Resumen Estadística

1. Modelo y regresión lineal

La idea es, a partir del vector de observaciones X (que no es v.a.), obtener un modelo lineal $Y = X\beta + \varepsilon$ para la variable Y. Para obtener esta función lineal se deben estimar los coeficientes β según algún criterio. En dimensión uno, usando error cuadrático, se resuelve

$$\min_{\alpha,\,\beta\in\mathbb{R}} \sum_{i=1}^{n} (y_i - \alpha - \beta\, x_i)^2\,,$$

de donde

$$\hat\beta = \frac{\sigma_{XY}}{\sigma_X^2} \quad \text{y} \quad \hat\alpha = \overline{y} - \hat\beta\,\overline{x}\,,$$

con

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \qquad \sigma_{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i\, y_i - \overline{x}\,\overline{y}\,.$$

Def 35. (Residuos) $\hat\varepsilon_i := y_i - \hat y_i$, donde $\hat y_i = \hat\alpha + \hat\beta\, x_i$.

Def 36. (Coeficiente de correlación lineal o de bondad de ajuste)

$$R^2 := \frac{\displaystyle\sum_{i=1}^{n}(y_i - \overline{y})^2 - \sum_{i=1}^{n}\hat\varepsilon_i^2}{\displaystyle\sum_{i=1}^{n}(y_i - \overline{y})^2}.$$

Notemos que $0 \le r^2 \le 1$. El coeficiente de correlación también puede ser escrito como

$$R^2 = \frac{\sigma_{XY}}{\sigma_X^2\,\sigma_Y^2} = \hat\beta^2\,\frac{\sigma_X^2}{\sigma_Y^2}.$$

Supuestos probabilísticos

- Errores centrados: $\mathbb{E}(\varepsilon) = 0$.
- Independencia: $\forall i, j \in [n], \varepsilon_i \perp \varepsilon_j$.
- Homocedasticidad: $\forall\, i \in [n], \mathrm{Var}(\varepsilon_i) = \sigma^2 < +\infty$.

Modelo lineal general
Sea $Y = (Y_1, \ldots, Y_n)$, X la matriz de $n \times p$ de las observaciones, $\beta = (\beta_1, \ldots, \beta_p)$ y $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$. Luego, el modelo es $Y = X\,\beta + \varepsilon$ y el problema a resolver se escribe como

$$\min_{\beta\in\mathbb{R}^p} \|Y - X\,\beta\|^2 \quad \text{o} \quad \min_{\beta\in\mathbb{R}^p} \|\varepsilon\|^2.$$

De forma geométrica, esto equivale a buscar $\hat\beta$ tal que la proyección ortogonal de Y sobre $\langle X\rangle = \{X\,\beta : \beta \in \mathbb{R}^n\}$ sea $X\hat\beta$. Si X es de rango p (modelo lineal de rango completo), entonces el estimador de mínimos cuadrados $\hat\beta = (X'X)^{-1}X'Y$ es la única solución del problema. En este caso $\hat\beta$ es insesgado para β.

Con lo anterior $\hat Y = X(X'X)^{-1}X'Y$. El proyector sobre $\langle X\rangle$ es $P_{\langle X\rangle} = X(X'X)^{-1}X'$ y el proyector sobre el ortogonal es $P_{\langle X\rangle^\perp} = I - X(X'X)^{-1}X'$, por lo que el vector de residuos es $\hat\varepsilon = Y - \hat Y = P_{\langle X\rangle^\perp}(Y)$.

Bajo los supuestos pobabilísticos se tiene $\mathrm{Var}(\hat\beta) = \sigma^2\,(X'X)^{-1}$. Un estimador de σ^2 basado en $\hat\varepsilon$ es $\hat\sigma^2 = \frac{1}{n-p}\|\hat\varepsilon\|^2$ que resulta ser insesgado para σ^2.

Teo 5. (Gauss-Markov) Sea Y un n-vector aleatorio de la forma $Y = X\beta + \varepsilon$, donde X es una $n\times p$ matriz conocida de rango p, β un p-vector desconocido y ε es el vector de errores con $\mathrm{E}(\varepsilon) = 0$ y $\mathrm{Var}(\varepsilon) = \sigma^2\, I$ con σ^2 desconocido. Sea $\hat\beta$ el único estimador de mínimos cuadrados de β. Sea $\phi = a\beta$ una función lineal de β, donde a es un vector conocido. Entonces $a\hat\beta$ es el mejor estimador lineal insesgado para ϕ, es decir, para todo estimador lineal insesgado $\tilde\phi$ de ϕ se tiene $\mathrm{Var}(a\hat\beta) \le \mathrm{Var}(\tilde\phi)$.

Modelo lineal general con errores correlacionados

Tenemos el modelo $Y = X\beta + \varepsilon$ con X de rango completo (p) y $\mathrm{E}(\varepsilon) = 0$, pero ahora $\mathrm{E}(\varepsilon\,\varepsilon') = \sigma^2\Sigma$, donde Σ es una matriz de $n \times n$ definida positiva.

El estimador de mínimos cuadrados ordinarios es

$$\hat\beta_{MCO} = (X'X)^{-1}X'Y$$

y escogiendo A no singular tal que $\Sigma = (A'A)^{-1}$, definimos el estimador de mínimos cuadrados generalizado por

$$\hat\beta_{MCG} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y\,,$$

que también es insesgado para β, y tiene menor varianza.

2. Modelo lineal: Teoría Normal

Además de los supuestos probabilísticos usuales y que la matriz sea de rango completo, se supondrá que $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$. Recordmos el modelo $Y = X\beta + \varepsilon$. Como los ε_i son v.a. $\mathcal{N}(0, \sigma^2)$ independientes, entonces $Y \sim \mathcal{N}(X\beta, \sigma^2 I)$. Por lo que: $\hat\beta \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1})$.Cada $\hat\beta_i \sim \mathcal{N}(\beta_i, \sigma^2 c_{ii})$, donde c_{ii} es el elemento i de la diagonal de $(X'X)^{-1}$.

Podemos construir **intervalos de confianza**, definiendo el pivote

$$\frac{\hat\beta_i - \beta_i}{\sigma\,\sqrt{c_{ii}}} \sim \mathcal{N}(0, 1).$$

Si σ es conocido se usa el pivote para encontrar $z_{\alpha/2}$ tal que $\mathrm{P}_{\beta_i}(-z_{\alpha/2} \le \hat\beta_i \le z_{\alpha/2}) = 1 - \alpha$. Si σ no es conocido, entonces se ocupa el pivote

$$\frac{\hat\beta_i - \beta_i}{\hat\sigma\,\sqrt{c_{ii}}} \sim t(n-p) \quad \text{con} \quad \hat\sigma^2 = \frac{1}{n-p}\sum_{i=1}^{n}\varepsilon_i^2.$$

Podemos aplicar **test de hipótesis** a este modelo, usando el TRV con $\Theta = \mathbb{R}^p \times (0, +\infty)$ y $\theta = (\beta, \sigma)$. Para este modelo se tienen las siguientes hipótesis usuales:

(1) $H_0 : \beta = 0$ v/s $H_1 : \beta \neq 0$. H_0 equivale a decir que las variables X son irrelevantes linealmente para Y.

(2) $H_0 : \beta_i = 0$ v/s $H_1 : \beta_i \neq 0$. H_0 indica que la i-ésima variable X es irrelevante linealmente para Y.

(3) $H_0 : \beta_1 = \beta_2$ v/s $H_1 : \beta_1 \neq \beta_2$. Esto se puede representar mediante tomando $h = [1, -1, 0, .., 0]^t$ de forma que $H_0 : h^t\beta = 0$ y $H_1 : h^t\beta \neq 0$.

Ocupando TRV de nivel α, k_α se puede calcular a partir de

$$\sup_{\theta\in\Theta_0} \mathrm{P}_\theta\left(\frac{\|\tilde\varepsilon\|^2}{\|\hat\varepsilon\|^2} \ge k_\alpha\right) = \alpha,$$

donde $\tilde\varepsilon = (Y - X\tilde\beta)'(Y - X\tilde\beta)$ con $\tilde\beta$ el óptimo de $\sup_{\theta\in\Theta_1} f_\theta(y)$ y $\hat\varepsilon = (Y - X\hat\beta)'(Y - X\hat\beta)$ con $\hat\beta$ el óptimo de $\sup_{\theta\in\Theta_0} f_\theta(y)$.

3. Recuerdos

3.1. Matrices

- $U = (U_1, \ldots, U_k)$ con U_i v.a, entonces $\mathrm{E}(U) = (\mathrm{E}(U_1), \ldots, \mathrm{E}(U_k))$.
- Si $M = (M_{ij})$ matriz con M_{ij} v.a, $\mathrm{E}(M) = (\mathrm{E}(M_{ij}))$.
- La matriz de covarianzas de U es $\mathrm{Var}(U)$

  $= (\mathrm{Cov}(U_i, U_j))_{i,j=1,\ldots k} = \mathrm{E}[(U - \mathrm{E}(U))\,(U - \mathrm{E}(U))']$, esta matriz siempre es semidefinida positiva.

3.2. Normal Multivariada

- Sea $U = (U_1, \ldots, U_n)$ un vector aleatorio con componentes independientes $U_i \sim \mathcal{N}(0,1)$, entonces

$$f_U(u) = \left(\frac{1}{2\pi}\right)^{n/2} e^{-\frac{1}{2}\sum_{i=1}^n u_i^2}.$$

- Sea $Z = AU + \mu$ con A matriz invertible y $\mu \in \mathbb{R}^n$, entonces $Z \sim \mathcal{N}(\mu, \Sigma)$ con $\Sigma = AA'$, es decir,

$$f_Z(z) = \frac{(2\pi)^{-n/2}}{\sqrt{\det(\Sigma)}}\, e^{-\frac{1}{2}(z-\mu)'\Sigma^{-1}(z-\mu)}.$$

- Sea $V = BZ + \nu$, B matriz apropiada, que puede tener $p < n$ filas, entonces $V \sim \mathcal{N}(B\mu + \nu, B\Sigma B')$.
- Sea $U \sim \mathcal{N}_n(0, \sigma^2 I)$, entonces $\frac{\|U\|^2}{\sigma^2} \sim \chi^2(n)$.

3.3. Distribuciones

(1) (Uniforme) La v.a. X se dice $X \sim U([a,b])$ si tiene función de densidad:

$$f_X(t) = \begin{cases} \frac{1}{b-a} & si \;\; t \in [a,b] \\ 0 & si \;\; t \notin [a,b] \end{cases}$$

entonces $F_X(t) = \displaystyle\int_{-\infty}^{t} f_X(s)ds = \begin{cases} 0 & si \;\; t < a \\ 1 & si \;\; t > b \\ \frac{t-a}{b-a} & si \;\; t \in [a,b] \end{cases}$

Obs. Si $Y \sim U(0,1)$ y $X = a + (b-a)Y$, entonces $X \sim U(a,b)$.

(2) (Exponencial) La v.a. X se dice $X \sim \text{exponencial}(\lambda)$ si tiene función de densidad: $f_X(t) = \lambda\, e^{-\lambda\, t}\, 1_{\{t>0\}}$.

entonces $F_X(t) = \displaystyle\int_{-\infty}^{t} \lambda\, e^{-\lambda\, s}\, 1_{\{s>0\}} ds = (1 - e^{-\lambda\, t})\, 1_{\{t>0\}}$.

(3) (Gamma) Diremos que la v.a. X tiene distribución Gamma de parámetros λ, $k > 0$ si la densidad de X es

$$f_X(x) = \lambda\, e^{-\lambda\, x}\, \frac{(\lambda\, x)^{k-1}}{\Gamma(k)}.$$

donde $\Gamma(n) = (n-1)!$ para $n \in \mathbb{N}$. Además, tenemos la identidad $\Gamma(p+1) = p\,\Gamma(p), \quad \forall\, p > 0$.

Obs. Si $X \sim \mathcal{N}(0,1)$, entonces $\mu + \sigma\, X \sim \mathcal{N}(\mu, \sigma^2)$

(5) (Chi-cuadrado) La v.a. Z se dice $Z \sim \mathcal{X}^2(n)$ (n grados de libertad) si tiene densidad:

$$f_X(x) = \frac{1}{2^{\frac{n}{2}}\,\Gamma(\frac{n}{2})}\, x^{\frac{n}{2}-1}\, e^{-\frac{x}{2}}\, \mathbb{1}_{(0,+\infty)}(x).$$

Obs. Si X es v.a. $\mathcal{N}(0,1)$, entonces $X^2 \sim \mathcal{X}^2(1)$

(6) (t-student) Sea U una v.a. $\mathcal{N}(0,1)$ y V un v.a. $\chi^2(n)$ independiente de U, entonces decimos que $T = \frac{U}{\sqrt{\frac{V}{n}}}$ tiene una distribución t-student con n grados de libertad si su densidad es

$$\Phi_n(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$