PSM Spring 2019, Exercise Sheet 2 - Solutions

Problem 1. If X: $\Omega \rightarrow \{1, 2\}$, Y: $\Omega \rightarrow \{\text{m, f, o}\}$, Z: $\Omega \rightarrow \{[0, 10], (10, 100]\}$, what is the size of the sample space of the RV $W = X \otimes Y \otimes Z$ (that is, how many elements does that sample space have)?

Solution: The sample spaces of X, Y, Z have sizes 2, 3, 2 respectively, thus the product space contains 2 x 3 x 2 = 12 elements.

Problem 2 (a formalization exercise) Consider a handwriting recognition system which does the job illustrated in Fig. 1 in the lecture notes. The raw input to this system is a high-resolution grayscale photographic image of an entire page of some handwritten document. Assume that this raw input is delivered by a RV X: $\Omega \rightarrow$ <the set of all grayscale images of size 2000 × 1000 pixels>. In a preprocessing stage, from such raw input images small rectangular fields like the one shown in Figure 1 are cropped. In addition to mere cropping, an "area of interest" inside the crop window is identified by the preprocessor. The complement of this area of interest is indicated in Figure 1 by the dark gray background color. Formally, this preprocessor is a transformation operation π which acts on X, giving a new RV $Y = \pi \circ X$. Your task: give a *formal* specification of the sample spaces of X and Y (not using plain English but math formalism). Your formalization should take into account the area of interest.

Solution: The sample space S_X for X, described in plain English in the problem statement as *<the set of all grayscale images of size 2000 × 1000 pixels>*, can be made formal as

$$S_X = [0\ 1]^{2000000},$$

with the convention that a grayscale value is a number in the real interval [0 1]. Since there are 2000000 pixels, each of which has a specific grayscale value, an "image" can be seen as 2000000-dimensional vector whose entries are numbers between 0 and 1. Notice that $[0\ 1]^{2000000}$ is a product of sets (all of which are the same, namely [0 1]).

There are more (and better) ways to formalize S_X. For instance, one would like to make the width and height coordinates of a pixel visible in the specification of S_X. There are 2000 height and 1000 width coordinate values (call them the y and x coordinates). The coordinate space is $C = \{1, \ldots, 2000\} \times \{1, \ldots, 1000\}$. A picture is a function p that assigns to each pixel coordinate value (y, x) a grayscale value in [0 1]. The set of all pictures then would be formalized as the set of all such functions:

$$S_X = \{p \mid p \text{ is a function from } C \text{ to } [0\ 1]\}.$$

It is perfectly admissible to use plain English in math formulas, as long as you use precisely defined math concepts in your English phrases.

Now let us proceed to formalizing the data value space S_Y for Y. You want to make formal the concept *"any rectangle within C, whose coordinate points either are* *assigned a grayscale value, or a background indicator marker".* Grayscale values are numbers in [0 1], so the background marker must be something different – we opt for the number –1. A formalization of S_Y then might look like this:

$S_Y = \{r \mid r$ is a function from $\{n, n+1,\ldots, n+h\} \times \{m, m+1\ldots, m+w\}$ to $[0\ 1] \cup \{-1\}$, where $1 \le n \le n+h \le 2000$ and $1 \le m \le m+w \le 1000\}$

Problem 3 (setting up a model for a super-complex temporal system) Global economists try to model **the global economy system** (of course, what else should they do) – just like meteorologists try to model the global atmospheric system. This is a temporal system of stunning complexity, and modeling it formally as a stochastic process is an extraordinarily difficult task. One difficulty is the heterogeneity of relevant information that has an impact on, or should even be considered part of, the global economical system. These relevant components not only comprise standard financial indicators but also factors like natural catastrophes, wars, elections, inventions... almost everything that happens on this planet.

Your task: Describe in English a suitable RSOI. (hint: think of the Weather Forecast example in the LN). Specifically, what are elementary events ω?

Solution. The global economical system (GES) is stochastic in the sense that at any point in time, for any observer whosoever, there is no certainty about how the GES will develop in the future. The reason for this uncertainty is that the current state of the GES can only be know imperfectly – it is impossible to have a *complete* description of our planet's business affairs at some point in time. This is similar to the situation in weather forecasting. The only way to go (which is exactly analog to the Weather Forecast example) is to rely on computer simulations of the GES.

In order to set them up, one needs a formal, programmable model (as rich as possible) of the GES. This model should incorporate all the laws of economics (and of politics, media, natural forces, and everything one deems relevant for economics) that one can pull together. The model must be a *dynamical* model, that is, it must yield itself to simulate "historical" evolutions of the GES; and it must be *stochastic*, that is, it must yield different "histories of the GES" if it is run multiple times. The stochasticity of the GES simulation engine can come from two sources: (i) changes in the initial state description (similar to Weather Forecasting), and/or (ii) temporal update steps that contain a random component (this is not done for weather forecasting, where the atmosphere model evolves deterministically).

The RSOI is the collection of all possible runs of the GES simulation engine. The elementary events ω are those possible runs (= the OOs in our private terminology). Note that only a few of the possible runs can actually be executed – those are the OAs.