Quantization of Discrete Probability Distributions

Yuriy A. Reznik

yreznik@ieee.org

Qualcomm Incorporated

5775 Morehouse Drive

San Diego, CA 92121, USA

Outline

1. Description of the problem
   - Where it appears?
   - Why it is relevant?
2. Connection to the Covering Radius problem
   - High-rate regime asymptotic results
3. Proposed algorithm for quantization of distributions
   - Description of the algorithm
   - Performance analysis
4. Discussion and Conclusions

The Problem

Consider:

- $A = \{\alpha_1, \ldots, \alpha_m\}$, $m < \infty$ – a finite set of events;
- Ω_m – set of probability distributions over A:

$$\Omega_m = \left\{ [\omega_1, \ldots, \omega_m] \in \mathbb{R}^m \,\middle|\, \forall i : \omega_i \geqslant 0\,,\ \textstyle\sum_i \omega_i = 1 \right\}$$

(*unit* $(m-1)$*-simplex*)

We receive:

- $p \in \Omega_m$ – input distribution;

and our task is to *encode* p *with some given fidelity criterion*.

Applications

Universal source coding

- 1960's: Lynch-Davisson codes (lossless coding of types)
- 1970's: “Rice machine” (coding of variance)
- 1980's: Rissanen’s two-part universal codes (parametric models)

  code = <quantized distribution> <encoded sample>

  ... but coding of distributions is never handled on its own!

Image recognition (SIFT/SURF/CHoG algorithms – $2004+$)

- work with “histograms of gradients” in images
- task is to quantize histograms to simplify search and retrieval

Quantization (conventional setting)

Consider:

- $d\,(p, q)$ – distance between $p, q \in \Omega_m$; p – input, q– reconstruction
- $Q \subset \Omega_m$ – a set of reconstruction points;

Fixed-rate case:

- $R(Q) = \log_2 |Q| = \mathrm{const.}$

If we further know that $p \sim \theta$, where θ is some density over Ω_m, then the problem becomes:

$$\bar{d}(\Omega_m, \theta, R) = \inf_{\substack{Q \subset \Omega_m \\ |Q| \leqslant 2^R}} \mathbf{E}_{\substack{p \in \Omega_m \\ p \sim \theta}} \min_{q \in Q} d(p, q)\,,$$

I.e., the task is to minimize the *expected distance* to the reconstruction point.

Quantization (cont'd)

Conventional setting (θ is a density over Ω_m):

$$\bar{d}(\Omega_m, \theta, R) = \inf_{\substack{Q \subset \Omega_m \\ |Q| \leqslant 2^R}} \mathbf{E}_{\substack{p \in \Omega_m \\ p \sim \theta}} \min_{q \in Q} d(p, q) \, ,$$

However, in practice, we usually:

- have no information about θ; and/or
- need to transmit/use quantized distribution instantaneously!!!
  - in two-part universal code quantized distribution is used right away to encode a block;
  - in image recognition histograms of a query image are created/used once.

Hence, finding minimal *expected* distance $\bar{d}(\Omega_m, \theta, R)$ is not exactly what we need!

Quantization (Minimax setting)

Let's minimize *worst-case distance*:

$$d^*(\Omega_m, R) = \inf_{\substack{Q \subset \Omega_m \\ |Q| \leqslant 2^R}} \max_{p \in \Omega_m} \min_{q \in Q} d(p, q) \, .$$

The problem is now purely geometric!

- it is equivalent to a problem of *covering* of the space Ω_m with at most 2^R balls of the same radius.

Dual problem can also be formulated:

$$R(\varepsilon) = \inf_{Q \subset \Omega_m : \; \max_{p \in \Omega_m} \min_{q \in Q} d(p,q) \leqslant \varepsilon} \log |Q| \, ,$$

Also a special case of a known problem:

- $R(\varepsilon)$ is the Kolmogorov's ε*-entropy* for metric space (Ω_m, d).

Known Results for Covering Radius Problem

Let $A \subset \mathbb{R}^k$ – compact, with positive Jordan measure $\lambda^k(A) > 0$.

Theorem 1 (S.Graf & H.Luschgy, 2000)**.** *With* $R \to \infty$*:*

$$d^*_\alpha(A, R) \sim C_{k,\alpha} \sqrt[k]{\lambda^k(A)}\; 2^{-R/k}$$

where:

$$C_{k,\alpha} = \inf_{R>0} 2^{R/k}\, d^*_\alpha([0,1]^k, R)$$

is a constant (covering coefficient for the unit cube).

The exact value of $C_{k,\alpha}$ depends on the distance

$$d_\alpha(p, q) = ||p - q||_\alpha = \Big(\sum_i |p_i - q_i|^\alpha\Big)^{1/\alpha}, \quad \alpha \geqslant 1\,.$$

For example: $C_{k,\infty} = \frac{1}{2}$ (for any k), $C_{2,1} = \frac{1}{\sqrt{2}}$, $C_{2,2} = \sqrt{\frac{2}{3\sqrt{3}}}$, etc.

Achievable Covering Radius for Probability Distributions

By replacing A with simplex Ω_m, and noticing that:

$$\mathrm{Vol}(\Omega_m) = \left. \frac{a^k}{k!}\sqrt{\frac{k+1}{2^k}} \,\right|_{\substack{k=m-1\\ a=\sqrt{2}}} = \frac{\sqrt{m}}{(m-1)!}\,,$$

we arrive at the following statement.

Corollary 1. *With $R \to \infty$:*

$$d^*_\alpha(\Omega_m, R) \sim C_{m-1,\alpha} \sqrt[m-1]{\frac{\sqrt{m}}{(m-1)!}}\, 2^{-\frac{R}{m-1}}\,,$$

where $C_{m-1,\alpha}$ are some known constants.

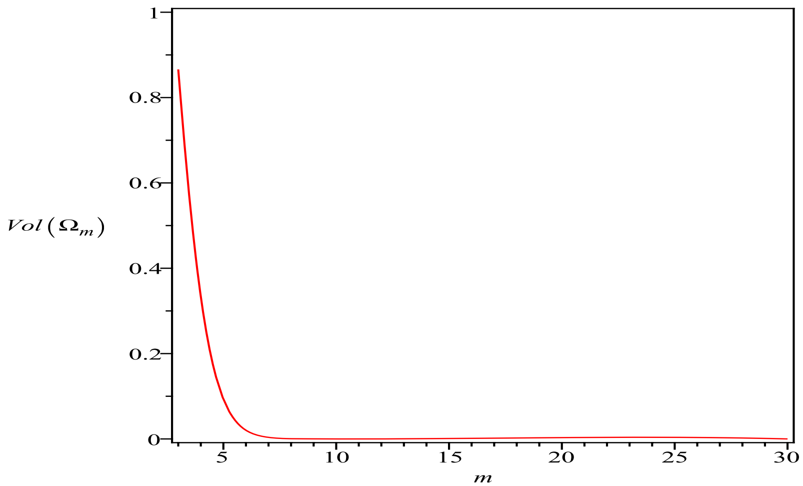
Achievable Covering Radius for Probability Distributions

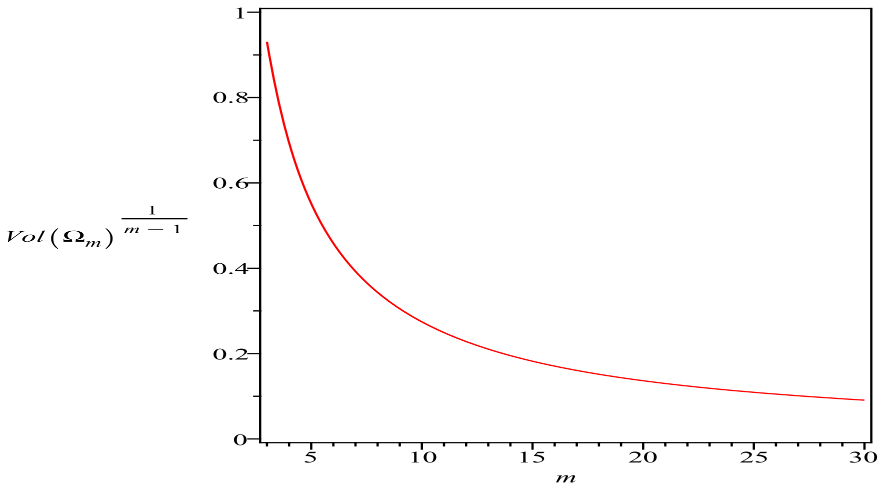
So what's special about our problem?

$$d^*_\alpha(\Omega_m, R) \sim C_{m-1,\alpha} \sqrt[m-1]{\mathrm{Vol}(\Omega_m)}\, 2^{-\frac{R}{m-1}}.$$

Leading term decays as the number of dimensions m increases:

$$\sqrt[m-1]{\mathrm{Vol}(\Omega_m)} = \sqrt[m-1]{\frac{\sqrt{m}}{(m-1)!}} = \frac{e}{m} + O\left(\frac{1}{m^2}\right)$$





Quantization of Distributions

Design of a practical algorithm:

- Choice of lattice
- Algorithm for finding nearest reconstruction point
- Enumeration of lattice points
- Encoding

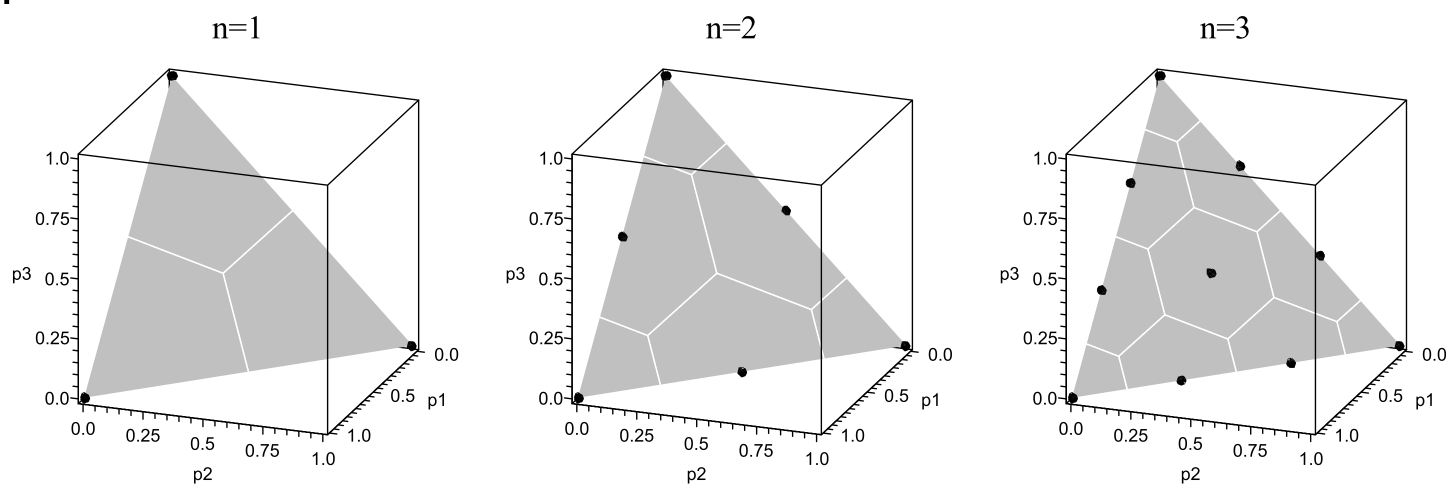
Type Lattice

Given some integer $n \geqslant 1$, we define a lattice $Q_n \subset \Omega_m$:

$$Q_n = \left\{ [q_1, \ldots, q_m] \in \mathbb{Q}^m \;\middle|\; \forall i : q_i = \frac{k_i}{n}\,, \quad k_i, n \in \mathbb{Z}^+\,; \quad \sum_i k_i = n \right\}\,.$$

Lattice points $q \in Q_n$ coincide with *memoryless types*!

Examples in $m = 3$ dimensions:



NB: in this example Q_n is equivalent to a hexagonal lattice. With $m > 3$ it is equivalent to a bounded subset of *lattice* A_n (cf. SPLAG, Chapter 4).

Quantization Algorithm

Algorithm 1. Given $p \in \Omega_m$ and n find nearest type $\left\{\frac{k_1}{n}, \ldots, \frac{k_m}{n}\right\}$:

1. Compute numbers (best unconstrained approximation):

$$k'_i = \left\lfloor np_i + \tfrac{1}{2} \right\rfloor, \quad n' = \textstyle\sum_i k'_i \,.$$

2. If $n' = n$ we are done. Otherwise, compute $\delta_i = k'_i - np_i$, and sort them:

$$-\tfrac{1}{2} < \delta_{j_1} \leqslant \delta_{j_2} \leqslant \ldots \leqslant \delta_{j_m} \leqslant \tfrac{1}{2} \,,$$

3. Let $\Delta = n' - n$. If $\Delta > 0$ then we decrement d values k'_i with largest errors

$$k_{j_i} = \begin{bmatrix} k'_{j_i}, & i=1,\ldots,m-\Delta-1 \,, \\ k'_{j_i}-1, & i=m-\Delta,\ldots,m \,, \end{bmatrix}$$

otherwise, we increment $|\Delta|$ values k'_i with smallest errors:

$$k_{j_i} = \begin{bmatrix} k'_{j_i}+1, & i=1,\ldots,|\Delta| \,, \\ k'_{j_i}, & i=|\Delta|+1,\ldots,m \,. \end{bmatrix}$$

Enumeration of Types

The number of points in Q_n is essentially the number of integers $k_1, \ldots, k_m$ with total n, which is:

$$|Q_n| = \binom{n+m-1}{m-1}.$$

Indices of types with frequencies $k_1, \ldots, k_m$ can be computed by:

$$\xi(k_1, \ldots, k_n) = \sum_{j=1}^{n-2} \sum_{i=0}^{k_j - 1} \binom{n - i - \sum_{l=1}^{j-1} k_l + m - j - 1}{m - j - 1} + k_{n-1}.$$

This formula follows by induction (starting with $m = 2, 3$, etc.), and performs lexicographic enumeration of types. For example:

$$\begin{aligned}
\xi(0, 0, \ldots, 0, n) &= 0\,, \\
\xi(0, 0, \ldots, 1, n-1) &= 1\,, \\
&\ldots \\
\xi(n, 0, \ldots, 0, 0) &= \tbinom{n+m-1}{m-1} - 1\,.
\end{aligned}$$

Encoding

We simply compute type indices $\xi(k_1, \ldots, k_n)$, and transmit them by using fixed-rate codes.

The rate of such code satisfies (for large n):

$$R(n) = \lceil \log_2 |Q_n| \rceil = (m-1)\log_2 n - \log_2 (m-1)! + O\left(\frac{1}{n}\right) .$$

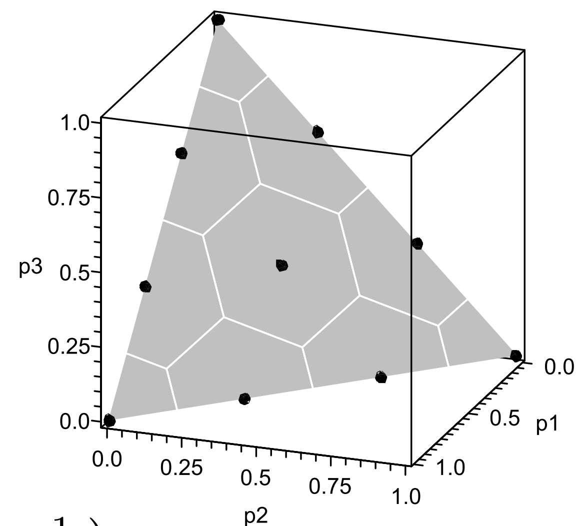
The entire algorithm is remarkably simple:

- $O(m)$ steps to compute nearest type
- $O(n)$ steps to compute lexicographic index
- $O(1)$ steps to create and transmit the code

Analysis: Properties of Voronoi Cells

Vertices of Voronoi cells (or *holes*) in type lattice are located at

$$q_i^* = q + v_i, \quad q \in Q_n, \quad i = 1, \ldots, m-1,$$



where

$$v_i = \frac{1}{n}\left[\underbrace{\frac{m-i}{m}, \ldots, \frac{m-i}{m}}_{i \text{ times}}, \underbrace{\frac{-i}{m}, \ldots, \frac{-i}{m}}_{m-i \text{ times}}\right].$$

This implies that (with $a = \lfloor m/2 \rfloor$):

$$\max_{p \in \Omega_m} \min_{q \in Q_n} d_\infty(p, q) = \frac{1}{n}\left(1 - \frac{1}{m}\right),$$

$$\max_{p \in \Omega_m} \min_{q \in Q_n} d_2(p, q) = \frac{1}{n}\sqrt{\frac{a(m-a)}{m}},$$

$$\max_{p \in \Omega_m} \min_{q \in Q_n} d_1(p, q) = \frac{1}{n}\frac{2a(m-a)}{m}.$$

Analysis: Performance of Type Quantization

Theorem 2. *The following holds (with large R):*

$$\min_{n:|Q_n|\leqslant 2^R} \max_{p\in\Omega_m} \min_{q\in Q_n} d_\infty(p,q) \sim \left(1-\tfrac{1}{m}\right) \frac{1}{\sqrt[m-1]{(m-1)!}}\, 2^{-\frac{R}{m-1}},$$

$$\min_{n:|Q_n|\leqslant 2^R} \max_{p\in\Omega_m} \min_{q\in Q_n} d_2(p,q) \sim \sqrt{\frac{a(m-a)}{m}} \frac{1}{\sqrt[m-1]{(m-1)!}}\, 2^{-\frac{R}{m-1}},$$

$$\min_{n:|Q_n|\leqslant 2^R} \max_{p\in\Omega_m} \min_{q\in Q_n} d_1(p,q) \sim \frac{2a(m-a)}{m} \frac{1}{\sqrt[m-1]{(m-1)!}}\, 2^{-\frac{R}{m-1}}.$$

In all cases the decay rate $2^{-\frac{R}{m-1}}$ is optimal. Furthermore, the factor

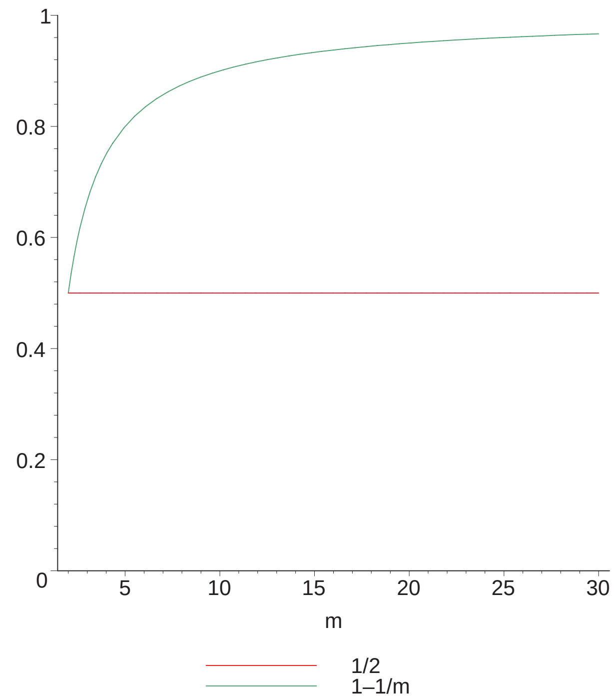
$$\frac{1}{\sqrt[m-1]{(m-1)!}} = \frac{e}{m} + O\left(\frac{1}{m^2}\right),$$

matches the decay rate w.r.t. m predicted for probability quantization problem. The only differences are in *leading factors.*

Analysis: Leading Factors

L_∞ - distance case:

- optimal: $C_{m-1,\infty} = \frac{1}{2}$
- type quantizer: $1 - \frac{1}{m}$



NB: Maximum L_∞-error of type quantizer is within a factor of 2 from minimum possible.

Type Quantization: Summary

Have shown that:

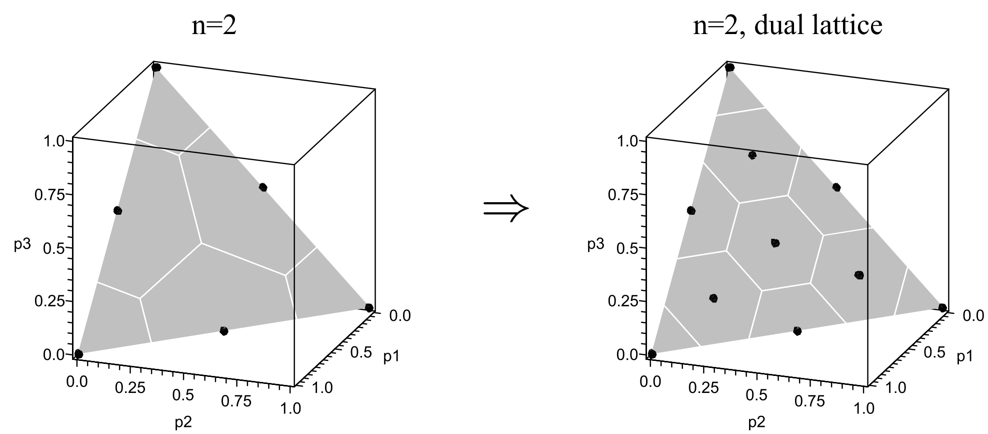
- There exists a remarkably simple algorithm for quantization of probability distributions
- It uses types with fixed total as quantization lattice.
- It is asymptotically optimal in high-rate regime
  - the only difference is in the leading factor. E.g. for L_∞-norm it is shown to be within a factor of 2 from minimum possible.

Beyond Types

Dual type lattice:

$$Q_n^* = \cup_{i=0}^{m-1} (Q_n + v_i) \; ; \quad v_i = \frac{1}{n} \Bigg[\underbrace{\frac{m-i}{m}, \ldots, \frac{m-i}{m}}_{i \text{ times}}, \underbrace{\frac{-i}{m}, \ldots, \frac{-i}{m}}_{m-i \text{ times}} \Bigg] .$$

I.e. we simply put additional points in holes of Q_n.



Dual type lattice achieves (asymptotically with $m \to \infty$):

- factor of 2 reduction in L_1 and L_∞ radii, and
- factor of $\sqrt{3}$ reduction in L_2 radius.

Conclusions & Open Problem

- We have shown that type-lattice can be used for quantization of distributions
  - very simple algorithm was developed for that purpose
- But, we also noted that thinner lattices exist!!!
  - Dual type lattice Q_n^*
  - E_8, Λ_{24}, and other lattices in “lucky dimensions”
- This brings a question:
  - Is there a better way to sample data and map them to probability estimates?
  - Better than types in covering-radius sense?