Correct Models

René Blacher

Correct Models

René BLACHER

Laboratory LJK
Université Joseph Fourier
Grenoble
France

Summary : In previous reports, we have show how to transform a text y_n in a random sequence by using functions of Fibonacci T_q. Now, in this report, we obtain a clearer result by proving that $T_q(y_n)$ has the IID model as correct model. But, it is necessary to define correctly a correct model. Then, we study also this problem.

Key Words : Fibonacci sequence, Random numbers, Random generator, Correct models.

1 Introduction

1.1 Building of IID sequences

In [6] and [7], we studied a method to obtain IID [1] sequences x_n of random numbers. With this aim, we have transformed sequences $y_n \in \{0/m, 1/m, ..., (m-1)/m\}$, n=1,2,....,N, provided by texts or machines.

One can thus suppose that y_n is the realization of a sequence of random variables Y_n defined on a probability space (Ω, A, P) : $y_n = Y_n(\omega)$ where $\omega \in \Omega$ and where Y_n is a correct model of y_n.

As a matter of fact, there exist an infinity of correct models of y_n. It is thus necessary to be placed in the set of all the possible random variables.

Hypothesis 1.1 *Let $m \in \mathbb{N}^*$. One considers the sequences of random variables Y_n^θ, n=1,.....,N, defined on the probabilities spaces $(\Omega, \mathcal{A}, P_\theta)$, $\theta \in \Theta$: $(Y_1^\theta,\ Y_2^\theta,\ Y_3^\theta,\ Y_4^\theta, ..., Y_N^\theta) : \Omega \to \{0/m, 1/m, ..., (m-1)/m\}^N$. One assumes that $Y_n^\theta = Y_n$ for all $\theta \in \Theta$. In order to simplify the presentations of results, we set $P\{(Y_1^\theta, ..., Y_N^\theta) \in Bo\} = P_\theta\{(Y_1, ..., Y_N) \in Bo\}$ for all Borel set Bo.*

For example, one can assume that $\Omega = \{0/m, 1/m, ..., (m-1)/m\}^N$ and $(Y_1, ..., Y_N) = (Id, ..., Id)$.

It will thus raise the question to define correctly what is a correct model because, even if a model Y_n^θ is not correct, it is however always possible that $y_n = Y_n^\theta(\omega)$ as it is the case for the increasing sequences when Y_n^θ is the IID model.

In the case where the model Y_n^θ is IID, to define a correct model is a generalization of the already very complex problem of the definition of an IID sequence. However we will be able to give satisfactory answers in the particular case which interests us. In this one, we use the functions of Fibonacci to obtain IID sequences.

Definition 1.1 *Let fi_n be the Fibonacci sequence : $fi_1 = fi_2 = 1$, $fi_{n+2} = fi_{n+1} + fi_n$. Let T be a congruence $T(x) \equiv ax$ modulo m such that there exists $n_1 > 2$ satisfying $a = fi_{n_1}$ and $m = fi_{n_1+1}$. Then T is said a Fibonacci's congruence.*

[1] Independent Identically Distributed

Definition 1.2 *Let $q \in \mathbb{N}^*$. Let T be the congruence of Fibonacci modulo m. We define the function of Fibonacci T_q by $T_q = Pr_q \circ \widehat{T}$ where*
1) $\widehat{T}(x) = \overline{T}(mx)/m$, when $\overline{z} \equiv z$ modulo m and $0 \leq \overline{z} < m$ if $z \in \mathbb{Z}$,
2) $Pr_q(z) = \overline{\overline{0, b_1 b_2 .... b_q}}$ when $z = \overline{\overline{0, b_1 b_2 ...}}$ is the binary writing of z.

These functions T_q make IID the sequences of random variables $Y_n^\theta \in \{0/m, 1/m, ..., (m-1)/m\}$. Indeed, we have proved in [7] and [6] that, by choosing m and q correctly, for most models Y_n^θ (including the bad ones), for all Borel set Bo, for all $n \in \{0, 1, ..., N\}$, for all $p \in \{0, 1, ..., N\}$, for all injective sequence $j_s \in \mathbb{Z}$, s=1,2,...,p, such that $j_1 = 0$,

$$P\big\{(X_{n+j_1}^\theta, ..., X_{n+j_p}^\theta) \in Bo\big\} = L(Bo)[1 + Ob(1)\epsilon] \ ,$$

where $X_n^\theta = T_q(Y_n^\theta)$, where ϵ is small enough, where L(Bo) means the Borelian measure of Bo, and where Ob(.) means the classical "O(.)" with the additional condition $|Ob(1)| \leq 1$.

Now, if $P\big\{(X_{n+j_1}^\theta, ..., X_{n+j_p}^\theta) \in Bo\big\} = L(Bo)[1 + Ob(1)\epsilon]$, one cannot differentiate x_n from an IID sequence. More generally, if there is a correct model $Y_n^{\theta_1}$ of a sequence y_n, n=1,2,....,N, and if another model $Y_n^{\theta_2}$ checks, for all n, for all p, for all sequence j_s, for all Borel set Bo,

$$P\big\{(Y_{n+j_1}^{\theta_2}, ..., Y_{n+j_p}^{\theta_2}) \in Bo\big\} = P\big\{(Y_{n+j_1}^{\theta_1}, ..., Y_{n+j_p}^{\theta_1}) \in Bo\big\}[1 + Ob(1)\epsilon] \ ,$$

$Y_n^{\theta_2}$ will be also a correct model of y_n. We will prove this result in section 4.

One will deduce from it that there is a correct model $Y_n^{\theta_c}$ of y_n such that, for all Borel set Bo,

$$P\big\{(T_q(Y_1^{\theta_c}), ..., T_q(Y_N^{\theta_c})) \in Bo\big\} = L(Bo) \ .$$

Thus the model $X_n^{\theta_c} = T_q(Y_n^{\theta_c})$ will be exactly the IID model and it is a correct model of $x_n = T_q(y_n)$. That gives a simpler proof of the results of [7] and [6].

But, of course, for showing well what that means, it is necessary to give at first a definition of a correct model, and thus also of an IID model.

2 Definitions of a random sequence

It thus raises the question to define correctly what is a correct model. In fact, to define a correct model is a question of the same order as to define an IID sequence. Indeed, in the IID case, to say that Y_n^θ will be a correct model amounts saying that y_n is an IID sequence : it is thus well the problem of the definition of an IID sequence which one finds here.

2.1 P-distributed sequences

Many studies were made to have reasonable definitions of IID sequences : there is a good summary of these studies in chapter 3-5 of Knuth : cf [1]. A first method to define an IID sequence x_n consists in using the p-distributed sequences.

Definition 2.1 *: Let $x_n \in \{0/m, 1/m, ..., (m-1)/m\}$, n=1,2,....,N, be a sequence of real numbers such that $m \in \mathbb{N}^*$. For all finite sequence of intervals $I_s \subset [0, 1]$, we denote by P_e the empirical probability : $P_e = (1/[N-p]) \sum_{n=1}^{N-p} 1_{I_1}(x_n) 1_{I_2}(x_{n+1}) .....1_{I_p}(x_{n+p})$.*

The sequence $\{x_n\}$ is said p-distributed if $|P_e - L(I)| \leq (N-p)^{-1/2}$ for all $I = I_1 \otimes I_2 \otimes ... \otimes I_p$.

Definition 2.2 *The sequence x_n is random if it is p-distributed for all $p \leq Log_2(N-p)$.*

Unfortunately, this definition does not take into account the randomness of subsequences $x_{t_1}, x_{t_2}, ...x_{t_h}$. However, it is known that one cannot extend this definition to all the transformations $s \to t_s$ which define these subsequences : for example, this definition cannot be satisfied by the sequences x_{t_s} increasing. It is necessary thus that the application $s \to t_s$ is too not complicated. Also Knuth proposes the following definition.

Definition 2.3 *: The sequence x_n is random with respect to a set of algorithms A, if for all sequence $x_{t_1}, x_{t_2}, ...x_{t_h}$, determined by A, it is p-distributed for all $p \leq Log_2(N-p)$.*

These definitions summarize those given by Knuth, [1] page 108. This type of definition was the subject of many studies. In any case, none of these definitions is fully satisfactory. Knuth speaks philosophical debate on this subject.

In any case, these definitions have gaps from the statistical point of view or from the point of view of Borel sets. In [6], one studied these problems and gave definitions more adapted, but unfortunately still too undetermined. It is not serious: the idea was to circumvent the associated problems by using the models of x_n, i.e. the sequences of random variables $X_n^\theta = T_q(Y_n^\theta)$ defined on probability spaces.

2.2 Probabilistic definition

Indeed, another method consists in using a probabilistic definition.

Definition 2.4 *: The sequence x_n is random if there exists an IID sequence of random variables, $X_n \in \{0/m, 1/m, ..., (m-1)/m\}$, defined on a probability space (Ω, A, P) such that $x_n = X_n(\omega)$ where $\omega \in \Omega$.*

But there is a problem with this definition : for example, x_n could be increasing. Then, Franklin proposed another definition : [8].

Definition 2.5 *: The sequence x_n is random if it has each property that is shared by all samples of an IID sequence of random variables.*

But, this definition is not precise and one could even deduce from it that no really random sequence exists (cf [1], Knuth page 149). Finally, one can also use the definition of a sample.

Definition 2.6 *: The sequence x_n is said random if it is known a priori that there exists an IID sequence of random variables, $X_n^{\theta_c}$, defined on a probability space (Ω, A, P) such that $x_n = X_n^{\theta_c}(\omega)$ and such that $X_n^{\theta_c}$ is a correct model of the sequence x_n.*

In this case, the definition of a correct model IID is equivalent to the definition of a random sequence. Of course, in order to understand this definition, it is necessary to know what is a correct model.

3 Correct models

3.1 General study

It thus raises the question to define correctly what is a correct model. Indeed, if a model Y_n^θ is not correct, it is however possible that $y_n = Y_n^\theta(\omega)$, where Y_n^θ is a sequence of random variable defined on a probability space $(\Omega, \mathcal{A}, P)$.

One has just understood that, to define a correct model it is a question of the same order as to define an IID sequence. It is thus extremely complex. But one can have a solution because one wants only to prove that the correct models $T_q(Y_n^\theta)$ will be close to the IID model.

Models with continuous density Because y_n is discrete, one can suppose that the random vector $(Y_1^\theta, ..., Y_N^\theta)$ has a continuous density function with respect to the uniform discrete measure of $\{0/m, 1/m, ..., (m-1)/m\}^N$. Therefore it is also generally the case for the conditional distribution of $Y_{\phi(n)}$ given $Y_{\phi(n-1)} = y'_1$, $Y_{\phi(n-2)} = y'_2$,.... when ϕ is a permutation de $\{1, 2, ..., N\}$.

For example, one can show that, if $y_n = [\ \overline{e(n) + rand_0(n)}\ ]/m$ (cf step c) page 92 of [7]) where e(n) means texts and $rand_0(n)$ is a pseudo random sequence, the conditional probabilities have a Lipchitz coefficient K'_0 small enough. As a matter of fact, it is encore easier to prove if $y_n = [\overline{e(n) + rand_0(n)/m + e'(n) + rand_1(n)}]/m$ where e'(n) represent texts witten backward (cf sections 11.2.4 of [6] and appendix B).

We deduce that the models Y_n^θ which have a Lipchitz coefficient K'_0 small enough are correct models.

A scientific assumption Generally, one feels well that correct models exist. In fact, it is a traditional assumption in science. In weather for example, the researchers seek a correct model, which implies its existence (if not, why to try to make forecasts?). One could thus admit that like a conjecture or a postulate without defining exactly what is a correct model.

To predict the future In fact, a correct model depends on its usefulness. For example, in meteorology, its usefulness is to predict weather.

One can transpose that to unspecified sequences of real numbers y_n, n=1,2,....,N. The usefulness of a model will be in general to predict the future. That applies perfectly to the research which we carry out in order to obtain IID sequences : if a sequence is IID random, one will not be able to predict the future knowing the past.

One could thus admit like definition of a correct model this one : a correct model is a model such as, knowing the past $Y_{n-s}^\theta = y'_{n-s}$, s=1,2,..., this one makes possible to predict the best possible the future. To be more complete, it is necessary to extend this definition to the sequences $y_{\phi(n)}$ where ϕ is a permutation of $\{1, 2, ..., N\}$.

It is necessary thus that the forecast is good : it has to be the most precise possible, but, if knowing the past, one predicts the future in a too precise way and that it is not real, the model will be bad.

Let us notice, that, under this condition, we suppose that one does not know the future $y_{\phi(n+s)}$, s=1,2,... : if not, the empirical probability would be a correct model.

Mathematical definition Mathematically, one can thus specify that: it will be said that Y_n^θ is a correct model, if, for any permutation ϕ of $\{1, 2, ..., N\}$, for all sequence y'_s, for all n, it makes possible to give the conditional probability of $Y_{\phi(n)}^{\theta_c}$ knowing the past $Y_{\phi(n-1)}^{\theta_c} = y'_1$, $Y_{\phi(n-2)}^{\theta_c} = y'_2$,....., which is the best possible one.

It will be thus true in particular when $y'_s = y_{\phi(n-s)}$ for s=1,2,3,.... It will thus be known that $P\{Y_{\phi(n)}^{\theta_c} \in Bo \mid Y_{\phi(n-1)}^{\theta_c} = y_{\phi(n-1)}, Y_{\phi(n-2)}^{\theta_c} = y_{\phi(n-2)}, ...\}$ will be the most precise possible by taking account of what one really knows, i.e the sequence $y_{\phi(n-s)}$.

Therefore, one can nothing object to this conditional probability in order to define the future when what one really knows, it is the sequence y_n. Of course it is in question conditional probabilities which one could really deduce from the sample y_n if all the mathematical properties were known and if one had an infinite computing power.

Some difficulties Unfortunately, in these definitions, one made only to move the problem: mathematically, what means "probabilities the most precise possible" and "the best possible"? One understands well what one wishes. But to define it mathematically seems extremely complicated.

However, one can do our study without knowing it. Indeed, which interests us, it is that the $T_q(Y_n^\theta)$ have a law close to an IID distribution.

Now, if $Y_n^{\theta_c}$ is a correct model, $P\{Y_{\phi(n)}^{\theta_c} \in Bo \mid Y_{\phi(n-1)}^{\theta_c} = y'_1, Y_{\phi(n-2)}^{\theta_c} = y'_2, ...\}$ defines the future $Y_{\phi(n)}^{\theta_c} \in Bo$ sufficiently well for all Borel set Bo, when, which one knows, it is the sequence $y_{\phi(n)}$. It will be thus true in particular for $P\{T_q(Y_{\phi(n)}^{\theta_c}) \in Bo' \mid Y_{\phi(n-1)}^{\theta_c} = y'_1, Y_{\phi(n-2)}^{\theta_c} = y'_2, ...\}$, and, therefore, for $P\{X_{\phi(n)}^{\theta_c} \in Bo' \mid X_{\phi(n-1)}^{\theta_c} = x'_1, X_{\phi(n-2)}^{\theta_c} = x'_2, ...\}$ (cf proposition A.1). Therefore, this conditional probability defines a good forecast of the future. That means that if one knows $x_{\phi(n-s)}$, s=1,2,.., a good prediction of $x_{\phi(n)}$ will be given by this conditional probability.

However we have proved in [6] that $P\{X_{\phi(n)}^{\theta_c} \in Bo' \mid X_{\phi(n-1)}^{\theta_c} = x'_1, X_{\phi(n-2)}^{\theta_c} = x'_2, ...\} = L(Bo')[1+Ob(1)\epsilon]$ where ϵ is small enough for the models with a continuous density and a coefficient of Lipschitz K'_0 not too large. Moreover, one has just understood above that one can admit that such models are correct if y_n is represents a text to which one adds a pseudo-random sequence. At last, we shall prove in section 5 that, in this case, there exists a correct model $Y_n^{\theta_c}$ such that $P\{X_{\phi(n)}^{\theta_c} \in Bo' \mid X_{\phi(n-1)}^{\theta_c} = x'_1, X_{\phi(n-2)}^{\theta_c} = x'_2, ...\} = L(Bo')$ if ϵ is small enough.

That means that if one knows $x_{\phi(n-s)}$, s=1,2,.., a good prediction of $x_{\phi(n)}$ will be given by uniform probability. Then, we have proved that, there exists a correct model $Y_n^{\theta_c}$ such that $T_q(Y_n^{\theta_c})$ is exactly the IID random sequence.

All the correct models One could think that another correct model could have different results and that poses a problem. But it is especially a philosophical problem: indeed, that would mean that two correct models would have incomptatible results. That seems impossible (cf also section 6).

One can thus conclude : because it is admitted that the model with continuous density and with a coefficient of Lipschitz K'_0 not too large is correct, the correct models of $T_q(Y_n^\theta)$ will be close to model IID [2].

Definition by negation One could also want to specify the definition in a negative way. In particular y_n should not fail for too many tests of the hypothesis H_0 : "$y_n = Y_n^\theta(\omega)$".

Another manner of specifying the definition in a negative way would be to say that a model is not incorrect. For example, for an IID sample, a model AR (1): $X_{n+1} = aX_n + \eta_n$ can be regarded as incorrect if a is large enough with respect to N.

One could thus say that a model is correct if one finds no mathematical property which shows that the conditional probabilities could be different from those obtained when one knows all the mathematical properties of the model and when one has a infinite capacity of computation.

This negative definition thus means that the sample y_n will not check the properties that one expects from a sample having a given law. But in these examples, they are only certain properties. In a more precise way, it is necessary to wonder which properties exactly one expects to find. One finds the problems of the definition of Franklin.

Some other correct models To obtain an acceptable definition of a correct model is maybe a problem which can be resolved.

Indeed, it is understood well that to say that there exists a correct model is a reasonable proposition. As one has pointed out it, it is a traditional scientific assumption.

In fact, for finite sequences y_n, it is a certainty that there exist correct models. For example: to y_n, one associates an independent sequence of random variables $Y_n^{\theta_0}$, with uniform law on intervals containing y_n, dependent on n, width small enough, for example about 100/N or 1/N, etc. One can thus also easily imagine a such correct model with continuous density.

It is noticed that our definition of a model correct is not contradicted by this example : this one is close to the empirical probability and we excluded this case.

[2] One can find other reasons to consolidate this assertion in chapter 13 of [6].

Several correct models There undoubtedly exist several models correct especially when the sequences y_n are finished. Thus, if y_n is an IID sample, the Y_n^θ's satisfying $P\{Y^\theta_{\phi(n)} \in [a,b] | Y^\theta_{\phi(n-s)} = y'_s, s = 1,2,...\} = (b-a)[1+\epsilon]$, for all $a < b$, for all permutation ϕ, for all sequence y'_s, s=1,2,..., and where ϵ is small enough are also correct models (cf section 4).

Conclusion In conclusion, a correct model would make possible to obtain the best conditional probability for all sequences $y_{\phi(n)}$, not knowing the future.

Does there exist such a model? Presumably, because generally y_n represents a physical phenomenon. It is thus normal to suppose its existence. Moreover, in certain cases, one can show such models : it is the case for texts.

3.2 Texts

Now, we consider the particular case where the y_n's result from texts.

A priori, a correct model would be a model which makes possible to predict the following letters (y_n, y_{n+1},.....) with a satisfactory probability if one knows the preceding letters y_{n-1}, y_{n-2} ..... One could thus say that the model will predict all the possible texts which follows the beginning of the text.

However such a model is too precise: indeed, for sequences representing a text, to suppose that one is in an English text is a priori which is wrong : cf 6) page 307 of [6]. For example, one could logically predict words invented not existing. A model in modern English language would be a correct model. But a model in a possible evolution of the English language would be it too.

These model can be refined besides: if a novel is used, it would be astonishing to find texts speaking about mathematical theorem. Therefore, there are models which make possible to better predict the continuation than others. But it is necessary that is explained by the text which precedes. If one takes only 100 words, one will not deduce from it the style of the author.

In fact in order to admit that only the English texts can represent the y_n, it would be necessary that the used text is preceded by a very large number of books which make possible to decode the language: for example preceded of all the books written in English and of all the texts of the author in order to know his style. In this case, it is possible that the only correct models are texts, even texts of the author.

Let us suppose that it is the case. That makes possible to define precise correct models. Indeed, in this case, one can admit that the correct model will be that representing all the possible texts written according to the style of the author and speaking about the subject introduced by the first pages. Of course, there is an almost infinite number of possible texts as soon as N, the sample size of y_n is large.

Concerning the associated probabilities, one can suppose that all the texts are equiprobable. That seems a correct model.

But it is not the alone one. One can take other probabilities than the equiprobable probability, for example a close probability, even another. Indeed, it seems that certain text are likely more to exist than the different ones. The equiprobable model is thus not the best inevitably. In order to find the best models it would be necessary to find those whose probabilities correspond the best to all which one knows about texts of the author. That seems impossible to realize. But theoretically, it could exist. In fact, there are several suitable models.

It thus seems difficult to find exactly all the possible correct models and especially to find a better model. However, it is felt well that these models including all the texts which the author can write seems rather correct and that there are from them which are better than others.

Therefore, for the texts, one can show correct models. All the possible texts of the author with an about uniform probability seems be a good model. Then this model defines conditional probabilities $P\{Y^{\theta_t}_{\phi(n)} \in Bo | Y^{\theta_t}_{\phi(n-1)} = y'_1, Y^{\theta_t}_{\phi(n-2)} = y'_2, ...\}$ for all n, for all y'_s, s=1,2,..., and for all permutation ϕ.

3.3 Conclusion

Thus in certain cases, there exist correct models which enable us to predict the future correctly. One can suppose that the method described for the texts is good and can be generalized.

If this assumption is refused, one will be obliged to admit that there exists such correct models defining correctly the conditional probabilities without more precise details as one does it in weather and elsewhere. It was understood that it is enough in order to prove that the IID model is a correct model of $x_n = T_q(y_n)$.

4 Models equivalent with a margin of ϵ

4.1 The problem

Let $Y_n^{\theta_2}$ and $Y_n^{\theta_1}$ be two sequences of random variables such that, for all Borel set Bo,

$$P\big\{(Y_1^{\theta_2}, ..., Y_N^{\theta_2}) \in Bo\big\} = P\big\{(Y_1^{\theta_1}, ..., Y_N^{\theta_1}) \in Bo\big\}\big[1 + Ob(1)\epsilon_{(N)}\big] \ ,$$

where $\epsilon_{(N)} = N\epsilon_{(1)}$ is small enough (cf proposition 6.3.2 of [6]). One supposes that $Y_n^{\theta_1}$ is a correct model of the sequence y_n, n=1,2,....,N. One wants to prove that $Y_n^{\theta_2}$ is also a correct model of y_n if $\epsilon_{(N)}$ is small enough (e.g. $\epsilon_{(N)} \leq 1/10$).

4.2 Example

Let us suppose that we have a really IID sequence of random variables X_n^ϵ with uniform distribution on [0,1/2] and [1/2,1] and with a probability such as $P\{X_n^\epsilon \in [1/2, 1]\} = 0,500[1 + \epsilon]$ where $\epsilon = 0,001$. Then, this sequence has not the uniform distribution on [0,1]. However, if we have a sample with size 10, we will absoluetely not understand that X_n^ϵ has not the uniform distribution on [0,1]. It is wellknown that one need samples with size larger than N=10000 minimum (and even more) in order to test this difference.

More precisely, by the CLT (Central Limit Theorem), $P\Big\{\frac{|\sum_{n=1}^N (\mathbb{1}_{[1/2,1]}(X_n^\epsilon) - 1/2 - \epsilon/2)|}{\sqrt{N(1-\epsilon^2)/4}} \geq b\Big\} \approx \Gamma(b)$ where $\Gamma(b) = P\{|X_G| \geq b\}$ when $X_G \sim N(0,1)$. Then, $P\Big\{\frac{|\sum_{n=1}^N (\mathbb{1}_{[1/2,1]}(X_n^\epsilon) - 1/2)|}{\sqrt{N/4}} \geq b\Big\} \approx \Gamma\big(b[1 - \eta(\epsilon)]\big)$ where η is continuous with $\eta(0) = 0$.

More generally, one cannot test significantly H_0 : "X_n^θ has the uniform distribution" against $H_1(\epsilon)$: "$P\{X_n^\theta \in Bo\} = L(Bo)[1 + Ob(1)\epsilon]$ " if $\sqrt{N}\ \epsilon \leq 1/10$.

For example, if $\sqrt{N}\ \epsilon = 1/10$ and b=2, the probability of obtaining $\frac{\sum_{n=1}^N [1_{[1/2,1]}(X_n^\theta) - 1/2]}{\sqrt{N/4}} \geq 2$ is about 0.0466 under $H_1(\epsilon)$ and about 0.0455 under H_0 : i.e. the probability of rejecting the assumption IID, H_0, under $H_1(\epsilon)$ is not much bigger than that of rejecting H_0 if X_n^θ is really IID.

Indeed, under, $H_1(\epsilon)$,

$$P\Big\{\frac{\sum_{n=1}^N (\mathbb{1}_{[1/2,1]}(X_n^\epsilon) - 1/2)}{\sqrt{N(1-\epsilon^2)/4}} \geq b\Big\} + P\Big\{\frac{\sum_{n=1}^N (\mathbb{1}_{[1/2,1]}(X_n^\epsilon) - 1/2)}{\sqrt{N(1-\epsilon^2)/4}} \leq -b\Big\}$$

$$= P\Big\{\frac{\sum_{n=1}^N (\mathbb{1}_{[1/2,1]}(X_n^\epsilon) - 1/2 - \epsilon/2)}{\sqrt{N(1-\epsilon^2)/4}} \geq b - \frac{\sqrt{N}\epsilon}{\sqrt{1-\epsilon^2}}\Big\}$$

$$+P\Big\{\frac{\sum_{n=1}^N (\mathbb{1}_{[1/2,1]}(X_n^\epsilon) - 1/2 - \epsilon/2)}{\sqrt{N(1-\epsilon^2)/4}} \leq -b - \frac{\sqrt{N}\epsilon}{\sqrt{1-\epsilon^2}}\Big\}$$

$$\approx (1/2)\Gamma\Big(b - \frac{\sqrt{N}\epsilon}{\sqrt{1-\epsilon^2}}\Big) + (1/2)\Gamma\Big(b + \frac{\sqrt{N}\epsilon}{\sqrt{1-\epsilon^2}}\Big) \ .$$

4.3 IID models with a margin of ϵ

These results hold in dimension p , i.e. for $\frac{1}{N-p}\sum_n \mathbb{1}_{Bo_1}(Y^{\theta_1}_{n+j_1})$.....$\mathbb{1}_{Bo_p}(Y^{\theta_1}_{n+j_p})$. One deduces from what precedes that, if x_n is the realization of a sequence of random variables X^θ_n such that $P\{(X^\theta_1,....,X^\theta_N)\in Bo\} = L(Bo)[1+Ob(1)N\epsilon]$ for all Borel set Bo, one will not be able to differentiate this model from an IID model if ϵ is rather small with respect to N.

Reciprocally, if x_n, n=1,2,....,N, is really an IID sample, a model such that $P\{(X^\theta_1,....,X^\theta_N)\in Bo\} = L(Bo)[1+N\epsilon]$ is also a correct model of the sequence x_n.

Because we shall obtain $P\{(X^\theta_1,....,X^\theta_N)\in Bo\} = L(Bo)[1+N\epsilon]$ in [6] and [7], one will be able to admit that the IID model is a correct model of the sequences x_n which we built in these reports.

4.4 Case where the CLT holds

One can adopt assumptions more general than those of the IID case by only supposing than the CLT is checked. For example, assume that the CLT holds and that the $Y^{\theta_1}_n$'s have the same distribution for n=1,2,....,N. Let $P_{Y_1}(I) = P\{Y^{\theta_1}_n \in I\}$ where I is an interval. Let $P^1_e = (1/N)\sum_n \mathbb{1}_I(Y^{\theta_1}_n)$ and $P^2_e = (1/N)\sum_n \mathbb{1}_I(Y^{\theta_2}_n)$. Let σ^2_B the variance of P^1_e. Then, if N is big enough, by the CLT,

$$P\{|P^1_e - P_{Y_1}(I)| > \sigma_B b\} \approx \Gamma(b)\ ,$$

where $\Gamma(b) = P\{|X_G|\geq b\}$ when $X_G \sim N(0,1)$. We recall that

$$P\{|P^1_e - P_{Y_1}(I)| > \sigma_B b\} = P_{\theta_1}\Big\{\big|(1/N)\sum_n \mathbb{1}_I(Y_n) - P_{Y_1}(I)\big| > \sigma_B b\Big\}\ .$$

Now there exists a Borel set $Bo^1 \subset \{0/m, 1/m, ..., (m-1)/m\}^N$ such that

$$\Big\{\omega\in\Omega\ \Big|\ \big|(1/N)\sum_n \mathbb{1}_I(Y_n(\omega)) - P_{Y_1}(I)\big| > \sigma_B b\Big\} = \Big\{(Y_1,....,Y_n)\in Bo^1\Big\}\ .$$

Then,

$$P\{|P^2_e - P_{Y_1}(I)| > \sigma_B b\} = P_{\theta_2}\Big\{\big|(1/N)\sum_n \mathbb{1}_I(Y_n) - P_{Y_1}(I)\big| > \sigma_B b\Big\}$$

$$= P_{\theta_1}\Big\{\big|(1/N)\sum_n \mathbb{1}_I(Y_n) - P_{Y_1}(I)\big| > \sigma_B b\Big\}[1+Ob(1)N\epsilon_{(1)}]$$

$$= P\{|P^1_e - P_{Y_1}(I)| > \sigma_B b\}[1+Ob(1)N\epsilon_{(1)}]\ .$$

Then,

$$P\{|P^2_e - P_{Y_1}(I)| > \sigma_B b\} \approx \Gamma(b)[1+Ob(1)N\epsilon_{(1)}]\ .$$

Then, there will not be possible to conclude that y_n is a realization of $Y^{\theta_1}_n$ rather than of $Y^{\theta_2}_n$ by testing $P_{Y_1}(I)$. For example, let us suppose $N = 10^4$, $\epsilon_{(1)} = 0.00001$. In this case, for b=2,

$$P\{|P^1_e - P_{Y_1}(I)| > 2\sigma_B\} \approx 0.0455,$$

$$P\{|P^2_e - P_{Y_1}(I)| > 2\sigma_B\} \leq c_2,\ \textit{where}\ c_2 \approx 0.0500\ .$$

Now, if y_n is a realization of $Y^{\theta_1}_n$, it is known that $(1/N)\sum_n \mathbb{1}_I(y_n)$ is close to $P_{Y_1}(I)$ with a certain probability : it is completely possible that $(1/N)\sum_n \mathbb{1}_I(y_n)$ is enough different from $P_{Y_1}(I)$, but the probability that occurs is weak.

Moreover, if y_n is a realization of $Y^{\theta_2}_n$, it is also possible that $(1/N)\sum_n \mathbb{1}_I(y_n)$ is enough different from $P_{Y_1}(I)$, but that is not likely much more to occur than if y_n is a realization of $Y^{\theta_1}_n$.

Then, for the test associated to $P_{Y_1}(I)$, it will be thus impossible to differentiate the model $Y_n^{\theta_1}$ and $Y_n^{\theta_2}$ as good model for the sequence y_n.

These results are not only true for the estimate of only one $P_{Y_1}(I)$, but of several : $P_{Y_1}(I_s)$, s=1,2,....,D. I.e. one can generalize them to the chi squared-test. Indeed, in this case, one is thus interested to the probability

$$P\Big\{N\sum_s\Big[\frac{1}{N}\sum_n \mathbb{1}_{I_s}(Y_n^{\theta_1})-p_s\Big]^2\geq a\Big\}$$

where $p_s = P_{Y_1}(I_s)$. In this case, one uses the Borel sets

$$Bo^2=\Big\{\omega\in\Omega\ \Big|\ N\sum_s\Big[\frac{1}{N}\sum_n \mathbb{1}_{I_s}(Y_n)-p_s\Big]^2\geq a\Big\}.$$

Therefore,

$$P\Big\{N\sum_s\Big[\frac{1}{N}\sum_n \mathbb{1}_{I_s}(Y_n^{\theta_2})-p_s\Big]^2\geq a\Big\}=P\Big\{N\sum_s\Big[\frac{1}{N}\sum_n \mathbb{1}_{I_s}(Y_n^{\theta_1})-p_s\Big]^2\geq a\Big\}[1+Ob(1)\epsilon_{(N)}]\ .$$

Then, if $\epsilon_{(N)}$ is small enough, one cannot differentiate $Y_n^{\theta_1}$ and $Y_n^{\theta_2}$ by this chi squared test.

One can generalize these results in dimension 2: for example

$$P\Big\{\frac{1}{N-1}\sum_n \mathbb{1}_A(Y_n^{\theta_2})\mathbb{1}_B(Y_{n+1}^{\theta_2})=k\Big\}=P\Big\{\frac{1}{N-1}\sum_n(\mathbb{1}_A(Y_n^{\theta_1})\mathbb{1}_B(Y_{n+1}^{\theta_1})=k\Big\}\big[1+Ob(1)\epsilon_{(N)}\big]\ .$$

In dimension p , one uses $\sum_n \mathbb{1}_{Bo_1}(Y_{n+j_1}^{\theta_1})$.....$\mathbb{1}_{Bo_p}(Y_{n+j_p}^{\theta_1})$. Of course, one can also generalize to other functions, i.e. to about the totality of the known tests. Because of it, it seems impossible to differentiate $Y_n^{\theta_1}$ and $Y_n^{\theta_2}$ as models of y_n .

Then, we have just studied the tests associated to these models. In order to be able to apply them it is useful to be able to use the CLT. Now, in general, the sequences y_n which we use are asymptotically independent (for example texts or numbers provided by machines). The models where the CLT is checked are thus correct. The conclusions that we deduce of it are thus correct too : it is impossible to differentiate $Y_n^{\theta_1}$ and $Y_n^{\theta_2}$ as models of y_n .

Now, even if y_n is asymptotically independent, a model not asymptotically independent could maybe be a correct model. What could one say in this case? It seems that one would arrive to the same conclusion because two correct models cannot give different conclusions : cf reasoning of section 6.

4.5 Another case

As a matter of fact, the relation $P\{(Y_1^{\theta_2},.....,Y_N^{\theta_2})\in Bo\}=P\{(Y_1^{\theta_1},.....,Y_N^{\theta_1})\in Bo\}$ $[1+Ob(1)\epsilon_{(N)}]$ for all Borel set $Bo\subset\{0/m,1/m,....,(m-1)/m\}^N$ is a very strong relation. Because of it, it seems impossible to differentiate $Y_n^{\theta_1}$ and $Y_n^{\theta_2}$ as models of y_n in other cases than the case where the CLT holds.

For example, this results holds also if only the Weak Law of Large Number holds. Indeed one does not know the exact law of $P_e - P_{Y_1}(I)$. But it exists theoretically. However, to know this law is not important : it is enough that one has the relation $P\{|P_e^2 - P_{Y_1}(I)| > b\} =$

$P\{|P_e^1 - P_{Y_1}(I)| > b\}[1 + Ob(1)\epsilon_{(N)}]$ for all $b > 0$ in order to be able to conclude from it that one will cannot differentiate the models $Y_n^{\theta_1}$ and $Y_n^{\theta_2}$.

Moreover, the inequality of Bienaymé-Tschebischeff shows that the sums divided by the variance are normalized. One deduced from it that one cannot differentiate the effects of these models.

4.6 General Case

One now asks if to prove this result in the general case is possible, i.e. if, whatever the model $Y_n^{\theta_1}$ (for example without tests), the relation $P\{(Y_1^{\theta_2}, ..., Y_N^{\theta_2}) \in Bo\} = P\{(Y_1^{\theta_1}, ..., Y_N^{\theta_1}) \in Bo\}[1 + Ob(1)\epsilon_{(N)}]$ implies always that one cannot differentiate $Y_n^{\theta_1}$ and $Y_n^{\theta_2}$. It is maybe the case. But, in order to prove it, there is likely philosophical or other problems of the type of the definition of the randomness of Franklin. That is thus likely a complicated study.

But one can say still a certain number of thing in the general case.

Is the problem it resolvable? At first, in the general case, a first question is : how for an unspecified Borelien Bo, can $(Y_1^{\theta}, ..., Y_N^{\theta}) \in Bo$ it be depending on a sample of size N? A priori, it is difficult to establish a connection, considering, in this case, the sample is precisely $(y_1, ..., y_N)$: in space $\{0/m, 1/m, ., (m-1)/m\}^N$, one has a sample of size 1.

Then it seems that it is not possible to have many informations on the model Y_n^{θ} if there is just the sample y_n. However yes : we are interested to the equality $P\big\{(Y_1^{\theta_2}, ..., Y_N^{\theta_2}) \in Bo\big\} = P\big\{(Y_1^{\theta_1}, ..., Y_N^{\theta_1}) \in Bo\big\}[1 + \epsilon_{(N)}]$. It is a strong relation especially for a sample of size 1. Moreover, in this case, there is no problem

Can one use it in the general case? Why not? Indeed, let us suppose that one has a correct model $Y_n^{\theta_1}$. To check $P\big\{(Y_1^{\theta_2}, ..., Y_N^{\theta_2}) \in Bo\big\} = P\big\{(Y_1^{\theta_1}, ..., Y_N^{\theta_1}) \in Bo\big\}[1 + Ob(1)\epsilon_{(N)}]$ leads to admit that the models $Y_n^{\theta_2}$ is correct: that intuitively seems a logical conclusion. But one does not understand how, in the state of our knowledge, one can prove it. In order to prove it, it would be necessary, for example, to study completely all the models asymptotically independent, at first checking the CLT and to know if it is enough that almost all the tests have good results. Moreover, all the possible models would have to be studied.

Empirical probability It is observed now that, if a model $Y_n^{\theta_1}$ is correct and a model $Y_n^{\theta_2}$ is not correct, it would be necessary that a variation of the probability which would be smaller than $P\big\{(Y_1^{\theta_1}, ..., Y_N^{\theta_1}) \in Bo\big\}\epsilon_{(N)}$ exchange something sufficiently important so that one understands a difference of the models with respect to the sample. Therefore, the probability in question will be close to the empirical probability (the empirical probability with any dimension, including N). Thus the model would be very close to the empirical model.

However, the empirical model is in general a bad theoretical model. Thus, in the case of texts, it is known a priori that the empirical probability is not the good model because it will fail as soon as one increases N. One thus arrives at a contradiction.

Then, even if the empirical probability can be selected like correct model, a probability of a model $Y_n^{\theta_2}$ where one changes only a little this probability [3] is also correct.

It would be thus astonishing that a model as special as the empirical model $Y_n^{\theta_1}$ satisfies effectively that, if $Y_n^{\theta_1}$ is correct, an approximate model $Y_n^{\theta_2}$ will be it also and that an unspecified

[3] In the case of empirical probability or of probability concentrated in only some points, the assumption of the probability chosen randomly in hypothesis 6.3.4 of [7] and used in order to define the quantity of models has a negligible probability to be checked : one is in a case $Proba(\Omega") = 0$ where $\Omega" \subset \Omega$ (cf Hypothesis 6.3.4 of [7]). In the case of probability concentrated in some points, it is better to choose the continuous case with a large K_0 in order to study the relations $P\{Y_n^{\theta_2} \in Bo\} = P\{Y_n^{\theta_1} \in Bo\}[1 + \epsilon]$: property 6.3.5 and appendix A of [7]

model does not check this implication. In particular, it would be astonishing for models with continuous density and coefficient of Lipschitz not too large. It would be even astonishing for models with unspecified coefficient of lipschitz, i.e. in the general case. Of course astonishing means that this is intuitive.

Presentation of the problem In fact, this intuition is based on the following reasoning: if $Y_n^{\theta_1}$ is a correct model for the sequence y_n, that means that the event "the sequence y_n is the result of a choice at random of ω where $y_n = Y_n^{\theta_1}(\omega)$" is an event which has reasonable probability to be carried out. Then, it is not understood what can prevent that $y_n = Y_n^{\theta_2}(\omega)$ is a realization equally probable if one changes only a little the probabilities.

The only cases where they could have problem seem those of the probability concentrated close to some points like the empirical probability. But one has just understood that even in this case, it is still true.

One thus understands well what leads to think that, in all the cases, one will not be able to differentiate $Y_n^{\theta_1}$ and $Y_n^{\theta_2}$.

4.7 A problem

Non transitivity But it is necessary to add something to these assertions. If the model $Y_n^{\theta_1}$ is correct and that the model $Y_n^{\theta_2}$ is also correct, a model $Y_n^{\theta_3}$ equivalent with a margin of $\epsilon_{(N)}$ to $Y_n^{\theta_2}$ would be it also correct with the relation $P\{(Y_1^{\theta_3}, ..., Y_N^{\theta_3}) \in Bo\} = P\{(Y_1^{\theta_1}, ..., Y_N^{\theta_1}) \in Bo\}[1 + \mathbf{2}\mathrm{Ob}(1)\epsilon_{(\mathrm{N})}]$? A priori not inevitably!

If it is admitted, one would manage to find that the models $Y_n^{\theta_p}$ checking $P\{(Y_1^{\theta_p}, ..., Y_N^{\theta_p}) \in Bo\} = P\{(Y_1^{\theta_1}, ..., Y_N^{\theta_1}) \in Bo\}[1 + \mathbf{p}\mathrm{Ob}(1)\epsilon_{(\mathrm{N})}]$ would be also correct for all p. One would end up finding models which would not be correct.

Therefore, there is no reason that $Y_n^{\theta_3}$ is also correct. It cannot be differentiated of $Y_n^{\theta_2}$, but not of $Y_n^{\theta_1}$. In other words, this relation is not transitive.

The problem That thus poses a problem because if one uses for example a realization y_n of the IID model, and that if one takes for sequence $Y_n^{\theta_1}$ a model checking $P\{(Y_1^{\theta_1}, ..., Y_N^{\theta_1}) \in Bo\} = L(Bo)[1 + Ob(1)\epsilon^1]$ where ϵ^1 is small enough but not very small, there are no reasons a priori that $Y_n^{\theta_2}$ is a correct model. Indeed, in order that $Y_n^{\theta_2}$ is not correct, it is enough that $Y_n^{\theta_1}$ is in extreme cases of the correct models, i.e. it is enough that ϵ^1 is in extreme cases of the possible values of the ϵ's such that $P\{(Y_1^{\theta}, ..., Y_N^{\theta}) \in Bo\} = L(Bo)[1 + Ob(1)\epsilon]$, $sup_{Bo}(Ob(1)) = 1$, imply that Y_n^{θ} is a correct model.

Answers One could want to answer to this objection by taking the ϵ more and more small. But it would thus be necessary that the ϵ^0 equal to the upper limit of the possible ϵ's cannot also give a correct model.

One can also introduce a second definition, that of models perfectly correct where the conditional probabilities would be the most concentrated possible, but where $Y_n^{\theta_1}$ would remain a valid model. But there is a problem : in the case of an IID sample, the model where the conditional probabilities are less concentrated possible is the IID model, therefore, a priori, one of the best possible models. It is thus a contradiction in our definitions.

It is thus necessary to return to the definition which we have given, that of correct models where the conditional probabilities are the best possible ones and to use differently it. In particular, it is necessary that ϵ^1 is not the limit of the possible values in the case of an IID sample.

Finally, a simple solution will be indeed to give a new definition, that of perfectly correct models by using the relation $P\{(Y_1^{\theta_2}, ..., Y_N^{\theta_2}) \in Bo\} = P\{(Y_1^{\theta_1}, ..., Y_N^{\theta_1}) \in Bo\}[1 + Ob_{Bo}(1)\epsilon]$ where $|Ob_{Bo}(1)| \leq 1$ and $sup_{Bo}\big(|Ob_{Bo}(1)|\big) = 1$.

Perfectly correct model It is said that a model $Y_n^{\theta_{pfc}}$ is perfectly correct if

1) It is a correct model : $Y_n^{\theta_{pfc}} \in \mathcal{MC}(y_n)$, the set of the correct models of y_n.

2) $Y_n^{\theta} \in \mathcal{MC}(y_n)$ if $P\{(Y_1^{\theta},...,Y_N^{\theta}) \in Bo\} = P\{(Y_1^{\theta_{pfc}},...,Y_N^{\theta_{pfc}}) \in Bo\}[1 + Ob(1)\epsilon_{\theta^S_{pfc}}] \Longrightarrow Y_n^{\theta} \in \mathcal{MC}(y_n)$ where $\epsilon^S_{\theta_{pfc}} = sup_{\theta_c \in \mathcal{MC}(y_n)}(\epsilon_{\theta_c})$ when $\epsilon_{\theta_c} = sup\Big\{\epsilon \;\Big|\; P\{(Y_1^{\theta},...,Y_N^{\theta}) \in Bo\} = P\{(Y_1^{\theta_c},...,Y_N^{\theta_c}) \in Bo\}[1 + Ob_{Bo}(1)\epsilon] \Longrightarrow Y_n^{\theta} \in \mathcal{MC}(y_n)\Big\}$.

Let us take again the above example about X_n^{ϵ}. It seems that a perfectly correct model would be that where $P\{X_n^{\epsilon} \in [1/2,1]\} = P_e\{[1/2,1]\}$, the empirical probability of [1/2,1] if the sample x_n is such as the conditions showing independence and that the law is the same one for all n are checked.

But, for a such model with a sample of size 10, the model where $P\{X_n^{\epsilon} \in [1/2,1]\} = P_e\{[1/2,1]\}[1 + 1/1000]$ will be probably correct, but not perfectly correct.

Remark Let us notice that the existence of perfectly correct models, seems not to pose a problem : as soon as there exist two correct models close with a margin of ϵ, it is likely that there exist perfectly correct models. The only difficulties are those which one could meet if the upper limit was not reached. One could maybe prove that this case is not posed by taking the points of accumulations. But there is no utility to make such mathematical proofs, the more so as the mechanism of the reasoning and the goal (cf following section) are easy to understand. This proves to be useless more especially as the definition of the correct models is undetermined mathematically. In any case, one could circumvent the difficulties by modifying the definitions a little. One could same introduce models locally perfectly correct , etc

Another answer There is another answer to the problem in order to prove that there exists a correct model Y_n^{θ} such that X_n^{θ} is the IID model. We introduce it in the following section. Then, the use of perfectly correct models will be useless in the following section. But it will be usefull latter.

5 Exact IID model

Generally, if Y_n^{θ} is a correct model such as $T_q(Y_n^{\theta})$ cannot be differentiated with the IID model, one will be able to choose another correct model $Y_n^{\theta_0}$ close to Y_n^{θ} and such that $T_q(Y_n^{\theta_0})$ is exactly the IID model. It is a manner simpler to show that $T_q(y_n)$ is an IID sequence.

Property 5.1 *One assumes that m is large enough. Let $Y_n^{\theta_c}$ be a correct model of the sequence y_n. One assumes that there exists $\epsilon_Y > 0$ such that if Y_n^{θ} is a model satisfying, for all Borel set Bo, $P\big\{(Y_1^{\theta},.....,Y_N^{\theta}) \in Bo\big\} = P\big\{(Y_1^{\theta_c},.....,Y_N^{\theta_c}) \in Bo\big\}[1 + Ob(1)\epsilon_Y]$, then Y_n^{θ} is a correct model of y_n.*

One assumes also that, for all $(k_1,....,k_N)$,

$$P\big\{\{T_q(Y_1^{\theta_c}) = k_1/2^q\} \cap ..... \cap \{T_q(Y_N^{\theta_c}) = k_N/2^q\}\big\} = \frac{1}{2^{qN}}[1 + \epsilon_{k_1,....,k_N}(q)]$$

where $sup_{k_1,....,k_N}|\epsilon_{k_1,....,k_N}(q)| = \epsilon_X(q)$. One assumes that $\epsilon_X(q)$ is increasing, that $\epsilon_X(1) << \epsilon_Y$ and that there exists $q_1 \in \mathbb{N}^*$ such that $\epsilon_X(q_1)$ is small enough.

Then, there exists $q_0 \in \mathbb{N}^*$ and a correct model $Y_n^{\theta_0}$ of the sequence $\{y_n\}_{n=1,...,N}$ such that, for all $(k_1,....,k_N)$,

$$P\big\{\{T_{q_0}(Y_1^{\theta_0}) = k_1/2^q\} \cap ..... \cap \{T_{q_0}(Y_N^{\theta_0}) = k_N/2^q\}\big\} = \frac{1}{2^{q_0 N}} \ .$$

Proof There exists $q_0 \leq q_1$ such that $\epsilon_X(q_0) \leq (1/2)\epsilon_Y$. Then, one uses the model $Y_n^{\theta_0}$ such that, for all $(k_1, ..., k_N)$,

$$P\big\{(Y_1^{\theta_0}, ..., Y_N^{\theta_0}) = (y'_1, ..., y'_N)\big\} = \frac{P\big\{(Y_1^{\theta_c}, ..., Y_N^{\theta_c}) = (y'_1, ..., y'_N)\big\}}{1 + \epsilon_{k_1,....,k_N}(q_0)}$$

for all $y'_1 \in T_{q_0}^{-1}(k_1/2^{q_0}), ..., y'_N \in T_{q_0}^{-1}(k_N/2^{q_0})$. It ckecks

$$P\big\{\{T_{q_0}(Y_1^{\theta_0}) = k_1/2^{q_0}\} \cap ..... \cap \{T_{q_0}(Y_N^{\theta_0}) = k_N/2^{q_0}\}\big\} = \frac{1}{2^{q_0 N}} \ .$$

It ckecks also : for all $(y'_1, ..., y'_N)$,

$$P\big\{(Y_1^{\theta_0}, ..., Y_N^{\theta_0}) = (y'_1, ..., y'_N)\big\} = P\big\{(Y_1^{\theta_c}, ..., Y_N^{\theta_c}) = (y'_1, ..., y'_N)\big\}[1 + Ob(1)\epsilon'_Y]$$

where $|\epsilon'_Y| \leq C_0 \approx \epsilon_X(q_0)$. Then, $|\epsilon'_Y| < \epsilon_Y$. Then, for all Borel sets Bo,

$$P\big\{(Y_1^{\theta_0}, ..., Y_N^{\theta_0}) \in Bo\big\} = P\big\{(Y_1^{\theta_c}, ..., Y_N^{\theta_c}) \in Bo\big\}[1 + Ob(1)\epsilon'_Y] \ .$$

Then, $Y_n^{\theta_0}$ is a correct model of y_n. Moreover $T_{q_0}(Y_n^{\theta_0})$ is the IID model. ■

Now, it is known that one can find models correct Y_n^{θ} such that $P\{(X_1^{\theta_c}, ..., X_N^{\theta_c}) \in Bo\} = L(Bo)[1 + Ob(1)\epsilon']$ where ϵ' is increasingly small if q decreases. Indeed, by property 6.3.5 of [7], it is known that it is true for the models $Y_n^{\theta_t}$ with a coefficient of Lipschitz K'_0 not too large : $P\{T_q(Y_n^{\theta_t}) = k/2^q \mid T_q(Y_{n-s}^{\theta_t}) = x'_s, s = 1, 2, ..., p\} = (1/2^q)\big[1 + \frac{O(1)K'_0}{m/2^q}\big]$. For the more general models there are similar results into proposition 6.3.5 of [7].

Remark that the value of the ϵ_Y's such that $P\{(Y_1^{\theta}, ..., Y_N^{\theta}) \in Bo\} = P\{(Y_1^{\theta_c}, ..., Y_N^{\theta_c}) \in Bo\}[1 + Ob_{Bo}(1)\epsilon_Y] \Longrightarrow Y_n^{\theta} \in \mathcal{MC}(y_n)$ depends very little on m. For example, if $\epsilon_Y \leq 1/100$, it is likely that almost all the tests checked by $Y_n^{\theta_c}$ will be it by Y_n^{θ} (cf section 4.4).

Thus **there exists indeed m sufficiently large and q sufficiently small and a correct model $Y_n^{\theta_0} \in \{0/m, 1/m, ..., (m-1)/m\}$ such that $T_q(Y_n^{\theta_0})$ is the IID model**.

In fact, there exist an infinity of correct models Y_n^{θ} such that $P\big\{(X_{n+j_1}^{\theta}, ..., X_{n+j_p}^{\theta}) \in Bo\big\} = L(Bo)$. In particular, if y_n means texts, it is true for the models with a coefficient of Lipchitz not too large. This result thus makes possible to have a very clear conclusion : x_n has like correct model the IID model.

That means that x_n behaves like any IID sample : a priori, x_n can check not the properties which one awaits from a IID sample like certain tests, but that occurs only with a probability equal to that of any IID sample.

6 The sequence x_n is IID for all the correct models

In this section, one will understand that one can deduce from the results about the Y_n^{θ} that if y_n n=1,2,., N, represents a text, $T_q(Y_n^{\theta})$ is indifferentiable from an IID sequence for all the correct models of y_n.

Indeed, one understood in [7] that, for some models with continous density and coefficient of Lipschitz K_0 not too large, X_n^{θ} cannot be differentiated with an IID sequence. If the y_n are provided by texts, one can admit this assumption.

There thus exists a correct model such as X_n^{θ} is IID. That means, according to the definition which we have given, that one cannot predict $X_{\phi(n)}^{\theta}$ knowing $X_{\phi(n-1)}^{\theta} = x'_1$, $X_{\phi(n-2)}^{\theta} = x'_2$,.... with a distribution other than the uniform distribution.

Also, it seems well that, if, for another correct model, one obtained a different result, there would be fatally one of the two models which would not be correct. Indeed, at the most, only one of the conclusions will be satisfied. For example, on the models of climatic warming, some predict a strong warming and others more restricted : both will not be checked.

Therefore, one cannot have a different result for various correct model. Therefore, for all the correct models Y_n^θ, a priori $T_q(Y_n^\theta)$ cannot be differentiated with an IID sequence.

On the other hand, we wanted to know the effects of T_q on the texts. Then, on all the texts which we have tested, we always have found that the empirical probability P_e checked the following condition : for all p not too large with respect to N, for all intervals I_1, I_2, ..., I_p,

$$P_e\big\{\{X_n \in I_1\} \cap \{X_{n+1} \in I_2\} \cap ..... \cap \{X_{n+p} \in I_p\}\big\} \approx L(I_1)L(I_2)....L(I_p) \ .$$

That corresponds completely with which we have already noted : a text and the $T_q^{-1}(I)$ where I is an interval are independent events. The conditional probability of the X_n^θ are thus sums chosen randomly, and thus asymptotically normal (cf section 6.1.2 of [7]).

The empirical probabilities thus check the fundamental equalities of IID sequences. It would be thus astonishing that the models associated to these sequences does not check the equalities of IID sequences, at least with a margin of ϵ. It is thus normal to think that the correct models will be thus those which will check this fundamental equality.

Other arguments are in sections 13.2, 13.3.2 and 13.3.3 of [6]

A Conditional probabilities of X_n^θ

A correct model Y_n^θ of a sequence y_n is thus a model which represents the associated phenomenon well.

Therefore, it makes possible to predict correctly the future knowing the past, i.e. to define the best possible $P\{Y_n^\theta \in Bo \mid Y_{n-s}^\theta = y'_s, s = 1, 2, 3, ...\}$. It will be thus true also for $P\{T_q(Y_n^\theta) \in Bo' | Y_{n-s}^\theta = y'_s, s = 1, 2, 3, ...\}$. The following proposition shows that it is true also for $P\{X_n^\theta \in Bo' | X_{n-s}^\theta = x'_s, s = 1, 2, 3, ...\}$.

Proposition A.1 *Let $Y_n \in \{0/m, 1/m, ..., (m-1)/m\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{A}, P)$ and let $X_n = T_q(Y_n)$. Then, for all Borel set Bo,*

$$P\{X_n \in Bo \mid X_{n-s} = x_s, s = 1, 2, ..., p\}$$

$$= \sum_{y_{s_1} \in T_q^{-1}(x_1)} .... \sum_{y_{s_p} \in T_q^{-1}(x_p)} \eta_{y_{s_1}, ..., y_{s_p}} P\{X_n \in Bo | Y_{n-j} = y_{s_j}, j = 1, 2, ..., p\}$$

where

$$\sum_{y_{s_1} \in T_q^{-1}(x_1)} .... \sum_{y_{s_p} \in T_q^{-1}(x_p)} \eta_{y_{s_1}, ..., y_{s_p}} = 1 \ .$$

Proof We have :

$$P\{X_n \in Bo | X_{n-s} = x_s, s = 1, 2, ..., p\}$$

$$= \frac{P\big\{\{X_n \in Bo\} \cap \{X_{n-1} = x_1\} \cap .... \cap \{X_{n-p} = x_p\}\big\}}{P\big\{\{X_{n-1} = x_1\} \cap .... \cap \{X_{n-p} = x_p\}\big\}}$$

$$= \frac{P\Big\{\{X_n \in Bo\} \cap \big\{ \cup_{y_{s_1}} \{Y_{n-1} = y_{s_1}\}\big\} \cap .... \cap \big\{ \cup_{y_{s_p}} \{Y_{n-p} = y_{s_p}\}\big\}\Big\}}{P\Big\{\big\{ \cup_{y_{s_1}} \{Y_{n-1} = y_{s_1}\}\big\} \cap .... \cap \big\{ \cup_{y_{s_p}} \{Y_{n-p} = y_{s_p}\}\big\}\Big\}}$$

where $\cup_{y_{s_t}} \{Y_{n-t} = y_{s_t}\} = \cup_{y_{s_t} \in T_q^{-1}(x_t)} \{Y_{n-t} = y_{s_t}\}$.

Then,

$$P\{X_n \in Bo | X_{n-s} = x_s, s = 1, 2, ..., p\}$$

$$= \frac{P\Big\{ \cup_{y_{s_1}} .... \cup_{y_{s_p}} \{X_n \in Bo\} \cap \{Y_{n-1} = y_{s_1}\} \cap .... \cap \{Y_{n-p} = y_{s_p}\}\Big\}}{P\Big\{ \cup_{y_{s_1}} .... \cup_{y_{s_p}} \{Y_{n-1} = y_{s_1}\} \cap .... \cap \{Y_{n-p} = y_{s_p}\}\Big\}}$$

$$= \frac{\sum_{y_{s_1}} .... \sum_{y_{s_p}} P\Big\{\{X_n \in Bo\} \cap \{Y_{n-1} = y_{s_1}\} \cap .... \cap \{Y_{n-p} = y_{s_p}\}\Big\}}{\sum_{y_{s_1}} .... \sum_{y_{s_p}} P\Big\{\{Y_{n-1} = y_{s_1}\} \cap .... \cap \{Y_{n-p} = y_{s_p}\}\Big\}}$$

$$= \sum_{y_{s_1}} .... \sum_{y_{s_p}} \eta_{y_{s_1}, ..., y_{s_p}} \frac{P\Big\{\{X_n \in Bo\} \cap \{Y_{n-1} = y_{s_1}\} \cap .... \cap \{Y_{n-p} = y_{s_p}\}\Big\}}{P\Big\{\{Y_{n-1} = y_{s_1}\} \cap .... \cap \{Y_{n-p} = y_{s_p}\}\Big\}}$$

where

$$\eta_{y_{s_1}, ..., y_{s_p}} = \frac{P\Big\{\{Y_{n-1} = y_{s_1}\} \cap .... \cap \{Y_{n-p} = y_{s_p}\}\Big\}}{\sum_{y_{s_1}} .... \sum_{y_{s_p}} P\Big\{\{Y_{n-1} = y_{s_1}\} \cap .... \cap \{Y_{n-p} = y_{s_p}\}\Big\}} \ .$$

Of course,

$$\sum_{y_{s_1}} .... \sum_{y_{s_p}} \eta_{y_{s_1},...,y_{s_p}} = 1 \ . \ \blacksquare$$

For example assume that the y_n's are provided by texts. Assume that the model with uniform probability is the best (all the english texts are equiprobable).

Therefore, it makes possible to predict correctly the future knowing the past, i.e. to define $P\{T_q(Y_n^\theta) \in Bo | Y_{n-s}^\theta = y'_s, s = 1, 2, 3, ...\}$ which is the best possible forecast for all the Borel set Bo, all the y'_{n-s} and all n. Indeed, for texts a such model exists (cf section 3.2).

Then, proposition A.1 which affims that $P\{X_n^\theta \in Bo | X_{n-s}^\theta = x'_s, s = 1, 2, 3, ...\}$ is a sum of $P\{X_n^\theta \in Bo | Y_{n-s}^\theta = y'_s, s = 1, 2, 3, ...\}$ which are the best forecast of the future for all the y'_s, s=1,2,3,..... shows that $P\{X_n^\theta \in Bo | X_{n-s}^\theta = x'_s, s = 1, 2, 3, ...\}$ is the best possible forecast.

Therefore, if Y_n^θ is a correct model, X_n^θ is also a correct model and $P\{X_n^\theta \in Bo \mid X_{n-s}^\theta = x'_s, s = 1, 2, ..., p\}$ defines correctly the conditional probabilities.

Now, if one uses one alone english text, one does not know that y_n derives from texts. Then, the conditional probabilities $P\{Y_n^{\theta_1} \in Bo | Y_{n-s}^{\theta_1} = y'_s, s = 1, 2, 3, ...\}$ are less concentrated than those associated to the model of the English texts. Then, if one uses one alone english text, the conditional probabilities $P\{X_n^{\theta_1} \in Bo \mid X_{n-s}^{\theta_1} = x'_s, s = 1, 2, ..., p\}$ are less concentrated than the conditional probabilities $P\{X_n^{\theta_0} \in Bo \mid X_{n-s}^{\theta_0} = x'_s, s = 1, 2, ..., p\}$ associated to the model of the transformation of the English texts.

The best way in order to explain it, it is to assume that there exists a continous density with a Lipchitz coefficient $K_0'^T$ for texts. Then if one does not know that one is an English text, one has a Lipschitz coefficient $K_0'^{nT}$ such that $K_0'^{nT} \leq K_0'^T$. Now we recall that, by property 6.3.5 of [7], $P\{X_n^{\theta_t} \in I \mid X_{n-s}^{\theta_t} = x'_s, s = 1, 2, ..., p\} \approx L(I)\big[1 + \frac{6Ob(1)K_0'^t}{m/2^q}\big]$ where $K_0'^0 = K_0'^T$ and $K_0'^1 = K_0'^{nT}$. On the other hand, it is no possible to have conditional probabilities less concentrated than the uniform distribution. We deduce $P\{X_n^{\theta_1} \in Bo \mid X_{n-s}^{\theta_1} = x'_s, s = 1, 2, ..., p\} \approx L(Bo)$.

Another way in order to explain this result is to write that $P\big\{(Y_1^{\theta_0}, ..., Y_N^{\theta_0}) \in Bo\big\} = L(Bo)[1 + Ob(1)\epsilon^0]$ when one assumes that one is in an English text, and $P\big\{(Y_1^{\theta_1}, ..., Y_N^{\theta_1}) \in Bo\big\} = L(Bo)[1 + Ob(1)\epsilon^1]$ when one does not know it, where $\epsilon^1 \leq \epsilon^0$.

B Use of text witten backward

B.1 Use of texts

Now, we suppose that we use sequences y_n obtained from texts.

In an obvious way, the texts are realizations of sequences of random variables: for example, one can take as model, the set of the possible texts provided with the uniform probability. In this model, if one knows a text until the letter "n-1", there are a large number of alternatives for the r following letters as soon as r is rather large. That means indeed that the conditional probability of Y_n^θ knowing the past, is not concentrated in a too small number of points.

However there is a problem for some subsequences $y_{\phi(n)}$: if one knows a text until the letter "n-1" and the text after the letter "n+r", (for example r=18), there will be much less possibilities for the r letters ranging between the two parts of texts than if only the past is known. To answer this point, in sections 11.2.4 of [6], we have added modulo m a text and a text written backward.

But that seems exaggerated because it is not known a priori that we are in an English text if one has only a few texts [4]. Moreover, a priori all the words possible of the English language are not known : one cannot thus predict them. That does not prevent from concluding : if the conditional probabilities of the texts are not concentrated in some points in a model of English text, a fortiori, it is also the case if it is not known that one is in a English text.

Moreover, a pseudo-random sequence is added to used texts (step c, page 93 of [7]). That makes possible to have sequences y_n which have a good randomness (cf [9], or chapter 3 of [6]).

Moreover, it is encore easier to prove that the conditional probability of Y_n^θ knowing the past, is not concentrated in a too small number of points if $y_n = [\overline{e(n) + rand_0(n) + e'(n) + rand_1(n)}]/m$ where e'(n) represent a text witten backward and $rand_j(n)$ pseudo-random sequences for j=0,1, (cf sections 11.2.4 of [6]). In this case, one can show that this condition is correct.

Indeed suppose that the sequences x_n and y_n represents two texts at which one adds to each one a pseudo-random sequences. Let Y_n and X_n be two correct models. One is interested to the sequence $\overline{X_{n+s} + Y_{n-s}}$, $s = 0, \pm 1, \pm 2, ...$ As matter of fact, one adds a text to a text written backward

Then, we will understand that the probability that $\overline{X_n + Y_n} = a_0$ given $\overline{X_{n+s} + Y_{n-s}} = a_s$ for s=1,-1, will be about that of $\overline{X_n + Y_n} = a_0$ given $X_{n-1} = b_1$ et $Y_{n-1} = c_1$.

B.2 Theorem

We have the following theorem

Theorem 1 *Let Y_n and X_n be two independent sequences of random variables defined on a probability space $(\Omega, \mathcal{A}, P)$ such that $X_n,\ Y_n \in \{0/m, 1/m, ..., (m-1)/m\}$. Then,*

$$P\{X_n + Y_n \equiv a_0 \mid X_{n-1} + Y_{n+1} \equiv a_1, X_{n+1} + Y_{n-1} \equiv a_2\}$$
$$= \sum_{x_1, y_1} \eta_{x_1,y_1} \alpha_{x_1,y_1} P\big\{X_n + Y_n \equiv a_0 \ \big|\ Y_{n+1} \equiv a_1 - x_1, X_{n+1} \equiv a_2 - y_1\big\} \ ,$$

where

$$\alpha_{x_1,y_1} = \frac{P\big\{X_{n-1} = x_1, Y_{n-1} = y_1 \ \big|\ X_n + Y_n \equiv a_0, Y_{n+1} \equiv a_1 - x_1, X_{n+1} \equiv a_2 - y_1\big\}}{P\big\{X_{n-1} = x_1, Y_{n-1} = y_1 \ \big|\ Y_{n+1} \equiv a_1 - x_1, X_{n+1} \equiv a_2 - y_1\big\}} \ ,$$

[4] Let us recall difficulties in order to discover the meaning of certain languages in archeology : all are not identifed. Let us recall also the hieroglyphs on the Rosetta Stone whose one had however 3 translations.

$$\eta_{x_1,y_1} = \frac{P\big\{\{X_{n-1}=x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}\ P\big\{\{Y_{n+1}\equiv a_1-x_1\}\cap\{Y_{n-1}=y_1\}\big\}}{\sum_{x_1,y_1} P\big\{\{X_{n-1}=x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}\ P\big\{\{Y_{n+1}\equiv a_1-x_1\}\cap\{Y_{n-1}=y_1\}\big\}}\ ,$$

$$\sum_{x_1,y_1} = \sum_{(x_1,y_1)\in\{0/m,1/m,....,(m-1)/m\}^2}\ ,\ \sum_{x_1,y_1}\eta_{x_1,y_1}=1 \textit{ and } a\equiv b \textit{ if } ma\equiv mb \textit{ modulo } m.$$

Proof We have

$$P\{X_n+Y_n\equiv a_0 \mid X_{n-1}+Y_{n+1}\equiv a_1, X_{n+1}+Y_{n-1}\equiv a_2\}$$

$$= \frac{P\big\{\{X_n+Y_n\equiv a_0\}\cap\{X_{n-1}+Y_{n+1}\equiv a_1\}\cap\{X_{n+1}+Y_{n-1}\equiv a_2\}\big\}}{P\big\{\{X_{n-1}+Y_{n+1}\equiv a_1\}\cap\{X_{n+1}+Y_{n-1}\equiv a_2\}\big\}}$$

$$= \frac{P\Big\{\{X_n+Y_n\equiv a_0\}\cap\Big\{\cup_{x_1}\{X_{n-1}=x_1\}\cap\{X_{n-1}+Y_{n+1}\equiv a_1\}\Big\}\cap\Big\{\cup_{y_1}\{Y_{n-1}=y_1\}\cap\{X_{n+1}+Y_{n-1}\equiv a_2\}\Big\}\Big\}}{P\Big\{\Big\{\cup_{x_1}\{X_{n-1}=x_1\}\cap\{X_{n-1}+Y_{n+1}\equiv a_1\}\Big\}\cap\Big\{\cup_{y_1}\{Y_{n-1}=y_1\}\cap\{X_{n+1}+Y_{n-1}\equiv a_2\}\Big\}\Big\}}$$

$$= \sum_{x_1,y_1}\frac{P\Big\{\{X_n+Y_n\equiv a_0\}\cap\Big\{\{X_{n-1}=x_1\}\cap\{X_{n-1}+Y_{n+1}\equiv a_1\}\Big\}\cap\Big\{\{Y_{n-1}=y_1\}\cap\{X_{n+1}+Y_{n-1}\equiv a_2\}\Big\}\Big\}}{\sum_{x_1,y_1}P\Big\{\Big\{\{X_{n-1}=x_1\}\cap\{X_{n-1}+Y_{n+1}\equiv a_1\}\Big\}\cap\Big\{\{Y_{n-1}=y_1\}\cap\{X_{n+1}+Y_{n-1}\equiv a_2\}\Big\}\Big\}}$$

$$= \sum_{x_1,y_1}\frac{P\Big\{\{X_n+Y_n\equiv a_0\}\cap\Big\{\{X_{n-1}=x_1\}\cap\{X_{n-1}+Y_{n+1}\equiv a_1\}\Big\}\cap\Big\{\{Y_{n-1}=y_1\}\cap\{X_{n+1}+Y_{n-1}\equiv a_2\}\Big\}\Big\}}{P\Big\{\Big\{\{X_{n-1}=x_1\}\cap\{X_{n-1}+Y_{n+1}\equiv a_1\}\Big\}\cap\Big\{\{Y_{n-1}=y_1\}\cap\{X_{n+1}+Y_{n-1}\equiv a_2\}\Big\}\Big\}}$$

$$= \sum_{x_1,y_1}\eta_{x_1,y_1}\frac{P\Big\{\{X_n+Y_n\equiv a_0\}\cap\Big\{\{X_{n-1}=x_1\}\cap\{Y_{n+1}\equiv a_1-x_1\}\Big\}\cap\Big\{\{Y_{n-1}=y_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\Big\}\Big\}}{P\Big\{\Big\{\{X_{n-1}=x_1\}\cap\{Y_{n+1}\equiv a_1-x_1\}\Big\}\cap\Big\{\{Y_{n-1}=y_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\Big\}\Big\}}$$

$$= \sum_{x_1,y_1}\eta_{x_1,y_1}P\Big\{X_n+Y_n\equiv a_0\ \Big|\ X_{n-1}=x_1, Y_{n+1}\equiv a_1-x_1, Y_{n-1}=y_1, X_{n+1}\equiv a_2-y_1\Big\}\ .$$

On the other hand,

$$P\Big\{X_n+Y_n\equiv a_0\ \Big|\ X_{n-1}=x_1, Y_{n+1}\equiv a_1-x_1, Y_{n-1}=y_1, X_{n+1}\equiv a_2-y_1\Big\}$$

$$= Cx_1,y_1P\big\{X_n+Y_n\equiv a_0\big|Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\big\}\ ,$$

where

$$c_{x_1,y_1} = \frac{P\Big\{X_n+Y_n\equiv a_0\ \Big|\ X_{n-1}=x_1, Y_{n+1}\equiv a_1-x_1, Y_{n-1}=y_1, X_{n+1}\equiv a_2-y_1\Big\}}{P\Big\{X_n+Y_n\equiv a_0\ \Big|\ Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\Big\}}$$

$$= \frac{\frac{P\Big\{\{X_n+Y_n\equiv a_0\}\cap\Big\{\{X_{n-1}=x_1\}\cap\{Y_{n+1}\equiv a_1-x_1\}\Big\}\cap\Big\{\{Y_{n-1}=y_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\Big\}\Big\}}{P\Big\{\Big\{\{X_{n-1}=x_1\}\cap\{Y_{n+1}\equiv a_1-x_1\}\Big\}\cap\Big\{\{Y_{n-1}=y_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\Big\}\Big\}}}{\frac{P\Big\{\{X_n+Y_n\equiv a_0\}\cap\{Y_{n+1}\equiv a_1-x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\Big\}}{P\Big\{\{Y_{n+1}\equiv a_1-x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\Big\}}}$$

$$= \frac{\frac{P\Big\{\{X_n+Y_n\equiv a_0\}\cap\Big\{\{X_{n-1}=x_1\}\cap\{Y_{n+1}\equiv a_1-x_1\}\Big\}\cap\Big\{\{Y_{n-1}=y_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\Big\}\Big\}}{P\Big\{\{X_n+Y_n\equiv a_0\}\cap\{Y_{n+1}\equiv a_1-x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\Big\}}}{\frac{P\Big\{\Big\{\{X_{n-1}=x_1\}\cap\{Y_{n+1}\equiv a_1-x_1\}\Big\}\cap\Big\{\{Y_{n-1}=y_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\Big\}\Big\}}{P\Big\{\{Y_{n+1}\equiv a_1-x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\Big\}}}$$

$$= \frac{P\Big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\Big|\; X_n+Y_n\equiv a_0, Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\Big\}}{P\Big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\Big|\; Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\Big\}} \;.\; \blacksquare$$

B.3 Application

Let us suppose again that the sequences x_n and y_n represents texts at which one adds to each one a pseudo-random sequence. It is supposed that Y_n and X_n are two correct models. One is interested by $\overline{X_{n+s}+Y_{n-s}}$, $s=0,\pm1,\pm2,....$: one adds a text and a text written backward.

Study of $P\big\{X_n+Y_n\equiv a_0 \;\big|\; Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\big\}$ We know that $P\big\{X_n+Y_n\equiv a_0 \;\big|\; Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\big\}$ is the conditional probability that $X_n+Y_n\equiv a_0$ given the futures Y_{n+1} and X_{n+1}.

There will be thus a probability which will not be more concentrated that of a text knowing the future. But it is an increase: the probability of the sum $\overline{X_n+Y_n}$ knowing the future $Y_{n+1}\equiv a_1-x_1$ and $X_{n+1}\equiv a_2-y_1$ is probably less concentrated than, for example, the probability of X_n knowing the future $X_{n+1}\equiv a_2-y_1$.

In fact, the conditional probability will be much less concentrated than that: it is not known that one is in a text. Moreover, because a pseudo-random generator is added, this probability will be rather close to that of independence : $P\big\{X_n+Y_n\equiv a_0 \;\big|\; Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\big\}$ is not too distant from $P\big\{X_n+Y_n\equiv a_0\big\}$ which is not too distant from 1/m (cf pages 199-202 of [6]).

Therefore, the probability of the sum X_n+Y_n knowing the future is not concentrated close to some points. That means that there will be no points where it is close to 0, and not points where it is close to 1. That means that, in the case of models with continuous density, the coefficient of Lipschitz will not be too large.

Study of $P\{X_{n-1}=x_1, Y_{n-1}=y_1 \mid Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\}$ Now considering the independence of texts X_n and Y_n, $P\big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\big|\; Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\big\} = P\big\{X_{n-1}=x_1 \;\big|\; X_{n+1}\equiv a_2-y_1\big\}P\big\{Y_{n-1}=y_1 \;\big|\; Y_{n+1}\equiv a_1-x_1\big\}$.

However, for the texts, as soon as one takes as sequence y_n a sequence of group of Q=10 or 20 letters for example, one finds the Q-dependence statistically (chapter 10 of [6]).

Therefore, $P\Big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\Big|\; Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\Big\}\approx 1/m^2$ if m is large enough.

Study of $P\big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\big|\; X_n+Y_n\equiv a_0, Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\big\}$ One understands, by simulation, that $P\big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\big|\; X_n+Y_n\equiv a_0, Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\big\}$ is not too different from $P\big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\big|\; X_n+Y_n\equiv a_0\big\}$.

It is not astonishing: X_{n-1} is almost independent of X_{n+1}. Therefore, $P\big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\big|\; X_n+Y_n\equiv a_0, Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\big\}$ depends especially on X_n+Y_n [5].

One can also understand it because of following relations

$$P\big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\big|\; X_n+Y_n\equiv a_0, Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\big\}$$

$$=\frac{P\big\{\{X_{n-1}=x_1\}\cap\{Y_{n-1}=y_1\}\cap\{X_n+Y_n\equiv a_0\}\cap\{Y_{n+1}\equiv a_1-x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}}{P\big\{\{X_n+Y_n\equiv a_0\}\cap\{Y_{n+1}\equiv a_1-x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}}$$

$$=\frac{P\big\{\{X_{n-1}=x_1\}\cap\{Y_{n-1}=y_1\}\cap\big\{\cup_{x_0}\{X_n=x_0\}\cap\{X_n+Y_n\equiv a_0\}\cap\{Y_{n+1}\equiv a_1-x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}}{P\big\{\big\{\cup_{x_0}\{X_n=x_0\}\cap\{X_n+Y_n\equiv a_0\big\}\cap\{Y_{n+1}\equiv a_1-x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}}$$

$$=\sum_{x_0}\frac{P\big\{\{X_{n-1}=x_1\}\cap\{Y_{n-1}=y_1\}\cap\big\{\{X_n=x_0\}\cap\{Y_n\equiv a_0-x_0\}\big\}\cap\{Y_{n+1}\equiv a_1-x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}}{\sum_{x_0}P\big\{\big\{\{X_n=x_0\}\cap\{Y_n\equiv a_0-x_0\}\big\}\cap\{Y_{n+1}\equiv a_1-x_1\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}}$$

$$=\sum_{x_0}\frac{P\big\{\{X_{n-1}=x_1\}\cap\{X_n=x_0\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}P\big\{\{Y_{n-1}=y_1\}\cap\{Y_n\equiv a_0-x_0\}\cap\{Y_{n+1}\equiv a_1-x_1\}\big\}}{\sum_{x_0}P\big\{\{X_n=x_0\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}P\big\{\{Y_n\equiv a_0-x_0\}\cap\{Y_{n+1}\equiv a_1-x_1\}\big\}}$$

$$=\sum_{x_0}\beta_{x_0}\frac{P\big\{\{X_{n-1}=x_1\}\cap\{X_n=x_0\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}P\big\{\{Y_{n-1}=y_1\}\cap\{Y_n\equiv a_0-x_0\}\cap\{Y_{n+1}\equiv a_1-x_1\}\big\}}{P\big\{\{X_n=x_0\}\cap\{X_{n+1}\equiv a_2-y_1\}\big\}P\big\{\{Y_n\equiv a_0-x_0\}\cap\{Y_{n+1}\equiv a_1-x_1\}\big\}}$$

$$=\sum_{x_0}\beta_{x_0}P\big\{X_{n-1}=x_1|X_n=x_0, X_{n+1}\equiv a_2-y_1\big\}P\big\{Y_{n-1}=y_1|Y_n\equiv a_0-x_0, Y_{n+1}\equiv a_1-x_1\big\}$$

where $=\sum_{x_0}\beta_{x_0}=1$.

It is not too difficult to understand, that, for example, $P\{X_{n-1}=x_1|X_n=x_0, X_{n+1}\equiv a_2-y_1\big\}$ is hardly more concentrated than $P\{X_{n-1}=x_1|X_n=x_0\big\}$ if x_n represents only texts. It is even truer if x_n represents a text to which one adds a pseudo random sequence, and it is even truer in the case which interests us considering than one summons on all the x_0.

Then, it is not astonishing that $P\big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\big|\; X_n+Y_n\equiv a_0, Y_{n+1}\equiv a_1-x_1, X_{n+1}\equiv a_2-y_1\big\}$ is not too different from $P\big\{X_{n-1}=x_1, Y_{n-1}=y_1 \;\big|\; X_n+Y_n\equiv a_0\big\}$.

[5] In the general case, that could be false : e.g. cf the properties of higher order correlation coefficients (cf [5])

Now, $P\{X_n + Y_n \equiv a_0\} \approx 1/m$ because one adds a pseudo random sequence to text (cf pages 199-202 of [6]). Therefore,

$$P\big\{X_{n-1} = x_1, Y_{n-1} = y_1 \;\big|\; X_n + Y_n \equiv a_0\big\} = \frac{P\{X_{n-1} = x_1, Y_{n-1} = y_1, X_n + Y_n \equiv a_0\}}{P\{X_n + Y_n \equiv a_0\}}$$

$$\approx m.P\big\{X_{n-1} = x_1, Y_{n-1} = y_1\big\}\frac{P\big\{X_{n-1} = x_1, Y_{n-1} = y_1, X_n + Y_n \equiv a_0\big\}}{P\big\{X_{n-1} = x_1, Y_{n-1} = y_1, \big\}}$$

$$= m.P\big\{X_{n-1} = x_1\big\}P\big\{Y_{n-1} = y_1\big\}P\big\{X_n + Y_n \equiv a_0\big|X_{n-1} = x_1, Y_{n-1} = y_1, \big\}$$

$$\approx (1/m)P\big\{X_n + Y_n \equiv a_0\big|X_{n-1} = x_1, Y_{n-1} = y_1, \big\} \; .$$

Of course, $P\big\{X_n + Y_n \equiv a_0\big|X_{n-1} = x_1, Y_{n-1} = y_1, \big\}$ is, this time, the conditional probability knowing the past. There are thus about the same results that above for $P\big\{X_n + Y_n \equiv a_0 \;\big|\; Y_{n+1} \equiv a_1 - x_1, X_{n+1} \equiv a_2 - y_1\big\}$. Therefore, $P\big\{X_n + Y_n \equiv a_0\big|X_{n-1} = x_1, Y_{n-1} = y_1, \big\}$ will be not too different from 1/m.

Conclusion By joining together all these results, one understands that

$$\alpha_{x_1,y_1}P\big\{X_n + Y_n \equiv a_0 \;\big|\; Y_{n+1} \equiv a_1 - x_1, X_{n+1} \equiv a_2 - y_1\big\}$$

will be not too different from 1/m.

Now,

$$\eta_{x_1,y_1} = \frac{P\big\{\{X_{n-1} = x_1\} \cap \{X_{n+1} \equiv a_2 - y_1\}\big\}P\big\{\{Y_{n+1} \equiv a_1 - x_1\} \cap \{Y_{n-1} = y_1\}\big\}}{\sum_{x_1,y_1} P\big\{\{X_{n-1} = x_1\} \cap \{X_{n+1} \equiv a_2 - y_1\}\big\}P\big\{\{Y_{n+1} \equiv a_1 - x_1\} \cap \{Y_{n-1} = y_1\}\big\}}$$

$$\approx \frac{P\{X_{n-1} = x_1\}P\{X_{n+1} \equiv a_2 - y_1\}P\{Y_{n+1} \equiv a_1 - x_1\}P\{Y_{n-1} = y_1\}}{\sum_{x_1,y_1} P\{X_{n-1} = x_1\}P\{X_{n+1} \equiv a_2 - y_1\}P\{Y_{n+1} \equiv a_1 - x_1\}P\{Y_{n-1} = y_1\}}$$

$$\approx \frac{1/m^4}{\sum_{x_1,y_1}(1/m^4)} \approx 1/m^2 \; .$$

Therefore,

$$\sum_{x_1,y_1} \eta_{x_1,y_1}\alpha_{x_1,y_1}P\big\{X_n + Y_n \equiv a_0 \;\big|\; Y_{n+1} \equiv a_1 - x_1, X_{n+1} \equiv a_2 - y_1\big\}$$

is not too different from

$$(1/m^2)\sum_{x_1,y_1} P\big\{X_n + Y_n \equiv a_0 \;\big|\; Y_{n+1} \equiv a_1 - x_1, X_{n+1} \equiv a_2 - y_1\big\} \; .$$

However in general, to make a sum on x_1, y_1 standardizes the probabilities (it is true as soon as one can consider that they are randomly selected cf section 6.1.2 of of [7]). Therefore, in most case, $P\{X_n + Y_n \equiv a_0 \mid X_{n-1} + Y_{n+1} \equiv a_1, X_{n+1} + Y_{n-1} \equiv a_2\}$ will be even more close to (1/m) that the previous reasonings which is carry out without the sums $\sum_{x_1,y_1}$ did not let it suppose.

Finally, all this confirms that $P\{X_n + Y_n \equiv a_0 \mid X_{n-1} + Y_{n+1} \equiv a_1, X_{n+1} + Y_{n-1} \equiv a_2\}$ is not too different from 1/m. One deduces from it that the coefficient of Lipschitz will not be too large. Then, it is enough to apply T_q in order to have sequences proved IID.

References

[1] KNUTH D.E. (1998) the Art of Computer Programming; Vol 2. Third Edition Addison-Wesley, Reading, Massachusetts.

[2] GENTLE J. (1984) Random Number Generation and Monte Carlo Method, Springer 13, 61-81.

[3] MENEZES A., VAN OORSCHOT P. , VANSTONE S. (1996) Handbook of Applied Cryptography, CRC Press, 1996.

[4] SCHNEIER B (1996) Applied Cryptography 2nd Edition, John Wiley and sons, Inc

[5] BLACHER R. (1993) Higher Order Correlation Coefficients. Statistics 25, 1-15.

[6] BLACHER R. (2009) A Perfect Random Number Generator. Rapport de Recherche LJK Universite de Grenoble. http://hal.archives-ouvertes.fr/hal-00426555/fr/

[7] BLACHER R. (2010) A Perfect Random Number Generator II. Rapport de Recherche LJK Universite de Grenoble. http://hal.archives-ouvertes.fr/hal-00443576/fr/.

[8] FRANKLIN J. N. (1963). Deterministic simulation of random processes. Math. Comp., 17, 28-59.

[9] DENG L. Y. and GEORGE, E. O. (1992) Some characterizations of the uniform distribution with applications to random number generation. Ann. Inst. Statist. Math. Vol. 44, No. 2, pp. 379-385