บทที่ ๓

W Test เมื่อ n มีค่ามาก

ในบทนี้เราจะกล่าวถึงการกระจายของ W เมื่อ n มีค่ามาก เราจะพบว่าสถิติ W มีการแจกแจงใกล้เคียงกับการแจกแจงปกติ ดังนั้นเราจึงอาจใช้ตารางการแจกแจงปกติประมาณค่า W_∞ ได้

๓.๑ หฤษฎีบท

ให้ W เป็นตัวแปรสุ่มดังนิยามไว้ใน (๒.๑.๑), (๒.๑.๒),(๒.๑.๓), และ (๒.๑.๔) ถ้า

(๓.๑.๑) $$U = \frac{W - E(W)}{\sqrt{Var\ (W)}}$$

เราจะได้ว่าสำหรับจำนวนจริง x ใดๆ

(๓.๑.๒) $$\lim_{n\to\infty} Prob(U \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$$

พิสูจน์

จาก (๒.๑.๒) และ (๒.๑.๓) เราจะได้ว่า W_j อาจมีค่าเป็นได้ตั้งแต่ ๐ ถึง k-1 และจาก (๒.๒.๒) ได้

(๓.๑.๓) $$E(W_j) = 0.\ p_0 + 1.\ p_1 + 2.\ p_2 + \dots + (k-1)\ p_{k-1}$$

$$= \sum_{i=0}^{k-1} i\ p_i$$

เนื่องจาก W_j มีการแจกแจงเดียวกันและไม่ขึ้นต่อกัน เพราะฉะนั้น

(๓.๑.๔) $$E(W) = E\left(\sum_{j=1}^{n} W_j\right)$$

$$= \sum_{j=1}^{n} E(W_j)$$

$$= n \sum_{i=0}^{k-1} i\, p_i$$

เพราะว่า

(๓.๑.๕) $$E(W_j^2) = 0^2 \cdot p_0 + 1^2 \cdot p_1 + 2^2 \cdot p_2 + \ldots + (k-1)^2 \cdot p_{k-1}$$

$$= \sum_{i=0}^{k-1} i^2 p_i$$

เพราะฉะนั้น

(๓.๑.๖) $$\mathrm{Var}(W_j) = E(W_j^2) - \left[E(W_j)\right]^2$$

$$= \sum_{i=0}^{k-1} i^2 p_i - \left(\sum_{i=0}^{k-1} i\, p_i\right)^2$$

ดังนั้น

(๓.๑.๗) $$\mathrm{Var}(W) = \mathrm{Var}\left(\sum W_j\right)$$

$$= \sum_{j=1}^{n} \mathrm{Var}(W_j)$$

$$= n\, \mathrm{Var}(W_j)$$

ให้

(๓.๑.๘) $$\sum_{i=0}^{k-1} i\, p_i = A$$

(๓.๑.๙) $$\sum_{i=0}^{k-1} i^2 p_i - \left(\sum_{i=0}^{k-1} i\, p_i\right)^2 = B$$

ดังนั้น

(๓.๑.๑๐) $$E\,(W) = n\,A$$

(๓.๑.๑๑) $$Var\,(W) = n\,B$$

จาก (๓.๑.๑) เราได้ moment generating function ของ U

(๓.๑.๑๒) $$M_u\,(t) = e^{-\frac{E(W)}{\sqrt{Var(W)}}t}\; M_w\left(\frac{1}{\sqrt{Var(W)}}t\right)$$

เพราะว่า w_j ไม่ขึ้นต่อกัน เพราะฉะนั้น

(๓.๑.๑๓) $$M_w\left(\frac{t}{\sqrt{Var(W)}}\right) = M_{\sum_{j=1}^{n} w_j}\left(\frac{t}{\sqrt{Var\,(W)}}\right)$$

$$= \prod_{j=1}^{n} M_{w_j}\left(\frac{t}{\sqrt{Var\,(W)}}\right)$$

$$= \left[M_{w_j} \left(\frac{t}{\sqrt{Var(\overline{W})}} \right) \right]^n$$

$$= \left[\sum_{i=0}^{k-1} p_i \, e^{\frac{i\,t}{\sqrt{Var(\overline{W})}}} \right]^n$$

ดังนั้น

(๓.๑.๑๔) $$M_u(t) = e^{-\frac{E(\overline{W})}{\sqrt{Var(\overline{W})}} t} \left[\sum_{i=0}^{k-1} p_i \, e^{\frac{i\,t}{\sqrt{Var(\overline{W})}}} \right]^n$$

แทนค่า $E(\overline{W})$ และ $Var(\overline{W})$ จาก (๓.๑.๑๐) และ (๓.๑.๑๑) จะได้

(๓.๑.๑๕) $$M_u(t) = e^{-\frac{nA}{\sqrt{nB}} t} \left[\sum_{i=0}^{k-1} p_i \, e^{\frac{it}{\sqrt{nB}}} \right]^n$$

$$= \left[\sum_{i=0}^{k-1} p_i \, e^{\frac{(-A+i)}{\sqrt{nB}} t} \right]^n$$

$$\ln M_u(t) = n \ln \left[\sum_{i=0}^{k-1} p_i \, e^{\frac{(-A+i)}{\sqrt{nB}} t} \right]$$

$$= \ln\left(\sum_{i=0}^{k-1} p_i \, e^{\frac{(-A+i)t}{\sqrt{nB}}}\right) \Bigg/ \frac{1}{n}$$

ใช้ L' Hospital's Rule เราจะได้

$$\lim_{n\to\infty} \ln M_u(t) = \lim_{n\to\infty} \frac{\left(\sum_{i=0}^{k-1} p_i \, e^{\frac{(-A+i)t}{\sqrt{nB}}} \frac{(-A+i)}{\sqrt{B}} t \left(-\frac{1}{2n^{3/2}}\right)\right) \Bigg/ \left(\sum_{i=0}^{k-1} p_i e^{\frac{(-A+i)t}{\sqrt{nB}}}\right)}{-\frac{1}{n^2}}$$

$$= \lim_{n\to\infty} \frac{t}{2\sqrt{B}} \frac{\left(\sum_{i=0}^{k-1} p_i e^{\frac{(-A+i)t}{\sqrt{nB}}} (-A+i)\right) \Bigg/ \left(\sum_{i=0}^{k-1} p_i e^{\frac{(-A+i)t}{\sqrt{nB}}}\right)}{n^{-1/2}}$$

$$= \frac{t}{2\sqrt{B}} \lim_{n\to\infty} \frac{-A + \left(\sum_{i=0}^{k-1} i \, p_i e^{\frac{(-A+i)t}{\sqrt{nB}}}\right) \Bigg/ \left(\sum_{i=0}^{k-1} p_i e^{\frac{(-A+i)t}{\sqrt{nB}}}\right)}{n^{-1/2}}$$

ใช้ L' Hospital's Rule อีกครั้งหนึ่ง เราจะได้

$$\lim_{n\to\infty} \ln M_u(t) = \frac{t}{2\sqrt{B}} \lim_{n\to\infty} \left\{ \left(\sum_{i=0}^{k-1} p_i e^{\frac{(-A+i)t}{\sqrt{nB}}}\right)^{-1} \left[\sum_{i=0}^{k-1} i \, p_i e^{\frac{(-A+i)t}{\sqrt{nB}}} \frac{(-A+i)t}{\sqrt{B}} \left(-\frac{1}{2} n^{-3/2}\right)\right]\right.$$

$$+ \left(\sum_{i=0}^{k-1} i\, p_i e^{\frac{(-A+i)t}{\sqrt{nB}}} \right) \left(\sum_{i=0}^{k-1} p_i e^{\frac{(-A+i)t}{\sqrt{nB}}} \right)^{-2} .$$

$$\left. \left(\sum_{i=0}^{k-1} p_i e^{\frac{(-A+i)t}{\sqrt{nB}}} \frac{(-A+i)t}{\sqrt{B}} \left(-\frac{1}{2} n^{-3/2} \right) \right) \right\} \Big/ \left(-\frac{1}{2} n^{-3/2} \right)$$

$$= \frac{t^2}{2B} \left(\sum_{i=0}^{k-1} i\, p_i(-A+i) - \sum_{i=0}^{k-1} i\, p_i \sum_{i=0}^{k-1} (-A+i)\, p_i \right)$$

$$= \frac{t^2}{2B} \left(\sum_{i=0}^{k-1} i^2\, p_i - \left\{ \sum_{i=0}^{k-1} i\, p_i \right\}^2 \right)$$

$$= \frac{t^2}{2B} \cdot B$$

$$= \frac{1}{2} t^2$$

นั่นคือ $\lim_{n\to\infty} M_u(t) = e^{\frac{t^2}{2}}$

ดังนั้น $\lim_{n\to\infty} \text{Prob}(W \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

๓.๒ การใช้การแจกแจงปกติประมาณค่า W_α

จาก (๓.๑) เราสรุปได้ว่าเมื่อ n มีค่ามากนั้น W มีการแจกแจงใกล้เคียงกับการแจกแจงปกติ ดังนั้นเราจึงอาจประมาณค่าของ W_α ได้โดยใช้การแจกแจงปกติ ถ้า W_α เป็นค่าวิกฤติ เราจะได้

$$\text{Prob}\,(W > W_\alpha) = \text{Prob}\left[\frac{W - E(W)}{\sqrt{\text{Var}(W)}} > \frac{W_\alpha - E(W)}{\sqrt{\text{Var}(W)}}\right]$$

$$= \text{Prob}\left[U > \frac{W_\alpha - E(W)}{\sqrt{\text{Var}\,(W)}}\right]$$

ดังนั้นเราอาจประมาณ W_α โดยใช้ W_α^* ซึ่ง

(๓.๒.๑) $$\frac{W_\alpha^* - E(W)}{\sqrt{\text{Var}\,(W)}} = Z_\alpha$$

โดยที่ Z_α เป็นจำนวนซึ่ง

(๓.๒.๒) $$\int_{Z_\alpha}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt = \alpha$$

จาก (๓.๒.๑) เราได้

(๓.๒.๓) $$W_\alpha^* = E(W) + Z_\alpha\sqrt{\text{Var}(W)}$$

ถ้า H_o จริงเราได้ $p_i = \frac{1}{k}$

ดังนั้น

(๓.๒.๔) $$E(W) = n \sum_{i=0}^{k-1} i\, p_i$$

$$= n \sum_{i=0}^{k-1} i.\ \frac{1}{k}$$

$$= \frac{n}{k} \sum_{i=0}^{k-1} i$$

$$= \frac{n}{k}\, \frac{(k-1)(k)}{2}$$

$$= \frac{n\,(k-1)}{2}$$

และได้

(๓.๒.๕) $$Var(W) = n \left\{ \sum_{i=0}^{k-1} i^2\, p_i - \left(\sum_{i=0}^{k-1} i\, p_i \right)^2 \right\}$$

$$= n \left\{ \sum_{i=0}^{k-1} i^2\ \frac{1}{k} - \left(\sum_{i=0}^{k-1} i\ \frac{1}{k} \right)^2 \right\}$$

$$= \frac{n}{k} \left\{ \sum_{i=0}^{k-1} i^2 - \frac{1}{k} \left(\sum_{i=0}^{k-1} i \right)^2 \right\}$$

$$= \frac{n}{k} \left\{ \frac{(k-1)(k)(2k-1)}{6} - \frac{(k-1)^2\, k^2}{4k} \right\}$$

$$= \frac{n\ k\ (k-1)}{12\ k} \left\{ 2\ (2k-1) - 3\ (k-1) \right\}$$

$$= \frac{n\ (k-1)\ (k+1)}{12}$$

แทนค่า (๓.๒.๔) และ (๓.๒.๕) ใน (๓.๒.๓)

(๓.๒.๖) $$W^*_{\alpha} = Z_{\alpha} \sqrt{\frac{n\ (k-1)\ (k+1)}{12}} + \frac{n\ (k-1)}{2}$$

ตารางต่อไปนี้เปรียบเทียบค่าวิกฤติที่แท้จริง (W_{α}) กับค่าวิกฤติที่ได้จากการประมาณการแจกแจงของ W ด้วยการแจกแจงปกติ (W^*_{α})

ตารางที่ ๓

ค่าวิกฤติ W_α และ W^*_α

n.	$W_{.01}$ (exact)	$W^*_{.01}$ (normal approx)	$W_{.05}$ (exact)	$W^*_{.05}$ (normal approx)	n.	$W_{.01}$ (exact)	$W^*_{.01}$ (normal approx)	$W_{.05}$ (exact)	$W^*_{.05}$ (normal approx)
6	11	11	10	10	44	63	57	60	53
8	14	14	13	12	46	65	59	63	56
10	17	17	15	15	48	68	62	65	58
12	19	19	18	17	50	73	64	71	60
14	22	22	20	20	52	75	66	73	62
16	24	24	22	22	54	77	68	76	64
18	27	27	25	24	56	79	71	78	67
20	29	29	27	26	58	81	73	80	69
22	32	31	29	29	60	84	75	82	71
24	34	34	31	31	62	86	77	84	73
26	37	36	34	33	64	89	80	87	75
28	39	39	36	36	66	93	82	91	77
30	44	41	41	38	68	97	84	94	80
32	47	43	44	40	70	100	86	98	82
34	50	46	47	42	72	103	89	100	84
36	52	48	50	45	74	105	91	103	86
38	55	50	53	47	76	108	93	106	88
40	58	53	55	49	78	111	95	108	90
42	60	55	58	51	80	112	98	111	93