# CS 373 HW 1- Gelei Chen

## Q1: Random Variables and Probability

**1.**

Ω (all possible events) = 36(6*6) possibilities
E(Sum of the dice is odd) = 18. E will occur half of the time. 18=36/2
F(At least one of the dice landed on 1) = 11. (1,1) + (1,2~6) + (2~6,1)
G(sum is 5) = 4. (1,4) + (4,1) + (2,3) + (3,2)

- **a.** P (E ∩ F) = $\frac{E \cap F}{\Omega}$ = $\frac{6}{36}$ $\approx$0.1667
- **b.** P (E ∪ F) = $\frac{E \cup F}{\Omega}$ = $\frac{18+5}{36}$ $\approx$0.6389
- **c.** P (F ∩ G) = $\frac{F \cap G}{\Omega}$ = $\frac{2}{36}$ $\approx$0.0556
- **d.** P (E ∩ (Ω \ F)) = $\frac{E \cap (\Omega - F)}{\Omega}$ = $\frac{12}{36}$ $\approx$0.3333
- **e.** P (E ∩ F ∩ G) = $\frac{E \cap F \cap G}{\Omega}$ = $\frac{2}{36}$ $\approx$0.0556

**2.** Bayes' rule
Let H be the event that a person has the disease
Let T be the test result.
Given:
P(H) = 0.005, P(T+ | H) = 0.95, P(T- | ¬H) = 0.95
Question:
P(H | T+)?

$P(H \cap T+)$ = P(T+ | H) * P(H) = 0.95*0.005 = 0.00475

P(T) = P(H) * P(T+ | H) + P(T | ¬H) * P(¬H) = 0.005 * 0.95 + 0.05 * 0.995 = 0.00475 + 0.04975 = 0.0545
P(H|T+) = $\frac{P(H \cap T+)}{P(T+)}$ = 0.00475 / 0.0545 ≈ 0.0872

**3.** Joint and Conditional Probabilities

- **a.** E will occurs if two or more disks fail.
  E = $d_{1f}d_{2f}d_3$ + $d_{1f}d_2d_{3f}$ +$d_1d_{2f}d_{3f}$ + $d_{1f}d_{2f}d_{3f}$ , because the disks fail independently
  P(E) = 0.01 * 0.03 * (1 - 0.05) + 0.01 * (1 - 0.03) * 0.05 + (1 - 0.01) * 0.03 * 0.05 + 0.01 * 0.03 * 0.05 = 0.00227

- **b.** F will occurs if (i)$d_1$ fails (ii) $d_2$ and $d_3$ both fail
  F = $d_{1f}$ + $d_{2f}d_{3f}$ - $d_{1f}d_{2f}d_{3f}$
  P(F) = 0.01 + 0.03 * 0.05 - 0.01 * 0.03 * 0.05 = 0.011485
- **c.** P(F|$d_{3f}$) = $\frac{P(F \cap d3f)}{P(d3f)}$ = $\frac{0.99 * 0.03 * 0.05 + 0.01 * 0.03 * 0.05 + 0.01 * 0.97 * 0.05}{0.05}$ = 0.001985 / 0.05 = 0.0397

**4.** Independence

**a.** Given E,F and G are independent events

i. Suppose B = F $\cup$G, then B and E are also independent events

ii. Prove P[E$\cap$(F $\cup$G)] = P(E) * P(F $\cup$G)

iii. Left hand-side = P(E$\cap$B) $\leftarrow$ substitute B

= P(E) * P(B) $\leftarrow$ independent events

= P(E) * P(F $\cup$G) $\leftarrow$ substitute B again

= right hand-side

So, I prove that P[E$\cap$(F $\cup$G)] = P(E) * P(F $\cup$G)

**b.** Given P(A,B) = P(A) * P(B)

Let $\neg$A = $\Omega \setminus A$

Prove that C and B are also independent

If I can prove P($\neg$A,B) = P($\neg$A)P(B), then I will show that $\Omega \setminus A$ and B are independent events as well.

P(B) = P(A, B) + P($\neg$A, B)

so P($\neg$A, B) = P(B) - P(A,B)

= P(B) - P(A)P(B) $\leftarrow$ A and B are independent events

= P(B) [ 1 - P(A)]

= P(B)P($\neg$A) $\leftarrow$ here I proved it.

So, $\Omega \setminus A$ and B are also independent.

**c.** They are not independent.

Prove by contraction:

Assume X and Y are independent, then P(X|Y) = P(X) and P(Y|X) = P(Y).

P(X) = 0.25, P(Y) = 0.5 by enumeration

P(X|Y) = P(X,Y) / P(Y)

= 0.25 / 0.5

= 0.5 **!=** P(X) 0.25

P(Y|X) = P(X,Y) / P(X)

= 0.25 / 0.25

= 1 **!=** P(Y) 0.5

So, X and Y are not independent.

**5.** Linearity of Expectation

Prove that E[X+Y] = E[X] + E[Y]

$$E[X+Y] = \sum_x \sum_y (x+y)P_{xy}(x,y) \leftarrow \text{by definition of expected values}$$

$$= \sum_x \sum_y xP_{xy}(x,y) + \sum_y \sum_x yP_{xy}(x,y) \leftarrow \text{distributive law of multiplication}$$

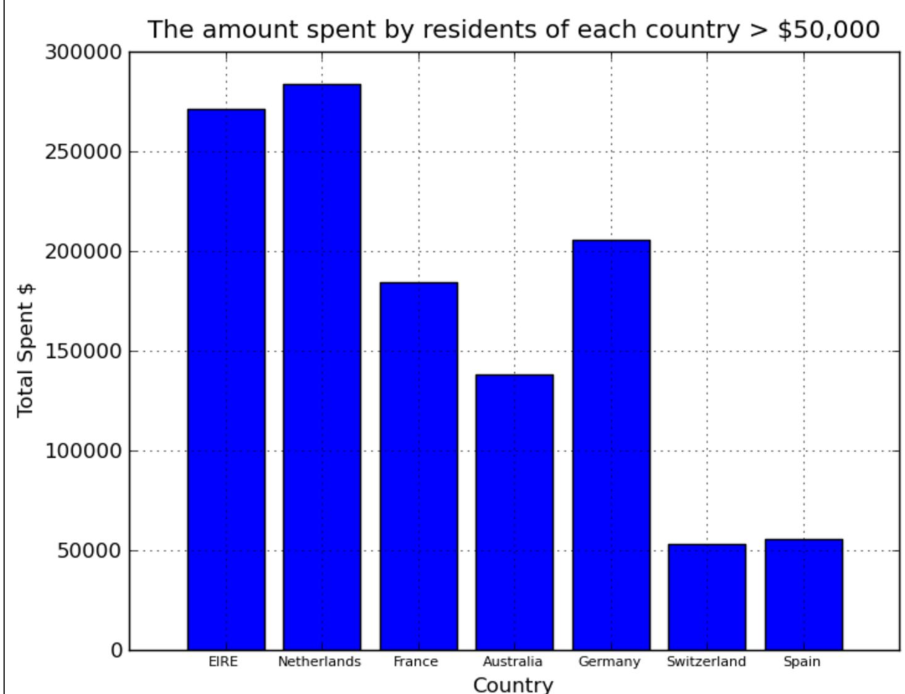
$$= \sum_x x \sum_y P_{xy}(x,y) + \sum_y y \sum_x P_{xy}(x,y)$$

$$= \sum_x x P_x(x) + \sum_y y P_y(y)$$

$$= E(X) + E(Y) \leftarrow \text{I prove that } E[X+Y] = E[X] + E[Y]$$

## Q2: Working with Python

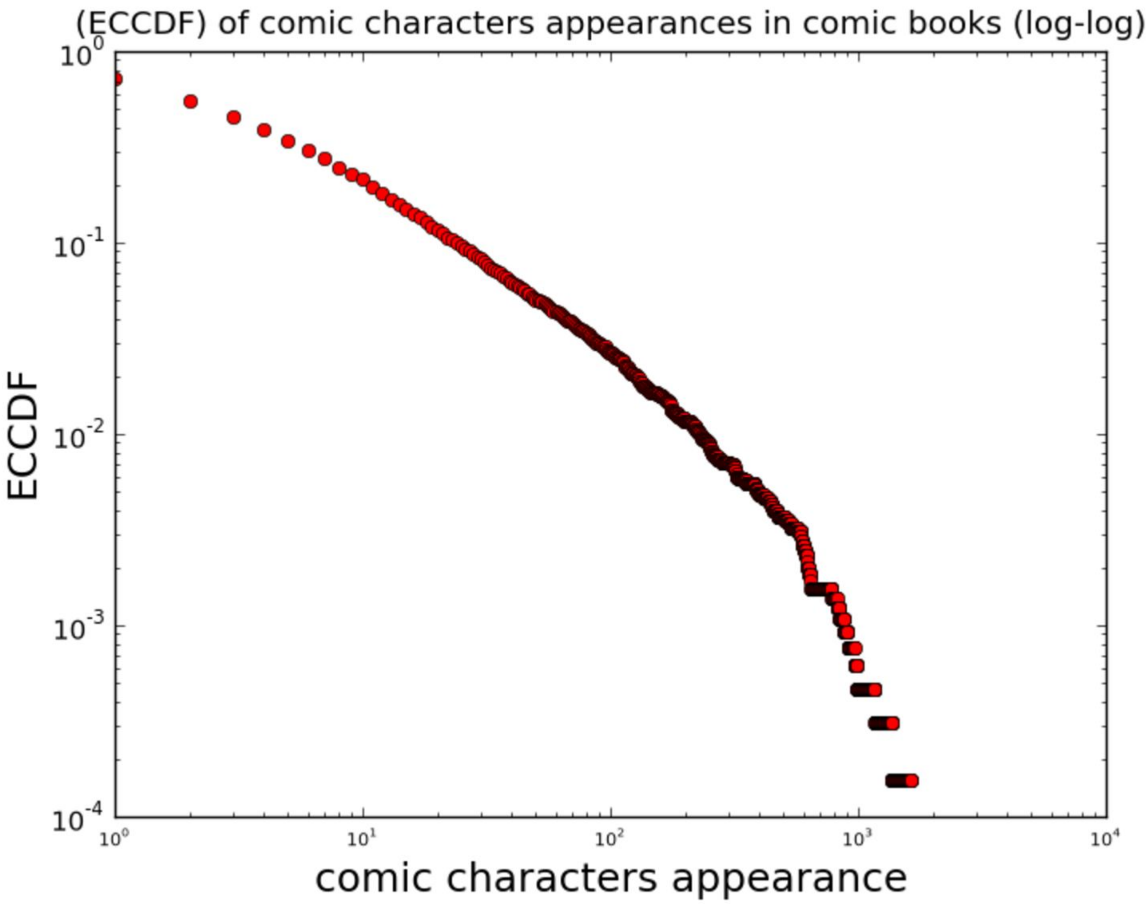
1. **I/O**
   a. total_number_of_rows is **43756**
   b. number_of_unique_items based on unique stock code are **2785**
2. **Data Processing**
   a. average unit price for the product with stock code 20685 is **7.08958333333**
   b. hour in the day are most items sold in the given data set is **10AM** and the maximum quantity is 166603
   c. generate bar graph using {'EIRE': 271119.30000000237, 'Netherlands': 283724.34000000003, 'France': 184582.7399999999, 'Australia': 138171.30999999979, 'Germany': 205529.59000000093, 'Switzerland': 53087.900000000009, 'Spain': 55705.609999999921}




3. **Writing (nothing goes to pdf)**

## Q3: Working with matrices in Python

a. Plot




b.

i. QUILL appear in **4** distinct comic books
ii. comic book that has the most number of characters is **COC 1**. Its id is 6496 with 111 # of characters
iii. (Nothing goes to pdf)

## Q4:

a.

**[ 5.84115794e-01 4.04326940e-01 2.58149738e-02 -2.87254130e-03**
**-4.17261214e-03 -2.82537022e-03 -1.70679154e-03 -1.00783093e-03**
**-5.97762852e-04 -3.58945304e-04 -2.16822098e-04 -1.27160479e-04**

**-6.26749638e-05 -3.46998422e-06 7.05345592e-05 1.87730020e-04**
**3.97047945e-04 7.89692276e-04 1.54026784e-03 -3.33277398e-03]**

b. inner_product:**0.801301030151**