A <u>probability distribution</u> is a function $\mu : \mathbb{Z}_p \to [0,1]$ satisfying

$$\sum_{t=0}^{p-1} \mu(t) = 1$$

and we imagine that μ is telling us the likelihoods of outcomes t from some random procedure: $0 \leq \mu(t) \leq 1$ is the probability that t is the outcome of the random procedure. We could ask, given $A \subset \mathbb{Z}_p$, what the likelihood of the outcome belonging to A is? It is

$$\mu(A) = \sum_{t \in A} \mu(t)$$

for any $A \subset \mathbb{Z}_p$. Our main goal today is to describe and relate two ways of measuring how "uncertain" a probability distribution is: how difficult it is to predict the outcome of the random procedure. At the extremes we have the Lebesgue distribution (most difficult to predict) and the dirac distributions (easy to predict). The two ways of measuring uncertainty are <u>total variation distance</u> and <u>entropy</u>.

Total Variation Distance

Given probability distributions μ and ν the <u>total variation distance</u> between μ and ν is

$$d(\mu,\nu) = \max\{|\mu(A) - \nu(A)| : A \subset \mathbb{Z}_p\}$$

i.e. the farthest apart $\mu(A)$ and $\nu(A)$ can be. This distance is in fact a metric on the set of all probability distributions on $\mathbb{Z}_p$. Given a probability distribution μ on $\mathbb{Z}_p$ we are especially interested in $d(\lambda,\mu)$ where (as usual) λ is the Lebesgue distribution. The

Lebesgue distribution is difficult to predict. We consider another distribution μ to be harder to predict the closer it is to λ.

L^1 identity

Calculating a max is not very easy: there are many subsets of $\mathbb{Z}_p$. The following theorem gives us an easier way to calculate $d(\mu,\nu)$.

Theorem For any probability distributions μ, ν on $\mathbb{Z}_p$ one has

$$d(\mu,\nu) = \frac{1}{2}\sum_{t=0}^{p-1} |\mu(t) - \nu(t)|$$

Example Calculate $d(\lambda, \delta_0)$.

Solution: By the theorem

$$\begin{aligned} d(\lambda,\delta_0) &= \tfrac{1}{2}\sum_{t=0}^{p-1} |\lambda(t) - \delta_0(t)| \\ &= \tfrac{1}{2}\sum_{t=0}^{p-1} \left|\tfrac{1}{p} - \delta_0(t)\right| \\ &= \tfrac{1}{2}\left|\tfrac{1}{p} - \delta_0(0)\right| + \tfrac{1}{2}\sum_{t=1}^{p-1} \left|\tfrac{1}{p} - \delta_0(t)\right| \\ &= \tfrac{1}{2}\left|\tfrac{1}{p} - 1\right| + \tfrac{1}{2}\sum_{t=1}^{p-1} \left|\tfrac{1}{p}\right| \\ &= \tfrac{1}{2}\left(1 - \tfrac{1}{p} + \tfrac{p-1}{p}\right) \\ &= \frac{p-1}{p} \end{aligned}$$

□

Are they close or far apart? By the triangle inequality

$$\begin{aligned} d(\mu,\nu) = \tfrac{1}{2}\sum_{t=0}^{p-1} |\mu(t) - \nu(t)| &\le \tfrac{1}{2}\sum_{t=0}^{p-1} |\mu(t)| + |\nu(t)| \\ &= \tfrac{1}{2}\sum_{t=0}^{p-1} \mu(t) + \tfrac{1}{2}\sum_{t=0}^{p-1} \nu(t) = \tfrac{1}{2} + \tfrac{1}{2} = 1 \end{aligned}$$

so distributions μ and ν on $\mathbb{Z}_p$ cannot be farther than a distance of one apart. If p is large then $\frac{p-1}{p}$ is close to 1, so we can think of λ and δ_0 as being far apart. This agrees with the idea that λ is unpredictable while δ_0 is very predictable: being close to λ should mean unpredictable, but δ_0 is predictable thus far from λ.

Entropy Entropy, which originated in Shannon's work on statistical mechanics, provides a second means of measuring uncertainty. Given a probability distribution $\mu : \mathbb{Z}_p \to [0,1]$ its <u>entropy</u> is

$$H(\mu) = -\sum_{t=0}^{p-1} \mu(t) \log \mu(t)$$

a number which we think of as measuring how surprised one can expect to be at the outcome of a random procedure with distribution μ.

<u>Example</u> Calculate $H(\lambda)$ and $H(\delta_0)$.

Solution: $H(\lambda) = -\sum_{t=0}^{p-1} \lambda(t) \log \lambda(t) = -\sum_{t=0}^{p-1} \frac{1}{p} \log \frac{1}{p} = -p \cdot \frac{1}{p} \log \frac{1}{p} = \log p$

$$H(\delta_0) = -\sum_{t=0}^{p-1} \delta_0(t) \log \delta_0(t) = -\delta_0(t) \log \delta_0(t) = -0 \log(0) = 0$$

We should not be surprised at all by the outcome of a random procedure with distribution δ_0, but should be very surprised by the outcome of a random procedure with distribution λ.

A relation We have two notions of uncertainty for a distribution on $\mathbb{Z}_p$

- distance $d(\lambda, \mu)$ from λ
- entropy

It turns out they are related.

<u>Theorem</u> For every probability distribution $\mu : \mathbb{Z}_p \to [0,1]$ we have

$$\frac{1}{2(H(\lambda)+1)} |H(\mu) - H(\lambda)| \leq d(\mu, \lambda) \leq \sqrt{2 |H(\mu) - H(\lambda)|}$$

If μ is δ_0 then $H(\delta_0) = 0$ and $d(\mu, \lambda) = \frac{p-1}{p}$ and plugging in gives

$$\frac{\log p}{2(\log p + 1)} \leq \frac{p-1}{p} \leq \sqrt{2 \log p}$$

There is another way to calculate the total variation distance $d(\mu, \nu)$.

Given a function $f : \mathbb{Z}_p \to \mathbb{R}$ write $\|f\|_\infty = \max\{|f(t)| : t \in \mathbb{Z}_p\}$ for its L^∞ norm.

Theorem

$$d(\mu, \nu) = \frac{1}{2} \max\{|\mu(f) - \nu(f)| : \|f\|_\infty \leq 1, f : \mathbb{Z}_p \to \mathbb{R}\}$$