CSE 569 Fundamental of Statistical Learning HW1 Solutions

Solution1

Part 1

We resolve the inequality of P(error) here

The probability of error is given by the equation

$$P(error) = \int_{-\infty}^{\infty} P(error, x)dx$$

$$= \int_{-\infty}^{\infty} \min(P(\omega_1 | x), P(\omega_2 | x)p(x)dx$$

By Bayes rule we get,

$$P(error) = \int_{-\infty}^{\infty} min(P(x | \omega_1)P(\omega_1), P(x | \omega_2)P(\omega_2))dx$$

By using $\min[a, b] \leq \sqrt{ab}$,

$$P(error) \leq \int_{-\infty}^{\infty} \sqrt{P(x | \omega_1)P(x | \omega_2)P(\omega_1)P(\omega_2)}dx$$

$$\leq \sqrt{P(\omega_1)P(\omega_2)} \int_{-\infty}^{\infty} \sqrt{P(x | \omega_1)P(x | \omega_2)}dx$$

We can use the Bhattacharya coefficient ρ to get the equation in the form

$$P(error) \leq \sqrt{P(\omega_1)P(\omega_2)}\rho$$

Part2

We resolve the inequality of second constraint

In case of two-category Bayes classifier we can say that

Let $P(\omega_1) = a, P(\omega_2) = b,\ P(\omega_1 + \omega_2) = 1$

We can infer that $2ab \leq 1/2$, which gives $ab \leq 1/4$

Taking square root on both the sides, we get,

$$\sqrt{ab} \leq 1/2$$

$$\sqrt{P(\omega_1)P(\omega_2)} \leq 1/2$$

Multipling by ρ on both the sides we get

$$\sqrt{P(\omega_1)P(\omega_2)}\rho \leq \rho/2$$

Thus we prove the inequalities.

Thus we can say that

$$P(error) \leq \sqrt{P(\omega_1)P(\omega_2)}\rho \leq \frac{\rho}{2}$$

Solution 2

Since we are dealing with two-category dimensional classification problem, we can use the discriminant function for linear machine which is given by

$$w^{\intercal}(x - x_0) = 0 \qquad \text{- (1)}$$

where

$w = \mu_i - \mu_j$ and

$$x_0 = \frac{(\mu_i + \mu_j)}{2} - \frac{\sigma^2}{||\mu_i - \mu_j||^2} \ln \frac{P(\omega_i)}{P(\omega_j)}(\mu_i - \mu_j) \qquad \text{- (2)}$$

Given values of the mean for the distribution are given as

$\mu_i = [0,0]^{\intercal}$ and $\mu_j = [1, -1]^{\intercal}$, $P(\omega_1) = P(\omega_2) = 1/2$, $\Sigma_i = \Sigma_j = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

a) Finding the decision boundary

Substituting the values in the equation (2) we get,

$x_0 = 0.5(\begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix}) - 0$ (since priors are equal)

$$x_0 = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}$$

Substituting the values in the equation (1) we get,

$$(\begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix})^{\intercal}(x - 0.5 \begin{bmatrix} 1 \\ -1 \end{bmatrix}) = 0$$

$$(\begin{bmatrix}-1\\1\end{bmatrix})^{\intercal}(x-0.5\begin{bmatrix}1\\-1\end{bmatrix})=0$$

Representing x as a column vector of 2 dimensions we get,

$$(\begin{bmatrix}-1\\1\end{bmatrix})^{\intercal}(\begin{bmatrix}x_1\\x_2\end{bmatrix}-0.5\begin{bmatrix}1\\-1\end{bmatrix})=0$$

$$([-1 \quad 1])(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}0.5\\-0.5\end{bmatrix})=0$$

$$(-x_1+x_2+1)=0$$

$x_1-x_2=1$ **-> Bayes Decision boundary**

b) To find the Bhattarcharya error bound, we can use the expression for P(error) as

$$P(error)=\sqrt{P(\omega_1)P(\omega_2)}e^{-}k(1/2)$$, where

$$k\frac{1}{2}=\frac{1}{8}(\mu_2-\mu_1)^{\intercal}\frac{[\Sigma1+\Sigma2]}{2})^{-}1(\mu_2-\mu_1)+\frac{1}{2}\ln\frac{\frac{|[\Sigma1+\Sigma2]|}{2}}{\sqrt{|\Sigma_1||\Sigma_2|}}$$

$$k\frac{1}{2}=\frac{1}{8}(\begin{bmatrix}1\\-1\end{bmatrix}-\begin{bmatrix}0\\0\end{bmatrix})^{\intercal}[\frac{\begin{bmatrix}1&0\\0&1\end{bmatrix}+\begin{bmatrix}1&0\\0&1\end{bmatrix}}{2}]^{-}1(\begin{bmatrix}1\\-1\end{bmatrix}-\begin{bmatrix}0\\0\end{bmatrix})+(x-0.5\begin{bmatrix}1\\-1\end{bmatrix})+\frac{1}{2}\ln\frac{\frac{|[\begin{bmatrix}1&0\\0&1\end{bmatrix}+\begin{bmatrix}1&0\\0&1\end{bmatrix}]|}{2}}{\sqrt{|\begin{bmatrix}1&0\\0&1\end{bmatrix}||\begin{bmatrix}1&0\\0&1\end{bmatrix}|}}$$

$$=\frac{1}{8}(\begin{bmatrix}1\\-1\end{bmatrix})^{\intercal}[\frac{\begin{bmatrix}2&0\\0&2\end{bmatrix}}{2}]^{-1}\begin{bmatrix}1\\-1\end{bmatrix}++\frac{1}{2}\ln\frac{|\begin{bmatrix}1&0\\0&1\end{bmatrix}|}{\sqrt{1\times1}}$$

$$=\frac{1}{8}(\begin{bmatrix}1\\-1\end{bmatrix})^{\intercal}[\begin{bmatrix}1&0\\0&1\end{bmatrix}]^{-1}\begin{bmatrix}1\\-1\end{bmatrix}+0$$

$$= \frac{1}{8}([1 \quad -1])\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \frac{1}{8}([1 \quad -1])\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \frac{1}{8}[2] = \frac{1}{4}$$

$$k\frac{1}{2} = \frac{1}{4}$$

Substituting the above value in the P(error) equation we get

$$P(error) = \sqrt{0.5 \times 0.5}e^{-}\frac{1}{4}$$

$$P(error) = 0.389$$

Thus the error for Bhattacharya bound is **0.389**

Solution 3

We are given two gaussian distributions with different mean and equal variance and equal priors

Using the given information, we can infer that threshold for the classifier is midway between the two means, which can be also proven by the following analysis of estimating decision boundary between two gaussians

$$p(x\,|\,\omega_1) = p(x\,|\,\omega_2)$$

$$\frac{1}{\sqrt{2}\pi\sigma}e^{(}\frac{-1}{2}(\frac{x-\mu_1}{\sigma})^2) = \frac{1}{\sqrt{2}\pi\sigma}e^{(}\frac{-1}{2}(\frac{x-\mu_2}{\sigma})^2)$$

$$(x-\mu_1)^2 = (x-\mu_2)^2$$

$$x^2 + \mu_1^2 - 2x\mu_1 = x^2 + \mu_2^2 - 2x\mu_2$$

$$x = \frac{\mu_1 + \mu_2}{2}$$

The probability of error is given by the following expression

$$P(error) = \int_{R_1} p\left(\mathbf{x}|\omega_2\right) P\left(\omega_2\right) d\mathbf{x} + \int_{R_2} p\left(\mathbf{x}|\omega_1\right) P\left(\omega_1\right) d\mathbf{x}$$

According to given information about gaussian distributions, we can say that the region of misclassification for either of the classes goes through the computed decision boundary(connecting the means)

We use this information to construct the limits of the integral as

$$P(error) = \frac{1}{2}\int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} \frac{1}{\sqrt{2\pi\sigma}} e^{(\frac{-1}{2}(\frac{x-\mu_1}{\sigma})^2)} dx + \frac{1}{2}\int_{\frac{\mu_1+\mu_2}{2}}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{(\frac{-1}{2}(\frac{x-\mu_2}{\sigma})^2)} dx$$

Let $\frac{(x-\mu_1)}{\sigma} = u$ for the first integral

Similarly $\frac{(x-\mu_2)}{\sigma} = u$ for the second integral

And also updating the limits for u (using change of variables method)

$$P(error) = \frac{1}{2}\int_{-\infty}^{\frac{\mu_2-\mu_1}{2\sigma}} \frac{1}{\sqrt{2\pi\sigma}} e^{(\frac{-u^2}{2})} \sigma du + \frac{1}{2}\int_{\frac{\mu_1-\mu_2}{2\sigma}}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{(\frac{-u^2}{2})} \sigma du$$

Consider $y = \frac{\mu_1-\mu_2}{2\sigma}$

Since the above integral depicts standard distribution, it's symmetric about the means, so we can say

$\phi(-y) = 1 - \phi(y)$ and it's an even function , thus $\int_{-\infty}^{-y} = \int_{y}^{\infty}$, using this we get,

$$P(error) = \frac{1}{2}\int_{\frac{\mu_1-\mu_2}{2\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{(\frac{-u^2}{2})} du + \frac{1}{2}\int_{\frac{\mu_1-\mu_2}{2\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{(\frac{-u^2}{2})} du$$

$$P(error) = \int_{\frac{\mu_1-\mu_2}{2\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{(\frac{-u^2}{2})} du$$

Since we know that $|-x| = x$, we can say that $\frac{\mu_1-\mu_2}{2\sigma} = \frac{|\mu_2-\mu_1|}{2\sigma}$

$$P(error) = \int_{\frac{|\mu_2 - \mu_1|}{2\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{(}\frac{-u^2}{2})du$$

Solution 4

$$P(S=1|L=1) = \frac{p(S,L)}{p(L)}$$

$$= \sum_{R,G} \frac{p(S,L,R,G)}{p(L)}$$

$$= \sum_{R,G} \frac{P(L).P(G).P(R|L,G).P(S|G)}{P(L)}$$

$$\begin{aligned}
&= P(G=0).P(R=0|L=1,G=0).P(S=1|G=0)+ \\
&\quad P(G=0).P(R=1|L=1,G=0).P(S=1|G=0)+ \\
&\quad P(G=1).P(R=0|L=1,G=1).P(S=1|G=1)+ \\
&\quad P(G=1).P(R=1|L=1,G=1).P(S=1|G=1)
\end{aligned}$$

$$\begin{aligned}
&= \delta \times 0.2 \times (1-\gamma) + \delta \times 0.8 \times (1-\gamma) + \\
&\quad (1-\delta) \times 0.9 \times (1-\beta) + (1-\delta) \times 0.1 \times (1-\beta) \\
&= 1 - \delta(\gamma - \beta) - \beta
\end{aligned}$$

Similarly we can compute the probability of $P(S=1|L=0)$

$$\begin{aligned}
P(S=1|L=0) &= P(G=0).P(R=0|L=0,G=0).P(S=1|G=0)+ \\
&= P(G=0).P(R=1|L=0,G=0).P(S=1|G=0)+ \\
&= P(G=1).P(R=0|L=0,G=1).P(S=1|G=1)+ \\
&\quad P(G=1).P(R=1|L=0,G=1).P(S=1|G=1) \\
&= \delta \times 0.6 \times (1-\gamma) + \delta \times 0.4 \times (1-\gamma) + \\
&\quad (1-\delta) \times 0.7 \times (1-\beta) + (1-\delta) \times 0.3 \times (1-\beta) \\
&= 1 - \delta(\gamma - \beta) - \beta
\end{aligned}$$

Solution 5

a) Bayes boundary and the Bayes Error

Bayes boundary

As per the given distribution, the region of misclassification is essentially the space that lies between the blue and red lines which have points passing through (0.5,2), (1,0) and (0.5,0), (1.5,1) respectively, so if we solve the line equations, we shall get the decision boundary(intersecting point)

We can say that

$$\begin{aligned} p(x|\omega_1) &= 4x, x \geq 0 \;\&\; x \leq 0.5 \\ &= 4 - 4x, x \geq 0.5 \;\&\; x \leq 1 \\ p(x|\omega_2) &= x - 0.5, x \geq 0.5 \;\&\; x \leq 1.5 \\ &= -x + 2.5, x \geq 1.5 \;\&\; x \leq 2.5 \end{aligned}$$

Bayes classifier will classify ω_1 if $p(\omega_1|x) > p(\omega_2|x)$ and ω_2 otherwise

We can get the decision boundary by solving for x in the distribution function given above with the region greater than 0.5 and less than 1.

Solve for x ,

$$4x + y - 4 = 0$$
$$x - y - 0.5 = 0$$

We get, $x = 0.9$

Thus Bayes decision boundary will be $x = 0.9$

Bayes error

According to bayes error expression, we can write

$$P(error) = \int_{R_1} p\left(\mathbf{x}|\omega_2\right) P\left(\omega_2\right) d\mathbf{x} + \int_{R_2} p\left(\mathbf{x}|\omega_1\right) P\left(\omega_1\right) d\mathbf{x}$$

$$= \int_{0.5}^{0.9} (x - 0.5)dx + \int_{0.9}^{1} (4 - 4x)dx$$

$$= 0.5x^2 - 0.5x \Bigg|_{0.5}^{0.9} + 4x - 2x^2 \Bigg|_{0.9}^{1}$$

$$= 0.5 \times 0.9^2 - 0.5 \times 0.9 - 0.5^3 + 0.5^2 + 4 \times 1 - 2(1^2) - 4 \times 0.9 + 2(0.9^2)$$

$$= 0.05$$

Bayes error is **0.05.**

b) Minimax boundary and minimax error

Minimax decision boundary

For minimax decision boundary, we need to estimate a x which minimises the maximum error between the region of misclassification (as stated in the explanation for finding Bayes decision boundary)

So we get

$$\int_{0.5}^{x} (x - 0.5)dx = \int_{x}^{1} (4 - 4x)dx$$

$$0.5x^2 - 0.5x \Big|_{0.5}^{x} = 4x - 2x^2 \Big|_{x}^{1}$$

$$0.5x^2 - 0.5x + 0.125 = 2 - 4x^2 + 2x^2$$

$$1.5x^2 - 3.5x + 1.875 = 0$$

Solving this quadratic equation we get, $x = 0.833$ or $x = 1.5$, however we need to discard 1.5.

Thus minimax decision boundary will be $x = 0.833$

Minimax error

Minimax error can be estimated while using the value of x for the region 0.5 to 0.833

$$P(error) = \int_{0.5}^{0.833} (x - 0.5)dx$$

$$= 0.5x^2 - 0.5x \Big|_{0.5}^{0.833}$$

$$= 0.5 \times 0.833^2 - 0.5 \times 0.833 - 0.5^3 + 0.5^2$$

$$P(error) = 0.055$$

c) We are given that maximum acceptable error rate for misclassifying ω_2 as ω_1 is 0.03 and we need to estimate the optimal decision boundary.

This use case matches well with the definition of Neyman Pearson criterion as we need to minimise the type 2 error(false positive) such that type 1 error(false negative) is under a threshold.

Error for classifying ω_2 as ω_1 is given as

$$\int_{0.5}^{x} p(\omega_2 | x) dx < 0.03$$

$$P(\omega_2) \int_{0.5}^{x} p(x | \omega_2) dx < 0.03$$

$$0.5 \int_{0.5}^{x} (x - 0.5) dx < 0.03$$

$$0.5x^2 - 0.5x \Big|_{0.5}^{x} < 0.06$$

$$0.5x^2 - 0.5x - 0.5^3 + 0.5^2 < 0.06$$

Solving for x, feasible value of $x = 0.84$

Similarly if the maximum acceptable error rate is 0.05, we can update our equation as follows

$$0.5 \int_{0.5}^{x} (x - 0.5) dx < 0.05$$

$$0.5x^2 - 0.5x \Big|_{0.5}^{x} < 0.1$$

$$0.5x^2 - 0.5x - 0.5^3 + 0.5^2 < 0.1$$

Solving for x, feasible value of $x = 0.94$

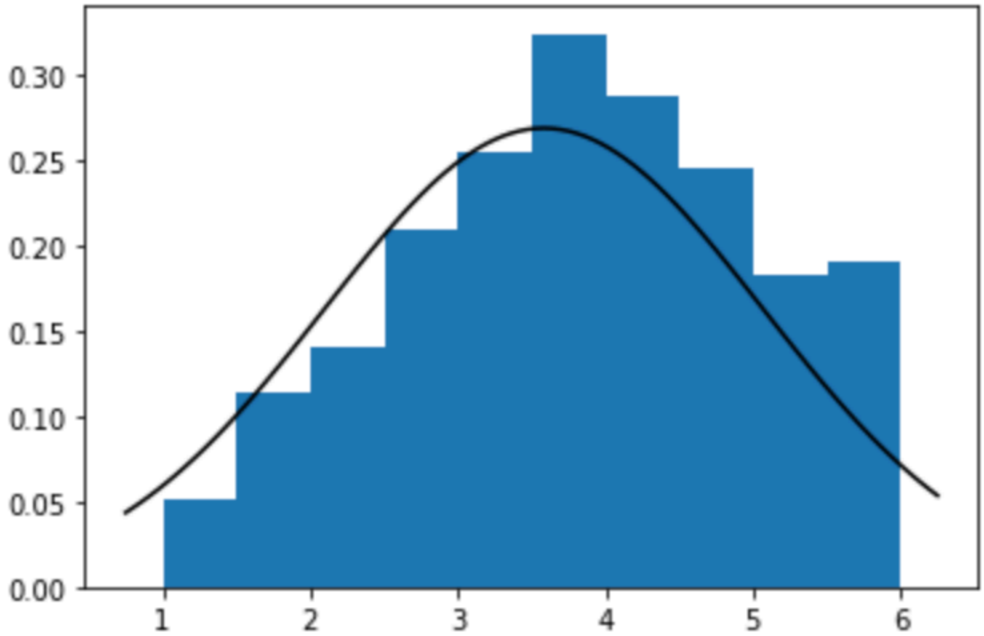
Solution 6

Code

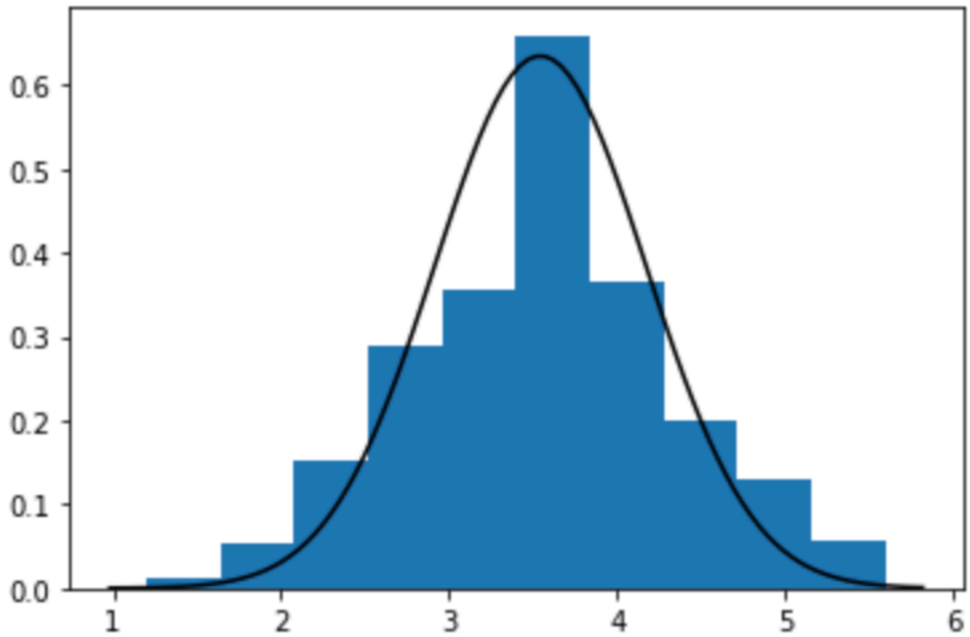
```
import random
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
import math
from collections import defaultdict

real_mean = np.mean(dice_values)
real_variance = np.var(dice_values)

print("Real mean",real_mean)
print("Real variance",real_variance)

mean_map = defaultdict(list)
for j in n:
    for i in range(N):

        n_sampled_values = np.random.choice(dice_values,j)
        mean_map[j].append(np.mean(n_sampled_values))

    mu_x = np.mean(mean_map[j])
    sigma_x2 = np.var(mean_map[j])

    print("For n = %s mu= %s,sigma= %s" %(j,mu_x,sigma_x2))

    plt.hist(mean_map[j], normed=1)
    xmin,xmax = plt.xlim()
    x = np.linspace(xmin,xmax,100)
    plt.plot(x, stats.norm.pdf(x, mu_x, sigma_x2),'k')
    #plt.ylabel('Sampling frequency')
    plt.show()
```

Figures

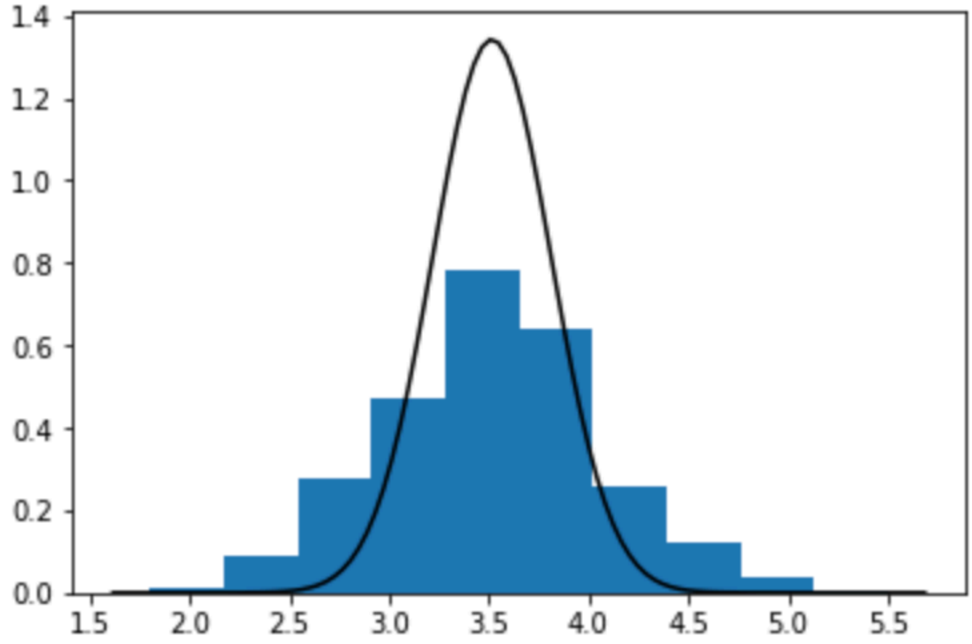
For n=2



For n=5



For n=10



Empirical Analysis

We computed theoretical mean and variance for the sampled dice values and we get mean = 3.5 and variance equal to 2.91667. By the results obtained from the code using CLT theorem and plotting histogram of mean sampled distribution for N counts for each n sampled values from the original distribution, we get a distribution that is gaussian and has mean around 3.5 and variance of about theoretical variance/n as shown by the numbers below

We have $\mu = 3.5$, $\sigma^2 = 2.91667$

For n = 2 , $\mu_x = 3.501$ and $\sigma_x^2 = 1.46$

For n = 5 , $\mu_x = 3.52$ and $\sigma_x^2 = 0.53$

For n = 10 , $\mu_x = 3.501$ and $\sigma_x^2 = 0.282$

We can observe that $\sigma_x^2 \approx \dfrac{\sigma^2}{n}$ and $\mu_x \approx \mu$

Thus we are able to show that using Central Limit theorem, we are able to convert a random discrete distribution of dice sample values to a gaussian distribution while sampling the means and effectively reducing the variance as n increases