Review of Applications of Urn Models, with Special Emphasis on Polya-Eggenberger Urn Models

Kanwar Sen* and Sonali Bhattacharya**

This paper attempts to review the urn model and its applications in various fields of engineering, science, social science, mathematics and finance. Some of the models and applications mentioned in this paper are recent and did not find mention in Feller (1968), Kotz and Balakrishnan (1997), and Balakrishnan and Koutras (2002). This paper would provide researchers in various fields with new ideas of modeling problems using the urn model.

Keywords: Urn model, Polya-Eggenberger sampling, Distribution of runs, Predetermined strategy

Introduction

The first mention of urn models seems to have been made by Bernoulli (1713), who in the third book of his *Ars Conjectandi* discussed the problem of drawing calculi out of urns. Bernoulli used the urn concept to model the underlying causes and observed the effects. It was first a quantitative attempt to construe a chance mechanism and only later the concept of lotteries, dice game and coin tosses were applied in explaining more specific kinds of problems. Stigler (1986, p. 124) made a beautiful treatment of Bernoulli's urn model to explain the Bayes' structure for treating the problem of Bernoulli directly. Other early evidences of the study of urn models are found in the works of De-Moivre (1667-1754), Mountmort (1678-1719) and Laplace (1749-1827), Quetelet (1794-1874), Ostrogadski (1801-1862), Poisson (1781-1840), Lexis (1837-1914) and Tschuprow (1905). Since the publication of the book *Urn Models and Their Applications* by Johnson and Kotz (1977), the theory and applications of urn models have received increased attention and intensive research among probabilists, statisticians and applied scientists alike. This paper discusses the significance of urn model as visualized by various statisticians in Section 2 and application of urn models in Section 3. Section 4 is devoted to the discussion of Polya-Eggenberger model, its generalizations, applications and modifications using predetermined strategy.

* Former Professor, Department of Statistics, University of Delhi, Delhi 110007, India. Email: kanwarsen2005@yahoo.com

** Associate Professor, Symbiosis Center of Management and Human Resource Development, Pune 411057, India; and is the corresponding author. E-mail: sonali_bhattacharya@scmhrd.edu

2. Significance of Urn Models

The significance of urn models is due to the easy accessibility in a majority of chance experiments, particularly those with a countable sample space. It is advantageous over other things like dice having six faces, card pack having 52 cards and chess board having 64 squares. Its significance is due to the following reasons:

- It is efficient in describing the concept of random choice which can be tested for *posteriori*.
- Urn models and chance experiments can easily be compounded into new hyperurn models corresponding to compounded experiments. Polya (1954) has displayed that random process such as the course and pattern of weather can be simulated as a sequence of run by using urn model.
- Even the term simulation can be interpreted as statistical equivalent to the basic concept of isomorphism which is intrinsically associated with urn model.
- Freudenthal (1960) helped in further clarifying the point: No statistician present at this moment will have been in doubt about the meaning of my words when I mentioned the common statistics model. It must be stochastic device producing random results. Tossing of coins or a dice or playing with cards are not flexible enough. The most general chance instrument is the urn filled with balls of different colors with tickets bearing some ciphers or letters. This model is continuously used in our courses as a didactic tool, and in our statistical analysis as a means of translating realistic problems into mathematical ones. In statistical language 'urn model' is a standard expression.

The significance of urn model can again be expressed in words of Feudenthal as below:

- The urn model is to be an expression of three postulates: (1) the constancy of a probability distribution, ensured by solidity of the vessel, (2) the random character of the choice, ensured by the narrowness of the mouth, which is to prevent visibility of the contents and any consciously selective choice, (3) the independence of successive choices, whenever the drawn balls are put into the urn. Of abstract probability and statistics, the word 'choice' can be avoided and all can be done without reference to such a model. But as soon as the abstract theory is applied, random choice plays an essential role.

The point is supported by Heitele (1975) and Polya (1954). Heilete remarked, "In principle it is possible to assign urn models with greater part of chance experiments, at least with those countable sample spaces." Polya asserted, "Any problem of probability appears comparable to a suitable problem about bags containing balls, and any random mass phenomenon appears as similar in certain essential respect to successive drawings of balls from a suitably combined bags."

Markov and Polya are among the first few who used urn models for deriving several probability distributions. A number of results in the probability theory have been derived by using different urn models. They have also been used theoretically to justify the Personian system of frequency curves and to explain modes of genesis for certain Langragian distributions (Consul, 1974; Consul and Mittal, 1975; and Janardhan and Schaeffer, 1977).

3. Applications of Urn Models

3.1 Application in Genetics

Mathematical models of population genetics can be considered equivalent to urn models, as genes in the population correspond to balls in the urn, and genetic type of a gene corresponds to the color of the ball. For example, consider the genetic model as follows: Let there be a population comprising a fixed number m of genes in any generation. Each gene is one out of two types A_1 and A_2. In other words, each generation will consist of i genes of type A_1 and $(m - i)$ genes of type A_2. Suppose the genetic composition of a daughter generation is derived by binomial sampling from the genes of parent generation, then in terms of urns and balls it is translated as follows:

Let there be an urn containing m balls, each ball being white or black. A second urn is then filled with m balls as follows. Balls are drawn one-by-one with replacement from the first urn. As each ball is drawn, a new ball of the same color is placed in the second urn. Then the probability that the second urn will have j white balls, given that the first urn contained i white balls (or the probability that in the daughter generation there will be j genes of type A_1 given that in the parent generation there are i such genes) is given by:

$$P_{ij} = \binom{m}{j}\left(\frac{i}{m}\right)^j\left(1-\frac{i}{m}\right)^{m-j} \qquad ...(1)$$

Limiting cases in genetics corresponds in the urn context to an infinite number of colors. A model of such a case is the following. Balls are drawn one by one with replacement from the parent urn. As each ball is drawn, with probability $(1 - u)$ a ball of the same color is placed in the daughter urn, whereas with probability u, a ball of entirely new color, not previously or currently seen during the history of the process, is placed in the daughter urn. The latter corresponds to a mutation in which an individual with completely new properties appears.

3.2 Estimation of Population Size

The classic capture-recapture models to estimate the size of natural population stem from the work of Schnabel (1938) and Chapman (1952), which discuss the estimation of the number of fish f in the lake. In terms of urns and balls, the model can be described as:

Start with an urn containing f indistinguishable white balls. Take n_0 balls from the urn (without replacement), dye each one black and return the black balls to the urn. Again take a random sample (without replacement) of n_0 balls and observe the number of T_1 of black balls. The aim is to estimate f, the number of balls originally in the urn. Witts *et al.* (1974) used a model which allows for random variation in the number of balls included in each sample which is not prespecified but is represented by a binomial random variable. Darling and Ribbons (1967) described the method with a stopping rule which is designed so that the proportional error in the estimation of sample size has a specified value. Binns (1976) described a sequential method to estimate f to avoid continuation of sampling in the capture-recapture model. In terms of urn models, the method can be explained as:

Suppose that f balls are distributed randomly among m urns. The urns are selected one at a time until the total number of balls in the first t chosen urns is at least $\phi\left(1-\frac{t}{m}\right)$, where ϕ is a constant, then $\tilde{f}=\frac{m}{t}$ (total number of balls when sampling ceases).

3.3 Learning Processes

Urn models serve as a useful tool at elementary level in determining the learning curves. Learning curves measure the performance of an individual as a function of training time or trials. The two elementary mathematical models of learning involving learning curves are:

1) The simple replacement model

2) The simple accumulative model

In replacement learning, some habits are replaced by others. The total number of entries is constant. In terms of an urn model, consider an urn U with m balls of two colors white and black. A white ball corresponds to a correct response and black ball to an error. According to the replacement idea, m is fixed and learning is regarded as a process of replacing black balls by white ones.

In the accumulation model, there are two urns U_1 and U_2. Here no balls are removed from the urn U_1 and a constant k of balls are transferred from urn U_2 to U_1 in each trial.

3.4 Military Application

In situations of military nature, targets are identified with urns, and missiles are identified with balls. Suppose that n missiles are aimed at the target in such a way that each target has an equal chance of being hit. Then, the number of targets that are hit or escape being hit has classical occupancy distribution. In 'air-battle theory' (or predator-prey models), bombers are identified with urns and fighters with balls; 'kill' parameters correspond to the probability that a fighter is able to intercept and destroy a bomber, and the number of occupied urns to the numbers of bombers destroyed.

3.5 Application in Computer Theory

Fagin (1975) provided application of urn model in the storage of data or instructions in computer memory. For example, computer's memory can be regarded as urn and pages of data as balls. Consider n balls numbered 1, 2, ..., n. The probabilities with which these numbers can be chosen are supposed to be $p_1, p_2, ..., p_n$ respectively. There is an urn that contains β_n of the balls (i.e., $n\beta_n$ balls in all). Initially, $n\beta_n$ balls are chosen at random with appropriate probabilities and placed in the urn and the ball that has been longest is removed from the urn. If a ball is already in the urn when its number is selected, there is no transfer of balls into or out of the urn, but the ball becomes the most recently selected and cannot be removed until at least $n\beta_n$ further selections have been made. Fagin (1975) considered the sum of selection probabilities for the number on the balls in the urn after n selections and called this sum the weight of the urn and obtained the expected value of the weight. For computer application, weight is the probability that a chosen page is already in the memory.

Yet another technique for organizing data in computer files, called 'perfect hashing', can be explained using the urn model. Consider random distribution of n balls in m urns where each urn has a maximum capacity of b balls. Each ball is randomly tossed so that the probability of a ball falling in an urn is $1/m$ and independent of other tossings. If an urn already contains b balls, any subsequent ball tossed into the urn is said to overflow. Let $P(n, m, b)$ denote the probability of a random distribution of n balls into m urns of size b resulting in no overflows. Let X be the random variable denoting the number of balls in the urn (or urns) containing the maximum number of balls. Then it is evident that $Pr\{X \leq b\} = P(n,m,b)$. Barton and David (1959) and David and Barton (1962) discussed a combinatorial extreme-value problem in this case. The exact computation of this probability distribution has been discussed by Monahan (1987) and Ramakrishna (1987).

An ingenious single urn model motivated by an imperfect debugging scheme, in which the ball represents flaws in the system was described by Siegrist (1987).

A single urn model is considered in which the number of balls in the urn at time $n + 1$ is determined as follows:

First, each ball in the urn at time n, independently of all others is removed with a probability $1 - p_{n+1} (0 \le p_{n+1} \le 1)$. The balls are first removed with probability that depends on the time value.

Next a random number of new balls are added to the urn, independently of the number of balls remaining after the first step. The equation given by Siegrist (1987) is:

$$X_{n+1} = (X_n - U_{n+1}) + V_{n+1} \qquad n = 0, 1, \ldots \qquad \ldots(2)$$

where X_n is the number of balls in the urn at time $n(n = 0, 1, \ldots)$ and $U_{n+1}(V_{n+1})$ is the number of balls removed (added) at time $n+1$. Evidently, $X_0 = V_0$ (the number of balls in the urn initially). Moreover, the distribution of $X_n - U_{n+1}$ given X_n is binomial with parameters n and p_{n+1}, V_{n+1} is independent of $X_n - U_{n+1}$ for $n = 0, 1, 2, \ldots$. This leads to the equation:

$$F_n(s) = \prod_{k=0}^{n} G_k \left(1 - (1 - s) \prod_{i=k+1}^{n} p_i \right) \qquad \ldots(3)$$

where $F_n(s) = E\left[s^{X_n}\right]$ and $G_n(s) = E\left[s^{V_n}\right]$ are the probability generating functions of X_n and V_n, respectively. The limiting cases of these distributions were also obtained.

Simha and Majumdar (1997) used the urn model to find probability distribution of number of distinct sites accessed by transaction in a distributed database and also probability distribution of the number of block accesses.

3.6 Application in Labor Market

Relationship between multiple applications of job seekers and vacancies in various firms was obtained using the urn model by Gautier (2002). Suppose there are v vacancies (urns) which are at disposal of u unemployed workers (balls). Each worker (ball) applies randomly to one vacancy (urn). The application process is undirected in the sense that any application is equally likely to go to any one of the vacancies (urns), i.e., each ball is randomly thrown into one of the urns. Next, one ball is drawn from each of these urns which will correspond to the worker who fetches the job in the firm (or application that has been processed). The rest are returned to workers as unsuccessful. It takes one period to process the chosen application.

3.7 Application in Biophysical and Pharmacological Modeling

In Pharmacological experimentation, a blood sample taken from the patient is distributed in compartments (urns) depending upon the combination of components. For example, To measure the total radioactivity in say 1 mL of blood (Asselin *et al.*, 2002), it is distributed between four physical compartments: the parent radioactivity in plasma (C_{1p}) and erythrocytes (C_{1e}), and similarly the metabolite radioactivity in plasma (C_{2p}) and erythrocytes (C_{2e}). Following a bolus injection, the parent radiotracer that is exchanged between plasma and erythrocytes with some rate constants, say k_1 and k_2, are transformed irreversibly (without replacement) into its main metabolite with rate constant k_3 which may be considered as balls.

Chattopadhyay (2004) used the urn model to develop a Randomized Play the Winner (RPW) rule for polychotomous treatments, for sequentially entering the patient. It was assumed that an urn contains a balls of two types A and B. When a patient enters a system, the patient is assigned a treatment by drawing a ball at random (with replacement). If the patient has a response in category j , we add an additional $(L-j)b$ balls of the same kind and $(j-1)b$ balls of the opposite kind in the urn. The rule for given (a, b) is denoted by RPWO (a, b), where 'O' stands for 'ordinal'. On an average, this rule allows more patients to be treated by better treatment in the course of decision making.

Biswas *et al.* (2006) designed Covariate Adjusted Continuous Drop the Loser (CCDL) rule for treatment of patients using the urn model. They assumed an urn having one ball each of type A, B, and I, where I is the immigration ball. For the $(i+1)^{st}$ entering patient, $i \geq 0$, we draw a ball from the urn, and treat the patient by treatment A or B if the drawn ball is of type A or B. On the other hand, if the drawn ball is of type I, one ball each of the types A and B is added to the urn, the ball of type I is replaced, and one ball is drawn from the urn afresh. This procedure is continued until a ball of A or B to treat the patient is drawn. Let the response of the patient be Y_{i+1}, the covariate vector is x_{i+1}, and the indicator of allocation is T_{i+1}. We then replace the drawn ball with a probability $p_{i+1} = p_{i+1}(Y_{i+1}, T_{i+1}, x_{i+1})$, which is also a function of all the accumulated data up to the first $(i+1)$ patients. The procedure is continued for the next entering patient.

3.8 Application in Portfolio Optimization

The urn based hidden Markov Model was used by Elliott and Hinz (2003) for explaining the optimum allocation of portfolio. Suppose there is a finite set S of urns containing balls of white and black colors of different sizes. The urns are hidden. In the first draw the urn is selected at random. In the first step a random machine is used for selection of an urn, a ball is drawn from it , its color and size are recorded and it is returned to the urn. In the next step, again the random device is used for the selection of the urn. A ball is drawn from it and shown. This process is repeated n times. The entire process generates a sequence $(x_k)_{k=0}^{n}$ of selected urns which is not observed, and an observed sequence $(y_k = (v_k, z_k))_{k=0}^{n}$ of ball color $(v_k)_{k=0}^{n}$ and ball size $(z_k)_{k=0}^{n}$. It can be used to interpret the portfolio optimization by assuming at any step $k \in \{1, 2, ..., n\} v_k$ instead of ball color represents the last observed price movement up and down and z_k represents the number of trades (or trade volume) between time period $\{\tau_{k-2}, \tau_{k-1}\}$. Let $\{S_t\}_{t\geq 0}$ is the asset price defined in the positive-valued continuous stochastic process defined in the probability space $\{\Omega, F, P\}$. A strategy is set up wherein trading occurs at predefined times $\tau_1 < \tau_2 < ... < \tau_n$ such that $\tau_{k+1} := inf\{t \geq \tau_k : S(t) \notin]\ d \cdot S(\tau_k), u \cdot S(\tau_k) [\}\ 0 < d < 1 < u$ are percentage price changes which trigger portfolio reallocation.

The optimal logarithmic portfolio $\pi^* = (\pi_k^*)_{k=0}^{n-1}$ is given by

$$\pi_k^* = \frac{P(v_{k+1} = u | I_k)}{1-d} = \frac{P(v_{k+1} = d | I_k)}{u-1} \text{ for all } k = 0, ..., n-1 \quad ...(4)$$

where d is the decision based on information I_k.

3.9 Modeling Credit Default Distribution

Amerio (2004) used reinforced urn model to design the distribution for credit default. Considering the Polya urn model in which the urn contains a white balls and b black balls. At every trial, a ball is drawn, its color is noted and it is returned to the urn with s additional balls before the next drawing. For simplicity, we discretize time, i.e., $T_1 ...; T_N \in S = \{u_0, u_1, u_2, ..., u_n, ...\}$ where $0 = u_0 < u_1 < u_2 < ... < u_n ...$; the meaning of observing $T_j = u_{i+1} \in S$ being that default of the j^{th} member of the Moody's rated class under scrutiny happened during the time interval $(u_i; u_{i+1}]$. Default in the future in a time period $(u_i; u_{i+1}]$ is based on the observations of defaults that happened in the past in the same time period. Moreover, the probability of default (non-default) in the time period $(u_i; u_{i+1}]$ is reinforced whenever a default (non-default) has been observed in the same time period. Therefore, it seems natural to control occurrences of defaults in each time period $(u_i; u_{i+1}]$ by means of different Polya's urns.

To each time point u_i a Polya's urn U_i initially containing w_i^0 white balls and b_i^0 black balls is assumed; with $w_0^0 = 0$; $b_i^0 > 0$; while w_i^0 and b_i^0 are nonnegative for $i = 1, 2, ...$. Reinforcement is set to be $s > 0$ for every urn. A process $\{X_n\}$ is iteratively defined on S in the following way: set $X_0 = u_0 = 0$: For $n \geq 1$; if $X_{n-1} = u_i$ we take a sample from the urn U_i: if the ball extracted is black, we set $X_n = u_{i+1}$ otherwise set $X_n = u_0$: (Recall that urn U_i's composition is updated each time the urn is sampled by the rule which adds in it m balls of the same color as that of the sampled ball). The sequence $\{X_n\}$ is a RUP. A less formal description of the process $\{X_n\}$ is in order: starting from urn U_0, we sequentially take a sample from urns U_0, U_1, U_2, ... until a white ball is produced; at that time we move back to urn U_0 and we start the urns sequential sampling once again.

3.10 Birthday and Coupon Collection Problem

Inoue and Aki (2008) assumed random supply of n balls into m distinguishable urns and with probability for any ball to drop into the i^{th} urn is p_i, $i = 1, 2, \ldots, m$. The two probabilities, which are closely related to a generalization of the well-known combinatorial and probabilistic problems are: (1) the probability that at least one urn contains r or more balls, (2) the probability that all urns contain r or more balls. The Problems (1) and (2) have been named as birthday problem and coupon collector's problem, respectively.

Thus if X_n^i denotes the number of balls in the i^{th} urn and F_{n,r_i}^i denotes that i^{th} urn contains r_i or more balls $F_{n,r_i}^i = P(X_n^i \geq r_i)$, then the classical birthday problem can be defined by the probability distribution $a_n(r) = P(X_n^1 > r_1 \text{ or } X_n^2 > r_2, \text{or} \ldots, X_n^m > r_m)$, where $r = (r_1, r_2, \ldots, r_m)$ and coupon collection problem is given as $d_n(r) =$ $P(X_n^1 > r_1, X_n^2 > r_2, \ldots, X_n^m > r_m)$. Other researchers who have worked on similar problems are Feller (1968), Flajolet *et al.* (1992), Diaconis and Holmes (2002), and Charalambides (2005).

3.11 Statistical and Mathematical Application

An interesting urn model with its applications to modeling outliers was given by Small (1985) wherein an urn containing c white and c black balls is considered. If on trial m, a white ball is selected, set $U_m = 1$. If it is black, set $U_m = 2$ before the $(m + 1)^{th}$ selection, a ball of opposite color that of m^{th} trial is put in the urn. The sequence $U_1, U_2, \ldots$ can be continued indefinitely. Now $S_{ni} = \{m: U_m = i, m \leq n\}$ for $i = 1, 2, \ldots$. Small (1985) allowed that such urn models can be used to model outliers in data, wherein the behavior of outliers is governed by quantities S_{ni}.

Paik (1983) used an urn model construction to clarify some paradoxes associated with the concept of infinity. Suppose on day 1, 10 balls numbered 1 to 10 are put in an urn and a ball numbered n_1 is withdrawn. On day 2, 10 more balls numbered 11-20 are put in the urn and the ball numbered n_2 is withdrawn. The problem is: what would be the number of balls after the process is continued for infinitely large number of times. The paradox is that in the experiment where $n = 10i$, the answer would be infinitely many, where $n = i$, the answer would be none. The difference in the result is disturbing due to the singular behavior of a function similar to the sequence of functions on the positive real line.

$$f_n(x) = \begin{cases} 1 & \text{if } x \in [n+1, 10n] \\ 0 & \text{otherwise} \end{cases} \quad \ldots(5)$$

Hence, the limit of the integral $\int_0^\infty f_n(x)\,dx$ is ∞, but the integral of the limit $f(x) = \lim_{n\to\infty} f_n(x)$ is 0. Urn model interpretation of Paik (1983) gives more precise interpretation of this breakdown which occurs at ∞.

Urn model found its application in describing the order statistics wherein the urn containing N balls numbered $\alpha_1 < \alpha_2 < \ldots < \alpha_N$ is considered, a sample of n balls are being drawn from it without replacement, and X_i's (i = 1, 2, ..., n) be the number of balls, and $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be the arrangement of chosen variables while X_i's be the random variables better known as 'order statistics'.

Urn models have also found applications in describing Markov Chain, wherein N urns marked 1, 2, ..., N are considered containing balls marked i (i = 1, 2, ..., N). Proportion of balls marked j in i^{th} urn be $p_{ij}, \sum_{j=1}^{n} p_{ij} = 1$. An urn marked say U_1 is randomly chosen, a ball (marked say α_1) is drawn from it randomly, then the urn marked α_1 is selected and a ball is drawn from it, and let us suppose it is marked α_2, then the urn marked α_2 is selected and the process is continued. Then U_1, α_1, α_2 forms a Markov-Chain.

4. Distributions Based on Urn Models

An urn model is constructed by imagining a number of urns, containing balls of one or more colors. Then balls are drawn from the urns according to certain rules. These rules may include the addition of balls from certain rules. These rules may include the addition of balls to or removal of balls from certain urns at various stages of the trial. Some of the rules for an urn containing two colors of balls are mentioned here.

At each stage, s' balls of the opposite color that are chosen are added (as well as s balls of the same color) s and s' may depend on the color of the chosen ball.

s (or s') may be negative, e.g., if $s = -1$ and $s' = 0$, we have sampling without replacement.

The values of s and s' may vary with the number of trials, i.e., we have a prespecified value st and st' on the t^{th} trial. The st and st' may be random variables.

For urns containing more than two colors of balls, Rules (1) through (5) can be modified. The urn model with more than two colors of balls leads to multivariate distributions.

By considering the limiting cases as certain parameters (e.g., number of runs, proportion of balls of certain color, number of trials) are varied, a number of continuous distributions are regarded as approximations to discrete ones (Bartlett, 1937).

Among the principal investigators who used urn models for developing probability models of contagion of certain events, Markov (1917) and Eggenberger and Polya (1923) are pioneers in the field. The urn model considered is widely known as Polya-Eggenberger (P-E) model by some authors like Janardhan and Schaeffer (1977).

Polya-Eggenberger Model is a single urn model with balls of two colors, i.e., a white and b black balls. A ball is drawn from the urn and then replaced along with s balls of the same color. Then two types of sampling schemes have been defined:

1. **Direct or Binomial Sampling Scheme:** A sample of fixed size n is drawn, so that if X represents the number of white balls in n drawings, then the distribution of the random variable X when the procedure of drawing of balls is repeated n (fixed) times is given by:

$$P(X=x)=\binom{n}{x}\frac{a^{(x,s)}b^{(n-x,s)}}{(a+b)^{(n,s)}},$$

$$x=0,\ 1,\ 2,\ \ldots,\ n \qquad \ldots(6)$$

is the P-E distribution which yields

Binomial distribution for $s = 0$

Hypergeometric distribution for $s = -1$

Discrete rectangular distribution for $s = a = b$

2. **Inverse Binomial Sampling Scheme:** Further, if X = number of white balls preceding n^{th} black ball, the waiting time distribution is given by:

$$P(X=x)=\binom{n+x-1}{x}\frac{a^{(x,s)}b^{(n,s)}}{(a+b)^{(n+x,s)}},$$

$$x=0,\ 1,\ 2,\ \ldots \qquad \ldots(7)$$

is known as the Inverse Polya-Eggenberger Sampling Scheme.

4.1 Key Developments on Polya-Eggenberger Model

Friedman (1949) generalized the P-E model by generalizing the P-E sampling scheme. In Friedman's urn model, each ball is returned together with s balls of the same color and s balls of the opposite color. He noted the following special cases:

$s = 0$, $s' = 0$ gives binomial model,

$s = -1$, $s' = 0$ gives sampling without replacement,

$s' = 0$ gives safety campaign model in which each draw of a safety campaign model in which each draw of a white ball is penalized, and

$s = -1, s' = 1$ gives the Ehrenfest Model of Heat Exchange.

A limiting case of the distribution when the limit $n \to \infty, \dfrac{a}{a+b} \to 0, \dfrac{s}{a+b} \to 0$ so that $\dfrac{na}{a+b} \to \lambda, \dfrac{s}{a+b} \to \dfrac{1}{\theta}$, was applied by Polya for explaining the epidemic of smallpox, and Newbold (1927) for accident analysis when certain people have higher sustainability under uniform conditions of risks.

Woodbury (1949) studied a quite different model in which $s = 0, s' = 0$ if a white ball appears, and $s' = -s > 0$ if a black ball appears. This means that the drawing of a white uninfected ball has no effect on the composition of the urn, but a black (infected) ball causes s uninfected ball to become infected.

Naor's (1957) urn model was studied in the context of machine-minding problem. Naor considered an urn containing n balls of which one is white and remaining are black. Sampling is continued until a white ball is obtained, and every time black ball is drawn it is replaced with a white ball.

Kotz and Balakrishnan (1997) gave a review of key developments that have occurred pertaining to two urn models, with special emphasis on P-E model in probabilistic, statistical and biological literature during the preceding two decades.

4.2 Generalized Polya-Eggenberger Model

Sen and Mishra (1996) obtained a generalized P-E model by using a usual urn model with a white and b black balls based on the sampling scheme which was termed as 'unified sampling scheme'. Let X_i be the number of white balls in the first i balls and Y_i be number of black balls in the first i balls and the event A as

$$A = \left[X_{n+(\mu+1)x} = x;\ Y_i < n + \mu X_i \left(i = 1, 2, \ldots, n + (\mu+1)x - 1\right); \quad Y_{n+(\mu+1)x} = n + \mu x \right] \qquad \ldots(8)$$

where n and μ are the given integers. Sampling scheme used is as follows. The balls are drawn from the urn one-by-one at random by P-E sampling scheme until conditions of the event A are satisfied.

By using the above sampling scheme, Mohanty (1981) and Sen and Mishra (1996) obtained the following probability model (by using combinatorial methods):

$$P_x = \frac{n}{n+(\mu+1)x} \binom{n+(\mu+1)x}{x} \frac{a^{(x,s)}\, b^{(n+\mu x,s)}}{(a+b)^{(n+(\mu+1)x,s)}}; \quad x = 0, 1, 2, \ldots \qquad \ldots(9)$$

where $(n > 0)$, $s(\geq -1)$ and $\mu(\geq -1)$ are integers.

This model is a generalized P-E model which generates many discrete distributions for various values of n, μ, s, a, b and including P-E distribution for $\mu = -1$ and the inverse P-E distribution for $\mu = 0$.

In other words, the sampling scheme used by Mohanty (1981) and Sen and Mishra (1996) unifies the binomial and inverse binomial sampling schemes with the parameter μ and thus was called the 'unified sampling scheme'.

Chung *et al*. (2003) considered finitely many bins each containing one ball, additional balls are allowed to arrive one at a time (Table 1). For each new ball, with probability p, a new bin is created and the ball is placed in that bin; with probability $1 - p$, or in an existing bin, such that the probability the ball is placed in a bin is proportional to m^g, where m is the number of balls in that bin. The model leads to the following distributions: f_i is the limit of the fraction of bins with i balls and

$$K = \frac{p}{1-p} \sum_{i=1}^{\infty} f_i i^{\gamma}$$

Table 1: Particular Cases of Chung *et al*.'s (2003) Generalized Polya-Eggenberger Model

	Finite Polya Process $p = 0$	Infinite Polya Process $0 < p < 1$	
$\gamma = 1$	One bin dominates	One bin dominates	
$\gamma = 1$	Polya's urn model	Power law distribution	$f_i \alpha\, i^{(-1+1/(1-p))}$
$0 < \gamma < 1$	All the bins grow at the same rate asymptotically	Exponentially decreasing	$f_i \alpha\, i^{-\gamma} e^{-Ki^{1-\gamma/(1-\gamma)}}$
$\gamma = 0$			$f_i \alpha (K+1)^{-i}$
$\gamma < 0$			$f_i = 0(((i-1)!)^{\gamma} / K^{i})$

4.3 Polya-Eggenberger Model for Distribution of Runs

1. Panaretos and Xekalaki (1986) defined X as the number of non-overlapping runs of white balls of length k in the sample size n, and then the probability distribution of X is given as:

$$P(X = x) = \sum_{m=0}^{k-1} \sum\nolimits_1 \binom{\sum_{i=1}^{k} x_i + x}{x_1, x_2, \ldots, x_k, x} \frac{b^{\left(\sum_{i=1}^{k} x_i, s\right)} a^{\left(n - \sum_{i=1}^{k} x_i, s\right)}}{(a+b)^{(n,s)}} . \; 0 \le x \le \left[\frac{n}{k}\right] \qquad \ldots(10)$$

where the inner summation Σ_1 is overall non-negative integers $x_1, x_2, \ldots, x_k$ such that

$$x_1 + 2x_2 + \ldots + kx_k = n - m - kx \qquad \ldots(11)$$

It yields Binomial distribution of order k (Philippou and Makri, 1986) and for $s = 0, \dfrac{a}{a+b} = p$, it yields hypergeometric distribution of order k (Godbole, 1990) for $s = -1$.

Under the same sampling scheme, the distribution of waiting time for n-overlapping white ball runs of length k was obtained by Panaretos and Xekalaki (1986)

$$P(X = x) = \sum \binom{\sum_{i=1}^{k} x_i + r - 1}{x_1, x_2, \ldots, x_k, r-1} \frac{b^{\left(\sum_{i=1}^{k} x_i, s\right)} a^{\left(x + kr - \sum_{i=1}^{k} x_i, s\right)}}{(a+b)^{(x+kr, s)}} \qquad \ldots(12)$$

where the inner summation Σ is overall non-negative integers $x_1, x_2, \ldots, x_k$ such that

$$x_1 + 2x_2 + \ldots + kx_k = x \qquad \ldots(13)$$

which generalizes the negative binomial distribution of order k (Philippou *et al.*, 1983) and waiting time hypergeometric distribution of order k (Godbole, 1990). Sen *et al.* (2002a and 2002b) used lattice path approach to give unified models for distributions of number of non-overlapping success runs of length k in a finite sample of size n, overlapping success runs of length k and their corresponding waiting time distributions of success. The Type-II P-E Distribution of the order k based on overlapping success run of length k (Sen *et al.*, 2002b) is given as

$$P_x^{(k,II)} = \sum_{m \geq \gamma} \sum_{r=0}^{\min\{m - x(k-1), n + \mu m + 1\}} \Sigma_{II} \frac{a^{(m,s)} b^{(n + \mu m, s)}}{(a+b)^{(n + \overline{\mu + 1} m, s)}} \qquad \ldots(14)$$

where the inner summation is over Σ_{II} is over $L : (l_1, l_2, \ldots, l_r, L_1, \ldots, L_r, L_{r+1})$ such that

(a) $(1-\delta)L_1 < n$

$\beta L_1 + (1-\delta)L_2 < n + \mu l_1$

$L_1 + \beta L_2 + (1-\delta)L_3 < n + \mu(l_1 + l_2)$

$\vdots \quad \vdots \quad \vdots$

$L_1 + L_2 + \ldots + \beta L_{r-1} + (1-\delta)L_r < n + \mu(l_1 + l_2 + \ldots + l_{r-1})$

(b) $L_1 + L_2 + \ldots + \beta L_r + (1-\delta) L_{r+1} < n + \mu (l_1 + l_2 + \ldots + l_r)$

(c) $\sum_{i=1}^{r} l_i = m$

(d) $\max(0, l_1 - k + 1) + \max(0, l_2 - k + 1) + \ldots + \max(0, l_r - k + 1) = x$

and $\delta = \begin{cases} 1 & \text{if } X_1 = 1 \\ 0 & \text{otherwise} \end{cases}$

$$\delta = \begin{cases} 0 & \text{if } X_1 = 1 \text{ and the last ball drawn is white} \\ 1 & \text{otherwise} \end{cases}$$

$$\gamma = \gamma(x) = \begin{cases} 0 & \text{if } x = 0 \\ x + k - 1 & \text{if } x > 0 \end{cases} \qquad \ldots(15)$$

which yield Type-II Polya distribution of order k and Type-II Inverse Polya Distribution of order k as particular case.

2. Sen *et al*. (2003) obtained circular distributions of order k based on non-overlapping success runs, overlapping success runs, exact success run of length k and of at least length k, longest and shortest run of and their corresponding waiting time distributions and their joint distribution by using P-E sampling scheme. So, if $N_{n,k}^{(C)}$, $M_{n,k}^{(C)}$, $G_{n,k}^{(C)}$ and $E_{n,k}^{(C)}$ represent respectively the number of non-overlapping success runs of length k, overlapping success runs of length k, success runs of at least length k and success runs of exact length k and their joint pdf is given as:

$$P_{n,x,y,u,v}^{(k)} = \sum_{m=uk+u-v}^{\min(n-u+1,n-2)} \sum_{c=0}^{\min(x,x+u-1)} \sum_{d=0}^{\min(x,x-u+1)-c}$$

$$\left[(1-\delta_{uv}) + \delta_{uv} \sum_{i=0}^{\min(k-1,m-vk-(u-v-1)(k+1)-(c+d)k)} \sum_{j=0}^{\min(k-1,m-vk-(u-v-1)(k+1)-(c+d)k)-i} (1) \right] \times$$

$$\sum_1 \binom{n-m-1}{n-m-1-u+\varepsilon_{ij}, v-\pi_{ij}, x_{m-(c+d+u-\varepsilon_{ij})k-i-j-(u-\varepsilon_{ij}-v+\pi_{ij})+\delta_{uv}}} \times$$

$$\left[\sum_{l=0}^{\left[\frac{m-k(u-\varepsilon_{ij})-y+((c+d-1)k+i+j)+u-\varepsilon_{ij}}{k}\right]} (-i)^j \binom{n-m-1-u+\varepsilon_{ij}}{l} \times \right.$$

$$\left.\left.\binom{n-k\left(u-\varepsilon_{ij}\right)-y+\left(\left(c+d-1\right)k+i+j\right)-lk-2}{m-k\left(u-\varepsilon_{ij}\right)-y+\left(\left(c+d-1\right)k+i+j\right)+u+\varepsilon_{ij}-lk}\right]\right]\times$$

$$\frac{a^{(m,s)}b^{(n-m,s)}}{(a+b)^{(n,s)}}+\beta_1\frac{b^{(n,s)}}{(a+b)^{(n,s)}}+n\beta_2\frac{ab^{(n-1,s)}}{(a+b)^{(n,s)}}$$

$$0\le v\prec u\le x\le y\le n-k+1 \qquad ...(16)$$

where the inner summation Σ_1 is over $x_1, x_2, \ldots x_{m-u(k+1)+v+1}$ such that:

$$x_1+x_2+\ldots+x_{m-\left(c+d+u-\varepsilon_{ij}\right)k-i-j-\left(u-\varepsilon_{ij}-v+\pi_{ij}\right)}+\delta_{uv}=u-\varepsilon_{ij}-v+\pi_{ij}$$

$$\left[\frac{k+1}{k}\right]x_1+\left[\frac{k+2}{k}\right]x_2+\ldots+\left[\frac{m-\left(c+d+u-\varepsilon_{ij}+1\right)k-i-j-\left(u-\varepsilon_{ij}-v+\pi_{ij}\right)+\delta_{uv}}{k}\right].$$

$$\left[x_m-\left(c+d+u-\varepsilon_{ij}\right)k-i-j-\left(u-\varepsilon_{ij}-v+\pi_{ij}\right)+\delta_{uv}\right]=x-\lambda_{ij}-v+\pi_{ij} \qquad ...(17)$$

$$2x_1+3x_2+\ldots+\left(m-\left(c+d+u-\varepsilon_{ij}\right)k-i-j-\left(u-\varepsilon_{ij}-v+\pi_{ij}\right)+\delta_{uv}+1\right).$$

$$\left(x_m-\left(c+d+u-\varepsilon_{ij}\right)k-i-j-\left(u-\varepsilon_{ij}-v+\pi_{ij}\right)+\delta_{uv}\right)=y-\mu_{ij}-v+\pi_{ij} \qquad ...(18)$$

$$\pi_{ij}=\begin{cases} & \text{if } c=1, d=0 \text{ and } i=j=0\\ 1 & \text{if } c=0, d=1 \text{ and } i=j=0\\ & \text{if } c=0, d=0 \text{ and } i+j=k\\ & \\ 0 & \text{otherwise}\end{cases} \qquad ...(19)$$

$$\varepsilon_{ij}=\begin{cases} & \text{if } c\ge 1\\ 1 & \text{or if } d\ge 1, \text{ and } i=j=0\\ & \text{or if } c=0, d=0 \text{ and } i+j\ge k\\ & \\ 0 & \text{otherwise}\end{cases} \qquad ...(20)$$

$$\mu_{ij}=\begin{cases}\left(c+d-1\right)k+i+j+1 & \text{if } \varepsilon_{ij}=1\\ & \\ 0 & \text{otherwise}\end{cases} \qquad ...(21)$$

$$\beta_1 = \begin{cases} 1 & \text{if } x = \left[\frac{n}{k}\right],\ y = n-k+1 \text{ and } u = v-1, \text{ given } k = n \\ & \text{or } \ x = \left[\frac{n}{k}\right],\ y = n-k+1 \text{ and } u = 1,\ v = 0, \text{ given } k < n \\ 0 & \text{otherwise} \end{cases} \quad ...(22)$$

$$\beta_2 = \begin{cases} 1 & \text{if } x = \left[\frac{n-1}{k}\right],\ y = n-k \text{ and } u = v = 1, \text{ given } k = n-1 \\ & \text{or } x = \left[\frac{n-1}{k}\right],\ y = n-k \text{ and } u = 1,\ v = 0, \text{ given } k < n-1 \\ 0 & \text{otherwise} \end{cases} \quad ...(23)$$

$$\delta_{uv} = \begin{cases} 0 & \text{if } u = v \\ 1 & \text{otherwise} \end{cases} \quad ...(24)$$

Sen *et al*. (2003) also obtained the probability function of marginal and joint distributions of 'inverse circular distribution' of order k. Sen *et al*. (2006) extended the study to obtain joint distributions of success and failure runs of length (k_1, k_2) and its corresponding sooner or later waiting time distributions.

3. Makri *et al*. (2007a) used P-E sampling scheme to obtain the distribution of the sum of the lengths of the success runs (i.e., the total number of successes in all the success runs) of length greater than or equal to a prespecified length, and the waiting time that the above-mentioned statistic equals or exceeds a predetermined level. Makri *et al.* (2007b) also obtained Polya, inverse Polya, circular Polya distributions of l-overlapping success runs which computes number of $(l - 1)$-overlapping occurrences of success runs of length k until the n^{th} overlapping occurrence of a success run of length l. Erylmaz (2008) used urn model with multicolor scheme to obtain joint distributions of runs. In this multicolor urn scheme, a ball is drawn from the urn initially containing m_j balls of color j, $j = 1, 2, \ldots, t$ and its color is noted. If a ball of color j is drawn at a stage, s balls of color j, $j = 1, 2, \ldots, t$ are added to the urn. Drawing a ball of color j is considered as a trial of type j, $j = 1, 2, \ldots, t$. This scheme is repeated n times and a sequence consisting of trials, namely $\{1, 2, \ldots, t\}$, is derived. Let

$R_n^{(j)}$ be the total number of runs of type j,

$\theta_n^{(j)}$ be the length of the i^{th} run of type j,

$E_{n,k_j}^{(j)}$ be the total number of runs of type j with length exactly equal to k_j,

$G_{n,k_j}^{(j)}$ be the total number of runs of type j with length at least k_j,

$L_n^{(j)}$ be the length of the longest run of type j, and

L_n be the length of the longest run of any type in $Z_1, Z_2, \ldots, Z_n$

The joint probability function of $E_{n,k_1}^{(1)}, E_{n,k_2}^{(2)}, \ldots, E_{n,k_t}^{(t)}$ was given as:

$$P\left(E_{n,k_1}^{(1)} = x_1, E_{n,k_2}^{(2)} = x_2, \ldots, E_{n,k_t}^{(t)} = x_t\right) = \sum_r \sum_n \prod_{m=1}^{t} \binom{r_m}{x_m}$$

$$A\left(n_m - r_m - x_m\left(k_m - 1\right), r_m - x_m, x_{m,} - 1\right) \; F_t\left(r_1, r_2, \ldots, r_t\right) p_n\left(n_1, n_2, \ldots, n_t\right)$$

where $A\left(\alpha, r, k\right) = \sum_{j=0}^{[\alpha/k]} (-1)^j \binom{r}{j} \binom{\alpha - (k+1)j + r - 1}{\alpha - jk}$

$$F_t\left(r_1, r_2, \ldots, r_t\right) = (-1)^r \sum_{m_1=1}^{r_1} \sum_{m_2=1}^{r_2} \ldots \sum_{m_t}^{r_t} (-1)^m \binom{r_1 - 1}{m_1 - 1} \ldots \binom{r_t - 1}{m_t - 1} \binom{m}{m_1, m_2, \ldots, m_t},$$

$$p_n\left(n_1, n_2, \ldots, n_t\right) = \frac{m_1^{(n_1)} m_2^{(n_2)} \ldots m_t^{(n_t)}}{m^{(n_1)} \left(m - n_1\right)^{(n_2)} \ldots \left(m - \sum_{i=1}^{t-1} n_i\right)^{(n_t)}}$$

$$r = \sum_{r=1}^{t} r_i \leq n, \;\; m = \sum_{r=1}^{t} m_i \;\text{ and }\; x^{(a)} = x(x-1)\ldots(x-a+1) \qquad \ldots(25)$$

Erylmaz (2008) also gave joint distribution of $G_{n,k_1}^{(1)}, G_{n,k_2}^{(2)}, \ldots, G_{n,k_t}^{(t)}$

4. The applications of discrete distributions of runs can be found in computation of reliability of consecutive k-out-of n: F system (Aki and Hirano, 1996), a consecutive k-out-of-r-from n: F system (Griffith, 1986; Papastavridis and Sfakianakis, 1991; Sfakianakis *et al.*, 1992; Papastavridis and Koutras, 1993; and Cai, 1994), in start-up demonstration test (Hahn and Gage, 1983), in molecular biology (Goldstein, 1990; and Huang and Tsai, 1991), in the study of theory of radar detection, time sharing systems and quality control (Greenberg, 1970; Nelson, 1978; Saparstein, 1973; Mirstik, 1978; and

Glaz, 1983) and Statistical Hypothesis Testing (Koutras and Alexandrou, 1997). Applications of runs and scans can be found in a more detailed way in Balakrishnan and Koutras (2002).

4.4 Distributions Based on Urn Models with Predetermined Strategies

Janardhan (1973) and Consul (1974) seem to be the pioneers in introducing urn models based on predetermined strategies.

Consul (1974) and Consul and Mittal (1975) described a two urn model and a four urn model respectively which are based on the strategy of the player and obtained distributions defined as quasi-binomial and quasi-Polya distribution using direct sampling scheme. Compositions of the urn for both the models are given below, where t and n are known positive integers, but x is chosen by the person to decide the strategy (Table 2).

Table 2: Urn Model with Predetermined Strategy

Pre-Strategy			Post-Strategy		
Urns	Balls		Urns	Balls	
	White	Black		White	Black
Two Urn Model					
1	a	–	1	A	Xt
2	a	B	2	$a + xt$	$b + (n - x)t$
Four Urn Model					
1	a	B	1	a	b
2	–	B	2	$(n - x)t$	b
3	a	–	3	a	xt
4	a	B	4	$a + xt$	$b + (n - x)t$

Janardhan (1975) showed Markov-Polya models with predetermined strategy with direct and inverse in both univariate and multivariate situations, which provide new probability distributions. He developed quasi-Polya and quasi-multivariate Polya distributions based on the urn models with predetermined strategy using direct sampling scheme. Janardhan gave recurrence relations for probabilities as well as expressions for the mean. He considered certain special and limiting cases. Quasi inverse Polya distributions were obtained by using inverse sampling scheme.

Consul and Mittal (1977) considered a $(s + 2)$ urn model with balls of s different colours where the strategy of the player alters the contents of $(s - 1)$ urns and hence the probability of success. The model provides a quasi-multinomial model.

Janardhan and Schaeffer (1977) and Janardhan (1978) introduced new urn models dependent upon a predetermined strategy for the development of probability model for voting in small groups where contagion is present within each group and group leader devises some new strategies for bringing success to candidates. Janardhan (1978) obtained Markov-Polya distributions along with recurrence relation for the probabilities as well as considered the special and limiting distributions and maximum likelihood estimations. Rao and Janardhan (1984) studied the use of generalized Markov-Polya distributions as random damage model.

Mishra *et al*. (1992) mentioned the moments of Quasi-Binomial Distributions (QBD) appear in terms of series which cannot be summed up easily. The method of moments thus fail to provide quick estimates of parameters involved. They defied a class of QBDs by using three urn model with strategy, which includes a number of QBDs in addition to Consul's (1974) QBD and Consul and Mittal's (1975) QBD.

Sen and Jain (1996) obtained three generalized urn models with predetermined strategy based on the unified sampling scheme, which were called Generalized Markov-Polya (GMP) models. The GMP model-I was obtained by using two urn model which yielded quasi-Polya and inverse Polya distributions and all particular cases of Janardhan (1975) and Consul (1974). The GMP Model-II was obtained by using four-urn model which verified the result of Consul and Mittal (1975) and Janardhan and Schaeffer (1977). Both these models yielded the generalized P-E model of Mohanty (1981) and Sen and Mishra (1996). The GMP model-III unified GMP model-I and model-II. It also yielded the class of quasi-binomial distribution. ✌

References

1. Aki S and Hirano K (1996), "Lifetime Distribution and Estimation Problems of Consecutive-*k*-out-of *n*: F System", *Annals of the Institute of Statistical Mathematics*, Vol. 48, No. 1, pp. 185-199.

2. Amerio E (2004), "Reinforced Urn Processes for Modeling Credit Default Distribution", *International Journal of Theoretical Applied Finance*, Vol. 7, No. 4, pp. 407-423.

3. Asselin M C, Wahl L M, Amano S and Nahmias C (2002), "Vivo Metabolism and Partitioning of Fluoro-L-Meta-Tyrosine in Whole Blood: A Unified Compartment Model", *Physics in Medicine and Biology,* Vol. 47, No. 11, pp. 1961-1977.

4. Balakrishnan N and Koutras M V (2002), *Runs and Scans with Applications*, 1st Edition, John Wiley, New York.

5. Bartlett M S (1937), "Natural", *Mathematics, Math. Gaz.*, Vol. 21, pp. 44-45.

6. Barton D E and David F N (1959), "Combinatorial Extreme Value Distributions", *Mathematika*, Vol. 6, pp. 153-161.

7. Bernoulli J (1713), *Arks Conjectandi*, Impensis Thernisiorm, Fratran, Balsileue.

8. Binns M R (1976), "A Sequential Counting Procedure for Estimating the Total Number of Randomly Distributed Individuals", *J. Am. Statist. Assoc.*, Vol. 71, No. 353, pp. 74-79.

9. Biswas A, Huang H-H and Huang W-T (2006), "Covariate-Adjusted Adaptive Designs for Continuous Responses in a Phase III Clinical Trial: Recommendation for Practice", *J. of Biopharm. Statist.,* Vol. 16, No. 2, pp. 227-239.

10. Cai J (1994), "Reliability of a Large Consecutive-*k*-out-of *r* from *n*: F System with Unequal Component Reliability", *IEEE Transaction on Reliability*, Vol. 43, No. 1, pp. 107-111.

11. Chapman D G (1952), "Inverse Multiple and Sequential Sample Censuses", *Biometric*s, Vol. 8, No. 4, pp. 286-306.

12. Charalambides Ch A (2005), *Combinatorial Methods in Discrete Distributions*, pp. 416+xiv, John Wiley & Sons, Hoboken, New Jersey.

13. Chattopadhyay G (2004), "Randomized Play-the-Winner Rule for Ordered Categorical Data", *Commun. Statist-Theor. and Methd.*, Vol. 33, No. 11, pp. 2683-2708.

14. Chung F, Handjani S and Jungreis D (2003), "Generalisations of Polya's Urn Problem", *Ann. Combin.*, Vol. 7, No. 2, pp. 141-153.

15. Consul P C (1974), "A Simple Urn Model with Predetermined Strategy", *Sankhya B*, Vol. 36, No. 4, pp. 391-399.

16. Consul P C and Mittal S P (1975), "A New Urn Model with Predetermined Strategy", *Biom. Zeit.*, Vol. 17, No. 2, pp. 67-75.

17. Consul P C and Mittal S P (1977), "Some Discrete Multinomial Probability Models with Predetermined Strategy", *Bio-Zeitschr*, Vol. 19, No. 3, pp. 161-173.

18. Darling D A and Robbins H (1967), "Finding the Size of Finite Population", *Annals of Mathematical Statistics,* Vol. 38, No. 5, pp. 1392-1397.

19. David F N and Barton D E (1962), *Combinatorial Chance*, Griffin, London.

20. De-Moivre A (1756), *The Doctrine of Chances, Or A Method of Calculating Probabilities in Events of Play*, 3rd Edition, A Millar, London.

21. Diaconis P and Holmes S (2002), "A Bayesian Peek into Feller Volume I", *Sankhya, Series A,* Special Issue in Memory of D Basu, Vol. 64, pp. 820-841.

22. Eggenberger F and Polya G (1923), "*Über die Statistik Verketteter Vorgänge*", *Zeitschrift, für angewandte Mathematik and Mechanik*, Vol. 3, pp. 279-289.

23. Elliott R and Hinz J (2003), "A Method for Portfolio Choice", *Appl. Stochastic Model Bus. Ind.*, Vol. 19, No. 1, pp. 1-11.

24. Erylmaz S (2008), "Run Statistics Defined on Multicolour Urn Model", *J. Appl. Prob.,* Vol. 45, No. 4, pp. 1007-1023.

25. Fagin R (1975), "Asymptotic Miss Ratios Over Independent References", IBM Research Report, Rc 5415, Yorktown Heights, NY.

26. Feller W (1968), *An Introduction to Probability Theory and Its Applications*, 3rd Edition, Vol. 1, pp. 1-499, Wiley, New York.

27. Flajolet P, Gardy D and Thimonier L (1992), "Birthday Paradox, Coupon Collectors, Cashing Algorithms and Self-Organizing Search", *Discrete Applied Mathematics,* Vol. 39, pp. 207-229.

28. Freudenthal H (1960), "Models in Applied View on Fundamental Stochastic Ideas", *Educ. Stud. Math.,* Vol. 6, pp. 187-205.

29. Friedman B (1949), "A Simple Urn Model", *Commun in Pure and Applied Math*., Vol. 2, No. 1, pp. 59-70.

30. Gautier P (2002), "Non-Sequential Search, Screening Externalities and the Public Good Role of Recruitment Offices", *Economic Modeling,* Vol. 19, No. 2, pp. 179-196.

31. Glaz J (1983), "Moving Window Detection for Discrete Data", *IEEE Transactions on Information Theory*, IT-29, pp. 457-462.

32. Godbole A P (1990), "On Hypergeometric and Related Distributions of Order *k*", *Commun. Statist. Theory Methods*, Vol. 19, No. 4, pp. 1291-1301.

33. Goldstein L (1990), "Poisson Approximation and DNA Sequence Matching", *Commun. Statist. Theory Methods*, Vol. 19, No. 11, pp. 4167-4179.

34. Greenberg I (1970), "The First Occurrences of Successes", in N Trials (Ed.), *Technometrics*, Vol. 12, No. 3, pp. 627-634.

35. Griffith W S (1986), "On Consecutive-*k*-out-of *n*: Failure Systems and Their Generalisations", *Reliability and Quality Control*, Vol. 20, pp. 157-165.

36. Hahn G J and Gage J B (1983), "Evaluation of a Start-Up Demonstration Test", *J. Qual. Tech*., Vol. 15, pp. 103-106.

37. Heitele D (1975), "An Epistemological View on Fundamental Stochastic Ideas", *Educational Studies in Mathematics*, Vol. 6, No. 2, pp. 187-205.

38. Huang W T and Tsai C S (1991), "On a Modified Binomial Distribution of Order *k*", *Statist. and Prob. Letter*, Vol. 11, No. 2, pp. 125-131.

39. Inoue K and Aki S (2008), "Method for Studying Generalized Birthday and Coupon Collection Problems", *Commun. Statist -Simulations and Computation*, Vol. 37, No. 5, pp. 844-862.

40. Janardhan K G (1973), "The Four Urn Models with Predetermined Strategy", *Technical Report*, Vol. 37, No. 1, Dept. Math. Systems Sanamon State Univ., Springfield, Illinois.

41. Janardhan K G (1975), "Markov-Polya Urn Models with Predetermined Strategies-I", *Gujarat Statist. Rev.*, Vol. 2, No. 1, pp. 17-32.

42. Janardhan K G (1978), "On a Generalized Markov-Polya Distributions", *Gujarat Statist. Rev*., Vol. 5, No. 1, pp. 16-32.

43. Janardhan K G and Schaeffer D (1977), "A Generalization of the Markov-Polya Distribution: Its Extensions and Applications", *Biom. Zeit*., Vol. 19, pp. 87-106.

44. Johnson N L and Kotz S (1977), *Urn Models and their Applications*, 1st Edition, John Wiley and Sons, New York.

45. Kotz S and Balakrishnan N (1997), "Advances in Urn Models During the Past Two Decades", *Advances in Combinatorial Methods and Applications to Probability and Statistics*, pp. 203-257, Birkhauser, Boston.

46. Koutras M V and Alexandrou V A (1997), "Non-Parametric Tests Based on Success Runs of Fixed Length", *Statist. Prob. Lett.,* Vol. 32, No. 4, pp. 393-404.

47. Laplace Pierre-Simon (1921), *Essai philiosophique sur les probabilités*, 2 Volumes, Paris, Gauthier-Villars, Vol. 1, pp. 22-23.

48. Lexis W (1877), *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft*, Fr. Wagner'sche Buchhandlung, Freiburg i. B.

49. Makri F S, Philippou A N and Psillakis Z M (2007a), "Success Run Statistics Defined on an Urn Model", *Adv. Appl. Prob.,* Vol. 39, pp. 991-1019.

50. Makri F S, Phillipou A N and Psillakis Z M (2007b), "Polya, Inverse Polya, and Circular Polya Distributions of Order *k* for *l*-Overlapping Success Runs", *Communications in Statistics: Theory and Methods*, Vol. 36, No. 1, pp. 657-668.

51. Markov A A (1917), "A Generalisation of a Sequential Exchange of Balls (in Russian)", in *Collected Works*, p. 589, Presented at a Meeting of Physico-Mathematical Section of Academy of Sciences, October 8, 1917.

52. Mirstik A V (1978), "Multistatic Radar Binomial Detection", *IEEE Transactions on Aerosp. Electron. Syst.,* AES-14, pp. 103-108.

53. Mishra A, Tiwary D and Singh S K (1992), "A Class of Quasi-Binomial Distributions", *Sankhya, Series B,* Vol. 1, No. 58, pp. 67-76.

54. Mohanty S G (1981), "On Some Distributions Related to Lattice Path Combinatorics", in Chaubey Y P and Dwivedi T D (Eds.), *Topics in Applied Statistics*, Proceedings of Statistics 1981 Canada Conference Held at Concordia University, Montreal, Canada, April 29-May 1, 1981.

55. Monahan J F (1987), "An Alternative Method for Computing Overflow Probabilities", *Commun. Statist. Theo. and Meth.*, Vol. 16, pp. 3355-3357.

56. Naor P (1957), "Normal Approximation to Machine Interface with Many Repairmen", *J. Royal Statist. Soc. B*, Vol. 1, pp. 334-341.

57. Nelson J B (1978), "Minimal Order Models for False Alarm Calculations on Sliding Windows", *IEEE Transactions on Aerosp. Electron. Syst*., AES-14, No. 2, pp. 351-363.

58. Newbold E M (1927), "Practical Application of Statistics of Repeated Events: Particularly to Industrial Accidents", *J. Royal Statist. Soc*., Vol. 90, No. 3, pp. 487-547.

59. Paik M K (1983), "Empty or Infinitely Full", *Mathematics Magazine*, Vol. 56, No. 4, pp. 221-223.

60. Panaretos J and Xekalaki E (1986), "On Some Distributions Arising from Generalized Sampling Schemes", *Commun. Statist. Theo. Meth*., Vol. 15, No. 3, pp. 873-891.

61. Papastavridis S G and Koutras M V (1993), "Bounds for Reliability of a Consecutive-*k*-within-*m*-out-of *n*: F System", *IEEE Transactions on Reliability*, Vol. 42, pp. 156-160.

62. Papastavridis S G and Sfakianakis M E (1991), "Optimal Arrangement and Importance of the Components in a Consecutive-*k*-out-of-r-from *n*: F System", *IEEE Transactions on Reliability*, Vol. 40, pp. 277-279.

63. Philippou A N and Makri F S (1986), "Successes, Runs and Longest Runs", *Statist. and Probab. Lett.*, Vol. 4, pp. 211-215.

64. Philippou A N, Geoghiou C and Phillipou G N (1983), "A Generalized Geometric Distribution and Some of Its Properties", *Statist. and Prob. Letters*, Vol. 1, pp. 171-175.

65. Poisson S D (1837), *Recherches sur la probabilité des jugements en matières criminelles et matière civile* (4to, 1837), All published at Paris, A translation of Poisson's Treatise on Mechanics was published in London in 1842.

66. Polya G (1954), *Mathematics and Plausible Reasoning*, Princeton University Press, 2nd Edition (1963), Vol. 1 and 2, Princeton, NJ.

67. Quetelet A (1835), *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale*, 2 Volumes.

68. Ramakrishna M V (1987), "Computing the Probability of Hash Table/Urn Overflow", *Commun. Statist-Theor. and Method*, Vol. 16, pp. 3343-3353.

69. Rao B R and Janardhan K G (1984), "Use of the Generalized Markov-Polya Distribution as a Random Change Model and Its Identifiability", *Sankhya, Series A*, Vol. 58, pp. 458-462.

70. Saparstein B (1973), "On the Occurrence of *n* Successes within *N* Bernoulli Trials", *Technometrics,* Vol. 15, pp. 809-818.

71. Schnabel Z E (1938), "The Estimation of Total Fish of a Lake", *Am. Math. Mon.*, Vol. 45, pp. 348-352.

72. Sen K and Jain R (1996), "Generalized Markov-Polya Urn Models with Predetermined Strategies", *J. of Statist. Plann. and Infer.*, Vol. 54, pp. 119-133.

73. Sen and Mishra (1996), "A Generalized Polya-Eggenberger Model Generating Various Discrete Distributions", *Sankhya, Series A*, Vol. 58, No. 2, pp. 243-251.

74. Sen K, Agarwal M and Chakraborty S (2002a), "Generalized Polya Eggenberger Model of Order *k* via Lattice Path Approach", *J. Statist. Plan. Inf.*, Vol. 102, pp. 467-476.

75. Sen K, Agarwal M and Chakraborty S (2002b), "Type-II Generalized Polya Eggenberger Model of Order *k*", *Assam Statist. Rev.*, Vol. 16, pp. 124-136.

76. Sen K, Agarwal M and Bhattacharya S (2003), "On Circular Distributions of Order *k* Based on Polya-Eggenberger Sampling Scheme", *J. Math. Sci. (N.S.)*, Vol. 2, pp. 34-54, Delhi.

77. Sen K, Agarwal M and Bhattacharya S (2006), "Polya Eggenberger F-S Model of Order (k_1, k_2)", *Studia Scientiarum Mathematicarum Hungarica,* Vol. 43, No. 1, pp. 1-31.

78. Sfakianakis M, Kounias S and Hillaris A (1992), "Reliability of Consecutive-*k*-out-of *r* from *n*: F system", *IEEE Transactions on Reliability*, Vol. 41, No. 3, pp. 442-447.

79. Siegrist K (1987), "An Urn Model with Bernoulli Removals and Independent Additions", *Stochastic Processes and Their Applications*, Vol. 25, No. 2, pp. 315-324.

80. Simha R and Majumdar A (1997), "An Urn Model with Applications to Database Performance Evaluation", *Computers and OR*, Vol. 24, No. 4, pp. 289-300.

81. Small C G (1985), "Decomposition of Models Whose Marginal Distributions are Mixtures", *Cand. J. of Statist*., Vol. 13, No. 2, pp. 131-136.

82. Stigler S M (1986), *The History of Statistics: The Measurement of Uncertainties Before 1900*, Harvard University Press, Cambridge, MA.

83. Witts J T, Collin T and Sidel V W (1974), "Capture-Recapture Methods for Assessing the Completeness of Case Ascertainment When Using Multiple Information Sources", *J. Cron. Dis*., Vol. 20, pp. 311-313.

84. Woodbury M A (1949), "On Probability Distribution", *Ann. Math. Statist*., Vol. 20, pp. 311-313.

Reference # 61J-2010-06-03-01