Probabilities and Associations, Capacities and Causation

Types of Variables

- Discrete / Categorical
  - Only finitely many values
- Continuous / Real-valued
  - Infinitely many values
    - Despite the name, rarely continuum many values
- Variable type can be determined by…
  - Underlying objects (gender is discrete)
  - Measurement device (air pressure is continuous)
  - Pragmatics (height is sometimes just tall / short)

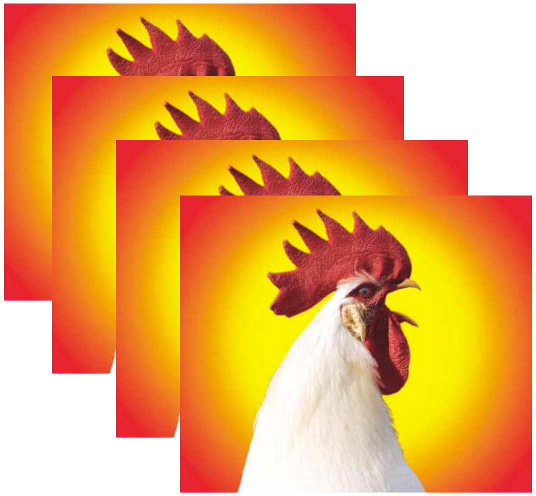
Changing Variable Types

- Discretization: Continuous → Discrete
  - Requires points of “meaningful difference” among the continuous values
    - Otherwise, very small differences (in continuous space) might translate into large differences (in discrete space)
- Interpolation: Discrete → Continuous
  - Requires a meaningful ordering of the values
    - Trivial for binary values (e.g., male / female)
    - Works for “really” continuous values that are discretized for measurement (e.g., high / medium / low)
    - But not for other sets of values (e.g., red / blue / green)

Frequency

- Given some (measured) population of individuals, the frequency of variable X having a particular value x is:
  - The number of individuals in the population with $X = x$; divided by
  - The number of individuals in the population

Frequency



- {Rooster, Apple, Frog}
  - – Fr(Picture = Frog) = 1/2

- {Plant, Animal}
  - – Fr(Picture = Plant) = 1/6

- {Alive, Dead}
  - – Fr(Picture = Alive) = 1

Probability Distribution

- Observed frequency is a noisy measure of the “true” probability
- Discrete variable ⇒ Probability distribution
  - $P(X = x)$ is defined for all possible x
- Example:
  - P(Picture = Apple) = 0.19
  - P(Picture = Frog) = 0.48
  - P(Picture = Rooster) = 0.33

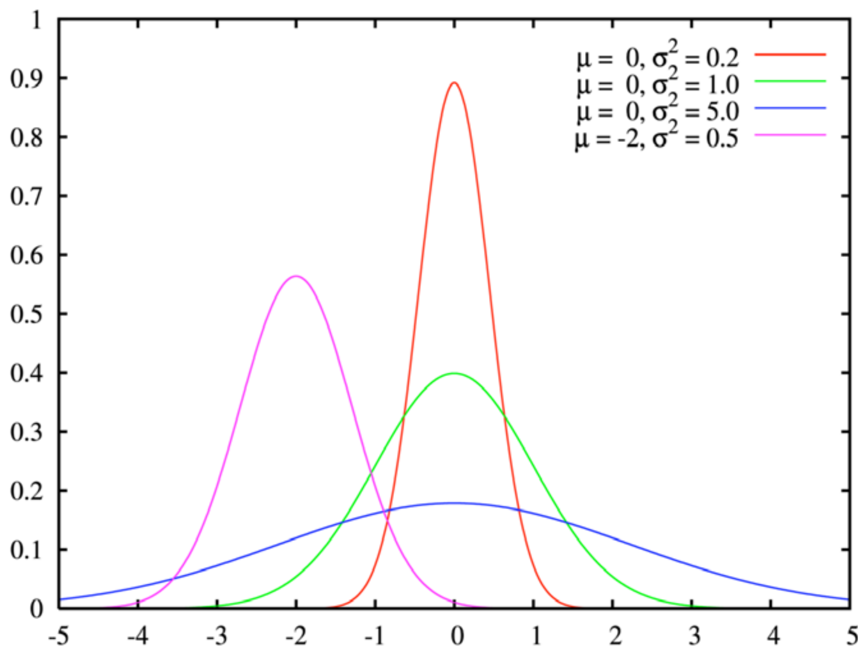
Probability Density

- Continuous variable $\Rightarrow$ Probability density
  - Typically, $P(X = x) = 0$ for all particular x
  - $P(X \in [x, y]) = \int_x^y f(a)da$ for some function $f(a)$
    - f is called the *probability density function*
- Example:
  - Uniform in [0,1] $\Rightarrow$ $f(a) = 1$ for $a \in [0, 1]$, else 0

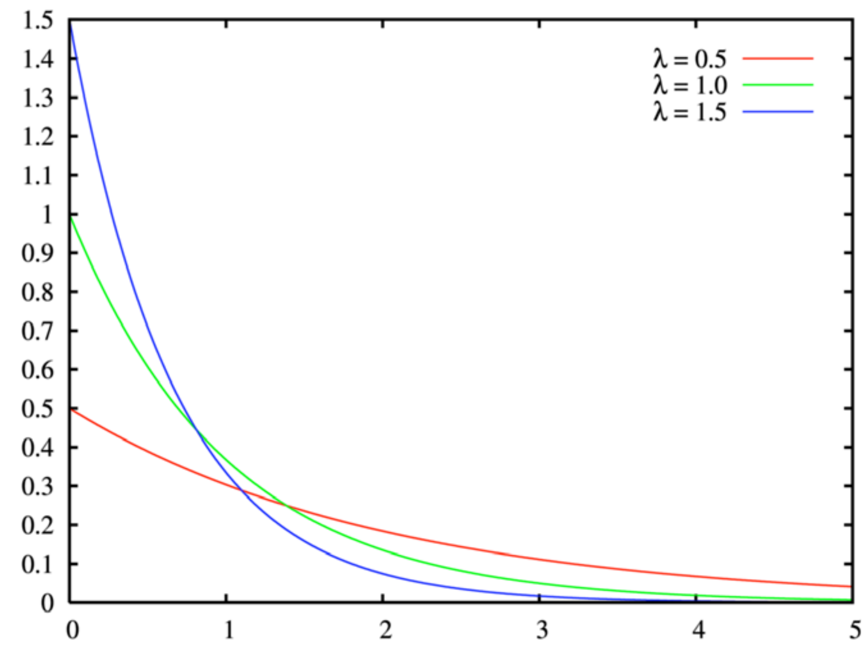
Probability Density

- Two common probability densities

Gaussian / Normal



Exponential



Interpretations of Probability

- Frequentist
  - Probability is frequency in a (hypothetical) infinite pop.
  - But what is the relevant population?
- Propensity
  - Probability captures underlying randomness and symmetry in the world
  - Nice in theory, tough in practice
- Subjectivist
  - Probability statements are really claims about our beliefs about the world
  - But doesn’t seem to fit intuitions about basic cases

Conditional Probability

- (Or conditional frequency)
- $P(X = x \mid Y = y)$ is just $P(X = x)$ in the sub-population of individuals with Y = y
  - Conditional probabilities are still probabilities

Conditional Probability

- (Or conditional frequency)
- $P(X = x \mid Y = y)$ is just $P(X = x)$ in the sub-population of individuals with Y = y
  - Conditional probabilities are still probabilities
- Mathematically, $P(X = x \mid Y = y) = \dfrac{P(X = x \,\&\, Y = y)}{P(Y = y)}$

Conditional Probability

- Except in special cases,

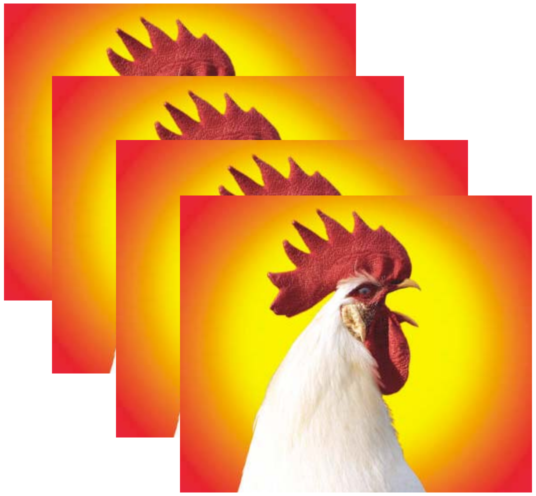
  $$P(X = x \mid Y = y) \neq P(Y = y \mid X = x)$$

  – Specifically, they're equal iff

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

  – Simple examples of the inequality:
    - P(Pregnant | Female) ≠ P(Female | Pregnant)
    - P(Over 6' | Male) ≠ P(Male | Over 6')

Conditional Probability



- {Rooster, Apple, Frog}

  - P(P = F | P = F or R) = 3/5

  - P(P = F or R | P = F) = 1

Independence

- *X* and *Y* are independent iff learning the value of *X* provides no information about the value of *Y*
  - In particular, learning about one does not change the predictability of the other
- *X* and *Y* are associated iff they are not independent
- Independence/association are symmetric!

Independence

- Formally, *X* and *Y* are independent iff
  - For all x, y, $P(X = x) = P(X = x \mid Y = y)$
  - For all x, y, $P(X=x \,\&\, Y=y) = P(X=x) \times P(Y=y)$
  - For all x, y_1, y_2, $P(X=x \mid Y=y_1) = P(X=x \mid Y=y_2)$
    - For “typical” distributions, these are all equivalent
    - For continuous variables, only “almost all x, y”

- *X* and *Y* are associated iff
  - There exists x, y such that…

Independence

- For conditional independence, must be independent in every (relevant) population
  - *X* independent of *Y* given *Z* iff
  - For all *x*, *y*, *z*,
    $P(X = x \mid Z = z) = P(X = x \mid Y = y \;\&\; Z = z)$
  - and similarly for the other definitions…
- Note: *Z* can be more than one variable!

- Conditional association is analogous

Independence

- Simple notation:
  - X independent of Y $\Rightarrow$ $X \perp\!\!\!\perp Y$
  - X independent of Y given Z $\Rightarrow$ $X \perp\!\!\!\perp Y \mid Z$
  - X associated with Y $\Rightarrow$ $X \not\perp\!\!\!\perp Y$
  - X associated with Y given Z $\Rightarrow$ $X \not\perp\!\!\!\perp Y \mid Z$

Independence

- And all of these notions extend to sets
  - Set $\boldsymbol{X}$ is independent of set $\boldsymbol{Y}$ given $\boldsymbol{Z}$ iff For all $X \in \boldsymbol{X}$, $Y \in \boldsymbol{Y}$, $X \perp\!\!\!\perp Y \mid \boldsymbol{Z}$
    - I.e., for every possible setting of variables in $\boldsymbol{Z}$, each X is independent of each Y
  - Note: Might require a large number of tests!

Bayes (and Bayesianism)

- Bayes' Theorem: $P(T \mid D) = \dfrac{P(D \mid T)P(T)}{P(D)}$
  - proof is trivial…
- General strategy:
  - Let *D* be the data and *T* be the theory
  - $\Rightarrow$ Bayes' theorem says how to update beliefs about the probability of various theories

Bayes (and Bayesianism)

- Terminology:

$$P(T \mid D) = \frac{P(D \mid T)P(T)}{P(D)}$$

Bayes (and Bayesianism)

- Terminology:

$$P(T \mid D) = \frac{P(D \mid T)\,P(T)}{P(D)}$$

Prior distribution

Bayes (and Bayesianism)

- Terminology:

Likelihood function → $P(D \mid T)$

$$P(T \mid D) = \frac{P(D \mid T)\,P(T)}{P(D)}$$

Bayes (and Bayesianism)

- Terminology:

Posterior distribution

$$\boxed{P(T \mid D)} = \frac{P(D \mid T)P(T)}{P(D)}$$

Bayes (and Bayesianism)

- Terminology:

$$P(T \mid D) = \frac{P(D \mid T)\,P(T)}{\boxed{P(D)}}$$

Data distribution

Bayes (and Bayesianism)

- Terminology:

$$P(T \mid D) = \frac{P(D \mid T)\,P(T)}{P(D)}$$

Posterior distribution — $P(T \mid D)$

Likelihood function — $P(D \mid T)$

Prior distribution — $P(T)$

Data distribution — $P(D)$

Linear Models

- For continuous variables, the relationship between two variables is often expressed in terms of a linear model:

$$X = aY + \varepsilon_X$$

  - *a* is some real-valued coefficient
  - ε_X is a "noise" term

  - Obvious multi-variate generalization…
  - Use regression to find the best-fitting models

Nature of Causation

- Token causal claims: Claims about causation between particular tokens, not populations
  - Event A caused event B
    - “This light switch flip caused the lights to turn on”
  - Having property A caused X to have property B
    - “The glass broke because it was brittle”
  - Thing 1 having property A caused Thing 2 to have property B
    - “I went to comfort my daughter because she was crying.”

Nature of Causation

- Type causal claims: About causation that occurs “in general”, or “in the population”
  - Events of type A cause events of type B
    - “Light switches turn on lights”
  - Having property A causes things of type X to have property B
    - “Some glasses break because they are brittle”
  - Thing 1 having property A caused Thing 2 to have property B
    - “Parents frequently go to comfort their children when the children cry.”

Nature of Causation

- Metaphysical primacy is ambiguous
  - Token-causation is primary, and type-level claims hold because of causation in the individuals

  OR

  - Type-causation is primary, and token-level claims (sometimes) hold for members of that population

- In general, we will focus on type causal claims
  - Token causal inference is much harder, and frequently dependent on type-causal prior knowledge

Nature of Causation

- What *is* causation? Is it based on:
  - Possible worlds?
  - Conservation of physical quantities?
  - Hypothetical experiments?
  - Purely pragmatic conventions?

- We'll talk more tomorrow afternoon…

Association vs. Causation

“*Association is symmetric;*
Causation is asymmetric”

- X associated w/ $Y \Rightarrow Y$ associated w/ X
  X causes $Y \not\Rightarrow Y$ causes X
  - In fact, for token-causation, we think we have:
  - X causes $Y \Rightarrow Y$ does not cause X

Association *and* Causation

- Although different, they are connected
  - In general,
    - If *X* causes *Y*, then *X* will be associated with *Y*
    - If *X* and *Y* are associated, then there is some sort of causal connection between them
  - Statistics is relevant to science precisely because the two are connected
  - Causal inference is really the problem of moving between these two types of claims

Causation and Intervention

- Causal claims support counterfactuals
  - In particular, those about interventions
    - “If I had flipped the switch, the light would have turned on.”
    - “If she hadn’t dropped the plate, then it wouldn’t have broken.”
    - Etc.

Causation and Intervention

- One of the central causal asymmetries
  - Interventions on cause lead to changes in the effect
    - Flipping the switch turns off the light
  - Interventions on the effect do not lead to changes in the cause
    - Breaking the light bulb doesn’t flip the switch

- Some have argued that this is the paradigmatic feature of causation (Woodward, Hausman)

Observation vs. Intervention

- Association vs. Causation distinction maps onto Observation vs. Intervention
  - Symmetry of observation: If observing *X* tells us about *Y*, then observing *Y* tells us about *X*
  - Asymmetry of intervention: If intervening on *X* affects *Y*, then intervening on *Y* might not affect *X*

- So which is better for learning the true causal structure? It depends…

Observation vs. Intervention

- Observations give you information about the causal structure in its "natural state"
  - Benefit: All of the causal relationships are (potentially) active, and so you can (i) try to learn the full causal structure; and (ii) draw more inferences from partial observations
  - Drawback: Hard (but not impossible!) to determine the directionality of a particular causal connection (e.g., does *X* cause *Y*? Does *Y* cause *X*? Is there a common cause?)

Observation vs. Intervention

- Manipulations give you information about an altered causal structure
  - In particular, the “normal” causes of the manipulated variable are no longer causes
  - <u>Benefit:</u> More information (locally) about causal structure; in particular, directionality is often clear
  - <u>Drawback:</u> Information is not about the full “normal” causal structure; also, manipulations can be costly on many different dimensions