E1.1

Exercise session 10/2 | Probability refresher

Exercise 1. = decaan

A dean says that the average number of students per class in his faculty is 20 students. A student says that this cannot be, because he is almost always in a class with more than 90 students, and his friends have the same experience. How can they both be right? Give an illustrative example.

Exercise 2.

You want to invest your capital $c \in \mathbb{N}$ in some funds. You are given two investment strategies:

(1) invest all of your capital in a single fund with interest X,

(2) invest your capital in c different funds with interests $X_1, X_2, \ldots, X_c$.

Assume that the interests $X_1, X_2, \ldots, X_c$ are mutually independent instances of X. Which of the two strategies is the least risky?

Exercise 3.

Let A and B be independent (discrete) random variables. Prove or disprove the following statement:

$$\mathrm{E}\left[\frac{A}{B}\right] = \frac{\mathrm{E}[A]}{\mathrm{E}[B]}.$$

Exercise 4.

Few things in computing are as vital as the lowly (humble) hard drive. If your memory goes bad or your processor blows, it's easy enough to switch out; when a hard drive gives up the ghost, your precious files expire along with it. Hard drives either fail from errors during the production process or from wear during normal use. Therefore, we assume that the distribution $F(t)$ of the lifetime X is a mixture of exponential distributions:

$$F(t) = \Pr[X \le t] = p(1 - \exp(-\lambda_1 t)) + (1-p)(1 - \exp(-\lambda_2 t)),$$

$= 1 - p \cdot \exp(-\lambda_1 \cdot t) - (1-p) \cdot \exp(-\lambda_2 \cdot t)$

with $p = 0.1$, $\lambda_1 = 10$ and $\lambda_2 = 1$.

(a) Calculate the failure rate $\lambda(t) = F'(t)/(1 - F(t))$. Explain why $\lambda(t)$ is called failure rate.

(b) Find the distribution of the remaining lifetime of the hard drive, given that the hard drive has been operational for a time t.

(c) Calculate the mean of the remaining lifetime in terms of t.

Exercise 5.

You are told that the average file size in a database is 6K.

(a) Explain why it follows that fewer than half of the files can have size > 12K.

(b) You are now given the additional information that the minimum file size is 3K. Derive a tighter upper bound on the percentage of files that have size > 12K.

Exercise 6.

A company pays a fine if the time to process a request exceeds 7 seconds. Processing a request consists of two tasks: (i) retrieving the file—which takes some time X that is exponentially distributed with mean 5, and (b) parsing the file—which takes some time Y that is distributed uniformly over the interval [1, 3]. Given that the mean time to process a request is clearly 7 seconds the company views the fine as unfair, because it will have to pay the fine on half its requests. Is this right? What is the actual fraction of time that the fine will have to be paid?

Exercise 7.

Events A and B are *positively correlated* if

$$\Pr[A|B] > \Pr[A].$$

Prove that this implies that

$$\Pr[B|A] > \Pr[B].$$

Exercise 8.

A chip manufacturer produces 95% good chips and 5% bad chips. The good chips fail with probability 0.0001 each day. The bad chips fail with probability 0.01 each day. You buy a random chip. Let T be the time until your chip fails. Compute the expected time until failure $\mathrm{E}[T]$ and variance of the time until failure $\mathrm{var}[T]$