Student Information

Name : Feyza Özlü

ID : e2580868

Answer 1

a)

To find the value of k, we can use the fact that the joint probability density function $f_{X,Y}(x,y)$ must integrate to 1 over the entire range of possible values for x and y, which is $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Mathematically, this can be expressed as:

$$\int_0^1 \int_0^1 (x + ky^3)\, dx\, dy = 1$$

Let's calculate this integral and solve for k:

$$\begin{aligned}
\int_0^1 \int_0^1 (x + ky^3)\, dx\, dy &= \int_0^1 \left(\frac{x^2}{2} + kxy^3\right)\Big|_0^1 dy \\
&= \int_0^1 \left(\frac{1}{2} + ky^3\right) - \left(\frac{0}{2} + k \cdot 0 \cdot y^3\right) dy \\
&= \int_0^1 \left(\frac{1}{2} + ky^3\right) dy \\
&= \left(\frac{y}{2} + \frac{k}{4}y^4\right)\Big|_0^1 \\
&= \left(\frac{1}{2} + \frac{k}{4}\right) - (0 + 0) \\
&= \frac{1}{2} + \frac{k}{4}
\end{aligned}$$

We know that this integral must equal 1:

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

Now, let's solve for k:

$$\begin{aligned}
\frac{k}{4} &= 1 - \frac{1}{2} = \frac{1}{2} \\
k &= \frac{4}{2} = 2
\end{aligned}$$

So, the value of k is 2.

b)

We know that, from the textbook page 75 "For all continuous variables, the probability mass function is always equal to zero, $P(x) = 0$ for all x." Since X is a continuous random variable, the probability that X is equal to a specific value (like $x = \frac{1}{2}$ in our case) is zero. So, $P(X = \frac{1}{2}) = 0$ for continuous random variable X.

c)

To find $P(0 \le X \le \frac{1}{2}, 0 \le Y \le \frac{1}{2})$, we need to integrate the joint probability density function $f_{X,Y}(x, y)$ over the specified region.

Given that the joint probability density function is:

$$f_{X,Y}(x,y) = \begin{cases} x + 2y^3 & 0 \le x \le 1,\, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

So, the probability $P(0 \le X \le \frac{1}{2}, 0 \le Y \le \frac{1}{2})$ can be calculated by integrating $f_{X,Y}(x, y)$ over this region:

$$P(0 \le X \le \frac{1}{2}, 0 \le Y \le \frac{1}{2}) = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} (x + 2y^3)\, dx\, dy$$

Let's compute this integral:

$$\begin{aligned}
&\int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} (x + 2y^3)\, dx\, dy \\
&= \int_0^{\frac{1}{2}} \left(\frac{x^2}{2} + 2xy^3\right)\Big|_0^{\frac{1}{2}} dy \\
&= \int_0^{\frac{1}{2}} \left(\frac{1}{8} + 2 \cdot \frac{1}{2} \cdot y^3\right) - \Big(0 + 2 \cdot 0 \cdot y^3\Big) dy \\
&= \int_0^{\frac{1}{2}} \left(\frac{1}{8} + y^3\right) dy \\
&= \left(\frac{1}{8} \cdot y + \frac{1}{4} \cdot y^4\right)\Big|_0^{\frac{1}{2}} \\
&= \left(\frac{1}{8} \cdot \frac{1}{2} + \frac{1}{4} \cdot \left(\frac{1}{2}\right)^4\right) - \left(\frac{1}{8} \cdot 0 + \frac{1}{4} \cdot 0^4\right) \\
&= \frac{1}{16} + \frac{1}{64} \\
&= \frac{5}{64}
\end{aligned}$$

So, $P(0 \le X \le \frac{1}{2}, 0 \le Y \le \frac{1}{2}) = \frac{5}{64}$.

Answer 2

a)

To find the marginal pdf of Y we can integrate the joint pdf over all possible values of X:

$$f_Y(y) = \int_0^\infty f_{X,Y}(x,y)\,dx$$

$$= \int_0^\infty \frac{e^{(-y-\frac{x}{y})}}{y}\,dx$$

Let's make the substitution $u = \frac{x}{y}$, we get $du = \frac{dx}{y}$

When x is 0, u will be 0 and when x is ∞ u will be also ∞. So our boundaries will not change.

$$= \int_0^\infty e^{(-y-u)}\,du$$

$$= e^{-y}\int_0^\infty e^{-u}\,du$$

$$= e^{-y}\Big(-e^{-u}\Big)\Big|_0^\infty$$

$$= e^{-y}\Big((-e^{-\infty}) - (-e^{-0})\Big)$$

$$= e^{-y}(0-(-1))$$

$$= e^{-y}(0+1)$$

$$= e^{-y}$$

So, the marginal pdf of Y is $f_Y(y) = e^{-y}$ for $0 < y < \infty$.

From the textbook page 84 we know that pdf of an exponential distribution can be calculated by this formula : $f_X(x) = \lambda e^{-\lambda x}, x > 0$. When we look at the formula we derived for marginal pdf of Y we can clearly see that, this is the pdf of an exponential distribution with parameter $\lambda = 1$. Therefore, Y follows exponential distribution family.

b)

Again from the textbook page 84 we know that we can calculate expected value of an exponential distribution by this formula : $E(x) = \frac{1}{\lambda}$. For our case, we knot that $\lambda = 1$ from part b. So:

$$E(y) = \frac{1}{\lambda} = \frac{1}{1} = 1$$

So, the expected value of Y is 1.

Answer 3

a)

For this question I will use Central Limit Theorem. We can think this distribution like a binomial distribution which is a soldier can be either a naval soldier or not-naval soldier. And then we can use binomial to normal distribution to approximate this possibility. Firstly we know that possibility of a soldier is a naval soldier is $p = 0.1$ and possibility of a soldier not-naval (army soldier or air soldier) is $q = 1 - p = 0.9$. Since n=1000 which is a very large number than 30 we can use central limit theorem.

In this case our expected value is $\mu = n \cdot p = 1000 \cdot 0.1 = 100$ and standard deviation is $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{1000 \cdot 0.1 \cdot (1-0.1)} = \sqrt{90}$.

Since we want at least 9% of the soldiers belong to the naval forces which means we want at least $1000 \cdot 9\% = 1000 \cdot \frac{9}{100} = 90$ soldiers belong to the naval forces.

Now we can make binomial to normal distribution change:

$$
\begin{aligned}
P\{X \geq 90\} = 1 - P\{X < 90\} &= 1 - P\{X < 89.5\} \quad \text{explanation-1 below} \\
&= 1 - P\{Z < \frac{X-\mu}{\sigma}\} \\
&= 1 - P\{Z < \frac{89.5-100}{\sqrt{90}}\} \\
&= 1 - P\{Z < -1.106\} \\
&= 1 - \phi(-1.106) \\
&\approx 0.866
\end{aligned}
$$

explanation-1 : I applied correction continuity here, as binomial distribution is discrete but normal distribution is continuous, this is needed.

If we look at the value of $1 - \phi(-1.106)$ from octave online by using this code $1 - \mathit{normcdf}(-1.106)$ we get the result as $0.8656 \approx 0.866$.

b)

If we do the same calculation with only changes is 2000 soldiers instead of 1000:

In this case our expected value is $\mu = n \cdot p = 2000 \cdot 0.1 = 200$ and standard deviation is $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{2000 \cdot 0.1 \cdot (1-0.1)} = \sqrt{180}$.

Since we want at least 9% of the soldiers belong to the naval forces which means we want at least $2000 \cdot 9\% = 2000 \cdot \frac{9}{100} = 180$ soldiers belong to the naval forces.

Now we can make binomial to normal distribution change:

$$
\begin{aligned}
P\{X \geq 180\} = 1 - P\{X < 180\} &= 1 - P\{X < 179.5\} \quad \text{explanation-2 below} \\
&= 1 - P\{Z < \frac{X - \mu}{\sigma}\} \\
&= 1 - P\{Z < \frac{179.5 - 200}{\sqrt{180}}\} \\
&= 1 - P\{Z < -1.528\} \\
&= 1 - \phi(-1.528) \\
&\approx 0.937
\end{aligned}
$$

explanation-2 : again I applied correction continuity here, as binomial distribution is discrete but normal distribution is continuous, this is needed.

If we look at the value of $1 - \phi(-1.528)$ from octave online by using this code $1 - normcdf(-1.528)$ we get the result as $0.9367 \approx 0.937$.

Since we can clearly see the probability is increase. The reason for this is with a larger sample size (2000 soldiers instead of 1000 soldiers in our case), effect of randomness will decrease.

Answer 4

a)

For this question let me represent the lifespan of an randomly selected African bush elephant with continuous random variable X. It is given that the mean of X is 65 and standard deviation of X is 6. So we know that $\mu = 65$ and $\sigma = 6$. Now we need to find $P\{60 < X < 75\}$. In order to find it let's standardize X to Z.

$$
\begin{aligned}
P\{60 < X < 75\} &= P\left\{\frac{60-\mu}{\sigma} < Z < \frac{75-\mu}{\sigma}\right\} \\
&= P\left\{\frac{60-65}{6} < Z < \frac{75-65}{6}\right\} \\
&= P\left\{\frac{-5}{6} < Z < \frac{10}{6}\right\} \\
&= \phi\Big(\frac{10}{6}\Big) - \phi\Big(\frac{-5}{6}\Big) \\
&= \phi(1.67) - \phi(-0.83) \qquad \text{explanation-3} \\
&= 0.9525 - 0.2033 \\
&= 0.7492 \\
&\approx 0.75
\end{aligned}
$$

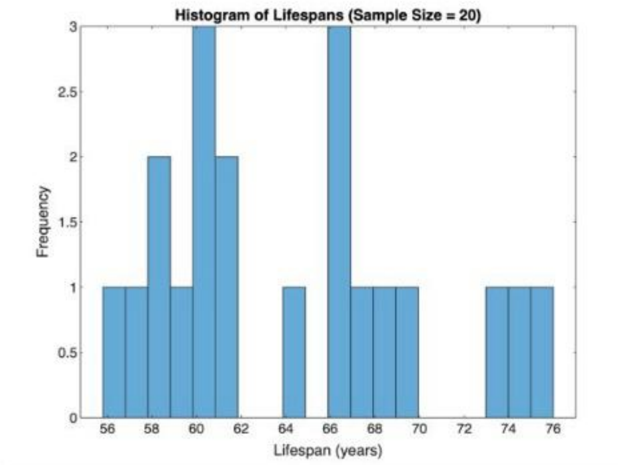
explanation-3 : either we can check these values from the textbook Table A4 Standard Normal distribution on page-417 or we can use octave code normcdf(x).
Hence, the probability that a randomly selected elephant will live more than 60 years, but less than 75 years is 0.75.

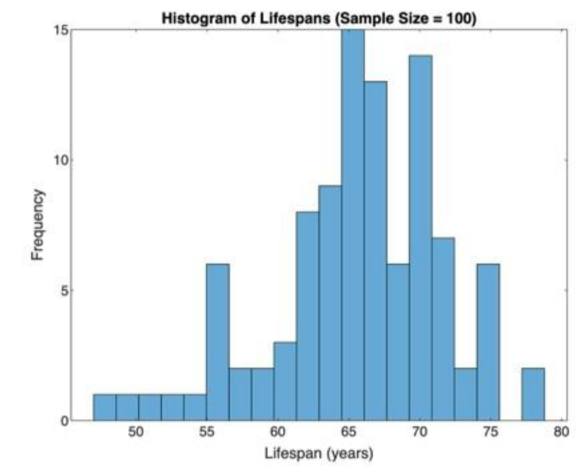
b)

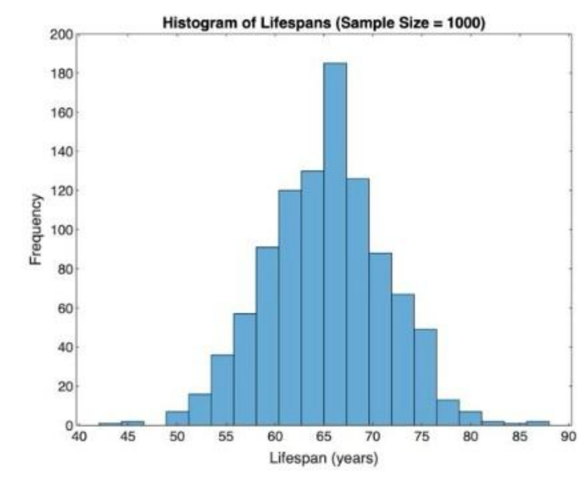
For this question I have used Matlab to create histograms. In these histograms x-axis represents the lifespan (in years) and y-axis represents the frequency.
Here is the code for histograms:

```
mu = 65;
sigma = 6;
sample_sizes = [20, 100, 1000];

for i = 1:length(sample_sizes)
    sample = normrnd(mu, sigma, [1, sample_sizes(i)]);

    figure;
    histogram(sample, 20);
    title(['Histogram of Lifespans (Sample Size = ' num2str(sample_sizes(i)) ')']);
    xlabel('Lifespan (years)');
    ylabel('Frequency');
end
```

And the histograms:


Histogram of Lifespans (Sample Size = 20)
Frequency
Lifespan (years)
3
2.5
2
1.5
1
0.5
0
56
58
60
62
64
66
68
70
72
74
76


Histogram of Lifespans (Sample Size = 100)
Frequency
Lifespan (years)
15
10
5
0
50
55
60
65
70
75
80


Histogram of Lifespans (Sample Size = 1000)
Frequency
Lifespan (years)
200
180
160
140
120
100
80
60
40
20
0
40
45
50
55
60
65
70
75
80
85
90

c)

From part-a we know that the probability that a randomly selected elephant will live more than 60 years, but less than 75 years is 0.75. So we expect that in most of the iterations at least 70% of the elephants will have a lifespan in the given range but the number of iterations that shows at least 85% of the elephants had a lifespan in the given range will be much less. In order to simulate that I used the below Matlab code:

```
mu = 65;
sigma = 6;
sample_size = 100;
iterations = 1000;
range_min = 60;
range_max = 75;
count_70_percent = 0;
count_85_percent = 0;

for iter = 1:iterations
    sample = normrnd(mu, sigma, sample_size, 1);
    count_in_range = sum(sample > range_min & sample <= range_max);
    percentage_in_range = count_in_range / sample_size * 100;
    if percentage_in_range >= 70
        count_70_percent = count_70_percent + 1;
    end
    if percentage_in_range >= 85
        count_85_percent = count_85_percent + 1;
    end
end

disp(['Number of simulations with at least 70% of elephants in the given range: ',
num2str(count_70_percent)]);

disp(['Number of simulations with at least 85% of elephants in the given range: ',
num2str(count_85_percent)]);
```

When i run this code I get these results:

```
Number of simulations with at least 70% of elephants in the given range: 899
Number of simulations with at least 85% of elephants in the given range: 8
```

So as we expected, the number of simulations with at least 70% of elephants had a lifespan in the given range is very high but the number of simulations with at least 85% of elephants had a lifespan in the given range is very low. This is because for a single randomly selected elephant the probability that its lifespan is between 60 and 75 years is 75% as calculated in part-a.