Sorpresa: el problema de la comunicación con la realidad

Semana 3

La estructura invariante del dato científico. Construcción de sistemas de comunicación con la realidad. Tasa de información. Evaluación de sistemas alternativos por su tasa de sorpresa.

Bibliografía Semana 3

- MacKay. *Information theory, inference and learning algorithms.*. Cambridge university press; 2003 (Descargar). (lecturas 1.1, 2.4-6, 4.1) Lecturas muy cortas. MacKay tiene su curso en Youtube (Ver).

Otras:

- Kelly. *A New Interpretation of Information Rate.*; Bell System Technical Journal. 1956. (Descargar). (lectura paper)

Base empírica

¿Cuál es el dato de la realidad?

¿Cuál es el conjunto de elementos del mundo que sirven de evidencia indubitable (dato)?

Base empírica

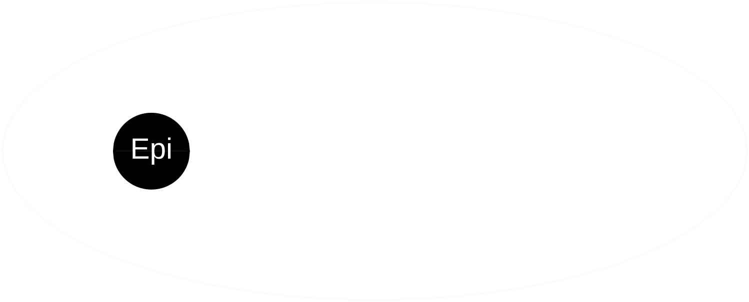
¿Cuál es el dato de la realidad?

- BE Filosófica: $\emptyset$

Base empírica

¿Cuál es el dato de la realidad?

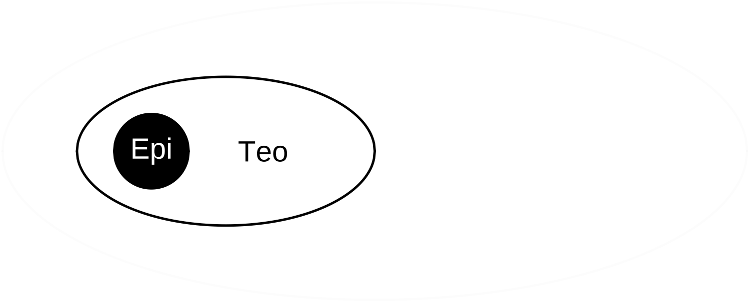
- BE Filosófica: $\emptyset$
- BE Epistemológica: Precios



Base empírica

¿Cuál es el dato de la realidad?

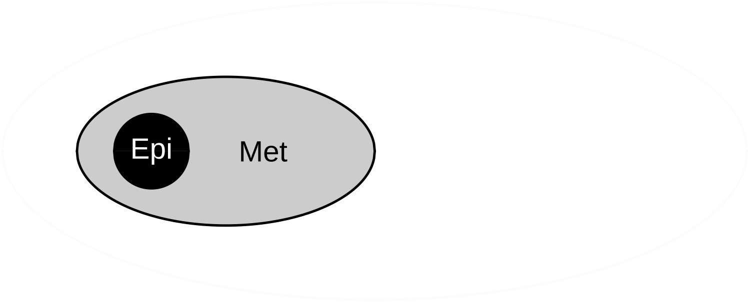
- BE Filosófica: $\emptyset$
- BE Epistemológica: Precios
- Teoría: Inflación



Base empírica

¿Cuál es el dato de la realidad?

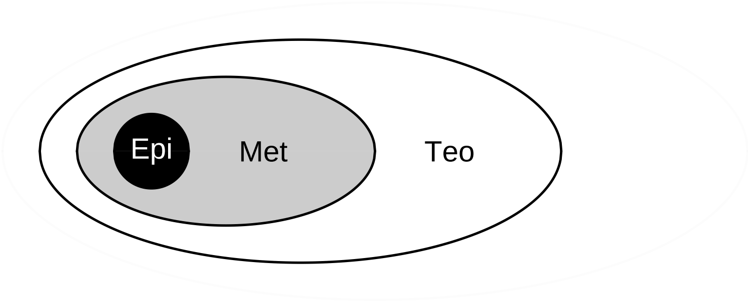
- BE Filosófica: $\emptyset$
- BE Epistemológica: Precios
- BE Metodológica$_1$: Inflación



Base empírica

¿Cuál es el dato de la realidad?

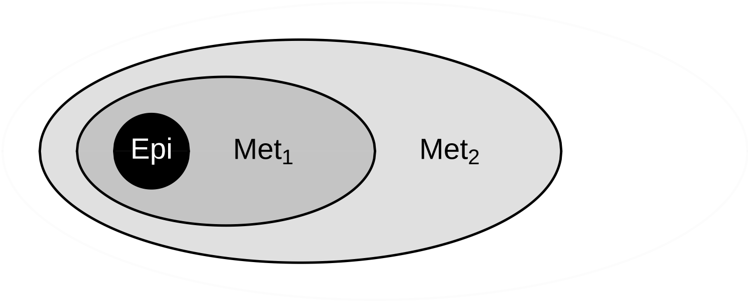
- BE Filosófica: $\emptyset$
- BE Epistemológica: Precios
- BE Metodológica$_1$: Inflación
- Teoría: Producto Bruto Interno



Base empírica

¿Cuál es el dato de la realidad?

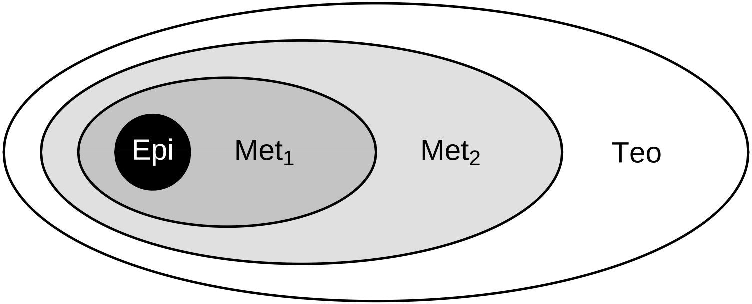
- BE Filosófica: $\emptyset$
- BE Epistemológica: Precios
- BE Metodológica$_1$: Inflación
- BE Metodológica$_2$: Producto Bruto Interno



Base empírica

¿Cuál es el dato de la realidad?

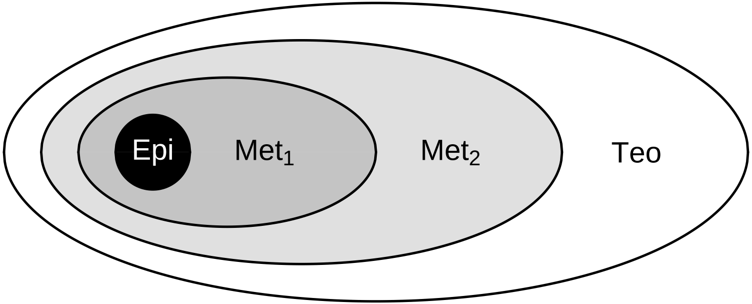
- BE Filosófica: $\emptyset$
- BE Epistemológica: Precios
- BE Metodológica$_1$: Inflación
- BE Metodológica$_2$: Producto Bruto Interno
- Teoría: Política pública



Base empírica

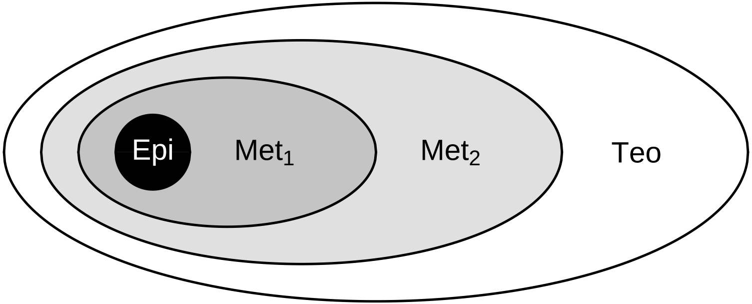
¿Cuál es el dato de la realidad?

Depende del conjunto de supuestos
que una comunidad no pone en duda!



Base empírica

El dato se construye



Los datos como funciones proposicionales

$$f(x) = y$$

- x **Unidad de análisis** (UA)
- f **Variable** de la unidad de análisis (V)
- y **Resultado** o valor de la variable (R)

Los datos como funciones proposicionales

$$f(x) = y$$

x **Unidad de análisis** (UA)

f **Variable** de la unidad de análisis (V)

y **Resultado** o valor de la variable (R)

$$\textit{Altura}(\text{Gustavo}) = 1.77\text{m}$$

Los datos como funciones proposicionales

$$f(x) = y$$

- x **Unidad de análisis** (UA)
- f **Variable** de la unidad de análisis (V)
- y **Resultado** o valor de la variable (R)

Ideología(Partido Comunista) = Comunista

Los datos como funciones proposicionales

$$f(x) = y$$

x **Unidad de análisis** (UA)

f **Variable** de la unidad de análisis (V)

y **Resultado** o valor de la variable (R)

Ideología(Partido Comunista) = Comunista

P(*Ideología*(Partido Comunista) = Comunista) = 0.8

Los datos como funciones proposicionales

$$f(x) = y$$

- x **Unidad de análisis** (UA)
- f **Variable** de la unidad de análisis (V)
- y **Resultado** o valor de la variable (R)

$$\textit{Habilidad}(\text{Maradona}) > \textit{Habilidad}(\text{Messi})$$

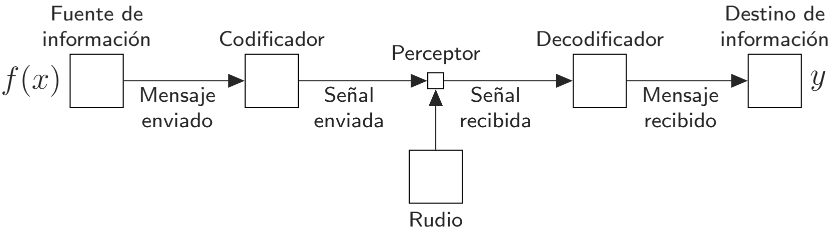
Los datos como funciones proposicionales

$$f(x) = y$$

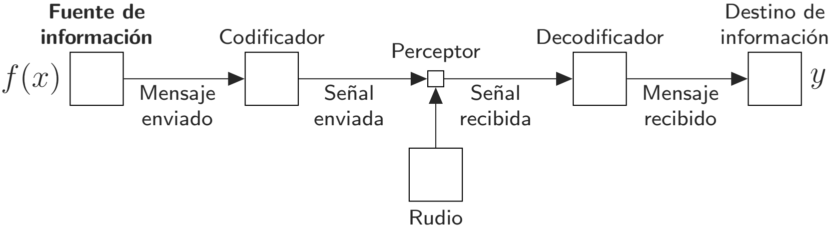
x **Unidad de análisis** (UA)
f **Variable** de la unidad de análisis (V)
y **Resultado** o valor de la variable (R)

El significado preciso de la función depende de la **operacionalización**

La estructura invariante del dato empírico

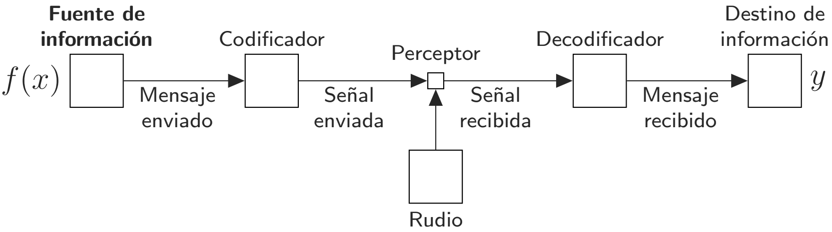


La estructura invariante del dato empírico



- **Fuente**: el estado *real* de la variable

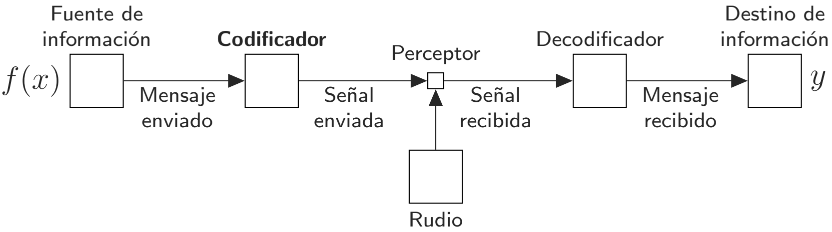
La estructura invariante del dato empírico



- **Fuente**: el estado *real* de la variable *habilidad*(Messi)

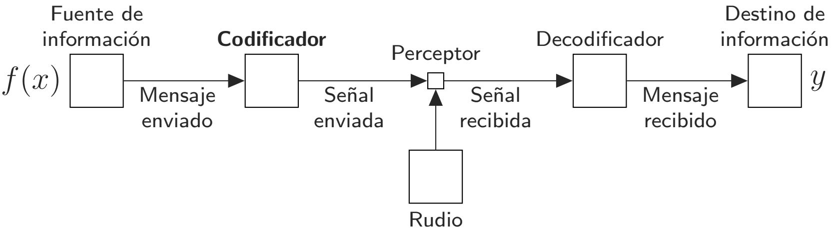
Pregunta de investigación (el problema de conocimiento)

La estructura invariante del dato empírico



- Fuente: el estado *real* de la variable *habilidad*(Messi)
- **Codificador**: epifenómeno

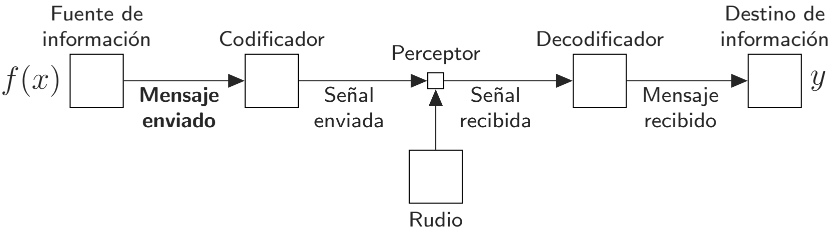
La estructura invariante del dato empírico



- Fuente: el estado *real* de la variable — *habilidad*(Messi)
- **Codificador**: epifenómeno — Ganar/Perder

Examen de representatividad: que la dimensión exprese lo relevante de la variable.

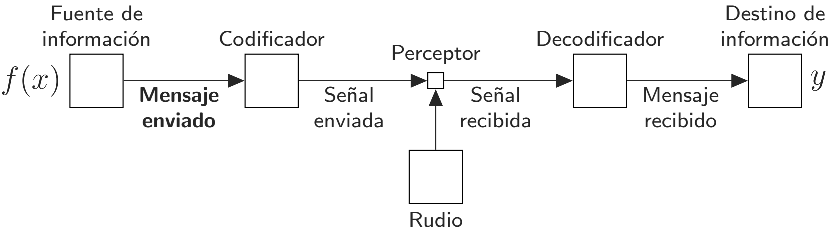
La estructura invariante del dato empírico



- Fuente: el estado *real* de la variable — *habilidad*(Messi)
- Codificador: epifenómeno — Ganar/Perder
- ◦ **Mensaje enviado**

Examen de representatividad: que la dimensión exprese lo relevante de la variable.

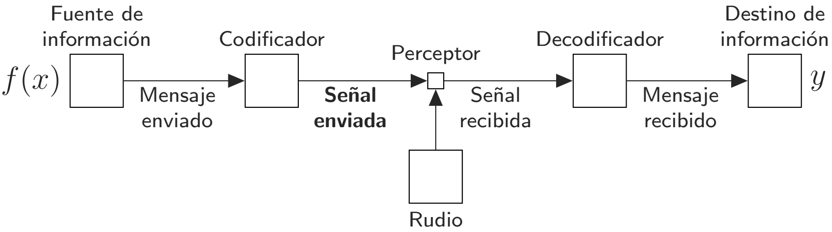
La estructura invariante del dato empírico



- Fuente: el estado *real* de la variable — *habilidad*(Messi)
- Codificador: epifenómeno — Ganar/Perder
- ◦ **Mensaje enviado** — Realidad Causal

Examen de representatividad: que la dimensión exprese lo relevante de la variable.

La estructura invariante del dato empírico

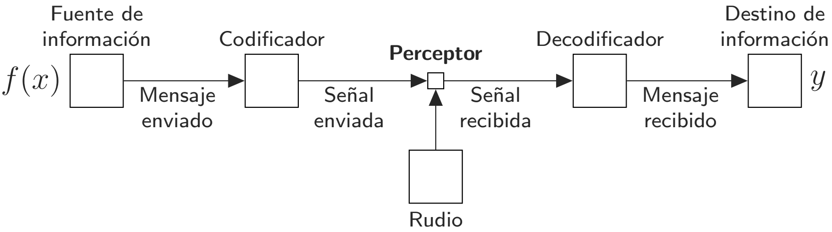


- Fuente: el estado *real* de la variable — *habilidad*(Messi)
- Codificador: epifenómeno — Ganar/Perder

○ Mensaje enviado**, señal enviada** — Realidad Causal

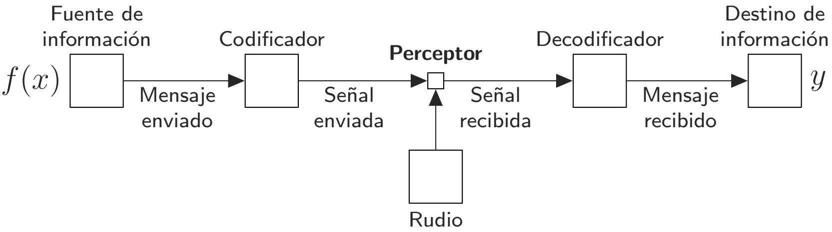
Examen de representatividad: que la dimensión exprese lo relevante de la variable.

La estructura invariante del dato empírico



- Fuente: el estado *real* de la variable — *habilidad*(Messi)
- Codificador: epifenómeno — Ganar/Perder
- ○ Mensaje enviado, señal enviada — Realidad Causal
- **Perceptor**: el instrumento de medición

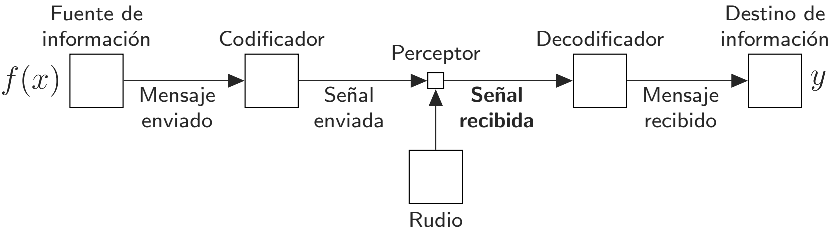
La estructura invariante del dato empírico



- Fuente: el estado *real* de la variable — *habilidad*(Messi)
- Codificador: epifenómeno — Ganar/Perder
- ○ Mensaje enviado, señal enviada — Realidad Causal
- **Perceptor**: el instrumento de medición — *Scraper* de `fifa.com`

Examen de confiabilidad: que el procedimiento detecte fielmente la señal enviada.

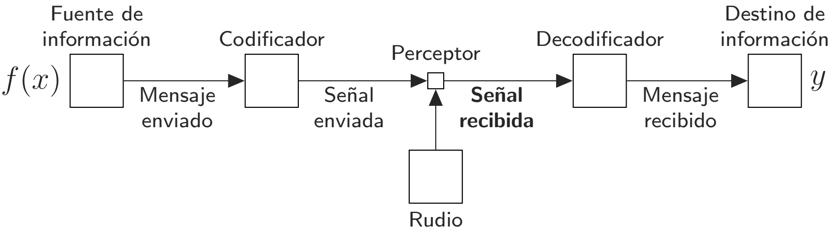
La estructura invariante del dato empírico



- Fuente: el estado *real* de la variable — *habilidad*(Messi)
- Codificador: epifenómeno — Ganar/Perder
  - Mensaje enviado, señal enviada — Realidad Causal
- Perceptor: el instrumento de medición — *Scraper* de `fifa.com`
  - **Señal recibida**: base empírica

Examen de confiabilidad: que el procedimiento detecte fielmente la señal enviada.

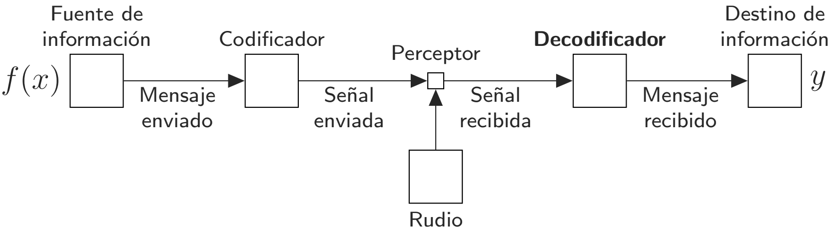
La estructura invariante del dato empírico



- Fuente: el estado *real* de la variable — *habilidad*(Messi)
- Codificador: epifenómeno — Ganar/Perder
- ◦ Mensaje enviado, señal enviada — Realidad Causal
- Perceptor: el instrumento de medición — *Scraper* de `fifa.com`
- ◦ **Señal recibida**: base empírica — True/False

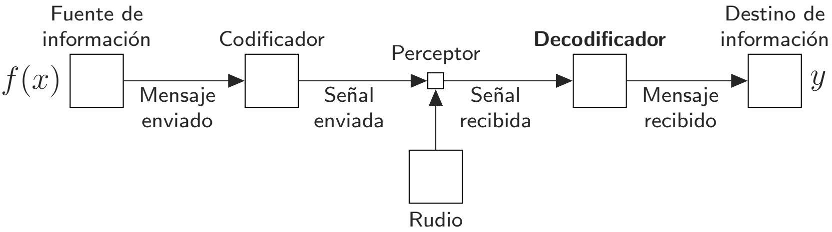
Examen de confiabilidad: que el procedimiento detecte fielmente la señal enviada.

La estructura invariante del dato empírico



• Fuente: el estado *real* de la variable	*habilidad*(Messi)
• Codificador: epifenómeno	Ganar/Perder
○ Mensaje enviado, señal enviada	Realidad Causal
• Perceptor: el instrumento de medición	*Scraper* de `fifa.com`
○ Señal recibida: base empírica	True/False
• **Decodificador**: la interpretación	

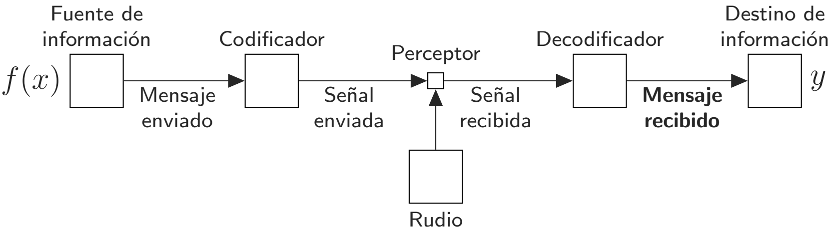
La estructura invariante del dato empírico



- Fuente: el estado *real* de la variable — *habilidad*(Messi)
- Codificador: epifenómeno — Ganar/Perder
- ○ Mensaje enviado, señal enviada — Realidad Causal
- Perceptor: el instrumento de medición — *Scraper* de `fifa.com`
- ○ Señal recibida: base empírica — True/False
- **Decodificador**: la interpretación — Modelo Causal

Hipótesis indicadora: los elementos necesarios para hacer inferencia.

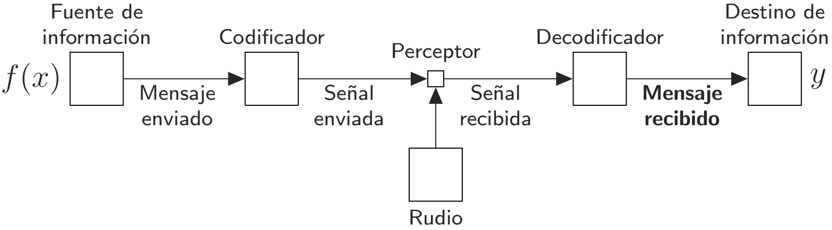
La estructura invariante del dato empírico



- Fuente: el estado *real* de la variable — *habilidad*(Messi)
- Codificador: epifenómeno — Ganar/Perder
- ◦ Mensaje enviado, señal enviada — Realidad Causal
- Perceptor: el instrumento de medición — *Scraper* de `fifa.com`
- ◦ Señal recibida: base empírica — True/False
- Decodificador: la interpretación — Modelo Causal
- ◦ **Mensaje recibido**: la inferencia

Hipótesis indicadora: los elementos necesarios para hacer inferencia.

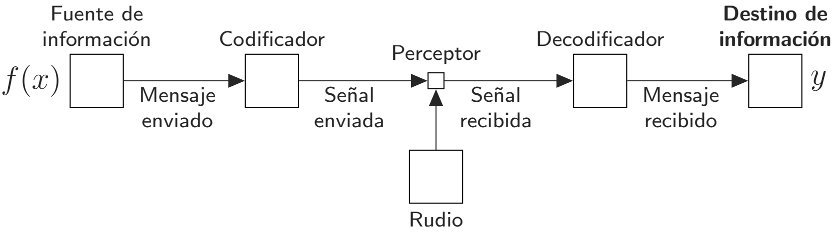
La estructura invariante del dato empírico



• Fuente: el estado *real* de la variable	*habilidad*(Messi)
• Codificador: epifenómeno	Ganar/Perder
○ Mensaje enviado, señal enviada	Realidad Causal
• Perceptor: el instrumento de medición	*Scraper* de `fifa.com`
○ Señal recibida: base empírica	True/False
• Decodificador: la interpretación	Modelo Causal
○ **Mensaje recibido**: la inferencia	Verosimilitud: $P(\text{Indicador}\mid h, \text{M})$

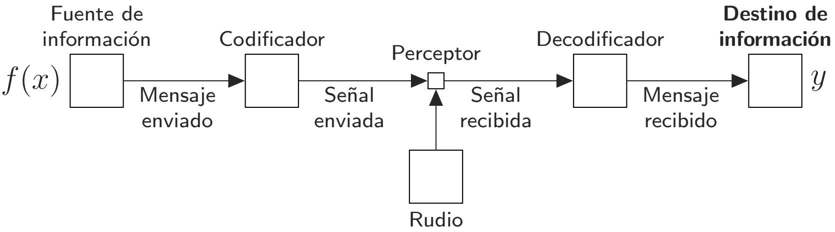
Hipótesis indicadora: los elementos necesarios para hacer inferencia.

La estructura invariante del dato empírico



• Fuente: el estado *real* de la variable	*habilidad*(Messi)
• Codificador: epifenómeno	Ganar/Perder
∘ Mensaje enviado, señal enviada	Realidad Causal
• Perceptor: el instrumento de medición	*Scraper* de `fifa.com`
∘ Señal recibida: base empírica	True/False
• Decodificador: la interpretación	Modelo Causal
∘ Mensaje recibido: la inferencia	Verosimilitud: $P(\text{Indicador}\mid h, \text{M})$
• **Destino**: la estimación	

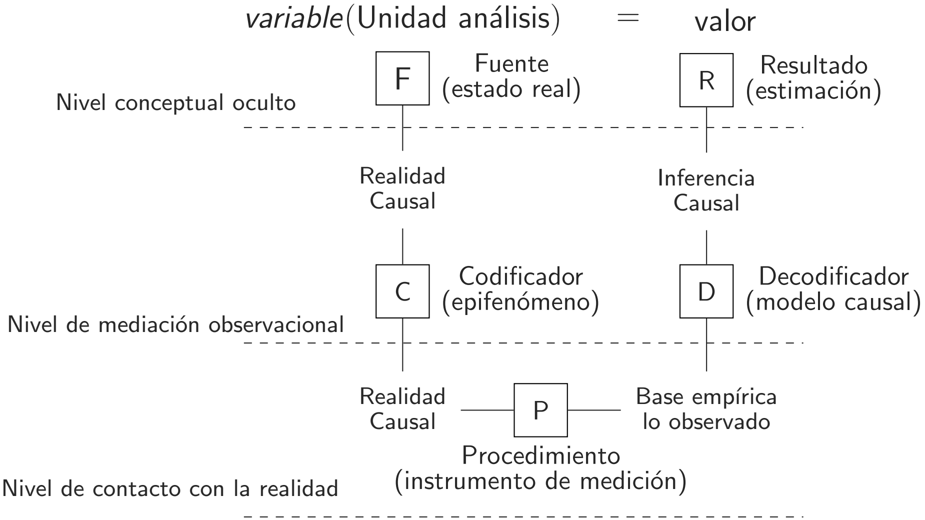
La estructura invariante del dato empírico



• Fuente: el estado *real* de la variable	*habilidad*(Messi)
• Codificador: epifenómeno	Ganar/Perder
◦ Mensaje enviado, señal enviada	Realidad Causal
• Perceptor: el instrumento de medición	*Scraper* de `fifa.com`
◦ Señal recibida: base empírica	True/False
• Decodificador: la interpretación	Modelo Causal
◦ Mensaje recibido: la inferencia	Verosimilitud: $P($Indicador$\|h,$ M$)$
• **Destino**: la estimación	Posterior: $P($*habilidad*(Messi) $= h\|$I, M$)$

Sistema de comunicación

con la realidad



El problema de la comunicación con la realidad

Teoría de la información

El problema de la comunicación con la realidad

Quisiéramos que el mensaje recibido sea igual al enviado

Teoría de la información

El problema de la comunicación con la realidad

Quisiéramos que el mensaje recibido sea igual al enviado

Soluciones

Física: acercarme para escuchar mejor
Inferencia: interpretar la señal con ruido

Teoría de la información

El problema de la comunicación con la realidad

Quisiéramos que el mensaje recibido sea igual al enviado

Soluciones

Física: acercarme para escuchar mejor

Inferencia: interpretar la señal con ruido

Teoría de la información
El problema de la comunicación con la realidad

Quisiéramos que el mensaje recibido sea igual al enviado

Soluciones

Física: acercarme para escuchar mejor

Inferencia: interpretar la señal con ruido

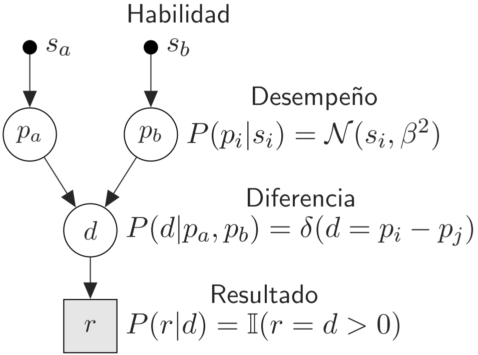
¿Cómo exactamente?

Estimación de habilidad

De la federación internacional de ajedrez

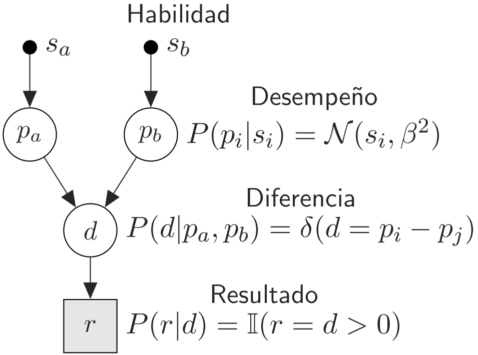
Estimación de habilidad

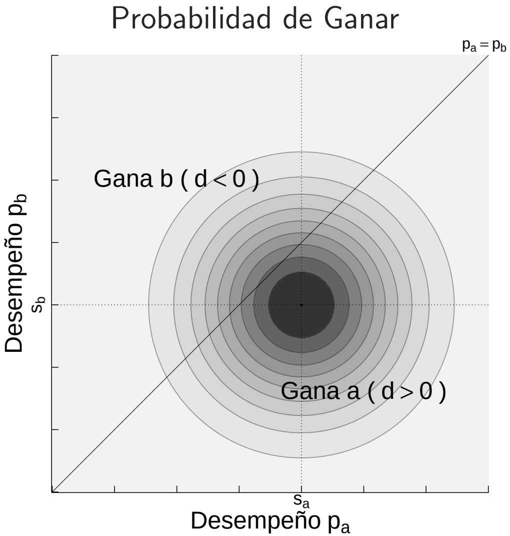
De la federación internacional de ajedrez



Estimación de habilidad

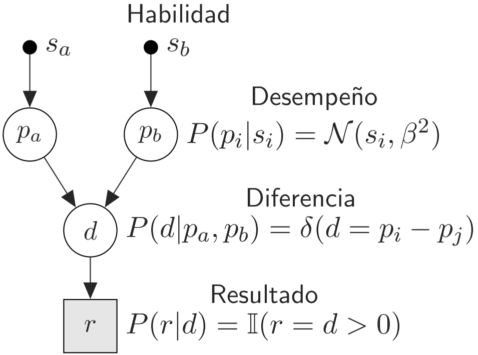
De la federación internacional de ajedrez



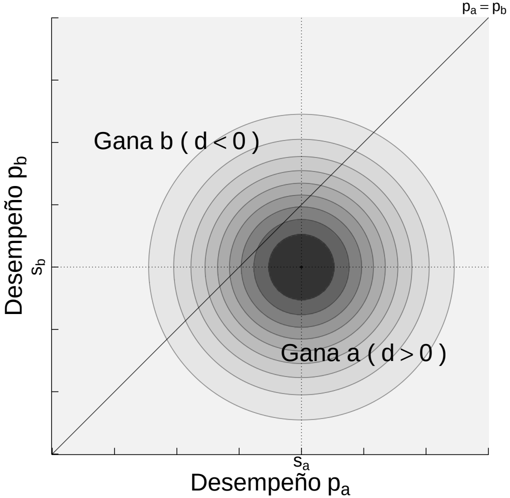


Estimación de habilidad

De la federación internacional de ajedrez



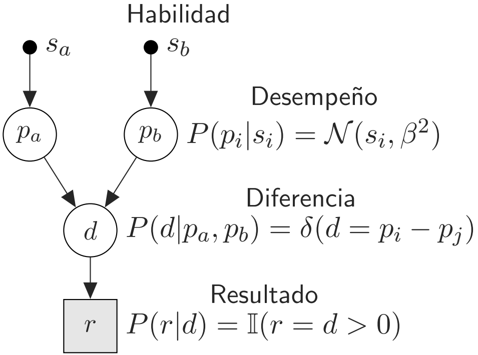
$$P(d > 0 \,|\, s_a, s_b, \mathrm{M})$$

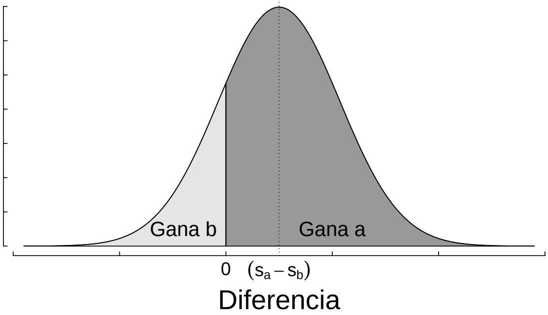


Estimación de habilidad

De la federación internacional de ajedrez

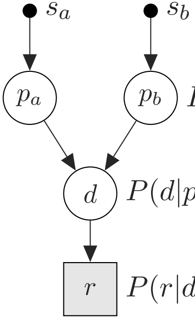
$$P(d > 0 \,|\, s_a, s_b, \mathsf{M}) = 1 - \Phi\left(\frac{s_a - s_b}{\sqrt{2}\beta}\right)$$





Estimación de habilidad

De la federación internacional de ajedrez



Habilidad

Desempeño

$$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$$

Diferencia

$$P(d|p_a, p_b) = \delta(d = p_i - p_j)$$

Resultado

$$P(r|d) = \mathbb{I}(r = d > 0)$$

$$P(d > 0 \,|\, s_a, s_b, \mathsf{M}) = 1 - \Phi\left(\frac{s_a - s_b}{\sqrt{2}\beta}\right)$$

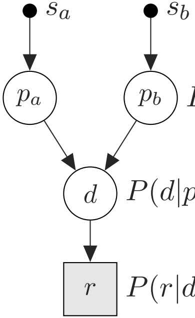
$$s_a^{\mathsf{new}} = s_a^{\mathsf{old}} + \Delta$$

Estimación de habilidad

De la federación internacional de ajedrez

$$P(d > 0 \,|\, s_a, s_b, \mathsf{M}) = 1 - \Phi\left(\frac{s_a - s_b}{\sqrt{2}\beta}\right)$$

Habilidad



$$s_a^{\text{new}} = s_a^{\text{old}} + \Delta$$

Desempeño

$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

$$\Delta = \underbrace{(1 - P(d > 0 \,|\, s_a,\ s_b))}_{\text{Sorpresa}}$$

Diferencia

$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

Resultado

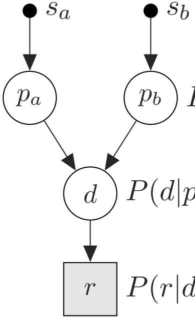
$P(r|d) = \mathbb{I}(r = d > 0)$

Estimación de habilidad

De la federación internacional de ajedrez

$$P(d > 0 \,|\, s_a, s_b, \mathsf{M}) = 1 - \Phi\left(\frac{s_a - s_b}{\sqrt{2}\beta}\right)$$

Habilidad



Desempeño

$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

Diferencia

$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

Resultado

$P(r|d) = \mathbb{I}(r = d > 0)$

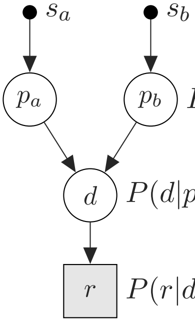
$$s_a^{\mathsf{new}} = s_a^{\mathsf{old}} + \Delta$$

$$\Delta = \underbrace{(1 - P(d > 0 \,|\, s_a,\, s_b))}_{\text{Sorpresa}}\ \underbrace{(-1)^{(1-r)}}_{\text{Signo}}$$

Estimación de habilidad

De la federación internacional de ajedrez

Habilidad



Desempeño

$$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$$

Diferencia

$$P(d|p_a, p_b) = \delta(d = p_i - p_j)$$

Resultado

$$P(r|d) = \mathbb{I}(r = d > 0)$$

$$P(d > 0 \,|\, s_a, s_b, \mathsf{M}) = 1 - \Phi\left(\frac{s_a - s_b}{\sqrt{2}\beta}\right)$$

$$s_a^{\text{new}} = s_a^{\text{old}} + \Delta$$

$$\Delta = \underbrace{(1 - P(d > 0 \,|\, s_a,\ s_b))}_{\text{Sorpresa}}\ \underbrace{(-1)^{(1-r)}}_{\text{Signo}}$$

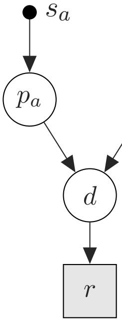
Problemas:

- Actualización simétrica

Estimación de habilidad

De la federación internacional de ajedrez

Habilidad



Desempeño

$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

Diferencia

$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

Resultado

$P(r|d) = \mathbb{I}(r = d > 0)$

$$P(d > 0 \,|\, s_a, s_b, \mathsf{M}) = 1 - \Phi\left(\frac{s_a - s_b}{\sqrt{2}\beta}\right)$$

$$s_a^{\text{new}} = s_a^{\text{old}} + \Delta$$

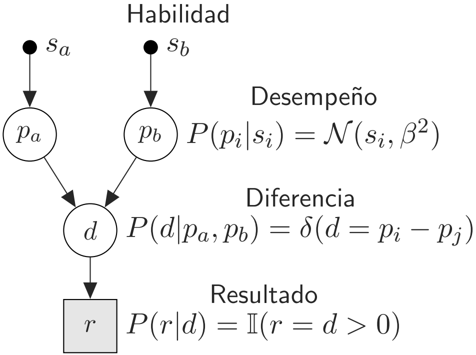
$$\Delta = \underbrace{(1 - P(d > 0 \,|\, s_a,\, s_b))}_{\text{Sorpresa}}\ \underbrace{(-1)^{(1-r)}}_{\text{Signo}}$$

Problemas:

- Actualización simétrica
- Jugadores nuevos desconocidos

Estimación de habilidad

De la federación internacional de ajedrez



$$P(d > 0 \,|\, s_a, s_b, \mathsf{M}) = 1 - \Phi\left(\frac{s_a - s_b}{\sqrt{2}\beta}\right)$$

$$s_a^{\mathsf{new}} = s_a^{\mathsf{old}} + \Delta$$

$$\Delta = \underbrace{(1 - P(d > 0 \,|\, s_a,\ s_b))}_{\text{Sorpresa}}\ \underbrace{(-1)^{(1-r)}}_{\text{Signo}}$$

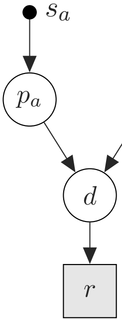
Problemas:

- Actualización simétrica
- Jugadores nuevos desconocidos
- No tiene en cuenta la incertidumbre

Estimación de habilidad

De la federación internacional de ajedrez

Habilidad



Desempeño

$$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$$

Diferencia

$$P(d|p_a, p_b) = \delta(d = p_i - p_j)$$

Resultado

$$P(r|d) = \mathbb{I}(r = d > 0)$$

$$P(d > 0 \,|\, s_a, s_b, \mathsf{M}) = 1 - \Phi\left(\frac{s_a - s_b}{\sqrt{2}\beta}\right)$$

$$s_a^{\text{new}} = s_a^{\text{old}} + \Delta K_a$$

$$\Delta = \underbrace{(1 - P(d > 0 \,|\, s_a,\ s_b))}_{\text{Sorpresa}}\ \underbrace{(-1)^{(1-r)}}_{\text{Signo}}$$

Problemas:

- Actualización simétrica
- Jugadores nuevos desconocidos
- No tiene en cuenta la incertidumbre

Estimación de habilidad

¿Cómo se resuelven todos esos problemas?

Estimación de habilidad

Aplicación estricta de las reglas de la probabilidad

$$P(\text{Hipótesis}|\text{Datos}, \text{Modelo}) = \frac{\overbrace{P(\text{Hipótesis}, \text{Datos}|\text{Modelo})}^{\substack{\text{Creencia compatible} \\ \text{con los datos}}}}{P(\text{Datos}|\text{Modelo})}$$

Estimación de habilidad

Aplicación estricta de las reglas de la probabilidad

$$P(\text{Hipótesis}|\text{Datos},\text{Modelo}) = \frac{\overbrace{P(\text{Hipótesis},\text{Datos}|\text{Modelo})}^{\substack{\text{Creencia compatible}\\ \text{con los datos}}}}{P(\text{Datos}|\text{Modelo})}$$

$$P(\text{Hipótesis},\text{Datos}|\text{Modelo}) = P(\text{Hipótesis}|\text{Modelo})\,P(\text{Datos}|\text{Hipótesis},\text{Modelo})$$

Estimación de habilidad

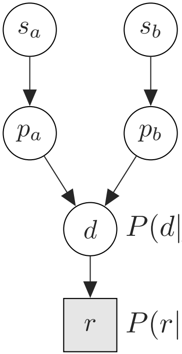
Aplicación estricta de las reglas de la probabilidad

$$P(\text{Hipótesis}|\text{Datos}, \text{Modelo}) = \frac{\overbrace{P(\text{Hipótesis}, \text{Datos}|\text{Modelo})}^{\substack{\text{Creencia compatible} \\ \text{con los datos}}}}{P(\text{Datos}|\text{Modelo})}$$

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

Estimación de habilidad

Siglo 21: Aplicación estricta de las reglas de la probabilidad



Habilidad

$p(s_i) = \mathcal{N}(s_i|\mu_i, \sigma_i^2)$

Desempeño

$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

Diferencia

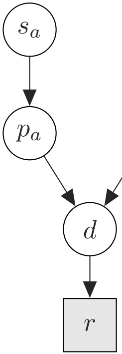
$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

Resultado

$P(r|d) = \mathbb{I}(r = d > 0)$

Estimación de habilidad

Siglo 21: Aplicación estricta de las reglas de la probabilidad



Habilidad

$p(s_i) = \mathcal{N}(s_i|\mu_i, \sigma_i^2)$

Desempeño

$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

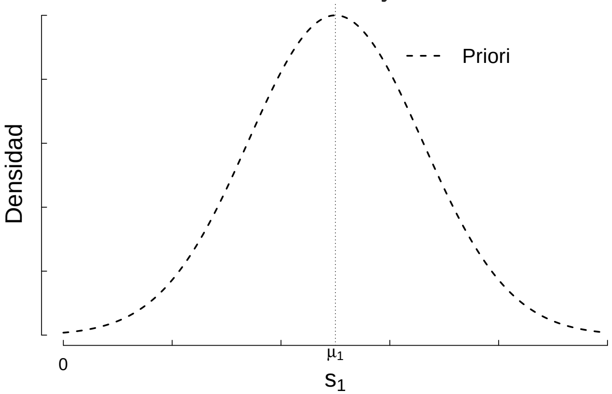
Diferencia

$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

Resultado

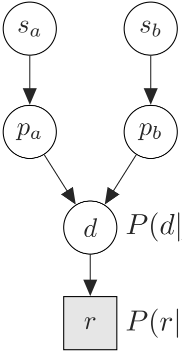
$P(r|d) = \mathbb{I}(r = d > 0)$

Teorema de Bayes.



Estimación de habilidad

Siglo 21: Aplicación estricta de las reglas de la probabilidad



Habilidad

$p(s_i) = \mathcal{N}(s_i|\mu_i, \sigma_i^2)$

Desempeño

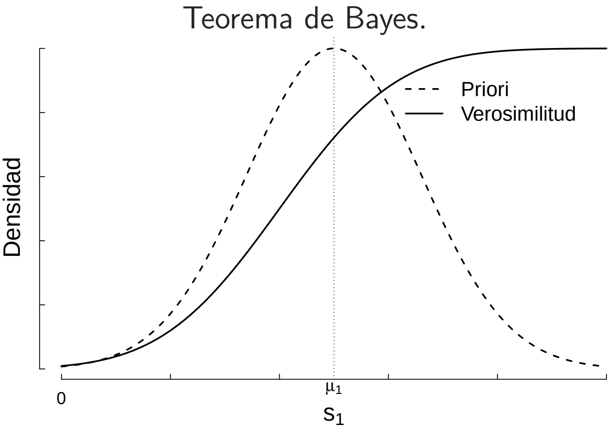
$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

Diferencia

$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

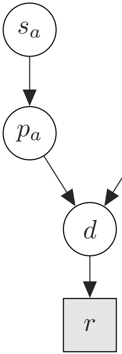
Resultado

$P(r|d) = \mathbb{I}(r = d > 0)$



Estimación de habilidad

Siglo 21: Aplicación estricta de las reglas de la probabilidad



Habilidad

$p(s_i) = \mathcal{N}(s_i|\mu_i, \sigma_i^2)$

Desempeño

$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

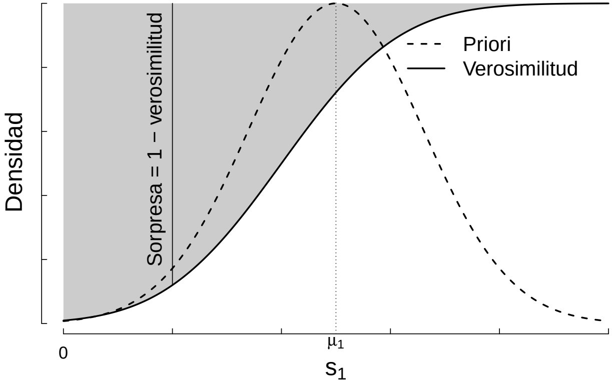
Diferencia

$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

Resultado

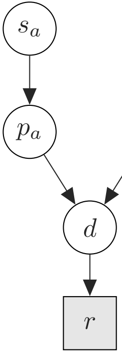
$P(r|d) = \mathbb{I}(r = d > 0)$

Teorema de Bayes.



Estimación de habilidad

Siglo 21: Aplicación estricta de las reglas de la probabilidad



Habilidad

$p(s_i) = \mathcal{N}(s_i|\mu_i, \sigma_i^2)$

Desempeño

$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

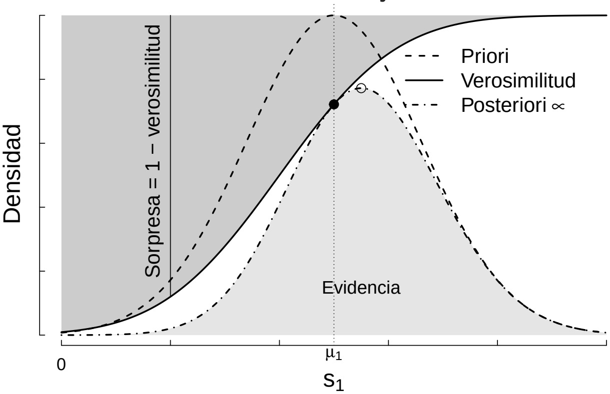
Diferencia

$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

Resultado

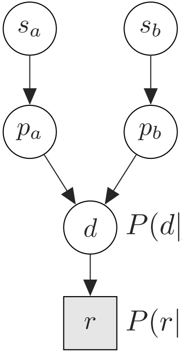
$P(r|d) = \mathbb{I}(r = d > 0)$

Teorema de Bayes.



Estimación de habilidad

Siglo 21: Aplicación estricta de las reglas de la probabilidad



Habilidad

$p(s_i) = \mathcal{N}(s_i|\mu_i, \sigma_i^2)$

Desempeño

$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

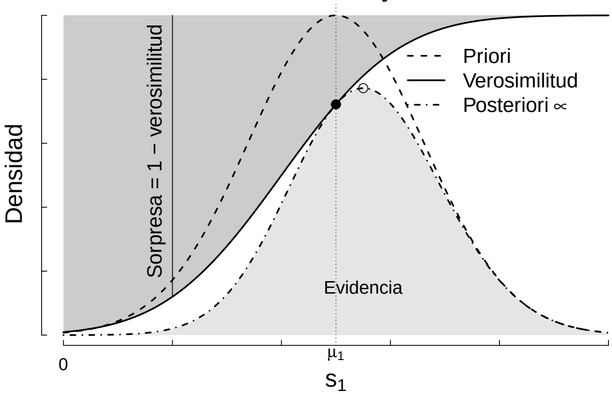
Diferencia

$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

Resultado

$P(r|d) = \mathbb{I}(r = d > 0)$

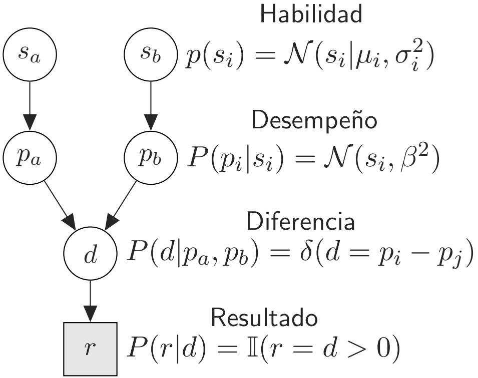
Teorema de Bayes.

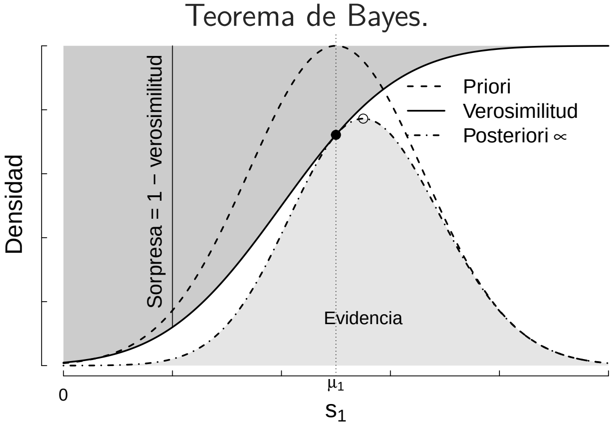


$$P(\text{Gana}|\text{Habilidad} = +\infty) = 1$$

Estimación de habilidad

Siglo 21: Aplicación estricta de las reglas de la probabilidad



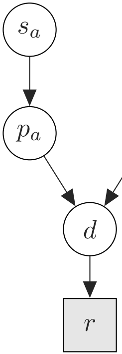


$P(\text{Gana}|\text{Habilidad} = -\infty) = 0$

$P(\text{Gana}|\text{Habilidad} = +\infty) = 1$

Estimación de habilidad

Siglo 21: Aplicación estricta de las reglas de la probabilidad



Habilidad

$p(s_i) = \mathcal{N}(s_i|\mu_i, \sigma_i^2)$

Desempeño

$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

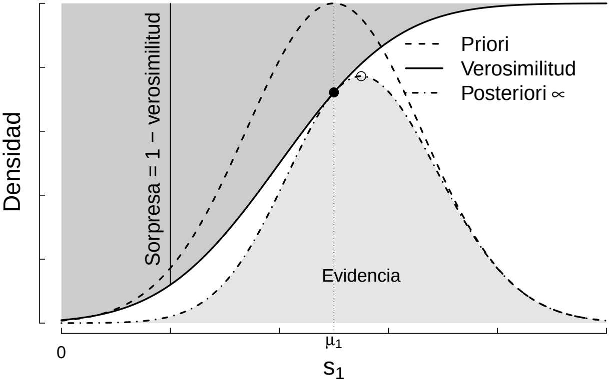
Diferencia

$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

Resultado

$P(r|d) = \mathbb{I}(r = d > 0)$

Teorema de Bayes.



$P(\text{Gana}|\text{Habilidad} = -\infty) =$

$P(\text{Gana}|\text{Habilidad} = +\infty) =$

Sorpresa
Fuente de información

Sorpresa
Fuente de información

Información para las hipótesis.

$$P(\text{Datos}|\text{Hipótesis}_i, \text{Modelo})$$

Sorpresa
Fuente de información

Información para las hipótesis.

$$P(\text{Datos}|\text{Hipótesis}_i, \text{Modelo})$$

Información para los modelos.

$$P(\text{Datos}|\text{Modelo}_i)$$

Sorpresa
Fuente de información

$P(\text{Datos}_T = \{d_1, d_2, \dots, d_T\} | \text{Modelo}_i)$

Sorpresa
Fuente de información

$$P(\mathsf{Datos}_T = \{d_1, d_2, \dots, d_T\} | \mathsf{Modelo}_i) = P(d_1 | \mathsf{Modelo}_i) P(d_2 | d_1, \mathsf{Modelo}) \dots$$

Sorpresa

Fuente de información

$$P(\text{Datos}_T = \{d_1, d_2, \ldots, d_T\} | \text{Modelo}_i) = P(d_1 | \text{Modelo}_i) P(d_2 | d_1, \text{Modelo}) \ldots \approx 0$$

Sorpresa

Fuente de información

$$P(\mathsf{Datos}_T = \{d_1, d_2, \ldots, d_T\}|\mathsf{Modelo}_i) = P(d_1|\mathsf{Modelo}_i)P(d_2|d_1, \mathsf{Modelo})\ldots \approx 0$$

Predicción en órdenes de magnitud

$$\log P(\mathsf{Datos}_T|\mathsf{Modelo}_i) = \log P(d_1|\mathsf{Modelo}_i) + \log P(d_2|d_1, \mathsf{Modelo})\ldots$$

Sorpresa

Fuente de información

$$P(\mathsf{Datos}_T = \{d_1, d_2, \dots, d_T\}|\mathsf{Modelo}_i) = P(d_1|\mathsf{Modelo}_i)P(d_2|d_1, \mathsf{Modelo})\dots \approx 0$$

Predicción en órdenes de magnitud

$$\log P(\mathsf{Datos}_T|\mathsf{Modelo}_i) = \log P(d_1|\mathsf{Modelo}_i) + \log P(d_2|d_1, \mathsf{Modelo})\dots$$

Tasa de predicción

$$\underbrace{P(\mathsf{Datos}_T|\mathsf{Modelo}_i)^{1/T}}_{\text{Media geométrica}}$$

Predicción en órdenes de magnitud

$\log_2 P(\text{Datos})$

Predicción en órdenes de magnitud

$\log_2 P(\text{Datos})$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

Predicción en órdenes de magnitud

$\log_2 P(\text{Datos})$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) $x \bmod 16 \geq 8$ 

Predicción en órdenes de magnitud

$\log_2 P(\text{Datos})$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$ 

2) x mod $8 \geq 4$ 

Predicción en órdenes de magnitud

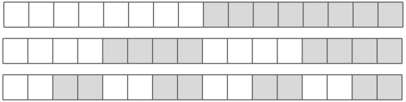
$\log_2 P(\text{Datos})$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$

2) x mod $8 \geq 4$

3) x mod $4 \geq 2$

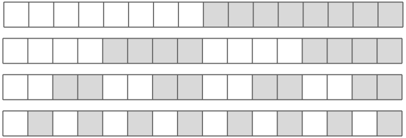


Predicción en órdenes de magnitud

$\log_2 P(\text{Datos})$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$
2) x mod $8 \geq 4$
3) x mod $4 \geq 2$
4) x mod $2 \geq 1$



Predicción en órdenes de magnitud

$$\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\})$$

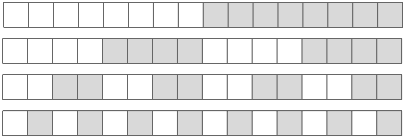
¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$ — $d_1 = 0$

2) x mod $8 \geq 4$ — $d_2 = 1$

3) x mod $4 \geq 2$ — $d_3 = 1$

4) x mod $2 \geq 1$ — $d_4 = 0$



Predicción en órdenes de magnitud

$$\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\})$$

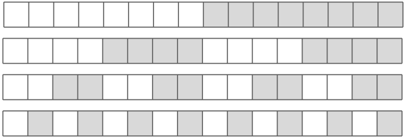
¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$ $\quad d_1 = 0$

2) x mod $8 \geq 4$ $\quad d_2 = 1$

3) x mod $4 \geq 2$ $\quad d_3 = 1$

4) x mod $2 \geq 1$ $\quad d_4 = 0$



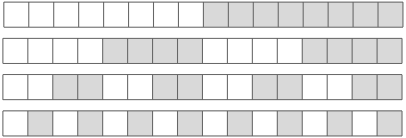
¿Qué número es?

Predicción en órdenes de magnitud

$$\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\})$$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$ — $d_1 = 0$
2) x mod $8 \geq 4$ — $d_2 = 1$
3) x mod $4 \geq 2$ — $d_3 = 1$
4) x mod $2 \geq 1$ — $d_4 = 0$



$$d_1 2^3 + d_2 2^2 + d_3 2^1 + d_4 2^0$$

Predicción en órdenes de magnitud

$$\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\})$$

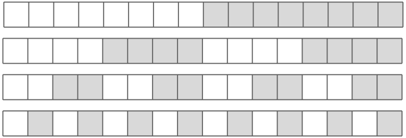
¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$ — $d_1 = 0$

2) x mod $8 \geq 4$ — $d_2 = 1$

3) x mod $4 \geq 2$ — $d_3 = 1$

4) x mod $2 \geq 1$ — $d_4 = 0$



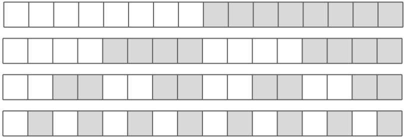
$$d_1 2^3 + d_2 2^2 + d_3 2^1 + d_4 2^0 = 0 + 4 + 2 + 0 = 6$$

Predicción en órdenes de magnitud

$$\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\}) = \sum_i \log_2 P(d_i)$$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$ $\quad d_1 = 0$
2) x mod $8 \geq 4$ $\quad d_2 = 1$
3) x mod $4 \geq 2$ $\quad d_3 = 1$
4) x mod $2 \geq 1$ $\quad d_4 = 0$



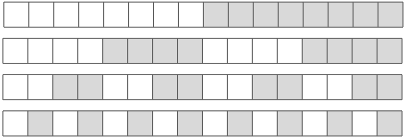
$$d_1 2^3 + d_2 2^2 + d_3 2^1 + d_4 2^0 = 0 + 4 + 2 + 0 = 6$$

Predicción en órdenes de magnitud

$$\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\}) = \sum_i \log_2 P(d_i)$$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$
2) x mod $8 \geq 4$
3) x mod $4 \geq 2$
4) x mod $2 \geq 1$



$d_1 = 0$
$d_2 = 1$
$d_3 = 1$
$d_4 = 0$

$$d_1 2^3 + d_2 2^2 + d_3 2^1 + d_4 2^0 = 0 + 4 + 2 + 0 = 6$$

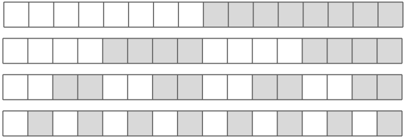
$$\log_2 P(\text{Dato} = (x \text{ mod } N \geq N/2)) = \log_2 \frac{1}{2}$$

Predicción en órdenes de magnitud

$$\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\}) = \sum_i \log_2 P(d_i) = -4$$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$ — $d_1 = 0$
2) x mod $8 \geq 4$ — $d_2 = 1$
3) x mod $4 \geq 2$ — $d_3 = 1$
4) x mod $2 \geq 1$ — $d_4 = 0$



$$d_1 2^3 + d_2 2^2 + d_3 2^1 + d_4 2^0 = 0 + 4 + 2 + 0 = 6$$

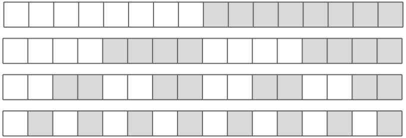
$$\log_2 P(\text{Dato} = (x \text{ mod } N \geq N/2)) = \log_2 \frac{1}{2}$$

Predicción en órdenes de magnitud
- **Información en bits**

$$\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\}) = \sum_i \log_2 P(d_i) = -4$$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$
2) x mod $8 \geq 4$
3) x mod $4 \geq 2$
4) x mod $2 \geq 1$



$d_1 = 0$
$d_2 = 1$
$d_3 = 1$
$d_4 = 0$

$$d_1 2^3 + d_2 2^2 + d_3 2^1 + d_4 2^0 = 0 + 4 + 2 + 0 = 6$$

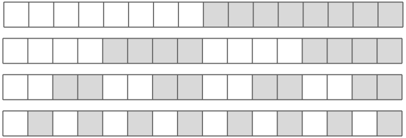
$$\log_2 P(\text{Dato} = (x \text{ mod } N \geq N/2)) = \log_2 \frac{1}{2}$$

Predicción en órdenes de magnitud
- **Información en bits**

$$-\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\}) = \sum_i \log_2 P(d_i) = 4$$

¿Cuál es la menor cantidad de preguntas Sí/No que se necesitan para identificar un entero entre 0 y 15?

1) x mod $16 \geq 8$
2) x mod $8 \geq 4$
3) x mod $4 \geq 2$
4) x mod $2 \geq 1$



$d_1 = 0$
$d_2 = 1$
$d_3 = 1$
$d_4 = 0$

$$d_1 2^3 + d_2 2^2 + d_3 2^1 + d_4 2^0 = 0 + 4 + 2 + 0 = 6$$

$$\log_2 P(\text{Dato} = (x \text{ mod } N \geq N/2)) = \log_2 \frac{1}{2}$$

Predicción en órdenes de magnitud
- **Información en bits**

$$-\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\}) = \sum_i \log_2 P(d_i) = 4$$

Información de Shannon

$$-\log_2 P(\text{Datos}| \dots)$$

Predicción en órdenes de magnitud
- **Información en bits**

$$-\log_2 P(\text{Datos} = \{d_1 = 0, d_2 = 1, d_3 = 1, d_4 = 0\}) = \sum_i \log_2 P(d_i) = 4$$

Información de Shannon

$$-\log_2 P(\text{Datos}|\ldots)$$

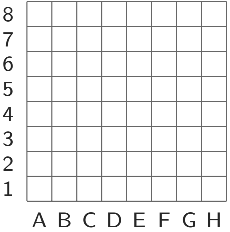
$$=$$

$$\log_2 P(d_1|\ldots) + \log_2 P(d_2|d_1\ldots) + \log_2 P(d_3|d_2, d_1\ldots) + \log_2 P(d_4|d_3, d_2, d_1\ldots)$$

Información en órdenes de magnitud

Información de Shannon

Submarino:

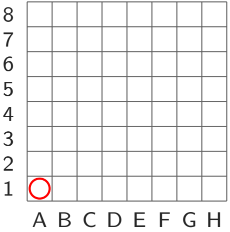


Información en órdenes de magnitud

Información de Shannon

$$P(A1 = 1) = 1/64$$

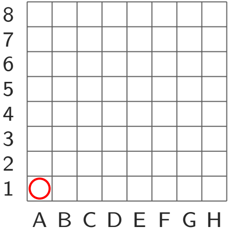
Submarino:



Información en órdenes de magnitud
Información de Shannon

$$-\log_2 P(A1 = 1) = 6 \text{ bits}$$

Submarino:



Información en órdenes de magnitud

Información de Shannon

$$-\log_2 P(A1 = 0) = \log_2 64/63 = 0.0227 \text{ bit}$$

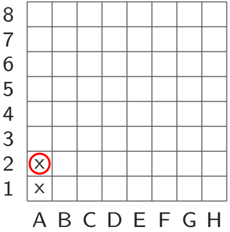
Submarino:

8								
7								
6								
5								
4								
3								
2								
1	x							
	A	B	C	D	E	F	G	H

Información en órdenes de magnitud

Información de Shannon

Submarino:



$$-\log_2 P(A1 = 0) = \log_2 64/63 = 0.0227 \text{ bit}$$

$$-\log_2 P(A2 = 0 | A1 = 0) = \log_2 63/62$$

Información en órdenes de magnitud
Información de Shannon

Submarino:

	A	B	C	D	E	F	G	H
8								
7								
6								
5								
4								
3								
2	x							
1	x							

$$-\log_2 P(A1 = 0) = \log_2 64/63 = 0.0227 \text{ bit}$$

$$-\log_2 P(A2 = 0|A1 = 0) = \log_2 63/62 = 0.0230 \text{ bit}$$

Información en órdenes de magnitud

Información de Shannon

Submarino:

8	×	×	×	×				
7	×	×	×	×				
6	×	×	×	×				
5	×	×	×	×				
4	×	×	×	×				
3	×	×	×	×				
2	×	×	×	×				
1	×	×	×	×				
	A	B	C	D	E	F	G	H

$$-\log_2 P(A1=0) = \log_2 64/63 = 0.0227 \text{ bit}$$

$$-\log_2 P(A2=0|A1=0) = \log_2 63/62 = 0.0230 \text{ bit}$$

$$\log_2 64/63 + \log_2 63/62 + \cdots + \log_2 33/32$$

Información en órdenes de magnitud

Información de Shannon

Submarino:

	A	B	C	D	E	F	G	H
8	×	×	×	×				
7	×	×	×	×				
6	×	×	×	×				
5	×	×	×	×				
4	×	×	×	×				
3	×	×	×	×				
2	×	×	×	×				
1	×	×	×	×				

$$-\log_2 P(A1=0) = \log_2 64/63 = 0.0227 \text{ bit}$$

$$-\log_2 P(A2=0|A1=0) = \log_2 63/62 = 0.0230 \text{ bit}$$

$$\log_2 64/63 + \log_2 63/62 + \cdots + \log_2 33/32$$

$$0.0227 \quad + \quad 0.0230 \quad + \cdots + \quad 0.0430 \quad = 1.0 \text{ bit}$$

Información en órdenes de magnitud
Información de Shannon

Submarino:

	A	B	C	D	E	F	G	H
8	×	×	×	×			×	×
7	×	×	×	×			×	×
6	×	×	×	×			×	×
5	×	×	×	×			×	×
4	×	×	×	×			×	×
3	×	×	×	×			×	×
2	×	×	×	×			×	×
1	×	×	×	×			×	×

$$-\log_2 P(A1=0) = \log_2 64/63 = 0.0227 \text{ bit}$$

$$-\log_2 P(A2=0|A1=0) = \log_2 63/62 = 0.0230 \text{ bit}$$

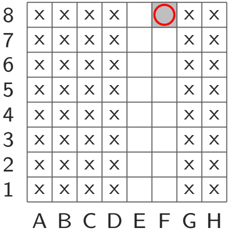
$$\log_2 64/63 + \log_2 63/62 + \cdots + \log_2 33/32$$

$$\log_2 32/31 + \log_2 31/30 + \cdots + \log_2 17/16 = 2.0 \text{ bit}$$

Información en órdenes de magnitud

Información de Shannon

Submarino:



$$-\log_2 P(A1 = 0) = \log_2 64/63 = 0.0227 \text{ bit}$$

$$-\log_2 P(A2 = 0 | A1 = 0) = \log_2 63/62 = 0.0230 \text{ bit}$$

$$\log_2 64/63 + \log_2 63/62 + \cdots + \log_2 33/32$$

$$\log_2 32/31 + \log_2 31/30 + \cdots + \log_2 17/16 + \underbrace{\log_2 16/1}_{4.0 \text{ bit}} = 6.0 \text{ bit}$$

Información en órdenes de magnitud
Información de Shannon

Submarino:

	A	B	C	D	E	F	G	H
8	×	×	×	×			×	×
7	×	×	×	×			×	×
6	×	×	×	×			×	×
5	×	×	×	×			×	×
4	×	×	×	×			×	×
3	×	×	×	×			×	×
2	×	×	×	×			×	×
1	×	×	×	×			×	×

$$-\log_2 P(A1=0) = \log_2 64/63 = 0.0227 \text{ bit}$$

$$-\log_2 P(A2=0|A1=0) = \log_2 63/62 = 0.0230 \text{ bit}$$

$$\log_2 64/63 + \log_2 63/62 + \cdots + \log_2 33/32$$

$$\log_2 32/31 + \log_2 31/30 + \cdots + \log_2 17/16 + \underbrace{\log_2 16/1}_{4.0 \text{ bit}} = 6.0 \text{ bit}$$

$$\log_2 \frac{64}{63} + \log_2 \frac{63}{62} + \cdots + \log_2 \frac{n+1}{n} + \log_2 \frac{n}{1}$$

Información en órdenes de magnitud

Información de Shannon

Submarino:

8	×	×	×	×			×	×
7	×	×	×	×			×	×
6	×	×	×	×			×	×
5	×	×	×	×			×	×
4	×	×	×	×			×	×
3	×	×	×	×			×	×
2	×	×	×	×			×	×
1	×	×	×	×			×	×
	A	B	C	D	E	F	G	H

$$-\log_2 P(A1=0) = \log_2 64/63 = 0.0227 \text{ bit}$$

$$-\log_2 P(A2=0|A1=0) = \log_2 63/62 = 0.0230 \text{ bit}$$

$$\log_2 64/63 + \log_2 63/62 + \cdots + \log_2 33/32$$

$$\log_2 32/31 + \log_2 31/30 + \cdots + \log_2 17/16 + \underbrace{\log_2 16/1}_{4.0 \text{ bit}} = 6.0 \text{ bit}$$

$$\log_2 \frac{64}{\cancel{63}} + \log_2 \frac{\cancel{63}}{62} + \cdots + \log_2 \frac{n+1}{n} + \log_2 \frac{n}{1}$$

Información en órdenes de magnitud

Información de Shannon

Submarino:

	A	B	C	D	E	F	G	H
8	×	×	×	×			×	×
7	×	×	×	×			×	×
6	×	×	×	×			×	×
5	×	×	×	×			×	×
4	×	×	×	×			×	×
3	×	×	×	×			×	×
2	×	×	×	×			×	×
1	×	×	×	×			×	×

$$-\log_2 P(A1=0) = \log_2 64/63 = 0.0227 \text{ bit}$$
$$-\log_2 P(A2=0|A1=0) = \log_2 63/62 = 0.0230 \text{ bit}$$

$$\log_2 64/63 + \log_2 63/62 + \cdots + \log_2 33/32$$
$$\log_2 32/31 + \log_2 31/30 + \cdots + \log_2 17/16 + \underbrace{\log_2 16/1}_{4.0 \text{ bit}} = 6.0 \text{ bit}$$

$$\log_2 \frac{64}{\cancel{63}} + \log_2 \frac{\cancel{63}}{\cancel{62}} + \cdots + \log_2 \frac{\cancel{n+1}}{\cancel{n}} + \log_2 \frac{\cancel{n}}{1}$$

Información en órdenes de magnitud
Información de Shannon

Submarino:

	A	B	C	D	E	F	G	H
8	×	×	×	×			×	×
7	×	×	×	×			×	×
6	×	×	×	×			×	×
5	×	×	×	×			×	×
4	×	×	×	×			×	×
3	×	×	×	×			×	×
2	×	×	×	×			×	×
1	×	×	×	×			×	×

$$-\log_2 P(A1=0) = \log_2 64/63 = 0.0227 \text{ bit}$$
$$-\log_2 P(A2=0|A1=0) = \log_2 63/62 = 0.0230 \text{ bit}$$

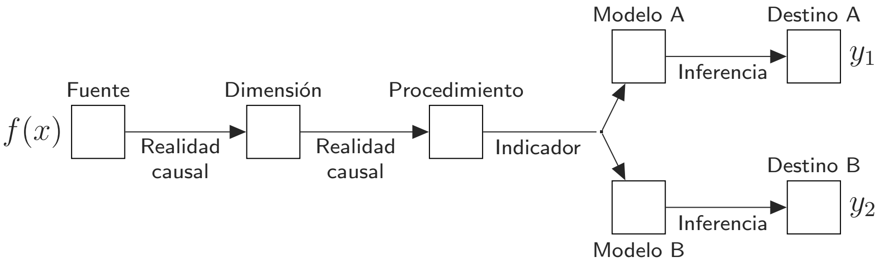
$$\log_2 64/63 + \log_2 63/62 + \cdots + \log_2 33/32$$
$$\log_2 32/31 + \log_2 31/30 + \cdots + \log_2 17/16 + \underbrace{\log_2 16/1}_{4.0 \text{ bit}} = 6.0 \text{ bit}$$

$$\log_2 \frac{64}{\cancel{63}} + \log_2 \frac{\cancel{63}}{\cancel{62}} + \cdots + \log_2 \frac{\cancel{n+1}}{\cancel{n}} + \log_2 \frac{\cancel{n}}{1} = \log_2 \frac{64}{1} = 6.0 \text{ bit}$$

Información en órdenes de magnitud
Información de Shannon

Hasta ahora medimos la información
contenida en observaciones individuales.

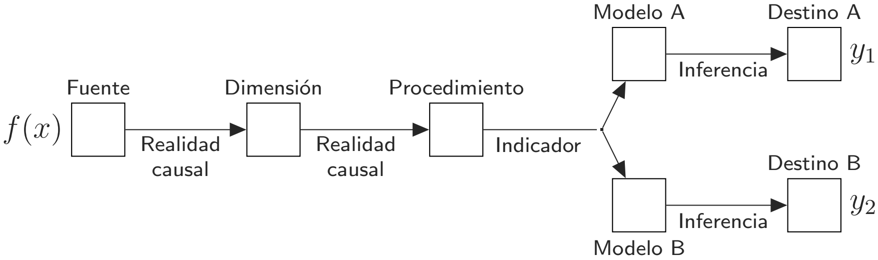
Información en órdenes de magnitud
Información de Shannon



¿Podemos calcular la tasa de información de los sistemas de comunicación alternativos?

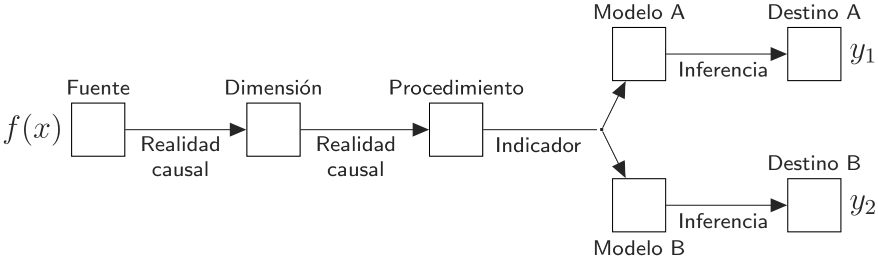
Tasa de información

de los sistemas de comunicación



¿Podemos calcular la tasa de información de los sistemas de comunicación alternativos?

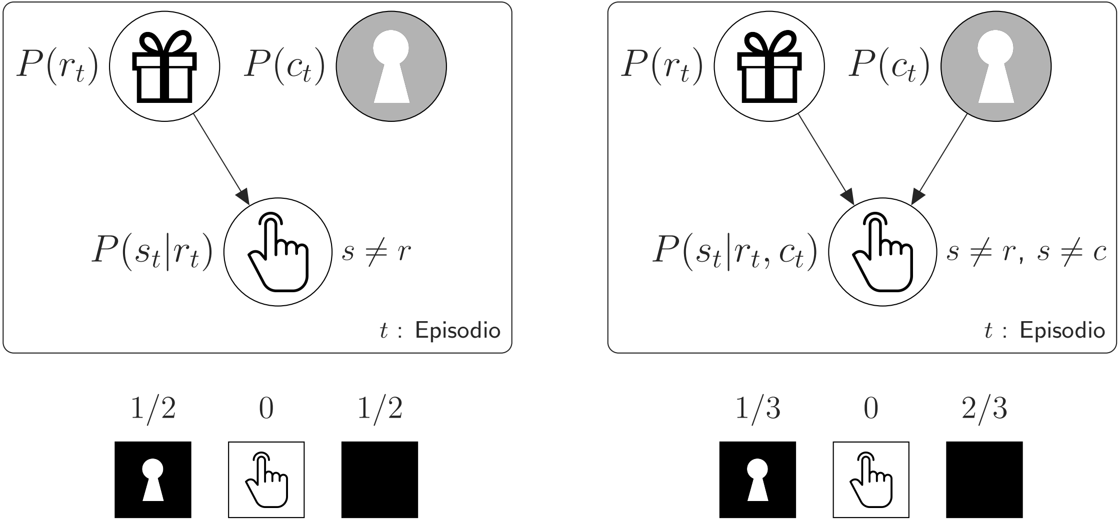
Tasa de información
de los sistemas de comunicación



¿Preferimos sistemas de comunicación con mayor o con menor tasa de información?

Tasa de información

de los sistemas de comunicación



Tasa de información
de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i)\, \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \dots\}|\mathsf{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

Tasa de información

de los sistemas de comunicación

$$P(\textsf{Modelo}_i, \textsf{Datos}_T) = P(\textsf{Modelo}_i)\, \underbrace{P(\textsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \dots\} | \textsf{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$P(\textsf{Datos}_T | \textsf{Modelo} = M) \quad = \prod_{t}^{T} P(c_t, s_t, r_t | M)$$

Tasa de información

de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i)\, \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \ldots\} | \mathsf{Modelo}_i)}_{\text{La } \textbf{información} \text{ está en la } \textbf{sorpresa} \text{ de la predicción}}$$

$$\begin{aligned} P(\mathsf{Datos}_T | \mathsf{Modelo} = M) &= \prod_t^T P(c_t, s_t, r_t | M) \\ &= \prod_{i,j,k} P(c = i, s = j, r = k | M)^{n_{ijk}} \end{aligned}$$

Tasa de información
de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i) \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \dots\} | \mathsf{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$\begin{aligned} P(\mathsf{Datos}_T | \mathsf{Modelo} = M) &= \prod_{t}^{T} P(c_t, s_t, r_t | M) \\ &= \prod_{i,j,k} P(c = i, s = j, r = k | M)^{n_{ijk}} \end{aligned}$$

con n_{ijk} la cantidad de observaciones de $(c = i, s = j, r = k)$.

Tasa de información
de los sistemas de comunicación

$$P(\text{Modelo}_i, \text{Datos}_T) = P(\text{Modelo}_i) \underbrace{P(\text{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \dots\} | \text{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$\begin{aligned} P(\text{Datos}_T | \text{Modelo} = M) &= \prod_{t}^{T} P(c_t, s_t, r_t | M) \\ &= \prod_{i,j,k} P(c = i, s = j, r = k | M)^{n_{ijk}} \\ &= \texttt{tasaDePrediccion}^T \end{aligned}$$

Tasa de información

de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i)\, \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \ldots\}|\mathsf{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$\begin{aligned} P(\mathsf{Datos}_T|\mathsf{Modelo} = M) &= \prod_t^T P(c_t, s_t, r_t|M) \\ &= \prod_{i,j,k} P(c = i, s = j, r = k|M)^{n_{ijk}} \\ &= \mathtt{tasaDePrediccion}^T \end{aligned}$$

la `tasaDePrediccion` es la media geométrica $(P(d_1|M)P(d_2|d_1, M)\ldots)^{1/T}$.

Tasa de información
de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i)\, \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \ldots\} | \mathsf{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$\begin{aligned}
P(\mathsf{Datos}_T | \mathsf{Modelo} = M) &= \prod_t^T P(c_t, s_t, r_t | M) \\
&= \prod_{i,j,k} P(c = i, s = j, r = k | M)^{n_{ijk}}
\end{aligned}$$

$$\underbrace{P(\mathsf{Datos}_T | \mathsf{Modelo} = M)^{\frac{1}{T}}}_{\text{Tasa de predicción}} = \texttt{tasaDePrediccion}$$

Tasa de información

de los sistemas de comunicación

$$P(\text{Modelo}_i, \text{Datos}_T) = P(\text{Modelo}_i)\, \underbrace{P(\text{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \ldots\}|\text{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$\begin{aligned}
P(\text{Datos}_T|\text{Modelo} = M) &= \prod_t^T P(c_t, s_t, r_t|M) \\
&= \prod_{i,j,k} P(c = i, s = j, r = k|M)^{n_{ijk}} \\
\underbrace{P(\text{Datos}_T|\text{Modelo} = M)^{\frac{1}{T}}}_{\text{Tasa de predicción}} &= \prod_{i,j,k} P(c = i, s = j, r = k|M)^{n_{ijk}/T}
\end{aligned}$$

Tasa de información

de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i)\, \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \dots\} | \mathsf{Modelo}_i)}_{\text{La } \textbf{información} \text{ está en la } \textbf{sorpresa} \text{ de la predicción}}$$

$$\begin{aligned}
P(\mathsf{Datos}_T | \mathsf{Modelo} = M) \quad &= \prod_t^T P(c_t, s_t, r_t | M) \\
&= \prod_{i,j,k} P(c = i, s = j, r = k | M)^{n_{ijk}} \\
\lim_{T\to\infty} \underbrace{P(\mathsf{Datos}_T | \mathsf{Modelo} = M)^{\frac{1}{T}}}_{\text{Tasa de predicción temporal}} &= \prod_{i,j,k} P(c = i, s = j, r = k | M)^{p_{ijk}}
\end{aligned}$$

Tasa de información
de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i)\, \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \ldots\}|\mathsf{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$\begin{aligned} P(\mathsf{Datos}_T|\mathsf{Modelo} = M) &= \prod_t^T P(c_t, s_t, r_t|M) \\ &= \prod_{i,j,k} P(c = i, s = j, r = k|M)^{n_{ijk}} \\ \lim_{T\to\infty} \underbrace{P(\mathsf{Datos}_T|\mathsf{Modelo} = M)^{\frac{1}{T}}}_{\text{Tasa de predicción temporal}} &= \prod_{i,j,k} P(c = i, s = j, r = k|M)^{p_{ijk}} \end{aligned}$$

con $p_{ijk} = P(c = i, s = j, r = k|\mathsf{Realidad\ Causal})$

Tasa de información
de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i)\, \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \dots\}|\mathsf{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$\begin{aligned} P(\mathsf{Datos}_T|\mathsf{Modelo} = M) \quad &= \prod_t^T P(c_t, s_t, r_t|M) \\ &= \prod_{i,j,k} P(c=i, s=j, r=k|M)^{n_{ijk}} \\ \log \lim_{T\to\infty} \underbrace{P(\mathsf{Datos}_T|\mathsf{Modelo} = M)^{\frac{1}{T}}}_{\substack{\text{Tasa de predicción temporal} \\ \text{en órdenes de magnitud}}} &= \log \prod_{i,j,k} P(c=i, s=j, r=k|M)^{p_{ijk}} \end{aligned}$$

con $p_{ijk} = P(c=i, s=j, r=k|\text{Realidad Causal})$

Tasa de información

de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i)\, \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \dots\} | \mathsf{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$\begin{aligned}
P(\mathsf{Datos}_T | \mathsf{Modelo} = M) \quad &= \prod_t^T P(c_t, s_t, r_t | M) \\
&= \prod_{i,j,k} P(c = i, s = j, r = k | M)^{n_{ijk}} \\
\log \lim_{T \to \infty} \underbrace{P(\mathsf{Datos}_T | \mathsf{Modelo} = M)^{\frac{1}{T}}}_{\substack{\text{Tasa de predicción temporal} \\ \text{en órdenes de magnitud}}} &= \sum_{i,j,k} p_{ijk} \log P(c = i, s = j, r = k | M)
\end{aligned}$$

con $p_{ijk} = P(c = i, s = j, r = k | \text{Realidad Causal})$

Tasa de información
de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i)\, \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \ldots\} | \mathsf{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$\begin{aligned} P(\mathsf{Datos}_T | \mathsf{Modelo} = M) &= \prod_t^T P(c_t, s_t, r_t | M) \\ &= \prod_{i,j,k} P(c = i, s = j, r = k | M)^{n_{ijk}} \\ \log \lim_{T\to\infty} \underbrace{P(\mathsf{Datos}_T | \mathsf{Modelo} = M)^{\frac{1}{T}}}_{\substack{\text{Tasa de predicción temporal} \\ \text{en órdenes de magnitud}}} &= \underbrace{\sum_{i,j,k} p_{ijk} \log P(c = i, s = j, r = k | M)}_{-\ \substack{\textbf{Valor esperado de la informa-} \\ \textbf{ción en órdenes de magnitud}}} \end{aligned}$$

Tasa de información

de los sistemas de comunicación

$$P(\mathsf{Modelo}_i, \mathsf{Datos}_T) = P(\mathsf{Modelo}_i)\, \underbrace{P(\mathsf{Datos}_T = \{(c_0, s_0, r_0), (c_1, s_1, r_1), \ldots\} | \mathsf{Modelo}_i)}_{\text{La \textbf{información} está en la \textbf{sorpresa} de la predicción}}$$

$$\begin{aligned}
P(\mathsf{Datos}_T | \mathsf{Modelo} = M) &= \prod_t^T P(c_t, s_t, r_t | M) \\
&= \prod_{i,j,k} P(c = i, s = j, r = k | M)^{n_{ijk}} \\
\log \lim_{T \to \infty} \underbrace{P(\mathsf{Datos}_T | \mathsf{Modelo} = M)^{\frac{1}{T}}}_{\substack{\textbf{Tasa de predicción tempo-} \\ \textbf{al en órdenes de magnitud}}} &= \underbrace{\sum_{i,j,k} p_{ijk} \log P(c = i, s = j, r = k | M)}_{-\ \substack{\textbf{Valor esperado de la informa-} \\ \textbf{ción en órdenes de magnitud}}}
\end{aligned}$$

Tasa de información

de los sistemas de comunicación

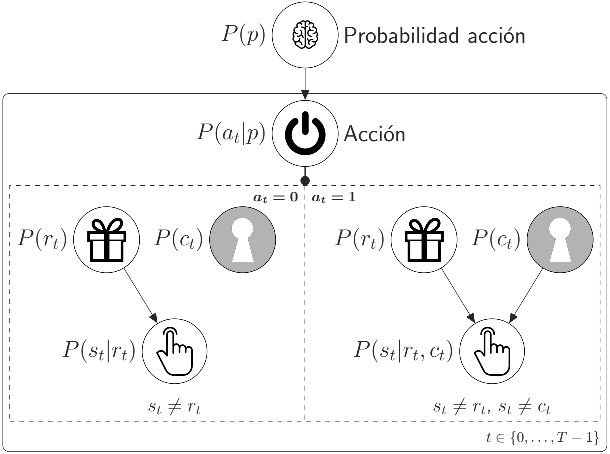
$$\overbrace{-\log \lim_{T\to\infty} P(\text{Datos}_T|\text{Modelo Causal})^{1/T}}^{\substack{\text{Tasa de predicción en}\\ \text{órdenes de magnitud}}}$$

$$=$$

$$\underbrace{\sum_{c,s,r} \underbrace{P(c,s,r|\text{Realidad Causal})}_{\substack{\text{Probabilidad de que}\\ \text{se genere el dato}}} \cdot \underbrace{\big(-\log P(c,s,r|\text{Modelo Causal})\big)}_{\substack{\text{Información en}\\ \text{órdenes de magnitud}}}}_{\substack{\text{Entropía cruzada}\\ \text{Tasa de información}\\ \text{del sistema de comunicación}}}$$

Inteligencia Bayesiana vs Inteligencia Artificial

¿Quién gana?



ChatGPT $\infty.0$

Entropía cruzada

Tasa de información del sistema de comunicación

$$\mathcal{H}(\underset{R}{\text{Realidad}}\text{ causal}, \underset{M}{\text{Modelo}}\text{ causal}) = \sum_{c,s,r} P(c,s,r|R) \cdot \big(-\log P(c,s,r|M) \big)$$

Entropía cruzada

Tasa de información del sistema de comunicación

$$\mathcal{H}(\underset{R}{\text{Realidad}}\text{ causal}, \underset{M}{\text{Modelo}}\text{ causal}) = \sum_{c,s,r} P(c,s,r|R) \cdot \big(-\log P(c,s,r|M)\big)$$

- Caso: $P(c,s,r|R) = P(c,s,r|M)$

Entropía cruzada

Tasa de información del sistema de comunicación

$$\mathcal{H}(\underset{R}{\text{Realidad}}\text{ causal}, \underset{M}{\text{Modelo}}\text{ causal}) = \sum_{c,s,r} P(c,s,r|R) \cdot \big(-\log P(c,s,r|M)\big)$$

- Caso: $P(c,s,r|R) = P(c,s,r|M)$

$$\mathcal{H}(R,M) = \mathcal{H}(R) = \mathcal{H}(M) = \underbrace{\sum_{c,s,r} P(c,s,r) \cdot \big(-\log P(c,s,r)\big)}_{\text{Entropía}}$$

Entropía cruzada

Tasa de información del sistema de comunicación

$$\mathcal{H}(\underset{R}{\text{Realidad causal}}, \underset{M}{\text{Modelo causal}}) = \sum_{c,s,r} P(c,s,r|R) \cdot \big(-\log P(c,s,r|M) \big)$$

- Caso: $P(c,s,r|R) = P(c,s,r|M)$

$$\mathcal{H}(R,M) = \mathcal{H}(R) = \mathcal{H}(M) = \underbrace{\sum_{c,s,r} P(c,s,r) \cdot \big(-\log P(c,s,r) \big)}_{\text{Entropía}}$$

- Caso: $P(c,s,r|R) \neq P(c,s,r|M)$

Entropía cruzada

Tasa de información del sistema de comunicación

$$\mathcal{H}(\underset{R}{\text{Realidad}}\text{ causal}, \underset{M}{\text{Modelo}}\text{ causal}) = \sum_{c,s,r} P(c,s,r|R) \cdot \big(-\log P(c,s,r|M) \big)$$

- Caso: $P(c,s,r|R) = P(c,s,r|M)$

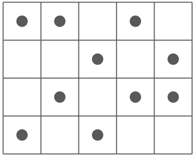
$$\mathcal{H}(R,M) = \mathcal{H}(R) = \mathcal{H}(M) = \underbrace{\sum_{c,s,r} P(c,s,r) \cdot \big(-\log P(c,s,r) \big)}_{\text{Entropía}}$$

- Caso: $P(c,s,r|R) \neq P(c,s,r|M)$

$$\mathcal{H}(R,M) = \mathcal{H}(R) + \underbrace{\sum_{c,s,r} P(c,s,r|R) \cdot \big(-\log \frac{P(c,s,r|M)}{P(c,s,r|R)} \big)}_{\text{Divergencia KL } \mathcal{D}_{KL}(R||M)}$$

Entropía

En la física estadística

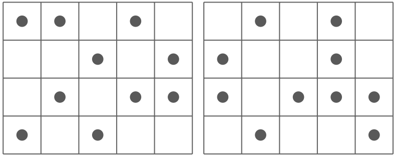


$$N_A = 10$$
$$V_A = 20$$

$$W_A = \binom{20}{10}$$

Entropía

En la física estadística



$N_A = 10$
$V_A = 20$

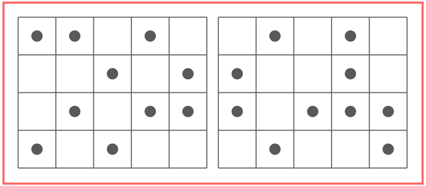
$N_B = 10$
$V_B = 20$

$$W_A = \binom{20}{10}$$

$$W_B = \binom{20}{10}$$

Entropía

En la física estadística



$N_A = 10$
$V_A = 20$

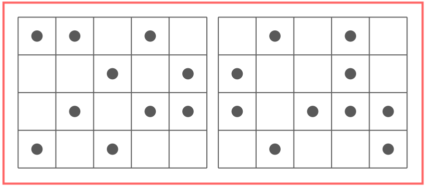
$N_B = 10$
$V_B = 20$

$$W_A = \binom{20}{10}$$

$$W_B = \binom{20}{10}$$

Entropía

En la física estadística



$$N_A = 10$$
$$V_A = 20$$

$$N_B = 10$$
$$V_B = 20$$

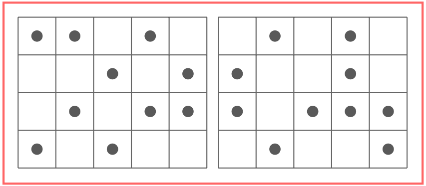
$$W_A = \binom{20}{10}$$

$$W_B = \binom{20}{10}$$

$$N_{AB} = N_A + N_B$$
$$V_{AB} = V_A + V_B$$
$$\boldsymbol{W_{AB} = W_A \quad W_B}$$

Entropía

En la física estadística



$$N_{AB} = N_A + N_B$$
$$V_{AB} = V_A + V_B$$
$$\log \boldsymbol{W_{AB}} = \boldsymbol{W_A} + \boldsymbol{W_B}$$

$N_A = 10$
$V_A = 20$

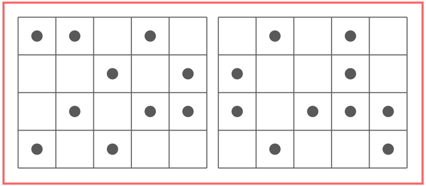
$N_B = 10$
$V_B = 20$

$$W_A = \binom{20}{10}$$

$$W_B = \binom{20}{10}$$

Entropía

En la física estadística



$$N_{AB} = N_A + N_B$$
$$V_{AB} = V_A + V_B$$
$$\log \boldsymbol{W_{AB}} = \boldsymbol{W_A} + \boldsymbol{W_B}$$

$N_A = 10$
$V_A = 20$

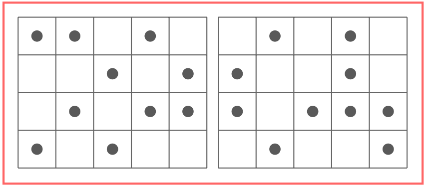
$N_B = 10$
$V_B = 20$

$$P(\text{contraiga}|N = 20, V = 40) = \frac{2}{W_{AB}} \approx 0$$

$$W_A = \binom{20}{10} \qquad W_B = \binom{20}{10}$$

Entropía

En la física estadística



$$N_{AB} = N_A + N_B$$
$$V_{AB} = V_A + V_B$$
$$\log \boldsymbol{W_{AB}} = \boldsymbol{W_A} + \boldsymbol{W_B}$$

$N_A = 10$
$V_A = 20$

$N_B = 10$
$V_B = 20$

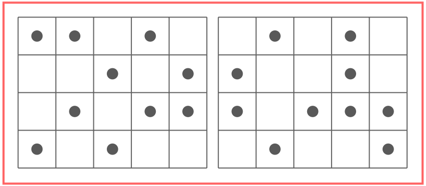
$$W_A = \binom{20}{10}$$

$$W_B = \binom{20}{10}$$

Entropía $\propto - \cdot \log \boldsymbol{W_{AB}}$

Entropía

En la física estadística



$N_A = 10$
$V_A = 20$

$$W_A = \binom{20}{10}$$

$N_B = 10$
$V_B = 20$

$$W_B = \binom{20}{10}$$

$$N_{AB} = N_A + N_B$$
$$V_{AB} = V_A + V_B$$
$$\log \boldsymbol{W_{AB}} = \boldsymbol{W_A} + \boldsymbol{W_B}$$

$$\underbrace{\text{Entropía} \propto - \cdot \log \boldsymbol{W_{AB}}}_{\text{Cantidad de posibles estados en órdenes de magnitud}}$$

Entropía
En la física estadística

Entropía
En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

e_i	n_i
6	
5	
4	
3	
2	
1	
0	

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

e_i	n_i
6	
5	
4	
3	
2	
1	6
0	

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

e_i	n_i	
6		
5		
4		
3		
2		1
1	6	4
0		1

Entropía
En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

e_i	n_i		
6			
5			
4			
3			
2		1	2
1	6	4	2
0		1	2

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

e_i	n_i			
6				
5				
4				
3				1
2		1	2	
1	6	4	2	3
0		1	2	2

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

e_i	n_i										
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

e_i	n_i										
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

$$W(n_0, n_1, n_2, n_3, n_4, n_5, n_6) = \frac{N!}{\prod_i n_i!}$$

e_i	n_i										
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

$$W(n_0, n_1, n_2, n_3, n_4, n_5, n_6) = \frac{N!}{\prod_i n_i!}$$

$$W(0, 6, 0, 0, 0, 0, 0) = 1$$

e_i	n_i										
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

e_i n_i											
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

$$W(n_0, n_1, n_2, n_3, n_4, n_5, n_6) = \frac{N!}{\prod_i n_i!}$$

$$W(0, 6, 0, 0, 0, 0, 0) = 1$$

$$W(1, 4, 1, 0, 0, 0, 0) = 30$$

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

$$W(n_0, n_1, n_2, n_3, n_4, n_5, n_6) = \frac{N!}{\prod_i n_i!}$$

$$W(0,6,0,0,0,0,0) = 1$$
$$W(1,4,1,0,0,0,0) = 30$$
$$W(2,2,2,0,0,0,0) = 90$$

e_i \ n_i											
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

e_i \ n_i											
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

$$W(n_0, n_1, n_2, n_3, n_4, n_5, n_6) = \frac{N!}{\prod_i n_i!}$$

$$W(0,6,0,0,0,0,0) = 1$$

$$W(1,4,1,0,0,0,0) = 30$$

$$W(2,2,2,0,0,0,0) = 90$$

$$\dots$$

$$W(5,0,0,0,0,0,1) = 6$$

Entropía

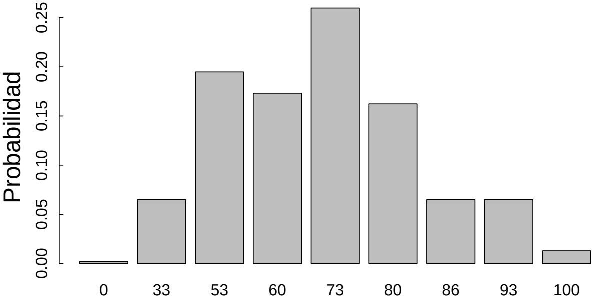
En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

$$W(n_0, n_1, n_2, n_3, n_4, n_5, n_6) = \frac{N!}{\prod_i n_i!}$$

e_i \ n_i											
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100



Entropía

En la física estadística

$$S = \log W = \log \binom{N}{n_1, n_2, \ldots} =$$

Entropía

En la física estadística

$$S = \log W = \log \binom{N}{n_1, n_2, \ldots} = \log \frac{N!}{\prod_i n_i!}$$

Entropía

En la física estadística

$$S = \log W = \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i!$$

Entropía

En la física estadística

$$S = \log W = \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i!$$

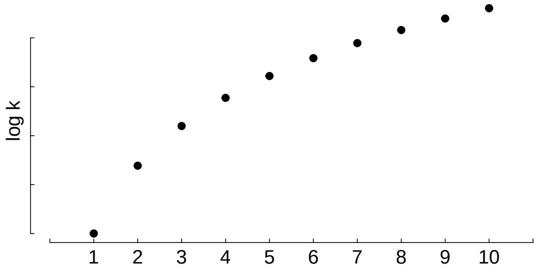
$$\log N! = \sum_{k=1}^{N} \log k$$

Entropía

En la física estadística

$$S = \log W = \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i!$$

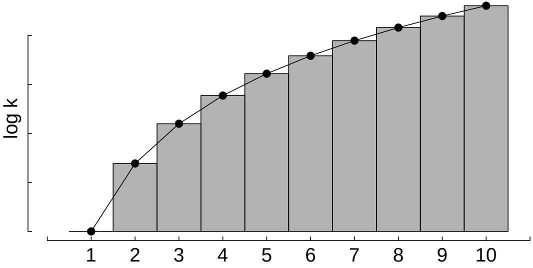
$$\log N! = \sum_{k=1}^{N} \log k \approx \int_1^N \log k \; dk$$



Entropía
En la física estadística

$$S = \log W = \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i!$$

$$\log N! = \sum_{k=1}^{N} \log k \approx \int_1^N \log k \; dk$$
$$= N \log N - N \cancel{+1}$$



Entropía

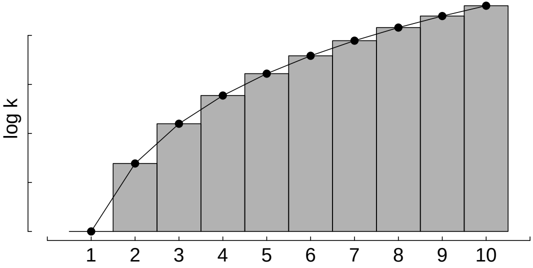
En la física estadística

$$S = \log W = \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i!$$

$$= N \log N - N - \big(\sum_i n_i \log n_i - n_i\big)$$

$$\log N! = \sum_{k=1}^{N} \log k \approx \int_1^N \log k \; dk$$

$$= N \log N - N \cancel{+1}$$



Entropía

En la física estadística

$$S = \log W = \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i!$$

$$= N \log N - N - \big(\sum_i n_i \log n_i - n_i \big)$$

Entropía

En la física estadística

$$S = \log W = \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i!$$

$$= N \log N - N - \sum_i n_i \log n_i - \sum_i n_i$$

Entropía

En la física estadística

$$S = \log W = \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i!$$

$$= N \log N - \cancel{N} - \sum_i n_i \log n_i + \cancel{\sum_i n_i}$$

Entropía

En la física estadística

$$
\begin{aligned}
S = \log W &= \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i! \\
&= N \log N - \cancel{N} - \sum_i n_i \log n_i + \cancel{\sum_i n_i} \\
&= (\sum_i n_i) \log N - \sum_i n_i \log n_i
\end{aligned}
$$

Entropía

En la física estadística

$$
\begin{aligned}
S = \log W &= \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i! \\
&= N \log N - \cancel{N} - \sum_i n_i \log n_i + \cancel{\sum_i n_i} \\
&= (\sum_i n_i) \log N - \sum_i n_i \log n_i \\
&= \sum_i n_i (\log N - \log n_i)
\end{aligned}
$$

Entropía

En la física estadística

$$
\begin{aligned}
S = \log W &= \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i! \\
&= N \log N - \cancel{N} - \sum_i n_i \log n_i + \cancel{\sum_i n_i} \\
&= (\sum_i n_i) \log N - \sum_i n_i \log n_i \\
&= \sum_i n_i \log \frac{N}{n_i}
\end{aligned}
$$

Entropía

En la física estadística

$$
\begin{aligned}
S = \log W &= \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i! \\
&= N \log N - \cancel{N} - \sum_i n_i \log n_i + \cancel{\sum_i n_i} \\
&= (\sum_i n_i) \log N - \sum_i n_i \log n_i \\
&= -\sum_i n_i \log \frac{n_i}{N}
\end{aligned}
$$

Entropía

En la física estadística

$$
\begin{aligned}
S = \log W &= \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i! \\
&= N \log N - \cancel{N} - \sum_i n_i \log n_i + \cancel{\sum_i n_i} \\
&= (\sum_i n_i) \log N - \sum_i n_i \log n_i \\
&= -\sum_i n_i \log \frac{n_i}{N} \\
\frac{S}{N} &= -\sum_i \frac{n_i}{N} \log \frac{n_i}{N}
\end{aligned}
$$

Entropía: promedio de la cantidad de posibles estados en órdenes de magnitud

Entropía

En la física estadística

$$
\begin{aligned}
S = \log W &= \log \binom{N}{n_1, n_2, \ldots} = \log N! - \sum_i \log n_i! \\
&= N \log N - \cancel{N} - \sum_i n_i \log n_i + \cancel{\sum_i n_i} \\
&= (\sum_i n_i) \log N - \sum_i n_i \log n_i \\
&= -\sum_i n_i \log \frac{n_i}{N} \\
\frac{S}{N} &= -\sum_i p_i \log p_i
\end{aligned}
$$

Entropía: promedio de la cantidad de posibles estados en órdenes de magnitud

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

$$\mathcal{H}(W(n_0, n_1, n_2, n_3, n_4, n_5, n_6)) = -\sum \frac{n_i}{N} \log \frac{n_i}{N}$$

e_i	n_i										
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

$$\mathcal{H}(W(n_0, n_1, n_2, n_3, n_4, n_5, n_6)) = -\sum \frac{n_i}{N} \log \frac{n_i}{N}$$

$$\mathcal{H}(W(0, 6, 0, 0, 0, 0, 0)) = 0$$

e_i	n_i										
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

$$\mathcal{H}(W(n_0, n_1, n_2, n_3, n_4, n_5, n_6)) = -\sum \frac{n_i}{N} \log \frac{n_i}{N}$$

e_i	n_i										
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

$$\mathcal{H}(W(0, 6, 0, 0, 0, 0, 0)) = 0$$

$$\mathcal{H}(W(1, 4, 1, 0, 0, 0, 0)) \approx 0.867$$

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

$$\mathcal{H}(W(n_0, n_1, n_2, n_3, n_4, n_5, n_6)) = -\sum \frac{n_i}{N} \log \frac{n_i}{N}$$

e_i	n_i										
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

$$\mathcal{H}(W(0, 6, 0, 0, 0, 0, 0)) = 0$$

$$\mathcal{H}(W(1, 4, 1, 0, 0, 0, 0)) \approx 0.867$$

$$\mathcal{H}(W(2, 2, 2, 0, 0, 0, 0)) \approx 1.099$$

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

$$\mathcal{H}(W(n_0, n_1, n_2, n_3, n_4, n_5, n_6)) = -\sum \frac{n_i}{N} \log \frac{n_i}{N}$$

e_i \ n_i											
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

$$\mathcal{H}(W(0, 6, 0, 0, 0, 0, 0)) = 0$$

$$\mathcal{H}(W(1, 4, 1, 0, 0, 0, 0)) \approx 0.867$$

$$\mathcal{H}(W(2, 2, 2, 0, 0, 0, 0)) \approx 1.099$$

$$\cdots$$

$$\mathcal{H}(W(5, 0, 0, 0, 0, 0, 1)) \approx 0.451$$

Entropía

En la física estadística

Restricciones:

- Partículas totales $N = 6$
- Energía total $E = \sum_i e_i \cdot n_i = 6$

$$\mathcal{H}(W(n_0, n_1, n_2, n_3, n_4, n_5, n_6)) = -\sum \frac{n_i}{N} \log \frac{n_i}{N}$$

e_i \ n_i											
6										1	1
5									1		
4							1				
3				1		1		2			
2		1	2		3	1			1		
1	6	4	2	3		1	2			1	
0		1	2	2	3	3	3	4	4	4	5
Gini:	0	33	53	60	60	73	80	80	86	93	100

$$\mathcal{H}(W(0,6,0,0,0,0,0)) = 0$$
$$\mathcal{H}(W(1,4,1,0,0,0,0)) \approx 0.867$$
$$\mathcal{H}(W(2,2,2,0,0,0,0)) \approx 1.099$$
$$\dots$$
$$\mathcal{H}(W(5,0,0,0,0,0,1)) \approx 0.451$$

Entropía cómo medida de la incertidumbre

Maximizar entropía

$$\underset{a=\text{Acción}}{\arg\,\mathbf{max}}\ \mathcal{H}(\text{Realidad}(a)) = \underset{a=\text{Acción}}{\arg\,\mathbf{min}}\ P(\text{Datos}_\infty|\text{Realidad}(a))$$

Maximizar entropía

$$\underset{a=\text{Acción}}{\arg\,\mathbf{max}}\ \mathcal{H}(\text{Realidad}(a)) = \underset{a=\text{Acción}}{\arg\,\mathbf{min}}\ P(\text{Datos}_\infty|\text{Realidad}(a))$$

Maximizar entropía (o información esperada)

es

Maximizar incertidumbre (minimizar la tasa de predicción)

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

Seleccionar el plan de acción a que use la balanza la menor cantidad de veces para determinar cuál es la pelota y si pesa más o menos que el resto.

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

$P($Balanza$\|$Acción$)$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}($Acción$)$
6 vs 6 (0)				
5 vs 5 (2)				
4 vs 4 (4)				
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

$P(\text{Balanza}\|\text{Acción})$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}(\text{Acción})$
6 vs 6 (0)		0		
5 vs 5 (2)				
4 vs 4 (4)				
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

P(Balanza|Acción)	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}$(Acción)
6 vs 6 (0)	$1/2$	0	$1/2$	
5 vs 5 (2)				
4 vs 4 (4)				
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

$P(\text{Balanza}\vert\text{Acción})$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}(\text{Acción})$
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)				
4 vs 4 (4)				
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

$P(\text{Balanza}\|\text{Acción})$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}(\text{Acción})$
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)		¿ ?		
4 vs 4 (4)				
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

P(Balanza|Acción)	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}$(Acción)
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)		$2/12$		
4 vs 4 (4)				
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

$P(\text{Balanza}\|\text{Acción})$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}(\text{Acción})$
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)		$1/6$		
4 vs 4 (4)				
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

P(Balanza|Acción)	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}$(Acción)
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)	$5/12$	$1/6$	$5/12$	
4 vs 4 (4)				
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

$P(\text{Balanza}\|\text{Acción})$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}(\text{Acción})$
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)	$5/12$	$1/6$	$5/12$	1.48 bits
4 vs 4 (4)				
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

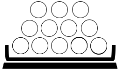
Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

P(Balanza|Acción)	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}$(Acción)
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)	$5/12$	$1/6$	$5/12$	1.48 bits
4 vs 4 (4)	$4/12$	$4/12$	$4/12$	
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

P(Balanza|Acción)	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}$(Acción)
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)	$5/12$	$1/6$	$5/12$	1.48 bits
4 vs 4 (4)	$1/3$	$1/3$	$1/3$	
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

$P(\text{Balanza}\|\text{Acción})$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}(\text{Acción})$
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)	$5/12$	$1/6$	$5/12$	1.48 bits
4 vs 4 (4)	$1/3$	$1/3$	$1/3$	1.58 bits
3 vs 3 (6)				
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

$P(\text{Balanza}\|\text{Acción})$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}(\text{Acción})$
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)	$5/12$	$1/6$	$5/12$	1.48 bits
4 vs 4 (4)	$1/3$	$1/3$	$1/3$	1.58 bits
3 vs 3 (6)	$1/4$	$1/2$	$1/4$	
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

$P(\text{Balanza}\vert\text{Acción})$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}(\text{Acción})$
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)	$5/12$	$1/6$	$5/12$	1.48 bits
4 vs 4 (4)	$1/3$	$1/3$	$1/3$	1.58 bits
3 vs 3 (6)	$1/4$	$1/2$	$1/4$	1.5 bits
2 vs 2 (8)				
1 vs 1 (10)				

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

$P(\text{Balanza}\|\text{Acción})$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}(\text{Acción})$
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)	$5/12$	$1/6$	$5/12$	1.48 bits
4 vs 4 (4)	$1/3$	$1/3$	$1/3$	1.58 bits
3 vs 3 (6)	$1/4$	$1/2$	$1/4$	1.5 bits
2 vs 2 (8)	$1/6$	2/3	$1/6$	1.25 bits
1 vs 1 (10)	1/12	5/6	1/12	0.82 bits

Maximizar entropía

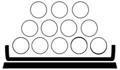
Maximizar incertidumbre es maximizar la información esperada.

Una de las 12 pelotas tiene un peso distinto al resto.

P(Balanza|Acción)	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\mathcal{H}$(Acción)
6 vs 6 (0)	$1/2$	0	$1/2$	1 bit
5 vs 5 (2)	$5/12$	$1/6$	$5/12$	1.48 bits
4 vs 4 (4)	$1/3$	$1/3$	$1/3$	1.58 bits
3 vs 3 (6)	$1/4$	$1/2$	$1/4$	1.5 bits
2 vs 2 (8)	$1/6$	2/3	$1/6$	1.25 bits
1 vs 1 (10)	$1/12$	5/6	$1/12$	0.82 bits

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.



	1^+	
$(1,2,3,4)$	2^+	$\uparrow$:
	$\dots$	
	12^+	
4 vs 4 (4)		$\leftrightarrow$:
	1^-	
	2^-	
$(5,6,7,8)$	$\dots$	$\downarrow$:
	12^-	

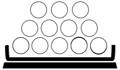
Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

	1^+		
$(1,2,3,4)$	2^+	$\uparrow$:	?
	$\dots$		
	12^+		
4 vs 4 (4)		$\leftrightarrow$:	?
	1^-		
	2^-		
$(5,6,7,8)$	$\dots$	$\downarrow$:	?
	12^-		

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.



	1^+									
$(1,2,3,4)$	2^+	$\uparrow$:	1^+	2^+	3^+	4^+	5^-	6^-	7^-	8^-
	$\dots$									
	12^+									
4 vs 4 (4)		$\leftrightarrow$:	?							
	1^-									
	2^-									
$(5,6,7,8)$	$\dots$	$\downarrow$:	?							
	12^-									

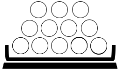
Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

	1^+									
$(1,2,3,4)$	2^+	$\uparrow$:	1^+	2^+	3^+	4^+	5^-	6^-	7^-	8^-
	$\dots$									
	12^+									
4 vs 4 (4)		$\leftrightarrow$:	9^+	10^+	11^+	12^+	9^-	10^-	11^-	12^-
	1^-									
	2^-									
$(5,6,7,8)$	$\dots$	$\downarrow$:	?							
	12^-									

Maximizar entropía

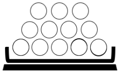
Maximizar incertidumbre es maximizar la información esperada.



$(1,2,3,4)$	1^+, 2^+, $\dots$	$\uparrow$:	1^+	2^+	3^+	4^+	5^-	6^-	7^-	8^-
4 vs 4 (4)	12^+, 1^-	$\leftrightarrow$:	9^+	10^+	11^+	12^+	9^-	10^-	11^-	12^-
$(5,6,7,8)$	2^-, $\dots$, 12^-	$\downarrow$:	1^-	2^-	3^-	4^-	5^+	6^+	7^+	8^+

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.



Ejercicio.

Elegir un plan de acción para el siguiente paso.

Maximizar entropía

Maximizar incertidumbre es maximizar la información esperada.

Ejercicio.

Elegir un plan de acción para el siguiente paso.

Si maximizan la información esperada (entropía)
pueden identificar la pelota en 2 pasos más.

Máxima entropía
dada información disponible

La definición formal del principio de “no mentir”.

Máxima entropía
dada información disponible

Distribuciones de creencias
(por máxima entropía dadas restricciones)

$$\text{max} \;\; -\int P(x|\theta) \log P(x|\theta)\, dx$$

$$\text{dado} \;\; \int P(x|\theta)\, dx = 1$$

$$\int \phi_k(x) P(x|\theta)\, dx = F_k \;\; \forall_k$$

Máxima entropía
dada información disponible

Distribuciones de creencias
(por máxima entropía dadas restricciones)

$$\text{max} \;\; -\int P(x|\theta)\log P(x|\theta)\,dx$$

$$\text{dado} \;\; \int P(x|\theta)\,dx = 1$$

$$\int \phi_k(x)P(x|\theta)\,dx = F_k \;\; \forall_k$$

La familia exponencial

$$P(x|\theta) = \frac{1}{Z(\theta)}h(x)\text{exp}(\theta^T\phi(x))$$

Con:

$x \in \mathbb{X}^m$ hipótesis

$\theta \in \mathbb{R}^d$ parámetros naturales

$\phi(x) \in \mathbb{R}^d$ estadísticos suficientes

Máxima entropía
dada información disponible

Distribuciones de creencias
(por máxima entropía dadas restricciones)

$$\text{max} \ -\int P(x|\theta) \log P(x|\theta)\, dx$$

$$\text{dado} \ \int P(x|\theta)\, dx = 1$$

$$\int \phi_k(x) P(x|\theta)\, dx = F_k \quad \forall_k$$

La familia exponencial

$$P(x|\theta) = \frac{1}{Z(\theta)} h(x) \text{exp}(\theta^T \phi(x))$$

- Bernoulli$(x|p)$
- Beta$(x|\alpha, \beta)$
- Geométrica$(x|p)$
- Multinomial$(x|p, n)$
- Hipergeométrica$(x|n, N, K)$
- Poisson$(x|\lambda)$
- Gaussiana$(x|\mu, \sigma^2)$
- Dirichlet$(x|\alpha_1, \dots, \alpha_K)$
- $\dots$

Divergencia Kullback-Leibler

$$-\log \lim_{T \to \infty} P(\text{Datos}_T | \text{Modelo Causal})^{1/T}$$

Tasa de predicción (temporal) en órdenes de magnitud (log) en signo opuesto

Divergencia Kullback-Leibler

$$-\log \lim_{T\to\infty} P(\text{Datos}_T|\text{Modelo Causal})^{1/T}$$

$$=$$

$$\sum_{c,s,r} P(c,s,r|\text{Realidad Causal}) \cdot \big(-\log P(c,s,r|\text{Modelo Causal})\big)$$

Información esperada del sistema de comunicación

Divergencia Kullback-Leibler

$$-\log \lim_{T\to\infty} P(\mathsf{Datos}_T|\mathsf{Modelo\ Causal})^{1/T}$$

$$=$$

$$\sum_{c,s,r} P(c,s,r|\mathsf{Realidad\ Causal}) \cdot \big(-\log P(c,s,r|\mathsf{Modelo\ Causal})\big)$$

$$=$$

$$\mathcal{H}(R) + \underbrace{\sum_{c,s,r} P(c,s,r|R) \cdot \big(-\log \frac{P(c,s,r|M)}{P(c,s,r|R)}\big)}_{\mathsf{Divergencia\ KL}\ \mathcal{D}_{KL}(R||M)}$$

Tasa de información del sistema de comunicación perfecto (entropía)
$+$ Tasa de información debida a la imperfección del sistema (divergencia KL)

Divergencia Kullback-Leibler

$$\sum_{c,s,r} P(c,s,r|R) \cdot \big(-\log \frac{P(c,s,r|M)}{P(c,s,r|R)} \big)$$

Tasa de información debida a la imperfección del sistema (divergencia KL)

Divergencia Kullback-Leibler

$$\sum_{c,s,r} P(c,s,r|R) \cdot \big(\log P(c,s,r|R) - \log P(c,s,r|M) \big)$$

Tasa de información debida a la imperfección del sistema (divergencia KL)

Divergencia Kullback-Leibler

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

Tasa de información debida a la imperfección del sistema (divergencia KL)

Divergencia Kullback-Leibler

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

Tasa de información debida a la imperfección del sistema (divergencia KL)

Divergencia Kullback-Leibler

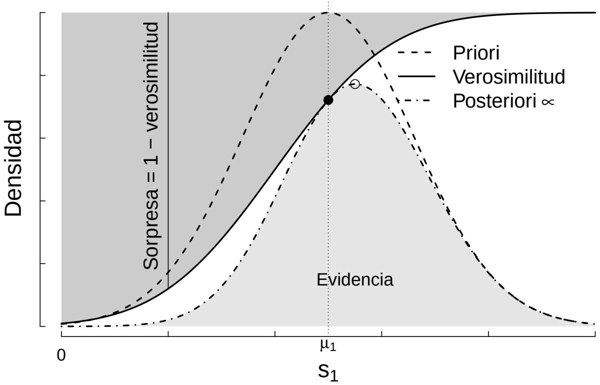
$$\sum_{c,s,r} P(c,s,r|R) \cdot \Big(\underbrace{\log P(c,s,r|R) - \log P(c,s,r|M)}_{\substack{\text{La diferencia de predicción} \\ \text{en órdenes de magnitud}}} \Big)$$

¿Para qué sirve si no conozco la realidad?

Tasa de información debida a la imperfección del sistema (divergencia KL)

TrueSkill

Posterior exacto



TrueSkill

Posterior aproximado

$$\underbrace{\widehat{p}(s_a|\text{Gana}, M)}_{\text{Aproximado}} = \underset{\mu,\sigma}{\text{arg min}}\ \mathcal{D}_{KL}(\ \underbrace{p(s_a|\text{Gana}, M)}_{\text{Exacto}}\ ||\ \underbrace{\mathcal{N}(s_a|\mu, \sigma^2)}_{\text{Familia}}\ )$$

Mínima Divergencia

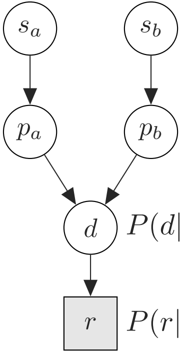
Posterior aproximado

$$\underbrace{\widehat{p}(s_a|\text{Gana}, M)}_{\text{Aproximado}} = \underset{\mu,\sigma}{\text{arg min}}\ \mathcal{D}_{KL}(\ \underbrace{p(s_a|\text{Gana}, M)}_{\text{Exacto}}\ ||\ \underbrace{\mathcal{N}(s_a|\mu, \sigma^2)}_{\text{Familia}}\ )$$

Mínima Divergencia

Posterior aproximado

$$\underbrace{\widehat{p}(s_a|\text{Gana}, M)}_{\text{Aproximado}} = \underset{\mu,\sigma}{\text{arg min}}\ \mathcal{D}_{KL}(\ \underbrace{p(s_a|\text{Gana}, M)}_{\text{Exacto}}\ ||\ \underbrace{\mathcal{N}(s_a|\mu,\sigma^2)}_{\text{Familia}}\ )$$



Habilidad

$p(s_i) = \mathcal{N}(s_i|\mu_i, \sigma_i^2)$

Desempeño

$P(p_i|s_i) = \mathcal{N}(s_i, \beta^2)$

Diferencia

$P(d|p_a, p_b) = \delta(d = p_i - p_j)$

Resultado

$P(r|d) = \mathbb{I}(r = d > 0)$

Mínima Divergencia

Posterior aproximado

$$\underbrace{\widehat{p}(s_a|\text{Gana}, M)}_{\text{Aproximado}} = \underset{\mu,\sigma}{\text{arg min}}\ \mathcal{D}_{KL}(\ \underbrace{p(s_a|\text{Gana}, M)}_{\text{Exacto}}\ ||\ \underbrace{\mathcal{N}(s_a|\mu, \sigma^2)}_{\text{Familia}}\ )$$

$$\underbrace{p(d, \text{Gana} \mid \text{Modelo})}_{\text{Marginal exacta}} = \int_{s_a}\int_{p_a}\int_{s_b}\int_{p_b} p(s_a)p(s_b)p(p_a|s_a)p(p_b|s_b)p(d|p_a, p_b)\ P(\text{Gana}|d)$$

Mínima Divergencia

Posterior aproximado

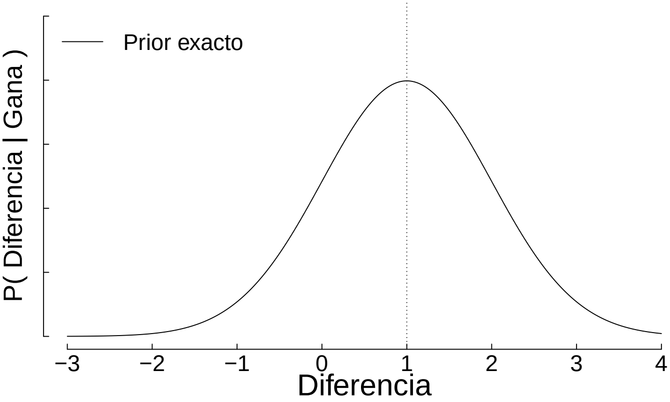
$$\underbrace{\widehat{p}(s_a|\text{Gana}, M)}_{\text{Aproximado}} = \underset{\mu,\sigma}{\text{arg min}}\ \mathcal{D}_{KL}(\ \underbrace{p(s_a|\text{Gana}, M)}_{\text{Exacto}}\ ||\ \underbrace{\mathcal{N}(s_a|\mu,\sigma^2)}_{\text{Familia}}\ )$$

$$\underbrace{p(d, \text{Gana} \mid \text{Modelo}) =}_{\text{Marginal exacta}} \underbrace{\int_{s_a}\int_{p_a}\int_{s_b}\int_{p_b} p(s_a)p(s_b)p(p_a|s_a)p(p_b|s_b)p(d|p_a,p_b)}_{\text{Prior exacto } p(d) = \mathcal{N}(d|\ \underbrace{\mu_a - \mu_b}_{\text{Diferencia de medias}},\ \underbrace{\sigma_a^2 + \sigma_b^2 + 2\beta^2}_{\text{Suma de incertidumbres}})}\ P(\text{Gana}|d)$$

Mínima Divergencia

Posterior aproximado

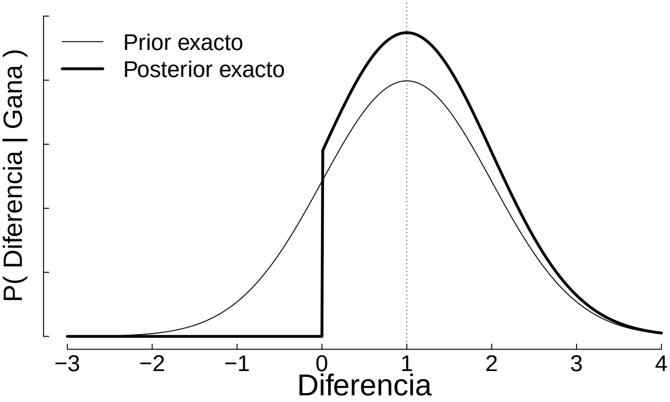
$$p(d, \text{Gana}) = \mathcal{N}(d|\mu_a - \mu_b, \sigma_a^2 + \sigma_b^2 + 2\beta^2)\mathbb{I}(d > 0)$$



Mínima Divergencia

Posterior aproximado

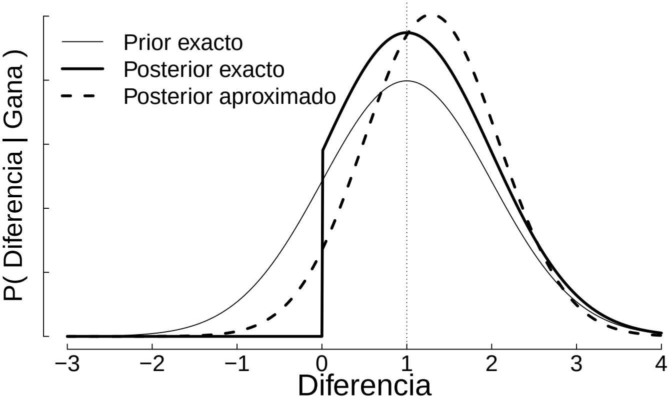
$$p(d, \text{Gana}) = \mathcal{N}(d|\mu_a - \mu_b, \sigma_a^2 + \sigma_b^2 + 2\beta^2)\mathbb{I}(d > 0)$$



Mínima Divergencia

Posterior aproximado

$$p(d, \text{Gana}) = \mathcal{N}(d|\mu_a - \mu_b, \sigma_a^2 + \sigma_b^2 + 2\beta^2)\mathbb{I}(d > 0)$$



Mínima Divergencia
Posterior aproximado

$$p(d, \mathsf{Gana}) = \mathcal{N}(d|\mu_a - \mu_b, \sigma_a^2 + \sigma_b^2 + 2\beta^2)\mathbb{I}(d > 0)$$

$$\widehat{p}(d, \mathsf{Gana}) \propto \mathcal{N}\Big(d \mid \underbrace{E(d, \mathsf{Gana}|d > 0)}_{\substack{\text{Misma} \\ \text{media}}}, \underbrace{V(d, \mathsf{Gana}|d > 0)}_{\substack{\text{Misma} \\ \text{varianza}}} \Big)$$

Mínima Divergencia
Posterior aproximado

$$p(d, \mathsf{Gana}) = \mathcal{N}(d|\mu_a - \mu_b, \sigma_a^2 + \sigma_b^2 + 2\beta^2)\mathbb{I}(d > 0)$$

$$\widehat{p}(d, \mathsf{Gana}) \propto \mathcal{N}\Big(d \mid \underbrace{E(d, \mathsf{Gana}|d > 0)}_{\text{Misma media}}, \underbrace{V(d, \mathsf{Gana}|d > 0)}_{\text{Misma varianza}}\Big)$$

La divergencia es mínima cuando tienen mismos momentos.
(Expectation Propagation)

Inferencia aproximada

Inferencia aproximada

Aproximaciones analíticas
(Sesgadas pero rápidas y eficientes)

Inferencia aproximada

Aproximaciones analíticas
(Sesgadas pero rápidas y eficientes)

- Expectation propagation: arg min $\mathcal{D}_{KL}(p||q)$

Inferencia aproximada

Aproximaciones analíticas
(Sesgadas pero rápidas y eficientes)

- Expectation propagation: arg min $\mathcal{D}_{KL}(p||q)$
- Variational Inference: arg min $\mathcal{D}_{KL}(q||p)$
- Otras

Inferencia aproximada

Aproximaciones analíticas
(Sesgadas pero rápidas y eficientes)

- Expectation propagation: arg min $\mathcal{D}_{KL}(p||q)$
- Variational Inference: arg min $\mathcal{D}_{KL}(q||p)$
- Otras

Aproximaciones por Monte Carlo
(No sesgadas pero lentas y ineficientes)

Inferencia aproximada

Aproximaciones analíticas
(Sesgadas pero rápidas y eficientes)

- Expectation propagation: arg min $\mathcal{D}_{KL}(p||q)$
- Variational Inference: arg min $\mathcal{D}_{KL}(q||p)$
- Otras

Aproximaciones por Monte Carlo
(No sesgadas pero lentas y ineficientes)

- Gibbs Sampler
- Metropolis Hasting
- Hamiltonian Monte Carlo
- Otras

Teoría de la información
El problema de la comunicación con la realidad

Quisiéramos que el mensaje recibido sea igual al enviado

Teoría de la información

El problema de la comunicación con la realidad

Quisiéramos que el mensaje recibido sea igual al enviado

Soluciones

Física: acercarme para escuchar mejor

Inferencia: interpretar la señal con ruido

Teoría de la información

El problema de la comunicación con la realidad

Quisiéramos que el mensaje recibido sea igual al enviado

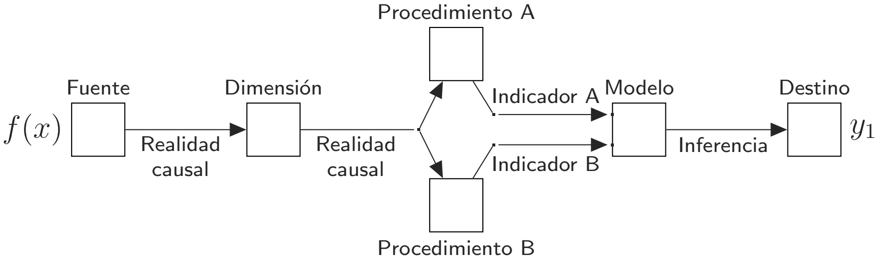
Soluciones

Física: acercarme para escuchar mejor

Inferencia: interpretar la señal con ruido

Teoría de la información

El problema de la comunicación con la realidad



Evaluar procedimientos alternativos.

Evaluación de test diagnóstico

Test rápido de chagas

Evaluación de test diagnóstico

Test rápido de chagas

$P(\text{Diagnóstico}|\text{Estado})$

	D = Positivo	D = Negativo
E = Verdadero	**Sensibilidad**	1-Sensibilidad
E = Falso	1-Especificidad	**Especificidad**

Evaluación de test diagnóstico

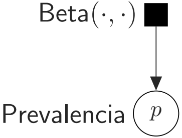
Test rápido de chagas

$P(\text{Diagnóstico}|\text{Estado})$

	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

Evaluación de test diagnóstico

Test rápido de chagas



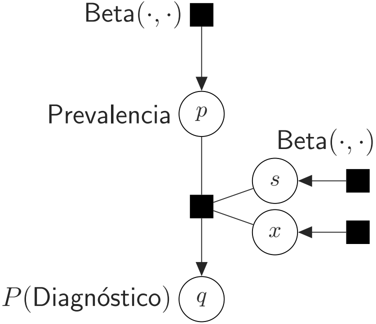
P(Diagnóstico|Estado)

	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

$p \sim \text{Beta}(\alpha, \beta)$

Evaluación de test diagnóstico

Test rápido de chagas



P(Diagnóstico|Estado)

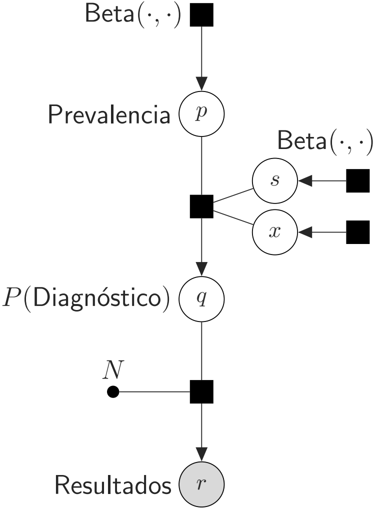
	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

$p \sim \text{Beta}(\alpha, \beta)$

$q = p\, s + (1-p)\,(1-x)$

Evaluación de test diagnóstico

Test rápido de chagas



$P(\text{Diagnóstico}|\text{Estado})$

	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

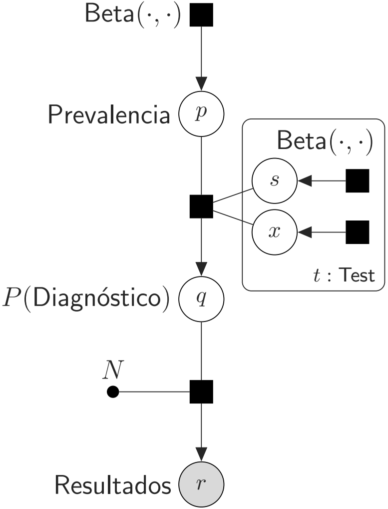
$$p \sim \text{Beta}(\alpha, \beta)$$

$$q = p\,s + (1-p)\,(1-x)$$

$$r = (\underbrace{n_0}_{-}, \underbrace{n_1}_{+}) \sim \text{Binomial}(q, N)$$

Evaluación de test diagnóstico

Test rápido de chagas



P(Diagnóstico|Estado)

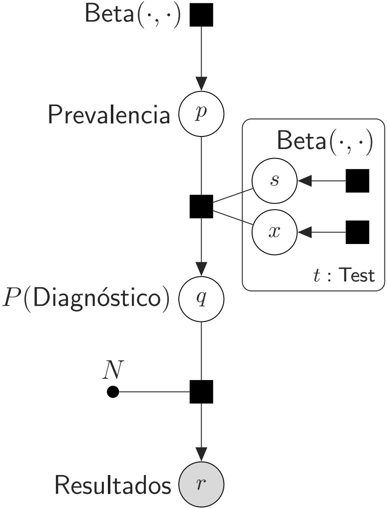
	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

$p \sim$ Beta(α, β)

$$r = (\underbrace{n_0}_{--}, \underbrace{n_1}_{-+}, \underbrace{n_2}_{+-}, \underbrace{n_3}_{++}) \sim \text{ Multinomial}(q, N)$$

Evaluación de test diagnóstico

Test rápido de chagas



P(Diagnóstico|Estado)

	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

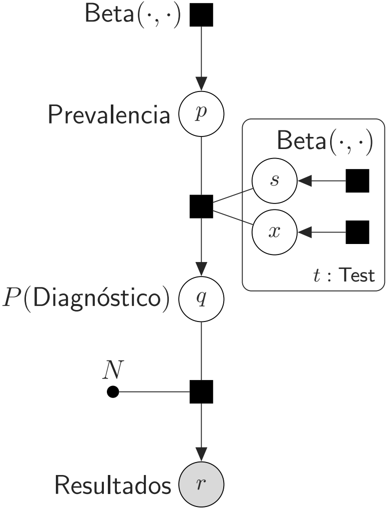
$p \sim$ Beta(α, β)

$$r = (\underbrace{n_0}_{--}, \underbrace{n_1}_{-+}, \underbrace{n_2}_{+-}, \underbrace{n_3}_{++}) \sim \text{ Multinomial}(q, N)$$

$$q = (\ q_0\ ,\ q_1\ ,\ q_2,\ q_3\ )$$

Evaluación de test diagnóstico

Test rápido de chagas



P(Diagnóstico|Estado)

	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

$p \sim \text{Beta}(\alpha, \beta)$

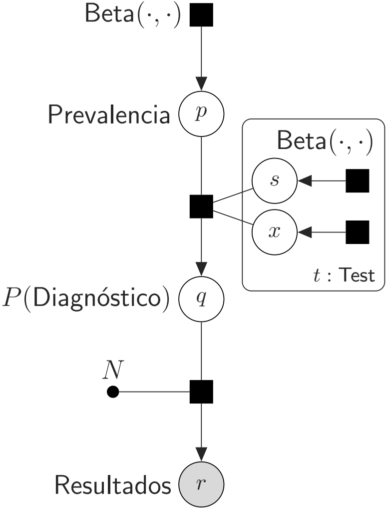
$$r = (\underbrace{n_0}_{--}, \underbrace{n_1}_{-+}, \underbrace{n_2}_{+-}, \underbrace{n_3}_{++}) \sim \text{Multinomial}(q, N)$$

$$q = (\ q_0\ ,\ q_1\ ,\ q_2,\ q_3\ )$$

$$q_0 = p\,(1-s_a)\,(1-s_b) + (1-p)\,x_a\,x_b$$

Evaluación de test diagnóstico

Test rápido de chagas



P(Diagnóstico|Estado)

	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

$p \sim \text{Beta}(\alpha, \beta)$

$r = (\underbrace{n_0}_{--}, \underbrace{n_1}_{-+}, \underbrace{n_2}_{+-}, \underbrace{n_3}_{++}) \sim \text{ Multinomial}(q, N)$

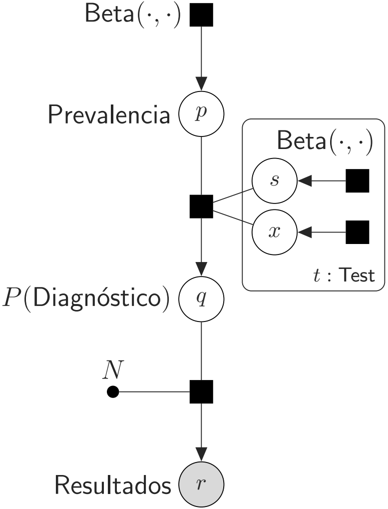
$q = (\ q_0\ ,\ q_1\ ,\ q_2,\ q_3\ )$

$q_0 = p\,(1-s_a)\,(1-s_b) + (1-p)\,x_a\,x_b$

$q_1 = p\,(1-s_a)\,s_b + (1-p)\,x_a\,(1-x_b)$

Evaluación de test diagnóstico

Test rápido de chagas



P(Diagnóstico|Estado)

	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

$p \sim \text{Beta}(\alpha, \beta)$

$r = (\underbrace{n_0}_{--}, \underbrace{n_1}_{-+}, \underbrace{n_2}_{+-}, \underbrace{n_3}_{++}) \sim \text{ Multinomial}(q, N)$

$q = (\ q_0\ ,\ q_1\ ,\ q_2,\ q_3\ )$

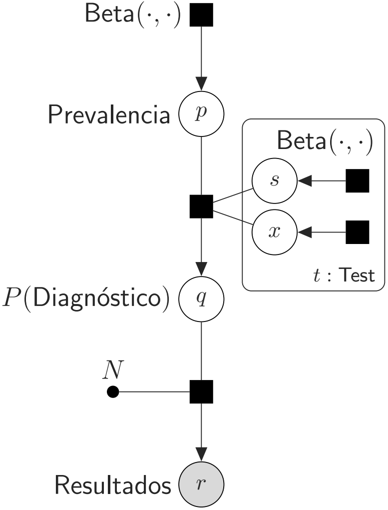
$q_0 = p\,(1-s_a)\,(1-s_b) + (1-p)\,x_a\,x_b$

$q_1 = p\,(1-s_a)\,s_b + (1-p)\,x_a\,(1-x_b)$

$q_2 = \dots$

Evaluación de test diagnóstico

Test rápido de chagas



P(Diagnóstico|Estado)

	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

$p \sim$ Beta(α, β)

$$r = (\underbrace{n_0}_{--}, \underbrace{n_1}_{-+}, \underbrace{n_2}_{+-}, \underbrace{n_3}_{++}) \sim \text{ Multinomial}(q, N)$$

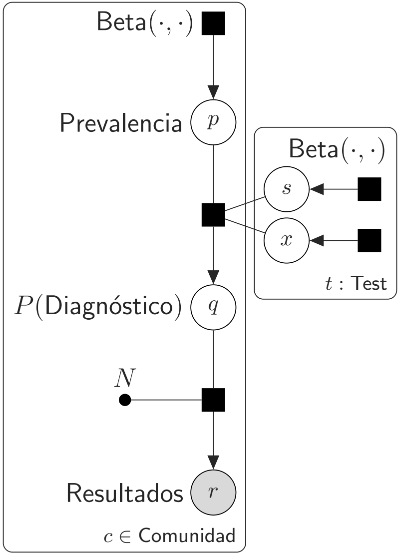
$$q = (\ q_0\ ,\ q_1\ ,\ q_2,\ q_3\ )$$

Problema no identificable

5 hipótesis (p, s_a, x_a, s_b, x_b) y 3 datos (n_0, n_1, n_2)

Evaluación de test diagnóstico

Test rápido de chagas



P(Diagnóstico|Estado)

	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

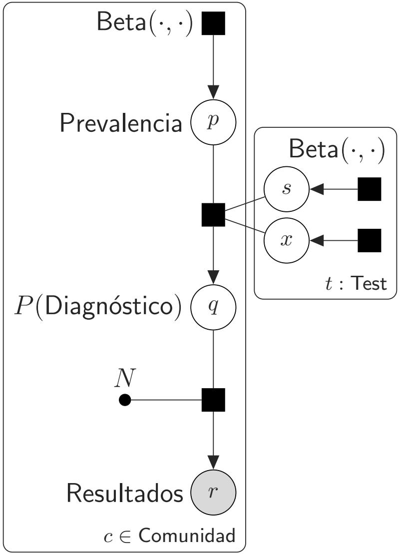
$p \sim \text{Beta}(\alpha, \beta)$

$r = (\underbrace{n_0}_{--}, \underbrace{n_1}_{-+}, \underbrace{n_2}_{+-}, \underbrace{n_3}_{++}) \sim \text{Multinomial}(q, N)$

$q = (\ q_0\ ,\ q_1\ ,\ q_2,\ q_3\ )$

Evaluación de test diagnóstico

Test rápido de chagas



P(Diagnóstico|Estado)

	D = Positivo	D = Negativo
E = Verdadero	s	$1-s$
E = Falso	$1-x$	x

$p \sim$ Beta(α,β)

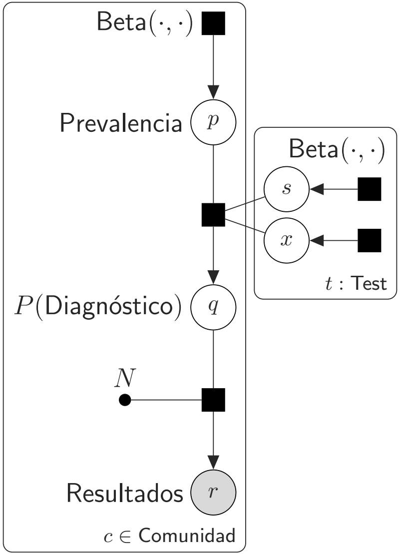
$r = (\underbrace{n_0}_{--}, \underbrace{n_1}_{-+}, \underbrace{n_2}_{+-}, \underbrace{n_3}_{++}) \sim$ Multinomial(q,N)

$q = (\ q_0\ ,\ q_1\ ,\ q_2,\ q_3\ )$

Problema identificable

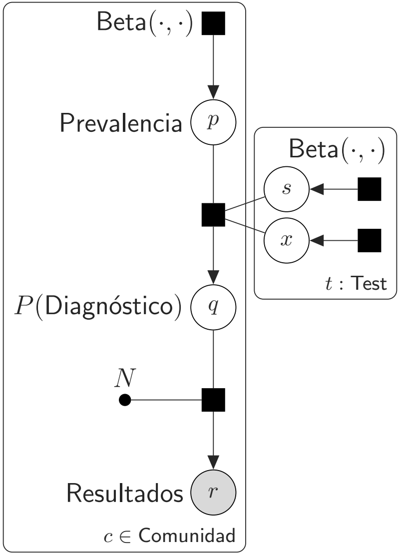
6 hipótesis $(p_A, p_B, s_a, x_a, s_b, x_b)$ y 6 datos (r_A, r_B)

Evaluación de test diagnóstico



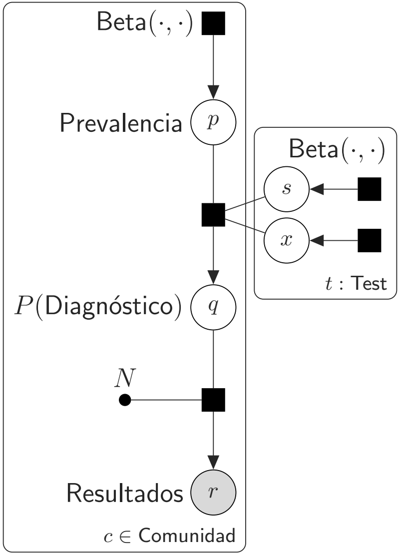

```
modelo <- "model{
  s[1] ~ dbeta(1, 1)
  s[2] ~ dbeta(1, 1)
  x[1] ~ dbeta(1, 1)
  x[2] ~ dbeta(1, 1)

  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
}"
```

Evaluación de test diagnóstico



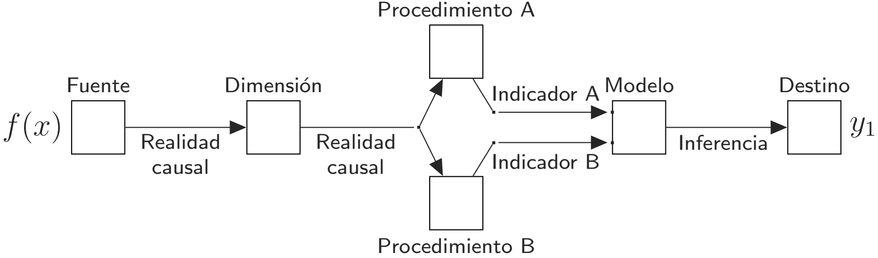

```
library("runjags") # JAGS: Just Another Gibbs Sampler

p = c(0, 0.15, 0.3, 0.45, 0.7) # REALIDAD
sensitivity = c(0.9, 0.6)
specificity = c(0.95, 0.9)

r <- matrix(nrow=4, ncol=length(p)) # DATA (simulada)
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

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)
```

Evaluación de test diagnóstico




```
     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
     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
```

Tasa de información
de los sistemas de comunicación



¿Con qué procedimiento nos quedamos?

$p = b$

Laboratorios de
Métodos Bayesianos

Bibliografía Unidad 4

- Bishop, C. **Pattern Recognition and Machine Learning**. 2006. (Descargar). (lectura 8.1, 8.2 y 8.4)
- Neal. **Pattern Recognition and Machine Learning**. 2020 (Draft). (Descargar). (lectura capítulo 1 y 3)