Module 6 – Poisson and Other Distributions

Reading: Sections 3.6, 3.8, 4.1 – 4.4

Geometric Distribution

- Consider an experiment with Bernoulli trials.
- Instead of a fixed number of trials, n, the trials will be conducted until a success is obtained.
- Let X be the number of trials until the first success.
- This is known as the geometric distribution and its PMF is:

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

Its mean and variance are:

$$\mu = \frac{1}{p}$$

$$\sigma^2 = \frac{(1-p)}{p^2}$$

Example: Wafer Contamination

Suppose you are inspecting a set of wafers. The probability that a wafer is contaminated is 0.01. What is the probability that you will inspect 125 wafers before selecting a contaminated wafer?

- This is a geometric distribution.
- $f(125) = (1-0.01)^{125-1}(0.01) = 0.0029$

Negative Binomial Distribution

- For a series of Bernoulli trials, the negative binomial distribution is one where the random variable equals the number of trials until r successes.
- Its PMF is:

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$$

Its mean and variance are:

$$\mu = \frac{r}{p}$$

$$\sigma^2 = \frac{r(1-p)}{p^2}$$

Example: Wafer Contamination

For the previous example, determine the probability that you have to inspect 375 wafers to find 3 contaminated wafers.

- This is a negative binomial distribution with $r = 3$
- $f(375) = \binom{375-1}{3-1}(1-0.01)^{375-3}0.01^3 = 0.00166$

Poisson Distribution

- Siméon Denis Poisson was a French mathematician, engineer, and physicist
- He formulated the Poisson process which is an experiment that satisfies the following
    - For an interval T, the average number of successes per unit interval is λ.
    - Therefore, we expect that $\mu = \lambda T$.
    - For an interval $\Delta t \to 0$, the probability of more than one event tends to zero.
    - For the same interval, the probability of one event tends to $\lambda\Delta t$
    - The events in that interval are independent of other intervals.
- These intervals are usually time intervals, though they can also represent space intervals.
- The random variable for this process is called a Poisson random variable and has the following PMF:

$$f(x) = \frac{e^{-\lambda T}(\lambda T)^x}{x!}$$

- It is implied that the intervals of an experiment are approximated as Bernoulli trials with $n = \frac{T}{\Delta t}$ number of trials and $p = \lambda\Delta t = \frac{\lambda T}{n}$ probability of an event occurring (success).
- Its mean and variance are:

$$\mu = \lambda T$$

$$\sigma^2 = \lambda T$$

Example: Wire Flaws

Suppose you are analyzing a wire to determine the number of flaws along a given length of wire. After measuring your data, you determine that the number of flaws in the wire follows a Poisson distribution with a mean of 2.3 flaws per millimeter. Determine the probability that you will find 10 flaws in 5 millimeters of wire.

- This is a Poisson distribution with $\lambda = 2.3\ mm^{-1}$ and $T = 5\ mm$
- The mean is $\lambda T = (2.3)(5) = 11.5$
- Therefore

$$f(10) = \frac{e^{-11.5}(11.5)^{10}}{10!} = 0.113$$

Continuous Random Variables

Continuous random variables don't have gaps between each outcome, and will require the use of Calculus to be studied. Three important distinctions are evident in continuous variables:

- They are always infinite.
- The probability is determined for an interval of the variable and not a value.
- In practice, where discrete measures are added, continuous measures are integrated.

Probability Density Function
Continuous random variables will have continuous probability distributions. In this case, the distribution can be described with a probability density function (PDF), $f(x)$. The PDF satisfies the following:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$

Cumulative Distribution Function of a Continuous Random Variable

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

Using the fundamental theorem of calculus, we can relate the cumulative distribution to the probability density function:

$$f(x) = \frac{dF(x)}{dx}$$

Measures of Continuous Probability Distributions

- As mentioned before, where discrete measures are added, continuous measures are integrated.
- The mean of a continuous variable is:

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

- The variance is:

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

- The standard deviation is:

$$\sigma = SD(X) = \sqrt{\sigma^2}$$

Continuous Uniform Distribution

- A random variable with equally likely outcomes will have a continuous uniform distribution.
- Its PDF is

$$f(x) = \frac{1}{b-a}$$

where a and b represent the interval of the random variable. Its mean and variance are:

$$\mu = \frac{a+b}{2}$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$

Example: Uniform Current
Let X denote the current measured in a thin cooper wire. Assume that the current ranges from 4.9 to 5.1 with a uniform PDF of $f(x) = 5$ for that range. Sketch the PDF (i.e., plot $f(x)$ vs. X) and determine the probability that the current measurement is between 4.95 and 5.0.

- The probability that the current measurement is between 4.95 and 5.0 is:

$$P(4.95 \leq x \leq 5.0) = \int_{4.95}^{5.0} 5\,dx = 0.25$$

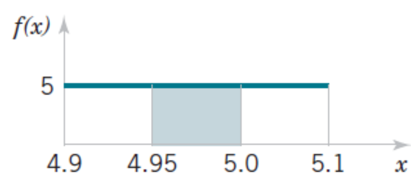
Graphical Approach to Continuous Probability Distributions

Recall that the integral of a function gives the area under the curve of that function for a given range. We can use this definition to approach continuous probability distributions graphically.

Since $P(a \leq X \leq b) = \int_a^b f(x)dx$, then the probability that an outcome will be contained within an interval is equal to the area under the PDF curve for that interval.

We could solve the previous example this way:

- The probability that the current measurement is between 4.95 and 5.0 is equal to the area under the PDF curve between 4.95 and 5 (shaded in the figure below).



Homework: 3.6.1, 3.6.9, 3.8.5, 4.1.3, 4.1.7, 4.2.5, 4.2.9, 4.3.1, 4.3.7, 4.4.2

Project Part 4: Recall the “Tools in a Bin” example from Module 3. Restate the problem in the form of a Geometric Distribution to determine the probability that the first 15 tools you select are Black & Decker. Indicate if sampling will be done with or without replacement for this experiment to have a Geometric Distribution. For your procedure, be sure to express p, x, and find the mean and standard deviation of the distribution. Show all your work and describe your steps.

References

Montgomery, Runger, *Applied Statistics and Probability for Engineers*