Probability and Statistics – Midterm Solutions

1.

```
clear all;
close all;
d1=[ 2.0 3.0 0.3 3.3 1.3 0.4];% 1st row
d2=[ 0.2 6.0 5.5 6.5 2.3];% 2nd row
d3=[ 1.5 4.0 5.9 1.8 4.7 0.7];% 3rd row
d4=[ 4.5 0.3 1.5 0.5 2.5 5.0];% 4th row
d5=[ 1.0 6.0 5.6 1.2 0.2];% 5th row
d=[d1 d2 d3 d4 d5];
% plot distribution
% use 7 bins for the histogram
[n,xout] = hist(d,[ 0.45 :1:6.45 ]);
bar(xout,n/sum(n)); %relative frequency is n/sum(n) frequency is n/sum(n)
set(gca,'XTick',[0.45:1:6.45]);
title('relative frequecy histogram','fontsize',18);
xlabel('years','fontsize',18);
ylabel('relative frequency','fontsize',18);
```

2.

Consider the events:

A: two nondefective components are selected,

N: a lot does not contain defective components, $P(N) = 0.6$, $P(A \mid N) = 1$,

O: a lot contains one defective component, $P(O) = 0.3$, $P(A \mid O) = \frac{\binom{19}{2}}{\binom{20}{2}} = \frac{9}{10}$,

T: a lot contains two defective components, $P(T) = 0.1$, $P(A \mid T) = \frac{\binom{18}{2}}{\binom{20}{2}} = \frac{153}{190}$.

$$P(O \mid A) = \frac{(9/10)(0.3)}{0.9505} = 0.2841;$$

3.

$$P\left(\tfrac{1}{4} < X < \tfrac{1}{2}, Y > \tfrac{1}{3}, Z < 2\right) = \tfrac{4}{9} \int_1^2 \int_{1/3}^1 \int_{1/4}^{1/2} xyz^2 \, dx \, dy \, dz = \tfrac{7}{162}.$$

4.

: For the discrete case, we can write

$$\begin{aligned}\sigma_{XY} &= \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f(x,y) \\ &= \sum_x \sum_y xy f(x,y) - \mu_X \sum_x \sum_y y f(x,y) \\ &\quad - \mu_Y \sum_x \sum_y x f(x,y) + \mu_X \mu_Y \sum_x \sum_y f(x,y).\end{aligned}$$

Since

$$\mu_X = \sum_x x f(x,y), \quad \mu_Y = \sum_y y f(x,y), \text{ and } \sum_x \sum_y f(x,y) = 1$$

for any joint discrete distribution, it follows that

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y = E(XY) - \mu_X\mu_Y.$$

For the continuous case, the proof is identical with summations replaced by integrals. ∎

5.

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

$$\begin{aligned}\sigma^2 &= E[(X-\mu)^2] \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\ &= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \underline{\int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)dx} + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx \\ &\geq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx\end{aligned}$$

now, if $|x-\mu| \geq k\sigma \therefore (x-\mu)^2 \geq k^2\sigma^2$

$$\Rightarrow \sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)dx$$

$$\Rightarrow \int_{-\infty}^{\mu-k\sigma} f(x)dx + \int_{\mu+k\sigma}^{\infty} f(x)dx \leq \tfrac{1}{k^2}$$

$$\therefore P(\mu-k\sigma < X < \mu+k\sigma) = \int_{\mu-k\sigma}^{\mu+k\sigma} f(x)dx \geq 1 - \tfrac{1}{k^2}$$

2

6.

6.1

1. The experiment consists of repeated trials.
2. Each trial results in an outcome that may be classified as a success or a failure.
3. The probability of success, denoted by p, remains constant from trial to trial.
4. The repeated trials are independent.

6.2

1. The number of outcomes occurring in one time interval or specified region of space is independent of the number that occur in any other disjoint time interval or region. In this sense we say that the Poisson process has no memory.
2. The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region and does not depend on the number of outcomes occurring outside this time interval or region.
3. The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.

7.

Using binomial distribution, there are n (n=12) repeated trials that are independent, and the probability of success in each trial remains constant (p=0.7), in which the results from each trial can classified as either a success (exercise daily) or a failure (not exercise daily). Therefore, n is 12, p is 70%(0.7). In (a), It takes r = 9 and r = 6 to calculate the number of success range from 7 to 9 by referring to Table A.1. In(b), the probability when the number of success is less than 8 is equivalent to 1 – P(x<=7) and, therefore, r = 7 is used to find the corresponding probability.

From Table A.1 with $n = 12$ and $p = 0.7$, we have

7.1

$P(7 \le X \le 9) = P(X \le 9) - P(X \le 6) = 0.7472 - 0.1178 = 0.6294.$

7.2

$P(X \ge 8) = 1 - P(X \le 7) = 1 - 0.2763 = 0.7237.$

8.

Using Poisson distribution.

The number of outcomes (the average number of mice per acre is estimated to be 12) occurring in a specified region of space (5-acre wheat field) is independent of the number of outcomes that occurred in other disjoint region. A single outcome occurs in a small region which is proportional to size of it. And the probability which is more than one outcome occurred in a small region is negligible. The average number of field mice per acre is estimated to be 12. Therefore, μ = 12 in this case, and the probability in question is when fewer than 7 mice are found, which includes all conditions of finding 0~6 mice. Therefore r = 6 is used to find the probability value from Table A.2.

Using the Poisson distribution with $\mu = 12$, we find from Table A.2 that $P(X < 7) = P(X \leq 6) = 0.0458$.

9.

9.1

Let X represent the number of defects found in 500 samples. The problem can calculated using a binomial distribution *b*(X;500,0.01), whose probability values are unavailable from the Binomial Probability Sums table. However, because n is large and p is close to zero, the binomial distribution *b*(X;500,0.01) can be approximated by a Poisson distribution *P*(X;500*0.01). From the Poisson probability sums table, we can check the probability values for μ=5.

Therefore,

P(X=15) = P(X≤15) - P(X≤14) = 0.9999 - 0.9998 = 0.0001

9.2

From 9.1, the probability of 0.0001 is very rare. Therefore, the presumption that the defective rate is 0.01 is questionable.