Assignment 2

Ashutosh Mittal (amittal@kth.se)
DD2434 Machine Learning, Advanced Course

December 29, 2016

Graphical Models

Task 2.1: Qualitiative effects in a Directed Graphical Model (DGM)

1. $p(t^1) > p(t^1|d^1)$
2. $p(d^1|t^0) > p(d^1)$
3. $p(h^1|e^1, f^1) = p(h^1|e^1)$
4. $p(c^1|f^0) = p(c^1)$
5. $p(c^1|h^0) > p(c^1)$
6. $p(c^1|h^0, f^0) = p(c^1|h^0)$
7. $p(d^1|h^1, e^0) = p(d^1|h^1)$
8. $p(d^1|e^1, f^0, w^1)\,?\,p(d^1|e^1, f^0)$
9. $p(t^1|w^1, f^0) = p(t^1|w^1)$

Question 1

1, 2 and 5 have inequality.

Question 2

3, 4, 6, 7 and 9 have equality.

Question 3

8 cannot be compared based on the provided information about the DAG.

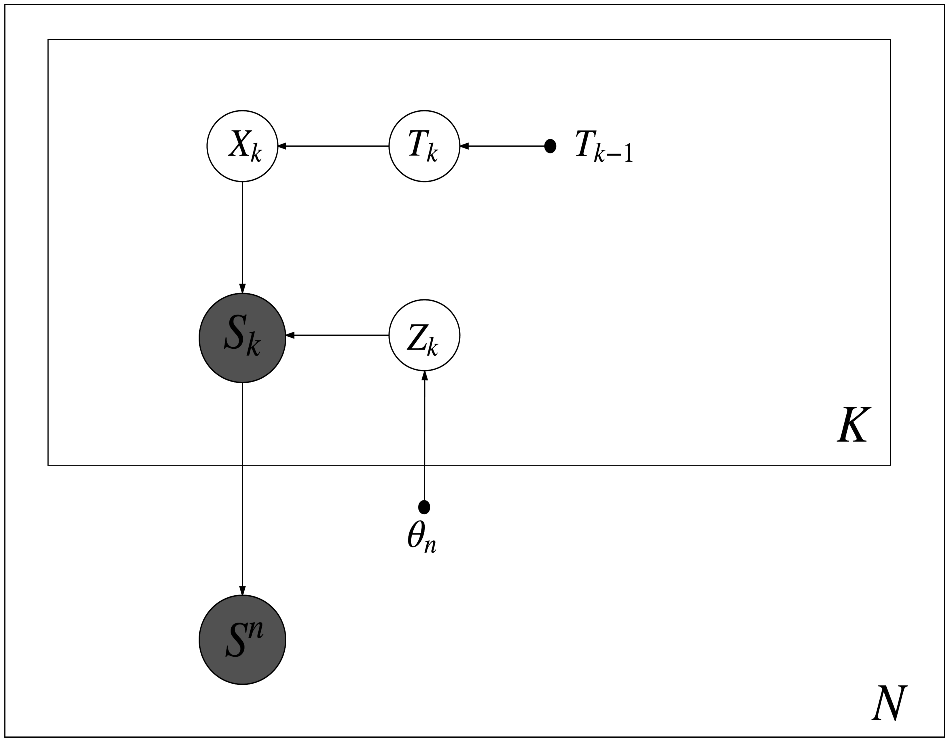


Figure 1: Graphical model

Task 2.2: Casino Model

Question 4

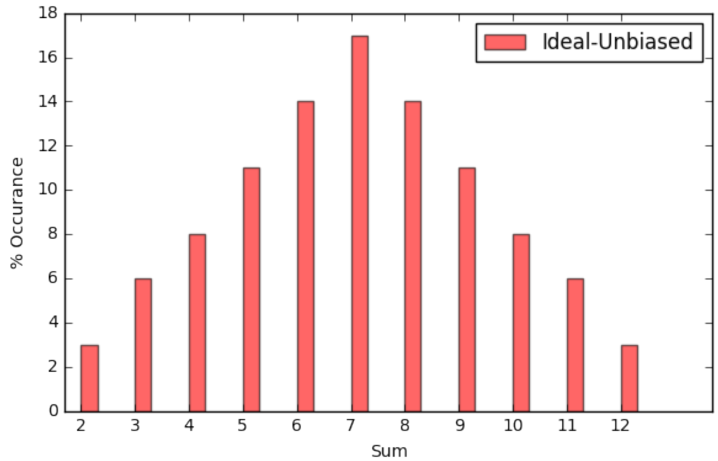
Fig. 1 shows the graphical model for the presented casino problem. While other symbols are same as used in the question, θ_n represents the parameters of the probability distribution of each user's personal dice outcome, and T_k is a random variable which represents the table user chooses given the previous table was T_{k-1}.

Question 5

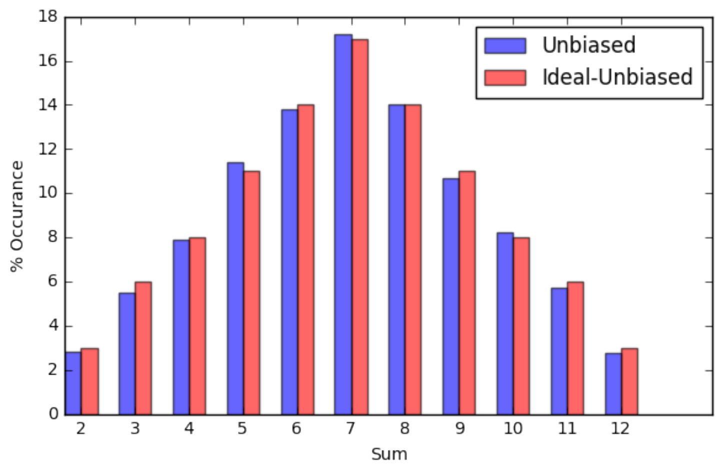
Done! [Python, 60 lines]

Question 6

Fig. 2a shows the probability distribution of values of sum of two unbiased dices that would be obtained if they were rolled for infinite number of times. We will use this as reference for the following cases. For all the following cases we are assuming there are 100 T' tables, 100 T tables and 100 players.



(a) Ideal Unbiased distribution for sum values



(b) Case 1: Unbiased distribution for sum values

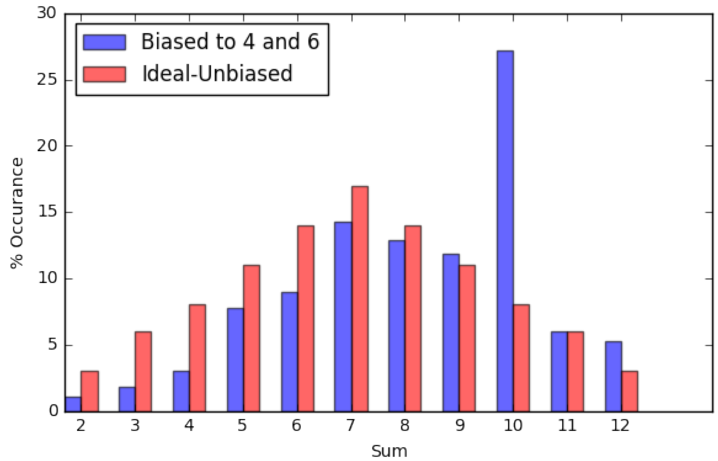
Figure 2: Casino Model

Case 1

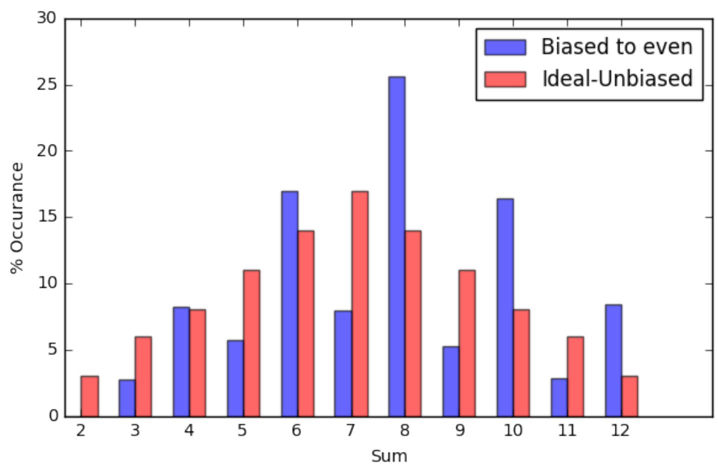
In this case we are assuming all dices are unbiased.

$$T' = [1/6, 1/6, 1/6, 1/6, 1/6, 1/6]$$
$$T = [1/6, 1/6, 1/6, 1/6, 1/6, 1/6]$$
$$Z = [1/6, 1/6, 1/6, 1/6, 1/6, 1/6]$$

Fig. 2b shows the distribution for this case compared to ideal unbiased case. We can see that they are approximately same, however the difference is due to the fact that the we are not rolling the dice for infinite number of times.



(a) Case 2: Players' dice bias to 4, Tables' dice biased to 6



(b) Case 3: Players' dice strictly bias to even, Tables' dice slightly bias to even

Figure 3: Casino Model

Case 2

In this case we are assuming players' dices are biased to 4 while those of the tables are biased to 6.

$$T' = [1/10, 1/10, 1/10, 1/10, 1/10, 1/2]$$
$$T = [1/10, 1/10, 1/10, 1/10, 1/10, 1/2]$$
$$Z = [1/10, 1/10, 1/10, 1/2, 1/10, 1/10]$$

Fig. 3a shows the distribution for this case compared to ideal unbiased case. We can see two effects here. Firstly, the probability of occurrence of 10 has risen up sharply and secondly, probability for sum values 4 and below have taken a serious blow.

Case 3

In this case we are assuming all the dices all biased towards even values.

$$T' = [1/4, 1/12, 1/4, 1/12, 1/4, 1/12]$$
$$T = [1/4, 1/12, 1/4, 1/12, 1/4, 1/12]$$
$$Z = [0, 1/3, 0, 1/3, 0, 1/3]$$

Fig. 3b shows the distribution for this case compared to ideal unbiased case. We can observe the drastic decrease in probability of odd sums as all the dices are biased towards even values. Probability of 8 has shot up even higher than 6.

Task 2.3: Simple VI

Question 7

Using Eq. 10.24 - 10.33 from the coursebook bishop (2006):

$$q(\mu, \tau) = q(\mu)q(\tau)$$

$$q(\mu^*) = \mathcal{N}(\mu_N, (\lambda_N)^{-1})$$
$$\mu_N = \frac{\mu_0\lambda_0 + N\overline{x}}{\lambda_0 + N}$$
$$\lambda_N = (\lambda_0 + N)\mathtt{E}(\tau)$$

$$q(\tau^*) = \mathrm{Gam}(a_N, b_N)$$
$$a_N = a_0 + \frac{N+1}{2}$$
$$b_N = b_0 + \frac{1}{2}\big(\lambda_0\mu_0^2 + \sum_n x_n^2\big) + \frac{N+\lambda_0}{2}(\mu_N^2 + 1/\lambda_N) - \big(\sum_n x_n + \lambda_0\mu_0\big)\mu_N$$
$$\mathtt{E}(\tau) = \frac{a_N}{b_N}$$

We can iterate and find the optimal value for $q(\mu, \tau)$.
Coded in python (110 lines)

Question 8

$$p(\mu, \tau | D, a_0, b_0, \mu_0, \tau_0, \lambda_0) \propto p(D, \mu, \tau | a_0, b_0, \mu_0, \tau_0, \lambda_0) \quad (1)$$

$$
\begin{aligned}
\ln(D, \mu, \tau | a_0, b_0, \mu_0, \tau_0, \lambda_0) &= \ln p(D|\mu, \tau) p(\mu|\mu_0, \lambda_0, \tau) p(\tau|a_0, b_0) \\
&= \ln \ \mathcal{N}(D|\mu, \tau)\mathcal{N}(\mu|\mu_0, (\lambda_0\tau)^{-1})\mathrm{Gam}(\tau|a_0, b_0) \\
&= \frac{N}{2}\ln\tau - \frac{\tau}{2}\sum_n (x_n - \mu)^2 + \frac{1}{2}\ln\tau - \frac{\lambda_0\tau}{2}(\mu - \mu_0)^2 \\
&\quad - (b_0 + \frac{\lambda_0\mu_0^2}{2} + \frac{\sum_n x_n^2}{2})\tau + (a_0 - 1)\ln\tau + C \\
&= \frac{1}{2}\ln\tau - \frac{\mu^2}{2}(\lambda_0 + N)\tau + \mu\tau\big(\mu_0\lambda_0 + \sum_n x_n\big) \\
&\quad - (b_0 + \frac{\lambda_0\mu_0^2}{2} + \frac{\sum_n x_n^2}{2})\tau + (a_0 + \frac{N}{2} - 1)\ln\tau + C \\
&= \ln \ \mathcal{N}(\mu|\mu', (\lambda'\tau)^{-1}) \ \mathrm{Gam}(\tau|a', b') + C
\end{aligned}
$$

$$
\begin{aligned}
\mu' &= \frac{\mu_0\lambda_0 + \sum_n x_n}{\lambda_0 + N} \\
\lambda' &= (\lambda_0 + N) \\
a' &= a_0 + \frac{N}{2} \\
b' &= b_0 + \frac{\lambda_0\mu_0^2}{2} + \frac{\sum_n x_n^2}{2}
\end{aligned}
$$

Normalizing the joint $p(D, \mu, \tau)$, we can find exact posterior which would also be a Normal-Gamma function as well.

$$
p(\mu, \tau|D) = \mathcal{N}(\mu|\mu', (\lambda'\tau)^{-1}) \ \mathrm{Gam}(\tau|a', b') \tag{2}
$$

Question 9

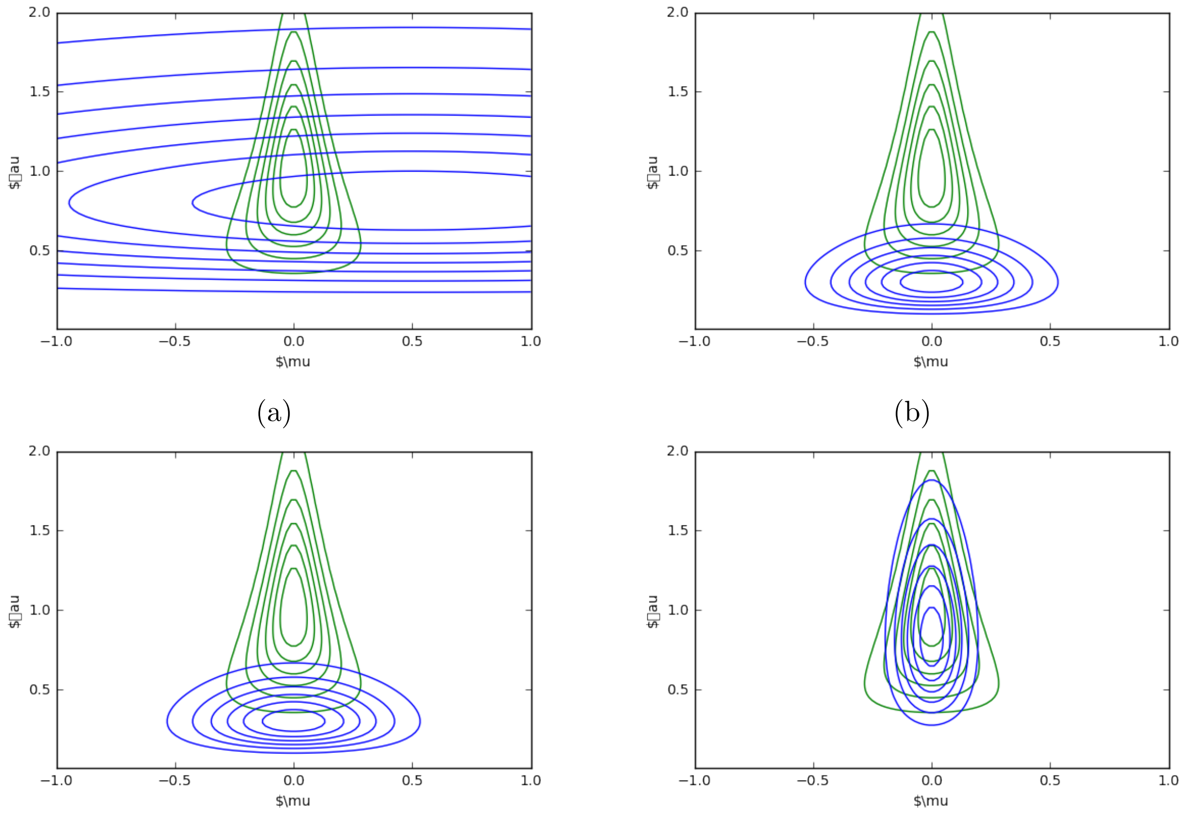
Fig. 4 shows how $q(\mu, \tau)$ converges to the exact posterior $p(\mu, \tau|D)$ (derivation and expression for exact posterior is shown in question 8). The generated data in this case is with zero mean and unit variance.

Task 2.4: Sampling tables given dice sums

Question 10

Given:

1. Θ: All categorical distributions corresponding to the dice
2. $r_1 ... r_K$ sequence of tables visited
3. $s_1 ... s_K$ observation of dice sums



(a)

(b)

(c)

(d)

Figure 4: Illustration of variational inference for the mean μ and precision τ of a univariate Gaussian distribution. Contours of the true posterior distribution $p(\mu, \tau|D)$ are shown in green. (a) Contours of the initial factorized approximation $q(\mu, \tau)$ are shown in blue. (b) After re-estimating the factor $q(\mu)$ (c) After re-estimating the factor $q(\tau)$ (d) Contours of the optimal factorized approximation, to which the iterative scheme converges, are shown in red

Algorithm to find $p(r_1, ..r_K|s_1, ...s_K, \Theta)$

$$p(r_1, ..r_K|s_1, ...s_K, \Theta) = \frac{p(r_1, ..r_K, s_1, ...s_K|\Theta)}{p(s_1, ...s_K|\Theta)} \tag{3}$$

We will separately evaluate the numerator and denominator of Eq. 3.

Let:

$$q = p(r_1|\Theta) \tag{4}$$
$$b_i = p(s_i|r_i, \Theta) \tag{5}$$
$$a_i = p(r_i|r_{i-1}, \Theta) \tag{6}$$

Based on the information provided in the question:

$$
\begin{aligned}
q &= p(r_1|\Theta) = 0.5 \\
b_i &= p(\text{Player's dice} + \text{Table's dice}|\text{Table } r_i) \\
\Rightarrow b_i &= p(\text{Player's dice}) * p(\text{Table's dice}|\text{Table } r_i) \\
a_i &= p(r_i|r_{i-1}, \Theta) = \begin{cases} 0.25, & \text{if } r_i, r_{i-1} \in \{t_i, t_{i-1}\} \text{ or } r_i, r_{i-1} \in \{t'_i, t'_{i-1}\} \\ 0.75, & \text{otherwise} \end{cases}
\end{aligned}
$$

<u>Numerator</u>:
We will start from $p(r_1, s_1|\Theta)$ to iterate and find the desired expression $p(r_1, ..r_K, s_1, ...s_K|\Theta)$.

$$
\begin{aligned}
&\text{Step 1: } P_1 && p(r_1, s_1|\Theta) = p(s_1|r_1, \Theta)p(r_1|\Theta) \\
& && \Rightarrow p(r_1, s_1|\Theta) = b_i q \\
&\text{Step 2: } P_2 && p(r_1, s_1, r_2, s_2|\Theta) = p(s_1, r_1|\Theta)p(r_2, s_2|r_1, s_2, \Theta) \\
& && \Rightarrow p(r_1, s_1, r_2, s_2|\Theta) = p(s_1, r_1|\Theta)p(r_2, s_2|r_1, \Theta) \\
& && \Rightarrow p(r_1, s_1, r_2, s_2|\Theta) = p(s_1, r_1|\Theta)p(r_2|r_1, \Theta)p(s_2|r_2, \Theta) \\
& && \Rightarrow p(r_1, s_1, r_2, s_2|\Theta) = a_2 b_2 P_1 \\
&... \\
&\text{Step j: } P_j && p(r_1, ...r_j, s_1, ...s_j|\Theta) = a_j b_j P_{j-1} \\
&... \\
&\text{Step K: } P_K && p(r_1, ...r_K, s_1, ...s_K|\Theta) = a_K b_K P_{K-1}
\end{aligned}
$$

Using the induction shown above, we can compute the numerator as follows:

$$
p(r_1, ...r_K, s_1, ...s_K|\Theta) = b_1 q \prod_{i=2}^{K} a_i b_i \tag{7}
$$

<u>Denominator</u>:

Let:

$$
\begin{aligned}
A_i &= p(s_i, r_i|s_{i-1}, ...s_1, \Theta) \\
C_i &= p(s_i|s_{i-1}...s_1, \Theta) \\
B_i &= p(r_i|s_i...s_1, \Theta) = \frac{A_i}{C_i}
\end{aligned}
$$

Step 1: $$A_1 = p(r_1, s_1|\Theta) = p(s_1|r_1, \Theta)p(r_1|\Theta)$$
$$\Rightarrow A_1 = b_i q$$
$$C_1 = \sum_{r_1} p(r_1, s_1|\Theta) = \sum_{r_1} A_1$$

Step 2: $$A_2 = p(r_2, s_2|s_1, \Theta) = p(s_2|r_2, \Theta)p(r_2|s_1, \Theta)$$
$$\Rightarrow A_2 = b_2 \sum_{r_1} p(r_2, r_1|s_1, \Theta)$$
$$\Rightarrow A_2 = b_2 \sum_{r_1} p(r_2|r_1, \Theta)p(r_1|s_1, \Theta)$$
$$\Rightarrow A_2 = b_2 \sum_{r_1} a_2 B_1$$
$$\Rightarrow A_2 = b_2 \sum_{r_1} p(r_2, r_1|s_1, \Theta)$$
$$\Rightarrow A_2 = b_2 \sum_{r_1} p(r_2|r_1, \Theta)p(r_1|s_1, \Theta)$$
$$\Rightarrow A_2 = b_2 \sum_{r_1} a_2 \frac{A_1}{C_1}$$
$$C_2 = \sum_{r_2} p(r_2, s_2|s_1, \Theta) = \sum_{r_2} A_2$$

...

Step j: $$A_j = b_j \sum_{r_{j-1}} a_j B_{j-1}$$
$$\Rightarrow A_j = b_j \sum_{r_{j-1}} a_j \frac{A_{j-1}}{C_{j-1}}$$
$$C_j = \sum_{r_j} A_j$$

...

Step K: $$A_K = b_K \sum_{r_{K-1}} a_K \frac{A_{K-1}}{C_{K-1}}$$
$$C_K = \sum_{r_K} A_K$$

From the induction based derivation shown above we can compute the denominator:

$$p(s_1, ...s_K|\Theta) = \prod_{i=1}^{K} C_i \tag{8}$$

Solution:

$$p(r_1, ..r_K|s_1, ...s_K, \Theta) = \frac{b_1 q \prod_{i=2}^{K} a_i b_i}{\prod_{i=1}^{K} C_i} \tag{9}$$

Question 11

$$\begin{aligned} p(R_1, ..., R_K|s_1, ..., s_K, \Theta) = p(R_K|s_1, ..., s_K, \Theta)p(R_{K-1}|R_{K-1}, s_1, ..., s_K, \Theta)... \\ ...p(R_1|R_K, ...R_2, s_1, ..., s_K, \Theta) \end{aligned}$$

We can see that in order to sample for R_i ($i \in \{1, K\}$) we need $\{R_{i-1}, ..., R_K\}$.
Thus we can progressively sample $\{r_K, ..., r_i\}$ by starting the sampling from R_K and use it's value to sample R_{K-1} and so on.

<u>Step 1: Sampling R_K</u>:

$$\begin{aligned}
p(R_K|s_1, ..., s_K, \Theta) &= \frac{p(R_K, s_1, ..., s_K|\Theta)}{p(s_1, ..., s_K|\Theta)} \\
&= \frac{p(R_K, s_1, ..., s_K|\Theta)}{\sum_{R_K} p(R_K, s_1, ..., s_K|\Theta)} \\
p(R_K, s_1, ..., s_K|\Theta) &= p(R_K, s_1, ..., s_{K-1}|\Theta)p(s_K|R_K, \Theta) \\
&= p(R_K, s_1, ..., s_{K-1}|\Theta)b_K \\
&= \sum_{R_{K-1}} p(R_K, R_{K-1}, s_1, ..., s_{K-1}|\Theta)b_K \\
&= \sum_{R_{K-1}} p(R_K|R_{K-1})p(R_{K-1}, s_1, ..., s_{K-1}|\Theta)b_K \\
&= \sum_{R_{K-1}} p(R_K|R_{K-1})p(s_{K-1}|R_{K-1})p(R_{K-1}, s_1, ..., s_{K-2}|\Theta)b_K \\
&= b_K \sum_{R_{K-1}} a_K b_{K-1} p(R_{K-1}, s_1, ..., s_{K-2}|\Theta) \\
&= b_K \sum_{R_{K-1}} a_K b_{K-1} \sum_{R_{K-2}} ... \sum_{R_1} a_2 b_1 q
\end{aligned}$$

We can compute this probability because we know all a, b and q. Using this probability we can sample R_K

Step 2: Sampling R_{K-1}:

$$\begin{aligned}
p(R_{K-1}|r_K, s_1, ..., s_K, \Theta) &= \frac{(R_{K-1}, r_K, s_1, ..., s_K|\Theta)}{p(r_K, s_1, ..., s_K|\Theta)} \\
&= \frac{(s_K|r_K, \Theta)p(r_K|R_{K-1}, \Theta)p(R_{K-1}, s_1, ..., s_{K-1}|\Theta)}{p(r_K, s_1, ..., s_K|\Theta)} \\
&= \frac{a_K \ b_K \ p(R_{K-1}, s_1, ..., s_{K-1}|\Theta)}{p(r_K, s_1, ..., s_K|\Theta)}
\end{aligned}$$

$p(r_K, s_1, ..., s_K|\Theta)$ can be used as obtained in the previous step and we already know a_K, b_K. $p(R_{K-1}, s_1, ..., s_{K-1}|\Theta)$ can be found in the same way as $p(R_K, s_1, ..., s_K|\Theta)$ was calculated in the first step. Using the value of probability thus obtained, we can sample R_{K-1}.

Expectation-Maximization (EM)

Task 2.5: EM

Question 12

We can start by assuming some initial parameters for the categorical distribution θ, π and update their values iteratively such that expectation of joint probability distribution is maximized.

$$\begin{aligned}
S_k^n &: \quad \text{Sum of } n^{th} \text{ player's dice and } k^{th} \text{ table's dice} \\
X_k^n &: \quad k^{th} \text{ table's dice value for } n^{th} \text{ player} \\
Z_k^n &: \quad n^{th} \text{ player's dice value for } k^{th} \text{ table} \\
\pi^k &: \quad \text{Categorical distribution parameters of } k^{th} \text{ table} \\
\theta^n &: \quad \text{Categorical distribution parameters of } n^{th} \text{ player} \\
I(\text{condition}) &: \quad \text{Indicator function}
\end{aligned}$$

Joint probability distribution:

$$\begin{aligned}
\ln p(S_{1:K}^{1:N}, X_{1:K}^{1:N}, Z_{1:K}^{1:N}|\theta^{1:N}, \pi^{1:K}) &= \sum_k \sum_n I(Z_k^n + X_k^n = S_k^n) \ln \left(p(Z_k^n|\theta^n)p(X_k^n|\pi^k)\right) \\
&= \sum_k \sum_n \sum_i \sum_j I(i + j = S_k^n) \ln(\theta_i^n \pi_j^k) \\
&= \sum_k \sum_n \sum_i \sum_j I(i + j = S_k^n) \ln(\theta_i^n) \\
&\quad + \sum_k \sum_n \sum_i \sum_j I(i + j = S_k^n) \ln(\pi_j^k)
\end{aligned}$$

Let $\gamma_{i,j}^{n,k}$ be the responsibility parameter. For a certain value of S_k^n, $\gamma_{i,j}^{n,k}$ assigns the responsibility that $Z_k^n = i$ and $X_k^n = j$ such that $i + j = S_k^n$.

$$\gamma_{i,j}^{n,k} = \frac{I(i + j = S_k^n) p(\theta_i^n) p(\pi_j^k)}{\sum_l \sum_m I(l + m = S_k^n) p(\theta_l^n) p(\pi_m^k)}$$

Expectation of Joint probability distribution:

$$\begin{aligned}
\mathbb{E} \ln p(S_{1:K}^{1:N}, X_{1:K}^{1:N}, Z_{1:K}^{1:N} | \theta^{1:N}, \pi^{1:K}) &= \mathbb{E}\Big[\sum_k \sum_n \sum_i \sum_j I(i + j = S_k^n) \ln(\theta_i^n) \\
&\quad + \sum_k \sum_n \sum_i \sum_j I(i + j = S_k^n) \ln(\pi_j^k) \Big] \\
&= \sum_k \sum_n \sum_i \sum_j \gamma_{i,j}^{n,k} \ln(\theta_i^n) \ + \ \sum_k \sum_n \sum_i \sum_j \gamma_{i,j}^{n,k} \ln(\pi_j^k) \\
&= \sum_n \sum_i \Big(\sum_k \sum_j \gamma_{i,j}^{n,k} \Big) \ln(\theta_i^n) \ + \ \sum_k \sum_j \Big(\sum_n \sum_i \gamma_{i,j}^{n,k} \Big) \ln(\pi_j^k)
\end{aligned}$$

In order to maximize this expectation, we can update the values of parameters θ, π:

$$\begin{aligned}
\theta_i^n &= \frac{\sum_k \sum_j \gamma_{i,j}^{n,k}}{\sum_i \sum_k \sum_j \gamma_{i,j}^{n,k}} = \frac{\gamma_i^n}{\gamma^n} \\
\pi_j^k &= \frac{\sum_n \sum_i \gamma_{i,j}^{n,k}}{\sum_j \sum_n \sum_i \gamma_{i,j}^{n,k}} = \frac{\gamma_j^k}{\gamma^k}
\end{aligned}$$

Using these updated parameter values, we can again sample $S_{1:K}^{1:N}$ and keep on iterating until convergence.

Question 13

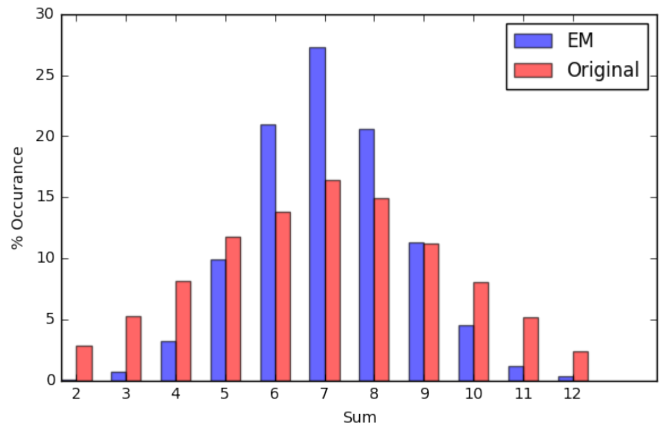
Here we generate frequency distribution of sum values for original dice distributions and then compare it with the same for dice distributions generated by EM algorithm. The cases considered are similar to ones taken in the Task 2.2 (Simple VI). Please keep in mind that the frequency distribution of sum values for original dice probability is also variable as we are rolling the dices limited number of times.

Case 1

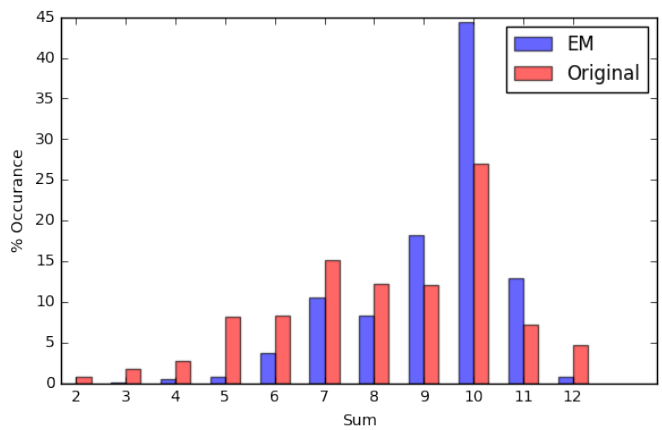
In this case we are assuming all dices are unbiased.

$$\begin{aligned}
X &= [1/6, 1/6, 1/6, 1/6, 1/6, 1/6] \\
Z &= [1/6, 1/6, 1/6, 1/6, 1/6, 1/6]
\end{aligned}$$

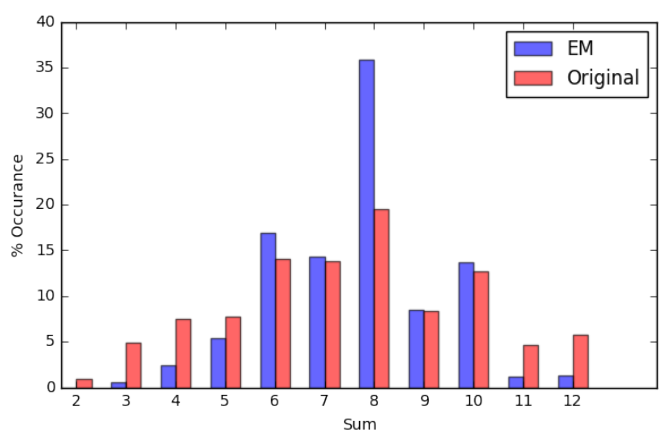
Fig. 5a shows the distribution for this case compared to original dice's distribution. We can see the one generated by EM follows similar trend as the one by original distribution.



(a) Case 1: Unbiased distribution of dices



(b) Case 2: Players' dice bias to 4, Tables' dice biased to 6



(c) Case 3: Players' dice strictly bias to even, Tables' dice slightly bias to even

Figure 5: EM: Casino Model

Case 2

In this case we are assuming players' dices (Z) are biased to 4 while those of the tables (X) are biased to 6.

$$X = [1/10, 1/10, 1/10, 1/10, 1/10, 1/2]$$
$$Z = [1/10, 1/10, 1/10, 1/2, 1/10, 1/10]$$

Fig. 5b shows the distribution for this case compared to original dice's distribution. We can see in both of them that the probability of occurrence of 10 has risen up sharply and probability for sum values 4 and below have taken a serious blow. EM has been able to rightly predict the bias.

Case 3

In this case we are assuming all the dices all biased towards even values.

$$X = [1/4, 1/12, 1/4, 1/12, 1/4, 1/12]$$
$$Z = [1/4, 1/12, 1/4, 1/12, 1/4, 1/12]$$

Fig. 5c shows the distribution for this case compared to original dice's distribution. We can see both of them show clear bias towards even values, thus we can say EM has been able to predict the bias.

Variational Inference

Task 2.6: VI

$$\begin{aligned} X_k &\sim \mathcal{N}(\mu_k, \lambda_k^{-1}) \quad \forall k \in [1, K] \\ \mu_k &\sim \mathcal{N}(\mu, \lambda^{-1}) \quad \forall k \in [1, K] \end{aligned}$$

$$\begin{aligned} Z_n &\sim \mathcal{N}(\epsilon_n, \iota_n^{-1}) \quad \forall n \in [1, N] \\ \epsilon_n &\sim \mathcal{N}(\epsilon, \iota^{-1}) \quad \forall n \in [1, N] \end{aligned}$$

$$\begin{aligned} \bar{\epsilon} =& \epsilon_1, ..., \epsilon_N \\ \bar{\mu} =& \mu_1, ..., \mu_K \\ q(\bar{\mu}, \bar{\epsilon}) =& \prod_k q(\mu_k) \prod_n q(\epsilon_n) \end{aligned}$$

$$\begin{aligned} D &= s^1, ..., s^N \\ s^n &= s_1^n, ..., s_n^K \end{aligned}$$

$$\begin{aligned} s_k^n &= X_k^n + Z_k^n \\ s_k^n &\sim \mathcal{N}(\mu_k + \epsilon_n, \lambda_k^{-1} + \iota_k^{-1}) \\ &\qquad \forall k \in [1, K] \ \& \ \forall n \in [1, N] \end{aligned}$$

$$
\begin{aligned}
p(D,\bar{\mu},\bar{\epsilon}) &= p(D|\bar{\mu},\bar{\epsilon})p(\bar{\mu}|\mu)p(\bar{\epsilon}|\epsilon) \\
\ln p(D,\bar{\mu},\bar{\epsilon}) &= \sum_k\sum_n \big[\ln p(s_k^n|\mu_k,\epsilon_n) \ + \ln p(\mu_k|\mu) \ + \ln p(\epsilon_n|\epsilon)\big] \\
&= \sum_k\sum_n \Big[-\frac{1}{2}\frac{\lambda_k\iota_n}{\lambda_k+\iota_n}(s_k^n-\mu_k-\epsilon_n)^2 \ - \ \frac{1}{2}\lambda(\mu_k-\mu)^2 \ - \ \frac{1}{2}\iota(\epsilon_n-\epsilon)^2\Big] \\
\text{Let } \gamma_k^n &= \frac{\lambda_k\iota_n}{\lambda_k+\iota_n} \\
p(D,\bar{\mu},\bar{\epsilon}) &= \sum_k\sum_n \Big[-\frac{1}{2}\gamma_k^n(s_k^n-\mu_k-\epsilon_n)^2 \ - \ \frac{1}{2}\lambda(\mu_k-\mu)^2 \ - \ \frac{1}{2}\iota(\epsilon_n-\epsilon)^2\Big]
\end{aligned}
$$

$$
\begin{aligned}
\ln q(\mu_j) &= \mathbb{E}_{(\bar{\epsilon},\bar{\mu}/\mu_j)} \ln p(D,\bar{\mu},\bar{\epsilon}) + C \\
&= \mathbb{E}_{(\bar{\epsilon},\bar{\mu}/\mu_j)}\Big[\sum_k\sum_n \big[-\frac{1}{2}\gamma_k^n(s_k^n-\mu_k-\epsilon_n)^2 \ - \ \frac{1}{2}\lambda(\mu_k-\mu)^2 \ - \ \frac{1}{2}\iota(\epsilon_n-\epsilon)^2\big]\Big] + C \\
&= \mathbb{E}_{(\bar{\epsilon},\bar{\mu}/\mu_j)}\Big[\sum_n \big[-\frac{1}{2}\gamma_j^n(s_j^n-\mu_j-\epsilon_n)^2 \ - \ \frac{1}{2}\lambda(\mu_j-\mu)^2\big]\Big] + C \\
&= \mathbb{E}_{(\bar{\epsilon},\bar{\mu}/\mu_j)}\Big[\sum_n \big[-\frac{1}{2}(\gamma_j^n+\lambda)\mu_j^2 \ + \ \big(\gamma_j^n(s_j^n-\epsilon_n)+\lambda\mu\big)\mu_j\big]\Big] + C \\
&= \sum_n \big[-\frac{1}{2}(\gamma_j^n+\lambda)\mu_j^2 \ + \ \big(\gamma_j^n(s_j^n-\mathbb{E}(\epsilon_n))+\lambda\mu\big)\mu_j\big] + C \\
&= -\frac{1}{2}\mu_j^2\sum_n(\gamma_j^n+\lambda) \ + \ \mu_j\sum_n\big(\gamma_j^n(s_j^n-\mathbb{E}(\epsilon_n))+\lambda\mu\big) + C \\
&= -\frac{1}{2}\mu_j^2(N\lambda+\sum_n\gamma_j^n) \ + \ \mu_j\Big(N\lambda\mu+\sum_n\gamma_j^n(s_j^n-\mathbb{E}(\epsilon_n))\Big) + C
\end{aligned}
$$

$$
\boxed{
\begin{aligned}
\mathbb{E}(\mu_j) &= \frac{\lambda\mu \ + \ \frac{1}{N}\sum_n\gamma_j^n(s_j^n-\mathbb{E}(\epsilon_n))}{\lambda+\frac{1}{N}\sum_n\gamma_j^n} && \text{(10a)} \\
\mathrm{var}(\mu_j) &= (N\lambda+\sum_n\gamma_j^n)^{-1} && \text{(10b)} \\
q(\mu_j) &\sim \mathcal{N}(\mathbb{E}(\mu_j),\mathrm{var}(\mu_j)) && \text{(10c)}
\end{aligned}
}
$$

$$
\boxed{
\begin{aligned}
\mathbb{E}(\epsilon_j) &= \frac{\iota\epsilon + \ \frac{1}{K}\sum_k\gamma_k^j(s_k^j-\mathbb{E}(\mu_k))}{\iota+\frac{1}{K}\sum_k\gamma_k^j} && \text{(11a)} \\
\mathrm{var}(\epsilon_j) &= (K\iota+\sum_k\gamma_k^j)^{-1} && \text{(11b)} \\
q(\epsilon_j) &\sim \mathcal{N}(\mathbb{E}(\epsilon_j),\mathrm{var}(\epsilon_j)) && \text{(11c)}
\end{aligned}
}
$$

where $\gamma_k^n = \frac{\lambda_k\iota_n}{\lambda_k+\iota_n}$

Task 2.7: VI

$$\begin{aligned} Z_n \ &\sim \mathcal{N}(\epsilon_n, \iota_n^{-1}) \quad \forall n \in [1, N] \\ \epsilon_n \ &\sim \mathcal{N}(\epsilon, \iota^{-1}) \quad \forall n \in [1, N] \end{aligned}$$

$$\begin{aligned} \bar{\epsilon} &= \epsilon_1, ..., \epsilon_N \\ q(\bar{\epsilon}) &= \prod_n q(\epsilon_n) \end{aligned}$$

$$\begin{aligned} D &= s^1, ..., s^L \\ s^l &= s_1^l, ..., s_K^l \\ s_k^g &= \sum_{i \in g} Y_k^i \\ s_k^g &\sim \mathcal{N}\Big(\sum_{i \in g} \epsilon_i, \sum_{i \in g} \iota_i^{-1}\Big) \\ &\qquad \forall k \in [1, K] \ \& \ \forall g \in \{G_1, ..G_L\} \end{aligned}$$

$$\begin{aligned} p(D, \bar{\epsilon}) &= p(D|\bar{\epsilon}) p(\bar{\epsilon}) \\ \ln p(D, \bar{\epsilon}) &= \sum_k \sum_g \Big[\ln p(s_k^g | \bar{\epsilon_n}) \ + \ln p(\bar{\epsilon_n}) \Big] \\ &= \sum_k \sum_g \Big[- \frac{1}{2 \ \sum_{i \in g} \iota_i^{-1}} \big(s_k^n - \sum_{i \in g} \epsilon_i \big)^2 \ - \sum_{i \in g} \frac{1}{2} \iota (\epsilon_i - \epsilon)^2 \Big] \end{aligned}$$

$$\text{Let } I(j \in g) = \begin{cases} 1 & P_j \in g \\ 0 & P_j \notin g \end{cases}$$
$$\forall j \in [1, N] \ \& \forall g \in \{G_1, ..G_L\}$$

$$
\begin{aligned}
\ln q(\epsilon_j) &= \mathbb{E}_{(\bar{\epsilon}/\epsilon_j)} \ln p(D, \bar{\epsilon}) + C \\
&= \mathbb{E}_{(\bar{\epsilon}/\epsilon_j)} \Big[\sum_k \sum_g \big[- \frac{1}{2\ \sum_{i\in g} \iota_i^{-1}} \big(s_k^n - \sum_{i \in g} \epsilon_i \big)^2 \ - \sum_{i\in g} \frac{1}{2} \iota (\epsilon_i - \epsilon)^2 \big] \Big] + C \\
&= \mathbb{E}_{(\bar{\epsilon}/\epsilon_j)} \Big[\sum_k \sum_g I(j \in g) \big[- \frac{1}{2\ \sum_{i\in g} \iota_i^{-1}} \big(s_k^g - \sum_{i \in g} \epsilon_i \big)^2 \ - \frac{1}{2} \iota (\epsilon_j - \epsilon)^2 \big] \Big] + C \\
&= \mathbb{E}_{(\bar{\epsilon}/\epsilon_j)} \Big[\sum_k \sum_g I(j \in g) \big[- \frac{1}{2} \epsilon_j^2 (\iota + \frac{1}{\sum_{i\in g} \iota_i^{-1}}) \ + \ \epsilon_j (\iota \epsilon + \frac{s_k^g - \sum_{i \in g/j} \epsilon_i}{\sum_{i\in g} \iota_i^{-1}}) \big] \Big] + C \\
&= \sum_k \sum_g I(j \in g) \big[- \frac{1}{2} \epsilon_j^2 (\iota + \frac{1}{\sum_{i\in g} \iota_i^{-1}}) \ + \ \epsilon_j (\iota \epsilon + \frac{s_k^g - \sum_{i \in g/j} \mathbb{E}(\epsilon_i)}{\sum_{i\in g} \iota_i^{-1}}) \big] + C \\
&= -\frac{1}{2} \epsilon_j^2 \sum_k \sum_g I(j \in g) \Big[\iota + \frac{1}{\sum_{i\in g} \iota_i^{-1}} \Big] \ + \ \epsilon_j \sum_k \sum_g I(j \in g) \Big[\iota \epsilon + \frac{s_k^g - \sum_{i \in g/j} \mathbb{E}(\epsilon_i)}{\sum_{i\in g} \iota_i^{-1}} \Big] + C \\
&= -\frac{1}{2} \epsilon_j^2 K \sum_g I(j \in g) \Big[\iota + \frac{1}{\sum_{i\in g} \iota_i^{-1}} \Big] \ + \ \epsilon_j K \sum_g I(j \in g) \Big[\iota \epsilon + \frac{\frac{1}{K} \sum_k s_k^g - \sum_{i \in g/j} \mathbb{E}(\epsilon_i)}{\sum_{i\in g} \iota_i^{-1}} \Big] + C
\end{aligned}
$$

$$
\boxed{
\begin{aligned}
\mathbb{E}(\epsilon_j) &= \frac{\sum_g I(j \in g) \Big[\iota \epsilon + \frac{\frac{1}{K} \sum_k s_k^g - \sum_{i \in g/j} \mathbb{E}(\epsilon_i)}{\sum_{i\in g} \iota_i^{-1}} \Big]}{\sum_g I(j \in g) \Big[\iota + \frac{1}{\sum_{i\in g} \iota_i^{-1}} \Big]} && \text{(12a)} \\
\mathrm{var}(\epsilon_j) &= \Big(K \sum_g I(j \in g) \Big[\iota + \frac{1}{\sum_{i\in g} \iota_i^{-1}} \Big] \Big)^{-1} && \text{(12b)} \\
q(\epsilon_j) &\sim \mathcal{N}(\mathbb{E}(\epsilon_j), \mathrm{var}(\epsilon_j)) && \text{(12c)}
\end{aligned}
}
$$