Total Variation is well-adapted to discrete probability measures all on the same discrete space. E.g. Bernoulli, Poisson...

Theorem: (The Law of Rare Events)

Let $\{X_j\}_{j=1}^n$ be independent, $X_j \overset{d}{=} \text{Bernoulli}(p_j)$.

Set $S_n = X_1 + \cdots + X_n$. Let $N \overset{d}{=} \text{Poisson}(p_1 + \cdots + p_n)$.

Then

$$d_{TV}(S_n, N) \leq \sum_{j=1}^n p_j^2 .$$

Cor: If $\lambda > 0$ and $n > 1/\lambda$

then $\text{Binom}(n, p = \lambda/n) \overset{d}{\approx} \text{Poisson}(\lambda)$

Lemma: Let $\{\mu_j, \nu_j\}_{j=1}^{n}$ be probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then

$$d_{TV}(\mu_1 * \mu_2 * \cdots * \mu_n, \nu_1 * \nu_2 * \cdots * \nu_n) \leq \sum_{j=1}^{n} d_{TV}(\mu_j, \nu_j)$$

Pf. [HW]

Pf. #1 of Law of Rare Events:

Let $N_j \overset{d}{=} \mathrm{Poiss}(p_j)$ be independent Poisson rv's.

Coupling

Given probability measures μ, ν on $(S, \mathcal{B})$, a **coupling** is a pair (X, Y) of random variables on a common probability space, taking values in $(S, \mathcal{B})$, such that $\mu_X = \mu$, $\mu_Y = \nu$.

In other words, a coupling is a probability measure on $(S \times S, \mathcal{B} \otimes \mathcal{B})$ whose marginals are μ and ν.

E.g. $\mu \otimes \nu$ is a (silly) coupling of μ, ν.

Lemma: (Coupling Estimate)
If (X, Y) is any coupling of the Borel probability measures μ, ν, then $d_{TV}(\mu, \nu) \leq \mathbb{P}(X \neq Y)$.

Pf.:

Pf. #2 of Law of Rare Events:

$S_n = X_1 + \cdots + X_n$, $X_j \overset{d}{=} \text{Bernoulli}(p_j)$

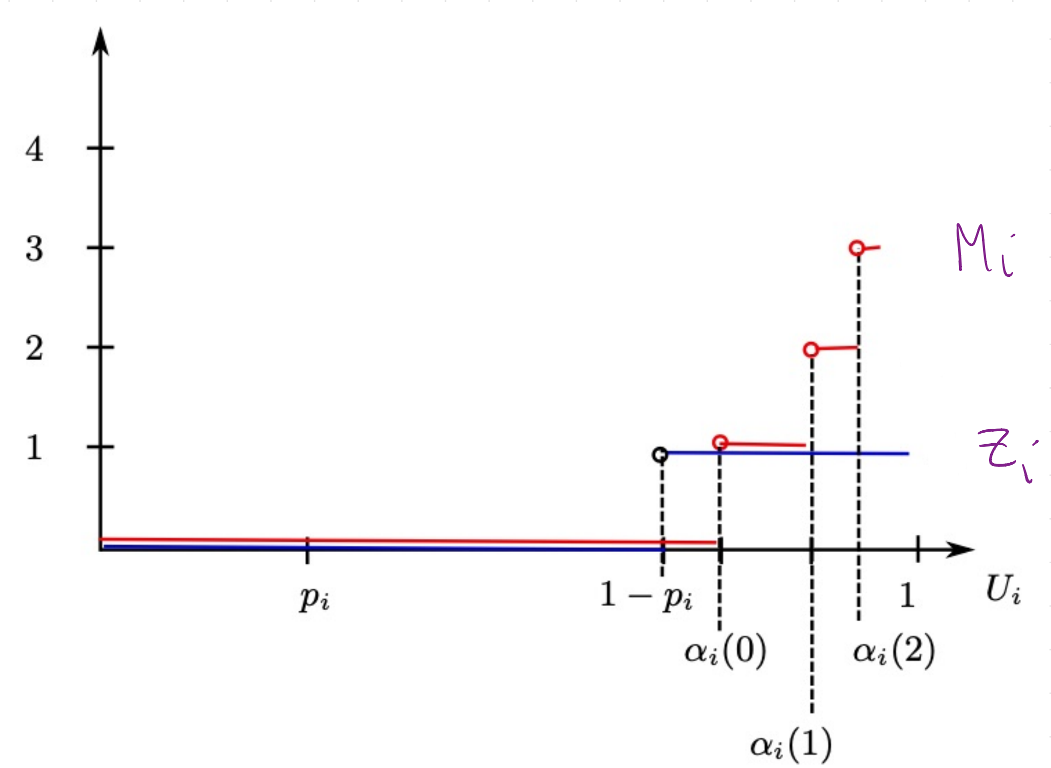
$N \overset{d}{=} \text{Poisson}(p_1 + \cdots + p_n)$

We construct a coupling of μ_{S_n}, μ_N.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with $\{U_j\}_{j=1}^n$ iid $\text{Unif}[0,1]$ rv's.

By the coupling estimate:

$$d_{TV}(\mu_{S_n}, \mu_N) \leq \mathbb{P}(T \neq M)$$



Couplings are important in many parts of probability.
They lead, for example, to other natural metrics on probability measures, adapted to problems in large deviations, optimal transport, stochastic analysis, etc.

Def: For $p \geq 1$, $\mathrm{Prob}^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is the set of Borel probability measures μ on $\mathbb{R}^d$ with finite p^{th} moment: $\int_{\mathbb{R}^d} |x|^p \mu(dx) < \infty$.

The p-Wasserstein distance between two such measures μ, ν is:

$$W_p(\mu,\nu) := \left(\inf\{ E[|X-Y|^p] : (X,Y) \text{ is a coupling of } \mu,\nu\}\right)^{1/p}$$

Generally weaker than d_{TV}.