0.0.1. Definiciones

- Un ***producto*** es la *unidad básica discreta de los datos*, se define como un ítem de un nomenclador (SITC). Se representa como un vector unitario. Definimos el superíndice i del vector como el i-ésimo producto del nomenclador y el i-ésimo elemento del vector. El V-ésimo producto del nomenclador es el vector w, tal que w^v=1 y w^u=0, $u \neq v$
- Un ***país*** es una secuencia de **N** productos, definido como $W = (w_1, w_2, ..., w_N)$
- Nuestro **corpus** es la colección de **M** países, definido como $D = (d_1, d_2, ..., d_M)$
- Un ***componente*** es una dimensión latente sobre el corpus, y suponemos una cantidad fija k de los mismos.
- Nuestro objetivo es obtener:

1. Una distribución de componentes sobre cada país.
2. Una distribución de los productos sobre los componentes.

0.1. Dimensión latente

swissRoll.png

Figura 1: Swiss Roll

- Nuestros datos están definidos en una dimensión alta

$$\mathcal{R}^{N*P*Y*T}$$

- Podemos encontrar un espacio de menor dimensión en el que se logra representarlos bien

0.2. Proceso generativo

Intuitivamente, suponemos el siguiente proceso generativo de datos:

- Para cada país del corpus, imaginamos que las exportaciones surgen de un proceso de dos etapas:
    - Elegimos aleatoriamente una distribución sobre los componentes
    - Para cada dólar exportado:
        - Elegimos aleatoriamente el componente al que pertence, dada la distribución definida en el paso anterior
        - Elegimos aleatoriamente un producto de la distribución correspondiente a dicho componente

1. Para cada Componente $k \in \{1, 2, ...K\}$

- Generar una distribución sobre los componentes $\beta \sim Dir_v(\eta)$ con $\eta \in \mathcal{R}_{>\prime}$ un parámetro fijo

2. Para cada país $d \in \{1, 2, ...D\}$

- generar un vector de proporciones de componentes $\theta_d \sim Dir_K(\alpha)$ con $\alpha \in \mathcal{R}^{\mathcal{K}}_{>\prime}$ un parámetro fijo
- Para cada dólar exportado:
    - I. generar una asignación del componente $z_{dn} \sim Mult(\theta_d)$
    - II. asignar el producto $w_{dn} \sim Mult(\beta_{zn})$

0.3. Representación gráfica

Figura 2: [?]

- Cada nodo representa una distribución de probabilidad.
- La arista significa que la distribución de salida define los parámetros de la distribución de entrada
- Los recuadros significan replicación:
  - El recuadro interior representa que el proceso se realiza para cada dólar exportado en el país.
  - El recuadro exterior representa que el proceso se realiza para cada país en el corpus.

0.4. Distribución de Dirichlet

Un Proceso de Dirichlet es una familia de procesos estocásticos donde **las realizaciones son ellas mismas distribuciones de probabilidad.**

Es decir, el rango de esta distribución (así como en una normal son los reales) son distribuciones de probabilidad.

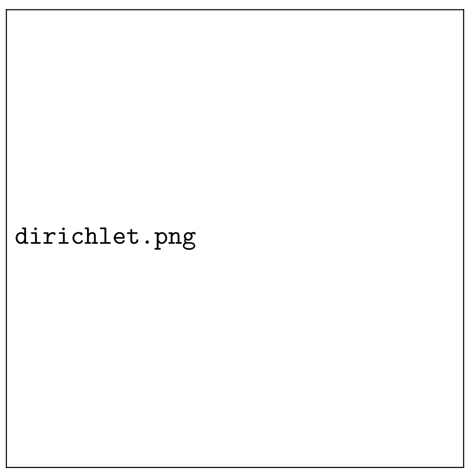

Figura 3: [**?**]

Para interpretarlo geométricamente, podemos pensar un ejemplo de la distribución de densidad para **3 productos** y **4 componentes**.

- El triángulo representa todas las distribuciones (multinomiales) posibles sobre los tres productos
- Cada uno de los vértices del triángulo es una distribución de probabilidad que asigna una probabilidad de 1 a uno de los productos.
- El punto medio de cada lado, es una distribución con probabilidad 0.5 a dos componentes.

- El cuarto componente, el centroide del triángulo, asigna probabilidad de $\frac{1}{3}$ a cada producto.
- Los cuatro puntos marcado con x son las distribuciones multinomiales de p(w|z) para cada uno de los cuatro componentes.
- La altura en el eje z es una posible distribución de densidad sobre el simplex, es decir, sobre las distribuciones de densidad multinomiales, dada por LDA

0.5. Inferencia

Cuando observamos los datos, no contamos con los tópicos ni con su distribución, sino con los productos y países. El objetivo es realizar inferencia sobre las variables latentes, mediante el Teorema de Bayes:

$$p(\theta, z|w, \alpha, \beta) = \frac{p(\theta, z, w|\alpha, \beta)}{p(w|\alpha, \beta)}$$

Nuestra función objetivo es:

$$\ell(\alpha, \beta) = \sum_{d=1}^{M} \log p(w_d|\alpha, \beta)$$

El problema es que esta ecuación es intratable, por la interacción entre θ y β

Por ello, la inferencia se realiza sobre una familia de modelos que se sabe que son una cota inferior de probabilidad, y que son tratables.

- Estos modelos tienen parámetros variacionales, que se ajustan para obtener el modelo que más se acerca a la cota inferior.
- La forma de obtener una familia de modelos tratables es considerar algunas modificaciones sobre el modelo gráfico original, removiendo nodos y aristas.

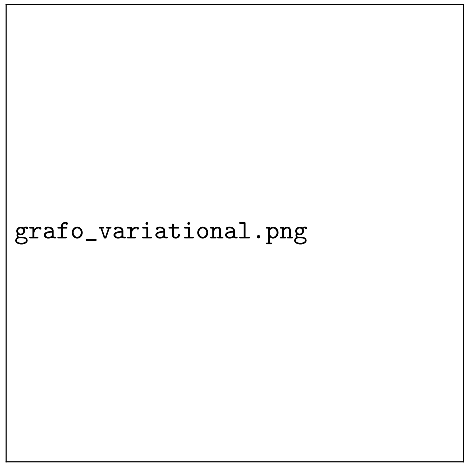

Figura 4: [?]

0.6. Estimación de parámetros

Podemos aproximar los parámetros alternando el proceso de *variational Expectation Maximization* (EM):

- **paso E**: Optimizamos los parámetros variacionales γ, φ
- **paso M**: Para los valores fijos γ, φ, maximizamos la cota inferior respecto a los parámetros del modelo, α, β

Estos dos pasos se alternan hasta converger.

Por último, se realiza un suavizado sobre las probabilidades que un componente asigna a un producto, para que sean siempre mayores a 0.