EECE5643 – Homework 4

1.

a.

Two-Pass Algorithm:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{3}(1+6+2) = 3$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{1}{3}((1-3)^2 + (6-3)^2 + (2-3)^2) \approx 4.667$$

$$s = \sqrt{s^2} = \sqrt{4.667} \approx 2.16$$

One-Pass Algorithm:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{3}(1+6+2) = 3$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \; - \; \overline{x}^2 = \frac{1}{3}(1^2 + 6^2 + 2^2) - (3)^2 = \frac{41}{3} - 9 \approx 4.667$$

$$s = \sqrt{s^2} = \sqrt{4.667} \approx 2.16$$

Welford's algorithm:

$$\overline{x}_1 = \overline{x}_{i-1} + \frac{1}{i}\left(x_i - \overline{x}_{i-1}\right) = \overline{x}_0 + \frac{1}{1}\left(x_1 - \overline{x}_0\right) = 0 + \frac{1}{1}\left(1 - 0\right) = 1$$

$$v_1 = v_{i-1} + \left(\frac{i-1}{i}\right)\left(x_i - \overline{x}_{i-1}\right)^2 = v_0 = 0$$

$$\overline{x}_2 = \overline{x}_1 + \frac{1}{2}\left(x_2 - \overline{x}_1\right) = 1 + \frac{1}{2}\left(6 - 1\right) = 3.5$$

$$v_2 = v_1 + \frac{1}{2}\left(x_2 - \overline{x}_1\right)^2 = 0 + \frac{1}{2}\left(6 - 1\right)^2 = 12.5$$

$$\overline{x}_3 = \overline{x}_2 + \frac{1}{3}\left(x_3 - \overline{x}_2\right) = 3.5 + \frac{1}{3}\left(2 - 3.5\right) = 3$$

$$v_3 = v_2 + \frac{2}{3}\left(x_3 - \overline{x}_2\right)^2 = 12.5 + \frac{2}{3}\left(2 - 3.5\right)^2 = 14$$

$$\overline{x} = \overline{x}_3 = 3$$

$$s^2 = \frac{v_3}{3} = \frac{14}{3} \approx 4.667$$

$$s = \sqrt{s^2} = \sqrt{4.667} \approx 2.16$$

b.

Two-Pass Algorithm:

$$\overline{x} \;=\; \frac{1}{\tau}\int_0^{\tau} x(t)\,dt \;=\; \frac{1}{5}\big(\int_0^2 3\,dt + \int_2^5 8\,dt\big) = 6$$

$$s^2 \;=\; \frac{1}{t_n}\sum_{i=1}^{n}\big(x_i-\overline{x}\big)^2\,\delta_i \;=\; \frac{1}{5}((3-6)^2\cdot(2-0)+(8-6)^2\cdot(5-2)) = 6$$

$$s = \sqrt{s^2} = \sqrt{6} \approx 2.45$$

One-Pass Algorithm:

$$\overline{x} \;=\; \frac{1}{\tau}\int_0^{\tau} x(t)\,dt \;=\; \frac{1}{5}\big(\int_0^2 3\,dt + \int_2^5 8\,dt\big) = 6$$

$$s^2 \;=\; \Big(\frac{1}{t_n}\sum_{i=1}^{n} x_i^2\,\delta_i\Big) \;-\; \overline{x}^2 = \frac{1}{5}(3^2\cdot 2 + 8^2\cdot 3) - 6^2 = 6$$

$$s = \sqrt{s^2} = \sqrt{6} \approx 2.45$$

Welford's algorithm:

$$\overline{x}_i = \overline{x}_{i-1} + \frac{\delta_i}{t_i}\Big(x_i - \overline{x}_{i-1}\Big)$$

$$v_i = v_{i-1} + \delta_i\,\frac{t_{i-1}}{t_i}\Big(x_i - \overline{x}_{i-1}\Big)^2$$

$$s^2 = \frac{v_n}{t_n}$$

$$x_1 = 3, \quad \delta_1 = 2$$

$$\overline{x}_1 = x_1 = 3, \quad t_1 = \delta_1 = 2, \quad v_1 = 0$$

$$x_2 = 8, \quad \delta_2 = 3$$

$$t_2 = \delta_1 + \delta_2 = 2 + 3 = 5$$

$$\overline{x}_2 = \overline{x}_1 + \frac{\delta_2}{t_2}\Big(x_2 - \overline{x}_1\Big) = 3 + \frac{3}{5}(8-3) = 3 + \frac{3}{5}\cdot 5 = 3 + 3 = 6$$

$$v_2 = v_1 + \delta_2\,\frac{t_1}{t_2}\Big(x_2 - \overline{x}_1\Big)^2 = 0 + 3\cdot\frac{2}{5}\Big(8-3\Big)^2 = 3\cdot\frac{2}{5}\cdot(5)^2 = 3\cdot\frac{2}{5}\cdot 25 = 3\cdot\frac{50}{5} = 3\cdot 10 = 30$$

$$\overline{x} = \overline{x}_2 = 6$$

$$s^2 = \frac{v_2}{t_2} = \frac{30}{5} = 6$$

$$s = \sqrt{s^2} = \sqrt{6} \approx 2.45$$

2.

a.

S is the collection of ball counts for each of the 1000 boxes:

$$S = \{x_1, x_2, \ldots, x_{1000}\}$$

X is the set of all possible values that the data in S can take.As placing 2000 balls into 1000 boxes, X can be all integers from 0 to 2000.

$$X = \{0, 1, 2, \ldots, 2000\}$$

value	count	proportion
0.0	141	0.141
1.0	252	0.252
2.0	282	0.282
3.0	184	0.184
4.0	95	0.095
5.0	28	0.028
6.0	14	0.014
7.0	4	0.004

```
sample size ........... = 1000
mean .................. = 2.000
standard deviation .... = 1.399
```

b.

S is the collection of ball counts for each of the 1000 boxes:

$$S = \{x_1, x_2, \ldots, x_{1000}\}$$

X is the set of all possible values that the data in S can take.As placing 10,000 balls into 1000 boxes, X can be all integers from 0 to 10000.

$$X = \{0, 1, 2, \ldots, 10000\}$$

value	count	proportion
2.0	1	0.001
3.0	7	0.007
4.0	11	0.011
5.0	48	0.048
6.0	64	0.064
7.0	75	0.075
8.0	119	0.119
9.0	138	0.138
10.0	130	0.130
11.0	125	0.125
12.0	86	0.086
13.0	63	0.063
14.0	46	0.046
15.0	29	0.029
16.0	25	0.025
17.0	19	0.019
18.0	9	0.009
19.0	1	0.001
20.0	3	0.003
21.0	1	0.001

```
sample size ........... = 1000
mean .................. = 10.000
standard deviation .... = 3.118
```

c.

2000 balls:
Histogram mean $\overline{x} = 2$
Histogram standard deviation s = 1.399

10000 balls:
Histogram mean $\overline{x} = 10$
Histogram standard deviation s = 3.118

3.

a.

```
value   count   proportion

19.0    3       0.000
20.0    2       0.000
21.0    8       0.000
22.0    18      0.000
23.0    48      0.000
24.0    148     0.001
25.0    218     0.002
26.0    389     0.004
27.0    758     0.008
28.0    1224    0.012
29.0    1978    0.020
30.0    3053    0.031
31.0    4210    0.042
32.0    5699    0.057
33.0    7220    0.072
34.0    8629    0.086
35.0    9708    0.097
36.0    10480   0.105
37.0    10459   0.105
38.0    9456    0.095
39.0    8409    0.084
40.0    6599    0.066
41.0    4837    0.048
42.0    3097    0.031
43.0    1809    0.018
44.0    962     0.010
45.0    392     0.004
46.0    142     0.001
47.0    42      0.000
48.0    3       0.000

sample size .......... = 100000
mean ................. = 35.992
standard deviation .... = 3.765

Probability of passing (score >= 36): 0.567
```

$$E[\text{score} \mid \text{Class I}] = 4 \times 0.6 \ + \ 3 \times 0.3 \ + \ 2 \times 0.1 = 2.4 + 0.9 + 0.2 = 3.5$$

$$E[\text{score} \mid \text{Class II}] = 3 \times 0.1 \ + \ 2 \times 0.4 \ + \ 1 \times 0.4 \ + \ 0 \times 0.1 = 0.3 + 0.8 + 0.4 + 0 = 1.5$$

$$P(\text{Class I}) = \frac{90}{120} = 0.75$$

$$P(\text{Class II}) = \frac{30}{120} = 0.25$$

Each question, when chosen independently with Class I probability 0.75 and Class II probability 0.25, has an expected score of

$$E[\text{score}] = 0.75 \times 3.5 \ + \ 0.25 \times 1.5 = 2.625 + 0.375 = 3.0$$

For 12 questions, the expected total score is

$$12 \times 3.0 = 36$$

The histogram shows a mean of $35.992 \approx 36$.Additionally, the histogram has most of its mass between scores of about 30 and 42, with the highest frequencies around scores 36-37. This symmetry around 36 is consistent with the expectation as 36 is the mean.

S is the set of total scores obtained by taking the sum of the scores on 12 questions for each test. In the simulation, for $N=$ 100000 replications:

$$S = \{s_1, s_2, \ldots, s_N\}$$

where each $s_i = \sum_{j=1}^{12} a_{ij}$ and a_{ij} is the points for the j-th question in the i-th test. Each a_{ij} is determined by the question's class where for a Class I question, $a_{ij} \in \{2, 3, 4\}$ with probabilities 0.1, 0.3, and 0.6 and for a Class II question, $a_{ij} \in \{0, 1, 2, 3\}$ with probabilities 0.1, 0.4, 0.4, and 0.1.

X is the set of all possible values that the data in S can take. As each question is graded from 0 to 4 and there are 12 questions, the possible total scores range from 0 to $12 \times 4 = 48$

$$X = \{0, 1, 2, \ldots, 48\}$$

b.

Probability of passing (score ≥ 36): 0.567

4.

bin	midpoint	count	proportion	density
1	0.050	56	0.006	0.056
2	0.150	140	0.014	0.140
3	0.250	265	0.026	0.265
4	0.350	352	0.035	0.352
5	0.450	403	0.040	0.403
6	0.550	540	0.054	0.540
7	0.650	658	0.066	0.658
8	0.750	733	0.073	0.733
9	0.850	901	0.090	0.901
10	0.950	1009	0.101	1.009
11	1.050	970	0.097	0.970
12	1.150	804	0.080	0.804
13	1.250	717	0.072	0.717
14	1.350	625	0.062	0.625
15	1.450	548	0.055	0.548
16	1.550	467	0.047	0.467
17	1.650	369	0.037	0.369
18	1.750	230	0.023	0.230
19	1.850	168	0.017	0.168
20	1.950	45	0.004	0.045

```
sample size ........... = 10000
mean ........... = 1.000
stdev .......... = 0.409
```

$$S = \{x_1, x_2, \ldots, x_{10000}\}$$

$$X = [0, 2)$$