분산안정화 변환 예제 (포아송분포)

- 이번에는 분산안정화 변환에 대한 예제를 다룬다.
- 김우철 수리통계학 참고함.

연습문제 5.12.

- $X_1 \dots X_n \sim Poi(\lambda)$.
- $\sqrt{n}\big(g(\bar{X}) - g(\lambda)\big) \xrightarrow{d} Z, \quad Z \sim N(0,1)$ 을 만족하는 변환 g를 구하라.

- CLT를 쓰면

$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{d} N(0, \lambda)$$

- $g(\cdot)$이 미분가능한 함수라고 하자.
- $g(\cdot)$는 미분가능한 어떠한 형태도 가능하다. 예를들어 아래와 같은 형태들이 가능하다고 하자.

$$g(x) = x^2$$

혹은

$$g(x) = 2\sqrt{x}$$

- 예를들어서 $g(x) = 2\sqrt{x}$와 같은 형태라고 하자. 그렇다면 $x = a$에서 $g(x)$를 테일러 전개하면 아래와 같이 표현가능하다.

$$g(x) = g(a) + (x - a)g'(a) + (x - a)^2 O(1)$$

- x대신에 $\bar{X}_n$ a대신에 λ를 대입하면

$$g(\bar{X}_n) = g(\lambda) + (\bar{X}_n - \lambda)g'(\lambda) + (\bar{X}_n - \lambda)^2 O(1).$$

양변에 $\sqrt{n}$을 곱하고 정리하면

$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) = \sqrt{n}(\bar{X}_n - \lambda)g'(\lambda) + \sqrt{n}(\bar{X}_n - \lambda)^2 O(1).$$

그런데 $\sqrt{n}(\bar{X}_n - \lambda)^2 O(1) = O_p(1)o_p(1)O(1) = o_p(1)$ 이므로

$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) = \sqrt{n}(\bar{X}_n - \lambda)g'(\lambda) + o_p(1)$$

그런데 $g(x) = 2\sqrt{x}$이므로 $g'(x) = 1/\sqrt{x}$이다. 따라서 $g'(\lambda) = 1/\sqrt{\lambda}$.

- 정리하면

$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) = \sqrt{n}(\bar{X} - \lambda)1/\sqrt{\lambda} + o_p(1)$$

CLT의 결과를 쓰면

$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \xrightarrow{d} N(0, \lambda)1/\sqrt{\lambda}$$

따라서

$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \xrightarrow{d} N(0, 1)$$

(결론) 만약에 $g(x) = 2\sqrt{x}$로 설정한다면

$$\sqrt{n}\big(g(\bar{X}) - g(\lambda)\big)$$

의 극한이 표준정규분포로 수렴하므로 문제에서 원하는 답을 구할 수 있다.

- 이는 바꾸어 말하면 포아송분포에 한정하여 (1) 관측한 확률변수에서 제곱근을 취한뒤 2배를 하고 (2) 그상태에서 CLT를 쓴다면 표준정규분포로 수렴한다는 것이다.
- 이 경우 그냥 포아송분포자체에서 CLT를 쓰는것보다 이득되는 경우가 있다. (일단 이렇게만 알아둘것)

- 그렇다면 어떻게 $g(x) = 2\sqrt{x}$임을 알 수 있을까?

- g가 미분가능하다면

$$g(x) \approx g(a) + (x-a)g'(a)$$

이고 다시 정리하면

$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \approx \sqrt{n}(\bar{X}_n - \lambda)g'(\lambda)$$

따라서

$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \xrightarrow{d} N(0,\lambda)g'(\lambda)$$

이고 정규분포의 특성에 따라서

$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \xrightarrow{d} N\big(0, \lambda(g'(\lambda))^2\big)$$

이다. 따라서

$$\lambda(g'(\lambda))^2 = 1$$

이어야 할 조건은

$$g'(\lambda) = \frac{1}{\sqrt{\lambda}}$$

일 조건이다. 즉 $g(x)$의 형태는 모르겠지만 $g'(x)$는 아래와 같아야 함을 알 수 있다.

$$g'(x) = \frac{1}{\sqrt{x}}$$

이러부터 $g(x)$를 유추하면

$$g(x) = \int \frac{1}{\sqrt{x}}dx = 2\sqrt{x}.$$

- 이제 λ의 점근적인 신뢰구간을 구해보자. 아래의 식을 관찰하자.

$$2\sqrt{n}\big(\sqrt{\bar{X}_n} - \sqrt{\lambda}\big) \xrightarrow{d} N(0,1)$$

이때 $2\sqrt{n}\big(\sqrt{\bar{X}_n} - \sqrt{\lambda}\big)$자체를 하나의 통계량 Y_n으로 본다면

$$Y_n \xrightarrow{d} N(0,1)$$

Y_n이 정규분포를 따른다면 아래가 성립한다.

$$P\big(-Z_{\alpha/2} < Y_n < Z_{\alpha/2}\big) = 1-\alpha$$

Y_n이 정규분포는 아니지만 점근적으로 정규분포를 따른다면 아래가 성립한다.

$$\lim_{n\to\infty} P\big(-Z_{\alpha/2} < Y_n < Z_{\alpha/2}\big) = 1-\alpha$$

따라서

$$\lim_{n\to\infty} P\left(-Z_{\alpha/2} < 2\sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\lambda}) < Z_{\alpha/2}\right) = 1-\alpha$$

이는

$$\lim_{n\to\infty} P\left(-Z_{\alpha/2} < 2\sqrt{n}(\sqrt{\lambda} - \sqrt{\bar{X}_n}) < Z_{\alpha/2}\right) = 1-\alpha$$

정리하면

$$\lim_{n\to\infty} P\left(\frac{-Z_{\alpha/2}}{2\sqrt{n}} + \sqrt{\bar{X}_n} < \sqrt{\lambda} < \frac{Z_{\alpha/2}}{2\sqrt{n}} + \sqrt{\bar{X}_n}\right) = 1-\alpha$$

따라서

$$\lim_{n\to\infty} P\left(\Big(\frac{-Z_{\alpha/2}}{2\sqrt{n}} + \sqrt{\bar{X}_n}\Big)^2 < \lambda < \Big(\frac{Z_{\alpha/2}}{2\sqrt{n}} + \sqrt{\bar{X}_n}\Big)^2\right) = 1-\alpha$$

이 성립한다.

- 위의 식이 의미하는것은 λ가 구간

$$\left(\Big(\frac{-Z_{\alpha/2}}{2\sqrt{n}} + \sqrt{\bar{X}_n}\Big)^2, \Big(\frac{Z_{\alpha/2}}{2\sqrt{n}} + \sqrt{\bar{X}_n}\Big)^2\right)$$

에 포함될 확률이 점근적으로 $1-\alpha$에 수렴한다는 뜻이다. 이는 바꾸어 말하면 매우 큰 n에 대하여 $1-\alpha$의 확률로 모수가 위의 구간에 있음을 확신할 수 있다는 의미이다. 이를 $1-\alpha$의 점근적 신뢰구간이라고 부른다.

- 예를들어 λ의 점근적인 95퍼센트 신뢰구간은 아래와 같이 구할 수 있다.

$$\left(\Big(\frac{-1.96}{2\sqrt{n}} + \sqrt{\bar{X}_n}\Big)^2, \Big(\frac{1.96}{2\sqrt{n}} + \sqrt{\bar{X}_n}\Big)^2\right)$$