**Problem 1:**

1. False, a random variable that is discrete has PMF (Probability Mass Function). It is a statistical function that describes the probabilities of discrete random variables assigning a probability to each possible outcome.
2. True, an individual probably value in a probability mass function must be between 0 and 1 inclusive. However, the sum of all the probabilities must equal 1.
3. False, the probability density function is the describing the likelihood of a continuous random variable taking on a particular value or range of values. The set of all possible realizations (outcomes) is the sample space.
4. False, the expected value E[X] of a discrete random variable is no part of its support Rx​, which is the set of values that the random variable can take with nonzero probability.
5. True, you can imagine a continuous random variable X representing height of a random person. X maps people to real number (their height). The continuous aspect is that the mapped height could be between a range with an infinite number of measured heights.
6. False, the statement is reversed.
7. **False**, Bayes posterior is a probability density function over the parameter space *θ*∈RD. By definition, a PDF must integrate to 1 over its entire domain for any observed data y. The statement given says that the PDF doesn’t integrate to 1.
8. True, the is because in a continuous random variable there are an infinite number of possibilities.

**Problem 2:**

1. My current expectation is that the market will go up. Personally, I would not invest in this market. If we do it based on expected value, we find that the resulting value is -1.2. We can calculate this by using the discrete values that we are given. There are only two probabilities with their given outcome.

One limitation of expectations is that it doesn’t account for risk. You could most likely make a return on your investment, however if you end up taking a loss there is no guarantee of recuperation of your losses in the future. We also do not know have variable the chances are.

**Problem 3:**

1. Prove
2. Prove
3. Affine transformation can be used

**Problem 4**

1. is the probability of the hypothesis before any evidence considered. Since the hypothesis is that the first islander told the truth, we can say the prior is 1/3.
2. is the probability of new evidence under all possible hypotheses. Since it is stated that the second islander is always saying yes, it means we derive 2 outcomes. The first is that the first islander tells the truth, which in turn implies that the second islander’s “yes” is truthful. This gives us 1/3\*1/3 since the likelihood of an islander telling the truth is 1/3. The second is that the first islander lies, which in turn implies that the second islander’s “yes” is a lie. This gives us 2/3\*2/3 since the likelihood of an islander lying is 2/3.
3. is the probability that the evidence is true, given that the hypothesis is true. Thus, we can ignore the 2/3 probability that the second islander lies since the hypothesis is true.

**Problem 5**

\*Note: code can be found in homework\_2/Q5.py in the repo

*The Case of the Missing Blueprint*

Professor Alden had been working on groundbreaking energy-efficient power system design, but one morning, the key blueprint vanished from his secure lab. Only three people had access:

* Dr. Evelyn Carter – A meticulous researcher who was last seen in the lab at 7:00 PM.
* Sarah Lin – A lab technician who left at 6:15 PM.

The only evidence available: The lab’s security system detected unauthorized access at 8:00 PM.

*Prior Distribution*

Before considering the security log, we assign prior probabilities based on access and potential motivation:

* Dr. Evelyn Carter: 65% (She had access and expertise.)
* Sarah Lin: 35% (She had access but no motive, she is paid well at her job.)

*Likelihood (Security Log Evidence)*

Suspect Items with fingerprints

* USB drive left on the computer
  + 25% for Dr. Evelyn Carter
  + 35% for Sarah Lin
* Access Panel
  + 10% for Dr. Evelyn Carter
  + 15% for Sarah Lin
* Computer Keyboard
  + 65% for Dr. Evelyn Carter
  + 50% for Sarah Lin

*Results:*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Samples | E-A | E-C | E-U | S-A | S-C | S-U |
| 100 | 0.05 | 0.42 | 0.2 | 0.05 | 0.19 | 0.09 |
| 1,000 | 0.064 | 0.434 | 0.173 | 0.048 | 0.172 | 0.109 |
| 10,000 | 0.0611 | 0.4259 | 0.1592 | 0.0525 | 0.178 | 0.1233 |
| 100,000 | 0.06434 | 0.4213 | 0.16249 | 0.05315 | 0.17951 | 0.12281 |
| 1,000,000 | 0.06503 | 0.42243 | 0.16264 | 0.05212 | 0.17512 | 0.12266 |
| Baseline | 0.065 | 0.4225 | 0.1625 | 0.0525 | 0.175 | 0.1225 |

N should be a million samples as the values are within the ten thousandths of error.

**Problem 6:**

Answer can be found in homework\_2/Q6.ipynb within the repo

**Problem 7:**

Answer can be found in homework\_2/Q7.ipynb within the repo