ASEN 5044, Fall 2018

Statistical Estimation for Dynamical Systems

Lecture 9: Conditional Probabilities, Random Variables, Distributions and Density Functions

Prof. Nisar Ahmed (Nisar.Ahmed@Colorado.edu)

Tues 9/25/2018

University of Colorado Boulder

Aerospace Engineering

Announcements

- **HW 3 Due Thurs 9/28 at 11 am (before start of next lecture)**
- Submit to Canvas

- **First advanced topic lecture: this Friday 9/28 at <<TBD>>**
  - **Optional: targeted at PhD students, but all welcome to attend**
  - **Will post recorded lecture + slides to watch online**

- **Midterm 1: next Thursday 10/4**
  - One week long take home exam posted to Canvas
  - Due Thurs 10/11/2017 on Cavas
  - Open book/notes – honor code applies (must complete by yourself)
  - Will cover HW 1-4 (HW 4 Out 9/28, Due 10/4)

Overview

<u>Last time:</u> Intro to Probability

- Motivation
- Formal definitions: sample spaces, event spaces, axioms
- Joint probabilities
- Marginal probabilities

<u>Today:</u> other important fundamental concepts

- Conditional probabilities
- Bayes' Rule
- Dependent/independent probabilities
- Random variables (RVs)
- Probability distributions for RVs (discrete/continuous)
- Probability density functions (pdfs) for continuous RVs

READ: Chapter 2.4 in Simon book

Last Time: Marginal Probabilities

- 6-sided die example again: A: roll is even # (1=yes, 0 = no)
  B: roll is prime # (1=yes, 0 = no)

P(A & B)	B=0	B=1
A=0	1/6	2/6
A=1	2/6	1/6

$P(A=0) = 1/2$
$P(A=1) = 1/2$

$P(B=0) = 1/2$ $P(B=1) = 1/2$

Marginal Prob of $A=0$: $P(A=0) = \sum_b P(A=0, B=b)$
$= P(A=0, B=0) + P(A=0, B=1) = \frac{1}{6} + \frac{2}{6} = \frac{1}{2}$

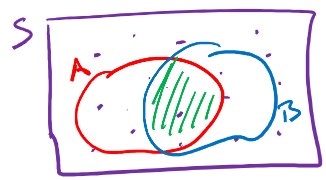
Marginal Prob of $A=1$: $P(A=1) = \sum_b P(A=1, B=b)$
$= P(A=1, B=0) + P(A=1, B=1) = \frac{2}{6} + \frac{1}{6} = \frac{1}{2}$

Likewise, can show that:

$P(B=0) = \sum_a P(A=a, B=0) = \frac{1}{2}$ $\qquad$ $P(B=1) = \sum_a P(A=a, B=1) = \frac{1}{2}$

Conditional Probabilities

- Are events A and B related, such that knowing whether/not B occurs alters prob. of A?



$P(A=a$ GIVEN $B=b$ is known to occur$)$

$= P(A=a \mid B=b)$

↳ "GIVEN" or "conditioned on" the event(s)

$$\triangleq \frac{P(A=a \,\&\, B=b)}{P(B=b)} = \frac{P(A=a \,\&\, B=b)}{\sum_a P(A=a \,\&\, B=b)}$$

→ Basically: restricting sample/outcome space S such that we only care about part where event $B=b$ must occur [renormalizing the joint probability to be consistent w/ given statement $B=b$)

→ For more than 2 events, e.g. $P(A,B,C)$

$$P(A=a \mid B=b, C=c) = \frac{P(A=a, B=b, C=c)}{P(B=b, C=c)} = \frac{P(A=a, B=b, C=c)}{\sum_a P(A=a, B=b, C=c)}, \text{ etc.}$$

$$P(A=a, C=c \mid B=b) = \frac{P(A=a, B=b, C=c)}{P(B=b)}$$

Example: Conditional Probabilities

- 6-sided die again:

A: roll is even # (1=yes, 0 = no)

B: roll is prime # (1=yes, 0 = no)

P(A & B)	B=0	B=1
A=0	1/6	2/6
A=1	2/6	1/6

$$P(B=0|A=1) = \frac{P(A{=}1\ \cap\ B{=}0)}{P(A{=}1)} = \frac{(2/6)}{(1/2)} = \frac{2}{3}$$

$$P(B=1|A=1) = \frac{P(A{=}1\ \cap\ B{=}1)}{P(A{=}1)} = \frac{(1/6)}{(1/2)} = \frac{1}{3}\ [= 1 - P(B=0|A=1)]$$

$$P(B=0|A=0) = \frac{P(A{=}0\ \cap\ B{=}0)}{P(A{=}0)} = \frac{(1/6)}{(1/2)} = \frac{1}{3}$$

$$P(A=0|B=0) = \frac{P(A{=}0\ \cap\ B{=}0)}{P(B{=}0)} = \frac{(1/6)}{(1/2)} = \frac{1}{3}$$

Consequences of Conditioning

- FACT #1: P(A,B) can always be **conditionally factored** in two ways

$$P(A,B) = P(A)\cdot \boxed{P(B|A)} = P(B)\cdot \boxed{P(A|B)}$$

prob. of B if A were hypothetically known

prob. of A if B were hypothetically known

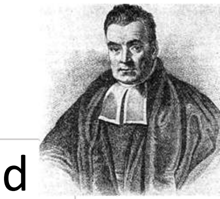
$$\left(P(A)\cdot \boxed{\frac{P(A,\ B)}{P(A)}} = P(B)\cdot \boxed{\frac{P(A,\ B)}{P(B)}}\right)$$

- FACT #2: **the law of total probability**: because of FACT #1 and definition of marginal distributions, we have

$$P(A=a) = \sum_b P(A=a, B=b) = \sum_b P(B=b)\cdot P(A=a|B=b)$$

$$P(B=b) = \sum_a P(A=a, B=b) = \sum_a P(A=a)\cdot P(B=b|A=a)$$

Bayes’ Rule for Reverse Conditioning



- Very handy for “**inverse problems**”, where we see the “effects” B=b (evidence) and want to infer the “cause” A (explanation), <u>based only on knowing</u> P(A) and P(B=b|A)
  - i.e. useful in cases where P(A) and P(B|A) are easy to specify, but P(A|B) is not…
- Allows us to update P(A) [prior belief in A] given new data B=b
  - P(A) [a priori, old belief before data] → P(A|B=b) [a posteriori, new belief given data]
- **Derivation:** start with FACT #1 from previous slide:

$$P(A,B) = P(A)\cdot P(B|A) = P(B)\cdot P(A|B)$$

Now, since $P(A)\cdot P(B|A) = P(B)\cdot P(A|B)$, re-arrange and solve for

$$P(A|B=b) = \frac{P(A)\cdot P(B{=}b|A)}{P(B{=}b)}$$

$\rightarrow$ but: $P(B=b) = \sum_a P(A=a, B=b) = \sum_a P(A=a)\cdot P(B=b|A=a)$, so:

Bayes' Rule ⊛

prior

$$\boxed{P(A|B=b) = \frac{P(A)\cdot P(B{=}b|A)}{\sum_a P(A{=}a)P(B{=}b|A{=}a)}}$$

posterior

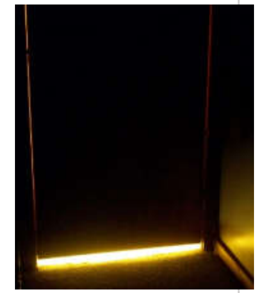
sometimes called "observation likelihood" or "likelihood"

Bayes' Rule Example: "Bayesian Inference"

- What is the probability that Prof. Ahmed is in his office given that lights are on?

A: Ahmed is in his office (0 = no, 1 = yes)

B: Lights are on in his office (0 = no, 1 = yes)



P(A=0)	P(A=1)
0.5	0.5

P(B=0|A=0)	P(B=1|A=0)
0.8	0.2

P(B=0|A=1)	P(B=1|A=1)
0.1	0.9

Want to use this data to find $P(A=1|B=1)$.

From Bayes' rule, we get

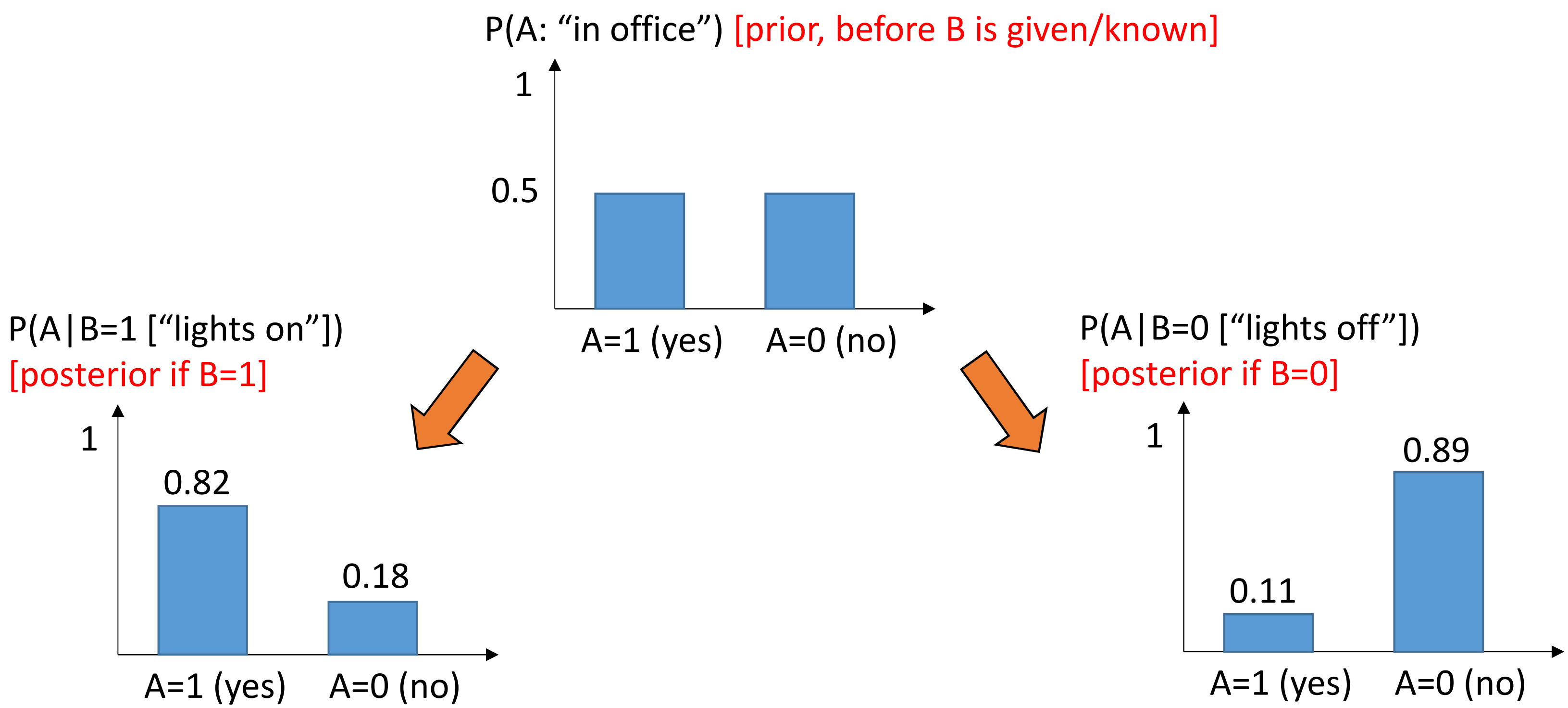
$$P(A=1|B=1) = \frac{P(A{=}1)\cdot P(B{=}1|A{=}1)}{\underbrace{\sum_a P(A{=}a)P(B{=}1|A{=}a)}} = \frac{P(A{=}1)\cdot P(B{=}1|A{=}1)}{P(A{=}0)\cdot P(B{=}1|A{=}0)+P(A{=}1)\cdot P(B{=}1|A{=}1)}$$

$0.55 = P(B=1)$: prob. of seeing lights on regardless of actual A value

$$= \frac{0.5\cdot 0.9}{\underbrace{(0.5\cdot 0.2)+(0.5\cdot 0.9)}} = \frac{0.9}{0.2+0.9} \approx 0.82$$

Bayes' Rule Example

- Can also apply Bayes' rule to compute full posterior distribution of A given B=0, or given B=1
- Compare posterior (probs of A after Bayes' rule) to prior (probs of A before Bayes' rule, i.e. NOT given B)



Independent Events and Independent Probabilities

- If knowledge of the occurrence of event B <u>never</u> alters P(A), then we say that A and B are **independent events**

if $A \perp\!\!\!\perp B$ (independent), then $P(A=a \,\&\, B=b) = P(A=a)\cdot P(B=b)$ [product of marginals] $\forall a, b$

[i.e. events A & B have no information about each other]

$\longrightarrow$ so if $A \perp\!\!\!\perp B$, then $P(A|B) = \dfrac{P(A \,\&\, B)}{P(B)} = \dfrac{P(A)\cdot \cancel{P(B)}}{\cancel{P(B)}} = P(A)$

(*) must hold hold for EVERY POSSIBLE VALUE $A=a$ & $B=b$!

$\Rightarrow \boxed{P(A|B) = P(A) \text{ if } A \perp\!\!\!\perp B}$

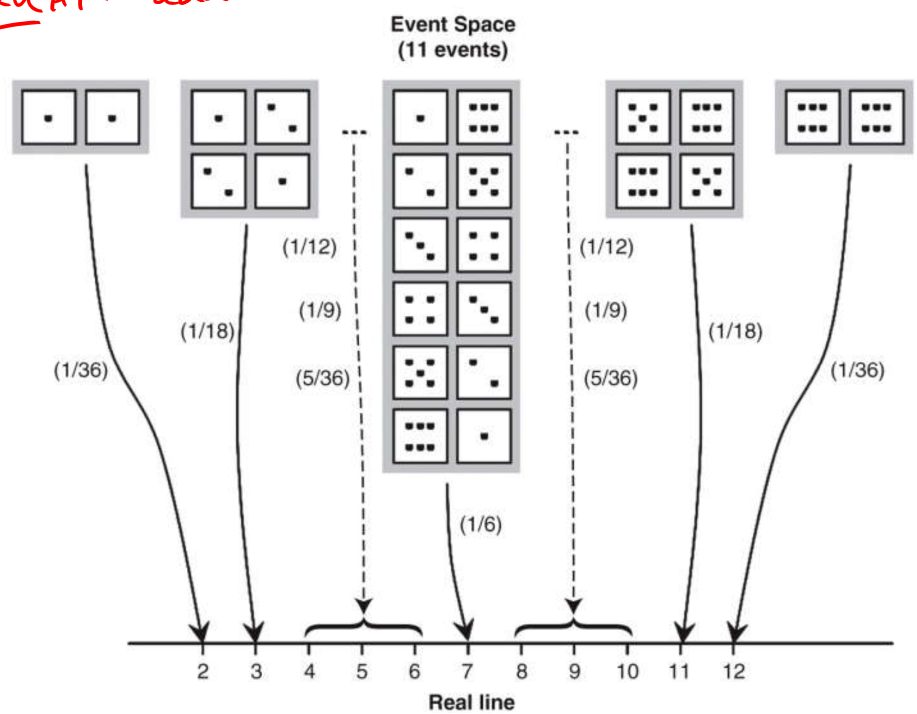
& likewise: $\boxed{P(B|A) = P(B) \text{ if } B \perp\!\!\!\perp A}$

Ex.:
- prob. of getting "heads" on 1st & 2nd flips of a coin are $\perp\!\!\!\perp$
- prob. of getting prime & even die rolls are not $\perp\!\!\!\perp$

Random Variables

- A **random variable (RV)** is a function that maps every point in an event space $\{A_i\}$ to points on the real line
- Example: RVs for two dice

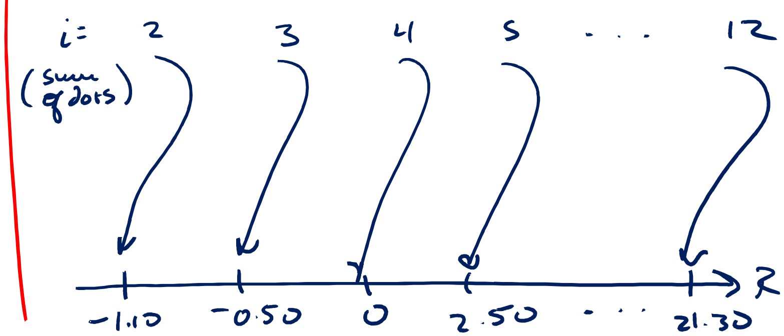
Exm #1: add the dots on the die



Exm #2

Suppose you get paid some arbitrary reward R for getting certain # of dots

→ can assign $R = r$ to be a random var.



What's the Point of Defining RVs?

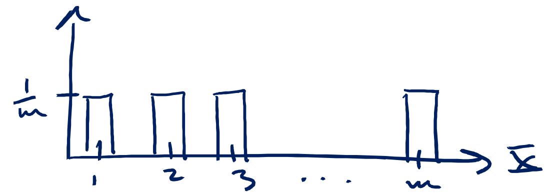
- Much easier to work with/visualize probabilities on RVs than "raw" events and outcomes
- Think of **"random quantity"** as another name for a "random variable"
- Other examples of random quantities (or RVs) that can be readily assigned to otherwise non-quantitative outcomes/events for random experiments:
  - Select a person in this room at random & then **measure** their height **h**, or weight **w**, or age **a**, or GPA **g**,…
    → any particular person is now "quantified" by a number on the real line
    → example of **continuous random quantity (i.e. a continuous RV)**
  - Flip a coin 5 times & then count (**i.e. measure**) number of heads → any particular outcome (e.g. THHHH, HTHHH, HHTHH,…) now maps to a number on the real line (integer in this case)
    → example of **discrete random quantity (i.e. a discrete RV)**
  - Take a reading from a Geiger counter and report the value you see on the dial
    → continuous RV (identity mapping)

Discrete Random Variables

X is a discrete RV if X maps outcomes/events to integer quantities

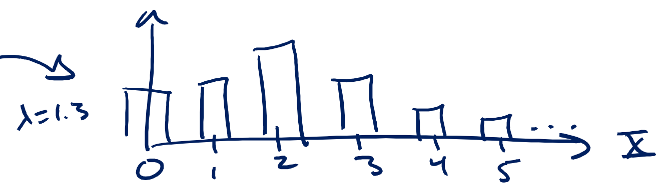
- Can be **finite or countably infinite** (e.g. number of e-mails between now and midnight)
- A fxn that assigns a single probability to each possible realization x of X, i.e. P(X=x), is called a **probability mass function (pmf)**
- The **pmf** is also sometimes called a **discrete probability distribution**
- **Example discrete probability distributions (pmfs):**

Uniform: $P(X=x) = \frac{1}{m},\ \text{for } x \in \{1, ..., m\}$



Poisson: $P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!},\ \text{for } x \in \{0, 1, 2, ...,\},\ \lambda \geq 0$

[countably infinite]



Bernoulli: $P(X=x) = p^x (1-p)^{1-x},\ \text{for } x \in \{0, 1\}$

(binary outcomes, e.g. coin flips: prob getting "1" = p)

Binomial: $P(X=x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x},\ \text{for } x \in \{0, 1, ..., n\}$ → prob. on total # of "1's" in sequence of n Bernoulli.

Continuous Random Variables

X is a continuous RV if it maps to continuous quantities (real-valued, for our purposes)

- **Uncountably infinite** (e.g. there is a continuum of numbers between 100 and 100.1)
- We need to be careful about assigning and defining what P(X=x) really means!!!
- Example: spinning the pointer on a wheel

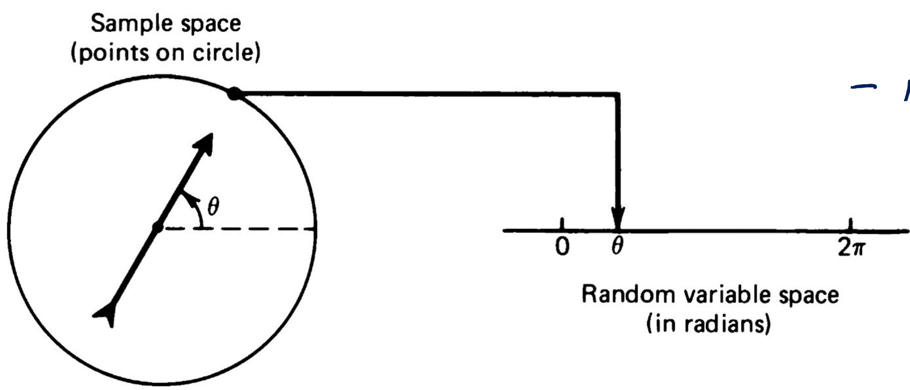


Figure 1.3a

– All possible θ are likely [if wheel is fair...]

– Naive "$\frac{1}{N}$" relative frequency calc of prob:

$$\text{Prob}(\theta) = \lim_{N \to \infty} \frac{1}{N} = 0$$

→ But then sum over Prob(θ) $\forall \theta$ = 0

→ Violates our axioms! need sum over Prob(θ) $\forall \theta$ = 1!

- **Recall:** probabilities defined on **events** for outcome space -- so we **need a way to properly define events over a continuum of outcomes,** and then assign probabilities to such events …
- Most natural way: define events to be **intervals (lengths) on continuous real line**
- So then we **need a way to assign probabilities to arbitrary intervals (events) on real line**

Probability Density Function (pdf)

- A fxn that assigns a single probability to each possible interval (x_1,x_2) of X=x, i.e. $P(x_1 < x < x_2)$, is called a **probability density function (pdf)**
  - The **pdf** is also sometimes called a **continuous probability distribution**
  - **Since probability is dimensionless, it follows that the pdf must have units = 1/[units of X]**
  - **Example pdfs:** uniform, Gaussian, exponential, Gamma, Beta, Rayleigh, Student's-t, Laplace, Weibull…
- Formally:
  - Event: $\{x \overset{\text{"such that"}}{:} \xi - d\xi < x < \xi\} \rightarrow$ ie either x falls inside this interval, or it does not
  - The probability density function (pdf) of a scalar RV:

$$\lim_{d\xi \to 0} \frac{P(\xi - d\xi < x < \xi)}{d\xi} \triangleq p_x(\xi) = p_X(x) = p(x)$$

  - From axioms of probability, it follows that

$$P(\eta < x \leq \xi) = \int_{\eta}^{\xi} p(x)\,dx = c(\xi) - c(\eta)$$

  - **Cumulative distribution function (cdf)**:

$$P(-\infty < x \leq \xi) = \int_{-\infty}^{\xi} p(x)\,dx = c(\xi) \longrightarrow p(x) = \frac{d\,P(-\infty < x \leq \xi)}{dx}$$

(ie pdf is derivative of cdf, when cdf is continuous & differentiable)

PDF Example: Spinning Pointer on Wheel

- If spinner fairly constructed, then θ has uniform pdf : $\Theta \sim \mathcal{U}[a,b]$, $a = 0$, $b = 2\pi$

distributed as

$$\mathcal{U}[0,2\pi] = \begin{cases} \frac{1}{2\pi}, & x \in [0, 2\pi] \\ 0, & \text{o'wise} \end{cases}$$