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 **Ø**
- Probability of an event **A** occurring \( P(A) \)
  - \( P(D) = 1 \)
  - \( P(Ø) = 0 \)
  - for any **A**, \( 0 \leq P(A) \leq 1 \)
  - if \( A \subseteq B \), 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

<table>
<thead>
<tr>
<th>Probability notation and definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>- <strong>D</strong> set of all events, Null event <strong>Ø</strong></td>
</tr>
<tr>
<td>- Probability of an event <strong>A</strong> occurring ( P(A) )</td>
</tr>
<tr>
<td>- ( P(D) = 1 )</td>
</tr>
<tr>
<td>- ( P(Ø) = 0 )</td>
</tr>
<tr>
<td>- for any <strong>A</strong>, ( 0 \leq P(A) \leq 1 )</td>
</tr>
<tr>
<td>- if ( A \subseteq B ), then ( P(A) \leq P(B) )</td>
</tr>
<tr>
<td>- ( P(A \cup B) = P(A) + P(B) - P(A \cap B) )</td>
</tr>
<tr>
<td>- Probability of either of two events occurring</td>
</tr>
</tbody>
</table>

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 \( \mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \)

<table>
<thead>
<tr>
<th>Probability Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Instead of single events we look at now a large collection of events.</td>
</tr>
<tr>
<td>- Assume that these events can be characterized by a number</td>
</tr>
<tr>
<td>- &quot;take to the limit&quot; and look at values of probability for values of ( x ) along the real line</td>
</tr>
<tr>
<td>- probabilities associated with ( x ) taking on a range of values. ([a,b] ) ((a,b] ) ((-\infty, \infty) ) etc.</td>
</tr>
<tr>
<td>- Convenient to look at two distribution functions</td>
</tr>
<tr>
<td>- probability density function ( p(a &lt; x &lt; b) = \int_{a}^{b} p(x)dx )</td>
</tr>
<tr>
<td>- cumulative density function ( F(a) = \int_{-\infty}^{a} p(x)dx = P(-\infty &lt; x \leq a) )</td>
</tr>
<tr>
<td>- For continuous density functions ( P(x=a) = 0 )</td>
</tr>
<tr>
<td>- Example density function: gaussian ( \mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} )</td>
</tr>
</tbody>
</table>

Working with distributions
- \( E(x) \) is the expected value of a random variable \( \sum_{x \text{values}} x p(x) \) \( E(x) = \int_{D} x p(x)dx \) \( E(g(x)) = \int_{D} g(x)p(x)dx \)
- \( E(x) \) 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 \( E(x)^2 \)
- Estimates departures from the mean \( \frac{\int_{D} x^2 p(x)dx}{\int_{D} x p(x)dx} - \left( \frac{\int_{D} x p(x)dx}{\int_{D} p(x)dx} \right)^2 \)
- 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.

<table>
<thead>
<tr>
<th>Working with distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>- ( E(x) ) is the expected value of a random variable ( \sum_{x \text{values}} x p(x) ) ( E(x) = \int_{D} x p(x)dx ) ( E(g(x)) = \int_{D} g(x)p(x)dx )</td>
</tr>
<tr>
<td>- ( E(x) ) is nothing but the mean or average of ( x )</td>
</tr>
<tr>
<td>- Variance ( \text{var}(x) = E[x^2] - (E(x))^2 )</td>
</tr>
<tr>
<td>- Variance is the difference between the expected value of the square and ( E(x)^2 )</td>
</tr>
<tr>
<td>- Estimates departures from the mean ( \frac{\int_{D} x^2 p(x)dx}{\int_{D} x p(x)dx} - \left( \frac{\int_{D} x p(x)dx}{\int_{D} p(x)dx} \right)^2 )</td>
</tr>
<tr>
<td>- Knowing the distribution and how to integrate functions of ( x ) with respect to it we can compute probabilities</td>
</tr>
<tr>
<td>- Sampling techniques -- attempt to compute probabilities by approximating the integral.</td>
</tr>
<tr>
<td>- Use known values at a few sample points.</td>
</tr>
</tbody>
</table>

Computing expectations with samples
- Distribution is a device to compute expectations
- Given a distribution of \( u \) and a distribution on these points \( f(u) \)
- 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
  \( \int g(U)p_f(U)dU \approx \sum_{i=1}^{N} g(U^i)p_f(U^i) \sum_{i=1}^{N} w^i \)

<table>
<thead>
<tr>
<th>Computing expectations with samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Distribution is a device to compute expectations</td>
</tr>
<tr>
<td>- Given a distribution of ( u ) and a distribution on these points ( f(u) )</td>
</tr>
<tr>
<td>- Represent a probability distribution ( p_f(X) = \frac{f(X)}{\int f(U)dU} )</td>
</tr>
<tr>
<td>- by a set of ( N ) weighted samples ( { (u^i, w^i) } )</td>
</tr>
<tr>
<td>- where ( u^i \sim s(u) ) and ( w^i = f(u^i)/s(u^i) ).</td>
</tr>
<tr>
<td>- Compute expectations using the sample points and weights</td>
</tr>
<tr>
<td>( \int g(U)p_f(U)dU \approx \sum_{i=1}^{N} g(U^i)p_f(U^i) \sum_{i=1}^{N} w^i )</td>
</tr>
</tbody>
</table>
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
  - Approximate integral somehow
- Von Neumann while working on the Manhattan project, approximated integral as

\[
\int f \, dv = V \left( \frac{\langle f \rangle}{N} \pm V \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}} \right)
\]

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 \( P(\text{measurements|parameters}) \)
- Prior: We have an estimate of \( P(\text{parameters}) \)
- Posterior: Given measurements, we want to update our estimate of the parameters. \( P(\text{parameters|measurements}) \)
- Bayesian inference formula is

\[
P(\text{parameters|measurements}) = \frac{P(\text{measurements|parameters})P(\text{parameters})}{P(\text{measurements})}
\]

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

Representing the posterior using samples

- Bayes rule (again)

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

- Evaluating \( K \)

\[
K = \int p(U|V = v_0) \, dU = \mathbb{E} \left[ \sum_{i=1}^{N} p(V = v_i | U) u_i \right] = \sum_{i=1}^{N} \frac{p(V = v_i | U) u_i}{\sum_{i=1}^{N} u_i}
\]

- Evaluate the posterior

\[
p(U|V = v_0) = \frac{1}{K} p(U|V = v_0) p(U) \, dU = \frac{1}{K} \mathbb{E} \left[ \sum_{i=1}^{N} p(U|V = v_i | U) u_i \right] = \frac{1}{K} \sum_{i=1}^{N} \frac{p(U|V = v_i | U) u_i}{\sum_{i=1}^{N} u_i}
\]

- Equiv. to computing \( E \) with weight \( u_i^2 = p(U = v_0 | u_i) u_i \)

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 - 2

Correction: Represent $P(X_t|y_{1:t})$ by

$$\{ (s^{i-}_t, w^{i-}_t) \}$$

where

$$s^{i+}_t = s^{i-}_t$$
$$w^{i+}_t = P(Y_t = y_t|X_t = s^{i-}_t) w^{i-}_t$$

Resampling: Normalize the weights so that $\sum_i w^{i+}_t = 1$ and compute the variance of the normalized 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 - 3: A practical particle filter resamples the posterior.

Improving the algorithm

- Make the distribution of sample points “better”
- Recall error estimate of MC method
  $$\hat{f} = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$
  $$\hat{f} = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$
  $$\left( \hat{f} \right)^2 = \frac{1}{N} \sum_{i=1}^{N} \left( f(x_i) - \hat{f} \right)^2$$
- 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