SOLUTION OF THE FINAL TEST

Prof. R. Liptser and Pavel Chigansky

9:00-13:00, 2nd of July, 1999

Problem 1. *Convergence Of Random Sequences*

(a) Direction 'if': it must be shown that $\mathbf{P}\{|\xi_n - \eta_n| > \varepsilon\} \xrightarrow{\mathbf{P}} 0$ implies $\mathbf{P}\{\xi \neq \eta\} = 0$. For any $\varepsilon > 0$:

$$\begin{aligned} &\mathbf{P}\{|\xi - \eta| > \varepsilon\} \leq \mathbf{P}\{|\xi - \xi_n| + |\xi_n - \eta_n| + |\eta_n - \eta| > \varepsilon\} \leq \\ \leq\ &\mathbf{P}\{|\xi - \xi_n| > \varepsilon/3\} + \mathbf{P}\{|\xi_n - \eta_n| > \varepsilon/3\} + \mathbf{P}\{|\eta_n - \eta| > \varepsilon/3\} \to 0 \end{aligned}$$

Since $\{\omega : |\xi - \eta| > \varepsilon\}$ does not depend on n we conclude that $\mathbf{P}\{|\xi - \eta| > \varepsilon\} \equiv 0$ for any $\varepsilon > 0$, i.e. $\mathbf{P}\{\xi \neq \eta\} = 0$

Direction 'only if': show that $\mathbf{P}\{\xi \neq \eta\} = 0$ implies $\mathbf{P}\{|\xi_n - \eta_n| > \varepsilon\} \xrightarrow{\mathbf{P}} 0$:

$$\begin{aligned} &\mathbf{P}\{|\xi_n - \eta_n| > \varepsilon\} \leq \mathbf{P}\{|\xi_n - \xi| + |\xi - \eta| + |\eta - \eta_n| > \varepsilon\} \leq \\ &\leq \mathbf{P}\{|\xi_n - \xi| > \varepsilon/3\} + \mathbf{P}\{|\xi - \eta| > \varepsilon/3\} + \mathbf{P}\{|\eta - \eta_n| > \varepsilon/3\} = \\ &= \mathbf{P}\{|\xi_n - \xi| > \varepsilon/3\} + \mathbf{P}\{|\eta - \eta_n| > \varepsilon/3\} \to 0 \end{aligned}$$

(b) If $a = 0$ or/and $b = 0$ the statement is trivial. If $a \neq 0$ and $b \neq 0$:

$$\begin{aligned} &\mathbf{P}\{|a\xi_n + b\eta_n - [a\xi + b\eta]| > \varepsilon\} \leq \mathbf{P}\{|a||\xi_n - \xi| + |b||\eta_n - \eta| > \varepsilon\} \leq \\ &< \mathbf{P}\{|a||\xi_n - \xi| > \varepsilon/2\} + \mathbf{P}\{|b||\eta_n - \eta| > \varepsilon/2\} \to 0, \quad \forall \varepsilon > 0 \end{aligned}$$

(c) By definition $f(x)$ is continuous if for any $\varepsilon > 0$ there exists $\delta > 0$, such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$:

$$\mathbf{P}\{|\xi_n - \xi| < \delta\} \leq \mathbf{P}\{|f(\xi_n) - f(\xi)| < \varepsilon\}$$

which in turn implies:

$$\mathbf{P}\{|f(\xi_n) - f(\xi)| > \varepsilon\} \leq \mathbf{P}\{|\xi_n - \xi| > \delta\} \to 0, \quad \forall \varepsilon > 0$$

(d) For discontinuous function the statement is incorrect. E.g. let ξ_n be a binary sequence:

$$\mathbf{P}\{\xi_n = \pm 1/n\} = 1/2$$

Clearly $\xi_n \xrightarrow{\mathbf{P}} 0$. Let $f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$. Then $\mathbf{P}\{f(\xi_n) = 0\} = \mathbf{P}\{f(\xi_n) = 1\} = 1/2$ for all n and thus for any $0 < \epsilon < 1$ we have

$$\mathbf{P}\{|f(\xi_n) - f(0)| > \epsilon\} = \mathbf{P}\{|f(\xi_n)| > \epsilon\} = 1/2 \not\to 0$$

Problem 2. *Gaussian Processes*

(a) To prove that $(X_n, Y_n)_{n\ge 0}$ is a Gaussian process we have to show that the function:

$$\varphi(\mu_0^n, \lambda_0^n) = \mathbf{E}\exp\Big\{i\sum_{i=0}^n \mu_i X_i + i\sum_{j=0}^n \lambda_j Y_j\Big\}$$

is the exponent of a quadratic form of $\mu_0^n = [\mu_0, ..., \mu_n]^\top$ and $\lambda_0^n = [\lambda_0, ..., \lambda_n]^\top$. The recursion of X_n and Y_n can be rewritten as:

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} = \underbrace{\begin{bmatrix} a & 0 \\ A & 0 \end{bmatrix}}_{\triangleq \Gamma} \begin{bmatrix} X_{n-1} \\ Y_{n-1} \end{bmatrix} + \underbrace{\frac{1}{\sqrt{X_{n-1}^2 + Y_{n-1}^2}} \begin{bmatrix} X_{n-1} & Y_{n-1} \\ -Y_{n-1} & X_{n-1} \end{bmatrix}}_{\triangleq U(X_{n-1}, Y_{n-1})} \begin{bmatrix} \varepsilon_n \\ \xi_n \end{bmatrix}$$

Note that the random matrix $U(X_{n-1}, Y_{n-1})$ is unitary, i.e. $UU^\top = I$.

$$\varphi_n(\mu_0^n, \lambda_0^n) = \mathbf{E}\mathbf{E}\Big[\exp\Big\{i\sum_{i=0}^n \mu_i X_i + i\sum_{j=0}^n \lambda_j Y_j\Big\}\Big| Y_0^{n-1}, X_0^{n-1}\Big] =$$

$$= \mathbf{E}\exp\Big\{i\sum_{i=0}^{n-1} \mu_i X_i + i\sum_{j=0}^{n-1} \lambda_j Y_j\Big\}\mathbf{E}\Big[\exp\Big\{i\mu_n X_n + i\lambda_n Y_n\Big\}\Big| Y_0^{n-1}, X_0^{n-1}\Big] \quad (1)$$

Further:

$$\begin{aligned}
&\mathbf{E}\Big[\exp\Big\{i\mu_n X_n + i\lambda_n Y_n\Big\}\Big| Y_0^{n-1}, X_0^{n-1}\Big] = \\
&= \exp\Big\{i\mu_n a X_{n-1} + i\lambda_n A X_{n-1}\Big\}\cdot \\
&\cdot \mathbf{E}\Big[\exp\Big\{i[\mu_n, \lambda_n] U(X_{n-1}, Y_{n-1})[\varepsilon_n, \xi_n]^\top\Big\}\Big| X_{n-1} Y_{n-1}\Big] = \\
&= \exp\Big\{i\mu_n a X_{n-1} + i\lambda_n A X_{n-1}\Big\}\exp\Big\{-1/2(\mu_n^2 + \lambda_n^2)\Big\} \qquad (2)
\end{aligned}$$

where the latter equality is due to the fact that $U(X_{n-1}, Y_{n-1})$ is unitary and $[\varepsilon_n, \xi_n]$ is a standard Gaussian vector, independent of $[X_{n-1}, Y_{n-1}]$. Proceed by induction: if $\varphi_{n-1}(\mu_0^{n-1}, \lambda_0^{n-1})$ is an exponent of quadratic function of its arguments then from (2) it follows that $\varphi_n(\mu_0^n, \lambda_0^n)$ is also exponent of quadratic form. Given that the initial condition $[X_0, Y_0]$ is Gaussian, we conclude that (X_n, Y_n) is a Gaussian process.

Remark: alternative proof would be to define a pair of sequences:

$$\widetilde{\varepsilon}_n = \frac{X_{n-1}\varepsilon_n + Y_{n-1}\xi_n}{\sqrt{X_{n-1}^2 + Y_{n-1}^2}}, \quad \widetilde{\xi}_n = \frac{-Y_{n-1}\varepsilon_n + X_{n-1}\xi_n}{\sqrt{X_{n-1}^2 + Y_{n-1}^2}}$$

and to show that $(\widetilde{\varepsilon}_n, \widetilde{\xi}_n)_{n\ge 1}$ is a Gaussian process. This implies that $(X_n, Y_n)_{n\ge 0}$ is a Gaussian process too.

(b) Since $[X_n, Y_n]^\top$ is a Gaussian vector it is sufficient to find its mean and covariance

$$m_n = \mathbf{E}\begin{bmatrix} X_n \\ Y_n \end{bmatrix}, \quad V_n = \mathbf{E}\begin{bmatrix} X_n \\ Y_n \end{bmatrix}\begin{bmatrix} X_n & Y_n \end{bmatrix}$$

which is easily found from the recursion for $[X_n, Y_n]$:

$$\begin{aligned} m_n &= \Gamma m_{n-1}, \quad m_0 = 0 \Rightarrow m_n \equiv 0 \\ V_n &= \Gamma V_{n-1}\Gamma^\top + I, \quad \Gamma_0 = Q \end{aligned}$$

so that the required density (assuming it exists):

$$f(x,y) = \frac{1}{2\pi\sqrt{\det\{V_n\}}} \exp\Big\{ -1/2[x,y]V_n^{-1}[x,y]^\top \Big\}$$

(c) Define a new pair of processes $(\widetilde{X}_n, \widetilde{Y}_n)$ by means of the recursion:

$$\begin{aligned} \widetilde{X}_n &= a\widetilde{X}_{n-1} + \varepsilon_n, \quad n \ge 1 \\ \widetilde{Y}_n &= A\widetilde{X}_{n-1} + \xi_n \\ \widetilde{X}_0 &\equiv X_0, \quad \widetilde{Y}_0 \equiv Y_0 \end{aligned} \tag{3}$$

The calculations as in (1) and (2) show that this pair has the same distribution as (X_n, Y_n). By Markov property of $(\widetilde{X}_n, \widetilde{Y}_n)_{n\ge 0}$:

$$\begin{aligned} f(x_0^n, y_0^n) &= \frac{1}{2\pi\sqrt{\det\{Q\}}} \exp\Big\{ -1/2[x_0,y_0]Q^{-1}[x_0,y_0]^\top \Big\} \times \qquad (4) \\ &\times \prod_{k=1}^n \frac{1}{\sqrt{2\pi}} \exp\Big\{ -1/2(x_k - ax_{k-1})^2 \Big\} \times \\ &\times \prod_{k=1}^n \frac{1}{\sqrt{2\pi}} \exp\Big\{ -1/2(y_k - Ax_{k-1})^2 \Big\} \end{aligned}$$

Remark: It was also possible to write the density for the original model, which after a number of simplifications could be reduced to (4)

(d) As it was mentioned above the system (3) has the same distribution as the original system. So their conditonal expectations also coincide (almost surely). The conventional Kalman filter generates the desired estimate $(n \ge 1)$

$$\begin{aligned} \widehat{X}_n &= a\widehat{X}_{n-1} + \underbrace{\frac{AaP_{n-1}}{A^2P_{n-1}+1}}_{\triangleq G_n}(Y_n - A\widehat{X}_{n-1}), \quad \widehat{X}_0 = Q_{xy}/Q_{yy}Y_0 \qquad (5) \\ P_n &= a^2P_{n-1} + 1 - \frac{[AaP_{n-1}]^2}{A^2P_{n-1}+1}, \quad P_0 = Q_{xx} - Q_{xy}^2/Q_{yy} \end{aligned}$$

Remark: alternatively one can use the orthogonal projection method. Once the (X_n, Y_n) is known to be Gaussian, the orthogonal projection coincides with the conditonal expectation. E.g.

$$\widehat{X}_{n|n-1} = \widehat{\mathbf{E}}(X_n|Y_0^{n-1}) = \mathbf{E}(X_n|Y_0^{n-1}) = a\mathbf{E}(X_{n-1}|Y_0^{n-1}) = a\widehat{X}_{n-1}$$

and hence

$$\begin{aligned} P_{n|n-1}^x &= \mathbf{E}(X_n - \widehat{X}_{n-1})^2 = \\ &= \mathbf{E}\left(a(X_{n-1} - \widehat{X}_{n-1}) + \frac{X_{n-1}\varepsilon_n + Y_{n-1}\xi_n}{\sqrt{X_{n-1}^2 + Y_{n-1}^2}} \right)^2 = \\ &= a^2P_{n-1} + 1 \end{aligned}$$

Similarly the expressions for $\widehat{Y}_{n|n-1}$, $P^{xy}_{n|n-1}$ and $P^{y}_{n|n-1}$ are obtained and the we arrive at the final result (5).

(e) Using the property of conditional expectation:

$$\begin{aligned}\widehat{X}_{n+k|n} &= \mathbf{E}(X_{n+k}|Y_0^n) = \mathbf{E}(\mathbf{E}(X_{n+k}|Y_0^{n+k})|Y_0^n) = \mathbf{E}(\widehat{X}_{n+k}|Y_0^n) = \\ &= \mathbf{E}\Big[a\widehat{X}_{n+k-1} + G_n(Y_{n+k} - A\widehat{X}_{n+k-1})\Big|Y_0^n\Big] = \\ &= a\mathbf{E}(\widehat{X}_{n+k-1}|Y_0^n) = ... = a^k\widehat{X}_n\end{aligned}$$

which can also be written as a recursion:

$$\widehat{X}_{n+i|n} = a\widehat{X}_{n+i-1|n}, \quad i = 1, ..., k$$

subject to $\widehat{X}_{n|n} = \widehat{X}_n$. In turn from the signal equation we have:

$$X_{n+i} = aX_{n+i-1} + \varepsilon_{n+i}, \quad i = 1, ..., k$$

subject to X_n.

To calculate the prediction error [1] introduce $D_{n+i} = X_{n+i} - \widehat{X}_{n+i|n}$, then:

$$D_{n+i} = aD_{n+i-1} + \varepsilon_{n+i}, \quad i = 1, ..., k$$

Squaring and taking the expectation we find:

$$P_{n+i|n} = a^2P_{n+i-1|n} + 1, \quad i = 1, ..., k$$

subject to $P_{n|n} = P_n$. The latter generates $P_{n+k|n}$ after k iterations.

(f) If X_0 and Y_0 dependent and at least one of them is non Gaussian, the estimate $\widehat{X}_0 = Q_{xy}/Q_{yy}Y_0$ is no longer optimal, so generally $\widehat{X}_n$ is not optimal. However in this case the original model, given in the problem, and the system (3) have the same distributions. Since for the system (3) the filter is still optimal among all the linear estimates, we conclude that optimality in the class of linear estimates is preserved.

The above considerations are true also for X_0 and Y_0 are independent and X_0 is non Gaussian.

If X_0 and Y_0 are independent and Y_0 is non Gaussian, whereas X_0 is Gaussian the $\widehat{X}_n$ remains optimal. In this case $\widehat{X}_0 = Q_{xy}/Q_{yy}Y_0 = 0 = \mathbf{E}(X_0|Y_0)$. Also the distribution of Y_0 does not affect the distribution of $[X_0, ..., X_n, Y_1, ..., Y_n]$ (i.e. it is remains Gaussian) and hence the conditional expectation is still generated by the same filter.

The above holds also for predicting estimate.

Problem 1. *Comparison of linear and non linear filters*

(a) Introduce a signal equation ($n \geq 1$):

$$\theta_n = \theta_{n-1}, \quad \theta_0 = \theta$$

[1] was not required in the test

Clearly $\theta_n \equiv \theta$ and hence $\widehat{\theta}_n = \mathbf{E}(\theta_n|Y_0^n)$ and $Y_n = \theta_n + \xi_n$ The latter is readily obtained by Kalman filter:

$$\begin{aligned}\widehat{\theta}_n &= \widehat{\theta}_{n-1} + \frac{P_{n-1}}{P_{n-1}+\sigma^2}(Y_n - \widehat{\theta}_{n-1})\\ P_n &= P_{n-1} - \frac{P_{n-1}^2}{P_{n-1}+\sigma^2}\end{aligned}$$

subject to $P_0 = \pi_0(1-\pi_0)$ and $\widehat{\theta}_0 = \pi_0$.

(b) $\widehat{\theta}$ converges to 0 in mean square sense (and hence also in the mean and in probability). In fact, P_n can be found explicitly:

$$P_n = \frac{P_{n-1}\sigma^2}{P_{n-1}+\sigma^2}$$

Let $Q_n = 1/P_n$ then:

$$Q_n = Q_{n-1} + 1/\sigma^2$$

or

$$Q_n = 1/P_0 + n/\sigma^2$$

Hence

$$P_n = \frac{P_0\sigma^2}{\sigma^2 + nP_0} \to 0$$

That is $\mathbf{E}(\widehat{\theta}-\theta)^2 \to 0$.

(c) It is also possible to treat θ_n as a degenerate Markov chain, i.e.:

$$\lambda_{j,i} = \mathbf{P}\{\theta_n = i|\theta_{n-1} = j\} = I(i=j)$$

Then using the formulas, derived in class we obtain the following non linear filter:

$$\pi_n = \frac{f(Y_n-1)\pi_{n-1}}{f(Y_n-1)\pi_{n-1} + f(Y_n)(1-\pi_{n-1})}, \quad n = 1, ...$$

subject to π_0.

(d)

$$
\begin{aligned}
V_1 &\triangleq \mathbf{E}(\pi_1-\theta)^2 = \mathbf{E}\mathbf{E}\big[(\pi_1-\theta)^2|\theta\big] = \\
&= \pi_0\mathbf{E}\big[(\pi_1-1)^2|\theta=1\big] + (1-\pi_0)\mathbf{E}\big[(\pi_1-0)^2|\theta=0\big] = \\
&= \pi_0\mathbf{E}\left\{\left(\frac{f(Y_1)(1-\pi_0)}{f(Y_1-1)\pi_0+f(Y_1)(1-\pi_0)}\right)^2 \Big| \ \theta=1\right\} + \\
&\quad +(1-\pi_0)\mathbf{E}\left\{\left(\frac{f(Y_1-1)\pi_0}{f(Y_1-1)\pi_0+f(Y_1)(1-\pi_0)}\right)^2 \Big| \ \theta=0\right\} = \\
&= \pi_0\mathbf{E}\left(\frac{f(\xi_1+1)(1-\pi_0)}{f(\xi)\pi_0+f(\xi_1+1)(1-\pi_0)}\right)^2 + \\
&\quad +(1-\pi_0)\mathbf{E}\left(\frac{f(\xi_1-1)\pi_0}{f(\xi-1)\pi_0+f(\xi_1)(1-\pi_0)}\right)^2 = \\
&= \pi_0\int_{-\infty}^{\infty}\left(\frac{f(x+1)(1-\pi_0)}{f(x)\pi_0+f(x+1)(1-\pi_0)}\right)^2 f(x)dx + \\
&\quad +(1-\pi_0)\int_{-\infty}^{\infty}\left(\frac{f(x-1)\pi_0}{f(x-1)\pi_0+f(x)(1-\pi_0)}\right)^2 f(x)dx = \\
&= \pi_0\int_{-\infty}^{\infty}\left(\frac{f(z)(1-\pi_0)}{f(z-1)\pi_0+f(z)(1-\pi_0)}\right)^2 f(z-1)dx + \\
&\quad +(1-\pi_0)\int_{-\infty}^{\infty}\left(\frac{f(x-1)\pi_0}{f(x-1)\pi_0+f(x)(1-\pi_0)}\right)^2 f(x)dx = \\
&= \pi_0(1-\pi_0)\int_{-\infty}^{\infty}\frac{f(x)f(x-1)}{f(x-1)\pi_0+f(x)(1-\pi_0)}dx
\end{aligned}
$$

By the way (was <u>not</u> required in the test) it is possible to derive an expression for V_n. Note the following fact:

$$\pi_n^{-1} = 1 + f(Y_n)/f(Y_n-1)(\pi_{n-1}^{-1}-1)$$

Let $\psi_n = \pi_n^{-1} - 1$ then

$$\psi_n = f(Y_n)/f(Y_n-1)\psi_{n-1}, \quad \psi_0 = \pi_0^{-1}-1$$

or

$$\psi_n = (\pi_0^{-1}-1)\prod_{k=1}^{n}\frac{f(Y_k)}{f(Y_k-1)}, \quad n=1,...$$

Returning to π_n:

$$\pi_n = \frac{1}{1+\psi_n} = \frac{\pi_0\prod_{k=1}^n f(Y_k-1)}{\pi_0\prod_{k=1}^n f(Y_k-1) + (1-\pi_0)\prod_{k=1}^n f(Y_k)}$$

From here similarly to the case $n=1$ (solved above) we can derive the formula:

$$\mathbf{E}(\pi_n-\theta)^2 = \pi_0(1-\pi_0)\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}\frac{\prod_{k=1}^n f(x_k)f(x_k-1)}{\pi_0\prod_{k=1}^n f(x_k-1)+(1-\pi_0)\prod_{k=1}^n f(x_k)}dx_1...dx_k$$

(e)

$$V_1 = \pi_0(1-\pi_0)\int_0^1 \frac{1/2\cdot 1/2}{1/2\pi_0+1/2(1-\pi_0)}dx = \frac{\pi_0(1-\pi_0)}{2}$$

FIGURE 1. $P_1 \geq V_1$

The linear filter gives the following error:

$$P_1 = P_0 - P_0^2/(P_0+\sigma^2) = P_0\sigma^2/(P_0+\sigma^2) = \frac{1/3\pi_0(1-\pi_0)}{\pi_0(1-\pi_0)+1/3}$$

Clearly $P_1 > V_1$ for $\pi_0 \in (0,1)$ and $P_1 = V_1 = 0$ for $\pi_0 = 0$ or $\pi_0 = 1$

(f) For the case $f(x) = 1/2I(x \in [0,1])$ the estimate is:

$$\pi_n = \begin{cases} 0, & Y_n \in [-1,0) \\ \pi_{n-1}, & Y_n \in [0,1) \\ 1, & Y_n \in [1,2] \end{cases}$$

Then $(n \geq 2)$

$$\begin{aligned} V_n &= \mathbf{E}(\pi_n - \theta)^2 = \pi_0 \mathbf{E}[(\pi_n - 1)^2 | \theta = 1] + (1-\pi_0)\mathbf{E}[(\pi_n)^2|\theta = 0] = \\ &= \pi_0 \Big[\mathbf{E}(\pi_{n-1} - 1)^2 \cdot \mathbf{P}\{\xi_n \in [0,1]\}\Big] + (1-\pi_0)\Big[\mathbf{E}\pi_{n-1}^2 \cdot \mathbf{P}\{\xi_n \in [0,1]\}\Big] = \\ &= 1/2\mathbf{E}(\pi_{n-1} - \theta)^2 = 1/2V_{n-1} \end{aligned}$$

subject to $V_1 = \pi_0(1-\pi_0)/2$. Clearly $V_n = (1/2)^n \pi_0(1-\pi_0) \to 0$. The sequence P_n also converges to 0 (both estimates are consistent). But V_n decreases faster (exponentially!), compared to P_n for which the convergence is linear (see (b))