MATHEMATICAL FINANCE IN PRACTICE

MOHSIN JAVED

1. Introduction

Mathematical finance, as applied to the derivatives market in particular.

2. Books, Articles and Topics We Should Know

- Black Scholes Model (Read Hull or Shreve)
- Change of measure
- Change of numeraire
- Forward measure
- Read Chapter 2 of Brigo/Mercurio
- For Quanto/Compo Read paper by Derman on Foreign indices
- How is a PDE equivalent to an expectation? (Feynman-Kac and all that)
- Forward and backward Kolmograv equations
- Local volatility Model (Dupire equation, in particular)
- How the denominator and numerator of the local vol formula imply arbitrage conditions.
- Stochastic volatility model.
- conditional variance
- *Price is the cost of hedging*
- Understand the following two:
  - Analogy 1: Volatility – Theta – Gamma
  - Analogy 2: Correlation – Theta – Cross Gamma
- Understand the financial implications of the following and how much of a difference they drive from the simple Black Scholes model.
  - Funding Spreads
  - Dividends
  - Interest Rates
- Funding Valuation Adjustment (FVA), read paper by Piterbarg

Date: September 29, 2017.

3. Risk, Return and Sharpe Ratio

Excess return, (i.e. actual return minus the risk free return) is proportional to risk (i.e. volatility).

Let r be the risk free return, μ and μ' be the actual returns of two portfolios with volatilites σ and σ'. It can be shown that the Sharpe ratio of each portfolio is the same [1]. That is,

$$\frac{\mu' - r}{\sigma'} = \frac{\mu - r}{\sigma} := \lambda \tag{3.1}$$

4. Black Scholes Equation with Replication

At time t, a portfolio Π is formed by selling a call option for price C, which is delta hedged by buying Δ of the underling stock, at a cost of ΔS:

$$\Pi = C - \Delta S.$$

At time $t + dt$, we have by Ito's formula:

$$\begin{aligned}
d\Pi &= dC - \Delta dS \\
r\Pi dt &= \frac{\partial C}{\partial t}dt + \frac{\partial C}{\partial S}dS + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}dS^2 - \Delta dS \\
r(C - \Delta S)dt &= \frac{\partial C}{\partial t}dt + \left(\frac{\partial C}{\partial S} - \Delta\right)dS + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2}dt
\end{aligned}$$

Now let $\Delta = \frac{\partial C}{\partial S}$, and we get the Black Scholes PDE by cancelling out dt on both sides:

$$r\left(C - S\frac{\partial C}{\partial S}\right) = \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} \tag{4.1}$$

5. Black Scholes Equation in terms of Greeks

An easy way to intuitively think of the Black Scholes equation is as follows: roughly speaking, the PnL of an option is the sum of θ and Γ, note that θ is negative and Γ is positive. This PnL should be matched by the growth at the risk free rate of the initial hedged portfolio and hence we get,

$$\theta + \frac{1}{2}\sigma^2 S^2 \Gamma = r(C - S\Delta). \tag{5.1}$$

Writing the same equation with the partial derivatives we get the more well known form of the Black Scholes equation:

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} = r(C - S\frac{\partial C}{\partial S}) \tag{5.2}$$

Now suppose that you have delta hedged the option. This means the right hand side of Equation 5.1 is 0, i.e.,

$$\theta + \frac{1}{2}\sigma_I^2 S^2 \Gamma = 0, \tag{5.3}$$

wher σ_I is the implied volatility of the underlying asset. If we own the option and the realized volatility turns out to be σ_R, then our P&L in a short time period Δt

is

$$\text{P\&L} = \frac{1}{2} S^2 \Gamma \left(\sigma_R^2 - \sigma_I^2 \right). \tag{5.4}$$

$S^2\Gamma = S^2 \frac{\partial^2 C}{\partial S^2}$ is known as the *Dollar Gamma* and has the same unit as the price of the option C, hence the name Dollar Gamma.

The net P&L is given by the equation:

$$\text{Net P\&L} = \frac{1}{2} \int_0^T S^2 \Gamma \left(\sigma_R(t)^2 - \sigma_I^2 \right) \, dt. \tag{5.5}$$

6. Price and Value

According to Derman

> Price is simply what you have to pay to acquire a security; value is what it is worth. The price is fair when it is equal to the value.

Motivation for Risk Neutral Pricing

Suppose there is a horse race with two horses. People place bets of total amounts x and y on the two horses, respectively. The bookmaker has total deposits amounting to $x + y$.

Based on the bets placed, the bookmaker implies and quotes the odds as x/y for the two horses. (This is artificial, as the odds must be known in advance for the people to bet the money).

The total amount of money the bookmaker has to pay if the first horse wins is $x + (y/x)x = x + y$, i.e., return the original amount x and the odds times the original amount.

Similarly, the total amount of money the bookmaker has to pay if the second horse wins is $y + (x/y)y = x + y$.

In both cases the bookmaker has to pay $x + y$, which is exactly what he had as a deposit. The bookmaker has no interest in the outcome of the race!

Risk Neutral Probability

The Risk neutral probability of a certain event, where the event is described by a financial contract, can be thought of as the *market price probability*, i.e. the probability inferred from the price that the market is willing to pay for that contract.

Let B be a binary contract, which pays \$1 at time T if an event E occurs and nothing if E does not occur, then the risk neutral probability of E is:

$$P(E) = \frac{\text{Price(Contract paying \$1 at time } T \text{ if } E \text{ occurs)}}{\text{Price(Contract paying \$1 dollar at time } T \text{ no matter what)}}.$$

Risk Neutral Price of a Stock. Let us assume that the price $S(t)$ of a stock follows geometric Brownian motion. The stochastic differential equation (SDE) followed by the stock price is given by

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW_t,$$

where $W_t \sim N(0, t)$.

Using Ito's formula we can solve the above SDE and show that for $u > t$,

$$S(u) = S(t) \exp\left((\mu - \frac{1}{2}\sigma^2)(u - t) + \sigma W_{u-t} \right).$$

We can write the log-normal random variable S as

$$S = e^X, \tag{6.1}$$

where $X(u)$ is normally distributed:

$$X(u) \sim \mathcal{N}\left(\ln S(t) + \left(\mu - \frac{1}{2}\sigma^2 \right)(u - t), \sigma^2(u - t) \right). \tag{6.2}$$

In the Black-Scholes world, we obtain the result $\mu = r$. However, a lot of people find this result puzzling. How does μ get completely eliminated in the final formula for derivative pricing? We try to explain this by applying the basic principle of rsisk neutral pricing on the most basic if all securities, the stock itself.

Let $\mathbb{Q}$ be the risk neutral measure, then by the fundamental theorem of asset pricing, the stock price at time t is the expected value of the stock at any future time $T > t$, discounted back to time t:

$$S(t) = E^{\mathbb{Q}}[S(T)]e^{-r(T-t)}.$$

Using 6.1, we can write the above as

$$\begin{aligned}
S(t) &= E^{\mathbb{Q}}[e^{X(T)}]e^{-r(T-t)} \\
S(t) &= e^{\ln S(t) + \left(\mu - \frac{1}{2}\sigma^2\right)(T-t) + \frac{\sigma^2(T-t)}{2}} e^{-r(T-t)} \\
S(t) &= S(t)e^{(\mu - r)(T-t)} \\
1 &= e^{(\mu - r)(T-t)},
\end{aligned}$$

and since $T - t \neq 0$, the last equation implies,

$$\mu = r.$$

The insight is that the special form of the SDE and its solution coupled with the no arbitrage asset pricing theorem forces the remarkable equality of μ and r.

Note that to help with the computation of $E^{\mathbb{Q}}[e^X]$, we can make use of the moment generating function ϕ_X of a normal random variable X with mean m and variance v^2:

$$\phi_X(s) = E[e^{sX}] = e^{ms + \frac{v^2 s^2}{2}}.$$

Evaluating ϕ_X at $s = 1$, we get the special result,

$$\phi_X(1) = E[e^X] = e^{m + \frac{v^2}{2}}.$$

In particular, the above identity expresses the mean value of the log-normal random variable $S = e^X$ in terms of the mean and variance of its underlying normal variable X.

Price of a Binary Option. What is the price of a European binary option which pays \$1 if $S(T) > K$ and nothing otherwise?

Let $B(t)$ be the price of the binary option at time $t < T$. We have,

$$\begin{aligned} B(t) &= E^{\mathbb{Q}}[I\,(S(T) \geq K)]e^{-r(T-t)} \\ &= E^{\mathbb{Q}}[I\left(e^{X(T)} \geq K\right)]e^{-r(T-t)} \\ &= E^{\mathbb{Q}}[I\,(X(T) \geq \ln K)]e^{-r(T-t)} \\ &= \frac{e^{-r(T-t)}}{\sqrt{2\pi\sigma^2(T-t)}} \int_{\ln K}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2(T-t)}}\,dx \\ &= \frac{e^{-r(T-t)}}{\sqrt{2\pi}} \int_{-\frac{\ln\frac{S(t)}{K}+(r-\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}}^{\infty} e^{\frac{-z^2}{2}}\,dz \\ &= e^{-r(T-t)}\mathcal{N}(d_2), \end{aligned}$$

where

$$d_2 = \frac{\ln\frac{S(t)}{K} + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}.$$

What is the risk neutral probability that the stock S at time $T > t$ is greater than or equal to the strike price K? $\mathcal{N}(d_2)$.

$$\text{Digital Call:} \qquad P^{\mathbb{Q}}(S(T) \geq K) = \mathcal{N}(d_2).$$

Similarly,

$$\text{Digital Put:} \qquad P^{\mathbb{Q}}(S(T) \leq K) = 1 - \mathcal{N}(d_2) = \mathcal{N}(-d_2).$$

$$\text{Digital Call} + \text{Digital Put} = \text{Riskless Bond} \tag{6.3}$$

Price of a European options

The price of a European call option is given by the equation:

$$C(t) = S(t)\mathcal{N}(d_1) - e^{-r(T-t)}K\mathcal{N}(d_2), \tag{6.4}$$

where $\mathcal{N}$ is the standard normal cumulative distribution function, while

$$d_1 = \frac{\ln(S(t)/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \qquad d_2 = \frac{\ln((S(t)/K) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}. \tag{6.5}$$

Note that as $t \to T$, $C_t \to S_T - K$.

Also useful is the relationship

$$d_2 = d_1 - \sigma\sqrt{T-t}. \tag{6.6}$$

A more intuitive way to think about d_1 and d_2 is by rewriting them as

$$d_1 = \frac{\ln\left(\frac{S(t)e^{r(T-t)}}{K}\right) + \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}}, \qquad d_2 = \frac{\ln\left(\frac{S(t)e^{r(T-t)}}{K}\right) - \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}}. \tag{6.7}$$

The price of a European put option is given by:

$$P(t) = -S(t)\mathcal{N}(-d_1) + e^{-r(T-t)}K\mathcal{N}(-d_2), \tag{6.8}$$

The put–call parity

For European options, the put–call parity is given by the equation

$$S(t) + P(t) - C(t) = Ke^{-r(T-t)}. \tag{6.9}$$

Another way of remembering the put–call parity is by the phrase: *long call and short put is the same as a forward.*

$$C(t) - P(t) = S(t) - Ke^{-r(T-t)}. \tag{6.10}$$

The put–call parity only holds for European options.

6.1. **Price of a forward contract.** A forward contract is an over the counter (OTC) agreement to buy a stock S at time T and price K. The value F of this forward contract as a function of time $t \leq T$ is:

$$F(t) = S(t) - e^{-r(T-t)}K. \tag{6.11}$$

Note that

$$F(t) = 0 \iff K = S(t)e^{r(T-t)}.$$

The economic interpretation of the last equation is simple: it should be free to enter into an at the money forward contract. All other prices will lead to arbitrage.

Why American calls have the same price as European calls?

An American call option of a stock which pays no dividend has the same price as that of the European option. Let the strike price of the call be K and its maturity be T. The optimal strategy for a holder of an American call is to exercise it when the value of the option is the same as its intrinsic value.

At time T, the payout of the call plus a bond that pays K at time T is

$$(S_T - K)^+ + K = \max\{S_T, K\} \geq S_T$$

So at time t, if we setup a portfolio that consists of the above call and the above bond, then we have to spend

$$X_t = V_t + Ke^{-r(T-t)},$$

where V_t is the price of the call option at time t. At time T, the value of this portfolio will dominate the stock price S_T. As a result, no arbitrage implies that at time t,

$$X_t > S_t$$

Otherwise, we can short[1] one share of stock at time t, and use the proceeds to setup this portfolio; at time T, we have zero probability of losing money, and

[1]Short the over-priced asset, go long on the under-priced

have a positive probability $P(S_T < K)$ of making money. This is an arbitrage! Combining the last two equations, we have that

$$V_t > S_t - Ke^{-r(T-t)} > S_t - K.$$

Which shows that the value of the option is always strictly greater than its intrinsic value $S_t - K$, therefore the holder should not exercise this option before its maturity T.

7. VTS: Volatility Trading Strategies

7.1. Volatility, Skew, Smile and the Vol Surface.

- The volatility σ, commonly called vol, is usually expressed on an annualized basis. We can therefore assume that the unit of volatility is per square root of time. Similarly, the unit of variance (var) is per unit of time. Correspondingly, the quantities $\sigma\sqrt{T}$ and $\sigma^2 T$ are dimensionless quantities and represent the total vol and total var attained in time T.
- Vol Surface: Volatility σ as a function of strike K and maturity T, $\sigma(K, T)$.
- ATMF: At the money forward.
- Normalized Strike (NS): For a given maturity T and strike K, if the ATMF vol is σ_{ATMF} and the forward value of the underlying asset is F, then the normalized strike NS is defined by the equation,
$$NS = \frac{1}{\sigma_{ATMF}\sqrt{T}} \log\left[\frac{K}{F}\right].$$
- Vol Skew: For a given maturity T, the slope of the vol surface slice, defined by the equation,
$$Skew = -\frac{1}{\sigma_{ATMF}} \frac{d\sigma(NS)}{dNS}.$$
- Smile: For a given maturity T, the curvature of the vol slice defined by the equation,
$$Smile = -\frac{1}{\sigma_{ATMF}} \frac{d^2\sigma(NS)}{dNS^2}.$$
- Vol Term Structure: Vol as a function of time to maturity, $\sigma(T)$.
- In the Black Scholes world, $\sigma(K, T)$ is a constant.
- In the FX options market, the vol smile actually looks like a human smile :).
- In the Equity derivatives market, the vol smile is a smirk with a downward slope.
- Sticky by Strike Vol: The vol does not change as a function of the spot S. The vol still changes as a function of the strike K. For example, if at the money vol is σ_0, one sticky by strike vol model can be

$$[TODO]\sigma(S, K, T) = \sigma(K, T) - b(K - S_0) \tag{7.1}$$

7.2. **A Short Summary of the Local Vol Model.** The Local Vol model assumes one factor geometric Brownian motion for the underlying asset where the volatility is a deterministic function of spot and time. Crucially, the following are also assumed to be deterministic:

- Interest rates
- Equity funding spreads
- Dividend yield

The aim of the model is to match the non-arbitrageable input implied vol surface at all strikes and maturities. The model can be thought of as a very good interpolator of the implied volatility surface, and allows us to accurately price European styles payoffs.

The local vol surface is analogous to the forward rates. Given two zero coupon bonds, with maturities T_1 and T_2, such that $T_1 < T_2$, the forward rate r of a zero coupon starting at time T_1 and maturing at time T_2 is given by the equation

$$r_2 T_2 = r_1 T_1 + r(T_2 - T_1).$$

This forward rate r is the only rate that is consistent with our market rates r_1 and r_2.

Similarly, given vols of various maturities and strikes (the discrete vol surface formed by market quotes), the *forward vol* aka the local vol is the vol surface which is consistent with the existing discrete vol surface of market quotes.

7.3. **A Short Summary of the Risky Log-OU model.** The Risky Log-OU enhances the local vol model in that it no longer assumes a deterministic funding spread. The model simulates the funding spread of an entity as a stochastic process. The aim of this model is to capture the correlation between equity and funding spreads, giving us conditional (path-wise) risky discounting based on equity levels, making it a richer model compared to a simple local volatility model which has fixed risky discount factors.

7.4. **General Rules.**

- There are four basic *reserves*
  (1) Vega
  (2) Skew
  (3) Funding Spreads
  (4) Dividend
- *Funding Spreads*: Let $r_C(t)$ be the CSA backed short rate, i.e., the agreed overnight rate paid on collateral according to the CSA. We now consider an asset, whose price is some process denoted by $S(t)$. Let $r_R(t)$ be the repo rate of this asset, i.e., the short interest rate we can get if we use the asset as the collateral. Piterbarg calls the difference $r_R(t) - r_C(t)$ as the stock lending fee [3]. Following Piterbarg [3], we also define the short rate

for unsecured funding by $r_F(t)$. Now comes the funding spread, which is defined as [3]

$$s_F(t) = r_F(t) - r_C(t).$$

In essence funding spread of an entity represents the market view of credit default risk of that entity.

In the context of equities, *Stock lending fee* is the extra interest rate charged on top of say LIBOR (or the CSA interest rate) if one wants to borrow money to buy a stock. One typically puts the stock as the collateral for the funding, which is a risky asset and therefore the lender wants to be rewarded with a risk premium for taking this risk. This is reflected via the stock lending fee.

It is expected that

$$r_C(t) \leq r_R(t) \leq r_F(t) \tag{7.2}$$

Negative Funding Spreads: Funding spreads can be negative. This can especially happen for short dated maturities. If we have very good credit rating, a short term loan can be obtained at a rate very close to LIBOR. Suppose we now use the cash we obtained to buy a stock and then lend the stock in the repo market. Effectively, this makes our borrowing cost lower than LIBOR, i.e., we have a negative funding spread.

- One wrong argument goes like this. Call prices should go down as funding spreads increase. This is because the large funding spreads are an indication that the equity is being considered very risky. After all, as a lender, if the money you gave away has the stock as the collateral, the spread you charge is proportional to the risk you imagine for the stock. This is also the reason that Put prices go up as funding spreads increase.

  The above argument is wrong. In reality, call prices go up as funding spreads go up and put prices go down as funding spreads go up. The reason – option pricing is done via hedging – not via the market view argument presented above. If you are selling a call option, you hedge it with buying delta. The cost of funding to buy your delta increases as the funding spreads increase and as a result, the cost of the option you are selling goes up as well. Similarly, if you are selling a put option, you are going to short some delta. However, when you short a stock, you have to pay a borrowing fee which moves in opposite direction to the funding spread [check this last sentence].
- Call prices go down as dividends increase. The equity spot will be down if dividends increase which results in a lower call price. Correspondingly, put prices go up if dividends go up.

7.5. **Miscellaneous.**

- *Call Overwrite (Buy-Write)*: Go long the stock and short a slightly OTM short dated Call. This caps your profit but the proceeds of the premium contribute towards reducing the cost of stock buying and therefore enhance the yield.

7.6. **Out-performance Options.**

- An out-performance option of an asset A over another asset B pays $\max\{A-B, 0\}$ at expiration. By convention, out-performance options are always quoted as *a Call of A over B.*
- Out-performance options are usually European.
- Out-performance options are short correlation. Extreme case: suppose you have two assets, X and Y:

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\rho_{XY}\sigma_X\sigma_Y, \tag{7.3}$$

If ρ_{XY} goes up to 100% the variance on the left is minimum, i.e. the option has the least price. If ρ_{XY} goes all the way down to -100%, we the variance σ^2_{X-Y} is maximized.

7.7. **Worst-Of and Best-Of Options.** We will use W for worst-of and B for best-of, and the subscripts C and P for Calls and Puts respectively. Assume there are n underlying assets. Then, the payoff worst-of and best-of options is given by

$$W_C = \max\{\min_{i=1,2,\ldots,n}\{S_i - K\}, 0\}, \tag{7.4}$$

$$W_P = \max\{\max_{i=1,2,\ldots,n}\{K - S_i\}, 0\}, \tag{7.5}$$

$$B_C = \max\{\max_{i=1,2,\ldots,n}\{S_i - K\}, 0\}, \tag{7.6}$$

$$B_P = \max\{\min_{i=1,2,\ldots,n}\{K - S_i\}, 0\}. \tag{7.7}$$

Worst-of call options are traded much more than best-of call options because worst-of call options are very cheap.

Worst-of put options are traded much more than best-of put options despite being quite expensive. This is due to the fact that worst-of put options provide protection and are high in demand.

7.7.1. *Worst-Of Calls.*

- If we lower the correlation between the assets, the price of worst-of call decreases. Intuitive reason, the stocks are more dispersed and the chance that one of them is below the strike is relatively higher. Indeed, worst-of call will be worthless even if a single asset is below the strike. Hence, worst-of call have a lower price if the correlation is low.
- If the correlation goes up, the price of worst-of call also goes up. Extreme case, if we have two assets which are 100% correlated, the price of the worst-of call on the two assets is the same as the price of cheaper of the two.
- Related to the above, a worst-of call is long correlation.
- Typically, clients buy worst-of calls as it gives them a cheaper way of getting exposure to the upside. As a result of the trade, the clients go long correlation. When a trading desk sells worst-of calls, it gets a short correlation exposure, typical position of an exotic derivative desk.

7.7.2. *Worst-Of Puts.*

- Worst-of put options are very expensive but still traded as they provide protection. The price of a worst-of put is higher than any of the puts on the underlying assets.
- If we lower the correlation between the assets, the price of worst-of put increases. Intuitive reason, the stocks are more dispersed and the chance that one of them is deep down below the strike is relatively higher. Indeed, worst-of put will pay according to the worst asset which has gone down the most. Hence, worst-of puts have a higher price if the correlation is low.
- If the correlation goes up, the price of worst-of puts goes down (but only relatively, they are already quite expensive). Extreme case, if we have two assets which are 100% correlated, the price of the worst-of puts on the two assets is the same as the price of more expensive of the two.
- Related to the above, worst-of puts are short correlation.
- When we sell worst-of puts, we are long correlation, i.e. we want the correlation to go up so that we have to pay less on the short put positions.
- A trading desk usually buys worst-of-puts (clients like to sell these since they look expensive when compared to basket puts, or individual puts). The desk therefore ends up short correlation (classic exotic position). Due to correlation skew the Bid price can be quite high compared to other buyers in the market. Therefore, there is sometimes pressure to bid lower to be competitive since other market participants might not be charging as much correlation skew.

7.7.3. *Worst-Of Options and the Correlation Skew.* Suppose we are long a WO put, which gives us a short correlation exposure. If the spot goes down, the correlation will go up and because of the correlation skew, the correlation is likely to go up high enough such that the increase in value of our put due to spot going down will be offset to a great extent by the correlation going up.

On the other hand, if the spot goes up the correlation goes down, but because of the correlation skew, it goes down very little. Our put goes cheaper due to the spot going up and the offsetting benefit we get from the correlation going down is small due to the correlation skew.

In either case, the correlation skew is hurting us. Hence, if we are buying a WO put, we should charge extra for the correlation skew.

7.8. **Convexity.** $9 = 5^2 - 4^2 > 4^2 - 3^2 = 7$ and this is called convexity. If you are long a convex payoff, you make more money on the up move than you would lose on a down move of equal magnitude.

Mathematically, for $x > 0$, $(x + \delta x)^2 - (x - \delta x)^2 \simeq 4x\delta x > 0$.

Derivatives on Foreign Indices

Let X be the value of 1 JPY in GBP as a function of time. We also suppose that X_0 is the value of X at $t = 0$ and X_T is the value of X at expiration time T. Let

K be the strike price in JPY and let S be the price of a stock in JPY, again as a function of time.

The underlying asset in all of the following cases is a stock with the value S, denominated in JPY.

(1) *Foreign-market derivatives*: Buy JPY denominated derivative by converting your GBP into JPY today. At expiry, you will convert the JPY payoff back into GBP at the prevailing exchange rate. You are exposed to both the asset and the FX.

$$\max\{0, X_T(S_T - K)\}. \tag{7.8}$$

(2) *Compo*: The derivative in this case derives its value by converting the value of the Japanese asset from JPY to GBP at the prevailing exchange rate. You are still exposed to the asset and the FX but this exposure is different from the previous case.

$$\max\{0, X_T S_T - K_{\$})\}. \tag{7.9}$$

(3) *Quanto*: The payoff of the derivative in this case is the JPY value of the derivative at the time of expiry converted to GBP at a guaranteed exchange rate decided at the start of the contract. You are still exposed to the performance of the underlying asset.

$$\max\{0, X_0(S_T - K)\}. \tag{7.10}$$

8. Credit

8.1. Glossary.

- Spread: Spread is the constant (absolute) shift to the zero-coupon discount curve in all scenarios that is required to ensure that the model value of the bond (average value over all scenarios) equals the observed market price [?].
- Rally: When bond prices go up and yields go down.
- Long CDS: When you have sold protection via a credit default swap. In this case you are long the credit of the company you have sold the protection on. You are short a put option on the company's credit rating.
- Long Credit: Same as a long CDS position. You are short put on a company's credit rating.
- DTS: Duration times spread. This is a measure of credit risk.
- Bond Price–Yield–Duration–Convexity: Let y be the yield of a bond with price B. Then

$$\frac{1}{B} dB = -D dy + \frac{1}{2} C dy^2, \tag{8.1}$$

where the duration D and the convexity C are defined as

$$D = -\frac{1}{B}\frac{dB}{dy},$$
$$C = \frac{1}{B}\frac{d^2B}{dy^2}.$$

- CDS Spread: Credit default swaps are priced in terems of a spread usually expressed in basis points. A CDS quote of 4.55 means that the CDS is pricing at a spread of 4.55 bp, or .0455% i.e., to buy \$10000 of protection, you have to pay \$4.55 per year.

8.2. **Hazard Rate.** The hazard rate (also called default intensity) is defined as a number h such that the probability of default in a certain time interval $[t, t + \Delta t]$, *conditional* on no earlier default is given by $h\Delta t$. In general, the hazard rate will be different when different time intervals are considered, and in those cases $h(t)$ is defined as a time dependant hazard rate.

Let X be the random variable representing the time (in years) when the company defaults. For a simple hazard rate model, we assume that h is the average hazard rate and the the time to default is an exponentially distributed random variable X:

$$P(X \leq t) = 1 - \exp(-ht),$$

Note that h is playing the same role that the more familiar λ plays in a classically defined exponential random variable X, with pdf:

$$f_X(x) = \lambda \exp(-\lambda x),$$

and CDF:

$$F_X(x) = 1 - \exp(-\lambda x)$$

The interpretation of h is given by the equation,

$$\begin{aligned}
P(t \leq X \leq t + \Delta t | X > t) &= \frac{P(t \leq X \leq t + \Delta t \text{ and } X > t)}{P(X > t)} \\
&= \frac{P(t \leq X \leq t + \Delta t)}{P(X > t)} \\
&= \frac{(1 - \exp(-h(t + \Delta t)) - (1 - \exp(-ht))}{\exp(-ht)} \\
&= 1 - \exp(-h\Delta t) \\
&\approx h\Delta t.
\end{aligned}$$

In other words, for this model, the hazard rate (or the default intensity) determines that the probability of default in a small time interval Δt is approximately $h\Delta t$

Note that

(1) The default probabilities backed out of bond prices or credit default swap spreads are risk-neutral default probabilities.

(2) The default probabilities backed out of historical data are real-world default probabilities.

8.3. **Put-Call Parity for the Merton Model.** The fundamental balance sheet equation of a firm is given by

$$V(t) = E(t) + D(t),$$

that is, at any time t, the firm's assets $V(t)$ are a sum of the firm's equity $E(t)$ and the firm's liabilities (or Debt) $D(t)$.

The Merton model shows that the firm's equity is a call option on the firm's assets, expiring at time T, having a strike F:

$$E(t) := (V(T) - F)^{+} = \max\{V(T) - F, 0\}.$$

The strike F is implied by the initial value of the debt $D(0)$ and a risky interest rate k_D:

$$F := D(T) = D(0)\exp(k_D T) = D(t)\exp(k_D(T-t))$$

.

The put-call parity implies:

$$V(t) + P(t) - C(t) = Fe^{-r(T-t)},$$

where r is the risk-free rate. Since $C(t) = E(t)$, using the balance sheet equation gives $V(t) - C(t) = D(t)$ and substituting this in the put-call parity above gives us:

$$D(t) + P(t) = Fe^{-r(T-t)}.$$

Now use the value $F = D(t)e^{k_D(T-t)}$:

$$\begin{aligned} D(t) + P(t) &= D(t)e^{k_D(T-t)}e^{-r(T-t)} \\ P(t) &= D(t)\left(e^{(k_D - r)(T-t)} - 1\right) \\ \frac{1}{T-t}\ln\left(1 + \frac{P(t)}{D(t)}\right) &= k_D - r \end{aligned}$$

which implies that the *credit spread* is

$$k_D - r = \frac{1}{T-t}\ln\left(1 + \frac{P(t)}{D(t)}\right).$$

9. Probability and Stochastic Calculus

Definition 1 (Probability Space)**.** *. A triple* $(\Omega, \mathcal{F}, \mathcal{P})$. Ω *is a set,* $\mathcal{F}$ *is a sigma-algebra on* Ω *and* $\mathcal{P} : \mathcal{F} \to [0, 1]$*, such that*

(1) $P(\Omega) = 1$

(2) *For any countable union of mutually disjoint sets in* $\mathcal{F}$*, the function* $\mathcal{P}$ *is additive.*

Definition 2 (Probability Measure)**.** *The function* $\mathcal{P}$ *above is called a probability measure.*

9.1. **Modes of convergence of a sequence of random variable.** Let X_n be a sequence of random variables defined on a probability space.

We say that

(1) $X_n \to X$ almost surely if for sufficiently large n $P(|X_n(\omega) - X(\omega)| < \epsilon) = 1$.

(2) Mean square convergence, which is stronger that convergence in probability. [TODO]

(3) $X_n \to X$ in probability if

(4) $X_n \to X$ in distribution if for all $z \in \mathbb{R}$, $F_n(z) \to F(z)$, where F_n is the distribution function of X_n and F is distribution function of X.

For a Venn diagram of modes of convergence, see Figure 7-5 of [2].

9.2. **Transformation of Random Variables.** Let X and Y be two random variables with joint density function f_{XY}. Given,

$$\begin{aligned} U &= g(X, Y), \\ V &= h(X, Y), \end{aligned}$$

what is the joint density of f_{UV}?

We assume that we can *invert* the transformation and express X and Y as,

$$\begin{aligned} X &= \phi(U, V), \\ Y &= \psi(U, V). \end{aligned}$$

Secondly,

$$dA = dxdy = |J(x, y)|\, dudv,$$

where,

$$J(x, y) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix}.$$

Then,

$$\begin{aligned} P((U, V) \in A) &= \int_A f_{UV}(u, v)\, du\, dv \\ &= P((X, Y) \in B) \\ &= \int_B f_{XY}(x, y)\, dx\, dy \\ &= \int_B f_{XY}(\phi(u, v), \psi(u, v))\, dx\, dy \\ &= \int_B f_{XY}(\phi(u, v), \psi(u, v))\, |J(x, y)|\, du\, dv. \end{aligned}$$

Therefore,

$$f_{UV}(u, v) = f_{XY}(\phi(u, v), \psi(u, v))\, |J(x, y)|\,.$$

Brownian Motion. Brownian motion is a stochastic process $B(t)$, such that

(1) $B(0) = 0$.

(2) For any $t_1 < t_2$, $B(t_2) - B(t_1)$ is independent of $B(t_1)$ and normal with zero mean and variance $t_2 - t_1$.

(3) $B(t)$ is continuous almost surely.

Stopping Time. Intuitively speaking, a stopping time is the time at which a stochastic process satisfies a certain rule. However, for a stopping time, the rule must be defined in a way that at any given time it may be tested by looking at only the past and present values of the stochastic process. For example: buy Microsoft as soon as the stock price goes below \$100 is a valid rule for a stopping time. On the other hand: sell Microsoft first thing in the morning if the closing price that day is less than \$50 per share is not a stopping time, since in the morning we can not tell what closing price will be attained at the end of the day.

Let $\tau_m := \inf\{t : B(t) = m\}$, i.e., the random variable τ_m is the time when Brownian motion B hits the level m for the first time.

Notice that

$$\tau_m = \inf\{t : B(t) \geq m\} \tag{9.1}$$

What is the probability $P(\tau_m \leq T)$?

$$P(\tau_m \leq T) = P(\tau_m \leq T \cap B(T) \geq m) + P(\tau_m \leq T \cap B(T) < m).$$

Now,

$$P(\tau_m \leq T \cap B(T) \geq m) = P(B(T) \geq m), \tag{9.2}$$

and by using the reflection principle,

$$P(\tau_m \leq T \cap B(T) < m) = P(B(T) \geq m), \tag{9.3}$$

Therefore,

$$P(\tau_m \leq T) = 2P(B(T) \geq m). \tag{9.4}$$

$$P(\tau_m \leq t) = 2P(B(t) \geq m) \tag{9.5}$$

$$= 2\frac{1}{\sqrt{2\pi}} \int_m^\infty e^{-\frac{x^2}{2t}}\, dx \tag{9.6}$$

$$\tag{9.7}$$

Therefore, the density function of τ_m is obtained by differentiating the last expression with respect to t:

$$f_{\tau_m}(t) = \tag{9.8}$$

9.3. **Copula.** A copula is a distribution function $C : \mathbb{R}^n \to [0,1]$ such that all the marginals are standard uniform random variables.

By definition,

$$C(1, 1, \ldots, 1, u_j, 1, \ldots, 1) = u_j.$$

The probability density function c associated with C is given by the usual formula,

$$c = \frac{\partial^n C}{\partial u_n \partial u_{n-1} \ldots \partial u_1}$$

9.4. **Probability Integral Transform.** Let $X \sim F$. Then $Y = F(X)$ is uniform.

Proof.

$$\begin{aligned} P(Y \leq y) &= P(F(X) \leq y) \\ &= P(X \leq F^{-1}(y)) \\ &= F(F^{-1}(y)) \\ &= y. \end{aligned}$$

□

9.5. **Probability Quantile Transform.** Let U be uniform. Then, $X = F^{-1}(U) \sim F$

Proof.

$$\begin{aligned} P(X \leq x) &= P(F^{-1}(U) \leq x) \\ &= P(U \leq F(x)) \\ &= F(x). \end{aligned}$$

□

Log-normal random variable. If the log of a random variable X is Normal, then X is called a log-normal random variable.

$$\ln X = \sigma Z + \mu, \tag{9.9}$$

where Z is a standard normal: $Z \sim \mathcal{N}(0,1)$.

$$X = e^{\sigma Z + \mu}. \tag{9.10}$$

The expected value of X is given by the formula

$$E[X] = e^{\sigma^2/2 + \mu}. \tag{9.11}$$

9.6. **Ito's Formula.**

9.7. **Ito's Formula for Brownian Motion.** Let $W(t)$ be a Brownian motion. Let $f : \mathbb{R} \to \mathbb{R}$ be a function with continuous second derivative, then

$$df(W(t)) = f'(W(t))dW(t) + \frac{1}{2}f''(W(t)), \qquad \text{(Differential Form)} \tag{9.12}$$

or

$$f(W(t)) - f(W(0)) = \int_0^t f'(W(s))dW(s) + \frac{1}{2}\int_0^t f''(W(s))dW(s), \qquad \text{(Integral Form).} \tag{9.13}$$

9.8. **Ito's Formula for a useful stochastic process.** Let $W(t)$ be a Brownian motion. Let us define the stochastic process $X(t)$ via the stochastic differential equation

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dW(t). \tag{9.14}$$

We also assume that $f(t, x)$ is a function with continuous derivatives f_t, f_x and f_{xx}. Then, the differential form of Ito's formula expresses the differential increment of $f(X(t), t)$ by

$$df(t, X(t)) = f_t(t, X(t))dt + f_x(t, X(t))dX(t) + \frac{1}{2}f_{xx}(t, X(t))dX(t)^2, \tag{9.15}$$

while the integral form is

$$f(T, X(T)) - f(0, X(0)) = \int_0^T f_t(t, X(t))dX(t) + \int_0^T f_x(t, X(t))dX(t) + \frac{1}{2}\int_0^t f_{xx}(t, X(t))dX(t). \tag{9.16}$$

For Geometric Brownian Motion (GBM), the SDE can be obtained by setting $\mu(X(t), t) = \mu_0 X(t)$ and $\sigma(X(t), t) = \sigma_0 X(t)$ in Equation (9.14). This is usually written as (abusing notation a little bit by identifying μ and σ as constants,

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t), \tag{9.17}$$

where $X(t)$ is the price of the stock.

The solution of the above SDE is

$$X(t) = X(0)\exp\left[(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)\right] \tag{9.18}$$

$W(t)$ is a martingale. Indeed, for any $t > s$,

$$\begin{aligned} E[W(t+s)|\mathcal{F}_s] &= E[W(t+s) - W(s) + W(s)|\mathcal{F}_s] \\ &= W(s) + E[W(t+s) - W(s)|\mathcal{F}_s] \\ &= W(s). \end{aligned}$$

$e^{\sigma W(t)}$ is not a martingale. One can check this by applying the Ito's Lemma. Let $f(W(t)) = e^{\sigma W(t)}$, then

$$df(W(t)) = \sigma W(t)dW(t) + \frac{1}{2}\sigma^2 dt. \tag{9.19}$$

Because of the non-zero drift term, f is not a martingale.

$e^{\left(\sigma W(t)-\frac{1}{2}\sigma^2 t\right)}$ is a martingale. Again, we can check this using Ito's Lemma. Let $f(W(t)) = e^{\sigma W(t)-\frac{1}{2}\sigma^2 t}$, then

$$df(W(t)) = -\frac{1}{2}\sigma^2 W(t)dt + \sigma W(t)dW(t) + \frac{1}{2}\sigma^2 W(t)dt = \sigma W(t)dW(t). \tag{9.20}$$

There is no drift term, hence f is a martingale.

The moment generating function of $X \sim N(\mu, \sigma)$ is

$$\phi(s) = E[e^{sX}] = e^{s\mu}e^{\frac{s^2\sigma^2}{2}} \tag{9.21}$$

Using the above result, we see that

$$\begin{aligned}
E[e^{\sigma W(t+s)}|\mathcal{F}_s] &= E[e^{\sigma(W(t+s)-W(s)+W(s))}|\mathcal{F}_s] \\
&= e^{\sigma W(s)}E[e^{\sigma(W(t+s)-W(s))}|\mathcal{F}_s] \\
&= e^{\sigma W(s)}e^{\frac{1}{2}\sigma^2 t^2} \\
&= e^{\sigma W(s)+\frac{1}{2}\sigma^2 t^2} \\
&\neq e^{\sigma W(s)}.
\end{aligned}$$

Therefore, $e^{\sigma W(s)}$ is not a martingale. Note that we have used the fact that

$$W(t+s) - W(s) \sim N(0, t), \tag{9.22}$$

and also the formula for the moment generating function for $N(0, t)$.

Change of measure. Consider X, a standard normal random variable defined on the probability space $(\mathbb{R}, \mathcal{F}, \mu_X)$, where $\mathcal{F}$ is the σ-algebra of open sets in $\mathbb{R}$ and for a Borel set B,

$$\mu_X(B) := \frac{1}{\sqrt{2\pi}}\int_B e^{-\frac{x^2}{2}}\,dx.$$

The expectation of a Borel measurable function f under the probability measure μ_X is given as

$$E_0[f(X)] = \int_{-\infty}^{\infty} f(X)d\mu_X,$$

where the subscript 0 in the expectation is to emphasize that the mean of X is 0. What if we now want to change the mean of the distribution to a new number say m. How is the expectation of f under the old measure related to the expectation under the new measure? The answer is easy in this case:

$$\begin{aligned}
E_0[f(X)] &= \int_{-\infty}^{\infty} f(X)d\mu_X \\
&= \int_{-\infty}^{\infty} f(X)e^{-X^2/2}\,dX \\
&= \int_{-\infty}^{\infty} f(X)e^{-mX+\frac{m^2}{2}}e^{-\frac{(X-m)^2}{2}}\,dX \\
&\equiv E_m\left[f(X)e^{-mX+\frac{m^2}{2}}\right].
\end{aligned}$$

The collection of theorems that tell us how to make drift disappear is commonly called Girsanov theory [4, Ch. 13].

10. Time Series Analysis

We assume that ϵ_t is white noise:

$$\begin{aligned} \epsilon_t &\sim \text{ i.i.d } N(0, \sigma^2), \\ \mathrm{E}(\epsilon_t) &= 0, \\ \gamma_\epsilon(t+h, t) &= \mathrm{Cov}(\epsilon_{t+h}, \epsilon_t) = 0, \qquad h \neq 0. \end{aligned}$$

11. Portfolio Theory

Suppose we have a universe of n assets, described by a n-random vector $A = [a_1, a_2, \ldots, a_n]^T$.

We assume that their mean vector is given by

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix},$$

while the covariance matrix is given by

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \sigma_{ij} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}.$$

Note that $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$.

Minimum Variance Portfolio. Our portfolio P is a linear combination of assets:

$$P = w \cdot a,$$

where $w = [w_1, w_2, \ldots, w_n]^T$ is a weight vector such that

$$w \cdot \mathbf{1} = \sum_{i=1}^{n} w_1 + w_2 + \cdots + w_n = 1.$$

However, we want to choose the weight vector w which minimizes the variance of our portfolio.

Recall that in general for a matrix A and a random vector X, $\mathrm{Var}(AX) = A\Sigma A^T$. Therefore,

$$\mathrm{Var}(P) = \mathrm{Var}(w^T a) = w^T \Sigma w.$$

We have the following standard minimization problem:

$$\begin{aligned} \text{minimize:} \quad & \frac{1}{2} w^T \Sigma w, \\ \text{subject to:} \quad & w \cdot \mathbf{1} = 1. \end{aligned}$$

The associated Lagrangian is given by,

$$L(w, \lambda) = \frac{1}{2} w^T \Sigma w - \lambda (w \cdot \mathbf{1} - 1).$$

Differentiating with respect to w and setting the result to zero gives,

$$\Sigma w - \lambda \mathbf{1} = 0,$$

which implies that the best weights vector w^* is given by,

$$w^* = \lambda \Sigma^{-1} \mathbf{1}.$$

Since w^* should satisfy the constraint as well,

$$\begin{aligned} w^* \cdot \mathbf{1} &= 1, \\ \lambda \mathbf{1}^T \Sigma^{-1} \mathbf{1} &= 1, \\ \lambda &= \frac{1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}, \\ w^* &= \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}. \end{aligned}$$

The minimum variance is given by,

$$\begin{aligned} \sigma^2_{min} &= w^{*T} \Sigma w, \\ &= \left(\frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \right)^T \Sigma \left(\frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \right), \\ &= \left(\frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \right)^T \frac{\mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}, \\ &= \left(\frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \right)^T \frac{(\mathbf{1}^T)^T}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}, \\ &= \left(\frac{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \right)^T \frac{1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}, \\ &= \frac{1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} = \lambda. \end{aligned}$$

Linear Regression

We assume that the output Y is a random variable and related to the input X through a regression function

$$Y = r(X).$$

Linear regression assumes that the best estimate of Y, given X is the expected value of Y where this expectation is driven by a function linear in X:

$$\mathrm{E}[Y|X = x] = X\beta + \epsilon.$$

The simplest case of linear regression is when we assume that both X and Y are one dimensional real random variables and we have a set of discrete observations,

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

such that $\mathrm{E}[\epsilon_i | X] = 0$ and $\mathrm{Var}[\epsilon_i | X] = \sigma^2$.

In matrix notation, we can write the above as,

$$y = \begin{bmatrix} \mathbf{1} & x \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \epsilon,$$

where x, y and ϵ are vectors in $\mathbb{R}^n$ and $\mathbf{1}$ is a n-vector of all ones.

$$\begin{bmatrix} \mathbf{1} & x \end{bmatrix}^T y = \begin{bmatrix} \mathbf{1} & x \end{bmatrix}^T \begin{bmatrix} \mathbf{1} & x \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \mathbf{1} & x \end{bmatrix}^T \epsilon,$$

or,

$$\begin{bmatrix} \mathbf{1}^T \\ x^T \end{bmatrix} y = \begin{bmatrix} \mathbf{1}^T \\ x^T \end{bmatrix} \begin{bmatrix} \mathbf{1} & x \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \mathbf{1}^T \\ x^T \end{bmatrix} \epsilon.$$

This gives,

$$\begin{bmatrix} \mathbf{1}^T y \\ x^T y \end{bmatrix} = \begin{bmatrix} \mathbf{1}^T \mathbf{1} & \mathbf{1}^T x \\ \mathbf{1}^T x & x^T x \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \mathbf{1}^T \epsilon \\ x^T \epsilon \end{bmatrix}.$$

Let's put hats on β_0 and β_1, as we are going to drop the terms involving ϵ and apply Cramer's rule to solve the system.

$$\begin{aligned} \hat{\beta}_0 &= \frac{\begin{vmatrix} \mathbf{1}^T y & \mathbf{1}^T x \\ x^T y & x^T x \end{vmatrix}}{\begin{vmatrix} \mathbf{1}^T \mathbf{1} & \mathbf{1}^T x \\ \mathbf{1}^T x & x^T x \end{vmatrix}} \\ &= \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - \sum x_i \sum x_i} \end{aligned}$$

$$\begin{aligned} \hat{\beta}_1 &= \frac{\begin{vmatrix} \mathbf{1}^T \mathbf{1} & \mathbf{1}^T y \\ \mathbf{1}^T x & x^T y \end{vmatrix}}{\begin{vmatrix} \mathbf{1}^T \mathbf{1} & \mathbf{1}^T x \\ \mathbf{1}^T x & x^T x \end{vmatrix}} \\ &= \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \sum x_i \sum x_i} \\ &= \frac{\frac{1}{n} \sum x_i y_i - \left(\frac{1}{n} \sum x_i\right) \left(\frac{1}{n} \sum y_i\right)}{\frac{1}{n} \sum x_i^2 - \left(\frac{1}{n} \sum x_i\right) \left(\frac{1}{n} \sum x_i\right)} \\ &= \frac{\mathrm{Cov}(x, y)}{\mathrm{Cov}(x, x)} \\ &= \rho \frac{\sigma_y}{\sigma_x} \end{aligned}$$

Here is the key idea to remember, the slope of the regression line β_1 is estimated by dividing the sample covariance $\mathrm{Cov}(x, y)$ with the sample covariance $\mathrm{Cov}(x, x)$.

Notice that the explicit solution of β_0 above does not have a clear interpretation. However, given we first solve for $beta_1$, we can find $beta_0$ by,

$$\beta_0 =$$

12. Statistics

12.1. The trio: Z-χ^2-t.

12.1.1. *Z*. This is just the central limit theorem. If $X_1, X_2, \ldots, X_n$ are i.i.d., with the mean and variance of each being μ and σ^2, respectively, then,

$$\lim_{n\to\infty} \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = Z,$$

where Z is a standard normal.

12.1.2. χ^2. If $Z_1, Z_2, \ldots, Z_n$ are independent standard normals, then,

$$\chi^2(n) = \sum_i^n Z_i^2,$$

has a chi-square distribution with n degrees of freedom.

12.1.3. *t*. Let Z and $Z_1, Z_2, \ldots, Z_n$, be independent standard normals, then

$$t = \frac{Z}{\sqrt{\frac{\sum_{i=1}^n Z_i^2}{n}}} = \frac{Z}{\sqrt{\frac{\chi(n)^2}{n}}},$$

has a t-distribution.

12.2. Hypothesis testing.
Type-I error is the error of rejecting a null hypothesis when it is in fact true. The probability of this error is usually denoted by α. A typical value of α used in practice is .05.

12.3. The Law of Total Expectation.

$$\mathrm{E}(X) = \mathrm{E}_y(\mathrm{E}_x(X|Y)))$$

12.4. The Law of Total Variance.

$$\mathrm{Var}(X) = \mathrm{Var}(\mathrm{E}(X|Z)) + \mathrm{E}(\mathrm{Var}(X|Z))$$

Proof. By definition,

$$\mathrm{Var}(X|Z) = \mathrm{E}(X^2|Z) - \mathrm{E}(X|Z)^2.$$

Taking expectation w.r.t Z, we get,

$$\begin{aligned}
\mathrm{E}(\mathrm{Var}(X|Z)) &= \mathrm{E}(\mathrm{E}(X^2|Z)) - \mathrm{E}(\mathrm{E}(X|Z)^2) \\
&= \mathrm{E}(X^2) - \mathrm{E}(\mathrm{E}(X|Z)^2) \\
&= \mathrm{E}(X^2) - \mathrm{E}(X)^2 - \big(\mathrm{E}(\mathrm{E}(X|Z)^2) - \mathrm{E}(X)^2\big) \\
&= \mathrm{Var}(X) - \big(\mathrm{E}(\mathrm{E}(X|Z)^2) - \mathrm{E}(\mathrm{E}(X|Z))^2\big) \\
&= \mathrm{Var}(X) - \mathrm{Var}(\mathrm{E}(X|Z)).
\end{aligned}$$

□

13. Matrices

13.1. **Rank Theorem.**

Theorem 1. *Given a matrix A, the row rank of A is the same as the column rank of A.*

Proof. Convert A to its row reduced echelon from R. This is done through elementary matrix operations on A,

$$\begin{aligned}
E_n \dots E_2 E_1 A = UA &= R, \\
A &= U^{-1} R.
\end{aligned}$$

Therefore, rows of R can be obtained as a linear combination of the rows of A and vice-versa, hence,

$$\mathrm{Row}(A) = \mathrm{Row}(R).$$

The columns of A and R are related as follows,

$$\begin{aligned}
UA &= U \begin{bmatrix} C_1 & C_2 & \dots & C_n \end{bmatrix} \\
&= \begin{bmatrix} UC_1 & UC_2 & \dots & UC_n \end{bmatrix} \\
&= \begin{bmatrix} C_1^R & C_2^R & \dots & C_n^R \end{bmatrix} \\
&= R.
\end{aligned}$$

If the row rank of A is r, then the matrix R has r rows with leading 1s. The key insight is that the columns of R with leading 1s form a basis of $\mathrm{Col}(R)$.

We claim that the columns of A for which R has a leading 1, form a basis of $\mathrm{Col}(A)$. That is, the subset of columns of A,

$$\phi = \{C_{j_1}, C_{j_2}, \dots, C_{j_r}\},$$

where $j_1, j_2, \dots, j_r$ are the indices of columns where R has a leading 1, is a basis of $\mathrm{Col}(A)$.

To prove the linear independence of ϕ, let

$$a_1 C_{j_1} + a_2 C_{j_2} + \dots + a_r C_{j_r} = 0,$$

then

$$a_1 UC_{j_1} + a_2 UC_{j_2} + \ldots + a_r UC_{j_r} = 0,$$
$$a_1 C^R_{j_1} + a_2 C^R_{j_2} + \ldots + a_r C^R_{j_r} = 0.$$

Since the columns $\{C^R_{j_1}, C^R_{j_2}, \ldots, C^R_{j_r}\}$ of R are a basis of $\text{Col}(R)$. the last equation implies that

$$a_1 = a_2 = \ldots = a_r = 0.$$

Therefore, ϕ is a linearly independent set.

To prove that ϕ spans $\text{Col}(A)$, consider $b \in \text{Col}(A)$. Then, there exists x such that,

$$b = Ax = U^{-1}Rx.$$

Since $Rx \in \text{Col}(R)$ and $\{C^R_{j_1}, C^R_{j_2}, \ldots, C^R_{j_r}\}$ is a basis of $\text{Col}(R)$, we can write the last equation as,

$$\begin{aligned} b &= U^{-1}(t_{j_1} C^R_{j_1} + t_{j_2} C^R_{j_2} \ldots + t_{j_r} C^R_{j_r}) \\ &= t_{j_1} C_{j_1} + t_{j_2} C_{j_2} \ldots + t_{j_r} C_{j_r}, \end{aligned}$$

and hence $b \in \text{Span}(\phi)$.

This proves that ϕ is a basis of $\text{Col}(A)$ and therefore,

$$\dim(\text{Col}(A)) = r = \dim(\text{Row}(A))$$
$$\text{Row rank}(A) = r = \text{Col Rank}(A).$$

□

Lemma 1. *A and A^T have the same rank.*

13.2. **Left Right Inverses.** Let A be a square matrix. If A has a left inverse, prove that A is invertible and that the left inverse is indeed the inverse.

Proof. Let B_L be the left inverse of A. Then, by definition,

$$B_L A = I.$$

WRONG PROOF: Assumes that for each y, we can express it as Ax. What if y is not in the column space of A?

As a result, for any y, the equation, $Ax = y$ has a solution $x = B_L y$. This solution is unique, since,

$$Ax_1 = y = Ax_2$$
$$x_1 = B_L y = x_2.$$

Choosing, $y = e_i$, the above implies the existence of a unique matrix B_R where,

$$\begin{aligned} B_R &= \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \\ &= \begin{bmatrix} B_L e_1 & B_L e_2 & \ldots & B_L e_n \end{bmatrix}, \end{aligned}$$

such that,

$$AB_R = I.$$

This proves the existence of a right inverse B_R.

To show $B_L = B_R$, compute the product $B_L A B_R$ in the following two ways,

$$B_L(AB_R) = B_L I = B_L,$$
$$(B_L A)B_R = I B_R = B_R,$$

which proves $B_L = B_R$. □

Therefore, A is invertible since $B := B_L = B_R$ is the unique inverse of A.

13.3. **LU Factorization.** Given a square $n \times n$ matrix A, we can find the LU decomposition of A such that

$$PA = LU, \tag{13.1}$$

where P is a permutation matrix, L is a lower triangular and U is an upper triangular.

The algorithm to compute the matrices L and U, given A is essentially the row reduced echelon form computation of A. We want to apply a sequence of matrices L_i such that A gets transformed into U.

$$L_1 L_2 \ldots L_{n-1} PA = U, \tag{13.2}$$

where each L_i is a lower triangular matrix which introduces zeros in the i^{th} column below the diagonal.

13.4. **Orthogonally Diagonalizable Matrices.** An orthogonally diagonalizable matrix is called a *Normal* matrix.

$$A \text{ is normal } \Leftrightarrow A = Q\Lambda Q^*,$$

where Q is orthogonal and Λ is diagonal.

A normal matrix commutes with its complex conjugate, and this property also characterizes a normal matrix.

$$A \text{ is normal } \Leftrightarrow A^*A = AA^*.$$

Orthogonally Diagonalizable Matrices	
Complex	Real
Unitary	Orthogonal
Hermitian	Symmetric
Skew-Hermitian	Skew-Symmetric

13.5. **Cholseky Decomposition in Finance.** The fundamental use of Cholesky decomposition in finance is to generate a vector of correlated samples from a vector of uncorrelated samples, where the correlations are specified by the correlation matrix. This correlation matrix, say A is by construction real, symmetric, and positive semi-definite.

13.5.1. *Example:* Suppose we have two independent zero mean random variables Z_1 and Z_2 with variances σ_1 and σ_2, respectively, and we want to construct two correlated random variables X_1 and X_2, again with variances σ_1 and σ_2 respectively, but now we want the correlation between X_1 and X_2 to be ρ.

The correlation matrix in this simple 2×2 case is

$$A = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{13.3}$$

The Cholesky decomposition is very simple (think RREF):

$$A = \begin{bmatrix} 1 & 0 \\ \rho & 1 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ 0 & 1-\rho^2 \end{bmatrix}. \tag{13.4}$$

We can write this in the form where diagonal of L and U both are replaced by the squar root of their products, and we get

$$A = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ 0 & \sqrt{1-\rho^2} \end{bmatrix}. \tag{13.5}$$

It can be shown, that applying L on the vector z gives us the required correlated vector x:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{13.6}$$

.

Perturbation of the Identity Matrix.

$$\left(I + uv^T\right)^{-1} = I - \frac{uv^T}{1 - u^T v}.$$

Projection Matrices. A square matrix P such that $P^2 = P$ is called a projection matrix.

A typical example of a projection matrix is the rank-1 matrix,

$$\frac{aa^T}{a^T a},$$

where a is a vector in $\mathbb{R}^n$. This matrix projects any vector onto the vector a.

The outer product of u and v, defined as uv^T, is a rank-1 matrix. The eigenvectors of uv^T are all vectors orthogonal to v (with eigenvalue 0) and the vector u (with eigenvalue $v^T u$).

14. Using Chebyshev Methods

(1) write a basic core

(2) must have a root finder

Where can it be used?

(1) Corr Bump: It's a 1d root finding problem

Questions

(1) We are given an array of integers starting from 0 and ending at 100. We start from 0 and start flipping a fair coin. If a heads come down, we advance by 1 and if a tail comes down we advance by 2. For example, if head comes down, we move to index 1, otherwise, we skip 1 and move to index 2. We continue flipping the coin until we reach hundred or beyond.

The question is, which index in the array has the largest probability of being visited?

(2) Let X, Y and Z be random variables. If the correlations between them satisfy the following relationship,

$$\rho_{XY} = \rho = \rho_{YZ}$$

what can you say about the correlation ρ_{XZ}?

Thinking in terms of vectors, let α be the angle between X and Y, which must be the same as the angle between Y and Z. Then

$$\cos\alpha = \rho,$$

[TODO]: This can be done through looking at the positive semi-definite property of the correlation matrix as well.

[TODO]: Draw a picture of X, Y and Z. If X and Y are orthogonal, i.e., $\rho = 0$, then Y must be normal to the plane formed by Z and X. In this case, ρ_{XZ} can by any number between -1 and 1.

If X, Y are parallel, i.e. $\rho = 1$, and Y, Z are parallel, then X, Z must be parallel as well. (Draw three vectors pointing in the same direction). Therefore, $\rho_{XZ} = 1$.

If X, Y are anti-parallel, i.e. $\rho = -1$, and Y, Z are anti-parallel, then X, Z must be parallel and therefore, $\rho_{XZ} = 1$.

For the general case, if X and Y make an angle α between each other and Y and Z make the same angle α between each other, then the angle between X and Z can be between 0 and 2α.[TODO: Draw cones]

Therefore, the range of correlation between X and Z must satisfy

$$\begin{aligned} &\cos(2\alpha) \leq \rho_{XZ} \leq \cos(0)\,, \\ \implies &\cos\left(2\cos^{-1}(\rho)\right) \leq \rho_{XZ} \leq 1. \end{aligned}$$

(3) Suppose you are trading and you define your *hit ratio* as the number of trades you have won divided by the total number of trades you participated in. You checked your hit ratio in the morning and it was less than 0.9. You checked later in the day and it was larger than 0.9. Is it the case that at some point during the day, the hit ratio was exactly 0.9?

Yes. Let k_1 be the number of trades lost out of a total of n_1 trades. Then, in the morning you had

$$\frac{n_1 - k_1}{n_1} < 0.9,$$

which implies

$$n_1 < 10k_1,$$

and similarly, later in the day you had

$$\frac{n_2 - k_2}{n_2} > 0.9,$$

$$n_2 > 10k_2.$$

where $n_2 \geq n_1$ and $k_2 \geq k_1$.

The variables $n_1 \leq n_2$ being discrete must be equal to $10k$ at some intermediate time at which point your hit ratio must be 0.9.

There is nothing special about 0.9, the same can be concluded for any hit ratio r as long as r has a specific form. Let

$$\frac{n-k}{n} = r,$$

which implies,

$$n = \frac{1}{1-r}k.$$

Therefore, as long as $\frac{1}{1-r}$ is an integer, the problem structure remains the same.

Let $\frac{1}{1-r} = j$ where $j \in \mathbb{Z}^+$, then the permissible values of r are:

$$r = 1 - \frac{1}{j}, \qquad j \in \mathbb{Z}^+.$$

(4) Given a data series, $x_1, x_2, \ldots, x_n$, for a large n, how do you test if the data is normal? Given another data series, $y_1, y_2, \ldots, y_n$, how do you compare the normality of the two data series with each other, i.e. which data series is *more* normal than the other?

(5) You have a jury of three judges. Two of the judges make the correct decision, each with probability p, while the third judge tosses a fair coin to make the decision. All judges think/toss coins independently. What is the probability that the jury makes the right decision?

Let C be the event that the jury makes the correction decision. Then, conditioning on whether the judge with the coin makes the correct decision (j_3), we have

$$P(C) = P(C|j_3)P(j_3) + P(C|\neg j_3)P(\neg j_3).$$

This boils down to

$$P(C) = \frac{1}{2}\left(1 - (1-p)^2\right) + \frac{1}{2}p^2 = p.$$

Thanks to the coin tossing, a three person jury has effectively been reduced to a one person jury.

(6) What is the relationship of $\mathrm{Var}(\mathrm{E}(X \mid Y))$ with $\mathrm{E}(\mathrm{var}(X \mid Y))$?

The law of total variance:

$$\mathrm{Var}(Y) = \mathrm{E}[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(\mathrm{E}[Y \mid X]).$$

(7) Find the value of α which minimizes,

$$\sum_{i}^{n} |\alpha w_i - n_i|,$$

where w_i and n_i are positive real numbers.

Notice that $\alpha w_i - n_i$ has a zero at $\frac{n_i}{w_i}$. Define

$$a := \min\left\{\frac{n_i}{w_i} : i = 1, 2, \dots, n\right\},$$

and

$$b := \max\left\{\frac{n_i}{w_i} : i = 1, 2, \dots, n\right\}.$$

Now apply a bisection on the interval $[a, b]$.

(8) Estimate the derivative of x^x at $x = 2$.

We have:

$$\begin{aligned}\frac{d}{dx}x^x &= \frac{d}{dx}e^{x\ln x} \\ &= e^{x\ln x}\frac{d}{dx}(x\ln x) \\ &= x^x\frac{d}{dx}(\ln x + 1).\end{aligned}$$

At $x = 2$, the last expression is roughly equal to $2^2(.7 + 1) = 6.8$.

(9) What are the eigenvalues of an $n \times n$ matrix all of whose entries are a constant?

The rank of this matrix is 1, so 0 is an eigenvalue with multiplicity $n - 1$. The only non-zero eigenvalue is n and the corresponding eigenvector is a vector of all ones.

(10) Let X be a random vector in $\mathbb{R}^n$ whose $n \times n$ variance-covariance matrix is Σ. What is the variance-covariance matrix of AX where A is an $n \times n$ matrix?

$$Var(AX) = A\Sigma A^T$$

An important example of the above is when $X = Z$ i.e., a random vector of i.i.d standard normals. In that case $\Sigma = I$ and the variance-covariance matrix of AZ is AA^T.

(11) You roll a hundred sided die and note the number that appears. How many rolls you need on average to see the same number again?

The answer is 100. You roll the first time and get some number, say 12. Your probability of getting a 12 on any following roll is $1/100$. You therefore need 100 more rolls on average to get the same number again.

(12) Roll 3 standard six-sided dice together. What is the probability that the max is less than or equal to 3.

Max less than or equal to 3 is the same as each of the three dice independently showing a number less than or equal to 3. Therefore, the probability is $(1/2)(1/2)(1/2) = 1/8$. (We have used the fact that the probability of a six-sided dice showing a number less than or equal to 3 is $1/2$.)

(13) Let X be a discrete random variable which takes values in $\{1, 2, 3\}$ with equal probability. What is the standard deviation of X? If we take three independent samples X_1, X_2 and X_3, what is the standard deviation of the sum $S = X_1 + X_2 + X_3$?

By calculation, $\sigma_X = \sqrt{2/3}$ and $\sigma_S = \sqrt{2}$

(14) Let X and Y be two random variables with $\sigma_X = 2$, $\sigma_Y = 3$. If $\sigma_{X+Y} = 5$, then what is the correlation ρ_{XY}?

We use the formula:

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho_{XY}\sigma_X\sigma_Y,$$

which implies

$$2\sigma_X\sigma_Y = 2\rho_{XY}\sigma_X\sigma_Y$$

i.e.

$$\rho_{XY} = 1.$$

(15) Generic version of the above question. If $\sigma_{X+Y} = \sigma_X + \sigma_Y$, then what is the value of ρ_{XY}?

We have:

$$(\sigma_X + \sigma_Y)^2 = \sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho_{XY}\sigma_X\sigma_Y,$$

where the first equality is given and the second is by definition. This simplifies to

$$2\sigma_X\sigma_Y = \rho_{XY}2\sigma_X\sigma_Y,$$

which implies $\rho_{XY} = 1$.

Intuition: The standard deviation does not add linearly for uncorrelated variables. If correlation between two variables is zero, then the *variance* of the sum is indeed additive but we have the basic identity for $a, b > 0$:

$$\sqrt{a+b} < \sqrt{a} + \sqrt{b},$$

from which it follows that the standard deviation of the sum would be strictly less than the sum of individual standard deviations. Given that the standard deviations added up linearly in our case gives us a hint that the variables must be highly positively correlated.

(16) Approximate $\log_7 1250$

(17) Approximate $4^{3.6}$

(18) the mean and variance of a random variable X is 10. What is $E[X^2]$?

(19) You pick three uniform i.i.d samples from $[0, 1]$. What the expected value of the product of the maximum and the minimum.

(20) Consider the following game with a six-sided standard dice. If you roll $1, 2$ or 3, you roll again. If you roll 4 or 5, you get the number of dollars equal to the number of rolls you have rolled so far, including the last roll. If you roll a 6, the game ends and you get nothing. On average, how many dollars are you expected to make by playing this game?

Let's first solve an easier game: in the modified game, you can roll again if you get $1, 2$ or 3 as before but if you roll $4, 5$ or 6 on your k^{th} roll, you get k dollars. In this case, you can map the game to a fair coin, since the probability of getting a value in the set $\{1, 2, 3\}$ or in the set $\{4, 5, 6\}$ is each $1/2$. The expected number of throws to get an outcome in the set $\{4, 5, 6\}$ is therefore $1/(1/2) = 2$ and on average you make \$2 playing this game.

Now, we come back to the original question. One thing to notice is that on average, the original game must pay less than the modified game above because in the original game you either get paid nothing or you get paid the same amount that the modified game would pay. Since the good outcomes are $2/3$ times the outcomes of the modified game, the expected value of the original game is also $2/3$ times the expected value of the modified game. The final answer is $(2/3)(2) = 4/3$.

We can also solve this problem in the classical mathematical way.

Let X be the random variable that you win k dollars by rolling a 4 of 5 at the k^{th} roll. Then $P(X = k)$ is given by the expression,

$$P(X = k) = (1/2)^{k-1}(2/6),$$

i.e. for the first $(k - 1)$ rolls you got $1, 2$ or 3, each independently with probability $1/2$ and then on the k^{th} go, you rolled a 4 or a 5 with probability $2/6$.

The expected value of X is given by:

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} kP(X = k) \\ &= (1/3) \sum_{k=1}^{\infty} k(1/2)^{k-1} \end{aligned}$$

(21) What are the eigenvalues of an $n \times n$ matrix all of whose off diagonal entries are a constant and the diagonal has ones?

Let A be the matrix in question and c be the off diagonal constant. We can write A as

$$A = cW + (1-c)I, \tag{14.1}$$

where W is a matrix of all ones and I is the identity matrix.

The eigenvalue problem can now be transformed as

$$A - \lambda I = 0 \iff W - \left(\frac{\lambda + c - 1}{c}\right) I = W - \omega I = 0, \tag{14.2}$$

where $\omega = \frac{\lambda+c-1}{c}$.

Now, the eigenvalues of W are 0 and n, where n is the dimension of W. This is true since W is of rank 1 which implies zero is an eigenvalue of multiplicity $n-1$. Also, a vector of all ones is an eigenvector of W with n as the eigenvalue.

Therefore,

$$\omega = 0 \iff \frac{\lambda + c - 1}{c} = 0 \implies \lambda = 1 - c. \tag{14.3}$$

And

$$\omega = n \iff \frac{\lambda + c - 1}{c} = n \implies \lambda = 1 + (n-1)c. \tag{14.4}$$

(22) A first order Taylor approximation of a convex function always underestimates the function:

$$e^x - (1+x) > 0, \qquad x \in \mathbb{R}. \tag{14.5}$$

Can we prove this? [TODO]

(23) Let X be a standard normal variable, $a \in \mathbb{R}$ and $p = P(X \le a)$. Define a new random variable $Y = -X$. What is $P(Y \le a)$?

$$\begin{aligned} P(Y \le a) &= P(-X \le a) && (14.6)\\ &= P(X \ge -a) && (14.7)\\ P(Y \le a) &= P(X \le a) = p. && (14.8) \end{aligned}$$

In the last step, we have used the symmetry property of the cumulative distribution function of a standard normal variable.

Intuitively, the result makes sense[2]. A normal random variable is symmetric around the origin. Its distribution does not change when we reflect it across the origin.

(24) Let us consider the Indicator function $I(X > 30)$, where X is a standard normal variable. Suppose you want to naively compute the expected value of $I(X > 30)$. Note that this number is strictly greater than 0, since

$$E[I(X > 30)] = P(X > 30) = \phi(-30) > 0.$$

[2]It does not have to, some results in probability are counter intuitive to most people. Also, what some people think intuitive might not be intuitive to others. Intuition is subjective.

However, if we apply the *standard* approach of approximating the expected value via the discrete sampling

$$E[I(X > 30)] \simeq \frac{1}{N}\sum_{i=1}^{N} I(X_i > 30),$$

what is the expected value of N to make the sum greater than zero?

Answer: Convert the problem to a geometric random variable, with the success probability $p = \phi(-30)$. Therefore $E[N] = 1/\phi(-30)$.

(25) What is $\frac{1}{\sqrt{2\pi}}$ approximately?

Answer: 0.4.

Polya's Urn: An urn has $n_b = 95$ black balls and $n_w = 5$ white balls. Therefore, the ratio r_b of black balls to the total number of balls in the urn is

$$r_b = \frac{n_b}{n_b + n_w} = 0.95. \tag{14.9}$$

You start a game, where you pick a ball randomly and if the ball is black, you add some more black balls and if it's white you add some more white balls. Suppose you have done this a few million times and you ask now, what is the expected value of the ratio r_b now?

Answer .95. The ratio is a martingale with respect to the stochastic process. Let r_{bk} be the ratio at the k^{th} turn. At the $(k+1)^{st}$ turn, we have

$$\begin{aligned}
E[r_{b(k+1)}] &= \left(\frac{n_{bk}}{n_{bk}+n_{wk}}\right)\frac{n_{bk}+f(k)}{n_{bk}+n_{wk}+f(k)} + \left(\frac{n_{wk}}{n_{bk}+n_{wk}}\right)\frac{n_{bk}}{n_{bk}+n_{wk}+f(k)} && (14.10)\\
&= \left(\frac{n_{bk}}{n_{bk}+n_{wk}}\right)\frac{n_{bk}+n_{wk}+f(k)}{n_{bk}+n_{wk}+f(k)} && (14.11)\\
&= \frac{n_{bk}}{n_{bk}+n_{wk}} && (14.12)\\
&= r_{bk}. && (14.13)
\end{aligned}$$

Maxims

Frictions are of great importance in financial markets; they are in many ways the krill that feed the financial Leviathans.

The first step in either finance or mechanics is to consider models that are free of frictions.

Let s be the sentiment function of gaining wealth. An obvious question is to compare $s(x)$ with $s(-x)$. In other words, how is gaining a hundred pounds different from loosing a hundred pounds? Of course, one is making you happier and the other is making you sad but can the happiness and sadness be quantified? and if so, are they equal?

One can argue that s is linear close to the origin. But what happens as $|x|$ increases? If you win a billion you are very happy, but if you loose a billion you may very well be broke. You can always gain unlimited amount of money but you always have a limited amount of money to loose, so

$$s(x) \neq -s(-x). \tag{14.14}$$

Another way of thinking about this is the following game. Suppose your annual salary is a hundred pounds and you are invited to play a fair game in which you might win or lose a hundred pounds. Will you play that game? How about a game in which you might win or lose only one pound?

15. Glossary

- Bond Yield: A single discount number, under which the sum of the present values of all the cash flows of a bond equal its market price. Let t be the current time, C_i the cash flow generated by the bond at time $t_i \geq t$, $P(t)$ the current market price of the bond. Then the bond yield y is defined by the equation,

$$P(t) = \sum_i C_i e^{-y(t_i - t)}, \qquad t \leq t_i. \tag{15.1}$$

- Par Value: The bond's principal, also known as the face value.
- Par Yield: The coupon rate that causes the bond price to match its Par Value.
- Strip(s): Zero coupon bonds that are synthetically created by selling or buying the coupon of a treasury bond separately from the principal.
- Systematic Risk: The risk related to the market as a whole. This risk can not be diversified away. Another way to think of this risk is in terms of correlation. The market components are strongly correlated when the market is going through a bad time.
- Non-systematic Risk: Risk unique to an asset.
- CAPM: Capital Asset Pricing Model. The main argument of the model is that the return should depend only on systematic risk. There is a simple equation describing the CAPM model:

$$E(r_a) = r_f + \beta(r_m - r_f), \tag{15.2}$$

  where r_a is the return of the asset, r_m is the return of the market, r_f is the risk free rate and $E(.)$ is the expectation operator.

- Local Vol Model:

References

[1] Derman, E. Lectures on Smile, Lecture 01.
[2] Papoulis, A. and Pillai, S. U., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Education, 4th Edition, 2002
[3] Piterbarg, V., Funding beyond discounting:collateral agreements and derivatives pricing, Asia Risk, March 2010
[4] Steele, M. J., *Stochastic Calculus and Financial Applications*, Springer, 2001

(London) Pret A Manger, Marble Arch Station,
Oxford Street, London

Email address, Mohsin Javed: mhsnjvd@gmail.com

URL: http://www.zindajaved.com