

Sesion 13: Diversificación de riesgos vía modelos jerárquicos

Juan Carlos Martínez-Ovando

Otoño 2019

1. Motivación

Hemos platicado reiteradamente en el curso acerca de la dos propiedades asociadas con los modelos actuariales:

- ▶ homogeneidad,
- ▶ no ambigüedad.

Estas propiedades son de suma relevancia para el calculo de probabilidades y para el proceso inferencial asociado.

Sin embargo, aunque estas dos propiedades son deseables y proveen simplificaciones estructurales en los modelos que resultan adecuados, *modelos de probabilidad basados en estas dos propiedades pueden ser sobrepasados por aspectos estilizados de los datos en problemas reales.*

Circunstancias donde los *datos puedan segmentarse de manera natural (intrínseca conocida)* son un ejemplo donde estas dos propiedades se ven limitadas.

Tal segmentación puede obedecer a:

- a) Patrones geográficos
- b) Aspectos socioeconómicos
- c) Aspectos temporales
- d) Relaciones de asociación/conectividad/interacción
- e) etc.

Con el *propósito de proveer un esquema de modelación que cubra aspectos conceptuales intrínsecos como los anteriores*, se han propuesto **modelos de probabilidad segmentados**:

- ▶ **Modelos estructurales** (en estadística frecuentista y econometría clásica)
- ▶ **Modelos jerárquicos** (en el contexto bayesiano)
- ▶ **Modelos de grafos dirigidos estructurales** (en el contexto de ingeniería en computación y sistemas cognitivos)

La **idea fundamental** detrás de estos modelos es la combinar y balancear dos aspectos de modelación:

- i) Preservar no ambigüedad y homogeneidad al interior de los segmentos
- ii) Distinguir elementos de heterogeneidad entre segmentos.

Con esto se garantiza una medición conciliada del desconocimiento del *mismo problema/fenómeno*, aunque de manera diferenciada entre segmentaciones.

Con base en esto, definimos la siguiente especificación del modelo.

2. Especificación

Consideremos el caso donde Y mide eventos binarios con $\mathcal{X} = \{0, 1\}$, *pensemos en la incidencia individual de siniestros de un portafolio de seguros.*

Siendo Y la variable de interés, consideremos a su vez que podemos indizar *observaciones* y *variables aleatorias* en dos dimensiones ki , donde

- ▶ $k = 1, \dots, K$ hace referencia a la etiqueta de K *segmentos intrínsecos* del problema, e
- ▶ $j = 1, \dots, J_k$ hace referencia a la etiqueta de las *unidades de medición* dentro de cada segmento. En el contexto de seguros, J_k representa la suscripción en el segmento k .

El **aspecto común** entre los segmentos reside en imponer un **aspecto estructural común** entendido como la forma de medir de la incertidumbre (a.k.a. forma estructural del modelo de probabilidad); en este caso mediante el modelo Bernoulli.

El **aspecto direfenciador entre segmentos** reside en distinguir que el aspecto que completa la especificación de los modelos al interior de los segmentos (a.k.a. parámetros) son distintos entre sí.

De esta forma, en el contexto descrito, resulta:

$$Y_{kj}|\theta_k \sim Ber(y|\theta_k),$$

con

$$\mathbb{P}(Y_{kj} = 1|\theta_k) = \theta_k,$$

para $j = 1, \dots, J_k$ y $k = 1, \dots, K$, donde

$$0 < \theta_k < 1.$$

Dos implicaciones derivadas de la especificación anterior se siguen:

- l) **Dentro de los segmentos**, los componentes del bloque $\mathbf{Y}_k = (Y_{kj})_{j=1}^{J_k}$ son condicionalmente i.i.d. dado θ_k , con

$$\mathbb{P}(Y_{k1} \in A_1, \dots, Y_{kJ_k} \in A_{J_k}) = \prod_{j=1}^{J_k} \text{Ber}(A_j | \theta_k),$$

siendo los conjuntos $(A_j)_{j=1}^{J_k}$ cualquier subconjunto de $\mathcal{X} = \{0, 1\}$.

Para $k = 1, \dots, K$.

II) **Entre segmentos**, los bloques

$$\mathbf{Y}_k, \dots, \mathbf{Y}_K,$$

son ind condicionalmente en

$$\theta_1, \dots, \theta_K,$$

(el bloque completo de parámetros).

Es decir, combinando (I) y (II) tenemos,

$$\mathbb{P}(\mathbf{Y}_1, \dots, \mathbf{Y}_K | \theta_1, \dots, \theta_K) = \prod_{k=1}^K \prod_{j=1}^{J_k} \text{Ber}(Y_{kj} | \theta_k).$$

Los parametros del modelo, en este caso, son

$$\theta_1, \dots, \theta_K,$$

en lugar de un solo θ como en el *modelo completamente homogeneo*.

2.1. Simetría estocastica

La dualidad referida en (I) y (II) da origen a la noción de **intercambiabilidad parcial**, en la que

$$\begin{aligned} & \mathbb{P}(Y_{kj} \in A_{kj} : k = 1, \dots, K; j = 1, \dots, J_k) = \\ &= \int_{\Gamma} \int_{(0,1)^K} \prod_{k=1}^K \left(\prod_{j=1}^{J_k} \text{Ber}(Y_{kj} \in A_{kj} | \theta_k) \right) Q(d\theta_1, \dots, d\theta_K | \gamma) \Pi(d\gamma), \end{aligned}$$

donde

- ▶ $Q(d\theta_1, \dots, \theta_k | \gamma)$ denota la incertidumbre entorno a las características particulares de los segmentos,
- ▶ $\gamma \in \Gamma$ representa el conocimiento experto característico común entre los segmentos,
- ▶ $\Pi(d\gamma)$ cuantifica el **conocimiento suplementario** común a todos los segmentos (i.e. conocimiento suplementario sobre el fenómeno común).

En particular, en el contexto que nos compete,

$$Q(d\theta_1, \dots, d\theta_K | \gamma) = \prod_{k=1}^K Be(\theta_k | \alpha, \beta),$$

siendo

$$\gamma = (\alpha, \beta),$$

con

$$\Gamma = (0, \infty)^2,$$

el espacio parametral.

Adicionalmente, la informacion/conocimiento acerca de (α, β) se cuantifica con una distribucion de probabilidad con,

$$\Pi(\alpha, \beta) = \Pi(\alpha) \times \Pi(\beta),$$

la cual es referida como **prior** o **distribución inicial**.

Esta distribucion tiene soporte en \mathbb{R}_+^2 .

2.2. Comentarios

- ▶ Los parámetros $\theta_1, \dots, \theta_K$ se convierten en parámetros/variables latentes, por tener una distribución intrínseca dada por $Q(\cdot|\gamma)$.
- ▶ Aunque diferenciado, la **versión marginal/reducida** del modelo para cada segmento es la misma, i.e. para todo $k = 1, \dots, K$ tenemos el mismo modelo,

$$\mathbb{P}(Y_{k1} \in A_1, \dots, Y_{kn} \in A_n) = \mathbb{P}(Y_{k'1} \in A_1, \dots, Y_{k'n} \in A_n),$$

para todo $k \neq k'$, y para todo n .

- ▶ **Sin embargo**, es posible ver que condicional en γ el modelo es diferente entre segmentos, i.e.

$$\mathbb{P}(Y_{k1} \in A_1, \dots, Y_{kn} \in A_n | \gamma) \neq \mathbb{P}(Y_{k'1} \in A_1, \dots, Y_{k'n} \in A_n | \gamma),$$

para todo $k \neq k'$ y para todo n .

- ▶ *Así, el modelo permite dar una visión diferenciada de la incertidumbre entre segmentos, condicional en (α, β) .*

3. Inferencia

La inferencia en esta clase de modelos con base en un conjunto de **datos**

$$(y_{kj})_{j=1,\dots,J_k},$$

para $k = 1, \dots, K$, se basa en estimar o hacer inferencia sobre los **parámetros**

$$\gamma = (\alpha, \beta),$$

y las **variables latentes**

$$\theta = (\theta_1, \dots, \theta_K).$$

Los datos, en este caso son típicamente leídos en R o Python como un `data.frame` donde una columna representa las etiquetas de los J segmentos y otra columna representa las mediciones y_{ji} s.

Revisen la estructura de los datos en la tabla `AllState`.

3.1. Verosimilitud integrada

La estimación se basa en la **verosimilitud integrada** que se define de manera simple como,

$$l(\gamma, \boldsymbol{\theta} | \mathbf{y}) = \prod_{k=1}^K \prod_{j=1}^{J_k} \text{Ber}(y_{kj} | \theta_k).$$

- Noten que ésta no depende del parámetro γ explícitamente, sino implícitamente a través de las θ_j s.

3.2. Aprendizaje

El aprendizaje, bajo la perspectiva bayesiana, consiste en actualizar el conocimiento acerca de γ y θ , empleando el teorema de Bayes, como

$$q(\gamma, \theta | \mathbf{y}) \propto l(\gamma, \theta | \mathbf{y}) \text{Be}(\theta_1 | \gamma) \cdots \text{Be}(\theta_J | \gamma) \cdot \pi(\gamma), \quad (1)$$

con $\gamma = (\alpha, \beta)$.

- La constante de normalización, en este caso, no puede obtenerse de manera analítica cerrada; requiriendo el uso de algoritmos numéricos. —Lo mismo sucede en el enfoque frecuentista de inferencia—.

3.3. Algoritmo numérico

Bajo el enfoque bayesiano de inferencia, la actualización $q(\gamma, \boldsymbol{\theta}|\mathbf{y})$ se puede obtener empleando el algoritmo *Gibbs sampler*, en el que iterativamente se obtienen valores

$$\left\{ \gamma^{(m)}, \boldsymbol{\theta}^{(m)} \right\}_{m=1}^M,$$

para un M grande fijo.

Los datos que el algoritmo genera presumiblemente representan una *muestra estocástica* de

$$q(\gamma, \boldsymbol{\theta}|\mathbf{y}).$$

La contraparte frecuentista de este algoritmo es el algoritmo E-M (*expectation*, para las variables latentes, y *maximization*, para los parámetros).

3.4. Gibbs sampler

p.0) Definimos valores iniciales (arbitrarios) para θ y γ , i.e.

$$\theta^{(0)} \in (0, 1)^J,$$

y

$$\gamma^{(0)} \in (0, \infty)^2,$$

recordemos que $\gamma = (\alpha, \beta)$ son los parametros de la distriucion Beta para las θ_{ks} .

p.1) Iterativamente, para $m \geq 1$ se siimulan los **parametros** y **variables latentes** como,

$\gamma^{(m)}$ con base en \mathbf{y} y $\theta^{(m-1)}$ mediante

$$\gamma | \mathbf{y}, \theta^{(m-1)} \sim \pi(\gamma | \mathbf{y}, \theta^{(m-1)}) \propto \prod_{k=1}^K \text{Be} \left(\theta_k^{(m-1)} | \gamma \right) \pi(\gamma),$$

- Nota:- En este caso, no se cumple homogeneidad, por lo que usamos un subalgoritmo **slice sampler**

Con base en \mathbf{y} y $\gamma^{(m)}$ se genera $\boldsymbol{\theta}^{(m)}$ mediante el siguiente ciclo para $k = 1, \dots, K$,

$$\theta_k^{(m)} | \boldsymbol{\theta}_{<k}^{(m)}, \boldsymbol{\theta}_{>k}^{(m-1)}, \gamma^{(m)}, \mathbf{y} \sim q(\theta_k | \mathbf{y}, \gamma^{(m)}),$$

donde

$$q(\theta_k | \mathbf{y}, \gamma^{(m)}) \propto \prod_{j=1}^{J_k} \text{Ber}(y_{kj} | \theta_k) \text{Be}(\theta_k | \gamma^{(m)}).$$

- En este caso, $q(\theta_j | \mathbf{y}, \gamma^{(m)})$ es conocida de manera analítica cerrada por ser $\text{Ber}(y_{ji} | \theta_j)$ y $\text{Be}(\theta_j | \gamma^{(m)})$ una **pareja conjugada**.

El orden de las subiteraciones para **parametros** y **variables latentes** es indistinto, el algoritmo garantiza convergencia en este todo caso.

Con base en los datos simulados

$$\left\{ \gamma^{(m)}, \boldsymbol{\theta}^{(m)} \right\}_{m=1}^M,$$

podemos generar inferencia y, sobre todo, prediccion con base en aproximaciones de Monte Carlo, i.e.

$$\mathbb{E}_{\mathbb{P}} \{ \phi(Y, \boldsymbol{\theta}, \gamma) \} \approx \frac{1}{M} \sum_{m=1} \phi \left(Y^{(m)}, \boldsymbol{\theta}^{(m)}, \gamma^{(m)} \right),$$

siendo $\{Y^{(m)}\}$ datos simulados complementariamente, en caso de ser requerido.

4. Comentarios

- ▶ El proceso de diversificación que acabamos de describir permite diversificar la exposición al riesgo en subconjuntos de un portafolio de seguros.
- ▶ La diversificación brinda la posibilidad de tener una lectura depurada del riesgo, por lo que la previsión de riesgos es más confiable.
- ▶ La diversificación de riesgos que comentamos es extendible a los montos de siniestros, no solo a la incidencia de siniestros.
- ▶ Esta diversificación está en función de un conjunto de **variables intrínsecas conocidas**. Este procedimiento se conoce como aprendizaje supervisado. Existirá la posibilidad de diversificar el riesgo con **características no observables**, en cuyo caso estaremos trabajando con aprendizaje no supervisado.

5. Ejercicios

1. Revisen el código en R que implementa el *Gibbs sampler* dentro de la carpeta `scripts`
2. Experimenten cambiando datos o especificaciones del código
3. Contextualicen un modelo jerárquico semejante para **montos individuales** o **frecuencias agregadas en el contexto de riesgo colectivo**.

Lecturas suplementarias

- ▶ Porrini, *Risk Classification Efficiency and the Insurance Market Regulation*. `act11302_sesion_13_suplemento.pdf`