VARIABLE-RATE VQ (AKA VQ WITH ENTROPY CODING)

- Variable-Rate VQ = Quantization + Lossless Variable-Length Binary Coding
- A range of options -- from simple to complex
  a. Uniform scalar quantization with variable-length coding, one index at a time.
    \[
    \begin{array}{cccccccccc}
    1100 & 1101 & 100 & 00 & 01 & 101 & 1110 & 1111 \\
    \end{array}
    \]
    \[
    w_1 \quad w_2 \quad \Delta \quad w_N
    \]
  b. Nonuniform scalar quantization with variable-length coding -- one index at a time.
  c. Scalar quantization with higher-order variable-length coding -- either block coding of \( n \) indices at a time or \( n \)-th order conditional coding of the indices.
  d. \( k \)-dimen'l VQ with variable-length coding, one index at a time.
  e. \( k \)-dimen'l VQ with higher-order variable-length coding -- either block coding of \( n \) indices at a time or \( n \)-th order conditional coding of the indices.
  f. \( k \)-dimen'l VQ with other types of lossless coding

- We study e. with block coding. Conditional coding is just a slight variation. a.-d. are special cases of e.

D. VQ WITH 1ST-ORDER LOSSLESS CODING

- Key characteristics: \( k \)-dimensional VQ with partition \( S = \{S_1, \ldots, S_M\} \), codebook \( C = \{w_1, \ldots, w_M\} \), quantization rule \( Q \), and binary prefix codebook \( B = C_b = \{c_1, \ldots, c_M\} \) with lengths \( \{L_1, \ldots, L_M\} \)
- Decompose encoder into "partition" and "binary encoder"
  Given \( x \), the partition produces index \( i \) when \( x \in S_i \)
  The binary encoder outputs binary codeword \( c_i \) with length \( L_i \)
- Decompose decoder into "binary decoder" and "codebook"
  The binary decoder decodes the bits into the index \( i \).
  The codebook outputs \( w_i \).
- Rate = \( R = \frac{1}{k} \sum_{x} L(x) \ p(x) = \frac{1}{k} \times \text{rate of binary encoder} \)
  we often assume \( R = \frac{1}{k} H(I) = \frac{1}{k} H(Q(X)) \) ("VQ with entropy coding")
- Distortion = \( D = \frac{1}{k} \sum_{x} ||X - Q(X)||^2 \) (not affected by choice of lossless coder)
E. VQ WITH BLOCK LOSSLESS CODING

- k-dimen'l VQ with variable-length coding of indices in blocks of n.
- Decompose encoder into "partition" and "binary encoder":
  "Partition" n successive k-dim'l source vectors \( X_1, \ldots, X_n \) into indices \( I_1, \ldots, I_n \), where \( X_j = (X_{j,1}, \ldots, X_{j,k}) \). Losslessly encode n indices at once, \((I_1, \ldots, I_n)\), using FVL block lossless code with prefix codebook \( C_b \) containing \( M^n \) binary codewords.
- Decompose decoder into "binary decoder" and "codebook":
  Decode binary codeword into n indices \( I_1, \ldots, I_n \). Output corresponding quantization vectors \( w_{i_1}, \ldots, w_{i_n} \) as reproductions of \( X_1, \ldots, X_n \), respectively.
- "Block lossless binary coding" is an easy to analyze paradigm for studying the benefits of variable-length coding (a.k.a entropy coding).
- **We usually assume source is stationary**, so that \( X_j \) has the same pdf for all \( j \), which will be denoted \( f_{X}(x) \), \( f(x) \) or \( f_k(x) \).

**SUMMARY OF CHARACTERISTICS**

**Quantizer** = k-dimensional VQ
- \( k = \) dimension
- \( M = \) size (not necessarily power of 2, not so important, may be infinite)
- \( S = \{S_1, \ldots, S_M\} = \) k-dimensional partition
- \( C = \{w_1, \ldots, w_M\} = \) k-dimensional codebook
- \( Q(x) = \) quantization rule

**Binary Encoder:**
- \( n = \) order of binary encoder (i.e. input blocklength)
- \( C_b = \{c_j : j \in J\} = \) binary prefix codebook -- one cdwrd for each seq. of \( n \) indices
  where \( J = \) set of cell index n-tuples = \( \{j = (i_1, \ldots, i_n) : 1 \leq i_1 \leq M, \ldots, 1 \leq i_n \leq M\} \)
  \( c_j = (c_{i_1}, \ldots, c_{i_n}) = \) binary codeword of length \( L_j \) for \( j \)

**Derivative Characteristics:**
- quantization rule: \( Q(x_j) = w_i \) when \( x_j \in S_i \)
- encoding rule: \( e(x_1, \ldots, x_n) = c_j \), when \( j = (i_1, \ldots, i_n) \) and \( x_j \in S_i \), \( j = 1, \ldots, n \)
- decoding rule: \( d(c_j) = (w_1, w_2, \ldots, w_n) \)
**PERFORMANCE**

**Distortion** (same as usual)

\[ D = \frac{1}{k} E \|X - Q(X)\|^2 = \frac{1}{k} \sum_{i=1}^{M} \int_{S_i} \|x - y_i\|^2 f_k(x) \, dx \]

where \( X = (X_1, \ldots, X_k) \) and \( f_k(x) \) is its density. Dist'n depends on \( S \) and \( C \) but not \( C_B \).

**Rate**

\[ R = \frac{1}{k n} \bar{L} = \frac{1}{k n} \sum_j P_j L_j \text{ bits/sample} \]

where \( \bar{L} \) = average length of binary codewords

\[ P_j = \text{probability of binary codeword } c_j = \Pr(X_1 \in S_{i1}, \ldots, X_n \in S_{in}), \quad j \in I \]

From lossless coding theorem

\[ H(I) \leq \bar{L}_n \leq H(I) + 1 \]

where \( \bar{L}_n \) = least avg. length of prefix code for given VQ & \( n \)

\[ H(I) = -\sum_j P_j \log_2 P_j = \text{entropy of } I \text{ (or of } (Y_1, \ldots, Y_n)) \]

From now on we assume (unless otherwise stated) that

\[ R = \frac{1}{k n} H(I) = \frac{1}{k} H_n(I) = \frac{1}{k n} H(Y_1, \ldots, Y_n) = H_{kn}(Y) \]

We call this "**VQ with nth-order entropy coding (EC)**".

---

**IMPLEMENTATION AND COMPLEXITY**

- Quantizer -- same issues as with fixed-rate coding
- Lossless Coder -- table lookup is the brute force method
  - Table stores \( M^n \) binary codewords of various lengths
  - \( M = 2^{nkR_f} \) where \( R_f = \frac{1}{k} \log_2 M \) is "fixed-rate" rate
  - Complexity of brute force implementation of entropy increases exponentially with \( nkR_f \).

**OPTIMAL PERFORMANCE**

- OPTA functions we seek

\[ \delta(k,n,R) \overset{\Delta}{=} \text{least MSE of k-dim'l VQ w. nth-order entropy coding and rate } R \text{ or less} \]

\[ S(k,n,R) \overset{\Delta}{=} \text{max SNR of k-dim'l VQ's w. nth-order entropy coding and rate } R \text{ or less} \]

\[ \delta(R) \overset{\Delta}{=} \inf_{k,n} \delta(k,n,R) = \text{least MSE of VQ with EC and rate } R \text{ or less (any } k,n) \]

\[ S(R) \overset{\Delta}{=} \sup_{k,n} \delta(k,n,R) = \text{max SNR of VQ with EC and rate } R \text{ or less (any } k,n) \]
HIGH-RESOLUTION ANALYSIS

- As before, assume the VQ has mostly small cells, negligible overload distortion, large \( M \), neighboring cells with similar sizes & shapes, point density approx'ly \( \Lambda(x) \), inertial profile approx'y \( m(x) \)
- Since quantizer size is unimportant (e.g. its not related to rate), and can even be infinite, we use unnormalized point density, \( \Lambda(x) \), which is a function such that
  
  1. \( \int_A \Lambda(x) \, dx \equiv \) number of codevectors (or cells) in region \( A \)
  
  2. If \( A \) is small, but much larger than the cells in the vicinity of \( x \),
      \( \Lambda(x) \, |A| \equiv \# \) points\cells in \( A \)
  
  3. \( \Lambda(x) \geq 0, \int \Lambda(x) \, dx = M = \) total number of quantization points (can be \( \infty \))
  
  4. Ordinarily \( \Lambda(x) \) is a smooth or piecewise smooth function.
  
  5. \( \Lambda(x) \equiv \frac{1}{|S_i|} \) when \( x \in S_i \)

DISTORTION: BENNETT'S INTEGRAL

Under high-resolution conditions, a derivation like that for the original Bennett shows

\[
D \equiv \int \frac{m(x)}{\Lambda^{2/k}(x)} \, f_k(x) \, dx
\]

RATE: ASYMPTOTIC ENTROPY FORMULA

Fact: If \( X_1, \ldots, X_n \) are identically distributed, then under high-resolution conditions,

\[
R = \frac{1}{kn} H(l) \equiv h_{kn} + \frac{1}{k} \int f_k(x) \log_2 \Lambda(x) \, dx
\]

where

\[
\bar{x} = (x_1 \ldots x_k)
\]

\[
h_{kn} = \frac{1}{kn} h(X_1, \ldots, X_{kn}) = \text{kn-th order differential entropy}
\]

\[
= -\frac{1}{kn} \int f_{kn}(x_1 \ldots x_{kn}) \log_2 f_{kn}(x_1 \ldots x_{kn}) \, dx_1 \ldots dx_{kn}
\]
Most Important Example:
Uniform scalar quantizer with step size $\Delta$, infinite support, and infinitely many levels

$$\Lambda(x) \equiv \frac{1}{\Delta}$$

Then from the approximate rate formula

$$R \approx h_k n + \frac{1}{k} \int f_k(x) \log_2 \Lambda(x) \, dx$$

$$= h_n - \log \Delta = h_n - \frac{1}{2} \log 12 \frac{\Delta^2}{12}$$

$$\approx h_n - \frac{1}{2} \log 12 \, D$$

Equivalently

$$D \approx \frac{1}{12} 2^{2h_n} 2^{-2R}$$

---

**DERIVATION OF ASYMPTOTIC FORMULA FOR $H(I)$**

First case: $n = 1$ (for simplicity)

$$H(I) = - \sum_i P_i \log P_i , \text{ where } P_i = \Pr(X_1 \in S_i) = \int_{S_i} f_k(x) \, dx$$

$$= - \sum_i \left( \int_{S_i} f_k(x) \, dx \right) \log \left( \int_{S_i} f_k(x) \, dx \right)$$

$$\approx - \sum_i f_k(w_i) |S_i| \log \left( f_k(w_i) |S_i| \right) \text{ because cells are small}$$

$$= - \sum_i f_k(w_i) |S_i| \log \left( f_k(w_i) |S_i| \right)$$

$$= - \sum_i \left( f_k(w_i) \log f_k(w_i) \right) |S_i| - \sum_i \left( f_k(w_i) \log \frac{1}{\Lambda(w_i)} \right) |S_i|$$

$$\approx - \int f_k(x) \log f(x) \, dx + \int f_k(x) \log \Lambda(x) \, dx$$

$$= h_k + \frac{1}{k} \int f_k(x) \log \Lambda(x) \, dx$$
General case: \( n \geq 1 \)

\[
H(I) = - \sum_{i} P_i \log P_i
\]

where \( \bar{x} = (x_1 \ldots x_{kn}) \)

\[
\approx - \sum_{i} \left( \int_{S_i} f_{kn}(w_i) \, \log f_{kn}(w_i) \right) \mid S_i \mid \log ( \mid S_i \mid )
\]

The first summation above can be approximated by the integral

\[
- \int f_{kn}(x) \log f_{kn}(x) \, dx = kn \ h_{kn}
\]  

Before approximating the second sum, note that

\[
\log |S_i| = \log ( |S_1| IS_2 | \ldots IS_n| ) = \sum_{j=1}^{n} \log |S_j| \equiv - \sum_{j=1}^{n} \log \Lambda(w_j)
\]

Substitute the above into the second summation:

\[
\approx - \sum_{i} \left( f_{kn}(w_i) \log |S_i| \right) \mid S_i \mid
\]

\[
\approx \int f_{kn}(x_1 \ldots x_n) \sum_{j=1}^{n} \log \Lambda(x_j) \, dx_1 \ldots dx_k
\]

\[
= \sum_{j=1}^{n} \int f_{kn}(x_1 \ldots x_n) \log \Lambda(x_j) \, dx_1 \ldots dx_k
\]

\[
= \sum_{j=1}^{n} \int f_k(x_j) \log \Lambda(x_j) \, dx_j
\]

\[
= n \int f_k(x_1) \log \Lambda(x) \, dx_1 \quad \text{because } X_1, \ldots X_n \text{ are identical}
\]

Substituting (\( \ast \)) and (\( \ast \ast \)) into the expression for \( H(I) \) gives

\[
\frac{1}{kn} H(I) = \frac{1}{kn} \left( kn \ h_{kn} + n \int f_k(x_1) \log \Lambda(x) \, dx_1 \right)
\]

\[
= h_{kn} + \frac{1}{k} \int f_k(x) \log_2 \Lambda(x) \, dx
\]
ZADOR-GERSHO THEOREM FOR VARIABLE-RATE VQ

For a stationary source and large $R$, the least distortion of $k$-dim'l VQ with nth-order entropy coding and rate $R$ or less is

$$\delta(k,n,R) \equiv m^*_k \sigma^2 \eta_{kn} 2^{-2R} \triangleq Z(k,nR)$$

where

- $Z(k,n,R)$ = Zador-Gersho function for $k$-dim'l VQ with nth-order entropy coding
- $m^*_k$ = best inertial profile = least NMI of any tessel'g polytope (Gersho's conj.)
- $\eta_{kn} = \frac{1}{\sigma^2} 2^{2h_{kn}}$

Equivalently, $S(k,n,R) \approx 6.02 R - 10 \log_{10} m^*_k \eta_{kn}$. (Again, 6 dB per bit.)

Notes:

- $k$-dimensional VQ-EC with $n = 1$ is at least as good as VQ-FR, because the latter is a special case of the former. Later we show directly that $\eta_k \leq \beta_k$, which implies $Z(k,1,R) \leq Z(k,R)$.
- The proof shows that an approximately optimal $k$-dimensional VQ with nth-order entropy coding can be constructed with a partition that is a tesselation of the best $k$-dimensional polytope, scaled to volume $2^{k(h_{kn}-R)}$. This has constant inertial profile $m(x) = m^*_k$, constant point density $\Lambda(x) = \Lambda^*_k = 2^{k(R-h_{kn})}$, distortion $D \equiv m^*_k (\Lambda^*_k)^{-2/k} = Z(k,nR)$.

PROOF OF ZADOR-GERSHO THEOREM

We begin with

$$\delta(k,n,R) \equiv \min_{m(x),\Lambda(x)} \int \frac{m(x)}{\Lambda^{2/k}(x)} f(x) \, dx$$

where $x = (x_1...x_k)$ and the minimization is over all inertial profiles $m(x)$ and all point densities $\Lambda(x)$ such that

$$h_{kn} + \frac{1}{R} \int f(x) \log_2 \Lambda(x) \, dx \leq R \quad \text{(such a } \Lambda \Rightarrow \text{ rate } \leq R)$$

- Best inertial profile:
  - We assume Gersho's conjecture -- In the high rate, small distortion regime, most cells of an optimal quantizer are, approximately, congruent to the tesselating polytope with least NMI.
  - We conclude: The best inertial profile is $m^*_k(x) = m^*_k \triangleq$ least NMI of $k$-dimen'l tess'ng polytopes
  - It follows that

$$\delta(k,n,R) \equiv m^*_k \min_{\Lambda(x)} \int \frac{1}{\Lambda^{2/k}(x)} f(x) \, dx$$

where the min is taken over functions $\Lambda(x)$ such that $\Lambda(x) \geq 0$ and

$$h_{kn} + \frac{1}{R} \int f(x) \log_2 \Lambda(x) \, dx \leq R$$
Best point density:

Suppose $\Lambda(x)$ satisfies

$$h_{kn} + \frac{1}{k} \int f(x) \log_2 \Lambda(x) \, dx \leq R \quad (\star)$$

Then by convexity of the logarithm and Jensen's inequality

$$\log_2 \int \frac{1}{\Lambda^{2/k}(x)} f(x) \, dx \geq \int \log_2 \left( \frac{1}{\Lambda^{2/k}(x)} \right) f(x) \, dx \quad \text{equality iff } \Lambda(x) \text{ is constant with probability one}$$

$$= -\frac{2}{k} \int f(x) \log \Lambda(x) \, dx \geq 2h_{kn} - 2R$$

Hence,

$$\int \frac{1}{\Lambda^{2/k}(x)} f(x) \, dx \geq 2^{2h_{kn} - 2R} \quad (\star\star)$$

with equality iff $\Lambda(x)$ is a constant with probability one.

$(\star)$ and $(\star\star) \Rightarrow \delta(k,n,R) \equiv m_k^* 2^{2h_{kn}} 2^{-2R} = m_k^* \sigma^2 \eta_{kn} 2^{-2R}$

Moreover, we have shown that the optimal point density is a constant. The constant must be such that $(\star)$ holds with equality. Therefore,

$$\Lambda(x) = 2^{k(R-h)} \triangleq \Lambda_k^*$$

We see from the proof that an approximately VQ can be constructed with a partition that is a tesselation of the best $k$-dimensional polytope, scaled to volume $2^{k(h_{kn}-R)}$.

Indeed, the tesselation need only cover the region where $f(x)$ is not small.

**SUMMARY OF FIXED- AND VARIABLE-RATE VQ**

Let 0th-order entropy coding denote fixed-rate coding.

Then, given a stationary source and large $R$, the least distortion of VQ with rate $R$ or less, dimension $k$, and nth-order order entropy coding is

$$\delta(k,n,R) \equiv \mathcal{Z}(k,n,R) = m_k^* \sigma^2 \eta_{k,n} \sigma^2 2^{-2R},$$

where

$$\eta_{k,n} = \begin{cases} 
\beta_k, & n = 0 \\
2^{2h_{kn}}/\sigma^2, & n \geq 1 
\end{cases}$$

Notice the ",," in $\eta_{k,n}$ but not in $h_{kn}$!