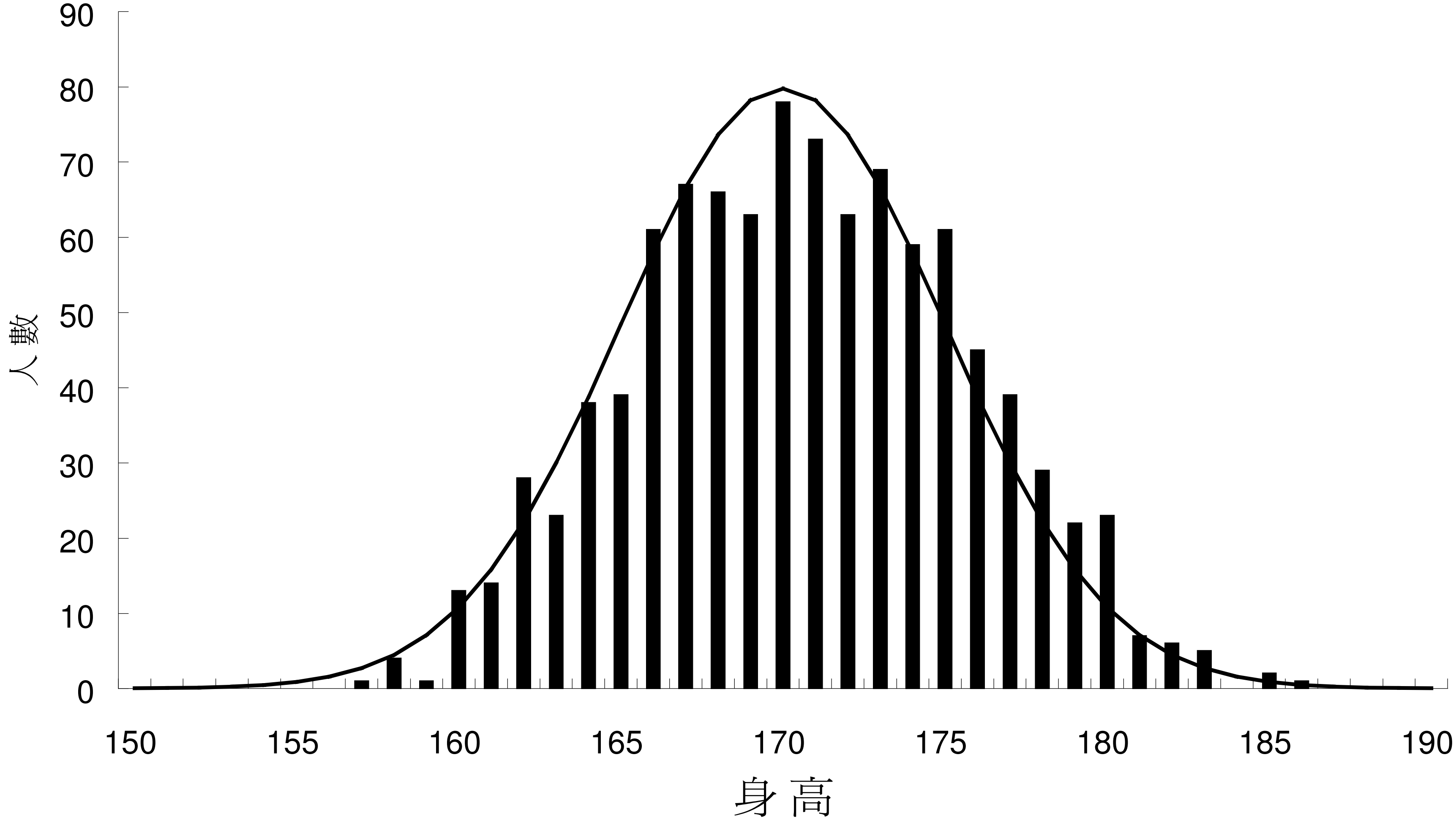
常態（高斯）分佈

- 常態分佈
  - 調查台灣 1000 名成年男性的身高，發現身高特別高或特別矮的人數很少，大部份的人數集中在中間 (e.g. 170 cm) 左右
  - 這 1000 人的平均身高為 170cm ，則離 170cm 越遠的人所佔的比例越少，畫成分析圖會發現，分佈圖大略以 170 為中心，往兩旁遞減，如下圖：

常態（高斯）分佈 (Cont.)



90
80
70
60
50
40
30
20
10
0
人數
150
155
160
165
170
175
180
185
190
身高

常態（高斯）分佈 (Cont.)

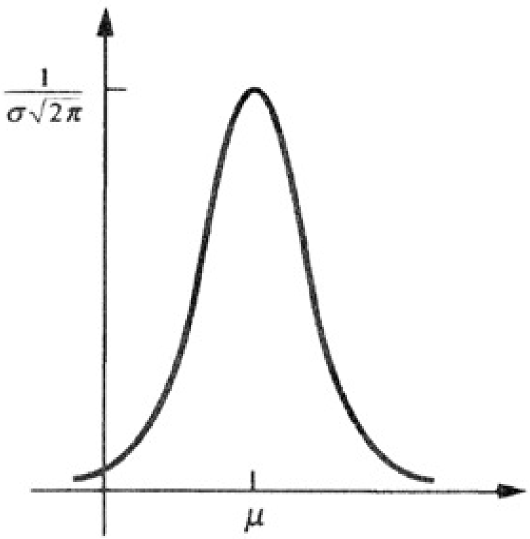
- 常態分佈的特性
  - 平滑線的左右是對稱的，像一座山或一口『銅鐘』
  - 由中間往兩邊遞減
  - 由於是左右對稱，因此最高處是為平均數、眾數也是中位數
  - 一般而言，當樣本數很大時曲線會很接近常態分佈
  - 德國天文學家 Carl Friedrich Gauss, 1777~1855 將該曲線應用於天文觀察，故又稱之為 『高斯分佈』
  - 著名的數學家及統計學家 Karl Pearson (1857~1936) 將高斯分佈稱之為常態分佈

常態(高斯)分佈 (Cont.)

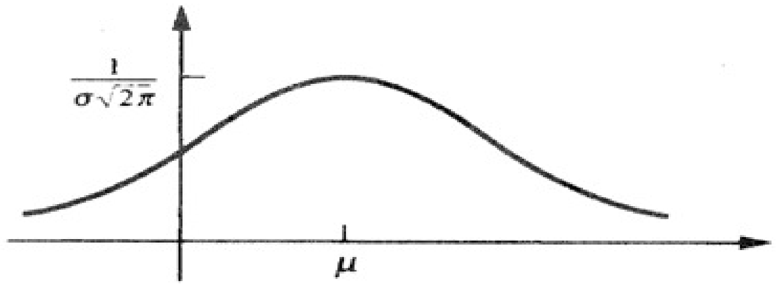
- 常態密度
  - 令 $\sigma > 0$
  - 若 X 具有密度 $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty$
  - 則說 X 是具有參數 μ 與 σ^2 的常態分佈
  - **常態**分佈又可稱為**高斯**分佈
- 符號說明
  - μ ：平均值 (mean value)
  - σ^2 ：變異數 (variance)
  - σ ：標準差 (standard deviatioin)

常態 (高斯) 分佈 (Cont.)

- 下方的兩張分佈圖說明了不同的 σ^2值呈現的分佈圖形也不一樣
  - 變異數 (variance) → 量測觀測值離平均數差異



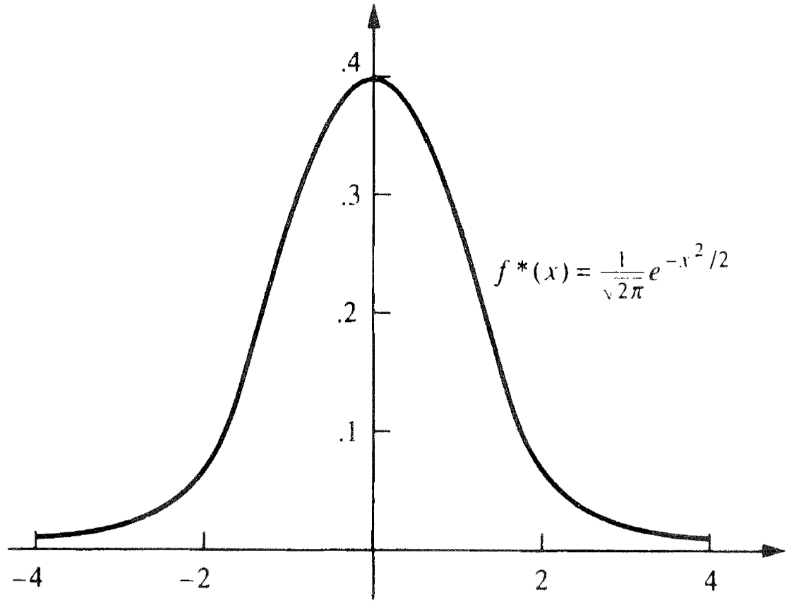
較小的 σ^2



較大的 σ^2

常態（高斯）分佈 (Cont.)

- 單位常態
  - 當常態分佈的 $\mu=0, \sigma^2=1$ 時稱之為單位或標準常態分佈
  - 一般使用 X^* 表示標準常態分佈下的隨機變數
  - f^* 表示為標準常態分佈下的密度函數
  - F^* 表示為標準常態分佈下的分佈函數
  - 下表列示部份標準常態分佈函數 $F^*(x)$ 的值



常態（高斯）分佈 (Cont.)

表 1　單位常態分佈函數

x	$F^*(x)$	x	$F^*(x)$	x	$F^*(x)$	x	$F^*(x)$
−3.0	0.001	−1.5	0.067	0.1	0.540	1.64	0.950
−2.9	0.002	−1.4	0.081	0.2	0.579	1.7	0.955
−2.8	0.003	−1.3	0.087	0.3	0.618	1.8	0.964
−2.7	0.004	−1.28	0.100	0.4	0.655	1.9	0.971
−2.6	0.005	−1.2	0.115	0.5	0.691	1.96	0.975
−2.5	0.006	−1.1	0.136	0.6	0.726	2.0	0.977
−2.4	0.008	−1.0	0.159	0.7	0.758	2.1	0.982
−2.3	0.011	−0.9	0.184	0.8	0.788	2.2	0.986
−2.2	0.014	−0.8	0.212	0.9	0.816	2.3	0.989
−2.1	0.018	−0.7	0.242	1.0	0.841	2.33	0.990
−2.0	0.023	−0.6	0.274	1.1	0.864	2.4	0.992
−1.96	0.025	−0.5	0.309	1.2	0.885	2.5	0.994
−1.9	0.029	−0.4	0.345	1.28	0.900	2.6	0.995
−1.8	0.036	−0.3	0.382	1.3	0.903	2.7	0.996
−1.7	0.045	−0.2	0.421	1.4	0.919	2.8	0.997
−1.64	0.050	−0.1	0.460	1.5	0.933	2.9	0.998
−1.6	0.055	0	0.500	1.6	0.945	3.0	0.999

常態 (高斯) 分佈 (Cont.)

- 有了標準 (單位) 常態分佈便可借用上表來計算出其他各式各樣的標準常態分佈值
- 例如： $P(X^* \geq 0.3) = 1 - P(X^* \leq 0.3) = 1 - F^*(0.3) = 1 - 0.618 = 0.382$
- 另一種作法 (使用對稱性)

$$P(X^* \geq 0.3) = P(X^* \leq -0.3) = F^*(-0.3) = 0.382$$

- 另外一個例子

$$P(|X^*| \leq 2) = P(-2 \leq X^* \leq 2) = F^*(2) - F^*(-2) = 0.954$$

常態（高斯）分佈 (Cont.)

- 新隨機變數 $aX+b$
  - 若 X 是平均值 μ 與變異數 σ^2 的常態分佈
  - 則

$X+b$ 是具有平均值 $\mu+b$ 與變異數 σ^2 的常態

aX 是具有平均值 $a\mu$ 與變異數 $a^2\sigma^2$ 的常態

$aX+b$ 是具有平均值 $a\mu+b$ 與變異數 $a^2\sigma^2$ 的常態

由非單位常態轉變為單位常態

- 設 X 為具有參數 μ 與 σ^2 的常態
  - 則 $(X-\mu)/\sigma$ 的平均值為 0
  - 變異數為 1
  - 意即 $(X-\mu)/\sigma$ 是單位常態分佈 X^*
- 因為 $(X-\mu)/\sigma$ 是 $aX+b$ where $a=1/\sigma, b=-\mu/\sigma$ 的特例
- 函數 $aX+b$

$$\text{mean}=a\mu+b=\frac{1}{\sigma}\mu-\frac{\mu}{b}=0$$

$$\text{variance}=a^2\sigma^2=(\frac{1}{\sigma})^2\sigma^2=1$$

由非單位常態轉變為單位常態 (Cont.)

- 例：身高的隨機變數 X 是常態分佈，且 μ=66 吋及 σ^2=9
- 隨機從人群中挑一人，其身高不足 6 呎的機率為何 ?(請使用查表)
  - 你要計算的機是 X<72
  - 對不等式兩邊同時減去平均值並被標準差除，可得到新的式子為 $\dfrac{X-\mu}{\sigma}<\dfrac{72-\mu}{\sigma}$

$$P(X<72)=P\left(\frac{X-66}{3}<\frac{72-66}{3}\right)$$
$$=P(X^*<2)$$
$$=F^*(2)$$
$$=0.977$$

度量衡換算
1 吋 (inches) = 2.540 公分 (cm)
1 英呎 (feet) = 12 吋 = 30.48 公分 (cm)
66 吋 = 66*2.540 = 167.64 公分 (cm)
6 英呎 = 6* 12 吋 = 72 吋 = 6*30.48 公分 = 182.88 公分

由非單位常態轉變為單位常態 (Cont.)

- 續上例，若你能找出 b 值，則可以 95% 確認你隨機挑的人是從平均值附近且與 b 相關的範圍中找到的，如何找 b?

$$P(66-b \le X \le 66+b) = 0.95$$

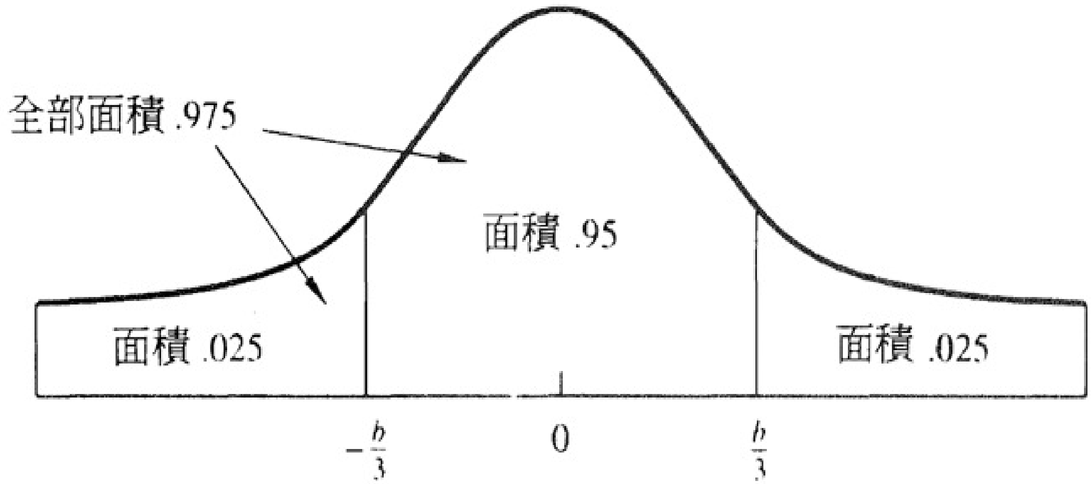
$$P(-\frac{b}{3} \le \frac{X-66}{3} \le \frac{b}{3}) = 0.95$$

$$P(-\frac{b}{3} \le X^* \le \frac{b}{3})$$

$P(X^* \le \frac{b}{3}) = 0.975$ (ps. 利用對稱性)

$\frac{b}{3} = 1.96$ 查表

$b = 5.88$



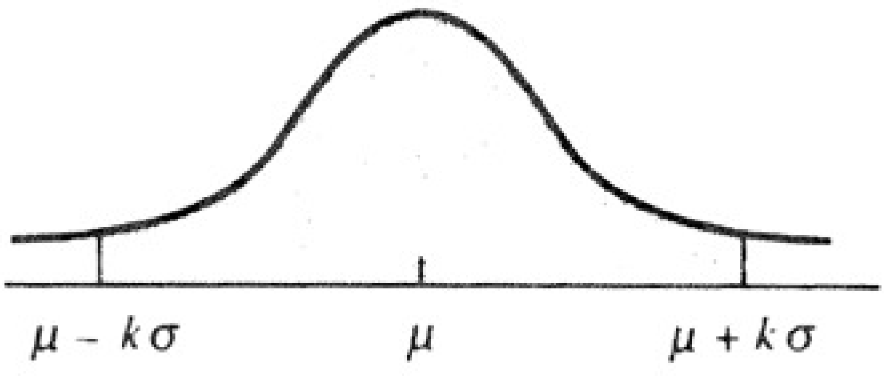
常態隨機變數在平均值的 k 個標準差的機率

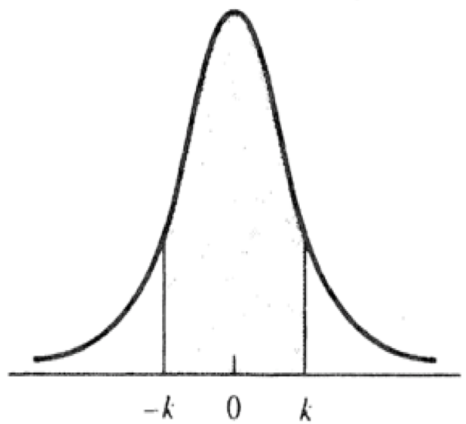
- 設 X 為具有參數 μ 與 σ^2 的常態
- 如果你進行實驗得到 X 在平均值 k 個標準差，則你有

$$P(\mu - k\sigma \le X \le \mu + k\sigma) = P\left(-k \le \frac{X-\mu}{\sigma} \le k\right)$$

$$P(\mu - k\sigma \le X \le \mu + k\sigma) = P(-k \le X^* \le k)$$

- 換言之，對所有常態分佈而言，X 是在平均值的 k 個標準差內的機率都是相同的，值得一提的是它介於 $-k$ 與 k 之間的單位常態 X^* 的機率相同



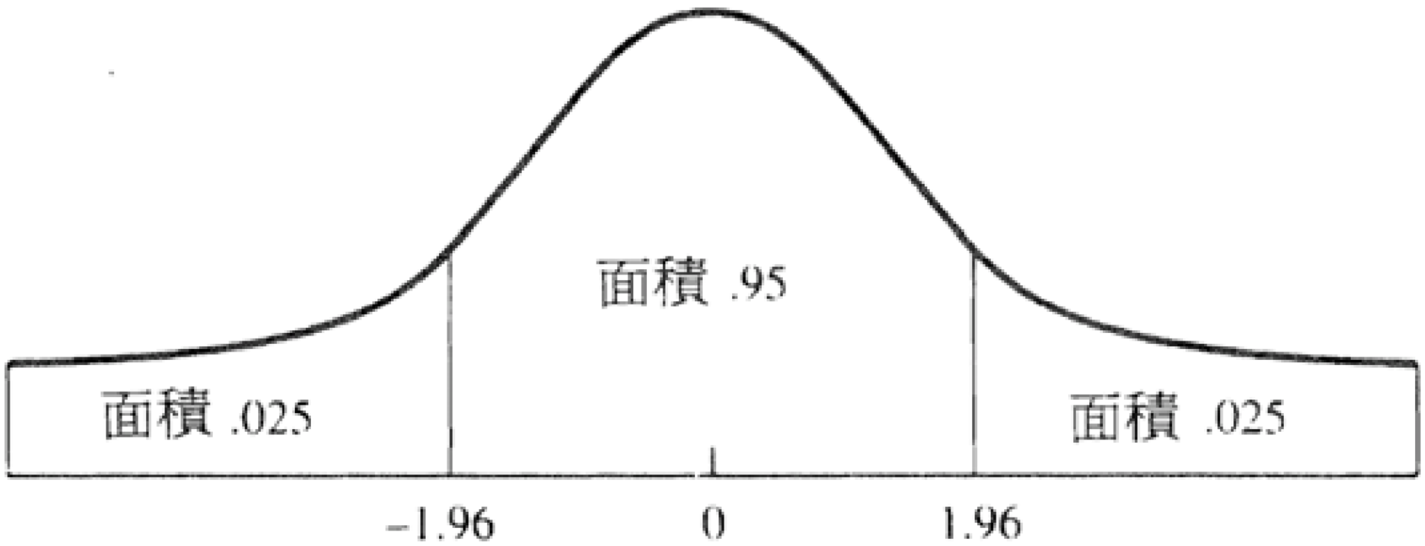


單位常態

常態隨機變數在平均值的 k 個標準差的機率 (Cont.)

- 標準差 → 距離單位
- 距離量測 → 常用來量測與平均值 μ 的距離，比較少用來量測標準常態分佈下與 0 的距離
- 例子：單位常態

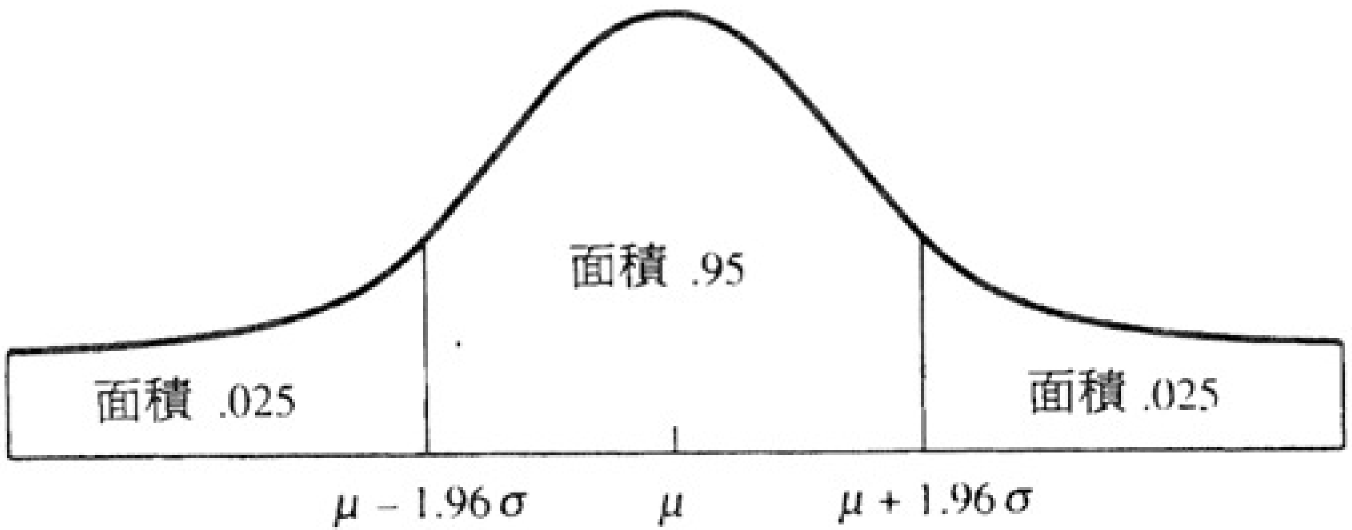
$$P(-1.96 \le X^* \le 1.96) = F^*(1.96) - F^*(-1.96) = 0.95$$



常態隨機變數在平均值的 k 個標準差的機率 (Cont.)

- 由上例得知，具參數 μ 與 σ^2 的常態密度，95% 的面積介於 $\mu x-1.96\sigma$ 與 $\mu x+1.96\sigma$ 之間，亦即對任一常態隨機變數 X 都有同樣的情況

$$P(\mu x-1.96\sigma x\leq X\leq\mu x+1.96\sigma x)=0.95$$



常態隨機變數在平均值的 k 個標準差的機率 (Cont.)

- 同理 $P(-2.6 \leq X^* \leq 2.6) = F^*(2.6) - F^*(-2.6) = 0.995 - 0.005 = 0.99$
- 同樣可整理成 $P(\mu x - 2.6\sigma x \leq X \leq \mu x + 2.6\sigma x) = 0.99$
- 意思是，任一常態分佈的隨機變數有 99% 的機率會介於平均值的 2.6 個標準差內

常態近似於二項式

- 令 X 是 n 次百努利試驗裡的成功次數，且 $P($ 成功 $)=p$ ，即 X 是離散隨機變數裡的二項式分佈
- 首先，利用圖式說明 X 的機率函數與鐘形常態密度之相似性
- 令 $n=100$ ， $p=½$ ， X 是 0, 1, 2, ..., 100 中的任意值

$$P(X=50)=\binom{100}{50}(\frac{1}{2})^{100}=0.0796$$

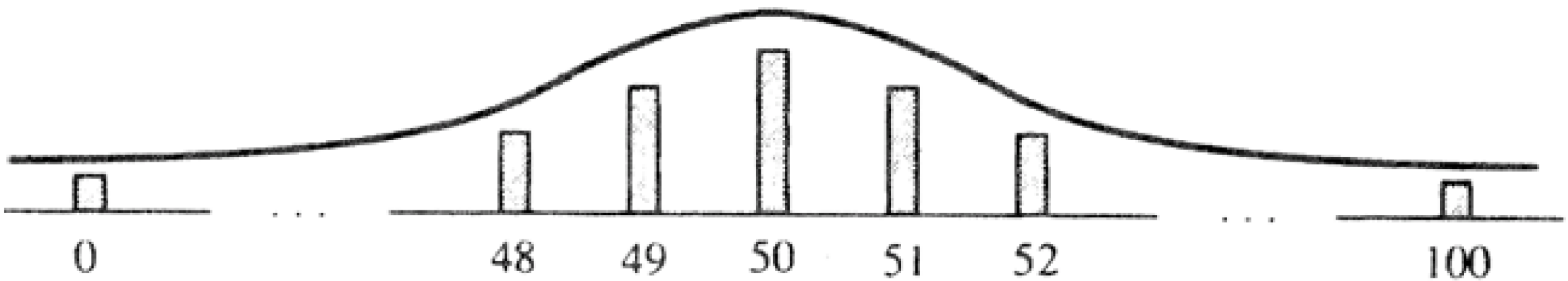
$$P(X=49)=P(X=51)=\binom{100}{49}(\frac{1}{2})^{100}=0.078$$

常態近似於二項式 (Cont.)

$$P(X=48)=P(X=52)=\binom{100}{48}\left(\frac{1}{2}\right)^{100}=0.074$$

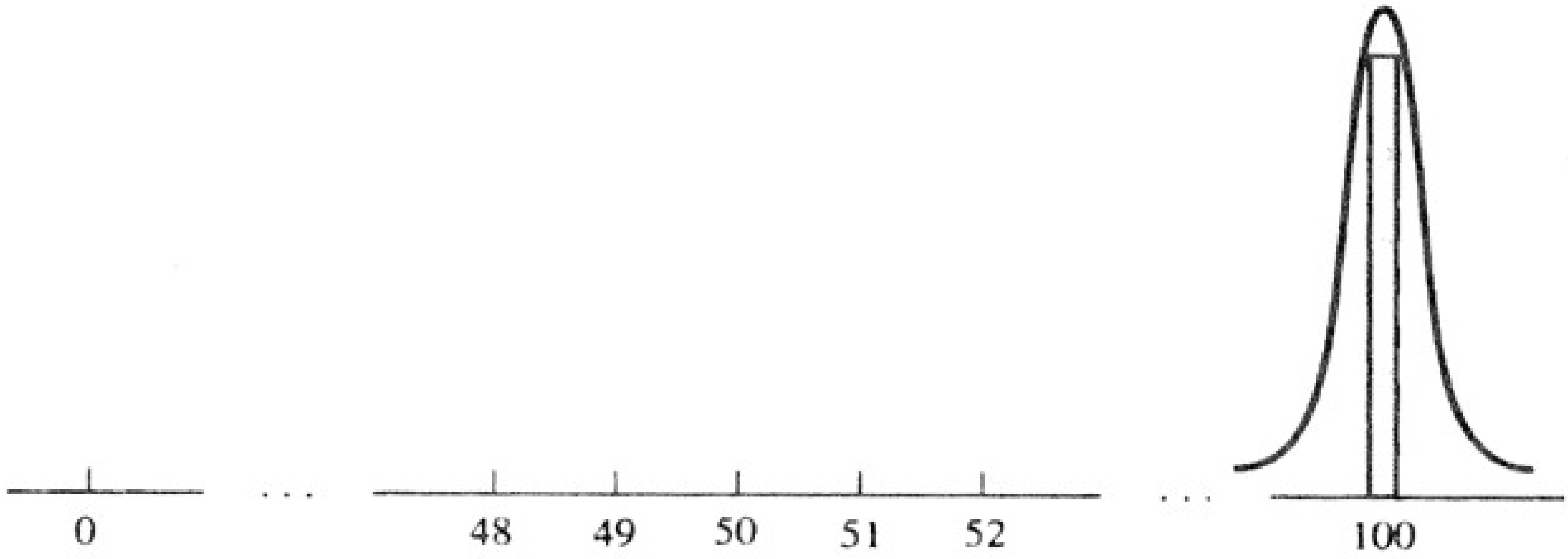
$$P(X=0)=P(X=100)=\left(\frac{1}{2}\right)^{100}=7.9\times 10^{-31}$$

- 將上述機率結果繪成圖形(見下圖)，二項式隨機變數的平均是 $np = 100 * ½ = 50$ ，再者，平均值兩邊的分佈是對稱，整體看來是常態分佈



常態近似於二項式 (Cont.)

- 另一個極端的例子，假設 $n = 100$ ，$p = 1$ ，則所有的機率將集中於 $X = 100$ 的情況，其分佈圖如下 (形狀亦很像常態分佈)



常態近似於二項式 (Cont.)

- 常態分佈與二項式之間的關聯
  - 常態分佈被用來做近似二項式分佈
  - 若 x 有一具參數 n 與 p 的二項式分佈，當 n 很大時，X 會很近似具平均數 np ，及變異數 npq 的常態分佈

$$\mu = np \text{ and } \sigma^2 = npq$$

- 例子：一歪斜銅板 P(人頭) = 0.3 ，丟擲銅板 1000 次，則丟到人頭期望值為 300 ，試求人頭數等於或多於 400 的機率
  - 設 x 是丟擲銅板出現人頭的次數，求 P(X ≧ 400)

常態近似於二項式 (Cont.)

- 方法一：
  - X 是具有參數 n=1000 且 p=0.3 的二項式分佈
  - 所以

$$P(X \ge 400) = P(X = 400) + P(X = 401) + P(X = 402) + \ldots + P(X = 1000)$$

$$= \binom{1000}{400}(0.3)^{400}(0.7)^{600} + \binom{1000}{401}(0.3)^{401}(0.7)^{599} + \ldots + (0.3)^{1000}$$

- 方法二：
  - X 是一個具有參數 $\mu = np = 300, \sigma^2 = npq = 210$ 近似常態分佈
  - $P(X \ge 400) = P\left(\frac{X - 300}{\sqrt{210}} \ge \frac{400 - 300}{\sqrt{210}}\right)$

$$= P(X^* \ge 6.9) \qquad \left(\frac{X - \mu}{\sigma} \text{ is the unit normal}\right)$$

$$= 1 - P(X^* \le 6.9)$$

$$= 0$$

常態近似於二項式 (Cont.)

- 方法三：
  - X是具有 $\mu 300, \sigma^2 = 210 =, \sigma \sim 14.5$ 的近似常態。400的值不同於平均值300加上100，而其標準差為6.9。由方程式 (7) 我們可知當標準差是2.6時，對來說只有1% 的機會是遠離平均值。所以從平均值遠離標準差6.9時，的機率差不多是0。