ARL-unbiased geometric control charts for high-yield processes

Manuel Cabral Morais

Department of Mathematics & CEMAT — IST, ULisboa, Portugal

IST — Lisboa, March 13, 2018
Warm up
- A few historical facts
- Charts for monitoring high-yield industrial processes
- Some variants

Eliminating the bias of the ARL function
- Relating ARL-unbiased charts and UMPU tests
- Illustrations

Final thoughts

ARL-unbiased geometric control charts for high-yield processes
Quality

- Fitness for use and conformance to requirements are the shortest and most consensual definitions of quality (Juran and Godfrey, 1999, p. 27; Crosby, 1979, p. 17).

- A curious fact that escapes most consumers nowadays: concerns about quality can be traced back to the Babylonian Empire (Gitlow et al., 1989, pp. 8–9).

Code of Hammurabi

Law 229: If a builder builds a house for a man and does not make its construction firm, and the house which he has built collapse and cause the death of the owner of the house, that builder shall be put to death.

This eye for an eye approach to quality was also adopted by Phoenicians inspectors, who eliminated any repeated violations of quality standards by chopping off the hand of the maker of the defective product (Gitlow et al., 1989, p. 9).

---

1 A Babylonian law code, dating back to about 1772BC, named after the sixth Babylonian king, who enacted it. It consists of 282 laws dealing with matters of contracts, terms of transactions or addressing household and family relationships such as inheritance, divorce, paternity and sexual behavior.
The founder of statistical process control (SPC)

- We have to leap to the 20th. century to meet the father of modern quality control.

Walter A. Shewhart (1891–1967)

- Shewhart altered the course of industry by celebrating a perfect marriage between statistics, engineering, and economics.

- In a historic memorandum of May 16, 1924, to his superiors at Bell Laboratories, we can find the lasting and tangible evidence of that marriage, the quality control chart.

- Control charts are used to track process performance over time and detect changes in process parameters, by plotting the observed value of a statistic against time and comparing it with a pair of control limits.

An obs. beyond the control limits indicate potential assignable causes responsible for changes in those parameters, thus, should be investigated and eliminated.
Nonconformity; defective item

- Each specification that is not satisfied by a unit of a product is considered a defect or nonconformity (Montgomery, 2009, p. 308).
- A unit with at least one defect is called a defective or nonconforming item.

Examples of high-yield industrial processes

Many processes produce defectives items at a rate less than 100 ppm.

- Wire bonding process in an integrated circuit assembly provides an electrical connection between a semiconductor die and the external leads (Chang and Gan, 2001).
- In the filling process in the manufacture of low voltage liquid crystal display units, an incompletely filled unit is regarded as a nonconforming item (Chan et al., 2003).
- Exhaust valves seats are force fitted by insertion into the head of an engine; incorrect installation can lead to an engine failure; the target defective rate is less than 50 ppm (Steiner et al., 2004).
Charts for monitoring high-yield industrial processes

Geometric (or CCC) chart with 3-σ limits

- The most popular procedure to monitor high-yield industrial processes can be traced back to Calvin (1983).

- **Control statistic:** cumulative count of conforming (CCC) items between the $(t - 1)^{th}$ and $t^{th}$ nonconforming units, $X_t$.

- **Distribution:** $X_t \sim \text{geometric}(p)$; $P_p(x) = (1 - p)^x p$, $x \in \mathbb{N}_0$.

- **(Unknown) parameter:** $p$ (fraction nonconforming).

- **Target value of $p$:** $p_0$.

- **True value of $p$:** $\rho \times p_0$ (0 < $\rho$ < $1/p_0$; $\rho$ magnitude of the shift).

- **3 – $\sigma$ control limits:**

$$LCL = (1 - p_0)/p_0 - 3 \sqrt{\frac{1 - p_0}{p_0^2}}; \quad UCL = (1 - p_0)/p_0 + 3 \sqrt{\frac{1 - p_0}{p_0^2}}.$$

- Triggers a signal and we **deem the process (mean) out-of-control** at sample $t$ if $X_t \notin [LCL, UCL]$.

- Plot the number of conforming units (between consecutive defective units) on a **logarithmic scale** to accommodate large values of $X_t$. 
Example 1

- Wire bonding process in an integrated circuit
  \( p_0 = 10^{-4} \) (target fraction nonconforming)

- Simulated data: first 50 obs. of \( X_t \) — process in-control;
  last 20 obs. of \( X_t \) — process out-of-control (\( p = 2 \times p_0 \)).

- \( LCL = \left[ \max \left\{ 0, \frac{1-p_0}{p_0} - 3 \sqrt{\frac{1-p_0}{p_0^2}} \right\} \right] = 0 \)

- \( UCL = \left[ \frac{1-p_0}{p_0} + 3 \sqrt{\frac{1-p_0}{p_0^2}} \right] = 39997 \)

- Geometric chart

One false alarm: unit 506313 + 45 \( (x_{45} = 70728) \)

One valid signal: unit 678446 + (50 + 10) \( (x_{60} = 55600) \)
Example 1 (cont’d)

Parallels with a repeated hypothesis test...

- \( H_0 : p = p_0 \) (process in-control); \( H_1 : p \neq p_0 \) (process out-of-control)
- Control statistic: \( X \sim \text{geometric}(p), \ t \in \mathbb{N} \)
- Rejection region: \( W = (0, LCL) \cup (UCL, +\infty) \)
- Exact power function: \( \xi(\rho) = P_{\rho \ p_0}(X \in W), \ \rho \in (0, 1/p_0) \)

Problems

- minimum of \( \xi(\rho) \) not achieved at \( \rho = 1; \ \xi(\rho) < \xi(1), \ \rho > 1 \)
- significance level: \( \xi(1) \approx 0.018316 \neq 0.0027 \approx 1 - [\Phi(3) - \Phi(-3)] \).
Performance

- **Run length (RL)** — number of units inspected taken until a signal is triggered: \( RL(\rho) \sim \text{Geometric}(\xi(\rho)) \).

- The performance is frequently assessed in terms of \( ARL(\rho) = 1/\xi(\rho) \).
  It is desirable that false alarms (resp. valid signals) are rarely triggered (resp. emitted as quickly as possible), corresponding to a large in-control (resp. small out-of-control) ARL.

- It is crucial to swiftly detect not only increases but also decreases in \( p \).
  Increases in \( p \) mean process deterioration.
  Decreases in \( p \) represent process improvement (to be noted and possibly incorporated). It can also mean that a new inspector may not have been trained properly to inspect the process output.

- Keep in mind that \( LCL = \left\lceil \frac{(1 - p_0)/p_0 - k \sqrt{(1 - p_0)/p_0^2}}{p_0} \right\rceil = 0 \Leftrightarrow \sqrt{1 - p_0} < k \) and holds as long as \( k \geq 1 \).
  Thus, \( ARL(\rho) > ARL(1) \), for some \( \rho \), i.e., the chart triggers false alarms more frequently than some valid signals.
Variants to mitigate the poor performance of the geometric chart

- Xie and Goh (1997) recommended the use of exact probability limits:
  \[ L_{CL\alpha} = \frac{\ln(1 - \alpha/2)}{\ln(1 - p_0)}, \quad \text{and} \quad U_{CL\alpha} = \frac{\ln(\alpha/2)}{\ln(1 - p_0)}, \]
  where \( \alpha \) represents the acceptable risk of false alarm.

- Zhang et al. (2004) suggested:
  - taking \( L \in A = \{1, \ldots, L_{CL\max}\} \), with \( L_{CL\max} = \lfloor \frac{\ln(1 - \alpha)}{\ln(1 - p_0)} \rfloor \);
  - finding \( U : P_{\rho p_0}(X < L) + P_{\rho p_0}(X > U) \approx \alpha \), for each \( L \in A \);
  - defining the set \( C \) of all such pairs of control limits \((L, U)\);
  - choosing \((L^*, U^*) \in C\) that most nearly equalizes the tail probab., i.e.,
    \[
    |P_{p_0}(X < L^*) - P_{p_0}(X > U^*)| = \min_{(L, U) \in C} |P_{p_0}(X < L) - P_{p_0}(X > U)|.
    \]
  \[
  ARL^*(\rho) = \left[ P_{\rho p_0}(X < L^*) + P_{\rho p_0}(X > U^*) \right]^{-1}, \quad 0 < \rho < 1/p_0.
  \]

Both charts are ARL-biased: ARL does not attain a maximum at \( \rho = 1 \).
Moreover, the in-control ARL does not coincide with \( \alpha^{-1} \), a pre-specified value.
Yet Zhang et al. (2004) calls it a nearly ARL-unbiased geometric chart.
Basic facts

- A size $\alpha$ test for $H_0 : p = p_0$ against $H_1 : p = \rho p_0 \neq p_0$, with power function $\xi(\rho)$, is said to be unbiased if $\xi(1) \leq \alpha$ and $\xi(\rho) \geq \alpha$, for $\rho \neq 1$. The test is at least as likely to reject under any alternative as under $H_0$;

$$ARL(1) \geq \alpha^{-1} \quad \text{and} \quad ARL(\rho) \leq \alpha^{-1}, \rho \neq 1.$$

- If we consider $C$ a class of tests for $H_0 : p = p_0$ against $H_1 : p \neq p_0$, then a test in $C$, with power function $\xi(\rho)$, is a uniformly most powerful (UMP) class $C$ test if $\xi(\rho) \geq \xi'(\rho)$, for every $p \neq p_0$ and every $\xi'(\rho)$ that is a power function of a test in class $C$.

- In this situation there is no UMP test, but there is a test which is UMP among the class of all unbiased tests — the uniformly most powerful unbiased (UMPU) test.

- The concept of an ARL-unbiased Shewhart-type chart is related to the notion of UMP test.
The **UMPU test** derived by Lehmann (1986, pp. 135–136) for a real-valued parameter \( p \) in the exponential family uses the critical function

\[
\phi(x) = P(\text{Reject } H_0 : p = p_0 \mid X = x) = \begin{cases} 
1 & \text{if } x < LCL \text{ or } x > UCL \\
\gamma_{LCL} & \text{if } x = LCL \\
\gamma_{UCL} & \text{if } x = UCL \\
0 & \text{if } LCL < x < UCL 
\end{cases}
\]

\( LCL, UCL, \gamma_{LCL}, \) and \( \gamma_{UCL} \) are such that:

\[
E_{p_0}[\phi(X)] = \alpha \quad \text{(prob. of false alarm = \( \alpha \))}
\]

\[
E_{p_0}[X \phi(X)] = \alpha E_{p_0}(X) \quad \text{(unbiased ARL)}.
\]

In the geometric case, these eq. are equivalent to

\[
\gamma_{LCL} \times P_{p_0}(LCL) + \gamma_{UCL} \times P_{p_0}(UCL) = \alpha - \left[ 1 - \sum_{x=LCL}^{UCL} P_{p_0}(x) \right] \quad (1)
\]

\[
\gamma_{LCL} \times LCL \times P_{p_0}(LCL) + \gamma_{UCL} \times UCL \times P_{p_0}(UCL) = \alpha \times E_{p_0}(X) - \left[ E_{p_0}(X) - \sum_{x=LCL}^{UCL} x \times P_{p_0}(x) \right]. \quad (2)
\]

However, equations (1) and (2) are not sufficient to define two control limits and two randomization probabilities.
Characterizing the ARL-unbiased geometric chart

Inspired by this UMPU test, we defined a *geometric chart that triggers a signal* with:

- **probability one** if the obs. number of conforming units between two consecutive nonconforming units, \( x \), is below LCL or above UCL;
- **probability** \( \gamma_{LCL} \) (resp. \( \gamma_{UCL} \)) if \( x = LCL \) (resp. \( x = UCL \)).

Furthermore,

- **the randomization probabilities** are the solutions of the system of two linear equations (1) and (2):

\[
\gamma_{LCL} = \frac{de - bf}{ad - bc} \quad \text{and} \quad \gamma_{LCL} = \frac{af - ce}{ad - bc},
\]

where

\[
a = P_{p_0}(LCL), \quad b = P_{p_0}(UCL), \quad c = LCL \times P_{p_0}(LCL),
\]
\[
d = UCL \times P_{p_0}(UCL), \quad e = \alpha - 1 + \sum_{x=LCL}^{UCL} UCL x \times P_{p_0}(x),
\]
\[
f = \alpha \times E_{p_0}(X) - E_{p_0}(X) + \sum_{x=LCL}^{UCL} UCL x \times P_{p_0}(x).
\]
Characterizing the ARL-unbiased geometric chart (cont’d)

- To rule out pairs of control limits leading to \((\gamma_{LCL}, \gamma_{UCL}) \notin (0, 1)^2\), the useful \((LCL, UCL)\) are restricted to the following grid of non-negative integer numbers:
  \[\{(LCL, UCL) : L_{min} \leq LCL \leq L_{max}, U_{min} \leq UCL \leq U_{max}\}\]

- The search for admissible values for \((\gamma_{LCL}, \gamma_{UCL})\) starts with \((LCL, UCL) = (L_{min}, U_{min})\) and stops as soon as an admissible solution is found.
  \[L_{min} = \max \left\{F^{-1}(\max\{0, F(U_{min} - 1) - 1 + \alpha\}), G^{-1}(\max\{0, G(U_{min} - 1) - 1 + \alpha\})\right\}\]
  \[L_{max} = \min \left\{\bar{F}^{-1}(\alpha), \bar{G}^{-1}(\alpha)\right\}\]
  \[U_{min} = \max \left\{F^{-1}(1 - \alpha), G^{-1}(1 - \alpha)\right\}\]
  \[U_{max} = \min \left\{\bar{F}^{-1}(\min\{1, F(L_{max}) + 1 - \alpha\}), \bar{G}^{-1}(\min\{1, G(L_{max}) + 1 - \alpha\})\right\}\]

- \[F(x) = P_{p_0}(X \leq x)\]
- \[G(x) = \frac{1}{E_{p_0}(X)} \sum_{i=0}^{x} i \times P_{p_0}(X = i)\]
- \[F^{-1}(\alpha) = \min\{x \in \mathbb{N}_0 : F(x) \geq \alpha\}\]
- \[\bar{F}^{-1}(\alpha) = \min\{x \in \mathbb{N}_0 : F(x) > \alpha\}\]
- \[G^{-1}(\alpha) = \min\{x \in \mathbb{N}_0 : G(x) \geq \alpha\}\]
- \[\bar{G}^{-1}(\alpha) = \min\{x \in \mathbb{N}_0 : G(x) > \alpha\}\]

(For the rationale and detailed derivation of this grid see Paulino et al., 2016a, Appendix C.)
**ARL function**

A signal is triggered by the ARL-unbiased geometric chart with probability

\[
\xi_{\text{unbiased}}(\rho) = \left[ 1 - \sum_{x=LCL}^{UCL} P_{\rho \ p_0}(x) \right] + \gamma_{LCL} \times P_{\rho \ p_0}(LCL) + \gamma_{UCL} \times P_{\rho \ p_0}(UCL).
\]

Its corresponding ARL function:

\[
\text{ARL}_{\text{unbiased}}(\rho) = \frac{1}{\xi_{\text{unbiased}}(\rho)}.
\]

**Randomization of the emission of the signal**

Can be done in practice by incorporating the generation of a pseudo-random number from a Bernoulli distribution with parameter \(\gamma_{LCL}\) (resp. \(\gamma_{LCL}\)) in the software used to monitor the data fed from the production line, whenever the observed number of nonconforming items is equal to \(LCL\) (resp. \(UCL\)).
Example 2 — ARL-unbiased geometric charts

- \( p_0 = 10^{-i}, \ i = 5, 4, 3, 2 \)
- in-control ARL = 200, 370.4 \( (\alpha = 0.005, 0.0027) \)
- Limits of the search grid, control limits and randomization probabilities

<table>
<thead>
<tr>
<th>( p_0 )</th>
<th>( L_{\text{min}} )</th>
<th>( L_{\text{max}} )</th>
<th>( L_{\text{CL}} )</th>
<th>( U_{\text{min}} )</th>
<th>( U_{\text{max}} )</th>
<th>( U_{\text{CL}} )</th>
<th>( \gamma_{\text{LCL}} )</th>
<th>( \gamma_{\text{UCL}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00001</td>
<td>441</td>
<td>501</td>
<td>441</td>
<td>743009</td>
<td>743294</td>
<td>743230</td>
<td>0.792137</td>
<td>0.754626</td>
</tr>
<tr>
<td>0.0001</td>
<td>44</td>
<td>50</td>
<td>44</td>
<td>74298</td>
<td>74326</td>
<td>74319</td>
<td>0.177234</td>
<td>0.318435</td>
</tr>
<tr>
<td>0.001</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>7426</td>
<td>7430</td>
<td>7428</td>
<td>0.415872</td>
<td>0.349557</td>
</tr>
<tr>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>739</td>
<td>739</td>
<td>739</td>
<td>0.440987</td>
<td>0.207035</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( p_0 )</th>
<th>( L_{\text{min}} )</th>
<th>( L_{\text{max}} )</th>
<th>( L_{\text{CL}} )</th>
<th>( U_{\text{min}} )</th>
<th>( U_{\text{max}} )</th>
<th>( U_{\text{CL}} )</th>
<th>( \gamma_{\text{LCL}} )</th>
<th>( \gamma_{\text{UCL}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00001</td>
<td>240</td>
<td>270</td>
<td>240</td>
<td>812554</td>
<td>812706</td>
<td>812674</td>
<td>0.736799</td>
<td>0.103324</td>
</tr>
<tr>
<td>0.0001</td>
<td>24</td>
<td>27</td>
<td>24</td>
<td>81252</td>
<td>81267</td>
<td>81263</td>
<td>0.072600</td>
<td>0.166090</td>
</tr>
<tr>
<td>0.001</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8122</td>
<td>8123</td>
<td>8122</td>
<td>0.406312</td>
<td>0.224264</td>
</tr>
<tr>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>808</td>
<td>808</td>
<td>808</td>
<td>0.240561</td>
<td>0.010422</td>
</tr>
</tbody>
</table>

Expectedly, the search grid/control limits grow comparatively larger, as we deal with smaller values of the target fraction nonconforming.
**Example 3 — Nearly ARL-unbiased vs. ARL-unbiased geometric charts**

- $p_0 = 10^{-3}$, $\alpha = 0.005$

- **Relative gain** in the ARL when we replace the nearly ARL-unbiased design with the ARL-unbiased chart: $[1 - \frac{\text{ARL}(\rho, \gamma_{LCL}, \gamma_{UCL})}{\text{ARL}^*(\rho)}] \times 100\%$.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>nearly unbiased</th>
<th>unbiased</th>
<th>% gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>29.6309</td>
<td>37.6573</td>
<td>-27.0878%</td>
</tr>
<tr>
<td>0.8</td>
<td>138.8971</td>
<td>162.7097</td>
<td>-17.1440%</td>
</tr>
<tr>
<td>0.9</td>
<td>178.4955</td>
<td>191.8332</td>
<td>-7.4723%</td>
</tr>
<tr>
<td>1.0</td>
<td>199.9869</td>
<td>200.0000</td>
<td>-0.0066%</td>
</tr>
<tr>
<td>1.1</td>
<td>204.1971</td>
<td>194.9502</td>
<td>4.5285%</td>
</tr>
<tr>
<td>1.2</td>
<td>198.2454</td>
<td>184.4424</td>
<td>6.9626%</td>
</tr>
<tr>
<td>1.5</td>
<td>166.1584</td>
<td>151.0359</td>
<td>9.1013%</td>
</tr>
</tbody>
</table>

- The randomization of the emission of a signal allows the ARL-unbiased geometric chart to take less time to detect decreases in $p$ than to trigger a false alarm, even if we are dealing with a null LCL.

- The out-of-control ARL performance of the nearly ARL unbiased and the ARL-unbiased geometric charts differ markedly when $p_0 \geq 10^{-3}$.

- The ARL-unbiased geometric chart is more (resp. less) sensitive to proc. deterioration (resp. improvement) than the nearly ARL-unbiased geom. chart.
**ARL-unbiased geometric chart**

As opposed to the geometric chart with 3-sigma control limits and its variants...

- The ARL-unbiased geometric chart can take a **pre-specified in-control ARL**.
- The associated **ARL curve attain a maximum when \( p \) is on target**.
- It **tackles the curse of the null LCL** and detects decreases in \( p \) in a timely fashion, by relying on the **randomization probabilities**.
- The ARL-unbiased geometric chart is also **RL-unbiased**: all out-of-control RL quantile curves are pointwise below the corresponding in-control RL quantile curve, i.e., all out-of-control RL are stochastically smaller (in the usual sense) than the in-control RL.
**ARL-unbiased CCC₆ chart**

The use of the geometric chart is limited to situations where units of product are inspected sequentially...

The cumulative count conforming (CCC) chart under group (G) inspection uses as its control statistic the cumulative number of samples of size \( n \) inspected until a nonconforming sample is encountered \((Y)\).

\[
P(Y = y) = P_{p,n}(y) = [(1 - p)^{n}]^{y-1} [1 - (1 - p)^{n}], \quad y \in \mathbb{N}.
\]

\( Y \) is geometric r.v., taking values in \( \mathbb{N} \) and with param. \( p_n = 1 - (1 - p)^n \).

To define an ARL-unbiased CCC₆ chart, we need to:

- obtain an ARL-unbiased geometric chart with false alarm rate \( \alpha \) and a target value \( 1 - (1 - p_0)^n \);
- add one unit to the control limits of the ARL-unbiased geometric chart and consider the same randomization probabilities.

ARL-unbiased geometric control charts for high-yield processes
By relying on the long and successful history of control charting, we believe that the **ARL-unbiased geometric and \( CCC_G \) charts** have the potential to play a major role in the **timely detection of the deterioration and improvement of real high-yield processes**.

**Future work**

Derive...

- **ARL-unbiased versions** of the existing **\( CCC - r \) charts** to monitor the cumulative count of items until the \( r^{th} \) nonconforming item,\(^2\) suchlike the ones discussed by Xie *et al.* (1998), ..., Albers (2010).

- **ARL-unbiased designs** of the **CUSUM** (cumulative sum) and **EWMA** (exponentially weighted moving average) charts proposed by Chang and Gan (2001) and Yeh *et al.* (2008) for geometric output.

\(^2\) \( X \) has a negative binomial distribution with parameters \( r \) and \( p \).
Related SPC papers by submission date


ARL-unbiased geometric control charts for high-yield processes


