Austrian Journal of Statistics
2024, Volume 53, 87–101.
http://www.ajs.or.at/
doi:10.17713/ajs.v53i3.1776

Yet Another Attempt to Classify Positive Univariate Probability Distributions

Yves Tillé
University of Neuchatel
Switzerland

Abstract

We propose an original classification of several discrete and continuous probability distributions. We establish links between these distributions, in particular the little known relationship between the negative hypergeometric distribution and the beta distribution. These relations allow us to propose a structure of relations which is summarized in graphic form. Our classification emphasises the analogy between certain discrete and continuous distributions. This analogy makes it possible to establish relations between the theory of point processes and the theory of survey sampling. It also makes it possible to envisage the use of link functions that are little used in generalised regression.

Keywords: continuous distribution, discrete distribution, hypergeometric, negative distribution.

1. Introduction

The graphs of relationships between probability distributions like those proposed by Leemis and McQueston (2008) and Song (2005) sometimes look a little like spaghetti dishes in which it is difficult to glimpse a structure. Morris and Lock (2009) proposed a structured map of the main distributions. Our aim is also to propose a general framework in the sense where we establish links between discrete and continuous distributions, distributions related to sampling with and without replacement, distributions and their negative (or inverse distributions).

Our objective is to make as complete a list as possible of the links between discrete and continuous positive probability distributions. In Section 2, we list some discrete distributions, most of which are special cases of the Pólya–Eggenberger distribution. In Section 3, we present some continuous distributions and describe the links between these distributions. In Section 4, we recall that several distributions can be defined as compound probability distributions where the parameters of an original distribution are themselves random variables. Finally, in Section 5, we establish links between discrete and continuous distributions. In particular, we show that the negative hypergeometric distribution can be considered as the discrete analogue of the beta distribution.

In section 6, we show that these analogies between discrete and continuous distributions make it possible to establish links between very different areas of statistics. We can show that sampling algorithms can be seen as point processes. In addition, we can imagine the use of link functions that are unusual in generalised regression. In Section 7, we draw a brief conclusion.

2. Discrete probability distributions

A list of discrete distributions is described in Table 1.

2.1. The most common probability distributions

Several common discrete distributions of probability are described here rapidly. Their distributions of probability, expectations, and variances are given in Table 1.

- The *Bernoulli random variable*, denoted by $\mathcal{Bern}(p)$ is the discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $1 - p$.

- The *binomial random variable*, denoted by $X \sim \mathcal{Bin}(n, p)$, is the number of successes in n independent trials when each success has a probability p to occur. The sum of n Bernoulli variables with the same parameter p is a binomial variable. When $n = 1$, the binomial random variable reduces to a Bernoulli random variable.

- Suppose that drawings are repeated independently in a population that has a proportion p of favorable events. The drawings are repeated until the first favorable event appears. The *geometric random variable*, $X \sim \mathcal{Geom}(1 - p)$, counts the number of failures before the first favorable event occurs.

- Suppose that drawings are repeated independently in a population that has a proportion p of favorable events. The *negative binomial random variable*, $X \sim \mathcal{NB}(r, p)$ counts the number of failures until obtaining r favorable events. This distribution can be generalized for non-integer values of $r > 0$. When $r = 1$ the negative binomial is a geometric random variable. Parameter r can be non–integer.

- The *Poisson random variable*, denoted by $X \sim \mathcal{Poiss}(\lambda)$, can be applied to systems with a large number of possible events, each of which is rare. The number of such events that occur has a Poisson distribution.

2.2. Hypergeometric distribution

Consider a population of size N that has a proportion p of favorable events. Thus Np units are ‘favorable’ and $N(1 - p)$ are ‘not favorable’. The hypergeometric random variable, $X \sim \mathcal{H}(N, n, Np)$, counts the number favorable events obtained in n drawings with equal probabilities from the population without replacement.

When N tends to infinity, the hypergeometric random variable reduces to the binomial random variable. As for the binomial random variable, when $n = 1$, the hypergeometric random variable also reduces to a Bernoulli random variable. The hypergeometric distribution can be generalized to non–integer values for Np

$$\begin{aligned}\Pr(X = x) &= p(x; N, n, Np)\\ &= \frac{\Gamma(Np+1)\Gamma(N(1-p)+1)}{x!\Gamma(Np-x+1)(n-x)!\Gamma(N(1-p)-n+x+1)}\binom{N}{n}^{-1},\end{aligned}$$

with $x \in \{\mathbb{N}|\max(0, n + Np - N) \le x \le \min(Np, n)\}$.

Table 1: Discrete distributions of probability

Name	Notation	PMF	Support	Parameters	Mean	Variance
Bernoulli	$\mathcal{Bern}(p)$	$p^x(1-p)^{1-x}$	$x \in \{0,1\}$	$p \in [0,1]$	p	$p(1-p)$
Binomial	$\mathcal{Bin}(n,p)$	$\binom{n}{x} p^x(1-p)^{n-x}$	$x \in \{0,1,\dots,n\}$	$p \in [0,1], n \in \mathbb{N}$	np	$np(1-p)$
Geometric	$\mathcal{Geom}(1-p)$	$p(1-p)^x$	$x \in \mathbb{N}$	$p \in [0,1]$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	$\mathcal{NB}(r,p)$	$\frac{\Gamma(r+x)}{x!\,\Gamma(r)} p^r(1-p)^x$	$x \in \mathbb{N}$	$p \in [0,1], r \in \mathbb{R}_{>0}$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Poisson	$\mathcal{Poiss}(\lambda)$	$\frac{e^{-\lambda}\lambda^x}{x!}$	$x \in \mathbb{N}$	$\lambda \in \mathbb{R}_{>0}$	λ	λ
Hypergeometric	$\mathcal{H}(N,n,Np)$	$\frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}}$	$x \in \{\mathbb{N} \|$ $\max(0, n+Np-N)$ $\le x \le \min(Np,n)\}$	$Np, N, n \in \mathbb{N}_{>0}$, $Np \le N$ $n \le N$	np	$np(1-p)\frac{N-n}{N-1}$
Negative Hypergeometric	$\mathcal{NH}(N,r,Np)$ $= \mathcal{BetaBin}(N-Np,$ $r, Np-r+1)$	$\frac{\Gamma(Np+1)\Gamma(x+r)\Gamma(N-x-r+1)\Gamma(N-Np+1)}{x!N!\Gamma(r)\Gamma(Np-r+1)\Gamma(N-Np-x+1)}$	$x \in \{0,1,\dots,N(1-p)\}$	$N \in \mathbb{N}_{>0}$ $Np, r \in \mathbb{R}_{>0}$ $Np \le N$ $r \le Np$	$\frac{Nr(1-p)}{Np+1}$	$\frac{rN(1-p)(N+1)(Np-r+1)}{(Np+1)^2(Np+2)}$
Beta-binomial	$\mathcal{BetaBin}(n,\alpha,\beta)$ $= \mathcal{NH}(n+\alpha+\beta-1,$ $\alpha, \alpha+\beta-1)$	$\frac{\Gamma(n+1)\Gamma(x+\alpha)\Gamma(n-x+\beta)\Gamma(\alpha+\beta)}{\Gamma(x+1)\Gamma(n-x+1)\Gamma(n+\alpha+\beta)\Gamma(\alpha)\Gamma(\beta)}$	$x \in \{0,1,\dots,n\}$	$n \in \mathbb{N}_{>0}$ $\alpha, \beta \in \mathbb{R}_{>0}$ $r \le Np$	$\frac{n\alpha}{\alpha+\beta}$	$\frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Discrete Uniform	$\mathcal{DU}(A)$	$\frac{1}{A+1}$	$x \in \{0,1,\dots,A\}$	$A \in \mathbb{N}_{>0}$	$\frac{A}{2}$	$\frac{(A+1)^2-1}{12}$
Pólya–Eggenberger	$\mathcal{PE}(x,N,n,Np,c)$	$\binom{n}{x}\frac{\prod_{z=0}^{x-1}(Np+cz)\prod_{z=0}^{n-x-1}(N-Np+cz)}{\prod_{z=0}^{n-1}(N+cz)}$	$x \in \{0,1,\dots,n\}$	$n, N \in \mathbb{N}_{>0}, n < N$, $p \in [0,1], c \in \mathbb{R}$	np	$np(1-p)\frac{N+cn}{N+c}$

In this table, PMF means probability mass function, $\Gamma(t) = \int_0^\infty u^{t-1}e^{-u}\,\mathrm{d}u$, $\binom{n}{x} = n!/[x!(n-x)!]$, $\mathbb{N} = \{0,1,2,\dots\}$, $\mathbb{N}_{>0} = \{1,2,3,\dots\}$, $\mathbb{R}_{>0}$ denotes the set of positive real numbers (without 0).

2.3. Negative hypergeometric or beta-binomial distribution

The negative (or inverse) hypergeometric variable is curiously still little known to such an extent that it is presented as a "forgotten distribution" by Miller and Fridell (2007). This distribution is however the counterpart of the negative binomial distribution when sampling is without replacement.

Consider a population of size N that has a proportion p of favorable events. Thus Np units are 'favorable' and $N(1-p)$ are not 'favorable'. Suppose that drawings are repeated without replacement until r favorable events occur. The negative hypergeometric random variable, $X \sim \mathcal{NH}(N, r, Np)$, counts the number of failures until obtaining r favorable events.

The PMF of a negative hypergeometric random variable, denoted by $X \sim \mathcal{NH}(N, r, Np)$, is

$$\begin{aligned}
\Pr(X = x) &= p(x; N, r, Np) = \frac{\binom{x+r-1}{x}\binom{N-x-r}{Np-r}}{\binom{N}{Np}} \\
&= \binom{x+r-1}{x} \frac{[N(1-p)]^{\underline{x}}(Np)^{\underline{r}}}{N^{\underline{r+x}}} \qquad (1) \\
&= \binom{N-Np}{x} \frac{(x+r-1)^{\underline{x}}(N-x-r)^{\underline{N-Np-x}}}{N^{\underline{N-Np}}} \\
&= \binom{N-Np}{x} \frac{(r)^{\overline{x}}(Np-r+1)^{\overline{N-Np-x}}}{(Np+1)^{\overline{N-Np}}} \qquad (2) \\
&= \frac{\Gamma(Np+1)\Gamma(x+r)\Gamma(N-x-r+1)\Gamma(N-Np+1)}{x!N!\Gamma(r)\Gamma(Np-r+1)\Gamma(N-Np-x+1)} \qquad (3) \\
&= \frac{1}{\mathrm{B}(r, Np-r+1)} \frac{(x+1)^{\overline{r-1}}(N-Np-x+1)^{\overline{Np-r}}}{(N-Np+1)^{\overline{Np}}}, \qquad (4)
\end{aligned}$$

with $x \in \{0, 1, \ldots, N(1-p)\}$, $N \in \{1, 2, \ldots\}$, and $Np \in \{1, 2, \ldots, N\}$, and $r \in \{1, 2, \ldots, Np\}$. The notation $N^{\underline{r}} = N!/(N-r)!$ is the falling factorial. The notation $N^{\overline{r}} = (N+r-1)!/(N-1)!$ is the rising factorial. Moreover

$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

$$\mathrm{E}(X) = \frac{Nr(1-p)}{Np+1}, \mathrm{var}(X) = \frac{rN(1-p)(N+1)(Np-r+1)}{(Np+1)^2(Np+2)}.$$

When N tends to infinity, the negative hypergeometric random variable reduces to the negative binomial random variable. The similarity between these two variables is striking by comparing Expression (1) with the distribution of probability of the negative binomial. Using Expression (3), non–integer values can be used for Np and r. The negative hypergeometric distribution can be generalized for non–integer values for r and Np.

The beta-binomial distribution is a negative hypergeometric distribution with another parametrization that is frequently used in Bayesian statistics. The beta-binomial distribution is the binomial distribution $\mathcal{Bin}(n, p)$ in which the probability of success p has a beta distribution. If X has a beta-binomial distribution, then

$$\Pr(X = x) = \frac{\Gamma(n+1)\Gamma(x+\alpha)\Gamma(n-x+\beta)\Gamma(\alpha+\beta)}{\Gamma(x+1)\Gamma(n-x+1)\Gamma(n+\alpha+\beta)\Gamma(\alpha)\Gamma(\beta)}, \qquad (5)$$

where $n \in \mathbb{N}_{>0}, \alpha, \beta \in \mathbb{R}_{>0}$. If we take, $n = N - Np, \alpha = r, \beta = Np - r + 1$, the right-hand-side of (5) is equal to (3). We then have

$$\mathrm{E}(X) = \frac{n\alpha}{\alpha+\beta}, \text{ and } \mathrm{var}(X) = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

2.4. Negative hypergeometric random variable: case $r = 1$

When $r = 1$, the PMF of a negative hypergeometric random variable with $r = 1$, denoted by $\mathcal{NH}(N, 1, Np)$, is

$$\begin{aligned}\Pr(X = x) &= p(x; N, 1, Np) = \frac{\binom{N-x-1}{Np-1}}{\binom{N}{Np}} \\ &= p\frac{(N-x-1)^{\underline{Np-1}}}{(N-1)^{\underline{Np-1}}} = p\frac{(N-Np)^{\underline{x}}}{(N-1)^{\underline{x}}},\end{aligned}$$

with $x \in \{0, 1, \ldots, N(1-p)\}$, $N \in \{0, 1, 2, \ldots\}$, and $N(1-p) \in \{0, 1, \ldots, N\}$.

$$\mathrm{E}(X) = \frac{N(1-p)}{Np+1}, \mathrm{var}(X) = \frac{N(1-p)(N+1)(Np)}{(Np+1)^2(Np+2)}.$$

2.5. Discrete uniform distribution

The PMF of the Discrete uniform random variable $\mathcal{DU}(A)$ is

$$\Pr(X = x) = p(x; 0, A) = \frac{1}{A+1}, x \in \{0, 1, \ldots, A\},$$

with $A = 0, 1, 2, 3, \ldots$.

$$\mathrm{E}(X) = \frac{A}{2}, \mathrm{var}(X) = \frac{(A+1)^2 - 1}{12}.$$

The discrete uniform distribution is a special case of the negative hypergeometric distribution. Indeed,

$$\mathcal{NH}(N, 1, 1) = \mathcal{DU}(N-1).$$

2.6. Pólya–Eggenberger distribution

The Pólya–Eggenberger random variable was first introduced by Eggenberger and Pólya (1923, 1928) (see also Johnson, Kemp, and Kotz 2005, pp. 258–259). Mahmoud (2008) has dedicated a book to the *Pólya Urn Models*. Suppose that we have an urn containing N balls with Np white balls and $N - Np$ black balls. At each trial a ball is selected and is replaced in the urn with c balls of the same color. Value c can be negative. In this case $-c$ balls of the same color are removed from the urn after the drawing.

The Pólya–Eggenberger random variable, $\mathcal{PE}(x, N, n, Np, c)$, counts the number of white balls selected after n trials:

$$\begin{aligned}\Pr(X = x) &= p(x; N, n, Np, c) \\ &= \binom{n}{x} \frac{\prod_{z=0}^{x-1}(Np + cz) \quad \prod_{z=0}^{n-x-1}(N - Np + cz)}{\prod_{z=0}^{n-1}(N + cz)},\end{aligned}$$

$x = 0, \ldots, n$.

$$\mathrm{E}(X) = np, \mathrm{var}(X) = np(1-p)\frac{N+cn}{N+c}.$$

The special cases of the Pólya–Eggenberger distribution are presented in Table 2.

2.7. Link between the discrete variables

Consider a sequence of m independent and identically distributed random variables and com-

Table 2: Some special cases of the Pólya–Eggenberger distribution

$c=-1$	$\mathcal{PE}(N,n,Np,-1)=\mathcal{H}(N,n,Np)$	Hypergeometric
$c=0$	$\mathcal{PE}(N,n,Np,0)=\mathcal{Bin}(n,p)$	Binomial
$c=1$	$\mathcal{PE}(N,n,Np,1)=\mathcal{NH}(N+n-1,N-1,Np)$	Negative hypergeometric
	or equivalently	
$c=1$	$\mathcal{PE}(Np+1,N-Np,r,1)=\mathcal{NH}(N,r,Np)$	Negative hypergeometric

pute the conditional distribution of X_j with respect to the sum:

$$\begin{aligned}\Pr\left(X_j=x\left|\sum_{i=1}^m X_i=S\right.\right) &= \frac{\Pr\left(\sum_{i=1}^m X_i=S|X_j=x\right)}{\Pr(\sum_{i=1}^m X_i=S)}\Pr(X_j=x)\\ &= \frac{\Pr\left(\sum_{i=1}^m X_i-X_j=S-x|X_j=x\right)}{\Pr(\sum_{i=1}^m X_i=S)}\Pr(X_j=x)\\ &= \frac{\Pr\left(\sum_{i=1}^m X_i-X_j=S-x\right)}{\Pr(\sum_{i=1}^m X_i=S)}\Pr(X_j=x).\end{aligned}$$

The distributions of the sums and the distributions of the conditional distributions are presented in Table 3 for several discrete random variables. These results are well–known and are cited in Johnson *et al.* (2005). The link between the negative binomial and the negative hypergeometric is due to Patil and Seshadri (1964) and Kagan, Linnik, and Rao (1973).

Table 3: Let $X_1,\ldots,X_m$ a sequence of m independent random variables. Computation of the distributions of the sums of these variables and of variables X_j conditionally to this sum.

Variable X_i	$S=\sum_{i=1}^m X_i$	$X_j\vert S$
$\mathcal{Poiss}(\lambda)$	$\mathcal{Poiss}(m\lambda)$	$\mathcal{Bin}(S,\frac{1}{m})$
$\mathcal{Bern}(p)$	$\mathcal{Bin}(m,p)$	$\mathcal{H}(m,S,1)$
$\mathcal{Bin}(n,p)$	$\mathcal{Bin}(mn,p)$	$\mathcal{H}(nm,S,m)$
$\mathcal{Geom}(1-p)$	$\mathcal{NB}(m,p)$	$\mathcal{NH}(S+m-1,1,m-1)$
$\mathcal{NB}(r,p)$	$\mathcal{NB}(mr,p)$	$\mathcal{NH}(S+mr-1,r,mr-1)$

In the other direction, the results presented in Table 4 show that the distributions presented in the last column of Table 3 converge in distribution to the distributions of the first column of this table. Most of these results are well–known except the negative binomial approximation of the negative hypergeometric distribution studied by López-Blázquez and Salamanca-Miño (2001) and the binomial approximation of the negative hypergeometric distribution that can be obtained by approximating the raising factorial by a power function in Expression (2) (see, for instance Terrell 1999, p. 182).

Following Miller and Fridell (2007), we can summarize these discrete distributions into two tables. Table 5 presents the usual distributions for sampling with and without replacement with the special case $n=1$. Table 6 presents the corresponding negative (or inverse) distributions.

3. Continuous distributions

3.1. The most usual continuous distributions

The most usual continuous distributions are described in Table 7. The *exponential random*

Table 4: Asymptotic results between discrete distributions

$\mathcal{H}(N, n, Np)$	$\xrightarrow[N\to\infty]{\mathcal{D}}$	$\mathcal{Bin}(n, p)$
$\mathcal{NH}(N, 1, Np)$	$\xrightarrow[N\to\infty]{\mathcal{D}}$	$\mathcal{Geom}(1-p)$
$\mathcal{NH}(N, r, Np)$	$\xrightarrow[N\to\infty]{\mathcal{D}}$	$\mathcal{NB}(r, p)$
$\mathcal{NH}(N + nr - 1, zr, rn - 1)$	$\xrightarrow[r\to\infty]{\mathcal{D}}$	$\mathcal{Bin}\left(N, \frac{z}{n}\right)$
$\mathcal{NB}\left(r, \dfrac{r}{\lambda + r}\right)$	$\xrightarrow[r\to\infty]{\mathcal{D}}$	$\mathcal{Poiss}(\lambda)$
$\mathcal{Bin}\left(n, \dfrac{\lambda}{n}\right)$	$\xrightarrow[n\to\infty]{\mathcal{D}}$	$\mathcal{Poiss}(\lambda)$

Table 5: Basic distributions for sampling with and without replacement. When $n = 1$, sampling with and without replacement are confounded.

	Sampling without replacement	Sampling with replacement
any $n \in \mathbb{N}$	Hypergeometric	Binomial
$n = 1$	Bernoulli	Bernoulli

Table 6: Negative (or inverse) distributions corresponding to Table 5

	Sampling without replacement	Sampling with replacement
any $r \in \mathbb{N}$	Negative Hypergeometric	Negative Binomial
$r = 1$	Negative Hypergeometric $r = 1$	Geometric

variable is denoted by $\mathcal{Exp}(\lambda)$, where $\lambda > 0$. The *gamma random variable* is denoted by $\mathcal{Gamma}(r, \theta)$, where $r > 0, \theta > 0$. When $r = 1$ and $\theta = 1/\lambda$ the gamma variable reduces to the exponential variable. The *beta random variable* ($\in [0, 1]$) is denoted by $\mathcal{Beta}(\alpha, \beta)$, where $\alpha > 0, \beta > 0$. The *continuous uniform random variable* (in $[0, 1]$) is denoted by $\mathcal{CU}(0, 1)$ and is is the special case of the beta distribution when $\alpha = 1$ and $\beta = 1$.

Table 7: Continuous distributions of probability

Notation	PDF	Support	Parameters	Mean	Variance
$\mathcal{Gamma}(r, \theta)$	$\frac{x^{r-1}e^{-x/\theta}}{\Gamma(r)\theta^r}$	$x \in \mathbb{R}_{>0}$	$r, \theta \in \mathbb{R}_{>0}$	$r\theta$	$r\theta^2$
$\mathcal{Exp}(\lambda)$	$\lambda e^{-\lambda x}$	$x \in \mathbb{R}_{>0}$	$\lambda \in \mathbb{R}_{>0}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathcal{Beta}(\alpha, \beta)$	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$	$x \in [0, 1]$	$\alpha, \beta \in \mathbb{R}_{>0}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
$\mathcal{CU}(a, b)$	$\frac{1}{b-a}$	$x \in [a, b]$	$a, b \in \mathbb{R}_{>0}, a < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

3.2. Relationships between the continuous variables

There also exists links between these continuous variables. As for the discrete variables (see Table 3), consider the sequence $X_1, \dots, X_m$ of independent continuous variables. We can derive the conditional distribution of one of these variables to their sum. The resulting relations are given in Table 8

Table 8: Let $X_1, \dots, X_m$ a sequence of m independent random variables. Computation of the distributions of the sum S of these variables and of variable X_j conditionally to this sum.

X_i	$S = \sum_{i=1}^m X_i$	$X_j \mid S$
$\mathcal{Exp}(\theta)$	$\mathcal{Gamma}(m, \theta)$	$\mathcal{Beta}(1, (m-1))$
$\mathcal{Gamma}(r, \theta)$	$\mathcal{Gamma}(mr, \theta)$	$\mathcal{Beta}(r, (m-1)r)$

As in the discrete case, it is possible to return from the last column of Table 8 to the first column by asymptotic results given in Table 9.

Table 9: Asymptotic results between continuous univariate distributions related to the conditioning on the sum presented in Table 8

$m\dfrac{\mathcal{Beta}(1, m)}{\lambda}$	$\xrightarrow[m\to\infty]{\mathcal{D}}$	$\mathcal{Exp}(\lambda)$
$m\theta\ \mathcal{Beta}(r, m)$	$\xrightarrow[m\to\infty]{\mathcal{D}}$	$\mathcal{Gamma}(r, \theta)$

4. Using prior distribution for the parameters

A compound probability distribution is the distribution obtained when one of the parameter of an original distribution is itself a random variable. In the framework of Bayesian inference, a prior probability distribution is a probability distribution used to model the parameter of another distribution, to produce a posterior distribution. For instance, for a Bernoulli random variable $\mathcal{Bern}(p)$, the prior for p can be a beta random variable $\mathcal{Beta}(\alpha, \beta)$. In this case, the posterior distribution is

$$\int_0^1 p^x (1-p)^{1-x} \frac{p^\alpha (1-p)^\beta)}{\mathrm{B}(\alpha, \beta)} dp = \frac{\mathrm{B}(b-x+1, a+x)}{\mathrm{B}(\alpha, \beta)}. \tag{6}$$

Expression (6) is the PMF of the beta-binomial distribution with $n = 1$ ($\mathcal{BetaBin}(n = 1, \alpha, \beta)$). A list of composition is given in Table 10.

Table 10: Composition of some discrete distributions

Original distribution	Distribution of the parameter	Marginal distribution
$\mathcal{Poiss}(\lambda)$	$\lambda \sim \mathcal{Gamma}(r, \theta)$	$\mathcal{NB}\left(r, \frac{\theta}{1+\theta}\right)$
$\mathcal{Poiss}(\lambda)$	$\lambda \sim \mathcal{Exp}(1/\theta)$	$\mathcal{Geom}\left(\frac{\theta}{1+\theta}\right)$
$\mathcal{Bern}(p)$	$p \sim \mathcal{Beta}(\alpha, \beta)$	$\mathcal{BetaBin}(1, \alpha, \beta)$
$\mathcal{Bin}(n, p)$	$p \sim \mathcal{Beta}(\alpha, \beta)$	$\mathcal{BetaBin}(n, \alpha, \beta)$

5. Relations between the discrete and the continuous variables

Relations between random variables are described in Leemis and McQueston (2008), Leemis, Luckett, Powell, and Vermeer (2012), Johnson, Kotz, and Balakrishnan (2000) and Johnson *et al.* (2005). Some relations are however unheralded. The geometric distribution is usually said to be a discrete analogue of the exponential distribution (Johnson *et al.* 2005, p. 210). The negative binomial is also considered as the discrete analogue of the gamma distribution (Young 1970; Adell and de la Cal 1994). The analogy between these two variables can be seen directly by comparing the probability distributions of these variables in Figure 1.

A probably less known relation is the link between the negative hypergeometric distribution and the beta distribution (Särndal 1968; Bowman, Kastenbaum, and Shenton 1992). A proof of the convergence of distribution, due to Eggenberger and Pólya (1923), is presented in Mahmoud (2008, p. 53) for the Pólya urn model where the negative hypergeometric distribution arises as a special case. The analogy between these two variables can be seen directly when we compare their distributions of probability in Figure 2. These relations are presented in Table 11. An interesting special case is that $\mathcal{Beta}(1,1)$ has a continuous uniform distribution in $[0,1]$. Analogously, $\mathcal{NH}(N,1,1)$ has a discrete uniform distribution on $\{0,1,\ldots,N-1\}$.

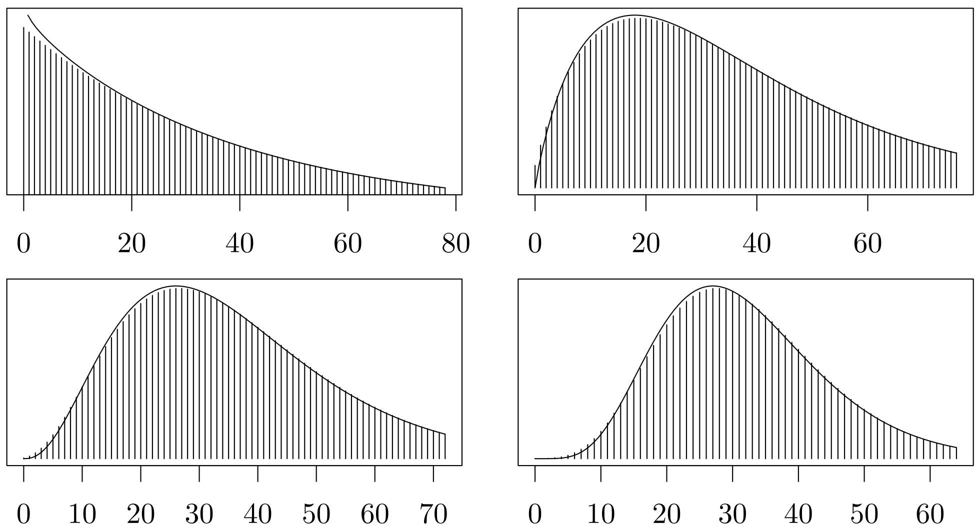

Figure 1: Adjustment of negative binomial distributions by continuous gamma distributions. The parameters of the negative binomial variables are $(r=1,p=0.025)$, $(r=2,p=0.05)$, $(r=4,p=0.1)$ and $(r=8,p=0.2)$.

Table 11: Analogy between discrete and continuous distributions

$\dfrac{\mathcal{NH}(N,r,Np)}{N-Np}$	$\xrightarrow[\substack{N\to\infty\\ p\to 0\\ Np-r+1\to\beta}]{\mathcal{D}}$	$\mathcal{Beta}(r,\beta)$
$\dfrac{\mathcal{NH}(N,1,Np)}{N-Np}$	$\xrightarrow[\substack{N\to\infty\\ p\to 0\\ Np\to\beta}]{\mathcal{D}}$	$\mathcal{Beta}(1,\beta)$
$p\theta\mathcal{NB}(r,p)$	$\xrightarrow[p\to 0]{\mathcal{D}}$	$\mathcal{Gamma}(r,\theta)$
$p\dfrac{\mathcal{Geom}(1-p)}{\lambda}$	$\xrightarrow[p\to 0]{\mathcal{D}}$	$\mathcal{Exp}(\lambda)$
$\dfrac{\mathcal{DU}(A)}{A}$	$\xrightarrow[A\to\infty]{\mathcal{D}}$	$\mathcal{CU}(0,1)$

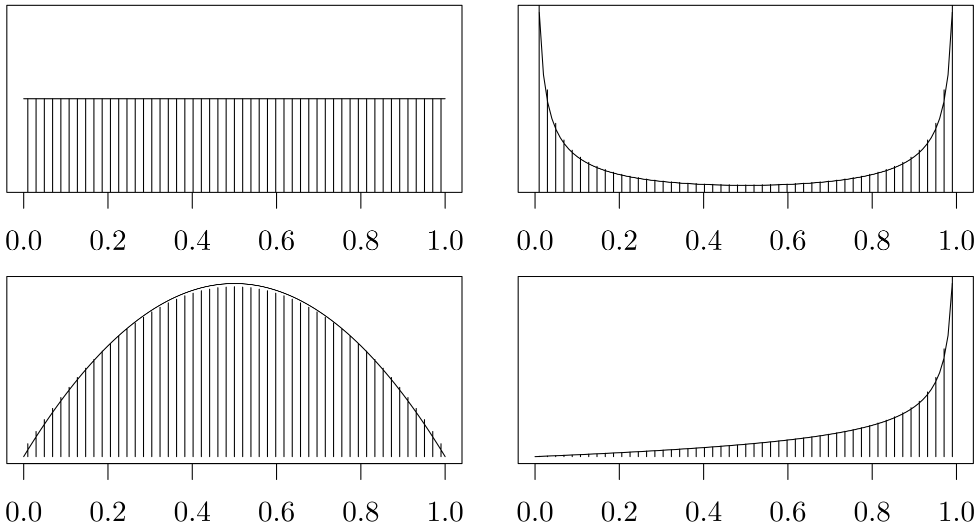

Figure 2: Adjustment of beta-binomial distributions by continuous beta distributions. The parameters of the beta-binomial variables are $(n = 50, \alpha = 1, \beta = 1)$, $(n = 50, \alpha = 0.5, \beta = 0.5)$, $(n = 50, \alpha = 2, \beta = 2)$, $(n = 50, \alpha = 2, \beta = 0.5)$.

The relations between the variables contained in Tables 3, 4, 6, 8, 9, and 8 are summarized in Figure 3. Other relations are presented in Morris and Lock (2009) (see also Morris 1982, 1983).

6. The importance of links between variables

6.1. Relation between point processes and sampling algorithms

Our classification of probability distributions makes it possible to establish links between very different areas of statistics. The analogy between continuous and discrete distributions is known in the theory of point processes. For example, the discrete analogue of the well-known continuous Poisson process is the Bernoulli process (Daley and Vere-Jones 2007). For the Poisson process, the time elapsing between two increments of the counting process follows an exponential distribution. The Bernoulli process is a sequence of independent Bernoulli random variables with the same parameter. The number of trials required to obtain a success follows a geometric distribution, which is the discrete analogue of the exponential distribution. In sampling theory, the Bernoulli design is a sampling design in which each unit is selected with the same probability independently of the other units (Wilhelm, Qualité, and Tillé 2017).

Let us also consider a sequence of n continuous random variables $U_1, \dots, U_n$ with a uniform distribution in $[0, 1]$. The order statistic $U_{(1)}, \dots, U_{(n)}$ is the sequence of values sorted in increasing order. The difference $U_{(t+1)} - U_{(t)}$ follows a beta distribution (see, for instance David and Nagaraja 2004). In the theory of sampling from finite population $U = \{1, \dots, k, \dots, N\}$, suppose that we select a simple random sample without replacement of fixed size n. The number of non-selected units between two selected units has a negative hypergeometric distribution. This property had already been used by Vitter (1987) to propose a rapid skip algorithm for implementing simple random sampling. We can jump over a certain number of units in the population to land directly on the next unit to be selected. The analogy between the negative hypergeometric and the beta distributions therefore make it possible to establish a link between non-parametric statistics and survey sampling theory.

Wilhelm *et al.* (2017) studied continuous processes in which the distribution of the interval between two increments is either a gamma distribution or a beta distribution. By adjusting the parameters of the distributions, it is possible to produce either a systematic sampling effect or a clustering effect. Tillé, Wilhelm, and Qualité (2018) then studied discrete processes. They showed that we can pass from a systematic design to a simple design and finally to a design with a cluster effect by simply adjusting the parameters of a negative hypermeometric distribution. Second-order inclusion probabilities are even computable.

6.2. Generalized regression and link function

In generalized linear models, a link function is used to adapt to the specificity of the dependent variables. If a variable is dichotomous, logistic regression can be used. If the variable is discrete and non-negative, Poisson regression can be used. Regression with a gamma link function is the continuous analogue of regression with a negative binomial link function. For continuous variables between 0 and 1, Kieschnick and McCullough (2003) and Ferrari and Cribari-Neto (2004) have used beta variable for the link function.

For a discrete variable between 0 and n, the link function could be a beta-binomial (or negative hypergeometric) variable. Beta-binomial regression is little known, but it is the discrete analogue of beta regression. This relatively little-known method has nevertheless been the subject of several publications (Forcina and Franconi 1988; Martin, Witten, and Willis 2020; Najera-Zuloaga, Lee, and Arostegui 2018; Martin *et al.* 2020). It would certainly suit the analysis of data such as Likert rating scales (Likert 1932) whose values are integer between 1 and n.

7. Conclusions

The classification we propose draws a parallel between discrete and continuous distributions. The sampling algorithms can then be seen as discrete point processes. We can also envisage the use of link functions that are rarely used in generalised regression.

Appendix: Compound distributions

Binomial with n Poisson is Poisson

If $X \sim \mathcal{Bin}(n,p)$, with $n \sim \mathcal{Poiss}(\lambda)$ is $\mathcal{Poiss}(p\lambda)$. Indeed,

$$\begin{aligned}\sum_{n=x}^{\infty}\binom{n}{x}p^x(1-p)^{n-x}\frac{e^{-\lambda}\lambda^n}{n!} &= \frac{p^x e^{-\lambda}\lambda^x}{x!}\sum_{n=x}^{\infty}\frac{(1-p)^{n-x}\lambda^{n-x}}{(n-x)!}\\ &= \frac{p^x e^{-\lambda}}{x!}\lambda^x e^{\lambda(1-p)} = \frac{(p\lambda)^x e^{-p\lambda}}{x!}.\end{aligned}$$

Binomial with n Geometric is Geometric

If $X \sim \mathcal{Bin}(n,p)$, with $n \sim \mathcal{Geom}(\pi)$ is $\mathcal{Geom}(\pi/(p+\pi(1-p)))$. Indeed,

$$\begin{aligned}\sum_{n=x}^{\infty}\binom{n}{x}p^x(1-p)^{n-x}\pi(1-\pi)^n &= \frac{\pi(1-\pi)^x p^x}{x!}\sum_{n=x}^{\infty}\frac{n!}{(n-x)!}(1-p)^{n-x}(1-\pi)^{n-x}\\ &= \frac{\pi(1-\pi)^x p^x}{x!}\frac{x!}{(p+\pi(1-p))^{x+1}}\\ &= \left[\frac{(1-\pi)p}{(p+\pi(1-p))}\right]^x\frac{\pi}{(p+\pi(1-p))}.\end{aligned}$$

Binomial with n Negative Binomial is Negative Binomial

If $X \sim \mathcal{Bin}(n,p)$, with $n \sim \mathcal{NB}(\pi,r)$ is $\mathcal{NB}(\pi/(p+\pi(1-p)),r)$.

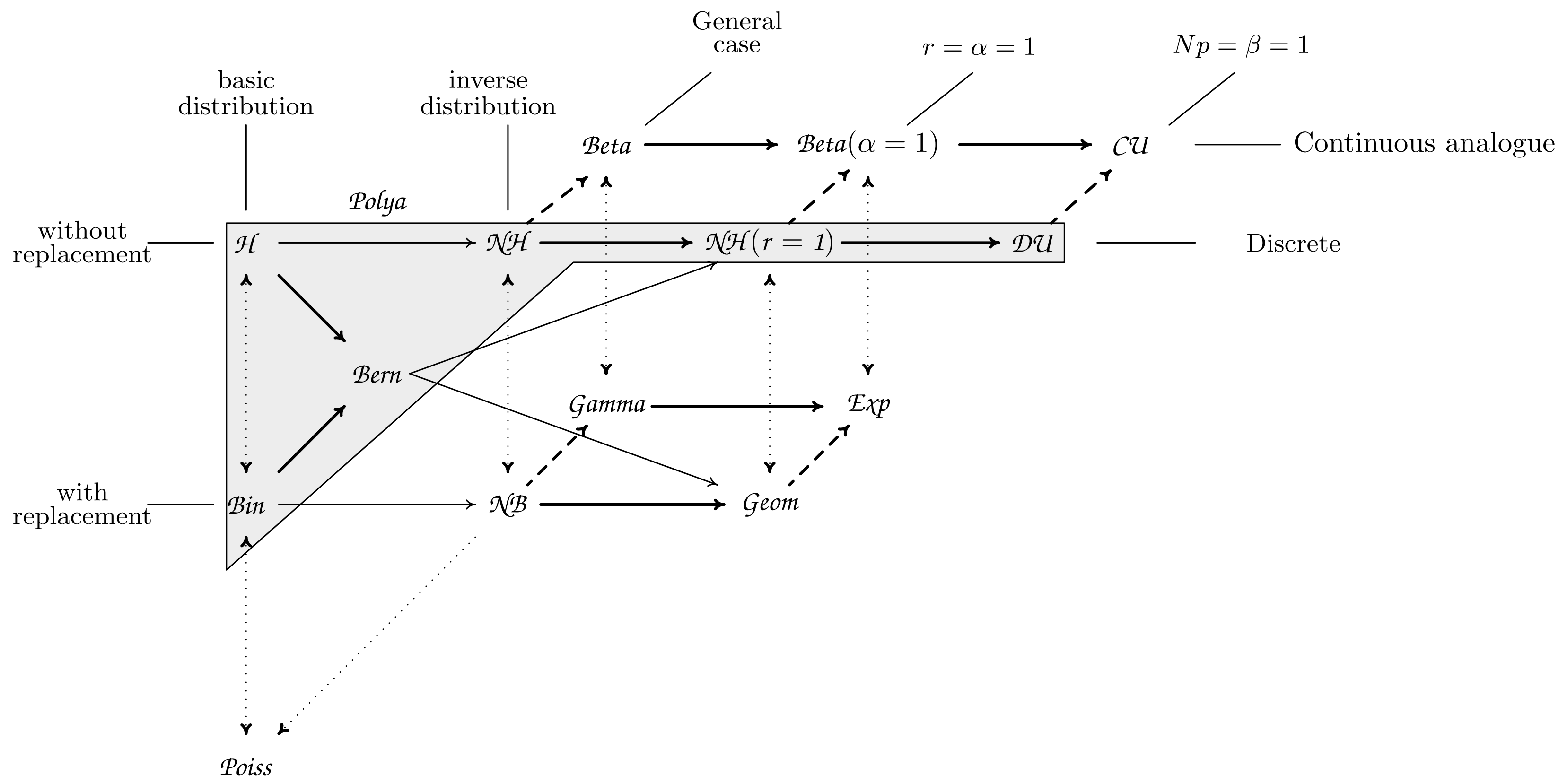

Figure 3: Summary of the relationships between the distributions. Thick arrows indicate a special case. Dashed arrows link the discrete with the continuous analogue distributions. Thin arrows link a distribution with an inverse (or negative) distribution. Dotted double arrows link the cases with and without replacement. In the reverse sense, they indicate convergence in distribution. Dotted arrows indicate convergence in distribution. The grey region contains the special cases of the Pólya–Eggenberger distribution.

Poisson with λ Gamma is negative binomial

If $X \sim \mathcal{Poiss}(\lambda)$, with $\gamma \sim \mathcal{Gamma}(r, (1-p)/p)$ is $\mathcal{NB}(r, p)$. Indeed,

$$\begin{aligned}
&\int_0^\infty \frac{\lambda^x}{x!} e^{-\lambda} \cdot \lambda^{r-1} \frac{e^{-\lambda(1-p)/p}}{\left(\frac{p}{1-p}\right)^r \Gamma(r)} \, \mathrm{d}\lambda \\
&= \frac{(1-p)^r p^{-r}}{x!\,\Gamma(r)} \int_0^\infty \lambda^{r+x-1} e^{-\lambda/p} \, \mathrm{d}\lambda \\
&= \frac{(1-p)^r p^{-r}}{x!\,\Gamma(r)} \, p^{r+x}\, \Gamma(r+x) = \frac{\Gamma(r+x)}{x!\,\Gamma(r)} \, p^x (1-p)^r.
\end{aligned}$$

References

Adell JA, de la Cal J (1994). "Approximating Gamma Distributions by Normalized Negative Binomial Distributions." *Journal of Applied Probability*, **31**(2), 391–400.

Bowman KO, Kastenbaum MA, Shenton LR (1992). "The Negative Hypergeometric Distribution and Estimation by Moments." *Communications in Statistics: Simulation and Computation*, **21**, 301–332.

Daley D, Vere-Jones D (2007). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Probability and Its Applications. Springer, New York.

David H, Nagaraja H (2004). *Order Statistics*. Wiley Series in Probability and Statistics. Wiley.

Eggenberger F, Pólya G (1923). "Über die Statistik Verketteter Vorgänge." *ZAMM – Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, **3**(4), 279–289.

Eggenberger F, Pólya G (1928). "Sur l'Interprétation de Certaines Courbes de Fréquence." *Comptes Rendus de l'Académie des Sciences, Paris*, **187**, 870–872.

Ferrari S, Cribari-Neto F (2004). "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics*, **31**(7), 799–815.

Forcina A, Franconi L (1988). "Regression Analysis with the Beta-binomial Distribution." *Rivista di Statistica Applicata*, **21**(1).

Johnson NL, Kemp AW, Kotz S (2005). *Univariate Discrete Distributions*. Wiley, New York.

Johnson NL, Kotz S, Balakrishnan N (2000). *Continuous Univariate Distributions. Volume 1 Models and applications*. Wiley, New York.

Kagan AM, Linnik YV, Rao CR (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York.

Kieschnick R, McCullough BD (2003). "Regression Analysis of Variates Observed on (0, 1): Percentages, Proportions and Fractions." *Statistical Modelling*, **3**(3), 193–213.

Leemis LM, Luckett DJ, Powell AG, Vermeer PE (2012). "Univariate Probability Distributions." *Journal of Statistics Education*, **20**(3), 1–11.

Leemis LM, McQueston JT (2008). "Univariate Distribution Relationships." *The American Statistician*, **62**(1), 45–53.

Likert R (1932). "A Technique for the Measurement of Attitudes." *Archives of Psychology.*

López-Blázquez F, Salamanca-Miño B (2001). "Exact and Approximated Relations between Negative Hypergeometric and Negative Binomial Probabilities." *Communications in Statistics – Theory Methods*, **30**(5), 957–967.

Mahmoud H (2008). *Pólya Urn Models.* Taylor & Francis, Hoboken.

Martin BD, Witten D, Willis AD (2020). "Modeling Microbial Abundances and Dysbiosis with Beta-binomial Regression." *The Annals of Applied Statistics*, **14**(1), 94.

Miller GK, Fridell SL (2007). "A Forgotten Discrete Distribution? Reviving the Negative Hypergeometric Model." *The American Statistician*, **61**(4), 347–350.

Morris CN (1982). "Natural Exponential Families with Quadratic Variance Functions." *The Annals of Statistics*, **10**, 65–80.

Morris CN (1983). "Natural Exponential Families with Quadratic Variance Functions: Statistical Theory." *The Annals of Statistics*, **11**, 515–529.

Morris CN, Lock KF (2009). "Unifying the Named Natural Exponential Families and Their Relatives." *The American Statistician*, **63**(3), 247–253.

Najera-Zuloaga J, Lee DJ, Arostegui I (2018). "Comparison of Beta-binomial Regression Model Approaches to Analyze Health-related Quality of Life Data." *Statistical Methods in Medical Research*, **27**(10), 2989–3009.

Patil GP, Seshadri V (1964). "Characterization Theorems for Some Univariate Probability Distributions." *Journal of the Royal Statistical Society*, **B26**, 286–292.

Särndal CE (1968). "Some Properties of the Negative Hypergeometric Distribution and Its Limit Distributions." *Metrika*, **13**(1), 170–190. ISSN 0026-1335.

Song WT (2005). "Relationships among Some Univariate Distributions." *IIE Transactions*, **37**(7), 651–656.

Terrell GR (1999). *Mathematical Statistics: A Unified Introduction.* Springer, New York.

Tillé Y, Wilhelm M, Qualité L (2018). "Sampling Designs from Finite Populations With Spreading Control Parameters." *Statistica Sinica*, **28**, 471–504.

Vitter JS (1987). "An Efficient Algorithm for Sequential Random Sampling." *ACM Transactions on Mathematical Software*, **13**, 58–67. ISSN 0098-3500.

Wilhelm M, Qualité L, Tillé Y (2017). "Quasi-Systematic Sampling from a Continuous Population." *Computational Statistics and Data Analysis*, **105**, 11–23.

Young DH (1970). "The Order Statistics of the Negative Binomial Distribution." *Biometrika*, **57**, 181–186. ISSN 0006-3444.

Affiliation:

Yves Tillé
Institute of Statistics
University of Neuchatel
Bellevaux, 51, CH-2000 Neuchatel, Switzerland
E-mail: yves.tille@unine.ch
URL: https://www.unine.ch/yves.tille

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 53 *Submitted:* 2023-08-08
2024 *Accepted:* 2023-12-21