Random Variables

Random variables
- Defined by functions mapping outcomes to values
- By choice, whatever we are interested in
- Typically denoted by uppercase letters (e.g., X)
- Generic values are corresponding lower case letters
- Shorthand: P(x) = P(X=x)
- Value "type" is arbitrary (typically categorical or real)

Example (from K&F)
- Outcomes are student grades (A,B,C)
- Random variable G=f_{GRADE}(student)

$$P(A) \equiv P(G = A) \equiv P(\{w \in \Omega : f_{GRADE}(w) = A\})$$

Joint Distributions

Joint distribution of random variables

$$P(X,Y) \equiv P(X = x, Y = y) \equiv P(\{w \in \Omega : X(w) = x \text{ and } Y(w) = y\})$$

Conditional definition, Bayes rule, chain rule all apply.

Marginal distributions ("sum rule")

$$P(X) = \sum_Y P(X,Y)$$ *

Chain (product) rule (two variable case of chain rule)

$$P(X,Y) = P(X|Y)P(Y)$$ *

Basic terminology and rules

Conditional probability

$$P(X|Y) = \frac{P(X,Y)}{\sum_X P(X,Y)}$$ *

Bayes

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} = \frac{P(Y|X)P(X)}{\sum_X P(X,Y)}$$ *

*

$$P(X|Y) \propto P(Y|X)P(X)$$ (when Y is constant, i.e., evidence)

Normalization

Often we will deal with quantities or functions which are proportional to probabilities (OK if we just want argmax)

$$p(x) \propto P(X = x)$$

To get probabilities we normalize: $P(X = x) = \frac{p(x)}{\sum_x p(x)}$

Example: $P(X|Y) \propto P(X,Y)$

$$P(X|Y) = \frac{P(X,Y)}{\sum_X P(X,Y)}$$

Independence

$X \perp Y \quad \Leftrightarrow \quad P(X|Y) = P(X)$ or P(Y)=0

$X \perp Y \quad \Leftrightarrow \quad P(X,Y) = P(X)P(Y)$

Conditional Independence

$X \perp Y \mid Z \quad \Leftrightarrow \quad P(X,Y|Z) = P(X \mid Z)P(Y \mid Z)$

Probabilistic Queries

Bold face because these are vectors of variables

Organize variables into
- Evidence (observed), **E**
- Query (what you want to know), **Y**
- Hidden (leftover), **X** (for completeness)

Generic Query: P(**Y**|**E**)

MAP Query: $MAP(\mathbf{W} \mid \mathbf{e}) = \underset{w}{\text{Argmax}}\ P(w,e)$

Example

X \ Y	y_1	y_2
x_1	0.04	0.36
x_2	0.30	0.30

		Y		
		y_1	y_2	
X	x_1	0.04	0.36	**0.4**
	x_2	0.30	0.30	**0.6**
		0.34	0.66	

$P(x_1)=P(x_1,y_1)+P(x_1,y_1)$
[i.e., sum across]

		Y	
		y_1	
X	x_1	0.04	0.04 / 0.34
	x_2	0.30	0.30 / 0.34
		0.34	

$P(x|y_1)$

		Y		
		y_1	y_2	
X	x_1	0.04	0.36	**0.4**
	x_2	0.30	0.30	**0.6**
		0.34	0.66	

Arg max P(X,Y) is (x_1, y_2)

Arg max P(X) is (x_2)

Arg max P(Y) is (y_2)

Arg max P(X,Y) is **not** (Arg max P(X), Arg max P(Y))

Discrete Distributions (Bernoulli)

$x \in \{0,1\}$ (e.g., 1 is "heads" and 0 is "tails")

$\mathrm{p}(x=1|\mu)=\mu$

$Bern(x \mid \mu)=\mu^{x}(1-\mu)^{(1-x)}$

Code for sampling a Bernoulli

```
a=rand()

if (a<u) return heads
else return tails
```

Discrete Distributions (Binomial)

How likely it is that we get m "heads" in N tosses?

$$Bin(m|N,\mu) = \binom{N}{m} \mu^m (1-\mu)^{N-m}$$

where $\binom{N}{m} \equiv \frac{N!}{(N-m)!m!}$