水文諸量의 豫測이나 分析에 있어서 時系列論 應用에 對하여 (I)

崔 榮 博*

머리말

地表面부근에 있어서 風速, 氣溫 또는 太陽黑點數等과 같은 氣象諸量, 또한 河川流域各測點에 있어서 降水量, 水位, 流量과 같은 水文諸量은 모두 時系列(time series)의 한 예이다. 特히 우리 나라 水資源開發이나 治水立場에서 볼때 河川 流出이 가장 主된 時系列의 課題가 된다. 流出이라는 것은 두말 할것도 없이 河川에 있어서 雨水의 流出이며 降水는 河川流出의 源泉이 되는 것이다. 流出은 表面流出, 中間流出 및 地下水流出의 3成分으로 나누어지며 平水時에서는 地下水流出分이 高水時에는 表面流出과 中間流出을 포함한 直接流出分이 卓越하다는 것은 周知의 事實이다.

最近, 長期經濟開發計劃과 함께 產業의 高度化, 人口의 都市集中, 生活水準의 向上은 물의 需要를 急增하고 있고 一面 年平均 63億원이 되는 水災額을 가져오는 水害等을 감안할때 水資源의 保全 및 開發은 緊急을 要하는 重要問題로서 關心거리가 되었다. 1965年부터 시작한 UNESCO의 國際水文10個年計劃 (International Hydrological Decade)는 이와같은 要求에서 誕生된 것도 우리는 記憶에 새롭고 韓國에서도 建設部水資源局에서 微弱하나마 이에 呼應하고 있는줄 안다. 이 計劃中에는 물收支라는 課題가 크게 取扱되어 있고 또 水資源의 實態把握이라는 우리 現實立場에서도 基本的 問題가 되므로 筆者는 特히 降水와 河川流出에 對한 解明에 時系列論을 試圖한 바 있다. 河川計劃에 있어서 利水上이나 治水上의 여러 計劃樹立過程에 있어서 河川流出量의 將來豫測이 基本이 된다는 것도 두말할 것도 없다. 이경우 普通 取扱되는 手法은 過去에 發生한 事象을 解析해서 그性狀을 明白히 하고 이를 基礎로 해서 將來 生起되는 事象을 推測하는 時系列的方法이 있다. 이것은 特히 美國등에서 發展된 바 있다. 여기서는 水文諸量豫測이나 分析에 있어서 必要한 時系列論에 對하여 于先 論述하고 次後 그應用法을 例示코저 하는 바이다.

1. 時系列論概要

一般으로 時系列變動中에는 觀測器機나 觀測統計方法에 依存하는 變動因子와 함께 一般으로 偶然에 支配되는 因子가 포함 되고 있다. 따라서 現實로 우리가 얻어진 時系列은 하나의 標本이다. 이 母集團에 相當하는 것을 確率過程(stochastic process) $X(t,\omega)$ $(a\leqq t\leqq b)$ 또는 確率系列(stochastic sequence) $X_K(\omega)$ $(k=\cdots,-1,0,+1,+2\cdots)$이다. 여기서 ω는 可能한 몇個函數 또는 數列의 하나하나를 指示하는 標識로서 ω를 固定하면 $X(t,\omega)$, $X_K(\omega)$는 各各 하나의 函數 또는 數列을 表示하고 任意의 $t=t_1,t_2\cdots,t_n$을 固定하면 $X(t_1,\omega)$, $X(t_2,\omega)$, $\cdots$, $X(t_n,\omega)$는 n次元의 確率變數이다. 時系列을 基本으로 해서 그 母集團인 確率過程(系列)의 性質을 研究하고 또將來의 豫測을

* 技術士(建設部門)
理學博士
高麗大學校理工大教授

確率的으로 行하는 것이 時系列의 目的이다. 앞으로는 簡便을 위하여 $X(t,\omega)$, $X_K(\omega)$를 $X(t)$, X_K로 쓰기로 한다.

只今 X_1, $X_2\cdots$가 서로 獨立(純偶發的)일때는 獨立確率變數의 理論에 基本을 둔 統計法이 適用되나 이들은 獨立이 아니고 어느 種類의 依存關係가 있을 경우가 많다. 이와같은 確率的 依存關係가 이 時系列論의 主된 對象이 된다. 그 依存關係의 가장 간단한 것은 X_n의 確率法測이 X_{n-1}이 取한 값만큼에 依存하고 X_{n-2}, X_{n-3}, $\cdots$의 取한 값에는 無關係인 것으로서 이것을 單純마르코프系列(simple Markoff sequence)이라 한다.

任意의 t_1, $t_2\cdots$에 對하여 $X(t_1)$, $X(t_2)$, $\cdots$이 마르코프系列을 이룰때 $X(t)$를 單純마르코프過程(simple Markoff procces)라 한다. 確率過程 $X(t)$에 있어서 任意의 時點(個數도 任意) t_1, t_2, $\cdots$, t_n에 對한 $X(t_1)$, $X(t_2)$, $\cdots X(t_n)$의 同時分布函數를 $F(t_1,\ x_1;\ t_2,\ x_2;\cdots;\ t_n,\ x_n)$로 할때 時點 t_1, t_2, $\cdots t_n$을 모두 같은 t만큼 늦추워도 이 確率法則이 不變의 경우, 即

$F(t_1+t,\ x_1;\ t_2+t,\ x_2;\cdots;\ t_n+t,\ x_n)=F(t_1,\ x_1;\ t_2,\ x_2;\cdots t_n,\ x_n)$ 일때 確率過程 $X(t)$는 '定常(stationary)이라고 한다.

定常일때 任意의 時刻 t에 있어서 $X(t)$의 分布函數는 t에 無關係로 된다.

이것을 主分布函數 (principal distribution function)이라 하고 $F(x)$로서 表示한다. $F(x)$에 의해서 定해지는 特性量이로서는

平均値(mean): $m=E(X(t))=\int_{-\infty}^{\infty}x\,dF(x)$

分 散(variance): $\delta^2=E\{(X(t)-m)^2\}$

$$=\int_{-\infty}^{\infty}(x-m)^2dF(x)$$

이것도 t에는 關係없다.

다음에 任意의 2時點 t_1, t_2에 있어서 $X(t_1)$, $X(t_2)$의 同時分布函數는 時間間隔 $\tau=t_1-t_2$만에 依存하므로 이것은 $F(|\tau|\ x_1,x_2)$으로 쓰고 原點의 周圍의 2次自己積率 (automoment)

$$\mu_2'(\tau)=E\{X(t)\times X(t+\tau)\}$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2\ dF(|\tau|\ x_1,\ x_2)$$

도 τ만에 依存하고 t에 依存하지 않는다. 따라서 自己相關係數 (autocorrelation coefficient)

$$\rho(\tau)=\frac{E\{(X(t)-m)(X(t+\tau)-m)\}}{\sqrt{E\{(X(t)-m)^2\}\cdot E\{(X(t+\tau)-m)^2\}}}$$

또 τ안에 依해서 定해진다.

(明白히 $|\rho(\tau)|\leqq 1$, $\rho(0)=1$이다)

$\rho(\tau)$가 原點에서 連續일때, 即 $\rho(\pm 0)=\rho(0)=1$일때는 모든 ρ에 對하여 連續이 된다. 이때 그 定常的 確率過程은 連續이라고 한다. 任意의 2時點 s, t에 對해서 $X(s)$와 $X(t)$가 獨立할때 그 定常的確率過程은 純偶發的 (purely random)이라고 한다. 純偶發的過程은 不連續이다. ($\because\rho(\pm 0)=0$, $\rho(0)=1$) 確率系列에 對해서는 連續性은 問題가 안된다.

定常時系列은 定常確率過程 x_1, x_2, $\cdots$, x_N에서의 標本으로 생각되는 時系列이라고 한다. 定常確率過程에 있어서 위에 記述한 特性量에 各各 對應하는 統計量을 計算해서 이것을 研究할 수 있다. 即

系列平均値(serial mean): $\overline{x}=\frac{1}{N}\sum_{i=1}^{N}x_i$

系列分散(serial variance):

$$S^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2=\frac{1}{N}\sum_{i=1}^{N}x_i^2-\overline{x}^2$$

系列相關係數 (serial correlation coefficient):

$$r_K=\frac{1}{N-k}\sum_{i=1}^{N-k}(x_i-\overline{x}_1)(x_{i+k}-\overline{x}_2)/S_1S_2$$

$$=\left(\frac{1}{N=k}\sum_{i=1}^{N-k}x_ix_{i+k}-\overline{x}_1\overline{x}_2\right)/S_1S_2$$

但

$$\overline{x}_1=\sum_{i=1}^{N-k}x_i/(N-k)$$

$$\overline{x}_2=\sum_{i=k+1}^{N}x_i/(N-k)$$

$$S_1^2=\sum_{i=1}^{N-k}(x_i-\overline{x}_1)^2/(N-k)$$

$$S_2^2=\sum_{i=k+1}^{N}(x_i-\overline{x}_2)^2/(N-k)$$

이들의 統計量과 母集團過程이 特性量의 關係와 ergodicity 性으로서 주어진다. 即 一般으로 正常時系列에 있어서는 그統計量의 長時間平均은 集團平均에 거의 같다는 것이다. 即

$$\lim_{N\to\infty} \bar{x} = m, \quad \lim_{N\to\infty} S^2 = \delta^2, \quad \lim_{N\to\infty} r_k = \rho(k)$$

이들은 $\bar{x}$, S^2, r_k이 각각 m, δ^2, $\rho(k)$의 一致評價量이라는 것을 나타낸다.

또한 定常時系列의 經驗的分布函數($x_i \leqq x$인 x_i의 個數의 比率)를 $F_N(x)$이라고 하면 이것은 $N\to\infty$일때 主分布函數 $F(x)$에 매우 가까와진다.

定常時系列解析에 있어서 그 各項의 依存性을 나타내는 系列相關係數는 特히 重要한 役割을 한다. $N \gg k$일때 이것은

$$r_k = \left(\frac{1}{N-k}\sum_{i=1}^{N-k} x_i x_{i+k} - \bar{x}^2\right) / S^2$$

에 의해서 計算하여도 支障이 없다.

예를 들면 r_1, r_2, r_3까지 計算하는데는 다음 方式이 좋다.

$i \downarrow$	$k\to$		0	1	2	3	4'
	x_i'	x_i	x_i^2	$x_i x_{i+1}$	$x_i x_{i+2}$	$x_i x_{i+3}$	$x_i x_{i+4}'$
1	x_1'	x_1	x_1^2	x_1x_2	x_1x_3	x_1x_4	x_1x_5'
2	x_2'	x_2	x_2^2	x_2x_3	x_2x_4	x_2x_5	x_2x_6'
3	x_3'	x_3	x_3^2	x_3x_4	x_3x_5	x_3x_6	x_3x_7'
4	x_4'	x_4	x_4^2	x_4x_5	⋮	⋮	⋮
5	x_5'	x_5	x_5^2	x_5x_6	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
N−1	x_{N-1}'	x_{N-1}	x_{N-1}^2	$x_{N-1}x_N$	—	—	—
N	x_{N}'	x_N	x_N^2	—	—	—	—
合 計	—	Σx_i	Σx_i^2	$\Sigma x_i \Sigma x_{i+1}$	$\Sigma x_i \Sigma x_{i+2}$	$\Sigma x_i x_{i+3}$	$\Sigma x_i x_{i+4}'$
平 均	—	$\bar{x}$	$\frac{1}{N}\Sigma x_i^2$	$\frac{1}{N-1}\Sigma x_i x_{i+1}$	$\frac{1}{N-2}\Sigma x_i x_{i+2}$	$\frac{1}{N-3}\Sigma x_i x_{i+3}$	—
$-\bar{x}^2$	—	$(\bar{x}^2)$	S^2	$\frac{1}{N-k}\Sigma x_i x_{i+k} - \bar{x}^2$			—
$\div S^2$	—	—	—	r_1	r_2	r_3	—

여기서 x_i'와 $x_i x_{i+1}'$의 欄은 檢算을 위한 것으로서 $x_N' = x_N$, $x_{N-1}' = x_N' + x_{N-1}$, $x_{N-2}' = x_{N-1}' + x_{N-2}$ ………, $x_1' = x_2' + x_1 = \sum_{1}^{N} x_i$ (아래에서 累計함)이다.

그리고 x_i中에 正負가 混在할때는 最小數의 絶對値를 加해서 負數를 없애든지, 또는 $x_i x_{i+k}$의 欄을 正負 2欄으로 나누어서 처리한다.

系列相關係數의 有意性을 檢定하는 表는 다음 表—1과 같다. r_k의 그래프를 correlogram 이라 한다.

表—1 系列相關係數의 Significance Point

N	Negative 1%	tail 5%	Positive 5%	tail 1%	N	Negative 1%	tail 5%	positive 5%	tail 1%
[illegible]	[illegible]	[illegible]	0.371	0.531	13	0.634	0.497	0.341	0.495
[illegible]	[illegible]	[illegible]	0.366	0.533	14	0.615	0.479	0.335	0.485
[illegible]	[illegible]	[illegible]	[illegible]	0.525	15	0.597	0.462	0.328	0.475
[illegible]	[illegible]	[illegible]	[illegible]	0.515	16	0.582	0.447	0.322	0.464
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	17	0.566	0.433	0.315	0.455

18	0.550	0.421	0.309	0.446	70	0.287	0.207	0.178	0.259
19	0.538	0.410	0.304	0.439	80	0.271	0.197	0.170	0.242
20	0.524	0.399	0.299	0.432	90	0.257	0.185	0.162	0.231
21	0.502	0.380	0.289	0.427	100	0.248	0.173	0.155	0.221
24	0.483	0.363	0.280	0.404	110	0.233	0.615	0.149	0.213
26	0.465	0.350	0.272	0.392	120	0.222	0.158	0.142	0.205
28	0.449	0.337	0.264	0.380	130	0.214	0.151	0.137	0.196
30	0.443	0.325	0.257	0.370	140	0.198	0.146	0.132	0.190
32	0.420	0.315	0.251	0.360	160	0.190	0.137	0.124	0.176
34	0.408	0.305	0.245	0.351	180	0.180	0.130	0.117	0.167
36	0.398	0.295	0.240	0.344	200	0.169	0.122	0.111	0.159
38	0.388	0.286	0.234	0.336	220	0.160	0.116	0.106	0.153
40	0.376	0.279	0.229	0.329	250	0.150	0.108	0.101	0.145
45	0.356	0.262	0.218	0.314	300	0.137	0.098	0.092	0.133
50	0.339	0.248	0.208	0.301	350	0.129	0.091	0.086	0.124
55	0.324	0.236	0.199	0.298	400	0.119	0.085	0.080	0.114
60	0.310	0.225	0.191	0.278	500	0.107	0.076	0.071	0.103
65	0.298	0.216	0.182	0.268	600	0.095	0.069	0.066	0.093

連續定常인 確率過程 $X(t)$의 自己相關係數 $\rho(\tau)$는

$$\rho(\tau)=\int_{-0}^{\infty}\cos\,\tau\lambda dV(\lambda)$$

의 形으로 표시된다. 逆으로 $V(\lambda)$는

$$V(\lambda)=\frac{1}{\pi}\int_{0}^{\infty}\rho(\tau)\frac{\sin\,\tau\lambda}{\tau}$$

로서 주어진다. 여기서 $V(\lambda)$는 $0\leqq\lambda<\infty$ 의 分布函數로서 確率過程 $X(t)$의 스펙트름分布函數(spectrum distribution function)이라고 부른다. 特히 모든 λ에 對해서

$$v(\lambda)=V'(\lambda)=\frac{2}{\pi}\int_{0}^{\infty}\rho(\tau)\,\cos\tau\lambda d\tau$$

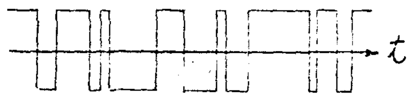
가 存在할때 이것을 스펙트름密度(spectrum density)이라고 한다.

여기서 $X(t)$는 다음式과 같다.

$$X(t)=\sum_{i=1}^{\infty}(A_i\,\cos\lambda_i t+B_i\,\sin\lambda_i t)$$

例로서 그림 1과 같은 波形이 放出된다 하고 이 스펙토르를 求하여 보기로 한다.

단 여기서 $P_r\{X(t)=1\}=P_r\{X(t)=-1\}=\frac{1}{2}$ 로 한다. 따라서 $E\{X^2(t)\}=1$로 하고 $X(t)$의 부호의 변화가 Poisson 過程에 따라서 일어나는 것으



로 한 것이다. 즉, t時間에 n回 부호 變化가 일어나는 確率은 $P_n(t)=(at)^n e^{-at}/n!$ 으로서 주어지는 것으로한다. 따라서 부호변화가

t시간에 奇數回 일어나는 確率

$$P_{(1)}(t)=\sum_{k=0}^{\infty}P_{2k+1}(t)=\frac{1}{2}(1-e^{-2at})$$

t시간에 偶數回 일어나는 確率

$$P_{(2)}(t)=\sum_{k=0}^{\infty}P_{2k}(t)=\frac{1}{2}(1+e^{-2at})$$

따라서 $\rho(\tau)=E\{X(t)X(t+\tau)\}$

$$=1\cdot P_{(2)}(\tau)+(-1)P_{(1)}(\tau)=e^{-2a\tau}$$

이경우

$$v(\lambda)=\frac{1}{\pi}\int_{0}^{\infty}\rho(\tau)\cos\tau\lambda d\tau$$ 의 式을 적용하면

<51 P에 繼續>