1. Introduction.

Whereas the classical theory of statistical inference in finite dimensional gaussian models is almost completely developed, in the infinite dimensional case many problems are unsolved. It is the purpose of this paper to provide some methods of statistical inference on gaussian infinite dimensional models. Our approach is based on recent developments of the theory of infinite dimensional statistical spaces and on the use of techniques from the theory of gaussian measures on Banach spaces.

Let us now be more specific and introduce the basic notations and conventions in order to summarize the results.

Let \( x = (x(t))_{t \in T} \) be an observation of a real random function \( (X(t))_{t \in T} \) such that
\[
X = \{ X(t) = m(t) + X_\circ(t) ; t \in T \}
\]
where \( T \) is a compact metric space and \( (X_\circ(t))_{t \in T} \) is a real gaussian function with zero mean and known covariance \( K \) on \( T \times T \). We shall assume that the mean function \( m \) belongs to the reproducing kernel Hilbert space \( H(K) \) of \( K \). The statistical space corresponding to this model is

\[
\left( \mathbb{R}^T, \mathcal{E}(\mathbb{R}), \left\{ \mathcal{N}_{\mathbb{R}^T}(m, K) ; m \in H(K) \right\} \right)
\]

where \( \mathcal{N}_{\mathbb{R}^T}(m, K) \) denotes the gaussian measure on \( \mathbb{R}^T \) with mean \( m \) and covariance \( K \). A statistical space is a triplet \( (E, \mathcal{C}, \mathcal{P}) \), that is, the mathematical model associated with a statistical experiment where \( E \) is a real vector space representing the set of all the possible observations \( x \), \( \mathcal{C} \) is the \( \sigma \)-field of subsets of \( E \) generated by the observable events and \( \mathcal{P} \) is a family of hypotheses concerning the probability distribution of the observed random varia-
ble $X$ in $(E, \mathcal{E})$. Assuming the model (1.1), statistics for estimating the mean function $m$ are given and tests of hypotheses of the form "$m \in V$", where $V$ is a finite dimensional subset of $\mathcal{H}(K)$, are performed.

An often adopted model for $V$ is to regard it as the linear span of a family of known functions on $T$. The tests performed are optimal when the dimension of $V$ is known. Thus, the next problem we are dealing with, is to define an estimation procedure for the dimension of $V$.

The representation of our model in terms of gaussian measures on Banach spaces is discussed in section 2.

In section 3, we survey some results of [1] which we shall need, concerning the estimation of the mean and the tests. Section 4 is devoted to the estimation of the dimension, which appears as an application of a result valid for models more general than model (1.1).


This section covers the basic definitions and notations necessary to the following work. All random variables considered from now on will be defined on a complete probability space $(\Omega, \mathcal{F}, P)$.

If $B$ is a real separable Banach space with norm $\| \|$, $B^*$ denotes the topological dual of $B$ and the symbol $(\cdot, \cdot)$ denotes the duality between $B$ and $B^*$. As usual $\mathcal{B}$ denotes the Borel $\sigma$-field of $B$.

Let $\mu$ be a centered probability measure on $(B, \mathcal{B})$ such that

\[ \int_B \|x\|^2 \, d\mu(x) < +\infty \]

and let $K$ denote the covariance function of $\mu$ defined by

\[ K(f,g) = \int_B (f,x) (g,x) \, d\mu(x) \quad (f,g \in B^*) \]

Then according to lemma 2.1 of [3], the reproducing kernel hilbert space $\mathcal{H}(K)$ of $K$ can be realized as a subset of $B$ and the natural inclusion map of $\mathcal{H}(K)$ into $B$, say $J$, is linear, one to one and continuous. The same properties are true for the adjoint map $J^*$ from $B^*$ into $\mathcal{H}(K)$ if we identify $\mathcal{H}(K)$ and $\mathcal{H}(K)$ in the usual way. Furthermore, $J^* (B^*)$ is dense in