

A los asistentes a los cursos del Centro de Educación

Continua

La Facultad de Ingeniería, por conducto del Centro de Educación Continua, otorga constancia de asistencia a quienes cumplan con los requisitos establecidos para cada curso. Las personas que deseen que aparezca su título profesional precediendo a su nombre en el diploma, deberán entregar copia del mismo o de su cédula profesional a más tardar 15 días antes de la terminación del curso, en las oficinas del Centro, con la Sra. Sánchez.

El control de asistencia se efectuará al terminar la primera hora de cada día de clase, mediante listas especiales en las que los interesados anotarán personalmente su asistencia. Las ausencias serán computadas por las autoridades del Centro.

Se recomienda a los asistentes participar activamente con sus ideas y experiencias, pues los cursos que ofrece el Centro están planeados para que los profesores expongan una tesis, pero sobre todo para que coordinen las opiniones de todos los interesados constituyendo verdaderos seminarios.

Al finalizar el curso se hará una evaluación del mismo a través de un cuestionario diseñado para emitir juicios anónimos por parte de los asistentes.

Las personas comisionadas por alguna institución deberán pasar a inscribirse en las oficinas del Centro en la misma forma que los demás asistentes.

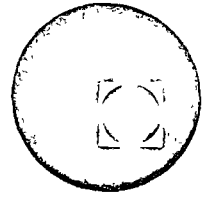
2

3

4



centro de educación continua
facultad de ingeniería, unam



GEOESTADISTICA

DR. HENRI SANS.



CONSTRUCCIÓN DE UN VARIOGRAMA EXPERIMENTAL

Aplicación práctica

Los datos provienen de la simulación de una realización de un variograma lineal $\gamma(h) = \bar{c}|h|$, hecha con una computadora.

Los datos son colocados en los nudos de una reticula de malla a . Se calculará el variograma en las 4 direcciones principales $\alpha = 1, 2, 3, 4$ y por valores de h tales como $a, 2a, 3a$ en las direcciones 1 y 2, $a\sqrt{2}, 2a\sqrt{2}, 3a\sqrt{2}$ en las direcciones 3 y 4.

El variograma experimental que se calcula en la dirección α es

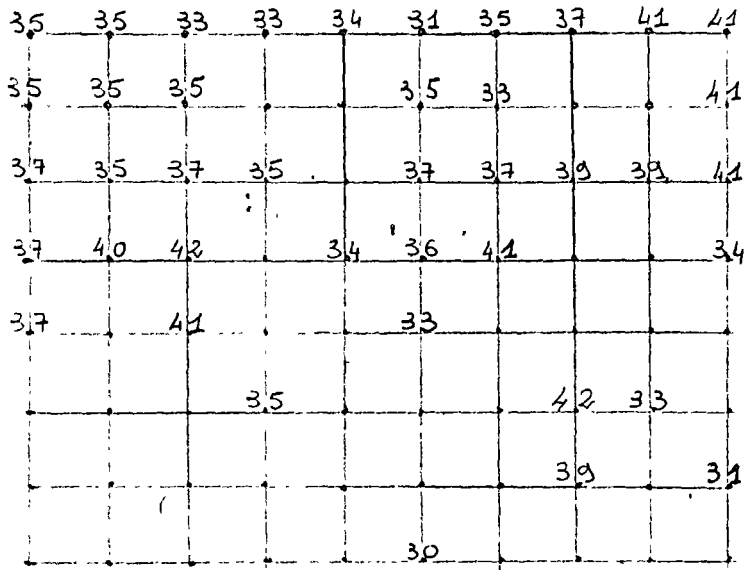
$$\gamma_{\alpha}^*(h) = \frac{1}{2} \frac{\sum_{x_i} \{ [f(x_i+h) - f(x_i)]^2 \}}{n(h)}$$

- $f(x_i+h)$ es el valor de la variable en el punto experimental x_i+h
- $f(x_i)$ es el valor de la variable en el punto experimental x_i
- $n(h)$ es el número de diferencias $[f(x_i+h) - f(x_i)]$

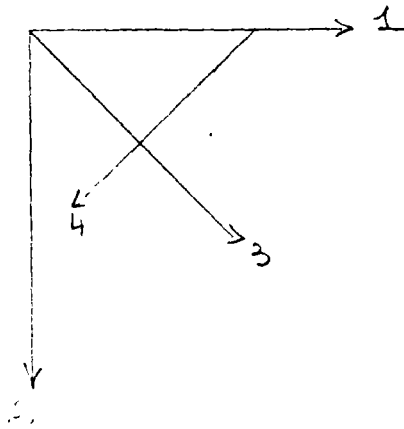
Se nota que el variograma no es nada más que el valor medio de $[f(x+h) - f(x)]^2$ en la dirección α .

El número de datos es tan pequeño que los cálculos se pueden hacer a mano.

Se disimularán los variogramas experimentales en cada dirección -
¿sus conclusiones se pueden sacar?



Los nudos de la red que no tienen valores son puntos desconocidos. En los cálculos de los variogramas no se deben tomar en cuenta.



1, 2, 3 y 4 son las cuatro direcciones principales.

ESTIMACION GLOBAL DEL YACIMIENTO

DE ITY - MONT - FLOTOUO

Estudio real

Presentación del problema

El yacimiento de ITY está constituido por una corteza aurífera de latente y arrolla con una potencia que varía entre 3 y 5 m. El reconocimiento se hizo por medio de una red de barrenos, rectangular, con separaciones de 20×30 m. Cada barreno fue muestreado a lo largo de la potencia mineralizada.

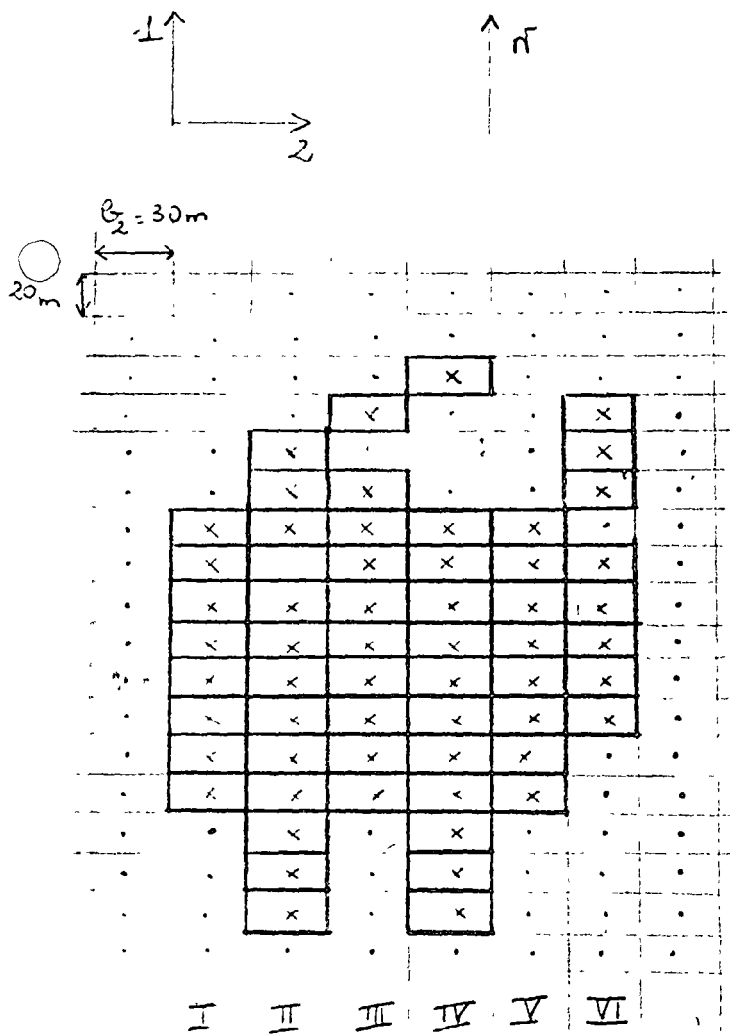


Fig 1

La localización de los barrenos positivos (es decir en los cuales se obtuvo mineralización) se indica en la fig 1. La superficie que se considera mineralizada es delimitada por el contorno poligonal obtenido por la "reunión de todos los "adocquinos" de 20×30 m² positivos". Al N.W se nota una laguna de mineralización. Hacia el Norte, la mineralización es bastante irregular y las zonas de influencia pueden ser objeto de discusión.

En el caso de este pequeño yacimiento, la geología no puede definir los límites de la mineralización. En lo que se refiere a la estimación de la superficie mineralizada, el problema es de calcular la varianza de estimación a partir de una red regular: estamos pues dentro del dominio de aplicación de la Teoría Transitiva.

La variable regionalizada que se estudia es la acumulación relativa $a(x)$ (1) u(x), es la acumulación medida en el punto x .

Se tiene $a(x) = p(x) \cdot t(x)$

donde $p(x)$ representa la potencia mineralizada y $t(x)$ la ley de la muestra.

a es la acumulación media obtenida con los $n = 58$ barrenos positivos: a no es otro que el estimador de la acumulación media A real, desconocida del yacimiento, dentro del contorno poligonal S^*

El valor $a = \frac{\sum a(x_i)}{n}$ no es comunicado por razón de discreción, pero esto no afecta nuestro estudio. Solamente, las varianzas de estimación que vamos a calcular ya serán varianzas relativas:

$$\frac{D^2[a(x)]}{a^2} = D^2\left[\frac{a(x)}{a}\right] = \frac{\sigma_a^2}{a^2}$$

Con los $n = 58$ barrenos positivos, se obtuvo el semi-variograma γ^* -S de las acumulaciones relativas:

h	$\gamma^*(h)$
20m	0.64
40	0.98
60	1.00
80	1.10
100	1.15
120	1.05
140	1.02

$$\gamma^*(h) = \frac{1}{2n(h)a^2} \sum_{x_i} \left\{ [a(x_i+h) - a(x_i)]^2 \right\}$$

La varianza a priori de nuestra VR, o varianza relativa de las acumulaciones dentro de todo el campo mineralizado (S^*) se estima a partir de los barrenos a $\sigma^2 = 1,0$

Se verificó la isotropía aproximativa de $\gamma^*(h)$

(1) Se un problema en 3 dimensiones (estudio de la ley), pasamos a un problema en 2 dimensiones (estudio de la acumulación) = ejemplo de la operación de subida que no modifica las formulas de las Varianzas.

Se buscan:

- ① La estimación de la superficie mineralizada con la ayuda del contorno poligonal.
- ② La varianza relativa de estimación de esta superficie, sea $\frac{Q_{\text{est}}^2}{S^2}$
- ③ El ajuste del variograma experimental a un modelo. Se precisarán los parámetros característicos (efecto de pepita, meseta, alcance).
- ④ La varianza relativa de estimación de la acumulación media en el caso de que el campo mineralizado sea delimitado efectivamente por el contorno poligonal. Este cálculo se hará de dos maneras diferentes:
 - A) utilizando el principio de composición de los términos de línea y de banda.
 - B) utilizando el principio de composición de las varianzas de extensión elementales.
- ⑤ La varianza relativa de estimación de la acumulación media tomando en cuenta el efecto de borde, debido al error que se comete en la estimación del campo efectivamente mineralizado.
- ⑥ La varianza relativa de la estimación global de la cantidad de metal.

① Estimación de la superficie mineralizada con la ayuda del contorno

por áreas

$$S^* = n \beta_1 \beta_2 = 58 \times 20 \times 30 = 34800 \text{ m}^2$$

② Variación relativa de estimación de esta superficie

Se admite que las direcciones principales de anisotropía son las direcciones de la red N-S y E-O. En este caso, se puede aplicar la fórmula

$$\frac{\sigma_s^2}{S^2} = \frac{1}{n^2} \left[0.167 N_2 + 0.061 \frac{N_1^2}{N_2} \right] \quad (N_2 \leq N_1)$$

• n indica el número de los terrenos positivos.

• $2N_1$ y $2N_2$ indican los números de los elementos paralelos a cada uno de los ejes, constituyendo el contorno poligonal, sin olvidar los de la longitud. N_2 es el más pequeño de los dos.

Aquí tenemos:

$$\text{en la dirección N-S} \quad 2N_1 = 42 + 2 = 44$$

$$\text{en la dirección E-O} \quad 2N_2 = 18 + 2 = 20$$

$$\text{Por tanto} \quad N_1 = 22$$

$$N_2 = 10$$

$$\frac{\sigma_s^2}{S^2} = \frac{1}{58^2} \times \left[0.167 \times 10 + 0.061 \times \frac{22^2}{10} \right] = 0.00137$$

$$\frac{\sigma_s}{S} = 0.00137 \quad \text{sea} \quad \frac{\sigma_s}{S} = 0.037 \quad \text{y} \quad \sigma_s = 0.037 \times 34800 = 1300 \text{ m}^2$$

Por tanto, un riesgo de 15% según la ley normal ($P\{S < S^* - \sigma_s\}$)

$$\text{no debería} \quad \Rightarrow S^* - \sigma_s = 33500 \text{ m}^2$$

Nota relativa a la red de barrenos -

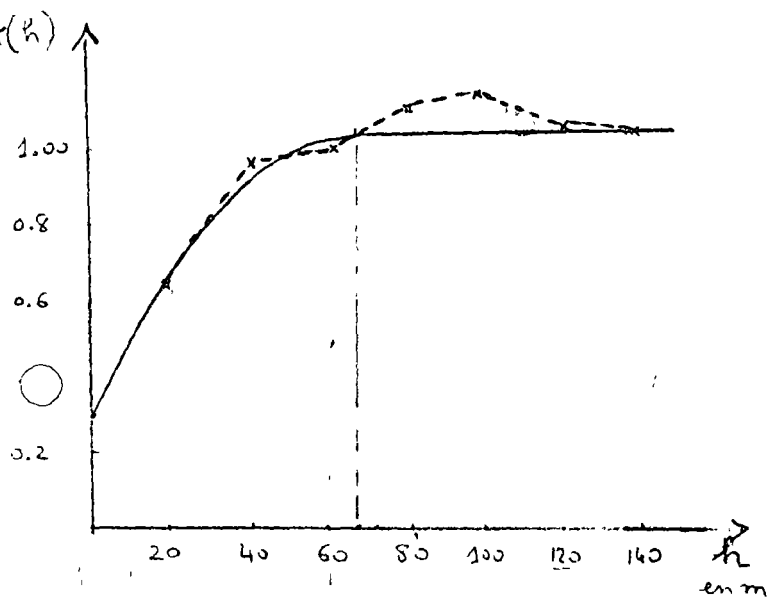
Las variaciones diametrales son

- en la dirección N-S : $D_1 = r_1 b_1 = 22 \times 20 = 440 \text{ m}$

- en la dirección E-O : $D_2 = r_2 b_2 = 10 \times 30 = 300 \text{ m}$

Al lado más largo de la red le corresponde la variación diametral más pequeña ($2D_2$) - Por lo tanto se ve que esta red rectangular no fue la adecuada y que hubiera sido mejor una red cuadrada.

③ Ajuste de $\gamma^*(h)$ a un modelo.



Se procede a un ajuste "de visu". El crecimiento muy rápido del variograma experimental nos lleva a ajustarlo a un esquema esférico con efecto de pepita.

$$\gamma(h) = C_0 + \begin{cases} C \left[\frac{3}{2} \frac{h}{a} - \frac{1}{2} \frac{h^3}{a^3} \right] & h < a \\ C & h \geq a \end{cases}$$

efecto de pepita -

Parámetros del modelo

- efecto de pepita C_0 : se estima la tangente al origen del esquema por los 2 primeros puntos experimentales (los mejores conocidos).

$$p = \frac{0.98 - 0.64}{20} = 0.017$$

Extrapolando al eje de las ordenadas, se obtiene $C_0 = 0.30$.

- Meseta : se estimó la meseta a $C_0 + C = 1.05$, sea $C = 0.75$

Sabiendo que la tangente al origen al esquema de Matérn cita la meseta en un punto de abscisa igual a los $\frac{2}{3}$ del alcance

$$\frac{2}{3} a = 44 \text{ m} \Rightarrow \underline{a = 66 \text{ m}}$$

El esquema queda pues completamente definido.

Problemas se la varianza a priori $\sigma^2(0/S^*)$ correspondiente a este modelo se acerca a la varianza a priori experimental. Por eso, le da la forma a la superficie mineralizada de un rectángulo $200 \times 240 \text{ m}^2$, sea, ya que $a = 66 \text{ m}$, a un rectángulo $3a \times 3,5a$.

$\sigma^2 = F(l, h) = 0.30 + 0.75 \times 0.95$ con la ayuda del ábaco de la función del esquema esférico.

fues $\sigma^2 = 1,01$ valor muy cercano al valor experimental, $1,0$. Entonces se puede considerar que este ajuste es correcto.

Notemos el efecto de pepita fuente: $\epsilon = \frac{C_0}{C} = \frac{0.3}{0.75} = 0.4$, lo que es normal por un yacimiento de oro.

(4) Varianza de estimación de la acumulación media.

Se supone aquí que la superficie mineralizada es conocida - se conoce el estimador a de la acumulación media A real, desconocida - a es, por ejemplo, el promedio aritmético de los 58 valores experimentales positivos.

A) Principio de composición de los términos de línea y de banda

Se calculará la varianza de estimación de A por a utilizando el principio de composición:

- de un término de línea (correspondiente a las líneas de mayor densidad dirección 1 o NS) reduciendo la extensión de los pozos centrales a sus segmentos de influencia ($b_1 = 20 \text{ m}$)

- de un término de banda reduciendo la extensión de las líneas l_i supuestas asociadas a sus familias de influencia (rectángulo $l_i \times b_2$)

Término de línea

Varianza de extensión de un pozo central a su segmento $b_1 = 20 \text{ m}$ de influencia $C_0 + C_1 \sigma_L^2(b_1)$

Se lee en el ábaco: $\frac{b_1}{a} = \frac{20}{66} = 0.30 \Rightarrow \sigma_L^2(b_1) = 0.08$.

... da de $C + C \cdot \sigma_b^2(b_2) = 0.3 + 0.75 \times 0.08 = 0.36$.

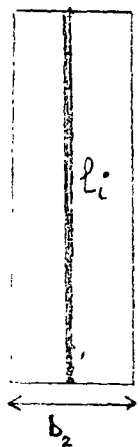
Temos, que el efecto de pepita tiene la mayor importancia.

Suponendo la independencia de las extensiones de cada pzo a su segmento b_2 de influencia, la varianza de estimación de la totalidad de las líneas o término de línea es:

$$T_L = \frac{0.36}{n} = \frac{0.36}{58} = 0.0062$$

Término de Banda.

Varianza de extensión de una línea l_i reconocida como antes a su banda de influencia ($l_i \times b_2$): $C \times \sigma_b^2(b_2; l_i)$



- el término $\sigma_b^2(b_2; l_i)$ se lee en un abaco por cada l_i

- el efecto de pepita no interviene ya que desaparece debido a la subida bajo la potencia constante l_i .

- Las longitudes l_i siendo diferentes, tenemos que estudiarlas una por una y luego ponderar por l_i^2

Por todas las líneas: $\frac{b_2}{a} = \frac{30}{66} = 0.45$ y $C = 0.75$

línea i	l_i	l_i/a	$\sigma_b^2(b_2; l_i)$	$C \sigma_b^2(b_2; l_i)$	$l_i^2 C \sigma_b^2(b_2; l_i)$
I	160m	2.4	0.010	0.0075	192
II	240	3.6	0.008	0.0060	345
III	200	3.0	0.009	0.0067	268
IV	240	3.6	0.008	0.0060	345
V	160	2.4	0.010	0.0075	192
VI	160	2.4	0.010	0.0075	192

$\Sigma = 1460m$

$\Sigma = 1534$

Término de Banda

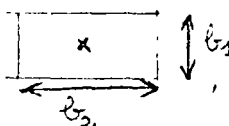
$$T_b = \frac{C \sum_i p_i^2 \sigma_b^2(b_i, p_i)}{(\sum_i p_i)^2} = \frac{1.524}{(1.160)^2} = 0.0011.$$

De donde la varianza de estimación de la acumulación dentro de S^*

$$\frac{D^2[a-A]}{a^2} = T_L + T_b = 0.0062 + 0.0011 = 0.0073$$

B) Principios de composición de las varianzas de extensión elom. ribs

Extensión de un pozo central a su rectángulo de influencia:


$$C_0 + C \sigma_e^2(b_1, b_2)$$
$$\frac{b_1}{a} = 0.30 \quad \frac{b_2}{a} = 0.45$$

Después de usar el álgebra $\sigma_e^2 = 0.15$

$$\text{De donde } (C_0 + C \sigma_e^2(b_1, b_2)) = 0.30 + 0.75 \times 0.15 = 0.40$$

El efecto del pepita tiene la mayor importancia.

Suponiendo la independencia de las extensiones de cada pozo a su rectángulo de influencia, la varianza de estimación de la ley relativa es

$$\frac{D^2[a-A]}{a^2} = \frac{0.40}{58} = 0.0068$$

Los 2 principios dan casi el mismo resultado

③ Efecto de Borde

Estimaremos como varianza de estimación de la acumulación media a superficie no corregida supuesta conocida la que se calculó según el principio de composición de términos de línea y de banda. De hecho, no es perfectamente conocida la superficie no corregida y el error geométrico repercute sobre la estimación de la acumulación media.

Se debe de usar pues un término suplementario que se añade a los términos

$$T_0 = \frac{\sigma^2(O/S^*)}{a^2} \times \frac{\sigma_s^2}{S^2} = 1,0 \times 0.00137 = 0.0014$$

○ En fin resulta,

$$\frac{D^2[a-A]}{a^2} = T_l + T_b + T_0 = 0.0087.$$

⑥ Varianza de estimación de la cantidad de metal Q

$$Q = aS$$

La hipótesis de independencia interna entre la VR y su campo nos permite escribir

$$\frac{D^2[Q]}{Q^2} = \frac{D^2[a]}{a^2} + \frac{D^2[S]}{S^2}$$

$$= 0.0087 + 0.0014 = 0.01$$

El error sobre la acumulación es mucho más grande que el error sobre la superficie.

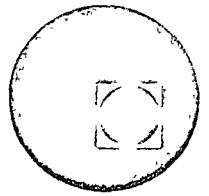
○ $\frac{\sigma(Q)}{Q} = 0.1$ con un riesgo de 15% y según una ley normal

$$Q > Q^* - \sigma(Q) = 0.9 Q^*$$





centro de educación continua
facultad de ingeniería, unam



GEOESTADISTICA

PROBABILIDAD Y ESTADISTICA

ING. SERGIO E. ARREDONDO GALLARDO.

a

NOTAS SOBRE

PROBABILIDAD

Y

ESTADISTICA

SERGIO E. ARREDONDO GALLARDO

OBJETIVO

CONOCER LOS FUNDAMENTOS DE PROBABILIDAD Y ESTADISTICA NECESARIOS PARA EL ESTUDIO Y COMPRESION DE LOS TEMAS DE GEOESTADISTICA.

CONTENIDO

PAG.

I. CONCEPTOS FUNDAMENTALES

1. ELEMENTOS DE TEORIA DE CONJUNTOS 2
 - A. DEFINICION DE CONJUNTO. NOTACION.
 - B. OPERACIONES ENTRE CONJUNTOS.
LEYES.
 - C. DIAGRAMAS DE VENN-EULER
 - D. CONJUNTOS FINITOS Y NUMERABLES.
 - E. CONJUNTO PRODUCTO.
 - F. PARTICION DE UN CONJUNTO.
 - G. PROBLEMAS.
2. TECNICAS DE CONTAR 12
 - A. PERMUTACIONES
 - B. COMBINACIONES
 - C. MUESTRAS ORDENADAS
 - D. DIAGRAMAS DE ARBOL
 - E. PROBLEMAS

II. PROBABILIDAD

1. EXPERIMENTO. ESPACIO MUESTRAL. 19
EVENTOS.
2. DIFERENTES CONCEPTOS DE PROBABILIDAD 21
 - A. DEFINICION CLASICA.

B. DEFINICION DE FRECUENCIA RELATIVA.	
C. DEFINICION AXIOMATICA.	
3. ESPACIOS DE PROBABILIDAD	26
4. PROBABILIDAD CONDICIONAL. INDEPENDENCIA.	28
5. PROBLEMAS	33

III. VARIABLES ALEATORIAS

1. UNA VARIABLE ALEATORIA	36
A. DEFINICION DE VARIABLE ALEATORIA.	
A.1. VARIABLES ALEATORIAS DISCRETAS.	
A.2. VARIABLES ALEATORIAS CONTINUAS.	
2. DOS VARIABLES ALEATORIAS	42
A. VARIABLES ALEATORIAS DISCRETAS.	
B. VARIABLES ALEATORIAS CONTINUAS.	
3. FUNCIONES DE VARIABLES ALEATORIAS	52
A. FUNCION DE UNA VARIABLE ALEATORIA	
B. FUNCION DE DOS VARIABLES ALEATORIAS	
4. VALORES ESPERADOS Y MOMENTOS	59
A. VALORES ESPERADOS	
B. MOMENTOS	
C. VALOR ESPERADO CONDICIONAL.	
5. DISTRIBUCIONES ESPECIALES	67
A. VARIABLES ALEATORIAS DISCRETAS	
A.1. DISTRIBUCION BINOMIAL	
A.2. DISTRIBUCION DE POISSON	
B. VARIABLES ALEATORIAS CONTINUAS	
B.1. DISTRIBUCION UNIFORME	

e

PAG.

B.2. DISTRIBUCION NORMAL

B.3. DISTRIBUCION LOG-NORMAL

6. PROBLEMAS

71

IV. REDUCCION DE DATOS

ANEXO A.

BIBLIOGRAFIA

(1)
I. CONCEPTOS

FUNDAMENTALES

OBJETIVO

RECORDAR LOS CONCEPTOS NECESARIOS
PARA EL ESTUDIO DE LA PROBABI-
LIDAD Y ESTADISTICA.

(2)

1. ELEMENTOS DE TEORIA DE CONJUNTOS

A. DEFINICION DE CONJUNTO. NOTACION.

- * CONJUNTO DE NUMEROS ENTEROS POSITIVOS

$$N = \{1, 2, 3, \dots\}$$

- * CONJUNTO DE LAS LETRAS DEL ALFABETO

$$A = \{a, b, \dots, y, z\}$$

- * CONJUNTO DE PUNTOS DE UNA CIRCUNFERENCIA DE RADIO 4 Y CENTRO EN EL ORIGEN

$$B = \{(x, y) \mid x^2 + y^2 = 16\}$$

- * CONJUNTO DE PUNTOS DEL PLANO CARTESIANO

$$C = \{(x, y) \mid x \in \mathbb{R}, y \in \mathbb{R}\}$$

CONJUNTO ES UNA COLECCION DE OBJETOS BIEN DEFINIDOS

LOS CONJUNTOS COMUNMENTE SE DENOTAN CON LETRAS MAYUSCULAS.

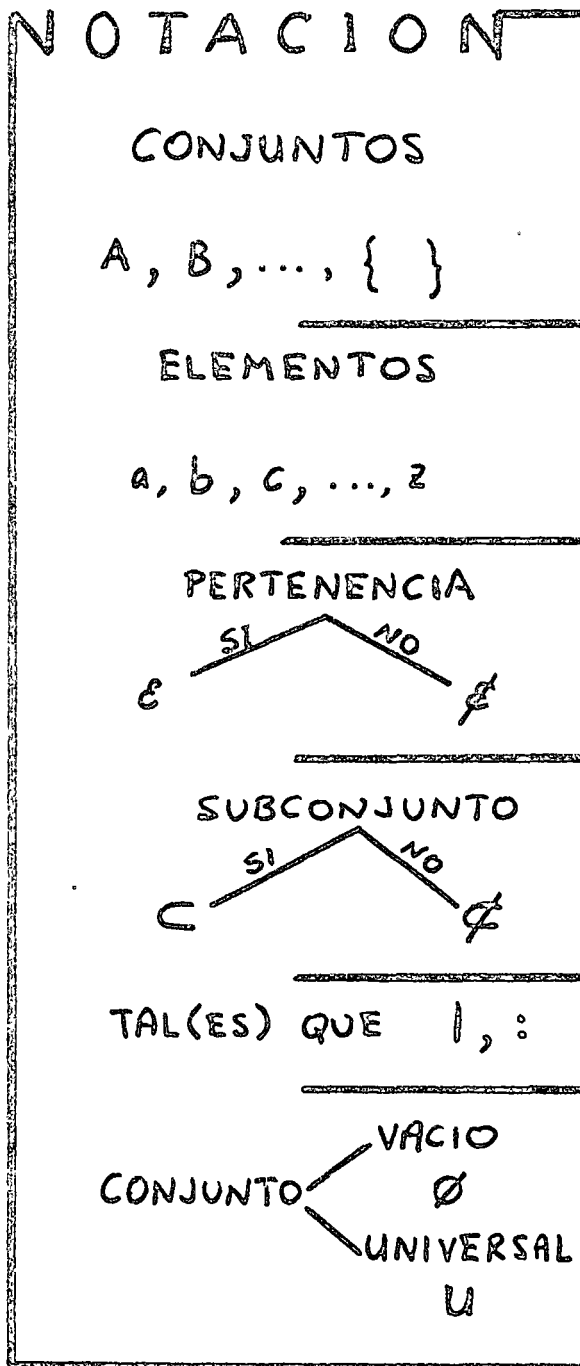
LOS OBJETOS QUE CONSTITUYEN A LOS CONJUNTOS SE LES CONOCE CON EL NOMBRE DE ELEMENTOS.

LOS ELEMENTOS SE DENOTAN COMUNMENTE CON LETRAS MINUSCULAS.

CUANDO LOS ELEMENTOS ESTAN EN UN CONJUNTO SE DICE QUE PERTENECEN A DICHO CONJUNTO.

UN CONJUNTO A SE DICE QUE ES SUBCONJUNTO DE OTRO B, SI TODOS LOS ELEMENTOS DE A, PERTENECEN TAMBIEN A B. SE DENOTA POR: $A \subset B$. SI EXISTE ALGUN ELEMENTO DE A, QUE NO PERTENEZCA A B, ENTONCES A NO ES SUBCONJUNTO DE B. SE DENOTA POR

$$A \not\subset B$$



(4)

CONJUNTO
VACIO, \emptyset

ES AQUEL QUE NO CONTIENE
ELEMENTOS. SE LE CONSIDERA
SUBCONJUNTO DE TODO CONJUNTO.

TODOS LOS CONJUNTOS QUE SE
MANEJEN EN UN DETERMINADO
ESTUDIO SON SUBCONJUNTOS DE
OTRO LLAMADO

CONJUNTO
UNIVERSAL

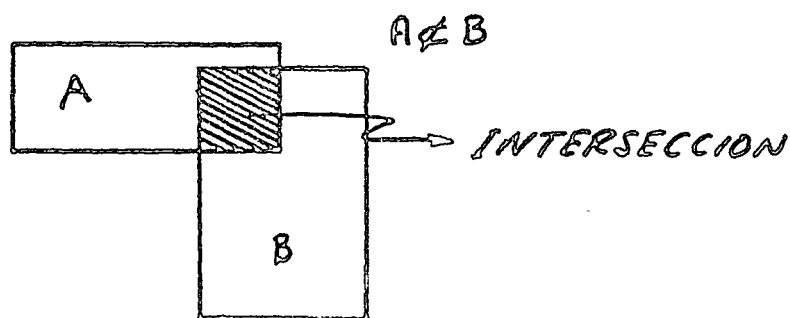
U

TEOREMA. 1. SEAN DOS CONJUNTOS A Y B.
SI $A \subset B$ Y $B \subset A$, ENTON-
CES $A = B$

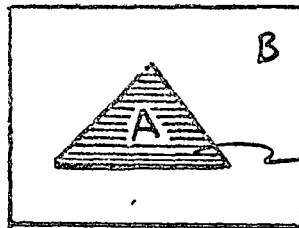
TEOREMA. 2. SEAN LOS CONJUNTOS A, B Y C.
SI $A \subset B$ Y $B \subset C$ ENTON-
CES $A \subset C$.

B. OPERACIONES ENTRE CONJUNTOS. LEYES.

SEAN DOS CONJUNTOS A Y B. LA
INTERSECCION ENTRE ESTOS CONJUNTOS
ES

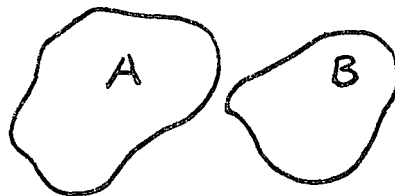


(5)



$A \subset B$

INTERSECCION



A Y B DISJUNTOS
O AJENOS

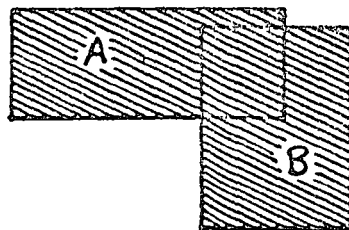
OTRO CONJUNTO FORMADO POR LOS ELEMENTOS QUE ESTAN EN A Y TAMBIEN EN B. SE DENOTA POR: $A \cap B$.

ASI:

$$A \cap B = \{x \mid x \in A, x \in B\}$$

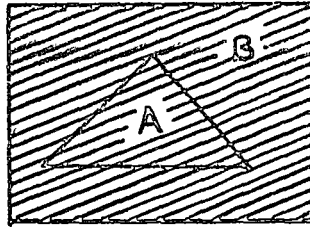
A LAS FIGURAS EMPLEADAS PARA LA REPRESENTACION DE LOS CONJUNTOS SE LES CONOCE COMO DIAGRAMAS DE VENN-EULER.

SEAN DOS CONJUNTOS A Y B. LA UNION ENTRE ESTOS CONJUNTOS ES

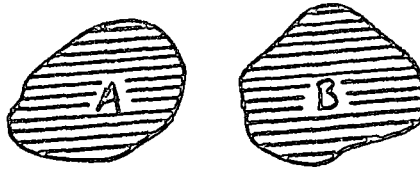


$A \cup B$

(6)



$A \subset B$



A Y B DISJUNTOS

OTRO CONJUNTO FORMADO POR LOS ELEMENTOS QUE ESTAN EN A, QUE ESTAN EN B, O EN AMBOS. SE DENOTA POR: $A \cup B$
ASI:

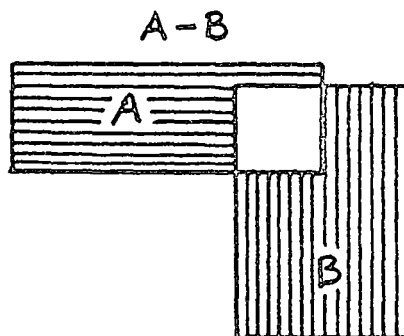
$$A \cup B = \{x \mid x \in A, x \in B, x \in (A \cap B)\}$$

O

$$A \cup B = \{x \mid x \in A \text{ O } x \in B\}$$

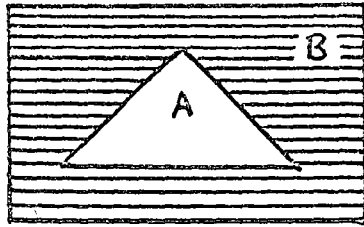
LA DIFERENCIA ENTRE LOS CONJUNTOS
A Y B ES

$A \not\subset B$



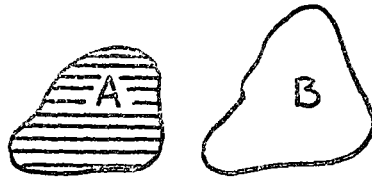
$B - A$

(7)



$A \subset B$

$B - A$



A Y B DISJUNTOS

$A - B$

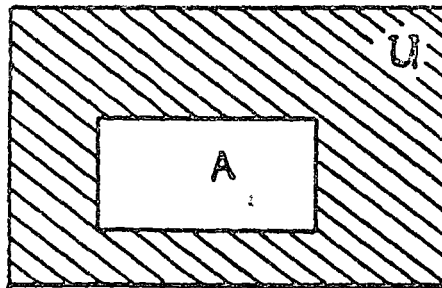
EL CONJUNTO FORMADO POR LOS ELEMENTOS QUE ESTAN EN EL PRIMER CONJUNTO PERO NO EN EL SEGUNDO.
ASI:

$$A - B = \{x \mid x \in A, x \notin B\}$$

$$B - A = \{x \mid x \in B, x \notin A\}$$

EL COMPLEMENTO DE UN CONJUNTO A

ES



(8)

EL CONJUNTO FORMADO POR LOS ELEMENTOS QUE ESTAN EN EL CONJUNTO UNIVERSAL Y QUE NO ESTAN EN A. SE DENOTA POR: A^c ó A' . ASÍ:

$$A^c = \{x \mid x \notin A\}$$

OBSERVESE QUE $A' = U - A$

LEYES DEL ALGEBRA DE CONJUNTOS

a)

$$A \cup A = A \quad , \quad A \cap A = A$$

b) ASOCIATIVAS

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

c) CONMUTATIVAS

$$A \cup B = B \cup A \quad , \quad A \cap B = B \cap A$$

d) DISTRIBUTIVAS

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

(9)

e) IDENTIDAD

$$\begin{aligned} A \cup \emptyset &= A & , & & A \cap U &= A \\ A \cup U &= U & , & & A \cap \emptyset &= \emptyset \end{aligned}$$

f) COMPLEMENTO

$$\begin{aligned} A \cup A^c &= U & , & & A \cap A^c &= \emptyset \\ (A^c)^c &= A & , & & U^c &= \emptyset & , & & \emptyset^c &= U \end{aligned}$$

g) DE MORGAN

$$\begin{aligned} (A \cup B)^c &= A^c \cap B^c \\ (A \cap B)^c &= A^c \cup B^c \end{aligned}$$

C. DIAGRAMAS DE VENN-EULER

VER PAGINA 5

D. CONJUNTOS FINITOS Y NUMERABLES

<p>CONJUNTO FINITO</p>

ES AQUEL QUE ES VACIO O CONSISTE DE EXACTAMENTE n ELEMENTOS, DONDE n ES UN ENTERO POSITIVO.

EN CASO CONTRARIO, EL CONJUNTO ES INFINITO.

<p>CONJUNTO NUMERABLE</p>

ES AQUEL QUE ES FINITO O CUYOS ELEMENTOS PUEDEN SER ARREGLADOS EN FORMA DE SECUENCIA. EN ESTE CASO SE DICE QUE ES INFINITO NUMERABLE

E. CONJUNTOS PRODUCTO

SEAN A Y B DOS CONJUNTOS. EL CONJUNTO PRODUCTO DE A Y B , DENOTADO POR $A \times B$, ESTA FORMADO POR TODAS LAS PAREJAS ORDENADAS (a, b) , DONDE $a \in A$ Y $b \in B$; ES DECIR,

$$A \times B = \{(a, b) \mid a \in A, b \in B\}$$

POR EJEMPLO, SI $A = \{1, 2\}$ Y $B = \{1, 2, 3\}$,

$$A \times B = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3)\}$$

(9-B)

F. PARTICION DE UN CONJUNTO

SEA UN CONJUNTO A. UNA PARTICION DE A ES UNA SUBDIVISION DE A EN SUB-CONJUNTOS NO VACIOS, LOS CUALES SON DISJUNTOS Y SU UNION ES IGUAL AL CONJUNTO A.

POR EJEMPLO, EL CONJUNTO $A = \{1, 2, \dots, 8, 9\}$ SE PUEDE PARTIR EN

$$\{1, 3, 5\}, \{2, 4, 6, 8\}, \{7, 9\}$$

G. PROBLEMAS

1. SEAN $U = \{ \text{LETRAS DEL ALFABETO} \}$

$$A = \{ a, b, c, d, e, f \}$$

COMPLETE CADA UNA DE LAS SIGUIENTES PROPOSICIONES:

1. a) $\{ a \} \underline{\quad} A$

1. b) $\{ a, b, c, d, e, f \} \underline{\quad} A$

1. c) $a \underline{\quad} A$

1. d) $g \underline{\quad} A$

1. e) $\{ h \} \underline{\quad} A$

1. f) $a \underline{\quad} \emptyset$

1. g) $\emptyset \underline{\quad} \{ b \}$

1. h) $U = A^c \underline{\quad} A$

1. i) $A^c = U \cap \underline{\quad}$

1. j) $A = U - \underline{\quad}$

(11)

2. SI $A \subset B$, ENTONCES :

2.a) $A \cap B =$

2.b) $A \cup B =$

2.c) $B^c \subset A^c =$

2.d) $A \cap B^c =$

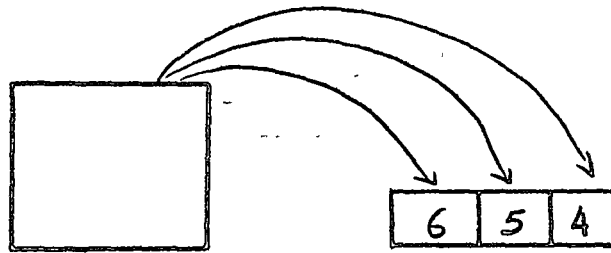
2.e) $B \cup A^c =$

2. TECNICAS DE CONTAR

EL PRINCIPIO FUNDAMENTAL DE CONTAR

DICE

SI SE TIENE UNA URNA CON 6 ESFERAS NUMERADAS DEL 1 AL 6, SE DESEA CONOCER EL NUMERO TOTAL DE CANTIDADES QUE SE PUEDEN FORMAR CON 3 DIGITOS DIFERENTES.



$$n = 6 \cdot 5 \cdot 4 = 120 \text{ CANTIDADES DIFERENTES}$$

"SI ALGUN EXPERIMENTO PUEDE SER REALIZADO EN N_1 FORMAS DIFERENTES, Y SI, SIGUIENDO EL MISMO PROCESO, UN SEGUNDO EXPERIMENTO PUEDE SER REALIZADO EN N_2 FORMAS DIFERENTES, Y ASI SUCESIVAMENTE; ENTONCES EL NUMERO DE FORMAS EN QUE EL EXPERIMENTO PUEDE SER REALIZADO EN EL ORDEN INDICADO, ES EL PRODUCTO $N_1 \cdot N_2 \cdot \dots$ " [3]

A. PERMUTACIONES

* SEA EL CONJUNTO $\{a, b, c, d\}$
 SUPONGASE QUE INTERESA EL ORDEN
 EN QUE APARECEN LAS LETRAS. SE
 PUEDEN FORMAR LOS SIGUIENTES
 ARREGLOS

- CON LAS 4 LETRAS:

$abcd, abdc, adbc, adcb, \dots$

- CON TRES LETRAS:

$abc, abd, acd, bcd, \dots$

TODOS LOS ARREGLOS ANTERIORES
 SON PERMUTACIONES.

CUALQUIER ARREGLO DE n OBJETOS EN
 UN ORDEN DADO. UN ARREGLO DE r
 OBJETOS, TAL QUE $r \leq n$, ES UNA
 PERMUTACION DE n OBJETOS TOMADOS
 r A LA VEZ. SE DENOTA POR:

$$P(n, r) \text{ ó } {}_n P_r$$

TEOREMA 3. EL NUMERO DE PERMUTACIONES DE n OBJETOS TOMANDO r A LA VEZ, ESTA DADO POR

$$P(n, r) = \frac{n!}{(n-r)!}$$

(14)

EJEMPLO 1. ; CUANTAS PERMUTACIONES HAY
DE 4 OBJETOS TOMADOS 3 A
LA VEZ?

$$P(4,3) = {}_4P_3 = \frac{4!}{(4-3)!} = 24$$

B. UNA COMBINACION ES

* SEA EL CONJUNTO $\{a, b, c, d\}$ Y
SUPONGASE QUE NO INTERESA EL
ORDEN EN QUE APARECEN LAS LETRAS.
LOS SIGUIENTES ARREGLOS

- abcd

- abc, abd, acd, bcd

SON COMBINACIONES TOMANDO 4
Y 3 OBJETOS A LA VEZ, RESPEC-
TIVAMENTE.

CUALQUIER SUBCONJUNTO DE OBJETOS.
SI SE TIENEN n OBJETOS, CUALQUIER
SUBCONJUNTO DE r OBJETOS, TAL QUE
 $r \leq n$, ES UNA COMBINACION DE n
TOMANDO r A LA VEZ.

UNA COMBINACION SE DENOTARA POR:

$$C(n, r) \text{ ó } \binom{n}{r}$$

TEOREMA. 4. EL NUMERO DE COMBINACIONES ESTA DADO POR

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{r!(n-r)!}$$

C. MUESTRAS ORDENADAS

SUPONGASE QUE SE TIENE UNA URNA CONTENIENDO n BOLAS NUMERADAS DE LA 1 A LA n . SI SE EXTRAEN BOLAS DE LA URNA, UNA A LA VEZ, HASTA COMPLETAR r , SE DICE QUE SE HA OBTENIDO UNA MUESTRA DE TAMAÑO r . POR SUPUESTO, SE TIENEN DOS CASOS:

- a) CON REEMPLAZO. EN ESTE CASO, LA BOLA SE COLOCA NUEVAMENTE DENTRO DE LA URNA, ANTES DE EXTRAER LA SIGUIENTE BOLA. DE ACUERDO CON EL PRINCIPIO FUNDAMENTAL DE CONTAR, SE TIENEN

$$\underbrace{n \cdot n \cdot \dots \cdot n}_{r \text{ VECES}} = n^r$$

DIFERENTES MUESTRAS DE TAMAÑO r .

- b) SIN REEMPLAZO. EN ESTE CASO, LA BOLA NO SE RESTITUYE Y EN CONSECUENCIA, NO HAY REPETICIONES EN LA MUESTRA. ASI, SE TIENEN

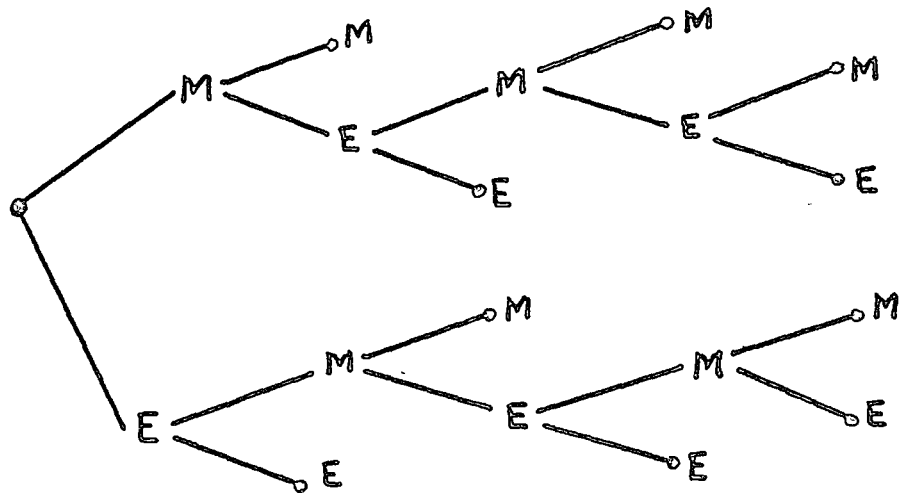
$$n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!} = P(n,r)$$

DIFERENTES MUESTRAS DE TAMAÑO r .

D. DIAGRAMAS DE ARBOL

UN DIAGRAMA DE ARBOL ES

* MARIO Y ERNESTO VAN A PARTICIPAR EN UN TORNEO DE TENIS. LA PRIMERA PERSONA DOS JUEGOS SEGUIDOS O TRES EN TOTAL, GANA EL TORNEO. EL NUMERO DE POSIBLES RESULTADOS SE PUEDE OBSERVAR



QUE ES IGUAL A 10

"UN DISPOSITIVO PARA ENUMERAR TODOS LOS POSIBLES RESULTADOS DE UNA SECUENCIA DE EXPERIMENTOS DONDE CADA EXPERIMENTO OCURRE EN UN NUMERO FINITO DE FORMAS."

PROBLEMAS [3]

1. CONSIDERENSE LAS CIUDADES A, B Y C. EXISTEN 6 CARRETERAS ENTRE A Y B Y 4 ENTRE B Y C.
 - a); DE CUANTAS FORMAS SE PUEDE MANEJAR DE A HACIA C, PASANDO POR B?
 - b); DE CUANTAS MANERAS SE PUEDE EFECTUAR UN VIAJE REDONDO ENTRE A Y C, SIN UTILIZAR LA MISMA CARRETERA MAS DE UNA VEZ?

2. UNA CLASE CONTIENE 9 HOMBRES Y 3 MUJERES. (a); EN CUANTAS FORMAS SE PUEDE ESCOGER UNA COMISION DE 4 PERSONAS? (b); CUANTAS COMISIONES TENDRIAN AL MENOS UNA MUJER? (c); CUANTAS TENDRIAN EXACTAMENTE UNA MUJER?

3. UN HOMBRE SE ENCUENTRA EN EL ORIGEN SOBRE EL EJE X Y SE MUEVE HACIA LA IZQUIERDA O LA DERECHA, CON PASOS UNITARIOS. EL PARA DESPUES DE 5 PASOS O SI LOGRA UN 3 O UN -2. CONSTRUIR EL DIAGRAMA DE ARBOL PARA DESCRIBIR TODAS LAS POSIBLES TRAYECTORIAS QUE PUEDE RECORRER EL HOMBRE.

II. PROBABILIDAD

OBJETIVO

ESTABLECER LOS DIFERENTES CONCEPTOS DE PROBABILIDAD Y LA IMPORTANCIA DE LA DEFINICION AXIOMATICA, ASI COMO LOGRAR COMPRENDER LOS CONCEPTOS DE PROBABILIDAD CONDICIONAL E INDEPENDENCIA.

1. EXPERIMENTO. ESPACIO MUESTRAL. EVENTOS.

SE CONSIDERARA QUE UN EXPERIMENTO ES

* SI SE LANZA UN DADO, SE PUEDE OBSERVAR EL NUMERO QUE APARECE EN LA CARA SUPERIOR. EN ESTE CASO, SON 6 POSIBLES NUMEROS:

$$S = \{1, 2, 3, 4, 5, 6\}$$

* SI SE LANZA UNA MONEDA 3 VECES, SE PUEDE OBSERVAR LA SECUENCIA DE AGUILAS Y SOLES QUE APARECE. POR EJEMPLO:

$$S = \{AAA, AAS, ASA, ASS, SSS, SSA, SAS, SAA\}$$

CUALQUIER PROCESO DE ENSAYO Y OBSERVACION

ESPACIO
MUESTRAL

ES EL CONJUNTO S DE TODOS
LOS POSIBLES RESULTADOS DE UN
EXPERIMENTO.

UN SUBCONJUNTO DEL ESPACIO MUESTRAL
SE LLAMARA _____

EVENTO

EVENTO
ELEMENTAL

ES AQUEL SUBCONJUNTO DE S
CONSTITUIDO POR UN SOLO ELEMENTO.

EL CONJUNTO VACIO, \emptyset , ES EL

EVENTO IMPOSIBLE

EVENTO SEGURO

ES EL PROPIO ESPACIO MUESTRAL, S .

SI DOS EVENTOS SON DISJUNTOS,
ES DECIR, SI $A \cap B = \emptyset$, ENTON-
CES SE DICE QUE SON _____

EVENTOS
MUTUAMENTE
EXCLUSIVOS

EL QUE DOS EVENTOS SEAN MUTUAMENTE EXCLUSIVOS
IMPLICA QUE NO PUEDEN OCURRIR SIMULTANEAMENTE.

EJEMPLO 2. CONSIDERECE EL EXPERIMENTO DE
LANZAR DOS DADOS. EL ESPACIO
MUESTRAL ESTARA FORMADO POR

$$S = \{(1,1), (1,2), \dots, (6,5), (6,6)\}$$

EL EVENTO FORMADO POR AQUE-
LLAS PAREJAS CUYA SUMA SEA 10

(21)

ES

$$A = \{(4,6), (5,5), (6,4)\}$$

EL EVENTO FORMADO POR LAS PA-
REJAS CUYO PRIMER ELEMENTO SEA 3
ES

$$B = \{(1,2), (1,1), (1,3), (1,4), (1,5), (1,6)\}$$

EL EVENTO FORMADO POR LAS PARE-
JAS CUYA SUMA SEA 10 Y EL PRIMER
ELEMENTO SEA 3 ES

$$C = A \cap B = \emptyset$$

ES DECIR, EL EVENTO IMPOSIBLE

2. DIFERENTES CONCEPTOS DE PROBABILIDAD

A. DEFINICION CLASICA. SI UN EVENTO PUEDE
OCURRIR EN n
FORMAS MUTUAMENTE EXCLUSIVAS E IGUALMEN-
TE PROBABLES, Y SI n_A DE ESOS RESULTA-
DOS TIENEN UNA CARACTERISTICA, A , ENTON-
CES LA PROBABILIDAD DE A , $P(A)$, ES
LA FRACCION n_A/n .

PROBABILIDAD
A PRIORI

ES LA PROBABILIDAD DETERMINA-
DA POR LA DEFINICION CLASICA.

CALCULAR LAS PROBABILIDADES A PARTIR DE LA DEFINICION CLASICA IMPLICA VARIOS PROBLEMAS:

1^º NO ES POSIBLE TENER SIEMPRE, EXPERIMENTOS CUYOS EVENTOS TENGAN LA MISMA POSIBILIDAD DE OCURRIR.

EJEMPLO : EL LANZAMIENTO DE UN DADO CARGADO.

2^º EN MUCHOS PROBLEMAS, EL NUMERO DE RESULTADOS ES INFINITO. ESTO NO ESTA DE ACUERDO CON LA DEFINICION QUE SUPONE SOLO n RESULTADOS.

3^º ; COMO SE PODRIAN CONTESTAR LAS SIGUIENTES PREGUNTAS ?

a) ; CUAL ES LA PROBABILIDAD DE QUE EN VERACRUZ NACERA UN NIÑO ?

b) ; CUAL ES LA PROBABILIDAD DE QUE UNA MUJER MUERA ANTES DE LOS 35 AÑOS ?

ETC.

ESTAS DIFICULTADES DIERON LUGAR A OTRA DEFINICION DE PROBABILIDAD QUE SE LE CONOCE COMO PROBABILIDAD DE FRECUENCIA.

B. DEFINICION DE FRECUENCIA RELATIVA [1]

" CONSIDERESE QUE UN EXPERIMENTO SE REPITE n VECES. SI UN EVENTO A OCURRE n_A VECES, ENTONCES SU PROBABILIDAD SE DEFINE COMO

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}$$

PROBABILIDAD
A POSTERIORI

ES LA PROBABILIDAD DETERMINADA CON LA DEFINICION DE FRECUENCIA.

ESTA DEFINICION DE PROBABILIDAD ES FUNDAMENTAL EN LA APLICACION DE LA TEORIA DE PROBABILIDAD AL MUNDO FISICO, PUES ES DE GRAN VALOR EN LA DETERMINACION DE DATOS PROBABILISTICOS Y DA UN CONTENIDO FISICO AL CALCULO DE PROBABILIDADES. NO OBTANTE, TIENE UNA FUERTE LIMITACION PARA SER UTILIZADA COMO BASE DE UNA TEORIA DEDUCTIVA. ESTA AFIRMACION SE BASA EN EL HECHO DE QUE SE SUPONE UNA IGUALDAD ENTRE LA FRACCION n_A/n Y LA PROBABILIDAD; ES DECIR, SE SUPONE

$$P(A) = \frac{n_A}{n}$$

EL NUMERO DE RESULTADOS n PUEDE SER MUY GRANDE, PERO SIEMPRE ES FINITO.

C. DEFINICION AXIOMATICA

LAS DEFINICIONES DE PROBABILIDAD ANTERIORMENTE TRATADAS, AUN CUANDO TIENEN CIERTAS OBJECIONES, SON DE GRAN AYUDA PARA RESPONDER A IMPORTANTES PROBLEMAS EN LA CIENCIA EXPERIMENTAL.

LA DEFINICION AXIOMATICA, QUE ELIMINA TODOS LOS PROBLEMAS DE LAS DEFINICIONES ANTERIORES, PERMITE EL DESARROLLO DE LA TEORIA MODERNA DE PROBABILIDAD EN BASE, UNICAMENTE, EN LOS AXIOMAS SIGUIENTES:

SEA S UN ESPACIO MUESTRAL, Y SEA A CUALQUIER EVENTO EN S . ENTONCES P SERA LLAMADA 'FUNCION DE PROBABILIDAD' DEFINIDA EN EL ESPACIO MUESTRAL S , SI SATISFACE

1. $P(A)$ ES UN NUMERO REAL TAL QUE

$$P(A) \geq 0 \text{ PARA TODO } A \subset S$$

2. $P(S) = 1$

3. SI S_1, S_2, \dots ES UNA SECUENCIA DE EVENTOS MUTUAMENTE EXCLUSIVOS EN S ; ES DECIR, SI $S_i \cap S_j = \emptyset$ PARA $i \neq j = 1, 2, \dots$, ENTONCES

$$P(S_1 \cup S_2 \cup \dots) = P(S_1) + P(S_2) + \dots$$

TEOREMA. 5. SI \emptyset ES EL CONJUNTO VACIO,
ENTONCES

$$P(\emptyset) = 0$$

TEOREMA. 6. SI A^c ES EL COMPLEMENTO DE UN
EVENTO A , ENTONCES

$$P(A^c) = 1 - P(A)$$

TEOREMA. 7. SI A ES CUALQUIER EVENTO EN
 S , ENTONCES

$$0 \leq P(A) \leq 1$$

TEOREMA. 8. SEAN DOS EVENTOS A Y B DE S ,
SI $A \subset B$, ENTONCES

$$P(A) \leq P(B)$$

TEOREMA. 9. SEAN A Y B DOS EVENTOS CUAL-
QUIERA DE S , ENTONCES

$$P(A - B) = P(A) - P(A \cap B)$$

TEOREMA. 10. SEAN A Y B DOS EVENTOS CUAL-
QUIERA DE S , ENTONCES

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

3. ESPACIOS DE PROBABILIDAD

SEA $S = \{a_1, a_2, \dots, a_n\}$ UN ESPACIO MUESTRAL FINITO. UN ESPACIO FINITO DE PROBABILIDAD

CONSIDERECE EL EXPERIMENTO DE LANZAR 3 MONEDAS (SIN IRREGULARIDADES) Y OBSERVAR EL NUMERO DE AGUILAS QUE APARECEN. EN ESTE CASO, EL ESPACIO MUESTRAL ES

$$S = \{AAA, AAS, ASA, ASS, SSS, SSA, SAS, SAA\}$$

COMO CADA RESULTADO SE CONSIDERA QUE TIENE LA MISMA POSIBILIDAD DE OCURRIR, ENTONCES LA PROBABILIDAD DE CADA RESULTADO ES $1/8$. POR LO TANTO

$$P(0) = \text{PROBABILIDAD DE OBTENER } \emptyset \text{ AGUILAS} \\ = \frac{1}{8}$$

$$P(1) = \text{ " " " 1 AGUILA} \\ = \frac{3}{8}$$

$$P(2) = \text{ " " " 2 AGUILAS} \\ = \frac{3}{8}$$

$$P(3) = \text{ " " " 3 AGUILAS} \\ = \frac{1}{8}$$

DE ESTA FORMA, EL ESPACIO MUESTRAL PUEDE CONVERTIRSE EN

$$S = \{0, 1, 2, 3\}$$

(27)

EL CONJUNTO DE PROBABILIDADES $P(0)$, $P(1)$, $P(2)$ Y $P(3)$, ASOCIADAS A LOS RESULTADOS 0, 1, 2 Y 3, CONSTITUYE UN ESPACIO FINITO DE PROBABILIDAD

SE OBTIENE ASIGNANDO A CADA PUNTO $a_i \in S$ UN NUMERO REAL p_i LLAMADO PROBABILIDAD DE a_i Y TAL QUE SATISFAGA LAS SIGUIENTES PROPIEDADES:

$$(1) \text{ CADA } p_i \geq 0$$

$$(2) \sum_{i=1}^n p_i = 1$$

LA PROBABILIDAD $P(A)$ DE CUALQUIER EVENTO A , SE DEFININE COMO LA SUMA DE LAS PROBABILIDADES DE LOS PUNTOS DE A . ASI, EN EL EJEMPLO ANTERIOR, LA PROBABILIDAD DE OBTENER CUANDO MAS 2 AGUILAS ES $7/8$, YA QUE:

$$A = \{0, 1, 2\}$$

$$P(A) = P(0) + P(1) + P(2) = \frac{7}{8}$$

ESPACIO
IGUALMENTE
PROBABLE

ES TODO ESPACIO FINITO DE PROBABILIDAD QUE SATISFACE:

$$(1) P(a_1) = P(a_2) = \dots = P(a_n) = 1/n$$

(2) SEA A UN EVENTO QUE CONTIENE n_A PUNTOS DE S . ENTONCES

$$P(A) = \frac{n_A}{n}$$

4. PROBABILIDAD CONDICIONAL. INDEPENDENCIA.

SEA B UN EVENTO ARBITRARIO DE S ,
TAL QUE $P(B) > 0$. LA PROBABILIDAD CONDI-
CIONAL SE DEFINE COMO

CONSIDERESE EL LANZAMIENTO DE DOS
DADOS (SIN DEFECTOS). SI LA SUMA ES 6,
LA PROBABILIDAD DE QUE UNO DE LOS DADOS
SEA 2, ES IGUAL A $2/5$. EN EFECTO
EL ESPACIO MUESTRAL S
ESTA FORMADO POR 36
PUNTOS (RESULTADOS). ASI

$$S = \{(1,1), (1,2), \dots, (6,5), (6,6)\}$$

DE ESTOS RESULTADOS, LOS
QUE SUMAN 6 SON

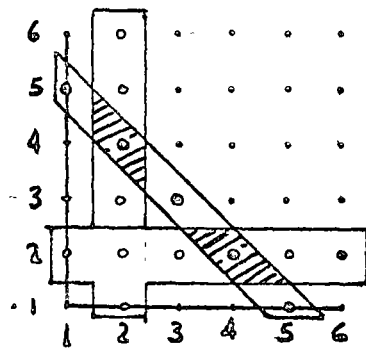
$$B = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$$

Y LOS QUE TIENEN AL MENOS UN 2 SON

$$A = \{(1,2), (2,2), (3,2), (4,2), (5,2), (6,2), \\ (2,1), (2,3), (2,4), (2,5), (2,6)\}$$

INTERESAN AQUELLOS PUNTOS QUE SUMEN 6
Y QUE ADEMAS TENGAN UN 2, ES DECIR

$$A \cap B = \{(2,4), (4,2)\}$$

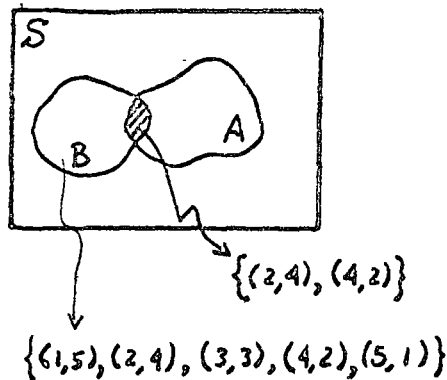


(29)

LA PROBABILIDAD DE $A \cap B$ ES

$$P(A \cap B) = \frac{2}{36}$$

APARENTEMENTE, ESTA ES LA PROBABILIDAD BUSCADA. SIN EMBARGO, EL EVENTO $A \cap B$ ESTA SUJETO A QUE B HAYA OCURRIDO (LA CONDICION ES QUE LA SUMA SEA 6).



EN OTRAS PALABRAS, SI B OCURRE (5 RESULTADOS), SOLO HAY 2 POSIBILIDADES, DE S , PARA QUE $A \cap B$ OCURRA; ES DECIR, LA PROBABILIDAD ES $\frac{2}{5}$

ESTE MISMO RESULTADO SE OBTIENE SI SE DIVIDE LA PROBABILIDAD DE $A \cap B$ ENTRE LA DEL EVENTO B ($P(B) = 5/36$):

$$\frac{P(A \cap B)}{P(B)} = \frac{2/36}{5/36}$$

LA PROBABILIDAD DE QUE UN EVENTO A OCURRA DADO QUE EL EVENTO B HA OCURRIDO. SE DENOTA POR $P(A|B)$ Y SE EXPRESA COMO LA RAZON DE PROBABILIDADES DE $A \cap B$ Y DE B ; ES DECIR

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

TEOREMA DE MULTIPLICACION

DE ACUERDO A LA LEY CONMUTATIVA DE INTERSECCION DE CONJUNTOS, $A \cap B = B \cap A$.
POR LO TANTO

$$P(A \cap B) = P(B \cap A)$$

UTILIZANDO LA EXPRESION PARA PROBABILIDAD CONDICIONAL, SE OBTIENE

$$\text{TEOREMA. II. } P(B \cap A) = P(B)P(A|B)$$

COROLARIO. 1. PARA CUALESQUIERA EVENTOS
 A_1, A_2, \dots, A_n , DE S ,

$$P(A_1 \cap A_2 \cap \dots \cap A_n) =$$

$$= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

EJEMPLO 3. UN LOTE CONTIENE 12 RADIOS DE LOS CUALES 4 ESTAN DEFECTUOSOS. TRES RADIOS SON SACADOS AL AZAR*, UNO POR UNO. LA PROBABILIDAD DE QUE LOS TRES SEAN NO DEFECTUOSOS ES $14/55$. EN EFECTO, SEAN LOS EVENTOS

$$A_1 = \{\text{EL PRIMER RADIO ES NO DEFECTUOSO}\}$$

$$A_2 = \{\text{EL SEGUNDO RADIO ES NO DEFECTUOSO}\}$$

$$A_3 = \{\text{EL TERCER RADIO ES NO DEFECTUOSO}\}$$

* LA EXPRESION "AL AZAR" SE UTILIZARA SOLAMENTE CON ESPACIOS IGUALMENTE PROBABLES.

(31)

EL EVENTO DESEADO ES $A_1 \cap A_2 \cap A_3$.
LA PRIMERA VEZ, HAY 8 POSIBILIDADES DE 12, DE SACAR UN RADIO NO DEFECTUOSO. POR LO TANTO

$$P(A_1) = \frac{8}{12}$$

LA SEGUNDA VEZ, DADO QUE UN RADIO NO DEFECTUOSO SE HA SACADO, HAY 7 POSIBILIDADES DE 11, DE SACAR OTRO RADIO NO DEFECTUOSO. POR TANTO

$$P(A_2 | A_1) = \frac{7}{11}$$

LA TERCERA VEZ, DADO QUE SE HAN SACADO DOS RADIOS NO DEFECTUOSOS, A_1 Y A_2 ($A_1 \cap A_2$), HAY 6 POSIBILIDADES DE 10, DE SACAR UN TERCER RADIO NO DEFECTUOSO. POR TANTO

$$P(A_3 | A_1 \cap A_2) = \frac{6}{10}$$

FINALMENTE, DE ACUERDO AL TEOREMA DE MULTIPLICACION

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \\ &= \frac{14}{55} \end{aligned}$$

SI DOS EVENTOS NO ESTAN RELACIONADOS EN ALGUNA FORMA, LA PROBABILIDAD DE QUE OCURRA UNO NO CAMBIA SI SE SABE QUE HA OCURRIDO EL OTRO. EN OTRAS PALABRAS, SE DICE QUE ESTOS EVENTOS SON INDEPENDIENTES.

(32)

DOS EVENTOS A Y B SON INDEPENDIENTES
SI Y SOLO SI

$$P(A|B) = P(A)$$

DE LA DEFINICION ANTERIOR SE TIENE

a) $P(B|A) = P(B)$

b) $P(A \cap B) = P(A)P(B)$

EN GENERAL, LOS EVENTOS A, B, ..., N SON
MUTUAMENTE INDEPENDIENTES SI Y SOLO SI

$$P(A \cap B \cap \dots \cap N) = P(A)P(B) \dots P(N)$$

5. PROBLEMAS

1.[4]. UN INGENIERO ESTA DISEÑANDO UN GRAN DEPOSITO PARA AGUA QUE LLEGA DE DOS AREAS SEPARADAS. LA CANTIDAD DE AGUA QUE LLEGA DE A PUEDE SER 0, 10, 20, 30 pies³/seg, Y LA QUE LLEGA DE B PUEDE SER 0, 20, 40, 60 pies³/seg. ENCUENTRE EL ESPACIO MUESTRAL PARA A Y B EN FORMA CONJUNTA, Y PARA A Y B EN FORMA SEPARADA. DEFINA LOS SIGUIENTES EVENTOS:

- a) $C = A > 10$, pies³/seg
- b) $D = B > 20$ "
- c) $E = C^c$
- d) $F = C \cap D$
- e) $G = (C \cup D)^c$

2. EN UN CIERTO PUEBLO, 40% DE LA GENTE TIENE PELO CAFE, 25% OJOS CAFES, Y 15% TIENE PELO Y OJOS CAFES. SE SELECCIONA AL AZAR UNA PERSONA DEL PUEBLO.

- a) SI TIENE PELO CAFE, ¿ CUAL ES LA PROBABILIDAD DE QUE TAMBIEN TENGA OJOS CAFES?
- b) SI TIENE OJOS CAFES, ¿ CUAL ES LA PROBABILIDAD DE QUE NO TENGA PELO CAFE?
- c) ¿ CUAL ES LA PROBABILIDAD DE QUE NO TENGA PELO CAFE NI OJOS CAFES?

3. EN UN COLEGIO, 25% DE LOS HOMBRES Y 10% DE LAS MUJERES ESTAN ESTUDIANDO MATEMATICAS. LAS MUJERES CONSTITUYEN EL 60% DE LOS ESTUDIANTES. SI SE SELECCIONA UN ESTUDIANTE AL AZAR Y ESTA ESTUDIANDO MATEMATICAS, DETERMINE LA PROBABILIDAD DE QUE EL ESTUDIANTE SEA MUJER.
4. CONSIDERESE EL LANZAMIENTO DE 3 MONEDAS (SIN DEFECTOS). SEAN:
- $$A = \{ \text{TODOS SON O AGUILAS, O SOLES} \}$$
- $$B = \{ \text{AL MENOS DOS AGUILAS} \}$$
- $$C = \{ \text{CUANDO MAS DOS AGUILAS} \}$$
- DE LAS PAREJAS (A, B) , (A, C) Y (B, C) ,
 ¿CUALES SON INDEPENDIENTES Y CUALES DEPENDIENTES?
5. LA PROBABILIDAD DE QUE 'A' PEGUE EN EL BLANCO ES $\frac{1}{4}$ Y LA PROBABILIDAD DE QUE 'B' PEGUE EN EL BLANCO ES $\frac{1}{3}$.
- a) SI CADA UNO DISPARA DOS VECES,
 ¿CUAL ES LA PROBABILIDAD DE QUE EL BLANCO SEA ALCANZADO AL MENOS UNA VEZ?
- b) SI CADA UNO DISPARA UNA VEZ Y EL BLANCO ES ALCANZADO UNA VEZ,
 ¿CUAL ES LA PROBABILIDAD DE QUE A LE HAYA PEGADO AL BLANCO?
- c) SI A PUEDE DISPARAR SOLO DOS VECES,
 ¿CUANTAS VECES DEBE DISPARAR B PARA QUE TENGAN, AL MENOS, UN 90% DE PROBABILIDAD DE PEGARLE AL BLANCO?

III. VARIABLES

ALEATORIAS

OBJETIVO

COMPRENDER LOS CONCEPTOS
RELATIVOS A LAS VARIABLES
ALEATORIAS PARA APLICARLOS
AL ESTUDIO DE LA GEOESTADIS-
TICA.

1. UNA VARIABLE ALEATORIA.

A. DEFINICION DE VARIABLE ALEATORIA

SEA UN ESPACIO MUESTRAL S CON UNA FUNCION DE PROBABILIDAD DEFINIDA EN EL. UNA VARIABLE ALEATORIA ES

CONSIDERESE EL LANZAMIENTO DE UNA MONEDA (SIN ALTERACIONES), DOS VECES. EL ESPACIO MUESTRAL ESTA FORMADO POR LOS PUNTOS: $s_1 = (A, A)$, $s_2 = (A, B)$, $s_3 = (B, B)$, $s_4 = (B, A)$. SEA X UNA FUNCION DEFINIDA EN S Y VALUADA EN LOS REALES, TAL QUE

$$X(s_i) = \text{NUMERO DE AGUILAS EN } s_i$$

DE ESTA FORMA

$$\begin{aligned} X(s_1) &= 2, & X(s_2) &= 1. \\ X(s_3) &= 0, & X(s_4) &= 1 \end{aligned}$$

PUESTO QUE ES UN ESPACIO IGUALMENTE PROBABLE

$$P(\{s \mid X(s) = 0\}) = P(\{s_3\}) = P(X=0) = \frac{1}{4}$$

$$P(\{s \mid X(s) = 1\}) = P(\{s_2, s_4\}) = P(X=1) = \frac{1}{2}$$

$$P(\{s \mid X(s) = 2\}) = P(\{s_1\}) = P(X=2) = \frac{1}{4}$$

LA FUNCION X ASI DEFINIDA ES UNA VARIABLE ALEATORIA.

UNA FUNCION X DEFINIDA EN S Y VALUADA EN LOS REALES.

B. VARIABLES ALEATORIAS Y DISTRIBUCIONES [4]

EL COMPORTAMIENTO DE UNA VARIABLE ALEATORIA SE DESCRIBE POR SU

LEY DE PROBABILIDAD

DISTRIBUCION DE PROBABILIDAD

DE UNA VARIABLE ALEATORIA ES LA FORMA MAS COMUN DE CARACTERIZAR SU LEY DE PROBABILIDAD.

1. VARIABLES ALEATORIAS DISCRETAS

CONSIDERESE EL LANZAMIENTO DE 4 MONEDAS SIMETRICAS PARA EL QUE SE REGISTRA EL NUMERO DE AGUILAS. SEA X LA VARIABLE ALEATORIA DEFINIDA COMO

$X(s_i) = \text{NUMERO DE AGUILAS EN } s_i$
DONDE $s_i \in S$. ASI

$$\{s \mid X(s) = 0\} = \{(L, L, L, L)\}$$

$$\{s \mid X(s) = 1\} = \{(A, L, L, L), (L, A, L, L), (L, L, A, L), (L, L, L, A)\}$$

(38)

$$\{s | X(s) = 2\} = \{(A, A, L, L), \dots, (L, L, A, A)\}$$

$$\{s | X(s) = 3\} = \{(A, A, A, L), \dots, (L, A, A, A)\}$$

$$\{s | X(s) = 4\} = \{(A, A, A, A)\}$$

SEGUN ESTO

$$P(X=0) = \frac{1}{16}, \quad P(X=2) = \frac{3}{8}$$

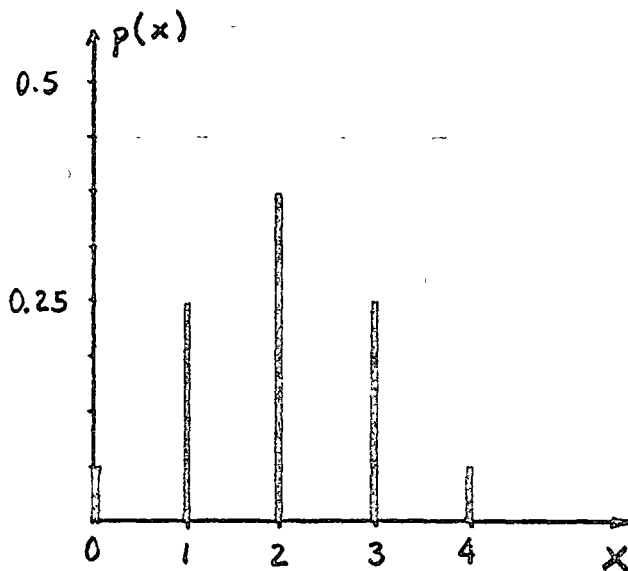
$$P(X=1) = \frac{1}{4}, \quad P(X=3) = \frac{1}{4}$$

$$P(X=4) = \frac{1}{16}$$

GRAFICANDO

EN GENERAL:

$$p(x) = \frac{\binom{4}{x}}{2^4}$$



SE PUEDE ESTAR INTERESADO EN OBTENER

$$P(X \leq x)$$

ES DECIR

$$P(\{s | X(s) \leq x\})$$

EN ESTE CASO:

$$P(\{s | X(s) < 0\}) = P(X < 0) = P(\emptyset) = 0$$

$$P(\{s | X(s) = 0\}) = P(X = 0) = \frac{1}{16}$$

$$P(\{s | X(s) < 1\}) = P(X < 1) = \frac{1}{16}$$

$$P(\{s | X(s) < 2\}) = P(X < 2) = \frac{5}{16}$$

$$P(\{s | X(s) < 3\}) = P(X < 3) = \frac{11}{16}$$

$$P(\{s | X(s) < 4\}) = P(X < 4) = \frac{15}{16}$$

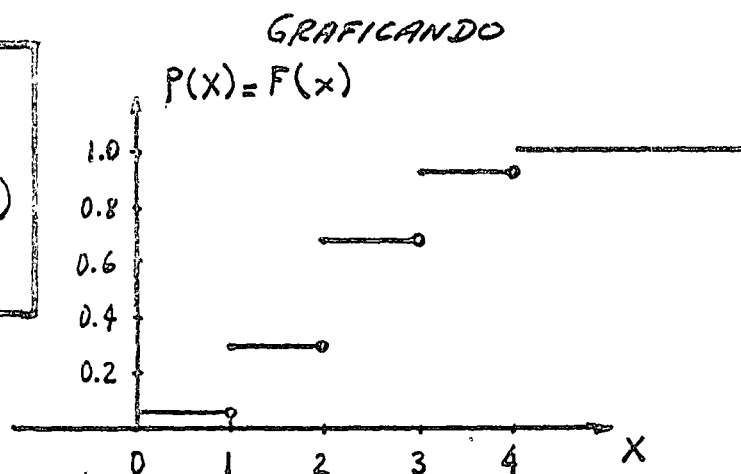
$$P(\{s | X(s) = 4\}) = P(X = 4) = \frac{1}{16}$$

$$P(\{s | X(s) < i\}) = P(X < i) = 1 \quad \text{PARA TODA } i > 4$$

EN GENERAL

$$F(x) =$$

$$= P(X < x) = \sum_{n=0}^{x-1} p(x)$$



UNA VARIABLE ALEATORIA DISCRETA ES AQUELLA QUE SOLO PUEDE TOMAR UN NUMERO FINITO O NUMERABLE DE VALORES. SU LEY DE PROBABILIDAD ESTA PRESENTADA, COMUNMENTE, EN LA FORMA DE UNA FUNCION DE MASA DE PRO-

PROBABILIDAD (FUNCION DE DENSIDAD). SE DENOTARA POR $p(x)$ Y DEFINIRA DE ACUERDO A LA EXPRESION

$$p(x) = P(X=x)$$

CUALQUIER FUNCION $p(x)$ QUE SATISFAGA

$$(a) \quad p(x_i) \geq 0, \quad i=1, 2, \dots$$

$$(b) \quad \sum p(x_i) = 1$$

ES UNA FUNCION DE MASA DE PROBABILIDAD.

OTRA FORMA DE DESCRIBIR LA DISTRIBUCION DE PROBABILIDAD ES A TRAVES DE LA FUNCION DE DISTRIBUCION ACUMULATIVA. SE DENOTARA POR $F(x)$ Y SE DEFINIRA POR

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \sum_{x_i \leq x} p(x_i) \end{aligned}$$

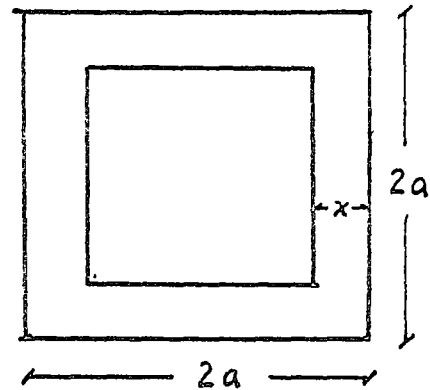
2. VARIABLES ALEATORIAS CONTINUAS

SUPONGASE QUE SE DESEA CONOCER LA DISTRIBUCION DE LA DISTANCIA X DE LA CARGA AL SOPORTE DE LA CARA MAS CERCA, DE UNA PLATAFORMA CUADRADA DE TAMAÑO $2a \times 2a$, SUJETA A CARGAS CONCENTRADAS, LOCALIZADAS ARBITRARIAMENTE; ES DECIR, AL AZAR. ESTO IMPLICA QUE LA PROBABILIDAD DE QUE LA CARGA CAIGA EN CUALQUIER REGION DEL PISO, ES PROPORCIONAL SOLAMENTE AL AREA DE LA REGION. [4]

(41)

DE ACUERDO A LO ANTERIOR

$$F(x) = P(X \leq x) \\ = \frac{4a^2 - (2a - 2x)^2}{4a^2}$$



$$F(x) = 1 - \frac{(a-x)^2}{a^2}; \quad 0 \leq x \leq a$$

LA CORRESPONDIENTE FUNCION DE DENSIDAD ES LA DERIVADA DE $F(x)$; ESTO ES,

$$f(x) = \frac{dF(x)}{dx} = \frac{2(a-x)}{a^2}, \quad 0 \leq x \leq a$$

UNA VARIABLE ALEATORIA CONTINUA, A DIFERENCIA DE LA DISCRETA, PUEDE TOMAR CUALQUIER VALOR SOBRE LOS REALES. TODAS LAS VARIABLES FISICAS DE INGENIERIA SON TRATADAS, COMUNMENTE, COMO CANTIDADES CONTINUAS. LA LEY DE PROBABILIDAD ESTA REPRESENTADA POR LA FUNCION DE DENSIDAD $f(x)$. LA PROBABILIDAD DE QUE X ESTE EN UN INTERVALO $[x, dx]$ ES IGUAL A $f(x)dx$.

CUALQUIER FUNCION $f(x)$ QUE SATISFAGA

(a) $f(x) \geq 0$

(b) $\int_{-\infty}^{\infty} f(x)dx = 1$

ES UNA FUNCION DE DENSIDAD.

AL IGUAL QUE EN EL CASO DISCRETO, OTRA FORMA DE DESCRIBIR LA DISTRIBUCION DE PROBABILIDAD DE UNA VARIABLE ALEATORIA CONTINUA, ES LA FUNCION DE DISTRIBUCION ACUMULATIVA. ESTA DEFINIDA COMO

$$F(x) = P(X \leq x) = P(-\infty \leq X \leq x) = \int_{-\infty}^x f(u) du$$

SI SE CONOCE $F(x)$, ENTONCES

$$f(x) = \frac{dF(x)}{dx}$$

2. DOS VARIABLES ALEATORIAS

CUANDO DOS O MAS VARIABLES ALEATORIAS ESTAN SIENDO CONSIDERADAS EN FORMA SIMULTANEA, SU COMPORTAMIENTO ESTA DETERMINADO POR UNA

LEY
DE
PROBABILIDAD
CONJUNTA

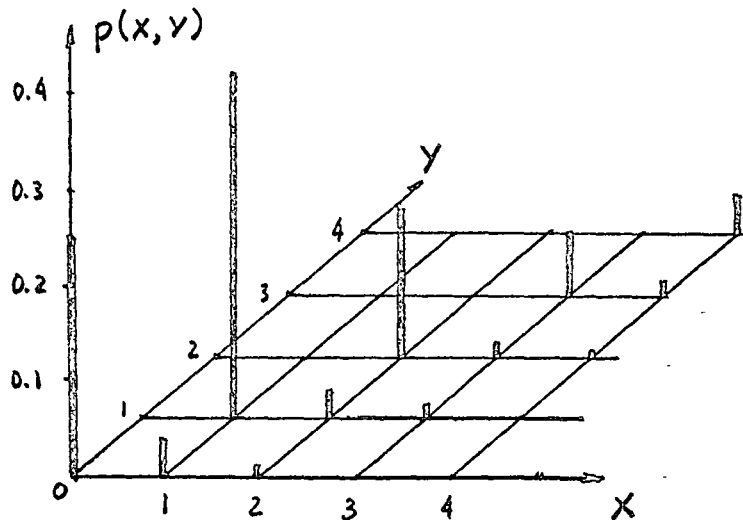
A. VARIABLES ALEATORIAS DISCRETAS

CONSIDERESE EL SIGUIENTE PROBLEMA DE INGENIERIA DE TRANSITO: REPRESENTESE POR X , EL NUMERO ALEATORIO DE VEHICULOS QUE PASAN POR UN PUNTO EN UN INTERVALO DE 30 SEG. LA VARIABILIDAD EN EL FLUJO DE TRAN-

SITO ES LA CAUSA DE DEMORAS Y CONGESTIONAMIENTO. LA OTRA VARIABLE ALEATORIA Y, ES EL NUMERO DE VEHICULOS EN EL MISMO INTERVALO DE 30 SEG. REGISTRADO POR UN CONTADOR DE TRANSITO, IMPERFECTO. ESTE DISPOSITIVO RESPONDE A LA PRESION EJERCIDA SOBRE UN CABLE COLOCADO A TRAVES DE UNA O MAS LINEAS DE TRANSITO, Y REGISTRA EL NUMERO TOTAL DE TALEZ APLICACIONES DE PRESION, DURANTE INTERVALOS SUCESIVOS DE 30 SEG. SIN EMBARGO, DEBIDO A EFECTOS DINAMICOS Y A INHABILIDAD MECANICA DEL CONTADOR PARA RESPONDER A ESOS PULSOS, EL NUMERO REAL DE VEHICULOS X Y EL NUMERO REGISTRADO Y , NO COINCIDEN SIEMPRE. DE LOS DATOS REUNIDOS PARA DETERMINAR LA NATURALEZA Y MAGNITUD DE ESTA FALTA DE CONFIABILIDAD EN EL CONTADOR, SE ADOPTARON DIRECTAMENTE LAS FRECUENCIAS RELATIVAS OBSERVADAS COMO PROBABILIDADES. ESTAS FUERON:

$P[X=0, Y=0] = 0.25$	$P[X=3, Y=2] = 0.02$
$P[X=1, Y=0] = 0.04$	$P[X=4, Y=2] = 0.0025$
$P[X=2, Y=0] = 0.01$	$P[X=0, Y=3] = 0$
$P[X=3, Y=0] = 0$	$P[X=1, Y=3] = 0$
$P[X=4, Y=0] = 0$	$P[X=2, Y=3] = 0$
$P[X=0, Y=1] = 0$	$P[X=3, Y=3] = 0.07$
$P[X=1, Y=1] = 0.36$	$P[X=4, Y=3] = 0.015$
$P[X=2, Y=1] = 0.03$	$P[X=0, Y=4] = 0$
$P[X=3, Y=1] = 0.01$	$P[X=1, Y=4] = 0$
$P[X=4, Y=1] = 0$	$P[X=2, Y=4] = 0$
$P[X=0, Y=2] = 0$	$P[X=3, Y=4] = 0$
$P[X=1, Y=2] = 0$	$P[X=4, Y=4] = 0.0325$
$P[X=2, Y=2] = 0.16$	

EN FORMA DE GRAFICA



LA PROBABILIDAD DESEADA, ESTO ES, LA PROBABILIDAD DE QUE EL CONTADOR ESTE CORRECTO ES:

$$\begin{aligned}
 \sum_{x_i} P[X=x_i, Y=x_i] &= P(X=Y) \\
 &= p(0,0) + p(1,1) + p(2,2) + \\
 &\quad p(3,3) + p(4,4) \\
 &= 0.25 + 0.36 + 0.16 + 0.07 + 0.0325 \\
 &= 0.8725
 \end{aligned}$$

LA PROBABILIDAD DE QUE EL CONTADOR NO ESTE CORRECTO ES

$$1 - P(X=Y) = 0.1275$$

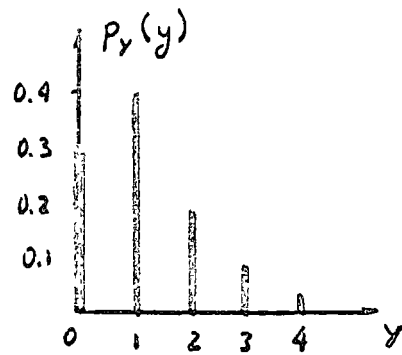
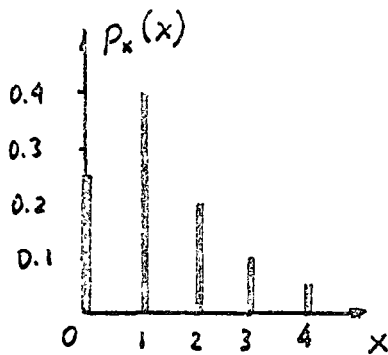
OTRAS PROBABILIDADES QUE PUEDEN SER DE INTERES SON:

* PROBABILIDAD DE QUE EL NUMERO REAL DE CARROS SEA 3 ES

$$P(X=3) = p_x(3) = \sum_{y=0}^4 p(3,y) = 0.01 + 0.02 + 0.07 = 0.1$$

A ESTA SE LE LLAMA PROBABILIDAD MARGINAL, DE $X=3$.

LAS PROBABILIDADES MARGINALES DE X E Y SON :

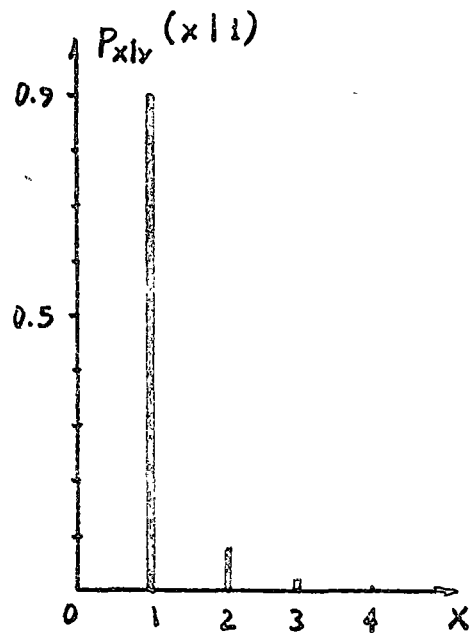


* SUPONGASE QUE EN EL CONTADOR SE LEE $Y=1$ EN UN INTERVALO EN PARTICULAR.

EL NUMERO REAL DE CARRROS NO SE CO-NOCE CON CERTEZA, PERO LAS FRECUENCIAS RELATIVAS DE DIFERENTES VALORES ESTAN DADAS POR $p(x,1)$, PARA $x=0,1,2,3,4$.

NORMALIZADAS POR LA SUMA, ESAS PROBABILIDADES LLEGAN A SER LA DISTRIBUCION CONDICIONAL

$P_{x|y}(x,1)$.



LA FUNCION DE MASA DE PROBABILIDAD CONJUNTA $p(x,y)$ DE DOS VARIABLES ALEATORIAS DISCRETAS X E Y SE DEFINE COMO

$$\begin{aligned}
 p(x, y) &\equiv P[(X=x) \cap (Y=y)] \\
 &= P[\{s \mid X(s)=x\} \cap \{s \mid Y(s)=y\}]
 \end{aligned}$$

LA FUNCION DE DISTRIBUCION ACUMULATIVA CONJUNTA ESTA DEFINIDA COMO

$$F(x, y) \equiv P[(X \leq x) \cap (Y \leq y)] = \sum_{x_i \leq x} \sum_{y_i \leq y} p(x_i, y_i)$$

LAS FUNCIONES DE MASA DE PROBABILIDAD MARGINALES SE DEFINEN COMO

$$p(x) \equiv P(X=x) = \sum_{y_i} p(x, y_i)$$

$$p(y) \equiv P(Y=y) = \sum_{x_i} p(x_i, y)$$

Y LAS DISTRIBUCIONES ACUMULATIVAS

$$F(x) = \sum_{x_i \leq x} \sum_{y_i} p(x_i, y_i) = \sum_{x_i \leq x} p(x_i)$$

$$F(y) = \sum_{y_i \leq y} \sum_{x_i} p(x_i, y_i) = \sum_{y_i \leq y} p(y_i)$$

LA FUNCION DE MASA DE PROBABILIDAD CONDICIONAL ESTA DEFINIDA COMO

$$\begin{aligned}
 P_{X|Y}(x, y) &= P[X=x \mid Y=y] = \frac{P[(X=x) \cap (Y=y)]}{P[Y=y]} \\
 &= \frac{p(x, y)}{\sum_{x_i} p(x_i, y)} = \frac{p(x, y)}{p(y)}
 \end{aligned}$$

(47)

$$P_{x|y>y} (x,y) = P[X=x | Y>y] = \frac{\sum_{y_i>y} p(x,y_i)}{\sum_{y_i>y} P_y(y_i)}$$

$$P_{y|a<x\leq b} (y) = P[Y=y | a < X \leq b]$$

$$= \frac{\sum_{a < x_j \leq b} p(x_j, y)}{\sum_{a < x_j \leq b} P_x(x_j)}$$

EN LA PRACTICA, LAS PROBABILIDADES CONDICIONALES SE PUEDEN DETERMINAR MAS RAPIDAMENTE QUE LAS PROBABILIDADES CONJUNTAS. JUNTO CON LA PROBABILIDAD MARGINAL, UNA CONDICIONAL SE PUEDE UTILIZAR PARA CALCULAR LA CONJUNTA; ES DECIR,

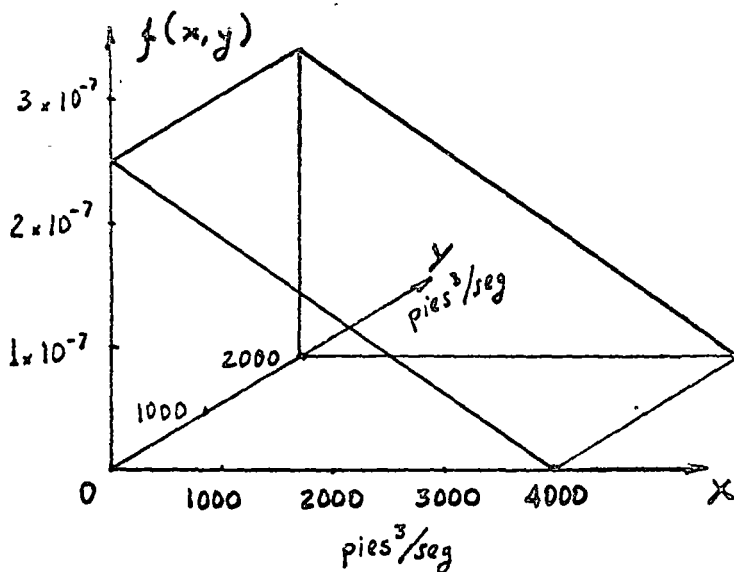
$$p(x,y) = P_{x|y}(x,y) P(y) = P_{y|x}(x,y) P(x)$$

B. VARIABLES ALEATORIAS CONTINUAS

CONSIDERESE EL FLUJO DE DOS CORRIENTES DIFERENTES X E Y, EN EL MISMO DIA. SE ESTA INTERESADO EN EL COMPORTAMIENTO PROBABILISTICO CONJUNTO DEBIDO A QUE ALIMENTAN EL MISMO DEPOSITO.

(48)

SUPONGASE QUE LA DISTRIBUCION CONJUNTA ES



$$f(x, y) = C \frac{4000 - x}{4000} \quad \begin{array}{l} 0 \leq x \leq 4000 \\ 0 \leq y \leq 2000 \end{array}$$

DONDE $C = 2.5 \times 10^{-7}$

POR EJEMPLO, LA PROBABILIDAD DE QUE EL FLUJO DE X SEA MAS DEL DOBLE DE Y, ES:

$$P[X \geq 2Y] = \int_0^{4000} \int_0^{0.5x} \frac{C}{4000} (4000 - x) dy dx = \frac{1}{3}$$

LAS DISTRIBUCIONES MARGINALES SON.

$$\begin{aligned} f(x) &= \int_0^{2000} C \frac{4000 - x}{4000} dy \\ &= C \frac{4000 - x}{2} \quad 0 \leq x \leq 4000 \end{aligned}$$

(49)

$$f(y) = \int_0^{4000} c \frac{4000-x}{4000} dx = 2000 c \quad 0 \leq y \leq 2000$$

LA PROBABILIDAD CONDICIONAL DE X DADO Y, ES

$$\begin{aligned} f_{x|y}(x,y) &= \frac{f(x,y)}{f(y)} = \frac{c(4000-x)/4000}{2000 c} \\ &= \frac{1}{2000} \frac{4000-x}{4000} ; 0 \leq x \leq 4000 \end{aligned}$$

LAS FUNCIONES ASOCIADAS CON VARIABLES ALEATORIAS DISTRIBUIDAS CONJUNTAMENTE (CONTINUAS), SON ANALOGAS A LAS ASOCIADAS CON VARIABLES ALEATORIAS DISCRETAS.

LA PROBABILIDAD DE LA OCURRENCIA CONJUNTA DE X E Y ES

$$P[x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2] = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x,y) dx dy$$

LA FUNCION DE DENSIDAD CONJUNTA DEBE SATISFACER LAS CONDICIONES:

(a) $f(x,y) \geq 0$

(b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = 1$

LA FUNCION DE DISTRIBUCION ACUMULATIVA

ES:

$$\begin{aligned} F(x, y) &\equiv P[(X \leq x) \cap (Y \leq y)] \\ &= P[(-\infty \leq X \leq x) \cap (-\infty \leq Y \leq y)] \\ &= \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv \end{aligned}$$

A PARTIR DE ESTA ES POSIBLE OBTENER

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

LA PROBABILIDAD MARGINAL ES

$$f(x) \equiv \int_{-\infty}^{\infty} f(x, y) dy$$

o

$$f(y) \equiv \int_{-\infty}^{\infty} f(x, y) dx$$

Y SU FUNCION ACUMULATIVA

$$F(x) = P[X \leq x] = \int_{-\infty}^x f(u) du$$

o

$$F(y) = P[Y \leq y] = \int_{-\infty}^y f(v) dv$$

(51)

LA PROBABILIDAD CONDICIONAL ES

$$f_{x|y}(x,y) = \frac{f(x,y)}{f(y)}$$

Y SU DISTRIBUCION ACUMULATIVA

$$F_{x|y}(x,y) = P[X \leq x | Y=y] = \int_{-\infty}^x f_{x|y}(u,y) du$$

FINALMENTE, SI X E Y SON VARIABLES ALEATORIAS INDEPENDIENTES, SE CUMPLE

$$f_{x|y}(x,y) = f(x) \quad ; \quad f_{y|x}(y,x) = f(y)$$

$$f_{x,y}(x,y) = f(x) f(y)$$

$$F(x,y) = F(x) F(y)$$

$$F_{x|y}(x,y) = F(x)$$

3. FUNCIONES DE VARIABLES ALEATORIAS.

A. FUNCION DE UNA VARIABLE ALEATORIA.

SUPONGASE QUE SE DESEA CONSTRUIR UNA ALBERCA Y SE QUIERE CONOCER LA DISTRIBUCION DEL COSTO TOTAL.

ESTE COSTO VA A DEPENDER DEL COSTO DE MATERIALES, QUE APROXIMADAMENTE SERA DE \$10000, MAS EL COSTO DE LA CUADRILLA DE TRABAJADORES A RAZON DE \$100/hr. NO SE CONOCE CON CERTEZA EL NUMERO DE HORAS H , PARA REALIZAR ESE TRABAJO; POR LO TANTO, H SERA UNA VARIABLE ALEATORIA. SU FUNCION DE DISTRIBUCION ACUMULATIVA ES

$$F_H(h) = 1 - \left(\frac{h-110}{10} \right)^2 \quad (1)$$

LA FUNCION DE DENSIDAD ES

$$\begin{aligned} \frac{dF(h)}{dh} &= -0.2 \left(\frac{h-110}{10} \right) \\ &= 2.2 - 0.02h \quad (2) \end{aligned}$$

(53)

EL COSTO TOTAL C , SE PUEDE EXPRESAR
COMO

$$C = 10\,000 + 100 H \quad (3)$$

INTERESA CONOCER $F_C(c)$ Y $f_C(c)$. POR
TANTO

$$\begin{aligned} F_C(c) &= P(C \leq c) = P(10\,000 + 100 H \leq c) \\ &= P\left(H \leq \frac{c - 10\,000}{100}\right) \\ &= F_H\left(\frac{c - 10\,000}{100}\right) \end{aligned} \quad (4)$$

DE (1)

$$\begin{aligned} F_C(c) &= 1 - \left[\frac{(c - 10\,000)/100 - 110}{10} \right]^2 \\ &= 1 - \left(\frac{c - 21\,000}{1000} \right)^2 \quad 20\,000 \leq c \leq 21\,000 \end{aligned} \quad (5)$$

DERIVANDO (5)

$$f_C(c) = -\frac{2}{1000} \left(\frac{c - 21\,000}{1000} \right)$$

EN MUCHOS PROBLEMAS DE INGENIERIA
SE TRABAJAN RELACIONES FUNCIONALES QUE
PREDECEN EL VALOR DE UNA VARIABLE
(DEPENDIENTE), DADO CUALQUIER VALOR DE

OTRA VARIABLE (INDEPENDIENTE). CUANDO LA VARIABLE INDEPENDIENTE ES UNA VARIABLE ALEATORIA, ESTA ALEATORIEDAD SE TRASMITE A LA VARIABLE DEPENDIENTE.

A CONTINUACION SE VERA LA FORMA DE DETERMINAR LA LEY DE PROBABILIDAD DE LA VARIABLE ALEATORIA DEPENDIENTE.

SEA UNA VARIABLE ALEATORIA X , Y UNA FUNCION REAL DE VARIABLE X , $g(x)$. SEA LA VARIABLE ALEATORIA Y , TAL QUE

$$Y = g(X)$$

RECUERDESE QUE PARA CADA RESULTADO $s \in S$, EXISTE UN NUMERO REAL $X(s)$. Y , ESTA DEFINIDA PARA TODO $s \in S$.
POR TANTO

$$Y(s) = g[X(s)]$$

DADO UN NUMERO REAL y , SEA R_y EL CONJUNTO DE TODOS LOS REALES x TAL QUE $g(x) \leq y$. ENTONCES

$$\{Y \leq y\} = \{X \in R_y\}$$

SEGUN ESTO

$$F_Y(y) = P(Y \leq y) = P(X \in R_y)$$

TEOREMA. 12. PARA ENCONTRAR $f_Y(y)$, SE RESUELVE LA ECUACION

$$y = g(x)$$

PARA x EN FUNCION DE y . SI

$x_1, x_2, \dots, x_n, \dots$ SON TODAS RAICES REALES, ENTONCES

$$f_Y(y) = \frac{f_X(x_1)}{|g'(x_1)|} + \dots + \frac{f_X(x_n)}{|g'(x_n)|} + \dots$$

DONDE
$$g'(x) = \frac{dg(x)}{dx}$$

SI PARA UNA y DADA, LA ECUACION

$$y = g(x)$$

NO TIENE RAICES REALES, ENTONCES

$$f_Y(y) = 0$$

EJEMPLO 4. SEAN LAS VARIABLES ALEATORIAS X E Y , TAL QUE

$$Y = aX^2, \quad a > 0$$

(56)

SEGUN EL TEOREMA

$$x = \pm \sqrt{y/a}$$

SI $y < 0$, NO EXISTEN RAICES REALES
Y POR TANTO $f_Y(y) = 0$

SI $y > 0$, SE TIENEN DOS RAICES REALES

$$x_1 = \sqrt{y/a}, \quad x_2 = -\sqrt{y/a}$$

AHORA, $g'(x) = 2ax$

$$\therefore g'(x_1) = 2ax_1 = 2\sqrt{ay}$$

$$g'(x_2) = 2ax_2 = -2\sqrt{ay}$$

POR TANTO

$$f_Y(y) = \frac{1}{2\sqrt{ay}} \left[f_x(\sqrt{y/a}) + f_x(-\sqrt{y/a}) \right] u(y)$$

$$\text{DONDE } u(y) = \begin{cases} 1 & \text{si } y \geq 0 \\ 0 & \text{si } y < 0 \end{cases}$$

LA FUNCION DE DISTRIBUCION ACUMULATIVA
ES

$$F_Y(y) = P[-\sqrt{y/a} \leq x \leq \sqrt{y/a}]$$

(57)

O SEA

$$F_y(y) = F_x(\sqrt{y/a}) - F_x(-\sqrt{y/a})$$

PARA TODA $y > 0$,

$$F_y(y) = 0 \quad \text{PARA TODA } y < 0$$

B. FUNCION DE DOS VARIABLES ALEATORIAS

SIGUIENDO UN RAZONAMIENTO ANALOGO AL CASO DE FUNCION DE UNA SOLA VARIABLE, SE PUEDEN OBTENER LAS FUNCIONES DE DENSIDAD Y DE DISTRIBUCION ACUMULATIVA, DE UNA FUNCION DE DOS VARIABLES ALEATORIAS.

DADAS DOS VARIABLES ALEATORIAS X E Y , Y UNA FUNCION $g(x, y)$ DE LAS VARIABLES REALES x e y , ENTONCES

$$Z = g(X, Y)$$

ES UNA VARIABLE ALEATORIA Y SU VALOR $Z(s)$, PARA $s \in S$, ESTA DADO POR

$$Z(s) = g[X(s), Y(s)]$$

(58)

SEA D_2 LA REGION DEL PLANO xy
TAL QUE

$$g(x, y) \leq z$$

ENTONCES

$$\{Z \leq z\} = \{(X, Y) \in D_2\}$$

POR TANTO

$$F_2(z) = P(Z \leq z) = P\{(X, Y) \in D_2\}$$

$$= \iint_{D_2} f_{xy}(x, y) dx dy$$

NUEVAMENTE

$$f_2(z) = \frac{dF_2(z)}{dz}$$

EJEMPLO 5. SEAN X E Y DOS VARIABLES
ALEATORIAS, TAL QUE

$$Z = X + Y$$

$$D_2 = \{(x, y) \mid x + y \leq z\}$$

ENTONCES

$$F_2(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{xy}(x, y) dx dy$$

(59)

$$f_z(z) = \int_{-\infty}^{\infty} f_{x,y}(z-y, y) dy$$

TEOREMA. 13. SI LAS VARIABLES ALEATORIAS X E Y SON INDEPENDIENTES, ENTONCES LA DENSIDAD DE

$$Z = X + Y$$

ES IGUAL A LA CONVOLUCION DE SUS RESPECTIVAS DENSIDADES :

$$\begin{aligned} f_z(z) &= \int_{-\infty}^{\infty} f_x(z-y) f_y(y) dy \\ &= \int_{-\infty}^{\infty} f_x(x) f_y(z-x) dx \end{aligned}$$

4. VALORES ESPERADOS Y MOMENTOS.

A. VALORES ESPERADOS

SEA X UNA VARIABLE ALEATORIA CON DENSIDAD $f_x(x)$. ENTONCES

(59-A)

CONSIDERESE EL LANZAMIENTO DE TRES MONEDAS IDEALES. LA DISTRIBUCION DEL NUMERO DE AGUILAS ES

$$p(x) = f(x) = \binom{3}{x} \left(\frac{1}{2}\right)^3 ; x = 0, 1, 2, 3$$

LAS PROBABILIDADES PARA LOS POSIBLES RESULTADOS SON

$$p(0) = \frac{1}{8} ; p(1) = \frac{3}{8}$$

$$p(2) = \frac{3}{8} ; p(3) = \frac{1}{8}$$

EL PROMEDIO DEL NUMERO DE AGUILAS ESPERADO ES

$$0 \times p(0) + 1 \times p(1) + 2 \times p(2) + 3 \times p(3) = 1.5$$

ESTE VALOR ASI OBTENIDO SE LLAMA VALOR ESPERADO (MEDIO) DE X. ESTE VALOR ES TEORICO YA QUE, POR EJEMPLO EN ESTE CASO, NO ES POSIBLE OBTENER EN PROMEDIO, UNA AGUILA Y MEDIA EN EL LANZAMIENTO DE LAS TRES MONEDAS.

EL VALOR ESPERADO DE X, $E(X)$, ES

$$E(X) = \sum_{x_i} x_i f(x_i)$$

(60)

SI X ES DISCRETA, Y

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

SI X ES CONTINUA

TEOREMA. 14. SEA X UNA VARIABLE ALEATORIA CON DENSIDAD $f(x)$. EL VALOR ESPERADO DE UNA FUNCION g DE LA VARIABLE X ES

$$E[g(x)] = \sum_{x_i} g(x_i) f(x_i)$$

SI X ES DISCRETA, Y

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

SI X ES CONTINUA.

EJEMPLO 6. SEA $f(x) = \binom{3}{x} \left(\frac{1}{2}\right)^3$, $x=0,1,2,3$
Y LA FUNCION

$$g(x) = x^2 + 1$$

(61)

ENTONCES

$$E[g(x)] = \sum_{x=0}^3 (x^2+1) f(x) = 4$$

SEAN X E Y DOS VARIABLES ALEATORIAS
CON DENSIDAD $f(x,y)$. ENTONCES

$$E(X) = \sum_{x_i} \sum_{y_i} x_i f(x_i, y_i)$$

$$E(Y) = \sum_{x_i} \sum_{y_i} y_i f(x_i, y_i)$$

EN CASO DE QUE X E Y SEAN DISCRETAS, Y

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

$$E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y) dx dy$$

SI X E Y SON CONTINUAS

TEOREMA. 15. SEAN X E Y DOS VARIABLES
ALEATORIAS CON DENSIDAD $f(x,y)$.

EL VALOR ESPERADO DE UNA FUNCION
 g DE LAS VARIABLES ALEATORIAS ES

$$E[g(x,y)] = \sum_x \sum_y g(x,y) f(x,y)$$

(62)

SI LAS VARIABLES SON DISCRETAS, Y

$$E[g(x,y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dx dy$$

SI LAS VARIABLES SON CONTINUAS

EJEMPLO 7. SEA X UNA VARIABLE ALEATORIA
CON DENSIDAD

$$f(x,y) = 2xy; 0 < x < 1, 0 < y < 1 \\ = 0 \quad \text{EN OTRO CASO}$$

SEA LA FUNCION

$$g(x,y) = 3x + 2y$$

EL VALOR ESPERADO ES

$$E[g(x,y)] = \int_0^1 \int_0^1 (3x + 2y) 2xy = \frac{5}{3}$$

TEOREMA. 16. SEAN X E Y DOS VARIABLES
ALEATORIAS CON DENSIDAD $f(x,y)$.
SEAN c_1 Y c_2 CONSTANTES Y g_1 Y g_2
FUNCIONES DE LAS VARIABLES ALEATORIAS.
ENTONCES

$$E[c_1 g_1(x,y) + c_2 g_2(x,y)] =$$

$$= c_1 E[g_1(x,y)] + c_2 E[g_2(x,y)]$$

(63)

EJEMPLO 8. SEAN LAS VARIABLES X E Y
Y LAS FUNCIONES $g_1(x) = 3x$,
 $g_2(y) = 2y$. ENTONCES

$$E[3x+2y] = 3E[x] + 2E[y]$$

B. MOMENTOS

LOS MOMENTOS DE UNA DISTRIBUCION SON LOS VALORES ESPERADOS DE LAS POTENCIAS DE LA VARIABLE ALEATORIA QUE TIENE LA DISTRIBUCION DADA. EL r th MOMENTO DE X , DENOTADO POR μ'_r , SE DEFINE COMO

$$\mu'_r = E(x^r) = \int_{-\infty}^{\infty} x^r f(x) dx$$

EL PRIMER MOMENTO, μ'_1 , SE LLAMA MEDIA DE X .

LOS MOMENTOS ACERCA DE CUALQUIER PUNTO a ESTAN DEFINIDOS POR

$$E[(x-a)^r] = \int_{-\infty}^{\infty} (x-a)^r f(x) dx$$

(64)

CUANDO a ES IGUAL A LA MEDIA, LOS MOMENTOS SE DENOTAN POR μ_r

$$\mu_r = E[(x - \mu_1)^r] = \int_{-\infty}^{\infty} (x - \mu_1)^r f(x) dx$$

EL PRIMER MOMENTO RESPECTO A LA MEDIA ES

$$\begin{aligned} \mu_1 &= \int_{-\infty}^{\infty} x f(x) dx - \mu_1 \int_{-\infty}^{\infty} f(x) dx \\ &= \mu_1 - \mu_1 = 0 \end{aligned}$$

EL SEGUNDO MOMENTO ES

$$\begin{aligned} \mu_2 &= \int_{-\infty}^{\infty} (x - \mu_1)^2 f(x) dx \\ &= \mu_2' - 2\mu_1 \mu_1 + (\mu_1')^2 \end{aligned}$$

$$\mu_2 = \mu_2' - (\mu_1')^2$$

QUE SE LLAMA VARIANZA DE X .

(65)

EN ESTADISTICA APLICADA, LOS PRIMEROS DOS MOMENTOS SON DE GRAN IMPORTANCIA, PERO EL TERCER MOMENTO (MEDIDA DE ASIMETRIA) Y LOS MOMENTOS MAYORES SON RARAMENTE DE UTILIDAD.

EJEMPLO 9. SEA LA VARIABLE ALEATORIA X CON DENSIDAD

$$f(x) = \begin{cases} 2(1-x) & 0 < x < 1 \\ = 0 & \text{EN OTRO CASO} \end{cases}$$

EL rth MOMENTO ES

$$\mu_r' = E(x^r) = \int_0^1 x^r [2(1-x)] dx$$

$$\mu_r' = \frac{2}{(r+1)(r+2)}$$

LA MEDIA ES

$$\mu = \mu_1' = \frac{1}{3}$$

Y LA VARIANZA

$$\mu_2 = \sigma^2 = \mu_2' - \mu^2 = \frac{1}{18}$$

C. VALOR ESPERADO CONDICIONAL

SEAN X E Y DOS VARIABLES ALEATORIAS CON DISTRIBUCION $f(x, y)$. SEA $g(y|x)$ LA DENSIDAD CONDICIONAL DE Y DADA $X=x$, Y SEA $h(x)$ LA DENSIDAD MARGINAL DE X . EL VALOR ESPERADO CONDICIONAL DE Y DADA $X=x$, $E(y|x)$ SE DEFINE COMO

$$E(y|x) = \int_{-\infty}^{\infty} y g(y|x) dy = \int_{-\infty}^{\infty} y \frac{f(x, y)}{h(x)} dy$$

EJEMPLO 10. SEAN LAS VARIABLES ALEATORIAS X E Y CON DISTRIBUCION

$$f(x, y) = 8xy \quad ; \quad 0 < y < x < 1$$

$$\text{ENTONCES: } h(x) = \int_0^x 8xy dy = 4x^3 \quad ; \quad 0 < x < 1$$

$$g(y|x) = \frac{f(x, y)}{h(x)} = \frac{2y}{x^2}$$

$$E(y|x) = \int_0^x y g(y|x) dy = \int_0^x y \frac{2y}{x^2} dy = \frac{2}{3} x$$

5. DISTRIBUCIONES ESPECIALES.

A. VARIABLES ALEATORIAS DISCRETAS

1. DISTRIBUCION BINOMIAL

SI LA PROBABILIDAD DE QUE UN EVENTO OCURRA ES p , Y LA PROBABILIDAD DE QUE NO OCURRA ES $q=1-p$, ENTONCES LA DENSIDAD DE LA VARIABLE ALEATORIA X (NUMERO DE OCURRENCIAS) ES

$$f(x) = p^x q^{1-x} ; x=0,1$$

ESTA ES LA DISTRIBUCION BINOMIAL PUNTUAL.

EN n PRUEBAS, SEA p LA PROBABILIDAD DE UN EVENTO QUE OCURRE EN CADA PRUEBA, Y SEAN TODAS LAS PRUEBAS INDEPENDIENTES. LA DENSIDAD DE LA VARIABLE ALEATORIA X (NUMERO DE OCURRENCIAS EN n PRUEBAS) ES

$$f(x) = \binom{n}{x} p^x q^{n-x} ; x=0,1,2,\dots,n$$

ESTA ES LA DISTRIBUCION BINOMIAL.

2. DISTRIBUCION DE POISSON

LA DISTRIBUCION DE POISSON TIENE UTIL APLICACION EN SITUACIONES DONDE UN GRAN NUMERO DE OBJETOS ESTAN DISTRIBUIDOS SOBRE UNA GRAN AREA.

SEA X UNA VARIABLE ALEATORIA. SE DICE QUE ESTA DISTRIBUIDA COMO POISSON SI LA DENSIDAD ES

$$f(x) = \frac{e^{-m} m^x}{x!} ; x=0,1,2,\dots$$

DONDE m ES CUALQUIER NUMERO POSITIVO.

B. VARIABLES ALEATORIAS CONTINUAS1. DISTRIBUCION UNIFORME

SEA X UNA VARIABLE ALEATORIA. SE DICE QUE ESTA DISTRIBUIDA UNIFORMEMENTE SI SU DENSIDAD ES

$$f(x) = \frac{1}{b-a} ; a < x < b$$

SE UTILIZA CUANDO TODOS LOS POSIBLES VALORES DE LA VARIABLE ALEATORIA TIENEN LA MISMA PROBABILIDAD.

2. DISTRIBUCION NORMAL

MUCHAS DE LAS TECNICAS UTILIZADAS EN ESTADISTICA APLICADA SE BASAN EN LA DISTRIBUCION NORMAL. FRECUENTEMENTE SE ADOPTA COMO APROXIMACIONES CONVENIENTES A OTRAS DISTRIBUCIONES. SE DEFINE DE LA FORMA SIGUIENTE:

SEA X UNA VARIABLE ALEATORIA. SE DICE QUE ESTA DISTRIBUIDA NORMALMENTE SI SU DENSIDAD ESTA DADA POR

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad ; -\infty < x < \infty$$

CAMBIANDO EL PARAMETRO μ , SE PUEDE DESPLAZAR LA CURVA A LA IZQUIERDA O DERECHA SIN CAMBIAR LA FORMA.

LA TRANSFORMACION $y = (x-\mu)/\sigma$, HACE QUE TODAS LAS DISTRIBUCIONES NORMALES TENGAN LA MISMA FORMA, LLAMADA FORMA ESTANDAR O NORMALIZADA.

3. DISTRIBUCION LOG-NORMAL

SEA X UNA VARIABLE ALEATORIA,
Y SEA

$$Y = \ln X$$

SI Y ES UNA VARIABLE CON DISTRIBUCION
NORMAL, ENTONCES SE DICE QUE X
TIENE UNA DISTRIBUCION LOG-NORMAL

ESTA DISTRIBUCION FUE UTILIZADA,
DESDE UN PRINCIPIO, EN ESTUDIOS
ESTADISTICOS DE DATOS HIDROLOGICOS Y
DE FALLAS POR FATIGA.

LA FUNCION DE DENSIDAD ES

$$f(x) = \frac{1}{xb\sqrt{2\pi}} e^{-\frac{1}{2b^2}(\ln x - \ln a)^2} \quad ; \quad x > 0$$

= 0 EN OTRO CASO

DONDE $a > 0$, $b > 0$.

IV. REDUCCION DE DATOS

ANEXO A. [4]

OBJETIVO

CONOCER LOS CONCEPTOS FUNDAMENTALES UTILIZADOS EN EL ANALISIS DE DATOS EXPERIMENTALES.

Data Reduction

A necessary first step in any engineering situation is an investigation of available data to assess the nature and the degree of the uncertainty. An unorganized list of numbers representing the outcomes of tests is not easily assimilated. There are several methods of organization, presentation, and reduction of observed data which facilitate its interpretation and evaluation.

It should be pointed out explicitly that the treatment of data described in this chapter is in no way dependent on the assumptions that there is randomness involved or that the data constitute a random sample of some mathematical probabilistic model. These are terms which we shall come to know and which some readers may have encountered previously. The methods here are simply convenient ways to reduce raw data to manageable forms.

In studying the following examples and in doing the suggested problems, it is important that the reader appreciate that the data are "real" and that the variability or scatter is representative of the magnitudes of variation to be expected in civil-engineering problems.

1.1 GRAPHICAL DISPLAYS

Histograms A useful first step in the representation of observed data is to reduce it to a type of bar chart. Consider, for example, the data presented in Table 1.1.1. These numbers represent the live loads observed in a New York warehouse. To anticipate the typical, the extreme, and the long-term behavior of structural members and footings in such structures, the engineer must understand the nature of load distribution. Load variability will, for example, influence relative settlements of the column footings. The values vary from 0 to 229.5 pounds per square foot (psf). Let us divide this range into 20-psf intervals, 0 to 19.9, 20.0 to 39.9, etc., and tally the number of occurrences in each interval.

Plotting the frequency of occurrences in each interval as a bar

Table 1.1.1 Floor-load data*

Bay	Base- ment	1st	2d	3d	4th	5th	6th	7th	8th	9th
A	0	7.8	36.2	60.6	64.0	64.2	79.2	88.4	38.0	72.7
B	72.2	72.6	74.4	21.8	17.1	48.5	16.8	105.9	57.2	75.7
C	225.7	42.5	59.8	41.7	39.9	55.5	67.2	122.8	45.2	62.9
D	55.1	55.9	87.7	59.2	63.1	58.8	67.7	90.4	43.3	55.2
E	36.6	26.0	90.5	23.0	43.5	52.1	102.1	71.7	4.1	37.3
F	129.4	66.4	138.7	127.9	90.9	46.9	197.5	151.1	157.3	197.0
G	134.6	73.4	80.9	53.3	80.1	62.9	150.8	102.2	6.4	45.4
H	121.0	106.2	94.4	139.6	152.5	70.2	111.8	174.1	85.4	83.0
I	178.8	30.2	44.1	157.0	105.3	87.0	50.1	198.0	86.7	64.6
J	78.6	37.0	70.7	83.0	179.7	180.2	60.6	212.4	72.2	86.0
K	94.5	24.1	87.3	80.6	74.8	72.4	131.1	116.1	53.6	99.1
L	40.2	23.4	8.4	42.6	43.4	27.4	63.8	15.4	16.2	58.7
M	92.2	49.8	50.9	116.4	122.9	132.3	105.2	160.3	28.7	46.8
N	99.5	106.9	55.9	136.8	110.4	123.5	92.4	160.9	45.4	96.3
O	88.5	48.4	62.3	71.3	133.2	92.1	111.7	67.9	53.1	39.7
P	93.2	55.0	80.8	143.5	122.3	184.2	150.0	57.6	6.8	53.3
Q	96.1	54.8	63.0	228.3	139.3	59.1	112.1	50.9	158.6	139.1
R	213.7	65.7	90.3	198.4	97.5	155.1	163.4	155.3	229.5	75.0
S	137.6	62.5	156.5	154.1	134.3	81.6	194.4	155.1	89.3	73.4
T	79.8	68.7	85.6	141.6	100.7	106.0	131.1	157.4	80.2	65.0
U	78.5	118.2	126.4	33.8	124.6	78.9	146.0	100.3	97.8	75.3
V	24.8	55.6	135.6	56.3	66.9	72.2	107.4	98.9	101.7	58.2

* Observed live loads (in pounds per square foot), bay size 400 ± 6 ft²

Source: J. W. Dunham, G. N. Brekke, and G. N. Thompson [1952], "Live Loads on Floors in Buildings," Building Materials and Structures Report 133, National Bureau of Standards, p. 22

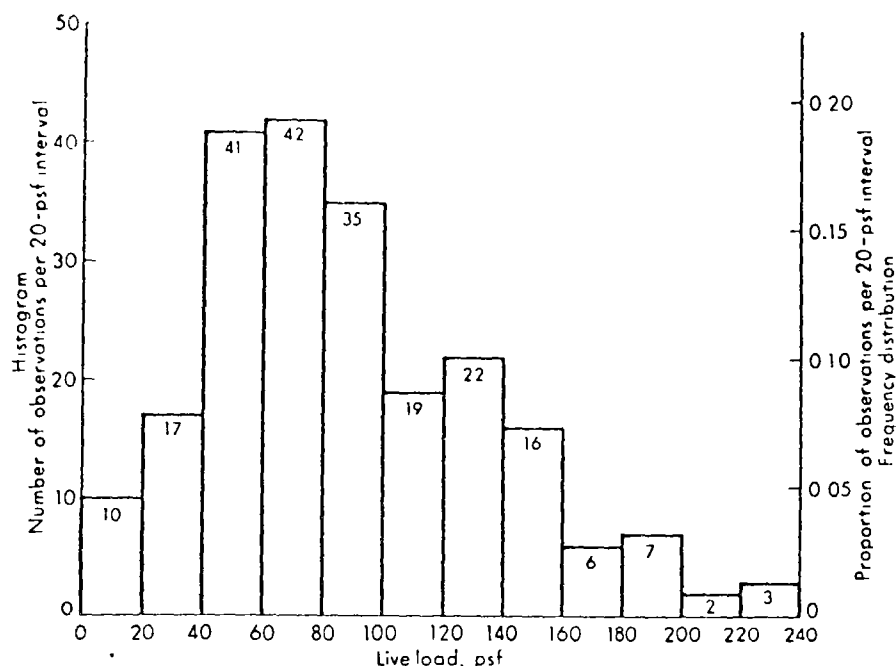


Fig. 1.1.1 Histogram and frequency distribution of floor-load data.

yields a *histogram*, as shown in Fig. 1.1.1. The height, and more usefully, the area, of each bar are proportional to the number of occurrences in that interval. The plot, unlike the array of numbers, gives the investigator an immediate impression of the range of the data, its most frequently occurring values, and the degree to which it is scattered about the central or typical values. We shall learn in Chap. 2 how the engineer can predict analytically from this shape the corresponding curve for the *total load* on a column supporting, say, 20 such bays.

If the scale of the ordinate of the histogram is divided by the total number of data entries, an alternate form, called the *frequency distribution*, results. In Fig. 1.1.1, the numbers on the right-hand scale were obtained by dividing the left-hand scale values by 220, the total number of observations. One can say, for example, that the proportion of loads observed to lie between 120 and 139.9 psf was 0.10. If this scale were divided by the interval length (20 psf), a *frequency density distribution* would result, with ordinate units of "frequency per psf." The area under this histogram would be unity. This form is preferred when different sets of data, perhaps with different interval lengths, are to be compared with one another.

The *cumulative frequency distribution*, another useful graphical representation of data, is obtained from the frequency distribution by calculating the successive partial sums of frequencies up to each interval division point. These points are then plotted and connected by straight lines to form a nondecreasing (or monotonic) function from zero to unity.

In Fig. 1.1.2, the cumulative frequency distribution of the floor-load data, the values of the function at 20, 40, and 60 psf were found by forming the partial sums $0 + 0.0455 = 0.0455$, $0.0455 + 0.0773 = 0.1230$, and $0.1230 + 0.1864 = 0.3090$ †. From this plot, one can read that the proportion of the loads observed to be equal to or less than 139.9 psf was 0.847. After a proper balancing of initial costs, consequences of poor performance, and these frequencies, the designer might conclude that a beam supporting one of these bays must be stiff enough to avoid deflections in excess of 1 in. in 99 percent of all bays. Thus the design should be checked for deflections under a load of 220 psf.

Some care should be taken in choosing the width of each interval

† When constructing the cumulative frequency distribution, one can avoid the arbitrariness of the intervals by plotting one point per observation, that is, by plotting i/n versus $x^{(i)}$, where $x^{(i)}$ is the i th in ordered list of data (see Fig. 1.2.1).

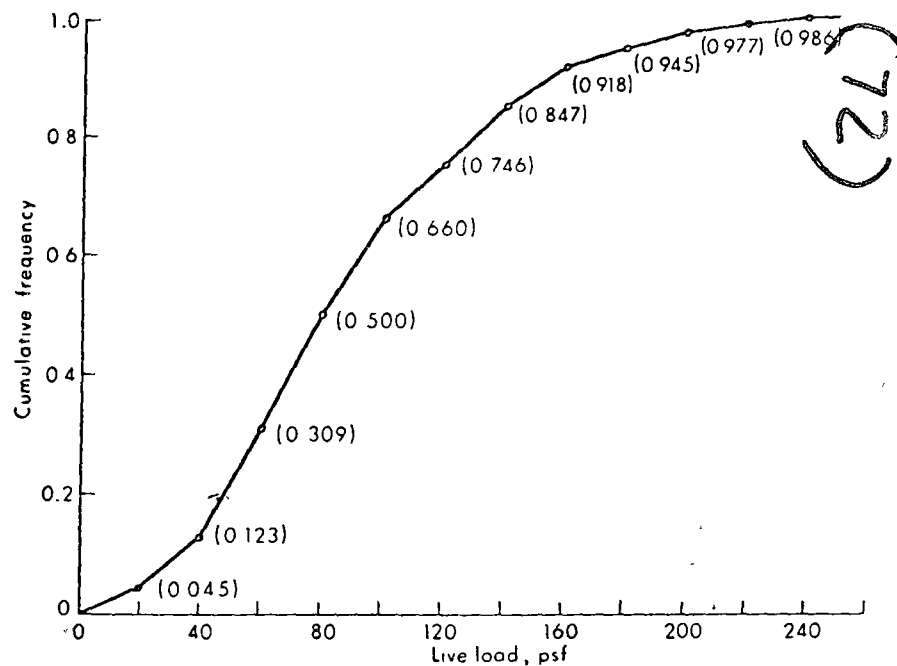


Fig. 1.1.2 Cumulative frequency distribution of floor-load data.

in these diagrams † A little experimentation with typical sets of data will convince the reader that the choice of the number of class intervals can alter one's impression of the data's behavior a great deal. Figure 1.1.3 contains two histograms of the data of Table 1.1.1, illustrating the influence of interval size. An empirical practical guide has been suggested by Sturges [1926]. If the number of data values is n , the number of intervals k between the minimum and maximum value observed should be about

$$k = 1 + 3.3 \log n \quad (1.1.1)$$

in which logarithms to the base 10 should be employed. Unfortunately, if the number of values is small, the choice of the precise point at which the interval divisions are to occur also may alter significantly the appearance of the histogram. Examples can be found in Sec. 1.2 and in the problems at the end of this chapter. Such variations in shape may at † If advantageous, unequal interval widths may be preferable (see Sec. 4.4).

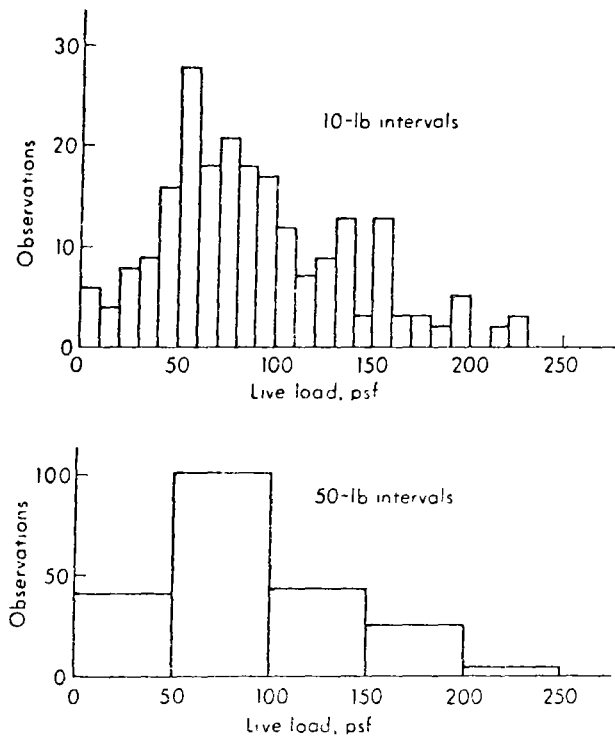


Fig. 1.1.3 Influence of interval size on appearance of histogram (data of Table 1.1.1)

first be disconcerting, but they are indicative of a failure of the set of data to display any sharply defined features, a piece of information which is in itself valuable to the engineer. This failure may be because of the inadequate size of the data set or because of the nature of the phenomenon being observed.

1.2 NUMERICAL SUMMARIES

Central value measures The single most helpful number associated with a set of data is its average value, or arithmetic mean. If the sequence of observed values is denoted x_1, x_2, \dots, x_n , the *sample mean* \bar{x} is simply

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.2.1)$$

Fifteen reinforced-concrete beams built by engineering students to the same specifications and fabricated from the same batch of concrete were tested in flexure. The observed results of first-crack and ultimate loads, recorded to the nearest 50 lb, are presented in Table 1.2.1. Their cumulative frequency distributions are shown in Fig. 1.2.1. (They were

Table 1.2.1 Tests of identical reinforced concrete beams

Beam number	Load at which the first crack was observed, lb	Failure load, lb
1	10,350	10,350
2	8,450	9,300
3	7,200	9,600
4	5,100	10,300
5	6,500	9,400
6	10,600	10,600
7	6,000	10,100
8	6,000	9,900
9	9,500	9,500
10	6,500	10,200
11	9,300	9,300
12	6,000	9,550
13	6,000	9,550
14	5,800	10,500
15	6,500	10,200

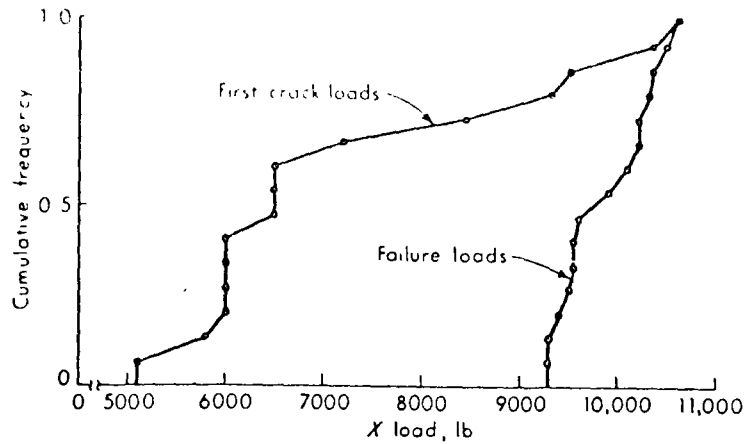


Fig. 1.2.1 Cumulative frequency distributions, beam data

plotted by the method suggested in the footnote on page 7.)† The scatter of the data might be attributed to unrecorded construction and testing differences, inconsistent workmanship, human errors, and inherent material variability as well as observation and measurement errors. The mean value of the failure loads is computed to be

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{148,350}{15} = 9890 \text{ lb}$$

The sample mean is frequently interpreted as a typical value or a central value of the data. If required to give only a single number, one would probably use this sample mean as his "best prediction" of the load at which a nominally identical beam would fail.

Other measures of the central tendency of a data set include *the mode*, the most frequently occurring value in the data set, and *the median*, the middle value in an ordered list (the middle value if n is odd, or the average of the two middle values if n is even) (see Table 1.2.2). The order of observing the values is usually not important, and they may be arranged in any way which is convenient. The median of the failure data is 9900 lb; the mode is not unique, since several values appear twice and none appears more times. These two terms occur commonly in the literature of other fields and consequently should be understood, but they only infrequently prove meaningful in engineering problems.

† The reader would do well to plot a number of different histograms of this failure-load data, both to judge the value of this means for data interpretation in this example and to illustrate the remarks about histogram intervals made in Sec. 1.1

Table 1.2.2 Ordered first-crack and ordered failure-load data

First-crack load	Failure load
5,100	9,300
5,800	9,300
6,000	9,400
6,000	9,500
6,000	9,550
6,000	9,550
6,500	9,600
6,500	9,900
6,500	10,100
7,200	10,200
8,450	10,200
9,300	10,300
9,500	10,350
10,350	10,500
10,600	10,600

Measures of dispersion Given a set of observed data, it is also desirable to be able to summarize in a single number something of the variability of the observations. In the past the measure most frequently occurring in engineering reports was the *range* of the data. This number, which is simply the difference between the maximum and minimum values observed, has the virtue of being easy to calculate, but certain obvious weaknesses have led to its being replaced or supplemented. The range places too much emphasis on the extremes, which are often suspected of experimental error, and neglects the bulk of the data and experimental effort which lies between these extremes. The range is also sensitive to the size of the sample observed, as will be demonstrated in Sec. 3.3.3.

A far more satisfactory measure of dispersion is found in the *sample variance*. It is analogous to the moment of inertia in that it deals with squares of distances from a center of gravity, which is simply the sample mean. The sample variance s^2 is defined here to be

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1.2.2a)$$

To eliminate the dependence on sample size, the squared distances are divided by n to yield an average squared deviation. There are sound reasons favoring division by $n - 1$, as will be shown in Chap. 4, but the

intuitively more satisfactory form will not be abandoned until the reader can appreciate the reasons. Similarly, small computational changes in other definitions given in this chapter may be found desirable in the light of later discussions on the estimation of *moments of random variables* (Secs. 2.4 and 4.1).

Expansion of Eq. (1.2.2a) yields an expression which will be found far more convenient for computation of s^2 :

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \right)$$

But, by Eq. (1.2.1),

$$\sum_{i=1}^n x_i = n\bar{x}$$

Therefore,

$$s^2 = \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \quad (1.2.2b)$$

The positive square root s of the sample variance of the data is termed the *sample standard deviation*. It is analogous to the radius of gyration of a structural cross section; they are both shape- rather than size-dependent parameters. The addition of a constant to all observed values, for example, would alter the sample mean but leave the sample standard deviation unchanged. This number has the same units as the original data, and, next to the mean, it conveys more useful information to the engineer than any other single number which can be computed from the set of data. Roughly speaking, the smaller the standard deviation of the sample, the more clustered about the sample mean is the data and the less frequent are large variations from the average value.

For the beam-failure data the sample variance and sample standard deviation are computed as follows.

$$\begin{aligned} s^2 &= \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \\ &= \frac{1}{15} (1,470,064,450 - 1,467,181,500) \\ &= 192,200 \text{ lb}^2 \\ s &= \sqrt{192,200} = 440 \text{ lb} \end{aligned}$$

(Notice that, owing to the required subtraction of two nearly equal numbers, many significant digits must be carried in the sums to maintain accuracy.) In this example, the sample standard deviation might

be used to compare the variability of the strength of lab beams with that of field-constructed beams.

When comparing the relative dispersion of more than one kind of data, it is convenient to have a dimensionless description such as the commonly quoted *sample coefficient of variation*. This quantity v is defined as the ratio of the sample standard deviation to the sample mean

$$v = \frac{s}{\bar{x}} \quad (1.2.3)$$

The sample coefficient of variation of the beam-failure data is

$$v = \frac{s}{\bar{x}} = \frac{440}{9890} = 0.0446$$

while that of the first observed crack is much larger, being

$$v = \frac{s}{\bar{x}} = \frac{1690}{7310} = 0.233$$

The engineer might interpret the difference in magnitudes of these coefficients as an indication that first-crack loads are "more variable" or more difficult to predict closely than failure loads. Such information is important when appearance as well as strength is a design criterion.

Measure of asymmetry One other numerical summary† of observed data is simply a logical extension of the reasoning leading to the formula for the sample variance. Where the variance was an average second moment about the mean, so the *sample coefficient of skewness* is related to the third moment about the mean. To make the coefficient nondimensional, the moment is divided by the cube of the sample standard deviation.

The coefficient of skewness g_1 provides a measure of the degree of

† A fourth numerical summary, the *coefficient of kurtosis*, may also be employed. Without large sample sizes, however, its use is seldom recommended.

The sample coefficient of kurtosis g_2 is related to the "peakedness" of the histogram

$$g_2 = \frac{(1/n) \sum_{i=1}^n (x_i - \bar{x})^4}{s^4}$$

Traditionally the value of this coefficient is compared to a value of $g_2 = 3.0$, the kurtosis coefficient of a commonly encountered bell-shaped continuous curve which we shall learn to call the *normal curve* (Sec. 3.3.3). For the data here,

$$g_2 = \frac{\frac{1}{15} (836 \times 10^7)}{(440)^4} = 1.61 < 3.0$$

75

asymmetric about the mean of the data

$$g_1 = \frac{(1/n) \sum_{i=1}^n (x_i - \bar{x})^3}{s^3} \quad (1.2.4)$$

The coefficient is positive for histograms skewed to the right (i.e., with longer tails to the right) and negative for those skewed to the left. Zero skewness results from symmetry but does not necessarily imply it.

For the beam-failure data of Table 1.2.1,

$$g_1 = \frac{1/15(120,000,000)}{(440)^3} = 0.103 > 0$$

indicating mild skewness to the right. The implication is that there were fewer but larger deviations to the high side than to the low side of the average value. (The sample coefficient of skewness should be calculated using $x_i - \bar{x}$, since an expansion similar to Eq. (1.2.2b) for the sample variance does not prove useful.)

In this case, if the students reported their experimental results in the form of only three numbers, \bar{x} , s , and g_1 , it would already be sufficient to gain appreciation for the shape of the histogram. The economy in the use of such sample averages to transmit information about data becomes even more obvious as the amount of data increases.

1.3 DATA OBSERVED IN PAIRS

If paired samples of two items of related interest, such as the first-crack load and the failure load of a beam (Table 1.2.1), are available, it is often of interest to investigate the correlation between them. A graphical picture is available in the so-called *scattergram*, which is simply a plot of the observed pairs of values. The scattergram of the reinforced-concrete-beam data is shown in Fig. 1.3.1, where the x_i are values of first-crack loads and the y_i are values of failure loads. There is no clearly defined functional relationship between these observations, even though an engineer might expect larger-than-average values of one load generally to pair with larger-than-average values of the other, and similarly with low values.

A numerical summary of the tendency towards high-high, low-low pairings is provided by the *sample covariance* $s_{X,Y}$, defined by

$$s_{X,Y} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (1.3.1)$$

Clearly, if larger (smaller) than average values of x are frequently paired with large (smaller) than average values of y , most of the terms will be

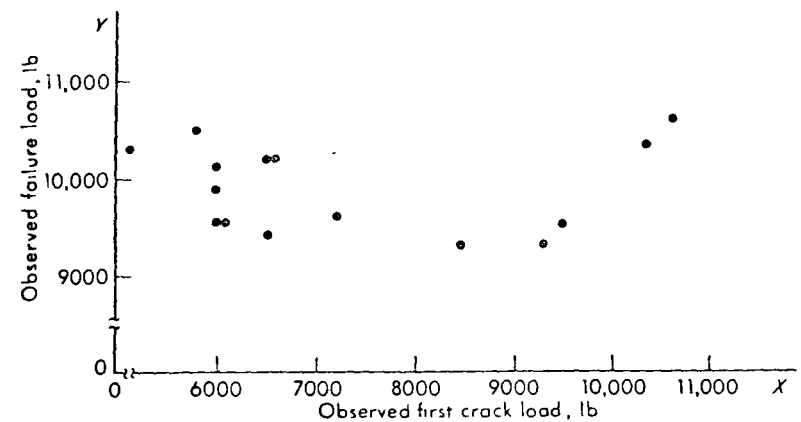


Fig. 1.3.1 Beam-data scattergram, plot shows lack of linear correlation

positive, while small-large pairings will tend to yield negative values of $s_{X,Y}$.

It is common to normalize the sample covariance by the sample standard deviations, denoted now with subscripts, s_X and s_Y . The result is called the *sample correlation coefficient* $r_{X,Y}$:

$$r_{X,Y} = \frac{s_{X,Y}}{s_X s_Y} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_X} \right) \left(\frac{y_i - \bar{y}}{s_Y} \right) \quad (1.3.2)$$

It can be shown that $r_{X,Y}$ is limited in value to $-1 \leq r_{X,Y} \leq 1$ and that the extreme values are obtained if and only if the points in the scattergram lie on a perfectly straight line, that is, only if

$$y_i = a + bx_i$$

the sign of $r_{X,Y}$ depending only on the sign of b . In this case the factors are said to be *perfectly correlated*. For other than perfectly linear relationships $|r_{X,Y}|$ is less than 1, the specific value $r_{X,Y} = 0$ being said to indicate that the x 's and y 's are *uncorrelated*. The x 's and y 's may, in fact, lie on a very well-defined nonlinear curve and hence be closely, perhaps functionally, related (for example, $y_i = bx_i^2$); in this case, the absolute value of the sample correlation coefficient will be less than one. The coefficient is actually a measure of only the *linear* correlation between the factors sampled.

For the beam data of Table 1.2.1, the sample covariance is

$$s_{X,Y} = \frac{1}{n} \sum_{i=1}^n (x_i - 7310)(y_i - 9890) = 94,200 \text{ lb}^2$$

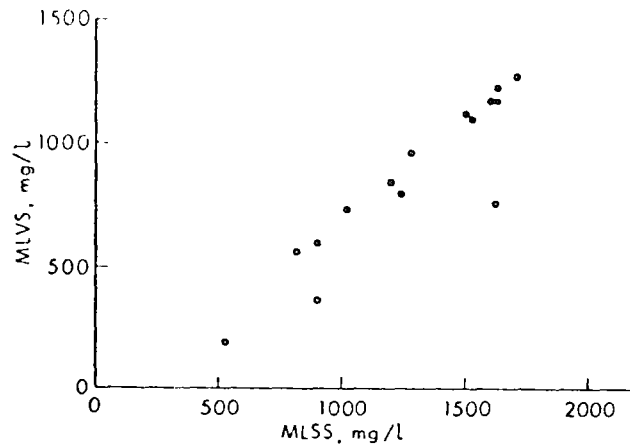


Fig. 13.2 Scattergram of MLVS against MLSS

while the sample correlation coefficient is

$$r_{X,Y} = \frac{s_{XY}}{s_X s_Y} = \frac{94,200}{(1690)(440)} = +0.12$$

The small value of this coefficient summarizes the qualitative conclusion reached by observing the scattergram, that is, that the first-crack loads and failure loads are not closely related. To the engineer who must judge the ultimate strength of a (moment) cracked beam, the implications are that the crack does not necessarily mean that (moment) failure is imminent, and also that he cannot successfully use the first-crack load to help predict the ultimate load better (Sec. 4.3).

Illustration: Large correlation As a second example consider a problem in mixed-liquor analysis. Suspended solids (MLSS) can be readily measured, but volatile solids (MLVS) prove more difficult. It would be convenient if a measure of MLSS could be interpreted in terms of MLVS. A scattergram between MLVS and MLSS is shown in Fig. 13.2. The sample correlation coefficient for the data shown is $r_{X,Y} = 0.92$. This relatively large value of the correlation coefficient follows from the nearly linear relationship between the two factors in the scattergram. It also suggests that the value of the MLVS could be closely estimated given a measurement of MLSS. Methods for carrying out and evaluating this prediction will be investigated in Sec. 4.3.

Discussion In addition to providing efficient summaries of groups of numbers, the data-reduction techniques discussed in this chapter play an important role in the formulation and implementation of probability-based engineering studies. The data in the illustrations and problems in

this chapter might be interpreted by the engineer as repeated observations of a random or *probabilistic* mathematical model that he formulated to aid him in describing and predicting some natural physical phenomenon of concern.

The concepts introduced in Chap. 2 are fundamental to the construction and manipulation of these probabilistic models. In Chap. 4 we shall be concerned with the use made of physical-world observed data, such as that seen in this chapter, in fitting the mathematical models, and in verifying their appropriateness.

The importance of the histogram in choosing and in verifying the mathematical model will become evident in Chaps. 2 to 4. Parameters of the mathematical models which are analogous to the sample mean, sample variance, and sample correlation coefficient will be discussed in Sec. 2.4. The relationship between sample averages and these model parameters will be treated in detail in Chap. 4.

1.4 SUMMARY FOR CHAPTER 1

In Chap. 1 we introduced graphical and numerical ways to reduce sets of observed data. The former include various forms of relative frequency and cumulative frequency diagrams. The latter include measures of central tendency, dispersion, skew, and correlation. The most commonly encountered are the sample mean \bar{x} , the sample standard deviation s , the sample coefficient of variation v , and the sample correlation coefficient r .

REFERENCES

General

The student will find it desirable to refer to other texts to amplify this chapter. In particular, Hald [1952] will prove most useful.

- Bowker, A. H. and G. J. Lieberman [1959] "Engineering Statistics," Prentice-Hall, Inc., Englewood Cliffs, N. J.
- Bryant, E. C. [1960] "Statistical Analysis," McGraw-Hill Book Company, New York.
- Hahn, G. J. and S. S. Shapiro [1967] "Statistical Models in Engineering," John Wiley & Sons, Inc., New York.
- Hald, A. [1952] "Statistical Theory with Engineering Applications," John Wiley & Sons, Inc., New York.
- Neville, A. M. and J. B. Kennedy [1964] "Basic Statistical Methods for Engineers and Scientists," International Textbook Company, Scranton, Pa.
- Wine, R. L. [1964] "Statistics for Scientists and Engineers," Prentice-Hall, Englewood Cliffs, N. J.

Specific text references

- Sturges, H. A. [1926] "The Choice of a Class Interval," *J. Am. Statist. Assoc.*, vol. 21, pp. 65-66.

77

PROBLEMS

11. Ten timber beams were tested on a span of 1 ft by a single concentrated load at mid span. The Douglas fir beams were 2 by 4 in. in nominal section (actual section 1.63 by 3.50 in.). The purpose of the study was to compare ultimate load and load allowable by building code (394 lb), to compare actual and calculated deflection ($E = 1,600,000$ psi), and to determine if a relationship exists between rigidity and ultimate strength.

Specimen	Deflection at working load, in	Ultimate load, lb
1	0.160	1750
2	0.130	2350
3	0.155	2050
4	0.134	2100
5	0.135	1525
6	0.123	2000
7	0.168	1450
8	0.130	2100
9	0.150	1475
10	0.132	1675

Compute the sample mean and variance for each set of data. Construct histograms and relative frequency distributions for each set. Plot the scatter diagram and compute the correlation coefficient. What are your conclusions?

12. Ten bolted timber joints were tested using a $\frac{1}{2}$ -in. bolt placed in double shear through three pieces of 2 by 4 in. (nominal size) Douglas fir. The results of the tests were:

Specimen	Load at yield of $\frac{1}{16}$ in., lb	Load at yield of $\frac{1}{8}$ in., lb	Ultimate load, lb
1	3500	3900	4000
2	3400	3500	3500
3	3100	3600	3800
4	2700	3500	4150
5	2900	3680	4050
6	3400	4250	4500
7	2700	3500	4400
8	3100	3700	4250
9	3100	4100	4550
10	2300	3200	3650

Calculate the sample standard deviations and coefficients of variation. Plot a scatter diagram and compute the sample correlation coefficient between load at yield of $\frac{1}{8}$ in. and ultimate load. Might this simpler, nondestructive test be considered as a prediction of ultimate joint capacity?

13. Jack and anchor forces were measured in a prestressed-concrete lift-slab job for both straight and curved cables. The objective of the study was to obtain estimates of the influence of friction and curvature. The data are given in terms of the ratio of jack to anchor force.

Straight cables	Curved cables
1 6609	2 0464
1 5497	1 7496
1 7190	2 0464
1 6534	1 8487
1 7446	2 1328
1 4954	1 9250
1 6836	2 5022
1 4882	2 0830
2 0811	2 4280
1 9819	2 1375
2 1898	2 6610
2 0061	2 3679
	2 6045
	2 3011
	2 8134
	2 4206

Compute the sample mean and variance of each set of data. Construct histograms, frequency distributions, and cumulative frequency distributions. Assuming that the jack forces are constant, what can be said about the influence of nominal curvature on friction loss?

14. The following values of shear strength (in tons per square foot) were determined by unconfined compression tests of soil from Waukell Creek, California. Plot histograms of this data using four, six, and eight intervals. In the second case consider three different locations of the interval division points. With a few extreme points, as here, it may be advantageous to consider unequal intervals, longer in the right-hand tail.

0.12	0.21	0.36	0.37
0.39	0.46	0.47	0.50
0.50	0.51	0.53	0.58
0.61	0.62	0.77	0.81
0.93	1.05	1.59	1.73

15. Compute the individual sample means, standard deviations, and coefficients of variation for the basement, fourth floor, and ninth floor of the floor-load data in

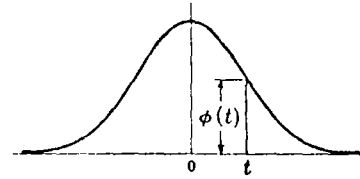
BIBLIOGRAFIA

- [1] A. PAPOULIS. *PROBABILITY, RANDOM VARIABLES, AND STOCHASTIC PROCESSES*. 1965. MC GRAW HILL BOOK COMPANY. NEW YORK.
- [2] L. MAISEL. *PROBABILIDAD Y ESTADISTICA*. 1973. FONDO EDUCATIVO INTERAMERICANO, S.A. MEXICO.
- [3] S. LIPSCHUTZ. *PROBABILITY*. 1968. SCHAUM'S OUTLINE SERIES. MC GRAW-HILL BOOK COMPANY. NEW YORK.
- [4] J.R. BENJAMIN, C.A. CORNELL. *PROBABILITY, STATISTICS, AND DECISION FOR CIVIL ENGINEERS*. 1970. MC GRAW-HILL BOOK COMPANY. NEW YORK.

79

STANDARD NORMAL CURVE ORDINATES

This table gives values $\phi(t)$ of the standard normal distribution ϕ at $t \geq 0$ in steps of 0.01.



t	0	1	2	3	4	5	6	7	8	9
0.0	.3989	.3989	.3989	.3988	.3986	.3984	.3982	.3980	.3977	.3973
0.1	.3970	.3965	.3961	.3956	.3951	.3945	.3939	.3932	.3925	.3918
0.2	.3910	.3902	.3894	.3885	.3876	.3867	.3857	.3847	.3836	.3825
0.3	.3814	.3802	.3790	.3778	.3765	.3752	.3739	.3725	.3712	.3697
0.4	.3683	.3668	.3653	.3637	.3621	.3605	.3589	.3572	.3555	.3538
0.5	.3521	.3503	.3485	.3467	.3448	.3429	.3410	.3391	.3372	.3352
0.6	.3332	.3312	.3292	.3271	.3251	.3230	.3209	.3187	.3166	.3144
0.7	.3123	.3101	.3079	.3056	.3034	.3011	.2989	.2966	.2943	.2920
0.8	.2897	.2874	.2850	.2827	.2803	.2780	.2756	.2732	.2709	.2685
0.9	.2661	.2637	.2613	.2589	.2565	.2541	.2516	.2492	.2468	.2444
1.0	.2420	.2396	.2371	.2347	.2323	.2299	.2275	.2251	.2227	.2203
1.1	.2179	.2155	.2131	.2107	.2083	.2059	.2036	.2012	.1989	.1965
1.2	.1942	.1919	.1895	.1872	.1849	.1826	.1804	.1781	.1758	.1736
1.3	.1714	.1691	.1669	.1647	.1626	.1604	.1582	.1561	.1539	.1518
1.4	.1497	.1476	.1456	.1435	.1415	.1394	.1374	.1354	.1334	.1315
1.5	.1295	.1276	.1257	.1238	.1219	.1200	.1182	.1163	.1145	.1127
1.6	.1109	.1092	.1074	.1057	.1040	.1023	.1006	.0989	.0973	.0957
1.7	.0940	.0925	.0909	.0893	.0878	.0863	.0848	.0833	.0818	.0804
1.8	.0790	.0775	.0761	.0748	.0734	.0721	.0707	.0694	.0681	.0669
1.9	.0656	.0644	.0632	.0620	.0608	.0596	.0584	.0573	.0562	.0551
2.0	.0540	.0529	.0519	.0508	.0498	.0488	.0478	.0468	.0459	.0449
2.1	.0440	.0431	.0422	.0413	.0404	.0396	.0387	.0379	.0371	.0363
2.2	.0355	.0347	.0339	.0332	.0325	.0317	.0310	.0303	.0297	.0290
2.3	.0283	.0277	.0270	.0264	.0258	.0252	.0246	.0241	.0235	.0229
2.4	.0224	.0219	.0213	.0208	.0203	.0198	.0194	.0189	.0184	.0180
2.5	.0175	.0171	.0167	.0163	.0158	.0154	.0151	.0147	.0143	.0139
2.6	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110	.0107
2.7	.0104	.0101	.0099	.0096	.0093	.0091	.0088	.0086	.0084	.0081
2.8	.0079	.0077	.0075	.0073	.0071	.0069	.0067	.0065	.0063	.0061
2.9	.0060	.0058	.0056	.0055	.0053	.0051	.0050	.0048	.0047	.0046
3.0	.0044	.0043	.0042	.0040	.0039	.0038	.0037	.0036	.0035	.0034
3.1	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026	.0025	.0025
3.2	.0024	.0023	.0022	.0022	.0021	.0020	.0020	.0019	.0018	.0018
3.3	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014	.0013	.0013
3.4	.0012	.0012	.0012	.0011	.0011	.0010	.0010	.0010	.0009	.0009
3.5	.0009	.0008	.0008	.0008	.0008	.0007	.0007	.0007	.0007	.0006
3.6	.0006	.0006	.0006	.0005	.0005	.0005	.0005	.0005	.0005	.0004
3.7	.0004	.0004	.0004	.0004	.0004	.0004	.0003	.0003	.0003	.0003
3.8	.0003	.0003	.0003	.0003	.0003	.0002	.0002	.0002	.0002	.0002
3.9	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001

Table 6.1

STANDARD NORMAL CURVE AREAS										
This table gives areas under the standard normal distribution ϕ between 0 and $t \geq 0$ in steps of 0.01.										
t	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

Table 6.2

VALUES OF $e^{-\lambda}$										
λ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$e^{-\lambda}$	1.000	.905	.819	.741	.670	.607	.549	.497	.449	.407
λ	1	2	3	4	5	6	7	8	9	10
$e^{-\lambda}$.368	.135	.0498	.0183	.00674	.00248	.000912	.000335	.000123	.000045

Table 6.3

Solved Problems

BINOMIAL DISTRIBUTION

6.1. Find (i) $b(2; 5, \frac{1}{3})$, (ii) $b(3; 6, \frac{1}{2})$, (iii) $b(3; 4, \frac{1}{4})$.

Here $b(k; n, p) = \binom{n}{k} p^k q^{n-k}$ where $p + q = 1$.

$$(i) \quad b(2; 5, \frac{1}{3}) = \binom{5}{2} (\frac{1}{3})^2 (\frac{2}{3})^3 = \frac{5 \cdot 4}{1 \cdot 2} (\frac{1}{3})^2 (\frac{2}{3})^3 = \frac{80}{243}.$$

$$(ii) \quad b(3; 6, \frac{1}{2}) = \binom{6}{3} (\frac{1}{2})^3 (\frac{1}{2})^3 = \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3} (\frac{1}{2})^3 (\frac{1}{2})^3 = \frac{5}{16}.$$

$$(iii) \quad b(3; 4, \frac{1}{4}) = \binom{4}{3} (\frac{1}{4})^3 (\frac{3}{4}) = \binom{4}{1} (\frac{1}{4})^3 (\frac{3}{4}) = \frac{4}{1} (\frac{1}{4})^3 (\frac{3}{4}) = \frac{3}{64}.$$

6.2. A fair coin is tossed three times. Find the probability P that there will appear (i) three heads, (ii) two heads, (iii) one head, (iv) no heads.

Method 1. We obtain the following equiprobable space of eight elements:

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

(i) Three heads (HHH) occurs only once among the eight sample points; hence $P = \frac{1}{8}$.

(ii) Two heads occurs 3 times (HHT, HTH, and THH); hence $P = \frac{3}{8}$.

(iii) One head occurs 3 times (HTT, THT and TTH); hence $P = \frac{3}{8}$.

(iv) No heads, i.e. three tails (TTT), occurs only once; hence $P = \frac{1}{8}$.

Method 2. Use Theorem 6.1 with $n = 3$ and $p = q = \frac{1}{2}$.

$$(i) \quad \text{Here } k = 3 \text{ and } P = b(3; 3, \frac{1}{2}) = \binom{3}{3} (\frac{1}{2})^3 (\frac{1}{2})^0 = 1 \cdot \frac{1}{8} \cdot 1 = \frac{1}{8}.$$

$$(ii) \quad \text{Here } k = 2 \text{ and } P = b(2; 3, \frac{1}{2}) = \binom{3}{2} (\frac{1}{2})^2 (\frac{1}{2}) = 3 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{3}{8}.$$

$$(iii) \quad \text{Here } k = 1 \text{ and } P = b(1; 3, \frac{1}{2}) = \binom{3}{1} (\frac{1}{2})^1 (\frac{1}{2})^2 = 3 \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{8}.$$

$$(iv) \quad \text{Here } k = 0 \text{ and } P = b(0; 3, \frac{1}{2}) = \binom{3}{0} (\frac{1}{2})^0 (\frac{1}{2})^3 = 1 \cdot 1 \cdot \frac{1}{8} = \frac{1}{8}.$$

6.3. Team A has probability $\frac{2}{3}$ of winning whenever it plays. If A plays 4 games, find the probability that A wins (i) exactly 2 games, (ii) at least 1 game, (iii) more than half of the games.

Here $n = 4$, $p = \frac{2}{3}$ and $q = 1 - p = \frac{1}{3}$.

$$(i) \quad P(2 \text{ wins}) = b(2; 4, \frac{2}{3}) = \binom{4}{2} (\frac{2}{3})^2 (\frac{1}{3})^2 = \frac{8}{27}.$$

NOTAS SOBRE LA INTEGRAL

DE CONVOLUCION

ANTONIO ROLDAN G.

Introducción. - Estas notas pretenden hacer comprender al lector el significado de la integral de convolución, mediante un tratamiento heurístico, desprovisto del rigorismo matemático. También constituye un intento de interiorizarse en el tópico introduciendo ideas de orden implícitas en la teoría de la aproximación.

Aproximación. - En la práctica es frecuente disponer del registro de una forma de onda (función) en un intervalo de tiempo finito, pero no de su expresión analítica. En éstos casos es necesario aproximar la forma de onda usando un conjunto de funciones conocidas y adecuadas para el análisis.

El tipo de funciones es arbitrario formado con un tipo de funciones seleccionadas, poseé una estructura ordenada. Por ejemplo: los polinomios tienen el orden implícito en las potencias. Los conjuntos de funciones senoidales, lo tienen en sus períodos.

(2)

Así, el problema de la aproximación se puede resumir diciendo que si se tiene una forma de onda $f(t)$ en un intervalo de tiempo finito, se desea aproximar la función, en ese mismo intervalo de tiempo, mediante una combinación lineal $f_0(t)$ de funciones convenientemente escogidas.

$$f_0(t) = a_0 g_0(t) + a_1 g_1(t) + \dots + a_n g_n(t)$$

Un método comúnmente empleado para este fin es el de los mínimos cuadrados. Este método nos asegura el error de aproximación más pequeño relativo a la función que se emplea como aproximante.

El sistema de ecuaciones que se debe resolver para una aproximación por mínimos cuadrados es el siguiente

$$\left. \begin{aligned} \int_{t_0}^{t_1} g_0(t)g_0(t)dt + \int_{t_0}^{t_1} g_0(t)g_1(t)dt + \dots + \int_{t_0}^{t_1} g_0(t)g_n(t)dt &= \int_{t_0}^{t_1} g_0(t)f(t)dt \\ \int_{t_0}^{t_1} g_1(t)g_0(t)dt + \int_{t_0}^{t_1} g_1(t)g_1(t)dt + \dots + \int_{t_0}^{t_1} g_1(t)g_n(t)dt &= \int_{t_0}^{t_1} g_1(t)f(t)dt \\ \vdots & \vdots \\ \int_{t_0}^{t_1} g_n(t)g_0(t)dt + \int_{t_0}^{t_1} g_n(t)g_1(t)dt + \dots + \int_{t_0}^{t_1} g_n(t)g_n(t)dt &= \int_{t_0}^{t_1} g_n(t)f(t)dt \end{aligned} \right\}$$

Para ilustrar este tipo de aproximación resolveremos el siguiente problema:

(3)

Problema.- Aproximar la función $f(t) = \cos \frac{\pi t}{2}$ mediante el polinomio $f_2(t) = a_0 + a_1 t + a_2 t^2$, en el intervalo $-1 < t < 1$ (para propósitos de ilustración y comparación se supone conocida la función $f(t)$)

Para este ejemplo $g_0(t) = 1$, $g_1(t) = t$, $g_2(t) = t^2$ y el sistema de ecuaciones por resolver es el siguiente

$$a_0 \int_{-1}^1 dt + a_1 \int_{-1}^1 t dt + a_2 \int_{-1}^1 t^2 dt = \int_{-1}^1 \cos \frac{\pi t}{2} dt$$

$$a_0 \int_{-1}^1 t dt + a_1 \int_{-1}^1 t^2 dt + a_2 \int_{-1}^1 t^3 dt = \int_{-1}^1 t \cos \frac{\pi t}{2} dt$$

$$a_0 \int_{-1}^1 t^2 dt + a_1 \int_{-1}^1 t^3 dt + a_2 \int_{-1}^1 t^4 dt = \int_{-1}^1 t^2 \cos \frac{\pi t}{2} dt.$$

haciendo las integraciones indicadas se obtiene

$$2a_0 + \frac{2}{3} a_2 = \int_{-1}^1 \cos \frac{\pi t}{2} dt = \frac{4}{\pi}$$

$$\frac{2}{3} a_1 = \int_{-1}^1 t \cos \frac{\pi t}{2} dt = 0$$

$$\frac{2}{3} a_0 + \frac{2}{5} a_2 = \int_{-1}^1 t^2 \cos \frac{\pi t}{2} dt = \frac{4}{\pi^3} (\pi^2 - 8)$$

cuando se resuelve este sistema los valores encontrados

son
$$a_0 = \frac{3}{\pi^3} (20 - \pi^2); a_1 = 0; a_2 = \frac{15}{\pi^3} (\pi^2 - 12)$$

o bien
$$a_0 = 0.970; a_1 = 0; a_2 = -1.030$$

y la función aproximante $f_2(t)$ es:

$$f_2(t) = 0.970 - 1.030 t^2$$

el error que se comete en óste tipo de aproximación se calcula con la expresión

$$E^2 = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} [\varepsilon(t)]^2 dt$$

en el que $\varepsilon(t) = f(t) - f_2(t)$

para el ejemplo dado, $E^2 = 0.017$

La figura 1 muestra da gráfica de la solución encontrada

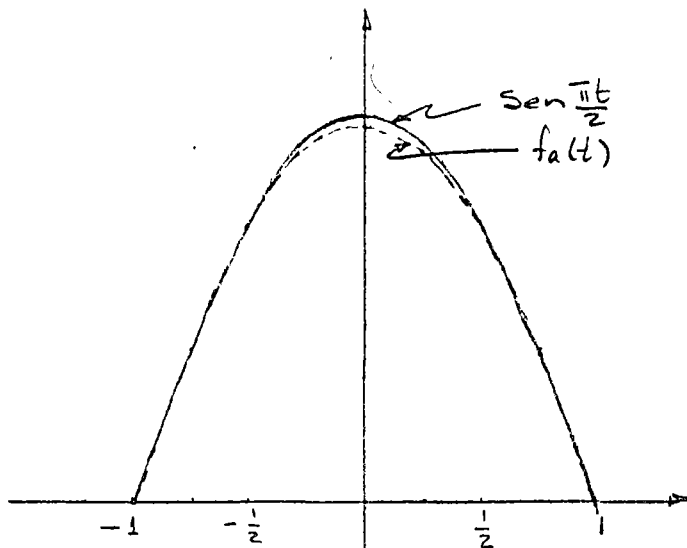


Fig. 1.- Grafica de las funciones $\text{sen } \frac{\pi t}{2}$ y $f_2(t)$

De entre los diferentes tipos de funciones que pueden escogerse existe un tipo de funciones llamadas ortogonales y que poseén la propiedad siguiente:

(5)

$$\int_{t_0}^{t_1} g_k(t) g_j(t) dt = \begin{cases} c & k=j \\ 0 & k \neq j \end{cases} \quad (2)$$

Cuando se emplea este tipo de funciones para aproximación, el sistema de ecuaciones (1) se simplifica notablemente, pudiéndose obtener una expresión sencilla para el cálculo de los coeficientes. Esta expresión es

$$a_k = \frac{\int_{t_0}^{t_1} g_k(t) f(t) dt}{\int_{t_0}^{t_1} [g_k(t)]^2 dt}$$

Cuando las funciones escogidas son complejas entonces la expresión para los coeficientes, que llamaremos ahora D_k , toma la forma

$$D_k = \frac{\int_{t_0}^{t_0+T} f(t) [g_k(t)]^* dt}{\int_{t_0}^{t_0+T} g_k(t) [g_k(t)]^* dt} \quad (3)$$

se emplea la notación D_k para representar cantidades complejas.

Un ejemplo de aproximación cuando las funciones escogidas son exponenciales complejas, es el siguiente.

Hagamos

$$g_k = e^{j \frac{2k\pi t}{T}} \quad \text{y} \quad g_k^* = e^{-j \frac{2k\pi t}{T}}$$

que son funciones periódicas de período T .

La expresión equivalente a (2) para las funciones complejas es:

$$\int_{t_0}^{t_1} g_k(t) g_j^*(t) dt = \begin{cases} C, & k=j \\ 0, & k \neq j \end{cases} \quad (4)$$

y para las funciones escogidas

$$\int_{t_0}^{t_0+T} e^{i \frac{2k\pi t}{T}} e^{-i \frac{2j\pi t}{T}} dt = \begin{cases} T, & k=j \\ 0, & k \neq j \end{cases}$$

substituyendo éste resultado en (3) se obtiene para D_k

$$D_k = \frac{1}{T} \int_{t_0}^{t_0+T} f(t) e^{-i \frac{2k\pi t}{T}} dt$$

la función aproximante $f_a(t)$, es entonces

$$f_a(t) = D_0 + \sum_{k=1}^N \left\{ D_k e^{i \frac{2k\pi t}{T}} + [D_k e^{i \frac{2k\pi t}{T}}]^* \right\} \quad (5)$$

ésta función es real por que

$$D_k e^{i \frac{2k\pi t}{T}} + [D_k e^{i \frac{2k\pi t}{T}}]^* = 2 \operatorname{Re} [D_k e^{i \frac{2k\pi t}{T}}]$$

puesto que D_k es un número complejo, también puede representarse por

$$2D_k = C_k e^{i\varphi_k} = A_k - iB_k$$

y por consiguiente

$$2 \operatorname{Re} [D_k e^{i \frac{2k\pi t}{T}}] = C_k \cos\left(\frac{2k\pi t}{T} + \varphi_k\right)$$

además debe hacerse notar que

$$[D_k e^{i \frac{2k\pi t}{T}}]^* = D_k^* e^{-i \frac{2k\pi t}{T}} = D_k^* e^{i(-k) \frac{2\pi t}{T}}$$

por consiguiente, es posible definir

$$D_k^* = D_{-k}$$

con esta notación la función aproximante se convierte en

$$f_0(t) = D_0 + D_1 e^{i \frac{2\pi t}{T}} + D_2 e^{i \frac{4\pi t}{T}} + \dots + D_k e^{i \frac{2k\pi t}{T}} + \dots + D_N e^{i \frac{2N\pi t}{T}} + D_{-1} e^{-i \frac{2\pi t}{T}} + D_{-2} e^{-i \frac{4\pi t}{T}} + \dots + D_{-k} e^{-i \frac{2k\pi t}{T}} + \dots + D_{-N} e^{-i \frac{2N\pi t}{T}}$$

que puede representarse en forma compacta por

$$\begin{aligned} f_0(t) &= \sum_{-N}^N D_k e^{i \frac{2k\pi t}{T}} \\ &= D_0 + 2 \operatorname{Re} \sum_{-N}^N D_k e^{i \frac{2k\pi t}{T}} \\ &= C_0 + \sum_{k=1}^N C_k \cos\left(\frac{2k\pi t}{T} + \varphi_k\right) \quad (6) \end{aligned}$$

Donde $C_0 = D_0$ y $2D_k = C_k e^{i\varphi_k}$; $k = 0, 1, 2, 3, \dots$

Así cuando se escoge como intervalo de aproximación $-\frac{T}{2} \leq t \leq \frac{T}{2}$, entonces una función cualquiera $f(t)$ puede aproximarse, en ese mismo intervalo, por una función aproximante del tipo (6) en el sentido de los mínimos cuadrados si los coeficientes D_k se escogen de manera que

$$D_k = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) e^{-i \frac{2k\pi t}{T}} dt \quad (7)$$

(8)

cuando los coeficientes D_k se calculan de la manera antes mencionada, $f_a(t)$ se llama serie de Fourier finita.

En este caso el error de aproximación se calcula con la expresión

$$E^2 = F^2 - D_0^2 - 2 \sum_{k=1}^N D_k^2$$

$$\text{Donde } F^2 = \frac{1}{T} \int_{t_0}^{t_0+T} [f(t)]^2 dt.$$

Lo anterior se ilustra con el siguiente ejemplo.

Ejemplo.- Se desea aproximar la función $f(t) = e^{-t}$ mediante la serie $f_a(t) = \sum_{k=-2}^2 D_k e^{i \frac{2k\pi t}{T}}$ en el intervalo $-\frac{T}{2} = -1 < t < \frac{T}{2} = 1$

Solución.- Para este ejemplo $T=2$, por consiguiente

$$D_k = \frac{1}{2} \int_{-1}^1 e^{-t} e^{-ik\pi t} dt$$

integrando y substituyendo límites se obtiene

$$D_k = \frac{1}{2} \frac{e^{+ik\pi} - e^{-(1+ik\pi)}}{1+ik\pi}$$

pero se sabe que $e^{\pm ik\pi} = (-1)^k = 1 (e^{180^\circ k})$ y

$$D_k = \frac{(-1)^k}{1+ik\pi} \cdot \frac{e^{-1} - e^{-1}}{1}$$

$$= 1.18 \frac{(-1)^k}{1+ik\pi}$$

$$D_k = \frac{1.18 (-1)^k}{\sqrt{1+k^2\pi^2}} \angle \tan^{-1}(k\pi)$$

y la función aproximada será

$$f_a(t) = \sum_{k=-2}^2 \frac{(-1)^k 1.18}{1+ik\pi} e^{ik\pi t}$$

(9)

o bien

$$f_0(t) = 1.18 + 2.36 \left[\frac{1}{\sqrt{1+\pi^2}} \cos(\pi t + 180^\circ + \tan^{-1} \pi) + \frac{1}{\sqrt{1+4\pi^2}} \cos(2\pi t + \tan^{-1} 2\pi) \right]$$

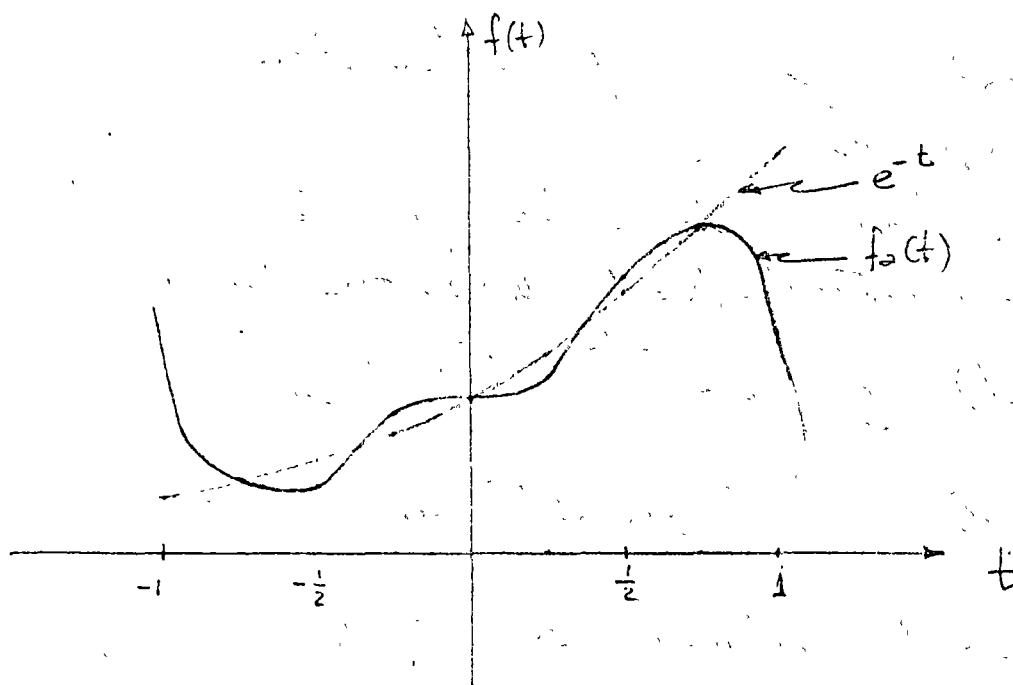
y el error medio cuadrático será

$$\begin{aligned} F^2 &= \frac{1}{2} \int_{-1}^1 e^{-2t} dt \\ &= \frac{1}{2} \frac{e^{-2} - e^{-2}}{2} = 1.81 \end{aligned}$$

$$y, \bar{F}_0^2 = D_0^2 + 2D_1^2 + 2D_2^2 = 1.68$$

$$\therefore E^2 = 0.13$$

$$\text{o en porcentaje } \frac{E^2}{F^2} \times 100 \% = 7.1 \%$$



Se puede ver de (9) que $f_0(t)$ es una función periódica para toda t , puesto que:

$$\sum_{-N}^N D_k e^{i \frac{2k\pi(t+T)}{T}} = \sum_{-N}^N D_k e^{i \frac{2k\pi t}{T}} e^{i 2k\pi} = \sum_{-N}^N D_k e^{i \frac{2k\pi t}{T}}$$

consecuente, si la función $f(t)$ es periódica, con período T y está se aproxima por

$$f_0(t) = \sum_{-N}^N D_k e^{i \frac{2k\pi t}{T}}$$

entonces, ésta aproximación ya no se restringe al intervalo $t_0 < t < t_0 + T$ sino que ahora es válida para todos los valores de t .

Si en lugar de usar una serie de Fourier finita se emplea una infinita, entonces se observa que todas las frecuencias son múltiplos enteros de la frecuencia $\frac{1}{T}$.

La frecuencia cuyos múltiplos enteros son las frecuencias de las componentes, se conoce como frecuencia fundamental; las demás se conocen como armónicas.

Lo expuesto anteriormente, puede formalizarse de la manera siguiente:

Cualquier función periódica $f(t)$ y periodo T integrable en cualquier intervalo, puede representarse por la serie

$$f_a(t) = \sum_{-\infty}^{\infty} D_k e^{i \frac{2\pi k t}{T}} \quad (8) \text{ donde}$$

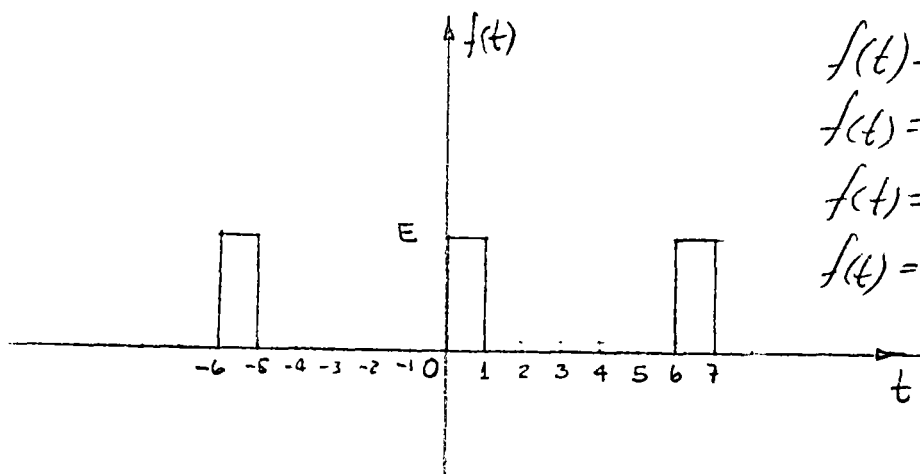
$$D_k = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) e^{-i \frac{2\pi k t}{T}} dt. \quad (9)$$

Si $f(t)$ no solo es integrable, sino también lo es $[f(t)]^2$, entonces la serie (8) converge y además converge al valor $f(t)$ en todos los puntos donde $f(t)$ es continua. Si en un punto $t = t_1$ la función $f(t)$ no es continua, es decir, $f(t^+) \neq f(t^-)$ entonces en ese punto, la serie converge al valor medio aritmético $\frac{1}{2} [f(t_1^+) + f(t_1^-)]$

como ilustración de lo anterior, daremos el siguiente:
Ejemplo.- Encontrar la serie de Fourier de la función

(11)

$f(t)$ mostrada en la figura y definida por



$$f(t) = 0 \quad -3 < t < 0$$

$$f(t) = E \quad 0 < t < 1$$

$$f(t) = 0 \quad 1 < t < 3$$

$$f(t) = f(t+6)$$

Solución.-

$$D_k = \frac{1}{6} \left[\int_{-3}^0 0 \cdot e^{-i \frac{2k\pi t}{6}} dt + \int_0^1 E e^{-i \frac{2k\pi t}{6}} dt + \int_1^3 0 \cdot e^{-i \frac{2k\pi t}{6}} dt \right]$$

$$= \frac{E}{6} \int_0^1 e^{-i \frac{2k\pi t}{6}} dt$$

integrando la ecuación anterior se obtiene

$$D_k = \frac{E}{6} \frac{\mathcal{L}^{-i \frac{2k\pi t}{6}}}{-\frac{ik\pi}{3}} \Big|_0^1 = E \frac{\mathcal{L}^{-i \frac{2k\pi t}{3}} - 1}{-i3k\pi}$$

$$= E \frac{\mathcal{L}^{-i \frac{k\pi}{6}}}{k\pi} \frac{\mathcal{L}^{-i \frac{k\pi}{6}} - \mathcal{L}^{-i \frac{k\pi}{6}}}{-3i} = \frac{E}{6} \frac{\text{Sen} \frac{k\pi}{6}}{\frac{k\pi}{6}} \mathcal{L}^{-i \frac{k\pi}{6}} \quad k \neq 0$$

y para $k=0$ $D_0 = \frac{1}{6} \int_0^1 E dt = \frac{E}{6}$

Finalmente

$$f(t) = \frac{1}{6} \left[1 + 2 \sum_{k=1}^{\infty} \frac{\text{Sen} \frac{k\pi}{6}}{\frac{k\pi}{6}} \cos \left(k \frac{\pi t}{3} - k \frac{\pi}{6} \right) \right]$$

Por consiguiente $C_0 = \frac{1}{6}$, $C_1 = \frac{1}{3} \left(\frac{\text{Sen} \frac{\pi}{6}}{\frac{\pi}{6}} \right)$, $C_2 = \frac{1}{3} \left(\frac{\text{Sen} \frac{2\pi}{6}}{\frac{2\pi}{6}} \right)$,

$C_k = \frac{1}{3} \left(\frac{\text{Sen} \frac{k\pi}{6}}{\frac{k\pi}{6}} \right)$; $\varphi_1 = -\frac{\pi}{6}$, $\varphi_2 = -\frac{2\pi}{6}$, ..., $\varphi_k = -\frac{k\pi}{6}$

Como es evidente, el ordenamiento introducido es en las frecuencias y en las fases φ_k (Ver figura)

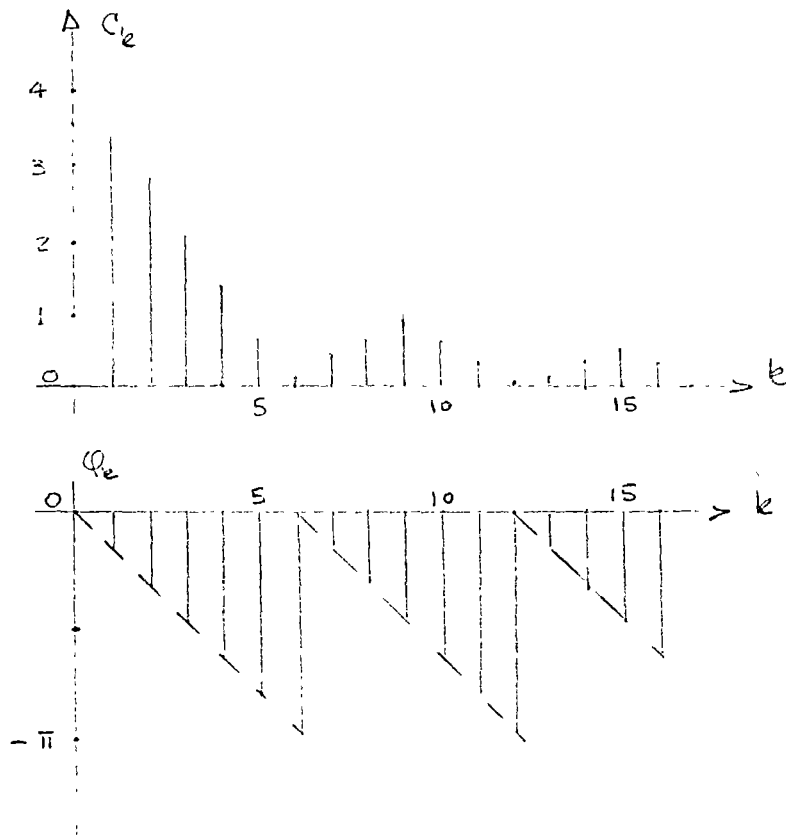


Fig. 2 Graficas de c_k y φ_k respectivamente.

Como puede verse de todo lo anterior, el concepto de orden es de gran importancia, porque es posible describir cuantitativamente a una función $f(t)$ en términos de cantidades conocidas y manejables. Además introducido un ordenamiento, se pueden ya definir operaciones matemáticas para resolver problemas específicos.

Una de éstas operaciones (quizás la más importante) es el producto de coeficientes D_k de dos funciones diferentes y que se ilustrará con un ejemplo sencillo.

Ejemplo. - Supongamos una función $f_1(t)$ dada por su serie

de Fourier

$$f_1(t) = 100(\cos t + \cos 3t + \cos 5t); \text{ los coeficientes}$$

De son

$$C_k = C_k e^{i\varphi_k} = 100; \quad k=1, 3, 5$$

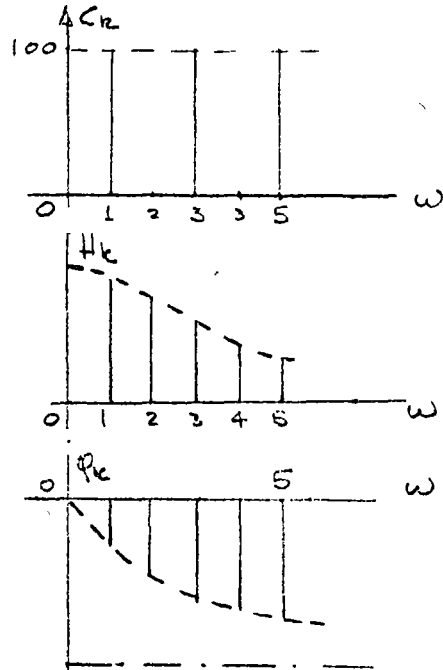
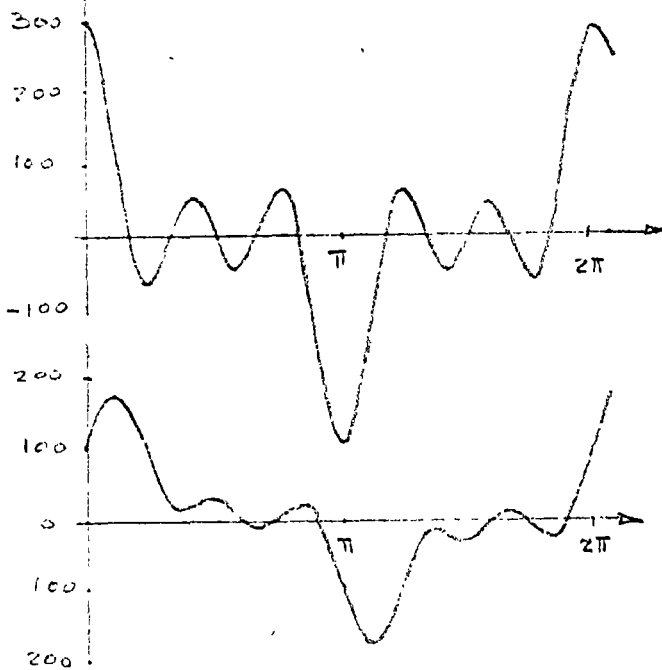
y $\varphi_k = 0$ para toda k

$$\text{y otra } f_2(t) = 1 + \sum_{k=1}^{\infty} \frac{1}{\sqrt{(\frac{k}{2})^2 + 1}} \cos(kt - \tan^{-1} \frac{k}{2})$$

con coeficientes

$$H_k = H_k e^{i\varphi_k}; \quad H_0 = 1, \quad H_1 = \frac{1}{\sqrt{\frac{5}{4}}}, \quad H_2 = \frac{1}{\sqrt{2}} + \dots$$

$$\varphi_0 = 0; \quad \varphi_1 = \tan^{-1} \frac{1}{2}, \quad \varphi_2 = \tan^{-1} 1, \quad \varphi_3 = \tan^{-1} \frac{3}{2}$$



El producto $H_k C_k = C_k H_k e^{i\varphi_k} e^{i\varphi_k} = C_k H_k e^{i(\varphi_k + \varphi_k)}$
 es el k -ésimo coeficiente de Fourier de la función $f(t)$ re-
 sultante. Por consiguiente

$$f(t) = 100 \left[\frac{1}{\sqrt{\frac{5}{4}}} \cos(t - \tan^{-1} \frac{1}{2}) + \frac{1}{\sqrt{\frac{13}{4}}} \cos(3t - \tan^{-1} \frac{3}{2}) + \frac{1}{\sqrt{\frac{29}{4}}} \cos(5t - \tan^{-1} \frac{5}{2}) \right]$$

Es obvio que sería muy difícil de las gráficas de $f_1(t)$ y $f_2(t)$ imaginarse como interaccionan para producir $f(t)$. Sin embargo en el dominio ordenado de las frecuencias, esto no solo es sencillo sino que puede cuantificarse.

En el ejemplo anterior se puede decir que la función resultante $f(t)$ se obtuvo atenuando las altas frecuencias y desplazando cada componente ϕ_k radianes del origen.

La sencillez con la que se opera en el dominio de las frecuencias, impulsa a extender este concepto a funciones no periódicas, que son las que se presentan con mayor frecuencia en la naturaleza. Para el propósito de estas notas se comenzará por considerar que una función ^{no} periódica $f(t)$ es el límite de una función periódica cuando el período tiende al infinito. Esta consideración nos permite interpretar en (8), la suma como integral, obteniéndose expresiones equivalentes a D_k y $f(t)$, para funciones no periódicas. Estas expresiones son:

$$F(i\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \quad (10) \quad \text{para } D_k$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\omega) e^{i\omega t} d\omega \quad (11) \quad \text{para } f(t)$$

A $F(i\omega)$ se le llama al espectro complejo de Fourier y es una función continua, a diferencia de D_k que es una función discreta.

A las expresiones (10) y (11) se les conoce también como las transformada directa e inversa de Fourier

Siendo $F(i\omega)$ una cantidad compleja también se le puede representar por

$$F(i\omega) = F(\omega) e^{i\varphi(\omega)}$$

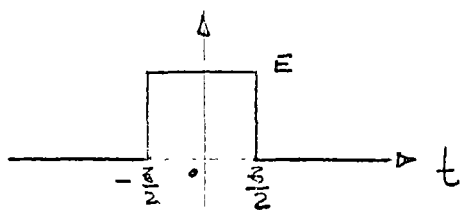
donde $\varphi(\omega)$ es también una función continua.

Las cantidades $F(\omega)$ y $\varphi(\omega)$ se les conoce como el espectro de amplitudes y de fases de $f(t)$.

Un ejemplo de lo anterior es el siguiente.

Se desea encontrar el espectro de amplitudes y el espectro de fases de la función.

$$f(t) = \begin{cases} 0 & -\infty < t < -\frac{\tau}{2} \\ E & -\frac{\tau}{2} < t < \frac{\tau}{2} \\ 0 & \frac{\tau}{2} < t < \infty \end{cases}$$



Solución. -

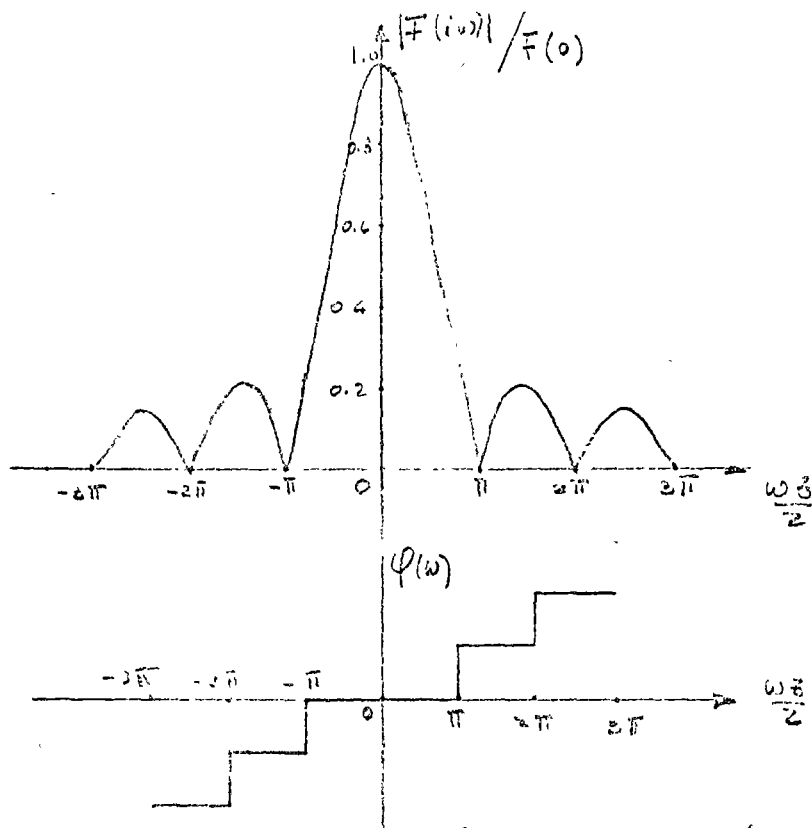
$$F(i\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

$$= E \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} e^{-i\omega t} dt$$

$$= E \frac{e^{-i\omega t}}{-i\omega} \Big|_{-\frac{\tau}{2}}^{\frac{\tau}{2}}$$

$$= \frac{2E}{\omega} \text{Sen} \frac{\tau}{2} \omega$$

o bien $F(i\omega) = E \tau \frac{\text{Sen} \frac{\tau}{2} \omega}{\tau}$



La operación de producto en el dominio de las frecuencias definida para funciones periódicas, también puede definirse para funciones no periódicas. Así, que si $H(i\omega)$ es el espectro complejo de una función $h(t)$ y $F(i\omega)$ es el espectro complejo de $f(t)$, entonces la operación de producto nos dará

$$G(i\omega) = H(i\omega)F(i\omega) \quad (12)$$

para saber cual es el resultado de esta operación en el dominio del tiempo hay que tomar la transformada inversa de (12)

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(i\omega) e^{i\omega t} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(i\omega) F(i\omega) e^{i\omega t} d\omega$$

en la expresión anterior se puede substituir $H(i\omega)$ por su definición dada por (10), quedando

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\omega) \int_{-\infty}^{\infty} h(z) e^{-i\omega z} dz e^{i\omega t} d\omega$$

(11)

El tipo de funciones considerado aquí, permite intercambiar el orden de integración. Esto produce

$$g(t) = \int_{-\infty}^{\infty} h(z) \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\omega) e^{i\omega(t-z)} dt \right] dz$$

En la expresión anterior reconocemos que

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\omega) e^{i\omega(t-z)} dt = f(t-z)$$

con lo que

$$\boxed{g(t) = \int_{-\infty}^{\infty} h(z) f(t-z) dz}$$

que se conoce como la integral de convolución

Por consiguiente, podemos decir que la convolución de dos funciones de tiempo, nos da la función resultante del producto de sus espectros complejos en el dominio de las frecuencias.

Ilustremos lo anterior con el ejemplo siguiente

Sean $h(t) = e^{-t} u(t)$ y $f(t) = e^{-2t} u(t)$ dos funciones de tiempo. Calcular su convolución.

Solución.

$$g(t) = \int_{-\infty}^{\infty} e^{-z} u(z) e^{-2(t-z)} u(t-z) dz$$

en el integrando de la función anterior se ve el factor $u(z)u(t-z)$, y puesto que $u(z) = 0$ para $z < 0$ y $u(t-z) = 0$ para $z > t$, el integrando se convierte en

$$\text{Integrando} = \begin{cases} e^{-z} e^{z(t-z)} & 0 < z < t \\ 0 & \text{para toda otra } z. \end{cases}$$

De consiguiente

$$\begin{aligned} f(t) &= \left[\int_0^t e^{-z} e^{-z(t-z)} dz \right] u(t) = e^{-zt} \left[\int_0^t e^z dz \right] u(t) \\ &= e^{-zt} (e^z) \Big|_{z=0}^{z=t} u(t) = e^{-zt} (e^t - 1) u(t) \end{aligned}$$

$$\underline{f(t) = (e^{-t} - e^{-zt}) u(t)}$$

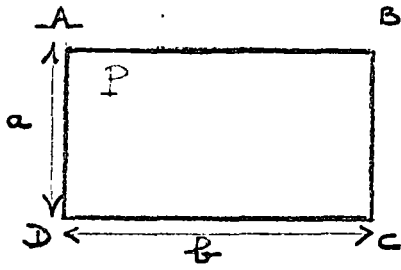
Como puede verse, esta operación nos ahorra el tener que conocer los espectros complejos, hacer el producto y calcular la transformada inversa.

Obramente que si se desea definir la operación en forma precisa, ésta se hace en el dominio de las frecuencias, como se vio con anterioridad.

PRACTICAS DE FUNCIONES AUXILIARES

Sea P un panel rectangular de ley media desconocida Z .

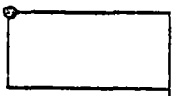
Se supone que P se coloca dentro de una regionalización isotrópica de semi-variograma γ



El problema es de calcular las varianzas de estimación de P en los casos siguientes:



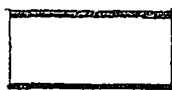
Panel reconocido por un sondeo central



Panel reconocido por uno de sus extremos



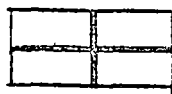
Panel reconocido por sus 4 extremos



Panel reconocido por 2 galerías a lo largo de los lados AB y CD

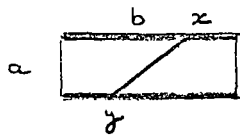


Panel reconocido por su periferia

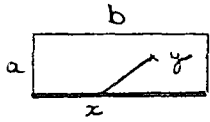


Panel reconocido por 2 galerías perpendiculares que se cruzan en el punto medio.

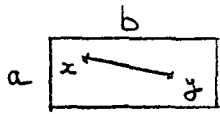
Se usarán las funciones auxiliares siguientes



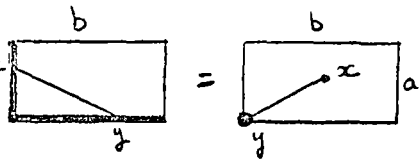
$\chi_b(a) =$ valor promedio de $f(x-y)$ cuando x e y describen independientemente los 2 lados paralelos de longitud \bigcirc del rectángulo $a \times b$.



$\chi_b(a) =$ valor promedio de $f(x-y)$, cuando x describe b e, independientemente de x , y describe el rectángulo $a \times b$



$F(a,b) = F(b,a) =$ valor promedio de $f(x-y)$ cuando x e y describen independientemente uno de otro el rectángulo $a \times b$



$Q(a,b) =$ valor promedio de $f(x-y)$ cuando x describe un lado de longitud a e y el lado de longitud b , o también, cuando x describe el rectángulo $a \times b$ e y queda en cualquier extremo.

Las varianzas que se buscan se expresarán con la ayuda de estas funciones auxiliares.

Nota

Propómosnos de estudiar cada $\bar{f}(F, F')$, valor promedio de $f(x-y)$ cuando x e y describen independientemente uno de otro los conjuntos F y F' , de la manera siguiente

$$\bar{f}(F, F') = \frac{1}{FF'} \int_F \int_{F'} f(x-y) dx dy$$

$$FF' \bar{f}(F, F') = \int_F \int_{F'} f(x-y) dx dy = S(F, F')$$

La linealidad de $S(F, F')$ permite de descomponerla en elementos más sencillos y fáciles de calcular.

SOLUCION

Usaremos la fórmula de la varianza de estimación

$$\sigma_{Est}^2 (F \rightarrow F') = 2 \bar{\gamma} (F, F') - \bar{\gamma} (F, F) - \bar{\gamma} (F', F')$$

- ① F' = sondeo central, sea un punto
 F = rectangulo $a \times b$.

$$\bar{\gamma} (F, F') = 2 Q(a/2, b/2)$$

$$\bar{\gamma} (F, F) = F(a, b)$$

$$\bar{\gamma} (F', F') = 0$$

$$\sigma_1^2 = 2 Q(a/2, b/2) - F(a, b)$$

- ② F' = sondeo
 F = rectangulo $a \times b$

$$\bar{\gamma} (F, F') = Q(a, b)$$

$$\bar{\gamma} (F, F) = F(a, b)$$

$$\bar{\gamma} (F', F') = 0$$

$$\sigma_2^2 = 2 Q(a, b) - F(a, b)$$

- ③ F' = 4 sondeos sea 4 puntos
 F = rectangulo $a \times b$

$$\bar{\gamma} (F, F') = \frac{1}{4 \times ab} \sum_{i=1}^4 \int_F \gamma(x_i, y_i) dy_i \quad \left. \vphantom{\bar{\gamma} (F, F')} \right\} \bar{\gamma} (F, F') = Q(a, b)$$

$$\forall x_i: \frac{1}{ab} \int_F \gamma(x_i, y_i) dy_i = Q(a, b)$$

$$\bar{\gamma} (F, F) = F(a, b)$$

$$\bar{\gamma} (F', F') = \frac{1}{16} \sum_{i=1}^4 \sum_{j=1}^4 \gamma(x_i, x_j) = \frac{1}{16} \times 4 \left[\gamma(0) + \gamma(b) + \gamma(a) + \gamma(\sqrt{a^2 + b^2}) \right] = \frac{1}{4} \left[\gamma(b) + \gamma(a) + \gamma(\sqrt{a^2 + b^2}) \right]$$

$$\sigma_3^2 = 2 Q(a, b) - F(a, b) - \frac{1}{4} \left[\gamma(b) + \gamma(a) + \gamma(\sqrt{a^2 + b^2}) \right]$$

- ④ $F' =$ los 2 lados de longitud b
 $F =$ el rectángulo $a \times b$

$$\bar{Y}(F, F') = \frac{1}{(AB+DC)F} \int_{AB+DC} \int_F \gamma(x-y) dx dy$$

$$(AB+DC)F \bar{Y}(F, F') = \int_{AB} \int_F \gamma(x-y) dx dy + \int_{DC} \int_F \gamma(x-y) dx dy$$

$$\int_{AB} \int_F \gamma(x-y) dx dy = AB \times F \chi_b(a)$$

$$\int_{DC} \int_F \gamma(x-y) dx dy = DC \times F \chi_b(a)$$

entonces
$$\bar{Y}(F, F') = \frac{AB \times F \times \chi_b(a) + DC \times F \times \chi_b(a)}{(AB+DC)F} = \frac{2b \chi_b(a)}{2b} = \chi_b(a)$$

$$\bar{Y}(F, F) = F(a, b)$$

$$\bar{Y}(F', F') = \frac{1}{(AB+CD)^2} \int_{AB+CD} \int_{AB+CD} \gamma(x-y) dx dy$$

$$(AB+CD)^2 \bar{Y}(F', F') = \int_{AB} \int_{AB} + 2 \int_{AB} \int_{CD} + \int_{CD} \int_{CD} = 2 \int_{AB} \int_{AB} + 2 \int_{AB} \int_{CD}$$

$$= 2b^2 F(0, b) + 2b^2 \chi_b(a)$$

$$\bar{Y}(F', F') = \frac{2b^2 F(0, b) + 2b^2 \chi_b(a)}{4b^2} = \frac{1}{2} (F(0, b) + \chi_b(a))$$

$$\sigma_4^2 = 2 \chi_b(a) - F(a, b) - \frac{1}{2} [F(0, b) - \chi_b(a)]$$

- ⑤ $F' =$ los 4 lados del rectángulo
 $F =$ el rectángulo $a \times b$

$$\bar{Y}(F, F') = \frac{1}{(2AB+2BC)F} \int_{(2AB+2BC)} \int_F \gamma(x-y) dx dy$$

$$2(a+b)F \bar{Y}(F, F') = 2 \int_{AB} \int_F \gamma(x-y) dx dy + 2 \int_{BC} \int_F \gamma(x-y) dx dy$$

$$- 2b \times F \chi_b(a) + 2a F \chi_a(b)$$

$$\bar{Y}(F, F') = \frac{b \chi_b(a) + a \chi_a(b)}{(a+b)}$$

$$\bar{Y}(F, F) = F(a, b)$$

$$\bar{Y}(F', F') = \frac{1}{(AB+BC+CD+DA)^2} \int_{AB+BC+CD+DA} \int_{AB+BC+CD+DA} \gamma(x-y) dx dy$$

que se puede descomponer en 16 términos

$$4(a+b)^2 \bar{y}(F', F') = 2 \iint_{AB|AB} + 2 \iint_{BC|BC} + 8 \iint_{AB|BC} + 2 \iint_{AB|bc} + 2 \iint_{AD|BC}$$

$$= 2b^2 F(0,b) + 2a^2 F(a,0) + 8Q(a,b) + 2\gamma_b(a) + 2\gamma_a(b)$$

$$\bar{y}(F', F') = \frac{b^2 F(0,b) + a^2 F(a,0) + 4Q(a,b) + 2\gamma_b(a) + 2\gamma_a(b)}{2(a+b)^2}$$

$$\sigma_5^2 = 2 \frac{b\chi_b(a) + a\chi_a(b)}{a+b} - F(a,b) - \frac{b^2 F(0,b) + a^2 F(a,0) + 4Q(a,b) + 2\gamma_b(a) + 2\gamma_a(b)}{2(a+b)^2}$$

⑥ $F' =$ los 2 segmentos medios
 $F =$ rectángulo $a \times b$

$$\bar{y}(F, F') = \frac{1}{(l_a + l_b)F} \int_{l_a}^{l_b} \int_F \gamma(x-y) dx dy$$

$$(l_a + l_b)F \bar{y}(F, F') = \int_{l_a}^F \int_F \gamma(x-y) dx dy + \int_{l_b}^F \int_F \gamma(x-y) dx dy$$

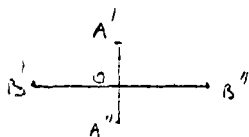
$$= aF\chi_a(b/2) + bF\chi_b(a/2)$$

$$\bar{y}(F, F') = \frac{a\chi_a(b/2) + b\chi_b(a/2)}{a+b}$$

$$\bar{y}(F, F) = F(a, b)$$

$$\bar{y}(F', F') = \frac{1}{(l_a + l_b)^2} \int_{l_a+l_b} \int_{l_a+l_b} \gamma(x-y) dx dy$$

$$(l_a + l_b)^2 \bar{y}(F', F') = \int_{l_a}^{l_a} \int_{l_a}^{l_a} + \int_{l_b}^{l_b} \int_{l_b}^{l_b} + 2 \int_{l_a}^{l_b} \int_{l_a}^{l_b} = a^2 F(0,a) + b^2 F(0,b) + 2 \int_{l_a}^{l_b} \int_{l_a}^{l_b}$$



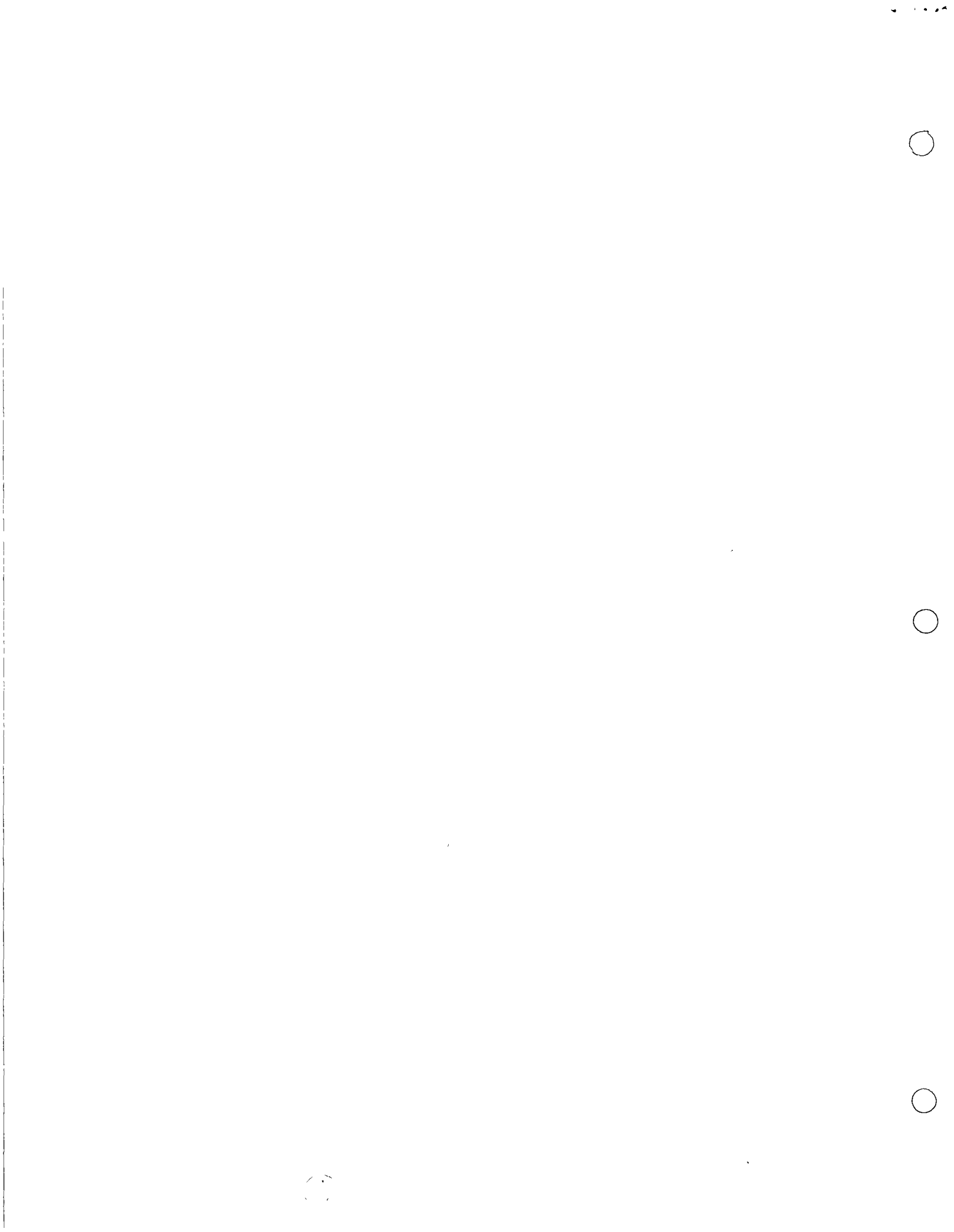
$$\int_{l_a}^{l_b} \int_{l_a}^{l_b} = Q(a/2, b/2)$$

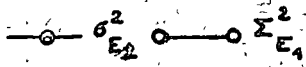
$$\int_{0A'+0A''} \int_{0B'+0B''} = \int_{0A'} \int_{0B'} + \int_{0A'} \int_{0B''} + \int_{0A''} \int_{0B'} + \int_{0A''} \int_{0B''}$$

$$= [a/2 \cdot b/2 Q(a/2, b/2)] \times 4 = ab Q(a/2, b/2)$$

$$\bar{y}(F', F') = \frac{a^2 F(0,a) + b^2 F(0,b) + 2ab Q(a/2, b/2)}{(a+b)^2}$$

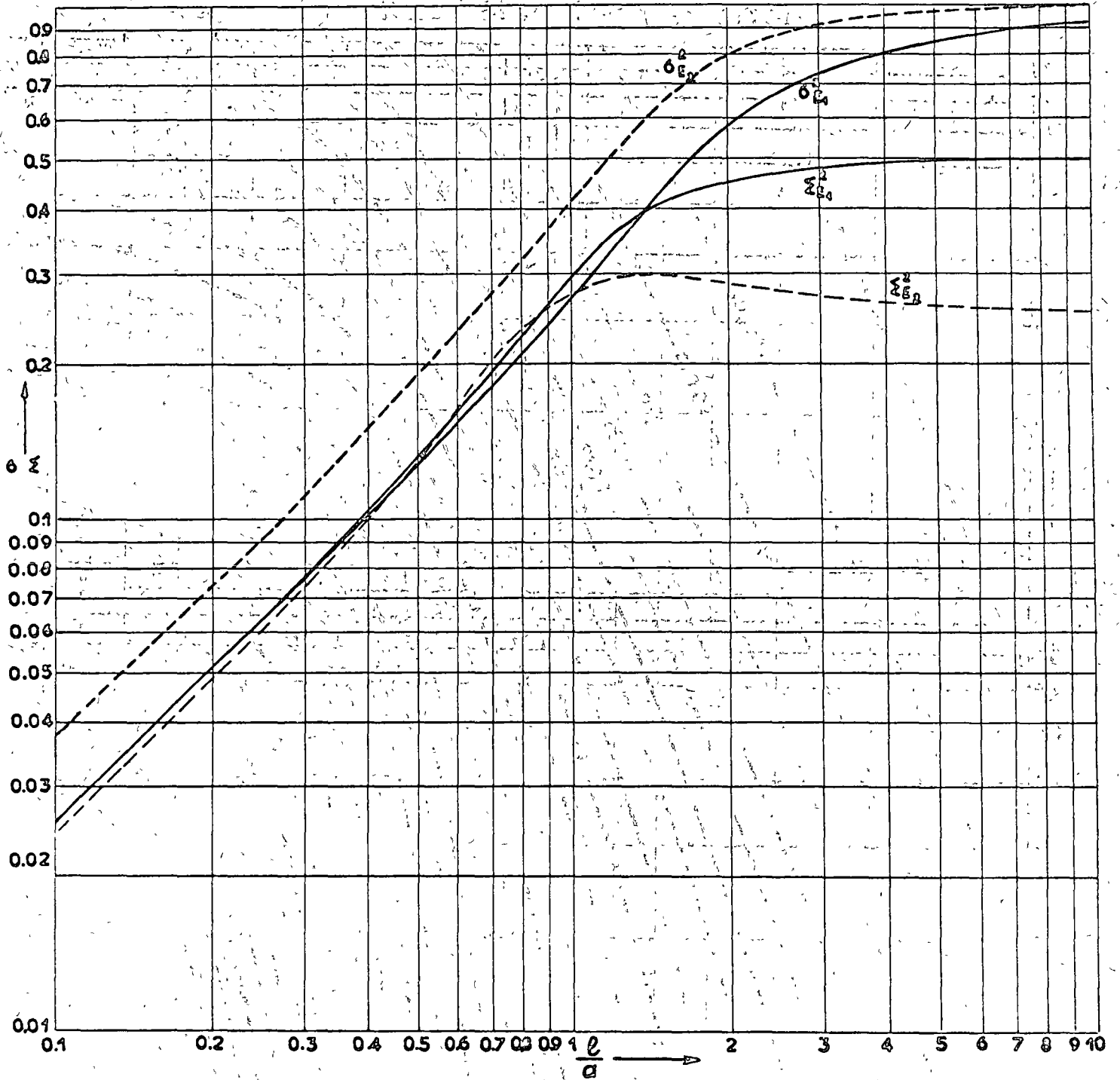
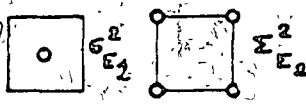
$$\sigma_5^2 = 2 \frac{a\chi_a(b/2) + b\chi_b(a/2)}{a+b} - F(a,b) - \frac{a^2 F(0,a) + b^2 F(0,b) + 2ab Q(a/2, b/2)}{(a+b)^2}$$

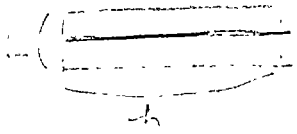




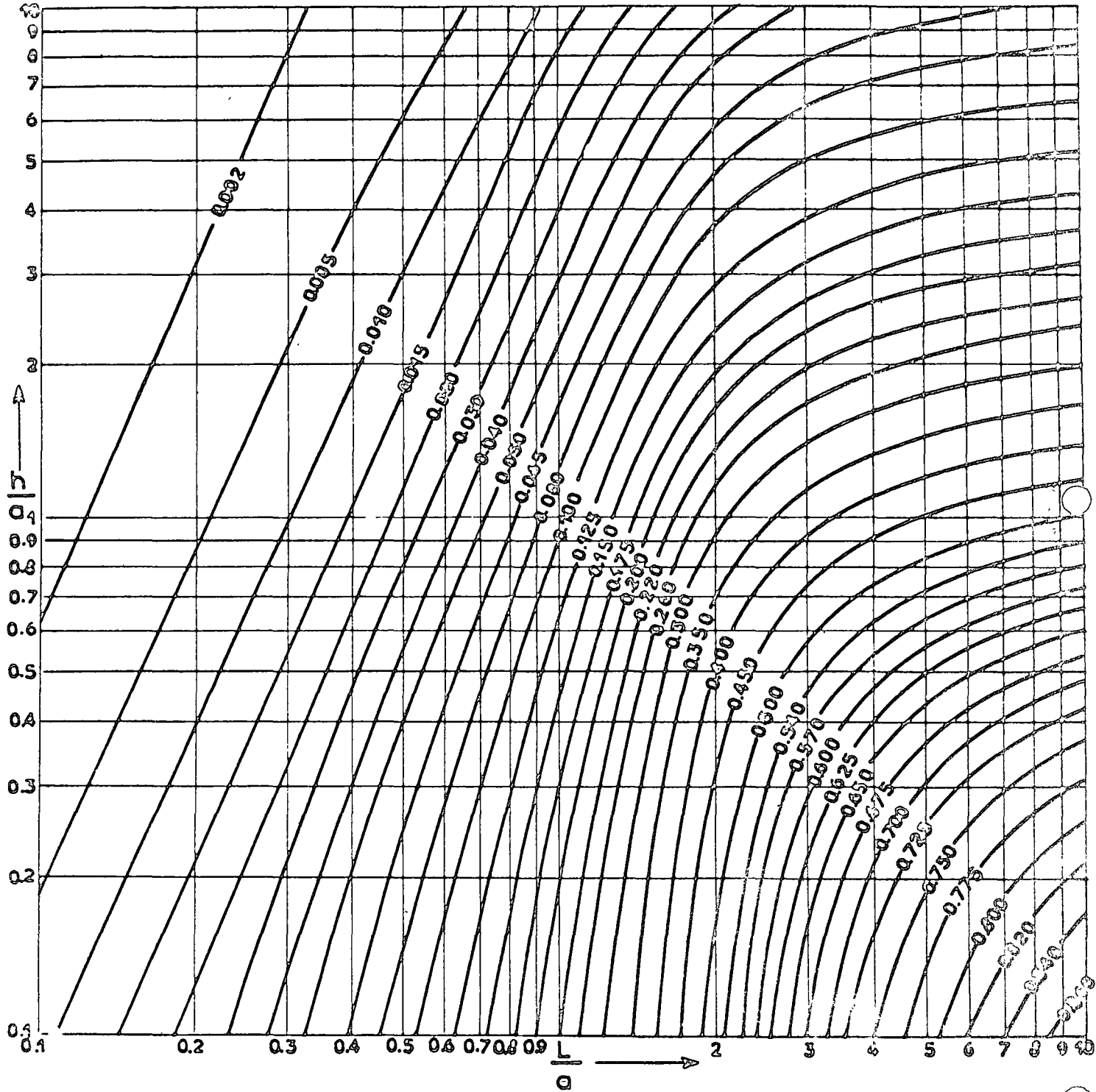
ABAUQUE 8 schéma sphérique.

variance d'extension diverses.





variance d'extension du tracage median h
dans le rectangle L x h



Presentación audio visual de la Geoestadística

Minerales

2 conferencias 2 x 2 h

Problemas tipos de Reconocimiento minero

Subvención de las técnicas utilizadas: 3 escuelas

El lenguaje de la Geoestadística.

1^a conferencia

Problemas tipos y Geoestadística

- Reconocimiento geológico

- Análisis estructural

2^a conferencia

- Estimación de recursos - Kriging

- Estimación de reservas - Selección

- Estudios de dispersiones - Simulaciones

- Correcciones

Presentación de la Geoestadística.

Curso 1.

Primitivamente la Geoestadística estaba definida como la aplicación práctica de la Teoría de las Variables regionalizadas al reconocimiento y a la estimación de los yacimientos mineros. El carácter muy general del formalismo teórico, y las numerosas aplicaciones extra-mineras ya realizadas (gravi-metría - hidrogeología - reconocimiento forestal - ecología ...) nos han conducido a redefinir la Geoestadística como la Práctica de la Teoría de las Variables Regionalizadas.

Esta práctica deriva de un conocimiento seguro de la teoría (que, insistamos sobre este punto, es completamente suficiente con un mínimo de conocimientos estadísticos clásicos), pero necesita, sobretudo:

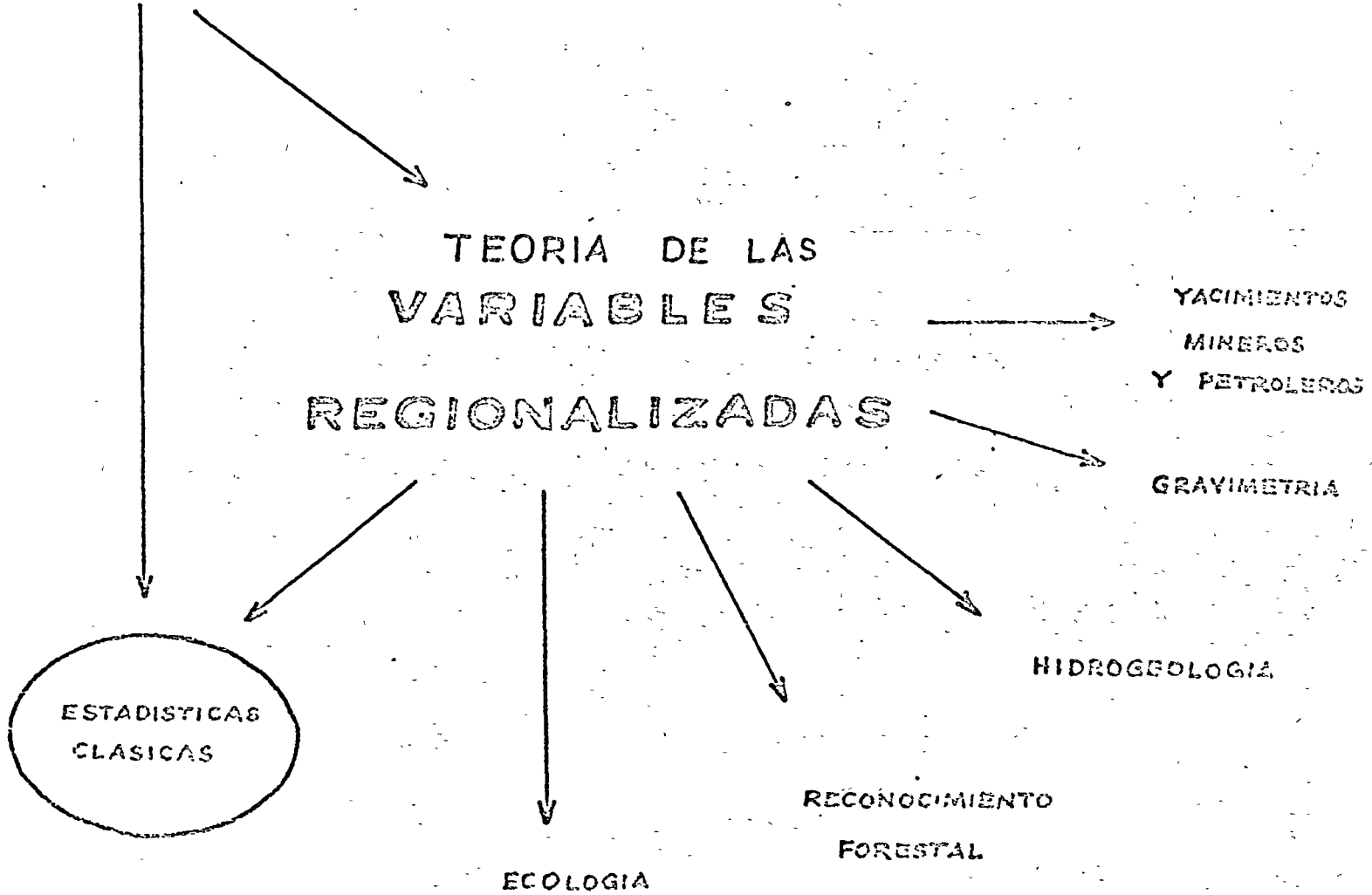
- un buen conocimiento físico del fenómeno estudiado
- un buen sentido, es decir el arte de hacer aproximaciones, de adaptar la Teoría.

El mejor geoestadístico forestal es primero un ingeniero forestal, el mejor geoestadístico minero es en un primer tiempo un geólogo, luego un minero finalmente un concentrador o un químico. Su herramienta común es el lenguaje probabilístico de la Teoría de las Variables Regionalizadas.

En esta óptica práctica, no presentaremos en forma abstracta la Teoría, sino más bien indicaremos como este útil teórico permite formalizar y resolver un cierto número de problemas-tipos que se presentan en el Reconocimiento de un yacimiento minero.

Uds. observarán, en el curso de la exposición, que un cierto número de problemas-tipos mineros son fácilmente transponibles a diversos dominios extra-mineros (petróleo, ecología, ...).

TEORIA MODERNA
DE PROBABILIDADES



2

Cartón 1

Problemas-tipos en Reconocimiento minero:

Consideramos un yacimiento de este tipo
 - en primer reconocimiento geológico: reconocimiento ciego o a partir de índices favorables.
 - hasta su explotación industrial día a día.

I-1 Primer Reconocimiento geológico

En una provincia, o en una vecindad metalogénica favorable, se busca revelar la existencia de posibles yacimientos. Para esto se procede

- generalmente a un primer reconocimiento a gran escala, la implantación de los sondajes es a menudo guiada sea por estudios anteriores (sísmicos por ejemplo) sea por conjeturas o hipótesis geológicas.
- a veces, a un primer reconocimiento sistemático y ciego de la zona favorable, y se trata de contar el número de sondajes positivos:

Al final de este primer reconocimiento, se trata de interpretar los resultados: ¿tal agrupación o alineamiento de sondajes positivos es característico de un posible yacimiento? Si la respuesta es afirmativa ¿cuál es la extensión en superficie o en volumen?

- en el caso del primer reconocimiento geológico preferencial, el papel del geólogo es adaptar sus hipótesis anteriores o transformarlas. La Geostatística no puede intervenir.
- en el caso de un reconocimiento ciego pero sistemático, la respuesta puede requerir de la competencia del geólogo y del geostatístico. En efecto, guiado por el geólogo, el geostatístico puede construir un modelo a priori de implantación y de dispersión de tamaños de los yacimientos posibles. Una vez adoptado el modelo, el geostatístico puede deducir criterios de decisión: tal agrupación de sondajes positivos es significativo. Del modelo a priori, y de los resultados del primer reconocimiento se podrán deducir primeros estimadores de los recursos mineralizados con varianzas de estimación.

RECONOCIMIENTO
GEOLOGICO PREFERENCIAL

- ESTUDIOS ANTERIORES
- HIPOTESIS GENETICA
- EXTENSION DE YACIMIENTOS VECINOS
(alineamientos, afloramientos)

ZONA DESERTICA

VARIOS TIPOS DE SONDAJES

- TOMADOS EN EL TERRENO
- CORTES GEOLOGICOS

RECONOCIMIENTO
SISTEMATICO Y CIEGO

ZONA SUB-MARINA
O ECUATORIAL

UN SOLO TIPO DE SONDAJES

MODELO DETERMINISTA

- CONTINUIDAD
- ESTRATIFICACIONES
- CUERPOS LENTICULARES
- FALLAS

TRAZO ESTIMADO DEL CONTORNO DE
LA MINERALIZACION

ESTIMACION DE RECURSOS

SIN VARIANZA DE ESTIMACION

MODELO PROBABILISTA

DISPERSION POISSONIANA
DE YACIMIENTOS

DISPERSION LOGNORMAL DE
LAS TALLAS DE LOS YACIMIENTOS

TRAZO ESTIMADO DEL CONTORNO

POLIGONAL DE LA MINERALIZACION

ESTIMACION DE RECURSOS

VARIANZA DE ESTIMACION

Sin embargo insistamos sobre el hecho que estos primeros conocimientos están eminentemente ligados al modelo probabilístico o geológico adoptado a priori, ellos son válidos solamente cuando el modelo mismo es válido. Este hecho es una característica de todos los procedimientos de reconocimiento ciego que abundan en la literatura y a menudo no es reconocido en forma clara.

Es claro, que si se adopta para la descripción (número, inclinaciones y tréllas de formas) de los yacimientos posibles un modelo a priori - sea determinista, probabilista o los dos a la vez - se puede definir el primer reconocimiento óptimo a efectuar. Pero este óptimo no es absoluto, es relativo al modelo adoptado, que no es la realidad desconocida.

- En la práctica se preferirán los modelos de tipo determinístico si las extrapolaciones de yacimientos vecinos a un conocimiento geológico preciso permite hacer este modelo plausible.

(-Un curso será dedicado a estos problemas de Reconocimiento ciego. Presentaremos aquí la Teoría Transitiva, es decir la Geostatística de la Estimación de Superficies y Volúmenes. Ejemplos concretos de los dos tipos de modelos serán discutidos).

2. Reconocimiento sistemático del yacimiento mineral.

Siendo localizada el yacimiento, se trata de estimar sus diversas características con la ayuda de reconocimientos más y más finos y secuenciales.

Esquemáticamente, se debe entregar:

- una estimación global de los recursos in situ: tonelaje mineral - tonelaje de metal - Ley en impurezas...

- estimaciones locales de estos recursos.

Estas estimaciones deben ser afectadas de una varianza de estimación que cuantifique su intervalo de confianza. Por otra parte, el estudio de estas varianzas de dispersión debe cifrar previamente la ganancia de información que aporta un cierto suplemento de reconocimiento, o definir el reconocimiento necesario para tener una precisión fijada de antemano.

En una etapa ulterior se define un cierto número de criterios de selección de recursos in situ, se trata entonces de estudiar estos recursos seleccionados o "reservas":

- estimaciones locales de los paneles o zonas seleccionadas;
- influencia del criterio de selección sobre las reservas: curva tonelaje - criterio de selección, por ejemplo.

curva tonelaje - { ley de corte.
potencia límite de explotabilidad
dimensiones de la unidad de selección.

Si la diferencia entre las 2 nociones de "recursos in situ" y de "reservas" son claras para todo el mundo, pocos autores analizan los sesgos graves de estimación que son provocados por el hecho que se seleccione tal o tal panel,

- raramente con la ayuda de la ley verdadera que es siempre conocida.

Sino a menudo con la ayuda de una ley estimada, esta estimación varía con el nivel de reconocimiento. En efecto la correlación efectiva, dispondría de una información más o menos fina, es decir de una ley estimada diferente a la disponible en el momento del estudio de las reservas.

Si no se corrigen estos sesgos, corremos en forma irremediable en los hechos bien conocidos en la práctica (ver Krige + autores americanos) que se va a:

- sobre-estimar la proporción de leyes ricas
- sub-estimar la proporción de leyes pobres.

Lo que, financieramente es catastrófico en ambos casos.

La ventaja principal de la Geoestadística, y que por sí sola justifica el empleo de la Geoestadística, es de haber:

- formalizado rigurosamente el problema de la selección
- propuesto estimadores simples (medias móviles lineales), insesgados de las reservas (Kriging y selección).

Todos estos problemas concierne la varianza de estimación. Hay otra varianza que la Geoestadística define claramente, es la varianza de dispersión que caracteriza una vez las fluctuaciones de las características (ley-pureza-impurezas etc...) al interior de un yacimiento o de una zona seleccionada.

Se confunde a menudo las 2 varianzas:

$$\sigma_{\hat{Z}}^2(Z^*(x_i) \rightarrow Z(x_i)) = E\{[Z^*(x_i) - Z(x_i)]^2\}$$

Varianza de estimación de $Z(x_i)$ por su estimador $Z^*(x_i)$ o error cuadrático medio, esta varianza depende:

- del panel particular $Z(x_i)$ estimado, implantado en x_i .
- del estimador formado $Z^*(x_i)$, es decir de la información disponible.

$\sigma^2(v|G)$, varianza de dispersión de las leyes reales desconocidas $Z(x)$ de los paneles de volumen v contenidos en el yacimiento G .

Esta varianza depende:

- de las dimensiones del soporte v y del dominio G .

Esta varianza no depende de la información disponible.

Se ve que en un problema de estimación, el minero tratará de formar el mejor estimador posible $Z^*(x)$ de la ley verdadera desconocida $Z(x)$, es decir el estimador $Z^*(x)$ que minimiza la varianza de estimación; el hace entonces un Kriging.

2 NOCIONES DE VARIANZAS

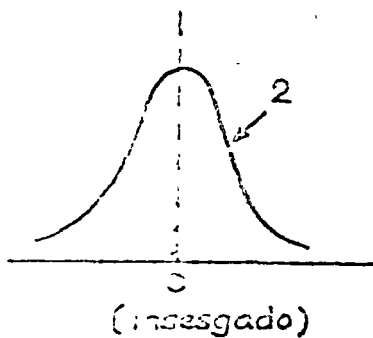
x_1	
x_2	x_4
x_3	x_5

i	Verdadera Z	estimada Z^{**}	$(Z - Z^{**})^2$	$(Z - m)^2$	$(Z^{**} - m^{**})^2$
1	5	6	1	0	1
2	7	6	1	4	1
3	6	4	4	1	1
4	2	4	4	9	1
5	5	5	0	0	0
Suma	25	25	10	14	4
$\frac{\text{Suma}}{5}$	$m = 5$	$m^{**} = 5$	2.0	2.8	0.8

Varianza
de estimación

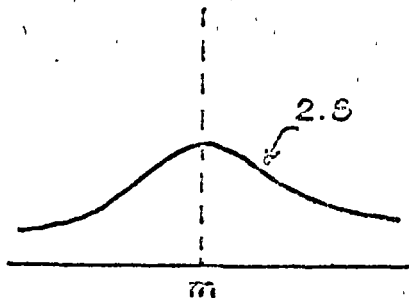
Dispersión de los
errores

$$E(x_i) = Z(x_i) - Z^{**}(x_i)$$

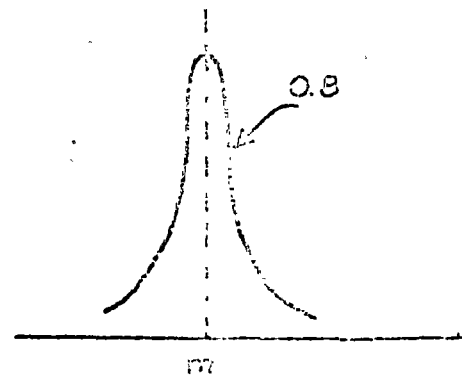


Varianzas de dispersión

Dispersión de las leyes
verdaderas: $Z(x_i)$



Dispersión de las leyes
estimadas: $Z^{**}(x_i)$



Mientras que en un problema de tratamiento de mineral, el concentrador de mineral se interesará más bien en la dispersión de las leyes (impurezas, por ejemplo) del mineral que el va a recibir:

v correspondiente a la unidad de selección de la mina.
G correspondiente a la producción de una jornada, o de un mes, o de un trimestre.

El concentrador de mineral tratará de conocer la varianza de dispersión $D^2(v|G)$, y eventualmente a disminuirla, jugando sobre la talla v de la unidad de selección o sobre el criterio de selección mismo.

El segundo aporte principal de la Geoestadística es de haber diferenciado netamente estas 2 varianzas y de permitir su cálculo.

El estudio de las varianzas de dispersión $D^2(v|G)$ deducidas de diferentes simulaciones de selección y de producción (movimiento-almacenamiento-homogenización) constituye otro punto práctico esencial de la Geoestadística.

Al término de estos estudios de reservas, de las simulaciones de producción o estudios de fluctuaciones de características a la entrada de la planta o usina, el yacimiento se pone en explotación.

Esta explotación dispondrá de una información día a día, muy precisa (pozos de tiro por ejemplo). La Geoestadística puede entonces, con la ayuda de esta última información y de un pequeño programa de computador simple, mejorar sus estimaciones anteriores y entregar sobre estas últimas bases un proyecto óptimo de producción o de selección.

Hemos hecho así un sobrevuelo cronológico rápido de los principales problemas-tipos que presenta el reconocimiento minero.

Se trata ahora de hacer un sobrevuelo de las técnicas utilizadas para resolver estos problemas, partiendo de técnicas empíricas que no son siempre las más malas, pasando por todas las técnicas de la estadística clásica que abundan en la literatura para llegar a la Geoestadística, o Práctica de la Teoría de las Variables Regionalizadas.

ii. Sobrevelo de las técnicas utilizadas

Muy groseramente, las técnicas de estimación se pueden clasificar en 2 grupos:

II-1. En primer lugar todas las técnicas empíricas, que suponen una perfecta y larga experiencia del fenómeno mineralizado. Estas técnicas comprenden también el olfato del geólogo-minero y todos los procedimientos clásicos de la geología minera (petrografía, tectónica, metalografía...). A menudo estas técnicas hacen uso de una serie de informaciones cualitativas y se caracterizan por el hecho de que son incapaces de dar el "intervalo de confianza" de sus estimaciones. Aunque estas estimaciones "empíricas" sean a veces, si no las únicas, al menos las "mejores" que se puede obtener. Estas técnicas empíricas no necesitan una metodología precisa, luego las dejaremos momentáneamente de lado. Momentáneamente, porque toda información cualitativa puede y debe orientar un estudio cuantitativo.

— en seguida las técnicas que necesitan una metodología precisa de tratamiento de datos cifrados. En este grupo tres escuelas son reconocidas y citadas en la literatura.

II-2 La primera escuela, o Escuela Norteamericana (S.W. Hazen, R.D. Berkenkotter, R.F. Hewlett, G.S. Koch, R.F. Link...) se apoya esencialmente sobre la Estadística clásica de variables independientes. Esta Estadística clásica ha sido elaborada hace mucho tiempo y utilizada con éxito en diferentes dominios extramineros (demografía - biología - economía, etc...). Su principal defecto, que es reconocido por estos autores, es que se apoya sobre hipótesis muy restrictivas para rendir cuenta de la complejidad de los fenómenos mineros (multi-dimensiones, anisotropías, no estacionaridad, tendencias, correlaciones espaciales). La realidad no debe ser forzada por una hipótesis inadecuada y aun más alterarla organizando una observación sesgada: proponer planos de muestra tales que los datos verifiquen las condiciones de independencia o limitando el reconocimiento solamente a las zonas mineralizadas que son ricas; ¡se trata de procedimientos no recomendables! Cualquiera que sea el fenómeno mineralizado, se le reduce así a un fenómeno muy particular sin correlaciones espaciales y por consiguiente particularmente simple a estudiar. ¿Pero, de qué vale el

resultado ?

En presencia de no-estacionaridad y de tendencias, los autores proponen distintos ajustes arbitrarios de polinomios para considerar el fenómeno de "trend" (tendencia) que no definen en forma clara: ¿es un valor esperado probabilístico, una media móvil, u otra cosa? Algunos autores, particularmente F.P. Agterberg ([1], F.P. Agterberg, 1970 y [12], W.A. Hodgson, 1972), han aproximado el riguroso formalismo de las correlaciones espaciales mediante el correlograma, que les permite en particular definir la noción geoestadística esencial de zona de influencia. En efecto, cuando el correlograma existe, esta herramienta es equivalente al variograma de la Geoestadística. Sin embargo hay numerosos casos en que el correlograma no existe mientras que el variograma puede existir (por ejemplo el proceso del movimiento Browniano); en este caso el estimador del correlograma es un puro artefacto sin significación.

En resumen, digamos que la Escuela Norteamericana parece confusa de la partida con la hipótesis de independencia espacial. Por otra parte la Geoestadística encuentra todos los resultados clásicos de la Estadística de Variables independientes como casos particulares; además el ajuste polinomial por mínimos cuadrados, o por el método de máxima verosimilitud son casos particulares de la teoría de Krigeage Universal. Estos casos particulares son muy simples, luego peligrosamente atractivos.

La segunda escuela, la Escuela Sudafricana (Krein - Sichel - M.T. Watson) nació del estudio muy pragmático del inmenso archivo de datos que representan las minas de oro sudafricanas. En primer lugar Sichel, luego Krige, han trabajado (1947 → 1960):

- las distribuciones lognormales de la ley en oro de sus muestras.

- que estas distribuciones son función de las dimensiones del soporte de la muestra. Citemos la relación de Krige, que la Geoestadística demostrará formalmente: $D^2(V|G) = D^2(V|V) + D^2(V|G)$. Con estas consideraciones, Krige y Sichel definieron estimadores $\hat{\mu}$ insesgados de la ley media de un panel a partir de las características media y variancia de los logaritmos de las leyes de las muestras.

Después, estudiando las regresiones entre las leyes verdaderas de bloques ya explotados y las medias móviles de las leyes de muestras disponibles a priori, Krige y Ueckermann (1963) definieron nuevos estimadores óptimos que ya no están ligados a la hipótesis restrictiva de la lognormalidad. Estos estimadores de Krige, que dieron nombre a la técnica de Krigeage formalizada más tarde por Matheron, permiten resolver sin sesgos el problema difícil de la estimación de reservas después de selección.

En resumen, la escuela sudafricana, poniendo al día las nociones claves de:

- existencia de correlaciones espaciales
- influencia de las dimensiones de la muestra o del panel
- problema de los sesgos de estimación cuando se efectúan selecciones sobre el mineral

está en el origen de la Geoestadística. Pero es necesario esperar las primeras obras (1955-1962) de Matheron para tener una visión de conjunto formalizada de estas nociones:

Heimos dicho que las medias móviles de Krige son, por definición, Krigeages. Respecto a los estimadores $\hat{\mu}$ de Sichel, solo son válidos para los minerales que verifican las hipótesis de lognormalidad de leyes, es decir precisamente los minerales con leyes muy débiles ($\lambda > 0, U$) estudiados por estos autores.

1955. - Llegamos finalmente a la tercera escuela, última en fecha, que ha aprovechado de los defectos de la Escuela Nortamericana y todo el trabajo anterior de la Escuela Sudafricana: la Escuela Francesa (esencialmente G. Matheron y también paralelamente Bertil Matérn en el dominio forestal). La razón por la cual esta escuela pudo ir más lejos que las otras dos se debe tal vez a que su iniciador, G. Matheron es un matemático probabilista que aprovechó el enorme conocimiento experimental de la escuela Sudafricana; su formación teórica le permitió formalizar en un lenguaje riguroso las observaciones experimentales de sus predecesores. Agreguemos también que Matheron ha estado rodeado de Ingenieros de Minas y de Geólogos, prácticos de la "cosa mineralizada", que han adaptado sin cesar su lenguaje probabilístico a la realidad concreta.

Toda la Geostatística se basa sobre la observación que toda variable (ley, potencia, acumulación, etc...) de un fenómeno mineralizado posee una función de autocorrelación espacial, por compleja que sea. Las leyes en un mismo metal en 2 puntos x y $x+h$ están autocorrelacionadas, esta correlación depende del vector \vec{h} (en módulo y dirección) que separa los 2 puntos. La independencia solo es un caso particular de autocorrelación y se trata en forma idéntica. Se caracteriza esta autocorrelación por un momento de orden 2 particularmente simple, llamado variograma, que es la herramienta fundamental de la Geostatística. Todo el problema es poder estimar este variograma, a partir de informaciones disponibles. Para esto se necesita la hipótesis de estacionaridad, que desarrollaremos más tarde. Es evidente que 2 mineralizaciones heterogéneas tienen 2 funciones de autocorrelación, 2 variogramas diferentes. Sin embargo si se tiene una no-estacionaridad debida a una tendencia (empobrecimiento progresivo de leyes por ejemplo), una vez quitada esta tendencia (Kriging Universal) es posible estimar el variograma del fenómeno estacionario residual.

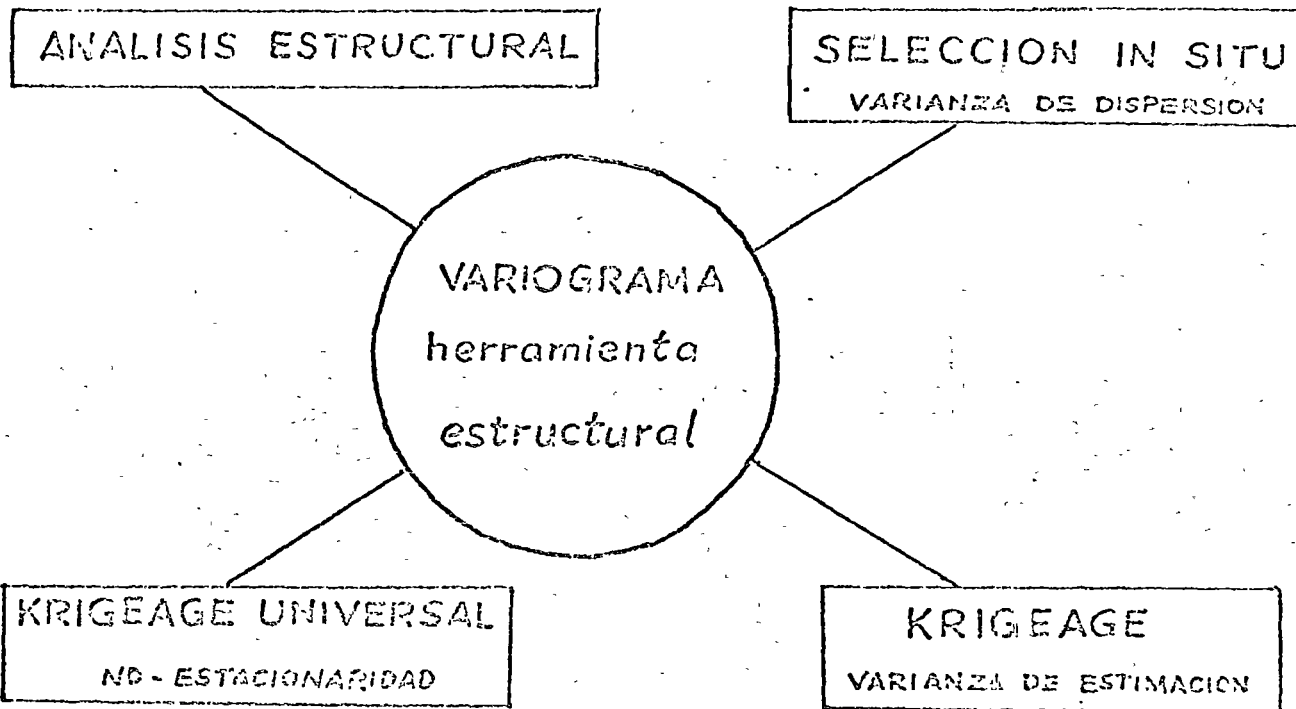
Es necesario ver bien que esta hipótesis de estacionaridad o mejor de cuasi-estacionaridad no es prohibitiva. Existe siempre una escala práctica donde se la encuentra. Por otra parte es intuitiva: a ningún minero se le ocurrirá por la mente estimar directamente las muy fuertes leyes de un filón, a partir de las leyes casi nulas del granito encajante; se pondrá a la escala interior del filón para encontrar la estacionaridad. Señalemos que

COPRLOGE 01A

NOCIÓN DE SOPORTE

ANÁLISIS ESTRUCTURAL
LIMITADO

RELACION DE KRIGE



13

cartón 1

TENDENCIAS POLINOMIALES

ESTUDIO DE SESGOS

RESIDUOS INDEPENDIENTES

LEY DE DISPERSION

○ GAUSSIANOS

ESTIMADORES DE SICHEL Y KRIGE .

esta hipótesis de estacionaridad, aun cuando no es siempre postulada, es implícita en todo el razonamiento estadístico de la Escuela Americana. en efecto, los histogramas y las dispersiones experimentales que se pueden calcular a partir de n muestras solo tienen significación cuando estas n muestras son sacadas de una misma población estacionaria.

Una vez en posesión de un estimador de esta función de autocorrelación "variograma", se le ajusta un modelo teórico particular, que sera entonces característica de la mineralización estudiada. A partir de este modelo teórico, la Teoría de las Variables Regionalizadas formaliza y resuelve los diversos problemas tipos que hemos citado.

Insistamos sobre la extrema homogeneidad de toda la Teoría, con el mismo lenguaje probabilístico y la misma herramienta de base variograma, se resuelven problemas tan diferentes como la estimación local y la homogenización de minerales.

Otro hecho esencial es que la Geoestadística ha encontrado y formalizado las principales nociones claves de las 2 escuelas que la han precedido; solo recusa las nociones falsas y probadas como falsas, las pruebas son a la vez teóricas y experimentales.

Una desventaja de la Geoestadística, ligada a su rigor, es que no se presta para una programación integrada (un pack Fortran unico de estimación por ejemplo); en efecto, la Geoestadística particulariza cada uno de los fenómenos estudiados para tratar mejor la información disponible. Por supuesto, existen subprogramas de Krigeage, de selección, de simulaciones, de cálculo de variogramas, pero a cada etapa es necesaria la interpretación o las aproximaciones con buen sentido del minero o del geólogo geoestadístico.

Otra desventaja de la Geoestadística es de no ser una exploración !!! La Geoestadística no crea la información, sino que la trata lo mejor, lo que consiste finalmente en economizar gastos de reconocimiento.

Ahora expondremos brevemente el lenguaje probabilístico de la Teoría de las Variables Regionalizadas, y mostraremos como este lenguaje unico permite expresar y resolver los problemas-tipos citados.

el lenguaje de la Geostatística.

III-1 - Noción de Variable Regionalizada.

Un fenómeno mineralizado cuantificable se caracteriza por el desplazamiento en el espacio (o el tiempo) de ciertas magnitudes, que llamaremos las Variables Regionalizadas. Citemos como ejemplos simples de variables regionalizadas:

- En el espacio clásico a 3 dimensiones, la ley en metal.
- En el espacio horizontal a 2 dimensiones, la potencia de un depósito sedimentario subhorizontal.
- En el espacio tiempo (1 dimensión t), el precio de un metal $p(t)$.

Corton 5

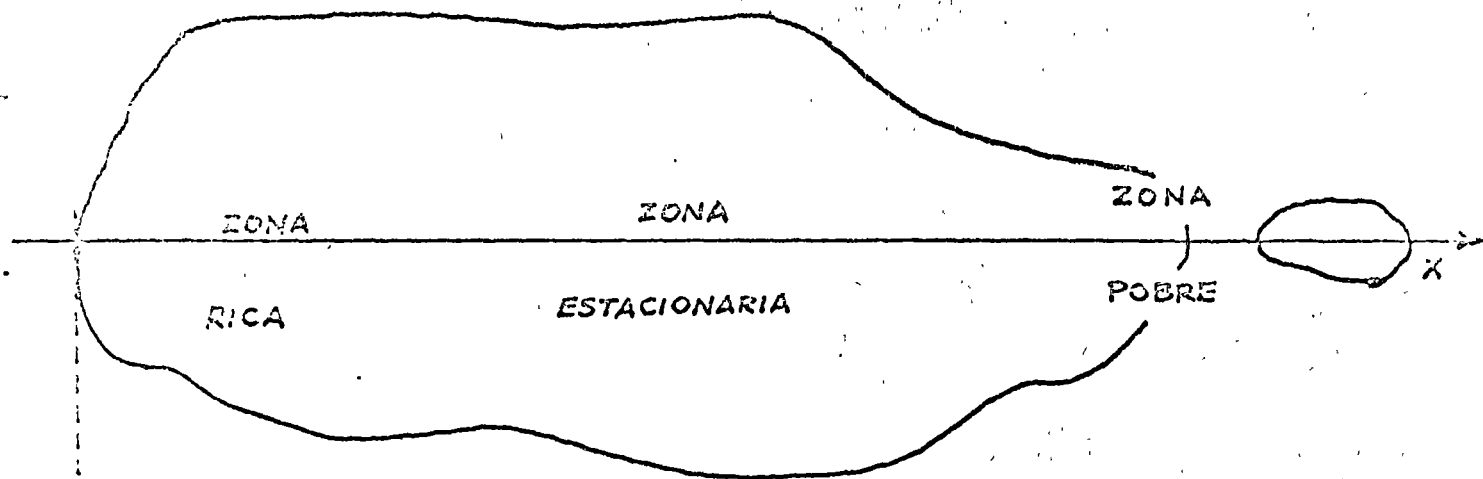
Del punto de vista matemático, una variable regionalizada (V.R.) es una función $y(x)$ del punto de coordenadas x . La variación de esta función en el espacio es a menudo muy irregular — la curva en dientes de sierra, dibujada a partir de los valores conocidos sub-estima las variaciones reales del fenómeno — entonces está excluido estudiar directamente la función matemática $y(x)$. Sin embargo bajo este aspecto caótico se esconde una cierta estructura: a pesar de las fluctuaciones locales, existe en casi todos los yacimientos:

- zonas ricas y zonas pobres, un sondaje implantado en una zona rica será — en promedio — más rico que un sondaje implantado en una zona pobre: el valor $y(x)$ depende de la implantación x .
- fenómenos de enriquecimiento y de empobrecimiento progresivo.

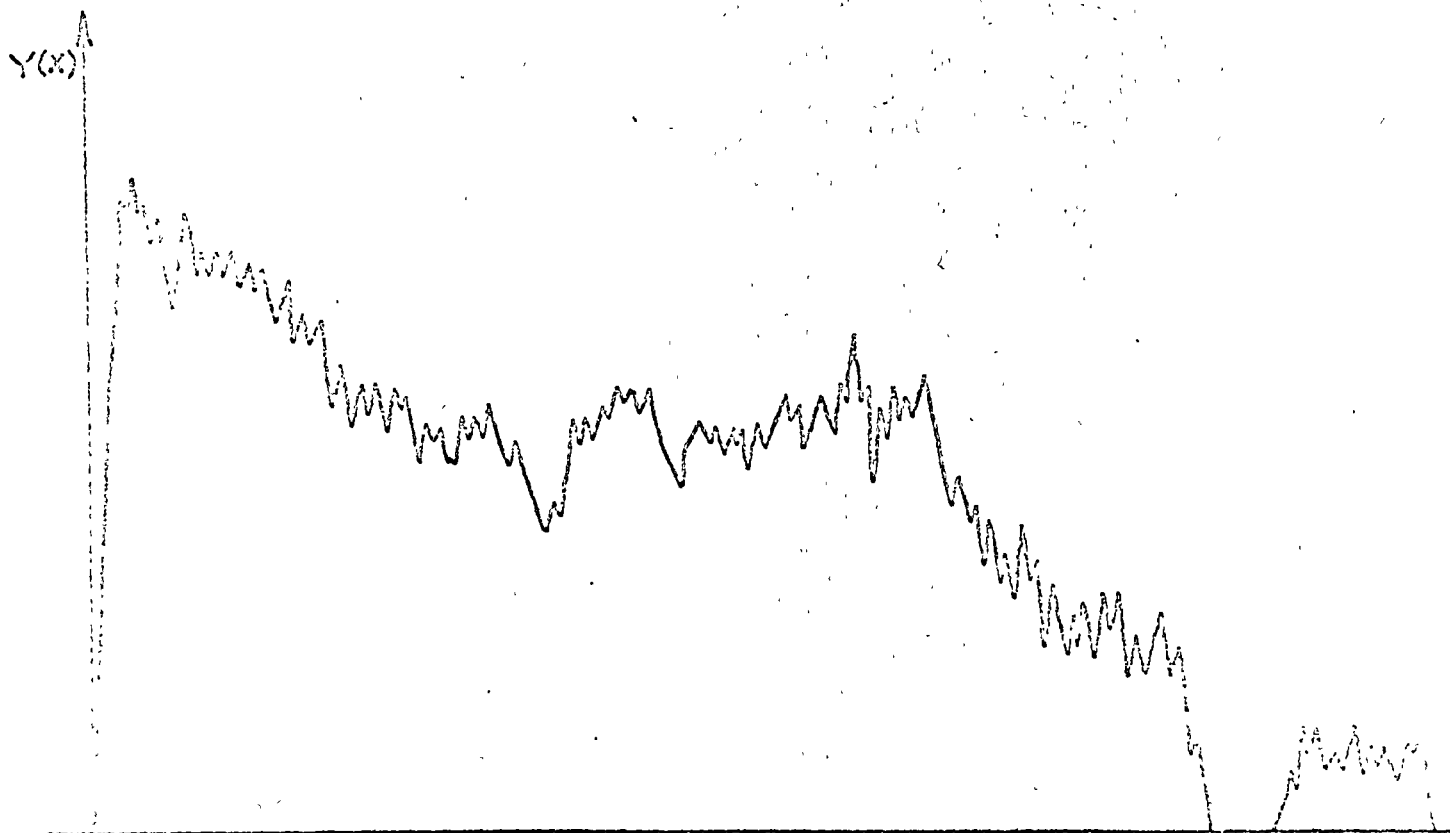
Una formalización correcta debe entonces tener en cuenta estos dos caracteres, aparentemente contradictorios de las variables regionalizadas:

- el carácter desordenado o "aleatorio" que sugiere recurrir a la noción de variable aleatoria, sin suponer para esto la independencia de estas variables.
- el carácter estructural, particular a cada fenómeno mineralizado o "regionalización" $y(x)$.

El lenguaje que permite considerar este doble aspecto aleatorio y estructural de las V.R. es el lenguaje probabilista de las Funciones Aleatorias.



ENRIQUECIMIENTO ESTACIONARIDAD EMPOBRECIMIENTO



La conceptualización probabilista de la realidad compleja es válida porque nos permite alcanzar los objetivos apuntados:

- caracterización de la estructura
- formalización homogénea y solución de los problemas tipos citados.

2. Noción de Función Aleatoria.

Cada uno sabe lo que es una Variable Aleatoria, es una variable que puede tomar un cierto número de valores con una cierta ley de probabilidad. Así el resultado de arrojar un dado no cargado es una variable aleatoria que puede tomar 6 valores (1 a 6) con una probabilidad ($1/6$) para cada uno de estos valores.

Si un tiro particular da 5 por ejemplo, diremos que 5 es una realización particular de la Variable Aleatoria "resultado de tirar un dado".

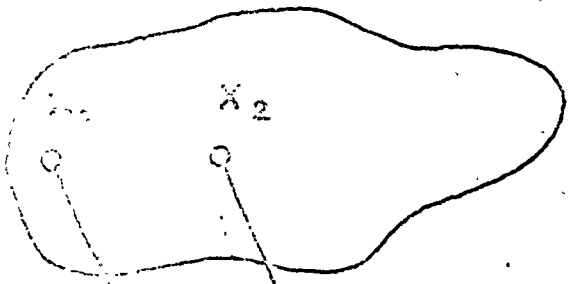
Analogamente, consideremos la ley $y(x_1) = 1.8\%$ Cu. en un punto preciso x_1 . Diremos que esta ley 1.8 es una realización particular de una cierta variable aleatoria $Y(x_1)$ implantada en el punto x_1 . Así el conjunto de leyes $y(x)$ por $x \in$ yacimiento, es decir la variable regionalizada $y(x)$ es una realización particular del conjunto de todas las variables aleatorias $Y(x)$, $x \in$ yacimiento. Este conjunto infinito de variables aleatorias se llama "Función Aleatoria" y usaremos la notación $Y(x)$.

La expresión misma: "Función Aleatoria" contiene el doble aspecto aleatorio y estructural de una Variable Regionalizada:

- localmente $Y(x_1)$ es una Variable Aleatoria.
- pero $Y(x)$ es una Función Aleatoria, es decir que 2 variables aleatorias $Y(x_1)$ e $Y(x_1+h)$ están ligadas por correlaciones.

Inferencia probabilista y estacionaridad.

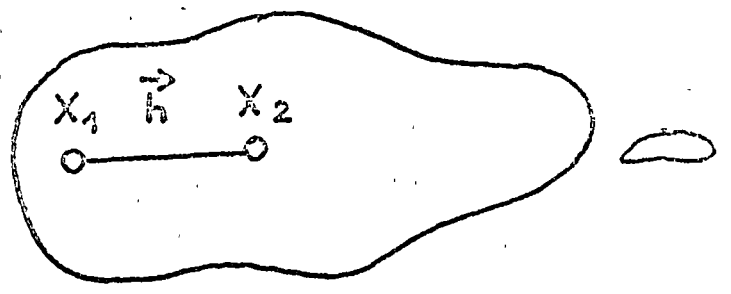
De este modo, se interpreta una Variable Regionalizada como 1 realización particular $y(x)$ de una cierta Función Aleatoria $Y(x)$. Pero el conocimiento de una realización única $y(x)$, además limitada a un cierto número de sondajes $x_i \in$ información, no permite remontar a la ley de $Y(x)$. Así el conocimiento de un tiro único no permite remontar a la ley de probabilidad del lanzamiento de un dado, es decir de responder la pregunta: "el dado está cargado?" Para esto es necesario repetir varios tiros del dado. Analogamente sería necesario conocer varias realizaciones $y_1(x)$, $y_2(x)$, ..., $y_n(x)$ de nuestra Función Aleatoria $Y(x)$ para inferir su ley, al menos una parte interesante de su ley. Por otra parte disponemos solamente de una realización



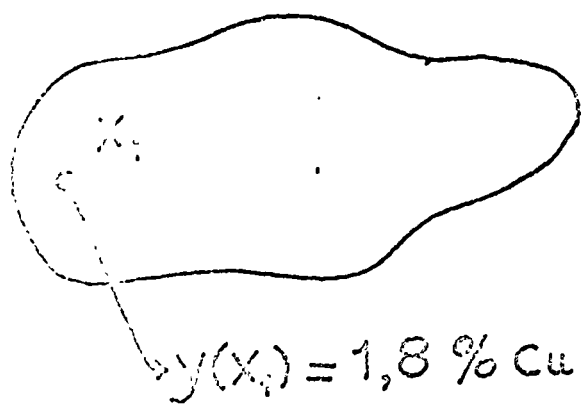
$y(x_2) \rightarrow$ VARIABLE ALEATORIA $Y(x_2)$
 $y(x_1) \rightarrow$ VARIABLE ALEATORIA $Y(x_1)$

FUNCION ALEATORIA
 $Y = \{ Y(x_1), Y(x_2), \dots \}$

CORRELACION ENTRE
 $Y(x_1)$ e $Y(x_2)$
 FUNCION DE \vec{h}



REALIZACION $y(x_1)$
 EN EL PUNTO x_1
 DE UNA VARIABLE ALEATORIA
 $Y(x_1)$



$y(x_1) = 1,8 \% Cu$

única limitada $y(x_i)$, $x_i \in I \subset G$. Es aquí donde interviene la hipótesis estacionaria:

Muy a menudo, en la práctica, la Variable Registrada en estudio $y(x)$ se repite en el espacio. 2 valores $y(x_0)$ e $y(x_0+h)$ pueden ser considerados como 2 realizaciones diferentes de la misma variable aleatoria $Y(x)$, por ejemplo. Esta hipótesis de estacionariedad unida a la hipótesis muy fuerte de ausencia de autocorrelación: ($y(x_0)$ independiente de $y(x_0+h)$) es la que justifica la estimación del histograma de la V.A. $Y(x)$ por la dispersión experimental de los datos disponibles: $y(x_i)$, $x_i \in I \subset G$.

La Geoestadística necesita una hipótesis mucho menos fuerte. No se interesa en los valores $y(x)$ mismos, sino solamente en sus incrementos $[y(x) - y(x+h)]$. Y, es la estacionariedad de estos incrementos lo que postula la Geoestadística:

2 parejas de datos $[y(x_i) - y(x_i+h)]$ e $[y(x_j) - y(x_j+h)]$ distantes del mismo vector \vec{h} se consideran como 2 realizaciones diferentes del mismo incremento aleatorio $[y(x) - y(x+h)]$.

Esta hipótesis estacionaria limitada a los incrementos se llama hipótesis intrínseca.

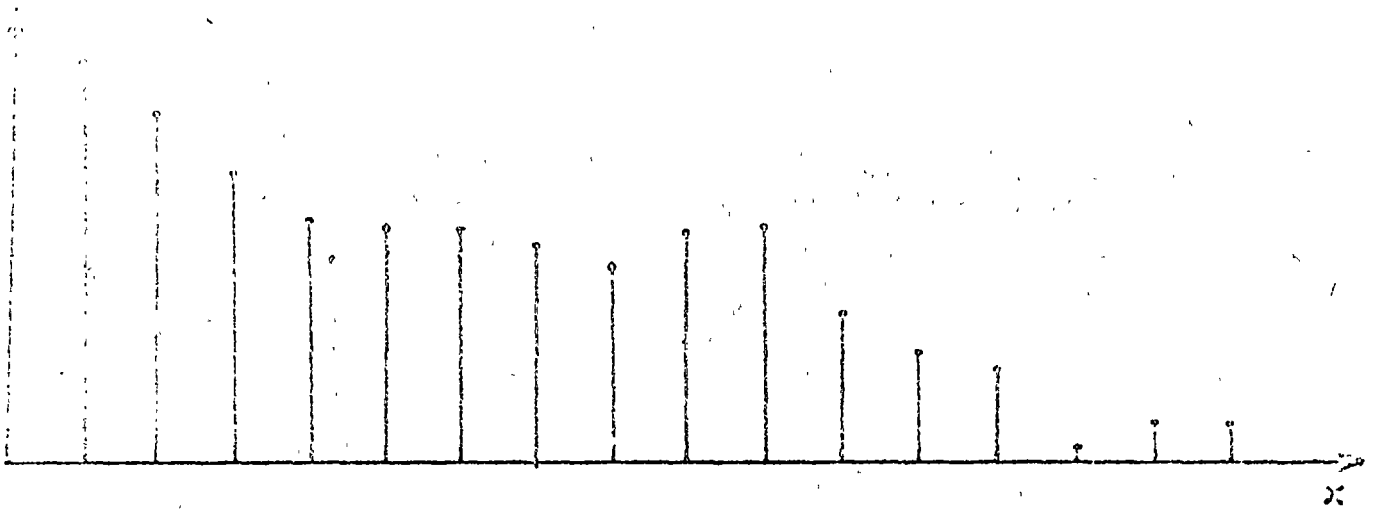
Bajo esta hipótesis intrínseca, los datos disponibles $\{y(x_i), x_i \in I \subset G\}$ permiten la estimación del momento de orden 2 de los incrementos, es decir la estimación del variograma:

$$2\gamma(\vec{h}) = E\{[Y(x+\vec{h}) - Y(x)]^2\}.$$

Valor medio de la desviación cuadrática $[Y(x+\vec{h}) - Y(x)]^2$.

Además la Geoestadística no exige que la hipótesis intrínseca se verifique en todo el yacimiento G , es decir \forall vector \vec{h} . Para los cálculos basta tener la hipótesis intrínseca sobre pequeños dominios ($|\vec{h}| < \text{limite}$) deslizantes sobre el yacimiento G .

— Remarquemos que la no-estacionariedad de los incrementos o la no estacionariedad en general puede deberse a 2 causas:

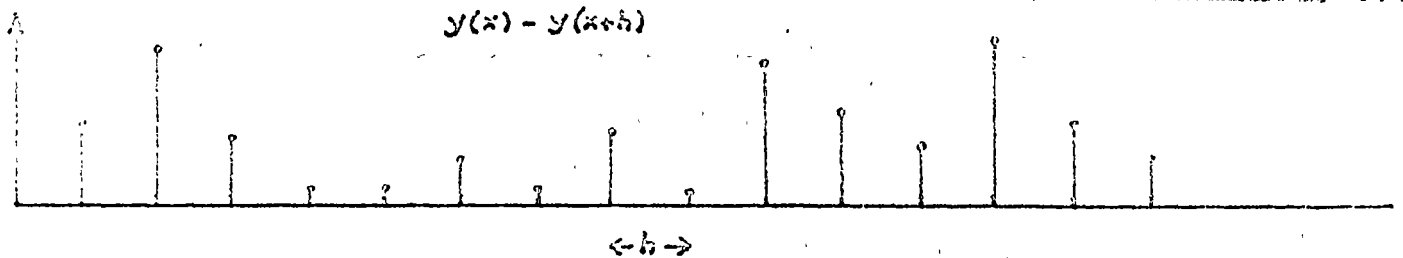


$\{y(x_1), \dots, y(x_n)\} = K$ realizaciones de K variables aleatorias diferentes

HIPOTESIS ESTACIONARIA



$\{y(x_1), \dots, y(x_n)\} = K$ realizaciones de 1 variable aleatoria única



HIPOTESIS INTRINSECA



$\{y(x_2) - y(x_1), \dots, y(x_n) - y(x_{n-1})\} = K = 1$ realizaciones de un incremento aleatorio único

- Existencia de una deriva progresiva de las características del yacimiento (tendencia): en promedio $y(x)$ decrece cuando se va de A a B. $E\{Y(x)\} = m(x) \downarrow$. Basta entonces quitar esta deriva funcional o aleatoria $m(x)$ y estudiar el residuo que es estacionario (lo que no quiere decir que no hay correlaciones espaciales propias):

$$Y(x) = m(x) + R(x)$$

- Heterogeneidad profunda entre 2 partes del mismo yacimiento. Es necesario entonces distinguir imperativamente las 2 mineralizaciones y estudiarlas aparte.

Señalemos que la herramienta variograma, construido a partir de los datos brutos disponibles permite en general revelar y diferenciar los dos casos de no-estacionariedad.

Volvamos ahora a los problemas-tipos que hemos citado e indiquemos brevemente como el lenguaje de las F.A. permite expresar estos problemas y como la herramienta variograma permite resolverlos.

Problemas-tipo y geoestadística

Primer Reconocimiento geológico.

Hemos dicho antes que en esta primera fase de reconocimiento, en ausencia de información anterior sistemática y cifrada, la orientación de este primer reconocimiento geológico como la interpretación de sus resultados dependen eminentemente del modelo adoptado a priori:

- modelo genético del geólogo
- modelo determinista o probabilista de implantación de yacimientos, de dispersiones de sus formas, etc...

Existe sin embargo un problema importante donde la Geoestadística aporta una respuesta puramente geométrica luego perfectamente general, en ausencia de cualquier hipótesis probabilística: es el problema de estimación de superficies (depósito sedimentario) o de volumen por una maila sistemática regular o de densidad regular.

Cantor 3

En este problema, la variable regionalizada interesante es la variable en todo o nada.

$$k(x) = \begin{cases} 1 & \Leftrightarrow x \in G \\ 0 & \Leftrightarrow x \notin G \end{cases}$$

Y se interesa en la estimación de la integral:

$$S = \int_{-\infty}^{+\infty} k(x) dx$$

La Geoestadística define la función estructural "covariograma geométrico":

$$K(h) = \int_{-\infty}^{+\infty} k(x) k(x+h) dx$$

Observamos que $S = K(0)$.

La superficie real desconocida S se estima por la unión de las m superficies de influencia de los m sondeos positivos; es decir por:

$$S^*(x_0) = a \sum_{p=-\infty}^{+\infty} k(x_0 + pa)$$

escritura unidimensional, $a \rightarrow$ superficie de la maila

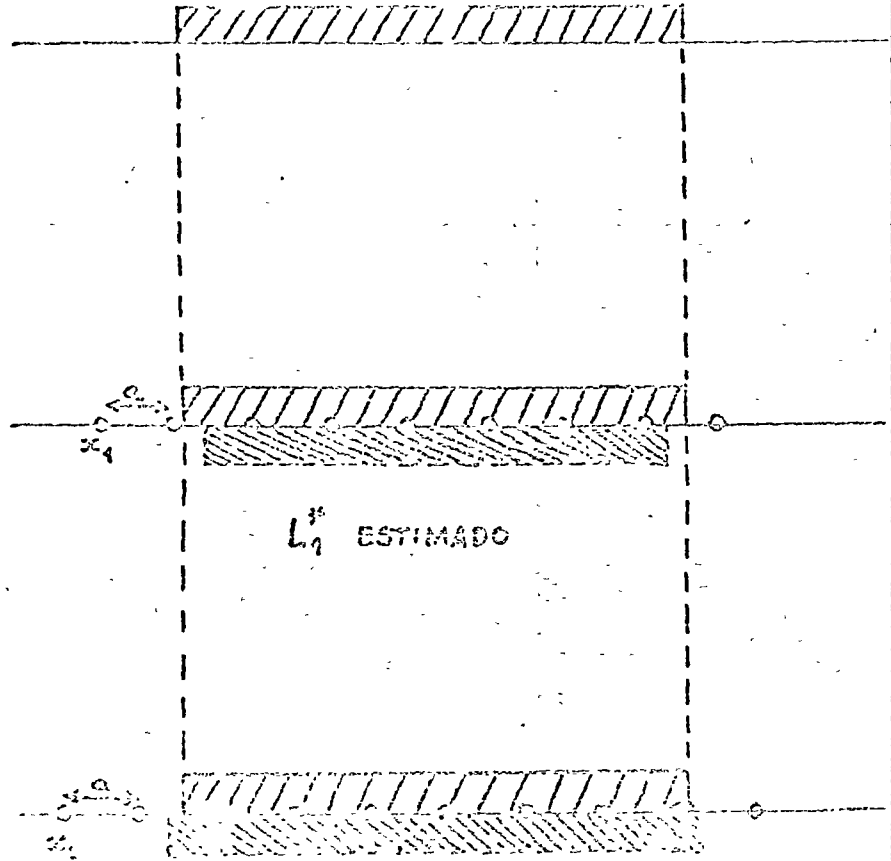
en que x_0 es el origen de la maila.

Se ve que el estimador S^* depende de la posición de este origen x_0 en la maila de base. Por consiguiente,

RECONOCIMIENTO GEOLOGICO

EXTENSION L DESCONOCIDA

$k(x)=0$ $k(x)=1$ $k(x)=0$



L_1 ESTIMADO

L_2 ESTIMADO

ESTIMACION DE SUPERFICIES CON LA METODOLOGIA DISPERSA

$n = n^\circ$ de sondeos

$2N_1 = n^\circ$ de elementos $ll a_1$



$n = 38$
 $2N_1 = 24$
 $2N_2 = 20$

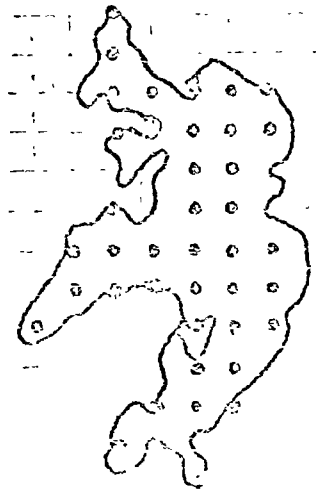
$$S^* = 38 a_1 a_2$$

$$\frac{\sigma}{S} = 1\%$$

$$S = (38 \pm 3) a_1 a_2$$

24

$n = 40$



$n = 40$
 $2N_1 = 32$
 $2N_2 = 26$

$$S^* = 40 a_1 a_2$$

$$\frac{\sigma}{S} = 5\%$$

$$S = (40 \pm 2) a_1 a_2$$

La varianza de estimación de S por S^* , es decir el error cuadrático medio, se escribe:

$$\sigma_E^2(a) = \frac{1}{a} \int_a [S^*(x_0) - S]^2 dx_0$$

Los dos términos $S^*(x_0)$ y S que figuran en esta integral se expresan en función de la v.r. $k(x)$. Al desarrollar esta integral se encuentra:

$$\sigma_E^2(a) = a \sum_{p=-\infty}^{+\infty} K(pa) - \int_{-\infty}^{+\infty} K(h) dh$$

La varianza de estimación por la malla, a se expresa únicamente con la función estructural $K(h)$. Pero esta última es tan desconocida, como la verdadera superficie real S . Sin embargo, cualquiera que sea la forma y la extensión de S , todo covariograma geométrico posee una pendiente en el origen $K'(0)$ lineal. Esta propiedad geométrica general permite conocer el primer término del desarrollo en el origen ($a \rightarrow 0$) de la varianza de estimación. Así, en un problema a 2 dimensiones de estimación de superficie, se demuestra la fórmula de primer orden siguiente:

$$\sigma_E^2 = \frac{1}{m^2} \left[\frac{1}{6} N_2 + 0.06 \frac{(N_1)^2}{N_2} \right] + \dots \quad (N_2 \leq N_1)$$

varianza de estimación relativa

(válida cuando $m \geq 10$)

m : número de sondajes positivos.

$2N_1$ y $2N_2$ son los números de elementos paralelos a los 2 lados de la malla (a_1, a_2), que constituyen el contorno de la unión S^* de las m zonas de influencia positivas.

Para la estimación de volúmenes (3 dimensiones) por una malla en forma de paralelepípedo, también se dispone de una fórmula simple.

Insistimos sobre el hecho que estas fórmulas de primer orden son perfectamente generales y de un carácter puramente geométrico.

Una vez localizado el yacimiento, se trata ahora de evaluarlo y en primer lugar de caracterizarlo. Vale la pena mostrar como la herramienta variograma estimado permite esta caracterización.

Bajo la hipótesis de estacionaridad local de los incrementos $[Y(x+h) - Y(x)]$, es posible estimar el variograma, $\gamma(h)$, definido como la esperanza del cuadrado de los incrementos.

$$2\gamma(h) = E\{[Y(x+h) - Y(x)]^2\}$$

— Es necesario ver bien que esta escritura unidimensional se generaliza a un espacio cualquiera a m dimensiones. En el espacio habitual a 3 dimensiones, el argumento h representa un vector (h_1, h_2, h_3) , y la notación condensada $\gamma(h)$ representa en efecto la función vectorial $\gamma(h_1, h_2, h_3) = \gamma(|h|, \alpha)$, siendo α la dirección del vector h y $|h|$ su módulo.

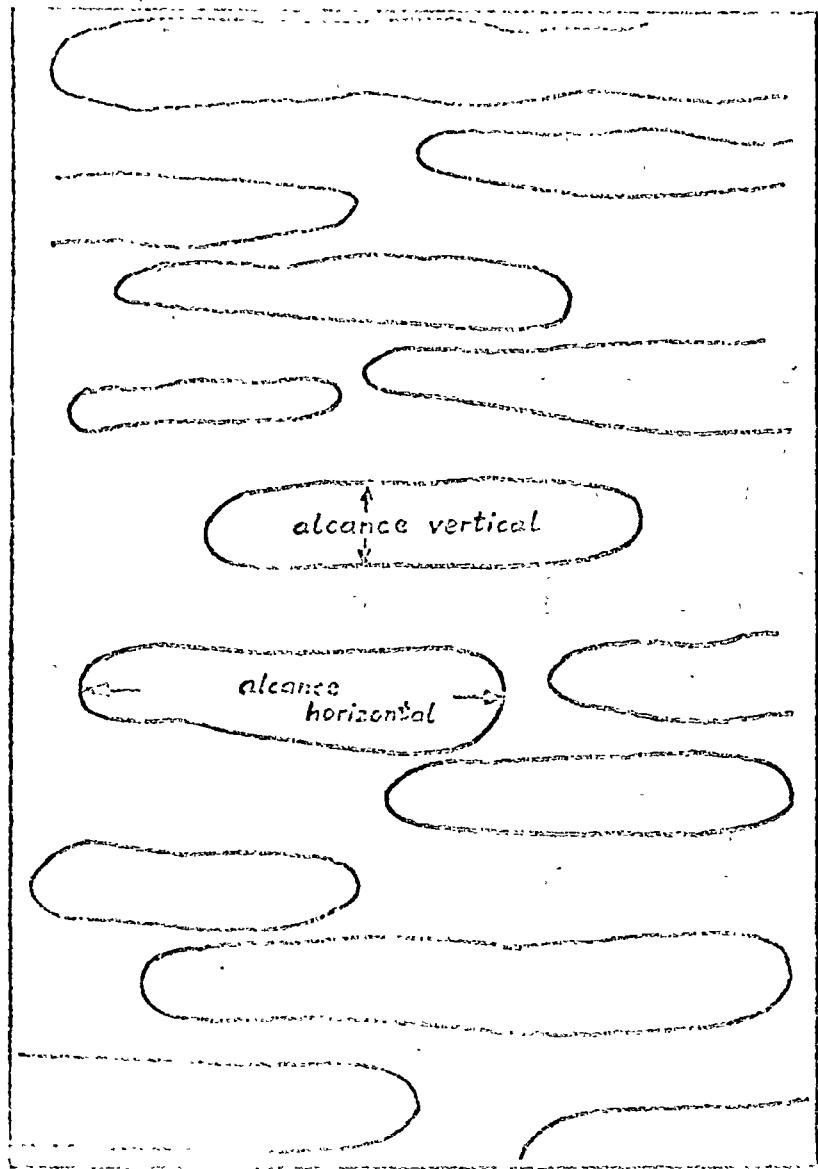
Zona de influencia: Fijemos la dirección α .

En general el variograma es una función creciente del módulo $|h|$: Los valores tomados en 2 puntos distintos serán en promedio más diferentes cuando estos dos puntos sean más alejados uno del otro. El aspecto de este crecimiento da una significación precisa a la tradicional noción de "zona de influencia de una muestra". Sucede a menudo que este crecimiento se atenúa de a poco, y más allá de una cierta distancia a que llamaremos "alcance" $\gamma(h)$ llega a ser una constante: la variación promedio de las leyes de 2 muestras no depende más de su distancia; se alcanza entonces el estado de independencia de las muestras postulado por la estadística sin correlaciones espaciales. No basta con la meseta del variograma, es decir prácticamente con la variancia de dispersión de las muestras, para caracterizar nuestro fenómeno, sino con toda la curva $\gamma(h)$ y sobre todo la parte $|h| < a$ dentro de la zona de influencia.

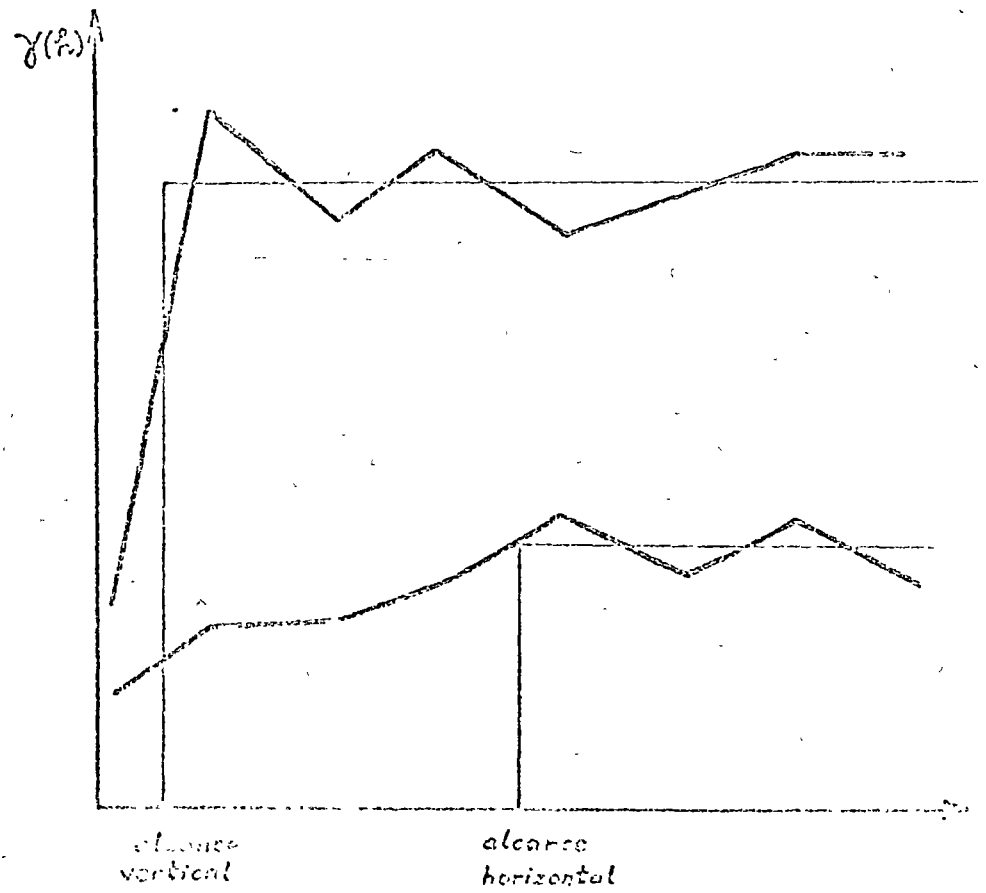
Es esencial remarcar que la Geoestadística encuentra todos los resultados clásicos de la Estadística en correlaciones espaciales como casos particulares, haciendo $\gamma(h) = \text{Constante } \forall |h| \neq 0$.

Anisotropías: Nos interesamos ahora en la dirección α del vector h . Si se sospechan anisotropías direccionales, hay que construir el variograma según varias direcciones del espacio. El estudio de las deformaciones de $\gamma(|h|)$, cuando se hace variar la dirección α de h , permite poner en evidencia y caracterizar estas eventuales anisotropías. En este caso particular, con referencia a un modelo del tipo:

ZONA DE INFLUENCIA



DESCRIPCION MINERA



VARIOGRAMA

27

10-10-77

$$\gamma(h_1, h_2, h_3) = \gamma_1(\sqrt{h_1^2 + h_2^2}) + \gamma_2(h_3, i)$$

El variograma tridimensional es la suma de una estructura isotrópica horizontal γ_1 y de una estructura vertical correspondiente a la dirección de las veins mineralizantes.

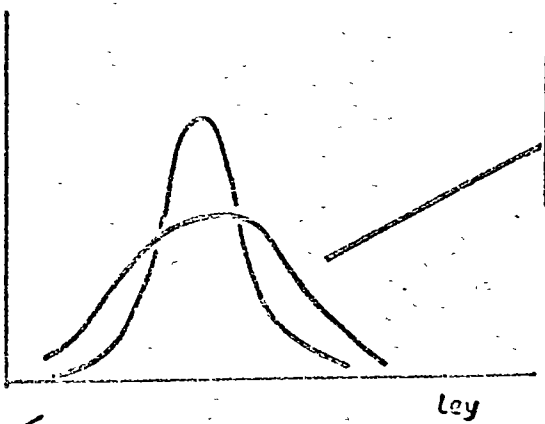
Estacionariedad local: Muy a menudo en la práctica, no es necesario conocer el variograma $\gamma(h)$ para todas las distancias $|h|$. En general basta conocer su comportamiento en el origen y su crecimiento (con o sin mesetas) hasta una cierta distancia $|h|=b$, útil para las estimaciones. Nos conformamos entonces con la estacionariedad local de los incrementos $[Y(x+h) - Y(x)]$ sobre vecindades deslizantes de extensión b . En particular, un efecto de deriva (no-estacionariedad) que se manifiesta más allá de b no tiene importancia.

Carier 10. Soporte: El variograma experimental y los datos a partir de los cuales ha sido calculado, está definido sobre un cierto soporte, por ejemplo un trozo de testigo de sección y longitud fijas. El cambio de soporte corresponde a un formalismo simple que permite a la Geostatística deducir γ_p (paneles) a partir de γ_c (testigos) o γ (puntual) del mismo γ_c . Es fácil darse cuenta de la importancia práctica extrema de este formalismo: un yacimiento no se explota por testigos, y es bien evidente que las leyes de correlaciones y de dispersiones entre paneles de algunos miles de toneladas no tienen nada que ver con las leyes correspondientes relativas a testigos de algunos kilos. Esta evidencia de simple buen sentido parece no haber sido bien comprendida por los autores que pretenden efectuar selecciones, o estimar reservas seleccionadas a partir de las propiedades de dispersión de sus muestras disponibles. Volveremos sobre este punto capital.

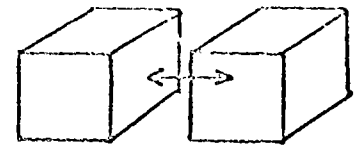
SOPORTE

EXPLOTACION

frecuencia

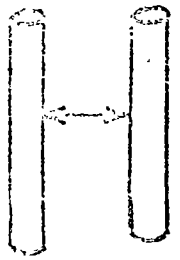


γ_p



PANELES

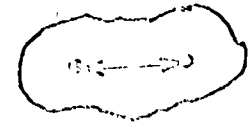
ESTIMACION



DES MEDOS

γ_c^*

FORMALIZACION



PUNTUAL

γ

IV-3 Estimación de recursos (in situ)

Código 14

Es claro que la elaboración de un estimador de recursos globales de un yacimiento depende:

- de la información disponible, no solamente en cantidad N sino también en su repartición relativa: una malla regular sistemática no entrega la misma información que una malla preferencial.
- de la estructura subyacente de la característica estudiada, por ejemplo: una potencia mineralizada presentará a menudo un carácter de continuidad más favorable a la estimación que la variable acumulación de metal o ley media correspondiente.

La Geostatística, al caracterizar el grado de continuidad de cada mineralización particular y las diversas zonas de influencia de la información disponible, permite construir — sin sesgos — estos estimadores globales de los recursos in situ. Pero sobre todo, la Geostatística calcula las varianzas de estimación, es decir los intervalos de confianza de estos estimadores:

Quando se estima la ley media $Y_V(x)$ de un panel de volumen V por la ley media Y_v de un conjunto cualquiera v de informaciones disponibles, el error cometido $Y_V - Y_v$ aparece como una variable aleatoria de la cual se puede calcular la varianza:

$$\sigma_E^2 = E\{[Y_V(x) - Y_v]^2\}$$

Esta varianza es independiente de los valores particulares de la información utilizada y se expresa unicamente con la ayuda del variograma γ característico de la ley considerada.

$$\sigma_E^2 = 2\bar{\gamma}(V, v) - \bar{\gamma}(V, V) - \bar{\gamma}(v, v)$$

$\bar{\gamma}(V, v)$ designa el valor medio de $\gamma(\vec{h}) = \gamma(MM')$ cuando los 2 puntos de apoyo M y M' del vector \vec{h} describen, independientemente uno del otro, respectivamente los 2 volúmenes o conjuntos V y v .

- Los valores medios $\bar{\gamma}$ se calculan sea directamente si los conjuntos V y v son de geometría simple, si no por integración numérica con la ayuda de un computador. Así la fórmula totalmente general da la varianza de estimación

VARIABLES DE ESTIMACIÓN

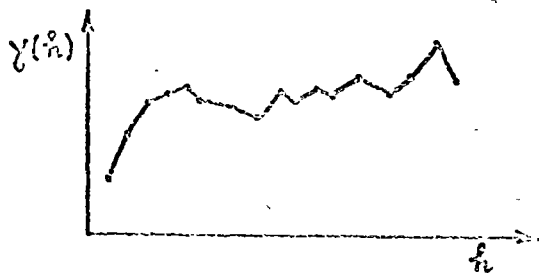
INFORMACION DISPONIBLE

ESTRUCTURA SUB-YACENTE

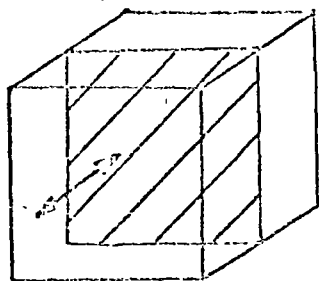
$\{v\}$ ESTIMADOR DE $\{V\}$

$$\sigma_E^2 = 2 \bar{\gamma}(v, V) - \bar{\gamma}(V, V) - \bar{\gamma}(v, v)$$

31

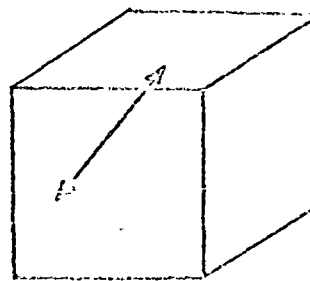


ESTRUCTURA DEL FENÓMENO MINERALIZADO



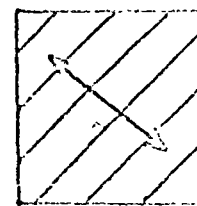
$2\bar{\gamma}(v, V)$

GEOMETRIA RELATIVA ESTIMADO V / ESTIMANTE v



$\bar{\gamma}(V, V)$

GEOMETRIA DEL ESTIMADO V



$\bar{\gamma}(v, v)$

GEOMETRIA DEL ESTIMANTE v

ESTIMACION

de una media desconocida sobre V por la media conocida v . V y v son conjuntos cualquiera.

- Señalamos que en el caso particular de independencia espacial: $\gamma(h) = \begin{cases} \text{Constante } \forall h > 0 \\ 0 & \text{para } h = 0 \end{cases}$, esta fórmula encuentra las varianzas en $\frac{\sigma^2}{N}$ de la Estadística clásica.

- Observemos aún que la escritura simple de esta fórmula general considera los hechos esenciales - y por otra parte intuitivos - que condicionan toda estimación:

La calidad de una estimación depende:

- 1) de la geometría del volumen a estimar (término $\bar{\gamma}(V, V)$)
- 2) de las distancias entre estimado y estimante: $\bar{\gamma}(V, v)$
- 3) de la geometría relativa de los estimantes: $\bar{\gamma}(v, v)$
- 4) del grado de regularidad del fenómeno estudiado: utilización de la característica estructural γ .

Composición de términos.

Cartón 12

En una estimación global, V representa todo el yacimiento o una parte importante de este yacimiento y v un conjunto complejo de informaciones de diferente naturaleza (sondajes a percusión, pozos, galerías, etc...). La fórmula exacta anterior conduce a cálculos muy largos. Se definen entonces fórmulas de aproximación que reposan sobre principios de composición de estimaciones elementales: así, en la estimación de un cuerpo por niveles, se distingue y después se compone:

- la estimación de cada galería por las muestras
- la estimación de cada nivel por sus galerías supuestas conocidas
- finalmente, la estimación de cada tajada del cuerpo por su nivel central (o sus 2 niveles extremos) supuestas conocidas

Para la estimación global, no serviría de nada multiplicar las muestras en las galerías si no se multiplican el número de tales galerías y también el número de niveles que reconocer el cuerpo mineralizado. Conociendo el precio del metro de avance, y el precio de una muestra, este principio de composición de términos permite optimizar un presupuesto de reconocimiento dado, es decir cuál es la razón óptima entre las densidades de galerías y muestras.

VARIANZA DE ESTIMACION :+ TERMINO DE LINEA
+ TERMINO DE NIVEL
+ TERMINO DE TAJADA

CALCULABLE A PRIORI

COMPARACION DE DIVERSOS ESQUEMAS DE
RECONOCIMIENTO

DEFINICION Y GESTION
DEL PRESUPUESTO DE RECONOCIMIENTO

33

Conf. 01. 18

Insistimos sobre el hecho que el cálculo de una varian-
za de estimación solo necesita el conocimiento de la
función estructural γ y las geometrías del estimado y
estimante v . Esta varianza de estimación puede calcu-
larse antes de realizar la información v . Así entonces
se podrá poner en balance el aporte de información y
el costo correspondiente antes de realizar el reconoci-
miento que se ha proyectado, analogamente se podrán com-
parar diversos esquemas de reconocimiento posibles, es
decir administrar mejor el presupuesto de reconocimiento.

Krigeage.

Carácter 12

La posibilidad de calcular la varianza de esti-
mación para todo estimador permite, en particular,
definir el mejor estimador que se puede deducir de
una información dada: es el formalismo de "Krigeage".
Presentemos un ejemplo:

Sea a estimar el panel P . La información
disponible está constituida de sondajes interiores y
exteriores a P . En nuestro ejemplo se dispone:

- de un trozo de testigo interior : S_i
- de 7 trozos de testigos exteriores: Se_1, Se_2, \dots, Se_7

¿Qué peso dar a cada una de estas 8 infor-
maciones para definir el mejor estimador de la Ley media
desconocida Z del panel P ?

El Krigeage entrega un estimador Z^* , combina-
ción lineal de las ocho informaciones anteriores dis-
ponibles:

$$Z^* = \lambda_i S_i + \sum_{j=1}^7 \lambda_{e_j} Se_j$$

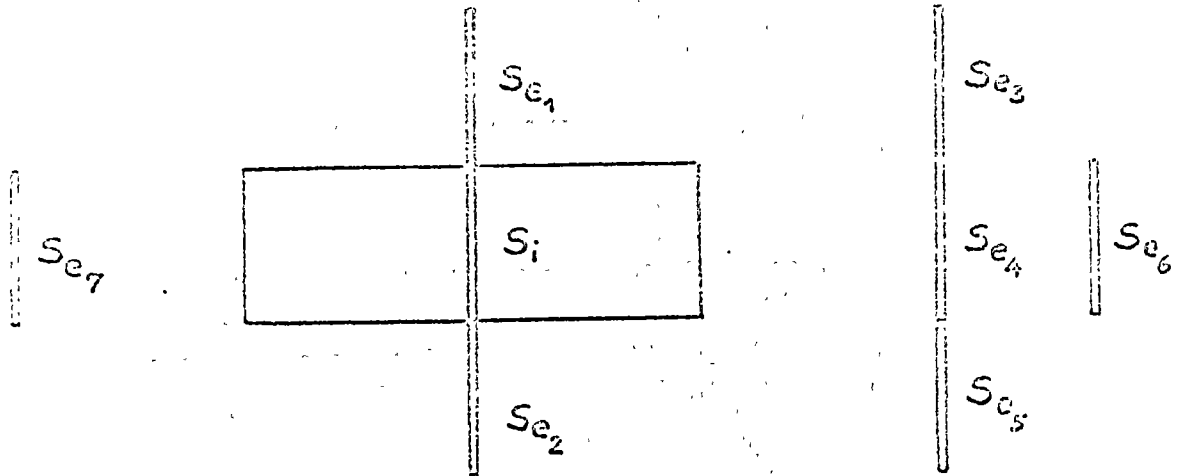
Este estimador presenta las 2 características
siguientes:

- es insesgado, es decir que en promedio el error es nulo:

$$E\{Z - Z^*\} = 0$$

KRIGEAGE

MEJOR ESTIMADOR LINEAL



$$Z^* = \lambda_1 S_i + \sum_{j=1}^7 \lambda_j S_{e_j}$$

INSESGADO

$$E\{Z - Z^*\} = 0$$

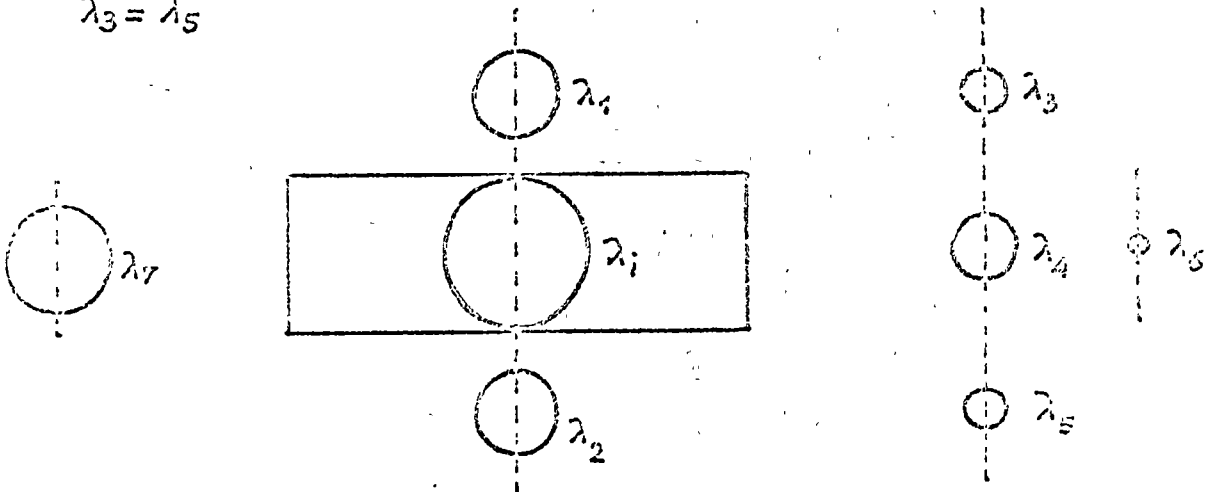
OPTIMO

$$E\{(Z - Z^*)^2\} = \sigma_K^2 \text{ MINIMA}$$

SIMETRIAS

$$\lambda_1 = \lambda_2$$

$$\lambda_3 = \lambda_5$$



CONDICION DE LA INFLUENCIA

$$\lambda_1 > \lambda_4$$

EFFECTO DE PANTALLA

$$\lambda_6 \ll \lambda_{e1}$$

es óptimo, es decir que los ponderadores λ se determinan de manera que la varianza de estimación $E\{(Z-Z^*)^2\}$ sea mínima.

El sistema de kriging, que entrega además de los ponderadores λ , la varianza de estimación mínima correspondiente, es un sistema lineal de ecuaciones que solamente hace intervenir integrales de γ , igual que la fórmula general:

En nuestro ejemplo, y en la hipótesis de isotropía estructural, el kriging encontrará

- las simetrías evidentes tales como:
$$\begin{cases} \lambda_{e_1} = \lambda_{e_2} \\ \lambda_{e_3} = \lambda_{e_5} \end{cases}$$

- las desigualdades tales como: $\lambda_i \geq \lambda_{e_k} \quad \forall k$

$\lambda_{e_3} \leq \lambda_{e_4}$: el sondaje S_4 hace pantalla a la influencia del sondaje S_3 .

$\lambda_{e_7} \leq \lambda_{e_4}$: el sondaje S_4 ha transferido una parte de su influencia sobre los sondajes próximos $S_{e_3}, S_{e_5}, S_{e_6}$.

La importancia de fenómenos como el "efecto de pantalla" o de transferencia de influencia depende de la regularidad de la regionalización, por consiguiente la del variograma 2γ .

La práctica del kriging permite:

- programar (día a día o al mes o al año, según las dimensiones de los paneles estimados) una producción minera.

- abordar el problema de la selección, es decir la definición de reservas.

IV-4 Estimación de reservas (selección)

El conjunto de los recursos in situ de un yacimiento presenta en muy pocos casos, características suficientes para permitir una explotación no selectiva. Para juzgar la rentabilidad del yacimiento o de una zona, o para decidir el mejor método de explotación, es necesario conducir a estudiar diversas selecciones posibles. El interés se centra entonces sobre la estimación de estas "reservas" seleccionadas.

Para abordar el estudio de las selecciones, es necesario guardar una primera noción esencial.

"La realidad es desconocida, y un estimador no se comporta como la realidad."

Caso 1:

Cada uno sabe que la ley de un testigo Y_L no es igual a la ley Z del panel que lo contiene: en particular, la dispersión de las leyes Y_L de los testigos, en el yacimiento G es bien superior a la dispersión de las leyes Z de los paneles en el mismo G . Así, en un yacimiento de oro, la ley Y_L de un testigo puede variar de 0 a 100% (pepita), mientras que la ley media Z de un panel de varios m^3 en el yacimiento variará mucho menos, por ejemplo entre 10^{-7} y 3×10^{-6} (0.1 y 3 gr/tonelada). Las dos centradas sobre la misma media:

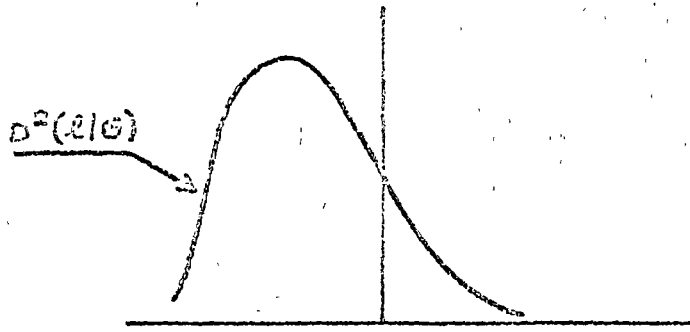
$E(Y_L) = E(Z) =$ ley media del yacimiento $= m$; las dos leyes de dispersión de Y_L y Z son fundamentalmente diferentes, solamente sus varianzas de dispersión están relacionadas por la relación de Krige, encontrada experimentalmente por Krige en Africa del Sur y formalizada por Matheron:

$$D^2(L|G) = D^2(L|P) + D^2(P|G) \quad \text{hipótesis estacionaria}$$

$$D^2(L|G) = D^2(Y_L)$$

$$D^2(P|G) = D^2(Z)$$

La varianza de dispersión de las leyes Y_L de los testigos L en el yacimiento G es igual a la varianza de dispersión de las leyes de estos testigos L en un panel P + varianza de dispersión de las leyes Z de estos paneles P en el yacimiento G .

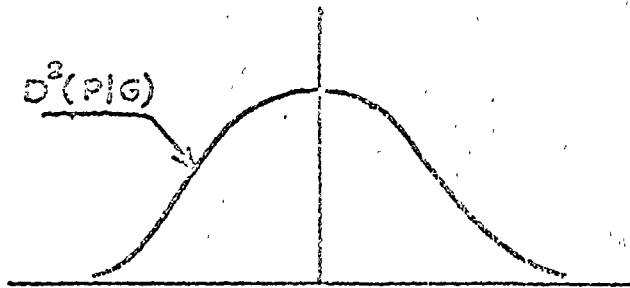


testigos

histograma experimental

$$\sigma^2(z|G) = \sigma^2(z|P) + \sigma^2(P|G) \quad \text{relación de Krige}$$

soportes diferentes



ley desconocida

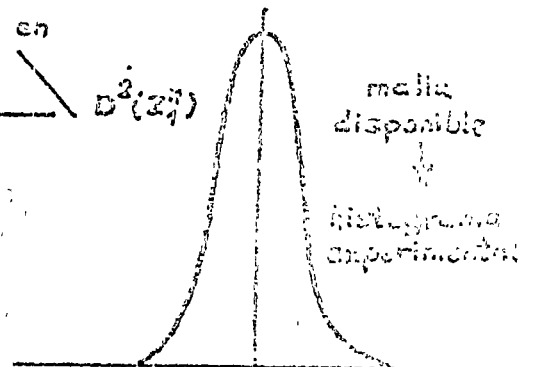
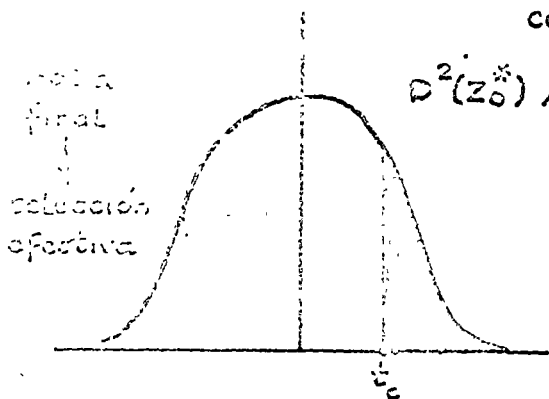
$D^2(z)$

mismo soporte

conservación de la ley en el límite

$D^2(z_0^*)$

$D^2(z_1^*)$



$$D^2(P|G) = D^2(z) > D^2(z_0^*) > D^2(z_1^*)$$

Así entonces, se puede estimar la media $E(Z)$ de las leyes de los paneles P , se puede conocer la dispersión $D^2(P|G)$ de estas leyes,

— sea, por cálculo geoestadístico formal, se demuestra:

$$D^2(P|G) = \bar{\gamma}(G,G) - \bar{\gamma}(P,P)$$

— sea, con la ayuda de la relación de Kriga, estimando experimentalmente los 2 términos $D^2(L|G)$ y $D^2(L|P)$.

Pero no se sabe nada de la ley de dispersión de las Z .

Algunos autores ajustan ciertas leyes teóricas (lognormales, gama, etc...) al histograma experimental de las leyes Y_2 de las muestras disponibles, luego, postulando un principio de conservación de ley, pretenden que la ley de las leyes Z de los paneles es del mismo tipo. En particular, el postulado de conservación de la normalidad es muy frecuente: en efecto en muchos yacimientos a base ley (Au, U, Cu, Ni) la dispersión experimental de las leyes de los testigos se ajusta bastante bien a una ley lognormal, además esta ley solo depende de 2 parámetros: media y varianza, y hemos visto que es posible calcular estos 2 parámetros en el caso de la dispersión de las leyes Z de los paneles. Pero no existe ninguna justificación teórica de este postulado de conservación de la lognormalidad; al contrario, numerosos ejemplos experimentales lo rechazan: las leyes Z de los paneles tienen tendencia a distribuirse según una ley normal truncada.

Si Y_2 , ley media conocida del testigo L no es la ley media desconocida Z del panel P , análogamente todo estimador Z^* formado a partir de la información $\{Y_2\}$ tampoco es la ley verdadera desconocida Z .

Si el estimador Z^* es insesgado, se tiene por definición:

$$E(Z) = E(Z^*) = m$$

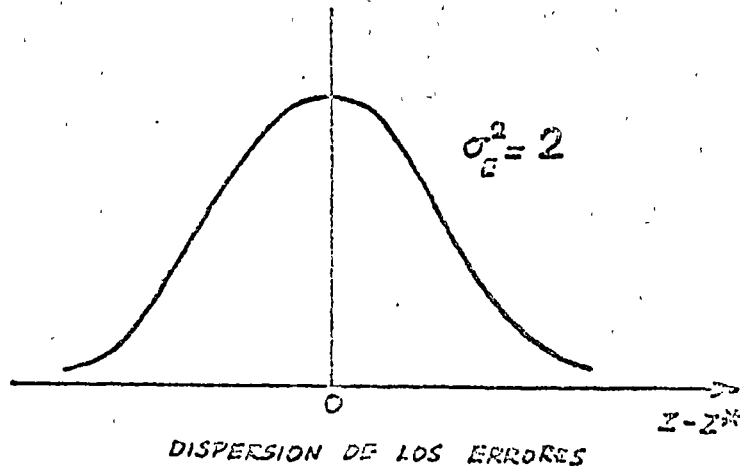
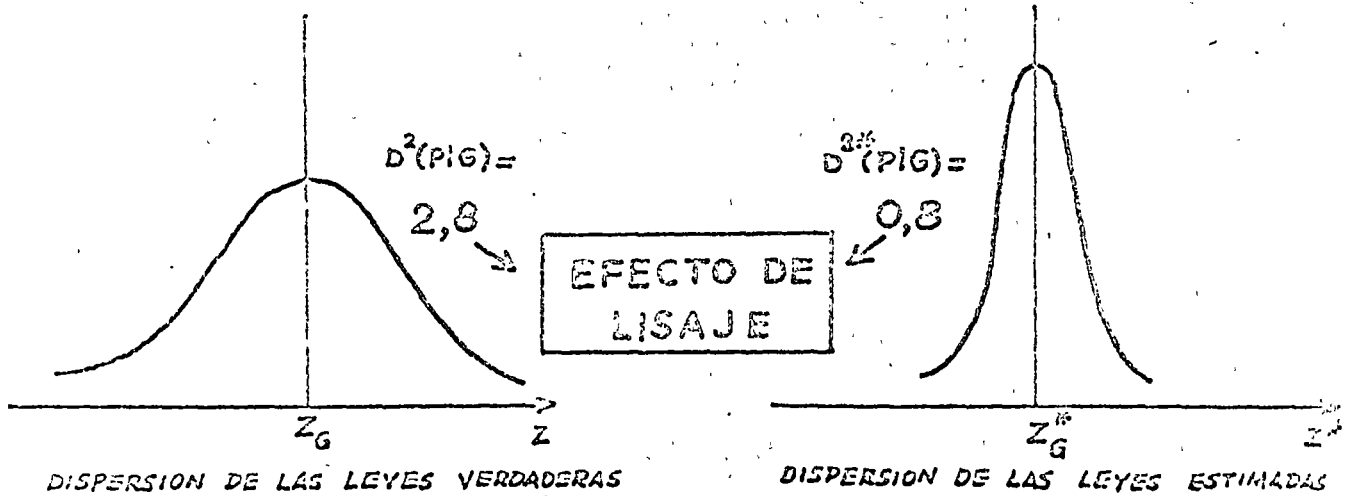
Las 2 leyes de dispersión de Z y de Z^* , aunque estrictamente centradas sobre la misma media, son diferentes en tipo

6	PANELES	
3	7	
3	6	

7	SOPORTE	
7	5	
5	6	

VALORES VERDADEROS Z DE LOS PANELES

VALORES ESTIMADOS Z^* (sobre el soporte) DE LOS PANELES



ley y en varianzas.

Respecto a las varianzas de dispersión, si y solo si el estimador Z_K^* es el estimador óptimo de Kriging Z_K^* , de varianza de estimación mínima, se demuestra la relación de lisaaje:

$$D^2(Z) = D^2(Z_K^*) + \sigma_K^2 - \sigma_M^2 \quad \text{hipótesis estacionaria.}$$

$D^2(Z) = D^2(P/G)$: varianza de dispersión de las leyes verdaderas Z
 $D^2(Z_K^*) = D^2(Z_K^*|G)$: varianza de dispersión de las leyes krigeadas Z_K^*
 $\sigma_K^2 = E\{[Z - Z_K^*]^2\}$: varianza de estimación de Z por Z_K^*
 $\sigma_M^2 =$ varianza de estimación de la media $E(Z)$, en general suficientemente pequeña para ser despreciada:

$$D^2(Z) = D^2(Z_K^*) + \sigma_K^2 \quad D^2(Z) \geq D^2(Z_K^*)$$

Vemos así como la dispersión experimental disponible $D^2(Z_K^*)$ queda alisada respecto de la dispersión verdadera $D^2(Z)$, y será tanto más alisada cuando la estimación de Z por Z_K^* sea más mala, es decir cuando σ_K^2 sea más grande.

En el límite, cuando la información disponible es insuficiente para permitir una estimación indiferenciada de cada uno de los paneles P , todos los estimadores Z_K^* son idénticos e iguales a la media de las informaciones disponibles. Queda entonces:

$$Z_K^* = m^* \quad \forall P \quad \Rightarrow \quad D^2(Z_K^*) = 0$$

$$\sigma_K^2 = E\{[Z - m^*]^2\} = E\{[(Z - m) + (m - m^*)]^2\} = D^2(Z) + \sigma_M^2$$

Inversamente en el otro límite, la estimación perfecta, es decir que: $Z_K^* = Z \quad \forall P$, queda entonces

$$\sigma_K^2 = \sigma_M^2 = 0 \quad \Rightarrow \quad D^2(Z) = D^2(Z_K^*)$$

Las 2 dispersiones real y krigeadas se confunden

Respecto a las varianzas de estimación, se tiene entonces, en el caso general:

$$D^2(Z) \geq D^2(Z_K^*)$$

En cuanto a los tipos de las leyes de dispersión, si bien se puede ajustar una ley teórica al histograma experimental disponible de los estimadores Z_k^* , nada se sabe del tipo de la ley de dispersión de los valores verdaderos Z .

Es necesario ver bien que la ley de dispersión de los estimadores Z_k^* cambia con el nivel de información. A medida que esta información se completa, se elaborarán krigesajes más y más precisos ($\sigma_k^2 \downarrow$), es decir que su dispersión $D^2(Z_k^*)$ tendrá tendencia a aproximarse más y más a la dispersión verdadera $D^2(Z)$. Cuando se alcanza un nivel de información tal que la diferencia relativa: $\frac{D^2(Z) - D^2(Z_k^*)}{D^2(Z)} \approx \frac{\sigma_k^2}{D^2(Z)}$ sea inferior a 20%,

se justifica esta vez de hacer la hipótesis de conservación del tipo de la ley de dispersión entre Z_k^* y Z .

Carson 26.

Hechas estas observaciones preliminares sobre las diferencias entre Y_2 , Z y Z^* , volvamos al problema preciso de la selección. Tomemos, a título de ejemplo, una selección por la ley de corte t_c : todo panel P , de geometría definida, es retenido si su ley media es superior a t_c . ¿Pero de cuál ley media se trata?

- no de la ley media verdadero Z desconocida
- tampoco de la ley estimada Z_k^* a partir de la información actual I , aún cuando esta estimación es óptima: Krigesaje $Z_{k_1}^*$.

- Sino la última ley estimada Z_0^* disponible en el momento de la selección efectiva: es en este último instante que se decidirá si explotar o no el panel P de acuerdo a que su estimador Z_0^* es superior o inferior a la ley de corte t_c .

Si se quiere entonces en el instante ①, estimar la recuperación correspondiente a la ley de corte t_c sobre unidades definidas P , se debe considerar solamente la dispersión $D^2(Z_0^*)$ y no $D^2(Z)$ o $D^2(Z_k)$ y aún menos $D^2(Y_2)$.

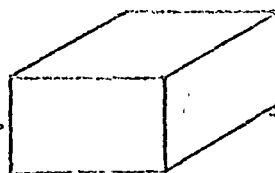
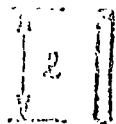
Pero solo se dispone experimentalmente de

- la ley de dispersión ajustada de los Y_2 , siendo I por ejemplo el soporte de algunos kilos de un testigo.
- la ley de dispersión ajustada de los $Z_k^* = Z_{k_1}$, siendo el soporte de estos estimadores precisamente el soporte

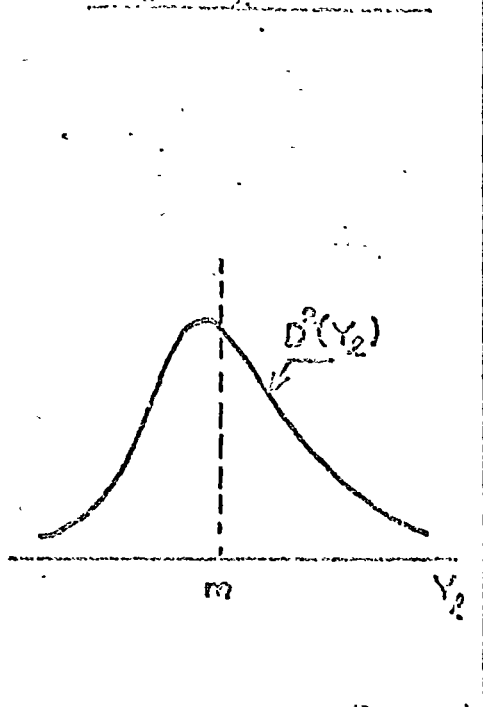
SELECCION Y RESERVAS RECUPERADAS

Unidad de Selección P

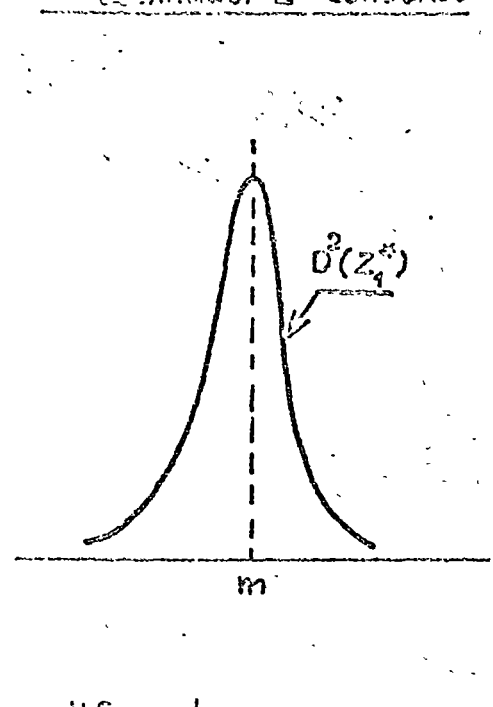
Testigo



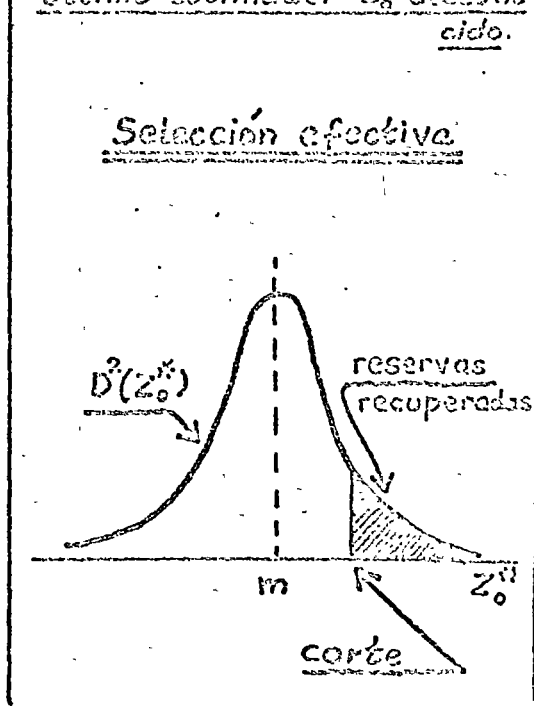
Ley Y_2 conocida



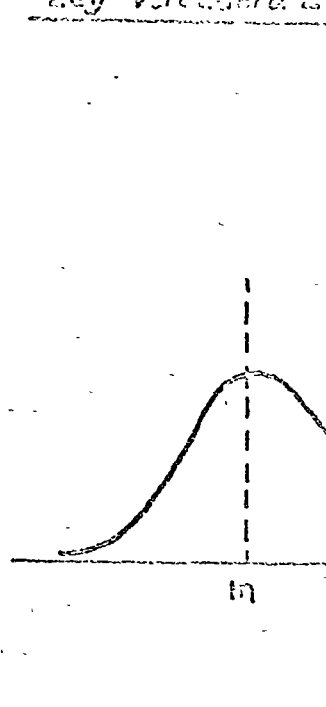
1er estimador Z_1^* conocido



último estimador Z_0^* desconocido



Ley verdadera Z



Selección efectiva

Soportes diferentes

Conservación de ley en el límite

mismo soporte

Conservación de ley en el límite

43

de la unidad de selección (algunas toneladas).
 de las relaciones, de Krige y de lisaje.

Como remontar a la ley de dispersión útil, es decir la de los Z_0^* :

- Se evalúa la última información disponible en el momento de la selección efectiva, se deduce la varianza de krigage correspondiente (Hemos visto que el cálculo de $\sigma_{K_0}^2$ puede realizarse antes de la obtención de la última información).

- Se calcula la varianza de dispersión $D^2(Z_0^*)$ de las leyes verdaderas, sea con la ayuda de la relación de Krige, sea directamente con la ayuda del variograma.

- Se deduce de la relación de lisaje:

$$D^2(Z_0^*) = D^2(Z) - \sigma_{K_0}^2$$

- Si $D^2(Z_1^*)$ difiere poco de $D^2(Z_0^*)$, se hace la hipótesis justificada de conservación de tipo de la ley de dispersión entre Z_1^* y Z_0^* .

- Se traza la ley de dispersión teórica buscada de los Z_0^* , de media $E(Z_0^*) = E(Z_1^*)$ insesgada de varianza $D^2(Z_0^*)$.

- Se aplica la ley de corte t_c sobre esta ley $D^2(Z_0^*)$ para evaluar la proporción de ponetes P que conservará la selección efectiva t_c .

Se repite un trabajo análogo para una serie de hipótesis de selección:

- diversos parámetros de selección: ley - potencia
- destecho - impurezas etc...
- diversos niveles de información última disponible.
- diversas unidades P de selección

La comparación de los resultados que salen de estas diversas hipótesis de selección permite:

- juzgar la rentabilidad del yacimiento
- decidir sobre el tipo de selección a operar
- si esta selección está impuesta a priori, calcular la información última necesaria para optimizarla.

Declaro: Las reservas del yacimiento, es decir el resultado de una selección sobre los recursos dependen:

- no solamente de la selección considerada.
- sino también del nivel de información disponible.

Insistamos sobre la formalización rigurosa que la Geostatística ha aportado a este difícil problema de estimación de reservas seleccionadas. En estos procedimientos el rigor es necesario, bajo la pena de sesgar gravemente (sobre-estimación sistemática) la estimación de reservas.

Nº 5 Estudio de dispersiones.

Falta por abordar un último problema-tipo, es el del estudio de las fluctuaciones de las características de las reservas explotadas:

- Las fluctuaciones de las leyes en impurezas, por ejemplo, interesan directamente al preparador de mineral o al químico.
- Las fluctuaciones en potencia de deshecho, por ejemplo, interesan directamente al ingeniero de explotación.

Para esto, la Geoestadística propone la técnica de las simulaciones condicionales:

Hemos visto que el conjunto de leyes reales $\{y(x), x \in G\}$ puede ser interpretado como una realización particular de una Función Aleatoria $Y(x)$, de la cual se conoce, por estimación, los 2 primeros momentos estructurales:

- la media $E\{Y(x)\}$.
- el variograma $2\gamma(h) = E\{[Y(x+h) - Y(x)]^2\}$.

Continúa

Entonces la idea es de simular una nueva realización $\{y_s(x), x \in G\}$ de esta F.A. $Y(x)$. Esta nueva realización $y_s(x)$ no será la realidad $y(x)$, ni tampoco un estimador $y^*(x)$ de esta realidad, sino una nueva variable regionalizada, "semejante" a la realidad $y(x)$ de acuerdo a que:

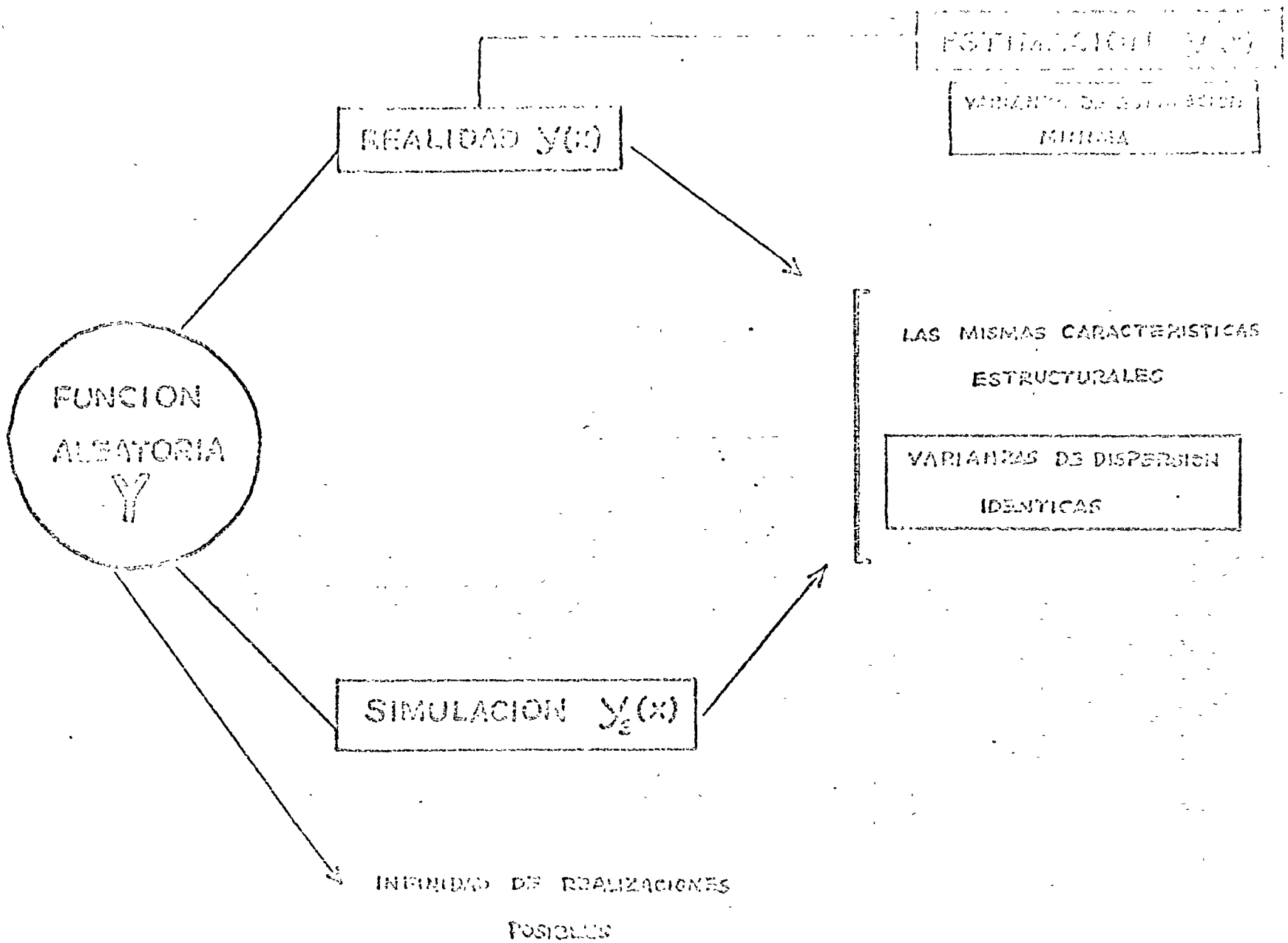
- presenta las mismas características estructurales, al menos hasta el 2º orden:

mismo momento de 1º orden, media o derivada (tendencia);
mismo momento de 2º orden de los incrementos (variograma).

- encuentra en los puntos x_i de implantación de la información disponible los valores reales:

$$y_s(x_i) = y(x_i) \quad \forall x_i \in I$$

- se puede además, imponer al histograma de valores simulados $y_s(x)$ de encontrar al histograma experimental de los datos reales disponibles $\{y(x_i), x_i \in I\}$.



49

capítulo 17

La simulación $y_s(x)$ puede ser hecha sobre una malla tan densa como se quiera, las únicas limitaciones son las del computador. Por ejemplo, condicionadamente a una malla disponible de $100m \times 100m$, se puede realizar una simulación a la malla de $5m \times 5m$, u otra malla aún más fina. Sobre esta nueva realidad simulada y conocida en forma casi perfecta, se puede aplicar un esquema de producción (desoxidación-selección-acarreo-almacenamiento etc...) y estudiar las consecuencias de este esquema sobre el aval: fluctuaciones de las leyes a la entrada de la planta, o capacidad de homogenización de un área de almacenamiento, o necesidad de explotar varias canteras a la vez etc...

Corollary 14

Insistamos sobre el hecho que una simulación $y_s(x)$ no es una estimación $y^*(x)$. Se le demanda al estimador $y^*(x)$ de ser en promedio lo más próximo posible de la realidad desconocida $y(x)$, es decir que se busca minimizar la varianza de estimación o error cuadrático medio: $E\{[Y(x) - Y^*(x)]^2\}$ minimal \rightarrow kriging.

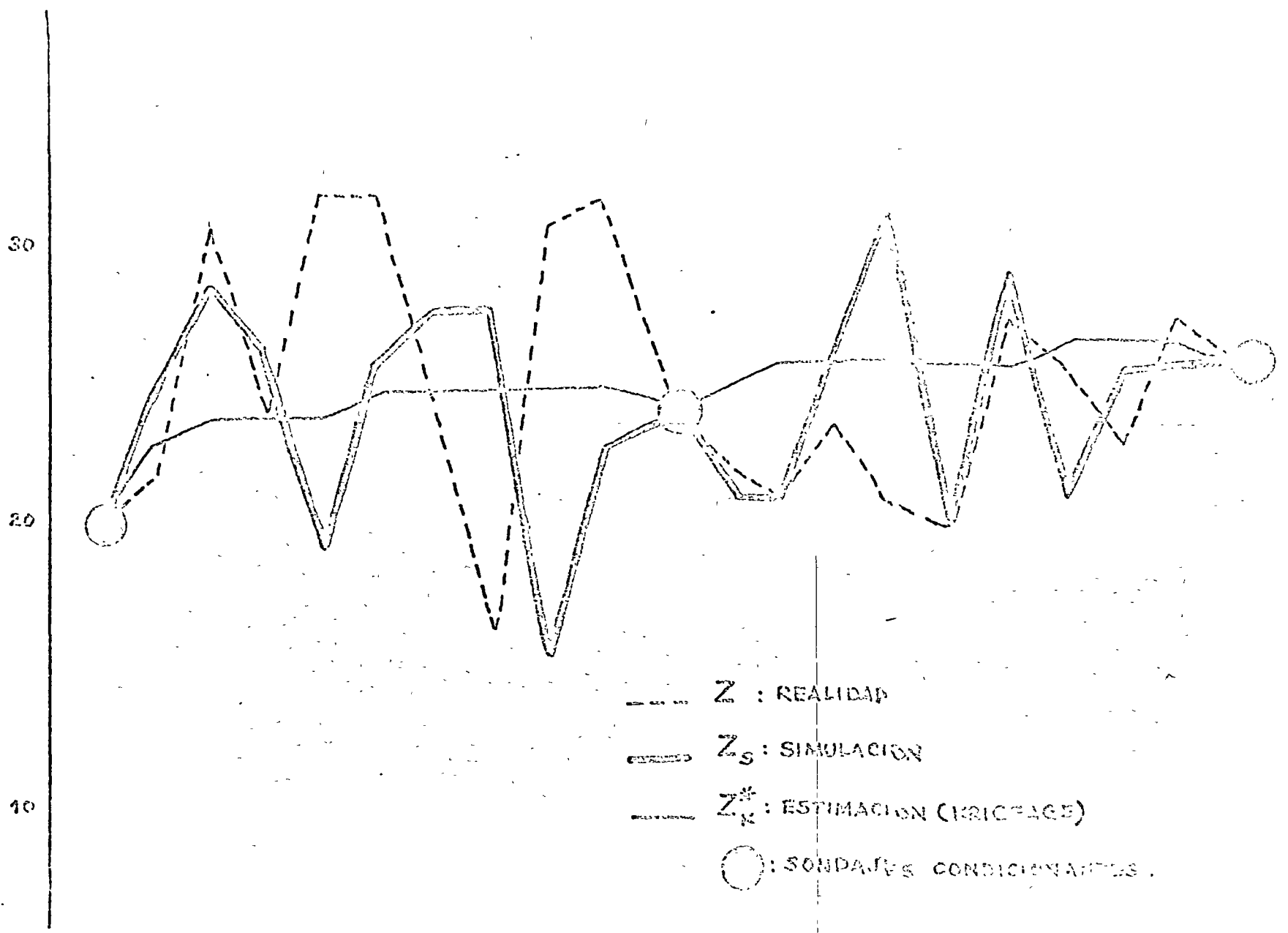
Pero este estimador $y^*(x)$, siendo óptimo, no presenta las mismas características estructurales que la realidad $y(x)$, en particular el fenómeno estimado $y^*(x)$ es mucho más alisado que la realidad $y(x)$:

$$D^2(y^*) \leq D^2(y)$$

Al contrario la simulación $y_s(x)$ no pretende ser un estimador óptimo de la realidad $y(x)$, pero va a presentar las mismas características estructurales que esta realidad, en particular la dispersión de los valores simulados es igual a la dispersión de los valores reales: $D^2(y_s) = D^2(y)$.

En los problemas de estimación, interesa la varianza de estimación que se busca a minimizar (kriging).

En los problemas de fluctuaciones, interesa la varianza de dispersión que se busca a identificar (simulaciones condicionales).



- Z : REALIDAD
- ... Z_S : SIMULACION
- Z_K* : ESTIMACION (KRICFACE)
- : SONDAJE'S CONDICIONANTES.

57

IV-6 Corregionalizaciones.

Hasta aquí, nos hemos interesado solamente en una sola variable regionalizada a la vez, por ejemplo, la ley $y(x)$ en el punto x . Se caracterizaba entonces la autocorrelación de las variables $y(x)$ por la función estructural "variograma":

$$2\gamma(h) = E\{[Y(x) - Y(x+h)]^2\}$$

Sin embargo se puede generalizar todo el formalismo y los resultados anteriores al estudio simultáneo de 2 variables regionalizadas diferentes, por ejemplo:

- La ley en Pb $y_1(x)$ en el punto x
- La ley en Zn $y_2(x)$ en el mismo punto x .

Cada una de estas 2 variables posee su propia función de autocorrelación variograma:

$$2\gamma_1(h) = E\{[Y_1(x) - Y_1(x+h)]^2\} \quad \text{para el Plomo}$$

$$2\gamma_2(h) = E\{[Y_2(x) - Y_2(x+h)]^2\} \quad \text{para el Zinc}$$

Pero se sabe que existen correlaciones cruzadas entre las leyes de Pb y en Zn; se definirá entonces, y se estimará el variograma cruzado Pb-Zn:

$$2\gamma_{12}(h) = E\{[Y_1(x) - Y_1(x+h)][Y_2(x) - Y_2(x+h)]\}$$

que caracterizará la correlación espacial cruzada entre los incrementos de las 2 variables Pb y Zn.

Si bien los autores clásicos admiten la existencia de una correlación local ($h=0$) entre las 2 variables $y_1(x)$ e $y_2(x)$:

$$r_{12} = \frac{E\{Y_1(x)Y_2(x)\}}{\sqrt{E(Y_1^2)E(Y_2^2)}} \quad \text{coeficiente de correlación,}$$

ellos ignoran en general, la correlación espacial entre $y_1(x)$ e $y_2(x+h)$ del mismo modo que ignoraban la autocorrelación espacial entre $y_1(x)$ e $y_1(x+h)$. Del mismo modo que el variograma directo $2\gamma_1(h)$ caracteriza la autocorrelación espacial de las variables y_1 , el variograma cruzado $2\gamma_{12}(h)$ caracteriza la correlación espacial de las 2 variables y_1 e y_2 .

COVARIANZA

	COVARIANZA	CORRELACIONAL	SEMI-VARIANZA
<p>UNA SOLA FUNCION ALEATORIA Y</p>	$C(h) = E[Y(x)Y(x+h)] - m^2$	$\rho(h) = \frac{C(h)}{C(0)}$ $\rho(0) = 1$	$\gamma(h) = C(0) - C(h)$
<p>caso particular independencia espacial $\forall h \neq 0$</p>	$C(h) = 0$	$\rho(h) = 0$	$\gamma(h) = C(0)$
<p>DOS FUNCIONES ALEATORIAS Y_1 e Y_2</p>	$C(h) = \begin{bmatrix} C_1(h) & C_{12}(h) \\ C_{12}(h) & C_2(h) \end{bmatrix}$ $C_{12}(h) = E[Y_1(x)Y_2(x+h)] - m_1 m_2$	$\rho(h) = \begin{bmatrix} \rho_1(h) & \rho_{12}(h) \\ \rho_{12}(h) & \rho_2(h) \end{bmatrix}$ $\rho_{12}(h) = \frac{C_{12}(h)}{\sqrt{C_1(0)C_2(0)}}$ $\rho_{12}(0) \leq 1$	$\gamma(h) = \begin{bmatrix} \gamma_1(h) & \gamma_{12}(h) \\ \gamma_{12}(h) & \gamma_2(h) \end{bmatrix}$ $\gamma_{12}(h) = C_{12}(0) - C_{12}(h)$
<p>caso particular independencia espacial $\forall h \neq 0$</p>	$C_{12}(h) = 0$	$\rho_{12}(h) = 0$	$\gamma_{12}(h) = C_{12}(0)$

51

El conocimiento de la matriz estructural

$$\begin{bmatrix} \gamma_1 & \gamma_{12} \\ \gamma_{12} & \gamma_2 \end{bmatrix} \quad \text{o más general} \quad \begin{bmatrix} \gamma_{ij} \end{bmatrix} \quad i, j = 1, m$$

con m variables

permite las co-estimaciones cruzadas. Por ejemplo:

- La estimación de la ley en Pb $y_1(x)$ desconocida en el punto x con la ayuda de las leyes en Pb y en Z en vecinas $\{y_1(x_i), y_2(x_i), x_i \neq x\}$

- En el caso de un reconocimiento heterogéneo, la estimación de una misma variable con la ayuda de 2 conjuntos heterogéneos de información:

se dispone de 2 tipos de análisis en F_a

$$\begin{cases} y_1(x_i) & \text{Ley de testigos, precisa} \\ y_2(x_i) & \text{Ley de pozos, muy imprecisa} \end{cases}$$

$x_i \in I$, información

En resumen, si se trata de la regionalización simultánea de varias variables, o de la regionalización directa de una sola variable, la Geostatística encuentra todos los resultados de la Estadística clásica como casos extremadamente particulares correspondientes a la independencia espacial.

- BIBLIOGRAPHIE -

- BIBLIOGRAFIA -

Indice des travaux géostatistiques les plus importants

* Indica los trabajos geostatísticos más importantes.

- [1] - AGTERBERG, P.P.- Autocorrelations functions in geology (1970) in "Geostatistics, a colloquium" - pp. 113-141.- Ed.: D.P. Merriam, Plenum Press, New York.
- [2] - HAAS, R. and CARLIER, A.- Applications of Geostatistics in ore valuation, ore reserve estimation and grade control (1968) in CIMM, Special Vol. N° 9, pp. 41-68, Montreal.
- [3] - SUBNICEK, L. and HAAS, A.- Method of calculation of the iron ore reserve in the Lorraine deposit (1969).- AFME Special Vol. "A decade of digital Computing" pp. 179-210, New York.
- [4] - CARLIER, A.- Contribution aux méthodes d'estimation des gisements d'uranium (1964).- Doctoral thesis - 359 p. French Commissariat à l'Energie Atomique Report n° 2332 - Fontenay-aux-Roses.
- [5] - DAVID, M.- The notion of extension variance and its application to the grade estimation of stratiform deposits (1969) AFME Special Vol. "A decade of digital computing" pp. 63-81, New York.
- [6] - DAVID, M.- Tools for planning variances and conditional simulations.- 11th APCOM Symp.- Ed.: College of Mines, University of Arizona, 1973.
- [7] - DESHOMME, J.P. and DEIMINER, P.- Application du krigeage à l'optimisation d'une campagne pluviométrique en zone aride.- Ed. : UNESCO - Proceedings of Symposium Dewarpid, Madrid, June 1973.
- [8] - BOWD, P.- A digigraphic console program for kriging techniques (1973).- 11th APCOM Symp. Ed.: College of Mines, University of Arizona.
- [9] - GUILAL, D.- Simulation de scénarios intrinsèques (1972).- Ed.: CIMM, Ecole Nat. Sup. des Mines, Paris.
- [10] - GUILAL, D.- L'estimation des okoués du Gabon (1971).- Ed.: CIMM, Ecole Nat. Sup. des Mines de Paris.
- [11] - HARRBAUGH, J.W.- Computer simulation in Geology (1970).- Ed.: Wiley and Sons, New York.
- [12] - HODGSON, W.A.- Optimum spacing for soil sample traverses (1972) 10th APCOM Symp. pp. 75-78.- Ed.: South Afr. Inst. of Min. and Metall., Johannesburg.

- [13] - HUIJBREGTS, Ch. and MATHERON, G.- Universal Kriging, an optimal approach to trend surface analysis (1970) CIMM Special Vol. N° 12, pp. 159-169 - Montreal.
- [14] - HUIJBREGTS, Ch. and JOURNEL, A.- Simulation du gisement de Nickel latéritique de Prony (1972).- Unpublished Research Note - PENAMAX Co. and Centre de Morphologie Mathématique, Fontainebleau.
- [15] - HUIJBREGTS, Ch. and SINGOVIA, R.- Geostatistics for the valuation of a copper deposit in Chile (1973) 11th APCOM Symp.- Ed.: College of Mines, University of Arizona.
- [16] - HUIJBREGTS, Ch.- Regionalized Variables and applications to quantitative analysis and estimation of spatial data (1973).- NATO Advanced Study Institute - Nottingham Conf. July 1973.
- [17] - JOURNEL, A.G.- Etudes sur l'estimation d'une variable régionalisée, application à la cartographie sous-marine (1969).- Ed.: Service Hydrographique de la Marine, Paris.
- [18] - JOURNEL, A.G.- Geostatistics and Sequential Exploration (1973) in "Mining Engineering", Oct. 1973
- [19] - JOURNEL, A.G. and HUIJBREGTS, Ch.- Estimation of lateritic-type deposit (1972).- 10th APCOM Symp.- pp. 207-212.- Ed.: South Afr. Inst. of Min. and Metall. Johannesburg.
- [20] - JOURNEL, A.G.- Le formalisme des relations ressources-Réserves - Simulation de gisements miniers (1973) in "Revue de l'Industrie Minière" N° 4, 1973.
- [21] - KOCH, G.S. and LINK, R.F.- Statistical Analysis of Geological data (1970).- Ed.: Wiley and Sons, New York.
- [22] - KOCH, G.S. and LINK, R.F.- An exploration model for tabular orebodies (1972).- 10th APCOM Symp.- pp. 189-192, Ed.: South Afr. Inst. of Min. and Metall., Johannesburg.
- [23] - KRIGE, D.G.- Two-dimensional weighted moving average trend surfaces for ore evaluation (1966).- Journal of the South Afr. Inst. of Min. and Metall.- March 1966
- [24] - KRIGE, D.G.- The role of mathematical statistics in improved ore valuation techniques in South African gold mines (1970) in "Topics in Mathematical Geology" Ed.: Consultants Bureau, New York, London.
- [25] - KRIGE, D.G.- Computer applications in investment analysis, ore valuation and planning for the Prichard copper mine (1971).- 11th APCOM Symp., Ed.: College of

- [26] - MARCHEL, A. and SERNA, J.- Random kriging (1970).- in "Geostatistics, a Colloquium" pp. 91-112.- Ed.: D.F. Merriam, Plenum Press, New York.
- [27] - MARCHEL, A. and ROULHIER, J.P.- Etude géostatistique des gisements de bauxite (1970).- Revue de l'Industrie Minière, Vol. 52 N° 7, pp. 492-507.
- [28] - MARCHEL, A.- Boletín de Geoestadística, Vol. 1-2-3-4-5 (1972-1973).- Ed.: Universidad de Santiago, Chile.
- [29] - MATHÉRON, G.- Application des méthodes statistiques à l'estimation des gisements (1955).- in Annales des Mines.- Dec. 1955, pp. 50-75.
- [30] - MATHÉRON, G.- Théorie géométrique de l'échantillonnage systématique des gisements (1957).- in Annales des Mines, Vol. 9, pp. 566-584.
- [31] - MATHÉRON, G.- Precision of exploring a stratified formation by boreholes with rigid spacing.- Application to a bauxite deposit (1962).
- [32] - MATHÉRON, G.- Principles of Geostatistics (1963).- in Economic Geology, Vol. 58, pp. 1246-1256.
- [33] - *MATHÉRON, G.- Traité de Géostatistique appliquée.- Tome 1 (1962) Tome 2 (1963).- Ed.: Technip, Paris.
- [34] - MATHÉRON, G. et FORMERY, Ph.- Recherche d'optima dans la reconnaissance et la mise en exploitation de gisements miniers.- in Annales des Mines, Mai-Juin 1963.
- [35] - *MATHÉRON, G.- Les Variables Régionalisées et leur estimation (1965).- Ed.: Masson, Paris.- 305 pp.
- [36] - MATHÉRON, G.- Osnovy Prikladnoi geostatistiki (1968).- Ed.: Mir, Moscow.
- [37] - *MATHÉRON, G.- Le Krigeage Universel (1969).- Les Cahiers du Centre de Morphologie Mathématique.- Booklet N° 1, 83 pp.- Ed.: CMM, Ecole Nat. Sup. des Mines, Paris.
- [38] - *MATHÉRON, G.- The Theory of Regionalized Variables, and its applications (1971).- Les Cahiers du Centre de Morphologie Mathématique, booklet N° 5, 211 pp.- Ed.: CMM; Ecole Nat. Sup. des Mines, Paris.
- [39] - MATHÉRON, G.- The intrinsic random functions and their applications.- in Advances in Applied Probability, Dec. 1975.
- [40] - KRIGEN, M.- Spatial Variation (1951).- Almqvist Forlaget, Stockholm.
- [41] - KRIGEN, M.- A new statistical treatment of gravity anomalies (1977) - Journal of the Int. Assoc. of Geodesy

[38] - NEFFEN, M.S.- The application of Geostatistics to mine sampling patterns (1975).- 11th APCOM Symp.- Ed.: College of Mines, University of Arizona.

[39] - OLGA, R.- Application of regionalized variable theory to automatic contouring (1972).- Special Report to the Amer. Petrol. Inst. Research project N° 131.

[40] - RANDBO, J.M.- Some applications of Geostatistics to decision-making in exploration (1970).- OIOM Special Vol. N° 12, pp. 179-184.

[41] - ROYCE, A.G.- A practical introduction to Geostatistics (1971).- Course of the University of Leeds, England.

[42] - *SARRA, J.- Méthantillonnage et Estimation locale des phénomènes de transition miniers (1967).- Doctorate Thesis (2 Tomes, 650 pp). Ed.: University of Nancy, France.

[43] - TERKI, J.- Les structures gigognes : morphologie mathématique et interprétation métallogénique (1968).- in Mineral. Deposita (Berlin) 3, pp. 135-154.

[44] - THIEL, H.S.- The estimation of means and associate confidence limits for small samples from lognormal populations (1966).- Journal of the South Afr. Inst. of Min. and Metall.- March 1966.

[45] - THOMAS, H.S.- Statistical valuation of diamondiferous deposits (1972).- 10th APCOM Symp., pp. 17-25.- Ed. : South Afr. Inst. of Min. and Metall., Johannesburg

[46] - SINCLAIR, A.J.- A geostatistical study of the Eagle copper vein, N. British Columbia (1973).- Canadian Inst. of Mining and Metallurgy.

[47] - VERMA, B.K.- Efficiency of regression techniques for ore valuation at Kolar gold fields, India (1972).- Journal of the IAMG, Vol. 4, N° 1, March 1972.

[48] - WATSON, G.S.- Trend surface analysis (1971).- Mathematical Geology, Vol. 3 N° 3, pp. 215-226.

[49] - WATSON, G.S.- Trend surface analysis and spatial correlation (1972).- The Geological Society of America, special paper n° 146.

[50] - WELSH, A. and KRIGER, D.G.- Reports on Exploration sessions (1971) 10th APCOM Symp., pp. 271.- Ed.: South Afr. Inst. of Min. and Metall.- Johannesburg.

~~... géostatistiques appliquées, contenant des~~
~~... publiées par les journaux de la Société Géologique de France~~

... géostatistiques appliquées, contenant des

DIRECTORIO DE ASISTENTES AL CURSO DE GEOESTADISTICA

DEL 22 AL 27 DE ABRIL DE 1974

<u>NOMBRE Y DIRECCION</u>	<u>EMPRESA Y DIRECCION</u>
1. ING. RAYMUNDO BARCENA L. Gómez Pérez 107 Pachuca, Hgo.	INSTITUTO NACIONAL DE ENERGIA NUCLEAR Insurgentes No. 1180 México, D. F. Tel: 5-75-94-02
2. SR. JUAN MANUEL BERLANGA GUTIERREZ Duque Job No. 11 2a. Col. Periodista México 13, D. F. Tel: 5-39-55-93	INSTITUTO MEXICANO DEL PETROLEO Av. Cien Metros No. 152 México, D. F. Tel: 5-67-55-00
3. ING. FRANCISCO JOSE ESCANDON VALLE Av. del Parque No. 210 San Luis Potosi, S. L. P.	ASARCO MEXICANA, S. A. Av. Baja California No. 200 7o. Piso México 7, D. F. Tel: 5-74-67-67
4. ING. MARIO A. GUTIERREZ ESTRADA Av. Rio Churubusco 274 Depto. C-8 Col. Carmen, Coyoacán México 21, D. F. Tel: 5-44-59-13	CENTRO CIENCIAS MARINAS, UNAM Ciudad Universitaria México 20, D. F. Tel: 5-48-27-66
5. SR. ALEJANDRO JOSE LIRA BLANCO Quemada 477-102 Col. Vertiz Narvarte México, D. F. Tel: 5-39-04-34	COMISION DE FOMENTO MINERO Puente de Tecamachalco No. 26 México 10, D. F. Tel: 5-40-34-00
6. ING. OMAR MARTINEZ TREJO Av. Industria Militar 305-F-12 México 10, D. F. Tel: 5-89-49-53	INSTITUTO MEXICANO DEL PETROLEO Av. 100 Metros 152 México, D. F. Tel: 5-67-66-00 Ext. 239
7. ING. GUSTAVO DE J. PEREZ DURAN Av. San Antonio 42-9 Col. Nápoles México, D. F.	INSTITUTO MEXICANO DEL PETROLEO Av. 100 Metros No. 152 Col. Lindavista México, D. F. Tel: 5-67-66-00 Ext. 242

DIRECTORIO DE ASISTENTES AL CURSO DE GEOESTADISTICA

DEL 22 AL 27 DE ABRIL DE 1974

NOMBRE Y DIRECCION

EMPRESA Y DIRECCION

8. ING. ALFONSO SALAZAR CARDENAS
Canoras 47
Las Arboledas
México, D. F.
Tel: 5-65-47-68

SOSA TEXCOCO, S. A.
Paseo de la Reforma 213-6o.P.
México, D. F.
Tel: 5-91-06-99

9. ING. HEBERTO A. VILLARREAL CONTLA
Av. Coyoacán 1116-404
México 12, D. F.

COMISION DE ESTUDIOS DEL
TERRITORIO NACIONAL
Av. San Antonio Abad 124
México 8, D. F.
Tel: 5-78-62-00