Characteristic Functions

In English one uses the term "characterise" to talk about a description of all the essential features of something or somebody. In mathematics, we similar define the characters or characteristics of some mathematical object for the same purpose. The hope is that the characteristics are easier to handle than the objects themselves!

We have already seen how the mean and variance of a random variable give us some idea of the nature of its distribution. The charactersitic function of a distribution (as we shall see) defines it completely.

Recall that for a (real-valued) random variable X and a function g of one real variable, we defined the notion of mathematical expectation $E(g(X))$.

We first deal with the case of a discrete random variable X. In that case, there is a discrete set D of real numbers (such as the set of integers) so that $P(X = a) = p_a$ for $a \in D$ and $\sum_{a \in D} p_a = 1$. In this case, we have the formula

$$E(g(X)) = \sum_{a \in D} g(a) p_a$$

On the other hand suppose that Y is a random variable for which $P(Y \leq a) = \int_{-\infty}^{a} f_Y(s) ds$ for a suitable (density) non-negative function f_Y. In that case, we have the formula

$$E(g(Y)) = \int_{-\infty}^{\infty} g(s) f_Y(s) ds$$

The definitions for other random variables can be given via limits of either of these two cases.

We note that, as in the case of the mean, the expectation $E(g(X))$ only depends on the *distribution* of X so that if X and Y have the same distribution, then $E(g(X)) = E(g(Y))$.

Given a random variable X, its characteristic function is defined as the function that assigns to a real number t, the expectation $\phi_X(t) = E(e^{tX\sqrt{-1}})$. In case, you feel uncomfortable with expectations of complex numbers you can also think of this as the pair of real numbers

$$\phi_X(t) = (E(\cos(tX)), E(\sin(tX))) = E(\cos(tX)) + E(\sin(tX))\sqrt{-1}$$

We check that

$$|\phi_X(t)|^2 = E(\cos(tX))^2 + E(\sin(tX))^2 \leq E(\cos(tX)^2 + \sin(tX)^2) = 1$$

Hence, we can think of ϕ as a map from the real line to the unit disk; further, we note that $\phi_X(0) = 1$.

One can show that ϕ is a uniformly continuous function of t. Recall that this means that, for every $c > 0$, there is a constant $d > 0$ so that distance between $\phi_X(t_1)$ and $\phi_X(t_2)$) is less than c whenever $|t_2 - t_1| < d$.

We note that

$$\phi_X(-t) = E(\cos(-tX)) + E(\sin(-tX))\sqrt{-1} = E(\cos(tX)) + E(-\sin(tX))\sqrt{-1} = \overline{\phi_X(t)}$$

which is the conjugate complex number of the complex number $\phi_X(t)$. Similarly, $\phi_{-X}(t) = \overline{\phi_X(t)}$.

By arguments similar to the ones used to prove the results about variance, one can show that if X_i for $i = 1, \ldots, n$ are *independent* random variables, then

$$E(\prod_i e^{tX_i\sqrt{-1}}) = \prod_i E(e^{tX_i\sqrt{-1}})$$

It follows easily that

$$\phi_{\sum_i X_i}(t) = E(\exp(t\sum_i X_i\sqrt{-1})) = E(\prod_i e^{tX_i\sqrt{-1}}) = \prod_i \phi_{X_i}(t)$$

Thus, the sum of independent random variables has as characteristic function the product of the individual characteristic functions.

Combining the results of two identical independent experiments X and Y, we see that $\phi_{X-Y}(t) = |\phi_X(t)|^2$ is a non-negative real-valued uniformly continuous function with values between \$0 and 1.

In the case where X is a discrete random variable with distribution given by $P(X = a) = p_a$ for a in a *discrete* set D. We see that the characteristic function is:

$$\phi_X(t) = \sum_{a\in D} p_a \exp(at\sqrt{-1})$$

where $p_a \geq 0$ are the probabilities; in particular, $\sum_{a\in D} p_a = 1$. Thus,

$$\left|\sum_{a\in D} p_a \exp(at\sqrt{-1})\right| \leq \sum_{a\in D} \left|p_a \exp(at\sqrt{-1})\right| \leq \sum_{a\in D} p_a = 1$$

so the above sum converges absolutely.

Now, if D has only one element a, then $X = H_a$ is a “random” variable which takes the value a with probability 1, then its distribution function is the Heaviside function J_a that is 0 below a and 1 for a and above. In this case $\phi_{H_a}(t) = \exp(at\sqrt{-1})$ is the standard periodic function with period $2\pi/a$.

Given functions g and h and a non-negative number p with $0 \leq p \leq 1$, we can form the function $pg + (1-p)h$. This is called a *convex* linear combination of f and g; the terminology, comes from geometry where a point on the line segment joining vectors $\vec{v}$ and $\vec{u}$ is given by $p\vec{v} + (1-p)\vec{u}$ for varying values of p with $0 \leq p \leq 1$.

We see that the distribution function F_X of a general discrete random variable X takes the form

$$F_X = \sum_{a\in D} p_a J_a$$

In other words, it is a convex linear combination of the Heaviside distribution. The formula given above shows that the characteristic function is the analogous convex linear combination of the characteristic functions ϕ_{H_a}.

This is true in general. Suppose that the distribution function F is a convex linear combination $F = \sum_i p_i F_i$ of distribution functions with p_i non-negative and $\sum_i p_i = 1$. We then then have $\phi_F = \sum_i p_i \phi_{F_i}$.

Since the characteristic function is supposed to determine the distribution, we expect it to determine the mean, variance and other moments, if they exist. To start with, we note that the derivative (w.r.t. t) of $\exp(at\sqrt{-1})$ is $a\sqrt{-1}$. It follows easily that for a discrete random variable X, we have $d\phi_X/dt = E(X)\sqrt{-1}$. Similarly, $d^2\phi_X/dt^2 = -E(X^2)$. This can then be generalised to other distributions by limiting arguments.

Using the Taylor approximation, we see that

$$\phi_X(t) = 1 + E(X)t\sqrt{-1} - E(X^2)t^2 + o(t^2)$$

as t goes to 0. In particular, we see that if m is the expectation of X and σ its standard deviation, and $Y = (X - m)/\sigma$, then

$$\phi_Y(t) = 1 - t^2 + o(t^2)$$

If X_i are independent random variables with the same distribution as X and $Y_i = (X_i - m)/\sigma$, then we see that

$$\phi_{Y_1+\cdots+Y_n}(t) = (1 - t^2 + o(t^2))^n$$

So, if we put $Z_n = (Y_1 + \cdots + Y_n)/\sqrt{n}$ with is $\sqrt{n}$ *times* the rolling average $(Y_1 + \cdots + Y_n)/n$, then

$$\phi_{Z_n}(t) = \left(1 - \frac{t^2 + o(t^2)/\sqrt{n}}{n}\right)^n$$

One can then show quite easily that the right-hand side converges to $e^{-t^2/2}$ for a fixed t as n goes to infinity. This is the key to the proof of the Central Limit Theorem.

Bochner's Theorem

As usual, for a matrix M with complex entries, let $M^\dagger$ denote the transpose of the matrix whose entries are the complex conjugates; in other words $(M^\dagger)_{i,j} = \overline{M_{j,i}}$. If M is a $p \times q$ matrix then $M^\dagger$ is a $q \times p$ matrix.

Recall that a square matrix is said to be *Hermitian* if $M^\dagger = M$. We can think of $\vec{v}$ as an $1 \times r$ matrix. It follows that $\vec{v}^\dagger \cdot \vec{v}$ is an $r \times r$ Hermitian matrix.

Recall that a Hermitian matrix M is said to be positive semi-definite if the complex number (1×1 matrix) $\vec{w} \cdot M\vec{w}^{\dagger}$ is non-negative for all vectors $\vec{w}$ of the appropriate dimension. We check

$$\vec{w} \cdot \left(\vec{v}^{\dagger} \cdot \vec{v}\right) \cdot \vec{w}^{\dagger} = \langle w, v\rangle\langle v, w\rangle = |\langle w, v\rangle|^{2} \geq 0$$

This shows that $H = \vec{v}^{\dagger}\vec{v}$ is a positive semi-definite matrix.

Returning to the study of $\exp(at\sqrt{-1}$, note that for any tuple $t = (t_1, \ldots, t_r)$ of real numbers we get a vector of complex numbers

$$\vec{v_t} = \left(\exp\left(at_1\sqrt{-1}\right), \ldots, \exp\left(at_r\sqrt{-1}\right)\right)$$

It follows that $H_t = \vec{v_t}^{\dagger}\vec{v_t}$ is a positive definite matrix. We note that H_t is can directly be described as

$$(H_t)_{i,j} = \exp\left(a\left(t_i - t_j\right)\sqrt{-1}\right)$$

More generally, we can introduce the matrix $H_{t,X}$ with entries given by

$$(H_{t,X})_{i,j} = \phi_X(t_i - t_j)$$

As a *convex* linear combination of the semi-definite matrices above, it too is semi-definite.

It turns out that this is a fundamental property of characteristic functions. Bochner's theorem states that if a function $\phi(t)$ that takes values in the unit disk (in the complex plane) has the property that the matrix $H_{t,X}$ is positive semi-definite for any tuple $t_1, \ldots, t_r)$ (and for any r), and $\phi(0) = 1$ is a point of continuity of ϕ, then ϕ is the characteristic function of a probability distribution.

Moreover, one can show that:

```
- the characteristic function uniquely determines the distribution

- pointwise convergence of distribution functions is equivalent to
  pointwise convergence of the characteristic functions
```

Summary

To every random variable X we associate a characteristic function ϕ_X. This has the following properties:

- The characteristic function depends *only* on the distribution function F_X. Conversely, if $\phi_X = \phi_Y$ then X and Y have the same distribution.
- The characteristic function takes values in the unit disk in the complex plane and $\phi_X(0) = 1$.

- The characteristic function is a uniformly continuous function from the real line to the unit disk.
- In the case of the "constant" random variable H_a that takes the value a with probability 1, the characteristic function is $\exp(at\sqrt{-1})$.
- If the distribution function F_X is a convex linear combination $\sum_i p_i F_{X_i}$ of distribution functions, then we have $\phi_X = \sum_i p_i \phi_{X_i}$.
- The characteristic function of a *sum* of independent random variables is the product of their individual characteristic functions.
- We have $\phi_{aX}(t) = \phi_X(at)$.
- A function ϕ from the real line to the unit disk with $\phi(0) = 1$is the characteristic function of a random variable if and only if, for any tuple $t = (t_1, \ldots, t_j)$, the matrix with (i, j)-th entry $\phi(t_i - t_j)$ is Hermitian and positive semi-definite.
- A sequence of random variables X_i converges *in distribution* to a random variable X if and only if the characteristic functions ϕ_{X_i} converge to the characteristic function ϕ_X.
- The expectation of a random variable X can be computed from its characteristic function by the formula $d\phi_X/dt = E(X)\sqrt{-1}$. The second moment can be computed using $d^2\phi_X/dt^2 = -E(X^2)$.

The best way to understand characteristic functions is to compute them for some standard distributions. This has been given in the assignment associated with these lectures.