Monitoring in Changing Environments: Discrete Case

Ansgar Steland[0000−1111−2222−3333] and
Ewaryst Rafajłowicz[1111−2222−3333−4444]

Abstract Abstract

1 Introduction

Let U_t denote a control statistics, e.g. the Shewhart chart statistic $U_t = \frac{X_t - \bar{\mu}_0}{\sigma_0}$, where $\bar{\mu}_0$ denotes the in-control mean and σ_0 the standard deviation asssumed to be known. Suppose the monitoring procedure gives a signal if U_t exceeds a threshold c. A common approach is to select c such that the false alarm rate, p_f, defined as the probability to give a signal although the in-control situation of no change holds, equals a prespecified value $\alpha \in (0, 1)$ at each time instant t, i.e.

$$p_f = P_\infty(U_t > c) = \alpha$$

where notation P_∞ indicates that the probability is calculated under the no-change null hypothesis corresponding to normal conditions. For i.i.d. Gaussian data one has $p_f = 1 - \Phi(c)$ where $\Phi(x), x \in \mathbb{R}$, denotes the distribution function of the standard normal law, so that the choice $c = \Phi^{-1}(1 - \alpha)$ leads to a false-alarm rate α, where $\Phi^{-1}(x), x \in (0, 1)$, stands for the quantile function associated to $\Phi(x)$. The classical Shewhart chart monitors the sequence $X_1, X_2, \cdots$ and signals an alarm, if the alarm event $\{U_t > c\}$ occurs for the first time. The also called average run length (ARL) or mean time to a signal, which is false alarm under P_∞, is given by ARL $= \frac{1}{\alpha}$. This simple charting procedure is known to sensitive with respect to change-points where the mean of the data jumps to a new value. (REFERENCES)

Institute of Statistics and AI Center
Pontdriesch 14-16, e-mail: steland@stochastik.rwth-aachen.de

Ewaryst ...
Name, Address of Institute e-mail: name@email.address

Frequently, however, there exist external random variabes, Z_t, representing the environment (or framework) of the monitoring procedure and potentially affecting the law of the observables X_t. Further, the environment often also influences how we want to evaluate a deviation U_t from normal behaviour, in the sense that we would like to change the design of the decision procedure, which goes beyond taking into account of the dependence of the distribution of X_t on Z_t by appropriately standardizing the measurements X_t. For example, relevant measurements of a safety system for a vehicle which intervenes in the steering may not only depend on the vehicle's velocity, but at high velocities the system should be substantially more sensitive to departures from normal behaviour indicating a safety problem than for low velocities, since even small misinterventions may result in serious accidents. Another interesting example is watermark detection for LLNs to identify AI-generated text. Here, the token generated in each step is drawn only from a randomly generated green set of tokens, whereas tokens from the complimentary red set are avoided. The green and red sets are created randomly using a random seed depending on last token, so that these sets are reproducible by a user. Hence, one may check for the presence of the watermark by considering the fraction of tokens lying in the green set. For high entropy parts of a text, the watermark effect on the generated output is negligible, whereas for low entropy parts, where the next token is almost deterministic, the output can be quite poor. Thus, the sensitivity of a watermark detector should be concentrated on high entropy parts, since the watermark is not detectable in low entropy parts anyway.

In such cases it is reasonable to select the threshold as a function of Z_t, in order to take the available information on the environment into account. Let us assume that the in-control distribution of X_t is Gaussian with mean function $\mu_0(Z_t)$ and standard deviation σ_0. This means, $\mu_0(z)$ describes the mean of X_t given the environment $Z_t = z$. To this end, we assume for simplicity of exposition that σ_0 is known. The corresponding Shewhart chart statistic now reads as $U_t = \frac{X_t - \mu_0(Z_t)}{\sigma_0}$ and follows under normal conditions a standard normal law given Z_t. The environment-adapted procedure gives a signal, if

$$U_t > c(Z_t)$$

for the first time. The question arises how one can now choose the threshold function $c(\cdot)$ in such a way that the procedure guarantees a certain marginal false alarm rate, so that

$$p_f = P(U_t > c(Z_t)) = \alpha.$$

In this general formulation, there are infinitely many functions $c(z)$ which solve this constraint. By expressing p_f in terms of the conditional alarm rate

$$\pi(z) = P(U_t > c(z) | Z_t = z),$$

which is governed by a normal law, and the distribution of Z_t, we discuss this problem in Section 1.1.

Let us assume that the law of our measurements depends on Z_t also in terms of the (conditional) variance, such that now $\sigma_0 = \sigma_0(Z_t)$ becomes a function of

Z_t. Obviously, this setting is covered by the above approach. Specifically, now the constant threshold $c(z) = c$ leads again to the choice $c = \Phi^{-1}(1-\alpha)$. Nevertheless, if we want to adapt the design of the decision procedure to behave differently for different values of Z_t, we are again led to the problem to find a function $c(z)$ such that $P(U_t > c(Z_t)) = \alpha$ holds, where now $U_t = \frac{X_t - \mu_0(Z_t)}{\sigma_0(Z_t)}$.

LLN EXAMPLE (long version): Another interesting example is watermark testing and monitoring in large language models (LLNs). Here the aim is to be able to detect reliably a watermark to identify AI-generated content, e.g., in social media. In [**?**] the authors propose to implement a watermark in LLNs by the following procedure. Recall that LLNs generate output by iteratively adding a new token s_t to the previous tokens $s_{t-N}, \ldots, s_{t-1}$ by drawing s_t from a distribution μ over the alphabet $\mathcal{A}$ of tokens, which depends on $s_{t-N}, \ldots, s_{t-1}$. The mapping $(s_{t-N}, \ldots, s_{t-1}) \mapsto \mu((s_{t-N}, \ldots, s_{t-1})$ is learned from a training sample. A watermark is now implemented by dividing randomly in each step the alphabet $\mathcal{A}$ in two sets $\mathcal{A}_r$ (red set) and $\mathcal{A}_g$ (green set) of equal size and drawing s_t only from the green set $\mathcal{A}_g$. The sets are randomly generated using a random seed determined from a hash function of the past tokens $(s_{t-N}, \ldots, s_{t-1})$. Thus, when checking whether an image or text comes from an LLNs, i.e. has a watermark, the sets $\mathcal{A}_r$ and $\mathcal{A}_g$ can be reproduced from the past tokens, and one can check whether the watermark rule $s_t \in \mathcal{A}_g$ is violated or now. Under the null hypothesis that the text is not generated from the LLN, $s_t \in \mathcal{A}_g$ holds with probability $\frac{1}{2}$. If we group the token sequence in successive blocks of length M and let X_t be the fraction of green set tokens, the standardized statistic $U_t = 2\sqrt{M}\frac{X_t - 1/2}{2}$ is approximately normal for large M, and we decide in favor of the alternative that a watermark is present, if $U_t > c$. The influence of the watermark on the quality of the generated output depends on the entropy of the token sequence. For high entropy parts the effect is negligible in practice, whereas for low entropy parts of the sequence the next token s_t is almost deterministic, i.e., the distribution μ has a peak at s_t, such that implementing the watermark leads to poor performance of the LLN. This suggests to let Z_t be an estimator of the current block's entropy and to select $c = c(Z_t)$ such that the watermark detector has high sensitivity for large values of Z_t and low sensitivity for high values.

1.1 Level α Threshold Functions

Let us assume that Z_t follows a discrete distribution on a set $\{z_1, \ldots, z_K\}$ with density function $p(z)$, $z \in \mathbb{R}$, and probabilities $p_k = P(Z_t = z_k)$, $1 \leq k \leq K$, for some $K \in \mathbb{N}$. Let $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$. Then

$$
\begin{aligned}
p_f &= P\left(\frac{X_t - \bar{\mu}_0}{\sigma_0} > c(Z_t)\right) \\
&= \sum_{k=1}^{K} p_k P\left(\frac{U_t - \bar{\mu}_0}{\sigma_0} > c(z)\right) \\
&= 1 - \sum_{k=1}^{K} p_k \Phi(c(z_k))
\end{aligned}
$$

If we let $c_k = \Phi(c(z_k))$, $1 \le k \le K$, then we arrive at the problem to determine K nonnegative numbers $c_1, \dots, c_K$ such that

$$
\sum_{k=1}^{K} p_k c_k = 1 - \alpha. \tag{1}
$$

There exist infinitely many solutions and the following lemma collects some special solutions and two general approaches to construct c-functions solving (1). There exists a constant solution and a solution with $c_k \sim p_k$. Further, one can consider sub probability densities summing up to $1 - \alpha$, and one can construct solutions geometrically.

For the geometric approach, recall the following facts. The angle between two K-vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as $\angle(\mathbf{u}, \mathbf{v}) = \arccos\left(\frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2}\right)$. The angle, γ, between the given vector $\mathbf{p} = (p_1, \dots, p_K)^\top$ and the vector $\mathbf{c} = (c_1, \dots, c_K)^\top$ to be determined needs to satisfy

$$
\gamma = \angle(\mathbf{p}, \mathbf{c}) = \arccos\left(\frac{1 - \alpha}{\|\mathbf{p}\|_2 \|\mathbf{c}\|_2}\right).
$$

If $\boldsymbol{c}$ is a unit vector, then γ depends only on α and $\boldsymbol{p}$, and one obtains solutions by rotating $\boldsymbol{p}$.

Lemma 1 *Given a probability vector* $\mathbf{p}$ *equation* (1) *can be solved as follows.*

(i) The constant solution given by

$$
\boldsymbol{c}_\alpha^* = (1 - \alpha)\mathbf{1}_K,
$$

leads to the c-function

$$
c_\alpha^*(z_k) = \Phi^{-1}(1 - \alpha), \qquad 1 \le k \le K.
$$

(ii)A proportional solution is given by

$$
\boldsymbol{c}_{prop}^* = \frac{1 - \alpha}{\|\boldsymbol{p}\|_2^2} \boldsymbol{p}
$$

with c-function

$$c^*_{prop}(z_k) = \Phi^{-1}\left(\frac{1-\alpha}{\|\boldsymbol{p}\|_2^2} p_k\right), \qquad 1 \le k \le K.$$

(iii) If $1 - \|\boldsymbol{p}\|_2 < \alpha < 1 + \|\boldsymbol{p}\|_2$, then unit vectors solve equation (1)*. Fix some $K \times K$ dimensional rotation matrix $\mathbf{A}_K(\gamma)$ and let*

$$\mathbf{c}^* = \mathbf{A}_K(\gamma)\mathbf{p}/\|\mathbf{p}\|_2.$$

The associated function $c(z)$ is then determined at all points z_k by

$$c(z_k) = \Phi^{-1}(c^*_k), \qquad 1 \le k \le K.$$

(iv) Let $\mathbf{g} = (g_1, \dots, g_K)^\top$ be a sub probability density dominated by $\boldsymbol{p}$, i.e., a vector of nonnegative real numbers with $0 < g_k < p_k$ for all k and $\sum_{k=1}^K g_k = 1 - \alpha$. Then $\mathbf{c}^*_g = (c^*_{g1}, \dots, c^*_{gK})^\top$ with entries defined by

$$c^*_{gk} = \Phi^{-1}\left(\frac{g_k}{p_k}\right)\mathbf{1}_{\{p_k>0\}} + \Phi^{-1}(0)\mathbf{1}_{\{g_k=0\}}, \qquad 1 \le k \le K,$$

provides a solution vector $\mathbf{c}^*_g$, and the associated c-function, c^*, is determined at all possible realizations z_k of Z_t by

$$c^*(z_k) = \Phi(c^*_{gk}), \qquad 1 \le k \le K.$$

Proof. Obviously, the constant solution (i) ensures $\boldsymbol{c}_\alpha^{*\top}\boldsymbol{p} = (1-\alpha)\sum_{k=1}^K p_k = 1-\alpha$. Similarly, the proportional solution satisfies $\boldsymbol{c}_{prop}^{*\top}\boldsymbol{p} = \frac{(1-\alpha)}{\|\boldsymbol{p}\|_2^2}\sum_{i=1}^K p_i^2 = 1-\alpha$. To see (iii), first note that the condition $1 - \|\boldsymbol{p}\|_2 < \alpha < 1 + \|\boldsymbol{p}\|_2$ ensures that $\frac{1-\alpha}{\|\boldsymbol{p}\|2} \in (-1, 1)$. Any unit vector $\boldsymbol{c}$ with $\angle(\boldsymbol{c}, \boldsymbol{p}) = \gamma := \arccos\left(\frac{1-\alpha}{\|\boldsymbol{p}\|}\right)$, i.e., with $\arccos\left(\frac{\boldsymbol{c}^\top\boldsymbol{p}}{\|\boldsymbol{p}\|}\right) = \gamma$, necessarily satisfies $\boldsymbol{c}^\top\boldsymbol{p} = 1 - \alpha$. Hence, any such $\boldsymbol{c}$ solves the given problem. Now, the specific choice $\boldsymbol{c}^* = \boldsymbol{A}_K(\gamma)\frac{\boldsymbol{p}}{\|\boldsymbol{p}\|}$ satisfies $\angle(\boldsymbol{c}^*, \boldsymbol{p}) = \gamma$ by construction and thus provides a solution. It remains to verify (iv). We have

$$\sum_{k=1}^K p_k\Phi(c^*_{gk}) = \sum_{1\le k\le K: p_k>0} p_k \Phi\circ\Phi^{-1}\left(\frac{g_k}{p_k}\right) = \sum_{1\le k\le K: p_k>0} p_k\frac{g_k}{p_k} = \sum_{k=1}^K g_k = 1-\alpha.$$

□

Remark 1 If in (i) and (ii) the constant $1-\alpha$ is replaced by some larger value, one still obtains a level α test, since then (1) holds with = replaced by ≥, so that $p_f \le \alpha$. The same applies, if in (iv) one chooses $0 < g_k < p_k$, $1 \le k \le K$, with $\sum_{k=1}^K g_k \ge 1-\alpha$.

Note that the constant solution selects in each state of the system the same threshold to guarantee the false alarm rate α. This solution is agnostic to the environment as it ignores the information Z_t. Let us consider the geometric approach. If there are two states, i.e. $K = 2$, there is a unique rotation matrix given by

$$\mathbf{A}_2(\gamma) = \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) \\ \sin(\gamma) & \cos(\gamma) \end{bmatrix}.$$

For $K > 2$ one may use any proper rotation from the special orthognal group, $SO(K)$, of degree K. Especially, one can simply embed the 2-dimensional rotation $\mathbf{A}_2$ as follows: Fix two coordinates $j, k \in \{1, \dots, K\}$ with $p_j p_k \neq 0$ and let $\mathbf{P}$ be a permutation matrix mapping (j, k) to $(1, 2)$. Recall that any permutation matrix is obtained by permuting the columns of the identity matrix. Now let

$$\mathbf{C}(\gamma) = \begin{bmatrix} \mathbf{A}_2(\gamma) & \mathbf{0}_{K-2}, \\ \mathbf{0}_{K-2}^\top & \mathbf{I}_{K-2} \end{bmatrix},$$

where for $l \in \mathbb{N}$ we put $\mathbf{0}_l = (0, \dots, 0)^\top \in \mathbb{R}^l$ and denote by $\mathbf{I}_l$ the l-dimensional unit matrix. The matrix

$$\mathbf{A}_K(\gamma) = \mathbf{P}^\top \mathbf{C}(\gamma) \mathbf{P}$$

performs a rotation by the angle γ.

However, the method of solution provided by (iv) may be most appropriate in applications. By selecting the vector $\mathbf{g}$ of probabilities which sum up to $1 - \alpha$ we can tailor the decision threshold $c(z)$.

2 Threshold designs controlling sensitivity

Now we study the problem to design the procedure in such a way that its sensitivity depends on the environmental information Z_t but keeping the overall false alarm rate α. Indeed, in many application one may be interested in specifying the decision procedure in such a way that its sensitivity to detect a change is higher for certain values of Z_t than for others. Here the idea is not to consider the sensitivity in terms of the detection power under alternatives. This issue is studied in next section. Instead, the aim is design the threshold function such that the false-alarm rate, which determines the sensitivity, varies depending on the value of Z_t. We consider two approaches. The first one guarantees the overall (marginal) false-alarm rate α and distributes it over a partition of the sampling space of Z_t by selecting appropriate threshold functions for each set of the partition. This is done by using prespecified probabilities assigned to the partioning sets, which sum up to the nominal false-alarm rate. The second approach conditions on the information Z_t and thus controls the conditional false-alarm rate at a given prespecified level α.

The details of the first approach are as follows. Select disjoint non-empty intervals I_l, $1 \leq l \leq L$, which partition the sampling space of the $Z_t's$, such that

$$\text{supp}(p) = \{z : p(z) > 0\} = I_1 \cup \dots \cup I_L.$$

Next, select false-alarm probabilities $\alpha_1, \dots, \alpha_L$ with

$$\sum_{l=1}^{L} \alpha_l \leq \alpha \tag{2}$$

satisfying

$$\alpha_l < p_l := \sum_{z \in I_l} p(z) \tag{3}$$

for $1 \leq l \leq L$. Let $g_l : \mathbb{R} \to \mathbb{R}$ be functions with $0 < g_l(z_k) < p_k$ for all $1 \leq k \leq K$ and

$$\sum_{z \in I_l} g_l(z) = p_l - \alpha_l, \tag{4}$$

$1 \leq l \leq L$. The proposed distributing threshold function is then defined by

$$c^*_{distr}(z) = \Phi^{-1}\left(\frac{g_l(z)}{p(z)}\right) \mathbf{1}_{\{p>0\}}(z) + \Phi^{-1}(0) \mathbf{1}_{\{p=0\}}(z), \qquad z \in I_l, 1 \leq l \leq L.$$

The following lemma shows that the resulting procedure defines a level α test with associated monitoring Shewhart chart attaining at least an average run length of $1/\alpha$.

Lemma 2 *Assume that conditions (2)–(4) hold. Then the threshold function c^*_{distr} has false-alarm rate*

$$P(U_t > c^*_{distr}(Z_t)) \leq \alpha.$$

Proof. Using the identity $c^*_{distr}(z) = \sum_{l=1}^{L} \mathbf{1}_{I_l}(z) \Phi^{-1}\left(\frac{g_l(z)}{p(z)}\right) = \Phi^{-1}\left(\sum_{l=1}^{L} \mathbf{1}_{I_l}(z) \frac{g_l(z)}{p(z)}\right)$ for all $z \in \operatorname{supp}(p)$, we obtain

$$\begin{aligned} P(U_t > c^*_{distr}(Z_t)) &= 1 - \sum_z P(U_t \leq c^*_{distr}(z)) p(z) \\ &= 1 - \sum_{\{z : p(z) > 0\}} \Phi(c^*_{distr}(z)) p(z) \\ &= 1 - \sum_{l=1}^{L} \sum_{z \in I_l \cap \{p>0\}} \frac{g_l(z)}{p(z)} p(z) \\ &= 1 - \sum_{l=1}^{L} \sum_{z \in I_l} g_l(z) \\ &= 1 - \sum_{l=1}^{L} (p_l - \alpha_l). \end{aligned}$$

Using $\sum_{l=1}^{L} p_l = 1$ and $\sum_{l=1}^{L} \alpha_l = \alpha$, we obtain $P(U_t > c^*_{distr}(Z_t)) = \alpha$. □

The second strategy is to control the conditional false-alarm rate given $Z_t \in I_k$ is observed for some $1 \leq k \leq L$. Fix target (conditional) false-alarm probabilities

$$\alpha_1, \ldots, \alpha_L \in (0, 1) \tag{5}$$

and let $g_l : \mathbb{R} \to \mathbb{R}$ be functions with $0 < g(z_k) < p(z_k)$, $1 \le k \le K$, and

$$\sum_{z \in I_l} g_l(z) = \alpha_l p_l, \tag{6}$$

where as above $p_l = \sum_{z \in I_l} p(z)$. Define

$$c^*_{cond}(z) = (1 - \Phi)^{-1} \left(\frac{g_l(z)}{p(z)} \right), \qquad z \in I_l, 1 \le l \le L.$$

Lemma 3 *Suppose that conditions (5) and (6) hold. Then the threshold function c^*_{cond} ensures the conditional false-alarm rates*

$$p_f(k) = P(U_t > c^*(Z_t) | Z_t \in I_k) = \alpha_k, \qquad 1 \le k \le L.$$

Proof. Clearly, $p_f(k) = \frac{P(U_t > c^*_{cond}(Z_t), Z_t \in I_k)}{p_k}$. We have

$$\begin{aligned} P(U_t > c^*_{cond}(Z_t), Z_t \in I_k) &= \sum_{z \in \{p>0\}} P(U_t > c^*_{cond}(z), z \in I_k) p(z) \\ &= \sum_{z \in \{p>0\}} P(U_t > c^*_{cond}(z)) \mathbf{1}_{I_k}(z) p(z) \\ &= \sum_{z \in I_k} (1 - \Phi)(c^*_{cond}(z)) p(z) \\ &= \sum_{z \in I_k \cap \{p>0\}} \frac{g_k(z)}{p(z)} p(z) \\ &= \alpha_k p_k. \end{aligned}$$

Hence $p_f(k) = \alpha_k$, $1 \le k \le K$. □

Acknowledgements This work was done during a research stay of the first author at Wrocław University of Science and Technology. He thanks the second author for his hospitality.

Appendix

3 Styling of References

References may be *cited* in the text either by number (preferred) or by author/year.[1] If the citatiion in the text is numbered, the reference list should be arranged in ascending order. If the citation in the text is author/year, the reference list should be

---

[1] Make sure that all references from the list are cited in the text. Those not cited should be moved to a separate *Further Reading* section or chapter.

sorted alphabetically and if there are several works by the same author, the following order should be used:

1. all works by the author alone, ordered chronologically by year of publication
2. all works by the author with a coauthor, ordered alphabetically by coauthor
3. all works by the author with several coauthors, ordered chronologically by year of publication.

The *styling* of references[2] depends on the subject of your book:

- The *two* recommended styles for references in books on *mathematical, physical, statistical and computer sciences* are depicted in [1, 2, 3, 4, 5] and [6, 7, 8, 9, 10].
- Examples of the most commonly used reference style in books on *Psychology, Social Sciences* are [11, 12, 13, 14, 15].
- Examples for references in books on *Humanities, Linguistics, Philosophy* are [16, 17, 18, 19, 20].
- Examples of the basic Springer Nature style used in publications on a wide range of subjects such as *Computer Science, Economics, Engineering, Geosciences, Life Sciences, Medicine, Biomedicine* are [21, 22, 24, 23, 25].

References

1. Broy, M.: Software engineering — from auxiliary to key technologies. In: Broy, M., Dener, E. (eds.) Software Pioneers, pp. 10-13. Springer, Heidelberg (2002)
2. Dod, J.: Effective substances. In: The Dictionary of Substances and Their Effects. Royal Society of Chemistry (1999) Available via DIALOG. http://www.rsc.org/dose/title of subordinate document. Cited 15 Jan 1999
3. Geddes, K.O., Czapor, S.R., Labahn, G.: Algorithms for Computer Algebra. Kluwer, Boston (1992)
4. Hamburger, C.: Quasimonotonicity, regularity and duality for nonlinear systems of partial differential equations. Ann. Mat. Pura. Appl. **169**, 321–354 (1995)
5. Slifka, M.K., Whitton, J.L.: Clinical implications of dysregulated cytokine production. J. Mol. Med. (2000) doi: 10.1007/s001090000086
6. J. Dod, in *The Dictionary of Substances and Their Effects*, Royal Society of Chemistry. (Available via DIALOG, 1999), http://www.rsc.org/dose/title of subordinate document. Cited 15 Jan 1999
7. H. Ibach, H. Lüth, *Solid-State Physics*, 2nd edn. (Springer, New York, 1996), pp. 45-56
8. S. Preuss, A. Demchuk Jr., M. Stuke, Appl. Phys. A **61**
9. M.K. Slifka, J.L. Whitton, J. Mol. Med., doi: 10.1007/s001090000086
10. S.E. Smith, in *Neuromuscular Junction*, ed. by E. Zaimis. Handbook of Experimental Pharmacology, vol 42 (Springer, Heidelberg, 1976), p. 593
11. Calfee, R. C., & Valencia, R. R. (1991). *APA guide to preparing manuscripts for journal publication.* Washington, DC: American Psychological Association.
12. Dod, J. (1999). Effective substances. In: The dictionary of substances and their effects. Royal Society of Chemistry. Available via DIALOG. http://www.rsc.org/dose/Effective substances. Cited 15 Jan 1999.

[2] Always use the standard abbreviation of a journal's name according to the ISSN *List of Title Word Abbreviations*, see http://www.issn.org/en/node/344

13. Harris, M., Karper, E., Stacks, G., Hoffman, D., DeNiro, R., Cruz, P., et al. (2001). Writing labs and the Hollywood connection. *J Film* Writing, 44(3), 213–245.
14. O'Neil, J. M., & Egan, J. (1992). Men's and women's gender role journeys: Metaphor for healing, transition, and transformation. In B. R. Wainrig (Ed.), *Gender issues across the life cycle* (pp. 107–123). New York: Springer.
15. Kreger, M., Brindis, C.D., Manuel, D.M., Sassoubre, L. (2007). Lessons learned in systems change initiatives: benchmarks and indicators. *American Journal of Community Psychology*, doi: 10.1007/s10464-007-9108-14.
16. Alber John, Daniel C. O'Connell, and Sabine Kowal. 2002. Personal perspective in TV interviews. *Pragmatics* 12:257–271
17. Cameron, Deborah. 1997. Theoretical debates in feminist linguistics: Questions of sex and gender. In *Gender and discourse*, ed. Ruth Wodak, 99–119. London: Sage Publications.
18. Cameron, Deborah. 1985. *Feminism and linguistic theory.* New York: St. Martin's Press.
19. Dod, Jake. 1999. Effective substances. In: The dictionary of substances and their effects. Royal Society of Chemistry. Available via DIALOG.
http://www.rsc.org/dose/title of subordinate document. Cited 15 Jan 1999
20. Suleiman, Camelia, Daniel C. O'Connell, and Sabine Kowal. 2002. 'If you and I, if we, in this later day, lose that sacred fire...': Perspective in political interviews. *Journal of Psycholinguistic Research*. doi: 10.1023/A:1015592129296.
21. Brown B, Aaron M (2001) The politics of nature. In: Smith J (ed) The rise of modern genomics, 3rd edn. Wiley, New York
22. Dod J (1999) Effective Substances. In: The dictionary of substances and their effects. Royal Society of Chemistry. Available via DIALOG.
http://www.rsc.org/dose/title of subordinate document. Cited 15 Jan 1999
23. Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. J Mol Med, doi: 10.1007/s001090000086
24. Smith J, Jones M Jr, Houghton L et al (1999) Future of health insurance. N Engl J Med 965:325–329
25. South J, Blass B (2001) The future of modern genomics. Blackwell, London