Stat 170 Spring 2023 Notes

Contents

Chapter 1

1/23/2023

Random walks, completing the square (for MGF expectations)

Chapter 2

1/25/2023

2.1 More on Random Walks and CLT

Consider the random walk that starts at 0 and at each step increases or decreases by 1, each with probability 1/2. More formally, we have $S_n = \sum_{i=1}^n x_i$ where each x_i is independently either 0 or 1 with probability 1/2 for $n \in [1, \cdots, N]$ for some $n \in \mathbb{Z}^+$. Note that $\operatorname{Var}(x_i) = 1$, so by CLT, for sufficiently large N, S_N approaches $\mathcal{N}(0, N)$. Thus, we can make an approximately 95% confidence interval of S_N using the two standard deviation estimation: $[-2\sqrt{N}, 2\sqrt{N}]$.

> **Theorem 2.1** Central Limit Theorem
>
> If X_i are i.i.d. with $\mathbb{E}[X_i] = \mu$ and $\operatorname{Var}(X_i) = \sigma^2$, then the distribution of the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is approximately $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

There's a bunch of nice things about this theorem: it's independent of the starting distribution of X_i, we can control the spread of the mean $\operatorname{Var}(\overline{X}) = \sigma^2/n$, the resulting distribution is normal. The normal distribution itself also has many nice properties: symmetry, maximizing entropy (among distributions with same mean and variance), etc.

2.2 Generalizing the Normal Distribution to Higher Dimensions

We can define two variables X and Y which are given by a bivariate distribution:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \underbrace{\begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}}_{\text{covariance matrix } \mathbf{\Sigma}}\right).$$

Since the covariance matrix is $\mathbf{\Sigma}$, its eigenvalues are all real. It is also positive definite – its eigenvalues are all positive. Also, $\operatorname{Cov}(X, Y) \leqslant \sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}$, and

$$\operatorname{Corr}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}.$$

Note that correlation is unitless and is between -1 and 1 inclusive. The analog of the standard normal distribution in two dimensions is

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

where the correlation $\rho \in [-1, 1]$. Note that since the variances are both 1, correlation and covariance are equal here. Note that $\rho = 0$ implies that X and Y are independent. This is not trivial – correlation of 0 does not always imply independence. Furthermore, if we generate X and Y using this bivariate distribution, then the eigenvectors of $\mathbf{\Sigma}$ are the axes of symmetry of the plot of the joint distribution of X and Y, and this is the key idea behind PCA. One other nice property is that if

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right),$$

then

$$\mathbf{\Sigma}^{1/2} \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Sigma}\right).$$

Chapter 3

1/30/2023: Introduction to Market Making

Brian Yates – Trader at Clear Street Markets

3.1 Market Making

Market Making Model: Check out paper by Avellaneda and Stoikov: "High-frequency trading in a limit order book."

3.2 ETF Market making

Just do sumproducts.

Chapter 4

2/1/2023

4.1 Linear Regression Review

Many packages by default do not include an intercept. Unless there's a compelling reason why you shouldn't include one, you should add it manually if necessary. Furthermore, no causalities can be drawn from a regression; it's a purely correlational thing. The generally important quantities to look at in the output of the regression:

- Betas: association between the predictor and response, holding all the other predictors constant
- p-values: the significance of the difference of the betas from 0
- R^2; the percent variance (unexplained by an intercept/mean-only model) explained by the model

A simple regression example is in CAPM, which regresses $r_i = \alpha + \beta Z_i + \varepsilon_i$, where r_i is the return of some stock on day i, and Z_i is the return of "the market" (generally some index or ETF such as SPY) on day i.

One common measure for risk-adjusted returns is the Sharpe ratio, calculated as $S_a = \mathbb{E}[R - R_F]/\sigma_R$, where R is the return of the asset, R_F is the returns of a risk-free asset, and σ_R is the standard deviation of the asset excess return.

Chapter 5

2/6/2023

Portfolio:

$$\mathbf{P}_t = \begin{pmatrix} P_{t,1} \\ P_{t,2} \\ \vdots \\ P_{t,q} \end{pmatrix}$$

and weights $\boldsymbol{\alpha}^\top = (\alpha_{t,1}, \alpha_{t,2}, \cdots, \alpha_{t,q})$. The initial value is

$$\Pi_t = \boldsymbol{\alpha}_t^\top \mathbf{P}_t = \sum \alpha_{t,j} P_{t,j}.$$

A long position has $\alpha_{t,j} > 0$ while a short position has $\alpha_{t,j} < 0$. We define

$$w_{t,j} = \frac{\alpha_{t,j} P_{t,j}}{\Pi_t}.$$

Clearly, $\sum_{j=1}^q w_{t,j} = 1$ for all t. The total long position is

$$\text{long}_t = \sum_{j=1}^q w_{t,j} \mathbf{1}[w_{t,j} > 0]$$

and similarly the total short position is

$$\text{short}_t = -\sum_{j=1}^q w_{t,j} \mathbf{1}[w_{t,j} < 0].$$

By construction, $\text{long}_t - \text{short}_t = 1$. Assuming no change in weights, the portfolio returns are then

$$r_{t+1}^\Pi = \frac{\Pi_{t+1} - \Pi_t}{\Pi_t}.$$

We can verify that this equals

$$\sum_{j=1}^q w_{t,j} r_{t+1,j},$$

i.e. that the arithmetic returns of the portfolio is the weighted average of the returns of the individual substituents of the portfolio.

Chapter 6

2/8/2023: Bonds

6.1 Bonds

Suppose we have P_T amount of capital and invest it at the risk free rate r^F. After one year, our capital will grow to $P_T(1 + r^F)$. If we have a different risk free rate in each year, then our principal will compound to

$$P_T = P_0 \prod_{t=1}^{T} (1 + r_t^F).$$

For example, suppose that we have have an asset that will pay $Q_T = 100$ at time T. Then the price we would pay $P_0 = M_T Q_T$, where $M_T = \frac{1}{1+r^F}$ or whatever the interest rate compounds to over time. In this case, M_T is essentially the discount rate. The time T is called the time to maturity. Then the **yield** is the single interest rate that will give the same total compound interest the combined effects of the r_t^Fs:

$$(1 + \hat{i}_T)^T = \prod_{t=1}^{T} (1 + r_t^F) = \frac{1}{M_T}.$$

If we take the log of both sides, then we get

$$\begin{aligned} T \log(1 + \hat{i}_T)^T &= \sum_{t=1}^{T} \log(1 + r_t^F) \\ \hat{i}_T &= \exp\left\{ \frac{1}{T} \sum_{t=1}^{T} \log(1 + r_t^F) \right\} - 1 \\ &= \exp\left\{ \frac{1}{T} \log\left(\frac{Q_T}{P_0} \right) \right\} - 1. \end{aligned}$$

Then if we plot $\hat{i}_T$ vs. T, we get the **yield curve**.

6.2 PCA

We can transform our observed data by projecting the dataset onto the space defined by the top m PCA components, which are given by the eigenvectors of the covariance matrix of the data.

The key idea is so that each PC explains as much of the variance in the data as possible. Often times the first few principal components the vast majority of the variance in the data, and thus PCA can be a dimensionality reduction technique.

Notably with respect to bonds, we can look at the yields for the treasuries of various maturities to get a sense of what is happening in the treasuries market. The first principal component is mostly an average of the rates at all expiries, and is called the level. The second principal component is essentially the rates of the treasuries with long expiry minus the rates of those with short expiry; this check for an "inversion" of the yield curve has become an indicator of recession. This second principal component is called the slope. The third principal component roughly tells you the difference of the slope at high expiries and the slope at low expiries and is called the curvature. This third principal component explains much less of the variance than the first two PCs do. Note that this principal component analysis is totally unsupervised – we did not provide any labels for the data; the PCs simply give the vectors that explain the most variance in the data itself.

Chapter 7

2/13/2023: Martingales

7.1 Some Background

First recall Adam's Law, or the Law of Total Expectation, which states that $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$. Martingales are based on this concept in a slightly different language.

We will be thinking about conditional expectation using the framework of *information*. Define $\mathcal{F}_Y$ to be the information contained in Y. Then we can write $\mathbb{E}[X|\mathcal{F}_X] = X$; this is essentially the more formal ("mathematically correct") way to write $\mathbb{E}[X|X] = X$. Then we can rewrite Adam's Law as

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}_Y]] = \mathbb{E}[X].$$

Similar to the logic behind Adam's Law, we can also have $\mathbb{E}[\mathbb{E}[X|Y,Z]|Z] = \mathbb{E}[X|Z]$. Written in the information framework, we have

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}_Y]|\mathcal{F}_Z] = \mathbb{E}[X|\mathcal{F}_Z]$$

given that $\mathcal{F}_Y \supseteq \mathcal{F}_Z$, meaning that Y contains at least as much information as Z. For example, if $Y = X^2$, then $\mathcal{F}_X \supseteq \mathcal{F}_Y$.

Also recall Markov Chains, which have the Markov property

$$\mathbb{P}[X_{t+1}|X_t, X_{t-1}, \cdots, X_1] = \mathbb{P}[X_{t+1}|X_t].$$

7.2 Martingale Definition

Definition 7.1: Martingale

$(M_t, \mathcal{F}_t)$ is a **martingale** if

- $\mathcal{F}_t \supseteq \mathcal{F}_{t-1}$ where $F_t = \text{Inf}(M_1, \cdots, M_t)$ is the information contained in the previous M_i values; this means that we have at least as much information at time t as we do at time $t-1$
- $\mathbb{E}[M_t|\mathcal{F}_{t-1}] = M_{t-1}$ for all t
- $\mathbb{E}[|M_T|] < \infty$

Concerning the last point, variance can be infinite, but infinite expected values would break everything. Note that the key difference from a Markov Chain is that Martingales make no claims about distributions and instead only care about means.

Example 7.1

Suppose that X_i are i.i.d. with $\mathbb{E}[X_i] = 0$, and define

$$\begin{aligned} M_0 &= 0 \\ M_n &= \sum_{k=1}^{n} X_k \\ \mathcal{F}_n &= \mathrm{Inf}(X_1, \cdots, X_n). \end{aligned}$$

Then $(M_n, \mathcal{F}_n)$ is a martingale since we have

$$\mathbb{E}[M_n|\mathcal{F}_{n-1}] = \mathbb{E}[X_n + M_{n-1}|\mathcal{F}_{n-1}] = \mathbb{E}[X_n|\mathcal{F}_{n-1}] + \mathbb{E}[M_{n-1}|F_{n-1}] = \mathbb{E}[X_n] + M_{n-1} = M_{n-1}.$$

Theorem 7.1

If $m < n$, then $\mathbb{E}[M_n|\mathcal{F}_m] = M_m$.

Proof. By the generalization of Adam's law from before,

$$\begin{aligned} \mathbb{E}[M_n|\mathcal{F}_{n-2}] &= \mathbb{E}[\mathbb{E}[M_n|\mathcal{F}_{n-1}]|\mathcal{F}_{n-2}] \\ &= \mathbb{E}[M_{n-1}|\mathcal{F}_{n-2}] \\ &= M_{n-2}. \end{aligned}$$

Then it's pretty clear we can formalize this iterative argument using induction. □

Corollary 7.1

$\mathbb{E}[M_n] = \mathbb{E}[M_m]$.

Proof. Take unconditional expectations on both sides of the previous theorem and use Adam's Law.

Consider another example of a security that pays off M_T at end time T. Then at times $t < T$, our expected end payoff is $P_t = \mathbb{E}[M_T|\mathcal{F}_t]$. We claim that $(P_t, \mathcal{F}_t)$ is a Martingale. The proof is by definition and generalized Adam's law:

$$P_t = \mathbb{E}[M_T|\mathcal{F}_t] = \mathbb{E}[\mathbb{E}[M_T|\mathcal{F}_{t+1}]|\mathcal{F}_t] = \mathbb{E}[P_{t+1}|\mathcal{F}_t].$$

Notably, being a Martingale does not necessarily just mean "unpredictable"; however, it does say that the mean is unpredictable. In particular, for some Martingale $(P_t, \mathcal{F}_t)$, then the expectation of the returns is 0:

$$\begin{aligned}
\mathbb{E}\left[\frac{P_{t+1}-P_t}{P_t}\middle|\mathcal{F}_t\right] &= \mathbb{E}\left[\frac{P_{t+1}}{P_t}\middle|\mathcal{F}_t\right]-1 \\
&= \frac{1}{P_t}\mathbb{E}[P_{t+1}|\mathcal{F}_t]-1 \\
&= \frac{1}{P_t}\cdot P_t - 1 \\
&= 0.
\end{aligned}$$

Definition 7.2: Martingale Difference

Let $(Y_t, \mathcal{F}_t)$ be a Martingale. Then $\varepsilon_t = Y_t - Y_{t-1}$ is a **Martingale Difference**, sometimes denoted as $\varepsilon_t = \Delta Y_t$. This is essentially the discrete first derivative of the sequence.

We can also rewrite this definition to get

$$\begin{aligned}
Y_{t+1} &= Y_t + \varepsilon_{t+1} \\
&= Y_{t-1} + \varepsilon_t + \varepsilon_{t+1} \\
&= \vdots
\end{aligned}$$

The core properties of a Martingale and its differences are:

1. $\mathbb{E}[Y_{t+s}|\mathcal{F}_t] = Y_t \qquad \forall s \geqslant 0$
2. $\mathbb{E}[\varepsilon_{t+s}|\mathcal{F}_t] = 0 \qquad \forall s \geqslant 0$
3. $\mathrm{Cor}(\varepsilon_{t+i}, \varepsilon_{t+j}|\mathcal{F}_t) = 0 \qquad i, j \geqslant 0, i \neq j$
4. $\mathrm{Var}(Y_{t+s} - Y_t|\mathcal{F}_t) = \sum_{j=1}^{s} \mathrm{Var}(\varepsilon_{t+j}|\mathcal{F}_t)$.

Note that conditions 2 and 3 are useful when doing a regression (probably some sort of autoregressive model predicting $Y_{t+1} = \beta Y_t + \varepsilon_{t+1}$. All four of these properties are also true unconditionally, for example:

1. $\mathbb{E}[Y_{t+s}] = \mathbb{E}[Y_t]$
2. $\mathbb{E}[\varepsilon_{t+s}] = 0$
3. $\mathrm{Cor}(\varepsilon_{t+i}, \varepsilon_{t+j}) = 0$.

Using property 3, we can see

$$\begin{aligned}
\mathrm{Var}(Y_{t+s} - Y_t|\mathcal{F}_t) &= \sum_{j=1}^{s}\sum_{i=1}^{s} \mathrm{Cov}(\varepsilon_{t+j}, \varepsilon_{t+i}|\mathcal{F}_t) \\
&= \sum_{i=1}^{s} \mathrm{Var}(\varepsilon_{t+j}|\mathcal{F}_t) + 2\sum_{i\neq j} \mathrm{Cov}(\varepsilon_{t+i}, \varepsilon_{t+j}|\mathcal{F}_t) \\
&= \sum_{j=1}^{s} \mathrm{Var}(\varepsilon_{t+j}|\mathcal{F}_t),
\end{aligned}$$

which gives property 4.

Chapter 8

2/15/2023: Martingales Continued

8.1 Computations

We will first prove the third property of Martingales from last class:

Claim 8.1

For $i \neq j$, we have $\mathrm{Cov}(\varepsilon_{t+j}, \varepsilon_{t+i}|\mathcal{F}_t) = 0$.

Proof: Utilizing the definition of covariance, we can write

$$\mathrm{Cov}(\varepsilon_{t+i}, \varepsilon_{t+j}|\mathcal{F}_t) = \mathbb{E}[\varepsilon_{t+i}\varepsilon_{t+j}|\mathcal{F}_t] - \mathbb{E}[\varepsilon_{t+i}|\mathcal{F}_t]\mathbb{E}[\varepsilon_{t+j}|\mathcal{F}_t].$$

By property 2 from last class, we have $\mathbb{E}[\varepsilon_{t+i}|\mathcal{F}_t]\mathbb{E}[\varepsilon_{t+j}|\mathcal{F}_t] = 0$. Then without loss of generality assume that $i < j$. By Generalized Adam's Law, we then get

$$\mathbb{E}[\varepsilon_{t+i}\varepsilon_{t+j}|\mathcal{F}_t] = \mathbb{E}[\mathbb{E}[\varepsilon_{t+i}\varepsilon_{t+j}|\mathcal{F}_{t+i}]\mathcal{F}_t] = \mathbb{E}[\varepsilon_{t+i}\mathbb{E}[\varepsilon_{t+j}|\mathcal{F}_{t+i}]|\mathcal{F}_t].$$

Again by property 2, we get $\mathbb{E}[\varepsilon_{t+j}|\mathcal{F}_{t+i}] = 0$, which means that our original covariance is 0. $\square$

Next, suppose we have a self-financing portfolio of stocks over time with prices $\{\mathbf{P}_t\}_{t=1}^T$ and histories $\{\mathcal{F}_t\}_{t=1}^T$, and suppose at time t we have weights $\boldsymbol{\alpha}_t$ in each of the stocks. Because this portfolio is self-financing, we do not want any cash to flow into or out of the portfolio and, assuming that we can instantaneously reweight our portfolio, this means that $\boldsymbol{\alpha}_{t-1}^\top \mathbf{P}_t = \boldsymbol{\alpha}_t^\top \mathbf{P}$.

Let $\Pi_t = \boldsymbol{\alpha}_t^\top \mathbf{P}_t$ be the total value of the portfolio at time t. Then the profit gained between days t and $t+1$ is

$$\begin{aligned}
\Pi_{t+1} - \Pi_t &= \boldsymbol{\alpha}_t^\top \mathbf{P}_{t+1} - \boldsymbol{\alpha}_{t-1}^\top \mathbf{P}_t \\
&= \boldsymbol{\alpha}_t^\top \mathbf{P}_{t+1} - \boldsymbol{\alpha}_t^\top \mathbf{P}_t \\
&= \boldsymbol{\alpha}_t^\top (P_{t+1} - P_t).
\end{aligned}$$

This is relatively obvious, it just says your total profit is the weighted average of the profits from your portfolio.

Claim 8.2

If $(\mathbf{P}_t, \mathcal{F}_t)$ is a Martingale, then $(\Pi_t, \mathcal{F}_t)$ is also a Martingale.

Proof: This follows fairly directly from above:

$$\mathbb{E}[\Pi_{t+1} - \Pi_t | \mathcal{F}_t] = \mathbb{E}[\boldsymbol{\alpha}_t^\top (\mathbf{P}_{t+1} - \mathbf{P}_t) | \mathcal{F}_t] = \boldsymbol{\alpha}_t^\top \mathbb{E}[\mathbf{P}_{t+1} - \mathbf{P}_t | \mathcal{F}_t].$$

□

This shows us that we can construct new Martingales from old ones; in particular, in this example we are creating the new Martingale $(\Pi_t, \mathcal{F}_t)$ from the Martingale differences of the original Martingale $(\mathbf{P}_t, \mathcal{F}_t)$. More generally, if we have Martingale $(M_t, \mathcal{F}_t)$ and random variables $\{A_t\}_{t=1}^n$ with $\mathbb{E}[A_t | \mathcal{F}_t] = A_t$ (i.e. $\mathcal{F}_t$ encodes all possible information about A_t? I think), then if we define

$$Z_n = \sum_{t=1}^{n} A_t (M_t - M_{k-1}),$$

then $(Z_t, \mathcal{F}_t)$ is a Martingale. If we have continuous time periods, then this becomes the integral $\int A_t \mathrm{d}M_t$, which is the basis of stochastic calculus.

8.2 Doob's Optional Stopping Theorem

Definition 8.1: Martingale Stopping Times

The **stopping time** τ of martingale $(M_n, \mathcal{F}_n)$ is a random variable whose value is interpreted as the time at which a given stochastic process exhibits a specific behavior of interest. In particular, we always need to be able to answer whether or not time τ has already occurred by time t given information $\mathcal{F}_t$.

Example 8.1 (Stopping Times and Not Stopping Time)

Here are a few examples to emphasize what defines a stopping time and examples of random variables that do not satisfy this condition.

- If τ is the first time that our portfolio is down \$5, then it is a stopping time.
- If τ is the time at which a stock reaches its daily minimum, then it is not a stopping time, because we can never know if it happened or not without looking into the future.

Theorem 8.1 Doob's Optional Stopping Theorem

If we have stopping time τ with $\mathbb{E}[\tau] < \infty$ and some bound B such that $|M_n - M_{n-1}| \leqslant B$, then $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$.

To show how nice this theorem is, we'll look at a famous example.

Example 8.2 (Gambler's Ruin)

Let $X_i \sim \text{Bern}(1/2)$ be a sequence of random variables and let $M_n = \sum_{i=1}^n X_i$ so that M_n when paired with the relevant information is a Martingale. Let τ be the first time we hit either A or $-B$; this is a stopping time since we can always verify whether this has happened and since the bounding conditions holds. Then by Doob, $\mathbb{E}[M_\tau] = 0$, which tells us that the probability of M_τ being A (i.e. that the sequence hits A first) is $B/(A+B)$.

Note that if our stopping condition was only the first time we hit A (and not bounding the other side), then Doob's Theorem would not hold because $\mathbb{E}[\tau]$ would not be finite.

Suppose we have some drunk monkeys (Chris) with typewriters jamming away at the keys. What's the expected time it will take to type ABRACADABRA? The answer is $26^{11} + 26^4 + 26$, and this problem can be solved relatively easily using Martingales.

For a solution idea: suppose a gambler comes along and bets on the monkey's letters. The gambler bets \$1, and if the monkey types the letter A, then they win \$26, and if they lose they get nothing. If they win, then they will bet their \$26 of winnings on the next letter and get the same odds, etc. If the gambler loses any of their bets, they leave because they have no money. Notably, every second a new gambler comes along and starts betting on the first letter. Before the stopping time, the house gets \$1 per time period, since each gambler puts in \$1 in the beginning and contributes no cash inflows or outflows unless they win the whole thing. Thus, by the stopping time τ (when the monkey has typed the whole word), the house will have $\$\tau$ at which it will have to pay out $\$26^{11}$ to the winner. The house will have also paid $\$26^4$ to the person on the fourth letter (since ABRA is at the beginning and the end), and similarly will have to pay \$26 to the person who has won one bet. Since this is a fair game, Doob's or whatever which gives the expected value.