DISCRETE DISTRIBUTIONS OF RANDOM VARIABLES

3. LESSON

Xabier Erdocia
Itsaso Leceta

OBJECTIVES

- ✓ Be able to identify the discrete distribution that the random variable is following
- ✓ After identifying the discrete distribution that a random variable is following, be able to calculate different probabilities using probability or distribution function
- ✓ Have the ability to calculate and interpret the moments of different discrete distributions
- ✓ Know the conditions that are necessary to approximate the hypergeometric distribution through the binomial distribution and the binomial distribution through Poisson distribution and be able to perform properly the approximation

INDEX

3.1. Binary distribution

- When there are only two possible results when performing a single random test, the distribution the random variable is following is binary distribution or Bernouilli distribution.

- The two possible results are $X = 1$ (Success) or $X = 0$ (Failure), being the probabilities $P(X = 1) = p$ and $P(X = 0) = q = 1 - p$.

- The distribution that the variable follows is given by:

$$X \sim \text{Binary}(p)$$

- The probability function is given by:

$$p(x) = p^x (1-p)^{1-x} = p^x q^{1-x}$$

- The distribution function is given by:

$$F(x) = P(X \le x) = \begin{cases} q & x = 0 \\ 1 & x = 1 \end{cases}$$

- The mean of a random variable that follows a binary distribution is:

$$E(X) = p$$

- Knowing the probability function, other moments and measurements can be calculated. For example, the value of the variance is:

$$\sigma^2 = p \cdot q$$

3.2. Binomial distribution

eman ta zabal zazu

UPV EHU

- By repeating in n occasions the random test performed in the binary distribution (being each test independent from the other), the distribution that follows the random variable which considers the results obtained in all the tests is a binomial distribution.

- In each test the two possible results are "Success" with p probability and "Failure" with $q = 1 - p$ probability.

- The distribution that the variable follows is given by :

$$X \sim \mathrm{B}(n, p)$$

- The probability function is given by:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

- The distribution function is given by:

$$F(x) = P(X \leq x) = \sum_{i=0}^{x} p(i) = \sum_{i=0}^{x} \binom{n}{i} p^i q^{n-i}$$

- The mean of a random variable that follows a binomial distribution is:

$$E(X) = n \cdot p$$

- Knowing the probability function, other moments and measurements can be calculated. For example, the value of the variance is:

$$\sigma^2 = n \cdot p \cdot q$$

3.3. Geometric distribution

- Taking into account the random test defined in the binary distribution (being each test independent from the other), the distribution that follows the random variable which considers the number of assays performed until the first success (not including the successful test) is geometric distribution.

- In each test the two possible results are "Success" with p probability and "Failure" with $q = 1 - p$ probability.

- The distribution that the variable follows is given by:

$$X \sim \mathrm{G}(p)$$

- The probability function is given by:

$$p(x) = q^x p \quad x = 0, 1, 2, \ldots$$

- The distribution function is given by:

$$F(x) = P(X \le x) = \sum_{i=0}^{x} p(i) = \sum_{i=0}^{x} q^i p$$

- The mean of a random variable that follows a geometric distribution is:

$$E(X) = \frac{q}{p}$$

- Knowing the probability function, other moments and measurements can be calculated. For example, the value of the variance is:

$$\sigma^2 = \frac{q}{p^2}$$

3.4. Negative binomial distribution

eman ta zabal zazu

UPV EHU

- Taking into account the random test defined in the binary distribution (being each test independent from the other), the distribution that follows the random variable which considers the number of tests performed until *n* successes (not including the *n* successful test) is geometric distribution

- In each test the two possible results are “Success” with p probability and “Failure” with $q = 1 - p$ probability

- The distribution that the variable follows is given by:

$$X \sim \mathrm{BN}(n, p)$$

- The probability function is given by:

$$p(x) = \binom{n+x-1}{x} q^x p^n \qquad x = 0,1,2,\ldots$$

- The distribution function is given by:

$$F(x) = P(X \le x) = \sum_{i=0}^{x} p(i) = \sum_{i=0}^{x} \binom{n+i-1}{i} q^i p^n$$

- The mean of a random variable that follows a negative binomial distribution is:

$$E(X) = \frac{nq}{p}$$

- Knowing the probability function, other moments and measurements can be calculated. For example, the value of the variance is:

$$\sigma^2 = \frac{nq}{p^2}$$

3.5. Hypergeometric distribution

eman ta zabal zazu

UPV EHU

- The hypergeometric distribution is similar to the binomial distribution but the sampling is not independent, that is, observations are made in a finite population without replacement.

- As a finite population, not returning an element to the population after observing it, influences the subsequent observation. Proportions do not remain constant as the observations advance.

- In each observation of a population of size N there are two possible results, "Success" and "Failure". In the first observation (test) the probability of success is p and the probability of failure is $q = 1 - p$.

- The distribution that the variable follows is given by:

$$X \sim \mathrm{H}(N, n, p)$$

- The probability function is given by:

$$p(x) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} \qquad \max(0, n - Nq) \le x \le \min(n, Np)$$

- The distribution function is given by:

$$F(x) = P(X \leq x) = \sum_{i=\max(0,n-Nq)}^{x} p(i) = \sum_{i=\max(0,n-Nq)}^{x} \frac{\binom{Np}{i}\binom{Nq}{n-i}}{\binom{N}{n}}$$

- The mean of a random variable that follows a negative binomial distribution is:

$$E(X) = n \cdot p$$

- Knowing the probability function, other moments and measurements can be calculated. For example, the value of the variance is:

$$\sigma^2 = n \cdot p \cdot q \frac{(N-n)}{(N-1)}$$

Approximation of hypergeometric distribution through binomial distribution:

- In a random variable that follows an hypergeometric distribution, if the size of the population, N, is much larger than the number of observations, n, this hypergeometric distribution can be approximated through a binomial distribution.
- Therefore, if $N \to \infty$ and $n \to 0$, $H(N, n, p) \cong B(n, p)$
- This approach is acceptable when: $N > 10 \cdot n$

3.6. Poisson distribution

- When an X random variable defines the number of times an event is repeated in a continuous interval (time, space...), it follows a Poisson distribution. The events occurred in an interval are independent of those produced in another interval, provided that the intervals do not overlap.

- The number of events will always be positive, the parameter λ will define the number of events expected in a given interval.

- The distribution that the variable follows is given by:

$$X \sim \mathscr{P}(\lambda)$$

- The probability function is given by:

$$p(x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!} \qquad \lambda, x \geq 0$$

- The distribution function is given by:

$$F(x) = P(X \leq x) = \sum_{i=0}^{x} p(i) = \sum_{i=0}^{x} \frac{e^{-\lambda} \cdot \lambda^{i}}{i!}$$

- The mean of a random variable that follows a negative binomial distribution is:

$$E(X) = \lambda$$

- Knowing the probability function, other moments and measurements can be calculated. For example, the value of the variance is:

$$\sigma^2 = \lambda$$

Approximation of binomial distribution through Poisson distribution:

- In a random variable that follows a binomial distribution, when the number of observations or tests, n, is very high and the probability of success, p, is low, this binomial distribution can be approaches through Poisson distribution.

- Therefore, if $n \to \infty$ and $p \to 0$, $B(n, p) \cong \mathscr{P}(n \cdot p = \lambda)$

- This approach is acceptable when: $n > 30$ and $p \leq 0.1$