Chapter 3

An approximation to a percentage point of the distribution of the kurtosis statistic b_2

3.1. Introduction

In a statistical inference, an optimal procedure based on normal samples is not always good in the case where the population distribution is not normal. In order to improve it in such a case, the robustness is variously discussed (see, e.g. Andrews et al.([ABH72])). Nevertheless, as a base of the inference, it is important to test departure from normality. In such a case, skewness and kurtosis statistics are well known as useful ones in testing it ([PR96]). Then it is necessary to use percentage point of their distributions, but it is difficult to obtain their exact distributions. So, the approximate values of percentage points are given by Pearson ([P65]) and Pearson and Hartley ([PH76]) using Pearson type curves.

In this chapter, in a similar way to Akahira ([Ak95]) who obtained a percentage point of the non-central t-distribution, we get approximation formulae of a percentage point of the distribution of the kurtosis statistic b_2 using the Cornish-Fisher expansion after a $1/k$-th power transformation for $k > 0$ ([TA97]). We also compare its approximate values with those of Pearson and Hartley ([PH76]). It consequently shows that the values in case of $k = 2$ are comparatively stable ([TA97]).

3.2. Skewness and kurtosis statistics

In test for departure from normality, it is necessary to use percentage points of the distributions of skewness and kurtosis statistics. Suppose that $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables according to a distribution function

$F(x)$. Then the skewness statistic $\sqrt{b_1}$ and the kurtosis one b_2 are given by $\sqrt{b_1} = m_3 \big/ m_2^{3/2}$ and $b_2 = m_4 / m_2^2$, respectively, where $m_r = (1/n) \sum_{i=1}^n (X_i - \overline{X})^r$ $(r = 2, 3, 4)$ with $\overline{X} = (1/n) \sum_{i=1}^n X_i$. Let $\sqrt{\beta_1} = \mu_3 \big/ \mu_2^{3/2}$ and $\beta_2 = \mu_4 / \mu_2^2$ with $\mu_r = E[(X_1 - \mu)^r]$ $(r = 2, 3, 4)$. Then the statistics $\sqrt{b_1}$ and b_2 have the consistency for $\sqrt{\beta_1}$ and β_2, respectively. Note that $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$ if the underlying distribution is normal.

Now we consider a problem of testing the hypothesis $H : F(x) \equiv \Phi((x - \mu)/\sigma)$, where $\Phi(\,\cdot\,)$ is the standard normal distribution function. In order to treat the problem it is convenient to use the statistics $\sqrt{b_1}$ and b_2. Since, under the hypothesis H, the distributions of $\sqrt{b_1}$ and b_2 are independent of μ and σ, they are determined in accordance with only n. In testing the hypothesis H, we see that $|\sqrt{b_1}| > c$ or $|b_2 - 3| > c'$ are usually taken as rejection regions, where c and c' are certain constants. In such a case, it is necessary to use percentage points of the distributions of $\sqrt{b_1}$ and b_2. Where we directly apply the Cornish-Fisher expansion for $\sqrt{b_1}$ and b_2, it is seen that a percentage of the distribution of $\sqrt{b_1}$ seems to be better, but that of b_2 not so (see, e.g. [S81]). It is also known that the distributions can be approximated by Pearson type curves and Johnson's S_U distribution by equating the first four moments (see, e.g. [Jo49], [P65]) and [PH76]).

3.3. Approximation formulae of a percentage point of the distribution of b_2

In this section we consider the distribution of b_2 under the hypothesis H. Since the distribution of b_2 is independent of μ and σ, we assume that $\mu = 0$ and $\sigma = 1$. Let k be a fixed positive number. Then we consider a $1/k$-th power transformation so that

$$P\{b_2 \le c\} = P\left\{\frac{m_4}{m_2^2} \le c\right\} = P\left\{\left(\frac{m_4}{m_2^2}\right)^{1/k} \le c^{1/k}\right\} \tag{3.1}$$
$$= P\left\{m_4^{1/k} - c^{1/k} m_2^{2/k} \le 0\right\}.$$

Putting

$$Y := m_4^{1/k} - c^{1/k} m_2^{2/k}, \tag{3.2}$$

we obtain the moments of Y about the origin. Since b_2 and m_2 are stochastically independent under the hypothesis H, it follows that for any $\alpha \ge 0$ and any $\beta > 0$

$$E\left[m_4^\alpha m_2^\beta\right] = E\left[b_2^\alpha m_2^{2\alpha+\beta}\right] = E[b_2^\alpha]\, E\left[m_2^{2\alpha+\beta}\right]. \tag{3.3}$$

By Pearson ([P30]) and Hsu and Lawley ([HL39]) we have

$$\mu_0 := E(b_2) = \frac{3(n-1)}{n+1} = 3 - \frac{6}{n+1},$$

$$\gamma_2 := E\left[\left(\frac{b_2-\mu_0}{\mu_0}\right)^2\right] = \frac{8n(n-2)(n-3)}{3(n-1)^2(n+3)(n+5)},$$

$$\gamma_3 := E\left[\left(\frac{b_2-\mu_0}{\mu_0}\right)^3\right] = \frac{64n(n-2)(n-3)(n^2-5n+2)}{(n-1)^3(n+3)(n+5)(n+7)(n+9)},$$

$$\gamma_4 := E\left[\left(\frac{b_2-\mu_0}{\mu_0}\right)^4\right]$$

$$= \frac{64n(n-2)(n-3)(n^5+207n^4-1707n^3+4105n^2-1902n+720)}{3(n-1)^4(n+3)(n+5)(n+7)(n+9)(n+11)(n+13)},$$

$$\gamma_5 := E\left[\left(\frac{b_2-\mu_0}{\mu_0}\right)^5\right] = O\left(\frac{1}{n^3}\right).$$

Hence we obtain for any $\alpha > 0$

$$E\left(b_2^\alpha\right) = E\left[\mu_0^\alpha\left(1+\frac{b_2-\mu_0}{\mu_0}\right)^\alpha\right]$$
$$= \mu_0^\alpha\left\{1 + \frac{\alpha(\alpha-1)}{2}\gamma_2 + \frac{\alpha(\alpha-1)(\alpha-2)}{6}\gamma_3\right.$$
$$\left. + \frac{1}{24}\alpha(\alpha-1)(\alpha-2)(\alpha-3)\gamma_4 + O\left(\frac{1}{n^3}\right)\right\}. \quad (3.4)$$

On the other hand, we have for any $\ell \geq 0$

$$E\left[m_2^\ell\right] = \left(\frac{2}{n}\right)^\ell \frac{\Gamma\left(\frac{n-1}{2}+\ell\right)}{\Gamma\left(\frac{n-1}{2}\right)}, \quad (3.5)$$

where $\Gamma(\,\cdot\,)$ is the Gamma function. Let $a_k = c^{1/k}$. Then it follows from (3.2) and (3.3) that the moments of Y about the origin are given by

$$\mu_1 := E(Y) = E\left(m_2^{2/k}\right)\left\{E\left(b_2^{1/k}\right) - a_k\right\},$$
$$\mu_2 := E(Y^2) = E\left(m_2^{4/k}\right)\left\{E\left(b_2^{2/k}\right) - 2a_kE\left(b_2^{1/k}\right) + a_k^2\right\},$$
$$\mu_3 := E(Y^3) = E\left(m_2^{6/k}\right)\left\{E\left(b_2^{3/k}\right) - 3a_kE\left(b_2^{2/k}\right) + 3a_k^2E\left(b_2^{1/k}\right) - a_k^3\right\},$$
$$\mu_4 := E(Y^4) = E\left(m_2^{8/k}\right)\left\{E\left(b_2^{4/k}\right) - 4a_kE\left(b_2^{3/k}\right) + 6a_k^2E\left(b_2^{2/k}\right) - 4a_k^3E\left(b_2^{1/k}\right) + a_k^4\right\}.$$

and their actual values are approximated by (3.4) and (3.5).

From (3.1) we have for any α $(0 < \alpha < 1)$

$$1 - \alpha = P\{b_2 \leq c_\alpha\} = P\{Y \leq 0\} = P\left\{\frac{Y-\mu_1}{\sigma_Y} \leq -\frac{\mu_1}{\sigma_Y}\right\}, \quad (3.6)$$

where $\sigma_Y := \sqrt{\sigma_Y^2} = \sqrt{\mathrm{Var}(Y)} = \sqrt{\mu_2 - \mu_1^2}$. Putting

$$W := \frac{Y - \mu_1}{\sigma_Y} \tag{3.7}$$

we obtain $E(W)=0$, $\mathrm{Var}(W)=1$, and the third and fourth cumulants are given by

$$\kappa_3(W) = \frac{1}{\sigma_Y^3}\left(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3\right),$$
$$\kappa_4(W) = \frac{1}{\sigma_Y^4}\left(\mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4\right).$$

Using the Cornish-Fisher expansion for W, we can obtain an approximation formula of a percentage point of the distribution of b_2.

Theorem 3.1 *The upper 100α percentile c_α of the distribution of b_2 can be derived from the formula*

$$\begin{aligned} -\frac{\mu_1}{\sigma_Y} &= u_\alpha + \frac{1}{6}\kappa_3(W)\left(u_\alpha^2 - 1\right) \\ &\quad + \frac{1}{24}\kappa_4(W)\left(u_\alpha^3 - 3u_\alpha\right) - \frac{1}{36}\kappa_3^2(W)\left(2u_\alpha^3 - 5u_\alpha\right) + O\left(\frac{1}{n^3}\right), \end{aligned} \tag{3.8}$$

where u_α is the upper 100α percentile of the standard normal distribution.

The proof straightforwardly follows from (3.6), (3.7) and the Cornish-Fisher expansion for W with the above cumulants.

Corollary 3.1 *The upper 100α percentile c_α of the distribution of b_2 can be approximately derived from the formula*

$$\left\{\mu_1\sigma_Y^2 + \frac{1}{6}\sigma_Y^3\kappa_3(W)\left(u_\alpha^2 - 1\right)\right\}^2 = \sigma_Y^6 u_\alpha^2. \tag{3.9}$$

The formula (3.9) easily follows from (3.8).

Remark The lower 100α percentile of the distribution of b_2 can be derived from the formulae (3.8) and (3.9) with $1-\alpha$ instead α in them.

3.4. Numerical comparison

First, in case of $k = 2, 4, 6, 8$, we calculate the values of upper (lower) 1 and 5 percentile using the formula (3.9). In order to compare the values by Pearson and Hartley ([PH76]) who obtained for $n = 20(10)50$, $75(25)200$, 250, $300(100)1000$, 2000 with ours, we also get differences between them and ours (see Figures 3.1 to 3.8). As is seen from the Figures, the values in case of $k = 2$ are seen to be comparatively stable.

Table 3.1: The values of upper percentile ($k = 2$)

n	upper 1 percentile			upper 5 percentile		
	[PH76]	(3.9)	Difference	[PH76]	(3.9)	Difference
20	5.38	4.58207	−0.798	4.18	4.10029	−0.080
30	5.20	4.59818	−0.602	4.12	4.07380	−0.046
40	5.04	4.56571	−0.474	4.06	4.02811	−0.032
50	4.88	4.51570	−0.364	4.00	3.98064	−0.019
75	4.59	4.37668	−0.213	3.87	3.87473	0.005
100	4.39	4.25199	−0.138	3.77	3.79075	0.021
125	4.24	4.14828	−0.092	3.70	3.72435	0.024
150	4.13	4.06259	−0.067	3.65	3.67086	0.021
175	4.04	3.99114	−0.049	3.61	3.62686	0.017
200	3.98	3.93079	−0.049	3.57	3.58999	0.020
250	3.87	3.83450	−0.035	3.52	3.53148	0.011
300	3.79	3.76092	−0.029	3.47	3.48685	0.017
400	3.67	3.65518	−0.015	3.41	3.42261	0.013
500	3.60	3.58210	−0.018	3.37	3.37795	0.008
600	3.54	3.52802	−0.012	3.34	3.34470	0.005
700	3.50	3.48606	−0.014	3.31	3.31875	0.009
800	3.46	3.45235	−0.008	3.29	3.29778	0.008
900	3.43	3.42453	−0.005	3.28	3.28040	0.000
1000	3.41	3.40110	−0.009	3.26	3.26569	0.006
2000	3.28	3.27646	−0.004	3.18	3.18626	0.006

Table 3.2: The values of lower percentile ($k = 2$)

n	lower 1 percentile			lower 5 percentile		
	[PH76]	(3.9)	Difference	[PH76]	(3.9)	Difference
20	1.64	1.37546	−0.265	1.83	1.58608	−0.244
30	1.79	1.56360	−0.226	1.98	1.78452	−0.195
40	1.89	1.72798	−0.162	2.07	1.93079	−0.139
50	1.95	1.85523	−0.095	2.15	2.03849	−0.112
75	2.08	2.06060	−0.019	2.27	2.21051	−0.059
100	2.18	2.17941	−0.001	2.35	2.31224	−0.038
125	2.24	2.25732	0.017	2.40	2.38067	−0.019
150	2.29	2.31334	0.023	2.45	2.43075	−0.019
175	2.34	2.35627	0.016	2.48	2.46954	−0.010
200	2.37	2.39070	0.021	2.51	2.50081	−0.009
250	2.42	2.44350	0.024	2.55	2.54879	−0.001
300	2.46	2.48296	0.023	2.59	2.58450	−0.005
400	2.52	2.53974	0.020	2.64	2.63529	−0.005
500	2.57	2.57984	0.010	2.67	2.67054	0.001
600	2.60	2.61035	0.010	2.70	2.69695	−0.003
700	2.62	2.63468	0.015	2.72	2.71773	−0.002
800	2.65	2.65471	0.005	2.74	2.73465	−0.005
900	2.66	2.67162	0.012	2.75	2.74879	−0.001
1000	2.68	2.68615	0.006	2.76	2.76084	0.001
2000	2.77	2.76880	−0.001	2.83	2.82760	−0.002

Table 3.3: The values of upper percentile ($k = 4$)

	upper 1 percentile			upper 5 percentile		
n	[PH76]	(3.9)	Difference	[PH76]	(3.9)	Difference
20	5.38	4.27664	-1.103	4.18	3.92368	-0.256
30	5.20	4.29872	-0.901	4.12	3.92393	-0.196
40	5.04	4.31683	-0.723	4.06	3.90484	-0.155
50	4.88	4.31751	-0.562	4.00	3.88010	-0.120
75	4.59	4.26532	-0.325	3.87	3.81246	-0.058
100	4.39	4.18537	-0.205	3.77	3.74939	-0.021
125	4.24	4.10564	-0.134	3.70	3.69509	-0.005
150	4.13	4.03374	-0.096	3.65	3.64910	-0.001
175	4.04	3.97074	-0.069	3.61	3.61005	0.000
200	3.98	3.91586	-0.064	3.57	3.57661	0.007
250	3.87	3.82585	-0.044	3.52	3.52239	0.002
300	3.79	3.75551	-0.034	3.47	3.48026	0.010
400	3.67	3.65276	-0.017	3.41	3.41865	0.009
500	3.60	3.58089	-0.019	3.37	3.37529	0.005
600	3.54	3.52740	-0.013	3.34	3.34278	0.003
700	3.50	3.48574	-0.014	3.31	3.31729	0.007
800	3.46	3.45220	-0.008	3.29	3.29663	0.007
900	3.43	3.42449	-0.006	3.28	3.27947	-0.001
1000	3.41	3.40113	-0.009	3.26	3.26492	0.005
2000	3.28	3.27658	-0.003	3.18	3.18603	0.006

Table 3.4: The values of lower percentile ($k = 4$)

	lower 1 percentile			lower 5 percentile		
n	[PH76]	(3.9)	Difference	[PH76]	(3.9)	Difference
20	1.64	1.12160	-0.518	1.83	1.50829	-0.322
30	1.79	1.32690	-0.463	1.98	1.70198	-0.278
40	1.89	1.51761	-0.372	2.07	1.85739	-0.213
50	1.95	1.67371	-0.276	2.15	1.97656	-0.173
75	2.08	1.93836	-0.142	2.27	2.17146	-0.099
100	2.18	2.09453	-0.085	2.35	2.28669	-0.063
125	2.24	2.19561	-0.044	2.40	2.36307	-0.037
150	2.29	2.26659	-0.023	2.45	2.41807	-0.032
175	2.34	2.31965	-0.020	2.48	2.46006	-0.020
200	2.37	2.36121	-0.009	2.51	2.49351	-0.016
250	2.42	2.42313	0.003	2.55	2.54415	-0.006
300	2.46	2.46799	0.008	2.59	2.58134	-0.009
400	2.52	2.53058	0.011	2.64	2.63361	-0.006
500	2.57	2.57361	0.004	2.67	2.66954	-0.000
600	2.60	2.60580	0.006	2.70	2.69631	-0.004
700	2.62	2.63118	0.011	2.72	2.71729	-0.003
800	2.65	2.65194	0.002	2.74	2.73434	-0.006
900	2.66	2.66935	0.009	2.75	2.74856	-0.001
1000	2.68	2.68425	0.004	2.76	2.76067	0.001
2000	2.77	2.76822	-0.002	2.83	2.82759	-0.002

Table 3.5: The values of upper percentile ($k = 6$)

n	upper 1 percentile			upper 5 percentile		
	[PH76]	(3.9)	Difference	[PH76]	(3.9)	Difference
20	5.38	4.18014	−1.200	4.18	3.88490	−0.295
30	5.20	4.19615	−1.004	4.12	3.88550	−0.234
40	5.04	4.22436	−0.816	4.06	3.86991	−0.190
50	4.88	4.23921	−0.641	4.00	3.84980	−0.150
75	4.59	4.21682	−0.373	3.87	3.79224	−0.078
100	4.39	4.15460	−0.235	3.77	3.73556	−0.034
125	4.24	4.08507	−0.155	3.70	3.68518	−0.015
150	4.13	4.01930	−0.111	3.65	3.64170	−0.008
175	4.04	3.96019	−0.080	3.61	3.60432	−0.006
200	3.98	3.90790	−0.072	3.57	3.57204	0.002
250	3.87	3.82094	−0.049	3.52	3.51929	−0.001
300	3.79	3.75225	−0.038	3.47	3.47802	0.008
400	3.67	3.65108	−0.019	3.41	3.41731	0.007
500	3.60	3.57991	−0.020	3.37	3.37440	0.004
600	3.54	3.52677	−0.013	3.34	3.34214	0.002
700	3.50	3.48532	−0.015	3.31	3.31681	0.007
800	3.46	3.45191	−0.008	3.29	3.29625	0.006
900	3.43	3.42428	−0.006	3.28	3.27916	−0.001
1000	3.41	3.40097	−0.009	3.26	3.26466	0.005
2000	3.28	3.27656	−0.003	3.18	3.18596	0.006

Table 3.6: The values of lower percentile ($k = 6$)

n	lower 1 percentile			lower 5 percentile		
	[PH76]	(3.9)	Difference	[PH76]	(3.9)	Difference
20	1.64	1.07359	−0.566	1.83	1.48775	−0.342
30	1.79	1.27566	−0.514	1.98	1.67920	−0.301
40	1.89	1.46805	−0.422	2.07	1.83609	−0.234
50	1.95	1.62832	−0.322	2.15	1.95785	−0.192
75	2.08	1.90497	−0.175	2.27	2.15888	−0.111
100	2.18	2.07032	−0.110	2.35	2.27816	−0.072
125	2.24	2.17760	−0.062	2.40	2.35706	−0.043
150	2.29	2.25276	−0.037	2.45	2.41367	−0.036
175	2.34	2.30873	−0.031	2.48	2.45673	−0.023
200	2.37	2.35238	−0.018	2.51	2.49091	−0.019
250	2.42	2.41700	−0.003	2.55	2.54247	−0.008
300	2.46	2.46348	0.003	2.59	2.58018	−0.010
400	2.52	2.52783	0.008	2.64	2.63298	−0.007
500	2.57	2.57174	0.002	2.67	2.66915	−0.001
600	2.60	2.60444	0.004	2.70	2.69605	−0.004
700	2.62	2.63014	0.010	2.72	2.71711	−0.003
800	2.65	2.65111	0.001	2.74	2.73421	−0.006
900	2.66	2.66868	0.009	2.75	2.74846	−0.002
1000	2.68	2.68370	0.004	2.76	2.76059	0.001
2000	2.77	2.76805	−0.002	2.83	2.82758	−0.002

Table 3.7: The values of upper percentile ($k = 8$)

n	upper 1 percentile			upper 5 percentile		
	[PH76]	(3.9)	Difference	[PH76]	(3.9)	Difference
20	5.38	4.13500	−1.245	4.18	3.86949	-0.311
30	5.20	4.14659	−1.053	4.12	3.86883	-0.251
40	5.04	4.17822	−0.862	4.06	3.85396	-0.206
50	4.88	4.19915	−0.681	4.00	3.83554	-0.164
75	4.59	4.19099	−0.399	3.87	3.78236	-0.088
100	4.39	4.13783	−0.252	3.77	3.72870	-0.041
125	4.24	4.07369	−0.166	3.70	3.68024	-0.020
150	4.13	4.01123	−0.119	3.65	3.63799	-0.012
175	4.04	3.95424	−0.086	3.61	3.60145	-0.009
200	3.98	3.90336	−0.077	3.57	3.56975	-0.000
250	3.87	3.81811	−0.052	3.52	3.51774	-0.002
300	3.79	3.75034	−0.040	3.47	3.47689	0.007
400	3.67	3.65007	−0.020	3.41	3.41664	0.007
500	3.60	3.57931	−0.021	3.37	3.37395	0.004
600	3.54	3.52638	−0.014	3.34	3.34182	0.002
700	3.50	3.48505	−0.015	3.31	3.31656	0.007
800	3.46	3.45171	−0.008	3.29	3.29606	0.006
900	3.43	3.42413	−0.006	3.28	3.27901	-0.001
1000	3.41	3.40086	−0.010	3.26	3.26454	0.005
2000	3.28	3.27655	−0.003	3.18	3.18592	0.006

Table 3.8: The values of lower percentile ($k = 8$)

n	lower 1 percentile			lower 5 percentile		
	[PH76]	(3.9)	Difference	[PH76]	(3.9)	Difference
20	1.64	1.05357	−0.586	1.83	1.47822	−0.352
30	1.79	1.25399	−0.536	1.98	1.66862	−0.311
40	1.89	1.44654	−0.443	2.07	1.82604	−0.244
50	1.95	1.60816	−0.342	2.15	1.94889	−0.201
75	2.08	1.88961	−0.190	2.27	2.15269	−0.117
100	2.18	2.05898	−0.121	2.35	2.27391	−0.076
125	2.24	2.16907	−0.071	2.40	2.35404	−0.046
150	2.29	2.24618	−0.044	2.45	2.41144	−0.039
175	2.34	2.30351	−0.036	2.48	2.45503	−0.025
200	2.37	2.34815	−0.022	2.51	2.48959	−0.020
250	2.42	2.41406	−0.006	2.55	2.54161	−0.008
300	2.46	2.46131	0.001	2.59	2.57958	−0.010
400	2.52	2.52651	0.007	2.64	2.63265	−0.007
500	2.57	2.57084	0.001	2.67	2.66895	−0.001
600	2.60	2.60379	0.004	2.70	2.69591	−0.004
700	2.62	2.62965	0.010	2.72	2.71702	−0.003
800	2.65	2.65072	0.001	2.74	2.73414	−0.006
900	2.66	2.66836	0.008	2.75	2.74841	−0.002
1000	2.68	2.68343	0.003	2.76	2.76055	0.001
2000	2.77	2.76797	−0.002	2.83	2.82757	−0.002