RANDOM PROCESSES. THE FINAL TEST SOLUTION.
June, 26th, 2000

Problem 1.

(a) $\mathbf{E}\big[\mathbf{E}(X|Y)|X\big] \stackrel{?}{=} X$. Wrong. E.g. if X and Y are independent, then $\mathbf{E}(X|Y) = \mathbf{E}X$ and $\mathbf{E}\big[\mathbf{E}(X|Y)|X\big] = \mathbf{E}X \neq X$

(b) $\mathbf{E}(X|Y) \equiv \mathbf{E}X \quad \stackrel{?}{\Longrightarrow} \quad X$ and Y indepedent. Wrong. E.g. let ξ be a r.v. with zero mean and $Y \in \{0,1\}$ with prob. $\{1-p,p\}$. ξ and Y are independent. Set $X = \xi Y$ Consider the pair (X,Y). Clearly:

$$\mathbf{E}(X|Y) = \mathbf{E}(\xi Y|Y) = Y\mathbf{E}\xi = 0 \equiv \mathbf{E}X$$

Let us show that X and Y depend:

$$\mathbf{E}|X|Y = \mathbf{E}|\xi Y|Y = \mathbf{E}|\xi|Y^2 = \mathbf{E}|\xi|Y = p\mathbf{E}|\xi|$$

on the other hand:

$$\mathbf{E}|X| \cdot \mathbf{E}Y = \mathbf{E}|\xi Y| \cdot p = p^2\mathbf{E}|\xi|$$

That is:

$$\mathbf{E}|X|Y \neq \mathbf{E}|X|\mathbf{E}Y$$

(c) $\mathbf{E}(X|Y) \stackrel{?}{=} \mathbf{E}\big[X|\mathbf{E}(X|Y)\big]$. Correct. By definition

$$\mathbf{E}\big[X|\mathbf{E}(X|Y)\big] = \phi\big(\mathbf{E}(X|Y)\big)$$

such that:

$$\mathbf{E}\big[X - \phi\big(\mathbf{E}(X|Y)\big)\big]g\big(\mathbf{E}(X|Y)\big) = 0 \tag{1}$$

for all bounded g. Take $\phi(x) = x$ and note that $g\big(\mathbf{E}(X|Y)\big)$ is actually a function of Y, so that (1) holds. Due to uniqueness of cond. expectation with prob. 1, we conclude that the statement is correct.

(d) $\{X,Y,Z\}$ is Gaussian, such that $\mathbf{E}X = 0$ and Y and Z are independent, then

$$\mathbf{E}(X|Y,Z) = \mathbf{E}(X|Y) + \mathbf{E}(X|Z)$$

This is correct and verified e.g. by explicit calculation (see also lecture notes)

$$\begin{aligned}\mathbf{E}(X|Y,Z) &= \mathbf{Cov}(X,Y)/\mathbf{Cov}(Y,Y)(Y - \mathbf{E}Y) + \\ &+ \mathbf{Cov}(X,Z)/\mathbf{Cov}(Z,Z)(Z - \mathbf{E}Z) = \mathbf{E}(X|Y) + \mathbf{E}(X|Z)\end{aligned} \tag{2}$$

(e) If $\{X,Y,Z\}$ is non Gaussian, then (2) is generally false. Assume $\mathbf{E}X = 0$ and let $X = ZY$, so that Z and Y are independent and with zero mean. Then $\mathbf{E}(X|Y,Z) = ZY \neq \mathbf{E}(X|Y) + \mathbf{E}(X|Z) = 0$

(f) (I) If $\mathbf{E}(X|Y) = c_0 + c_1 Y$, then $\mathbf{E}(X|Y) = \widehat{\mathbf{E}}(X|Y)$ with prob. 1. Let us show that

$$\mathbf{E}\big(\mathbf{E}(X|Y) - \widehat{\mathbf{E}}(X|Y)\big)^2 = 0 \tag{3}$$

$$\mathbf{E}\big(\mathbf{E}(X|Y) - \widehat{\mathbf{E}}(X|Y)\big)^2 = \mathbf{E}\big(\mathbf{E}(X|Y) - X + X - \widehat{\mathbf{E}}(X|Y)\big)^2 =$$
$$= \mathbf{E}\big(\mathbf{E}(X|Y) - X\big)^2 - 2\mathbf{E}\big(\mathbf{E}(X|Y) - X\big)\big(X - \widehat{\mathbf{E}}(X|Y)\big) +$$
$$+\mathbf{E}\big(X - \widehat{\mathbf{E}}(X|Y)\big)^2$$

But (why?)

$$\mathbf{E}\big(\mathbf{E}(X|Y) - X\big)\big(X - \widehat{\mathbf{E}}(X|Y)\big) = \mathbf{E}\big(\mathbf{E}(X|Y) - X\big)\big(X - \mathbf{E}(X|Y)\big)$$

so

$$\mathbf{E}\big(\mathbf{E}(X|Y) - \widehat{\mathbf{E}}(X|Y)\big)^2 = \mathbf{E}\big(X - \widehat{\mathbf{E}}(X|Y)\big)^2 - \mathbf{E}\big(X - \mathbf{E}(X|Y)\big)^2$$

Clearly $\mathbf{E}\big(X - \widehat{\mathbf{E}}(X|Y)\big)^2 \geq \mathbf{E}\big(X - \mathbf{E}(X|Y)\big)^2$.

But since $\mathbf{E}(X|Y)$ is linear in Y, we have $\mathbf{E}\big(X - \widehat{\mathbf{E}}(X|Y)\big)^2 \leq \mathbf{E}\big(X - \mathbf{E}(X|Y)\big)^2$ (recall that orthogonal projection is the best *linear* estimate). This implies (3).

(II) Since $\mathbf{E}X^2 < \infty$ and $\mathbf{E}Y^2 < \infty$ for any *linear* function $\ell(x)$

$$\mathbf{E}\big(X - \mathbf{E}(X|Y)\big)\ell(Y) = 0$$

Since $\mathbf{E}(X|Y) = c_0 + c_1 Y$ (i.e. linear (affine) in Y) and by uniqueness of the orthogonal projection we conclude $\mathbf{E}(X|Y) = \widehat{\mathbf{E}}(X|Y)$.

(g) (I) E.g. let ξ be a r.v. with $\mathbf{E}\xi = 1$ and Y be a r.v. with $\mathbf{E}Y = 0$, $\mathbf{E}Y^2 < \infty$. ξ and Y are independent. Define $X = \xi Y$. Then

$$\mathbf{E}(X|Y) = \mathbf{E}(\xi Y|Y) = Y\mathbf{E}\xi = Y$$

Note that $\mathbf{E}X = 0$ and

$$\widehat{\mathbf{E}}(X|Y) = \frac{\mathbf{Cov}(X,Y)}{\mathbf{Cov}(Y,Y)}(Y - \mathbf{E}Y) = \frac{\mathbf{E}XY}{\mathbf{E}Y^2}Y = \frac{\mathbf{E}\xi Y^2}{\mathbf{E}Y^2}Y = Y \cdot \mathbf{E}\xi = Y$$

(II) Simply pick any independent $X = c_0 + c_1 Y$. Or X and Y independent (in this case $c_0 = \mathbf{E}X$ and $c_1 = 0$.)

(h) $X > Y \overset{?}{\Longrightarrow} \widehat{\mathbf{E}}(X|Z) > \widehat{\mathbf{E}}(Y|Z)$. Wrong. A simple example is $Y \equiv C < 0$ and $X = |\xi|$, where ξ is e.g. Gaussian. Clearly $X > Y$. Moreover $\widehat{\mathbf{E}}(Y|Z) \equiv C$ and $\widehat{\mathbf{E}}(X|Z) = \alpha + \beta Z$, where α and β are some constants ($\alpha = \mathbf{E}|\xi|$, etc.). Clearly Z can be chosen so that $\mathbf{P}\{\alpha + \beta Z < C\} > 0$ (e.g. choose Z Gaussian), which means that $\mathbf{P}\big\{\widehat{\mathbf{E}}(X|Z) < \widehat{\mathbf{E}}(Y|Z)\big\} > 0$.

In several particular cases, the property holds, e.g. X and Y are orthogonal to Z:

$$\widehat{\mathbf{E}}(X|Z) = \mathbf{E}X, \quad \widehat{\mathbf{E}}(Y|Z) = \mathbf{E}Y$$

But
$$X > Y \implies \mathbf{E}X > \mathbf{E}Y$$

Problem 2.

(a) Given θ, the process $(X_n, Y_n)_{n\ge 1}$ is Gaussian.

Introduce Gaussian processes $(X^{(i)}_n, Y^{(i)}_n)_{n\ge 0}$, $i = 1, ..., d$, generated by:

$$\begin{aligned} X^{(i)}_n &= a(i)X_{n-1} + b(i)\varepsilon_n, \quad X^{(i)}_0 = X_0 \\ Y^{(i)}_n &= X^{(i)}_{n-1} + \sigma\xi_n, \quad n \ge 1 \end{aligned}$$

and let $\phi^i_n(\lambda^n_0, \mu^n_1)$ denote its characteristic function:

$$\phi^i_n(\lambda^n_0, \mu^n_1) = \mathbf{E}\exp\left\{ i\sum_{\ell=0}^n \lambda_\ell X^{(i)}_\ell + i\sum_{\ell=1}^n \mu_\ell Y^{(i)}_\ell \right\}$$

where λ_i and μ_i are real numbers.

Let $\phi_n(\lambda^n_0, \mu^n_1; \theta)$ denote the *conditional* characteristic function of (X_n, Y_n), i.e.

$$\phi_n(\lambda^n_0, \mu^n_1; \theta) = \mathbf{E}\exp\left\{ i\sum_{\ell=0}^n \lambda_\ell X_\ell + i\sum_{\ell=1}^n \mu_\ell Y_\ell \Big| \theta \right\}$$

Clearly

$$\phi_n(\lambda^n_0, \mu^n_1; \theta) = \sum_{i=1}^d \phi^i_n(\lambda^n_0, \mu^n_1) I(\theta = i)$$

so that $\phi_n(\lambda^n_0, \mu^n_1; \theta)$ has a form of a Gaussian characteristic function (depending of θ, of course)

(b) Set $m_n := \mathbf{E}(X_n|\theta)$, then:

$$\begin{aligned} m_n &= \mathbf{E}(X_n|\theta) = \mathbf{E}\big(a(\theta)X_{n-1}|\theta\big) + \mathbf{E}\big(b(\theta)\varepsilon_n|\theta\big) = \\ &= a(\theta)\mathbf{E}(X_{n-1}|\theta) = a(\theta)m_{n-1} \end{aligned}$$

Similarly:

$$V_n = a^2(\theta)V_{n-1} + b^2(\theta)$$

(c) $(X_n, Y_n)_{n\ge 1}$ is not a Gaussian process, e.g. the distribution of X_1 is non Gaussian, in fact it is a *Gaussian mixture*:

$$f(x) = \frac{d\mathbf{P}\{X_1 \le x\}}{dx} = \sum_i p_i \varphi_i(x)$$

where $\varphi_i(x)$ is the density of a Gaussian r.v. with zero mean and variance $a^2(i) + b^2(i)$.

(d) Note that $(X_n, Y_n)_{n\ge 1}$ is Gaussian, conditioned on $\{\theta = i\}$. So the optimal estimate $\widehat{X}_n(i) = \mathbf{E}(X_n|Y^n_1, \theta = i)$ is given by the Kalman filter ($n \geq 1$):

$$\begin{aligned} \widehat{X}_n(i) &= a(i)\widehat{X}_{n-1}(i) + \frac{aP_{n-1}(i)}{P_{n-1}(i)+\sigma^2}(Y_n - \widehat{X}_{n-1}(i)) \\ P_n(i) &= a^2(i)P_{n-1}(i) + b^2(i) - \frac{a^2(i)P^2_{n-1}(i)}{P_{n-1}(i)+\sigma^2} \end{aligned} \tag{4}$$

subject to $\widehat{X}_0(i) = 0$ and $P_0(i) = 1$, $i \in S$.

I.e.

$$\widehat{X}_n(\theta) = \sum_{i=1}^{d} \widehat{X}_n(i) I(\theta = i)$$

(e) Clearly:

$$\widehat{X}_n = \mathbf{E}(X_n|Y_1^n) = \sum_j \mathbf{P}\{\theta = j|Y_1^n\}\mathbf{E}(X_n|Y_1^n, \theta = j) = \sum_j \pi_n(j)\widehat{X}_n(j)$$

i.e. the optimal on-line filter in this case can be constructed by a combination of a bank of d Kalman filters and a Wonham filter (as we will see shortly)

(f) The conditional probability $\pi_n(i)$ is found as a function $G(x; Y_1^{n-1})$, such that:

$$\mathbf{E}\big[I(\theta = i)h(Y_n)|Y_1^{n-1}\big] = \mathbf{E}\big[G(Y_n; Y_1^{n-1})h(Y_n)|Y_1^{n-1}\big] \tag{5}$$

for any bounded h.

The left hand side:

$$\begin{aligned} &\mathbf{E}\big(I(\theta = i)\big[h(Y_n)|\theta = i, Y_1^{n-1}\big]|Y_1^{n-1}\big) = \\ &= \mathbf{E}\Big(I(\theta = i)\int h(x)\varphi_i(x)dx|Y_1^{n-1}\Big) = \\ &= \pi_{n-1}(i)\int h(x)\varphi_i(x)dx \end{aligned}$$

where $\varphi_i(x)$ is a Gaussian density with mean $\widehat{X}_{n-1}(i)$ and variance $P_{n-1}(i) + \sigma^2$, i.e.

$$\varphi_i(x) = \frac{1}{\sqrt{2\pi(P_{n-1}(i)+\sigma^2)}} \exp\left\{-\frac{\big(x - \widehat{X}_{n-1}(i)\big)^2}{2(P_{n-1}(i)+\sigma^2)}\right\}$$

This follows from the fact that given $\theta = i$, the conditional distribution of Y_n given Y_1^{n-1} is Gaussian. Calculating the right hand side of (5) and using the arbitrariness of $h(x)$ we arrive at:

$$\pi_n(i) = \frac{\pi_{n-1}(i)\varphi_i(Y_n)}{\sum_j \pi_{n-1}(j)\varphi_j(Y_n)} \tag{6}$$

subject to $\pi_0(i) = p(i)$.

Problem 3.

(a) See lecture note 9 (optimal filtering of finite state Markov chain)
(b) See lecture note 9 (optimal linear filtering of finite state Markov chain)
(c) (I) Let I_n be a vector with elements $I(\theta_n = a_i)$, $a_i \in S$. Then (these formulae have been derived in class)

$$I_n = \Lambda^\top I_{n-1} + \nu_n$$

where $(\nu_n)_{n\geq 1}$ is a vector sequence such that:

$$\mathbf{E}\nu_n \equiv 0, \quad \mathbf{E}\nu_n\nu_m^\top = \delta(n-m)D_n$$

$$D_n = \mathbf{diag}(V_n) - \Lambda^\top \mathbf{diag}(V_{n-1})\Lambda$$

and

$$V_n = \Lambda^\top V_{n-1}$$

subject to $V_n = p$.
Introduce an *augmented* state vector (in $\mathbb{R}^{d+1}$):

$$X_n := \begin{pmatrix} I_n \\ -- \\ \xi_n \end{pmatrix}$$

Then

$$\begin{aligned} X_n &= \underbrace{\begin{pmatrix} \Lambda^\top & 0 \\ 0 & \gamma \end{pmatrix}}_{:=\Gamma} X_{n-1} + \widetilde{\varepsilon}_n \\ Y_n &= \widetilde{S}^\top X_n = \widetilde{S}^\top \Gamma X_{n-1} + \widetilde{S}^\top \widetilde{\varepsilon}_n \end{aligned}$$

where

$$\widetilde{S} = \begin{pmatrix} H(a_1) \\ H(a_2) \\ \vdots \\ H(a_d) \\ 1 \end{pmatrix} \in \mathbb{R}^{d+1}$$

and $(\widetilde{\varepsilon}_n)_{n\geq 1}$ is an $\mathbb{R}^{d+1}$ valued sequence of zero mean r.v. such that:

$$\mathbf{E}\widetilde{\varepsilon}_n\widetilde{\varepsilon}_m = \delta(n-m)\begin{pmatrix} D_n & 0 \\ 0 & 1 \end{pmatrix} := Q_n$$

The linear optimal estimate is given by the Kalman filter:

$$\begin{aligned} \widehat{X}_n &= \Gamma\widehat{X}_{n-1} + (\Gamma P_{n-1}\Gamma^\top \widetilde{S}^\top + Q_n\widetilde{S}) \cdot \\ &\quad \cdot(\widetilde{S}^\top \Gamma P_{n-1}\Gamma^\top \widetilde{S} + \widetilde{S}^\top Q_n \widetilde{S})^+ (Y_n - \widetilde{S}\Gamma\widehat{X}_{n-1}) \end{aligned} \tag{7}$$

$$\begin{aligned} P_n &= \Gamma P_{n-1}\Gamma^\top + Q_n - (\Gamma P_{n-1}\Gamma^\top \widetilde{S}^\top + Q_n\widetilde{S}) \cdot \\ &\quad \cdot(\widetilde{S}^\top \Gamma P_{n-1}\Gamma^\top \widetilde{S} + \widetilde{S}^\top Q_n \widetilde{S})^+ (\Gamma P_{n-1}\Gamma^\top \widetilde{S}^\top + Q_n\widetilde{S})^\top \end{aligned} \tag{8}$$

and $\widehat{\theta}_n = \widehat{\mathbf{E}}(\theta_n | Y_1^n) = \sum_j a_j \widehat{X}_n(j)$.

(II) Let $\mathcal{H}$ be a vector with elements $H(a_i)$, $\nu_n := I_n - \Lambda^\top I_{n-1}$ and J denote the identity matrix:

$$
\begin{aligned}
&Y_n = \theta_n + \xi_n = \mathcal{H}^\top I_n + \gamma \xi_{n-1} + \varepsilon_n = \mathcal{H}^\top I_n + \gamma(Y_{n-1} - \mathcal{H}^\top I_{n-1}) + \varepsilon_n \\
&= \mathcal{H}^\top (\Lambda^\top I_{n-1} + \nu_n) + \gamma(Y_{n-1} - \mathcal{H}^\top I_{n-1}) + \varepsilon_n = \\
&= \mathcal{H}^\top (\Lambda^\top - \gamma J) I_{n-1} + \gamma Y_{n-1} + \mathcal{H}^\top \nu_n + \varepsilon_n
\end{aligned}
$$

Together with $I_n = \Lambda^\top I_{n-1} + \nu_n$, a linear model, suitable for the Kalman filter is obtained.

(d) Following the standard technique, we look for a function $G(x; Y_1^{n-1})$ such that:

$$\mathbf{E}\big[I(\theta_n = a_i) h(Y_n) | Y_1^{n-1}\big] = \mathbf{E}\big[G(Y_1; Y_1^{n-1}) h(Y_n) | Y_1^{n-1}\big] \tag{9}$$

First calculate:

$$
\begin{aligned}
&\mathbf{E}\big[I(\theta_n = a_i) h(Y_n) | \theta_{n-1}, Y_1^{n-1}\big] = \\
&= \mathbf{E}\Big[\sum_\ell I(\theta_{n-1} = a_\ell) I(\theta_n = a_i) h(H(a_i) + \gamma \xi_{n-1} + \varepsilon_n) | \theta_{n-1}, Y_1^{n-1}\Big] = \\
&= \sum_\ell I(\theta_{n-1} = a_\ell) \lambda_{\ell i} \int h(H(a_i) + \gamma(Y_{n-1} - a_\ell) + x) f(x) dx
\end{aligned}
$$

Taking the conditional expectation with respect to Y_1^{n-1} of the latter equation we arrive at an expression for the left hand side of (9):

$$\sum_\ell \pi_{n-1}(\ell) \lambda_{\ell i} \int h(x) f(x - H(a_i) - \gamma(Y_{n-1} - H(a_\ell))) dx$$

By similar calculations one obtains an expression for the right hand side, which finally lead to the filter:

$$\pi_n(i) = \frac{\sum_\ell f\big(Y_n - H(a_i) - \gamma(Y_{n-1} - H(a_\ell))\big) \lambda_{\ell i} \pi_{n-1}(\ell)}{\sum_i \sum_\ell f\big(Y_n - H(a_i) - \gamma(Y_{n-1} - H(a_\ell))\big) \lambda_{\ell i} \pi_{n-1}(\ell)}, \quad n \geq 2 \tag{10}$$

Since $\xi_0 = 0$, $Y_1 = \theta_1 + \xi_1 = \theta_1 + \varepsilon_1$:

$$\pi_1(i) = \frac{\sum_\ell f\big(Y_1 - H(a_i)\big) \lambda_{\ell i} p(\ell)}{\sum_i \sum_\ell f\big(Y_1 - H(a_i)\big) \lambda_{\ell i} p(\ell)} \tag{11}$$

Note that for $\gamma = 0$, this filter is reduced to the conventional Wonham filter.