STAT 626: Outline of Lecture 4
Correlation and Dependence (§2.1)

1. Mean Function of a Time Series (Stochastic Process),

2. Covariance Function of a Time Series,

3. Stationary Time Series

4. NOTES: Review of Mean, Variance, Covariance of Lin Comb. of RVs.

Topics From Chapter 1

5. White Noise: The Building Blocks

6. Autoregression: The Birth of Modern Time Series Analysis

7. Random Walks: The Engine of Financial Engineering

8. Signal + Noise: For Other Engineering

DEFINITIONS

1. A Time Series $\{x_t\}$ is **stationary** if

   (a) the mean function $E(x_t)$ does not depend on the time t,

   (b) the covariance function $\text{cov}(x_s, x_t)$ depends on the times s, t only through the distance $|s-t|$.

2. **Autocovariance Function** of a Stationary Time Series:

$$\gamma(h) = \text{cov}(x_{t+h}, x_t), \quad h = 0, 1, \ldots.$$

   NOTE:

$$\gamma(0) = \text{cov}(x_t, x_t) = \text{var}(x_t).$$

3. **The Autocorrelation Function (ACF)**

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad h = 0, 1, \ldots,$$

   is a **symmetric** function of the lag h. **Correlogram** is the plot of $\rho(h)$ vs h.

4. **Multivariate Time Series:**
   Cross-Covariance Function (CCF) of Two Time Series:

$$\gamma_{xy}(h) = \text{Cov}(x_{t+h}, y_t), h = 0, \pm 1, \pm 2, \ldots$$

   Why ACF is symmetric and CCF is not?
   Proof without words!

$$\gamma_{xx}(h) = \text{Cov}(x_{t+h}, x_t) = \text{Cov}(x_t, x_{t+h}) = \gamma_{xx}(-h), \quad h = 1, 2, \ldots.$$

Chapter 2

Correlation and Stationary Time Series

2.1 Measuring Dependence

We now discuss various measures that describe the general behavior of a process as it evolves over time. The material on probability in Appendix B may be of help with some of the content in this chapter. A rather simple descriptive measure is the mean function, such as the average monthly high temperature for your city. In this case, the mean is a *function of time*.

Definition 2.1. *The* **mean function** *is defined as*

$$\mu_{xt} = E(x_t) \tag{2.1}$$

provided it exists, where E denotes the usual expected value operator. When no confusion exists about which time series we are referring to, we will drop a subscript and write μ_{xt} as μ_t.

Example 2.2. Mean Function of a Moving Average Series
If w_t denotes a white noise series, then $\mu_{wt} = E(w_t) = 0$ for all t. The top series in Figure 1.8 reflects this, as the series clearly fluctuates around a mean value of zero. Smoothing the series as in Example 1.8 does not change the mean because we can write

$$\mu_{vt} = E(v_t) = \tfrac{1}{3}[E(w_{t-1}) + E(w_t) + E(w_{t+1})] = 0. \qquad \diamond$$

Example 2.3. Mean Function of a Random Walk with Drift
Consider the random walk with drift model given in (1.4),

$$x_t = \delta\, t + \sum_{j=1}^{t} w_j, \qquad t = 1, 2, \dots .$$

Because $E(w_t) = 0$ for all t, and δ is a constant, we have

$$\mu_{xt} = E(x_t) = \delta\, t + \sum_{j=1}^{t} E(w_j) = \delta\, t$$

which is a straight line with slope δ. A realization of a random walk with drift can be compared to its mean function in Figure 1.10. $\diamond$

Example 2.4. Mean Function of Signal Plus Noise

A great many practical applications depend on assuming the observed data have been generated by a fixed signal waveform superimposed on a zero-mean noise process, leading to an additive signal model of the form (1.5). It is clear, because the signal in (1.5) is a fixed function of time, we will have

$$\begin{aligned}\mu_{xt} &= E\left[2\cos(2\pi\tfrac{t+15}{50}) + w_t\right]\\ &= 2\cos(2\pi\tfrac{t+15}{50}) + E(w_t)\\ &= 2\cos(2\pi\tfrac{t+15}{50}),\end{aligned}$$

and the mean function is just the cosine wave. ♢

The mean function describes only the marginal behavior of a time series. The lack of independence between two adjacent values x_s and x_t can be assessed numerically, as in classical statistics, using the notions of covariance and correlation. Assuming the variance of x_t is finite, we have the following definition.

Definition 2.5. *The* **autocovariance function** *is defined as the second moment product*

$$\gamma_x(s,t) = \text{cov}(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)], \tag{2.2}$$

for all s and t. When no possible confusion exists about which time series we are referring to, we will drop the subscript and write $\gamma_x(s,t)$ as $\gamma(s,t)$.

Note that $\gamma_x(s,t) = \gamma_x(t,s)$ for all time points s and t. The autocovariance measures the *linear* dependence between two points on the same series observed at different times. Recall from classical statistics that if $\gamma_x(s,t) = 0$, then x_s and x_t are not linearly related, but there still may be some dependence structure between them. If, however, x_s and x_t are bivariate normal, $\gamma_x(s,t) = 0$ ensures their independence. It is clear that, for $s = t$, the autocovariance reduces to the (assumed finite) *variance*, because

$$\gamma_x(t,t) = E[(x_t - \mu_t)^2] = \text{var}(x_t). \tag{2.3}$$

Example 2.6. Autocovariance of White Noise

The white noise series w_t has $E(w_t) = 0$ and

$$\gamma_w(s,t) = \text{cov}(w_s, w_t) = \begin{cases} \sigma_w^2 & s = t, \\ 0 & s \neq t. \end{cases} \tag{2.4}$$

A realization of white noise is shown in the top panel of Figure 1.8. ♢

We often have to calculate the autocovariance between filtered series. A useful result is given in the following proposition.

Property 2.7. *If the random variables*

$$U = \sum_{j=1}^{m} a_j X_j \quad \text{and} \quad V = \sum_{k=1}^{r} b_k Y_k$$

are linear filters of (finite variance) random variables $\{X_j\}$ *and* $\{Y_k\}$*, respectively, then*

$$\operatorname{cov}(U, V) = \sum_{j=1}^{m} \sum_{k=1}^{r} a_j b_k \operatorname{cov}(X_j, Y_k). \tag{2.5}$$

Furthermore, $\operatorname{var}(U) = \operatorname{cov}(U, U)$.

An easy way to remember (2.5) is to treat it like multiplication:

$$(a_1 X_1 + a_2 X_2)(b_1 Y_1) = a_1 b_1 X_1 Y_1 + a_2 b_1 X_2 Y_1$$

Example 2.8. Autocovariance of a Moving Average
Consider applying a three-point moving average to the white noise series w_t of the previous example as in Example 1.8. In this case,

$$\gamma_v(s, t) = \operatorname{cov}(v_s, v_t) = \operatorname{cov}\left\{ \tfrac{1}{3}\left(w_{s-1} + w_s + w_{s+1}\right), \tfrac{1}{3}\left(w_{t-1} + w_t + w_{t+1}\right) \right\}.$$

When $s = t$ we have

$$\begin{aligned} \gamma_v(t, t) &= \tfrac{1}{9}\operatorname{cov}\{(w_{t-1} + w_t + w_{t+1}), (w_{t-1} + w_t + w_{t+1})\} \\ &= \tfrac{1}{9}[\operatorname{cov}(w_{t-1}, w_{t-1}) + \operatorname{cov}(w_t, w_t) + \operatorname{cov}(w_{t+1}, w_{t+1})] \\ &= \tfrac{3}{9}\sigma_w^2. \end{aligned}$$

When $s = t + 1$,

$$\begin{aligned} \gamma_v(t+1, t) &= \tfrac{1}{9}\operatorname{cov}\{(w_t + w_{t+1} + w_{t+2}), (w_{t-1} + w_t + w_{t+1})\} \\ &= \tfrac{1}{9}[\operatorname{cov}(w_t, w_t) + \operatorname{cov}(w_{t+1}, w_{t+1})] \\ &= \tfrac{2}{9}\sigma_w^2, \end{aligned}$$

using (2.4). Similar computations give $\gamma_v(t-1, t) = 2\sigma_w^2/9$, $\gamma_v(t+2, t) = \gamma_v(t-2, t) = \sigma_w^2/9$, and 0 when $|t - s| > 2$. We summarize the values for all s and t as

$$\gamma_v(s, t) = \begin{cases} \frac{3}{9}\sigma_w^2 & s = t, \\ \frac{2}{9}\sigma_w^2 & |s - t| = 1, \\ \frac{1}{9}\sigma_w^2 & |s - t| = 2, \\ 0 & |s - t| > 2. \end{cases} \tag{2.6}$$

◊

Example 2.9. Autocovariance of a Random Walk
For the random walk model, $x_t = \sum_{j=1}^{t} w_j$, we have

$$\gamma_x(s,t) = \operatorname{cov}(x_s, x_t) = \operatorname{cov}\left(\sum_{j=1}^{s} w_j, \sum_{k=1}^{t} w_k\right) = \min\{s,t\}\,\sigma_w^2\,,$$

because the w_t are uncorrelated random variables. For example, with $s = 2$ and $t = 4$,

$$\operatorname{cov}(x_2, x_4) = \operatorname{cov}(w_1 + w_2, w_1 + w_2 + w_3 + w_4) = 2\sigma_w^2\,.$$

Note that, as opposed to the previous examples, the autocovariance function of a random walk depends on the particular time values s and t, and not on the time separation or lag. Also, notice that the variance of the random walk, $\operatorname{var}(x_t) = \gamma_x(t,t) = t\,\sigma_w^2$, increases without bound as time t increases. The effect of this variance increase can be seen in Figure 1.10 where the processes start to move away from their mean functions $\delta\, t$ (note that $\delta = 0$ and .3 in that example). ◊

As in classical statistics, it is more convenient to deal with a measure of association between -1 and 1, and this leads to the following definition.

Definition 2.10. *The* **autocorrelation function (ACF)** *is defined as*

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}. \tag{2.7}$$

The ACF measures the linear predictability of the series at time t, say x_t, using only the value x_s. And because it is a correlation, we must have $-1 \leq \rho(s,t) \leq 1$. If we can predict x_t *perfectly* from x_s through a linear relationship, $x_t = \beta_0 + \beta_1 x_s$, then the correlation will be $+1$ when $\beta_1 > 0$, and -1 when $\beta_1 < 0$. Hence, we have a rough measure of the ability to forecast the series at time t from the value at time s.

Often, we would like to measure the predictability of another series y_t from the series x_s. Assuming both series have finite variances, we have the following definition.

Definition 2.11. *The* **cross-covariance function** *between two series, x_t and y_t, is*

$$\gamma_{xy}(s,t) = \operatorname{cov}(x_s, y_t) = E[(x_s - \mu_{xs})(y_t - \mu_{yt})]. \tag{2.8}$$

We can use the cross-covariance function to develop a correlation:

Definition 2.12. *The* **cross-correlation function (CCF)** *is given by*

$$\rho_{xy}(s,t) = \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\gamma_y(t,t)}}. \tag{2.9}$$

2.2 Stationarity

Although we have previously not made any special assumptions about the behavior of the time series, many of the examples we have seen hinted that a sort of regularity may exist over time in the behavior of a time series. Stationarity requires regularity in the mean and autocorrelation functions so that these quantities (at least) may be estimated by averaging.

Definition 2.13. *A* **stationary** *time series is a finite variance process where*

(i) the mean value function, μ_t, defined in (2.1) is constant and does not depend on time t, and

(ii) the autocovariance function, $\gamma(s,t)$, defined in (2.2) depends on times s and t only through their time difference.

As an example, for a stationary hourly time series, the correlation between what happens at 1AM and 3AM is the same as between what happens at 9PM and 11PM because they are both two hours apart.

Example 2.14. A Random Walk is Not Stationary
A random walk is not stationary because its autocovariance function, $\gamma(s,t) = \min\{s,t\}\sigma_w^2$, depends on time; see Example 2.9 and Problem 2.5. Also, the random walk with drift violates both conditions of Definition 2.13 because the mean function, $\mu_{xt} = \delta t$, depends on time t as shown in Example 2.3

Because the mean function, $E(x_t) = \mu_t$, of a stationary time series is independent of time t, we will write

$$\mu_t = \mu. \tag{2.10}$$

Also, because the autocovariance function, $\gamma(s,t)$, of a stationary time series, x_t, depends on s and t only through time difference, we may simplify the notation. Let $s = t + h$, where h represents the time shift or *lag*. Then

$$\gamma(t+h,t) = \text{cov}(x_{t+h}, x_t) = \text{cov}(x_h, x_0) = \gamma(h,0)$$

because the time difference between $t+h$ and t is the same as the time difference between h and 0. Thus, the autocovariance function of a stationary time series does not depend on the time argument t. Henceforth, for convenience, we will drop the second argument of $\gamma(h,0)$.

Definition 2.15. *The* **autocovariance function of a stationary time series** *will be written as*

$$\gamma(h) = \text{cov}(x_{t+h}, x_t) = E[(x_{t+h} - \mu)(x_t - \mu)]. \tag{2.11}$$

Definition 2.16. *The* **autocorrelation function (ACF) of a stationary time series** *will be written using (2.7) as*

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}. \tag{2.12}$$

STAT 626: Review of Mean and Variance of Lin. Combination of Random Variables

Mean and Variance of Sum of Random Variables

1. If X is a random variable with mean $E(X) = \mu$ and variance

$$\sigma^2 = E(X-\mu)^2 = E(X^2) - \mu^2.$$

Then for c a constant, we have

(a) $E(cX) = cE(X) = c\mu$,

(b) $\mathrm{Var}(cX) = c^2\sigma^2$.

2. If X, Y are two random variables with means μ_1, μ_2, variances σ_1^2, σ_2^2 and covariance

$$\sigma_{12} = \mathrm{Cov}(X, Y) = E(X - \mu_X)(Y - \mu_Y),$$

then

(a). $E(X+Y) = \mu_1 + \mu_2$.

(b). $\mathrm{Var}(X+Y) = \sigma_1^2 + \sigma_2^2 + 2\sigma_{12}$.

Exercise 1 : If X, Y are random variables as above and c_1, c_2 are constants, write the formulas for

$$E(c_1X + c_2Y), \quad \mathrm{Var}(c_1X + c_2Y),$$

and $Q(b) = \mathrm{Var}(Y - bX)^2$, where b is a scalar.

Exercise 2 : If X, Y are random variables with

$$\mu_1 = 2, \quad \mu_2 = -5, \quad \sigma_1^2 = 4, \quad \sigma_2^2 = 17, \quad \sigma_{12} = -3,$$

and $c_1 = -2, \quad c_2 = 1$, find the numerical values of

$$E(c_1X + c_2Y), \quad \mathrm{Var}(c_1X + c_2Y), \quad Q(b).$$

Compute the correlation coefficient ρ between X and Y.

Exercise 3: If $X_1, \ldots, X_n$ are independent random variables with $E(X_i) = \mu_i, \mathrm{Var}(X_i) = \sigma_i^2$, and c_1, c_2 are constant scalars, find

$$E(c_1X_1 + c_2X_2), \quad \mathrm{Var}(c_1X_1 + c_2X_2),$$

and $\mathrm{Var}(\bar{X})$, where $\bar{X}$ is the sample mean.

Exercise 4: Show that for any scalars a_1, a_2, a_3 and random variables X_1, X_2, X_3: $\mathrm{Var}(a_1X_1 + a_2X_2 + a_3X_3) =$

$a_1^2\mathrm{Var}(X_1)+a_2^2\mathrm{Var}(X_2)+a_3^2\mathrm{Var}(X_3)+2a_1a_2\mathrm{Cov}(X_1, X_2)+2a_1a_3\mathrm{Cov}(X_1, X_3)+2a_2a_3\mathrm{Cov}(X_2, X_3)$.

(a) If Cov (X_1, X_2)= Cov (X_2, X_3)=ρ, Cov $(X_1, X_3) = \rho^2$ and $\mathrm{Var}(X_i) = 1, i = 1, 2, 3$, then write the 3×3 covariance matrix of the random vector $X = (X_1, X_2, X_3)$.

(b) Compute $\mathrm{Var}(X_1 + X_2 + X_3)$ when $\rho = 0.6$.

(c) Mark T is interested in forecasting X_3 using the linear predictor $\widehat{X}_3 = b_2X_2 + b_1X_1$. He realizes the forecast error is $X_3 - \widehat{X}_3 = X_3 - b_2X_2 - b_1X_1$ and a great way to find the predictor coefficients b_1, b_2, is by *minimizing the variance of forecast error*

$$Q(b_1, b_2) = \mathrm{Var}(X_3 - b_2X_2 - b_1X_1),$$

which turns out to be a quadratic function of b_1, b_2 (as in least-squares estimation in regression). Help Mark to minimize this function or derive the normal equations.

(d) Solve the normal equations and observe that $b_1 = 0$, regardless of the value of ρ.

Exercise 5: Now consider the random variables X_1, X_2, X_3.

(a) If Cov (X_1, X_2)= Cov (X_2, X_3)= Cov $(X_1, X_3) = \rho$ and $\mathrm{Var}(X_i) = 1, i = 1, 2, 3$, write the 3×3 covariance matrix of the random vector $X = (X_1, X_2, X_3)$.

(b) Express $\mathrm{Var}(X_1 + X_2 + X_3)$ in terms of ρ.

(c) In forecasting X_3 using the linear predictor $\widehat{X}_3 = b_2X_2 + b_1X_1$, the forecast error is $X_3 - \widehat{X}_3 = X_3 - b_2X_2 - b_1X_1$, find the predictor coefficients b_1, b_2, by *minimizing the variance of forecast error*

$$Q(b_1, b_2) = \mathrm{Var}(X_3 - b_2X_2 - b_1X_1),$$

which turns out to be a quadratic function of b_1, b_2. Minimize this function or derive the normal equations.

(d) Solve the normal equations and express b_1 and b_2 in terms of ρ. Compare these predictor coefficients with those in the previous exercise.

Exercise 6: Suppose all pairwise covariances between the (past) random variables $X_1, X_2, \ldots, X_p$ and a (future) random variable X_{p+1} are known and given by

$$\mathrm{Cov}(X_i, X_j) = \rho^{|i-j|}.$$

(a) Organize the above pairwise covariance information in a $(p+1) \times (p+1)$ matrix. For $p = 4, \rho = 0.5$, write the form of this matrix.

(b) Explain in words why you might be interested in choosing the b_i's so that

$$Q(b_1, b_2, \ldots, b_p) = \mathrm{Var}(X_{p+1} - b_1X_1 - b_2X_2 - \ldots - b_pX_p),$$

is *minimized*. (Hint: Think in terms of prediction or forecasting).

(c) Derive the *normal equations* for minimizing $Q(b_1, b_2, \ldots, b_p)$.

(d) Write the equations in (c) in matrix form. What does it take to solve it for the b_i's.

Exercise 7: Suppose all pairwise covariances between the (past) random variables $X_1, X_2, \ldots, X_p$ and a (future) random variable Y are known and given by Cov $(X_i, X_j) = \rho, \quad i \neq j$, and Cov $(X_i, Y) = 0$, for all i.

(a) Organize the above pairwise covariance information in a $(p+1) \times (p+1)$ matrix. For $p = 4, \rho = 0.5$, write the form of this matrix.

(b) One is interested in choosing the b_i's so that

$$Q(b_1, b_2, \ldots, b_p) = \mathrm{Var}(Y - b_1X_1 - b_2X_2 - \ldots - b_pX_p),$$

is *minimized*. Derive the *normal equations* for minimizing $Q(b_1, b_2, \ldots, b_p)$.

(c) Write the equations in (b) in matrix form and solve it to find the b_i's.