Bayesian Methods

Peter Komar

November 24, 2018

Contents

1 Foundations

1.1 Definitions, identities

Notation

- Upper-case letters (A, B, C, X, Y): random variables
- Lower-case letters (a, b, c, x, y): real numbers
- $P(X = x) \ =: \ P(x)$
- $P(X = x \text{ and } Y = y) \ =: \ P(x, y)$
- $P(A = a, \text{ given } B = b) \ =: \ P(a \mid b)$
- $\int_{\infty}^{+\infty} [\ldots] da \ =: \ \sum_{a \in \mathbb{R}} [\ldots] \ =: \ \sum_a [\ldots]$

Conditional probability

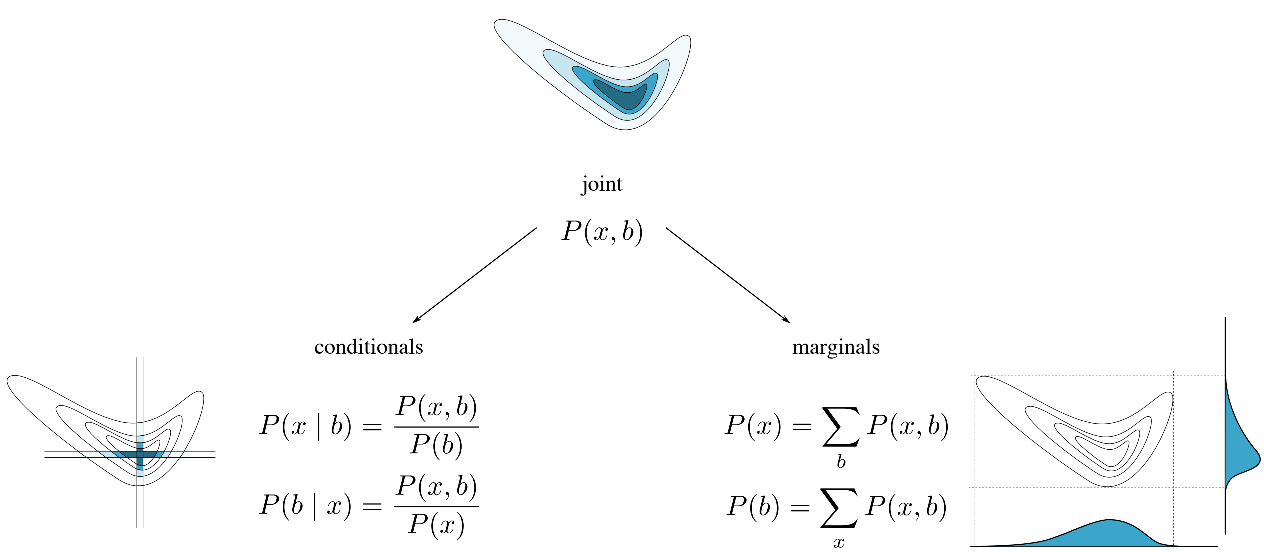
- $P(a \mid b) \ = \ \frac{P(a,b)}{P(b)}$
- $P(a, b) \ = \ P(a \mid b) \, P(b)$
- $P(a, b \mid c) \ = \ P(a \mid b, c) \, P(b \mid c)$
- $\sum_a P(a \mid b) \ = \ 1, \quad$ but $\ \sum_b P(a \mid b) \neq 1$, in general
- $\sum_b P(a \mid b) \, P(b) \ = \ \sum_b P(a, b) \ = \ P(a)$

Marginal

- $P(a) \ = \ \sum_b P(a \mid b) \, P(b)$

Bayes theorem

- $P(b \mid x) \ = \ \frac{1}{P(x)} P(x \mid b) \, P(b)$



1.2 Bayesian inference

Prior, likelihood, posterior

- Data: $D = \{x_1, x_2, \ldots x_n\}$, independent measurements.
- Model: M with θ: parameter(s) to estimate
- Prior: $P(\theta)$
- Likelihood: $P(D \mid \theta) = P(d_1 \mid \theta)\, P(d_2 \mid \theta)\ \ldots\ P(d_n \mid \theta) = \prod_{i=1}^{n} P(d_i \mid \theta)$
- Unnormalized posterior: $P^*(\theta \mid D) \ = \ P(D \mid \theta)\, P(\theta)$
- Normalization: $Z = \sum_\theta P^*(\theta \mid D)$
- Posterior: $P(\theta \mid D) = \frac{1}{Z} P^*(\theta \mid D)$

Example

"Three light bulbs of the same make lasted 1, 2 and 5 months of continuous use. Let us estimate the lifetime of this kind of light bulb."

- $D = \{t_1, t_2, t_3\} = \{1, 2, 5\}$
- M: Light bulbs have average lifetime of T months.
- $P(T) = \frac{1}{1000}$, uniform on [0,1000].
- $P(t \mid T) = \frac{1}{T} \exp\left(-\frac{t}{T}\right)$
- $P(D \mid T) = \prod_i P(t_i \mid T) = \prod_{i=1}^{3} \frac{1}{T} \exp\left(-\frac{t_i}{T}\right) = \frac{1}{T^3} \exp\left(-\frac{1+2+5}{T}\right)$
- $P^*(T \mid D) = \frac{1}{T^3} \exp\left(-\frac{8}{T}\right)$
- Z and $P(T \mid D)$ can be determined numerically:

```python
import numpy as np

T_arr = np.linspace(0.1, 1000, 10_000)
Pstar_arr = 1.0/T_arr**3 * np.exp(-8/T_arr)
Z = np.sum(Pstar_arr)
P_arr = Pstar_arr / Z
```

  yielding $Z = 0.1562$

- $\mathbb{E}(T \mid D) = \sum_T T\ P(T \mid D)$
- $\mathrm{std}(T \mid D) = \sqrt{\sum_T (T - \mathbb{E}(T))^2\ P(T \mid D)}$

```python
T_ev = np.sum(T_arr * P_arr)
T_std = np.sqrt(np.sum((T_arr - T_ev)**2 * P_arr))
```

  yielding $\mathbb{E}(T \mid D) = 7.937, \quad \mathrm{std}(T \mid D) = 14.48.$

1.3 Model comparison

New definition: Evidence

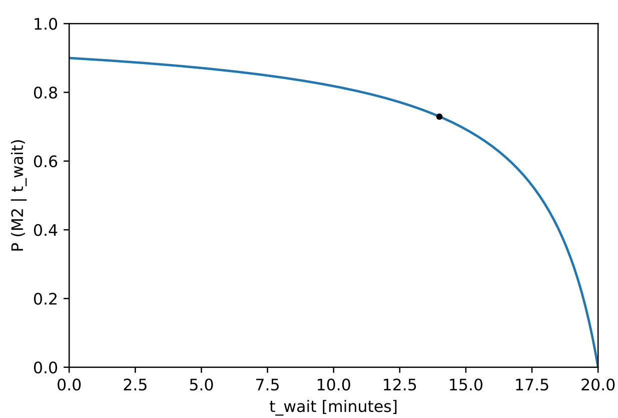
- Data: D
- Model 1: M_1 with parameter θ_1 and prior $P(\theta_1 \mid M_1)$, and likelihood $P(D \mid \theta_1, M_1)$
- Model 2: M_2 with parameter θ_2 and prior $P(\theta_2 \mid M_2)$, and likelihood $P(D \mid \theta_2, M_2)$
- Prior on models: $P(M_1) = 0.5, \quad P(M_2) = 0.5$.
- Evidence for each model: $P(D \mid M_i) = \sum_\theta P(D \mid \theta_i, M_i) P(\theta_i \mid M_i)$
- Unnormalized posterior: $P^*(M_1 \mid D) = P(D \mid M_1)\, P(M_1)$, and $P^*(M_2 \mid D) = P(D \mid M_2)\, P(M_2)$
- Normalization: $Z = P^*(M_1 \mid D) + P^*(M_2 \mid D)$.

Example

"Waiting for my baggage at the airport carousel, there are two possibilities: 1) It could miss the plane, and will never come, or 2) It was on the plane and it had a 1/20 chance of arriving within any of the 1-minute intervals between 0 and 20 minutes. Now, given what is the posterior probability of model 2 given that 14 minutes has passed and the bas has not arrived?"

- $D = \{\text{Bag has not arrived after } t_\text{wait} = 14 \text{ minutes}\}$
- M_1: It never arrives, $P(D \mid M_1) = 1$
- M_2: It shows up some time between 0 and 20 minutes, $P(t_\text{bag} \mid M_2) = 1/20$ for $t_\text{bag} \in [0, 20]$, and the likelihood is $P(D \mid t_\text{bag}, M_2) = 1$, if $t_\text{bag} > t_\text{wait}$, and 0 otherwise.
- $P(M_1) = 0.1, \quad P(M_2) = 0.9$
- $P(D \mid M_1) = 1$
- $P(D \mid M_2) = \sum_{t_\text{bag}} P(D \mid t_\text{bag}, M_2)\, P(t_\text{bag} \mid M_2) = \sum_{t_\text{bag}} [t_\text{bag} > 14] \times \frac{1}{20} = \frac{20-14}{20}$
- $P^*(M_1 \mid D) = 1 \times 0.1, \quad P^*(M_2 \mid D) = \frac{20-14}{20} \times 0.9$
- $Z = 0.1 + \frac{3}{10} \times 0.9 = 0.37$
- $P(M_2 \mid D) = P^*(M_2 \mid D)/Z = 0.7297$.

We can also plot $P(M_2 \mid t_\text{wait})$ for all waiting times between 0 and 20 minutes.



1.4 Prediction

New definition: Predictive distribution

- Data: $D = \{x_1, x_2, \dots x_n\}$
- Model: M with parameter θ, prior $P(\theta)$ and likelihood $P(x \mid \theta)$
- Posterior: $P(\theta \mid D) = P^*(\theta \mid D)/Z = \dots$ (see previous sections)
- Predictive distribution: $P(X_{n+1} = x \mid D) = \sum_\theta P(x \mid\theta)\ P(\theta \mid D)$
- Customized prediction: $P(f(\theta) \mid D) = \sum_\theta f(\theta)\ P(\theta \mid D)$

Example

"Two player, A and B are playing a game of luck, where at the beginning of the game a ball is rolled on a pool table to divide the table in two un-equal halves: A's side and B's side. In each subsequent round, a ball is rolled. A point is given to the player on whose side the ball stops. A and B are playing this game until one of them reaches 6 points. The current score is 5 to 3 in favor of A. What is the chance that A will win this game?"

- $D = \{n_A = 5, n_B = 3\}$
- M, first ball: $P(b) = 1$ in $[0, 1]$
- $P(\text{A scores} \mid b) = b$
- $P(D \mid b) = \text{Binomial}(5 \mid 5 + 3,\ b)$
- $P^*(b_0 \mid D) = \text{Binomial}(5 \mid 8, b) \times 1$
- $Z = \sum_b \text{Binomial}(5 \mid 8, b)$ can be calculated numerically

```
import numpy as np
from scipy.stats import binom

b_arr = np.linspace(0, 1, 1000)
Pstar_arr = binom.pmf(5, 8, b_arr)
Z = np.sum(Pstar_arr)
```

- $P(\text{A wins} \mid b, D) = 1 - P(\text{B wins} \mid b, D) = 1 - (1 - b)^3 = f(b)$
- $P(\text{A wins} \mid D) = \sum_b f(b) P^*(b \mid D)/Z$

```
P_arr = Pstar_arr / Z
P_Awins = np.sum((1 - (1 - b_arr)**3) * P_arr)
```

yielding $P(\text{A wins} \mid D) = 0.909$