Distribution

Sreeja Bolla

May 30, 2025

1 Random Variables

Random Variables: For an experiment that has a sample space S, a random variable is a function from S to $\mathbb{R}$. For example, the random variable X assigns a value X(s) to each outcome s in S.

> **Example:** A coin is flipped two times. Define the random variable X as the number of times a tails is flipped. For the sample space S = {HH, TT, HT, TH}, X(HH) = 0, X(TT) = 2, X(HT) = 1, and X(TH) = 1. By looking at the sample space in terms of this random variable S = {(0, 0), (1, 1), (0, 1), (1, 0)} and X(s_1, s_2) = s_1 + s_2.

Discrete Random Variables: A random variable X is discrete if there is a list of values, either finite or infinite, a_1, a_2, ..., such that P(X = a_i) = 1. If X is discrete, the set of values x such that P(X = x) > 0 is the support of X.

2 Probability Mass Functions

Probability Mass Function: The PMF of a discrete random variable X is the function p_X(x) = P(X = x). This is positive if x is in the support of X and 0 otherwise.

> **Example:** Continuing with the previous example, consider that a coin is flipped two times and X is the number of times a tails is flipped. Thus $p_X(0) = \frac{1}{4}$, $p_X(1) = \frac{1}{2}$, and $p_X(2) = \frac{1}{4}$.

PMF Criteria: Let X be a discrete random variable with the support x_1, x_2, ... The PMF p_X must satisfy the following:

1. $p_X(x) > 0$ for $x = x_i$ and $p_X(x) = 0$ otherwise.
2. $\sum_{i=1}^{\infty} p_X(x_i) = 1$

3 Bernoulli, Binomial, and Geometric

Bernoulli Distribution: A random variable X has the Bernoulli distribution with parameter p if P(X = 1) = p and P(X = 0) = 1 - p, where $0 < p < 0$. If this holds, then we can write $X \sim \text{Bern}(p)$.

Indicator Random Variable: An indicator random variable for an event A equals 1 when the event occurs and 0 if it doesn't. The indicator random variable for A can be written I_A or I(A). For such a variable, $I_A \sim \text{Bern}(p)$ for p = P(A).

Bernoulli Trial: An experiment that can result in either a success or a failure but not both is considered a Bernoulli trial. In such a trial, a Bernoulli random variable can be thought of as an indicator of success.

> **Bernoulli Example Problem:** A chef serves food to a critic who likes 63% of it. Consider that the random variable X = 1 if the critic enjoys his dish and equals 0 if she does not. What is the distribution of this variable? What is its PMF? Find the expected value and variance of X.
>
> **Solution:** $X \sim \text{Bern}(p = 0.63)$, since the chance that the trial will be a success is 63%. Since the two possible values are 1 and 0 for X, $p_X(x) = \begin{cases} 0.63 \text{ if } x = 1 \\ 0.37 \text{ if } x = 0 \end{cases}$. The expected value of X is the probability that the critic will like the food served to him, which is 0.63. The variance can be calculated using the formula $V(X) = E(X^2) - E(X)^2 = E(X) - E(X)^2 = 0.63 - 0.3969 = 0.2331$.

Binomial Distribution: Suppose n Bernoulli trials are conducted, each with the same success rate p. Let X be the number of successes. Then, this random variable is binomially distributed with parameters n and p, where n is a positive integer and $0 < p < 1$, and we can write $X \sim \text{Bin}(n, p)$.

Binomial PMF: If X $\sim$ Bin(n, p), then the PMF of X is P(X = k) = $\binom{n}{k}$p^k(1 - p)$^{n-k}$ for k = 0, 1, ..., n.

Binomial Failure Distribution: Let X $\sim$ Bin(n, p). Then n - X $\sim$ Bin(n, 1 - p).

Binomial Example Problem: A company manufactures 5,000 bags of flour per week, with a 3% rate of contamination due to improper handling. Consider that the random variable X is the number of contaminated bags per week. What is the distribution of this variable? What is its PMF? Find the expected value, variance, and standard deviation of X.

Solution: X $\sim$ Bin(p = 0.03, n = 5,000), since there are 5,000 bags produced per week and the chance of each of them being contaminated is 3%. Using the Binomial PMF formula, p$_X$(x) = $\binom{5000}{x}$(0.03)^x(0.97)$^{5000-x}$. The expected value of X is the average number of contaminated bags per week, which is 0.03*5,000 = 150 bags. The variance can be calculated using the formula V(X) = E(X^2) - E(X)^2. Since the variance is calculated by added the expected values at each X value, V(X) = 5,000(0.03 - 0.0009) = 145.5 bags. Since variance is standard deviation squared, the standard deviation is $\sqrt{145.5}$ bags of flour.

Geometric Distribution: Consider a series of independent Bernoulli trials, each with the same success rate p, with trials performed until a success occurs. Let X be the number of failures until the first success. Then X has a Geometric distribution with parameter p, and we can write X $\sim$ Geom(p).

Example: For an experiment where a coin is tossed twice and X is the number of tails (successes), X $\sim$ Bin(2, $\frac{1}{2}$). Let I be an indicator of the first coin landing on tails. Thus, I $\sim$ Bern($\frac{1}{2}$).

4 Hypergeometric and Discrete Uniform

Hypergeometric Distribution: Consider a bowl with w white marbles and b black marbles. We draw n marbles from the marble, such that there are $\binom{w+b}{n}$ possible samples. Let the random variable X be the number of white marbles in the sample. Then, X has a hypergeometric distribution with parameters w, b, n and X $\sim$ HGeom(w, b, n). Additionally the distribution is identical to HGeom(n, w+b-n, w)

Hypergeometric PMF: If X $\sim$ HGeom(w, b, n), then the PMF for X is P(X = k) = $\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$ for 0 $\leq$ k $\leq$ w and 0 $\leq$ n - k $\leq$ b.

Discrete Uniform Distribution and PMF: Let C be a finite, nonempty set of numbers. Choose one number at random, andlet that number be X. The X has a Discrete Uniform distribution with parameter C such that X $\sim$ DUnif(C). The PMF of X is P(X = k) = $\frac{1}{|C|}$ for k $\in$ C and 0 otherwise.

5 Normal and Poisson

Normal Distribution A random variable X has a Normal distribution if it is distributed symmetrically around its mean, and its mean, mode, and median are equal.

> **Normal Example Problem:** Before a middleweight boxing event, competitors have their weights measured to make sure they are within regulations. Their weights were normally distributed around the mean of 157 pounds, with a standard deviation of 2 pounds. Let the random variable X be the weight of a randomly selected boxer in the event. How is X distributed? Find the expected value and variance of X.
>
> **Solution:** We are given the mean of the data, 157, so the expected value is also 157 pounds. Since the variance is the standard deviation squared, V(X) = 2^2 = 4 pounds. Since the weights are distributed normally, X $\sim$ N(μ = 157, σ = 4).

Poisson Distribution A random variable X has the Poisson distribution wit parameter $\lambda > 0$ if the PMF of X is P(X = k) = $\frac{e^{-\lambda}\lambda^k}{k!}$ for k = 0, 1, 2, ... If this holds, we can write X $\sim$ Pois(λ)