Data Science, 2022
Tut 6: Machine Learning 1

1. [Probability] Assume that the probability of obtaining heads when tossing a coin is λ.
a. What is the probability of obtaining the first head at the (k + 1)-th toss?
b. What is the expected number of tosses needed to get the first head?

2. [Probability] Assume X is a random variable.
a. We define the variance of X as: $\text{Var}(X) = E[(X - E[X])^2]$. Prove that $\text{Var}(X) = E[X^2] - E[X]^2$.
b. If $E[X] = 0$ and $E[X^2] = 1$, what is the variance of X? If $Y = a + bX$, what is the variance of Y ?

3. [Probability] Your friend Aku is a great predictor about winning horse race. Assume that we know three facts: 1) If Aku tells you that a horse name black beauty will win, it will win with probability 0.99. 2) If Aku tells you that a black beauty will not win, it will not win with probability 0.99999. 3) With probability 10^{-5}, Aku predicts that a black beauty is a winning horse. This also means that with probability $1 - 10^{-5}$, Aku predicts that a black beauty will not win.
a. Given a horse, what is the probability that it wins?
b. What is the probability that Aku correctly predicts a black beauty is winning ?

Make Assumptions about values when it is necessary in consistent manner. Refer necessary table from following link when necessary.

https://www.sheffield.ac.uk/polopoly_fs/1.43999!/file/tutorial-10-reading-tables.pdf

Testing a Proportion of small samples

1. H_0: $p = p_0$

2. One of the alternatives H\: $p < p_{n,}$ $p > p_{o,}$ or $p \neq p_o$

3. Choose a level of significance equal to a.

4. Test statistic: Binomial variable *X* with $p = p_0$.

5. Computations: Find x, the number of successes, and compute the appropriate P-value.

6. Decision: Draw appropriate conclusions based on the P-value

Ex. 1

A builder claims that air-conditions are installed in 70% of all homes being constructed today in the city of Mumbai. Would you agree with this claim
if a random survey of new homes in this city shows that 8 out of 15 had air-conditions installed? Use a 0.10 level of significance

DS Tutorial 6 (Ex 1)

(a) $P(k \text{ tails in the first } k \text{ tosses, then 1 head}) = (1-\lambda)^k \lambda$

(b) For n no. of tosses

$X = E[n]$

$X = \lambda \times 1 + (1-\lambda)(X+1)$ [If tosses are independent]

$X = \frac{1}{\lambda}$

Ex.2

A commonly prescribed drug for relieving nervous tension is believed to be only 60% effective. Experimental results with a new drug administered to a random sample of 100 adults who were suffering from nervous tension show that 70 received relief. Is this sufficient evidence to conclude that the new drug is superior to the one commonly prescribed? Use a 0.05 level of significance.

Ex 2

a $\text{Var}(X) = E[(X - E[X])^2]$

We know that

$= E[X^2 - 2XE[X] + E[X]^2]$

$= E[X^2] - 2E[XE[X]] + E[X]^2$

$= E[X^2] - 2E[X]^2 + E[X]^2$

$= E[X^2] - E[X]^2$

Hence proved.

b $E[X] = 0 \quad E[X^2] = 1$

$y = a + bx$

$\text{Var}(X) = E[X^2] - E[X]^2$

$= 1 - 0$

$= 1$

$\text{Var}(y) = \text{var}(a + bx)$

$= E[y^2] - E[y]^2$

$E[y^2] = E[(a+bx)^2]$

$= E[a^2 + 2abx + b^2x^2]$

$= a^2 + 2ab\,E[x] + b^2E[x^2]$

$= a^2 + b^2$

$$E[y] = E[a+bX]$$
$$= a + bE[X]$$
$$= a$$
$$Var(y) = a^2 + b^2 - (a)^2$$
$$= b^2$$

Ex.3

A vote is to be taken among the residents of a Mumbai and the surrounding area to determine whether a proposed Nuclear plant should be constructed. The construction site is within the Mumbai limits, and for this reason many voters in the surrounding area feel that the proposal will pass because of the large proportion of Mumbai voters who favor the construction. To determine if there is a significant difference in the proportion of Mumbai voters and surrounding area voters favoring the proposal, a poll is taken. If 120 of 200 Mumbai voters favor the proposal and 240 of 500 surrounding area residents favor it, would you agree that the proportion of Mumbai voters favoring the proposal is higher than the proportion of surrounding area voters? Use an a = 0.05 level of significance.

Ex 3

Let E be the event "Aku predicts that the horse is a winning horse"

$\therefore \neg E$ will be the event "Aku predicts that the horse is not a winning horse"

Let W be the event that the horse is a winning horse

$\neg W$ be the event that the horse is not a winning horse

Given $P(W|A) = 0.99$

$P(\neg W | \neg A) = 0.99999$

$P(A) = 10^{-5}$

a) $P(W) = P(W, A) + P(W, \neg A)$

$= P(W|A)\,P(A) + P(W|\neg A)\,P(\neg A)$

$= 0.99 \times 10^{-5} + (1 - 0.99999)(1 - 10^{-5})$

$= 1.99 \times 10^{-5}$

b) Prob that Aku predicts winning correctly

$$P(A|W) = \frac{P(A, W)}{P(W)}$$
$$= \frac{P(W|A)\,P(A)}{P(W)}$$
$$= \frac{0.99 \times 10^{-5}}{1.99 \times 10^{-5}}$$
$$\approx 0.4975$$