Estimating Uncertainty using 3-pts estimation

Steve Webb
stevewebb@swestimation.co.uk

3rd June 2014

SW Estimation Ltd

- Cover all aspects of software (i.e. sw) estimation
- Provide: training, CMMI level 5 compliant estimation assets
- Advice on parametric models
- Advice on software sizes e.g. Story Points, SLOC, Function Points
- COSMIC Function Points accurately estimate real-time software effort
- Considerable software estimation experience

Presentation will cover:

- Estimating uncertainty for software & systems engineering
- Show why single point estimation should be avoided
- Deterministic 3-pt estimation
- Some 3-pt estimation issues & misunderstandings
- Monte Carlo simulation
- Correlation

What is a good estimate

- A good estimate should express the possible range of outcomes i.e. 3-pts
  e.g. a minimum (i.e. best case) of 15 units
  most likely 20 units
  max (i.e. worst case) of 35 units

- Most likely = most frequent occurring or modal point of the estimate

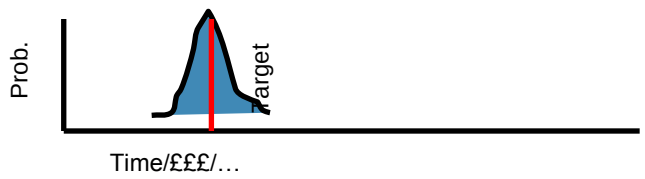
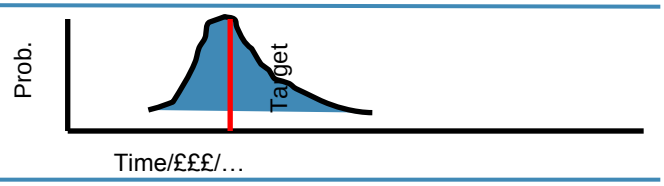
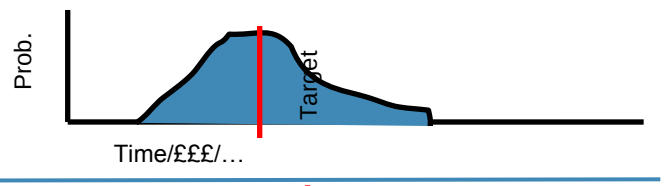
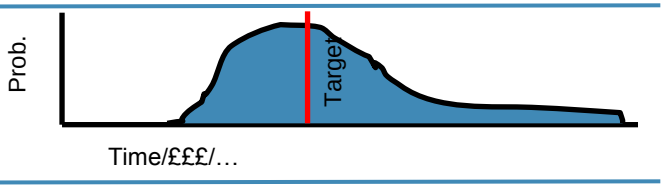
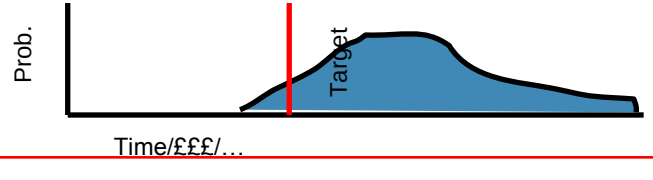
- Min and max are the extremes which could happen

- Also need probabilities of occurrence and the Mean

- Probabilities normally provided by Monte Carlo tools e.g. Arrisca, Crystal Ball

Need an Input distribution that fits the data

Software development tends to follow a Beta distribution which is normally skewed to the right

CMMI Level / Type	Predicted Performance
5 Optimising **Initial estimates within 5% to 10%** **& centred on the actual median / mean.** **The red line is the initial estimate**	
4 Managed	
3 Defined	
2 Repeatable	
1 Initial	

Which input distribution to use

- For each 3 points need to decide on its input distribution

- For Uncertainty in software development:
  - Beta (PERT) is the most widely used distribution for software development
  - If significant amount of unknowns (e.g. new type of business with poor requirements) then triangular distribution could be used

- For Software Risk:
  - Triangular is normally used for software risk
  - Gives a pessimistic estimate compared to Beta – an example later
  - Risks typically less understood compared to development so a more pessimist viewpoint is ok

Beta (PERT) for uncertainty in software development

- Beta (PERT) is an approximation of the mathematical Beta distribution

- Many software authorities (e.g. Boehm, Stutzke, McConnell) use Beta (PERT)

- All estimate the Mean in the same way
- Mean = expected value = unbiased estimate
- The outcome has an equal chance of being above or below this Mean

- To estimate the probabilities the standard deviation (i.e. σ) is required

- Some authorities (e.g. SEER-SEM, Boehm) use a non-standard formula to increase the size of σ which makes the probability values more pessimistic
  - Making the prob. values more pessimistic is not unreasonable

Beta (PERT) & Triangular

- E.g. min = 15, most likely (ml) = 20, max = 35

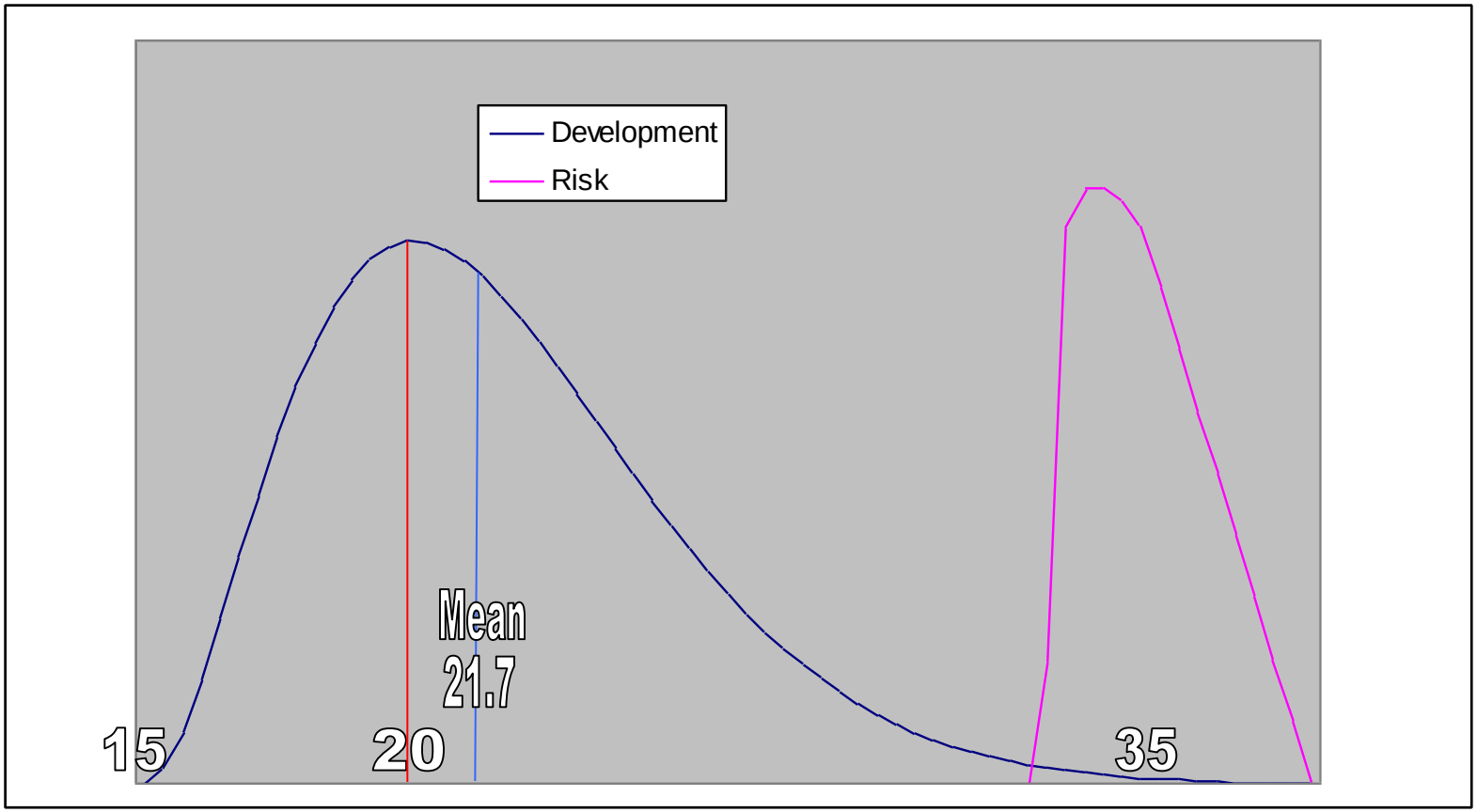
- Beta Mean = (min + (4 * ml) + max) / 6
  = (15 + (4 * 20) + 35) / 6 = 21.7
- So an equal chance that the outcome will be above or below 21.7

- Note the most likely point (i.e. 20) is an underestimate

- Standard Beta (PERT) 1 σ = (max – min) / 6 = (35 – 15) / 6 = 20/6 = 3.3
- Note (max – min) is divided by 6 which is discussed later

- Triangular Mean = (min + ml + max) / 3 = (15 + 20 + 35) / 3 = 23.3
- Triangular 1 σ = *4.2*
- Triangular Mean & σ are more pessimistic than Beta

Beta (PERT)



- Software & Systems size & effort costs tend to follow a right skewed Beta distr.

	Min	Most Likely	Max	Individual Mean
Estimate	**15**	**20**	**35**	**21.7**

3-pt estimation deterministic workbooks

- Within software & systems the use of single points is common practice
- Single points are often an underestimate & provide no probabilities
- SW Estimation Ltd have developed 3-pt estimation deterministic workbooks which use Beta (PERT) & Triangular
- They calculate the Mean & probabilities
- Deterministic Means are always completely accurate
- If Monte Carlo number of simulation iterations is not large enough then Monte Carlo Mean is less accurate than Deterministic Mean

Deterministic 3-point workbook

Item	Min	Most Likely	Max	Mean	Variance	Std. Dev
Overall Totals	**185**	**300**	**505**	**315.0**	**190**	13.8
Item 1 - Gather information	15	20	35	21.7	11.1	3.3
Item 2 - Write draft document 1	10	15	25	15.8	6.3	2.5
Item 3 - Review doc 1	5	10	20	10.8	6.3	2.5
Item 4 - Correct defects for doc 1	20	30	40	30.0	11.1	3.3
Item 5 - Write draft document 2	15	20	40	22.5	17.4	4.2
Item 6 - Review doc 2	5	10	20	10.8	6.3	2.5
Item 7 - Correct defects for doc 2	20	30	40	30.0	11.1	3.3
Item 8 - Write draft document 3	10	20	40	21.7	25.0	5.0
Item 9 - Review doc 3	5	10	20	10.8	6.3	2.5
Item 10 - Correct defects for doc 3	15	25	45	26.7	25.0	5.0
Item 11 - Write draft document 4	10	15	25	15.8	6.3	2.5
Item 12 - Review doc 4	5	10	20	10.8	6.3	2.5
Item 13 - Correct defects for doc 4	20	30	40	30.0	11.1	3.3
Item 14 - Write draft document 5	10	20	35	20.8	17.4	4.2
Item 15 - Review doc 5	5	10	20	10.8	6.3	2.5
Item 16 - Correct defects for doc 5	15	25	40	25.8	17.4	4.2

- User inputs mins, most likely's, maxs into green areas & rest is calculated
- Overall mean = sum of individual means = 315
- An equal chance of 16 work items being done above or below 315

Spreadsheet formula

- This slide only for those who want to understand the workbook calculations

Individual standard deviation (σ) for Beta (PERT) = (max – min) / 6

Individual variance for Beta (PERT) = σ * σ

Individual variance for Triangular

= ((max - min)*(max - min) + (mean - min)*(mean - max)) / 18

Overall variance = Sum of the individual variances

Overall standard deviation = Square root of the overall variance

What is Correlation?

- Positive correlation for items 1 and 2 means:
  - increase in cost of item 1 means an increase the cost in item 2
  - increase in cost of item 2 means an increase the cost in item 1

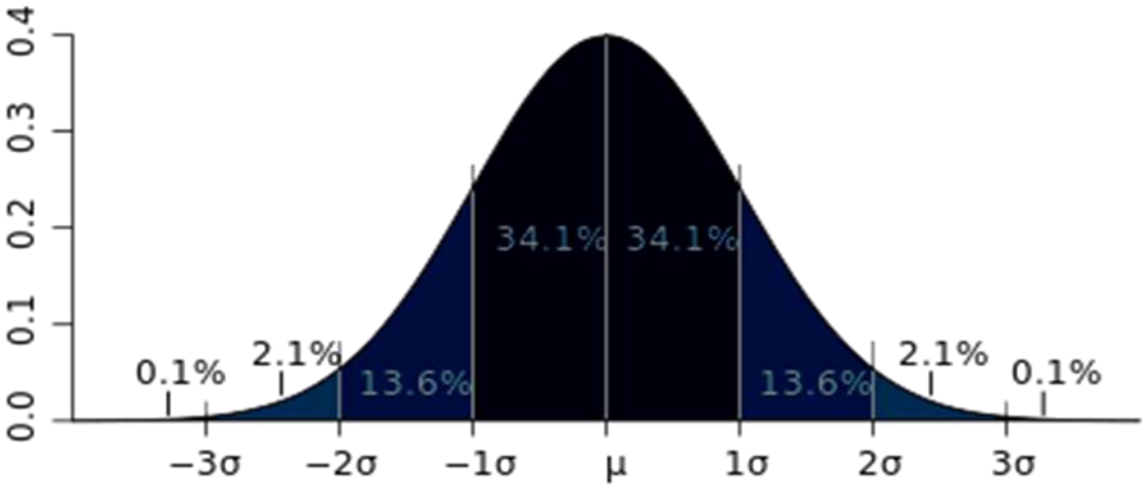
  - decrease in cost of item 1 means a decrease the cost in item 2
  - decrease in cost of item 2 means a decrease the cost in item 1

- No correlation means an increase/decrease in item 1 does not effect item 2

Standard normal distribution for 0% correlation

- This slide shows why workbook uses the standard normal for the output distribution
- Central limit theorem states: “the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed”
- So no matter what shape of the input distributions (e.g. Beta, Triangular) by adding 3 or 4 individual means the output distribution approximates to a Normal distribution (ND) when there is no correlation
- Workbook uses the ND as its output distribution for non-correlated work
- Convergence is very rapid, with just 5 items its close to a ND
- For correlated items the output distribution is not a Normal distribution

6 *sigma (σ)*



- The above is a Normal distribution
- 99.7% of area covered by -3 *σ* to + 3 *σ* i.e. 6 *σ* range

- For Beta (PERT), 1 individual *σ* = (max – min) / 6
- So 6 *σ* = (max – min)
- So max & min should cover 99.7% of occurrences
- So max & min should be extreme values

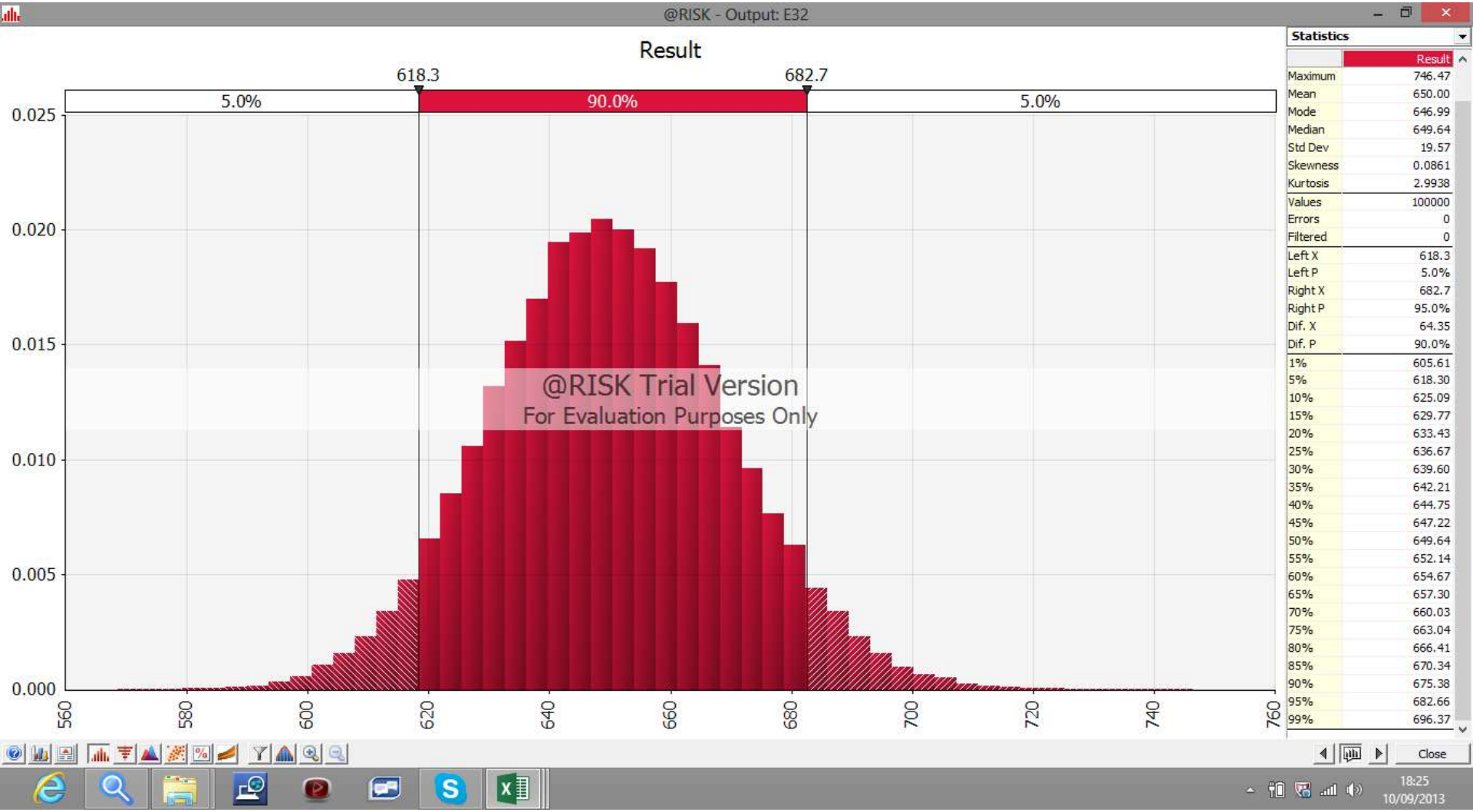
No (i.e. 0%) correlation spreadsheet example

- Beta (PERT) contrived e.g. of 30 items with same min, ml, max of (15, 20, 35)
- With 0% correlation the output is a standard normal distribution e.g.

0% Correlation				
1% probability point =	608			
5% probability point =	620			
10% point =	627			
Overall Mean =	650			
70% point =	660			
90% point =	673			
95% point =	680			
99% point =	692			
Item	**Min**	**Most Likely**	**Max**	**Mean**
Overall Totals	**450**	**600**	**1,050**	**650**
Item 1	15	20	35	21.7
Item 2	15	20	35	21.7
……				
Item 29	15	20	35	21.7
Item 30	15	20	35	21.7

- Note Triangular mean = 700, Triangular 90% pt = 730

Monte Carlo simulation also gives Standard Normal



- Note 50% pt = Overall mean = 650 - only happens when no correlation

Most Likely

0% Correlation	
1% probability point =	608
10% point =	627
Overall Mean =	650
70% point =	659
90% point =	673
99% point =	692

Item	Min	Most Likely	Max	Mean
Overall Totals	450	600	1,050	650

- If single pt estimation was used then estimate = sum of most likely's = 600
- For 0% correlation, only 1% probability that work done for less than 608!
- Clearly sum of most likely's is a serious underestimate

- Using schedule most likely's normally causes serious underestimation
- Conclusion: Sum of the most likely's is often a misleading statistic

Sum of Mins and Maxs

0% Correlation	
1% probability point =	608
10% point =	627
Overall Mean =	650
70% point =	659
90% point =	673
99% point =	692

Item	Min	Most Likely	Max	Mean
Overall Totals	450	600	1,050	650

- 1% probability pt = 608
- 0.1% probability pt = 594
- Sum of the Mins (i.e. 450) is practically impossible to achieve

- 99% probability pt = 692
- 99.9% probability point = 706
- Sum of the Maxs (i.e. 1050) is practically impossible to achieve

Workbook 0% & 100% correlated results

0% Correlation	
1% probability point =	608
10% point =	627
Overall Mean =	650
70% point =	659
90% point =	673
99% point =	692

100% Correlation	
1% point =	470
10% point =	509
Mean =	650
70% point =	708
90% point =	802
99% point =	917

Item	Min	Most Likely	Max	Mean
Overall Totals	450	600	1,050	650

- 100% correlation results = all 30 items fully correlated with each other
- In practice very unlikely to have all 30 items fully correlated
- Note big differences in 1% prob pt values
- Note big differences in 99% prob pt values

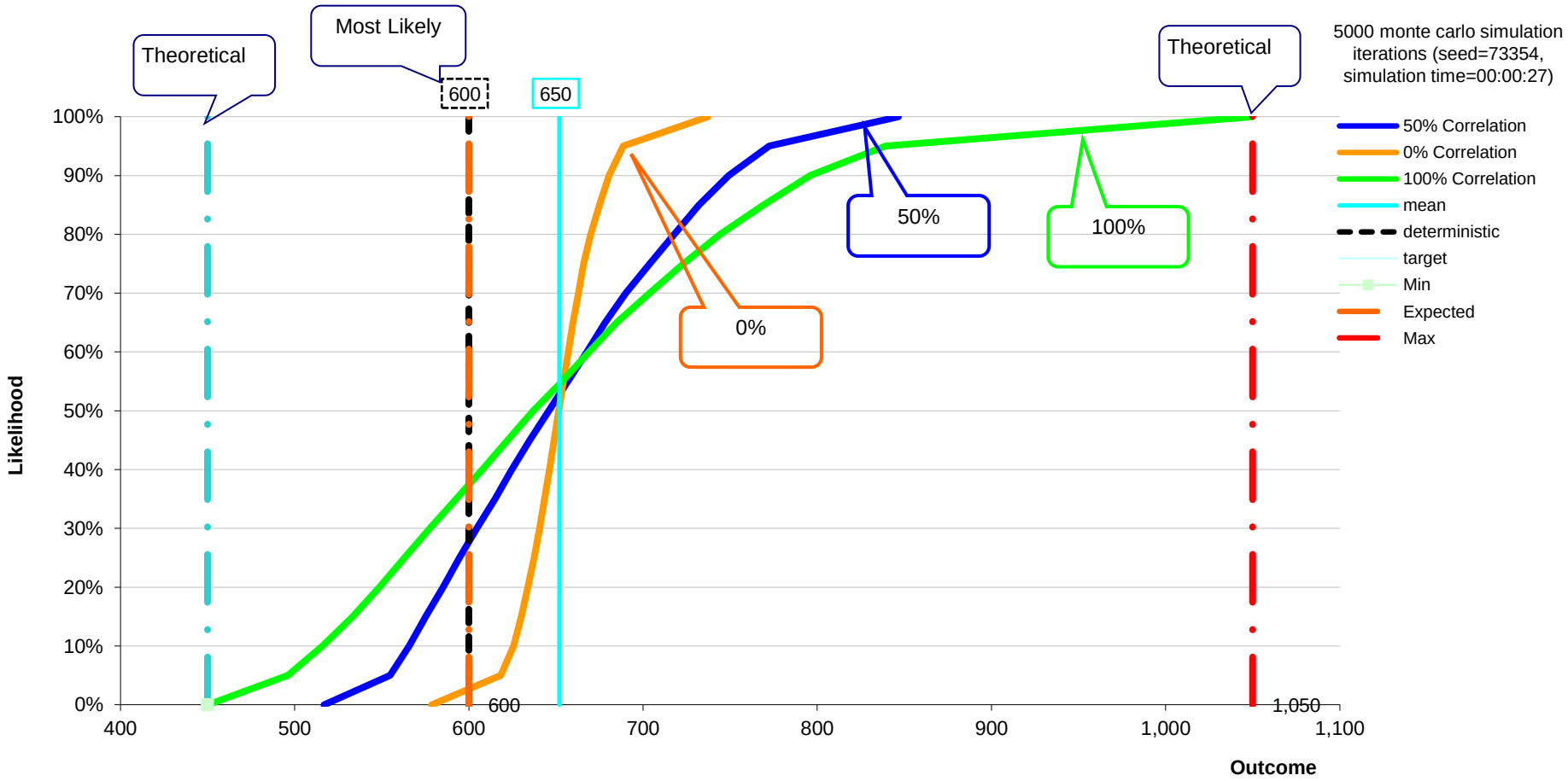
- Sum of the Mins (i.e. 450) still practically impossible to achieve
- Sum of the Maxs (i.e. 1050) still practically impossible to achieve

Adding up Mins, Most Likely's, Mins

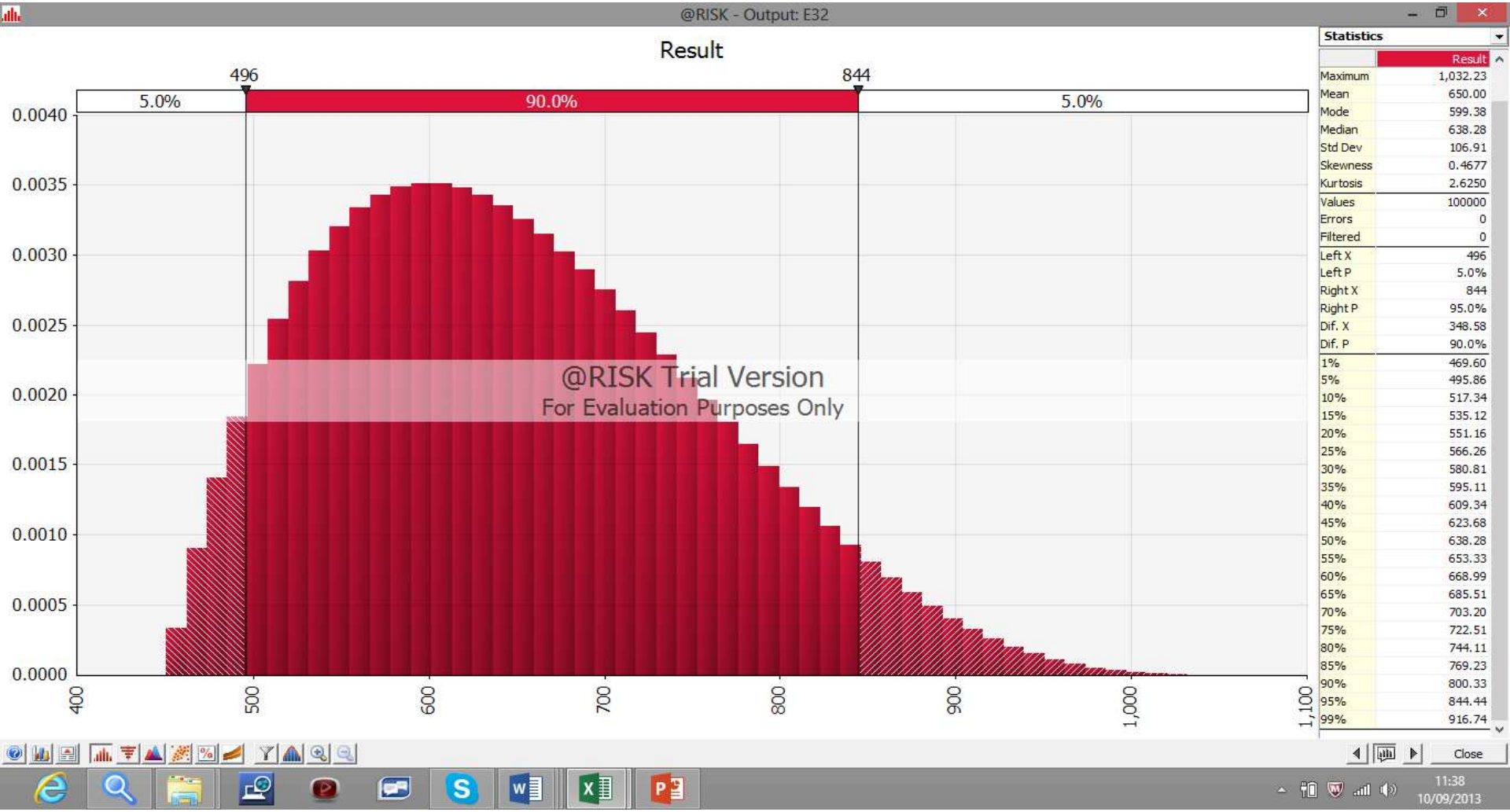
- To determine the range some software or systems engineers:
  - add up all min's together
  - add up all the max's together
- By adding up all mins & maxs it is making everything 100% correlated resulting in predicted values that's impossible to achieve e.g. max = 1050
- They also add up all the most likely's to determine the predicted "outcome"
- Even some estimation "experts" add up the mins, ml, maxs!!
- Adding up any of the above is not useful
- The ONLY thing useful is to add up are the Means

RiskHive (Arrisca) 0%, 50%, 100% correlation output

- Overall Mean = 650 regardless on the amount of correlation



Monte Carlo 100% correlation output results



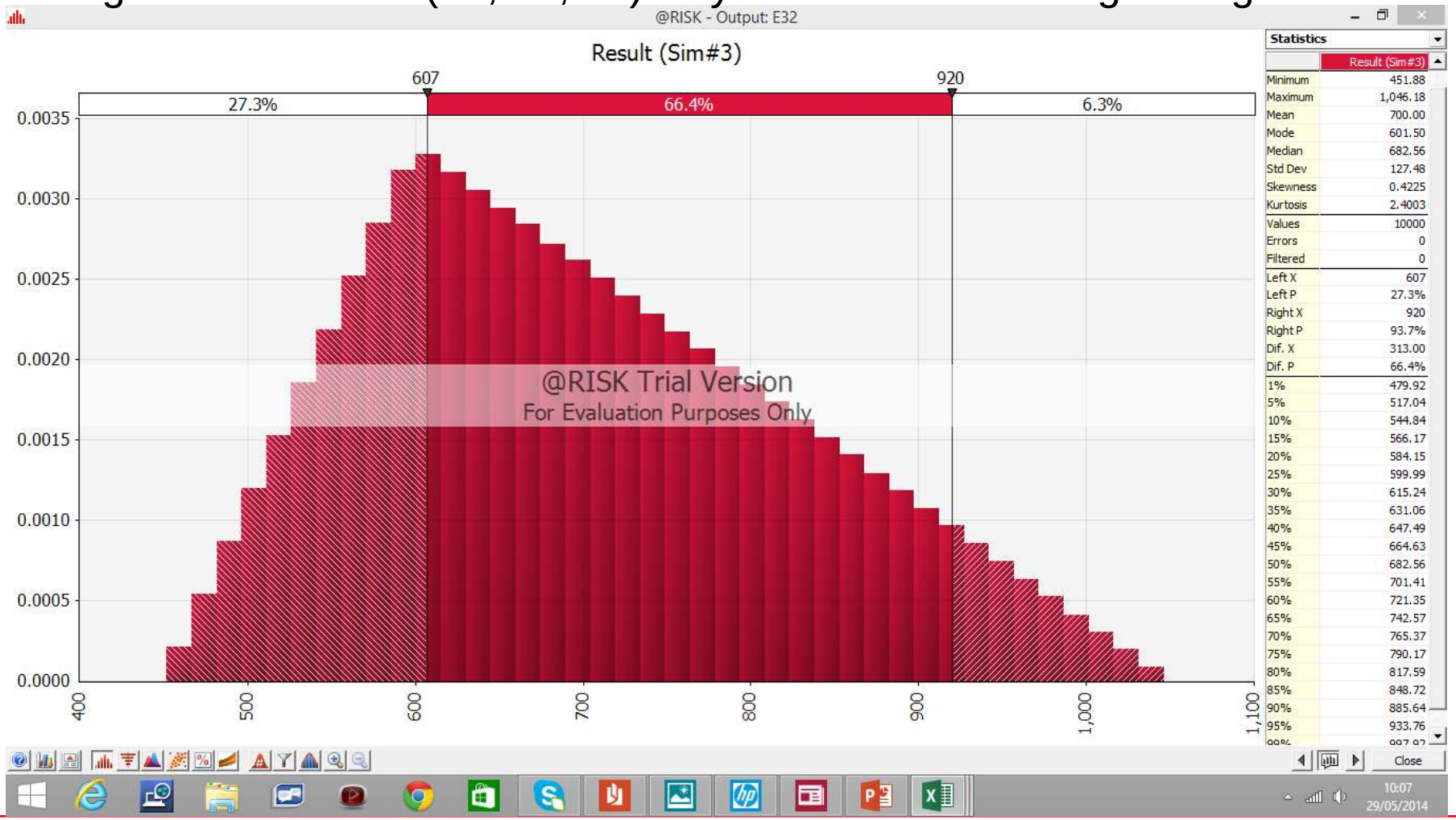
- All 30 Beta items (15, 20, 35) fully correlated creates Beta (PERT) output
- Note 50% pt = 638 but Mean = 650

Mean & 50% pt

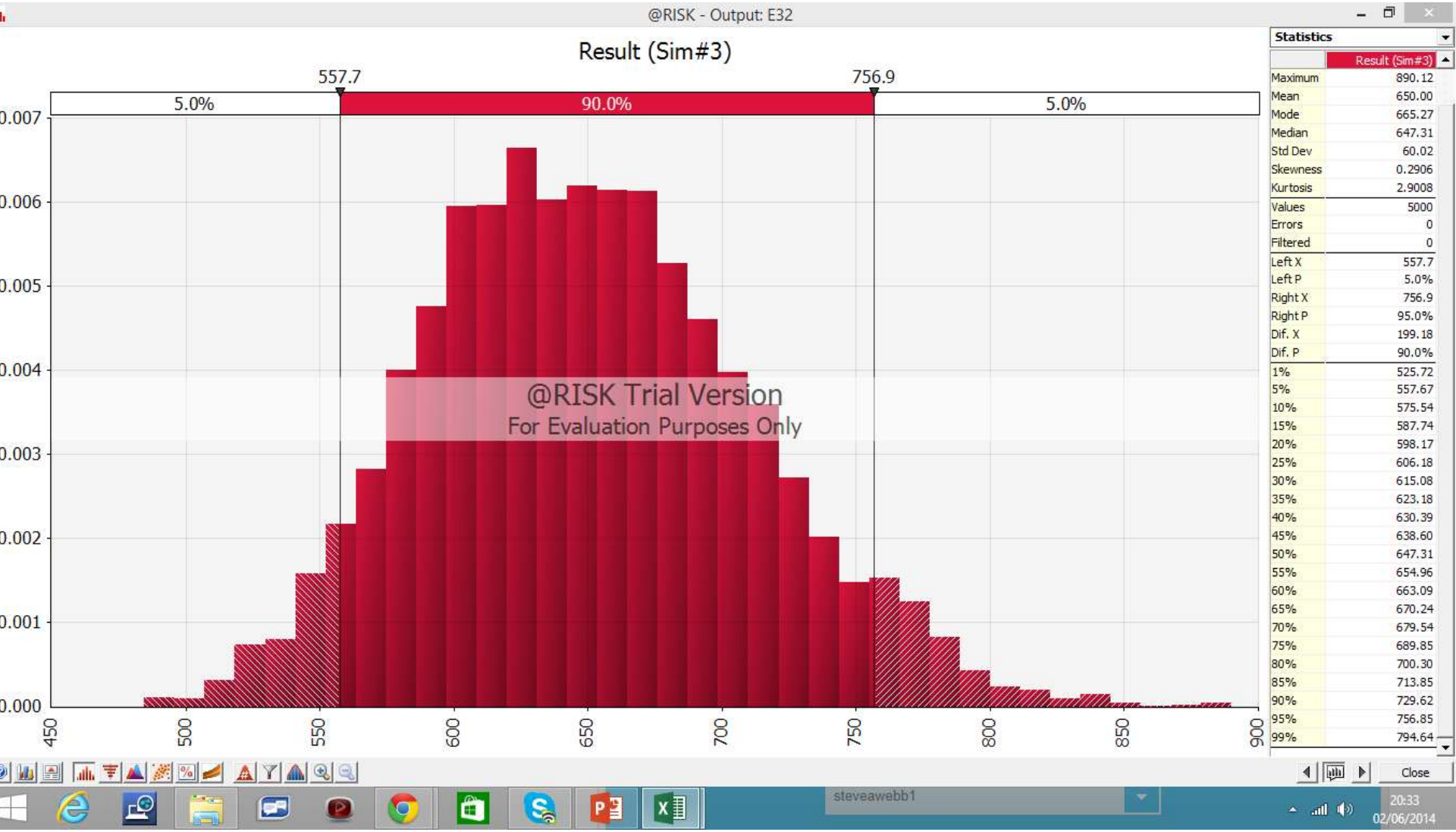
- For zero correlation the 50% pt = Mean = 650
- Mean = unbiased estimate
- For non-zero correlation the 50% pt does not equal the Mean
  E.g. 638 (50% pt) ≠ 650 (Mean)
- Monte Carlo % pts are simulation sample pts
- E.g. 50% pt breaks the number of samples into 2 equal chunks

Monte Carlo 100% correlation Triangular output results

- Again all 30 items (15, 20, 35) fully correlated but using Triangular



A more typical output i.e. 25% correlation



- Sum of most likely's = 600. 20% probability = 598. Note 99% pt = 795 ≠ 1050

Software development correlation

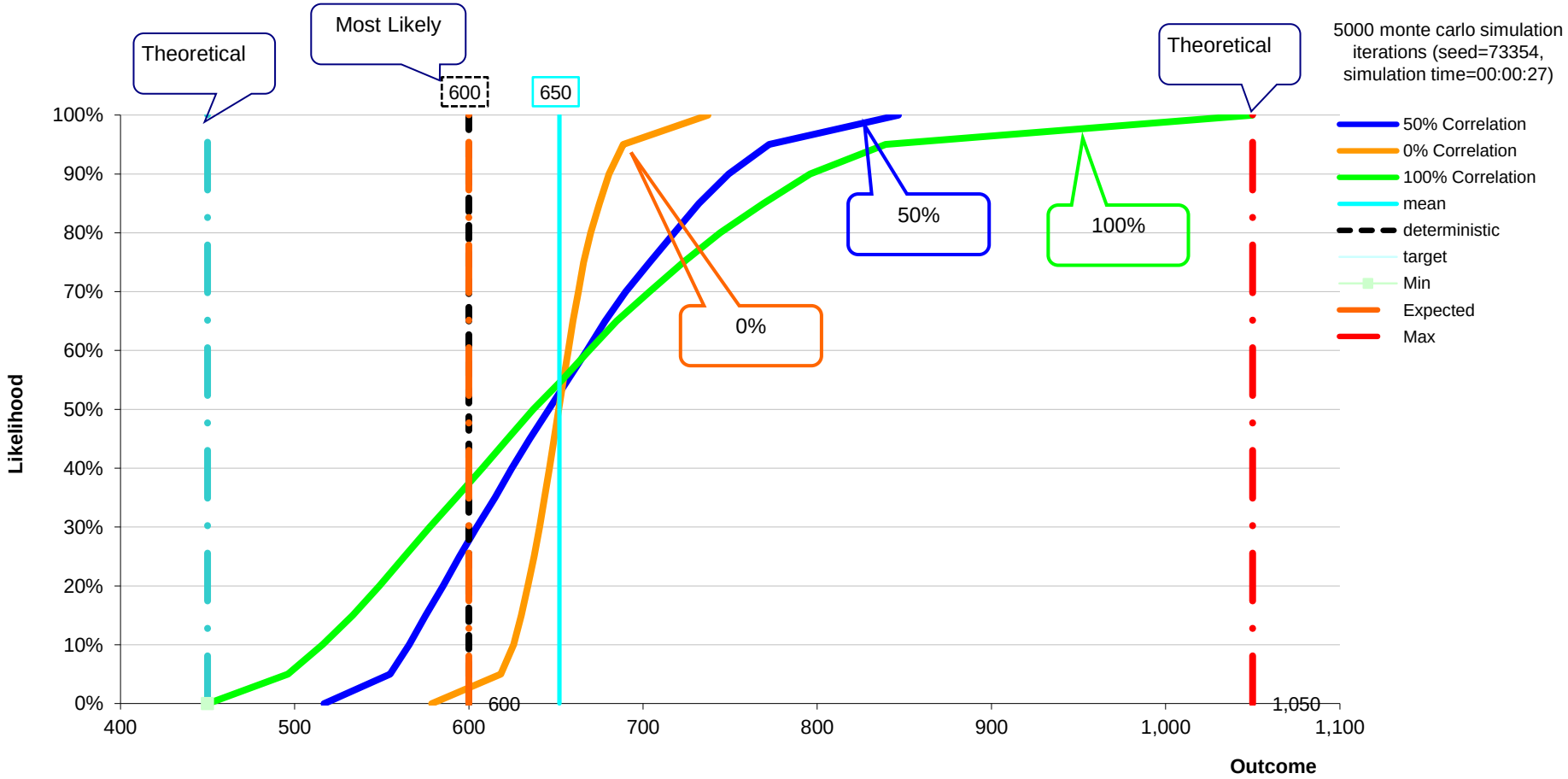
- The amount of correlation within software is not well understood
- Empirical evidence suggests a default average value of 25% to 30% correlation is reasonable
- However typically some items will have no correlation whilst other items have higher correlation than 30%
- Also some types of software has much higher correlation than others
- E.g. radar tracking system has much higher correlation compared to a database system where each component can be fully tested on its own

Simulation tools & correlation

- Not all Monte Carlo simulation tools calculate correlation in the same way
- Tools give similar correlated output but its not exactly the same output
- Creating correlation output is challenging so some tools have defects in them especially for large simulations

RiskHive (Arrisca) 0%, 50%, 100% correlation output

- Note 50% correlation curve roughly mid-distance between 0 & 100 % curves



Workbook uses Linear interpolation

- Most Monte Carlo tools allow different amounts of correlation per item
- Workbook only allows the same amount of correlation throughout
- Workbook uses linear interpolation on 0% & 100% correlation figures to obtain other correlation % figures which is reasonable
- For example, type 30% into the green box

Linear Interpolation	
Reqd Corr. %	30%
1% point =	566
10% point =	591
Mean =	650
70% point =	674
90% point =	712
99% point =	760

- Monte Carlo tools give slightly more accurate results

Underestimation Issues

- For Beta (PERT) the min & max values should cover 99.7% of occurrences
- That is 997 out of 1,000 occurrences
- How many estimators realise this?

- Studies have shown estimated 90% pt is actually the 70% pt
- This is often due to the estimators being too optimistic
- Often the max has been seriously underestimated
- Min value might be realistic because often all 3 pts were underestimated

- Estimators start with the most likely & then estimate min & max
- Studies have shown that estimators appear to be “anchored” to their first estimated value i.e. the most likely value
- Anchoring is a well known phenomena

Poorly estimated 3 points

- Max values are often seriously underestimated
- Most likely is often underestimated
- If it was poorly estimated as:

	Min	Most Likely	Max	Mean	σ
Item 1	15	20	25	20.0	1.7

and it should have been estimated as:

	Min	Most Likely	Max	Mean	σ
Item 1	15	23	40	24.5	4.2

- Note the difference in the Mean and σ values
- The huge difference in σ will radically change the probability values

- Last estimate has more typical 2:1 ratio for (max - ml : ml – min = 17:8)

Be careful when using percentiles

- To allow for underestimation some tools allow different min & max values
- E.g. @Risk RiskTrigen(15, 20, 35,10, 90) specifies a triangular distribution with a 10th percentile value of 15, a most likely value of 20 and a 90th percentile value of 35
- Above settings might not help because the issue is the max not the min
- So RiskTrigen(15, 20, 35, 2, 80) might be more appropriate

Estimation Accuracy

- Max to min range is meant to cover 99.7% of the situations i.e. 6 σ
- It takes a lot of practice to estimate 3 pts with that accurately
- The workbook allows the range too be changed e.g. 4 σ = 67.7%
- But range change should only be used as an interim solution

Software Estimation: Perfect Practice Makes Perfect

- Read the June 2002 Crosstalk article by David Henry with the above title
- Can be found at http://www.crosstalkonline.org/back-issues/
- Estimators used 3-pts & Beta (PERT) workbook to determine the Mean
- Their workbook did not include probabilities or correlation
- The developers were not use to estimating
- Weekly estimates on development work of small tasks (i.e. 3 days or less) on effort not size
- To improve estimation accuracy estimators compared estimates with actuals

David Henry results

- For 6 months most estimators (inc. subject matter experts) significantly underestimated the max
- At the start the difference between estimate & actuals averaged 75%
- After 3 months down to 35%
- After 6 months of weekly practice was within 20% of actuals
- Most useful workbook feature was estimation history i.e. estimates, actuals
- Estimators need this history & estimation practice to improve their accuracy

Workbook monitor sheet

Feed-back information:	
Completed activities mean =	1,717
Completed activities actual =	1,962
Total Difference (compared to Expected) =	14%

Task Name	Min	Most Likely	Max	Individual Mean	Estimated items completed	Actual	% Difference	Comments
Item 1	500	1000	2100	1100.0	1100.0	1221.0	11%	Expert Judgement
Item 2	300	350	450	358.3	358.3	376.0	5%	Analogy estimate
Item 3	200	250	350	258.3	258.3	365.0	41%	Expert Judgement

- Input estimates and actuals
- Also comment on each estimate

Conclusion

- Software & systems estimates should use 3 pt estimation
- Deterministic workbook very accurate for non-correlated work
- Workbook gives reasonable estimates for correlated work
- If possible use a Monte Carlo tool
- Must use Monte Carlo for schedule estimates
- Not all Monte Carlo tools give same output values for correlated work
- Some Monte Carlo tools do not always simulate correlation correctly
- Suggest a default of 25% to 30% correlation for software development
- Need an estimation history feedback loop to become good at estimating

Any Questions

Contact Details

Steve Webb

SW Estimation Ltd
www.swestimation.co.uk

Email: stevewebb@swestimation.co.uk or steve.a.webb@talk21.com

Mobile: 077 181 283 99