Inferencia Estadística

Dra. Graciela González Farías
Dr. Ulises Márquez Urbina
Dr. Edgar Jimenez Peña

Maestría en Cómputo Estadístico

CIMAT Monterrey

Agradecimientos

La mayor parte de estas notas fueron preparadas por la Dra. Graciela González Farías.

En forma de agradecimiento, se enlistan personas que han contribuido de una u otra forma en la construcción de estas notas a través de los años:

- Víctor Muñiz
- Rogelio Ramos Quiroga
- Juan Antonio López
- Sigfrido Iglesias González
- Rodrigo Macías Paéz
- Benito Hernández Chaudary
- Todos los estudiantes que han colaborado con sugerencias y comentarios sobre estas notas desde 1998.

Temario

Objetivos del curso: Proporcionar las bases de la estadística inferencial, orientadas al manejo y análisis de datos.

1. Variables aleatorias y distribuciones de probabilidad.
    1. Distribuciones de probabilidad de variables aleatorias discretas.
    2. Procesos de Poisson.
    3. Distribuciones de probabilidad de variables aleatorias continuas.
    4. Métodos gráficos para la identificación de distribuciones.
    5. Estimación de densidades vía kernels.
    6. Desigualdades para variables aleatorias.
    7. Distribuciones de probabilidad de vectores aleatorios.
    8. Esperanzas condicionales y regresión.
    9. Modelos jerárquicos, compuestos y mezclas de variables aleatorias.
    10. Transformaciones de variables aleatorias.
    11. Simulación de variables aleatorias.
    12. Convergencia de variables aleatorias y el Teorema del Límite Central.

Temario

2. Distribuciones muestrales y métodos de estimación.
    1. Estimación puntual.
    2. Distribuciones muestrales.
    3. Propiedades de los estimadores.
    4. Estimadores de momentos.
    5. Estimadores de máxima verosimilitud y sus propiedades.
    6. Estimación por intervalos.
    7. Método Bootstrap.
3. Pruebas de Hipótesis e intervalos de confianza.
    1. Definición de conceptos.
    2. Potencia de la prueba.
    3. Pruebas para dos poblaciones normales independientes.
    4. Pruebas para medias en muestras pareadas.
    5. Pruebas básicas de varianzas.
    6. Pruebas para proporciones.

Temario

4. Lecturas complementarias:
    1. Cambio de variables para transformaciones de más de una variable
    2. Maximización de la función de verosimilitud.
    3. Otras que los profesores del curso consideren apropiadas y que se le proporcionarán a lo largo del curso.
5. Temas optativos de modelos para presentaciones finales, por ejemplo:
    1. Pruebas no-paramétricas clásicas.
    2. Pruebas de permutaciones.
    3. Estimación no paramétrica (suavizadores y splines).
    4. Pruebas de bondad de ajuste.

    Entre muchas otras opciones que se discutirán con el grupo un mes antes de la fechas de exámenes finales.

Evaluación y acreditación

- Examen parcial: 25 %.
- Evaluación de las tareas (de 2 tipos) y actividades en clase y asistencia: 40 %.
- Un examen final, consistente en una exposición donde se entrega un reporte y se hace una presentación de 1/2 hora. La presentación debe incluir antecedentes, metodología, un ejemplo práctico y compartir el código. Deberán entregar a los instructores y a sus compañeros el resumen. Adicionalmente, deberán dejar un ejercicio sobre el tema a sus compañeros que calificarán en forma honesta: 35 %.

Las tareas tienen una frecuencia quincenal e incluyen TODOS los ejercicios dejados en las notas y requerirán en general el uso de recursos computacionales.

Textos

- **Larry Wasserman (2004) . All of Statistics, A concise course in Statistical Inference. Springer.**
- F.M. Dekking, C. Kraaikamp, H.P. Lopuhaa L.E. Meester (2005). A Modern Introduction to Probability and Statistics, Understanding Why and How. Springer text in Statistics.
- John A. Rice (1995). Mathematical Statistics and Data Analysis, Second Edition. Duxbury Press.
- Casella & Berger. (2002). Statistical Inference, Second Edition . Duxbury Press.
- Richard J. Larsen and Morris L. Marx (2011). An Introduction to Mathematical Statistics and its Applications. Fifth Edition. Prentice Hall.

$\bigcup_{\infty} \Sigma \quad \bigcap^{\infty} \Sigma$

Kolmogrof

$0 \le P \le 1$

Modelos probabilísticos

$(\Omega, \mathcal{F}, P)$

SIGMA ALGEBRA DE BOREL

1

$F(x)$

$0 \quad -\infty \quad \infty$

$F(x) = P(X \le x)$

$(-\infty, x]$

Recta $\mathbb{R}$

POBLACIÓN

LEY RIGE EL ASAR DE COMO SE GOBIERNA

Modelos probabilísticos

V.A $\equiv X$: función medible
$X: \Omega \to \mathbb{R}^+; \mathbb{Z}, \mathbb{N}, \mathbb{R}$

Recordemos que una variable aleatoria discreta es aquella que sólo toma un número contable de valores (finito o infinito). Por ejemplo:

- El número de plantas con daños visibles producidos por una plaga.
- El número de individuos a favor de un partido político.
- El número de televisores con defectos en su selector de canales de un lote de 100 televisores. $\Omega \leftarrow ?$
- El número de personas en la fila en un centro de servicio al público, entre las 9 y 10 de la mañana. ; 11 – 12 ; 1 – 3

Notamos que en cada una de esas situaciones uno lleva a cabo algún tipo de **conteo**.

Modelos probabilísticos

¿?. POBLA

"MUESTRAS"

SESGOS

I
II
III
IV
V

En un principio uno debería:

- Examinar cada caso;
- Ver cuáles son las condiciones específicas en que se realiza el muestreo;
- Establecer los supuestos de simplicidad que sean factibles; y,
- Determinar el modelo probabilístico que mejor describa el comportamiento de la característica bajo estudio (verificando su validez).

"CRISIS DE LA REPLICABILIDAD"

Modelos probabilísticos

farias@cimat.mx $X_1, X_2, \ldots, X_n$ indep. i.d. iid

Este procedimiento general ha dado lugar a un cierto número de modelos que aparecen frecuentemente en las aplicaciones. Así, lo que haremos aquí, es construir estos modelos particulares formando un catálogo básico que nos permita referenciar nuestras situaciones particulares a alguno de estos. En la construcción del catálogo, contemplamos varios puntos:

1. Supuestos necesarios para identificar el uso del modelo. JUST.
2. Construcción del modelo: función de probabilidad y de probabilidad acumulada.
   $p(x)$; $f(x)$ $F(x)$
3. Momentos: Media, Varianza, Función generatriz de momentos (cuando aplique).
   D.S.

$$\infty = \int_{-\infty}^{\infty} x f(x)\,dx = E(X), \quad E(X^2)$$

Cauchy $t=0$

Modelos de variables discretas

Distribución uniforme discreta

Cuando se tiene un número **finito** de resultados de un experimento, $x_1, x_2, \ldots, x_n$, cada uno de ellos **igualmente probable**, se dice que se tiene una variable aleatoria que puede ser modelada por una distribución uniforme discreta. Notación:

$$X \sim U(x_1, x_2, \ldots, x_n).$$

Se lee: X se distribuye como una variable aleatoria uniforme.

Este tipo de modelo aparece típicamente en la selección de **muestras al azar**.

El único parámetro de la distribución es n, el número posible de resultados.

Distribución uniforme discreta

Como la suma de todas las probabilidades debe ser uno y todos los valores son equiprobables, entonces la función de masa está dada por

$$f(x) = \begin{cases} 1/n & \text{para } x = x_1, x_2, \ldots, x_n \\ 0 & \text{en otro caso} \end{cases}$$

Bajo esta definición, claramente $f(x) \geq 0$ para toda x, y

$$\sum_{i=1}^{n} f(x_i) = n\left(\frac{1}{n}\right) = 1.$$

El caso más común de esta distribución es cuando la variable toma los valores enteros

$$\{1, 2, \ldots, n\}.$$

El resto de los resultados los estableceremos para cuando la variable aleatoria uniforme toma estos valores.

Distribución uniforme discreta

El parámetro (valor que identifica unívocamente al modelo) de la distribución es n, el número total de objetos. Se dice en este caso que se trata de un espacio de probabilidad equiprobable.

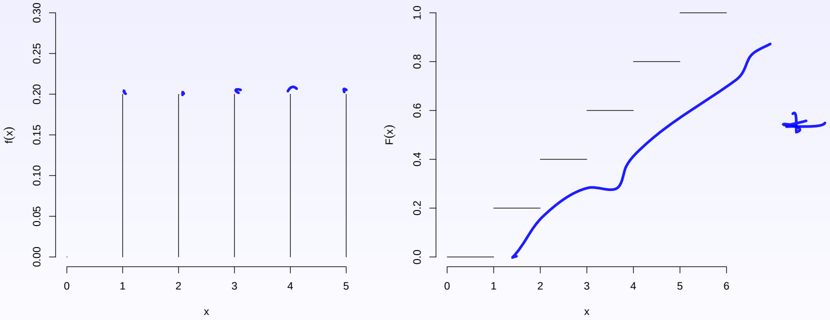

Figura: Distribución uniforme para $n = 5$.

Distribución uniforme discreta

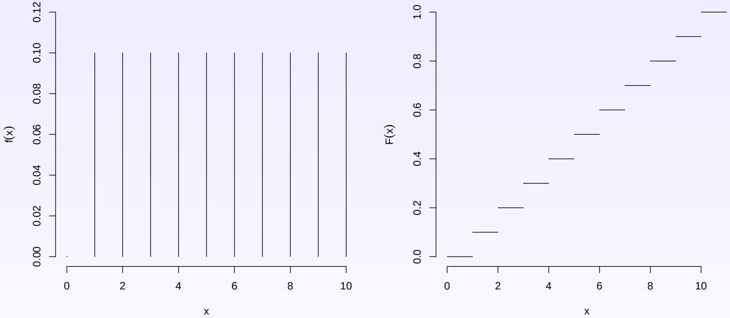

Figura: Distribución uniforme $n = 10$.

Distribución uniforme discreta

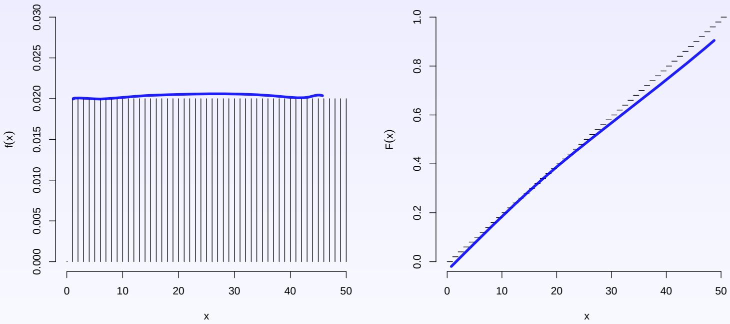

Figura: Distribución uniforme $n = 50$.

Distribución uniforme discreta

Media y Varianza:

$$E(X) = \sum_{x=1}^{n} x\frac{1}{n} = \frac{1+2+\cdots+n}{n} = \frac{n+1}{2},$$

y

$$V(X) = E(X^2) - [E(X)]^2 = \sum_{x=1}^{n} x^2\frac{1}{n} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{2n^2+3n+1}{6} - \frac{n^2+2n+1}{4}$$

$$= \frac{2(2n^2+3n+1) - 3(n^2+2n+1)}{12} = \frac{4n^2+6n+2-3n^2-6n-3}{12}$$

$$= \frac{n^2-1}{12}.$$

Distribución uniforme discreta

Función Generatriz de Momentos:

$\varphi(t)$ $\qquad M_x(t)$ ✗

$$\cancel{M_X}(t) = \mathrm{E}\left(e^{itX}\right) = \sum_{x=1}^{n} e^{tx} \cdot f(x) = \sum_{x=1}^{n} e^{tx} \cdot \frac{1}{n}$$

$$= \frac{1}{n} \sum_{x=1}^{n} \left(e^{t}\right)^{x} = \frac{1}{n}\left[\frac{e^{t} - (e^{t})^{n+1}}{1 - e^{t}}\right] = \underbrace{\frac{1}{n_{-M}} \cdot \frac{e^{t}(1 - e^{\overset{m}{n}t})}{1 - e^{t}}}, \quad \forall t.$$

Nota: $M_X(0) = 0$, aplicando la regla de L'Hopital. $\Leftarrow$ ✓

POB Univ.

Se puede consultar más información sobre la función característica en:
en.wikipedia.org/wiki/Characteristic_function_(probability_theory)

$\Leftarrow$

Distribución Bernoulli

Definamos un experimento en el cual hay únicamente dos posibles resultados: a uno de ellos le llamamos "éxito", al otro "fracaso". Este tipo de variables aparece frecuentemente en nuestros conteos, por ejemplo, si pensamos en clasificar nuestros productos como: defectuoso, no defectuoso; grande o pequeño; azul o blanco; sí o no, etc.

Generalmente le asignamos un valor de 1 al "éxito" y un valor de 0 al "fracaso". Notemos que la asignación de los valores numéricos es arbitraria. Este procedimiento sirve de base para la construcción de otras distribuciones de gran utilidad.

Distribución Bernoulli

"POBLACIÓN"

Definamos $X = \#$ de "éxitos", entonces

$$x = \begin{cases} 1 & \text{si } X = \text{“éxito”}, \text{ esto con probabilidad } p \\ 0 & \text{si } X = \text{“fracaso”}, \text{ esto con probabilidad } (1-p) = q \end{cases}$$

o bien,

$$f(x) = p^x(1-p)^{1-x}, \quad \text{para} \quad x = 0, 1; \qquad \text{(Distribución Bernoulli).}$$

Notación: $X \sim Bernoulli(p)$. El único parámetro de la distribución de ésta variable aleatoria es p, la probabilidad de éxito.

Distribución Bernoulli

Por ejemplo:

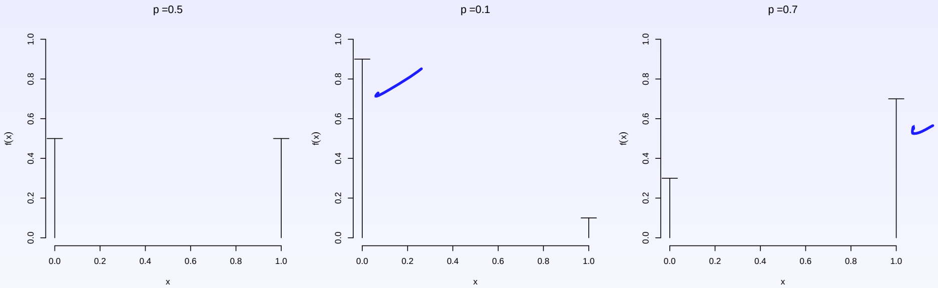

Figura: Función de densidad para $X \sim Bernoulli(p)$.

El experimento que guió al modelo, dos posibles resultados con probabilidades p y q $(p + q = 1)$, se denomina experimento Bernoulli.

Distribución Bernoulli

Media y Varianza:

1 x. "Y = log(x)" 1

$$\mathrm{E}(X) = \sum_{x=0}^{1} x \cdot f(x) = 0 \cdot (1-p) + 1 \cdot p = p$$

$$\mathrm{E}(X^2) = \sum_{x=0}^{1} x^2 \cdot f(x) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

$$\mathrm{V}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = p - p^2 = p(1-p) = pq$$

$(1-p)$

$p - p^2$

Función Generatriz de Momentos:

$$M_X(t) = \mathrm{E}\left(e^{tX}\right) = \sum_{x=0}^{1} e^{tx} \cdot f(x) = e^{t \cdot 0}(1-p) + e^{t \cdot 1} p$$

$$= (1-p) + pe^t = q + pe^t, \quad \forall t.$$

"$Y = g(X)$"

Distribución Binomial e Hipergeométrica

Para las siguentes distribuciones definiremos nuestra variable aleatoria de la siguiente manera:

POBL.

$$X = \# \text{ de “éxitos” en } n \text{ observaciones.}$$

X es el número de unos o “éxitos” que se presenten en la muestra de tamaño n. Lo que asigna un patrón distinto de comportamiento, es la **forma como se realizan (condiciones) las n observaciones**.

Distribución Binomial

Supongamos que podemos hacer:

1. n observaciones independientes.
2. la probabilidad de éxito en cada observación permanece constante, esto es, siempre es p.

En otras palabras, estamos asumiendo que nuestra población es suficientemente grande como para tomarla como "infinita" y que podemos garantizar la independencia entre observaciones, esto es, no obtenemos información adicional para predecir el siguiente resultado sólo porque ya observamos al (los) anterior (es). Esto refleja un comportamiento de procesos que lo podríamos denominar estable. Estas condiciones deben estar presentes al menos durante el período en que se realiza el estudio.

Distribución Binomial

Por ejemplo, pensemos en un proceso industrial, producción de mangueras para gas. Las clasificaremos como defectuosas (éxito) o no defectuosas, de acuerdo a si cumplen o no con el tamaño requerido. Se toma una muestra de tamaño tres y se hacen las mediciones de cada una de las mangueras.

- Asumimos que nuestra población son todas las mangueras que pasan por ese proceso (# muy grande),
- Se asume que no hay ninguna causa que motive desperfectos sistemáticos,
- También notemos que, bajo estas condiciones, la probabilidad de que una manguera no sea de las medidas requeridas deberá permanecer constante para las 3 observaciones y, más aún, esta probabilidad está dada por otro mecanismo independiente al conteo que nos atañe en este momento.

Distribución Binomial

Por ejemplo, p se podría determinar como una proporción observada a través del tiempo, o bien porque conozcamos la ley de fallas en cortes, de la maquinaria que se emplea. Para nuestros propósitos, p es un valor dado.

Notemos primero que el número posible de valores que puede tomar X es $\{1, 2, \ldots, n\}$. Ahora obtengamos la ley de comportamiento asociada con esta variable:

$$\begin{aligned} P(X = 0) &= P(\text{ninguna manguera defectuosa en la muestra de } n = 3) \\ &= P(X_1 = 0 \overset{y}{,} X_2 = 0 \overset{y}{,} X_3 = 0) \quad \longleftarrow P(X_1 \cap X_2 \cap X_3) \end{aligned}$$

donde X_i denota la i-ésima observación. Cabe notar que cada X_i toma el valor de 0 ó 1 con probabilidades constantes q y p, respectivamente. Esto es, cada X_i **representa una variable Bernoulli**(p) **y son independientes entre sí**.

Distribución Binomial

Entonces:

$$
\begin{aligned}
P(X_1 = 0, X_2 = 0, X_3 = 0) &= P(X_1 = 0)P(X_2 = 0)P(X_3 = 0) \\
&= qqq = q^3 = \left(\begin{array}{c} 3 \\ 0 \end{array}\right) q^3
\end{aligned}
$$

A manera de ejercicio, puedes verificar que:

$$
\begin{aligned}
P(X = 1) &= P(\{X_1 = 1, X_2 = 0, X_3 = 0\} \text{ ó } \{X_1 = 0, X_2 = 1, X_3 = 0\} \text{ ó } \{X_1 = 0, X_2 = 0, X_3 = 1\}) \\
&= 3pqq = 3pq^2 = \left(\begin{array}{c} 3 \\ 1 \end{array}\right) pq^2, \qquad \textbf{1} \ \textbf{“éxito”}; \\
P(X = 2) &= P(\{X_1 = 1, X_2 = 1, X_3 = 0\} \text{ ó } \{X_1 = 1, X_2 = 0, X_3 = 1\} \text{ ó } \{X_1 = 0, X_2 = 1, X_3 = 1\}) \\
&= 3ppq = 3p^2q = \left(\begin{array}{c} 3 \\ 2 \end{array}\right) p^2q, \qquad \textbf{2} \ \textbf{“éxitos”}; \\
P(X = 3) &= P(\{X_1 = 1, X_2 = 1, X_3 = 1\}) \\
&= ppp = 3p^3 = \left(\begin{array}{c} 3 \\ 3 \end{array}\right) p^3, \qquad \textbf{3} \ \textbf{“éxitos”}.
\end{aligned}
$$

Distribución Binomial

Esto es, tenemos la Distribución Binomial

$$f(x) = \begin{cases} \underbrace{\binom{n}{x} p^x (1-p)^{n-x}} & \text{para } x = 0, 1, \ldots, n \\ 0 & \text{en otro caso} \end{cases}$$

En este caso se dice que X sigue una distribución binomial con parámetros (n, p). Notación: $X \sim B(n, p)$.

Distribución Binomial