Random variables

let \( x \) be a r.v. — an outcome of some random event

\( x \): Will it rain tomorrow?

OR

"a specific datum about which we may have incomplete knowledge

\( x \): do I have cancer?

A random variable has a domain,

\( x \in \{ \text{true, false} \} \)

\( x \in \mathbb{R} \)

Probability

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

\[ \sum_{i} p(x=i) = 1, \text{ and } \forall i \; p(x=i) > 0. \]

Discrete Probabilities

e.g. dice, \( p(x=1) = \frac{1}{6} \)

Continuous

We can plot "probability":

\( p(x) ?? \)

\( p(x) \)

\( x \)

\( \Rightarrow \) What is \( p(x=0) \)? 0!

These familiar plots are of prob. density,

\( p(x=0) = \int_{-\infty}^{0} \text{pdf}(x=t) \, dt = 0. \Leftrightarrow \) no support.
What restrictions on pdf(x)?
\[ \text{pdf}(x) \geq 0 \quad \text{but not} \quad \text{pdf}(x) \leq 1. \]

Is \( p(x=t) \) always zero for continuous r.v. \( x \)?
\( \text{No} \Rightarrow \delta \text{Deltas} \)

PDFs usually written \( p(x) \)!
We’ll have to live with this. Usually clear in context.

**Multi-valued r.v.'s**

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

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

Important to consider multi-valued r.v.'s when the datums are related =) only way to capture correlations.

\[ x \in \begin{bmatrix} \text{weight} \\ \text{height} \end{bmatrix} \]

**Joint Distribution**

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

> Not always an efficient representation

\[
\begin{array}{ccc}
X_1 & X_2 & p(X_1=\_ \text{and} \ X_2=\_)\\
\hline
\text{t} & \text{t} & 0.09 & 0.3 \times 0.3 \\
\text{t} & \text{h} & 0.21 & 0.3 \times 0.7 \\
\text{h} & \text{t} & 0.21 & 0.7 \times 0.3 \\
\text{h} & \text{h} & 0.49 & 0.7 \times 0.7
\end{array}
\]
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 \(\{0, 1\}^n\)?: \(2^n - 1\)

**Independence** allows factorizations

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

If \(p(a, b) = p(a) p(b) \Rightarrow "a \& b \text{ are independent}"\)

How much storage if \(a, b\) can be factored?

**Marginal Distributions**

Joint

\[ p(x_1, x_2) \Rightarrow \text{Marginalization} \Rightarrow p(x_1) \]

(integration over \(x_2\))

"Unmarginalizing"

**Conditional Distributions**

\[ p(\text{weight} | \text{height}) \]

\[ p(\text{weight} | \text{height} = h_0) \]

\[ \neq p(\text{weight} | \text{height} = h_1) \]

\(\Rightarrow "\text{weight} \& \text{height} \text{ are dependent}"\)

\[ p(a, b | c) = p(a | c) p(b | c) \Rightarrow "a \& b \text{ are conditionally independent given } c" \]

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. \( \mathbf{X} \in \{\text{blue}_1, \text{red}_1, \ldots\} \)
   - Bad choice - ignores structure!
     \[ \mathbf{X} \in \{\text{red}, \text{blue}, \text{yellow}\} \]
     \[ \{\text{square}, \text{circle}, \text{triangle}\} \]

Q2. Tally joint.

Q3. Marginal \( \mathbf{P}(\text{blue}) = \sum_{i} \mathbf{P}(\text{blue}, i) \)
    \[ i \in \{\text{square}, \text{circle}, \text{triangle}\} \]

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

---

Critical Prob. Rules

Product Rule
\[
\mathbf{P}(A, B) = \mathbf{P}(A|B) \mathbf{P}(B)
\]

Derive Bayes' rule from Product Rule

- Allows reversal of diagnostic & causal info!
  \[
  \mathbf{P}(\text{fever}|\text{cold}) \quad \mathbf{P}(\text{cold}|\text{fever})
  \]
  Causal \quad \text{Diagnostic}

  "easy" to measure \quad \text{harder to measure}

Lowest total probability
\[
\int p(x) \, dx = 1.
\]

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

E.g. \( p(a|e) = \frac{p(a,e)}{p(e)} \times p(a,e) \)

E.g. maybe computing \( p(b) \) is annoying, but normalizing \( p(b) \) is easy.
Let's postulate $p(\theta) = 2 - 2\theta$.

\[
p(\theta | \text{light}) = \frac{p(\theta) p(\text{light} | \theta)}{p(\text{light})} = \frac{2 - 2\theta}{2 - 2\theta} = 1,
\]

so $\hat{\theta}$ is a minimum.

The other solution:

\[
p(\theta | \text{dark}) = \frac{p(\theta) p(\text{dark} | \theta)}{p(\text{dark})} = \frac{2 - 2\theta}{2 - 2\theta} = 1.
\]

Let's postulate $p(\theta) = \frac{2 - \theta}{2 - \theta}$.

\[
p(\theta | \text{light}) = \frac{p(\theta)}{p(\text{light})} = \frac{\theta}{2 - 2\theta} = \frac{\theta}{2 - \theta}.
\]

As $\theta \to 0$, $p(\theta | \text{light}) \to 1$, so we can't use the ML estimate independent of $\theta$.

Prior

\[
p(\text{light}) = p(\theta | \text{light}) p(\theta).
\]

Is easy. Bayer's rule.

Contrast $p(\theta | \text{light})$ and $p(\theta)$, conditioned on evidence.

What is $p(\theta | \text{light} = \text{light})$? Loss = $\ell(\theta, 0) = 0$.

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

\[
\text{Checkpoint - Bayers.}
\]
Expectation

\[ E[x] = \int_{-\infty}^{\infty} x \cdot p(x) \, dx \]  

"weighted average"

Properties (prove):

\[ \begin{align*}
E[A] &= A \\
E[Ax] &= A \cdot E[x] \\
E[A+x] &= A + E[x] \\
E[x+y] &= E[x] + E[y]
\end{align*} \]

\[ \int (x+y) p(x,y) \, dx \, dy \]

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

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

\[ \underline{p(y)} \quad \underline{(marginalization)} \quad \underline{p(y)} \]

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

\[ = E[x] + E[y] \]

Anti-properties

\[ E[xy] \neq E[x] \cdot E[y] \]

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:

\[ \text{VAR}(x+y) = \text{VAR}(x) + \text{VAR}(y) \]

\[ \text{VAR}(Ax) = A^2 \text{VAR}(x) \quad \text{prove!} \]
Variance

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

defined & useful even for non-Gaussian distributions!

Checkpoint:

Mean & Variance of \( f(x) \) over \( (a, b) \)

\[ E(x) = \frac{a+b}{2} = \int_a^b x \ p(x) \ dx = \int_a^b x \ \frac{1}{b-a} \ dx = \frac{1}{b-a} \ \frac{x^2}{2} \bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2} \]

by inspection

\[ \text{Var}(x) = E\left[(x - \frac{a+b}{2})^2\right] = \int_a^b (x - \frac{a+b}{2})^2 \ \frac{1}{b-a} \ dx \]

\[ = \frac{1}{b-a} \left[ \frac{b^3}{3} - \frac{(a+b)x^2}{2} + \frac{(a+b)^2}{4} \right] \bigg|_a^b \]

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

\[ + \frac{a^2b+2ab^2+b^3}{4} - \frac{a^3}{4} - 2ab^2 - \frac{a^2b^2}{4} \]

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

Computing sample variances in one pass

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

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

\[ = E[x^2] - E[x]^2 \]