Statistics Fall 2005

Marcus Kracht

Assignments, Part 2.

Ex 2.1 Imagine a rotating wheel (like roulette) with 10 cells, numbered 0 to 9, each of equal size. The probability of each cell is 1/10. Let $W = \{0, 1, \dots, 9\}$. What is the probability space? (Give a description only.) Now let us group the first seven cells into group A and the other three into group B. You can now only bet on either A or B. What are the odds? What is the probability space?

Solution The first space consists of all subsets U of $\{0, 1, \dots, 9\}$ with $P(U) := |U|/10$. In other words, we have a Laplace space with 10 elements. The odds of A against B are 7 : 3. The probability space is a Bernoulli space with $p = 0.7$. (A represents a win.)

Ex 2.2 Suppose you bet five times in a row on the wheel, where you can bet only A or B. Outcomes are: 0 times A, 1 time A, 2 times A, 3 times A, 4 times A and 5 times A. In general, when you are betting n times, we look at the outcome e^n_k: "the wheel turned A exactly k out of n times". This number is

$$(1) \quad p(e^n_k) := \binom{n}{k} 0.7^k 0.3^{n-k}$$

Compute the distribution of e^5. Plot the distribution of e^{10}, e^{20} and e^{50}. What do you see?

Solution The command that generates the function e^5 is `dbinom(0:5, 5, 0.7, log=FALSE)`. It gives

(2) `[1] 0.00234 0.02835 0.13230 0.36015 0.16807`

Similarly for the other cases. If one wants to plot them, there are two choices: open a file using `pdf ()`, inserting a filename if you want, and then issuing the plot commands one after the other. This will result in a pdf-file that contains one image per plot. If one wants to see them in one single graphics, one can in principle add the option `add=TRUE`, or equivalently, `add=T`. However, I did not manage to use `plot` with that option. So I decided to use `curve`, which works well. To have them in one graphics I decided to also rescale the x-coordinates to span the interval from 0 to 10. The whole input is now as follows:

(3)
```
pdf (file="overlay.pdf")
dist <- function (n, x) y <- (n/10) * x;
    choose (n, y) * 0.7 ** y * 0.3 ** (n-y)
curve (dist (5, x), from=0, to=10, 6
curve (dist (10, x), from=0, to=10, 11, add=T, col="blue"
curve (dist (20, x), from=0, to=10, 21, add=T, col="green"
curve (dist (50, x), from=0, to=10, 51, add=T, col="red"
dev.off ()
```

Notice that `curve`, unlike `plot`, wants an expression that contains `x` free, and it thinks this is the value on the x-axis and the puts the corresponding values on the y-scale.

The oberservation is that the y-coordinates shrink. This is so since the function gets "fatter": it lives on a wider interval. The sum of all values must still be 1; so they have shrink accordingly. However, it becomes more symmetrical and gets closer to the typical bell shape one is accustomed to.

Ex 2.3 Suppose you know that the wheel showed an even number. What is the probability that the outcome was A? What is the probability that it was B? What is probability that the number is even when you know that it is in A?

Solution There are 5 cases, and the numbers 0, 2, 4, 6 belong to A, while the number 8 does not. So with E denoting the set of even outcomes,

$$P(A|E) = \frac{P(E \cap A)}{P(E)} = \frac{4/10}{5/10} = \frac{4}{5} \quad (4)$$

Thus, $P(O|A) = 1 - 4/5 = 1/5$. Now a direct computation gives

$$P(E|A) = \frac{P(E \cap A)}{P(A)} = \frac{4/10}{7/10} = \frac{4}{7} \quad (5)$$

Using inverse probabilities, we get

$$(6)\qquad P(E|A) = P(A|E)\frac{P(E)}{P(A)} = \frac{4}{5}\cdot\frac{5/10}{7/10} = \frac{4}{5}\cdot\frac{5}{7} = \frac{4}{7}$$

Ex 2.4 Now we change the setup somewhat. Suppose that all you know is that the cells have been divided into two groups, one called A and the other called B. You also know that the results have recently been

(7) A B B A B A B A B B

Of course it is impossible to know exactly which cells are called A and B; the only thing we can reliably find out is how many of them are called A. For the numbers i between 0 and 10 compute the probability (a) that the outcome is as given when there are exactly i A cells, (b) the probability that there are exacty i A cells given that the outcome is as given above. For (b) assume additionally that the a priori probabilities are evenly distributed.

Solution Let $p = i/10$. The outcome is 6 times A ($E^{10}6$). It has probability $\binom{10}{6}p^4(1-p)^6$. Direct solution using R:

(8)
```
> x <- choose (10, 6) * (0:10/10) ** 4 *
    (1 - (0:10/10)) ** 6
> x
```

This gives ("beautifying" the output somewhat):

(9)
```
> x
[1] 0.000000000 0.011160261 0.088080384 0.200120949
[5] 0.250822656 0.205078125 0.111476736 0.036756909
[9] 0.005505024 0.000137781 0.000000000
```

This solves (a). Now for (b) we use the law of inverse probabilities:

$$(10)\quad P(p = k/10|E^{10}_6) = P(E^{10}_6|p = k/10)\frac{P(E^{10}_6)}{P(p = k/10)}$$

What we need is the number $P(E^{10}_6)$. This is actually the overall probability for our outcome, given that all p are equally likely *a priori*. Thus, each of the numbers above has weight 1/11 and so $P(E_6^{10})$ is simply the mean of x and $P(p = k/10) = 1/11$. Thus we calculate

(11)
```
> y <- x / (11 * (mean (x))
```

And we get

(12)
```
> y
[1] 0.0000000000 0.0122756401 0.0968833159 0.2201214418
[5] 0.2758903801 0.2255740481 0.1226179467 0.0404304689
[9] 0.0060552072 0.0001515511 0.0000000000
```