

估算是一个范围，不是一个数

唐工：你估计要完成开发用户登录模块要多少天？

小李：三天。

唐工：能在三天完成的可能性有多高？

小李：可能性很高。

唐工：可否量化一点？

小李：估计 50%-60%。

唐工：所以很有机会不止三天，要四天了。

小李：对的，其实也有可能甚至要五、六天，但我估计机会不大。

唐工：你信心有多少？

小李：难说，有 95% 的信心可以在六天之内完成。

唐工：所以有可能甚至要用上七天了？

小李：这样说吧，如果所有可能出问题的都出了问题，甚至会 10 或 11 天，但这种概率很低。

(最终管理者唐工还是要求有一个承诺，而不是一个估算)

所以唐工再问小李：是能否给我一个确实能完成这个模块的日期？

小李：正如我前面说，很可能三天可以完成，但有可能四天。

唐工追问：你可以说四天吗？

小李：也有可能五六天。

唐工结束对话：OK，请你尽力六天之内完成这个模块。

唐工貌似请求，但实际是要求小李承诺这个模块要在六天之内开发完。假如这个模块的开发时间超过六天，唐工就有依据说小李没有尽力导致延误了。

所以从以上对话，可以看到作为开发专业人员，必须分清估算和承诺。作为专业人士，我们不应该给一些没有把握的承诺，误导对方。中国老话说一诺千金就是这个道理。

从单点到 3 点估算

从上面的例子可以看到，一般的单点估算是很容易被误导，以为那个天数是有把握达成的，所以我们最好从单点估算变成三点估算，除了估算最可能的天数，还有最佳和最差共三点。但项目是由一系列的任务组成（如第二任务依赖于第一个任务的完成），如何计算所有任务的总天数？下面用例子说明如何用 3 种使用 3 点估算估计的方法 (A,B,C) 估算总天数：

A) 假定都是正态分布，用模型估计：

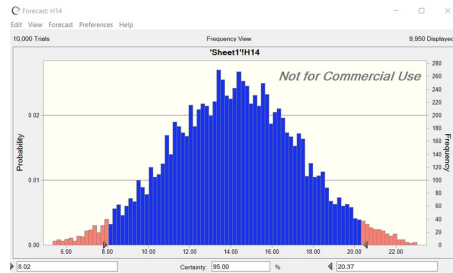
先用 PERT 方程式计算每一步的预计值与标准差：

:: 预计值 Expected Value EV = (Best + 4xMost Likely + Worst) / 6

:: 标准差 Sigma = (Worst - Best) / 6

Step	Best 最佳	Most Likely 最可能	Worst 最差	EV	Sigma
1	1	3	12	4.167	1.833
2	1	1.5	14	3.5	2.167
3	3	6.25	11	6.5	1.333
		10.75		14.168	

如果假定是正态分布，按以上预计值和标准差，使用蒙地卡罗模拟，从下图可看到，95% 范围是 8.02 ~ 20.37



B) 直接用 PERT 方程式计算总天数的均值与标准差:

如不用模拟, 直接把 3 步的均值加起来

$$::4.2 + 3.5 + 3.6 = 14$$

计算 3 步总方差: (方差 = Sigma^2)

假定: 总方差 = 每步方差的总和

总方差 = 9.77

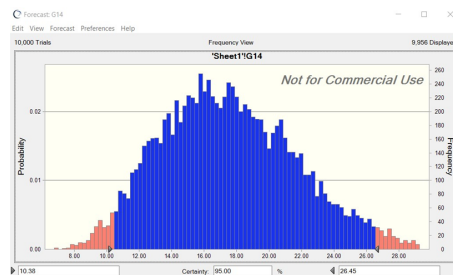
Sigma $\sigma = 3.13$

95% 范围计算公式为: 均值的总和 $\pm 2\sigma = (4.2 + 3.5 + 3.6) \pm 2 \times 3.13 = 14 \pm 6.26 = 7.74 \sim 20.26$

结果与蒙地卡罗模拟预测类似。

C) 假定都是三角形分布, 用模型估计:

如果用三角形分布, 95% 范围是 10.38 ~ 26.45



总结 + 解读分析结果

- 如果假定每一步的分布都是一个正态分布, 就可以用头两个方程式计算每一步的平均值跟标准差和方差, 用方程式可计算 3 步的总均值大概是 14。也可以用方程式计算标准差, 总的 sigma (标准差) 是 3.13 左右。
- 也可用蒙地卡罗模型 (假定步骤都是正态分布), 得出很类似的正态分布, 总的平均也接近 14, 95% 的范围是从 8.02 到 20.37, 接近上面算出的均值 \pm 两个标准差数值。
- 但因三个步骤都是明显往右偏, 所以不能假设它们是正态分布, 更合适的是使用三角形分布, 然后用蒙地卡罗估算 “加” 起来的分布, 看见最后一个图明显是类似往右有个尾巴, 能更正确反应三个步骤加起来的的天数的估计分布。

- 跟假定正态分布的结果比较，很明显看到用三角形分布结果往右偏，上限是 26.45（比正态分布的 20.37 高）。不是正态分布的话，左面就没有长尾巴，所以就会比本来正态分布的下限高，下限是 10.38（比正态分布的 8 高）。
- 从这简单例子看到，如果我们要把三点估算加起来，尤其是非正态分布的话，就不能用简单的方程式，或者假定它是正态分布来计算，需要用蒙地卡罗模型假设三角形分布才能反应总体的分布。

从这三个偏左分布步骤例子看起来好像有些偏差，但不是很严重。如果我们看见用十个步骤都是偏一边分布，总分布会如何？是否相差会更远？

利用蒙地卡罗模拟 10 个步骤（三角形分布）的总分布

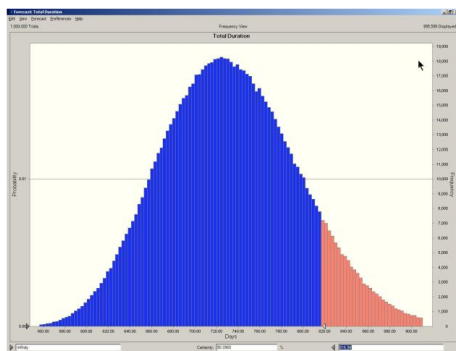
如果每步都估算天数，10 个步骤的总天数就是 500 天（把 10 个估算值加起来）。

但如果每个步骤都是三点估算：

Process 过程	Durations 天数		
step 步骤	Best 最短	Most Likely 最可能	Worst 最长
1	27	30	75
2	45	50	125
3	72	80	200
4	45	50	125
5	81	90	225
6	23	25	63
7	32	35	88
8	41	45	113
9	63	70	175
10	23	25	63
		500	

很明显看到每一步都是偏左的分布，所以可预计总天数应不止 500 天，但估多少才合适？

假定每步骤是三角形分布，用模型估计，重复五千次，得出下面分布：



Process 过程	Durations 天数				
step 步骤	Best 最佳	Most Likely 最可能	Worst 最差	Expected	Sigma
1	27	30	75	37	8
2	45	50	125	61.66	13.33
3	72	80	200	98.66	21.33
4	45	50	125	61.66	13.33
5	81	90	225	111	24
6	23	25	63	31	6.66
7	32	35	88	43.33	9.33
8	41	45	113	55.66	12
9	63	70	175	86.33	18.66
10	23	25	63	31	6.66
		500		617.33	

A) 用 PERT 方程式计算每一步的预计值与标准差:

Step	Best 最佳	Most Likely 最可能	Worst 最差	Expected	Sigma
1	27	30	75	37	8
2	45	50	125	61.666667	13.333333
3	72	80	200	98.666667	21.333333
4	45	50	125	61.666667	13.333333
5	81	90	225	111	24
6	23	25	63	31	6.666667
7	32	35	88	43.333333	9.333333
8	41	45	113	55.666667	12
9	63	70	175	86.333333	18.666667
10	23	25	63	31	6.666667
		500		617.333333	

假定是正态分布，按以上预计值和标准差，使用蒙地卡罗模拟，得出的总分布与上面用三角形分布的结果几乎一致，都是左右平均分布的正态形。

分析 10 步骤模拟结果

- 为什么用三角形模拟出来不是偏左的分布（类似前面 3 步结果），而是一个正态分布

以上实验验证了“中心极限定理”，无论本来是什么形状分布，如果随机抽样够多，样本的平均值分布接近正态分布。所以如果本来只是三个步骤的时候还是可以看出是三角形偏左，但到了用十个步骤相加的时候，就跟每个都是用正态分布去估算的结果没有什么区分。

(中心极限定理会在后面数据分析里用上，例如通过画控制图判断过程是否稳定)

- 实验结果也验证了 PERT 三点估算法公式，无论任何分布都可以用 PERT 公式计算每一步的预计值和标准差，然后计算总结果的分布（不需要蒙地卡罗模拟），当步骤越多差异就越小，如果是像上面的例子，只是希望求十个步骤的总分布（无论每步本身是怎样分布），都不需要用模拟，用 PERT 公式计算便可。

(这不表示蒙地卡罗模型没有用，到后面根因分析部分，我们还是会继续用它来比较不同的搭配选择最优)

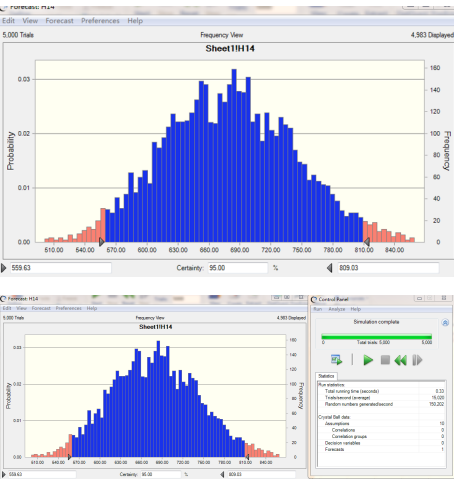
附件

蒙地卡罗 (Monte Carlo) 模拟

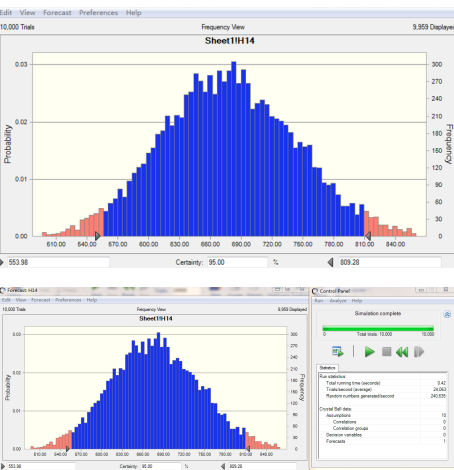
当结果不能用数学公式计算的时候（例如是三角形分布），可以用电脑随机模拟结果。例如：

- 计算三个步骤的总共人天，每个步骤都是三角形分布，我们就用电脑的随机功能模拟，让随机功能的结果按三角形分布：
- 第一次模拟：步骤 1 得出 1.3，步骤 2 得出 1.2，步骤 3 得出 2.0，得出三个步骤的总工期是 4.5 人天。
- 第二次。。。
- 如果我们模拟 1000 次，10000 次就有，便能模拟出总分布

5000 次



10000 次



-----<< END >>=====