# AN OVERVIEW OF "UNIFORM AND MALLOWS RANDOM PERMUTATIONS: INVERSIONS, LEVELS & SAMPLING"

PETER RABINOVITCH

A *permutation* of a set of objects is an arrangement of those objects into a particular order. So, for example, $(3,1,4,5,2)$ is a permutation of $(1,2,3,4,5)$. In combinatorics, an area of mathematics, permutations have been studied intensely for a long time. More recently, mathematicians have studied the properties of permutations that were generated randomly, where each permutation of a given size has the same chance of being selected. In this thesis, I investigated permutations that are not selected with equal probabilities, but rather those that are selected with probabilities given by the Mallows measure.

The Mallows measure has a parameter $p$, where $0 < p$. When $p < 1$ permutations that are close to being in order, like $(1,3,2,4,5)$, are more likely, and when $p > 1$ permutations that are close to being in reverse order, like $(5,4,2,3,1)$, are more likely. If $p = 1$, then all permutations have equal chance of being chosen.

There are many interesting characteristics of permutations that mathematicians discuss. Here we focus on two: the inversions and the level.

The *inversions* is the smallest number of swaps that are needed to take the permutation back to being in order, where a *swap* is simply switching two adjacent elements. For example, in $(3,1,4,5,2)$ we can swap the third and fourth element to get $(3,1,5,4,2)$. And we can form the following sequence of swaps to get $(3,1,4,5,2)$ back in order: $(3,1,4,5,2)\to(1,3,4,5,2)\to(1,3,4,2,5)\to(1,3,2,4,5)\to(1,2,3,4,5)$. So we see that 4 swaps were sufficient. To show that 4 were actually necessary (i.e. that there is no shorter sequence of swaps that does the job), and that therefore the permutation $(3,1,4,5,2)$ has 4 inversions, is a little more complicated and won't be done here.

So, how does this relate to the Mallows measure? The *Mallows measure* assigns probability $cp^i$ to a permutation with $i$ inversions, where $p$ is the parameter mentioned above, and $c$ is a constant that is chosen so that when you add up all the probabilities of choosing all the possible permutations, they sum to 1.

We next define the level of a permutation. The *level* is the furthest that any element of the permutation has to move (to the left) to get back in order. In our example of $(3,1,4,5,2)$, the 3 doesn't have to move at all, the 1 has to move one place to the left, the 4 doesn't have to move and neither does the 5. But the 2 has to move three places to the left, and so the level of this permutation is 3.

There are all sorts of interesting questions one could ask about these things, such as if you have a permutation of $n$ objects, how many of them have $i$ inversions and level $l$? A simple question with an ugly answer:

$$\begin{aligned}\psi(n,i,l) \;=\; & \sum_{t=\left(i-\frac{l(l-1)}{2}\right)\vee l}^{(n-l)l\wedge i} \phi(l,i-t)\left(\sum_{j=0}^{(n-l)\wedge\frac{t}{l+1}\wedge\frac{n-l+t-1}{l+1}} (-1)^j\binom{n-l}{j}\binom{n-l+t-j(l+1)-1}{n-l-1}\right)\\ & -\sum_{t=\left(i-\frac{l(l-1)}{2}\right)\vee l}^{(n-l)l\wedge i} \phi(l,i-t)\left(\sum_{j=0}^{(n-l)\wedge\frac{t}{l}\wedge\frac{n-l+t-1}{l}} (-1)^j\binom{n-l}{j}\binom{n-l+t-jl-1}{n-l-1}\right)\end{aligned}$$

where $\phi(n,i)$ is the number of permutations of $n$ objects with $i$ inversions. $\phi$ is relatively well known and understood, but $\psi$ is not. There are two problems with this formula: it is ugly, and it is hard to understand what it means. Fortunately, computers care about neither and so we can plot it in Figure 0.1.

This helps a little, but not much. What I really want to understand is the following: say $n$ is very large, so my permutation is on a large number of objects, and let us for the moment assume that we're choosing permutations at random with probabilities according to the Mallows measure with $p < 1$. What I'd like to know is the average number of inversions, and the average level. Also, I'd like to know the same thing, but now with $p > 1$. The reason for looking at this in two parts ($p < 1$ and $p > 1$) is that the results are

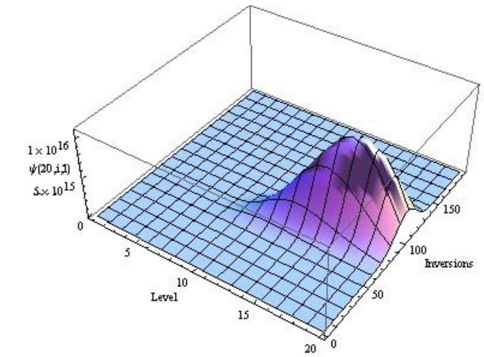


FIGURE 0.1. $\psi\left(20,i,l\right)$

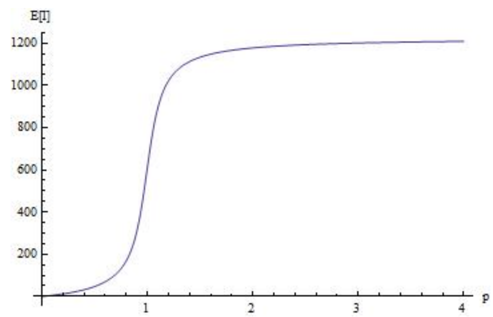


FIGURE 0.2. $\mathbb{E}_p\left[\mathbf{I}\right]$ for $n=50$

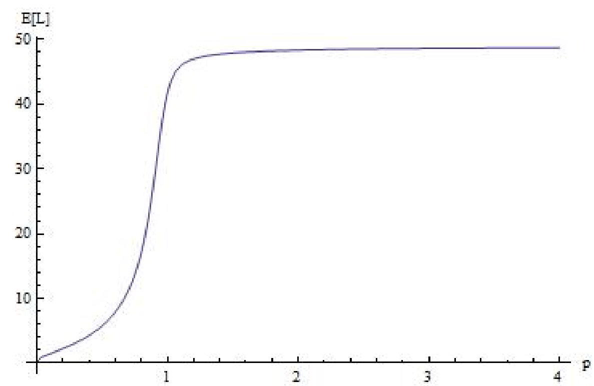


FIGURE 0.3. $\mathbb{E}_p\left[\mathbf{L}\right]$ for $n=50$

different. We can explicitely calculate them, but they are ugly - not as ugly as the big formula earlier, but still pretty ugly. So, let's plot them in Figure 0.2 and Figure 0.3. The vertical axes are $\mathbb{E}_p\left[\mathbf{I}\right]$ and $\mathbb{E}_p\left[\mathbf{L}\right]$, which is the mathematical notation for the *expected value*, or average, of the $I$nversions and the $L$evel. The horizontal axis is the parameter $p$, and the plots were done for permutations of length 50. You can clearly see the change in behaviour as $p$ crosses 1.

The next question to ask is if we can describe this behavior mathematically, and indeed we can. After much work, we prove that for fixed $p<1$ we have that $\mathbb{E}_p\left[\mathbf{I}_n\right]\sim\frac{np}{1-p}$, and $\mathbb{E}_p\left[\mathbf{L}_n\right]\ll n$ and for fixed $p>1$ we have that $\mathbb{E}_p\left[\mathbf{I}_n\right]\sim\frac{n^2}{2}$ and $\mathbb{E}_p\left[\mathbf{L}_n\right]\sim n$, where informally, $a\sim b$ means $a$ is approximately equal to $b$, and $a\ll b$ means $a$ is much less than $b$.

Finally, the question arises of how you would choose a permutation from the Mallows measure on your computer....but that's another part of the story, for another time.