NA 568 - Winter 2024

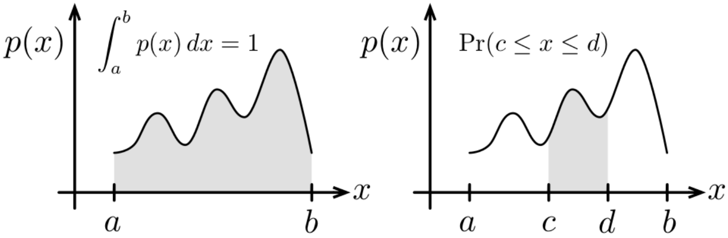
Bayes Filters

Minghan Zhu and Maani Ghaffari

January 12, 2024

UNIVERSITY OF MICHIGAN

Probability Density Functions



Courtesy: T. Barfoot

Joint and Conditional Distribution

Let X and Y be two random variables.

- The joint distribution of X and Y is:
  $p(x,y) = p(X = x \text{ and } Y = y)$;

- The conditional probability of X given Y is:
  $p(x|y) = \frac{p(x,y)}{p(y)} \qquad p(y) > 0.$

- If X and Y are independent then $p(x,y) = p(x)p(y)$

Marginalization

- Given $p(x,y)$, the marginal distribution of X can be computed by summing (integration) over Y.

$$p(x) = \sum_{y \in \mathcal{Y}} p(x,y)$$

- The law of total probability is its variant which uses the conditional probability definition

$$p(x) = \sum_{y \in \mathcal{Y}} p(x|y)p(y)$$

and for continuous random variables, it is

$$p(x) = \int_{y \in \mathcal{Y}} p(x,y)dy = \int_{y \in \mathcal{Y}} p(x|y)p(y)dy$$

- $p(x,y) = p(x|y)p(y) = P(y|x)p(x)$
- $$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\sum_{x \in \mathcal{X}} p(y|x)p(x)}$$
  $$p(\text{hypothesis}|\text{data}) = \frac{p(\text{data}|\text{hypothesis})p(\text{hypothesis})}{p(\text{data})}$$
- $$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence (Marginal Likelihood)}}$$

Causal vs. Diagnostic Reasoning

- $p(\text{hypothesis}|\text{data})$ is **diagnostic**.
- $p(\text{data}|\text{hypothesis})$ is **causal**.
- Often causal knowledge is easier to obtain.
- Bayes rule allows us to use causal knowledge for diagnostic reasoning:

$$p(\text{hypothesis}|\text{data}) = \frac{p(\text{data}|\text{hypothesis})p(\text{hypothesis})}{p(\text{data})}$$

Example

An autonomous car is approaching a traffic light which can be either green, yellow, or red. The car is programmed to be conservative and thus it will stop if it detects a yellow or red light; otherwise it will continue driving. Previous tests have demonstrated that due to sensor imperfections, the car will drive through (without stopping) 10% of yellow lights, 95% of green lights, and 1% of red lights. The traffic light is on a continuous cycle (30 seconds green, 5 seconds yellow, 25 seconds red). You are riding in the car and are busy working on your Mobile Robotics project (i.e., not watching the road, light, etc.). You feel the car stop as it approaches the traffic light described above. What is the probability that the traffic light was yellow when the vehicle sensed it?

Answer

Let S represent the event that the vehicle stopped, G the event that the light was green, Y that it was yellow, R that is was red.

- *Given:* $P(S|Y) = 0.90, P(S|G) = 0.05, P(S|R) = 0.99, P(Y) = 5/60, P(R) = 25/60, P(G) = 30/60$
- *Find:* $P(Y|S)$

Answer

Let S represent the event that the vehicle stopped, G the event that the light was green, Y that it was yellow, R that is was red.

- *Given:* $P(S|Y) = 0.90, P(S|G) = 0.05, P(S|R) = 0.99, P(Y) = 5/60, P(R) = 25/60, P(G) = 30/60$
- *Find:* $P(Y|S)$

$$P(Y|S) = \frac{P(S|Y)P(Y)}{P(S)}$$

$$P(Y|S) = \frac{P(S|Y)P(Y)}{P(S|Y)P(Y) + P(S|R)P(R) + P(S|G)P(G)}$$

$$P(Y|S) = \frac{0.90(5/60)}{0.90(5/60) + 0.99(25/60) + 0.05(30/60)} = 14.63\%$$

Bayes' Rule with Prior Knowledge

- Given three random variables X, Y, and Z, Bayes' rule relates the prior probability distribution, $p(x|z)$, and the likelihood function, $p(y|x,z)$, as follows.

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

- Given Z, if X and Y are **conditionally independent** then $p(x,y|z) = p(x|z)p(y|z)$

Example

Height and vocabulary are not independent; but they are conditionally independent if you add age.

https://en.wikipedia.org/wiki/Conditional_independence#Examples

The univariate (one-dimensional) *Gaussian (or normal) distribution* with mean μ and variance σ^2 has the following Probability Density Function (PDF).

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2})$$

We often write $X \sim \mathcal{N}(\mu, \sigma^2)$ or $\mathcal{N}(x; \mu, \sigma^2)$ to imply that X follows a Gaussian distribution with mean $\mu = \mathbb{E}[X]$ and variance $\sigma^2 = \mathbb{V}[X]$.

Multivariate Normal Distribution

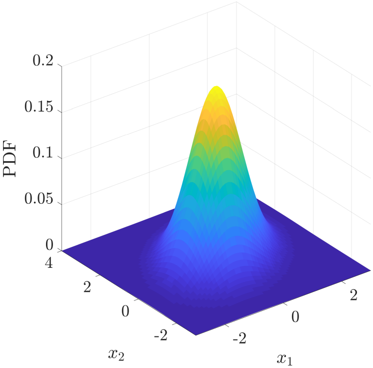
The multivariate Gaussian (normal) distribution of an n-dimensional random vector $X \sim \mathcal{N}(\mu, \Sigma)$, with mean $\mu = \mathbb{E}[X]$ and covariance $\Sigma = \mathrm{Cov}[X] = \mathbb{E}[(X-\mu)(X-\mu)^\mathsf{T}]$ is

$$p(x) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2}(x-\mu)^\mathsf{T}\Sigma^{-1}(x-\mu))$$

Visualizing multivariate Gaussian

Let $x = \mathrm{vec}(x_1, x_2)$ and $X \sim \mathcal{N}(\mu, \Sigma)$ where

$$\mu = \begin{bmatrix} 0.0 \\ 0.5 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.8 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}$$



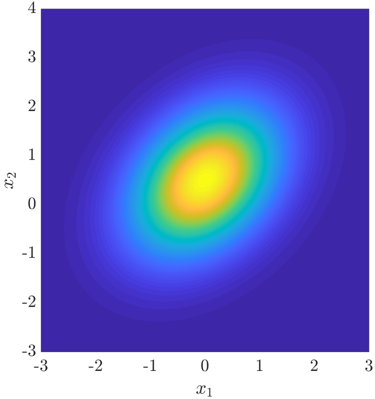


Figure: Left, two-dimensional PDF; right, top view of the first plot.

Marginalization and Conditioning of Normal Distribution

Let X and Y be jointly Gaussian random vectors

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} A & C \\ C^\mathsf{T} & B \end{bmatrix})$$

then the marginal distribution of X is

$$X \sim \mathcal{N}(\mu_x, A)$$

and the conditional distribution of X given Y is

$$X|Y_{=y} \sim \mathcal{N}(\mu_x + CB^{-1}(y - \mu_y), A - CB^{-1}C^\mathsf{T})$$

Visualizing multivariate Gaussian

Let $x = \mathrm{vec}(x_1, x_2)$ and $X \sim \mathcal{N}(\mu, \Sigma)$ where

$$\mu = \begin{bmatrix} 0.0 \\ 0.5 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.8 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}$$

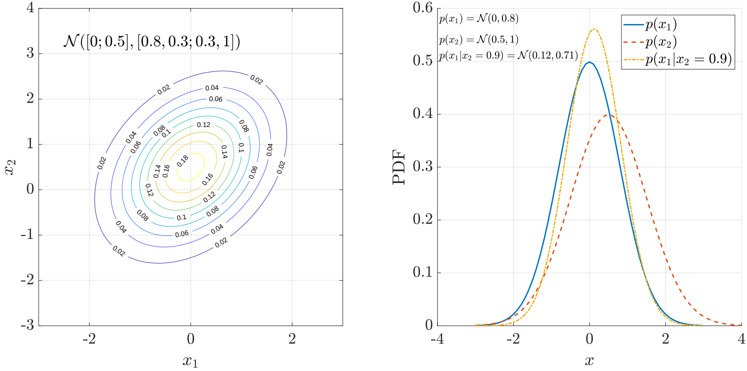


Figure: Left, the contour plot of the PDF; right, the marginals and the conditional distribution of $p(x_1|x_2 = 0.9)$.

Affine Transformation of a Multivariate Gaussian

Suppose $X \sim \mathcal{N}(\mu, \Sigma)$ and $Y = AX + b$.

Then $Y \sim \mathcal{N}(A\mu + b, A\Sigma A^\mathsf{T})$.

$$\mathbb{E}[Y] = \mathbb{E}[AX + b] = A\mathbb{E}[X] + b = A\mu + b$$

$$\begin{aligned}
&\mathrm{Cov}[Y] = \mathbb{E}[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^\mathsf{T}] \\
&= \mathbb{E}[(AX - A\mu)(AX - A\mu)^\mathsf{T}] = A\mathbb{E}[(X - \mu)(X - \mu)^\mathsf{T}]A^\mathsf{T} \\
&= A\Sigma A^\mathsf{T}
\end{aligned}$$

Bayes Filters: Framework

- Given:
  - Stream of observations $z_{1:t}$ and action data $u_{1:t}$
  - Sensor/measurement model $p(z_t|x_t)$
  - Action/motion/transition model $p(x_t|x_{t-1},u_t)$
- Wanted:
  - The state X_t of dynamical system
  - The posterior of state is called belief $bel(x_t) = p(x_t|z_{1:t}, u_{1:t})$

Bayes Filter

Algorithm `Bayes-filter`

Require: Belief $bel(x_{t-1}) = p(x_{t-1} | z_{1:t-1}, u_{1:t-1})$, action u_t, measurement z_t;

for all state variables **do**

    $\overline{bel}(x_t) = \int p(x_t | x_{t-1}, u_t) bel(x_{t-1}) \mathrm{d}x_{t-1}$ // Predict using action/control input u_t

    $bel(x_t) = \eta p(z_t | x_t) \overline{bel}(x_t)$ // Update using perceptual data z_t

return $bel(x_t)$

Algorithm `Bayes-filter`

Require: Belief $bel(x_{t-1}) = p(x_{t-1}|z_{1:t-1}, u_{1:t-1})$, action u_t, measurement z_t;

1: **for** all state variables **do**

2: $\overline{bel}(x_t) = \int p(x_t|x_{t-1}, u_t) bel(x_{t-1}) \mathrm{d}x_{t-1}$ // Predict using action/control input u_t

3: $bel(x_t) = \eta p(z_t|x_t)\overline{bel}(x_t)$ // Update using perceptual data z_t

4: **return** $bel(x_t)$

Ingredients:

- Bayes' rule
- Conditional independence
- Law of total probability

Bayes Filters: Implementation Examples

Linear:

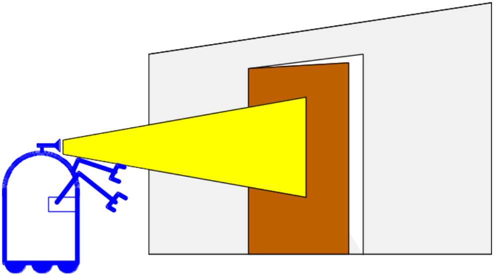
- Kalman Filter: unimodal linear filter
- Information Filter: unimodal linear filter

Nonlinear:

- Extended Kalman Filter: unimodal nonlinear filter with Gaussian noise assumption
- Extended Information Filter: unimodal nonlinear filter with Gaussian noise assumption
- Particle Filter: multimodal nonlinear filter

Simple Example of State Estimation

- Suppose a robot obtains measurement z, e.g., using its camera;
- What is $p(\text{open}|z)$?



Causal vs. Diagnostic Reasoning

- $p(\text{open}|z)$ is **diagnostic**.
- $p(z|\text{open})$ is **causal**.
- Often causal knowledge is easier to obtain.
- Bayes rule allows us to use causal knowledge:

$$p(\text{open}|z) = \frac{p(z|\text{open})p(\text{open})}{p(z)}$$

Example

Sensor model (likelihood):

- $p(z = \text{sense\_open}|\text{open}) = 0.6$
- $p(z = \text{sense\_open}|\neg\text{open}) = 0.3$

Prior knowledge (non-informative in this case):

- $p(\text{open}) = p(\neg\text{open}) = 0.5$

Update/Correction:

$$p(\text{open}|z) = \frac{p(z|\text{open})p(\text{open})}{p(z|\text{open})p(\text{open}) + p(z|\neg\text{open})p(\neg\text{open})}$$

$$p(\text{open}|z = \text{sense\_open}) = \frac{0.6 \times 0.5}{0.6 \times 0.5 + 0.3 \times 0.5} = 0.6667$$

Remark

z raises the probability that the door is open.

Combining Evidence

- Suppose our robot obtains another observation z_2.
- How can we integrate this new information?
- More generally, how can we estimate $p(x|z_1, \dots, z_n)$?

$$p(x|z_1,\dots,z_n) = \frac{p(z_n|x,z_1,\dots,z_{n-1})p(x|z_1,\dots,z_{n-1})}{p(z_n|z_1,\dots,z_{n-1})}$$

Assumption (Markov Assumption)

Z_n *is* **independent** *of* $Z_1,\dots,Z_{n-1}$ **if** *we* **know** $X = x$.

$$p(x|z_1, \ldots, z_n) = \frac{p(z_n|x, z_1, \ldots, z_{n-1})p(x|z_1, \ldots, z_{n-1})}{p(z_n|z_1, \ldots, z_{n-1})}$$

Assumption (Markov Assumption)

Z_n *is* **independent** *of* $Z_1, \ldots, Z_{n-1}$ **if** *we* **know** $X = x$.

or equivalently we can state:

Assumption (Markov Property)

*The Markov property states that "***the future is independent of the past if the present is known.***" A stochastic process that has this property is called a* **Markov process***.*

Recursive Bayesian Updating

$$p(x|z_1,\dots,z_n) = \frac{p(z_n|x, z_1,\dots,z_{n-1})p(x|z_1,\dots,z_{n-1})}{p(z_n|z_1,\dots,z_{n-1})}$$

Assumption (Markov Assumption)

Z_n *is* **independent** *of* $Z_1,\dots,Z_{n-1}$ **if** *we* **know** $X = x$.

$$p(x|z_1,\dots,z_n) = \frac{p(z_n|x)p(x|z_1,\dots,z_{n-1})}{p(z_n|z_1,\dots,z_{n-1})}$$

$$= \eta_n \; p(z_n|x)p(x|z_1,\dots,z_{n-1}) = \eta_{1:n} \; \prod_{i=1}^{n} p(z_i|x)p(x)$$

where $\eta_{1:n} := \eta_1\eta_2\cdots\eta_n$.

- Probabilistic Robotics: Ch. 1 and 2, Understand Example 2.4.2
- State Estimation for Robotics: Ch. 2
- Lecture Notes for Mobile Robotics: Ch. 1

Next Time

- Kalman Filtering
- Readings:
  - Probabilistic Robotics: Ch. 3
  - State Estimation for Robotics: Ch. 3
  - Lecture Notes for Mobile Robotics: Ch. 2 and 6