14. Lectures 14.

14.1. Continuous Random Variables and Density Functions.

Definition. A random variable X is called <u>continuous</u> if there exists a function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that for all $c, d \in \mathbb{R}$ with $c \leq d$ the following equality holds:

$$P(c \leq X \leq d) = \int_c^d f(x)dx.$$

The function f with this property is called a <u>density function of X</u> (or probability density function, abbreviated as PDF).

Example 1. *Fix $a, b \in \mathbb{R}$, with $a < b$, and let X be a random point of the interval $[a, b]$, where all points are equally likely to occur. Then X is a continuous random variable, and its density function f is given by*

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise,} \end{cases} \qquad (***)$$

Remark: Density function f of X is not uniquely defined; for instance, if we define $f(x) = \frac{1}{b-a}$ for $a < x < b$ (replacing non-strict inequalities by strict inequalities in the above definition) and $f(x) = 0$ otherwise, this new f would still be a density function of X.

To see why f given by (***) is a density of X we first need to give a more formal definition of X. We interpret the description of X given in the example as follows: $P(a \leq X \leq b) = 1$, and for any $a \leq c \leq d \leq b$, the probability $P(c \leq X \leq d)$ should be proportional to the length of interval $[c, d]$. Since $P(a \leq X \leq b) = 1$, we must have $P(c \leq X \leq d) = \frac{c-d}{b-a}$ whenever $a \leq c \leq d \leq b$.

Now it is clear why $P(c \leq X \leq d) = \int_c^d f(x)dx$ for f given by (***) at least when $a \leq c \leq d \leq b$. Indeed, in this case $\int_c^d f(x)dx = \int_c^d \frac{1}{b-a}dx = \frac{x}{b-a}\Big|_c^d = \frac{d}{b-a} - \frac{x}{b-a} = \frac{d-c}{b-a}$. Alternatively, we can argue geometrically: $\int_c^d f(x)dx = \int_c^d \frac{1}{b-a}dx$ is the area of the rectangle which has width $d - c$ and height $\frac{1}{b-a}$, so the area is $\frac{d-c}{b-a}$.

It is easy to see that the formula $P(c \leq X \leq d) = \int_c^d f(x)dx$ remains true even without the assumption $a \leq c \leq d \leq b$. This can be formally proved using case-by-case analysis. For instance, consider the case $c < a \le d \le b$. In this case $P(c \le X \le d) = P(c \le X < a) + P(a \le X \le d) = 0 + \frac{d-a}{b-a}$ (here $P(c \le X < a)$ since the range of X is contained on $[a, b]$ and $P(a \le X \le d) = \frac{d-a}{b-a}$ by the result of the previous paragraph). On the other hand, $\int_c^d f(x)dx = \int_c^a f(x)dx + \int_a^d f(x)dx = \int_c^a 0dx + \int_a^d \frac{1}{b-a}dx = \frac{d-a}{b-a}$ as well. Other cases are treated similarly.

Definition. The random variable X in Example 1 is called the **uniform random variable on** $[a, b]$.

Before discussing the next example we state a complete characterization of density functions. Note that this characterization is very similar to the corresponding characterization of probability mass functions.

Theorem 14.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function. Then f is a density function of some random variable X if and only if*

(i) $f(x) \ge 0$ *for all* $x \in \mathbb{R}$

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$.

Example 2. *Fix $\lambda > 0$ and define $f : \mathbb{R} \to \mathbb{R}$ by*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Then f is a density function of some random variable. The corresponding random variable is called the exponential random variable with parameter λ.

We use Theorem 14.1 to check that f is indeed a density function. The condition $f(x) \ge 0$ is clear. To compute $\int_{-\infty}^{\infty} f(x)dx$ we use substitution $u = \lambda x$:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{\infty} f(x)dx = 0 + \int_{0}^{\infty} \lambda e^{-\lambda x}dx$$

$$= \int_{0}^{\infty} \lambda e^{-u}du = -e^{-u}\Big|_0^{\infty} = e^{-u}\Big|_{\infty}^{0} = e^{-0} - e^{-\infty} = 1 - 0 = 1.$$

Note that in the above computation the integration limits did not change when we switched from x to u since if $x = 0$, then to $u = \lambda \cdot 0 = 0$, and if $x \to \infty$, then $u = \lambda x \to \infty$ as well (since $\lambda > 0$).

14.2. **Expectation (mean) and variance of continuous random variables.** If we know the definition of some general concept dealing with discrete random variables, there is a general principle which can be used at least to guess what should be the right definition of the corresponding concept for continuous random variables. Informally speaking, we take the definition in the discrete case, replace all sums by integrals and replace all occurrences of PMFs by density functions. We shall now use this principle to motivate the definition of expectation of a continuous random variable (a more conceptual motivation will be discussed later in the course).

Recall that for a discrete random variable X, the expectation $E[X]$ is defined by $E[X] = \sum\limits_{x \in R(X)} x \cdot P(X = x)$ where $R(X)$ denotes the range of X. Note that we can rewrite this definition as $E[X] = \sum\limits_{x \in \mathbb{R}} x \cdot p_X(x)$ where p_X is the PMF of X. Note that adding values of x outside of $R(X)$ does not affect the sum since for all such x we have $p_X(x) = P(X = x) = 0$.

Based on the above principle, if X is a continuous random variable, the expectation $E[X]$ should be defined as $\int\limits_{\mathbb{R}} x f(x)\, dx$. This turns out to be the right definition, but we shall convert to more traditional calculus notation and write $\int\limits_{-\infty}^{\infty}$ instead of $\int\limits_{\mathbb{R}}$.

Definition. Let X be a continuous random variable with density function f.

(a) The expectation of X is the number $E[X] = \int\limits_{-\infty}^{\infty} x f(x)\, dx$ (if this integral converges)
(b) The variance of X is the number $Var(X) = E[(X - E[X])^2]$.

Note that the definition of variance in terms of expectation is the same in discrete and continuous cases.

The following theorem provides a formula for computing expectation of a function of a random variable:

Theorem 14.2. *Let X be a continuous random variable with density f, and let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then*

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

In particular, $E[X^2] = \int\limits_{-\infty}^{\infty} x^2 f(x)\, dx$.

A formal proof of this theorem is considerably more involved than the proof of the corresponding theorem in the discrete case.

The following three formulas involving expectation and variance of continuous random variables are identical to the corresponding formulas for discrete random variables:

Theorem 14.3. *Let X and Y be continuous random variables and let $a, b \in \mathbb{R}$ be some constants. The following hold:*

(a) $E[X+Y] = E[X] + E[Y]$
(b) $E[aX+b] = aE[X] + b$
(c) $Var(X) = E[X^2] - (E[X])^2$.

14.3. **Cumulative distribution functions.**

Definition. Let X be a random variable. Its cumulative distribution function (abbreviated as CDF) is the function $F_X : \mathbb{R} \to \mathbb{R}$ given by

$$F_X(x) = P(X \leq x).$$

Note that unlike the notions of probability mass functions (which are only defined for discrete random variables) and density functions (which are only defined for continuous random variables), CDFs are defined for arbitrary random variables (even the ones which are neither continuous nor discrete). However, if X is discrete or continuous, the general formula for CDF can be made more explicit in both cases.

Case 1: Let X be a discrete random variable. Then

$$F_X(x) = \sum_{y \leq x} P(X = y) \tag{14.1}$$

Example 3. *Let X be a discrete random variable with $R(X) = \{1, 3, 5\}$ and $P(X = 1) = \frac{1}{4}$, $P(X = 3) = \frac{1}{2}$ and $P(X = 5) = \frac{1}{4}$. Based on* (14.1)*, the corresponding CDF is given as follows:*

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{4} & \text{if } 1 \leq x < 3 \\ \frac{3}{4} = \frac{1}{4} + \frac{1}{2} & \text{if } 3 \leq x < 5 \\ 1 = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} & \text{if } x \geq 5 \end{cases}$$

The general principle for obtaining CDF of a discrete random variable X from its PMF is very simple – CDF will always be a piecewise-constant function, with jumps occurring at the points of the range of X. The magnitude of the jump at a point $x \in R(X)$ is equal to $P(X = x)$.

Case 2: Let X be a continuous random variable with density function f. Then

$$F_X(x) = \int_{-\infty}^{x} f(t)dt \tag{14.2}$$

Example 4. *Let X be a continuous random variable with density $f(X) = \frac{1}{\pi(x^2+1)}$ (the fact that this f is indeed a density function is verified similarly to the example with exponential random variable). Based on* (14.2)*, the CDF in this example is given by*

$$F_X(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} \frac{1}{\pi(t^2+1)}\,dt$$
$$= \frac{\arctan(t)}{\pi}\Big|_{-\infty}^{x} = \frac{\arctan(x)}{\pi} - \left(\frac{-\pi/2}{\pi}\right) = \frac{1}{2} + \frac{\arctan(x)}{\pi}.$$

We now state a complete characterization of CDFs:

Theorem 14.4. *Let $F : \mathbb{R} \to \mathbb{R}$ be a function. Then F is a CDF of some random variable X if and only if*

(i) $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to\infty} F(x) = 1$

(ii) *F is increasing (that is, $x \le y$ implies $F(x) \le F(y)$).*

(iii) *F is right-continuous at every $x \in \mathbb{R}$ (by definition this means that the right-hand limit $\lim_{y\to x^+} F(y)$ is equal to $F(x)$).*

About the proof. In [Durrett, Theorem 5.1] it is shown that CDF of any random variable satisfies all properties (i)-(iii). The proof in the reverse direction (that any function satisfying (i)-(iii) is a CDF of some random variable) is more involved and beyond the level of the course.

The fact that any CDF must satisfy (i) is actually clear: if $x \le y$, then the event $X \le x$ is a subset of the event $X \le y$, so we must have $F_X(x) = P(X \le x) \le P(X \le y) = F_X(y)$. Property (ii) takes more work to prove, but intuitively is clear as well: if $x \to \infty$, then $F(x) = P(X \le x)$ should converge to $P(X < \infty) = 1$, and similarly if $x \to -\infty$, then $F(x) = 1 - P(X > x)$ should converge to $1 - P(X > -\infty) = 1 - 1 = 0$. Finally, we shall provide an informal explanation of property (iii). Let us think of our random variable X as a mass distribution on the real line. Then part of the overall mass will be concentrated at isolated points, and part of the mass will be distributed continuously over certain intervals. The CDF F_X will be continuous (not just right-continuous) everywhere except for points with isolated masses. If x is a point containing an isolated mass m, then that mass m will contribute to the value $F_X(y)$ for every $y \ge x$ and will not contribute to $F_X(y)$ for every $y < x$. This is why $\lim_{y\to x^+} F(y)$ will equal $F(x)$, but $\lim_{y\to x^-} F(y)$ will not necessarily equal $F(x)$.

$\square$