STAT 26100 - Review

Solutions

Exercise 1: Define autocovariance, autocorrelation, cross-covariance, and cross-correlation.

Solution 1:

Autocovariance: $\gamma(i,j) = \mathrm{Cov}(X_i, X_j)$.

Autocorrelation: $\rho(i,j) = \frac{\gamma(i,j)}{\sqrt{\gamma(i,i)\gamma(j,j)}}$.

Cross-covariance: $\gamma_{XY}(i,j) = \mathrm{Cov}(X_i, Y_j)$.

Cross-correlation: $\rho_{XY}(i,j) = \frac{\gamma_{XY}(i,j)}{\sqrt{\gamma_X(i,i)\gamma_Y(j,j)}}$.

Exercise 2: Define strong and weak stationarity. Define joint stationarity.

Solution 2:

Strong stationarity: $\mathbb{P}\{X_{t_1}, \cdots, X_{t_k}\} = \mathbb{P}\{X_{t_1+h}, \cdots, X_{t_k+h}\}$.

Weak stationarity: $\mathbb{E}(X_t) = \mathbb{E}(X_{t+h})$ is constant and $\mathrm{Cov}(X_t, X_{t'}) = \mathrm{Cov}(X_{t+h}, X_{t'+h})$ depends only on the time difference h.

Joint stationarity: X_t and Y_t are jointly stationary if

1. X_t and Y_t are both stationary, and
2. the cross-covariance $\gamma_{XY}(t+h,t) = \mathrm{Cov}(x_{t+h}, y_t)$ depends only on the time difference h.

Exercise 3: Define short and long memory.

Solution 3:
A stationary series has short memory if $\sum_{j=-\infty}^{\infty} |\gamma(h)| < \infty$, and has long memory if $\sum_{j=-\infty}^{\infty} |\gamma(h)| = \infty$.

Exercise 4: Show that if $\lim_{k\to\infty} \gamma(k) = 0$, then $\bar{X}_n \to \mu$ in probability.

Solution 4:
Use Gauss + triangle inequality.

Exercise 5: What is Chebyshev's inequality?

Solution 5:
For any $\epsilon > 0$, $\mathbb{P}\{|X| \geq \epsilon\} \leq \frac{\mathbb{E}|X|^2}{\epsilon^2}$ (if the second moment exists).

Exercise 6: Define mean square convergence, convergence in probability, and convergence in distribution. Show that mean square convergence implies convergence in probability.

Solution 6:
Mean square convergence is when a sequence of random variables X_n is such that $\mathbb{E}|X_n - X|^2 \to 0$ as $n \to \infty$. Convergence is when a sequence of random variables is such that for any $\epsilon > 0$, $\mathbb{P}\{|X_n - X| \geq \epsilon\} \to 0$ as $n \to \infty$. Convergence in distribution is when a sequence of random variables with distribution functions f_n are such that $f_n \to f_X$ as $n \to \infty$.

By Chebyshev,

$$\mathbb{P}\{|X_n - X| \geq \epsilon\} \leq \frac{\mathbb{E}|X_n - X|^2}{\epsilon^2} \to 0.$$

Exercise 7: State the normal CLT, the M-dependent CLT, and the linear process CLT.

Solution 7:
The normal CLT. If $x_1, \cdots, x_n$ are iid with mean μ and variance σ^2, let $\bar{x}_n$ be the sample mean. Then

$$\bar{x}_n \sim AN(\mu, \sigma^2/n).$$

The M-dependent CLT. If $\{x_t\}$ is a strictly stationary M-dependent series with mean μ, then if $V_M = \sum_{j=-M}^{M} \gamma_x(j)$, then

$$\bar{x}_n \sim AN(\mu, V_m/n).$$

The linear process CLT. Let $x_t = \mu_x + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}$, where $w_t \sim$ i.i.d.$(0, \sigma_w^2)$, and $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Then

$$\begin{aligned} \bar{x}_n &\sim AN(\mu_x, n^{-1} \sum_{j=-\infty}^{\infty} \gamma_x(j)) \\ &\sim AN(\mu_x, n^{-1}\sigma_w^2 (\sum_{j=-\infty}^{\infty} \psi_j)^2) \end{aligned}$$

Exercise 8: State the weak and strong law of large numbers.

Solution 8:
Weak. $\bar{X} \to \mu$ in probability.
Strong. If $\{X_t\}$ are iid, then $\bar{X} \to \mu$ almost surely.

Exercise 9: Let $X_t = \mu + W_t + \theta W_{t-1}$, where $W_t \sim_{\text{i.i.d.}} N(0, \sigma^2)$ and $|\theta| < 1$. Let $\tilde{\mu}$ be the BLUE of μ. Show that $\tilde{\mu} \to \mu$ in probability.

Solution 9:
Let $\epsilon > 0$. By Chebyshev's inequality, we have

$$\mathbb{P}\{|\tilde{\mu} - \mu| \geq \epsilon\} \leq \frac{\mathbb{E}|\tilde{\mu} - \mu|^2}{\epsilon^2}.$$

Since $\tilde{\mu}$ is the BLUE of μ, $\text{Var}(\tilde{\mu}) = \mathbb{E}|\tilde{\mu} - \mu|^2$. Consider the sample mean $\bar{\mu}$, which has variance $\frac{\sigma^2}{n}$ and is also unbiased. Since $\text{Var}(\bar{\mu}) = \mathbb{E}|\bar{\mu} - \mu|^2 \to 0$ as $n \to \infty$ and $\tilde{\mu}$ is the BLUE, it must be that

$$\mathbb{E}|\tilde{\mu} - \mu|^2 \leq \mathbb{E}|\bar{\mu} - \mu|^2 \to 0.$$

Exercise 10: Define $AR(p)$. What is the autoregressive polynomial / operator?

Solution 10:
The AR(p) model is $\phi(B)X_t = w_t$, where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and B is the backshift operator.

Exercise 11: Define $MA(q)$. What is the moving average polynomial / operator?

Solution 11:
The MA (q) model is $X_t = \theta(B)w_t$, where $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ and B is the backshift operator.

Exercise 12: Define a causal $ARMA(p, q)$ model. Define an invertible $ARMA(p, q)$ model. What are the conditions for both? How would you find the causal / invertible form?

Solution 12:
An $ARMA(p, q)$ model is causal if there exists a representation $x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$ where $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $\psi_0 = 1$. A model is causal if and only if all the roots of the autoregressive polynomial lie outside of the unit circle. The causal parameters can be found through the equation

$$\psi(z) = \frac{\theta(z)}{\phi(z)}.$$

An $ARMA(p, q)$ model is invertible if there exists a representation $w_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}$, where $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and $\pi_0 = 1$. A model is invertible if and only if all the roots of the moving average polynomial lie outside of the unit circle. The inverted parameters can be found through the equation

$$\pi(z) = \frac{\phi(z)}{\theta(z)}.$$

Exercise 13: Define the PACF, and describe in intuitive terms. What can you say about the best linear predictors $\hat{x}_{t+h}$ and $\hat{x}_t$ (based on values $x_{t+1}, \cdots, x_{t+h-1}$)?

Solution 13:
The PACF with lag h is the correlation between x_{t+h} and x_t with the linear dependence removed. In particular, let

$$\hat{x}_{t+h} = \alpha_1 x_{t+h-1} + \cdots + \alpha_{h-1} x_{t+1}$$
$$\hat{x}_t = \beta_1 x_{t+h-1} + \cdots + \beta_{h-1} x_{t+1}.$$

Then

$$\phi_{11} = \rho(1)$$
$$\phi_{hh} = \text{Corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t).$$

Due to stationarity, the best coefficients are equivalent i.e.

$$\forall i, \alpha_i = \beta_i.$$

Exercise 14: Find the PACF for an AR(1) process. Find the PACF of an MA(1) process.

Solution 14:
For AR(1), $\phi_{11} = \phi$ and $\phi_{kk} = 0$ for $k \geq 2$. For MA(1), $\phi_{hh} = -\frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2(h+1)}}$ for $h \geq 1$.

Exercise 15: What is the minimum mean square error predictor of x_{n+m}, where you have access to $x_n, \cdots$?

Solution 15:
The best predictor is the conditional expectation $\mathbb{E}(x_{n+m}|x_n, \cdots)$.

Exercise 16: How do you find the best m-step ahead linear predictor? What is the prediction error for the best one-step ahead linear predictor?

Solution 16:
Denote $x^n_{n+m} = \beta_0 + \sum_{i=1}^n \beta_i x_i$ as the best m-step ahead linear predictor. We have the following equation

$$x^n_{n+m} = \arg \min_{\beta_0, \cdots, \beta_{n-1}} \mathbb{E}[x_{n+m} - \beta_0 + \sum_{i=1}^n \beta_i x_i]^2.$$

The best parameters can be found with the following: for $k = 0, \cdots, n$ (where $x_0 = 1$)

$$\mathbb{E}[(x_{n+m} - x_{n+m}^n)x_k] = 0.$$

The $k = 0$ equation gives that $\beta_0 = \mu$. Minusing out the mean, expanded into matrix form, and applying the expectation, we get

$$\underbrace{\begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix}}_{\Gamma_n} \underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}}_{\mathbf{B}_n} = \underbrace{\begin{bmatrix} \gamma(m + (n-1)) \\ \vdots \\ \gamma(m) \end{bmatrix}}_{\gamma_m^{m+n-1}}.$$

The parameters for the best m-step ahead linear predictor can then be found with $\mathbf{B}_n = \Gamma_n^{-1}\gamma_m^{m+n-1}$. If $x = [x_1, ...x_n]^T$, the prediction error is given by

$$\begin{aligned} \mathbb{E}(x_{n+1} - \mathbf{B}_n^T x)^2 &= \mathbb{E}(x_{n+1}^2 - 2\mathbf{B}_n x_{n+1} x + \mathbf{B}_n^T x x^T \mathbf{B}_n) \\ &= \gamma(0) - 2\mathbf{B}_n \gamma_1^n + \mathbf{B}_n^T \Gamma_n \mathbf{B}_n \\ &= \gamma(0) - (\gamma_1^n)^T \Gamma_n^{-1} \gamma_1^n. \end{aligned}$$

Exercise 17: What is the best one-step ahead linear predictor of $AR(p)$ with autoregressive polynomial $\phi(\cdot)$, assuming that $n \geq p$? What is the prediction error based on MSE?

Solution 17:
The best one-step ahead linear predictor of $AR(p)$ is $\phi(B)x_n$. Assuming the white noise component has variance σ_w^2, the prediction error is

$$\begin{aligned} \mathbb{E}(x_{n+1} - \phi_1 x_n - \cdots - \phi_p x_1)^2 &= \mathbb{E}(w_{n+1})^2 \\ &= \sigma_w^2. \end{aligned}$$

Exercise 18: Explain why using the prediction equations for the best linear predictor may not be optimal when n is large.

Solution 18:
Inverting a matrix is computationally expensive (although since Γ_n is symmetric positive semidefinite, I don't know why you wouldn't use Cholesky if its definite, LU, or some sort of block matrix thing?)

Exercise 19: Derive the truncated m-step ahead predictor for ARMA(p, q) that is both causal and invertible. Explain the relationship between the following:

1. $x_{n+m}^n = \mathbb{E}(x_{n+m} | x_n, \cdots, x_1)$
2. $\tilde{x}_{n+m} = \mathbb{E}(x_{n+m} | x_n, \cdots)$
3. $\tilde{x}_{n+m}^n =$ truncated predictor of $\tilde{x}_{n+m}$.

Solution 19:
We first explain the relationship between the three predictors. In the optimal case, we want x_{n+m}^n. The idea is that $\tilde{x}_{n+m}$ is a good estimate of x_{n+m}^n if n is large. But we usually don't have access to the infinite past, so the idea is that $\tilde{x}_{n+m}^n$ is a good estimate of $\tilde{x}_{n+m}$. Hence, denoting $\to$ as "is an estimator of", we have

$$\tilde{x}_{n+m}^n \to \tilde{x}_{n+m} \to x_{n+m}^n.$$

Given these relationships, we first derive $\tilde{x}_{n+m}$. Write x_t in its causal and invertible forms:

$$\begin{aligned} x_t &= \sum_{j=0}^{\infty} \psi_j w_{t-j}, \psi_0 = 1 \\ w_t &= \sum_{j=0}^{\infty} \pi_j x_{t-j}, \pi_0 = 1. \end{aligned}$$

Taking the conditional mean, we have from causality and invertibility that

$$\tilde{x}_{n+m} = \sum_{j=0}^{\infty} \psi_j \tilde{w}_{n+m-j} = \boxed{\sum_{j=m}^{\infty} \psi_j w_{n+m-j}}.$$

The mean square prediction error is then

$$\begin{aligned} \mathbb{E}(x_{n+m} - \tilde{x}_{n+m})^2 &= \mathbb{E}\left[\sum_{j=0}^{m-1} \psi_j w_{n+m-j}\right]^2 \\ &= \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2. \end{aligned}$$

Notice that as $m \to \infty$, $\tilde{x}_{n+m}$ quickly converges to the mean, and the mean square prediction error converges to $\sigma_w^2 \sum_{j=0}^{\infty} \psi_j^2 = \gamma_x(0)$.

A recursive algorithm for $\tilde{x}_{n+m}$ can be found from the invertible form:

$$0 = \tilde{w}_{n+m} = \tilde{x}_{n+m} + \sum_{j=1}^{\infty} \pi_j \tilde{x}_{t-j}.$$

This means

$$\begin{aligned} \tilde{x}_{n+m} &= -\sum_{j=1}^{\infty} \pi_j \tilde{x}_{n+m-j} \\ &= -\sum_{j=1}^{m-1} \pi_j \tilde{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j}. \end{aligned}$$

To find $\tilde{x}_{n+m}^n$, we just replace the values that we don't have access to with 0 i.e. all nonpositive indexed values are set to zero. In particular, this means

$$\tilde{x}^n_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{x}^n_{n+m-j} - \sum_{j=m}^{n+m-1} \pi_j x_{n+m-j}.$$

Written explicitly, the truncated predictor is

$$\begin{aligned}\tilde{x}^n_{n+m} =&\phi_1 \tilde{x}^n_{n+m-1} + \cdots + \phi_p \tilde{x}^n_{n+m-p} + \\ &\theta_1 \tilde{w}^n_{n+m-1} + \cdots + \theta_q \tilde{w}^n_{n+m-q}, \\ &\text{where } \tilde{x}^n_t = \begin{cases} 0, & t \leq 0 \\ x_t, & 1 \leq t \leq n \end{cases}, \\ &\tilde{w}^n_t = \begin{cases} 0, & t \leq 0, t > n \\ \phi(B)\tilde{x}^n_t - \theta_1 \tilde{w}^n_{t-1} - \cdots - \theta_q \tilde{w}^n_{t-q} & 1 \leq t \leq n. \end{cases}\end{aligned}$$

Exercise 20: Find the m-step ahead truncated predictor for MA(q). What is the prediction error? Use to find 95% confidence interval.

Solution 20:
For an invertible MA(q) where $q > m$, it is clear that

$$\tilde{x}^n_{n+m} = \theta_m w_n + \cdots \theta_q w_{n+m-q}.$$

The prediction error is

$$\begin{aligned}\mathbb{E}(x_{n+m} - \tilde{x}^n_{n+m})^2 &= \mathbb{E}(w_{n+m} + \theta_1 w_{n+m-1} + \cdots + \theta_{m-1} w_{n+1})^2 \\ &= \sigma^2_w \sum_{j=0}^{m-1} \theta^2_j, \ \theta_0 = 1.\end{aligned}$$

If the white noise is Gaussian, then an approximate 95% confidence interval would be $\tilde{x}^n_{n+m} \pm 1.96 \cdot \sqrt{\sigma^2_w \sum_{j=0}^{m-1} \theta^2_j}$.

Exercise 21: Find the one-step ahead predictions under the following models: AR(2), MA(1), ARMA(1,1), based on the infinite past. What is the mean-square prediction error of the conditional estimator (conditional on the infinite past)?

Solution 21:
For AR(2), assume the model $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$. Assuming we have $x_t, \cdots$, we want to find

$$\hat{\beta}_1, \hat{\beta}_2 = \arg\min_{\beta_1, \beta_2} \mathbb{E}(x_{t+1} - \beta_1 x_t - \beta_2 x_{t-1})^2.$$

Differentiating and equalling to zero, this is equivalent to solving

$$\begin{aligned}\gamma(0)\hat{\beta}_1 + \gamma(1)\hat{\beta}_2 &= \gamma(1) \\ \gamma(1)\hat{\beta}_1 + \gamma(0)\hat{\beta}_2 &= \gamma(2).\end{aligned}$$

Given that $\rho(1) = \frac{\phi_1}{1-\phi_2}$, we find that $\hat{\beta}_1 = \phi_1$ and $\hat{\beta}_2 = \phi_2$. Hence the best linear predictor is $x_{t+1} = \phi_1 x_t + \phi_2 x_{t-1}$. We could have also found this by taking the conditional expectation:

$$\begin{aligned}\mathbb{E}(x_{t+1}|x_t,\cdots) &= \mathbb{E}(\phi_1 x_t + \phi_2 x_{t-1} + w_{t+1}|x_t,\cdots) \\ &= \phi_1 x_t + \phi_2 x_{t-1}.\end{aligned}$$

For MA(1), assume the model $x_t = w_t + \theta w_{t-1}$ and further assume that it is invertible. Then assuming we have $x_t, \cdots$, we have

$$\begin{aligned}\mathbb{E}(x_{t+1}|x_t,\cdots) &= \mathbb{E}(w_{t+1} + \theta \sum_{j=0}^{\infty} \pi_j x_{t-j}|x_t,\cdots) \\ &= \theta \sum_{j=0}^{\infty} \pi_j x_{t-j} = \theta w_t.\end{aligned}$$

For ARMA(1,1), assume the model $x_t + \phi x_{t-1} = w_t + \theta w_{t-1}$, and further assume causality and invertibility.

Exercise 22: Let $X_t = \mu + W_t + \theta W_{t-1}$, where $W_t \sim_{\text{i.i.d.}} N(0, \sigma^2)$ and $|\theta| < 1$. Let $\tilde{\mu}$ be the BLUE of μ. Based on $X_1, \cdots, X_T$, construct an approximate 0.95 confidence interval for X_{T+1}. Do the same for X_{T+2}.

Solution 22:

Exercise 23: Describe three ways of finding parameter estimates for ARMA(p, q). What are the asymptotics? Use the three methods to find estimators for AR(1).

Solution 23:
The three primary ways are method of moments, conditional LSE, and MLE. The asymptotics for conditional LSE and MLE, initialized with method of moments estimators...

Exercise 24: What are the Yule-Walker equations? How are they related to the estimation methods above?

Solution 24:
The Yule-Walker equations are basically a way to find the method of moment estimators for AR(p). They happen to be equivalent to the conditional LSE method, so the method of moment estimator is optimal in this case.

Exercise 25: Construct a method of moments estimator for MA(1). Show how you would construct a more optimal parameter estimate for MA(q), and show which of the three estimation methods are related to this more-optimal method.

Solution 25:
The innovations algorithm can be used to find a more optimal parameter estimate - it is an iterative method for finding the conditional LSE estimate.

Exercise 26: Show that the asymptotic distributions of $\hat{\phi}$ and $\hat{\theta}$ from AR(1) and MA(1) (respectively) are of the same form.

Exercise 27: What is the spectral distribution function? What is the spectral density function? Find the spectral distribution of $X_t = U_1 cos(2\pi\omega_0 t) + U_2 sin(2\pi\omega_0 t)$, where U_1, U_2 are uncorrelated zero-mean random variables with equal variance σ^2.

Solution 27:
If $\{x_t\}$ is stationary with autocovariance γ_x, there exists a monotonically increasing right-continuous-left-limit function $F(\omega)$ with $F(-\infty) = F(-1/2) = 0$ and $F(\infty) = F(1/2) = \gamma(0)$ such that

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i\omega h} dF(\omega).$$

The spectral density function is $F'(\theta) = f(\theta)$.

Define $c_t = \cos(2\pi\omega_0 t)$ and $s_t = \sin(2\pi\omega_0 t)$. Then

$$\begin{aligned}\gamma(h) &= \sigma^2(c_{t+h}c_t + s_{t+h}s_t)\\ &= \sigma^2 \cos(2\pi\omega_0 h)\\ &= \sigma^2 \left(\frac{\exp(2\pi i\omega_0 h)}{2} + \frac{\exp(-2\pi i\omega_0 h)}{2}\right).\end{aligned}$$

This implies that

$$F(\omega) = \begin{cases} 0, & \omega < -\omega_0 \\ \frac{\sigma^2}{2}, & -\omega_0 \le \omega < \omega_0 \\ \sigma^2, & \omega_0 \le \omega. \end{cases}$$

Exercise 28: What is the Discrete Fourier Transform (DFT) and its inverse? Decompose it into its sine and cosine parts.

Solution 28:
Given $X_1, \cdots, X_n$, and fundamental frequencies $\omega_j = j/n$ for $j = 0, \cdots, n-1$, the DFT is defined as

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} X_t \exp(-2\pi i\omega_j t).$$

The inverse DFT is given by

$$X_t = n^{-1/2} \sum_{j=0}^{n-1} d(\omega_j) \exp(2\pi i\omega_j t).$$

The cosine and sine parts of the DFT are

$$d_c(\omega_j) = n^{-1/2} \sum_{t=1}^{n} X_t \cos(2\pi\omega_j t)$$
$$d_s(\omega_j) = n^{-1/2} \sum_{t=1}^{n} X_t \sin(2\pi\omega_j t).$$

This implies that $d(\omega_j) = d_c(\omega_j) - i \cdot d_s(\omega_j)$.

Exercise 29: What is the relationship between the spectral density function and the autocovariance for a stationary process with weak memory?

Solution 29:
If $\gamma_x(\cdot)$ of a stationary series $\{x_t\}$ has weak memory, then

$$\gamma(h) = \int_{-1/2}^{1/2} \exp(2\pi i\omega h) f(\omega) d\omega.$$

The inverse transform is

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-2\pi i\omega h).$$

Exercise 30: Suppose $y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j}$, and $\sum_{j=-\infty}^{\infty} |a_j| < \infty$. Let $A(\omega) = \sum_{j=-\infty}^{\infty} a_j \exp(-2\pi i\omega_j)$. Show that

$$f_y(\omega) = |A(\omega)|^2 f_x(\omega).$$

Exercise 31: Find the spectral density functions of AR(1), MA(1), and ARMA(1,1).

Solution 31:
Let $\rho = \exp(-2\pi i\omega)$. For AR(1), suppose $x_t = \phi x_{t-1} + w_t$. Then write

$$w_t = x_t - \phi x_{t-1}$$
$$w_0 = x_0 - \phi x_{-1}.$$

Then $\mathbb{E}(w_t w_0) = \sigma_w^2$ if and only if $t = 0$. Hence

$$\sigma_w^2 = f_x(\omega)(1 + \phi^2 - \phi\rho^{-1} - \phi\rho).$$

This means

$$f_x(\omega) = \frac{\sigma_w^2}{1 - \phi\rho^{-1} - \phi\rho + \phi^2}.$$

For MA(1), suppose $x_t = w_t + \theta w_{t-1}$. Then write

$$\begin{aligned} x_t &= w_t + \theta w_{t-1} \\ x_0 &= w_0 + \theta w_{-1}. \end{aligned}$$

Then using the same summing trick, we get

$$f_x(\omega) = \sigma_2^2(1 + \theta^2 + \theta\rho + \theta\rho^{-1}).$$

For ARMA(1,1), suppose $x_t - \phi x_{t-1} = w_t + \theta w_{t-1}$. Then using the same summing trick, we get

$$f_x(\omega) = \sigma_2^2 \frac{1 + \theta^2 + \theta\rho + \theta\rho^{-1}}{1 + \phi^2 - \phi\rho - \phi\rho^{-1}}.$$

Exercise 32: Show that if $\{X_T\} \sim ARMA(p, q)$ with autoregressive polynomial ϕ and moving average polynomial θ, and $\rho(\omega) = e^{-2\pi i\omega}$, then the spectral density function is given by

$$f_x(\omega) = \sigma_w^2 \frac{|\theta(\rho)|^2}{|\phi(\rho)|^2}.$$

Solution 32:
Use the summing trick.

Exercise 33: What is the periodogram? Decompose into sine and cosine components.

Solution 33:
Given $X_1, \cdots X_n$ and the fundamental frequencies $\omega_j = j/n$ for $j = 0, \cdots, n-1$, the periodogram is defined as

$$\begin{aligned} I(\omega_j) &= |d(\omega_j)|^2 \\ &= n^{-1} \left| \sum_{t=1}^{n} X_t \exp(-2\pi i \omega_j t) \right|^2 \\ &= (d_c(\omega_j) - i \cdot d_s(\omega_j))(d_c(\omega_j) + i \cdot d_s(\omega_j)) \\ &= d_c(\omega_j)^2 + d_s(\omega_j)^2. \end{aligned}$$

Decomposed into its sine and cosine components:

$$\begin{aligned} I(\omega_j) &= |d(\omega_j)|^2 \\ &= d_c(\omega_j)^2 + d_s(\omega_j)^2. \end{aligned}$$

Exercise 34: Show the relationship between the periodogram and the ANOVA table.

Exercise 35: Give three nonparametric estimators of the spectral density function, and derive their inference properties.

Solution 35:
Three nonparametric estimators:

1. Periodogram
2. Smoothed periodogram
3. Lagged-window periodogram

Periodogram. Choose $\omega_j \approx \omega$. We can write the periodogram as follows:

$$I(\omega_j) = \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) \exp(-2\pi i \omega_j h).$$

The periodogram is hence the naive spectral density estimator - it just substitutes estimators for autocovariance. To derive some inference properties, notice that if $x_t \sim_{\text{i.i.d.}} N(0,1)$, then for any $j = 0, \cdots, n-1$, $d_c(\omega_j), d_s(\omega_j) \sim N(0, 1/2)$. Also, the two are independent. This means

$$2I(\omega_j) = \underbrace{(\sqrt{2} d_c(\omega_j))^2}_{\sim N(0,1)^2} + \underbrace{(\sqrt{2} d_s(\omega_j))^2}_{\sim N(0,1)^2} \sim \chi_2^2.$$

(Note that the spectral density is 1 if x_t are iid.) In general, if x_t is stationary, then for any $j = 0, \cdots, n-1$,

$$\frac{2I(\omega_j)}{f(\omega_j)} \to_d \chi_2^2 \iff \frac{I(\omega_j)}{f(\omega_j)} \to_d \exp(1).$$

In fact,

$$\frac{I(\omega_j)}{f(\omega_j)} \to_{\text{i.i.d.}} \exp(1).$$

The periodogram is unbiased (just note that expectation of exp(1) is 1), but it has shitty variance.

Smoothed Periodogram. Let $\kappa_m(\cdot)$ be a kernel function. Choose $\omega_j \approx \omega$. The smoothed periodogram is defined as

$$\bar{I}(\omega) = \sum_{k=-m}^{m} \kappa_m(k) \cdot I\left(\omega_j + \frac{k}{n}\right).$$

The intuition is that most spectral densities change very little over small intervals - hence by averaging, we can hopefully reduce variability at the cost of increasing bias.

For bias, we have

$$\begin{aligned}
\mathbb{E}(\bar{I}(\omega)) &= \mathbb{E}\left[\sum_{k=-m}^{m} \kappa_m(k) \cdot I\left(\omega_j + \frac{k}{n}\right)\right] \\
&\approx \sum_{k=-m}^{m} \kappa_m(k) \cdot f\left(\omega + \frac{k}{n}\right) \\
&\approx \sum_{k=-m}^{m} \kappa_m(k) \cdot \left[f(\omega) + \frac{k}{n} f'(\omega) + \frac{1}{2}\frac{k^2}{n^2} f''(\omega)\right] \\
&= f(\omega) + \frac{f''(\omega)}{2n^2} \sum_{k=-m}^{m} \kappa_m(k) \cdot k^2.
\end{aligned}$$

This means bias is approximated by $\frac{f''(\omega)}{2n^2} \sum_{k=-m}^{m} \kappa_m(k) \cdot k^2$.

For variance, recall that the periodogram at the fundamental frequencies are approximately independently distributed according to $f(\omega_j) \cdot \exp(1)$. This means

$$\begin{aligned}
\mathrm{Var}(\bar{I}(\omega)) &\approx \sum_{k=-m}^{m} \kappa_m(k)^2 \mathrm{Var}\left(I\left(\omega_j + \frac{k}{n}\right)\right) \\
&\approx \sum_{k=-m}^{m} \kappa_m(k)^2 f(\omega)^2 \\
&= f(\omega)^2 \sum_{k=-m}^{m} \kappa_m(k)^2.
\end{aligned}$$

From now on, assume the rectangular kernel i.e. $\kappa_m(k) = \frac{1}{2m+1}$ for $-m \leq k \leq m$. This kernel gives

$$\begin{aligned}
\mathrm{Bias}(\bar{I}(\omega)) &\approx \frac{f''(\omega)}{n^2} \cdot \frac{m(m+1)}{6} \propto (m/n)^2, \\
\mathrm{Var}(\bar{I}(\omega)) &\approx \frac{f(\omega)^2}{2m+1}.
\end{aligned}$$

The bias converges to zero as $n \to \infty$ as long as $m/n \to 0$. The variances converges to zero as $m \to \infty$. For asymptotics, we can obtain two asymptotic distributions. For the first, notice that the periodograms are roughly independent, and assume that the smoothed periodogram is roughly unbiased. By the CLT, we have

$$\begin{aligned}
\bar{I}(\omega) - f(\omega) &\to N\left(0, \frac{f(\omega)^2}{2m+1}\right) \text{ or} \\
\sqrt{2m+1}(\bar{I}(\omega) - f(\omega)) &\to N(0, f(\omega)^2).
\end{aligned}$$

However, notice also that $\bar{I}(\omega)$ is roughly the sum of approximately independent chi-square distributions. This suggests

$$\frac{2(2m+1)\bar{I}(\omega)}{f(\omega)} \to \chi^2_{2(2m+1)}.$$

For both asymptotics, we use log to stabilize the variance. In particular, the confidence intervals are

$$\text{CLT: } \log(f(\omega)) \pm \left[\log(\bar{I}(\omega)) \pm \frac{z_{1-\alpha/2}}{\sqrt{2m+1}}\right]$$

$$\text{Chi: } \log(\bar{\omega}) + \log\frac{2(2m+1)}{\chi_{1-\alpha/2}} \leq \log(f(\omega)) \leq \log(\bar{\omega}) + \log\frac{2(2m+1)}{\chi_{\alpha/2}}$$

Lagged-window periodogram. This is exactly like the periodogram, except for large lags, we estimate $\hat{\gamma}$ to be zero. The lagged-window periodogram is

$$\tilde{I}(\omega) = \sum_{h=-B}^{B} \hat{\gamma}(h)\exp(-2\pi i\omega_j h), \text{ where } B \sim n^{-1/3}.$$

Exercise 36: Give a parametric estimator of the spectral density function.

Solution 36:
Because we know the form of the spectral density function for a ARMA(p,q) process, we can estimate the parameters of the process to get

$$f_x(\omega) = \hat{\sigma}_w^2 \frac{|\hat{\theta}(\rho)|^2}{|\hat{\phi}(\rho)|^2}, \text{ where } \sigma_w^2 \text{ is the sample variance.}$$

Exercise 37: Define cross-spectrum and cross-coherence. What is an estimator for the cross-spectrum?

Solution 37:
The cross-spectrum is defined as $f_{XY}(\omega) = \sum_{h=-\infty}^{\infty} \gamma_{XY}(h)\rho^h$, where $\rho = \exp(-2\pi i\omega)$. We can impute $\hat{\gamma}$ for an estimator. Cross-coherence is defined as $\rho_{xy} = \frac{|f_{XY}(\omega)|}{\sqrt{f_{XX}(\omega)}\sqrt{f_{YY}(\omega)}}$.

Exercise 38: How would you estimate the parameters for $x_t = a|x_{t-1}| + w_t$?

Solution 38:
Define the model

$$x_t = a\mathbb{1}_{x_{t-1}\geq 0} - a\mathbb{1}_{x_{t-1}\leq 0} + w_t.$$

Then we can estimate a in the same way as linear regression.

Exercise 39: What is the ARCH(1,1) model? State some properties about the model. How would you estimate the parameters?

Solution 39:

The ARCH(1,1) model is defined as

$$x_t = \epsilon_t \sqrt{a_0^2 + a_1^2 x_{t-1}^2}, \epsilon_t \sim_{\text{i.i.d.}} N(0,1).$$

Some properties:

1. The conditional distribution $x_t | x_{t-1} = x \sim N(0, a_0^2 + a_1^2 x^2)$.
2. x_t is a martingale.
3. $\mathbb{E}(x_t^2) < \infty$ if and only if $a_1^2 < 1$.
4. $\mathbb{E}(\log |a_1 \epsilon_1|) < 0$ if and only if x_t is stationary.
5. If $a_1^4 < \frac{1}{3}$, then the fourth moment is finite.

We can estimate using MLE on the distribution of $x_t | x_{t-1} = x$. In particular, we have

$$a_0, a_1 = \arg\max \prod_{t=2}^{n} f(x_t | x_{t-1}).$$

Exercise 40: Define the negative binomial. What is the gamma function?

Solution 40:
If $a \in \mathbb{Q}$, then

$$\binom{a}{k} = \frac{a(a-1)\cdots(a-(k-1))}{k!}.$$

The gamma function is

$$\Gamma(\lambda) = \int_0^\infty e^{-x} x^{\lambda-1} dx.$$

Notice that $\Gamma(n) = (n-1)!$ if $n \in \mathbb{N}$ and $\Gamma(\lambda) = \Gamma(\lambda + 1)$.

Exercise 41: Define ARFIMA(p, d, q).

Solution 41:
Same as ARIMA, but with fractional differencing based on the negative binomial.