L02 probability

Random variables

let x be a r.v. — "an outcome of some random event"

x : Will it rain tommorrow?

OR

"a specific datum about which we may have incomplete knowledge

x: do I have cancer?

A random variable has a domain,

$x \in \{\text{true}, \text{false}\}$

$x \in \mathbb{R}^?$

Probability

Our belief about an r.v., assigning a prob. to each value in the domain such that

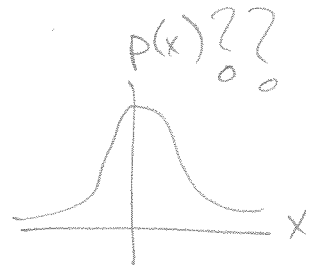
$$\sum_i p(x=i) = 1, \text{ and } \forall_i\, p(x=i) \geq 0,$$

Discrete Probabilities

e.g. dice. $p(x=1) = 1/6$.

Continuous

We can plot "probability":



- What is $p(x=0)$? 0!

  These familar plots are of prob. density.

- $p(x=0) = \int_{t=0}^{0} \text{pdf}(x=t)\, dt = 0$. ⟸ no support.

- What restrictions on pdf(x)?

  $pdf(x) \geq 0$ but NOT $pdf(x) \leq 1$.

- Is $p(x=t)$ always zero for continuous r.v. x?

  NO $\Rightarrow$ Dirac Deltas.

- PDFs usually written $p(x)$!

  We'll have to live w/ this. Usually clear in context.

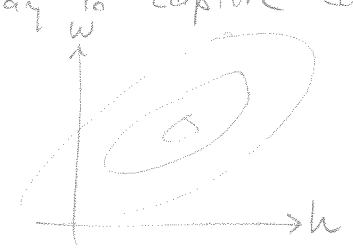
Multi-valued r.v.'s

"A set of (possibly related) datums about which we may have incomplete knowledge."

e.g. Where am I? $x \in \mathbb{R}^2$

Important to consider multi-valued R.V.'s when the datum are related $\Rightarrow$ only way to capture correlations.

$$x \in \begin{bmatrix} \text{weight} \\ \text{height} \end{bmatrix}$$



Joint Distribution

Specifies $p(\vec{x})$ completely. Everything we could ever want is in joint distribution.

> Not always an efficient representation

Consider

x_1	x_2	$p(x_1 = \_\_, x_2 = \_\_)$	
t	t	0.09	0.3×0.3
t	h	0.21	0.3×0.7
h	t	0.21	0.7×0.3
h	h	0.49	0.7×0.7

Could have written:

$$p(x_1, x_2) = p(x_1)\, p(x_2)$$

> How much memory to store a joint distribution for an r.v. with domain $\{0,1\}^N$? $\quad 2^N - 1$

why? $\sum p() = 1$!

<u>Independence</u> – allows factorizations

is important — we'll see this repeatedly during term.

If $p(a, b) = p(a)\,p(b) \Rightarrow$ "a & b are independent"

How much storage if a, b can be factored?

<u>Marginal Distributions</u>

Joint $\quad\quad\quad\quad\quad\quad\quad\quad$ Marginal

$p(x_1, x_2) \Rightarrow$ Marginalization $\Rightarrow p(x_1)$

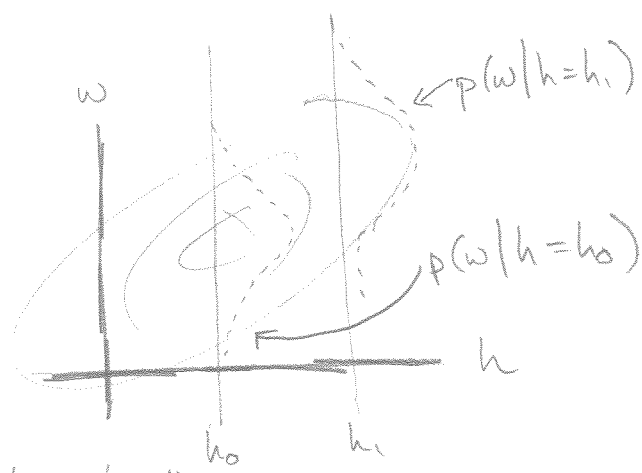
(integration over x_2)

"Unmarginalizing"

[show example with coins above].

<u>Conditional Distributions</u>

$p(\text{weight} \mid \text{height})$



$p(\text{weight} \mid \text{height} = h_0)$

$\neq p(\text{weight} \mid \text{height} = h_1)$

$\Rightarrow$ "weight & height are dependent"

$p(a, b \mid c) = p(a \mid c)\, p(b \mid c) \Leftrightarrow$ "a & b are conditionally independent given c"

Q: does this imply a & b are C.I.? No!

Evidence (c) can render dependent r.v's. C.I.

Color / Shape cards exercise.

Q1. How to represent?

- Could have one r.v. $x \in \{$ bluesquare, redsquare, ... $\}$
- Bad choice — ignores structure!
- $x \in \begin{bmatrix} \{\text{red, blue, yellow}\} \\ \{\text{square, circle, triangle}\} \end{bmatrix}$

Q2. Tally joint.

Q3. Marginal $p(\text{blue}) = \sum_{i \in \{\text{square, circle, triangle}\}} p(\text{blue}, i)$

Q4. Conditional $P(\text{yellow} | \text{circle})$

Critical Prob. Rules

Product Rule

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

When valid? Always!

Derive Bayes' rule from Product Rule

> Allows reversal of diagnostic & causal info!

$p(\text{fever} \mid \text{cold})$	$p(\text{cold} \mid \text{fever})$
CAUSAL	DIAGNOSTIC
"easy" to measure	harder to measure

Law of total Probability

$$\int p(x)\,dx = 1.$$

Useful b/c we can ignore scale factors & normalize after the fact.

e.g. $p(a|e) = \frac{p(a,e)}{p(e)} \propto p(a,e)$

e.g. maybe computing $p(b)$ is annoying, but normalizing $p(a|e)$ is easy.

Checkpoint - Bayes.

A coin (possibly crooked) comes up heads, heads, tails.

What is $p(\underbrace{q}_{\text{Prob of heads}} \mid \underbrace{\text{toss}_1 = h,\ \text{toss}_2 = h,\ \text{toss}_3 = t}_{\text{conditioned on evidence}})$?

- Contrast $P(q \mid hht)$ and $P(q)$

  no evidence — our prior!

Not sure how to compute $P(q \mid hht)$, but $p(hht \mid q)$ is easy. Bayes' rule!

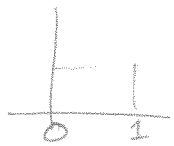
$$P(q \mid hht) = \frac{P(hht \mid q)\, P(q)}{P(hht)}$$

$q^2(1-q)$

our subjective prior.

independent of q — won't affect ML estimate of q. Can also view as $\int p(hht \mid q)\, p(q)\, dq$.

We now know $p(q \mid hht)$ and could do useful things like compute the likelihood of the next toss. e.g. $E[t_4 \sim p(q \mid hht)] = \int_0^1 q\,(q^2 - q^3)\,dq = 3/5$.

Let's posit $p(q) =$ 

$$P(q \mid hht) = \frac{q^2(1-q)}{p(hht)} = \frac{q^2 - q^3}{p(hht)} \propto q^2 - q^3$$

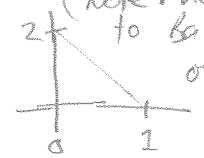
ML solution: (Why is ML useful if we know $p(q \mid hht)$?)

$$\frac{\partial\, p(q \mid hht)}{\partial q} = 2q - 3q^2 = 0 \quad \Rightarrow q = 2/3\ !$$

(Note: not equal to Bayesian estimate of 3/5)

The other solution, $q = 1$, is a minimum.

Let's posit instead $p(q) = 2 - 2q$ ...



$$q^2(1-q)(2-2q) = (q^2 - q^3)(2-2q) = 2q^2 - 2q^3 - 2q^3 + 2q^4$$

Diff. $\Rightarrow 4q - 12q^2 + 8q^3 = 0 \Rightarrow 2q^2 - 3q + 1 = 0 \Rightarrow \frac{+3 \pm \sqrt{9-8}}{4}$

$q = 1/2$

Expectation

$$E[x] = \int_{-\infty}^{\infty} x \; p(x) \; dx \qquad \text{"weighted average"}$$

Properties (prove):

$$\begin{cases} E[A] = A & \text{(easy)} \\ E[Ax] = A\,E[x] & \text{(easy)} \\ E[A+x] = A + E[x] & \text{(easy)} \\ E[x+y] = E[x] + E[y] & \end{cases}$$

$$\iint (x+y)\,p(x,y)\,dx\,dy$$

$$= \iint x\,p(x,y)\,dx\,dy + \iint y\,p(x,y)\,dx\,dy$$

$$= \int x \underbrace{\int p(x,y)\,dy}_{p(x)} dx + \int y \underbrace{\int p(x,y)\,dx}_{p(y)} dy$$

$p(x) \leftarrow$ (marginalization) $\rightarrow p(y)$

$$= \int x\,p(x)\,dx + \int y\,p(y)\,dy$$

$$= E[x] + E[y]$$

Anti-properties $\qquad E[xy] \neq E[x]\,E[y]$

Expectation of multi-dimensional r.v.'s:

$$E\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} E(x) \\ E(y) \end{bmatrix}$$

VAR properties: $\mathrm{VAR}(x+y) = \mathrm{VAR}(x) + \mathrm{VAR}(y) \qquad \mathrm{VAR}(Ax) = A^2 x \quad \leftarrow$ prove!

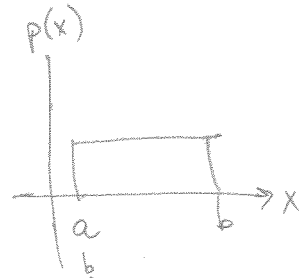
Variance

$$\mathrm{Var}(x) = E\left[\left(x - E(x)\right)^2\right]$$

defined & useful even for non-Gaussian distributions!

Checkpoint:

Mean & Variance of



?

$$E(x) = \frac{a+b}{2} = \int_{-\infty}^{\infty} x\, p(x)\, dx = \int_a^b x \cdot \frac{1}{b-a}\, dx = \frac{1}{b-a} \cdot \frac{x^2}{2}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

by inspection!

$$\mathrm{Var}(x) = \int_a^b \Big(x - \underbrace{\frac{a+b}{2}}_{E(x)}\Big)^2 \underbrace{\frac{1}{b-a}}_{p(x)}\, dx = \frac{1}{b-a}\int_a^b x^2 - (a+b)x + \frac{(a+b)^2}{4}\, dx$$

$$= \frac{1}{b-a} \cdot \left[\frac{x^3}{3} - \frac{(a+b)x^2}{2} + \frac{\overbrace{(a+b)^2}^{a^2+2ab+b^2}}{4} x\right]\Bigg|_a^b$$

$$= \frac{1}{b-a} \cdot \left[\frac{b^3}{3} - \frac{a^3}{3} - \frac{ab^2}{2} - \frac{b^3}{2} + \frac{a^3}{2} + \frac{a^2b}{2} + \frac{a^2b}{4} + \frac{2ab^2}{4} + \frac{b^3}{4} - \frac{a^3}{4} - \frac{2a^2b}{4} - \frac{ab^2}{4}\right]$$

$$= \frac{1}{b-a} \cdot \left[\frac{b^3}{12} - \frac{a^3}{12} + \frac{a^2b}{4} - \frac{ab^2}{4}\right] = \frac{1}{12}(b-a)^2$$

Computing sample variances in one pass

Don't need to compute $E(x)$ first:

$$E\left[(x - E(x))^2\right] = E\left[x^2 - 2x\,E(x) + E(x)^2\right] = E[x^2] - 2E(x)E(x) + E(x)^2$$

$$\boxed{= E[x^2] - E[x]^2}$$