Some Constraints on Complexity Reduction

Within the context of this text, complexity reduction means the process of modelling a given complex environment, problem or system in a simplified way in order to make it easier understandable and manageable. Doing so, it is normally tried to tackle the important features of the given system and to omit the unimportant ones, thus hoping that the conclusions and solutions derived from the simplified system also hold when applied to the real complex system.

Seen from this viewpoint, complexity reduction is obviously a very ubiquitous and essential process for us human beings when handling our extremely complex world. It is applied, for instance:

- When we try to understand others humans,
- When we try to understand environmental behaviour, starting from how simple organisms grow and ending up how climate changes,
- When we try to resolve day-to-day challenges, such as planning for the day, finding a needed product in shops or deciding how to travel from A to B,
- When we try to construct a machine or software code to resolve whatever given task,
- When we vote for political solutions of challenges in society.

Whenever doing so, we focus on some aspects regarded as important and omit the unimportant. This is necessarily so because the given systems are usually much more complex and contain much more information than our brain is able to process. The process of complexity reduction is so natural with us that in many cases we aren’t even aware of it, e.g. when we try to find a parking lot and don’t care for other details of streets, houses or cars we may see while doing so.

Because complexity reduction is of such a high importance to us, it is interesting to analyze the constraints to it in terms of accuracy and predictability of the real system’s behaviour when simpifying it. Obviously, if accuracy and predictability are bad, we’ll not be able to really understand and our solutions won’t work well. This analysis, taken from a formal and mathematical viewpoint, is the purpose of this article.

Let **X** be a vector of stochastic processes on a given probabilty space $(\Omega, \Im, P)$:

(1.1) $\mathbf{X} = X_i^t : (\Omega, \Im, P) \rightarrow \Re,\ t \in \Re^+,\ t \geq t_0,\ i = 1 \ldots N,\ N \in \aleph$

with P-almost sure continuous paths and deterministic start values

(1.2) $X_i^{t_0} \in \Re,\ i = 1 \ldots N.$

X shall represent the real system's inputs, constituted by its input variables X_i which may change their values randomly over time. The components X_i^t are assumed as being stochastically independent, but for now, no other assumption is made about their distributions.

The independence ssems to be a rather intuitive assumption, though. If, for instance, we observe input variables X_i^t, X_j^t not being stochastically independent, but not fully dependent as well (meaning $E(X_j^t \mid X_i^t) \neq X_j^t$ with $P > 0$) – which should be the case if we need to take X_j^t into account at all – , in many cases there might exist another stochastic process Y^t (or more than one, if needed) independent from X_i^t such that X_j^t could be expressed as $X_j^t = f(X_i^t, Y^t)$, f being a deterministic function. In this case, X_j^t could be replaced by Y^t to have two independent input processes X_i^t and Y^t. This process could be sequentially completed, finding the "independent random source" of X_i^t independent of all X_j^t, $j \leq i$, until we arrive at a completely mutually independent set $\mathbf{X} = \{ X_j^t \}$ of stochastic processes. Of course, in practice, the derivation of such a Y^t is easier said than done.

Continuing our definition set, there shall be a deterministic continuous output function:

(2) $\mathbf{F} : \Re^N \rightarrow \Re^K, K \in \aleph$

The combination of the input values, provided by **X,** with the resulting output values, provided by **F**(**X**)**,** determines the system's behaviour. Let's look at some examples:

- When analyzing the global climate change, **X** would represent the input variables being modeled (as CO_2 emissions, absorption capabilities of forests, and so on), while **F**(**X**) models the resulting average global temperature (maybe supplemented with some other outputs).
- When writing a piece of software code, **X** would represent the input data being processed, while **F**(**X**) models the output data delivered by the software.

However, when we evaluate systems to behave in a satisfying or unsatisfying way – which is normally the purpose of analysis when serving to support decisions – there are also further aspects to be considered. If a piece of software shall be written, the effort doing it plays a significant role in the decision whether it shall be done. This effort (from an economic viewpoint, to be regarded as investment costs), as well as the day-to-day effort running the software (operating costs, at least power supply) are to be compared with the added value the software is generating (the utility). When calculating the overall profit of the software, the "prices" of working hours as well as of the output data would need to be taken into account and be included into **X,** modeled as input variables too. **F**(**X**) would then include the overall profit of writing the software.

Similarly, the climate change model can be extended such that the impact of a rising global temperature is measured – e.g. floodings and the costs of it. The "prices" of floodings would then need to be included into **X**, while **F**(**X**) would model the impact of a rising global temperature on humanity. This seems to be a very impractical and over-inclusive modelling approach, but actually, it's reality: When deciding whether and how to manage the climate change, such "prices" of rising temperatures actually affect whether such decisions are good or bad. If, for instance, humanity would eventually be able to manage a living in settlements completely unaffected by floodings (if that would be possible somehow), it could be less justified to invest high amounts of effort, resources and money to prevent a temperature rise.

It is therefore not surprising that within most systems in reality, *N* is an extremely large number and the need for complexity reduction is absolutely mandatory. The degree of complexity is increasing by magnitudes when moving from the pure functionality of a system ("works as designed" vs. "fails") to its economic or social impact ("good" vs "bad").

Let now $\varepsilon \in \Re$, $\varepsilon > 0$, be an output threshold and $\mathbf{y} = y_i$ be a corresponding vector of real-valued input deviation thresholds derived as follows:

(3.1) $$y_i := \inf (|x - X_i^{t_0}| \;\Big|\; x \in \Re, \; \exists (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N) \in \Re^{N-1} : \| \mathbf{F}(x_1, x_2, \ldots, x_{i-1}, \mathbf{x}, x_{i+1}, \ldots, x_N) - \mathbf{F}(x_1, x_2, \ldots, x_{i-1}, X_i^{t_0}, x_{i+1}, \ldots, x_N) \| > \varepsilon),$$

$i = 1 \ldots N$

($\|\cdot\|$ being the euclidic metric on $\Re^K$). As we assumed F to be continuous and thus the metric on the right side of (3.1) as function of x, the exact match $| X_i^t - X_i^{t_0} | = y_i$ itself can only lead to an output change of maximal ε. This is also true for all X_i^t which are closer to $X_i^{t_0}$ than y_i. This y_i is the upper boundary for the range of values for X_i^t who do not effect the output $\mathbf{F}(\mathbf{X})$ in a way regarded as relevant (changing the difference to $X_i^{t_0}$ beyond ε) while allowing for all possible combinations with the other components of **X**. The y_i can be interpreted in a way that if these deviation thresholds are exceeded by X_i^t, the system under observation would change its behaviour such that it needs to be taken into account when deriving a simplified model – or more formally, the output $\mathbf{F}(\mathbf{X})$ would be affected significantly. If, on the other hand, $| X_i^t - X_i^{t_0} |$ would not exceed its respective threshold, the development of X_i^t could be assumed as irrelevant. Obviously, $X_i^{t_0}$ can be 0, but doesn't need to. The latter would be the case e.g. for the the power supply voltage for the computer running the software; while undergoing minimal fluctuations, the software would run, but higher fluctuations could lead to power failures causing a hard shutdown of the computer (possibly including uncontrolled data loss).

y_i could be be undefined if the threshold ε is never exceeded for no x – or it could be defined but outside the possible output range of X_i^t. In this case, the variable X_i^t is irrelevant by construction because it cannot affect the outcome in any significant way. We assume those variables to be excluded from the system. Finally, we assume that y_i can't be 0 – a neighborhood of $X_i^{t_0}$ must exist where ε cannot be exceeded if only x is varied and stays close enough. We can therefore assume all y_i to be well-defined real numbers > 0 whose values *can*, at least theoretically, be produced by the respective variables X_i^t.

Let p_i^t be the respective probabilities of **X** exceeding a threshold within interval $[t_0, t]$:

(3.2) $$p_i^t := P(\max_{s \in [t_0, t]} | X_i^s - X_i^{t_0} | > y_i), \; i = 1 \ldots N$$

Because of what was said above, we have $0 < p_i^t < 1$ for all $i = 1 \ldots N$.

Finally, to achieve complexity reduction, let's form a subset of the components $i = 1 \ldots N$. It would be defined by a number $n > 0$, $n < N$. The simplified model would now only observe the system given by:

(4) $$\mathbf{x} = X_i^s : (\Omega, \Im, P) \rightarrow \Re, \; s \in [t_0, t], \; i = 1 \ldots n.$$

Consequently, a good accuracy of the simplified model, compared to the real model, would require the p_i^t being very small (and thus neglectable) when i > *n*, ensuring that **X** will hopefully not exceed the deviation thresholds when i > *n*.

In other words, the simplified model is derived by omiting *N-n* variables of the real model which are regarded as unimportant. Not much more is assumed here, and we obviously do have a quite generalistic approach which covers many of the real world problems. From a modelling perspective, it's irrelevant whether the modeler of the simplified model has conciously attributed the i > *n* variables to be irrelevant or just wasn't aware of them.

The interesting measure now is the probability under which at least one of the X_i^t , i > *n*, exceeds its deviation threshold within the given interval. This probability, call it P_{fail}^t , corresponds to the probability of complexity reduction not being adequate and thus possibly not being successful.

Obviously, as the components are assumed to be stochastically independent, P_{fail}^t is to be calculated as follows:

(5) $$P_{fail}^t := 1 - \prod_{i=n+1}^{N} (1 - p_i^t) .$$

We now calculate the mean threshold exceeding probability among the variables excluded from the simplified model:

(6) $$\overline{p}^{\,t} := \sum_{i=1}^{N-n} p_i^t / (N\text{-}n)$$

From (5) follows, because of (6) and due to the fact that the average mean is always equal or larger than the geometric mean:

$$P_{fail}^t = 1 - \left(\left(\prod_{i=n+1}^{N} (1 - p_i^t) \right)^{1/(N-n)} \right)^{N-n}$$

$$\geq 1 - \left(\left(\sum_{i=1}^{N-n} (1 - p_i^t) / (N\text{-}n) \right)^{N-n} \right.$$

(7) $$= 1 - (1 - \overline{p}^{\,t})^{N-n} .$$

Because of $0 < \overline{p}^{\,t} < 1$, this expression would converge monotonously to 1 when (*N-n*) grows while $\overline{p}^{\,t}$ remains constant, which is logical: Out of an extremely large number of independent variables all with a constant probability > 0 to exceed their thresholds, there is nearly 100% probability at least some of the variables doing that.

The power term from (7) can be simplified as follows: Let's assume that

(8.1) $$\overline{p}^{\,t} \leq 4 / (N\text{-}n\text{-}3)$$

(we'll consider the other case below). From (7) follows, because of the binomial theorem,

$$P_{fail}^t \geq 1 - (1 - \overline{p}^{\,t})^{N-n}$$

$$= 1 - \sum_{i=0}^{N-n} \binom{N-n}{i} (-\overline{p}^{\,t})^{i}$$

$$= 1 - \sum_{i=0}^{N-n} \binom{N-n}{i} (-1)^{i} (\overline{p}^{\,t})^{i}$$

(8.2) $$= \sum_{i=1}^{N-n} \binom{N-n}{i} (-1)^{i+1} (\overline{p}^{\,t})^{i} .$$

If we now look at summands i=3 and i=4 of the alternating sum from (8.2) (provided these exist), based on (8.1) we have

$$(N-n)(N-n-1)(N-n-2) (\overline{p}^{\,t})^3 / 6 - (N-n)(N-n-1)(N-n-2)(N-n-3) (\overline{p}^{\,t})^4 / 24$$

$$= (N-n)(N-n-1)(N-n-2) (\overline{p}^{\,t})^3 (1/6 - (N-n-3)\, p^{t}_{N} / 24)$$

$$\geq (N-n)(N-n-1)(N-n-2) (\overline{p}^{\,t})^3 (1/6 - (N-n-3)(4 / (N-n-3)) / 24)$$

$$= 0 .$$

The same inequation obviously also holds for all higher pairs (provided these exist) of summands like i=5 + i=6, i=7 + i=8, and so on (because if $\overline{p}^{\,t} \leq 4 / (N-n-3)$ holds, also $\overline{p}^{\,t} \leq (i+1) / (N-n-i)$ holds for $i \geq 5$). It's also valid if the negative summand is missing due to uneven *N-n*. We can therefore omit all summands from (8.2) except i=1 and i=2 and have

(8.3) $$\mathrm{P}^{t}_{fail} \geq (N-n)\, \overline{p}^{\,t} - (N-n)(N-n-1) (\overline{p}^{\,t})^2 / 2 .$$

If, on the other hand, (8.1) would not hold, then, in particular, $\overline{p}^{\,t} > 2 / (N-n-1)$ which would make the right side of (8.3) negative and therefore (8.3) trivially true as well. In this case, (8.3) obviously wouldn't provide a helpful lower boundary and better boundaries should be calculated. However, it is intuitively clear that with such a rather high value of $\overline{p}^{\,t}$, the constraints on complexity reduction derived below would arise even stronger, because the power term from (7) would be even larger and closer to 1.

Of course everything now depends on the relation between $\overline{p}^{\,t}$ and (N-n). One plausible approach to model this proportion could be the following:

- Let *N* be fixed and *n* be the flexible part, decided by the modeller of the simplified model,
- let's assume the p^{t}_{i} to be arranged descending and to fulfil a power law like $p^{t}_{i} = c\, i^{-k}$, $c > 0$, $k \geq 1$,
- let's (very roughly) conclude that $\overline{p}^{\,t}$, based on its definition from (6), fulfils

(9.1) $$\overline{p}^{\,t} = c\, ((N + n + 1)/2)^{-k} ,\ c > 0,\ k \geq 1.$$

In case of *n* = *N*-1 (being the most accurate nontrivial complexity reduction imaginable), the above kind of approximation of the average would obviously be accurate according to the definition of the p^{t}_{i}, while with a larger distance between *n* and *N*, $\overline{p}^{\,t}$ would be underestimated because of $2\, (i + 1)^{-k} < 1/ i^{-k} + 1/ (i + 2)^{-k}$ for any i *[proof needed]*. As we search for lower boundarys of $\overline{p}^{\,t}$ and P^{t}_{fail} anyway, this appears acceptable.

It follows that if n is close to N, $\overline{p}^t$ defined by (9.1) approximately fulfils the same power law regarding N like the p_i^t regarding i, but increases with decreasing n (because the p_i^t covered by the average increase as well). However, it can't exceed p_{n+1}^t , the first variable excluded from the complexity reduction.

If the p_i^t wouldn't be strictly descending but more descending by groups – e.g. with $p_j^t = c\, i^{-k}$ for $j = i, i+1, \dots, i+k$; $p_j^t = c\,(i+k+1)^{-k}$ for $j = i+k+1, \dots$; and so on – (9.1) would still serve as a reasonable approcimation for the average. The factor c could counterbalance whether the basis for the power law (in the above example: i) is more taken from the beginning or from the end of the group range.

Because of (9.1), (8.3) now turns to

$$P_{fail}^t \geq c\, 2^k (N-n) / (N + n + 1)^k - (1/2)\, c^2\, 2^{2k} (N-n)(N-n-1) / (N + n + 1)^{2k} \quad \Rightarrow$$

(9.2) $$P_{fail}^t \geq \alpha - \alpha^2 / 2 , \quad \alpha := c\, 2^k (N-n) / (N + n + 1)^k \quad .$$

This function of α has its maximum at $\alpha = 1$, delivering $P_{fail}^t \geq 1/2$. If α is smaller due to a small n, (9.2) does not provide a helpful boundary (due to the fact that in this case, the omitted summands of the binomial sum become more relevant and omitting them gives away too much). However, from (9.1) and (7) it is easily seen that if starting from the value of n which leads to $\alpha = 1/2$, reducing n while fixing all other parameters leads to an increasing boundary for P_{fail}^t (which is logical).

The power law term above resembles Zipf's Law (although derived in a different context). In the special case of Zipf's classical approach, we would assume $c = 1$ and $k = 1$. If we then additionally assume that only a smaller fraction of the variables have been included into the simplified model, fulfilling

(9.3) $$n \leq (N-1) / 3 \ ,$$

then, because (i) if we had equality in (9.3), we would have $\alpha = 1$ in (9.2), and because (ii) the lower boundary for P_{fail}^t increases with reducing n because of (9.1) and (7), (9.3) would lead to

(9.4) $$P_{fail}^t \geq 1/2 \quad .$$

Such a failure probability of at least 50% ("failure" in the meaning of having omitting variables which should not have been omitted) would probably hardly be regarded as satisfactory for a complexity reduction approach. It can be concluded that if such a "Zipf classic" proportion is the case for (9.1) (with $c = 1$ and $k = 1$), complexity must not be reduced too much, providing a n which is large enough related to N, being more an $O(N)$ instead of $O(\sqrt{N})$. To be more specific, according to (9.1) and (9.2) n would need to be at least equal to $2N/3$ to ensure an $\alpha \leq 2/5$ and thus a failure probability $\leq$ 30%; or in other words, given (9.1) with a "Zipf classic" setting, you need to include a majority (!) of all variables into your simplified model to reach a failure probability at least less that 30% . If this is not possible because even the simplified model would become too complex, there is no easy way out. **You cannot include "only some" of the missing variables into the model, because you just don't know beforehand which of the many will be the problematic one(s) exceeding its threshold.**

This could be an unreasonable bad case, so let's try to improve parameters in favour of the complexity reduction, assuming e.g. $k = 2$ instead of $k = 1$, making $\overline{p}^t$ much smaller. We can see that (except trivial systems with a small N), α too becomes rather small and, importantly, depends proportionally on 1/N. If we assume something like N = 100, no value for n in (9.2) and also in (7) (even n = 0) delivers a lower boundary for P^t_{fail} higher than 4%. It could be assumed that this might be too optimistic because the switch from the geometric mean to the arithmetic mean between (5) and (7) gives away too much, but also the exact calculation from (5) using $p^t_i = c\, i^{-k}$ delivers no P^t_{fail} higher than 8,2% , if $n \geq 10$. It appears we're on the safe side when $k \geq 2$.

To assume $c < 1$ also improves the situation (not as strongly as k, though). It turns out that with a value for *c* equal or less to 1/2, an α = 1/2 is no more possible with any value for *n* and *k*, even *n* = 0 and k = 1. A value for c of $c \leq 0{,}1$ ensures a lower boundary for P^t_{fail} smaller than 20%.

Summary so far: Given a power law for $\overline{p}^{\,t}$ as in (9.1), the chances for complexity reduction are:

- bad when $c = 1$, $k = 1$,
- good when $k \geq 2$,
- at least reasonable with $c \leq 0{,}1$,
- with values in between, it depends.

We've so far already found several significant constraints on complexity reduction while still regarding t being constant; if we make optimistic assumptions about the value distribution of p^t_i and thus of $\overline{p}^{\,t}$, complexity reduction still seems to be feasible, though. However, when analyzing or designing a system by reducing complexity, this approach should normally be sustainable during a longer period and not only until one specific timestamp t.

What happens with the lower boundary for P^t_{fail} from (7) when t increases?

Without any further constraints there seem little conclusions to be made. Let's therefore make the following assumptions:

There is at least one variable X^t_w , $n < w \leq N$, among those excluded by the complexity reduction which fulfils the following:

(10.1) $p^t_w \leq \overline{p}^{\,t}$ for all $t \geq t_0$.

(10.2) $p^t_w \geq 1 - (1 - \overline{p}^{\,t})^{1/2^k}$ for at least one $t \geq t_0$.

(10.3) Let

- $Z^{s,t} := X^t_w - X^s_w$ be the increments of X^t_w , $t_0 < s < t$,
- $Z^{s,t}$, $Z^{s,u}$ being mutually independent and identically distributed for any $t_0 \leq s < t < u$ with $u - t = t - s$,
- the $Z^{s,t}$ being symmetrically distributed around 0, meaning $P(Z^{s,t}_i \geq x) = P(Z^{s,t}_i \leq -x)$ for any $x \in \Re$, for any s, t,
- the $Z^{s,t}$ having distribution densities $f^{t-s}(x)$ monotonously centered around 0, meaning $f^{t-s}(y) \leq f^{t-s}(x)$ for $0 < x < y$ as well as for $y < x < 0$, for any s, t.

In other words: Among the (N-n) variables excluded from the complexity reduction, we assume at least one with a threshold exceeding probability constantly equal or below the average (10.1), with a threshold exceeding probability at some timestamp t with approximately at least a value fulfilling (10.2), and finally satisfying some regularity requirements about independent increments, symmetry and divisibility as fulfilled e.g. by the Wiener process (10.3).

In this case, much more can be said. Obviously, there is

(10.4) $P(\max_{s\in[t_0,\,t]} | X_w^s - X_w^{t_0} | > y_w) = P(\max_{s\in[t_0,\,t]} | Z^{t,s} | > y_w)$.

Let now be $t_0 \leq t < u$. $Z^{t,u}$ is independent from $Z_i^{t_0,t}$. We can therefore use the density convolution for the sum of these two variables and have, together with the other assumptions from (10.3) (here abbreviating y_w with y):

(10.5) $P(| X_w^u - X_i^{t_0} | \leq y) = P(|Z_i^{t_0,t} + Z^{t,u} | \leq y)$

$$= \int_{-y}^{y} \int_{-\infty}^{\infty} f^{t-t_0}(x)\, f^{u-t}(z-x)\, dx\, dz$$

$$= \int_{-\infty}^{\infty} f^{t-t_0}(x) \int_{-y}^{y} f^{u-t}(z-x)\, dz\, dx$$

$$\leq \int_{-\infty}^{\infty} f^{t-t_0}(x) \int_{-y}^{y} f^{u-t}(z)\, dz\, dx$$

$$= P(| Z^{t,u} | \leq y) .$$

As *u* was arbitrarily selected, this implies for the complement set of the maximum that

(10.6) $P(\max_{s\in[t,\,u]} | X_i^s - X_i^{t_0} | > y_w) \geq P(\max_{s\in[t,\,u]} | Z_i^{t,s} |) > y_w)$.

(10.5) and (10.6) hold the same if the sets on both sides are intersected with $\{ \max_{s\in[t_0,\,t]} | X_w^s - X_w^{t_0} |) \leq y_w \} = \{ \max_{s\in[t_0,\,t]} | Z^{t,s} | \leq y_w \}$ *[proof needed]*. Based on that, it follows that for each $t_0 < t < u$

$$
\begin{aligned}
p_w^u = \;& P(\max_{s\in[t_0,\,u]} | X_w^s - X_w^{t_0} |) > y_w) \\
= \;& P(\max_{s\in[t_0,\,t]} | X_w^s - X_w^{t_0} |) > y_w) \\
& + P((\max_{s\in[t_0,\,t]} | X_w^s - X_w^{t_0} |) \leq y_w) \cap (\max_{s\in[t,\,u]} | X_w^s - X_w^{t_0} |) > y_w)) \\
\geq \;& P(\max_{s\in[t_0,\,t]} | Z^{t_0,s} |) > y_w) \\
& + P(\max_{s\in[t_0,\,t]} | Z^{t_0,s} |) \leq y_w) \cap (\max_{s\in[t,\,u]} | Z^{t,s} |) > y_w) \\
\geq \;& P(\max_{s\in[t_0,\,t]} | Z^{t_0,s} |) > y_w) \\
& + P(\max_{s\in[t_0,\,t]} | Z^{t_0,s} |) \leq y_w)\; P(\max_{s\in[t,\,u]} | Z^{t,s} |) > y_w) \\
= \;& p_w^t + (1 - p_w^t)\, p_w^{u-t}
\end{aligned}
$$

(10.7) $\quad = \quad 1 - (1 - p^{t}_{w}) (1 - p^{u-t}_{w})$.

The right-hand side of (10.7) is just the same formula as for the complement of the intersection if p^{t}_{i} and p^{u-t}_{w} would represent independent events. If we set $u = t_0 + m(t - t_0)$, $m \in \aleph$, it follows by total induction

(10.8) $\quad p^{u}_{w} \geq 1 - (1 - p^{t}_{w})^{m}$.

Finally, inserting (10.8) into (7) for u, we have (thanks to (10.1) and (10.2))

$$
\begin{aligned}
(10.9) \quad P^{u}_{fail} &\geq 1 - (1 - \overline{p}^{\,u})^{(N-n)} \\
&\geq 1 - (1 - p^{u}_{w})^{(N-n)} \\
&\geq 1 - (1 - p^{t}_{w})^{m(N-n)} \\
&\geq 1 - (1 - \overline{p}^{\,t})^{(m/2^{k})(N-n)} \quad , u = t_0 + m(t - t_0) \quad .
\end{aligned}
$$

Provided $(m/2^{k})$ as integer, increasing $(t - t_0)$ by a factor m therefore would produce at least $1/2^{k}$ of the effect on P^{t}_{fail} from (9.2) as increasing (N-n) by the same factor m and leaving $\overline{p}^{\,t}$ unchanged – we just have (9.2) again with an α multiplied by $(m/2^{k})$:

(10.10) $\quad P^{t}_{fail} \geq \alpha - \alpha^{2} / 2, \quad \alpha := m\, c\, (N\text{-}n) / (N + n + 1)^{k}$.

Remember that k is a constant small number determining the degree of the power law from (9.1), while *m* increases unrestrictedly with time. In particular, by assuming

(10.11) $\quad m = (N + n + 1)^{k-1} / c$,

m would equalize all effects of any $c < 1$, $k > 1$ and thus recreate the (unsatisfying) situation from (9.2) with $c = 1$, $k = 1$.

Admittedly, depending on the time unit $(t - t_0)$, defining *m* like this could still mean a quite long time (roughly proportional to *N* if $k = 2$). The reason for this a bit paradoxial behaviour of our lower boundary for P^{t}_{fail} is that for $k \geq 2$, the arithmetic mean contained in (7) decreases strongly with *N*, consequenty the lower boundary for P^{t}_{fail} , while the exact formula from (5) does not. So, with $k \geq 2$, moving from (5) to (7) seems not to deliver a very helpful lower boundary any more. Instead, the trivial lower boundary resulting from (5), using only the first factor from the product (provided the power law as in (9.1))

(10.12) $\quad P^{t}_{fail} \geq p^{t}_{n+1} = c\, (n+1)^{-k}$

could be even better, if *n* is small. (10.8) and (10.9), if we replace $\overline{p}^{\,t}$ by p^{t}_{n+1} , then lead to

(10.13) $\quad P^{u}_{fail} \geq 1 - (1 - c\, (n+1)^{-k})^{m/2^{k}}$, if, again, we assume to find another variable X_w^t which fulfils (10.1)-(10.3) with $\overline{p}^t$ replaced by p_{n+1}^t .

Using the same methods as from (7) to (8.3) – with ($m/2^k$) instead of (N-n) and p_{n+1}^t instead of $\overline{p}^t$ – we can estimate:

(10.14) $P_{fail}^t \geq (m/2^k)\ c\ /\ (n+1)^k - (1/2)\ (m/2^k)\ ((m/2^k)-1)\ c^2\ /\ (n+1)^{2k}$

$\geq \alpha - \alpha^2/2, \quad \alpha := (m/2^k)\ c\ /\ (n+1)^k$.

Again, this lower boundary reaches 1/2 if we have $\alpha = 1$, implying

(10.15) $m = 2^k\ (n+1)^k\ /\ c$.

Provided *n* to be rather small compared to *N*, this could require a shorter time period (or *m*) compared to (10.11). On the other hand, if *n* is also large, the time span from (10.11) might be shorter (because the exponent is *k*-1 instead of *k).* We have found two alternative methods to estimate *m* and could switch to the one more appropriate depending on the circumstances.

We can conclude that, very likely, any system based on complexity reduction – if not already from the start, at least by letting it run long enough – eventually fails: At least some of the omitted variables will probably exceed their threshold at least once. (Note that "failure", as mentioned already, here means that variables were omitted which should not have been omitted. The system itself does not inevitably need to fail, whatever that means. However, there is probably a correlation.). This is not surprising in itself; what might be surprising is that the time interval in which the failure probability is becoming too high could be much shorter – given (10.11) or (10.15), which might be still too optimistic – than human intuition would assume.