PART 1: Bayes Theorem

A new analytical method to detect pollutants in the water is being tested. This new method of chemical analysis is important because it can detect three different contaminants: Organic pollutants, volatile solvents, and chlorinated compounds. The makers of the test claim that it can detect organics with 99.7% accuracy, volatile solvents with 99.95% accuracy and chlorinated compounds with 89.7% accuracy. If pollutants aren't present, the test does not signal.

Samples are prepared for the calibration of the test and 60% of them are contaminated with organics, 27% with volatile solvents, and 13% with traces of chlorinated compounds.

A test sample is selected randomly.

1a. What is the probability that the test will signal?

1b. If the test signals what is the probability that chlorinated compounds are present?

Draw a tree diagram to outline the experiment. Solve using Bayes Theorem.

Car Fault Assignment (Software)

You have been instructed to quantify and improve the failure rates associated with components made in a car assembly plant that's just coming online (still working out a lot of bugs). Below are the incident rates of faults (flaw or problem, but still functions) and failures for selected systems and subsystems that are known to have the highest rates at the plant. You can assume the failures occur independent of each other (e.g., A wiring failures does not cause an alternator failure).

System	Subsystem	P(fault)	P(fault and failure)	
Braking				
	brake pads	1/500	1/100000	
	calipers	1/1000	1/2000	
	brake line	1/100	1/20000	
	master cylinder	1/1500	1/1500	
Drive train				
	Engine			
	head gasket	1/2000	1/400000	
	engine block	1/2500	1/800000	
	Transmission			
	each gear	1/1500	1/7000	
	Rear Differential	1/500	1/3000	
	Front Differential	1/550	1/3000	
Electrical				
	wiring	1/60	1/600	
	alternator	1/100	1/150	

Note: To input these values into an R dataframe you can use the following syntax-

```
car.df=data.frame(subsystem=c("brake.pads", "calipers", "brake.line",
               "master.cylinder", "head.cover","engine.block",
               "gear", "rear.diff", "front.diff", "wiring",
               "alternator"),
```

```
p.fault=c(1/500, 1/1000, 1/100, 1/1500, 1/2000, 1/2500,
          1/1500, 1/500, 1/550, 1/60, 1/100),
p.failure=c(1/100000, 1/2000, 1/20000, 1/1500,
          1/400000, 1/800000, 1/7000, 1/3000, 1/3000,
          1/600, 1/150))
```

A. What is the probability of all the gears having faults if the transmission has **six** gears?
B. The transmission engineers have been working on a new system that would guarantee one gear failure wouldn't cause a catastrophic failure, but two or more would (system redundancy). They have only been able to prototype a system with two "gears" so far. What is the probability of a 1 "gear" failure? Assuming the same probability of "gear" failure, 1/7000.

Extra Credit: If they can ramp the new transmission system up to 6 "gears", what percentage of failures could be avoided i.e. P(1 failure)/P(any failure) for 6 gears? Show your work.

C. One of the gear machines was damaged during production and the P(gear failure) is now 1/30. Write a function to compute the distribution of failures over 100000 transmissions (each transmission has 6 gears) eg. 5000 transmissions with single gear failures, 400 transmissions with two gear failures, etc.
D. What is the probability of a gear failing given it has a fault (use original data in table)?
E. What is the probability of an electrical failure?
F. What is the probability of a failure in the wiring given a failure has been determined in the electrical system? <u>Use a probability of an electrical failure of 8/900 opposed to the answer from part E.</u>
G. What is the probability of no faults in the braking system?
H. What is the probability of no failures in the braking system?
I. What is the probability of at least 1 fault in the braking system?
J. What is the probability of at least 1 failure in the braking system?
K. What is the probability of a master cylinder failure given there is a failure in the braking system? <u>Use a probability of a braking failure of 11/5000 opposed to the answer from part J.</u>
L. Determine what the probability of a fault being present in any car. (Show the work by hand or code used, this should be an analytical approach either way: **HINT-use the prod function to take the product of a vector. If you created a dataframe with all the faults and failures, this is only a few characters on 1 line of code**)
M. Determine what the probability of a failure being present in any car. (Show the work by hand or code used, this should be an analytical approach either way: **HINT-Same as L**)
N. Create code that provides the probability of each subsystem given there is a failure in the car, ie. P(brake pads| a failure in the car), P(calipers| a failure in the car), etc.
O. Simulate 100 million cars and provide the number of expected failures in each system. Hint: check out the documentation on rbinom. The example code runs in rbinom 10 billion cars in 0.04 seconds utilizing the rbinom function.

```
nums = 100000000
car.df$failures = 0
for (subsystem.index in 1:length(car.df$subsystem)){
  car.df$failures[subsystem.index] <- rbinom(1, nums,
car.df$p.failure[subsystem.index])
}
```

P. Create a proportional table, `prop.table(table(results))`, to compare your answers from N to the simulated values.

PART 2 – Probability: Slot Machine

You are working on a slot machine for your company. The slot machine uses a three drum design. A person pulls the lever and the drums begin to spin. Each drum is designed to stop at random in one of 12 resting positions. The resting positions include: 1 cherry, 1 bell, 1 lemon, 1 bar, 1 seven, and 7 blank spaces. Each position has an equal probability and each drum's outcome resting position is independent of the other two. Use this information to answer the following questions (make sure to provide your work and code used to complete the assignment):

a) The big pay out occurs if all drums land on cherries. What is the probability that this occurs?
b) A lesser pay out occurs if any of the other symbols listed above (blank spaces result in no payout) occur on all drums. What is the probability of this small payout occurring?
c) If the big payoff is $3000, the slot machine costs $2 per play, and the other payoffs are smaller but equal. What is the highest amount these payouts can be without the slot machine losing the owner money, breakeven point or profit margin=0 (it's a very hospitable casino)?
d) Create a function to simulate this slot machine and provide your code.
e) Simulate a considerable number of plays and confirm the payout outcomes are as expected (approx. breakeven). Provide a table of your outcomes and any additional results to support the function produces the desired behavior.

Notes:

1- margin = (total_cost_to_play -total_payouts) / total_cost_to_play can sway significantly even at runs>>1000 so consider a high number of plays.

If the margin is greater than ±10%, try a few rounds of testing. This is a random process but you should hover around 0 margin if you've coded it right.

2- If a simulation of 10,000 plays takes longer than a minute or two to run check for inefficiencies in your code. As a benchmark, the example solution runs 10,000 plays in 0.25 seconds using a for loop and memory allocating the result array beforehand.

If you use a loop of some sort, make sure to allocate the memory of any vector opposed to growing on iteration, eg.

```
array=rep(NA,1000) #allocating the memory prior to loop
for (i in 1:1000) {
array[i] <- someFunction()
}.
```

Allocating memory considerable speeds up an iterative process in R.

3- If things are running for more than a few minutes without result, cut the number of plays by an order of magnitude until you can provide some results supporting the code from d. performs near to expected.