

1. A continuous random variable X has cumulative distribution function given by

$$F_X(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ c \arctan(x) & \text{for } 0 \leq x \end{cases}$$

(a) What is the value of c ?

$$\lim_{x \rightarrow \infty} F_X(x) = 1, \text{ so } \lim_{x \rightarrow \infty} c \arctan(x) = c \frac{\pi}{2} = 1. \text{ Then } \boxed{c = \frac{2}{\pi}}.$$

(b) Find (as a piecewise function) the probability density function $f_X(x)$ of X .

$$\boxed{f_X(x) = \frac{d}{dx} F_X(x) = \frac{2}{\pi} \frac{1}{1+x^2} \text{ for } x \geq 0 \text{ and } f_X(x) = 0 \text{ for } x < 0.}$$

(c) Find the expected value of X .

$$\begin{aligned} E[X] &= \int_0^{\infty} x \frac{2}{\pi} \frac{1}{1+x^2} dx = \frac{1}{\pi} \int_0^{\infty} \frac{d(x^2+1)}{x^2+1} \\ &= \ln(x^2+1) \Big|_0^{\infty} = \infty \end{aligned}$$

2. The inhabitants of a certain town volunteer to do charity work at a rate of 2 volunteers per 100,000 inhabitants per month on the average. Assume the number of volunteers at any time is independent of the number of volunteers at any other time.

(a) In a town of 200,000, what is the probability that at most six people will volunteer during the next two months.

Volunteering here can be modeled as a Poisson process. The expected number of volunteers per month in a town of 200,000 is $\lambda = 4$. Then the number of volunteers in an interval of size 2 months is a Poisson random variable with parameter $\lambda = 4 \cdot 2 = 8$. Then the probability of at most 6 volunteers in the

next two months is $\boxed{e^{-8} \left(1 + \frac{8}{1!} + \frac{8^2}{2!} + \frac{8^3}{3!} + \frac{8^4}{4!} + \frac{8^5}{5!} + \frac{8^6}{6!} \right) = 0.3134}.$

An alternate model is binomial, $n = 200,000$ and $p = \frac{8}{200000} = 0.00004$, and as usual $q = 1 - p$. The Poisson random variable is a good approximation to this binomial random variable since p is very much smaller than n and np is of reasonable size. The binomial distribution gives the probability of at most 6 volunteers as

$$\sum_{i=0}^6 \binom{n}{i} p^i q^{n-i}$$

This takes a little more work to compute than the Poisson approximation, and I suspect many calculators might not be up to it. And it buys you almost nothing over the Poisson approximation. Maple computed $P\{\text{at most 6 volunteers}\} =$

0.3133693919 using the binomial distribution, whereas the Poisson computation came out as $P\{\text{at most 6 volunteers}\} = .3133742775$.

A normal distribution is not a good approximation with these numbers. Notice the comment on page 206 of the text that the normal distribution is a quite good approximation to a binomial distribution if npq is quite large, specifically mentioning $npq \geq 10$ which is not the case here. The continuity correction is essential in this approximation by a normal random variable. Maple computes $\Phi\left(\frac{(6.5 - 8)}{\sqrt{8 \cdot * (1 - 8/200000)}}\right) = .2979378689$ which is still less than the binomial cdf at 6. With p much smaller than large n but np moderate, you should use Poisson to approximate the binomial distribution. If npq is large, use normal.

- (b) What is the probability in this town of 200,000, that precisely one month of the year will have precisely 6 volunteers.

The number of volunteers in one month is a Poisson random variable with parameter $\lambda = 4$ per month. The probability that a given month will have precisely 6 volunteers and the other 11 will have a number distinct from 6 is $e^{-4} \cdot \frac{4^6}{6!} \cdot \left(1 - e^{-4} \cdot \frac{4^6}{6!}\right)^{11}$, and there are 12 distinct ways to pick the 6 volunteer month so we have 12 disjoint events, and thus the probability of precisely one

6 volunteer month is $12 \cdot e^{-4} \cdot \frac{4^6}{6!} \cdot \left(1 - e^{-4} \cdot \frac{4^6}{6!}\right)^{11} = 0.3727$.

- (c) What is the probability that, in this town of 200,000, the first person to volunteer in a given year will volunteer during the month of February.

The time when the first occurrence of a Poisson event occurs is an exponential random variable with parameter λ . Here λ is 4 per month, so we are asked the probability that the first occurrence of a person volunteering is between time

$t = 1$ months and $t = 2$ months. This probability is $\int_1^2 4 e^{-4t} dt = 1.798017626 \times 10^{-2}$

3. A fair die is tossed 6000 times. Approximate the probability that a 5 or a 6 appears on top between 1950 and 2025 times.

The requested probability is a binomial random variable with parameters $p = \frac{1}{3}$ and $n = 6000$. Its mean is $np = \frac{6000}{3} = 2000$, and its variance is $npq = \frac{6000}{3} \cdot \frac{2}{3} = \frac{12000}{9}$. We approximate this probability with a normal random variable X with mean $\mu = 2000$ and standard deviation $\sigma = \sqrt{\frac{12000}{9}} = 36.51$. Assuming that “between” refers to an open interval rather than a closed one, (this continuity correction very small with respect to the numbers involved so it should make almost no difference whether we ignore it all together or include the endpoints in the interval) we get

$$\begin{aligned} 1950.5 < X < 2024.5 &\iff -39.5 < X - 2000 < 24.5 \\ &\iff \frac{-39.5}{36.51} = -1.082 < \frac{X - 2000}{36.51} < \frac{24.5}{36.51} = 0.6710 \end{aligned}$$

so the desired probability is $\Phi(0.6710) - (1 - \Phi(1.082)) = .7486 - (1 - .85993) = 0.6085$.

4. A discrete random variable X has probability mass function

$$p(1) = .3; \quad p(3) = .2; \quad p(-1) = .1; \quad p(-2) = .25; \quad p(-3) = c$$

(a) Find the value of c .

The total probability is 1, so $c = 1 - .3 - .2 - .1 - .25 = 0.15$

(b) Find $E[X]$.

By definition $E[X] = (.3)(1) + (.2)(3) + (.1)(-1) + (.25)(-2) + (.15)(-3) = -0.15$

(c) Find the variance $\text{var}(X)$

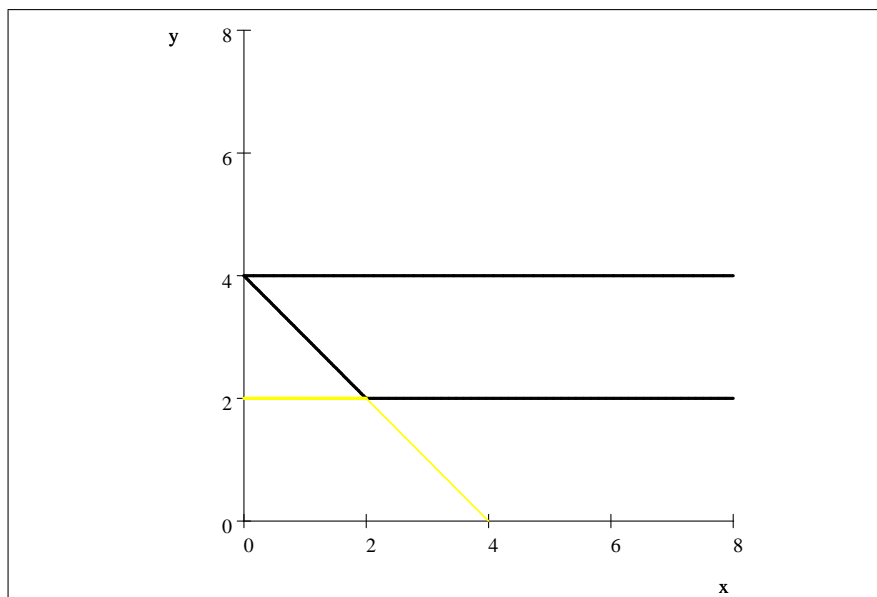
$$\begin{aligned} \text{var}(X) &= E[X^2] - (E[X])^2 \\ &= ((.3)(1)^2 + (.2)(3)^2 + (.1)(-1)^2 + (.25)(-2)^2 + (.15)(-3)^2) - (-0.15)^2 \\ &= 4.55 - 0.0225 = 4.5275 \end{aligned}$$

5. The jointly distributed random variables X and Y are independent random variables. X is exponential with parameter $\lambda = 2$, and Y has density function $f_Y(y) = \frac{y}{6}$ for $2 \leq y \leq 4$ (and f_Y is zero elsewhere). Let Z be the random variable

$$Z = X + Y.$$

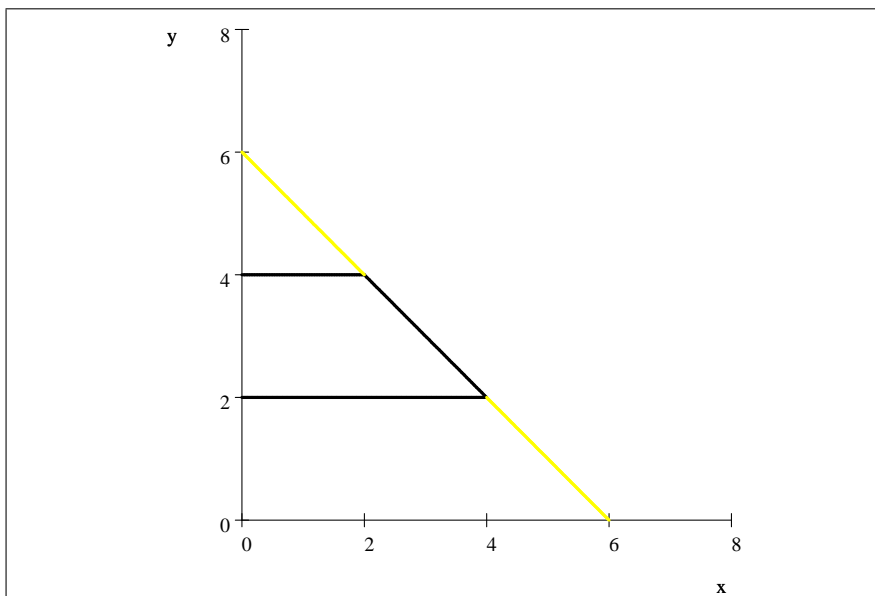
(a) Sketch the region in the plane where $4 \leq Z$ and Z has nonzero probability density.

y must be between 2 and 4 (horizontal thick lines) and x to the right of the line $x + y = 4$.



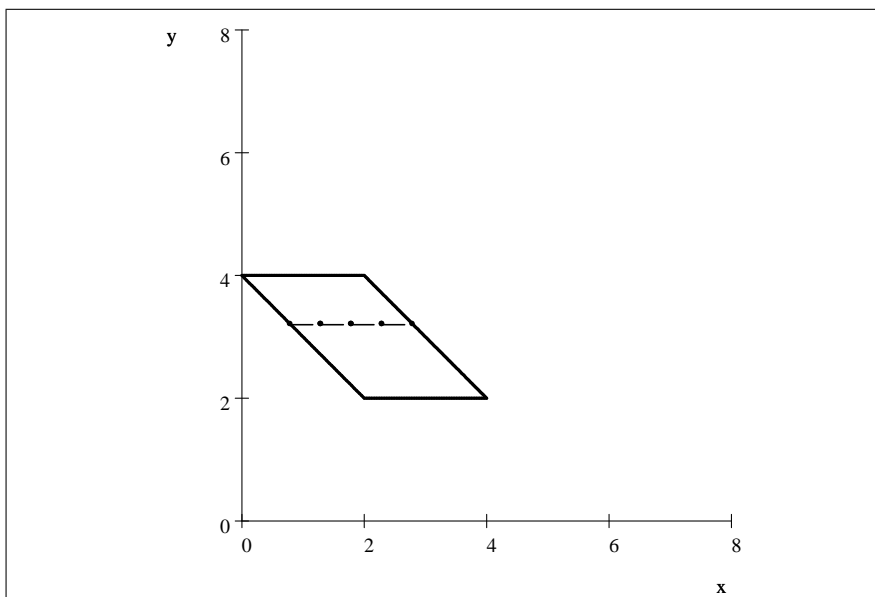
(b) Sketch the region in the plane where $Z \leq 6$ and Z has nonzero probability density.

x and y must lie between the lines $y = 2$ and $y = 4$ and between $x = 0$ and $x + y = 6$.



- (c) Find the probability that $4 \leq Z \leq 6$. In case you do not remember, integration by parts gives $\int t e^{\alpha t} dt = \left(\frac{t e^{\alpha t}}{\alpha} - \frac{e^{\alpha t}}{\alpha^2} \right)$.

We first draw a picture of the region over which to integrate based on the sketches above.



Observe that if one integrates with respect to x first, for every y , the smallest value of x is on the line $x + y = 4$ and the largest on the line $x + y = 6$. Hence

$$\begin{aligned}
 P\{4 \leq Z \leq 6\} &= \int_2^4 \int_{4-y}^{6-y} 2e^{-2x} \frac{y}{6} dx dy \\
 &= \int_2^4 \frac{y}{6} \left(\int_{4-y}^{6-y} 2e^{-2x} dx \right) dy \\
 &= \int_2^4 \frac{y}{6} [e^{-2(4-y)} - e^{-2(6-y)}] dy = 0.284077
 \end{aligned}$$

Here the integration was done using a Computer Algebra System to evaluate it numerically.

Most credit for this part was for setting up a correct iterated integral or sum of integrals.

There happens to be an alternate way of doing this particular example, which works here for three reasons but in general fails to work so is very unadvisable. The first two reasons it works are rather general. The r.v.'s X and Y are independent, and Z is their sum. The third is extremely special to this example and as written would fail to work if the roles of x and y were reversed. In this problem you can integrate with respect to x first with a single integration. You can get from what I did above to what this method does by adding and subtracting 1 inside the expression in square brackets and breaking the outer integral into a sum of two integrals with the same limits. From page 260, if the random variables X and Y are independent:

$$F_{X+Y}(a) = \int_{-\infty}^{\infty} F_X(a-y) f_Y(y) dy \quad \text{so}$$

$$F_Z(a) = \int_2^4 (1 - e^{-2(a-y)}) \frac{y}{6} dy \quad (*)$$

Then in this particular case, $P\{4 \leq Z \leq 6\} = F_Z(6) - F_Z(4) = \int_2^4 (1 - e^{-2(6-y)}) \frac{y}{6} dy - \int_2^4 (1 - e^{-2(4-y)}) \frac{y}{6} dy = 0.284077$

WARNING: the only reason this works is that integrating with respect to x first does not require splitting the double integral into a sum of more than one iterated integral. For example, what happens if you try to compute $P\{3 \leq Z \leq 6\}$ for the same X, Y, Z by using this convolution technique. If we replace a by 3 in the equation labeled $(*)$ we get $F_Z(3) \stackrel{??}{=} \int_2^4 (1 - e^{-2(3-y)}) \frac{y}{6} dy = -1.138224$ which is clearly not correct since any cumulative distribution function has values between 0 and 1. What happened? The problem is that the cumulative distribution function for the exponential random variable X with parameter 2 at the real number $3 - y$ is not $1 - e^{-2(3-y)}$, but rather

$$F_X(3-y) = \begin{cases} 1 - e^{-2(3-y)} & \text{if } 0 < 3-y \\ 0 & \text{if } 0 \geq 3-y \end{cases}$$

When $y \in (3, 4)$, $F_X(3-y)$ is zero, whereas $1 - e^{-2(3-y)} < 0$. I did not feel that it was fair to penalize students who got the correct answer here using convolution even though they did not justify it, but do not use convolution on the final without justification or using the correct piecewise c.d.f. if needed.