

특정분포에 따른 확률 Plot들의 정규성과 Bimodality 비교

이 제 영¹⁾, 이 성 원²⁾

요 약

특정분포에 대한 P-P plot과 Q-Q plot의 특징을 분석하고 두 가지 그래프 사이의 관계를 비교해 보았다. NTV의 본질적인 특징을 알아보고 각 특정분포에 대한 NTV plot의 특징을 분석했다. 전형적인 bimodal 분포가 나타날 때 P-P 혹은 Q-Q plot은 뚜렷한 break-point를 갖는다는 것을 알아보고, 단지 skewed된 분포나 skewed된 것처럼 보이는 bimodal 분포 사이에서 발생하는 판단의 어려움에 대하여 여러 plot고찰해 보았다.

1. 서 론

실험에서 얻어진 표본이 '정규분포를 따르는가 또는 따르지 않는가'하는 문제는 주어진 특정결론을 이끌어내는데 있어서 매우 중요하다. 그래서 분포의 정규성에 관한 고찰은 계속 연구되어왔다. 이와 같은 분포의 정규성을 검정하는 방법으로는 Kolmogorov-Smirnov 검정, Shapiro와 Wilk (1965)의 W 검정, Dyer (1974)의 Anderson과 Darling A_n^2 검정 등이 있다. 그 외에도 Shapiro와 Francia (1972) 그리고 Weisberg와 Bingham (1975) 등에 의해서 계속 연구되어 왔으며, 더 나아가 D'Agostino (1971)는 대표본에 대한 검정통계량으로 D를 제안했었고, Spiegelhalter (1977)는 베이지안 S 검정통계량을 제안해서 extreme 분포들에 대한 적합성을 밝혔었다. 또한, Lin과 Mudholker (1980)는 Z-test를 기초로 한 검정을 제안하기도 하였고, Looney (1995)는 다중정규성(multivariate normality)에 대한 연구를 하였었다. 이와 같이 대수적인 방법에 의한 통계량으로 분포의 정규성을 검정하는 방법 외에도, 그래프에 의하여 시각적으로 분포의 정규성을 검정하는 방법이 계속 연구되어 왔다. 한편, 의학실험에서는 bimodal 분포에 대한 문제에도 많은 관심을 갖게 되어서 Jackson et al. (1989a, b), Endrenyi과 Patel (1991a, b), 그리고 Lee et al. (1996) 등에 의해 토의되었다. 본 논문에서는 정규성 검정에 대한 대표적인 그래프인 P-P plot과 Q-Q plot, 그리고 최근에 발표된 NTV(Normal Test Variable) plot을 중심으로 각 plot의 특징(2절)과 특정한 분포에 따라 이들 plot들이 갖는 특성을 논의해 보고자 한다. 이를 위하여 확률표본으로부터 generated된 표본들(n=50, 100, 200)을 사용해서 plot들의 특징을 서로 비교해 보고자 한다(3절). 더 나아가 정규분포에 의해 생성된 표본의 조합으로 이루어진 bimodal 분포에 대한 특성도 함께 조사해 볼 것이다(4절).

1) (712-749) 경상북도 경산시 대동 214-1 영남대학교 통계학과 조교수

2) (712-749) 경상북도 경산시 대동 214-1 영남대학교 통계학과

2. 정규 확률 PLOT

2.1 P-P Plot

확률표본 X_1, \dots, X_n 의 순서통계량을 $X_{1:n}, \dots, X_{n:n}$ 이라 하고, 이 확률변수 X 의 누적분포함수(CDF)를 $\Phi(x)$ 라고 하자. 그리고 $U_i = \Phi(X_i)$ 로 두자. 그러면 확률적분변환(Probability Integral Transformation)에 의하여 U_i 는 $\text{UNIF}(0, 1)$ 을 따르게 된다. 그리고 U_1, \dots, U_n 의 순서통계량을 $U_{1:n}, \dots, U_{n:n}$ 로 두자. 그러면, $U_{i:n} = \Phi(X_{i:n})$ 가 된다. 이때 순서통계량 $U_{i:n}$ 의 확률밀도함수 g_i 는 아래와 같이 유도된다.

$$\begin{aligned} g_i(u_{i:n}) &= \frac{n!}{(i-1)!(n-i)!} \left[\int_{-\infty}^{u_{i:n}} f(x) dx \right]^{i-1} f(u_{i:n}) \left[\int_{u_{i:n}}^{\infty} f(x) dx \right]^{n-i} \\ &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} (u_{i:n})^{i-1} (1-u_{i:n})^{n-i} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} (u_{i:n})^{\alpha-1} (1-u_{i:n})^{\beta-1} \end{aligned}$$

단, 여기서 $f(x)$ 는 $\text{UNIF}(0, 1)$ 분포의 확률밀도함수이다. 그리고 $\alpha=i, \beta=n-i+1, 0 \leq u_{i:n} \leq 1$ 이다. 이렇게 유도된 확률밀도함수 g_i 는 BETA 분포의 확률밀도함수와 같다. 그래서, 우리는 $U_{i:n}$ 가 $\text{BETA}(i, n-i+1)$ 을 따르게 된다는 것을 유도할 수 있다. 그리고 표본분포함수를 $\Phi_n(\cdot)$ 라고 하자. 그러면 $\Phi_n(X_{i:n})$ 는 $\frac{i}{n}$ 가 된다. P-P plot은 위의 성질을 이용하여 (x, y) 평면상에,

$$\left(\Phi(X_{i:n}), \Phi_n(X_{i:n}) \right)$$

를 표시하는 그래프로 정의된다. 이때, $\Phi(X_{i:n})$ 는 $\text{BETA}(i, n-i+1)$ 을 따르므로 그 기대값을 구해보면,

$$E[\Phi(X_{i:n})] = \frac{i}{n+1}$$

가 된다. 그래서 위의 그래프는 기울기 $\frac{n+1}{n}$ 인 직선의 형태가 될 것이라고 기대되고 이 직선에서 벗어난 정도를 정규성 검정의 측도로 삼는다.

2.2 Q-Q Plot

확률표본 X_1, \dots, X_n 의 순서통계량을 $X_{1:n}, \dots, X_{n:n}$ 이라 하고, 이 확률변수 X 의 누적분포함수(CDF)를 $\Phi(x)$ 라고 하고, 그 역함수를 $\Phi^{-1}(\cdot)$ 라 두자. Q-Q plot은 (x, y) 평면상에,

$$\left(\text{분포 } \Phi \text{의 } \left(\frac{i}{n} \right) \text{분위수}, \text{ 표본의 } \left(\frac{i}{n} \right) \text{분위수} \right)$$

를 표시하는 그래프로 정의된다. 이 때, 분포 Φ 의 $\left(\frac{i}{n} \right)$ 분위수는 $\Phi^{-1}\left(\frac{i}{n}\right)$ 가 되고, 표본의

$\left(\frac{i}{n}\right)$ 분위수는 $X_{i:n}$ 가 된다. 다시 말해서, Q-Q plot은 (x, y) 평면상에,

$$\left(\Phi^{-1}\left(\frac{i}{n}\right), X_{i:n} \right)$$

를 표시하는 그래프가 된다. 여기서 이 그래프의 성질을 알아보기 위하여 $X_{i:n}$ 의 기대값을 구해 본다. 먼저 $U_i = \Phi(X_i)$ 로 두고, U_1, \dots, U_n 의 순서통계량을 $U_{1:n}, \dots, U_{n:n}$ 로 두자. 그러면, $U_{i:n} = \Phi(X_{i:n})$ 가 된다. 우리는 이미 $U_{i:n}$ 가 BETA($i, n-i+1$) 분포를 따름을 알고 있다. 그래서 $U_{i:n}$ 의 확률밀도함수 $b_i(u_{i:n})$ 는 아래와 같이 모수가 각각 $\alpha=i, \beta=n-i+1$ 인 BETA 분포의 확률밀도함수와 같다.

$$b_i(u_{i:n}) = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} (u_{i:n})^{i-1} (1-u_{i:n})^{n-i}$$

$X_{i:n}$ 의 기대값, $E[X_{i:n}]$ 를 구하기 위해 변수변환기법(transformation technique)을 사용하여, $X_{i:n}$ 의 확률밀도함수 $h_i(x_{i:n})$ 를 구해본다.

$$X_{i:n} = \Phi^{-1}(U_{i:n})$$

$$h_i(x_{i:n}) = b_i(\Phi(x_{i:n})) \cdot |\Phi'(x_{i:n})|$$

이것을 이용하여 $E[X_{i:n}]$ 를 구한다.

$$\begin{aligned} E[X_{i:n}] &= \int_{-\infty}^{\infty} x_{i:n} h_i(x_{i:n}) dx_{i:n} \\ &= \int_{-\infty}^{\infty} x_{i:n} b_i(\Phi(x_{i:n})) \cdot |\Phi'(x_{i:n})| dx_{i:n} \\ &= \int_0^1 \Phi^{-1}(u_{i:n}) b_i(u_{i:n}) du_{i:n} \\ &\cong \int_0^1 \Phi^{-1}\left(\frac{i-c}{n-2c+1}\right) b_i(u_{i:n}) du_{i:n} \\ &= \Phi^{-1}\left(\frac{i-c}{n-2c+1}\right) \int_0^1 b_i(u_{i:n}) du_{i:n} \\ &= \Phi^{-1}\left(\frac{i-c}{n-2c+1}\right) \end{aligned}$$

이때 c 는 $[0,1]$ 에 속하며, 이미 여러 학자들이 몇가지의 값들을 제시했었다. 즉, $X_{i:n}$ 의 기대값은,

$$E[X_{i:n}] \cong \Phi^{-1}\left(\frac{i-c}{n-2c+1}\right)$$

가 된다. 그래서 Q-Q plot에서는 x 좌표에 $\Phi^{-1}\left(\frac{i}{n}\right)$ 대신에 $\Phi^{-1}\left(\frac{i-c}{n-2c+1}\right)$ 를 사용한다. 그러면 그래프는 y 좌표의 기대값과 x 좌표의 값이 같게되어 기울기가 1인 직선이 될 것이 기대된다.

그러나 실제로 정규성을 검정하려고 하는 Q-Q plot에서는 x 좌표에 $\Phi^{-1}\left(\frac{i-c}{n-2c+1}\right)$ 값 대신

에 표준정규분포의 $\left(\frac{i}{n}\right)$ 분위수 값이 들어가기 때문에, 결국 그래프에서 기대되는 직선은 $y = \sigma x + \mu$ 이 된다. 그래서 이 직선에서 벗어나는 정도가 정규성 검정의 측도가 된다. 본 논문에서는 c 를 Blom이 제시한 $3/8$ 을 사용하여 plot을 그려보았다.

2.3 NTV Plot

NTV plot은 Endrenyi와 Patel에 의해 제시된 것으로 정규성 검정뿐만 아니라, bimodality 검정에도 유용하다고 알려져 의학통계에서 많이 사용되고 있다. NTV plot을 구현하는 절차는 다음과 같다. 먼저, 표본 x_1, \dots, x_n 의 중위수 m 과 표준편차 s 를 계산하고, 표본의 순서통계량 $x_{1:n}, \dots, x_{n:n}$ 을 구하여

$$y_{i:n} = \frac{x_{i:n} - m}{s},$$

$$f = \frac{i}{n},$$

$$z_i = f - (1-f)e^{1.6y_{i:n}}$$

를 계산한 후, NTV(Normal Test Variable)를 다음과 같이 정의한다:

$$NTV_i = \begin{cases} -z_i & , f \leq 0.5 \\ \frac{z_i}{e^{1.6y_{i:n}}} & , f > 0.5 \end{cases}$$

이때, NTV plot은 (x, y) 평면상에,

$$(f, NTV_i)$$

를 표시하는 그래프를 말하며, 이론적으로 NTV_i 의 값이 대략 0과 0.02사이의 양수 값을 갖는다면 표본이 정규분포를 따른다고 기대된다. 그래서 NTV_i 의 최저값과 최고값이 정규성 검정의 측도가 된다고 알려져 있으며, 특히 NTV_i 의 값이 -0.03보다 작은 값을 가질 때에는 주어진 분포가 bimodality를 나타낸다고 알려져 있다.

3. 특정분포에 대한 분석

본 3절에서는 정규분포(Normal distribution)를 중심으로 짧은 꼬리(short-tails)를 갖는 균일분포(Uniform distribution), 긴 꼬리(long-tails)를 갖는 코쉬분포(Cauchy distribution), 왼쪽으로 치우친(skewed-left) 지수분포(Exponential distribution), 그리고 오른쪽으로 치우친(skewed-right) 음의 지수분포(Negative Exponential distribution) 등과 같은 특정분포에 대한 확률 plot들의 특징을 비교 분석한다. 이미 많은 사람들이 특정분포에 대한 P-P plot과 특히, Q-Q plot의 특성에 관해서 연구를 했었고 잘 알려져 있다. 본 논문에서는 NTV plot과의 비교를 위하여 이들 두 plot들을 새로운 시각에서 관찰하여 본다. 여기에 사용된 graph들은 각각의 분포에 대하여 난수발생함수를 이

용하여 생성된 크기가 200인 표본에 대한 것이다.

3.1 P-P plot에서의 분석

먼저 정규분포에서는 직선에 잘 적합되어 있음을 보여주고 있다[그림1-a]. 여기서 우리는 분포의 꼬리부분을 정규분포와 비교하여 [표 1]의 현상을 관찰할 수 있었다. 양쪽 꼬리부분이 짧은 균일분포에서는 plot의 형태가 직선에서 벗어나 기울기가 감소하다가 0.5부근을 지나면서 다시 기울기가 증가하는 형태(N자 형태)가 되었고[그림1-c], 양쪽 꼬리부분이 긴 코쉬분포에서도 역시 plot의 형태가 직선에서 벗어났지만 벗어난 형태는 균일분포에서와는 반대로 0에서부터 점점 기울기가 증가하다가 1에 가까워질수록 다시 기울기가 감소하는 형태(S자 형태)가 되었다[그림1-e]. 지수분포는 왼쪽 꼬리부분이 짧고 오른쪽 꼬리부분이 길기 때문에, P-P plot의 왼쪽 부분에서 기울기가 감소하고 오른쪽 부분에도 기울기가 감소하는 현상을 보여주었다[그림1-g]. 그리고 음의 지수분포는 왼쪽 꼬리부분이 길고 오른쪽 꼬리부분이 짧기 때문에, P-P plot의 왼쪽 부분에서 기울기가 증가하고 오른쪽 부분에도 기울기가 증가하는 현상을 보여주었다[그림1-i]. 그래서 어떤 표본의 P-P plot을 보고 그 표본의 정규성뿐만 아니라 분포의 형태도 어느 정도 추측할 수도 있다. 하지만 위와 같은 특정분포가 아닌 일반적인 표본에서는 형태를 구별하기가 쉽지 않다.

3.2 Q-Q plot에서의 분석

P-P plot과 같은 차례로 분석을 하기로 하자. 먼저 정규분포에서의 그래프의 형태는 이론적으로 제시된 직선에 잘 적합되어 있음을 보여주고 있다[그림1-b]. 여기서도 P-P plot에서처럼 plot 형태의 규칙을 발견할 수 있다[표 1]. 그 형태는 P-P plot과는 반대로 나타나고 있다. 균일분포에서의 plot의 형태는 plot의 기울기가 왼쪽에서부터 점점 증가해 오다가 중간을 지나면서 다시 점점 감소해 가는 형태(S자 형태)가 되고[그림1-d], 코쉬분포에서의 plot의 형태는 균일분포에서와는 반대로 plot의 기울기가 점점 감소하다가 중간을 지나면서 다시 증가하는 형태(N자 형태)가 되었다[그림1-f]. 지수분포는 왼쪽 꼬리부분이 짧아서 Q-Q plot의 왼쪽 부분에서는 기울기가 증가하고, 오른쪽 꼬리부분이 길기 때문에 Q-Q plot 오른쪽 부분에서도 기울기가 감소하는 현상을 보여주었다[그림1-h]. 그리고 음의 지수분포는 왼쪽 꼬리부분이 길고 오른쪽 꼬리부분이 짧기 때문에, Q-Q plot의 왼쪽 부분에서 기울기도 감소하고 오른쪽 부분에서의 기울기도 감소하는 현상을 보여주었다[그림1-j]. 그래서 P-P plot에서처럼 어떤 표본의 Q-Q plot을 보고 그 표본의 정규성뿐만 아니라 분포의 형태도 어느 정도 추측할 수도 있다. 하지만 역시 일반적인 표본에서는 형태를 구별하기가 쉽지 않다.

[표 1] P-P plot과 Q-Q plot에 대한 특징

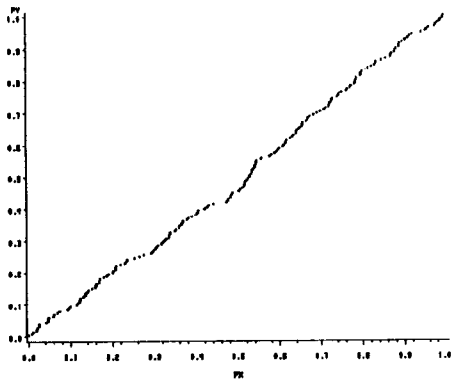
분포 형태	P-P plot	Q-Q plot	분포 형태	P-P plot	Q-Q plot
중심에서 왼쪽 꼬리부분이 정규 분포보다 짧다.	기울기 감소	기울기 증가	중심에서 오른쪽 꼬리부분이 정규 분포보다 짧다.	기울기 증가	기울기 감소
중심에서 왼쪽 꼬리부분이 정규 분포보다 길다.	기울기 증가	기울기 감소	중심에서 오른쪽 꼬리부분이 정규 분포보다 길다.	기울기 감소	기울기 증가

위의 분석에서 보는 바와 같이 P-P plot과 Q-Q plot은 그 형태에 있어서 어느 정도 규칙성을 가지고 있다. 이 규칙들을 서로 비교해 보면 P-P plot과 Q-Q plot은 직선 $y=x$ 에 대해 서로 대칭인 성향을 가진다. 그래서 P-P plot에서 x축과 y축을 서로 바꾸어 표시하면 각 축의 scale만 다를 뿐, P-P plot과 Q-Q plot은 서로 같은 규칙을 가지게 된다[그림1].

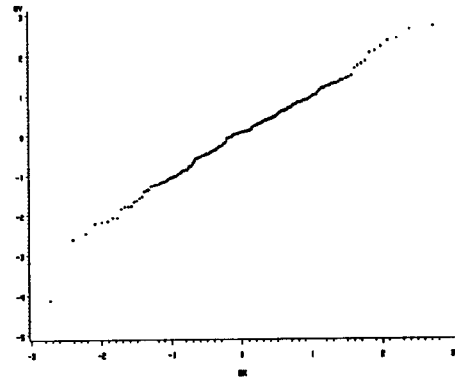
3.3 NTV plot에서의 분석

표본이 정규분포를 따를 때, 이론적인 NTV plot의 NTV 값은 대략 0과 0.02사이의 양수가 될 것이라고 기대된다[그림2-a]. 실제로 정규분포로부터 generating한 표본을 가지고 NTV plot을 그려보니 NTV값이 0 근처에서 불규칙하게 분포되어 있음을 알 수 있었다[그림2-b]. 이것은 이론적인 NTV plot에서는 정규분포의 순서통계량의 기대값을 이용하여 plot을 그렸기 때문에 난수발생 함수에 의하여 생성된 표본과는 차이가 생기게 된 것이다. 양쪽 꼬리부분이 짧은 균일분포에서의 NTV plot은 x좌표의 시작점($f=0$)에서부터 NTV 값이 양수에서 점점 감소하다가 0을 지나서는 불

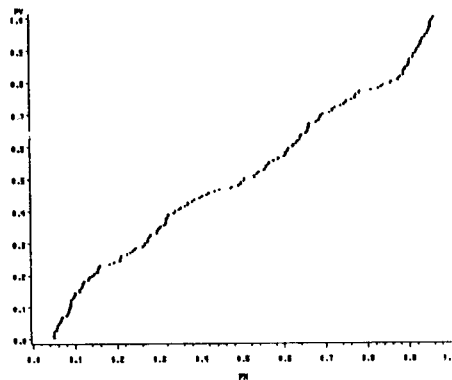
a 정규분포 p-value=0.8888



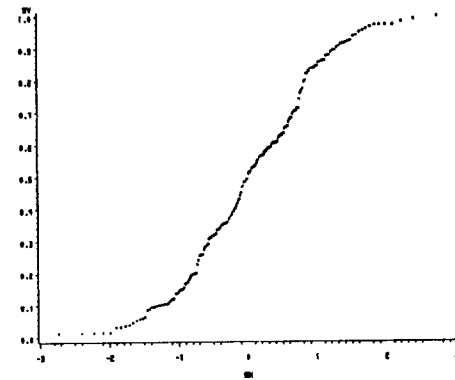
b 정규분포 p-value=0.8888



c 균일분포 p-value<0.0001

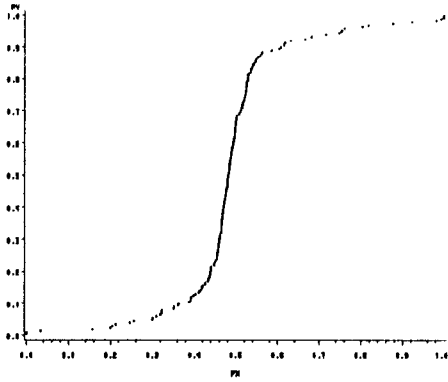


d 균일분포 p-value<0.0001

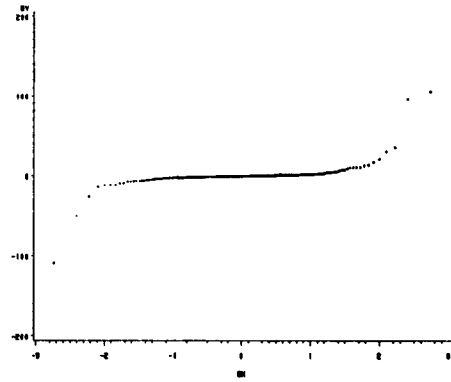


[그림1] 특정분포에서 생성된 표본에 관한 P-P Plot(a,c,e,g,i)과 Q-Q Plot(b,d,f,h,j)
(모든 그래프 상단의 P-value는 정규성 검정에 대한 Shapiro-Wilk 검정의 P-value임.)

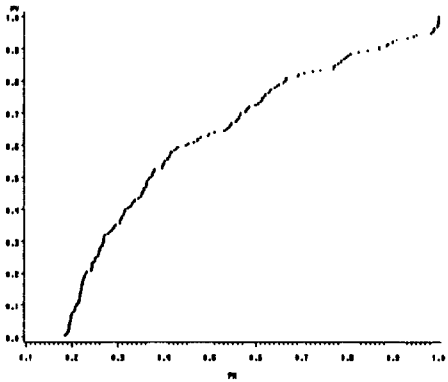
e 코쉬분포 p-value<0.0001



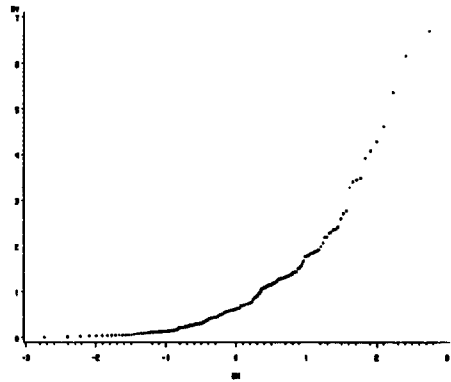
f 코쉬분포 p-value<0.0001



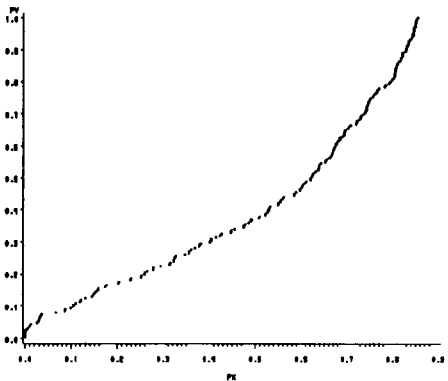
g 지수분포 p-value<0.0001



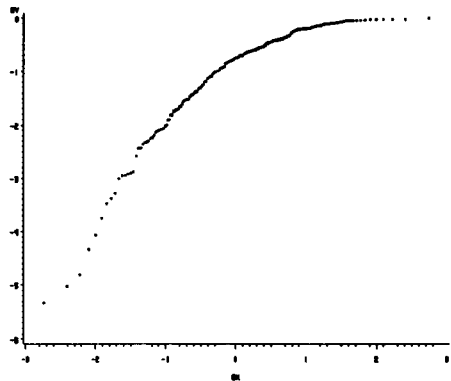
h 지수분포 p-value<0.0001



i 음의 지수분포 p-value<0.0001

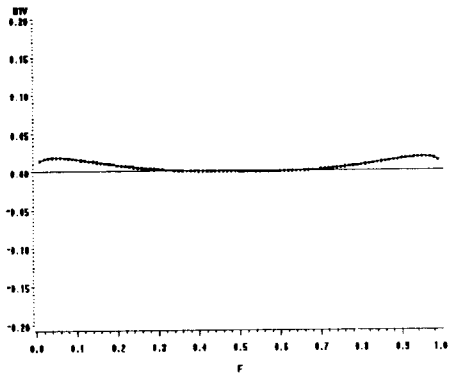


j 음의 지수분포 p-value<0.0001

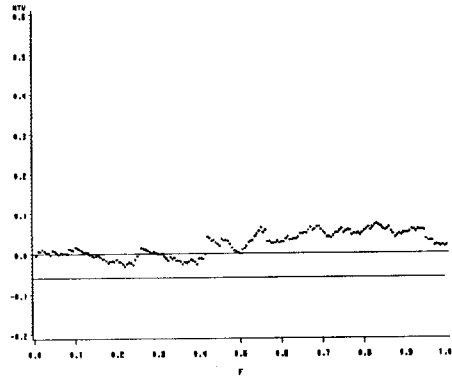


[그림1] (continued)

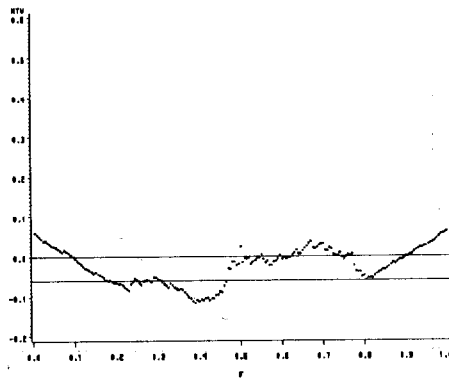
a 정규분포에 대한 이론적인 NTV plot



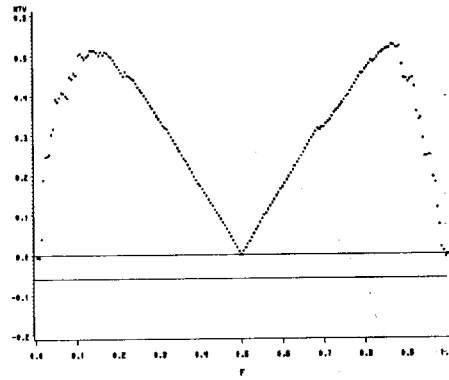
b 정규분포 p-value=0.8888



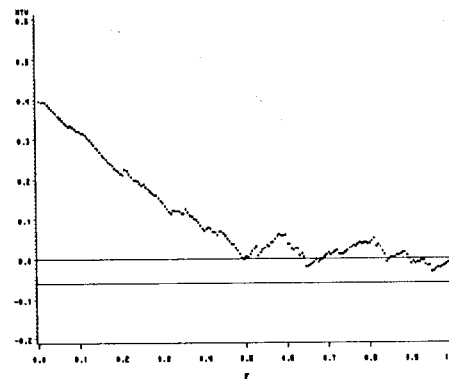
c 균일분포 p-value<0.0001



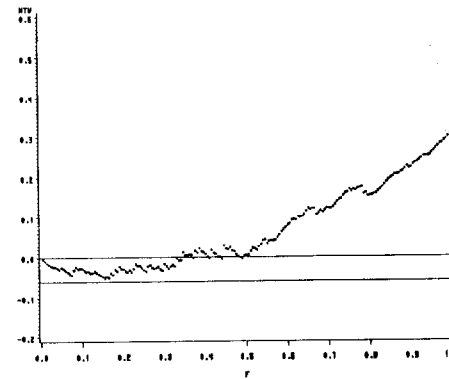
d 코쉬분포 p-value<0.0001



e 지수분포 p-value<0.0001



f 음의 지수분포 p-value<0.0001



[그림2] 특정분포에서 생성된 표본에 관한 NTV Plot

규칙하게 변화하다가 다시 x좌표의 끝점($f=1$)에 가까워 질 때 0에서 양수로 증가하는 형태를 가진다[그림2-c]. 특히 이 plot은 $f=0.5$ 부근에서는 0에 접근하는 형태가 된다. 양쪽 꼬리부분이 긴 코쉬분포에서의 NTV plot은 $f=0$ 에서부터 NTV 값이 양수에서 갑자기 상승하다가 조금 후 직선의 형태로 감소하여 $f=0.5$ 에서 0 부근으로 접근하다가 다시 선형으로 증가하다가 $f=1$ 에 가까워지면서 갑자기 감소하는 형태(M자 형태)를 가진다[그림2-d]. 오른쪽 꼬리부분이 긴 지수분포에서는 $f=0$ 에서 $f=0.5$ 까지는 NTV 값이 양수에서 0으로 거의 선형으로 감소하다가 $f=0.5$ 에서 $f=1$ 까지에서는 정규분포에서처럼 0 중심으로 불규칙하게 분포되어 있는 형태이다[그림2-e]. 왼쪽 꼬리부분이 긴 음의 지수분포에서는 앞의 지수분포에서의 plot의 형태와는 $f=0.5$ 를 중심으로 좌우 대칭인 형태를 가진다[그림2-f]. 이들 plot들을 관찰해 보니 NTV plot은 표본의 값에 매우 민감하게 변화하는 것을 알 수 있었다.

4. Bimodal 분포에 대한 분석

이번 4절에서 취급하려고 하는 bimodal 분포는 여러 가지 bimodal 분포들 가운데에서 두 개의 정규분포의 조합으로 이루어진 분포들이다.

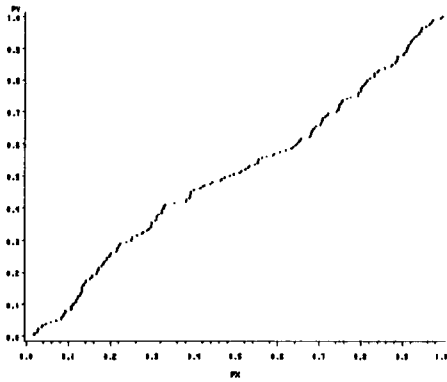
4.1 P-P plot과 Q-Q plot에서의 분석

먼저 모분산은 같고 모평균이 다른 두 정규 모집단에서 같은 비율로 생성된 표본에 대하여 분석해 본다. 두 모평균이 멀리 떨어져 있을수록 표본은 명확한 두 개의 peak를 보여주고 있다. 여기서는 두 모평균의 차가 각각 모분산의 2배와 4배가 되는 표본에 대하여 조사해 보았다. 모평균의 차가 모분산의 2배인 경우에서의 표본에 대한 plot들은 거의 정규분포와 비슷하게 즉, 직선에서 크게 벗어나지 않은 형태를 가지고 있다[그림3-a, b]. 이 plot의 형태에서는 beak-point를 명확히 파악할 수 없다. 한편, 모평균의 차가 모분산의 4배인 표본의 plot형태는 거의 균일분포에서의 형태와 비슷하게 되며, 실제로 여기에서는 뚜렷한 beak-point를 찾을 수 있다[그림3-c, d]. 다음으로, 모분산은 같고 모평균이 서로 3배의 모분산만큼 떨어져 있는 두 정규 모집단에서 서로 다른 비율로 생성된 표본에 대하여 분석해 본다. 두 모집단으로부터 3:1로 생성한 표본에서는 x축의 양쪽 끝점 부근에서만 제외하면 지수분포의 특징을 보이고 있다[그림3-e, f]. 두 모집단으로부터 5:1로 생성한 표본 역시 x축의 양쪽 끝점 부근에서만 제외하면 지수분포의 특징을 보이고 있으나 앞의 표본보다 왼쪽으로 치우친 특징을 많이 가지고 있다[그림3-g, h]. 이와 같이 bimodal 분포에서 생성된 표본에 대한 P-P plot과 Q-Q plot의 분석은 어디까지나 보는 사람의 주관적인 기준에 따라 결론을 내려야 하는 경우가 많기 때문에 매우 어렵다.

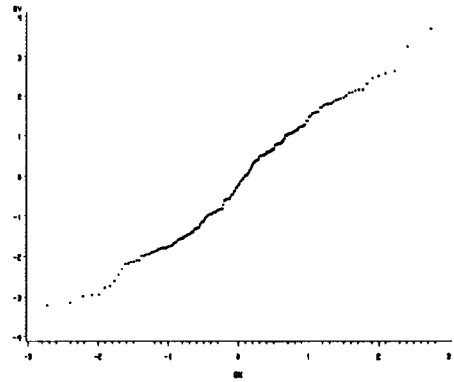
4.2 NTV plot에서의 분석

역시, 먼저 모분산은 같고 모평균이 다른 두 정규 모집단에서 같은 비율로 생성된 표본에 대하여 분석해 본다. 모평균의 차가 모분산의 2배인 경우에서의 표본에 대한 plot은 균일분포와는 다르게 $f=0.5$ 를 중심으로 좌우에 아래쪽으로 크게 하나씩의 peak를 가진 모양(W자 형태)을 형성하고 있다[그림4-a]. 모평균의 차가 모분산의 4배가 되면 plot에서 형성되는 peak들이 앞의 경우보다 더 크게 된다[그림4-b]. 다음으로, 모분산은 같고 모평균이 서로 3배의 모분산만큼 떨어져 있는 두 정규 모집단으로부터 3:1의 비율로 생성한 표본에서의 NTV plot은 지수분포에서와 비슷하게 $f=0$

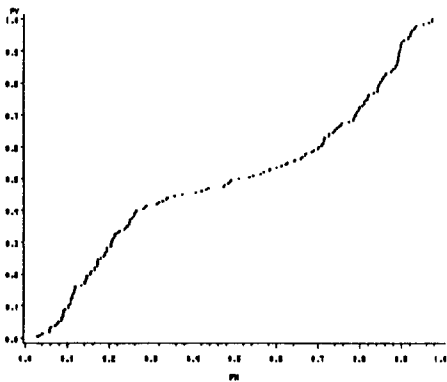
a $\frac{1}{2}N(-1,1) + \frac{1}{2}N(1,1)$ p-value=0.0019



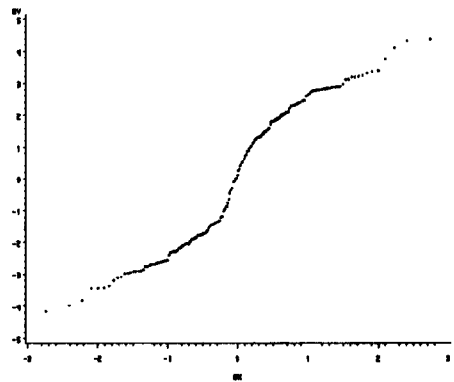
b $\frac{1}{2}N(-1,1) + \frac{1}{2}N(1,1)$ p-value=0.0019



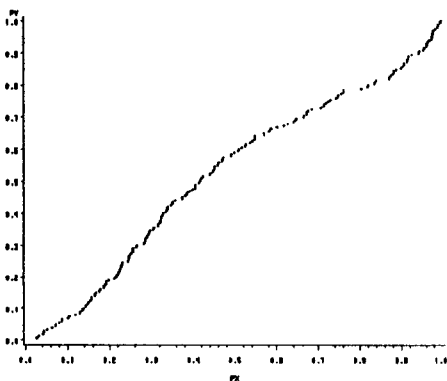
c $\frac{1}{2}N(-2,1) + \frac{1}{2}N(2,1)$ p-value<0.0001



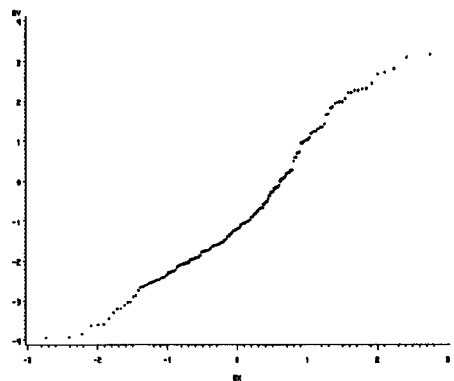
d $\frac{1}{2}N(-2,1) + \frac{1}{2}N(2,1)$ p-value<0.0001



e $\frac{3}{4}N(-1.5,1) + \frac{1}{4}N(1.5,1)$ p-value<0.0001

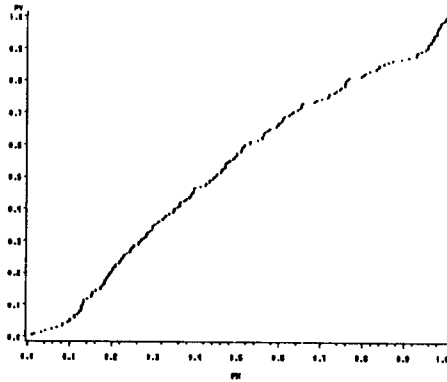


f $\frac{3}{4}N(-1.5,1) + \frac{1}{4}N(1.5,1)$ p-value<0.0001

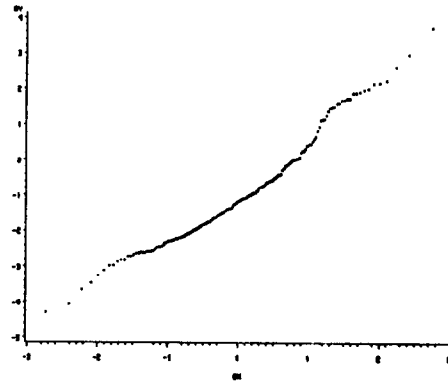


[그림3] Bimodal 분포에서 생성된 표본에 관한 P-P Plot(a,c,e,g)과 Q-Q Plot(b,d,f,h)

g $\frac{5}{6}N(-1.5,1)+\frac{1}{6}N(1.5,1)$ p-value<0.0001

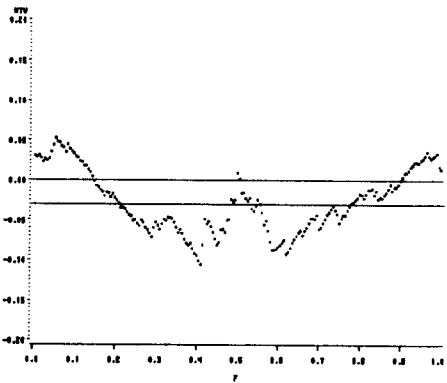


h $\frac{5}{6}N(-1.5,1)+\frac{1}{6}N(1.5,1)$ p-value<0.0001

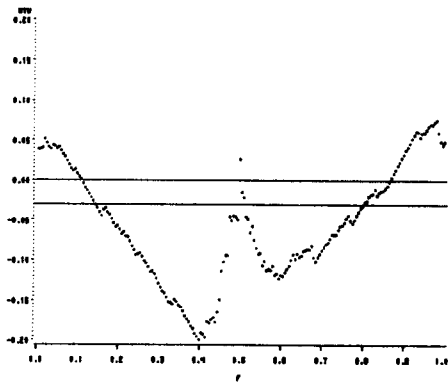


[그림3] (continued)

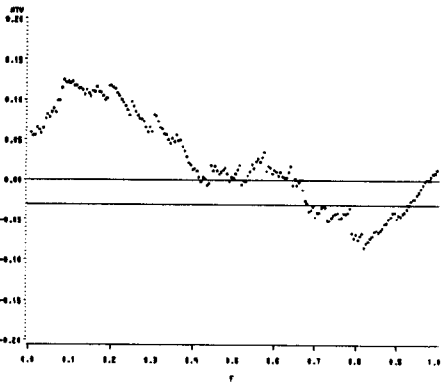
a $\frac{1}{2}N(-1,1)+\frac{1}{2}N(1,1)$ p-value=0.0019



b $\frac{1}{2}N(-2,1)+\frac{1}{2}N(2,1)$ p-value<0.0001



c $\frac{3}{4}N(-1.5,1)+\frac{1}{4}N(1.5,1)$ p-value<0.0001



d $\frac{5}{6}N(-1.5,1)+\frac{1}{6}N(1.5,1)$ p-value<0.0001



[그림4] Bimodal 분포에서 생성된 표본에 관한 NTV Plot

과 $f=0.5$ 사이에서 양수이지만 지수분포에서와는 달리 하나의 peak를 형성하고, $f=0.5$ 와 $f=1$ 사이에서는 불규칙한 모양에서 아래쪽으로 하나의 peak을 음수 영역에서 형성한다[그림4-c]. 두 모집단으로부터 5:1의 비율로 생성한 표본에서는 앞의 plot에서의 나타난 peak들이 각각 양쪽 끝으로 조금씩 더 이동한 형상이다[그림4-d]. 이처럼 P-P plot나 Q-Q plot에서 잘 감지할 수 없었던 bimodality를 NTV plot에서는 좀더 쉽게 알 수 있다. 하지만 이와 같은 분석도 객관적인 기준을 찾기는 쉽지 않다.

참고 문헌

- [1] Arnold, B.C. and Groeneveld, R.A. (1995). Measuring skewness with respect to the mode, *The American Statistician*, Vol. 49, 34-38.
- [2] D'Agostino, R.B. (1971). An omnibus test of normality for moderate and large sample size, *Biometrika*, Vol. 58, 341-348.
- [3] Dyer, A.R. (1974). Comparison of tests for normality with a cautionary note, *Biometrika*, Vol. 61, 185-189.
- [4] Endrenyi, L. and Patel, M. (1991a). A new, sensitive graphical method for detecting deviations from the normal distribution of drug responses: the NTV plot, *Br. J. clin. pharmac.*, Vol. 32, 159-166.
- [5] Endrenyi, L. and Patel, M. (1991b). Evaluation of two assumptions: single straight line, and single normal distribution. *Trends. Pharmacol. Sci.*, Vol. 12, 293-296.
- [6] Jackson, P.R., Tucker, G.T. and Woods, H.F. (1989a). Testing for bimodality in frequency distributions of data suggesting polymorphisms of drug metabolism—histograms and probit plots, *Br. J. clin. pharmac.*, Vol. 28, 647-653.
- [7] Jackson, P.R., Tucker, G.T. and Woods, H.F. (1989b). Testing for bimodality in frequency distributions of data suggesting polymorphisms of drug metabolism—hypothesis testing, *Br. J. clin. pharmac.*, Vol. 28, 655-662.
- [8] Lee, J.-Y., Hah, J.O. and Rhee, S.-W. (1997). Normal Test Variable(NTV) plots and Normal Probability Plots by Generating Normal Random Numbers, *Br. J. clin. Pharmac.*, To appear.
- [9] Looney, S.W. (1995). How to use tests for univariate normality to assess multivariate normality, *The American Statistician*, Vol. 49, 64-70.
- [10] Shapiro, S.S. and Wilk, M.B. (1965). An analysis-of-variance test for normality (complete sample), *Biometrika*, Vol. 52, 591-611.
- [11] Shapiro, S.S. and Francia, R.S. (1972). An approximation analysis of variance test for normality, *J. Am. Stat. Assoc.*, Vol. 67, 215-216.
- [12] Spiegelhalter, D.J. (1977). A test for normality against symmetric alternatives, *Biometrika*, Vol. 64, 415-418.
- [13] Weisberg, S. and Bingham, C. (1975). An approximation analysis of variance test for non-normality suitable for machine calculation, *J. Am. Stat. Assoc.*, 17, 133-134.