Tarea Examen

Inferencia Bayesiana
Semestre 2025-1

Diego Leipen Lara

Notación. Escribimos $\mathcal{N}_n(\mu, P)$ para denotar una normal multivariada de dimensión n con media μ y matriz de precisión P. Escribimos $N_n(\mu, \Sigma)$ para denotar una normal multivariada de dimensión n con media μ y matriz de covarianza Σ.

Lema 1. Si $y \sim N_n(\mu, \Sigma)$ y A es una matriz $(m \times n)$, entonces

$$Ay \sim N_m(A\mu, A\Sigma A^T).$$

1. *Regresión lineal.*

En clase vimos que

$$p(\beta|\tau, y) = \mathcal{N}_p(\beta|\mu_1, \tau P_1).$$

(a) Sea $\gamma = C\beta$ donde C es una matriz $(r \times p)$ de rango r $(r \leq p)$. ¿Cuál es la correspondiente distribución final de γ condicional en τ, $p(\gamma|\tau, y)$?

(b) Una observación futura del modelo se puede describir a través de

$$y_* = z_*^T \beta + e,$$

donde $e \sim \mathcal{N}(0, \tau)$. Usando el resultado del inciso (a) y el hecho de que β y e son independientes dados τ y y, encuentra la distribución predictiva final de Y_*, $p(y_*|\tau, y)$.

(c) ¿Qué relación guarda esta distribución predictiva condicional con la distribución predictiva discutida en la página 16 de las notas “Regresión.pdf”?

Demostración.

(a) Por el lema 1,

$$p(\gamma|\tau, y) = N_r(\gamma|C\mu_1, C(\tau P_1)^{-1}C^T).$$

(b) Por (a),

$$p(z_*^T\beta|\tau, y) = N(z_*^T\beta|z_*^T\mu_1, z_*^T(\tau P_1)^{-1}z_*).$$

Además, como

- $e \sim N(0, \tau^{-1})$,
- β y ϵ son independientes dado τ y y,
- $y_* = z_*^T\beta + e$, y
- la suma de normales independientes es normal con media igual a la suma de las medias y varianza igual a la suma de las varianzas,

entonces

$$p(y_*|\tau, y) = N(y_*|z_*^T\mu_1 + 0, z_*^T(\tau P_1)^{-1}z_* + \tau^{-1}) = N\left(y_*|z_*^T\mu_1, \tau^{-1}(z_*^T P_1^{-1} z_* + 1)\right).$$

(c) La relación que guarda esta distribución predictiva condicional con la distribución predictiva discutida en la página 16 de las notas "Regresión.pdf" es

$$p(y^*|y) = \int p(y^*|\tau, y)p(\tau|y)\, d\tau$$

Además,

$$p(\tau|y) = \text{Gamma}(\tau|\alpha_1, \delta_1)$$

(cf. página 15 de las notas "Regresión.pdf"). Entonces

$$\begin{aligned}
\int p(y^*|\tau, y)p(\tau|y)\, d\tau &= \int N\left(y_*|z_*^T\mu_1, \tau^{-1}(z_*^T P_1^{-1} z_* + 1)\right) \text{Ga}(\tau|\alpha_1, \delta_1)\, d\tau \\
&= \int \text{NormalGamma}(y_*, \tau|z_*^T\mu_1, (z_*^T P_1^{-1} z_* + 1)^{-1}, \alpha_1, \delta_1)\, d\tau \\
&= \text{St}\left(y_*|z_*^T\mu_1, \frac{\alpha_1}{\delta_1}(z_*^T P_1^{-1} z_* + 1)^{-1}, 2\alpha_1\right).
\end{aligned}$$

Lo cual coincide con el resultado mencionado en la página 16.

$\square$

2. *Modelos lineales generalizados*

(a) Sea $Y_1, \ldots, Y_m$ una muestra de observaciones i.i.d. de una distribución Geométrica(λ), con $0 < \lambda < 1$, de manera que $\overline{Y}$ tiene una distribución Binomial Negativa escalada con funcion de probabilidad

$$p(y|\lambda) = \binom{m + my - 1}{my} \lambda^{my}(1-\lambda)^m, \qquad y = 0, 1/m, 2/m, \ldots$$

(i) Escribe $p(y|\lambda)$ en su forma exponencial, identificando el parámetro canónico y el parámetro de dispersión.

(ii) Encuentra la media $\mu = \mathbb{E}[\overline{Y}|\lambda]$ y escribe la varianza de $\overline{Y}$ en términos de μ.

(iii) Encuentra la función liga canónica para este modelo.

(b) Considera la clase de modelos lineales generalizados y supón que no se puede o no se desea especificar una función liga particular. Discute la posibilidad de hacer inferencias sobre g. ¿De qué manera se puede incorporar una función liga *desconocida* al modelo?

Demostración.

(a) (i) Sean

$$\begin{aligned}
\theta &= \log\lambda, \\
a(\theta) &= -\log(1-e^\theta) = -\log(1-\lambda), \\
\sigma^2 &= 1, \\
\phi &= m/\sigma^2 = m, \\
b(y,\phi) = b(y,m) &= \binom{m+my-1}{my} \mathbb{1}_{\{k/m | k \in \mathbb{Z}_{\geq 0}\}}
\end{aligned}$$

Como

$$\exp\left\{\phi[y\theta - a(\theta)]\right\} = \exp\left\{m[y\log\lambda + \log(1-\lambda)]\right\} = \lambda^{my}(1-\lambda)^m,$$

entonces

$$p(y|\lambda) = b(y,\phi)\exp\left\{\phi[y\theta - a(\theta)]\right\}.$$

(ii) Como $m\overline{Y} \sim \text{BinomNeg}(r = m, p = 1-\lambda)$, entonces

$$\begin{aligned}
\mathbb{E}[m\overline{Y}] &= \frac{r(1-p)}{p} = \frac{m\lambda}{1-\lambda}, \\
\text{Var}[m\overline{Y}] &= \frac{r(1-p)}{p^2} = \frac{m\lambda}{(1-\lambda)^2}.
\end{aligned}$$

En particular,

$$\begin{aligned}
\mathbb{E}[\overline{Y}] &= \frac{\lambda}{1-\lambda}, \\
\text{Var}[\overline{Y}] = \frac{\lambda}{m(1-\lambda)^2} &= \frac{1}{m(1-\lambda)}\mathbb{E}[\overline{Y}].
\end{aligned}$$

Esto coincide con las expresiones de la pagina 18 de las notas "Regresión.pdf". Específicamente,

$$\begin{aligned}
\mathbb{E}[\overline{Y}|\theta] = a'(\theta) &= \frac{d}{d\theta}\Big(-\log(1-e^\theta)\Big) = \frac{e^\theta}{1-e^\theta} = \frac{\lambda}{1-\lambda}, \\
\text{Var}[\overline{Y}] = \frac{\sigma^2}{m}a''(\theta) &= \frac{1}{m}\frac{d}{d\theta}\left(\frac{e^\theta}{1-e^\theta}\right) = \frac{1}{m}\frac{e^\theta}{(1-e^\theta)^2} = \frac{\lambda}{m(1-\lambda)^2}.
\end{aligned}$$

(iii) Sea $g(\lambda) \coloneqq \log\left(\frac{\lambda}{\lambda+1}\right)$. Como $a'(\theta) = \frac{e^\theta}{1-e^\theta}$, entonces

$$a'(g(\lambda)) = \frac{\frac{\lambda}{\lambda+1}}{1-\frac{\lambda}{\lambda+1}} = \frac{\frac{\lambda}{\lambda+1}}{\frac{1}{\lambda+1}} = \lambda,$$

$$g(a'(\theta)) = \log\left(\frac{\frac{e^\theta}{1-e^\theta}}{\frac{e^\theta}{1-e^\theta}+1}\right) = \log\left(\frac{\frac{e^\theta}{1-e^\theta}}{\frac{1}{1-e^\theta}}\right) = \log e^\theta = \theta.$$

Por lo tanto, g es la función liga canónica para este modelo.

(b) Una posibilidad es restringir g a una familia paramétrica de funciones que sea adecuada para el problema dado. Por ejemplo, si se sabe que $\mathbb{E}[Y|z] \in [a,b]$, es razonable escoger a la familia paramétrica de manera que sus elementos sean biyecciones de $[a,b]$ a $\mathbb{R}$. Si denotamos por $\{g_\theta \mid \theta \in \Theta\}$ a la familia paramétrica, tendríamos que

$$Y \sim N_n(g_\theta^{-1}(X\beta), \tau I_n).$$

El objetivo es inferir los valores de los parámetros θ, β, y τ. Para ello es necesario asignar distribuciones iniciales $p(\theta)$, $p(\beta)$, $p(\tau)$ y luego encontrar las distribuciones finales $p(\theta|y)$, $p(\beta|y)$ $p(\tau|y)$.

$\square$

3. *Modelos jerárquicos* Considera el modelo jerárquico Poisson dado por:

I.
$$p(y|\theta) = \text{Poisson}(y|\theta), \qquad \theta > 0 \text{ desconocida.}$$

II.
$$p(\theta|\phi) = \text{Gamma}(\theta|\alpha, \phi), \qquad \alpha > 0 \text{ conocida, } \phi > 0 \text{ desconocida.}$$

III.
$$p(\phi) = B(a,b)\phi^{a-1}(1+\phi)^{-(a+b)}, \qquad a > 0,\, b > 0 \text{ conocidas.}$$

donde
$$B(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}.$$

(a) Demuestra que $p(y|\phi)$ es una distribución Binomial Negativa con parámetros $\pi = \frac{\phi}{1+\phi}$ y α.

(b) Encuentra $\mathbb{E}[Y|\phi]$ y $\mathrm{Var}[Y|\phi]$.
¿Cómo se relacionan estas cantidades entre sí?
Compara esta relación con la que tienen $\mathbb{E}[Y|\theta]$ y $\mathrm{Var}[Y|\theta]$.

(c) Finalmente, encuentra la distribución final de π, esto es, $p(\pi|y)$.

Demostración.

(a) En la pagina 19 de las notas "Regresión.pdf" vimos que
$$p(y|\theta,\phi) = p(y|\theta)$$
pues el hiperparámetro ϕ afecta a y solo a través de θ. De esto se sigue que
$$\begin{aligned}
p(y|\phi) &= \int_0^\infty p(x|\theta,\phi)p(\theta|\phi)\,d\theta = \int_0^\infty p(x|\theta)p(\theta|\phi)\,d\theta \\
&= \int_0^\infty \text{Poisson}(y|\theta)\text{Gamma}(\theta|\alpha,\phi)\,d\theta = \int_0^\infty \left(\frac{\theta^y e^{-\theta}}{y!}\right)\left(\frac{\phi^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\phi\theta}\right)d\theta \\
&= \frac{1}{y!}\frac{\phi^\alpha}{\Gamma(\alpha)}\int_0^\infty \theta^{y+\alpha-1}e^{-(1+\phi)\theta}\,d\theta = \frac{1}{y!}\frac{\phi^\alpha}{\Gamma(\alpha)}\int_0^\infty \text{Gamma}(\theta|y+\alpha, 1+\phi)\,d\theta \\
&= \frac{1}{y!}\frac{\phi^\alpha}{\Gamma(\alpha)}\frac{\Gamma(y+\alpha)}{(1+\phi)^{y+\alpha}} = \frac{(y+\alpha-1)!}{y!(\alpha-1)!}\left(\frac{\phi}{1+\phi}\right)^\alpha\left(\frac{1}{1+\phi}\right)^y \\
&= \binom{y+\alpha-1}{y}\left(\frac{\phi}{1+\phi}\right)^\alpha\left(1-\frac{\phi}{1+\phi}\right)^y \\
&= \text{BinomNeg}\left(y|r=\alpha, \pi = \frac{\phi}{1+\phi}\right).
\end{aligned}$$

(b) Sabemos que si $X \sim \text{BinomNeg}(r,\pi)$, entonces
$$\mathbb{E}[X] = \frac{r(1-\pi)}{\pi}, \qquad \mathrm{Var}[X] = \frac{r(1-\pi)}{\pi^2} = \frac{1}{\pi}\mathbb{E}[X].$$

Entonces
$$\begin{aligned}
\mathbb{E}[Y|\phi] &= \frac{\alpha\left(1-\frac{\phi}{1+\phi}\right)}{\frac{\phi}{1+\phi}} = \frac{\alpha\frac{1}{1+\phi}}{\frac{\phi}{1+\phi}} = \frac{\alpha}{\phi}, \\
\mathrm{Var}[Y|\phi] &= \frac{1}{\pi}\mathbb{E}[Y|\phi] = \frac{1}{\frac{\phi}{1+\phi}}\frac{\alpha}{\phi} = \frac{\alpha(1+\phi)}{\phi^2}.
\end{aligned}$$

En contraste, como $Y|\theta \sim \text{Poisson}(\theta)$, entonces $\mathbb{E}[Y|\theta] = \lambda = \mathrm{Var}[Y|\theta]$. En comparación, $\mathbb{E}[Y|\phi]$ y $\mathrm{Var}[Y|\phi]$ se relacionan a través de un múltiplo constante que solo depende de ϕ.

(c) Calculando directamente,

$$\begin{aligned}
p_\phi(\phi|y) &\propto p(y|\theta)p(\phi) \\
&\propto \left(\phi^{\alpha}(1+\phi)^{-(y+\alpha)}\right)\left(\phi^{a-1}(1+\phi)^{-(a+b)}\right) \\
&\propto \phi^{\alpha+a-1}(1+\phi)^{-(y+\alpha+a+b)}.
\end{aligned}$$

Sea $g(\phi) = \pi = \frac{\phi}{1+\phi}$. Entonces

$$\begin{aligned}
g^{-1}(\pi) = \phi &= \frac{\pi}{1-\pi}, \\
\frac{d}{d\pi}g^{-1}(\pi) &= \frac{1}{(1-\pi)^2}.
\end{aligned}$$

Entonces

$$\begin{aligned}
p(\pi|y) &= p_\phi(g^{-1}(\pi))\left|\frac{d}{d\pi}g^{-1}(\pi)\right| \\
&\propto \left(\frac{\pi}{1-\pi}\right)^{\alpha+a-1}\left(1+\left(\frac{\pi}{1-\pi}\right)\right)^{-(y+\alpha+a+b)}\left|\frac{1}{(1-\pi)^2}\right| \\
&= \left(\frac{\pi}{1-\pi}\right)^{\alpha+a-1}\left(\frac{1}{1-\pi}\right)^{-(y+\alpha+a+b)}\left|\frac{1}{(1-\pi)^2}\right| \\
&= \pi^{\alpha+a-1}(1-\pi)^{y+b+1}(1-\pi)^{-2} \\
&= \pi^{\alpha+a-1}(1-\pi)^{y+b-1} \\
&\propto \text{Beta}(\pi|\alpha+a, y+b).
\end{aligned}$$

□