Solutions to Assignment 12

1. A box contains 3 coins C_1, C_2 and C_3 with probability of head as 1/4, 1/2 and 3/4 respectively. You pick a coin out of the box but you don't know which one it is so you assign a *a priori* probability of 1/3 for each choice of coin. You flip the coin 120 times and get 70 heads. What is the *a posteriori* probability that you will assign to each choice of coin.

Solution: Let X be the random variable that counts the number of heads and C be the random variable that indicates the choice of coin. We have the following calculations

$$\begin{aligned} P(X = 70|C = C_1) &= \frac{3^{50}}{4^{120}} \\ P(X = 70|C = C_2) &= \frac{1}{2^{120}} \\ P(X = 70|C = C_3) &= \frac{3^{70}}{4^{120}} \end{aligned}$$

We combine this with $P(C = C_i) = 1/3$ to get

$$P(X = 70) = (1/3)\frac{3^{50} + 2^{120} + 3^{70}}{4^{120}}$$

We now calculate using Bayes rule

$$P(C = C_i|X = 70) = \frac{P(X = 70|C = C_i)P(C = C_i)}{P(X = 70)}$$

This gives us

$$\begin{aligned} P(C = C_1|X = 70) &= \frac{3^{50}}{3^{50} + 2^{120} + 3^{70}} \\ P(C = C_2|X = 70) &= \frac{2^{120}}{3^{50} + 2^{120} + 3^{70}} \\ P(C = C_3|X = 70) &= \frac{3^{70}}{3^{50} + 2^{120} + 3^{70}} \end{aligned}$$

Since the event $X = 70$ has already taken place, this gives the new probability estimates for the choice of the coin.

2. In order to check a coin for bias, it is flipped a large number of times. In 10000 flips it shows heads 4844 times. Calculate the likelihood ratio between the maximum likelihood value for p and 0.5. Would you take the coin as unbiased or not? Justify your answer.

Solution: The maximum likelihood value for p is .4844. We calculate the log-likelihood as

$$4844 \log(p) + (10000 - 4844) \log(1 - p)$$

Note that this is the same as $10000 \cdot S_p$ where

$$S_p = p \log(p) + (1 - p) \log(1 - p)$$

On the other hand, if the coin was unbiased, the log-likelihood of the event is

$$4844 \log(0.5) + (10000 - 4844) \log(0.5) = 10000 \log(0.5)$$

Thus, the difference between the log-likelihoods is

$$10000(S_p - \log(0.5)$$

The likelihood ratio is $\exp(10000(S_p - \log(0.5)))$. We can calculate this to be approximately 130.

To understand this ratio we compare it with the ratios for a completely biased coin. In that case, the ratio turns out to be 2^k for k flips of the coin. Thus, taking logarithm to the base 2. We have

$$k' = 10000(S_p/\log(2) - \log(0.5)/\log(2)) = 10000(s_p + 1)$$

where $s_p = p \log_2(p) + (1 - p) \log_2(1 - p)$. In our case this is roughly 7.

Thus, calling our coin biased is like deciding that a coin shows only heads (or is extremely likely to show heads) after seeing this other coin getting 7 heads in a row.

3. In order to check a coin for bias, it was flipped a large number of times. With 10000 flips it turns up 4913 heads.
   (a) What is an assertion that we can make with 70% confidence?
   (b) What is an assertion that we can make with 95% confidence?
   (c) What is an assertion that we can make with 99% confidence?

Solution: Let X_i denotes the random variable that gives 1 for Head and 0 for Tail on the i-th toss. Suppose there are r heads out of n tosses. The sample mean m of the X_i's is r/n. The sample variance can also be calculated as

$$\frac{r \cdot (1 - r/n)^2 + (n - r)(0 - r/n)^2}{n - 1}$$

This can be simplified to $\frac{rn - r^2}{n(n-1)}$. Its square root, the sample standard deviation σ can be written as $m\sqrt{t}$ where $t = \frac{n(n-r)}{r(n-1)}$. Now $t - 1 = \frac{n^2 - 2nr + r}{r(n-1)}$ is small in many cases and we can use the Binomial theorem to approximate *sqrtt* as

$$\sqrt{t} \simeq 1 + \frac{t-1}{2} = \frac{t+1}{2}$$

The latter is $(t+1)/2 = \frac{n^2-r}{2r(n-1)}$ which is easily calculated. Since n is large in most cases, we can also approximate it as $\frac{n^2-r}{2rn} = 1/2m - 1/2n$. To summarise, we can approxiate σ as $1/2 - m/2n = 1/2(1 - r/n^2)$. **Warning**: This approximation is only good for large n and when $t-1$ is small.

In our case $m = r/n = .4913$, so $\sigma \simeq 1/2(1 - 4913/20000) \simeq 0.4999$. We thus have $s = \sigma/sqrtn \simeq .0049$. The intervals we have to look at are $[m - ks, m + ks]$ for $k = 1, 2, 3$.

Now for $k = 1$, we see that .5 (unbiased coin) does not lie in this interval. Thus we can assert with 70% confidence that the coin is biased.

One the other hand for $k = 2$ (and hence for $k = 3!$) we see that .5 lies in this interval. Thus we can assert with 99% confidence that the hypothesis that the coin is unbiased cannot be rejected.

Think about why the above two statements do not entirely contradict each other!