Solution to Homework 7

Question. Why, out of all possible representation of probabilities, the cumulative distribution function (cdf) is the most appropriate for decision making?

Answer. There are many different representations of probability: cdf, probability density function (pdf), moments, etc. To decide which representation is most appropriate for decision making, we need to recall how probabilities are used in decision making.

How are probabilities used in decision making. According to decision theory, a rational decision maker selects an alternative for which the expected utility is the largest. So, to make this decision, we need to estimate, for each possible action, its expected utility. For each action, the expected utility is determined as follows:

$$u = p_1 \cdot u(x_1) + p_2 \cdot u(x_2) + \ldots + p_n \cdot u(x_n),$$

where:

- index $i = 1, \ldots, n$ describes possible outcomes of this action,
- p_i is the probability of the i-th outcome,
- x_i is the i-th outcome, and
- $u(x_i)$ is the utility of the i-th outcome.

First case: when utility is a smooth function of the outcome. In many cases, small changes of outcome lead to equally small changes in utility, so the dependence $u(x)$ is smooth.

In such cases, it is usually convenient to expand this dependence in Taylor series and keep only several first terms in this expansion. This is, e.g., how functions like $\exp(x)$ and $\sin(x)$ are computed in the computers: we take into account that

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^N}{N!} + \ldots$$

and then estimate $\exp(x)$ as

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^N}{N!}.$$

So, we expand the dependence $u(x)$ in Taylor series and keep the first few terms in this expansion:

$$u(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \ldots + a_N \cdot x^N.$$

For this expression, the expected utility takes the following form:

$$u = p_1 \cdot (a_0 + a_1 \cdot x_1 + a_2 \cdot x_1^2 + \ldots + a_N \cdot x_1^N) +$$
$$p_2 \cdot (a_0 + a_1 \cdot x_2 + a_2 \cdot x_2^2 + \ldots + a_N \cdot x_2^N) +$$
$$\ldots +$$
$$p_n \cdot (a_0 + a_1 \cdot x_n + a_2 \cdot x_n^2 + \ldots + a_N \cdot x_n^N).$$

If we combine terms proportional to different powers of x_i, we get the following expression:

$$u = a_0 \cdot (p_1 + p_2 + \ldots + p_n) +$$
$$a_1 \cdot (p_1 \cdot x_1 + p_2 \cdot x_2 + \ldots + p_n \cdot x_n) +$$
$$a_2 \cdot (p_1 \cdot x_1^2 + p_2 \cdot x_2^2 + \ldots + p_n \cdot x_n^2) +$$
$$\ldots +$$
$$a_N \cdot (p_1 \cdot x_1^N + p_2 \cdot x_2^N + \ldots + p_n \cdot x_n^N).$$

The sum in the first parentheses is the sum of all the probabilities, i.e., 1. The sums in all other parentheses are expected values of some power of x_i. Such expected values are known as *moments*:

$$M_k = p_1 \cdot x_1^k + p_2 \cdot x_2^k + \ldots + p_n \cdot x_n^k.$$

In terms of moments, the above expression for the utility u takes the following form:

$$u = a_0 + a_1 \cdot M_1 + a_2 \cdot M_2 + \ldots + a_N \cdot M_N.$$

So: *in the case when utility is a smooth function of the outcome, to make a decision, it is desirable to known the moments.*

Second case, when the dependence of the utility on the outcome is discontinuous. In many practical situations, the dependence of utility on the outcome is discontinuous. For example, a chemical plant generates a profit as long as the level of its pollution x does not exceed the required threshold x_0. Once the pollution level is higher, the plant may be closed and the company fired, so utility will be 0 or even negative. In this case, the utility is proportional to the probability $\text{Prob}(x \le x_0)$ that $x \le x_0$. This probability is known as the value of the cumulative distribution function (cdf) $F(x)$ for $x = x_0$.

So: *in the case when utility is a discontinuous function of the outcome, to make a decision, it is desirable to known the cdf.*

First conclusion. To make a decision, it is desirable to know both the moments and the cdf.

A natural question. Both moments and cdf fully describe the probabilities. So, a natural question is: do we really need both? Maybe we can keep only one of these two characteristics?

If we know moments, it is possible but difficult to reconstruct the cdf. In principle, if we know moments, we can reconstruct cdf. We can view the definition of each moment M_k as an equation with the unknowns p_i and x_i, and solve the resulting system of non-linear equations. The problem is that solving such systems requires a lot of computations time.

On the other hand, if we know the cdf, then we can easily estimate the moments. Good news is that if we know the cdf, then we can easily reconstruct the probabilities of individual values – and thus, compute the moments. In general, if we know that the cdf has the following form, for some x_i and F_i:

- for $x < x_1$, we have $F(x) = F_0 = 0$,
- for $x_1 \le x < x_2$, we have $F(x) = F_1$,
- for $x_2 \le x < x_3$, we have $F(x) = F_2$,
- ...
- for $x_k \le x < x_{k+1}$, we have $F(x) = F_k$,
- ...
- for $x_{n-1} \le x < x_n$, we have $F(x) = F_{n-1}$, and
- for $x \ge X_n$, we have $F(x) = F_n = 1$,

this means that we have values $x_1, \ldots, x_n$ with probabilities $p_i = F_i - F_{i-1}$. Based on these values, we can easily compute each moment M_k by using the above formula for M_k. So: *if we know cdf, we can easily compute moments.*

General conclusion. For decision making, we need to know both moments and cdf. However, since the moments can be easily estimates based on cdf, it is sufficient to know cdf. Thus, indeed, *from the viewpoint of decision making, cdf is the most adequate description of probabilities.*