***Problem 1a*:**

Assume that the sensitivity and specificity of a routine screening polygraph are about 80% and 85%, respectively. That is, the probability that the polygraph report will indicate a concern is about 80% if the individual has committed a security violation, and the probability that the exam will not indicate a concern is about 85% if the individual has not committed a security violation.

a. Assume that about 1 in 1000 individuals in a given organization has committed a security violation. What is the posterior probability that an individual whose polygraph report indicates a concern has committed a security violation? Comment on the implications of these results for the use of routine screening polygraphs for individuals working positions requiring security clearances.

b. Suppose an item has been stolen. Investigators have concluded that it was an inside job. A suspect has been identified and is given a polygraph exam. The polygraph report indicates a concern on the topic of the theft. Assume the same sensitivity and specificity of the polygraph as in problem 1, but assume a prior probability of 25% that the individual stole the item. What is the posterior probability that this individual committed the theft? Explain the difference between this result and the result from Problem 1a.

***Solution to Part a*:**

For a person in the given organization, we have the following probabilities for the polygraph outcome:

| Condition | Concern | No concern |
|---|---|---|
| Guilty | 0.80 | 0.20 |
| Innocent | 0.15 | 0.85 |

Prior Probability: P(Guilty) = 0.001 and P(Innocent)=0.999

We can calculate the posterior probability that the individual is guilty as follows:

First, we calculate the probability of the test indicating a concern (this will be the denominator of Bayes Rule):[1]

P(Concern) = P(Concern |Guilty) * P(Guilty)+P(Concern | Innocent) * P(Innocent)
= 0.80*0.001 + 0.15*0.999 = 0.15065

Then, we calculate the posterior probability that the individual is guilty by applying Bayes rule:

P(Guilty | Concern) = P(Guilty and Concern)/P(Concern)
= P(Concern |Guilty) *P(Guilty)/ P(Concern)

[1] Note: In all problems, I carry enough significant figures to allow students to duplicate the calculations. Results should only be reported to as many significant figures as the inputs.

= 0.80*0.001/0.15065 = 0.0053

The posterior probability that an individual showing a concern is guilty is 0.0053.

*Discussion*: The probability of a security violation increases from 1 in 1,000 to 5 in 1,000 when the polygraph shows a concern. The probability increases by a factor of 5, but because the probability started out so small, the result is still a very small probability. Although only a small proportion of the innocent employees show a concern on the polygraph, and most of the guilty employees show a concern on the polygraph, there are so many more innocent employees in relation to the guilty employees that most of the positives will be false positives.

Students who have had prior experience with this kind of problem are not surprised at the low posterior probability of Guilty. Others are often surprised at how low the posterior probability is that the person is Guilty.

I have found that a good way to explain this problem to people who have trouble with it is to ask them to imagine a large organization with 20,000 employees. We would expect 20 (1 in 1000) to be guilty of a security violation. Of these 20, we would expect 16 to show a concern and 4 not to show a concern. So the polygraph does catch most of the guilty individuals. But 19,980 of the employees did not commit a security violation. Of these, we expect 85%, or 16,983, not to show a concern, and 15%, or 2,997, to show a concern. Even though 2,997 is only 15% of the Innocent employees, it is a much larger number than the 16 guilty individuals who showed a concern. That is, because there are so many more innocent than guilty people, there are many more false positives than true positives.

***Solution to Part b*:**

The sensitivity and specificity of the polygraph are the same as above:

| Condition | Concern | No concern |
|---|---|---|
| Guilty | 0.80 | 0.20 |
| Innocent | 0.15 | 0.85 |

But the prior probability is different:
Prior Probability: P(Guilty) = 0.25 and P(not Guilty) = 0.75

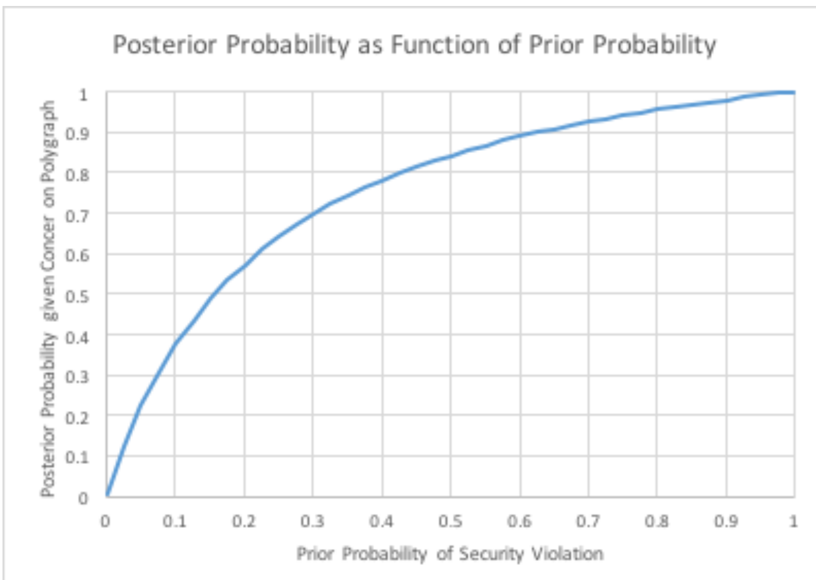
Again, we compute the probability of the polygraph indicating a concern:
P(Concern) = P(Concern |Guilty) * P(Guilty)+P(Concern | Innocent) * P(Innocent)
= 0.80*0.25 + 0.15*0.75 = 0.3125

The Posterior probability that the individual is guilty is:
P(Guilty | Concern) = P(Guilty and Concern)/P(Concern)
= P(Concern | Guilty) *P(Guilty)/ P(Concern)
= 0.80*0.25/0.3125 = 0.64

The posterior probability that an individual showing a concern is guilty is 0.64.

*Discussion*: This is called a *specific incident* situation, in contrast to the *screening* situation described in Part a. Compared to the screening case, the posterior probability of a security violation in the specific incident case is much higher. This is because the base rate is much higher than in the previous problem. Thus, there are more guilty individuals, and therefore we expect many more true positives and many fewer false positives than in the previous problem.

A useful visualization of the role of the prior probability is to plot the posterior probability as a function of the prior probability. For prior probabilities near zero, a polygraph showing a concern increases the probability of guilt, but it remains very low. The polygraph can be useful in specific incident settings, especially in conjunction with other evidence. However, it is not a strong enough source of evidence to be used on its own, and will yield too many false positives to be very useful in a screening situation.




***Problem 2*:**
Consider a situation in which an individual may have committed some kind of crime or security violation. Suppose the organization must decide whether to administer some kind of sanction (arrest; remove security clearance; reprimand). Suppose we can obtain evidence that has 80% sensitivity and 85% specificity. Find the policy that minimizes expected loss. In problems such as this, it is more natural to think in terms of losses than utilities. Assume the following losses:

- Do not administer sanction; individual is innocent loss = 0
- Administer sanction; individual is guilty loss = 1
- Administer sanction; individual is innocent loss = 10
- Do not administer sanction; individual is guilty loss = 100

For the decision of whether to administer sanctions to an individual who may have considered a security violation, find the range of prior probabilities for which considering the evidence results in lower expected loss than ignoring or not collecting the evidence. Discuss your results.

***Solution to Problem 2*:**

As before, we have the following sensitivity and specificity:

| Condition | positive | negative |
|---|---|---|
| Guilty | 0.80 | 0.20 |
| Innocent | 0.15 | 0.85 |

To calculate EVSI, we do the following:

1. Calculate the expected loss if we do not do a polygraph. To do this, we use the prior probability to calculate the expected loss if we administer sanctions and if we don't administer sanctions. The minimum of these is the expected loss of not doing a polygraph.

2. Calculate the expected loss if we do a polygraph.

3. Now, the EVSI is the how much doing the polygraph reduces our loss:
   EVSI = E(L|no poly) - E(L|poly)

We follow these steps, allowing the prior probability to vary:
Prior Probability: P(Guilty) = $p$ and P( Innocent ) = 1- $p$

Step 1. If we do not do the polygraph, we cannot use the polygraph result, so we have two choices: administer sanctions or do not administer sanctions.

The expected loss of administering sanctions is:
E(L|S) = 1*P(Guilty) + 10* P( Innocent) = 10 – 9$p$

The expected loss of not administering sanctions me is:
E(L|N) = 100*P(Guilty) + 0* P(Innocent) = 100*P(Guilty) = 100$p$

Step 2. Now consider the strategy of doing the polygraph and administering sanctions to those who show a concern. We get the following losses (compare with page 22 of the Unit 1 notes):

| Patient status | Test result | Action | Loss (L) | Probability |
|---|---|---|---|---|
| Innocent | No concern | no sanction | 0 | (1-$p$)*0.85 |
| Innocent | Concern | sanction | 10 | (1-$p$)*0.15 |
| Guilty | No concern | no sanction | 100 | $p$*0.20 |
| Guilty | Concern | sanction | 1 | $p$*0.80 |

The total expected loss of this strategy is:
E(L|T) = 0 + 10*0.15*(1-$p$) + 100*0.20*$p$ + 1*0.80$p$ = 1.5+19.3$p$

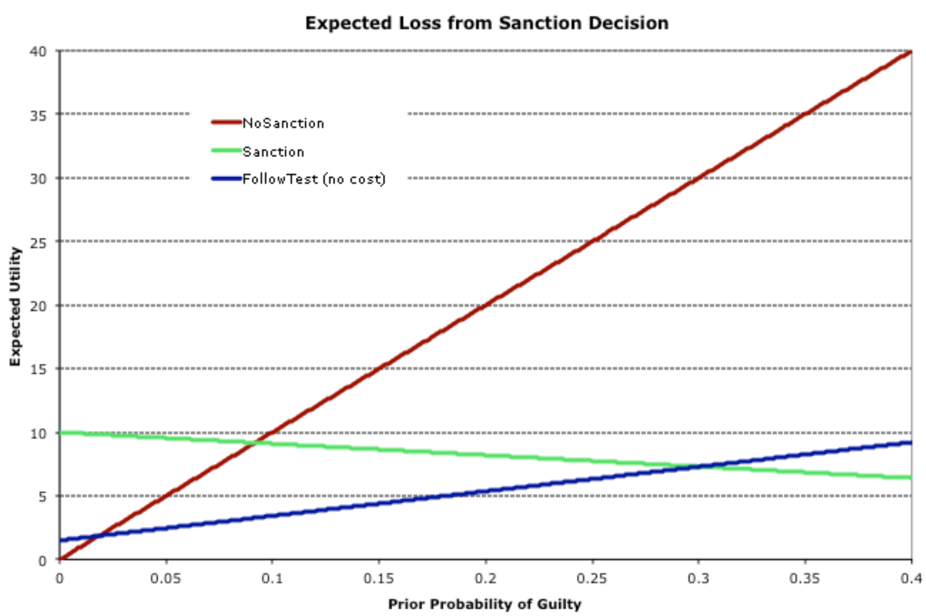
The graph below shows the plots of expected loss for these three policies. I show results only for probabilities up to 0.4 in order to bet a better view of the results for small probabilities.

With a little algebra, we can find the intersections between the lines and the regions of optimality. These are:

| | |
|---|---|
| $0 \le p < 0.019$ | Optimal policy is N, do not administer sanctions |
| $0.0186 < p < 0.30$ | Optimal policy is T, do polygraph and administer sanctions if polygraph shows concern. |
| $0.30 < p \le 1$ | Optimal policy is S, administer sanctions |

Therefore, administering the polygraph has lower expected loss than always or never administering sanctions for prior probabilities in the range $0.0186 < p < 0.30$.

For Problem 1a, the prior probability of 0.001 is less than 0.019, so it is not optimal to administer the polygraph. For Problem 1b, the prior probability of 0.25 is in the range of values for which it is optimal to administer the polygraph. This analysis confirms that the polygraph has little value in screening situations, but may be more useful for a specific incident situation.




***Problem 3*:**
Experimenters performed a study in which a polygraph was administered to subjects in a simulated theft scenario. Before being polygraphed, subjects waited in a room where $50 was left on a table in open view. Some subjects were instructed to take the money, while others were asked to leave it there. Both groups were asked to tell the polygrapher that they did not take the money, and were given a monetary reward if the polygrapher

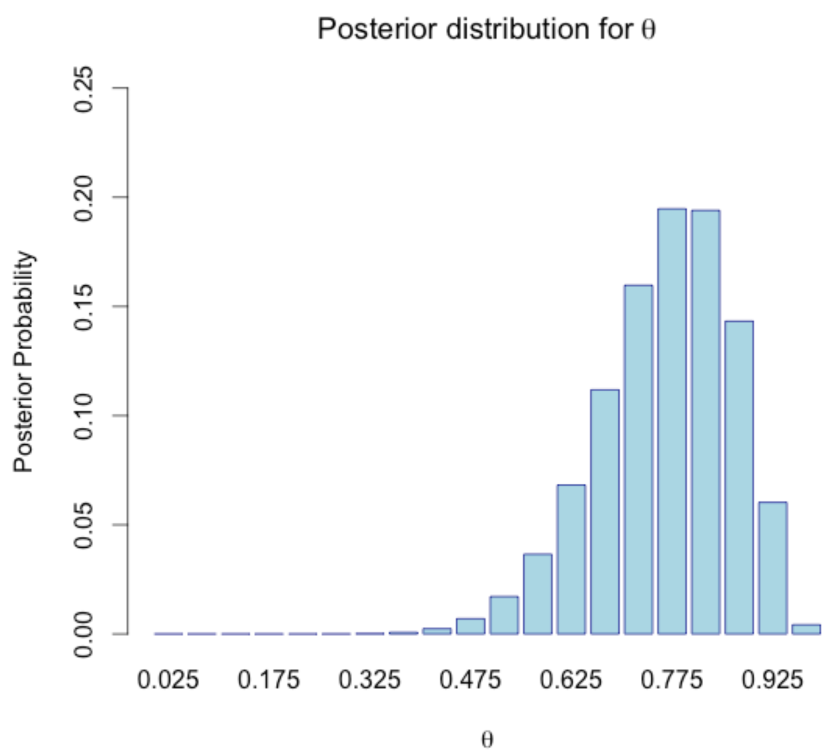
believed them. Assume the polygraph sensitivity Θ has 20 equally spaced possible values ranging from 0.025 to 0.975. Assume that polygraph results for people who took the $50 are independent and identically distributed, with probability Θ of the polygraph showing a concern. Assume that all 20 values of Θ are equally likely *a priori*. Suppose the polygrapher found a concern for 12 of the 15 subjects who took the $50. Use R to make a bar plot of the posterior distribution for Θ given the results of the study.

***Solution to Problem 3***

We can use the R code from the Unit 1 example to do this problem. The R file `Unit1_Example.R` was posted on Blackboard. Please see modified code below. The only changes that needed to be made were:

- Change the comments to refer to the polygraph example instead of the disease example
- Change lines 16 and 17 to 15 observations and 12 showing a concern
- Remove the lines plotting the prior pmf (this is not strictly necessary)
- Change the limits on the y axis to `ylim=c(0,0.25)`(or remove the ylim parameter and let the plot use the default y axes) so the bars do not run off the page.

Notice that the posterior distribution favors values near the observed frequency of 80%.



R Code:

```
# Assignment 1 Problem 3
# Calculate and plot posterior distribution
# for polygraph sensitivity:  Discretized uniform prior;
# sample of 15 cases, 12 showing a concern

# Assume theta can take on one of 20 evenly spaced values
# between 0.025 and 0.975
theta <- seq(length=20,from=0.025,to=0.975)

# Uniform prior distribution: 20 possibe values of theta
# are equally likely
priorDist <- array(1/20, 20)

# Calculate the posterior distribution for sample of
# 15 cases, 12 showing a concern
numobs=15    # Number of observations
numd=12      # Number showing a concern
postDist <- priorDist * theta^numd*(1-theta)^(numobs-numd) # prior times likelihood
postDist <- postDist/sum(postDist)

# Plot the posterior distribution
barplot(postDist,main=expression("Posterior distribution for"~theta),
xlab=expression(theta), ylab="Posterior Probability",names.arg=theta,
border="darkblue", col="lightblue",ylim=c(0,0.25))
```