From runs to weak runs: Generalizations of distributions of order k and applications

Spiros D. Dafnis [1]

[1]Department of Statistics and Actuarial – Financial Mathematics, University of the Aegean, Samos, Greece

International Workshop on “Discrete Distributions”

Outline of the talk

- Classical discrete distributions.
- The idea behind the use of runs - distributions of order k.
- Classical generalizations of runs: Scans.
- A recent generalization of runs: Weak runs.
- Indicative applications of weak runs.

Introduction

A variety of problems in different research areas can be modeled by considering binary trials (success - failure) and studying r.v.'s denoting

- the number of successes in a given number of trials.
- the waiting time for the appearance of a given number of successes.

Introduction

A variety of problems in different research areas can be modeled by considering binary trials (success - failure) and studying r.v.'s denoting

- the number of successes in a given number of trials.
- the waiting time for the appearance of a given number of successes.

Classical discrete distributions

- Binomial distribution
- Geometric distribution
- Negative binomial distribution

A first generalization

- In all the aforementioned cases, the researcher is interested in the occurrence of single successes.
- Are there theoretical tools when a specific application requires the study of the occurrence of consecutive successes?

A first generalization

- In all the aforementioned cases, the researcher is interested in the occurrence of single successes.
- Are there theoretical tools when a specific application requires the study of the occurrence of consecutive successes?

Theory of runs

We are now interested in the occurrence of runs of successes of length k (k consecutive successes) with $k \geq 1$. In that case, the classical discrete distributions are generalized to the distributions of order k.

Distributions of order k

- The name "distributions of order k" is due to Philippou et al. (1983), who studied the geometric distribution of order k.
- Philippou and Makri (1986) (see, also, Hirano (1986)) studied the binomial distribution of order k.
- For an, up to 2002, review on "distributions of order k" we refer to the excellent book of Balakrishnan and Koutras (2002).

Distributions of order k

- The name "distributions of order k" is due to Philippou et al. (1983), who studied the geometric distribution of order k.
- Philippou and Makri (1986) (see, also, Hirano (1986)) studied the binomial distribution of order k.
- For an, up to 2002, review on "distributions of order k" we refer to the excellent book of Balakrishnan and Koutras (2002).

Example: Binomial distribution of order $k = 3$.

$F\underbrace{SSS}\underbrace{SSS}F\underbrace{SSS}FSSF$

A classical generalization of runs: Scans

- The assumption of necessarily consecutive successes is relaxed.
- Instead, we consider windows of trials of length m, each of which contains at least k ($k \leq m$) successes (scans k/m).
- A k/k scan is a run of length k (a scan is a generalized run).

A classical generalization of runs: Scans

- The assumption of necessarily consecutive successes is relaxed.
- Instead, we consider windows of trials of length m, each of which contains at least k ($k \leq m$) successes (scans k/m).
- A k/k scan is a run of length k (a scan is a generalized run).

Example: All 2/3 scans

SSS, SSF, SFS, FSS

A classical generalization of runs: Scans

- The assumption of necessarily consecutive successes is relaxed.
- Instead, we consider windows of trials of length m, each of which contains at least k ($k \leq m$) successes (scans k/m).
- A k/k scan is a run of length k (a scan is a generalized run).

Example: All 2/3 scans

SSS, SSF, SFS, FSS

Glaz and Balakrishnan (1999), Glaz et al. (2001).

A recent generalization of runs: Weak runs in binary trials

Definition

Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of binary (or multistate) trials, numbered from 1 to n. The distance between two trials numbered i and j is $d(Z_i, Z_j) = |j - i|$, $1 \leq i, j \leq n$.

Definition

Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of binary trials ($Z_i = 0$ or $Z_i = 1$), numbered from 1 to n. Then, an r-weak 1-run of length (exactly/at least) k is a string of (exactly/at least) k ones (1s), where any two consecutive 1s (of the string) at trials i, j ($1 \leq i < j \leq n$) may have a maximum distance of $r + 1$, i.e. $\forall i, j$ with $Z_i = 1$, $Z_j = 1$ and $Z_{i+t} = 0$ for $t = 1, \ldots, j - i - 1$, then $d(Z_i, Z_j) \leq r + 1$.

Example: All 2^{r+1} r-weak runs of length k=3

$$1\underbrace{0\ldots0}_{\leq r}1\underbrace{0\ldots0}_{\leq r}1$$

Weak runs in binary trials and related distributions

Example: All 2^{r+1} r-weak runs of length $k{=}3$

$$1\underbrace{0\ldots0}_{\leq r}1\underbrace{0\ldots0}_{\leq r}1$$

For $r = 0$, a 0-weak run of length k reduces to a typical run of length k. Therefore, binomial-type or waiting time weak-run distributions are generalized run distributions.

Dafnis and Makri (2022, 2023) studied binomial-type weak-run distributions by employing the Markov chain imbedding technique.

Markov chain imbedding technique and imbeddable variables

Markov chain imbedding technique (Fu and Koutras (1994), Fu and Lou (2003)) projects the random variable of interest to appropriate subspaces of the state space of a properly defined Markov chain.

Dafnis and Makri (2023) introduced the Markov chain imbeddable variable of returnable - polynomial type, which generalizes the

- Markov chain imbeddable variable of binomial type (Koutras and Alexandrou, 1995).
- Markov chain imbeddable variable of returnable type (Han and Aki, 1999).
- Markov chain imbeddable variable of polynomial type (Antzoulakos et al., 2003).

Applications of weak-run related distributions

- System reliability (Dafnis et al. (2019)).
- Heat unit models in Agriculture (Dafnis and Makri (2022, 2023)).
- Statistical quality control, and, more specifically, statistical process monitoring (SPM, Dafnis et al. (2023, 2024)).

SPM - Chi-square control chart (Phase II)

- Simultaneous monitoring of m distinct quality characteristics.
- At each time t, the sample mean vector $\bar{\boldsymbol{X}}_{\boldsymbol{t}} = (\bar{\boldsymbol{X}}_{\boldsymbol{1t}}, \bar{\boldsymbol{X}}_{\boldsymbol{2t}}, \ldots, \bar{\boldsymbol{X}}_{\boldsymbol{mt}})$ of size n of m characteristics is collected.
- The m-related quality characteristics are jointly distributed according to an m-variate normal distribution with known in-control process mean vector $\boldsymbol{\mu}_0$, and covariance matrix, $\boldsymbol{\Sigma}_0$.
- OOC signal as soon as

$$D_t^2 = n(\bar{\mathbf{X}}_{\mathbf{t}} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\bar{\mathbf{X}}_{\mathbf{t}} - \boldsymbol{\mu}_0) > \mathrm{UCL}.$$

- D_t^2 follows a non-central chi-square distribution with m degrees of freedom ($x_m^2(\lambda)$), where $\lambda = nd^2$ is the non-centrality parameter and $d = \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)}$ is the Mahalanobis distance.
- $d = 0$ ($\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0$): the process is in-control.
- $d > 0$ ($\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_0$): the process is out-of-control .

SPM - Chi-square control chart (CSCC)

Design and performance measurement

- *RL*: the number of plotted points until the first one which is above UCL.
- Design: $ARL_{IC} = \frac{1}{\mathrm{P}(D_t^2 > UCL | \boldsymbol{\mu} = \boldsymbol{\mu}_0)}$.
- Performance measurement: $\mathrm{ARL}_{OC} = \frac{1}{\mathrm{P}(D_t^2 > UCL | \boldsymbol{\mu} = \boldsymbol{\mu}_1)}$.

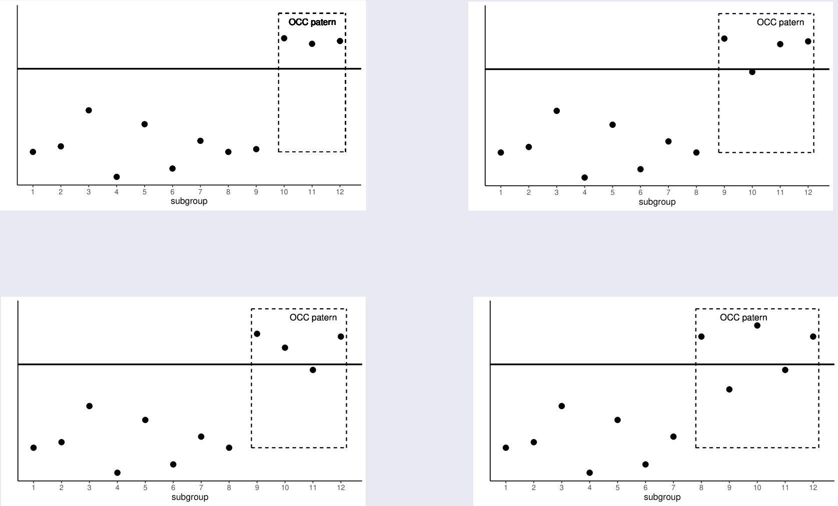
Chi-square control chart with run-rules

- Traditional CSCCs are preferable in detecting large shifts in the process mean vector.
- When a shift of small magnitude occurs during the process monitoring, their operation is less sensitive in detecting the shift in the mean vector.
- For that reason, CSCCs with run rules have been introduced and extensively studied in the literature.
- What if run rules are replaced by the (generalized) weak-run rules?

Weak runs and Statistical Process Monitoring

- Dafnis et al. (2023) introduced a new chi-square control chart (CSCC-$k(r)$) based on weak runs.
- An out-of-control signal is given if k consecutive plotted points fall above UCL, even if any two successive of these are separated by at most r points placed in $[0, \mathrm{UCL}]$.
- The authors compared the performance of the proposed chart to the performance of several charts with runs rules seen in the literature, proving its superiority for detecting small to moderate process mean vector shifts.
- For different shifts in the process mean vector, the optimal design was determined (the optimal values of k and r) and the chart's performance was measured by the aid of the mean value of the waiting time variable for the first occurrence of a weak run.

Demonstration of the CSCC-3(1)



ARL of CSCC-$k(r)$

The mean value of the r.v. $T_{k,r}$ $(k \geq 2, r \geq 1)$ is given by $E(T_{k,r}) = \mathbf{e}_1(\mathbf{I} - \mathbf{A})^{-1}\mathbf{1}'$, where $\mathbf{A}$ is an $((k-1)r+k) \times ((k-1)r+k)$ matrix which has all its entries 0 except for the entries:

- $(1,1)$, which is equal to q,
- $(k+jr,1)$, $j = 1, ..., k-1$, which are all equal to q,
- $(i, i+1)$, $1 \leq i \leq k-1$, which are all equal to p,
- $(i, k+(i-2)r+1)$, $2 \leq i \leq k$, which are all equal to q,
- $(k+jr+i, j+3)$ for $k \geq 3$, $1 \leq i \leq r$, $j = 0, 1, \ldots, max\{0, k-3\}$, which are all equal to p,
- $(k+i-1, k+i)$ for $r \geq 2$, $(j-2)r+2 \leq i \leq (j-1)r$, $j = 2, 3, \ldots, k$, which are all equal to q,

$\mathbf{e}_m$ denotes the m-th unit row vector of $\mathbb{R}^{(k-1)r+k}$, $\mathbf{1}$ denotes the row vector of $\mathbb{R}^{(k-1)r+k}$ with all its entries being equal to 1 and $\mathbf{I}$ denotes the $((k-1)r+k) \times ((k-1)r+k)$ identity matrix.

Performance comparison - indicative results

Performance comparisons between the proposed chart and the CS: r/m chart (Rakitzis and Antzoulakos, 2011) for $\mathrm{ARL}_{IC} = 200$

			CS: r/m chart				CSCC-$k(r)$	
m	n	d	scan	UCL	WCL	ARL_{OC}	(k, r, UCL)	ARL_{OC}
5	1	0.50	3/5	8.55	19.80	132.89	(6,5,8.69)	**126.74**
		1.00	3/5	8.63	19.34	50.93	(5,5,9.21)	**49.12**
		1.25	3/5	8.70	19.04	**30.07**	(4,4,9.63)	30.62
		1.50	3/5	8.79	18.71	**18.36**	(3,4,10.77)	19.89
		0.50	3/5	8.57	19.64	92.74	(6,5,8.69)	**87.22**
	2	1.00	3/5	8.76	18.83	**21.62**	(4,4, 9.63)	22.91
		1.25	3/5	8.92	18.36	**11.53**	(3,3,10.42)	13.01
		1.50	2/5	9.14	17.92	**6.96**	(2,2,12.22)	8.08
		0.50	3/5	8.66	19.20	39.63	(5,5,9.21)	**39.21**
	5	1.00	3/5	9.23	17.80	**6.06**	(2,2,12.22)	6.98
		1.25	2/5	11.46	18.61	**3.29**	(2,1,11.74)	3.97
		1.50	2/4	11.00	17.65	**2.13**	(2,1,11.74)	[illegible]

Improved Shewhart-type control charts - Weak runs in multistate trials

- Weak-run rules can also establish improved Shewhart-type (univariate) control charts.
- An appropriate extension of the definition of weak runs in the case of multistate trials is required (Dafnis et al. 2024).

Future research directions

- From a theoretical prospective, distributions related to weak runs in binary (or multistate) trials could be studied under different theoretical frameworks (for example, by considering dependent trials), in accordance with the specific field they will be applied on.
- From a practical prospective,run distributions have numerous applications in different fields than the ones currently presented. The incorporation of the extra parameter r provides flexibility in problems of higher complexity.
- When it comes to SPM, it is of interest to investigate the effect of adding warning limits to the new-proposed charts.

Weak-Run References

1. Dafnis SD, Makri FS (2022) Weak runs in sequences of binary trials. *Metrika*, 85, 573-603.
2. Dafnis SD, Makri FS (2023) Distributions related to weak runs with a minimum and a maximum number of successes: A unified approach. *Methodology and Computing in Applied Probability*, 24, 573-603.
3. Dafnis SD, Makri FS, Philippou AN (2019) The reliability of a generalized consecutive system. *Applied Mathematics and Computation*, 359, 187-193.
4. Dafnis SD, Perdikis T, Koutras MV (2024) Improved Shewhart-type control charts based on weak runs in multistate trials. *Quality Technology and Quantitative Management*, 10.1080/16843703.2024.2313240.
5. Dafnis SD, Perdikis T, Papadopoulos GK (2023) Improved Chi-square control charts with weak-run rules. *Communications in Statistics - Theory and Methods*, 10.1080/03610926.2023.2215357.

Thank you for your attention!!