Homework 1 Zhuo Liu

1. Problem 1
   (a) exactly 8 users experienced insomnia as a side-effect
   P(X=8)=0.1607
   (b) less than 4 users experienced insomnia as a side-effect
   P(X<4)=P(X=0)+P(X=1)+P(X=2)+P(X=3)=2.10297E-05+0.0003+0.0018+0.0076 =972.10297E-05
   (c) 5 or more users experienced insomnia as a side-effect
   P(X>=5)=1-P(X=0)-P(X=1)-P(X=2)-P(X=3)- P(X=4)=1-2.10297E-05-0.0003-0.0018 -0.0076-0.02224=0.9680389703
   (d) 20 users do not experience insomnia as a side-effect
   P(X=5)=0.0506
   (e) 2 or fewer users having insomnia
   P(X<=2)=P(X=0)+P(X=1)+P(X=2)=2.10297E-05+0.0003+0.0018=0.0021210297
   Because P(X<=2)<2%, the probability of observing 2 or fewer users having insomnia is an unusual event.
   In my opinion, we should use probabilities, because probability is a measurement of specific event. We can use probability as the reference.

2. Problem 2
   (a)

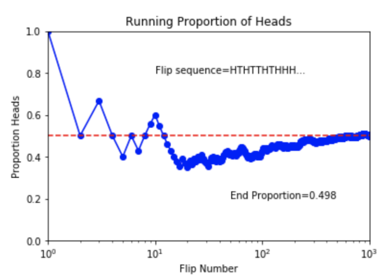

Figure 1: N=1000

   (b) We can discover from the graph that with the increase of flip numbers the probability would get closer to 0.5 eventually. Although during the first ten times, the probability would cause strong variations. The probability would become stable in the end.

(c)

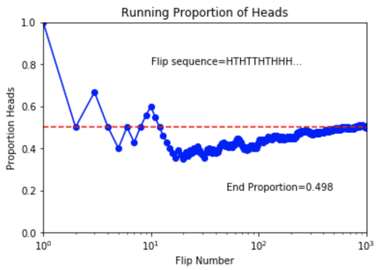

Figure 2: N=10^3

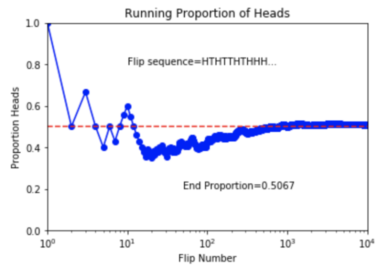

Figure 3: N=10^4

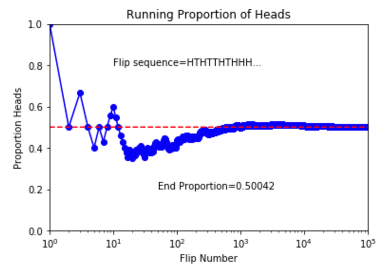

Figure 4: N=10^5

If we increase N, we can learn from graph that end proportion might get more closer to 0.5 and the error would reduce significantly.

(d)

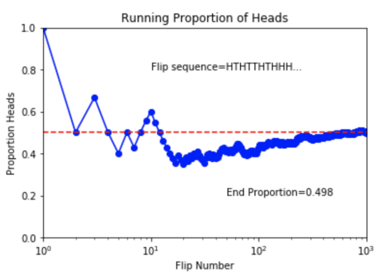

Figure 5: P(H=0.5)

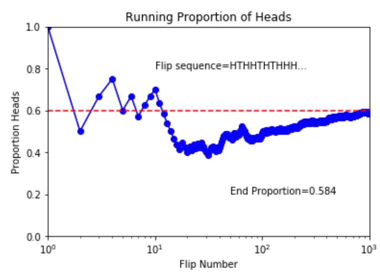

Figure 6: P(H=0.6)

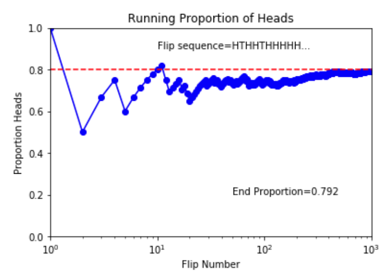

Figure 7: P(H=0.8)

If we change the probability of H to 0.6, the end proportion would get close to 0.6 in the end. If we change the probability of H to 0.8, the end proportion would get closer to 0.8 eventually.

3. Problem 3

(a)

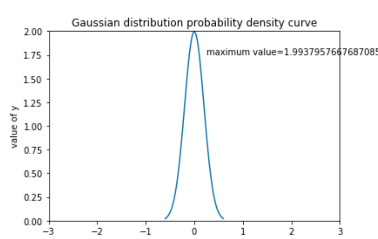

Figure 8: Gaussian distribution

First, let's begin with the mathematic perspective.Because when taking x=μ, we can get the maximum value, the value of exponential function $e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ would be 1. So the value of $\frac{1}{\sqrt{2\pi\sigma}}$ is greater than 1. So the maximum value is 1.9937 according to probability density function. Also, from my perspective, it is meaningless to calculate the value of mass density function. Because only the area of the density distribution represents the probability.

(b)

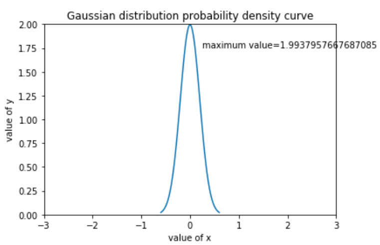

Figure 9: σ=0.2

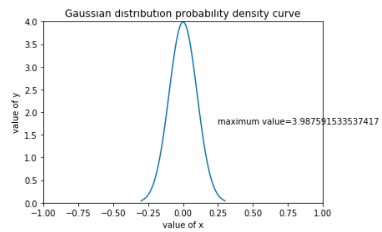

Figure 10: σ=0.1

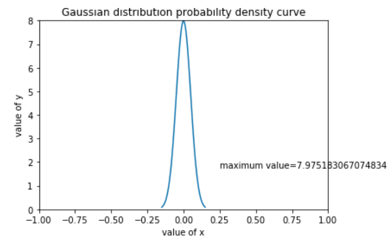

Figure 11: σ=0.05

If we reduce the value of σ, we can see from the graph that the density function of Gaussian distribution becomes narrower and the max value of probability function becomes greater as the figure shown above.

(c)

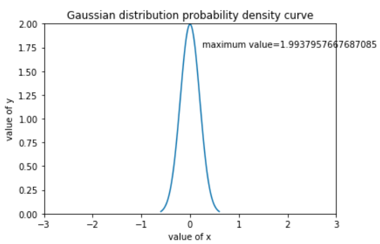

Figure 12: μ=0

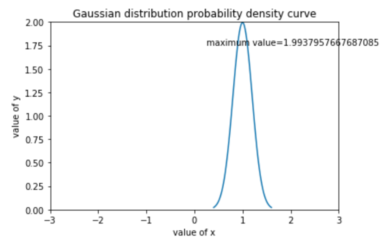

Figure 13: μ=1

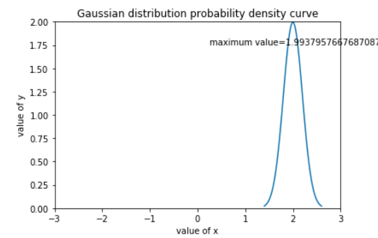

Figure 14: μ=2

If we change the value of mean, the entire Gaussian distribution would move accordingly. Because the Gaussian distribution is symmetric with respect to the value of mean.

4. Problem 4

(a) f(x)=$\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

Parameters: μ which is the mean, σ which is the variance

It models continuous variables.

Applications:If we want to take some measurements in statistic areas, we can take advantage of Gaussian distribution.

(b) f(k;p)=$\begin{cases} p & \text{if } k = 1 \\ q = 1 - p & \text{if } k = 0 \end{cases}$

Parameters: p which is the probability of success of the event, q which is the probability of failure of the same event.
It models discrete variables.
Applications: The special case of Binomial distribution is when N=1. There are only two possible results and the sum of the probability should be 1. For instance, success or failure, win or lose, etc.

(c) probability mass function
f(k,n,p)=P(x=k)=$\binom{n}{k}p^k(1-p)^{n-k}$
Parameters:k which is the frequency of event occurs which probability is p, p which represents the probability of specific event and 1-p should be the case this event can not happen, n is the number of trials in total.
It models discrete variables.
Applications: When there are fixed number of trials and these trials are independent, the sum of these probabilities is 1. For instance, we throw a dice and see whether we can get a six or not.

(d) probability mass function
f($x_1, ..., x_k; n, p_1, ..., p_k$)= $\begin{cases} \frac{n!}{x_1!...x_k!}p_1^{x^1} \times ... \times p_k^{x_k}, when \sum_{i=1}^{k} x_i = n \\ 0 \qquad\qquad\qquad\qquad otherwise \end{cases}$
Parameters: $x_1, ..., x_k$ which are k possible different results, $p_1, ..., p_k$ which are correspondingly probabilities, n which is number of trials
It models discrete variables.
Applications: If we want to get multiple results, like throw a dice, what is the probability of getting 1,2,3,4,5,6 respectively, we can apply multinomial distribution.

(e) $f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$
Parameters: λ which is the number of random events happened in unit time
It models continuous variable.
Applications:If we want to get the probability of random events happened in unit time. For instance, if a baby was born in one hour, what is the probability of no baby is born in next two hours.

(f) P(k events in interval)=$\frac{\lambda^k e^{-\lambda}}{k!}$
Parameters: λ which is the number of random events happened in unit time, k which is number of events
It models continuous variable.
Applications:If we want to get the probability of specific intervals between two events. For instance, the probability of time intervals of how many calls one staff can receive in the phone company or the time intervals one website can be visited and so forth.

(g) $f(x_1, ...x_K; \alpha_1, ..., \alpha_K) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$
Parameters:$B(\alpha)$ which is multivariate beta function and Lebesgue measure on the Euclidean space R^{K-1}.

It models continuous variable.
Applications: We can use the Dirichlet distribution combined with a Multinomial distribution to form a conjugate prior. We can apply this to the process of natural language and specifically the research on topic model.