Inferencia Estadística

Dra. Graciela González Farías
Dr. Ulises Márquez

Maestría en Cómputo Estadístico.

CIMAT Monterrey.

Agradecimientos

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
- Juan Antonio López
- Sigfrido Iglesias González
- Rodrigo Macías Paéz
- Edgar Jiménez
- Todos los estudiantes que han colaborado con sugerencias y comentarios sobre estas notas.

Estas notas son de uso exclusivo para enseñanza y no pretende la sustitución de los textos y artículos involucrados.

Temario

1. Variables aleatorias y distribuciones de probabilidad.
   a) Distribuciones de probabilidad de variables aleatorias discretas.
   b) Procesos de Poisson.
   c) Distribuciones de probabilidad de variables aleatorias continuas.
   d) Métodos gráficos para la identificación de distribuciones.
   e) Estimación de densidades.
   f) Distribuciones de probabilidad de vectores aleatorios.
   g) Esperanzas condicionales y regresión.
   h) Modelos jerárquicos, compuestos y mezclas de variables aleatorias.
   i) Transformaciones de variables aleatorias.
   j) Simulación de variables aleatorias.
   k) Convergencia de variables aleatorias y el Teorema del Límite Central.

Temario

2. Distribuciones muestrales y métodos de estimación.
   a) Propiedades de los estimadores.
   b) Estimadores no insesgados
   c) Distribuciones muestrales.
   d) Principio de máxima verosimilitud.
   e) Estimación puntual.
   f) Bootstrap y jackknife.

Temario

3. Pruebas de Hipótesis e intervalos de confianza.
   a) Definición de conceptos.
   b) Potencia de la prueba.
   c) Pruebas para dos poblaciones normales independientes.
   d) Pruebas para medias en muestras pareadas.
   e) Pruebas básicas de varianzas.
   f) Pruebas para proporciones.
   g) Conceptos de estimación bayesiana.
   h) 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. Modelos gráficos probabilístas.
      5. Entre muchos otros.

Evaluación y acreditación

- Dos exámenes parciales, 18 de septiembre y 6 de noviembre: 15 %, cada uno.
- 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: 30 %.

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.

Modelos probabilísticos

Modelos probabilísticos

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.
- El número de personas en la fila en un centro de servicio al público, entre las 9 y 10 de la mañana.

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

Modelos probabilísticos

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).

Modelos probabilísticos

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.
2. Construcción del modelo: función de probabilidad y de probabilidad acumulada.
3. Momentos: Media, Varianza, Función generatriz de momentos (cuando aplique).

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.

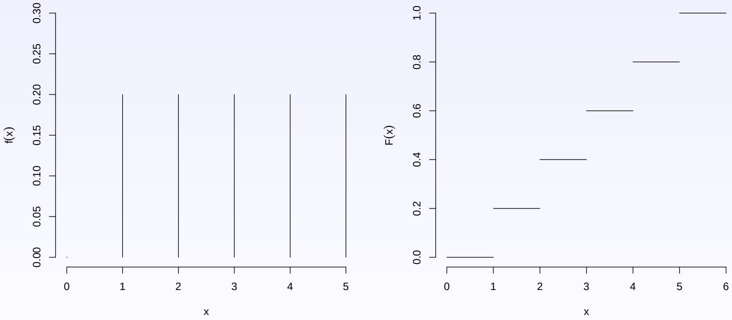

Figura: Distribución uniforme para $n = 5$.

Distribución uniforme discreta

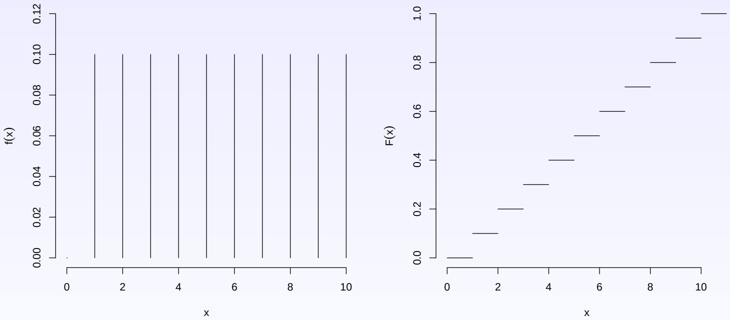

Figura: Distribución uniforme $n = 10$.

Distribución uniforme discreta

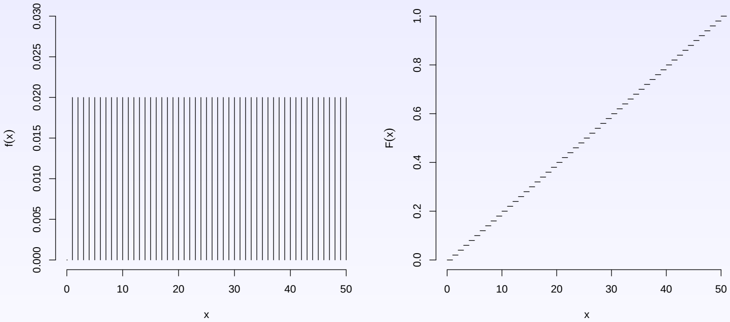

Figura: Distribución uniforme $n = 50$.

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

$$\begin{aligned}
V(X) = \; & E(X^2) - [E(X)]^2 = \sum_{x=1}^{n} x^2\frac{1}{n} - \left(\frac{n+1}{2}\right)^2 \\
= \; & \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{2n^2+3n+1}{6} - \frac{n^2+2n+1}{4} \\
= \; & \frac{2(2n^2+3n+1) - 3(n^2+2n+1)}{12} = \frac{4n^2+6n+2-3n^2-6n-3}{12} \\
= \; & \frac{n^2-1}{12}.
\end{aligned}$$

Distribución uniforme discreta

<u>**Función Generatriz de Momentos**</u>:

$$
\begin{aligned}
M_X(t) = \; & \mathrm{E}\left(e^{tX}\right) = \sum_{x=1}^{n} e^{tx} \cdot f(x) = \sum_{x=1}^{n} e^{tx} \cdot \frac{1}{n} \\
= \; & \frac{1}{n} \sum_{x=1}^{n} \left(e^{t}\right)^{x} = \frac{1}{n}\left[\frac{e^{t} - (e^{t})^{n+1}}{1-e^{t}}\right] = \frac{1}{n} \cdot \frac{e^{t}(1-e^{nt})}{1-e^{t}}, \quad \forall t.
\end{aligned}
$$

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