Quantization

Robert M. Haralick

Computer Science, Graduate Center
City University of New York
Outline

1. Quantizing
2. Number of Quantizing Levels
3. Quantizing Boundaries
4. Bin Probability
5. Optimization
6. Laplace Probability
Data is real-valued
Data is integer valued with large max value
To use a discrete Bayes rule the data has to be quantized
  Quantize each dimension to 10 or less quantized intervals
The Problem

- Assume input data in each dimension is discretely valued:
  - 0 – 255
  - 0 – 1023
- And now it must be quantized
- Determine the Number of Quantizing levels for each dimension
- Determine the Quantizing interval boundaries
- Determine the Probability associated with each quantizing bin
Quantizer and Bins

- $N$ dimensions
- $L$ quantized values per dimension
- $L^N$ bins in discrete measurement space
- Each bin has a class conditional probability
The Quantizer

Definition

A quantizer \( q \) is a monotonically increasing function that takes in a real number and produces a non-negative integer between 0 and \( K - 1 \) where \( K \) is the number of quantizing levels.
The quantizing interval $Q_k$ associated with the integer $k$ is defined by

$$Q_k = \{ x \mid q(x) = k \}$$
Let $Z = (z_1, \ldots, z_J)$ be a measurement tuple

Let $q_j$ be the quantizing function for the $j^{th}$ dimension

The quantized tuple for $Z$ is $(q_1(z_1), \ldots, q_J(z_J))$

The address for $Z$ is $a(q_1(z_1), \ldots, q_J(z_J))$
Definition

If $p_1, \ldots, p_K$ is a discrete probability function, its entropy $H$ is

$$H = - \sum_{k=1}^{K} p_k \log_2 p_k$$
Person A chooses an index from \{1, \ldots, K\} in accordance with probabilities \(p_1, \ldots, p_K\).

Person B is to determine the index chosen by Person A by guessing.

Person B can ask any question that can be answered Yes or No.

Assuming that Person B is clever in formulating the questions, it will take on the average \(H\) questions to correctly determine the index Person A chose.
Entropy

\[ H = - \sum_{k=1}^{K} p_k \log_2 p_k \]

If a message is sent composed of index choices sampled from a distribution with probabilities \( p_1, \ldots, p_K \), the average information in the message is \( H \) bits per symbol.
Entropy Estimation

Data is discrete
Observe $x_1, \ldots, x_N$, where each $x_n \in \{1, \ldots, K\}$
Count the number of occurrences

$$m_k = \#\{n \mid x_n = k\}$$

Estimate the probability

$$p_k = \frac{m_k}{N}$$

Number of zero counts $n_0 = \#\{k \mid m_k = 0\}$
Unbiased estimate of entropy

$$\hat{H} = - \sum_{k=1}^{K} p_k \log_2 p_k + \frac{n_0 - 1}{2N \log_e 2}$$
Each observation is a tuple $Z = (z_1, \ldots, z_J)$
Each $x_j$ is discretely valued
Let $M$ be the total number of quantizing bins over $J$ dimensions
How to determine the number $L_j$ of bins for the $j^{th}$ dimension?
$M = \prod_{j=1}^{J} L_j$
The Entropy Solution

Let $\hat{H}_j$ be the entropy of the $j^{th}$ component of $Z$. Let $L_j$ be the number of bins for the $j^{th}$ component of $Z$. Define

$$f_j = \frac{\hat{H}_j}{\sum_{i=1}^{J} \hat{H}_i}$$

$$L_j = M^{f_j}$$

$$\prod_{j=1}^{J} L_j = \prod_{j=1}^{J} M^{f_j}$$

$$= M^{\sum_{j=1}^{J} f_j}$$

$$= M$$
The Number of Quantizing Bins

Let $\hat{H}_j$ be the entropy of the $j^{th}$ component of $Z$.
Let $L_j$ be the number of bins for the $j^{th}$ component of $Z$.
Define

$$f_j = \frac{\hat{H}_j}{\sum_{i=1}^{J} \hat{H}_i}$$

$$L_j = \lceil M^{f_j} \rceil$$
Carrying out the Entropy Solution requires probability estimates.

For each variable, the A/D converter of the sensor might have say 10 bits.

There are then 1024 discrete possible values.

Assume that the total number of observations is more than 10 times 1024.

The probability for each value is estimated simply as the fraction of observations that have that value.
The sample is $Z_1, \ldots, Z_N$

Each tuple has $J$ components

The $n^{th}$ observed tuple: $Z_n = (z_{n1}, \ldots, z_{nJ})$

Let $z_{(1)j}, \ldots, z_{(N)j}$ be the $N$ values of the $j^{th}$ component of the observed tuples, ordered in ascending order.

The left quantizing interval boundaries are:
The sample is $Z_1, \ldots, Z_N$

Each tuple has $J$ components

The $n^{th}$ observed tuple: $Z_n = (z_{n1}, \ldots, z_{nJ})$

Let $z_{(1)j}, \ldots, z_{(N)j}$ be the $N$ values of the $j^{th}$ component of the observed tuples, ordered in ascending order.

The left quantizing interval boundaries are:
Example

Suppose $N = 12$ and $K = 4$.

- $j^{th}$ component $z_{1j}, \ldots, z_{12j}$
- ordered values of $j^{th}$ component
  - $z_{(1)j}, \ldots z_{(12)j}$

The quantizing intervals are:

$$[-\infty, z_{(4)j})$$
$$[z_{(4)j}, z_{(7)j})$$
$$[z_{(7)j}, z_{(10)j})$$
$$[z_{(10)j}, \infty)$$
The sample is $Z_1, \ldots, Z_N$

The $n^{th}$ observed tuple: $Z_n = (z_{n1}, \ldots, z_{nJ})$

Let $z_{(1)j}, \ldots, z_{(N)j}$ be the $N$ values of the $j^{th}$ component of the observed tuples, ordered in ascending order.

$k$ indexes quantizing interval: $k = 1, \ldots, K$

The $k^{th}$ quantizing interval $[c_{jk}, d_{jk})$ for the $j^{th}$ component is defined by

$$c_{jk} = Z((k-1)N/K + 1)j$$

$$d_{jk} = Z(kN/K + 1)j$$
Non-uniform Quantization
The sample is $Z_1, \ldots, Z_N$

The $n^{th}$ observed tuple: $Z_n = (z_{n1}, \ldots, z_{nJ})$

The quantized tuple for $Z_n$ is $(q_1(z_{n1}), \ldots, q_J(z_{nJ}))$

The address for $Z_n$ is $a(q_1(z_{n1}), \ldots, q_J(z_{nJ}))$

The bins are numbered $1, \ldots, M$

The number of observations falling into bin $m$ is $t_m$

The maximum likelihood estimate of the probability for bin $m$ is $p_m$

\[
t_m = \# \{ n \mid a(q_1(z_{n1}), \ldots, q_J(z_{nJ})) = m \} \]

\[
p_m = \frac{t_m}{N} \]
Density Estimation Using Fixed Volumes

- Total count $N$
- Fix a volume $v$
- Count the number $k$ of observations in the volume $v$
- Density is mass divided by volume
- Estimate the density in the volume as $\frac{k}{Nv}$
Density Estimation Using Fixed Counts

- Total count $N$
- Fix a count $k^*$
- Find the smallest volume $v$ around the point having a count $k$ just greater than $k^*$
- Density is mass divided by volume
- Estimate the density in the volume as $\frac{k}{Nv}$
Smoothed Estimates

- If the sample size is not large enough, the MLE probability estimates may have too large a variance.

bin smoothing

- Let bin \( m \) have volume \( v_m \) and count \( t_m \)
- Let \( m_1, \ldots, m_l \) be the indexes of the \( l \) closest bins to bin \( m \) satisfying

\[
\sum_{i=1}^{l} t_{m_i} \geq k \quad \sum_{i=1}^{l-1} t_{m_i} < k
\]

- \( b_m = \sum_{i=1}^{l} t_{m_i} \)
- \( V_m^* = \sum_{i=1}^{l} v_{m_i} \)
- Density in bin \( m \): \( \alpha b_m / V_m^* \)
- Set \( \alpha \) so that the density integrates to 1
Smoothed Estimates

- Density in bin $m$: $\alpha b_m / V_m^*$
- Volume of bin $m$: $v_m$
- Probability of bin $m$: $p_m = (\alpha b_m / V_m^*) v_m$
- Total probability: $1 = \sum_{m=1}^{M} \alpha b_m v_m / V_m^*$

$$\alpha = \frac{1}{\sum_{m=1}^{M} b_m v_m / V_m^*}$$

$$p_m = \frac{1}{\sum_{k=1}^{M} b_k v_k / V_k^*} b_m v_m / V_m^*$$
Optimization

- Fixed Sample Size $N$
- Sample $Z_1, \ldots, Z_N$
- Total number of bins $M$
- Bin smoothing: $k$
- Calculate the quantizer and probability associations
- Calculate a decision rule maximizing expected gain
- Everything depends on $M$ and $k$
Memorization and Generalization

- $k$ is too small: memorization, over-fitting
- $k$ is too large: over-generalization, under-fitting
- $M$ is too large: memorization, over-fitting
- $M$ is too small: over-generalization, under-fitting
Optimize The Probability Estimation Parameters

- Split the ground truthed sample into three parts
- Use the first part to calculate the quantizer and probabilities
- Calculate the decision rule
- Apply the decision rule to the second part so that an unbiased estimate of the expected economic gain given the decision rule can be computed
- Brute force optimization to find the values of $M$ and $k$ to maximize the estimated expected gain
- With $M$ and $k$ fixed, use the third part to determine an estimate the expected gain for the optimization
Once the parameters $M$ and $k$ have been optimized, the quantizer boundaries can be optimized.
Optimize The Quantizer Boundaries

Repeat until no change

- Randomly choose a component $j$ and quantizing interval $k$
- Randomly choose a small perturbation $\delta$ ($\delta$ can be positive or negative)
- Randomly choose a small integer $M$ (No collision with neighboring boundaries)
- $b_{kj}^{\text{new}} = b_{kj} - \delta(M + 1)$
- For $(m = 0; m \leq 2M + 1; m++)$
  - $b_{kj}^{\text{new}} \leftarrow b_{kj}^{\text{new}} + \delta$
  - Compute New Probabilities
  - Recompute Bayes Rule
  - Save expected gain
- Replace $b_{kj}$ by the value associated with the highest gain
Optimize The Quantizer Boundaries

Greedy Algorithm has a random component
  - Multiple runs will produce different answers

Repeat greedy algorithm T times
Keep track of best result so far
After T times, use the best result
Bayesian Perspective

- MLE: start bin counters from 0
- Bayesian: start bin counters from $\beta, \beta > 0$
The Observation

There are $K$ bins. Each time an observation is made, the observation falls into exactly one of the $K$ bins. The probability that an observation falls into bin $k$ is $p_k$. To estimate the bin probabilities $p_1, \ldots, p_K$, we take a random sample of $I$ observations. We find that of the $I$ observations,

- $I_1$ observations fall into bin 1
- $I_2$ observations fall into bin 2
- \ldots
- $I_K$ observations fall into bin $K$
Under the protocol of the random sampling, the probability of observing counts $I_1, \ldots, I_K$ given the bin probabilities $p_1, \ldots, p_K$ is given by the multinomial

$$P(I_1, \ldots, I_K \mid p_1, \ldots, p_K) = \frac{l!}{I_1! \cdots I_K!} p_1^{I_1} \cdots p_K^{I_K}$$
We have observed $l_1, \ldots, l_K$ we would like to determine the probability that an observation falls in bin $k$.

- Denote by $d_k$ the event that an observation falls into bin $k$.
- We wish to determine $P(d_k \mid l_1, \ldots, l_K)$
To do this we will need to evaluate two integrals over the $K - 1$-simplex

$$S = \{ (q_1, \ldots, q_K) \mid 0 \leq q_k \leq 1, k = 1, \ldots, K \text{ and } q_1 + q_2 + \ldots + q_K = 1 \}$$

- 0 Simplex: point
- 1 Simplex: line segment
- 2 Simplex: triangle
- 3 Simplex: Tetrahedron
- 4 Simplex: Pentachoron
$K - 1$ Simplex

**Definition**

A $K - 1$ Simplex is a $(K - 1)$-dimensional polytope which is the convex hull of its $K$ vertices.

$$ S = \{ (q_1, \ldots, q_K) \mid 0 \leq q_k \leq 1, k = 1, \ldots, K \text{ and } q_1 + q_2 + \ldots + q_K = 1 \} $$

The $K$ vertices of $S$ are the $K -$ tuples $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0)$, $(0, 0, 1, 0, \ldots, 0)$, $\ldots$, $(0, \ldots, 0, 1)$.
The $K-1$ Simplex

$$S = \{ (q_1, \ldots, q_K) \mid 0 \leq q_k \leq 1, k = 1, \ldots, K \text{ and } q_1 + q_2 + \ldots + q_K = 1 \}$$
Two Integrals

They are:

\[
\int_{(q_1, \ldots, q_K) \in S} dq_1, \ldots, dq_K = \frac{1}{(K - 1)!}
\]

\[
\int_{(q_1, \ldots, q_K) \in S} \prod_{k=1}^{K} q_k^{l_k} dq_1, \ldots, dq_K = \frac{\prod_{k=1}^{K} l_k!}{(l + K - 1)!}
\]

where

\[
\sum_{k=1}^{K} l_k = l
\]
Gamma Function

\[ \Gamma(z) = \int_0^\infty t^{z-1} e^{-t} \, dt \]

\[ \Gamma(n) = (n - 1)! \]
The derivation goes as follows: 

\[ \text{Prob}(d_k \mid l_1, \ldots, l_K) \]

\[ = \frac{\text{Prob}(d_k, l_1, \ldots, l_K)}{\text{Prob}(l_1, \ldots, l_K)} \]

\[ = \frac{\int_{(p_1, \ldots, p_K) \in S} \text{Prob}(d_k, l_1, \ldots, l_K, p_1, \ldots, p_K) \, dp_1 \cdots dp_K}{\int_{(q_1, \ldots, q_K) \in S} \text{Prob}(l_1, \ldots, l_K, q_1, \ldots, q_K) \, dq_1 \cdots dq_K} \]

\[ = \frac{\int_{(p_1, \ldots, p_K) \in S} \text{Prob}(d_k, l_1, \ldots, l_K \mid p_1, \ldots, p_K) \, P(p_1, \ldots, p_K) \, dp_1 \cdots dp_K}{\int_{(q_1, \ldots, q_K) \in S} \text{Prob}(l_1, \ldots, l_K \mid q_1, \ldots, q_K) \, P(q_1, \ldots, q_K) \, dq_1 \cdots dq_K} \]

\[ = \frac{\int_{(p_1, \ldots, p_K) \in S} \frac{\prod_{n=1}^{K} l_n!}{l!} \prod_{m=1}^{K} p_m^l \, p_k (K - 1)! \, dp_1 \cdots dp_K}{\int_{(q_1, \ldots, q_K) \in S} \frac{\prod_{n=1}^{K} l_n!}{l!} \prod_{m=1}^{K} q_m^l \, (K - 1)! \, dq_1 \cdots dq_K} \]
\[ \text{Prob}(d_k \mid l_1, \ldots, l_K) \]

\[ = \frac{\int_{p_1, \ldots, p_K} p_1^{l_1} p_2^{l_2} \cdots p_{k-1}^{l_{k-1}} p_k^{l_k+1} p_{k+1}^{l_{k+1}} \cdots p_K^{l_K} \, dp_1 \cdots dp_K}{\int_{q_1, \ldots, q_K} q_1^{l_1} q_2^{l_2} \cdots q_K^{l_K} \, dq_1 \cdots dq_K} \]

\[ = \frac{l_1! l_2! \cdots l_{k-1}!(l_{k+1}+1)! l_{k+1}! \cdots l_K}{(I+K)!} \]

\[ = \frac{l_1! l_2! \cdots l_K!}{(I+K-1)!} \]

\[ = \frac{l_k + 1}{I + K} \]
Prior Distribution

The prior distribution over the $K$-Simplex does not have to be taken as uniform. The natural prior distribution over the $K$-Simplex is the Dirichlet distribution.

$$P(p_1, \ldots, p_K | \alpha_1, \ldots, \alpha_K) = \frac{\Gamma \left( \sum_{k=1}^{K} \alpha_k \right)}{\prod_{k=1}^{K} \Gamma (\alpha_k)} \prod_{k=1}^{K} p_k^{\alpha_k-1}$$

$$\alpha_k > 0$$

$$0 < p_k < 1, \ k = 1, \ldots, K$$

$$\sum_{k=1}^{K-1} p_k < 1$$

$$p_K = 1 - \sum_{k=1}^{K-1} p_k$$
Dirichlet Distribution Properties

\[
E[p_k] = \frac{\alpha_k}{\sum_{j=1}^{K} \alpha_j}
\]

\[
V[p_k] = \frac{E[p_k](1 - E[p_k])}{1 + \sum_{j=1}^{K} \alpha_j}
\]

\[
C[p_i, p_j] = \frac{-E[p_i]E[p_j]}{1 + \sum_{k=1}^{K} \alpha_k}
\]

If \( \alpha_k > 1, k = 1, \ldots, K \), the maximum density occurs at

\[
p_k = \frac{\alpha_k - 1}{(\sum_{j=1}^{K} \alpha_j) - K}
\]
The uniform distribution on the \( K - 1 \)-Simplex is a special case of the Dirichlet distribution where \( \alpha_k = 1, \ k = 1, \ldots, K \).

\[
P(p_1, \ldots, p_K | \alpha_1, \ldots, \alpha_K) = \frac{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} p_k^{\alpha_k - 1}
\]

\[
P(p_1, \ldots, p_K | 1, \ldots, 1) = \frac{\Gamma(K)}{\Gamma(1)} \prod_{k=1}^{K} p_k
\]

\[
= (K - 1)!
\]
The Beta Distribution

The Beta distribution is a special case of the Dirichlet distribution for $K = 2$.

$$P(y) = \frac{1}{B(p, q)} y^{p-1} (1 - y)^{q-1}$$

where $p > 0$ and $q > 0$ and $0 \leq y \leq 1$

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p + q)}$$
In the case of the Dirichlet prior distribution, the derivation goes in a similar manner.

\[
Prob(d_k \mid l_1, \ldots, l_K) = \frac{Prob(d_k, l_1, \ldots, l_K)}{Prob(l_1, \ldots, l_K)}
\]

\[
= \frac{\int_{(p_1, \ldots, p_K) \in S} Prob(d_k, l_1, \ldots, l_K, p_1, \ldots, p_K) dp_1 \ldots dp_K}{\int_{(q_1, \ldots, q_K) \in S} Prob(l_1, \ldots, l_K, q_1, \ldots, q_K) dq_1 \ldots dq_K}
\]

\[
= \frac{\int_{(p_1, \ldots, p_K) \in S} Prob(l_1, \ldots, l_{k-1}, l_k + 1, l_{k+1}, \ldots, l_K \mid p_1, \ldots, p_K) P(p_1, \ldots, p_K) dp_1 \ldots dp_K}{\int_{(q_1, \ldots, q_K) \in S} Prob(l_1, \ldots, l_K \mid q_1, \ldots, q_K) P(q_1, \ldots, q_K) dq_1 \ldots dq_K}
\]
Dirichlet Prior

\[
\text{Prob}(I_1, \ldots, I_K) = \int_{(q_1, \ldots, q_K) \in S} \text{Prob}(I_1, \ldots, I_K, q_1, \ldots, q_K) dq_1 \ldots dq_K
\]

\[
= \int_{(q_1, \ldots, q_K) \in S} \left[ \frac{\prod_{n=1}^{K} I_n!}{l!} \prod_{m=1}^{K} q_m^{l_m} \right] \left[ \frac{\Gamma(\sum_{n=1}^{K} \alpha_n)}{\prod_{n=1}^{K} \Gamma(\alpha_n)} \prod_{m=1}^{K} q_m^{\alpha_m - 1} \right] dq_1, \ldots, dq_K
\]

\[
= \frac{\prod_{n=1}^{K} I_n!}{l!} \frac{\Gamma(\sum_{n=1}^{K} \alpha_n)}{\prod_{n=1}^{K} \Gamma(\alpha_n)} \int_{(q_1, \ldots, q_K) \in S} \prod_{m=1}^{K} q_m^{l_m + \alpha_m - 1} dq_1, \ldots, dq_K
\]

\[
= \frac{\prod_{n=1}^{K} I_n!}{l!} \frac{\Gamma(\sum_{n=1}^{K} \alpha_n)}{\prod_{n=1}^{K} \Gamma(\alpha_n)} \frac{\prod_{k=1}^{K} (l_k + \alpha_k - 1)!}{\prod_{n=1}^{N} \Gamma(\alpha_n) (l - 1 + \sum_{k=1}^{K} \alpha_k)!}
\]
Dirichlet Prior

\[
\text{Prob}(d_k, I_1, \ldots, I_K) = \int_{(q_1, \ldots, q_K) \in S} \text{Prob}(d_k, I_1, \ldots, I_K, q_1, \ldots, q_K) dq_1 \ldots dq_K
\]

\[
= \int_{(q_1, \ldots, q_K) \in S} \text{Prob}(d_k) \text{Prob}(I_1, \ldots, I_K|q_1, \ldots, q_K) \text{Prob}(q_1, \ldots, q_K) dq_1, \ldots, dq_K
\]

\[
= \int_{(q_1, \ldots, q_K) \in S} \prod_{n=1}^K \frac{I_n!}{l_n!} \prod_{m=1}^K q_m^{l_m} \left[ \Gamma \left( \sum_{n=1}^K \alpha_n \right) \prod_{n=1}^K \Gamma (\alpha_n) \prod_{m=1}^K q_m^{\alpha_m - 1} \right] dq_1, \ldots, dq_K
\]

\[
= \prod_{n=1}^K \frac{I_n!}{l_n!} \left[ \Gamma \left( \sum_{n=1}^K \alpha_n \right) \prod_{n=1}^K \Gamma (\alpha_n) \prod_{m=1}^K q_m^{l_m + \alpha_m - 1} \right] dq_1, \ldots, dq_K
\]

\[
= \prod_{n=1}^K \frac{I_n!}{l_n!} \left[ \Gamma \left( \sum_{n=1}^N \alpha_n \right) \prod_{n=1}^N \Gamma (\alpha_n) (l_n + \alpha_n)(l_n + \alpha_n - 1)! \right] \left( l + \sum_{n=1}^K \alpha_n \right) (l - 1 + \sum_{n=1}^K \alpha_n)!
Dirichlet Prior

\[
\text{Prob}(d_k, l_1, \ldots, l_K) = \frac{\prod_{n=1}^{K} l_n! \frac{\Gamma(\sum_{n=1}^{K} \alpha_n)}{\prod_{n=1}^{N} \Gamma(\alpha_n)} \frac{(l_k + \alpha_k) \prod_{n=1}^{K} (l_n + \alpha_n - 1)!}{(l + \sum_{n=1}^{K} \alpha_n)(l - 1 + \sum_{n=1}^{K} \alpha_n)!}}{l!}
\]

\[
\text{Prob}(l_1, \ldots, l_K) = \frac{\prod_{n=1}^{K} l_n! \frac{\Gamma(\sum_{n=1}^{K} \alpha_n)}{\prod_{n=1}^{N} \Gamma(\alpha_n)} \prod_{k=1}^{K} (l_k + \alpha_k - 1)!}{l! \prod_{n=1}^{N} \Gamma(\alpha_n) (l - 1 + \sum_{k=1}^{K} \alpha_k)!}
\]

\[
\text{Prob}(d_k \mid l_1, \ldots, l_K) = \frac{(l_k + \alpha_k) \prod_{n=1}^{K} (l_n + \alpha_n - 1)!}{(l + \sum_{k=1}^{K} \alpha_k)(l - 1 + \sum_{k=1}^{K} \alpha_k)!}
\]

\[
\frac{\prod_{n=1}^{K} (l_n + \alpha_n - 1)!}{(l - 1 + \sum_{k=1}^{K} \alpha_k)!}
\]

\[
= \frac{l_k + \alpha_k}{l + \sum_{n=1}^{K} \alpha_n}
\]