Dallin Romney
Design of Experiments

Attachments: figure, table, code

Summary
This memo is split into two parts. Part I contains two scenarios wherein Bayes Theorem is applicable; a chemical test for pollutant detection and failure rates for car components. Part II discusses slot machine probabilities. Problem introductions and questions are all included and paraphrased from the homework 2 handout.

Part I. Bayes Theorem

Scenario 1. Pollutant Detection Test

A chemical test is developed to detect pollutants in water. The makers claim they can detect organic pollutants with 99.7% accuracy, volatile solvents with 99.95% accuracy and chlorinated compounds with 89.7% accuracy. The test doesn't signal if pollutants aren't present. Samples are prepared for test calibration: 60% are contaminated with organic pollutants, 27% with volatile solvents, and 13% with traces of chlorinated compounds.

Figure 1 shows a tree diagram of the pollutant test, and shows Bayes Theorem applied to it. The probability that the test will signal at all is **98.47%**. If the test signals, the probability that chlorinated compounds are present is **11.84%**.

Scenario 2. Car Fault Assignment

A car assembly plant wishes to improve failure rates in component production. Table 1 shows the independent probabilities for fault and failure for the selected components. Any R code used to answer the following questions is attached. Answers are in blue.

A. What is the probability of all the gears having faults if the transmission has six gears?
   These are independent events, so the probability of all six gears having faults = P(gear fault)^6 = $(1/1500)^6$ = **$8.78 \cdot 10^{-20}$ → Very low probability!**

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.
   The probability of any gear failing is 1/7000, so the probability of only one of the two gears failing is (2 choose 1) · (1/7000) · (6999/7000) = 0.000286. The probability of both failing and thus the two-gear transmission failing is $(1/7000)^2$ = **$2.04 \cdot 10^{-8}$**.

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.

The probability of any gear failing is 1/7000, so the probability of only one of the six gears failing is (6 choose 1) · (1/7000) · $(6999/7000)^5$ = 0.0008565. The probability of two failing is (6 choose 2) · $(1/7000)^2$ · $(6999/7000)^4$, and so on (binomial theorem). Thus, the probability of the transmission failing is:

(6 choose 2)$(1/7000)^2$ · $(6999/7000)^4$ + (6 choose 3) · $(1/7000)^3$ · $(6999/7000)^3$ + (6 choose 4) · $(1/7000)^4$ · $(6999/7000)^2$ + (6 choose 5)$(1/7000)^5$ · (6999/7000) + $(1/7000)^6$ = $3.06 \cdot 10^{-7}$. This is on a similar order as the two-gear version. The percentage of failures that could be avoided is 0.0008565 / ($3.06 \cdot 10^{-7}$ + 0.0008565) = **99.9642% ≈ pretty much 100% of failures!**

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) e.g. 5000 transmissions with single gear failures, 400 transmissions with two gear failures, etc.

```
# PART I.C
# gearFailureDist returns a distribution of how many gears fail in a set of transmissions
# t is the number of transmissions
# P is the probability of failure for one gear
# N is the number of gears in each transmission
# distribution is a data frame with the results
gearFailureDist = function(t, P, N){

  numFailedGears = 0:N

  # Use binomial theorem to populate vector of number of transmissions failed for each # gears failed
  numTransmissions = t*((1-P)^N)
  for (i in 1:N) {
    numTransmissions = c(numTransmissions, t*choose(N, i)*(1-P)^(N-i)*(P^i))
  }
  distribution = data.frame(numFailedGears, round(numTransmissions)) # Compile the results into a frame

  return(distribution)
}
# Distribution for 100000 6-gear transmissions with a 1/30 probability of a gear failing
dist = gearFailureDist(100000, 1/30, 6)
dist
```

```
  numFailedGears round.numTransmissions.
1              0                   81594
2              1                   16882
3              2                    1455
4              3                      67
5              4                       2
6              5                       0
7              6                       0
```

D. What is the probability of a gear failing given it has a fault (use original data in table)?
P(A | B) = P(B | A) · P(A) / P(B) → P(failing | fault) = P(fault | failing) · P(failing) / P(fault) = 100% · (1/7000) / (1/1500) = 1500/7000 = 3/14 = **21.43%**.

E. What is the probability of an electrical failure?
Assuming the car has one "wiring" part and one alternator, P(any electrical failure) = 1 – P(no failure) = 1 – (599/600) · (149/150) = **0.832%**

F. What is the probability of a failure in the wiring given a failure has been determined in the electrical system? Use a probability of an electrical failure of 8/900 opposed to the answer from part E.
The probability of wiring failure is 1/600, so:
P(wiring | electrical failure) = P(electrical failure | wiring) · P(wiring) / P(electrical failure) = 100% · (1/600) / (8/900) = 3/16 = **18.75%**.

G. What is the probability of no faults in the braking system?
P(no faults) = P(no brake pad faults) · P(no caliper faults)… = (1 – P(brake pad faults)) · (1 – P(caliper faults))… = (499/500) · (999/1000) · (99/100) · (1499/1500) = **98.64%** (assuming one of each component)

H. What is the probability of no failures in the braking system?
P(no failures) = P(no brake pad failures) · P(no caliper failures)… = (1 – P(brake pad failures)) · (1 – P(caliper failures))… = (99999/100000) · (1999/2000) · (19999/20000) · (1499/1500) = **99.88%** (assuming one of each component)

I. What is the probability of at least 1 fault in the braking system?
P(at least one fault in the braking system) = 1 – P(no fault in the braking system) = 1 – 98.64% = **1.36%**

J. What is the probability of at least 1 failure in the braking system?
P(at least one failure in the braking system) = 1 – P(no failure in the braking system) = 1 – 99.88% = **0.12%**

K. What is the probability of a master cylinder failure given there is a failure in the braking system? Use a probability of a braking failure of 11/5000 instead of part J's answer.
The probability of master cylinder failure is 1/1500, so:
P(master cylinder failure | braking failure) = P(braking failure | master cylinder failure) · P(master cylinder failure) / P(braking failure) = 100% · (1/1500) / (11/5000) = 5000/16500 = 10/33 = **30.30%**.

L. Determine what the probability of a fault being present in any car.
The probability of any fault being present is the same as 1 minus the probability that there are no faults in any of the components. This is 1 - (499/500) · (999/1000) · (99/100) · (1499/1500) · (1999/2000) · (2499/2500) · (1499/1500) · (499/500) · (549/550) · (59/60) · (99/100) = **4.49% probability of a fault present in any car.**

M. Determine what the probability of a failure being present in any car.
The probability of any failure being present is the same as 1 minus the probability that there are no failures in any of the components. This is 1- (99999/100000) · (1999/2000) · (19999/20000) · (1499/1500) · (399999/400000) · (799999/800000) · (6999/7000) ·

(2999/3000) · (2999/3000) · (599/600) · (149/150) = **1.03% probability of a failure present in any car.**

N. Create code that provides the probability of each subsystem given there is a failure in the car, i.e. P(brake pads| a failure in the car), P(calipers| a failure in the car), etc.
P(component | failure) = P(failure | components) · P(component) / P(failure) → P(failure | components) = 1 but the sum of the component failure probabilities will be very slightly greater than 1 because there is a small chance that more than one component failed if the car failed. See code in part O.

O. Simulate 100 million cars and provide the number of expected failures in each system.
See table in part P.

P. Compare your answers from N to the simulated values.

```
> car.df
        subsystem      p.fault    p.failure failures  expected     prop
1      brake.pads 0.0020000000 0.0000100000     1043   1000.00 1.043000
2        calipers 0.0010000000 0.0005000000    50233  50000.00 1.004660
3      brake.line 0.0100000000 0.0000500000     4860   5000.00 0.972000
4 master.cylinder 0.0006666667 0.0006666667    66418  66666.67 0.996270
5      head.cover 0.0005000000 0.0000025000      233    250.00 0.932000
6    engine.block 0.0004000000 0.0000012500      116    125.00 0.928000
7            gear 0.0006666667 0.0001428571    14148  14285.71 0.990360
8       rear.diff 0.0020000000 0.0003333333    33572  33333.33 1.007160
9      front.diff 0.0018181818 0.0003333333    33696  33333.33 1.010880
10         wiring 0.0166666667 0.0016666667   166214 166666.67 0.997284
11     alternator 0.0100000000 0.0066666667   666779 666666.67 1.000169
```

I had difficulty applying the prop.table() function so I added the proportions as the last column in the data frame. The biggest discrepancy was the component with the lowest frequency of failure – the engine block, with 92.8% similarity.

Part II. Slot Machine Probability

A company makes slot machines with three drums that stop at random in one of 12 resting positions: 1 cherry, 1 bell, 1 lemon, 1 bar, 1 seven, and 7 blank spaces. All the positions are independent and of equal probability.

A. The big payout occurs if all drums land on cherries. What is the probability of this?
$(1/12)^3$ = **0.05787%**

B. A lesser payout occurs if any of the other symbols listed above (blank spaces results in no payout!) occur on all drums. What is the probability of this small payout occurring?
= P(any non-blank combo) – P(big payout) = $(5/12)^3 - (1/12)^3$ = **7.16593%**

C. If the big payoff is $3000, the slot machine costs $3 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 = 0)?
$3/play - ($3000 / big win) · (0.0005787 big wins / play) – ($x / small win) · (0.0716593 small wins / play) = 0 → x = **$17.64**

D. Create a function to simulate this slot machine and provide your code.

```r
slotMachine = function(numDrums, drumSize, numBlank){

  # Determine probabilities of each prize option (or no prize)
  pBig = (1/drumSize)^numDrums
  pSmall = ((drumSize - numBlank)/drumSize)^numDrums - pBig
  pNothing = 1 - pBig - pSmall
  probs = c(pBig, pSmall, pNothing)

  prizes = c("Big Prize", "Small Prize", "No Prize")

  # Randomly return a result!
  return(sample(prizes, 1, replace = TRUE, prob = probs))
}
```

E. Simulate a considerable number of plays and confirm the payout outcomes are as expected. Provide a table of your outcomes and any additional results to support that the function produces the desired behavior.

Frequency:

```
outputs
  Big Prize    No Prize Small Prize
        564      927768       71667
```

Proportion:

```
outputs
  Big Prize    No Prize Small Prize
   0.000564    0.927768    0.071667
```

```
> margin
[1] 0.01459804
```

With a sample of 1000000 plays, the margin was typically within 1-4%, and the slot machine results were almost exactly as expected. 0.0564% vs 0.0578% expected for the big prize, 7.17% as expected for the small prize, and 92.78% vs 92.77% expected for no prize. See attached code.

Conclusion

This assignment thoroughly introduced conditional and independent probability. Bayes theorem is useful for conditional probability, but every problem must be approached with caution, and the answers should all make sense!

Attachments

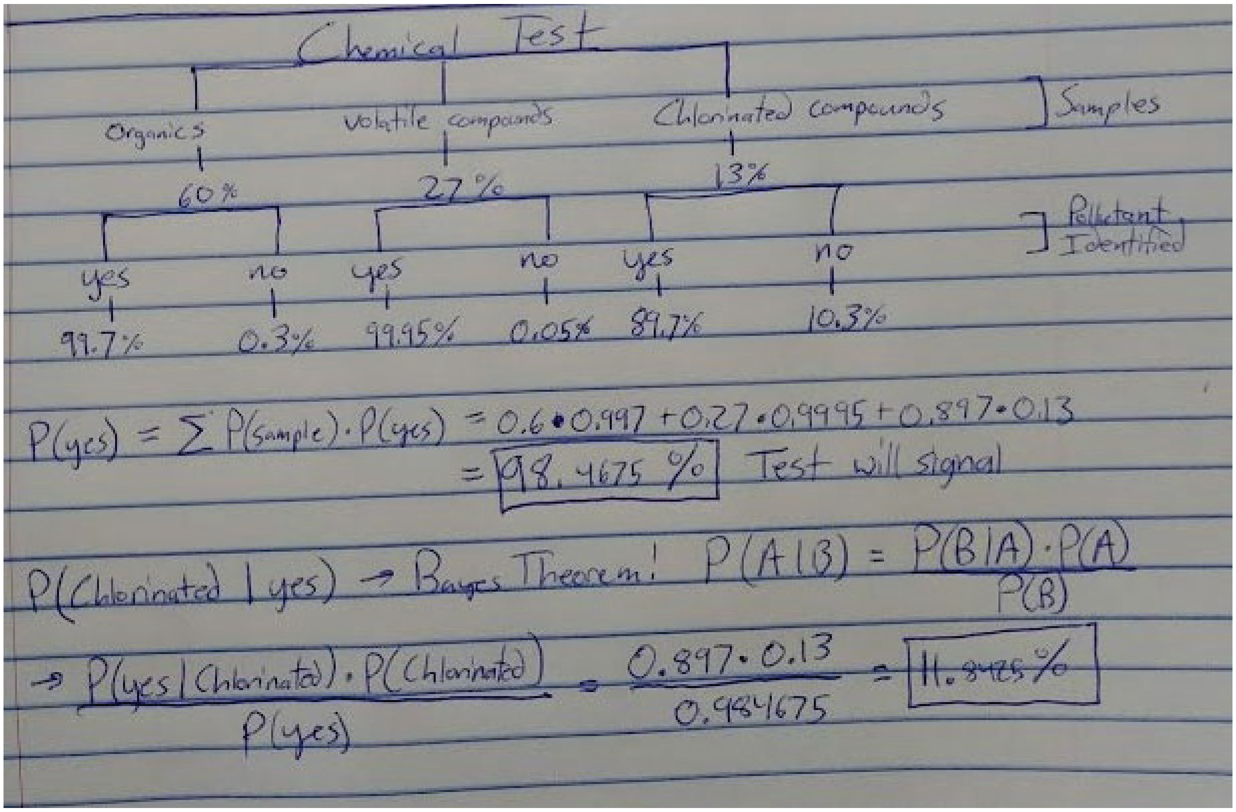

Figure 1. Tree diagram for pollutant detection probabilities

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

Table 1. Car part fault and failure probabilities