Segundo bimestre de 2025 | Clase 1 - Estadística Bayesiana

Modelado Estadístico

---

1. (Teórico)

   (a) Sea $\mathbf{X} = (X_1, \ldots, X_n)$ una muestra aleatoria con distribución P_θ. Suponga una distribución a priori T para θ, con $T \sim \tau$. Sea $\delta_\Lambda(\mathbf{X})$ el estimador Bayes para θ. Si es insesgado, pruebe que
   $$\mathbb{E}((\delta_\Lambda(\mathbf{X}) - T)^2) = 0.$$

   (b) Mostrar que $\bar{X}$ no es un estimador de Bayes de θ para ninguna distribución a priori Λ cuando $X|_{T=\theta} \sim N(\theta, 1)$ y se usa pérdida cuadrática.

2. (Calculo de posteriori) Sea una m.a. $X_1, \ldots, X_n$ donde $X_i|_{=\theta} \sim \mathcal{P}(\theta)$ y $\sim \Gamma(r, \lambda)$, $r, \lambda > 0$.

   (a) Encontrar el estimador Bayes δ_Λ y calcular el riesgo de Bayes $r(\delta_\Lambda, \Lambda)$.

   (b) Mostrar que δ_Λ puede escribirse como un promedio ponderado entre $\bar{X}$ y $\frac{r}{\lambda}$. Interprete.

3. (Empirical Bayes)

   (a) Usar el dataset `Batting` del paquete `Lahman` (`playerID`, `H`, `AB`). Obtenga el número total de hits e intentos por jugador.

   (b) ¿Elegiría como “mejor jugador” a quienes tienen promedio 1? ¿Por qué?

   (c) Para jugadores con más de 1000 intentos, estime la distribución de promedios usando una Beta. Llame a esto Λ.
   Sugerencia: use `fitdistr` de `MASS` o método de momentos.

   (d) Fije la distribución del punto anterior como prior para los valores p_i (“probabilidad REAL de hit” del jugador i). Para cada jugador, obtenga el estimador Bayes para p_i.

   (e) Rankee los jugadores usando los estimadores Bayes p_i obtenidos.

   (f) Grafique y compare la estimación frecuentista (p_i empírico) y la de Empirical Bayes para todos los jugadores.

4. (Modelado bayesiano con STAN). Se mide el efecto de un método educativo en 8 escuelas: el archivo `schools.csv` contiene estimaciones $Y_1, \ldots, Y_8$ y desvíos estándar $\sigma_1, \ldots, \sigma_8$. Se modela:

   - $\mu \sim N(0, 10000)$
   - $\theta_i \sim N(\mu, 1)$
   - $Y_i \sim N(\theta_i, \sigma_i^2)$

   (a) Implemente el modelo en STAN.

   (b) Fittear y obtenga estimaciones de los θ_i.

   (c) ¿Qué se esperaría si σ_i se multiplica por una constante muy pequeña? ¿Y si se multiplica por una constante grande?