

# Doing bayesian data analysis by Kruschke

Revisión

J. E. Alcalá

CEIC, Universidad de Guadalajara

4 de abril de 2019

## 1. Capítulo 2

**Ideas fundacionales:**

1. Reasignación de credibilidad a través de las posibilidades.
2. Las posibilidades se representan como parámetros de modelos matemáticos.

### 1.1. Reasignación a través de posibilidades

#### **Ejemplo 1:**

**A.** Observación: camino mojado.

**B.** Posibilidades-creencias: {llovió, borracho tiró su bebida, etc.}.

**C.** Nuevas observaciones: {humedad se extiende a lo lejos, humedad está localizada}.

**D.** Reasignación: la ocurrencia de un elemento del conjunto en **C** modifica la credibilidad en alguno de los elementos de **B** para los cuales el elemento en **C** podría haberlo causado.

En el **Ejemplo 1**, observar que la humedad se extiende a lo lejos podría reducir las posibilidades de “borracho tiró su bebida”, por otro lado, observar que la humedad está localizada reduce las posibilidades de “llovió”.

Las creencias depositadas en los elementos de **A** son *a priori* en el sentido de que, en el momento en el que se enuncian (junto con un valor numérico entre 0 y 1 de acuerdo al *grado* en el que se crea en ellas; la suma de la credibilidad debe dar 1), no se ha recolectado evidencia de su ocurrencia.

Las nuevas observaciones son datos. La reasignación de credibilidad es *a posteriori* porque sucede después de haber recolectado datos. Dado que estos datos redistribuyen

los valores numéricos (entre 0 y 1), una vez recolectamos datos no creemos en todas las posibilidades al mismo grado que antes de la recolección.

La credibilidad posterior puede servir como credibilidad *a priori*, es decir, el conjunto  $\mathbf{D} \rightarrow \mathbf{B}'$ , y el proceso se repite, lo que se conoce como *actualización bayesiana*.

### 1.1.1. Datos ruidosos: inferencia probabilística

Los datos tienen relaciones *probabilísticas* con las causas subyacentes.

#### Datos *ruidosos*

1. Los datos son producto de mediciones imperfectas.
2. Cada dato resultado de una medición imperfecta es diferente a otros datos. E.g.,  $x_1 \neq x_2 \neq x_3$ .
3. Además de las mediciones imperfectas, los datos en sí pudieron ser generados por un proceso no determinístico.
4. Esta variabilidad acota la exactitud con la que inferimos *qué* pudo haber generado esos datos (i.e., cuál es la causa subyacente).

#### **Ejemplo 2:**

La tasa de respuesta en un programa IV 30 s varía de animal en animal, sesión a sesión en el mismo animal y bajo diferentes condiciones, como grado de privación o tipo de reforzador utilizado.

Además, el operando puede descalibrarse, no detectar respuestas muy contiguas, o respuestas dadas con poca fuerza, etc.

#### **Nota:**

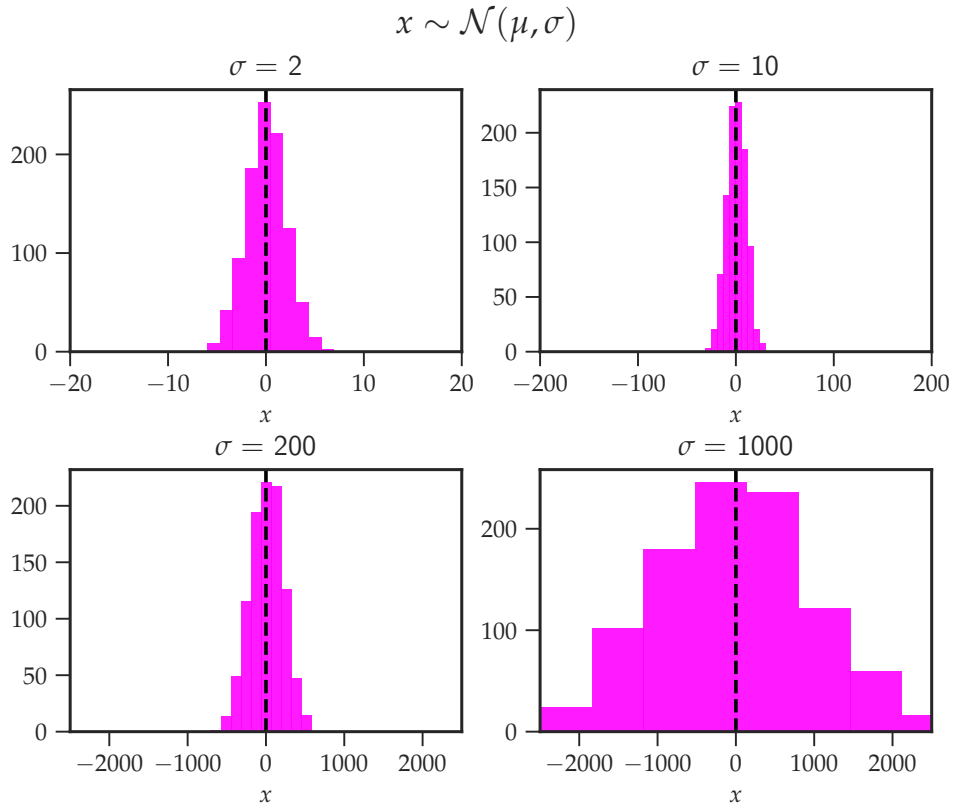
☞ Se asume que existe un proceso ‘verdadero’ que da lugar a la variable observada. Por ejemplo, que la tasa es  $\mathcal{R}$ , pero obtenemos  $\hat{R}$ , un estimador de la *tasa verdadera*, o la tasa que se tendría si no hubiera errores de medición.  $\mathcal{R}$  podría ser una variable aleatoria, pero los errores de medición añaden **incertidumbre** a la *forma* que esa variable toma.

#### **Resumen:**

1. La esencia de la inferencia bayesiana es la reasignación de la credibilidad.
2. La credibilidad inicial (*a priori*) refleja el conocimiento previo sobre las posibilidades, y puede ser muy vago (en cuyo caso no se beneficia a alguna posibilidad por encima del resto).
3. Las posibilidades inconsistentes con los datos pierden credibilidad.

### 1.2. Las posibilidades son parámetros de modelos

Dado que las mediciones producen datos variables, la mejor forma de representar las creencias del proceso generador de datos es con probabilidad.



*Figura 1:* Distribuciones normales con mismo  $\mu$  pero distinto  $\sigma$ , que refleja diferentes grados de ‘vaguedad’. Conforme  $\sigma \rightarrow \infty$ , las posibilidades son equiprobables. Notar que los ejes en  $x$  cambian de límites. En el panel inferior derecho, si restringimos los valores de  $x \in [-100, 100]$  la distribución se vería casi plana.

Distintas distribuciones (o densidades) de probabilidad implica que se debe elegir cuál de ellas (qué modelo) representa mejor las creencias. Luego de elegir el modelo probabilístico (e.g., distribución normal, binomial, etc), se debe elegir qué *qué parámetros* determinan la distribución (e.g., en distribución normal son  $\mu$  y  $\sigma$ ).

- ¿Cuánta confianza tenemos en las posibilidades sobre las que depositamos nuestras creencias? Si no es mucha, nuestras creencias son *vagas*.
- Si las creencias son vagas, debemos escoger un modelo *a priori* acorde, por ejemplo, uno donde todas las posibilidades tengan la misma probabilidad de ser verdad (e.g., distribución uniforme, o distribución normal con una desviación estándar muy (muy) grande; ver la Figura 1).

### 1.3. Pasos de la inferencia bayesiana

1. Identificar qué datos (o variables) responden a una pregunta.
2. Definir modelo descriptivo para los datos, con la forma matemática adecuada.
3. Especificar la distribución *a priori* de los parámetros.
4. Usar inferencia bayesiana para reasignar credibilidad a los parámetros.
  - Interpretar la distribución posterior con respecto a la teoría.
5. Revisar que las predicciones posteriores se asemejen a los datos con precisión razonable (*posterior predictive check*). Si no, volver al paso 2.