14 The Cumulative Distribution Function

Definition The cumulative distribution function of a random variable X is the function $F_X : \mathbb{R} \to \mathbb{R}$ defined by

$$F_X(r) = \mathbb{P}(X \leq r)$$

for all $r \in \mathbb{R}$.

Proposition 14.1 (Properties of the cumulative distribution function)**.** *Let X be a random variable. Then*

(a) $0 \leq F_X(r) \leq 1$ *for all* $r \in \mathbb{R}$.

(b) *For all* $a, b \in \mathbb{R}$ *with* $a \leq b$, *we have* $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$. *In particular* $F_X(a) \leq F_X(b)$ *so* F_X *is an increasing function.*

(c) $\lim_{r\to-\infty} F_X(r) = 0$ *and* $\lim_{r\to\infty} F_X(r) = 1$.

Proof The proof follows easily from the definition of F_X. •

Remark Suppose X is a discrete random variable. If we know one of the probability mass function and cumulative distribution function of X then we can determine the other. For example, if the range of X is $\{0, 1, 2, \ldots\}$, then, for all $r \in \mathbb{R}$,

$$F_X(r) = \sum_{0 \leq i \leq r} \mathbb{P}(X = i)$$

and, for all $k \in \{0, 1, 2, ...\}$,

$$\mathbb{P}(X = k) = \mathbb{P}(X \leq k) - \mathbb{P}(X \leq k - 1) = F_X(k) - F_X(k - 1).$$

15 Continuous Random Variables

Definition A random variable X is *continuous* if its cumulative distribution function F_X is a continuous function.

If X is a continuous random variable then we must have $\mathbb{P}(X = r) = 0$ for all $r \in \mathbb{R}$. This implies that the probability mass function gives no information on the distribution of X. It also implies that $\mathbb{P}(X < r) = \mathbb{P}(X \leq r)$.

Definition Let X be a continuous random variable. Then:

- a *median* of X is a number r such that $F_X(r) = 1/2$;
- a *lower quartile* of X is a number r such that $F_X(r) = 1/4$;
- an *upper quartile* of X is a number r such that $F_X(u) = 3/4$;
- for any number k with $0 \leq k \leq 100$, a *kth percentile* of X is a number r such that $F_X(r) = k/100$.

Remark The above definition also holds for discrete random variables. However, for a discrete random variable the median (and the quartiles and percentiles) may not exist. If the random variable is continuous they are guaranteed to exist. (Which result from calculus implies this?)

Definition The *probability density function* of a continuous random variable X is the function f_X we obtain by differentiating the cumulative distribution function F_X. So

$$f_X(r) = \frac{d}{dr} F_X(r).$$

Note f_X is not defined at points where F_X is not differentiable. We can either leave it undefined at these points or give it any reasonable values. It is a fact (from calculus) that the cumulative distribution function of a continuous random variable is differentiable except possibly at a few "corners", so whatever we do will make no difference to integrals involving f_X. Everything that follows will be unaffected by the value of f_X at these "bad" points.

Proposition 15.1 (Properties of the probability density function)**.** *Let X be a continuous random variable. Then:*

(a) $f_X(r) \geq 0$ *for all* $r \in \mathbb{R}$.

(b) $F_X(r) = \int_{-\infty}^{r} f_X(t)dt$ *for all* $r \in \mathbb{R}$.

(c) $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(r)dr$ *for all* $a, b \in \mathbb{R}$ *with* $a \leq b$.

(d) $\int_{-\infty}^{\infty} f_X(r)dr = 1$.

Proof (a) Proposition 14.1(b) tells us that F_X is an increasing function, hence its derivative is non-negative.
(b) Follows from the Fundamental Theorem of Calculus.

(c) Follows from (b).
(d) Follows from (c). •

The probability density function plays a similar role in the theory of continuous random variables as the probability mass function in the theory of discrete random variables. In particular we can use it to define the expectation and variance of a continuous random variable.

Definition Suppose X is a continuous random variable with probability density function f_X. Then

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} r f_X(r) dr$$

and, if $\mathrm{E}(X) = \mu$,

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (r - \mu)^2 f_X(r) dr.$$

The properties of expectation and variance that we proved in the discrete case (Propositions 11.2 and 11.3) also hold for continuous random variables. We also have the result that, if X is a continuous random variable and $g : \mathbb{R} \to \mathbb{R}$ is a continuous function, then $g(X)$ is also a continuous random variable and

$$\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(r) f_X(r) dr.$$

In particular we may use a similar proof to to that of Propositions 11.4 to show that

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2.$$

Note In the above definitions the integrals go from $-\infty$ to ∞. However, in practice the probability density function is often 0 outside a smaller range and so we can integrate over this smaller range only (see examples in notes and on problem sheets).