COMP 585 **Spring 2025**

Probabilistic Toolkit for Learning and Computing *Instructor: Maryam Aliakbarpour*

February 6, 2025 *Scriber: Maryam Aliakbarpour*

Lecture 7

Reducing the L_2 Norm of a Distribution via Flattening

A useful technique in distribution testing is reducing the L_2 norm of a distribution. This can be achieved through *flattening*, a process that transforms the original distribution p into a new distribution p'. The core idea is to distribute the probability mass of elements with high probability in p among multiple elements in p', effectively "flattening" the distribution and reducing its L_2 norm. This process is illustrated in Figure 1.

The transformation from p to p' involves determining, for each element i in the domain $[n]$, the number of elements in p' that will correspond to i. Let b_i denote this number for element i. We will discuss how to choose b_i shortly; for now, assume b_i is given. For each element i in the domain of p, we associate b_i elements in the domain of p' with i. We refer to these associated elements as "buckets." The probability mass of i in p is then distributed equally among its b_i buckets in p'.

Formally, we define:

$$\text{New domain of } p': \qquad D' \coloneqq \{(i,j) \mid i \in [n], j \in [b_i]\} \tag{1}$$

$$\text{New domain size:} \qquad |D'| = \sum_{i=1}^{n} b_i \tag{2}$$

$$\text{Probability of a domain element in } p': \qquad p'_{(i,j)} = \frac{p_i}{b_i} \tag{3}$$

It is straightforward to verify from this definition that the probabilities in p' sum to one.

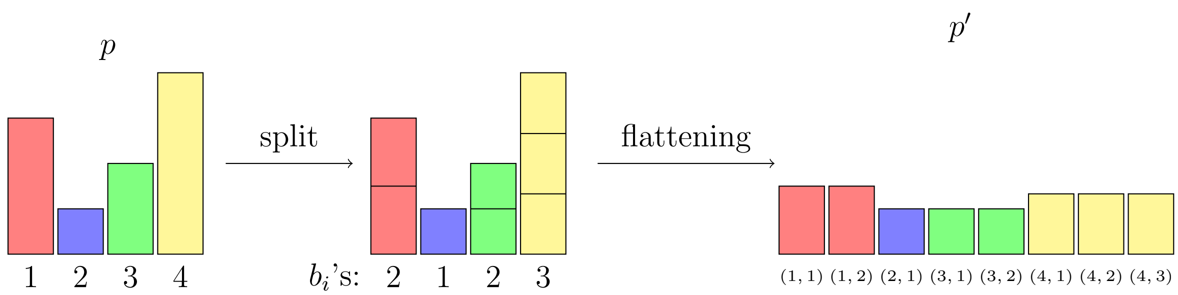

Figure 1: Flattening of a distribution p to p'

This transformation ideally has the following properties:

1. **Access:** For known b_i's, we have the same access to p' as to p. That is, if we know the probabilities of every element in p, we also know the probabilities of every element in p'. Similarly, if we have sample access to p, we have sample access to p'.

2. **Preservation of L_1 distance:** For a fixed set of b_i's, the transformation preserves the L_1 distance between distributions.

3. **Reduced L_2 norm:** We can choose b_i's such that the L_2 norm of p' is low.

The first two properties are straightforward to establish for any choice of b_i's. Regarding access, if p is known, the probability of each bucket (i, j) is given by Equation (3). To sample from p', we can use a sample from p. Specifically, if i is a sample drawn from p, we can then draw a uniform random sample j from $[b_i]$. The pair (i, j) then constitutes a sample from p'.

Next, we show that if we flatten p and q using the same set of b_i's, the resulting distributions p' and q' have the same L_1 distance as p and q. Formally,

$$\begin{aligned}\|p' - q'\|_1 &= \sum_{i\in[n]} \sum_{j\in[b_i]} |p'_{(i,j)} - q'_{(i,j)}| = \sum_{i\in[n]} \sum_{j\in[b_i]} \left|\frac{p_i}{b_i} - \frac{q_i}{b_i}\right| \\ &= \sum_{i\in[n]} |p_i - q_i| \sum_{j\in[b_i]} \frac{1}{b_i} \sum_{i\in[n]} |p_i - q_i| = \|p - q\|_1 \ .\end{aligned}$$

Determining the Number of Buckets

For the third property (reduced L_2 norm), we need to carefully choose the number of buckets, b_i, for each element i. Ideally, we want to decompose elements of p with high probability into smaller pieces, effectively "flattening" the distribution and making it more uniform. Thus, we aim for b_i to be proportional to p_i. Various methods exist for determining the b_i values. Here, we focus on the approach proposed in [DK16]. This approach is illustrated in Figure 2. Given a parameter k, the method proceeds as follows:

1. Draw k' from a Poisson distribution with mean k (i.e., $k' \leftarrow \text{Poi}(k)$).

2. Draw a set F of k' independent samples from p.

3. For each $i \in [n]$, let f_i denote the frequency of element i in F.

4. For each $i \in [n]$, set $b_i = f_i + 1$.

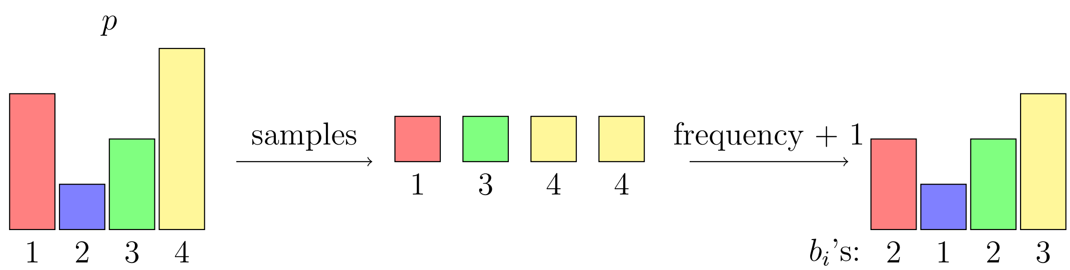

Figure 2: Calculating b_i's from samples of p

Let's analyze how this flattening affects the L_2^2 norm of p'. We focus on the expected value of the L_2^2 norm, where the expectation is taken over the randomness of the sample set F.

$$\begin{aligned}\mathbf{E}_F\left[\|p'\|_2^2\right] &= \mathbf{E}_F\left[\sum_{i=1}^{n}\sum_{j=1}^{b_i}(p'_{(i,j)})^2\right] = \mathbf{E}_F\left[\sum_{i=1}^{n}\sum_{j=1}^{b_i}\frac{p_i^2}{b_i^2}\right] \\ &= \sum_{i=1}^{n} p_i^2 \cdot \mathbf{E}_F\left[\frac{1}{b_i}\right] = \sum_{i=1}^{n} p_i^2 \cdot \mathbf{E}_F\left[\frac{1}{f_i+1}\right].\end{aligned}$$

Recall that we defined $b_i = f_i + 1$, where f_i is the frequency of element i in the sample set F. From the discussion on Poissonization, we know that f_i is a random variable drawn from $\mathrm{Poi}(p_i\, k)$. Let's focus on the expected value of $1/(f_i+1)$:

$$\begin{aligned}\mathbf{E}_{f_i\sim\mathrm{Poi}(p_i\cdot k)}\left[\frac{1}{f_i+1}\right] &= \mathbf{E}\left[\int_0^1 x^{f_i}\mathrm{d}x\right] = \int_0^1 \mathbf{E}\left[x^{f_i}\right]\mathrm{d}x && \text{(via linearity of expectation)} \\ &= \int_0^1\left(\sum_{t=0}^{\infty} x^t \cdot \frac{e^{-p_i k}(p_i\, k)^t}{t!}\right)\mathrm{d}x && \text{(via definition of Poisson dist.)} \\ &= \int_0^1 e^{-p_i\, k + p_i\, k\, x}\left(\sum_{t=0}^{\infty}\frac{e^{-p_i\, k\, x}(p_i\, k\, x)^t}{t!}\right)\mathrm{d}x\end{aligned}$$

Note that the terms in the sum are probabilities of $Z = t$, where Z is drawn from $\mathrm{Poi}(p_i\, k\, x)$. Thus, the sum of those terms is equal to one. Therefore, we have:

$$\mathbf{E}_{f_i\sim\mathrm{Poi}(p_i\cdot k)}\left[\frac{1}{f_i+1}\right] = \int_0^1 e^{p_i\, k\,(x-1)}\mathrm{d}x = \frac{1}{p_i\, k}\cdot e^{p_i\, k\,(x-1)}\,\Big|_{x=0}^{1} \leq \frac{1}{p_i\, k}$$

Now, returning to the bound for the L_2^2 norm of p', we have:

$$\mathbf{E}_F\Big[\|p'\|_2^2\Big] = \sum_{i=1}^{n} p_i^2 \cdot \mathbf{E}_F\left[\frac{1}{f_i+1}\right] \le \frac{1}{k}\sum_{i=1}^{n} p_i = \frac{1}{k}$$

Costs of flattening: While flattening significantly reduces the L_2^2-norm, it introduces some costs. First, the process of determining the b_i's requires drawing samples from p, thus increasing the overall sample complexity. Second, flattening inflates the domain size, which can indirectly increase the sample complexity of any subsequent algorithms that operate on the flattened distribution:

$$\text{new domain size:} \qquad |D'| = \sum_{i=1}^{n} b_i = \sum_{i=1}^{n} f_i + 1 = \mathrm{Poi}(k) + n$$

Other flattening schemes: There are various methods for determining the b_i values, allowing us to choose a flattening strategy tailored to the specific problem structure. For instance, some flattening techniques are designed for testing the independence of random variables. Here, we focused on a scheme suitable for an *unknown* distribution p. If p were *known* (i.e., all probabilities p_i were available), we could set $b_i = \lfloor np_i \rfloor + 1$. As an exercise, the reader can verify that this approach reduces the L_2^2 norm of p' to $O(1/n)$.

Back to Closeness Tester

Recall from our previous lecture that there exists an algorithm that, for two distributions over $[n]$, distinguishes whether $p = q$ or they are ϵ-far with a probability of at least 0.9 using the following number of samples:

$$s = O\left(\frac{n \cdot \max\left(\|p\|_2, \|q\|_2\right)}{\epsilon^2}\right).$$

Given the flattening technique introduced in this lecture, we can efficiently test closeness between p and q. For some k (to be determined), we draw $\mathrm{Poi}(k)$ samples from p and q and use them to create flattened distributions p' and q', respectively. We then apply the closeness tester to p' and q' to determine if they are equal or ϵ-far. This process is depicted in Figure 3. As shown previously, flattening preserves the L_1 distance between distributions. Thus, $p = q$ if and only if $p' = q'$, and if p is ϵ-far from q, then p' is ϵ-far from q'.

To reduce the L_2 norm of both p and q, we combine the flattening steps. Drawing two sets of $\mathrm{Poi}(k)$ samples (one from p and one from q), we set $b_i = f_i^{(p)} + f_i^{(q)}$, where $f_i^{(p)}$ and $f_i^{(q)}$ are the frequencies of element i in the respective sample sets. As long as $b_i \ge f_i^{(p)} + 1$, the L_2^2 norm reduction is guaranteed, as shown earlier.

To ensure that the testing step has a reduced sample complexity, we must show that the

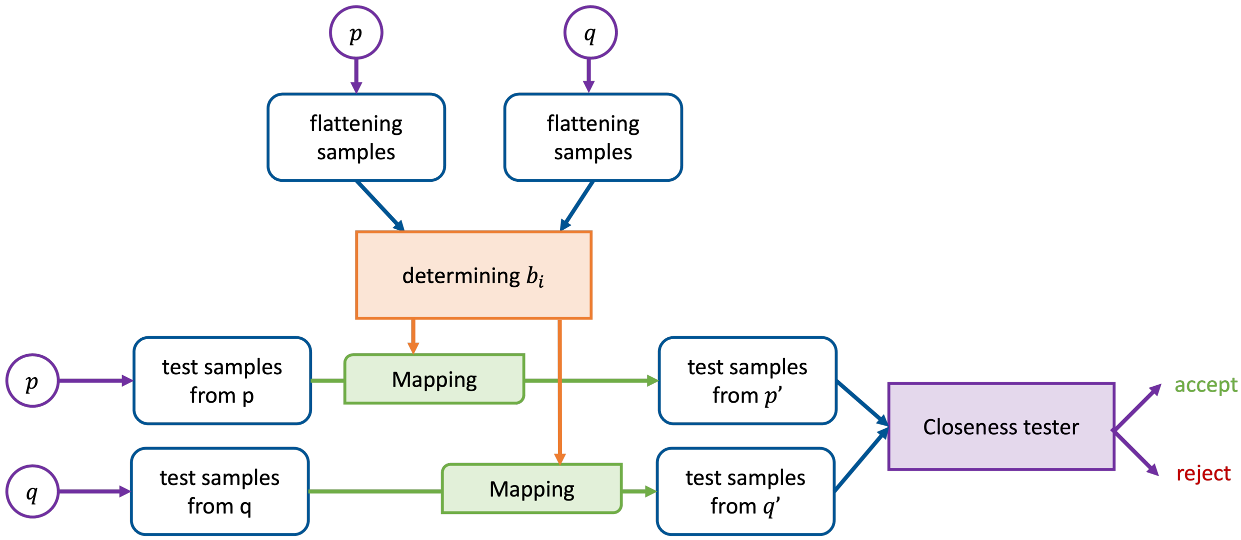

Figure 3: Diagram of the flattening and testing process

L_2 norm is reduced with high probability. Since the expected L_2 norm of p' is at most $1/k$, Markov's inequality implies that the L_2 norms of p' and q' are each at most $100/k$ with probability at least 0.99.

The optimal choice of k balances the sample complexity of the flattening and testing steps. The total sample complexity is

$$O(k+s) = O\left(k + \frac{n'}{\epsilon^2\sqrt{k}}\right) = O\left(k + \frac{n+k}{\epsilon^2\sqrt{k}}\right),$$

where $n' = O(n+k)$ is the size of the new domain.

If $k \geq n$, the sample complexity increases with k, so we need not consider $k > n$. When $k \leq n$, we have two competing terms: $O(k)$ (increasing) and $O\left(\frac{n}{\sqrt{k}\epsilon^2}\right)$ (decreasing). Minimizing their sum yields $k = n^{2/3}/\epsilon^{4/3}$. Since $k \leq n$, the optimal choice is $k = \min(n, n^{2/3}/\epsilon^{4/3})$. Substituting this value into the total sample complexity gives a final sample complexity of

$$O\left(\frac{n^{2/3}}{\epsilon^{4/3}} + \frac{\sqrt{n}}{\epsilon^2}\right).$$

This sample complexity is known to be optimal for this problem.

Bibliographic Note: The content of this lecture is based on [DK16]. Further applications of the flattening technique can be found in that work.

References

[DK16] Ilias Diakonikolas and Daniel M. Kane. A new approach for testing properties of discrete distributions. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 685–694, 2016.