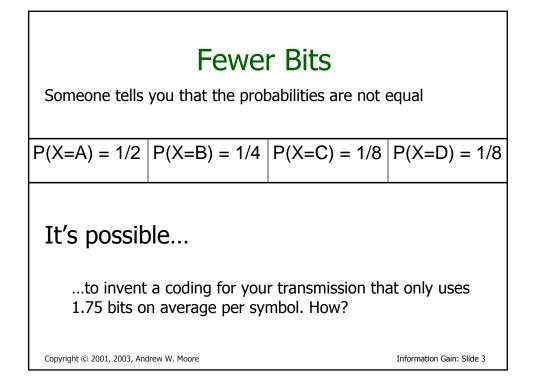
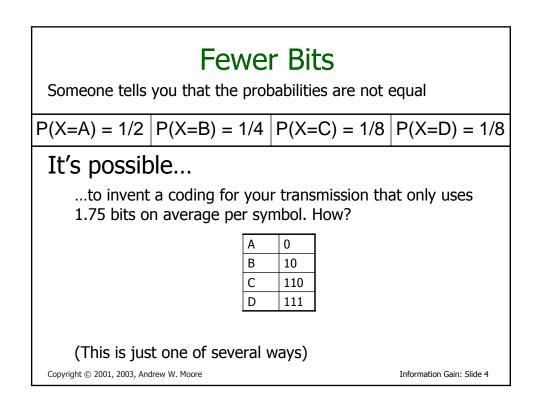
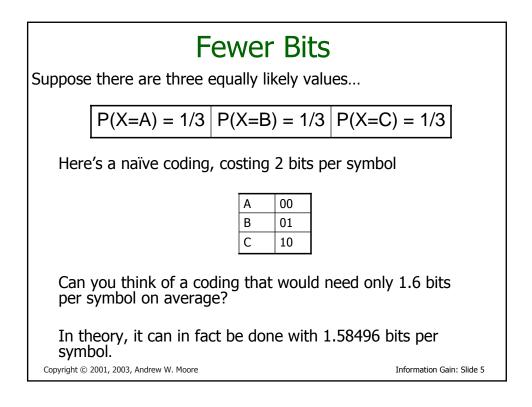
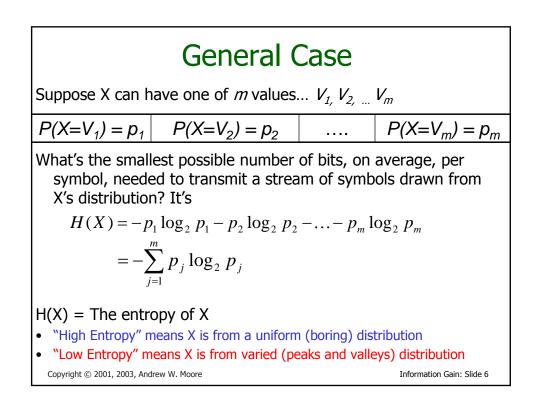


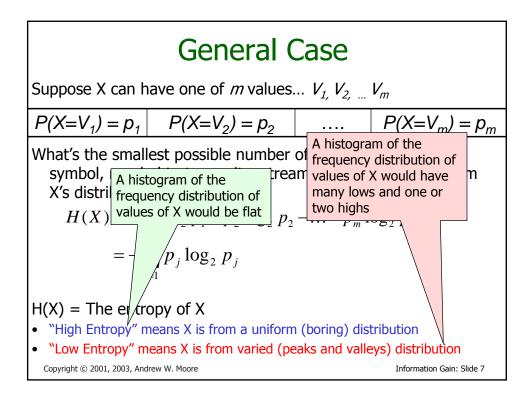
| Bits  |              |              |              |  |  |  |  |
|---|--------------|--------------|--------------|--|--|--|--|
| You are watching a set of independent random samples of X   |              |              |              |  |  |  |  |
| You see that X has four possible values   |              |              |              |  |  |  |  |
| P(X=A) = 1/4  | P(X=B) = 1/4 | P(X=C) = 1/4 | P(X=D) = 1/4 |  |  |  |  |
| So you might see: BAACBADCDADDDA<br>You transmit data over a binary serial link. You can encode<br>each reading with two bits (e.g. A = 00, B = 01, C = 10, D =<br>11)<br>0100001001001110110011111100<br>Copyright © 2001, 2003, Andrew W. Moore |              |              |              |  |  |  |  |

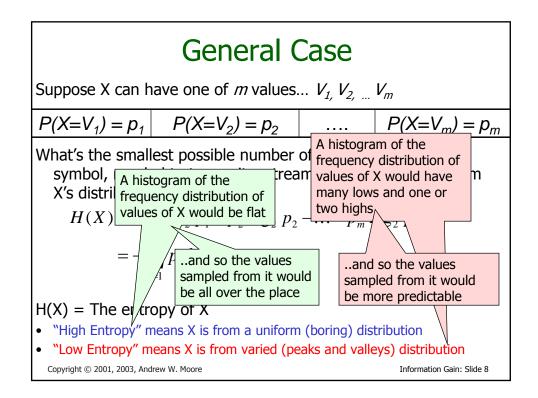












## Entropy in a nut-shell



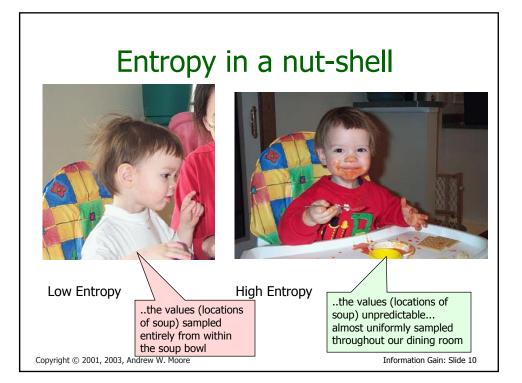
Low Entropy

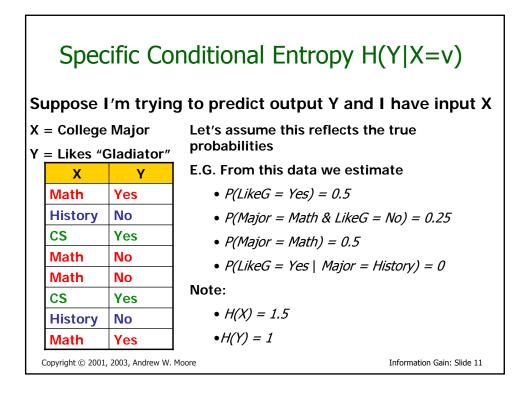


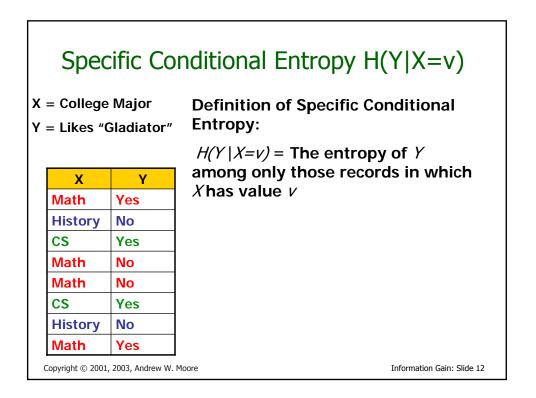
Information Gain: Slide 9

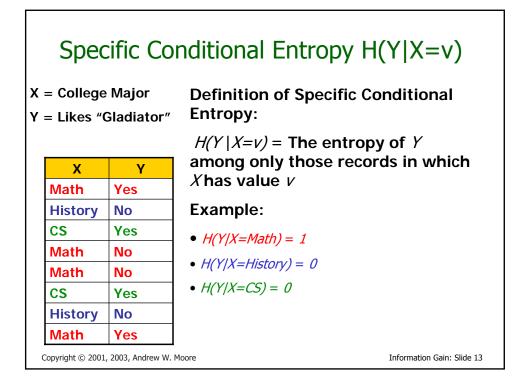
High Entropy

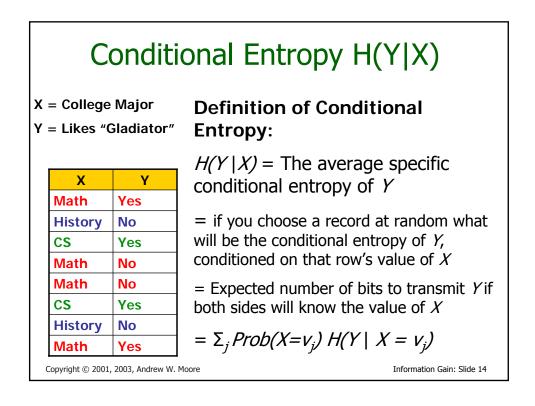
Copyright © 2001, 2003, Andrew W. Moore





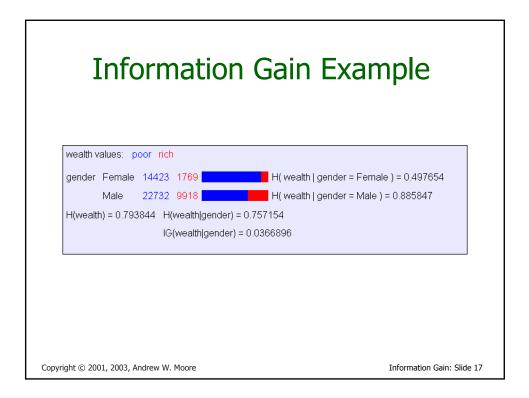


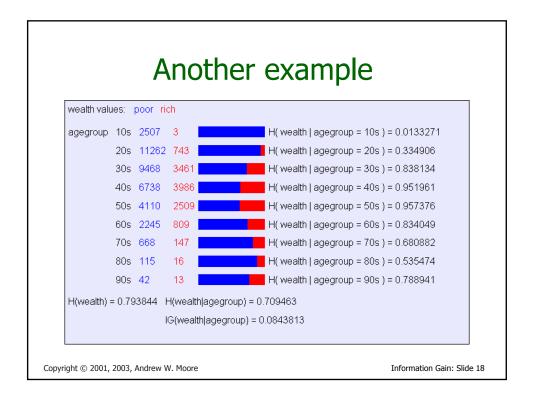




| Conditional EntropyX = College MajorDefinition of Conditional Entropy:Y = Likes "Gladiator" $H(Y X) =$ The average conditional<br>entropy of Y |  |     |  |         |               |                     |   |
|--|--|-----|--|---------|---------------|---------------------|---|
| $= \sum_{j} Prob(X = v_j) H(Y \mid X = v_j)$   |  |     |  |         |               |                     |   |
|  | X  | Y   |  | Example |               |                     |   |
|  | Math   | Yes |  |         |               |                     |   |
|  | History  | No  |  | Vi      | $Prob(X=v_i)$ | $H(Y \mid X = V_i)$ |   |
|  | CS   | Yes |  | Math    | 0.5           | 1                   | 1 |
|  | Math   | No  |  |         |               | 1                   |   |
|  | Math   | No  |  | History | 0.25          | 0                   |   |
|  | CS   | Yes |  | CS      | 0.25          | 0                   |   |
|  | History  | No  |  |         | -             | I                   | 1 |
|  | Math Yes $H(Y X) = 0.5 * 1 + 0.25 * 0 + 0.25 * 0 = 0.5$            |     |  |         |               |                     |   |
| C  | Copyright © 2001, 2003, Andrew W. Moore Information Gain: Slide 15 |     |  |         |               |                     |   |

| 0-11   | N/a:au |  |
|--|--------|--|
| <pre>&lt; = College Major</pre>  |        | Definition of Information Gain:                    |
| IG(Y   X) = I  must transmit  Y.<br>How many bits on average<br>would it save me if both ends of |        |  |
| Х  | Y      | the line knew $\lambda$ ?                          |
| Math   | Yes    | $IG(Y \mid X) = H(Y) - H(Y \mid X)$                |
| History  | No     |  |
| CS   | Yes    | Example:   |
| Math   | No     | • H(Y) = 1   |
| Math   | No     |  |
| CS   | Yes    | • $H(Y   X) = 0.5$                                 |
| History  | No     | <ul> <li>Thus IG(Y   X) = 1 – 0.5 = 0.5</li> </ul> |
| Math   | Yes    |  |





| <b>Relativ</b><br>X = College Major<br>Y = Likes "Gladiator" |                    | /e Information Gain<br>Definition of Relative Information<br>Gain:                               |  |
|--|--------------------|--|--|
|  |                    | <i>RIG(Y</i>   <i>X)</i> = I must transmit <i>Y</i> , what fraction of the bits on average would |  |
| Х  | Y                  | it save me if both ends of the line  |  |
| Math   | Yes                | knew X?  |  |
| History  | No                 | RIG(Y X) = H(Y) - H(Y X) / H(Y)  |  |
| CS   | Yes                |  |  |
| Math   | No                 | Example:   |  |
| Math   | No                 | • $H(Y X) = 0.5$   |  |
| CS   | Yes                |  |  |
| History  | No                 | • $H(Y) = 1$   |  |
| Math   | Yes                | • Thus $IG(Y X) = (1 - 0.5)/1 = 0.5$   |  |
| Copyright © 200  | 1, 2003, Andrew W. | Moore Information Gain: Slide 19   |  |

