Problem 1: Determine whether each of the following random variables as a binomial distribution. If so, identify the values of the shape parameters n and p.

(a) $X = $ the number of 4's in ten rolls of a fair die.

X is a binomial random variable. This is because we are concerned with a binary outcome (roll a 4, or not?), the outcomes of the rolls are all independent, and the probability of rolling a 4 is fixed. We have $X \sim \mathcal{B}in(10, 1/6)$.

(b) $X = $ the number of multiple-choice questions a student gets right on a 40-question test, when each question has four choices and the student is completely guessing.

X is a binomial random variable. This is because we are concerned with a binary outcome (get the question right, or not?), the outcome of each question is independent of all the others, and the probability of getting a question right is fixed. We have $X \sim \mathcal{B}in(40, 1/4)$.

(c) Same as part (b), but now half the questions have two possible answers and the other half four.

X is *not* a binomial random variable. This is because the probability of a successful outcome (getting a question right) changes from $1/4$ to $1/2$.

Problem 2: When circuit boards used in the manufacture of DVD players are tested, the long-run percentage of defectives is 5%. Let $X = $ the number of defective boards in a random sample of size $n = 25$, so $X \sim \mathcal{B}in(25, 0.05)$.

(a) Determine $P(X \leq 2)$.

We have
$$P(X \leq 2) = \sum_{x=0}^{2} \binom{25}{x}(0.05)^x(1 - 0.05)^{25-x} \approx 0.873.$$

(b) Determine $P(X \geq 5)$.

We have
$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{x=0}^{4} \binom{25}{x}(0.05)^x(1 - 0.05)^{25-x} \approx 0.07.$$

(c) Determine $P(1 \leq X \leq 4)$.

We have
$$P(1 \leq X \leq 4) = P(X \leq 4) - P(X = 0) = \sum_{x=0}^{4} \binom{25}{x}(0.05)^x(1 - 0.05)^{25-x} - (1 - 0.05)^{25} \approx 0.715.$$

(d) What is the probability that none of the 25 boards is defective?

We have
$$P(X = 0) = (1 - 0.05)^{25} \approx 0.277.$$

(e) Calculate the expected value and standard deviation of X.

We have

$$E(X) = 25(0.05) = 1.25 \quad \text{and} \quad \sigma_X = \sqrt{25(0.05)(1-0.05)} \approx 1.090.$$

Problem 3:   Suppose that 30% of the applicants for a certain industrial job possess advanced training in computer programming. Applicants are interviewed sequentially and are selected at random from the pool.

(a) Find the probability that the first applicant with advanced training in programming is found on the fifth interview.

Let X be the number of applicants that are interviewed before the first one with advanced training is found. Then $X \sim \mathcal{G}eo(0.30)$, and we are asked to compute:

$$P(X = 5) = (0.30)(1-0.30)^{5-1} \approx 0.072.$$

(b) Find the probability that the first applicant with advanced training in programming is found in one of the first five interviews.

We have

$$P(X \leq 5) = \sum_{x=1}^{5}(0.30)^x(1-0.30)^{x-1} \approx 0.832.$$

(c) What is the expected number of applicants who need to be interviewed in order to find the first one with advanced training?

We have

$$E(X) = \frac{1}{0.30} = 10/3 = 3.\bar{3}.$$

Problem 4:   Given that we have already tossed a fair coin ten times and obtained zero heads, what is the probability that we must toss it at least two more times to obtain the first head?

The way that the problem is phrased suggests that our task is to compute a conditional probability. Let's solve it this way, and then we'll come back to a simpler way. So, we have $X \sim \mathcal{G}eo(0.5)$, and we are asked to compute

$$P(X \geq 12 | X \geq 11) = \frac{P(X \geq 12, X \geq 11)}{P(X \geq 11)} = \frac{P(X \geq 12)}{P(X \geq 11)}.$$

But

$$P(X \geq 12) = 1 - P(X \leq 11) \approx 0.0005$$

and

$$P(X \geq 11) = 1 - P(X \leq 10) \approx 0.001,$$

so

$$P(X \geq 13 | X \geq 11) \approx \frac{0.0005}{0.001} = 0.5.$$

On the other hand, since the tosses are all independent of each other, the fact that we tossed the coin ten times and obtained zero heads has no bearing on the subsequent tosses. So, the probability of needing at least two more tosses after the 10th to obtain a head is the same as $P(X \geq 2)$. But

$$P(X \geq 2) = 1 - P(X = 1) = 1/2,$$

which confirms our computation above.

Problem 5:   Five individuals from an animal population thought to be near extinction in a region have been caught, tagged, and released to mix into the population. After they have had an opportunity to mix, a random sample of ten of these animals is selected. Let $X =$ the number of tagged animals in the second sample. If there are actually 25 animals of this type in the region, what is the probability that

(a) $X = 2$?

We have $M = 25$, $n = 5$, and $N = 10$, so $X \sim \mathcal{HGeo}(25, 5, 10)$ and we are asked to compute:

$$P(X = 2) = \frac{\binom{5}{2}\binom{25-5}{10-2}}{\binom{25}{10}} \approx 0.385.$$

(b) $X \leq 2$?

We have

$$P(X \leq 2) = \sum_{x=0}^{2} \frac{\binom{5}{x}\binom{25-5}{10-x}}{\binom{25}{10}} \approx 0.699.$$

Problem 6:   A geologist has collection 10 specimens of basaltic rock and 10 specimens of granite. The geologist instructs a laboratory assistant to randomly select 15 of the specimens for analysis.

(a) What is the PMF of the number of granite specimens selected for analysis?

Let $X =$ number of granite specimens selected. Then $X \sim \mathcal{HGeo}(20, 10, 15)$ and so the PMF is

$$p(x) = \frac{\binom{10}{x}\binom{20-10}{15-x}}{\binom{20}{15}}.$$

(b) What is the probability that all specimens of one of the two types of rock are selected for analysis?

The probability that all granite specimens are selected is

$$P(X = 10) = \frac{\binom{10}{10}\binom{20-10}{15-10}}{\binom{20}{15}} \approx 0.016.$$

Since the event that all granite specimens are selected and the event that all basaltic specimens are selected are disjoint, and since there are ten of each type of specimen, the desired probability is $0.016 \times 2 = 0.032$.

(c) What is the probability that the number of granite specimens selected for analysis is within one standard deviation of its mean value?

We have
$$\mu = \frac{15 \cdot 10}{20} = 7.5$$
and
$$\sigma = \sqrt{\left(\frac{20-15}{20-1}\right) \cdot 15 \cdot \frac{10}{20}\left(1 - \frac{10}{20}\right)} \approx 0.993.$$

Then, X will be within one standard deviation of its mean if it falls in the interval
$$[\mu - \sigma, \mu + \sigma] \approx [6.507, 8.493].$$

But X can only take integer values, so this translates to X being in the interval $[7, 8]$. Therefore, the probability that we must compute is
$$P(7 \leq X \leq 8) = P(X \leq 8) - P(X \leq 6) \approx 0.697.$$

Problem 7: A store owner believes that customers arrive at his store at a rate of 4.5 customers per hour on average.

(a) Describe a distribution that models the number X of customers that arrive at the store during a two-hour time interval.

We have $X \sim \mathcal{P}ois(9)$, since we are told $\lambda = 4.5$ and that the time interval is of length $t = 2$.

(b) Compute the probabilities that one customer arrives in the time interval, and that ten customers arrive.

We have
$$P(X = 1) = \frac{9^1}{1!}e^{-9} \approx 0.001 \quad \text{and} \quad P(X = 10) = \frac{9^{10}}{10!}e^{-9} \approx 0.119.$$

(c) Compute the probabilities that at most five customers arrive in the time interval, and that more than ten arrive.

We have
$$P(X \leq 5) = \sum_{x=0}^{5} \frac{9^x}{x!}e^{-9} \approx 0.117$$
and
$$P(X \geq 11) = 1 - P(X \leq 10) = 1 - \sum_{x=0}^{10} \frac{9^x}{x!}e^{-9} \approx 0.294.$$

Problem 8:   The number of customers X that visit a bank on a day is modeled by a Poisson distribution. It is known that the probability of no customers at all is $0.00001 = 10^{-5}$. What is the expected number of customers?

We are told $X \sim \mathcal{P}ois(\mu)$ where μ is the expected number of customers. We are also told that

$$10^{-5} = P(X = 0) = \frac{\mu^0}{0!}e^{-\mu} = e^{-\mu},$$

so that

$$\mu = -\ln 10^{-5} = 5\ln 10 \approx 11.513.$$

Problem 9:   A certain brand of copper wire has a flaw once about every 40 centimeters. Let X denote the number of flaws in a 1-meter length of wire.

(a) Describe an appropriate distribution for X.

   Assuming that the flaws occur randomly and independently of each other along the length of the wire, we may use $X \sim \mathcal{P}ois(\mu)$. To compute the parameter μ, we need to use the formula $\mu = \lambda \times \ell$, where λ is the mean rate at which the flaws occur (measured in flaws per meter) and ℓ is the length of the wire (measured in meters). But

$$\lambda = \frac{1 \text{ flaw}}{40 \text{ cm}} = \frac{2.5 \text{ flaws}}{1 \text{ m}}.$$

   Since $\ell = 1$ m, we therefore have $\mu = 2.5$.

(b) What is the probability of two flaws in the 1-meter length of wire? The probability of more than five?

   We have

$$P(X = 2) = \frac{(2.5)^2}{2!}e^{-2.5} \approx 0.257$$

   and

$$P(X \geq 6) = 1 - P(X \leq 5) = 1 - \sum_{x=0}^{5} \frac{(2.5)^x}{x!}e^{-2.5} \approx 0.042.$$

Problem 10:   The achievement scores X for a college entrance exam are normally distributed with mean 75 and standard deviation 10. What fraction of the scores lies between 80 and 90?

The desired fraction is a difference in normal CDFs:

$$P(80 \leq X \leq 90) = P(X \leq 90) - P(X \leq 80) \approx 0.242.$$

Problem 11:   Compute the following critical values.

(a) $z_{0.0055} \approx 2.542$
(b) $z_{0.09} \approx 1.341$
(c) $z_{0.663} \approx -0.421$

Problem 12: For each real number $\alpha \geq 0.5$, prove that $-z_\alpha = \Phi^{-1}(\alpha)$.

Since $\alpha \geq 0.5$, we have $z_\alpha \geq 0$. By symmetry, we have

$$\int_{-z_\alpha}^0 f(z)\,\mathrm{d}z = \int_0^{z_\alpha} f(z)\,\mathrm{d}z \quad \text{and} \quad \int_{-\infty}^0 f(z)\,\mathrm{d}z = \int_0^\infty f(z)\,\mathrm{d}z,$$

where $f(z)$ is the standard normal density. Hence

$$\Phi(-z_\alpha) = \int_{-\infty}^0 f(z)\,\mathrm{d}z - \int_{-z_\alpha}^0 f(z)\,\mathrm{d}z = \int_0^\infty f(z)\,\mathrm{d}z - \int_0^{z_\alpha} f(z)\,\mathrm{d}z = \int_{z_\alpha}^\infty f(z)\,\mathrm{d}z = \alpha.$$

Thus, $-z_\alpha = \Phi^{-1}(\alpha)$.

Problem 13: There are two machines available for cutting corks intended for use in wine bottles. The first produces corks with diameters that are normally distributed with mean 3 cm and standard deviation 0.1 cm. The second machine produces corks with diameters that have a normal distribution with mean 3.04 cm and standard deviation 0.02 cm. Acceptable corks have diameters between 2.9 and 3.1 cm. Which machine is more likely to produce an acceptable cork?

The standard deviation of the second machine is an order of magnitude smaller than the first, so clearly the second machine is more likely to produce acceptable corks. But let's check this with some computations: Let X_1 and X_2 be the diameters of the corks produced by the two machines. We are told

$$X_1 \sim \mathcal{N}(3, 0.1^2) \quad \text{and} \quad X_2 \sim \mathcal{N}(3.04, 0.02^2).$$

Then

$$P(2.9 \leq X_1 \leq 3.1) = P(X_1 \leq 3.1) - P(X_1 \leq 2.9) \approx 0.683$$

and

$$P(2.9 \leq X_2 \leq 3.1) = P(X_2 \leq 3.1) - P(X_2 \leq 2.9) \approx 0.999,$$

which confirms our answer above.

Problem 14: Historical evidence indicates that times X between fatal accidents on scheduled American domestic passenger flights have an approximately exponential distribution. Assume that the mean time between accidents is 44 days.

(a) If one of the accidents occurred on July 1 of a randomly selected year in the study period, what is the probability that another accident occurred that same month?

We are told $X \sim \mathcal{E}xp(\lambda)$ with $1/\lambda = 44$. We then compute

$$P(X \leq 31) = \frac{1}{44} \int_0^{31} e^{-x/44}\,\mathrm{d}x \approx 0.506.$$

(b) What is the variance of the times between accidents?

We compute:

$$V(X) = \frac{1}{(1/44)^2} = 1{,}936.$$

Problem 15: Suppose that when a type of transistor is subjected to an accelerated life test, the lifetime T (in weeks) has a gamma distribution with mean 24 weeks and standard deviation 12 weeks.

(a) What is the probability that a transistor will last between 12 and 24 weeks?

We are told $T \sim \Gamma(\alpha, \beta)$, along with
$$\frac{\alpha}{\beta} = 24 \quad \text{and} \quad \frac{\alpha}{\beta^2} = 12^2.$$
Solving this system of equations gives $\alpha = 4$ and $\beta = 1/6$. We then compute:
$$P(12 \leq T \leq 24) = P(T \leq 24) - P(T \leq 12) \approx 0.424.$$

(b) What is the probability that a transistor will last at most 24 weeks? Is the median of the lifetime distribution less than 24? Why or why not?

We compute:
$$P(T \leq 24) \approx 0.567.$$
The median is the 0.5-quantile of the distribution, which we compute as
$$F^{-1}(0.5) \approx 22.032,$$
where F is the distribution function of T. Thus, the median lifetime is less than 24 weeks.

(c) What is the 99th-percentile of the lifetime distribution?

The 99th percentile is another name for the 0.99-quantile. We compute:
$$F^{-1}(0.99) \approx 60.271.$$

(d) Suppose the test will actually be terminated after t weeks. What value of t is such that only 0.5% of all transistors would still be operating at termination?

The desired value of t is the one for which the area to the *right* under the density curve is $0.5\% = 0.005$. Thus, we want to solve the equation
$$F(t) = 1 - 0.005$$
for t. But:
$$t = F^{-1}(1 - 0.005) \approx 65.865.$$

Problem 16: A gasoline wholesale distributor has bulk storage tanks that hold fixed supplies and are filled every Monday. Of interest to the wholesaler is the proportion X of this supply that is sold during the week. Over many weeks of observation, the distributor found that this proportion could be modeled by a beta distribution with $\alpha = 4$ and $\beta = 2$. Find the probability that the wholesaler will sell at least 90% of her stock in a given week.

We are told $X \sim \mathcal{B}eta(4, 2)$. Then, we compute:
$$P(X \geq 0.9) = 1 - P(X \leq 0.9) \approx 0.081.$$