Statistics of models

David Bernick
Entropy

\[ encodingCost_M = -\sum_{L} \log_2(P_M(x_i)) \]

\[ H(X) = -\sum_i P(x_i) \log P(x_i) \]

- Entropy maximized when \( P(x) \) is equal (minimized) and \( H(X) = 0 \) when one quantity (of \( X \)) is certain.
Relative Entropy

\[ H(P \parallel Q) = \sum_i P(x_i) \log \frac{P(x_i)}{Q(x_i)} \]

- For 2 distributions, their “distance” in bits can be estimated with *relative entropy* *(aka Kulback-Liebler distance)*
- \( H(P \parallel Q) \) is always positive
- In cases where we compare 2 models M and R, using log odds ratios, \( H(P \parallel Q) \) is the expected value
P-value

• Probability of obtaining a result at least as extreme as the one observed
E-value

• A count of the number of times we would expect to see a value this extreme in multiple tests.
Continuous Distributions

- Probability density function (pdf) normal distribution
Continuous Distributions-2

- Cumulative Density Function (CDF)

$X \sim N(\mu, \sigma)$

$mean = \mu$

$variance = \sigma^2$

\[
\Phi_{\mu,\sigma^2}(x) = \begin{cases} 
\frac{1}{2} & \text{if } \mu = 0, \sigma^2 = 0.2, \\
\frac{1}{2} & \text{if } \mu = 0, \sigma^2 = 1.0, \\
\frac{1}{2} & \text{if } \mu = 0, \sigma^2 = 5.0, \\
\frac{1}{2} & \text{if } \mu = -2, \sigma^2 = 0.5, 
\end{cases}
\]
Discrete distributions - 2

- Cumulative Distribution Function (CDF) of Binomial \( B(n, p) \)
Discrete Distributions

- Probability Mass Function of a Binomial distribution $B(n,p)$

$K \sim B(n,p)$

$n = \# \text{trials}, n \in N_0$

$p = P_{\text{success}}$

$\text{mean} = np$

$\text{var} = np(1 - p)$

$$P(K = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$
Central Limit Theorem (CLT)

- Large numbers of randomly drawn samples from a distribution with finite mean and variance will have an approximately normal distribution.

- This allows us to fit one distribution to another if we have a sufficient number of i.i.d. samples
  - For example: a binomial distribution can be fit to a normal distribution and calculate p-values as though the distribution were Normal.
Class Exercise

• We see 2 coins, one of which is loaded (heavy on the tails side, the other is known to be fair. (What is our prior belief? )
• We pick one of the coins and spin it 100 times, and find that we get 10 heads.

• With what probability do we now believe that this is a fair coin?
• How many times would we expect to see a result this extreme?
Class Exercise

• It’s annoying to calculate 100!, but we can easily approximate mean and variance

\[ \mu = np = 100(0.5) = 50 \]
\[ \text{var} = np(1 - p) = 100(0.5)(0.5) = 25 \]
\[ \sigma = \sqrt{25} = 5 \]
\[ Z = \frac{10 - \mu}{5} = -8 \]
\[ P(Z \leq -8) \approx 6.221 \times 10^{-16} \]
\[ Evalue = 100 \times Pvalue \approx 6.221 \times 10^{-14} \]