CS 760 Homework 1

Kefan Zheng 908 617 5008

P1.1

According to the definition of subspace, for a subset $U \subseteq R^D$, if for every $a, b \in R$ and every $u, v \in U$, $au + bv \in U$, then U is a subspace.

We can use a D-dimension vector $[x_1, x_2, \cdots x_D]$ where $x_1, \cdots x_D \in R$ to present each point in the R^D. For every $a, b \in R$ and every $u, v \in R^D$, $au + bv = a[u_1, u_2, \cdots u_D]^T + b[v_1, v_2, \cdots v_D]^T = [au_1 + bv_1, \cdots, au_D + bv_D]^T$. $a, b \in R$, $u_D, v_D \in R$, so $au_D + bv_D \in R$, so $[au_1 + bv_1, \cdots, au_D + bv_D]^T \in R^D$, so R^D is a subspace

P1.2 (a)

For example, $[x_1, x_2, \cdots x_D]$ is an element of R^D, where $x_1, \cdots x_D \in R$. We let $x_1 = -1$ here, so this element is $[-1, x_2, \cdots x_D]$. After element-wise square roots, the result is $[\sqrt{-1}, \sqrt{x_2}, \cdots \sqrt{x_D}]$. $\sqrt{-1} = i \notin R$, so the result is not an element of R^D, which proves R^D is not closed under element-wise square roots.

P1.2 (b)

The subset $U \subseteq R^D$ and this U contains only one element – the D dimension zero vector.

First prove it is a subspace (closed under linear combinations). For every $a, b \in R$ and every $u, v \in U$, because U only contains zero vector, so $u = v = 0$, $au + bv = 0 \in U$, so U is a subspace. Then prove it's closed under element-wise square roots. For every $u \in U$, $u = 0$, so the element-wise square root of u is still 0, so it's still in the subspace U. So it's closed under element-wise square root.

P 1.3

According to the definition of Span, $\mathrm{span}[u_1, \cdots, u_R] := \{x \in R^D : x = \sum_{r=1}^{R} c_r u_r \text{ for some } c_1, \cdots, c_R \in R\}$, which means $U = \mathrm{span}[u_1, \cdots u_R]$ = the set of all linear combinations of $\{u_1, \cdots, u_R\}$. So for every $a, b \in R$ and every $u, v \in U$, u is a linear combination of $\{u_1, \cdots, u_R\}$, v is also a linear combination of $\{u_1, \cdots u_R\}$, so $au + bv$ is still a linear combination of $\{u_1, \cdots, u_R\}$, so $au + bv \in U$, so U is a subspace.

P 1.4 (a)

$P(D) = 0.093 \quad P(I|D) = 0.95$

$$P(D|I) = \frac{P(I|D)P(D)}{P(I)} = \frac{0.95 \times 0.093}{P(I)} = \frac{0.08835}{P(I)}$$

P 1.4(b)

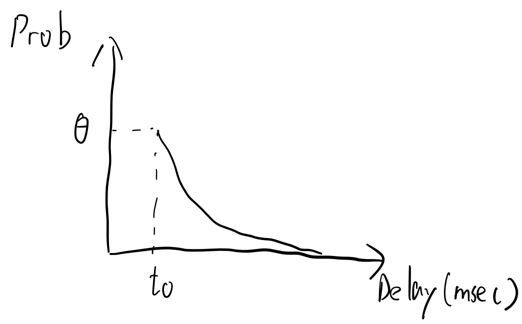
The probability of these genes being inactive in the entire population

P 1.4 (c)

The probability of these genes being inactive in the entire population is low and about the same as having diabetes, meaning that these genes are generally inactive only in people with diabetes.

P 1.5

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

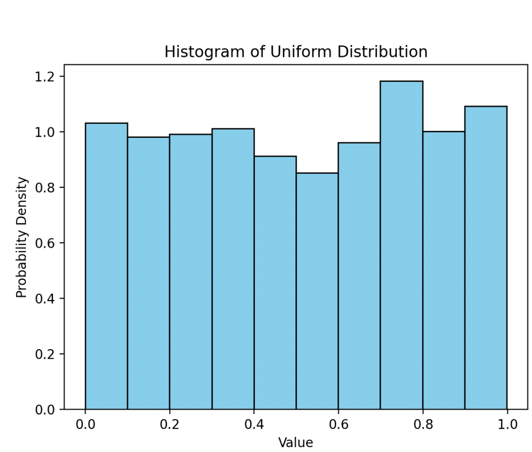


According to the description of the problem, we can draw the picture on the right (there is a minimal time delay t_0 and larger delays are rarer than shorter ones)

The curve is similar to the exponential distribution, so the model can be built by exp-distribution.

P 1.6 (a)

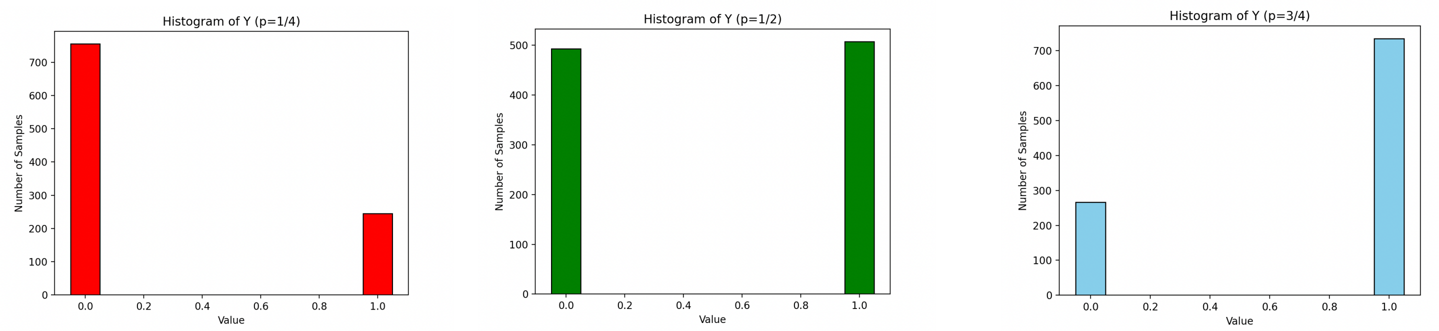
Yes. Because the uniform distribution implies that each random variable x_i has the same probability of falling into the value within the interval $[0, 1]$.



P1.6(b)

Bernoulli distribution. Because for Bernoulli distribution, $P(X=1)=p$ and $P(X=0)=1-p$. The distribution of $y_i = \begin{cases} 1 & \text{if } x \le p_i \\ 0 & \text{otherwise} \end{cases}$ is the same as it.

P1.6(c)



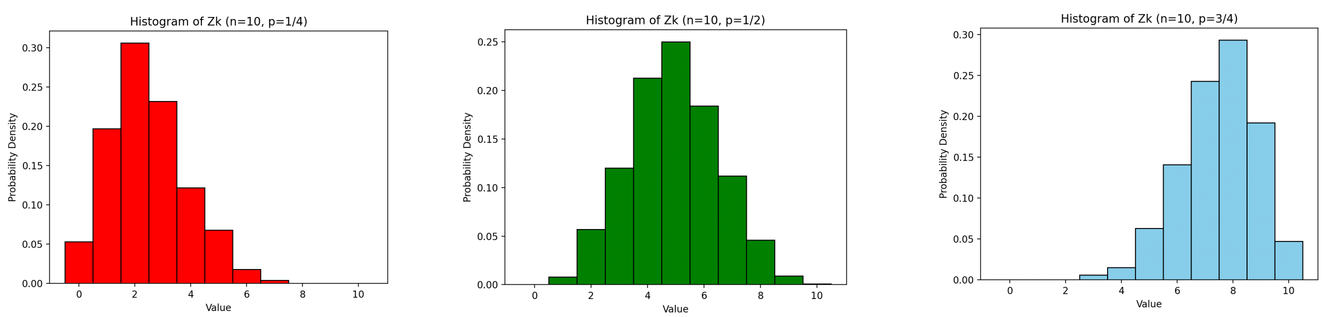
Yes, match.

P1.6(d)

Binomial distribution. Because for Z_k, $P(X=a) = \binom{n}{x} p^a (1-p)^{n-a}$, $a = 0, \dots, n$

And it's binomial distribution.

P1.6(e)



Yes, match.

P1.7(a)

$$L(\theta) = \sum_{i=1}^{N} \left[y_i \log\left(\frac{1}{1+e^{-\theta^T x_i}}\right) + (1-y_i) \log\left(\frac{1}{1+e^{\theta^T x_i}}\right) \right]$$

$$\frac{dL}{d\theta} = \sum_{i=1}^{N} \left[y_i (1+e^{-\theta^T x_i}) \cdot \frac{-e^{-\theta^T x_i} \cdot -x_i}{(1+e^{-\theta^T x_i})^2} + (1-y_i)(1+e^{\theta^T x_i}) \cdot \frac{-e^{\theta^T x_i} \cdot x_i}{(1+e^{\theta^T x_i})^2} \right]$$

$$= \sum_{i=1}^{N} \left[y_i \cdot \frac{x_i e^{-\theta^T x_i}}{1+e^{-\theta^T x_i}} - (1-y_i) \frac{x_i e^{\theta^T x_i}}{1+e^{\theta^T x_i}} \right]$$

$$= \sum_{i=1}^{N} \left[y_i \frac{x_i}{1+e^{\theta^T x_i}} - (1-y_i) \frac{x_i e^{\theta^T x_i}}{1+e^{\theta^T x_i}} \right]$$

$$= \sum_{i=1}^{N} x_i \left(y_i - \frac{1}{1+e^{-\theta^T x_i}} \right)$$

P1.7 (b)

$$\frac{dL}{d\theta} = \sum_{i=1}^{N} x_i \left(y_i - \frac{1}{1+e^{-\theta^T x_i}}\right), \quad \text{Let } \sigma(\theta) = \frac{1}{1+e^{-\theta^T x_i}}, \quad \sigma'(\theta) = \sigma(\theta)\left[1-\sigma(\theta)\right] x_i$$

$$\text{So } \frac{dL}{d\theta} = \sum_{i=1}^{N} x_i (y_i - \sigma(\theta))$$

$$H_{L(\theta)} = \frac{d^2 L}{d\theta^2} = \sum_{i=1}^{N} -x_i \sigma'(\theta) = -\sum_{i=1}^{N} \left[x_i x_i^T \cdot \sigma(\theta)\left[1-\sigma(\theta)\right]\right]$$

$$= -\sum_{i=1}^{N} \left[x_i \cdot x_i^T \cdot \frac{e^{-\theta^T x_i}}{(1+e^{-\theta^T x_i})^2}\right]$$

P1.7 (c)

$L(\theta)$: scalar, It's gradient: vector, Hessian: matrix