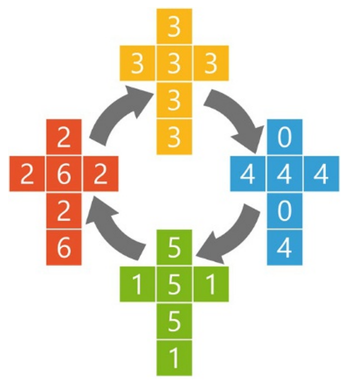
limited compute budget and an ample source of training data, it will often be better to apply maximum likelihood techniques, generally augmented with one or more forms of regularization, to a large neural network rather than apply a Bayesian treatment to a much smaller model.

Exercises

2.1 ($\star$) In the cancer screening example, we used a prior probability of cancer of $p(C = 1) = 0.01$. In reality, the prevalence of cancer is generally very much lower. Consider a situation in which $p(C = 1) = 0.001$, and recompute the probability of having cancer given a positive test $p(C = 1|T = 1)$. Intuitively, the result can appear surprising to many people since the test seems to have high accuracy and yet a positive test still leads to a low probability of having cancer.

2.2 ($\star\star$) Deterministic numbers satisfy the property of *transitivity*, so that if $x > y$ and $y > z$ then it follows that $x > z$. When we go to random numbers, however, this property need no longer apply. Figure 2.16 shows a set of four cubical dice that have been arranged in a cyclic order. Show that each of the four dice has a 2/3 probability of rolling a higher number than the previous die in the cycle. Such dice are known as *non-transitive dice*, and the specific examples shown here are called *Efron dice*.

Figure 2.16 An example of non-transitive cubical dice, in which each die has been 'flattened' to reveal the numbers on each of the faces. The dice have been arranged in a cycle, such that each die has a 2/3 probability of rolling a higher number than the previous die in the cycle.



2.3 ($\star$) Consider a variable $\mathbf{y}$ given by the sum of two independent random variables $\mathbf{y} = \mathbf{u} + \mathbf{v}$ where $\mathbf{u} \sim p_{\mathbf{u}}(\mathbf{u})$ and $\mathbf{v} \sim p_{\mathbf{v}}(\mathbf{v})$. Show that the distribution $p_{\mathbf{y}}(\mathbf{y})$ is given by

$$p(\mathbf{y}) = \int p_{\mathbf{u}}(\mathbf{u})p_{\mathbf{v}}(\mathbf{y} - \mathbf{u})\,\mathrm{d}\mathbf{u}. \tag{2.119}$$

This is known as the *convolution* of $p_{\mathbf{u}}(\mathbf{u})$ and $p_{\mathbf{v}}(\mathbf{v})$.

2.4 ($\star\star$) Verify that the uniform distribution (2.33) is correctly normalized, and find expressions for its mean and variance.

2.5 ($\star\star$) Verify that the exponential distribution (2.34) and the Laplace distribution (2.35) are correctly normalized.