Exercises with solutions (4)

1. Consider a symmetric scalar quantizer with 3 intervals,

\[ q(x) = \begin{cases} 
- b : & x < - a \\
0 : & |x| \leq a \\
b : & x > a 
\end{cases} \]

and a quantizer input with a zero-mean Laplace pdf,

\[ f(x) = \frac{1}{2m} e^{-\frac{|x|}{m}} \]

(a) Derive the optimal reconstruction value \( b \) as a function of the decision threshold \( a \) for MSE distortion.

Express the resulting distortion as function of \( a \) and the variance \( \sigma^2 = 2m^2 \).

Solution:

Due to the symmetry of the pdf and the quantizer design, the MSE distortion can be written as

\[
D_2(a, b) = 2 \int_0^a x^2 f(x) \, dx + 2 \int_a^\infty (x-b)^2 f(x) \, dx \\
= 2 \int_0^\infty x^2 f(x) \, dx - 4b \int_0^a x f(x) \, dx + 2b^2 \int_a^\infty f(x) \, dx
\]

Setting the first partial derivative with respect to \( b \) equal to 0,

\[
\frac{\partial}{\partial b} D_2(a, b) = 0 = -4 \int_a^\infty x f(x) \, dx + 4b \int_a^\infty f(x) \, dx
\]

yields the condition for the optimal reconstruction value \( b \),

\[
b = \frac{\int_a^\infty x f(x) \, dx}{\int_a^\infty f(x) \, dx},
\]

which is known as the centroid condition for MSE distortion.

Inserting this expression into the equation yields the following formula for the distortion (when the centroid condition is fulfilled),

\[
D_2^*(a) = 2 \int_0^\infty x^2 f(x) \, dx - 2 \left( \frac{\int_a^\infty x f(x) \, dx}{\int_a^\infty f(x) \, dx} \right)^2
\]

1
By defining

\[ A = 2 \int_0^\infty x^2 f(x) \, dx = \frac{1}{m} \int_0^\infty x^2 e^{-\frac{x}{m}} \, dx \]

\[ B(a) = 2 \int_a^\infty x f(x) \, dx = \frac{1}{m} \int_a^\infty x e^{-\frac{x}{m}} \, dx \]

\[ C(a) = 2 \int_a^\infty f(x) \, dx = \frac{1}{m} \int_a^\infty e^{-\frac{x}{m}} \, dx \]

we can also write

\[ b(a) = \frac{B(a)}{C(a)} \quad \text{and} \quad D_2^*(a) = A - \frac{B(a)^2}{C(a)} \]

For the integrals \( A, B(a) \) and \( C(a) \), we obtain

\[ A = \frac{1}{m} \int_0^\infty x^2 e^{-\frac{x}{m}} \, dx = \left[ -e^{-\frac{x}{m}} \left( x^2 + 2mx + 2m^2 \right) \right]_0^\infty = 2m^2 \]

\[ B(a) = \frac{1}{m} \int_a^\infty x e^{-\frac{x}{m}} \, dx = \left[ -e^{-\frac{x}{m}} \left( x + m \right) \right]_a^\infty = (a + m) e^{-\frac{a}{m}} \]

\[ C(a) = \frac{1}{m} \int_a^\infty e^{-\frac{x}{m}} \, dx = \left[ -e^{-\frac{x}{m}} \right]_a^\infty = e^{-\frac{a}{m}} \]

Then, we obtain for the optimal reconstruction value

\[ b(a) = \frac{B(a)}{C(a)} = \frac{(a + m) e^{-\frac{a}{m}}}{e^{-\frac{a}{m}}} = a + m = a + \frac{1}{2} \sqrt{2\sigma^2} \]

And the distortion for a centroidal quantizer is given by

\[ D_2^*(a) = A - \frac{B(a)^2}{C(a)} = 2m^2 - \frac{(a + m)^2 e^{-\frac{2a}{m}}}{e^{-\frac{a}{m}}} = 2m^2 - (a + m)^2 e^{-\frac{a}{m}} \]

By using the variance \( \sigma^2 = 2m^2 \), we obtain

\[ D_2^*(a) = \sigma^2 - \left( a + \frac{1}{2} \sqrt{2\sigma^2} \right)^2 e^{-\frac{2a}{\sqrt{2}\sigma^2}} \]
(b) What is the general centroid condition for MAE distortion?

Derive the optimal reconstruction value \( b \) as a function of the decision threshold \( a \) for MAE distortion and express the resulting distortion as function of \( a \) and \( m \).

**Solution:**

In the general case, the MAE distortion is given by

\[
D_1 = \sum_k^{u_{k+1}} \int_{u_k}^{s_k'} |x - s_k' f(x) dx
\]

\[
= \sum_k^{u_{k+1}} \left( \int_{s_k}^{s_k'} (s_k' - x) f(x) dx + \int_{s_k}^{u_k} (x - s_k') f(x) dx \right)
\]

\[
= \sum_k s_k' \left( \int_{u_k}^{s_k'} f(x) dx - \int_{s_k}^{u_k} f(x) dx \right) + \left( \int_{s_k}^{u_{k+1}} f(x) dx - \int_{s_k}^{s_k'} f(x) dx \right)
\]

Setting the partial derivative with respect to \( s_k' \) equal to 0,

\[
\frac{\partial}{\partial s_k'} D_1 = 0 = \left( \int_{u_k}^{s_k'} f(x) dx - \int_{s_k}^{u_k} f(x) dx \right)
\]

\[
+ s_k' \left( f(s_k') + f(s_k') \right) + \left( -s_k f(s_k') - s_k f(s_k') \right)
\]

yields the centroid condition for MAE distortion,

\[
\int_{u_k}^{s_k'} f(x) dx = \int_{s_k}^{u_k} f(x) dx,
\]

which specifies that the optimal reconstruction level divides the quantization cell into two parts with the same probability.

The distortion is then given by

\[
D_1^* = \sum_k \left( \int_{s_k}^{u_{k+1}} f(x) dx - \int_{s_k}^{u_{k+1}} f(x) dx \right)
\]

For our example, we have

\[
\int_a^b f(x) dx = \frac{1}{2m} \int_a^b e^{-\frac{x}{m}} \approx x = \left[ -\frac{1}{2} e^{-\frac{a}{m}} \right]_a^b = \frac{1}{2} \left( e^{-\frac{b}{m}} - e^{-\frac{a}{m}} \right)
\]

and

\[
\int_b^\infty f(x) dx = \frac{1}{2m} \int_b^\infty e^{-\frac{x}{m}} \approx x = \left[ -\frac{1}{2} e^{-\frac{x}{m}} \right]_b^\infty = \frac{1}{2} e^{-\frac{b}{m}}
\]
yielding the centroid condition
\[
\frac{1}{2} (e^{-\frac{a}{m}} - e^{-\frac{b}{m}}) = \frac{1}{2} e^{-\frac{a}{m}}
\]
\[
e^{-\frac{b}{m}} = \frac{1}{2} e^{-\frac{a}{m}}
\]
\[
-\frac{b}{m} = - \ln 2 - \frac{a}{m}
\]
\[
b = a + m \ln 2
\]

Furthermore, we have
\[
\int_y^z x f(x) \, dx = \frac{1}{2m} \int_y^z x e^{-\frac{x}{m}} \, dx = \left[ -\frac{1}{2} e^{-\frac{x}{m}} (m + x) \right]_y^z
\]
\[
= \frac{m + y}{2} e^{-\frac{x}{m}} - \frac{m + z}{2} e^{-\frac{x}{m}}
\]
Hence, the distortion for the centroidal quantizer for MAE distortion is given by
\[
D_1^* = 2 \left( \int_0^a x f(x) \, dx - \int_a^{a + m \ln 2} x f(x) \, dx + \int_{a + m \ln 2}^{\infty} x f(x) \, dx \right)
\]
\[
= m - (m + a) e^{-\frac{a}{m}} - (m + a) e^{-\frac{a + m \ln 2}{m}} + (m + a + m \ln 2) e^{-\frac{a + m \ln 2}{m}} - 0
\]
\[
= m - e^{-\frac{a}{m}} \left( -2(m + a) + 2(m(1 + \ln 2) + a) \frac{1}{2} \right)
\]
\[
= m - e^{-\frac{a}{m}} \left( m(1 - \ln 2) + a \right)
\]

(c) Determine the decision threshold \( a \) for both distortion measures in a way that a Lloyd quantizer is obtained.

Determine the distortion and rate for the Lloyd quantizers by assuming fixed-length coding \( R = \log_2 N \) and compare the obtained R-D point with the Shannon lower bound for MSE and the rate-distortion function for MAE.

Solution:
A Lloyd quantizer is the optimal quantizer (minimum distortion) for a given number of reconstruction levels. It fulfills two conditions: the centroid condition considered above and the so-called nearest neighbor condition:
\[
u_k = \frac{1}{2} (s_{k-1} + s_k)
\]
For our given 3-interval quantizer, we have
\[
a = \frac{1}{2} (0 + b) = \frac{b}{2}
\]
Hence, for MSE distortion, we obtain
\[ a = \frac{a + m}{2} \quad \Rightarrow \quad a^*_2 = m = \frac{1}{2}\sqrt{2\sigma^2} \]
and, for MAE distortion, we obtain
\[ a = \frac{a + m \ln 2}{2} \quad \Rightarrow \quad a^*_1 = m \cdot \ln 2 \]

It should be noted that these parameters are also obtained by setting the derivatives of the distortions \( D^*_1(a) \) and \( D^*_2(a) \) with respect to \( a \) equal to 0.

For the distortions, we obtain
\[
D^*_2 = \sigma^2 - \left( a^*_1 + \frac{1}{2}\sqrt{2\sigma^2} \right)^2 e^{-\frac{a^*_1}{\sqrt{2\sigma^2}}}
\]
\[
= \sigma^2 - \left( \frac{1}{2}\sqrt{2\sigma^2} + \frac{1}{2}\sqrt{2\sigma^2} \right)^2 e^{-\frac{1}{\sqrt{2\sigma^2}}\sqrt{2\sigma^2}}
\]
\[
= \sigma^2 - \frac{2}{e} \sigma^2 = \sigma^2 \left( 1 - \frac{2}{e} \right) \approx 0.26421 \cdot \sigma^2
\]

and
\[
D^*_1 = m - e^{-\frac{a^*_1}{\ln 2}} \left( m(1 - \ln 2) + a^*_1 \right)
\]
\[
= m - e^{-\frac{a^*_1}{\ln 2}} \left( m(1 - \ln 2) + m \ln 2 \right)
\]
\[
= m \left( 1 - e^{-\ln 2} \right) = \frac{1}{2} m
\]

The nominal rate \( R \) for both Lloyd quantizers is
\[ R_1 = R_2 = \log_2 3 \]

For MAE distortion, the rate-distortion function (which is equal to the Shannon lower bound for the Laplace pdf and MAE distortion) is given by
\[ D^{(I)}_1(R) = m 2^{-R} \quad \text{or} \quad R^{(I)}_1(D) = \log_2 \left( \frac{m}{D} \right) \]

Hence, the distortion bound \( D^{(I)}_1 \) at rate \( R_1 \) is
\[ D^{(I)}_1(R_1) = m 2^{-\log_2 \frac{m}{3}} = \frac{m}{3} = \left( \frac{1}{2} m \right) \cdot \frac{2}{3} = \frac{2}{3} D^*_1, \]
which means that the distortion for the Lloyd quantizer is factor 1.5 larger than the rate-distortion bound.

The rate bound \( R^{(I)}_1 \) at distortion \( D^*_1 \) is
\[ R^{(I)}_1(D^*_1) = \log_2 2 = 1 = \log_2 3 - (\log_2 3 - 1) \approx R_1 - 0.584963, \]
which means that the rate for the Lloyd quantizer is approximately 0.585 bits per symbol (or 58.5%) larger than the rate-distortion bound.

For MSE distortion, the Shannon lower bound is given by

\[ D_{2}^{\text{SLB}}(R) = \frac{e}{\pi} \sigma^2 2^{-2R} \quad \text{or} \quad R_{2}^{\text{SLB}}(D) = \frac{1}{2} \log_2 \left( \frac{\sigma^2}{D} \cdot \frac{e}{\pi} \right) \]

Hence, the distortion for the Shannon lower bound \( D_{2}^{\text{SLB}} \) at rate \( R_2 \) is

\[
D_{2}^{\text{SLB}}(R_2) = \frac{e}{\pi} \sigma^2 2^{-2 \log_2 \frac{3}{2}} = \frac{e}{9\pi} \sigma^2 = \left( \frac{e - 2}{e} \right) \frac{2}{3}
\]

\[
= 9\pi(e - 2) D_{2}^{\text{SLB}} \approx 0.363833 \cdot D_{2}^{\text{SLB}}
\]

which means that the distortion for the Lloyd quantizer is approximately factor 2.75 (or 4.39 dB) larger than the Shannon lower bound.

The rate for the Shannon lower bound \( R_{2}^{\text{SLB}} \) at distortion \( D_2^{\text{SLB}} \) is

\[
R_{2}^{\text{SLB}}(D_2) = \frac{1}{2} \log_2 \left( \frac{e^2}{\pi (e - 2)} \right) = \log_2 3 - \frac{1}{2} \log_2 \left( \frac{9\pi(e - 2)}{e^2} \right)
\]

\[
\approx R_2 - 0.729326
\]

which means that the rate for the Lloyd quantizer is approximately 0.729 bits per symbol (or 85.2%) larger than the Shannon lower bound.

(d) Can the derived optimal quantizers for fixed-length coding be improved by adding entropy coding (without changing the decision thresholds and reconstruction levels).

\textbf{Solution:}

For the MSE-optimized quantizer, we obtain reconstruction symbols with the following pmf \( \{p_0, p_1, p_2\} \):

\[
p_1 = 2 \int_0^a f(x) \, dx = \frac{1}{m} \int_0^m e^{-\frac{x}{m}} \, dx = \left[ -e^{-\frac{x}{m}} \right]_0^m = 1 - \frac{1}{e}
\]

\[
p_0 = p_2 = \frac{1 - p_1}{2} = \frac{1}{2e}
\]

Since the probability masses are not the same, the performance of the quantizer can be improved by entropy coding.

The minimum rate is given by

\[
R_2^* = H = -2 \cdot \frac{1}{2e} \cdot \log_2 \left( \frac{1}{2e} \right) - 1 \cdot \left( 1 - \frac{1}{e} \right) \cdot \log_2 \left( 1 - \frac{1}{e} \right)
\]

\[
= \frac{1}{e} \log_2(2e) - \left( 1 - \frac{1}{e} \right) \cdot \log_2 \left( 1 - \frac{1}{e} \right) \approx 1.316909
\]
Hence, by entropy coding, the rate (for same distortion) could be reduced by approximately 0.27 bits per sample or 16.9%.

For the MAE-optimized quantizer, the pmf for the reconstruction symbols is given by

\[ p_1 = \frac{1}{m} \int_0^{m \ln 2} e^{-\frac{x}{m}} \, dx = [-e^{-\frac{x}{m}}]_0^{m \ln 2} = 1 - \frac{1}{2} = \frac{1}{2} \]

\[ p_0 = 1 - p_1 = \frac{1}{4} \]

For the minimum rate, which can be obtained by a simple code as \{01, 1, 00\}, we obtain

\[ R^*_2 = H = -2 \cdot \frac{1}{4} \cdot \log_2 \left( \frac{1}{4} \right) - 1 \cdot \frac{1}{2} \cdot \log_2 \left( \frac{1}{2} \right) = \frac{3}{2} \]

Here, entropy coding can reduce the rate by about 0.085 bits per sample or 5.26%.
2. Given is a Centroidal quantizer (not necessarily a Lloyd quantizer) for MSE distortion and a source $X$. The quantizer has 5 reconstruction levels $\{-3, -1, 0, 1, 3\}$ which are chosen with probabilities $\{0.05, 0.1, 0.4, 0.3, 0.15\}$ and achieves an MSE of 1.05.

(a) Determine the mean $\mu$ and variance $\sigma^2$ of the source $X$.

**Solution:**

We know that the quantizer obeys the centroid condition for MSE. Hence, the reconstruction levels can be written as 

$$s'_k = \frac{\int_{u_k}^{u_{k+1}} x f(x) \, dx}{\int_{u_k}^{u_{k+1}} f(x) \, dx} = \frac{1}{p_k} \int_{u_k}^{u_{k+1}} x f(x) \, dx,$$

where $p_k$ is the probability that the reconstruction level $s'_k$ is chosen, i.e., the probability that the value of $X$ falls inside the $k$-th quantization interval $[u_k, u_{k+1})$.

Then, the mean value $\mu$ of $X$ can be written as 

$$\mu = \int_{-\infty}^{\infty} x f(x) \, dx = \sum_k \int_{u_k}^{u_{k+1}} x f(x) \, dx = \sum_k p_k s'_k$$

It should be noted that this is the definition of the mean of the quantizer output $Q(X)$. Hence, a centroidal quantizer for MSE distortion does not modify the mean of the source $X$, we have 

$$E\{X\} = E\{Q(X)\}$$

For the mean of the quantization error $e(X) = X - Q(X)$, we obtain 

$$E\{e(X)\} = E\{X - Q(X)\} = E\{X\} - E\{Q(X)\} = 0$$

The quantization error of a centroidal quantizer for MSE distortion has always zero mean.

For the given quantizer, we obtain 

$$\mu = -0.05 \cdot 3 - 0.1 \cdot 1 + 0.3 \cdot 1 + 0.15 \cdot 3 = 0.5$$

Another given value is the MSE distortion, which can be written as 

$$D = \sum_k \int_{u_k}^{u_{k+1}} (x - s'_k)^2 f(x) \, dx$$

$$= \sum_k \left( \int_{u_k}^{u_{k+1}} x^2 f(x) \, dx - 2s'_k \int_{u_k}^{u_{k+1}} x f(x) \, dx + s'^2_k \int_{u_k}^{u_{k+1}} f(x) \, dx \right)$$

$$= \int_{-\infty}^{\infty} x^2 f(x) \, dx - \sum_k (2s'_k \cdot (s'_k p_k) - s'^2_k \cdot p_k)$$

$$= E\{X^2\} - \sum_k s'^2_k p_k$$
The last term on the right side is the second moment \( E\{q(X)^2\} \) of the quantizer output and the distortion is the second moment \( E\{e(X)\} = E\{(X - q(X))^2\} \) of the quantization error. Hence, for a centroidal quantizer for MSE distortion, the second moment of the input \( X \) is equal to the sum of the second moments of the quantizer output \( q(X) \) and the quantization error \( e(X) = X - q(X) \),

\[
E\{X^2\} = E\{q(X)^2\} + E\{e(X)^2\}
\]

Since the mean of \( X \) and \( q(X) \) is the same and the mean of \( e(X) \) is zero, we also have

\[
E\{X^2\} = E\{q(X)^2\} + E\{e(X)^2\} = 0
\]

The variance of the source \( X \) is equal to the sum of the variances of the quantizer output and the quantization error. The variance can then be expressed according to

\[
\sigma^2 = E\{(X - \mu)^2\} = E\{X^2\} - 2\mu E\{X\} + \mu^2 = E\{X^2\} - \mu^2
\]

\[
= E\{q(X)^2\} + E\{(X - q(X))^2\} - \mu^2
\]

\[
= \sum_k s_k^2 p_k + D - \mu^2
\]

For the given quantizer, we obtain

\[
\sigma^2 = 0.05 \cdot 9 + 0.1 \cdot 1 + 0.3 \cdot 1 + 0.15 \cdot 9 + 1.05 - 0.5^2 = 3
\]

(b) With \( q(X) \) being the quantizer output and \( e(X) = X - q(X) \) being the quantization error, determine the correlations \( E\{X q(X)\} \), \( E\{X e(X)\} \), and \( E\{q(X) e(X)\} \).

**Solution:**

The correlation \( E\{X q(X)\} \) can be written as

\[
E\{X q(X)\} = \int_{-\infty}^{\infty} x q(x) f(x, q(x)) dx
\]

\[
= \sum_k \int_{u_k}^{u_{k+1}} x s_k^2 f(x) g(q(x), x) dx
\]

\[
= \sum_k s_k^2 \int_{u_k}^{u_{k+1}} x f(x) dx = \sum_k s_k^2 p_k
\]

\[
= E\{q(X)^2\}
\]

For centroidal quantizers for MSE distortion, the correlation between the input and the output signal is equal to the second moment of the quantizer output.
For our example quantizer, we obtain
\[ E\{X q(X)\} = 0.05 \cdot 9 + 0.1 \cdot 1 + 0.3 \cdot 1 + 0.15 \cdot 9 = 2.2 \]

For the correlation between the input signal \( X \) and the quantization error, we obtain
\[
E\{X e(X)\} = E\{X (X - q(X))\} = E\{X^2\} - E\{X q(X)\} \\
= E\{e(X)^2\} + E\{q(X)^2\} - E\{q(X)^2\} \\
= E\{e(X)^2\} = D
\]

For centroidal quantizers for MSE distortion, the correlation between the input and quantization error is equal to the second moment of the quantization error, i.e., it is equal to the MSE distortion. Except the quantizer yields a distortion of zero, i.e., it does not apply any quantization, the input signal and the quantization error are always correlated.

For the given quantizer, the correlation \( E\{X e(X)\} \) is equal to 1.05.

Finally, for the correlation between quantizer output and quantization error, we obtain
\[
E\{q(X) e(X)\} = E\{q(X) (X - q(X))\} = E\{q(X)^2\} - E\{q(X)^2\} \\
= E\{q(X)^2\} - E\{q(X)^2\} = 0
\]

For centroidal quantizers for MSE distortion, the quantizer output and the quantization error are always uncorrelated.
3. Consider uniform threshold quantization of an exponential pdf given by
\[ f(x) = a e^{-ax} \]. With \( \Delta \) denoting the quantization step size, the decision thresholds are given by \( u_k = k\Delta \), with \( k = 0, 1, 2, \ldots \).

(a) Determine the probability mass function of the quantization indices and calculate the rate (assuming optimal entropy coding) as function of the probability
\[ p = P(X > \Delta) = e^{-a\Delta} \]
that a symbol \( X \) is greater than the quantization step size \( \Delta \).

For which step size \( \Delta \) is the rate equal to 2 bit per symbol?

Describe an entropy coding scheme for the quantization indices that virtually achieves the entropy.

**Hint:** For \(|a| < 1\),

\[ \sum_{k=0}^{\infty} a^k = \frac{1}{1-a} \quad \sum_{k=0}^{\infty} k a^k = \frac{a}{(1-a)^2} \]

**Solution:**
The probability \( p_k = P(u_k \leq X < u_{k+1}) \) that a source symbol falls into the \( k\)-th quantization interval is given by

\[ p_k = \int_{u_k}^{u_{k+1}} f(x) \, dx = \int_{k\Delta}^{(k+1)\Delta} a e^{-ax} \, dx \]

\[ = \left[ -e^{-ax} \right]_{k\Delta}^{(k+1)\Delta} = e^{-ak\Delta} - e^{-ak\Delta-a\Delta} \]

\[ = (e^{-a\Delta})^k (1 - e^{-a\Delta}) \]

\[ = p^k (1 - p) \]

where

\[ p = P(X > \Delta) = e^{-a\Delta} \]

is the probability that a source symbol \( X \) is greater than the quantization step size \( \Delta \) (that directly follows from considering \( p_0 = 1 - p \)).

The quantization indices that are obtained by uniform threshold quantization of an exponential pdf have a geometric distribution.

When considering optimal entropy coding, the rate \( R \) is equal to the
entropy and we obtain

\[ R = -\sum_{k=0}^{\infty} p^k (1-p) \log_2 (p^k (1-p)) \]

\[ = -(1-p) \sum_{k=0}^{\infty} p^k \left( k \log_2 p + \log_2 (1-p) \right) \]

\[ = -(1-p) \log_2 (1-p) \left( \sum_{k=0}^{\infty} p^k \right) - (1-p) \log_2 p \left( \sum_{k=0}^{\infty} k p^k \right) \]

\[ = -(1-p) \log_2 (1-p) - p \log_2 p \]

\[ = \frac{1}{1-p} H_b(p) = \frac{1}{1 - e^{-a\Delta}} H_b(e^{-a\Delta}) \]

where

\[ H_b(x) = -x \log_2 x - (1-x) \log_2 (1-x) \]

is the binary entropy function.

When looking at the extrema \( \Delta \to \infty \) and \( \Delta \to 0 \), we obtain

\( \Delta \to \infty \implies p \to 0 \implies R \to 0 \)

\( \Delta \to 0 \implies p \to 1 \implies R \to \infty \)

The binary entropy function is equal to 1 if and only if \( p = 0.5 \). Then, we obtain a rate of

\[ R(p = 0.5) = \frac{1}{1 - 0.5} H_b(0.5) = 2 \cdot 1 = 2 \]

Consequently, the step size \( \Delta \) that corresponds to a rate of 2 bit per sample is given by

\[ p = e^{-a\Delta} = \frac{1}{2} \implies \Delta(R = 2) = \frac{\ln 2}{a} \]

We can map the quantization indices \( k \) to binary strings \( b_k \) using a simple unary code as shown in the following table.

<table>
<thead>
<tr>
<th>( k )</th>
<th>( p_k )</th>
<th>( b_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>((1-p))</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>(p(1-p))</td>
<td>01</td>
</tr>
<tr>
<td>2</td>
<td>(p^2(1-p))</td>
<td>001</td>
</tr>
<tr>
<td>3</td>
<td>(p^3(1-p))</td>
<td>0001</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>0</td>
<td>(p^k(1-p))</td>
<td>{0}_k1</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
</tbody>
</table>
For the quantization index $k$, we first transmit $k$ zeros followed by a one. The shortest codeword consists (for quantization index 0) consists of a single 1.

The pmf for all binary symbols is the same, namely $\{p, 1-p\}$. We can use binary arithmetic coding with a single pmf for virtually achieving the entropy of the quantization indices. If the decay parameter $a$ of the pmf is not given, it can be estimated during coding using a single adaptive context model.

For $p = 0.5$, which corresponds to a quantization step size of $\frac{\ln 2}{a}$ and a rate of 2 bits per symbol, a particularly simple code is obtained. Here, the binary strings can be directly used as codewords, without arithmetic coding.

(b) Derive a formula for the optimal reconstruction levels $s_k'$, for MSE distortion, as function of the quantization step size $\Delta$ and as function of the lower interval borders $u_k$ and the probability $p = e^{-a\Delta}$.

Consider the difference $s_k' - u_k$ for the extremum $\Delta \to 0$.

Solution:
The optimal levels $s_k'$ are given by the centroid condition

$$s_k' = \frac{\int_{u_k}^{u_{k+1}} x f(x) \, dx}{\int_{u_k}^{u_{k+1}} f(x) \, dx} = \frac{1}{p_k} \int_{u_k}^{u_{k+1}} ax e^{-ax} \, dx$$

$$= \frac{1}{p_k} \left[ -e^{-ax} \left( \frac{1}{a} + x \right) \right]_{u_k}^{u_{k+1}}$$

$$= \frac{1}{p_k} \left( e^{-a\Delta} \left( \frac{1}{a} + k\Delta \right) - e^{-a\Delta} \left( \frac{1}{a} + (k+1)\Delta \right) \right)$$

$$= \frac{1}{p_k} \left( e^{-a\Delta} \frac{1}{a} + k\Delta - e^{-a\Delta} \left( \frac{1}{a} + k\Delta - \Delta e^{-a\Delta} \right) \right)$$

$$= \frac{1}{p_k} \left( e^{-a\Delta} \frac{1}{a} + k\Delta - e^{-a\Delta} \left( \frac{1}{a} + k\Delta - \Delta e^{-a\Delta} \right) \right)$$

$$= \frac{1}{a} + k\Delta - \Delta e^{-a\Delta} = k\Delta + \Delta \left( \frac{1}{a\Delta} - \frac{e^{-a\Delta}}{1-e^{-a\Delta}} \right)$$

$$= u_k + \Delta b$$

with

$$b = \frac{1 - e^{-a\Delta} - (a\Delta) e^{-a\Delta}}{(a\Delta) (1 - e^{-a\Delta})} = \frac{1 - p + p \ln p}{(p - 1) \ln p}$$

The offset between the reconstruction levels and the lower interval boundaries is constant for all quantization intervals. For a given exponential pdf, it depends only on the quantization step size.

In the following, we look at the extremum $p \to 1$ (or $\Delta \to 0$). Using
L'Hôpital’s rule, we have
\[
\lim_{p \to 1} b = \lim_{p \to 1} \frac{1 - p + p \ln p}{(p - 1) \ln p} = \lim_{p \to 1} \frac{0 - 1 + p \frac{1}{p} + 1 \ln p}{(p - 1) \frac{1}{p} + 1 \ln p} = \lim_{p \to 1} \frac{\ln p}{p + \ln p} = \lim_{p \to 1} \frac{\frac{p}{p^2} + \frac{1}{p} + 1}{p + 1} = \lim_{p \to 1} \frac{1}{p + 1} = \frac{1}{2}
\]

And for the difference \(s'_k - u_k\), we obtain
\[
\lim_{p \to 1} s'_k - u_k = \Delta \lim_{p \to 1} b = \frac{\Delta}{2}
\]

For high rates (small quantization step sizes), the reconstruction values lie in the middle of the quantization intervals.

(c) Is the obtained quantizer an optimal entropy-constrained scalar quantizer?
What is the relationship between the Lagrange parameter \(\lambda\) and the quantization step size \(\Delta\)? And what is the relationship between the Lagrange parameter \(\lambda\) and probability \(p = e^{-\Delta}\)?

Solution:
For optimal entropy-constrained scalar quantizers, two conditions have to be fulfilled, the centroid condition for the reconstruction levels,
\[
s'_k = \frac{\int_{u_k}^{u_{k+1}} x f(x) \, dx}{\int_{u_k}^{u_{k+1}} f(x) \, dx},
\]
and the condition for the decision thresholds,
\[
u_k = \frac{1}{2} (s_k + s_{k-1}) + \frac{\lambda \ell_k - \ell_{k-1}}{2 s_k - s_{k-1}},
\]
where \(\lambda\) is the Lagrange parameter that determines the operation point and \(\ell_k\) are the optimal codeword length for the intervals \(k\) given by
\[
\ell_k = -\log_2 p_k = -\log_2 \left( \int_{u_k}^{u_{k+1}} f(x) \, dx \right)
\]
The centroid condition is fulfilled by setting the reconstruction levels as described above. Hence, we only have to check the condition for the decision thresholds.
The difference between two successive reconstruction levels \(s'_{k-1}\) and \(s'_k\), with \(k > 0\), is given by
\[
s'_k - s'_{k-1} = (k + d)\Delta - (k - 1 + d)\Delta = \Delta
\]
It is constant and equal to the quantization step size. The difference between the optimal codeword lengths $\ell_{k-1}$ and $\ell_k$ for two successive quantization intervals, with $k > 0$, is given by

$$\ell_k - \ell_{k-1} = - \log_2 p_k + \log_2 p_{k-1} = - \log_2 \left( \frac{p_k}{p_{k-1}} \right)$$

$$= - \log_2 \left( \frac{p_k (1 - p)}{p_{k-1} (1 - p)} \right) = - \log_2 p = - \log_2 \left( e^{-a\Delta} \right)$$

$$= \frac{a}{\ln 2} \Delta$$

Hence, the condition for the decision thresholds is given by

$$u_k = \frac{1}{2} \left( k\Delta + \Delta b + \Delta - \Delta + \Delta b \right) + \frac{\lambda}{2} \frac{a \Delta}{\Delta \ln 2}$$

$$= k\Delta + \Delta \left( b - \frac{1}{2} \right) + \frac{\lambda}{2} \frac{a \Delta}{\ln 2}$$

It is given that $u_k = k\Delta$. Consequently, the condition on the decision thresholds is fulfilled if

$$\lambda = \frac{2 \Delta \ln 2}{a} \left( \frac{1}{2} - b \right)$$

$$= \frac{2 \Delta \ln 2}{a} \left( \frac{1}{2} - \frac{1 - e^{-a\Delta} - a\Delta e^{-a\Delta}}{a\Delta (1 - e^{-a\Delta})} \right)$$

$$= \frac{2 \Delta \ln 2}{a} \cdot \frac{a\Delta - a\Delta e^{-a\Delta} - 2 + 2 e^{-a\Delta} + 2a\Delta e^{-a\Delta}}{2a\Delta (1 - e^{-a\Delta})}$$

$$= \frac{\ln 2}{a^2} \cdot \frac{a\Delta + a\Delta e^{-a\Delta} - 2 + 2 e^{-a\Delta}}{1 - e^{-a\Delta}}$$

$$= \frac{\ln 2}{a^2} \cdot \frac{a\Delta (1 + e^{-a\Delta}) - 2 (1 - e^{-a\Delta})}{1 - e^{-a\Delta}}$$

$$= \frac{\ln 2}{a} \cdot \left( \frac{1 + e^{-a\Delta}}{1 - e^{-a\Delta}} - \frac{2}{a} \right)$$

For the extreme $\Delta \to 0$, the Lagrange parameter approaches

$$\lim_{\Delta \to 0} \lambda = \frac{\ln 2}{a} \cdot \left( \lim_{\Delta \to 0} \frac{\Delta + \Delta e^{-a\Delta}}{1 - e^{-a\Delta}} \right) - \frac{2 \ln 2}{a^2}$$

$$= \frac{\ln 2}{a} \cdot \left( \lim_{\Delta \to 0} \frac{1 + e^{-a\Delta} - a\Delta e^{-a\Delta}}{a e^{-a\Delta}} \right) - \frac{2 \ln 2}{a^2}$$

$$= \frac{\ln 2}{a} \cdot \left( \lim_{\Delta \to 0} \frac{1 + 1 - 0 \cdot 1}{a \cdot 1} \right) - \frac{2 \ln 2}{a^2} = \frac{\ln 2}{a} \cdot \frac{2}{a} - \frac{2 \ln 2}{a^2} = 0$$

Here, we used L'Hôpital's rule in the first step.

For large step sizes, $e^{-a\Delta} \ll 1$, we obtain

$$\lim_{\Delta \to \infty} \lambda = \frac{\ln 2}{a} \cdot \Delta - \frac{2 \ln 2}{a^2}$$
Hence, for large step sizes the Lagrange parameter $\lambda$ grow approximately linearly with the quantization step size $\Delta$ and it approaches infinity if the step size approaches infinity.

Since $\lambda$ is a contiously growing function of $\Delta$ and approaches 0 and infinity at the extrema $\Delta \to 0$ and $\Delta \to \infty$, respectively, each value of $\lambda$ corresponds to a particular step size $\Delta$.

Hence, for each Lagrange parameter $\lambda$ we can find a quantization step size $\Delta$ for which the condition on the decision thresholds is fulfilled. Consequently, the uniform threshold quantizer with reconstruction levels at the centroids of the quantization intervals is an optimal entropy-constrained quantizer for the exponential pdf.

The relationship between $\lambda$ and $\Delta$ is given by

$$\lambda = \frac{\ln 2}{a} \cdot \left( \Delta \frac{1 + e^{-a\Delta}}{1 - e^{-a\Delta}} - \frac{2}{a} \right)$$

By using $p = e^{-ax}$, we obtain the following relationship between $\lambda$ and $p$

$$\lambda = -\frac{\ln 2}{a^2} \cdot \left( \frac{1 + p}{1 - p} \ln p + 2 \right)$$

In the following, the relationships between $\lambda$ and $\Delta$ and $\lambda$ and $p$ are illustrated.
(d) Determine the distortion in dependence of the quantization step size for the developed quantizer.

**Hint:** For $|a| < 1$:

$$\sum_{k=0}^{\infty} k^2 a^k = \frac{a (1 + a)}{(1 - a)^3}$$

**Solution:**

The distortion is given by

$$D = \sum_{k=0}^{\infty} \int_{u_k}^{u_{k+1}} (x - s_k')^2 f(x) \, dx$$

$$= \sum_{k=0}^{\infty} \int_{u_k}^{u_{k+1}} x^2 f(x) \, dx - 2s_k' \int_{u_k}^{u_{k+1}} xf(x) \, dx + s_k'^2 \int_{u_k}^{u_{k+1}} f(x) \, dx$$

The probability that a value $X$ falls inside the $k$-th quantization interval is given by

$$p_k = \int_{u_k}^{u_{k+1}} f(x) \, dx$$

And, if the centroid condition is fulfilled, the reconstruction levels are given by

$$s_k' = \frac{\int_{u_k}^{u_{k+1}} x f(x) \, dx}{\int_{u_k}^{u_{k+1}} f(x) \, dx}$$
Hence, we can write
\[
D = \sum_{k=0}^{\infty} \int_{u_k}^{u_{k+1}} x^2 f(x) \, dx - 2s'_k \int_{u_k}^{u_{k+1}} x f(x) \, dx + s''_k \int_{u_k}^{u_{k+1}} f(x) \, dx
\]
\[
= \sum_{k=0}^{\infty} \int_{u_k}^{u_{k+1}} x^2 f(x) \, dx - 2s'_k \cdot (s'_k p_k) + s''_k \cdot p_k
\]
\[
= \int_{0}^{\infty} x^2 f(x) \, dx - \sum_{k=0}^{\infty} s''_k p_k
\]
\[
= \left[ -e^{-ax} \left( \frac{2}{a^2} + \frac{2x}{a} + x^2 \right) \right]_{0}^{\infty} - \sum_{k=0}^{\infty} s''_k p_k
\]
\[
= \frac{2}{a^2} - \sum_{k=0}^{\infty} s''_k p_k
\]

The probabilities \( p_k \) have been already derived,
\[
p_k = (e^{-a\Delta})^k (1 - e^{-a\Delta})
\]

When deriving the optimal reconstruction levels, we obtained
\[
s'_k = \frac{1}{a} + k\Delta - \frac{\Delta e^{-a\Delta}}{1 - e^{-a\Delta}}
\]

Using \( p = e^{-a\Delta} \), we can also write
\[
p_k = p^k (1-p)
\]
\[
s'_k = \frac{1}{a} + k\Delta - \frac{p\Delta}{1-p}
\]

For the product \( s'_k p_k \), we get
\[
s'_k p_k = p^k (1-p) \left( \frac{1}{a} + k\Delta - \frac{p\Delta}{1-p} \right)
\]
\[
= p^k \cdot \left( (1-p) \left( \frac{1}{a} + k\Delta \right) - p\Delta \right)
\]

For the term \( s''_k p_k \), we then obtain
\[
s''_k p_k = p^k \left( (1-p) \left( \frac{1}{a} + k\Delta \right) - p\Delta \right) \left( \frac{1}{a} + k\Delta - \frac{p\Delta}{1-p} \right)
\]
\[
= p^k \left[ (1-p) \left( \frac{1}{a} + k\Delta \right)^2 - 2p\Delta \left( \frac{1}{a} + k\Delta \right) + \frac{p^2 \Delta^2}{1-p} \right]
\]
\[
= p^k \left[ \frac{1-p}{a^2} + \frac{2k\Delta (1-p)}{a} + k^2 \Delta^2 (1-p) - \frac{2p\Delta}{a} - 2kp\Delta^2 + \frac{p^2 \Delta^2}{1-p} \right]
\]

18
By reordering the terms, we obtain

\[
s_k^2 p_k = k^2 p^k \cdot \left( \Delta^2 (1 - p) + k p^k \cdot \frac{2\Delta(1 - p)}{a} - 2p\Delta^2 \right) + p^k \cdot \left( \frac{1 - p}{a^2} - \frac{2p \Delta}{a} + \frac{p^2 \Delta^2}{1 - p} \right)
\]

And for the sum over all term \( s_k^2 p_k \), we get

\[
\sum_{k=0}^{\infty} s_k^2 p_k = \left( \Delta^2 (1 - p) \right) \cdot \left( \sum_{k=0}^{\infty} k^2 p^k \right) + \left( \frac{2\Delta(1 - p)}{a} - 2p\Delta^2 \right) \cdot \left( \sum_{k=0}^{\infty} k p^k \right) + \left( \frac{1 - p}{a^2} - \frac{2p \Delta}{a} + \frac{p^2 \Delta^2}{1 - p} \right) \cdot \left( \sum_{k=0}^{\infty} p^k \right)
\]

\[
= \left( \Delta^2 (1 - p) \right) \cdot \left( \frac{p(1 + p)}{(1 - p)^2} \right) + \left( \frac{2\Delta(1 - p)}{a} - 2p\Delta^2 \right) \cdot \left( \frac{p}{(1 - p)^2} \right) + \left( \frac{1 - p}{a^2} - \frac{2p \Delta}{a} + \frac{p^2 \Delta^2}{1 - p} \right) \cdot \left( \frac{1}{1 - p} \right)
\]

\[
= \left( \frac{p\Delta^2}{(1 - p)^2} + \frac{p^2 \Delta^2}{(1 - p)^2} \right) + \left( \frac{2p \Delta}{a(1 - p)} - \frac{2p^2 \Delta^2}{(1 - p)^2} \right) + \left( \frac{1}{a^2} - \frac{2p \Delta}{a(1 - p)} + \frac{p^2 \Delta^2}{(1 - p)^2} \right)
\]

\[
= \frac{1}{a^2} + \frac{p\Delta^2}{(1 - p)^2}
\]

And for the distortion, we obtain

\[
D = \frac{2}{a^2} - \left( \frac{1}{a^2} + \frac{p\Delta^2}{(1 - p)^2} \right)
\]

\[
= \frac{1}{a^2} \left( 1 - \frac{(a\Delta)^2 p}{(1 - p)^2} \right)
\]

\[
= \sigma^2 \left( 1 - \frac{(a\Delta)^2 e^{-a\Delta}}{(1 - e^{-a\Delta})^2} \right)
\]

Here, we used the fact that the variance for the exponential pdf is equal to \( \sigma^2 = 1/a^2 \).
(e) Compare the operational rate-distortion function (using a parametric formulation) to the Shannon lower bound.
Plot the operational rate-distortion function and the Shannon lower bound in one diagram for rates from 0 to 5 bit per sample.
Plot the ratio of the distortion for the operational rate-distortion function and the Shannon lower bound for rates from 0 to 5 bit per sample.

Solution:
The Shannon lower bound for the exponential pdf (and MSE distortion) has been derived in a previous exercise and is given by

$$D_{SLB}(R) = \sigma^2 \cdot \frac{e}{2\pi} \cdot 2^{-2R}$$

For the operational rate-distortion function for optimal ECSQ we obtained a parametric formulation. With $t = a\Delta$, it can be written as

$$R(t) = -e^{-t} \log_2(e^{-t}) - \left(1 - e^{-t}\right) \log_2\left(1 - e^{-t}\right)$$

$$D(t) = \sigma^2 \left(1 - \frac{t^2 e^{-t}}{(1 - e^{-t})^2}\right)$$

The ratio between the distortion can be written as

$$\rho(t) = \frac{D(t)}{D_{SLB}(R(t))} = \frac{2\pi}{e} \left(1 - \frac{t^2}{(1 - e^{-t})^2}\right) 2^{2R(t)}$$

In the following figures, the operational rate-distortion function for ECSQ and the Shannon lower bound are compared in the linear and logarithmic domain. Furthermore, the ratio of the distortion (for the same rate) is plotted over the rate in both linear and logarithmic domain.

For high rates, where the Shannon lower bound coincides with the rate distortion function, the distortion for the operational rate-distortion function is by the factor $e\pi/6 \approx 1.42$ (or 1.53 dB) larger than the fundamental bound. It can only be further improved if vector quantization is used instead of scalar quantization.
distortion ratio (ECSQ vs SLB)
rate [bit/sample]
\[
\frac{\pi e}{6} \quad \frac{D_{ECSQ}}{D_{SLB}}
\]

\begin{align*}
\text{distortion ratio (ECSQ vs SLB)} \quad & \quad \text{rate [bit/sample]} \\
& \quad \frac{\pi e}{6} \\
& \quad \frac{D_{ECSQ}}{D_{SLB}} \\
& \quad 1.53 \text{ dB} \\
& \quad 10 \log_{10} \left( \frac{D_{ECSQ}}{D_{SLB}} \right)
\end{align*}
4. Consider a discrete Markov process $X = \{X_n\}$ with the symbol alphabet $A_X = \{0, 2, 4, 6\}$ and the conditional pmf

$$p_{X_n|X_{n-1}}(x_n|x_{n-1}) = \begin{cases} a : x_n = x_{n-1} \\ \frac{1}{3}(1-a) : x_n \neq x_{n-1} \end{cases},$$

for $x_n, x_{n-1} \in A_X$. The parameter $a$, with $0 < a < 1$, is a variable that specifies the probability that the current symbol is equal to the previous symbol. For $a = 1/4$, our source $X$ would be i.i.d.

Given is a quantizer of size 2 with the reconstruction levels $s'_0 = 1$ and $s'_1 = 5$ and the decision threshold $u_1 = 3$.

(a) Assume optimal entropy coding using the marginal probabilities of the quantization indices and determine the rate-distortion point of the quantizer.

**Solution:**

First, we determine the marginal pmf $p_X(x)$. By reasons of symmetry, it can easily be seen that the marginal pmf is given by

$$p_X(x) = \frac{1}{4}$$

for all symbols $x \in A_X$. In a more rigorous way, this can also be derived by

$$p_X(x) = \sum_{y \in A_X} p_{X_n|X_{n-1}}(x|y) \cdot p_X(y)$$

$$= a \cdot p_X(x) + \frac{1 - a}{3} \sum_{y \notin x \in A_X} p_X(y)$$

$$= a \cdot p_X(x) + \frac{1 - a}{3} \sum_{y \notin x \in A_X} p_X(y)$$

$$= a \cdot p_X(x) + \frac{1 - a}{3} \times (1 - p_X(x))$$

$$= \left(\frac{4a - 1}{3} \right) \cdot p_X(x) + \frac{1}{3} - a$$

$$3p_X(x) = (4a - 1) p_X(x) + (1 - a)$$

$$\frac{4 - 4a}{p_X(x)} = 1 - a$$

$$p_X(x) = \frac{1 - a}{4 - 4a} = \frac{1}{2}$$

Let $Y = \{Y_n\}$ denote the random sequence of quantization indices, with $Y_n \in \{0, 1\}$. With $C_k$ denoting the quantization cell with the reconstruction value $s'_k$, the marginal pmf $p_Y(y)$ can be written as

$$p_Y(k) = \sum_{x \in C_k} p_X(x) = 2 \cdot \frac{1}{4} = \frac{1}{2}$$

For the distortion $D$, we then obtain

$$D = \sum_{k=0}^1 \sum_{x \in C_k} (x - s'_k)^2 p_X(x) = 4 \cdot 1^2 \cdot \frac{1}{4} = 1$$

23
And the rate for optimal entropy coding is given by
\[ R = - \sum_{k=0}^{1} p_Y(k) \log_2 p_Y(k) = -2 \cdot \frac{1}{2} \cdot \log_2 \left( \frac{1}{2} \right) = 1 \]

An optimal entropy coding can be realized by a simple code that assigns a bit equal to 0 to the quantization index 0 and a bit equal to 1 to the quantization index 1 (or vice versa).

The rate-distortion point for the quantizer with scalar entropy coding is given by \( R = 1 \) and \( D = 1 \).

(b) Can the overall quantizer performance be improved by applying conditional entropy coding (e.g., using arithmetic coding with conditional probabilities)? How does it depend on the parameter \( a \)?

**Solution:**
The distortion is only dependent on the decision threshold and reconstruction levels. Hence, it does not change by modifying the entropy coding for the quantization indices and is \( D = 1 \).

For conditional entropy coding, the minimum achievable rate is given by the conditional entropy
\[ R = - \sum_{i=0}^{1} \sum_{j=0}^{1} p_{Y_n | Y_{n-1}}(i, j) \log_2 p_{Y_n | Y_{n-1}}(i | j) \]

For determining the rate, we first have to calculate the joint and conditional pmf for the quantization indices. The joint pmf is given by
\[ p_{Y_n Y_{n-1}}(i, j) = \sum_{x \in C_i} \sum_{y \in C_j} p_X(x, y) \]
\[ = \sum_{x \in C_i} \sum_{y \in C_j} p_{X_n | X_{n-1}}(x | y) p_X(y) \]

For the case \( i \neq j \), we obtain
\[ p_{Y_n Y_{n-1}}(i, j) = 2 \cdot 2 \cdot \frac{1 - a}{3} \cdot \frac{1}{4} = \frac{1 - a}{3} , \]
and for the case \( i = j \), we have
\[ p_{Y_n Y_{n-1}}(i, i) = 2 \cdot \left( a + \frac{1 - a}{3} \right) \cdot \frac{1}{4} = \frac{1}{6} \left( 6a + 1 - a \right) = \frac{1 + 2a}{6} \]

Hence, the joint pmf is
\[ p_{Y_n Y_{n-1}}(i, j) = \begin{cases} \frac{1}{3}(1 + 2a) & : i = j \\ \frac{1}{3}(1 - a) & : i \neq j \end{cases} \]

And for the conditional pmf, we obtain
\[ p_{Y_n | Y_{n-1}}(i | j) = \frac{p_{Y_n Y_{n-1}}(i, j)}{p_Y(j)} = \begin{cases} \frac{1}{3}(1 + 2a) & : i = j \\ \frac{1}{3}(1 - a) & : i \neq j \end{cases} \]
Consequently, the minimum rate achievable by conditional entropy coding is

\[
R = - \sum_{i=0}^{1} \sum_{j=0}^{1} p_{Y_n} \cdot p_{Y_{n-1}}(i, j) \cdot \log_2 p_{Y_n | Y_{n-1}}(i | j)
\]

\[
= -2 \left( \frac{1 + 2a}{6} \log_2 \left( \frac{1 + 2a}{3} \right) + \frac{1 - a}{3} \log_2 \left( \frac{2 - 2a}{3} \right) \right)
\]

\[
= - \left( \frac{1 + 2a}{3} \right) \log_2 \left( \frac{1 + 2a}{3} \right) - \left( \frac{2 - 2a}{3} \right) \log_2 \left( \frac{2 - 2a}{3} \right)
\]

\[
= H_b \left( \frac{1 + 2a}{3} \right),
\]

where \( H_b(p) \) represents the binary entropy function.

The rate \( R \) is maximized (equal to 1) if the argument of the binary entropy function is equal to 1/2, i.e., if

\[
\frac{1}{2} = \frac{1 + 2a}{3} \quad \Rightarrow \quad a = \frac{1}{4}
\]

In this case, the source samples are independent.

The larger the absolute difference \( |a - \frac{1}{4}| \), the more dependent successive source samples are and the lower the rate becomes. If \( a \) approaches 1 (i.e., if the difference \( |a - \frac{1}{4}| \) approaches the maximum of \( \frac{3}{4} \)), the rate approaches 0.

In the following diagram, the rate for optimal entropy coding using the conditional pmf of the quantization indices is plotted over the probability parameter \( a \).

In summary, we note that the performance of scalar quantizers for sources with memory can be improved if we apply entropy coding techniques that employ conditional (or joint) probabilities.
5. Calculate the gain of optimal 2-dimensional vector quantization relative to optimal scalar quantization for high rates on the example of a uniform pdf.

*Hint:* For high rates, border effects can be neglected. It can be assumed that the signal space for which the pdf is non-zero is completely filled with regular quantization cells.

*Solution:*

For a uniform pdf the optimal scalar quantizer is a uniform threshold quantizer with reconstruction levels at the center of the quantization intervals. This quantizer represents both, a Lloyd quantizer and an entropy-constrained Lloyd quantizer.

Without loss of generality, we write the uniform pdf according to

\[
 f(x) = \begin{cases} 
 \frac{1}{2a} & : |x| \leq a \\
 0 & : |x| > a 
\end{cases},
\]

where \( a \) determines the width of the distribution.

Let \( N \) denote the number of quantization cells. The width of the quantization cells is given by

\[
 \Delta = \frac{2a}{N}
\]

and the rate for the quantizer (all cells have the same probability) is given by

\[
 R = \log_2 N
\]

For the distortion \( D_k \) inside one interval, we obtain

\[
 D_k = \int_{s_k'}^{s_{k}'+\frac{\Delta}{2}} (x-s'_k)^2 f(x) \, dx = \frac{1}{2a} \int_{s_k'-\frac{\Delta}{2}}^{s_k'+\frac{\Delta}{2}} (x-s'_k)^2 \, dx
\]

\[
 = \frac{1}{2a} \left[ \frac{t^3}{3} \right]_{s_k'-\frac{\Delta}{2}}^{s_k'+\frac{\Delta}{2}} = \frac{1}{6a} \left[ \frac{\Delta}{8} + \frac{\Delta}{8} \right] = \frac{1}{24a} \Delta^3 = \frac{1}{24a} \left( \frac{2a}{N} \right)^3
\]

For the overall distortion, we then obtain

\[
 D = \sum_{k=0}^{N-1} D_k = N \cdot D_k = N \cdot \frac{a^2}{3 N^3} = \frac{a^2}{3 N^2}
\]

By using \( R = \log_2 N \) (and thus \( N = 2^R \)), finally obtain the operational rate-distortion function for scalar quantization

\[
 D_1(R) = \frac{a^2}{3} \cdot 2^{-2R}
\]
We now consider vector quantization with 2 dimensions. The joint pdf in 2 dimensions is given by

\[
f(x, y) = \begin{cases} 
\frac{1}{4a^2} & : |x| \leq a \land |y| \leq a \\
0 & : |x| > a \lor |y| > a 
\end{cases}
\]

If we ignore border effects, the optimal quantization cells in two dimensions are regular hexagons, as they provide the densest packing. At the borders of the non-zero range \([-a \cdots a] \times [-a \cdots a]\) of our joint pdf \(f(x, y)\) the shapes would be different, but these effects can be ignored if we consider high rates (i.e., a large number of quantization cells). For the uniform pdf, a quantizer with hexagonal cells and reconstruction values inside the center is a Lloyd quantizer as well as an entropy-constrained Lloyd quantizer.

In the high rate case, the number of quantization cells can be approximated by

\[
N = \frac{4a^2}{A_{\text{hexagon}}},
\]

where \(A_{\text{hexagon}}\) represents the area of a hexagonal quantization cell.

For determining the area of a hexagonal cell, we can divide it into 6 equilateral triangles as shown in the figure above. Let \(b\) denote the length of a side of the hexagon and let \(h\) denote the height of the triangles. Then, we have

\[
A_{\text{hexagon}} = 6 \cdot A_{\text{triangle}} = 6 \cdot \frac{1}{2} \cdot h \cdot b = 3 \cdot (b \cdot \cos(30^\circ)) \cdot b = 3 \cdot (b \cdot \frac{\sqrt{3}}{2}) \cdot b = \frac{3\sqrt{3}}{2} b^2
\]

The number of quantization cells becomes

\[
N = \frac{4a^2}{\frac{3\sqrt{3}}{2} b^2} = \frac{8\sqrt{3}}{9} \frac{a^2}{b^2}
\]
For calculating the distortion $D_k$ inside a quantization cell, we further divide one of the triangles (the top center one) into two right triangles and calculate the distortion of one of these triangles (the one to the right). The line at the right side of the triangle we consider is given by

$$y = \tan(60^\circ) \cdot x = \sqrt{3}x$$

The distortion for a quantization cell is

$$D_k = 12 \cdot D_{\text{triangle}} = 12 \int_{x=0}^{b/2} \int_{y=0}^{h} r^2 \cdot f(x, y) \, dy \, dx$$

$$= \frac{12}{4a^2} \int_{x=0}^{b/2} \int_{y=0}^{h} (x^2 + y^2) \, dy \, dx$$

$$= \frac{12}{4a^2} \left( \int_{x=0}^{b/2} \left( \int_{y=0}^{h} y^2 \, dy \right) \, dx + \int_{x=0}^{b/2} x^2 \left( \int_{y=0}^{h} \, dy \right) \, dx \right)$$

$$= \frac{12}{4a^2} \left( \int_{x=0}^{b/2} \left[ \frac{h^3}{3} - \frac{3\sqrt{3}x^3}{3} \right] \, dx + \int_{x=0}^{b/2} x^2 \left[ h - \sqrt{3}x \right] \, dx \right)$$

$$= \frac{12}{4a^2} \left( \int_{x=0}^{b/2} \left[ \frac{h^3}{3} \right] \, dx - \sqrt{3} \int_{x=0}^{b/2} x^3 \, dx + h \int_{x=0}^{b/2} x^2 \, dx - \sqrt{3} \int_{x=0}^{b/2} x^3 \, dx \right)$$

$$= \frac{12}{4a^2} \left( \frac{h^3}{3} \left( \frac{b}{2} \right) - \sqrt{3} \left( \frac{b^4}{4 \cdot 16} \right) + h \left( \frac{h^3}{3 \cdot 8} \right) - \sqrt{3} \left( \frac{b^4}{4 \cdot 16} \right) \right)$$

$$= \frac{12}{4a^2} \left( \frac{bh^3}{6} - \frac{\sqrt{3}b^4}{64} + \frac{b^3h^3}{24} - \frac{\sqrt{3}b^4}{64} \right)$$

With $h = b \cos(30^\circ) = \frac{\sqrt{3}}{2} b$, we obtain

$$D_k = \frac{12}{4a^2} \left( \frac{3\sqrt{3}b^4}{6 \cdot 8} - \frac{\sqrt{3}b^4}{64} + \frac{\sqrt{3}b^4}{24 \cdot 2} - \frac{\sqrt{3}b^4}{64} \right)$$

$$= \frac{12\sqrt{3}b^4}{4a^2} \left( \frac{3}{48} - \frac{1}{64} + \frac{1}{48} - \frac{1}{64} \right) = \frac{3\sqrt{3}b^4}{a^2} \left( \frac{4}{48} - \frac{2}{64} \right)$$

$$= \frac{3\sqrt{3}}{a^2} \left( \frac{1}{12} - \frac{1}{32} \right) = \frac{3\sqrt{3}}{a^2} \cdot \frac{8 - 3}{96} = \frac{5\sqrt{3}b^4}{32} \cdot \frac{81}{N^2} = \frac{64}{27} \cdot \frac{a^4}{N^2}$$

Using the previously derived relation

$$N = \frac{8\sqrt{3}}{9} \frac{a^2}{b^2} \quad \Rightarrow \quad b^4 = \frac{64 \cdot 3 \cdot a^4}{81 \cdot N^2} = \frac{64}{27} \cdot \frac{a^4}{N^2}$$
we get
\[ D_k = \frac{5\sqrt{3}}{32} \frac{1}{a^2} \cdot \frac{64}{27} \frac{a^4}{N^2} = \frac{10\sqrt{3}}{27} \frac{a^2}{N^2} \]

The overall distortion for 2 samples is then
\[ D(2) = \sum_{k=0}^{N-1} D_k = N \cdot D_k = \frac{10\sqrt{3}}{27} \frac{a^2}{N} \]

And for the distortion \( D \) per sample, we obtain
\[ D = \frac{D(2)}{2} = \frac{5\sqrt{3}}{27} \frac{a^2}{N} \]

The rate for 2 samples is given by
\[ R(2) = \log_2 N \]

The rate per sample is then
\[ R = \frac{R(2)}{2} = \frac{1}{2} \log_2 N \]

yielding the following expression for the number of samples
\[ N = 2^{2R} \]

Inserting this expression into the expression for the distortion per sample yields the operational rate-distortion function for 2-dimensional vector quantization
\[ D_2(R) = \frac{5\sqrt{3}}{27} \frac{a^2}{2^{-2R}} \cdot 2^{-2R} \]

For the ratio of the distortions for 2-d vector quantization and scalar quantization at the same rate, we obtain
\[ \frac{D_2(R)}{D_1(R)} = \frac{5\sqrt{3}}{27} \frac{a^2}{3} \cdot 2^{-2R} = \frac{5 \cdot 3 \cdot \sqrt{3}}{27} = \frac{5\sqrt{3}}{9} \approx 0.962250 \]

The signal-to-noise ratio for high rates is improved by
\[ \Delta \rho = -10 \log_{10} \left( \frac{D_2(R)}{D_1(R)} \right) = 10 \log_{10} \left( \frac{3\sqrt{3}}{5} \right) \approx 0.167119 \text{ dB} \]

The increase in SNR that is obtained by increasing the quantizer dimension from 1 to 2 at high rates (and using optimal entropy-constrained quantizers) is approximately 0.17 dB. Hence, the difference to the rate-distortion curve has been reduced from 1.533 dB to 1.366 dB.

By further increasing the quantizer dimension, the distance to the rate-distortion function can be further reduced. Asymptotically, the rate-distortion function is achieved if the quantizer dimension approaches infinity.