Continuity Correction

Rabi Bhattacharya, The University of Arizona, USA

According to the central limit theorem (CLT), the distribution function F_n of a normalized sum $n^{-1/2}(X_1 + ... + X_n)$ of n independent random variables $X_1, ..., X_n$, having a common distribution with mean zero and variance $\sigma^2 > 0$, converges to the distribution function Φ_σ of the normal distribution with mean zero and variance σ^2, as $n \to \infty$. We will write Φ for Φ_1 for the case $\sigma = 1$. The densities of Φ_σ and Φ are denoted by ϕ_σ and ϕ, respectively. In the case $X_j's$ are discrete, F_n has jumps and the normal approximation is not very good when n is not sufficiently large. This is a problem which most commonly occurs in statistical tests and estimation involving the normal approximation to the binomial and, in its multi-dimensional version, in Pearson's frequency chisquare tests, or in tests for association in categorical data. Applying the CLT to a binomial random variable T with distribution $B(n,p)$, with mean np and variance $npq(q = 1 - p)$, the normal approximation is given , for integers $0 \le a \le b \le n$, by

$$P(a \le T \le b) \approx \Phi((b - np)/\sqrt{npq}) - \Phi((a - np)/\sqrt{npq}). \tag{1}$$

Here $\approx$ indicates that the difference between its two sides goes to zero as $n \to \infty$. In particular, when $a = b$, the binomial probability $P(T = b) = C^n_b p^b q^{n-b}$ is approximated by zero. This error is substantial if n is not very large. One way to improve the approximation is to think graphically of each integer value b of T being uniformly spread over the interval $[b - \frac{1}{2}, b + \frac{1}{2}]$. This is the so called *histogram approximation*, and leads to the *continuity correction* given by replacing $\{a \le T \le b\}$ by $\{a - \frac{1}{2} \le T \le b + \frac{1}{2}\}$

$$P(a - \frac{1}{2} \le T \le b + \frac{1}{2}) \approx \Phi((b + \frac{1}{2} - np)/\sqrt{npq}) - \Phi((a - \frac{1}{2} - np)/\sqrt{npq}). \tag{2}$$

To give an idea of the improvement due to this correction, let $n = 20, p = .4$. Then $P(T \le 7) = .4159$, whereas the approximation (1) gives a probability $\Phi(-.4564) = .3240$, and the continuity correction (2) yields $\Phi(-.2282) = .4177$. Analogous continuity corrections apply to the Poisson distribution with a large mean.

For a precise mathematical justification of the continuity correction consider, in general, i.i.d. integer-valued random variables $X_1, ..., X_n$, with lattice span 1, mean μ, variance σ^2, and finite moments of order at least four. The distribution function $F_n(x)$ of $n^{-1/2}(X_1 + ... + X_n)$ may then be approximated by the *Edgeworth expansion* (See Bhattacharya and Ranga Rao (1976), p. 239, or Gnedenko and Kolmogorov (1954), p. 213)

$$F_n(x) = \Phi_\sigma(x) - n^{-\frac{1}{2}} S_1(n\mu + n^{\frac{1}{2}}x)\phi_\sigma(x) + n^{-\frac{1}{2}}\mu_3/(6\sigma^3)(1 - x^2/\sigma^2)\phi_\sigma(x) + O(n^{-1}), \tag{3}$$

where $S_1(y)$ is the right continuous periodic function $y - \frac{1}{2}$ (mod 1) which vanishes when $y = \frac{1}{2}$. Thus, when a is an integer and $x = (a - n\mu)/\sqrt{n}$, replacing a by $a + \frac{1}{2}$ (or $a - \frac{1}{2}$) on the right side of (3) gets rid of the discontinuous term involving S_1.

Consider next the continuity correction for the *(Mann-Whitney-) Wilcoxon two sample test.* Here one wants to test nonparametrically if one distribution G is stochastically larger than another distribution F, with distribution functions $G(.)$, $F(.)$. Then the null hypothesis is $H_0 : F(x) = G(x)$ for all x, and the alternative is $H_1 : G(x) \leq F(x)$ for all x, with strict inequality for some x. The test is based on independent random samples $X_1, ..., X_m$ and $Y_1, ..., Y_n$ from the two unknown continuous distributions F and G, respectively. The test statistic is $W_s =$ the sum of the ranks of the $Y_j' s$ in the combined sample of $m+n$ $X_i' s$ and $Y_j' s$. The test rejects H_0 if $W_s \geq c$, where c is chosen such that the probability of rejection under H_0 is a given level α. It is known (see Lehmann (1975), pp. 5-18) that W_s is asymptotically normal and $E(W_s) = \frac{1}{2}n(m + n + 1)$, $Var(W_s) = mn(m + n + 1)/12$. Since W_s is integer-valued, the continuity correction yields

$$P(W_s \geq c|H_0) = P(W_s \geq c - \frac{1}{2}|H_0) \approx 1 - \Phi(z), \qquad (4)$$

where $z = (c - \frac{1}{2} - \frac{1}{2}n(m + n + 1))/\sqrt{mn(m + n + 1)/12}$.

As an example, let $m = 5$, $n = 7$, $c = 54$. Then $P(W_s \geq 54 \mid H_0) = .101$, and its normal approximation is $1 - \Phi(1.380) = .0838$. The continuity correction yields the better approximation $P(W_s \geq 54 \mid H_0) = P(W_s \geq 53.5 \mid H_0) \approx 1 - \Phi(1.299) = .0097$.

The continuity correction is also often used in 2×2 *contingency tables* for testing for association between two categories. It is simplest to think of this as a two-sample problem for comparing two proportions p_1, p_2 of individuals with a certain characteristic (e.g., smokers) in two populations (e.g., men and women), based on two independent random samples of sizes n_1, n_2 from the two populations, with $n = n_1 + n_2$. Let r_1, r_2 be the numbers in the samples possessing the characteristic. Suppose first that we wish to test $H_0 : p_1 = p_2$, against $H_1 : p_1 < p_2$. Consider the test which rejects H_0, in favor of H_1, if $r_2 \geq c(r)$, where $r = r_1 + r_2$, and $c(r)$ is chosen so that the conditional probability (under H_0) of $r_2 \geq c(r)$, given $r_1 + r_2 = r$, is α. This is the uniformly most powerful unbiased (*UMPU*) test of its size (See Lehmann (1959), pp. 140-146, or Kendall and Stuart (1973), pp. 570-576). The conditional distribution of r_2 , given $r_1 + r_2 = r$, is multinomial, and the test using it is called *Fisher's exact test.* On the other hand, if $n_i p_i \geq 5$ and $n_i(1 - p_i) \geq 5$ $(i = 1, 2)$, the normal approximation is generally used to reject H_0. Note that the (conditional) expectation and variance of r_2 are $n_2 r/n$ and $n_1 n_2 r(n - r)/[n^2(n - 1)]$, respectively (See Lehmann (1975), p. 216). The normalized statistic t is then

$$t = [r_2 - n_2 r/n]/\sqrt{n_1 n_2 r(n - r)/[n^2(n - 1)]}, \qquad (5)$$

and H_0 is rejected when t exceeds $z_{1-\alpha}$, the $(1-\alpha)th$ quantile of Φ. For the continuity correction, one subtracts $\frac{1}{2}$ from the numerator in (5), and rejects H_0 if this adjusted t exceeds $z_{1-\alpha}$. Against the two-sided alternative $H_1 : p_1 \neq p_2$, Fisher's *UMPU* test rejects H_0 if r_2 is either too large or too small. The corresponding continuity corrected t rejects H_0 if either the adjusted t, obtained by subtracting $\frac{1}{2}$ from the numerator in (5), exceeds $z_{1-\alpha/2}$, or if the t adjusted by adding $\frac{1}{2}$ to the numerator in (5) is smaller than$-z_{1-\alpha/2}$. This may be compactly expressed as

$$\text{Reject } H_0 \text{ if } V \equiv (n-1)[\mid r_1 n_2 - r_2 n_1 \mid -\frac{1}{2}n]^2/(n_1 n_2 r(n-r)) > \chi^2_{1-\alpha}(1), \tag{6}$$

where $\chi^2_{1-\alpha}(1)$ is the $(1-\alpha)$th quantile of the chisquare distribution with 1 degree of freedom. This two-sided continuity correction was originally proposed by F.Yates in1934, and it is known as *Yates' correction.* For numerical improvements due to the continuity corrections above, we refer to Kendall and Stuart (1973), pp. 575-576, and Lehmann (1975), pp. 215-217. For a critique, see Connover (1974). If the sampling of n units is done at random from a population with two categories (men and women), then the *UMPU* test is still the same as Fisher's test above, conditioned on fixed marginals n_1,(and, therefore, n_2) and r.

Finally, extensive numerical computations in Bhattacharya and Chan (1996) show that the chisquare approximation to the distribution of *Pearson's frequency chisquare* statistic is reasonably good for degrees of freedom 2 and 3, even in cases of small sample sizes, extreme asymmetry, and values of expected cell frequencies much smaller than 5. One theoretical justification for this may be found in the classic work of Esseen (1945), which shows that the error of chisquare approximation is $O(n^{-d/(d+1)})$ for degrees of freedom d.

REFERENCES

Bhattacharya, R.N. and Chan, N.H. (1996). Comparisons of chisquare, Edgeworth expansions and bootstrap approximations to the distribution of the frequency chisquare. *Sankhya, Ser.A*, **58**,57-68.

Bhattacharya, R.N. and Ranga Rao, R.(1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

Connover, W.J.(1974). Some reasons for not using Yates' continuity correction on 2 × 2 contingency tables. *J. Amer. Statist. Assoc.* **69**, 374-376.

Esseen, C.G.(1945). Fourier analysis of distribution functions: A mathematical study of the Laplace-Gaussian law. *Acta Math.* **77**, 1-125.

Genedenko, B.V. and Kolmogorov, A.N. (1954). *Limit Distributions of Sums of Independent Random Variables.* English translation by K.L. Chung. Reading, Massachussetts.

Kendall, M.G. and Stuart, A. (1973). *The Advanced Theory of Statistics.* Vol. 2, 3rd edition. Griffin, London.

Lehmann, E.L.(1959). *Testing Statistical Hypotheses.* Wiley, New York.

Lehmann, E.L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* (With the special assistance of DA'brera, H.J.M.), Holden-Day, Oakland, California.