability $1-p$. The assumption of independence means that the outcome of trial T_k, $k = 1, 2, \ldots$ has no influence whatsoever on the outcome of any other trial.

In this setting, the counts that matter are the following:

1. The number k of outcomes $\mathcal{E}$ in n trials (n fixed, k random). From the definition of Bernoulli trials, it is easy to conclude that, with X denoting such a count random variable, $\mathbb{P}[X = k] = p_k = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, \ldots, n$. This random variable,

with probability mass function $\left\{p_k = \binom{n}{k} p^k (1-p)^{n-k}\right\}_{k=0}^{n}$, is called a binomial random variable with parameters $n \in \mathbb{N}^+$ and $p \in (0,1)$, and we shall denote it as $X \sim Binomial(n,p)$ or

$$X = \begin{cases} k, & k = 0, \ldots, n \\ p_k = \binom{n}{k} p^k (1-p)^{n-k} \end{cases}.$$

The expectation is $\mathbb{E}(X) = np$ and the variance $\text{var}(X) = np(1-p)$. Hence, the dispersion index is $\mathcal{D}(X) = \frac{\text{var}(X)}{\mathbb{E}(X)} = 1 - p < 1$. For that reason, we say that the random variable $X \sim Binomial(n,p)$ is underdispersed.

2. The number n of trials needed to observe k times the outcome $\mathcal{E}$ (k fixed, n random). The simple case is $k = 1$. As in this case, with W_1^* denoting such a count random variable, $\mathbb{P}[W_1^* = n] = \mathbb{P}[\overline{\mathcal{E}}_1 \ldots \overline{\mathcal{E}}_{n-1} \mathcal{E}_n] = (1-p)^{n-1} p$. W_1^* is called geometric (or sometimes Pascal) random variable, and we shall use the notation $W_1^* \sim Geometric(p)$, or

$$W_1^* = \begin{cases} n, & n = 1, 2, \ldots \\ p_n = p(1-p)^{n-1} \end{cases}.$$

Its expectation is $\mathbb{E}(W_1^*) = \frac{1}{p}$ and its variance is $\text{var}(W_1^*) = \frac{1-p}{p^2}$.
More generally, let W_k^* be the number of trials needed for $\mathcal{E}_n$ being the k-th $\mathcal{E}$ of outcomes. Obviously, due to the independence of the Bernoulli trials,

$$\mathbb{P}(W_k^* = n) = \mathbb{P}(k-1\ \mathcal{E}'\text{s in } n-1 \text{ trials}) \times \mathbb{P}(\mathcal{E}_n) = \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \times p.$$

W_k^* is a Negative Binomial random variable with parameters k and p, that we denote $W_k^* \sim NegativeBinomial(k,p)$, or

$$W_k^* = \begin{cases} n, & n = k, k+1, \ldots \\ p_n = \binom{n-1}{k-1} p^k (1-p)^{n-k} \end{cases}.$$

Obviously, $W_k^* \stackrel{\text{d}}{=} W_{1,1}^* + \ldots + W_{1,k}^*$, with the $W_{1,i}^* \sim Geometric(p)$, $i = 1, \ldots, k$, independent. So $\mathbb{E}(W_k^*) = \frac{k}{p}$ and $\text{var}(W_k^*) = \frac{k(1-p)}{p^2}$.
It is sometimes convenient to shift the Negative Binomial random variables to start at 0, i.e., to count the number of $\overline{\mathcal{E}}$s that precede the k-th $\mathcal{E}$. In other words, to define

$$W_k = W_k^* - k \sim NegativeBinomial(k,p,0), \quad W_k = \begin{cases} n, & n = 0, 1, \ldots \\ p_n = \binom{n+k-1}{k-1} p^k (1-p)^n \end{cases}.$$

Its dispersion index is $\mathcal{D}(W_n) = \frac{\text{var}(W_n)}{\mathbb{E}(W_n)} = \frac{\frac{k(1-p)}{p^2}}{k(\frac{1}{p}-1)} = \frac{1}{p} > 1$, and for that reason the Negative Binomial random variables are considered overdispersed.
Note also that using the gamma function extension of factorials, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \text{e}^{-x} \text{d}x$, $\alpha > 0$, satisfying the recurrence relation $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, so that $\Gamma(n+1) = n!$, we may consider Negative Binomial random variables with parameter $\alpha > 0$, $p \in (0,1)$:

$$W_\alpha \sim NegativeBinomial(\alpha,p,0) = \begin{cases} n, & n = 0, 1, \ldots \\ p_n = \frac{\Gamma(n+\alpha)}{n!\Gamma(\alpha)} p^\alpha (1-p)^n \end{cases}.$$

In many situations, asymptotic results are paramount in modelling decisions or simplifications. The first central limit theorem (a name coined by Pólya [10] in 1920), establishing that if $X \sim Binomial(n,p)$ and n is large, then $\frac{X-np}{\sqrt{np(1-p)}}$ can be approximated by $Z \sim Gaussian(0,1)$, appeared in the second edition of Abraham de Moivre's [11] *The Doctrine of Chances*.

The other important asymptotic result about Binomial sequences is Poisson's [12] law of rare events:

Let $\{X_n \sim Binomial(n, p_n)\}_{n\in\mathbb{N}}$, mean-stable in the sense that $\mathbb{E}(X_n) = np_n \underset{n\to\infty}{\longrightarrow} \mu > 0$ (observe that this implies that $p_n = O(\frac{1}{n}) \underset{n\to\infty}{\longrightarrow} 0$, and this is the rationale for the name "law of rare events"). Then,

$$X_n \underset{n\to\infty}{\xrightarrow{\mathrm{d}}} Y = \begin{cases} k, & k = 0, 1, \dots \\ p_k = \mathrm{e}^{-\mu}\frac{\mu^k}{k!} \end{cases};$$

Y is a Poisson random variable with parameter $\mu > 0$, that we denote $Y \sim Poisson(\mu)$.

$\mathbb{E}(Y) = \mathrm{var}(Y) = \mu$, and hence, in what concerns dispersion, the Poisson random variable is a yardstick, in the sense that its dispersion index is $\mathcal{D}(Y) = 1$. McCabe and Skeels (2020) [13] and Di Noia et al. (2024) [14] extensively investigated testing *Poissoness* vs. overdispersion or underdispersion; cf. also Mijburgh and Visagie (2020) [15]'s overview of goodness-of-fit tests for the Poisson distribution.

The Binomial, the Negative Binomial, and the Poisson random variables are the discrete members of Morris' [16] Natural Exponential Family (NEF) with quadratic variance function in the mean value (QVF). (Recall that X is a member of NEF if its probability density function can be written as $f_X(x|\theta) = h(x)\exp[\theta x - A(\theta)]$, and therefore its cumulant generating function has the simple form $\mathcal{K}_X(t) = A(\theta + t) - A(\theta)$.) In the sequel, Morris [17] treated "*topics that can be handled within this unified NEF-QVF formulation, including unbiased estimation, Bhattacharyya and Cramér-Rao lower bounds, conditional distributions and moments, quadratic regression, conjugate prior distributions, moments of conjugate priors and posterior distributions, empirical Bayes and G_2 minimax, marginal distributions and their moments, parametric empirical Bayes, and characterisations*", and this shows the relevance of Binomial, Negative Binomial, and Poisson count models in Statistical Inference.

It is also worth mentioning that Binomial, Negative Binomial, and Poisson random variables have relevant additive properties, in the sense that

1. If $X \sim Binomial(n, p)$ and $Y \sim Binomial(m, p)$ are independent, then $X + Y \sim Binomial(n + m, p)$.
2. If $X \sim NegativeBinomial(k, p)$ and $Y \sim NegativeBinomial(j, p)$ are independent, then $X + Y \sim NegativeBinomial(k + j, p)$.
3. If $X \sim Poisson(\mu)$ and $Y \sim Poisson(\theta)$ are independent, then $X + Y \sim Poisson(\mu + \theta)$.

Furthermore, if $X \sim Poisson(\mu)$ is subject to Binomial filtering or thinning, the resulting Y has probability mass function

$$\begin{aligned}\mathbb{P}(Y = k) &= \sum_{j=0}^{\infty} \mathbb{P}(Y = k|X = k + j)\,\mathbb{P}(X = k + j) \\ &= \sum_{j=0}^{\infty} \binom{k+j}{k} p^k (1-p)^j \mathrm{e}^{-\mu}\frac{\mu^{k+j}}{(k+j)!} = \mathrm{e}^{-p\mu}\frac{(p\mu)^k}{k!},\ k = 0, 1, \dots\end{aligned}$$

i.e., $Y \sim Poisson(p\mu)$.

Many natural phenomena are subject to filtering; for instance, using a Poisson model for the number of eggs laid by turtles, the number of hatched eggs, the number of surviving newborns until they reach the ocean, the number of turtles surviving the first year, and so forth, can be modelled by a chain of filtered Poisson random variables, a very useful tool in population dynamics modelling.

On the other hand, the Negative Binomial is a "random Poisson", obtained when its parameter is a random variable with a Gamma distribution: Let $X|\Lambda \sim Poisson(\Lambda)$ with $\Lambda \sim Gamma(\alpha, \delta)$. Then,

$$\begin{aligned}\mathbb{P}[X = k] &= \int_0^{\infty} \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}\frac{\lambda^{\alpha-1}\mathrm{e}^{-\lambda/\delta}}{\delta^\alpha\Gamma(\alpha)}\mathrm{d}\lambda = \frac{1}{k!\delta^\alpha\Gamma(\alpha)}\int_0^{\infty}\left(\frac{\delta y}{1+\delta}\right)^{\alpha+k-1}\frac{\delta\,\mathrm{e}^{-y}}{1+\delta}\mathrm{d}y \\ &= \frac{\Gamma(\alpha+k)}{k!\Gamma(\alpha)}\left(\frac{1}{1+\delta}\right)^{\alpha}\left(\frac{\delta}{1+\delta}\right)^{k} = \binom{\alpha+k-1}{k}\left(\frac{1}{1+\delta}\right)^{\alpha}\left(\frac{\delta}{1+\delta}\right)^{k},\end{aligned}$$

$k = 0, 1, \ldots$, so we obtain $X \sim NegativeBinomial\left(\alpha, \frac{1}{1+\delta}\right)$. For this reason, and as the Negative Binomial random variable is overdispersed, the Poisson being the yardstick, some authors in population dynamics modelling consider the Negative Binomial a more dispersed Poisson.

The above results on the sums of Morris' discrete variable have interesting consequences when conditioning is applied:

If $X \sim Poisson(\mu)$ and $Y \sim Poisson(\theta)$ are independent,

$$\begin{aligned}\mathbb{P}(X = k|X + Y = n) &= \frac{\mathbb{P}(X = k)\,\mathbb{P}(Y = n - k)}{\mathbb{P}(X + Y = n)} \\ &= \frac{\mathrm{e}^{-\mu}\frac{\mu^k}{k!}\,\mathrm{e}^{-\theta}\frac{\theta^{n-k}}{(n-k)!}}{\mathrm{e}^{-(\mu+\theta)}\frac{(\mu+\theta)^n}{n!}} = \binom{n}{k}\left(\frac{\mu}{\mu+\theta}\right)^k\left(\frac{\theta}{\mu+\theta}\right)^{n-k},\ k = 0, 1, \ldots,\end{aligned}$$

i.e., $X|X + Y = n \sim Binomial\left(n, \frac{\mu}{\mu+\theta}\right)$.

Reasoning in the same way, if $X \sim Binomial(n, p)$ and $Y \sim Binomial(m, p)$ are independent,

$$\begin{aligned}\mathbb{P}(X = k|X + Y = \ell) &= \frac{\mathbb{P}(X = k)\,\mathbb{P}(Y = \ell - k)}{\mathbb{P}(X + Y = \ell)} \\ &= \frac{\binom{n}{k}p^k(1-p)^{n-k}\,\binom{m}{\ell-k}p^{\ell-k}(1-p)^{m-\ell+k}}{\binom{n+m}{\ell}p^\ell(1-p)^{n+m-\ell}} \\ &= \frac{\binom{n}{k}\,\binom{m}{\ell-k}}{\binom{n+m}{\ell}},\ k = \max\{0, \ell - k\}, \ldots, \min\{\ell, n\}.\end{aligned}$$

This is the probability mass function of $X|X + Y = \ell \sim Hypergeometric(n + m, \ell, p)$ with $p = \frac{m}{m+n}$.

If $X \sim Hypergeometric(N, n, p)$, $\mathbb{E}(X) = np$ and $\mathrm{var}(X) = np(1-p)\frac{N-n}{N-1}$.

It is interesting to observe that if $X_P \sim Poisson(\mu)$, $Y_B \sim Binomial(n, p)$, and $T_H \sim Hypergeometric(N, n, p)$ have the same expectation np, then,

$$\mathrm{var}(X_P) = \mu = np > \mathrm{var}(Y_B) = np(1-p) > \mathrm{var}(T_H) = np(1-p)\frac{N-n}{N-1}.$$

In other words, for these three count models, the number of parameters implies a tradeoff between decreasing simplicity and increasing information. Augmenting the number of parameters supplies more information. However, one should always bear in mind that all models are wrong, but some are useful (as Box [18] judiciously stated), and that the parsimony principle hints that the simplest model that is useful enough should be used. Observe that in model choice, criteria such as the Akaike Information Criterion (AIC) [19] penalise the number of parameters of the model, and this often leads us to choose a model with fewer parameters, although it provides a better fit.

The Hypergeometric random variable is used in simple random sampling without replacement, while the Binomial is appropriate for simple random sampling with replacement. In the Hypergeometric context, we are dealing with exchangeability; in the Binomial setting, with a stronger independence concept.

The Hypergeometric random model is also the basis for estimating the size of a population with the technique of capture–recapture, cf. Seber and Schofield [20]. Suppose that in a controlled investigation, k animals are captured, marked, and released, and that after a while, there is a second instance of capturing, in which j of the n captured are marked (i.e., recaptured). If the unknown size of the population is N, k of which have been marked and $N - k$ are unmarked, j is the observed value of $Y \sim Hypergeometric(N, n, p)$, with $p = \frac{k}{N}$. The Peterson estimator $\widetilde{N} = \frac{nk}{j}$ of the size of the population is a method of

moments estimator (observe that we described a technique with just one recapture of j marked animals, and in a sample (y) of size one, this is the mean).

4. Katz Count Models and Extensions

The probability mass function of the discrete Morris NEF-QVF Binomial, Poisson, and Negative Binomial random variables share the interesting property

$$p_{k+1} = \left(a + \frac{b}{k+1}\right) p_k, \quad k = 0, 1, \ldots \tag{1}$$

Let $N_{a,b}$ denote the non-degenerate count random variables whose probability mass function satisfies the recursive relation (1).

Multiplying both sides of (1) by s^{k+1} and adding for $k \geq 0$, we find that the corresponding probability generating functions $\mathcal{G}_{a,b}(s) = \sum_{k=0}^{\infty} p_k s^k$ are the solutions of the differential equation $\dfrac{\mathcal{G}'_{a,b}(s)}{\mathcal{G}_{a,b}(s)} = \dfrac{a+b}{1-as}$; for details, cf. Pestana and Velosa (2004) [21]; namely:

1. $\mathcal{G}_{a,b}(s) = \left(1 - \frac{a}{a-1} + \frac{a}{a-1}s\right)^{-\frac{a+b}{a}}$, $a < 0$, $\frac{a}{a-1} = p \in (0,1)$, $-\frac{a+b}{a} = n \in \mathbb{N}$, i.e., $N_{a,b} \sim Binomial\left(-\frac{a+b}{a}, \frac{a}{a-1}\right)$.
2. $\mathcal{G}_{a,b}(s) = \mathrm{e}^{b(s-1)}$, $a = 0$, $b > 0$, i.e., $N_{0,b} \sim Poisson(b)$.
3. $\mathcal{G}_{a,b}(s) = \left(\frac{1-a}{1-as}\right)^{\frac{a+b}{a}}$, $a \in (0,1)$, $a + b > 0$, i.e., $N_{a,b} \sim NegativeBinomial\left(\frac{a+b}{a}, a\right)$.

The simple expression (1) for the successive probability atoms has been several times rediscovered in different contexts (McCabe and Skeels, 2020 [13]). Katz (1945, 1965) [2,22] used it to organise a family of discrete models in the same spirit of the continuous Pearson family, and for that reason, the family of random variables $\{N_{a,b}\}$ whose probability mass function satisfies (1) is referred to as the Katz family. In Risk Theory and Insurance, $\{N_{a,b}\}$ is known as Panjer's class, due to Panjer's (1981) [3] important breakthrough that the recurrence relation (1) implies that the distribution of aggregate claims can be iteratively computed or approximated:

Let $S_1, S_2, \ldots$ denote the identically distributed discrete claim sizes, with $\pi_k = \mathbb{P}[S_i = k]$, $k \in \mathbb{N}_0$, and assume that the S_i are mutually independent and independent of the claim number $N_{a,b}$. Then,

$$\mathbb{E}\left(S_1 \middle| \sum_{i=1}^{n} S_i = j\right) = \frac{j}{n}, \quad \text{and} \quad \mathbb{P}\left(S_1 = k \middle| \sum_{i=1}^{n} S_i = j\right) = \frac{\pi_k \, \pi^{*(n-1)}_{j-k}}{\pi^{*n}_j},$$

where $\{\pi^{*n}_k\}$ denotes the n-th convolution of $\{\pi_k\}$. Consequently, the probabilities $p_{X,k} = \mathbb{P}[X = j]$ of the aggregate claim amount $X = \sum_{i=1}^{N_{a,b}} S_i$ (understood as $X = 0$ if $N_{a,b} = 0$) can be obtained recursively using *Panjer's algorithm*:

$$\mathbb{P}[X = j] = \begin{cases} \mathcal{G}_{a,b}(\pi_0), & j = 0 \\ \dfrac{\sum_{k=1}^{j} \left(a + \frac{bk}{j}\right) \pi_k \, p_{X,j-k}}{1 - a\pi_0}, & j \in \mathbb{N} \end{cases}.$$

This requires only $\mathrm{O}(j)$ computations, while the traditional convolution method would require $\mathrm{O}(j^2)$. More generally, the cumulative distribution function $F_X = \sum_{k=0}^{\infty} p_k F_S^{*k}$ of the aggregate claim $X = \sum_{k=1}^{N_{a,b}} S_k$ for an arbitrary distribution function F_S of the individual

claim sizes, when the number of claims is $N_{a,b}$ with probability mass $\{p_k\}$, satisfies the integral equation

$$F_X(x) = p_0 + aF_S * F_X(x) + b\int_0^x v\int_0^{x-v} \frac{\mathrm{d}F_X(y)}{v+y}\mathrm{d}F_S(v)$$

for $x > 0$, if $F_S(0) = 0$. For nonnegative claim sizes with $F_S(0) > 0$,

$$F_X(0) = \begin{cases} \mathrm{e}^{b[F_S(0)-1]}, & a = 0 \\ \left[\dfrac{1-a}{1-aF_S(0)}\right]^{\frac{a+b}{a}}, & a \neq 0 \end{cases}.$$

For detailed proofs, cf. Rólski et al. [23] (pp. 118–124). Klugman et al. [24] (pp. 221–224) use an example to discuss in depth the use of the empiric dispersion index as guidance for model choice in the Panjer family.

If we relax the condition on the support, $p_{k+1} = (a + \frac{b}{k+1})\, p_k$, $k = 1, 2, \ldots$, as investigated by Jewell [25], Willmot [26], or Sundt [27], we also obtain the Logarithmic random variable

$$X_L = \begin{cases} k, & k = 1, 2, \ldots \\ p_k = \dfrac{\theta^k/k}{-\ln(1-\theta)} \end{cases}, \ \theta \in (0,1),$$

used by Fisher [28] to model species abundance, and the Engen Extended Negative Binomial (ENB) [29] random variable,

$$X_{ENB} = \begin{cases} k, & k = 1, 2, \ldots \\ p_k = \dfrac{1}{1-(1-\theta)^\beta}\binom{\beta+k-1}{\beta-1}\theta^k(1-\theta)^\beta \end{cases}, \ \theta \in (0,1), \beta \in (-1,0).$$

The dispersion indices of the Logarithmic and of the Engen random variables depend on θ. For $X_L \sim Logarithmic(\theta)$,

$$\mathcal{D}(X_L) = \frac{-\dfrac{\theta^2 + \theta\ln(1-\theta)}{(1-p)^2[\ln(1-\theta)]^2}}{-\dfrac{\theta}{(1-\theta)\ln(1-\theta)}} = \frac{\theta + \ln(1-\theta)}{(1-\theta)\ln(1-\theta)} \Longrightarrow \mathcal{D}(X_L)\begin{cases} < 1 & \text{if } \theta \in \left(0, 1-\dfrac{1}{\mathrm{e}}\right) \\ = 1 & \text{if } \theta = 1-\dfrac{1}{\mathrm{e}} \\ > 1 & \text{if } \theta \in \left(1-\dfrac{1}{\mathrm{e}}, 1\right) \end{cases},$$

and for $X_{ENB} \sim Engen(\beta, \theta)$

$$\mathcal{D}(X_{ENB}) = \frac{1 + \beta\theta - \frac{\beta\theta}{1-(1-\theta)^\beta}}{(1-\theta)} \Longrightarrow \mathcal{D}(X_E)\begin{cases} < 1 & \text{if } \theta \in \left(0, 1-\left(\frac{1}{\beta+1}\right)^{1/\beta}\right) \\ = 1 & \text{if } \theta = 1-\left(\frac{1}{\beta+1}\right)^{1/\beta} \\ > 1 & \text{if } \theta \in \left(1-\left(\frac{1}{\beta+1}\right)^{1/\beta}, 1\right) \end{cases}.$$

Hess et al. (2002) [4] investigated ν-Panjer classes, $\nu \in \mathbb{N}_0$, whose probability mass function satisfies $p_{k+1} = \left(a + \frac{b}{k+1}\right) p_k$, $k = \nu, \nu+1, \ldots$. Excluding the 0-Panjer family, which does not contain the logarithmic and ENB distributions, Hess et al. (2002) [4] established that any ν-Panjer distribution is the left endpoint truncation of an $(\nu-1)$-Panjer distribution, for $\nu \geq 2$. For this reason, they call the Binomial, Poisson, Negative Binomial, Logarithmic, and ENB distributions basic count models.

Klugman et al.'s [24] notation $N(a, b, n)$ has been used by Fackler (2024) [30] to supply a unified, practical, and intuitive representation of the Panjer distributions and their parameter space, and give an inventory of parameterisations used for Panjer distributions.

Panjer's recursion has been extended by Tzaninis and Bozikas (2024) [31] to mixed compound distributions, following the use of finite mixtures for modelling dispersion in count data by Ong et al. (2023) [32].

5. Other Families of Count Models

The Katz [2,22] organisation of discrete models using the difference equation (1), similar to Pearson's continuous model types description using a differential equation, inspired later essays on the rationale description of discrete models.

5.1. Power Series Distributions

The discrete random variable X_θ, $\theta > 0$ has a power series distribution (PSD) if

$$X_\theta = \begin{cases} k, & k \in \mathbb{N}_0 \\ p_k = \frac{a_k \theta^k}{A(\theta)} \end{cases}, \; a_k \geq 0, \; A(\theta) = \sum_{k \in \mathbb{N}_0} a_k \theta^k.$$

A power series distribution is a discrete linear exponential distribution.

If we replace $\mathbb{N}_0$ by $\mathcal{S}$, a nonempty enumerable subset of $\mathbb{N}_0$, we obtain a generalised power series distribution (GPSD):

$$X_\theta = \begin{cases} k, & k \in S \\ p_k = \dfrac{a_k \theta^k}{A(\theta)} \end{cases}, \; a_k \geq 0, \; A(\theta) = \sum_{k \in S} a_k \theta^k.$$

Observe that the probability generating function of X_θ is $\mathcal{G}_{X_\theta}(s) = \frac{A(\theta s)}{A(\theta)}$.

The random variables $X_B \sim Binomial(n, \theta)$, with $A(\theta) = (1+\theta)^n$, $Y_P \sim Poisson(\theta)$, with $A(\theta) = \mathrm{e}^\theta$, $W_{NB} \sim NegativeBinomial(k, \theta)$, with $A(\theta) = \frac{1}{(1-\theta)^k}$, $k > 0$, and $X_L \sim Logarithmic(\theta)$, $\theta \in (0, 1)$, with $A(\theta) = -\ln(1-\theta)$, have GPSD.

Obviously, if a generalised power series distribution is truncated, another generalised power series distribution is obtained. An important issue, seldom discussed in the literature: restricting the support may enlarge the parameter space Θ; for example truncating to the right of $N \in \mathbb{N}$ the support of $X_L \sim Logarithmic(\theta)$, $\theta \in (0, 1)$, the resulting truncated logarithmic random variable X_{TL}

$$X_{TL} = \begin{cases} k, \; k = 1, \ldots, N \\ p_k = \dfrac{\theta^k / k}{\sum_{k=1}^N \theta^k / k} \end{cases}, \; \theta > 0,$$

has a parameter space $\Theta = (0, \infty)$, since $\sum_{k=1}^N \frac{\theta^k}{k} < \infty$ for any $\theta > 0$.

5.2. Kemp's Generalised Hypergeometric Probability Distributions

The work by Johnson et al. (2005) [33] is a coordinated thorough description of useful discrete random variables, structured on Kemp's (1968) [5] use of hypergeometric functions, and the corresponding characterisation sections enhance their randomness patterns. Kemp's [5] generalised hypergeometric probability distributions, with probability generating function $\frac{{}_pF_q(\lambda z)}{{}_pF_q(\lambda)}$, where

$${}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda\, z] = \sum_{j=0}^{\infty} \frac{(a_1)_j \cdots (a_p)_j}{(b_1)_j \cdots (b_q)_j} \frac{(\lambda\, z)^j}{j!}$$

$((\alpha)_j = \alpha(\alpha+1)\cdots(\alpha+j-1)$ is the Pochhammer's symbol) is the generalised hypergeometric function with numeratorial parameters $a_1, \ldots a_p$ and denominatorial parameters $b_1, \ldots, b_q$. This huge family—Johnson et al. (2005) [33] (pp. 93–94) identify the name of 37 of the more than 70 studied in Patil et al. (1984) [34]—has many useful properties.

6. Uniformity and Power Law Randomness, a Striking Example of Opposite Patterns

The simplest equiprobability pattern of count randomness is modelled by the discrete uniform distribution $X \sim DiscreteUniform(n)$, $X = \begin{cases} k, \ k = 0, 1, \ldots n \\ p_k = \frac{1}{n+1} \end{cases}$.

Equiprobability is, however, the exception; natural phenomena are more prone to present other patterns of equilibrium, such as with Zipf's model, which is the special case $q = 0$ of the Zipf–Mandelbrot family of discrete models given by

$$X_{N,q,\alpha} = \begin{cases} k, \quad k \in \{1, 2, \ldots, N\} \\ p_k = \dfrac{1}{H_{N,q,\alpha}} \dfrac{1}{(k+q)^\alpha} \end{cases},$$

where $\alpha > 0$, $q \geq 0$, k is the rank of the data and $H_{N,q,\alpha} = \sum_{j=1}^{N} \frac{1}{(j+q)^\alpha}$ is a generalisation of the harmonic number $H_N = \sum_{j=1}^{N} \frac{1}{j}$. The cumulative distribution function is, for $x \in \{1, \ldots, N\}$, $F(x) = \dfrac{H_{\lfloor x \rfloor, q, \alpha}}{H_{N,q,\alpha}}$, where $\lfloor x \rfloor$ is the largest integer not greater than x.

Zipf's law has been introduced to model the equilibrium—in the sense that rank $\times$ frequency is approximately constant—in verbal communication, which employs mainly words of ordinary usage (social trend) and scarcely words tied to the user's vocabulary and thematic preferences (individual trend). In what concerns extra-pair paternity (EPP), there is some evidence that the observed variability of the number of extra-pair offspring (EPO) in the brood, in various species of passerines, results from a delicate equilibrium of needing a social partner to care for raising the brood and the tendency to have offspring sired by a stronger male. This is the reason why the Zipf–Mandelbrot family seems to be a plausible model, cf. Marques et al. (2005) [35].

Notice that Zipf's law is also a particular case of the extended parameter space of the right-truncated logarithmic distribution where $p_k = \frac{\theta^k/k}{\ln(1-\theta)}$, for $k = 1, 2, \ldots$ and $\theta \in (0, 1)$. In fact, when considering the truncated logarithmic model where $p_k = \frac{\theta^k/k}{\sum_{j=1}^{\nu} \theta^j/j}$ for $k = 1, 2, \ldots, \nu$, the parameter space can be extended to $\theta \in (0, \infty)$, as already observed. In particular, if $\theta = 1$, we obtain Zipf's law. For more information, cf., e.g., Johnson et al. (2005) [33] (Chapter 11) and references therein. It is a power law in the sense that $\mathbb{P}[X = k] \propto k^{-\alpha}$, for some positive α, and therefore its log-log plot exhibits a linear signature. When an almost linear signature is observed, it is considered a strong support for choosing a power law model, namely, in the discrete context, a Zipf–Mandelbrot law.

7. Hierarchical Models and Parameter Randomisation

A powerful tool to obtain models is to randomise parameters. For instance, let $X|P \sim Binomial(n, P)$ and $P \sim Uniform(0, 1)$. Since

$$\binom{n}{k} \int_0^1 p^k (1-p)^{n-k} \, \mathrm{d}p = \binom{n}{k} B(k+1, n+1-k) = \frac{1}{n+1},$$

we have $\mathbb{P}[X = k] = \frac{1}{n+1}$, $k = 0, 1, \ldots, n$, i.e., $X \sim DiscreteUniform(n)$. Observe that $P \sim Uniform(0, 1)$ can be considered as "null" information about the value of p, thus resulting in discrete equiprobability.

More generally, if $X|P \sim Binomial(n,P)$ with $P \sim Beta(\mu,\nu)$, we obtain Beta-Binomial distributions, with probability mass function

$$p_k = \mathbb{P}[X=k] = \binom{n}{k} \int_0^1 p^k (1-p)^{n-k} \, \frac{p^{\mu-1}(1-p)^{\nu-1}}{B(\mu,\nu)} \, \mathrm{d}p$$
$$= \binom{n}{k} \frac{\Gamma(k+\mu)\Gamma(n-k+\nu)\Gamma(\mu+\nu)}{\Gamma(n+\mu+\nu)\Gamma(\mu)\Gamma(\nu)}, \quad k=0,1,\ldots,n.$$

Let $X|P \sim Geometric(P)$ with $P \sim Uniform(0,1)$. As $\int_0^1 p\,(1-p)^{k-1}\,\mathrm{d}p = B(2,k)$, for $k=1,2,\ldots$ we obtain $\mathbb{P}[X=k] = \dfrac{1}{k(k+1)}$.

More generally, if $X|P \sim Geometric(P)$ with $P \sim Beta(\mu,\nu)$,

$$p_k = \int_0^1 p\,(1-p)^{k-1} \, \frac{p^{\mu-1}(1-p)^{\nu-1}}{B(\mu,\nu)} \, \mathrm{d}p = \frac{\mu\,\Gamma(\mu+\nu)\,\Gamma(\nu+k-1)}{\Gamma(\nu)\,\Gamma(\mu+\nu+k)}.$$

For example, if $\mu=2$ and $\nu=1$, for $k=1,2,\ldots$ we obtain

$$p_k = \frac{2\,\Gamma(k)\,\Gamma(3)}{\Gamma(k+3)} = \frac{4}{k\,(k+1)\,(k+2)}.$$

Let $X|P \sim Geometric(P)$ with $P \sim BetaBoop(1,1,1,2)$; cf. Brilhante and Pestana [36]. Using formula 1, $\int_0^1 x^{p-1}\,(1-x^r)^{q-1}(-\ln x)\,\mathrm{d}x = \frac{1}{r^2}\,B\left(\frac{p}{r},q\right)\left[\psi\left(\frac{p}{r}+q\right)-\psi\left(\frac{p}{r}\right)\right]$, from subsubsection 4.253 in Gradshteyn and Ryzhik (2007) [37] (p. 538), where $\psi(z) = \frac{\mathrm{d}}{\mathrm{d}z}\ln\Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ is the logarithmic derivative of the gamma function, we conclude that [37],

$$p_k = \int_0^1 p\,(1-p)^{k-1}(-\ln p)\,\mathrm{d}p = B(2,k)\,[\psi(2+k)-\psi(2)].$$

Observing that taking the derivative of $\Gamma(z+1) = z\,\Gamma(z)$, $\Gamma'(z+1) = z\,\Gamma'(z)+\Gamma(z)$, and dividing by $\Gamma(z+1)$, we obtain $\frac{\Gamma'(z+1)}{\Gamma(z+1)} = \frac{z\,\Gamma'(z)}{z\,\Gamma(z)} + \frac{\Gamma(z)}{z\,\Gamma(z)}$, and consequently $\psi(z+1) = \psi(z)+\frac{1}{z}$, we deduce that

$$p_k = \frac{1}{k\,(k+1)}\left[\psi(k)+\frac{1}{k}+\frac{1}{k+1}-\psi(2)\right].$$

Continuing to apply the recursive expression to ψ, we obtain

$$X = \begin{cases} k,\ k=1,2,\ldots \\ p_k = \dfrac{H_{k+1}-1}{k(k+1)} \end{cases},$$

where $H_n = \sum_{j=1}^{n} \frac{1}{j}$ is the n-th harmonic number.

8. Discrete Analogues of Continuous Random Variables

Discrete analogues of continuous random variables may inherit interesting patterns. For instance, if we define

$$W = \begin{cases} k, \quad k=0,1,\ldots \\ \mathbb{P}[W=k] = F_X(k+1) - F_X(k) \end{cases},$$

with $X \sim Exponential(\lambda)$, then $W \sim Geometric(1-\mathrm{e}^{-\lambda})$ inherits the memoryless property of the Exponential in the sense that $\mathbb{P}[W \geq k+\ell | W \geq \ell] = \mathbb{P}[W \geq k]$.

In this context, it is worth citing the discrete lognormal random variable:

$$X = \begin{cases} k, \quad k = 1, 2, \ldots \\ p_k = \dfrac{1}{k\, C(\mu, \sigma)} \exp\left[-\dfrac{1}{2} \left(\dfrac{\ln k - \mu}{\sigma} \right)^2 \right] \end{cases}$$

is a model for discrete skew populational data, that is not a power law but imitates a linear signature. For an overview of discrete analogues of continuous probability distributions, cf. Chakraborty (2015) [38].

9. Characterisations of Count Models

Families of discrete distributions have striking structural properties, with relevant consequences for model choice. Characterisation theorems are also helpful in the craft of probabilistic modelling.

As already observed, the dispersion index $\mathcal{D}(X) = \frac{\mathrm{Var}(X)}{\mathbb{E}(X)}$ of the Poisson distribution is 1. In fact, a PSD random variable X has $\mathcal{D}(X) = 1$ if and only if $X \sim Poisson(\mu)$, for some $\mu > 0$, a characterisation due to Kosambi (1949) [39].

Characterisations are a useful tool in model choice. For instance, Kapur (1989) [40] has shown that a discrete random variable X with fixed arithmetic mean has maximum Shannon entropy $-\sum_{k \in S} p_k \ln p_k$ if and only if X is a Geometric random variable. Observe that choosing the largest entropy model within a class of distributions amounts to selecting as default the least informative model, and so minimising the prior information.

A discrete random variable X with fixed geometric mean has maximum Shannon entropy if and only if it is a Zipf or Discrete Pareto random variable

$$X_Z = \begin{cases} k, \; k = 1, 2, \ldots \\ p_k = \dfrac{1}{\zeta(\eta)\, k^{\eta}} \end{cases},$$

where $\zeta(\eta)$ is the Riemann zeta function. A discrete random variable X with fixed arithmetic and geometric means has maximum Shannon entropy if and only if X is a Good random variable

$$X_{Go} = \begin{cases} k, \; k = 1, 2, \ldots \\ p_k = \dfrac{q^k}{k^{\eta}\, \phi(q, \eta, 1)} \end{cases},$$

where $\phi(q, \eta, \nu) = \sum_{k=0}^{\infty} \frac{q^k}{(\nu + k)^{\eta}}$, $\nu \neq 0, -1, -2, \ldots$, is the Lerch function (subsubsection 9.55 in Gradshteyn and Ryzhik (2007) [37]). Observe that the Geometric random variable with support $1, 2, \ldots$ is the special case when $\eta = 0$, and the Logarithmic random variable is the case $\eta = 1$.

10. Concluding Remarks

As in the continuous case, asymptotic results also play an important role in modelling; for instance, the Poisson limit in discrete settings is, to a certain extent, comparable to the central limit Gaussian law in general; cf. Steutel and van Harn (1979) [41]. In the same spirit, extremal discrete laws do arise naturally in order statistics contexts; cf. Hall (1996) [42]. In addition, useful models can originate other patterns of randomness via truncation (with the eventual broadening of the parameter space), randomly stopped sums, or randomly stopped extremes, and the randomisation of parameters.

Moreover among all continuous probability distributions with support $[0, \infty)$ and mean μ, the exponential distribution with $\lambda = 1/\mu$, modelling interarrival times in the Poisson process, has the largest differential entropy

$$h(f) = -\int_{-\infty}^{\infty} f_X(x) \ln f_X(x) \mathrm{d}x,$$

meaning that the Poisson process has the greatest λt-dimensional entropy in the interval $(0, t)$ among all the homogeneous point processes with the same given intensity λ. For that reason, Poisson streams of events are generally interpreted as representing the pattern of unconstrained randomness. This is a rationale for considering Poisson modelling, since entropy is non-decreasing, and therefore the Poisson random variable is a weak limit in very general frameworks, and in particular of Binomial and of Negative Binomial sequences with a mean stability condition. In a sense, the Poisson distribution models the unconstrained scattering of the type of throwing rice grains when seeding.

In Population Dynamics, dispersion is a clue to model choice. Zhang et al. (2018) [43] investigated dispersion patterns of finite mixtures, and Cahoy et al. (2021) [44] introduced flexible fractional Poisson distributions to model underdispersed and overdisersed count data. Recent advances in what concerns underdispersed count models were made by Huang (2022) [45] and Seck et al. (2022) [46]. Rana et al. (2023) [47] investigated the influence of outliers on overdispersion, and Sengupta and Roy (2023) [48] the role of under-reported count data and zero-inflated models, an issue also investigated by Aswi et al. (2022) [49].

From the additive properties of Binomial random variables shown in Section 3, $X \sim Binomial(n, p)$ is n-divisible, in the sense that it can be decomposed as the sum of n independent and identically distributed $B_j \sim Bernoulli(p)$, $j = 1, \ldots, n$, random variables; and $X \sim NegativeBinomial(k, p)$ is k-divisible, in the sense that it can be decomposed as the sum of k independent and identically distributed $W_j \sim Geometric(p), j = 1, \ldots, k$, random variables. On the other hand $X \sim Poisson(\mu)$ is infinitely divisible, since for any $n = 2, 3, \ldots$, it can be decomposed as the sum of n independent and identically distributed $W_k \sim Poisson(\mu/n)$ random variables.

Moreover, from Raikov's theorem [50], if $X \sim Poisson(\mu)$ is decomposed as the sum of two independent random variables, $X = Y + Z$, then Y and Z are also Poisson random variables. This means that the Poisson random variables are extreme points of the convex set of infinitely divisible random variables. A similar result with the Gaussian random variables, conjectured by Lévy [51] and proved by Cramér [52], shows that they are also extreme points of the set of infinitely divisible random variables. From this, Johansen (1966) [53]) supplied a structural proof of the integral representations of infinitely divisible characteristics functions. In fact, the Poisson random variables are the building blocks of infinitely divisible random variables (de Finetti (1929, 1932) [54,55], which is a strong modelling asset, since many observed phenomena are the result of several contributing effects.

Author Contributions: Conceptualization, S.M., A.A.O., D.P. and M.L.R.; methodology, S.M. and D.P.; investigation, S.M., A.A.O., D.P. and M.L.R.; writing—original draft preparation, S.M. and D.P.; writing—review and editing, S.M., A.A.O., D.P. and M.L.R.; supervision, S.M. and D.P.; project administration, S.M. and D.P.; funding acquisition, S.M., D.P. and M.L.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly financed by national funds through FCT—Fundação para a Ciência e a Tecnologia, Portugal, under the project UIDB/00006/2020 (https://doi.org/10.54499/UIDB/00006/2020, accessed on 1 August 2024) and the research grant UIDB/00685/2020 of the CEEAplA—Centre of Applied Economics Studies of the Atlantic, School of Business and Economics of the University of the Azores.

Acknowledgments: The authors thank the anonymous referees for inspiring comments and suggestions that contributed to improve this entry.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the writing of the manuscript; or in the decision to publish the results.

References

1. Gani, J. *The Craft of Probabilistic Modelling: A Collection of Personal Accounts*; Springer: New York, NY, USA, 1986. [CrossRef]
2. Katz, L. Characteristics of Frequency Functions Defined by First Order Difference Equations. Ph.D. Thesis, University of Michigan, Ann Arbor, MI, USA, 1945.

3. Panjer, H.H. Recursive evaluation of a family of compound distributions. *ASTIN Bull.* **1981**, *12*, 22–26. [CrossRef]
4. Hess, K.T.; Liewald, A.; Schmidt, K.D. An extension of Panjer's recursion. *ASTIN Bull.* **2002**, *32*, 283–297. [CrossRef]
5. Kemp, A.W. A wide class of discrete distributions and the associated differential equations. *Sankhyā Indian J. Stat. Ser. A* **1968**, *30*, 401–410.
6. Zipf, G.K. The Unity of Nature, Least-Action, and Natural Social Science. *Sociometry* **1942**, *5*, 48–62. [CrossRef]
7. Bernoulli, J. *The Art of Conjecturing, Together with Letter to a Friend on Sets in Court Tennis*; John Hopkins University Press: Baltimore, MD, USA, 2005.
8. Keynes, J.M. *A Treatise on Probability*; Macmillan: London, UK, 1921.
9. Laplace, P.S. *Essai Philosophique sur les Probabilités*; Courcier: Paris, France, 1814.
10. Pólya, G. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.* **1920**, *8*, 171–181. [CrossRef]
11. De Moivre, A. *The Doctrine of Chances*; Chelsea: New York, NY, USA, 2000.
12. Poisson, S.D. *Probabilité des Jugements en Matière Criminelle et en Matière Civile, Précédés des Règles Générales du Calcul des Probabilités*; Bachelier: Paris, France, 1837.
13. McCabe, B.P.; Skeels, C.L. Distributions you can count on . . . But what's the point? *Econometrics* **2020**, *8*, 9. [CrossRef]
14. Di Noia, A.; Marcheselli, M.; Pisani, C.; Pratelli, L. A family of consistent normally distributed tests for Poissonity. *AStA Adv. Stat. Anal.* **2024**, *108*, 209–223. [CrossRef]
15. Mijburgh, P.A.; Visagie, I.J.H. An overview of goodness-of-fit tests for the Poisson distribution. *S. Afr. Stat. J.* **2020**, *54*, 207–230. [CrossRef]
16. Morris, C.N. Natural exponential families with quadratic variance function. *Ann. Stat.* **1982**, *10*, 65–80. [CrossRef]
17. Morris, C.N. Natural Exponential Families with Quadratic Variance Functions: Statistical Theory. *Ann. Stat.* **1983**, *11*, 515–529. [CrossRef]
18. Box, G.E.P. Science and Statistics. *J. Am. Stat. Assoc.* **1976**, *71*, 791–799. [CrossRef]
19. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]
20. Seber, G.; Schofield, M. *Capture-Recapture: Parameter Estimation for Open Animal Populations*; Springer: Cham, Switzerland, 2019.
21. Pestana, D.; Velosa, S. Extensions of Katz-Panjer families of discrete distributions. *REVSTAT—Stat. J.* **2004**, *2*, 145–162. [CrossRef]
22. Katz, L. Unified treatment of a broad class of discrete probability distributions. In *Classical and Contagious Discrete Distributions*; Patil, G.P., Ed.; Pergamon: Oxford, UK, 1965; pp. 175–182.
23. Rólski, T.; Schmidli, H.; Schmidt, V.; Teugels, J. *Stochastic Processes for Insurance and Finance*; Wiley: New York, NY, USA, 1999.
24. Klugman, S.A.; Panjer, H.H.; Willmot, G.F. *Loss Models. From Data to Decisions*; Wiley: New York, NY, USA, 1998.
25. Sundt, B.; Jewell, W. Further results on recursive evaluation of compound distributions. *ASTIN Bull.* **1981**, *12*, 27–39. [CrossRef]
26. Willmot, G.E. Sundt and Jewell's family of discrete distributions. *ASTIN Bull.* **1988**, *18*, 17–29. [CrossRef]
27. Sundt, B. On some extensions of Panjer's class of counting distributions. *ASTIN Bull.* **1992**, *22*, 61–80. [CrossRef]
28. Fisher, R.A. A theoretical distribution for the apparent abundance of different species. *J. Anim. Ecol.* **1943**, *12*, 54–57.
29. Engen, S. On species frequency models. *Biometrika* **1974**, *61*, 263–270. [CrossRef]
30. Fackler, M. Panjer class revisited: One formula for the distributions of the Panjer (a, b, n) class. *Ann. Actuar. Sci.* **2023**, *17*, 145–169. [CrossRef]
31. Tzaninis, S.M.; Bozikas, A. Extensions of Panjer's recursion for mixed compound distributions. *arXiv* **2024**, arXiv:2406.17726.
32. Ong, S.H.; Sim, S.Z.; Liu, S.; Srivastava, H.M. A family of finite mixture distributions for modelling dispersion in count data. *Stats* **2023**, *6*, 942–955. [CrossRef]
33. Johnson, N.L.; Kemp, A.W.; Kotz, S. *Univariate Discrete Distributions*; Wiley: Hoboken, NJ, USA, 2005. [CrossRef]
34. Patil, G.P.; Boswell, M.T.; Joshi, S.W.; Ratnaparkhi, M.V. *Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work*; International Cooperative Publishing House: Fairland, MD, USA, 1884.
35. Marques, T.A.; Pestana, D.; Velosa, S. Count data models in Biometry and randomness patterns in birds extra-pair paternity. *Biom. Lett.* **2005**, *42*, 81–112.
36. Brilhante, M.F.; Pestana, P. BetaBoop function, BetaBoop random variables and extremal population growth. In *International Encyclopedia of Statistical Science*, 2nd ed.; Lovric, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2024.
37. Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series, and Products*, 7th ed.; Elsevier/Academic Press: Amsterdam, The Netherlands, 2007.
38. Chakraborty, S. Generating discrete analogues of continuous probability distributions—A survey of methods and constructions. *J. Stat. Distrib. Appl.* **2015**, *2*, 6. [CrossRef]
39. Kosambi, D.D. Characteristic properties of series distributions. *Proc. Natl. Inst. Sci. India* **1949**, *15*, 109–113.
40. Kapur, J.N. *Maximum-Entropy Models in Science and Engineering*; Wiley Eastern Limited: New Delhi, India, 1989.
41. Steutel, F.W.; van Harn, K. Discrete analogues of self-decomposability and stability. *Ann. Probab.* **1979**, *7*, 893–899. [CrossRef]
42. Hall, A.O. Maximum term of a particular autoregressive sequence with discrete margins. *Commun. Stat. Theory Methods* **1996**, *25*, 721–736. [CrossRef]
43. Zhang, H.; Tan, K.; Li, B. COM-negative binomial distribution: Modeling overdispersion and ultrahigh zero-inflated count data. *Front. Math. China* **2018**, *13*, 967–998. [CrossRef]
44. Cahoy, D.; Di Nardo, E.; Polito, F. Flexible models for overdispersed and underdispersed count data. *Stat. Pap.* **2021**, *62*, 2969–2990. [CrossRef]

45. Huang, A. On arbitrarily underdispersed discrete distributions. *Am. Stat.* **2022**, *77*, 29–34. [CrossRef]
46. Seck, N.K.G.; Ngom, A.; Noba, K. Modelling underdispersed count data: Relative performance of Poisson model and its alternatives. *Afr. J. Math. Stat. Stud.* **2022**, *5*, 16–32.
47. Rana, S.; Al Mamun, A.S.M.; Rahman, F.A.; Elgohari, H. Outliers as a source of overdispersion in Poisson regression modelling: Evidence from simulation and real data. *Int. J. Stat. Sci.* **2023**, *23*, 31–37. [CrossRef]
48. Sengupta, D.; Roy, S. Modelling zero inflated and under-reported count data. *J. Stat. Comput. Simul.* **2023**, *93*, 29–34. [CrossRef]
49. Aswi, A.; Astuti, S.A.; Sudarmin, S. Evaluating the performance of zero-inflated and hurdle Poisson models for modelling overdispersion in count data. *Inferensi* **2022**, *5*, 17–22. [CrossRef]
50. Raikov, D. On the decomposition of Poisson laws. *Dokl. Acad. Sci. URSS* **1937**, *14*, 9–11.
51. Lévy, P. Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchaînées. *J. Math. Pures Appl.* **1935**, *14*, 347–402.
52. Cramér, H. Über eine Eigenschaft der normalen Verteilungsfunktion. *Math. Z.* **1936**, *41*, 405–414. [CrossRef]
53. Johansen, S. An application of extreme point methods to the representation of infinitely divisible distributions. *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* **1966**, *5*, 304–316. [CrossRef]
54. De Finetti, B. Sulle funzioni ad incremento aleatorio. *Rend. Accad. Naz. Lincei* **1929**, *X*, 163–168.
55. De Finetti, B. Funzione caratteristica di un fenomeno aleatorio. In Proceedings of the Atti del Congresso Internazionale dei Matematici, Bologna, Italy, 3–10 September 1928; Volume VI, pp. 179–190.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.