Problem 1: Variance Calculation Algorithms

a. Discrete Data Set

Data points: $x_1 = 1, x_2 = 6, x_3 = 2$. Number of samples $n = 3$.

1. Two-Pass Algorithm

$$\begin{aligned}
\overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{3}(1+6+2) = 3 \\
s^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{3}\left[(1-3)^2 + (6-3)^2 + (2-3)^2\right] = \frac{1}{3}[4+9+1] = \frac{14}{3} \approx 4.667 \\
s &= \sqrt{4.667} \approx 2.16
\end{aligned}$$

2. One-Pass Algorithm

$$\begin{aligned}
\overline{x} &= 3 \\
s^2 &= \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \overline{x}^2 = \frac{1}{3}(1^2+6^2+2^2) - (3)^2 = \frac{1}{3}(1+36+4) - 9 = \frac{41}{3} - 9 = \frac{14}{3} \approx 4.667 \\
s &= \sqrt{4.667} \approx 2.16
\end{aligned}$$

3. Welford's Algorithm

Update formulas: $\overline{x}_i = \overline{x}_{i-1} + \frac{1}{i}(x_i - \overline{x}_{i-1})$ and $v_i = v_{i-1} + \frac{i-1}{i}(x_i - \overline{x}_{i-1})^2$.

Step 1 ($x_1 = 1$):

$$\begin{aligned}
\overline{x}_1 &= 0 + \frac{1}{1}(1-0) = 1 \\
v_1 &= 0
\end{aligned}$$

Step 2 ($x_2 = 6$):

$$\begin{aligned}
\overline{x}_2 &= 1 + \frac{1}{2}(6-1) = 3.5 \\
v_2 &= 0 + \frac{1}{2}(6-1)^2 = 12.5
\end{aligned}$$

Step 3 ($x_3 = 2$):

$$\begin{aligned}
\overline{x}_3 &= 3.5 + \frac{1}{3}(2-3.5) = 3 \\
v_3 &= 12.5 + \frac{2}{3}(2-3.5)^2 = 12.5 + \frac{2}{3}(2.25) = 12.5 + 1.5 = 14
\end{aligned}$$

Final Result:

$$s^2 = \frac{v_3}{3} = \frac{14}{3} \approx 4.667$$
$$s = 2.16$$

b. Time-Weighted Data

Data: $x(t) = 3$ for $t \in [0, 2)$, $x(t) = 8$ for $t \in [2, 5)$. Total time $\tau = 5$.

1. Two-Pass Algorithm

$$\bar{x} = \frac{1}{\tau}\int_0^\tau x(t)\,dt = \frac{1}{5}[3(2) + 8(3)] = \frac{6+24}{5} = 6$$
$$s^2 = \frac{1}{\tau}\sum \delta_i (x_i - \bar{x})^2 = \frac{1}{5}[2(3-6)^2 + 3(8-6)^2] = \frac{1}{5}[2(9) + 3(4)] = \frac{18+12}{5} = 6$$
$$s = \sqrt{6} \approx 2.45$$

2. One-Pass Algorithm

$$s^2 = \left(\frac{1}{\tau}\sum \delta_i x_i^2\right) - \bar{x}^2 = \frac{1}{5}[2(3^2) + 3(8^2)] - 6^2 = \frac{1}{5}[18 + 192] - 36 = \frac{210}{5} - 36 = 42 - 36 = 6$$
$$s = \sqrt{6} \approx 2.45$$

3. Welford's Algorithm (Time-Weighted)

Formulas: $t_i = t_{i-1} + \delta_i$, $\bar{x}_i = \bar{x}_{i-1} + \frac{\delta_i}{t_i}(x_i - \bar{x}_{i-1})$, $v_i = v_{i-1} + \delta_i \frac{t_{i-1}}{t_i}(x_i - \bar{x}_{i-1})^2$.
Step 1 ($x_1 = 3, \delta_1 = 2$):

$$t_1 = 2, \quad \bar{x}_1 = 3, \quad v_1 = 0$$

Step 2 ($x_2 = 8, \delta_2 = 3$):

$$t_2 = 2 + 3 = 5$$
$$\bar{x}_2 = 3 + \frac{3}{5}(8-3) = 3 + 3 = 6$$
$$v_2 = 0 + 3\left(\frac{2}{5}\right)(8-3)^2 = 3(0.4)(25) = 30$$

Final Result:

$$s^2 = \frac{v_2}{t_2} = \frac{30}{5} = 6$$
$$s = 2.45$$

Problem 2: Balls in Boxes Simulation

a. 2000 Balls

Let S be the collection of ball counts for each of the 1000 boxes:

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

The set of possible values X ranges from 0 to the total number of balls:

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

```
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
```

Figure 1: Histogram of ball counts (2000 balls).

b. 10,000 Balls

Similarly, for 10,000 balls:

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

```
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
```

Figure 2: Histogram of ball counts (10,000 balls).

c. Statistics Comparison

- **2000 Balls:** Mean $\bar{x} = 2$, Std Dev $s \approx 1.399$.
- **10000 Balls:** Mean $\bar{x} = 10$, Std Dev $s \approx 3.118$.

Note that for a Poisson distribution (which approximates this process), the variance equals the mean ($\sigma^2 \approx \mu$). Here $\sqrt{2} \approx 1.414$ and $\sqrt{10} \approx 3.162$, which match the observed standard deviations closely.

Problem 3: Exam Score Simulation

a. Expected Value and Histogram

```
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
```

Figure 3: Distribution of Total Exam Scores.

Expected Value Calculation:

$$
\begin{aligned}
E[\text{score} \mid \text{Class I}] &= 4(0.6) + 3(0.3) + 2(0.1) = 2.4 + 0.9 + 0.2 = 3.5 \\
E[\text{score} \mid \text{Class II}] &= 3(0.1) + 2(0.4) + 1(0.4) + 0(0.1) = 0.3 + 0.8 + 0.4 = 1.5 \\
P(\text{Class I}) &= \frac{90}{120} = 0.75, \quad P(\text{Class II}) = 0.25 \\
E[\text{question}] &= 0.75(3.5) + 0.25(1.5) = 2.625 + 0.375 = 3.0 \\
E[\text{Total Score}] &= 12 \times 3.0 = 36
\end{aligned}
$$

The histogram shows a mean of $35.992 \approx 36$, consistent with the calculation.

Definitions:

- $S = \{s_1, \ldots, s_N\}$ is the set of total scores from $N = 100,000$ simulations.
- $X = \{0, 1, \ldots, 48\}$ is the set of possible total scores (12 questions $\times$ max 4 points).

b. Passing Probability

The probability of passing (score ≥ 36) is observed to be **0.567**.

Problem 4: Continuous Distribution Simulation

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
```

Figure 4: Histogram of simulated values.

Data Definition:

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

The range of the variable appears to be $X = [0, 2)$.