1 Introduction

In this lecture, we discussed continuous random variables. We have learned how to use the probability mass function (PMF) to describe a discrete random variable. Now, we will learn how to use the probability density function (PDF) to characterize a continuous random variable.

2 Continuous Random Variables

Continuous random variables can take a continuous range of values as outcomes. We can use continuous random variables to approximate discrete random variables. Using powerful calculus tools, we can significantly speed up the computation through approximation. (ex. Poisson vs Binominal)

2.1 Probability Density Function

For a continuous random variable X, its probability density function $f_X(x)$ describes its probability at any real interval. Perhaps the most striking property of PDF is that at any particular point, we have:

$$P(X = k) = \int_k^k f_X(x)dx = F_X(k) - F_X(k) = 0 \tag{1}$$

Another property of PDF is:

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_X(x)dx = 1$$

2.2 Cumulative Distribution Function

F_X mentioned in equation (1) is the cumulative distribution function of X. It computes the probability that X is less than or equal to a given value: $P(X \leq k) = F_X(k)$.

The probability of X in a real interval can be considered as the *area* under the PDF graph within that interval. $F_X(k)$ represents the area within the interval $(-\infty, k]$. It can be computed through integration:

$$F_X(k) = \int_{-\infty}^k f_X(x)dx$$

Note that one use case of CDF is to analyze the maxima of independent random variables:

$$P(max(X_1, X_2) \leq k) = P(X_1 \leq k \text{ and } X_2 \leq k)$$
$$= P(X_1 \leq k)P(X_2 \leq k)$$
$$= F_{X_1}(k)F_{X_2}(k)$$

2.3 Expected Value & Variance

For a continuous random variable X, its expected value can be computed as the following:

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

Its variance can be computed as the following:

$$Var(X) = E((x - E(X))^2) = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) dx$$

Notice that $E(X)$ is a constant, we can always move it to the outside of the integration like other constants:

$$\int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) dx = \int_{-\infty}^{\infty} (x^2 - 2xE(X) + E(X)^2) f_X(x) dx$$
$$= \int_{-\infty}^{\infty} x^2 f_X(x) dx - 2E(X) \int_{-\infty}^{\infty} x f_X(x) dx + E(X)^2 \int_{-\infty}^{\infty} f_X(x) dx$$
$$= E(X^2) - 2E(X)E(X) + E(X)^2 \times 1$$
$$= E(X^2) - E(X)^2$$

which means:

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

3 Practice Problems

1. Let X be a continuous random variable with probability density function given by:

$$f_X(x) = \begin{cases} 3x^2 & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

Find F_X and draw the graph for both f_X and F_X. Is f_X a valid probability density function?

2. Given that

$$f_X(x) = \begin{cases} cx^2 & 0 \leq x \leq 2 \\ 0 & otherwise \end{cases}$$

a) Find the value of c for which f(x) is a valid density function.
b) After determining the value of c, find $P(1 \leq x \leq 2)$ and $P(1 < x < 2)$.
c) Find $E(X)$ and $Var(X)$.

2

3. The proportion of time per day that all checkout counters in a supermarket are busy is a random variable X with density function

$$f_X(x) = \begin{cases} cx^2(x-1)^4 & 0 \le x \le 1 \\ 0 & elsewhere \end{cases}$$

a) Find the value of c that makes $f_X(x)$ a probability density function.
b) Find $E(X)$ and $Var(X)$. (hint: you can use your calculator)

4. If X is a continuous random variable such that $E((X-k)^2) < \infty$ for all k, show that $E((X-k)^2)$ is minimized when $k = E(X)$.

4 Answers

1. Yes, it is a valid probability density function because:

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 3x^2 dx = 1$$

2. a)

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^2 cx^2 dx = \frac{8c}{3} = 1$$

$$c = \frac{3}{8}$$

b)

$$P(1 < X < 2) = P(1 \le X \le 2) = \int_1^2 \frac{3}{8}x^2 dx = \frac{7}{8}$$

c)

$$E(X) = \int_{-\infty}^{\infty} xf_X(x)dx = \int_0^2 \frac{3}{8}x^3 dx = \frac{3}{2}$$

$$Var(X) = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx - (\int_{-\infty}^{\infty} xf_X(x)dx)^2 = \int_0^2 \frac{3}{8}x^4 dx - (\frac{3}{2})^2 = \frac{3}{20}$$

3. a) The trick is to let $u = x - 1$, then we can substitute x with $1 + u$ and dx with du.

$$\int_0^1 cx^2(x-1)^4 dx = \int_{-1}^0 c(1+u)^2 u^4 du = \int_{-1}^0 c(u^6 + 2u^5 + u^4)du = c\frac{1}{105}$$

$$c = 105$$

b)

$$E(X) = \frac{3}{8}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{1}{6} - (\frac{3}{8})^2 = \frac{5}{192}$$

4. Since k is a constant, we can put it outside of the integration as the following:

$$E((x-k)^2) = \int_{-\infty}^{\infty} (x-k)^2 f_X(x) dx$$

$$= \int_{-\infty}^{\infty} (x^2 - 2kx + k^2) f_X(x) dx$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x) dx - 2k \int_{-\infty}^{\infty} x f_X(x) dx + k^2 \int_{-\infty}^{\infty} f_X(x) dx$$

$$= E(X^2) - 2kE(X) + k^2 \times 1$$

$$= E(X^2) + (k - E(X))^2 - E(X)^2$$

Since $(k - E(X))^2 \geq 0$, the $E((x-k)^2)$ is minimized when $(k - E(X))^2 = 0$.