Bio5488 Midterm Equations

Equations

Poisson Distribution
\[ P(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \]

where
- \( e \) is the base of the natural logarithm (\( e = 2.71828... \))
- \( x \) is the number of occurrences of an event - the probability of which is given by the function
- \( x! \) is the factorial of \( x \)
- \( \lambda \) is a positive real number, equal to the expected number of occurrences that occur during the given interval

Binomial Distribution
\[ P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \]

where
\[ \binom{n}{k} = \frac{n!}{k! (n-k)!} \]
- \( n \) = number of trials
- \( k \) = number of successes
- \( p \) = probability of success

Gaussian or Normal Distribution
\[ P(x, \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

where
- \( \mu \) = mean
- \( \sigma \) = standard deviation

Equation for calculating the log odds score for entries in a BLOSUM Matrix
\[ score = \log_2 \left( \frac{P(\text{residues align} \mid \text{homology model})}{P(\text{residues align} \mid \text{random model})} \right) \]

Pearson Correlation Coefficient
\[ r = 1 - \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right) \]

where
- \( n \) = number of conditions
- \( \bar{x} \) = average expression of gene \( x \) in all \( n \) conditions
- \( \bar{y} \) = average expression of gene \( y \) in all \( n \) conditions
- \( s_x \) = sample standard deviation of \( x \)
- \( s_y \) = sample standard deviation of \( y \)
Sample standard deviation

\[ s_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}} \]

where
- \( n \) = number of data points
- \( \bar{x} \) = average of the \( x_i \)
- \( x_i \) = each of the values of the data

Extreme value distribution

\[ P(S \geq x) = 1 - e^{-\lambda(x-\mu)} \]

Bonferroni Correction

\[ \alpha = \frac{\alpha}{g} \]

False Discovery Rate

\[ n_{DE} \ast \alpha \]

Information content

\[ I_{\text{seq}} = \sum_{j} \sum_{b} f(b, j) \log_2 \frac{f(b, j)}{p(b)} \]