ARKIV FÖR MATEMATIK Band 5 nr 31

1.64041 Communicated 26 February 1964 by Harald Cramér and Otto Frostman

Problems of reconstruction in connection with addition of independent stochastic processes

By Ove Frank

1. Introduction

Consider two independent stochastic processes $X=\{X_t : t\in T\}$ and $Y=\{Y_t : t\in T\}$ where the index set T is a subset of real numbers and the components X_t and Y_t take their values in a locally compact topological group G. Suppose that a new stochastic process $Z=\{Z_t=X_t+Y_t : t\in T\}$ is obtained by composition of the components of X and Y according to the group operation which we will name addition and denote by a plus sign. With knowledge of the probabilistic structures of the processes X and Y and with a realization available of the process Z it is desired to reconstruct the belonging realizations of the processes X and Y.

This type of problem is fundamental in information theory where it is desired to reconstruct a message which has been received as an additively disturbed signal. The method of reconstruction which is given in the literature is based upon the work of Wiener [11]. The realization of X is supposed to be a realvalued function which loosely speaking is approximated by linear combinations of the components of the realization of Z.

In this paper we will mostly take G as a countable group. The set of possible reconstructions will be characterized by a central value, the reconstruction, and a measure of dispersion, the entropy. The choise of central value is made according to the principles of decision theory as given in Wald [10] or Lehman [7] chapter 1. The central value which maximizes the probability of correct reconstruction, i.e. the mode, will be dealt with in most cases but when T is discrete the central value which maximizes the expected number of correctly reconstructed components of X will be considered too.

The paper is arranged in the following way. In section 2 the problem of reconstruction is given a formulation according to decision theory. In section 3 the entropy is defined and some properties of entropy are reported. In section 4 the problem of reconstruction is considered for Markov chains. A recursive method to get the mode is given and the possibilities to use the theory of optimal trajectories (Bellman and Dreyfus [2]) in this connection is shown. The entropy is given. When G is commutative and the Markov chains have independent increments the mode and the entropy are given in more explicite forms. A number of examples are given too. In section 5 is discussed addition of independent jump processes which are generated by renewal processes and independent jumps in a commutative countable group G.

The problem of reconstruction is considered for exponentially and geometrically distributed waiting times of renewal points. In the first case the entropy is calculated in general and in the second case under the assumption of identically distributed jumps. At last section 6 is devoted to Gaussian processes and to comparisons between the mode and the reconstruction of Wiener.

2. Formulation of the problem according to decision theory

Denote by $x^*(z)$ the reconstruction of the X-realization when the Z-realization is z. Introduce a real valued loss function $L(x, x^*(z))$ which measures the "loss" when the X-realization is x and the reconstruction is $x^*(z)$. If needed conditions of measurability and integrability are assumed the risk is $R(x, x^*) = E(L(x, x^*(Z)) \mid X = x)$ and the Bayes' risk is $B(x^*) = EL(X, x^*(Z))$. The Bayes' risk is minimized for x^* satisfying $E(L(X, x^*(z)) \mid Z = z) = \min_{u \in G^T} E(L(X, u) \mid Z = z)$, i.e. $x^*(z)$ is an element of G^T which minimizes the expected loss according to the a posteriori distribution of X. If $x^*(z)$ exists it is not necessarily unique and in that case we will consider $x^*(z)$ as a many valued function. This is completely analogous to point estimation of a parameter with given a priori distribution.

With discrete distribution of $(X \mid Z = z)$

and
$$L(x, u) = 1 - \delta_{x,u} = \begin{cases} 1 & \text{if} \quad x \neq u \\ 0 & \text{if} \quad x = u \end{cases}$$

the reconstruction will be the mode, $x^*(z) = \text{mode}(X \mid Z = z)$, which we may consider as a many valued function too.

If G is countable, the index set $T = \{1, 2, \dots, n\}$ and $L(x, u) = \sum_{i=1}^{n} (1 - \delta_{x_i, u_i})$ for $x = (x_1, \dots, x_n)$ and $u = (u_1, \dots, u_n)$ the reconstruction will be $x^*(z) = (\text{mode}(X_1 \mid Z = z), \dots, \text{mode}(X_n \mid Z = z))$. This is the reconstruction which maximizes the expected number of correctly reconstructed components of the X-realization.

If G^T is a metrical space with distance d and $L(x, u) = d(x, u)^k$ for some positive natural number k the reconstruction will be the generalised Fréchet-meanvalue of the distribution of $(X \mid Z = z)$. Fréchet [5] defines the meanvalue m of a stochastic element X in a metrical space D with distance d by $\min_{u \in D} E d(X, u)^2 = E d(X, m)^2$ and generalised meanvalues are obtained by changing the exponent 2 to some positive natural number k. For questions of existence and uniqueness we refer to the mentioned work by Fréchet. If G^T is an n-dimensional Euclidean space with Euclidean distance the Fréchet-meanvalue coincides with ordinary expectation; i.e. $x^*(z) = E(X \mid Z = z) = (E(X_1 \mid Z = z), \dots, E(X_n \mid Z = z))$.

Instead of using a single value in G^T as the reconstruction we may give some subset K of G^T as a region of reconstruction. If we suppose that the Haar measure exists and if we impose the condition that the Haar measure of the region of reconstruction has a fixed value $\varepsilon > 0$ and if we use the loss function

$$L(x, K) = \begin{cases} 1 & \text{if} \quad x \notin K \\ 0 & \text{if} \quad x \in K \end{cases}$$

we will get the region of reconstruction $x^*(z)$ as the subset K of G^T which has Haar measure $\mu(K)=\varepsilon$ and maximizes $P(X\in K|Z=z)=\int_K \pi(x|z)\mu(dx)$. Here we have denoted by $\pi(x|z)$ the Radon-Nikodym derivative with regard to Haar measure of $P(X\in K|Z=z)$ considered as a measure of K. This is possible if and only if the distributions of X and Y are absolutely continuous with regard to Haar measure. According to the Neyman–Pearson fundamental lemma the region of reconstruction will be $x^*(z)=\{x:\pi(x|z)>c\}$ where the constant c depends on ε which ought to be carefully chosen to avoid unessential complications. For details we refer to Lehman [7] p. 63. In more suggestive words we may say that the region of reconstruction consists of the elements of G^T which are most probable according to the *a posteriori* distribution of X.

3. The entropy

If X is a discrete stochastic variable taking the value x_i with probability $p_i>0$ $(i=1,2,...;\ \sum_{i=1}^{\infty} p_i=1)$ the entropy of X is defined as $H(X)=-\sum_{i=1}^{\infty} p_i \log p_i$. Then $0\leqslant H(X)\leqslant +\infty$ and $H(X)=0$ if and only if X has a one point distribution. If X takes on only finitely many values, say n, the maximum entropy is $\log n$ and corresponds to the uniform distribution. If X and Y are discrete stochastic variables $H(X, Y)=H(X)+\sum_{i=1}^{\infty} p_i H(Y|X=x_i)=H(X)+EH(Y|X)$ and more generally with three discrete stochastic variables X, Y, Z $EH(X, Y|Z)=EH(X|Z)+EH(Y|X, Z)$. (Pinsker [8] p. 36). Another relation is $0\leqslant EH(Y|X)\leqslant H(Y)$ where there is equality to the left if and only if Y is a univalent function of X and equality to the right if X and Y are stochastically independent. If there is equality to the right and if $H(Y)<+\infty$ then X and Y are stochastically independent.

Let $Y_1, Y_2, \ldots$ be discrete stochastic variables and define a stochastic variable Y as the mixing of $Y_1, Y_2, \ldots$ with weights $p_1, p_2, \ldots$; i.e. $Y=Y_i$ with probability $p_i>0$; $i=1, 2, \ldots$.

Then if we suppose all entropies finite $\sum_i p_i H(Y_i)\leqslant H(Y)\leqslant -\sum_i p_i \log p_i+\sum_i p_i\cdot H(Y_i)$ with equality to the left if and only if $Y_1, Y_2, \ldots$ are identically distributed and equality to the right if and only if the ranges of values of different Y_i are disjoint. By introducing a stochastic variable X independent of $Y_1, Y_2, \ldots$ and taking the value x_i with probability p_i for $i=1, 2, \ldots$ and by noticing that $Y_i=(Y|X=x_i)$ the given inequalities may be written $EH(Y|X)\leqslant H(Y)\leqslant H(X)+EH(Y|X)=H(X, Y)=H(Y)+EH(X|Y)$. Here if we suppose all entropies finite it is equality to the left if and only if X and Y are independent; i.e. if and only if $Y_1, Y_2, \ldots$ are identically distributed. It is equality to the right if and only if X is a univalent function of Y; i.e. if and only if the ranges of values of different Y_i are disjoint.

If X and Y are stochastic elements in a discrete group G it is known (Grenander [6]) that $X+Y$ has a distribution which is more flat than the distributions of X or Y. A simple way to state this which does not seem to have been stressed in the literature is

$$\max\{\min_i p_i,\ \min_i q_i\}\leqslant \min_i r_i\leqslant \max_i r_i\leqslant \min\{\max_i p_i\ \max_i q_i\}.$$

Here p_i, q_i, r_i are the point masses of $X, Y, X+Y$ and the minima are supposed to exist. Then if G is infinite $\min_i p_i=\min_i q_i=0$ and the left bound is trivial. If G is finite and X or Y has probability mass over all G the left bound is non-trivial. From

$$r_j = \sum_i p_{j-i} q_i \leqslant (\max_i p_i) \sum_i q_i = \max_i p_i$$

and
$$r_j = \sum_i p_i q_{-i+j} \leqslant (\max_i q_i) \sum_i p_i = \max_i q_i$$

is obtained
$$r_j \leqslant \min \{\max_i p_i \max_i q_i\}$$

and as this is valid for all $j \in G$ the right-hand inequality is proved. The left-hand inequality may be derived in an analogous manner.

That the distribution of $X + Y$ is more flat than the distributions of X or Y may also be expressed in terms of entropy by the relation $H(X+Y) \geqslant \max\{H(X), H(Y)\}$. This relation follows from $H(X+Y) \geqslant EH(X+Y \mid Y) = H(X)$ and $H(X+Y) \geqslant EH(X+Y \mid X) = H(Y)$.

4. Addition of independent Markov chains

Let $(X_n)_{n=1}^{\infty}$ and $(Y_n)_{n=1}^{\infty}$ be independent Markov chains with a countable group G as the state space. Introduce the notations

$$P(X_1 = i) = p_i \quad i \in G$$

$$P(X_{n+1} = j \mid X_n = i) = P_{ij}(n) \quad i \in G, \quad j \in G, \quad n = 1, 2, \ldots$$

$$P(Y_1 = i) = q_i \quad i \in G$$

$$P(Y_{n+1} = j \mid Y_n = i) = Q_{ij}(n) \quad i \in G, \quad j \in G, \quad n = 1, 2, \ldots$$

Let $(z_n)_{n=1}^{\infty}$ be a realization of $(Z_n = X_n + Y_n)_{n=1}^{\infty}$ and define on G the sequence of functions $(M_n)_{n=1}^{\infty}$ by the recursive formula

$$\begin{cases} M_1(i) = p_i q_{-i+z_1} \quad i \in G \\ M_{n+1}(j) = \sup_{i \in G} M_n(i) P_{ij}(n) Q_{-i+z_n, -j+z_{n+1}}(n) \\ j \in G \quad n = 1, 2, \ldots \end{cases}$$

Theorem 1. *The supremum in the recursive formula above is a maximum which occurs for say* $i = \xi_n(j)$, $n = 1, \ldots, N-1$. *Moreover,* $\max_{j \in G} M_N(j)$ *exists and occurs for* $j = x_N^*$ *say. Then* $\operatorname{mode}(X_1, \ldots, X_N \mid Z_1 = z_1, \ldots, Z_N = z_N) = (x_1^*, \ldots, x_N^*)$ *with* $x_n^* = \xi_n(x_{n+1}^*)$ *for* $n = 1, \ldots, N-1$.

Proof: $M_n(i)$ is bounded by 0 and 1 and 0 is the only limit point when i varies in G. Thus $\max_{i \in G} M_n(i)$ exists. That 0 is the only limit point of $M_1(i)$, $i \in G$, is immediately clear. Suppose it holds for n. Then take a finite $C \subset G$ such that $M_n(i) < \varepsilon$ for $i \notin C$. To every $i \in C$ choose a finite $C_i \subset G$ such that $P_{ij}(n) < \varepsilon$ for $j \notin C_i$. Then $M_{n+1}(j) < \varepsilon$ for $j \notin \bigcup_{i \in C} C_i$, i.e. 0 is the only limit point of $M_{n+1}(j)$, $j \in G$. From this it also follows that the supremum in the recursive formula is a maximum.

$$\max_{x_1,\ldots,x_N} P(X_n=x_n,\ Y_n=-x_n+z_n;\ n=1,\ldots,N)=\max_{x_N\in G} M_N(x_N)=M_N(x_N^*)$$

$$=\max_{x_{N-1}\in G} M_{N-1}(x_{N-1})\,P_{x_{N-1},x_N^*}(N-1)\,Q_{-x_{N-1}+z_{N-1},-x_N^*+z_N}(N-1),$$

where the maximum occurs for $x_{N-1}=x_{N-1}^*=\xi_{N-1}(x_N^*)$. By iteration $x_n^*=\xi_n(x_{n+1}^*)$ for $n=1, \ldots, N-1$ and the proof is complete.

By theorem 1 the problem to determine a maximum over G^N is reduced to a series of determinations of maxima over G. When a digital computer is used for the calculations this method is very advantageous. A systematic account of recursive methods for determinations of maxima is given in Bellman [1] and Bellman and Dreyfus [2]. In the last book it is also shown how to arrange the calculations for the computer.

In principle it is possible to determine the maximum with the graphical method which is known as the optimal trajectory technique and on which there exists an extensive literature. In a few words this technique as applied to the present problem may be described as follows. The given realization of $(Z_n)_{n=1}^N$ is plotted in a (n, z_n)-plane together with all the possible realizations of $(X_n)_{n=1}^N$. These possible realizations are associated numbers. The segment from (n, i) to $(n+1, j)$ is given the number $P_{ij}(n)Q_{-i+z_n,-j+z_{n+1}}(n)$. The point $(n+1, j)$ is given the number $M_{n+1}(j)$ and an arrow is put on the segment (or segments) from $(n, \xi_n(j))$ to $(n+1, j)$. When x_N^* is determined the optimal trajectory (or trajectories) is obtained by following the arrows backwards from (N, x_N^*). A detailed accounts of the method is given in chapter VI of Bellman and Dreyfus [2]. The method is demonstrated in example 1 below.

Theorem 2. *Provided all entropies are finite*

$$EH(X_1, \ldots, X_N|Z_1, \ldots, Z_N)$$
$$=H(X_1)+\sum_{n=1}^{N-1} EH(X_{n+1}|X_n)$$
$$+H(Y_1)+\sum_{n=1}^{N-1} EH(Y_{n+1}|Y_n)-H(Z_1,\ldots,Z_N)$$

Proof: $EH(X_1, \ldots, X_N|Z_1, \ldots, Z_N)$

$$=H(X_1, \ldots, X_N, Z_1 \ldots, Z_N)-H(Z_1, \ldots, Z_N)$$
$$=H(X_1, \ldots, X_N, Y_1, \ldots, Y_N)-H(Z_1, \ldots, Z_N)$$
$$=H(X_1, \ldots, X_N)+H(Y_1, \ldots, Y_N)-H(Z_1, \ldots, Z_N)$$

and $\quad H(X_1, \ldots, X_N)$

$$=H(X_1, \ldots, X_{N-1})+EH(X_N|X_1, \ldots, X_{N-1})$$
$$=H(X_1, \ldots, X_{N-1})+EH(X_N|X_{N-1})$$

and an analogous expression for $H(Y_1, \ldots, Y_N)$ will give the desired formula by iteration.

The sum of two Markov chains is generally not Markovian as is shown in the following example. (Another example is given in example 1 below.) Take $G=\{0, 1\}$ with addition modulo 2 and let $(X_n)_{n=1}^{\infty}$ and $(Y_n)_{n=1}^{\infty}$ be stationary Markov chains with matrixes of transition probabilities

$$\begin{pmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{pmatrix} \text{ and } \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \text{ respectively.}$$

Then for instance

$$P(Z_{n+2}=0 \mid Z_{n+1}=0, Z_n=i) = \begin{cases} 61/108 & \text{if } i=0 \\ 47/84 & \text{if } i=1 \end{cases}$$

and consequently $(Z_n)_{n=1}^{\infty}$ is not Markovian.

A special case giving Markovian sum process is mentioned in the following theorem the proof of which is obvious.

Theorem 3. *If G is commutative under addition and if $(X_n)_{n=1}^{\infty}$ and $(Y_n)_{n=1}^{\infty}$ are processes with independent increments, i.e. $X_n=\xi_1+\ldots+\xi_n$ and $Y_n=\eta_1+\ldots+\eta_n$, where $\xi_1, \xi_2, \ldots, \eta_1, \eta_2, \ldots$ are independent stochastic elements of G, then $(Z_n)_{n=1}^{\infty}$ will be a process of independent increments and it is possible to write $Z_n=\zeta_1+\ldots+\zeta_n$ with $\zeta_n=\xi_n+\eta_n$ for $n=1, 2, \ldots$.*

Thus when G is commutative a sufficient condition for the sum process to be Markovian is that the added Markov processes have independent increments. That this condition is not necessary follows from the following example. Take G finite. Let $(X_n)_{n=1}^{\infty}$ be a Markov process but not with independent increments and let $(Y_n)_{n=1}^{\infty}$ be a degenerate Markov process consisting of independent components which are uniformly distributed over G. Then also $(Z_n)_{n=1}^{\infty}$ becomes a degenerate Markov process with independent components which are uniformly distributed over G.

Theorem 4. *With the assumptions and notations of Theorem 3*

$$\text{mode } (X_1, \ldots, X_N \mid Z_1=z_1, \ldots, Z_N=z_N) = (x_1^*, \ldots, x_N^*)$$

with
$$x_n^* = \sum_{v=1}^{n} \text{mode } (\xi_v \mid \zeta_v = z_v - z_{v-1}) \; n=1, \ldots, N;$$

$z_0=0$ (zero-element) and provided all entropies are finite

$$EH(X_1, \ldots, X_N \mid Z_1, \ldots, Z_N) = \sum_{n=1}^{N} EH(\xi_n \mid \zeta_n) = \sum_{n=1}^{N} [H(\xi_n) + H(\eta_n) - H(\zeta_n)].$$

Morover, for $n=1, \ldots, N$ the distribution of $(X_n \mid Z_1=z_1, \ldots, Z_N=z_N)$ is the convolution of the distributions of $(\xi_v \mid \zeta_v = z_v - z_{v-1})$ for $v=1, \ldots, n$.

Proof:
$$P(X_n = x_n;\, n = 1, \ldots, N \mid Z_n = z_n;\, n = 1, \ldots, N)$$
$$= \frac{P(\xi_n = x_n - x_{n-1},\, \eta_n = z_n - x_n - z_{n-1} + x_{n-1};\, n = 1, \ldots, N)}{P(\zeta_n = z_n - z_{n-1};\, n = 1, \ldots, N)}$$
$$= \prod_{n=1}^{N} P(\xi_n = x_n - x_{n-1} \mid \zeta_n = z_n - z_{n-1}),$$

where $x_0 = z_0 = 0$ and the maximum occurs for $x_n - x_{n-1} = x_n^* - x_{n-1}^* = \text{mode}(\xi_n \mid \zeta_n = z_n - z_{n-1})$ which proves the first part of the theorem. From Theorem 2, the Markovian structure of $(Z_n)_{n=1}^{\infty}$ and the relation $EH(X_{n+1} \mid X_n) = EH(X_n + \xi_{n+1} \mid X_n) = H(\xi_{n+1})$ and analogous expressions for $EH(Y_{n+1} \mid Y_n)$ and $EH(Z_{n+1} \mid Z_n)$ it follows that

$$EH(X_1, \ldots, X_N \mid Z_1, \ldots, Z_N)$$
$$= \sum_{n=1}^{N} [H(\xi_n) + H(\eta_n) - H(\zeta_n)] = \sum_{n=1}^{N} [H(\xi_n, \eta_n) - H(\zeta_n)]$$
$$= \sum_{n=1}^{N} [H(\xi_n, \zeta_n) - H(\zeta_n)] = \sum_{n=1}^{N} EH(\xi_n \mid \zeta_n).$$

The last part of the theorem is a consequence of the relation

$$(X_n \mid Z_1 = z_1, \ldots, Z_N = z_N)$$
$$= \sum_{v=1}^{n} (\xi_v \mid \zeta_1 = z_1,\, \zeta_2 = z_2 - z_1, \ldots, \zeta_N = z_N - z_{N-1})$$
$$= \sum_{v=1}^{n} (\xi_v \mid \zeta_v = z_v - z_{v-1}).$$

Example 1. Let G be the integers with addition and put

$$P(X_{n+1} = i + 1 \mid X_n = i) = p_1;\quad i = 0, \pm 1, \ldots;\quad n = 1, 2, \ldots$$
$$P(X_{n+1} = i - 1 \mid X_n = i) = 1 - p_1 = q_1;\quad i = 0, \pm 1, \ldots;\quad n = 1, 2, \ldots$$
$$P(X_1 = 1) = p_1,\ P(X_1 = -1) = q_1$$
$$P(Y_{n+1} = 1 \mid Y_n = i) = p_2;\quad i = \pm 1;\quad n = 0, 1, \ldots;\quad Y_0 = 0$$
$$P(Y_{n+1} = -1 \mid Y_n = i) = 1 - p_2 = q_2;\quad i = \pm 1;\quad n = 0, 1, \ldots;\quad Y_0 = 0.$$

Thus the process $(X_n)_{\infty}^{n=1}$ is a random walk process and the process $(Y_n)_{n=1}^{\infty}$ consists of identically distributed independent components. To a given realization $(z_n)_{n=1}^{N}$ of the sum process the possible realizations $(x_n)_{n=1}^{N}$ of $(X_n)_{n=1}^{N}$ are given by the conditions

$$x_1 = 1 \quad \text{if} \quad z_1 = 2$$
$$x_1 = -1 \quad \text{if} \quad z_1 = -2$$
$$x_n = z_n + 1 \quad \text{and} \quad x_{n+1} = z_{n+1} - 1 \quad \text{if} \quad z_{n+1} - z_n = 3;\quad n = 1, \ldots, N - 1$$
$$x_n = z_n - 1 \quad \text{and} \quad x_{n+1} = z_{n+1} + 1 \quad \text{if} \quad z_{n+1} - z_n = -3;\quad n = 1, \ldots, N - 1$$
$$|x_n - z_n| = 1;\quad n = 1, \ldots, N.$$
$$|x_{n+1} - x_n| = 1;\quad n = 1, \ldots, N - 1.$$

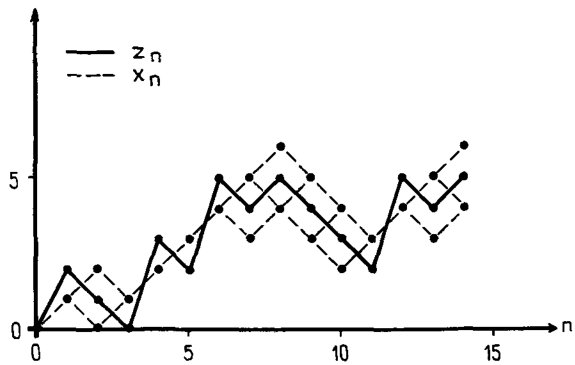

Fig. 1.

If the possible realizations are represented in a trajectory diagram of the type mentioned after Theorem 1 it is seen that the sections of the optimal trajectory which are not uniquely determined are situated above or below $(z_n)_{n=1}^N$ as $p_2 \leqslant q_2$ or $p_2 \geqslant q_2$. If x_N is not uniquely determined by $(z_n)_{n=1}^N$ one may compare the disjoint sections of the two optimal trajectories ending with $x_N = z_N + 1$ and $x_N = z_N - 1$ in order to decide which one is the optimal trajectory. In the case represented in Fig. 1 this comparison gives

$$1)\quad \text{when} \quad p_2 \leqslant q_2 \; x_N^* = \begin{cases} z_N + 1 & \text{if } p_1 q_2 \geqslant p_2 q_1 \text{ i.e. if } q_1 \leqslant q_2 \\ z_N - 1 & \text{if } p_1 q_2 \leqslant p_2 q_1 \text{ i.e. if } q_1 \geqslant q_2 \end{cases}$$

$$2)\quad \text{when} \quad p_2 \geqslant q_2 \; x_N^* = \begin{cases} z_N + 1 & \text{if } p_1 q_2 p_1 q_2 \geqslant p_2 q_1 p_1 p_2 \text{ i.e. if } \dfrac{p_1}{q_1} \geqslant \left(\dfrac{p_2}{q_2}\right)^2 \\ z_N - 1 & \text{if } p_1 q_2 p_1 q_2 \leqslant p_2 q_1 p_1 p_2 \text{ i.e. if } \dfrac{p_1}{q_1} \leqslant \left(\dfrac{p_2}{q_2}\right)^2. \end{cases}$$

Example 2. Take $G = \{0, 1, 2\}$ with addition modulo 3 and take $(X_n)_{n=1}^\infty$ and $(Y_n)_{n=1}^\infty$ as identically distributed independent stationary Markov chains with the common matrix of transition probabilities

$$\begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

As the matrix is cyclical, i.e. of the form

$$\begin{pmatrix} p_0 & p_1 & p_2 \\ p_2 & p_0 & p_1 \\ p_1 & p_2 & p_0 \end{pmatrix}$$

the processes have independent increments and according to Theorem 3 the sum process $(Z_n = X_n + Y_n)_{n=1}^\infty$ is Markovian with the easily found matrix of transition probabilities

$$\begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

The entropy calculations are straightforward according to Theorem 4

$$EH(X_1, \dots, X_N \mid Z_1, \dots, Z_N) = 2\,[\log 3 + (N-1) \log 2]$$
$$- \log 3 - \frac{N-1}{2} \log \tfrac{1}{2} - 2\,\frac{N-2}{4} \log \tfrac{1}{4} = \log 3 + \frac{N-1}{2} \log 2.$$

In this simple example the logarithm of the number of possible realizations of

$$(X_1, \dots, X_N \mid Z_1 = z_1, \dots, Z_N = z_N) \text{ is } \log 3 + \sum_{n=1}^{N-1} \delta_{z_n, z_{n+1}} \log 2$$

which may be realized from the following facts:

1) There are three alternatives of x_1
2) There is one or two alternatives of x_{n+1} for every choice of x_n depending on whether $z_{n+1} \neq z_n$ or $z_{n+1} = z_n$; $n = 1, \dots, N-1$.

This is a consequence of the given probability structure according to which repeated components are not allowed. This implies $x_n \neq x_{n+1}$ and $z_n - x_n \neq z_{n+1} - x_{n+1}$ for $n = 1, \dots, N-1$.

The expected value of the logarithm of the number of possible alternatives is then easily seen to be $\log 3 + (N-1)/2 \log 2$, i.e. equal to the entropy.

Example 3. This example is a discussion of the addition of two independent random walk processes. Let G be the integers with addition and $X_n = \xi_1 + \dots + \xi_n$ with $P(\xi_n = 1) = p_1$, $P(\xi_n = -1) = 1 - p_1 = q_1$ and $Y_n = \eta_1 + \dots + \eta_n$ with $P(\eta_n = 1) = p_2$, $P(\eta_n = -1) = 1 - p_2 = q_2$. The variables $\xi_1, \xi_2, \dots, \eta_1, \eta_2, \dots$ are independent. Put $\zeta_n = \xi_n + \eta_n$. Then

$$\zeta_n = \begin{cases} 2 & \text{with probability } p_1 p_2 \\ 0 & \text{with probability } p_1 q_2 + p_2 q_1 \\ -2 & \text{with probability } q_1 q_2 \end{cases}$$

and $(\xi_n \mid \zeta_n = 2) = 1$, $(\xi_n \mid \zeta_n = -2) = -1$

and $$(\xi_n \mid \zeta_n = 0) = \begin{cases} 1 & \text{with probability } p = \dfrac{p_1 q_2}{p_1 q_2 + p_2 q_1} \\ -1 & \text{with probability } 1 - p = q \end{cases}$$

i.e. $(\xi_n \mid \zeta_n = 0) = -1 + 2$ binomial $(1, p)$. Put for abbreviation

$$\sum_{v=1}^{n} \delta_{z_v - z_{v-1}, i} = a_{in};\ i = -2, 0, 2;\ n = 1, \dots, N;\ z_0 = 0.$$

Now it follows that there is $2^{a_{0N}}$ possible realizations of $(X_1, ..., X_N | Z_1 = z_1, ..., Z_N = z_N)$. If $p_1 = p_2$ they all have the same conditional probability. If $p_1 \neq p_2$ there is a unique mode and according to Theorem 4

$$\text{mode}(X_1, \ldots, X_N \mid Z_1 = z_1, \ldots, Z_N = z_N) = (x_1^*, \ldots, x_N^*)$$

with
$$x_n^* = \begin{cases} a_{2n} - a_{-2n} + a_{0n} & \text{if } p_1 > p_2 \\ a_{2n} - a_{-2n} - a_{0n} & \text{if } p_1 < p_2. \end{cases}$$

Thus
$$P(X_n = x_n^*;\ n = 1, \ldots, N \mid Z_n = z_n;\ n = 1, \ldots, N) = \begin{cases} p^{a_{0N}} & \text{if } p_1 \geqslant p_2 \\ q^{a_{0N}} & \text{if } p_1 \leqslant p_2. \end{cases}$$

From Theorem 4 it also follows that

$$(X_n \mid Z_1 = z_1, \ldots, Z_N = z_N) = \sum_{v=1}^{n} (\xi_v \mid \zeta_v = z_v - z_{v-1})$$
$$= a_{2n} - a_{-2n} - a_{0n} + 2 \text{ binomial } (a_{0n}, p).$$

This implies $x_n^{**} = \text{mode}(X_n | Z_1 = z_1, ..., Z_N = z_N) = a_{2n} - a_{-2n} - a_{0n} + 2[(1 + a_{0n})p]$. If the number inside the whole number part brackets is a whole number the mode is two-valued. For details we refer to Feller [4] chapter VI Theorem 1. The reconstruction $(x_1^*, ..., x_N^*)$ maximizes the probability that the reconstruction is correct and the reconstruction $(x_1^{**}, ..., x_N^{**})$ maximizes the expected number of correctly reconstructed components. It is interesting to observe that when N increases the expected number of correctly reconstructed components is bounded for $(x_1^*, ..., x_N^*)$ but is increasing at least as fast as $\sqrt{N}$ for $(x_1^{**}, ..., x_N^{**})$. The boundedness may be proved as follows for $p_1 \geqslant p_2$ and analogously for $p_1 \leqslant p_2$. Put $A_n = \sum_{v=1}^{n} \delta_{Z_v - Z_{v-1}, 0} = \text{binomial}$ $(n, p_1 q_2 + p_2 q_1)$. Then

$$E \sum_{n=1}^{N} \delta_{X_n, x_n^*} = E \sum_{n=1}^{N} P(X_n = x_n^* \mid Z_1, \ldots, Z_N) = E \sum_{n=1}^{N} p^{A_n}$$
$$= \sum_{n=1}^{N} \sum_{k=0}^{n} p^k \binom{n}{k} (p_1 q_2 + p_2 q_1)^k (1 - p_1 q_2 - p_2 q_1)^{n-k}$$
$$= \sum_{n=1}^{N} (1 - p_2 q_1)^n = (1 - p_2 q_1) \frac{1 - (1 - p_2 q_1)^N}{p_2 q_1} \nearrow \frac{1 - p_2 q_1}{p_2 q_1}$$

when $N \to \infty$.

The statement
$$\varliminf_{N \to \infty} \frac{1}{\sqrt{N}} E \sum_{n=1}^{N} \delta_{X_n, x_n^{**}} \geqslant c,$$

where c is a positive constant independent of N may be proved as follows.

$$E \sum_{n=1}^{N} \delta_{X_n, x_n^{**}} = E \sum_{n=1}^{N} P(X_n = x_n^{**} \mid Z_1 \ldots, Z_N) = E \sum_{n=1}^{N} \max_{0 \leqslant v \leqslant A_n} \binom{A_n}{v} p^v q^{A_n - v}$$

where $A_n = \text{binomial}(n, p_1 q_2 + p_2 q_1)$. By application of the inequality involving point masses of convolutions given in section 3 is obtained

$$\max_{0 \leqslant v \leqslant A_n} \binom{A_n}{v} p^v q^{A_n - v} \geqslant \max_{0 \leqslant v \leqslant n} \binom{n}{v} p^v q^{n-v}.$$

Moreover, for arbitrary small $\varepsilon > 0$ and sufficiently large n, say $n \geqslant N_0$,

$$\max_{0 \leqslant v \leqslant n} \binom{n}{v} p^v q^{n-v} \geqslant (1-\varepsilon) \frac{1}{\sqrt{2\pi n p q}}$$

(see Feller [4] Chapter VII, Theorem 1).

Thus
$$E \sum_{n=1}^{N} \delta_{X_n, x_n^{**}} \geqslant \sum_{n=N_0}^{N} \frac{1-\varepsilon}{\sqrt{2\pi n p q}} \geqslant \frac{1-\varepsilon}{\sqrt{2\pi p q}} \int_{N_0}^{N} \frac{du}{\sqrt{u}}$$

$$= (1-\varepsilon) \sqrt{\frac{2}{\pi p q}} (\sqrt{N} - \sqrt{N_0}) \text{ for } N \geqslant N_0$$

and the statement follows. By some very tedious calculations it is possible to prove the stronger relation

$$\lim_{N \to \infty} \frac{1}{\sqrt{N}} E \sum_{n=1}^{N} \delta_{X_n, x_n^{**}} = \sqrt{\frac{2}{\pi} \left(\frac{1}{p_1 q_2} + \frac{1}{p_2 q_1} \right)}.$$

Example 4. Again let G be the integers with addition. With the notation of Theorem 3 let $\xi_n = \text{Poisson}(p\lambda_n)$ and $\eta_n = \text{Poisson}(q\lambda_n)$ where $0 < p < 1$ and $p + q = 1$. Then $\zeta_n = \text{Poisson}(\lambda_n)$ and $(\xi_n \mid \zeta_n = k) = \text{Binomial}(k; p)$ and $(X_n \mid Z_1 = z_1, \ldots, Z_N = z_N) = \sum_{v=1}^{n} (\xi_v \mid \zeta_v = z_v - z_{v-1}) = \text{Binomial}(z_n; p)$. Ignoring possible cases of non-uniqueness the reconstruction which maximizes the probability of correct reconstruction is

$$\text{mode}\,(X_1, \ldots, X_N \mid Z_1 = z_1, \ldots, Z_N = z_N) = (x_1^*, \ldots, x_N^*)$$

with
$$x_n^* = \sum_{v=1}^{n} [(z_v - z_{v-1} + 1)\, p]$$

and the reconstruction which maximizes the expected number of correctly reconstructed components is

$$(x_1^{**}, \ldots, x_N^{**}) \text{ with } x_n^{**} = \text{mode } (X_n \mid Z_1 = z_1, \ldots, Z_N = z_N) = [(z_n + 1)\, p].$$

5. Addition of independent jump processes

Introduce two independent stochastic processes $(U_n)_{n=1}^{\infty}$ and $(\xi_n)_{n=1}^{\infty}$. Here $U_n - U_{n-1}$ for $n = 1, 2, \ldots$ and $U_0 = 0$ are positive identically distributed independent stochastic variables, i.e. $(U_n)_{n=1}^{\infty}$ is a renewal process. Further $\xi_1, \xi_2, \ldots$ are independent stochastic elements in a discrete commutative group G. Put $P(\xi_n=i)=P_i(n)$ for $i\in G$ and $n=1, 2, \ldots$ and suppose that the zero element in G has probability zero for every n; i.e. $P_0(n)\equiv 0$. Define the jump process $\{X_t:t\geqslant 0\}$ by

$$X_t=\sum_{v\leqslant A_t}\xi_v \quad \text{where} \quad A_t=\max_{U_n\leqslant t} n \quad \text{and} \quad \{A_t:t\geqslant 0\}$$

is the counting process associated with the renewal process $(U_n)_{n=1}^{\infty}$ of jump times.

Introduce another jump process $\{Y_t:t\geqslant 0\}$ which is independent of $\{X_t:t\geqslant 0\}$ and is defined by $Y_t=\sum_{v\leqslant B_t}\eta_v$ with renewal process $(V_n)_{n=1}^{\infty}$, counting process $\{B_t:t\geqslant 0\}$ and jump value process $(\eta_n)_{n=1}^{\infty}$. Put $P(\eta_n=i)=Q_i(n)$, $i\in G$; $n=1, 2, \ldots$ with $Q_0(n)\equiv 0$.

For the sum process $\{Z_t=X_t+Y_t:t\geqslant 0\}$ we denote the jump time process by $(W_n)_{n=1}^{\infty}$, the counting process by $\{C_t:t\geqslant 0\}$ and the jump value process by $(\zeta_n)_{n=1}^{\infty}$. With a realization of $\{Z_t:0\leqslant t\leqslant T\}$ available it is desired to reconstruct the corresponding realization of $\{X_t:0\leqslant t\leqslant T\}$. If the renewal processes $(U_n)_{n=1}^{\infty}$ and $(V_n)_{n=1}^{\infty}$ are deterministic with the same period (for instance $U_n=V_n=n$) the problem is in principle the same as that discussed in the Theorems 3 and 4. Another special case is obtained with G as the integers with addition and $\xi_n=\eta_n=1$. Then $X_t=A_t$ and $Y_t=B_t$. The addition of counting processes is dealt with in Cox and Smith [3] where the probability structure of the sum process is discussed when there is a great number of added processes. In general the jump times of the sum process are not constituting a renewal process but in the cases discussed below it does happen.

Case 1. If $P(U_n-U_{n-1}\leqslant t)=1-e^{-at}$ and $P(V_n-V_{n-1}\leqslant t)=1-e^{-bt}$ for $n=1, 2, \ldots$ then $(W_n)_{n=1}^{\infty}$ is a renewal process with $P(W_n-W_{n-1}\leqslant t)=1-e^{-(a+b)t}$; $n=1, 2, \ldots$; $W_0=0$ and $(\zeta_n)_{n=1}^{\infty}$ is a process with components given by

$$\zeta_n=\begin{cases}\xi_k \quad \text{with probability} \quad \binom{n-1}{k-1}\left(\dfrac{a}{a+b}\right)^k\left(\dfrac{b}{a+b}\right)^{n-k}; \quad k=1, \ldots, n\\ \eta_k \quad \text{with probability} \quad \binom{n-1}{k-1}\left(\dfrac{b}{a+b}\right)^k\left(\dfrac{a}{a+b}\right)^{n-k}; \quad k=1, \ldots, n.\end{cases}$$

The first part of the statement is well known and the second part depends on that $\{X_t:t\geqslant 0\}$ and $\{Y_t:t\geqslant 0\}$ with probability 1 have no coincident time points of jumps.

Define a sequence of functions $(M_n)_{n=1}^{\infty}$ by the recursive formula

$$\begin{cases}M_n(0)=b^n\prod_{k=1}^{n}Q_{\zeta_k}(k); \quad n=1, 2, \ldots\\ M_n(k)=\max\{M_{n-1}(k)\,b\,Q_{\zeta_n}(n-k);\ M_{n-1}(k-1)\,a\,P_{\zeta_n}(k)\};\\ \qquad\qquad\qquad\qquad\qquad\qquad k=1, \ldots, n-1;\ n=2, 3, \ldots\\ M_n(n)=a^n\prod_{k=1}^{n}P_{\zeta_k}(k); \quad n=1, 2, \ldots\end{cases}$$

and put $I_n(k)$ equal to 0 or 1 as in the recursive formula the first or the second quantity in brackets is the greatest. If the two quantities are equal we take $I_n(k)$

two-valued. Put $I_n(0)=0$ and $I_n(n)=1$ for $n=1, 2, \dots$. Suppose that $\{Z_t : 0\leqslant t\leqslant T\}$ has a realization $\{z_t : 0\leqslant t\leqslant T\}$ with the jumps $\zeta_1, \dots, \zeta_N$ at time points $w_1, \dots, w_N$ respectively.

Theorem 5. mode$(X_t : 0\leqslant t\leqslant T\,|\,Z_t=z_t : 0\leqslant t\leqslant T)$ *has* k^* *jump time points and* w_n *is such a point with jump* ζ_n *if* $I_n(k_n)=1$; $n=1, \dots, N$. *Here* k^* *is (not necessarily uniquely) determined by* $\max_{0\leqslant k\leqslant N} M_N(k)=M_N(k^*)$ *and* k_n *is given recursively by* $k_n=k_{n+1}-I_{n+1}(k_{n+1})$; $n=1, \dots, N-1$ *with* $k_N=k^*$.

Proof: Among the 2^N possible realizations of $(X_t : 0\leqslant t\leqslant T\,|\,Z_t=z_t : 0\leqslant t\leqslant T)$ the mode is determined by maximizing

$$a^k \prod_{v=1}^{k} P_{\zeta_{i_v}}(v)\, b^{N-k} \prod_{v=1}^{N-k} Q_{\zeta_{j_v}}(v),$$

on one hand, in regard to k, the number of jumps, and, on the other hand, in regard to the choice of k jump points among the N possible points. Here $(i_1, \dots, i_k, j_1, \dots, j_{N-k})$ is a permutation of $(1, \dots, N)$ satisfying $1\leqslant i_1<i_2<\dots<i_k\leqslant N$ and $1\leqslant j_1<j_2<\dots<j_{N-k}\leqslant N$. By fixing k and maximizing in regard to the permutations the maximal value becomes $M_N(k)$ and then $M_N(k)$ is maximized for $k=k^*$. From the definition of I_n it follows that $I_n(A_{w_n})=A_{w_n}-A_{w_{n-1}}$; $n=1, \dots, N$; $w_0=0$. Thus with $A_{w_n}=k_n$

$$I_n(k_n)=k_n-k_{n-1};\quad n=1, \dots, N;\quad k_0=0$$

and the mode has a jump at w_n if $I_n(k_n)=1$.

With the technique of trajectories the mode is determined in the following way. In the (n, A_{w_n})-plane sections from $(n-1, k)$ to (n, k) are given the numbers $b\cdot Q_{\zeta_n}(n-k)$ and sections from $(n-1, k-1)$ to (n, k) are given the numbers $a\cdot P_{\zeta_n}(k)$. The points (n, k) are given the numbers $M_n(k)$ and arrows are given to sections from $(n-1,\ k-I_n(k))$ to (n, k). By following the arrows backwards from (n, k^*) the optimal trajectory (or trajectories) is obtained which corresponds to the mode.

If the jump value processes consist of identically distributed components, i.e. if $P_i(n)=P_i$ and $Q_i(n)=Q_i$, the determination of the mode is considerably simplified. Then the mode has a jump ζ_n at time w_n if $a\cdot P_{\zeta_n}>b\cdot Q_{\zeta_n}$ but has no jump if $a\cdot P_{\zeta_n}<b\cdot Q_{\zeta_n}$. If $a\cdot P_{\zeta_n}=b\cdot Q_{\zeta_n}$ the mode is not unique. We omit the details. Notice that this type of processes is identical with the homogeneous processes on G.

Theorem 6.

$$EH(X_t\!:0\leqslant t\leqslant T\,|\,Z_t\!:0\leqslant t\leqslant T)\leqslant -aT\ \log\frac{a}{a+b}-bT\ \log\frac{b}{a+b}$$

with equality if and only if $P_i(n)=Q_i(m)$ *for* $n=1, 2, \dots$; $m=1, 2, \dots$; $i\in G$.

Proof: Putting $X=(X_t : 0\leqslant t\leqslant T)$, $A=(A_t : 0\leqslant t\leqslant T)$, $\xi=(\xi_1, \dots, \xi_{A_T})$, $Z=(Z_t : 0\leqslant t\leqslant T)$, $C=(C_t : 0\leqslant t\leqslant T)$, $\zeta=(\zeta_1, \dots, \zeta_{C_T})$ and using the properties of entropy we obtain $EH(X\,|\,Z)=EH(A, \xi\,|\,C, \zeta)=EH(A\,|\,C, \zeta)+EH(\xi\,|\,A, C, \zeta)=EH(A\,|\,C, \zeta)\leqslant EH(A\,|\,C)$ with equality if and only if A and ζ are independent when conditioned by C, which means that the jumps of the sum process are independent of knowledge about the component processes generating them. Thus $\xi_1, \xi_2, \dots, \eta_1, \eta_2, \dots$ are all identically distributed; i.e. $P_i(n)=Q_i(m)$.

The direct method to evaluate $EH(A|C)$ is to observe that in the conditional distribution the jumps are independent and belongs to A with probability $a/(a+b)$

Thus
$$EH(A\,|\,C)=E\,C_T\left(-\frac{a}{a+b}\log\frac{a}{a+b}-\frac{b}{a+b}\log\frac{b}{a+b}\right)$$
$$=-aT\log\frac{a}{a+b}-bT\log\frac{b}{a+b}.$$

Another method would be to write $EH(A|C)=H(A)+H(B)-H(C)$ provided that we had defined entropy for such variables in a consistent way. It may be of some interest to notice that

$$H(A)=H(U_1,\dots,U_{A_T})=H(A_T)+EH(U_1,\dots,U_{A_T}\,|\,A_T)$$
$$=-E\log\left[e^{-aT}\frac{(aT)^{A_T}}{A_T!}\right]+E\log\frac{T^{A_T}}{A_T!}=aT\log\frac{e}{a}$$

is obtained if $\log(T^k/k!)$ is taken as the entropy of $(U_1,\dots,U_{A_T}|A_T=k)$ which is distributed as an ordered k-sample from the uniform distribution over the interval $(0, T)$. Compare Tacács [9] p. 39. Then

$$EH(A\,|\,C)=aT\log\frac{e}{a}+bT\log\frac{e}{b}-(a+b)T\log\frac{e}{a+b}$$
$$=-aT\log\frac{a}{a+b}-bT\log\frac{b}{a+b}.$$

Case 2. When the renewal processes $(U_n)_{n=1}^{\infty}$ and $(V_n)_{n=1}^{\infty}$ have discrete distributions they may have coincident renewal points with positive probability. This makes the problem of reconstruction more complicated and here we will only discuss the case when the two renewal processes have geometrical distributions with $P(U_n-U_{n-1}=k)=p_1q_1^{k-1}$; $p_1+q_1=1$; $k=1, 2, \dots$; $n=1, 2, \dots$,; $U_0=0$; $P(V_n-V_{n-1}=k)=p_2q_2^{k-1}$; $p_2+q_2=1$; $k=1, 2, \dots$; $n=1, 2, \dots$; $V_0=0$. Then also $(W_n)_{n=1}^{\infty}$ is such a process with $P(W_n-W_{n-1}=k)=pq^{k-1}$; $q=q_1q_2$; $p+q=1$; $k=1, 2, \dots$; $n=1, 2, \dots$; $W_0=0$.

Further

$$\zeta_n=\begin{cases}\xi_k & \text{with probability } \binom{n-1}{k-1}\left(\frac{p_1}{p}\right)^{k-1}\left(\frac{p_2q_1}{p}\right)^{n-k}\frac{p_1q_2}{p};\quad k=1,\dots,n\\[2ex] \eta_k & \text{with probability } \binom{n-1}{k-1}\left(\frac{p_2}{p}\right)^{k-1}\left(\frac{p_1q_2}{p}\right)^{n-k}\frac{p_2q_1}{p};\quad k=1,\dots,n\\[2ex] \xi_i+\eta_j & \text{with probability } \dfrac{(n-1)!}{(n-i)!\,(n-j)!\,(i+j-n-1)!}\left(\frac{p_1q_2}{p}\right)^{n-j}\\[2ex] & \quad\times\left(\frac{p_1p_2}{p}\right)^{i+j-n-1}\left(\frac{p_2q_1}{p}\right)^{n-i}\frac{p_1p_2}{p}\quad i=1,\dots,n\quad j=1,\dots,n\end{cases}$$

and we assume $P(\zeta_n = 0) = 0$ which is satisfied for instance if G is the integers with addition and ξ_n and η_n are positive variables.

Define the sequence of functions $(M_n)_{n=1}^{\infty}$ by the recursive formula

$$\begin{cases} M_n(0,n) = (p_2 q_1)^n \prod_{k=1}^{n} Q_{\zeta_k}(k); \quad n = 1, 2, \dots \\ M_n(n,0) = (p_1 q_2)^n \prod_{k=1}^{n} P_{\zeta_k}(k); \quad n = 1, 2, \dots \\ M_n(n,n) = (p_1 p_2)^n \prod_{k=1}^{n} \max_{i \in G} P_i(k) Q_{\zeta_k - i}(k); \quad n = 1, 2, \dots \\ M_n(a,b) = \max \begin{Bmatrix} M_{n-1}(a-1,b) p_1 q_2 P_{\zeta_n}(a) \\ M_{n-1}(a,b-1) p_2 q_1 Q_{\zeta_n}(b) \\ M_{n-1}(a-1,b-1) p_1 p_2 \max_{i \in G} P_i(a) Q_{\zeta_n - i}(b) \end{Bmatrix} \\ \qquad a = 1, \dots, n; \quad b = 1, \dots, n; \\ \qquad n \leqslant a + b < 2n; \quad n = 2, 3, \dots . \end{cases}$$

Put $I'_n(a,b)$ equal to 1 if the first or third quantity is the greatest among the three quantities in the maximum brackets above and put $I'_n(a,b)$ equal to 0 otherwise. Put $I''_n(a,b)$ equal to 1 if the second or third quantity is the greatest among the three quantities in the maximum brackets above and put $I''_n(a,b)$ equal to 0 otherwise.

Put
$$I'_n(0,n) = 0, \quad I'_n(n,0) = 1, \quad I'_n(n,n) = 1,$$
$$I''_n(0,n) = 1, \quad I''_n(n,0) = 0, \quad I''_n(n,n) = 1.$$

Suppose that $\{Z_t : 0 \leqslant t \leqslant T\}$ has a realization $\{z_t : 0 \leqslant t \leqslant T\}$ with the jumps $\zeta_1, \dots, \zeta_N$ at the time points $w_1, \dots, w_N$. Then the following theorem may be proved in much the same way as Theorem 5.

Theorem 7. mode$(X_t : 0 \leqslant t \leqslant T \mid Z_t = z_t : 0 \leqslant t \leqslant T)$ *has* a^* *jump time points and* w_n *is such a point if* $I'_n(a_n, b_n) = 1$ *and then the jump is* ζ_n *if* $I''_n(a_n, b_n) = 0$ *and the jump is equal to the element* $i \in G$ *which maximizes* $P_i(a_n) Q_{\zeta_n - i}(b_n)$ *if* $I''_n(a_n, b_n) = 1$. *Here* a^*, a_n *and* b_n *are given by* $\max_{a,b} M_N(a,b) = M_N(a^*, b^*)$ *and*

$$\begin{cases} a_n = a_{n+1} - I'_{n+1}(a_{n+1}, b_{n+1}) \\ b_n = b_{n+1} - I''_{n+1}(a_{n+1}, b_{n+1}) \\ n = 1, \dots, N-1; \quad a_N = a^*; \quad b_N = b^*. \end{cases}$$

If the jumps ξ_n are independent and identicallly distributed it is possible to write $X_n = \xi_1 + \dots + \xi_{A_n} = \xi'_{U_1} + \dots + \xi'_{U_{A_n}} = \xi'_1 + \dots + \xi'_n$ where ξ'_n are independent and identically distributed and

$$\xi_1' = \begin{cases} 0 & \text{with probability} \quad q_1 \\ \xi_1 & \text{with probability} \quad p_1 \end{cases}$$

i.e.
$$P(\xi_1' = i) = \begin{cases} q_1 + p_1 P_0 & \text{for} \quad i = 0 \\ p_1 P_i & \text{for} \quad i \neq 0. \end{cases}$$

In the same way $Y_n = \eta_1 + \ldots + \eta_{B_n} = \eta_1' + \ldots + \eta_n'$ where η_n' are independent and identically distributed and

$$\eta_1' = \begin{cases} 0 & \text{with probability} \quad q_2 \\ \eta_1 & \text{with probability} \quad p_2 \end{cases}$$

i.e.
$$P(\eta_1' = i) = \begin{cases} q_2 + p_2 Q_0 & \text{for} \quad i = 0 \\ p_2 Q_i & \text{for} \quad i \neq 0. \end{cases}$$

Then $Z_n = \zeta_1' + \ldots + \zeta_n'$ with $\zeta_n' = \xi_n' + \eta_n'$ and

$$\zeta_1' = \begin{cases} 0 & \text{with probability} \quad q_1 q_2 \\ \xi_1 & \text{with probability} \quad p_1 q_2 \\ \eta_1 & \text{with probability} \quad q_1 p_2 \\ \xi_1 + \eta_1 & \text{with probability} \quad p_1 p_2. \end{cases}$$

By application of Theorem 4

$$\text{mode} \quad (X_1, \ldots, X_N \mid Z_1 = z_1, \ldots, Z_N = z_N) = (x_1^*, \ldots, x_N^*)$$

with
$$x_n^* = \sum_{v=1}^{n} \text{mode } (\xi_v' \mid \zeta_v' = z_v - z_{v-1}); \quad n = 1, \ldots, N; \quad z_0 = 0.$$

Moreover, provided all entropies are finite

$$EH(X_1, \ldots, X_N \mid Z_1, \ldots, Z_N) = N\,[H(\xi_1') + H(\eta_1') - H(\zeta_1')]$$

and by use of the inequalities for entropies of mixed distributions it follows that this entropy is strictly smaller than

$$\begin{aligned} N[&-p_1 \log p_1 - q_1 \log q_1 + p_1 H(\xi_1) - p_2 \log p_2 \\ &\quad - q_2 \log q_2 + p_2 H(\eta_1) - p_1 q_2 H(\xi_1) - q_1 p_2 H(\eta_1) - p_1 p_2 H(\xi_1 + \eta_1)] \\ &= N[p_1 p_2 EH(\xi_1 \mid \xi_1 + \eta_1) - p_1 \log p_1 - q_1 \log q_1 - p_2 \log p_2 - q_2 \log q_2]. \end{aligned}$$

Now if we impose the restrictions $P_0 = Q_0 = \sum_i P_i Q_{-i} = 0$ we get for $j \neq 0$

$$\operatorname{mode}(\xi_v' \mid \zeta_v' = j) = \begin{cases} j \\ \operatorname{mode}(\xi_1 \mid \zeta_1 = j) \\ 0 \end{cases}$$

according as $p_1 q_2 P_j$ or $p_1 p_2 \max_i P_i Q_{j-i}$ or $p_2 q_1 Q_j$ is the greatest of the three quantities. If the greatest quantity is not uniquely determined the mode is not unique either. For the entropies we get

$$H(\xi_1') = -p_1 \log p_1 - q_1 \log q_1 + p_1 H(\xi_1)$$
$$H(\eta_1') = -p_2 \log p_2 - q_2 \log q_2 + p_2 H(\eta_1)$$
$$H(\zeta_1') = -p \log p - q \log q + pH(\zeta''),$$

$$\text{where } \zeta'' = \begin{cases} \xi_1 & \text{with probability } \dfrac{p_1 q_2}{p} \\ \eta_1 & \text{with probability } \dfrac{p_2 q_1}{p} \\ \xi_1 + \eta_1 & \text{with probability } \dfrac{p_1 p_2}{p} \end{cases}$$

and
$$H(\zeta'') \geqslant \frac{p_1 q_2}{p} H(\xi_1) + \frac{p_2 q_1}{p} H(\eta_1) + \frac{p_1 p_2}{p} H(\xi_1 + \eta_1)$$

with equality if and only if ξ_1, η_1 and $\xi_1 + \eta_1$ are identically distributed, i.e. $P_i = Q_i = \sum_j P_j P_{i-j}$ which can never hold together with $P_0 = Q_0 = 0$. It follows that

$$EH(X_1, \ldots, X_N \mid Z_1, \ldots, Z_N) < N \left[p_1 p_2 EH(\xi_1 \mid \xi_1 + \eta_1) - p_1 p_2 \log \frac{p_1 p_2}{p} - p_1 q_2 \log \frac{p_1 q_2}{p} - p_2 q_1 \log \frac{p_2 q_1}{p} \right].$$

6. Addition of independent Gaussian processes

Let $(X_n)_{n=-\infty}^{+\infty}$ and $(Y_n)_{n=-\infty}^{+\infty}$ be independent Gaussian processes with $EX_n = 0$, $EX_n \overline{X}_m = A_{nm}$, $EY_n = 0$, $EY_n \overline{Y}_m = B_{nm}$. A realization $(z_n)_{n=1}^{N}$ of the sum process $(Z_n = X_n + Y_n)_{n=1}^{N}$ is available and it is required to reconstruct the corresponding realization of $(X_n)_{n=1}^{N}$. If we denote by A and B the covariance matrixes (assumed nonsingular) of $X = (X_1, \ldots, X_N)$ and $Y = (Y_1, \ldots, Y_N)$ respectively we obtain for $(X \mid Z = z)$ the frequency function

$$k \cdot \exp - \tfrac{1}{2} [x' A^{-1} x + (z - x)' B^{-1} (z - x)],$$

where k is a constant independent of $x = (x_1, \ldots, x_N)$. The reconstruction mode $(X \mid Z = z)$ defined as the element which maximizes the frequency function is determined by maximizing $z'B^{-1}x + x'B^{-1}z - x'(A^{-1}+B^{-1})x$ in regard to x which after some calculations gives

$$\mathrm{mode}(X \mid Z=z) = (A^{-1}+B^{-1})^{-1}B^{-1}z = (I+BA^{-1})^{-1}z = A(A+B)^{-1}z.$$

The method of reconstruction of Wiener only works when the processes are supposed stationary and when the given realizations have infinitely many components. Then $x_n^* = \sum_{m=0}^{\infty} c_m z_{n-m}$ is taken as the reconstruction of x_n if $c_0, c_1, \dots$ can be determined so that $E\,|X_n - \sum_{m=0}^{\infty} c_m Z_{n-m}|^2$ is minimized. A natural formulation of this linear method in our case with finitly many components and without demand for stationarity is to reconstruct x_n with $x_n^* = \sum_{m=1}^{N} L_{nm} z_m$ where the L_{nm} are determined by minimizing $E\,|X_n - \sum_{m=1}^{N} L_{nm} Z_m|^2$. If d is the ordinary Euclidean distance in the N-dimensional Euclidean space we may say that $x = (x_1, \dots, x_N)$ is reconstructed by $x^* = Lz$ where the matrix L is determined by minimizing $Ed(X, LZ)^2$. From the relation

$$E\,|X_k - \sum_{m=1}^{N} L_{km} Z_m|^2 = A_{kk} + \sum_{n=1}^{N} \sum_{m=1}^{N} L_{kn} \overline{L_{km}} (A_{nm} + B_{nm}) - \sum_{m=1}^{N} L_{km} A_{mk} - \sum_{m=1}^{N} \overline{L_{km}} A_{km}$$

and straightforward calculations the optimal choise of L is seen to be $L = A(A+B)^{-1}$, i.e. the linear reconstruction Lz is equal to the mode.

ACKNOWLEDGEMENT

I am very grateful to my teacher Professor Ulf Grenander for his comments and encouragement. Also I am deeply indebted to my colleague Per Martin-Löf who has critically read the manuscript and made some very valuable comments.

University of Stockholm

REFERENCES

1. BELLMAN, R., Dynamic Programming. Princeton University Press 1957.
2. BELLMAN, R., and DREYFUS, S., Applied Dynamic Programming. Princeton University Press 1962.
3. COX, D. R., and SMITH, W. L., On the superposition of renewal processes. Biometrika 1954, *41*, 91–99.
4. FELLER, W., An Introduction to Probability Theory and its Applications. Wiley 1957.
5. FRÉCHET, M., Les éléments aléatoires de nature quelconque dans un espace distancié. Ann. Inst. H. Poincaré X, pp. 215–310.
6. GRENANDER, U., Probabilities on Algebraic Structures. Wiley 1963.
7. LEHMAN, E. L., Testing Statistical Hypotheses. Wiley 1959.
8. PINSKER, M. S., Information und Informationsstabilität zufälliger Grössen und Prozesse. Arbeiten zur Informationstheorie V. VEB Deutscher Verlag der Wissenschaften. Berlin 1963.
9. TACÁCS, L., Stochastic Processes. Methuen's Monographs. London 1960.
10. WALD, A., Statistical Decision Functions. Wiley 1950.
11. WIENER, N., The extrapolation, interpolation and smoothing of stationary time series with engineering applications. Technology Press 1949.

Tryckt den 10 november 1964

Uppsala 1964. Almqvist & Wiksells Boktryckeri AB