ENGR 510 Neural Networks Homework

Anthony Su

November 8, 2024

1 Intro to Probability

1.a

Using a brute force enumeration technique, the result is

$$\boxed{\mathrm{P}(S = 10) = 0.125}$$

1.b

Due to the central limit theorem, this system can be modeled as normally distributed. For a single roll with $\omega = \{1, 2, 3, 4, 5, 6\}$,

$$\begin{aligned} \mathrm{E}(X_i) &= \frac{1}{6}\sum_{i=1}^{6} \omega_i \\ \mathrm{E}(X_i) &= 3.5 \end{aligned}$$

and the variance is

$$\begin{aligned} \mathrm{Var}(X_i) &= \frac{1}{6}\sum_{i=1}^{6} (\omega_i - 3.5)^2 \\ \mathrm{Var}(X_i) &= \frac{35}{12} \end{aligned}$$

Thus, the mean and variance for the average of 100 rolls can be modeled by $\mathcal{N}\left(3.5, \sqrt{\frac{35}{12}}\right)$. For the sum of 100 rolls, this becomes $\mathcal{N}\left(350, \sqrt{100 \cdot \frac{35}{12}}\right)$.

Integrating the PDF of this distribution from 299.5 to 300.5 (one unit around 300) yields a probability of approximately 0.03% to roll a sum of exactly 300. Integrating the PDF of this distribution from $-\infty$ to 299.5 yields a probability of approximately 0.16% to roll a sum of ≤ 300.

$$\boxed{\mathrm{P}(S = 300) \approx 0.03\%}$$

$$\boxed{\mathrm{P}(S < 300) \approx 0.16\%}$$

1.c

There are 6^5 ways to roll 5 dice. There are 5^5 ways to roll no ones for each of of the 5 dice. There are 5^4 ways to roll 1 one with any given die. Thus, there are 5×5^4 ways to roll 1 one out of 5 rolls.

The probability of rolling 2 or more ones is

$$\text{P(snake eyes)} = 1 - \text{P(0 ones)} - \text{P(1 one)}$$

Substituting in values,

$$\text{P(snake eyes)} = 1 - \frac{5^5}{6^5} - \frac{5 \cdot 5^4}{6^5}$$
$$\text{P(snake eyes)} = 1 - 2 \cdot \left(\frac{5}{6}\right)^5$$
$$\boxed{\text{P(snake eyes)} \approx 19.62\%}$$

1.d

There are $\binom{6}{2} = 15$ ways to place the two numbers within the license plate. For each placement of letters and numbers, there are $26^4 * 10^2$ ways to assign letters and numbers. Thus, the total number of possible license plates is

$$|\Omega| = \binom{6}{2} \cdot 26^4 \cdot 10^2$$
$$\boxed{|\Omega| = 685,464,000}$$

1.e

The number of license plates with no repeated characters is

$$|\Omega'| = \binom{6}{2} \cdot \frac{26!}{(26-4)!} \cdot \frac{10!}{(10-2)!}$$

Thus, the number of license plates with repeated characters is

$$|\Omega| - |\Omega'| = \binom{6}{2} \left(26^4 \cdot 10^2 - \frac{26!}{(26-4)!} \cdot \frac{10!}{(10-2)!}\right)$$
$$\boxed{|\Omega| - |\Omega'| = 201,084,000}$$

1.f

There are four ways for a license plate to have the word "CAT":

Table 1: Locations of "CAT" on license plates

C A T _ _ _
_ C A T _ _
_ _ C A T _
_ _ _ C A T

For each of those ways, the remaining letter can be placed in one of the three remaining slots; there are $\binom{3}{1} = 3$ ways to do this. For each of those ways, the number of outcomes for the remaining one letter and two numbers is $26 * 10 * 10 = 2600$. The total number of ways for the license plate to have the word "CAT" is

$$|\Omega'| = 4 \cdot 3 \cdot 2600$$

Thus, the number of license plates without the word "CAT" is

$$|\Omega| - |\Omega'| = \binom{6}{2} \cdot 26^4 \cdot 10^2 - 4 \cdot 3 \cdot 2600$$
$$\boxed{|\Omega| - |\Omega'| = 685,432,800}$$

2 Birthday Problem

2.a

$$P(\geq 2 \text{ share}) = \prod_{k=0}^{N-1} \frac{365 - k}{365}$$

There must be at least 32 people for the chance of a shared birthday to be over 75%.

2.b

There are many ways for people to share birthdays without forming a triplet. This precludes the traditional birthday problem approach of further subtracting the probabilities of all outcomes with exclusively birthday pairs. Thus, in lieu of an exact/analytical solution,

I would approach this problem with a Monte Carlo simulation.

3 Monte Carlo

3.a

The chance of landing in the largest sphere that fits inside the cube is equal to the ratio of the volumes of the sphere and the cube:

$$P(\text{inside}) = \frac{(1/6)\pi L^3}{L^3}$$
$$\boxed{P(\text{inside}) = \frac{\pi}{6}}$$

3.b

```
def sample_hypercube(n):
    x = np.random.random_sample(size=n) * 2 - 1  # rescale to [-1, 1]
    r = np.linalg.norm(x)
    if r < 1:
        return True
    elif r > 1:
        return False
```

3.c

```
N_samples = 10000
N_in = 0
for i in range(N_samples):
    if sample_hypercube(3):
        N_in += 1
print(f'Number of points within sphere for 3D cubic domain: {N_in}')
```

Out of 10000 samples, 5118 were within the sphere, which is 51.18%.

3.d

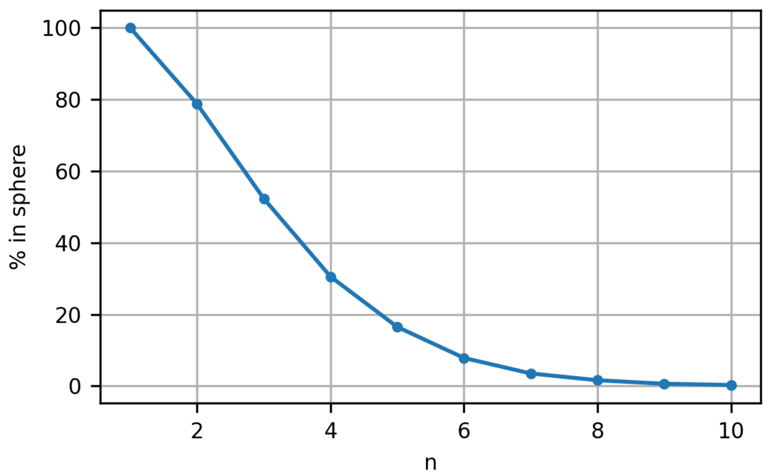

Figure 1: Chance of being in the unit sphere for n dimensions

As n increases, the probability of sampling a point within a unit radius of the origin rapidly approaches zero.

3.e

Each value X_i in Figure 1 is a realization of a binomially distributed random variable X:

$$X \sim \text{Binomial}(N, p)$$

From the law of large numbers, we know $\mu = p = \bar{X}_i$. The variance can be computed for a given n by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

This variance can then be used in the normal approximation to compute 95% confidence intervals.

Table 2: 95% confidence intervals for various n

n	σ^2	upper bound (%)	lower bound (%)
1	0.0	100	100
2	1.64e-05	78.62	80.20
3	2.49e-05	51.73	53.69
4	2.16e-05	30.63	32.45
5	1.41e-05	16.32	17.80
6	7.35e-06	7.46	8.52
7	3.69e-06	3.46	4.22
8	1.55e-06	1.33	1.81
9	6.66e-07	0.51	0.83
10	2.39e-07	0.14	0.34

Note that σ^2 is approximated as $\hat{\sigma}^2/N$.

Note that at higher values of n, the normal distribution becomes a poorer approximation to the binomial distribution since $p \ll 1$. However, with the large sample size of $N = 10,000$, this issue is somewhat mitigated.

4 Central Limit Theorem

4.a

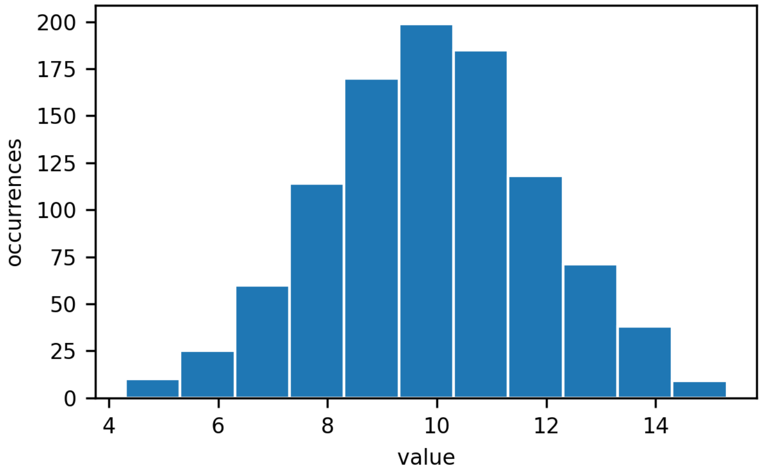

Figure 2: Example normally-distributed sequence with $n = 1000$

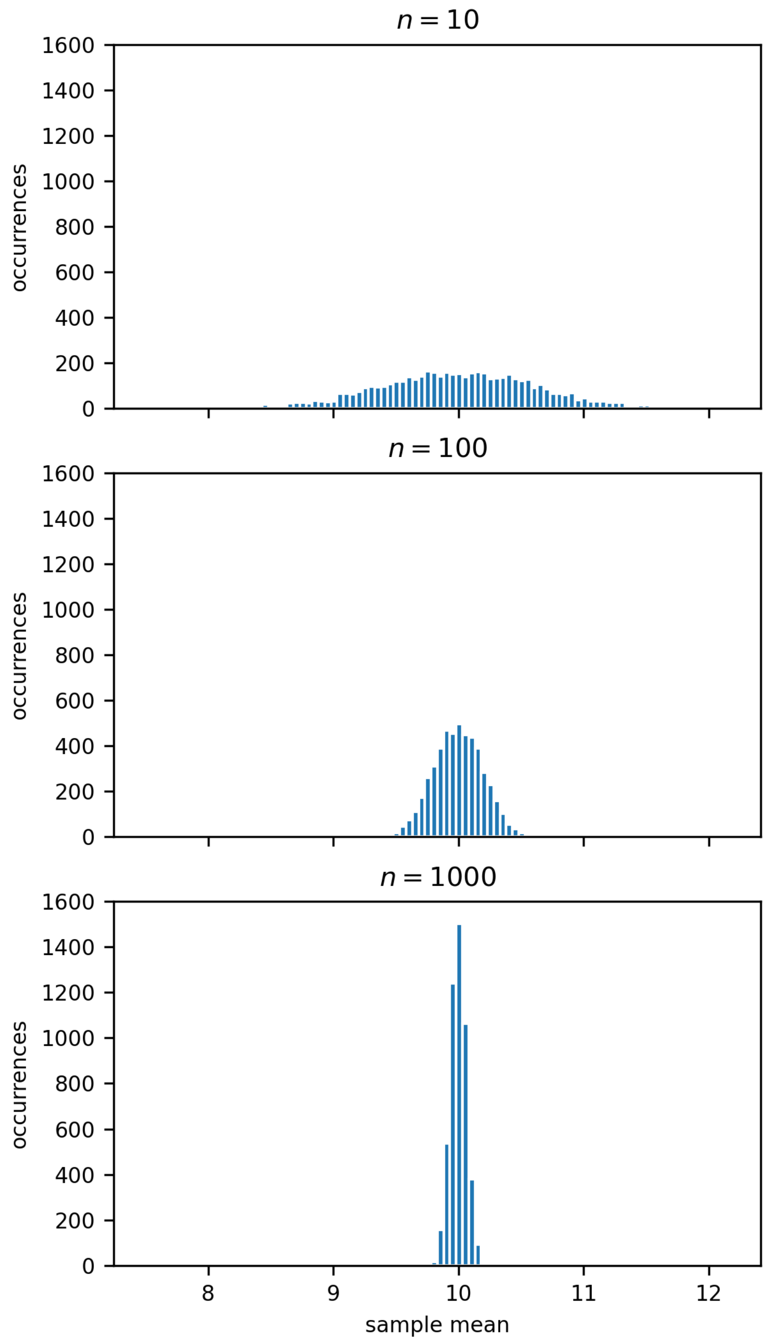

Figure 3: Sample means for normally-distributed sequences with $n = \{10, 100, 1000\}$

4.b

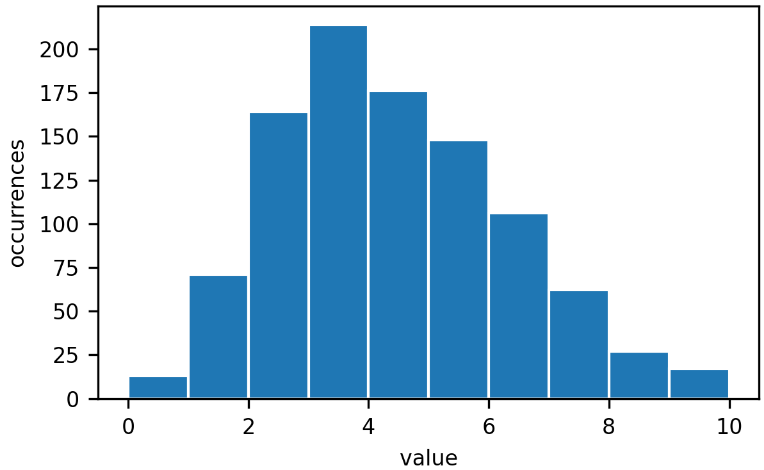

Figure 4: Example Poisson-distributed sequence with $n = 1000$

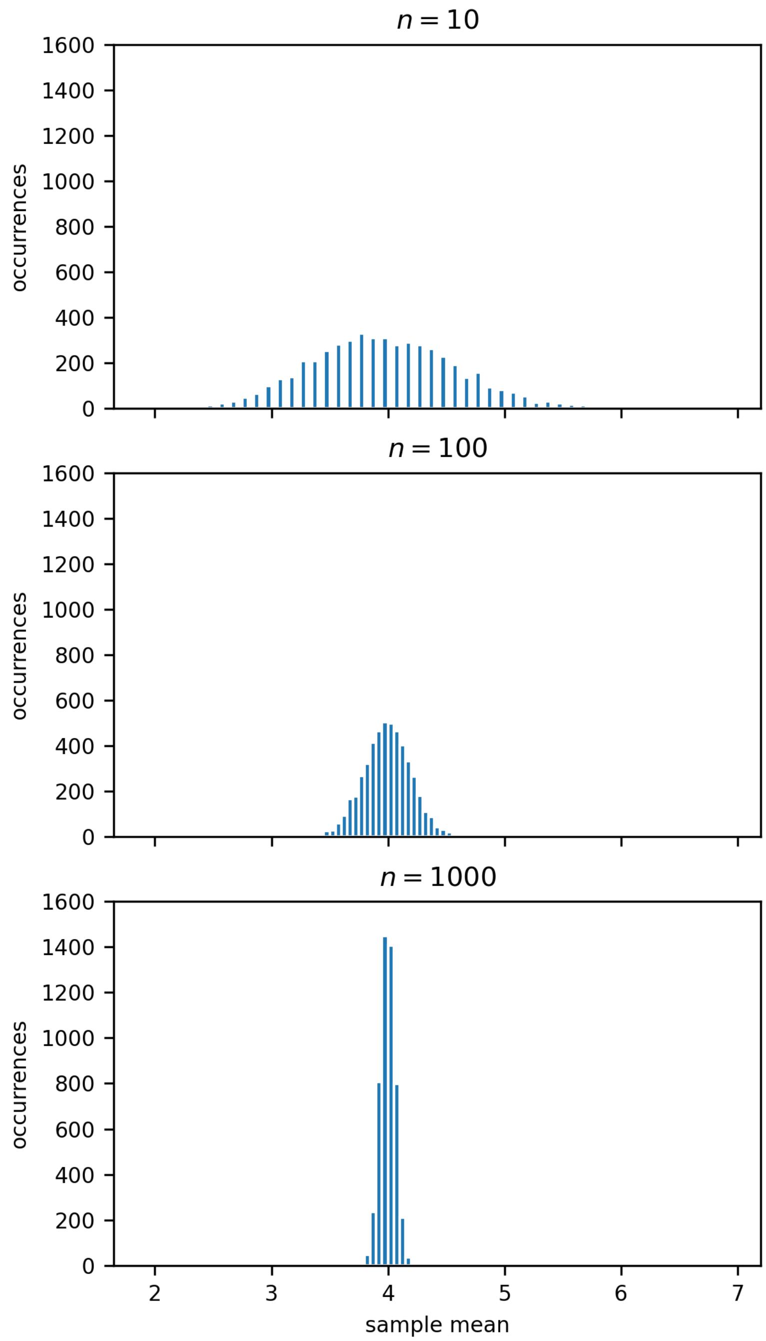

Figure 5: Sample means for Poisson-distributed sequences with $n = \{10, 100, 1000\}$

4.c

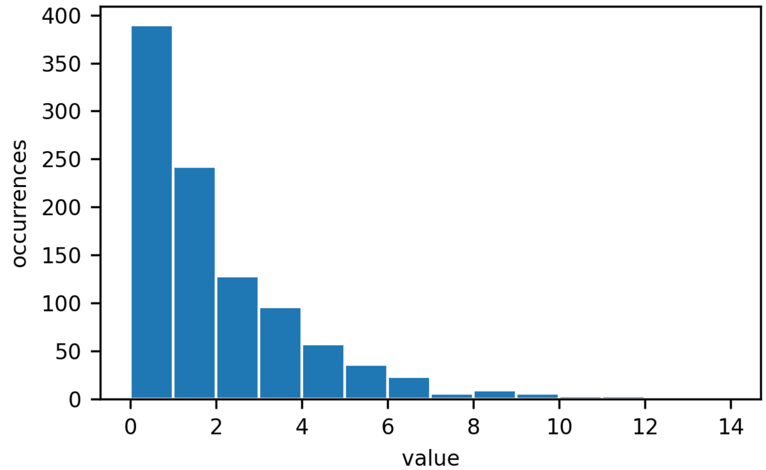

Figure 6: Example Gamma-distributed sequence with $n = 1000$

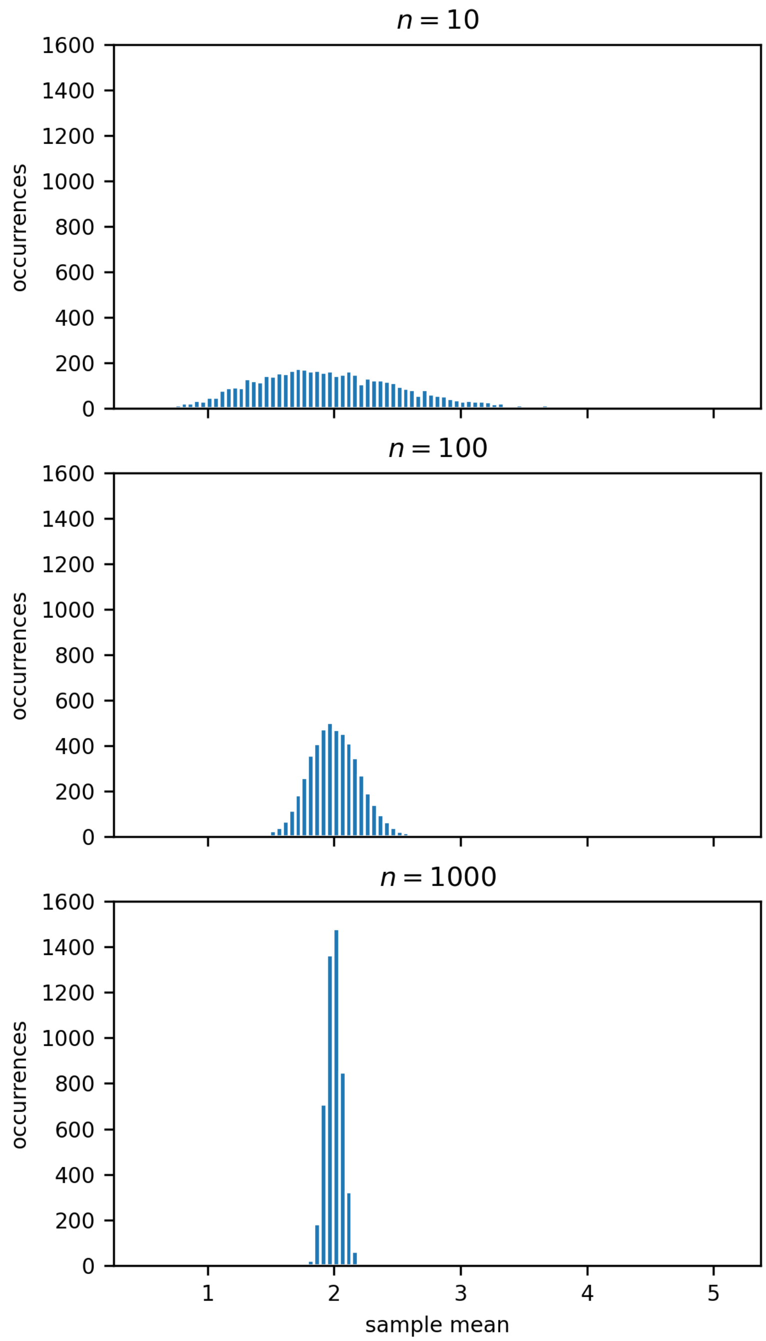

Figure 7: Sample means for Gamma-distributed sequences with $n = \{10, 100, 1000\}$

5 Conditional Probability

$$P(\text{defect}) = 0.05 \qquad P(\text{no defect}) = 0.95$$
$$P(\text{positive}|\text{defect}) = 0.9 \qquad P(\text{negative}|\text{defect}) = 0.1$$
$$P(\text{positive}|\text{no defect}) = 0.1 \qquad P(\text{negative}|\text{no defect}) = 0.9$$

5.a

$$P(\text{defect}|\text{positive}) = \frac{P(\text{positive}|\text{defect}) \cdot P(\text{defect})}{P(\text{positive}|\text{defect}) \cdot P(\text{defect}) + P(\text{positive}|\text{no defect}) \cdot P(\text{no defect})}$$
$$P(\text{defect}|\text{positive}) = \frac{0.9 \cdot 0.05}{0.9 \cdot 0.05 + 0.1 \cdot 0.95}$$
$$\boxed{P(\text{defect}|\text{positive}) = 32.14\%}$$

5.b

$$P(\text{no defect}|\text{positive}) = 1 - P(\text{defect}|\text{positive})$$
$$\boxed{P(\text{no defect}|\text{positive}) = 67.86\%}$$

5.c

As given,

$$\boxed{P(\text{positive}|\text{no defect}) = 0.1}$$

6 Statistical Data Analysis

6.a

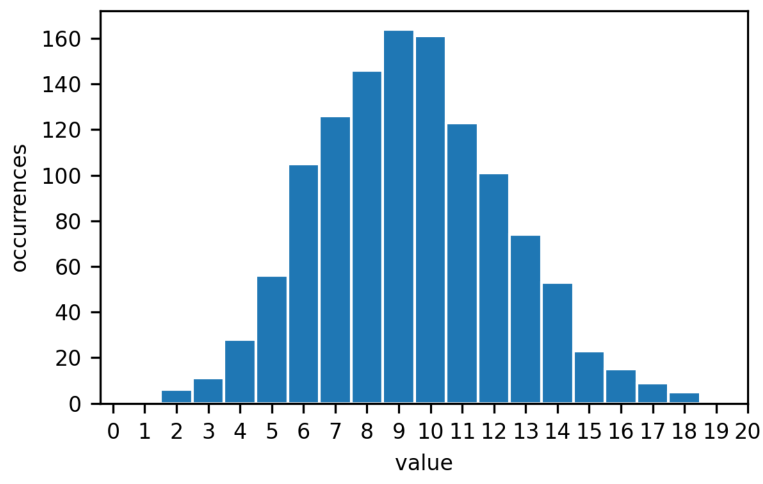

Figure 8: Problem 6 data

6.b

The sample mean is $\bar{X} = 9.36$ and the sample variance is $\hat{\sigma}^2 = 8.62$.

6.c

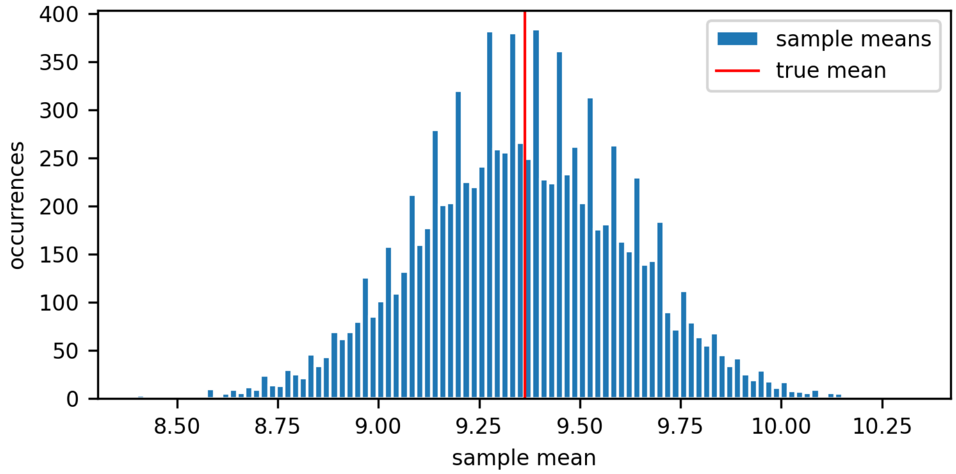

Figure 9: 10% subset sample means

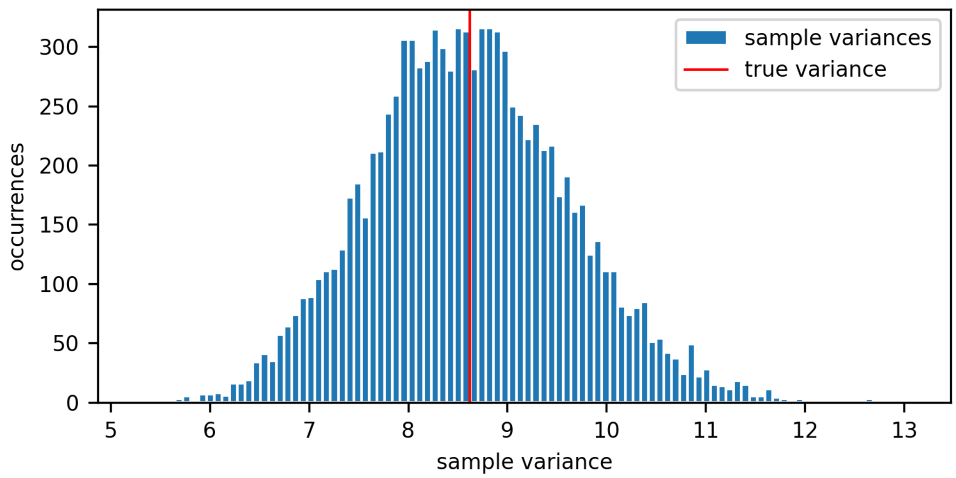

Figure 10: 10% subset sample variances

This is consistent with the expectation from the central limit theorem because the sample means and variances are normally distributed about the true mean and variance.

6.d

6.d.1 Method of Moments

Assume a sample X_i is distributed as $X_i \sim \text{Poisson}(\lambda)$.

λ is defined as

$$\lambda = \mu_1$$
$$\lambda = \mathrm{E}(X_i)$$

Due to the law of large numbers, this is

$$\lambda = \bar{X}_i$$
$$\lambda = 9.36$$

Using the method of moments, $X_i \sim \text{Poisson}(9.36)$

6.d.2 MLE

Assume a sample X_i is distributed as $X_i \sim \text{Poisson}(\lambda)$.

The distribution function is

$$f(x_1, \ldots, x_n|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

The log-likelihood function is

$$l(X_1, \ldots, X_n|\lambda) = \log\left(f(X_1, \ldots, X_n|\lambda)\right)$$
$$l(X_1, \ldots, X_n|\lambda) = \log\left(\prod_{i=1}^{n} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}\right)$$
$$l(X_1, \ldots, X_n|\lambda) = \sum_{i=1}^{n} \left[X_i \log(\lambda) - \lambda - \log(X_i!\,)\right]$$

Taking the derivative,

$$\frac{\partial l}{\partial \lambda}(X_1, \ldots, X_n|\lambda) = \sum_{i=1}^{n} \left[\frac{1}{\lambda} X_i - 1\right]$$
$$\frac{\partial l}{\partial \lambda}(X_1, \ldots, X_n|\lambda) = \frac{n}{\lambda}\bar{X}_i - n$$

Equating the derviative to zero,

$$\frac{\partial l}{\partial \lambda}(X_1, \ldots, X_n|\lambda) = 0$$
$$\frac{n}{\lambda}\bar{X}_i - n = 0$$
$$\bar{X}_i = \lambda$$

Thus, $\lambda = \bar{X}_i$ is a critical point of the distribution log-likelihood function.

Taking the second derivative,

$$\frac{\partial^2 l}{\partial \lambda^2 \partial(} X_1, \ldots, X_n|\lambda) = \frac{n}{\lambda}\bar{X}_i - n$$
$$\frac{\partial^2 l}{\partial \lambda^2 \partial(} X_1, \ldots, X_n|\lambda) = -\frac{n}{\lambda^2}\bar{X}_i$$

Evaluating the second derivative at the critical point,

$$\frac{\partial^2 l}{\partial \lambda^2 \partial(} X_1, \ldots, X_n | \lambda = \bar{X}_i) = -\frac{n}{\bar{X}_i^2} \bar{X}_i$$
$$\frac{\partial^2 l}{\partial \lambda^2 \partial(} X_1, \ldots, X_n | \lambda = \bar{X}_i) = -\frac{n}{\bar{X}_i}$$

Since $\bar{X}_i > 0$, this second derivative is negative. Since $\lambda = \bar{X}_i$ is a critical point of $l(X_1, \ldots, X_n|\lambda)$ with negative second derivative, it is a maximum.

Thus, Poisson(9.36) is the maximum-likelihood estimate of X_i.

6.e

Since the data is already integer-valued, treat each integer value (starting from 0) as a 1-wide bin. One exception is that $[18, \infty)$ will be treated as one bin. Define the test χ^2 statistic as

$$\chi^2 = \sum_{i=0}^{18} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the number of observed i values and E_i is the value of the poisson distribution at i.

For the full dataset, $\chi^2 = 7.92$ which corresponds to $p = 0.97$.

This means that the result is consistent with the null hypothesis.

6.f

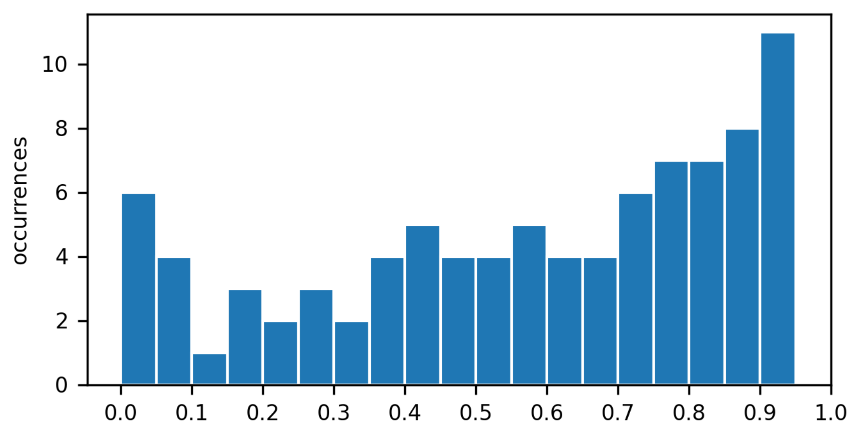

Figure 11: 10% subset sample p-values

98 out of 100 samples confirm the null hypothesis while 2 out of 100 samples reject the null hypothesis.

This is consistent with the full dataset's p-value of $p = 0.97$ which indicates a false positive rate of about 3%.