Class12-14

September 21, 2020

0.1 Expectation, variance, and distribution of Random variables

Random variables assign numerical values to the outcome of an experiment or otherwise random phenomena. The study of random variables is central to describing the majority of subjects that arise in population health. The number of patients who experienced an adverse event, the proportion of a population susceptible to an infectious disease, and the association between cigarette sales and lung cancer in a community are all examples of assigning a numerical value to the outcome of a random phenomena—they are all examples of describing our world with random variables.

0.1.1 Discrete and continuous random variables

We can classify random variables into two types: (i) discrete and (ii) continuous. A **discrete** random variable takes a finite, distinct set of numerical values. For example, we can define a discrete r.v. to be the number of individuals infected with a respiratory disease in a fixed population. A **continuous** random variable takes an infinite number of numerical values. For example, we can defined a continuous r.v. to be the creatinine clearance of a patient, a continuous measurement of the rate of waste eliminated by the body through the kidneys.

Discrete random variables are often easier to conceptualize that continuous random variables.

0.1.2 Expectation and variance for discrete random variables

The **expectation** of a random variable is the sum of each value of the random variable weighted by the probability that value will occur.

If X is a random variable then the expectation of X is

$$E(X) = x_1 p(x_1) + x_2 p(x_2) + \cdots + x_n p(x_n) \tag{1}$$

$$\sum_{i=1}^{N} x_i p(x_i) \tag{2}$$

where the random variable X can take any of n values from x_1 to x_n.

The **variance** of a random variable is the sum of squared differences between each value of the random variable and it's expectation weighted by the probability that value will occur.

$$\text{Var}(X) = [x_1 - E(X)]^2 \, p(x_1) + [x_2 - E(X)]^2 \, p(x_2) + \cdots + [x_n - E(X)]^2 \, p(x_n) \tag{3}$$

$$= \sum_{i=1}^{N} [x_i - E(X)]^2 \, p(x_i) \tag{4}$$

0.1.3 Probability distribution of random variables

Just like a set of disjoint outcomes and their associated probabilities defined a probability distribution, the values of a random variable and their associated probabilities defined a probability distribution of that r.v. Some probability distributions of random variables are so common that they have been standardized. Below we will discuss five random variables, their expectation and variance, and an example of how to apply them in real-world examples using Python.

0.1.4 Bernoulli

Definition A random variable X follows a **Bernoulli** distribution if X takes either the value 0 or 1, and

$$p(x) = \begin{cases} 0 & (1 - \theta) \\ 1 & \theta \end{cases} \tag{5}$$

We write that $X \sim \text{Bern}(\theta)$, in words, that X follows a Bernoulli distribution with parameter theta.

Expectation and variance The expected value of X, if it is distributed Bernoulli is

$$E(X) = 1 \times p(X = 1) + 0 \times p(X = 0) \tag{6}$$
$$= 1 \times \theta + 0 \times (1 - \theta) \tag{7}$$
$$= \theta \tag{8}$$

The variance of X is

$$\text{Var}(X) = (1 - \theta)^2 \times p(X = 1) + (0 - \theta)^2 \times p(X = 0) \tag{9}$$
$$= (1 - \theta)^2 \times \theta + \theta^2 \times (1 - \theta) \tag{10}$$
$$= \theta(1 - \theta) \left[(1 - \theta) + \theta \right] \tag{11}$$
$$= \theta(1 - \theta) \tag{12}$$

Application The US Food and Drug Administration (FDA) allows the voluntary addition of adverse event records for devices. This data base is called MAUDE: Manufacturer and User Facility Device Experience. MAUDE is a searchable database, and I searched and extracted events reported on the RESOLUTE ONYX Drug-eluting stent, manufactured by Medtronic.

Let's load the database into Python and look at how we can defined Bernoulli random variables to describe our data.

```python
# Load Data from a uniform resource locator (url)
import pandas as pd

maudeData = pd.read_csv("")

# Pandas is a Python module for working with data
#, and it is common to abbreviate the module pandas as pd.

# Above, i used the function read_csv() from pd to load our dataset.

# We can view the first few rows of our data by using the function "head"
maudeData.head()
```

0.1.5 Geometric

0.1.6 Normal

0.1.7 Binomial

0.1.8 Poisson

```python

```

```python

```

```python

```

```python

```

```python

```