EE2340/EE5847: Information Sciences/Information Theory **2020**

Homework 1: 3rd Jan 2020

Instructor: Shashank Vatedka

Instructions: *You are encouraged to discuss and collaborate with your classmates. However, you* must explicitly *mention at the top of your submission who you collaborated with. Copying is NOT permitted, and solutions must be written independently and in your own words.*

Homeworks must be submitted on Google classroom. Please scan a copy of your handwritten assignment and upload as pdf with filename <your ID>_HW<homework no>.pdf. Example: EEB19BTECH00000_HW1.pdf.

For programming questions, submit as separate files. Please use the naming convention <your ID>_HW<homework no>_problem<problem no>.. Example: EEB19BTECH00000_HW1_problem1.c*

Exercise 1.1 (Probability refresher, nothing to submit)**.** Review the following concepts from your probability course:

- Probability mass function
- Probability density function, probability distribution function
- Random variables: continuous, discrete
- Functions of random variables
- Common random variables: Bernoulli, Binomial, Poisson, Exponential, Gaussian
- Bayes theorem
- Expectation, expected values
- Moments, moment generating function
- Markov's inequality, Chebyshev's inequality, Chernoff bounds

Exercise 1.2. Consider n independent tosses of a biased coin which lands heads with probability p.

1. Find an expression for the probability that the number of heads is exactly m.
2. Let $p = 0.3$. Write a program (C/C++/python/matlab) or otherwise compute the probability that in $n = 10$ tosses, the number of heads is at most $m = 5$. You do not have to submit this program. Just write down the answer (value of probability that you have computed).
3. Find ~~upper~~ *lower* bounds for the probability in part 2 using Markov's inequality, Chebyshev's inequality, and Chernoff bound (simplified version for Bernoulli rvs). Compare with the value you have computed in part 2.
4. Repeat (3) for $n = 1000$ and $m = 350$ (only Markov, Chebyshev and Chernoff).
5. Find an expression for the probability that out of n tosses, there are at most k heads.

Exercise 1.3 (Bernoulli random number generation)**.** Find a built-in function in the programming language of your choice (C/C++/python/matlab) which generates uniformly distributed random numbers in $[0, 1]$. For e.g., if you use the numpy package in python, this is numpy.random.rand().

Use this function to generate Bernoulli(p) random variables. Your program should have a global variable `num_trials`, and it should print `num_trials` many independent random numbers.

Exercise 1.4 (Random number generation)**.** Show that if F_X is the cumulative distribution function of a (continuous) random variable X, then $F_X(X)$ is uniformly distributed in $[0, 1]$.

This gives a natural method for generating random numbers: If U is a uniformly distributed in $[0, 1]$, then $F_X^{-1}(U)$ has the same distribution as X.

Use this approach to write a program to generate random numbers in $\{0, 1, 2, 3\}$ according to the probability mass function $[0.1, 0.3, 0.1, 0.5]$. Your program must print 10 independent samples according to this distribution. Also, for this pmf, draw the CDF and find F_X^{-1} (I mean, $F_X^{-1}(y)$ is the set of all values that map to y by F_X).

Exercise 1.5 (Bounds on the binomial coefficient)**.** Let us prove some results that will be useful in the rest of the course.

1. Show that for all positive integers (n, k, l) with $n > k > l$, we have $\frac{n-l}{k-l} \geqslant \frac{n}{k}$.
2. Prove the following bounds:
$$\left(\frac{n}{k}\right)^k \leqslant \binom{n}{k} \leqslant \left(\frac{n^k}{k!}\right).$$
3. Prove that for all positive integers $n > k > 0$, we have
$$\frac{\sqrt{2\pi}}{e^2} \frac{1}{\sqrt{np(1-p)}} 2^{nH_2(p)} \leqslant \binom{n}{k} \leqslant \frac{e}{2\pi} \frac{1}{\sqrt{np(1-p)}} 2^{nH_2(p)}$$
where $p = k/n$ and $H_2(p) = -p \log_2 p - (1-p) \log_2(1-p)$.

   You may use (without proof) Stirling's approximation:
$$\sqrt{2\pi} n^{n+1/2} e^{-n} \leqslant n! \leqslant e n^{n+1/2} e^{-n}.$$
4. Use the above to conclude for any $0 < p \leqslant 1/2$,
$$\lim_{n\to\infty} \frac{1}{n} \log_2 \binom{n}{\lfloor np \rfloor} = H_2(p).$$

Exercise 1.6. Write a program (in C/C++/python/matlab) to do the following: given an input (English) text file (named inputfile.txt, stored in the same directory as the program), compute the empirical pmf/histogram of the characters in the file (i,e., the fraction of 'a', 'b', etc.). The program should print the empirical pmf as a 26-length vector corresponding to letters in the English alphabet.

The empirical pmf of a sequence is also called the type of the sequence.