A Graphical Model for Ordinal Data - MVA 2023/2024

Bruno Costa Alves Freire brunocaf123@gmail.com

December 22, 2023

1 Introduction and contributions

In this project we make a study of the paper [1], which proposes a novel probabilistic model for ordinal data. In the context of parametric statistics, with applications to unsupervised learning (most notably clustering tasks), ordinal data is a kind of categorical data for which there is a total order structure among the categories.

The aim of the original paper is to develop the *Binary Ordinal Search* (BOS) model and present some of its interesting properties in the context of ordinal data, as opposed to other models. The main motivation behind it is to propose a model that is parsimonious, identifiable, and most of all, that does not rely on artifically encoding any sort of arbitrary algebraic structures in the data, thus depending solely on the order structure.

The paper also proposes algorithms for parameter estimation, both in the univariate and multivariate cases, as well as clustering algorithms based on the BOS model, and then applies those to analyse two datasets through an unsupervised learning approach.

For this project in particular, we decided to present alternative proofs for some of the propositions presented in [1], seeking to establish more general arguments for the validity of the properties of the BOS model, i.e., to provide proofs for any number of categories. Besides this theoretical effort, we also implement the algorithms for parameter estimation in the univariate case, as well as well as the direct estimation of the distribution through Monte Carlo simulations. **??**

Since the project was done by a single student, there was no repartition of work. All of the code used to implement the algorithms presented in the original paper was produced by the student from scratch. The code is comprised of a presentation Jupyter notebook, whereas the implementation of the algorithms are organized in separate Python scripts.

2 Method - BOS model

We begin by defining ordinal data and summarizing the BOS model presented in the original paper. An ordinal data can be thought of as realizations of random variables whose state space is a *finite, totally ordered set* $(\mathcal{M}, \prec)$. For ease of representation, we can take $\mathcal{M}$ to be $\{1, \dots, m\}$, for some $m \in \mathbb{N}^*$, and take $\prec$ as the usual order relation among natural numbers. However, we shall not exploit any algebraic relations between those numbers, as they just represent category labels.

The BOS model is a probability distribution on $\mathcal{M}$ parameterized by a position parameter μ, and a precision parameter π. This distribution is characterized by the distribution of the outcome of a stochastic binary search algorithm, described in algorithm 1. The algorithm produces a sequence of decreasing nested intervals until it reaches a singleton, whose sole element will be the outcome.

The auxiliary variables, namely the break points $\{Y_j\}_j$ and accuracies $\{Z_j\}_j$, can be interpreted as missing values for the model.

Algorithm 1 Stochastic Binary Search

Require: Initial interval $E_1 = \{1, \ldots, m\}$, position μ and precision π.
Ensure: Category $x \in \{1, \ldots, m\}$ with probability $\mathrm{p}(x; \mu, \pi)$.

for $j = 1$ **to** $m - 1$ **do**
  Draw break point $Y_j \sim \mathcal{U}(E_j)$
  Draw accuracy $Z_j \sim \mathcal{B}(\pi)$
  if $Z_j = 1$ (perfect comparison) **then**
    Pick $E_{j+1} = \arg\min\{\delta(E, \mu) \mid E \in \{E_j^-, E_j^=, E_j^+\}\}$
  else
    Draw $E_{j+1} \in \{E_j^-, E_j^=, E_j^+\}$ randomly with probability $\frac{|E_{j+1}|}{|E_j|}$
  end if
end for
return $X = Y_m \sim \mathcal{U}(E_m)$ $(E_m = \{X\})$.

To describe and compute the resulting distribution of X from the Stochastic Binary Search algorithm, we can model the conditional probabilities at each step. At each iteration we have the choice of a break point $y_j \in e_j$, the accuracy $z_j \in \{0, 1\}$, and finally the next subinterval $e_{j+1} \subset e_j$, with the following conditional distributions

$$\mathrm{p}(y_j | e_j) = \frac{1}{|e_j|} \mathbb{I}(y_j \in e_j), \tag{1}$$

$$\mathrm{p}(z_j; \pi) = \pi \cdot \mathbb{I}(z_j = 1) + (1 - \pi) \cdot \mathbb{I}(z_j = 0), \tag{2}$$

$$\mathrm{p}(e_{j+1} | y_j, e_j, z_j = 0) = \frac{|e_{j+1}|}{|e_j|} \mathbb{I}(e_{j+1} \in \{e_j^-, e_j^=, e_j^+\}), \tag{3}$$

$$\mathrm{p}(e_{j+1} | y_j, e_j, z_j = 1; \mu) = \mathbb{I}(e_{j+1} = \arg\min\{\delta(e, \mu) \mid e \in \{e_j^-, e_j^=, e_j^+\}\}). \tag{4}$$

In summary, the break point y_j is uniform in e_j, the accuracy z_j follows a Bernoulli distribution with parameter π. As for the subinterval e_{j+1}, it is chosen among the three subintervals $e_j^-, e_j^=, e_j^+ \subset e_j$, defined as $e_j^- = \{b \in e_j \mid b \prec y_j\}$, $e_j^+ = \{b \in e_j \mid b \succ y_j\}$, and $e_j^= = \{y_j\}$, which perform a partition of e_j. If any of them happen to be empty, the choice of e_{j+1} considers only the non-empty ones. Moreover, if the comparison is blind ($z_j = 0$), the choice is random with probability proportional to the length of the subintervals. If the comparison is perfect $z_j = 1$, we deterministically take the interval that contains μ, if any, or the one "closest" to μ among the non-empty options. This distance relation is encoded in $\delta(e, \mu)$.

By combining those expressions, we can express $\mathrm{p}(x; \mu, \pi)$ by marginalizing on y_j, z_j at each iteration and then exploiting the Markovian property of the sequence of subintervals, leading to

$$\mathrm{p}(x; \mu, \pi) = \sum_{e_{m-1}, \ldots, e_1} \prod_{j=1}^{m-1} \mathrm{p}(e_{j+1} | e_j; \mu, \pi) \mathrm{p}(e_1). \tag{5}$$

This distribution will be further explored when we present the proofs for some of the propositions in the paper.

Regarding the computational aspect of computing that distribution, the original paper takes the approach of formal calculus, which yields for every value of m, μ and x, a polynomial in the variable π representing the probability $\mathrm{p}(x;\mu,\pi)$. In this project, we adopted the same approach, by recursively calculating the polynomials for all values of $(\mu, x) \in [m]^2$, and for m up to a small number M, and storing all the coeficients in a table to exploit the memoization technique. The space complexity to store such a table is $\sum_{m=1}^{M} O(m^3) = O(M^4)$, which limits the maximum number of categories M for which we can employ this strategy.

The code implementing this algorithm is in the associated Python scripts and a showcase of its usage is presented in a Jupyter notebook. Moreover, we also compare this direct computation of $\mathrm{p}(x;\mu,\pi)$ to the Monte Carlo estimation of it through the simulation of the Stochastic Binary Search. Results in the appendix see figure **??**.

2.1 Alternative proofs of the model's properties

The original paper highlights some interesting properties of the BOS model distribution, namely: unimodality, precision parameter related to how pronounced the mode is, and identifiability of the parameters. These are more precisely defined through 7 propositions presented in the paper. However, the proofs presented there are essentially computational, and cover only a limited number of cases, i.e., for values of m up to 8. In this context, we found interesting to present alternative proofs for some of those propositions, covering all possible values of $m \in \mathbb{N}^*$. Despite our best efforts, we failed to provide a general proof for proposition A.3, which states that μ is indeed the mode of the distribution $\mathrm{p}(\cdot;\mu,\pi)$.

Proposition 1. *($\pi = 0$: Uniformity) If $\pi = 0$, then $\forall m \in \mathbb{N}^*$, $\forall(\mu, x) \in \{1, \dots, m\}^2$, $\mathrm{p}(x;\mu,\pi) = m^{-1}$.*

Proof. The idea is to show that $\forall 1 \leq j \leq m$, $Y_j \sim \mathcal{U}([m])$. We proceed by induction. Indeed, Y_1 is taken uniformly at random in $E_1 = [m]$. Now consider the conditional distribution of Y_{j+1} given E_j. One has

$$\begin{aligned}
\mathbb{P}[Y_{j+1} = y|E_j = e_j] &= \sum_{y_j}\sum_{e_{j+1}} \mathbb{P}[Y_{j+1} = y|E_{j+1} = e_{j+1}, Y_j = y_j, E_j = e_j]\mathrm{p}(e_{j+1}|y_j, e_j)\mathrm{p}(y_j|e_j) \\
&= \frac{1}{|e_j|}\sum_{y_j \in e_j}\sum_{e_{j+1}} \mathbb{P}[Y_{j+1} = y|E_{j+1} = e_{j+1}]\mathrm{p}(e_{j+1}|y_j, e_j) \\
&= \frac{1}{|e_j|}\sum_{y_j \in e_j}\Bigg(\mathbb{P}[Y_{j+1} = y|E_{j+1} = e_j^-(y_j)]\frac{|e_j^-(y_j)|}{|e_j|} \\
&+ \mathbb{P}[Y_{j+1} = y|E_{j+1} = e_j^=(y_j)]\frac{|e_j^=(y_j)|}{|e_j|} \\
&+ \mathbb{P}[Y_{j+1} = y|E_{j+1} = e_j^+(y_j)]\frac{|e_j^+(y_j)|}{|e_j|}\Bigg)
\end{aligned}$$

Now we use the fact that $E_{j+1} \subset E_j$ and the definitions of $E_j^{-,=,+}$ to obtain

$$\begin{aligned}
\mathbb{P}[Y_{j+1} = y|E_j = e_j] &= \frac{1}{|e_j|}\sum_{y_j \in e_j}\left(\frac{\mathbb{I}(y \in e_j^-(y_j))}{\cancel{|e_j^-(y_j)|}}\frac{\cancel{|e_j^-(y_j)|}}{|e_j|} + \frac{\mathbb{I}(y \in e_j^=(y_j))}{\cancel{|e_j^=(y_j)|}}\frac{\cancel{|e_j^=(y_j)|}}{|e_j|} + \frac{\mathbb{I}(y \in e_j^+(y_j))}{\cancel{|e_j^+(y_j)|}}\frac{\cancel{|e_j^+(y_j)|}}{|e_j|}\right) \\
&= \frac{1}{|e_j|^2}\sum_{y_j \in e_j}\mathbb{I}(y \in e_j) = \frac{1}{|e_j|}\mathbb{I}(y \in e_j) = \mathbb{P}[Y_j = y|E_j = e_j].
\end{aligned}$$

To conclude, we take the expectation w.r.t. to E_j and obtain that $Y_{j+1} =_d Y_j$, i.e., Y_{j+1} and Y_j have the same distribution, establishing the inductive step, and thus we have $X = Y_m =_d \cdots =_d Y_1 \sim \mathcal{U}([m])$. $\square$

Proposition 2. *($\pi = 1$: Dirac in μ) If $\pi = 1$, then $\forall m \in \mathbb{N}^*$, $\forall (\mu, x) \in \{1, \ldots, m\}^2$, $\mathrm{p}(x; \mu, \pi) = \mathbb{I}(x = \mu)$.*

Proof. If $\pi = 1$, that means that $Z_j = 1\ \forall 1 \leq j < m$, and thus at each step we have $E_{j+1} = \arg\min\{\delta(E, \mu) \mid E \in \{E_j^-, E_j^=, E_j^+\}\}$. By induction, since $\mu \in E_1 = [m]$, we'll have that $\mu \in E_j$ for all j almost surely. Moreover, $|E_{j+1}| < |E_j|$ for all j as long as $|E_j| > 1$. So, since $|E_1| = m$, we have that E_m will almost surely have length 1 and contain μ, i.e., $E_m = \{\mu\}$, and thus $X = \mu$ almost surely. $\square$

Now we present a general proof for A.4.

Proposition 3. *(Absolute growing of $\mathrm{p}(\mu; \mu, \pi)$ with π) $\forall m \in \mathbb{N}^*$, $\forall \mu \in [m]$, $\mathrm{p}(\mu; \mu, \pi)$ is an increasing function of π.*

Proof. The key to this proof is to use induction of m, i.e., the size of the initial interval. Moreover, we observe that, when conditioned to a subinterval e at some iteration of the binary search, the distribution $\mathrm{p}(\cdot|e; \mu, \pi)$ behaves exactly as if we had $e_1 = e$ (thus $m = |e|$) and the position of μ adjusted correspondingly. This will allow us to apply the inductive hypothesis.

We begin by observing that when $m = 2$, the distribution takes the form

$$\mathrm{p}(x; \mu = 1, \pi) = \frac{1 + \pi}{2}\mathbb{I}(x = 1) + \frac{1 - \pi}{2}\mathbb{I}(x = 2)$$
$$\mathrm{p}(x; \mu = 2, \pi) = \frac{1 - \pi}{2}\mathbb{I}(x = 1) + \frac{1 + \pi}{2}\mathbb{I}(x = 2).$$

In both cases, we have $\mathrm{p}(\mu; \mu, \pi) = \frac{1+\pi}{2}$, which is increasing with π.

Now, suppose that $\mathrm{p}(\mu; \mu, \pi)$ is increasing with π for all $m \geq M$ and for all $\mu \in [m]$. We compute then $\mathrm{p}(\mu; \mu, \pi)$ in the case where $m = M + 1$, by conditioning on Y_1 and E_2:

$$\begin{aligned}
\mathrm{p}(\mu; \mu, \pi) &= \frac{1}{M+1} \sum_{y_1 \in [M+1]} \sum_{e_2} \sum_{z_1 \in \{0,1\}} \mathrm{p}(\mu|e_2(y_1))\mathrm{p}(e_2|y_1, z_1, e_1)\mathrm{p}(z_1; \pi) \\
&= \frac{1}{M+1} \sum_{z_1 \in \{0,1\}} \left(\sum_{y_1 \prec \mu} \mathrm{p}(\mu|e_1^+(y_1))\mathrm{p}(e_1^+|y_1, z_1, e_1) \right. \\
&\quad \left. + \ \mathrm{p}(\mu|e_1^=(\mu))\mathrm{p}(e_1^=|y_1 = \mu, z_1, e_1) + \sum_{y_1 \succ \mu} \mathrm{p}(\mu|e_1^-(y_1))\mathrm{p}(e_1^-|y_1, z_1, e_1) \right).
\end{aligned}$$

Next, we unravel the conditioning on Z_1:

$$
\begin{aligned}
\mathrm{p}(\mu;\mu,\pi) &= \frac{\pi}{M+1}\left(\sum_{y_1\prec\mu}\mathrm{p}(\mu|e_1^+(y_1))\tfrac{1}{|e_1|}+\mathrm{p}(\mu|e_1^=(\mu))\tfrac{1}{|e_1|}+\sum_{y_1\succ\mu}\mathrm{p}(\mu|e_1^-(y_1))\tfrac{1}{|e_1|}\right)\\
&+\frac{(1-\pi)}{M+1}\left(\sum_{y_1\prec\mu}\mathrm{p}(\mu|e_1^+(y_1))\tfrac{|e_1^+(y_1)|}{|e_1|}+\mathrm{p}(\mu|e_1^=(\mu))\tfrac{|e_1^=(\mu)|}{|e_1|}+\sum_{y_1\succ\mu}\mathrm{p}(\mu|e_1^-(y_1))\tfrac{|e_1^-(y_1)|}{|e_1|}\right)\\
&= \frac{\pi}{M+1}\left(\sum_{y_1\prec\mu}\mathrm{p}(\mu|e_1^+(y_1))\tfrac{|e_1\setminus e_1^+|}{|e_1|}+\mathrm{p}(\mu|e_1^=(\mu))\tfrac{|e_1\setminus e_1^=|}{|e_1|}+\sum_{y_1\succ\mu}\mathrm{p}(\mu|e_1^-(y_1))\tfrac{|e_1\setminus e_1^-|}{|e_1|}\right)\\
&+\frac{1}{M+1}\left(\sum_{y_1\prec\mu}\mathrm{p}(\mu|e_1^+(y_1))\tfrac{|e_1^+(y_1)|}{|e_1|}+\mathrm{p}(\mu|e_1^=(\mu))\tfrac{|e_1^=(\mu)|}{|e_1|}+\sum_{y_1\succ\mu}\mathrm{p}(\mu|e_1^-(y_1))\tfrac{|e_1^-(y_1)|}{|e_1|}\right)
\end{aligned}
$$

In the above expression, all the intervals $e_1^{-,=,+}(y_1)$ are strict subintervals of e_1, i.e., $|e_2| < |e_1| \leq M+1$, and thus we can apply the inductive hypothesis, meaning that $\mathrm{p}(\mu|e_1^{-,=,+}(y_1))$ are all increasing and positive functions of π. That yields $\mathrm{p}(\mu;\mu,\pi) = \pi \cdot F(\pi) + G(\pi)$, where F, G are positive nondecreasing functions, and then $\mathrm{p}(\mu;\mu,\pi)$ for $m = M+1$ is increasing with π, thus establishing the inductive step. Since we have already proven the base case $m = 2$, and for $m = 1$ we have $\mathrm{p}(\mu;\mu,\pi)$ constant, i.e., nondecreasing, the induction is complete and thus $\mathrm{p}(\mu;\mu,\pi)$ is increasing with π for all $m \in \mathbb{N}^*$ and all $\mu \in [m]$. $\square$

2.2 Parameter estimation via EM Maximum Likelihood estimation

In the original paper, the method proposed for estimating the parameters μ and π from an n-sample $(x_i)_{i=1}^n$ is to consider the break-points, accuracies and subintervals Y_j, Z_j, E_{j_j} as missing values, and run an EM algorithm. This algorithm consists in alternating expectation (E) and maximization (M) steps until convergence of the log-likelihood $\ell(\mu, \pi)$. In practice, we define a threshold for the change in the absolute value of the log-likelihood, and optionally set a maximum number of iterations.

At each iteration $q \geq 1$, the E step consists in computing the conditional distribution of the missing values for all possible paths $c_i = \{y_{ij}, z_{ij}, e_{ij}\}_{j=1}^{m-1}$ that result in x_i, for all $i \in [n]$. The M step then updates the parameters μ, π according to the equations:

$$
\pi^{[q+1]} = \frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m-1}\mathrm{p}(z_{ij}=1|x_i;\mu^{[q]},\pi^{[q]}) \tag{6}
$$

$$
\mu^{[q+1]} = \underset{\mu'\in[m]}{\arg\max}\,\bar{\ell}_c(\mu'), \tag{7}
$$

where the expected conditional completed log-likelihood $\bar{\ell}_c(\mu')$ is defined as

$$
\bar{\ell}_c(\mu') := \sum_{i=1}^{n}\sum_{c_i\in C_i}\mathrm{p}(c_i|x_i;\mu^{[q]},\pi^{[q]})\ln \mathrm{p}(x_i;\mu',\pi^{[q+1]}). \tag{8}
$$

We find important to point out that equation 8 is incorrect in the original paper. The mistaken replacement of $\mathrm{p}(x_i;\mu',\pi^{[q+1]})$ by $\mathrm{p}(c_i, x_i;\mu',\pi^{[q+1]})$ in the formula can make the algorithm incapable of finding the correct μ.

For this project we implemented the (corrected) EM algorithm in Python, with slight modifications. In order to speed up the update steps of μ, π, we precomputed the joint distributions of (c_i, x_i) for all possible values of μ and for all $i \in [n]$, where the dependency in π is encoded as polynomial coefficients, thus doing the E step only once. Moreover, we decided to keep only y_{ij} and z_{ij} as missing values in c_i, since the sequence of y_{ij} values already determines the sequence of subintervals.

2.3 Extension to multivariate ordinal data and clustering

With the purpose of applying the BOS model to clustering tasks, it is necessary to first extend it to multivariate ordinal data. We consider then $\mathbf{x} = (x_h)_{1\leq h\leq d}$ a d-dimensional ordinal variable, where the h-th component x_h lies in $[m_h]$. On top of that, the multivariate model also includes a group indicator variable $\mathbf{w} = (w_1, \ldots, w_g) \in \{0, 1\}^g$, where g is the number of groups under consideration, which works as a one-hot encoding of the group, i.e., $w_k = 1$ only for the cluster k to which the observation belongs. The basic assumption behind the proposed extension is that the components are conditionally independent given the group they belong to:

$$\mathrm{p}(\mathbf{x}|w_k = 1; \boldsymbol{\mu}_k, \boldsymbol{\pi}_k) = \prod_{h=1}^{d} \mathrm{p}(x_h; \mu_k^h, \pi_k^h), \tag{9}$$

where $\boldsymbol{\mu} = (\mu_k^1, \ldots, \mu_k^d), \boldsymbol{\pi} = (\pi_k^1, \ldots, \pi_k^d)$ contain the position and precision parameters for each of the components of $\mathbf{x}$ in group k, whereas w has a one order multinomial distribution $w \sim \mathcal{M}(1, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_g)$ is the mixture parameter of the clusters. We can then express the marginal distribution of $\mathbf{x}$ by taking the expectation on w:

$$\mathrm{p}(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\pi}) = \sum_{k=1}^{g} \alpha_k \mathrm{p}(\mathbf{x}|w_k = 1; \boldsymbol{\mu}_k, \boldsymbol{\pi}_k) \tag{10}$$

2.4 AECM algorithm for Maximum Likelihood estimation

In this multivariate setting, we can estimate the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\pi}$ of the multivariate distributions as well as the mixture proportions $\boldsymbol{\alpha}$ through the Alternating Expectation-Conditional Maximization (AECM) algorithm, proposed in the original paper. We consider a n-sample $(\mathbf{x_i})_{i=1}^n$ of multivariate ordinal data in the space $[m_1] \times \cdots \times [m_d]$, independently drawn from the distribution $\mathrm{p}(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\pi})$. We then proceed to maximize the log-likelihood $\ell(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\pi})$ through an EM-like algorithm with missing values. The missing values in this case are appended of the group indicators $(\mathbf{w}_i)_{i=1}^n$, and the other missing values are adjusted to the multivariate setting, yielding $c_i = \{\mathbf{w}_i, \{y_{ijk}^h, z_{ijk}^h, e_{ijk}^h\}_{1\leq h\leq d, 1\leq j\leq m^h-1, 1\leq k\leq g}\}$.

The algorithm starts with initial parameters $(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\pi})^{[0]}$ and alternates between E steps, where one computes the conditional probabilities:

$$\mathrm{p}(w_{ik}|\mathbf{x}_i, \boldsymbol{\alpha}^{[q]}, \boldsymbol{\mu}^{[q]}, \boldsymbol{\pi}^{[q]}) = \frac{\alpha_k^{[q]} \mathrm{p}(\mathbf{x}_i|w_{ik} = 1; \boldsymbol{\mu}_k^{[q]}, \boldsymbol{\pi}_k^{[q]})}{\sum_{k'\in[g]} \alpha_{k'}^{[q]} \mathrm{p}(\mathbf{x}_i|w_{ik'} = 1; \boldsymbol{\mu}_{k'}^{[q]}, \boldsymbol{\pi}_{k'}^{[q]})}, \tag{11}$$

and M steps, where one updates the model parameters. The update for the mixing proportions is

$$\alpha_k^{[q+1]} = \frac{1}{n} \sum_{i\in[n]} \mathrm{p}(w_{ik} = 1|\mathbf{x}_i; \boldsymbol{\alpha}^{[q]}, \boldsymbol{\mu}^{[q]}, \boldsymbol{\pi}^{[q]}). \tag{12}$$

For the mode and precision parameters, $(\mu_k^h, \pi_k^h)^{[q+1]}$, the update is done through an internal univariate EM algorithm, ran independently for each component $h \in [d]$. To take into account cluster membership, the k-th set of parameters is updated by weighing the observations by $\mathrm{p}(w_{ik} = 1|\mathbf{x}_i; \boldsymbol{\alpha}^{[q]}, \boldsymbol{\mu}^{[q]}, \boldsymbol{\pi}^{[q]})$.

Appendix

In this appendix we gather all the figures produced for this report. More will be produced for the poster version of this project.

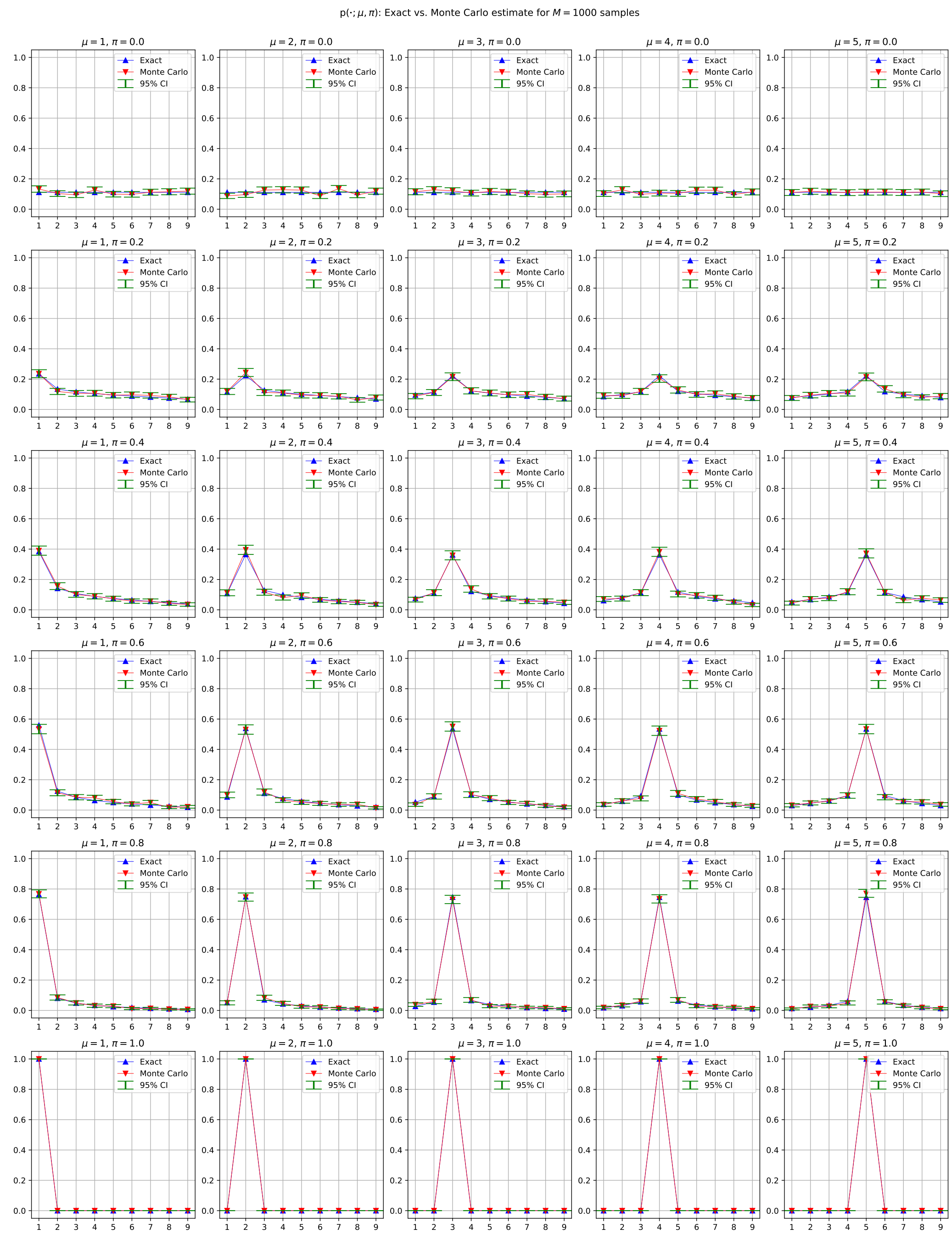

Figure 1: Shape of the distribution of the BOS model $\mathrm{p}(x; \mu, \pi)$ for $m = 7$, for different values of μ, π. In blue we have the exact distribution plotted agains the Monte Carlo estimation for $M = 1000$ samples.

References

[1] Christophe Biernacki and Julien Jacques. Model-Based Clustering of Multivariate Ordinal Data Relying on a Stochastic Binary Search Algorithm. *Statistics and Computing*, 26(5):929–943, 2016.