STATISTICS FOR ENGINEERS
REVIEW PROBLEMS SOLUTIONS

Problem This is a quiz

A quiz is given to the class. It consists of 5 True-False questions and 5 multiple-choice question, with four options each. How many ways can the quiz be answered? Assuming that a certain student guesses on all of the questions, what is the probability that the student will Ace the quiz (get 9 or 10 correct)?

Solution: This is an interesting question, because there are two types of questions a person may miss. Thus, this requires breaking apart into two parts and using conditional probability.

First, let us figure out the probability that the student gets all 10 questions correct. This is just the number of ways of getting all questions correct (1) and dividing by all the possible ways of answering the 10 questions: $2^5 \times 4^5 = 32768$.

Now, let us assume that the student missed a single question. If the student missed a TF question, then the number of ways of answering 9 questions correctly is 1×5. If the student missed a MC question, then the number of ways of answering 9 questions correctly is 3×5.

Now, we note that the probability of randomly missing a MC question is more than that of missing a TF question: $\mathbb{P}\left[W|TF\right] = 0.500$, whereas $\mathbb{P}\left[W|MC\right] = 0.750$. So, let us put it all together now:

$$\mathbb{P}\left[C \geq 9\right] = \left(\frac{1}{32768}\right) + \left(0.500\ \frac{5}{32768} + 0.750\ \frac{15}{32768}\right) \approx 0.000450$$

This is about a half of what it would be if there were just 10 TF questions (0.000977). You may want to check how I got that answer. ⋄

Problem A license to drive

The license plates for the great state of Oregon consist of six spaces. The first space is letter that signifies the month the plate expires. Possible values are A, B, C, D, E, F, G, H, J, K, L, and M. The next two spaces consist of letters other than O and I. The final three spaces consist of digits (zero through nine). How many license plates can Oregon produce under this scheme? If every license plate is created, what is the probability that I will be issued `MER 101` as my plate? What is the probability that I will be issued either `OLE 007` or `OLE 666` as my plate?

Solution: For the first question, using the Fundamental Principle of Counting, we have $12 \times 24 \times 10 \times 10 \times 10 = 6,912,000$.

The second question assumes that all plates are available at once, and that I have an equal chance of getting any plate available. Thus, the probability of getting that plate is

$$\mathbb{P}\left[\text{MER 101}\right] = \frac{1}{6,912,000} \approx 1.45 \times 10^{-7}$$

The answer to the third question is zero, as the letter 'O' is *not allowed* in Oregon license plates. This is a good type of question to test if you know the **support** of a distribution. Be aware. ◇

Problem The blackbird

In general, let us denote having the disease by D, not having the disease by $\sim D$, testing positive by $+$ and testing negative by $-$.

Doctor researchers have just completed trials on a new clinical test, called SR-71. They concluded that the false-positive rate is $\mathbb{P}[+|\sim D] = 0.100$, and the false-negative rate is $\mathbb{P}[-|D] = 0.010$. The prevalence of the disease in society is $\mathbb{P}[D] = 0.0015$. What is the true-negative rate ($\mathbb{P}[-|\sim D]$)? What is the true-positive rate ($\mathbb{P}[+|D]$)?

Now, using Bayes' Rule (page 80), what is the probability of having the disease, given that the person tested positive ($\mathbb{P}[D|+]$)? What is the probability of having the disease, given that the person tested negative ($\mathbb{P}[D|-]$)?

Solution: The first two questions are simply trying to determine which you subtract from one. Thus, we have $\mathbb{P}[-|\sim D] = 1 - \mathbb{P}[+|\sim D] = 1 - 0.100 = 0.900$, and $\mathbb{P}[+|D] = 1 - \mathbb{P}[-|D] = 0.990$.

Now for the important question. This is a question that doctors (researchers, not clinicians) answer daily. Merely having a small false negative rate and false positive rate does not mean too much when it comes to rare diseases, such as this one ($\mathbb{P}[D] = 0.0015$). Bayes' Rule is very useful in this problem, and its ilk.

$$\begin{aligned}
\mathbb{P}[D|+] &= \frac{\mathbb{P}[+|D]\ \mathbb{P}[D]}{\mathbb{P}[+|D]\ \mathbb{P}[D] + \mathbb{P}[+|\sim D]\ \mathbb{P}[\sim D]} \\
&= \frac{(0.990) \times (0.0015)}{(0.990) \times (0.0015) + (0.100) \times (0.9985)} \approx 0.147
\end{aligned}$$

That is, you still have only a 14.7% chance of having the disease, even if you test positive for it.

$$\begin{aligned}
\mathbb{P}[D|-] &= \frac{\mathbb{P}[-|D]\ \mathbb{P}[D]}{\mathbb{P}[-|D]\ \mathbb{P}[D] + \mathbb{P}[-|\sim D]\ \mathbb{P}[\sim D]} \\
&= \frac{(0.010) \times (0.0015)}{(0.010) \times (0.0015) + (0.900) \times (0.9985)} \approx 1.669 \times 10^{-5}
\end{aligned}$$

◇

Problem Declaration of dependence

Molecular Biophysicists conduct the following experiment. Light from one of three lasers (100Å, 4000Å, and 7000Å) is shown on a DNA base pair, and the event of disintegration of the base pair is measured. A laser is categorized according to its light output. The results of the experiments are provided in the following table.

	100Å	4000Å	7000Å
Disintegrations	2	6	3
Non-disintegrations	8	24	12

Are the effects of the lasers independent of the disintegration of the DNA base pairs? (**Note**: This is the opposite question as asking if the lasers affected the disintegrations differently.)

Solution: We know if A and B are independent events, then $\mathbb{P}[A \cap B] = \mathbb{P}[A]\ \mathbb{P}[B]$. So, let us test this:

As we can start anywhere, let us start in the upper-left cell. Let A be the event of disintegrating. So, $\mathbb{P}[A] = \frac{11}{55} = 0.20$. Now, Let B be the event of having a 100Ålaser light shown on it. So, $\mathbb{P}[B] = \frac{10}{55}$. A simple calculation shows $\mathbb{P}[A \cap B] = \frac{2}{55} = \frac{11}{55} \times \frac{10}{55}$. So, this cell checks out.

On to the second cell (remember, you must test it for all six cells). Let A be as before, but B is the even of having a 4000Ålaser shown on it: $\mathbb{P}[B] = \frac{30}{55}$. So, $\mathbb{P}[A \cap B] = \frac{6}{55} = \frac{11}{55}\frac{30}{55}$, which is also true.

...I will let you check the rest. As a side note, there is a theorem that says we are done in this example, but the theorem is not in your book, so play it safe and check all six cells.

So, we have determined that the events are independent. ◇

Problem Risk it all

Rolling a fair die is a Bernoulli trial, depending on how we determine the definition of success. In the game of Risk, the winner of a battle is determined by rolling dice. If the challenger rolls *higher* numbers, he or she wins the battle. Otherwise, the challenger loses. Julie challenges Tom to a battle. Julie decides to roll one die; Tom, also. What is the probability that Julie wins the battle?

Solution: Perhaps the most straight-forward way of solving this is to list out all pairs of possible outcomes of the two each rolling a die, then determine in which ones Julie wins. In the following array, Julie is along the side and Tom is along the top. **Bold** entries are wins by Julie:

1, 1	1, 2	1, 3	1, 4	1, 5	1, 6
2,1	2, 2	2, 3	2, 4	2, 5	2, 6
3,1	**3,2**	3, 3	3, 4	3, 5	3, 6
4,1	**4,2**	**4,3**	4, 4	4, 5	4, 6
5,1	**5,2**	**5,3**	**5,4**	5, 5	5, 6
6,1	**6,2**	**6,3**	**6,4**	**6,5**	6, 6

Thus, there are 15 ways Julie wins out of 36 total possible outcomes. Thus, the probability Julie wins is $\mathbb{P}\left[W\right] = \frac{15}{36} = \frac{5}{12} \approx 0.4167$.

Another way is to note that of the 36 possible outcomes, 6 are ties (losses for Julie). The remaining 30 are evenly split between Julie wins (15) and Tom wins. Thus, the probability of Julie winning is that $\frac{15}{36}$, as before. ◇

Problem Under a little pressure

Let us suppose that air enters a compressor at pressure $P_1 = 10.1 \pm 0.3$ MPa and leaves the compressor at $P_2 = 20.1 \pm 0.3$ MPa. From our knowledge of fluids, we know that the intermediate pressure in the line is the geometric mean, $P_3 = \sqrt{P_1 P_2}$. Find the intermediate pressure, including the appropriate uncertainty. Which would provide a greater reduction in uncertainty in your estimate of P_3: reducing the uncertainty in your measurement of P_1 to 0.2 MPa or reducing the uncertainty in your measurement of P_2 to 0.2 MPa?

Solution: Here is an example of when the geometric means is appropriate or useful. We need to remember that functions of uncertainties need the (partial) derivative.

The best guess of the intermediate pressure is $P_3 = \sqrt{P_1\ P_2} = \sqrt{10.1 \times 20.1} \approx$ 203.01. As for the uncertainties, we use the partial derivatives evaluated at the given values.

$$
\begin{aligned}
P_3 &= (P_1 \times P_2)^{1/2} \\
\frac{\partial P_3}{\partial P_1} &= \frac{1}{2}\left(\frac{P_2}{P_1}\right)^{1/2} \\
\frac{\partial P_3}{\partial P_2} &= \frac{1}{2}\left(\frac{P_1}{P_2}\right)^{1/2} \\
\Rightarrow \qquad (\Delta P_3)^2 &= (\Delta P_1)^2 \frac{1}{2}\left(\frac{P_2}{P_1}\right)^{1/2} + (\Delta P_2)^2 \frac{1}{2}\left(\frac{P_1}{P_2}\right)^{1/2} \\
&= (0.3)^2 \frac{1}{2}\left(\frac{20.1}{10.1}\right)^{1/2} + (0.3)^2 \frac{1}{2}\left(\frac{10.1}{20.1}\right)^{1/2} \\
&\approx 0.72425\ldots
\end{aligned}
$$

Thus, the solution is $P_3 = 203 \pm 0.7$. ◇

Problem Einsteinium

A radioactive isotope of Einsteinium, ^{253}Es, emits an alpha particle from time to time according an exponential distribution with rate parameter 0.049 day^{-1}. What is the probability density function (pdf) for this isotope? What is its cumulative distribution function (cdf)? What is the expected time between emissions? What is the variance of the time between emissions? What is the probability that this isotope will *not* emit an alpha particle in the next 40 days? What is the probability that this isotope will not emit an alpha particle in the next 40 days *given that it has not in the last 120 days*?

Solution: As we are told this is an exponential distribution, and we are given $\lambda = 0.049$, the pdf is $f_T(t) = 0.049e^{-0.049t}$. The cdf is just the integral of the pdf:

$$\begin{aligned} F_T(t) &= \int_{-\infty}^{t} 0.049e^{-0.049x} \, \mathrm{d}x \\ &= \int_{0}^{t} 0.049e^{-0.049x} \, \mathrm{d}x \\ &= 1 - e^{-0.049x} \end{aligned}$$

We can either calculate the mean and variance if we do not remember the formulae, or we can apply the formulae.

$$\begin{aligned} \mathbb{E}[T] &= \frac{1}{0.049} \approx 20.4 \\ \mathbb{V}[T] &= \frac{1}{0.049^2} \approx 416 \end{aligned}$$

As the next question asks for a probability, we can either integrate the pdf, or we can use the cdf:

$$\mathbb{P}[T > 40] = 1 - \mathbb{P}[T \leq 40] = \exp[-0.049 \times 40] \approx 0.141$$

Because of the memoryless property of the exponential distribution, we know the answer is also 0.141. ◇

Problem Fits you to a T (cell)

In a medical test, Human T-Cells are suspended in a medium. After 145 minutes, samples are taken from the medium, and the life-status of the T-Cells, which have not increased in size or number, is determined. The researcher originally placed 1.50×10^6 T-Cells in 1.510L of media. She then stirred the medium to achieve uniform density of T-Cells in the medium. At the end of the experiment, 145 minutes later, she removed 1.000mL of medium. How many T-Cells would we expect her to have withdrawn along with the medium? What is the probability that she removed none?

Solution: This problem screams Poisson to me. Why?

The number of T-Cells removed will be distributed Poisson, $N \sim \mathcal{P}(\lambda = \frac{1.50\times10^6}{1510})$. Here, $\lambda \approx 993$. Thus, the number we would expect her to remove would be $\mathbb{E}[N] = 993$. Why are we dividing by 1510 and not 1.510?

Calculating the probability that she removed none is a straight-forward application of the pmf:

$$f_N(n = 0) = \frac{e^{-993} 993^0}{0!} \approx 0$$

In other words, if she did not remove *any* T-Cells, she would have to conclude that there was something wrong with her apparatus or with her assumptions.

◇

Problem More positioning tubes

In a certain production process, the lengths of manufactured positioning tubes are independently distributed normally with mean 12.10cm and standard deviation 1.00cm. In one shift, 5000 are produced. You select one at random and measure its length. What is the probability that it measure less than 11.99cm? You select another at random from the remaining 4999. Given that the first one measured 12.00cm, what is the probability that this second tube measures less than 11.99cm in length?

Solution: We are given $L \sim \mathcal{N}(\mu = 12.10cm, \sigma = 1.00)$.

Now, if we convert all measurements to z-scores, we can look up in the table the probabilities (as there is no explicit formula for the cdf of the Normal distribution). And so, 11.99cm in z-units is $\frac{11.99-12.10}{1.00} \approx -0.11$. Looking in our tables, we see $\mathbb{P}[Z < -0.11] = 0.456$. We could also have used our calculators. The TI calculators would use '`normalcdf(-99999999, -0.11)`'. And yes, I will provide a table if problems like this are on the test.

As we are told the positioning tubes are independently distributed, the answer to this will also be 0.456. ◇

Problem Take a chance?

A certain professor makes a deal with his students on an upcoming True-False examination. If any student answers all 20 True-False questions *incorrectly*, then the student receives a score of 150% on the examination (30/20). Otherwise, the student receives the number correct out of 20. Let us assume that a certain student's knowledge about Question i is independent of his or her knowledge about Question j. Under this assumption, we are looking at a series of 20 independent Bernoulli trials, each with probability of success being p. Student Abel believes he has a 50% chance of answering all of the questions correctly. What chance does he have of answering a single question correctly? Student Bourbaki thinks he has a 99.5% of answering each question correctly, what is his probability of answering all 20 questions correctly?

Solution: Well, Student Abel should not take the bet. He would be better off answering correctly than incorrectly. But, this is not the question. The probability that he answers all of the questions correctly is p^{20}. He thinks he has a 50% percent chance of answering correctly, thus $p^{20} = 0.500 \Rightarrow p = (0.500)^{1/20} \approx 0.966$.

Student Bourbaki has a $0.995^{20} \approx 0.905$ of answer all questions correctly. ◇

Problem Missed the bus

Let us assume that bus arrival time is exponentially distributed, with a rate parameter of 0.05 (min^{-1}). You arrive at the bus stop, just seeing a bus leaving as you approach the stop. How long should you expect to wait for the next bus? Let us make this problem a bit more interesting. Let us suppose that there are five bus lines that service your stop. All of the busses look identical from the rear. Thus, you do not know which bus you just missed. Now, how long should you expect to wait for *your* bus?

Solution:

We are given $T \sim Exp(\lambda = 0.05)$. We know, for an exponential distribution, $\mathbb{E}[T] = \frac{1}{\lambda} = 20$. Thus, you should expect to wait 20 minutes for your bus.

The answer to the second question is also 20 minutes. Why?

◇

Problem Short-lived connections

The Exponential distribution is often used in survival time analysis. It has one parameter, λ. It has been discovered that the connector component of a circuit board has a lifetime in minutes, T, that is distributed as $T \sim Exp(\lambda = 0.01)$. How long should we expect the connector to last?

Solution:

In an exponential distribution, $\mathbb{E}[T] = \frac{1}{\lambda} = 100$. Thus, we should expect the connector to last 100 minutes.

◇

Problem A shot in the dark

A flashlight requires two functional batteries to operate correctly. You go spelunking in the mountains of southwestern Oklahoma searching for peace and relaxation. Spelunking requires a flashlight. You pack four (4) batteries with your flashlight, thinking that will be sufficient for your exploration. The lifetime (in minutes) of these batteries are independently distributed $T \sim Exp(\lambda = 0.01)$. Assuming you are able to change out dead batteries instantly, how long will your flashlight work?

Solution: We are given $T \sim Exp(\lambda = 0.01)$. As this is an exponential distribution, it has that wonderful memoryless property. If the first battery dies in 10 minutes, the second battery still has an expected lifetime of 20 minutes, as do the other two batteries. In fact, because of the memoryless property, this problem reduces to trying to find the expected time for three batteries to die.

Thus, the expected number of minutes you can expect to go with a light is $3 \times \frac{1}{0.01} = 300$. And so, you can expect to explore the caves for 5 hours before being left in the dark. $\diamond$