# Chapter 6. Metrics on Groups, and Their Statistical Uses

In working with data, it is often useful to have a convenient notion of distance. Statisticians have used a number of different measures of closeness for permutations. This chapter begins by analyzing some applications. Then a host of natural metrics (and basic properties) is provided. Next, some abstract principles for constructing metrics on any group are shown to yield the known examples. Finally, the ideas are carried from groups to homogeneous spaces.

## A. Applications of metrics.

*Example 1. Association.* Let $\rho$ be any metric on the permutations in $S_n$. Thus, $\rho(\pi,\pi) = 0, \rho(\pi,\sigma) = \rho(\sigma,\pi)$ and $\rho(\pi,\eta) \leq \rho(\pi,\sigma) + \rho(\sigma,\eta)$. Many possible metrics will be described in Section B. To fix ideas, one might think of $\rho$ as Spearman's footrule: $\rho(\pi,\sigma) = \sum_i |\pi(i) - \sigma(i)|$. One frequent use is calculation of a measure of nonparametric association between two permutations. A standard reference is the book by Kendall (1970).

As an example, consider the draft lottery example in Figure 2 of Chapter 5. The data consists of 12 pairs of numbers, $(i, Y_i)$, and $Y_i$ being the rank of the average lottery number in month $i$. It is hard to get the value of $Y_i$ out of the figure, but easy to get the rank of $Y_i$ (i.e., biggest, next biggest, etc.). I get

| $\pi$ Month | J | F | M | A | M | J | J | A | S | O | N | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma$ Rank $Y_i$ | 5 | 4 | 1 | 3 | 2 | 6 | 8 | 9 | 10 | 7 | 11 | 12 |

The two rows can be thought of as two permutations in $S_{12}$. Are they close together? Taking $\rho$ as the footrule, $\rho(\pi, \sigma) = 18$. Is this small? The largest value $\rho$ can take is 72. This doesn't help much. One idea is to ask how large $\rho(\pi, \sigma)$ would be if $\sigma$ were chosen at random, uniformly. Diaconis and Graham showed the following result (proved in Section B below).

**Theorem 1.** *Let $\rho(\pi, \sigma) = \sum |\pi(i) - \sigma(i)|$. If $\sigma$ is chosen uniformly in $S_n$ then*

$$AV(\rho) = \frac{1}{3}(n^2 - 1)$$

$$Var(\rho) = \frac{1}{45}(n+1)(2n^2+7)$$

$$P\{\frac{\rho - AV}{SD} \leq t\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx + o(1).$$

In the example, $AV \doteq 47.7, SD \doteq 9.23$. The value 18 is more than 3 standard