On ARL-unbiased charts to monitor the traffic intensity of a single server queue

Manuel Cabral Morais(1) and **Sven Knoth**(2)

(1) Department of Mathematics & CEMAT — IST, ULisboa, Portugal

(2) Department of Economics and Social Sciences — Helmut Schmidt University, Germany

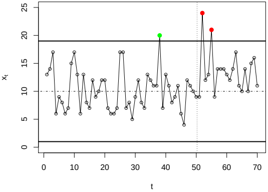
IST — Lisboa, May 14, 2020

The founder of statistical process control (SPC)

- By proposing a quality control chart to his superiors (Bell Laboratories), in a memorandum on May 16, 1924, the American physicist, engineer and statistician Walter Andrew Shewhart (1891–1967) altered the course of industry, brought together statistics, engineering, and economics and became known as the father of modern quality control.

Quality control charts

- Are used to track process performance over time and detect changes in a process parameter, by plotting the observed value of a statistic against time and comparing it with a pair of control limits. An observation beyond the control limits indicate potential assignable causes responsible for changes...



- Applications

  Computer intrusion detection, finance, health care, queueing systems, staff management, water monitoring, etc.

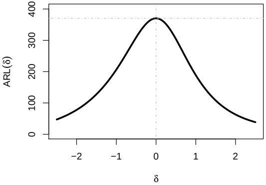
Performance of control charts

- **Average run length** (ARL) — expected number of samples taken until a signal is triggered by the chart.

  It is desirable that valid signals/false alarms are emitted as quickly/rarely as possible, corresponding to small out-of-control/large in-control **ARL**.

- Ideally, the **ARL function should achieve a maximum when the process is in-control** (the chart takes longer, in average, to trigger a false alarm than to detect any shifts), i.e., the chart is **ARL-unbiased**

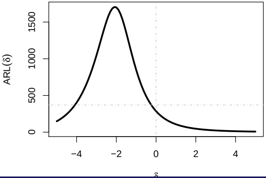
  (Pignatiello *et al.*, 1995; Basseville and Nikiforov, 1993).



Disadvantages of most control charts

- ARL-biased!

  Inability to have a pre-specified in-control ARL, $ARL^\star$, when the control statistic is a discrete r.v. (c.d.f. is a step function).



Queueing systems

Customers eventually wait for service and congestion occurs due to the random character of the arrival process and the service times.

Pioneering work in queueing systems by Erlang (1909, 1917, 1920).

Relevant parameters

- rate of arrivals — λ
- rate of service — μ
- number of parallel servers — s
- **traffic intensity** — if the queueing system has an unlimited waiting room then $\rho = \lambda/(s\mu)$ is a measure of congestion and represents the load offered to each server if the work is divided equally among servers.

Important

The effective operation of a queueing system requires maintaining a desired level of traffic intensity, thus the use of regulation techniques for ρ.

- Bhat and Rao (1972) — seminal work proposing an unusual control chart for the traffic intensity of $M/G/1$ and $GI/M/1$ systems;

  a signal is triggered if the statistic falls beyond the control limits longer than a preassigned number of consecutive observations.

- Approaches to monitoring ρ can be divided in categories depending on:
  - the information we collect
    - the no. of customers in the system at departure/arrival epochs
    - the no. of arrivals while the n^{th} customer is being served, etc.;
  - the statistical technique used to detect changes in ρ
    - control charts (1972, 2000, 2006, 2007, 2011, 2012, 2015)
    - sequential probability ratio tests (1984, 1987, 1989, 2000, 2013).

- (Potential) applications

  Bear in mind that production, computer and transportation systems are often modelled as queueing systems.

  E.g.: serial production line; call center performance monitoring.

To monitor the traffic intensity of a single server queue and keep it at a target level ρ_0, we shall use:

- $X_n =$ no. of customers left behind in the $M/G/1$ system by the n^{th} departing customer;
- $\hat{X}_n =$ no. of customers seen in the $GI/M/1$ system by the n^{th} arriving customer;
- $W_n =$ waiting time of the n^{th} arriving customer to the $GI/G/1$ system.

These three control statistics were chosen for their simplicity, recursive and Markovian character.

System	Control statistic
$M/G/1$	$X_{n+1} = \max\{0,\, X_n - 1\} + Y_{n+1}$
$GI/M/1$	$\hat{X}_{n+1} = \max\{0, \hat{X}_n + 1 - \hat{Y}_{n+1}\}$
$GI/G/1$	$W_{n+1} = \max\{0,\, W_n + S_{n+1} - A_{n+1}\}$ (Lindley equation)

Increments

- $Y_{n+1} =$ no. of customers arriving during the service of the $(n+1)^{th}$ customer.
  $Y_n \overset{i.i.d.}{\sim} Y$, $n \in \mathbb{N}$; $P_Y(i) = \alpha_i$, $i \in \mathbb{N}_0$.
- $\hat{Y}_{n+1} =$ no. of customers served between the arrivals of customers n and $n+1$.
  $\hat{Y}_n \overset{i.i.d.}{\sim} \hat{Y}$, $n \in \mathbb{N}$; $P_{\hat{Y}}(i) = \hat{\alpha}_i$, $i \in \mathbb{N}_0$.
- $S_{n+1} - A_{n+1}$, where $S_{n+1} =$ service time of the n^{th} customer,
  $A_{n+1} =$ time between the arrivals of customers n and $(n+1)$, for $n \in \mathbb{N}_0$.
  $S_n - A_n \overset{i.i.d.}{\sim} S - A$, $n \in \mathbb{N}$.

- **Transition probability matrix of X_n**

$$\mathbf{P} = \begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \cdots \\ \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \cdots \\ 0 & \alpha_0 & \alpha_1 & \alpha_2 & \ddots \\ 0 & 0 & \alpha_0 & \alpha_1 & \ddots \\ 0 & 0 & 0 & \alpha_0 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where α_i denotes the probability that exactly i customers arrive during a service time S,

$$\alpha_i = P_Y(i) = \int_0^{+\infty} e^{-\lambda s} \frac{(\lambda s)^i}{i!} \, dF_S(s), \; i \in \mathbb{N}_0.$$

- **Special case**: $M/E_k/1 \;\rightarrow\; Y \sim \text{NegBinomial}^*(k, k(k+\rho)^{-1}), \; k \in \mathbb{N}.$

- **TPM of $\hat{X}_n$**

$$\hat{\mathbf{P}} = \begin{bmatrix} \hat{p}_{00} & \hat{\alpha}_0 & 0 & 0 & 0 & \cdots \\ \hat{p}_{10} & \hat{\alpha}_1 & \hat{\alpha}_0 & 0 & 0 & \cdots \\ \hat{p}_{20} & \hat{\alpha}_2 & \hat{\alpha}_1 & \hat{\alpha}_0 & 0 & \cdots \\ \hat{p}_{30} & \hat{\alpha}_3 & \hat{\alpha}_2 & \hat{\alpha}_1 & \hat{\alpha}_0 & \ddots \\ \vdots & \cdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

where $\hat{\alpha}_i$ denotes the probability of serving i customers during an interarrival time A (given that the server remains busy during this interval)

$$\hat{\alpha}_i = P_{\hat{Y}}(i) = \int_0^{+\infty} e^{-\mu a} \frac{(\mu a)^i}{i!} \, dF_A(a), \; i \in \mathbb{N}_0,$$

and $\hat{p}_{i0} = 1 - \sum_{j=0}^{i} \hat{\alpha}_j, \; i \in \mathbb{N}_0$.

- **Special case**: $E_k/M/1 \;\rightarrow\; \hat{Y} \sim \text{NegBinomial}^*(k, \rho\,(k^{-1} + \rho)^{-1})$.

- **Approximate TPM of W_n**

$$\tilde{\mathbf{P}} = \begin{bmatrix} F(0) & F(\Delta) - F(0) & F(2\Delta) - F(\Delta) & \cdots \\ F\left(-\frac{\Delta}{2}\right) & F\left(\frac{\Delta}{2}\right) - F\left(-\frac{\Delta}{2}\right) & F\left(\frac{3\Delta}{2}\right) - F\left(\frac{\Delta}{2}\right) & \cdots \\ F\left(-\frac{3\Delta}{2}\right) & F\left(-\frac{\Delta}{2}\right) - F\left(-\frac{3\Delta}{2}\right) & F\left(\frac{\Delta}{2}\right) - F\left(-\frac{\Delta}{2}\right) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where $F \equiv F_{S-A}$.

Discretized approximating DTMC:

  - state space $\mathbb{N}_0$;
  - first state corresponds to the singleton $\{0\}$;
  - state j is associated with interval $E_j = ((j-1)\Delta, j\Delta]$, for $j \in \mathbb{N}$, where Δ denotes the common range of all the intervals;
  - $E_j = ((j-1)\Delta, j\Delta]$ is represented by its midpoint $(j-1/2)\,\Delta$, for $j \in \mathbb{N}$;
  - $P(W_{n+1} \in E_j \mid W_n \in E_i) \simeq P[W_{n+1} \in E_j \mid W_n = (i-1/2)\,\Delta]$.

- **Special case**: $M/M/1 \;\rightarrow\; F_{S-A}(x) = \begin{cases} \mu\, e^{\lambda x}/(\lambda+\mu), & x \leq 0 \\ 1 - \lambda\, e^{-\mu x}/(\lambda+\mu), & x > 0. \end{cases}$

Motivation

- Downward (resp. upward) shifts in the traffic intensity can correspond to a decreasing (resp. increasing) interest in the offered services, thus calling for a timely detection.

Main goal

- Design a control chart with a peak of the ARL curve at $\rho = \rho_0$ — i.e., an *ARL-unbiased chart* for ρ — and a pre-specified in-control ARL.

Challenges

- The discrete (resp. mixed) character of the control statistics $X_n, \hat{X}_n$ (resp. W_n).
- The high frequency of zero values when compared to other values of these control statistics.

  If we are to design a chart to monitor ρ with a reasonably large in-control ARL, we have to set $LCL \equiv 0$ and deal with an inherently upper one-sided chart.

A solution for the $M/G/1$ and $GI/M/1$ systems

- The **ARL-unbiased X_n-chart** used to monitor the traffic intensity of the $M/G/1$ system should trigger a **signal** at the n^{th} departure with:
  - **probability one** if the **number of customers left behind** by the n^{th} departing customer, x_n, is **larger than the upper control limit** U;
  - **probability** γ_L (resp. γ_U) **if** x_n **is equal to** $L \equiv 0$ (resp. U).

- Randomizing the emission of a signal $\Rightarrow$ using the sub-stoch. matrix $\mathbf{Q}$

$$\begin{bmatrix} p_{L\,L} \times (1-\gamma_L) & p_{L\,L+1} & \cdots & p_{L\,U-1} & p_{L\,U} \times (1-\gamma_U) \\ p_{L+1\,L} \times (1-\gamma_L) & p_{L+1\,L+1} & \cdots & p_{L+1\,U-1} & p_{L+1\,U} \times (1-\gamma_U) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{U-1\,L} \times (1-\gamma_L) & p_{U-1\,L+1} & \cdots & p_{U-1\,U-1} & p_{U-1\,U} \times (1-\gamma_U) \\ p_{U\,L} \times (1-\gamma_L) & p_{U\,L+1} & \cdots & p_{U\,U-1} & p_{U\,U} \times (1-\gamma_U) \end{bmatrix}.$$

- Performance measure: $ARL^0 = \underline{\mathbf{e}}_0^\top \times (\mathbf{I}-\mathbf{Q})^{-1} \times \underline{\mathbf{1}}$.
  Search for L, γ_L, etc. follows the same lines of the algorithm used to derive the ARL-unbiased $c-$chart for the mean of INAR(1) Poisson counts (Paulino et al., 2019).

- The **ARL-unbiased $\hat{X}_n-$chart** is obtained similarly...

A solution for the $GI/G/1$ system

- The **ARL-unbiased W_n-chart** to control the traffic intensity of the $GI/G/1$ system should trigger a **signal** at the n^{th} arrival with:
  - **probability one** if the **waiting time** of the n^{th} arriving customer, w_n, is **larger than the upper control limit** U;
  - **probability γ_L if w_n is equal to $L \equiv 0$.**
- Randomizing the emission of a signal $\Rightarrow$ using the sub-stoch. matrix $\tilde{\mathbf{Q}}$

$$\begin{bmatrix} \tilde{p}_{L\,L} \times (1-\gamma_L) & \tilde{p}_{L\,L+1} & \cdots & \tilde{p}_{L\,\frac{U}{\Delta}-1} & \tilde{p}_{L\,\frac{U}{\Delta}} \\ \tilde{p}_{L+1\,L} \times (1-\gamma_L) & \tilde{p}_{L+1\,L+1} & \cdots & \tilde{p}_{L+1\,\frac{U}{\Delta}-1} & \tilde{p}_{L+1\,\frac{U}{\Delta}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \tilde{p}_{\frac{U}{\Delta}-1\,L} \times (1-\gamma_L) & \tilde{p}_{\frac{U}{\Delta}-1\,L+1} & \cdots & \tilde{p}_{\frac{U}{\Delta}-1\,\frac{U}{\Delta}-1} & \tilde{p}_{\frac{U}{\Delta}-1\,\frac{U}{\Delta}} \\ \tilde{p}_{\frac{U}{\Delta}\,L} \times (1-\gamma_L) & \tilde{p}_{\frac{U}{\Delta}\,L+1} & \cdots & \tilde{p}_{\frac{U}{\Delta}\,\frac{U}{\Delta}-1} & \tilde{p}_{\frac{U}{\Delta}\,\frac{U}{\Delta}} \end{bmatrix},$$

- Performance measure: $ARL^0 = \underline{\mathbf{e}}_0^\top \times (\mathbf{I}-\mathbf{Q})^{-1} \times \underline{\mathbf{1}}$; alternatively, ARL can be obtained by solving an integral equation.
  Search for U and γ_L involves a nested secant rule as the algorithm used to derive the ARL-unbiased EWMA$-S^2$ chart (Knoth and Morais, 2013).

To detect downward and upward shifts in the traffic intensity...

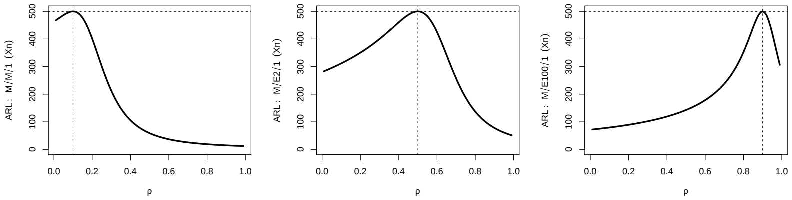
- Target value: $\rho_0 = 0.1, 0.5, 0.9$
- In-control ARL: $ARL^\star = 500$
- ARL-unbiased designs
  - X_n: $M/M/1$, $M/E_2/1$ and $M/E_{100}/1$;
  - $\hat{X}_n$: $M/M/1$, $E_2/M/1$ and $E_5/M/1$;
  - W_n: $M/M/1$, $M/E_2/1$ and $E_2/M/1$

    either with fixed arrival rate or with fixed service rate.

ARL obtained using: the Markov chain approach (X_n- and $\hat{X}_n-$charts); the collocation method to solve the integral equation (W_n-chart).

ARL-unbiased X_n −chart

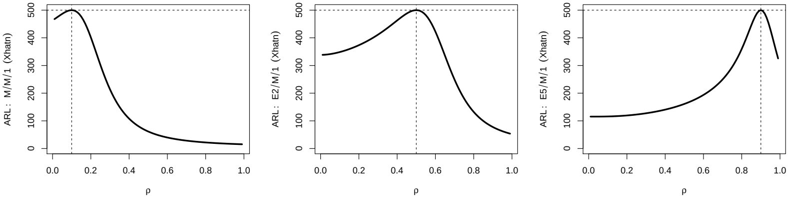
System	ρ_0	$[L, U]$	(γ_L, γ_U)	$ARL(0.95\,\rho_0)$	$ARL(\rho_0)$	$ARL(1.05\,\rho_0)$
$M/M/1$	0.1	[0, 4]	(0.002160, 0.629778)	499.816	500.000	499.805
	0.5	[0, 10]	(0.003568, 0.609947)	496.526	500.000	495.881
	0.9	[0, 30]	(0.013043, 0.709996)	462.258	500.000	455.964
$M/E_2/1$	0.1	[0, 3]	(0.002152, 0.068181)	499.838	500.000	499.829
	0.5	[0, 8]	(0.003566, 0.320705)	496.497	500.000	495.810
	0.9	[0, 24]	(0.013475, 0.066710)	457.401	500.000	447.720
$M/E_{100}/1$	0.1	[0, 3]	(0.002147, 0.328369)	499.855	500.000	499.848
	0.5	[0, 6]	(0.003558, 0.170932)	496.514	500.000	495.797
	0.9	[0, 19]	(0.014002, 0.943674)	450.843	500.000	434.972



Similar ARL profiles for the $M/M/1$, $M/E_2/1$ and $M/E_{100}/1$ systems and a fixed ρ_0.

ARL-unbiased $\hat{X}_n$−chart

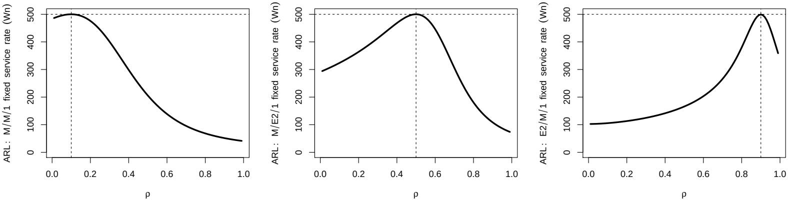
System	ρ_0	$[L, U]$	(γ_L, γ_U)	$ARL(0.95\ \rho_0)$	$ARL(\rho_0)$	$ARL(1.05\ \rho_0)$
$M/M/1$	0.1	[0, 4]	(0.002160, 0.634850)	499.816	500.000	499.805
	0.5	[0, 10]	(0.003567, 0.651244)	496.545	500.000	495.914
	0.9	[0, 29]	(0.012936, 0.221365)	463.558	500.000	458.852
$E_2/M/1$	0.1	[0, 3]	(0.002039, 0.876869)	499.898	500.000	499.886
	0.5	[0, 7]	(0.002955, 0.065346)	496.559	500.000	495.751
	0.9	[0, 24]	(0.010323, 0.532068)	458.093	500.000	449.697
$E_5/M/1$	0.1	[0, 2]	(0.002004, 0.238163)	499.973	500.000	499.967
	0.5	[0, 6]	(0.002600, 0.408281)	496.673	500.000	495.704
	0.9	[0, 20]	(0.008666, 0.133624)	453.910	500.000	441.644



$\gamma_{L\equiv 0}$ also tends to be much smaller than γ_U, to achieve a fairly large in-control ARL in the presence of very frequent zero values of the statistic.

ARL-unbiased W_n-chart — fixed service rate

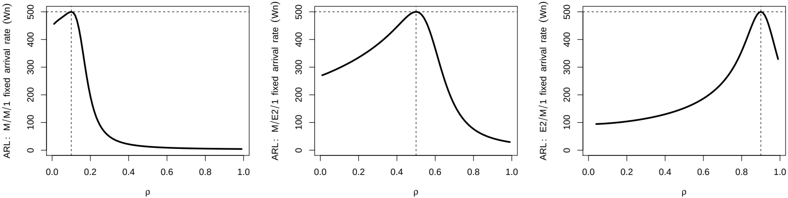
System	ρ_0	U	γ_L	$ARL(0.95\ \rho_0)$	$ARL(\rho_0)$	$ARL(1.05\ \rho_0)$
$M/M/1$	0.1	7.077585	0.002068	499.949	500.000	499.948
	0.5	11.912665	0.003335	497.823	500.000	497.533
	0.9	29.461491	0.012315	467.940	500.000	463.249
$M/E_2/1$	0.1	4.738168	0.002095	499.925	500.001	499.923
	0.5	8.863481	0.003433	497.519	500.141	497.091
	0.9	24.001537	0.013020	462.104	500.351	454.016
$E_2/M/1$	0.1	6.423954	0.002006	499.989	500.000	499.988
	0.5	9.757652	0.002741	498.137	500.000	497.836
	0.9	24.234818	0.009750	463.359	499.176	455.744



Replacing the $\hat{X}_n-$chart with a W_n-chart does not pay-off in terms of ARL, when the service rate has been fixed.

ARL-unbiased W_n−chart — fixed arrival rate

System	ρ_0	U	γ_L	$ARL(0.95\ \rho_0)$	$ARL(\rho_0)$	$ARL(1.05\ \rho_0)$
$M/M/1$	0.1	0.911543	0.002198	499.505	500.000	499.400
	0.5	6.773410	0.003706	494.831	500.000	493.402
	0.9	28.254818	0.013579	457.637	500.000	451.070
$M/E_2/1$	0.1	0.590423	0.002200	499.457	500.005	499.316
	0.5	4.957663	0.003733	494.626	500.274	492.561
	0.9	22.823032	0.014126	451.630	500.372	440.513
$E_2/M/1$	0.1	0.807873	0.002049	499.785	500.000	499.733
	0.5	5.554376	0.003021	495.473	500.000	494.093
	0.9	23.098971	0.010632	452.652	498.926	442.630



The $\hat{X}_n$−chart compares unfavourably to the W_n−chart in terms of ARL, when the arrival rate has been fixed.

- **ARL-unbiased charts for the traffic intensity**
  - Their control statistics have a recursive and Markovian character.
  - The associated ARL curves attain a maximum when ρ is on target.
  - Their in-control ARL take a pre-stipulated value.
  - Tackle the curse of the null LCL and detect decreases in ρ in a timely fashion, by relying on the randomization probabilities.
- **On-going and future work**
  - Derive ARL-unbiased designs referring to other interarrival time distributions such as the hyperexponential and hypoexponential, commonly used in QT and in practice.
  - Additional comparisons between the $\hat{X}_n-$ and W_n-charts, based on the RL percentage points and SDRL.
  - Derivation of ARL-unbiased versions of existing/sophisticated charts (WZ and CUSUM).

Bibliography

- Basseville, M., & Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall.
- Bhat, U. N. (1987). A statistical technique for the control of traffic intensity in Markovian queue. *Annals of Operations Research, 8*, 151–164.
- Bhat, U. N., & Rao, S. S. (1972). A statistical technique for the control of traffic intensity in the queuing systems M/G/1 and GI/M/1. *Operations Research, 20*, 955–966.
- Brook, D., & Evans, D. A. (1972). An approach to the probability distribution of CUSUM run length, *Biometrika, 59*, 539–549.
- Chen, N., Yuan, Y., & Zhou, S. (2011). Performance analysis of queue length monitoring of M/G/1 systems. *Naval Research Logistics, 58*, 782–794.
- Chen, N., & Zhou, S. (2015). CUSUM statistical monitoring of M/M/1 queues and extensions. *Technometrics, 57*, 245–256.
- Cohen, J. W. (1982). *The Single Server Queue* (revised edition). Amsterdam: North-Holland Publishing Company.
- Erlang, A. K. (1909). Sandsynlighedsregning og Telefonsamtaler. *Nyt Tidsskrift for Matematik B (Copenhagen), 20*, 33–41. Translation: The theory of probabilities and telephone conversations. In: Brockmeyer, Halstrøm & Jensen (1948, pp. 131–137).
- Erlang, A. K. (1917). Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler. *Elektrotkeknikeren (Copenhagen), 13*, 5–13. Translation: Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. In: Brockmeyer, Halstrøm & Jensen (1948, pp. 138–155).
- Erlang, A. K. (1920). Telefon-Ventetider. Et Stykke Sandsynlighedsregning. *Matematisk Tidsskrift B (Copenhagen), 31*, 25–42. Translation: Telephon waiting times: an example of probability calculus. In: Brockmeyer, Halstrøm & Jensen (1948, pp. 156–171).
- Feller, W. (1971). *An Introduction to Probability Theory and its Applications* (2nd. edition). New York: John Wiley & Sons.
- Greenberg, I. (1997). Markov chain approximation methods in a class of level-crossing problems. *Operations Research Letters, 21*, 153–158.
- Hung, Y.-C., Michailidis, G., & Chuang, S.-C. (2012). Estimation and monitoring of traffic intensities with application to control of stochastic systems. *Applied Stochastic Models in Business and Industry, 30*, 200–217.

- Jain, S. (2000). An autoregressive process and its application to queueing model. *Metron-International Journal of Statistics, 58*, 131–138.
- Jain, S., & Templeton, J. G. C. (1989). Problem of statistical inference to control the traffic intensity. *Sequential Analysis, 8*, 135–146.
- Kendall, D. G. (1951). Some problems in the theory of queues. *Journal of the Royal Statistical Society, Series B (Methodological), 13*, 151–185.
- Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics, 24*, 338–354.
- Kim, S.-H., Alexopoulos, C., Tsui, K.-L., & Wilson, J. R. (2007). A distribution-free tabular CUSUM chart for autocorrelated data. *IIE Transactions, 39*, 317–330.
- Kleinrock, L. (1975). *Queueing Systems, Volume I: Theory*. New York: John Wiley & Sons.
- Knoth, S. (2005). Accurate ARL computation for EWMA-S^2 control charts. *Statistics and Computing, 15*, 341–352.
- Knoth, S. (2010). Control charting normal variance – reflections, curiosities, and recommendations. In: H.-J. Lenz. & P.-T. Wilrich (Eds.), *Frontiers in Statistical Quality Control* (Vol. 9, pp. 3–18). Heidelberg: Physica.
- Knoth, S., & Morais, M. C. (2015). On ARL-unbiased control charts. In: S. Knoth, & W. Schmid (Eds.), *Frontiers in Statistical Quality Control* (Vol. 11, pp. 95–117). Switzerland: Springer International Publishing.
- Lindley, D. V. (1952). The theory of queues with a single server. *Mathematical Proceedings of the Cambridge Philosophical Society, 48*, 277–289.
- Montgomery, D. C. (2009). *Introduction to Statistical Quality Control* (6th. edition). New York: John Wiley & Sons.
- Morais, M. C. (2016). An ARL-unbiased np-chart. *Economic Quality Control, 31* 11–21.
- Morais, M. C. (2017). ARL-unbiased geometric and CCC_G control charts. *Sequential Analysis, 36*, 513–527.
- Morais, M. C. & Knoth, S. (2018). On ARL-unbiased charts to monitor the traffic intensity of a single server queue. In: S. Knoth and W. Schmid (Eds.), *Frontiers in Statistical Quality Control* (Vol. 12, pp. 87–112). Switzerland: Springer International Publishing.

- Morais, M.C. and Knoth, S. (2020). Improving the ARL profile and the accuracy of its calculation for Poisson EWMA charts. *Quality and Reliability Engineering International, 36*, 876–889.
- Morais, M.C., Knoth, S. and Weiß, C.H. (2018). A thinning-based EWMA chart to monitor counts. *Sequential Analysis, 37*, 487–510.
- Morais, M. C., & Pacheco, A. (2016). On stochastic ordering and control charts for the traffic intensity. *Sequential Analysis, 35*, 536–559.
- Nadarajah, S., & Kotz, S. (2005). On the linear combination of exponential and gamma random variables. *Entropy, 7*, 161–171.
- Paulino, S., Morais, M. C., & Knoth, S. (2016). An ARL-unbiased c-chart. *Quality and Reliability Engineering International, 32*, 2847–2858.
- Paulino, S., Morais, M. C. & Knoth, S. (2019). On ARL-unbiased c-charts for INAR(1) Poisson counts. *Statistical Papers, 60*, 1021–1038.
- Pignatiello, J. J., Jr., Acosta-Mejía, C. A., & Rao, B. V. (1995). The performance of control charts for monitoring process dispersion. In *4th Industrial Engineering Research Conference* (pp. 320–328).
- R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. `http://www.R-project.org`
- Rao, S S., Bhat, U. N., & Harishchandra, K. (1984). Control of traffic intensity in a queue – a method based on SPRT. *Opsearch, 21*, 63–80.
- Santos, M. D. M. (2016). ??On Control Charts and the Detection of Increases in the Traffic Intensity of Queueing Systems. M. Sc. thesis, Instituto Superior Técnico, Universidade de Lisboa.
- Santos, M., Morais, M. C. & Pacheco, A. (2018). Comparing short-memory charts to monitor the traffic intensity of single server queues. *Stochastics and Quality Control, 33*, 1–21.
- Santos, M., Morais, M. C. & Pacheco, A. (2019). Comparing short and long-memory charts to monitor the traffic intensity of single server queues. *Stochastics and Quality Control, 34*, 9–18.
- Shore, H. (2000). General control charts for attributes. *IIE Transactions, 32*, 1149–1160.
- Shore, H. (2006). Control charts for the queue length in a G/G/s system. *IIE Transactions, 38*, 1117–1130.
- Zobu, M. & Sağlam, V. (2013). Control of traffic intensity in hyperexponential and mixed Erlang queueing systems with a method based on SPRT. *Mathematical Problems in Engineering*. Article ID 241241, 9 pages. Accessed from `http://www.hindawi.com/journals/mpe/2013/241241/` on 2015-11-11.