Signal vs. Noise

Jared Fisher

Lecture 0b

Announcements

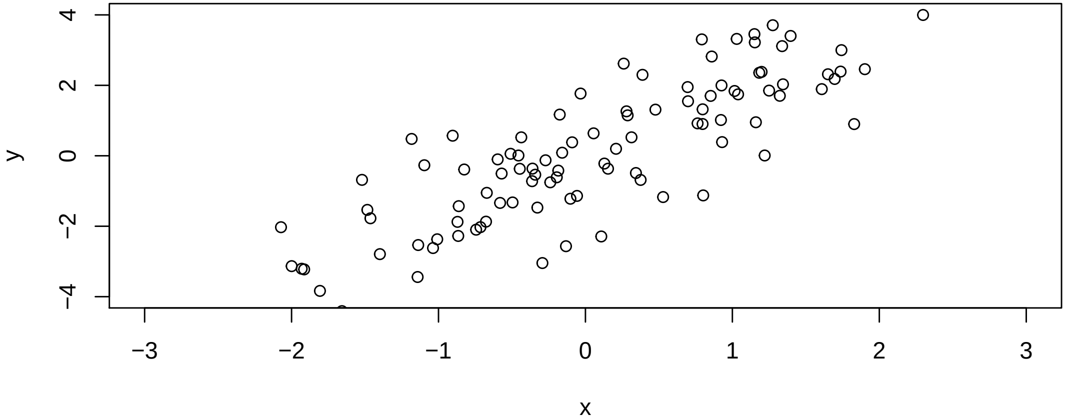
- “Practice Homework 0” has been posted. **This is not graded!** But it will give a chance to those who want to practice turning something in before turning in a “real” assignment. It also reviews pre-requisite material.
- Homework 0 is “due” January 27, but it’s not graded so it’s fine if it’s late too. Again, it’s for practice if you want.
- Must be submitted as a PDF on Gradescope.

Accomodations and Schedule Conflicts

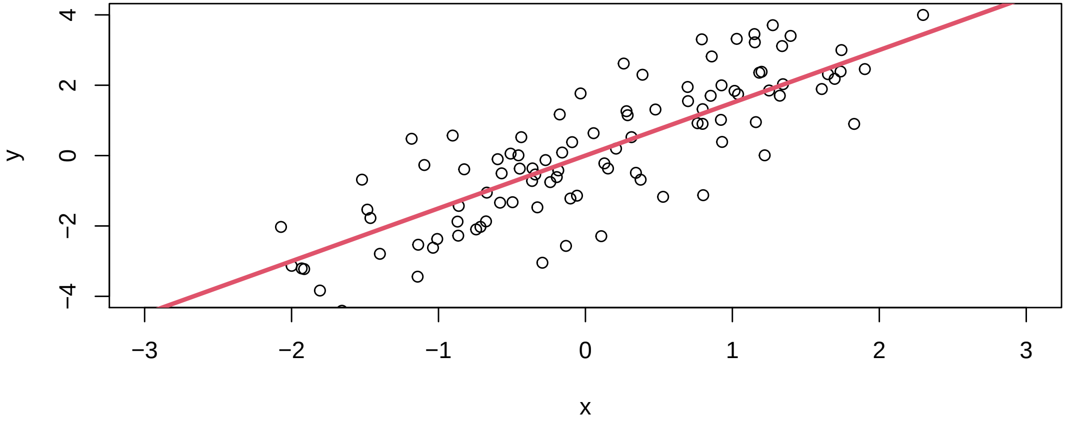
- Please let me know of any conflicts or accommodations (religious, DSP, or otherwise) as soon as possible.

Parallels between Ideas behind Time Series and Regression

Regression



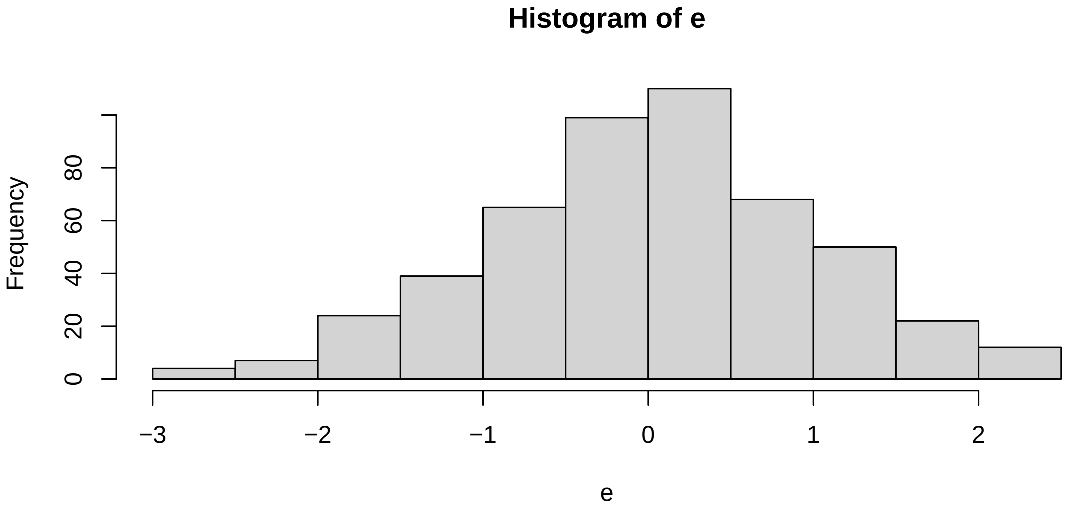
Regression



Regression $=$ Signal $+$ Noise

- $y_i = f(\mathbf{x_i}) + \epsilon_\mathbf{i}$
- $f(\mathbf{x_i}) \Rightarrow$ "Signal"
- $\epsilon_i \Rightarrow$ "Noise"
- Often $\epsilon_i \sim N(0, \sigma^2)$

Gaussian Errors

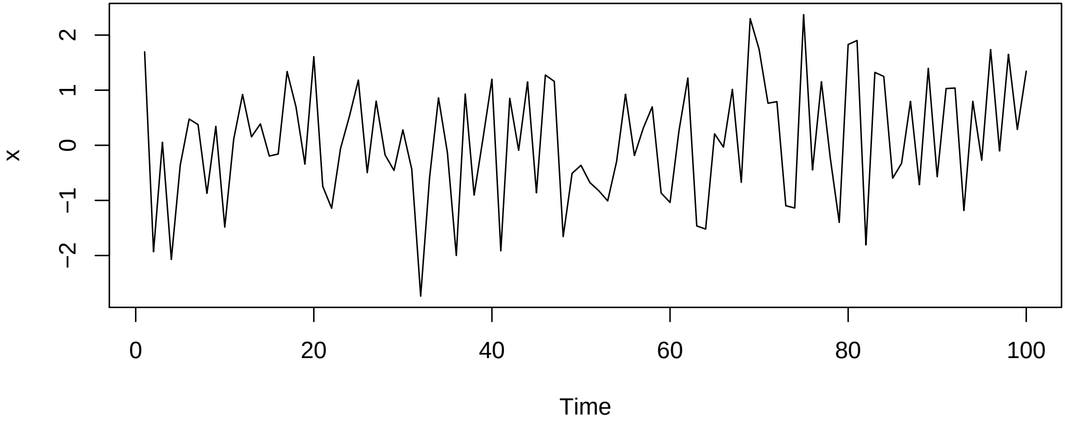


But. . .

- In this class, we don't have data on n subjects at 1 point in time, we will have data on 1 subject at n points in time. . .
- So, instead think of the model as $y_t = f(t) + \epsilon_t$, and now view the ϵ's over time
- Or, with no signal: $y_t = \epsilon_t$

Gaussian Errors, over time

```
plot.ts(x)
```



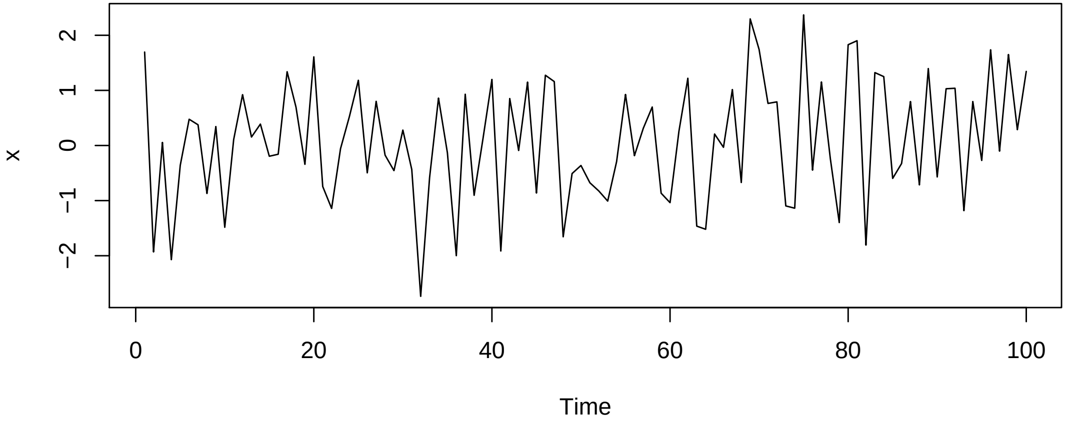
We call this "Gaussian noise"

Definitions (TSA4e Example 1.8)

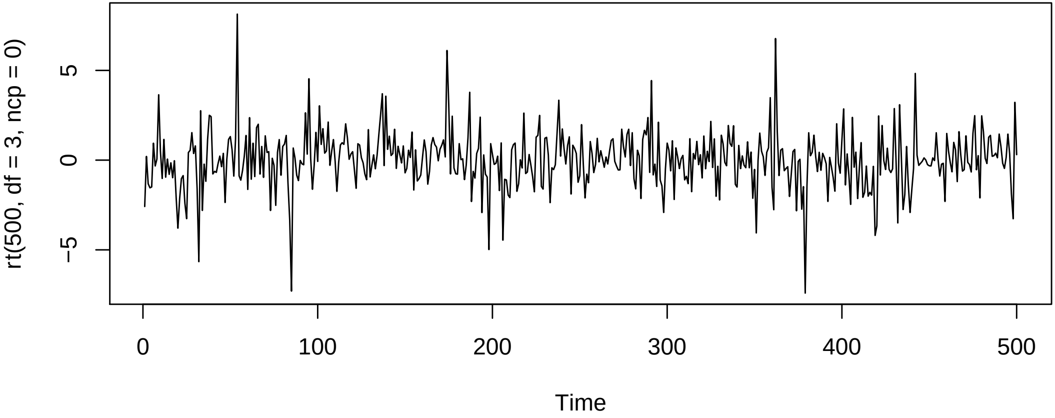
Random variables $X_1, \ldots, X_n$ will be denoted as

- White noise: if they have mean zero, variance σ^2, and are uncorrelated
- IID noise: if they are white noise AND are independent and identically distributed (IID).
- Gaussian [white] noise: if they are IID noise AND are normally distributed, $X_i \sim N(0, \sigma^2)$

Gaussian Noise



IID Noise (T distribution)



But wait!

- Didn't we say that this class is about things that are correlated over time?
- White noise has no varying structure over time, so for most time series data it's not a good model
- BUT it is the basis for many time series models
- So how do we check to see if white noise is an appropriate model?

Tools we'll need

- Autocovariance (Definition 1.2):

$$\begin{aligned} \gamma_x(s,t) &= cov(X_s, X_t) \\ &= E\left[(X_s - E[X_s])(X_t - E[X_t])\right] \end{aligned}$$

- [mental aside: let $s > t$ and $h = s - t$. h is the number of "lags"]
- Sample autocovariance (Definition 1.14):

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

Tools we'll need

- Autocorrelation function "ACF" (Definition 1.3):

$$\rho(s,t) = \frac{\gamma_x(s,t)}{\sqrt{\gamma_x(s,s)\gamma_x(t,t)}} = \frac{cov(X_s, X_t)}{\sqrt{var(X_s)var(X_t)}}$$

- Sample autocorrelation (Definition 1.15):

$$r_h = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

A note:

- The book uses $\hat{\rho}$. I'll accept either r or $\hat{\rho}$ for sample correlations.

Properties of White noise

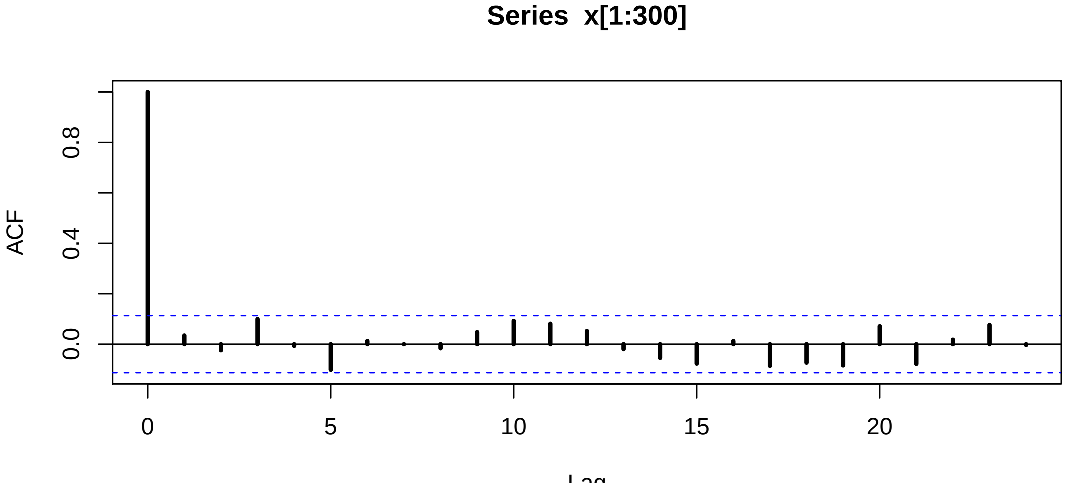
- $E(X_t) = 0$
- $var(X_t) = \sigma^2$ (constant)
- $\rho(s, t) = 0$ for all $s \neq t$
- $\rho(t, t) = 1$ (by obvious)
- How do we check if white noise is a reasonable model for a time series?

Evaluating White noise

- How do we check if white noise is a reasonable model for a time series?
- Is the average effectively 0?
- Is the variance constant?
- $r_k \approx \rho(k) = 0$ for all $k \neq 0$?

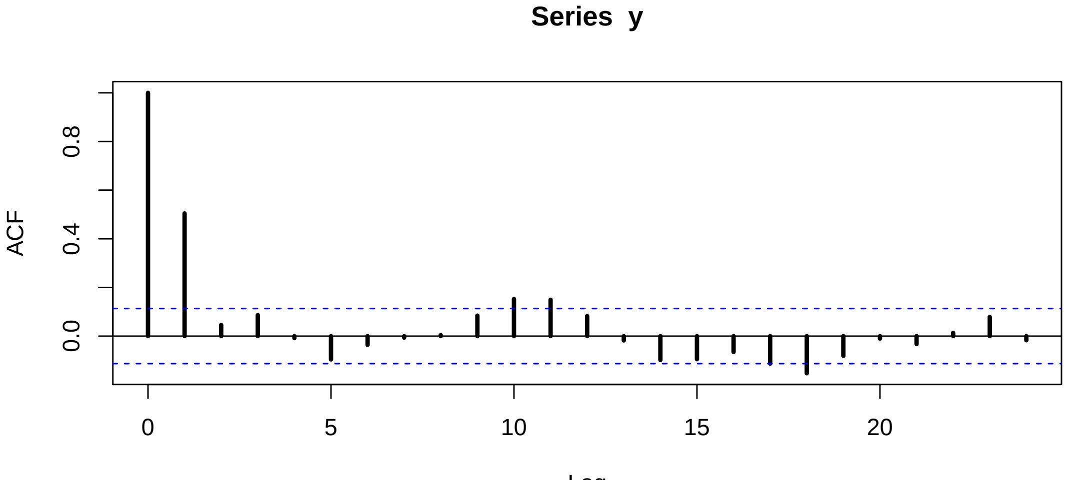
ACF plot

```
x = rnorm(301)
acf(x[1:300],lwd=3)
```



ACF plot - simple Moving Average

```
y = .5*(x[1:300] + x[2:301])
acf(y,lwd=3)
```



CI for Sample Correlations

- Wouldn't it be great if those dashed blue lines were the appropriate confidence interval?

Simplified Theorem A.7 (see Property 1.2)

- Under general conditions, if x_t is white noise, then for n large, and with arbitrary but fixed H, then the sample autocorrelations

$$r_1, \ldots, r_H \overset{iid}{\sim} N(0, 1/n)$$

- In other words

$$\sqrt{n} \begin{pmatrix} r_1 \\ \vdots \\ r_H \end{pmatrix} \to N(0, I) \quad \text{as } n \to \infty$$

- Key takeaway: $var(r_h) = 1/n$ (Equation 1.38)

Confidence Interval

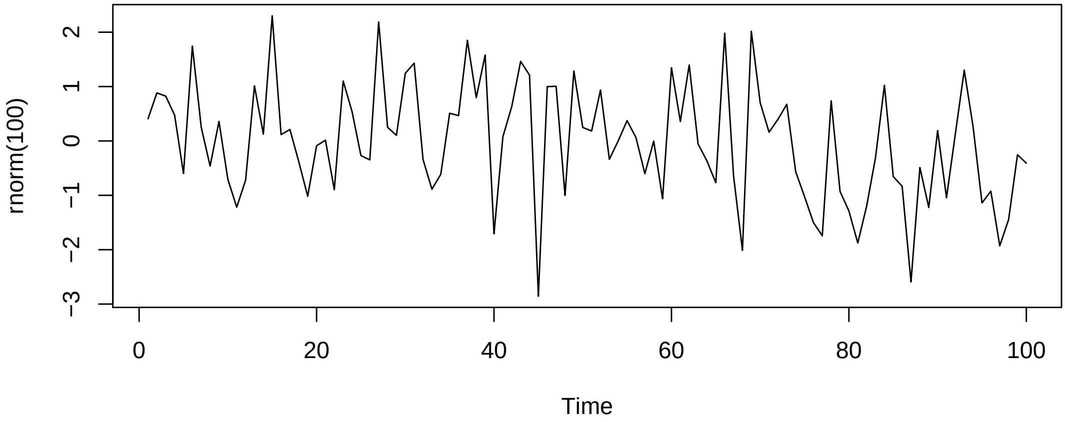
- For white noise $r_1, \ldots, r_H \overset{iid}{\sim} N(0, 1/n)$, for

$$P\left(|r_h| > 1.96 n^{-\frac{1}{2}}\right) \approx P\left(|N(0,1)| > 1.96\right) = 5\%$$

- So for $n = 100$, $1.96 n^{-\frac{1}{2}} = 1.96/\sqrt{100} = .196$

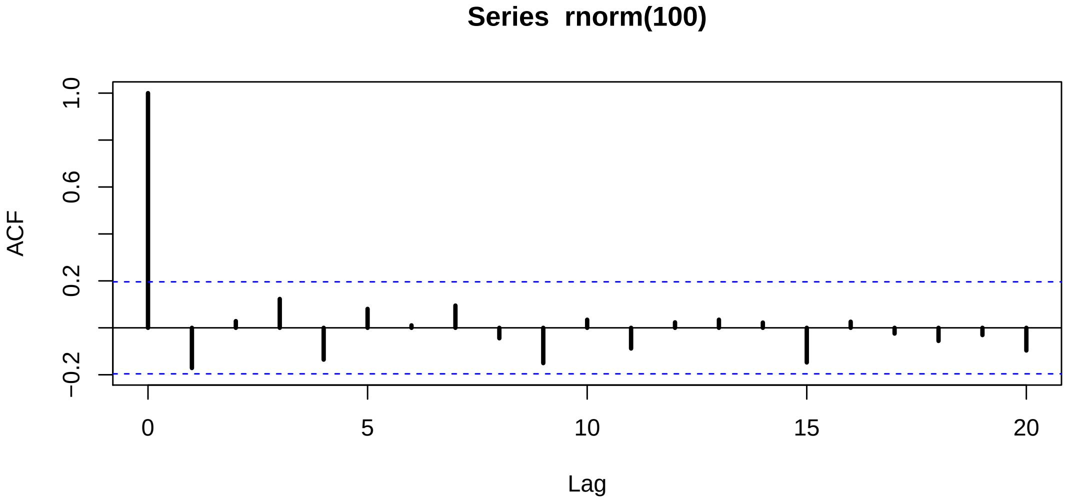
Gaussian Noise

```
plot.ts(rnorm(100))
```



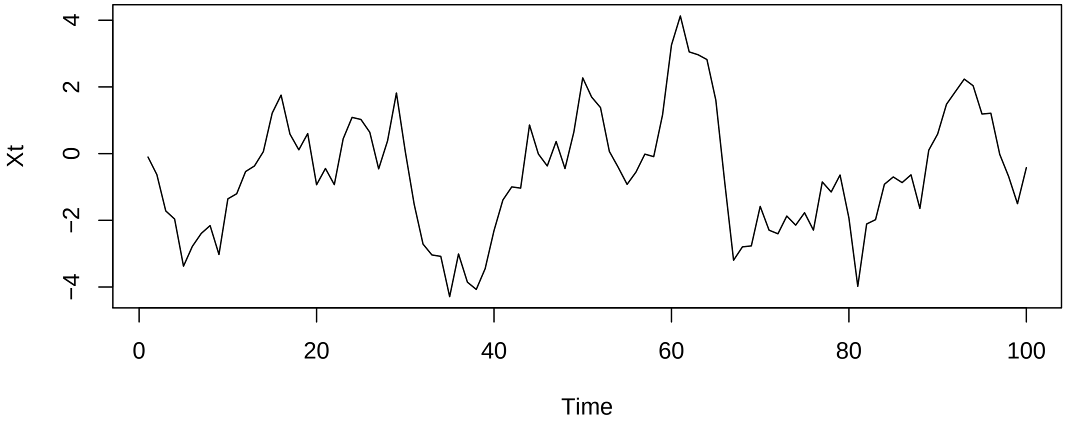
ACF plot - Dashes at .196?

```
acf(rnorm(100),lwd=3)
```



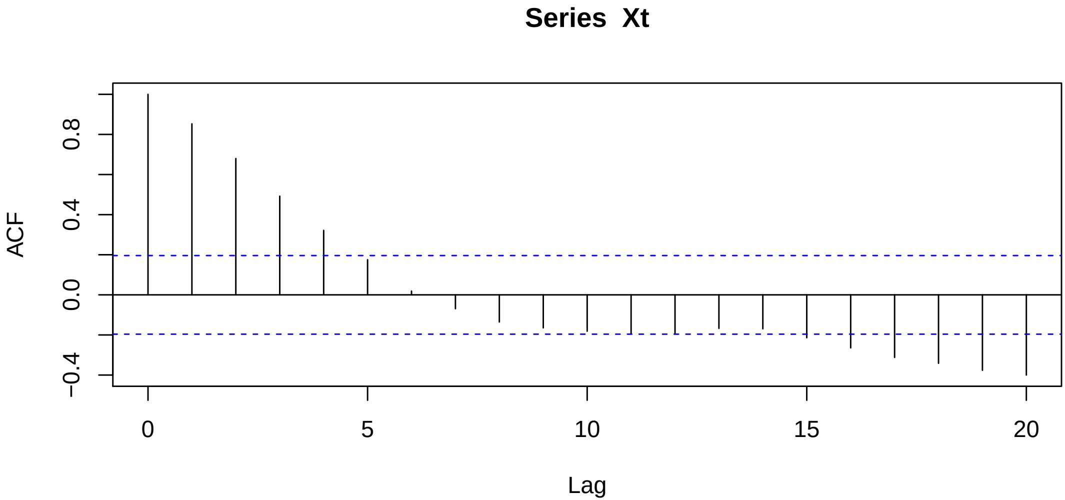
AR(1) Process (we'll learn about later)

```r
Xt = arima.sim(model=list(ar=.9),n=100)
plot.ts(Xt)
```



ACF Plot of AR(1)

```
acf(Xt)
```



Final Thought

- The ACF plot is called a “Correlogram”.
- We will use the correlogram of the ACF as a diagnostic to evaluate whether or not a *stationary* process can be assumed to be white noise.
- We will define “stationary” next time.

Textbook Alignment

- Section 1.1 presents time series examples
- Section 1.2 overviews the semester
- Section 1.3 defines autocovariance, etc.
- Section 1.5 defines sample autocovariance, etc.
- Remember that the textbook is intended as a reference, the readings are supplementary/optional (but probably helpful!), and some parts of the sections are beyond the scope of this course. All needed information will be in the lecture slides, lab material, etc.