### The Gamma Distribution

- \( X \sim \text{Gamma}(\alpha, \beta) \)
- \( f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \) for \( x > 0 \).
- \( E(X) = \frac{\alpha}{\beta} \)
- \( \text{Var}(X) = \frac{\alpha}{\beta^2} \)

### A Model for the Data from a Two-Treatment Experiment

- Assume there are \( J \) genes indexed by \( j = 1, 2, ..., J \).
- Data for gene \( j \) is \( x_j = (x_{j1}, x_{j2}, ..., x_{jI}) \) where \( x_{ji} \) is the normalized measure of expression on the original scale for the \( j \)th gene and \( i \)th experimental unit.
- Let \( s_1 \) denote the subset of the indices \( \{1, ..., I\} \) corresponding to treatment 1.
- Let \( s_2 \) denote the subset of the indices \( \{1, ..., I\} \) corresponding to treatment 2.

### The Model (continued)

- If gene \( j \) is differentially expressed, then

\[
\{x_{ji} : i \in s_1\} \sim \text{Gamma}(\alpha, \lambda_j) \text{ with mean } \frac{\alpha}{\lambda_j} ,
\]

where \( \lambda_j \sim \text{Gamma}(\alpha_0, \nu) \), and

\[
\{x_{ji} : i \in s_2\} \sim \text{Gamma}(\alpha, \lambda_{j2}) \text{ with mean } \frac{\alpha}{\lambda_{j2}} ,
\]

where \( \lambda_{j2} \sim \text{Gamma}(\alpha_0, \nu) \).

- All random variables are assumed to be independent.

- \( p, \alpha, \alpha_0, \) and \( \nu \) are unknown parameters to be estimated from the data.
An example of how the model is imagined to generate the data for the jth gene.

• Suppose \( p = 0.05, \alpha = 12, \alpha_0 = 0.9, \) and \( \nu = 36. \)
• Generate a Bernoulli random variable with success probability 0.05. If the result is a success the gene is DE, otherwise the gene is EE.
• If EE, generate \( \lambda_j \) from Gamma(\( \alpha_0 = 0.9, \nu = 36 \))
• Then generate i.i.d. expression values from Gamma(\( \alpha = 12, \lambda_j \)).

Example Continued

• If the gene is DE, generate \( \lambda_{j1} \) and \( \lambda_{j2} \) independently from Gamma(\( \alpha_0 = 0.9, \nu = 36 \)).
• Then generate treatment 1 expression values i.i.d. from Gamma(\( \alpha = 12, \lambda_{j1} \)), and
• generate treatment 2 expression values i.i.d. from Gamma(\( \alpha = 12, \lambda_{j2} \)).

Coefficient of Variation is Constant across Gene-Treatment Combinations

• Coefficient of Variation \( CV = \text{sd} / \text{mean} \)
• Conditional on the mean for a gene-treatment combination, say \( \alpha / \lambda_{jk} \), the CV for the expression data is the CV of Gamma(\( \alpha, \lambda_{jk} \)).
• CV of Gamma(\( \alpha, \lambda_{jk} \)) is \( \frac{\alpha^{1/2}}{\alpha / \lambda_{jk}} = \frac{1}{\alpha^{1/2}}. \)
• Note that \( \alpha \) is assumed to be the same for all gene-treatment combinations.
Marginal Density for Gene j

\[ f(x_j) = p \cdot f_{DE}(x_j) + (1-p) \cdot f_{EE}(x_j) \]

Marginal Likelihood for the Observed Data

\[ f(x_1) \cdot f(x_2) \cdots f(x_J) \]

Use the EM algorithm to find values of \( p, \alpha, \alpha_0, \) and \( v \) that make the log likelihood as large as possible.

The posterior probability of differential expression for gene j is obtained by replacing \( p, \alpha, \alpha_0, \) and \( v \) in

\[ \frac{p \cdot f_{DE}(x_j)}{p \cdot f_{DE}(x_j) + (1-p) \cdot f_{EE}(x_j)} \]

with their maximum likelihood estimates.

Software for EBArrays is available at http://www.biostat.wisc.edu/~kendzior.

Extension to Multiple Treatment Groups

• If there are 3 treatment groups, each gene can be classified into 5 categories rather than just the two categories EE and DE:
  a) 1=2=3
  b) 1=2\neq3
  c) 1\neq2=3
  d) 1=3\neq2
  e) 1\neq2, 2\neq3, 1\neq3.

• Extensions to more than 3 groups can be handled similarly.