Journal of Statistical
Theory & Methods
1998, Vol. 9, No. 1, pp. 47 ~ 56

이변량 영과잉-포아송 분포의 적률 [1]

김 경 무 [2] · 이 성 호 · 김 종 태 [3]

요약

영과잉-포아송모형는 포아송분포와 베르누이 분포의 혼합모형으로 볼 수 있다. 최근 기 술의 발달로 생산공정에서 불량품이 거의 나타나지 않는 경우가 많아 기존의 포아송 분포 보다 영과잉-포아송 분포가 많이 응용되어 진다. 일변량 영과잉-포아송 분포를 이변량 영과잉-포아송 분포로 확장하는 일은 다변량으로 확장하기 위한 전초작업으로 중요하다. 본 논문에서는 세가지 형태의 이변량 영과잉-포아송 분포를 제시하고 이들 분포의 적률 을 구하여 보았다. 또한 적률을 이용하여 세가지 분포를 비교하여 보았다.

주제어: 적률, 영과잉-포아송분포, 이변량 영과잉-포아송분포

1. 서 론

영과잉-포아송(Zero-Inflated Poisson:이후 *ZIP*로 표기함)분포는 이산형 확률분포에서 정상적인 포아송 확률분포보다 영의 값이 과잉 관측되는 경우를 묘사하기 위하여 기존의 확률분포를 변형시킨 것이다. 포아송분포는 생산공정과정에서 단위당 나타나는 불량품의 수에 관한 확률분포로서 지금까지 중요한 분포로 이용되어 왔다. 그러나 문명의 발달과 제품을 만들어 내는 기술의 고급화로 인하여 불량률은 현저하게 감소되어 가고 있다. 예를 들면, 반도체 분야에서 컴퓨터에 내장되어 있는 메모리 칩들은 생산공정의 정확도와 기술의 발달로 불량품으로 판정되는 경우가 극히 드물다. 다시말하면 여러번 실시한 표본검사 중 단위당 불량품의 수는 대부분 영이고 극히 적은 경우에 한하여 불량품이 발생한다고 볼 수 있다. 이러한 경우, 기존의 포아송분포에 적합시켜 통계적인 추정 및 검정을 한다면 이는 제 3종의 통계적 오류를 범하는 결과를 초래할 것이다.

이러한 *ZIP*분포는 처음으로 Singh(1963)에 의해 처음으로 소개되었으나 수학적인 모형으로만 인식되어 응용분야가 다양하지 못하였다. 그러나 최근 Lambert(1992)는 *ZIP*분포를

[1] 이 논문은 1997학년도 대구대학교 학술연구비 지원에 의한 논문임.
[2] 대구대학교 자연과학대학 통계학과 교수, (712-714) 경북 경산시 진량읍 내리리 15
[3] 대구대학교 자연과학대학 통계학과 부교수

이용한 ZIP 회귀모형을 제시하였고 공변량(covariate)들의 효과를 추정하고 이를 실제의 자료에 응용하였다. 그러나 지금까지는 확률변수가 한 개인 일변량 ZIP분포만 연구되어 왔으나 이를 확장하여 서 로 종속관계가 있는 두 확률변수의 이변량 ZIP분포를 설계하려 한다. 또한 이변량 ZIP분포는 유 일하게 결정되지 않기 때문에 세가지 형태의 이변량 ZIP분포를 제시하고 이들의 적률을 구하여 보았다. 또한 적률을 이용하여 세가지 형태의 이변량 ZIP분포들을 비교하여 보았다. 이변량분포 로 확장하는 일은 다변량으로 일반화시키기 위한 전초작업으로 중요한 일일 것이다.

2. 이변량 영과잉-포아송 모형

확률변수 Y는 생산공정에서 일정 단위당 불량품이 나타나는 수로서 일변량 ZIP 분포를 따른 다고 하자. ZIP는 포아송분포와 베르누이분포와의 혼합모형으로 볼 수 있다. 즉,

$$\begin{aligned} Y &\sim 0, & p \text{ 의 확률로} \\ &\sim Poisson(\lambda), & 1-p \text{ 의 확률}, \end{aligned}$$

여기에서 $0 \le p \le 1$는 불량품이 전혀 나타나지 않는 상태(perfect state)의 확률이며, $\lambda \ge 0$는 포아송분포의 평균이다. 이때 확률질량함수(pmf)는 아래와 같이 된다.

$$\begin{aligned} P(Y=k) &= p + (1-p)e^{-\lambda}, & k = 0 \\ &= (1-p)\lambda^k e^{-\lambda}/k!\ , & k = 1, 2, \cdots. \end{aligned}$$

앞으로 위 분포를 $ZIP(p,\lambda)$로 표기하기로 한다. 이를 이변량 ZIP분포로 확장하려 한다. 그러나 이변량 ZIP분포는 유일하지 않기 때문에, 본 논문에서는 세가지 형태의 이변량 ZIP분포를 제 시하려 한다.

2.1 모형 I

서로 종속인 두 확률변수 Y_1, Y_2가 각각ZIP분포를 따른다고 하자. 특히 이들 두 변수를 이용 한 이변량 ZIP분포의 주변분포(marginal distribution)가 역시 일변량 ZIP분포가 되도록 모형을 설계하려 한다. (Y_1, Y_2)가 모형 (1)을 따른다 하자.

$$(Y_1, Y_2) \sim \begin{cases} (0,0), & p_{00} \text{ 의 확률로}, \\ (0,V), & p_{01} \text{ 의 확률로}, \\ (U,0), & p_{10} \text{ 의 확률로}, \\ (U,V), & p_{11} \text{ 의 확률로} \end{cases} \tag{1}$$

여기에서 $p_{00} + p_{10} + p_{01} + p_{11} = 1$ 이다. 또한 U와 V는 각각 평균이 λ_1 및 λ_2의 일변량 포아송분포를 따르고 (U,V)는 이변량 포아송분포 $Poisson(\lambda_{10}, \lambda_{01}, \lambda_{11})$를 따르고 그 분포식

은 (2)와 같다.

$$P(U=u,V=v) \;=\; \sum_{a=0}^{min(u,v)} \frac{\lambda_{11}^{a}\lambda_{10}^{u-a}\lambda_{01}^{v-a}}{a!(u-a)!(v-a)!e^{-(\lambda_{10}+\lambda_{01}+\lambda_{11})}}, \qquad (2)$$
$$u=0,1,2,\cdots, \quad v=0,1,2,\cdots.$$

이러한 이변량 ZIP분포를 $BZIP_I(p_{00},p_{10},p_{01},\lambda_{10},\lambda_{01},\lambda_{11})$ 로 나타내고 6개의 모수를 포함하고 있다.

모형 I의 BZIP분포 특성을 알아보기 위하여 확률질량함수, 적률, 조건부확률 및 조건부 기대 값 등을 구하여 보았다. $BZIP_I(p_{00},p_{10},p_{01},\lambda_{10},\lambda_{01},\lambda_{11})$ 분포의 확률질량함수는 다음과 같게 된다.

$$\begin{aligned}
P(Y_1=0,Y_2=0) &= p_{00}+p_{10}e^{-\lambda_1}+p_{01}e^{-\lambda_2}+p_{11}e^{-\lambda},\\
P(Y_1=k_1,Y_2=0) &= p_{10}\frac{\lambda_1^{k_1}}{k_1!}e^{-\lambda_1}+p_{11}\frac{\lambda_{10}^{k_1}}{k_1!}e^{-\lambda},\\
P(Y_1=0,Y_2=k_2) &= p_{01}\frac{\lambda_2^{k_2}}{k_2!}e^{-\lambda_2}+p_{11}\frac{\lambda_{01}^{k_2}}{k_2!}e^{-\lambda},\\
P(Y_1=k_1,Y_2=k_2) &= p_{11}\sum_{a=0}^{min(k_1,k_2)}\frac{\lambda_{11}^{a}\lambda_{10}^{k_1-a}\lambda_{01}^{k_2-a}}{a!(k_1-a)(k_2-a)!}e^{-\lambda},
\end{aligned}$$

여기에서 $\lambda_1=\lambda_{10}+\lambda_{11}$, $\lambda_2=\lambda_{01}+\lambda_{11}, \lambda=\lambda_{10}+\lambda_{01}+\lambda_{11}$이다. 물론 위 $BZIP_I$의 확률질량함수는 다음의 성질을 만족하게 된다.

$$\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}P(Y_1=k_1,Y_2=k_2)=1.$$

그리고 이변량 $BZIP_I$에 대한 Y_1 및 Y_2의 주변분포는 다음과 같이 구하여진다.

$$\begin{aligned}
P(Y_1=k_1) &= (p_{00}+p_{01})+(p_{10}+p_{11})e^{-\lambda_1}, & k_1&=0,\\
&= (p_{10}+p_{11})\frac{\lambda_1^{k_1}}{k_1!}e^{-\lambda_1}, & k_1&>0\\
P(Y_2=k_2) &= (p_{00}+p_{10})+(p_{01}+p_{11})e^{-\lambda_2}, & k_2&=0,\\
&= (p_{01}+p_{11})\frac{\lambda_2^{k_2}}{k_2!}e^{-\lambda_2}, & k_2&>0
\end{aligned}$$

여기에서 특이할만한 점은 위 Y_1 및 Y_2의 주변분포가 각각 일변량 $ZIP(p_{00}+p_{01},\lambda_1)$ 그리고 $ZIP(p_{00}+p_{10},\lambda_2)$가 된다는 것이다. 다시말하면

$$\sum_{k_1=0}^{\infty} P(Y_1 = k_1, Y_2 = k_2) = P(Y_2 = k_2).$$

$BZIP_I$의 주변분포는 일변량 ZIP가 되는 좋은 성질을 가지고 있다. 또한 Y_1 및 Y_2 의 평균, 분산 및 공분산은 다음과 같다.

$$\begin{aligned}
E(Y_1) &= (p_{10} + p_{11})\lambda_1 \\
E(Y_2) &= (p_{01} + p_{11})\lambda_2 \\
Var(Y_1) &= (p_{10} + p_{11})\lambda_1(1 + \lambda_1 - (p_{10} + p_{11})\lambda_1) \\
Var(Y_2) &= (p_{01} + p_{11})\lambda_2(1 + \lambda_2 - (p_{01} + p_{11})\lambda_2) \\
cov(Y_1, Y_2) &= p_{11}(\lambda_1\lambda_2 + \lambda_{11}) - (p_{10} + p_{11})(p_{01} + p_{11})\lambda_1\lambda_2.
\end{aligned}$$

또한 실수 s, t에 대하여 (Y_1, Y_2)의 적률모함수(mgf)는

$$\begin{aligned}
m(s,t) \;&=\; p_{00} + p_{10}\exp((e^s - 1)\lambda_1) + p_{01}\exp((e^t - 1)\lambda_2) \\
&\quad + p_{11}\exp(\lambda_{10}e^s + \lambda_{01}e^t + \lambda_{11}e^{s+t} - \lambda)
\end{aligned}$$

와 같게 된다. 예를 들면 Y_2의 평균은 위 적률모함수를 이용하면 다음과 같이 될 것이다.

$$\left\{\frac{\partial\, m(s,t)}{\partial\, t}\right\}_{s=t=0} = E(Y_2) = (p_{01} + p_{11})\lambda_2$$

그리고 $Y_1|Y_2 = k_2$의 조건부확률(conditional probability)은 $k_2 = 0$일 때,

$$\begin{aligned}
P(Y_1 = k_1|Y_2 = k_2) \;&=\; \frac{p_{00} + p_{10}e^{-\lambda_1} + p_{01}e^{-\lambda_2} + p_{11}e^{-\lambda}}{(p_{00} + p_{10}) + (p_{01} + p_{11})e^{-\lambda_2}}, \qquad k_1 = 0 \\
&=\; \frac{p_{10}\frac{\lambda_1^{k_1}e^{-\lambda_1}}{k_1!} + p_{11}\frac{\lambda_{10}^{k_1}e^{-\lambda}}{k_1!}}{(p_{00} + p_{10}) + (p_{01} + p_{11})e^{-\lambda_2}}, \qquad k_1 > 0
\end{aligned}$$

그리고 $k_2 > 0$일 경우는 다음과 같다.

$$\begin{aligned}
P(Y_1 = k_1|Y_2 = k_2) \;&=\; \frac{p_{01}\lambda_2^{k_2} + p_{11}\lambda_{01}^{k_2}e^{-\lambda_{10}}}{(p_{01} + p_{11})\lambda_2^{k_2}}, \qquad k_1 = 0 \\
&=\; \frac{p_{11}k_2!\sum_{a=0}^{min(k_1,k_2)} \frac{\lambda_{11}^{a}\lambda_{10}^{k_1-a}\lambda_{01}^{k_2-a}}{a!(k_1-a)!(k_2-a)!}e^{-\lambda}}{(p_{01} + p_{11})\lambda_2^{k_2}e^{-\lambda_2}}, \qquad k_1 > 0
\end{aligned}$$

위 조건부확률을 이용하면 조건부기대값(conditional expectation)이 다음과 같이 구하여진다.

$$E(Y_1|Y_2=k_2) = \frac{p_{10}\lambda_1 + p_{11}\lambda_{10}e^{-\lambda_2}}{p_{00}+p_{10}+(p_{01}+p_{11})e^{-\lambda_2}}, \quad k_2 = 0 \quad (3)$$
$$= \frac{p_{11}}{p_{01}+p_{11}}(\lambda_{10}+\frac{\lambda_{11}}{\lambda_2}k_2), \quad k_2 > 0.$$

2.2 모형 II

이변량 확률변수 (Y_1, Y_2) 가 모형 (1)을 따른다고 하자. 이때 U와 V는 각각 모형 I에서와 같이 포아송분포를 따르지 않고 $ZIP(1-e^{-(\lambda_{01}+\lambda_{11})}, \lambda_{10})$ 그리고 $ZIP(1-e^{-(\lambda_{10}+\lambda_{11})}, \lambda_{01})$ 를 따른다고 하자. 그리고 (U,V)는 이변량 포아송분포 $Poisson(\lambda_{10}, \lambda_{01}, \lambda_{11})$ 를 따르도록 설계한다. 이러한 이변량 ZIP분포를 $BZIP_{II}(p_{00}, p_{10}, p_{01}, \lambda_{10}, \lambda_{01}, \lambda_{11})$로 표기하기로 한다. 물론 모형 I과 마찬가지로 $p_{00}+p_{10}+p_{01}+p_{11}=1$, $\lambda_1 = \lambda_{10}+\lambda_{11}$, $\lambda_2 = \lambda_{01}+\lambda_{11}$ 그리고 $\lambda = \lambda_{10}+\lambda_{01}+\lambda_{11}$ 을 만족한다. $BZIP_{II}(p_{00}, p_{10}, p_{01}, \lambda_{10}, \lambda_{01}, \lambda_{11})$ 분포의 확률질량함수는 다음과 같다.

$$P(Y_1=0, Y_2=0) = p_{00} + p_{10}(1-e^{-\lambda_2}+e^{-\lambda}) + p_{01}(1-e^{-\lambda_1}+e^{-\lambda}) + p_{11}e^{-\lambda},$$
$$P(Y_1=k_1, Y_2=0) = (p_{10}+p_{11})\frac{{\lambda_{10}}^{k_1}e^{-\lambda}}{k_1!},$$
$$P(Y_1=0, Y_2=k_2) = (p_{01}+p_{11})\frac{{\lambda_{01}}^{k_2}e^{-\lambda}}{k_2!},$$
$$P(Y_1=k_1, Y_2=k_2) = p_{11}\sum_{a=0}^{min(k_1,k_2)}\frac{\lambda_{11}^a\lambda_{10}^{k_1-a}\lambda_{01}^{k_2-a}}{a!(k_1-a)(k_2-a)!}e^{-\lambda}.$$

그리고 $BZIP_{II}$에 대한 Y_1및 Y_2의 주변분포는 다음과 같이 구하여진다.

$$P(Y_1=k_1) = p_{00}+p_{01}+p_{10}(1-e^{-\lambda_2}+e^{-\lambda})+p_{11}e^{-\lambda_1} \quad ,k_1=0,$$
$$= p_{10}\frac{\lambda_{10}^{k_1}}{k_1!}e^{-\lambda}+p_{11}\frac{\lambda_1^{k_1}}{k_1!}e^{-\lambda_1}, \quad k_1>0$$
$$P(Y_2=k_2) = p_{00}+p_{10}+p_{01}(1-e^{-\lambda_1}+e^{-\lambda})+p_{11}e^{-\lambda_2} \quad ,k_2=0,$$
$$= p_{01}\frac{\lambda_{01}^{k_2}}{k_2!}e^{-\lambda}+p_{11}\frac{\lambda_2^{k_2}}{k_2!}e^{-\lambda_2}, \quad k_2>0.$$

또한 Y_1 및 Y_2의 평균, 분산 및 공분산은 다음과 같다.

$$E(Y_1) = p_{10}\lambda_{10}e^{-\lambda_2} + p_{11}\lambda_1$$
$$E(Y_2) = p_{01}\lambda_{01}e^{-\lambda_1} + p_{11}\lambda_2$$
$$Var(Y_1) = p_{10}\lambda_{10}e^{\lambda_2}(\lambda_{10}+1) + p_{11}(\lambda_1+1)\lambda_1 - (p_{10}\lambda_{10}e^{-\lambda_2}+p_{11}\lambda_1)^2$$

$$
\begin{aligned}
Var(Y_2) &= p_{01}\lambda_{01}e^{\lambda_1}(\lambda_{01}+1) + p_{11}(\lambda_2+1)\lambda_2 - (p_{01}\lambda_{01}e^{-\lambda_1} + p_{11}\lambda_2)^2 \\
cov(Y_1, Y_2) &= p_{11}(\lambda_1\lambda_2 + \lambda_{11}) - (p_{10}\lambda_{10}e^{-\lambda_2} + p_{11}\lambda_1)(p_{01}\lambda_{01}e^{-\lambda_1} + p_{11}\lambda_2).
\end{aligned}
$$

또한 실수 s, t에 대하여 Y_1, Y_2 의 적률모함수(mgf)는

$$
\begin{aligned}
m(s,t) \;\; = \;\; & p_{00} + p_{10}(1 - e^{-\lambda_2} + \exp(\lambda_{10}e^{s} - \lambda)) + p_{01}(1 - e^{-\lambda_1} + \exp(\lambda_{01}e^{t} - \lambda)) \\
& + p_{11}\exp(\lambda_{10}e^{s} + \lambda_{01}e^{t} + \lambda_{11}e^{s+t} - \lambda)
\end{aligned}
$$

와 같게 된다. 그리고 $Y_1|Y_2 = k_2$의 조건부확률(conditional probability)은 $k_2 = 0$일 때,

$$
\begin{aligned}
P(Y_1 = k_1|Y_2 = k_2) &= \frac{p_{00} + p_{10}(1 - e^{-\lambda_2} + e^{-\lambda}) + p_{01}(1 - e^{-\lambda_1} + e^{-\lambda}) + p_{11}e^{-\lambda}}{p_{00} + p_{10} + p_{01}(1 - e^{-\lambda_1} + e^{-\lambda}) + p_{11}e^{-\lambda_2}}, \quad k_1 = 0 \\
&= \frac{(p_{10} + p_{11})\frac{\lambda_{10}^{k_1}e^{-\lambda}}{k_1!}}{p_{00} + p_{10} + p_{01}(1 - e^{-\lambda_1} + e^{-\lambda}) + p_{11}e^{-\lambda_2}}, \quad k_1 > 0.
\end{aligned}
$$

그리고 $k_2 > 0$ 일 경우는 다음과 같다.

$$
\begin{aligned}
P(Y_1 = k_1|Y_2 = k_2) &= \frac{(p_{01} + p_{11})\frac{\lambda_{01}^{k_2}e^{-\lambda}}{k_2!}}{p_{01}\frac{\lambda_{01}^{k_2}e^{-\lambda}}{k_2!} + p_{11}\frac{\lambda_{2}^{k_2}e^{-\lambda_2}}{k_2!}}, \quad k_1 = 0 \\
&= \frac{p_{11}\sum_{a=0}^{min(k_1,k_2)} \frac{\lambda_{11}^{a}\lambda_{10}^{k_1-a}\lambda_{01}^{k_2-a}}{a!(k_1-a)!(k_2-a)!}e^{-\lambda}}{\frac{p_{01}\lambda_{01}^{k_2}e^{-\lambda}}{k_2!} + \frac{p_{11}\lambda_{2}^{k_2}e^{-\lambda_2}}{k_2!}}, \quad k_1 > 0.
\end{aligned}
$$

위 조건부확률을 이용하면 조건부기대값(Conditional expectation)이 다음과 같이 구하여진다.

$$
\begin{aligned}
E(Y_1|Y_2 = k_2) &= \frac{(p_{10} + p_{11})\lambda_{10}e^{-\lambda_2}}{p_{00} + p_{10} + p_{01}(1 - e^{-\lambda_1} + e^{-\lambda}) + p_{11}e^{-\lambda_2}}, \quad k_2 = 0 \\
&= \frac{p_{11}\lambda_2^{k_2}e^{-\lambda_2}}{p_{01}\lambda_{01}^{k_2}e^{-\lambda} + p_{11}\lambda_2^{k_2}e^{-\lambda_2}}(\lambda_{10} + \frac{\lambda_{11}}{\lambda_2}k_2), \quad k_2 > 0.
\end{aligned}
$$

2.3 모형 III

모형 III은 I, II에서 불량품수가 없는 확률, 즉 $P(Y_1 = 0, Y_2 = 0)$이 다르게 되게끔 설계하려한다. 이때 모형 III에 대한 이변량 ZIP를 $BZIP_{III}$ 라 표기하면 $BZIP_{II}(p_{00}, p_{10}, p_{01}, \lambda_{10}, \lambda_{01}, \lambda_{11})$분포의 확률질량함수는 다음과 같다.

$$
P(Y_1 = 0, Y_2 = 0) \;\; = \;\; p_{00} + p_{10}e^{-\lambda_{10}} + p_{01}e^{-\lambda_{01}} + p_{11}e^{-\lambda},
$$

$$
\begin{aligned}
P(Y_1=k_1, Y_2=0) &= p_{10}\frac{\lambda_{10}{}^{k_1}e^{-\lambda_{10}}}{k_1!} + p_{11}\frac{\lambda_{10}{}^{k_1}e^{-\lambda}}{k_1!},\\
P(Y_1=0, Y_2=k_2) &= p_{01}\frac{\lambda_{01}{}^{k_2}e^{-\lambda_{01}}}{k_2!} + p_{11}\frac{\lambda_{01}{}^{k_2}e^{-\lambda}}{k_2!},\\
P(Y_1=k_1, Y_2=k_2) &= p_{11}\sum_{a=0}^{min(k_1,k_2)}\frac{\lambda_{11}^{a}\lambda_{10}^{k_1-a}\lambda_{01}^{k_2-a}}{a!(k_1-a)(k_2-a)!}e^{-\lambda},
\end{aligned}
$$

여기에서 앞절과 마찬가지로 $p_{00}+p_{10}+p_{01}+p_{11}=1$, $\lambda_1=\lambda_{10}+\lambda_{11}$, $\lambda_2=\lambda_{01}+\lambda_{11}$ 그리고 $\lambda=\lambda_{10}+\lambda_{01}+\lambda_{11}$ 을 만족한다. 그리고 $BZIP_{III}$에 대한 Y_1 및 Y_2의 주변분포는 다음과 같이 구하여진다.

$$
\begin{aligned}
P(Y_1=k_1) &= p_{00}+p_{01}+p_{10}e^{-\lambda_{10}}+p_{11}e^{-\lambda}, && k_1=0,\\
&= p_{10}\frac{\lambda_{10}^{k_1}e^{-\lambda_{10}}}{k_1!}+p_{11}\frac{\lambda_1^{k_1}e^{-\lambda_1}}{k_1!}, && k_1>0\\
P(Y_2=k_2) &= p_{00}+p_{10}+p_{01}e^{-\lambda_{01}}+p_{11}e^{-\lambda_2}, && k_2=0,\\
&= p_{01}\frac{\lambda_{01}^{k_2}e^{-\lambda_{01}}}{k_2!}+p_{11}\frac{\lambda_2^{k_2}e^{-\lambda_2}}{k_2!}, && k_2>0.
\end{aligned}
$$

또한 Y_1 및 Y_2의 평균, 분산 및 공분산은 다음과 같다.

$$
\begin{aligned}
&E(Y_1)=p_{10}\lambda_{10}+p_{11}\lambda_1\\
&E(Y_2)=p_{01}\lambda_{01}+p_{11}\lambda_2\\
&Var(Y_1)=p_{10}\lambda_{10}(\lambda_{10}+1)+p_{11}\lambda_1(\lambda_1+1)-(p_{10}\lambda_{10}+p_{11}\lambda_1)^2\\
&Var(Y_2)=p_{01}\lambda_{01}(\lambda_{01}+1)+p_{11}\lambda_2(\lambda_2+1)-(p_{01}\lambda_{01}+p_{11}\lambda_2)^2\\
&cov(Y_1,Y_2)=p_{11}(\lambda_1\lambda_2-\lambda_{11})-(p_{10}\lambda_{10}+p_{11}\lambda_1)(p_{01}\lambda_{01}+p_{11}\lambda_2)
\end{aligned}
$$

또한 실수 s,t에 대하여 Y_1,Y_2의 적률모함수(mgf)는

$$
\begin{aligned}
m(s,t) &= p_{00}+p_{10}\exp\Big(\lambda_{10}(e^{s}-1)\Big)+p_{01}\exp\Big(\lambda_{01}(e^{t}-1)\Big)\\
&\quad +p_{11}\exp\Big(\lambda_{10}e^{s}+\lambda_{01}e^{t}+\lambda_{11}e^{s+t}-\lambda\Big)
\end{aligned}
$$

와 같게 된다. 그리고 $Y_1|Y_2=k_2$의 조건부확률은 $k_2=0$일 때,

$$
\begin{aligned}
P(Y_1=k_1|Y_2=k_2) &= \frac{p_{00}+p_{10}e^{-\lambda_{10}}+p_{01}e^{-\lambda_{01}}+p_{11}e^{-\lambda}}{p_{00}+p_{10}+p_{01}e^{-\lambda_{01}}+p_{11}e^{-\lambda_2}}, && k_1=0\\
&= \frac{p_{10}\frac{\lambda_{10}^{k_1}e^{-\lambda_{10}}}{k_1!}+p_{11}\frac{\lambda_{10}^{k_1}e^{-\lambda}}{k_1!}}{p_{00}+p_{10}+p_{01}e^{-\lambda_{01}}+p_{11}e^{-\lambda_2}}, && k_1>0
\end{aligned}
$$

그리고 $k_2 > 0$ 일 경우는 다음과 같다.

$$
\begin{aligned}
P(Y_1 = k_1 | Y_2 = k_2) &= \frac{p_{01}\lambda_{01}^{k_2} e^{-\lambda_{01}} + p_{11}\lambda_{01}^{k_2} e^{-\lambda}}{p_{01}\lambda_{01}^{k_2} e^{-\lambda_{01} + p_{11}\lambda_2^{k_2}} e^{-\lambda_2}}, \qquad k_1 = 0 \\
&= \frac{p_{11}\sum_{a=0}^{min(k_1,k_2)} \frac{\lambda_{11}^{a}\lambda_{10}^{k_1-a}\lambda_{01}^{k_2-a}}{a!(k_1-a)!(k_2-a)!} e^{-\lambda}}{p_{01}\frac{\lambda_{01}^{k_2} e^{-\lambda_{01}}}{k_2!} + p_{11}\frac{\lambda_2^{k_2} e^{-\lambda_2}}{k_2!}}, \qquad k_1 > 0
\end{aligned}
$$

위 조건부확률을 이용하면 조건부기대값(conditional expectation)은 다음과 같이 구하여진다.

$$
\begin{aligned}
E(Y_1 | Y_2 = k_2) &= \frac{p_{10}\lambda_{10} + p_{11}\lambda_{10} e^{-\lambda_2}}{p_{00} + p_{10} + p_{01} e^{-\lambda_{01}} + p_{11} e^{-\lambda_2}}, \qquad k_2 = 0 \\
&= \frac{p_{11}\lambda_2^{k_2} e^{-\lambda_2}}{p_{01}\lambda_{01}^{k_2} e^{-\lambda_{01}} + p_{11}\lambda_2^{k_2} e^{-\lambda_2}} (\lambda_{10} + \frac{\lambda_{11}}{\lambda_2} k_2), \qquad k_2 > 0.
\end{aligned}
$$

3. BZIP분포의 비교

세가지 형태의 이변량 영과잉-포아송분포를 비교하면 다음과 같은 사실을 알 수 있다.

1) 세가지 형태의 이변량 영과잉-포아송 모형 중 불량품의 수가 없는 확률, 즉 $P(Y_1 = 0, Y_2 = 0)$은 모형 II에서 가장 크게된다.

$P(Y_1 = 0, Y_2 = 0)$ 이 가장 큰 경우의 모형을 찾아내기 위하여, 모형 I, II 및 III에 대한 $P(Y_1 = 0, Y_2 = 0)$의 확률을 각각 P_I, P_{II}, P_{III} 라고 한다면,

$$
\begin{aligned}
P_I &= p_{00} + p_{10}e^{-\lambda_1} + p_{01}e^{-\lambda_2} + p_{11}e^{-\lambda}, \\
P_{II} &= p_{00} + p_{10}(1 - e^{-\lambda_2} + e^{-\lambda}) + p_{01}(1 - e^{-\lambda_1} + e^{-\lambda}) + p_{11}e^{-\lambda}, \qquad (4) \\
P_{III} &= p_{00} + p_{10}e^{-\lambda_{10}} + p_{01}e^{-\lambda_{01}} + p_{11}e^{-\lambda}
\end{aligned}
$$

이 된다. (4)식에서 모형간의 확률차이를 계산하면 (5)식과 같다.

$$
\begin{aligned}
P_{II} - P_{III} &= p_{10}(1 - e^{-\lambda_{10}})(1 - e^{-\lambda_2}) + p_{01}(1 - e^{-\lambda_{01}})(1 - e^{-\lambda_1}), \qquad (5) \\
P_{III} - P_I &= p_{10}(e^{-\lambda_{10}} - e^{-\lambda_1}) + p_{01}(e^{-\lambda_{01}} - e^{-\lambda_2}) \ > \ 0
\end{aligned}
$$

(5)식 첫 번째에서 함수 $f(x) = 1 - e^{-x}$는 $x > 0$인 경우에 항상 양의 값을 갖게 되므로 $P_{II} > P_{III}$가 된다. 또한 두 번째 식에서 $\lambda_1 > \lambda_{10}$, $\lambda_2 > \lambda_{01}$이므로 $P_{III} > P_I$된다. 그러므로 $P_I < P_{III} < P_{II}$가 되어 모형 II가 최대가 됨을 알 수 있다.

2) 주변분포

모형 I에서 이변량 영과잉-포아송분포의 주변분포는 일변량 영과잉-포아송분포가 되어 좋은 성질을 갖고 있으나 모형 II와 III은 그렇치 못하다.

3) 조건부 기대값

모형 I에서 조건부 기대값, $E(Y_1|Y_2 = k_2)$는 $k_2 > 0$인 경우 식 (3)에 의해 k_2의 선형함수로 표현된다. 또한 $k_2 = 0$인 경우, 즉 어느 한 변수의 불량품수가 없을 경우 다른 변수의 불량품수에 대한 조건부 기대값은 모형 I에서 가장 크고 모형 III 그리고 모형 II의 순으로 작아진다.

참고 문헌

1. Charalambides, CH. A. & Parageorgiou, H. (1981). Bivariate Poisson Binomial Distributions.*Biometrical Journal*, 23, pp. 437-450.

2. Lambert, D.(1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing, *Technometrics*, 34, pp. 1-14.

3. Marshall, A.M. & Olkin, I. (1985). A Family of Bivariate Distributions Generated by the Bivariate Bernoulli Distribution, *Journal of the American Statistical Association*, 80, pp. 332-338.

4. Singh, S.N.(1963). A Note on Inflated Poisson Distribution, *Journal of the Indian Statistical Association* 1, pp. 140-144

Moments of the Bivariate Zero-Inflated Poisson Distributions

Kyungmoo Kim [4] , Sungho Lee , Jong Tae Kim [5]

Abstract

Zero-Inflated Poisson models are mixed models of the Poisson and Bernoulli models. Recently Zero-Inflated Poisson distributions have been used frequently rather than previous Poisson distributions because the developement of industrial technology make few defects in manufacturing process. It is important that univariate Zero-Inflated Poisson distributions are extended to bivariate distributions to generalize the multivariate distributions. In this paper we proposed three types of the bivariate Zero-Inflated Poisson distributions and obtained these moments. We compared the three types of distributions by using the moments.

[4]Professor, Department of Statistics, Taegu University, Kyungpook, 712-714 Korea.
[5]Associate Professor