STAT 3911 - STOCHASTIC PROCESSES

ANDREW TULLOCH

Contents

1. Stochastic Processes

Definition 1.1 (Markov process)**.** A process $X = \{X_t : t \in T\}$ is a **Markov process** if

$$\mathbb{P}(X(t) \leq x \,|\, X(t_1) = x_1, X(t_2) = x_2, \ldots, X(t_n) = x_n) = \mathbb{P}(X(t) \leq x \,|\, X(t_n) = x_n)$$

whenever $t_1 < t_2 < \cdots < t_n < t$.

In other words, a Markov process is a stochastic process satisfying the Markov property: the conditional distribution of future values is independent of the past.

Definition 1.2 (Markov chain)**.** A *Markov chain* is a stochastic process with a finite or countable state space is a Markov chain if it satisfies the Markov property. Define the transition probabilities P_{ij} by

$$P_{ij} = \mathbb{P}(X_{n+1} = j \,|\, X_n = i)$$

Theorem 1.3 (Chapman-Kolmogorov equation)**.** *Let P_{ij}^n denote the probability of going from state i to state j in n steps, that is,*

$$P_{ij}^n = \mathbb{P}(X_{n+m} = j \,|\, X_m = i) = \mathbb{P}(X_n = j \,|\, X_0 = i)$$

Then we have, for any $m, n = 0, 1, 2, \ldots$ and $i, j \in S$,

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^m P_{kj}^n.$$

In terms of matrix multiplication, this is equivalent to

$$P^{n+m} = P^n P^m$$

Proposition 1.4 (Initial distribution)**.** *At time t, let π_n denote the column vector such that $\pi_n(j) = \mathbb{P}(X_n = j)$ for all $j \in S$.*

$$\pi_n = \pi_0 P^n$$

2. Classification of states

Definition 2.1 (Classification of states)**.** Define an equivalence relation $\sim$ on the set of states such that $i \leftrightarrow j$ if and only if $P_{ij}^n > 0$ for some $n \geq 0$. Two states i and j which are accessible to each other are said to communicate and we write $i \leftrightarrow j$

Lemma 2.2. *Communication is an equivalence relation on the set of states.*

If two states communicate with each other we say they belong to the same **communicating class**.

Definition 2.3 (Irreducible)**.** A Markov chain is said to be **irreducible** if there is only one class, i.e., all states communicate with each other. A class C is closed if no state outside C can be reached from any state in C. A state i is called **absorbing** if the set $\{i\}$ is a closed class, i.e., $P_{ii} = 1$.

Definition 2.4 (Period of a state)**.** A state i is said to have **period** d_i if $P_{ii}^n = 0$ except when $n = d_i, 2d_i, 3d_i, \ldots$.

Proposition 2.5. *If $i \leftrightarrow j$ then $d_i = d_j$.*

Definition 2.6 (First transition time)**.** For any states i and j, define f_{ij}^n as the probability that starting in i, the first transition into j occurs at time n, that is

$$f_{ij}^0 = 0$$
$$f_{ij}^n = \mathbb{P}(X_n = j, X_k \neq j, k = 1, 2, \ldots, n-1 \,|\, X_0 = i)$$

Let T_j be the random time of the **first transition** into state j. Thus we have

$$\mathbb{P}(T_j = n \,|\, X_0 = i) = f_{ij}^n$$

If $X_0 = j$ then T_j is the **first return** time to j. If $X_0 = i \neq j$ then T_j is the **first hitting** time of j.

We define the quantity f_{ij} by the equality

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^n$$

f^{ij} is the probability of ever making a transition into state j, given that the process starts in state i. We also have that $f_{ij} > 0$ if and only if state j is accessible from state i.

Definition 2.7 (Recurrent states)**.** A state is **recurrent** if

$$\mathbb{P}(X_n = i \text{ for some } n \geq 1 \,|\, X_0 = i) = 1$$

A state i is recurrent if and only if $f_{ii} = 1$. A state i is transient if and only if $f_{ii} < 1$.

Proposition 2.8. *A state i is recurrent if and only if*

$$\sum_{n=1}^{\infty} P_{ii}^n = \infty$$

A state i is transient if and only if

$$\sum_{n=1}^{\infty} P_{ii}^n < \infty$$

Corollary. *If i is recurrent and $i \leftrightarrow j$ then state j is recurrent.*

Corollary. *If $i \leftrightarrow j$ and j is recurrent then $f_{ij} = 1$.*

3. Limit theorems for Markov chains

Let m_{jj} be the **mean recurrence time**, that is, the expected number of transitions until a Markov chain starting in j returns to that state, so

$$m_j j = \begin{cases} \infty & \text{if } j \text{ is transient} \\ \sum_{n=1}^{\infty} n f_{jj}^n & \text{if } j \text{ is recurrent} \end{cases}$$

Definition 3.1 (Ergodic state)**.** Let j be recurrent. Then a state j is called **positive recurrent** if $m_{jj} < \infty$ and **null recurrent** if $m_{jj} = \infty$. A positive recurrent, aperiodic state is called **ergodic**.

We have $\frac{1}{m_{jj}} = 0$ if a state j is either transient of null recurrent. If j is positive recurrent then $0 < \frac{1}{m_{jj}} < \infty$.

Theorem 3.2. *Assume $i \leftrightarrow j$. If a state j is aperiodic, then*

$$\lim_{n\to\infty} P_{ij}^n = \frac{1}{m_{jj}}$$

If a state j has period d_j, then

$$\lim_{n\to\infty} P_{ij}^n = \frac{d_j}{m_{jj}}$$

Definition 3.3 (Stationary distributions)**.** A probability distribution $\pi^\star$ is called **stationary** for for a Markov chain X with transition matrix P if

$$\pi^\star P = \pi^\star$$

We note that the stationary distribution $\pi^\star$ associated with a Markov chain is not necessarily unique, nor does it always exist.

Theorem 3.4 (Irreducible aperiodic Markov chains)**.** *Let X be an irreducible aperiodic Markov chain. Then*

- *Either all the states are transient or null recurrent; in that case we have that $\lim_{n\to\infty} P_{ij}^n = 0$ for all i, j and there exists no stationary distribution.*
- *or else all states are positive recurrent - in that case we have $\Pi_j = \lim_{n\to\infty} P_{ij}^n > 0$ for all i, j; moreover, Π_j is a stationary distribution and the stationary distribution is unique. Moreover, for all j,*

$$\Pi_j = \frac{1}{m_{jj}}$$

We note that in a **finite-state**, irreducible, aperiodic Markov chain all states are positive recurrent/ergodic, and so the stationary distribution exists and is unique.

In a **finite state**, irreducible, **periodic** Markov chain all states are positive recurrent and the unique stationary distribution exists and is unique.

4. First step analysis

Consider the Markov chain X with transition probability matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ \alpha & \beta & \gamma \\ 0 & 0 & 1 \end{bmatrix}$$

We ask the following questions.

- What state will the process ultimately be absorbed in?
- How long, on average, will it take to reach one of these states?

Let

$$T = \min\{n \geq 0 \,|\, X_n = 0 \text{ or } X_n = 2\}$$

be the random **time of absorption** for the process X. We ask the questions

$$u = \mathbb{P}(X_T = 0 \,|\, X_0 = 1)$$

and

$$v = \mathbb{E}\,[T \,|\, X_0 = 1]$$

Thus u is the probability of begin absorbed in state 0 given we start in state 1, and v is the expected time of being absorbed given we start in state 1.

Now we have the following.

$$u = 1 \cdot \alpha + u \cdot \beta + 0 \cdot \gamma$$

Similarly, we have

$$v = 1 + a \cdot 0 + \beta \cdot v + \gamma \cdot 0$$

5. Branching processes

Definition 5.1 (Branching process)**.** let $X_0 = 1$ and for every $n = 0, 1, 2, \ldots$ define

$$X_{n+1} = \sum_{i=1}^{X_n} Z_i$$

where Z_i are independently, identically distributed random variables with probability distribution $\mathbb{P}(Z = k) = P_k, k = 0, 1, 2, \ldots$. Then $X = \{X_n, n \geq 0\}$ is a branching process

We have the following formulas.

Lemma 5.2. *We have*

$$\mathbb{E}\,[E(X_n)] = \mu \mathbb{E}\,[X_{n-1}] = \mu^n$$

and

$$Var(X_n) = \begin{cases} \sigma^2 \mu^{n-1} \left(\frac{\mu^n - 1}{\mu - 1} \right) & \mu \neq 1 \\ n\sigma^2 & \mu = 1 \end{cases}$$

Proposition 5.3 (Extinction probabilities)**.** *Define u_n to be the probability of extinction at or prior to the n^{th} generation, assuming $X_0 = 1$. Then*

$$u_n = \mathbb{P}(X_n = 0 \,|\, X_0 = 1)$$

Then we have

$$u_n = \sum_{k=0}^{\infty} P_k (u_{n-1})^k$$

where $P_k = \mathbb{P}(Z = k)$.

Proposition 5.4. *Let $\Pi_0 = \lim_{n\to\infty} \mathbb{P}(X_n = 0)$ be the probability of eventual extinction. Then, defininining the probability generating functions $G_n(s) = \mathbb{E}\left[s^{X_n}\right] = \sum_{j=0}^{\infty} s^j \mathbb{P}(X_n = j)$, we have that*

- *$\Pi_0 = 1$ if and only if $\mu \leq 1$,*
- *If $\mu > 1$ then $\Pi_0 < 1$ is the smallest positive number p satisfying $G(p) = p$ - that is,*

$$\Pi_0 = G(\Pi_0)$$

where $G(p) = \sum_{j=0}^{\infty} p^j P_j$.

6. Poisson process

Definition 6.1 (Poisson process)**.** A process $N = \{N(t), t \geq 0\}$ is a **Poisson process** if

- $N(0) = 0$,
- N is a process of independent increments.
- The number of events in any interval of length t has a Poisson distribution with parameter λt, that is

$$\mathbb{P}(N(t+s) - N(s) = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

The positive paramenter $\lambda > 0$ is called the **intensity** of the process. We have that $\mathbb{E}[N(t)] = \lambda t$ and $\mathrm{Var}(N(t)) = \lambda t$

Lemma 6.2. *The sum of two Poisson processes of intensities λ and μ is a Poisson process of intensity $\lambda + \mu$.*

7. Sojourn and waiting times

Definition 7.1 (Soujourn time)**.** For $n = 0, 1, 2, \ldots$, the **sojourn time** in state n equals

$$S_n = \inf\{t \in [0, \infty) \,|\, N(t + S_{n-1}) - N(S_{n-1}) = 1\}$$

Lemma 7.2. *The sojourn times S_n are independently, identically distributed exponential variables with parameter λ.*

Definition 7.3 (Waiting time)**.** For $n = 0, 1, 2, \ldots$, the **waiting time** until the n^{th} event equals

$$W_n = \inf\{t \in [0, \infty) \,|\, N(t) = n\}$$

Lemma 7.4. *If X_i are independently, identically distributed exponential random variables with parameter λ then $\sum_{i=1}^{n} X_i$ is distributed with a gamma distribution with parameters n, λ, that is, $\Gamma(n, \lambda)$.*

Proposition 7.5. *Given $N(t) = 1$, the waiting time $W_1 = S_0 = T_1$ has uniform distribution on $[0, t]$*

Theorem 7.6. *Given that $N(t) = 1$, the waiting times $(W_1, \ldots, W_n)$ have the same conditional joint probability density of n independently, identically distributed uniformly distributed random variables on the interval $[0, t]$.*

Lemma 7.7. *If $X_i, i = 1, 2$ are independent exponential random variables with parameters λ_i, then*

$$\mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

8. Continuous-time Markov chains

Consider a continuous time Markov process $X = \{X(t), t \geq 0\}$ where transition probabilities

$$P_{ij}(t) = \mathbb{P}(X(t + s) = j \,|\, X(s) = i)$$

are independent of s. This is a **time-homogenous**, or **stationary**, process.

Definition 8.1 (Pure birth process)**.** A **pure birth process** is a Markov process satisfying the following conditions.

- $P_{k,k+1}(h) = \lambda_k h + o(h)$,
- $P_{k,k}(h) = 1 - \lambda_k h + o(h)$,
- $\mathbb{P}(X(t + h) - X(t) < 0 \,|\, X(t) = k) = 0$,
- $X(0) = 0$ or $X(0) = 1$

Example 8.2. Let us review some interesting special cases for birth intensities λ_k.

- **Poisson process:** $\lambda_k = \lambda$ for all k.
- **Simple birth:** $\lambda_k = k\beta$. This model gives the probabiities

$$\mathbb{P}(X(t + h) - X(t) = m \,|\, X(t) = k) = \begin{cases} 1 - k\beta h + o(h) & m = 0 \\ k\beta h + o(h) & m = 1 \\ o(h) & \text{otherwise} \end{cases}$$

- **Simple birth with immigration:** $\lambda_k = k\beta + \nu$.

9. Differential equations for marginal probabilities

Consider a pure birth process. Let $P_n(t) = \mathbb{P}(X(t) = n)$. Assume $P_0(0) = 1$.

Theorem 9.1. *The functions P_n satisfy the following system of ordinary differential equations.*

$$P_0'(t) = -\lambda_0 P_0(t)$$
$$P_n'(t) = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t)$$

Lemma 9.2. *The marginal probability functions P_n satisfy*

$$P_0(t) = e^{-\lambda_0 t}$$

and

$$P_n(t) = \lambda_{n-1} e^{-\lambda_n t} \int_0^t e^{\lambda_n s} P_{n-1}(s) ds$$

As above, define S_k to be the time between the kth and $(k+1)$th birth. Then we have the following lemma.

Lemma 9.3. *The random variables S_k are independent and S_k has an exponential distribution with parameter λ_k*

10. Birth and death processes

Definition 10.1 (Pure birth process)**.** A **birth and death process** is a Markov process satisfying the following conditions.

- $P_{k,k+1}(h) = \lambda_k h + o(h)$,
- $P_{k,k}(h) = 1 - (\lambda_k + \mu_k h + o(h)$,
- $P_{k,k-1}(h)\mu_k h + o(h)$
- $\mathbb{P}(X(t+h) - X(t) < 0 \,|\, X(t) = k) = 0$,
- $X(0) = 0$ or $X(0) = 1$

Proposition 10.2 (Chapman-Kolmogorov equation)**.**

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ij}(t) P_{kj}(s).$$

Proposition 10.3 (Sojourn times)**.** *We wish to calculate the **sohourn time** in state i, denoted S_i. Defining $\mathbb{P}(S_i \geq t) = D_i(t)$. We have*

$$D_i(t) = e^{-(\lambda_i + \mu_i)t}$$

Thus S_i is exponential distributed with parameter $\lambda_i + \mu_i$.

Proposition 10.4. *The process stays in a given state i for a random length of time whose distribution function is exponential with parameter $\lambda_i + \mu_i$. The process leaves state i by entering state $i+1$ or $i-1$ with probability $\frac{\lambda_i}{\lambda_i+\mu_i}$ and $\frac{\mu_i}{\lambda_i+\mu_i}$.*

Proposition 10.5. *Transition probabilities $P_{ij}(t)$ satisfy a system of ordinary differential equations, known as the* ***backward Kolmogorov equations.*** *For a fixed j,*

$$P'_{0j}(t) = -\lambda_0 P_{0j}(t) + \lambda_0 P_{1j}(t)$$
$$P'_{ij}(t) = \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t) + \lambda_i P_{i+1,j}(t)$$

with initial conditions $P_{ij}(0) = \delta_{ij}$. In matrix notation,

$$P'(t) = AP(t),$$

with $P(0) = I$.

Proposition 10.6. *Transition probabilities $P_{ij}(t)$ satisfy a system of ordinary differential equations, known as the* ***forward Kolmogorov equations.*** *For a fixed j,*

$$P'_{i0}(t) = -\lambda_0 P_{i,0}(t) + \mu_1 P_{i,1}(t)$$
$$P'_{ij}(t) = \lambda_{j-1} P_{i,j-1}(t) - (\lambda_i + \mu_i) P_{ij}(t) + \mu_{j+1} P_{i,j+1}(t)$$

with initial conditions $P_{ij}(0) = \delta_{ij}$. In matrix notation,

$$P'(t) = P(t)A,$$

with $P(0) = I$.

From the forward Kolmogorov equation, we can find a system of ordinary differential equations for the marginal probabilities $P_k = \mathbb{P}(X(t) = j)$. We have

$$P'_0(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t)$$
$$P'_j(t) = \lambda_{j-1} P_{j-1}(t) - (\lambda_j + \mu_j) P_j(t) + \mu_{j+1} P_{j+1}(t)$$

with $P_j(0) = \pi_0(j)$. In matrix notation,

$$\pi'_t = \pi_t A$$

11. Time Series Analysis

Definition 11.1 (Time series)**.** Let $\{X_t\}$ be a time series with $\mathbb{E}\left[X_t^2\right] < \infty$. The **mean function** for $\{X_t\}$ is

$$\mu_X(t) = \mathbb{E}\left[X_t\right]$$

The **covariance function** of $\{X_t\}$ is

$$\gamma_X(r,s) = \mathrm{Cov}\left(() X_r, X_s\right) = \mathbb{E}\left[(X_r - \mu_X(r)(X_s - \mu_X(s)))\right]$$

Definition 11.2 (Stationary)**.** A time series $\{X_t\}$ is **weakly stationary** if

- $\mu_X(t)$ is independent of t,
- $\gamma_X(t+h,t)$ is independent of t for each h.

Definition 11.3 (Autocovariance function)**.** Let $\{X_t\}$ be a stationary time series. The **autocovariance function** (ACVF) of $\{X_t\}$ at lag h is

$$\gamma_X(t) = \mathrm{Cov}\left(X_{t+h}, X_t\right)$$

The **autocorrelation function** (ACF) of $\{X_t\}$ at lag h is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Cor}(X_{t+h}, X_t)$$

Definition 11.4 (Sample autocovariance function)**.** The **sample autocovariance function** is

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x})$$

The **sample autocorrelation function** is

$$\overline{\rho}_X(h) = \frac{\overline{\gamma}_X(h)}{\overline{\gamma}_X(0)}$$

12. ARMA(p,q) Models

Definition 12.1 (ARMA(p,q) process)**.** A time series $\{X_t\}$ is an **ARMA(p,q)** process if $\{X_t\}$ is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_t + \cdots + \theta_q Z_{t-q}$$

where $\{Z_t\} \sim WN(0,\sigma^2)$ and the polynomials $\theta(z)$ and $\phi(z)$ have no common factors.

Theorem 12.2 (Stationary)**.** *A stationary solution $\{X_t\}$ of the equations exists, and is the unique stationary solution if and only if*

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$$

for all $|z| = 1$

Theorem 12.3 (Causality)**.** *An ARMA(p, q) process is **causal**, if there exists constants* $\{\psi_j\}$ *with*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

for all t.

Causality is equivalent to the condition

$$\phi(z) = \phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$$

for all $|z| \leq 1$.

The coefficients ψ_j can be determined by the identity

$$\psi(z) = \psi_0 + \psi_1 z + \cdots = \frac{\theta(z)}{\phi(z)}$$

Definition 12.4 (Invertibility)**.** An ARMA(p, q) process $\{X_t\}$ is **invertible** if there exist constants π_j such that

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

Invertibility is equivalent to the condition

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

for all $|z| \leq 1$.

The coeffients π_j can be determined by the identity

$$\pi(z) = \pi_0 + \pi_1 z + \cdots = \frac{\phi(z)}{\theta(z)}$$

Definition 12.5 (Partial autocorrelation function)**.** The **partial autocorrelation function** $\alpha(\cdot)$ satisfies the following equations.

$$\alpha(0) = 1$$
$$\alpha(1) = \rho(1)$$
$$\alpha(2) = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}$$

and in general, is the last component of

$$\phi_h = \Gamma_h^{-1} \gamma_y$$

where Γ_h is the covariance matrix and γ_y is a vector of covariances.

Theorem 12.6. *The PACF of an AR(p) process is given by*

$$\alpha(p) = \phi_p$$

and

$$\alpha(h) = 0$$

for $h > p$

Theorem 12.7 (Stationary AR(2) process)**.** *An AR(2) process* $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$ *is stationary if and only if the following all hold.*

$$\alpha_1 + \alpha_2 < 1$$
$$\alpha_1 - \alpha_2 > -1$$
$$\alpha_2 > -1$$

13. ARIMA Models

Definition 13.1 (ARIMA process)**.** If d is a nonnegative integer, then $\{X_t\}$ is an **ARIMA**(p, d, q) **process** if $Y_t = (1 - B)^d X_t$ is a causal ARMA(p, q) process.

14. Model selection

To select a model, we must first transform the data until we are examining a stationary process. Then, we consider the ACF and PACF functions. A few general points.

- A slowly decaying (slower than exponential decay) ACF indicates the data needs to be differenced, i.e., that an ARIMA(p, d, q) model should be considered.
- A PACF that cuts out sharply after lag p indicates an AR(p) model. This can be quantified by Quenoille's Test.

  Theorem 14.1 (Quenouille's Test)**.** *If a process is AR(m), and if N is large, then*

  $$\hat{\pi}_k \sim \mathcal{N}(0, \frac{1}{N}), \quad k > m$$

  Thus any $\hat{\pi}_k$ *satisfying* $|\hat{\pi}_k| > \frac{2}{\sqrt{N}}$ *for* $k > m$*, indicates disagreement with* H_0 *at the* 5% *significance level.*

- Any MA process must have $\rho(1) < 0.5$
- An MA(q) process must have $\rho(k) = 0$ for all $k > q$.
- An AR(p) process will have the ACF function exhibiting exponential decay.

To test the fit of a model, we use the Box-Pierce test and consider the AIC statistic.

Theorem 14.2 (Box-Pierce test)**.** *Assume an ARIMA*(p, d, q) *model has been fitted to time series. Consider the series of residual* $\{s_t\}$*. Then, for a fixed maximum lag* K*, we consider the statistic*

$$BP = N \sum_{k=1}^{K} \hat{r}_k^2$$

where $\hat{r}_k^2$ is the squared sample autocorrelation coefficient for the residual time series.

This is distributed with according to a χ^2-distribution with $K-p-q$ degrees of freedom. This statistic is then used to test the hypothesis that the none of the autocorrelation coefficients for the residuals are different from zero against the alternate that at least one autocorrelation coefficient is greater than zero.

Theorem 14.3 (AIC)**.** *The AIC statistic for a fitted model is a sum of the likelihood (as estimate of model fit) and the number of parameters (p and q). The* best *model is the one that minimizes the value of the AIC.*

15. Prediction using ARIMA(p,d,q) models

Definition 15.1 (Best linear predictor)**.** The best linear predictor of Y in terms of $X_i, i=1,2,\ldots,n$ in the least squares sense, is the linear function of $\mathbf{X}$

$$\lambda(\mathbf{X}) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i$$

where $\beta_i, i=0,1,2,\ldots,n$ are chosen to minimise

$$\mathbb{E}\left[\left(Y - \beta_0 - \sum_{i=1}^{n}\beta_i X_i\right)^2\right]$$

Theorem 15.2 (Properties of the best linear predictor)**.** *If $\hat{\lambda}(\mathbf{X}) = \hat{\beta}_0 + \sum_{i=1}^{n}\hat{\beta}_i X_i$ is the best linear predictor for Y in terms of $X_i, i=1,2,\ldots,n$, the the residual*

$$U = Y - \hat{\lambda}(\mathbf{X})$$

has expectation 0 and is uncorrelated with ***every*** *linear function $\lambda(\mathbf{X})$.*

Theorem 15.3 (Uniqueness of the best linear predictor)**.** *Let $\hat{\lambda}(\mathbf{X}) = \hat{\beta}_0 + \sum_{i=1}^{N}\hat{\beta}_i X_i$ be the best linear predictor, and let $U = Y - \hat{\lambda}(\mathbf{X})$.*

Suppose $\lambda^\star(\mathbf{X}) = \beta_0^\star + \sum_{i=1}^{N}\beta_i^\star X_i$ and $U^\star = Y - \lambda^\star(\mathbf{X})$ has the properties:

(1) $\mathbb{E}[U^\star] = 0$

(2) $U^\star$ is uncorrelated with every linear function of $X_1, X_2, \ldots X_n$. Then

$$\lambda^\star(\mathbf{X}) = \hat{\lambda}(\mathbf{X})$$

Example 15.4 (Best linear predictor for AR(p))**.** Let $\{X_t\}$ be an AR(p) process. Then we have

$$X_{t-1}(1) = \mu + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p}$$

is the **best linear predictor** for X_t in terms of $X_{t-1}, \ldots, X_{t-p}$

Example 15.5 (Best linear predictor for ARMA(p,q))**.** Let $\{X_t\}$ be an ARIMA(p,q) process. Then we have

$$X_{t-1}(1) = \mu + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}$$

As we have

$$X_{N+1} - X_N(1) = Z_t \sim \mathcal{N}(0, \sigma_Z^2),$$

we then have a $(1-\alpha)100\%$ confidence interval for X_{N+1} is given by

$$X_N(1) \pm \Phi\left(1 - \frac{\alpha}{2}\right)\sigma_Z$$

Example 15.6 (Two step predictors for an AR(p) process)**.** We have the best linear predictor for $X_{t-2}(2)$ is given as

$$X_{t-2}(2) = \mu + \alpha_1 X_{t-2}(1) + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p}$$

where $X_{t-2}(1)$ is the one step predicted value for the AR(p) process above.

We can calculate that the $(1-\alpha)100\%$ confidence intervals for X_t are given by

$$X_{t-2}(2) \pm \Phi\left(1 - \frac{\alpha}{2}\right)\sqrt{(1+\alpha_1^2)\sigma_Z^2}$$