

CLASE

4

Teoría de la Información y la Comunicación

Entropía Avanzada y Redundancia

Educación de calidad



INTRO DUC CIÓN

DE LA CLASE

GRADO / POSGRADO / TECNOLOGÍAS

En esta semana se profundiza en el análisis cuantitativo de la información, abordando métricas que permiten comprender cómo se distribuye la incertidumbre en sistemas reales de datos. Mientras en semanas anteriores se introdujo la entropía como medida básica, ahora se amplía el enfoque hacia escenarios donde múltiples variables interactúan, lo cual es fundamental en entornos digitales complejos.

El estudio de la entropía conjunta y condicional permite analizar cómo se relacionan diferentes fuentes de información, identificando dependencias, patrones y estructuras ocultas. Estos conceptos son clave cuando se trabaja con datos reales, donde la información rara vez es completamente independiente. Asimismo, la información mutua se introduce como una herramienta para medir la cantidad de conocimiento compartido entre variables. Este concepto resulta esencial para detectar redundancias, correlaciones y posibles ineficiencias en la transmisión de datos. Finalmente, se aborda la redundancia desde una perspectiva aplicada, conectándola directamente con los límites teóricos de compresión.

4. Entropía Avanzada y Redundancia

En el análisis de sistemas de información, la entropía permite cuantificar el nivel de incertidumbre presente en una fuente de datos. Sin embargo, en escenarios reales, los datos no se presentan de manera aislada, sino que están interrelacionados, lo que hace necesario extender el análisis hacia modelos más complejos. En este contexto, la entropía avanzada proporciona herramientas para estudiar la interacción entre múltiples variables.

A partir de este enfoque, es posible comprender cómo se distribuye la información en sistemas donde existen dependencias, redundancias y patrones ocultos. Estas características son especialmente relevantes en áreas como la ciberseguridad, la compresión de datos y la transmisión de información, donde la eficiencia y la detección de anomalías dependen directamente de la estructura de los datos.

Además, el análisis de la redundancia permite identificar información repetitiva o predecible dentro de un conjunto de datos. Esta propiedad es fundamental para optimizar procesos de compresión, ya que permite eliminar información innecesaria sin afectar el contenido esencial.

El estudio de la entropía avanzada y la redundancia constituye la base para evaluar la eficiencia de los datos y establecer los límites teóricos de compresión.

En contextos modernos como el análisis de grandes volúmenes de datos (Big Data), estas métricas adquieren aún mayor relevancia, ya que permiten identificar estructuras ocultas dentro de conjuntos masivos de información. En particular, la combinación de entropía, dependencia y redundancia facilita la construcción de modelos predictivos más precisos.

Asimismo, en sistemas distribuidos y redes, comprender cómo fluye la información entre diferentes componentes permite optimizar el uso de recursos, reducir la latencia y mejorar la seguridad. La entropía avanzada no solo describe la incertidumbre, sino que también permite inferir el comportamiento dinámico de sistemas complejos.

Por tanto, su estudio no es únicamente teórico, sino altamente aplicable en entornos tecnológicos actuales.

4.1 Entropía Conjunta

La entropía conjunta es una extensión del concepto de entropía de una sola variable hacia sistemas donde intervienen múltiples fuentes de información. Permite cuantificar la incertidumbre total cuando se consideran simultáneamente dos variables aleatorias discretas. Formalmente, si X y Y son variables aleatorias discretas definidas sobre un mismo espacio de probabilidad, la entropía conjunta se expresa como:

$$H(X, Y) = - \sum_{x,y} p(x, y) \log_2 p(x, y)$$

donde $p(x, y)$ representa la probabilidad conjunta, es decir, la probabilidad de que ocurra simultáneamente el evento $X = x$ y $Y = y$.

Esta medida cuantifica la incertidumbre total asociada al sistema combinado (Alencar, 2015).

La entropía conjunta mide la incertidumbre total del sistema considerando la interacción entre variables, no solo su comportamiento individual.

1. Caso de independencia

Si X y Y son independientes:

$$p(x, y) = p(x)p(y)$$

Entonces:

$$H(X, Y) = H(X) + H(Y)$$

Cuando no hay relación entre variables, la incertidumbre total es simplemente la suma de ambas y es mayor.

2. Caso de dependencia

Si existe dependencia:

$$p(x, y) \neq p(x)p(y)$$

Entonces:

$$H(X, Y) < H(X) + H(Y)$$

La dependencia reduce la incertidumbre total porque una variable aporta información sobre la otra, en consecuencia, es menor.

Tabla 1

Comparación entre independencia y dependencia

Característica	Independencia	Dependencia
Relación entre variables	No existe	Existe
Probabilidad conjunta	$p(x, y) = p(x)p(y)$	$p(x, y) \neq p(x)p(y)$
Entropía conjunta	$H(X, Y) = H(X) + H(Y)$	$H(X, Y) < H(X) + H(Y)$
Incertidumbre	Máxima	Reducida

Nota. La tabla compara las propiedades de variables aleatorias independientes y dependientes en términos de probabilidad conjunta, entropía e incertidumbre. Adaptado de *Information theory: Concepts and applications*, por M. S. Alencar, 2015, Springer. (Miguel Ortiz Navarrete)

3. Interpretación en términos de información

La entropía conjunta puede entenderse como:

- Toda la información necesaria para describir el sistema completo
- Incluye:
 - Información individual de cada variable
 - Información compartida entre ellas

Esto conecta directamente con la información mutua (tema siguiente).

Desde una perspectiva práctica, la entropía conjunta permite modelar sistemas donde múltiples variables interactúan simultáneamente, como ocurre en bases de datos relacionales, sistemas de monitoreo o redes de sensores. En estos casos, analizar cada variable por separado puede llevar a conclusiones incompletas, mientras que la entropía conjunta captura la complejidad total del sistema. Además, esta métrica es clave en el diseño de algoritmos de compresión multivariable, donde la eficiencia depende de identificar dependencias entre fuentes de información.

Ejemplo aplicado

Escenario: Detección de ataques en un Centro de Operaciones de Seguridad SOC (Security Operations Center)

Se analizan dos variables:

- X : Tipo de evento

- NORMAL
- ALERTA
- Y: Origen del tráfico
 - INTERNO
 - EXTERNO

La distribución conjunta observada se evidencia en la tabla 1

Tabla 2

Distribución conjunta de eventos de seguridad según tipo de evento y origen del tráfico

X (Evento)	Y (Origen)	Probabilidad
NORMAL	INTERNO	0.50
NORMAL	EXTERNO	0.10
ALERTA	INTERNO	0.05
ALERTA	EXTERNO	0.35

Nota. Distribución conjunta utilizada para ilustrar el cálculo de la entropía conjunta en un contexto de análisis de eventos de ciberseguridad. Elaboración propia (Miguel Ortiz Navarrete).

Cálculo de la entropía conjunta

$$H(X, Y) = -\sum p(x, y) \log_2 p(x, y)$$

$$H(X, Y) = -[0.5 \log_2(0.5) + 0.1 \log_2(0.1) + 0.05 \log_2(0.05) + 0.35 \log_2(0.35)]$$

Resultado aproximado:

$$H(X, Y) \approx 1.68 \text{ bits}$$

Ahora calculemos las probabilidades marginales:

Para X(tipo de evento)

Sumamos por filas:

- $P(NORMAL) = 0.50 + 0.10 = 0.60$
- $P(ALERTA) = 0.05 + 0.35 = 0.40$

Para Y(origen)

Sumamos por columnas:

- $P(INTERNO) = 0.50 + 0.05 = 0.55$
- $P(EXTERNO) = 0.10 + 0.35 = 0.45$

Calculamos la entropía de X

$$H(X) = -\sum p(x) \log_2 p(x)$$

$$H(X) = -(0.6 \log_2 0.6 + 0.4 \log_2 0.4)$$

Cálculo:

- $\log_2(0.6) \approx -0.737$
- $\log_2(0.4) \approx -1.322$

$$H(X) = -(0.6(-0.737) + 0.4(-1.322))$$

$$H(X) = 0.442 + 0.529 \approx 0.971 \approx 0.99 \text{ bits}$$

Calculamos la entropía de Y

$$H(Y) = -(0.55 \log_2 0.55 + 0.45 \log_2 0.45)$$

Cálculo:

- $\log_2(0.55) \approx -0.862$

- $\log_2(0.45) \approx -1.152$

$$H(Y) = -(0.55(-0.862) + 0.45(-1.152))$$

$$H(Y) = 0.474 + 0.518 \approx 0.992 \approx 0.97 \text{ bits}$$

Si fueran independientes:

$$H(X, Y) = H(X) + H(Y)$$

- Entonces:

$$0.99 + 0.97 \approx 1.96$$

Pero vimos que:

$$H(X, Y) \approx 1.68$$

La diferencia entre 1.96 y 1.68 indica que existe dependencia entre las variables.

Es decir:

- Conocer el origen (IP) ayuda a predecir el tipo de evento
- Esto reduce la incertidumbre total

Enlaces complementarios

Título: Joint Entropy, Conditional Entropy, and Mutual Information

Descripción: Recurso académico que presenta de forma directa las definiciones matemáticas de entropía conjunta, entropía condicional e información mutua, incluyendo fórmulas y propiedades como independencia y regla de la cadena. Ideal para reforzar exactamente los contenidos del documento.

Enlace: https://classes.up-microlab.org/index.php?title=Joint_entropy%2C_conditional_entropy%2C_and_mutual_information

4.2 Entropía Condicional

La entropía condicional es una extensión del concepto de entropía que permite analizar sistemas donde existe dependencia entre variables. Mientras la entropía simple mide la incertidumbre total de una variable, la entropía condicional cuantifica cuánta incertidumbre permanece cuando ya se dispone de información parcial.

$$H(Y | X) = - \sum_{x,y} p(x,y) \log_2 p(y | x)$$

donde:

- $p(x, y)$: probabilidad conjunta
- $p(y | x)$: probabilidad condicional
- $H(Y | X)$: incertidumbre de Y dado X

Según Csiszár y Körner (2011), esta métrica es clave para entender la dependencia entre variables.

La entropía condicional mide cuánto “desconocemos todavía” después de haber observado parte del sistema.

Relación con la Regla de la Cadena

Una propiedad fundamental es:

$$H(X, Y) = H(X) + H(Y | X)$$

Esta ecuación indica que:

- Primero medimos la incertidumbre de X
- Luego añadimos la incertidumbre restante de Y , sabiendo X

La incertidumbre total se construye de forma secuencial: primero una variable, luego la incertidumbre adicional de la otra.

Esta propiedad es especialmente útil en modelos probabilísticos como las cadenas de Markov y las redes bayesianas, donde el conocimiento previo reduce progresivamente la incertidumbre. En estos modelos, la entropía condicional permite evaluar cuánto aporta cada nueva variable al conocimiento del sistema. De esta manera, se puede identificar qué variables son más relevantes para la predicción y cuáles aportan poca información adicional.

Ejemplo

Análisis de comportamiento en un sistema corporativo

Definimos:

- X : tipo de usuario
 - Administrador
 - Usuario normal
- Y : tipo de acción
 - Acceso a base de datos crítica
 - Navegación estándar

Caso 1: Alta dependencia (comportamiento estructurado)

Supongamos:

- Administradores → acceden casi siempre a bases de datos
- Usuarios normales → rara vez lo hacen

Entonces:

Conocer X reduce significativamente la incertidumbre sobre Y

Resultado:

$$H(Y | X) \text{ es bajo}$$

Esto indica que el comportamiento es predecible y estructurado.

Caso 2: Independencia (comportamiento aleatorio)

Si:

- Tanto administradores como usuarios normales realizan cualquier acción con igual probabilidad

Entonces:

$$H(Y | X) = H(Y)$$

El sistema no tiene patrones claros → mayor incertidumbre.

Caso 3: Dependencia total (escenario crítico)

Supongamos:

- Solo administradores acceden a bases críticas
- Usuarios normales nunca lo hacen

Entonces:

$$H(Y | X) = 0$$

Si observas una acción, puedes inferir el tipo de usuario. Esto puede ser útil para detección de anomalías.

La entropía condicional permite entender cuánto valor aporta conocer una variable para predecir otra.

Esto es fundamental porque:

- Los sistemas reales no son independientes
- La información útil está en las relaciones entre datos

4.3 Información Mutua

La información mutua es una de las métricas más importantes en la teoría de la información, ya que permite cuantificar el grado de dependencia entre dos variables aleatorias. A diferencia de la entropía, que mide incertidumbre, la información mutua mide cuánta incertidumbre se reduce al conocer otra variable. La información mutua mide cuánto conocimiento se comparte entre dos variables. (Gorzalka & Deloumeaux, 2012)

Formalmente se define como:

$$I(X; Y) = H(X) + H(Y) - H(X, Y)$$

También puede expresarse como:

$$I(X; Y) = \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}$$

La información mutua mide cuánta información comparten dos variables, es decir, cuánto saber una ayuda a predecir la otra.

Esto implica:

- Si conocer X no cambia lo que sabemos sobre $Y \rightarrow$ no hay información compartida
- Si conocer X reduce la incertidumbre de $Y \rightarrow$ existe dependencia
- Si conocer X determina completamente $Y \rightarrow$ máxima dependencia

Interpretación desde la entropía

La ecuación:

$$I(X; Y) = H(X) - H(X | Y)$$

también es válida y revela una idea clave:

La información mutua es la reducción de incertidumbre de X cuando se conoce Y

De forma equivalente:

$$I(X; Y) = H(Y) - H(Y | X)$$

La información mutua representa la parte de la entropía que deja de ser incertidumbre y se convierte en conocimiento.

Propiedades importantes

No negatividad

$$I(X; Y) \geq 0$$

Nunca puede ser negativa.

- $I(X; Y) = 0 \rightarrow$ independencia
- $I(X; Y) > 0 \rightarrow$ existe relación

Simetría

$$I(X; Y) = I(Y; X)$$

La relación es bidireccional.

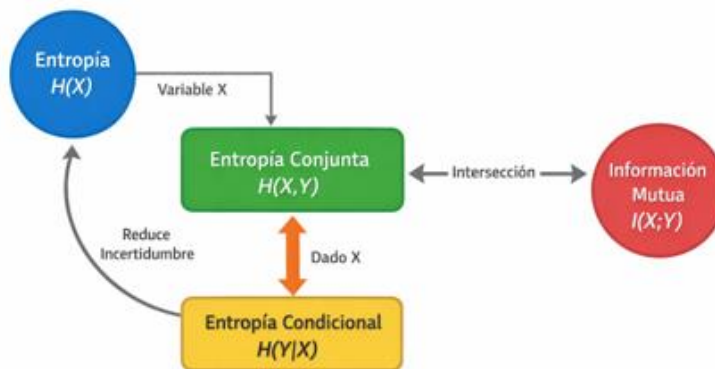
Relación con entropía conjunta

$$H(X, Y) = H(X) + H(Y) - I(X; Y)$$

La información mutua representa la “parte compartida” de la información entre variables.

Figura 1

Relación entre tipos de entropía



Nota: Elaborado de Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley. (Miguel Ortiz Navarrete)

Ejemplo

Sistema SOC (Security Operations Center)

Definimos:

- X : tipo de evento
 - LOGIN_OK
 - LOGIN_FAIL
- Y : dirección IP
 - IP interna
 - IP externa

Caso 1: Independencia (comportamiento normal)

- Los eventos ocurren aleatoriamente en distintas IPs
- No hay patrón

Resultado:

$$I(X; Y) = 0$$

No existe relación entre el tipo de evento y la IP.

Caso 2: Dependencia moderada

- LOGIN_FAIL ocurre más frecuentemente en IP externas
- LOGIN_OK ocurre más en IP internas

Resultado:

$$I(X; Y) > 0$$

Existe correlación parcial que puede indicar comportamiento sospechoso.

Caso 3: Alta dependencia (ataque dirigido)

- Todos los LOGIN_FAIL provienen de una sola IP
- Esa IP nunca genera LOGIN_OK

Resultado:

$$I(X; Y) \gg 0$$

Alta información mutua indica un patrón fuerte → posible ataque.

Las aplicaciones en ciberseguridad son variadas

1. Detección de anomalías

- Identifica relaciones inesperadas. Ejemplo: IP específica asociada a eventos críticos

2. Sistemas IDS/IPS

- Detectan patrones de tráfico
- Identifican correlaciones ocultas

3. Análisis de logs

- Relación entre:
 - Usuario $\leftrightarrow$ acciones
 - IP $\leftrightarrow$ eventos
 - tiempo $\leftrightarrow$ actividad

4. Machine Learning en seguridad

- Selección de características (feature selection)
- Identificación de variables relevantes

La información mutua permite detectar patrones ocultos que no son visibles analizando variables de forma individual.

Esto es fundamental porque:

- Los ataques no siempre son evidentes
- Se manifiestan como relaciones entre variables

En el ámbito del aprendizaje automático, la información mutua se utiliza ampliamente para la selección de características, permitiendo reducir la dimensionalidad de los datos sin perder información relevante. Esto mejora el rendimiento de los modelos y reduce el costo computacional. Además, es una herramienta clave en el análisis de dependencia no lineal, donde otras métricas tradicionales pueden fallar.

Tabla 3

Relación entre entropía conjunta, condicional e información mutua

Concepto	Fórmula	Interpretación
Entropía conjunta	$H(X,Y)$	Incertidumbre total
Entropía condicional	$H(Y X)$	
Información mutua	$I(X;Y)$	Información compartida

Nota. La tabla resume las principales medidas de la teoría de la información utilizadas para analizar la incertidumbre y la dependencia entre variables aleatorias. Adaptado de *Information theory: Coding theorems for discrete memoryless systems*, por I. Csiszár y J. Körner, 2011, Cambridge University Press.

Enlaces complementarios

Título: Conditional Entropy and Mutual Information

Descripción: Explicación técnica de cómo la entropía condicional y la información mutua permiten medir dependencia entre variables, incluyendo interpretación de independencia y análisis de información compartida.

Enlace: <https://www.classes.up-microlab.org/index.php?title=Conditional Entropy and Mutual Information>

4.4 Redundancia

La redundancia es un concepto fundamental en la teoría de la información que permite identificar qué parte de los datos no aporta información nueva. En términos generales, representa el grado en que un sistema contiene patrones repetitivos o predecibles. (Csiszár & Körner, 2011)

Formalmente, se define como:

$$R = 1 - \frac{H(X)}{H_{max}}$$

donde:

- $H(X)$: entropía real del sistema
- $H_{max} = \log_2(n)$: entropía máxima posible (cuando todos los eventos son equiprobables)
- n : número de símbolos posibles

Interpretación conceptual

La redundancia mide qué proporción de la información es repetitiva y, por tanto, potencialmente eliminable sin pérdida significativa.

Esto implica:

- Si la entropía es baja $\rightarrow$ alta redundancia
- Si la entropía es alta $\rightarrow$ baja redundancia

En otras palabras:

- Redundancia = ineficiencia informativa
- Entropía = contenido útil de información

Esto implica:

- Alta redundancia $\rightarrow$ mayor compresión posible
- Baja redundancia $\rightarrow$ datos eficientes

Entropía real

$$H(X)$$

- Es la entropía calculada con las probabilidades reales
- Refleja cómo se comporta el sistema en la práctica

La redundancia se define como:

$$R = \frac{H_{max} - H(X)}{H_{max}}$$

Esto se puede interpretar en dos partes:

Numerador: $H_{max} - H(X)$

Representa la cantidad de información que “falta” para llegar al máximo nivel de incertidumbre
Es decir:

- Cuánto le falta al sistema para ser completamente aleatorio
- Cuánto está “ordenado” o “estructurado”

Este término mide la cantidad de información repetitiva o predecible.

Denominador: H_{max}

Normaliza el valor para que:

$$0 \leq R \leq 1$$

- Permite comparar sistemas diferentes
- Expresa la redundancia como proporción

La redundancia mide qué tan lejos está un sistema de ser completamente aleatorio.

Puedes entenderlo como una distancia:

- $H_{max} \rightarrow$ punto ideal (máxima incertidumbre)
- $H(X) \rightarrow$ estado real

Entonces:

$$\text{Distancia} = H_{max} - H(X)$$

Y al dividirlo:

$$R = \text{distancia relativa}$$

La redundancia es una medida relativa de “orden” dentro del sistema.

En sistemas reales, la redundancia no siempre es negativa. En algunos casos, se introduce de forma intencional para mejorar la confiabilidad, como en sistemas de almacenamiento redundante o protocolos de comunicación con corrección de errores. Sin embargo, desde el punto de vista de la eficiencia, reducir la redundancia innecesaria es clave para optimizar el almacenamiento y la transmisión de datos.

Figura 2

Relación inversa entre entropía y redundancia en función de la distribución de probabilidad



Nota. La figura muestra la relación inversa entre la entropía y la redundancia: a medida que la distribución de probabilidad se vuelve más uniforme (aleatoria), la entropía aumenta y la redundancia disminuye; por el contrario, cuando la distribución es más sesgada (ordenada), la entropía disminuye y la redundancia aumenta. Adaptado de *A Mathematical Theory of Communication*, por C. E. Shannon, 1948, *Bell System Technical Journal*, 27(3), 379–423.

Ejemplo

Supongamos en un tráfico de red de 4 eventos (tráfico normal, errores, intentos fallidos, accesos sospechosos) un sistema con 4 eventos:

$$n = 4 \Rightarrow H_{max} = \log_2(4) = 2 \text{ bits}$$

Caso 1: Distribución uniforme

$$p = (0.25, 0.25, 0.25, 0.25)$$

No hay eventos dominantes, todos con el 25 %

La entropía de una variable discreta es:

$$H(X) = -\sum p(x_i) \log_2 p(x_i)$$

Entonces:

$$H(X) = -[0.25 \log_2(0.25) + 0.25 \log_2(0.25) + 0.25 \log_2(0.25) + 0.25 \log_2(0.25)]$$

$$H(X) = 2$$

$$R = \frac{2 - 2}{2} = 0$$

No hay redundancia → sistema completamente aleatorio

Caso 2: Distribución sesgada

$$p = (0.7, 0.1, 0.1, 0.1)$$

Hay un evento dominante (70%)

El resto son raros

$$H(X) \approx 1.36$$

$$R = \frac{2 - 1.36}{2} = 0.32$$

Existe redundancia del 32% → el sistema tiene patrones

Caso 3: Máxima predictibilidad

$$p = (1, 0, 0, 0)$$

Hay un solo evento en ocurrencia (100%)

$$H(X) = 0$$

$$R = \frac{2 - 0}{2} = 1$$

Redundancia total → sistema completamente predecible

Cuando las probabilidades no son uniformes, algunos eventos dominan el sistema, reduciendo la entropía y generando redundancia, lo cual permite identificar patrones en los datos.

Tabla 4

Nivel de redundancia según distribución

Distribución	Entropía	Redundancia	Interpretación
Uniforme	Alta	0	Sistema aleatorio
Sesgada	Media	Parcial	Existen patrones
Determinística	0	1	Totalmente predecible

Nota. La tabla muestra la relación entre la distribución de probabilidad, la entropía y el nivel de redundancia en un sistema de información. A medida que la distribución se vuelve más uniforme, la entropía aumenta y la redundancia disminuye; en distribuciones determinísticas ocurre lo contrario. Adaptado de *A mathematical theory of communication*, por C. E. Shannon, 1948, *Bell System Technical Journal*, 27(3), 379–423.

Tabla 5
Relación operativa entre métricas de teoría de la información

Métrica	Qué mide	Depende de otras variables	Aplicación práctica	Interpretación clave
Entropía ($H(X)$)	Incertidumbre de una variable	No	Evaluar aleatoriedad de una fuente	Mayor valor = más incertidumbre
Entropía conjunta ($H(X,Y)$)	Incertidumbre total de dos variables	Sí	Análisis multivariable (logs, sensores, redes)	Considera interacción entre variables
Entropía condicional ($H(Y X)$)	Incertidumbre restante de Y dado X	Sí	Predicción, modelos probabilísticos	Menor valor = mayor dependencia
Información mutua ($I(X;Y)$)	Información compartida entre variables	Sí	Detección de patrones, selección de variables	0 = independencia
Redundancia (R)	Información repetitiva o ineficiente	No (pero deriva de H)	Compresión de datos, optimización de transmisión	Mayor valor = más compresión posible

Nota: La tabla sintetiza las principales métricas de la teoría de la información, incluyendo entropía, entropía conjunta, entropía condicional, información mutua y redundancia, destacando su interpretación y aplicaciones en análisis de datos, compresión y sistemas de comunicación. Elaboración propia con base en Alencar (2015), Cover y Thomas (2006), y Csiszár y Körner (2011).

RE FE REN CIAS

- **Alencar, M. S. (2015).** Information theory. Momentum Press.
- **Csiszár, I., & Körner, J. (2011).** Information theory: Coding theorems for discrete memoryless systems (2nd ed.). Cambridge University Press.
- **Gorzalka, J. D., & Deloumeaux, P. (2012).** Information theory: New research. Nova Science Publishers.



síguenos en:



www.pucevirtual.puce.edu.ec