Software Size Growth and Uncertainty:

Both Affect Estimate Accuracy

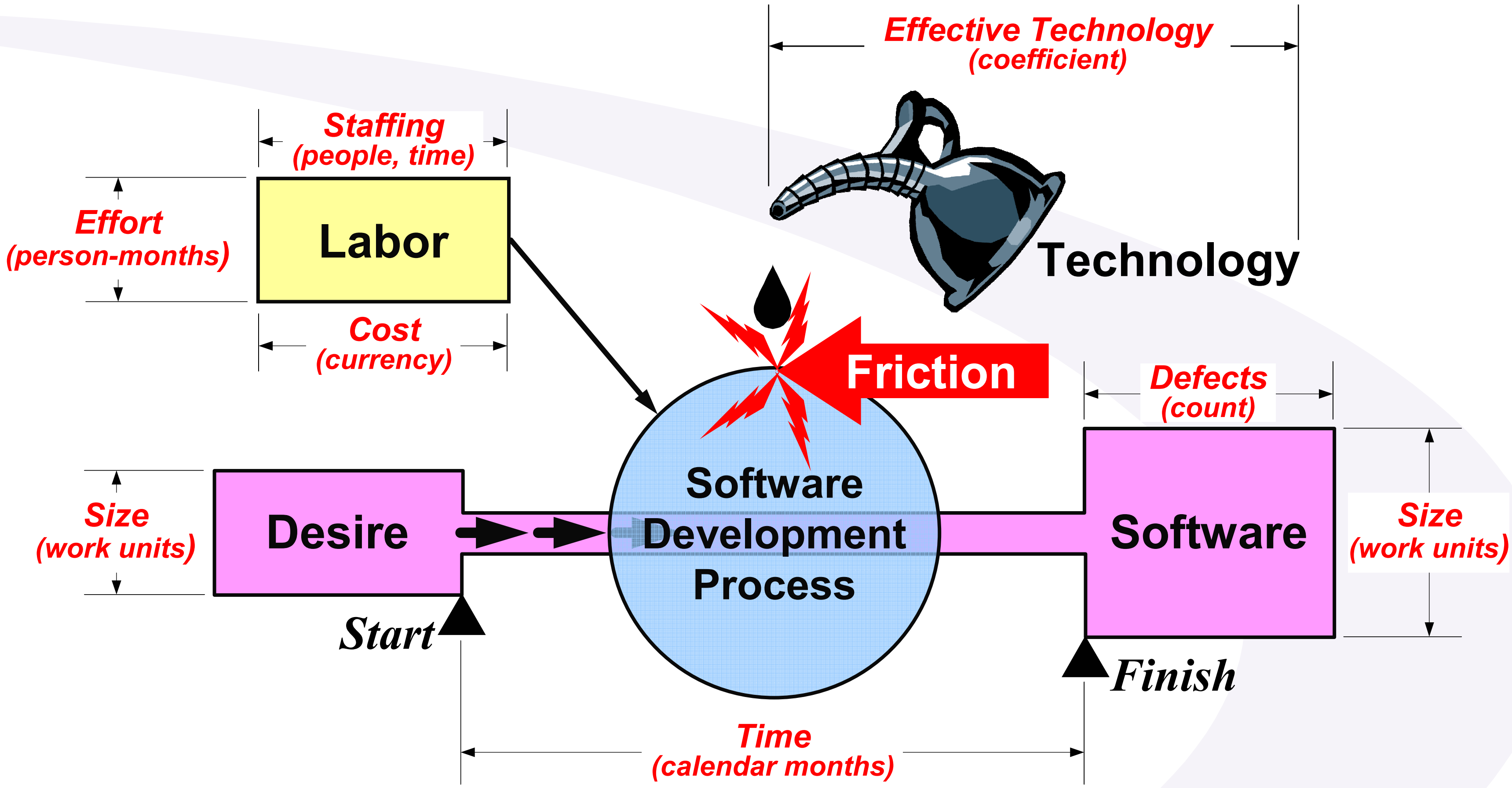
Presented by:
Mike Ross, Chief Engineer

Galorath Incorporated
100 North Sepulveda Boulevard
Suite 1801
El Segundo, California 90245
480.488.8366 (o)   480.488.8420 (f)
mross@galorath.com

Summary

- **Measurement objectifies management**
- **Estimation is a function of progress (continuous process)**
- **A well-formed estimate is specified as a probability distribution**
- **Uncertainty ←**
  - Variability
  - Risk
  - Opportunity
- **Software size estimates ←**
  - Size growth
  - Size estimation variability

Software Development and Measurement



Fundamental Measures

Size

Effective Technology

Time

Effort → Cost, Staffing

Defects

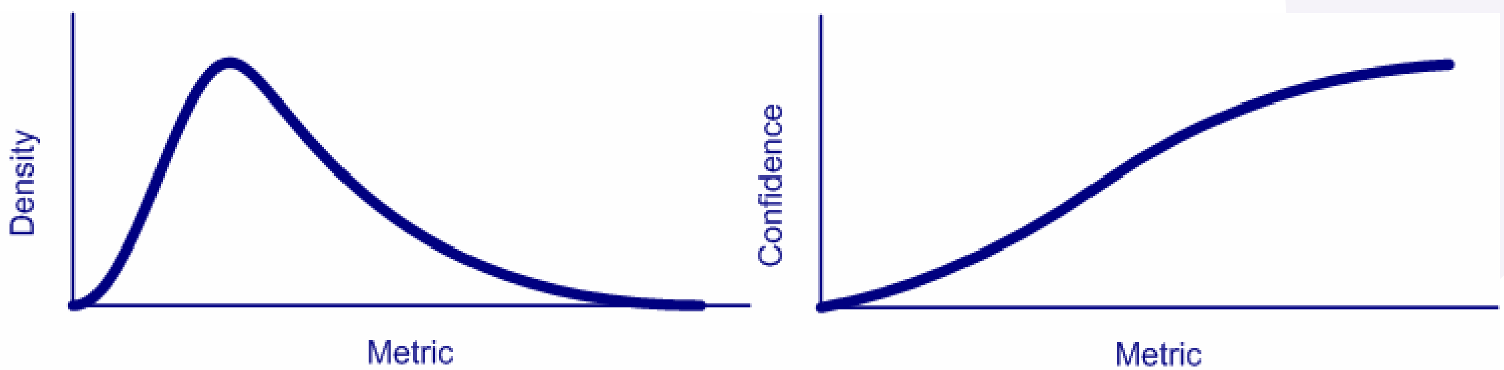
Estimate Defined

es·ti·mate (es′ti mit), *n.*

an approximate *judgment* or *calculation*, as of the value or amount of something

a prediction that is equally likely to be above or below the actual result (Tom DeMarco)

A WELL FORMED ESTIMATE IS A DISTRIBUTION



Two Key Drivers of Software Size Estimates

- **Size Growth**
  - Change in the baseline estimated software size due to:
    - Change in development and/or operating environment
    - Change in the required functionality
  - Technological and Programmatic risk
- **Size Estimation Variability**
  - Estimation process variability due to:
    - Human behavior
    - Model behavior

Size Growth

- **Operational Environment Volatility**
  - The mission changes.
  - The regulations that govern how this software should behave have changed.
- **Essence (Requirements) Volatility**
  - The customer doesn't know what he/she wants.
- **Essence Understanding (Requirements Completeness and Correctness)**
  - The customer doesn't understand the problem.
  - The specifications are vague.
- **Essence versus Implementation Correspondence**
  - The vendor adds a few extra features (gold plating).

Growth Factor Function

- **Yields Growth Factor as a function of normalized earned value**
- **Based on Galorath Incorporated analysis of historical data**
- **Embedded in *SEER-SEM*™'s Phase at Estimate parameter**

$$G(s) = -0.7s + 0.69$$

Growth Factor Function Distribution

- Triangular Distribution per (Book 2002)
- Skew per modified (Tarbet 2002)

$$\mathbf{G}(s) = \begin{bmatrix} L & M & H \end{bmatrix} = \begin{bmatrix} 0 & 0 & \mathrm{G}(s) \end{bmatrix}$$

Size Growth Impact Distribution

- Function of normalized earned value (progress)
- Product of best guess size and growth factor
- Triangular Distribution per (Book 2002) from growth factor
- Skew per (Tarbet 2002) from growth factor

$$\mathbf{S}_{\mathbf{G}}(s) = \mathrm{S}_{\mathrm{M}}(s)\mathbf{G}(s) = \begin{bmatrix} 0 & 0 & \mathrm{S}_{\mathrm{M}}(s)\mathrm{G}(s) \end{bmatrix}$$

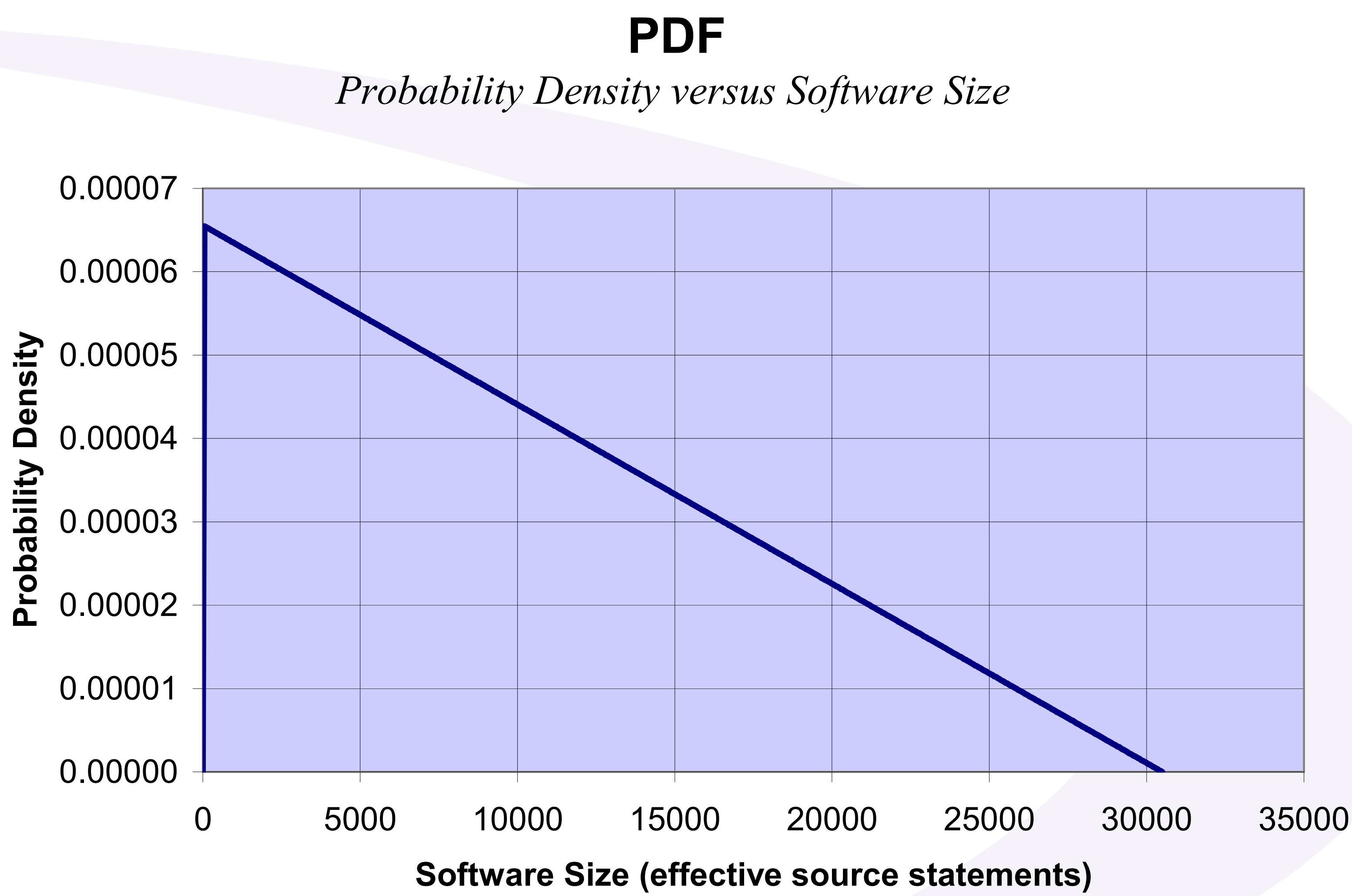
Size Growth Example Calculation

- **Assume a best guess size at SRR of 50,000 ESLOC**
- **Assume normalized earned value of 11.8% at SRR**

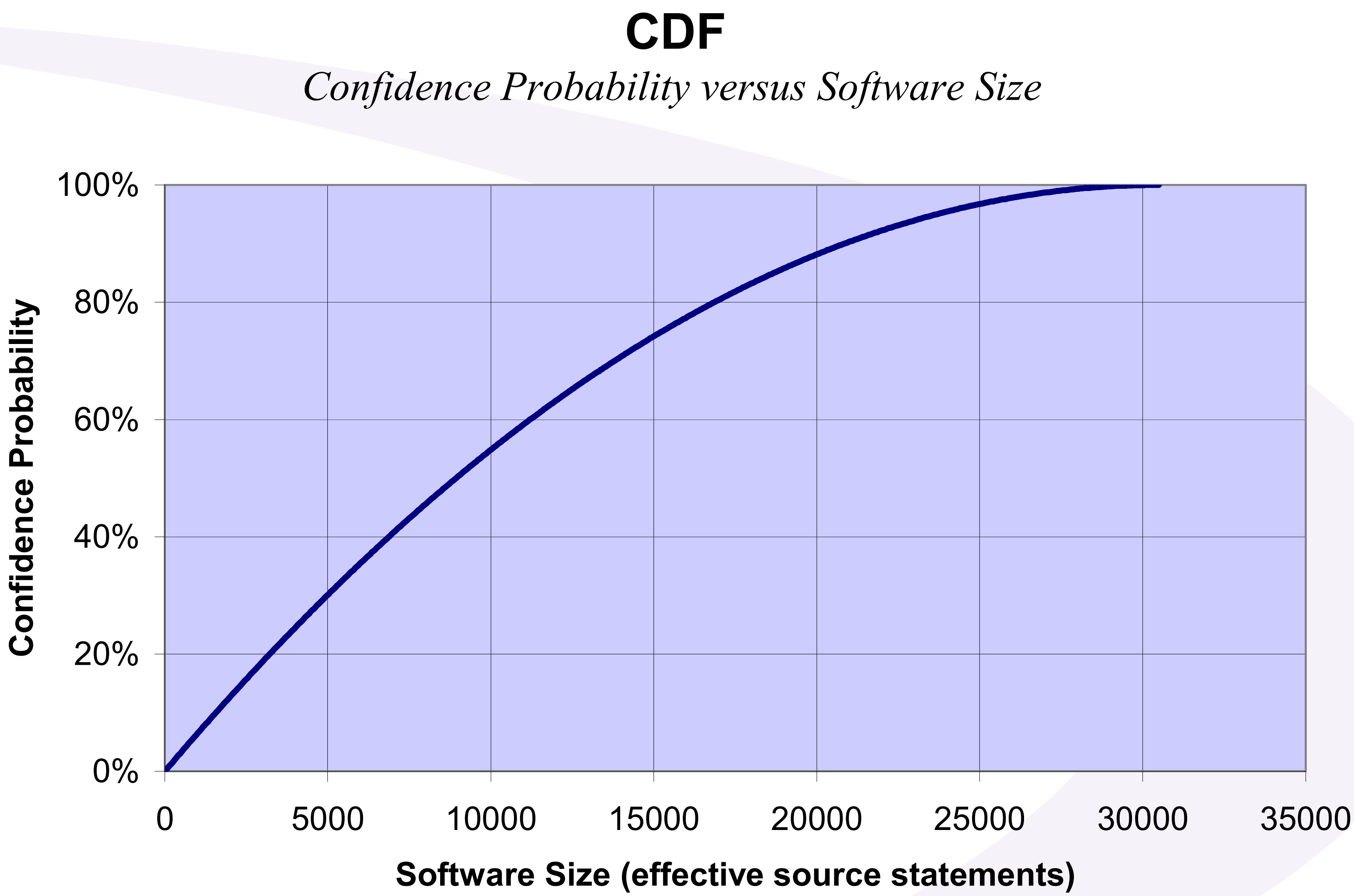
$$G_{SRR} = \mathrm{G}(11.8\%) = -0.7(11.8\%) + 0.69 = 0.61$$

$$\mathbf{S}_{\mathbf{G\_SRR}} = \begin{bmatrix} 0 & 0 & S_{M\_SRR}(0.61) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 30,500 \end{bmatrix}$$

Size Growth Example PDF



Size Growth Example CDF



Size Estimation Variability

- Uncertainty about the translation of essence to implementation
- Error and bias introduced by the estimation process
- Error and bias introduced by the estimation model / relationships
- Error and bias introduced by the people performing the process

Size Estimation Variability Impact Distribution

- Function of normalized earned value (progress)
- Normal (Gaussian) Distribution per (Book 2002)
- Variance per (Tarbet 2002)

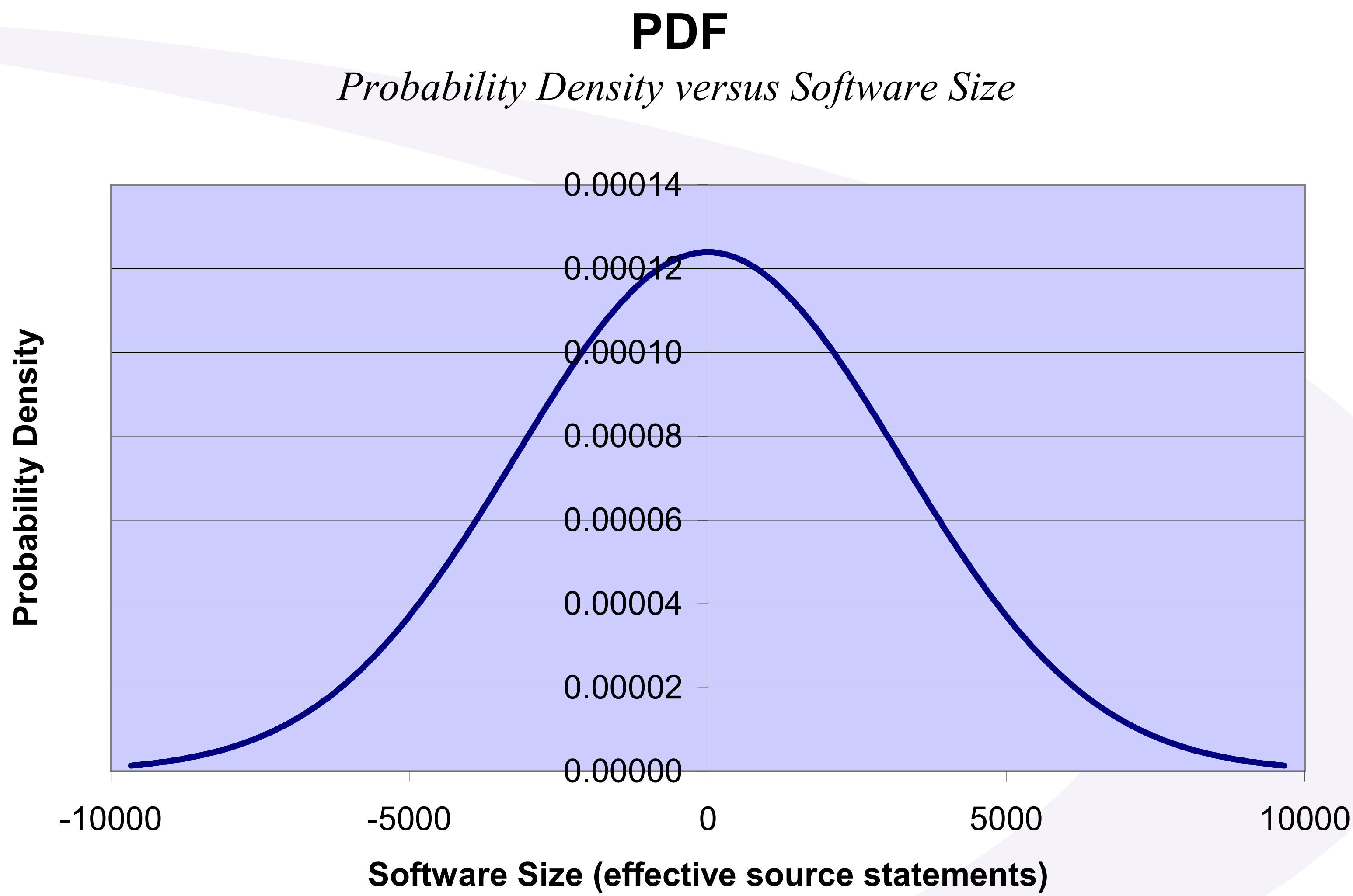
$$\mathbf{S}_{\mathbf{EV}} = \begin{bmatrix} \mu & \sigma \end{bmatrix} = \begin{bmatrix} 0 & \dfrac{(30\%) S_M}{(2)(2.33)} \end{bmatrix} = \begin{bmatrix} 0 & 3{,}219 \end{bmatrix}$$

Size Estimation Variability Example Calculation

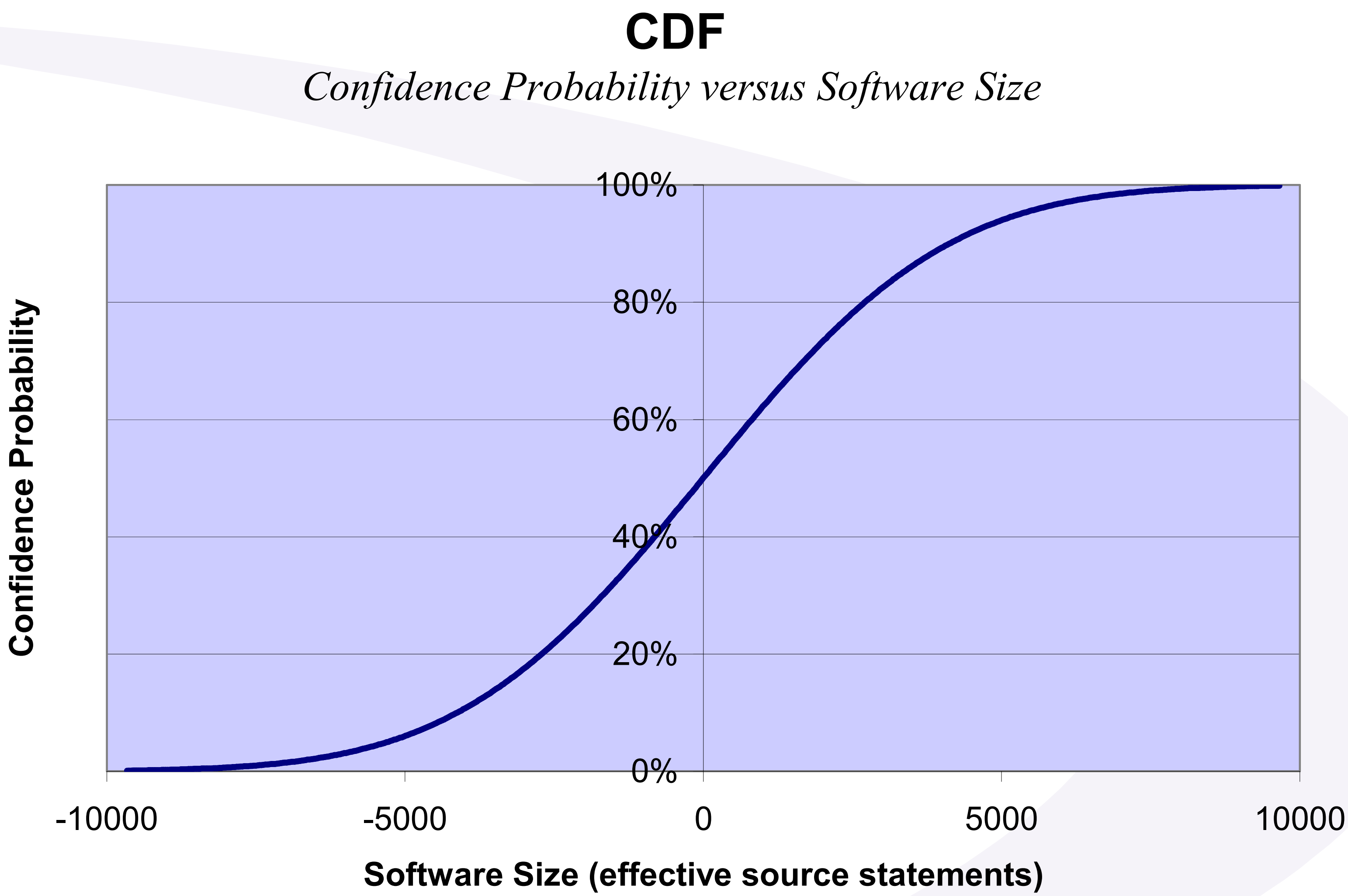
- **Assume a best guess size at SRR of 50,000 ESLOC**

$$\mathbf{S}_{\mathbf{EV\_SRR}} = \left[0 \quad \frac{(30\%) S_{M\_SRR}}{(2)(2.33)} \right] = [0 \quad 3,219]$$

Size Estimation Variability
Example PDF



Size Estimation Example CDF



Combining Size Growth and Size (Estimate) Uncertainty

- **The mean of the sum of a set of random variables is equal to the sum of the means of each random variable in the set**
- **The standard deviation of the sum of a set of independent random variables is equal to the square root of the sum of the squares of the standard deviations of each random variable in the set**

$$\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}\left(X_i\right)$$

$$\mathrm{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{V}\left(X_i\right)$$

Combining Size Growth and Size (Estimate) Uncertainty

- **Sum of the means:**

$$\mu_{S_M(s)} = S_M(s)$$

$$\mu_{S_G(s)} = \frac{0 + 0 + S_M(s)G(s)}{3} = \frac{S_M(s)G(s)}{3}$$

$$\mu_{S_{EV}(s)} = 0$$

$$\therefore \mu_{S(s)} = S_M(s) + \frac{S_M(s)G(s)}{3} = \frac{S_M(s)\left(G(s)+3\right)}{3}$$

Combining Size Growth and Size (Estimate) Uncertainty

- **Square root of the sum of the squares of the standard deviations:**

$$\sigma_{S_M(s)} = 0$$

$$\sigma_{S_G(s)} = \sqrt{\frac{L^2 + M^2 + H^2 - LH - LM - MH}{18}} = \sqrt{\frac{\left(S_M(s)G(s)\right)^2}{18}}$$

$$\sigma_{S_{EV}(s)} = \frac{(30\%)S_M(s)}{(2)(2.33)}$$

$$\therefore \sigma_{S(s)} = \sqrt{\frac{\left(S_M(s)G(s)\right)^2}{18} + \left(\frac{(30\%)S_M(s)}{(2)(2.33)}\right)^2}$$

Example Calculation

- Assume a best guess size at SRR of 50,000 ESLOC
- Assume a growth factor at SRR of 0.61

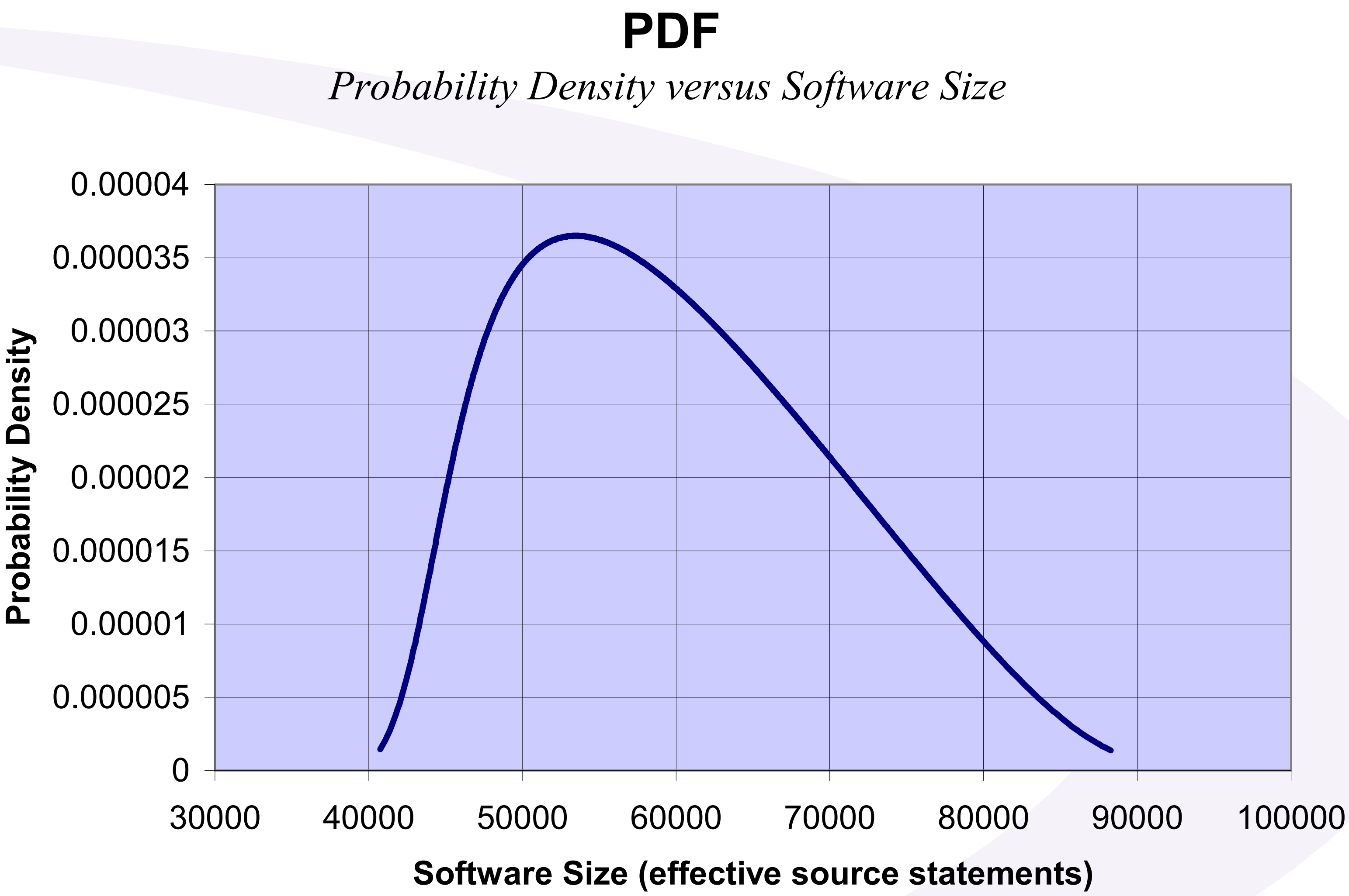
$$\mathbf{S}_{\mathbf{SRR}} = \begin{bmatrix} \mu & \sigma \end{bmatrix}$$

$$\mathbf{S}_{\mathbf{SRR}} = \left[\frac{S_{M\_SRR}\left(G_{SRR}+3\right)}{3} \quad \sqrt{\frac{\left(S_{M\_SRR}G_{SRR}\right)^2}{18} + \left(\frac{(30\%)S_{M\_SRR}}{(2)(2.33)}\right)^2} \right]$$

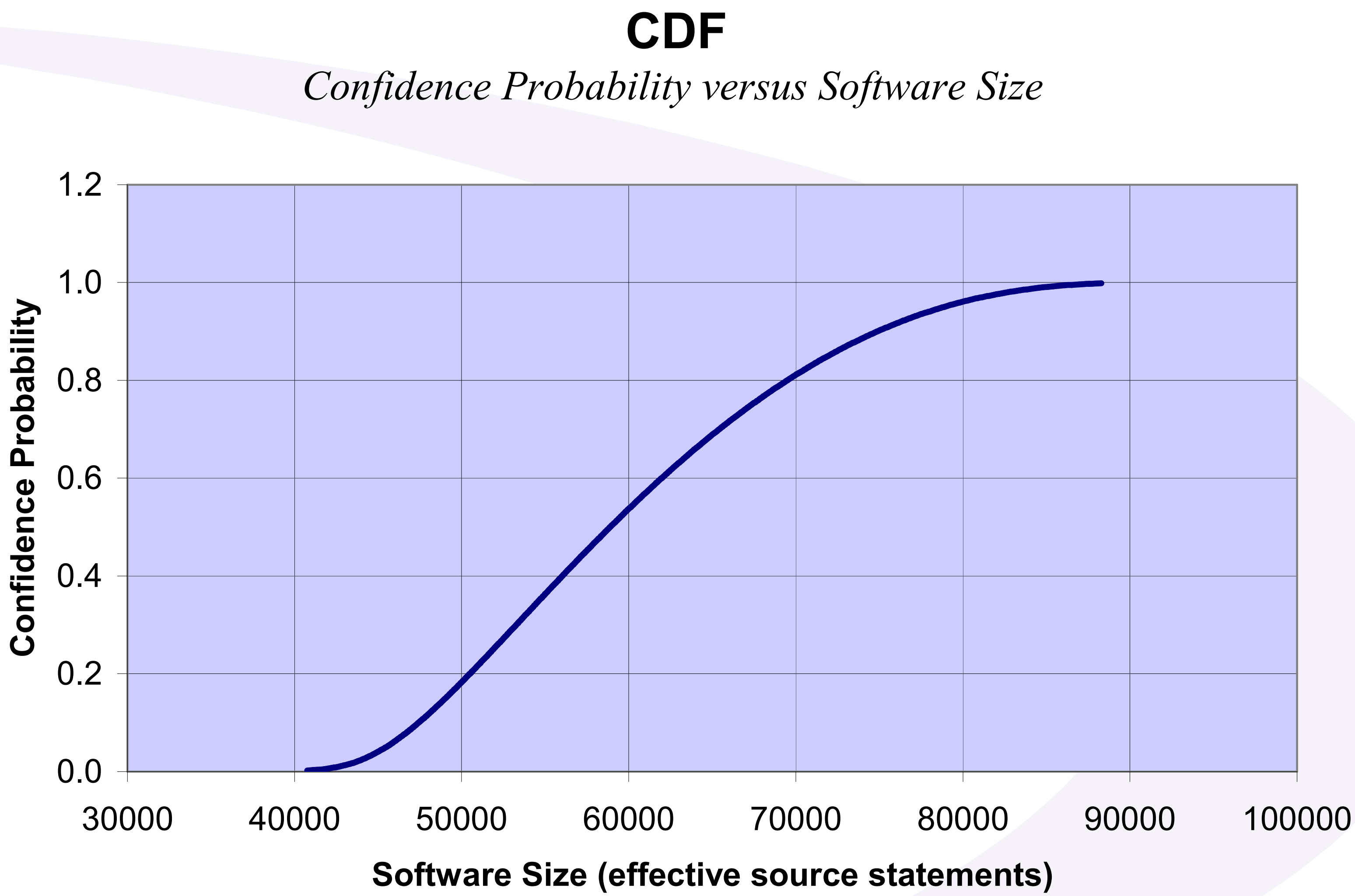
$$\mathbf{S}_{\mathbf{SRR}} = \left[\frac{50,000\left(0.61+3\right)}{3} \quad \sqrt{\frac{\left(\left(50,000\right)\left(0.61\right)\right)^2}{18} + \left(\frac{(30\%)\left(50,000\right)}{(2)(2.33)}\right)^2} \right]$$

$$\therefore \mathbf{S}_{\mathbf{SRR}} = \begin{bmatrix} 60,167 & 7,877 \end{bmatrix}$$

Combined PDF



Combined CDF



Review

- **Measurement objectifies management**
- **Estimation is a function of progress (continuous process)**
- **A well-formed estimate is specified as a probability distribution**
- **Uncertainty ←**
  - Variability
  - Risk
  - Opportunity
- **Software size estimates ←**
  - Size growth
  - Size estimation variability