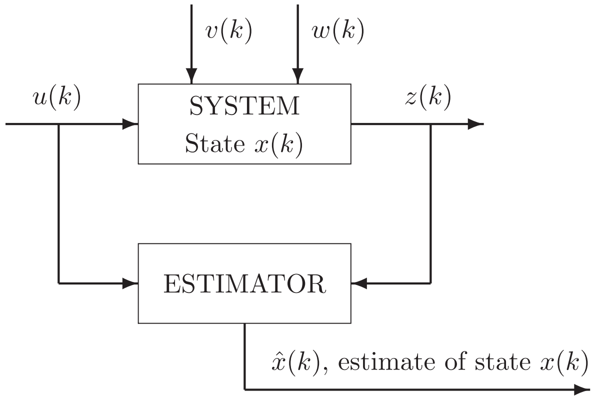
1. Introduction to State Estimation

What this course is about

Estimation of the state of a dynamic system based on a model and observations (sensor measurements), in a computationally efficient way.



- $u(k)$: known input
- $z(k)$: measured output
- $v(k)$: process noise
- $w(k)$: sensor noise
- $x(k)$: internal state

Systems being considered

Nonlinear, discrete-time system:

$$x(k) = q_k\big(x(k-1), u(k), v(k)\big) \qquad k = 1, 2, ...$$
$$z(k) = h_k\big(x(k), u(k), w(k)\big) \qquad x(0),\ \{v(\cdot)\},\ \{w(\cdot)\} \text{ have a } \textit{probabilistic} \text{ description.}$$

Focus on recursive algorithms

Estimator has its own state: $\hat{x}(k)$, the estimate of $x(k)$ at time k. Will compute $\hat{x}(k)$ from $\hat{x}(k-1)$, $u(k)$, $z(k)$, and model knowledge (dynamic model and probabilistic noise models). No need to keep track of the complete time history $\{u(\cdot)\}$, $\{z(\cdot)\}$.

Applications

- Generally, estimation is the dual of control. *State feedback control*: given state $x(k)$, what should input $u(k+1)$ be? So one class of estimation problems is *any* state feedback control problem where the state $x(k)$ is not available. This is a very large class of problems.
- Estimation without closing the loop: state estimates of interest in their own right (for example, system health monitoring, fault diagnosis, aircraft localization based on radar measurements, economic development, medical health monitoring).

Date compiled: February 26, 2013

Resulting algorithms

Will adopt a *probabilistic approach*.

- Underlying technique: Bayesian Filtering
- Linear system and Gaussian distributions: Kalman Filter
- Nonlinear system and (approximately) Gaussian distributions: Extended Kalman Filter
- Nonlinear system or Non-Gaussian (especially multi-modal) distributions: Particle Filter

The Kalman Filter is a special case where we have analytical solutions. Trade-off: tractability vs. accuracy.

2. Probability Review

Engineering approach to probability: rigorous, but not the most general. We will not go into measure theory, for example.

2.1 Probability: A Motivating Example

Only for intuition.

- A man has M pairs of pants and L shirts in his wardrobe. Over a long period of time, we observe the pants/shirt combination he chooses. In particular, out of N observations:

  $n_{ps}(i, j)$: number of times he wore pants i with shirt j
  $n_p(i)$: number of times he wore pants i
  $n_s(j)$: number of times he wore shirt j

- Define

  $f_{ps}(i, j) := n_{ps}(i, j)/N$, the *likelihood* of wearing pants i with shirt j
  $f_p(i) := n_p(i)/N$, the likelihood of wearing pants i
  $f_s(j) := n_s(j)/N$, the likelihood of wearing shirt j

  Note that $f_p(i) \geq 0$, $\displaystyle\sum_{i=1}^{M} f_p(i) = \sum_{i=1}^{M} \frac{n_p(i)}{N} = \frac{N}{N} = 1$. Similarly for $f_s(j)$.

- We notice a few things:

  $n_p(i) = \displaystyle\sum_{j=1}^{L} n_{ps}(i, j)$, all the ways in which he chose pants i

  $n_s(j) = \displaystyle\sum_{i=1}^{M} n_{ps}(i, j)$, all the ways in which he chose shirt j

  Therefore $f_p(i) = \displaystyle\sum_{j=1}^{L} f_{ps}(i, j)$ and $f_s(j) = \displaystyle\sum_{i=1}^{M} f_{ps}(i, j)$.
  Called the **marginalization**, or **sum rule**.

- Define

  $f_{p|s}(i,j) := n_{ps}(i,j)/n_s(j)$, the likelihood of wearing pants i given that he is wearing shirt j

  $f_{s|p}(j,i) := n_{ps}(i,j)/n_p(i)$, the likelihood of wearing shirt j given that he is wearing pants i

  Then

  $$\begin{aligned} f_{ps}(i,j) = \frac{n_{ps}(i,j)}{N} &= \frac{n_{ps}(i,j)}{n_p(i)}\frac{n_p(i)}{N} = f_{s|p}(j,i)\, f_p(i) \\ &= \frac{n_{ps}(i,j)}{n_s(j)}\frac{n_s(j)}{N} = f_{p|s}(i,j)\, f_s(j) \end{aligned}$$

  Called the **conditioning**, or **product rule**.

- Everything we do in this class stems from these two simple rules. Understand them well.

- "Frequentist" approach to probability: captured by this example. Intuitive. Relative frequency in a large number of trials. Great way to think about probability for physical processes such as tossing coins, rolling dice, and other phenomena where the physical process is essentially random.

- "Bayesian" approach. Probability is about beliefs and uncertainty. Measure of the state of knowledge.

2.2 Discrete Random Variables (DRV)

Formalize the motivating example.

- $\mathcal{X}$: the set of all possible outcomes, subset of the integers $\{..., -1, 0, 1, ...\}$.

- $f_x(\cdot)$: the **probability density function** (PDF), a real valued function that satisfies

  1. $f_x(\bar{x}) \geq 0 \; \forall \; \bar{x} \in \mathcal{X}$
  2. $\sum_{\bar{x}\in\mathcal{X}} f_x(\bar{x}) = 1$

- $f_x(\cdot)$ and $\mathcal{X}$ define a **discrete random variable** (DRV) x.

- The PDF can be used to define the notion of probability: the **probability** that a random variable x is equal to some value $\bar{x} \in \mathcal{X}$ is $f_x(\bar{x})$. This is written as $\Pr(x = \bar{x}) = f_x(\bar{x})$.

- In order to simplify notation, we often use x to denote a DRV *and* a specific value the DRV can take. So for example, we will write $\Pr(x) = f_x(x)$. Furthermore, if it is clear from context which DRV we are talking about, we simply write $f(x)$ instead of $f_x(x)$. *While this is convenient, it may confuse you at first. If so, we encourage you to use the more cumbersome notation until you are comfortable with the shorthand notation.*

Examples

- $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$, $f(x) = \frac{1}{6} \; \forall \; x \in \mathcal{X}$, captures a fair die.

- $\mathcal{X} = \{0, 1\}$, $f(x) = \begin{array}{ll} 1-h & \text{for } x = 0 \text{ ("tails")} \\ h & \text{for } x = 1 \text{ ("heads")} \end{array}$

  where $0 \leq h \leq 1$, captures the flipping of a coin, h captures the coin bias.

Multiple DRVs

- What if we have multiple random variables? **Joint PDF.**
  - Let x and y be two DRVs. Joint PDF satisfies:
    1. $f_{xy}(x, y) \geq 0 \;\; \forall\, x \in \mathcal{X},\; \forall\, y \in \mathcal{Y}$
    2. $\displaystyle\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} f_{xy}(x, y) = 1$
  - Example: $\mathcal{X} = \mathcal{Y} = \{1, 2, 3, 4, 5, 6\}$, $f_{xy}(x, y) = \frac{1}{36}$, captures the outcome of two fair dice.
  - Generalizes to an arbitrary number of random variables.
  - Short form: $f(x, y)$.

- **Marginalization, or Sum Rule** axiom:

$$\text{Given } f_{xy}(x, y), \text{ define } f_x(x) := \sum_{y\in\mathcal{Y}} f_{xy}(x, y).$$

  - This is a *definition*: $f_x(x)$ is fully defined by $f_{xy}(x, y)$. (Recall pants & shirts example.)

- **Conditioning, or Product Rule** axiom:

$$\text{Given } f_{xy}(x, y), \text{ define } f_{x|y}(x, y) := \frac{f_{xy}(x, y)}{f_y(y)} \;\; \text{(when } f_y(y) \neq 0\text{)}.$$

  - This is a *definition.* $f_{x|y}(x, y)$ can be thought of as a function of x, with y fixed. It is easy to verify that it is a valid PDF in x. "Given y, what is the probability of x?" (Recall pants & shirts example.)
  - Alternate, more expressive notation: $f_{x|y}(x|y)$.
  - Short form: $f(x|y)$
  - Usually written as $f(x, y) = f(x|y)\, f(y) = f(y|x)\, f(x)$.

- We can combine these to give us our first theorem, the **Total Probability Theorem:**

$$f_x(x) = \sum_{y\in\mathcal{Y}} f_{x|y}(x|y)\, f_y(y). \qquad \text{A weighted sum of probabilities.}$$

Multi-variable generalizations

Sometimes x is used to denote a collection (or vector) of random variables $x = (x^1, x^2, ..., x^N)$. So when we write $f(x)$ we implicitly mean $f\big(x^1, x^2, ..., x^N\big)$.

Marginalization: $f(x) = \displaystyle\sum_{y\in\mathcal{Y}} f(x, y)$ short form for

$$f\big(x^1, x^2, ..., x^N\big) = \sum_{(y^1,...,y^L)\in\mathcal{Y}} f\big(x^1, ..., x^N, y^1, ..., y^L\big).$$

Still a scalar!

Conditioning: Similarly, $f(x, y) = f(x|y)\, f(y)$ applies to collections of random variables.

2.3 Continuous Random Variables (CRV)

Very similar to discrete random variables.

- $\mathcal{X}$ is a subset of the real line, $\mathcal{X} \subseteq \mathbb{R}$ (for example, $\mathcal{X} = [0,1]$ or $\mathcal{X} = \mathbb{R}$).
- $f_x(\cdot)$, the PDF, is a real valued function satisfying:
  1. $f_x(x) \geq 0 \; \forall \, x \in \mathcal{X}$
  2. $\displaystyle\int\limits_{\mathcal{X}} f_x(x)\, dx = 1$
  3. $f_x(x)$ is bounded and piecewise continuous
     - Stronger than necessary, but will keep you out of trouble. No delta functions, things that go to infinity, etc. Adequate for most problems you will encounter.
- Relation to probability: doesn't make sense to say that the probability of x is $f_x(x)$. Look at the following limiting process: consider the integers $\{1, 2, \ldots, N\}$ divided by N; that is, the numbers $\{1/N, 2/N, \ldots, N/N\}$, which are in the interval [0,1]. Assume that all are of equal probability $1/N$. As N goes to infinity, the probability of any specific value i/N goes to 0.

  So instead we talk about probability of being in an interval:

  $$\Pr(x \in [a,b]) := \int\limits_a^b f_x(x)\, dx$$

- All other definitions, properties, etc. derived to date for DRVs apply to CRVs, just replace "$\sum$" by "$\int$."
- Can mix discrete and continuous random variables. Example:

  $$x \in \{0,1\},\; y \in [0,1],\quad f(x,y) = \begin{cases} 1-y & \text{for } x = 0 \\ y & \text{for } x = 1 \end{cases}$$

  - x: flip of a coin, a DRV.
  - y: coin bias, a CRV.
  - $f(y) = \displaystyle\sum_x f(x,y) = 1$, uniformly distributed.
  - $f(x) = \displaystyle\int\limits_0^1 f(x,y)\, dy = \frac{1}{2}$ for $x = 0$ and $x = 1$.
  - $f(x|y) = \dfrac{f(x,y)}{f(y)} = f(x,y)$.
  - $f(y|x) = \dfrac{f(x,y)}{f(x)} = 2f(x,y)$.