Quantization

Robert M. Haralick

Computer Science, Graduate Center
City University of New York

Outline

Quantizing

- Each dimension (feature) has the same quantizer
- The number of quantizing levels for each dimension can be different
- Quantizers are independent of class
- Limitations
  - Determine the number N of observations for the smallest class
  - Determine the size M of memory that can used for the class conditional probability tables
  - $M \leq N/10$ The variety in the memory must be much smaller than the variety in the training data
  - The Memory size M times the number of classes K must satisfy $MK \leq$ available memory
- Once M is known, how to choose the number of quantizing levels for each dimension?

Quantizing

- 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 fewer 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

Simple Quantizer and Bins

- J dimensions
- L quantized values per dimension
- L^J 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 bin associated with 0 is the first bin
- The bin associated with $K - 1$ is the K^{th} bin

Quantizing Interval

Definition

The quantizing interval Q_k associated with the integer k is defined by

$$Q_k = \{x \mid q(x) = k\}$$

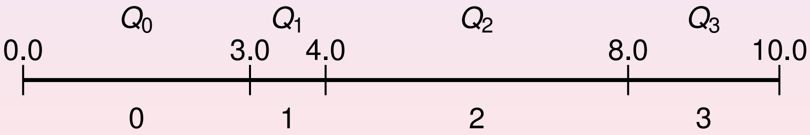


Table Lookup

- 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 quantized z is $a(q_1(z_1), \ldots, q_J(z_J))$
- $P(a(q_1(z_1), \ldots, q_J(z_J)) \mid c) = P(q_1(z_1), \ldots, q_J(z_J) \mid c)$

Entropy Definition

Definition

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

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

Entropy Meaning

- 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

- We need to estimate the entropy of the probability distribution in each dimension independent of class
- We need to do this to determine the number of quantizing levels being given to each dimension
- We do this by setting a large number N of quantizing levels, but not as large as the training set size
- $N \leq$ available memory
- We do this for one dimension at a time

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

$$\hat{H} = -\sum_{k=1}^{K} p_k log_2 p_k + \frac{n_0 - 1}{2Nlog_e 2}$$

G. Miller, *Note On The Bias Of Information Estimates*, In H. Quastler (Ed.) **Information theory in psychology II-B**, Free Press, Glencoe, IL, 1955, pp. 95-100.

Number of Quantizing Levels

- J dimensions
- Each observation is a tuple $z = (z_1, \ldots, z_J)$
- Each z_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

$$
\begin{aligned}
f_j &= \frac{\hat{H}_j}{\sum_{i=1}^{J} \hat{H}_i} \\
L_j &= M^{f_j} \\
\prod_{j=1}^{J} L_j &= \prod_{j=1}^{J} M^{f_j} \\
&= M^{\sum_{j=1}^{J} f_j} \\
&= M
\end{aligned}
$$

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

$$
\begin{aligned}
f_j &= \frac{\hat{H}_j}{\sum_{i=1}^{J} \hat{H}_i} \\
L_j &= \lceil M^{f_j} \rceil
\end{aligned}
$$

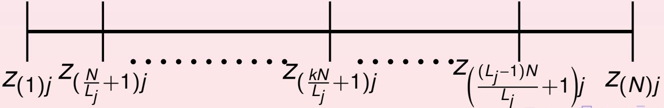
Now we have solved for the number of quantizing bins for each dimension.

How to Determine The Probabilities

- Memory size for each class conditional probability is M
- Real Valued Data
  - If the number of observations is N
  - Equal Interval Quantize each component to $N/10$ Levels
- Digitized Data
  - If each data item is I bits
  - Set the number of quantized levels to 2^I
- Determine the probability for each quantized level for each component
- Determine the entropy H_j for each component $j \in J$
- Set the number of quantized levels for component j to be $L_j = \lceil M^{f_j} \rceil$
- $f_j = \frac{\hat{H}_j}{\sum_{i=1}^{J} \hat{H}_i}$

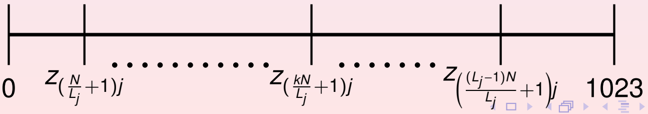
Initial Quantizing Interval Boundary

- The sample is $z_1, \ldots, z_N$
- Each tuple has J components
- Component j has L_j quantized levels
- 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:



Initial Quantizing Interval Boundary

- Equal Probability Quantizing
- The sample is $z_1, \ldots, z_N$
- Each tuple has J components
- Component j has L_j quantized levels
- 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 $N = 12$, Data is 10 bits, and $L_j = 4$.
- j^{th} component $z_{1j}, \ldots, z_{12j}$
- ordered values of j^{th} component: $z_{(1)j}, \ldots z_{(12)j}$
- $\frac{N}{L_j} + 1 = 4$
- $\frac{2N}{L_j} + 1 = 7$
- $\frac{3N}{L_j} + 1 = 10$

The quantizing intervals are:

$$[0, z_{(4)j})$$
$$[z_{(4)j}, z_{(7)j})$$
$$[z_{(7)j}, z_{(10)j})$$
$$[z_{(10)j}, 1023)$$

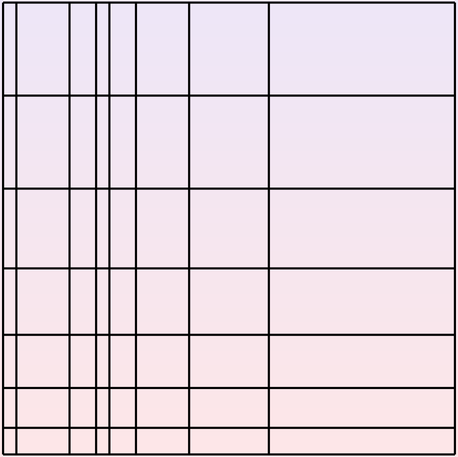
Initial Quantizing Interval Boundary

- 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: $k = 1, \ldots, L_j$
- The k^{th} quantizing interval $[c_{kj}, d_{kj})$ for the j^{th} component is defined by

For $k \in \{2, \ldots, L_j - 1\}$

$$\begin{aligned} c_{kj} &= z_{((k-1)N/L_j+1)j} \\ d_{kj} &= z_{(kN/L_j+1)j} \end{aligned}$$

Non-uniform Quantization



Maximum Likelihood Probability Estimation

- 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 $0, \ldots, M-1$
- The number of observations falling into bin m is t_m
- The maximum likelihood estimate of the probability for bin m is p_m

$$
\begin{aligned}
t_m &= \#\{n \mid a(q_1(z_{n1}), \ldots, q_J(z_{nJ})) = m\} \\
p_m &= \frac{t_m}{N}
\end{aligned}
$$

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 of each point in the volume by $\frac{k/N}{v}$

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 of each point in the volume by $\frac{k/N}{v}$

Smoothed Estimates

- If the sample size is not large enough, the MLE probability estimates may not be representative.
- bin smoothing
    - Let bin m have volume v_m and count t_m
    - Let $m_1, \ldots, m_I$ be the indexes of the I closest bins to bin m satisfying

$$\sum_{i=1}^{I} t_{m_i} \geq k \qquad \sum_{i=1}^{I-1} t_{m_i} < k$$

- $b_m = \sum_{i=1}^{I} t_{m_i}$
- $V_m^* = \sum_{i=1}^{I} v_{m_i}$
- Density of each point in bin m: $\alpha b_m / V_m^*$
- Set α so that the density integrates to 1

Smoothed Estimates

- Density of each point 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: $1 = \sum_{m=1}^{M} \alpha b_m v_m / V_m^*$
- $\alpha = \frac{1}{\sum_{m=1}^{M} b_m v_m / V_m^*}$
- $p_m = \frac{1}{\sum_{k=1}^{M} b_k v_k / V_k^*} b_m v_m / V_m^*$

Smoothed Estimates

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

If $v_m = v, \;\; m = 1, \ldots, M$, then $V_m^* = l_m v$

$$\begin{aligned} p_m &= \frac{b_m v / l_m v}{\sum_{k=1}^{M} b_k v / l_k v} \\ &= \frac{b_m / l_m}{\sum_{k=1}^{M} b_k / l_k} \end{aligned}$$

Optimization

- Fixed Sample Size N
- Sample $Z_1, \ldots, Z_N$
- Total number of bins M
- Calculate the quantizer
- Determine the probability for each bin
- Smooth the bin probabilities using smoothing k
- 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

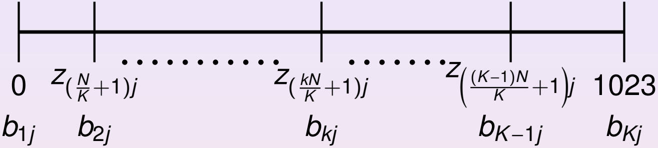
- Split the ground truthed sample into three parts
- Use the first part to calculate the quantizer and bin probabilities
- Calculate the discrete Bayes decision rule
- Apply the decision rule to the second part so that an unbiased estimate of the expected economic gain given the decision rule can be 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 for the expected gain for the optimization

Optimize The Probability Estimation Parameters

Once the parameters M and k have been optimized, the quantizer boundaries can be optimized.
Repeat until no change

- Use training data part 1
  - Choose a dimension
  - Choose a boundary
  - Change the boundary
  - Determine the adjusted probabilities
  - Determine the discrete Bayes decision rule
- Use training data part 2
  - Calculate the Expected Gain

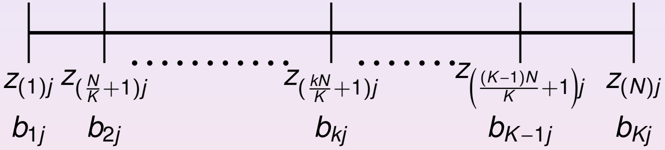
Optimize The Quantizer Boundaries



Repeat until no change

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

- Bayesians use the prior probability
  - Here prior probability is the prior probability of the bin
  - For each bin before we observe the data, the Bayesian must guess a prior density for the bin
  - What is the prior density for the bin being considered to be .135?
- MLE: start bin counters from 0
- Bayesian: start bin counters from β, $\beta > 0$
- Where does β come from?

The Observation

There are K bins. Each time an observation is made, the observation falls into exactly one of the K bins. The unknown 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,

$$
\begin{array}{c}
I_1 \text{ observations fall into bin } 1 \\
I_2 \text{ observations fall into bin } 2 \\
\vdots \\
I_K \text{ observations fall into bin } K
\end{array}
$$

Multinomial

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

$$P(I_1, \ldots, I_K \mid p_1, \ldots, p_K) = \frac{I!}{I_1! \cdots I_K!} p_1^{I_1} \cdots p_K^{I_K}$$

Bayesian Bin Probability

We have observed $I_1, \ldots, I_K$ 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 $P(d_k \mid I_1, \ldots, I_K)$

$K-1$ Simplex

To do this we will need to evaluate two integrals over the $K-1$-simplex

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

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

$K-1$ Simplex

Definition

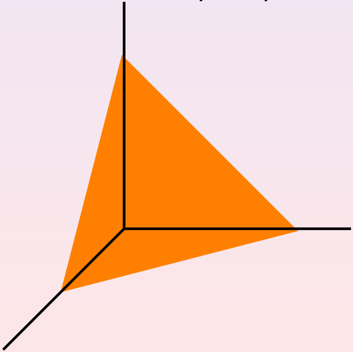
A $K-1$ Simplex is a $(K-1)$-dimensional polytope which is the convex hull of its K vertices.

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

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

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



Two Integrals

They are:

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

$$\int_{(q_1,\ldots,q_K)\in S} \prod_{k=1}^{K} q_k^{I_k} dq_1,\ldots,dq_K = \frac{\prod_{k=1}^{K} I_k!}{(I+K-1)!}$$

where

$$\sum_{k=1}^{K} I_k = I$$

Gamma Function

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

$$\Gamma(n) = (n-1)!$$

Derivation

The derivation goes as follows: $Prob(d_k \mid l_1, \ldots, l_K)$

$$= \frac{Prob(d_k, l_1, \ldots, l_K)}{Prob(l_1, \ldots, l_K)}$$

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

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

$$= \frac{\int_{(p_1,\ldots,p_K)\in S} \frac{\prod_{n=1}^K l_n!}{l!} \prod_{m=1}^K p_m^{l_m} p_k (K-1)! dp_1 \cdots dp_K}{\int_{(q_1\ldots,q_K)\in S} \frac{\prod_{n=1}^K l_n!}{l!} \prod_{m=1}^K q_m^{l_m} (K-1)! dq_1 \ldots dq_K}$$

Derivation

$$Prob(d_k \mid l_1, \ldots, l_K)$$

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

$$= \frac{\frac{l_1!l_2!\cdots l_{k-1}!(l_k+1)!l_{k+1}!\cdots l_K}{(l+K)!}}{\frac{l_1!l_2!\cdots l_K!}{(l+K-1)!}}$$

$$= \frac{l_k + 1}{l + K}$$

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.

$$\begin{aligned}
P(p_1,\ldots,p_K|\alpha_1,\ldots,\alpha_K) &= \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K p_k^{\alpha_k - 1} \\
\alpha_k &> 0 \\
0 < \; p_k &< 1, \; k = 1,\ldots,K \\
\sum_{k=1}^{K-1} p_k &< 1 \\
p_K &= 1 - \sum_{k=1}^{K-1} p_k
\end{aligned}$$

Dirichlet Distribution Properties

$$
\begin{aligned}
E[p_k] &= \frac{\alpha_k}{\sum_{j=1}^{K} \alpha_j} \\
V[p_k] &= \frac{E[p_k](1 - E[p_k])}{1 + \sum_{j=1}^{K} \alpha_j} \\
C[p_i, p_j] &= \frac{-E[p_i]E[p_j]}{1 + \sum_{k=1}^{K} \alpha_k}
\end{aligned}
$$

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

The uniform distribution on the $K-1$-Simplex is a special case of the Dirichlet distribution where $\alpha_k = 1,\ k = 1, \ldots, K$.

$$
\begin{aligned}
P(p_1, \ldots, p_K | \alpha_1, \ldots, \alpha_K) &= \quad = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K p_k^{\alpha_k - 1} \\
P(p_1, \ldots, p_K | 1, \ldots, 1) &= \frac{\Gamma(K)}{\Gamma(1)} \prod_{k=1}^K p_k^0 \\
&= (K-1)!
\end{aligned}
$$

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 $0 \leq y \leq 1$

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

Dirichlet Prior

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

$$\begin{aligned}
Prob(I_1, \ldots, I_K) &= \int_{(q_1,\ldots,q_K)\in S} Prob(I_1, \ldots, I_K, q_1, \ldots, q_K) dq_1 \ldots dq_K \\
&= \int_{(q_1,\ldots,q_K)\in S} \left[\frac{\prod_{n=1}^K I_n!}{I!} \prod_{m=1}^K q_m^{I_m} \right] \left[\frac{\Gamma(\sum_{n=1}^K \alpha_n)}{\prod_{n=1}^K \Gamma(\alpha_n)} \prod_{m=1}^K q_m^{\alpha_m - 1} \right] dq_1, \ldots, dq_K \\
&= \frac{\prod_{n=1}^K I_n!}{I!} \frac{\Gamma(\sum_{n=1}^K \alpha_n)}{\prod_{n=1}^K \Gamma(\alpha_n)} \int_{(q_1,\ldots,q_K)\in S} \prod_{m=1}^K q_m^{I_m + \alpha_m - 1} dq_1, \ldots, dq_K \\
&= \frac{\prod_{n=1}^K I_n!}{I!} \frac{\Gamma(\sum_{n=1}^K \alpha_n)}{\prod_{n=1}^N \Gamma(\alpha_n)} \frac{\prod_{k=1}^K (I_k + \alpha_k - 1)!}{(I - 1 + \sum_{k=1}^K \alpha_k)!}
\end{aligned}$$

Dirichlet Prior

$$
\begin{aligned}
Prob(d_k, I_1, \ldots, I_K) &= \int_{(q_1,\ldots,q_K)\in S} Prob(d_k, I_1, \ldots, I_K, q_1, \ldots, q_K) dq_1 \ldots dq_K \\
&= \int_{(q_1,\ldots,q_K)\in S} Prob(d_k) Prob(I_1, \ldots, I_K | q_1, \ldots, q_K) \\
&\qquad Prob(q_1, \ldots, q_K) dq_1, \ldots, q_K \\
&= \int_{(q_1,\ldots,q_K)\in S} q_k \frac{\prod_{n=1}^K I_n!}{I!} \prod_{m=1}^K q_m^{I_m} \left[\frac{\Gamma(\sum_{n=1}^K \alpha_n)}{\prod_{n=1}^K \Gamma(\alpha_n)} \prod_{m=1}^K q_m^{\alpha_m - 1} \right] dq_1, \ldots, dq_K \\
&= \frac{\prod_{n=1}^K I_n!}{I!} \frac{\Gamma(\sum_{n=1}^K \alpha_n)}{\prod_{n=1}^K \Gamma(\alpha_n)} \int_{(q_1,\ldots,q_K)\in S} q_k \prod_{m=1}^K q_m^{I_m + \alpha_m - 1} dq_1, \ldots, dq_K \\
&= \frac{\prod_{n=1}^K I_n!}{I!} \frac{\Gamma(\sum_{n=1}^K \alpha_n)}{\prod_{n=1}^N \Gamma(\alpha_n)} \frac{(I_k + \alpha_k) \prod_{n=1}^K (I_n + \alpha_n - 1)!}{(I + \sum_{n=1}^K \alpha_n)(I - 1 + \sum_{n=1}^K \alpha_n)!}
\end{aligned}
$$

Dirichlet Prior

$$
\begin{aligned}
Prob(d_k, I_1, \ldots, I_K) &= \frac{\prod_{n=1}^{K} I_n!}{I!} \frac{\Gamma(\sum_{n=1}^{K} \alpha_n)}{\prod_{n=1}^{N} \Gamma(\alpha_n)} \frac{(I_k + \alpha_k) \prod_{n=1}^{K} (I_n + \alpha_n - 1)!}{(I + \sum_{n=1}^{K} \alpha_n)(I - 1 + \sum_{n=1}^{K} \alpha_n)!} \\
Prob(I_1, \ldots, I_K) &= \frac{\prod_{n=1}^{K} I_n!}{I!} \frac{\Gamma(\sum_{n=1}^{K} \alpha_n)}{\prod_{n=1}^{N} \Gamma(\alpha_n)} \frac{\prod_{k=1}^{K} (I_k + \alpha_k - 1)!}{(I - 1 + \sum_{k=1}^{K} \alpha_k)!} \\
Prob(d_k \mid I_1, \ldots, I_K) &= \frac{\frac{(I_k + \alpha_k) \prod_{n=1}^{K} (I_n + \alpha_n - 1)!}{(I + \sum_{k=1}^{K} \alpha_k)(I - 1 + \sum_{k=1}^{K} \alpha_k)!}}{\frac{\prod_{n=1}^{K} (I_n + \alpha_n - 1)!}{(I - 1 + \sum_{k=1}^{K} \alpha_k)!}} \\
&= \frac{I_k + \alpha_k}{I + \sum_{n=1}^{K} \alpha_n}
\end{aligned}
$$