Discrete Uniform, Bernoulli, and Binomial distributions

Anastasiia Kim

February 12, 2020

Variance

The variance of a random variable X is a measure of dispersion or scatter in the possible values for X

$$Var(X) = E([X - E(X)]^2) = E(X^2) - (E(X))^2$$

- For any constants a and b, $Var(aX + b) = a^2 Var(X)$
- The standard deviation is $\sqrt{Var(X)}$

Discrete Uniform Distribution

A random variable X has a discrete uniform distribution if each of the n values in its range, $x_1, x_2, ..., x_n$, has equal probability. Then

$$p(x) = \frac{1}{n}$$

Suppose that the range of the discrete random variable X equals the consecutive integers $a, a+1, a+2, ..., b$ for $a \leq b$, then

- $E(X) = \frac{b+a}{2}$
- $Var(X) = \frac{(b-a+1)^2-1}{12}$

Example

A voice communication system for a business contains 48 external lines. At a particular time, the system is observed, and some of the lines are being used. Let the random variable X denote the number of lines in use. Then X can assume any of the integer values 0 through 48.

$$E(X) = \frac{48+0}{2} = 24, Var(X) = \frac{(48-0+1)^2-1}{12} \approx 200, SD(X) = \sqrt{200} = 14.14$$

Interpretation: The average number of lines in use is 24, but the dispersion (as measured by variance or std. dev.) is large. Therefore, at many times far more or fewer than 24 lines are used.

Bernoulli random variables and distribution

Suppose that a trial, or an experiment, whose outcome can be classified as either a success or a failure is performed. If we let $X = 1$ when the outcome is a success and $X = 0$ when it is a failure, then the probability mass function of X is given by

$$p(0) = P(X = 0) = 1 - p$$

$$p(1) = P(X = 1) = p$$

where $0 \leq p \leq 1$, is the probability that the trial is a success. A random variable X is said to be a Bernoulli random variable if its probability mass function defined by the equations above. $E(X) = p, Var(X) = p(1 - p)$.

Bernoulli trials

An experiment that can result in either a "success" or a "failure" (but not both) is called a Bernoulli trial. A Bernoulli random variable can be thought of as the indicator of success in a Bernoulli trial: it equals 1 if success occurs and 0 if failure occurs in the trial.

Here are some examples:

- You take a pass-fail exam. You either pass (resulting in $X = 1$) or fail ($X = 0$).
- You toss a coin. The outcome is either heads or tails.
- A child is born. The gender is either male or female.

Binomial Distribution

A random variable X is said to be a binomial random variable $X \sim Binomial(n, p)$, if its pmf is given by

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, 2, ..., n$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient, hence the name of the distribution.
$E(X) = np$, $Var(X) = np(1-p)$

- k successes occur with probability p^k and $n-k$ failures occur with probability $(1-p)^{n-k}$
- k successes can occur anywhere among the n trials, and there are $\binom{n}{k}$ different ways of distributing k successes in a sequence of n trials

Binomial theorem

For any real numbers x and y and integer $n \geq 0$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Special case when $x = y = 1$

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

Example

It is known that screws produced by a certain company will be defective with probability 0.01, independently of each other. The company sells the screws in packages of 10 and offers a money-back guarantee that at most 1 of the 10 screws is defective. What proportion of packages sold must the company replace?
Let $X \sim \mathit{Binomial}(10, 0.01)$ is the number of defective screws in a package. Then, the probability that a package will have to be replaced is

$$1 - P(X = 0) - P(X = 1) = 1 - \binom{10}{0}(0.01)^0(0.99)^{10} - \binom{10}{1}(0.01)^1(0.99)^9 = 0.004$$

Only 0.4 percent of the packages will have to be replaced.

Example

Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Let X = the number of samples that contain the pollutant in the next 18 samples analyzed.

- Find the probability that in the next 18 samples, exactly 2 contain the pollutant
- Determine the probability that at least four samples contain the pollutant
- Find $P(3 \leq X < 7)$