RT-MAT 2005 - 16

# SURE INFERENCE ANALYSIS

José Carlos Simon de Miranda

**Agosto 2005**

Esta é uma publicação preliminar ("preprint").

# SURE INFERENCE ANALYSIS

JOSÉ CARLOS SIMON DE MIRANDA

ABSTRACT In this work we define the new concepts of **standard deviation based inferential sequence, variance based inferential sequence, inferential sequence of random variables or random vectors** and **inferential sequence of real or vector valued stochastic processes** that enable us to estimate and draw conclusions about vectors and vector valued functions without knowing any other information about the probabilistic structure of these random variables and stochastic processes but their inferential sequences. We also define **m-th order at least probability $p$ intervals and bands for vectors and vector valued functions**. The intervals and bands thus obtained are extremely cautious and this lead us to call this subject **sure inference analysis**. The existence of orders of inference is also suggested. As an example an inferential sequence for estimators defined by integrals of random measures is given. We also suggest a way of calculating small sample intervals for the mean when we have some extra information, i.e., information that comes from outside the data.

## 1. Introduction

In this article the problem to be studied is that of finding confidence intervals for parameters, here real numbers or real vectors, and confidence bands for functions that may be real or vector valued ones. The analysis of inference that will be presented here is one which is extremely cautions. It is made in such a way to, as much as possible, avoid doubts and it is in this sense that we choose the name **sure inference analysis**.

The plan of this article is the following. In section 2 we present the first and central definitions and theorems for standard deviation based sure inference and variance based sure inference. No assumption about the probabilistic structure of random variables or stochastic processes but their inferential sequences are made till section 6. Section 3 is devoted to optimization of random sets (confidence intervals and bands) for real parameters and real functions. The optimal sets for vector parameters and vector valued functions are obtained in section 4. Examples are given in sections 5 and 6. Actually, section 6 is used to develop sure inference when some extra information is available such as relations between mean values and variances (standard deviations) or more complete information as distributions of random variables. We also suggest a way of calculating small sample confidence intervals for the mean of some random variables. In section 7 we close the work with some comments.

We will assume that $X : \Omega \to \mathbb{R}$ is an unbiased estimator for $x$ and that $\mathrm{Std}\,(X) = \sigma_1$. We will also assume that we have a sequence of non-negative estimators and finite standard deviations, $\hat{\sigma}_n$ and $\sigma_n$ for all $n \geq 1$ such that $\sigma_1 = \mathrm{Std}\,(X)$, $\sigma_{n+1} = \mathrm{Std}\,(\hat{\sigma}_n)$ and $E\hat{\sigma}_n = \sigma_n$. By Chebychev's

*Key words and phrases.* inferential sequence of random variables or of stochastic processes, m-th order at least probability $p$ confidence interval or confidence band, non-parametric confidence interval or confidence band, wavelet estimators.

The author thanks Our Lord and Savior Jesus Christ.

inequality, for $\lambda_1 > 0$, we have $P\{X(\omega) \in [x-\lambda_1\sigma_1, x+\lambda_1\sigma_1]\} \geq 1-1/\lambda_1^2$ and, equivalently, $P\{x \notin [X(\omega)-\lambda_1\sigma_1, X(\omega)+\lambda_1\sigma_1]\} \leq 1/\lambda_1^2$. Analogously we can write $P\{\sigma_n \notin [\hat{\sigma}_n(\omega)-\lambda_{n+1}\sigma_{n+1}, \hat{\sigma}_n(\omega)+\lambda_{n+1}\sigma_{n+1}]\} \leq 1/(\lambda_{n+1})^2$, for all $n \geq 1$. It may occur, and this is the most frequent situation in practice, that we don't know the value of $\sigma_1$ and we use $\tilde{\sigma}_1(\omega)$, some estimator of $\sigma_1$, and $X(\omega)$ to form confidence intervals for $x$ when the distribution of $X : \Omega \to \mathbb{R}$ is known. Note that $\tilde{\sigma}_1$ is not necessarily unbiased. In most of these situations, after some analysis has been done, we could almost conclude that, with probability $p$, $x$ belongs to the interval $[X(\omega)-\alpha\tilde{\sigma}_1(\omega), X(\omega)+\beta\tilde{\sigma}_1(\omega)]$ for some $\alpha, \beta \in \mathbb{R}_+$. We have said *almost conclude* because there is uncertainty associated with $\tilde{\sigma}_1$. On the other hand there are situations where we know the distribution of some statistic $S(X, \tilde{\sigma}_1, x)$ that allows us to conclude that with probability $p$, $x$ belongs to the set $S^{-1}(X, \tilde{\sigma}_1, \cdot)([\alpha, \beta])$ when $P\{\omega \in \Omega | \alpha \leq S(X, \tilde{\sigma}_1, x) \leq \beta\} = p$. As an example consider for $1 \leq i \leq n$, $Y_i \sim N(\mu, \sigma^2)$ normal random variables with mean $\mu$ and variance $\sigma^2$, $X = \sum_{i=1}^{n} Y_i/n$, $\tilde{\sigma}_1 = \sqrt{\sum_{i=1}^{n}(Y_i - X)^2/(n-1)}$, $x = \mu$, and $S(X, \tilde{\sigma}_1, x) = T(\frac{X-\mu}{\tilde{\sigma}_1})$ where $T$ is t-student distribution.

We are interested in the situation where not only we don't know the distribution of $X$ but also we want to decrease the uncertainty due to the substitution of $\tilde{\sigma}_1(\omega)$ or $\hat{\sigma}_1(\omega)$ by $\sigma_1$.

Since $P\{\sigma_1 \notin [\hat{\sigma}_1(\omega) - \lambda_2\sigma_2, \hat{\sigma}_1(\omega) + \lambda_2\sigma_2]\} \leq 1/\lambda_2^2$, we have

$$\begin{aligned} P\{\sigma_1 \leq \hat{\sigma}_1(\omega) + \lambda_2\sigma_2\} &= 1 - P\{\sigma_1 > \hat{\sigma}_1(\omega) + \lambda_2\sigma_2\} \\ &\geq 1 - P\{\sigma_1 \notin [\hat{\sigma}_1(\omega) - \lambda_2\sigma_2, \hat{\sigma}_1(\omega) + \lambda_2\sigma_2]\} \geq 1 - \frac{1}{\lambda_2^2}. \end{aligned}$$

Let $L(\omega, \lambda_1, \lambda_2) = \lambda_1(\hat{\sigma}_1(\omega) + \lambda_2\sigma_2)$, $A(\omega, \lambda_1, \lambda_2) = X(\omega) - L(\omega, \lambda_1, \lambda_2)$ and $B(\omega, \lambda_1, \lambda_2) = X(\omega) + L(\omega, \lambda_1, \lambda_2)$.

Let also

$$\begin{aligned} \Omega^+ &= \{\omega \in \Omega | \sigma_1 \leq \hat{\sigma}_1(\omega) + \lambda_2\sigma_2\}, \\ \Omega^0 &= \{\omega \in \Omega | x \in [X(\omega) - \lambda_1\sigma_1, X(\omega) + \lambda_1\sigma_1]\}, \\ \Omega^1 &= \{\omega \in \Omega | x \in [X(\omega) - L(\omega, \lambda_1, \lambda_2), X(\omega) + L(\omega, \lambda_1, \lambda_2)]\}. \end{aligned}$$

Thus we have $P(\Omega^0) \geq \left(1 - \frac{1}{\lambda_1^2}\right)$ and $P(\Omega^+) \geq \left(1 - \frac{1}{\lambda_2^2}\right)$.

Since $L(\omega, \lambda_1, \lambda_2) \geq \lambda_1\sigma_1$ when $\sigma_1 \leq \hat{\sigma}_1(\omega) + \lambda_2\sigma_2$, we have $(\Omega^+ \cap \Omega^1) \supset (\Omega^+ \cap \Omega^0)$ and we can write

$$\begin{aligned} &P\{x \in [A(\omega, \lambda_1, \lambda_2), B(\omega, \lambda_1, \lambda_2)]\} = P(\Omega^1) \\ &\geq P(\Omega^1 \cap \Omega^+) \geq P(\Omega^0 \cap \Omega^+) \\ &\geq P(\Omega^0) + P(\Omega^+) - 1 \geq 1 - \frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2}. \end{aligned}$$

The inequality above lets us draw such conclusions as: with at least probability $\left(1 - \frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2}\right)$, $x$ belongs to the interval $[X(\omega) - \lambda_1(\hat{\sigma}_1(\omega) + \lambda_2\sigma_2), X(\omega) + \lambda_1(\hat{\sigma}_1(\omega) + \lambda_2\sigma_2)]$.

In practice, this interval will be replaced by

$$[X(\omega) - \lambda_1(\hat{\sigma}_1(\omega) + \lambda_2\hat{\sigma}_2(\omega)), X(\omega) + \lambda_1(\hat{\sigma}_1(\omega) + \lambda_2\hat{\sigma}_2(\omega))]$$

and this substitution induces some uncertainty. In this situation we can almost conclude that, with at least probability $\left(1 - \frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2}\right)$, $x$ belongs to the later interval. We can continue this process of analyzing the worst case and obtain probabilities in the form $1-\sum_{i=1}^{n}\frac{1}{\lambda_i^2}$ for intervals of the form $[X(\omega)-L_m(\omega,\lambda_1,\ldots,\lambda_m), X(\omega)+L_m(\omega,\lambda_1,\ldots,\lambda_m)]$ where

$$L_m(\omega,\lambda_1,\ldots,\lambda_m)=\lambda_1(\hat{\sigma}_1(\omega)+\lambda_2(\cdots+\lambda_{m-1}(\hat{\sigma}_{m-1}(\omega)+\lambda_m\sigma_m)\cdots)).$$

## 2. Definitions and Theorems

The veracity of the following theorems is not changed by the substitution of m-th order inferential sets for inferential sequences in their hypothesis.

### 2.1. Standard Deviation Based Sure Inference.

#### 2.1.1. *Real Random Variables and Stochastic Process.*

**Definition 1.** *A 3-tuple $(X,(\sigma_n)_{1\le n\le m},(\hat{\sigma}_n)_{1\le n\le m})$ formed by a random variable $X:\Omega\to\mathbb{R}$, $m$ positive numbers $(\sigma_n)_{1\le n\le m}$ and $m$ random variables $(\hat{\sigma}_n:\Omega\to\mathbb{R})_{1\le n\le m}$, is a **standard deviation based m-th order inferential set** for $x\in\mathbb{R}$ if and only if*

(i) $EX=x$, $\sigma_1=\mathrm{Std}(X)$,
(ii) *if* $1\le n\le m-1$ *then* $\sigma_{n+1}=\mathrm{Std}(\hat{\sigma}_n)$,
(iii) *if* $1\le n\le m-1$ *then* $E\hat{\sigma}_n=\sigma_n$, *and* $E\hat{\sigma}_m\ge\sigma_m$
(vi) *if* $1\le n\le m$ *then* $\hat{\sigma}_n(\Omega)\subset\mathbb{R}_+$.

**Definition 2.** *A 3-tuple $(X,(\sigma_n)_{n\in\mathbb{N}^*},(\hat{\sigma}_n)_{n\in\mathbb{N}^*})$ formed by a random variable $X:\Omega\to\mathbb{R}$, a sequence of positive numbers $(\sigma_n)_{n\in\mathbb{N}^*}$ and a sequence of random variables $(\hat{\sigma}_n:\Omega\to\mathbb{R})_{n\in\mathbb{N}^*}$, is **a standard deviation based inferential sequence** for $x\in\mathbb{R}$ if and only if*

(i) $EX=x$, $\sigma_1=\mathrm{Std}(X)$,
(ii) $\forall n\in\mathbb{N}^*$ $\sigma_{n+1}=\mathrm{Std}(\hat{\sigma}_n)$,
(iii) $\forall n\in\mathbb{N}^*$ $E\hat{\sigma}_n=\sigma_n$,
(vi) $\forall n\in\mathbb{N}^*$ $\hat{\sigma}_n(\Omega)\subset\mathbb{R}_+$.

We will use the notation $(X,\ \sigma_n,\ \hat{\sigma}_n)$ to represent an inferential sequence and, occasionally, we will simply say that the sequences $\sigma_n$ and $\hat{\sigma}_n$ form an inferential sequence for $x$. Observe that this definition implies the fact that all random variables, that is, $X$ and $\hat{\sigma}_n$, $n\ge 1$, have finite expectations and variances, which is a necessary condition to apply Chebychev's inequality to each of them.

**Theorem 2.1. (Standard deviation inferential sequence of random variables' theorem.)** *Let $(X,\ \sigma_n,\ \hat{\sigma}_n)$ be an inferential sequence for $x\in\mathbb{R}$. If*

$$L_m(\omega,\lambda_1,\ldots,\lambda_m)=\lambda_1(\hat{\sigma}_1(\omega)+\lambda_2(\cdots+\lambda_{m-1}(\hat{\sigma}_{m-1}(\omega)+\lambda_m\sigma_m)\cdots)),$$

*$\lambda_i\in\mathbb{R}_+^*$ for all $1\le i\le m$, $m\in\mathbb{N}^*$, then*

$$P\{x\in[X(\omega)-L_m(\omega,\lambda_1,\ldots,\lambda_m),X(\omega)+L_m(\omega,\lambda_1,\ldots,\lambda_m)]\}\ge 1-\sum_{i=1}^{m}\frac{1}{\lambda_i^2}.$$

**Proof** (By induction.) If $m = 1$, then

$$P\{x \in [X(\omega) - \lambda_1\sigma_1, X(\omega) + \lambda_1\sigma_1]\} = 1 - P\{x \notin [X(\omega) - \lambda_1\sigma_1, X(\omega) + \lambda_1\sigma_1]\} \geq 1 - \frac{1}{\lambda_1^2},$$

by Chebychev's inequality.

For easy of notation, let $A_k(\omega) = X(\omega) - L_k(\omega, \lambda_1, \ldots, \lambda_k)$ and $B_k(\omega) = X(\omega) + L_k(\omega, \lambda_1, \ldots, \lambda_k)$. Under the assumption that the statement holds for $m-1$, we have

$$\begin{aligned} P\{x \in [A_m(\omega), B_m(\omega)]\} &\geq P\{x \in [A_m(\omega), B_m(\omega)] \wedge \sigma_{m-1} \leq \hat{\sigma}_{m-1}(\omega) + \lambda_m\sigma_m\} \\ &\geq P\{x \in [A_{m-1}(\omega), B_{m-1}(\omega)] \wedge \sigma_{m-1} \leq \hat{\sigma}_{m-1}(\omega) + \lambda_m\sigma_m\} \end{aligned}$$

since

$$[A_{m-1}(\omega), B_{m-1}(\omega)] \subset [A_m(\omega), B_m(\omega)]$$

when $\sigma_{m-1} \leq \hat{\sigma}_{m-1}(\omega) + \lambda_m\sigma_m$.

Thus, $P\{x \in [A_m(\omega), B_m(\omega)]\}$

$$\begin{aligned} &\geq P\{x \in [A_{m-1}(\omega), B_{m-1}(\omega)]\} + P\{\sigma_{m-1} \leq \hat{\sigma}_{m-1}(\omega) + \lambda_m\sigma_m\} - 1 \\ &\geq \left(1 - \sum_{i=1}^{m-1} \frac{1}{\lambda_i^2}\right) + \left(1 - \frac{1}{\lambda_m^2}\right) - 1 = 1 - \sum_{i=1}^{m} \frac{1}{\lambda_i^2}, \end{aligned}$$

since

$$\begin{aligned} P\{\sigma_{m-1} \leq \hat{\sigma}_{m-1}(\omega) + \lambda_m\sigma_m\} &\geq 1 - P\{\sigma_{m-1} \notin [\hat{\sigma}_{m-1}(\omega) - \lambda_m\sigma_m, \hat{\sigma}_{m-1}(\omega) + \lambda_m\sigma_m]\} \\ &\geq 1 - \frac{1}{\lambda_m^2}. \end{aligned}$$

■

If we substitute $\hat{\sigma}_m(\omega)$ for $\sigma_m$ some uncertainty will be introduced in our analysis. This kind of uncertainty may be eliminated if we know the value of $\sigma_J$ for some $J$ or some superior bound for $\sigma_J$.

We observe that by applying this **sure inference analysis** we can obtain these confidence intervals of "at least probability $p$", for $x$ whatever the distribution of $X$ is. Furthermore, this analysis is more conservative then that made if we assume some distribution to construct confidence intervals.

**Definition 3.** *We will call the interval* $[X(\omega) - L_m(\omega, \lambda_1, \ldots, \lambda_m), X(\omega) + L_m(\omega, \lambda_1, \ldots, \lambda_m)]$ *an m-th order standard deviation based at least probability* $p = 1 - \sum_{i=1}^{m} \frac{1}{\lambda_i^2}$ *confidence interval for* $x$.

Briefly we will call this intervals **sure inference intervals**.

As random variables are estimators of real numbers, stochastic processes can be understood as estimators of functions.

If $X : \Omega \times \mathbb{R} \to \mathbb{R}$ is a stochastic process which is an unbiased estimator for the function $x : \mathbb{R} \to \mathbb{R}$, that is, such that $EX$ and $x$ are equal, and we have sequences of non-negative estimators (stochastic process in this case) $\hat{\sigma}_n : \Omega \times \mathbb{R} \to \mathbb{R}$, and standard deviations (standard deviation functions), $\sigma_n : \mathbb{R} \to \mathbb{R}$ for $n \geq 1$, such that

$$\sigma_1 = \operatorname{Std} X : \begin{array}{l} \mathbb{R} \to \mathbb{R}, \\ t \to \operatorname{Std}(X(t) : \Omega \to \mathbb{R}) \end{array} \qquad \sigma_{n+1} = \operatorname{Std}(\hat{\sigma}_n) \text{ and } E\hat{\sigma}_n = \sigma_n,$$

we can develop a sure inference analysis to obtain "at least probability $p$" confidence bands in a completely similar way to that presented above for random variables.

From now on, $I$ is simply an arbitrary set.

**Definition 4.** *A 3-tuple $(X, (\sigma_n)_{n\in\mathbb{N}^*}, (\hat{\sigma}_n)_{n\in\mathbb{N}^*})$ formed by a stochastic process $X : \Omega \times I \to \mathbb{R}$, a sequence of functions $(\sigma_n : I \to \mathbb{R})_{n\in\mathbb{N}^*}$ and a sequence of stochastic processes $(\hat{\sigma}_n : \Omega \times I \to \mathbb{R})_{n\in\mathbb{N}^*}$ is a* **standard deviation based inferential sequence** *for $x : I \to \mathbb{R}$ if and only if*

(i) $EX = x$, $\sigma_1 = \mathrm{Std}\,(X)$,
(ii) $\forall n \in \mathbb{N}^* \quad \sigma_{n+1} = \mathrm{Std}\,(\hat{\sigma}_n)$,
(iii) $\forall n \in \mathbb{N}^* \quad E\hat{\sigma}_n = \sigma_n$,
(vi) $\forall n \in \mathbb{N}^* \quad \hat{\sigma}_n(\Omega \times I) \subset \mathbb{R}_+$.

Analogously we define m-th order standart deviation based inferential sets for $x : I \to \mathbb{R}$.

**Theorem 2.2. (Standard deviation inferential sequence of stochastic processes' theorem).** *Let $(X, \sigma_n, \hat{\sigma}_n)$ be an inferential sequence for $x : I \to \mathbb{R}$. Defining for all $m \in \mathbb{N}^*$, $L_m : \Omega \times I \times (\mathbb{R}_+^*)^m \to \mathbb{R}_+$ by*

$$L_m(\omega, t, \lambda_1, \ldots, \lambda_m) = \lambda_1(\hat{\sigma}_1(\omega, t) + \lambda_2(\ldots + \lambda_{m-1}(\hat{\sigma}_{m-1}(\omega, t) + \lambda_m\sigma_m(t))\ldots)),$$

*we have, for all $t \in I$, and all $m \in \mathbb{N}^*$,*

$$P\{x(t) \in [X(\omega, t) - L_m(\omega, t, \lambda_1, \ldots, \lambda_m), X(\omega, t) + L_m(\omega, t, \lambda_1, \ldots, \lambda_m)]\} \geq 1 - \sum_{i=1}^{m} \frac{1}{\lambda_i^2}.$$

**Proof** It is sufficient to observe that, for each fixed and arbitrary $t$, we have as a direct consequence of definitions (2) and (4) that $(X(t), (\sigma_n(t))_{n\in\mathbb{N}^*}, (\hat{\sigma}_n(t))_{n\in\mathbb{N}^*})$ is an inferential sequence for $x(t)$ and apply Theorem 2.1. ∎

**Definition 5.** *We will call the set*

$$\{[X(\omega, t) - L_m(\omega, t, \lambda_1, \ldots, \lambda_m), X(\omega, t) + L_m(\omega, t, \lambda_1, \ldots, \lambda_m)] | t \in I\} \tag{1}$$

*an* **m-th order standard deviation based at least probability $p$ confidence band.**

Briefly we will call this bands **sure inference bands**.

For example, if $I = \mathbb{R}$, the sure inference band (1) corresponds to the $t, x$ plane region delimited by the curves $x_1(t) = X(\omega, t) - L_m(\omega, t, \lambda_1, \ldots, \lambda_m)$ and $x_2(t) = X(\omega, t) + L_m(\omega, t, \lambda_1, \ldots, \lambda_m)$ that contains the curve $X(\omega, t)$. If $I = \mathbb{R}^2$, $t = (t_1, t_2)$, then the sure inference band (1) corresponds to the $\mathbb{R}^3$ region delimited, in a natural way, by the surfaces (not necessarily continuous) $x_1(t_1, t_2) = X(\omega, t_1, t_2) - L_m(\omega, t_1, t_2, \lambda_1, \ldots, \lambda_m)$ and $x_2(t_1, t_2) = X(\omega, t_1, t_2) + L_m(\omega, t_1, t_2, \lambda_1, \ldots, \lambda_m)$ that contains the graph of the mapping $X : \{\omega\} \times \mathbb{R}^2 \to \mathbb{R}$, that is, the surface $X(\omega, t_1, t_2)$. If $I = \mathbb{R}^m$ we have analogous interpretations.

2.1.2. *Random Vectors and Stochastic Process Over* $\mathbb{R}^q$. We will now study the case of random vectors.Let $X = (X_1, \ldots, X_q) : \Omega \to \mathbb{R}^q$, $X_i : \Omega \to \mathbb{R}$, $1 \leq i \leq q$, and $\mathrm{Std}\,(X) = (\mathrm{Std}\,(X_1), \ldots, \mathrm{Std}\,(X_q))$. We define m-th order inferential sets in an analogous way.

**Definition 6.** *A 3-tuple $(X,(\sigma_n)_{n\in\mathbb{N}^*},(\hat\sigma_n)_{n\in\mathbb{N}^*})$ formed by a random vector $X:\Omega\to\mathbb{R}^q$, a sequence of vectors $(\sigma_n)_{n\in\mathbb{N}^*}=(\sigma_{1,n},\ldots,\sigma_{q,n})_{n\in\mathbb{N}^*}$, where $(\sigma_{i,n})_{n\in\mathbb{N}^*}$ is a sequence of positive numbers for all $i$, $1\le i\le q$, and a sequence of random vectors $(\hat\sigma_n:\Omega\to\mathbb{R}^q)_{n\in\mathbb{N}^*}$, is a* ***standard deviation based inferential sequence*** *for $x\in\mathbb{R}^q$ if and only if*

(i) $EX=x$, $\sigma_1=\mathrm{Std}(X)$,
(ii) $\forall n\in\mathbb{N}^*$ $\sigma_{n+1}=\mathrm{Std}(\hat\sigma_n)$,
(iii) $\forall n\in\mathbb{N}^*$ $E\hat\sigma_n=\sigma_n$,
(vi) $\forall n\in\mathbb{N}^*$ $\hat\sigma_n(\Omega)\subset\mathbb{R}^q_+$.

**Theorem 2.3. (Standard deviation inferential sequence of random vectors' theorem.)** *Let $(X,\ \sigma_n,\ \hat\sigma_n)$ be an inferential sequence for $x\in\mathbb{R}^q$. If*

$$L_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m})=\lambda_{j,1}(\hat\sigma_{j,1}(\omega)+\lambda_{j,2}(\cdots+\lambda_{j,m-1}(\hat\sigma_{j,m-1}(\omega)+\lambda_{j,m}\sigma_{j,m})\cdots)),$$

*$\lambda_{j,i}\in\mathbb{R}^*_+$ for all $1\le j\le q$, $1\le i\le m$, $m\in\mathbb{N}^*$, then*

$$P\{x\in\prod_{j=1}^{q}[X_j(\omega)-L_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m}),X_j(\omega)+L_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m})]\}\ge 1-\sum_{j=1}^{q}\sum_{i=1}^{m}\frac{1}{\lambda_{j,i}^2}.$$

**Proof** Observe that for events $A_j$, such that $P(A_j)=1-\alpha_j$ we have $P(\bigcap_{j=1}^{q}A_j)\ge 1-\sum_{j=1}^{q}\alpha_j$ as can be easily checked by induction.

Let $A_j=\{\omega\in\Omega|x_j\in[X_j(\omega)-L_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m}),X_j(\omega)+L_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m})]\}$. From Theorem 2.1, for each $j$, $1\le j\le q$,

$$P\{A_j\}\ge 1-\sum_{i=1}^{m}\frac{1}{\lambda_{j,i}^2}.$$

Thus

$$P\{x\in\prod_{j=1}^{q}[X_j(\omega)-L_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m}),X_j(\omega)+L_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m})]\}=$$

$$P(\bigcap_{j=1}^{q}A_j)\ge 1-\sum_{j=1}^{q}(\sum_{i=1}^{m}\frac{1}{\lambda_{j,i}^2}).$$

■

**Definition 7.** $\prod_{j=1}^{q}[X_j(\omega)-L_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m}),X_j(\omega)+L_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m})]$ *is called an m-th order standard deviation based at least probability $p=1-\sum_{j=1}^{q}(\sum_{i=1}^{m}\frac{1}{\lambda_{j,i}^2})$ confidence $\mathbb{R}^q$-interval for $x\in\mathbb{R}^q$.*

We note that the $\mathbb{R}^q$ sure inference interval above does not dependent on the joint distribution of $X=(X_1,\ldots,X_p)$.

**Definition 8.** *A 3-tuple $(X,(\sigma_n)_{n\in\mathbb{N}^*},(\hat{\sigma}_n)_{n\in\mathbb{N}^*})$ formed by a vector valued stochastic process $X:\Omega\times I\rightarrow\mathbb{R}^q$, a sequence of vector valued functions $(\sigma_n:I\rightarrow\mathbb{R}^q)_{n\in\mathbb{N}^*}$ and a sequence of vector valued stochastic processes $(\hat{\sigma}_n:\Omega\times I\rightarrow\mathbb{R}^q)_{n\in\mathbb{N}^*}$ is a* **standard deviation based inferential sequence** for $x:I\rightarrow\mathbb{R}^q$ *if and only if*

(i) $EX=x$, $\sigma_1=\mathrm{Std}\,(X)$,
(ii) $\forall n\in\mathbb{N}^*\quad\sigma_{n+1}=\mathrm{Std}\,(\hat{\sigma}_n)$,
(iii) $\forall n\in\mathbb{N}^*\quad E\hat{\sigma}_n=\sigma_n$,
(vi) $\forall n\in\mathbb{N}^*\quad\hat{\sigma}_n(\Omega\times I)\subset\mathbb{R}^q_+$.

**Theorem 2.4. (Inferential sequence of vector valued stochastic processes' theorem).** *Let $(X,\sigma_n,\hat{\sigma}_n)$ be an inferential sequence for $x:I\rightarrow\mathbb{R}^q$. Defining for all $j$, $1\leq j\leq q$, and for all $m\in\mathbb{N}^*$, $L_{j,m}:\Omega\times I\times(\mathbb{R}^*_+)^m\rightarrow\mathbb{R}_+$ by*

$$L_{j,m}(\omega,t,\lambda_{j,1},\ldots,\lambda_{j,m})=\lambda_{j,1}(\hat{\sigma}_{j,1}(\omega,t)+\lambda_{j,2}(\ldots+\lambda_{j,m-1}(\hat{\sigma}_{j,m-1}(\omega,t)+\lambda_{j,m}\sigma_{j,m}(t))\ldots)),$$

*we have, for all $t\in I$, and all $m\in\mathbb{N}^*$,*

$$P\{x(t)\in\prod_{j=1}^{q}[X_j(\omega,t)-L_{j,m}(\omega,t,\lambda_{j,1},\ldots,\lambda_{j,m}),X_j(\omega,t)+L_{j,m}(\omega,t,\lambda_{j,1},\ldots,\lambda_{j,m})]\}$$

$$\geq 1-\sum_{j=1}^{p}\sum_{i=1}^{m}\frac{1}{\lambda_{j,i}^2}.$$

**Proof** It is sufficient to observe that, for each fixed and arbitrary $t$, we have as a direct consequence of definitions (6) and (8) that $(X(t),(\sigma_n(t))_{n\in\mathbb{N}^*},(\hat{\sigma}_n(t))_{n\in\mathbb{N}^*})$ is an inferential sequence for $x(t)$ and apply Theorem 2.3. ▪

**Definition 9.** *We will call the set*

$$\{\prod_{j=1}^{q}[X_j(\omega,t)-L_{j,m}(\omega,t,\lambda_{j,1},\ldots,\lambda_{j,m}),X_j(\omega,t)+L_{j,m}(\omega,t,\lambda_{j,1},\ldots,\lambda_{j,m})]|t\in I\}$$

*an* **m-th order standard deviation based at least probability $p$ confidence band** *for the vector valued function $x$.*

Briefly we will call this bands **sure inference bands.**

For example, if we take $I=\mathbb{R}$ and $q=2$ then the sure inference band is a tubular neighborhood contained in $\mathbb{R}^3$ of the curve $\widetilde{X}(\omega,t)=(t,X_1(\omega,t),X_2(\omega,t))$ which section at $t$ is the rectangle

$$[X_1(\omega,t)-L_{1,m}(\omega,t,\lambda_{1,1},\ldots,\lambda_{1,m}),X_1(\omega,t)+L_{1,m}(\omega,t,\lambda_{1,1},\ldots,\lambda_{1,m})]\times$$

$$[X_2(\omega,t)-L_{2.m}(\omega,t,\lambda_{2,1},\ldots,\lambda_{2,m}),X_2(\omega,t)+L_{2,m}(\omega,t,\lambda_{2,1},\ldots,\lambda_{2,m})].$$

If $I=\mathbb{R}^2$ and $q=2$ we can not visualize the band. Whenever we have $I=\mathbb{R}^r$, the bands will be tubular neighborhoods of the surface $\widetilde{X}(\omega,t)=(t,X(\omega,t))$ in $\mathbb{R}^{r+q}$ which section at $t\in\mathbb{R}^r$ is a $q$-dimensional parallelepiped.

**2.2. Variance Based Sure Inference.** We will assume that $X:\Omega\to\mathbb{R}$ is an unbiased estimator for $x$ and that $\mathrm{Var}(X)=\sigma_1^2$. We will also assume that we have a sequence of non-negative estimators and finite variances, $\hat{V}_n$ and $V_n$ for all $n\geq 1$ such that $V_1=\mathrm{Var}(X)$, $V_{n+1}=\mathrm{Var}\,\hat{V}_n$ and $E\hat{V}_n=V_n$. By Chebychev's inequality, for $\lambda_1>0$, we have $P\{X(\omega)\in[x-\lambda_1\sigma_1,x+\lambda_1\sigma_1]\}\geq 1-1/\lambda_1^2$ and, equivalently, $P\{x\notin[X(\omega)-\lambda_1\sigma_1,X(\omega)+\lambda_1\sigma_1]\}\leq 1/\lambda_1^2$. Let $\sigma_n=\sqrt{V_n}$.

Since $E\hat{V}_n=V_n$, by Chebychev's inequality, $P\{\hat{V}_n(\omega)\in[V_n-\lambda_{n+1}\sqrt{V_{n+1}},V_n+\lambda_{n+1}\sqrt{V_{n+1}}]\}\geq 1-1/\lambda_{n+1}^2$ and, equivalently, $P\{V_n\in[\hat{V}_n(\omega)-\lambda_{n+1}\sqrt{V_{n+1}},\hat{V}_n(\omega)+\lambda_{n+1}\sqrt{V_{n+1}}]\}\geq 1-1/\lambda_{n+1}^2$. So, we have

$$P\{\sigma_n\in\left[\sqrt{\max\{0,\hat{V}_n(\omega)-\lambda_{n+1}\sqrt{V_{n+1}}\}},\sqrt{\hat{V}_n(\omega)+\lambda_{n+1}\sqrt{V_{n+1}}}\right]\}\geq 1-1/\lambda_{n+1}^2$$

from which $P\{\sigma_n\leq\sqrt{\hat{V}_n(\omega)+\lambda_{n+1}\sqrt{V_{n+1}}}\}\geq 1-1/\lambda_{n+1}^2$ for all $n\in\mathbb{N}^*$.

Proceeding as before we start writing $P\{x\in[X(\omega)-\lambda_1\sigma_1,X(\omega)+\lambda_1\sigma_1]\wedge\sigma_1\leq\sqrt{\hat{V}_1(\omega)+\lambda_2\sqrt{V_2}}\}\geq 1-1/\lambda_1^2-1/\lambda_2^2$ that is $P\{x\in[X(\omega)-\lambda_1\sqrt{\hat{V}_1(\omega)+\lambda_2\sqrt{V_2}},X(\omega)+\lambda_1\sqrt{\hat{V}_1(\omega)+\lambda_2\sqrt{V_2}}]\geq 1-1/\lambda_1^2-1/\lambda_2^2$ and continue to obtain the conclusion that with at least probability $1-\sum_{i=1}^{m}1/\lambda_i^2$, $x$ belongs to the interval $[X(\omega)-L_m^v(\omega,\lambda_1,\ldots,\lambda_m),X(\omega)+L_m^v(\omega,\lambda_1,\ldots,\lambda_m)]$ where

$$L_m^v(\omega,\lambda_1,\ldots,\lambda_m)=\lambda_1\sqrt{\hat{V}_1(\omega)+\lambda_2\sqrt{\ldots+\lambda_{m-1}\sqrt{\hat{V}_{m-1}(\omega)+\lambda_m\sqrt{V_m}}}}.$$

In practice we will replace $L_m^v$ by $\hat{L}_m^v$,

$$\hat{L}_m^v(\omega,\lambda_1,\ldots,\lambda_m)=\lambda_1\sqrt{\hat{V}_1(\omega)+\lambda_2\sqrt{\ldots+\lambda_{m-1}\sqrt{\hat{V}_{m-1}(\omega)+\lambda_m\sqrt{\hat{V}_m(\omega)}}}}.$$

### 2.2.1. *Real Random Variables and Stochastic Process.*

**Definition 10.** *A 3-tuple $(X,(V_n)_{1\leq n\leq m},(\hat{V}_n)_{1\leq n\leq m})$ formed by a random variable $X:\Omega\to\mathbb{R}$, $m$ positive numbers $(V_n)_{1\leq n\leq m}$ and $m$ random variables $(\hat{V}_n:\Omega\to\mathbb{R})_{1\leq n\leq m}$, is an variance based m-th order inferential set for $x\in\mathbb{R}$ if and only if*

(i) $EX=x$, $V_1=\mathrm{Var}(X)$,
(ii) *if* $1\leq n\leq m-1$ *then* $V_{n+1}=\mathrm{Var}(\hat{V}_n)$,
(iii) *if* $1\leq n\leq m-1$ *then* $E\hat{V}_n=V_n$, *and* $E\hat{V}_m\geq V_m$
(vi) *if* $1\leq n\leq m$ *then* $\hat{V}_n(\Omega)\subset\mathbb{R}_+$.

**Definition 11.** *A 3-tuple $(X,(V_n)_{n\in\mathbb{N}^*},(\hat{V}_n)_{n\in\mathbb{N}^*})$ formed by a random variable $X:\Omega\to\mathbb{R}$, a sequence of positive numbers $(V_n)_{n\in\mathbb{N}^*}$ and a sequence of random variables $(\hat{V}_n:\Omega\to\mathbb{R})_{n\in\mathbb{N}^*}$, is a variance based inferential sequence for $x\in\mathbb{R}$ if and only if*

(i) $EX=x$, $V_1=Var(X)$,
(ii) $\forall n\in\mathbb{N}^*$ $V_{n+1}=Var(\hat{V}_n)$,
(iii) $\forall n\in\mathbb{N}^*$ $E\hat{V}_n=V_n$,
(vi) $\forall n\in\mathbb{N}^*$ $\hat{V}_n(\Omega)\subset\mathbb{R}_+$.

We will use the notation $(X,\ V_n,\ \hat{V}_n)$ to represent an inferential sequence and, occasionally, we will simply say that the sequences $V_n$ and $\hat{V}_n$ form an inferential sequence for $x$. Observe that this definition implies the fact that all random variables, that is, $X$ and $\hat{V}_n$, $n \geq 1$, have finite expectations and variances, which is a necessary condition to apply Chebychev's inequality to each of them.

**Theorem 2.5. (Variance based inferential sequence of random variables' theorem.)** *Let $(X,\ V_n,\ \hat{V}_n)$ be an inferential sequence for $x \in \mathbb{R}$. If*

$$L^v_m(\omega,\lambda_1,\ldots,\lambda_m) = \lambda_1\sqrt{\hat{V}_1(\omega)+\lambda_2\sqrt{\ldots+\lambda_{m-1}\sqrt{\hat{V}_{m-1}(\omega)+\lambda_m\sqrt{V_m}}}},$$

$\lambda_i \in \mathbb{R}^*_+$ *for all* $1 \leq i \leq m$, $m \in \mathbb{N}^*$, *then*

$$P\{x \in [X(\omega)-L^v_m(\omega,\lambda_1,\ldots,\lambda_m), X(\omega)+L^v_m(\omega,\lambda_1,\ldots,\lambda_m)]\} \geq 1-\sum_{i=1}^{m}\frac{1}{\lambda_i^2}.$$

**Proof** Substitute $\sqrt{\hat{V}_{m-1}(\omega)+\lambda_m\sqrt{V_m}}$ for $\hat{\sigma}_{m-1}(\omega)+\lambda_m\sigma_m$ and $L^v_k$ for $L_k$ in the demonstration of Theorem 2.1. ∎

**Definition 12.** *We will call the interval $[X(\omega)-L^v_m(\omega,\lambda_1,\ldots,\lambda_m), X(\omega)+L^v_m(\omega,\lambda_1,\ldots,\lambda_m)]$ an m-th order variance based at least probability $p = 1-\sum_{i=1}^{m}\frac{1}{\lambda_i^2}$ confidence interval for $x$.*

If $X : \Omega \times \mathbb{R} \to \mathbb{R}$ is a stochastic process which is an unbiased estimator for the function $x : \mathbb{R} \to \mathbb{R}$, that is, such that $EX$ and $x$ are equal, and we have sequences of non-negative estimators (stochastic process in this case) $\hat{V}_n : \Omega \times \mathbb{R} \to \mathbb{R}$, and variances (variance functions), $V_n : \mathbb{R} \to \mathbb{R}$ for $n \geq 1$, such that

$$V_1 = \operatorname{Var} X : \begin{array}{l} \mathbb{R} \to \mathbb{R}, \\ t \to \operatorname{Var}(X(t) : \Omega \to \mathbb{R}) \end{array} \qquad V_{n+1} = \operatorname{Var}(\hat{V}_n) \ \text{ and } \ E\hat{V}_n = V_n,$$

we can develop a sure inference analysis to obtain "at least probability $p$" confidence bands in a completely similar way to that presented above for random variables.

**Definition 13.** *A 3-tuple $(X,(V_n)_{n\in\mathbb{N}^*},(\hat{V}_n)_{n\in\mathbb{N}^*})$ formed by a stochastic process $X : \Omega\times I \to \mathbb{R}$, a sequence of functions $(V_n : I \to \mathbb{R})_{n\in\mathbb{N}^*}$ and a sequence of stochastic processes $(\hat{V}_n : \Omega \times I \to \mathbb{R})_{n\in\mathbb{N}^*}$ is a* **variance based inferential sequence** *for $x : I \to \mathbb{R}$ if and only if*

(i) $EX = x$, $V_1 = Var(X)$,
(ii) $\forall n \in \mathbb{N}^*$ $V_{n+1} = Var(\hat{V}_n)$,
(iii) $\forall n \in \mathbb{N}^*$ $E\hat{V}_n = V_n$,
(vi) $\forall n \in \mathbb{N}^*$ $\hat{V}_n(\Omega \times I) \subset \mathbb{R}_+$.

**Theorem 2.6. (Variance based inferential sequence of stochastic processes' theorem).** *Let $(X,V_n,\hat{V}_n)$ be an inferential sequence for $x : I \to \mathbb{R}$. Defining for all $m \in \mathbb{N}^*$, $L^v_m : \Omega \times I \times (\mathbb{R}^*_+)^m \to \mathbb{R}_+$ by*

$$L^v_m(\omega,t,\lambda_1,\ldots,\lambda_m) = \lambda_1\sqrt{\hat{V}_1(\omega,t)+\lambda_2\sqrt{\ldots+\lambda_{m-1}\sqrt{\hat{V}_{m-1}(\omega,t)+\lambda_m\sqrt{V_m(t)}}}}$$

*we have, for all $t \in I$, and all $m \in \mathbb{N}^*$,*

$$P\{x(t) \in [X(\omega,t) - L^v_m(\omega,t,\lambda_1,\dots\lambda_m), X(\omega,t) + L^v_m(\omega,t,\lambda_1,\dots,\lambda_m)]\} \geq 1 - \sum_{i=1}^{m} \frac{1}{\lambda_i^2}.$$

**Proof** It is sufficient to observe that, for each fixed and arbitrary $t$, we have as a direct consequence of definitions (11) and (13) that $(X(t),(V_n(t))_{n\in\mathbb{N}^*},(\hat{V}_n(t))_{n\in\mathbb{N}^*})$ is a inferential sequence for $x(t)$ and apply Theorem 2.5. ∎

**Definition 14.** *We will call the set*

$$\{[X(\omega,t) - L^v_m(\omega,t,\lambda_1,\dots,\lambda_m), X(\omega,t) + L^v_m(\omega,t,\lambda_1,\dots,\lambda_m)] | t \in I\} \tag{2}$$

*an* **m-th order variance based at least probability $p$ confidence band.**

**2.3. Random Vectors and Stochastic Process Over $\mathbb{R}^q$.** We will now study the case of random vectors. Let $X = (X_1,\dots,X_q) : \Omega \to \mathbb{R}^q$, $X_i : \Omega \to \mathbb{R}$, $1 \leq i \leq q$, and $\mathrm{Var}(X) = (\mathrm{Var}(X_1),\dots,\mathrm{Var}(X_q))$

**Definition 15.** *A 3-tuple $(X,(V_n)_{n\in\mathbb{N}^*},(\hat{V}_n)_{n\in\mathbb{N}^*})$ formed by a random vector $X : \Omega \to \mathbb{R}^q$, a sequence of vectors $(V_n)_{n\in\mathbb{N}^*} = (V_{1,n},\dots,V_{q,n})_{n\in\mathbb{N}^*}$, where $(V_{i,n})_{n\in\mathbb{N}^*}$ is a sequence of positive numbers for all $i$, $1 \leq i \leq q$, and a sequence of random vectors $(\hat{V}_n : \Omega \to \mathbb{R}^q)_{n\in\mathbb{N}^*}$, is a variance based inferential sequence for $x \in \mathbb{R}^q$ if and only if*

(i) $EX = x$, $V_1 = Var(X)$,
(ii) $\forall n \in \mathbb{N}^*$ $V_{n+1} = Var(\hat{V}_n)$,
(iii) $\forall n \in \mathbb{N}^*$ $E\hat{V}_n = V_n$,
(vi) $\forall n \in \mathbb{N}^*$ $\hat{V}_n(\Omega) \subset \mathbb{R}^q_+$.

**Theorem 2.7. (Variance based inferential sequence of random vectors' theorem.)** *Let $(X, V_n, \hat{V}_n)$ be an inferential sequence for $x \in \mathbb{R}^q$. If*

$$L^v_{j,m}(\omega,\lambda_{j,1},\dots,\lambda_{j,m}) = \lambda_{j,1}\sqrt{\hat{V}_{j,1}(\omega) + \lambda_{j,2}\sqrt{\dots + \lambda_{j,m-1}\sqrt{\hat{V}_{j,m-1}(\omega) + \lambda_{j,m}\sqrt{V_{j,m}}}}}$$

*$\lambda_{j,i} \in \mathbb{R}^*_+$ for all $1 \leq j \leq q$, $1 \leq i \leq m$, $m \in \mathbb{N}^*$, then*

$$P\{x \in \prod_{j=1}^{q}[X_j(\omega) - L^v_{j,m}(\omega,\lambda_{j,1},\dots,\lambda_{j,m}), X_j(\omega) + L^v_{j,m}(\omega,\lambda_{j,1},\dots,\lambda_{j,m})]\} \geq 1 - \sum_{j=1}^{q}\sum_{i=1}^{m}\frac{1}{\lambda_{j,i}^2}.$$

**Proof** Substitute $L^v_{j,m}$ for $L_{j,m}$ in Theorem's 2.3 demonstration. ∎

**Definition 16.** $\prod_{j=1}^{q}[X_j(\omega) - L^v_{j,m}(\omega,\lambda_{j,1},\dots,\lambda_{j,m}), X_j(\omega) + L^v_{j,m}(\omega,\lambda_{j,1},\dots,\lambda_{j,m})]$ *is called an m-th order variance based at least probability $p = 1 - \sum_{j=1}^{q}(\sum_{i=1}^{m}\frac{1}{\lambda_{j,i}^2})$ confidence $\mathbb{R}^q$-interval for $x \in \mathbb{R}^q$.*

**Definition 17.** *A 3-tuple $(X,(V_n)_{n\in\mathbb{N}^*},(\hat{V}_n)_{n\in\mathbb{N}^*})$ formed by a vector valued stochastic process $X:\Omega\times I\to\mathbb{R}^q$, a sequence of vector valued functions $(V_n:I\to\mathbb{R}^q)_{n\in\mathbb{N}^*}$ and a sequence of vector valued stochastic processes $(\hat{V}_n:\Omega\times I\to\mathbb{R}^q)_{n\in\mathbb{N}^*}$ is a* **variance based inferential sequence** *for $x:I\to\mathbb{R}^q$ if and only if*

(i) $EX=x$, $V_1=Var(X)$,
(ii) $\forall n\in\mathbb{N}^*\ \ V_{n+1}=Var(\hat{V}_n)$,
(iii) $\forall n\in\mathbb{N}^*\ \ E\hat{V}_n=V_n$,
(vi) $\forall n\in\mathbb{N}^*\ \ \hat{V}_n(\Omega\times I)\subset\mathbb{R}^q_+$.

**Theorem 2.8. (Variance based inferential sequence of vector valued stochastic processes' theorem).** *Let $(X,V_n,\hat{V}_n)$ be an inferential sequence for $x:I\to\mathbb{R}^q$. Defining for all $j$, $1\le j\le q$, and for all $m\in\mathbb{N}^*$, $L^v_{j,m}:\Omega\times I\times(\mathbb{R}^*_+)^m\to\mathbb{R}_+$ by*

$$L^v_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m})=\lambda_{j,1}\sqrt{\hat{V}_{j,1}(\omega,t)+\lambda_{j,2}\sqrt{\dots+\lambda_{j,m-1}\sqrt{\hat{V}_{j,m-1}(\omega,t)+\lambda_{j,m}\sqrt{V_{j,m}(t)}}}}$$

*we have, for all $t\in I$, and all $m\in\mathbb{N}^*$,*

$$\begin{aligned}P\{x(t) &\in \prod_{j=1}^{q}[X_j(\omega,t)-L^v_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m}),X_j(\omega,t)+L^v_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m})]\}\\ &\ge 1-\sum_{j=1}^{q}\sum_{i=1}^{m}\frac{1}{\lambda_{j,i}^2}.\end{aligned}$$

**Proof** Immediate. ∎

**Definition 18.** *We will call the set*

$$\{\prod_{j=1}^{q}[X_j(\omega,t)-L^v_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m}),X_j(\omega,t)+L^v_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m})]|t\in I\}$$

*an* **m-th order variance based at least probability** $p$ **confidence band** *for the vector valued function $x$.*

We observe that letting $\lambda_i$ or $\lambda_{j,i}$ depend on $t\in I$ and substituting $\lambda_i(t)$ for $\lambda_i$ or $\lambda_{j,i}(t)$ for $\lambda_{j,i}$ on Theorems 2.2, 2.4, 2.6 and 2.8 we obtain new true statements. If we perform the same substitutions on Definitions 5, 9, 14 and 18, we obtain at least probability $p(t)$ confidence bands, that is, "non-homogeneous" confidence bands.

## 3. Optimal intervals and bands for numbers and real functions

Now we are interested in finding the optimal $m$-th order sure confidence interval or band, given an at least probability level $p$. Let us start with the simplest case, i.e., that of intervals for $x\in\mathbb{R}$. All $m$-th order standard deviation based intervals are written as $[X(\omega)-L_m(\omega,\lambda_1,\dots,\lambda_m),X(\omega)+L_m(\omega,\lambda_1,\dots,\lambda_m)]$ where $L_m(\omega,\lambda_1,\dots,\lambda_m)=\lambda_1(\hat{\sigma}_1(\omega)+\lambda_2(\dots+\lambda_{m-1}(\hat{\sigma}_{m-1}(\omega)+\lambda_m\sigma_m)\dots))=\sum_{i=1}^{m-1}(\prod_{j=1}^{i}\lambda_j)\hat{\sigma}_i(\omega)+(\prod_{i=1}^{m}\lambda_i)\sigma_m$. An estimate for this interval is written as $[X(\omega)-\hat{L}_m(\omega,\lambda_1,\ldots,\lambda_m), X(\omega)+\hat{L}_m(\omega,\lambda_1,\ldots,\lambda_m)]$ where $\hat{L}_m(\omega,\lambda_1,\ldots,\lambda_m) = \sum_{i=1}^{m}(\prod_{j=1}^{i}\lambda_j)\hat{\sigma}_i(\omega)$. The later interval is the one that we can access in practice.

Analogously, we write $[X(\omega)-L_m^v(\omega,\lambda_1,\ldots,\lambda_m), X(\omega)+L_m^v(\omega,\lambda_1,\ldots,\lambda_m)]$. An estimate for this interval is written as $[X(\omega)-\hat{L}_m^v(\omega,\lambda_1,\ldots,\lambda_m), X(\omega)+\hat{L}_m^v(\omega,\lambda_1,\ldots,\lambda_m)]$.

**Definition 19.** *An m-th order at least probability p interval for $x \in \mathbb{R}$, $[X(\omega)-\hat{L}_m(\omega,\lambda_1,\ldots,\lambda_m), X(\omega)+\hat{L}_m(\omega,\lambda_1,\ldots,\lambda_m)]$ or $[X(\omega)-\hat{L}_m^v(\omega,\lambda_1,\ldots,\lambda_m), X(\omega)+\hat{L}_m^v(\omega,\lambda_1,\ldots,\lambda_m)]$, is optimal if it has minimum length.*

**Theorem 3.1.** *Let $\overline{p}(m)$ be the greatest integer less than or equal to m such that $\hat{\sigma}_{\overline{p}(m)}(\omega) > 0$ and for all $j, \overline{p}(m) < j \leq m, \hat{\sigma}_j(\omega) = 0$.*

*The optimal m-th order standard deviation based at least probability p, $p \neq 1$, confidence interval for $x \in \mathbb{R}$ is written as $[X(\omega)-\hat{L}_m(\omega,\lambda_1^*,\ldots,\lambda_m^*), X(\omega)+\hat{L}_m(\omega,\lambda_1^*,\ldots,\lambda_m^*)]$ where $(\lambda_1^*,\ldots,\lambda_m^*) \in (\mathbb{R}_+^*)^m$ satisfies:*

*$\forall j$, $\overline{p}(m) < j \leq m$, $\lambda_j^*$ is an arbitrary positive real number and $(\lambda_1^*,\ldots,\lambda_{\overline{p}(m)}^*)$ is within the solutions of the simultaneous system of $\overline{p}(m)+1$ equations*

$$\sum_{k=1}^{\overline{p}(m)} \frac{1}{(\lambda_k)^2} = 1-p \text{ and } (\lambda_k)^2\left(\sum_{i=k}^{\overline{p}(m)}\left(\prod_{j=1}^{i}\lambda_j\right)\hat{\sigma}_i(\omega)\right) = \gamma$$

*for all k, $1 \leq k \leq \overline{p}(m)$, and $\gamma$ a real constant.*

**Proof** We search for the minimum of $\hat{L}_m(\omega,\lambda_1,\ldots,\lambda_m)$ as a function of $\lambda = (\lambda_1,\ldots,\lambda_m) \in (\mathbb{R}_+^*)^m$ subjected to the constraint $g_m(\lambda) = \sum_{i=1}^{m} 1/\lambda_i^2 \leq 1-p$. Since for all $\omega \in \Omega$ and $i$, $1 \leq i \leq m$, $\hat{\sigma}_i(\omega) \geq 0$ the function $\hat{L}_m(\omega,\cdot) : (\mathbb{R}_+^*)^m \subset \mathbb{R}^m \to \mathbb{R}$ is non-decreasing in each of its variables $\lambda_i$ and $g_m(\lambda)$ is decreasing in each of its variables, the constraint $g_m(\lambda) \leq 1-p$ may be replaced by $g_m(\lambda) = 1-p$ in our search for the minimum.

Thus we have $\sum_{i=1}^{\overline{p}(m)} \frac{1}{\lambda_i^2} = 1-p-\left(\sum_{i=\overline{p}(m)+1}^{m} \frac{1}{\lambda_i^2}\right)$.

Now, $\hat{L}_{\overline{p}(m)}(\omega,\cdot) : (\mathbb{R}_+^*)^{\overline{p}(m)} \subset \mathbb{R}^{\overline{p}(m)} \to \mathbb{R}$ is an increasing function of all of its variables $\lambda_1,\ldots,\lambda_{\overline{p}(m)}$.

Note that $\hat{L}_m(\omega,\lambda_1,\ldots,\lambda_m) = \hat{L}_{\overline{p}(m)}(\omega,\lambda_1,\ldots,\lambda_{\overline{p}(m)})$, that is, this value does not dependent on $\lambda_{\overline{p}(m)+1},\ldots,\lambda_m$. Let $a = \sum_{i=\overline{p}(m)+1}^{m} \frac{1}{\lambda_i^2}$ and $\tilde{\lambda} = (\lambda_1,\ldots,\lambda_{\overline{p}(m)}) \in (\mathbb{R}_+^*)^{\overline{p}(m)}$.

Thus we will assume the constraint $g_{\overline{p}(m)}(\lambda_1,\ldots,\lambda_{\overline{p}(m)}) = 1-p-a$ to hold.

For all $p$ and $a$, $0 \leq (p+a) < 1$ the set $G_{p+a} = \{\lambda | g_{\overline{p}(m)}(\lambda) = 1-p-a\} \subset (\mathbb{R}_+^*)^{\overline{p}(m)} \subset \mathbb{R}^{\overline{p}(m)}$ is a $\overline{p}(m)-1$ dimensional $C^\infty$ manifold with empty border and, for all $\omega \in \Omega$, $\hat{L}_{\overline{p}(m)}(\omega,\cdot)$ is a $C^\infty$ function. Thus, the critical points of $\hat{L}_{\overline{p}(m)}(\omega,\cdot)$ under $g_{\overline{p}(m)}$ constraint satisfy Lagrange's multipliers equation $d\hat{L}_{\overline{p}(m)}(\omega,\lambda^*(a)) = \gamma_1(a) dg_{\overline{p}(m)}(\lambda^*(a))$, $\lambda^*(a) \in (\mathbb{R}_+^*)^{\overline{p}(m)}$ and $\gamma_1(a)$ a real constant.

Due to $\partial G_{p+n} = \emptyset$, all local minimum points are critical points. Since, $\lim_{\substack{\|\tilde{\lambda}\|\to\infty \\ \tilde{\lambda}\in G_{p+n}}} \hat{L}_{\overline{p}(m)}(\omega,\tilde{\lambda}) = \infty$ there is a global minimum for $\hat{L}_{\overline{p}(m)}$ under $g_{\overline{p}(m)}(\widetilde{\lambda}) = 1-p-a$ constraint and it is within the critical points.

Now, $\forall a \in [0, 1-p)$ we have $d\hat{L}_{\overline{p}(m)}(\omega, \lambda^*(a)) = \gamma_1(a) dg_{\overline{p}(m)}(\lambda^*(a))$ and this relation defines $(\lambda^*(a), \gamma_1(a))$ as an implicit function of $a$ which is also $C^\infty$ in $a$.

Thus $\lim_{a\to 0}(\lambda^*(a), \gamma_1(a)) = (\lambda^*(0), \gamma_1(0)) = (\lambda^*, \gamma_1)$ is the solution for $d\hat{L}_{\overline{p}(m)}(\omega, \lambda^*) = \gamma_1 dg_{\overline{p}(m)}(\lambda^*)$ under the constraint $g_{\overline{p}(m)}(\lambda^*) = 1-p$.

Suppose $0 < a < b$. The minimum of $\hat{L}_{\overline{p}(m)}(\omega, \mu)$ under the constraint $g_{\overline{p}(m)}(\mu) = 1-p-a > 0$ is less then its minimum under the alternative constraint $g_{\overline{p}(m)}(\mu) = 1-p-b > 0$ since, if $\mu = (\mu_1, \ldots, \mu_{\overline{p}(m)})$ satisfies $g_{\overline{p}(m)}(\mu) = 1-p-b$ then $\mu' = (\mu_1 - \alpha_1, \ldots, \mu_{\overline{p}(m)} - \alpha_{\overline{p}(m)})$ satisfies $g_{\overline{p}(m)}(\mu') = 1-p-a$ for some $\alpha = (\alpha_1, \ldots, \alpha_{\overline{p}(m)}) \in (\mathbb{R}_+)^{\overline{p}(m)}$ from which $\hat{L}_{\overline{p}(m)}(\omega, \mu') < \hat{L}_{\overline{p}(m)}(\omega, \mu)$.

Thus we will assume the constraint

$$\sum_{i=1}^{\overline{p}(m)} \frac{1}{\lambda_i^2} = 1 - p - \lim_{\substack{\lambda_i\to\infty \\ \overline{p}(m)+1\le i\le m}} \left( \sum_{i=\overline{p}(m)+1}^{m} \frac{1}{\lambda_i^2} \right) = 1-p$$

to hold.

Now, $d\hat{L}_{\overline{p}(m)}(\omega, \tilde{\lambda}) = \left(\frac{\partial \hat{L}_{\overline{p}(m)}(\omega,\tilde{\lambda})}{\partial \lambda_1}, \ldots, \frac{\partial \hat{L}_{\overline{p}(m)}(\omega,\tilde{\lambda})}{\partial \lambda_{\overline{p}(m)}}\right)$; $dg_{\overline{p}(m)}(\tilde{\lambda}) = \left(\frac{-2}{\lambda_1^3}, \ldots, \frac{-2}{\lambda_{\overline{p}(m)}^3}\right)$ and $\frac{\partial \hat{L}_{\overline{p}(m)}}{\partial \lambda_k}(\omega, \widetilde{\lambda}) = \frac{\partial}{\partial \lambda_k}(\sum_{i=1}^{\overline{p}(m)} (\prod_{j=1}^{i} \lambda_j)\hat{\sigma}_i(\omega)) = \sum_{i=k}^{\overline{p}(m)} (\prod_{\substack{j=1 \\ j\neq k}}^{i} \lambda_j)\hat{\sigma}_i(\omega)$.

Thus, $(\lambda_k^*)^2 (\sum_{i=k}^{\overline{p}(m)} (\prod_{j=1}^{i} \lambda_j^*)\hat{\sigma}_i(\omega)) = -2\gamma_1 = \gamma$ for all $k$, $1 \le k \le \overline{p}(m)$, and $\lambda^* = (\lambda_1^*, \ldots, \lambda_{\overline{p}(m)}^*, \ldots, \lambda_m^*)$ with $\lambda_i^*$, $\overline{p}(m) < i \le m$ arbitrary real positive numbers. ∎

**Theorem 3.2.** *Let $\overline{p}(m)$ be the greatest integer less than or equal to $m$ such that $\hat{V}_{\overline{p}(m)}(\omega) > 0$ and for all $j$, $\overline{p}(m) < j \le m$, $\hat{V}_j(\omega) = 0$.*

*The optimal m-th order variance based at least probability p, $p \neq 1$, confidence interval for $x \in \mathbb{R}$ is written as $[X(\omega) - \hat{L}_m^v(\omega, \lambda_1^*, \ldots, \lambda_m^*), X(\omega) + \hat{L}_m^v(\omega, \lambda_1^*, \ldots, \lambda_m^*)]$ where $(\lambda_1^*, \ldots, \lambda_m^*) \in (\mathbb{R}_+^*)^m$ satisfies: $\forall j$, $\overline{p}(m) < j \le m$, $\lambda_j^*$ is an arbitrary positive real number and $(\lambda_1^*, \ldots, \lambda_{\overline{p}(m)}^*)$ is within the solutions of the simultaneous system of $\overline{p}(m)+1$ equations $\sum_{k=1}^{\overline{p}(m)} \frac{1}{(\lambda_k)^2} = 1-p$ and*

$$\frac{(\lambda_k)^2 (\prod_{i=1}^{k} \lambda_i) \Delta_{m,k}(\omega, \lambda_1, \ldots, \lambda_m)}{2^k \prod_{i=1}^{k-1} \Delta_{m,i}(\omega, \lambda_1, \ldots, \lambda_m)} = \gamma$$

*where* $\Delta_{m,l}(\omega,\lambda_1,\dots,\lambda_m)=\sqrt{\hat{V}_l(\omega)+\lambda_{l+1}\sqrt{\dots+\lambda_m\sqrt{\hat{V}_m(\omega)}}}$ *for* $1\le l\le m$, *for all* $k$, $1\le k\le \overline{p}(m)$, *and* $\gamma$ *a real constant.*

**Proof** It suffices to observe that

$$\frac{\partial\hat{L}^v_m}{d\lambda_k}=\frac{\lambda_1}{2\Delta_{m,1}}\dots\frac{\lambda_{k-1}}{2\Delta_{m,k-1}}\Delta_{m,k}=\frac{(\prod_{i=1}^{k-1}\lambda_i)\Delta_{m,k}}{2^{k-1}\prod_{i=1}^{k-1}\Delta_{m,i}}$$

and

$$\left(\frac{\partial\hat{L}^v_m}{d\lambda_k}=\frac{-2\gamma_1}{\lambda_k^3}\right)\longrightarrow\left(\frac{(\lambda_k)^2(\prod_{i=1}^{k}\lambda_i)\Delta_{m,k}}{2^k\prod_{i=1}^{k-1}\Delta_{m,i}}=-\gamma_1\right).$$

Note that the void product is equal to one. ■

For $2^{nd}$ order sure inference analysis we have the following closed form for optimal intervals.

**Theorem 3.3.** *The optimal $2^{nd}$ order standard deviation based at least probability $p$ confidence interval for $x\in\mathbb{R}$ is* $[X(\omega)-\hat{L}^*_2(\omega),X(\omega)+\hat{L}^*_2(\omega)]$ *where* $\hat{L}^*_2(\omega)=\lambda_1^*(\hat{\sigma}_1(\omega)+\lambda_2^*\hat{\sigma}_2(\omega))$, $\lambda_1^*=\sqrt{\frac{(\lambda_2^*)^2}{(1-p)(\lambda_2^*)^2-1}}$,

$$\lambda_2^*=\sqrt[3]{\frac{r}{2(1-p)}+\sqrt{\left(\frac{r}{2(1-p)}\right)^2-\left(\frac{2}{3(1-p)}\right)^3}}+\sqrt[3]{\frac{r}{2(1-p)}-\sqrt{\left(\frac{r}{2(1-p)}\right)^2-\left(\frac{2}{3(1-p)}\right)^3}}$$

*and* $r=\hat{\sigma}_1(\omega)/\hat{\sigma}_2(\omega)$ *for* $\hat{\sigma}_2(\omega)\neq0$. *If* $\hat{\sigma}_2(\omega)=0$ *then* $\hat{L}^*_2(\omega)=(1/\sqrt{1-p})\,\hat{\sigma}_1(\omega)$.

**Proof** By Theorem 3.1 for $\overline{p}(m)=2$ we have $\frac{1}{\lambda_1^2}+\frac{1}{\lambda_2^2}=1-p$, $\lambda_1^2\sum_{i=1}^{2}(\prod_{j=1}^{i}\lambda_j)\hat{\sigma}_i(\omega)=\gamma$ and $\lambda_2^2\sum_{i=2}^{2}(\prod_{j=1}^{i}\lambda_j)\hat{\sigma}_i(\omega)=\gamma$ which reduces to $\lambda_1^2=\frac{\lambda_2^2}{(1-p)\lambda_2^2-1}$ and $\lambda_1^2(\hat{\sigma}_1(\omega)+\lambda_2\hat{\sigma}_2(\omega))=\lambda_2^2(\lambda_2)\hat{\sigma}_2(\omega)$ thus $\hat{\sigma}_1(\omega)+\lambda_2\hat{\sigma}_2(\omega)=\lambda_2((1-p)\lambda_2^2-1)\hat{\sigma}_2(\omega)$ from which

$$\frac{\hat{\sigma}_1(\omega)}{\hat{\sigma}_2(\omega)}+2\lambda_2=(1-p)\lambda_2^3$$

$$\lambda_2^3-\frac{2\lambda_2}{1-p}-\frac{r}{1-p}=0.$$

We remind that the roots of the cubic equation $x^3+px+q=0$ are $x=\sqrt[3]{-q/2+\sqrt{(q/2)^2+(p/3)^3}}+\sqrt[3]{-q/2-\sqrt{(q/2)^2+(p/3)^3}}$. Thus

$$\lambda_2^* = \sqrt[3]{\frac{r}{2(1-p)}+\sqrt{\left(\frac{r}{2(1-p)}\right)^2-\left(\frac{2}{3(1-p)}\right)^3}}+\sqrt[3]{\frac{r}{2(1-p)}-\sqrt{\left(\frac{r}{2(1-p)}\right)^2-\left(\frac{2}{3(1-p)}\right)^3}}$$

If $\hat{\sigma}_2(\omega)=0$ then we have $\sum_{k=1}^{1}\frac{1}{(\lambda_1^*)^2}=1-p$ from which $\lambda_1^*=\frac{1}{\sqrt{1-p}}$. ∎

**Theorem 3.4.** *The optimal $2^{nd}$ order variance based at least probability $p$ confidence interval for $x\in\mathbb{R}$ is $[X(\omega)-\hat{L}_2^{v*}(\omega), X(\omega)+\hat{L}_2^{v*}(\omega)]$ where $\hat{L}_2^{v*}=\lambda_1^*\sqrt{\hat{V}_1(\omega)+\lambda_2^*\sqrt{\hat{V}_2(\omega)}}$, $\lambda_1^*=\sqrt{\frac{(\lambda_2^*)^2}{(1-p)(\lambda_2^*)^2-1}}$,*

$$\lambda_2^* = \sqrt[3]{\frac{r}{(1-p)}+\sqrt{\left(\frac{r}{(1-p)}\right)^2-\left(\frac{1}{(1-p)}\right)^3}}+\sqrt[3]{\frac{r}{(1-p)}-\sqrt{\left(\frac{r}{(1-p)}\right)^2-\left(\frac{1}{(1-p)}\right)^3}}$$

*and $r=\hat{V}_1(\omega)/\sqrt{\hat{V}_2(\omega)}$ for $\hat{V}_2(\omega)\neq 0$. If $\hat{V}_2(\omega)=0$ then $L_2^{v*}(\omega)=(1/\sqrt{1-p})\sqrt{\hat{V}_1(\omega)}$.*

**Proof** Immediate for $\overline{p}_{(m)}=1$ and for $\overline{p}_{(m)}=2$ we have $\frac{1}{\lambda_1^2}+\frac{1}{\lambda_2^2}=1-p$ and

$$\frac{\lambda_1^3\sqrt{\hat{V}_1(\omega)+\lambda_2\sqrt{\hat{V}_2(\omega)}}}{(2)(1)}=\frac{\lambda_2^2(\lambda_1\lambda_2)\sqrt{\hat{V}_2(\omega)}}{2^2\sqrt{\hat{V}_1(\omega)+\lambda_2\sqrt{\hat{V}_2(\omega)}}}$$

that reduces to $\lambda_1^2=\frac{\lambda_2^2}{(1-p)\lambda_2^2-1}$ and $2\lambda_1^2(\hat{V}_1(\omega)+\lambda_2\sqrt{\hat{V}_2(\omega)})=\lambda_2^3\sqrt{\hat{V}_2(\omega)}$. Thus

$$\begin{aligned}2\lambda_2^2(\hat{V}_1(\omega)+\lambda_2\sqrt{\hat{V}_2(\omega)}) &= ((1-p)\lambda_2^2-1)\lambda_2^3\sqrt{\hat{V}_2(\omega)}\\ 2(r+\lambda_2) &= \lambda_2((1-p)\lambda_2^2-1)\\ \lambda_2^3-\frac{3\lambda_2}{(1-p)}-\frac{2r}{(1-p)} &= 0\end{aligned}$$

and the result follows. ∎

Observe that we can define a function $\Lambda^*:[0,1]\times\Omega\to(\mathbb{R}_+^*)^m$ where $\Lambda(p,\omega)$ is the vector, in some situations one of the possible vectors, $\Lambda^*(p,\omega)\in(\mathbb{R}_+^*)^m$ that makes the interval $[X(\omega)-\hat{L}_m(\omega,\Lambda^*(p,\omega)), X(\omega)+\hat{L}_m(\omega,\Lambda^*(p,\omega))]$ have minimum length at probability level $p$. Analogously we define $\Lambda^{v*}$.

Now, we turn our attention to sure inference bands. Define now $\hat{L}_m(\omega,t,\lambda_1,\ldots,\lambda_m)=\sum_{i=1}^{m}(\prod_{j=1}^{i}\lambda_j)\hat{\sigma}_i(\omega,t)$ and $\hat{L}_m^v(\omega,t,\lambda_1,\ldots,\lambda_m)=\lambda_1\sqrt{\hat{V}_1(\omega,t)+\lambda_2\sqrt{\cdots+\lambda_m\sqrt{\hat{V}_m(\omega,t)}}}$.

**Definition 20.** *An optimal m-th order at least probability p band for $x : I \to \mathbb{R}$ is a band $\{[X(\omega,t)-\hat{L}_m(\omega,t,\lambda_1,\dots,\lambda_m), X(\omega,t)+\hat{L}_m(\omega,t,\lambda_1,\dots,\lambda_m)] | t \in I\}$ or $\{[X(\omega,t)-\hat{L}^v_m(\omega,t,\lambda_1,\dots,\lambda_m)$ $X(\omega,t)+\hat{L}^v_m(\omega,t,\lambda_1,\dots,\lambda_m)] | t \in I\}$ such that for all $t \in I$ the corresponding intervals are optimal at least probability p intervals for $x(t)$.*

**Theorem 3.5.** *Let for each $t \in I$, $\overline{p}(t,m)$ be the greatest integer less than or equal to m such that $\hat{\sigma}_{\overline{p}(t,m)}(\omega,t) > 0$ and for all j, $\overline{p}(t,m) < j \le m, \hat{\sigma}_j(\omega,t) = 0$.*

*The optimal m-th order standard deviation based at least probability p ($p \neq 1$) confidence band for $x : I \to \mathbb{R}$ is written as $\{[X(\omega,t) - \hat{L}_m(\omega,t,\lambda_1^*,\dots,\lambda_m^*), X(\omega,t) + \hat{L}_m(\omega,t,\lambda_1^*,\dots,\lambda_m^*)] | t \in I\}$ where $(\lambda_1^*(t),\dots,\lambda_m^*(t)) \in (\mathbb{R}_+^*)^m$ satisfies, for all t, $\forall j$, $\overline{p}(t,m) < j \le m$, $\lambda_j^*(t)$ is an arbitrary positive real number and $(\lambda_1^*(t),\dots,\lambda^*_{\overline{p}(t,m)}(t))$ is within the solutions of the simultaneous systems of $\overline{p}(t,m)+1$ equations:*

$$\sum_{k=1}^{\overline{p}(t,m)} \frac{1}{(\lambda_k(t))^2} = 1-p \text{ and } (\lambda_k(t))^2 (\sum_{i=k}^{\overline{p}(m)} (\prod_{j=1}^{i} \lambda_j(t)) \hat{\sigma}_i(\omega,t)) = \gamma(t)$$

*for all k, $1 \le k \le \overline{p}(t,m)$, and $\gamma(t)$ a real constant for each t.*

**Proof** Direct application of Theorem 3.1 for each $t \in I$. ∎

**Theorem 3.6.** *Let for each $t \in I$, $\overline{p}(t,m)$ be the greatest integer less than or equal to m such that $\hat{V}_{\overline{p}(t,m)}(\omega,t) > 0$ and for all j, $\overline{p}(t,m) < j \le m, \hat{V}_j(\omega,t) = 0$.*

*The optimal m-th order variance based at least probability p ($p \neq 1$) confidence band for $x : I \to \mathbb{R}$ is written as $\{[X(\omega,t) - \hat{L}^v_m(\omega,t,\lambda_1^*,\dots,\lambda_m^*), X(\omega,t) + \hat{L}^v_m(\omega,t,\lambda_1^*,\dots,\lambda_m^*)] | t \in I\}$ where $(\lambda_1^*(t),\dots,\lambda_m^*(t)) \in (\mathbb{R}_+^*)^m$ satisfies, for all t, $\forall j$, $\overline{p}(t,m) < j \le m$, $\lambda_j^*(t)$ is an arbitrary positive real number and $(\lambda_1^*(t),\dots,\lambda^*_{\overline{p}(t,m)}(t))$ is within the solutions of the simultaneous systems of $\overline{p}(t,m)+1$ equations:*

$$\sum_{k=1}^{\overline{p}(t,m)} \frac{1}{(\lambda_k(t))^2} = 1-p \text{ and } \frac{(\lambda_k(t))^2 (\prod_{i=1}^{k} \lambda_i(t)) \Delta_{m,k}(\omega,t,\lambda_1(t),\dots,\lambda_m(t))}{2^k \prod_{i=1}^{k-1} \Delta_{m,i}(\omega,t,\lambda_1(t),\dots,\lambda_m(t))} = \gamma(t)$$

*where $\Delta_{m,l}(\omega,t,\lambda_1(t),\dots,\lambda_m(t)) = \sqrt{\hat{V}_l(\omega,t) + \lambda_{l+1}(t)\sqrt{\dots + \lambda_m(t)\sqrt{\hat{V}_m(\omega,t)}}}$ for $1 \le l \le m$, for all k, $1 \le k \le \overline{p}(t,m)$, and $\gamma(t)$ a real constant for each t.*

**Proof** Direct application of Theorem 3.2 for each $t \in I$. ∎

**Theorem 3.7.** *The optimal $2^{nd}$ order standard deviation based at least probability p confidence band for $x : I \to \mathbb{R}$ is $\{[X(\omega,t)-\hat{L}_2^*(\omega,t), X(\omega,t)+\hat{L}_2^*(\omega,t)] | t \in I\}$ where $\hat{L}_2^*(\omega,t) = \lambda_1^*(t)(\hat{\sigma}_1(\omega,t) + \lambda_2^*(t)\hat{\sigma}_2^*(\omega,t))$, $\lambda_1^*(t) = \sqrt{\frac{(\lambda_2^*(t))^2}{(1-p)(\lambda_2^*(t))^2-1}}$,*

$$\lambda_2^*(t) = \sqrt[3]{\frac{r(t)}{2(1-p)} + \sqrt{\left(\frac{r(t)}{2(1-p)}\right)^2 - \left(\frac{2}{3(1-p)}\right)^3}} + \sqrt[3]{\frac{r(t)}{2(1-p)} - \sqrt{\left(\frac{r(t)}{2(1-p)}\right)^2 - \left(\frac{2}{3(1-p)}\right)^3}}$$

*and* $r(t) = \hat{\sigma}_1(\omega,t)/\hat{\sigma}_2(\omega,t)$ *for* $\hat{\sigma}_2(\omega,t) \neq 0$. *If* $\hat{\sigma}_2(\omega,t) = 0$ *then* $L_2^*(\omega,t) = (1/\sqrt{1-p})\hat{\sigma}_1(\omega,t)$.

**Proof** Consequence of Theorems 3.3 and 3.5. ▪

**Theorem 3.8.** *The optimal* $2^{nd}$ *order variance based at least probability p confidence band for* $x : I \to \mathbb{R}$ *is* $\{[X(\omega,t)-\hat{L}_2^{v*}(\omega,t), X(\omega,t)+\hat{L}_2^{v*}(\omega,t)|t \in I\}]$ *where* $\hat{L}_2^{v*}(\omega.t) = \lambda_1^*(t)\sqrt{\hat{V}_1(\omega,t) + \lambda_2^*(t)\sqrt{\hat{V}_2(\omega,t)}}$,

$\lambda_1^*(t) = \sqrt{\frac{(\lambda_2^*(t))^2}{(1-p)(\lambda_2^*(t))^2-1}}$,

$$\lambda_2^*(t) = \sqrt[3]{\frac{r(t)}{(1-p)} + \sqrt{\left(\frac{r(t)}{(1-p)}\right)^2 - \left(\frac{1}{(1-p)}\right)^3}} + \sqrt[3]{\frac{r(t)}{(1-p)} - \sqrt{\left(\frac{r(t)}{(1-p)}\right)^2 - \left(\frac{1}{(1-p)}\right)^3}}$$

*and* $r(t) = \hat{V}_1(\omega,t)/\sqrt{\hat{V}_2(\omega,t)}$ *for* $\hat{V}_2(\omega,t) \neq 0$. *If* $\hat{V}_2(\omega,t) = 0$ *then* $\hat{L}_2^{v*}(\omega,t) = (1/\sqrt{1-p})\sqrt{\hat{V}_1(\omega,t)}$.

**Proof** Consequence of Theorems 3.4 and 3.6. ▪

## 4. Optimal intervals and bands for vectors and vector valued functions

In this section we are interested in finding optimal $\mathbb{R}^q$-intervals for vector values and optimal bands for $\mathbb{R}^q$-valued functions. In the previous case, that of $\mathbb{R}^1$, there was no other sensible way of defining an optimal interval but claiming that it must have minimum length and extend this property, in a point-wise manner to bands. Note that we are only dealing with symmetrical intervals. Now for $\mathbb{R}^q$, $q \geq 2$ there is freedom in choosing optimal criteria. For example we could say that an $\mathbb{R}^q$-interval is optimal if its volume (Lebesgue measure) is minimal, or if the sum of its edges is minimum, or choose weights for each coordinate interval to form an "weighted volume" to be minimized. This situation may happen in practice since it is possible that some of the coordinates are more important than others and demand a better resolution on their values.

Let $\hat{L}_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m}) = \sum_{i=1}^{m}(\prod_{l=1}^{i}\lambda_{j,l})\hat{\sigma}_{j,i}(\omega)$ and

$$\hat{L}_{j,m}^v = \lambda_{j,1}\sqrt{\hat{V}_{j,1}(\omega) + \lambda_{j,2}\sqrt{\cdots + \lambda_{j.m}\sqrt{\hat{V}_{j,m}(\omega)}}}.$$

**Definition 21.** *Let* $M : \mathbb{R}^q \to \mathbb{R}_+$ *be a function. An m-th order at least probability p confidence interval for* $x \in \mathbb{R}^q$, $\prod_{j=1}^{q}[X_j(\omega) - \hat{L}_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j.m}), X_j(\omega) + \hat{L}_{j,m}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m})]$ *(or* $\prod_{j=1}^{q}[X_j(\omega) - \hat{L}_{j,m}^v(\omega,\lambda_{j,1},\ldots,\lambda_{j,m}), X_j(\omega) + \hat{L}_{j,m}^v(\omega,\lambda_{j,1},\ldots,\lambda_{j,m})]$*) is said to be M-optimal if and only if* $M(\hat{L}_{1,m}(\omega,\lambda_{1,1},\ldots,\lambda_{1,m}),\ldots,\hat{L}_{q,m}(\omega,\lambda_{q,1},\ldots,\lambda_{q,m}))$ *(or* $M(\hat{L}_{1,m}^v(\omega,\lambda_{1,1},\ldots,\lambda_{1,m}),\ldots,\hat{L}_{q,m}^v(\omega,\lambda_{q,1},\ldots,\lambda_{q,m}))$*) is minimum.*

Clearly we are interested in functions $M$ that are non-decreasing in all of its variables.

It is important to note that the concept of $M$-optimality is quite general. For example, let $l = (l_1,\ldots,l_q) \in \mathbb{R}^q$, $L = (L_1,\ldots,L_q) \in (\mathbb{R}_+)^q$ and $\lambda_j = (\lambda_{j,1},\ldots,\lambda_{j,m}) \in (\mathbb{R}_+^*)^m$, $1 \leq j \leq q$, $\lambda = (\lambda_1,\ldots,\lambda_q)$. If $\{\mu_l : \mathcal{B}_{\mathbb{R}^q} \to \mathbb{R}_+ | l \in \mathbb{R}^q\}$ is a class of Borel measures then minimizing $F$ :

$\mathbb{R}^q \times (\mathbb{R}_+)^q \to \mathbb{R}_+$ given by $F(l, L) = \mu_l(\prod_{j=1}^q [l_j - L_j, l_j + L_j])$, for $l = X(\omega)$ and $L_j = \hat{L}_{j,m}(\omega, \lambda_j)$ (or $L_j = \hat{L}^v_{j,m}(\omega, \lambda_j)$), $1 \le j \le q$, with respect to $\lambda$ is the same as minimizing $M : (\mathbb{R}^q_+) \to \mathbb{R}_+$ given by $M(\lambda_1, \ldots, \lambda_q) = F(X(\omega), L)$ with respect to $\lambda$. That is, for optimal criteria as general as choosing Borel measures that may depend on the point $X(\omega)$ for measuring the $\mathbb{R}^q$ symmetrical intervals $\prod_{j=1}^q [X_j(\omega) - L_j(\omega, \lambda_j), X_j(\omega) + L_j(\omega, \lambda_j)]$, we can choose an appropriate function $M$ to accomplish the same minimization work.

Let $(X, \sigma_n, \hat{\sigma}_n)$, and $(X, V_n, \hat{V}_n)$ be inferential sequences for $x'$.

**Theorem 4.1.** *Let $M : (\mathbb{R}_+)^q \to \mathbb{R}_+$ be a $C^1$ function which is non-decreasing in all of its variables. Define, for all $j$, $\overline{p}_{(j,m)}$ as the greatest integer less than or equal to $m$ such that $\hat{\sigma}_{j,\overline{p}_{(j,m)}}(\omega) > 0$ and for all $k$, $\overline{p}_{(j,m)} < k \le m, \hat{\sigma}_{j,k}(\omega) = 0$.*

*The $M$-optimal $m$-th order standard deviation based at least probability $p$, $p \neq 1$, confidence interval for $x' \in \mathbb{R}^q$ is written as $\prod_{j=1}^q [X_j(\omega) - \hat{L}_{j,m}(\omega, \lambda^*_{j,1}, \ldots, \lambda^*_{j,m}), X_j(\omega) + \hat{L}_{j,m}(\omega, \lambda^*_{j,1}, \ldots, \lambda^*_{j,m})]$ where $(\lambda^*_{j,1}, \ldots, \lambda^*_{j,m}) \in (\mathbb{R}^*_+)^m$ satisfies, for all $j$, $1 \le j \le q$:*

*For all $k$ such that $\overline{p}_{(j,m)} < k \le m$, $\lambda^*_{j,m}$ is an arbitrary positive real number and $(\lambda^*_{j,1}, \ldots, \lambda^*_{j,\overline{p}_{(j,m)}})$ is within the solutions of the simultaneous system of $(\sum_{j=1}^q \overline{p}_{(j,m)}) + 1$ equations:*

$$\sum_{j=1}^{q} \sum_{l=1}^{\overline{p}_{(j,m)}} \frac{1}{(\lambda_{j,l})^2} = 1 - p$$

*and*

$$(\lambda_{j,l})^2 \Big( \sum_{i=l}^{\overline{p}_{(j,m)}} \big( \prod_{k=1}^{i} \lambda_{j,k} \big) \hat{\sigma}_{j,i}(\omega) \Big) \frac{\partial M}{\partial x_j}(x) = \gamma,$$

*$x = (x_1, \ldots, x_q)$ $(x_j = \hat{L}_{j,m}(\omega, \lambda^*_{j,1}, \ldots, \lambda^*_{j,m}))$ for all $j$ and $l$, $1 \le j \le q$, $1 \le l \le \overline{p}_{(j,m)}$ and $\gamma$ a real constant.*

**Proof** Follow the steps of Theorem's 3.1 demonstration. Let

$$\hat{L}_{[q,m]} = (\hat{L}_{1,m}, \ldots, \hat{L}_{q,m}) : \Omega \times (\mathbb{R}^*_+)^{mq} \to (\mathbb{R}^*_+)^q,$$

$$\hat{L}_{[q,m]}(\omega, \lambda_{1,1}, \ldots, \lambda_{1,m}, \ldots, \lambda_{q,1}, \ldots, \lambda_{q,m}) =$$

$$(\hat{L}_{1,m}(\omega, \lambda_{1,1}, \ldots, \lambda_{1,m}), \ldots, \hat{L}_{q,m}(\omega, \lambda_{q,1}, \ldots, \lambda_{q,m})).$$

Note that $M \circ L_{[q,m]}(\omega, \cdot)$ is non-decreasing in all of its variables and that the continuity of the implicit function argument is still valid since $M$ is a $C^1$ function. Let $g(\overline{\lambda}) = \sum_{j=1}^q \sum_{i=1}^m \frac{1}{(\lambda_{j,i})^2}$, $\overline{\lambda} = (\lambda_{1,1}, \ldots, \lambda_{q,m}) \in (\mathbb{R}^*_+)^{mq}$.

Now, Lagrange's equation is written $d(M \circ \hat{L}_{[q,m]}(\omega, \cdot)) = \gamma_1 dg$. Thus $\sum_{k=1}^q \frac{\partial M}{\partial x_k} \frac{\partial \hat{L}_{k,m}}{\partial \lambda_{j,l}} = \gamma_1 \frac{(-2)}{(\lambda_{j,l})^3}$ that reduces to $(\lambda_{j,l})^3 \frac{\partial M}{\partial x_j} \frac{\partial \hat{L}_{j,m}}{\partial \lambda_{j,l}} = -2\gamma_1 = \gamma$ since $\hat{L}_{k,m}$ does not depend on $\lambda_{j,l}$ for $j \neq k$.

$$\hat{L}_{j,m}(\omega,\lambda_{j,1},\ldots\lambda_{j,m})=\sum_{i=1}^{m}(\prod_{k=1}^{i}\lambda_{j,k})\hat{\sigma}_{j,i}(\omega)=\sum_{i=1}^{\overline{p}_{(j,m)}}(\prod_{k=1}^{i}\lambda_{j,k})\hat{\sigma}_{j,i}(\omega)\text{ and}$$

$$\frac{\partial\hat{L}_{j,m}}{\partial\lambda_{j,l}}=\sum_{i=l}^{\overline{p}_{(j,m)}}(\prod_{\substack{k=1\\k\neq l}}^{i}\lambda_{j,k})\hat{\sigma}_{j,i}(\omega)$$

from which

$$(\lambda_{j,l})^2(\sum_{i=l}^{\overline{p}_{(j,m)}}(\prod_{k=1}^{i}\lambda_{j,k})\hat{\sigma}_{j,i}(\omega))\frac{\partial M}{\partial x_j}|_{x=\hat{L}_{(q,m)}(\omega,\lambda_{1,1},\ldots,\lambda_{1,m},\ldots,\lambda_{q,1},\ldots,\lambda_{q,m})}=\gamma.$$

∎

**Theorem 4.2.** *Let $M:(\mathbb{R}_+)^q\to\mathbb{R}_+$ be a $C^1$ function which is non-decreasing in all of its variables. Define, for all $j$, $\overline{p}_{(j,m)}$ as the greatest integer less than or equal to $m$ such that $\hat{V}_{j,\overline{p}_{(j,m)}}(\omega)>0$ and for all $k$, $\overline{p}_{(j,m)}<k\leq m$, $\hat{V}_{j,k}(\omega)=0$.*

*The $M$-optimal $m$-th order variance based at least probability $p$, $p\neq 1$, confidence interval for $x\in\mathbb{R}^q$ is written as $\prod_{j=1}^{q}[X_j(\omega)-\hat{L}^v_{j,m}(\omega,\lambda^*_{j,1},\ldots,\lambda^*_{j,m}),X_j(\omega)+\hat{L}^v_{j,m}(\omega,\lambda^*_{j,1},\ldots,\lambda^*_{j,m})]$ where $(\lambda^*_{j,1},\ldots,\lambda^*_{j,m})\in(\mathbb{R}^*_+)^m$ satisfies, for all $j$, $1\leq j\leq q$:*

*For all $k$ such that $\overline{p}_{(j,m)}<k\leq m$, $\lambda^*_{j,m}$ is an arbitrary positive real number and $(\lambda^*_{j,1},\ldots,\lambda^*_{j,\overline{p}_{(j,m)}})$ is within the solutions of the simultaneous system of $(\sum_{j=1}^{q}\overline{p}_{(j,m)})+1$ equations:*

$$\sum_{j=1}^{q}\sum_{l=1}^{\overline{p}_{(j,m)}}\frac{1}{(\lambda_{j,l})^2}=1-p\text{ and}$$

$$(\lambda_{j,l})^2\frac{(\prod_{i=1}^{l}\lambda_i)\Delta_{j,m,l}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m})}{2^l\prod_{i=1}^{l-1}\Delta_{j,m,i}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m})}\frac{\partial M}{\partial x_j}(x)=\gamma$$

*where $\Delta_{j,m,l}(\omega,\lambda_{j,1},\ldots,\lambda_{j,m})=\lambda_{j,l}\sqrt{\hat{V}_{j,l}(\omega)+\lambda_{j,l+1}\sqrt{\cdots+\lambda_{j,m}\sqrt{\hat{V}_{j,m}(\omega)}}}$, $x=(x_1,\ldots,x_q)$, $x_j=\hat{L}^v_{j,m}(\omega,\lambda^*_{j,1},\ldots,\lambda^*_{j,m})$ for all $j$ and $l$, $1\leq j\leq q$, $1\leq l\leq\overline{p}_{(j,m)}$ and $\gamma$ a real constant.*

**Proof** Analogous to Theorem's 4.1 proof. ∎

If the criterium of minimum is that of minimum volume of confidence intervals them we can use $M=\prod_{j=1}^{q}x_j$. From practical instances, it may occur that some of the variables are more important than others. In this situation we can use weighted volumes and define $M = \prod_{j=1}^{q} x_j^{\alpha_j}$, for $\alpha_j > 0$ real constants that indicate the weights associated to each magnitude.

**Theorem 4.3.** *If $M = \prod_{j=1}^{q} x_j^{\alpha_j}$, the $M$-optimal $m$-th order standard deviation based at least probability $p$, $p \neq 1$, confidence interval may be obtained from the solutions of the system:*

$$\sum_{j=1}^{q} \sum_{l=1}^{\overline{p}_{(j,m)}} \frac{1}{(\lambda_{j,1})^2} = 1 - p \text{ and}$$

$$\frac{\alpha_j (\lambda_{j,l})^2 \left( \sum_{i=l}^{\overline{p}_{(j,m)}} \left( \prod_{k=1}^{i} \lambda_{j,k} \right) \hat{\sigma}_{j,i}(\omega) \right) \prod_{n=1}^{q} (\hat{L}_{n,m}(\omega, \lambda_{n,1}, \ldots, \lambda_{n,m}))^{\alpha_n}}{\hat{L}_{j,m}(\omega, \lambda_{j,1}, \ldots, \lambda_{j,m})} = \gamma$$

*for all $j$ and $l$, $1 \leq j \leq q$, $1 \leq l \leq \overline{p}_{(j,m)}$ and $\gamma$ a real constant.*

**Proof** Direct application of Theorem 4.1 for $M = \prod_{j=1}^{q} x_j^{\alpha_j}$

$$\frac{\partial M}{\partial x_j} = \frac{\partial}{\partial x_j} \prod_{n=1}^{q} x_n^{\alpha_n} = \frac{\alpha_j (\prod_{n=1}^{q} x_n^{\alpha_n})}{x_j} = \frac{\alpha_j}{x_j} M.$$

∎

**Theorem 4.4.** *If the $M = \prod_{j=1}^{q} x_j^{\alpha_j}$, the $M$-optimal $m$-th order variance based at least $p$, $p \neq 1$, confidence interval may be obtained from the solutions of the system:*

$$\sum_{j=1}^{q} \sum_{l=1}^{\overline{p}_{(j,m)}} \frac{1}{(\lambda_{j,l})^2} = 1 - p \text{ and}$$

$$\frac{\alpha_j (\lambda_{j,l})^2 (\prod_{i=1}^{l} \lambda_i) \Delta_{j,m,l}(\omega, \lambda_{j,1}, \ldots, \lambda_{j,m}) \prod_{n=1}^{q} (\hat{L}^v_{n,m}(\omega, \lambda_{n,1}, \ldots, \lambda_{n,m}))^{\alpha_n}}{2^l (\prod_{i=1}^{l-1} \Delta_{j,m,i}(\omega, \lambda_{j,1}, \ldots, \lambda_{j,m})) \hat{L}^v_{j,m}(\omega, \lambda_{j,1}, \ldots, \lambda_{j,m})} = \gamma$$

*for all $j$ and $l$, $1 \leq j \leq q$, $1 \leq l \leq \overline{p}_{(j,m)}$ and $\gamma$ a real constant.*

**Proof** Direct application of Theorem 4.2. ∎

Lets turn our attention to optimal bands. Note that now we have much more freedom for our choice of optimum criteria, since for each point $t \in I$ we can choose an $M$ function that will define an optimal interval associated to that particular point.

**Definition 22.** *Let $M : I\times \mathbb{R}^q \to \mathbb{R}_+$ be a function. An m-th order at least probability p confidence band for $x : I \to \mathbb{R}^q$,*

$$\{\prod_{j=1}^{q}[X(\omega,t)-\hat{L}_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m}),X(\omega,t)+\hat{L}_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m})]|t\in I\}$$

$$(or\ \{\prod_{j=1}^{q}[X(\omega,t)-\hat{L}^{v}_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m}),X(\omega,t)+\hat{L}^{v}_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m})]|t\in I\})$$

*is M-optimal if for each and every $t\in I$, the interval*

$$\prod_{j=1}^{q}[X(\omega,t)-\hat{L}_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m}),X(\omega,t)+\hat{L}_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m})]$$

$$(or\ \prod_{j=1}^{q}[X(\omega,t)-\hat{L}^{v}_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m}),X(\omega,t)+\hat{L}^{v}_{j,m}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m})])$$

*is $M(t)$-optimal.*

Clearly we are interested in functions $M : I\times\mathbb{R}^q \to \mathbb{R}_+$ such that for all $t\in I$ $M(t,\cdot)$ is non-decreasing in all of its variables.

**Theorem 4.5.** *Let for all $t\in I$, $M(t,\cdot):(\mathbb{R}_+)^q\to\mathbb{R}_+$ be a $C^1$ function which non-decreasing in all of its variables. Let for all $t\in I$ and for all $j$, $\overline{p}_{(t,j,m)}$ be the greatest integer less than or equal to m such that $\hat{\sigma}_{j,\overline{p}_{(t,j,m)}}(\omega,t)>0$ and for all $k$, $\overline{p}_{(t,j,m)}<k\le m$, $\hat{\sigma}_{j,k}(\omega,t)=0$.*

*The M-optimal m-th order standard deviation based at least probability p, $p\neq 1$, confidence band for $x : I\to\mathbb{R}^q$ is written as $\{\prod_{j=1}^{q}[X_j(\omega,t)-\hat{L}_{j,m}(\omega,t,\lambda^*_{j,1},\dots,\lambda^*_{j,m}),X_j(\omega,t)+\hat{L}_{j,m}(\omega,t,\lambda^*_{j,1},\dots,\lambda^*_{j,m})]\mid t\in I\}$ where $(\lambda^*_{j,1},\dots,\lambda^*_{j,m})=(\lambda^*_{j,1}(t),\dots,\lambda^*_{j,m}(t))\in(\mathbb{R}^*_+)^m$ satisfies, for all $t$ and for all $j$, $1\le j\le q$:*

*For all k such that $\overline{p}_{(t,j,m)}<k\le m$, $\lambda^*_{j,m}(t)$ is an arbitrary positive real number and $(\lambda^*_{j,1}(t),\dots,\lambda^*_{j,\overline{p}_{(t,j,m)}}(t))$ is within the solutions of the simultaneous system of $(\sum_{j=1}^{q}\overline{p}_{(t,j,m)})+1$ equations:*

$$\sum_{j=1}^{q}\sum_{l=1}^{\overline{p}_{(t,j,m)}}\frac{1}{(\lambda_{j,l}(t))^2}=1-p$$

*and*

$$(\lambda_{j,l}(t))^2\,(\sum_{i=l}^{\overline{p}_{(t,j,m)}}(\prod_{k=1}^{i}\lambda_{j,k}(t))\hat{\sigma}_{j,i}(\omega,t))\frac{\partial M}{\partial x_j}(t,x)=\gamma(t),$$

*$x=(x_1,\dots,x_q)$, $x_j=\hat{L}_{j,m}(\omega,\lambda^*_{j,1},\dots,\lambda^*_{j,m})$ for all $j$ and $l$, $1\le j\le q$, $1\le l\le\overline{p}_{(t,j,m)}$ and $\gamma(t)$ a real constant for each $t$.*

**Proof** Apply Theorem 4.1 for each $t\in I$. ∎

**Theorem 4.6.** *Let for all $t \in I$, $M(t,\cdot) : (\mathbb{R}_+)^q \to \mathbb{R}_+$ be a $C^1$ function which is non-decreasing in all of its variables. Let for all $t \in I$ and for all $j$, $\overline{p}_{(t,j,m)}$ be the greatest integer less than or equal to $m$ such that $\hat{V}_{j,\overline{p}_{(t,j,m)}}(\omega,t) > 0$ and for all $k$, $\overline{p}_{(t,j,m)} < k \le m$, $\hat{V}_{j,k}(\omega,t) = 0$.*

*The $M$-optimal $m$-th order variance based at least probability $p$, $p \neq 1$, confidence band for $x : I \to \mathbb{R}^q$ is written as $\{\prod_{j=1}^{q}[X_j(\omega,t) - \hat{L}^v_{j,m}(\omega,t,\lambda^*_{j,1},\dots,\lambda^*_{j,m}), X_j(\omega,t) + \hat{L}^v_{j,m}(\omega,t,\lambda^*_{j,1},\dots,\lambda^*_{j,m})]$ where $(\lambda^*_{j,1},\dots,\lambda^*_{j,m}) = (\lambda^*_{j,1}(t),\dots,\lambda^*_{j,m}(t)) \in (\mathbb{R}^*_+)^m$ satisfies, for all $t$ and for all $j$, $1 \le j \le q$:*

*For all $k$ such that $\overline{p}_{(t,j,m)} < k \le m$, $\lambda^*_{j,m}(t)$ is an arbitrary positive real number and $(\lambda^*_{j,1}(t),\dots,$ $\lambda^*_{j,\overline{p}_{(t,j,m)}}(t))$ is within the solutions of the simultaneous system of $(\sum_{j=1}^{q}\overline{p}_{(t,j,m)}) + 1$ equations:*

$$\sum_{j=1}^{q}\sum_{l=1}^{\overline{p}_{(t,j,m)}}\frac{1}{(\lambda_{j,l}(t))^2} = 1 - p \text{ and}$$

$$(\lambda_{j,l}(t))^2\frac{(\prod_{i=1}^{l}\lambda\Delta_{j,m,i}(\omega,\lambda_{j,1},\dots,\lambda_{j,m})}{2^l\prod_{i=1}^{l-1}\Delta_{j,m,i}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m})}\frac{\partial M}{\partial x_j}(t,x) = \gamma(t)$$

*where $\Delta_{j,m,l}(\omega,t,\lambda_{j,1},\dots,\lambda_{j,m}) = \lambda_{j,l}(t)\sqrt{\hat{V}_{j,l}(\omega,t) + \lambda_{j,l+1}(t)\sqrt{\dots + \lambda_{j,m}(t)\sqrt{\hat{V}_{j,m}(\omega,t)}}}$, $x = (x_1,\dots,x_q)$, $x_j = \hat{L}^v_{j,m}(\omega,t,\lambda^*_{j,1},\dots,\lambda^*_{j,m})$ for all $j$ and $l$, $1 \le j \le q$, $1 \le l \le \overline{p}_{(t,j,m)}$ and $\gamma(t)$ a real constant.*

**Proof** Apply Theorem 4.2 for each $t \in I$. ∎

We observe that we could have made the at least probability $p$ confidence level vary over $I$. Taking $p : I \to [0,1)$ a function that assigns to each point $t$ in $I$ an at least probability level $p(t)$ we can, substituting $p(t)$ for $p$, rewrite Theorems 4.5 and 5.6 and obtain two corresponding new theorems for the case where the at least probability level may not be constant on the set $I$. This new statements are true because everything that relates to $I$ is done in a pointwise manner.

## 5. An example

In this section we will construct an inferential sequence for the case where the unbiased estimator $X$ of $x$ can be written as an stochastic integral with respect to a special type of random measure.

Let $\xi$ be a non-negative random measure with state space $\mathcal{X}$, a complete separable metric space. Let also $D$ be the diagonal set $\{(x,x)|x \in \mathcal{X}\} \subset \mathcal{X}^2$ and $\pi_1$ the first projection $\pi_1 : \mathcal{X}^2 \to \mathcal{X}$, $\pi_1(x,y) = x$.

**Definition 23.** *A random measure $\xi$ is said to satisfy assumption $\mathcal{A}$ if and only if $E(\xi\times\xi)(A\cap D) = E\xi\pi_1(A\cap D)$ for all $A \in \mathcal{B}_{\mathcal{X}^2}$ Borel set of $\mathcal{X}^2$.*

**Definition 24.** *A random measure $\xi$ is said to satisfy assumption $\mathcal{B}$ if and only if $(E\xi\times E\xi)(D) = 0$.*

**Definition 25.** *A random measure $\xi$ is said to satisfy assumption $C$ if and only if $\forall A, B \in \mathcal{B}_{\mathcal{X}}$ $(A \cap B = \emptyset \to Cov(\xi(A), \xi(B)) = 0)$.*

**Theorem 5.1.** *Let $A \in \mathcal{B}_{\mathcal{X}}$ be a bounded Borel set and $f : \mathcal{X} \to \mathbb{R}$ be Borel measurable and bounded over bounded sets. Suppose that $\xi$ satisfies assumptions $\mathcal{A}, \mathcal{B}$ and $C$ and that $X = \int_A f d\xi : \Omega \to \mathbb{R}$ is an unbiased estimator for $x$. Then $(X, V_n, \hat{V}_n)$, where $\hat{V}_n = \int_A f^{2^n} d\xi$ and $V_n = E \int_A f^{2^n} d\xi$, is an inferential sequence of random variables for $x$.*

**Proof**

(i) $EX = x$ by hypothesis.

$$\begin{aligned}
\mathrm{Var}\,(X) &= E(\int_A f d\xi)^2 - (E\int_A f d\xi)^2 = E(\int_{A^2} f \otimes f d\xi \times d\xi) - (\int_A f E d\xi)^2 \\
&= \int_{A^2} f \otimes f E(d\xi \times d\xi) - \int_{A^2} f \otimes f (Ed\xi) \times (Ed\xi) \\
&= \int_{A^2 - D} f \otimes f \mathrm{Cov}(d\xi, d\xi) + \int_{A^2 \cap D} f \otimes f E(d\xi \times d\xi) - \int_{A^2 \cap D} f \otimes f (Ed\xi) \times (Ed\xi) \\
&= 0 + \int_A f^2 E d\xi + 0 = E\hat{V}_1 = V_1.
\end{aligned}$$

(ii) Substituting $f^{2^n}$ for $f$ in item's (i) argument we have $\mathrm{Var}\,\hat{V}_n = E(\int_A f^{2^n} d\xi)^2 - (E\int_A f^{2^n} d\xi)^2 = \int_A (f^{2^n})^2 E d\xi$. Thus $\mathrm{Var}\,\hat{V}_n = \int_A f^{2^{n+1}} E d\xi = V_{n+1}$.

(iii) $E\hat{V}_n = V_n$ by construction.

(vi) $\forall \omega \in \Omega\ \hat{V}_n(\omega) = \int_A f^{2^n} d\xi(\omega) \geq 0$ since $f^{2^n}(x) \geq 0$ for all $x \in \mathcal{X}$.

∎

Some estimators used in point processes intensity function estimation are special cases of the one presented above. For a detailed example of inferential sequence of random variables and inferential sequence of stochastic processes for the estimation of real values and functions accompanied by the presentation of some inference bands, see de Miranda and Morettin (2003).

## 6. Sure inference with extra information – Applications to small sample confidence intervals for the mean

The examples that follow illustrate the use of sure inference in situations where we have some information on the random variables distribution.

In Theorems 6.1 and 6.2 we construct variance based and standard deviation based inferential sequences for the mean of non-negative random variables.

**Theorem 6.1.** *Let $Y : \Omega_1 \to \mathbb{R}_+$ be a non negative random variable. Define $\nu \in \mathbb{R}$ by $\mathrm{Var}\,Y = \nu EY$; $\mu = EY$, and suppose $y_1, \ldots, y_n$ is an i.i.d. sample of $Y$. Then $X = (\sum_{i=1}^n y_i)/n$, $V_m = \left(\frac{\nu}{n}\right)^{(2^m - 1)} \mu$ and $\hat{V}_m = \left(\frac{\nu}{n}\right)^{(2^m - 1)} X$ is an inferential sequence for $\mu$.*

**Proof** (i) $EX = (\sum_{i=1}^n EY)/n = \mu$; $\mathrm{Var}\,X = \frac{1}{n}\mathrm{Var}\,Y = \frac{\nu}{n} EY = \frac{\nu}{n}\mu = V_1$.

(ii) $\mathrm{Var}\,(\hat{V}_m) = \left(\frac{\nu}{n}\right)^{(2^{m+1} - 2)} \mathrm{Var}\,X = \left(\frac{\nu}{n}\right)^{(2^{m+1} - 1)} \mu = V_{m+1}$.

(iii) $E\hat{V}_m = \left(\frac{\nu}{n}\right)^{(2^m-1)} EX = V_m$.
(iv) Since $Y$ is non-negative, so is $X$ and, consequently, $\hat{V}_m$ for all $m \in \mathbb{N}^*$. ∎

**Theorem 6.2.** *Let $Y : \Omega_1 \to \mathbb{R}_+$ be a non negative random variable. Define $\nu \in \mathbb{R}$ by $\mathrm{Std}(Y) = \nu EY$: $\mu = EY$ and suppose $y_1, \ldots, y_n$ is an i.i.d. sample of $Y$. Then $X = \left(\sum_{i=1}^n y_i\right)/n$, $\sigma_m = \left(\frac{\nu}{\sqrt{n}}\right)^m \mu$ and $\hat{\sigma}_m = \left(\frac{\nu}{\sqrt{n}}\right)^m X$ is a standard deviation based inferential sequence for $\mu$.*

**Proof** (i) $EX = \mu$; $\mathrm{Std}(X) = \frac{\mathrm{Std}(Y)}{\sqrt{n}} = \frac{\nu}{\sqrt{n}}\mu = \sigma_1$.
(ii) $\mathrm{Std}(\hat{\sigma}_m) = \left(\frac{\nu}{\sqrt{n}}\right)^m \mathrm{Std}(X) = \left(\frac{\nu}{\sqrt{n}}\right)^{m+1} \mu = \sigma_{m+1}$.
(iii) $E\hat{\sigma}_m = \left(\frac{\nu}{\sqrt{n}}\right)^m EX = \sigma_m$.
(iv) Immediate. ∎

**Theorem 6.3.** *Under the same hypothesis and notation as in theorem 6.1 (or 6.2), let $\hat{a} = T(y_1, \ldots, y_n) + c$, where $T$ is a linear transformation, $T = (c_1, \ldots, c_n)$, and $c$ is a real constant, be an unbiased estimator for a parameter $a \in \mathbb{R}$. Then, $\hat{a}$, $V_m = \frac{\|T\|^{2^m} \nu^{(2^m-1)}}{n^{(2^{m-1}-1)}}\mu$ and $\hat{V}_m = \frac{\|T\|^{2^m} \nu^{(2^m-1)}}{n^{(2^{m-1}-1)}} X$ (or $\hat{a}$, $\sigma_m = \frac{\|T\|\nu^m}{(\sqrt{n})^{m-1}}\mu$, $\hat{\sigma}_m = \frac{\|T\|\nu^m}{(\sqrt{n})^{m-1}} X$) is an inferential sequence for the parameter $a$.*

**Proof** For variance based inference.
(i) $E\hat{a} = a$; $\mathrm{Var}(\hat{a}) = \mathrm{Var}(T(y_1, \ldots, y_n) + c) = \left(\sum_{i=1}^n c_i^2\right) \mathrm{Var}\, Y = \|T\|^2 \nu \mu = V_1$.
(ii) $\mathrm{Var}(\hat{V}_m) = \frac{\|T\|^{2^{m+1}} \nu^{(2^{m+1}-2)}}{n^{(2^m-2)}} \mathrm{Var}\, X = \frac{\|T\|^{2^{m+1}} \nu^{(2^{m+1}-1)}}{n^{(2^m-1)}} \mu = V_{m+1}$.
(iii) and (iv) immediate.

For standard deviation based inference.
(i) $\mathrm{Std}(\hat{a}) = \mathrm{Std}(T(y_1, \ldots, y_n)) = \sqrt{\|T\|^2 \mathrm{Var}\, Y} = \|T\|\nu\mu = \sigma_1$.
(ii) $\mathrm{Std}(\hat{\sigma}_m) = \frac{\|T\|\nu^m}{(\sqrt{n})^{m-1}} \mathrm{Std}(X) = \frac{\|T\|\nu^{m+1}}{(\sqrt{n})^m}\mu = \sigma_{n+1}$.
(iii) and (iv) immediate. ∎

Since we don't know the distribution of $Y$, the sure inference intervals for its mean are calculated as before.

Now we will consider examples where we know the distribution of $Y$. This information will be used for constructing *sure inference intervals under knowledge of distribution.*

**Theorem 6.4.** *Let $Y$ be a Poisson distributed random variable with mean $\mu$ ($Y \sim$ Poisson($\mu$)); $y_1, \ldots, y_n$ be an i.i.d. sample of $Y$, $Z = \left(\sum_{i=1}^n y_i\right)$, $X = Z/n$, $V_m = \left(\frac{1}{n}\right)^{2^m} (n\mu)$ and $\hat{V}_m = \left(\frac{1}{n}\right)^{2^m} Z$. For $\alpha_1 \in (0,1)$ let $\lambda_1^- = \lambda_1^-(\alpha_1)$ and $\lambda_1^+ = \lambda_1^+(\alpha_1)$ be such that $P\{n\mu - \lambda_1^-\sqrt{n\mu} \le Z' \le n\mu + \lambda_1^+\sqrt{n\mu}\} = 1 - \alpha_1$, where $Z' \sim$ Poisson($n\mu$), and $\lambda_1^+ - \lambda_1^-$ is minimum. For $m \ge 1$, let $\lambda_{m+1} = \lambda_{m+1}(\alpha_{m+1})$ be such that $P\{Z' \ge n\mu - \lambda_{m+1}\sqrt{n\mu}\} = 1 - \alpha_{m+1}$. Then, for all $m \in \mathbb{N}^*$, we have*

$$P\left\{\mu \in \left[\frac{Z}{n} - \frac{\lambda_1^+}{n}\sqrt{Z + \lambda_2\sqrt{Z + \cdots + \lambda_m\sqrt{n\mu}}}, \frac{Z}{n} + \frac{\lambda_1^-}{n}\sqrt{Z + \lambda_2\sqrt{Z + \cdots + \lambda_m\sqrt{n\mu}}}\right]\right\}$$
$$\ge 1 - \sum_{i=1}^m \alpha_i.$$

**Proof** Since $Y \sim \text{Poisson}(\mu)$ we can use Theorem 6.1 with $\nu = 1$ which yields the inferential sequence $X, V_m, \hat{V}_m$. Observe that $Z = \sum_{i=1}^{n} y_i$ is Poisson distributed and we have $Z \sim \text{Poisson}(n\mu) \sim Z'$. Now we proceed as in theorem's 2.1 proof. Noting that we have freedom to choose asymmetrical intervals and that

$$P\Big\{X \in [\mu - \lambda_1^- \sqrt{V_1}, \mu + \lambda_1^+ \sqrt{V_1}]\Big\} = P\Big\{Z \in [n\mu - \lambda_1^- \sqrt{n\mu}, n\mu + \lambda_1^+ \sqrt{n\mu}]\Big\} = 1 - \alpha_1$$

and, for all $m \in \mathbb{N}^*$,

$$P\Big\{\hat{V}_m + \lambda_{m+1}\sqrt{V_{m+1}} \geq V_m\Big\} = P\left\{\left(\frac{1}{n}\right)^{2^m} Z \geq \left(\frac{1}{n}\right)^{2^m} n\mu - \lambda_{m+1}\sqrt{\left(\frac{1}{n}\right)^{2^{m+1}} n\mu}\right\}$$
$$= P\Big\{Z \geq n\mu - \lambda_{m+1}\sqrt{n\mu}\Big\} = 1 - \alpha_{m+1};$$

we have

$$P\left\{\mu \in \left[X - \lambda_1^+ \sqrt{\hat{V}_1 + \lambda_2 \sqrt{+ \cdots + \lambda_m \sqrt{V_m}}}, X + \lambda_1^- \sqrt{\hat{V}_1 + \lambda_2 \sqrt{+ \cdots + \lambda_m \sqrt{V_m}}}\right]\right\}$$
$$\geq 1 - \sum_{i=1}^{m} \alpha_i.$$

Now, substituting $Z/n$ for $X$, $\hat{V}_j$, $1 \leq j \leq m-1$ and $V_m$ for their expressions, the result follows.
∎

The theorem above furnishes an m-th order sure inference interval for the mean of a Poisson distributed random variable. Let us call it sure inference interval under knowledge of distribution. This interval is approximated for practical purposes by

$$\left[\frac{Z}{n} - \frac{\lambda_1^+}{n}\sqrt{Z + \lambda_2\sqrt{Z + \cdots + \lambda_m\sqrt{Z}}}, \frac{Z}{n} + \frac{\lambda_1^-}{n}\sqrt{Z + \lambda_2\sqrt{Z + \cdots + \lambda_m\sqrt{Z}}}\right].$$

We observe that one might think that theorem 6.4 is useless, since we could have simply stopped at the first order inference to obtain

(3) $$P\left\{\mu \in \left[\frac{Z}{n} - \frac{\lambda_1^+}{n}\sqrt{n\mu}, \frac{Z}{n} + \frac{\lambda_1^-}{n}\sqrt{n\mu}\right]\right\} = 1 - \alpha_1,$$

which reduces to

$$\mu \in \left|\frac{\sqrt{Z + \left(\frac{\lambda_1^+}{2}\right)^2} - \left(\frac{\lambda_1^+}{2}\right)}{n}, \frac{\sqrt{Z + \left(\frac{\lambda_1^-}{2}\right)^2} + \left(\frac{\lambda_1^-}{2}\right)}{n}\right|,$$

where $|a, b| = (a, b) \cup (b, a)$, with probability $1 - \alpha_1$, i.e., probably (with probability $1 - \alpha_2$) the smaller interval with the higher assured probability level within the former intervals; but this is not the case from a statistician's point of view. First, in practice, it is common not to know exactly what the distribution of the data is; so, assuming that the data is a realization of a Poisson distributed random variable is something that has to be checked, a task that will require, for a reasonable confidence level, some minimum size for the data set. Second, if we have some clue, for example from mathematical modelling of the situation under study, that the distribution is approximately a Poisson one, it may be wiser not to assume it to be exactly a Poisson distribution and use the exact relation (3) but prefer instead to use higher order inferential intervals.

So, for small sample from Poisson or Poisson like distributions the use of at least second order sure inference intervals may be appropriate.

Now let $Y \sim \text{Gamma}(\alpha,\beta)$ be a gamma distributed random variable. Its density function is written as $f(y) = \frac{y^{\alpha-1}e^{-y/\beta}}{\beta^{\alpha}\Gamma(\alpha)}$ for $y > 0$ and $f(y) = 0$ for $y \le 0$, for positive real constants $\alpha$ and $\beta$. We remind that $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$. Observe that for $\beta = 1$ we have $\mu = \alpha = \sigma^2$ and we are under theorem's 6.1 hypothesis. On the other hand, if $\alpha = 1$ then $\mu = \beta = \sigma$ and theorem 6.2 may be applied.

Using the same notation as in theorem 6.4 we write the following theorems.

**Theorem 6.5.** *Let* $Y \sim \text{Gamma}(\mu, 1)$ *and* $Z' \sim \text{Gamma}(n\mu, 1)$. *Then, for all* $m \in \mathbb{N}^*$, *we have*

$$P\left\{\mu \in \left[\frac{Z}{n} - \frac{\lambda_1^+}{n}\sqrt{Z+\lambda_2\sqrt{Z+\cdots+\lambda_m\sqrt{n\mu}}}, \frac{Z}{n} + \frac{\lambda_1^-}{n}\sqrt{Z+\lambda_2\sqrt{Z+\cdots+\lambda_m\sqrt{n\mu}}}\right]\right\}$$
$$\ge 1 - \sum_{i=1}^{m}\alpha_i.$$

**Proof** Just note that $Z \sim \text{Gamma}(n\mu, 1)$ and follow the steps on theorem's 6.4 proof. ∎

**Theorem 6.6.** *Let* $Y \sim \text{Gamma}(1,\mu) = \exp(\mu)$ *and* $Z' \sim \text{Gamma}(n,\mu)$. *Let also* $X = Z/n$, $\sigma_m = \left(\frac{1}{\sqrt{n}}\right)^m \mu$ *and* $\hat{\sigma}_m = \left(\frac{1}{\sqrt{n}}\right)^m \frac{Z}{n}$. *For* $\lambda_1^+$ *and* $\lambda_1^-$ *defined by* $P\left\{n\mu\left(1-\frac{\lambda_1^-}{\sqrt{n}}\right) \le Z' \le n\mu\left(1+\frac{\lambda_1^+}{\sqrt{n}}\right)\right\} = 1-\alpha_1$, *for which* $\lambda_1^+ - \lambda_1^-$ *is minimum and* $\lambda_{m+1}$ *such that* $P\left\{Z' \ge n\mu\left(1-\frac{\lambda_{m+1}}{\sqrt{n}}\right)\right\} = 1-\alpha_{m+1}$ *for all* $m \in \mathbb{N}^*$, *we have*

$$P\left\{\mu \in \left[\frac{Z}{n}\left(1-\frac{\lambda_1^+}{\sqrt{n}}\left(1+\frac{\lambda_2}{\sqrt{n}}\left(1+\frac{\lambda_3}{\sqrt{n}}\left(\cdots\left(1+\frac{\lambda_m}{\sqrt{n}}\left(\frac{n\mu}{Z}\right)\right)\cdots\right)\right)\right)\right),\right.\right.$$
$$\left.\left.\frac{Z}{n}\left(1+\frac{\lambda_1^-}{\sqrt{n}}\left(1+\frac{\lambda_2}{\sqrt{n}}\left(1+\frac{\lambda_3}{\sqrt{n}}\left(\cdots\left(1+\frac{\lambda_m}{\sqrt{n}}\left(\frac{n\mu}{Z}\right)\right)\cdots\right)\right)\right)\right)\right]\right\},$$
$$\ge 1 - \sum_{i=1}^{m}\alpha_i.$$

**Proof** Observe that $Z \sim \text{Gamma}(n,\mu)$ and follow the steps in theorem's 6.4 proof with appropriate modifications from variances to standard deviations. ∎

The corresponding "practical" intervals obtained by substituting $Z$ for $n\mu$ in both intervals above may be used for small sample inference of the mean for $Y \sim \text{Gamma}(\mu, 1)$ and $Y \sim \exp(\mu)$.

## 7. Conclusions

In this article we propose an analysis of inference that is based on the principle of systematically doubting all assumptions that are made and all intermediate results that are obtained in the path to the main conclusions we want to draw from the information we have. This approach may generate a series of questions about the assumptions, the intermediate results and also about their possible answers in a cyclic way and we will want to answer these questions in a most conservative way. We do so in order to obtain secure and cautious conclusions. These conclusions avoid to the maximum extent, while *it is still convenient*, all doubts. Clearly, we will want to use all relevant available information.

As a matter of fact, we have presented a procedure for calculating confidence intervals. This was done for situations under which either we cannot or we do not want to assume any distribution for the data as well as for situations where we know some extra information that comes from outside the data as, for example, model assumptions relating means and variances, means and standard deviations or even the distributions of the random variables.

Typically, as an intermediate answer is also an affirmation to be questioned and checked, this procedure generates an infinite sequence of questions of the same type. One way of capturing this feature is the use of inferential sequences. This analysis of inference will always assume the worst case to draw conclusions, i.e., as if the inference situations were such that they were always, as much as they could possibly be, against the conclusion we want to arrive at. This feature is reflected in the construction of sure inference intervals and sure inference intervals under extra information.

We observe that this work suggests the existence of orders of inference. In this line, information that comes form outside the data, like knowing the distribution of a random variable, knowing the value of $\nu = VarX/EX$ for non negative random variables, knowing a bound for some of the moments of a random variable, or other hypothesis or assumptions, may be regarded as an infinite order inferential statement since, informally, it brings certainty to our analysis which can not be reached by finite order inference.

This article is a first collection of definitions and theorems that fit in this general idea of sure inference analysis.

## Bibliography

de Miranda, J.C.S. (2003). On the estimation of the intensity of point processes via wavelets. PhD thesis. University of São Paulo. (IME/USP). São Paulo, Brazil. (In Portuguese.)

de Miranda, J.C.S. and Morettin, P.A. (2003). *On Estimation of the Intensity of Point Processes via Wavelets*. Preprint.

Feller, W. An introduction to probability theory and its applications. Vol.1-2.

Rudin, W. (1981). *Real and Complex Analysis*. Tata Mc Graw-Hill Publishing Co.

Universidade de São Paulo
*E-mail address*: simon@ime.usp.br

**TRABALHOS DO DEPARTAMENTO DE MATEMÁTICA**

**TÍTULOS PUBLICADOS**

2004-01 ASSEM, I., COELHO. F.U., LANZILOTTA, M., SMITH, D.and TREPODE, SONIA Algebras determined by their left and right parts. 34p.

2004-02 FUTORNY, V., MOLEV, A. and OVSIENKO, S. Harish-Chandra Modules for Yangians. 29p.

2004-03 COX, B. L. and FUTORNY, V. Intermediate Wakimoto modules for affine *sl(n+1, C)*. 35p.

2004-04 GRISHKOV, A. N. and ZAVARNITSINE, A. V. Maximal subloops of simple Moufang loops. 43p.

2004-05 GREEN, E.L. and MARCOS, E. δ-Koszul Algebras. 15p.

2004-06 GORODSKI, C. Taut Representantions of compact simple lie groups. 16p.

2004-07 ASPERTI, A.C. and VALÉRIO, B.C. Ruled helicoidal surfaces in a 3-dimensional space form. 9p.

2004-08 ASSEM, I., CASTONGUAY, D., MARCOS, E.N. and TREPODE, S. Strongly simply connected schurian algebras and multiplicative bases. 26p.

2004-09 RODRIGUES, A.A.M., MIRANDA FILHO, R.C. and SOUZA, E.G. Definability and Invariance in First Order Structures. 15p.

2004-10 GIAMBRUNO, A. and MILIES, C. P. Free groups and involutions in the unit group of a group algebra. 5p.

2004-11 RAO, S. ESWARA and FUTORNY, V. Classification of integrable modules for affine lie superalgebras. 20p.

2004-12 GRISHKOV, A. N. and ZAVARNITSINE, A. V. Groups with triality. 22p.

2004-13 BUSTAMANTE, J. C. and CASTONGUAY, D. Fundamental groups and presentations of algebras. 10p.

2004-14 FERNANDES, S.M. and MARCOS, E.N. On the Fundamental Derivation of a Finite Dimensional Algebra. 15p.

2004-15 CARRIÓN, H.; GALINDO, P. and LOURENÇO, M.L. Banach spaces whose bounded sets are bounding in the bidual. 10p.

2004-16 KOSZMIDER, P. Projections in weakly compactly generated Banach spaces and Chang's conjecture. 15p.

2004-17 KOSZMIDER, P. On a problem of Rolewicz about Banach spaces that admit support sets. 19p.

2004-18 GOODAIRE, E.G., MILIES, C.P. and PARMENTER, M.M. Central units in alternative loop rings. 8p.

2004-19 GIAMBRUNO, A. and MILIES, C. P. Free groups and involutions in the unit group of a group algebra. 5p.

2004-20 CARRIÓN, H. Entire Functions on $l_1$. 16p.

2004-21 SHESTAKOV, I. and ZHUKAVETS, N. Universal multiplicative envelope of free Malcev superalgebra on one odd generator. 28p.

2004-22 ASSEM, I., BUSTAMANTE, J.C., CASTONGUAY, D. and NOVOA, C. A note on the fundamental group of a one-point extension. 6p.

2004-23 BAHTURIN, Y. A., SHESTAKOV, I. P. and ZAICEV, M . V. Gradings on Simple Jordan and Lie Algebras. 25p.

2004-24 DOKUCHAEV. M., KIRICHENKO. V. and MILIES, C. P. Engel subgroups of triangular matrices over local rings. 16p.

2004-25 ADVÍNCULA, F. H. and MARCOS, E. N., Stratifications of algebras with radical square zero. 7p.

2004-26 ADVÍNCULA, F. H. and MARCOS, E. N., Algebras which are standardly stratified in all orders. 8p.

2004-27 DOKUCHAEV, M. A., KIRICHENKO, V. V., ZELENSKY, A. V. and ZHURAVLEV, V. N. Gorenstein matrices. 22p.

2004-28 PEREIRA, A. L. and PEREIRA, M. C. An extension of the method of rapidly oscillating solutions. 15p.

2004-29 DOKUCHAEV, M. and EXEL, R. Associativity of crossed products by partial actions, enveloping actions and partial representations. 24p.

2004-30 FUTORNY, V. and USTIMENKO, V. Small world semiplanes with generalised Schubert cells. 10p.

2004-31 FERRAZ, R. A. and MILIES, C. P. Idempotents in group algebras and minimal abelian codes. 11p.

2004-32 CRISTO, O. B. and MILIES, C. P. Symmetric elements under oriented involutions in group rings. 12p.

2005-01 PEREIRA, A. L. and PEREIRA, M. C. Generic simplicity for the solutions of a nonlinear plate equation. 22p.

2005-02 ARAGONA, J. Colombeau generalized functions on quasi-regular sets. 25p.

2005-03 PEREIRA. A. L. Global attractor and nonhomogeneous equilibria for a non local evolution equation in an unbounded domain. 26p.

2005-04 SHESTAKOV, I. and ZHUKAVETS, N. Malcev Poisson superalgebra of free Malcev superalgebra on one odd generator. 19p.

2005-05 SHESTAKOV, I. and ZHUKAVETS, N. Speciality of Malcev superalgebras on one odd generator. 15p.

2005-06 BILLIG, Y., FUTORNY, V. and MOLEV, A. Verma modules for Yangians. 18p.

2005-07 GORODSKI, C. Jacobi fields on normal homogeneous Riemannian manifolds. 6p.

2005-08 GRISHKOV, A. Notes on simple Lie algebras over a field of characteristic 2. 7p.

2005-09 MIRANDA, J. C. S. and MORETTIN, P. A. Estimation of the density of point processes on $\mathbb{R}^m$ via wavelets. 15p.

2005-10 KASHUBA, I., OVSIENKO, S. and SHESTAKOV, I. Jordan algebras versus associative algebras. 35p.

2005-11 IWAKI, E., JESPERS, E. and JURIAANS, S. O. The hypercentre of the unit group of an integral group ring. 11p.

2005-12 FUTORNY, V. and OVSIENKO, S. Theory of Galois algebras. 54p.

2005-13 MIRANDA, J. C. S. Characterization of the intensity for a class of point processes. 16p.

2005-14 MIRANDA, J. C. S. Neglecting higher order infinitesimals and the study of some measures. 16p.

2005-15 MIRANDA, J. C. S. Some wide classes of point processes interrelationships and three working properties. 15p.

2005-16 MIRANDA, J. C. S. Sure inference analysis. 27p.

**Nota: Os títulos publicados nos Relatórios Técnicos dos anos de 1980 a 2002 estão à disposição no Departamento de Matemática do IME-USP.**

**Cidade Universitária "Armando de Salles Oliveira"**
**Rua do Matão, 1010 - Cidade Universitária**
**Caixa Postal 66281 - CEP 05315-970**