Universidad de Chile - FCFM - DIM
MA34B 2009-01
Profesor: Julio Deride; Auxiliar: Gonzalo Contador.

Resumen: Intervalos de Confianza

Teóricos

1) Si $X_1, ..., X_r$ son variables aleatorias independientes, todas con distribución $Normal(0,1)$, se define la distribución que sigue la suma de sus cuadrados como una Chi-cuadrado a r grados de libertad. Se denota

$$\sum_{i=1}^{r} X_i^2 \sim \chi_r^2$$

2) Sean $Z \sim Normal(0,1)$, $W \sim \chi_r^2$, con Z y W independientes. La distribución que sigue la variable aleatoria $\frac{Z}{\sqrt{\frac{W}{r}}}$ se conoce como distribución t de Student a r grados de libertad. Escribimos

$$\frac{Z}{\sqrt{\frac{W}{r}}} \sim t_r$$

3) Dada una variable aleatoria con un parametro μ desconocido y una m.a.s. de ella, definimos el intervalo de confianza a nivel α como el intervalo $[a,b]$ de largo mínimo tal que $\mathbb{P}(\mu \in [a,b]) = \alpha$. Usualmente, este intervalo es simétrico con respecto a la media muestral $\overline{X}$.

Propiedad

Dada una m.a.s. de una variable aleatoria $X \sim Normal(\mu, \sigma^2)$, se cumple que

$$\frac{ns_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

Lo anterior viene del hecho que, como vimos en clase auxiliar

$$\frac{ns_n^2}{\sigma^2} = \sum_{i=1}^{n} (\frac{X_i - \mu}{\sigma})^2 - (\frac{\overline{X} - \mu}{\sigma})^2$$

se ve que el primer término es la suma de n v.a. $Normal(0,1)$ al cuadrado, luego es una χ_n^2, y ella se obtiene de sumar a $\frac{ns_n^2}{\sigma^2}$ el segundo término de la igualdad, que es una v.a. normal al cuadrado (o bien, una χ_1^2). Luego, $\frac{ns_n^2}{\sigma^2} \sim \chi_{n-1}^2$, osea sigue una distribución Chi-Cuadrado a $n-1$ grados de libertad.

Resumen

En esta parte del curso ya no buscamos estimar un parámetro μ que desconocemos de una distribución, sino dar un rango de valores en función de una muestra, dentro del cuál es altamente probable que se encuentre ese parámetro. En ese contexto, dado $\alpha \in (0,1)$, buscamos un intervalo del menor largo posible de manera de que la probabilidad de que μ se encuentre dentro de ese intervalo corresponda a α.

Cuando trabajamos con una distribución $Normal(\mu, \sigma^2)$, tenemos que esta distribución concentra la mayor probabilidad para valores cercanos a μ, y es simétrica respecto a este parámetro. Luego, para plantear un intervalo que concentre un cierto nivel de probabilidad, requerimos que este esté centrado en μ. Luego, dentro de muchos intervalos que podemos encontrar asociadps a un cierto nivel de probabilidad, el intervalo de confianza será el único para aquella probabilidad que sea centrado en μ.

En un contexto general, podemos plantearnos dos casos en los que queremos dar intervalos de confianza:

Caso 1: Intervalo de confianza para un parámetro μ o σ^2, siendo el otro parámetro conocido.

Sabemos, por el Teorema Central del Límite, que para n "grande" (en MA34A se tomaba n tendiendo a infinito, aquí nos basta con una muestra grande para hacer la aproximación), que si $X_1...X_n$ son variables de media μ y varianza σ^2 todas independientes entre sí, entonces

$$\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim Normal(0,1)$$

En el caso de que queramos dar un intervalo de confianza para μ a nivel α, tenemos

$$\mathbb{P}(a < \mu < b) = \alpha$$

si y solo si

$$\mathbb{P}(\frac{\sqrt{n}(\overline{X}-a)}{\sigma} > \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} > \frac{\sqrt{n}(\overline{X}-b)}{\sigma}) = \alpha$$

si $Z \sim Normal(0,1)$, como queremos un intervalo simétrico, lo anterior es equivalente a pedir

$$\mathbb{P}(-c < Z < c) = \alpha$$

por la simetría de la distribución, esto equivale a

$$\mathbb{P}(Z > c) = \frac{1-\alpha}{2}$$

con esto, despejamos el valor de c con la tabla de valores de la distribución normal. Finalmente, despejamos $\frac{\sqrt{n}(\overline{X}-b)}{\sigma} = -c$ y $\frac{\sqrt{n}(\overline{X}-a)}{\sigma} = c$, y obtenemos los valores de los límites del intervalo.

En el caso de la varianza, si conocemos la media, tenemos

$$\mathbb{P}(a < \sigma < b) = \alpha = \mathbb{P}(\frac{\sqrt{n}(\overline{X}-\mu)}{a} > \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} > \frac{\sqrt{n}(\overline{X}-\mu)}{b})$$

y de manera análoga a lo anterior despejamos a y b.

Caso 2: Intervalo de confianza para μ, con σ desconocido.

En este caso, lo anterior ya no es válido pues no podemos despejar valores del intervalo si no conocemos σ, por lo que necesitamos una distribución que sólo dependa de μ y la muestra. Como vimos, $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim Normal(0,1)$ y $\frac{ns_n^2}{\sigma^2} \sim \chi^2_{n-1}$. como además se tiene que estos estadísticos son independientes (si no me creen, no los culpo, pero es verdad que lo son. Piensenlo un rato, igual tiene sentido), se tiene

$$\frac{\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}}{\sqrt{\frac{ns_n^2}{\sigma^2(n-1)}}} = \frac{\sqrt{n-1}(\overline{X}-\mu)}{s_n} \sim t_{n-1}$$

El estadístico anterior nos sirve para construir un intervalo de confianza para μ ya que sólo depende de éste y de valores muestrales conocidos. Como la distribución t de Student también es simétrica respecto a cero, procedemos de igual manera que en el caso 1, pero buscando los valores en la tabla de la t de Student. (Nota: en esa tabla no vienen las distribuciones para todos los grados de libertad de la distribución, por lo que tienen que aproximar a la distribución con grados de libertad más cercanos a $n-1$ que encuentren)