 版权所有 不准抄袭翻印

第 5 章 概率分布

第 106 页 第 6 行

参数是确定值，在本章及下一章是已知值，在第 7 章以 后是未知值。

第 113 页

5.1.8 负二项分布

(1) 概率分布函数 $P(x)=\binom{x+k-1}{x}p^k(1-p)^x$, $x=0,1,2,\cdots$

第 113 页

5.2 连续型随机变量的概率分布

连续型随机变量的概率密度函数概率分布的关联性，请见：Chap16 总复习补充.pdf。

5.3 正态分布概率的计算

例题 5.4: $P(0\le Z\le 2)=0.5-P(Z\le 2)=0.5-0.023=0.477$

例题 5.5: $P(-1.34\le Z\le 1.68)=1-P(Z\ge 1.34)-P(Z\ge 1.68)=1-0.090-0.046=0.864$

例题 5.6: $P(1.34\le Z\le 2.38)=P(Z\ge 1.34)-P(Z\ge 2.38)=0.090-0.009=0.081$

例题 5.7: $P(-2.14\le Z\le -1.17)=P(Z\ge 1.17)-P(Z\ge 2.14)=0.121-0.016=0.105$

例题 5.8: $X\sim N(50,25)$， 求 $P(40\le X\le 65)$

解答： $P(40\le X\le 65)=P\left(\frac{40-50}{5}\le\frac{X-\mu}{\sigma}\le\frac{65-50}{5}\right)=P(-2\le Z\le 3)=0.976$

例题 5.9: $X\sim N(120,400)$， 求 $P(X\ge 135)$

解答： $P(X\ge 135)=P\left(\frac{X-\mu}{\sigma}\ge\frac{135-120}{20}\right)=P(Z\ge 0.75)=0.227$

例题 5.10: 大学生性向测验(GAT)的数学部份 X 是正态分布其均值 500，标准差 100，$X\sim N(500,100^2)$。求(1). 如果一个学生考 600 分，他应该在全部学生排名前百分之几？ (2). 第 80 个百分位(percentile) P_{80} 是多少？换句话说，如果一个学生要考到前面 20%，才能入学，那么他要考多少分？

解答：(1). $P(X\ge 600)=P(Z\ge\frac{600-500}{100})=P(Z\ge 1)=0.159$

(2). $P(Z\ge z)=0.2$, $z=0.84$, $\frac{P_{80}-500}{100}=0.84$, $P_{80}=584$

5.3.4　转换成标准正态分布求概率

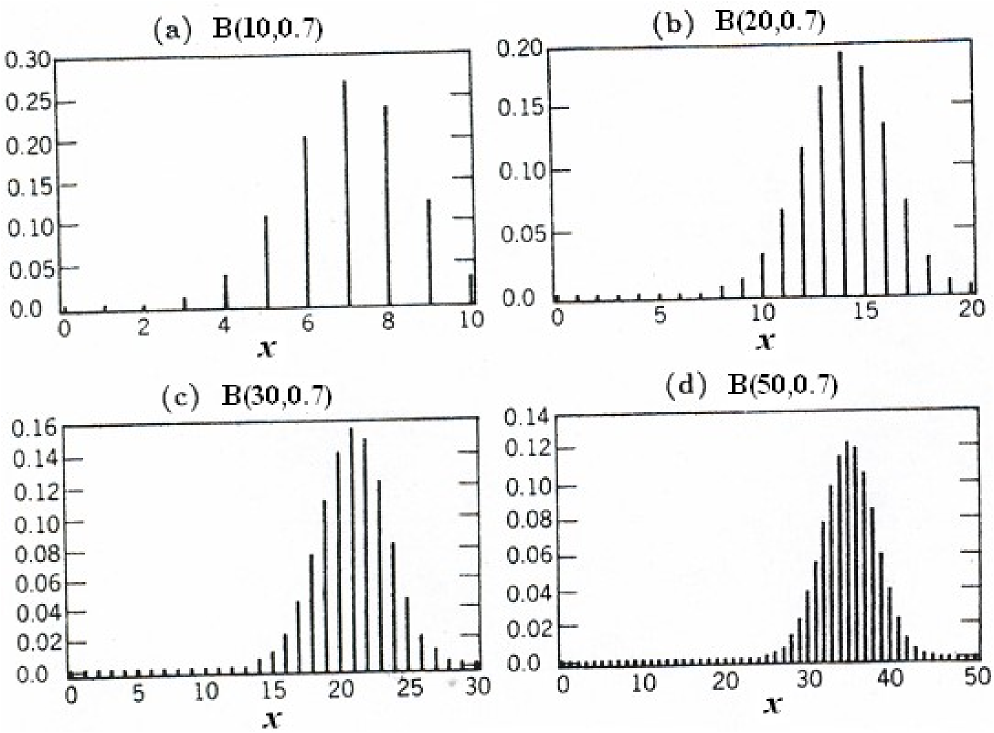


图 5.a　　二项分布近似正态分布　　注意：(a) $n \times (1-p) = 10 \times (0.3) = 3 < 5$

例题 5.11： 市民赞成某法案的比例是 70%。现在有 500 个市民，其中至少有 370 个市民赞成的概率是多少？

解答：$X \sim B(500, 0.7),\ \ np = 500(0.7) = 350 > 5,\ \ n(1-p) = 500(0.3) = 150 > 5$

$$Y \sim N(\mu, \sigma^2),\ \ \mu = np = 350,\ \ \sigma = \sqrt{np(1-p)} = \sqrt{350(0.3)} = 10.25$$

$$P(X \geq 370) \cong P(Y \geq 369.5) = P(Z \geq \frac{369.5 - 350}{10.25}) = P(Z \geq 1.90) = 0.0287$$

例题 5.12： $X \sim B(15, 0.6)$，求 $P(X=8) + P(X=9)$

1. 利用二项分布表。　2. 利用二项分布公式。　3. 利用正态分布。

解答：1. $P(X=8) + P(X=9) = 0.177 + 0.207 = 0.384$

2. $P(X=8) + P(X=9) = \binom{15}{8}(0.6)^8(0.4)^7 + \binom{15}{9}(0.6)^9(0.4)^6 = 0.177 + 0.207 = 0.384$

3. $\mu = np = (15)(0.6) = 9,\ \ \sigma = \sqrt{9(0.4)} = 1.897$

$$P(7.5 \leq Y \leq 9.5) = P(-0.79 \leq Z \leq 0.26) = 0.3878$$

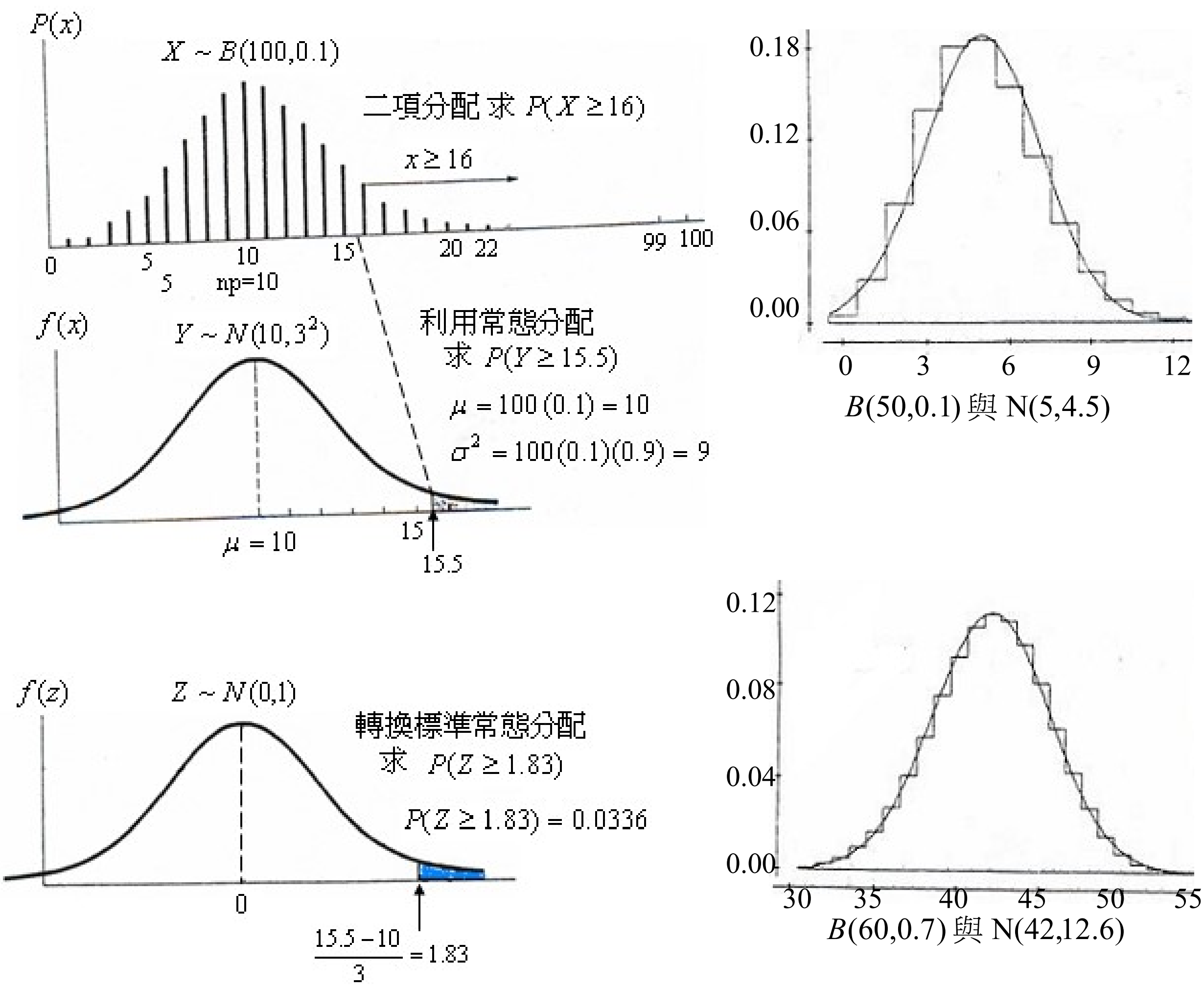


图 5.b 二项分布经过「连续性修正」，利用正态分布计算概率

5.3.5 正态分数

正态分布在统计检验扮演很重要的角色，许多检验方法(有母数检验)，需要总体数据服从正态分布(假设条件)，所以有时需要检查或检验样本数据是否符合正态分布，检验正态分布有一些方法，如卡方适合度检验、无母数 Lilliefors 检验等。

如果数据有 n 个，正态分数(normal score)是将正态概率分成 n+1 等分，再和原始数据画 XY 散布图，如果形成一直线，则该数据是正态分布的数据。其计算步骤如下：

1. 如果有 n 个数据 c，已经照大小次序，由小到大排列。

2. 令 $\alpha_i = \dfrac{i}{n+1}$，则 x_i 的正态分数是 $z_{1-\sigma_i}$，即 Z 是标准正态分布 $Z \sim N(0,1)$，

   令 $\Phi(x) = P(Z \le x)$, $\Phi^{-1}(p) = z_{1-p}$，所以 $P(Z \le z_{1-\alpha_i}) = \alpha_i$，$z_{1-\alpha_i} = \Phi^{-1}(\alpha_i)$

3. 如果 x_i 与 $z_{1-\sigma_i}$ 画出二度空间散布图是一「直线」，则 x_i 是正态分布的资料。

 版权所有 不准抄袭翻印

图 5.c 是 5 个资料与 10 个数据的正态分数，这些正态分数也可以用在第 14 章，卡方适合度检验样本数据是否正态分布。

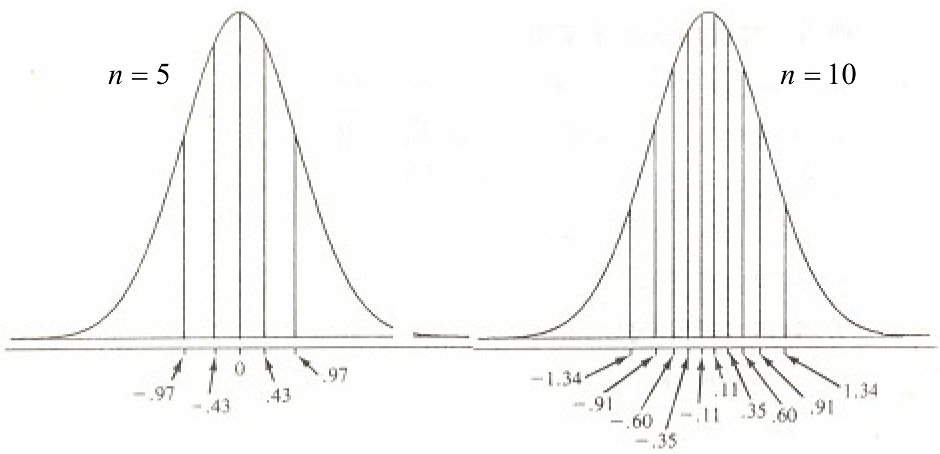
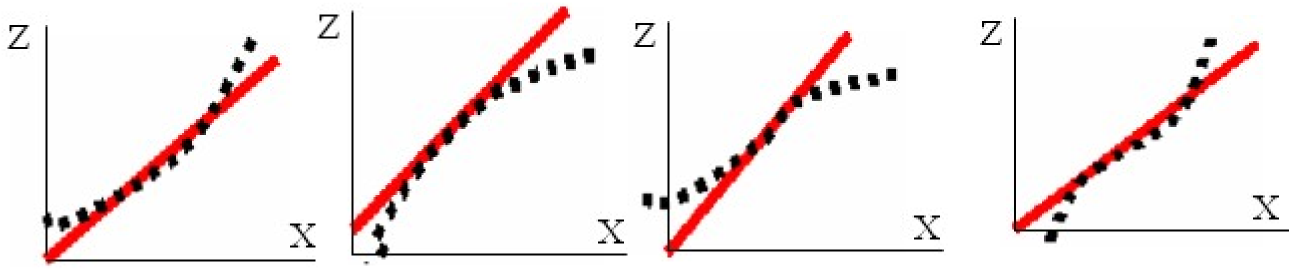


图 5.c 正态分数

正态概率图为一直线，则为正态。下列情况，可判定偏度及峰度。



左偏型　右偏型　高狭峰　低阔峰

图 5.d 正态分数与概率分布型态

5.3.6 正态分布的判断

在统计学的「有母数」统计，假设条件是：样本数据来自「正态」总体。但是，如何判定样本数据是否正态，有下列方法：

1. 经验法则；
   (1) 叙述统计的偏度和峰度系数
   (2) 直方图或多边形图判断其形态，注意：选择不同的组距，直方图形态不同
   (3) 枝叶图与盒须图
   (4) 样本均值加减 1、2、3 个标准差，概率应该是：68%、95%、99.7%
   (5) 四分位距 IQR ÷ S 标准差 ≈ 近似 1.3
   (6) 正态分数、正态概率图为一直线
2. 检验法则
   (1) 卡方适合度检验(第 12 章)
   (2) Lilliefors 检验(第 13 章)
   (3) Kolmogorov-Smirnov 检验
   (4) Shapiro–Wilk 检验或 Shapiro–Francia 检验
   (5) Jarque-Bera (Skewness-Kurtosis) 检验

(6) 其他检验方法

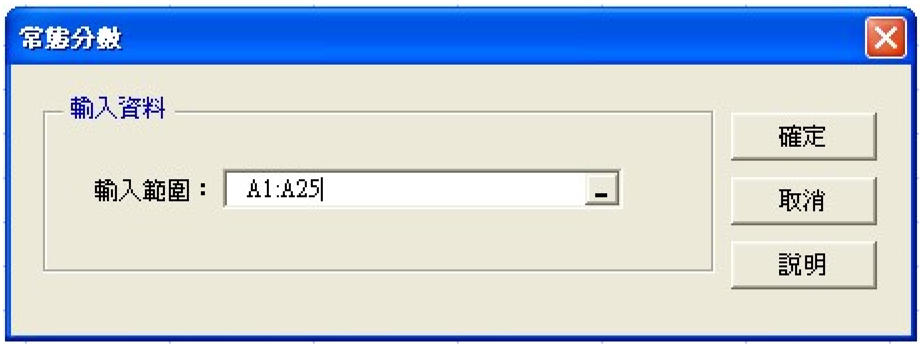
5.3.7 概率分布的偏度与峰度

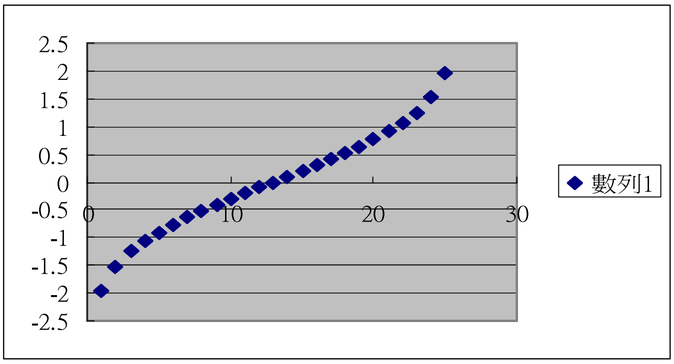
偏度 / 峰度	对称型	右偏型	左偏型
正态峰	正态	二项、Beta	二项、Beta
尖峰	二项、超几何 t 分配、Beta	二项、超几何、几何 负二项、卜松、指数 卡方、F 分配 对数常态、Beta	二项、超几何 Beta
平峰	二项、超几何、均匀 Beta、三角形	二项、超几何 Beta、三角形	二项、超几何 Beta、三角形

Beta 分布各种形态都有，因此适用连续型一般分布(如项目管理的三段时间估计 PERT 分析)。

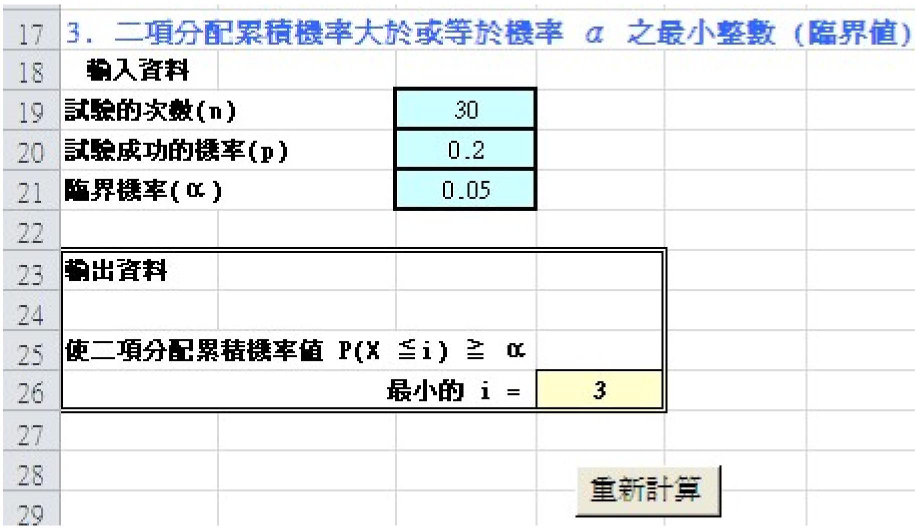
5.4 中文统计应用

3. 计算正态分数，选择「概率分布」→「正态分数」，出现正态分数窗体，以例题 2.1 的数据输入，得以正态分数与原始数据的散布图，这个图有些弯曲，没有完全一直线，只是中间数据接近正态。





4. 选择「概率分布」 → 「离散型概率分布」，出现模块工具的事算表清单。选择需要的概率分布，输入查询数据，可以得到：概率值、累积概率。



习题

1. 某一种疾病的死亡率为 40%，而在某一医院里，患此种疾病的最后十位病人中，有九位痊愈了，你认为这个结果和已知的死亡率一致吗？或者应该调查一下是否这个医院使用了新的治疗方法？
2. 可乐公司在 6 罐可乐，有一个免费再来一罐的拉环，如果要有 95% 的概率，得到免费一罐的拉环，要买多少罐可乐？
3. 三个人在餐厅吃完饭各掷硬币，出现不同面者付账。若三人掷出相同面，则再掷一次。计算：(1). 最多各掷三次即可决定付账者之概率；(2). 平均掷多少次可决定付账者。
4. 一项是非题以投掷铜板方式来作答，若正面则答是，若反面则答否。是非题共有一百题， 则至少对六十题的概率为多少？
5. 一个骰子投掷 180 次，记录出现的点数，则下列概率为何？
   (1). 六点出现二十次的概率？
   (2). 奇数点出现四十次以下的概率？
   (3). 三的倍数点出现至少五十次的概率？
6. 我在 10:00a.m. 到车站，知道公交车到达时间为 10:00a.m. 到 10:30a.m. 的连续均匀分布。
   (1). 公交车在 10:10a.m. 以后到达之概率？
   (2). 公交车在 10:20a.m. 以前到达之概率？
   (3). 我等到 10:15a.m. 还没有公交车，则至少再等 10 分钟以上之概率？
7. 根据「今日美国」的报导，接受大学教育的学生，1986 年平均每年费用为 5137，若在这些大学生中有 55%，每年费用在 5000 以上。如果费用的分布为正态分布，则其概率分布的标准差为多少？
8. 记忆测验的分数为正态分布，其标准差为 11.62 ，若分数在 70.34 以上的部份占 69.5 %，此项记忆测验分数的均值为多少？
9. 一台装填果汁的机器可设定装填均值 44.80 到 48.20 盎斯，不管均值多少，其标准差均为 0.08 盎斯，厂商标示的净重为 46 盎斯以上。
   (1). 若厂商希望每 10,000 罐中，少于 46 盎斯的罐头不超过 100 罐，则均值应设多少？
   (2). 若厂商希望每 100,000 罐中，少于 46 盎斯的罐头不超过 100 罐，则均值应设多少？

 版权所有 不准抄袭翻印

10. 某间高中的毕业生平均 60% 继续接受大学教育，今年毕业班学生共 150 人，若以过去资料为标准，继续上大学的毕业生少于 75 位的概率为多少？多于 100 位的概率为多少？ 刚好 100 位的概率为多少？
11. 一位爱好园艺者决定在他的车道两旁种植牵牛花，他买了一包混合的牵牛花种子，根据标签指出，五分之一的种子会长成粉红色的牵牛花，五分之四会长成红白相间的牵牛花。若从这个比例混合的总体中随机抽样，并且适用二项式定理，当二百个种子发芽时，长成粉红色的牵牛花至少 30 朵以上而小于 50 朵的概率为多少？
12. 一家公司希望对一项价格较高但包装造型较吸引顾客的产品进行市场测试活动，若购买产品的顾客中有 30% 选择高价格的包装，则测试结果视为成功。现在将两种包装的产品同时陈列在店中显眼的地方，并仔细记录不同包装的销售次数，假设在 200 次销售中有 53 次是购买高价格的包装：
    (1). 若实际比例为 30%，则对于这个结果出现的概率为多少？
    (2). 若顾客购买高价格包装的比例为 20%，则对于这个结果出现的概率为多少？
13. 一位生物学家需要一种消化系统不良的特殊蚯蚓样本至少十只以上，这种特殊蚯蚓在全数蚯蚓中占有 20% 比例。由于收集时不能立刻就测试出其消化系统特征不良，如果生物学家想迅速收集到足够样本，一次收集比需要量多的话会比较省时，因为收集过少可能需要再收集一次，其固定成本较高。因为 50 只的 20% 为 10 只，她的学生建议带回 50 只。
    (1). 若她接受学生建议的话，则消化系统不良的蚯蚓收集数量可能不足的概率为多少？
    (2). 若生物学家希望数量不足的概率少于 0.01，那么她应带回多少只蚯蚓？
14. 将同一副扑克牌的四张 A 放进一帽子中，任意取出一张，观察其花色，然后放回帽子里。假设每一次的抽出之间互相独立，欲达到至少出现一次黑桃 A 的概率大于 0.8 的结果，必须抽取多少次？
15. 每分钟打进来的电话是泊松分布，平均是 5 个电话。计算从现在算起，第 2 通电话在 1 分钟内打进来的概率。
16. 工厂生产电灯泡，其生产过程，经实验确定不良率为 0.03，现在于生产在线抽验 200 支，若发现 6 支不良品即停工检修，计算：
    (1). 不良品数目之概率密度函数
    (2). 停工检修的概率
17. 某人平均每五发射击可命中三发，问该人须射几发，才使至少命中一发之概率大于 0.999？
18. 袋中有 3 红球，7 白球，从袋中任取一球再放回袋中，求在 8 次的取出中，恰有 3 次为红球的概率
    (1). 利用二项分布计算概率
    (2). 利用泊松分布计算概率
    (3). 利用正态分布计算概率
19. 有一批 10 个产品，已知这批产品其中有 2 个有缺点。选取 2 个检验，若发现 1 个或 1 个以上之产品有缺点，则整批被销毁，否则全部被接受。如果每个产品成本为 65 元，可售得 120 元，问制造商获利的期望值为何？
20. 设一批 300 个电子零件中有 5% 是不良品，现由其中抽样 5 个加以试验，试求至少有 2 个不良品的概率。
21. 掷铜币 1000 次，其出现正面次数在 440 次与 500 次间之概率如何？
22. 设有二个独立变量随机变量 X, Y。令 $W = X + Y$，试依以下的条件，求算 W 分布的期望值 $E(W)$，

 版权所有 不准抄袭翻印

方差 V(W) 及随机变数 W = 1 时的概率 P(W=1)：

(1). X 及 Y 分布皆为二项分布，且 $X \sim B(6, 0.2), Y \sim B(4, 0.2)$

(2). X 及 Y 分布皆为正态分布，且 $X \sim N(27,16), Y \sim N(23,9)$

(3). X 及 Y 分布皆为泊松分布，且 $X \sim Pois(1.8), Y \sim Pois(1.2)$

23. 某公司生产一种工产品，包装方式分别有一箱 20 个装，40 个装与 100 个装。为了质量保证，每箱产品在销售之前，必定抽验 5% 的产品，抽验结果如果发现不良品，则整箱再做全部检验以换掉所有不良品，不然就封箱准备销售，问：

(1). 现有 20 个装的产品一箱，其中有 2 个不良品，抽验通过的概率为多少？

(2). 现有 40 个装的产品一箱，它含 4 个不良品，抽验通过的概率为多少？

(3). 现有 100 个装的产品一箱，它含 10 个不良品，抽验通过的概率为多少？

24. 一位保险推销员根据以往他推销保险的经验，发现每天推销出去的保险单数呈泊松分布，平均每天可以推销出去 1.2 张保险单，请问：

(1). 一天中他至少可以推销出去 3 张保险的概率为多少？

(2). 一天中他推销出去的保险单至少 1 张且至多 4 张的概率为多少？

(3). 每天推销出去的保险单数的标准差为多少？

25. 假定在台湾地区有 8% 的人经常服用维他命，为了进一步了解这些人经常服用维他命的动机，有一个维他命公司要做一次市场调查，在此地区随意抽出 1000 个人中，设 X 为抽出的 1000 人中经常服用维他命的人数，请问：

(1). X 的适当概率分布为什么？

(2). X 的概率分布是否会近似泊松分布？

(3). X 的概率分布是否会近似正态分布？

(4). 用适当概率分布或适当的近似概率分布，求抽出 1000 人中至少有 100 个人经常服用维他命的概率？

(5). 抽出的 1000 人，平均有多少人经服用维他命？标准差多少人？

26. 大学生的身高已知为正态分布，平均为 168 公分，标准差为 6 公分。随机抽样两个学生，试问两个学生都至少为 172 公分的概率有多少？

27. 某位统计老师想出一份选择题测验学生，每题有五个答案，单选。若此老师想使学生随机乱猜的答案，「答对题数超过半数」的概率为 10%，试求应出几题？

28. 船到港口的概率是泊松分布，平均每两天有一艘船，从现在起四天内没有船到达的概率是多少？

29. 若根据调查 70%走北二高的车辆，时速都会超过 90 公里的限制，假设你想在北二高做车辆时速的研究，随机抽样 10 辆车，定义随机变量 X 为时速超过 90 公里限制的汽车数，请根据这个随机抽样回答以下的问题。

(1) 请写出随机变数 X 的概率分布。

(2) 请问 $P(X=10)$ 与 $P(4<X<9)$ 分别为何？

(3) 请问 $P(X=2)$ 与 $P(3 \leqq X \leqq 6)$ 分别为何？

(4) 请找到 X 之期望值。

(5) 请找到 X 之标准差。

30. 若根据调查 5%的大学在学生抽烟，假设你想研究大学生抽烟与否跟健康的关系，随机抽样 20 位大学生，定义随机变量 X 为抽烟的大学生人数，请根据这个随机抽样回答以下的问题。
 (1) 请写出随机变数 X 的概率分布。
 (2) 请问 20 位大学生中刚好两位抽烟的概率？
 (3) 请问 20 位大学生中至少两位抽烟的概率？
 (4) 请问 20 位大学生中不到两位抽烟的概率？
 (5) 请问 20 位大学生中最多两位抽烟的概率？
 (6) 请问 20 位大学生中多于两位抽烟的概率？
 (7) 若将随机抽样数增加为 100 位大学生，请问 X 之期望值为何？
 (8) 若将随机抽样数增加为 100 位大学生，请问 X 之标准差为何？
31. 若平均每小时有 20 通电话打到市立美术馆的总机，假设打到市立美术馆总机的电话为随机且独立的，定义随机变量 X 为每小时打电话到市立美术馆的电话数，请根据这个随机抽样回答以下的问题。
 (1) 请写出随机变数 X 的概率分布。
 (2) 请问 5 分钟内有两通电话打到市立美术馆的总机的概率？
 (3) 请问 20 分钟内有八通电话打到市立美术馆的总机的概率？
 (4) 请问 4 分钟内没有任何电话打到市立美术馆的总机的概率？
32. 假设每天下午 2:00 到 5:00 到达某加油站加油的汽车数为一泊松分布，若平均有 10 辆车到达此加油站加油，定义随机变量 X 为每天下午 2:00 到 5:00 到达此加油站加油的汽车数，请根据这个随机抽样回答以下的问题。
 (1) 请写出随机变数 X 的概率分布。
 (2) 请问每天下午 2:00 到 5:00 至少有 12 辆车到达此加油站加油的概率？
 (3) 请问每天下午 2:30 到 4:00 至少有 6 辆车到达此加油站加油的概率？
 (4) 请问每天下午 2:00 到 3:00 正好有 3 辆车到达此加油站加油的概率？
 (5) 请问每天下午 2:00 到 5:00 有 3 辆车至 12 辆车到达此加油站加油的概率？
33. 假设到某大型综合医院急诊室就诊的病人，等待医生看诊的时间是一均匀分布，最少等 40 分钟最多等 3 小时，定义随机变量 X 为到此大型综合医院急诊室就诊病人等待医生看诊的时间，请根据这个随机变量回答以下的问题。
 (1) 请写出随机变量 X 的概率密度函数。
 (2) 请问随机变量 X 介于 1 小时到 2 小时间的概率？
 (3) 请问到此大型综合医院急诊室就诊病人等待医生看诊正好 1 小时的概率？
 (4) 请问到此大型综合医院急诊室就诊病人等待医生看诊不超过 1 小时的概率？
 (5) 请问到此大型综合医院急诊室就诊病人等待医生看诊超过 2 小时的概率？
34. 假设到某大型百货公司的顾客，等待电梯上下楼的时间是一均匀分布，最少等 1 分钟最多等 6 分钟，定义随机变量 X 为到此百货公司的顾客等待电梯上下楼的时间，请根据这个随机变量回答以下的问题。
 (1) 请写出随机变量 X 的概率密度函数。
 (2) 请问到此百货公司的顾客等待电梯上下楼的时间介于 1 分钟到 5 分钟的概率？
 (3) 请问到此百货公司的顾客等待电梯上下楼的时间不超过 3 分钟的概率？
 (4) 请问到此百货公司的顾客等待电梯上下楼的时间只需 1 分半钟的概率？

 版权所有 不准抄袭翻印

(5) 请问到此百货公司的顾客等待电梯上下楼的时间中位数？

35. 假设某省电灯泡的寿命是一正态分布，平均寿命为 3600 小时，标准差为 300 小时，定义随机变量 X 为到此省电灯泡的寿命(以小时计)，请根据这个随机变量回答以下的问题。
    (1) 请问到此省电灯泡至少会用超过 4000 小时的概率？
    (2) 请问到此省电灯泡不会用超过 3500 小时的概率？
    (3) 请问到此省电灯泡的寿命介于 3800 小时至 4100 小时的概率？
    (4) 若此省电灯泡的制造商想做广告推销此产品，请问应该用多少小时的寿命来广告只有 2%灯泡会烧坏的概率？
36. 假设大学资管系刚毕业的社会新鲜人的起薪是一正态分布，平均起薪每个月为 NT$37500，标准差为 NT$8500，定义随机变量 X 为大学资管系刚毕业的社会新鲜人的起薪(每个月)，请根据这个随机变量回答以下的问题。
    (1) 请问大学资管系刚毕业的社会新鲜人每个月起薪介于$40000 至$50000 的概率？
    (2) 请问大学资管系刚毕业的社会新鲜人每个月起薪超过$52000 的概率？
    (3) 请问大学资管系刚毕业的社会新鲜人每个月起薪少于$35000 的概率？
    (4) 请问 5%最低的大学资管系刚毕业的社会新鲜人每个月起薪为何？
    (5) 请问 3%最高的大学资管系刚毕业的社会新鲜人每个月起薪为何？
37. 某县一家二手车经销商的销售员平均约 4 天可以卖掉一辆车，假设此二手车经销商的销售员卖车的间隔时间是一指数分布，请回答以下的问题。
    (1) 请问此二手车经销商的销售员今天卖掉车之后，再等 6 天才又卖车的概率？
    (2) 请问此二手车经销商的销售员今天卖掉车之后，再等 3 到 8 天才又卖车的概率？
    (3) 请问此二手车经销商的销售员今天卖掉车之后，不到 2 天才又卖车的概率？
    (4) 请问二手车经销商的销售员卖车的间隔时间的方差为何？
38. 假设到达某得来快餐餐厅的车辆数为一泊松分布，平均每分钟有 0.5 辆车到达此得来快餐餐厅，请回答以下的问题。
    (1) 若定义随机变量 X 为此得来快餐餐厅到达车辆的间隔时间，请写出随机变量 X 的概率分布。
    (2) 请问若现在正好有一车辆到达此得来快餐餐厅，要再等 1 分半钟才有下一辆车到达的概率？
    (3) 请问若现在正好有一车辆到达此得来快餐餐厅，要再等 3 分钟才有下一辆车到达的概率？
    (4) 请问此得来快餐餐厅每分钟有 1 辆车到达的概率？
    (5) 请问此得来快餐餐厅每小时有 40 辆车到达的概率？