ENGI 4421 Probability and Statistics
Faculty of Engineering and Applied Science

Problem Set 7

Probability Distributions

---

1. A bin of 1,000 silicon chips is known to contain 750 good chips and 250 defective chips. A random sample of four chips is drawn from the bin. Let X represent the number of defective chips found in the sample.

(a) Can the binomial distribution be used to model the distribution of X? Exactly or approximately? *Justify your answer*.
(b) Find the probability, (correct to two significant figures), that not all of the chips in the sample are good.
(c) Find the **exact** probability, (expressed as a fraction), that not all of the chips in the sample are good.
(d) What are the population mean E[X] and approximate population variance V[X] of X ?

---

2. A box contains 8 good items and 2 defective items from a factory's production line. A manager selects four items at random from the box, without replacement. Let X = (the number of defective items in the random sample).

(a) Show that the probability distribution of X is **not** binomial.
(b) Find the exact probability mass function $p(x)$.
(c) Find E[X].
(d) Find V[X].
(e) Find P[$X = 3$] [*Note*: this part can be attempted without part (b)].
(f) Find P[$X < 2$] .
(g) If the random sample were drawn **with** replacement, would the probability mass function be binomial?
(h) If the random sample were drawn **with** replacement, what would the value of P[$X = 3$] be?

---

3. Consider a Poisson process with constant rate λ per unit time. This can be used to model the arrival of telephone calls, electronic messages, accidents in industry, customers in queuing models, to name but a few processes. The assumption of a constant rate may apply only for certain periods (for example those during peak demand). Under this condition, the time between consecutive arrivals follows an exponential distribution.

In a telephone system the assumptions just noted have been found to apply. There are on average 12 calls from a certain origin every five minutes.

(a) What is the value of λ (in units of calls per minute)?
(b) Find the probability that fewer than two calls arrive during the next minute.
(c) From this Poisson distribution, find the parameter of the exponential distribution that models the time in minutes between phone calls from the specified origin and write down the p.d.f.
(d) Find the probability that two successive calls arrive less than thirty seconds apart.
(e) Find the probability that two successive calls arrive more than a minute apart.
(f) Find the probability that the two successive calls are separated by an interval that is within two standard deviations of the mean. Sketch the pdf showing this situation.

---

4. [Navidi, Exercises 4.3, Question 11]

A microbiologist wants to estimate the concentration of a certain type of bacterium in a waste water sample. She puts a 0.5 mL sample of the waste water on a microscope slide and counts 39 bacteria in that sample. Estimate the concentration of bacteria per mL in this waste water and find the uncertainty in this estimate.

---

5. The lifetime in months X of an electronic component can be modelled by the exponential distribution

$$f_X(x) = 0.1\, e^{-0.1x}, \qquad (x \ge 0)$$

(a) Find the cumulative distribution function $F_X(x)$.
(b) What is the probability that the life of a component exceeds 6 months?
(c) What is the mean of the distribution?
(d) What is the standard deviation of the distribution?
(e) Find the half-life x_h, defined such that $\mathrm{P}\left[X \le x_h\right] = \frac{1}{2}$.
(f) Plot or sketch the probability distribution and the c.d.f.

---

6. [Navidi, Exercises 4.11, Question 9]

A surveyor estimates the height of a cliff by taking n independent measurements and averaging them. Each measurement is unbiased and has standard deviation $\sigma = 1$ m. How many measurements should be taken for the probability P[the average is within 0.25 m of the true value] = .95?

---

7. The concentration of pollutant in a sample taken from a lake, measured in mg/10,000*l*, is found to follow a normal distribution with a mean of 19.6 and a standard deviation of 4.6. Three events are considered to be of importance:
A = pollution in a sample being less than 15 mg/10,000*l*.
B = pollution in a sample being greater than 15 but less than 25 mg/10,000*l*.
C = pollution in a sample being greater than 25 mg/10,000*l*.

(a) What is the probability of individual events *A*, *B*, and *C* occurring?
(b) What is the probability that in 6 random samples, event *A* occurs once, *B* occurs 5 times and *C* does not occur?

---

8. Two machines are available for the manufacture of resistors rated at 100 ohm, for use in a particular type of electric circuit. A resistor is rejected if its actual resistance *R* is not in the range $99.90 < R < 100.10$ (ohm).
Machine ‘A’ produces resistors whose resistance is normally distributed with mean 100.00 ohm and standard deviation 0.05 ohm.
Machine ‘B’ produces resistors whose resistance is normally distributed with mean 100.05 ohm and standard deviation 0.02 ohm.

(a) Which machine is more likely to produce an acceptable resistor?
Show your working.
(b) The difference *C* of two normally distributed random quantities *A*, *B* is also normally distributed, with its mean equal to the difference of the two means:
$$\mu_C \; = \; \mu_A - \mu_B \,.$$
If the random quantities are uncorrelated, (which is true if they are independent), then the variance of the difference is the sum of the variances:
$$\text{V}[C] \; = \; \text{V}[A] + \text{V}[B] \,.$$
How likely is it that a randomly chosen resistor produced by machine ‘A’ will have a lower resistance than that of a randomly chosen resistor produced by machine ‘B’?

---

9. The vehicle speed distribution on a **two-way** road is $N(80, 20^2)$, where speed is measured in km/h.

(a) Determine the probability that the relative speed of two vehicles moving in opposite directions is between 150 and 200 km/hour.
(b) Upper and lower speed limits of 100 km/h and 50 km/h are imposed on the road. Sketch the new probability distribution of vehicle speeds assuming that the limits are obeyed and that the affected drivers change their vehicle speed by the minimum legal amount.
(c) Is the (idealized) distribution in part (b) continuous, discrete or something else? Discuss briefly.

---

10. Prove the formula at the foot of page 9.22 of the lecture notes:
The estimate of concentration c when n particles are counted in a volume v is

$$c = \frac{n}{v} \pm \frac{\sqrt{n}}{v}$$

---

11. To a good approximation, the strength at failure of a valve, (expressed in terms of the applied pressure that causes failure), is known to be normally distributed with a mean of 980 kN m^{-2} and a standard deviation of 60 kN m^{-2} . The maximum pressure to which the valve is subjected in any given week is known to be normally distributed with a mean of 725 kN m^{-2} and a standard deviation of 80 kN m^{-2}.

(a) Find the probability that a randomly chosen valve fails during the first week.
(b) Find the odds that the valve does not fail during the first 52 weeks, (that is, during the first year).
[Express the odds in the approximate form r **:** 1, with r rounded off to the nearest integer.]
State any assumptions that you have made.

---

12. ***Bonus Question*:**

A continuous random quantity that follows an exponential distribution is said to "have no memory", in that the probability of an event occurring during a specified future time interval is independent of the time that has elapsed so far.
Put another way, the probability that the event occurs during the next b seconds from now is the same as the probability that the event occurs during the first b seconds.
Mathematically, for all positive numbers a, b,

$$\mathrm{P}[T < a + b \mid T > a] \;=\; \mathrm{P}[T < b]$$

where the random quantity T has the cumulative distribution function

$$F(t) \;=\; \mathrm{P}[T \le t] \;=\; 1 - e^{-t/\mu} \qquad (t \ge 0)$$

and $\mu = \mathrm{E}[T]$ is the mean value of T.
Prove that this statement is true.

---

13. ***Bonus Question*:**

If two independent continuous random quantities U and V have probability density functions $g(u)$ and $h(v)$ respectively, then their sum, $X = U + V$, has a p.d.f. $f(x)$, given by the convolution

$$f(x) \;=\; (g * h)(x) \;=\; \int_{-\infty}^{\infty} g(t)\,h(x-t)\,dt$$

Show that the sum of two independent identically distributed random quantities with the p.d.f. of an exponential distribution $g(x; \lambda) \;=\; \lambda\, e^{-\lambda x}$, $(x > 0)$ is a random quantity that has the p.d.f. $f(x; 2, \lambda) \;=\; \lambda^2\, x\, e^{-\lambda x}$, $(x > 0)$, (which is the Gamma distribution $\Gamma(2, \lambda)$.)

☜ Back to the index of questions On to the solutions to this problem set ☞