Boston University Questrom School of Business
MF 793 – Fall 2021

Eric Jacquier

Problem Set 1
Due Sunday September 26th at 11pm Boston Time

Formatted version, do not move text

Problems turned in after the deadline are not graded

- Do the Problem Set in groups of four at the most – students can be in different sections.
- Turn in one single copy for the group on the Gradescope site.
- No email, no paper submission, will be accepted.
- Write solutions in this word file, insert figures from R and hand-written material as pdf graphics. Then save the file as PDF.
- **A properly formatted and spaced file will be posted in a couple days. Do not start filling this file.**
- **To get a check, you need to answer all the questions.**

If you do not do this, you can not get a check plus

- ALL discussion and math questions answered.
- All math questions hand-written
- All figures professionally made with X and Y axes labels and title and fig. numbers
- Tables must have row and column names, title and table number.
- Numbers in the tables must **not** contain too many useless or irrelevant digit, use your common sense as to how many digits to report in a Table. Otherwise it looks like you have no idea what matters.
- All R code as an appendix must be at the back of the homework, starting at the top of a new page.

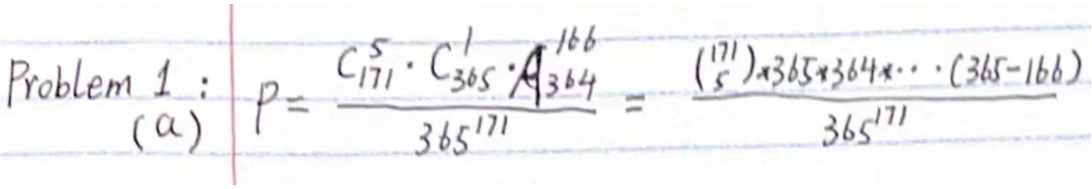
Type the (up to) four team member names below.
Make sure to also enter the names when you upload on Gradescope.

	Last Name	First Name	Section (D1 or D2)
1			
2			
3			
4			

<u>Problem 1:</u> <u>Counting</u>

The new MBA class has arrived, all 171 of them. The social events director wants to know the probability that exactly 5 students have the same birthday, and the rest **all** have different birthdays. Use a 365 day year.

a) Write the theoretical formula **by hand**.



Problem 1 : (a)

$$P = \frac{C_{171}^{5} \cdot C_{365}^{1} \cdot A_{364}^{166}}{365^{171}} = \frac{\binom{171}{5} * 365 * 364 * \cdots \cdot (365-166)}{365^{171}}$$

b) Explain in words how you came up with the numerator, and the denominator.

Numerator:

- C(5,171) means choosing 5 students from the class, C(1,365) means then choosing one day as their common birthday. A(166,364) means all the samples that the rest 166 students have separate birthdays on the rest of 364 days.

Denominator

- The denominator is the whole samples – everyone could be born on any day of a year.

c) Compute it in R and give the result. Show the code you used here below.

-
  ```
  p1 <- choose(171,5) * choose(365,1) * choose(364,166) / (365 ** 171)
  ```

  The result of this code is **0**, which seems unreasonable. So I advance my code to below ones.

-
  ```
  logp <- log(choose(171,5)) + log(choose(365,1)) + log(choose(364,166)) + lfactorial(166) - 171* log(365)
  pp1 <- exp(logp)
  ```

  The result of this code is about **1.025265e-21**

d) Did you encounter a problem computing the R solution ?

- As I mentioned before, I got the result 0 and it's unreasonable at first. The reason why I got it is that R computes 365 ** 171 as infinite result(or overflows). One way to solve it is to change them into the form of log(), which make it possible for R to compute.

e) What is the probability that at least (possibly more) 2 students share a birthday? Give the formula and the final result

- Since it's really hard to compute the possibility of 'at least', we just need to compute the probability that none of these students have the same birthday, and use 1 minus this probability to get the final result:

$$P = 1 - \frac{A\binom{171}{365}}{365^{171}} \approx 1$$

f) Write the definition of the Γ (gamma) function. What is its relation with the factorial function.

- The definition of gamma:

  Γ(x) = integral_0^Inf t^(x-1) exp(-t) dt

  And its relation with the factorial function:

  Γ (n) = Factorial(n-1)

Problem 2: Conditional probabilities, total probability theorem, forecasting stock returns

The folks at TAN Inc. (TisAllNoiz) don't know much finance. They do know that people really care whether the market return is positive. They want to start a weekly financial letter which (seems to) predict the direction of the weekly S&P500 return. They consider three strategies:
1) always predict >0 return,
2) always predict <0 return,
3) randomly predicts >0 return 60% of the time.
Your published research shows that the weekly market return is positive with probability p(U)=60%, and is unpredictable. They must have found this interesting, so they hired you as a consultant for this project.

a) Give them the weekly probability of success, and the expected number of correct predictions after 52 weeks for each rule..

Rule 1
:
Weekly: 0.6
Expected number: 52 * 0.6 = 31

Rule 2:

Weekly : 0.4
Expected number: 52 * 0.4 = 21

Rule 3:

Weekly : 0.6 * 0.6 + 0.4 *0.4= 0.52
Expected number: 52 * 0.52 = 27

b) Explain in words, what is wrong with strategy 3) aka the "randomizing strategy".

c) TAN is getting serious; they want to do **conditional** prediction. The rules in question a) were *unconditional* rules. They ask you to investigate conditional rules, maybe markets are not efficient ! Even though you already know the answer, you will be able to charge them for some data analysis. You hop on to Ken French's web site and download the **weekly** US stock market excess return over the risk free rate.[1] You use only data from Jan. 2012 to Dec. 2020. Of course, you need to assume that future returns will behave consistently with these data.

You will write a nice report for Tan; three versions of the two-way table:

1) simple counts
2) joint probabilities
3) conditional probabilities.

Simple Count Table

	$R_t < 0$	$R_t > 0$
$R_{t-1} < 0$	65	114
$R_{t-1} > 0$	114	176

Joint Probability Table

	$R_t < 0$	$R_t > 0$
$R_{t-1} < 0$	13.86%	24.31%
$R_{t-1} > 0$	24.31%	37.53%

Conditional Probability Table (R_t | R_{t-1})

	$R_t < 0$	$R_t > 0$
$R_{t-1} < 0$	36.31%	63.69%
$R_{t-1} > 0$	39.31%	60.69%

d) Use the numbers in the conditional table to give the **unconditional probability** of success of the rule that predicts that every week the S&P is up (down), it will go up (down).

- The unconditional probability equals to the joint probability here.
  So P(Rt<0 Rt-1<0) = 13.86%   P(Rt>0 Rt-1>0) = 37.53%
  P(Rt<0 Rt-1<0)+ P(Rt>0 Rt-1>0) = 51.39%

e) Given your tables, can a conditional rule improve on this rule?

NO.

f) Given these results, how efficient do you think the US market is?

[1] It is a good time to go on Ken French's web page and look at all the data available there. You can't live in the data based portfolio management space without knowing this data source. There you will find explanations on the data and collection process. Always read the details and description of the data

Problem 3: Bayes rule and conditioning properly

The Wales Cargo company company is in trouble for account manipulations. The North East region manager, Mrs Head Bump, has been instructed to shut down two of three branches in Metro Boston. She communicated to the branch managers of Burlington (Mr Bean), Natick (Mrs Natty), Needham (Mr Veegan), that two of them will be fired and their banks closed, only one will keep his/her job.
They all have the same probability of being fired: p(B) = p(N) = p(V) = 2/3.

Mrs. Head Bump knows which branch will remain open but obviously she must not tell. At a virtual not-so happy hour (and maybe over one too many Summer Ales), she confided to the Burlington manager that the Needham branch would close.

- Mr Bean tells himself: Good news! I now know that Needham will close, so either my branch or Natick will close. My probability of being fired is only ½, not 2/3.

- Mrs. Head Bump realizes that she broke the rule of silence, but she tells herself: "Either Natick or Needham must close anyway, so I have given Mr Bean no information on **his** branch, so he should still think his probability of being fired is 2/3".

Who is wrong, Mrs Head Bump or Mr Bean? It is a question of proper conditioning!

Call "HB" the event: Ms Head Bump tells Mr Bean that the Needham branch will close.

a) Compute p(B|HB). Was Mrs Head Bump or Mr Bean correct?

- This problem is a kind of derivative of Monty Hall problem: instead of compute it directly, we need to think about it from the other side.

  Since the probability of being fired is P(B) = P(N) = P(V) = 2/3, the probability of staying is P(1- B) = P(1- N) = P(1- V) = 1/3

  P(1-B|HB) = P(HB|1-B) / P(HB)
  Where P(HB) = P(1-B)P(HB|1-B) + P(1-N)P(HB|1-N) + P(1-V)P(HB|1-V) = 1/2
  So P(1-B|HB) = P(HB|1-B) / P(HB) = (1/6) / (1/2) = 1/3

  Which means P(B|HB) = 1 – 1/3 = 2/3. Thus, Mrs Head Bump is correct!

b) What "wrong ?" conditioning did Mr Bean use to come up with p(B| ?) = 1/2

- Mr Bean thinks that the back side of the Event HB is 'N is going to be remain'. However, in fact , the back side of the HB is 'Mrs Head Bump didn't tell him that N is going to be filed'

Problem 4: Conditional probabilities, fund performance

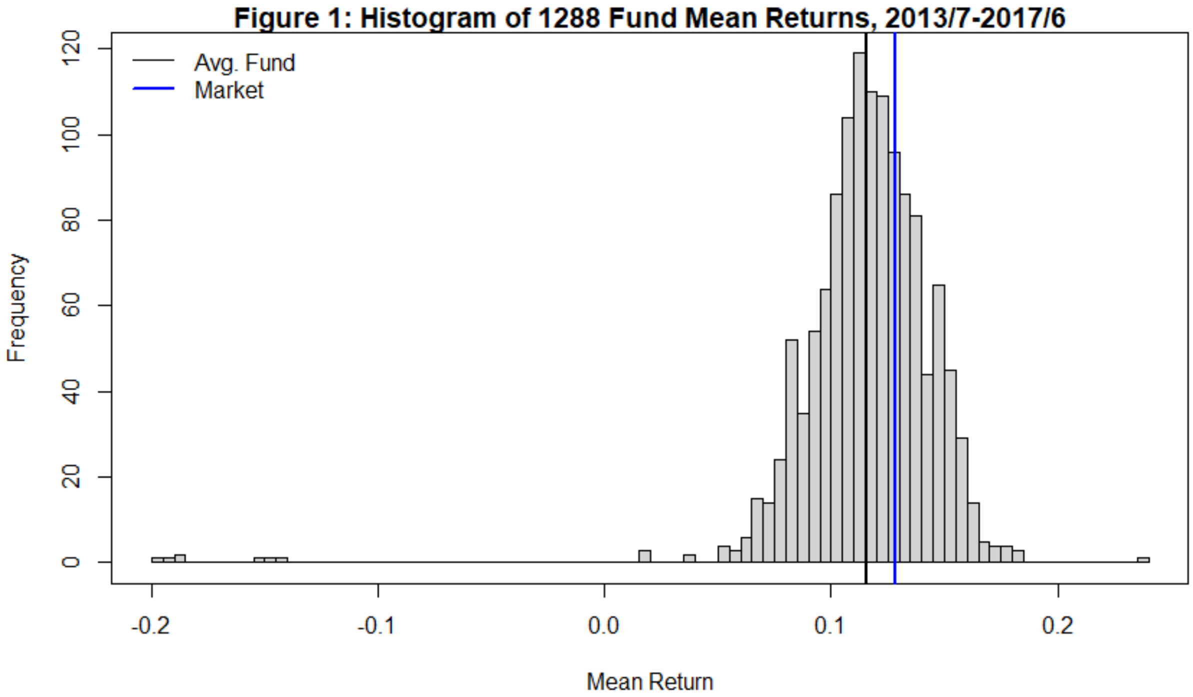
Your boss wants a quick and not dirty recent performance analysis of mutual funds. She reads that most funds don't beat the market but can't get a recent review of fund performance.

You collected monthly returns from Jan. 2010 to June. 2021 for 1288 funds fully invested in the stock market, file funds-1288g-monret.csv in the Data folder.

You know you can go to Ken French's data web site and get the monthly return on the market index for the exact same period, so you do it!

a) Compute the **average monthly return** for each fund for the period (2013/7-2017/6), and the market.

- In Fig. 1 Plot a histogram of these 1288 average fund returns. Annualize these monthly averages by multiplying them by 12 so they have kind of an annual magnitude to them. Add a vertical bar in black for the average **of the 1288 averages**, and a vertical bar in blue for the market average.



- What % of funds beat the market for that period? **31.36%** Answer Here

- What would you expect the result to be if the **fund managers were** randomly picking stocks ?

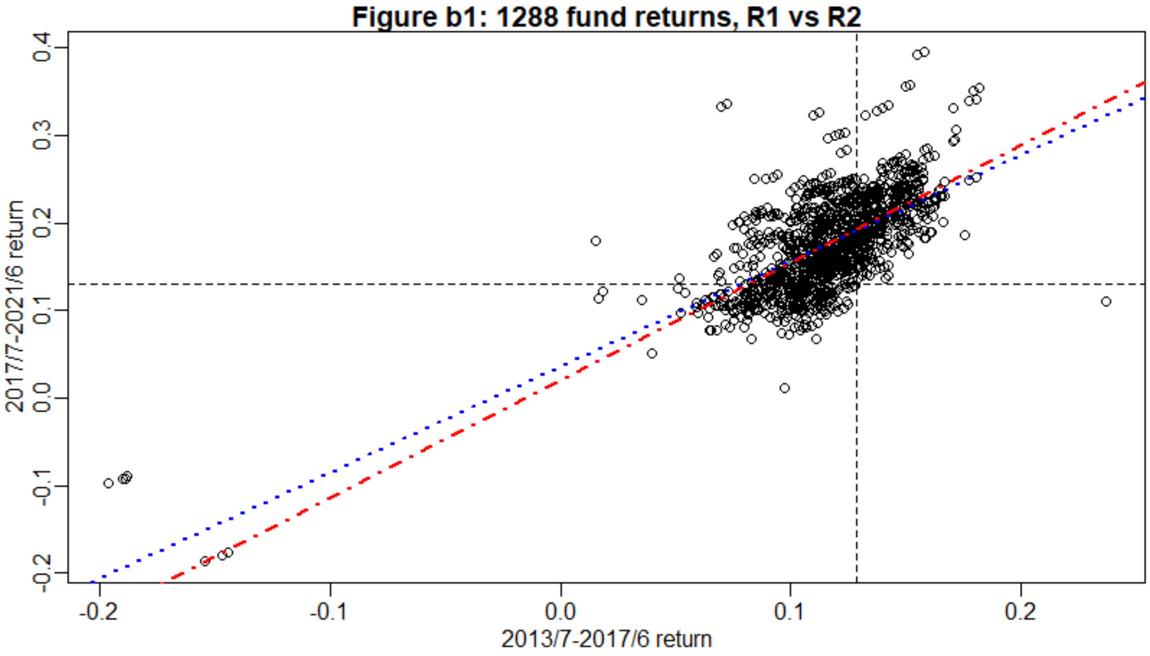
Of Course fund manager can't predict the future, so they still have 50% probability to beat the market if they pick stocks randomly

- In a couple **sentences** tell your boss what you think of mutual fund performance for the first period.

  First, from the big picture, thiese 1588 funds performed quite bad. Most of them were below the market returns. Second, it's really awful that some stock returns even below 0 while the market return is above 0.1. Choosing these mutual funds is a disaster.

b) Your boss says that this is all dandy but she wants you to find out if even a small number of funds can beat the market consistently. "Ahah" you say, "this is why I saved the 2017-21 period, I will prepare a persistence analysis using basic concepts of joint and conditional probabilities". Are some funds consistently the best?

- So you compute Period 1 (2013/7-2017/6) and Period 2 (2017/7-2021/6) mean returns for each fund. Plot R2 vs R1. This scatter plot has 1288 points.
  Add the Market as a point to the graph; choose a symbol that makes it visible (see help("par") and help("points") in R)
  Add a vertical and a horizontal dashed lines going through the market (see abline, and lty graphic parameter in R)



Do you see any pattern in this plot?

Most of these points assemble around the upper right corner as well as the market point, which means they might not have some persistence. However, I've drawn two regression lines, and the positive slope means there's some persistence between two periods. The bad stock remains bad and the good ones are still good.
Since the slope of the redline (positive return in both periods) is larger than the blue one, it seems that the boss's words is correct. Maybe these small number of funds have beaten the market consistently.

What would the plot look like if there was no persistence in performance, strong persistence?

The plot would be more like a circular scatter if there's no persistence in performance.

- Define W (win) as a fund being top 15%, L (lose) bottom 15%, and M (middle) the middle 70% range. Fill three versions of the two-way table for period 2 vs 1:
  1A) simple counts out of 1288
  1B) joint probabilities
  1C) conditional probabilities. Keep only 2 digits for probabilities

Table 1A: Persistence in fund ranking, total counts

	L2	M2	W2
L1	94	95	5
M1	99	702	99
W1	1	103	90

.

Table 1B: Persistence in fund ranking, joint probabilities

	L2	M2	W2
L1	7.30%	7.38%	0.39%
M1	7.69%	54.50%	7.69%
W1	0.08%	8.00%	6.99%

Table 1C: Persistence in fund ranking, conditional probabilities

	L2	M2	W2
L1	0.49	0.49	0.03
M1	0.11	0.78	0.11
W1	0.01	0.53	0.47

- Add a Table 1D, to show what Table 1C should be if funds had all no persistence in abilities.

Table 1D: Conditional probabilities under **no persistence** in ranking

	L2	M2	W2
L1	0.15	0.7	0.15
M1	0.15	0.7	0.15
W1	0.15	0.7	0.15

- In a few sentences summarize these results to your boss, explaining the evidence on the ability of the best funds to remain the best, the so-called "hot-hand".

Hot hand is a kind of persistence performance in capital market.
To be more specific, about 47% of the stock which win in the R1 would win in R2, and about 49% of the stocks which lose in R1 would still lose in R2. So does those in medium range.
In this case, we can see some persistence between two period, which revels the hot hand effect.

c) This was a nice effort but your boss is a bit bored and confused. "Anyway, what do I care that the best remain the best if they can't reliably beat the market! I will still advise our clients to buy DFA! Do me a table showing persistence in beating the market. Just show me the last one, the conditional thing you call it?"

- This time you prepare a two-by two set of three tables 2A, 2B, 2C,2D. You define Win / Lose by having a larger/ smaller return than the market.

Table 2A: Beating the Market, total counts in period 2 vs. 1

	L2	W2
L1	272	612
W1	1	403

Table 2B: Beating the Market, joint probabilities in period 2 vs. 1

	L2	W2
L1	21.12%	47.52%
W1	0.08%	31.23%

Table 2C: Beating the Market, conditional probabilities in period 2 vs. 1

	L2	W2
L1	30.77%	69.23%
W1	0.25%	99.75%

Table 2D: Conditional probabilities with no persistence

	L2	W2
L1	68.63%	31.37%
W1	68.63%	31.37%

- Conclude with respect to persistence in ability to beat the market.

In last section, we found an obvious hot hand effect in these funds. And in this section, we found strong persistence in these top funds. The conditional probability of W1W2 is 99.75% which is much higher than the unconditional probability 31.37%. But there's no evidence shows persistence in poor stocks. In this case, if one chooses the stock that beat the market in period 1, it will have a probability of 99.75% to beat the market again during P2

- As you conclude your presentation and your boss seems interested, Frankie pops his head in and quips "I know how you collected your data, performance is overblown, you have survival bias" He is really annoying, thinks he is a hot shot because he did this MSMF at BU.

Now you need to explain what survival bias is to your boss and how it can affect your results.

Answer:

The Survival Bias is a kind of selection bias, or a kind of logical error in sampling. Briefly, People just concentrate on some special group and can't see the whole things from a broader picture, which lead to an incorrect result.

First, in this case, our 1288 funds might be a special group, where the top funds performed well consistently. However, 1288 stocks are just the corner of the whole capital market. And there the fund might not follow the 'rules' as these 1288 stocks do. In other words, we can choose any 1288 stocks from the market randomly to form groups, and there're a mass of choices. Maybe some groups show top fund have persistence while others are not. However, all these results can't represent the whole market, since they are all the small part of financial world.

Second, the periods we choose is restricted. we just focus two periods. May there're thousands of periods showing the opposite result, but we just choose these two period by coincidence then conclude the wrong output.

In conclusion, our sample groups might be special and lack universality. They only reveal a part of the whole market. We choose them, analyze and get the result. This result only shows the performance of these specific 1288 stocks over this specific period. Thus, this 1288 funds are just like the survivor, and our attitude to them is the survival bias,

Problem 5: Attitude to risk and return

You total wealth is \$10,000. A project can earn 30% or lose 10% with probability 0.5. Your utility of wealth has the shape U(W) = -1/W. You consider whether to invest your total wealth into it!

a) What is the \$CE of your total wealth if you undertake the project?

- $CE = U^{-1}(EU(W)) = -1 / [(-1/13000 + -1/9000)/2] \approx 10636\$$
  CE > 10000 would raise our current wealth, so I will take this project. I will do it.

b) You can borrow at 0%. What is the maximum $\$B_{max}$ you can borrow without going bankrupt in the down case.

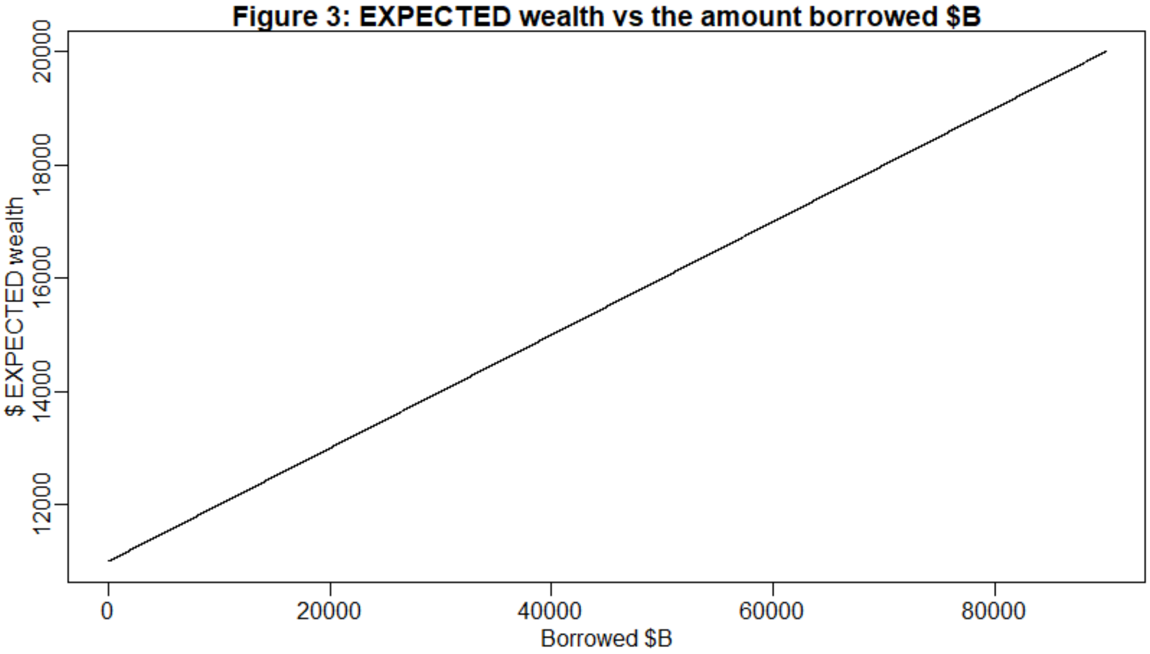
- This problem asks me to pay my debt even if I lose my money:

$$0.9 * (B + 10000) - B = 0$$
$$\rightarrow B=90000$$

  Therefore, The maximus \$B is 90000

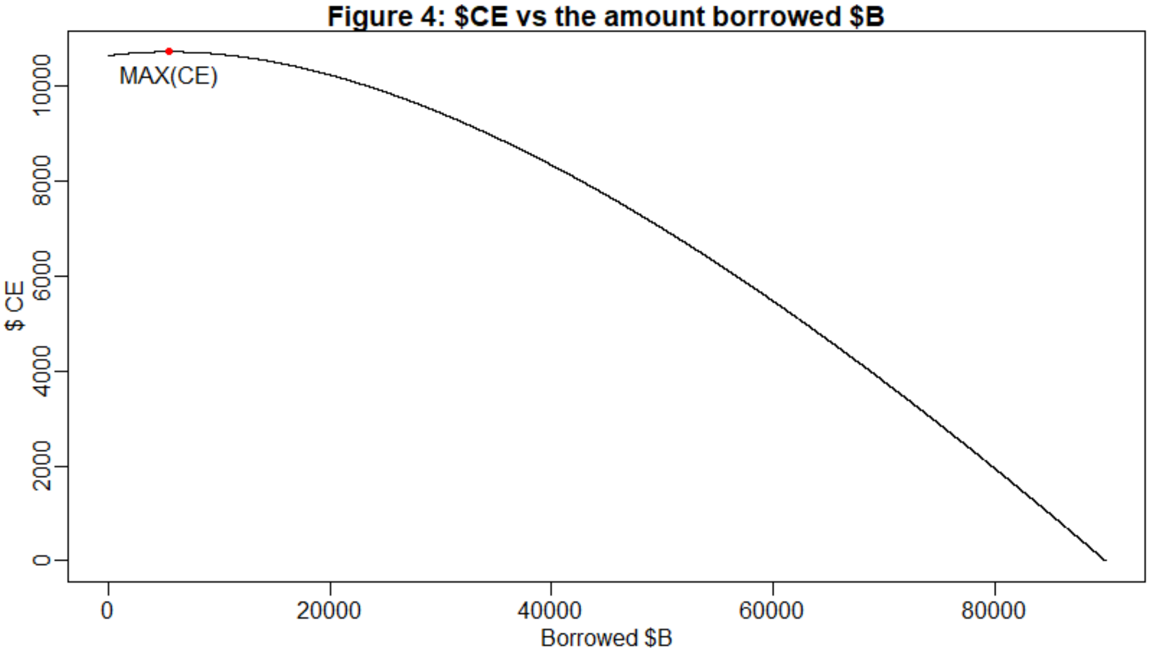
c) In Figure 3 plot your EXPECTED wealth vs the amount borrowed \$B in $[0 , B_{max}]$

- EW = 0.5 [(B+10,000) * 1.3 + (B+10,000) * 0.9] – B = 11,000 + 0.1 B



d) Write the simple formula of your \$CE as a function of \$B. In Figure 4, Use R to plot \$CE vs \$B. Again have B in $[0 , B_{max}]$

- CE(B) = -1 / [(-1/(13000+0.3B) + -1/(9000-0.1B))/2] -B



d) How much would you borrow optimally to increase your expected utility? Show that point on your Figure 4.

- So I will borrow 5470$ because at this point, we would get the highest CE = 10717.97. From a bigger picture, the CE would rise first and then reach the highest point then go down consistently. Borrow a proper amount of money is a wise choice,

R CODE BELOW

```
s  # Problem 1.b
p1 <- choose(171,5) * choose(365,1) * choose(364,166) / (365 ** 171)
# p1 reveals 0, because R computes 365 ** 171 as infinite result.

logp <- log(choose(171,5)) + log(choose(365,1)) + log(choose(364,166)) + lfactorial(166) - 171* log(365)
pp1 <- exp(logp)

# Problem 1.e
p2 <- 1 - exp(log(choose(365,171)) + lfactorial(171) - 171*log(365))



#problem 2
wkrf  <- read.csv("C:/Users/MSI_NB/Desktop/BU FALL 2021/MF793/HM1/F-F_Research_Data_Factors_weekly.csv",header=T)
rf = wkrf[4462:4931,2]
rf1 = wkrf[4462:4930,2]
rf2 = wkrf[4463:4931,2]

#simple count table
tablecount <- matrix(0,ncol=2,nrow=2)
tablecount[1,1] <-length(rf1[rf1<=0 & rf2 <=0])
tablecount[1,2] <-length(rf1[rf1<=0 & rf2 >0])
tablecount[2,1] <-length(rf1[rf1>0 & rf2 <=0])
tablecount[2,2] <-length(rf1[rf1>0 & rf2 >0])
tablecount

#joint probability
tablej = tablecount / length(rf1)
tablej

#condition probability
r1 = length(rf1[rf1<=0]) / length(rf1)
r2 = length(rf1[rf1>0]) / length(rf1)

tablecon = tablej / c(r1,r2)
```

```
tablecon


#unconditional
pu = length(rf[rf<=0]) / length(rf)
pd = length(rf[rf>0]) / length(rf)
pu
pd






#Problem 4


#a)
dataa <- read.csv("C:/Users/MSI_NB/Desktop/BU FALL 2021/MF793/HM1/funds-1288g-mon.csv",header = T)
mkrf  <- read.csv("C:/Users/MSI_NB/Desktop/BU FALL 2021/MF793/HM1/KF-Market-mon.csv",header=T)
rmret<-(mkrf[,2]+mkrf[,3])/100


rm = mean(rmret[1045:1092])
fund <- dataa[43:90,2:1289]
avreturn <- apply(fund,2,mean)


hist(avreturn*12,xlab="Mean Return",freq=T,nclass=100,main="")
title("Figure 1: Histogram of 1288 Fund Mean Returns, 2013/7-2017/6",line=0.2)
box()
abline(v=mean(avreturn)*12,col="black",lwd=2)
abline(v=rm*12,col="blue",lwd=2)
legend("topleft",c("Avg. Fund","Market"),bty="n",col=c("black","blue"),lwd=c(1,2),lty=1)


length(avreturn[avreturn>rm]) # 404 funds beat the market
length(avreturn[avreturn>rm])/1288 #31% of these funds beat the market




#b)


rm2 = mean(rmret[1093:1132])
```

```
fund2 <- dataa[91:138,2:1289]
avreturn2 <- apply(fund2,2,mean)

par(mgp=c(1.5,0.5,0))
plot(avreturn*12,avreturn2*12,xlab="2013/7-2017/6 return",ylab="2017/7-2021/6 return")

title("Figure b1: 1288 fund returns, R1 vs R2",line=0.2)
abline(v=rm*12,lty=2):abline(h=rm2*12,lty=2)
abline(lsfit(avreturn*12,avreturn2*12),lty=3,lwd=2,col='blue')
abline(lsfit(avreturn[avreturn>0]*12,avreturn2[avreturn>0]*12),lty=4,lwd=2,col="red")

#plot ranks
rank1<-rank(avreturn)
rank2<-rank(avreturn2)
plot(rank1,rank2,xlab="Rank1",ylab="Rank2")
title("Figure b2: Funds rank persistence from 2013/7-2017/6 to 2017/7-2021/6",line=0.5)
abline(lsfit(rank1,rank2))


#fill the tables
condition <- cbind(quantile(avreturn,c(0.15,0.85)),quantile(avreturn2,c(0.15,0.85)))

#table A
tableA <- matrix(0,ncol=3,nrow=3)

tableA[1,1]<-length(avreturn[avreturn<condition[1,1]&avreturn2<condition[1,2]])
tableA[1,3]<-length(avreturn[avreturn<condition[1,1]&avreturn2>condition[2,2]])
tableA[1,2]<-length(avreturn[avreturn<condition[1,1]]) - tableA[1,1] - tableA[1,3]

tableA[3,1]<-length(avreturn[avreturn>condition[2,1]&avreturn2<condition[1,2]])
tableA[3,3]<-length(avreturn[avreturn>condition[2,1]&avreturn2>condition[2,2]])
tableA[3,2]<-length(avreturn[avreturn>condition[2,1]]) - tableA[3,1] - tableA[3,3]

tableA[2,1]<-length(avreturn[avreturn>condition[1,1]&avreturn<condition[2,1]&avreturn2<condition[1,2]])
tableA[2,3]<-length(avreturn[avreturn>condition[1,1]&avreturn<condition[2,1]&avreturn2>condition[2,2]])
tableA[2,2]<-length(avreturn[avreturn>condition[1,1]&avreturn<condition[2,1]]) - tableA[2,1] - tableA[2,3]
```

```
tableA

#table B (joint probability)
tableB = tableA / 1288
tableB

#table C (conditional porbability)
condp = c(0.15,0.7,0.15)
tableC = tableB / condp
tableC

# add a tabldD to show the condition probability with no persisitence
tableD <- matrix(0,ncol=3,nrow=3)
tableD[,1] <- 0.15
tableD[,2] <- 0.7
tableD[,3] <- 0.15
tableD


#c)

table2A <- matrix(0,ncol=2,nrow=2)
table2A[1,1] <- length(avreturn[avreturn<rm&avreturn2<rm2])
table2A[1,2] <- length(avreturn[avreturn<rm&avreturn2>rm2])
table2A[2,1] <- length(avreturn[avreturn>rm&avreturn2<rm2])
table2A[2,2] <- length(avreturn[avreturn>rm&avreturn2>rm2])
table2A

table2B = table2A / 1288
table2B

table2C = table2B / c(length(avreturn[avreturn<rm]),length(avreturn[avreturn>rm])) * 1288
table2C
```

```
#problem 5

BR = seq(0,90000)
EW = 11000 + 0.1 * BR
plot(BR,EW,xlab="Borrowed $B",ylab="$ EXPECTED wealth",type="l")
title("Figure 3: EXPECTED wealth vs the amount borrowed $B",line=0.2)

BR2 = seq(0,90000)
CE<--1/(-0.5/(13000+0.3*BR2)-0.5/(9000-0.1*BR2))
plot(BR2,CE,xlab="Borrowed $B",ylab="$ CE",type="l")
title("Figure 4: $CE vs the amount borrowed $B",line=0.2)
BR2[CE==max(CE)] #the best point is this, when BR2 = 5470
points(BR2[CE==max(CE)], max(CE),pch=20,col="red")
text(BR2[CE==max(CE)], max(CE),"MAX(CE)",pos=1)
```

The Following R commands are of **great** interest

- Do help(“commandname”) for details on how to use a command

- Repeat after me: **“In R I will try to avoid loops as much as possible”**

- Counting: look at the commands choose, factorial, lfactorial
  # What if you need factorial(171) .... or bigger?
  # For ratios of overflow numbers, make partial divisions and take the product after
  # or remember that: ratio = exp(log(ratio)), and log(ratio) = log(numerator)-log(denominator)
  # some functions that can overflow can have their log directly computed, lfactorial, lgamma

- Reading Data
  Easiest is simple rectangular data sets with a header row in .csv format
  arrayname <- read.csv(“filename.csv”)
  # it assumes a header, first row contains name variables
  also: read.table(), scan()

- Writing Data for your results
  - Think of putting table results in a R array, then write the array to a .csv file. The .csv file can then be put directly in a .xls table and into a word document
  - There are many ways to prepare latex tables in R for those using that more sophisticated quantitative paper formatting package. Not needed for our problem sets.

```
write.csv(arrayname, “table.csv”)
also: write.table(). write()
```

- Choosing subsets of data

Say we have a matrix with 200 rows and 5 columns. We can select any subset we want

```
smallmat<- bigdatamat[101:200,3:5]
    # creates small mat as the indicated subset with columns 3 to 5, rows 101 to 200
smallmat<- bigdatamat[101:200,c(2,4)]
    # takes columns 2 and 4 of bigmat
smallmat<-bigdatamat[bigdatamat[,1]>20171231,]
    # takes rows with date after 12/13/2017
```

Can also use multiple conditions with the & (and) , | (or), and ! (not) signs. Do help(“Logic”) to see more

```
retmat[retmat[,1]>20171231 & retmat[,2]>0,]        #
```

selects all periods past 20171231 where the first return is >0.

Can directly count over a condition

```
     sum(rets[,2]>0)        # gives the number of positive returns in column 2
     length(rets[,2])       # total number of observations
```

- Building blocs

```
1:5            # a sequence of numbers from 1 to 5
c(1,5)         # the numbers 1 and 5
c(vector1,vector2)        # concatenates two vectors into 1
cbind(mat1,mat2)          # joins 2 matrices next to each other,         column bind
rbind(mat1,mat2)          # joins 2 matrices with mat1 on top of mat2,   row bind
length(vector)            # gives the length
dim(matrix)               # gives the dimension
```

Basic loop

```
for (i in 4:10){
     sales[i] <- price[i]*quantity[i]
     }
```

But we can (and should) avoid loops:

```
     sales<-price*quantity                  # Voila!
```

The simple * in R is not an inner product but the Hadamard (element-wise) product.. very convenient!

So.. what is the inner product of vectors or matrices?   a %*% b
There we need a to be a row and b a column

- prod and sum commands give the product or sums of elements of a vector.

```
prod(1+ ret)-1            # compound returns contained in vector ret
```

- Plots

```
plot(x,y, pch="O" ,xlab="This is x", ylab="This is y",col="blue")
     # pch allows you to choose the symbol.
     # By default the graph comes out as a scatter plot with points
```

Other useful qualifiers:

```
plot(x1,y1,  xlim= c(low,high),   ylim=     ) # axes limits,
points(x2,y2, col="red",pch="*")          # adds points at (x2,y2) to the graph with red
                                          # color, and a star as symbol
title("Figure 1: This is my title")       # adds title to the plot
```

Do help("par") for graphics tweaks

```
abline(h=10)              # adds horizontal line at chosen value of 10
abline(v=   )             # adds vertical line
abline(a=… ,  b= … )      # adds straight line with intercept a, slope b
abline(lsfit(x,y))        # adds regression line to plot
```

- Histograms

  ```
  hist(rets, nclass=40, prob=T)
        # always try to  increase the default number of bins (nclass) to make the
        # histogram  look more realistic. Otherwise it’s pretty useless
        # prob=T scales the histogram to be a density.
  ```

- Descriptive stats

  ```
  mean(myret)
  sd(myret)         # standard deviation
  var(myret)        # variance
  cor(ret1,ret2)    # correlation
  ```

  If myret is a matrix of returns, not a single vector, var(myret) computes the covariance matrix.
  Use the apply command below (we don’t loop) to get only the variances

- apply(matrix, 2, fun)
  computes the function fun for each column (2) or row (1) of a matrix. Result is a vector.

  For example:

  ```
  Per1mean<- apply(retmatrix[retmatrix[,1]<20140000 , 2: 501],2,mean)
                #       returns a vector of 500 means using data up to Dec 2013
                #       Use & for more than one conditions, like A & B  (A and B)
                #       Ummh, this can be useful!
  ```