Probability

Probability

- Random variables
- Atomic events
- Sample space

RVs: variables whose values are (potentially) uncertain
tomorrow’s weather (rain/sun), change in AAPL stock price (up/same/dn), grade on HW1 (0..100)
discrete for now

atomic event: setting for *all* rvs of interest
w=rainy & AAPL=down & HW1=93

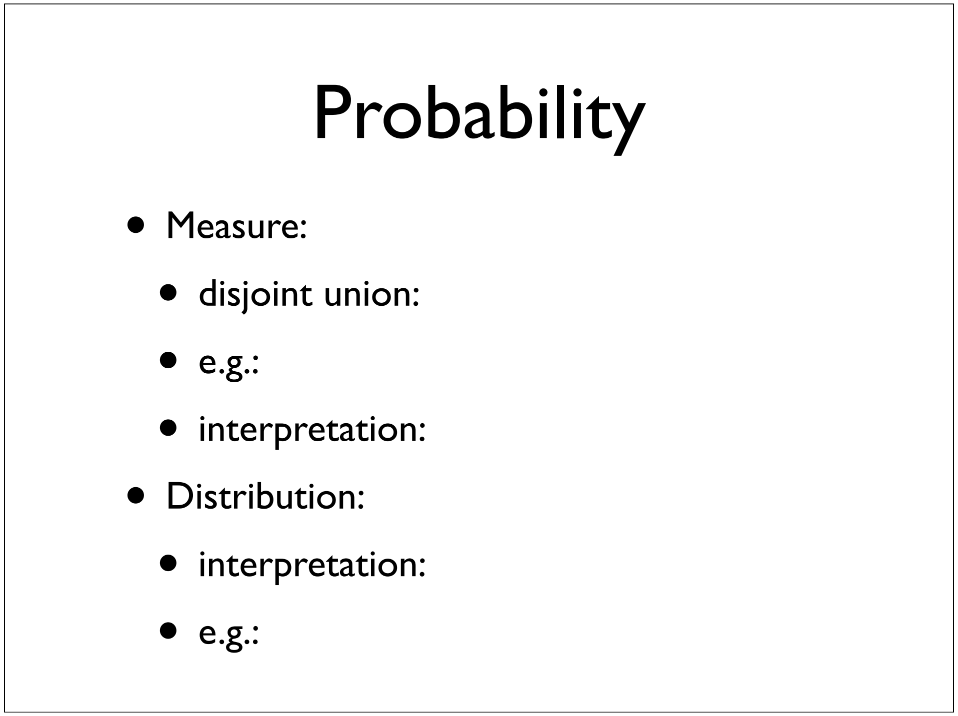
sample space: Omega = set of all atomic events

Probability

- Events
- Combining events

weather = rainy, grade = 93/100
grade >= 90
    set of atomic events

combining: and, or, not = iters., union, set diff
    w = rainy, AAPL != dn
    (note , means AND)

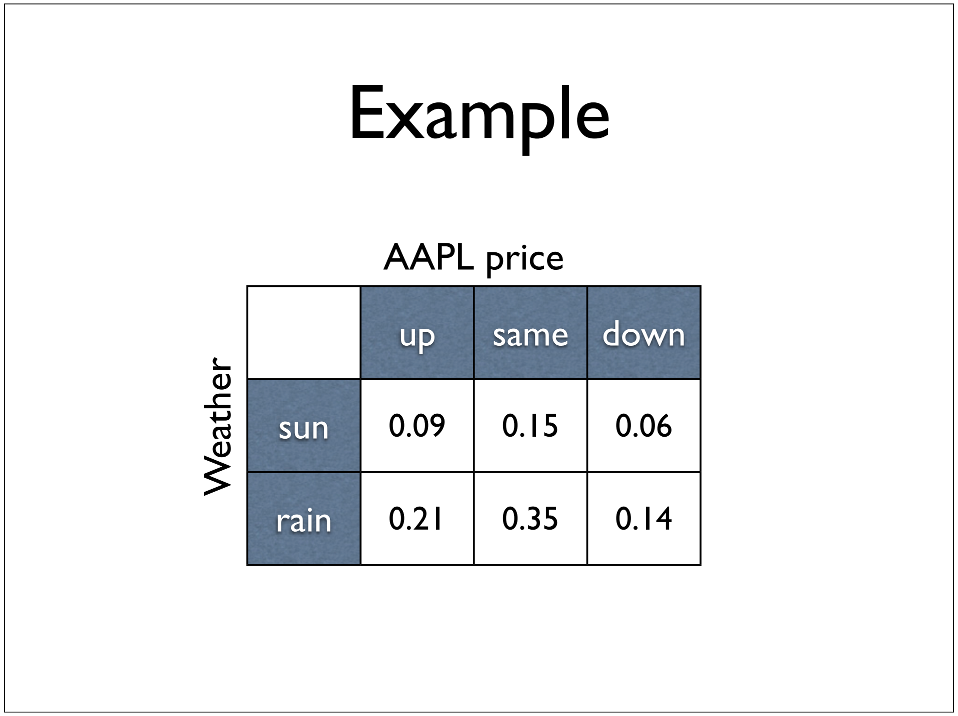


Probability

- Measure:
  - disjoint union:
  - e.g.:
  - interpretation:
- Distribution:
  - interpretation:
  - e.g.:

measure: fn mu from 2^Omega -> R+
subsets of sample space to reals >= 0
[note: R++ means +ve reals]
additive: events e1, e2, ..., e_k: mu(union e_i) = sum(mu(e_i))
implies mu(empty-set) = 0
e.g.: counting (mu(S) = |S|)
interp: “size” of set

dist’n: Omega measures 1; interp: probability of set
e.g.: uniform (1/|Omega| on each singleton)



Example

AAPL price

Weather

	up	same	down
sun	0.09	0.15	0.06
rain	0.21	0.35	0.14

note that they sum to 1
note we only need to list atomic events
work out P(sun & ~down) = .24
used disjoint union

=====

```
>> [.3; .7] * [.3 .5 .2]
    0.0900    0.1500    0.0600
    0.2100    0.3500    0.1400
```

Bigger example

AAPL price

PIT Weather

	up	same	down
sun	0.03	0.05	0.02
rain	0.07	0.12	0.05

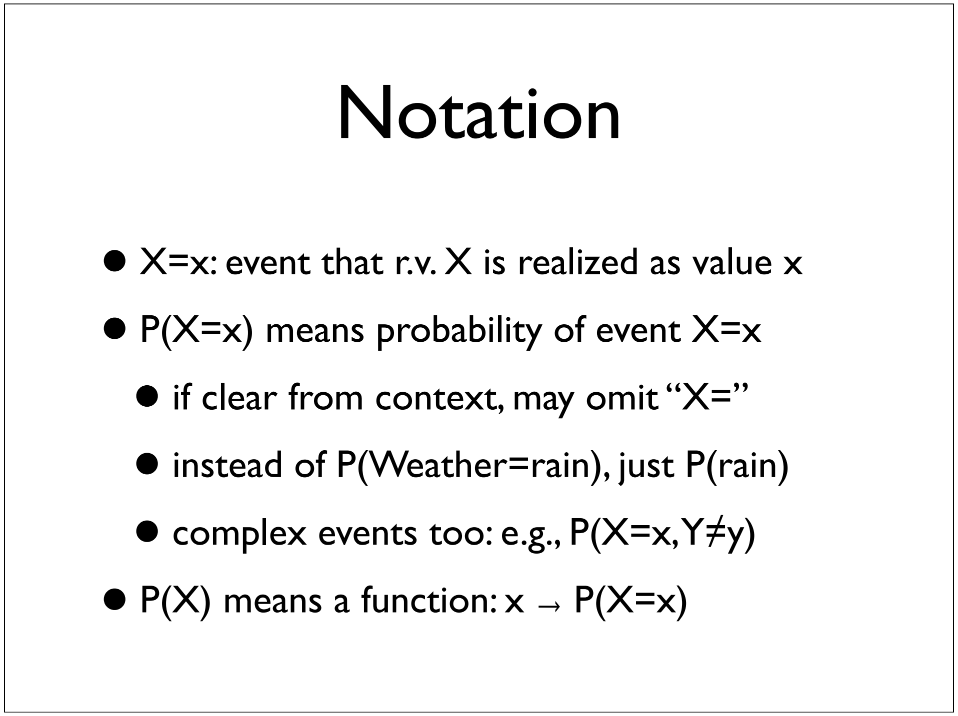
LAX Weather

	up	same	down
sun	0.14	0.23	0.09
rain	0.06	0.10	0.04

calculate P(up) = .03 + .07 + .14 + .06 = .3
P(down & sun) = .02 + .09 = .11

====

```
>> [.3; .7] * [.3 .5 .2] * (1/3)
ans =
   0.0300    0.0500    0.0200
   0.0700    0.1167    0.0467
>> [.7; .3] * [.3 .5 .2] * (2/3)
ans =
   0.1400    0.2333    0.0933
   0.0600    0.1000    0.0400
```



Notation

- $X=x$: event that r.v. X is realized as value x
- $P(X=x)$ means probability of event $X=x$
  - if clear from context, may omit "$X=$"
  - instead of P(Weather=rain), just P(rain)
  - complex events too: e.g., $P(X=x, Y\neq y)$
- $P(X)$ means a function: $x \to P(X=x)$

P: under some distribution understood from context -- may write P_theta if there are parameters theta

Functions of RVs

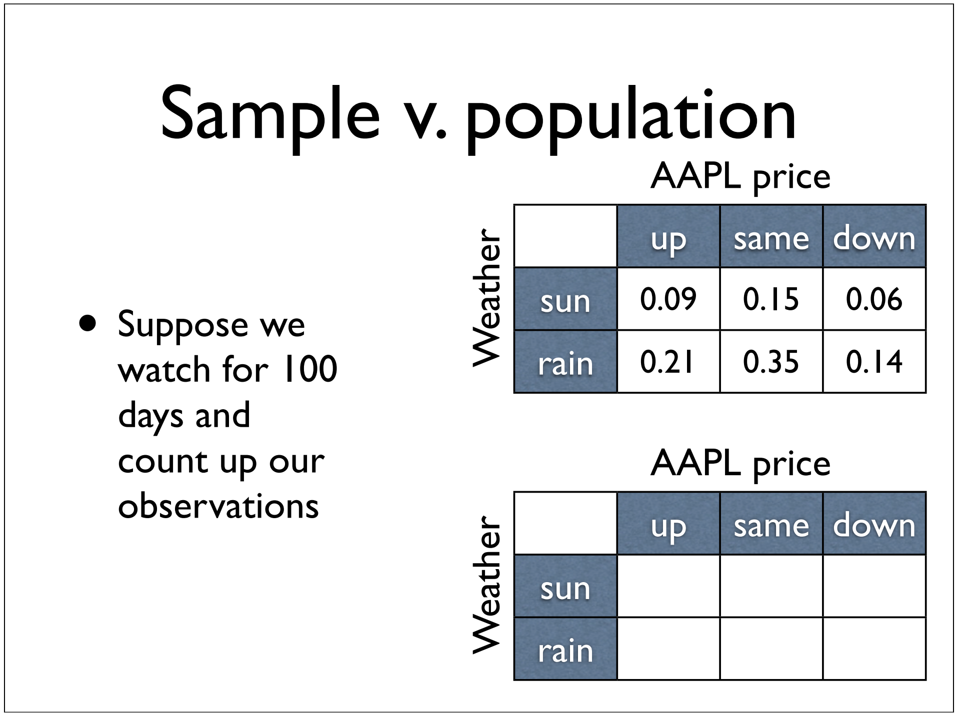
- Extend definition: any deterministic function of RVs is also an RV
- E.g.,

AAPL price

Weather

	up	same	down
sun	3	8	3
rain	0	5	0

eg: 3[sunny] + 5[same]
note bracket notation: *indicator* of event



Sample v. population

- Suppose we watch for 100 days and count up our observations

AAPL price

Weather

	up	same	down
sun	0.09	0.15	0.06
rain	0.21	0.35	0.14

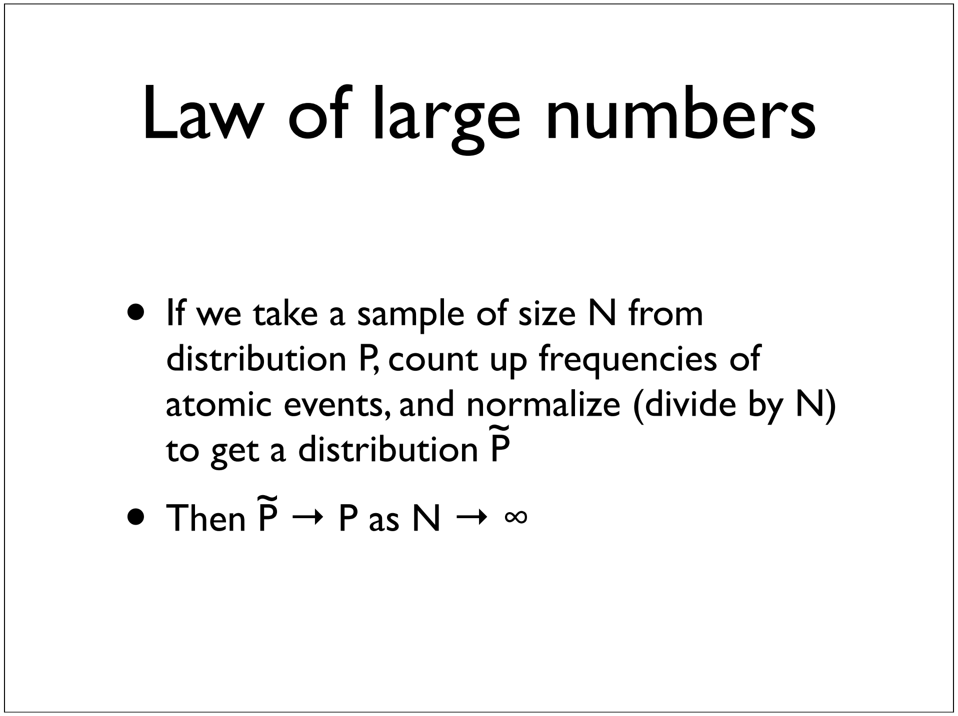
AAPL price

Weather

	up	same	down
sun			
rain			

write:
7 12 3
22 41 15
(actual matlab-generated sample)

note: if we normalize, get similar but not same dist’n as we started with



Law of large numbers

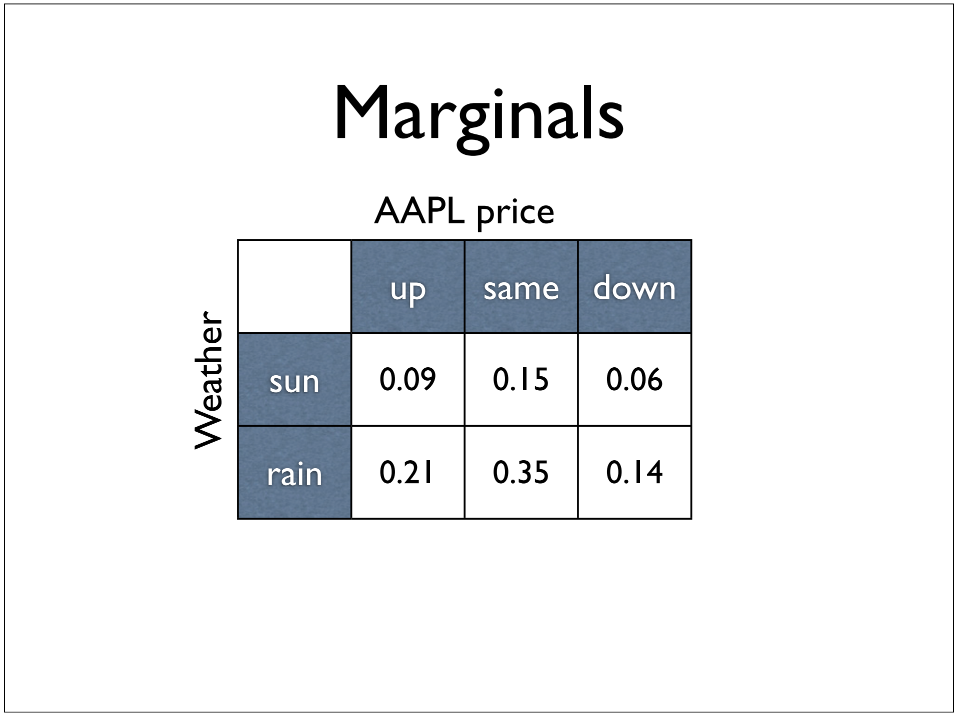
- If we take a sample of size N from distribution P, count up frequencies of atomic events, and normalize (divide by N) to get a distribution $\tilde{P}$
- Then $\tilde{P} \rightarrow P$ as $N \rightarrow \infty$

this and related properties are what allow learning from samples

Working w/ distributions

- Marginals
- Joint

marginal: get rid of an rv, get dist'n as if it weren't there
joint: before marginalization (to distinguish)

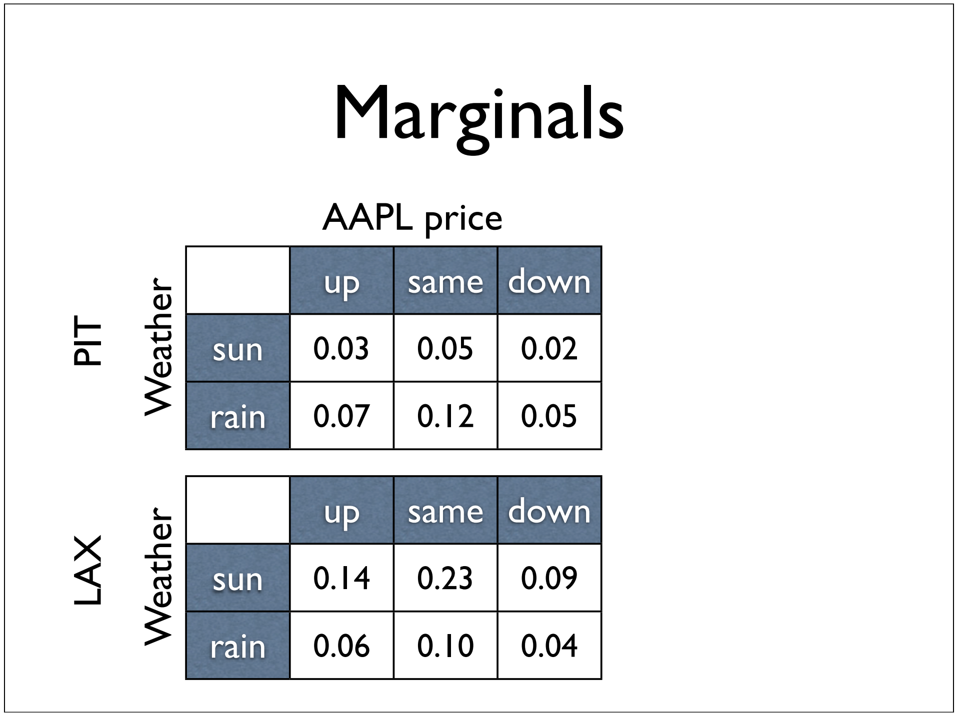


Marginals

AAPL price

Weather	up	same	down
sun	0.09	0.15	0.06
rain	0.21	0.35	0.14

[.3 .7] and [.3 .5 .2]
notation: P(Weather) or P(AAPL)



Marginals

AAPL price

PIT

Weather

	up	same	down
sun	0.03	0.05	0.02
rain	0.07	0.12	0.05

LAX

Weather

	up	same	down
sun	0.14	0.23	0.09
rain	0.06	0.10	0.04

marginalize out location, then AAPL

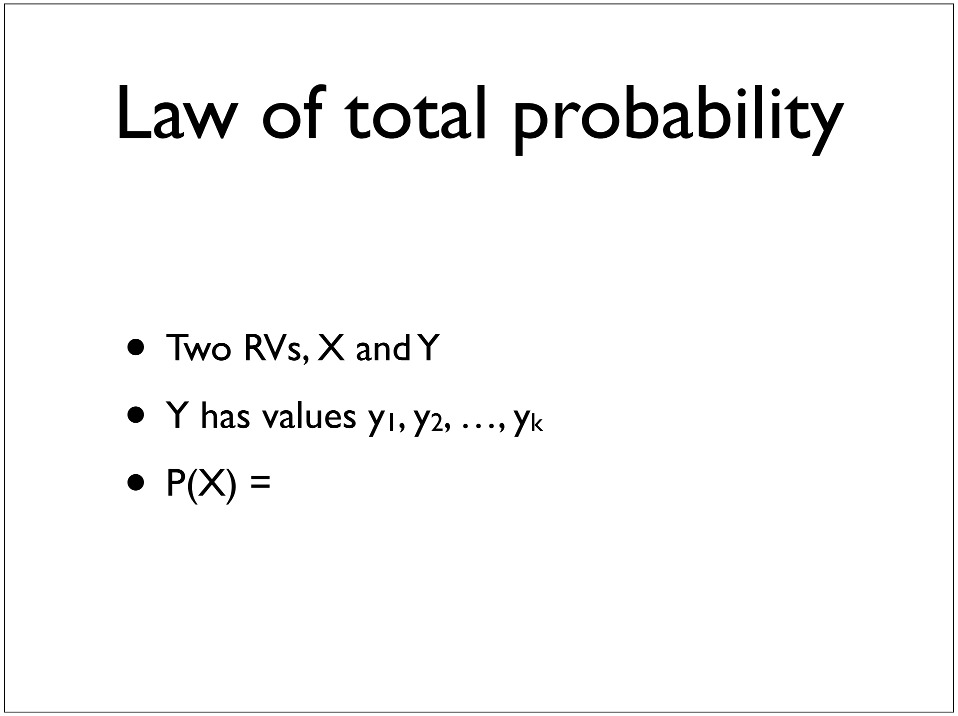
0.17	0.28	0.11
0.13	0.22	0.09

then [.56 .44]

===

if we had marginalized location then weather:

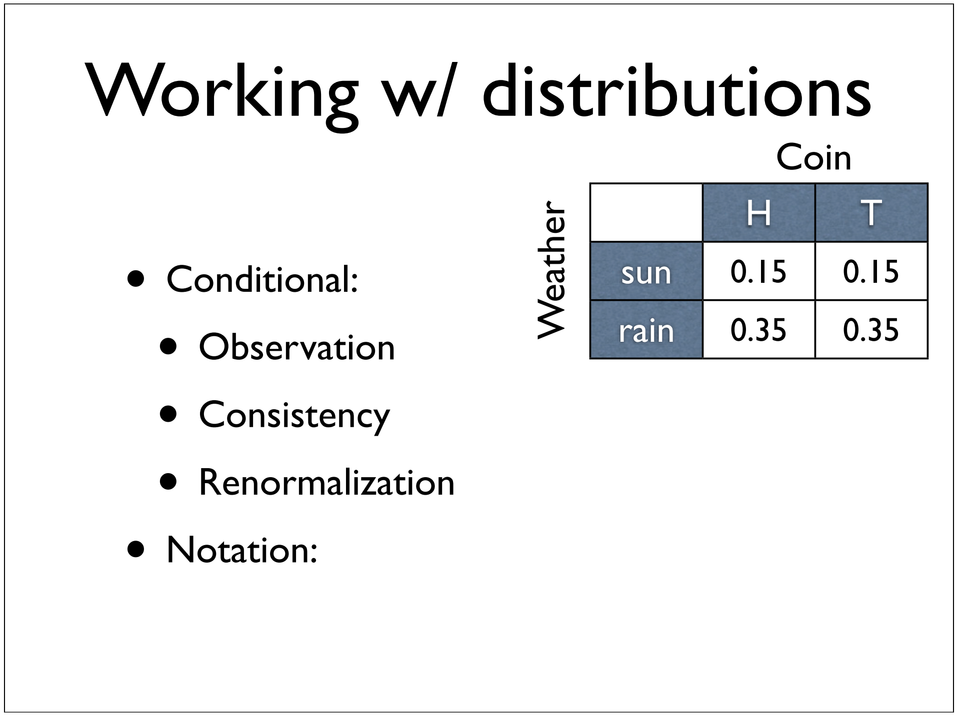
0.30	0.50	0.20



Law of total probability

- Two RVs, X and Y
- Y has values $y_1, y_2, \ldots, y_k$
- P(X) =

P(X) = P(X, Y=y1) + P(X, Y=y2) + …



Working w/ distributions

		Coin	
		H	T
Weather	sun	0.15	0.15
	rain	0.35	0.35

- Conditional:
  - Observation
  - Consistency
  - Renormalization
- Notation:

observation: an event that happened, or that we imagine happened -- e.g., coin H

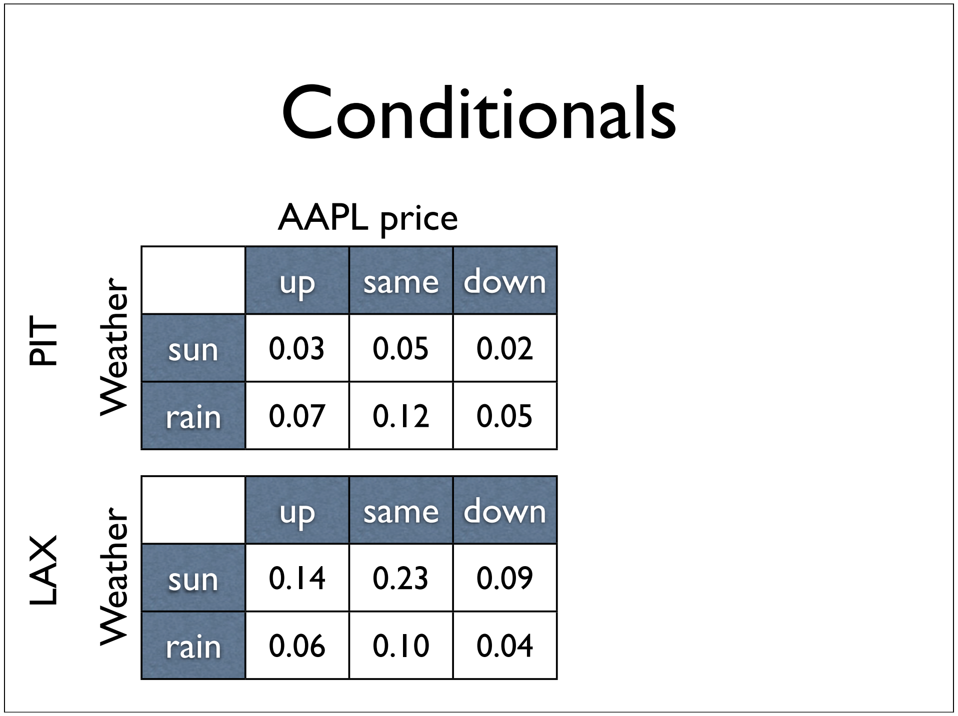
consistency: zero out impossibilities
note: every atomic event is either perfectly consistent or completely inconsistent w/ observed event

renorm: makes a distribution again

notation: P(Weather | Coin=H) or P(sun | H)
conditioning bar -- read as “given”

Conditionals in the literature

> When you have eliminated the impossible, whatever remains, however improbable, must be the truth.
>
> —Sir Arthur Conan Doyle, as Sherlock Holmes



Conditionals

AAPL price

PIT

Weather

	up	same	down
sun	0.03	0.05	0.02
rain	0.07	0.12	0.05

LAX

Weather

	up	same	down
sun	0.14	0.23	0.09
rain	0.06	0.10	0.04

condition on sun: P(sun) = .56

```
>> [.03 .05 .02; .14 .23 .09] / .56
ans = (table of location by AAPL)
    0.0536    0.0893    0.0357
    0.2500    0.4107    0.1607
```

now condition on AAPL=up
location: 1/6 5/6

In general

- Zero out all but some slice of high-D table
    - or an irregular set of entries
- Throw away zeros
    - unless irregular structure makes it inconvenient
- Renormalize
    - normalizer for P(. | event) is P(event)

Conditionals

- Thought experiment: what happens if we condition on an event of zero probability?

answer: undefined! Not useful to ask what happens in an impossible situation, so NaN is not a problem.

Notation

- $P(X \mid Y)$ is a function: $x, y \rightarrow P(X=x \mid Y=y)$
- As is standard, expressions are evaluated separately for each realization:
    - $P(X \mid Y)\ P(Y)$ means the function $x, y \rightarrow$

$P(X=x \mid Y=y)\ P(Y=y)$

Exercise

Monty Hall paradox
prize behind one door, other 2 empty (uniform)

say we pick #1; 3 cases: T1, T3, T3 (1/3 each)
T1: O2 or O3, equally
T2: O3
T3: O2

observe O2