Del modelado bayesiano al machine learning
Introducción práctica en Python

Parte 1: Modelado Bayesiano con PyMC
Juan Serrano Bell (ICIFI-UNSAM)

...

Parte 2: Machine learning
Carolina del Valle Garay (OAFA)

RAAA67

Modelado Bayesiano: Introducción práctica con Pymc

Contenidos

- Modelado probabilístico
- Inferencia bayesiana
- Por qué usar Bayes
- Probabilistic programming: PyMC
- Los 3 pasos del análisis de datos bayesiano
- HANDS-ON 1: Modelo lineal
- HANDS-ON 2: Pruébelo usted mismo

¿Por qué este tutorial?

- Motivación
- Qué es este tutorial (y qué no)

Modelo probabilístico

Podemos pensarlo como un conjunto de supuestos sobre el *proceso* que genera los datos que estamos modelando.

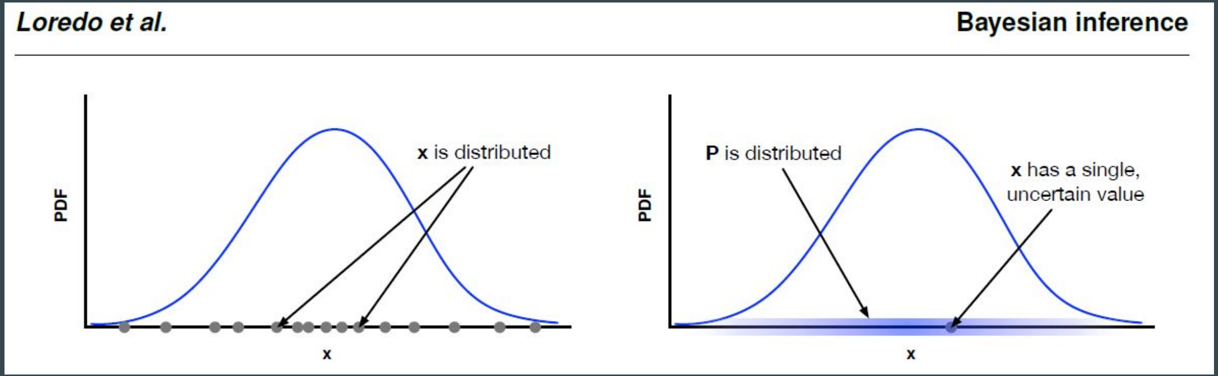
Una distribución conjunta de probabilidad de todas las variables observables y no observables del problema. Consistente con el conocimiento actual sobre el proceso que genera los datos y la forma en que fueron medidos.

Ej modelo lineal: $$Y \sim N\left(X\beta,\, \sigma^2\right)$$

Probabilidad bayesiana

- Frecuentista: Define la probabilidad como una propiedad de un *ensamble* de eventos.
  0: No sucede nunca
  1: Sucede siempre

- Bayesiana: Es una cuantificación de la incerteza, o, complementariamente, una medida de la *creencia*.
  0: No va a pasar
  1: Va a pasar



El resultado de la inferencia frecuentista es un estimador del parámetro.

El resultado de la inferencia bayesiana es una distribución de probabilidad, o de creencias respecto a los posibles valores del parámetro..

Inferencia Bayesiana

Permite expresar la distribución de probabilidad a posteriori de las hipótesis sobre un proceso de generación de datos como el producto (normalizado) de las probabilidades a priori y la función de verosimilitud.

$$P(A, B) = P(A \mid B)\, P(B)$$
$$P(A, B) = P(B \mid A)\, P(A)$$

$$\rightarrow P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Teorema de Bayes

Creencia que sobrevive

Predicción

Creencia previa

Creencia posterior

$$P\left(H_i \mid D_{obs}\right) = \frac{P\left(D_{obs} \mid H_i\right) P\left(H_i\right)}{P\left(D_{obs}\right)}$$

Evidencia (cte. norm)

$$P\left(H_i\right)$$

Prior probability ("prior to accounting for the observed data")

$$P\left(D_{obs} \mid H_i\right)$$

Likelihood for the hypothesis H_i. Quantifies how well each hypothesis predicts the observed data.

$$P\left(D_{obs}\right)$$

Prior predictive probability or marginal likelihood (also called evidence).

$$P\left(H_i \mid D_{obs}\right)$$

Posterior probability for hypothesis H_i. ("posterior: after taking into account the observed data")

Si consideramos un continuo de hipótesis...

$$\overbrace{P(\theta \mid D_{obs})}^{\text{Posterior}} = \frac{\overbrace{P(D_{obs} \mid \theta)}^{\text{Likelihood}}\ \overbrace{P(\theta)}^{\text{Prior}}}{\underbrace{P(D_{obs})}_{\text{Evidencia}}}$$

No depende de Θ

Integral multidimensional (generalmente intractable)

$$P(D_{obs}) = \int P(D_{obs} \mid \theta)\, P(\theta)\, d\theta$$

Si el objetivo es inferencia →

$$\boxed{P(\theta \mid D_{obs}) \propto P(D_{obs} \mid \theta)\, P(\theta)}$$

Posterior $\propto$ Likelihood * Prior

Tanto los datos como las hipótesis son variables, lo que las relaciona es un modelo:

$$P(\theta \mid D_{obs}, M) = \frac{P(D_{obs} \mid \theta, M)\, P(\theta \mid M)}{\underbrace{P(D_{obs} \mid M)}}$$

Es la predicción que hace el modelo
de los datos, con la contribución de
todas las hipótesis

El modelo que usamos también es una hipótesis.
¿Por qué no tener una distribución de creencias sobre los distintos modelos?

Evidencia = Likelihood

$$P(M \mid D_{obs}) = \frac{P(D_{obs} \mid M)\, P(M)}{P(D_{obs})}$$

Comparación de modelos

$$P(M_1 \mid D_{obs}) = \frac{P(D_{obs} \mid M_1)\, P(M_1)}{P(D_{obs})} \qquad P(M_2 \mid D_{obs}) = \frac{P(D_{obs} \mid M_2)\, P(M_2)}{P(D_{obs})}$$

$$\frac{P(M_1 \mid D)}{P(M_2 \mid D)} = \frac{P(D \mid M_1)\, P(M_1)}{P(D \mid M_2)\, P(M_2)}$$

Cociente de las evidencias

Bayes Factor

$$\boxed{\frac{P(M_1 \mid D)}{P(M_2 \mid D)} = \frac{P(D \mid M_1)}{P(D \mid M_2)}}$$

Jeffreys scale

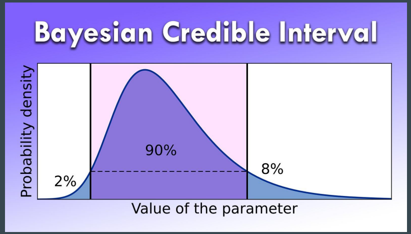
1 -> No hay evidencia
1-3.2 -> Débil
3.2-10 -> Moderada
10-100 -> Fuerte
>100 -> Decisiva

¿Por qué usar Bayes?

- Es intuitivo y natural a nuestra forma de conocer el mundo.
- Incorporar conocimiento previo, gran ventaja cuando tengo pocos datos.
- Tratamiento de la incerteza en los parámetros.
- Versatilidad en el modelado.

Teñer una distribución de probabilidad (posterior) permite aplicar las leyes de la probabilidad para encarar una gran variedad de tareas del análisis de datos.

HPD (highest posterior density)



- Credible regions:

$$p\,(\theta \,\epsilon\, R \,|\, D_{obs}) \;=\; \int_R d\theta\, p\,(\theta \,\epsilon\, R \,|\, \theta)\, p\,(\theta \,|\, D_{obs})$$

- Predicciones sobre datos futuros

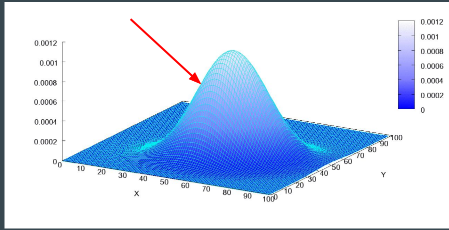
$$p\left(D' \,\middle|\, D_{obs}\right) \;=\; \int d\theta\, p\left(D' \,\middle|\, \theta\right) p\,(\theta \,|\, D_{obs})$$

Posterior predictive distribution

- Marginalizar sobre parámetros molestos (*nuisance)*
- Propagación de errores bayesiana
- Comparación de modelos
- Model averaging

Aproximando la posterior

Probabilidad



Prior space
N-dimensional

Cuando condicionamos a los datos observados, el espacio no cambia, lo que cambia es la *superficie*, reflejando la probabilidad posterior.

Simulation methods -> MCMC

Queremos explorar el espacio de la posterior para buscar el “pico” (nos interesa el área que lo rodea). Si queremos solo el pico, existe el método MAP.

La idea detrás de MCMC es hacer una búsqueda inteligente (eso esperamos) del espacio, recogiendo muestras a medida que lo exploramos. Una buena cantidad de muestras (del orden de las miles) nos permitirán aproximar la posterior verdadera.

MCMC es una familia de algoritmos

Markov Chain Monte Carlo

Pseudocódigo

1 - Start at the current position.

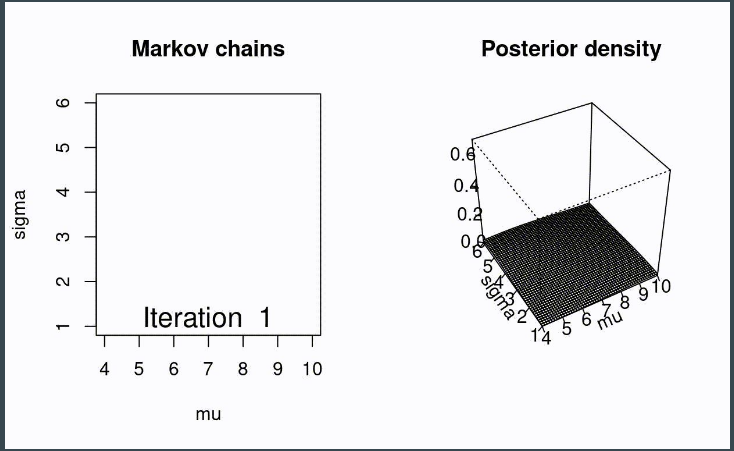
2 - Propose moving to a new position.

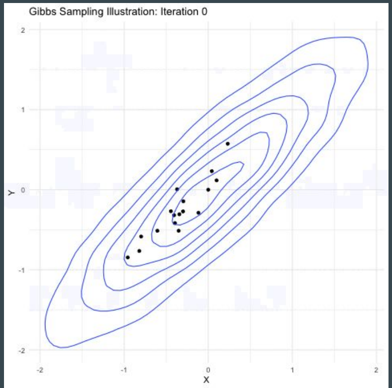
3 - Accept/Reject the new position based on the position’s adherence to the data and prior distributions.

4 - a) I you accept: Move to the new position. Return to Step 1.

b) Else: Do not move to the new position. Return to Step 1.

5 - After a large number of iterations, return all accepted positions.





Hamiltonian Monte Carlo

Less randomness -> More efficiency

Utiliza una analogía física, simulando una partícula moviéndose sin fricción en un espacio N-dimensional.

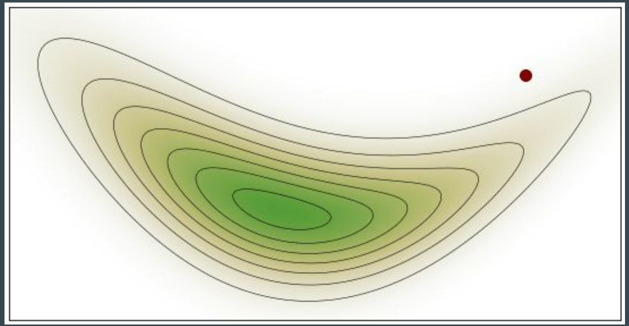
-> Propuestas más inteligentes, utiliza más información de la distribución. Variable auxiliar: momento.

-> Requiere computar el gradiente. Limitado a distribuciones continuas.

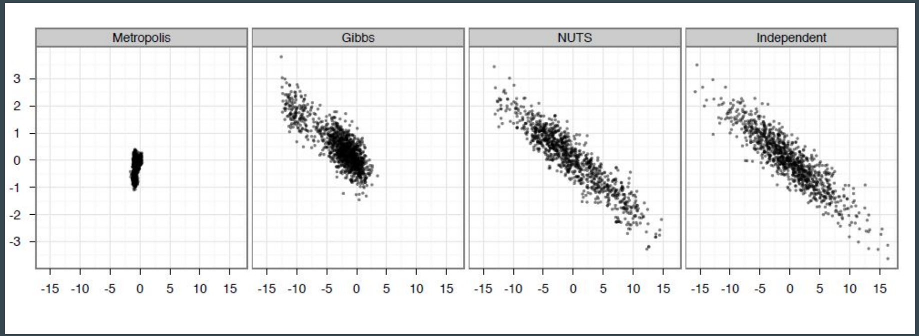
-> Mucho más eficiente, necesita menos muestras para lograr un buen sampleo.

-> Más costoso computacionalmente.

-> Brilla en modelos de alta dimensionalidad y con correlaciones.



NUTS: variante de HMC con auto-tuning



“Probabilistic programming”

PyMC

Permite construir modelos bayesianos y hacer inferencia con MCMC de forma muy fácil con una sintaxis de alto nivel.

Permite concentrarse en el problema y no en los métodos.

Objetos estocásticos

Distribuciones sobre valores

```
X ~ Normal(μ, σ)
x = X.random(n=100)
```

Distribuciones sobre funciones

```
Y ~ GaussianProcess(mean_func(x), cov_func(x))
y = Y.predict(x2)
```

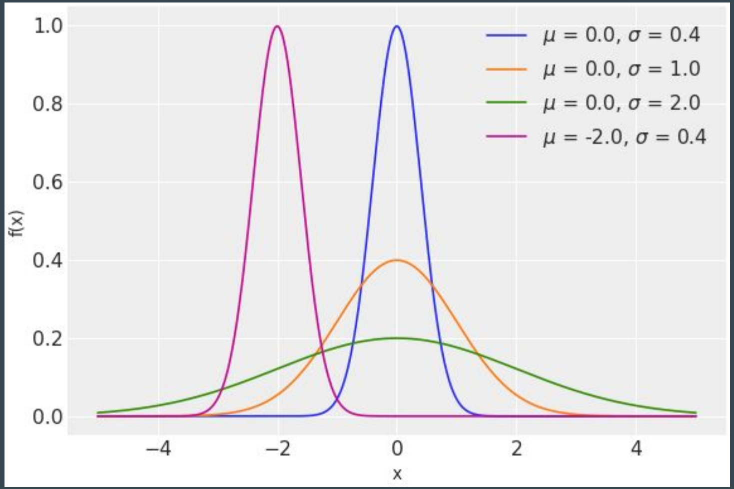
Los 3 pasos del modelado bayesiano

1. Construir el modelo probabilístico, una distribución conjunta de probabilidad para todas las variables observables y no observables.

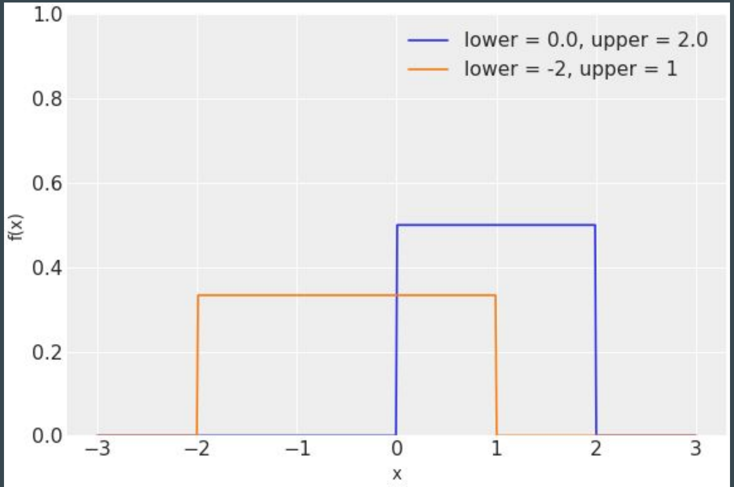
Los 3 pasos del modelado bayesiano

1. Construir el modelo probabilístico, una distribución conjunta de probabilidad para todas las variables observables y no observables.

PRIORS $P(\theta)$ Representan la información previa que tenemos de los parámetros.



$$\theta \sim N(\mu, \sigma)$$



$$\theta \sim U(a, b)$$

Los 3 pasos del modelado bayesiano

1. Construir el modelo probabilístico, una distribución conjunta de probabilidad para todas las variables observables y no observables.

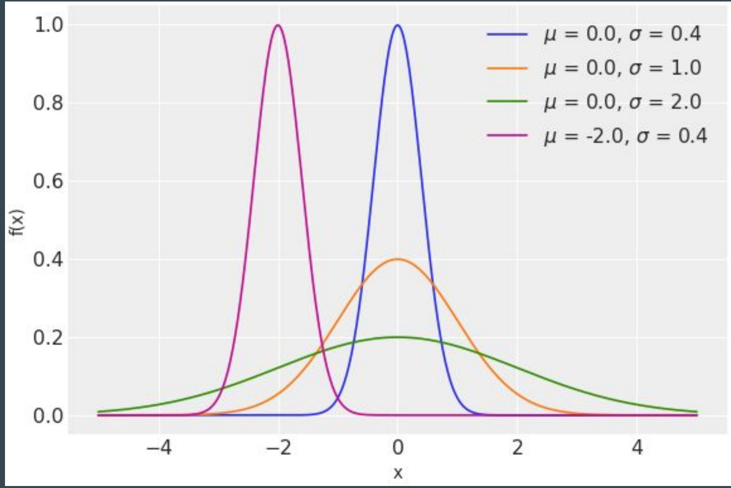
LIKELIHOOD

$$P\left(D_{obs} \mid \theta\right)$$

Condiciona el modelo a los datos observados.

Usualmente Gaussiana

$$x \sim N\left(\mu, \sigma\right)$$



Los 3 pasos del modelado bayesiano

2. Condicionar el modelo en las variables observadas, obteniendo la posterior (MCMC sampling).

Botón de inferencia:

```
idata = pymc.sample()
```

Los 3 pasos del modelado bayesiano

3. Evaluar el ajuste!

   - Convergence checks
   - Inspección visual
   - Inferencia de parámetros
   - Son las conclusiones razonables?
   - Reparametrizar? Ajustar priors? Volver al punto 1.

Fuentes:

- Bayesian inference: More than Bayes's theorem.(Loredo et al. 2024)
- Bayesian Methods for Hackers (Davidson-Pilon)
- Bayesian Data Analysis 3rd Ed. (Gelman et al.)
- Bayesian modeling and computation in Python (O. Martin et al.)
- Documentación PyMC

HANDS-ON