Announcements

- **Reading Report #2**
  - Due today
  - Your questions!

- **Assignment #2**
  - Early: Wednesday
  - Due: Friday

---

**Objectives**

- Understand 4 important discrete distributions
- Describe uncertain worlds with joint probability distributions

---

**Bernoulli Distribution**

- 2 possible outcomes
  
  \[
  p(x) = B(x; p) = \begin{cases} 
  p & \text{when } x = 1 \\
  1 - p & \text{when } x = 0 \\
  0 & \text{otherwise}
  \end{cases}
  \]

  “What’s the probability of a single binary event \( x \), if a ‘positive’ event has probability \( p \)?”

---

**Bernoulli Distribution**

- 2 possible outcomes
  
  \[
  p(x) = B(x; p) = \begin{cases} 
  p^x (1 - p)^{1-x} & \text{when } x \in [0, 1] \\
  0 & \text{otherwise}
  \end{cases}
  \]

  “What’s the probability of a single binary event \( x \), if a ‘positive’ event has probability \( p \)?”

---
Categorical Distribution

- Extension for $m$ possible outcomes

What’s the probability of a single event $x$ (containing a 1 in only one position), if outcomes 1, 2, ..., and $m$, are specified by $g = [p_1, p_2, ..., p_m]$?

- Note: $\sum_{i=1}^{m} p_i$ must be 1

Great for language models, where each value corresponds to a word or an $n$-gram of words. (e.g., value '1' corresponds to 'the')

1-of-$m$ Encoding

[0, 1, 0] -> [1, 0, 0]

[0, 0, 1] -> [0, 1, 0]

Categorical Distribution

- Extension for $m$ possible outcomes

\[
P(x) = C(x; p) = \begin{cases} p_i & \text{when } x \in \{1, ..., m\} \\ 0 & \text{otherwise} \end{cases}
\]

\[
P(x) = C(x; p) = \begin{cases} \prod_{i=1}^{m} p_i & \text{when } x_i \in \{0, 1\} \text{ and exactly one } x_i = 1 \\ 0 & \text{otherwise} \end{cases}
\]

Equivalently:

\[
P(x) = C(x; p) = \begin{cases} p_i & \text{when } x \in \{1, ..., m\} \\ 0 & \text{otherwise} \end{cases}
\]

- Great for language models, where each value corresponds to a word or an $n$-gram of words. (e.g., value '1' corresponds to 'the')

when $\{0, 1\}$ and exactly one 1
Binomial Distribution

- 2 possible outcomes; N trials

\[ P(x) = \binom{N}{x} p^x (1-p)^{N-x} \]

“What’s the probability in N independent Bernoulli events that x of them will come up ‘positive’, if a ‘positive’ event has probability p?”

Multinomial Distribution

- Extension for m possible outcomes; N trials

\[ P(\mathbf{x}) = \text{Mult}(\mathbf{x}; N, \mathbf{p}) = \frac{N!}{x_1! \cdots x_m!} p_1^{x_1} \cdots p_m^{x_m} \]

when \( \sum x_i = N \)

otherwise

“What’s the probability in N independent categorical events that value 1 will occur \( x_1 \) times...and that value m will occur \( x_m \) times, if the probabilities of each possible value are specified by \( \mathbf{p} = [p_1, p_2, \ldots, p_m] \)?

* Note: \( p_i \) must sum to 1

Normal Distribution

\[ n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

e.g., for modeling the length of a sentence?
Parametric Distributions

- The shape of the distribution is described by a few parameters.

Acknowledgments

Note to other teachers and users of the following slides: Andrew Moore would be delighted if you found this source material useful in giving your own lectures. Feel free to use these slides verbatim, or to modify them to fit your own needs. PowerPoint originals are available. If you make use of a significant portion of these slides in your own lecture, please include this message, or the following link to the source repository of Andrew’s tutorials: http://www.cs.cmu.edu/~awm/tutorials. Comments and corrections gratefully received.

For slides 20-24
Next

- Answer queries on joint distributions
- Address our efficiency problem by making independence assumptions!
- Use the Bayes Net methodology to build joint distributions in manageable chunks