Introducción al Análisis de Datos Bayesiano

Se consideran tres pasos:

- Establecer un modelo probabilístico completo: Distribución de Probabilidades conjunta para todas las cantidades observables y no-observables.
- Condicionar sobre los datos observados: Calcular e interpretar la distribución a posteriori (probabilidad condicional de las cantidades no-observadas, dados los datos observados).
- Evaluar el ajuste del modelo y las implicaciones de la distribución a posteriori resultante.

Si es necesario se altera o se expande el modelo resultante y se repiten los pasos 1 a 3.

Notación General para Inferencia Estadística

Objetivo: Obtener conclusiones a partir de datos númericos sobre cantidades no observadas.

Ejemplo: Prueba Clínica. Se desea estimar la probabilidad de sobrevivencia de pacientes con cancer en cinco años dada una nueva droga. Comparación de esta probabilidad con la probabilidad de sobrevivencia de los pacientes a un tratamiento estándard. Inferencias sobre estas probabilidades se basa en una muestra de pacientes (razones éticas).

Cantidades a estimar o estimandos (sobre lo que queremos hacer inferencia estadística:

- Cantidades no observables directamente (Ejemplo: Parámetros de un modelo de regresión)
- Potencialmente observables (Ejemplo: Posible resultado de un tratamiento no recibido)

- Parámetros, Datos y Predicciones:
  - θ: Cantidades vectoriales no observables o parámetros poblacionales de interés (Ej. Probabilidad de sobrevivencia a cada tratamiento)
  - y: Datos observados (Ej.: No. de sobrevivientes y muertes en cada grupo)
  - $\tilde{y}$: Cantidad desconocida pero potencialmente observable (Ej.: Resultados de los pacientes bajo el otro tratamiento o resultados de un nuevo paciente bajo tratamientos similares)
- Variables y Unidades Experimentales:

  Sea y el conjunto de n objetos o unidades $y = (y_1, \ldots, y_n)$. En la prueba clínica por ejemplo: $y_i = 1$ si el paciente i está vivo luego de cinco años, o $y_i = 0$ si el paciente muere. Si se miden varias variables para cada unidad y_i entonces y es una matriz. Los valores de y son los resultados aleatorios del experimento.

(Continuación:)

- Intercambiabilidad:

  La distribución de probabilidades conjunta $p(y_1, y_2, \ldots, y_n)$ es invariante a permutaciones en los índices y_i.

- Variables explicativas:

  Observaciones no aleatorias disponibles para cada unidad. Ej.: Edad y condición previa de salud. X es la matriz de covariables de dimensión $n \times k$ donde n es el número de unidades y k es el número de covariables.

- Modelos jerárquicos:

  También llamados modelos de niveles múltiples. Se utilizan cuando hay información disponible sobre distintos niveles de las unidades observacionales.

Principios de Inferencia Bayesiana

- Cantidades de interés: $p(\theta|y)$ y $p(\tilde{y}|y)$
- **Regla de Bayes**:

  Sea $p(\theta, y)$ = distribución conjunta de θ y y. $p(\theta, y) = p(\theta)\ p(y|\theta)$ donde $p(\theta)$ = distribución a priori de θ y $p(y|\theta)$ = distribución de muestreo. La regla de Bayes establece que la densidad a posteriori de θ condicionando sobre los valores conocidos de los datos y viene dada por:

  $$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}$$

  donde:

  - $p(y) = \sum_\theta p(\theta)p(y|\theta)$ (suma sobre todos los posibles valores de θ).
  - $p(y) = \int p(\theta)p(y|\theta)d\theta$ (caso de θ continuo).

  Para un y fijo decimos que : $p(\theta|y) \propto p(\theta)p(y|\theta)$ (densidad a posteriori no normalizada).

  Estas expresiones resumen las técnicas Bayesianas: la primera tarea es desarrollar un modelo para $p(\theta, y)$ y realizar los cáculos necesarios para resumir $p(\theta|y)$ de forma apropiada.

Predicción

Predicción es hacer inferencia sobre una cantidad observable pero desconocida. Usualmente se le llama inferencia predictiva. Si y es desconocida pero observable:

$$p(y) = \int p(y, \theta) d\theta = \int p(\theta) p(y|\theta) d\theta$$

Esta es la distribución marginal de y o distribución predictiva a priori (es una *distribución a priori* porque no está condicionada a una observación previa del proceso; es una *distribución predictiva* porque es la distribución de una cantidad observable).

Después de que y ha sido observada, podemos predecir una cantidad desconocida $\tilde{y}$ observable.

Ejemplo: Sea $y = (y_1, y_2, \ldots, y_n)$ el peso registrado de un objeto pesado n veces. Sea $\theta = (\mu, \sigma^2)$ el peso verdadero y la varianza de medición del peso respectivamente. Sea $\tilde{y}$ el peso del objeto para una nueva pesada.

La distribución predictiva a posteriori es:

$$p(\tilde{y}|y) = \int p(\tilde{y},\theta|y)d\theta \quad (1)$$

$$= \int p(\tilde{y}|\theta,y)p(\theta|y)d\theta \quad (2)$$

$$= \int p(\tilde{y}|\theta)p(\theta|y)d\theta \quad (3)$$

Se asume que $\tilde{y}$ y y son independientes.

Principio de verosimilitud

Sea $p(y|\theta)$ la función de verosimilitud como función de θ para y fijo.

La inferencia Bayesiana obedece el principio de verosimilitud, es decir, para una muestra de datos, dos modelos $p(y/\theta)$ con la misma función de verosimilitud producen la misma inferencia sobre θ. En la práctica no siempre se tiene el modelo correcto por lo que es necesario probar una variedad de modelos.

Verosimilitud y *Odds Ratio*

(Razón de Probabilidades o Razón de Productos Cruzados)

El *Odds Ratio* es el cociente de las densidades a posteriori evaluadas en los puntos θ_1 y θ_2 bajo un modelo dado. Usualmente el concepto se aplica a parámetros discretos donde θ_1 es el *complemento* de θ_2:

$$\frac{p(\theta_1|y)}{p(\theta_2|y)} = \frac{p(\theta_1)p(y|\theta_1)/p(y)}{p(\theta_2)p(y|\theta_2)/p(y)} = \frac{p(\theta_1)}{p(\theta_2)}\frac{p(y|\theta_1)}{p(y|\theta_2)}$$

Esta expresión es equivalente al producto de la razón de distribuciones a priori por la razón de verosimilitudes.

Ejercicio: Discutir el problema de la sección 1.4 del GCSR. El ejemplo es una aplicación sencilla del teorema de Bayes. Se considera el problema genético de determinar la probabilidad de que una mujer sea portadora del gen de la hemofilia ($\theta = 1$) o no ($\theta = 0$).

- Se obtiene una estimación de la probabilidad a priori para θ $Pr(\theta = 1)$ y $Pr(\theta = 0)$ a partir de la condición de su madre, padre y hermano.
- Se conocen el estado de sus dos hijos: y_1 y y_2 donde $y_i = 1$ si el hijo es afectado o $y_i = 0$ sino. En este ejemplo se asume $y_1 = y_2 = 0$, es decir, ninguno de los hijos está afectado por la enfermedad. Se calcula la verosimilitud de estos datos: $p(y_1 = 0, y_2 = 0|\theta)$.
- Se calcula la distribución a posteriori $Pr(\theta = 1|y)$ mediante el teorema de Bayes.
- Se actualiza la información a posteriori si se obtienen nuevos datos (Ejemplo: Condición de un nuevo hijo y_3)

Resultados útiles de teoría de probabilidades

- Factorización de una densidad conjunta:

$$p(u, v, w) = p(u|v, w)\ p(v|w)\ p(w)$$

- Medias y Varianzas de distribuciones condicionales:

$$E(u) = E(E(u|v))$$

$$var(u) = E(var(u|v)) + var(E(u|v))$$

  Ambas identidades son válidas si u es un vector. En este caso $E(u)$ es un vector y $var(u)$ es una matriz.

- Transformación de variables:

  Si $p_u(u)$ es una distribución contínua del vector u y $v = f(u)$ es una transformación uno a uno, entonces la densidad conjunta del vector transformado es:

  $$p_v(v) = |J| p_u(f^{-1}(v))$$

  donde $|J|$ es el determinante del Jacobiano de la transformación $u = f^{-1}(v)$ como una función de v. (El elemento (i, j)-ésimo de $J = \delta v_i / \delta u_j$)

  Transformaciones usuales en una dimensión:

  $(0, \infty) \rightarrow (-\infty, +\infty)$: logarítmica

  $(0, 1) \rightarrow (-\infty, +\infty)$: logit(u) = log $(\frac{u}{1-u})$

  $(0, 1) \rightarrow (-\infty, +\infty)$: probit $= \Phi^{-1}(u)$ donde Φ es la normal estándar.

Ejemplo

Sea $X : 2 \times 1$ con $X = (Y, Z)'$ donde Y y Z son escalares.

La función de densidad de X vienen dado por:

$$f(\mathbf{x}) = f(y, z) = \frac{1}{2\pi\sigma^2} exp\{-\frac{1}{2\sigma^2}[(y - \theta_y)^2 + (z - \theta_z)^2]\}$$

$$-\infty < y < \infty, \ -\infty < z < \infty$$

$$\Rightarrow X \sim N((\theta_y, \theta_z), \sigma^2 I)$$

Consideramos la transformación:

$$Y = \log U$$

$$Z = \log V$$

Por teorema de la transformación la densidad $g(u,v)$ vienen dada por la ecuación:

$$\begin{aligned} g(u,v) &= f(y(u,v),z(u,v)).J((y,z)\rightarrow(u,v)) \\ &= \frac{1}{2\pi\sigma^2}exp\{-\frac{1}{2\sigma^2}[(\log u-\theta_y)^2+(\log v-\theta_z)^2]\}.J((y,z)\rightarrow(u,v)) \end{aligned}$$

$$\begin{aligned} J &= \left|\begin{array}{cc} \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{array}\right| \\ &= \left|\begin{array}{cc} \frac{1}{u} & 0 \\ 0 & \frac{1}{v} \end{array}\right| \\ &= \frac{1}{u.v} \end{aligned}$$

$$\Rightarrow g(u,v) = \frac{1}{2\pi\sigma^2 uv}exp\{-\frac{1}{2\sigma^2}[(\log u-\theta_y)^2+(\log v-\theta_z)^2]\}$$

$$0<u<\infty \quad 0<v<\infty$$

Uso de la simulación para la inferencia

La simulación es fundamental en la estadística Bayesiana. Dada una muestra grande de una distribución de probabilidades, podemos estimar varios momentos muestrales, percentiles y otros estadísticos para estudiar muchos aspectos de la distribución.

Muestreo de una distribución de probabilidades

Sea F la función de distribución acumulada (cdf) de una función de densidad unidimensional $p(v)$.

$$\begin{aligned} F(v^*) &= Pr(v \leq v^*) \\ &= \begin{cases} \sum_{v \leq v^*} p(v) & \text{p es discreta} \\ \int_{-\infty}^{v^*} p(v)dv & \text{p es contínua} \end{cases} \end{aligned} \tag{4}$$

Para obtener una muestra aleatoria de una distribución $p(v)$ el procedimiento es el siguiente:

- Generar un valor aleatorio U de una distribución uniforme [0,1]
- Obtener $v = F^{-1}(U)$ donde v es una muestra aleatoria de p.

Ejercicio: Obtenga 10000 muestras de una distribución exponencial con parámetro $\lambda = 2$ usando R; grafique el histograma correspondiente y compárelo con el histograma que resulta de obtener 10000 muestras de la misma distribución usando el método de la función de distribución acumulada inversa descrito anteriormente.

Simulación de cantidades a posteriori y predictivas a posteriori

Simulación	Parámetros	Cantidades Predictivas
	$\theta_1, \ldots, \theta_k$	$\tilde{y}_1, \ldots, \tilde{y}_n$
1	$\theta_1^1, \ldots, \theta_k^1$	$\tilde{y}_1^1, \ldots, \tilde{y}_n^1$
$\vdots$	$\vdots$	$\vdots$
L	$\theta_1^L, \ldots, \theta_k^L$	$\tilde{y}_1^L, \ldots, \tilde{y}_n^L$

De los valores simulados podemos estimar por ejemplo: Distribución a posteriori de θ_1/θ_3; $Pr(\tilde{y}_1 + \tilde{y}_2 > \exp(\theta_1))$; el intervalo a posteriori del 95 % para el parámetro θ_j, etc.