Entropy and Uncertainty

Appendix C
Outline

• Random variables
• Joint probability
• Conditional probability
• Entropy (or uncertainty in bits)
• Joint entropy
• Conditional entropy
• Applying it to secrecy of ciphers
Random Variable

• Variable that represents outcome of an event
  • $X$ represents value from roll of a fair die; probability for rolling $n$: $p(=n) = 1/6$
  • If die is loaded so 2 appears twice as often as other numbers, $p(X=2) = 2/7$ and, for $n \neq 2$, $p(X=n) = 1/7$

• Note: $p(X)$ means specific value for $X$ doesn’t matter
  • Example: all values of $X$ are equiprobable
Joint Probability

• Joint probability of $X$ and $Y$, $p(X, Y)$, is probability that $X$ and $Y$ simultaneously assume particular values
  • If $X$, $Y$ independent, $p(X, Y) = p(X)p(Y)$

• Roll die, toss coin
  • $p(X=3, Y=\text{heads}) = p(X=3)p(Y=\text{heads}) = 1/6 \times 1/2 = 1/12$
Two Dependent Events

• $X = \text{roll of red die, } Y = \text{sum of red, blue die rolls}$

\[
p(Y=2) = \frac{1}{36} \quad p(Y=3) = \frac{2}{36} \quad p(Y=4) = \frac{3}{36} \quad p(Y=5) = \frac{4}{36} \\
p(Y=6) = \frac{5}{36} \quad p(Y=7) = \frac{6}{36} \quad p(Y=8) = \frac{5}{36} \quad p(Y=9) = \frac{4}{36} \\
p(Y=10) = \frac{3}{36} \quad p(Y=11) = \frac{2}{36} \quad p(Y=12) = \frac{1}{36}
\]

• Formula:

\[
p(X=1, Y=11) = p(X=1)p(Y=11) = \left(\frac{1}{6}\right)\left(\frac{2}{36}\right) = \frac{1}{108}
\]
Conditional Probability

- Conditional probability of $X$ given $Y$, $p(X \mid Y)$, is probability that $X$ takes on a particular value given $Y$ has a particular value.

- Continuing example ...
  - $p(Y=7 \mid X=1) = 1/6$
  - $p(Y=7 \mid X=3) = 1/6$
Relationship

- \( p(X, Y) = p(X \mid Y) \ p(Y) = p(X) \ p(Y \mid X) \)

- Example:
  \[ p(X=3, Y=8) = p(X=3 \mid Y=8) \ p(Y=8) = (1/5)(5/36) = 1/36 \]

- Note: if \( X, Y \) independent:
  \[ p(X \mid Y) = p(X) \]
Entropy

• Uncertainty of a value, as measured in bits
• Example: \( X \) value of fair coin toss; \( X \) could be heads or tails, so 1 bit of uncertainty
  • Therefore entropy of \( X \) is \( H(X) = 1 \)
• Formal definition: random variable \( X \), values \( x_1, \ldots, x_n \); so \( \sum_i p(X = x_i) = 1 \); then entropy is:

\[
H(X) = -\sum_i p(X = x_i) \log p(X = x_i)
\]
Heads or Tails?

• \( H(X) = - p(X=\text{heads}) \lg p(X=\text{heads}) - p(X=\text{tails}) \lg p(X=\text{tails}) \)
  
  \[
  = - (1/2) \lg (1/2) - (1/2) \lg (1/2) \\
  = - (1/2) (-1) - (1/2) (-1) = 1
  \]

• Confirms previous intuitive result
$n$-Sided Fair Die

\[ H(X) = -\sum_i p(X = x_i) \lg p(X = x_i) \]

As \( p(X = x_i) = 1/n \), this becomes

\[ H(X) = -\sum_i (1/n) \lg (1/n) = -n(1/n) (-\lg n) \]

so

\[ H(X) = \lg n \]

which is the number of bits in \( n \), as expected
Ann, Pam, and Paul

Ann, Pam twice as likely to win as Paul

$W$ represents the winner. What is its entropy?

- $w_1 = \text{Ann}, w_2 = \text{Pam}, w_3 = \text{Paul}$
- $p(W=w_1) = p(W=w_2) = 2/5, p(W=w_3) = 1/5$
- So $H(W) = -\sum_i p(W=w_i) \lg p(W=w_i)$
  $\quad = - (2/5) \lg (2/5) - (2/5) \lg (2/5) - (1/5) \lg (1/5)$
  $\quad = - (4/5) + \lg 5 \approx -1.52$
- If all equally likely to win, $H(W) = \lg 3 \approx 1.58$
Joint Entropy

• $X$ takes values from $\{ x_1, \ldots, x_n \}$, and $\sum_i p(X=x_i) = 1$
• $Y$ takes values from $\{ y_1, \ldots, y_m \}$, and $\sum_i p(Y=y_i) = 1$
• Joint entropy of $X$, $Y$ is:

$$H(X, Y) = -\sum_j \sum_i p(X=x_i, Y=y_j) \log p(X=x_i, Y=y_j)$$
Example

X: roll of fair die, Y: flip of coin

As X, Y are independent:

\[ p(X=1, Y=\text{heads}) = p(X=1) \cdot p(Y=\text{heads}) = \frac{1}{12} \]

and

\[ H(X, Y) = -\sum_j \sum_i p(X=x_i, Y=y_j) \log p(X=x_i, Y=y_j) \]
\[ = -2 \left[ 6 \left[ \frac{1}{12} \log \frac{1}{12} \right] \right] = \log 12 \]
Conditional Entropy

• $X$ takes values from \{ $x_1, \ldots, x_n$ \} and $\sum_i p(X=x_i) = 1$

• $Y$ takes values from \{ $y_1, \ldots, y_m$ \} and $\sum_i p(Y=y_i) = 1$

• Conditional entropy of $X$ given $Y=y_j$ is:

$$H(X \mid Y=y_j) = -\sum_i p(X=x_i \mid Y=y_j) \lg p(X=x_i \mid Y=y_j)$$

• Conditional entropy of $X$ given $Y$ is:

$$H(X \mid Y) = -\sum_j p(Y=y_j) \sum_i p(X=x_i \mid Y=y_j) \lg p(X=x_i \mid Y=y_j)$$
Example

• $X$ roll of red die, $Y$ sum of red, blue roll

• Note $p(X=1 | Y=2) = 1$, $p(X=i | Y=2) = 0$ for $i \neq 1$
  • If the sum of the rolls is 2, both dice were 1

• Thus

$$H(X | Y=2) = -\sum_i p(X=x_i | Y=2) \log p(X=x_i | Y=2) = 0$$
Example (con’t)

• Note $p(X=i, Y=7) = 1/6$
  - If the sum of the rolls is 7, the red die can be any of 1, ..., 6 and the blue die must be 7—roll of red die

• $H(X|Y=7) = -\sum_i p(X=x_i|Y=7) \log p(X=x_i|Y=7)$
  $= -6 \frac{1}{6} \log \frac{1}{6} = \log 6$
Perfect Secrecy

• Cryptography: knowing the ciphertext does not decrease the uncertainty of the plaintext
• $M = \{ m_1, ..., m_n \}$ set of messages
• $C = \{ c_1, ..., c_n \}$ set of messages
• Cipher $c_i = E(m_i)$ achieves perfect secrecy if $H(M \mid C) = H(M)$