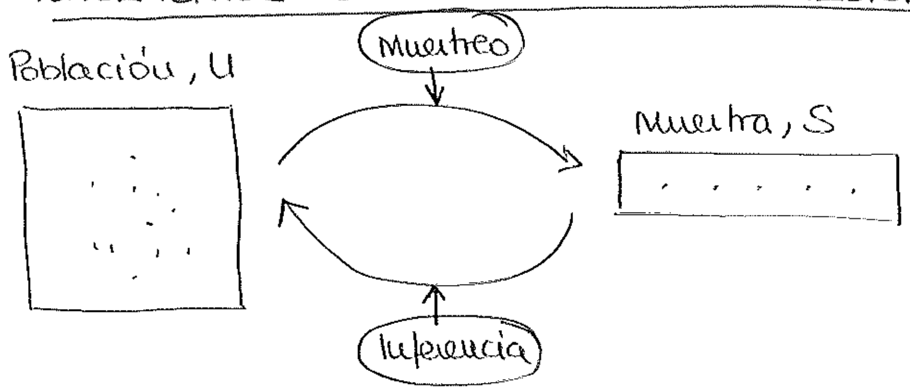
ESTAD – T18. FUNDAMENTOS de la INFERENCIA ESTAD.

CONCEPTO de MUESTRA ALEATORIA.
DISTRIBUCIÓN de la MUESTRA.
ESTADÍSTICOS y SU DISTRIBUCIÓN en el MUESTREO.
FUNCIÓN de DISTRIB. EMPÍRICA y SUS CARACT.
TEOREMA de GLIVENKO–CANTELLI

---

1. FUNDAMENTOS de la INFERENCIA ESTADÍSTICA



→ de la que se obtiene {una serie} características {cualit., cuantit.}.

Tenemos un problema sobre la población, y por falta de dinero, tiempo, etc. nos limitamos a estudiar una parte de dicha población, la muestra, con el objetivo de generalizar los resultados obtenidos en la muestra a toda la población.

Muestreo → Proceso de ~~la~~ selección de los elementos poblacionales que van a formar parte de la muestra.

Inferencia → Proceso de generalización de resultados

Población → cjto de elementos que se pretende estudiar. Fenómeno aleatorio.

Muestra → cjto de elementos que realmente estudiamos. Realización del experimento aleatorio (v.a. n-dim).

La finalidad ideal perseguida al tomar una muestra es la representación a escala de la población, por lo que el procedimiento seguido para obtener la muestra debe garantizar esta representación perfecta.

2_ CONCEPTO de MUESTRA ALEATORIA

Nótese la importancia de la muestra en el proceso inferencial, puesto que de ella dependen los resultados que se generaliza a la población.

Para garantizar la representatividad de la muestra, hay que estudiar los ≠ métodos de obtención de la muestra → muestreo.

Muestreo
- No probabilístico | • Probab. desconocida • No error
- Probabilístico | • Probab. conocida • Error medible y acotable

En el proceso inferencial los resultados obtenidos de la muestra diferirán de los verdaderos valores poblacionales.
El muestreo probabilístico garantiza que las diferencias se deben exclusivamente al azar (muestras ≠) y se puede cuantificar el error cometido → error de muestreo, partiendo del estudio de la probab. de extracción de la muestra.

Muestreo probabilístico
- probab =
- probab ≠

→ SR / CR

Dependiendo del muestreo utilizado, cambiará la distrib. de probabilidad de la muestra.

Muestreo aleatorio ⇒ muestra es v.a. n-dimensional ⇒
⇒ podemos utilizar la $T^{\underline{a}}$ Probab. en su estudio.

Generalmente, en Inferencia Estadística se utiliza muestreo probabilístico con probab. iguales y extracción con reemplazamiento (de igual poblac. finitas que infinitas) ⇒ la distrib. de probab. de la muestra se obtiene como el producto de las distrib. de probab. de las unid. muestrales, que coinciden con la distrib. de probab. de la población.

3_ DISTRIB. de la MUESTRA

Formalmente, el muestreo probabilístico se describe como un espacio de probabilidad:

$U = \{U_1 \ldots U_N\}$ población finita ó infinita

$S = \{S_1 \ldots S_n\}$ conjunto de sucesos elementales, espacio muestral → muestras posibles.

$$P : S \longrightarrow [0,1] \quad / \quad P(S_i) \geq 0$$
$$P(S_i) = P(U_1) \cdot P(U_2/U_1) \ldots \cdot P(U_n/U_1 \cap \ldots \cap U_{n-1})$$

(U, S, P) es un espacio de probabilidad y (S_i, P_i) distrib. de la muestra.

En el caso de muestreo probabilístico con probab iguales y CR, si observamos una característica cuantitativa X sobre la población, la muestra aleatoria simple se puede considerar como la realización del experimento aleatorio n veces de manera independiente:

distrib. conjunta de la muestra

$$P(S) = P(X_1 \ldots X_n) =$$

Caso discreto:

$$= P[(\xi_1 = X_1) \cap (\xi_2 = X_2) \cap \ldots \cap (\xi_n = X_n)] \overset{\text{indep.}}{=} P[\xi_1 = X_1] \cdot \ldots \cdot P(\xi_n = X_n)$$

$$\underset{\substack{\text{idénticamente} \\ \text{distribuidas}}}{=} P(\xi = X_1) \cdot P(\xi = X_2) \cdot \ldots \cdot P(\xi = X_n) \quad / \quad P(\xi = x) \equiv \text{f. cuantía de la poblac.}$$

Caso continuo: $P(x_i < \xi_i < x_i + dx_i) = f(x_i)dx_i \quad , \forall i = 1 \ldots n.$

$$f(x_1 \ldots x_n) = P[x_1 < \xi_1 < x_1 + dx_1 , \ldots , x_n < \xi_n < x_n + dx_n] = P(x_1 < \xi_1 < x_1 + dx_1) \cdot \ldots$$
$$\ldots \cdot P(x_n < \xi_n < x_n + dx_n) = f(x_1) \cdot \ldots \cdot f(x_n)\, dx_1 \ldots dx_n .$$

donde $f(x_i) \equiv$ f. densidad poblac.

En general, en mas se verifica:

$$F(X) = F(X_1) = \ldots = F(X_n)$$
$$F(X_1 \_\_ X_n) = F(X_1) F(X_2) \_\_ F(X_n)$$

v.a.i.i.d.

Si el muestreo hubiese sido sin reemplazamiento:

$$P(X_1=x_1 \ldots X_n=x_n) = P(X_1=x_1) \cdot P(X_2=x_2 / X_1=x_1) \cdot \ldots \cdot P(X_n=x_n / X_1=x_1 \ldots X_{n-1}=x_{n-1})$$

$$f(x_1 \ldots x_n) = f(x_1) \cdot f(x_2/x_1) \ldots f(x_n / x_1 \ldots x_{n-1}).$$

En este punto, cabe destacar la distinción entre modelo probabilístico y modelo muestral, como integrantes del modelo estadístico.

Modelo probabilístico $\longrightarrow$ distrib. probab. de la v.a.
$f(x, \theta)$ / $\theta \equiv$ parámetro/s poblacional/es.

Modelo muestral $\longrightarrow$ distrib. probab. de la muestra
tamaño muestral
tipo muestreo
$\mathcal{L}(X, \theta) \rightarrow X = (x_1 \ldots x_n)$ muestra
$\theta =$ parámetros poblac.

θ fijo, X variable | $\mathcal{L}(X, \theta) \equiv$ f. probab. conjunta de la muestra.

X fijo, θ variable | $\mathcal{L}(X, \theta) \equiv$ f. verosimilitud

4. ESTADÍSTICOS y su distrib. en el MUESTREO

Estadístico → cualquier función de los elementos muestrales que no contenga parámetros desconocidos.

Poblac. ξ → Muestra $X=\{x_1 \dots x_n\}$ → Estadísticos

$T_1(X)=\sum^n x_i$

$T_2(X)=\sum^n x_i^2$

$T_3(X)=\frac{1}{n}\sum^n x_i = \overline{X}$

$T_4(X)=\min\{x_1 - x_n\}=u_1$

$T_5(X)=\sum^n \ln x_i$.

Muestreo aleatorio $\Rightarrow x_i, i=1\dots n$, es v.a. $\Rightarrow$

$\Rightarrow T(X)$ es v.a. → campo de variación
→ distrib. de probab.

Campo de variación: conjunto de valores que toma en cada uno de los elementos del espacio muestral.

Distrib. de probabilidad: distrib. de probabilidad en el muestreo.

$T: X \rightarrow \mathbb{R}$
$(x_1 - x_n) \rightarrow t$ / $P(T(X)=t)=P(\text{Muestras} / T(X)=t)$.

t	$P(T=t)$

← Distrib. en el muestreo

5. FUNCIÓN de DISTRIB. EMPÍRICA y sus características. TEOREMA de GLIVENKO-CANTELLI.

Función de distribución empírica de la muestra, es la frecuencia relativa acumulada de cada elemento muestral de la muestra de tamaño n.

$F_n(x_i) = \frac{N_i}{n}$, $i = 1 \dots n$. | $F_n(x_i) = \frac{\text{nº elementos} \le x_i}{\text{tamaño muestral}}$

$X = (x_1 \dots x_n)$ mas. (n)

Es una función de distrib. de tipo discreto, por lo que su representación gráfica tiene forma de escalera, con tantos saltos como valores muestrales distintos haya y la cuantía del salto el nº de repeticiones de cada valor.

Al ser una función de distrib, hereda sus propiedades:

1 – $F_n(-\infty) = 0$, $F_n(+\infty) = 0$.

2 – $F_n(x)$ es monótona no decreciente

3 – $F_n(x)$ es continua por la derecha

CARACTERÍSTICAS

Momentos muestrales, respecto al origen o respecto a la media.

Sea una poblac. ξ / $E[\xi] = \mu$, $V[\xi] = \sigma^2$,

de la que extraemos una mas(n), $X = \{x_1 — x_n\}$,

definimos:

$$a_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r \longrightarrow E[a_r] = \alpha_r$$

$$V[a_r] = \frac{1}{n}(\alpha_{2r} - \alpha_r^2).$$

$a_1 = \bar{X}$ = media muestral)

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r \longrightarrow E[m_r] = \mu_r + O\left(\frac{1}{n}\right)$$

$$S^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} \ / \ E[S^2] = \frac{n-1}{n}\sigma^2$$

$$S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} \ / \ E[S_1^2] = \sigma^2$$

TEOREMA de GLIVENKO-CANTELLI:

Se considera el teorema fundamental de la estadística, puesto que establece que un tamaño muestral suficientemente grande garantiza que la información contenida en la muestra es prácticamente la misma que la inform. poblacional.

Teorema: $F_n(x) \xrightarrow[n \to \infty]{P} F(x)$

$$\forall \varepsilon > 0, \lim_{n \to \infty} P\left(|F_n(x) - F(x)| \geq \varepsilon \right) = 0 .$$

Este teorema implica que, bajo condiciones generales, las características muestrales convergen en probabilidad a las características poblacionales:

$$\lim_{n \to \infty} P\left(|a_r - \alpha_r| \geq \varepsilon \right) = 0$$

$$\lim_{n \to \infty} P\left(|m_r - \mu_r| \geq \varepsilon \right) = 0$$