LECTURE 19

Agenda:

1. Expected values of continuous random variables
2. Tchebysheff's inequality for continuous random variables
3. ~~[illegible]~~ Examples

EXPECTED VALUES OF CONTINUOUS RANDOM VARIABLES

Definition: If X is a continuous random variable with density f_X, then the expected value of X is defined as

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

(Assuming that $\int_{-\infty}^{\infty} |x| f_X(x)\, dx < \infty$).

Notice the parallels with discrete random variables. The sum is replaced by an integral, and the probability mass function p_X is replaced by the probability density function f_X.

RESULT: If X is a continuous random variable with density f_X, then for any function $g: \mathbb{R} \to \mathbb{R}$,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx.$$

(Assuming that $\int_{-\infty}^{\infty} |g(x)| f_X(x)\, dx < \infty$).

The variance of a continuous random variable is defined in a similiar fashion as a discrete random variable.

Definition: If X is a continuous random variable with probability density function f_X, then then variance of X is given by

$$V(X) = E[(X - E(x))^2]$$

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

Let us look at an example to understand these defintions and results.

Example: For a given teller in a bank, let X denote the proportion of time, out of a 40-hour work week, that he is directly serving the customers. Suppose that X has a probability density function given by

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

$\boxed{(a)}$ Find the mean proportion of time during a 40-hour work week the teller directly serves customers.

We [redacted] required $E(X)$.

By definition of the expected value,

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

$$= \int_0^1 x.3x^2 \; dx \qquad \left(\because f_X(x) \text{ is zero if } x \text{ is outside } [0,1]. \right)$$

$$= 3 \int_0^1 x^3 \, dx$$

$$= 3 \left[\frac{x^4}{4} \right]_0^1$$

$$= \frac{3}{4}(1-0)$$

$$= \frac{3}{4}.$$

Hence, the teller on an average spends 75% of his time directly serving customers.

$\boxed{(b)}$ Find the variance of the proportion of time during a 40-hour work week the teller directly serves customers.

Before we go forward, we state some linearity properties of $E(X)$ and $V(X)$, that were true for discrete random variables, and are also true for continuous random variables.

(i) If a, b are constants, $E(aX+b) = aE(X) + b$.
(ii) If a, b are constants, $V(aX+b) = a^2 V(X)$.
(iii) $V(X) = E(X^2) - (E(X))^2$.

By definition of the expected value of $g(X)$ with $g(x) = x^2$, we get that,

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

$$= \int_0^1 x^2 \cdot 3x^2\,dx$$

$$= 3\int_0^1 x^4\,dx$$

$$= 3\left[\frac{x^5}{5}\right]_0^1$$

$$= \frac{3}{5}(1-0)$$

$$= \frac{3}{5}.$$

Hence, $V(X) = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = 0.60 - (0.75)^2 = 0.0375.$

TCHEBYSHEFF'S THEOREM FOR CONTINUOUS RANDOM VARIABLES

The Tchebysheff's theorem holds for continuous random variables in the same way as for discrete random variables.

$$\boxed{\begin{array}{l}\underline{\text{RESULT:}}\ \text{If } X \text{ is a continuous random variable with } E(X) = \mu_X \text{ and } SD(X) = \sigma_X, \text{ then for any } k>0, \\ \qquad P(|X-\mu_X| < k\sigma_X) \geq 1 - \dfrac{1}{k^2}.\end{array}}$$

There is another way to determine expectations of $\boxed{\text{non-negative}}$ continuous random variables directly from their distribution functions.

RESULT: If X is a $\boxed{\text{non-negative}}$ continuous random variable, then

$$E[X] = \int_0^\infty P(X \geq x)\,dx = \int_0^\infty [1 - F_X(x)]\,dx.$$

Example: The distribution function of the random variable X, the time (in years) from the time a machine is serviced until it breaks down, is as follows:

$$F_X(x) = \begin{cases} 1 - e^{-4x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$\boxed{(9)}$ Find $E(X)$, $V(X)$.

$$E(X) = \int_0^\infty (1 - F_X(x))\,dx \quad \left(\because X \text{ is } \boxed{\text{non-negative}}.\right)$$

$$= \int_0^\infty e^{-4x}\,dx$$

$$= \left[-\frac{e^{-4x}}{4}\right]_0^\infty$$

$$= \left\{(-0) - \left(-\frac{1}{4}\right)\right\}$$

$$= \frac{1}{4}.$$

Since $f_X(x) = \frac{d}{dx} F_X(x)$, $f_X(x) = \begin{cases} 4e^{-4x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$

It can be established from integration by parts that,

$$E[X^2] = \int_0^{\infty} x^2 f_X(x)\,dx = 4\int_0^{\infty} x^2 e^{-4x}\,dx = \frac{1}{8}.$$

Hence,

$$V(X) = E[X^2] - (E[X])^2 = \frac{1}{8} - \left(\frac{1}{4}\right)^2 = \frac{1}{16}.$$

$\boxed{(b)}$ Find an interval such that the probability that X lies in the interval is atleast 75%.

By Tchebysheff's inequality, with $k=2$,

$$P\left(\left|X - \frac{1}{4}\right| < 2\,SD(X)\right) \geq 1 - \frac{1}{4} = 0.75$$

Since $V(X) = \frac{1}{16}$, $SD(X) = \sqrt{\frac{1}{16}} = \frac{1}{4}$

Hence $P\left(\left|X - \frac{1}{4}\right| < 2 \cdot \frac{1}{4}\right) \geq 0.75$

$$\Rightarrow P\left(\left|X - \frac{1}{4}\right| < \frac{1}{2}\right) \geq 0.75$$

$$\Rightarrow P\left(-\frac{1}{4} \leq X \leq \frac{3}{4}\right) \geq 0.75$$

Hence, X lies in the interval $\left[-\frac{1}{4}, \frac{3}{4}\right]$ with probability atleast 75%.