Inference and Representation

Rachel Hodos

New York University

Lab 1, September 2, 2015

Welcome!

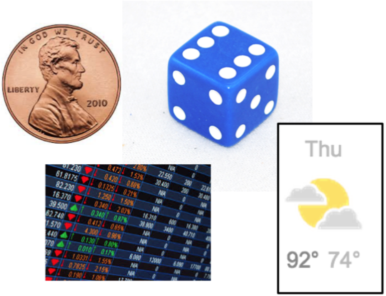
- Instructor: David Sontag, dsontag@cs.nyu.edu
- Lab instructor: Rachel Hodos, hodos@cims.nyu.edu
- Grader: Prasoon Goyal, pgoyal@nyu.edu
- Lecture: Tue 5:10-7pm WWH 102
- Lab: Wed 7:10-8pm WWH 102
- Attendance required for both lab and lecture (may be taken periodically)
- Lab material will be complementary to main lectures
- Slides and additional info posted on website (TBD)
- Python / pymc3

About Me

- 5th year PhD student in Computational Biology
- Based in math department
- Research: computational drug discovery
- Will draw many examples from computational biology

Random Variables

- Wikipedia: A variable whose values are subject to variations due to chance.
- Examples:



- Discrete or continuous

Probability Distributions

- Every random variable has some probability distribution (may be unknown)
- Describes the likelihood that a random variable will take a certain value (or set of values)
- Discrete variables: probability mass function (pmf):

$$\sum_x p(x) = 1$$

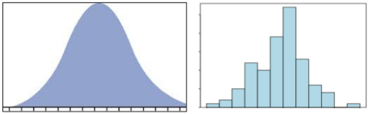
- Continuous variables: probability density functions (pdf):

$$\int_x p(x)dx = 1$$

- In both cases, $p(x) >= 0$ for all x.

Examples of Probability Distributions

- Discrete:
  - Binomial, Bin(p), coin toss
  - Multinomial, Multi($\theta_1, \theta_2, \ldots, \theta_k$), dice roll
  - Poisson, Poisson(λ), number of events in a given time window, where each event occurs with average rate λ
- Continuous:
  - Uniform, e.g. $U[a, b]$
  - Gaussian or Normal, $N(\mu, \sigma^2)$, white noise
    - "Standard normal" = $N(0, 1)$
  - Exponential, exp(λ), waiting times
- Theoretical vs. empirical distributions



Expectation

- Also called mean, average, first moment
- Defined as

$$\mathbf{E}(X) = \sum_x x p(x)$$

- Continuous:

$$\int_x x p(x) dx$$

- Expectation of empirical distribution is what we're used to:

$$\frac{1}{N} \sum_{i=1}^{N} x_i$$

- Linearity of expectation:

$$\mathbf{E}(aX + b) = a\mathbf{E}(X) + b$$

Multivariate Distributions

- Often want to understand relationships between variables
- Example: X = "sprinkler on", Y = "raining"

p(X,Y)	sprinkler off	sprinkler on
not raining	.6	.24
raining	.15	.01

- This is an example of a *multivariate* or *joint distribution*, i.e. $p(X, Y)$ or more generally $p(X_1, X_2, \ldots, X_n)$
- For discrete variables, simply defined as

$$p(X_1 = x_1, \ldots, X_n = x_n) :=$$
$$p(X_1 = x_1 \cap \ldots \cap X_n = x_n)$$

- Examples:
  - Discrete: series of coin tosses, rolling a pair of dice
  - Continuous: Multivariate Gaussian, Dirichlet

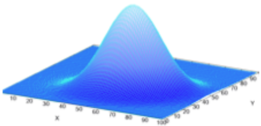
The Multivariate Gaussian Distribution

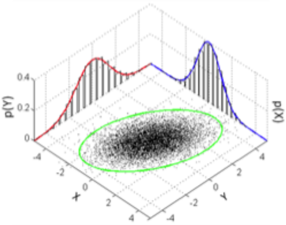
- One variable:

$$p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- Multiple variables:

$$p(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{n/2}\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x-\boldsymbol{\mu})^T\Sigma^{-1}(x-\boldsymbol{\mu})\right)$$





Marginalization

- Suppose we have joint distribution $p(X_1, \ldots, X_n)$ and we want to know $p(X_i = x_i)$
- Example: from p(X,Y), what is $p(Y = \text{raining})$?

p(X,Y)	sprinkler off	sprinkler on
not raining	.6	.24
raining	.15	.01

- Answer: .15 + .01 = .16
- More generally:

$$p(X_i = x_i) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \sum_{x_n} p(x_1, \ldots, x_n)$$

- Doing this for each x_i gives us the *marginal distribution* over X_i

Conditioning

- What if we want to know the distribution of X when Y is set to a particular value?
- Example: from p(X,Y), what is *p*(sprinkler on | raining)?

p(X,Y)	sprinkler off	sprinkler on
not raining	.6	.24
raining	.15	.01

- Answer: .01 / (.15 + .01) = 1/16
- More generally:

$$p(x|y) = \frac{p(x,y)}{p(y)}$$

- Note that this is a univariate distribution, defined for each value of y

Independence

- Intuitively: two variables are independent if they are unrelated
- One definition: $X \perp Y$ means that for all x and y, $p(x, y) = p(x)p(y)$
- Examples: a sequence of coin tosses
- Why do we care?
  - Simple representation of joint distribution
  - Tells you where not to look when building predictive models
  - But, if everything were independent, we couldn't make any predictions, and machine learning wouldn't exist...

Two Important Rules

- **Chain Rule:**

$$p(x_1, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \ldots p(x_n|x_1, x_2, \ldots, x_{n-1})$$

  - Intuition: generalizes conditioning on a single variable

- **Bayes' Rule**:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

  - Basis of Bayesian statistics
  - Useful when $p(y|x)$ is more natural to measure or estimate than $p(x|y)$
  - Proof: from chain rule and definition of conditional distribution

Questions?

- Probability review: http://cs229.stanford.edu/section/cs229-prob.pdf