INTRODUCTION

Zero-Mean Time Series Models

① i.i.d. noise

$\{X_i,\ i=1,\dots,n\}$, $E(X_i)=0$, each X_i is indep.

② Random walk

$\{S_T,\ t=0,1,\dots\}$, $S_t=\sum_{i=1}^{t} X_i$, X_i is i.i.d. noise.

NOT finite variance

③ White noise

$\{X_i,\ i=1,\dots,n\}$, where:

1. Each X_i not corr. (can still be dep.)
2. $E(X_i)=0$
2. $Var(X_i)=\sigma^2<\infty$

Relationships:

1. i.i.d. variables has finite variance $\Rightarrow$ white noise
2. White noise variables are indep. $\Rightarrow$ i.i.d. noise.

LOSS FUNCTIONS

Definitions

Loss function: $L(Y, f(x))$

Properties:

① $L>0$

② $L=0 \Leftrightarrow Y=f(x)$

Risk: $E_{X,Y}[L(Y,f(X))]$

Quadratic loss function

Defn: $L(Y, f(x)) = (Y - f(x))^2$

Conditional expectation theorem: f that minimizes risk is $f(x) = E[Y|X]$

Absolute loss function

Defn: $L(Y, f(x)) = |Y - f(x)|$

Median theorem: f that minimizes $L(Y, f(x))$ is Median $(Y|X)$

REGRESSION REVIEW

Parameter estimation

Let model be $f(X; \beta) = X\beta + \varepsilon$, $\varepsilon \sim MVN(0, \sigma^2 I)$

β that minimizes sum of squares loss: $\hat{\beta} = (X'X)^{-1} X' Y$

Hat matrix

Defn: $X'(X'X)^{-1}X'$ $\Rightarrow$ transforms $Y \rightarrow \hat{Y}$

Properties:

1. Symmetric
2. Idempotent
3. $\text{trace}(H) = p + 1$ # of variables

Inference

1. Distn. of $\hat{\beta}$:

$$\hat{\beta} \sim MVN(\beta, \sigma^2 (X'X)^{-1})$$

2. Variance estimator:

$$\hat{\sigma}^2 = \frac{1}{n-p-1} SSE$$

3. Test of param significance:

$H_0: \beta_j = 0$, $H_a: \beta_j \neq 0$

Discrepancy measure:

$$\frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{v_j}} \sim t_{n-p-1}$$

$v_j \rightarrow (j+1)^{th}$ diagonal elem. of $(X'X)^{-1}$

④ Test of of smaller model

H_0: model w/ p_0 variables is the same as model w/ p_1 variable

$$\frac{RSS_0 - RSS_1}{p_1 - p_0} \div \frac{RSS_1}{n-p_1-1} \sim F_{p_1-p_0,\, n-p_1-1}$$

Residual Diagnostics

Normality

① Graphical: Q-Q plot

② Test: Shapiro-Wilk test (H_0: $Y_1, \ldots, Y_n \sim N$)

Constant mean & variance

① Graphical:

1. Residuals vs. fitted values
2. " index / time

$\Rightarrow$ no pattern ideally

② Test: Fligner-Kileen test ($H_0: \sigma_1^2 = \ldots = \sigma_k^2$)

Uncorrelatedness

ACF $\Rightarrow$ 95% of bands should be in interval for no corr.

Randomness

① Diff. signs test

1. S = # of values s.t. $y_i - y_{i-1} > 0$

2. $S \sim N\left(\frac{n-1}{2}, \frac{n+1}{12}\right)$

② Runs signs test

1. $n_1 :=$ # of obs > median. $n_2 :=$ num. of obs < median.

2. $R :=$ # of consec. seq. of data where all < median

4. $$R \sim N\left(\underbrace{1+\frac{2n_1 n_2}{n_1+n_2}}_{\mu_R}, \frac{(\mu_R-1)(\mu_R-2)}{n_1+n_2-1}\right)$$

PREDICTION W/ LINEAR REGRESSION

Prediction interval

$$x_0'\hat{\beta} \pm c_\alpha \hat{\sigma}\sqrt{1+x_0'(X'X)^{-1}x_0}$$

Variance inflation factor

$$VIF(\hat{\beta}_j) = \frac{1}{1-R^2_{x_j|x_{-j}}}$$

Interpolation & Extrapolation

$$x'(X'X)^{-1}x \leq \max(H_{ii}) \Rightarrow \text{interpolation}$$

BIAS VARIANCE TRADEOFF

Decomposition

$$MSE(x_0) = Bias(\hat{y}_0)^2 + Var(y_0)$$

Prediction decomposition

$$Err(x_0) = \sigma_\varepsilon^2 + bias^2 + variance$$

VARIABLE SELECTION

R^2

$$R^2 = 1 - \frac{SSE}{SST}$$

$$\text{Adjusted } R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

Aikake's Information Criteria

$$AIC = -2\ell(\hat{\beta}) + 2N_p$$

$$AIC_c = AIC + \frac{2N_p(N_p+1)}{n - N_p + 1}$$

Bayesian Information Criteria

$$BIC = -2\ell(\hat{\beta}) + \log(n)N_p$$

Average prediction squared error

$$APSE = \frac{1}{|\mathcal{N}|}\sum_{\mathcal{N}}(y-\hat{y})^2, \quad \mathcal{N} \text{ is test set}$$

REGULARIZATION METHODS

Ridge Regression

$$(\hat{\beta}_0, \hat{\beta}_{ridge}) = \underset{\beta_0, \vec{\beta}}{\operatorname{argmin}} \left\{ (y_i - \beta_0 - X\vec{\beta})'(y_i - \beta_0 - X\vec{\beta}) + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$

$$= \underset{\beta_0, \vec{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 \quad \text{s.t.} \sum_{j=1}^{p} \beta_j^2 \le t$$

$$\hat{\beta}_0 = \bar{y} - \frac{1}{n}\sum_{i,j}\hat{\beta}_j x_{ij} \qquad \hat{\beta}_{ridge} = (X'X + \lambda I)^{-1}X'Y$$

Lasso Regression

$$(\hat{\beta}_0, \hat{\beta}_{ridge}) = \underset{\beta_0, \vec{\beta}}{\operatorname{argmin}} \left\{ (y_i - \beta_0 - X\vec{\beta})'(y_i - \beta_0 - X\vec{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

Elastic Net Regression

$$\hat{\beta}_{en} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{p}\left((1-\alpha)\beta_j^2 + \alpha|\beta_j|\right) \right\}$$

STATIONARY PROCESS

Strong stationarcy

$\forall$ finite subsets $\{t_1, \ldots, t_n\} \subseteq T$ & $\forall h$, $F(X_{t_1}, \ldots, X_{t_n}) = F(X_{t_1+h}, \ldots, X_{t_n+h})$

Weak stationary

① $E[X_t^2] < \infty$ $\forall t \in T$ & indep. of t OR $Var[X_t] < \infty$

② $E[X_t] = \mu$ $\forall t \in T$ & indep. of t

③ $\gamma_{X_t}(h) = Cov(X_t, X_{t+h})$ is indep. of t $\forall h$

ACF

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{\gamma_X(h)}{Var(X)}$$

Estimation of AVCF & ACF

① $$\hat{\gamma}(h) = \widehat{Cov}(X_{t+|h|}, X_t)$$
$$= \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x}_{t+h})(x_t - \bar{x})$$

② $$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\widehat{Var}(X_t)}$$

③ If $\{X_t\}$ is iid noise & finite variance

$$\tilde{\rho}(h) \sim N(0, 1/n)$$

SMOOTHING METHODS

Finite moving avg. filter

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t-j}$$

Exponential smoothing

$$\hat{m}_t = \begin{cases} m_t = \alpha X_t + (1-\alpha)\hat{m}_{t-1} \\ \hat{m}_1 = X_1 \end{cases}$$

$$\hat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j X_{t-j} + (1-\alpha)^{t-1} X_1$$

Differencing

Eliminates trends of poly. k $\Rightarrow$ $\nabla^k X_t = (1-B)^k X_t \Rightarrow$ diff. k times w/ lag 1

$\nabla_k X_t = (1-B^k) X_t \Rightarrow$ diff. 1 time w/ lag k

Decompose

① Trend estimation:

$$\hat{m}_t = \begin{cases} \dfrac{0.5x_{t-q} + x_{t-q+1} + \cdots + x_{t+q-1} + 0.5x_{t+q}}{2q} & \Rightarrow \text{period is even} \\ \dfrac{1}{2q+1}\sum_{i=-q}^{q} x_{t-i}, & \text{period is odd} \end{cases}$$

② Detrendize: $x_t - \hat{m}_t$

③ Seasonal estimation:

a) Take avg. of each season $\Rightarrow \{W_k\}$

b) Centralize each to overall avg.

$$\hat{S}_k = W_k - \frac{\sum W_j}{d}$$

Holt-Winters

$$\begin{cases} L_T \\ T_T \\ I_t \end{cases} \longrightarrow$$

Forecast

$$\hat{X}_{t+h} = L_t + hT_t + I_{\text{period of } t+h}$$

If $\gamma = \beta = 0 \rightarrow$ regular exp. smoothing, $\hat{X}_{t+h} = L_t$

If $\gamma = 0 \rightarrow$ double exp smoothing (no seasonality): $\hat{X}_{t+h} = L_t + hT_t$

STATIONARITY & LINEAR PROCESSES

MA(q)

① Defn:

$Z_t \sim WN(0, \sigma^2)$ & θ_i are non-zero constants. MA(q) is

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

② Theorem: MA(q) is stationary

③ AVCF

$$\gamma_{MA(q)}(h) = \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+h}$$

④ ACF

$$\rho_{X_t}(h) = \frac{\sum_{j=0}^{q-|h|} \theta_j \theta_{j+h}}{\sum_{j=0}^{q} \theta_j^2}$$

④ Signature:

$$\gamma(h) = \rho(h) = 0 \quad \forall h > q$$

⑤ q-dep

$\{X_t\}$ is q-dep. if X_t & X_s are dep. if $|t-s| = q$ & indep. $|t-s| > q$

⑥ q-corr.

" q-cor. " corr " uncorr. "