1. **Continuous Density Function**

   The length of time (in hours) taken by a student to complete a one hour exam is a continuous random variable X with probability density function f_X defined by $f_X(x) = cx^2 + x, 0 < x \leq 1$, for some constant c, and zero otherwise.

   (a) Find the value of c.

   the two conditions for a probability density function are, that the function has only positive values (which is true for all positive c in this example) and the integral over R is 1:

$$1 \stackrel{!}{=} \int_{-\infty}^{\infty} f(x)dx = \int_0^1 cx^2 + xdx = \frac{c}{3}x^3 + \frac{1}{2}x^2\Big|_0^1 = \frac{c}{3} + \frac{1}{2}$$
$$\Rightarrow \quad c = 1.5.$$

   For $c = 1.5$ *the above function is a probability mass function.*

   (b) By integration, find the distribution function of X, F_X.

$$F_X(t) = \int_{-\infty}^{t} f(x)dx = \begin{cases} 0 & t < 0 \\ 0.5t^3 + 0.5t^2 & 0 \leq t \leq 1 \\ 1 & t > 1 \end{cases}$$

   (c) Find the probability that a student completes the exam in less than half an hour.

   X is the time the student needs for the exam, $X \sim F_X$.

$$P(X < 0.5) = F_X(0.5) = 0.5 \cdot 0.5^3 + 0.5 \cdot 0.5^2 = 0.1875.$$

   (d) Given that a student takes longer than fifteen minutes to complete the exam, find the probability that he/she requires at least half an hour.

$$P(X > 0.5 \mid X > 0.25) = \frac{P(X > 0.25 \cap X > 0.5)}{1 - P(X < 0.25)} = \frac{P(X > 0.5)}{1 - P(X < 0.25)} = \frac{1 - F_X(0.5)}{1 - F_X(0.25)}$$
$$= \frac{1 - 0.1875}{1 - 0.039} = 0.8455.$$

   (e) In a class of sixty students, find the probability that at most three students complete the exam in fewer than fifty minutes.

   The probability for each student to finish in fewer than fifty minutes is $p = P(X \leq 5/6) = 0.64$

   Let Y be the number of students finishing in fewer than fifty minutes. Y then has a binomial distribution with $n = 60$ and $p = 0.64$.

$$\begin{aligned} P(Y \leq 3) &= B_{60,0.64}(3) \stackrel{\text{central limit theorem}}{\approx} N_{60\cdot 0.64, 60\cdot 0.64\cdot(1-0.64)}(3) = \\ &= N_{38.4,13.824}(3) = N_{0,1}\left(\frac{3 - 38.4}{\sqrt{13.824}}\right) = N_{0,1}(-9.52) = \\ &= 1 - N_{0,1}(9.5) = 0 \end{aligned}$$

2. **Applications to the Central Limit Theorem**

   (a) If there is, on average, a crack in the road every 5 m and number of cracks on the road can be modelled as a Poisson Process (with appropriate parameter), then what is the chance that you will find 5 cracks in a 60 m section of road? more than 750 cracks in 2.5 km (2500 m) of road?

   Define X_{60} *= # of cracks in 60 m of road.*
   X_{60} *then has a Poisson distribution with* $\lambda = 60/5 = 12$ *cracks per 60 m.*
   the probability to find (at least) 5 cracks in 60m is:

$$P(X_{60} \geq 5) = 1 - P(X_{60} \leq 4) = 1 - Po_{12}(4) = 1 - 0.0076 = 0.9924.$$

   Similarly, let X_{2500} *= # of cracks in 2500 m of road.*
   X_{2500} *then has a Poisson distribution with* $\lambda = 2500/5 = 500$ *cracks per 2500 m.*
   the probability to find more than 525 cracks in 2500 m of road is:

$$\begin{aligned} P(X_{2500} > 525) &= 1 - P(X_{2500} \leq 525) = 1 - Po_{500}(525) \approx \\ \approx &\quad 1 - N_{500,500}(525) = 1 - N_{0,1}\left(\frac{525-500}{\sqrt{500}}\right) = 1 - N_{0,1}(1.11) = 0.1335. \end{aligned}$$

   (b) A rookie is brought to a baseball club on the assumption (based on his minor league performance) that he has a .300 batting average (Batting average is the ratio of the number of hits to the number of times at bat). Assume that hits can be considered Bernoulli trials with probability .3 for success.

      i. Compute the probability that if he bats 300 times, his batting average turns out to be 0.283 or less (i.e, he has 85 or less hits in these 300 times).

      Let X be the number of hits in he bats 300 times.
      From the assumption, we have that $X \sim B_{300,0.3}$.
      For a batting average of .283 or less, we need to compute the probability that the rookie had 85 hits or less:

$$\begin{aligned} P(X \leq 85) &= B_{300,.3}(85) \approx N_{300\cdot.3,300\cdot.3\cdot.7}(85) = \\ &= N_{0,1}\left(\frac{85-90}{\sqrt{63}}\right) = 1 - N_{0,1}(0.63) = 0.2643. \end{aligned}$$

      ii. Suppose he actually bats 300 times in his first year and his average for the first year turns out to be 0.283 (=85 hits). Based on your answer for (a) discuss, whether such a low average could be considered just bad luck or whether the rookie should be sent back to the minor leagues.

      Assuming that his true average is 0.300, there is a 26.43% chance that the rookie had a bad year, which is not very unlikely (not a small chance). Based on this data, there is not sufficient reason to believe that his "true" average is lower than 0.300, and hence he should be given another chance.

(c) If height of an adult in a population has an unknown distribution with mean 160 and s.d 10 (in some units), and 30 people are sampled from the population, find
(a) Distribution of $\bar{X} = \frac{1}{30}(X_1 + \ldots + X_{30})$
From Central Limit Theorem (CLT), since n is large enough, $\bar{X} \sim N(160, (\frac{10}{\sqrt{30}})^2) = N(160, (1.83)^2)$ *approximately.*
(b) Distribution of $S = X_1 + \ldots + X_{30}$
From Central Limit Theorem (CLT), since n is large enough, $S \sim N(30 \times 160, (\sqrt{30} \times 10)^2) = N(4800, (54.77)^2)$ *approximately.*
(c) Find probability that this average of 30 height observations will turn out to be more than 164.
Using (a), we get $P(\bar{X} > 164) = 1 - N_{0,1}(\frac{164-160}{1.83}) = 1 - N_{0,1}(2.19) = 0.0143$

3. **Loading Trucks** Consider a sequence of packages. Each package is loaded independently onto either a red truck (with probability p) or a green truck (with probability $1 - p$).

(a) Let R be the total number of items selected for the red truck out of the first n trucks. Determine the pmf, expected value and variance of the random variable R.
Here, $R \sim Bin(n, p)$. *So,* $P(R = k) = \binom{n}{k} p^k q^{n-k}$, $k = 0, 1, \ldots, n$. *E(R)=np, Var(R)=npq*

(b) Let X denote the first package number that is loaded in the red truck (i.e $X = 5$ means the fifth package is the first package that is loaded in the red truck). Find the pmf, expected value and variance of the random variable X.
Here $X \sim Geo(p)$. *Hence,* $P(X = k) = q^{k-1}p, k = 1, \ldots, ; E(X) = 1/p, Var(X) = q/p^2$

4. **Sum of Variables** Let X and Y be two exponential random variables with densities

$$f_X(x) = 2e^{-2x} \qquad \text{and} \qquad f_Y(y) = e^{-y}$$

(a) Let $W = X + Y$. Determine the density function f_W.
As W is the sum of two independent random variables, we can find the density function by doing a convolution:

$$f_W(w) = \int_{all\ possible\ x} f_X(x) \cdot f_Y(w - x)dx = (*)$$

As X and Y are both exponential variables, we need to make sure that we are only dealing with positive values in the density functions. This gives two conditions: $x \geq 0$ *and* $w - x \geq 0$. *The possible values for x are therefore* $0 \leq x \leq w$.

$$\begin{aligned}
(*) &= \int_0^w f_X(x) \cdot f_Y(w - x)dx = \int_0^w 2e^{-2x}e^{-(w-x)}dx = \\
&= \int_0^w 2e^{-2x-w+x}dx = \int_0^w 2e^{-x-w}dx = \\
&= 2e^{-w}\int_0^w e^{-x}dx = 2e^{-w} \cdot (-e^{-x})|_0^w \\
&= 2e^{-w}(1 - e^{-w}).
\end{aligned}$$

(b) Let $Z = Y^2$. Determine the probability that $Z > 4$.

$$\begin{aligned}
P(Z > 4) &= P(Y^2 > 2) = P(Y > 2 \cup Y < -2) = P(Y > 2) = \\
&= 1 - P(Y \leq 2) = 1 - (1 - e^{-2}) = e^{-2}.
\end{aligned}$$

(c) If (X, Y) are jointly distributed as a Bivariate normal with parameters ($\mu_X = 5, \mu_Y = 6, \sigma_X = 1, \sigma_Y = 2, \varrho = 0.5$), then find the density, f_Z of $Z = X + Y$. What is $E(Z), Var(Z), P(Z > 12)$

Here using properties of Normal random variable we get $Z \sim N(5 + 6, 1^2 + 2^2 + 2(0.5) \cdot 1 \cdot 2) = N(11, 7)$. *Hence the density is*

$$f(z) = \frac{1}{\sqrt{2\pi 7}} e^{-\frac{(z-11)^2}{2\cdot 7}}, \ -\infty < z < \infty$$

And $E(Z) = 11, Var(Z) = 7, P(Z > 12) = 1 - N_{(0,1)}(\frac{12-11}{\sqrt{7}}) = 1 - N_{(0,1)}(0.38) = 0.3520.$

5. **Continuous random Variables:**

Evaluate the following quantities for different assumptions on the random variable X given in (a), (b) and (c) below:

$$E(X), \ \ Var(X), \ \ P(X > 1), \ \ P(X < 1)$$

(a) $X \sim Exp(\lambda = 1)$

$E(X) = 1/1 = 1, Var(X) = 1/(1^2) = 1, P(X > 1) = e^{-1\cdot 1} = e^{-1} = 0.3679, P(X < 1) = 1 - 0.3679 = 0.6321.$

(b) $X \sim N(0.5, 0.5)$

$E(X) = 0.5, Var(X) = 0.5, P(X < 1) = N_{(0,1)}(\frac{1-0.5}{\sqrt{0.5}}) = N_{(0,1)}(0.71) = 0.7611, P(X > 1) = 0.2389$

(c) $X \sim Erlang(k = 2, \lambda = 1)$

$E(X) = 2 \cdot 1 = 2, Var(X) = 2 \cdot 1 = 2, P(X < 1) = 1 - Po_{1\cdot 1}(2 - 1) = 1 - Po_1(1) = 0.7356, P(X < 1) = 1 - 0.7356 = 0.2644$