1 Probability and independence

a) Decomposition

We aim to validate

$$(X \perp Y, W \mid Z) \implies (X \perp Y \mid Z) \tag{1.1}$$

Proof. We suppose the statement $(X \perp Y,\ W \mid Z)$ is true. It follows from the definition of the conditional independence that $p(x, y, w \mid z) = p(x \mid z)p(y, w \mid z)$ for all $x \in \Omega_x$, $(y, w) \in \Omega_y \times \Omega_w$ and $z \in \Omega_z$. We then consider the marginalize $p(x, y, w \mid z)$:

$$\begin{aligned}
p(x, y|z) &= \sum_{w\, \in\, \Omega_w} p(x, y, w \mid z) \\
&= \sum_{w\, \in\, \Omega_w} p(x \mid z)p(y, w \mid z) \\
&= p(x \mid z) \sum_{w\, \in\, \Omega_w} p(y, w \mid z) \\
&= p(x \mid z)p(y \mid z)
\end{aligned}$$

from which we conclude that $(X \perp Y \mid Z)$ $\square$. By symmetry of the argument, we can show that $(X \perp W \mid Z)$ is true as well.

b)

We aim to validate

$$(X \perp Y \mid Z) \text{ and } (X, Y \perp W \mid Z) \implies (X \perp W \mid Z) \tag{1.2}$$

Proof. Suppose $(X, Y \perp W \mid Z)$ and $(X \perp Y \mid Z)$ are true. We know from the symmetry and decomposition properties of the conditional independence that $(X, Y \perp W \mid Z) \implies (W \perp X, Y \mid Z) \implies (X \perp W \mid Z)$. Therefore $(X \perp W \mid Z)$ is true $\square$.

c)

We aim to validate

$$(X \perp Y, W \mid Z) \text{ and } (Y \perp W \mid Z) \implies (X, W \perp Y \mid Z) \tag{1.3}$$

Proof. Suppose $(X \perp Y, W \mid Z)$ is true. Then it follows from the definition of conditional independence that

$$p(x, y, w \mid z) = p(x \mid z)p(y, w \mid z)$$

Then assume $(Y \perp W \mid Z)$ is true. The second factor can be factorized

$$p(x, y, w \mid z) = p(x \mid z)p(y \mid z)p(w \mid z)$$

From the decomposition property, we know $(X \perp W \mid Z)$ is true. Thus

$$p(x, y, w \mid z) = p(x, w \mid z)p(y \mid z)$$

From which we conclude $(X, W \perp Y \mid Z)$ is true $\square$.

d)

We aim to validate

$$(X \perp Y \mid Z) \text{ and } (X \perp Y \mid W) \implies (X \perp Y \mid Z, W) \tag{1.4}$$

Counter example. We consider the following R.V.

1. X: Person A arrive late for diner;
2. Y: Person B arrive late for diner;
3. W: They come from the same city;
4. Z: They work in the same city.

For this situation, we see that X and Y are conditionally independent when given either W or Z. If we know they are from the same city, then they might work in different cities and take different route home. Thus knowing person A was late doesn't inform us on the probability of person B to arrive late.

A similar argument can be made for $(X \perp Y \mid Z)$.

Thus the LHS of the proposition is true, yet the RHS is clearly false in our case. Assuming we were given that W and Z are true, then we are given the geolocalisation of person A and B. If we were given that person A would be late for diner, then we'd be able to make a good guess that person B would be late as well (they would both be impacted by the same traffic jam or whatnot). Thus the proposition is false.

2 Bayesian inferance and MAP

Let $\mathbf{X}_1, \dots, \mathbf{X}_n \mid \boldsymbol{\pi} \overset{\text{iid}}{\sim} \text{Multinomial}(1, \boldsymbol{\pi})$ on k element. The values are sampled from a set of cardinality 2, that is $x_j^{(i)} \in \{0, 1\}$. Each R.V. has only one non-zero entry for a given trial, that is $\sum_{j=1}^k x_j^{(i)} = 1$.

We assume a Dirichlet prior $\boldsymbol{\pi} \sim \text{Dir}(\boldsymbol{\alpha})$ with a PDF

$$p(\boldsymbol{\pi} \mid \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^k \alpha_j)}{\prod_{j=1}^k \Gamma(\alpha_j)} \prod_{j=1}^k \pi_j^{\alpha_j - 1}$$

a)

Since the data is IID, they are mutually independent by definition. Being given the parameters of their Multinomial distribution (or a subset for that matter) does not change the independence of the $\mathbf{X}$'s. Thus,

$$(\mathbf{X}_i \perp \mathbf{X}_j \mid \boldsymbol{\pi}) \; \forall \; (i,j) \; \in \{1, \dots, k\} \times \{1, \dots, k\}$$

Of course, none of the vector can be mutually nor conditionally independent to $\boldsymbol{\pi}$ since it contains information about the distribution of the one hot vectors $\mathbf{X}_i$. In this case $\boldsymbol{\pi}$ are the probabilities of one of the k entry to be equal to one. Even giving one of these away is enough to impact the posterior distribution of the conditional $p(x_i \mid x_\ell, \pi_j)$ for example.

b)

The posterior distribution $p(\boldsymbol{\pi} \mid x_1, \dots, x_n)$ is computed via the Bayes rule

$$p(\boldsymbol{\pi} \mid \mathbf{x}_{1:n}) = \frac{p(\mathbf{x}_{1:n} \mid \boldsymbol{\pi}) p(\boldsymbol{\pi})}{p(\mathbf{x}_{1:n})}$$

where $p(\boldsymbol{\pi}) = p(\boldsymbol{\pi} \mid \boldsymbol{\alpha})$ is the prior for $\boldsymbol{\pi}$ defined above. For the sakes of determining the posterior distribution nature, we can postpone the derivation of the marginal likelihood. The likelihood is given by the product of n Multinomial PMF given individually by

$$p(\mathbf{x}_i \mid \boldsymbol{\pi}) = \frac{1}{\prod_{\ell=1}^k x_\ell^{(i)}!} \prod_{j=1}^k \pi_j^{x_j^{(i)}}$$

Since $\sum_{j=1}^k x_j^{(i)} = 1$, then the product in the denominator is unity. Therefore, the posterior must be

$$p(\boldsymbol{\pi} \mid \mathbf{x}_{1:n}) \propto \prod_{i=1}^n \prod_{j=1}^k \pi_j^{x_j^{(i)}} \left(\frac{\Gamma(\sum_{\ell=1}^k \alpha_\ell)}{\prod_\ell^k \Gamma(\alpha_\ell)} \prod_{\ell=1}^k \pi_\ell^{\alpha_\ell - 1} \right)$$

We use the fact that we can swap around product operator for real numbers and ignore the Dirichlet normalizing constant for the sake of the argument

$$p(\boldsymbol{\pi} \mid \mathbf{x}_{1:n}) \propto \prod_{i=1}^n \prod_{j=1}^k \prod_{\ell=1}^k \pi_\ell^{\alpha_\ell - 1} \pi_j^{x_j^{(i)}}$$

We can readily see that the resulting distribution will be a Dirichlet with updated α_ℓ's.

The posterior distribution is a <u>Dirichlet</u> distribution with parameters $\alpha_j' = \alpha_j + \sum_{i=1}^n x_j^{(i)}$.

c) Marginal Likelihood

The marginal likelihood $p(\mathbf{x}_{1:n})$ is a normalizing constant defined as the integral of the numerator over all instantiation of $\boldsymbol{\pi}$

$$p(\mathbf{x}_{1:n}) = \int_{\boldsymbol{\Delta_k}} p(\mathbf{x}_{1:n} \mid \boldsymbol{\pi}) p(\boldsymbol{\pi}) d^{(k)}\boldsymbol{\pi}$$

where $\boldsymbol{\Delta}_k$ is the probability simplex. In term of the quantities defined above, this is

$$p(\mathbf{x}_{1:n}) = \int_{\boldsymbol{\Delta_k}} d^{(k)}\boldsymbol{\pi} \prod_{i=1}^{n}\prod_{j=1}^{k} \pi_j^{x_j^{(i)}} \left(\frac{\Gamma(\sum_{\ell=1}^{k}\alpha_\ell)}{\prod_{\ell=1}^{k}\Gamma(\alpha_\ell)} \prod_{\ell=1}^{k} \pi_\ell^{\alpha_\ell - 1} \right)$$

The π_j's are independent variables since the simplex $\boldsymbol{\Delta}_k$ is crucially defined as an affine plane in an Euclidian space which is supported by a set of orthonormal vectors. Thus, our task is to evaluate k identical integrals of the form

$$\int_0^1 d\pi_j \prod_{i=1}^{n} \pi_j^{x_j^{(i)} + \alpha_j - 1} = \left(\sum_{i=1}^{n} x_j^{(i)} + \alpha_j \right)^{-1}$$

Thus

$$\boxed{p(\mathbf{x}_{1:n}) = \frac{\Gamma(\sum_{\ell=1}^{k}\alpha_\ell)}{\prod_{\ell=1}^{k}\Gamma(\alpha_\ell)} \prod_{j=1}^{k} \left(\sum_{i=1}^{n} x_j^{(i)} + \alpha_j \right)^{-1}}$$