- Teoría de la Información -

Trabajo Práctico Integrador N°4

Integrantes:

Barriga, Nahuel

Bedini, Tomas

Hernandez, Julieta

Repositorio GitHub: https://github.com/NahuelBarriga/TeoDeLaInfo

Introducción

En el siguiente Trabajo Práctico desarrollaremos la explicación del procedimiento logrado. Se analizaron distintos canales de información a partir de las probabilidades de la fuente y de la matriz del canal brindado por la cátedra. Con estos datos, se calcularon sus propiedades, tales como la entropía a priori y posterior, la equivocación y la información mutua. Además, se simuló el envío de mensajes y, de ser solicitado, se aplicó el método de paridad cruzada al conjunto de mensajes a enviar. Finalmente, se determinó la cantidad de mensajes enviados correcta e incorrectamente y la cantidad de mensajes corregidos.

Canal de información

Un canal de información es un medio por el que se transmite la información desde la fuente de información al destino. Es importante analizarlos ya que en la transmisión de la información puede haber sucesos no deseados como la pérdida de información o el ruido

Entropías, Equivocación e Información Mutua

Marco Teórico

Todos los canales de información brindados por la cátedra, vienen determinados por un alfabeto de entrada *A = {ai} = 0,1* ; un alfabeto de salida *B = { bj} = 0, 1* ; un conjunto de probabilidades condicionales *P (bj/ai)* y las probabilidades de la fuente de entrada *P(ai).*

Para obtener los resultados, previamente se calcularon las siguientes probabilidades:

- **Probabilidad de salida**: Es la probabilidad de observar bj sin conocer el símbolo que entró.

$$P(bj) = \sum_{i=1}^{2} P(ai) * P(B/ai)$$

- **Probabilidad condicional**: Es la probabilidad de que haya entrado el símbolo ai sabiendo que en la salida ha aparecido el símbolo bj.

$$P(ai/bj) = \frac{P(bj/ai) * P(ai)}{P(bj)}$$

- **Probabilidad de suceso simultáneo**: es la probabilidad de que habiendo entrado ai, el simbolo de salida sea bj.

$$P(ai, bj) \quad = \quad P(ai/bi) \quad * \quad P(bi)$$

Entropía A priori

Es el número medio de binits necesarios para representar un símbolo de una fuente con una probabilidad a priori *P (xi)*..

$$H(A) = \sum_{A} P(A) * \log_2(\frac{1}{P(A)})$$

$$H(B) = \sum_{B} P(B) * \log_2(\frac{1}{P(B)})$$

Entropía A posteriori

Representa el número medio de binits necesarios para representar un símbolo de una fuente con una probabilidad a posteriori P(ai/bj). De manera análoga, utilizando la matriz del canal se puede definir la entropía de salidas conocidos los símbolos de entrada. Estas fórmulas son:

$$H(A/bj) = \sum_{A} P(a/bj) * \log_2(\frac{1}{P(a/bj)})$$

$$H(B/ai) = \sum_{B} P(b/ai) * \log_2(\frac{1}{P(b/ai)})$$

Entropía Afín

Mide la incertidumbre del suceso simultáneo (ai, bj) La probabilidad de este suceso es P(ai, bj), de modo que la entropía afín valdrá:

$$H(A,B) = \sum_{A,B} P(a,b) * \log_2(\frac{1}{P(a,b)})$$

Equivocación

Se denomina ruido al número mínimo de preguntas binarias en promedio para determinar la entrada conocida y la salida. Se obtiene mediante la fórmula:

$$H(A/B) = \sum_{B} P(b) * H(A/b)$$

Se lo denomina pérdida al número mínimo de preguntas binarias en promedio para determinar la salida conocida la entrada. Se obtiene mediante la fórmula:

$$H(B/A) = \sum_{B} P(a) * H(B/a)$$

Información mutua

Medida de la incertidumbre sobre la salida del canal que se resuelve enviando la entrada. Es un indicador de la información ganada debido al acople entre las variables A (entrada) y B (salida).

Cuanto mayor sea el valor de la información mutua, mayor será el canal de información ya que menos incertidumbre habrá en la entrada y salida de la información, pues en los símbolos de entrada y de salida habrá dependencia estadística.

$$I\ (A,B) = H(A) - H(A/B)$$

$$I\ (B,A) = H(B) - H(B/A)$$

Resultados obtenidos para los diferentes canales

Cálculos	Entropía A Priori		Entropía A-posteriori (a1=0, a2=1; b1=0, b2=1)				Entropía Afin	Equivocación Ruido - Pérdida		Información mutua	
Canales	**H(A)**	**H(B)**	**H(A/b1)**	**H(A/b2)**	**H(B/a1)**	**H(B/a2)**	**H(A,B)**	**H(A/B)**	**H(B/A)**	**I(A,B)**	**I(B,A)**
tp4_sample0	0,811	0,953	0,921	0,72	1	0,915	1.747	0,795	0,936	0,016	0,016
tp4_sample1	0	0	0	0	0	0	0	0	0	0	0
tp4_sample2	1	1	0	0	0	0	1	0	0	1	1
tp4_sample3	0	0	0	0	0	0	0	0	0	0	0
tp4_sample4	1	1	0	0	0	0	1	0	0	1	1
tp4_sample5	1	1	1	1	1	1	2	1	1	0	0
tp4_sample6	0,881	0,843	0,824	0,573	0,915	0,469	1,484	0,641	0,603	0,24	0,24
tp4_sample7	0,722	0,469	0,503	0	0	1	0,922	0,453	0,2	0,269	0,269

Tabla 1: Resultados obtenidos (A: fuente de entrada, B: salida)

En los casos donde el resultado no puede ser posible debido a que hay una división por cero, se los considera el resultado como cero.

Análisis de los resultados obtenidos

tp4_sample0

P(ai/bj)	b1	b2
a1	0,33	0,199
a2	0,664	0,8

Tabla 2.1: Probabilidades condicionales Canal 0

P(ai,bj)	b1	b2
a1	0,125	0,125
a2	0,248	0,503

Tabla 2.2: Probabilidades de sucesos simulteneos Canal 0

P(b1)	P(b2)
0,373	0,628

Tabla 2.3: Probabilidades de salida Canal 0

Como medida de calidad del canal hemos calculado las entropías a priori y de salida, dándonos como resultado para la entropía a priori H(A) = 0,811 el número medio de unidades de información necesarios para representar un símbolo de la fuente dada, y para la entropía de salida, H(B) = 0,953 el número mínimo de preguntas binarias en promedio para determinar la salida.

Las probabilidades de los sucesos simultáneos son todas distintas de cero, indicando que el canal presenta distorsiones, es decir, que dada una entrada todas las salidas son posibles y viceversa.

Al comparar la entropía de entrada con la entropía afín, siendo $H(A) < H(A,B)$, nos indica que la incertidumbre aumenta. Esto es lógico ya que en éstos canales el valor de la información mutua es muy bajo, y por definición esta propiedad es dependiente de la información mutua.

El número medio de binits necesarios para representar un símbolo de la fuente habiendo recibido un cero es mayor que al haber recibido un uno.

En cuanto a la equivocación tenemos como resultado que $H(A/B) < H(A)$ por lo tanto hay mayor incertidumbre y no se pierde información al observar la salida.

Respecto a la Información mutua, que es la cantidad de información sobre A que atraviesa el canal, verificamos que cumpla las siguientes propiedades y sacando las siguiente conclusiones:

- $I >= 0$: Los símbolos de entrada y salida NO son estadísticamente independientes. No se pierde en absoluto información por el hecho de observar la salida del canal.

- $I(A,B) = 0.016 = I(B,A) = 0.016$: Hay reciprocidad de la información mutua. Lo que esta propiedad establece es que la cantidad de información que se obtiene de A gracias al conocimiento de B, es igual a la cantidad de información que se obtiene de B gracias al conocimiento de A.

Por lo tanto, queda a la salida del canal 0,016 y hay un ruido de 0,795.

<u>**tp4_sample1 y tp4_sample3**</u>

P(a1)	P(a2)
1	0

Tabla 3.1: Probabilidades de entrada Canal 1

P(a1)	P(a2)
0	1

Tabla 4.1: Probabilidades de entrada Canal 3

La entropía de entrada H(A) y de salida H(B) son cero debido a que uno de los eventos tiene probabilidad 1 y el otro tiene probabilidad 0, es decir, no hay incertidumbre en la salida dada la una entrada específica.

Dado que $I(A,B) = 0$, esto significa que los símbolos de entrada y salida son estadísticamente independientes. En consecuencia, se deduce que es un buen canal, ya que al menor ser su valor, más independientes son los símbolos de entrada con los símbolos de salida, pues la gran cantidad de ruido del canal da incertidumbre de cuál es la entrada al observarse una salida determinada. Aún así, para que el canal sea confiable, el valor de la información mutua debería ser mayor.

tp4_sample2 y tp4_sample4

P(ai/bj)	b1	b2
a1	1	0
a2	0	1

Tabla 5.1: Probabilidades condicionales Canal 2

P(ai,bj)	b1	b2
a1	0,5	0
a2	0	0,5

Tabla 5.2: Probabilidades de sucesos simulteneos Canal 2

P(b1)	P(b2)
0,5	0,5

Tabla 5.3: Probabilidades de salida Canal 2

P(ai/bj)	b1	b2
a1	0	1
a2	1	0

Tabla 6.1: Probabilidades condicionales Canal 4

P(ai,bj)	b1	b2
a1	0	0,5
a2	0,5	0

Tabla 6.2: Probabilidades de sucesos simulteneos Canal 4

P(b1)	P(b2)
0,5	0,5

Tabla 6.3: Probabilidades de salida Canal 4

Como medida de calidad del canal hemos calculado las entropías a priori y de salida, dándonos como resultado para la entropía a priori H(A) = 1 el número medio de unidades de información necesarios para representar un símbolo de la fuente dada, y para la entropía de salida, H(B) = 1 el número mínimo de preguntas binarias en promedio para determinar la salida. Estos resultados implican que la fuente A y la salida B tiene el máximo nivel de incertidumbre posible. En este caso, los eventos 0 y 1 son igualmente probables, y no hay patrón o estructura discernible que haga que uno sea más probable que el otro.

Al comparar la entropía de entrada con la entropía afín, siendo H(A) = H(A,B) nos indica que la incertidumbre se mantiene.

Respecto a la Información mutua, que es la cantidad de información sobre A que atraviesa el canal, verificamos que cumpla las siguientes propiedades y sacando las siguiente conclusiones:

- I >= 0 : Los símbolos de entrada y salida NO son estadísticamente independientes. No se pierde en absoluto información por el hecho de observar la salida del canal.

- I(A,B)= 1 = I(B,A)= 1 : Hay reciprocidad de la información mutua. Lo que esta propiedad establece es que la cantidad de información que se obtiene de A gracias al conocimiento de B, es igual a la cantidad de información que se obtiene de B gracias al conocimiento de A.
- Es un gran canal de información debido a su alto valor de información mutua. Esto implica que habrá menos incertidumbre en la entrada y salida de la información, pues en los símbolos de entrada y de salida habrá dependencia estadística.

Por lo tanto, queda a la salida del canal 1 y no existe ruido.

tp4_sample5

P(ai/bj)	b1	b2
a1	0,5	0,5
a2	0,5	0,5

Tabla 7.1: Probabilidades condicionales Canal 5

P(ai,bj)	b1	b2
a1	0,25	0,25
a2	0,25	0,25

Tabla 7.2: Probabilidades de sucesos simulteneos Canal 5

P(b1)	P(b2)
0,5	0,5

Tabla 7.3: Probabilidades de salida Canal 5

Dado que todos los eventos son equiprobables, todas las entropías tendrán valor 1, es decir, que el nivel de incertidumbre es el máximo sobre el símbolo que emitirá la fuente.

La equivocación igual a 1 implica que la probabilidad de error en la predicción de la salida a partir de la entrada es máxima. Esto sugiere que la relación entre la entrada y la salida es compleja y difícil de prever.

Respecto a la Información mutua, que es la cantidad de información sobre A que atraviesa el canal, verificamos que cumpla las siguientes propiedades y sacando las siguiente conclusiones:

- I >= 0 : Los símbolos de entrada y salida NO son estadísticamente independientes. No se pierde en absoluto información por el hecho de observar la salida del canal.
- I(A,B) = 1 = I(B,A) = 1 : Hay reciprocidad de la información mutua. Lo que esta propiedad establece es que la cantidad de información que se obtiene de A gracias al

conocimiento de B, es igual a la cantidad de información que se obtiene de B gracias al conocimiento de A.
- Es un gran canal de información debido a su alto valor de información mutua. Esto implica que habrá menos incertidumbre en la entrada y salida de la información, pues en los símbolos de entrada y de salida habrá dependencia estadística.

tp4_sample6

P(ai/bj)	b1	b2
a1	0,742	0,136
a2	0,258	0,864

Tabla 8.1: Probabilidades condicionales Canal 6

P(ai,bj)	b1	b2
a1	0,201	0,099
a2	0,07	0,63

Tabla 8.2: Probabilidades de sucesos simulteneos Canal 6

P(b1)	P(b2)
0,271	0,729

Tabla 8.3: Probabilidades de salida Canal 6

Como medida de calidad del canal hemos calculado las entropías a priori y de salida, dándonos como resultado para la entropía a priori H(A) = 0,881 el número medio de unidades de información necesarios para representar un símbolo de la fuente dada, y para la entropía de salida, H(B) = 0,843 el número mínimo de preguntas binarias en promedio para determinar la salida.

Las probabilidades de los sucesos simultáneos son todas distintas de cero, indicando que el canal presenta distorsiones, es decir, que dada una entrada todas las salidas son posibles y viceversa.

Al comparar la entropía de entrada con la entropía afín, siendo H(A) < H(A,B), nos indica que la incertidumbre aumenta. Esto es lógico ya que en éstos canales el valor de la información mutua es muy bajo, y por definición esta propiedad es dependiente de la información mutua.

El número medio de binits necesarios para representar un símbolo de la fuente habiendo recibido un cero es mayor que al haber recibido un uno.

En cuanto a la equivocación tenemos como resultado que H(A/B) < H(A) por lo tanto hay mayor incertidumbre y no se pierde información al observar la salida.

Respecto a la Información mutua, que es la cantidad de información sobre A que atraviesa el canal, verificamos que cumpla las siguientes propiedades y sacando las siguiente conclusiones:

- I >= 0 : Los símbolos de entrada y salida NO son estadísticamente independientes. No se pierde en absoluto información por el hecho de observar la salida del canal.

- $I(A,B) = 0.24 = I(B,A) = 0.24$: Hay reciprocidad de la información mutua. Lo que esta propiedad establece es que la cantidad de información que se obtiene de A gracias al conocimiento de B, es igual a la cantidad de información que se obtiene de B gracias al conocimiento de A.

Por lo tanto, queda a la salida del canal 0,24 y hay un ruido de 0,641.

tp4_sample7

P(ai/bj)	b1	b2
a1	0,889	0
a2	0,111	1

Tabla 9.1: Probabilidades condicionales Canal 7

P(ai,bj)	b1	b2
a1	0,8	0
a2	0,1	1

Tabla 9.2: Probabilidades de sucesos simulteneos Canal 7

P(b1)	P(b2)
0,9	0,1

Tabla 9.3: Probabilidades de salida Canal 7

Como medida de calidad del canal hemos calculado las entropías a priori y de salida, dándonos como resultado para la entropía a priori H(A) = 0,722 el número medio de unidades de información necesarios para representar un símbolo de la fuente dada, y para la entropía de salida, H(B) = 0,469 el número mínimo de preguntas binarias en promedio para determinar la salida.

Las probabilidades de los sucesos simultáneos P(ai,bj) son distintas de cero para P(0,0), P(1,0) y para P(1,1), indicando que el canal presenta distorsiones, es decir, que para dichas entradas todas dichas salidas son posibles y viceversa. Distinto es el caso de la probabilidad P(a1,b2), que es cero ya que la probabilidad P(0,1) es nula.

Al comparar la entropía de entrada con la entropía afín, siendo $H(A) < H(A,B)$, nos indica que la incertidumbre aumenta. Esto es lógico ya que en éstos canales el valor de la información mutua es muy bajo, y por definición esta propiedad es dependiente de la información mutua.

El número medio de binits necesarios para representar un símbolo de la fuente habiendo recibido un cero es mayor que al haber recibido un uno.

En cuanto a la equivocación tenemos como resultado que $H(A/B) < H(A)$ por lo tanto hay mayor incertidumbre al observar la salida.

Respecto a la Información mutua, que es la cantidad de información sobre A que atraviesa el canal, verificamos que cumpla las siguientes propiedades y sacando las siguiente conclusiones:

- $I >= 0$: Los símbolos de entrada y salida NO son estadísticamente independientes. No se pierde en absoluto información por el hecho de observar la salida del canal.

- I(A,B) = 0.269 = I(B,A) = 0.269 : Hay reciprocidad de la información mutua. Lo que esta propiedad establece es que la cantidad de información que se obtiene de A gracias al conocimiento de B, es igual a la cantidad de información que se obtiene de B gracias al conocimiento de A.

Por lo tanto, queda a la salida del canal 0,269 y hay un ruido de 0,453.

Envío de N x M mensajes

Implementación

Para generar los los N mensajes de longitud M, se utilizan las probabilidades de la fuente de entrada P(ai). Se desarrolló un algoritmo que, para cada bit del mensaje, mediante la función random, se genera un valor aleatorio entre [0,1) y si este valor es menor o igual que la probabilidad P(a1) el bit a agregar en el mensaje será 0, caso contrario será 1.

- $0 \leq x \leq P(a1)$ → El bit a agregar será *0*
- $P(a1) < x \leq 1$ → El bit a agregar será *1*

Método paridad cruzada

Marco Teórico

El principal objetivo del método de paridad cruzada es detectar y corregir errores en la transmisión de datos, garantizando la precisión de la información recibida. Se busca lograr una mayor robustez en el envío de mensajes, especialmente en entornos propensos a interferencias o ruido.

El procedimiento consiste en agregar una fila y una columna a la matriz de mensajes y aplicar un criterio de paridad. Se debe elegir entre el criterio par, donde el bit de paridad será un 0 si el número total de "1" en el mensaje es par, caso contrario será un 1; y el criterio impar. donde el bit de paridad será un 0 si el número total de "1" en el mensaje es impar, caso contrario será un 1.

Finalmente, la nueva columna estará formada por los bits de paridad horizontal de todas las filas y la nueva fila estará formada por los bits de paridad vertical de todas las columnas. Adicionalmente se emplea un bit de paridad cruzada que se calcula a partir de los bits de paridad de filas y columnas.

En caso de requerir aplicar el método de paridades, el procedimiento será el siguiente:

- Generar los mensajes a enviar con las probabilidad de entrada.
- Aplicar el método de paridades al conjunto de mensajes a enviar.
- Enviar los mensajes junto con las paridades mediante la matriz del canal.
- Aplicar el método de paridades al conjunto de mensajes recibidos.
- Determinar la cantidad de mensajes correctos, incorrectos y corregidos.

Envío de mensajes

Para simular el envío de mensajes, se debe pasar la matriz de mensajes por la matriz del canal. Se desarrolló un algoritmo donde para cada bit *xi* de cada mensaje, se utilizan las probabilidades aposteriori de la fila *i* de la matriz del canal. Luego, mediante la función random, se genera un valor aleatorio entre [0,1) y si este valor es menor o igual que la probabilidad P(b1/ai) el bit a enviar será 0, caso contrario será 1.

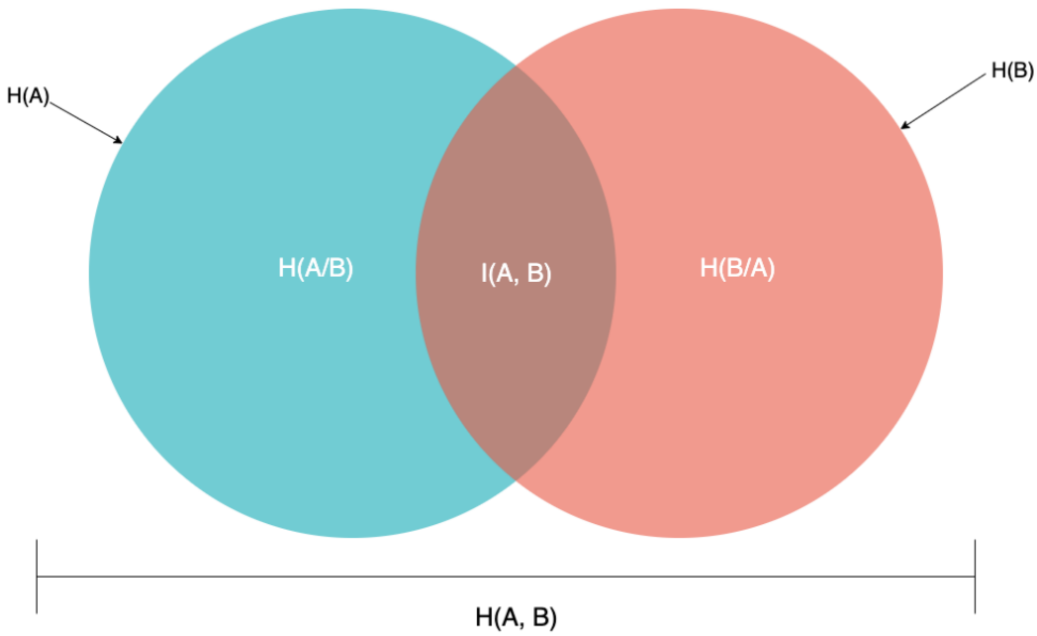
Mensajes correctos, incorrectos y corregidos

En ausencia de la solicitud explícita del flag -p, la estrategia adoptada para la detección de errores consistió en la comparación entre cada mensaje destinado a ser enviado y cada mensaje recibido. Esta resolución, aunque funcional, no representa el método óptimo, especialmente en contextos realistas, ya que el conjunto de mensajes enviado por la fuente no estaría disponible para su comparación directa con los mensajes recibidos.

De ser solicitado el flag -p, se aplicó el método de paridades cruzadas. Primeramente se calculan las paridades al conjunto de mensajes a enviar y se transmite esta matriz. En el receptor, se reciben los datos junto con los bits de paridad y se recalculan las paridades tanto por filas como por columnas utilizando los datos recibidos. Cualquier discrepancia entre los bits de paridad recalculados y los recibidos indica la presencia de errores. De presentarse discrepancia entre un solo bit de la fila de paridades y un solo bit de la columna de paridades, el error puede ser detectado y por ende corregido, caso contrario, no puede ser detectado debido a que son varios los posibles bit erróneos.

Conclusiones

Para concluir con el análisis de los Canales de Información, se puede resumir todo en el diagrama de Venn siguiente:



Se podría entender entonces a la **información mutua** como la zona de intersección entre entre las variables A y B, es decir, lo condicionadas que están una por la otra. En consecuencia, la información mutua sería igual a cero si las variables de entrada y salida fueran estadísticamente independientes. Hechos los cálculos de la equivocación, se pudo analizar la información mutua en cada caso, para todos los canales son valores mayores a cero. Este valor mientras más alto sea, mayor será la cantidad de información promedio de la entrada que resuelve el canal, es decir, que tan bueno es cada canal para representar la información que recibe en la entrada. Se puede concluir que cuánta más pérdida haya, menos cantidad de información se obtendrá de A gracias al conocimiento de B, lo cuál es lógico.

La **incertidumbre de sucesos simultáneos** es la suma de la entropía "a priori" más la entropía de salida menos la información mutua. Es decir, que si el valor de la información mutua es alto, se puede obtener mucho conocimiento de la entrada o de la salida a partir del conocimiento de una de ellas. Si este valor es bajo, los sucesos son poco dependientes, lo cuál genera mucha incertidumbre que se den dos sucesos de manera simultánea.

En promedio, nunca se pierde información al observar la salida ya que $H(A/B) < H(A)$.

En cuanto al **método de paridades cruzadas**, este no corrige los errores identificados, sino que simplemente los detecta. Esto implica que se requiere la implementación de mecanismos adicionales para abordar y corregir los errores una vez que han sido identificados. En entornos donde la integridad de los datos es crítica, esta limitación puede resultar insuficiente.

Otra desventaja a considerar es que el método puede volverse menos eficaz en situaciones donde se producen múltiples errores durante la transmisión de un mensaje. La capacidad de detección puede disminuir significativamente en tales casos, lo que pone en entredicho la robustez del método en escenarios de comunicación más complejos.