Stat 551 Exam 1 Formulas

Introduction

Covariances/Variances of Linear Combinations

$$\text{Var}\left(a_0 + \sum_{j=1}^{p} a_j X_j\right) = \sum_{j=1}^{p}\sum_{k=1}^{p} a_j a_k \text{Cov}(X_j, X_k)$$

$$\text{Cov}\left(a_0 + \sum_{j=1}^{p} a_j X_j, b_0 + \sum_{k=1}^{q} b_j Y_j\right) = \sum_{j=1}^{p}\sum_{k=1}^{q} a_j b_k \text{Cov}(X_j, Y_k)$$

Weakly, Strong Stationarity and White Noise

Definition. A time series $\{X_t\}$ is *weakly stationary* if for all integers t and h,

a) $\text{Var}(X_t) < \infty$

b) $\text{E}[X_t]$ does not depend on t

c) $\text{Cov}(X_t, X_{t+h})$ does not depend on t

In other words, first and second order moments do not change with time.

Definition. A time series $\{X_t\}$ is *strictly stationary* if for any positive integer k and integers $t_i, \dots, t_k$ and h, $(X_{t_1}, \dots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_k+h})$ where $\stackrel{d}{=}$ denotes equality in probability distribution. In other words, joint probability distributions do not change with time.

Definition. A time series is *white noise* if it is uncorrelated, zero mean, with finite variance.

row $\Rightarrow$ column

	WS	WN	SS	IID
WS	$\Rightarrow$	$\not\Rightarrow$	$\not\Rightarrow$ (4)	$\not\Rightarrow$
WN	$\Rightarrow$	$\Rightarrow$	$\not\Rightarrow$	$\not\Rightarrow$
SS	$\not\Rightarrow$ (1)	$\not\Rightarrow$	$\Rightarrow$	$\not\Rightarrow$
IID	$\not\Rightarrow$ (2)	$\not\Rightarrow$ (3)	$\Rightarrow$	$\Rightarrow$

(1) SS + finite variances $\Rightarrow$ WS (2) IID + finite variances $\Rightarrow$ WS (3) IID + finite variances + mean zero $\Rightarrow$ WN (4) WS implies SS if $\{X_t\}$ is Gaussian.

ACVF/ACF function properties

Definition. The Autocovariance function (ACVF) for a WS $\{X_t\}$ is defined as $\text{Cov}(X_t, x_{t+h}) = \gamma(h)$. This is a function of h only.

Properties of ACVF:(i)$\gamma(0) = \text{Var}(X_t)$
(ii)$|\gamma(h)| = |\text{Cov}(X_t, x_{t+h}| \leq \sqrt{\text{Var}(X_t)\text{Var}(X_t+h)} = \gamma(0)$
(iii)$\gamma(-h) = \text{Cov}(X_t, x_{t-h}) = \text{Cov}(X_{t+h}, x_t) = \gamma(h)$ Thus $\gamma(h)$ is an even function.

Definition. A function $\gamma : \mathbb{Z} \to \mathbb{R}$ is *nonnegative definite* iff for any n, all $(a_1, \dots, a_n) \in \mathbb{R}^n$, all $t_1, \dots, t_n \in \mathbb{Z}$ $0 \leq \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(t_i - t_j)$.

Result. $\gamma : \mathbb{Z} \to \mathbb{R}$ is NND iff γ is the ACVF of some WS time series.

Definition. Autocorrelation function (ACF) $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$

Properties of ACF (i) $\rho(0) = 1$ (ii) $|\rho(h)| \leq 1$ (iii) $\rho(\cdot)$ is even: $\rho(h) = \rho(-h) = \rho(|h|)$ (iv) $\rho(\cdot)$ is NND

Sample ACVF/ACF & large sample properties

Definition. Same ACVF based on $X_1, \dots, X_n$ is defined as for $|h| < n, \hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(X_t - \bar{X}_n)(X_{t+|h|} - \bar{X}_n$ where $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$. Estimates $\gamma(h) = \text{E}[(X_t - \mu)(X_{t+|h|} - \mu)]$ where $mu = \text{E}X_t$ for WS process.

Definition. Sample ACF: $\hat{\rho} = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$

Under mild conditions, $(\hat{\rho}(1), \dots, \hat{\rho}(h))' = \underset{\sim}{\hat{\rho}}_h$ is asymptotically normal, $\text{N}(\underset{\sim}{\rho}_h, \Sigma)$ where $\Sigma = \frac{1}{n}W$ and $W = [w_{ij}]_{i,j=1,\dots,h}$ and $w_{ij} = \sum_{k=1}^{\infty}\{\rho(k+i) + \rho(k-i) - 2\rho(k)\rho(i)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(k)\rho(j)\}$.

Definition. Bartlett Bounds: $P\left(\frac{-1.96}{\sqrt{n}} \leq \rho(\hat{k}) \leq \frac{1.96}{\sqrt{n}}\right)$

If $\{X_t\}$ are $\text{IID}(\mu, \sigma^2)$, about 5% of sample autocorrelations should exit the bounds by chance.

Removing Trend/Seasonality

Classical decomposition

$X_t = m_t + s_t + I_t$ where m_t is trend (non-random), s_t is seasonality component, and I_t is random/irregular with mean 0.

Removing Trend

To start, assume $s_t = 0$ and focus on removing m_t in $X_t = m_t + I_t$.
Linear filter (to estimate trend)
integer $q \geq 0$ and filter $\{a_{-q}, \dots, a_{-1}, a_0, a_1, \dots, a_q\}$. Then $\hat{m}_t = \sum_{k=-q}^{q} a_k X_{t-k}$.
Exponential smoothing (trend estimation)
$0 < a < 1, \hat{m}_1 = X_1, \hat{m}_t = aX_t + a(1-a)X_{t-1} + a(1-a)^2 X_{t-2} + \cdots$. In other words, exponentially decreasing weights on previous observations. $a \to 1$ less bias, $a \to 0$ less variance.
Differencing (trend elimination)

- Backshift operator: $B^k f(t) = f(t-k)$
- Difference operator: $1 - B$
- Difference once to kill linear trend, difference twice to kill quadratic trend.
- Differencing also helping remove "stochastic trends"

Removing seasonality

$X_t = m_t + s_t + I_t$, s_t cyclic behavior, known period d.

Definition. Deterministic seasonality: perfect repetition in the seasonal component $s_t = s_{t+d} = s_{t+2d} = \cdots$. The sum of effects over length of period will always be equal, so that WLOG, $\sum_{j=0}^{d-1} s_{t-j} = 0$.

Smoothing:

$$Y_t = \frac{X_t + X_{t-1} + \cdots + X_{t-(d-1)}}{d} = \frac{\sum_{j=1} m_{t-j}}{d} + \frac{\sum_{j=1} I_{t-j}}{d}$$

Differencing:
Use seasonal difference $1 - B^d$ to remove seasonality. In fact $1 - B^d - (1-B)(1 + B + B^2 + \cdots + B^{d-1}$ smooths seasonal summation filter and eliminates trends.
Regression (dummy variable, trigonometric polynomials)

Estimate s_t with $s_t = s_{t+d} = s_{t+2d} \cdots$ and $\sum_{j=0}^{d-1} s_{t-j} = 0$

Dummy variables: Let s_t be assigned dummy variable $\gamma_1, \dots, \gamma_{d-1}$. for $t = 1, 2+d$, etc and then for $t = d, 2d, 3d, \dots$ $s_t = \sum_{j=1}^{d-1}$. Then regression of X_t, dummy variables s_t over t. $\sum_{t=1}^{n}(X_t - s_t)^2$ minimize for $\gamma_1, \dots, \gamma_{d-1}$.

Trigonometric polynomials:
Let $f = \lfloor \frac{d}{2} \rfloor$ and $s_t = \sum_{j=1}^{f}\{a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)\}$ where $\lambda_j = \frac{2\pi j}{2}$ (frequency). Then estimate a_j and b_j by regression methods.
To summarize, could use regression to estimate seasonality or difference to eliminate.

Checks for WN or iid

Bartlett bounds in sample ACF for WN – 95% of $\hat{\rho}(h)$ should fall within the bounds.
Ljung-Box (portmanteau) for WN
$Q_{LB} = n(n+2)\sum_{k=1}^{h}[\hat{\rho}(k)]^2/(n-k) \dot{\sim} \chi^2_h$. Reject H_0 if $Q_{LB} > \chi^2_{h,1-\alpha}$.$\text{H}_0$: $\{Y_t\}$ is WN.
McLeod-Li for iid Gaussian
$Q_{ML} = n(n+2)\sum_{k=1}^{h}[\hat{\rho}_{X^2}(k)]^2/(n-k) \dot{\sim} \chi^2_h$, $\rho_{X^2}(k) = \text{corr}(X_t^2, X_{t+h}^2)$. Reject H_0 if $Q_{ML} > \chi^2_{h,1-\alpha}$.$\text{H}_0$: $\{Y_t\}$ is Gaussian WN.
Rank tests for iid/randomness
Replace data with ranks $\Rightarrow R_1, \dots, R_n$.

- Turning Point: $T \equiv \#$ of $i(2 \leq i \leq n-1)$ where $R_{i-1} < R_i$ and $R_i > R_{i+1}$ or $R_{i-1} > R_i$ and $R_i < R_{i+1}$.
  $\text{E}(T) = 2(n-2)/3$, $\text{Var}(T) = (16n - 29)/90$
- Positive Differences: $S \equiv \#$ of $i(2 \leq i \leq n)$ where $R_{i-1} < R_i$.
  $\text{E}(T) = (n-1)/2$, $\text{Var}(P) = (n+1)/12$
- Positive Pairs: $P \equiv \#$ of $(i,j)(2 \leq i \leq j \leq n)$ where $R_j > R_i$.
  $\text{E}(T) = \binom{n}{2}/2$, $\text{Var}(T) = n(n-1)(2n+5)/72$

$\{Y_t\}$ is IID. Low power on all three tests.

Prediction

Problem: Given $X_1, \dots, X_n$, predict X_{n+h}.

best MS prediction vs. best linear MS prediction

- Using MSE criterion, want to find a predictor $\tilde{X}_{n+h} = h(X_1, \dots, X_n)$ that minimizes $\text{MSE}(\tilde{X}_{n+h}) \equiv \text{E}\left[X_{n+h} - \tilde{X}_{n+h}\right]^2$.
- Well-known *best* MSE predictor given by $\hat{X}_{n+h} = \text{E}[X_{n+h}|X_1, \dots, X_n]$.
- Can compute $\text{E}[X_{n+h}|X_1, \dots, X_n]$ only when $\{X_t\}$ IID with $\text{E}X^2 < \infty$ or when $\{X_t\}$ is Gaussian WS.

Gaussian W.S. X_t

$\hat{X}n + h = \text{E}[X_{n+h}|X_1, \dots, X_n] = \mu + \underset{\sim}{\gamma}(h)'\Gamma_n^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})$ and $\text{MSE}\hat{X}n + h = \gamma(0) - \underset{\sim}{\gamma(h)}'\Gamma_n^{-1}\underset{\sim}{\gamma}(h)$.

best linear MS predictor

- Linear prediction means a predictor for X_{n+h} of the form $a_0 + \sum_{j=1}^{n} a_j X_{n+1-j}$ for real constants $a_0, \dots, a_n$.
- Best Linear prediction means choosing $a_0, \dots, a_n$ to minimize MSE.
- compute best linear predictor by pretending $\{X_t\}$ are WS Gaussian and calculating $E(X_{n+h}|X_1, \dots, X_n)$.

$\hat{X}_{n+1} = \mu + \sum_{j=1}^{n} \phi_{n,j}(X_{n+1-j} - \mu)$ where $\underset{\sim}{\phi}_n = \Gamma_n^{-1} \underset{\sim}{\gamma}(h)$ where $h = 1, \dots, n$.

Projection Theorem

For Time Series,

1. $X_1, \dots, X_n, X_{n+h} \in L_2 \equiv$ space of all r.v.s with finite 2nd moment.
2. Inner Product: $\langle X, Y \rangle = \mathrm{E}XY$ for $X, Y \in L_2$
3. Orthogonal: $\langle X, Y \rangle = 0$ for $X, Y \in L_2$.
4. Distance: $||X - Y|| = [\mathrm{E}(X - Y)^2]^{1/2}$ for $X, Y \in L_2$.
5. The best linear predictor, $P_n X_{n+h}$ is the linear combination of $(1, X_1, \dots, X_n)$ in L_2 which minimizes the distance to X_{n+h} and the residual $X_{n+h} - P_n X_{n+h}$ is orthogonal to 1 and each X_i.

Theorem (Projection Theorem). *If* $\mathcal{M} \equiv span(1, X_1, \dots, X - n) \subset L_2$ *and* $X_{n+h} \in L_2$,

(i) there exists a unique $P_n X_{n+h} \in \mathcal{M}$ *st* $||X_{n+h} - P_n X_{n+h}|| = \inf_{Z \in \mathcal{M}} ||X_{n+h} - Z||$

(ii) $\tilde{X}_{n+h} = P_n X_{n+h}$ *iff* $\tilde{X}_{n+h} \in \mathcal{M}$ *and* $X_{n+h} - \tilde{X}_{n+h} \perp \mathcal{M}$.

Durbin-Levinson algorithm

Let $\{X_t\}$ be WS, mean zero, ACVF $\gamma(\cdot)$ such that $\gamma(0) > 0$ and $\lim_{h\to\infty} \gamma(h) = 0$. Then:

1. Set $P_0 X_1 = EX_1 = 0$ and $v_0 \equiv \mathrm{E}(X_1 - P_0 X_1)^2 = \mathrm{Var}(X_1) = \gamma(0)$.
2. Set $P_1 X_2 = \frac{\gamma(1)}{\gamma(0)} X_1 = \phi_{1,1} X_1$, $\phi_{1,1} = \frac{\gamma(1)}{\gamma(0)}$ and $v_1 \equiv \mathrm{E}(X_2 - P_1 X2)^2 = (1 - \phi_{1,1}^2) v_0$.
3. For $k \geq 2$, set $P_k X_{k+1} = \sum_{j=1}^{k} \phi_{k,j} X_{k+1-j}$ where

$$\phi_{k,k} = \left[\gamma(k) - \sum_{j=1}^{k-1} \phi_{k-1,j} \gamma(k-j)\right] / v_{k-1}$$

$$\begin{bmatrix} \phi_{k,1} \\ \vdots \\ \phi_{k,k-1} \end{bmatrix} = \begin{bmatrix} \phi_{k-1,1} \\ \vdots \\ \phi_{k-1,k-1} \end{bmatrix} - \phi_{k,k} \begin{bmatrix} \phi_{k-1,1-1} \\ \vdots \\ \phi_{k,1} \end{bmatrix}$$

And $v_k \equiv \mathrm{E}(X_{k+1} - P_k X_{k+1})^2 = (1 - \phi_{k,k}^2) v_{k-1}$.

PACF

Definition. If $\{X_t\}$ is WS (as in D-L algorithm), the partial autocorrelation function is given as $\alpha(n) = \mathrm{Corr}(X_n + 1 - P_{\mathcal{K}_1} X_{n+1}, X_1 - P_{\mathcal{K}_1} X_1) = \phi_{n,n}$ for $n \geq 1$ where $\mathcal{K}_1 = \mathrm{span}(X_2, \dots, X_n$ for $n \geq 2$ and $\mathcal{K}_1 = \emptyset$ if $n = 1$.

PACF for AR(p)

Using $\{Z_t\} \sim WN(0, \sigma^2)$ (Z_t uncorrelated with $X_j, j < t$), define $X_t = \phi_1 X_{t-1} + \cdots \phi_p X_{t-p} + Z_t$. If $n = p$,

$$\begin{aligned} P_p X_{p+1} &= \phi_{p,1} X_p + \cdots + \phi_{p,p} X_1 && \text{(D-L)} \\ &= \phi_1 X_p + \cdots + \phi_p X_1 && \text{(projection thm)} \end{aligned}$$

So $\alpha(p) = \phi_p$. And if $n > p$, $\alpha(n) = 0$.

Mean square convergence & Linear Filters

Mean sequare convergence/Cauchy criterion

Definition. Suppose X is a r.v. and $\{X_n\}$ a sequence of r.v.s. We say X_n *converges to* X *in MSE*, $X_n \overset{MSE}{\to} X$ iff $||X_n - X||^2 = \mathrm{E}(X_n - X)^2 \to 0$ as $n \to \infty$ iff *Cauchy criterion* holds: $||X_n - X_m||^2 = \mathrm{E}(X_n - X_m)^2 \to 0$ as $n, m \to \infty$.

Sample mean variance formula & asymptotic normality

X_t WS, $\bar{X}_n \overset{MSE}{\to} \mu$ and $\bar{X}_n \dot\sim N\left(\mu, \frac{1}{n^2} \sum_{h=-n}^{n} (n - |h|)\gamma(h)\right)$ for large n or $\bar{X}_n \dot\sim N\left(\mu, \frac{1}{n} \sum_{h=-\infty}^{\infty} \gamma(h)\right)$.

Linear filter result WS

Result. Consider WS $\{Z_t\}$ (not necessarily WN) with mean 0 and ACVF $\gamma_Z(h) = \mathrm{Cov}(Z_t, Z_{t+h})$. Let $\psi \in \mathbb{R}$, any integer j, with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Then the process $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} = \left(\sum_{j=-\infty}^{\infty} \psi_j B^j\right) Z_t = \psi(B) Z_t$ is WS with mean 0 and ACVF $\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Z(h - k + j)$.

Linear processes

Use $Z_t \sim WN(0, \sigma^2)$, then $\gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2$

Causal linear representations/MA(∞), AR(1)

$\underline{MA(\infty)}$: Linear process where $\psi_j = 0$ for $j < 0$. $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ and $\gamma_X(h) = \gamma_X(|h|) = \sum_{j=1}^{\infty} \psi_j \psi_{j+|h|} \sigma^2$. Note X_t depends only on past Z_t's, this is a causal representation.
$\underline{AR(1)}$: For $Z_t \sim WN(0, \sigma^2)$ and $|\phi| < 1$, define $AR(1)$ process as $X_t = \sum_{j=0}^{\infty} Z_{t-j}$. Since $\sum_{j=0}^{\infty} |\phi|^j < \infty$, $\{X_t\}$ is WS with $\mathrm{E}X_t = 0$ and with $\gamma_X(h) = \frac{\sigma^2 \phi^{|h|}}{1 - \phi^2}$.
What happens if $|\phi| > 1$? 1. Explosive, non-stationary case, 2. Stationary case. $\{X_t\}$ not necessarily stationary when the relationship with Z_t doesn't hold for all t.

ARMA(p,q) models

Definition. A process $\{X_t\}$ is said to be ARMA(p, q) for integers p, q ≥ 0 with

AR polynomial $\phi(z) = 1 - \phi_1 z - \phi 2z^2 - \cdots - \phi_p z^p$

MA polynomial $\theta(z) = 1 + \theta_1 z + \theta 2z^2 + \cdots + \theta_p z^p$

if $\{X_t\}$ satisfies $\phi(B) X_t = \theta(B) Z_t$ for all integers t, wrt some $\{Z_t\} \sim WN(0, \sigma^2)$.

In the ARMA definition, assume that the polynomials $\phi(z)$ and $\theta(z)$ have no common factors.

WS criteria

There is a unique solution $\{X_t\}$ to $\phi(B) X_t = \theta(B) Z_t$ where $\{X_t\}$ is WS iff the AR polynomial $\phi(z)$ has no roots $z \in \mathbb{C}$ where $|z| = 1$.

Causal/invertible

Definition. For $\{X_t\} \sim$ ARMA(p, q), $\{X_t\}$ is said to be *causal*

- iff $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ holds for all integer t wrt some coefficients $\{\psi_j\}$ where $\sum_{j=0}^{\infty} |\psi_j| < \infty$.
- iff AR polynomial satisfies $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for any $z \in \mathbb{C}$ with $|z| \leq 1$.

Note, the coefficients $\{\psi_j\}$ are given by $\phi(z)\psi(z) = \theta(z)$.

Definition. For $\{X_t\} \sim$ ARMA(p, q), $\{X_t\}$ is said to be *invertible*

- iff $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ holds for all integer t wrt some coefficients $\{\pi_j\}$ where $\sum_{j=0}^{\infty} |\pi_j| < \infty$.
- iff MA polynomial satisfies $\theta(z) = 1 + \theta_1 z + \cdots + \theta_p z^p \neq 0$ for any $z \in \mathbb{C}$ with $|z| \leq 1$.

Note, the coefficients $\{\pi_j\}$ are given by $\theta(z)\pi(z) = \phi(z)$.

Yule-Walker equations for causal ARMA(p,q)

Let $m = \max\{p, q + 1\}$. Then the Yule-Walker equations are

$$\gamma(h) - \phi_1 \gamma(h-1) - \cdots \phi_p \gamma(h-p) = \begin{cases} \sigma^2 \sum_{j=1}^{m} \theta_{j+h} \psi_j & 0 \leq h < m \\ 0 & h \geq m \end{cases}$$

Where $\theta_0 = 1$ and $\theta_j = 0$ for $j > q$.

Methods for finding ACVF of causal ARMA

- Method 1: Direct Approach (must have all $\psi_0, \psi_1, \psi_2, \dots$ to be determined.

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \Rightarrow \gamma_X(h) = \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \sigma^2$$

- Method 2: explicit, non-recursive
  1. Find $\psi_0, \dots, \psi_q$
  2. Compute the k distinct roots $\xi_1, \dots, \xi_k$ of $\phi(z) = 0$ where root ξ_i has multiplicity $r_i \geq 1, i = 1, \dots, k$ and $r_1 + \cdots r_k = p$. Causal implies each root $|\xi_i| > 1$.
  3. $\gamma(h) = \sum_{i=1}^{k} \sum_{j=1}^{r_i - 1} \beta_{ij} \xi_i^{-h}$, $h \geq m - p$.
  4. Need to find $\{\beta_{ij}\}$ along with the first $m - p$ ACVF values. There are m unknowns, so use first m Y-W equations.
- Method 3: explicit, recursive
  1. Find $\psi_0, \dots, \psi_q$
  2. write down the first $p + 1$ Y-W equations $(h = 0, \dots, h = p)$.
  3. Solve the first $p + 1$ covariances
  4. $\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p)$ for all $h \geq m$ so we can solve for all other covariances recursively.