Notes On

HILBERT SPACE OF STOCHASTIC PROCESSES

Filtering, Prediction, and Smoothing

Stochastic processes, defined over a continuous time interval, say $[0,1]$, are parametrized random variables (parametrized in the time variable t) . Denoting a (scalar) stochastic process by $X(t,\omega)$, we have for each $t\in[0,1]$ a random variable defined on the probability space $(\Omega,\mathcal{F},\mathcal{P})$, and for each $\omega\in\Omega$ a square-integrable function on $[0,1]$, that is an element of the Hilbert space $L_2[0,1]$. Collecting all this together, we can view a stochastic process, X, as a mapping from $(\Omega,\mathcal{F})$ to $(L_2[0,1],\mathcal{B}_{L_2[0,1]})$, where $\mathcal{B}_{L_2[0,1]}$ is a sigma algebra of subsets of $L_2[0,1]$. Defining the inner product of any two such maps by

$$(X,Y):=E\{\int_0^1 X(t,\omega)Y(t,\omega)dt\}\equiv\int_0^1 E\{X(t,\omega)Y(t,\omega)\}dt\,, \qquad (o)$$

we generate a Hilbert space of all such stochastic processes on $(\Omega,\mathcal{F},\mathcal{P})$, denoted by $H=L_2(\Omega,\mathcal{P};L_2[0,1])$. Note that the inner product (o) involves a composition of the inner product in the Hilbert space $L_2[0,1]$ (for each fixed $\omega\in\Omega$), and the inner product in the Hilbert space $L_2(\Omega,\mathcal{P})$ (for each fixed $t\in T$), both of which were introduced in class. I now formulate and then discuss the solution to the counterpart for stochastic processes of the *linear least squares estimation* problem discussed in class for random variables.

Problem 1. *Given $X\in H$ and an observed process $Y\in H$, find linear least squares estimate for X given Y. That is, find $\hat{X}\in M$, where M is a closed subspace of H, generated by square-integrable functions $K\in L_2([0,1]\times[0,1])$, defined as*

$$M=\{m:m(t,\omega)=\int_0^1 K(t,s)Y(s,\omega)\,ds\,,\ t\in[0,1]\,,\ K\in L_2([0,1]\times[0,1])\}\,, \qquad (oo)$$

such that

$$\|X-\hat{X}\|=\inf_{m\in M}\|X-m\|\,, \qquad (*)$$

where the norm is the one induced by the inner product (o).

Solution and Analysis. For the problem formulated above to make sense, we first have to make sure that M in (oo) is a closed subspace of H. For this, it will be sufficient that every $m\in M$ is also in H (as by linearity, it is straightforward to see that $m_1,m_2\in M$ implies that any linear combination is also in M). Now, note that

$$\begin{aligned}(m,m)&=E\{\int_0^1 dt\,|\int_0^1 K(t,s)Y(s,\omega)ds|^2\ \}\\ &\le E\{\int_0^1 dt\int_0^1|K(t,s)|^2ds\int_0^1|Y(\tau,\omega)|^2d\tau\ \}\\ &=\int_0^1 dt\int_0^1 ds\,|K(t,s)|^2\ (Y,Y)<\infty\end{aligned}$$

where the inequality follows from Cauchy-Schwarz applied to elements of $L_2[0,1]$ with $\omega\in\Omega$ and $t\in[0,1]$ fixed, and boundedness of the last expression follows because $K\in L_2([0,1]\times[0,1])$ and $Y\in H$. Hence, all elements of M are bounded in the same norm as in H,

Now, coming back to the optimization problem $(*)$, the standard *Projection Theorem* applies, and we are assured of existence of a solution, say $\hat{X}$, which is a projection of X onto M, that is $X-\hat{X}\perp M$. We can write $\hat{X}$ as

$$\hat{X}(t,\omega)=\int_0^1\hat{K}(t,s)Y(s,\omega)ds,\ \text{for some }\hat{\mathrm{K}}\in\mathrm{L}_2([0,1]\times[0,1])\,.$$

Define

$$R_{YY}(t,s) := E\{Y(t,\omega)Y(s,\omega)\} \text{ and } R_{XY}(t,s) := E\{X(t,\omega)Y(s,\omega)\}\,,$$

where R_{YY} is the auto-correlation function of Y, and R_{XY} is the cross-correlation function of X and Y. Then, carrying out the condition $X - \hat{X} \perp M$, we have the following sequence of steps for all $K \in L_2([0,1]\times[0,1])$:

$$E\Big\{\int_0^1 X(t,\omega)\,dt \int_0^1 K(t,s)Y(s,\omega)\,ds\Big\} = E\Big\{\int_0^1 dt \int_0^1 \hat{K}(t,\tau)Y(\tau,\omega)\,d\tau \int_0^1 K(t,s)Y(s,\omega)\,ds\Big\}\,,$$

$$\iff \quad \int_0^1 dt \int_0^1 ds\, K(t,s)R_{XY}(t,s) = \int_0^1 dt \int_0^1 ds \int_0^1 d\tau\, \hat{K}(t,\tau)K(t,s)R_{YY}(\tau,s)$$

$$\iff \quad \int_0^1 dt \int_0^1 ds\, K(t,s)\big[R_{XY}(t,s) - \int_0^1 d\tau\, \hat{K}(t,\tau)R_{YY}(\tau,s)\big] = 0$$

and this last equality holds for all $K \in L_2([0,1]\times[0,1])$, if and only if $\hat{K}$ solves what is inside the brackets with equality, that is

$$R_{XY}(t,s) = \int_0^1 \hat{K}(t,\tau)R_{YY}(\tau,s)\,d\tau \qquad (**)$$

By the *Projection Theorem*, there exists $\hat{K} \in L_2([0,1]\times[0,1])$ that solves $(**)$. For *uniqueness* we need an additional condition on Y (which is the counterpart of the condition we had in the case of a sequence of random variables, $\{Y_1, Y_2, \ldots, Y_n\}$, discussed in class today (February 21st), which was that they had to be linearly independent). The condition is that the auto-correlation function of Y, $R_{YY}(t,s)$ is positive definite. This means that taking any partition $\{0 \le t_1 < t_2 < \ldots < t_n \le 1\}$ of the interval $[0,1]$, the corresponding symmetric $n\times n$ matrix $R = \{R(t_i,t_j),\, i,j = 1,\ldots,n\}$ is positive definite, this holding for any n and any partitioning. This is equivalent to saying that each corresponding sequence of random variables $\{Y(t_i,\omega),\, i = 1,\ldots,n\}$, obtained by time-sampling Y, is a linearly independent set.

The following theorem summarizes the result (that is, the solution to Problem 1).

Theorem 1 (Smoothing). *Consider Problem 1 formulated above, under the additional assumption that $R_{YY}(t,s)$ is positive definite over $[0,1]\times[0,1]$. Then, there is a unique $\hat{X}$ that solves the problem, that is satisfying $(*)$, which is given by*

$$\hat{X}(t,\omega) = \int_0^1 \hat{K}(t,s)Y(s,\omega)\,ds\,,$$

where $\hat{K}$ is the unique solution of $(**)$ in $L_2([0,1]\times[0,1])$.

Note the qualifier "smoothing" in the label of the theorem, which corresponds to an estimation problem where the *linear least squares estimate* of the stochastic process X makes use of the entire path of the observation process Y for each time t (that is, the *past* as well as the *future*). Hence, in that sense the solution is not *causal*. Alternatives to this are *prediction* and *filtering* estimation problems which use only past values of Y in the estimation of X. A general class of such problems can be formulated as one of finding the *linear least squares estimate* for $X(t+\lambda,\omega)$, with $0 \le \lambda < 1$, using $\{Y(s,\omega),\, s \le t\}$, and doing this for each $t \le 1-\lambda$. The counterpart of Problem 1 in this case would be the following; here $\lambda = 0$ corresponds to the *filtering* problem, and the estimation problem with $\lambda > 0$ is referred to as the λ-*step ahead prediction.*

Problem 2 (Prediction and Filtering). *Given $X \in H$, an observed process $Y \in H$, and a fixed $\lambda \in [0,1)$, find, for each $t \in [0,1-\lambda)$, a linear least squares estimate for $X(t+\lambda,\omega)$ given $Y(s,\omega)$, $s \le t\}$. That is, find $\hat{X}_{t+\lambda} \in M_t$, where M_t is a closed subspace of H, generated by square-integrable functions $K \in L_2([0,1]\times[0,1])$, defined as*

$$M_t = \big\{m : m(t,\omega) = \int_0^t K(t,s)Y(s,\omega)\,ds\,,\ t \in [0,1]\,,\ K \in L_2([0,1]\times[0,1])\big\}\,, \qquad (ooo)$$

such that, for each t,

$$\|X(t+\lambda) - \hat{X}(t+\lambda)\| = \inf_{m(t)\in M_t} \|X(t+\lambda) - m(t)\|\,, \qquad (***)$$

where the norm is the one induced by the inner product in $L_2(\Omega, \mathcal{P})$.

Again by the *Projection Theorem*, we have, for each t, $X(t+\lambda,\omega) - \hat{X}(t+\lambda,\omega) \perp M_t$, which, following the steps in the solution of Problem 1, leads to the following result. [**Prove it!**].

Theorem 2 (Prediction and Filtering). *Consider Problem 2 formulated above, under the additional assumption that $R_{YY}(t,s)$ is positive definite over $[0,1]\times[0,1]$. Then, there is a unique $\hat{X}$ that solves the problem, that is satisfying $(***)$, which is given by*

$$\hat{X}(t+\lambda,\omega) = \int_0^t \hat{K}(t,s)Y(s,\omega)\,ds\,,$$

where $\hat{K}$ is the unique solution of

$$R_{XY}(t+\lambda,s) = \int_0^t \hat{K}(t,\tau)R_{YY}(\tau,s)\,d\tau$$

for each $t \in [0, 1-\lambda]$.

♢ ♢ ♢