

Contents

Clase 3: Ajuste de Distribuciones en R con <code>fitdistrplus</code> - Profundización y Pruebas de Bondad de Ajuste	1
Ejemplo: Ajuste de distribuciones a tiempos de espera en una cola	2
Cargar los datos (asumimos que están en un archivo llamado "tiempos_espera.csv")	2
Exploración de los datos	2
Ajustar distribuciones	3
Distribución Exponencial	3
Distribución Gamma	3
Distribución Weibull	3
Pruebas de Bondad de Ajuste (K-S)	3
Pruebas de Bondad de Ajuste (Chi-cuadrado - ejemplo con Gamma)	3
Agrupar los datos en intervalos (ejemplo: 5 intervalos)	3
Criterios de Información	3
Interpretación: Comparar los valores p de las pruebas K-S y Chi-cuadrado, así como los AIC y BIC.	3
Seleccionar la distribución con el mejor ajuste según estas métricas.	3
Ejemplo: Validación con simulación (continuando con el ejemplo anterior)	4
Simular datos a partir de la distribución Gamma ajustada	4
Comparación gráfica	4
Comparación estadística (ejemplo: comparar medias)	4
Si las distribuciones y las medias son similares, el modelo Gamma es una buena aproximación.	4

“markdown

Clase 3: Ajuste de Distribuciones en R con `fitdistrplus` - Profundización y Pruebas de Bondad de Ajuste

Objetivos de la Clase:

- Profundizar en el uso del paquete `fitdistrplus` en R para el ajuste de distribuciones a datos de entrada.
- Comprender y aplicar diferentes pruebas de bondad de ajuste para evaluar la calidad del ajuste de una distribución.
- Interpretar los resultados de las pruebas de bondad de ajuste y seleccionar la distribución más adecuada.
- Aplicar técnicas de simulación para validar los modelos ajustados.

Contenido Teórico Detallado:

1. Repaso de `fitdistrplus` y Métodos de Estimación:

- Breve repaso de la función `fitdist` del paquete `fitdistrplus`.
- Recordatorio de los métodos de estimación de parámetros: Máxima Verosimilitud (MLE) y Método de los Momentos (MME). Énfasis en sus diferencias y cuándo es más apropiado usar cada uno. Consideraciones sobre el impacto de la elección del método en los resultados.

2. Pruebas de Bondad de Ajuste:

- **Prueba de Kolmogorov-Smirnov (K-S):**
 - Explicación del principio subyacente: compara la función de distribución empírica (ECDF) de los datos con la función de distribución acumulada (CDF) teórica de la distribución ajustada.
 - Hipótesis nula y alternativa.
 - Interpretación del estadístico D y del valor p.
 - Implementación en R con `ks.test()`.
- **Prueba de Chi-cuadrado:**
 - Revisión del principio subyacente: compara las frecuencias observadas en los datos con las frecuencias esperadas según la distribución ajustada.
 - Cálculo de grados de libertad.
 - Consideraciones sobre el tamaño de muestra y la agrupación de datos.
 - Implementación en R con `chisq.test()` y necesidad de agrupar los datos en intervalos.
- **Criterios de Información (AIC y BIC):**
 - Repaso de AIC (Akaike Information Criterion) y BIC (Bayesian Information Criterion) como medidas de la calidad del ajuste que penalizan la complejidad del modelo (número de parámetros).
 - Interpretación de los valores de AIC y BIC: valores más bajos indican un mejor ajuste.
 - Uso de `fitdistrplus` para obtener AIC y BIC automáticamente después del ajuste.

3. Implementación en R con `fitdistrplus`:

- **Ejemplo práctico detallado:** Ajuste de distribuciones a un conjunto de datos (ej. tiempos de espera en una cola).
 - Cargar los datos en R.
 - Explorar los datos con histogramas y otros gráficos (como en clases anteriores).
 - Ajustar varias distribuciones (Exponencial, Gamma, Weibull, Normal) usando `fitdist`.
 - Realizar las pruebas de bondad de ajuste (K-S y Chi-cuadrado) para cada distribución.
 - Calcular AIC y BIC para cada distribución.
 - Interpretar los resultados y seleccionar la distribución más adecuada.
 - Código de ejemplo en R (completo y comentado).

“R

Ejemplo: Ajuste de distribuciones a tiempos de espera en una cola

Cargar los datos (asumimos que están en un archivo llamado "tiempos_espera.csv")

```
tiempos_espera <- read.csv("tiempos_espera.csv")
```

Exploración de los datos

```
hist(tiempos_espera$tiempo_espera, main="Histograma de Tiempos de Espera", xlab="Tiempo de Espera")
```

Ajustar distribuciones

```
library(fitdistrplus)
```

Distribución Exponencial

```
fit_exp <- fitdist(tiempos_espera$tiempo_espera, distr = "exp", method = "mle") summary(fit_exp)
```

Distribución Gamma

```
fit_gamma <- fitdist(tiempos_espera$tiempo_espera, distr = "gamma", method = "mle") summary(fit_gamma)
```

Distribución Weibull

```
fit_weibull <- fitdist(tiempos_espera$tiempo_espera, distr = "weibull", method = "mle") summary(fit_weibull)
```

Pruebas de Bondad de Ajuste (K-S)

```
ks.test(tiempos_espera$tiempo_espera, "pexp", rate = fit_exp$estimate["rate"]) # Exponencial
ks.test(tiempos_espera$tiempo_espera, "pgamma", shape = fit_gamma$estimate["shape"], rate =
fit_gamma$estimate["rate"]) # Gamma
ks.test(tiempos_espera$tiempo_espera, "pweibull", shape =
fit_weibull$estimate["shape"], scale = fit_weibull$estimate["scale"]) # Weibull
```

Pruebas de Bondad de Ajuste (Chi-cuadrado - ejemplo con Gamma)

Agrupar los datos en intervalos (ejemplo: 5 intervalos)

```
breaks <- seq(min(tiempos_espera$tiempo_espera), max(tiempos_espera$tiempo_espera), length.out =
6)
observed_counts <- hist(tiempos_espera$tiempo_espera, breaks = breaks, plot = FALSE)$counts
expected_counts <- pgamma(breaks[-1], shape = fit_gamma$estimate["shape"], rate = fit_gamma$estimate["rate"])
- pgamma(breaks[-length(breaks)], shape = fit_gamma$estimate["shape"], rate = fit_gamma$estimate["rate"])
expected_counts <- expected_counts * length(tiempos_espera$tiempo_espera)
chisq.test(observed_counts, p = expected_counts/sum(expected_counts))
```

Criterios de Información

```
AIC(fit_exp); BIC(fit_exp) AIC(fit_gamma); BIC(fit_gamma) AIC(fit_weibull); BIC(fit_weibull)
```

Interpretación: Comparar los valores p de las pruebas K-S y Chi-cuadrado, así como los AIC y BIC.

Seleccionar la distribución con el mejor ajuste según estas métricas.

““

4. Validación de Modelos Ajustados mediante Simulación:

- Generación de datos simulados a partir de la distribución ajustada. Uso de funciones `r` (ej. `rexp`, `rgamma`, `rweibull`) para generar datos aleatorios.
- Comparación de los datos simulados con los datos reales:
 - Gráficamente (histogramas, Q-Q plots).
 - Estadísticamente (comparación de medias, varianzas, etc.).
- Si los datos simulados se parecen a los datos reales, el modelo ajustado es una buena representación de los datos de entrada.

“R

Ejemplo: Validación con simulación (continuando con el ejemplo anterior)

Simular datos a partir de la distribución Gamma ajustada

```
n_simulaciones <- 1000
datos_simulados <- rgamma(n_simulaciones, shape = fit_gamma$estimate["shape"],
rate = fit_gamma$estimate["rate"])
```

Comparación gráfica

```
hist(tiempos_espera$tiempo_espera, main="Datos Reales", xlab="Tiempo de Espera")
hist(datos_simulados, main="Datos Simulados (Gamma)", xlab="Tiempo de Espera")
```

Comparación estadística (ejemplo: comparar medias)

```
mean(tiempos_espera$tiempo_espera)
mean(datos_simulados)
```

Si las distribuciones y las medias son similares, el modelo Gamma es una buena aproximación.

“

Ejemplos y Casos de Estudio:

1. **Modelado del tiempo de fallo de un componente electrónico:** Ajuste de distribuciones (Weibull, Exponencial) a datos de tiempo hasta el fallo. Uso de pruebas de bondad de ajuste para seleccionar la distribución más adecuada. Predicción de la vida útil del componente.
2. **Modelado de la demanda en un centro de llamadas:** Ajuste de distribuciones (Poisson, Binomial Negativa) al número de llamadas recibidas por hora. Uso de la distribución ajustada para planificar la capacidad del centro de llamadas.

Problemas Prácticos o Ejercicios con Soluciones:

1. **Análisis de datos de tiempos de reparación de máquinas:** Dado un conjunto de datos de tiempos de reparación, ajustar distribuciones (Exponencial, Lognormal, Gamma) usando R. Realizar pruebas de bondad de ajuste y seleccionar la distribución más adecuada. Interpretar los resultados y determinar el tiempo de reparación promedio. **Solución:** (Código R detallado similar al ejemplo anterior, con la explicación paso a paso de cada comando y la interpretación de los resultados). Incluir la comparación de los AIC y BIC para la selección del mejor modelo.
2. **Modelado del número de clientes que llegan a una tienda por minuto:** Recolectar datos sobre el número de clientes que llegan a una tienda por minuto durante un período de tiempo. Ajustar distribuciones (Poisson, Binomial Negativa) a los datos. Evaluar el ajuste utilizando pruebas de bondad

de ajuste. Usar la distribución ajustada para simular la llegada de clientes y predecir la cantidad de personal necesaria. **Solución:** (Código R detallado y explicado, con la justificación de por qué se eligen esas distribuciones, y cómo se interpreta el resultado para la predicción de personal).

Materiales Complementarios Recomendados:

- Documentación del paquete `fitdistrplus` en R.
- Artículos sobre pruebas de bondad de ajuste (Kolmogorov-Smirnov, Chi-cuadrado).
- Libros de texto sobre simulación de eventos discretos y modelado de datos de entrada.
- Ejemplos de código R para el ajuste de distribuciones y la realización de pruebas de bondad de ajuste.