EL4005 Principios de Comunicaciones

Clase No.8: Procesos ARMA y Procesos de Markov

Patricio Parada

Departamento de Ingeniería Eléctrica
Universidad de Chile

9 de Noviembre de 2011

Contenidos de la Clase (1)

Otros Modelos de Procesos Aleatorios

- Existen dos modelos elementales que se utilizan para introducir “memoria” en un proceso, de manera diferente a los procesos de Markov.
    - Proceso auto-regresivo (AR)
    - Proceso de media móvil (MA: *moving average*)
- Estos modelos son empleados en forma extensiva en filtros digitales, análisis de secuencias aleatorias (series de tiempo), sistemas de control, extracción de características, etc.

Operador Backshift B

- Antes introducir los modelos vamos a introducir el operador B:

$$BX_n = X_{n-1} \tag{1}$$

- Por inducción es directo de mostrar que:

$$B^k X_n = X_{n-k} \quad k \in \mathbb{Z} \tag{2}$$

Proceso Auto-Regresivo (1)

- Un proceso es auto-regresivo de orden p - denotado como $AR(p)$ - si existe una dependencia directa entre el valor de una muestra en un instante arbitrario n y sus p valores anteriores:

$$X_n = \sum_{i=1}^{p} \alpha_i X_{n-i} + W_n \tag{3}$$

donde W_n es un proceso de ruido aditivo independiente de X.

- Si usamos la notación con el operador B entonces:

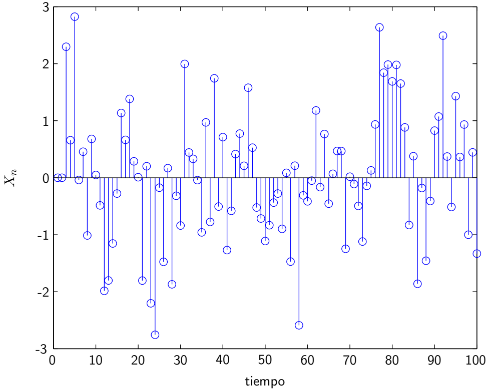
$$X_n = \sum_{i=1}^{p} \alpha_i B^i X_n + W_n$$

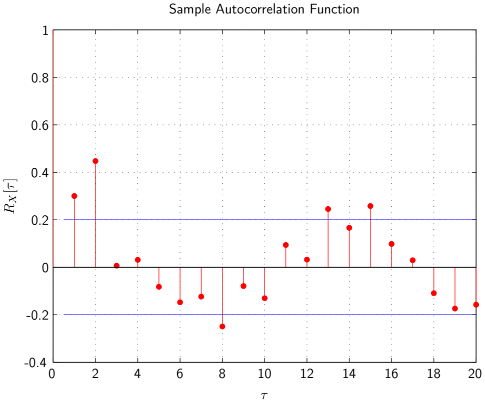
$$\Rightarrow \phi(B) X_n = W_n$$

Proceso Auto-Regresivo (2)

- donde
$$\phi(B) = 1 - \sum_{i=1}^{p} \alpha^i B^i. \tag{4}$$
- En la práctica hay dos formas en que uno puede utilizar un proceso auto-regresivo:
  - Si los parámetros del modelo $(\alpha_1, \ldots, \alpha_p)$ y el orden p son conocidos, entonces el modelo puede ser utilizado para generar un proceso auto-regresivo.
  - Ejemplo:
$$X_n = 0{,}1X_{n-1} + 0{,}5X_{n-2} + W_n \tag{5}$$
    con $W_n \sim \mathcal{N}(0, 1)$ y $X_0 = 0$.

Proceso Auto-Regresivo (3)





Proceso Auto-Regresivo (4)

- Desde el punto de vista de sistemas, un proceso auto-regresivo no es más que la salida de un sistema LTI sin ceros y cuyos polos son las raices del polinomio

$$\phi(z) = 1 - \sum_{i=1}^{p} \alpha_i z^i. \tag{6}$$

- Si los zeros de $\phi(z)$ están dentro del círculo unitario, entonces el proceso será estacionario.
- La segunda forma en que un modelo auto-regresivo puede ser empleado es para ser "ajustado" a una serie de observaciones de un proceso.
- Esta tarea es muy común en la práctica y su solución no es única.

Ajustes de Procesos Auto-Regresivos

- P: ¿Qué queremos decir con el término “ajustar”?
- R: Simplemente, que queremos determinar los parámetros del modelo p y $(\alpha_1, \ldots, \alpha_p)$ que mejor “explican” las observaciones.
- El criterio de ajuste más común se basa en la **minimización del error cuadrático medio** entre el modelo y los observaciones.
- El problema se expresa de la siguiente forma:

$$\min_{p \in \mathbb{N}} \min_{(\alpha_1, \ldots, \alpha_p) \in \mathbb{R}^p} \mathbb{E}[(X_n - \sum_{i=1}^{p} \alpha_i X_{n-i})^2]. \tag{7}$$

Coeficientes Óptimos $(\alpha_1, \ldots, \alpha_p)$ (1)

- Para encontrar los coeficientes que minimizan el error cuadrático medio podemos utilizar técnicas básicas de optimización.
- Sea $\epsilon_n = X_n + \sum_{i=1}^{p} \alpha_i X_{n-i}$. Luego

$$\mathbb{E}[\epsilon_n^2] = \mathbb{E}[X_n^2] - 2\sum_{i=1}^{p} \alpha_i \mathbb{E}[X_n X_{n-i}] + \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_i \alpha_j \mathbb{E}[X_{n-i} X_{n-j}] \qquad (8)$$

- Notar que $\mathbb{E}[X_n^2] = R_X[n;n]$, $\mathbb{E}[X_n X_{n-i}] = R_X[n; n-i]$ y $\mathbb{E}[X_{n-i} X_{n-j}] = R_X[n-i; n-j]$.

Coeficientes Óptimos $(\alpha_1, \ldots, \alpha_p)$ (2)

- Luego, en el óptimo se cumple que $\dfrac{\partial \mathbb{E}[\epsilon^2]}{\partial \alpha_k} = 0$:

$$\frac{\partial \mathbb{E}[\epsilon^2]}{\partial \alpha_k} = -2R_X[n; n-k] + 2\sum_{i=1}^{p} \alpha_i R_X[n-k; n-i] \tag{9}$$

- Si el proceso $\{X_n : n \in \mathbb{N}\}$ es estacionario en sentido amplio, las expresiones dejan de depender de n y se simplifican a lo siguiente:

$$\Rightarrow R_X[k] = \sum_{i=1}^{p} \alpha_i R_X[k-i], \quad k = 1, \ldots, p.$$

Coeficientes Óptimos $(\alpha_1, \ldots, \alpha_p)$ (3)

- Si expresamos estas ecuaciones en forma matricial obtenemos lo siguiente:

$$\underbrace{\begin{bmatrix} R_X[0] & R_X[-1] & \ldots & R_X[1-p] \\ R_X[1] & R_X[0] & \ldots & R_X[2-p] \\ & \ddots & & \\ R_X[k-1] & R_X[k-2] & \ldots & R_X[k-p] \\ & & \ddots & \\ R_X[p-1] & R_X[p-2] & \ldots & R_X[0] \end{bmatrix}}_{\mathbf{R}_X} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \\ \vdots \\ \alpha_p \end{bmatrix} = \underbrace{\begin{bmatrix} R_X[1] \\ R_X[2] \\ \vdots \\ R_X[k] \\ \vdots \\ R_X[p] \end{bmatrix}}_{\mathbf{r}_X} \quad (10)$$

Coeficientes Óptimos $(\alpha_1, \ldots, \alpha_p)$ (4)

- Notemos que por definición la función de autocorrelación es simétrica, pues

$$R_X[n;m] = \mathbb{E}[X[n]X[m]] = \mathbb{E}[X[m]X[n]] = R_X[m;n]$$

- Luego la matriz $\mathbf{R}_X$ es simétrica y puede ser diagonalizada - y por lo tanto invertida.
- Finalmente

$$\boldsymbol{\alpha}^* = \mathbf{R}_X^{-1}\mathbf{r}_X. \tag{11}$$

Análisis de la solución

- La solución encontrada para $\boldsymbol{\alpha}^*$ depende de dos elementos:
  - Que la función de autocorrelación $R_X[k]$ sea conocida para $k = 0, \dots, p$.
  - Del condicionamiento de la matriz $\mathbf{R}_X$.
- Si la matriz $\mathbf{R}_X$ está mal condicionada, esto es, si uno de sus valores propios es muy cercano a 0, la inversa va a estar llena de elementos pequeños excepto por los asociados a este valor propio que serán muy grandes.

Regularización del problema original (1)

- Una manera de mejorar la calidad de la solución es agregando una restricción sobre la magnitud del vector $\boldsymbol{\alpha}$.
- Esto lleva a reescribir el problema como

$$\min_{p \in \mathbb{N}} \min_{\substack{(\alpha_1, \dots, \alpha_p) \in \mathbb{R}^p \\ \|\boldsymbol{\alpha}\|^2 < A}} \mathbb{E}[(X_n - \sum_{i=1}^{p} \alpha_i X_{n-i})^2]. \tag{12}$$

donde A es una constante real positiva grande.

Regularización del problema original (2)

- El procedimiento de solución es el mismo que utilizamos para derivar las expresiones (11) pero se debe agregar al lagrangiano el término

$$+\theta(\|\boldsymbol{\alpha}\|^2 - A) \tag{13}$$

donde θ es el multiplicador de Lagrange asociado a la restricción.

- En la práctica, es preferible utilizar un término de **regularización** que permita, por ejemplo, controlar la magnitud del vector $\boldsymbol{\alpha}$.
- Regularización es un proceso que permite transformar un problema mal condicionado en uno que está bien condicionado, incorporando información adicional al problema.

Regularización del problema original (3)

- La forma más típica de regularización se conoce como **Regularización de Tikhonov** y modifica la función objetivo original del problema así:

$$\text{f.o.} = \mathbb{E}[(X_n - \sum_{i=1}^{p} \alpha_i X_{n-i})^2] \to \mathbb{E}[(X_n - \sum_{i=1}^{p} \alpha_i X_{n-i})^2] + \|\Gamma \boldsymbol{\alpha}\|^2 \tag{14}$$

  donde Γ es una matriz de transformación.
- En la mayoría de las aplicaciones $\Gamma = \gamma \mathbf{I}$ lo que simplifica la expresión.
- γ es un parámetro que uno controla a voluntad para obtener una solución que minimiza el error cuadrático medio condicionado a que la magnitud de $\boldsymbol{\alpha}$ permanezca acotada.

Regularización del problema original (4)

- Notemos que la nueva función objetivo tiene la forma:

$$\begin{aligned} f(\boldsymbol{\alpha}) &= \mathbb{E}[X_n^2] - 2\boldsymbol{\alpha}^T \mathbf{r}_X + \boldsymbol{\alpha}^T R_X \boldsymbol{\alpha} + (\Gamma\boldsymbol{\alpha})^T \Gamma\boldsymbol{\alpha} \\ &= \mathbb{E}[X_n^2] - 2\boldsymbol{\alpha}^T \mathbf{r}_X + \boldsymbol{\alpha}^T (R_X + \Gamma^T\Gamma)\boldsymbol{\alpha} \end{aligned}$$

- En el óptimo, $\nabla f(\boldsymbol{\alpha}) = 0$, es decir:

$$\boldsymbol{\alpha}^* = (\mathbf{R}_X - \Gamma^T\Gamma)^{-1}\mathbf{r}_X. \tag{15}$$

- Notar que si $\Gamma = \gamma I$, entonces $\boldsymbol{\alpha}^* = (\mathbf{R}_X - \gamma^2 I)^{-1}\mathbf{r}_X$.

Entrenamiento y Validación de los resultados (1)

- En la práctica ni la matriz $\mathbf{R}_X$ ni el vector $\mathbf{r}_X$ son conocidos.
- Uno deben estimarlos a partir de datos y empleando algunas suposiciones básicas que uno asume como verdaderas (como que el proceso sea estacionario).
- Esto puede llevar a errores en el resultado final, y que el modelo que se construye tenga un desempeño bajo.
- Una manera de resolver esta situación es estudiando el desempeño del modelo en dos escenarios distintos:
  - Explicando los datos utilizados para estimar $\mathbf{R}_X$ y $\mathbf{r}_X$.
  - Explicando datos nuevos, no utilizados en el proceso anterior.

Entrenamiento y Validación de los resultados (2)

- Estas dos etapas reciben el nombre de **entrenamiento** y **validación** del modelo, y son parte fundamental de una gran cantidad de algoritmos de aprendizaje estadístico.
- Para poder implementar esto en la práctica uno debe dividir el conjunto de datos en dos subconjuntos:
  - Datos de Entrenamiento
  - Datos de Validación
- En el caso del ajuste de un modelo $AR(p)$ vamos a observar en la práctica que si los datos provienen de un proceso estacionario, entonces, el error cuadrático medio de aproximación se reduce a medida que aumenta p.

Entrenamiento y Validación de los resultados (3)

- Sin embargo, el desempeño del mismo modelo con los datos de validación tenderá a empeorar con a medida que aumenta el orden del modelo.
- Este fenómeno se denomina **sobreajuste** y es un fenómeno no deseado.
- Para evitarlo, un criterio de diseño empleado habitualmente es seleccionar p de forma que se minimice conjuntamente tanto el error cuadrático medio para los datos de entrenamiento como para los de validación.

Proceso de Media Móvil (1)

- Un proceso de media móvil de orden q - denotado como $MA(q)$ - es un proceso estocástico X_n que "suaviza" o promedia los valores de otro proceso W_n.

$$X_n = \sum_{j=-m}^{q-1-m} \beta_j W_{n-j}. \tag{16}$$

para algún $m \in \mathbb{N}$.

- Mediante este promedio se logra remover las variaciones locales del proceso original (el ruido de la señal) y rescatar la tendencia o valor medio del proceso.
- Notemos que si el proceso W_n no es estacionario, entonces veremos que este promedio debería variar considerablemente en el tiempo (lo que justifica su nombre).

Proceso de Media Móvil (2)

- Al igual que un proceso auto-regresivo, el proceso de media móvil puede ser descrito en forma compacta mediante un operador:

$$X_n = \theta(B) W_n \tag{17}$$

con

$$\theta(B) = \sum_{j=-m}^{q-1-m} \beta_j B^j. \tag{18}$$

para algún $m \in \mathbb{N}$.

Proceso de Media Móvil (3)

- El modelo de media móvil es general, y no restringe sobre los parámetros β_j ni sobre su causalidad o falta de ella.
- Este enfoque es el que toman Box & Jenkins en su libro *Time Series Analysis: Forecasting and Control* (ver referencia bibliográfica al final).
- Sin embargo, en aplicaciones en finanzas y estadística es habitual imponer como condición que

$$\sum_{j=-m}^{q-1-m} \beta_j = 1. \tag{19}$$

- En este curso adoptaremos esta última convención, aunque deben tener presente que en general esta restricción no tiene porque cumplirse.

Proceso de Media Móvil (4)

- Otro tema que merece mención es si el modelo es causal o no.
- Esto tiene relación con el valor de m.
- Si $m = 1$, el modelo será causal y si $m > 1$ será no causal.
- Nuevamente, Box & Jenkins imponen como condición la causalidad, i.e.

$$X_n = \sum_{j=1}^{q} \beta_j W_{n-j} \tag{20}$$

- Por otro lado, en finanzas y estadística se tiende a imponer la no causalidad:

$$X_n = \sum_{j=-(q-1)/2}^{(q-1)/2} \beta_j W_{n-j} \tag{21}$$

para q impar.

Proceso de Media Móvil (5)

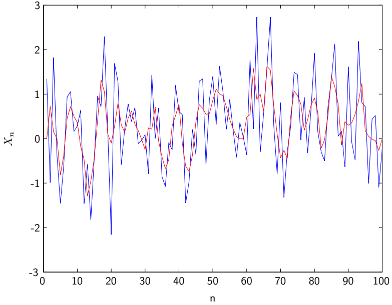
- Notar que si q es par entonces se define como proceso $MA(q)$ al siguiente proceso:

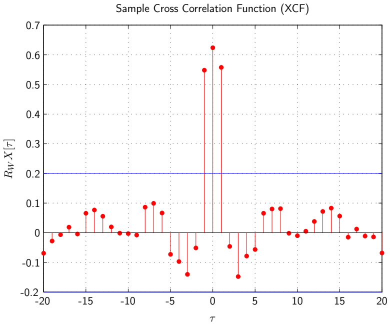
$$X_n = \frac{1}{2}\Big(\sum_{j=-q/2}^{q/2-1} \beta_j W_{n-j} + \sum_{j=-q/2+1}^{q/2} \beta_j W_{n-j} \Big)$$
$$= \frac{1}{2}\beta_{-q/2} W_{n+q/2} + \beta_{-q/2+1} W_{n+q/2-1} + \ldots + \beta_{q/2+1} W_{n-q/2+1} + \frac{1}{2}\beta_{q/2} W_{n-q/2}$$

Proceso de Media Móvil (4)

- Ejemplo: $MA(3)$

$$X_n = \frac{1}{2}W_{n-1} + \frac{1}{3}W_n + \frac{1}{3}W_{n+1} \tag{22}$$

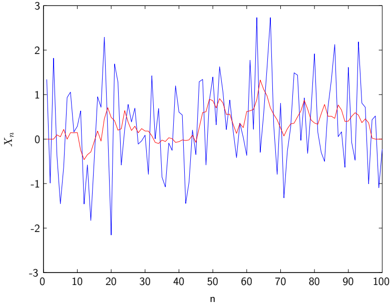


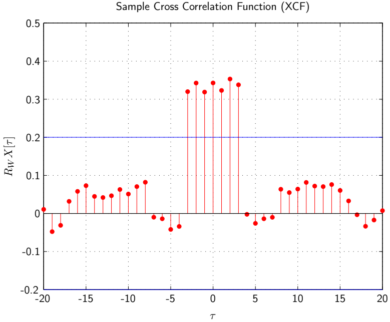


Proceso de Media Móvil (5)

- Ejemplo: $MA(7)$

$$X_n = \frac{1}{7}\sum_{j=-3}^{3} W_{n-j} \tag{23}$$

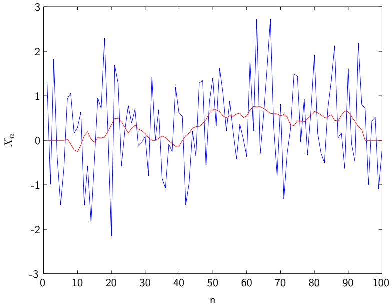


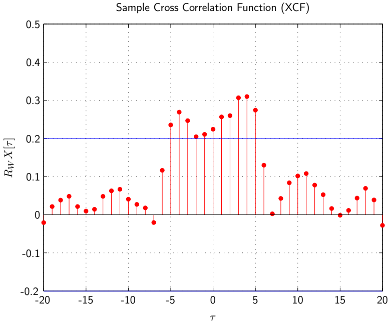


Proceso de Media Móvil (6)

- Ejemplo: $MA(12)$

$$X_n = \frac{1}{24} W_{n+6} + \frac{1}{12} \sum_{j=-5}^{5} W_{n-j} + \frac{1}{24} W_{n-6} \qquad (24)$$





Proceso Auto-Regresivo de Media Móvil (1)

- Un proceso auto-regresivo de media móvil de orden (p, q) -denotado como $ARMA(p, q)$ - es un proceso estocástico que combina los efectos de suavizado del ruido con la memoria propia del proceso bajo estudio.
- La expresión del proceso es la siguiente:

$$X_n = \sum_{i=1}^{p} \alpha_i X_{n-i} + \sum_{j=1}^{q} \beta_j W_{n-j} \tag{25}$$

Proceso Auto-Regresivo de Media Móvil (2)

- Este proceso puede ser representado en forma compacta de igual forma que los procesos AR y MA por si solos.
- Si definimos los operadores:

$$\phi(B) = 1 - \sum_{i=1}^{p} \alpha_i B^i \text{ y } \theta(B) = \sum_{j=1}^{q} \beta_j B^j \tag{26}$$

entonces

$$\phi(B) X_n = \theta(B) W_n. \tag{27}$$

Procesos de Markov (1)

- Un proceso de Markov corresponde a un proceso aleatorio dependiente o "con memoria".
- Es similar a los procesos auto-regresivos, excepto que la caracterización del proceso se realiza con las distribuciones de probabilidad del proceso en lugar de los valores que toma el proceso mismo.
- Los procesos de Markov puede ser de tiempo continuo o de tiempo discreto y pueden tomar valores continuos o discretos.
- En el caso particular que el proceso sea de tiempo discreto y tome valores discretos recibe el nombre de **Cadena de Markov**.

Procesos de Markov (2)

- Un proceso continuo de Markov de primer orden satisface que

$$f_X(x_n|x_{n-1}, \ldots, x_1; t_n, \ldots, t_1) = f_X(x_n|x_{n-1}; t_n, t_{n-1}) \tag{28}$$

  para todo $x_1, x_2, \ldots, x_n$ y $t_1 < t_2 < \ldots, < t_n$ y todo entero $n > 0$.

- Un proceso discreto de Markov de primer orden satisface que

$$P_X(x_n|x_{n-1}, \ldots, x_1; t_n, \ldots, t_1) = P_X(x_n|x_{n-1}; t_n, t_{n-1}) \tag{29}$$

  para todo $x_1, x_2, \ldots, x_n$ y $t_1 < t_2 < \ldots, < t_n$ y todo entero $n > 0$.

Ejemplo (1)

- Consideremos el caso de dos computadores independientes conectados a un router inalámbrico.
- Podemos modelar el número de computadores transmitiendo simultáneamente como un sistema con tres estados:
  - 0 computadores transmitiendo.
  - 1 computador transmitiendo.
  - 2 computadores transmitiendo.

Ejemplo (2)

- Consideremos las siguientes condiciones:
  - La duración de una trasmisión se distribuye de manera exponencial en el tiempo, con $\lambda > 0$ (el mismo para ambos equipos).
  - La duración de cada silencio se distribuye de manera exponencial en el tiempo, con $\mu > 0$.
  - Las duraciones de las transmisiones y silencios son independientes entre ellas y entre computadores.

Ejemplo (3)

- Sea $X(t)$ el estado del sistema en un tiempo t.
- Entonces $X = \{X(t) : t \in \mathbb{R}\}$ es un proceso de Markov de tiempo continuo.
- Por qué? Porque los tiempos entre transiciones se distribuyen exponecialmente y estos no tienen memoria.
- La transición de 0 a 1 ocurre cuando uno de los computadores comienza a transmitir.
- Denotaremos por
  - T_1: instante en que el PC_1 comienza a transmitir.
  - T_2: instante en que el PC_2 comienza a transmitir.

Ejemplo (4)

- Luego, la transición ocurre en

$$\tau_{\ 0\to 1} = \text{mín}\{T_1, T_2\} \tag{30}$$

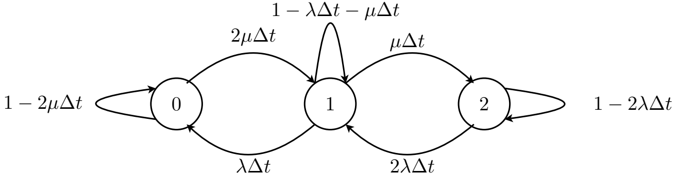
- Pero el mínimo entre dos exponenciales de parámetro μ es otra exponencial de parámetro 2μ.
- Luego, la transición de $0 \to 1$ ocurre a una tasa 2μ por segundo.
- La transición de $1 \to 2$, ocurre cuando el segundo equipo comienza a transmitir, es decir, lo hace una tasa μ.

Ejemplo (5)

- Similarmente, la transición de $1 \to 0$ ocurre a una tasa λ, ya que $1/\lambda$ es la duración promedio de una transmisión.
- Finalmente, la transición de $2 \to 1$ ocurre cuando uno de los dos equipos deja de transmitir, es decir, con tasa 2λ.

Diagrama de Transisión de Corto Plazo

- Un proceso de Markov de tiempo continuo puede ser representado de manera gráfica mediante una figura como la presentada en la figura



- Los pesos de cada rama son las probabilidades de transición de corto plazo, es decir, cuando $\Delta t \to 0$.

Cadenas de Markov (1)

- Una cadena de Markov es una secuencia aleatoria que toma valores discretos en un espacio de estado $\mathcal{S}$.
- La probabilidad condicional que define a la cadena está dada por:

$$P_X(X[n]|X[n-1], X[n-2], \ldots, X[n-N]) = P_X(X[n]|X[n-1]) \tag{31}$$

para todo n, todos los valores $X[k]$, y para cualquier entero $N > 1$.

Cadenas de Markov (2)

- Si el espacio de estado tiene cardinalidad finita - digamos M - las probabilidades condicionales pueden ser escritas en una matriz $\mathbf{P}[n]$ cuyas entradas son de la forma:

$$\begin{aligned} p_{ij}[n] &= P_{X[n]|X[n-1]}(j|i) \\ &= P(X[n] = j|X[n-1] = i), \end{aligned}$$

  para $1 \leq i, j \leq M$.

- Las filas de la matriz deben sumar 1.

Distribución de probabilidad de los estados de la cadena

- Para poder determinar el valor de la distribución de probabilidad de los estados de la cadena en un instante arbitrario es necesario conocer dos elementos:
  - La distribución de probabilidad de estados en $n = 0$, que denotaremos por $\mathbf{p}[0]$,
  - La matriz de transición de estados $\mathbf{P}[n]$ para todo $n > 0$.
- Luego

$$
\begin{aligned}
\mathbf{p}[1] &= \mathbf{p}[0]\mathbf{P}[1] \\
\mathbf{p}[2] &= \mathbf{p}[1]\mathbf{P}[2] \\
&\vdots \\
\mathbf{p}[n] &= \mathbf{p}[n-1]\mathbf{P}[n].
\end{aligned}
$$

Cadenas de Markov Homogéneas (1)

- Si la matriz $\mathbf{P}[n]$ es independiente del tiempo, decimos que la cadena es **homogénea**, y por lo tanto

$$\mathbf{p}[n] = \mathbf{p}[0]\mathbf{P}^n. \tag{32}$$

- Un caso particular en el que estamos interesados es en determinar la distribución de probabilidad del régimen permanente, si es que existe:

$$\mathbf{p}[\infty] = \lim_{n\to\infty} \mathbf{p}[n] \tag{33}$$

- De existir el límite, entonces tendríamos que

$$\mathbf{p}[\infty] = \mathbf{p}[\infty]\mathbf{P}. \tag{34}$$

Cadenas de Markov Homogéneas (2)

- Por lo tanto, $\mathbf{p}[\infty]$ es el vector propio (por la izquierda) asociado al valor propio $\lambda = 1$ de la matriz $\mathbf{P}$.
- Se obtiene entonces
$$\mathbf{p}[\infty](\mathbf{I} - \mathbf{P}) = \mathbf{0} \tag{35}$$
lo que nos entrega $M - 1$ ecuaciones linealmente independientes para calcular $\mathbf{p}[\infty]$.
- La última ecuación se obtiene de la condición de normalización que $\mathbf{p}[\infty]$ satisface.

Aplicaciones de las Cadenas de Markov

- El modelo de cadenas de Markov es aplicable en cualquier situación donde los valores que tome un determinado proceso aleatorio sean discretos, y existe sospecha que hay condicionalidad temporal.
- Algunos ejemplos clásicos son los siguientes:
  - Modelamiento del tráfico que llega a los servidores y equipos de red en una red de computadores.
  - Modelamiento de un canal de comunicaciones punto-a-punto.
  - Modelamiento de una fuente de información con memoria.

Resumen

Hemos revisado:

- Procesos auto-regresivos, de media móvil y ARMA.
- Análisis de estimación de MSE para modelo $AR(p)$.
- Procesos de Markov.
- Cadenas de Markov.

Lecturas

- Salehi & Proakis, *Communication Systems Engineering*, Capítulo 4.
- Box, Jenkins, & Reinsel, *Time Series Analysis: Forecasting and Control*, 4th Edition, Capítulo 3.
- Stark & Woods, *Probability and Random Processes with Applications to Signal Processing*, 3rd. Edition, Sección 9.1.