Review sheet

Preparation for Midterm 2

Probability mass function, cumulative distribution function

Exercise A. For each of the following discrete random variables N, find the probability mass function.

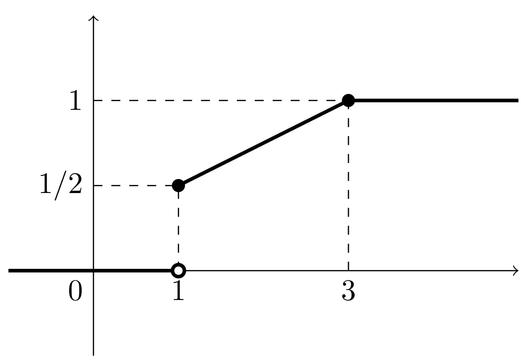
1. Select three people at random in Notre Dame. For each day of the week, count how many of them were born on this date. N is the maximum of these 7 numbers.

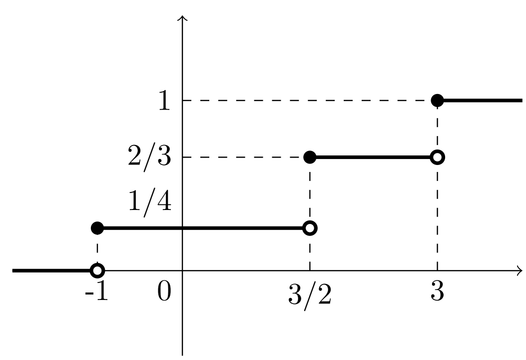
2. An ad company calls random families with 3 children, at least one of which is a girl. N is the number of girls among the children.

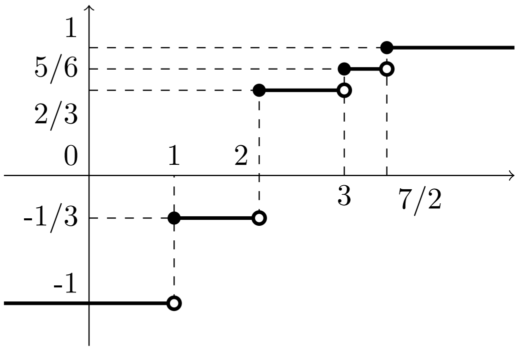
3. A cake is 10% dough, 60% filling, 30% topping. The filling can be an apple sauce (90% fruit), a cream with strawberries (50% fruit), or salted caramel (0% fruit); the topping can be a chocolate ganache (0% fruit) or a fruit salad (100% fruit). The dough is a classic crust (no fruit). A cake is selected uniformly at random among all six possible combinations. N is the proportion of fruit in the cake.

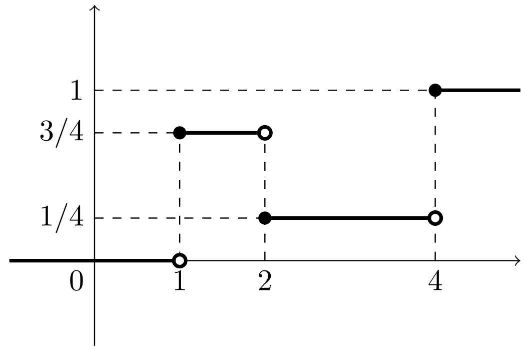
4. There are 3 independent secrets in a videogame, that you have a 30% chance of finding and that will grant you four stars each. Upgrades for your gear (initially of level zero) will cost 1 star for the first level, 2 for the second (i.e. a total of 3 stars), 3 stars for the third and so on. N is the level of your gear after you played the game.

Exercise B.









1. Which of the above graphs is the cumulative distribution function of a discrete random variable, and why?
2. Let us call N this variable. What are the probabilities $\mathbb{P}(N = 1)$, $\mathbb{P}(N = 3)$, $\mathbb{P}(N \geq 1)$?

   (In the midterm, there will be no question depending crucially on the previous one.)
3. What is the median of N? What is its 25th percentile?

   (Same remark.)

Exercise C. Let N be the outcome of a die throw.
What is the probability mass function of $(N - 2)^2$?

Usual distributions

Exercise D. Out of three variables X, Y and Z, there is one binomial, one Poisson and one negative binomial. However, we don't know which is which. What we do know is that

- $\mathbb{P}(X = 0) = 0$, $\mathbb{P}(X = 1) = 1/3$;
- $\mathbb{E}[Y] = 4$, $\mathrm{Var}(Y) = 3$;
- $\mathbb{P}(Z = 0 | Z \leq 1) = 1/4$.

Recognise each distribution, and give their parameters.

Exercise E. You store 20 loose socks in a drawer, from 10 pairs, and pick them at random, hoping to get matching ones. Identify the distribution and parameter(s) of the following random variables N.

1. You have a favourite pair there. You pick 5 socks without replacement, and N is the number of socks you get from this pair.
2. Same question, except there are only 19 socks in the drawer, because one of your favourite socks is missing.
3. You pick one sock that you wear on your left foot, then pick one at a time sock, without replacement, until you get the matching one. N is the number of socks you have picked (including the last one that makes up a pair, but not the initial one).

4. Same question with replacement: you pick one pair, then pick again until you get the matching one.

5. You don't actually want to wear socks, you're just looking for inspiration. You pick two socks together, see what emotions the (possibly non-matching) pair convey, then put them back. You do this 7 times, and N is the number of matching pairs you got.

Exercise F. You get to the bus station, and the sign says "3 buses an hour in average". Assuming buses can arrive at any time and are independent of each other, what is the probability that you don't get a bus in the next 20 minutes?

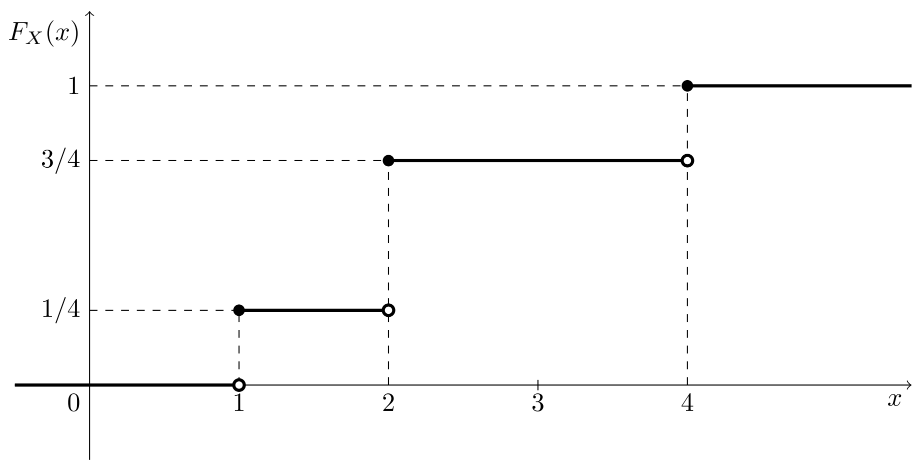
Expectation and variance of discrete variables

Exercise G. Let $X \sim \mathcal{B}in(2, 1/3)$. Compute the expectation

$$\mathbb{E}\left[\frac{1}{1+X}\right].$$

Exercise H. Compute the expectation and the variance of the following random variables.

1. X is a variable with the following cumulative distribution function:



2. X takes the following values with the following probabilities:

$$\mathbb{P}(X=-2)=\frac{1}{5}, \qquad \mathbb{P}(X=0)=\frac{3}{10}, \qquad \mathbb{P}(X=1)=\frac{1}{6}, \qquad \mathbb{P}(X=3)=\frac{1}{3}.$$

Exercise I. Let X and Y be two variables such that

$$\mathbb{E}[X]=1, \qquad \mathrm{Var}(X)=2,$$
$$\mathbb{E}[Y]=0, \qquad \mathrm{Var}(Y)=1,$$
$$\mathrm{Cov}(X,Y)=-1.$$

1. What is the covariance $\mathrm{Cov}(2X, -3Y)$?

2. Deduce the variance of $7+2X-3Y$.

Exercise J. You are hand lettering cards for the new year ('tis the season), and are considering two types of paper of different quality. The low quality one is only \$0.8 each, but it makes it easier for the ink to bleed, and for you to have to redo it all over again (you give it a 5% chance). The high quality one is \$1 for each card, but you are fairly confident it will be easy to nail it (you give it a 99% chance).

You really want to sent something great this year, so you will try as many times as necessary.

1. Which paper minimises the expected price?
2. Which paper minimises the uncertainty?

Moments, inequalities

Exercise K.

1. Let X_n be a sequence of random variables with distribution $\mathcal{B}in\left(n, \frac{1}{n}\right)$. Give an upper bound on $\mathbb{P}(X_n \geq 10)$ that does not depend on n.[1]
2. Let X_n be a sequence of random variables with distribution $\mathcal{B}in\left(n, \frac{1}{2}\right)$. Give an upper bound on
$$\mathbb{P}\left(\left|\frac{1}{n}X_n - \frac{1}{2}\right| \geq \frac{1}{10}\right)$$
that goes to zero as n goes to infinity.

Exercise L. An initial investment of \$10,000 grows by X_i% the i-th year, where $X_1, X_2, \ldots$ is a sequence of non-negative random variables. Moreover, we know that for all i,
$$\mathbb{E}\left[\ln\left(1 + \frac{X_i}{100}\right)\right] = 0.02.$$

1. Find the expected value of $\ln(Y)$, where Y is the balance after 10 years.
2. Deduce an upper bound for $\mathbb{P}(Y \geq 15,000)$.

Exercise M. Let X be a Poisson variable of parameter 1.

1. Recall that $\sum_{k=0}^{\infty} x^k/k! = \exp(x)$. Show that for any $\alpha > 0$,
$$\mathbb{E}[\alpha^X] = \exp(\alpha - 1).$$
2. Using Chebyshev's inequality, show that
$$\mathbb{P}(X \geq x) \leq \left(\frac{\mathrm{e}}{x}\right)^x.$$
Note that this goes to zero a lot faster than any exponential function.

   Hint: optimise over α.

[1] The trivial bound 1 doesn't count.