Basic Probability and Distributions
Sampling, Tracking
Tracking via the Particle Filter
CMSC 828D
Fall 2000
Probability notation and definitions

- $D$ set of all events, Null event $\emptyset$
- Probability of an event $A$ occurring $P(A)$
  - $P(D) = 1$
  - $P(\emptyset) = 0$
  - for any $A$, $0 \leq P(A) \leq 1$
  - if $A \subset B$, then $P(A) \leq P(B)$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probability of either of two events occurring
• Probability of both events occurring $P(A,B)$
  \[ P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A) \]

• Leads directly to Bayes Rule
  \[ P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \]
  • Way to transfer conditional probabilities

• Bayesian Inference

• Independence of two events $A$ and $B$
  \[ P(A \mid B) = P(A)P(B) \]

• Conditional independence
  \[ P(A,B \mid C) = P(A \mid C)P(B \mid C) \]
Probability Distributions

- Instead of single events we look at now a large collection of events.
- Assume that these events can be characterized by a number
- “take to the limit” and look at values of probability for values of $x$
  along the real line
- probabilities associated with $x$ taking on a range of values.
  $[a,b]$  $(a,b]$  $(-\infty, \infty)$ etc.

- Convenient to look at two distribution functions

  probability density function  
  $P(a < x < b) = \int_a^b p(x)dx$

  cumulative density function  
  $F(a) = \int_{-\infty}^a p(x)dx = P(-\infty < x \leq a)$

- For continuous density functions $P(x=a) = 0$

- Example density function: gaussian  
  $N(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$
Working with distributions

- $E(x)$ is the expected value of a random variable
  $$E[x] = \sum_{i \in \text{values}} x_i p(x_i)$$
  $$E[x] = \int_D xp(x)dx$$
  $$E(g(x)) = \int_{-\infty}^{\infty} g(x)p(x)dx$$

- $E(x)$ is nothing but the mean or average of $x$

- Variance $\text{var}(x) = E[x^2 - (E(x))^2]$  

- Variance is the difference between the expected value of the square and $E(x)^2$.  
  $$\text{var}(x) = \int_{-\infty}^{\infty} x^2 p(x)dx - \left[\int_{-\infty}^{\infty} xp(x)dx\right]^2$$

- Estimates departures from the mean

- Knowing the distribution and how to integrate functions of $x$ with respect to it we can compute probabilities

- Sampling techniques -- attempt to compute probabilities by approximating the integral.

- Use known values at a few sample points.
Computing expectations with samples

- Distribution is a device to compute expectations
- Given a distribution of points $u^i$ and a distribution on these points $f(u^i)$

Represent a probability distribution

$$p_f(X) = \frac{f(X)}{\int f(U)dU}$$

by a set of $N$ weighted samples

$$\{(u^i, w^i)\}$$

where $u^i \sim s(u)$ and $w^i = f(u^i)/s(u^i)$.

- Compute expectations using the sample points and weights

$$\int g(U)p_f(U)dU \approx \frac{\sum_{i=1}^{N} g(u^i)w^i}{\sum_{i=1}^{N} w^i}$$
Sampling

- Basic problem for Monte-Carlo Methods
  - Integrate a function $f$ over a region of volume $V$
  - Integral may be hard to calculate because
    - the function is not known explicitly,
    - region over which the integral is to be taken cannot be characterized
    - Integral is over many dimensions (e.g. 100s)
  - Approximate integral somehow

- Von Neumann while working on the Manhattan project, approximated integral as

$$\int f \, dV \approx V \langle f \rangle \pm V \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}}$$

$$\langle f \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} f(x_i) \quad \langle f^2 \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} f^2(x_i)$$

Figure 7.6.1. Monte Carlo integration. Random points are chosen within the area $A$. The integral of the function $f$ is estimated as the area of $A$ multiplied by the fraction of random points that fall below the curve $f$. Refinements on this procedure can improve the accuracy of the method; see text.
Bayesian Inference

- Convert the simple Bayes formula into a powerful way to look at any new piece of information.
- Probabilistic model with some parameters
- Fixing parameters allows predicting the probabilities of events. Can calculate $P(\text{measurements}|\text{parameters})$
- **Prior:** We have an estimate of $P(\text{parameters})$
- **Posterior:** Given measurements, we want to update our estimate of the parameters. $P(\text{parameters}|\text{measurements})$
- Bayesian inference formula is

$$P(\text{parameters}|\text{measurements}) = \frac{P(\text{measurements}|\text{parameters})P(\text{parameters})}{P(\text{measurements})}$$
Tracking

- Components:
  - a motion model that predicts the new state of the system.
    - Allows one to predict $y_i$
  - Measurement
    - Measure things that can also be predicted by your model
    - E.g. position of a point, or some other quantity
    - Measurement satisfies equation
  - Use Bayesian framework
  - Estimate posterior distribution of $y_i$

- When equations were linear and noise models were Gaussian, the Kalman filter applies
- When equations are nonlinear and noise is Gaussian we can use the Extended Kalman filter
- Another approach is to use sampling
Tracking as inference

- Given an estimate of parameters at the new step as $y$
- Use measurement and Bayes rule to improve the estimate
  \[ P(y \mid x) = \frac{P(x \mid y)P(y)}{P(x)} = \frac{P(x \mid y)P(y)}{\int P(x \mid y)dy} \]

- Denominator only depends upon the data, and not our estimates of $y$
- Thus it is constant w. r. to $y$ and we can write
  \[ P(y \mid x) \propto P(x \mid y)P(y) \]

- Often evaluating the denominator is hard and this proportionality equation is used
Representing the posterior using samples

- Bayes rule (again)

\[
p(U | V = v_0) = \frac{p(V = v_0 | U)p(U)}{\int p(V = v_0 | U)p(U) \, dU} = \frac{1}{K} p(V = v_0 | U)p(U)
\]

- Evaluating \( K \)

\[
K = \int p(V = v_0 | U)p(U) \, dU
\]

\[
= E \left[ \frac{\sum_{i=1}^{N} p(V = v_0 | u^i)w^i}{\sum_{i=1}^{N} w^i} \right] \approx \frac{\sum_{i=1}^{N} p(V = v_0 | u^i)w^i}{\sum_{i=1}^{N} w^i}
\]

\[
\int g(U)p(U | V = v_0) \, dU = \frac{1}{K} \int g(U)p(V = v_0 | U)p(U) \, dU
\]

\[
\approx \frac{1}{K} \frac{\sum_{i=1}^{N} g(u^i)p(V = v_0 | u^i)w^i}{\sum_{i=1}^{N} w^i}
\]

\[
\approx \frac{\sum_{i=1}^{N} g(u^i)p(V = v_0 | u^i)w^i}{\sum_{i=1}^{N} p(V = v_0 | u^i)w^i}
\]

- Evaluate the posterior

- Equiv. to computing \( E \) with weight \( w'^i = p(V = v_0 | u^i)w^i \)
Resampling

- Original points may not sample the posterior well
- Resample … distribute points according to the pdf of the posterior and compute new points $u_j$ and weights $w_j$. 
Algorithm

- Initialize
- Predict using the motion model
- Measure
- Use measurements to obtain new weights
- Resample to generate new points and new weights
- Loop
Algorithm

Initialization: Represent $P(X_0)$ by a set of $N$ samples

$$\left\{ (s_0^{k,-}, w_0^{k,-}) \right\}$$

where

$$s_0^{k,-} \sim P_s(S)$$

and

$$w_0^{k,-} = \frac{P(s_0^{k,-})}{P_s(S = s_0^{k,-})}$$

Ideally, $P(X_0)$ has a simple form and $s_0^{k,-} \sim P(X_0)$ and $w_0^{k,-} = 1$.

Prediction: Represent $P(X_i|y_0, y_{i-1})$ by

$$\left\{ (s_i^{k,-}, w_i^{k,-}) \right\}$$

where

$$s_i^{k,-} = f(s_{i-1}^{k,+}) + \xi_i$$
Correction: Represent \( P(X_i|y_0, y_i) \) by

\[
\{(s_i^{k,+}, w_i^{k,+})\}
\]

where

\[
s_i^{k,+} = s_i^{k,-} \\
w_i^{k,+} = P(Y_i = y_i|X_i = s_i^{k,-})w_i^{k,-}
\]

Resampling: Normalise the weights so that \( \sum_i w_i^{k,+} = 1 \) and compute the variance of the normalised weights. If this variance exceeds some threshold, then construct a new set of samples by drawing, with replacement, \( N \) samples from the old set, using the weights as the probability that a sample will be drawn. The weight of each sample is now \( 1/N \).

Algorithm 19.8: A practical particle filter resamples the posterior.
The Condensation algorithm

\[
p(X_{t-1} | Z_{t-1}) \\
p(X_t | Z_{t-1}) \\
p(Z_t | X_t) \\
p(X_t | Z_t)
\]

\[
s_{t-1}, \pi_{t-1} \\
\text{predict} \\
\text{measure} \\
s_t, \pi_t
\]
Improving the algorithm

• Make the distribution of sample points “better”

• Recall error estimate of MC method

\[ \int f \, dV \approx V \langle f \rangle \pm V \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}} \]

\[ \langle f \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} f(x_i) \quad \langle f^2 \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} f^2(x_i) \]

• Error can be reduced by
  • Increasing \( N \)
  • Reducing variance of \( f \) computed on the sampled points
  • Using deterministic sets of points called quasi-random points to do the sampling.
Conventional tracking algorithms

- Assume image motion model (e.g., affine)
- Compute flow for patches
- Obtain parameters of the transformation for patches
- Track …
- Not very robust … but could be important for applications.

- J. Shi and C. Tomasi. *Good Features to Track*. IEEE Conference on Computer Vision and Pattern Recognition, June 1994, pp. 593-600