

G E O E S T A D I S T I C A

Fecha	Duración	Tema	Profesor			
Lunes 14	8 a 12 y 13 a 15 Hs.	Repaso de Estadística y Probabilidad Introducción a la Geoestadística El Variograma: definición y propiedades. Tarea práctica: construcción de un variograma experimental.	Prof. Dr. Henri Sans			
Martes 15	8 a 12 y 13 a 15 Hs.	Las Varianzas de la Geoestadística: varianza de dispersión, varianza de extensión: definición y propiedades. Relación de KRIGE. Principios de aproximación. Ejercicios de aplicación. Tarea práctica (caso real): Estimación global del yacimiento de oro de Ity Monte-Flotuo (Africa)	"	"	"	"
Miércoles 16	8 a 12 y 13 a 15 Hs.	Estimación local: el método del Krigeage: definición y propiedades. Aplicación al cálculo de las reservas económicas de un yacimiento. Ejercicios.	"	"	"	"
Jueves 17	8 a 12y 13 a 15 Hs.	Las simulaciones de yacimientos minerales: objetivos y principios. Las decisiones secuenciales.				
Viernes 18	8 a 12 y 13 a 15 Hs.	Estudios detallados de casos reales: aplicación de la Geoestadística	"	"	"	"
Sábado 19	8 a 12 y 13 a 15 Hs.	Los ejes de desarrollo de la geoestadística, ejemplo: el control de la producción del yacimiento de fosfatos de TOGO (Africa)	"	"	"	"



CURSO INTENSIVO DE GEOESTADISTICA

Centro de Educación Continua

Octubre de 1974

Dr. Henri SANS

CONTENIDO

	Pág.
* La Hipótesis Intrínseca. El Variograma - - - - -	1
Aplicación práctica: Construcción de un Variograma experimental- - -	- 10
* Las Varianzas de la Geoestadística - - - - -	15
Varianza de Estimación. Varianza de Extensión.	
Fórmulas de aproximación.	
Aplicaciones prácticas: - Varianza de estimación - - - - -	23
- Funciones auxiliares - - - - -	27
* Estimación global del yacimiento de ITV-Monte Flotuo - - - - -	31
* El Krigeage - - - - -	40
Aplicación práctica: Krigeage de la ley promedio de un segmento -	45
* Las Decisiones Secuenciales - - - - -	48
* Estudio real: Veta de cobre de Eagle (CANADA) - - - - -	55
* Las Relaciones Recursos- Reservas - - - - -	59
* Las simulaciones CONDICIONALES - - - - -	66
* Estudio Real: Yacimiento de Niquel de Prony (Nueva Caledonia) -	71
Ejemplo de reconocimiento secuencial de yacimiento.	
* Estudio Real: Yacimiento de Fosfatos (TOGO) - - - - -	75
Ejemplo de control geoestadístico de yacimiento.	
* Bibliografía	

LA HIPOTESIS INTERSECA - EL VARIOGRAMA

I.- INTERPRETACIÓN DE UNA V.R.

* definición de una Función Aleatoria (F.A.)

una F.A. es un conjunto infinito de Variables Aleatorias (V.A.), generalmente no independientes.

- en cada punto x del espacio que se considera, hay una V.A.

$Y(x) \rightarrow$ aspecto ALEATORIO de la V.R.

- en dos puntos x_1 y x_2 , las V.A. $Y(x_1)$ e $Y(x_2)$ no son independientes \rightarrow aspecto ESTRUCTURAL de la V.R.

El conjunto $Y = \{ Y(x_1), Y(x_2), \dots \}$ es la F.A.

* definición de la REALIZACION

- Sea la V.A. X

Se hace un tiro: $X=5$, valor particular de la V.A. X , es una realización de la V.A. X .

Se hace otro tiro: $X=98$ es otra realización particular de la V.A. X

- Sea la F.A. Y

Se hace un tiro: $Y = \{ Y(x_1)=9, Y(x_2)=12, \dots \}$ es una función numérica llamada realización particular de la F.A. Y

Si se hace otro tiro, se obtiene otra función numérica, generalmente diferente, o sea otra realización de la F.A.

* Una V.R. es una realización de una cierta F.A. = es la hipótesis básica que hacemos.

II.- INFERENCIA ESTADISTICA

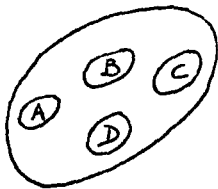
- V.A. X . A partir de una sola realización de X , es imposible alcanzar la ley de distribución de la V.A. X .

- F.A. Y . A partir de una sola realización $Y = \{ Y(x_1), Y(x_2), \dots \}$

es imposible de alcanzar la ley temporal de Y o sea de de-
terminar los parámetros probabilísticos de Y

Necesidad de hacer hipótesis compatibles con la realidad física
que se estudia a fin de tener acceso a la ley temporal o por lo
menos a la parte de la ley temporal realmente necesaria para lo-
grar nuestros objetivos

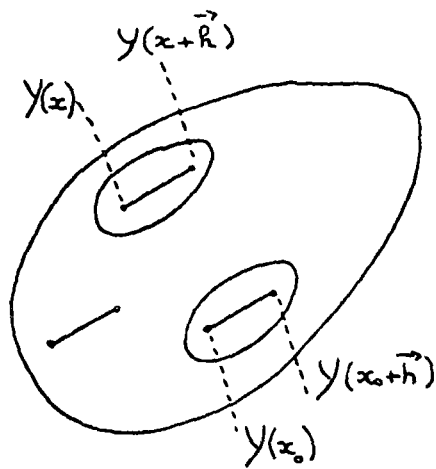
III.- ESTACIONARIDAD DE LA F.A. Y



Las características estadísticas de las zo-
nas A, B, C, D, son iguales.

Hipótesis fuerte que no se puede admitir --
siempre en la práctica

IV.- ESTACIONARIDAD DE LOS INCREMENTOS DE Y



No se consideran $Y(x)$, $Y(x+\vec{h})$ sino más
bien el incremento $\{Y(x+\vec{h})-Y(x)\}$.

El incremento $[Y(x+\vec{h})-Y(x)]$ es una V.A.

Hipótesis Intrínseca = los incrementos son
estacionarios o sea $[Y(x+\vec{h})-Y(x)]$ e

$[Y(x_0+\vec{h})-Y(x_0)]$ son dos realizaciones de
la misma variable aleatoria $[Y(x+\vec{h})-Y(x)]$

(o $[Y(x_0+\vec{h})-Y(x_0)]$ es igual)

La parte de la ley temporal que interesa es:

El momento de orden 1 $E \{Y(x+\vec{h})-Y(x)\} = m(\vec{h})$

El momento de orden 2 $D^2 \{[Y(x+\vec{h})-Y(x)]^2\} = 2\gamma(\vec{h})$

Estos momentos no dependen del punto particular $x \rightarrow$ estacionaridad
 $\gamma(\vec{h})$ se llama el variograma = es la herramienta básica de la
 geoestadística.

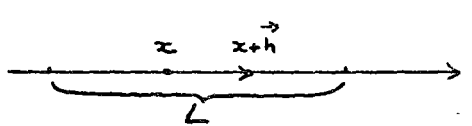
h es un vector del espacio que se considera, se define con su módulo y su dirección.

Siempre, usaremos $\gamma(\vec{h})$ o semivariograma.

V.- ESTUDIO DEL SEMIVARIOGRAMA. ANALISIS ESTRUCTURAL

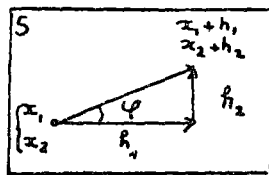
* Determinación del semivariograma

- V.R. de una dimensión



$$\gamma(\vec{h}) = \frac{1}{2(L-h)} \int_0^{L-h} [y(x+\vec{h}) - y(x)]^2 dx$$

- V.R. de 2 dimensiones



$$\gamma(\vec{h}_1, \vec{h}_2) = \frac{1}{2S} \int_0^S [y(x_1+\vec{h}_1, x_2+\vec{h}_2) - y(x_1, x_2)]^2 dx_1 dx_2$$

$\gamma(\vec{h}_1, \vec{h}_2)$ es una función de $h = \sqrt{h_1^2 + h_2^2}$ y φ

- V.R. de 3 dimensiones. Se considera, de la misma manera,
 $(\vec{h}_1, \vec{h}_2, \vec{h}_3)$.

Más generalmente, \vec{h} representará el vector del espacio que se considera.

* Propiedades del semivariograma

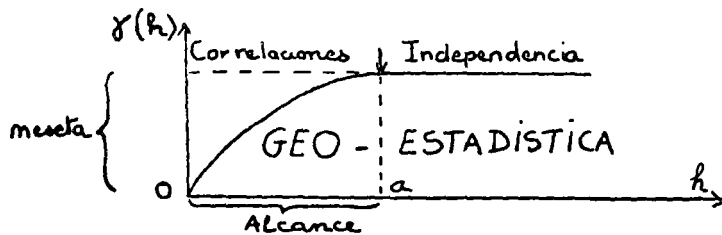
- simetría $\gamma(-\vec{h}) = \gamma(+\vec{h})$
- $\gamma(0) = 0$.
- $\gamma(\vec{h})$ es una función creciente de \vec{h} .

* Isotropía - Anisotropías.

Consideremos un fenómeno que se desplaza en el espacio con 2 dimensiones.

- Sea una dirección particular D_1

El fenómeno es un fenómeno de transición. Su semivariograma se presenta así:



A partir de una cierta distancia llamada alcance, el semivariograma ya no crece = alcanza su meseta.

Interpretación: 2 puntos alejados de una distancia mayor que 'a' se pueden considerar como independientes. 2 puntos alejados de una distancia menor que 'a' tienen correlaciones expresadas por la parte creciente del semivariograma.

La parte creciente del semivariograma expresa lo que aporta la -- Geoestadística. Empieza el dominio de la Estadística clásica, luego del alcance (Independencia de las V.A. $Y(x)$ e $Y(x+\vec{h})$)

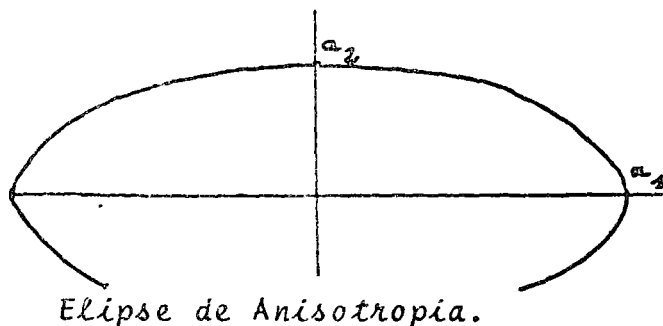
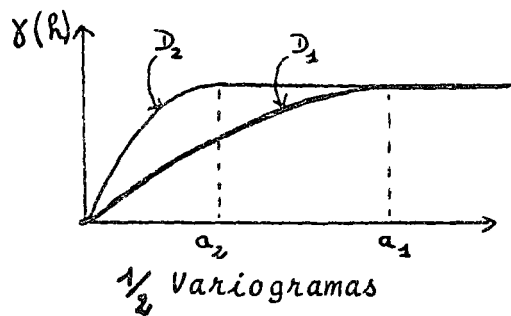
Además, el semivariograma cuantifica la manera más o menos grande como se degrada la influencia de un punto sobre los demás, en dicha dirección D_1 .

- Sean otras direcciones $D_2, D_3 - - -$

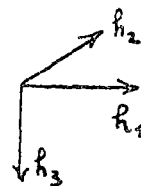
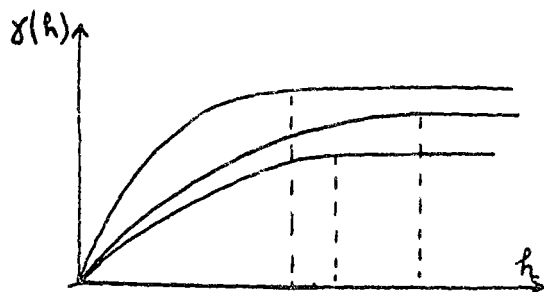
Isotropía = si el semivariograma es igual en cada dirección, el fenómeno que se estudia es isótropo. El semivariograma depende nada más del módulo $h = \sqrt{h_1^2 + h_2^2}$

Anisotropías = si el semivariograma presenta diferentes valores - del alcance, de la meseta, según la dirección, entonces la estructura del fenómeno es anisotrópica.

Anisotropía geométrica = misma meseta, alcances diferentes



Anisotropía Zonal = mesetas y alcances diferentes



Ejemplo en el espacio con 3 dimensiones

$V(x)$ depende de h_3 y queda constante en planos paralelos a (\vec{h}_1, \vec{h}_2)
Entonces $\gamma(\vec{h}_1, \vec{h}_2, \vec{h}_3) = \gamma(\vec{h}_3)$

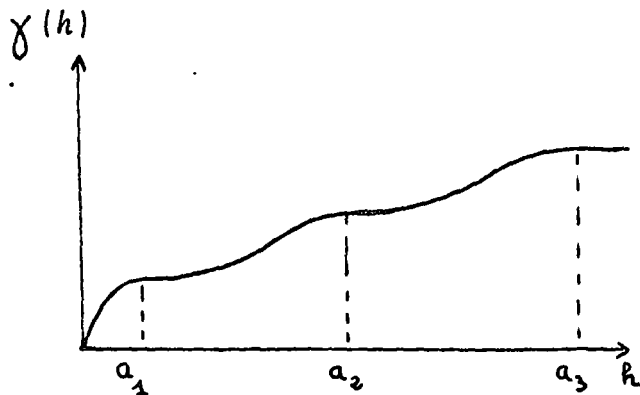
$V(x)$ tiene una fuerte anisotropía vertical

Entonces $\gamma(\vec{h}_1, \vec{h}_2, \vec{h}_3) = \gamma_1(\vec{h}_3) + \gamma_2(\vec{h}_1, \vec{h}_2, \vec{h}_3)$

El semivariograma es la suma de dos componentes γ_1 y γ_2

A veces, el modelo de anisotropía puede ser muy complejo y hacer intervenir a la vez las anisotropías geométrica y zonal.

* Estructuras encadenadas



Varias estructuras de escala diferente (metro, decena de metros, centena de metros) se suman en el semivariograma global

Ejemplo práctico: Tesis de J. SERRA
(Ver Bibliografía)

Resulta el semivariograma una herramienta muy adecuada al estudio de las estructuras geológicas.

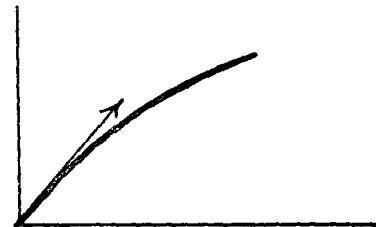
* Continuidad de la V.R.

El comportamiento del semivariograma al origen refleja el grado de regularidad más o menos grande de la V.R. en el espacio.

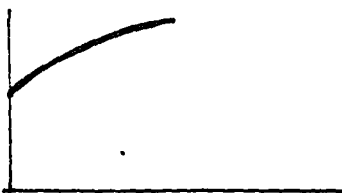


Comportamiento
parabólico:

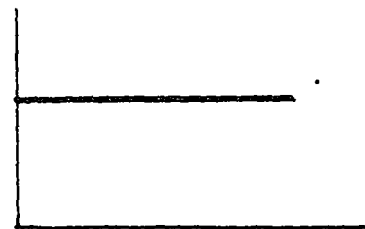
Alto grado de regularidad



Comportamiento lineal
Continuidad en $h=0$



Efecto de pepita
discontinuidad en $h=0$
Alto grado de irregularidad



Caso llmite
Caso Aleatorio puro

Nota importante:

En el caso isótropo, el semivariograma tiene un desarrollo limitado al origen del tipo siguiente:

$$2\gamma(h) = \underbrace{\sum a_{2n} h^{2n}}_{\text{parte regular}} + \underbrace{\sum c_{2n} h^{2n} \log h + \sum c_{\lambda} h^{\lambda}}_{\text{parte irregular}}$$

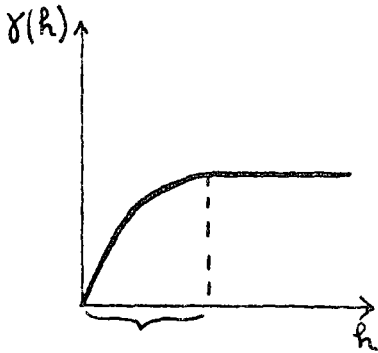
$$h = \sqrt{h_1^2 + h_2^2 + \dots}$$

a_{2n}, c, c_{2n} = coeficientes

λ diferente de un entero par.

La continuidad de la V.R. es ligada a la parte irregular del desarrollo.

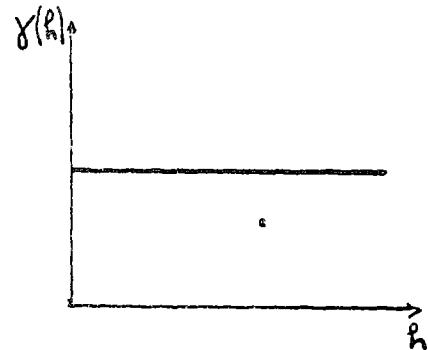
* Geoestadística y estadística clásica



parte creciente
CORRELACIONES



GEOESTADISTICA



No hay parte creciente
CASO ALEATORIO PURO
INDEPENDENCIA



ESTADISTICA CLASICA

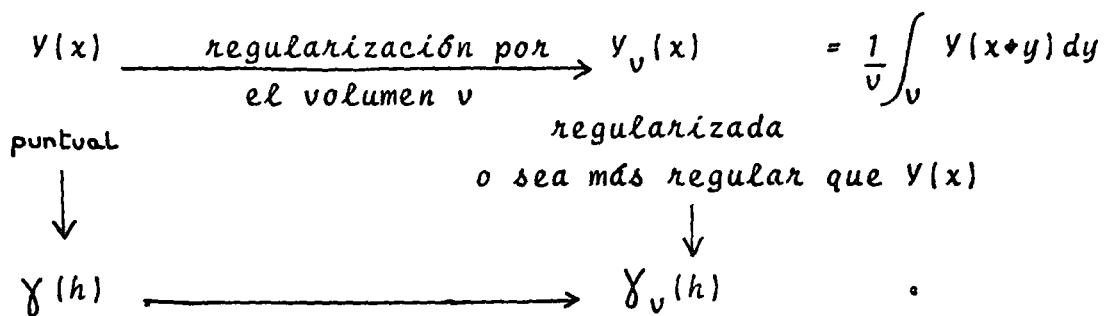
La Estadística clásica (Variables Aleatorias Independientes) es un caso particular de la Geoestadística ya que la Geoestadística encuentra los mismos resultados que la Estadística Clásica en el caso de un semivariograma reducido a su meseta.

VI.- EL SEMIVARIOGRAMA REFLEJA LAS CARACTERISTICAS ESTRUCTURALES DE LA V.R.

- CONTINUIDAD
- ANISOTROPIAS
- FENOMENO DE TRANSICION

VII.- REGULARIZACION

No se conocen los valores puntuales de la V.R. $\gamma(x)$ sino los valores promedios dentro de un cierto volumen v .



Se demuestra la fórmula siguiente:

$$\gamma_v(h) = \int \gamma(h+z) P(z) dz - \int \gamma(z) P(z) dz$$

o sea

$$\gamma_v(h) = \gamma * P - A$$

A es una constante, el símbolo * representa el producto de convolución. p es la función de ponderación asociada al volumen v y $P = p * p$

En la práctica, solo se conoce el semivariograma experimental de la V.R. $\gamma_V(x)$.

Con la ayuda de ábacos, se puede fácilmente deducir de $\gamma_V(h)$, el semivariograma $\gamma(h)$ de $V(x)$ o sea el semivariograma puntual que es la ley de dispersión intrínseca del fenómeno que se estudia.

Vemos así que la Geoestadística formaliza rigurosamente la influencia del soporte y la fórmula ya mencionada permite encontrar el semivariograma de las V.R. definidas sobre los paneles de explotación de varias toneladas.

VIII.- PRINCIPALES SEMIVARIOGRAMAS

* Semivariograma esférico o esquema de MATHERON

Caracteriza los fenómenos de transición.

Su expresión analítica es (caso isótropo):

$$\gamma(h) = C \left\{ \frac{3}{2} \cdot \frac{h}{a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right\} \quad 0 \leq h \leq a$$
$$\gamma(h) = C \quad h > a$$

'C' = meseta

'a' = alcance

* Semivariograma lineal

Su expresión analítica es (caso isótropo):

$$\gamma(h) = \bar{\omega} |h|$$

Son los 2 semivariogramas más usados en la práctica minera.

CONSTRUCCION DE UN VARIOGRAMA EXPERIMENTAL

Aplicación práctica

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

Los datos son colocados en los nudos de una retícula de malla a . Se calculará el semivariograma 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 semivariograma experimental que se calcula en la dirección α es

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

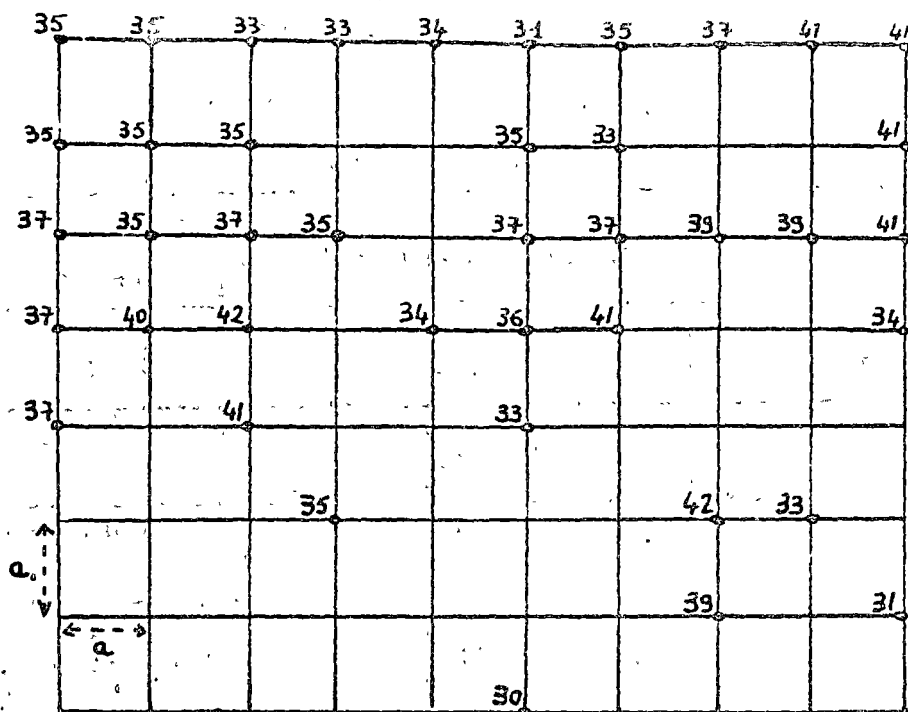
- $f(x_i+h)$ es el valor de la variable en el punto experimental (x_i+h) .
- $f(x_i)$ es el valor de la variable en el punto experimental (x_i) .
- $N_{\alpha}(h)$ es el número de diferencias $[f(x_i+h) - f(x_i)]$

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

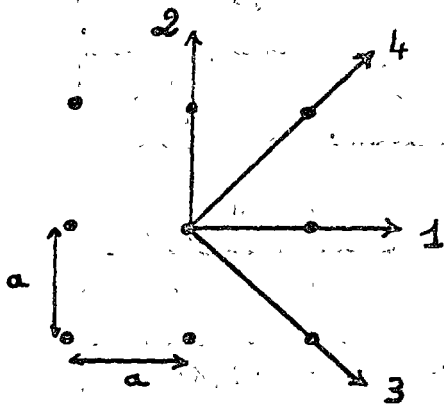
El número de datos es tan pequeño que los cálculos se pueden hacer a mano.

Se diseñarán los variogramas experimentales en cada dirección.

Qué conclusiones se pueden sacar?



Los nudos de la retícula 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.

Los resultados numéricos se encuentran en el cuadro A

Dirección α	incremento	número de incrementos	χ_{α}^*
1	a	24	4.18
	$2a$	20	8.40
	$3a$	18	13.03
2	a	22	4.25
	$2a$	18	8.69
	$3a$	15	10.90
3	$a\sqrt{2}$	19	5.02
	$2a\sqrt{2}$	16	11.90
	$3a\sqrt{2}$	10	17.25
4	$a\sqrt{2}$	18	6.47
	$2a\sqrt{2}$	14	11.25
	$3a\sqrt{2}$	8	15.44

Cuadro A

SEMI VARIOGRAMA EXPERIMENTAL

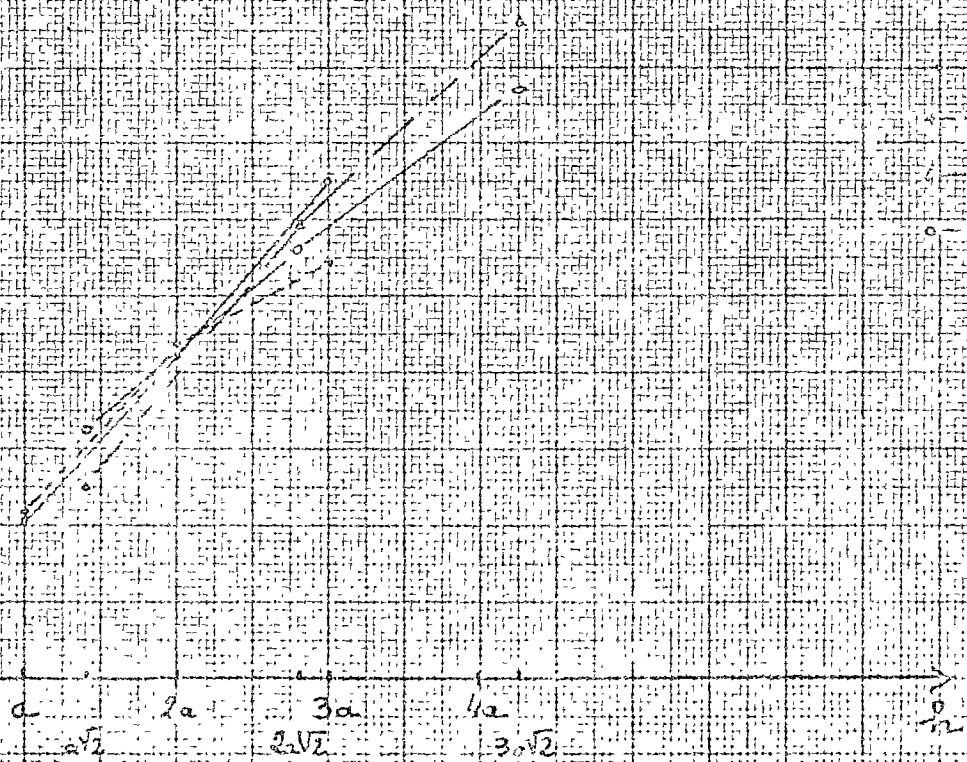
$\gamma(h)$

18
16
14
12
10
8
6
4
2
0

0 a 2a 3a 4a
 $a\sqrt{2}$ $2a\sqrt{2}$ $3a\sqrt{2}$

$\frac{h}{2}$

Dist. 1
Dist. 2
Dist. 3
Dist. 4



Se pueden sacar las conclusiones siguientes (ver la gráfica)

- Cualquiera que sea la dirección que se considera, el comportamiento del semivariograma queda igual. Se concluye a la isotropía de la regionalización.

- Este comportamiento es lineal. Se puede ajustar este semivariograma experimental al modelo teórico $\gamma(h) = \bar{\omega}|h|$

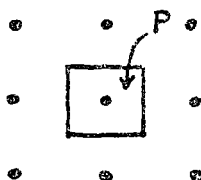
LAS VARIANZAS DE LA GEOESTADISTICA

OBJETIVOS

* Varianza de extensión - Varianza de estimación.

¿Cuál es el error que se comete cuando se estima la ley media $V(V)$ del volúmen V por la ley media $V(v)$ del volúmen v ?

La varianza de ese error se llama Varianza de extensión de v a V . Se escribe $\sigma^2(v, V)$.



En la práctica, se calculará la varianza de extensión del barreno central al panel P .

Si se disponen más barrenos (exteriores e interiores) la varianza de extensión toma el nombre de Varianza de estimación.

No hay ninguna diferencia conceptual entre una varianza de extensión y una varianza de estimación.

* Varianza de dispersión



¿Cuál es la varianza de la ley de v cuando v toma todas las ubicaciones posibles dentro del volúmen V ?

Esa varianza se llama Varianza de dispersión de v dentro de V .

Se escribe $\sigma^2(v/V)$.

La GEOESTADISTICA destaca estas dos varianzas y las calcula.

VARIANZA DE ESTIMACION

Estimación del valor promedio de la variable Y en el volúmen V por el valor promedio del volumen v .

Fórmula general.

$$\sigma^2(v, V) = 2\bar{\gamma}(v, V) - \bar{\gamma}(V, V) - \bar{\gamma}(v, v).$$

$\bar{\gamma}(v, V)$ = valor promedio del $\frac{1}{2}$ variograma $\gamma(\overrightarrow{MM'}) = \gamma(h)$ cuando M y M' describen, independientemente uno de otro, M el volumen v y M' el volumen V .

Definiciones idénticas para $\bar{\gamma}(V, V), \bar{\gamma}(v, v)$.

Dicha varianza de estimación depende de 4 factores:

- * la estructura subyacente del fenómeno que se estudia, o sea el $\frac{1}{2}$ variograma $\gamma(h)$, puntual, de la variable Y - - - - - $\gamma(h)$
- * la geometría del volumen V que se estima - - - - - $\bar{\gamma}(V, V)$
- * la geometría del volumen v que sirve a la estimación - $\bar{\gamma}(v, v)$
- * la geometría relativa de los volúmenes v y V - - - - $\bar{\gamma}(v, V)$

Aplicación práctica

Se hace la estimación con la ayuda de $i = 1, N$ informaciones experimentales Y_i disponibles. Sea v_i este conjunto experimental.

$$\sigma^2(v_i, V) = \frac{2}{NV} \sum_i \int \gamma(x_i - y) dy - \frac{1}{V^2} \int \int \gamma(x - y) dx dy - \frac{1}{N^2} \sum_i \sum_j \gamma(x_i - x_j)$$

VARIANZA DE DISPERSION

Se definen rigurosamente las varianzas de dispersión siguientes:

$$\sigma^2(v/V) \text{ cuando } \begin{cases} v \text{ es un punto que se nota } 0 \\ v \text{ es tal que } v = u \cdot v_i \text{ con } \begin{cases} v_i = v \\ v_i \cap v_j = \emptyset \end{cases} \end{cases}$$

Los resultados más importantes son:

$$\sigma^2(0/V) = \frac{1}{V^2} \int \int \gamma(x - y) dx dy = \bar{\gamma}(V, V)$$

$$(2) \quad \sigma^2(v/V) = \frac{1}{V^2} \int_V \int_V \gamma(x-y) dx dy - \frac{1}{V^2} \int_V \int_V \bar{\gamma}(x-y) dx dy = \bar{\gamma}(v,v) - \bar{\gamma}(V,V)$$

Una varianza de dispersión depende de:

- * la estructura subyacente o sea el $\frac{1}{2}$ variograma $\gamma(h)$
- * la geometría del volumen V - - - - $\bar{\gamma}(V,V)$
- la geometría del volumen v - - - - $\bar{\gamma}(v,v)$

La varianza de dispersión no depende de la información disponible.

Caso particular = V no es una partición exacta de v , pero la dimensión de v es muy chica en cuanto a la dimensión de V

La fórmula (2) queda válida.

Relaciones de aditividad

$$\sigma^2(v/V) = \sigma^2(v/v) - \sigma^2(v/V)$$

Relación de Krige

y también

$$\sigma^2(v/V^-) = \sigma^2(v/V) + \sigma^2(V/V^-)$$

LA ESTIMACION DE LAS V.R.:

FORMULAS DE APROXIMACION

Si la información disponible es muy numerosa y de implantación irregular, hace intervenir volúmenes de tamaño diferente o, en término geostatístico, soportes diferentes, la fórmula

$$\sigma^2_{Est} = \frac{2}{NV} \sum_{i=1}^{i=N} \int_V \gamma(x_i - y) dy - \frac{1}{V^2} \int_V \int_V \gamma(x - y) dx dy - \frac{1}{N^2} \sum_{i=1}^{i=N} \sum_{j=1}^{j=N} \gamma(x_i - x_j)$$

no es fácil de calcular. Los principios de aproximación siguientes permiten sacar, en las aplicaciones prácticas, varianzas de estimación representativas.

METODO DE APROXIMACION EN EL ESPACIO DE UNA DIMENSION

(1) Principio de correspondencia

El $\frac{1}{2}$ variograma $\gamma(h)$ admite a la vecindad del origen un desarrollo limitado cuya forma analítica comporta términos en h de potencia par (parte regular) y términos de potencia impar más términos en $\log h$ (parte irregular) o sea:

$$\gamma(h) = \underbrace{\sum_n a_{2n} h^{2n}}_{\text{parte regular}} + \underbrace{\sum_{\lambda} c_{\lambda} h^{\lambda} + \sum_n c_{2n} h^{2n} \log h}_{\text{parte irregular}}$$

→ a cada término de la parte irregular se puede asociar un término de la varianza de estimación.

→ la parte regular no interviene en los cálculos de la varianza de estimación.

Ejemplo

Estimación de la ley media de un segmento (una galería en la práctica) por una muestra central.



$$\frac{1}{2} \text{ variograma } \gamma(h) = h^{\lambda}$$

$$\text{varianza de estimación rigurosa } \sigma_{\text{Est}}^2 = \frac{2 a^{\lambda}}{\lambda+1} \left(\frac{1}{2^{\lambda}} - \frac{1}{\lambda+2} \right)$$

Entonces se ve que

al término h le corresponde el término $\frac{a}{6}$ de σ_{Est}^2

" " h^2 " " 0 " "

" " h^3 " " $-\frac{3}{80} a^3$ "

" " $h^2 \log h$ " " $-0.07385 a^2$ "

Dominio de validez de este principio = mallas pequeñas.

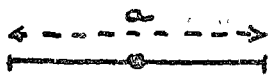
(2) Principio de composición.

* Funciones auxiliares

$$X(a) = \frac{1}{a} \int_0^a Y(x) dx$$

$$F(a) = \frac{1}{a^2} \int_0^a \int_0^a Y(x-y) dx dy$$

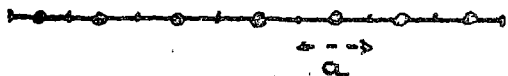
* Varianza de extensión elemental



Se estima el valor promedio del segmento por el valor en el punto central

$$\text{Se demuestra } \sigma_{Ext}^2 = 2 X(a/2) - F(a)$$

* Principio de composición.



Se estima el valor promedio de la línea por el conjunto de N muestras centrales

El principio admite que los errores elementales son independientes; pues la varianza de estimación final de la línea por N muestras centrales es igual a la varianza de extensión elemental dividida por el número de muestras.

$$\sigma_{Est}^2 = \frac{1}{N} \sigma_{Ext}^2 = \frac{1}{N} \left\{ 2 X\left(\frac{a}{2}\right) - F(a) \right\}$$

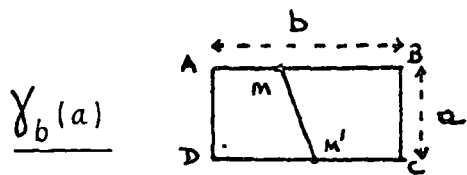
Comparación de los 2 principios:

a pequeño → los 2 principios son equivalentes

a grande → solamente se debe usar el segundo principio

METODO DE APROXIMACION EN EL ESPACIO DE DOS DIMENSIONES

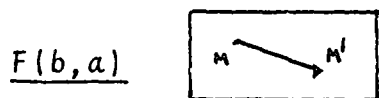
* Funciones auxiliares de dos dimensiones



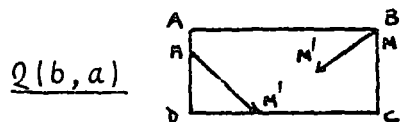
$$\gamma_b(a) = \bar{\gamma}(AB, DC)$$



$$x_b(a) = \bar{\gamma}(AB, ABCD)$$

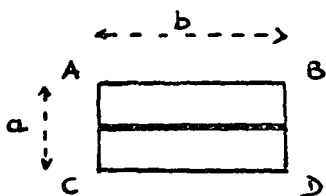


$$F(b, a) = \bar{\gamma}(ABCD, ABCD)$$



$$Q(b, a) = \bar{\gamma}(AD, DC) = \bar{\gamma}(B, ABCD)$$

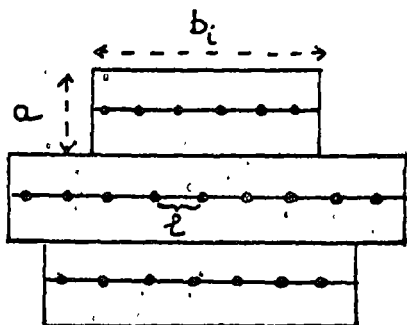
* Varianza de extensión de un segmento mediano a su rectángulo de influencia.



Se supone el segmento mediano perfectamente conocido

$$\sigma_{Ext}^2 = 2 x_b\left(\frac{a}{2}\right) - F(b, a) - \gamma_b(0)$$

(1) Principio de composición de términos de línea y de banda.



El yacimiento se compone de un conjunto de N rectángulos (a, b_i) , siendo cada rectángulo reconocido por unas muestras implantadas a malla l en el segmento mediano. Se necesita la relación $l \ll a$.

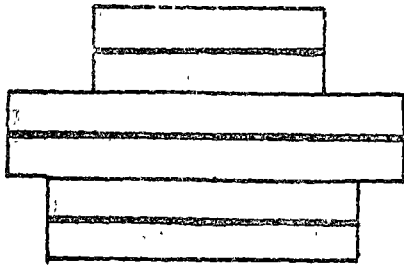
Término de línea

Se conoce solamente una estimación de cada línea b_i . La varianza de estimación de todas las líneas por las muestras elementales constituye el término de línea.

Si N' representa el número de muestras de todas las líneas b_i , el

término de línea vale $(\sigma_{Est}^2)_1 = \frac{1}{N'} \left\{ 2X\left(\frac{l}{2}\right) - F(l) \right\}$

Término de banda



Se suponen las líneas medianas perfectamente conocidas y se admite que los errores elementales de estimación de cada rectángulo por su línea mediana son independientes.

Por tanto el término de banda es:

$$(\sigma_{Est}^2)_2 = \frac{1}{(\sum b_i)^2} \sum_i b_i^2 \left\{ 2X_{b_i}\left(\frac{a}{2}\right) - F(b_i, a) - \chi_{b_i}(0) \right\}$$

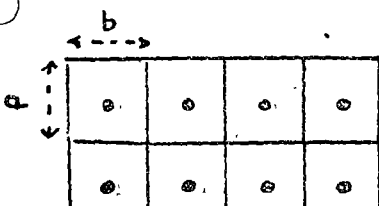
Caso particular $b_i = b_j \forall i, j$ $(\sigma_{Est}^2)_2 = \frac{1}{N} \left\{ 2X_b\left(\frac{a}{2}\right) - F(b, a) - \chi_b(0) \right\}$

con $N =$ número de rectángulos

La varianza de estimación total es la suma de estos 2 términos supuestos independientes.

$$\sigma_{Est}^2 = (\sigma_{Est}^2)_1 + (\sigma_{Est}^2)_2 = \frac{1}{N'} \left\{ 2X\left(\frac{l}{2}\right) - F(l) \right\} + \frac{1}{(\sum b_i)^2} \sum_i b_i^2 \left\{ 2X_{b_i}\left(\frac{a}{2}\right) - F(b_i, a) - \chi_{b_i}(0) \right\}$$

(2). Principio de composición en el caso de una malla cuadrada o rectangular.



La varianza de extensión elemental de un barrero a su rectángulo (o cuadrado) de influencia es:

$$\sigma_{Ext}^2 = 2 Q\left(\frac{a}{2}, \frac{b}{2}\right) - F(a, b).$$

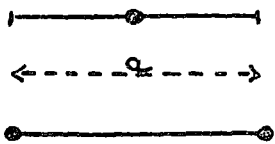
La varianza de estimación es:

$$\sigma_{Est}^2 = \frac{1}{N} \left\{ 2 Q \left(\frac{a}{2}, \frac{b}{2} \right) - F(a, b) \right\}$$

donde N representa el -

número de rectángulos o cuadrados que componen el yacimiento.

VARIANZA DE ESTIMACION - APLICACION PRACTICA



Este ejercicio tiene por objeto calcular la varianza de estimación (o varianza de extensión elemental) del segmento de longitud 'a' por:

- su muestra central \rightarrow varianza de estimación σ_E^2
- 2 muestras colocadas en sus extremidades \rightarrow varianza de estimación σ_E^2 .

Se efectuarán los cálculos con el semivariograma $\gamma(h) = h^\lambda$ ($0 < \lambda < 2$) y se compararán los resultados obtenidos según varios valores de λ .

Cálculo de las funciones auxiliares X y F del semivariograma $\gamma(h) = h^\lambda$

* $X(a) = \frac{1}{a} \int_0^a h^\lambda dh = \frac{1}{a} \left[\frac{h^{\lambda+1}}{\lambda+1} \right]_0^a = \frac{a^\lambda}{\lambda+1} \rightarrow \boxed{X(a) = \frac{a^\lambda}{\lambda+1}}$

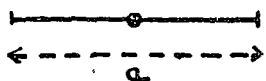
* $F(a) = \frac{2}{a^2} \int_0^a (a-h) \gamma(h) dh = \frac{2}{a^2} \int_0^a (a-h) h^\lambda dh = \frac{2}{a} \int_0^a h^\lambda dh - \frac{2}{a^2} \int_0^a h^{\lambda+1} dh$

$F(a) = \frac{2a^\lambda}{\lambda+1} - \frac{2}{a^2} \left[\frac{h^{\lambda+2}}{\lambda+2} \right]_0^a = \frac{2a^\lambda}{\lambda+1} - \frac{2a^\lambda}{\lambda+2} = \frac{2a^\lambda (\lambda+2 - \lambda - 1)}{(\lambda+1)(\lambda+2)} = \frac{2a^\lambda}{(\lambda+1)(\lambda+2)}$

$F(a) = \frac{2a^\lambda}{(\lambda+1)(\lambda+2)}$

Cálculo de σ_E^2

En este caso, se ha obtenido la fórmula general



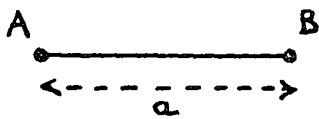
$\sigma_E^2 = 2 X \left(\frac{a}{2} \right) - F(a)$

con el semivariograma $\gamma(h) = h^\lambda$, se obtiene fácilmente

$$\sigma_E^2 = 2 \frac{a^\lambda}{2^\lambda (\lambda+1)} - \frac{2a^\lambda}{(\lambda+1)(\lambda+2)} = \frac{2a^\lambda}{\lambda+1} \left(\frac{1}{2^\lambda} - \frac{1}{\lambda+2} \right)$$

$$\sigma_E^2 = \frac{2a^\lambda}{\lambda+1} \left(\frac{1}{2^\lambda} - \frac{1}{\lambda+2} \right)$$

Cálculo de σ_E^2



Se usará la fórmula general

$$\sigma_{Est}^2 = \frac{2}{NV} \sum_i \int_V \gamma(x_i - y) dy - \frac{1}{V^2} \int_V \int_V \gamma(x - y) dx dy - \frac{1}{N^2} \sum_i \sum_j \gamma(x_i - x_j)$$

$N=2$, número de muestras

$V=a$, longitud del segmento

x_i, x_j = coordenadas de la muestra x_i, x_j

Calculemos cada término:

$$* \frac{2}{NV} \sum_i \int_V \gamma(x_i - y) dy = \frac{2}{2a} \left\{ \int_0^a \gamma(x_A - y) dy + \int_0^a \gamma(x_B - y) dy \right\}$$

Siendo iguales las 2 integrales anteriores, viene

$$\frac{2}{NV} \sum_i \int_V \gamma(x_i - y) dy = \frac{2}{a} \int_0^a \gamma(x_A - y) dy = \frac{2}{a} \int_0^a \gamma(y) dy = 2 X(a)$$

$$* \frac{1}{V^2} \int_V \int_V \gamma(x - y) dy dx = \frac{1}{a^2} \int_0^a \int_0^a \gamma(x - y) dx dy = F(a)$$

$$* \frac{1}{N^2} \sum_i \sum_j \gamma(x_i - x_j) = \frac{1}{4} \left\{ \gamma(x_A - x_A) + \gamma(x_A - x_B) + \gamma(x_B - x_A) + \gamma(x_B - x_B) \right\}$$

$$\frac{1}{N^2} \sum_i \sum_j \gamma(x_i - x_j) = \frac{1}{2} \gamma(a)$$

Por tanto, se obtiene: $-\sigma_{Est}^2 = \frac{2 X(a) - F(a) - \gamma(a)}{2}$

Con el semivariograma $\gamma(h) = h^\lambda$, viene

$$\sigma_{E'}^2 = \frac{2a^\lambda}{\lambda+1} - \frac{2a^\lambda}{(\lambda+1)(\lambda+2)} - \frac{a^\lambda}{2} = \frac{4a^\lambda(\lambda+2) - 4a^\lambda - (\lambda+1)(\lambda+2)a^\lambda}{2(\lambda+1)(\lambda+2)} = \frac{a^\lambda(\lambda-2)(\lambda+1)}{2(\lambda+1)(\lambda+2)}$$

$$\sigma_{E'}^2 = a^\lambda \left(\frac{2}{\lambda+2} - \frac{1}{2} \right)$$

Aplicación numérica

$\lambda = 1$

$\sigma_E^2 = \frac{a}{6}$

$\sigma_{E'}^2 = \frac{a}{6}$

$\sigma_E^2 = \sigma_{E'}^2$

$\lambda = \frac{3}{2}$

$\sigma_E^2 = 0.054 a^{3/2}$

$\sigma_{E'}^2 = 0.071 a^{3/2}$

$\sigma_E^2 < \sigma_{E'}^2$

$\lambda = \frac{1}{2}$

$\sigma_E^2 = 0.41 a^{1/2}$

$\sigma_{E'}^2 = 0.31 a^{1/2}$

$\sigma_{E'}^2 < \sigma_E^2$

Conclusiones

$\lambda = 1$ Los 2 muestreos dan la misma precisión

$\lambda > 1$ De una manera general, se obtiene $\sigma_E^2 < \sigma_{E'}^2$, o sea, la muestra central, única pero mejor colocada, proporciona una precisión mejor que el conjunto de las 2 - muestras. Cabe hacer notar que en tal caso, la estructura subyacente es muy regular y el semivariograma - tiene un comportamiento al origen parecido al parabólico.

$\lambda < 1$

Se obtiene en forma general el resultado contrario - al precedente sea $\sigma_{E'}^2 < \sigma_E^2$. Los semivariogramas correspondientes caracterizan una estructura más aleatoria, así que resulta más importante el número - -

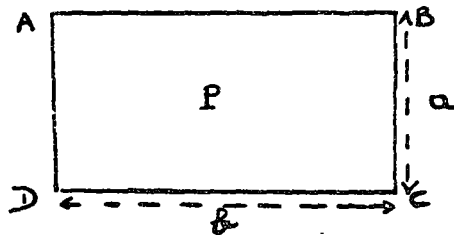
de muestras que su implantación.

En este ejemplo, vemos que la posición y el número de muestras no son los únicos factores que toma en cuenta el cálculo de la varianza de estimación. Importa mucho el tipo de semivariograma o sea, en otras palabras, el grado de regularidad y de continuidad de la estructura subyacente.

PRACTICAS DE FUNCIONES AUXILIARES

Sea P un panel rectangular de ley media desconocida Z .

Se supone que P se coloca dentro de una regionalización isotropa de semivariograma γ



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



①

Panel reconocido por un sondeo central



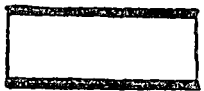
②

Panel reconocido por uno de sus extremos



③

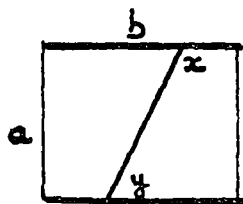
Panel reconocido por sus 4 extremos



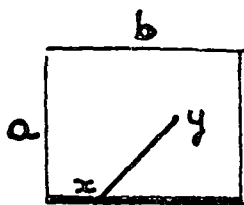
④

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

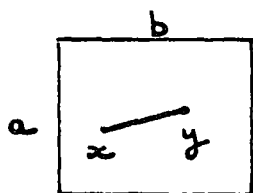
Se usarán las funciones auxiliares siguientes:



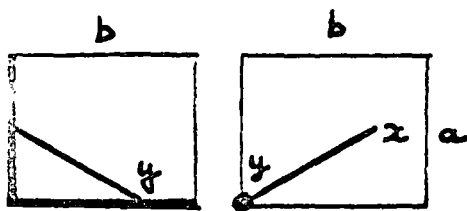
$\bar{y}_b(a)$ = valor promedio de $\gamma(x-y)$ cuando x e y describen independientemente los 2 lados paralelos de longitud del rectángulo a x b .



$X_b(a)$ = valor promedio de $\gamma(x-y)$, cuando x describe b e, independientemente de x , y describe el rectángulo a x b .



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



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

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

Nota

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

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

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

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

SOLUCION

Usaremos la fórmula de la varianza de estimación

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

① F' = sondeo central, sea un punto

F = rectángulo $a \times b$

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

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

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

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

② F' = sondeo

F = rectángulo $a \times b$

$$\bar{y}(F, F') = Q(a, b)$$

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

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

$$\sigma_2^2 = 2 Q(a, b) - F(a, b)$$

③ F' = 4 sondeos sea 4 puntos

F = rectángulo $a \times b$

$$\bar{y}(F, F') = \frac{1}{4ab} \sum_{i=1}^4 \int_F y(x_i - y) dy \quad \left. \vphantom{\bar{y}(F, F')} \right\} \bar{y}(F, F') = Q(a, b)$$

$$\forall x_i \frac{1}{ab} \int_F y(x_i - y) dy = Q(a, b)$$

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

$$\bar{y}(F', F') = \frac{1}{16} \sum_{i,j}^4 y(x_i - x_j) = \frac{1}{16} \times 4 \left[y(0) + y(b) + y(a) + y(\sqrt{a^2 + b^2}) \right] = \frac{1}{4} \left[y(b) + y(a) + y(\sqrt{a^2 + b^2}) \right]$$

$$\sigma_3^2 = 2 Q(a, b) - F(a, b) - \frac{1}{4} \left[y(b) + y(a) + y(\sqrt{a^2 + b^2}) \right]$$

④ $F' =$ los 2 lados de longitud b

$F =$ el rectángulo axb

$$\bar{\gamma}(F, F') = \frac{1}{(AB+DC) \times F} \int_{AB+DC} \int_F \gamma(x-y) dx dy$$

$$(AB+DC) \cdot F \cdot \bar{\gamma}(F, F') = \int_{AB} \int_F \gamma(x-y) dx dy + \int_{DC} \int_F \gamma(x-y) dx dy$$

$$\int_{AB} \int_F \gamma(x-y) dx dy = AB \times F \times X_b(a)$$

$$\int_{DC} \int_F \gamma(x-y) dx dy = DC \times F \times X_b(a)$$

$$\text{entonces } \bar{\gamma}(F, F') = \frac{AB \times F \times X_b(a) + DC \times F \times X_b(a)}{(AB+DC) \times F} = \frac{2bX_b(a)}{2b} = X_b(a)$$

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

$$\bar{\gamma}(F', F') = \frac{1}{(AB+CD)^2} \int_{AB+CD} \int_{AB+CD} \gamma(x-y) dx dy$$

$$(AB+CD)^2 \bar{\gamma}(F', F') = \int_{AB} \int_{AB} + 2 \int_{AB} \int_{CD} + \int_{CD} \int_{CD} = 2 \int_{AB} \int_{AB} + 2 \int_{AB} \int_{CD}$$

$$= 2 b^2 F(o, b) + 2 b^2 \gamma_b(a)$$

$$\bar{\gamma}(F', F') = \frac{2b^2 F(o, b) + 2b^2 \gamma_b(a)}{4b^2} = \frac{1}{2} (F(o, b) + \gamma_b(a))$$

$$\sigma_4^2 = 2 X_b(a) - F(a, b) - \frac{1}{2} [F(o, b) + \gamma_b(a)]$$

ESTIMACION GLOBAL DEL YACIMIENTO
DE ITY - MONT-FLOTUO

Estudio real

Presentación del problema

El yacimiento de ITY está constituido por una corteza aurífera de laterita y arcilla 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 20x30 m. Cada barreno fué muestreado a lo largo de la potencia mineralizada.

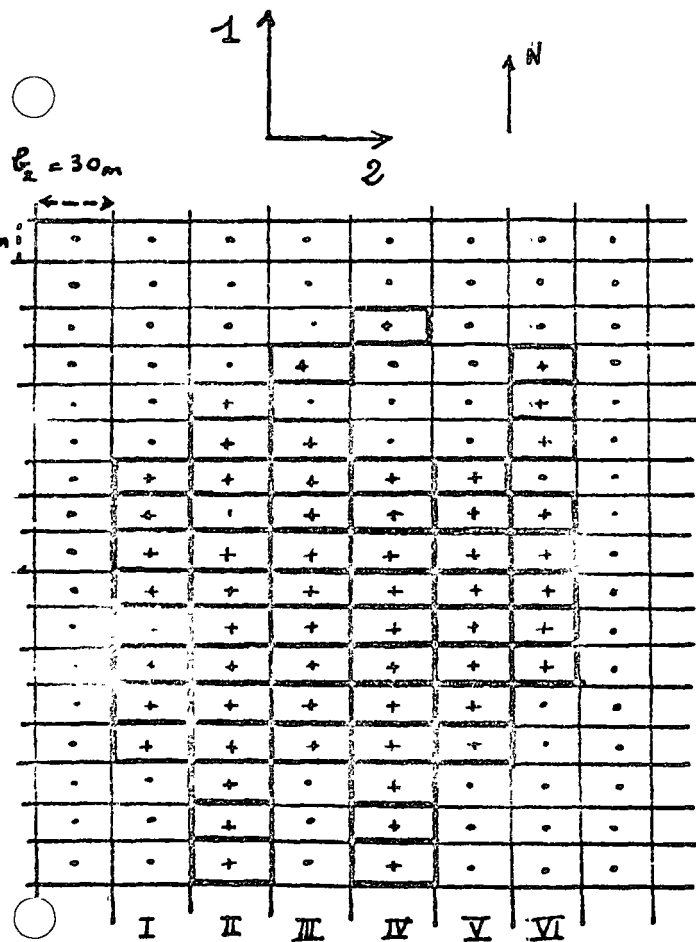


Fig 1

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 "adoquinos" de 20x30 m² positivos. Ae 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 la aplicación de la Teoría Transitiva.

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

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

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 semivariograma N-S de las acumulaciones relativas:

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

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

La varianza a priori de nuestra V.R., 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) de un problema con 3 dimensiones (estudio de la ley), pasamos a un problema con 2 dimensiones (estudio de la acumulación) = ejemplo de la operación de subida que no modifica las fórmulas 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:

$$\frac{\sigma_s^2}{s^2}$$

(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 poligonal.

$$S^* = n b_1 b_2 = 58 \times 20 \times 30 = 34800 \text{ m}^2$$

(2) Varianza relativa de estimación de esta superficie.

Se admite que las direcciones principales de anisotropía son las direcciones de la red sea NS y EO. 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 barrenos 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 laguna. N_2 es el más pequeño de los dos.

Aquí tenemos:

en la dirección N.S. $2 N = 42+2 = 44$

en la dirección E.O $2 N = 18+2 = 20$

por tanto $N_1 = 22$

$$N_2 = 10$$

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

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

Así sea un riesgo de 15% según la ley normal ($P \{S < S^* - \sigma_S\}$)

se obtiene $S > S^* - \sigma_S = 33500 \text{ m}^2$

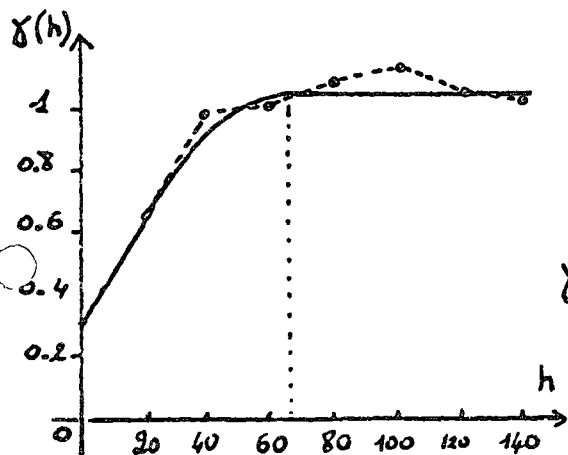
Nota relativa a la red de barrenos.

Las semivariaciones diametrales son:

- en la dirección N.S.: $D_1 = N_1 b_1 = 22 \times 20 = 440 \text{ m}$
- en la dirección E.O : $D_2 = N_2 b_2 = 10 \times 30 = 300 \text{ m}$

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 cuadriculada.

(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.

$$\gamma(h) = C_0 + \begin{cases} C \left[3/2 \frac{h}{a} - 1/2 \frac{h^3}{a^3} \right] & \dots \quad 0 \leq h \leq a \\ C & \dots \quad h > a \end{cases}$$

↑
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 estima la meseta a $C_0 + C = 1.05$, sea $C = 0.75$

Sabiendo que la tangente al origen al esquema de Matheron corta la meseta en un punto de abscisa igual a los 2/3 del alcance

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

El esquema queda pues completamente definido.

Probemos si 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 $3a \times 3,6a$.

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

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

Notemos el efecto de pepita fuerte: $E = \frac{C_0}{C} = \frac{0.3}{0.75} = 0.4$, 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) midiendo la extensión de los pozos centrales a sus segmentos de influencia ($b_1 = 20\text{m}$).

- de un término de banda midiendo la extensión de las líneas l_i supuestas conocidas a sus bandas 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 $\rightarrow C_0 + C \times \sigma_l^2(b_1)$

Se lee en el abaco = $\frac{b_1}{a} = \frac{20}{66} = 0.30 \Rightarrow \sigma_l^2(b_1) = 0.08$

De donde $C_0 + C_1 \sigma_L^2 (b_1) = 0.3 + 0.75 \times 0.08 = 0.36$

Vemos que el efecto de pepita tiene la mayor importancia.

Suponiendo la independencia de las extensiones de cada pozo a su segmento b_1 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	$\sigma_b^2 (b_2; L_i)$	$C \times \sigma_b^2 (b_2; L_i)$	$L_i^2 \times C \times \sigma_b^2 (b_2; L_i)$
I	160 m	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

$\Sigma = 1160 \text{ m}$

$\Sigma = 1534$

Término de banda

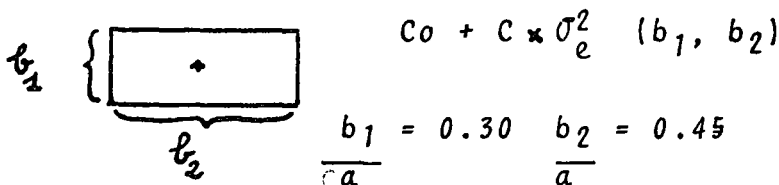
$$T = \frac{C \sum l_i^2 \sigma_b^2 (b_2; l_i)}{(\sum l_i)^2} = \frac{1.534}{(1.160)^2} = 0.0011$$

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) Principio de composición de las varianzas de extensión elementales

Extensión de un pozo central a su rectángulo de influencia:



luego de leer el ábaco $\sigma_e^2 = 0.15$

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

El efecto de pepita tiene la mayor importancia.

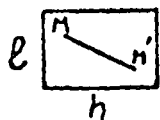
Suponiendo la independenciam de las extensiones de cada pozo a su rectángulo de influencia, la varianza de estimación de la ley relativa es:

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

Los 2 principios dan casi el mismo resultado.

(5) Efecto de borde

Detengamos como varianza de estimación de la acumulación media a su superficie mineralizada 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 mineralizada y el error geométrico repercute sobre la estimación de la acumulación media. Se introduce pues un término suplementario que se añade a los términos precedentes.



Fonction $\frac{1}{C} F\left(\frac{l}{a}, \frac{h}{a}\right) =$ Variance d'un point dans le rectangle $l \times h$

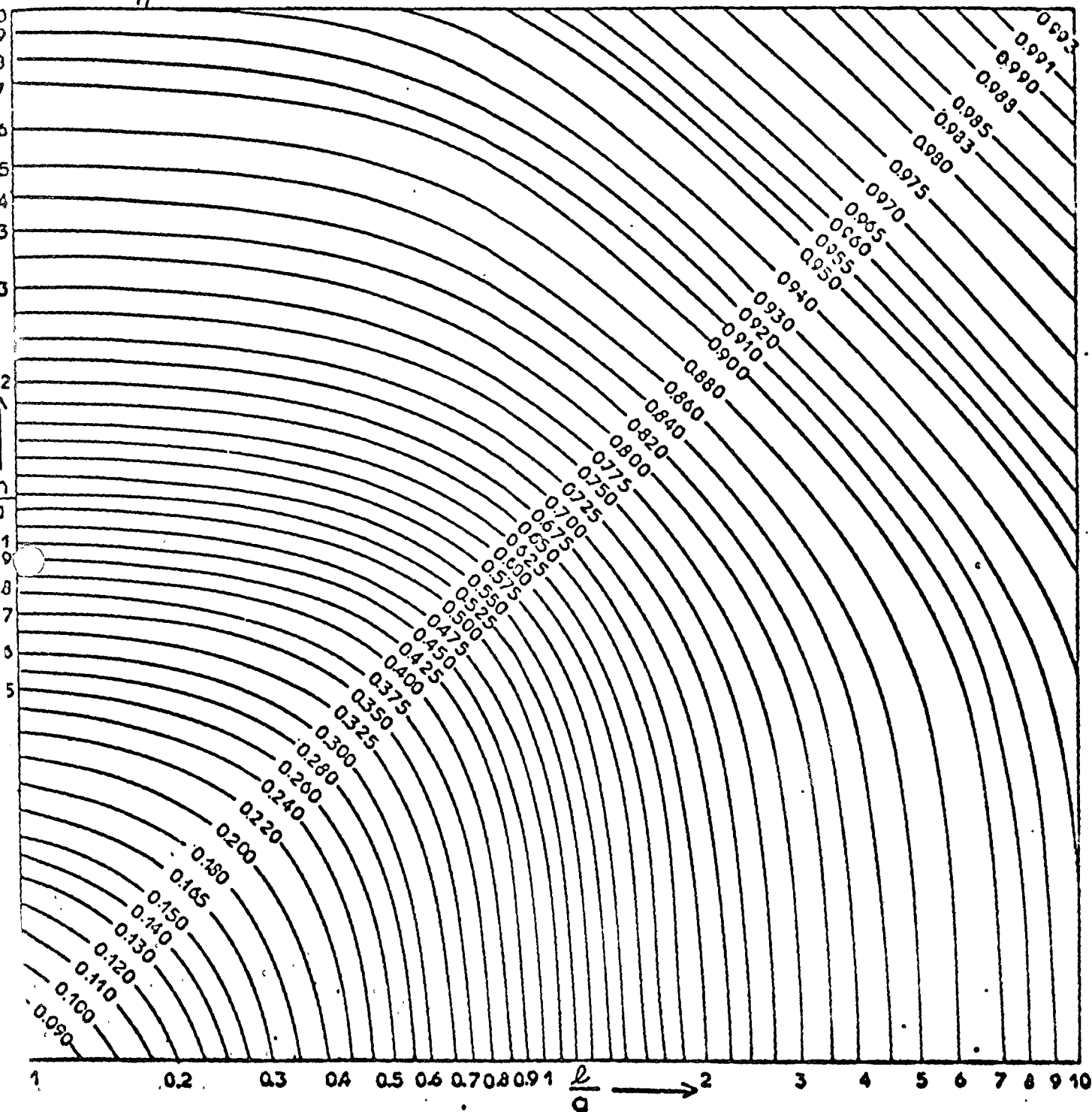


Fig:9 ..schéma de MATHERON

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

en fin resulta,

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

(6) Varianza de estimación de la cantidad de metal Q

$$Q = a S$$

La hipótesis de independencia interna entre la V R y su campo nos permite escribir:

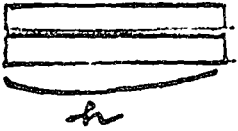
$$\begin{aligned} \frac{D^2 [Q]}{Q^2} &= \frac{D^2 [a]}{a^2} + \frac{D^2 [S]}{S^2} \\ &= 0.0087 + 0.0014 = 0.01 \end{aligned}$$

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

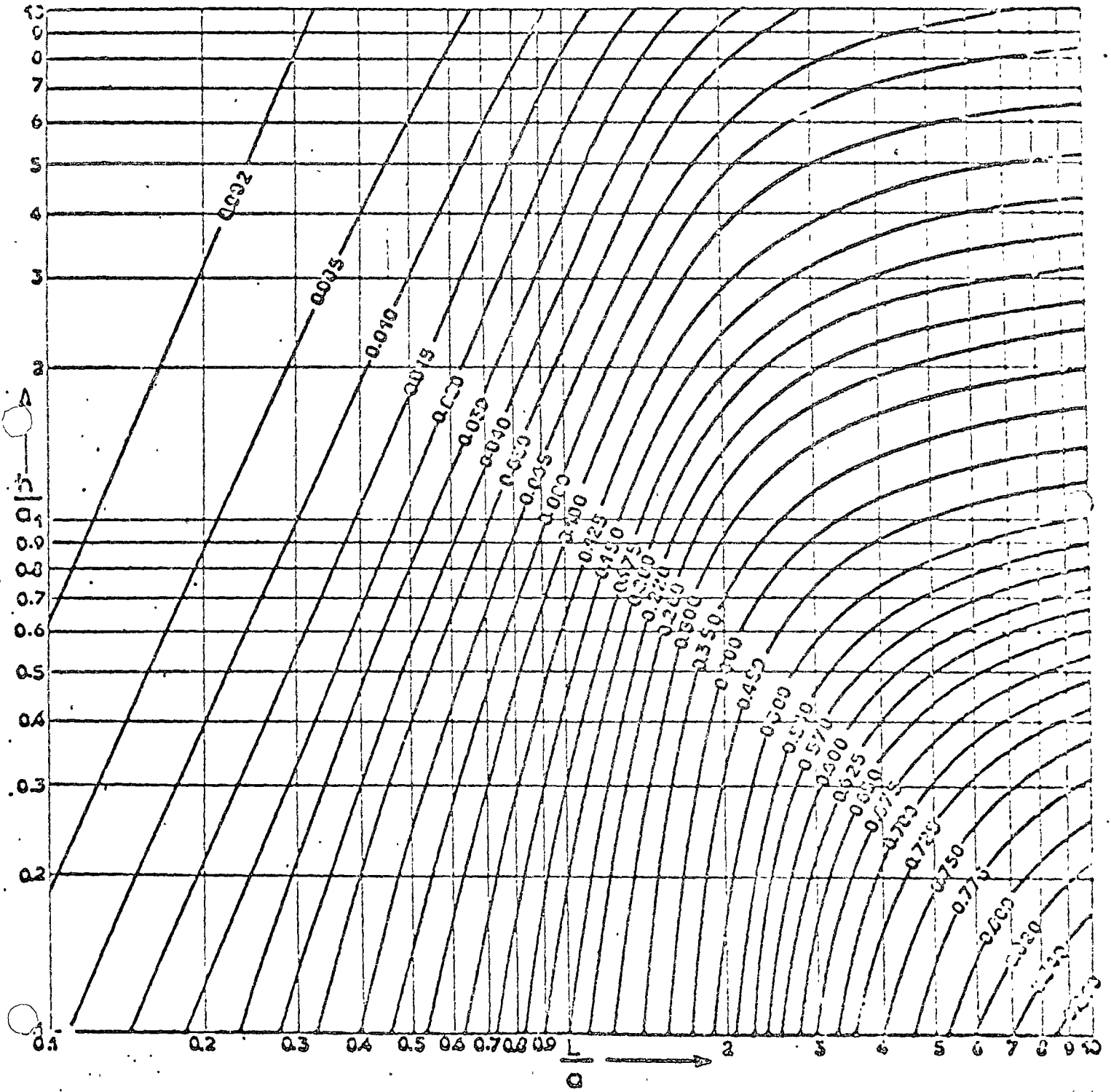
$\frac{\sigma(Q)}{Q} = 0.1$ con un riesgo de 15% y según una ley normal

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

ABAUUE : 9 SCHEMA SPHERIQUE.



variance d'extension du tracage median h dans le rectangle $L \times h$



Variances d'extension diverses.

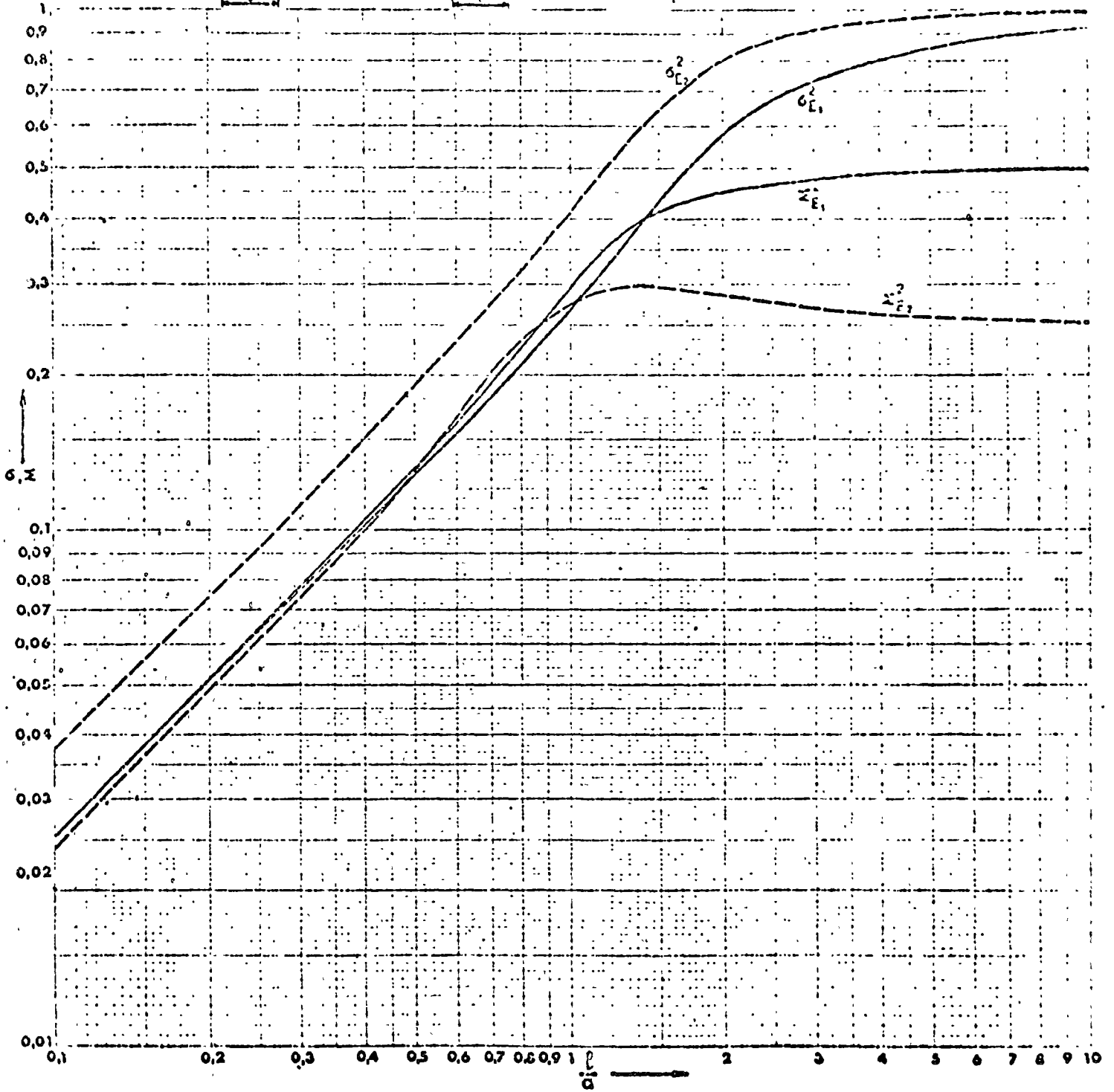
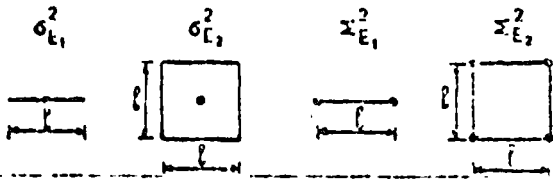
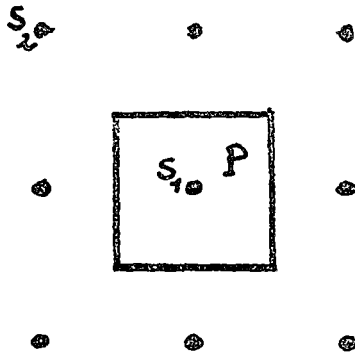


Fig.11-Schéma de Matheron

EL KRIGEAGE

I.- PRESENTACION DEL PROBLEMA

En términos mineros, el problema del KRIGEAGE -o estimación local- consiste en encontrar la manera cómo procesar la información disponible para sacar la estimación más acertada de un bloque. Consideremos la configuración siguiente:



- se dispone de una retícula cuadrada.

En cada nudo se conoce el valor de la ley de una muestra.

- se trata de estimar la ley promedio del bloque P, centrado en un nudo de la retícula.

* Para estimar el bloque P, se puede considerar

- la perforación interior S_1
- ó cualquier perforación exterior
- ó un conjunto de perforaciones exterior e interior.

* En cada caso, se puede calcular la varianza de estimación. La estimación que se debe tomar en cuenta es la que proporciona la varianza de estimación mínima; por ejemplo, la varianza de estimación del bloque P por S_1 es menor que la de P por S_2 . S_1 siendo mejor colocada (está dentro de P) representa mejor el bloque P que S_2 .

* Es cierto que obtener la estimación más precisa es de gran interés en la práctica; pero el Krigage permite también evitar un sesgo de estimación que puede conducir a consecuencias económicas graves; es la segunda propiedad fundamental del Krigage.

La práctica minera confirma que al usar la sola información interior en la estimación de un bloque

- se sobre-estima la proporción de bloques ricos
- se sub-estima la proporción de bloques pobres.

Dicho método, llamado Método por Polígonos de Influencia, no es adecuado a la estimación de un bloque, y por tanto, a la estimación de las reservas de un yacimiento.

Por tanto, los objetivos del KRIGEAGE son los siguientes:

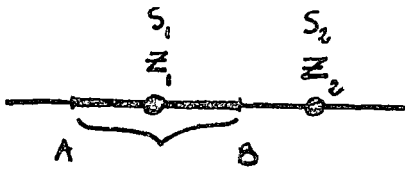
- EVITAR EL SESGO DE ESTIMACION \iff utilizar tanto la información interior como la información exterior.
- EFFECTUAR LA PONDERACION OPTIMA (o en otras palabras, afectar a cada perforación disponible su peso óptimo) de manera de obtener la VARIANZA DE ESTIMACION MINIMA.

Naturalmente el KRIGEAGE toma en cuenta

- el aspecto estructural del problema mediante el variograma: generalmente mientras más lejana es una perforación, menos es su peso (o influencia) en la estimación.
- la forma, el tamaño del panel que se estima
- la implantación relativa del bloque y de las muestras
- el soporte de la muestra. Esta noción es sumamente importante: se sabe que la varianza de las leyes de las muestras es menor que la varianza de las leyes reales de los bloques. Dicho fenómeno es el origen de las sobre-estimaciones de las explotaciones selectivas.

11.- SISTEMA DE KRIGEAGE

(1) Caso sencillo



Sea el problema de estimar la ley - promedio del segmento AB mediante las muestras S1 y S2.

Sea Z la ley real desconocida de AB, y Z1, Z2 las leyes de las muestras S1, S2.

¿Cómo efectuar la ponderación óptima?

- El estimador de KRIGEAGE es una combinación lineal de las informaciones disponibles. Sea Z* este estimador.

Se escribe:

$$Z^* = \lambda_1 Z_1 + \lambda_2 Z_2$$

con $\begin{cases} \lambda_1 \text{ peso a afectar a } S_1 \\ \lambda_2 \text{ peso a afectar a } S_2 \end{cases}$

λ_1, λ_2 , son desconocidos

- Este estimador Z* debe ser insesgado: entonces λ_1 y λ_2 verifican la condición siguiente:

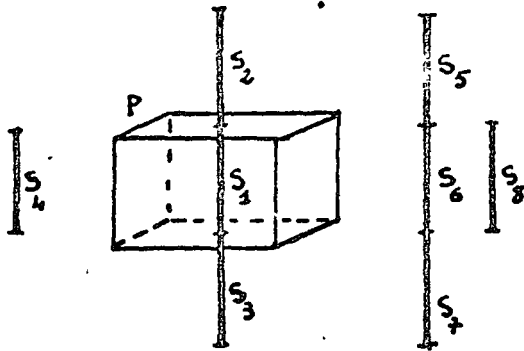
siendo $Z^* - Z = \lambda_1 Z_1 + \lambda_2 Z_2 - Z$ el error de estimación,

error nulo en promedio $\Rightarrow E [Z^* - Z] = 0 \Rightarrow \lambda_1 + \lambda_2 = 1$

- Bajo la condición $\lambda_1 + \lambda_2 = 1$, se determinan λ_1 y λ_2 minimizando la varianza de estimación o sea la varianza del error $[Z^* - Z]$

$$\sigma_E^2 \{ [Z^* - Z]^2 \} \text{ m\u00ednima} = \sigma_K^2 \text{ o varianza de Krigeage}$$

(2) Caso general



Se estima el bloque P (ver figura).

La información disponible se compone de barrenos interiores y exteriores a P: en este ejemplo se dispone de un tramo de barreno interior S_1 y de 7 tramos de barreno exteriores $S_2 \dots S_8$.

Se conocen precisamente la ley Z_i y la implantación de cada tramo S_i .

Qué peso afectara cada una de estas 8 informaciones para encontrar el mejor estimador de la ley promedio Z de P?

Según el párrafo anterior, el estimador Z^* se escribe:

$$Z^* = \sum_i \lambda_i Z_i$$

Estimador insesgado $\Rightarrow E[Z - Z^*] = 0 \Rightarrow \sum_{i=1}^8 \lambda_i = 1$

Estimador óptimo $\Rightarrow \lambda_i, i=1, 8$, tales como $E\{[Z - Z^*]^2\}$ mínima.

El sistema de Krigage siguiente entrega

- los valores de los ponderadores o pesos $\lambda_i, i=1, 8$
- la varianza de estimación mínima o varianza de krigage σ_K^2

Se escribe

$$\begin{cases} \sum_{i=1}^8 \lambda_i \bar{\gamma}(S_i, S_j) + \mu = \bar{\gamma}(S_j, P) \\ \sum_{i=1}^8 \lambda_i = 0 \\ \sigma_K^2 = \sum_i \lambda_i \bar{\gamma}(S_i, P) + \mu - \bar{\gamma}(P, P). \end{cases}$$

' μ ' = constante de Lagrange

El sistema de Krigage es un sistema de ecuaciones lineales en el cual intervienen integrales geométricas fácilmente calculables. - (iguales que las de la fórmula de la varianza de estimación).

En este ejemplo y bajo la hipótesis de isotropía estructural, el Krigage encontrará

- las simetrías evidentes $\lambda_2 = \lambda_3$
 $\lambda_5 = \lambda_7$

- las desigualdades tales como

$$\lambda_1 \geq \lambda_j, \quad j=2,8$$

$\lambda_6 \geq \lambda_8$ expresando que el barrenos S_6 hace pantalla a la influencia de S_8 (efecto de pantalla)

$\lambda_4 \leq \lambda_6$ expresando que el barrenos S_4 ha transferido una parte de su influencia sobre los barrenos S_5, S_7, S_8 .

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 .

III.- APLICACIONES

La práctica del Krigage permite

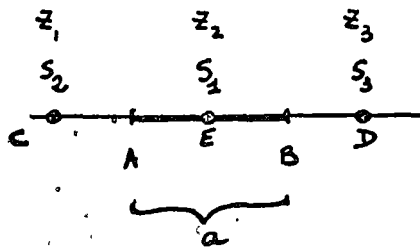
- determinar la curva tonelaje-ley que proporciona el porcentaje de toneladas de mineral de ley mayor que la ley de corte t_c .

- programar (día a día, 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.

APLICACION PRACTICA:

KRIGEAGE DE LA LEY PROMEDIO DE UN SEGMENTO



Se considera el segmento AB reconocido por la muestra central S_1 y dos muestras exteriores S_2 y S_3 (ver fig.). Se conocen las leyes Z_1, Z_2, Z_3 de las muestras S_1, S_2, S_3 .

Problema: efectuar el Krigeage del segmento AB, siendo el semivariograma lineal isótropo ($\gamma(h) = h$)

A la muestra S_1 , le corresponde el peso λ . Tomando en cuenta la simetría, al conjunto $(S_2 + S_3)$ que llamamos S_m , le corresponde el peso $1 - \lambda$, de manera que cumpla la condición de estimación insesgada.

$$\sum_i \lambda_i = \lambda + (1 - \lambda) = 1$$

El sistema de Krigeage se escribe

$$\begin{cases} \sum_i \lambda_i \bar{\gamma}(S_j, S_i) + \mu = \bar{\gamma}(S_j, AB) \\ \sum_i \lambda_i = 1 \end{cases}$$

$$\sigma_K^2 = \sum_i \lambda_i \bar{\gamma}(S_i, AB) + \mu - \bar{\gamma}(AB, AB) \quad \text{varianza de KRIGEAGE}$$

de donde en este ejemplo,

$$\lambda \bar{\gamma}(S_1, S_1) + (1 - \lambda) \bar{\gamma}(S_1, S_m) + \mu = \bar{\gamma}(S_1, AB)$$

$$\lambda \bar{\gamma}(S_m, S_1) + (1 - \lambda) \bar{\gamma}(S_m, S_m) + \mu = \bar{\gamma}(S_m, AB)$$

Se calcula cada $\bar{\gamma}$

$$\bar{\gamma}(S_1, S_1) = 0$$

$$\bar{\gamma}(S_1, S_m) = \bar{\gamma}(S_m, S_1) = \frac{1}{2} [\gamma(EC) + \gamma(ED)] = \gamma(a) = a$$

$$\bar{\gamma}(S_m, S_m) = \frac{1}{4} [\gamma(CC) + \gamma(CD) + \gamma(DC) + \gamma(DD)] = \frac{\gamma(2a)}{2} = a$$

$$\bar{\gamma}(S_1, AB) = \frac{1}{a} \int_0^a \gamma\left(\frac{a}{2} - x\right) dx = x\left(\frac{a}{2}\right) = \frac{a}{4}$$

$$\bar{\gamma}(S_m, AB) = \frac{1}{a} \int_{\frac{a}{2}}^{\frac{3a}{2}} \gamma(x) dx = \frac{3}{2} x\left(\frac{3a}{2}\right) - \frac{1}{2} x\left(\frac{a}{2}\right) = a$$

Entonces se obtiene

$$\begin{cases} (1-\lambda)a + \mu = \frac{a}{4} \\ \lambda a + (1-\lambda)a + \mu = a \end{cases} \Rightarrow \lambda a = a - \frac{a}{4} = \frac{3a}{4}$$

⇒

$\lambda = \frac{3}{4}$
$1-\lambda = \frac{1}{4}$
$\mu = 0$

- la muestra central S_1 tiene la mayor influencia: el kriging le atribuye el peso más importante $\lambda = \frac{3}{4}$

- cada muestra exterior, S_2 y S_3 , tiene el peso $\frac{1-\lambda}{2} = \frac{1}{8}$

Cálculo de la varianza de estimación

$$\sigma_K^2 = \lambda \bar{\gamma}(S_1, AB) + (1-\lambda) \bar{\gamma}(S_m, AB) + \mu - \bar{\gamma}(AB, AB)$$

con $\bar{\gamma}(AB, AB) = F(a) = \frac{a}{3}$, viene

$$\sigma_K^2 = \frac{3}{4} \times \frac{a}{4} + \frac{1}{4} a - \frac{a}{3} = \frac{5a}{48}$$

$\sigma_K^2 = \frac{5a}{48}$

Nota importante

Estimando el segmento AB afectando a cada información el mismo peso, es decir, sin kriging, se obtiene la varianza de estimación - como sigue, aplicando la fórmula general:

$$\sigma_{Est}^2 = \frac{2}{3a} \left\{ 2 \int_{a/2}^{3a/2} \gamma(x) dx + \int_0^a \gamma\left(\frac{a}{2} - x\right) dx \right\} - F(a) - \frac{1}{9} \{ 3\gamma(0) + 4\gamma(a) + 2\gamma(2a) \}$$

$$\sigma_{Est}^2 = \frac{2}{3} \left\{ 3 X\left(\frac{3a}{2}\right) - X\left(\frac{a}{2}\right) + X\left(\frac{a}{2}\right) \right\} - F(a) - \frac{4}{9} \gamma(a) - 2 \frac{\gamma(2a)}{9}$$

$$\sigma_{Est}^2 = 2X\left(\frac{3a}{2}\right) - F(a) - \frac{4}{9} \gamma(a) - \frac{2}{9} \gamma(2a)$$

con los valores particulares al variograma lineal, se obtiene

$$\sigma_{Est}^2 = \frac{5a}{18}$$

y la relación $\sigma_{Est}^2 > \sigma_K^2$

$$\frac{5a}{18} > \frac{5a}{48}$$

El Krigeage mejora mucho la precisión de la estimación

LAS DECISIONES SECUENCIALES

I. - INTRODUCCION

Consideremos un yacimiento de tonelaje mineral 'T' conocido, pero -
cuya ley promedio 'Z' es desconocida.

Un estudio económico preliminar permitió definir:

- la ley límite de explotabilidad: m_0
- el coeficiente 'A' o beneficio por tonelada de metal recuperada.
'A' depende de los costos de explotación, de las inversiones y del precio del metal.

Si Z fuera conocida, el beneficio total se escribiría:

$$B = AT(Z - m_0)$$

En la práctica minera Z es desconocida; se dispone solamente de su estimación geoestadística. Entonces se calculará el valor probable del beneficio o sea $E\{B\} = AT(E\{Z\} - m_0)$

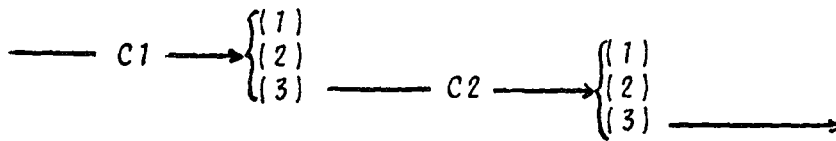
* La primera campaña de perforaciones C1 proporciona la estimación m_1 de la ley promedio Z y la Geoestadística cuantifica su precisión bajo la forma de la varianza de estimación $\sigma_{E_1}^2$.

* Dados estos resultados, cuál decisión tomar?

- (1) abandonar el yacimiento
- (2) explotar el yacimiento
- (3) efectuar una segunda campaña C2

Se tomará la decisión a la cual le corresponde el mayor $E\{B\}$.

Elegir la solución (3) es encontrar otra vez las mismas posibilidades.



De donde el nombre de esquema secuencial de decisiones que se da a este conjunto de eventualidades.

En esta presentación, nos limitaremos al problema que consiste en saber si, después de C2, se debe explotar o no el yacimiento.

Notas Importantes

1.- A veces, se plantea el problema contrario: Z conocida y T desconocido. En tal caso, se escribe el beneficio total:

$$B = AZ \{T - T_0\}$$

2.- No se tomarán en cuenta los factores políticos, sociales, financieros que intervienen en este tipo de problema.

II.- ESTUDIO DE LA PRIMERA CAMPANA

Se conocen m_1 y $\sigma_{E_1}^2$.

$$\underline{m_1 < m_0}$$

$$E [B_1] = AT(m_1 - m_0)$$

ya que se tiene $E [Z] = m_1$

$$E [B_1] < 0$$

(estimación insesgada)

El valor promedio del beneficio es negativo y nunca se decidirá explotar. Pero queda la alternativa:

- Abandonar el yacimiento
- Efectuar una segunda campaña

Dicha relación o fórmula de aditividad es sumamente general. Tiene un significado objetivo y no implica un modelo probabilístico - más o menos arbitrario.

La Geoestadística permite calcular $\sigma_K^2 (Y-Z)$ previamente a la realización de C2. En la práctica, C2 no es independiente de C1 y se conoce la implantación de las perforaciones de C2 (se tratará, por ejemplo, centrar una retícula).

Ley de Probabilidad de Y

Se conocen solamente los 2 primeros momentos de la variable aleatoria Y: su promedio m_1 y su varianza σ^2 . Entonces, se elige en la práctica una ley de probabilidad normal o lognormal.

Sea $f(y)$ dy la densidad de probabilidad de Y.

Valor promedio del beneficio

Si fuera conocido Y, se escribiría

$$B_2' = AT(Y - m_0)$$

Siendo desconocido Y, se calcula el valor promedio de B_2' , o sea

$$E[B_2'] = AT \int_{m_0}^{\infty} (y - m_0) f(y) dy.$$

Hay beneficio si $y > m_0$; por tanto y varía entre m_0 y ∞ .

Con $L = \frac{m_1 - m_0}{\sigma}$, y tomando en cuenta el costo de realización P de C2, se obtiene:

$$E[B_2] = E[B_2'] - P = AT\sigma \Psi(L) - P$$

IV.- TOMA DE DECISION

Se examina la situación después de C1

$$\underline{m_1 < m_0} \Rightarrow t < 0 \Rightarrow E[B_1] < 0$$

- Si $E[B_2] < 0$, se abandona el yacimiento
- Si $E[B_2] > 0$, se programa la segunda campaña.

$$\underline{m_1 > m_0} \Rightarrow t > 0 \Rightarrow E[B_1] > 0$$

Después de C1, la esperanza del beneficio es positiva. Para decidir entre explotar de inmediato y efectuar una segunda campaña, se comparan $E[B_1]$ y $E[B_2]$

- Si $E[B_2] - E[B_1] < 0$, se explota de inmediato el yacimiento.
- Si $E[B_2] - E[B_1] > 0$, se programa la segunda campaña.

V.- GRAFICA DE DECISION

$$\underline{m_1 - m_0 < 0} \Rightarrow t < 0$$

$$E[B_2] = A\sigma\varphi(t) - P. \text{ La expresión } K(t) = \frac{E[B_2]}{A\sigma} = \varphi(t) - \frac{P}{A\sigma}$$

cuantifica la ventaja que se obtiene ampliando el reconocimiento. Si $K(t) > 0$, se efectúa C2.

$$\underline{m_1 - m_0 > 0} \Rightarrow t > 0$$

$$E[B_2] - E[B_1] = A\sigma\varphi(t) - P - A\sigma t = A\sigma[\varphi(t) - t] - P$$
$$K'(t) = \frac{[\varphi(t) - t]}{A\sigma} \text{ representa la ventaja que se obtiene ampliando el reconocimiento.}$$

Si $K'(t) > 0$, se efectúa C2.

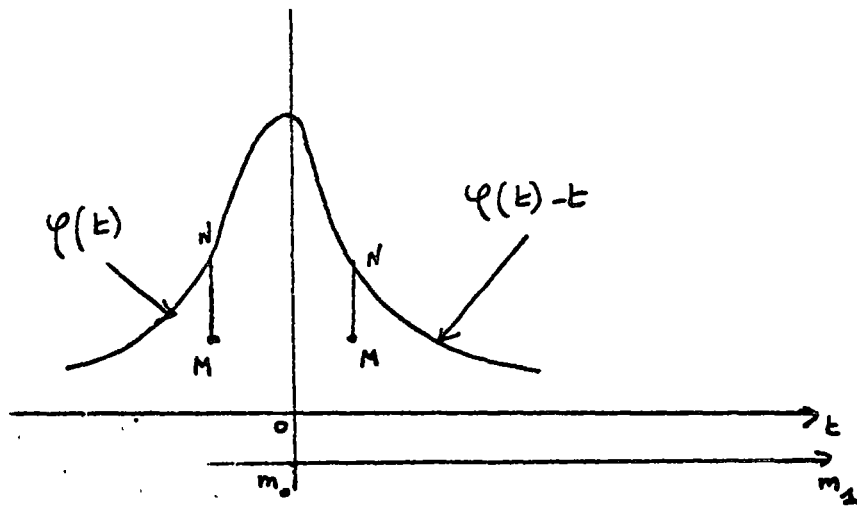
En la misma gráfica, se diseñan

para $t < 0$, la curva $K_1(t) = \varphi(t)$

para $t > 0$, la curva $K_2'(t) = \varphi(t) - t$

Sea el punto M de coordenadas $\left\{ \begin{array}{l} t = \frac{m_1 - m_0}{\sigma} \\ \text{ordenada } \frac{P}{AT\sigma} \end{array} \right.$

Sea el punto N de coordenadas $\left\{ \begin{array}{l} t = \frac{m_1 - m_0}{\sigma} \\ \text{ordenada } \begin{cases} \varphi(t) - t & \text{si } t > 0 \\ \varphi(t) & \text{si } t < 0 \end{cases} \end{array} \right.$



En esta gráfica, la distancia \overline{MN} (de M a N) representa la ventaja que se obtiene ampliando el reconocimiento.

Si $\overline{MN} > 0 \Rightarrow$ se efectúa C2

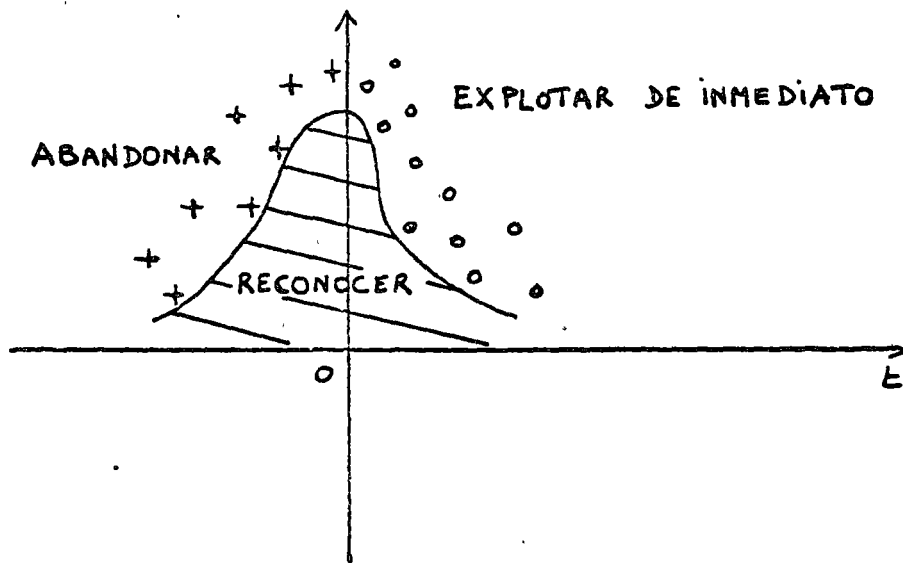
Si $\overline{MN} < 0 \Rightarrow$ - se abandona el yacimiento si $t < 0$

- se explota de inmediato si $t > 0$

VI. - CONCLUSION

Este estudio permite determinar las áreas correspondientes a las 3 decisiones posibles:

- abandonar
- explotar de inmediato
- reconocer

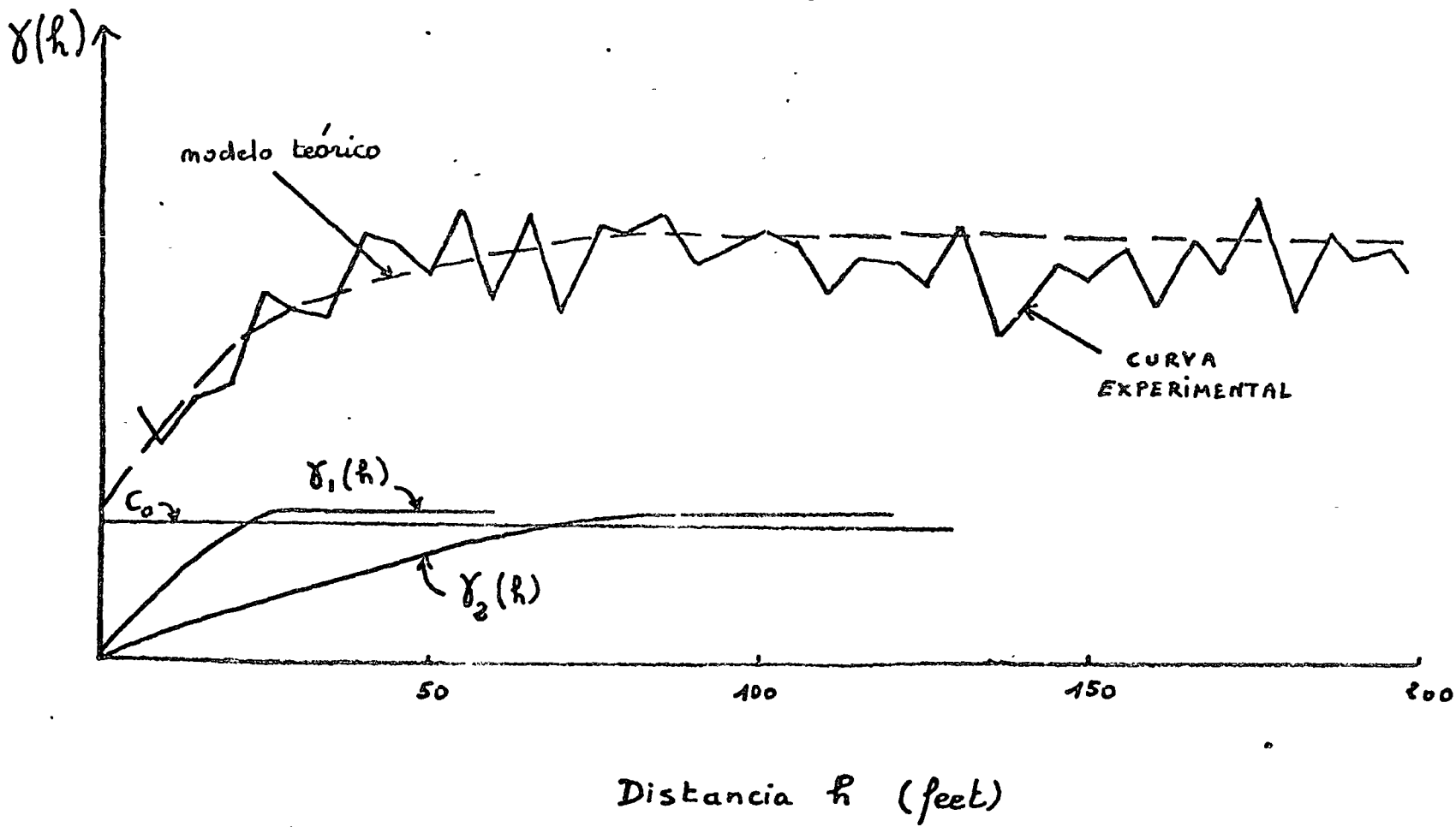


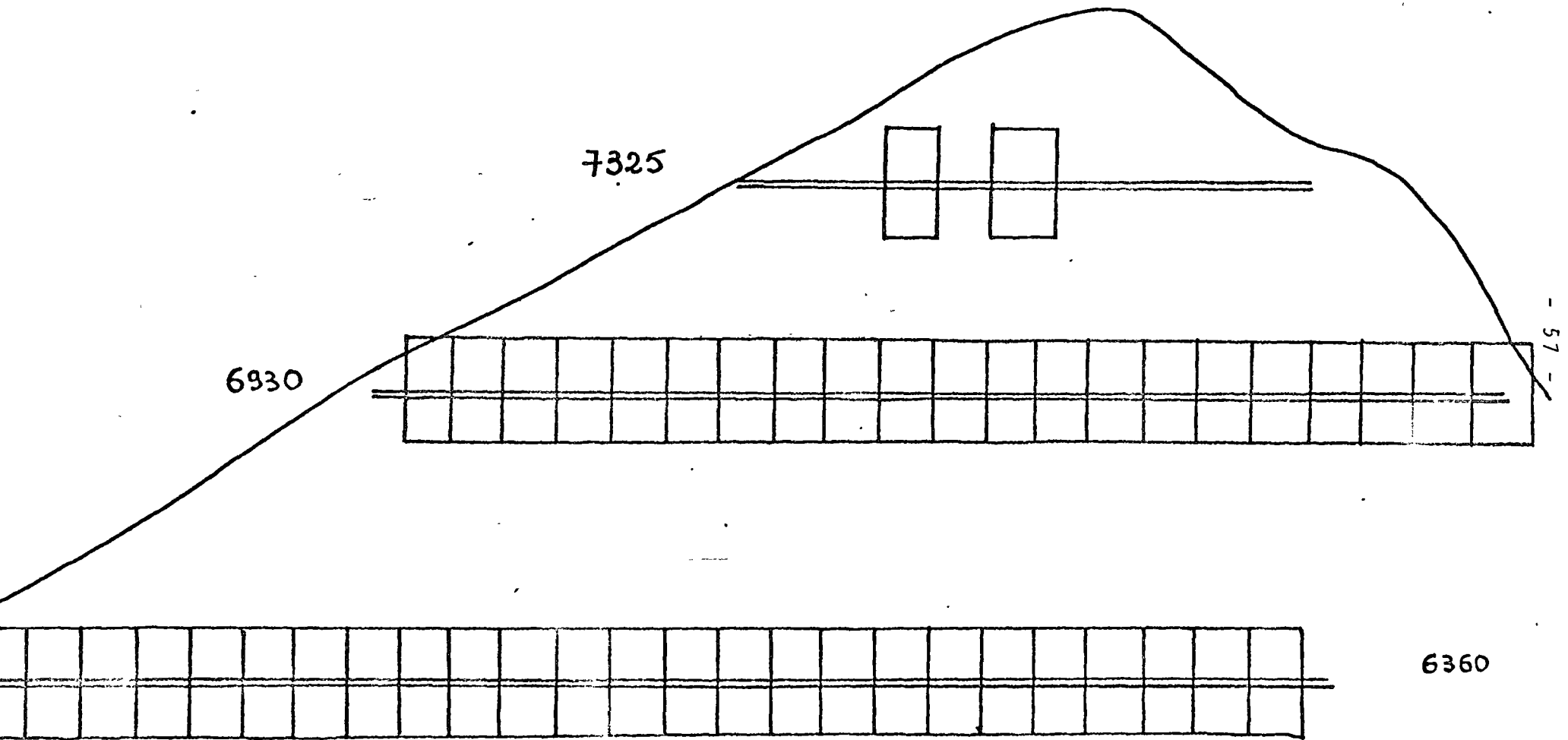
VETA DE COBRE DE EAGLE (CANADA)

El estudio tenía por fin, la Estimación Global de los Recursos in situ, así como la realización de un plan de Krigeage. Se trataba - sobretodo de examinar si el nivel de información disponible permitía alcanzar estos objetivos y, en caso negativo, comparar de manera cifrada la eficiencia de varios suplementos de Reconocimiento.

A continuación se enumeran los puntos claves del estudio:

- * Geología e información disponible
- * Variables Regionalizadas estudiadas
- * Análisis Estructural
 - Construcción de los semivariogramas experimentales; características estructurales mayores
 - semivariogramas teóricos.
- * Estimación Global: potencia y ley promedios así como sus varianzas de estimación. Conclusiones acerca del tipo actual de reconocimiento
- * Estimación local (KRIGEAGE)
 - Configuración de Krigeage: Bloques estimados
 - Precisión de la estimación local
- * Sugerencias para optimizar la Exploración
 - Cálculos guías con el fin de
 - mejorar la estimación global
 - mejorar la estimación local.



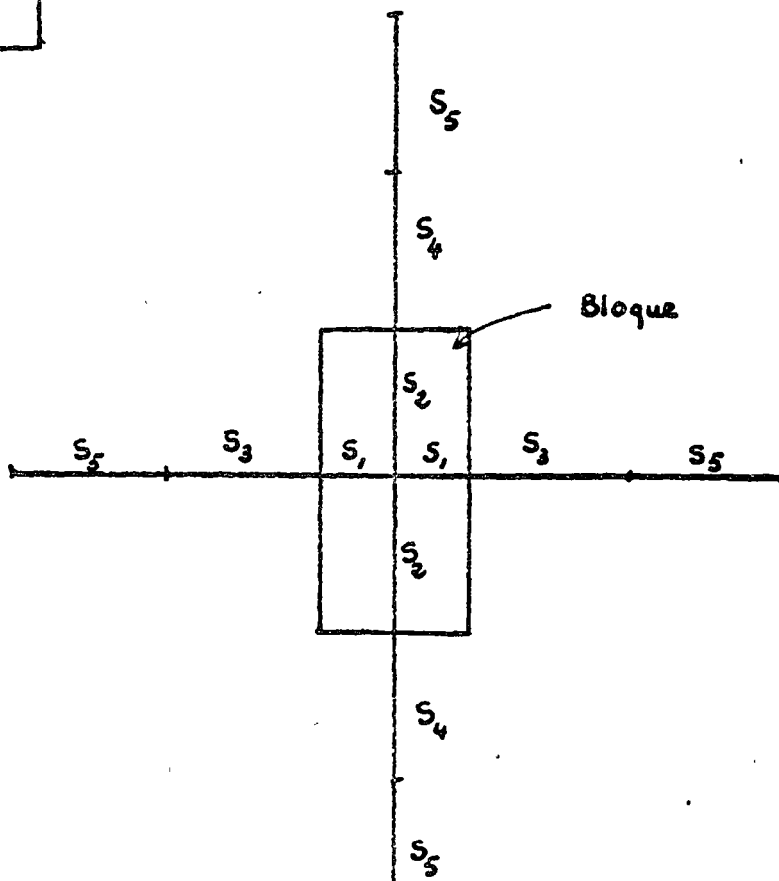
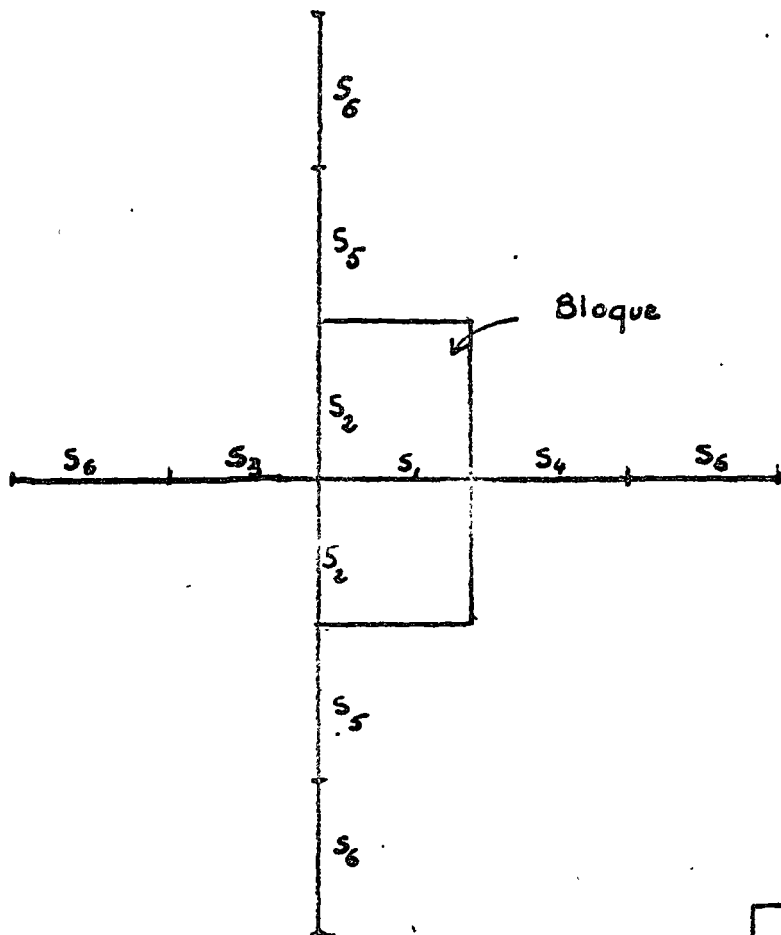
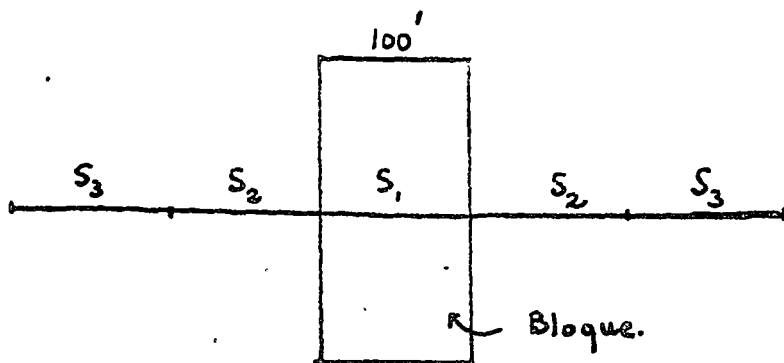


6930

7325

6360

- 57 -



Las Relaciones RECURSOS - RESERVAS

Siendo conocidos los RECURSOS R de un yacimiento G , el estudio de mayor importancia consiste en determinar la fracción $r \in R$ explotable en cierto contexto económico. Dicha fracción r constituye las RESERVAS.

Este problema depende de los factores siguientes:

- la regionalización, caracterizada por el semivariograma $\gamma(h)$. La zona de influencia, el efecto de pepita, las anisotropías, son los parámetros estructurales propios al yacimiento.
- el nivel de información disponible, siendo la malla sistemática o aleatoria estratificada o al azar. Si dada cierta malla, no es posible una estimación correcta de las reservas, es menester definir el suplemento mínimo de reconocimiento necesario.
- la forma, la geometría de los bloques o volúmenes a los cuales se interesa el minero, siendo tales factores bajo la dependencia del método de explotación que es conveniente adoptar.
- varios parámetros de corte tales como una ley de metal, una ley de impureza, una potencia límite de explotabilidad, una relación límite de mineral a escape, etc.
- una cierta función de beneficio que toma en cuenta los datos técnicos anteriores y cuya optimización entrega la selección efectiva.

I.- RECURSOS R del Yacimiento G

Se consideran

- las muestras de volumen v y de semivariograma $\gamma_v(h)$
- los bloques de explotación de volumen V y de semivariograma $\gamma_V(h)$

La ley Z_v de la muestra ^{es} conocida.

La ley Z_V del bloque es desconocida. Sea Z_V^* su estimador.

(1) Histograma de dispersión de las muestras v .

- Valor promedio de las leyes de las muestras $m^* \neq m = E\{Z_v\}$
- Varianza de dispersión $\sigma^2(v/G) = E\{[Z_v - m]^2\}$ m ley promedio del yacimiento
- Forma log-normal, muy a veces en la práctica

De este histograma se deduce la proporción de muestras de ley mayor que z_0 .

(2) Histograma de dispersión de las leyes reales de los bloques V

Supongamos conocidas las leyes reales Z_V

- Valor promedio de las leyes $Z_V : E\{Z_V\} = m$
- Varianza de dispersión $\sigma^2(V/G) = E\{[Z_V - m]^2\}$
- Cierta forma.

El área punteada representa la proporción de paneles de ley promedio $Z_V \geq z_0$

Comparemos estas dos variables Z_v y Z_V

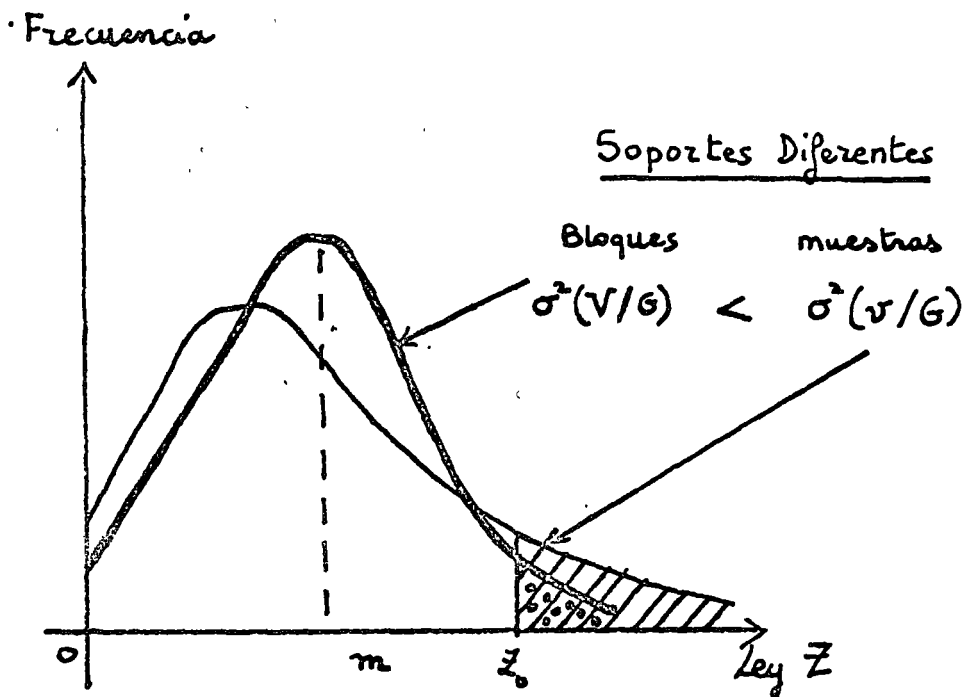
- * soporte diferente $V \gg v$
- * variogramas diferentes $\gamma_v(h)$ y $\gamma_V(h)$
- * varianza de dispersión diferentes $\sigma^2(v/G) < \sigma^2(V/G)$ como lo -- indica la relación de KRIGE

$$\sigma^2(v/G) = \sigma^2(v/V) + \sigma^2(V/G)$$

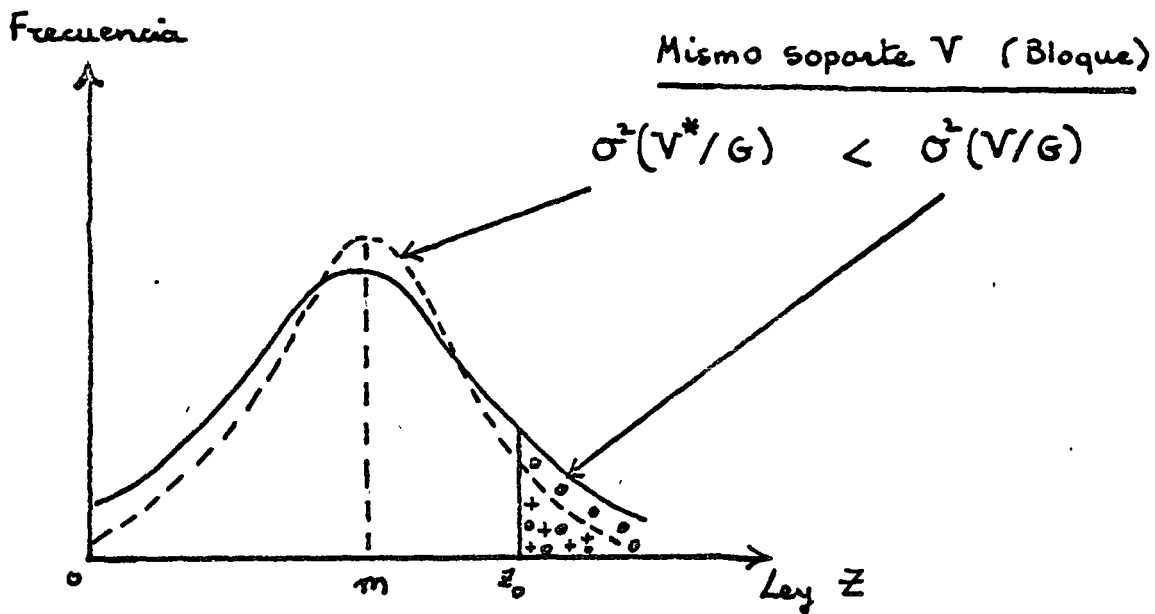
- * cada histograma tiene también su forma propia.

(3) Histograma de las dispersiones Krigeadas.

- valor promedio de los valores Krigeados $E\{Z_v - Z_v^*\} = 0 \Rightarrow E\{Z_v\} = E\{Z_v^*\} = m$
- varianza de dispersión: se demuestra solamente si Z_v^* es un esti -



Dispersiones Reales



Dispersiones Real y Krigada.

mador de Krigeage la relación siguiente:

$$\sigma^2(V/G) = \sigma^2(V^*/G) + \sigma_K^2$$

con $\sigma^2(V/G) = E\{[Z_V - m]^2\}$

$$\sigma^2(V^*/G) = E\{[Z_V^* - m]^2\}$$

$$\sigma_K^2 = E\{[Z_V - Z_V^*]^2\}$$

- una cierta forma

La selección efectiva se hace en base a las leyes estimadas Z_V^* .

Este histograma permite entonces preveer la proporción de paneles recuperados de ley $Z_V^* \geq z_0$

Los recursos in situ R se caracterizan por

- los semivariogramas $\gamma_V(h)$, $\gamma_{V^*}(h)$
- los histogramas de dispersión o sea las varianzas de dispersión $\sigma^2(V/G)$, $\sigma^2(V^*/G)$, $\sigma^2(V/G)$

Las nociones esenciales son:

- la noción de soporte
- la noción de dispersión
- la noción de estimación

II.- RESERVAS n del yacimiento G

(1) Definición de un criterio de selección.

Con el objeto de seleccionar las reservas n entre los recursos R, se trata controlar la calidad del mineral recuperado, definir las condiciones optimales de explotación o más generalmente, optimizar una cierta función.

$$B \{ X(t_1, t_2, \dots, t_n), Y(t_1, t_2, \dots, t_n), V(t_1, t_2, \dots, t_n), V^*(t_1, t_2, \dots, t_n) \}$$

en la cual X, Y, V, V' , son las características del mineral explotado.

Por ejemplo, se puede escribir B

$$B = f(X, Y) - \beta V - \beta' V' \quad \text{donde}$$

X representa la cantidad de metal útil extraído

Y representa la cantidad de metal impureza extraído

$f(X, Y)$ el valor económico del mineral extraído

βV el costo de explotación y del beneficio del volumen V de mineral

$\beta' V'$ el costo de descapote correspondiente a un volumen V' de encape.

X, Y, V, V' dependen ellos mismos de un cierto número de parámetros de corte t_1, t_2, \dots, t_n : ley de corte t_1 , relación de mineral a encape t_2 , etc.

A cada conjunto $\{t_1, t_2, \dots, t_n\}$ de parámetros de corte le corresponde un cierto mineral recuperado de características $X(t_1, t_2, \dots, t_n)$, $Y(t_1, t_2, \dots, t_n)$ y el beneficio $B(X, Y, \dots)$

* el problema es definir el conjunto $\{t_1^0, t_2^0, \dots, t_n^0\}$ óptimo que proporciona B máximo, bajo ciertas restricciones.

* llamaremos "criterio de selección" dicha función B de las características del mineral eventualmente recuperable

* el conjunto $\{t_1^0, t_2^0, \dots, t_n^0\}$ se llama parámetro óptimo de selección.

Generalmente, la Gerencia impone el criterio de selección. Para definir $\{t_1^0, t_2^0, \dots, t_n^0\}$ parámetro óptimo de selección, es menester estudiar cómo varían las características $X(t_1, t_2, \dots, t_n)$, $Y(t_1, t_2, \dots, t_n)$, $V(t_1, t_2, \dots, t_n)$... en función del parámetro de corte $\{t_1, t_2, \dots, t_n\}$.

Pero, en la práctica, nunca se conocen $X(t_1, t_2 \dots t_n)$, $Y(t_1, t_2 \dots t_n)$ etc. sino más bien sus estimaciones $X^*(t_1, t_2, \dots t_n)$, $Y^*(t_1, t_2 \dots t_n)$ etc. dado cierto nivel de información disponible

(2) Nivel de información disponible

Sea un yacimiento que se compone de un conjunto de bloques elementales de volumen V . Se considera el parámetro de corte: $t =$ ley de metal.

La explotación es selectiva: siendo t fijo, se explotarán los paneles cuya ley estimada $Z_V^* \geq t$. En el momento cuando se hace el estudio, se disponen de los estimadores Z_V^* de todos los bloques y del histograma de dispersión de las leyes Z_V^* . Por tanto, se puede preveer, para cada valor del parámetro t , la proporción de bloques recuperados: cada bloque tiene una ley real Z_V y su estimador es tal que $Z_V^* \geq t$. El mineral recuperado se constituye del conjunto de las leyes reales Z_V bajo la condición $Z_V^* \geq t$. La ley promedio del mineral recuperado es

$$E \{ Z_V / Z_V^* \geq t \} = X(t)$$

$X(t)$ es desconocida. Se estima $X(t)$ mediante el valor promedio de las leyes estimadas Z_V^* de los bloques recuperados o sea

$$X^*(t) = E \{ Z_V^* / Z_V^* \geq t \}$$

Solamente si el estimador Z_V^* es un estimador de krigage, la estimación de $X(t)$ por $X^*(t)$ experimentalmente disponible es insesgada.

La varianza de estimación de $X(t)$ por $X^*(t)$

$$\sigma_E^2(t) = E \{ [X(t) - X^*(t)]^2 \}$$

Se deduce de las varianzas de krigage elementales $\sigma_K^2 = E \{ [Z_V - Z_V^*]^2 \}$

Se admite entonces que $X(t)$ sigue una ley de probabilidad gaussiana de media $X^*(t)$ y de varianza $\sigma_E^2(t)$. Se conocen también la ley de las características $Y(t), V(t), \dots$ y se puede calcular la esperanza del beneficio $B(X, Y, V, \dots)$ sea $E\{B(t)\}$. Por lo tanto se determina el parámetro óptimo de selección $t=t_0$ que maximiza $E\{B(t)\}$:

- $X^*(t_0), Y^*(t_0), V^*(t_0)$ son los estimadores insesgados de las reservas recuperadas $r(t_0)$

- Se conocen las varianzas de estimación $\sigma_E^2(t_0)$ de dichas características.

LAS SIMULACIONES CONDICIONALES

- Aplicación al yacimiento de Niquel de Prony.

El estudio de las Reservas Recuperadas proporciona la estimación global de las características del mineral recuperado y permite a veces localizar dichas Reservas en el yacimiento.

A pesar de su importancia, estos datos no son suficientes para iniciar la explotación efectiva del yacimiento. Es menester aún, conocer las fluctuaciones de las características del mineral seleccionado a varias escalas (día, semana, mes, etc.):

- las fluctuaciones de la ley de metal son esenciales para dimensionar la planta de beneficio y pueden decidir de la necesidad de un área de homogeneización.

- las fluctuaciones de la relación mineral a encape influyen sobre la elección de la máquina de extracción.

- etc.

Desde luego en la práctica, no se conoce la "Realidad", sino la ESTIMACION de esta "Realidad". La "Estimación" no puede ser la solución del problema: un estimador alisa la Realidad y por tanto el yacimiento podría revelarse homogéneo de una manera ficticia.

Ya que no se puede alcanzar la Realidad y que la estimación no restituye las fluctuaciones reales de las características, la solución consiste en SIMULAR la Realidad.

Disponiendo de la simulación; se estudia la influencia de varios esquemas de extracción-almacenamiento-etc., sobre la calidad del mineral y se determina el proceso óptimo de extracción. etc.

Si el objetivo principal de una simulación es la identificación de las fluctuaciones, en la práctica será interesante que en los puntos experimentalmente disponibles se vuelva a encontrar los valores reales: la simulación se llamará entonces Simulación Condicional.

Insistimos aún en los puntos claves del problema:

- la "Realidad" z_0 se caracteriza por

su valor promedio $m = E\{z_0\}$

su semivariograma $\gamma_0(h)$

su varianza de dispersión $E\{[z_0 - m]^2\}$

- la "Simulación" z_s tiene por características

su valor promedio $m = E\{z_s\}$

su semivariograma $\gamma_s(h) = \gamma_0(h)$

su varianza de dispersión $E\{[z_s - m]^2\} = E\{[z_0 - m]^2\}$

su varianza de estimación $E\{[z_s - z_0]^2\}$

- Los valores Krigeados z_k^* se caracterizan por

su valor promedio $m = E\{z_k^*\}$

su semivariograma $\gamma_k(h) \neq \gamma_0(h) \neq \gamma_s(h)$

su varianza de dispersión $E\{[z_k^* - m]^2\} < E\{[z_0 - m]^2\} = E\{[z_s - m]^2\}$

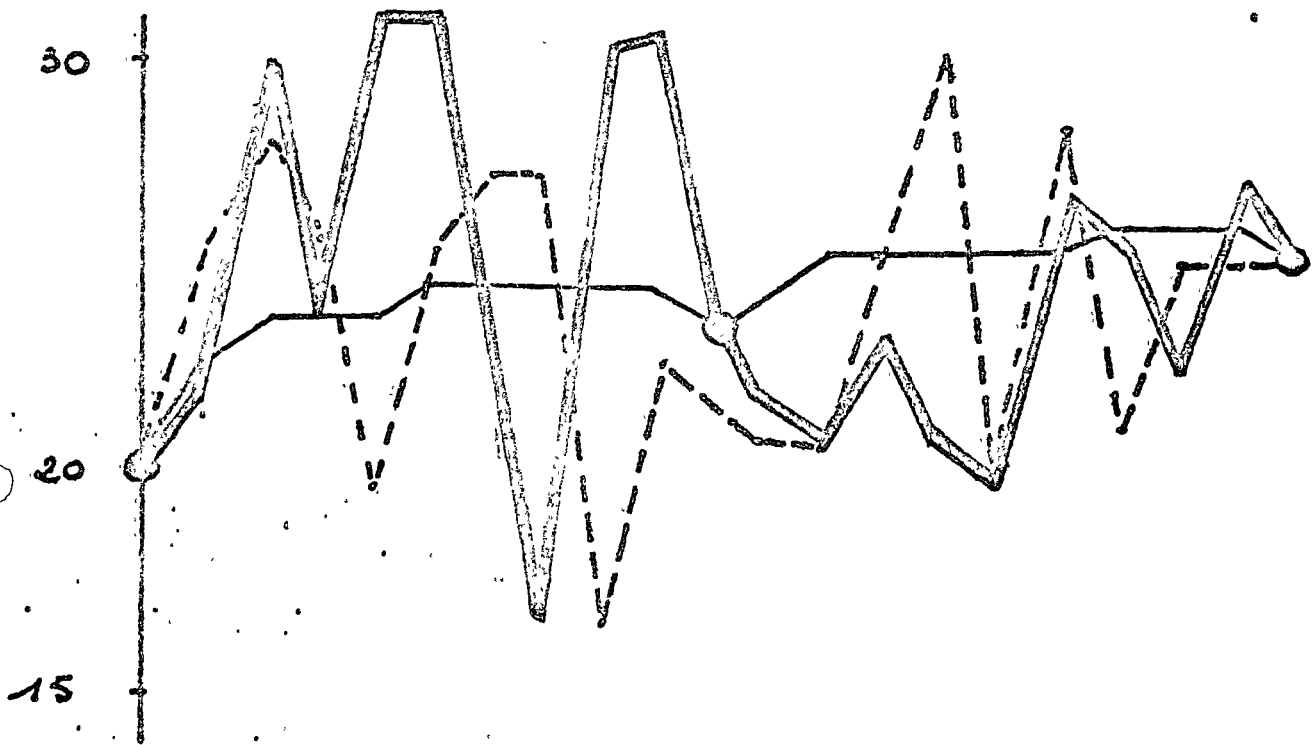
su varianza de Krigage $E\{[z_k^* - z_0]^2\}$ mínima $< E\{[z_s - z_0]^2\}$




La Simulación z_s identifica valor promedio, semivariograma, varianza de dispersión: es otro yacimiento pero de características estructurales iguales al que interesa.

z_s representa una estimación de z_0 pero no es la óptima, ya que se demuestra la relación $E\{[z_s - z_0]^2\} = 2 E\{[z_k^* - z_0]^2\}$

El Krigage z_k^* minimiza la varianza de estimación de z_0 . Es la mejor estimación posible de la "Realidad". Se usará con eficiencia en los-

REALIDAD - SIMULATION - KRIGEAGE



-  $Z_0(x)$ Realidad
-  $Z_s(x)$ Simulación Condicional
-  $Z_k^*(x)$ Krigage

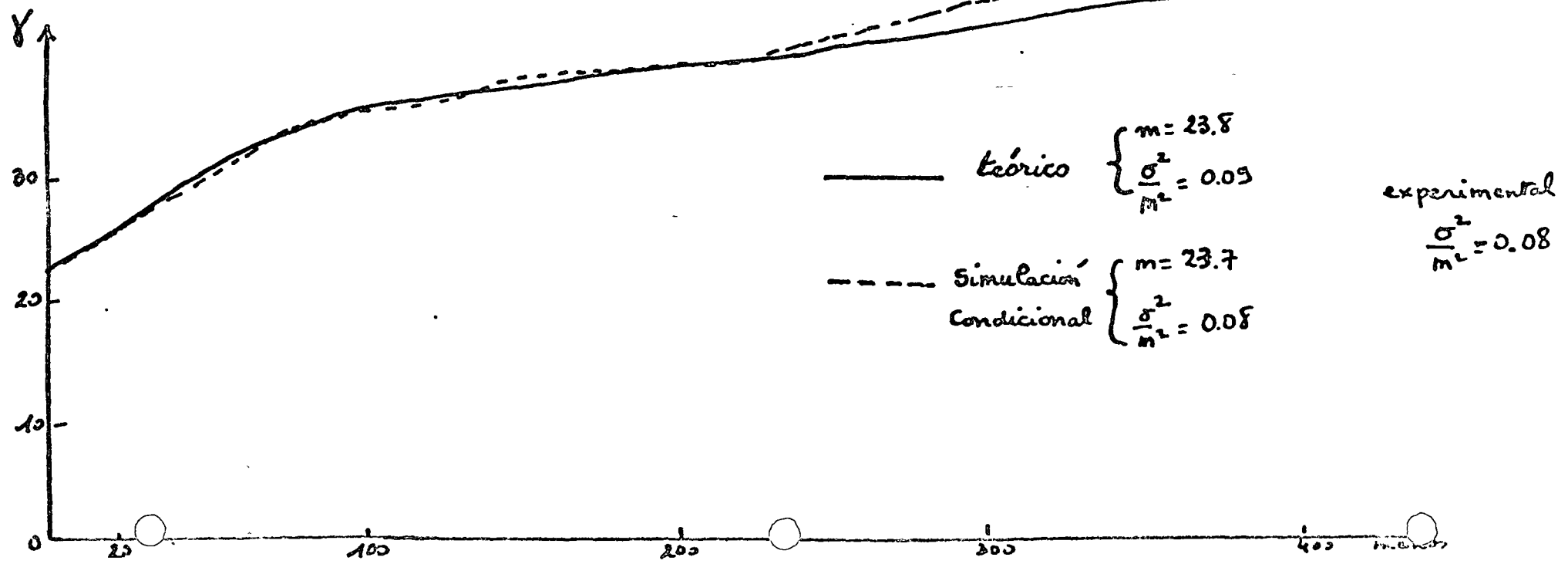
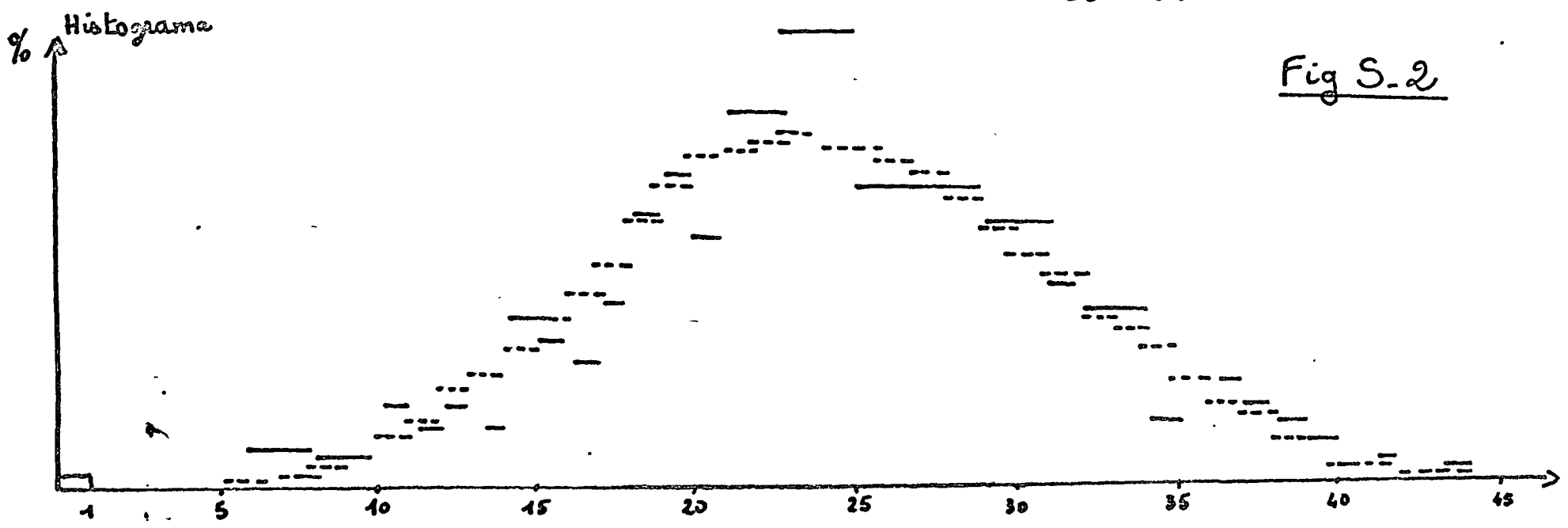
o Interpolación.

Fig 5-1

$$E_2 = 0.0$$

Potencia Mineral Lateritico

Fig S.2



problemas de estimación de Reservas.

La Fig. S.1, ilustra las diferencias entre Realidad - Simulación - Estimación (Kriging).

Aplicación al yacimiento de Níquel de Prong

En la primera zona de producción (cerca de 2 km²) se simuló la realización de una campaña de perforaciones a malla de 10x10 m², sea la simulación de 17600 perforaciones.

En cada perforación simulada, se obtuvieron los valores correspondientes a 24 variables simuladas.

La Simulación impone a cada una de las 24 variables simuladas, volver a encontrar las características experimentales globales que se hablan deducido de las perforaciones reales de la malla de 100 x 100.

Estas características son:

- las medias y varianzas de dispersión dentro de la zona simulada
- los semivariogramas e histogramas
- las correlaciones regionalizadas entre las Variables

Se indican en la Figura S.2

- los 2 histogramas de dispersión experimental y simulado
- los 2 semivariogramas teórico (deducido del experimental) y simulado de la variable potencia de mineral laterítico.

YACIMIENTO DE NIQUEL LATERITICO DE PRONY (NUEVA CALEDONIA)

El reconocimiento del yacimiento de PRONY se hizo por medio de etapas sucesivas, cada etapa cuantificando la eficiencia de las obras realizadas y definiendo el suplemento de reconocimiento que se debía programar para alcanzar el nivel de precisión siguiente. A lo largo del estudio se definieron y se resolvieron varios problemas de estimación de Recursos, Reservas, así como otros problemas de alcance técnico-económico.

Este estudio tiene la ventaja:

- presentar un caso-tipo de Reconocimiento Secuencial de yacimiento
- plantear problemas técnicos de estimación y simulación cuya generalidad no se limita al caso particular del Níquel laterítico.

A continuación se exponen brevemente los resultados de cada etapa:

Etapa 1 reconocimiento geológico preliminar-Realización de la primera campaña sistemática S_1 de perforaciones (malla $400 \times 400 \text{ m}^2$, 150 perforaciones)

Etapa 2 Con los datos de la campaña S_1 , se llevó a cabo un primer estudio geoestadístico G_1 . El análisis estructural de la regionalización en N_i permitió:

- una primera estimación global de los Recursos *in situ*
- un estudio de varianzas con el objeto de determinar la malla necesaria ($\rightarrow 200 \times 200 \text{ m}^2$) para obtener estimaciones locales satisfactorias.

En base a estos resultados, se decidió reducir la malla a una red de 200x200 m² (450 perforaciones) e implantar 2 cruces de perforaciones a malla de 20 m (82 perforaciones) para reconocer las estructuras de la mineralización a pequeña escala.

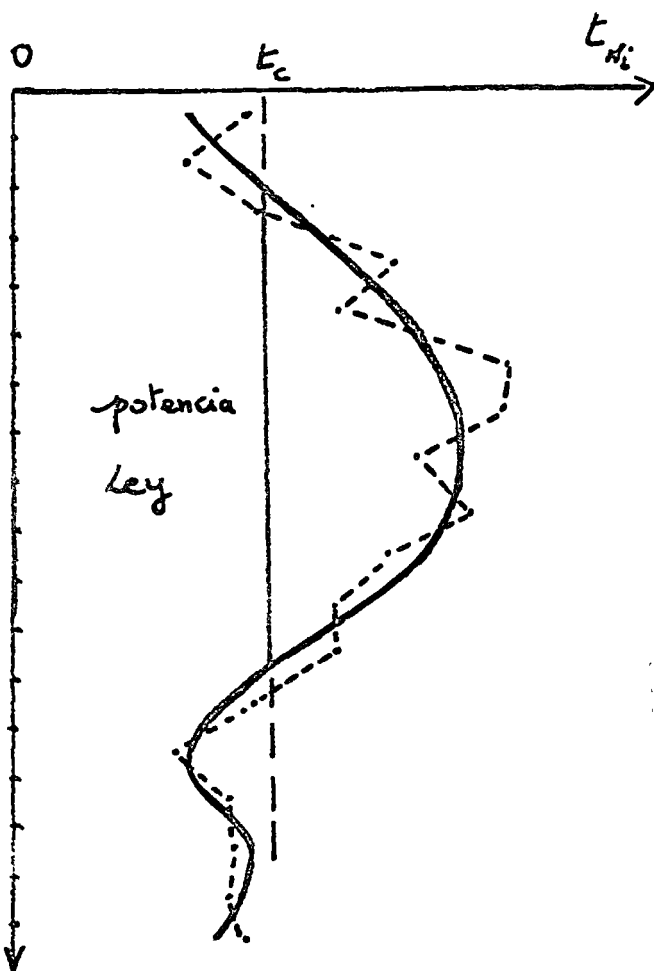
Etapa 3 El segundo estudio geoestadístico G₂

- afinó el análisis estructural
- proporcionó una segunda estimación global más precisa, de los Recursos *in situ*
- mejoró las estimaciones de los paneles 400x400 m² componiendo el yacimiento
- estudió la influencia de los criterios de selección y entrega la estimación de las Reservas Recuperadas

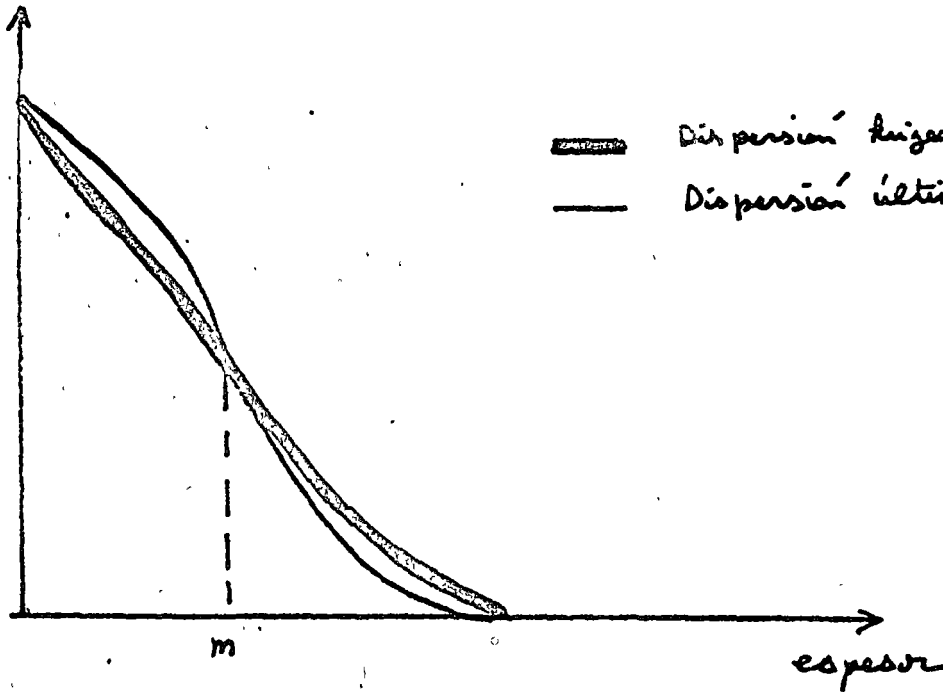
Etapa 4 Siendo elegida la primera zona de explotación, se programó una red de 100x100 m².

El tercer estudio geoestadístico G₃ realizó, en esta zona particular, la simulación de una red mucho más reducida de 10x10 m², con la cual posteriormente el minero puede simular un proceso de explotación y el químico prever las fluctuaciones diarias, semanales, mensuales etc. de las características del mineral al entrar en la planta de beneficio.

Etapa 5 (todavía no realizada pero previsible): para controlar la calidad de la producción, un cuarto estudio geoestadístico G₄, permitirá, mediante un programa sencillo de computadora, utilizando las últimas informaciones, estimar unidades de explotación tan pequeñas como bloques de 20x20 m².



frecuencia acumulada



Cabe hacer notar uno de los resultados esenciales de los 3 primeros estudios geoestadísticos: la Geoestadística necesitó solamente 682 perforaciones para cumplir con los objetivos de la planeación, cuando los métodos más clásicos requerían 2400 perforaciones!

El desarrollo del yacimiento de Prony es un ejemplo tipo de reconocimiento programado armoniosamente por todos los Ingenieros, geólogos, mineros, químicos, geoestadistas, etc. La Geoestadística se reveló muy eficiente y económica para abordar y resolver los problemas planteados, minimizando a cada etapa los gastos de reconocimiento.

YACIMIENTO DE FOSFATOS DE TOGO

El problema planteado por la Compañía COTOMIB era un problema de -- control de producción: se trataba de encontrar la manera óptima de explotar el yacimiento bajo una restricción de calidad. Más precisamente, la Compañía quería conocer los perfiles óptimos del techo y del piso de manera de recuperar el tonelaje metal máximo, siendo la ley de dicho tonelaje recuperado C_p fija de antemano (la Compañía - admitía el riesgo de extraer un mineral de ley inferior a C_p)

Tal problema nunca había sido abordado por los métodos geoestadísticos. Se ha llegado a descubrir, y a aplicar con éxito a este yacimiento, una nueva metodología cuyo carácter general hace posible que sea el campo de aplicaciones de la geoestadística más fructuoso en el futuro.

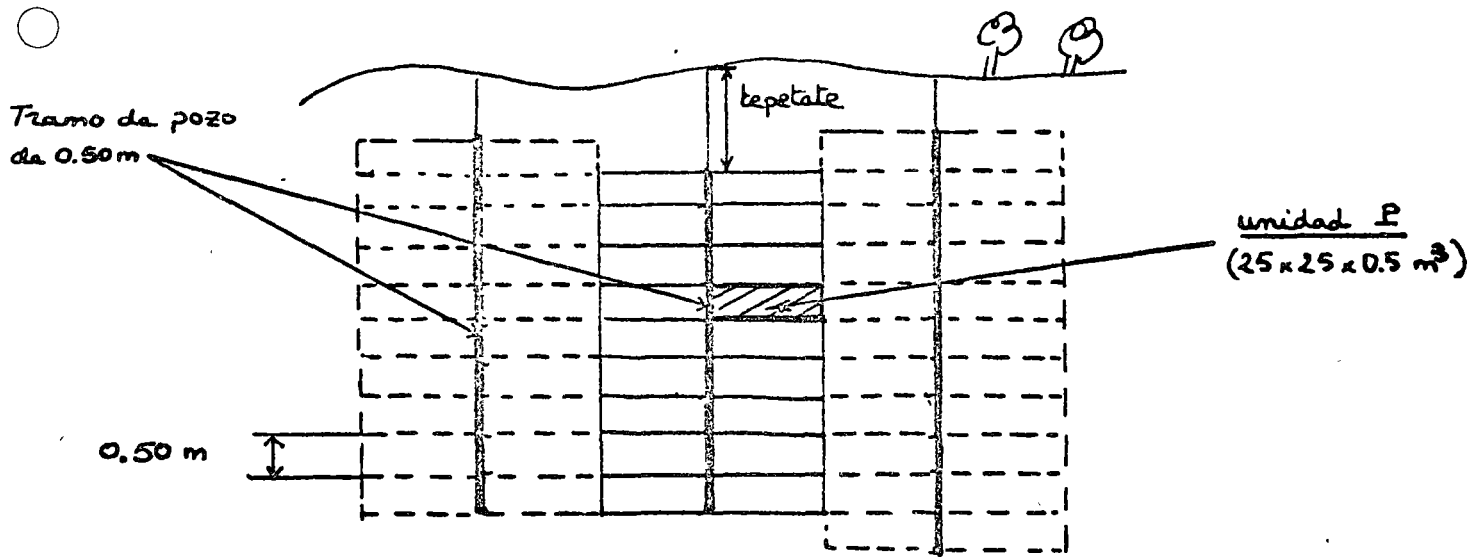
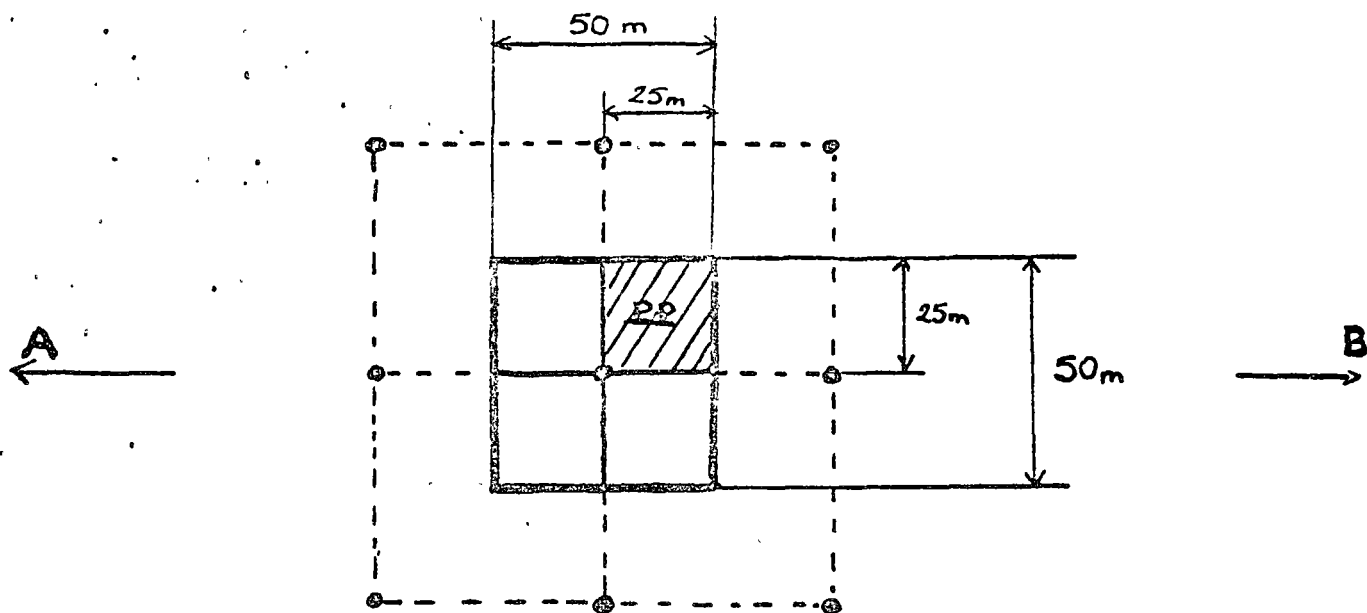
Las etapas del estudio fueron las siguientes:

Etapa 1: Estimación Global

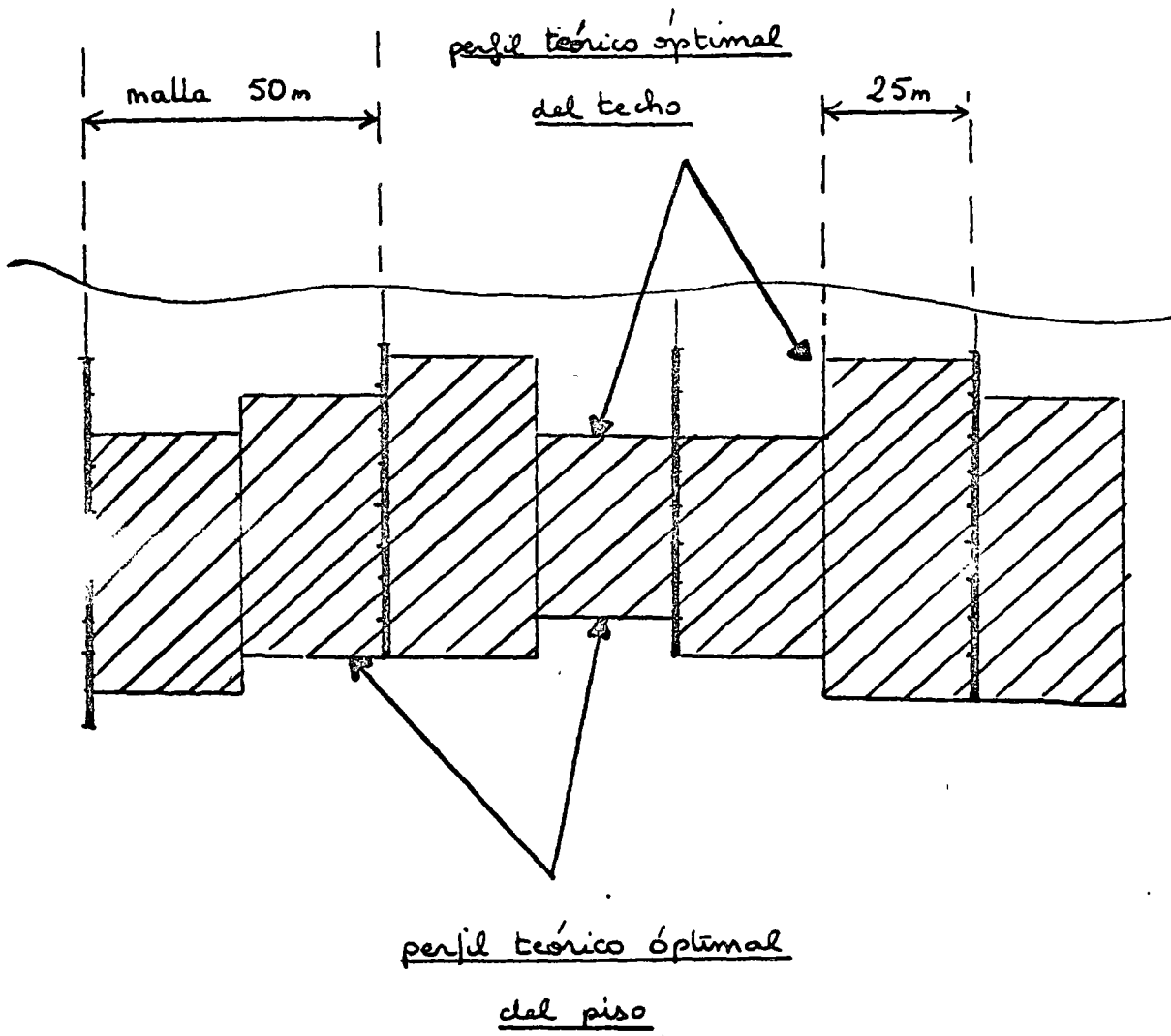
ley promedio
tonelaje metal
tonelaje mineral } así como su varianza de estimación

Etapa 2: KRIGEAGE de las unidades elementales P (ver Figura)

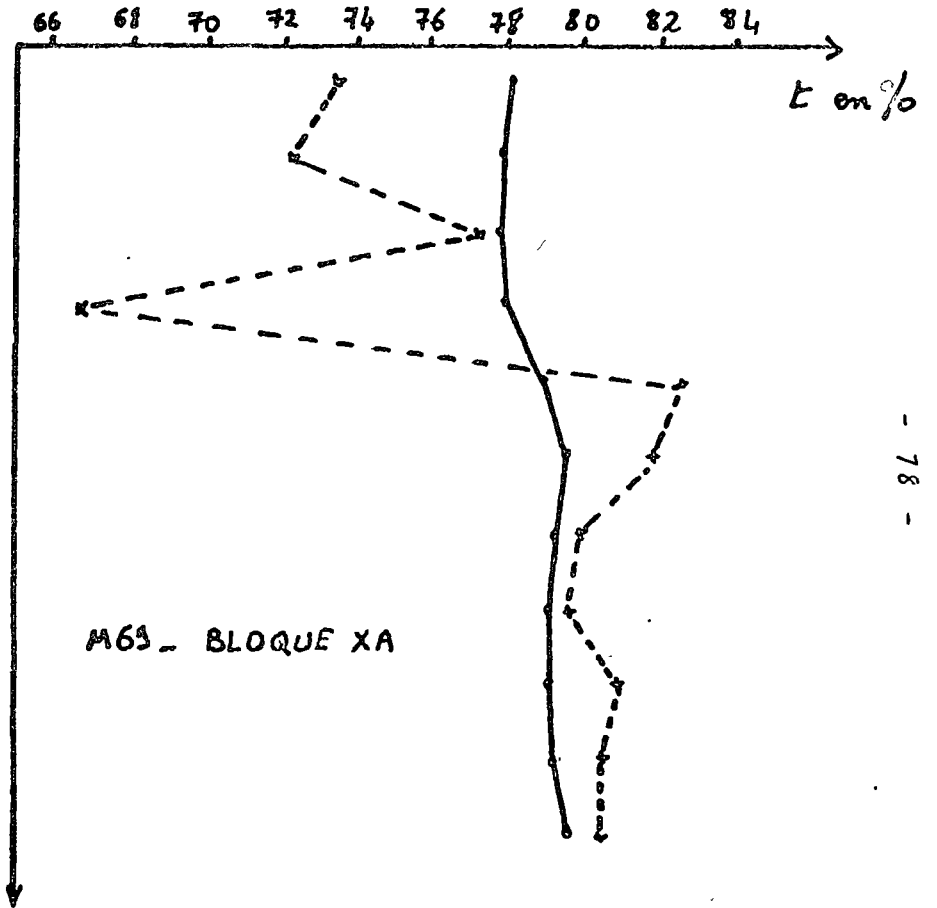
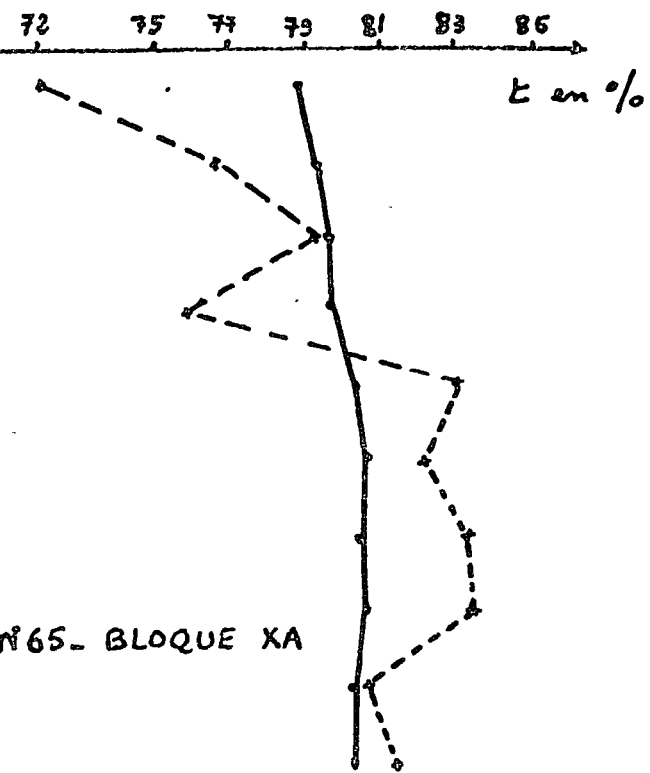
La realización del plano de Krigeage necesitaba un análisis estructural muy fino de la regionalización, el cual no se podía llevar a cabo sin información suplementaria. La Geoestadística permitió estudiarla minimizando dicho suplemento: solo una cruz de 54 perforación ha sido necesaria



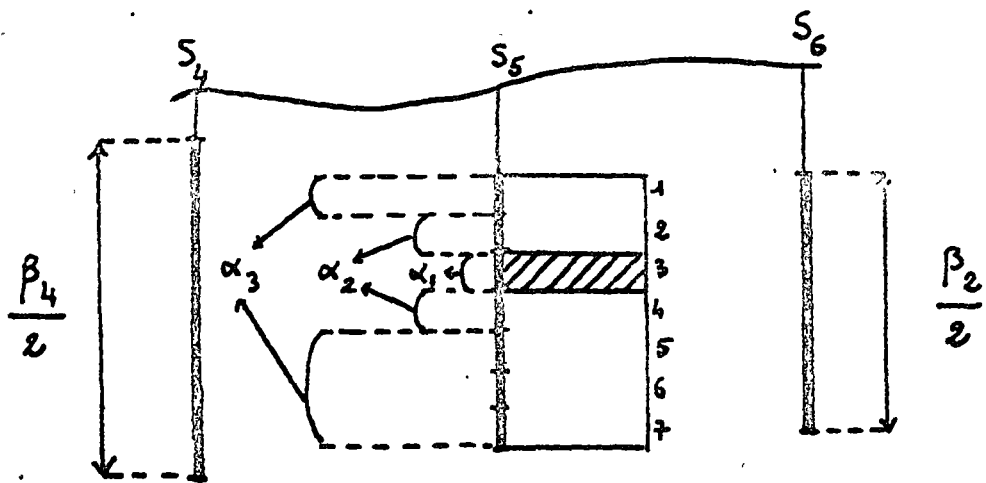
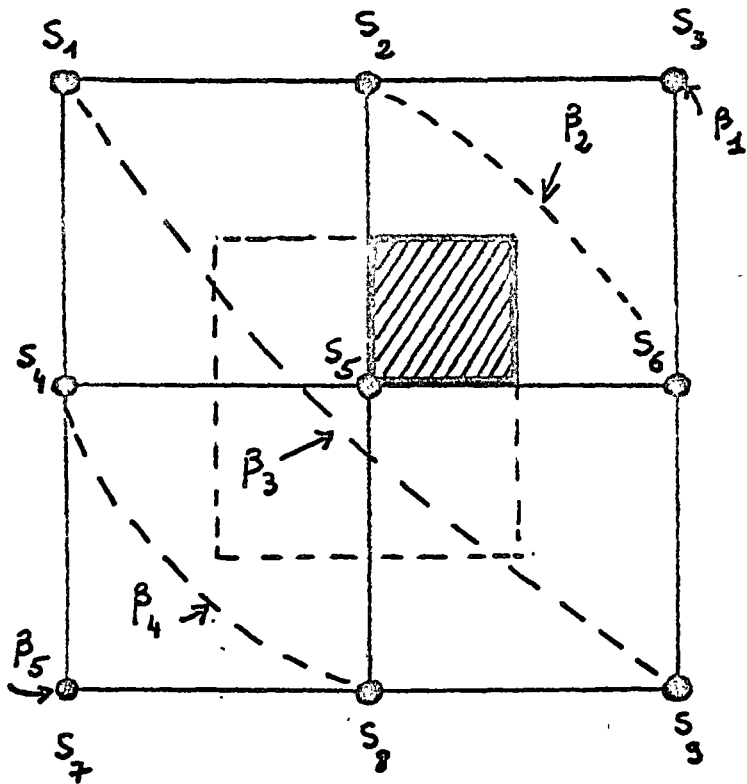
Sección AB



COMPARACION : LEY EXPERIMENTAL - LEY KRIGEADA



--- CURVA EXPERIMENTAL DE LAS LEYES DE LOS TRAMOS (0.5m)
 ——— CURVA DE LAS LEYES KRIGEADAS DE LOS BLOQUES P (25x25x0.5 m³)



CONFIGURACION DE KRIGEAGE

Este plano de krigéage, programado en la computadora, proporcionó va rianzas de krigéage satisfactorias.

Etapa 3: Control de Producción

Se define una zona de explotación y se considera el parámetro ω = ley. Para cada valor del parámetro, un proceso de optimización determina el perfil techo-piso que suministra el tonelaje metal T máximo: a este tonelaje le corresponde cierta ley t . Se determina entonces fácilmente el valor óptimo del parámetro ω_0 que entrega el tonelaje máximo T^0 cuya ley es ζ_2 según el riesgo α .

Ha sido elaborada la programación de la Metodología. Este programa funciona para 5 ζ_2 diferentes y un valor de α .

Por tanto, dado cierto contexto económico y en base a estos documentos, la Compañía determina la mejor explotación posible.

BIBLIOGRAFIA

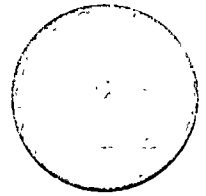
- 1.- AGTERBERG, F.P.- Autocorrelations functions in geology (1970) in "Geostatistics, a colloquium" - pp. 113-141.- Ed.: D.F. Merriam, Plenum Press, New York.
- 2.- DOWD, P.- A digigraphic console program for kriging techniques (1973).- 11th APCOM Symp. Ed.: College of Mines, University of Arizona.
- 3.- 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.
- 4.- HUIJBREGTS, Ch. and SEGOVIA, R.- Geostatistics for the valuation of a copper deposit in Chile (1973) 11th APCOM Symp.- Ed.: College of Mines, University of Arizona.
- 5.- JOURNEL, A.G.- Geostatistics and Sequential Exploration (1973) in "Mining Engineering", Oct. 1973.
- 6.- 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.
- 7.- MARECHAL, A. and ROULLIER, J.P.- Etude géostatistique des gisements de bauxite (1970).- Revue de l'Industrie Minérale, Vol. 52 N° 7, pp. 492-507.
- 8.- MATHERON, G.- Principles of Geostatistics (1963).- in Economic Geology, Vol. 58, pp. 1246-1256.
- 9.- MATHERON G.- Les Variables Régionalisées et leur estimation (1965).- Ed.: Masson, Paris.- 305 pp.
- 10.- MATHERON, G.- The Theory of Regionalized Variables, and its applications (1971).- Les Cahiers du Centre de Morphologie Mathématique, Pooklet No. 5, 211 pp. Ed.: CMM; Ecole Nat, Sup. des Mines, Paris.
- 11.- SANS, H.- Optimisation Geostatistique de la production du gisement de Phosphates du TOGO - 1973 Research Report, COTOMIB Co and CMM, Ecole des Mines, Paris.

- 12.- SERRA, J.- *Echantillonnage et Estimation Locale des phénomènes de transition miniers* (1967).- *Doctorate Thesis* (2 Tomos, 650 pp). Ed.: *University of Nancy, France.*
- 13.- SINCLAIR, A.J.- *A geostatistical study of the Eagle copper vein, N. British Columbia* (1973).- *Canadian Inst. of mining and Metallurgy.*





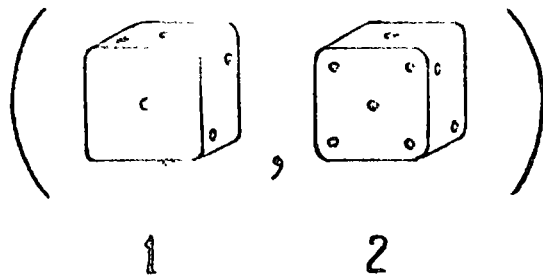
centro de educación continua
facultad de ingeniería, unam



GEOESTADISTICA

ING. SERGIO E. ARREDONDO GALLARDO

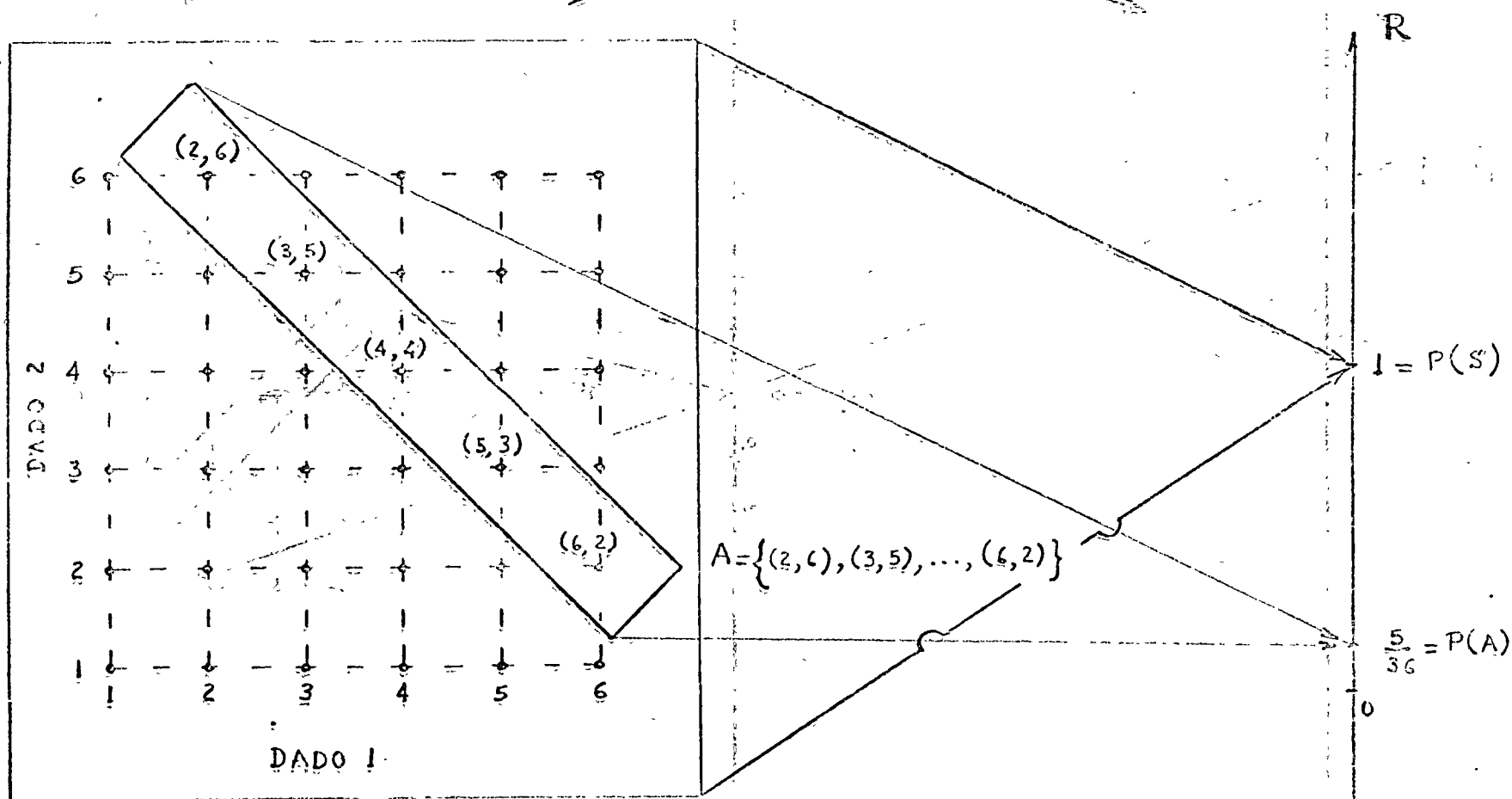
LANZAMIENTO DE DOS DADOS



		RESULTADOS POSIBLES								
		1	2	3	4	5	6			
DADO 2	6	+	-	+	-	+	-	+	-	+
	5	+	-	+	-	+	-	+	-	+
	4	+	-	+	-	+	-	+	-	+
3	+	-	+	-	+	-	+	-	+	
2	+	-	+	-	+	-	+	-	+	
1	+	-	+	-	+	-	+	-	+	
		1	2	3	4	5	6			
		DADO 1								

PROBABILIDAD

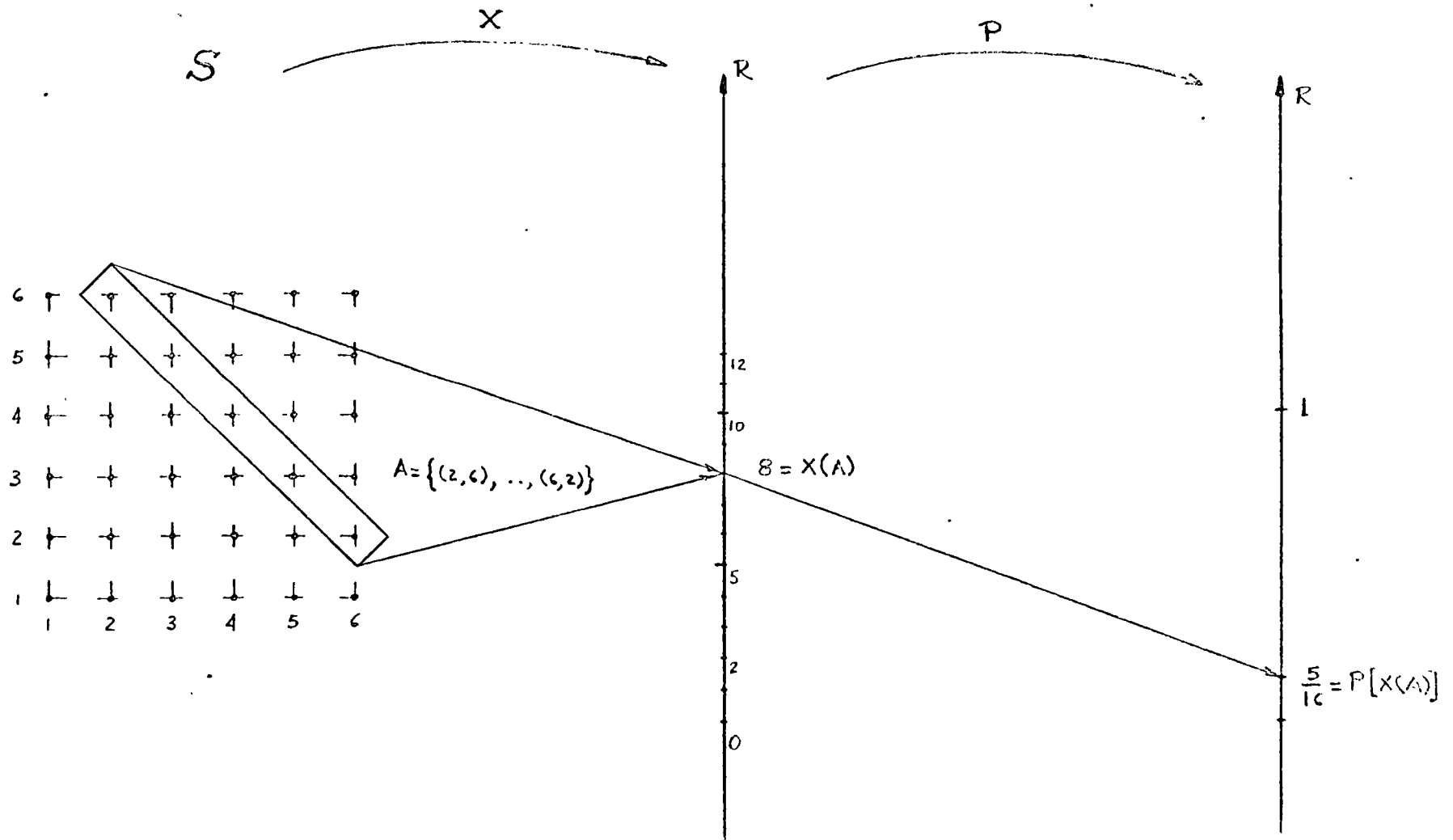
$$S = \{(1,1), \dots, (6,6)\}$$



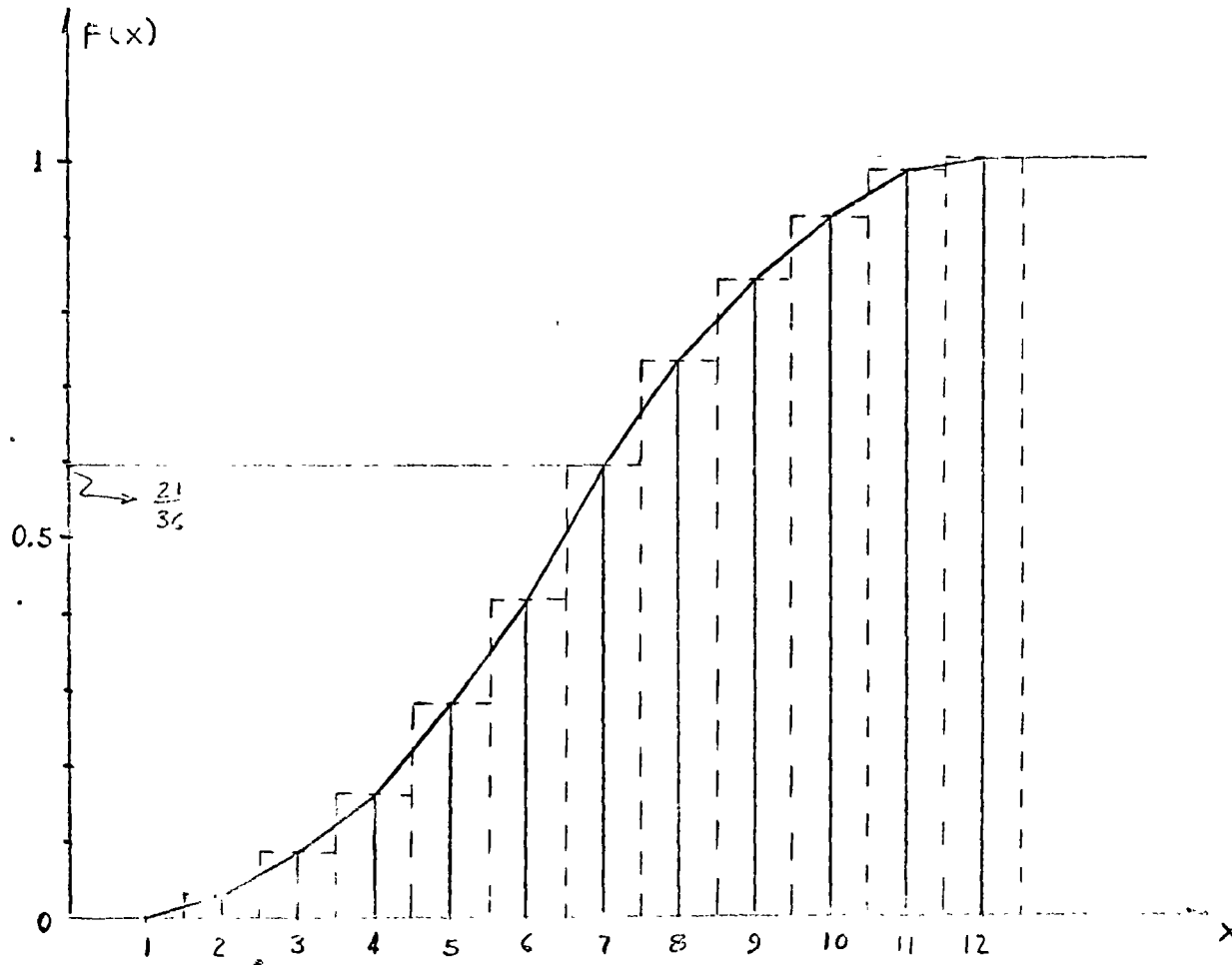
$$P_i = \frac{1}{36}, \quad i = 1, 36$$

$0 \leq P(A) \leq 1$, PARA TODO A QUE SE ENCUENTRE EN S

VARIABLE ALEATORIA (V.A.)



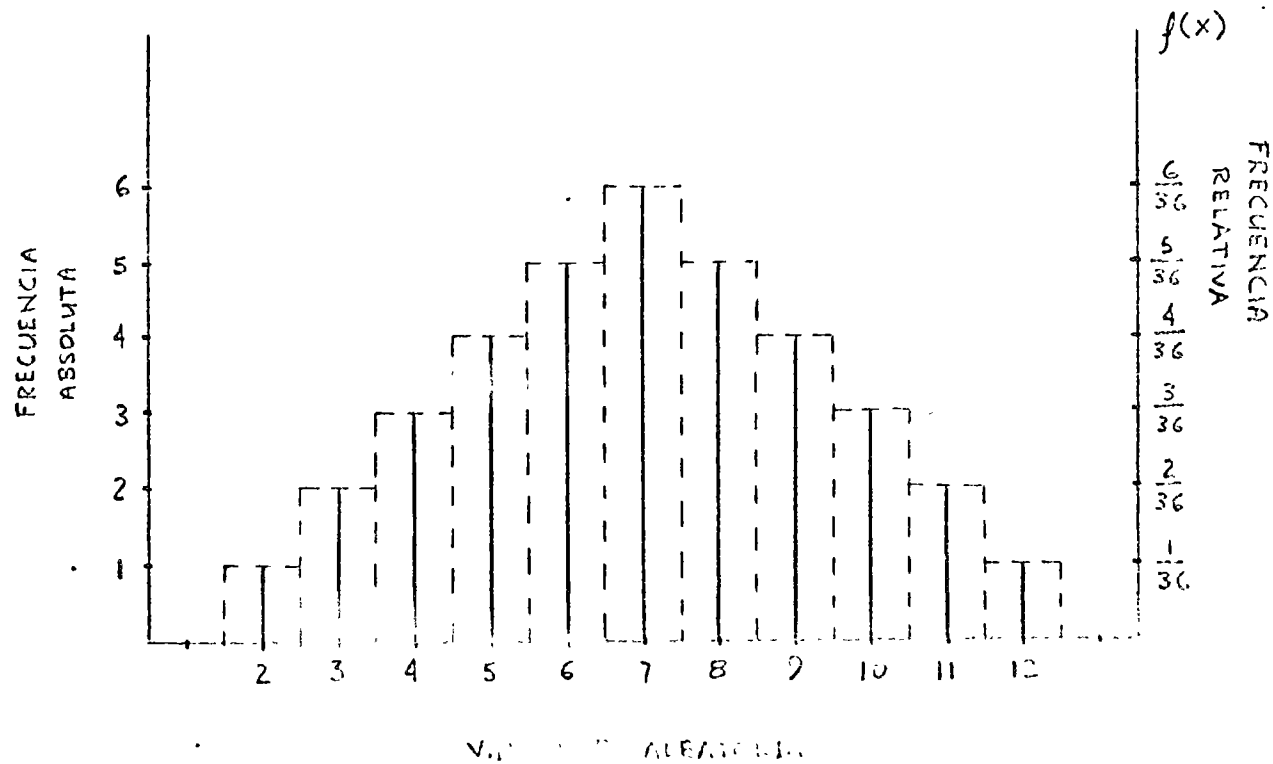
DISTRIBUCION DE FRECUENCIA
ACUMULATIVA



$$F(x_i) = P\{X \leq x_i\} = \sum_{x_i} p(x_i)$$

$$F(7) = P\{X \leq 7\} = p(2) + \dots + p(7) = \frac{21}{36}$$

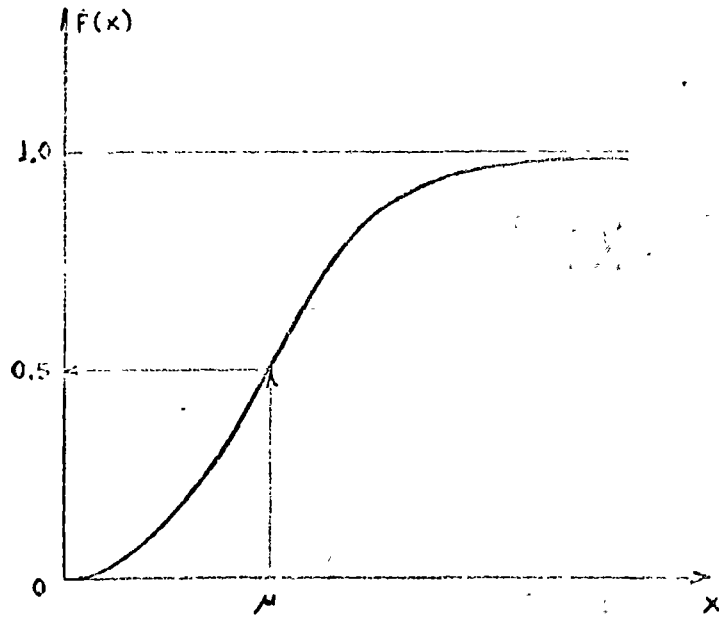
COMPORTAMIENTO DE UNA V.A. O SU LEY DE PROBABILIDAD



HISTOGRAMA

LEYES DE PROBABILIDAD

FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD

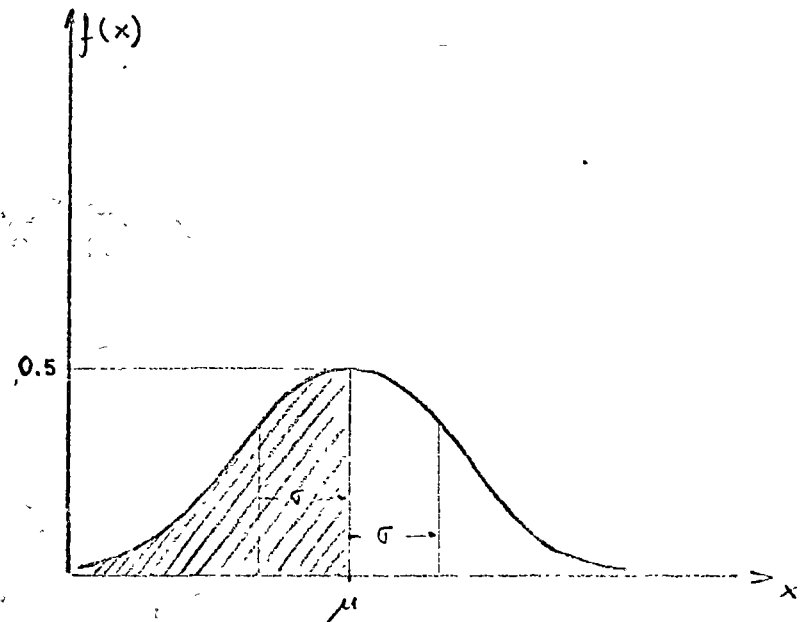


V.A. CONTINUA :

$$F(x) = \int_{-\infty}^{x} f(x) dx, \quad f(x) = \frac{dF(x)}{dx}$$

V.A. DISCRETA : $F(x_i) = \sum_{x_j} p(x_j)$

FUNCIÓN DE DENSIDAD DE PROBABILIDAD



VALOR ESPERADO

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

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

Ejemplo:

$$\begin{aligned} E(x) &= 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + \dots + 11\left(\frac{2}{36}\right) + 12\left(\frac{1}{36}\right) \\ &= 7.056 \end{aligned}$$

MOMENTOS DE PRIMER Y
SEGUNDO ORDEN

MEDIA

$$\mu = E[X] = \begin{cases} \int_{-\infty}^{+\infty} x f(x) dx, & \text{V.A. CONTINUA} \\ \sum_{x_i} x_i p(x_i), & \text{V.A. DISCRETA} \end{cases}$$

VARIANZA

$$\sigma^2 = E[(X - \mu)^2] = \begin{cases} \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx, & \text{V.A. CONTINUA} \\ \sum_{x_i} (x_i - \mu)^2 p(x_i), & \text{V.A. DISCRETA} \end{cases}$$

MOMENTOS

$$\begin{aligned} m'_r &= E[x^r] = \int_{-\infty}^{+\infty} x^r f(x) dx \\ &= \sum_{x_i} x_i^r p(x_i) \end{aligned}$$

$$\begin{aligned} m_r &= E[(x-a)^r] = \int_{-\infty}^{+\infty} (x-a)^r f(x) dx \\ &= \sum_{x_i} (x_i - a)^r p(x_i) \end{aligned}$$

EJEMPLO

MEDIA

$$\mu = 7.056 \Rightarrow 7$$

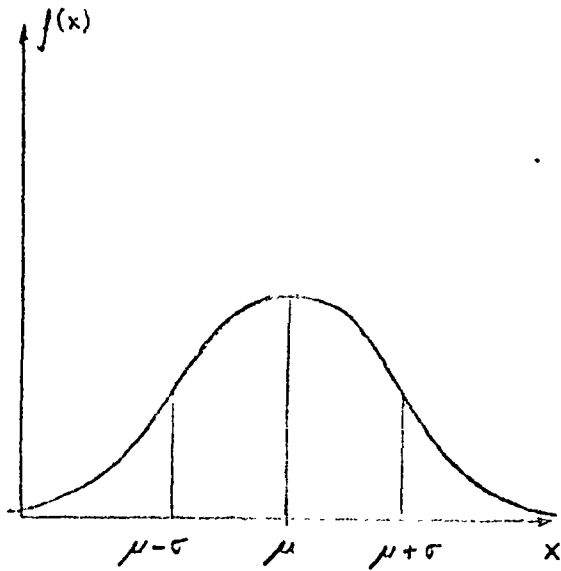
VARIANZA

$$\begin{aligned}\sigma^2 &= (2 - 7.056)^2 \frac{1}{36} + (3 - 7.056)^2 \frac{2}{36} + \dots + (12 - 7.056)^2 \frac{1}{36} \\ &= 18.05\end{aligned}$$

$$\sigma = \sqrt{18.05} = 4.24 \Rightarrow 4$$

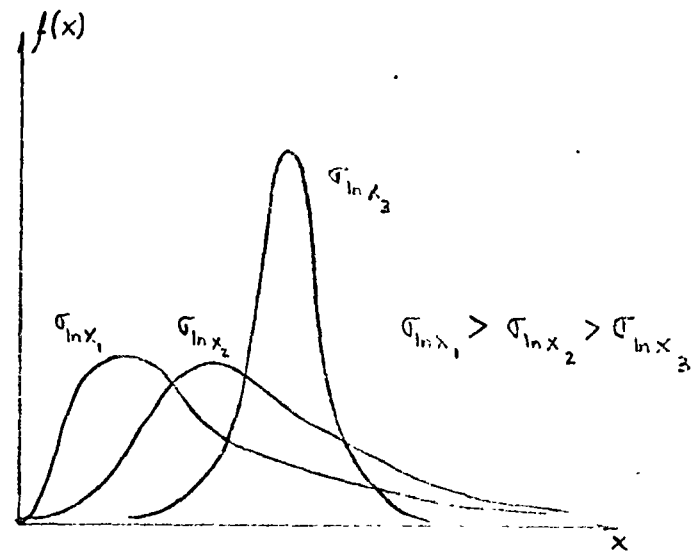
$$\text{DISPERSION} = \sqrt{\sigma^2} = \sigma$$

NORMAL



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

LOGNORMAL



$$f(x) = \frac{1}{x \sigma_{\ln x} \sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{1}{\sigma_{\ln x}} \ln\left(\frac{x}{m_x}\right)\right]^2}$$

$$x \geq 0$$

$$x = e^y$$

$$\text{or } y = \ln x$$

ESTIMADORES

Sea $S = \{1, 2, 3, 4\}$, $\mu = 2.5$, $\sigma^2 = 1.25 \Rightarrow \sigma = 1.11$

TOMENSE TODOS LOS POSIBLES SUBCONJUNTOS DE 3 ELEMENTOS:

$$S_1 = \{1, 2, 3\} , S_2 = \{1, 2, 4\} , S_3 = \{1, 3, 4\} , S_4 = \{2, 3, 4\}$$

ENTONCES:

$$m_1 = 2 , m_2 = 2.33 , m_3 = 2.67 , m_4 = 3 \quad \text{si } P(m_i) = \frac{1}{4}$$

$$\sigma_1^2 = (2-2.5)^2 \frac{1}{4} + (2.33-2.5)^2 \frac{1}{4} + (2.67-2.5)^2 \frac{1}{4} + (3-2.5)^2 \frac{1}{4} = \underline{\underline{0.14}}$$

DE LOS SUBCONJUNTOS FORMADOS, TOMENSE DE DOS EN DOS:

S_1, S_2	S_1, S_3	S_1, S_4	S_2, S_3	S_2, S_4	S_3, S_4
$m_1 = 2.165$	$m_2 = 2.335$	$m_3 = 2.5$	$m_4 = 2.5$	$m_5 = 2.665$	$m_6 = 2.885$

$$\sigma_2^2 = (2.165-2.5)^2 \frac{1}{6} + (2.335-2.5)^2 \frac{1}{6} + \dots + (2.885-2.5)^2 \frac{1}{6} = \underline{\underline{0.042}}$$

DE LOS SUBCONJUNTOS FORMADOS, TOMENSE DE TRES EN TRES

S_1, S_2, S_3	S_1, S_2, S_4	S_1, S_3, S_4	S_2, S_3, S_4
$m_1 = 2.33$	$m_2 = 2.44$	$m_3 = 2.56$	$m_4 = 2.67$

$$\sigma_3^2 = (2.33-2.5)^2 \frac{1}{4} + (2.44-2.5)^2 \frac{1}{4} + (2.56-2.5)^2 \frac{1}{4} + (2.67-2.5)^2 \frac{1}{4} = \underline{\underline{0.016}}$$

TOMENSE TODOS LOS SUBCONJUNTOS A LA VEZ

$$S_1, S_2, S_3, S_4$$

$$m = 2\left(\frac{1}{4}\right) + 2.33\left(\frac{1}{4}\right) + 2.67\left(\frac{1}{4}\right) + 3\left(\frac{1}{4}\right)$$

$$\underline{m = 2.5 /}$$

$$\sigma_4^{*2} = (2.5 - 2.5)^2 (1) = \underline{\underline{0 /}}$$

SE PUEDE OBSERVAR QUE

$$\sigma_1^{*2} > \sigma_2^{*2} > \sigma_3^{*2} > \sigma_4^{*2}$$

PARAMETRO VERDADERO θ

PARAMETRO ESTIMADO θ^*

$$\text{ERROR} = e = (\theta^* - \theta)^2$$

$$E[(\theta^* - \theta)^2] = \text{ERROR MEDIO CUADRATICO}$$

PROPIEDADES DE LOS ESTIMADORES

1. INSESGADO
2. CONSISTENTE
3. ASINTOTICAMENTE EFICIENTE
4. INSESGADO DE VARIANZA MINIMA

ESTIMADOR INSESGADO

θ^* ES INSESGADO SI $E[\theta^*] = \theta$

O EN FORMA EQUIVALENTE, SI $E[\theta^* - \theta] = 0$

EJEMPLO

SEAN x_1, \dots, x_n V.A CON DENSIDAD $f(x_i)$ Y MEDIA μ .

SI $\mu^* = \frac{1}{n} \sum_{i=1}^n x_i$ ENTONCES μ^* ES INSESGADO

DEMOSTRACION

$$\begin{aligned} E[\mu^*] &= E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} E\left[\sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] \\ &= \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} \cdot n \mu = \mu \end{aligned}$$

$$\therefore E[\mu^*] = \mu$$

O TAMBIEN

$$E[\mu^* - \mu] = E[\mu^*] - E[\mu] = E[\mu^*] - \mu = \mu - \mu = 0$$

$$\therefore E[\mu^* - \mu] = 0$$

ESTIMADORES INSESGADOS DE VARIANZA MINIMA

θ^* ES INSESGADO DE VAR. MIN. SI:

1) $E[\theta^*] = \theta$

2) $\sigma^2(\theta^*)$ ES MENOR QUE CUALQUIERA OTRA DE
ESTIMADORES INSESGADOS.

INTERVALOS DE CONFIANZA

SUPONGASE QUE $S = \{1, 2, 3, 4\}$ SE DISTRIBUYE NORMALMENTE
CON MEDIA $\mu = 2.5$ Y $\sigma^2 = 1.25 \Rightarrow \sigma = 1.11$. ESTO ES

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\text{SI } Y = \frac{x-\mu}{\sigma} \Rightarrow f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

LA MUESTRA $S_1 = \{1, 2, 3\}$ TIENE MEDIA $\mu^* = 2$, $\sigma^* = 0.816$

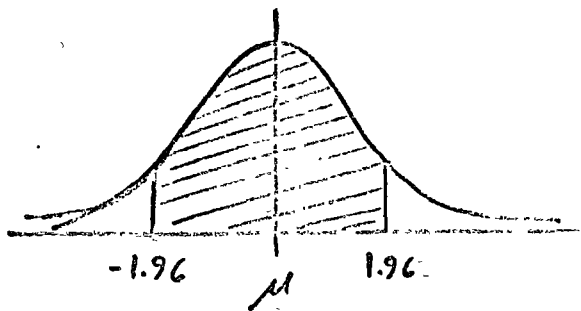
$$\text{SI } \sigma = \frac{\sigma^*}{\sqrt{n}}$$

$$\text{ENTONCES } y = \frac{\mu^* - \mu}{\frac{\sigma^*}{\sqrt{n}}} = \frac{2 - \mu}{\frac{0.816}{\sqrt{3}}} = \frac{2 - \mu}{0.472}$$

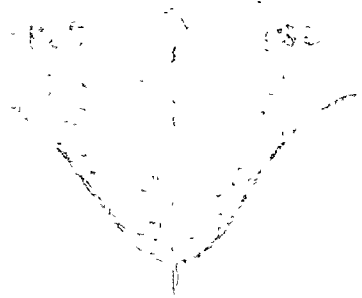
$$P\{-1.96 \leq y \leq 1.96\} = P\left\{-1.96 \leq \frac{2 - \mu}{0.472} \leq 1.96\right\}$$

$$= P\{2 - 1.96(0.472) \leq \mu \leq 2 + 1.96(0.472)\}$$

$$= P\{1.077 \leq \mu \leq 2.923\} = 0.95$$



$$P\{-1.96 \leq y \leq 1.96\} = 0.95$$



Handwritten text at the top right, possibly a title or introductory sentence.

Handwritten text below the first line, possibly a definition or a specific condition.

$$\int_{100}^{150} f(x) dx = \int_{100}^{150} f(x) dx$$

$$\frac{d}{dx} \left(\frac{1}{x} \right) = -\frac{1}{x^2}$$

$$\frac{d}{dx} \left(\frac{1}{x^2} \right) = -\frac{2}{x^3}$$

Handwritten text in the middle section, possibly a list of items or a continuation of the derivation.

$$\frac{d}{dx} \left(\frac{1}{x^3} \right) = -\frac{3}{x^4}$$

$$\frac{d}{dx} \left(\frac{1}{x^4} \right) = -\frac{4}{x^5}$$

Handwritten text in the lower middle section, possibly a conclusion or a note.

Handwritten text at the bottom of the page, possibly a signature or a final note.





centro de educación continua
facultad de ingeniería, unam



GEOESTADISTICA

PROBABILIDAD Y ESTADISTICA

ING. SERGIO E. ARRÉDONDO GALLO

a

NOTAS SOBRE

PROBABILIDAD

Y

ESTADISTICA

SERGIO E. ARREDONDO GALLARDO

OBJETIVO

CONOCER LOS FUNDAMENTOS DE PROBABILIDAD Y ESTADISTICA NECESARIOS PARA EL ESTUDIO Y COMPRENSION 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	

B.2. DISTRIBUCION NORMAL

B.3. DISTRIBUCION LOG-NORMAL

6. PROBLEMAS

IV. REDUCCION DE DATOS

ANEXO A.

BIBLIOGRAFIA

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

NOTACION

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

$$A \not\subset B$$

CONJUNTOS

$A, B, \dots, \{ \}$

ELEMENTOS

a, b, c, \dots, z

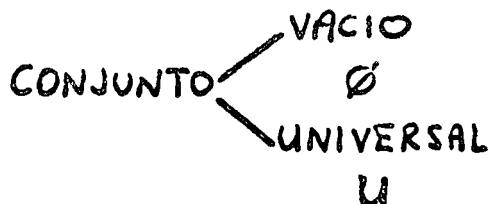
PERTENENCIA



SUBCONJUNTO



TAL(ES) QUE $|, :$



(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

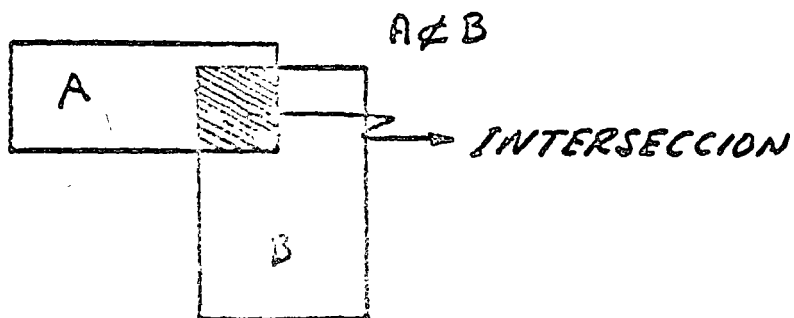
U

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

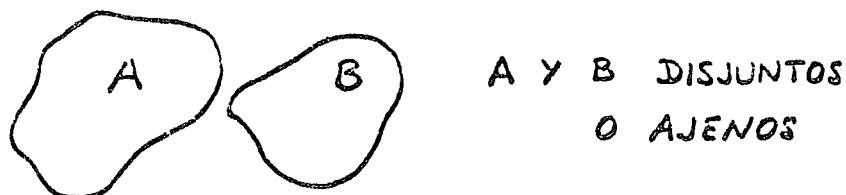
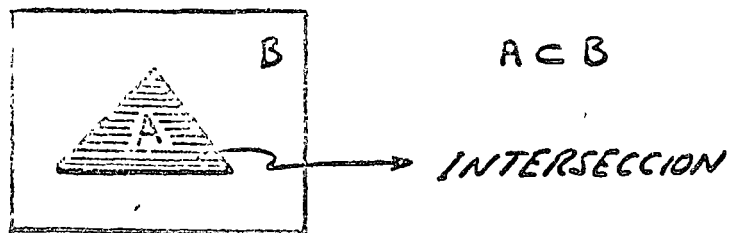
TEOREMA. 2. SEAN LOS CONJUNTOS A, B Y C.
SI $A \subset B$ Y $B \subset C$ ENTON-
CES $A \subset C$.

B. OPERACIONES ENTRE CONJUNTOS. LEYES.

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



(5)



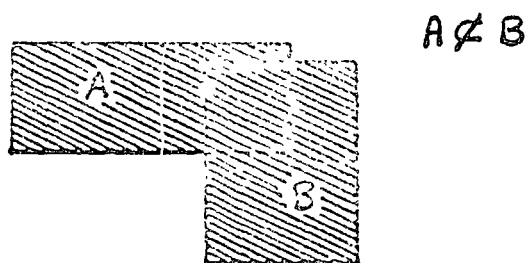
OTRO CONJUNTO FORMADO POR LOS ELEMENTOS QUE ESTAN EN A Y TAMBIEN EN B. SE DENOTA POR: $A \cap B$.

ASI:

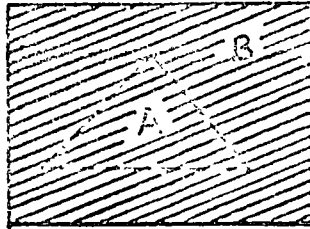
$$A \cap B = \{x \mid x \in A, x \in B\}$$

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

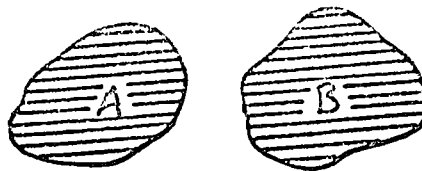
SEAN DOS CONJUNTOS A Y B. LA UNION ENTRE ESTOS CONJUNTOS ES



(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:

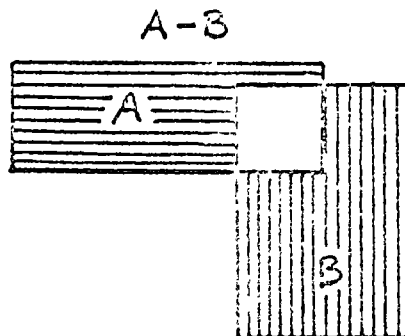
$$A \cup B = \{x \mid x \in A, x \in B, x \in (A \cap B)\}$$

o

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

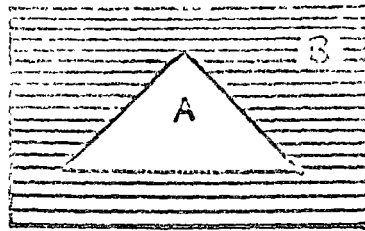
LA DIFERENCIA ENTRE LOS CONJUNTOS
A Y B ES

$A \not\subset B$



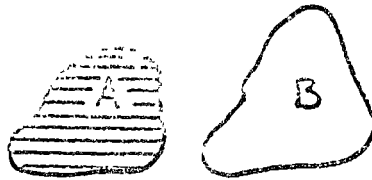
$B - A$

(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.

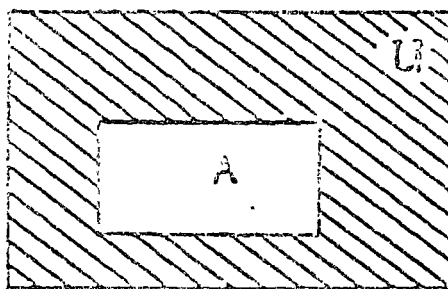
ASI:

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

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

EL COMPLEMENTO DE UN CONJUNTO A

ES



(3)

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' . ASI:

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

OBSERVESE QUE $A^c = U - A$

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 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



(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) $\{a\} \text{ --- } A$

1. b) $\{a, b, c, d, e, f\} \text{ --- } A$

1. c) $a \text{ --- } A$

1. d) $g \text{ --- } A$

1. e) $\{h\} \text{ --- } A$

1. f) $a \text{ --- } \emptyset$

1. g) $\emptyset \text{ --- } \{b\}$

1. h) $U = A^c \text{ --- } A$

1. i) $A^c = U \cap \text{ --- }$

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

(11)

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

2.a) $A \cap B =$

2.b) $A \cup B =$

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

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

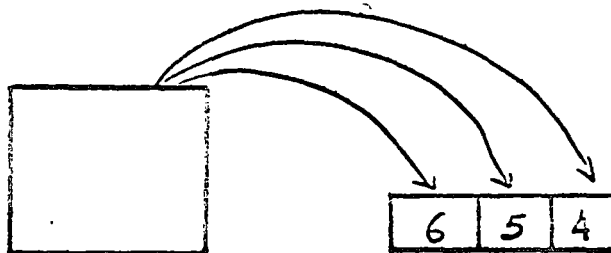
2.e) $B \cup A^c =$

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:

$$P(n, r) \text{ o } {}_n P_r$$

TEOREMA 3. EL NUMERO DE PERMUTACIONES
 DE n OBJETOS TOMAN-
 DO 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, RESPECTIVAMENTE.

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:

$$C(n, r) \text{ o } \binom{n}{r}$$

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

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 TENDRAN AL MENOS UNA MUJER? (c) ; CUANTAS TENDRAN 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.

I. 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 $A \cap B = \emptyset$, ENTONCES 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 AQUELLAS 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.

Faint, illegible text, possibly bleed-through from the reverse side of the page. The text is arranged in several paragraphs and is mostly illegible due to low contrast and blurring.

B. DEFINICION DE FRECUENCIA RELATIVA (1)

" CONSIDERERESE 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 OBSTANTE, 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 ARTERIORMENTE 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

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

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

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

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

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 $\omega_i \in S$ UN NUMERO REAL P_i LLAMADO PROBABILIDAD DE ω_i Y TAL QUE SATISFAGA LAS SIGUIENTES PROPIEDADES:

(1) 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) $P(\omega_1) = P(\omega_2) = \dots = P(\omega_n) = 1/n$

(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). ASÍ

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

DE ESTOS RESULTADOS, LOS
QUE SUMAN 6 SON

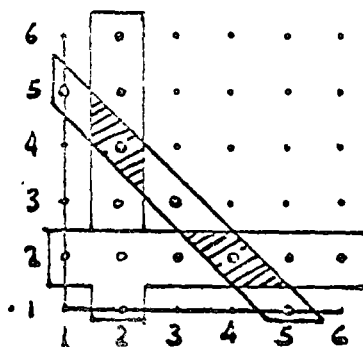
$$B = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$$

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 ADEMÁS TENGAN UN 2, ES DECIR

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

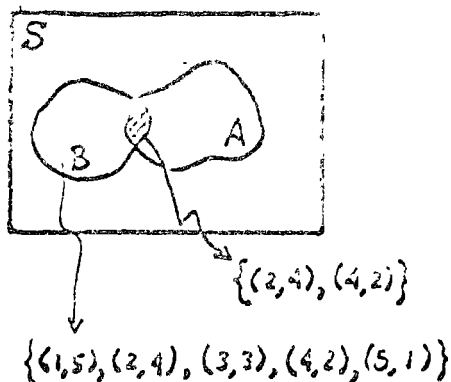


(29)

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

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

COROLARIO. I. PARA CUALESQUIERA EVENTOS

A_1, A_2, \dots, A_n , DE S ,

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

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, NA_2, NA_2 .

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, NA_2), HAY 6 POSIBILIDADES DE 10, DE SACAR UN TERCER RADIO NO DEFECTUOSO. POR TANTO

$$P(A_3 | A_1, NA_2) = \frac{6}{10}$$

FINALMENTE, DE ACUERDO AL TEOREMA DE MULTIPLICACION

$$\begin{aligned} P(A_1, NA_2, NA_3) &= P(A_1)P(A_2 | A_1)P(A_3 | A_1, NA_2) \\ &= \frac{14}{55} \end{aligned}$$

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.

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. (a) 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 pies^3/seg , Y LA QUE LLEGA DE B PUEDE SER 0, 20, 40, 60 pies^3/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) $C = A \geq 10$ pies^3/seg
- b) $D = B \geq 20$ "
- 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.

I. 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

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

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

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

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

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

SEGUN ESTO

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

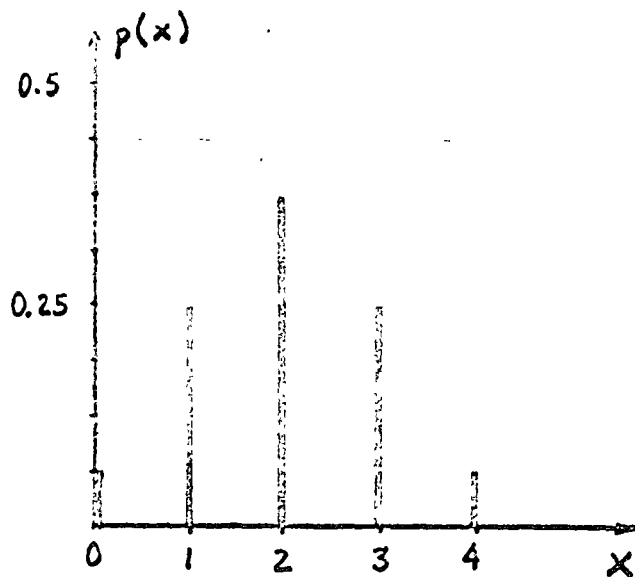
$$P(X=1) = \frac{1}{4}, \quad P(X=3) = \frac{1}{4}$$

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

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:

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

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

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

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

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

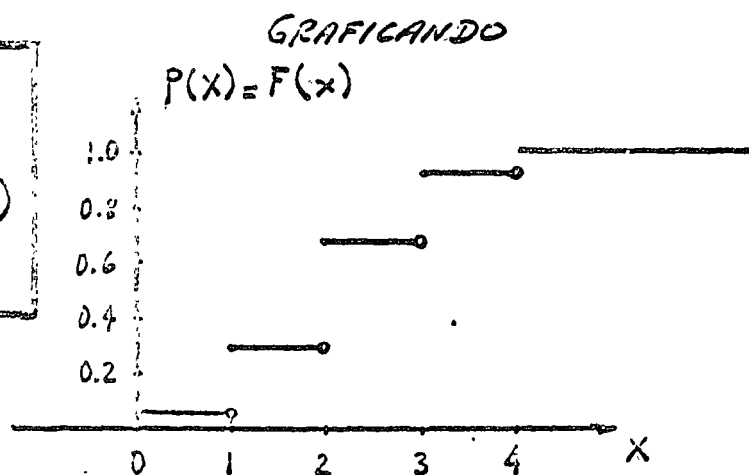
$$P(\{s | X(s) < 4\}) = P(X < 4) = \frac{15}{16}$$

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

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

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, COMUNNMENTE, 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

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

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

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

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

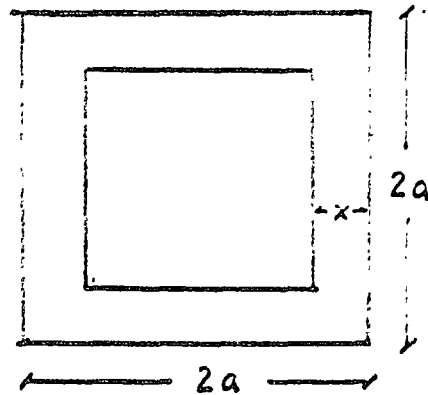
2. VARIABLES ALEATORIAS CONTINUAS

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 DIFE-
RENCIA 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 DENSI-
DAD $f(x)$. LA PROBABILIDAD DE QUE X
ESTE EN UN INTERVALO $[x, dx+x]$ 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-

(43)

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 TALEJ 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:

$$P[X=0, Y=0] = 0.25$$

$$P[X=1, Y=0] = 0.04$$

$$P[X=2, Y=0] = 0.01$$

$$P[X=3, Y=0] = 0$$

$$P[X=4, Y=0] = 0$$

$$P[X=0, Y=1] = 0$$

$$P[X=1, Y=1] = 0.36$$

$$P[X=2, Y=1] = 0.03$$

$$P[X=3, Y=1] = 0.01$$

$$P[X=4, Y=1] = 0$$

$$P[X=0, Y=2] = 0$$

$$P[X=1, Y=2] = 0$$

$$P[X=2, Y=2] = 0.16$$

$$P[X=3, Y=2] = 0.02$$

$$P[X=4, Y=2] = 0.0025$$

$$P[X=0, Y=3] = 0$$

$$P[X=1, Y=3] = 0$$

$$P[X=2, Y=3] = 0$$

$$P[X=3, Y=3] = 0.07$$

$$P[X=4, Y=3] = 0.015$$

$$P[X=0, Y=4] = 0$$

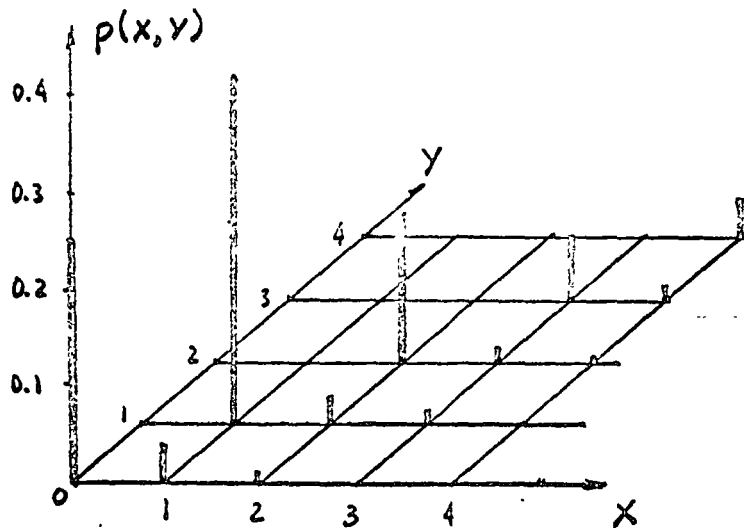
$$P[X=1, Y=4] = 0$$

$$P[X=2, Y=4] = 0$$

$$P[X=3, Y=4] = 0$$

$$P[X=4, Y=4] = 0.0325$$

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

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

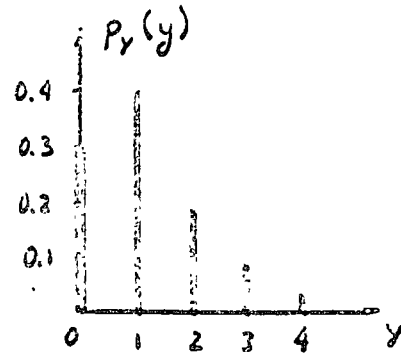
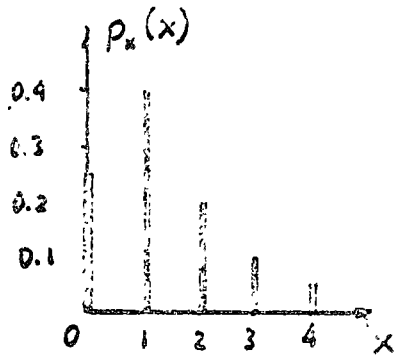
OTRAS PROBABILIDADES QUE PUEDEN SER DE INTERES SON:

* 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 :

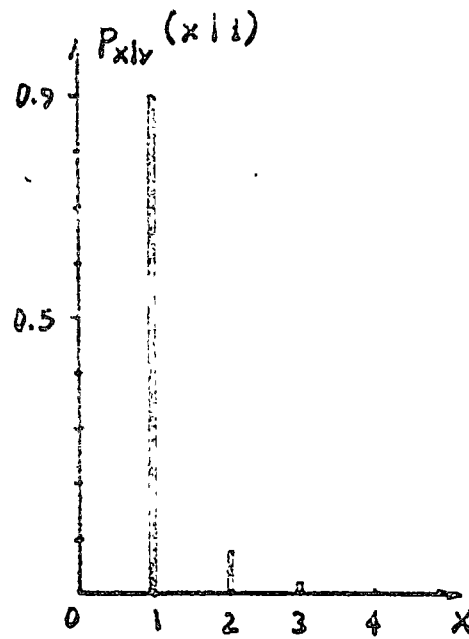


* SUPONGASE QUE EN EL CONTADOR SE LEE $Y=1$ EN UN INTERVALO EN PARTICULAR.

EL NUMERO REAL DE CARROS NO SE CONOCE 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

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

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

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

(47)

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

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

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

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,

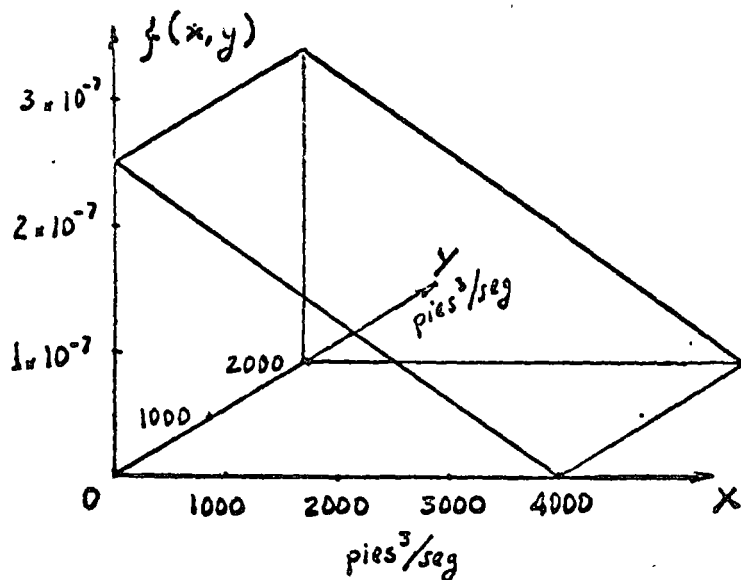
$$p(x,y) = P_{x|y}(x,y) p(y) = P_{y|x}(x,y) p(x)$$

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

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

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:

$$\begin{aligned} F(x, y) &\equiv P[(X \leq x) \cap (Y \leq y)] \\ &= P[(-\infty \leq X \leq x) \cap (-\infty \leq Y \leq y)] \end{aligned}$$

$$= \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv$$

A PARTIR DE ESTA ES POSIBLE OBTENER

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

LA PROBABILIDAD MARGINAL ES

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

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

Y SU FUNCION ACUMULATIVA

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

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

(5:)

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 \$10 000, 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

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

(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 $s \in \mathcal{S}$, EXISTE UN NUMERO REAL $X(s)$. Y, ESTA DEFINIDA PARA TODO $s \in \mathcal{S}$.
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

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

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

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 $y > 0$,

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

8. 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

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

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

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

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

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

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

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

EL PRIMER MOMENTO RESPECTO A LA MEDIA ES

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

EL SEGUNDO MOMENTO ES

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

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

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

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

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

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

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

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

= 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 outcome 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 problem, it is important that the reader appreciate that the data are not "raw" but are a result of some process of organization of the magnitude of variation in the data. The data are not "raw" but are a result of some process of organization of the magnitude of variation in the data.

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	91.4	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	80.0
K	94.5	24.1	87.5	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	18.4	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	65.9	53.1	39.7
P	93.2	55.0	80.8	143.5	122.3	181.2	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	151.1	131.3	81.6	191.4	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	121.6	78.9	116.0	100.3	97.8	75.3
V	24.8	55.6	135.6	56.2	66.9	72.2	107.4	98.9	101.7	58.2

* Observed live loads (in pounds per square foot), bay area = 400 ± 61 ft². See also J. W. Bond, G. N. Biddle, and C. N. Thompson (1962) "Live Loads in High-rise Buildings," *Journal of the American Institute of Steel Construction*, Report 111, New York, N.Y., p. 12.

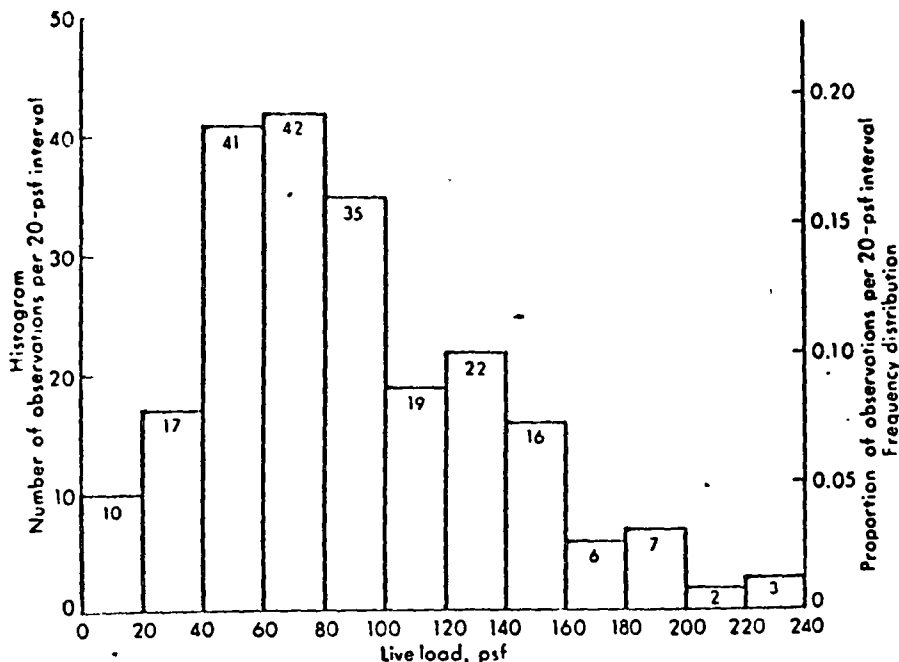


Fig. 1.1.1 Histogram and frequency distribution of floor-load data.

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

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).

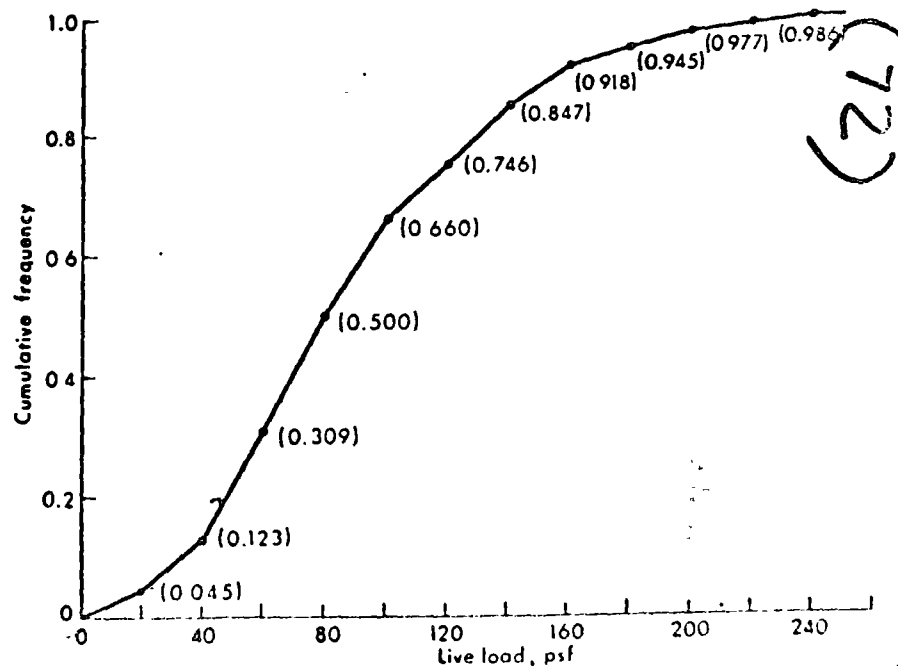


Fig. 1.1.2 Cumulative frequency distribution of floor-load data

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).

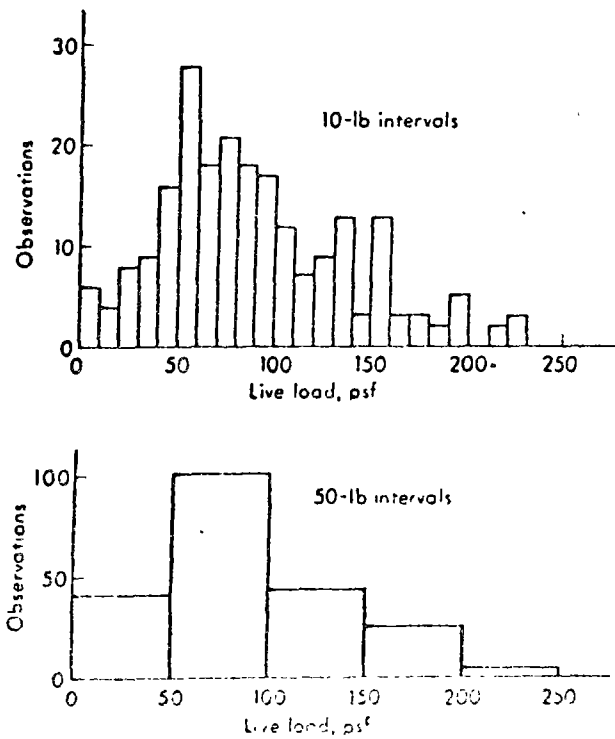


FIG. 1.1.3. Influence of interval size on appearance of histogram of live load (Table 1.1.1).

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

73

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	9,500
13	6,000	9,500
14	5,800	10,500
15	6,500	10,500

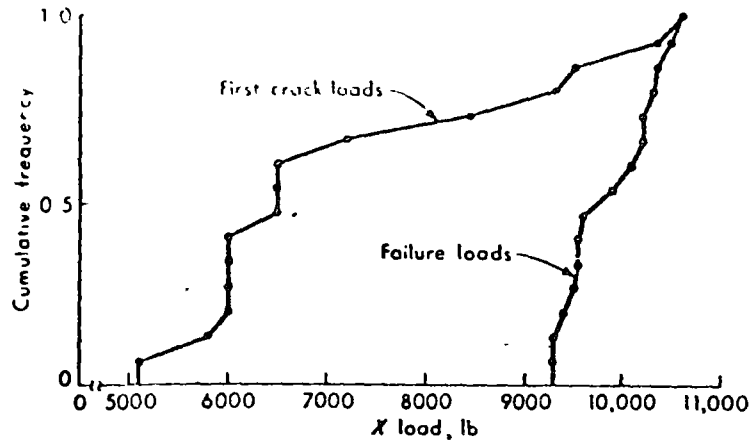


Fig. 1.2.1 Cumulative frequency distributions; beam data.

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 will do well to plot a number of different histograms of this failure load data to see the effect of the mean or data interpretation in this example. The reader will also do well to read about histograms 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.

$$\begin{aligned} s^2 &= \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \\ &= \frac{1}{15} (1,470,094.450 - 1,467,181,500) \\ &= 192,200 \text{ lb}^2 \\ s &= \sqrt{192,200} = 440 \text{ lb} \end{aligned}$$

(Notice that owing to the required subtraction of two nearly equal numbers, many digits must be carried in the sums to maintain accuracy. In this case, only 20 digits are needed to compute

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,

$$g_2 = \frac{4,150,000 \times 10^6}{11,000,000,000} = 1.64 < 3.0$$

asymmetry 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 \bar{x} or \bar{y} (or both) their average values of x are frequently paired with low y values, most of the terms will be

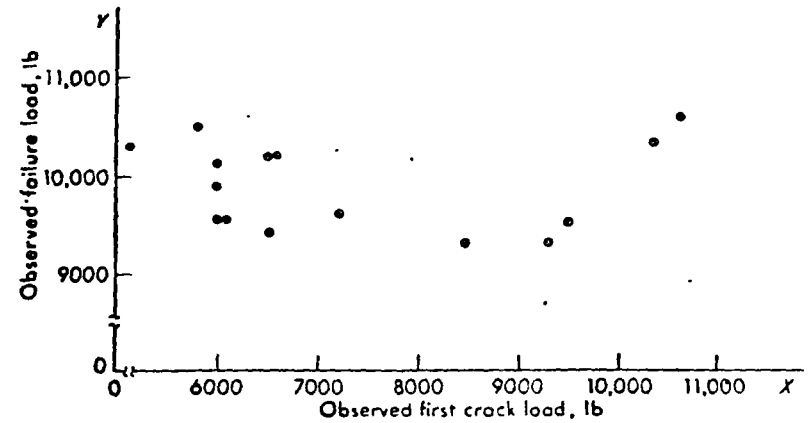


Fig. 1.3.1 Beam-data scattergram; plot shows lack of linear correlation

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

$$s_{X,Y} = \frac{1}{n} \sum_{i=1}^n (x_i - 7310)(y_i - 9890) = 91,200 \text{ lb}^2$$

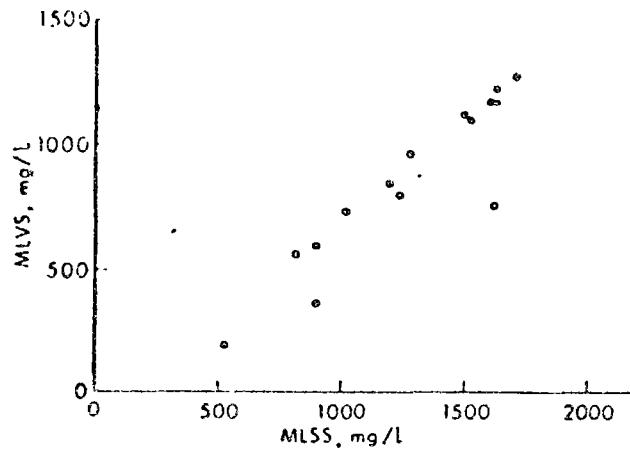


Fig. 1.3.2 Scattergram of MLVS against MLSS.

while the sample correlation coefficient is

$$r_{x,y} = \frac{s_{xy}}{s_x s_y} = \frac{94,200}{(1690)(410)} = +0.12$$

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. 1.3.2. 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. 4.7.

Discussion. In addition to providing efficient summaries of groups of numbers, the data reduction techniques described in this chapter apply a large part of the mathematical formulation and numerical concepts previously introduced in this study. The student should be able to solve 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.

Hann, 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. R. Kennedy [1964] "Basic Statistical Methods for Engineers and Scientists," International Textbook Company, Scranton, Pa.

Winer, B. L. [1964] "Statistics for Scientists and Engineers," Prentice-Hall, Englewood Cliffs, N. J.

Specific References

Stuart, R. A. [1962] "The Use of CUSUM Tests," *J. Roy. Stat. Soc. Ser. B*, Vol. 24, pp. 65-69.

PROBLEMS

1.1 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. 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 (391 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	0.131	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?

1.2 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	3100	4250	4500
7	2700	3500	4100
8	3100	3700	4250
9	3100	4100	4500
10	2400	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?

1.3 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	2.0461
1.5497	1.7496
1.7190	2.0461
1.6534	1.8187
1.7446	2.1328
1.4954	1.9250
1.6336	2.5022
1.4882	2.0830
2.0811	2.4280
1.9819	2.1375
2.1893	2.6819
2.0061	2.6815
	2.3011
	2.8131
	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?

1.4. 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

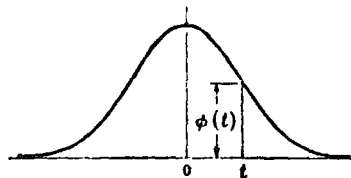
1.5. Compute the individual sample means, standard deviations, and coefficients of variation for the basement, fourth floor, and ninth floor of the floor-slabs. Do in

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.

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										
<p>This table gives areas under the standard normal distribution ϕ between 0 and $t \geq 0$ in steps of 0.01.</p>										
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	.4635
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	.4985	.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	.4993	.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

81

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 $p + q = 1$.

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

$$(ii) \quad b(3; 6, \frac{1}{2}) = \binom{6}{3} (\frac{1}{2})^3 (\frac{1}{2})^3 = \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3} (\frac{1}{2})^3 (\frac{1}{2})^3 = \frac{5}{18}.$$

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

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 $q = 1 - p = \frac{1}{3}$.

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

Ore-grade control in sub-horizontal deposits

519.272 622 013 34

Synopsis

A basic problem of quality control is the extraction of a maximum ore tonnage from a mine working under constraints of quality (imposed grades, for example). The direct solution to this problem may require an inextricable combination analysis of all possible total selections. Hence, it is of prime importance to design a practical and efficient selection procedure which avoids such an analysis, the working being divided into N mining units P_i of constant horizontal section, the best cut to select on each of these N units being deduced very simply from a *unique* cutoff grade. In case of indeterminacy, the cut which maximizes a quantity, interpreted as total benefit, is selected.

It is stressed that mining is by panels, never by drill-holes; hence, one must first proceed to the unbiased estimation (kriging) of the mean characteristics of these panels; in the best conditions kriging allows calculation of confidence limits. Cutoff performed on raw drill-hole data would irrevocably result in an overestimation of the recovered ore, which could be economically disastrous.

The theory of optimization is presented first and then its practical application to ore-grade control of the Togo phosphate deposit, West Africa.

Statement of the problem

Consider a sub-horizontal sedimentary deposit, containing a section W composed of a set of N panels P_i , the horizontal section s of each panel being constant (see Fig. 1). Note that the section W can comprise several separate sub-workings ($W_1 + W_2$ in Fig. 1). The

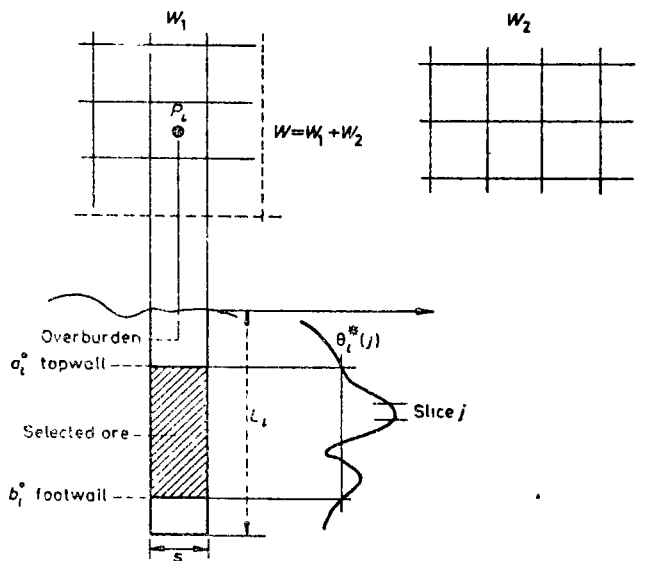


Fig. 1 Selective mining

mean mineralized thickness L_i is variable from one panel to another. Each panel P_i is subdivided into several slices of equal vertical thickness, which will be denoted $j = 1$ to L_i . Each of these slices has a mean characteristic $\theta_i^0(j)$, such as the grade, for example.

Section W is not rich enough for non-selective mining. Hence, the problem is to mine from W a maximum ore tonnage, the mean grade τ of which is guaranteed to be superior to a limiting grade τ_l .

On each of the N mining units P_i , an optimal topwall a_i^0 and footwall b_i^0 must be defined in such a way that

A. G. Journal Civil and Mining Engineer, Maitre de Recherches ENSMP
H. Sans Geologist ENSG
Both of the Centre de Morphologie Mathématique, Fontainebleau, France

the total recovered ore tonnage should be maximum and the mean grade τ of this tonnage should be equal or superior to the limiting grade τ_l , $\tau \geq \tau_l$.

In writing (a_i^0, b_i^0) the superscript designates the optimum cut on panel P_i , among all possible cuts (a_i, b_i) .

The problem does not consider the possibility of mining separate horizons in the same panel. The selected cut (a_i^0, b_i^0) is compact, and dirt layers within (a_i^0, b_i^0) are not distinguished.

In practice, the true values L_i , $\theta_i(j)$ and τ are unknown. They must be estimated without bias, and, preferentially, with minimum estimation variance (kriging). The particular kriging procedure applied to the Togo phosphate deposit is presented briefly later in this

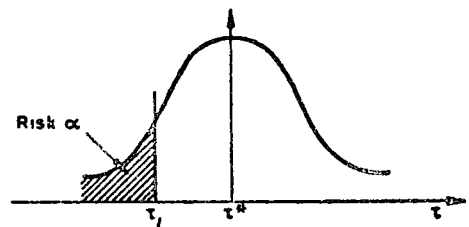


Fig. 2 Permitted risk

paper. As a certain error is made when the mean grade τ of the recovered ore is estimated by τ^* , the mean-grade constraint $\tau \geq \tau_l$ is written in practice

$$\tau^* \geq \tau_l + \beta \sigma_\tau$$

where $\sigma_\tau^2 = E\{[\tau - \tau^*]^2\}$ is the estimation variance of τ by τ^* . Kriging ensures the minimization of this variance. β is a parameter defined by the allowed risk α of recovering a mean grade τ inferior to the limiting grade τ_l . If a Gaussian law is classically adopted for the errors $\tau - \tau^*$, parameter β has the following values: $\beta = 2$ (for risk $\alpha = 2.5$ per cent); $\beta = 1.28$ (for risk $\alpha = 10$ per cent); and $\beta = 0$ (for risk $\alpha = 50$ per cent).

Technical optimum

The preceding optimum provides a set of N cuts $\{(a_i^0, b_i^0), i = 1$ to $N\}$ to be selected from section W . This optimum is independent of technical constraints; for example, the two neighbouring footwall elevations b_i^0, b_{i+1}^0 may be too different to allow haulage (see Fig. 3). Hence, a second technical optimum must be

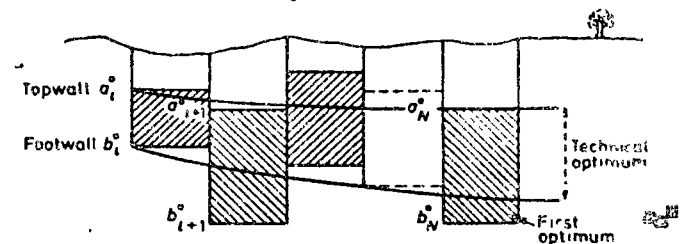


Fig. 3 Technical optimum

deduced from the first optimum, according to each particular set of technical constraints. Once this technical optimum is defined, kriging of its recovered ore provides the effectively recovered mean grade. As technical constraints can vary considerably from one mine to another, the theoretical exposé will be limited to the problem of defining the first optimum $\{(a_i^0, b_i^0), i = 1$ to $N\}$.

Theory of optimization

Notation

Consider panel P , let $c_i = b_i - a_i$ be the length of a possible cut (a_i, b_i) characterized by its topwall a_i and its footwall b_i . To simplify the nomenclature the product of *in situ* density and the section s of the panels will be taken as a constant equal to 1, the unit of length being the constant vertical thickness of the panel slices. Hence, we obtain for panel P ,

Recovered ore tonnage for the cut (a_i, b_i)

$$T_i = c_i$$

Recovered metal tonnage for the cut (a_i, b_i)

$$Q_i = \sum_{j=1}^{c_i} \theta_j(j)$$

$\theta_j(j)$ being the grade of slice j within the cut (a_i, b_i) . The mean grade of the recovered ore is

$$g_i = \frac{Q_i}{T_i}$$

One possible global selection is the set of the N possible cuts $\{(a_i, b_i), i = 1$ to $N\}$, one cut (a_i, b_i) being defined on each of the N panels P , comprising section W . For the total recovered ore

$$T = \sum_{i=1}^N T_i$$

Total recovered metal tonnage

$$Q = \sum_{i=1}^N Q_i$$

Mean grade

$$\tau = \frac{Q}{T}$$

Among the numerous possible global selections $\{(a_i, b_i), i = 1$ to $N\}$, which is the one $\{(a_i^0, b_i^0), i = 1$ to $N\}$ giving a maximum ore tonnage T , with a mean grade τ superior or equal to the grade limit τ_i ?

Combined analysis

Each panel P , is composed of a certain number L_i of slices of constant vertical thickness (see Fig. 1). Any possible cut (a_i, b_i) is composed of $c_i = b_i - a_i$ slices. Hence, on panel P , one has one possible cut of length 0, implying the panel is left, L_i possible cuts of length 1 slice, $L_i - 1$ possible cuts of length 2 slices and one possible cut of length L_i slices, or the whole panel mined. Thus, there are

$$\frac{L_i(L_i + 1)}{2} + 1$$

possible cuts on panel P_i .

As section W comprises N panels of respective total lengths $\{L_i, i = 1$ to $N\}$, the number of possible global selections is

$$A = \prod_{i=1}^N \left[\frac{L_i(L_i + 1)}{2} + 1 \right] = \left[\frac{L_1(L_1 + 1)}{2} + 1 \right] \left[\frac{L_2(L_2 + 1)}{2} + 1 \right] \dots \left[\frac{L_N(L_N + 1)}{2} + 1 \right]$$

With $N = 100$ panels and an average length of 10 slices there are $A = 56^{100}$ global possibilities—a number great enough to discourage any attempt to find the optimal possibility!

A can be reduced considerably by noting that, among the L_i possible cuts of length one slice on panel P_i , the cut that maximizes the metal tonnage Q_i is certainly

that which would be chosen. Let $\bar{Q}_i(1)$ be the maximum tonnage for a cut of length 1. In the same way, among the $L_i - 1$ possible cuts of length 2 slices, the cut that maximizes the metal tonnage Q_i is certainly that which would be retained. Let that maximum tonnage be denoted by $\bar{Q}_i(2)$, and so on, to $\bar{Q}_i(L_i)$. Hence the number of possible cuts of panel P_i is reduced from $\frac{L_i(L_i + 1)}{2} + 1$ to $(L_i + 1)$ 1 unique cut for each length c_i , varying from 0 to L_i slices. The number of possible global selections is then reduced to

$$\bar{A} = \prod_{i=1}^N [L_i + 1] = (L_1 + 1)(L_2 + 1) \dots (L_N + 1)$$

With $N = 100$ panels and the average length $L = 10$, there still remain 11^{100} possibilities to test!

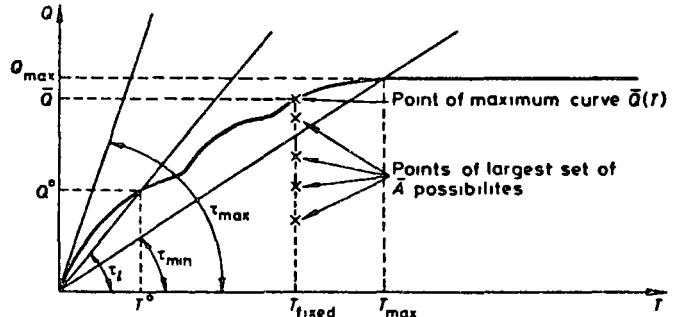


Fig. 4 Maximum curve

Let Q^k and T^k be the total metal tonnage and ore tonnage corresponding to the possibility number k among the \bar{A} . This set of \bar{A} possibilities $\{(Q^k, T^k), k = 1$ to $\bar{A}\}$ is represented in Fig. 4 by a cloud of \bar{A} points. Next, consider the subset conditioned by a constant total ore tonnage T , i.e. all the points of the cloud located on the fixed vertical line T (see Fig. 4). Obviously, if the optimum point is on that fixed line T , it must be the point that maximizes the total metal tonnage Q , let \bar{Q} be that maximum.

$$Q = \text{Max} \{Q^k / T^k = T_{\text{fixed}}\}$$

Hence, the optimum point is on the curve $\bar{Q}(T)$ which represents the upper limit of the cloud of \bar{A} possibilities. This curve is called the 'maximum curve'. Where exactly on the 'maximum curve' is the sought optimum point?

'Maximum curve' $\bar{Q}(T)$ (see Fig. 4)

The curve starts at the origin ($\bar{Q} = T = 0$), because rejecting all the N panels is a possibility corresponding to a limiting grade τ_i superior to the grade τ_{max} of the richest slice of all N panels.

The maximum curve never decreases, i.e.

$$\forall T' > T \Rightarrow \bar{Q}(T') \geq \bar{Q}(T)$$

In fact, if $\bar{Q}(T') < \bar{Q}(T)$, the maximum of $Q^k, T^k = T'$ being fixed, would have been $\bar{Q}(T)$ instead of $\bar{Q}(T')$, the increment of tonnage $T' - T$ corresponding to unmineralized ground with an average grade of zero.

It reaches its sill (Q_{max}) at the point ($Q_{\text{max}}, T_{\text{max}}$) corresponding to a limiting grade τ_{min} , which is nothing but the mean grade of the total ore *in situ*. For any limiting grade $\tau_i \leq \tau_{\text{min}}$, the maximum recovered metal tonnage is Q_{max} , total tonnage of metal *in situ*.

The limiting mean grade τ_i is shown by the line of slope τ_i passing through the origin O ($0, 0$) being a possible solution, there is at least one intersection of the line $Q = \tau_i T$ with the maximum curve $\bar{Q}(T)$. The optimum point (Q^0, T^0) which maximizes the recovered ore T^0 , while ensuring a mean grade

$$\tau^0 = \frac{Q^0}{T^0} \geq \tau_i$$

is the extreme right-hand intersection (see Fig. 4)

Hence the position of the optimum point on the maximum curve is known exactly. But, the problem of constructing the curve $\bar{Q}(T) = \text{Max}\{Q^*/T^k = T_{\text{fixed}}\}$ still remains.

Superior convex envelope of $\bar{Q}(T)$

Here lies the *key idea* of the optimization theory: some points of the maximum curve can be found directly without using the previous combined analysis, which consists of selecting the possibility (\bar{Q}, T) that maximizes the metal tonnage Q , from all the possibilities which give the same ore tonnage T . These particular points of the maximum curve $\bar{Q}(T)$ are those which are located on the superior convex curve of $\bar{Q}(T)$. Let them be denoted by $\bar{Q}(T)$ (see Fig 5). In simple terms $\bar{Q}(T)$ is composed of the convex parts of the maximum curve $\bar{Q}(T)$.

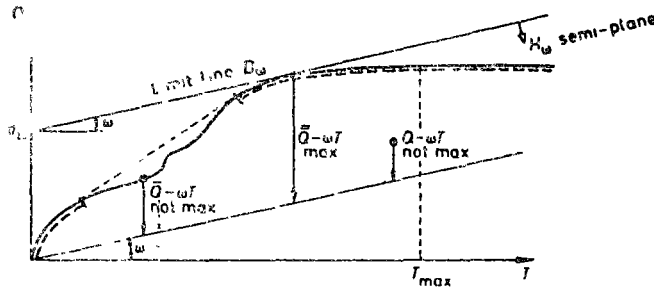


Fig. 5 $\bar{Q}(T)$ and its superior convex $Q(T)$

More precisely, consider all the lines of fixed slope ω . Their equations are $Q = \omega T + a$, parameter a being a variable. Now, consider the limit line D_ω of equation $Q = \omega T + a_\omega$, which is such that all the points (\bar{Q}, T) of the maximum curve are either below or on D_ω , i.e.

$$\forall \bar{Q}, T \in \text{Max curve} \Rightarrow \bar{Q} \leq \omega T + a_\omega \Rightarrow \bar{Q} - \omega T \leq a_\omega$$

(Note that the limit points, i.e. the contact points of D_ω with the maximum curve are such that $\bar{Q} - \omega T = a_\omega$, i.e. these limit points maximize the quantity $(\bar{Q} - \omega T)$.)

The limit line D_ω of slope ω defines a semi-plane

$$\{H_\omega, 0 \leq \omega T + a_\omega\}$$

which contains all the points of the maximum curve. Now, ω is made to vary from 0 to $+\infty$. The intersection of all the limit planes H_ω defines the superior convex envelope of the maximum curve $\bar{Q}(T)$, and the intersection of that envelope with the maximum curve $\bar{Q}(T)$ is the superior convex curve:

$$\bar{Q}(T) = \left\{ \bigcap_{\omega > 0} H_\omega \right\} \cap \{\bar{Q}, T\}$$

One characteristic of the superior convex curve $\bar{Q}(T)$ is that it contains all the points (\bar{Q}, T) of the maximum curve, and more generally all the points (Q, T) of the cloud of A global possibilities maximizing the quantity $(Q - \omega T)$, whatever the parameter $\omega \geq 0$ may be (see Fig 5).

Hence, to obtain directly a point of the convex (\bar{Q}, T) , without any tedious combined analysis, it is sufficient to fix the parameter $\omega \geq 0$ and to draw the possibility (Q, T) which maximizes the quantity $Q - \omega T$.

This can be drawn immediately, as the maximum of $Q - \omega T$ is given by the sum of the N local maxima $Q_i - \omega T_i$,

$$\text{Max}\{Q_i - \omega T_i\} = \sum_{i=1}^N \text{Max}\{Q_i - \omega T_i\}$$

ω being fixed, maximization on panel P_i of the quantity $Q_i - \omega T_i$ provides a unique cut (a_i^ω, b_i^ω) . The superscript ω merely indicates that the cut (a_i^ω, b_i^ω) corresponds to the fixed value ω of the parameter. The union of these N cuts $\{(a_i^\omega, b_i^\omega), i = 1 \text{ to } N\}$ provides a global point

$$(Q = \sum_i Q_i, T = \sum_i T_i)$$

which maximizes the quantity $Q - \omega T$. Hence the global point (Q, T) is on the convex curve $\bar{Q}(T)$, i.e. on the convex part of the maximum curve $\bar{Q}(T)$.

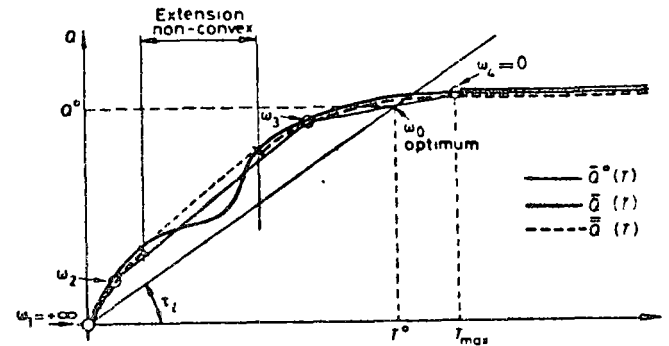


Fig. 6 $\bar{Q}(T)$ and its linear approximation $\bar{Q}^*(T)$

For each value ω , the corresponding point of the convex curve $\bar{Q}(T)$ can then be obtained. Linear interpolation between these points provides an approximation $Q^*(T)$ of the maximum curve $\bar{Q}(T)$. The unique intersection of the line of slope τ_i passing through the origin with the approximation $Q^*(T)$ provides the optimum value ω_0 of the parameter (see Fig 6). The required optimum point (Q^0, T^0) corresponds to the maximization of the quantity $Q - \omega T$, i.e. to the maximization of the N local quantities $\{Q_i - \omega_0 T_i\}, i = 1 \text{ to } N$.

This parametrical procedure with ω cannot by definition, provide any point (\bar{Q}, T) of the non-convex part of $\bar{Q}(T)$. This is not too serious, as, in practice, the extensions of these non-convex parts of $\bar{Q}(T)$ are negligible. In Figs 4, 5 and 6 the extensions of the non-convex parts have been greatly exaggerated to clarify the demonstration.

Definition of local cut (a_i^ω, b_i^ω)

The only problem left now is to determine on each panel P_i the cut (a_i^ω, b_i^ω) that maximizes the quantity $Q_i - \omega T_i$, whatever the parameter ω may be. (Note that $Q_i - \omega T_i = 0$, corresponding to rejection of panel P_i , this possibility is better than any cut that would provide a negative quantity $Q_i - \omega T_i$.)

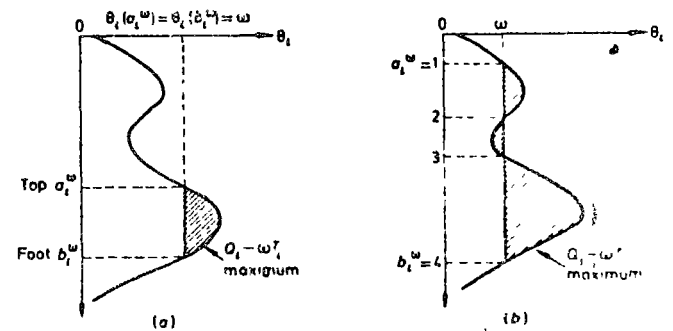


Fig. 7 Definition of local cut (a_i^ω, b_i^ω)

The parameter ω already appears as a grade $\frac{Q_i}{T_i} \geq \omega$. It will be shown that ω is precisely a cutoff grade. Consider the two limit grades $\theta_i(a_i^\omega)$ and $\theta_i(b_i^\omega)$ at the topwall and the footwall (see Fig 7(a)).

$\theta_i(a_i^\omega)$ cannot be less than ω —otherwise the limiting top slice would provide a negative contribution $\theta_i(a_i^\omega) - \omega < 0$ to the quantity $Q_i - \omega T_i$.

Similarly, if $\theta_i(a_i^\omega)$ were greater than ω the limiting slices of grades θ_i , included within $[\omega, \theta_i(a_i^\omega)]$, would have been retained for their positive contribution $\theta_i - \omega$ to the quantity $Q_i - \omega T_i$.

$\theta_i(b_i^\omega)$ must be equal to ω , which then works exactly as a cutoff grade

If the vertical curve of grades $\theta_i(j)$ is not convex, there may be several possible cuts, with limiting topwall and footwall grades, equal to ω for certain values of parameter ω (see Fig 7(b)). There are three possible cuts, (1-2), (3-4), and (1-4). The last cut (1-4) = (a_i^ω, b_i^ω) will be selected, as it maximizes the quantity $Q_i - \omega T_i$.

If the vertical grade function $\theta_i(j)$ could be differentiated (generally it has no reason to be so), a simple differentiation would have shown that ω is a cutoff grade $Q_i - \omega T_i$ is a function of the ordinate a_i of the topwall—hence, the maximum of $Q_i - \omega T_i$ corresponds to

$$dQ_i(a_i) - \omega dT_i(a_i) = 0$$

$$\left. \frac{d(Q_i - \omega T_i)}{da_i} \right|_{a_i^\omega} = \theta_i(a_i^\omega) = \omega$$

$$\left. \frac{d(Q_i - \omega T_i)}{db_i} \right|_{b_i^\omega} = \theta_i(b_i^\omega) = \omega$$

Summary

The problem is how to extract a maximum ore tonnage T with a mean grade constraint $\tau \geq \tau_i$ from a section W comprising N elementary panels

The desired optimum (Q°, T°) comprises the union of N particular local cuts $\{(a_i^\circ, b_i^\circ), i = 1 \text{ to } N\}$. N is usually too large to allow direct determination of the optimum by combined analysis

It is shown that the optimum point (Q°, T°) is certainly on the maximum curve $\bar{Q}(T)$ defined by

$$\bar{Q} = \text{Max} \{Q/T \text{ being fixed}\}$$

The intersection of the maximum curve $\bar{Q}(T)$ with the line $Q = \tau_i T$ provides the desired optimum point. The problem is now to draw the maximum curve $\bar{Q}(T)$ without any tedious combined analysis

It is shown that all the points of the convex parts of $\bar{Q}(T)$, i.e. all the points of the superior convex curve $\bar{Q}(T)$, maximize the global quantity

$$Q - \omega T = \sum_{i=1}^N [Q_i - \omega T_i]$$

Hence, to obtain a point on the convex curve $\bar{Q}(T)$ it is sufficient to fix a value of the parameter $\omega \geq 0$ and to determine on each panel the cut (a_i^ω, b_i^ω) which maximizes the local quantity $Q_i - \omega T_i$. This selection is obtained immediately by considering ω as a cutoff grade, unique to the N panels

For each value $\omega \geq 0$ a point of $\bar{Q}(T)$, i.e. a point of the maximum curve $\bar{Q}(T)$, is thus obtained. Linear interpolation between these points provides an estimation $\bar{Q}^*(T)$ of that maximum curve. The intersection of the approximation $\bar{Q}^*(T)$ with the constraint line $Q = \tau_i T$ provides the value ω_0 corresponding to the desired optimum point.

$$(Q^\circ, T^\circ) = \{(a_i^{\omega_0}, b_i^{\omega_0}), i = 1 \text{ to } N\}$$

Interpretation with benefit functions

The optimum point, provided by the preceding parametric procedure, maximizes the global quantity $B = Q - \omega_0 T$. This quantity can be interpreted as a global benefit B , which is proportional to the recovered metal tonnage Q , the extraction cost being proportional to the recovered ore tonnage $(-\omega_0 T)$. The value ω_0 of the parameter is determined by the mean grade constraint $\tau = \frac{Q}{T} \geq \tau_i$, more generally, ω_0 can be fixed by economic considerations, for example, ω_0 can be a marginal grade. Any unit of mean grade ω_0 would just pay for its mining and milling cost, thus providing a zero benefit

Generalization

The preceding benefit $B = Q - \omega_0 T$ corresponds to a linear profit formula—simplistic, but not unusual in practice

More generally, with a benefit function of the type $B(Q, T, T')$, T' being the overburden tonnage corresponding to the ore tonnage T , and if the optimum corresponds to the maximization of $B(Q, T, T')$, it can be shown that the optimum point $(Q^\circ, T^\circ, T'^\circ)$ is certainly on the maximum surface $\bar{Q}(T, T')$ defined by $\bar{Q} = \text{Max} \{Q/T \text{ and } T' \text{ being fixed}\}$. The superior convex envelope of the maximum surface can be obtained point by point with a multi-parametric (ω, v) procedure similar to that presented above, i.e. by maximization of the quantity $Q - \omega T - v T'$. Hence, an approximation $\bar{Q}^*(T, T')$ of the maximum surface $\bar{Q}(T, T')$ can be easily drawn without any combined analysis

The optimum point $(Q^\circ, T^\circ, T'^\circ)$ is at the contact of the approximation surface $\bar{Q}^*(T, T')$ with the iso-benefits surfaces $B(Q, T, T') = \text{Constant}$. The unique contact with the highest benefit provides the values (ω_0, v_0) corresponding to the required optimum

A more general description of the convex analysis of optimum selections has been given elsewhere³

Application to Togo phosphate deposit

Description of mine⁸

The phosphate formation of the sedimentary basin of Lower Togo, West Africa, is intercalated within clay strata. The average phosphate thickness is 10 m from top to bottom, oolitic phosphates, phosphated marls and, finally, gravelly phosphates can be distinguished

Open-pit mining is practised with bucket-wheels, bucket excavator and conveyor

The rather strict constraints of quality on the ore sold have repercussions on the mining system: the mined ore must contain a fixed tricalcium phosphate grade in the size 43–840 μm , which we will call 'pilot' size. To establish the annual exploitation programme the mine service must indicate, for each set of panels to be mined, the top- and footwalls that will guarantee the mean grade of tricalcium phosphate

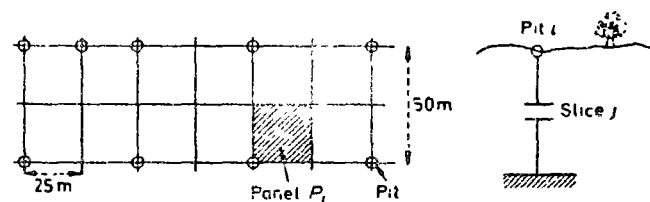


Fig 8 Available information

The available information consists of a regular square grid (50 m) of sampling pits, each 0.60-m diameter pit is sampled on every slice of 0.50-m vertical thickness (see Fig 8). The 'pilot' size of the ore is analysed. Hence, there are available from every pit i the vertical series of the two variables $p_i(j)$, the percentage of the size 43–840 μm and $k_i(j)$, the tricalcium phosphate grade in that size. Hence, the tricalcium phosphate grade of the raw ore mined from the slice j of pit i is

$$r_i(j) = \frac{p_i(j) \times k_i(j)}{100}$$

Presentation of optimization⁷

Consider a section W comprising N mining units P_i , each of constant horizontal area $s = 25 \times 25 \text{ m}^2$ (see Fig 8). Which are the optimum top- and footwalls (a_i°, b_i°) of each of the N panels that maximize the total recovered ore tonnage, the mean grade τ of tricalcium phosphate of the pilot size being superior to a limiting grade τ_i , with a defined risk α ?

This practical problem fits the exposed optimization

theory exactly in fact, on the volume ($0.50 \text{ m} \times$ section $25 \times 25 \text{ m}^2$) defined by a slice j of a panel P_i , let $\pi_i(j)$ be the mean percentage of the pilot size, $\lambda_i(j)$ be the mean tricalcium phosphate grade in that size and $\theta_i(j) = \frac{\pi_i(j) \times \lambda_i(j)}{100}$ the mean tricalcium phosphate grade of the *in situ* ore of slice j .

The addition of the N cuts $\{(a_i, b_i), i = 1 \text{ to } N\}$ defined on the N panels P_i provides the following. A total 'pilot' tonnage

$$T = \sum_{i=1}^N T_i = \sum_{i=1}^N \sum_{j \in (a_i, b_i)} \pi_i(j)$$

A total metal tonnage

$$Q = \sum_{i=1}^N Q_i = \sum_{i=1}^N \sum_{j \in (a_i, b_i)} \theta_i(j)$$

the mean grade of tricalcium phosphate in the pilot size

$$\bar{\tau} = \frac{Q}{T}$$

One has to maximize the total recovered pilot tonnage T under the constraint $\tau \geq \tau_i$, with the risk α , i.e. with a mean estimated grade $\tau^* \geq \tau_i = \tau_i + \beta \sigma_\tau$ (Fig 2)

Kriging of slices j

The two groups of variables $\pi_i(j), \lambda_i(j), \theta_i(j)$ and $\rho_i(j), r_i(j)$ are defined on two different volumes: the first group on panel slices of $0.5 \times 25 \times 25 = 312.5 \text{ m}^3$, the second group on pit slices of $0.5 \times 3.14 \times 0.30^2 = 0.14 \text{ m}^3$. The panel grade $\theta_i(j)$ is not the pit grade $r_i(j)$.

The unknown values $\pi_i(j)$ and $\theta_i(j)$ of each panel slice j must be estimated from the neighbouring pit data $\rho_i(j'), r_i(j')$. This problem is classically solved by kriging (see, in particular, Matheron, p. 115). First, the three-dimensional structures of the regionalized variables $\rho_i(j)$ and $r_i(j)$ are characterized with the help of variograms. Then, the unknown values $\pi_i(j)$ and $\theta_i(j)$ are kriged, the kriged estimators $\pi_i^*(j)$ and $\theta_i^*(j)$ being linear combinations of the pit data. Kriging in this way

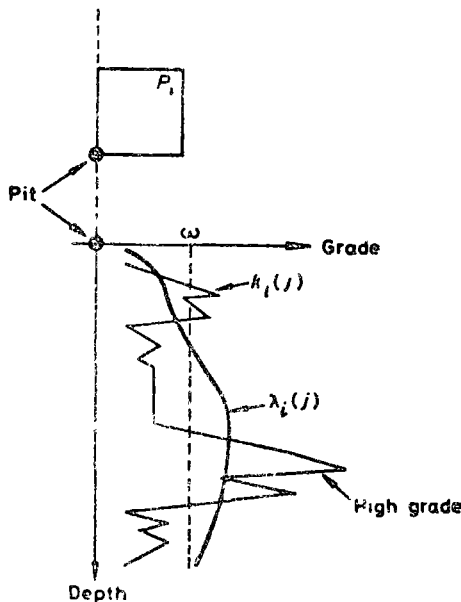


Fig. 9 Kriging

provides the estimation variances $E\{\pi_i - \pi_i^*\}^2$ and $E\{\theta_i - \theta_i^*\}^2$, which allow the calculation of the estimation variance of the mean grade τ of the recovered ore, that is

$$\sigma_\tau^2 = E\{\tau - \tau^*\}^2$$

This last variance is used directly in the mean grade constraint $\tau^* \geq \tau_i + \beta \sigma_\tau$. (It would be illusory to build an optimization procedure based on the pit values $\rho_i(j), k_i(j), r_i(j)$, instead of on the panel values $\pi_i(j), \lambda_i(j)$

$\theta_i(j)$, or at least on their best linear unbiased estimators $\pi_i^*(j), \lambda_i^*(j), \theta_i^*(j)$.)

The serrated aspect of the vertical series $k_i(j)$ is certainly not representative of the vertical fluctuations of the series $\lambda_i(j)$ or $\lambda_i^*(j)$ (see Fig 9). Any cutoff grade ω applied to the 'serrated' curve $k_i(j)$ would result in a falsely attractive precision of selection which, alas, will not be realized during subsequent production.

In particular, pit slices of 0.14 m^3 may sometimes present very high grades $k_i(j)$, whereas obviously, no panel slice of 312 m^3 will ever have such a high-grade $\lambda_i(j)$. Exploitation recovers panel slices not pit slices, so the high-grade assays found in pit slices will never be found to extend over panel-size slices.

Practical realization on Togo deposit

The geostatistical structural analysis has been performed with the help of the 50-m square grid (see Fig 8) and of two cross-shaped tighter grids of 5 and 10 m, which reveal the small-scale structures. Following this, a unique FORTRAN program gave the following: (1) the kriging ($\pi^*, \lambda^*, \theta^*$) of all slices j of each of the N panels P_i comprising the section W to be optimized, (2) the N local definitions of the optimum cut (a_i^0, b_i^0) to mine, and (3) the estimations with confidence limits of the various global mean characteristics of the recovered ore.

The limiting grade τ_i , as well as the accepted risk α are input parameters, which can vary according to market conditions.

The computer time was quite competitive with respect to a previous program with the classical procedure of 'cutoff on pit-grades and influence polygons'. Roughly, the cost of kriging and optimization of one $25 \times 25 \text{ m}$ panel with an average of 10 slices is 1 Franc or about 20 U.S. cents.

Verifications

The mining company suspected that the deposit could not provide high-grade mill feed in the pilot size range, containing more than 80 per cent, for example of tricalcium phosphate, unless they resorted to high-grading. In fact, the kriging and optimization procedures proved that to attain a high grade such that $\tau \geq 80$ per cent, 30–50 per cent of the ore *in situ* would have to be rejected. No phosphate company would accept a rejection of 30–50 per cent of its reserves, especially when the rejected ore contains an average pilot tricalcium phosphate grade of 76 per cent!

The first program to use the classical procedure with pit grades and influence polygons indicated that a maximum rejection rate of 10 per cent would provide the required mean grade $\tau \geq 80$ per cent. Of course, subsequent mining never achieved these optimistic forecasts.

On a particular section W where the effective top- and footwall profiles have been correctly determined, the actual mining results have been compared with the corresponding kriged forecasts. Whereas the first procedure, by influence polygons, systematically overvalued the grade of the recovered ore, the confidence limits of the kriged mean grade τ^* accounted for the realities of production: $\tau^* = 79.8 \pm 0.07$ per cent, $0.07 = \sigma_\tau$ being the simple standard deviation of estimation by kriging.

A systematic check is being done by sections, the mining of which will follow the top- and footwall profiles given by the new geostatistical optimization program.

Acknowledgement

The authors are indebted to Professor G. Matheron, Director of the Centre de Morphologie Mathématique at Fontainebleau, for his valuable theoretical advice throughout this optimization study. Mr Gabriel Directeur of the COTOMIB Company kindly afforded them the opportunity to apply their theoretical procedure to the real case of the Togo phosphate deposit.

References

- 1 Journel A Geostatistics and sequential exploration *Min Engng NY*, 25, Oct 1973 44-8
- 2 Journel A Le formalisme des relations ressources-réserves *Ind Minér Sér Mine* no 4 1973, 214-20
- 3 Journel A and Matheron G Analyse convexe et problèmes d'optima miniers Unpublished note, Centre de Morphologie Mathématique, Fontainebleau, 1973
- 4 Matheron G. and Formery Ph. Recherche d'optimum dans la reconnaissance et la mise en exploitation des gisements miniers *Annls Mines*, May 1963, 23-42, June 1963, 11-30
- 5 Matheron G. Le formalisme des coupures et des relations teneur-teneur Unpublished geostatistical note 119, Centre de Morphologie Mathématique Fontainebleau, 1971
- 6 Matheron G. The theory of regionalized variables and its applications *Cah Centre morphol Math, Fontainebleau* no 5, 1971, 211 p
- 7 Sans H Optimisation géostatistique de la production des phosphates du Togo Research Report, COTOMIB Co and Centre de Morphologie Mathématique, Fontainebleau, 1973
- 8 Robert M La mise en valeur des phosphates du Togo *Annls Mines*, March 1965, 11-34.