MAT2377 (Fall 2012) - Assignment 2

Total number of points: 28

Important:

- The assignments have to be handed in the lobby of the Department of Mathematics and Statistics (in the box marked MAT 2377) on the due date, no later than 14:30. Late assignments cannot be accepted. Since the assignments will not be distributed after marking, students are advised to make a copy of each assignment before submitting it.
- Electronic submission of assignments is not accepted!
- According to the university policy regarding the confidentiality of students' work, graded assignments cannot be brought to class for students to pick them up.
- Students should keep a copy of each assignment, check the solution posted on the Virtual Campus, and verify their grades (using also the Virtual Campus). If they have any doubts regarding the way they have been graded on any particular assignment, they are welcome to see their assignment in the instructor's office, during the office hours.
- Write clearly.

Q1. The sample space of a random experiment is $\{a, b, c, d, e, f\}$ and each outcome is equally likely. A random variable is defined as follows

outcome	a	b	c	d	e	f
x	0	0	1.5	1.5	2	3

Determine the probability mass function of X. Determine the following probabilities:

(a) $P(X = 1.5)$ (b) $P(0.5 < X < 2.7)$ (c) $P(X > 3)$

(d) $P(0 \le X < 2)$ (e) $P(X = 0 \text{ or } X = 2)$

Solution to Q1:

Probability mass function is

$$P(X=0) = P(\{a,b\}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6};\ P(X=1.5) = \frac{2}{6},\ P(X=2) = \frac{1}{6},\ P(X=3) = \frac{1}{6}.$$

(a) $P(X = 1.5) = \dfrac{2}{6}$

(b) $P(0.5 < X < 2.7) = P(X = 1.5) + P(X = 2) = \dfrac{3}{6}$

(c) $P(X > 3) = 0$

(d) $P(0 \le X < 2) = P(X = 0) + P(X = 1.5) = \dfrac{4}{6}$

(e) $P(X = 0 \text{ or } X = 2) = \dfrac{3}{6}$

Marking scheme for Q1:

Completely correct p.m.f. - 1 point, correct answer for each part - 1 point. Total - 6 points.

Q2. Determine the mean and the variance in Question Q1

Solution to Q2:

$\mathrm{E}(X) = 1.33$, $\mathrm{Var}(X) \approx 1.15$

Marking scheme for Q2:

1 point for the mean and the variance. Total - 2 points.

Q3. (3.54) We say that X has *uniform distribution* on a set of values $\{x_1, \ldots, x_k\}$ if

$$P(X = x_i) = \frac{1}{k}, \qquad i = 1, \ldots, k.$$

The thickness measurements of a coating process are *uniformly distributed* with values 0.15, 0.16, 0.17, 0.18, 0.19. Determine the mean and variance.

Solution to Q3:

We have $P(X = 0.15) = \cdots = P(X = 0.19) = 1/5$. Mean: 0.17; Variance:

$$\frac{1}{5}\left(0.01^2 + 0.02^2 + 0.02^2 + 0.01^2\right)$$

Marking scheme for Q3:

1 point for the mean and the variance. Total - 2 points.

Q4. Samples of rejuvenated mitochondria are mutated in 1% cases. Suppose 15 samples are studied and they can be considered to be independent for mutation. Determine the following probabilities:

(a) No samples are mutated

(b) At most one sample is mutated

(c) More than half the samples are mutated

Solution to Q4:

Let X be the number of no-mutated samples; then X has binomial distribution with $n = 15$ and $p = 0.99$ (success=no mutation; note that if choose success=mutation, then $p = 0.01$, then we cannot use the Table). To compute (use Tables)

(a) 0 mutated samples = 15 no-mutated samples, thus to compute $P(X = 15) = P(X \leq 15) - P(X \leq 14) = 1 - 0.13999 \approx 0.86$

(b) at most one sample mutated = at least 14 are not mutated; $P(X \geq 14) = 1 - P(X < 14) = 1 - P(X \leq 13) = 1 - 0.0096 = 0.9904$

(c) More than half the samples are mutated = less than or half samples are no-mutated; $P(X \leq 7.5) = P(X \leq 7) = 0$

Marking scheme for Q4:

Correct answer for each part - 1 point. Total - 3 points.

Q5. **A statistical process control**. Samples of 20 parts from a metal punching process are selected every hour. Typically, 1% of the parts require rework. Let X denote the number of parts in the sample that require rework. A process problem is suspected if X exceeds its mean by more than three standard deviations.

(a) What is the probability that there is a process problem?

(b) If rework percentage increases to 4%, what is the probability that X exceeds 1?

(c) If rework percentage increases to 4%, what is the probability that X exceeds 1 in at least one of the next five hours of samples?

Solution to Q5:

(a) We have $X \sim \mathcal{B}(n,p)$, $n = 20$, $p = 0.01$ (success = a part requires rework). $\mathrm{E}(X) = np = 0.2$, $\mathrm{Var}(X) = np(1-p) = 0.2 \times 0.99 = 0.198$, $\mathrm{SD}(X) \approx 0.44$. To compute

$$\begin{aligned} P(X > 0.2 + 3 \times 0.44) &= P(X > 1.535) = P(X \geq 2) = 1 - P(X < 2) = 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \binom{20}{0} 0.01^0 0.99^{20} - \binom{20}{1} 0.01^1 0.99^{19} \approx 0.017 \end{aligned}$$

Alternatively, you may use Table, however, you must work then with $Y \sim \mathcal{B}(n, 0.99)$, i.e. Y is the # of samples which do not require rework.

(b) We have $X \sim \mathcal{B}(n,p)$, $n = 20$, $p = 0.04$. To compute

$$\begin{aligned} P(X > 1) &= P(X \geq 2) = 1 - P(X < 2) = 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \binom{20}{0} 0.04^0 0.96^{20} - \binom{20}{1} 0.04^1 0.96^{19} \approx 0.19. \end{aligned}$$

Note: here you cannot use Table at all.

(c) Now, we have 5 hourly samples, and each of them consists of 20 items. let Y be the number of hourly samples, where number of items which require rework is bigger than 1 (Success = hourly sample has more than 1 defective item, i.e. $X > 1$). We have $Y \sim \mathcal{B}(5, p_0)$, where $p_0 = P(X > 1) = 0.19$ is the probability of success. Success = X exceeds 1. Thus, to compute

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{5}{0} 0.19^0 \times 0.81^5 \approx 0.651$$

Marking scheme for Q5:

For part a) - 3 points: 1 point for correct $\mathrm{E}(X)$ and $\mathrm{SD}(X)$, 1 point for correct formula to compute, i.e. $P(X > 0.2 + 3 \times 0.44)$, 1 point for the correct answer. 1 point for part b). 2 points for part c): 1 point for the correct identification of Y , 1 point for the correct answer. Total - 6 points.

Q6. In a clinical study, volunteers are tested for a gene that has been found to increase the risk for a disease. The probability that the person carries a gene is 0.1

(a) What is the probability that 4 or more people will have to be tested in order to detect one person with the gene?

(b) How many people are expected to be tested in order to detect one person with the gene?

(c) How many people are expected to be tested before 2 with a gene are detected?

Solution to Q6:

(a) If X is the number of steps before the 1st success, then X has geometric distribution with $p = 0.1$ (success = gene detection). To compute

$$P(X \geq 4) = \sum_{k=4}^{\infty} (1-p)^{k-1} p = p \times \frac{(1-p)^3}{1 - (1-p)} = (1-p)^3 = 0.729.$$

(b) $\mathrm{E}(X) = 1/p = 10$.

(c) $\mathrm{E}(X) = 20$.

Marking scheme for Q6:

Correct answer for each part - 1 point. Total - 4 points.

Q7. The number of failures of a testing instrument from contamination particles on the product is a Poisson random variable with a mean of 0.02 failure per hour.

(a) What is the probability that the instrument does not fail in an 8-hour shift?

(b) What is the probability of at least one failure in a 24-hour day?

Solution to Q7:

(a) The failure rate per 8 hours is $8 \times 0.02 = 0.16$. If X is Poisson random variable with $\lambda = 0.16$, we have to compute $P(X = 0) = \exp(-0.16)$.

(b) Now, X is Poisson with $\lambda = 24 \times 0.02 = 0.48$. To compute $P(X \geq 1) = 1 - P(X = 0) = 1 - \exp(-0.48)$.

Marking scheme for Q7:

1 point for each part. Total - 2 points.

Q8. `R-simulation:` Generate a sample from a binomial distribution or Poisson distribution (choose one of those, choose parameters). Use `R` to compute sample mean and sample variance. Compare to population mean and population variance.

Solution to Q8:

I simulated 1000 numbers from a binomial distribution with parameters $n = 10$ and $p = 0.3$. True mean and variance are, respectively, 3 and 2.1. The estimated values agree with the theoretical ones.

```
X=rbinom(1000,10,0.3);
mean(X); var(X);
[1] 3.036 [1] 2.084789
```

Marking scheme for Q8:

3 points: 1 point correct simulation command; 1 point for correct estimated values (0.5 points each); 1 point for correct theoretical values (0.5 points each).