

Práctica 3.

Sorpresa: el problema de la comunicación con la realidad.

Docente: Gustavo Landfried

Inferencia Bayesiana Causal 1
1er cuatrimestre 2025
UNSAM

Índice

1. La estructura de la función de costo epistémica	6
1.1. Calcular la esperanza de los recursos en los primeros dos pasos temporales con $b = 0,5$.	7
1.2. Simular 10000 personas, y graficar el promedio de sus recursos en los primeros 10 pasos temporales.	8
1.3. Graficar cómo se actualizan los recursos en el tiempo con una apuesta determinada.	8
1.4. Graficar la tasa de crecimiento temporal ($\lim_{T \rightarrow \infty}$) de los recursos para cada apuesta alternativa b	8
1.5. Buscar la apuesta que maximiza los recursos si se modifica el pago que ofrece la casa de apuestas.	9
1.6. Calcular analíticamente la diferencia de desempeño de dos apuestas b y d dado unos pagos fijos Q_c y Q_s	9
1.7. Calcular la diferencia de las tasas de crecimiento de dos apuestas b y d a largo plazo ($\lim_{T \rightarrow \infty}$) expresado en escala logarítmica.	9
1.8. ¿Cuándo se maximiza la tasa de crecimiento a largo plazo?	10
2. Evaluación de procedimientos alternativos de medición independientes.	10
2.1. Hui-Walter. Evaluación de instrumentos de medición independientes.	11
2.1.1. Generación de datos sintéticos de tests independientes	13
2.1.2. Realizar inferencia mediante algún PPL	13
2.1.3. Diagnosticar si el motor de inferencia funciona correctamente.	14
2.1.4. Repetir el procedimiento pero con datos generados a partir de dos tests con covarianzas.	14
2.1.5. Conclusiones.	15
2.2. Dendukuri. Evaluación de procedimientos alternativos de medición que co-varían.	15
2.2.1. Simular diagnósticos dependientes y estimar el desempeño de 2 tests con el modelo Dendukuri	17
2.3. Harmony. El error de especificación del modelo europeo	18
2.3.1. Evaluación empírica	19
2.4. Propuestas de extensión.	20
3. Trabajo Final 1 de la materia: Estimación de desempeño de test diagnósticos conociendo el estado de la muestras pero con datos falsificados o cargados con error.	20
3.1. Problema de investigación	21

3.2.	Especificar un modelo causal que incluye tanto la evaluación del test diagnóstico como la evaluación del desempeño de los operadores de los laboratorios.	21
3.3.	Utilizar algún método para realizar inferencia con el modelo causal propuesto. . . .	21
3.4.	Graficar el posterior sobre el desempeño de los operadores.	21
3.5.	Graficar el posterior sobre el desempeño de los test en el contexto de alta y baja concentración.	21
3.6.	Calcular de a pares la probabilidad que el desempeño de un test sea superior al otro en el contexto de concentraciones baja.	22
4.	Anexo	23
4.1.	Infinite probability distributions.	23

Sistemas de comunicación con la realidad

La realidad produce señales constantemente, pero como seres humanos percibimos apenas un pequeño subconjunto, de forma distorsionada a través de instrumentos de medición imperfectos como nuestros cuerpos o tecnologías, o indirectamente, a través de epifenómenos observables del fenómeno oculto de interés. En este contexto, el problema fundamental de la teoría de la información es diseñar sistemas de comunicación con la realidad que permitan reconstruir los mensajes distorsionados e indirectos que percibimos, reproduciendo lo más fielmente posible el mensaje enviado por la realidad. En palabras de Claude Shannon [1].

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. (. . .). The significant aspect is that the actual message is one selected from a set of possible messages.

The system must be *designed* to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

Es importante destacar dos puntos de esta cita. El primero es que Shannon recuerda, en pleno siglo 20, la importancia de que el sistema de comunicación no se limite a seleccionar un único mensaje o hipótesis, como se solía hacer en su época, sino que considere todos los posibles mensajes alternativos como emerge de la aplicación estricta de las reglas de la probabilidad.

$$\underbrace{P(\text{Hipótesis, Datos})}_{\text{Creencia inicial compatible con los datos}} = \underbrace{P(\text{Hipótesis})}_{\text{Creencia inicial}} \underbrace{P(\text{Datos}|\text{Hipótesis})}_{\text{Sorpresa, fuente de información.}}$$

Las creencias sobre las hipótesis alternativas se actualizan preservando la creencia previa que sigue siendo compatible con los nuevos datos. Allí surge naturalmente que la sorpresa es el único filtro de la creencia previa y por lo tanto única fuente de información.

El segundo punto a destacar es que Shannon subraya que los sistemas de comunicación no vienen dados sino que se diseñan (*designed*). Sistemas alternativos comunican mensajes alternativos de la realidad. Si el mensaje que recibimos de la realidad no está dado sino que se construye: ¿cuál es entonces el mensaje enviado por la realidad? Debido a que los sistemas de comunicación alternativos son también hipótesis, su evaluación depende de la sorpresa contenida en la predicción que hace de los datos, $P(\text{Datos}|\text{Sistema})$.

En la primera unidad ya evaluamos dos sistemas de comunicación alternativos, que se diferenciaban únicamente por su modelo causal (Monty Hall vs Base), y calculamos sus predicciones $P(\text{Datos}|\text{Modelo Causal})$ para actualizar la creencia sobre los modelos. En el ejemplo del modelo Monty Hall, todas sus hipótesis internas (c, s, r) son observadas al final de cada episodio. En la teórica hemos demostrado la siguiente igualdad.

$$\begin{aligned} & -\log_2 \lim_{T \rightarrow \infty} P(\text{Datos}_T | \text{Modelo Causal})^{1/T} \\ &= \sum_{c, s, r} P(c, s, r | \text{Realidad Causal} = R) \cdot (-\log_2 P(c, s, r | \text{Modelo Causal} = M)) = \mathcal{H}(R, M) \end{aligned}$$

Esto es lo que descubre Shannon: la tasa de predicción a largo plazo ($\lim_{T \rightarrow \infty}$) expresada en órdenes de magnitud (\log_2) del sistema de comunicación es el negativo de la tasa de información en órdenes de magnitud que recibe el sistema, lo que se llama entropía cruzada $\mathcal{H}(R, M)$.

Shannon muestra además que la mínima tasa de información a la que puede aspirar un sistema de comunicación ocurre cuando la predicción que hace el sistema coincide con la probabilidad de generación de los datos por la realidad causal, $P(c, s, r | \text{Realidad Causal}) = P(c, s, r | \text{Modelo Causal})$. En este caso, la entropía cruzada se transforma en entropía.

$$\mathcal{H}(R) = \sum_{c, s, r} P(c, s, r | R) \cdot (-\log P(c, s, r | R))$$

Todo el resto de los sistemas de comunicación reciben información adicional, conocida como divergencia Kullback-Leibler.

$$\sum_{c,s,r} P(c, s, r|R) \cdot \underbrace{(\log P(c, s, r|R) - \log P(c, s, r|M))}_{\substack{\text{La diferencia de predicción} \\ \text{en órdenes de magnitud}}}$$

Estructura invariante

Los datos se expresan como proposiciones funcionales pues siempre que midamos una cierta variable f (como la *habilidad*) sobre una unidad de análisis x (como Messi) deberíamos obtener el mismo valor h , al menos en un momento determinado.

$$Habilidad(Messi) = h$$

Sin embargo la habilidad es una **variable oculta** que debe inferirse a partir de sus consecuencias observables, como son el éxito o fracaso de las competencias en las que participó.

$$P(Habilidad(Messi) = h | \text{Datos})$$

La variable oculta de interés es la fuente desde donde la realidad causal emite señales que se codifican a través de sus **epifenómenos observables**, como Ganar/Perder. Debido a que existen otros epifenómenos potencialmente observables, su elección es una decisión de diseño del sistema de comunicación que se toma buscando que ellos transmitan información relevante de la variable oculta de interés.

Los epifenómenos no se observan efectivamente hasta que se aplica un determinado **procedimiento de medición**. Al existir muchos procedimientos alternativos. La elección de un procedimiento de medición particular es otra decisión de diseño del sistema de comunicación que involucra consideraciones relativas a la factibilidad de su aplicación y a su capacidad para detectar fielmente la señal producida por el epifenómeno (y no otros factores asociados). Para estudiar la habilidad de Messi, por ejemplo, sería razonable desarrollar una software que descargue todos los resultados de la página oficial fifa.com. Las señales efectivamente registradas (True/False) constituyen la base empírica sobre la que se evaluarán las hipótesis de habilidad alternativas.

La señal registrada del epifenómeno no sirve si no es interpretada (o decodificada) a través de un **modelo causal** que permita vincularla con la variable oculta, la habilidad de Messi. Por

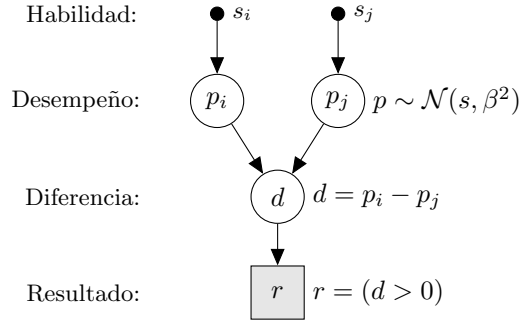


Figura 1: Modelo causal que vincula la habilidad oculta con el resultado observado.

ejemplo, considerar sólo la frecuencia de éxitos como indicador de la habilidad de los individuos puede conducir a aproximaciones erróneas, fundamentalmente porque el valor de un éxito o un fracaso depende de la dificultad de los desafíos. Por esta razón, todos los modelos de habilidad ampliamente utilizados consideran que el resultado de la competencia o resolución de problema se produce en relación a otra entidad oculta, que puede ser la habilidad de la persona contra la que compiten o la dificultad intrínseca del problema. En la figura 1 especificamos el modelo

causal generativo en el que las habilidades ocultas causan los resultados observables a través de la diferencia de los desempeños ocultos, $d = p_i - p_j$, ambas variables aleatorias centradas en la verdadera habilidad, $p \sim \mathcal{N}(s, \beta^2)$. Quien haya obtenido mayor rendimiento gana, $r = (d > 0)$.

Dado el modelo causal y la señal registrada cada hipótesis de habilidad (o posible mensajes del sistema de comunicación), tiene asociada una sorpresa, lo que permite actualizar su distribución de creencias (o **posterior**). La distribución de creencias a priori sobre la habilidad de un individuo i , s_i , se suele representar mediante una distribución gaussiana, $p(s_i) = \mathcal{N}(s_i | \mu_i, \sigma_i^2)$. Por las

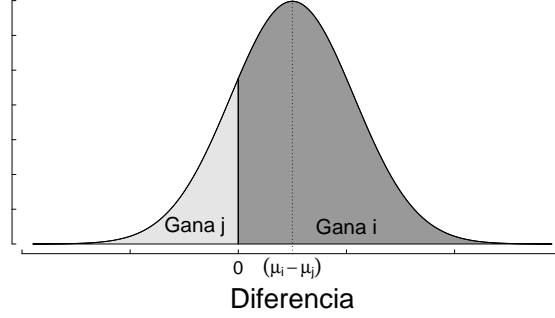


Figura 2: Predicción de los resultados

propiedades de la gaussiana (descritas en detalle en el anexo de la práctica anterior) ustedes pueden derivar el siguiente resultado: la distribución marginal de la variable que representa la diferencia de desempeños $p(d)$ (con distribución de probabilidad condicional determinística $p(d | p_i, p_j) = \mathbb{I}(d = p_i - p_j)$) es una gaussiana que está centrada en la diferencia de las medias de las distribuciones de habilidad a priori $(\mu_i - \mu_j)$, y que tiene como varianza la incertidumbre de ambas estimaciones (σ) y la varianza de ambos rendimientos (β), $\mathcal{N}(d | \mu_i - \mu_j, 2\beta^2 + \sigma_i^2 + \sigma_j^2)$. Por lo tanto, la predicción a priori o evidencia de un resultado ganador es la densidad acumulada (Φ) de todos los valores positivos de la gaussiana de diferencia de desempeños:

$$\underbrace{P(r)}_{\text{Evidencia}} = 1 - \Phi\left(0 \mid \underbrace{\mu_i - \mu_j}_{\text{Diferencia esperada}}, \underbrace{2\beta^2 + \sigma_i^2 + \sigma_j^2}_{\text{Incertidumbre total}}\right). \quad (1)$$

La evidencia es la verosimilitud marginal $p(r)$, la predicción a priori del resultado observado realizada con la contribución de todas las hipótesis elementales de habilidad.

Para calcular el posterior sobre la habilidad $p(s_i | r)$ necesitamos la verosimilitud, $p(r | s_i)$. En el anexo de la práctica anterior se ofrecen todas las propiedades de la distribución gaussiana que se requieren para llegar a la siguiente expresión analítica exacta.

$$\underbrace{p(s_i | r)}_{\text{Posterior}} \propto \underbrace{1 - \Phi(0 | s_i - \mu_j, 2\beta^2 + \sigma_j^2)}_{\text{Verosimilitud } P(r | s_i)} \underbrace{\mathcal{N}(s_i | \mu_i, \sigma_i^2)}_{\text{Prior } p(s_i)}. \quad (2)$$

Donde el posterior normalizado se obtiene dividiendo el lado derecho con la evidencia, $P(r)$.

En este momento me interesa que noten las similitudes y diferencias entre la verosimilitud y la evidencia. La verosimilitud $p(r | s_i)$, a diferencia de la evidencia $p(r)$, supone que la habilidad oculta s_i es conocida. Y si revisan con detalle la diferencia entre ambas expresiones, van a ver que lo que se expresaba usando la media de la estimación μ_i en la evidencia, se reemplaza directamente por el valor de habilidad s_i , y que la incertidumbre asociadas a esa estimación σ_i no aparece cuando reemplazamos μ_i por s_i . Cuando estudiemos flujo de inferencia en modelos causales volveremos a ver esta expresión y verán como emerge naturalmente.

El posterior no es más que la densidad del prior no filtrada por la verosimilitud. La sorpresa, definida como el complemento de la verosimilitud, funciona como un filtro para el prior. En la región de hipótesis de muy alta habilidad, donde el resultado ganador no nos hubiera generado casi ninguna sorpresa ($\lim_{s_i \rightarrow \infty} P(r | s_i) = 1$), el posterior recibe casi toda la densidad del prior.

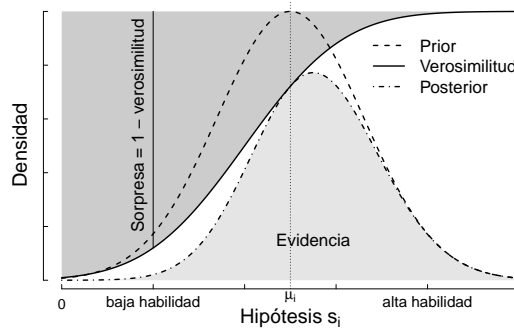


Figura 3: Actualización de creencias para el caso ganador.

En cambio, en la región de hipótesis de muy baja habilidad, donde el resultado habría generado mucha sorpresa ($\lim_{s_i \rightarrow -\infty} P(r|s_i) = 0$), el posterior no recibe casi nada de la densidad del prior.

Estos 5 elementos son invariante en el diseño de los sistemas de comunicación.

- **Fuente:** el estado *real* de la variable *habilidad*(Messi)
- **Codificador:** el epifenómeno Ganar/Perder
- **Canal:** la medición *Scrapper* de **fifa.com**
- **Decodificador:** la interpretación Modelo Causal
- **Mensaje:** el posterior $P(\text{habilidad}(\text{Messi}) = h|d, M)$

La estructura invariante de los sistemas de comunicación con la realidad puede resumirse en el siguiente esquema.

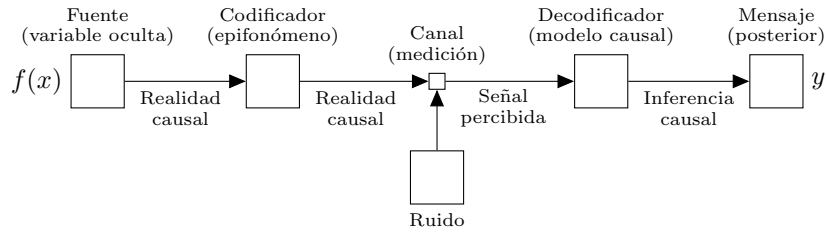


Figura 4: Esquema de sistema de comunicación general propuesto por Shannon.

Si modificamos cualquiera de esos 5 elementos, la variable oculta de interés, el epifenómeno observable, el procedimiento de medición, el modelo causal, y el método de estimación del mensaje, entonces tendremos sistemas de comunicación alternativos.

1. La estructura de la función de costo epistémica

En pleno siglo 20, Shannon recuerda la importancia de que el sistema de comunicación no se limite a seleccionar un único mensaje o hipótesis, como se solía hacer en su época, sino que considere todos los posibles mensajes alternativos como emerge de la aplicación estricta de las reglas de la probabilidad. A diferencia de los enfoques que seleccionan una única hipótesis mediante funciones de costo arbitrarias (e.g. por máxima verosimilitud), la aplicación estricta de las reglas de la probabilidad induce una función de costo para evaluar las hipótesis mutuamente contradictorias (A y no A) que tiene una estructura multiplicativa, en la que el filtro de las creencias previas, única fuente de información, se realiza mediante el producto de las sorpresas.

$$\underbrace{P(\text{Hipótesis}, \text{Datos} = \{d_1, d_2, \dots\})}_{\text{Creencia inicial compatible con los datos}} = \underbrace{P(\text{Hipótesis})}_{\text{Creencia inicial}} \underbrace{P(d_1|\text{Hipótesis})P(d_2|d_1, \text{Hipótesis})\dots}_{\text{Sorpresas, fuentes de información.}}$$

Si esta interpretación no había quedado clara, 8 años más tarde su colega John Kelly publica el artículo *A New Interpretation of Information Rate*, texto aprobado por el propio Shannon. El artículo comienza así.

The cost function approach [...] can actually be used to analyze nearly any branch of human endeavor. The utility theory of Von Neumann shows us one way to obtain such a cost function. The point here is that **an arbitrary combination of a statistical transducer (i.e., a channel) and a cost function does not necessarily constitute a communication system.** What can be done, however, is to take some real-life situation which seems to possess the essential features of a communication problem, and to analyze it without the introduction of an arbitrary cost function. The situation which will be chosen here is one in which a gambler uses knowledge of the received symbols of a communication channel in order to make profitable bets on the transmitted symbols. (Kelly 1956 *A New Interpretation of Information Rate* Bell Labs)

En este artículo, Kelly deriva las propiedades de los sistemas de comunicación usando un ejemplo de apuesta. Supongamos que una casa de apuestas paga $Q_c = 3,0$ si el lanzamiento de una moneda sale Cara y $Q_s = 1,2$ si el lanzamiento de una moneda sale Sello. Sin pérdida de generalidad podemos suponer que los individuos se ven obligados a apostar en cada paso temporal todos sus recursos, asignando una proporción $b_c = b$ a Cara y $b_s = 1 - b$ a Sello. Si nuestros recursos iniciales son ω_0 , los recursos que obtenemos con una apuestas $b \in [0, 1]$ después de observar una Cara y una Seca es el producto de nuestros resultados.

$$\omega(b|\text{Datos} = \{\text{Cara}, \text{Sello}\}) = \underbrace{\omega_0 \overbrace{b Q_c}^{\text{Cara}}}_{\omega(b|\text{Cara})} \overbrace{(1 - b) Q_s}^{\text{Sello}}$$

En el primer paso apostamos una proporción $\omega_0 b$ a Cara y $\omega_0(1 - b)$ a Sello. Como se observó Cara, perdemos lo apostado a Sello, y nos multiplican por Q_c lo apostado a Cara, por lo que nuestros recursos pasan a ser $\omega(b|\text{Cara})$. En el segundo paso ocurre algo parecido, apostamos una proporción $\omega_1(b|\text{Cara}) \cdot b$ a Cara y $\omega_1(b|\text{Cara}) \cdot (1 - b)$ a Sello. Como se observó Sello, perdemos lo apostado a Cara, y nos multiplican por Q_s lo apostado a Sello, por lo que nuestros recursos pasan a ser $\omega_2(b|\text{Datos})$.

1.1. Calcular la esperanza de los recursos en los primeros dos pasos temporales con $b = 0,5$.

Recordar que la definición matemática de la esperanza es de los recursos es,

$$\mathbb{E}(\omega_T(b)) = \sum_d^{\{\text{Cara}, \text{Sello}\}^T} P(d) \cdot \omega(b|d)$$

donde $\{\text{Cara}, \text{Sello}\}^T$ es el producto cartesiano de los posibles resultados en T pasos temporales. Es decir, en el primer paso temporal hay solo dos opciones.

$$\mathbb{E}(\omega_1(b)) = P(\text{Cara})\omega(b|\text{Cara}) + P(\text{Sello})\omega(b|\text{Sello})$$

Y en el segundo paso temporal hay 4 opciones.

$$\mathbb{E}(\omega_2(b)) = P(\text{Cara}, \text{Cara})\omega(b|\text{Cara}, \text{Cara}) + \dots + P(\text{Sello}, \text{Sello})\omega(b|\text{Sello}, \text{Sello})$$

Suponiendo que los recursos iniciales son $\omega_0 = 1$, ¿cuál es la tasa de crecimiento en cada paso temporal según la esperanza?

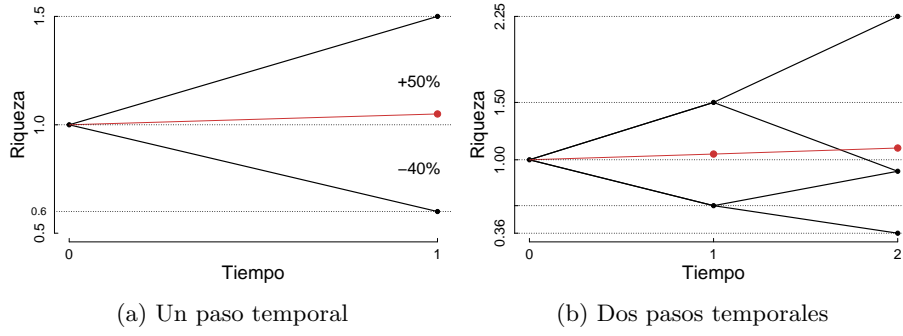
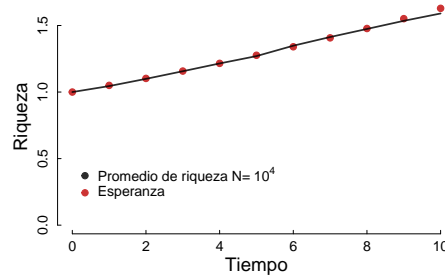


Figura 5: La línea negra representa las trayectorias individuales de los recursos. La línea roja representa la media.

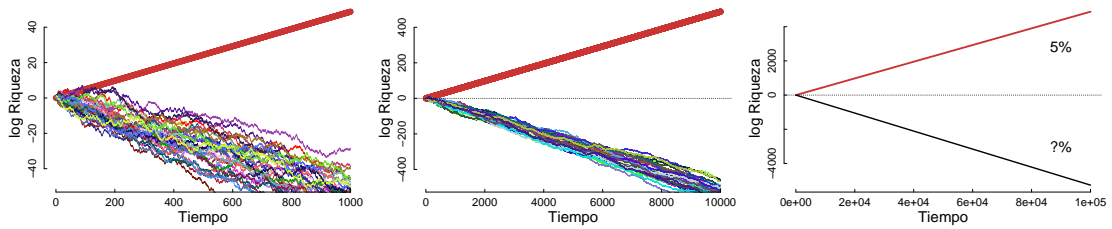
1.2. Simular 10000 personas, y graficar el promedio de sus recursos en los primeros 10 pasos temporales.

Mostrar que efectivamente el promedio de los recursos de los 10000 individuos crece siguiendo la tasa de crecimiento calculada en el punto anterior con la esperanza.



1.3. Graficar cómo se actualizan los recursos en el tiempo con una apuesta determinada.

Elegir una apuesta $b \in [0, 1]$ y calcular cómo se actualizan los recursos a medida que vamos observando datos hasta un tiempo T igual a 100 o 1000. Repetir el procedimiento varias veces. Guardar el valor de los recursos en cada paso temporal para graficar todas las trayectorias simuladas en el tiempo.



1.4. Graficar la tasa de crecimiento temporal ($\lim_{T \rightarrow \infty}$) de los recursos para cada apuesta alternativa b .

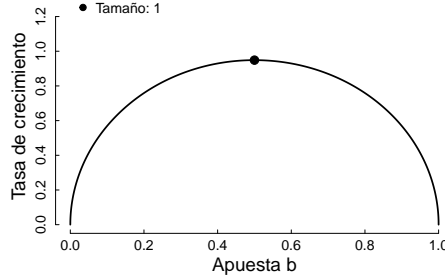
Sea n_c y n_s la cantidad de Caras y Secas observadas hasta el tiempo $T = n_c + n_s$, los recursos obtenidos hasta el tiempo T son,

$$\omega_T(b) = \omega_0(b Q_c)^{n_c} ((1-b) Q_s)^{n_s}$$

Calcular la tasa de crecimiento $r(b)$ de los recursos a largo plazo, cuando T tiende a infinito.

$$\lim_{T \rightarrow \infty} \omega_T(b) = r(b)^T$$

Hacer una grilla de valores de posibles apuestas (entre 0 y 1), y graficar la tasa de crecimiento de los recursos.



1.5. Buscar la apuesta que maximiza los recursos si se modifica el pago que ofrece la casa de apuestas.

Supongamos que la casa de apuestas cambia los pagos por Cara y por Sello a $Q_c = 10$ y $Q_s = 1$.

¿Cuál es la apuesta que maximiza los recursos?

1.6. Calcular analíticamente la diferencia de desempeño de dos apuestas b y d dado unos pagos fijos Q_c y Q_s .

Una forma de comparar el desempeño de dos apuestas alternativas b y d se evaluando el cociente entre sus recursos.

$$\frac{\omega_T(b)}{\omega_T(d)} = \frac{\omega_0(b Q_c)^{n_c} ((1-b) Q_s)^{n_s}}{\omega_0(d Q_c)^{n_c} ((1-d) Q_s)^{n_s}}$$

¿La elección de la apuesta depende del pago que ofrece la casa?

1.7. Calcular la diferencia de las tasas de crecimiento de dos apuestas b y d a largo plazo ($\lim_{T \rightarrow \infty}$) expresado en escala logarítmica.

Dado unos pagos fijos Q_c y Q_s , la tasa de crecimiento a largo plazo de una apuesta b es.

$$r(b) = \left(\lim_{T \rightarrow \infty} \omega_T(b) \right)^{1/T}$$

Calcular la diferencia entre

$$\log \frac{r(b)}{r(d)} = \log \frac{\lim_{T \rightarrow \infty} \omega_0(b Q_c)^{n_c/T} ((1-b) Q_s)^{n_s/T}}{\lim_{T \rightarrow \infty} \omega_0(d Q_c)^{n_c/T} ((1-d) Q_s)^{n_s/T}}$$

Notar que $\lim_{T \rightarrow \infty} n_c/T = p_c = p$ la probabilidad de que salga cara, y $p_s = 1 - p$.

¿Qué relación tiene con la entropía, entropía cruzada y divergencia?

1.8. ¿Cuándo se maximiza la tasa de crecimiento a largo plazo?

Seleccionar la apuesta que maximiza la tasa de crecimiento.

$$\arg \min_b \log \frac{r(b)}{r(d)} = \arg \min_b p \log b + (1 - p) \log(1 - b)$$

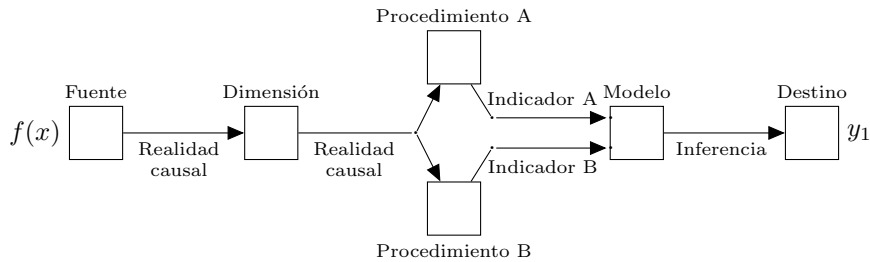
¿Qué relación tiene con la entropía?

2. Evaluación de procedimientos alternativos de medición independientes.

La enfermedad de Chagas es susceptible de tratamiento eficaz cuando se detecta en una etapa temprana. Sin embargo en la actualidad se cree que solo el 10 % de las personas infectadas conocen su estado. El principal cuello de botella para la identificación de nuevos casos es consecuencias del protocolos vigentes actualmente en América Latina para la detección de la enfermedad de Chagas. Debido a que no existe un test diagnóstico perfecto, se requieren la aplicación de varios tests que referencia, que deben ser analizados en laboratorios de alta complejidad que suelen solo en las grandes ciudades. Dado que la discrepancia entre los resultados de los test de referencia no es infrecuente, en caso de discordancia entre los diagnósticos de las dos primeras pruebas, se emplea un tercer test de referencia para desempatar.

Una de las ideas para resolver este problema propone modificar el protocolo actual, reemplazando los test diagnósticos de referencia, que son caros y solo pueden ser analizados en las grandes ciudades, por test rápidos baratos que puedan ser aplicados in situ, dando una respuesta concluyente al paciente de forma casi inmediata. Para ello es necesario evaluar el desempeño de los test diagnósticos. Como no existe ningún test diagnóstico perfecto, debemos evaluar el desempeño de los test sin saber qué individuo se encuentra infectado y que individuo se encuentra sano.

En medicina, los test diagnóstico son instrumentos de medición que se utilizan para medir epifenómenos, como por ejemplo antígenos, que están relacionados con la variable oculta de interés, por ejemplo el estado infeccioso del paciente. Dos procedimientos de medición alternativos producen dos sistemas de comunicación alternativos, si opta por una de ellos, o se pueden usar combinados para producir un único sistema de comunicación.



La evaluación del desempeño de un test diagnóstico es sencilla cuando disponemos de datos “gold standard”, que nos garantizan el valor real de la clase. En ese caso, podemos calcular la probabilidad de diagnóstico dado el estado real a través de la tabla 2 de contingencias.



	D = Positivo = 1	D = Negativo = 0
Estado = Verdadero	Sensibilidad , s	1-Sensibilidad, $1 - s$
Estado = Falso	1-Especificidad, $1 - x$	Especificidad , x

En el caso de la enfermedad de Chagas sabemos que no existe ningún test diagnóstico perfecto, por lo cual en general no podemos observar directamente esta tabla. En estos casos es necesario inferir las *variables ocultas* a partir de los diagnósticos reportados por los diferentes test: el estado real de las personas, la sensibilidad y especificidad de los test y la prevalencia de la enfermedad en la población. La prevalencia hace referencia a la proporción de la población que está infectada.

Los modelos bayesianos de evaluación de test diagnóstico se los conoce en la literatura médica y veterinaria como Bayesian Latent Class Analysis [2, 3]. Dos modelos B-LCA que tienen gran aceptación en la comunidad de evaluación de test diagnóstico el modelo Hui-Walter (1980) y el modelo Dendukuri (2001). El modelo Hui-Walter, en su versión bayesiana, es capaz de estimar los desempeños correctamente (y con baja incertidumbre) cuando los test son independientes entre sí. Cuando existen dependencias entre tests, sin embargo, el modelo Hui-Walter puede sobreestimar los desempeños de los tests. En los años 2000 se proponen los primeros modelos de clases latentes que incorporan una variable adicional para estimación la correlación entre pares de test [4, 5]. Como desventaja, este modelo solo permite evaluar el desempeño de a pares de tests, produciendo estimaciones con mucha incertidumbre.

Recientemente, la *European Cooperation in Science and Technology* una agencia europea para el financiamiento de redes de investigación y desarrollo inició el proyecto “Harmony (COST Action CA18208) Novel tools for test evaluation and disease prevalence estimation” (start date 25/10/2019, end date 04/04/2024)¹. La descripción del proyecto decía,

Despite the availability of appropriate statistical methods – Bayesian Latent Class Models (BLCMs) – that adjust for the imperfect accuracy of the diagnostic process and produce comparable prevalence estimates, the number of research studies and scientific reports that use them is small compared to the number of instances that use of such methods would have been optimal. The objective of this proposal is to coordinate and promote the implementation of BLCMs through networking and knowledge transfer between BLCM experts and researchers working in statistics, epidemiology, diagnostics and population health.

El *2nd Harmony training workshop on diagnostic test evaluation (Copenhagen 2021)*² fue impartido por Matthew Denwood, desarrollador del paquete `runjags` para el lenguaje de programación R³. Este paquete de R tenía hasta el año pasado una función interna `template.huiwalter()` que creaba los modelos BLCA automáticamente para que sean ejecutados por el lenguaje de programación probabilística `rjags`. Esta función fue usada en la sesión 4 del *2nd Harmony training workshop* para inferir el desempeño de 3 tests con covarianzas en datos simulados. En este ejercicio revisaremos el modelo Hui-Walter, el modelo Dendukuri y el error que tenía el modelo utilizado hasta el año pasado por la red europea *Harmony (COST Action CA18208) Novel tools for test evaluation and disease prevalence estimation* a través del método `template.huiwalter()` del paquete de `runjags`.

2.1. Hui-Walter. Evaluación de instrumentos de medición independientes.

El resultado de un test diagnóstico efectuado sobre N pacientes puede resumirse en un vector $\mathbf{r} = (n_0, n_1)$, donde n_0 y n_1 son la cantidad de diagnósticos negativos y positivos respectivamente, con $N = n_0 + n_1$. Los modelos suponen que la probabilidad de diagnósticos negativo y positivo, $\mathbf{q} = (q_0, q_1)$, depende de la sensibilidad s y especificidad x del test, pero también de la prevalencia

¹Reference: <https://www.cost.eu/actions/CA18208/> and <https://harmony-net.eu/wp-content/uploads/CA18208-e.pdf>

²Training materials at <https://harmony-net.eu/denmark-training-school-material/>, source code at <https://github.com/ku-awdc/HarmonyWS2> and fork at <https://github.com/glandfried/HarmonyWS2>

³Package description at <https://cran.r-project.org/web/packages/runjags/index.html> and source code at <https://github.com/ku-awdc/runjags>

de la enfermedad en la población, $q_1 = ps + (1 - p)(1 - x)$.

$$\begin{aligned} p(\mathbf{q}|p, s, x) &= \mathbb{I}(\mathbf{q} = (q_0, q_1)) \\ q_0 &= 1 - q_1 \\ q_1 &= ps + (1 - p)(1 - x) \end{aligned}$$

Es decir, la probabilidad de que un test sea positivo viene dada por una suma de los casos en los que el test detecta correctamente la enfermedad (prevalencia por sensibilidad, ps) más los casos que fueron detectados como positivos pero que no estaban infectados (Complementos de la prevalencia y especificidad $(1 - p)(1 - x)$). Luego, la probabilidad del resultado observado es $P(\mathbf{r}|\mathbf{q}, N) = \text{Binomial}(\mathbf{r}|\mathbf{q}, N)$.

$$p(\mathbf{r}|\mathbf{q}, N) = \text{Binomial}(\mathbf{r}|\mathbf{q}, N)$$

Finalmente, la prevalencia, la sensibilidad y la especificidad se modelen mediante distribuciones $\text{Beta}(\cdot, \cdot)$.

$$p(p) = \text{Beta}(p|\alpha_p, \beta_p) \quad p(s) = \text{Beta}(s|\alpha_s, \beta_s) \quad p(x) = \text{Beta}(x|\alpha_x, \beta_x)$$

En el figura 10 podemos ver la especificación del modelo usando la notación de *factor graphs*. Notar que este modelo tiene 3 grados de libertad (p , s y x) y un solo dato $n_0 = N - n_1$. Esto

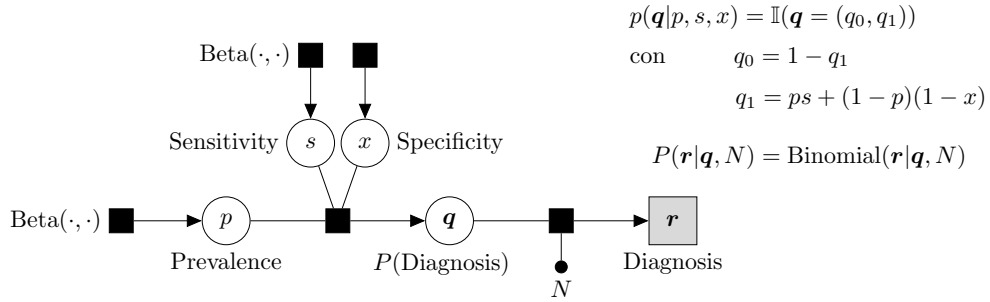


Figura 10: Modelo de una población y un test.

quiere decir que hay muchas combinaciones de p , s y x que son igualmente verosímiles y por lo tanto indistinguibles entre sí. Para poder identificar el valor real es necesario tener al menos la misma cantidad de datos que de hipótesis (mismas variables libres que variables restrictoras). Por lo tanto, este modelo básico no es capaz de identificar (o estimar) las hipótesis debido a que hay más variables libres que datos.

El más sencillo capaz de evaluar el desempeño de los test diagnósticos es el modelo Hui-Walter. El modelo Hui-Walter es igual al anterior, pero con al menos dos test y dos comunidades. El resultado observable de dos test es un vector, $\mathbf{r} = (n_0, n_1, n_2, n_3)$, donde n_0 , n_1 , n_2 , y n_3 representa la cantidad de resultados que son conjuntamente 00, 01, 10 y 11 respectivamente. El modelo Hui-Walter supone que la sensibilidad y especificidad de los test (s_t y x_t) son independientes entre sí. Luego, la probabilidad del resultado conjunto es un vector $\mathbf{q} = (q_0, q_1, q_2, q_3)$, donde $q_0 = p(1 - s_a)(1 - s_b) + (1 - p)x_a x_b$ (ver el resto de las probabilidades q_i en la figura 11). Y con N pacientes la probabilidad es $P(\mathbf{r}|\mathbf{q}, N) = \text{Multinomial}(\mathbf{r}|\mathbf{q}, N)$.

La especificación gráfica del modelo Hui-Walter es similar al modelo anterior, salvo que las estructuras que aparecen repetidas veces se representa de forma plegada usando placas que tienen por subíndice la variable que despliega la repetición, aquí por test y por comunidades.

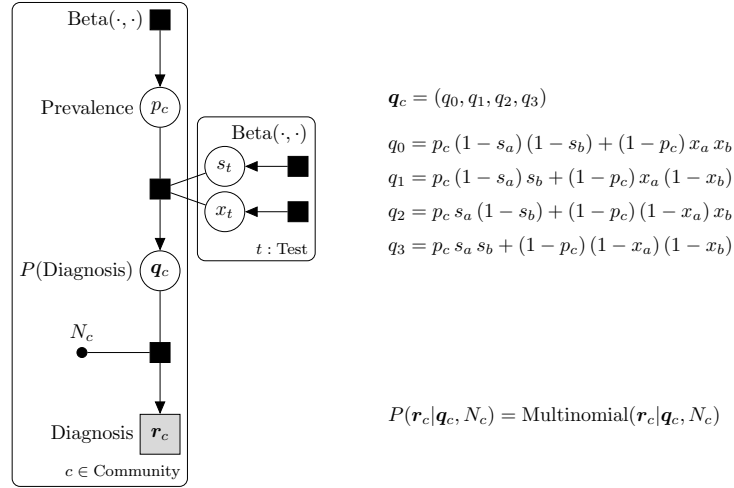


Figura 11: Modelo Hui-Walter [6]

Con dos poblaciones y dos test el modelo ya puede estimar las 6 hipótesis (p_c , s_t , x_t) gracias a que tenemos 6 datos (3 por cada r_c).

2.1.1. Generación de datos sintéticos de tests independientes

Para evaluar el modelo creamos un conjunto de datos simulados. En el siguiente código mostramos los datos generados a partir del modelo Hui-Walter, con valores de prevalencia $\mathbf{p} = (0, 0.15, 0.3, 0.45, 0.7)$ para 5 comunidades, y valores de sensibilidad y especificidad de dos test $\mathbf{s} = (s_a, s_b)$ y $\mathbf{x} = (x_a, x_b)$.

```
prevalence = c(0, 0.15, 0.3, 0.45, 0.7) # TRUE VALUES
sensitivity = c(0.9, 0.6)
specificity = c(0.95, 0.9)

r <- matrix(nrow=4, ncol=length(p)) # DATA (simulated)
r[,1] <- c(839, 104, 44, 7)
r[,2] <- c(754, 92, 85, 92)
r[,3] <- c(598, 99, 134, 151)
r[,4] <- c(503, 75, 187, 240)
r[,5] <- c(275, 61, 287, 373)
N <- apply(r, MARGIN=2, FUN=sum); Comunidades = 5
```

Simular los datos.

2.1.2. Realizar inferencia mediante algún PPL

Para realizar la inferencia, aquí usamos el lenguaje de programación probabilística JAGS, a través del paquete de R `runjags`. Ustedes usen el lenguaje de su preferencia. En JAGS la implementación del modelo es directa respecto de la especificación gráfica.

```
modelo <- "model{
  s[1] ~ dbeta(1, 1)
  s[2] ~ dbeta(1, 1)
  x[1] ~ dbeta(1, 1)
  x[2] ~ dbeta(1, 1)

  for(c in 1:Comunidades){
    p[c] ~ dbeta(1, 1)

    q[1,c] <- p[c]*(1-s[1])*(1-s[2]) + (1-p[c])*x[1]*x[2]
    q[2,c] <- p[c]*(1-s[1])*s[2] + (1-p[c])*x[1]*(1-x[2])
    q[3,c] <- p[c]*s[1]*(1-s[2]) + (1-p[c])*(1-x[1])*x[2]
    q[4,c] <- p[c]*s[1]*s[2] + (1-p[c])*(1-x[1])*(1-x[2])
  }
}
```

```

    r[1:4, c] ~ dmulti(q[1:4, c], N[c])
  }
  #data# r, N, Comunidades
  #monitor# p, s, x
  #inits# p, s, x
}"

```

Para estimar los valores de sensibilidad, especificidad y prevalencia necesitamos indicar los valores iniciales a partir de donde el lenguaje de programación probabilística comenzará la inferencia (mediante cadenas de Markov). Finalmente, activamos el proceso de inferencia.

```

library("runjags") # JAGS: Just Another Gibbs Sampler

s <- list(chain1=c(0.5,0.99), chain2=c(0.99,0.5)) # INITs
x <- list(chain1=c(0.5,0.99), chain2=c(0.99,0.5))
p <- list(chain1=c(0.1, 0.1, 0.1, 0.9, 0.9), chain2=c(0.9, 0.9, 0.9, 0.1, 0.1))

results <- run.jags(modelo, n.chains=2)
print(results)

```

Realizar la inferencia del desempeño de los tests usando el lenguaje de programación probabilística de su preferencia.

2.1.3. Diagnosticar si el motor de inferencia funciona correctamente.

Como referencia, compartimos los resultados obtenidos y los comparamos con los valores reales.

	Lower95	Median	Upper95	Real
p[1]	6e-07	0.0071	0.0184	0.00
p[2]	0.1213	0.1501	0.1792	0.15
p[3]	0.2478	0.2829	0.3207	0.30
p[4]	0.3922	0.4324	0.4718	0.45
p[5]	0.6573	0.7016	0.7428	0.70
s[1]	0.8822	0.9233	0.9646	0.95
s[2]	0.5556	0.5829	0.6129	0.60
x[1]	0.9429	0.9566	0.9705	0.95
x[2]	0.8755	0.8897	0.9034	0.90

Antes de sacar conclusiones siempre se debe revisar la estabilidad de las cadenas de Markov. Una evaluación básica consiste en revisar los valores de las variables **SSEff** y **psrf**. Como ayuda memoria, necesitamos **SSEff** > 2000 y **psrf** \approx 1.

	SSEff	AC.10	psrf
p[1]	5575	0.02143	1.0000
p[2]	6162	0.03274	1.0002
p[3]	5323	0.01990	1.0000
p[4]	4992	0.04256	0.9999
p[5]	3730	0.04781	1.0004
s[1]	2742	0.08044	1.0009
s[2]	7065	0.01934	1.0002
x[1]	4558	0.04616	1.0007
x[2]	4260	0.04574	1.0001

Revisar la salud de las cadenas de Markov.

2.1.4. Repetir el procedimiento pero con datos generados a partir de dos tests con covarianzas.

Si bien el modelo Hui-Walter es capaz de recuperar los verdaderos valores de sensibilidad y especificidad de los tests sin necesidad de introducir gold standard, para que funcione correctamente se requiere que los diagnósticos de los tests sean independientes entre sí dado el verdadero estado de los pacientes (infectado/no-infectado). Puede ocurrir que los resultados de los tests sean independientes entre sí cuando miden diferentes antígenos o usan métodos diferentes. En esos casos, el diagnóstico d_i de un test i no nos da información de lo que le va a ocurrir a otro test j cuando

condicionamos por el verdadero estado s de la infección del paciente, $P(d_b|d_a, i) = P(d_b|i)$. Sin embargo, es común que haya al menos dos tests que miden los mismo antígenos o que usan métodos similares, produciendo diagnósticos que están correlacionados entre sí, $P(d_b|d_a, i) \neq P(d_b|i)$.

Cuando se miden tests que tienen dependencias condicionales usando un modelo que supone independencias condicionales, se produce una sobre-estimación de la sensibilidad y especificidad. Para ejemplificar este problema, simulamos los diagnósticos realizados por dos tests con una co-varianza de 0,12 y 0,09 para estados infecciosos verdadero y falso respectivamente.

Estado = 1	$D_1 = 1$	$D_1 = 0$		Estado = 0	$D_1 = 1$	$D_1 = 0$	
$D_2 = 1$	0,72	0,03	0,75	$D_2 = 1$	0,1275	0,0225	0,15
$D_2 = 0$	0,08	0,17	0,25	$D_2 = 0$	0,1225	0,7275	0,85
	0,80	0,20			0,25	0,75	

Cuadro 1: Esta distribución conjunta ha sido creada con la siguiente información: $s_1 = 0,8$, $s_2 = 0,75$, $x_1 = 0,75$, $x_2 = 0,85$, $cov_1 = 0,12$, $cov_0 = 0,09$. En la siguiente sección discutiremos en detalle cómo construir este tipo de tablas.

Mostrar las estimaciones con sus intervalos de credibilidad y la diferencia respecto del valor real

En la siguiente figura se muestra un ejemplo.

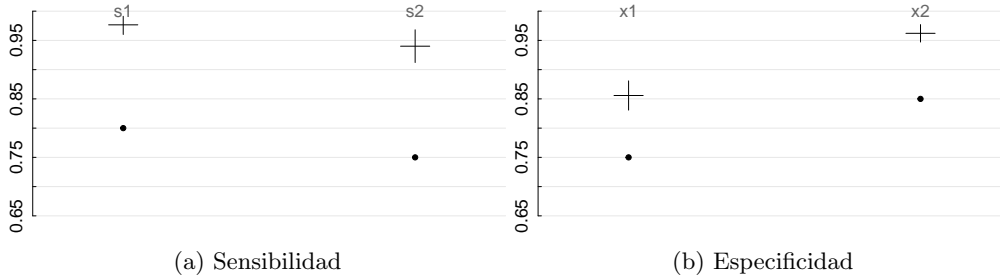


Figura 12: Estimaciones del desempeño de los tests que tienen covarianzas usando el modelo que asume independencias (Hui-Walter). Los puntos negros representan el valor real.

2.1.5. Conclusiones.

En estos casos en los que los test generan sus diagnósticos independientemente entre sí, ¿Somos capaces de detectar el desempeño de los test sin tener conocimiento del estado real de la persona (infectada/no-infectada), sin saber qué test miente y qué test dice la verdad?

Sacar conclusiones. ¿Se puede crear un detector de mentiras (y verdades) entre medios de comunicación que mienten (y dicen la verdad) con una frecuencia que es aleatoria (independiente) de los otros medios de comunicación?

2.2. Dendukuiri. Evaluación de procedimientos alternativos de medición que co-varían.

Hemos visto que usando existen dependencias entre tests, sin embargo, el modelo Hui-Walter puede sobreestimar los desempeños de los tests. Para evaluar la sensibilidad y especificidad en presencia de dependencias condicionales se han propuesto modelos alternativos [4, 7, 8], en los cuales se incorporan variables que estiman la co-varianza entre tests. Describiremos aquí el modelo de efectos fijos de la sección 2 del paper de Dendukuri [4].

Recordamos primero la definición de covariancia.

$$\begin{aligned}
\text{cov}(x, y) &= \mathbb{E}[(X - \mu_x)(Y - \mu_y)] \\
&= \mathbb{E}[X Y - \mu_x Y - X \mu_y + \mu_x \mu_y] \\
&= \mathbb{E}[X Y] - \mu_x \mathbb{E}[Y] - \mathbb{E}[X] \mu_y + \mu_x \mu_y \\
&= \mathbb{E}[X Y] - \mu_x \mu_y - \cancel{\mu_x \mu_y} + \cancel{\mu_x \mu_y} \\
&= \mathbb{E}[X Y] - \mathbb{E}[X] \mathbb{E}[Y]
\end{aligned}$$

Donde μ es la media de las variables, en este caso serán diagnósticos. Notar que la covarianza está definida sobre un par de variables, $\text{cov}(x, y)$. En el caso especial de variables binarias ($x, y \in \{0, 1\}$) la esperanza del producto es igual a la probabilidad conjunta cuando las dos variables son positivas,

$$\mathbb{E}[XY] = \sum_{\substack{x \in X \\ y \in Y}} P(x, y) x y = P(x = 1, y = 1)$$

Y de forma similar, la esperanza de una variable es igual a la probabilidad de que esa variable sea positiva,

$$\mathbb{E}[X] = \sum_{x \in X} P(x) x = P(x = 1)$$

Luego, esto es lo que nos permite escribir la covarianza en términos de la probabilidades conjunta y marginal de ambas variables positivas,

$$\begin{aligned}
\text{cov}(X, Y) &= P(x = 1, y = 1) - P(x = 1)P(y = 1) \\
P(x = 1, y = 1) &= P(x = 1)P(y = 1) + \text{cov}(X, Y)
\end{aligned} \tag{3}$$

Esta propiedad nos permite recuperar toda la distribución conjunta, $P(x, y)$, conociendo solamente las probabilidades marginales de ambas variables positivas, $P(x = 1)$ y $P(y = 1)$, y su covarianza $\text{cov}(x, y)$. La siguiente tabla representa la distribución conjunta de 2 variables binarias con las

$P(x, y)$	$x = 0$	$x = 1$	
$y = 0$			$1 - P(y = 1)$
$y = 1$		$P(x = 1)P(y = 1) + \text{cov}(x, y)$	$P(y = 1)$
	$1 - P(x = 1)$	$P(x = 1)$	1

Cuadro 2: Reconstructing joint distribution from marginals and covariance.

variables latentes que queremos estimar: marginales y covarianza. Dada la información de la marginales y la covarianza podemos reconstruir la distribución conjunta.

$$\begin{aligned}
P(x = 1, y = 1) &= P(x = 1)P(y = 1) + \text{cov}(x, y) \\
P(x = 0, y = 1) &= P(y = 1) - P(x = 1, y = 1) \\
&= P(y = 1) - P(x = 1)P(y = 1) - \text{cov}(x, y) \\
&= P(y = 1)(1 - P(x = 1)) - \text{cov}(x, y) \\
P(x = 0, y = 0) &= (1 - P(x = 1)) - P(x = 0, y = 1) \\
&= (1 - P(x = 1))(1 - P(y = 1)) + \text{cov}(x, y) \\
P(x = 1, y = 0) &= P(x = 1) - P(x = 1, y = 1) \\
&= P(x = 1)(1 - P(y = 1)) - \text{cov}(x, y)
\end{aligned}$$

El resultado final de estas cuatro ecuaciones es lo que se utiliza en el modelo Dendukuri para definir la variable \mathbf{q}_c en la figura 13. En el siguiente modelo gráfico agregamos dos variables c_s y

c_x que representan la covarianza de la sensibilidad y especificidad entre un par de test. Notar que en el modelo de Dendukuri hay dos distribuciones conjuntas independientes: la que corresponde con la sensibilidad $P(\text{diagnosis}_a, \text{diagnosis}_b | \text{state} = 1)$, y la que corresponde con la especificidad $P(\text{diagnosis}_a, \text{diagnosis}_b | \text{state} = 0)$.

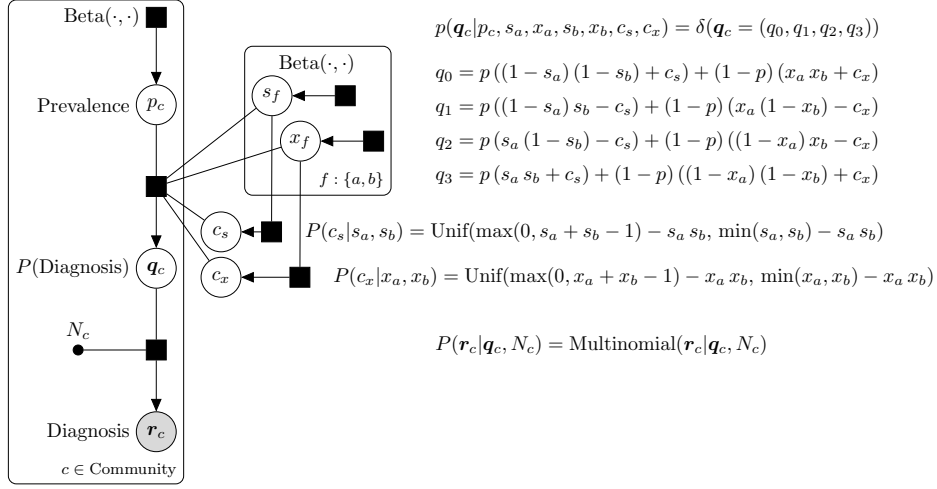


Figura 13: Modelo Dendukuri “fixed effect model” con covarianzas (2001)

2.2.1. Simular diagnósticos dependientes y estimar el desempeño de 2 tests con el modelo Dendukuri

A partir de los datos simulados utilizados para ilustrar las limitaciones del modelo de Hui-Walter en la sección anterior ($s_1 = 0,8, s_2 = 0,75, x_1 = 0,75, x_2 = 0,85, \text{cov}_1 = 0,12, \text{cov}_0 = 0,09$) mostrar que este nuevo modelo es capaz de recuperar las sensibilidades y especificidades verdaderas incluso en presencia de covarianzas entre pares de pruebas.

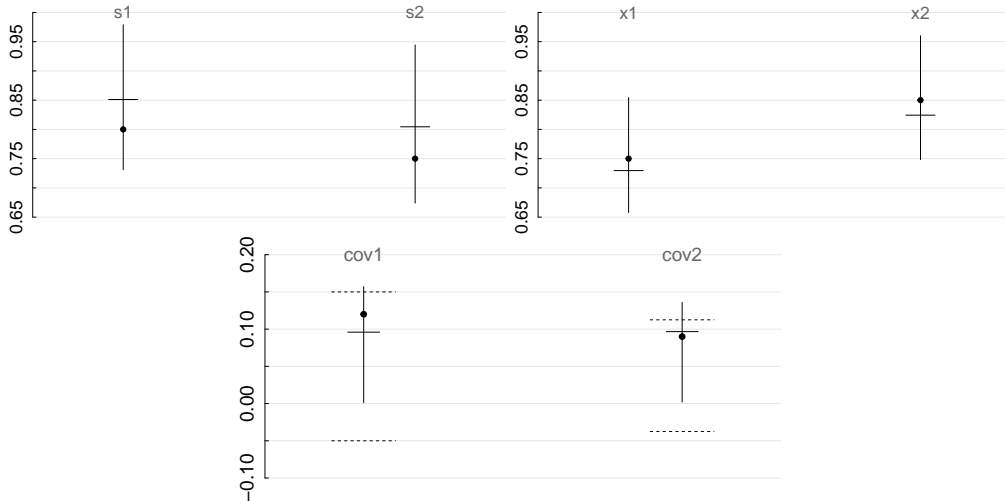


Figura 14: Estimación de las sensibilidades, especificidades y covarianza de dos tests en varias poblaciones.

A pesar de que las estimaciones de la figura 14 se realizaron usando 9 poblaciones, las incertidumbres son altas, similares a las que se obtienen observando una sola población. En particular, la incertidumbre de las covarianzas ocupan todo el rango (positivo) de valores de covarianzas posibles

dado las verdaderas sensibilidades y especificidades (rangos que han sido marcados con líneas punteadas). Posiblemente esto ocurre debido a que todas esas covarianzas tienen alguna sensibilidad y especificidad que hacen plausibles a los datos. En todo caso, a pesar de la alta incertidumbre, la mediana de las estimaciones están bastante cerca de los valores reales.

2.3. Harmony. El error de especificación del modelo europeo

El modelo Dendukuri (2001) ha sido adoptado ampliamente en la literatura de evaluación bayesiana de test diagnóstico. Como desventaja, este modelo ha sido definido para evaluar únicamente tests de a pares. ¿Es posible extender el modelo Dendukuri 2001 para estimar correctamente el desempeño de los tests diagnóstico en contextos de dependencia entre varios tests?.

The `runjags` package has an internal function `template_huiwalter()` that create a BLCA model to be executed by the probabilistic programming language `rjags`. This function was used at the session 4 of the 2nd Harmony training workshop to infer the performance of 3 tests in a simulate scenario in which there is a positive covariances between 2 of them. In the following code, we display the first part of the mentioned model. En el siguiente código mostramos la primera parte del modelo mencionado.

```
for(p in 1:Populations){
  # Probability of observing Test1- Test2- Test3- from a true positive::
  se_prob[1,p] <- prev[p] * ((1-se[1])*(1-se[2])*(1-se[3]) +covse12 +covse13 +covse23)
  # Probability of observing Test1- Test2- Test3- from a true negative::
  sp_prob[1,p] <- (1-prev[p]) * (sp[1]*sp[2]*sp[3] +covsp12 +covsp13 +covsp23)

  # Probability of observing Test1+ Test2- Test3- from a true positive::
  se_prob[2,p] <- prev[p] * (se[1]*(1-se[2])*(1-se[3]) -covse12 -covse13 +covse23)
  # Probability of observing Test1+ Test2- Test3- from a true negative::
  sp_prob[2,p] <- (1-prev[p]) * ((1-sp[1])*sp[2]*sp[3] -covsp12 -covsp13 +covsp23)
```

En términos generales, el modelo propuesto implementado en el la función `template_huiwalter()` define la distribución de probabilidad conjunta de dimensionalidad n , $P(\wedge_{i=1}^n X_i = v_i)$, función de las marginales, $P(X_i = v_i)$, y las covarianzas $\text{cov}(X_i, Y_i)$.

$$P(\wedge_{i=1}^n X_i = v_i) = \left(\prod_{i=1}^n P(X_i = v_i) \right) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n (-1)^{\mathbb{I}(v_i=v_j)} \text{cov}(X_i, Y_j) \quad (4)$$

donde $\mathbb{I}(v_i = v_j)$ representa la función indicadora que vale 1 si la se cumple la igualdad y 0 en caso contrario. Notar la diferencia de este modelo de covarianzas respecto del modelo descrito en la sección anterior. En la sección anterior derivamos el modelo de covarianza para el caso de dos variables binarias. Esta propuesta de extensión del modelo de Dendukuri es teóricamente incorrecta.

El principal problema teórico del modelo propuesto está relacionado con el hecho de que no se posible determinar los valores de las distribuciones de probabilidad conjunta de dimensionalidad mayor a 2 teniendo como datos únicamente las probabilidades marginales y las covarianzas. Por ejemplo, la distribución conjunta de tres variables binarias, $P(x, y, z)$, tiene 7 variables libres (8 probabilidades y la restricción de que sumen 1) y solo contamos con 6 datos (las 3 marginales $P(x = 1)$, $P(y = 1)$, $P(z = 1)$ y las 3 covarianzas $\text{cov}(x, y)$, $\text{cov}(x, z)$, $\text{cov}(y, z)$). Esto significa que existen infinitas distribuciones de probabilidad $P(x, y, z)$ compatible con los datos. Esto puede tener consecuencias negativas a la hora de identificar las variables latentes, pero no implica que el modelo sea teóricamente incorrecto.

Sin embargo, si el modelo propuesto es teóricamente correcto debe ocurrir que la distribución conjunta de dimensionalidad n sea compatible con la información de las marginales y sus covarianzas. Por ejemplo, en el caso de dimensionalidad 3 debe ocurrir $\sum_{k=0}^1 P(x, y, z = k) = P(x, y)$.

Si nos basamos en el modelo “Harmony” de covarianza, debe valer por definición que

$$\begin{aligned}
P(x = 1, y = 1) &= P(1, 1, 0) + P(1, 1, 1) \\
&= P_x P_y P_z + C_{xy} + \cancel{C_{xz}} + \cancel{C_{yz}} \\
&+ P_x P_y (1 - P_z) + C_{xy} - \cancel{C_{xz}} - \cancel{C_{yz}} \\
&= P_x P_y (1 - \cancel{P_z} + \cancel{P_z}) + 2C_{xy}
\end{aligned} \tag{5}$$

donde usamos los siguientes renombres: $P_x = P(x = 1)$ y $C_{xy} = \text{cov}(x, y)$. Notar que la solución de esta ecuación 5 tiene una covarianza adicional respecto del verdadero valor $P(x = 1, y = 1)$ que hemos derivado en la ecuación 3. Luego, el modelo propuesto implementado en el la función `template_huiwalter()` ni siquiera es compatible con las marginales.

2.3.1. Evaluación empírica

En esta sección evaluamos en datos simulados el desempeño de la generalización del modelo Dendukuri implementada en el paquete `runjags`. Para ello elegiremos una de las infinitas distribución de probabilidad conjunta $P(x, y, z)$ compatible con las siguientes distribuciones marginales y condicionales.

$$\begin{aligned}
P(x = 1) &= 0,80 & P(x = 1, y = 1) &= P(x = 1)P(y = 1|x = 1) = 0,80 \cdot 0,90 \\
P(y = 1) &= 0,75 & P(x = 1, z = 1) &= P(x = 1)P(z = 1) = 0,80 \cdot 0,70 \\
P(z = 1) &= 0,70 & P(y = 1, z = 1) &= P(y = 1)P(z = 1) = 0,75 \cdot 0,70
\end{aligned}$$

De aquí se desprende que las covarianzas son,

$$\begin{aligned}
\text{cov}(x, y) &= P(x = 1, y = 1) - P(x = 1)P(y = 1) = 0,72 - 0,60 = 0,12 \\
\text{cov}(x, z) &= \text{cov}(y, z) = 0
\end{aligned}$$

Esta información determina infinitas distribuciones conjuntas de a 3 variables ($P(v_1, v_2, v_3)$). En el apéndice se muestra como construir todas estas distribuciones conjuntas. Para la simulación de los datos tomamos la siguiente distribución conjunta.

P_{xy0}	x0	x1	P_{xy1}	x0	x1
y0	0.045	0.03	y0	0.125	0.05
y1	0.015	0.21	y1	0.015	0.51

Cuadro 3: Sensitivity/Specificity joint distribution used to create the synthetic data set.

Generamos 3000 pacientes en 4 poblaciones con diversas prevalencias, y generamos los resultados de los test diagnósticos mediante la distribución conjunta 3. Creamos el modelos pasándole los datos simulados al método `template_huiwalter()` del paquete `run.jags` y modificamos el modelo a mano para activar la estimación de covarianzas. Adjuntamos un archivo comprimido con todos los archivos para replicar la simulación y la inferencia. En la siguiente figura se puede observar el resultado de la inferencia obtenida con el paquete `runjags`.

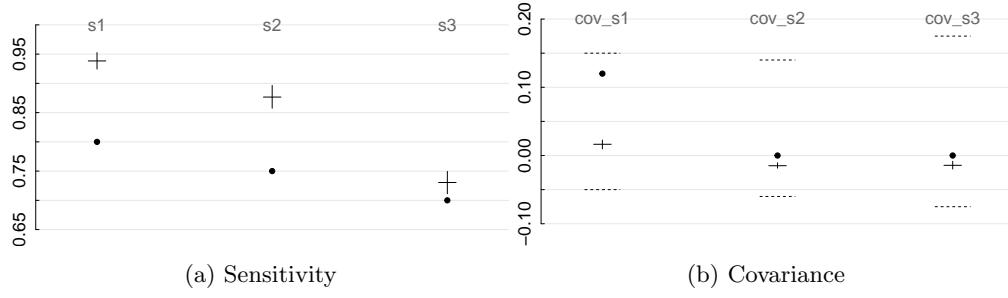


Figura 15: True values (points) and credibility intervals (ranges) obtained with the “Harmony” model for covariance estimation.

Se puede observar que las estimaciones del modelo Harmony no contienen los verdaderos valores con los que se generaron los datos simulados. Es decir, no solo hay un problema teórico con el modelo Harmony. El problema también es práctico.

2.4. Propuestas de extensión.

Hemos visto que la extensión del modelo Dendukuri 2001 implementada por el método `template_huiwalter()` es teóricamente incorrecta. En particular, no es compatible con las marginales. Pero además hemos visto, en datos simulados, que tampoco es capaz de estimar correctamente las variables latentes. Es necesario entonces encontrar un modelo alternativo que sea capaz de estimar las sensibilidades, especificidades y dependencias entre varios tests simultáneamente.

Recientemente, en el artículo *Modelling conditional dependence among multiple tests* [9] Wang, Dendukuri y otros presentan un modelo que sería capaz de estimar el comportamiento conjunto de todos los tests usando una generalización del concepto de covarianzas.

3. Trabajo Final 1 de la materia: Estimación de desempeño de test diagnósticos conociendo el estado de la muestras pero con datos falsificados o cargados con error.

En el ejercicio anterior hemos estudiado el desempeño de test diagnóstico imperfecto sin saber el verdadero estado de los individuos, usando como único observable la coincidencia o discrepancia entre tests. Vimos que podemos obtener buenas estimaciones de los desempeños a pesar de que no conocemos el estado real de los individuos cuando los datos analizados se generan a partir del mismo modelo causal usado para hacer la inferencia. En la vida real los datos pueden haber sido generados con otra realidad causal subyacente. En este ejercicio vamos a estudiar el desempeño de los test diagnóstico de Chagas en una base de datos real, generada por un importante grupo de investigación argentino.

Este grupo de investigación implementó una técnica capaz de agregar artificialmente una concentración específica del parásito a muestras de sangre de individuos que habitan comunidades sin prevalencia de la enfermedad. Para estudiar el desempeño de 3 test diagnósticos equivalentes de tres marcas distintas (DTU1, DTU2, DTU6), decidieron enviar a 14 laboratorios 3 muestras con una concentración baja del parásito (20 par.Eq/ml) y 3 muestras con una concentración alta del parásito (50 par.Eq/ml). Es decir, cada uno de los 3 tests (DTU1, DTU2, DTU6) reporta 6 diagnósticos (3 por concentración baja y 3 por concentración) en cada uno de los 14 laboratorios.

Uno de los problemas que hay en los estudios de evaluación del desempeño de los test diagnósticos es que los operadores (los empleados de los laboratorios) realizan a veces su trabajo de forma imprecisa, aplicando de forma errónea el protocolo requerido para el diagnóstico, o simplemente cargando los datos con error en la planilla.

3.1. Problema de investigación

Nuestro objetivo es proponer un model causal que considere la posibilidad de que los operadores trabajen con error, mostrando que efectivamente el desempeño de los 3 tests son significativamente diferentes en los caso en los que la concentración del parástio en la muestra es baja.

3.2. Especificar un modelo causal que incluye tanto la evaluación del test diagnóstico como la evaluación del desempeño de los operadores de los laboratorios.

En cualquier formato digitalizable. Recomendamos el template de latex.

3.3. Utilizar algún método para realizar inferencia con el modelo causal propuesto.

Probablemente usen algún lenguaje de programación probabilística. En ese caso deben verificar que la inferencia haya convergida y que las correlaciones entre las muestras sea baja.

3.4. Graficar el posterior sobre el desempeño de los operadores.

Queremos conocer el posterior respecto de que el dato haya sido cargado con error.

3.5. Graficar el posterior sobre el desempeño de los test en el contexto de alta y baja concentración.

Nos interesa la distribución de probabilidad a posteriori marginal de los desempeños de los tests.

$$p(s_t | \text{Datos})$$

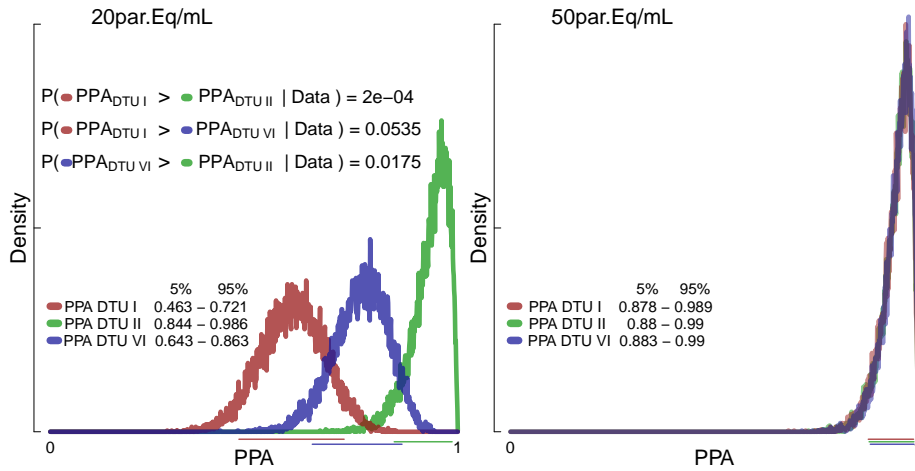


Figura 16: Estimaciones marginales de los desempeños

3.6. Calcular de a pares la probabilidad que el desempeño de un test sea superior al otro en el contexto de concentraciones baja.

Nos interesa conocer la probabilidad de que el desempeño de un test sea superior al otro en el contexto de concentraciones baja.

$$P(s_i > s_j | \text{Datos})$$

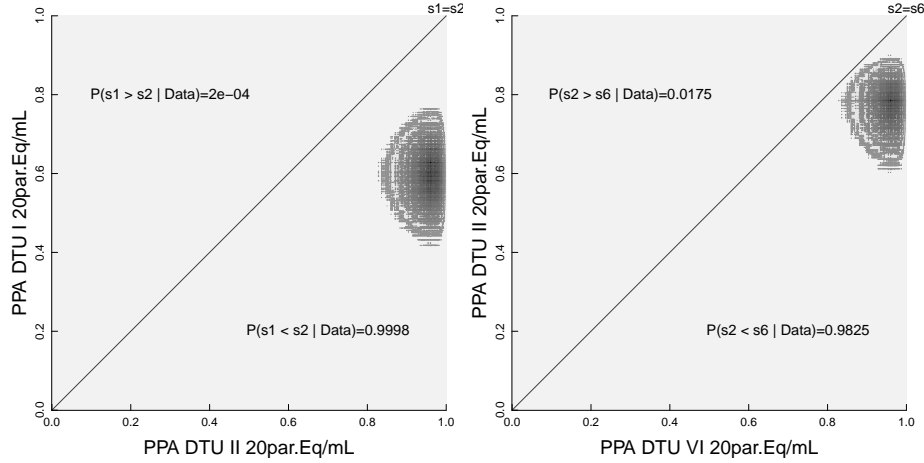


Figura 17: Estimaciones conjuntas de los desempeños

Referencias

- [1] Shannon CE. A mathematical theory of communication. Bell system technical journal. 1948;27(3):379–423.
- [2] Cheung A, Dufour S, Jones G, Kostoulas P, Stevenson M, Singanallur N, et al. Bayesian latent class analysis when the reference test is imperfect. Revue Scientifique et Technique (International Office of Epizootics). 2021;40(1):271–286.
- [3] Umemneku Chikere CM, Wilson K, Graziadio S, Vale L, Allen AJ. Diagnostic test evaluation methodology: a systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard—an update. PLoS One. 2019;14(10):e0223832.
- [4] Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. 2001;57(1):158–167.
- [5] Georgiadis MP, Johnson WO, Gardner IA, Singh R. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. Journal of the Royal Statistical Society Series C: Applied Statistics. 2003;52(1):63–76.
- [6] Hui SL, Walter SD. Estimating the error rates of diagnostic tests. Biometrics. 1980; p. 167–171.
- [7] Vacek PM. The effect of conditional dependence on the evaluation of diagnostic tests. Biometrics. 1985; p. 959–968.
- [8] Branscum A, Gardner I, Johnson W. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Preventive veterinary medicine. 2005;68(2-4):145–163.
- [9] Wang Z, Dendukuri N, Zar HJ, Joseph L. Modeling conditional dependence among multiple diagnostic tests. Statistics in Medicine. 2017;36(30):4843–4859.

4. Anexo

4.1. Infinite probability distributions.

Ahora nos preguntamos cómo construir las infinitas distribuciones conjuntas $P(x, y, z)$ tal que son compatibles con las marginales. Veamos un ejemplo con tres variables en el que solo hay covarianza entre dos variables, x e y .

$$\begin{aligned} P(x=1) &= 0,80 & P(x=1, y=1) &= P(x=1)P(y=1|x=1) = 0,80 \cdot 0,90 \\ P(y=1) &= 0,75 & P(x=1, z=1) &= P(x=1)P(z=1) = 0,80 \cdot 0,70 \\ P(z=1) &= 0,70 & P(y=1, z=1) &= P(y=1)P(z=1) = 0,75 \cdot 0,70 \end{aligned}$$

Luego, las covarianzas son

$$\begin{aligned} \text{cov}(x, y) &= P(x=1, y=1) - P(x=1)P(y=1) = 0,72 - 0,60 = 0,12 \\ \text{cov}(x, z) &= \text{cov}(y, z) = 0 \end{aligned}$$

Luego, las marginales en dos dimensiones son,

```
joint2D_from_marginals1D_and_covariance <- function(Px, Py, Cxy){
  P = array(NA, dim=c(2,2), dimnames = list(c("x0", "x1"), c("y0", "y1")))

  P["x0", "y0"] = (1-Px)*(1-Py) + Cxy
  P["x1", "y0"] = Px*(1-Py) - Cxy
  P["x0", "y1"] = (1-Px)*Py - Cxy
  P["x1", "y1"] = Px*Py + Cxy
  return(P)
}

Px = 0.80; Cxy = 0.12
Py = 0.75; Cxz = 0.0
Pz = 0.70; Cyz = 0.0

Pxy = joint2D_from_marginals1D_and_covariance(Px, Py, Cxy)
Pxz = joint2D_from_marginals1D_and_covariance(Px, Pz, Cxz)
Pyz = joint2D_from_marginals1D_and_covariance(Py, Pz, Cyz)
```

$P(x, y)$	$x = 0$	$x = 1$
$y = 0$	$P_{000} + P_{001} = 0,17$	$P_{100} + P_{101} = 0,08$
$y = 1$	$P_{010} + P_{011} = 0,03$	$P_{110} + P_{111} = 0,72$

$P(x, z)$	$x = 0$	$x = 1$
$z = 0$	$P_{000} + P_{010} = 0,06$	$P_{100} + P_{110} = 0,24$
$z = 1$	$P_{001} + P_{011} = 0,14$	$P_{101} + P_{111} = 0,56$

$P(y, z)$	$y = 0$	$y = 1$
$z = 0$	$P_{000} + P_{100} = 0,075$	$P_{010} + P_{110} = 0,225$
$z = 1$	$P_{001} + P_{101} = 0,175$	$P_{011} + P_{111} = 0,525$

Cuadro 4: Distribuciones conjuntas de a pares determinadas por las marginales y la covarianzas.

Here we have 12 equations (Table 4) and 7 unknowns. Then, the system of equation we must solve is,

```
library(matlib)

A <- cbind(
# 000 001 010 011 100 101 110 111
c(1, 1, 0, 0, 0, 0, 0, 0), #P(x=0,y=0)
c(0, 0, 0, 0, 1, 1, 0, 0), #P(x=1,y=0)
c(0, 0, 1, 1, 0, 0, 0, 0), #P(x=0,y=1)
```

```

c(0, 0, 0, 0, 0, 0, 1, 1), #P(x=1,y=1)

c(1, 0, 1, 0, 0, 0, 0, 0), #P(x=0,z=0)
c(0, 0, 0, 0, 1, 0, 1, 0), #P(x=1,z=0)
c(0, 1, 0, 1, 0, 0, 0, 0), #P(x=0,z=1)
c(0, 0, 0, 0, 0, 1, 0, 1), #P(x=1,z=1)

c(1, 0, 0, 0, 1, 0, 0, 0), #P(y=0,z=0)
c(0, 0, 1, 0, 0, 0, 1, 0), #P(y=1,z=0)
c(0, 1, 0, 0, 0, 1, 0, 0), #P(y=0,z=1)
c(0, 0, 0, 1, 0, 0, 0, 1) #P(y=1,z=1)
)

example_marginals <- c(
# from: p(x=1) = 0.8, p(y=1) = 0.75, p(y=1|x=1) = 0.9
# X,Y
0.17, #P(x=0,y=0)
0.08, #P(x=1,y=0)
0.03, #P(x=0,y=1)
0.72, #P(x=1,y=1)
# X,Z
0.06, #P(x=0,z=0)
0.24, #P(x=1,z=0)
0.14, #P(x=0,z=1)
0.56, #P(x=1,z=1)
# Y,Z
0.075, #P(y=0,z=0)
0.225, #P(y=1,z=0)
0.175, #P(y=0,z=1)
0.525, #P(y=1,z=1)
)

example_solution = Solve(t(A), example_marginals)
example_solution = c(0.555, -0.385, -0.495, 0.525, -0.48, 0.56, 0.72)

```

Desde donde obtenemos el siguiente sistema de ecuaciones,

$$\begin{aligned}
P_{000} + P_{111} &= 0,555 \\
P_{001} - P_{111} &= -0,385 \\
P_{010} - P_{111} &= -0,495 \\
P_{011} + P_{111} &= 0,525 \\
P_{100} - P_{111} &= -0,480 \\
P_{101} + P_{111} &= 0,560 \\
P_{110} + P_{111} &= 0,720
\end{aligned}$$

El rango de la variable P_{111} depende de la restricción $0 \leq P(x, y, z) \leq 1$

```

joint_distribution <-function(p, rs){
  array_tensor <- array(c(
    rs[1]-p, rs[2]+p, # p000, p001
    rs[3]+p, rs[4]-p, # p010, p011
    rs[5]+p, rs[6]-p, # p100, p101
    rs[7]-p, p        # p110, p111
  ), dim = c(2, 2, 2))

  dimnames(array_tensor) <- list(c("z0", "z1"), c("y0", "y1"), c("x0", "x1"))
  return(array_tensor)
}

joint1 = joint_distribution(0.495,example_solution)
joint2 = joint_distribution(0.510,example_solution)
joint3 = joint_distribution(0.525,example_solution)

```

Desde donde obtenemos las siguientes distribuciones conjuntas

P_{xy0}	x0	x1
y0	0.06	0.015
y1	0	0.225

P_{xy1}	x0	x1
y0	0.11	0.065
y1	0.03	<u>0.495</u>

P_{xy0}	x0	x1
y0	0.045	0.03
y1	0.015	0.21

P_{xy1}	x0	x1
y0	0.125	0.05
y1	0.015	<u>0.51</u>

P_{xy0}	x0	x1
y0	0.03	0.045
y1	0.03	0.195

P_{xy1}	x0	x1
y0	0.14	0.035
y1	0	<u>0.525</u>

Cuando aumentamos el valor de la variable P_{111}

P_{xy0}	x0	x1
y0	↓	↑
y1	↑	↓

P_{xy1}	x0	x1
y0	↑	↓
y1	↓	↑