(* BME 6770: Genomic Signal Processing *)

(* The Hypergeometric Probability Distribution and P-Value *)

(* Initialize *)

Clear["Global`*"]

(* Define the P-Value Function by Using the Built-In Binomial Coefficient *)

? :=

> **Symbol**   ⓘ
>
> *lhs* := *rhs* assigns *rhs* to be the delayed value of *lhs*. *rhs* is maintained in an
>
>   unevaluated form. When *lhs* appears, it is replaced by *rhs*, evaluated afresh each time.
>
> ⌄

? Binomial

> **Symbol**   ⓘ
>
> Binomial[*n*, *m*] gives the binomial coefficient $\binom{n}{m}$.
>
> ⌄

(* One Formulation *)

pValue1[k_, K_, m_, M_] := N[Sum[Binomial[K, i] * Binomial[M - K, m - i], {i, k, m}] / Binomial[M, m]];
pValue1[84, 86, 86, 251]
pValue1[84, 86, 86, 251] == pValue1[(251 - 86) - (86 - 84), 251 - 86, 251 - 86, 251]

8.01057×10^{-62}

True

(* Another Mathematically and Computationally Equivalent Formulation *)

pValue2[k_, K_, m_, M_] := N[Sum[Binomial[K, i] * Binomial[M - K, m - i], {i, k, K}] / Binomial[M, m]];
pValue2[84, 86, 86, 251]
pValue2[84, 86, 86, 251] == pValue1[84, 86, 86, 251]

8.01057×10^{-62}

True

(* Mathematically Equivalent but Computationally Inequivalent Formulation *)

pValue[k_, K_, m_, M_] := 1 - N[Sum[Binomial[K, i] * Binomial[M - K, m - i], {i, 0, k - 1}] / Binomial[M, m]];
pValue[84, 86, 86, 251]

0.

```
pValue1[84, 86, 86, 251]
1 - pValue1[84, 86, 86, 251]
1 - (1 - pValue1[84, 86, 86, 251])
```

8.01057×10^{-62}

1.

0.

(* Define the P-Value Function by Using the the Built-In Hypergeometric Distribution *)

? HypergeometricDistribution

Symbol	ⓘ
HypergeometricDistribution[n, n_{succ}, n_{tot}] represents a hypergeometric distribution.	
⌄	

? PDF

Symbol	ⓘ
PDF[$dist$, x] gives the probability density function for the distribution $dist$ evaluated at x.	
PDF[$dist$, {x_1, x_2, ...}] gives the multivariate	
probability density function for a distribution $dist$ evaluated at {x_1, x_2, ...}.	
PDF[$dist$] gives the PDF as a pure function.	
⌄	

```
probability[k_, K_, m_, M_] := N[PDF[HypergeometricDistribution[m, K, M], k]];
probability[84, 86, 86, 251] + probability[85, 86, 86, 251] + probability[86, 86, 86, 251] ==
 pValue1[84, 86, 86, 251]
```

True

? CDF

Symbol	ⓘ
CDF[$dist$, x] gives the cumulative distribution function for the distribution $dist$ evaluated at x.	
CDF[$dist$, {x_1, x_2, ...}] gives the multivariate cumulative	
distribution function for the distribution $dist$ evaluated at {x_1, x_2, ...}.	
CDF[$dist$] gives the CDF as a pure function.	
⌄	

(* Yet Another Mathematically and Computationally Equivalent Formulation *)

```
pValue3[k_, K_, m_, M_] := N[1 - CDF[HypergeometricDistribution[m, K, M], k - 1]];
pValue3[84, 86, 86, 251] == pValue1[84, 86, 86, 251]
```

True