http://www.computervisionmodels.com/
Descargar el contenido de los temas 2 y 5

Tema 4
Fundamentos de Probabilidad

Rafael Molina

Variables Aleatorias

- Una variable aleatoria x denota una cantidad que es incierta.
- Puede ser el resultado de un experimento (lanzar una moneda) o una medición del mundo real (medir la temperatura).
- Si observamos diferentes realizaciones de x obtenemos diferentes valores.
- Algunos valores ocurren más que otros y esta información es capturada por una distribución de probabilidad.

Variables Aleatorias Discretas

Dos representaciones distintas

Histograma

La altura de cada barra representa la probabilidad.

La suma de las alturas ha de ser uno

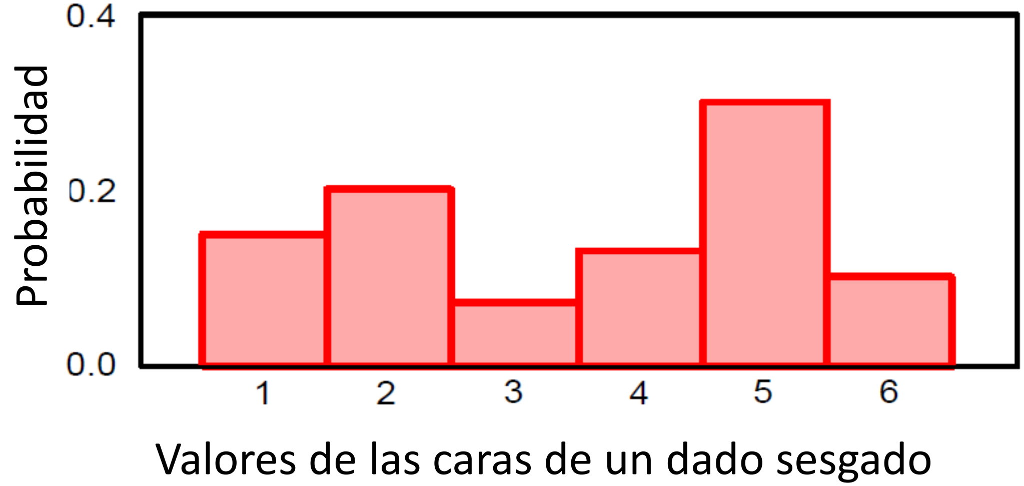
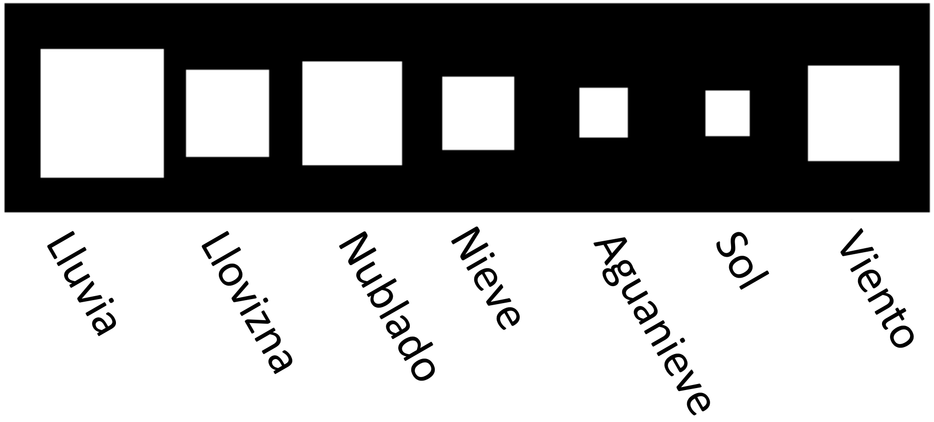


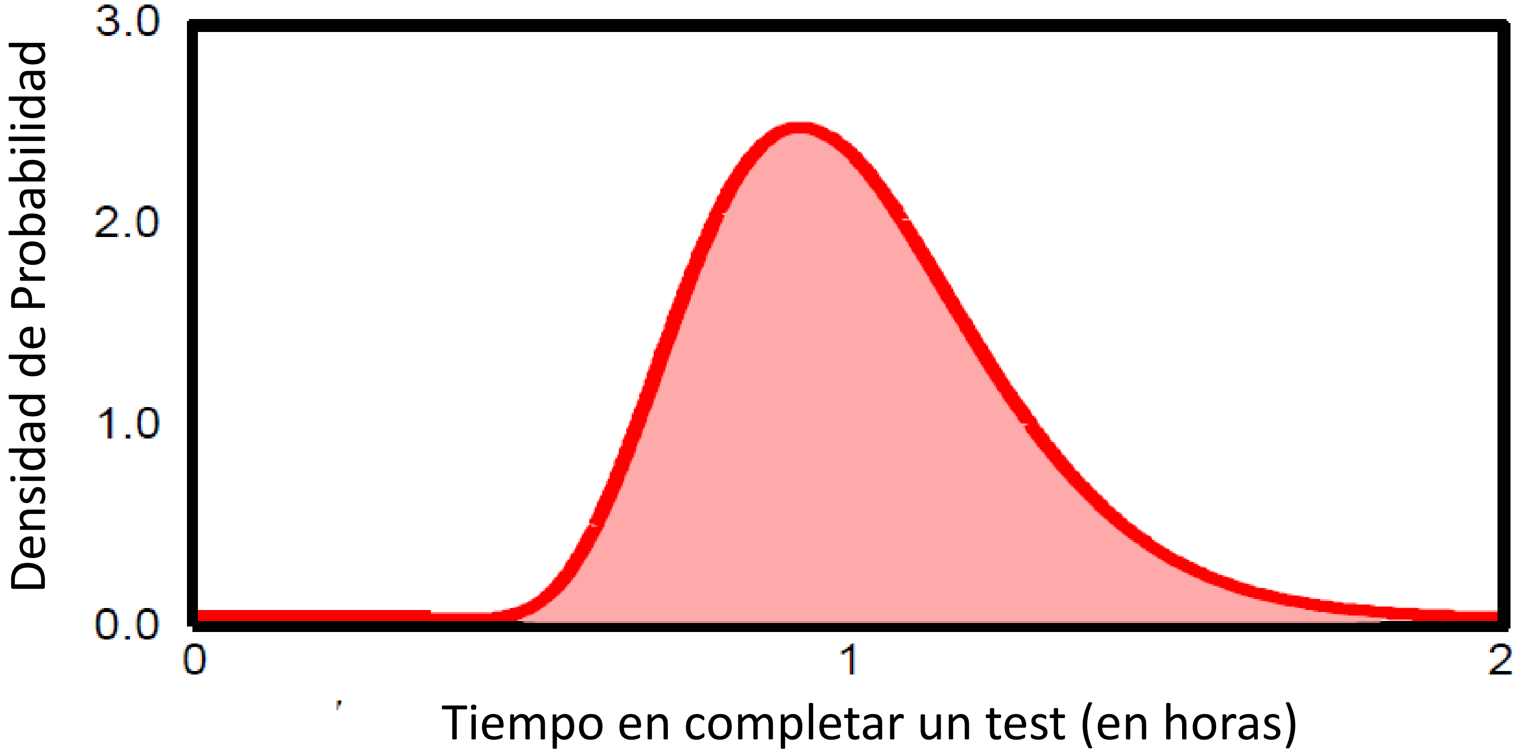
Diagrama de Hinton

El área de cada cuadrado representa una probabilidad.

La suma de las áreas ha de ser uno



Variables aleatorias continuas



Para variables continuas la densidad de probabilidad NO es la probabilidad de observar un valor. ¿Cuál es su interpretación?

Probabilidad Conjunta

- Consideremos dos variables aleatorias x e y
- Si observamos realizaciones múltiples de estas parejas entonces, normalmente, algunas combinaciones de resultados aparecen más frecuentemente que otras.
- Esto se captura en la distribución conjunta
- Se escribe como $\Pr(x,y)$
- $\Pr(x,y)$ se lee como "probabilidad de x e y"

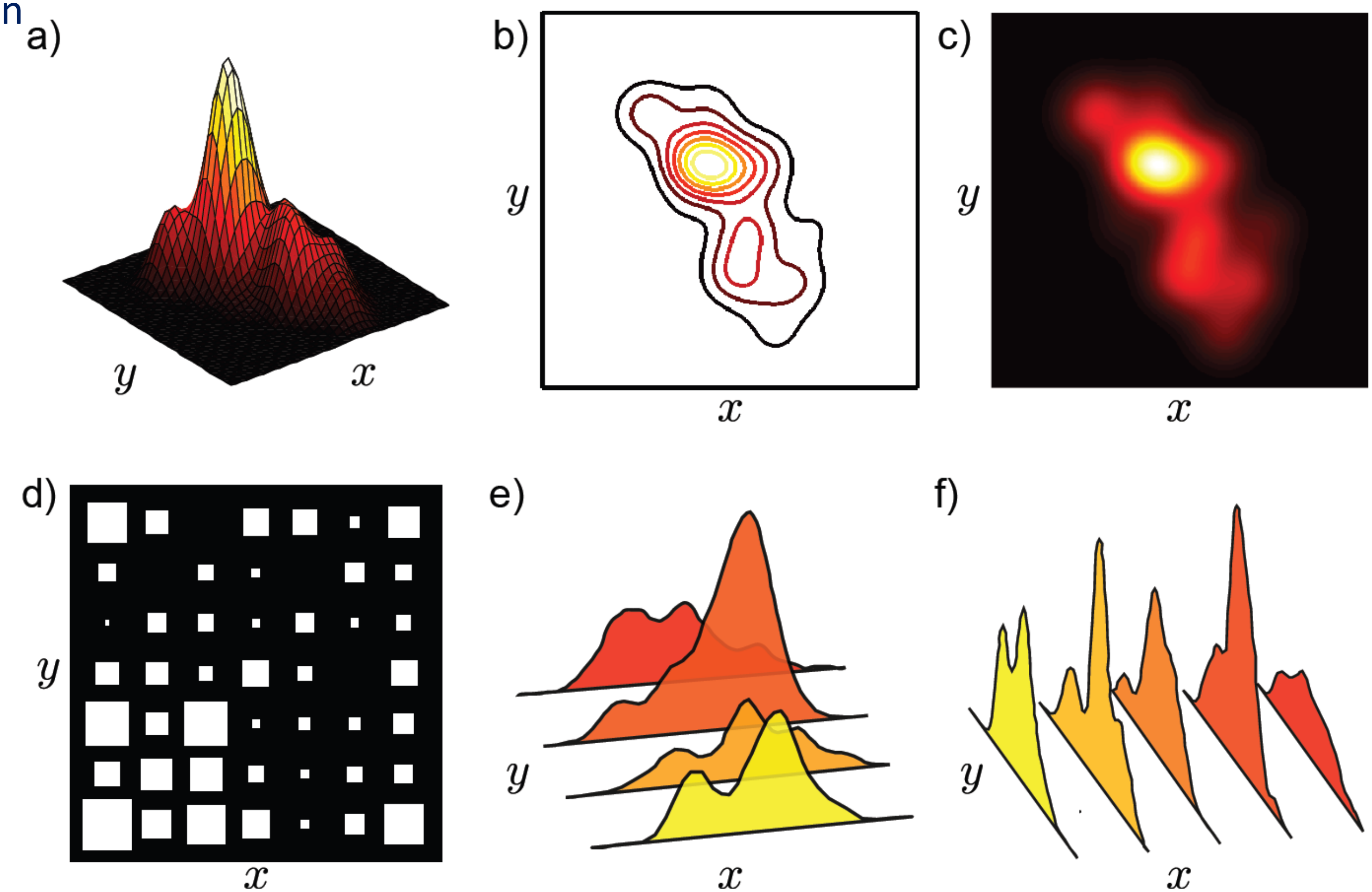
Probabilidad Conjunta

La misma distribución de probabilidad bidimensional vista como

- Superficie (a)
- Contornos (b)
- Imagen (c)

Ambas variables son continuas

- Variable bidimensional discreta (d)
- y discreta y x continua (e)
- x discreta e y continua



Marginalización

Podemos recuperar la distribución de probabilidad de cualquier variable en una distribución conjunta integrando (o sumando) en las otras variables

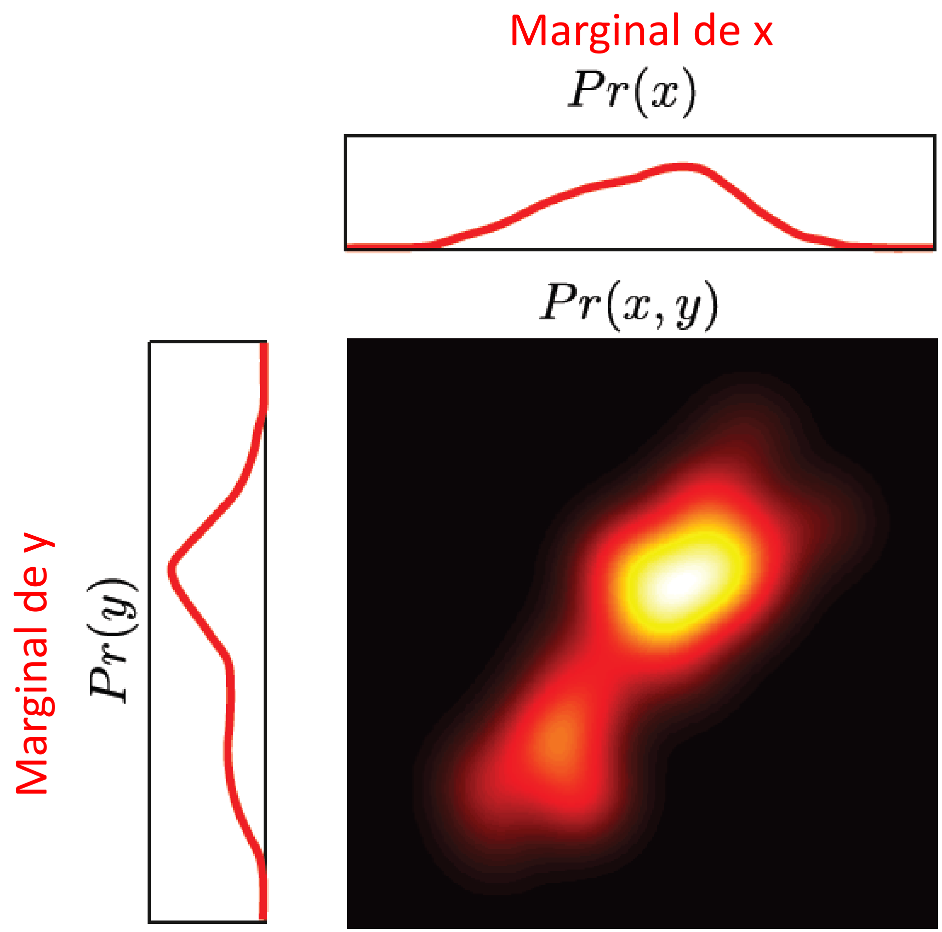
Ejemplo para ambas variables continuas

Marginalizamos sobre y

$$Pr(x) = \int Pr(x, y)\, dy$$

$$Pr(y) = \int Pr(x, y)\, dx$$

Marginalizamos sobre x



Marginalización

Podemos recuperar la distribución de probabilidad de cualquier variable en una distribución conjunta integrando (o sumando) en las otras variables

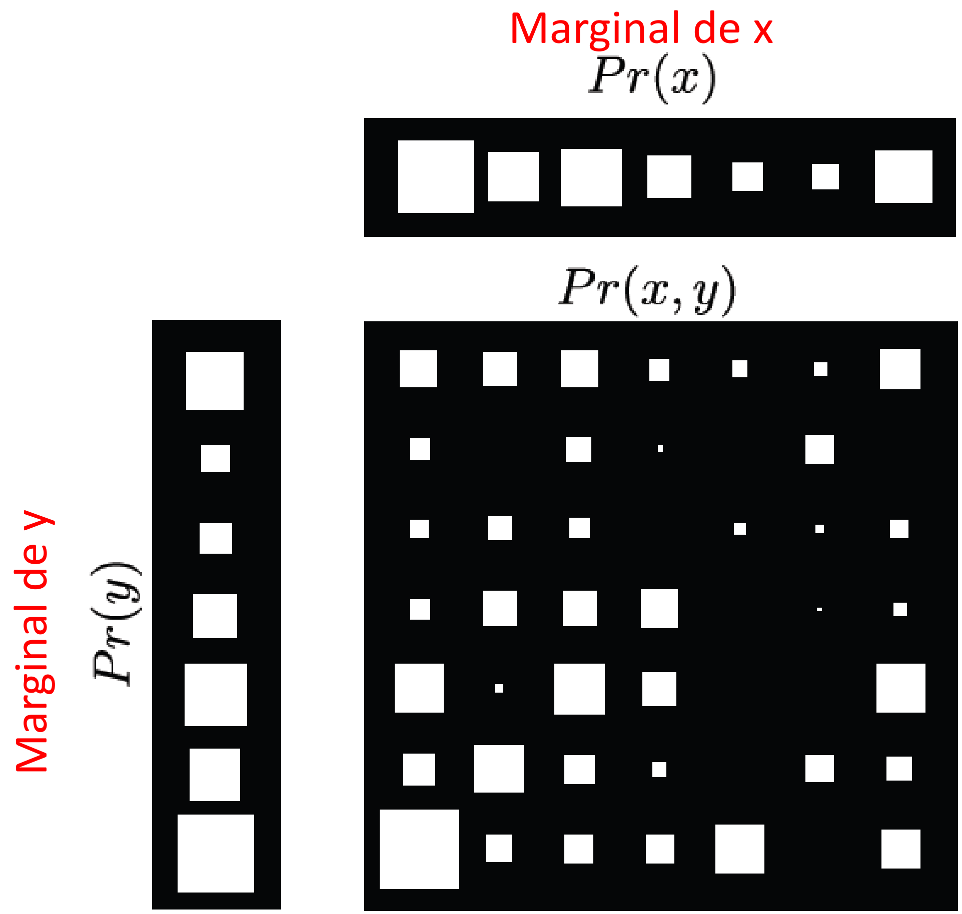
Ejemplo para ambas variables discretas

Marginalizamos sobre y

$$Pr(x) = \sum_{y} Pr(x, y)$$

$$Pr(y) = \sum_{x} Pr(x, y)$$

Marginalizamos sobre x

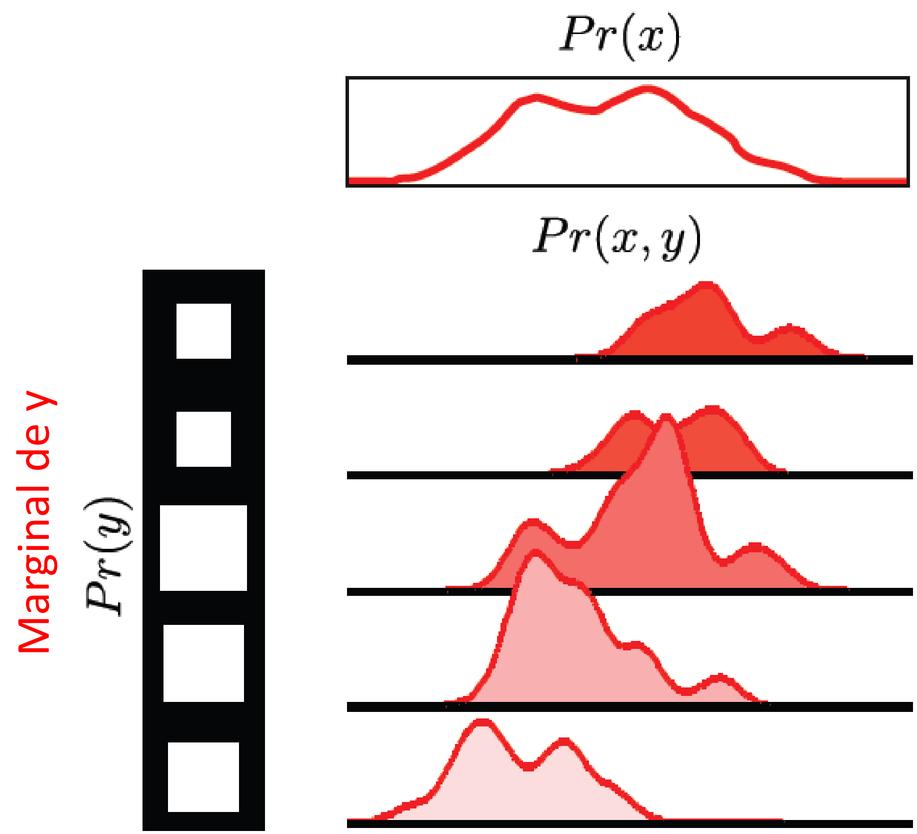


Marginalización

Podemos recuperar la distribución de probabilidad de cualquier variable en una distribución conjunta integrando (o sumando) en las otras variables

Ejemplo para una variable discreta (y) y otra continua (x)

Marginal de x



Marginalizamos sobre y

$$Pr(x) = \sum_y Pr(x,y)$$

$$Pr(y) = \int Pr(x,y)\,dx$$

Marginalizamos sobre x

Marginalización

Podemos recuperar la distribución de probabilidad de cualquier variable en una distribución conjunta integrando (o sumando) en las otras variables

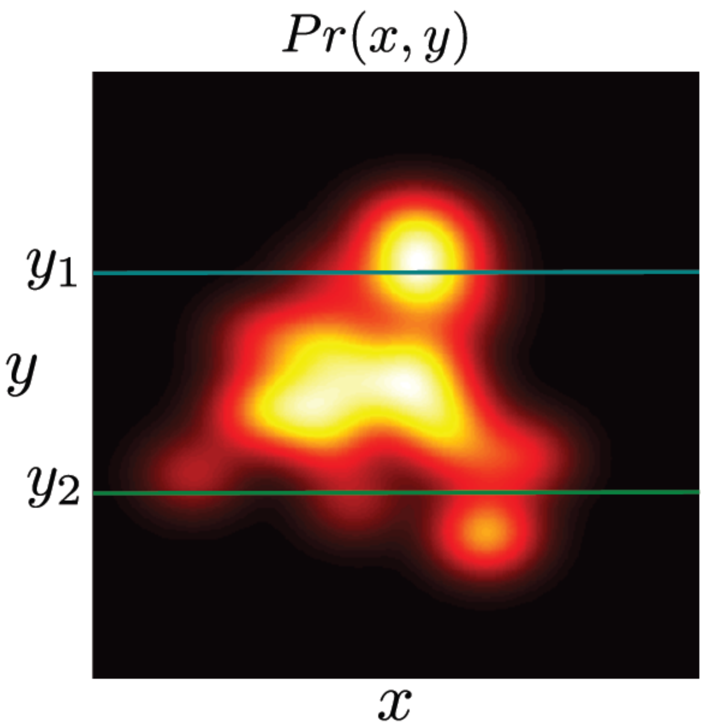
$$
\begin{aligned}
Pr(x) &= \int Pr(x,y)\,dy \\
Pr(y) &= \int Pr(x,y)\,dx
\end{aligned}
$$

Funciona también para más dimensiones. Obtenemos la distribución conjunta de las variables que quedan, es decir, sobre las que no hemos integrado o sumado.

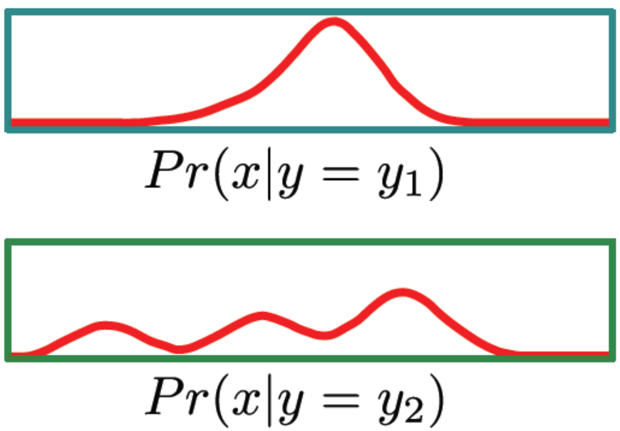
$$Pr(x,y) = \sum_{w} \int Pr(w,x,y,z)\,dz$$

Probabilidad Condicional

- La probabilidad condicional de x dado que $\mathrm{y=y_1}$ es la tendencia de la variable x a tomar diferentes valores dado que y está fijada a ser igual a $\mathrm{y_1}$.
- Se escribe como $\Pr(\mathrm{x}|\mathrm{y=y_1})$



Hemos observado $\mathrm{y_1}$ y analizamos la conducta de x

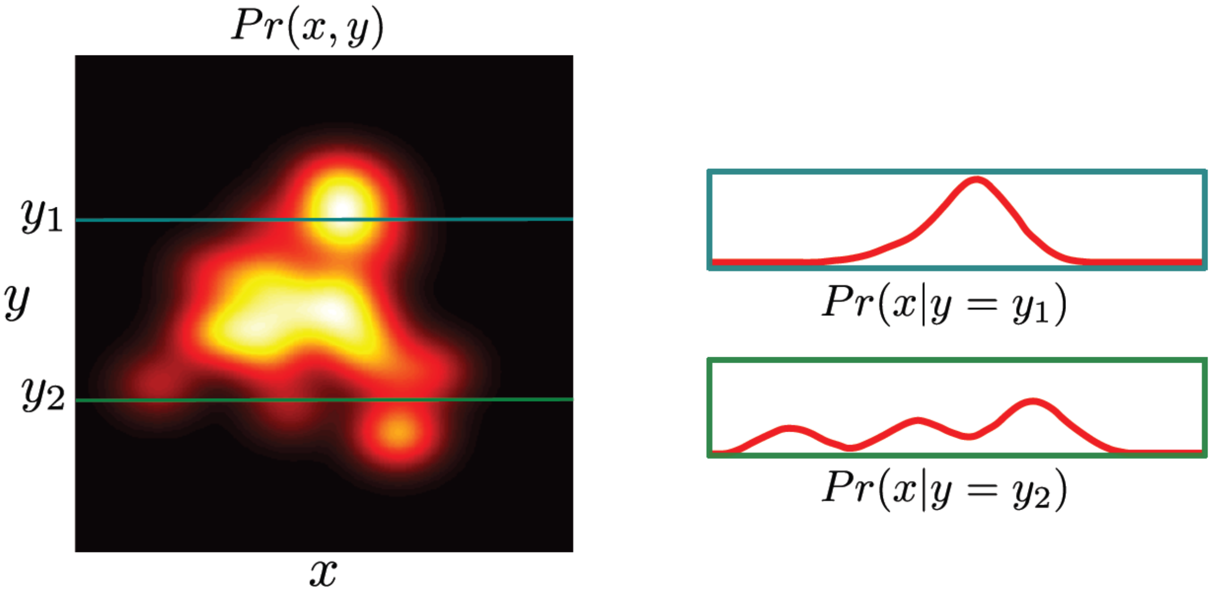


Hemos observado $\mathrm{y_2}$ y analizamos la conducta de x

Probabilidad Condicional

- La probabilidad condicional se extrae a partir de la probabilidad conjunta
- Extraer el corte correspondiente y normalizar

$$Pr(x|y=y^*) = \frac{Pr(x,y=y^*)}{\int Pr(x,y=y^*)dx} = \frac{Pr(x,y=y^*)}{Pr(y=y^*)}$$



Probabilidad Condicionada

$$Pr(x|y=y^*) = \frac{Pr(x,y=y^*)}{\int Pr(x,y=y^*)dx} = \frac{Pr(x,y=y^*)}{Pr(y=y^*)}$$

- Normalmente se escribe en forma compacta

$$Pr(x|y) = \frac{Pr(x,y)}{Pr(y)}$$

- Podemos reordenarla y escribir

$$Pr(x,y) = Pr(x|y)Pr(y)$$

$$Pr(x,y) = Pr(y|x)Pr(x)$$

Probabilidad Condicionada

$$Pr(x, y) = Pr(x|y)Pr(y)$$

- Esta idea puede extenderse a más de dos variables

$$\begin{aligned} Pr(w, x, y, z) &= Pr(w, x, y|z)Pr(z) \\ &= Pr(w, x|y, z)Pr(y|z)Pr(z) \\ &= Pr(w|x, y, z)Pr(x|y, z)Pr(y|z)Pr(z) \end{aligned}$$

Comprueba que estas igualdades son ciertas utilizando la definición de probabilidad condicionada

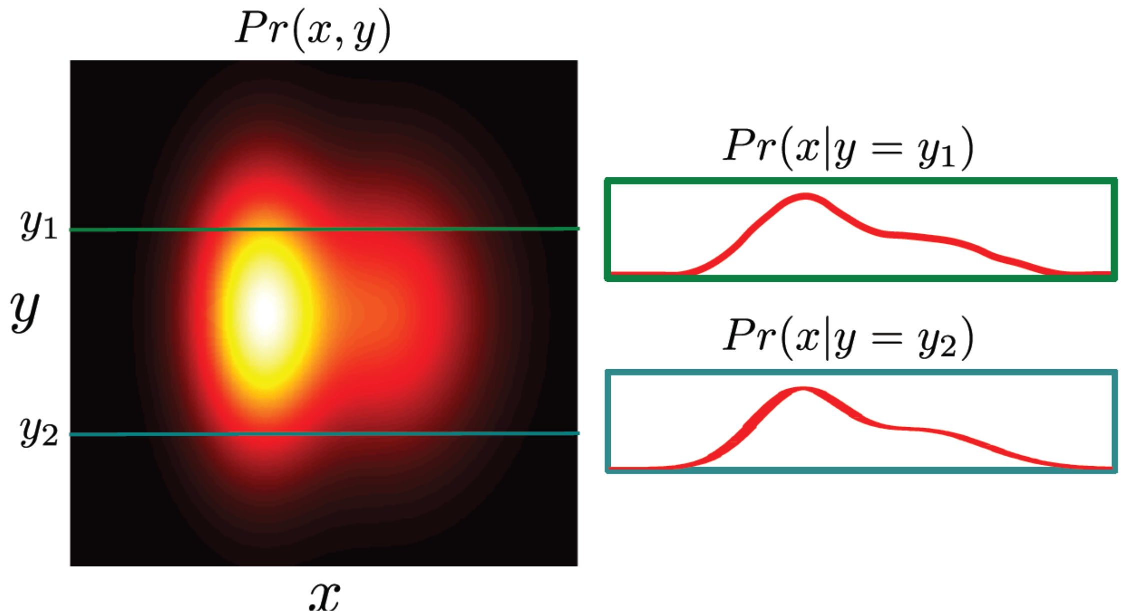
Independencia

- Si dos variables x e y son independientes entonces la variable x nos dice nada sobre la variable y (y viceversa)

Si una igualdad es cierta, la otra también lo es. Compruébalo

$$
\begin{aligned}
Pr(x|y) &= Pr(x) \\
Pr(y|x) &= Pr(y)
\end{aligned}
$$

Ejemplo de dos variables aleatorias continuas que son independientes



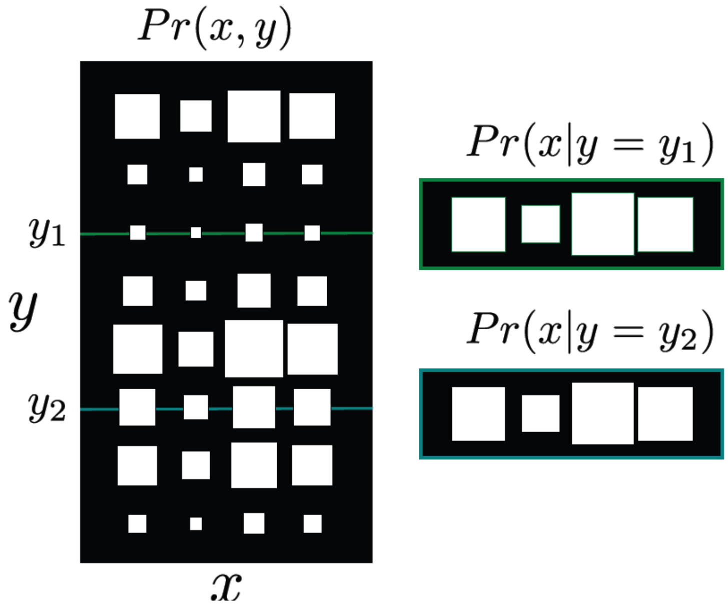
Independencia

- Si dos variables x e y son independientes entonces la variable x nos dice nada sobre la variable y (y viceversa)

Si una igualdad es cierta, la otra también lo es. Compruébalo.

$$
\begin{aligned}
Pr(x|y) &= Pr(x) \\
Pr(y|x) &= Pr(y)
\end{aligned}
$$

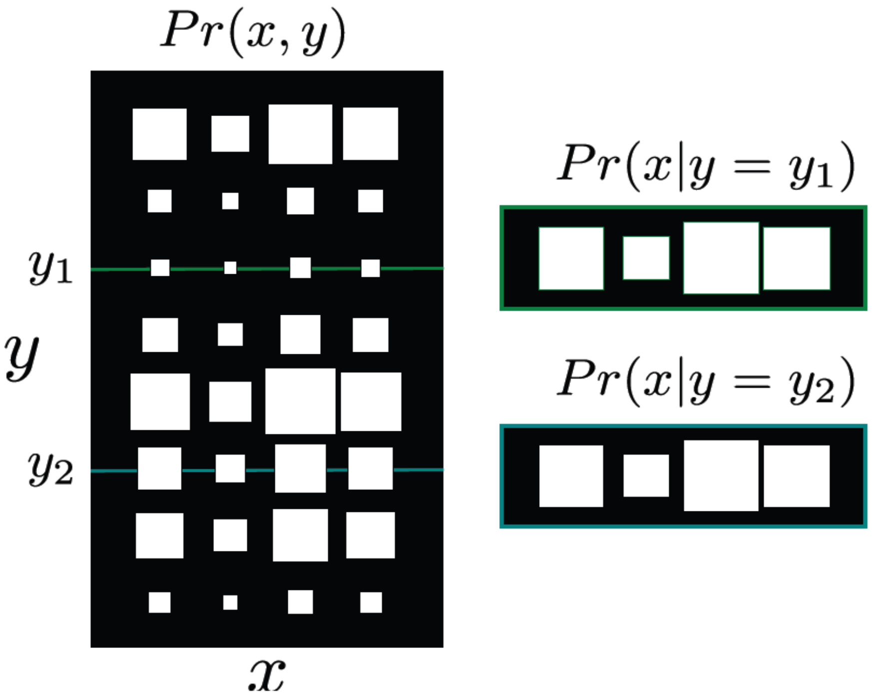
Ejemplo de dos variables aleatorias discretas que son independientes.



Independencia

- Cuando las variables son independientes, la distribución conjunta se factoriza como el producto de marginales:

$$
\begin{aligned}
Pr(x, y) &= Pr(x|y)Pr(y) \\
&= Pr(x)Pr(y)
\end{aligned}
$$



Esperanza

La esperanza nos dice el valor esperado o promedio de una función $f[x]$ teniendo en cuenta la distribución de x

Definición:

$$\mathrm{E}[f[x]] = \sum_x f[x] Pr(x)$$

$$\mathrm{E}[f[x]] = \int f[x] Pr(x)\, dx$$

Observa que estamos usando Pr(x) para notar tanto densidad como función masa de probabilidad

Esperanza

La esperanza nos dice el valor esperado o promedio de una función f [x] teniendo en cuenta la distribución de x

Definición en dos dimensiones:

$$\mathrm{E}[f[x,y]] \quad = \quad \iint f[x,y] Pr(x,y)\; dx\; dy$$

Observa que estamos usando Pr(x,y) para notar tanto densidad como función masa de probabilidad

Esperanza: Casos comunes

$$\mathrm{E}[f[x]] \quad = \quad \int f[x]Pr(x)\,dx$$

Función $f[\bullet]$	**Esperanza**
x	media , μ_x
x^k	k^{th} momento alrededor del cero
$(x-\mu_x)^k$	k^{th} momento alrededor de la media
$(x-\mu_x)^2$	varianza
$(x-\mu_x)^3$	skew
$(x-\mu_x)^4$	curtosis
$(x-\mu_x)(y-\mu_y)$	covarianza de x e y

Esperanza: Reglas

$$E[f[X]] \quad = \quad \int f[x] Pr(X = x) dx$$

Regla 1:

El valor esperado de una constante es una constante

$$\mathrm{E}[\kappa] = \kappa$$

Esperanza: Reglas

$$E[f[X]] \quad = \quad \int f[x]Pr(X = x)dx$$

Regla 2:

El valor esperado de una constante por una función es la constante por el valor esperado de la función

$$\mathrm{E}[kf[x]] = k\mathrm{E}[f[x]]$$

Esperanza: Reglas

$$E[f[X]] \quad = \quad \int f[x]Pr(X = x)dx$$

Regla 3:

La esperanza de una suma de funciones es la suma de las esperanzas de las funciones

$$\mathrm{E}[f[x] + g[x]] = \mathrm{E}[f[x]] + \mathrm{E}[g[x]]$$

Esperanza: Reglas

$$E[f[X]] = \int f[x]Pr(X = x)dx$$

Regla 4:

La esperanza del producto de funciones en las variables x e y, respectivamente, es el producto de las esperanzas de las funciones si x e y son independientes

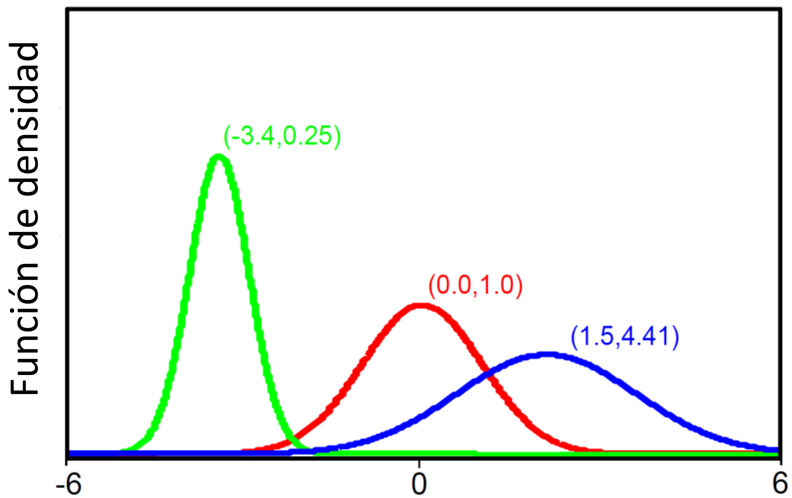
$$\mathrm{E}[f[x]g[y]] = \mathrm{E}[f[x]]\mathrm{E}[g[y]]$$ Si x e y son independientes

La distribución Normal Univariante

$$Pr(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp[-0.5(x-\mu)^2/\sigma^2]$$

que también escribiremos:

$$Pr(x|\mu,\sigma^2) = \mathrm{Norm}_x[\mu,\sigma^2]$$



La distribución Normal univariante describe una variable continua unidimensional.

Utiliza 2 parámetros μ y σ^2>0

Distribución Normal Multivariante

$$Pr(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}}\exp[-0.5(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]$$

D es el número de componentes de **x**

que también escribimos

$$Pr(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \mathrm{Norm}_{\mathbf{x}}[\boldsymbol{\mu},\boldsymbol{\Sigma}]$$

La distribución Normal Multivariante describe variables aleatorias múltiples. Utiliza dos parámetros:

- Un vector que contiene las medias, **μ**
- Una matriz de covarianzas simétrica “definida positiva” **Σ**

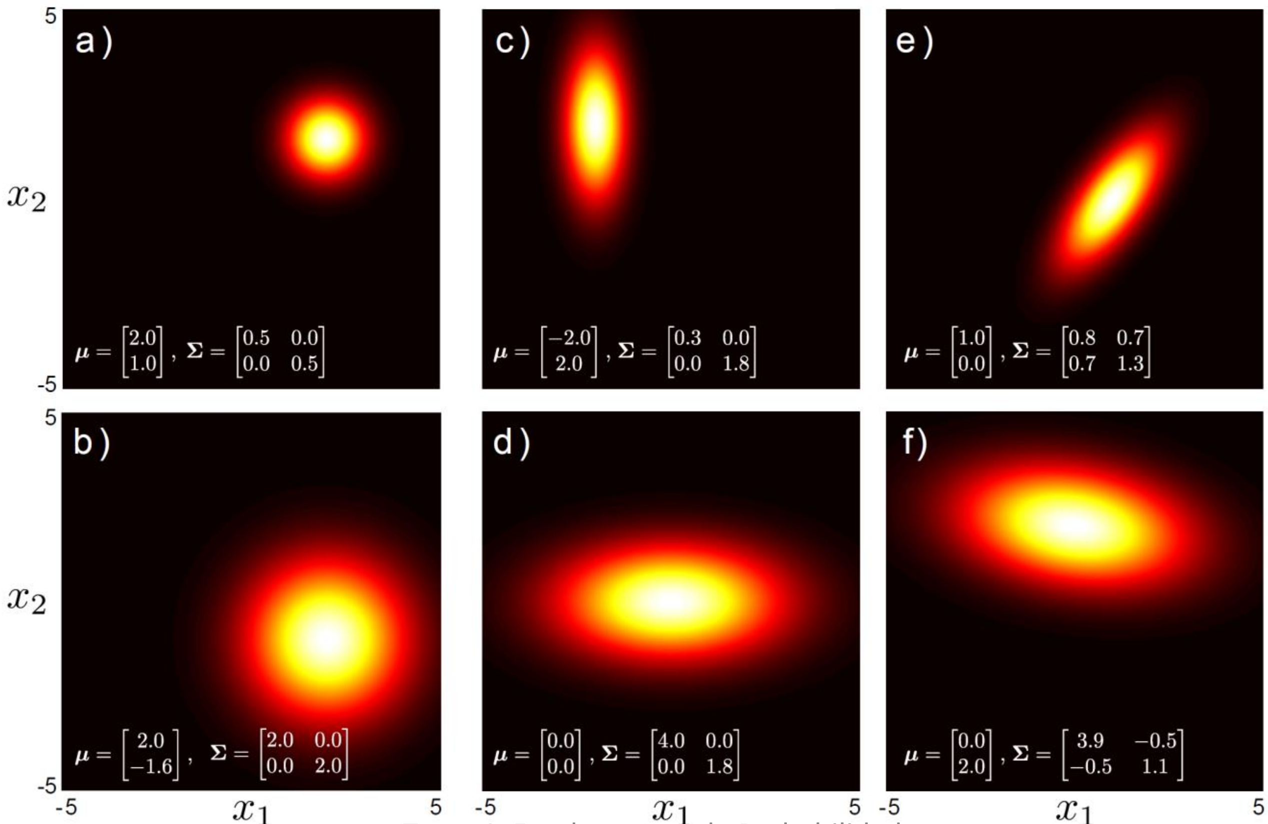
Matriz definida positiva: $\mathbf{z}^T\boldsymbol{\Sigma}\mathbf{z}$ es positivo para cualquier vector no nulo $\mathbf{z}$

Tipos de covarianzas

La matriz de covarianza tiene tres formas, llamadas esférica, diagonal y completa

$$\boldsymbol{\Sigma}_{\text{esfer}} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \quad \boldsymbol{\Sigma}_{\text{diag}} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \quad \boldsymbol{\Sigma}_{\text{comp}} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix}$$

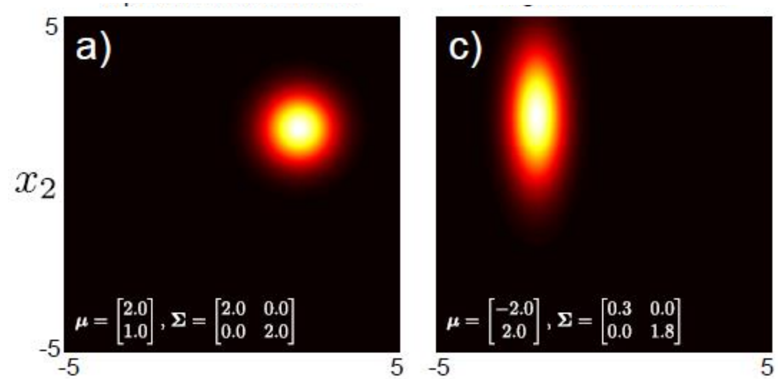
Covarianza esférica | Covarianza diagonal | Covarianza completa



Covarianza Diagonal= Independencia

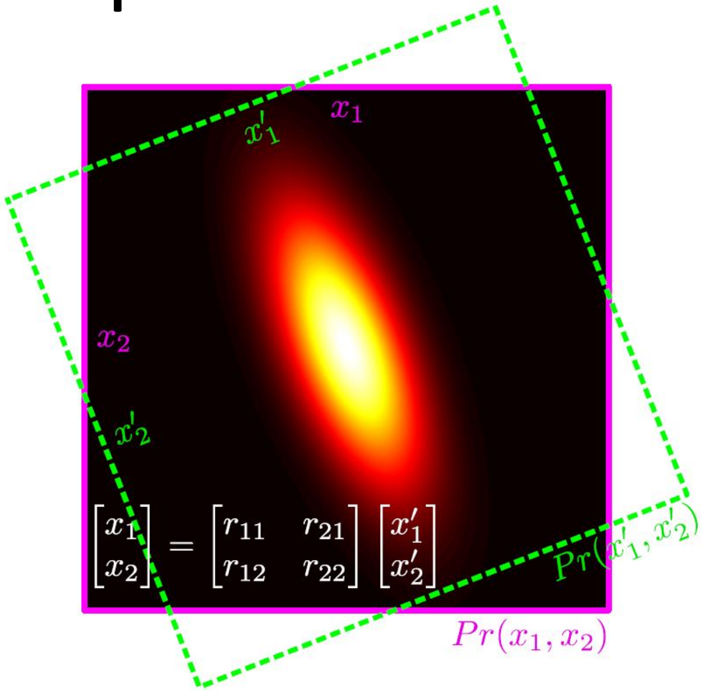
$$\mathbf{\Sigma}_{\text{esfer}} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \quad \mathbf{\Sigma}_{\text{diag}} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

Covarianza esférica   Covarianza diagonal



$$\begin{aligned}
Pr(x_1, x_2) &= \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}|}} \exp\left[-0.5\begin{pmatrix} x_1 & x_2 \end{pmatrix} \mathbf{\Sigma}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right] \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-0.5\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} \sigma_1^{-2} & 0 \\ 0 & \sigma_2^{-2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right] \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{x_1^2}{2\sigma_1^2}\right] \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{x_2^2}{2\sigma_2^2}\right] \\
&= Pr(x_1)Pr(x_2)
\end{aligned}$$

Descomposición de la Covarianza



Consideremos el sistema de referencia verde:

$$Pr(\mathbf{x}') = \frac{1}{(2\pi)^{K/2}|\Sigma'_{diag}|^{1/2}} \exp\left[-0.5\mathbf{x}'^T \Sigma'^{-1}_{diag}\mathbf{x}'\right]$$

La relación entre los sistemas de referencia verde y rosa es una rotación:

$$\mathbf{x}' = \mathbf{R}\mathbf{x}$$

Descomposición de la Covarianza

Sustituyendo:

$$
\begin{aligned}
Pr(\mathbf{x}) &= \frac{1}{(2\pi)^{K/2}|\Sigma'_{diag}|^{1/2}} \exp\left[-0.5(\mathbf{Rx})^T \Sigma'^{-1}_{diag} \mathbf{Rx}\right] \\
&= \frac{1}{(2\pi)^{K/2}|\mathbf{R}^T\Sigma'_{diag}\mathbf{R}|^{1/2}} \exp\left[-0.5\mathbf{x}^T\mathbf{R}^T \Sigma'^{-1}_{diag} \mathbf{Rx}\right]
\end{aligned}
$$

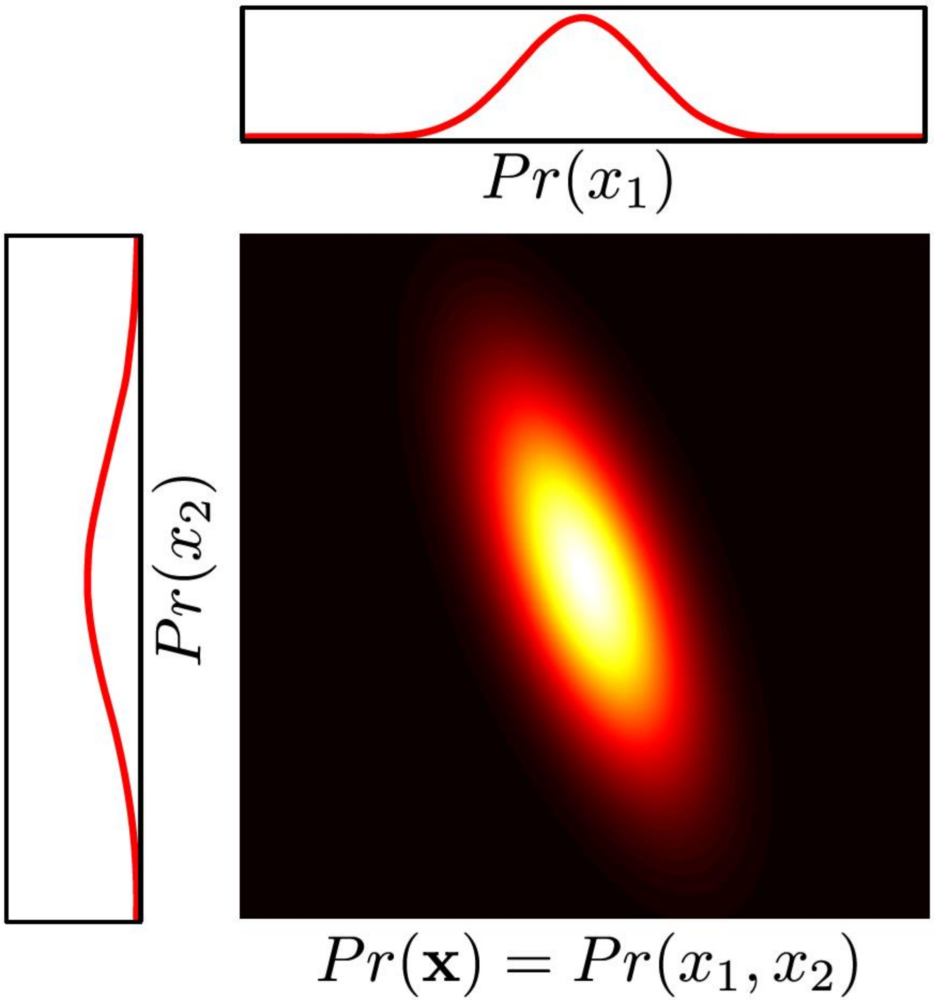
Conclusión:

$$\mathbf{\Sigma}_{\text{comp}} = \mathbf{R}^T \mathbf{\Sigma}'_{diag} \mathbf{R}$$

Una matriz de covarianza completa puede descomponerse en una matriz de rotación y una diagonal

Fíjate que este resultado nos indica como podemos generar realizaciones de una distribución normal con covarianza arbitraria a partir de realizaciones de una normal con covarianza la identidad

Distribuciones marginales



Las distribuciones marginales de una normal multivariante son también normales

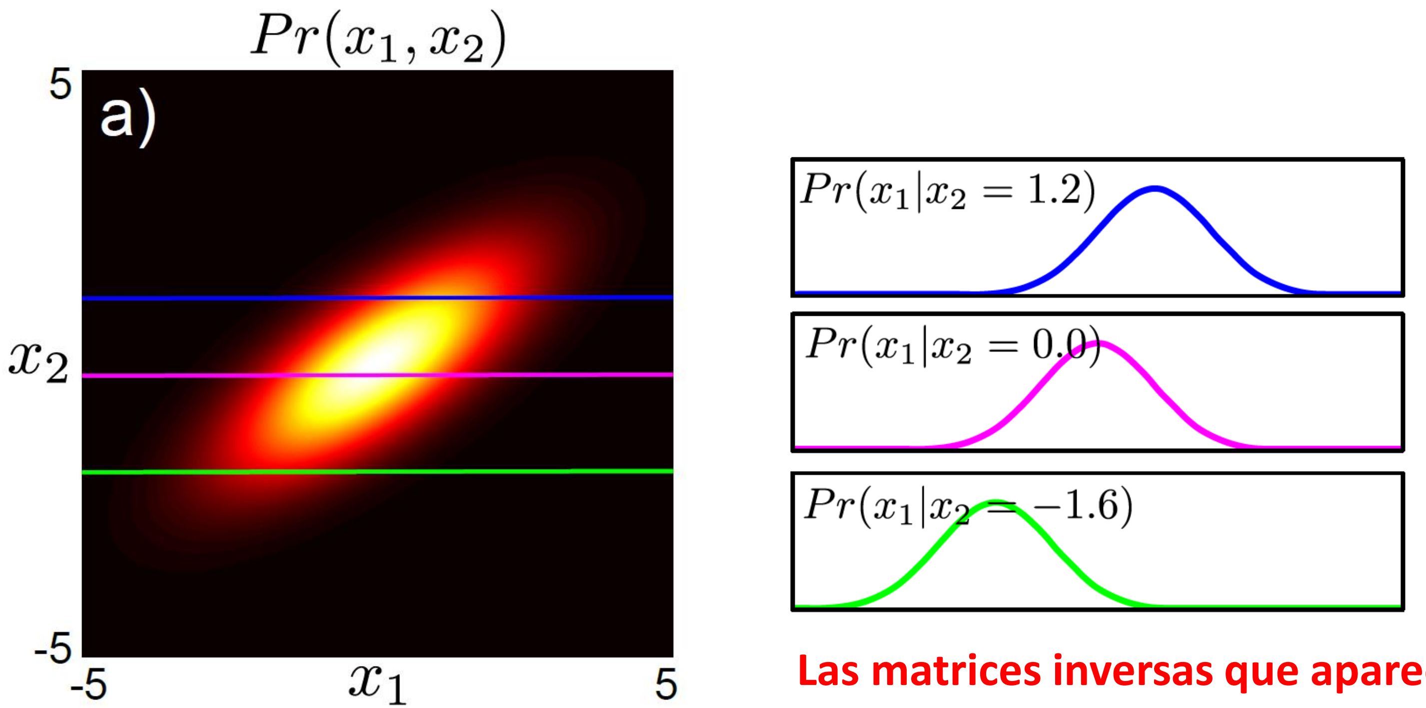
$$Pr(\mathbf{x}) = Pr\left(\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\right)$$

$$= \text{Norm}_{\mathbf{x}}\left[\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{21}^T\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right]$$

entonces

$$\begin{aligned} Pr(\mathbf{x}_1) &= \text{Norm}_{\mathbf{x}_1}[\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}] \\ Pr(\mathbf{x}_2) &= \text{Norm}_{\mathbf{x}_2}[\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}] \end{aligned}$$

Distribuciones condicionadas



Las matrices inversas que aparecen pueden traernos muchos problemas si son grandes

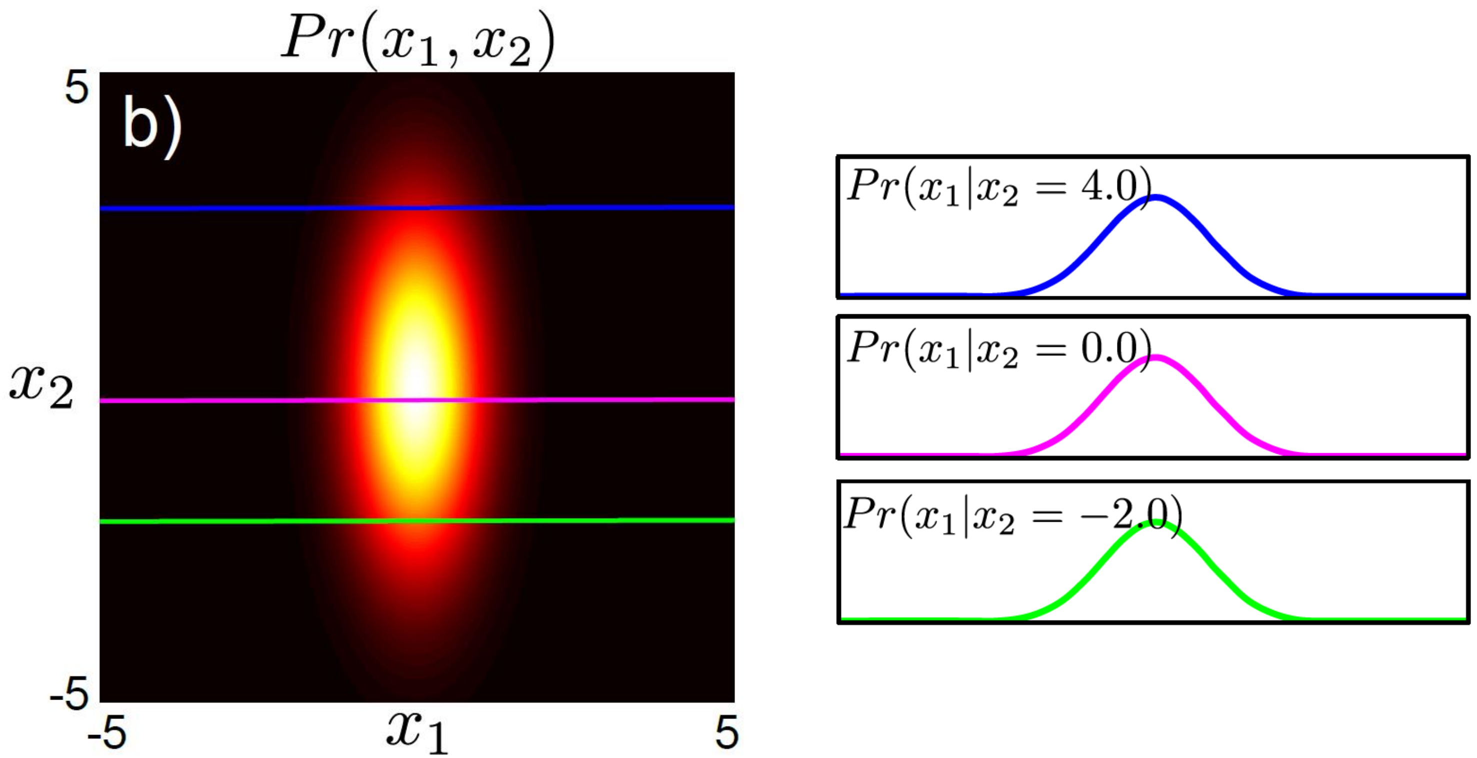
Si

$$Pr(\mathbf{x}) = Pr\left(\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\right) = \text{Norm}_{\mathbf{x}}\left(\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{21}^T\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right)$$

entonces

$$\begin{aligned} Pr(\mathbf{x}_1|\mathbf{x}_2) &= \text{Norm}_{\mathbf{x}_1}\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{21}^T\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{21}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)\\ Pr(\mathbf{x}_2|\mathbf{x}_1) &= \text{Norm}_{\mathbf{x}_2}\left(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{21}^T\right) \end{aligned}$$

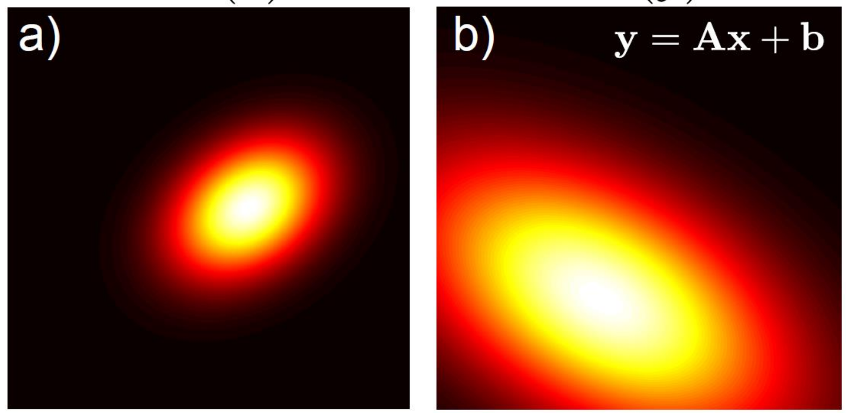
Distribuciones Condicionadas



Para el caso esférico/diagonal, x_1 y x_2 son independientes, de forma que todas las distribuciones condicionadas son la misma.

Transformación de Variables

$Pr(\mathbf{x})$ $Pr(\mathbf{y})$



Si

$$Pr(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \text{Norm}_{\mathbf{x}}[\boldsymbol{\mu}, \boldsymbol{\Sigma}]$$

y transformamos la variable en

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

El resultado es también una distribución normal

$$Pr(\mathbf{y}) = \text{Norm}_{\mathbf{y}}[\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$$

Transformación de Variables

Si

$$\mathrm{p}(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

con

$$\mathrm{p}(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{x}|A\mathbf{x}, \sigma^2\mathbf{I})$$

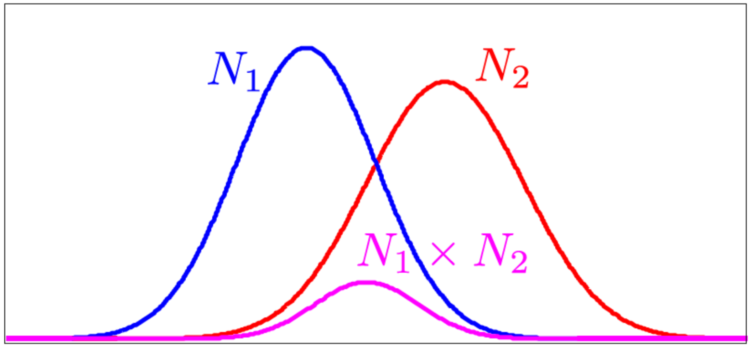
o dicho de otra forma

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon} \quad \mathrm{p}(\boldsymbol{\epsilon}) = \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \sigma^2\mathbf{I})$$

¿Cuál es la distribución de **y**? ¿y la de **x** dado **y**?

Producto de dos normales

Estos resultados se obtienen muy fácilmente derivando el exponente e igualando a cero. Lo veremos en clase



N_1 una normal
N_2 otra
N_1xN_2 producto sin normalizar
El producto normalizado (la normal correspondiente) tendrá menor varianza que cada una y la media estará entre las dos. Pruébalo

El producto de dos distribuciones normales sobre la misma variable es proporcional a una distribución normal sobre la misma variable

$$\mathrm{Norm}_{\mathbf{x}}[\mathbf{a},\mathbf{A}]\mathrm{Norm}_{\mathbf{x}}[\mathbf{b},\mathbf{B}] = \kappa \cdot \mathrm{Norm}_{\mathbf{x}}\left[\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}(\mathbf{A}^{-1}\mathbf{a}+\mathbf{B}^{-1}\mathbf{b}),\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}\right]$$

Una integral de (mucho) interés

Si **x** es un vector con M componentes e **y** tiene N y tenemos las distribuciones

$$\mathrm{q}(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

y

$$\mathrm{p}(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|A\mathbf{x}, \sigma^2\mathbf{I}) \quad \mathrm{p}(\mathbf{x}) = \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \sigma^2\mathbf{I})$$

¿Cuánto vale

$$\int \mathrm{q}(\mathbf{x}) \ln \frac{\mathrm{p}(\mathbf{x})\mathrm{p}(\mathbf{y}|\mathbf{x})}{\mathrm{q}(\mathbf{x})} \mathrm{d}\mathbf{x}$$

?

La regla de Bayes

- Habíamos visto que

$$Pr(x|y = y^*) = \frac{Pr(x, y = y^*)}{\int Pr(x, y = y^*)dx} = \frac{Pr(x, y = y^*)}{Pr(y = y^*)}$$

- Normalmente se escribe en forma compacta

$$Pr(x|y) = \frac{Pr(x, y)}{Pr(y)}$$

- Podemos reordenarla y escribir

$$Pr(x, y) = Pr(x|y)Pr(y)$$

$$Pr(x, y) = Pr(y|x)Pr(x)$$

La regla de Bayes

De lo que hemos visto antes:

$$Pr(x,y) = Pr(x|y)Pr(y)$$

$$Pr(x,y) = Pr(y|x)Pr(x)$$

Combinando ambas ecuaciones:

$$Pr(y|x)Pr(x) = Pr(x|y)Pr(y)$$

Reordenando:

$$\begin{aligned} Pr(y|x) &= \frac{Pr(x|y)Pr(y)}{Pr(x)} \\ &= \frac{Pr(x|y)Pr(y)}{\int Pr(x,y)\,dy} \\ &= \frac{Pr(x|y)Pr(y)}{\int Pr(x|y)Pr(y)\,dy} \end{aligned}$$

Observa que no hay nada nuevo. Pero esta ecuación tiene enormes implicaciones

Terminología de la regla de Bayes

Verosimilitud – tendencia a observar un cierto valor de x dado un cierto valor de y

A Priori – lo que sabemos sobre y antes de observar x

$$Pr(y|x) = \frac{Pr(x|y)Pr(y)}{\int Pr(x|y)Pr(y)\, dy}$$

A Posteriori – lo que conocemos de y después de observar x

Evidencia – una constante que nos asegura que la distribución a posteriori (sobre y) es una distribución válida

Recuerda

$$Pr(y|x) = \frac{Pr(x|y)Pr(y)}{\int Pr(x|y)Pr(y)\,dy}$$

Es muy importante tener claro lo siguiente, si hemos observado x y queremos estimar y ¿qué podemos hacer?

- Usar p(x/y), conduce a la llamada estimación de máxima verosimilitud

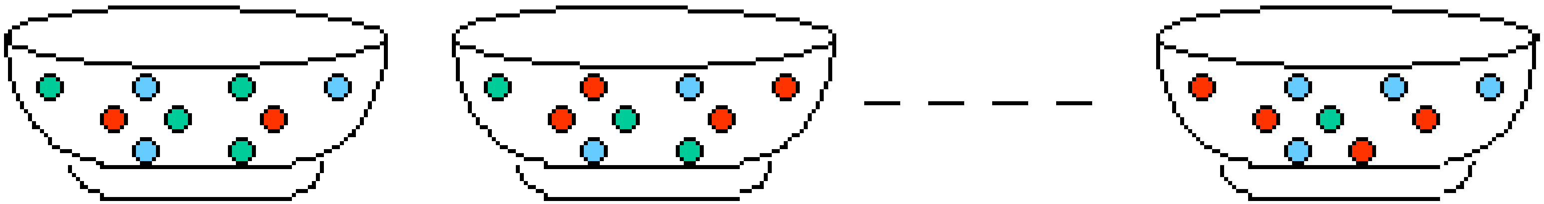
- Usar p(y)p(x/y) conduce a la estimación de la moda de la distribución a posteriori

- Calcular p(y/x) nos lleva a la aproximación bayesiana a la estimación

Para terminar,

tres problemas no iguales que usan la regla de Bayes

Un Problema
(que deberíais saber resolver)



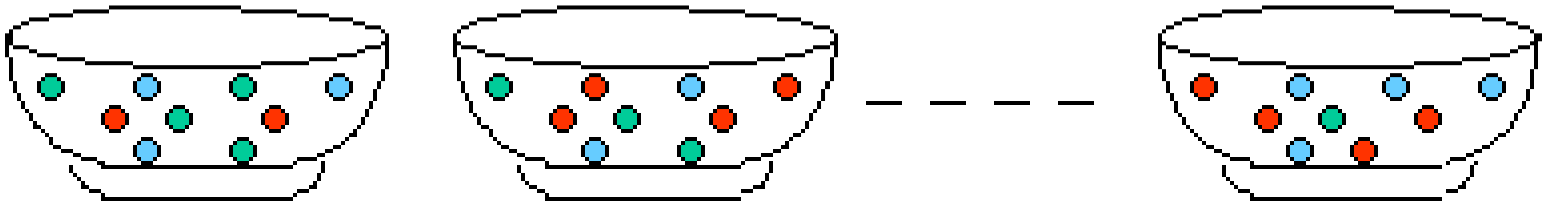
Cada urna (y) tiene una probabilidad de ser elegida

$Pr(y_1)$ $Pr(y_2)$ ----------- $Pr(y_N)$

Dentro de cada urna, cada tipo de bola (B) tiene una probabilidad de ser elegida

$Pr(B/y_1)$ $Pr(B/y_2)$ ----------- $Pr(B/y_N)$

Un Problema
(que deberíais saber resolver)

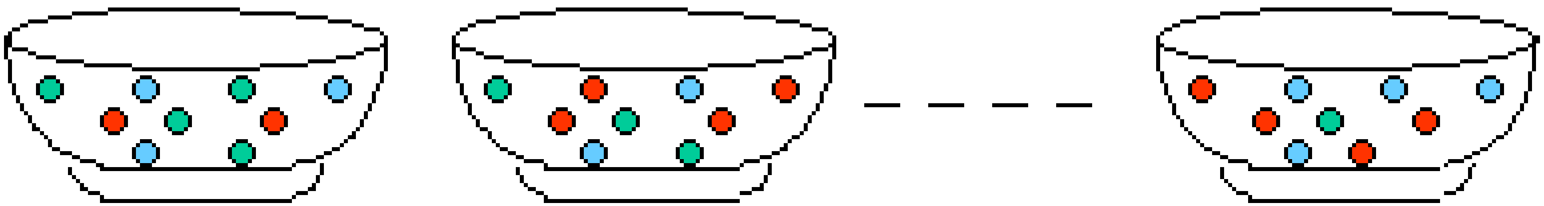


Sacamos una bola B. ¿Cuál es la probabilidad de que sea de la urna y_i?

La famosa regla de Bayes

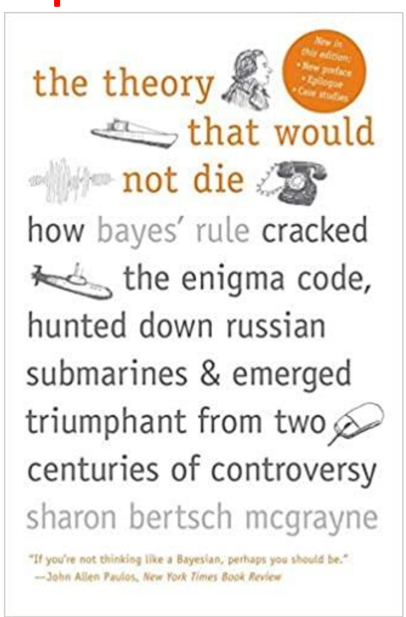
$Pr(y_i|B)=[Pr(y_i)Pr(B|y_i)]/[\Sigma_j Pr(y_j)Pr(B|y_j)]$

Un Problema
(que debería saber resolver)



¿Y si no conocemos $Pr(y_i)$ y/o $Pr(B/y_i)$?

El problema se vuelve mucho más complicado



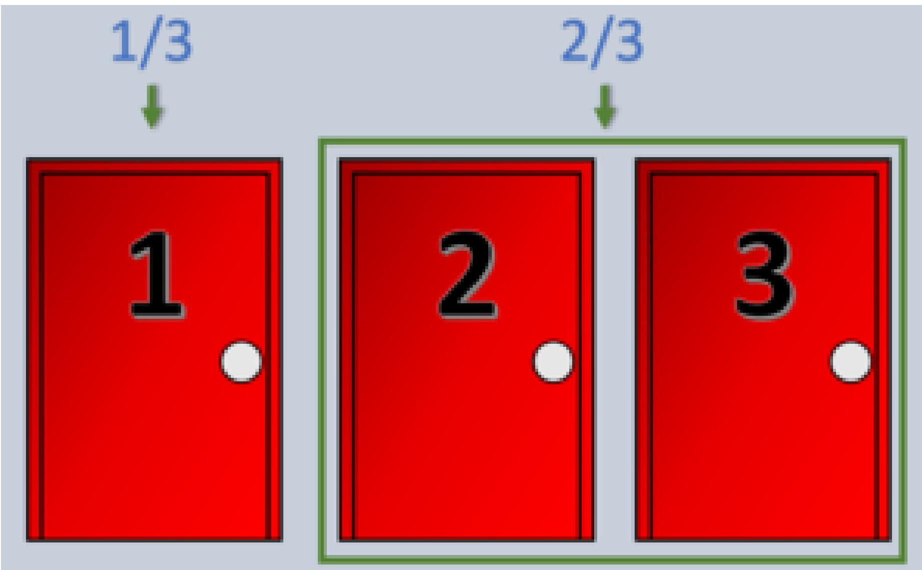
El problema de Monty Hall

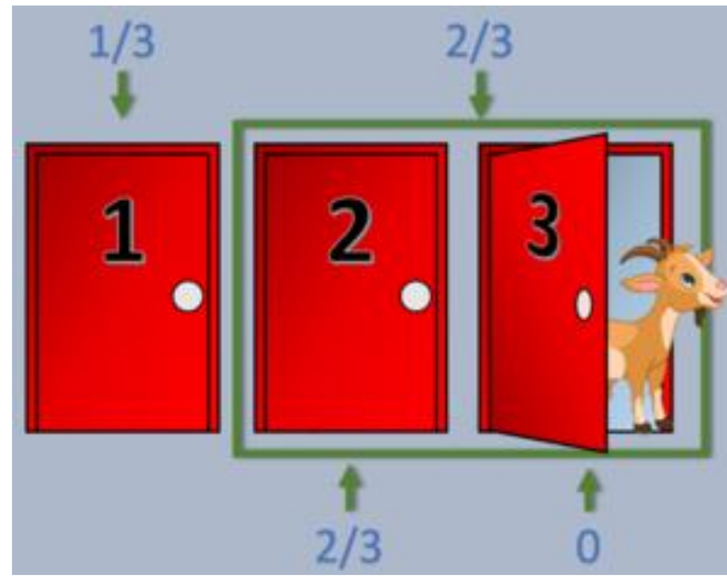
(https://es.wikipedia.org/wiki/Problema_de_Monty_Hall)

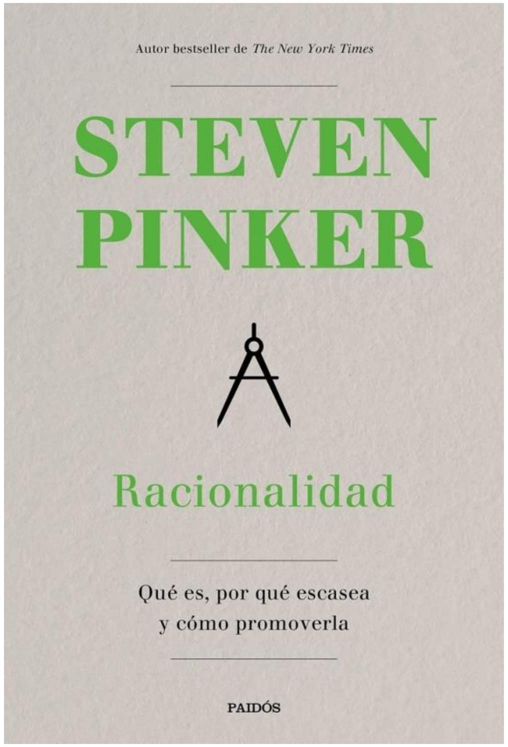
El concursante debe elegir una puerta entre tres (todas cerradas); el premio consiste en llevarse lo que se encuentra detrás de la elegida. Se sabe con certeza que tras una de ellas se oculta un coche, y tras las otras dos hay cabras (calabazas en nuestro caso).

- Una vez que el concursante haya elegido una puerta y comunicado su elección a los presentes, el presentador, que sabe lo que hay detrás de cada puerta, abrirá una de las otras dos en la que haya una calabaza.
- A continuación, le da la opción al concursante de cambiar, si lo desea, de puerta (tiene dos opciones).

¿Debe el concursante mantener su elección original o escoger la otra puerta? ¿Hay alguna diferencia?


1/3
2/3
1
2
3


1/3
2/3
1
2
3
2/3
0



- El problema de Monty Hall aparece en muchos sitios, visita el simulador del Prof. José Miguel Contreras García de la UGR.
- El problema se hizo famoso cuando Marylin vos Savant lo presentó y le dio solución en su columna **“Ask Marylin”** en la revista Parade en 1990.
- Existe una solución anterior, proporcionada en dos cartas al editor de American Statistician escritas por Steve Selvin (1975).

El problema aparece en estos dos libros que deberías leer.

Marylin vos Savant está en el libro Guiness de los Records por alcanzar la calificación más alta en un test de inteligencia.

Paul Erdös parece que necesitó de simulaciones para convencerse. ¿Tienes un número Erdös?

Algunas reacciones a la solución que correctamente había obtenido Marylin vos Savent fueron muy desafortunadas.

Visita la página de citas famosas de Marilyn vos Savent, son muy interesantes.

(https://www.azquotes.com/author/13031-Marilyn_vos_Savant)

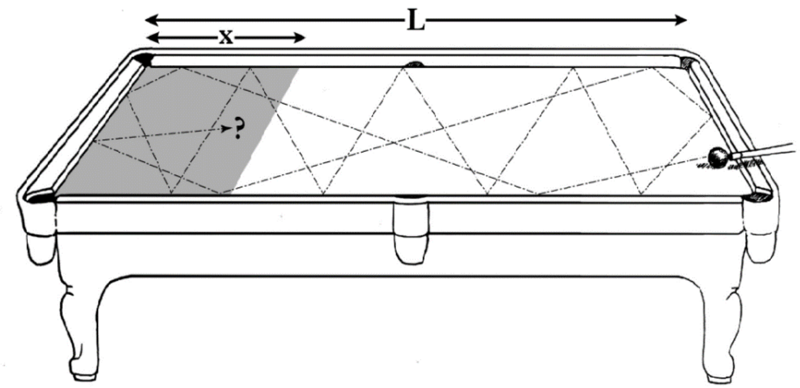
To acquire knowledge, one must study; but to acquire wisdom, one must observe.

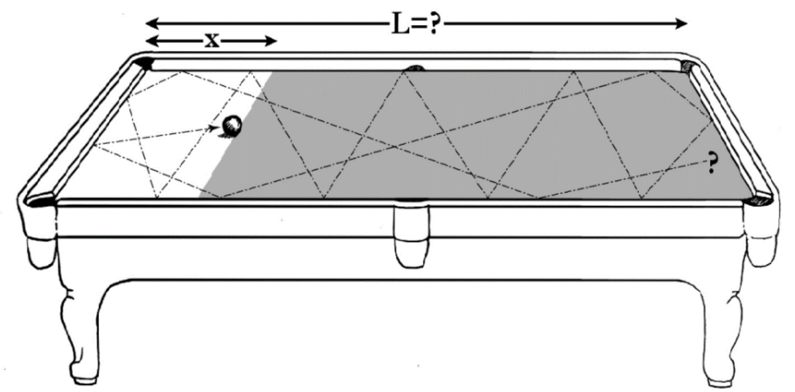
Por si queréis participar:

Desafío matemático de la Lotería de Navidad: ¿Se queda su décimo o me lo cambia?

https://elpais.com/loteria-de-navidad/2022-12-14/desafio-matematico-de-la-loteria-de-navidad-se-queda-su-decimo-o-me-lo-cambia.html 14/12/2022

Otro problema: (casi el origen de) La regla de Bayes





Versión ligeramente simplificada* del ejemplo que sugiere T. Bayes en su trabajo póstumo "An essay towards solving a Problem in the doctrine of Chance" (1763). Vió la luz gracias a Richard Price

Pregunta de probabilidad hacia delante (fácil): sabemos la longitud de la mesa y queremos calcular la probabilidad de que la bola se pare dentro de los x cms del final.

Pregunta de probabilidad hacia atrás (no tan fácil):

Habiendo observado x, ¿Cuál es la probabilidad** de que, digamos, la mesa tenga una longitud (L) de 30 mts?

Dibujo de Maayan Harel para el libro "The Book of Why".

*Ver "The Book of Why"
** densidad

Conclusiones

- Las reglas de la probabilidad son simples y compactas

- Conceptos como marginalización, distribución marginal y conjunta, la regla de Bayes y esperanza están en la base del aprendizaje automático.

- La distribución Normal se utiliza muy frecuentemente en Aprendizaje Automático. Propiedades importantes de la distribución normal son:
  - La distribución marginal de una normal es normal
  - La distribución condicionada de una normal es normal
  - El producto de distribuciones normales es proporcional a una normal
  - La normalidad se conserva tras un cambio de variable lineal

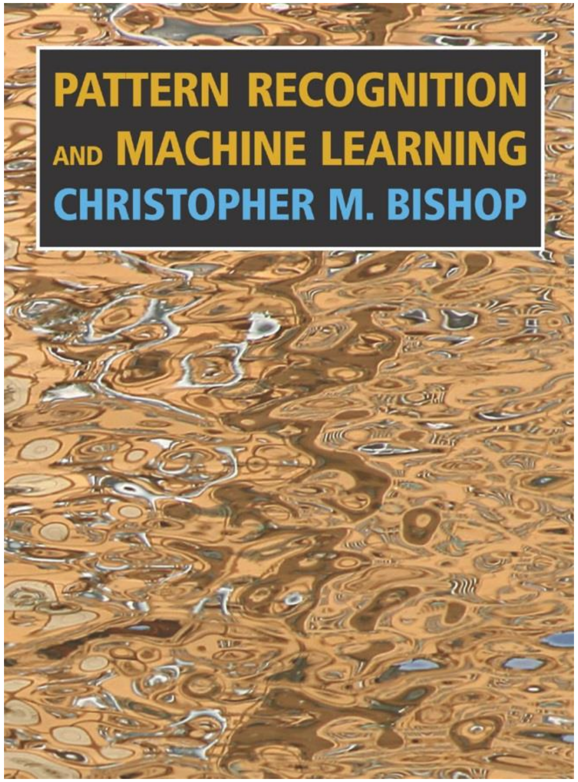
Ejercicios

Se recomienda

- Realizar los ejercicios 2.1 a 2.10 del tema II del libro http://www.computervisionmodels.com/

- Realizar los ejercicios 5.1-5.3, 5.6-5.8 del tema V del libro http://www.computervisionmodels.com/

Bibliografía



http://www.computervisionmodels.com/

http://research.microsoft.com/en-us/um/people/cmbishop/prml/