COMP30690 Information Theory: First Assignment

Patrick Keogh 19321326

Exercise 1: Probability Theory

Let *F* denote choosing the fair coin.

Let *B* denote choosing the biased coin.

Let *H* denote observing n consecutive heads.

The probability that you chose the fair coin given that you observed *n* consecutive heads is P(F | H)

We can compute this as:

$$P(F \mid H) = \frac{P(H \mid F)\,P(F)}{P(H)}$$, where

- → P(F) is the prior probability of choosing the fair coin, which is 0.5,
- → P(H | F) is the probability of observing n consecutive heads given that you chose the fair coin, which is (0.5)^n,
- → P(H) is the total probability of observing n consecutive heads, considering both coin choices:
  - P(H) = P(F)P(H | F) + P(B)P(H | B) (Law of Total Probability)
  - P(B) is the prior probability of choosing the biased coin, which is also 0.5, and
  - P(H | B) is the probability of observing n consecutive heads given that you chose the biased coin, which is q^n.

Given this:

P(H) = (0.5 * (0.5)^n) + (0.5 * q^n)

Now, we can compute P(F | H):

P(F | H) = (0.5 * (0.5)^n) / [(0.5 * (0.5)^n) + (0.5 * q^n)]

Simplified:

P(F | H) = (0.5^n) / [(0.5^n) + (0.5 * q^n)]

Exercise 2: Random Variables

a)

Support:

$\{1, -1, -3\}$

PMF:

$P(X = 1) = p + (1 – p) * p * p)) = 0.48 + (1 - 0.48) * 0.48 * 0.48 = 0.599808$

$P(X = -1) = ((1 - p) * (1 - p) * p) + ((1 - p) * p * (1 - p))$

$= ((1 - 0.48) * (1 - 0.48) * 0.48) + ((1 - 0.48) * 0.48 * (1 - 0.48)) = 0.259584$

$P(X = -3) = (1 - p) * (1 - p) * (1 - p) = (1 - 0.48) * (1 - 0.48) * (1 - 0.48) = 0.140608$

$P(X > 0) = P(X = 1) = 0.599808$

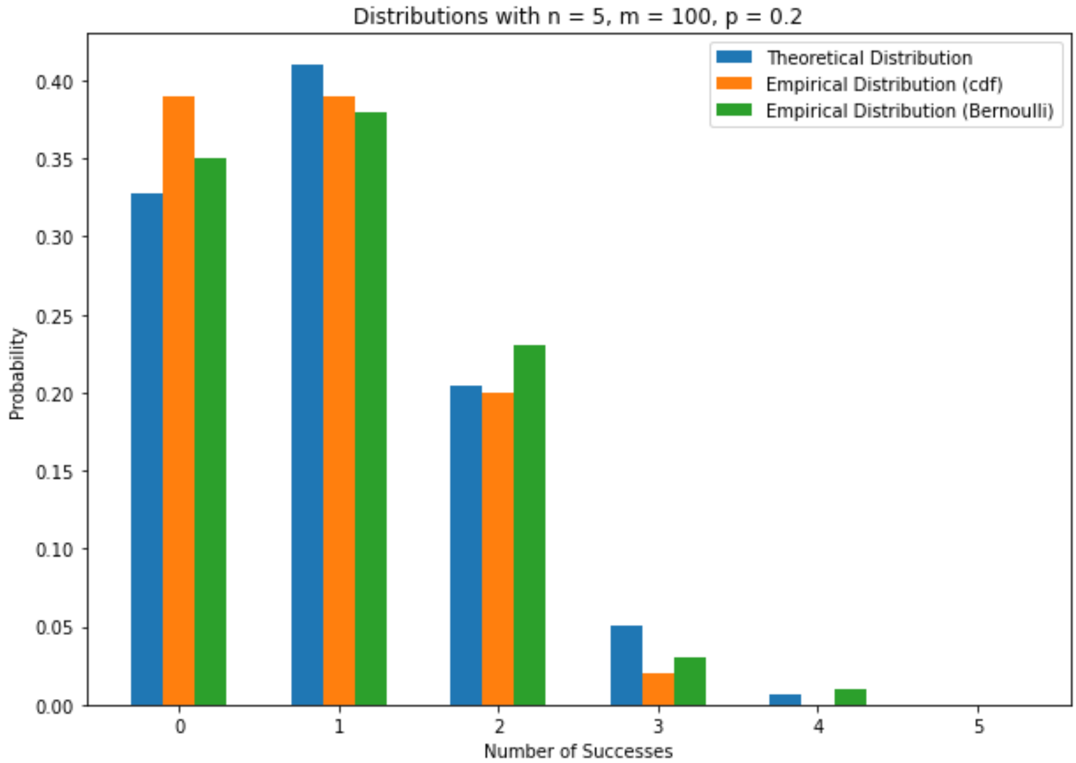
b)

Yes, as the probability of $X > 0$, i.e. making a profit, is greater than 0.5, meaning you are more likely to profit than not. It is also greater than the chance of profiting based on just one toss, even if you pick tails (0.52). However, this strategy is still far from infallible.

Exercise 3: Generation of Random Variables

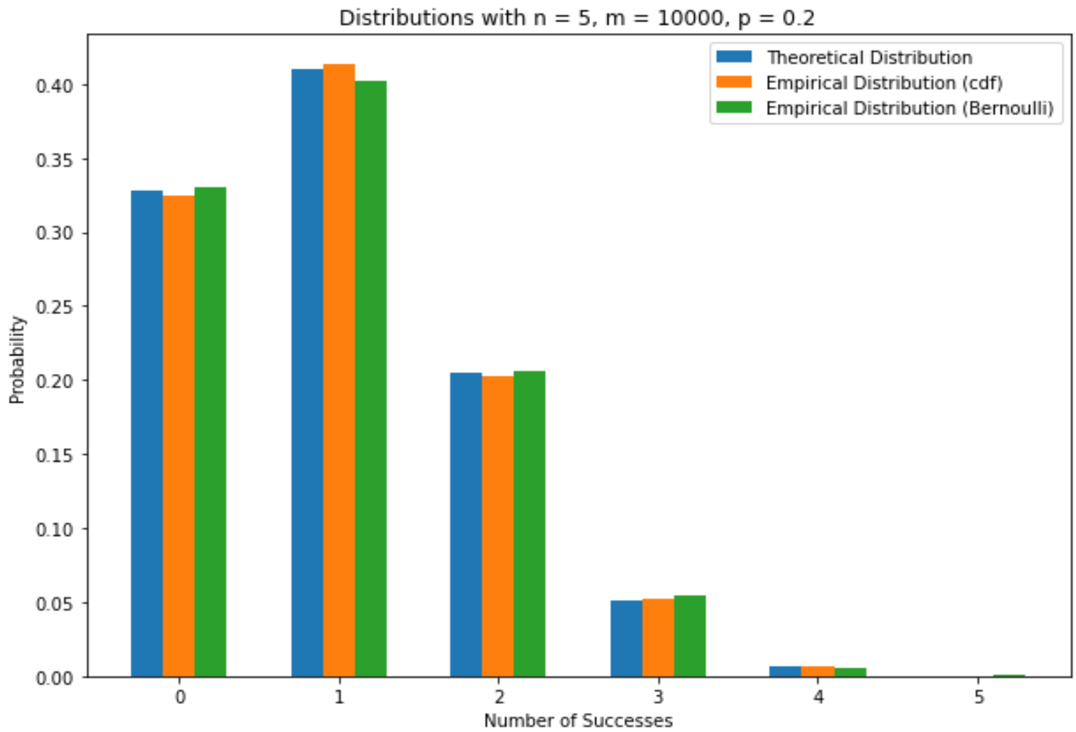
Please use the Jupyter notebook to run the provided Go source code

1) $m = 10^2$



Empirical Distribution vs Theoretical for $m = 10^2$

2) $m = 10^4$



Empirical Distribution vs Theoretical for $m = 10^4$

Analysis:

As would be expected from the Law of Large Numbers (since the trials are independent and identically distributed (i.i.d.)), as the sample size increase in the experiment increases, the empirical distribution of a random variable approaches the theoretical distribution,), due to a decrease in randomness or sampling error.

This convergence is also underlined by the Central Limit Theorem, which states that the distribution of the sum (or average) of a large number of i.i.d. random variables, will be approximately normally distributed.

We can also see that the Bernoulli method is closer to the theoretical values for a lower number of successes than the CDF method but the opposite for more successes. This may be due to the accumulation of rounding errors in the computation of multiple Bernoulli trials becoming more pronounced when the number of successes is higher, as the method requires more random number generations and more arithmetic operations, each of which introduces potential for a small error. However, both methods still converge at higher sample sizes.

Exercise 4: Mutual Information and Entropy

a)

To derive λ, we use the relationship between mutual information and entropy:

$I(X;Y) = H(X) - H(X|Y)$

Given the symmetry of X and Y, i.e. X and Y have the same pmf:

$I(X;Y) = H(Y) - H(Y|X)$

Given λ as:

$\lambda = 1 - (H(X)/H(Y|X))$

Rearranging gives:

$\lambda = H(X) - H(Y|X)/ H(X)$

Replacing $H(X) - H(Y|X)$ with $I(X;Y)$ gives:

$\lambda = I(X;Y)/H(X)$

b)

Given $\lambda = I(X;Y)/H(X)$, prove $0 \leq \lambda \leq 1$:

Lower bound

Mutual information measures the amount of information that one random variable contains about another, and the concept of 'negative information' does not exist, so it must be non-negative so,

$(X;Y) \geq 0$

We are also told that the entropy $H(Y) > 0$, and since X and Y have the same pmf so,

$H(X) = H(Y)$

$\Rightarrow H(X) > 0$

Since the mutual information and entropy are both non-negative, there ratio is also non-negative so,

$\lambda \geq 0$

Upper bound

Mutual information is bounded above by the entropy of either variable involved,

$I(X;Y) \leq \min(H(X),H(Y))$

Since $H(X) = H(Y)$,

$I(X;Y) \leq H(X)$

As the mutual information is at most the entropy, their ratio must be less than one so,

$\lambda \leq 1$

Combining these two bounds we get:

$0 \leq \lambda \leq 1$

c)

The definition of λ is:

$\lambda = I(X;Y)/H(X)$

If λ is 0, multiplying both sides by H(X) and rearranging gives:

$I(X;Y) = 0$

The mutual information, I(X;Y), measures the amount of information one random variable contains about another. A value of 0 indicates that the random variables X and Y are independent and there is no shared information between them.

d)

The definition of λ is:

$\lambda = I(X;Y)/H(X)$

If λ is 1, multiplying both sides by H(X) and rearranging gives:

$I(X;Y) = 1$

This is the maximum possible value for mutual information, meaning that all of the entropy of X is shared with Y. In other words, X and Y are perfectly related.

$I(X;Y) = H(X) - H(Y|X)$

$H(X) = H(X) - H(Y|X)$

$H(Y|X) = 0$

As the conditional entropy H(Y|X) is 0, Y can be determined with certainty from X.