

t and χ^2 : revisiting the classical tests for the 21st century classroom

Dimitris N. Politis

Consider the standard setup where X_1, \dots, X_n are i.i.d. from distribution F with mean $\mu = EX_i$, variance $\sigma^2 = E(X_i - \mu)^2 > 0$, skewness $\gamma = E(X_i - \mu)^3/\sigma^3$, and kurtosis $\kappa = E(X_i - \mu)^4/\sigma^4$ assumed finite. As usual, define the sample mean $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, the sample variance $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, and the t -statistic $T = \sqrt{n}(\bar{X} - \mu)/\hat{\sigma}$.

At the beginning of the 20th century, the classical theory for the t and χ^2 tests was developed by the statistical pioneers W.S. Gosset (“Student”), R.A. Fisher, etc. These were undoubtedly major breakthroughs at the time but are intimately tied to the assumption that F is Gaussian which is often hard to justify. Indeed, the true sampling distribution of T and $(n-1)\hat{\sigma}^2$ can be quite different from the textbook t_{n-1} and $\sigma\chi_{n-1}^2$ distributions respectively even under moderate non-normality.

Skewness and the t -statistic. Nowadays that samples are typically large, practitioners may be reluctant to start out with such a restrictive assumption as normality, relying instead on a Central Limit Theorem (CLT). It is well known that the accuracy of the CLT for \bar{X} depends on the skewness of the data. Under regularity conditions, an Edgeworth expansion yields

$$P(\sqrt{n}(\bar{X} - \mu)/\sigma \leq x) = \Phi(x) + b(x)\gamma/\sqrt{n} + O(1/n) \quad (1)$$

for all x , where $\Phi(\cdot)$ is the standard normal distribution, and $b(\cdot)$ a bounded function. Eq. (2) indicates that the distribution of \bar{X} has a skewness of order γ/\sqrt{n} that is not captured in the symmetric shape of the limit distribution.

What is perhaps lesser known is that skewed data induce a strong dependence between the numerator and denominator of the t -statistic thereby shedding doubt on the accuracy of the t -tables; Hesterberg [5] brings out this point. The corresponding Edgeworth expansion for the t -statistic reads

$$P(\sqrt{n}(\bar{X} - \mu)/\hat{\sigma} \leq x) = \Phi(x) + c(x)\gamma/\sqrt{n} + O(1/n) \quad (2)$$

where $c(\cdot)$ is another bounded function. Figure 1 depicts a scatterplot of \bar{X} vs. $\hat{\sigma}$ based on 250 simulated samples each of size $n = 30$ from $F = \frac{1}{2}\chi_2^2$, i.e., exponential; the correlation between \bar{X} and $\hat{\sigma}$ is unmistakable.

One might think that this is a phenomenon particular to the exponential distribution where \bar{X} and $\hat{\sigma}$ estimate the same parameter. In fact, the phenomenon holds true for any skewed distribution F , and the correlation persists regardless of sample size. To elaborate, the δ -method implies $\text{Corr}(\bar{X}, \hat{\sigma}) \approx \text{Corr}(\bar{X}, \hat{\sigma}^2) \rightarrow \gamma/\sqrt{\kappa - 1}$ as $n \rightarrow \infty$. When F is exponential, $\gamma = 2$, $\kappa = 9$, and $\text{Corr}(\bar{X}, \hat{\sigma}) \approx 0.7$. Incidentally, this large-sample correlation calculation shows as a by-product the nontrivial general inequality $\gamma^2 \leq \kappa - 1$ which is sharper than what Littlewood’s inequality implies; see p. 643 of DasGupta [2] with $r = 4$, $s = 3$ and $t = 2$. Interestingly, when F is Bernoulli(p), the identity $\gamma^2 = \kappa - 1$ holds for any $p \in (0, 1)$.

Figure 2 shows the true density of the t -statistic obtained via Monte-Carlo simulation from an exponential sample of size $n = 30$ with the t_{29} density superimposed. The true density is skewed left although visually this might not appear particularly striking. However, the true 2.5%

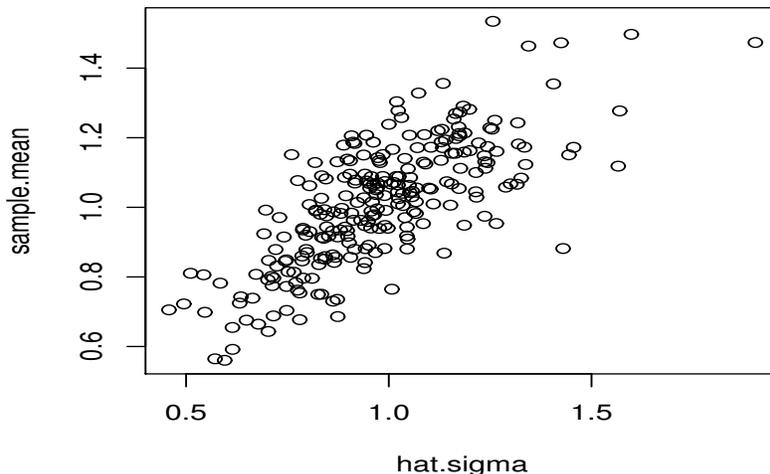


Figure 1: Scatterplot of \bar{X} vs. $\hat{\sigma}$ based on 250 exponential samples each of size $n = 30$.

and 5% quantiles of T are -2.84 and -2.28 respectively that are poorly approximated by the corresponding t_{29} quantiles -2.04 and -1.70 . Luckily, Efron's [3] bootstrap can be applied to the studentized sample mean, i.e., the statistic T , yielding a better approximation. As expounded upon by Hall [4], the studentized bootstrap captures the skewness, i.e.,

$$P^*(\sqrt{n}(\bar{X}^* - \bar{X})/\hat{\sigma}^* \leq x) = \Phi(x) + c(x)\gamma/\sqrt{n} + O_P(1/n) \quad (3)$$

where $\bar{X}^* = \frac{1}{n} \sum_{i=1}^n X_i^*$, $\hat{\sigma}^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i^* - \bar{X}^*)^2}$, and the re-sample X_1^*, \dots, X_n^* is obtained by i.i.d. sampling from the empirical distribution of X_1, \dots, X_n . Let $q^*(\alpha)$ denote the α -quantile of the LHS of eq. (3) which serves as an estimator of $q(\alpha)$, the α -quantile of the LHS of eq. (2). In the exponential example with $n = 30$, the expected values of $q^*(0.025)$ and $q^*(0.050)$ are (approximately) -2.90 and -2.31 , i.e., very close to the true ones.

Kurtosis and asymptotic failure of the χ^2 test. The saving point for Student's t approximation is that $\hat{\sigma} \approx \sigma$ for large n . Since a constant is independent to all random variables, $\hat{\sigma}$ will gradually, i.e., asymptotically, become independent to $\sqrt{n}(\bar{X} - \mu)$, and the t distribution will become standard normal. Hence, Student's t approximation to the distribution of T is asymptotically valid for non-Gaussian data (with finite variance).

The χ^2 test does not even have this excuse: it is asymptotically *invalid* when $\kappa \neq 3$. This might appear suprising since the χ_m^2 distribution is asymptotically normal when m is large; however, it has the wrong asymptotic variance. To see why, let $Y_i = (X_i - \mu)^2$ and $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$. Then, a CLT on the i.i.d. variables Y_1, \dots, Y_n yields

$$P(\sqrt{n}(\bar{Y} - \sigma^2) \leq x) \rightarrow \Phi(x/\tau) \quad \text{as } n \rightarrow \infty \quad (4)$$

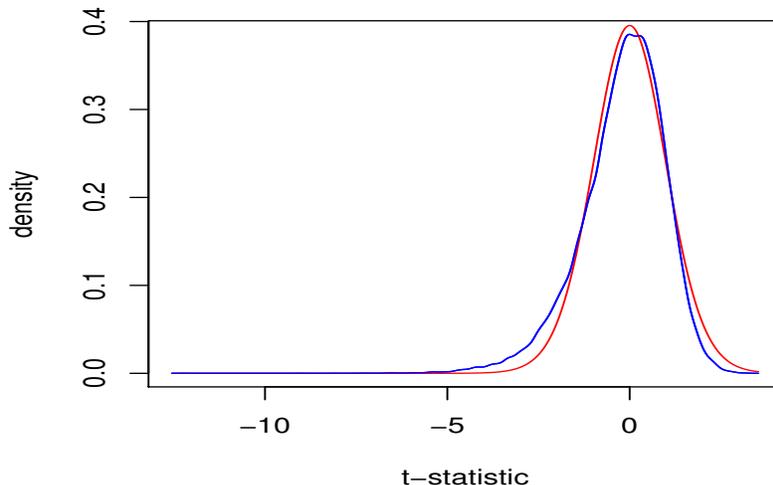


Figure 2: True density of the t -statistic calculated from an exponential sample of size $n = 30$ with the t_{29} density superimposed (in red).

where $\tau^2 = (\kappa - 1)\sigma^4$. But $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 + V_n$ with $V_n = O_p(1)$, i.e., bounded in probability. Hence, eq. (4) implies

$$P(\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \leq x) \rightarrow \Phi(x/\tau) \text{ as } n \rightarrow \infty. \quad (5)$$

Recall that a χ_m^2 variable has mean m and variance $2m$. Hence, the χ_{n-1}^2 approximation to the distribution of $(n-1)\hat{\sigma}^2/\sigma^2$ is associated with a large sample approximation analogous to eq. (5) with τ^2 replaced by $2\sigma^4$; but this is only admissible when the kurtosis κ equals 3 as it does in the Gaussian case. If $\kappa > 3$, the χ^2 approximation underestimates the variance of $\hat{\sigma}^2$ leading to confidence intervals with smaller level than nominal. In the exponential case, the χ^2 approximation effectively estimates the variance of $\hat{\sigma}^2$ to only be 1/4 of what it truly is.

Eq. (5) is valid in general but to be practically useful for tests and confidence intervals it needs a consistent estimate of τ^2 to be plugged in, e.g., letting $\hat{\tau}^2 = (\hat{\kappa} - 1)\hat{\sigma}^4$ where $\hat{\kappa}$ is the sample kurtosis. Furthermore, finite-sample accuracy can be improved via a logarithmic transformation; eq. (5) together with the δ -method yields

$$P(\sqrt{n}(\log \hat{\sigma}^2 - \log \sigma^2) \leq x) \rightarrow \Phi(x/v) \text{ as } n \rightarrow \infty \quad (6)$$

with $v^2 = \tau^2/\sigma^4 = \kappa - 1$. In addition, eq. (5) and (6) imply that the distribution of $\hat{\sigma}^2$ and/or $\log \hat{\sigma}^2$ can be bootstrapped. Focusing on eq. (6), one may readily approximate the quantiles of distribution $P(\sqrt{n}(\log \hat{\sigma}^2 - \log \sigma^2) \leq x)$ by the respective quantiles of the bootstrap distribution $P^*(\sqrt{n}(\log \hat{\sigma}^{*2} - \log \hat{\sigma}^2) \leq x)$. Studentized versions of eq. (5) and (6) can also be derived whose bootstrap approximations may give higher order refinements.

Classroom strategies. It is important to convey to the students that there is nothing wrong with the T and $\hat{\sigma}^2$ statistics *per se*; the issue is how to accurately approximate their sampling distributions without relying on the often unjustifiable assumption of Gaussianity. Luckily, simulation and resampling have been slowly finding their way into the undergraduate classroom. see e.g. the textbooks [1] [6] [7]. However, in their first course on Mathematical Statistics, students may find hard the notion of resampling from the empirical distribution.

The following steps may be found useful in order to ease undergraduates into the matter.

1. Introduce Monte-Carlo simulation from the start as an alternative, computer-intensive methodology to approximate probabilities and expectations, and create quantile tables.
2. Employ *parametric bootstrap*—which is easy to motivate—in order to wean students out of reliance on textbook tables.
3. Finally, depose of the parametric framework altogether, and show students how to generate quantiles for the T and $\hat{\sigma}^2$ statistics by resampling from a *histogram* of the X_1, \dots, X_n data as an approximation to the underlying true density (or mass) function.

If F is (absolutely) continuous, resampling from the histogram is closely related to the smoothed bootstrap; however, it still works when F is discrete as long as the bins are small enough. To elaborate, suppose the histogram of X_1, \dots, X_n is based on m bins centered at midpoints v_1, \dots, v_m that are collected in vector \mathbf{V} with their respective counts collected in vector $\mathbf{Vcounts}$. Basic resampling can then be performed in R letting `Xstar = sample(V, size=n, replace = TRUE, prob = Vcounts/sum(Vcounts))`; note that both the midpoints \mathbf{V} and the counts $\mathbf{Vcounts}$ are part of the list output of the R function `hist()`. If F is continuous, one can then add to each of the coordinates of `Xstar` a uniform “jitter”; i.e., the final bootstrap sample (from which T and $\hat{\sigma}^2$ are to be re-computed) would be given by `Xstar + Ujit` where `Ujit = runif(n, min=-b/2, max=b/2)` and $b = v_{j+1} - v_j$ is the bin width.

References

- [1] Chihara, L.M. and Hesterberg, T.C. (2012). *Mathematical Statistics with Resampling and R*, John Wiley, New York.
- [2] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*, Springer, New York.
- [3] Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Ann. Statist.*, 7, pp. 1-26.
- [4] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.
- [5] Hesterberg, T.C. (2015). What teachers should know about the bootstrap: resampling in the undergraduate statistics curriculum, *American Statistician*, vol. 69, no. 4., pp. 371-386.
- [6] Larsen, R.J. and Marx, M.L. (2012), *An Introduction to Mathematical Statistics and its Applications*, (5th ed.), Prentice Hall, New York.
- [7] Moore, D.S., McCabe, G.P. and Craig, B.A. (2012). *Introduction to the Practice of Statistics*, (8th ed.), WH Freeman, New York.