Day 4: Binomial Distribution I ★

Terms you'll find helpful in completing today's challenge are outlined below.

Random Variable

A random variable, X, is the real-valued function $X : S \to \mathbf{R}$ in which there is an event for each interval I where $I \subseteq \mathbf{R}$. You can think of it as the set of probabilities for the possible outcomes of a sample space. For example, if you consider the possible sums for the values rolled by 2 four-sided dice:

- $X = \{2, 3, 4, 5, 6, 7, 8\}$
- $P(X = 2) = P(\{(1, 1)\}) = \frac{1}{16}$
- $P(X = 3) = P(\{(1, 2), (2, 1)\}) = \frac{2}{16}$
- $P(X = 4) = P(\{(1, 3), (2, 2), (3, 1)\}) = \frac{3}{16}$
- $P(X = 5) = P(\{(1, 4), (2, 3), (3, 2), (4, 1)\}) = \frac{4}{16}$
- $P(X = 6) = P(\{(2, 4), (3, 3), (4, 2)\}) = \frac{3}{16}$
- $P(X = 7) = P(\{(3, 4), (4, 3)\}) = \frac{2}{16}$
- $P(X = 8) = P(\{(4, 4)\}) = \frac{1}{16}$

Note: When we roll two dice, the value rolled by each die is independent of the other.

Binomial Experiment

A binomial experiment (or Bernoulli trial) is a statistical experiment that has the following properties:

- The experiment consists of n repeated trials.
- The trials are independent.
- The outcome of each trial is either success (s) or failure (f).

Bernoulli Random Variable and Distribution

The sample space of a binomial experiment only contains two points, s and f. We define a Bernoulli random variable to be the random variable defined by $X(s) = 1$ and $X(f) = 0$. If we consider the probability of success to be p and the probability of failure to be q (where $q = 1 - p$), then the probability mass function (PMF) of X is:

$$p(x) = \begin{cases} 1 - p \equiv q & \text{if } x = 0 \\ p & \text{if } x = 1 \\ 0 & \text{otherwise.} \end{cases}$$

We can also express this as:

$$f(x) = p^x (1 - p)^{1-x}, \text{ for } x \in \{0, 1\}$$

Binomial Distribution

We define a binomial process to be a binomial experiment meeting the following conditions:

- The number of successes is x.

- The total number of trials is n.
- The probability of success of 1 trial is p.
- The probability of failure of 1 trial q, where $q = 1 - p$.
- $b(x, n, p)$ is the binomial probability, meaning the probability of having exactly x successes out of n trials.

The binomial random variable is the number of successes, x, out of n trials.

The binomial distribution is the probability distribution for the binomial random variable, given by the following probability mass function:

$$b(x, n, p) = \binom{n}{x} \cdot p^x \cdot q^{(n-x)}$$

Note: Recall that $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. For further review, see the Combinations and Permutations Tutorial.

Cumulative Probability

We consider the distribution function for some real-valued random variable, X, to be $F_X(x) = P(X \leq x)$. Because this is a non-decreasing function that accumulates all the probabilities for the values of X up to (and including) x, we call it the cumulative distribution function (CDF) of X. As the CDF expresses a cumulative range of values, we can use the following formula to find the cumulative probabilities for all $x \in [a, b]$:

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

Example

A fair coin is tossed 10 times. Find the following probabilities:

- Getting 5 heads.
- Getting at least 5 heads.
- Getting at most 5 heads.

For this experiment, $n = 10$, $p = 0.5$, and $q = 0.5$. The respective probabilities for the above three events are as follows:

- The probability of getting 5 heads is:

$$b(x = 5, n, p) = 0.24609375$$

- The probability of getting at least 5 heads is:

$$b(x \geq 5, n, p) = \sum_{r=5}^{10} b(x = r, n, p) = 0.623046875$$

- The probability of getting at most 5 heads is:

$$b(x \leq 5, n, p) = \sum_{r=0}^{5} b(x = r, n, p) = 0.623046875$$