Statistical Machine learning _ TH2 Sols

1)

a. lof (left order operation) $\longrightarrow$ sort columns based on their binary number representation

b.1 The Process will be exchangeable because lof is invariant to column operations.

i.e If we change the order of variable, the lof transformation wont change

b.2 Actually the Proof is so simple.
The first customer chooses a Pois(α) numbers of new dishes.
By exchangeability all customers must also choose Pois(α) number of new dishes
Since we can always specify an ordering on customers which begins with a particular customer ■

but you are welcome to solve the problem with other approaches like induction and ... .

b.3 we first try to solve the limited case, then calculate the limit $K \longrightarrow \infty$

$$E[\# \text{ones}] = E[1^T Z 1] = E\left[\sum_{i,K} z_{iK}\right]$$

$$E[\# \text{ones}] = K \cdot E[1^T z_j] \Rightarrow \text{since each column is independant}$$

$$E[1^T z_j] = E\left[\sum_i z_{ij}\right] = \sum_i E[z_{ij}] = \sum_i \underbrace{\int_0^1 \pi_j \, P(\pi_j) \, d\pi_j}_{\text{Beta distribution}} = N \frac{\alpha/K}{1+\alpha/K}$$

$\pi_K \sim Beta(\alpha/K, 1)$

$z_{iK} \sim Bernouli(\pi_K)$

$$\Longrightarrow \mathbb{E}[\#\text{ones}] = \frac{N\alpha}{1+\alpha/K} \rightsquigarrow \mathbb{E}[\#\text{ones}]\Big|_{K\to\infty} = \frac{N\alpha}{1+0} = N\alpha \quad \blacksquare$$

2) This is a very famous result and a lot of approaches have been suggested to prove this theorem.

e.g theorem 1. Ferguson / theorem 4.3 Sethuraman and ... which all of them are acceptable.

But here we try to suggest the easiest one.

Ghosal_ Van der vaart : $X_1, \dots, X_n \sim G$, $G \sim DP(\alpha) \to G \mid X_{1:n} \sim ?$

from the tail_free / self_similarity propertie of DP : $\mathbb{P}(G \mid X_{1:n}) \triangleq \mathbb{P}(G \mid N)$

$\triangleq$: same distribution

N: vector of counts.

$$N_j = \sum_i \delta_{X_i}(A_j)$$

Bayes rule $\to$ $\mathbb{P}(G \mid X_{1:n}) \triangleq \mathbb{P}(G \mid N) \propto \underbrace{\mathbb{P}(G)}_{Dp(\alpha)} \underbrace{\mathbb{P}(N \mid G)}_{\text{likelihood : Multinomial}}$

from congugacy propertie of Dirichlet distribution and Multinomial distribution.

if $P \sim Dir(\alpha)$ and $N/P \sim M(n) \Longrightarrow P/N \sim Dir(\alpha + N)$

$\to \mathbb{P}(G / X_{1:n}) \sim Dir(\alpha + N)$

$\hookrightarrow$ meaning that $\mathbb{P}(G \mid X_{1:n}) \sim Dir\left(\alpha + \sum_i \delta_{X_i}\right) \quad \blacksquare$

Statistical Machine Learning - TH2 Sols

3)

a. Calculating the Expected number of occupied tables

$P(\text{new table on ith draw}) = \frac{\alpha}{\alpha+i-1}$, $W_i = \begin{cases} 1 & \text{if ith draw is a new table} \\ 0 & \text{o.w} \end{cases}$

Indicator Random variable

$\Longrightarrow E[\#\text{ occupied tables}] = E[\sum_i W_i] = \sum_{i=1}^{n} \frac{\alpha}{\alpha+i-1} = \alpha \sum_{j=0}^{n-1} \frac{1}{\alpha+j}$ (If you want know to how solve such equations study Harmonic series)

$\sum_{j=0}^{n-1} \frac{1}{\alpha+j} = \Psi(n+\alpha) - \Psi(\alpha)$, $\Psi(x) \approx \log(x)$

$\hookrightarrow \Psi$: Digamma function

$\Longrightarrow E[\sum W_i] \approx \alpha \log\left(\frac{\alpha+n}{\alpha}\right)$ ■

b. $W = (1,1,0,0,1)$

$P(W=w) = \frac{\alpha}{\alpha} \cdot \frac{\alpha}{1+\alpha} \cdot \frac{2}{2+\alpha} \cdot \frac{3}{3+\alpha} \cdot \frac{\alpha}{4+\alpha} = \frac{6\alpha^3 \Gamma(\alpha)}{\Gamma(\alpha+n)}$ ■

c. $K_n = \#$ tables in n draws

$P(K_n=k) = \frac{\alpha}{\alpha+n-1} P(K_{n-1}=k-1) + \frac{n-1}{\alpha+n-1} P(K_{n-1}=k) \longrightarrow$ So, the problem is now to solve this recursive equation (of course by induction over n)

n-th client sits on a new table

n-th client sits on current table

For the base case: $P(K_1=k) = \begin{cases} 1 & k=1 \\ 0 & \text{o.w} \end{cases}$

$P(K_1=k) = S(1,k) \frac{\alpha^k \Gamma(\alpha)}{\Gamma(\alpha+1)}$ $\qquad S(1,k) = \begin{cases} 1 & k=1 \\ 0 & \text{o.w} \end{cases}$

$\longrightarrow P(K_1 = K) = P(K_1 = 1) = \frac{\alpha \Gamma(\alpha)}{\Gamma(\alpha+1)} = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1)} = 1$ ✓

Now, we assume that equation holds for $n-1$ and then use the calculated recursive equation

$$P(K_n = k) = \frac{\alpha}{\alpha+n-1} S(n-1, k-1) \frac{\alpha^{k-1}\Gamma(\alpha)}{\Gamma(\alpha+n-1)} + \frac{n-1}{\alpha+n-1} S(n-1,k) \frac{\alpha^k \Gamma(\alpha)}{\Gamma(\alpha+n-1)}$$

$$= \frac{\alpha^k}{\alpha+n-1} \frac{\Gamma(\alpha)}{\Gamma(\alpha+n-1)} \Big[\underbrace{S(n-1,k-1) + (n-1) S(n-1,k)}_{S(n,k) \,::\, \text{stirling Propertie}} \Big]$$

$$= \frac{\alpha^k \Gamma(\alpha)}{\Gamma(\alpha+n)} S(n,k) \blacksquare$$

But its fine to induct over K as well.