Tema 2

Distribuciones en el muestreo de poblaciones normales.

2.1. Introducción

En el tema 1 se obtuvieron algunos resultados relativos al comportamiento de algunas características muestrales como los momentos. En particular, se estudiaron:

- Resultados "exactos", acerca de momentos de la distribución en el muestreo de algunas características muestrales.

- Resultados "de aproximación" asintóticos, sobre el comportamiento límite (para "muestras grandes") de ciertas características.

Exceptuando los ejemplos específicos analizados, todas las consideraciones se hicieron sin presuponer ninguna distribución concreta en la población.

En este tema, se estudian algunos resultados fundamentales de la Inferencia Estadística relativos al muestreo de poblaciones normales, que se refieren a la distribución exacta (para muestras de cualquier tamaño) de estadísticos que surgen de forma natural en problemas concretos básicos de inferencia.

En este sentido, se analizan la distribución χ^2 de Pearson, t de Student y F de Snedecor, desde una doble aproximación:

1. Desde el punto de vista analítico, por particularización o construcción basada en distribuciones conocidas (normal y gamma).

2. Desde el punto de vista muestral, como distribuciones exactas relacionadas con el muestreo de poblaciones normales.

2.2. Distribuciones χ^2 de Pearson, t de Student y F de Snedecor

2.2.1. Distribución χ^2 de Pearson

Se dice que la v.a. X tiene una distribución χ^2 con n ($n \in \mathbb{N}$) grados de libertad si su función de densidad de probabilidad viene dada por

$$f(x) = \frac{1}{\Gamma\left(n/2\right) 2^{n/2}}\, x^{n/2-1} e^{-x/2}, \quad x > 0,$$

y se denota $X \rightsquigarrow \chi^2(n)$[1].

La distribución χ^2 es un caso particular de la distribución Gamma, $\Gamma(p, a)$, cuya densidad es:

$$f(x) = \frac{a^p}{\Gamma(p)}\, x^{p-1} e^{-ax}, \quad x > 0$$

En particular se verifica: $X \rightsquigarrow \chi^2(n) \Leftrightarrow X \rightsquigarrow \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.

Propiedades:

- FGM: $M(t) = \frac{1}{(1-2t)^{n/2}}, \quad t < 1/2$
- Momentos: $E[X^k] = 2^k \dfrac{\Gamma\left(\frac{n}{2} + k\right)}{\Gamma\left(\frac{n}{2}\right)}$. En particular, se tiene: $E[X] = n$, $E[X^2] = n^2 + 2n$, $Var[X] = 2n$.
- Reproductividad: Si $X_1, \ldots, X_n$ son v.a. independientes tales que $X_i \rightsquigarrow \chi^2(k_i)$. Entonces
$$\sum_{i=1}^{n} X_i \rightsquigarrow \chi^2\left(\sum_{i=1}^{n} k_i\right).$$
- Relación con la distribución normal $\mathcal{N}(0, 1)$: Si $X_1, \ldots, X_n$ son v.a.i.i.d. con distribución común $\mathcal{N}(0, 1)$. Entonces
$$\sum_{i=1}^{n} X_i^2 \rightsquigarrow \chi^2(n)$$

Tablas y aproximaciones: La distribución χ^2 está tabulada para valores de n pequeños. Para n grandes, su distribución se puede aproximar por $\mathcal{N}(n, 2n)$.

Ejemplo: Calcula el valor de k o la probabilidad indicada:

[1]La función Γ se define como $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$. Esta función cumple un par de propiedades: $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$; si α es un número entero, $\Gamma(\alpha) = (\alpha - 1)!$.

a) $P[\chi^2(10) \geq k] = 0.005$

b) $P[\chi^2(45) \leq k] = 0.005$

c) $P[\chi^2(14) \geq 21.06]$

d) $P[\chi^2(20) \leq 12.44]$

Gráfica de la función de densidad de $\chi^2(n)$: Cumple las siguientes propiedades

- Es asimétrica a la derecha y unimodal.
- Para $n = 1$, $\lim\limits_{x \to 0} f(x) = +\infty$ y la función de densidad es estrictamente decreciente.
- Para $n = 2$, $f(0) = 1/2$ y la función de densidad es estrictamente decreciente.
- Para $n \geq 3$, $f(0) = 0$, crece hasta la moda y luego decrece.

2.2.2. Distribución t de Student

Sean X e Y variables aleatorias independientes con distribuciones, $X \rightsquigarrow \mathcal{N}(0,1)$ e $Y \rightsquigarrow \chi^2(n)$. Entonces, la v.a.

$$T = \frac{X}{\sqrt{Y/n}}$$

se dice que tiene una distribución t de Student con n grados de libertad. Se denota $T \rightsquigarrow t(n)$.

Propiedades:

- Función de densidad:

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}$$

  Para obtener esta densidad se aplica el cambio de variable indicado en la propia construcción de la distribución, $T = X/\sqrt{Y/n}$, tomando como variable auxiliar $U = Y$.

- Momentos: Sea X una v.a. con distribución $t(n)$, con $n > 1$. Entonces se tiene que existen los momentos $E[X^r]$ para $r < n$ y se verifica que,

  - Si $r < n$ e impar, se tiene que $E[X^r] = 0$.

- Si $r < n$ y par, $E[X^r] = n^{r/2} \dfrac{\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}$.

  En particular, para $n > 2$, existen los momentos de primer y segundo orden, $E[X] = 0$ y $E[X^2] = Var[x] = \dfrac{n}{n-2}$.

Tablas y aproximaciones: La distribución t está tabulada para valores de n pequeños. Para n grandes, su distribución se puede aproximar por $\mathcal{N}(0,1)$.

Ejemplo: Calcula el valor de k o la probabilidad indicada:

a) $P[t(26) \geq k] = 0.05$

b) $P[t(20) \leq k] = 0.25$

c) $P[t(26) \geq k] = 0.9$

d) $P[t(21) \geq 1.721]$

e) $P[t(11) \leq 0.697]$

f) $P[t(8) \leq -2.306]$

Gráfica de la función de densidad de $t(n)$: Cumple las siguientes propiedades

- Es similar a la de la $\mathcal{N}(0,1)$, es decir, simétrica alrededor del cero y unimodal.
- Para $n \to +\infty$ se aproxima a la gráfica de la $\mathcal{N}(0,1)$.
- Tiene colas mayores que las de la normal que van reduciéndose y aproximándose a las de la normal conforme n crece.
- Es más aplastada que la de la normal, es decir es platicúrtica, y su zona central va creciendo y aproximándose a la de la normal conforme n crece.

2.2.3. Distribución F de Snedecor

Sean X e Y variables aleatorias independientes con distribuciones $X \rightsquigarrow \chi^2(m)$ e $Y \rightsquigarrow \chi^2(n)$. Entonces, la v.a.

$$F = \frac{X/m}{Y/n}$$

se dice que tiene una distribución F de Snedecor con (m,n) grados de libertad. Se denota $F \rightsquigarrow F(m,n)$.

Propiedades:

- Función de densidad:
$$g(f) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} f^{m/2-1} \left(1+\frac{m}{n}f\right)^{-\frac{m+n}{2}}, \quad f > 0.$$
Para obtener esta densidad se aplica el cambio de variable indicado en la propia construcción de la distribución, $F = \frac{X/m}{Y/n}$, tomando como variable auxiliar $U = Y$.

- Momentos: Sea X una v.a. con distribución $F(m,n)$. Entonces se verifica que,
$$E[X^r] = \left(\frac{n}{m}\right)^r \frac{\Gamma\left(\frac{m}{2}+r\right)\Gamma\left(\frac{n}{2}-r\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}, \quad \text{para } 0 < r < \frac{n}{2}.$$
En particular, $E[X] = \dfrac{n}{n-2}$ si $n > 2$, $E[X^2] = \dfrac{n^2(m+2)}{m(n-4)(n-2)}$ si $n > 4$ y $Var[X] = \dfrac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}$ si $n > 4$.

- Mediante un cambio de variable se puede comprobar que:
  - $X \rightsquigarrow F(m,n) \Leftrightarrow X^{-1} \rightsquigarrow F(n,m)$.
  - $X \rightsquigarrow t(n) \Leftrightarrow X^2 \rightsquigarrow F(1,n)$.

Tablas y aproximaciones: La distribución F está tabulada para valores de m y n pequeños. Las tablas que se os han proporcionado incluyen aproximaciones para valores grandes de m y n.

Ejemplo: Calcula el valor de k o la probabilidad indicada:

a) $P[F(7,3) \leq k] = 0.95$

b) $P[F(8,4) \geq k] = 0.01$

c) $P[F(2,2) \leq 19]$

d) $P[F(3,5) \geq 12.1]$

e) $P[F(60,40) \leq k] = 0.05$

Gráfica de la función de densidad de $F(m,n)$: Cumple las siguientes propiedades

- Es asimétrica a la derecha y unimodal.
- Para $m = 1$, $\lim\limits_{f \to 0} g(x) = +\infty$ y la función de densidad es estrictamente decreciente.
- Para $m = 2$, $g(0) = 1$ y la función de densidad es estrictamente decreciente.
- Para $m \geq 3$, $g(0) = 0$, crece hasta la moda y luego decrece.

2.3. Muestreo en una población normal unidimensional

Sea X una v.a. con distribución $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ y $\sigma^2 > 0$. Se considera una m.a.s. de tamaño n, $X_1, \ldots, X_n$. Sea $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ la media muestral y $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ la cuasivarianza muestral. El objetivo de este apartado es dar las distribuciones de ambos estadísticos muestrales, para lo cual es necesario el siguiente resultado:

Teorema: Sea $X_1, \ldots, X_n$ una m.a.s. de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Entonces, $\bar{X}$ y $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ son independientes.

Corolarios:

1. $\bar{X} \rightsquigarrow \mathcal{N}(\mu, \frac{\sigma^2}{n})$
2. (Lema de Fisher): $\bar{X}$ y S^2 son independientes.
3. $\dfrac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$
4. $\dfrac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t(n-1)$

2.3.1. Esquemas de resultados para una muestra y uso en inferencia

$$\text{- Inferencia sobre } \mu : \begin{cases} \sigma_0^2 \text{ conocida} & \dfrac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1) \\ \\ \sigma^2 \text{ desconocida} & \dfrac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t(n-1) \end{cases}$$

$$\text{- Inferencia sobre } \sigma^2 : \begin{cases} \mu_0 \text{ conocida} & \dfrac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2} \rightsquigarrow \chi^2(n) \\ \\ \mu \text{ desconocida} & \dfrac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1) \end{cases}$$

2.4. Muestreo en dos poblaciones normales unidimensionales

Sea X una v.a. con distribución $\mathcal{N}(\mu_1, \sigma_1^2)$, $\mu_1 \in \mathbb{R}$ y $\sigma_1^2 > 0$. Se considera una m.a.s. de tamaño n_1, $X_1, \ldots, X_{n_1}$. Sea $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ la media muestral y $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i -$

$\bar{X})^2$ la cuasivarianza muestral.

Sea Y otra v.a. con distribución $\mathcal{N}(\mu_2, \sigma_2^2)$, $\mu_2 \in \mathbb{R}$ y $\sigma_2^2 > 0$. Se considera una m.a.s. de tamaño n_2, $Y_1, \ldots, Y_{n_2}$. Sea $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ la media muestral y $S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$ la cuasivarianza muestral.

Se supone que $(X_1, \ldots, X_{n_1})$ e $(Y_1, \ldots, Y_{n_2})$ son independientes.

Extensión del Lema de Fisher: Los vectores $(\bar{X}, \bar{Y})$ y (S_1^2, S_2^2) son independientes.

Corolarios:

1. $\dfrac{n_2}{n_1}\dfrac{\sigma_2^2}{\sigma_1^2}\dfrac{\sum_{i=1}^{n_1}(X_i - \mu_1)^2}{\sum_{i=1}^{n_2}(Y_i - \mu_2)^2} \rightsquigarrow F(n_1, n_2)$

2. $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \rightsquigarrow F(n_1 - 1, n_2 - 1)$

   En particular, si $\sigma_1 = \sigma_2$, se tiene que $\dfrac{S_1^2}{S_2^2} \rightsquigarrow F(n_1 - 1, n_2 - 1)$

3. $\dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2}{\sigma_1^2} + \frac{(n_2-1)S_2^2}{\sigma_2^2}}}\sqrt{\dfrac{n_1 + n_2 - 2}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow t(n_1 + n_2 - 2)$

   En particular, si $\sigma_1 = \sigma_2$, se tiene que $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}}\sqrt{\frac{n_1+n_2-2}{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow t(n_1 + n_2 - 2)$

2.4.1. Esquemas de resultados para dos muestra y uso en inferencia

- Inferencia sobre $\mu_1 - \mu_2$ (comparación de medias):

$$\begin{cases} \sigma_1^2, \sigma_2^2 \text{ conocidas} & \dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0,1) \\[2em] \sigma_1^2 = \sigma_2^2 = \sigma^2 \text{ desconocidas} & \dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}}\sqrt{\dfrac{n_1 + n_2 - 2}{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow t(n_1 + n_2 - 2) \end{cases}$$

- Inferencia sobre $\dfrac{\sigma_1^2}{\sigma_2^2}$ (comparación de varianzas):

$$\begin{cases} \mu_1, \mu_2 \text{ conocidas} & \dfrac{n_1}{n_2}\dfrac{\sigma_1^2}{\sigma_2^2}\dfrac{\sum_{i=1}^{n_2}(Y_i - \mu_2)^2}{\sum_{i=1}^{n_1}(X_i - \mu_1)^2} \rightsquigarrow F(n_2, n_1) \\[2em] \mu_1, \mu_2 \text{ desconocidas} & \dfrac{\sigma_1^2}{\sigma_2^2}\dfrac{S_2^2}{S_1^2} \rightsquigarrow F(n_2 - 1, n_1 - 1) \end{cases}$$