6.081 - Lecture 11
Fall 06

Probability

Imagine that you Know the robot is somewhere in the Stata Center, and you have a map. Now, you read the sonars and get a reading of .5 on the left. What do you know about the robot's location?

Probability

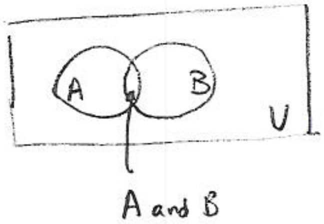
Sample space U of atomic events
Event: subset of U
} make a model

Axioms

$P(U) = 1$

$P(\{\}) = 0$

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$



Interpretations

- Frequentist: prob. is a long-run frequency
- Bayesian: prob is a subjective measure of belief

Where do we get a sample space?

Usually, a product of random variables.

A <u>discrete</u> random variable is a discrete set, and a mapping from elements of that set to probabilities.

Die: {1: .2, 2: .4, 3: .1, 4: .1, 5: .1, 6: .1}

Probabilities of all values of a R.V. sum to 1.

Sample space of a product of R.V.'s is the cartesian product of their sample spaces.

Cavity : {T, F}   Tooth Ache : {T, F}

Cavity × ToothAche : { (T, T), (T, F), (F, T), (F, F) }

Joint distribution of random variables is a function from elements of the product space to probability values. Sums to 1.

		Cavity		
		T	F	
Ache	T	.05	.05	.1
	F	.1	.8	.9
		.15	.85	

marginal dist'n on Ache

marginal dist'n on Cavity

Marginal distribution: $P(A=T) = P(A=T \text{ and } C=T) + P(A=T \text{ and } C=F)$
$P(A=F) = P(A=F \text{ and } C=T) + P(A=F \text{ and } C=F)$

$\boxed{\text{Can't compute joint from marginals}}$

... unless variables are independent

If two R.V.'s are independent, then

$$P(A=a \text{ and } B=b) = P(A=a) \cdot P(B=b)$$

What would that mean for the toothache/cavity distribution?

Evidence

What if a person walks into our dental office with a toothache? Now how likely do we think it is that they have a cavity?

Conditional probability: $P(C=T \mid A=T)$

Restrict the universe to $A=T$, and ask what the probability of $C=T$ is.

$$P(C=T \mid A=T) = \frac{P(C=T \text{ and } A=T)}{P(A=T)} = \frac{0.05}{0.1} = 0.5$$

Often, for medical diagnosis or characterizing the quality of a sensor, we experimentally measure $P(\text{symptom} = \text{true} \mid \text{disease} = \text{true})$

Bayes' Rule

But what we really want to know is $P(\text{Disease} = \text{true} \mid \text{Symptom} = \text{tr.}$

So

$$P(\text{Disease} = \text{true} \mid \text{Symptom} = \text{true}) = P(S = t \mid D = t)\, P(D = t) / P(S = t)$$

HMM-1

Simple state estimation problem:

Is the copy machine broken?

States: good, bad

Observations: smudged, all black, perfect

discrete time
hidden Markov model

Random variables

$S_0, S_1, S_2, \ldots$

$O_1, O_2, \ldots$

Problem: compute $P(S_t | O_1, \ldots, O_t)$

Assumptions: S_{t+1} depends <u>only</u> on S_t } and in the same way for all t

O_t depends <u>only</u> on S_t

prior

Given: $P(S_0 = s)$ for all s

good	bad
.9	.1

transition model

$P(S_{t+1} = s_i | S_t = s_j)$ for all s_i, s_j

		S_{t+1}	
		good	bad
S_t	good	.7	.3
	bad	.1	.9

Conditional, not joint models; <u>rows</u> sum to 1

sensor model

$P(O_t = o | S_t = s)$ for all o, s

		O_t		
		perfect	smudged	black
S_t	good	.8	.1	.1
	bad	.1	.7	.2

HMM-2

What if we print a page, and it's perfect? O_1 = perfect

We'd like to compute $P(S_1 = \text{good} \mid O_1 = \text{perfect})$.

Two steps: first, let's think about $P(S_1 = \text{good})$

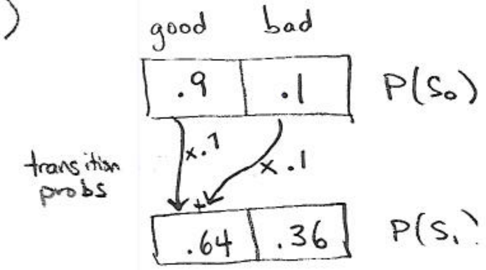
$$P(S_1=\text{good}) = P(S_1=\text{good}, S_0=\text{good}) + P(S_1=\text{good}, S_0=\text{bad})$$

$$= P(S_1=\text{good} \overset{\text{sensor model}}{\mid} S_0=\text{good})\,P(S_0=\overset{\text{prior}}{\text{good}}) +$$

$$P(S_1=\text{good} \mid S_0=\text{bad})\,P(S_0=\text{bad})$$

$$= .7 \cdot .9 + .1 \cdot .1$$

$$= .64$$



Now, incorporate the observation, using Bayes' rule

$$P(S_1=\text{good} \mid O_1=\text{perfect}) = \frac{P(O_1=\text{perfect} \overset{\text{sensor model}}{\mid} S_1=\text{good})\,P(S_1=\text{good})}{P(O_1=\text{perfect})}$$

What's $P(O_1=\text{perfect})$?

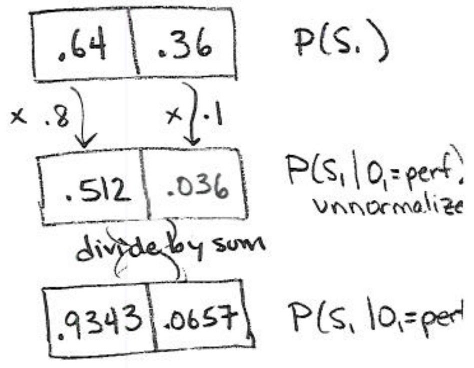
$$= P(O_1=\text{perfect}, S_1=\text{good}) + P(O_1=\text{perfect}, S_1=\text{bad})$$

$$= P(O_1=\text{perfect} \mid S_1=\text{good})\,P(S_1=\text{good}) +$$

$$P(O_1=\text{perfect} \mid S_1=\text{bad})\,P(S_1=\text{bad})$$

$$= .8 \cdot .64 + .1 \cdot .36$$

$$= .548$$



Now, back to

$$P(S_1=\text{good} \mid O_1=\text{perfect}) = \frac{.8 \cdot .64}{.548}$$

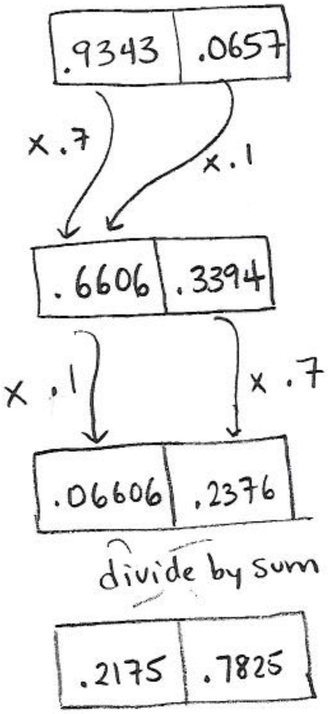
$$= \boxed{.9343}$$

HMM-3

Now, we print another page, and it's smudged.

Transition update

Observation update



$P(S_1 \mid O_1 = \text{perf})$

$P(S_2 \mid O_1 = \text{perf})$

$P(S_2 \mid O_1 = \text{perf}, O_2 = \text{smudged})$
unnormalized

$P(S_2 \mid O_1 = \text{perf}, O_2 = \text{smudged})$

More generically:

Let b_t be the current <u>belief state</u> $P(S_t)$

Step 1: $$b'_{t+1}(s_i) = \sum_{s_j} P(S_{t+1} = s_i \mid S_t = s_j)\, b_t(s_j)$$

Step 2: given obs o

$$b_{t+1}(s_i) = \frac{P(O_{t+1} = o \mid S_{t+1} = s_i)\, b'_{t+1}(s_i)}{\sum_{s_j} P(O_{t+1} = o \mid S_{t+1} = s_j)\, b'_{t+1}(s_j)}$$

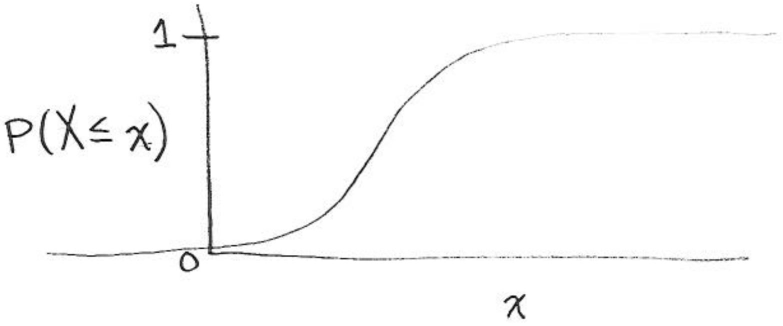
Continuous random variables

$X : \mathbb{R}$ takes on real values

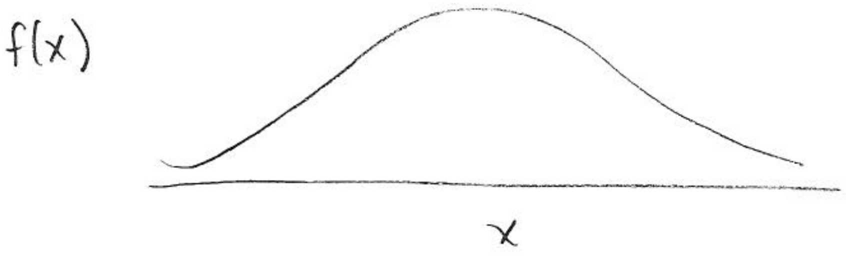
Events:

$P(X \leq c)$

Cumulative distribution function (CDF)



If $F_X(x)$ is the CDF, then $f_X(x) = \frac{d}{dx} F_X(x)$ is the probability density function (PDF).



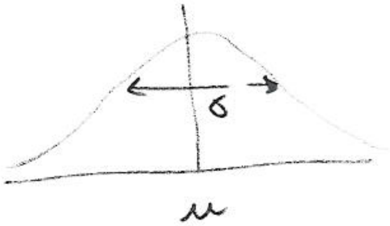
- $f(x) \geq 0$ for all x; can have values > 1
- $\int_{X} f(x)\, dx = 1$

Gaussian

PDF

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

No analytic form for CDF



Symmetric about μ

mean = median = mode = μ

variance = σ^2

63% of the area under the curve is within ± σ of μ

- Mathematically easy to work with
- Sum of many independent R.V.'s, no matter how distributed

  $\rightsquigarrow$ Gaussian