CS 760 Homework 1

Lingyun Xiao

February 4, 2022

1. $\mathbb{R}^D$ of course is a subset of itself. Also, we know that $\mathbb{R}$ is closed under addition and scalar multiplication. Therefore, for every $a, b \in \mathbb{R}$, and every $\mathbf{u}, \mathbf{v} \in \mathbb{R}^D$, we know that $a\mathbf{u} + b\mathbf{v} \in \mathbb{R}^D$.

2. (a) Consider $\mathbf{u} = [-1] \in \mathbb{R}^D$. $\sqrt{-1} = i \notin \mathbb{R}^D$.

   (b) Consider the subspace of all positive real numbers: $\mathbb{R}_+$. For every $\mathbf{u} \in \mathbb{R}_+$, we know that $\sqrt{\mathbf{u}} \in \mathbb{R}_+$.

3. Every $\mathbf{x}, \mathbf{y} \in \mathbb{U}$ can be written as: $\mathbf{x} = \sum_{i=1}^{R} a_i \mathbf{u_i}$ and $\mathbf{y} = \sum_{i=1}^{R} b_i \mathbf{u_i}$ for some $a_i, b_i \in \mathbb{R}$.

   For any $c, d \in \mathbb{R}$, we know that $ca_i + db_i \in \mathbb{R}$, so:

   $c\mathbf{x} + d\mathbf{y} = c\sum_{i=1}^{R} a_i \mathbf{u_i} + d\sum_{i=1}^{R} b_i \mathbf{u_i} = \sum_{i=1}^{R} (ca_i + db_i)\mathbf{u_i} \in \mathbb{U}$.

4. (a) Call the event that a person has diabetes D. Call the event that the genes being inactive as G. Then:

$$P(D|G) = \frac{P(G|D)P(D)}{P(G)} = \frac{0.95 * 0.093}{P(G)} = \frac{0.088}{P(G)}$$

   (b) I need to know the probability that any U.S. person has inactive genes.

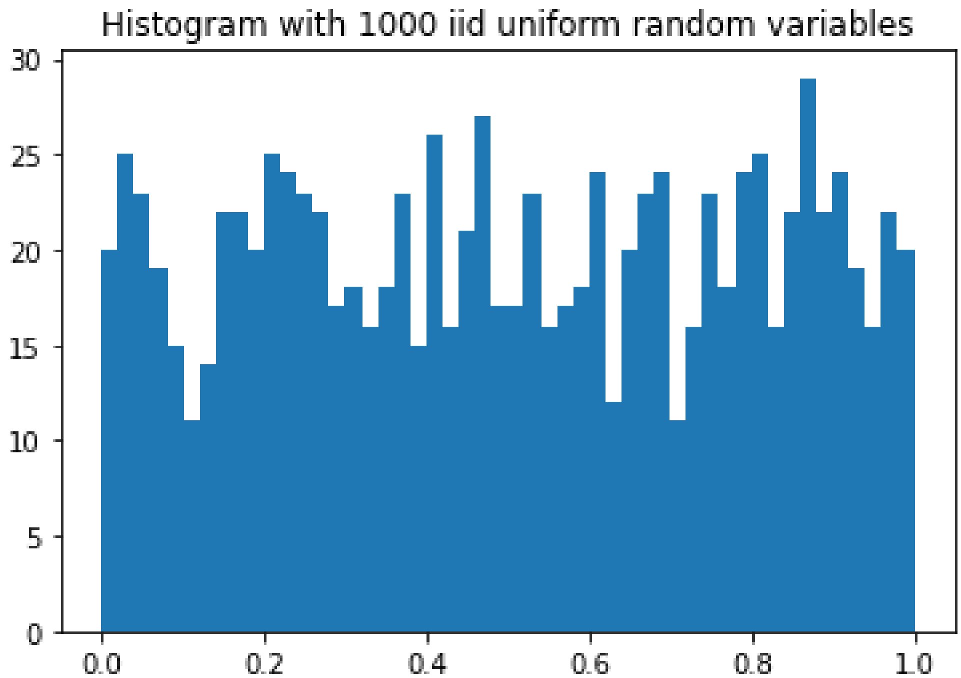
   (c) If $P(G)$ is close to 8.8%, then I should be concerned. If $P(G)$ is much higher than 8.8%, then I would not be so concerned.

5. Based on the properties of the delay, I would propose the following distribution:

$$P(\mathrm{x}|\theta) = \begin{cases} 0, & \text{for } \mathrm{x} < t_0 \\ \theta e^{-\theta(\mathrm{x}-t_0)}, & \text{for } \mathrm{x} \geq t_0 \end{cases}$$

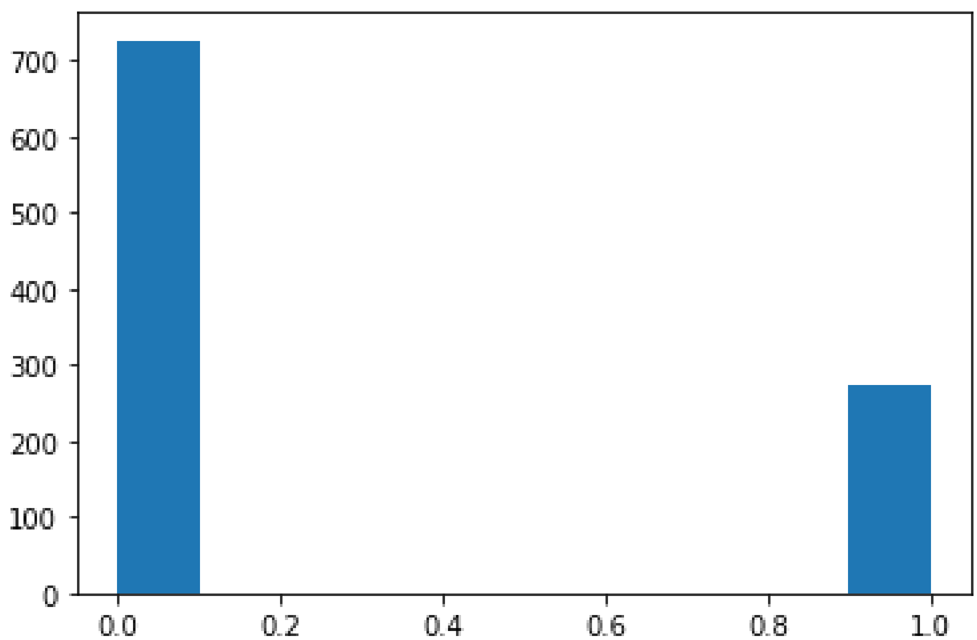
   The probability that $\mathrm{x} < t_0$ is zero. Also, the probability decays when x increases.

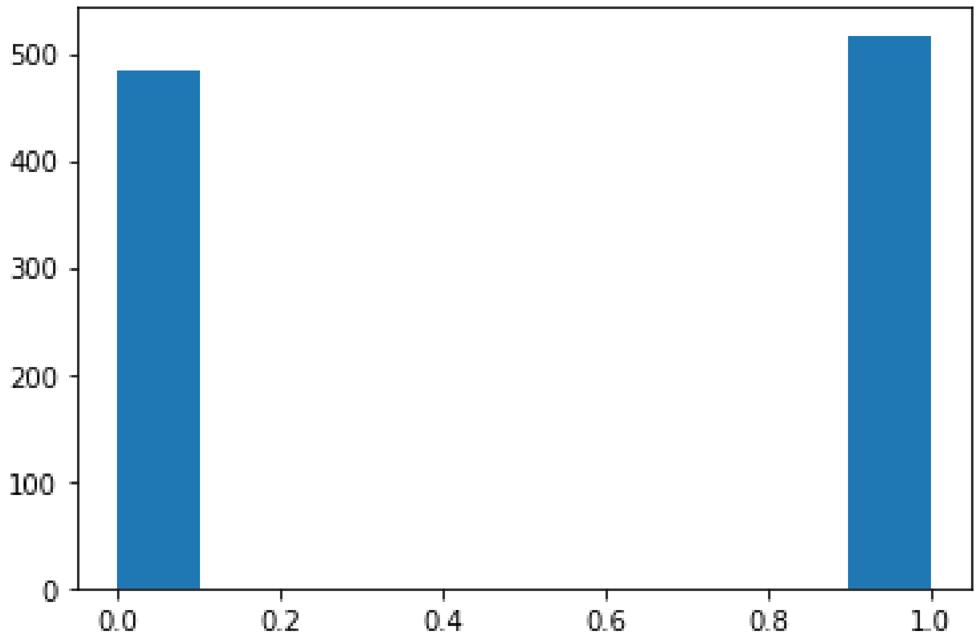
6. (a) The histogram shows the distribution of 1000 iid uniform random variables from $(0, 1)$ with bin size 50. The whole distribution looks fairly uniform with some zigzags.

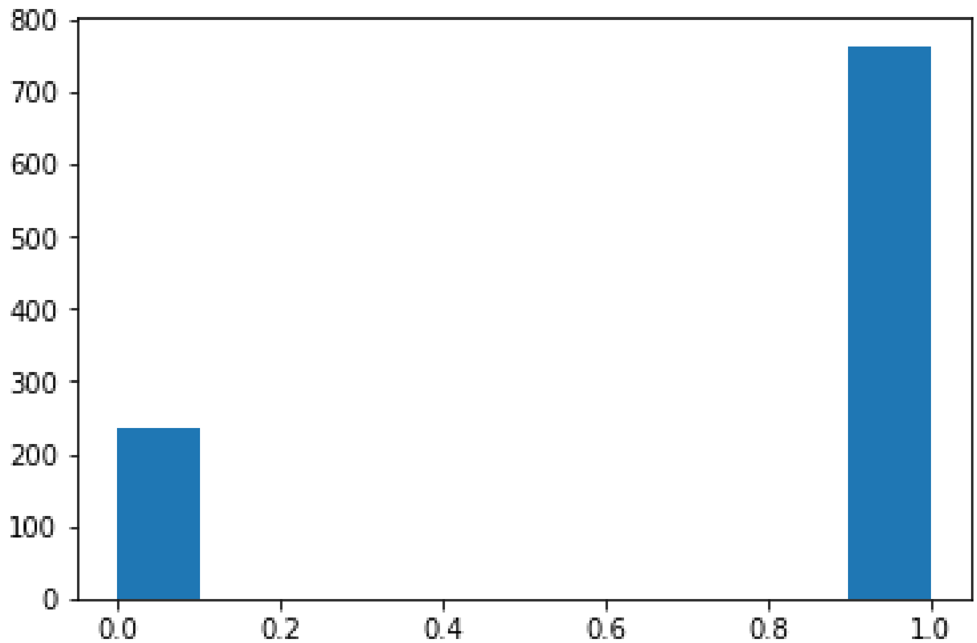


(b) $y_i \sim Bernoulli(p)$ because $\mathbb{P}(x_i \leq p) = p$ by uniform distribution.

(c) Yes. The histograms follow the distribution defined above.

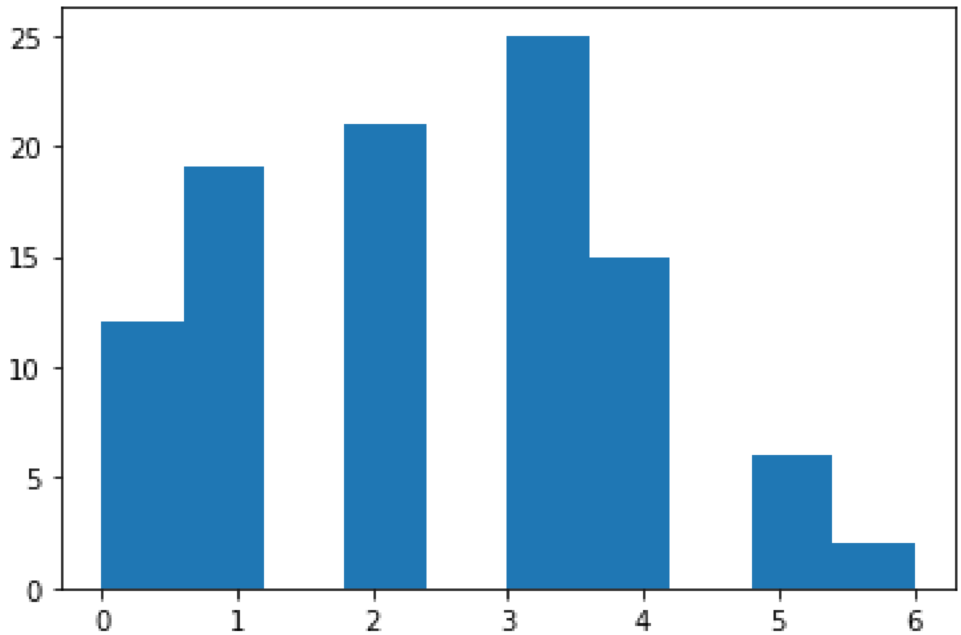


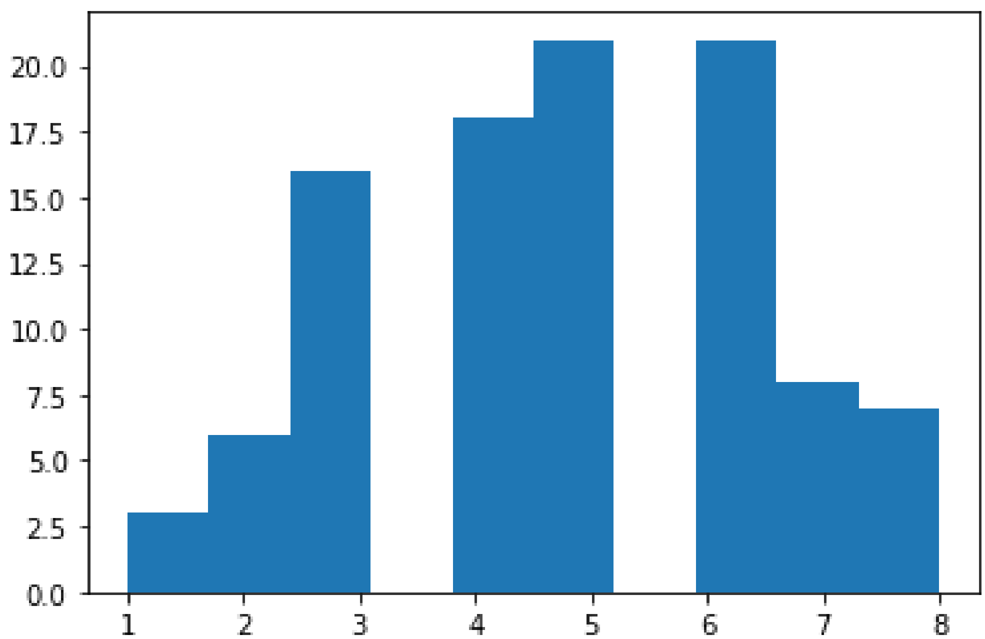


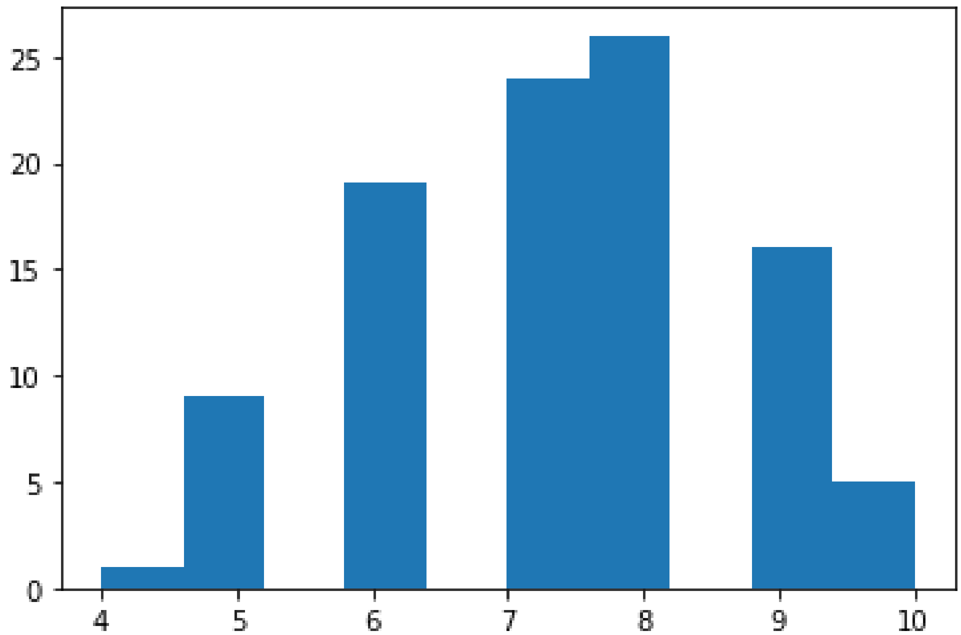


(d) $z_k \sim Binomial(n, p)$

(e) Yes. The histograms follow the distribution defined above.


0
5
10
15
20
25
0 1 2 3 4 5 6


0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
1 2 3 4 5 6 7 8



7. (a) Define:

$$\sigma(\boldsymbol{\theta}, x_i) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T x_i}}$$

Then we have:

$$\frac{\partial \sigma}{\partial \boldsymbol{\theta}} = \sigma(1 - \sigma) x_i$$

Therefore, we can calculate the gradient as follows:

$$\nabla_l(\boldsymbol{\theta}) = \sum_{i=1}^{m} \left(\frac{y_i}{\sigma} \frac{\partial \sigma}{\partial \boldsymbol{\theta}} + \frac{1 - y_i}{1 - \sigma} \frac{\partial \sigma}{\partial \boldsymbol{\theta}} \right) = \sum_{i=1}^{m} (y_i - \frac{1}{1 + e^{-\boldsymbol{\theta}^T x_i}}) x_i$$

(b) Calculate each entry of the Hessian as follows:

$$\nabla_l^2(\theta_i) = \frac{\partial \nabla_l}{\partial \theta_i} = \frac{\partial \sigma}{\partial \theta_i} x_i = \sum_{i=1}^{m} x_i x_i^T \sigma(1 - \sigma)$$

Define a diagonal matrix D where $D_{ii} = \sigma(\theta_i, x_i)(1 - \sigma(\theta_i, x_i))$, then:

$$\nabla_l^2(\boldsymbol{\theta}) = XDX^T$$

(c) The log-likelihood is a scalar. The gradient of the log-likelihood is a vector. The Hessian is a matrix.