Published in final edited form as:

J Am Stat Assoc. 2010 September 1; 105(491): 1059–1063. doi:10.1198/jasa.2010.tm10237.

Comment on “Correlated z-values and the accuracy of large-scale statistical estimates” by Bradley Efron

Armin Schwartzman[Assistant Professor]

Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute,
Boston, MA 02115

Armin Schwartzman: armins@hsph.harvard.edu

1 Introduction

In a series of recent articles, Bradley Efron has pointed out that in large-scale multiple testing problems, the observed distribution of the test statistics often does not match the theoretical null distribution (Efron, 2004, 2007a,b, 2008). The correction, which he termed “empirical null”, has been a subject of controversy in the statistical community. In Efron (2007a), he made the case that, even when the theoretical model is correct, the observed distribution of the test statistics can look different from the theoretical null distribution simply because of correlation between the test statistics.

Efron’s present article represents an important step forward in the understanding of this problem. As opposed to previous papers, where the effect of correlation was treated within the context of multiple testing problems and false discovery rates, Efron’s present article breaks through the confusion by separating these two concepts and focusing on the core issue, which is the behavior of a large collection of correlated normal variables. Only then, as applications, he presents the implications for false discovery rate analysis when the correlated normal variables are z-scores in a large-scale multiple testing problem. I think this separation is crucial and helps get us nearer a new theory of inference for high-dimensional data.

In what follows, I present my own interpretation of Efron’s results on how correlation affects the empirical distribution of normal variables. As a shortcut, I work in the continuous domain directly rather than with histogram bins and avoid the inclusion of unnecessary constraints such as normalization, which applies very specifically to microarray data. For simplicity, I assume all the variables are standard normal rather than belonging to a mixture model, but may have an arbitrary correlation structure.

Using these results, I hope to answer the question raised in Efron (2007a) of whether large-scale correlation can substantially widen the observed histogram, as in Figure 1 in that paper and Efron’s current one. The theoretical arguments below indicate that this is unlikely. On the contrary, large-scale correlation has to be mostly positive in order to satisfy positive definite constraints, and as a result, correlation will most likely narrow and shift the observed histogram rather than widen it. An important implication is that a wide histogram may be an indication of the presence of true signal, rather than an artifact of correlation.

Having a model for the observed distribution under correlation, I will suggest a new way of fitting an empirical null, which may be more appropriate in cases where the observed widening is larger than could be explained by correlation alone. I will conclude commenting briefly on the possibility of performing this kind of analysis with χ^2 variates rather than normal.

I will try to keep the notation close to that of Efron, but some discrepancies in notation will be inevitable.

2 The distribution of correlated normal variates

Let $Z_1, Z_2, \dots, Z_N$ be $N(0, 1)$ variables with pairwise correlations $\pi_{ii'} = \text{cor}(Z_i, Z_{i'})$. Using a similar notation to that in (2.1) of Efron's article, let $\hat{F}(x)$ denote the right-sided empirical cdf

$$\hat{F}(x) = \frac{1}{N} \sum_{i=1}^N 1(Z_i \geq x),$$

where $1(\cdot)$ denotes the indicator function. The empirical cdf $\hat{F}(x)$ is always an unbiased estimator of the marginal right-sided standard normal cdf

$$E[\hat{F}(x)] = \Phi^+(x) = 1 - \Phi(x).$$

Its covariance function is given by the following proposition. The result resembles Theorem 1 of Efron, but is given instead in continuous form. The proof is given in the appendix.

Proposition 1

For any $x, x' \in \mathbb{R}$,

$$\text{cov}[\hat{F}(x), \hat{F}(x')] = \text{Cov}_0(x, x') + \text{Cov}_1(x, x'),$$

where

$$\text{Cov}_0(x, x') = \frac{1}{N} [\Phi^+(\max(x, x')) - \Phi^+(x)\Phi^+(x')] \quad (1)$$

and

$$\text{Cov}_1(x, x') = \left(1 - \frac{1}{N}\right) \sum_{j=1}^{\infty} \frac{\alpha_j}{j!} \phi^{(j-1)}(x) \phi^{(j-1)}(x'), \quad (2)$$

where $\phi^{(j)}(x)$ denotes the j -th derivative of the standard normal density $\phi(x)$, and

$$\alpha_j = \frac{1}{M} \sum_{i < i'} \rho_{ii'}^j = \int_{-1}^1 \rho^j dG(\rho)$$

is the j -th moment according to the empirical measure G of the $M = N(N-1)/2$ pairwise correlations $\rho_{ii'}$.

In Proposition 1, as in Theorem 1 of Efron, Cov_0 is the covariance under independence and Cov_1 is the additional covariance as a result of correlation. The empirical measure G corresponds to the “density” $g(\rho)$ in (2.16) of Efron. Taking a second order approximation to (2), as Efron suggests in (2.33), Proposition 1 implies particularly that

$$\text{Var}[\hat{F}(x)] \doteq \frac{1}{N} \Phi^+(x)[1 - \Phi^+(x)] + \left(1 - \frac{1}{N}\right) \left(\alpha_1 + \frac{\alpha_2}{2} x^2\right) \phi^2(x).$$

Notice that α_1 need not be zero as in Efron’s approximations if the data does not come from a column-normalized matrix, but in general,

$$-\frac{1}{N-1} \leq \alpha_1 \leq 1, \quad 0 \leq \alpha_2 \leq 1, \quad (3)$$

where the lower bound on α_1 is a direct consequence of the fact that the matrix of correlations is non-negative definite (see appendix). The fact that α_1 is most likely non-negative for large N implies that the observed histogram will more likely be narrow than wide in most unnormalized situations, as we will see below.

When extended to the multi-class model, the above results are useful for assessing the variability of the empirical distribution $\hat{F}(x)$, leading to expressions such as (1.4) of Efron. I would like to study instead the effect of correlation on the shape that the observed distribution $\hat{F}(x)$ and the corresponding histogram may take. To keep things simple, I will work with the “one-class” null model rather than the multi-class model.

For large N , the covariance function of Proposition 1 approximates the covariance function

$$C(x, x') = \sum_{j=1}^{\infty} \frac{\alpha_j}{j!} \phi^{(j-1)}(x) \phi^{(j-1)}(x'). \quad (4)$$

Thus, $\hat{F}(x)$ approximates a random function $F_0(x)$ with mean $\Phi^+(x)$ and covariance function $C(x, x')$. This covariance structure suggests an expansion of $F_0(x)$ in terms of the basis $\phi^{(0)}(x), -\phi^{(1)}(x), \phi^{(2)}(x), \dots$ as

$$F_0(x) = \Phi^+(x) + \sum_{j=1}^{\infty} W_j (-1)^j \phi^{(j-1)}(x), \quad (5)$$

where $W_1, W_2, \dots$ are random variables with mean 0 and covariance

$$\text{cov}(W_j, W_{j'}) = \begin{cases} \alpha_j / j!, & j = j' \\ 0, & j \neq j' \end{cases} \quad (6)$$

obtained by setting $\text{cov}[F_0(x), F_0(x')] = C(x, x')$. Additional constraints exist between the weights $W_1, W_2, \dots$ in order to ensure that $F_0(x)$ is non-decreasing.

Equation (5) tells us how to generate likely outcomes of the empirical cdf $\hat{F}(x)$ when N is large. A better indication for what these distributions may look like may be obtained by differentiating (5) to get

$$f_0(x) = \phi(x) + \sum_{j=1}^{\infty} W_j (-1)^j \phi^{(j)}(x). \quad (7)$$

Equation (7) can be recognized as a Gram-Charlier series expansion of type A (Cramér, 1957; Blinnikov and Moessner, 1998). It says that the observed $N(0, 1)$ histogram is modified by adding the basis functions $-\phi^{(1)}(x)$, $\phi^{(2)}(x)$, ..., randomly scaled. Notice that these are not quite principal components because the basis functions are not orthogonal in an L_2 sense, but still give a description of the main modes of variation. In the special case of column-normalized data, $\alpha_1 = 0$ forces $W_1 = 0$ by (6), and (7) resembles Equation (49) of Efron (2007a).

Figure 1 shows the first three basis functions and a few examples of $f_0(x)$. The first basis function $-\phi^{(1)}(x)$ produces a shift in the density, to the left if $W_1 < 0$ and to the right if $W_1 > 0$. The second basis function $\phi^{(2)}(x)$ produces narrowing if $W_2 < 0$ or widening if $W_2 > 0$. It may also produce a bimodal distribution if $W_2 > 0$ is large enough. The higher order basis functions have less of an effect because the variance of $W_3, W_4, \dots$ decays as $1/j!$ by (6). However, they may be important for high correlation. In the extreme case of perfect correlation where $\rho_{ii'} = 1$ for all i, i' , we have that $\alpha_j = 1$ for all j , and $F_0(x)$ takes the shape of a step function where the location of the step has a standard normal distribution. The Gram-Charlier expansion is known to have poor convergence properties (Cramér, 1957; Blinnikov and Moessner, 1998), and it is not clear that it would converge in this case.

To answer the question of whether correlation can produce substantial widening of the $N(0, 1)$ density, it is convenient to write (7) as

$$f_0(x) = \phi(x) \sum_{j=0}^{\infty} W_j h_j(x), \quad (8)$$

where $W_0 = 1$ and $h_j(x)$ denotes the j -th Hermite polynomial $h_0(x) = 1$, $h_1(x) = x$, $h_2(x) = x^2 - 1$, ... as in (2.15) of Efron. The derivatives of $f_0(x)$ are given by

$$f_0^{(l)}(x) = (-1)^l \phi(x) \sum_{j=0}^{\infty} W_j h_{j+l}(x), \quad l=0, 1, 2, \dots \quad (9)$$

Consider now the moments of the random density $f_0(x)$, themselves random. Using (8) it is easy to show that the weights W_j can be interpreted as the ‘‘Hermite moments’’ of $f_0(x)$,

$$W_j = \frac{1}{j!} \int_{-\infty}^{\infty} h_j(x) f_0(x) dx, \quad j=0, 1, 2, \dots \quad (10)$$

In particular, we can compute the random mean and variance of $f_0(x)$:

$$\begin{aligned}\widehat{\mu}_0 &= \int_{-\infty}^{\infty} x f_0(x) dx = W_1, \\ \widehat{\sigma}_0^2 &= \int_{-\infty}^{\infty} x^2 f_0(x) dx - \left[\int_{-\infty}^{\infty} x f_0(x) dx \right]^2 = 2W_2 + 1 - W_1^2.\end{aligned}\tag{11}$$

We see that W_1 controls the mean $\widehat{\mu}_0$ of $f_0(x)$ directly. However, the variance $\widehat{\sigma}_0^2$ of $f_0(x)$ is controlled by both W_1 and W_2 . Its expected value is

$$E(\widehat{\sigma}_0^2) = E(2W_2 + 1 - W_1^2) = 1 - \alpha_1$$

by (6). If $\alpha_1 = 0$, as forced by the column normalization, then $E(\widehat{\sigma}_0^2) = 1$, meaning that the observed histogram can be sometimes wider or sometimes narrower than $N(0, 1)$. On the other hand, if $\alpha_1 > 0$, then $E(\widehat{\sigma}_0^2) < 1$ and we would expect a *narrower* histogram than $N(0, 1)$. In order to expect a wider histogram, we would need $\alpha_1 < 0$, but this is unlikely for large N by (3).

As shown in Figure 1, if $f_0(x)$ is indeed wider than $N(0, 1)$, it cannot be very wide before it becomes bimodal. As an approximation, discarding terms for $j \geq 4$ in (9), we get that the curvature of $f_0(x)$ at zero,

$$f^{(2)}(0) \doteq \phi(0)[h_2(0) + W_2 h_4(0)] = \phi(0)[-1 + 3W_2],$$

remains non-positive (concave) only for $W_2 \leq 1/3$. Plugging in (11), we get that the largest central spread for a unimodal density is about $\widehat{\sigma}_0 \approx 1.3$.

This result is important because it implies that the central spread $\widehat{\sigma}_0 = 1.68$ of the observed histogram for the leukemia data in Figure 1 of Efron's current paper cannot be explained on the basis of correlation alone. The same is true for the central spread $\widehat{\sigma}_0 = 1.55$ of the observed histogram for the breast cancer data in Figure 1 of Efron (2007a). The sampling variability in the estimation of the true correlations $\rho_{ii'}$ may affect the estimation of the correlation parameters α_j , but does not affect the spread of the histogram of z variates, which is specified by the random variables W_j .

A possible explanation for the unusually large central spread in could be the presence of unobserved covariates, as suggested in Efron (2007b). Otherwise, it may be that the non-null component of the mixture is stronger than shown and that the non-null counts (the red bars in Figure 1 of Efron's current paper) should extend further into the center of the histogram. This is impossible to detect if the empirical null is purposefully made to match the center of the histogram. Assuming a null distribution that is wider than it should be will result in a loss of detection power.

3 The empirical null

A natural question that arises from the above results is how to properly estimate an empirical null. Since the empirical null might be used for inference in multiple testing, it is important that it accurately represents the tail behavior of the distribution. While for small departures from $N(0, 1)$, $f_0(x)$ may be approximated by a $N(\widehat{\mu}_0, \widehat{\sigma}_0^2)$ density as in Efron's central matching method (Efron, 2007b), (7) and Figure 1 suggest that $f_0(x)$ may not be normal in general.

Rather than fitting a normal density to the bulk of the histogram, one could fit model (8) directly. Assuming the mixture model (4.10) of Efron

$$f(x) = p_0 f_0(x) + p_1 f_1(x)$$

and taking $f_0(x)$ as (8), we have that at $x = 0$,

$$\begin{aligned} f_0(x) &\doteq p_0 \phi(0) [1 + W_2 h_2(0)] = (p_0 / \sqrt{2\pi})(1 - W_2) \\ f_0'(x) &\doteq -p_0 \phi(0) [W_1 h_2(0)] = -(p_0 / \sqrt{2\pi}) W_1 \\ f_0^{(2)}(x) &\doteq p_0 \phi(0) [h_2(0) + W_2 h_4(0)] = (p_0 / \sqrt{2\pi})(-1 + 3W_2) \end{aligned}$$

where we have discarded the terms $j \geq 3$ in (8) and (9). Thus, if we fit a quadratic polynomial $y = b_0 + b_1 x + b_2 x^2$ to the observed histogram near $x = 0$, matching the above derivatives to those of the polynomial gives estimates

$$\widehat{W}_2 = \frac{2\widehat{b}_2 + \widehat{b}_0}{2\widehat{b}_2 + 3\widehat{b}_0}, \quad \widehat{W}_1 = -\frac{\widehat{b}_1}{\widehat{b}_0}(1 - \widehat{W}_2), \quad \widehat{p}_0 = \frac{\sqrt{2\pi}\widehat{b}_0}{1 - \widehat{W}_2} \quad (12)$$

and corresponding empirical null

$$\widehat{f}_0(x) = \phi(x) [1 + \widehat{W}_1 x + \widehat{W}_2 (x^2 - 1)]. \quad (13)$$

This idea is similar to the null adjustment made in Equation (57) of Efron (2007a), with $\widehat{W}_2$ playing the role of $\widehat{A}/\sqrt{2}$. A technical consideration is that for the empirical null to be well defined, the estimates $\widehat{W}_1$ and $\widehat{W}_2$ should be such that $\widehat{f}_0(x) \geq 0$ for all x . This is ensured if $\widehat{W}_1^2 \leq 4\widehat{W}_2(1 - \widehat{W}_2)$.

For a rough comparison without having access to the actual data, Figure 2 shows the empirical null of Efron's Figure 1 and the empirical null (13) obtained by fitting to the former a quadratic polynomial near the mode and computing the estimates (12). The estimates (12) are $\widehat{W}_1 = 0.024$, $\widehat{W}_2 = 0.244$, and $\widehat{p}_0 = 0.73$, which correspond to new moment estimates $\widehat{\mu}_0 = 0.024$ and $\widehat{\sigma} = 1.22$ by (11). Notice that the new empirical null is much narrower and that the estimated value of p_0 is much lower, indicating a stronger presence of non-null genes in the data.

4 The distribution of correlated χ^2 variates

As an additional issue, I would like to comment on Efron's recommendation to transform non-normal variates to a normal scale. In Section 5 of the current paper, Efron uses the central limit theorem to argue that both null and non-null non-normal variates can look remarkably normal when transformed to a normal scale via a quantile transformation such as (5.1) or (5.22). While this is true in the t and gamma examples presented, there are situations where the argument does not apply. For example, in likelihood ratio tests and the second order delta method, the asymptotic null distribution of the test statistics for large sample size is χ^2 rather than normal.

The question is whether a transformation like (5.22) should be applied and then the analysis performed on the z variables. Working with normal variates is easier and allows taking advantage of that theory. On the other hand, the quantile transformation (5.22) might substantially shift the mode and emphasize the wrong part of the distribution, particularly if an empirical null is to be fitted (Schwartzman, 2008). In addition, it is unclear how the correlation structure is mapped and whether the resulting normal variates would still be jointly normal.

An idea suggested in Schwartzman and Lin (2009) is to model the χ^2 test statistics directly. Let $f_v(t)$ and $F_v(t)$ denote the density and distribution functions of the $\chi^2(v)$ distribution with v d.f. Under the complete null, the pair of variables $(T_i, T_{i'})$ admits a Lancaster bivariate model where both T_i and $T_{i'}$ have the same marginal density f_v , their correlation is $\rho_{ii'}$, and their joint density is given by

$$f_v(t_i, t_{i'}; \rho_{ii'}) = f_v(t_i) f_v(t_{i'}) \sum_{k=0}^{\infty} \frac{\rho_{ii'}^k}{k!} \frac{\Gamma(v/2)}{\Gamma(v/2+k)} \mathcal{L}_k^{(v/2-1)}\left(\frac{t_i}{2}\right) \mathcal{L}_k^{(v/2-1)}\left(\frac{t_{i'}}{2}\right), \quad (14)$$

where $\mathcal{L}_k^{(v/2-1)}(t)$ are the generalized Laguerre polynomials of degree $v/2-1$:

$\mathcal{L}_0^{(v/2-1)}(t)=1$, $\mathcal{L}_1^{(v/2-1)}(t)=-t+v/2$, $\mathcal{L}_2^{(v/2-1)}(t)=t^2-2(v/2+1)t+(v/2)(v/2+1)$ and so on (Koudou, 1998). A derivation similar to the one above for the Gaussian case gives that, for large N , the covariance function (4) takes the form

$$C(x, x') = f_{v+2}(x) f_{v+2}(x') \sum_{j=1}^{\infty} \frac{\alpha_j}{j!} \frac{\Gamma(v/2)}{\Gamma(v/2+k)} \mathcal{L}_{k-1}^{(v/2-1)}\left(\frac{u}{2}\right) \mathcal{L}_{k-1}^{(v/2-1)}\left(\frac{v}{2}\right) \quad (15)$$

and the expansion (8) takes the form

$$f_0(x) = f_v(x) \sum_{j=0}^{\infty} W_j \frac{\Gamma(v/2)}{\Gamma(v/2+k)} \mathcal{L}_k^{(v/2-1)}\left(\frac{x}{2}\right), \quad (16)$$

where again $W_0 = 1$ and $W_1, W_2, \dots$ are uncorrelated random variables with mean 0 and variance $\alpha_j/j!$. An empirical null in this case could be fitted in a similar way to the Gaussian by matching $p_0 f_0(x)$ and its first two derivatives computed from (16) to a quadratic polynomial fitted to the modal region of the observed histogram of χ^2 variates.

5 Summary

In this comment, I have attempted to exploit Efron's results to shed some light on the effect that correlation has on the observed distribution of a large collection of variables, first standard normal and then χ^2 . To do this, I derived a simplified "complete null" version of Efron's Theorem 1 working directly with empirical distribution functions rather than histograms, and then suggested an approximation to the observed density as a Gram-Charlier expansion in terms of Hermite polynomials in the normal case and generalized Laguerre polynomials in the χ^2 case. Better approximations may exist based on Gauss-Hermite or Edgeworth expansions (Cramér, 1957; Blinnikov and Moessner, 1998).

One of the conclusions is that correlation modifies the common marginal density by adding or subtracting the density's derivatives multiplied by random coefficients. These coefficients have zero mean, are dependent but uncorrelated, and their variance is proportional to the empirical moments of the pairwise correlations, decaying as the order increases. A second order approximation might be enough in practice and could be used as a model for fitting an empirical null to the mode of the observed histogram.

While I have only considered the “one-class” model, the results extend to Efron's normal multi-class model under the assumption that the “null” class dominates the mixture and that the other classes have little overlap with the mode. This assumption is the same one made in Efron (2007a) in order to fit an empirical null. More generally, a multi-class version of (4) for normal variables would resemble the first term of Efron's (2.31) and a corresponding version of (7) could be proposed. Unfortunately, this could not be done easily for χ^2 variables as model (14) cannot be easily extended to the non-central case.

I want to conclude remarking that whether an empirical null should be used at all, regardless of whether it is Efron's central matching or (13) above, is still an open question. In Section 4 of the current article, Efron states that estimating an empirical null “is both necessary and feasible but can greatly increase variability”. The feasibility and increased variability are understood. The necessity, however, is a more fundamental question of marginal versus conditional inference: marginal over all possible realizations of the data or conditional on the particular data observed. This is a crucial question for large-scale hypothesis testing problems. Who better than Efron can answer such questions? I look forward to further developments in this area.

Acknowledgments

This work was partially supported by NIH grant PO1-CA134294. The author thanks Rebecca Betensky and Sarah Emerson for helpful discussions.

References

- Blinnikov S, Moessner R. Expansions for nearly Gaussian distributions. *Astron Astrophys Suppl Ser.* 1998; 130:193–205.
- Cramér, H. *Mathematical Methods of Statistics.* Princeton: Princeton Univ. Press; 1957.
- Efron B. Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *J Amer Statist Assoc.* 2004; 99:96–104.
- Efron B. Correlation and Large-Scale Simultaneous Hypothesis Testing. *J Amer Statist Assoc.* 2007a; 102:93–103.
- Efron B. Size, power and false discovery rates. *Ann Statist.* 2007b; 35:1351–1377.
- Efron B. Simultaneous inference: When should hypothesis testing problems be combined? *Ann Appl Statist.* 2008; 2:197–223.
- Kotz, S.; Balakrishnan, N.; Johnson, NL. *Continuous Multivariate Distributions, Vol. 1: Models and Applications.* New York: John Wiley & Sons; 2000. Bivariate and Trivariate Normal Distributions; p. 251-348.
- Koudou AE. Lancaster bivariate probability distributions with Poisson, negative binomial and gamma margins. *Test.* 1998; 7:95–110.
- Patel, JK.; Read, CB. *Handbook of the Normal Distribution. 2.* New York: Marcel Dekker Inc; 1996.
- Schwartzman A. Empirical Null and False Discovery Rate Inference for Exponential Families. *Ann Appl Statist.* 2008; 2:1332–1359.
- Schwartzman, A.; Lin, X. Working Paper 106. Harvard University; 2009. The effect of correlation in false discovery rate estimation. <http://www.bepress.com/harvardbiostat/paper106>

A Proofs and Derivations

Proof of Proposition 1

For two standard normal variables Z, Z' with correlation ρ , let $\Phi_{\rho}^{+}(x, x') = P(Z \geq x, Z' \geq x')$ be the corresponding right-tail bivariate cdf. Then

$$\begin{aligned} \text{cov}[\widehat{F}(x), \widehat{F}(x')] &= E[\widehat{F}(x)\widehat{F}(x')] - E[\widehat{F}(x)]E[\widehat{F}(x')] \\ &= \frac{1}{N^2} \sum_{i,i'} P(Z_i \geq x, Z_{i'} \geq x') - \Phi^{+}(x)\Phi^{+}(x') \\ &= \frac{1}{N^2} N \Phi^{+}(\max(x, x')) + \frac{1}{N^2} \sum_{i \neq i'} \Phi_{\rho_{ii'}}^{+}(x, x') - \Phi^{+}(x)\Phi^{+}(x') \\ &= \text{Cov}_0(x, x') + \text{Cov}_1(x, x') \end{aligned}$$

where $\text{Cov}_0(x, x')$ is given by (1) and

$$\text{Cov}_1(x, x') = \frac{1}{N^2} \sum_{i \neq i'} \Phi_{\rho_{ii'}}^{+}(x, x') - \left(1 - \frac{1}{N}\right) \Phi^{+}(x)\Phi^{+}(x'). \quad (17)$$

To obtain (2), we proceed as follows. As in (2.10) of Efron, let $\phi_{\rho}(x, x')$ denote the bivariate standard normal density with correlation ρ . Mehler's identity (Patel and Read, 1996; Kotz et al., 2000), also used in (2.15) of Efron, states that

$$\phi_{\rho}(x, x') = \phi(x)\phi(x') \sum_{j=0}^{\infty} \frac{\rho^j}{j!} h_j(x)h_j(x') = \phi(x)\phi(x') + \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \phi^{(j)}(x)\phi^{(j)}(x'),$$

where $h_j(x)$ is the j -th Hermite polynomial and the j -th derivative satisfies $\phi^{(j)}(x) = (-1)^j \phi(x)h_j(x)$. Integrating from u to ∞ and v to ∞ , we get that

$$\Phi_{\rho}^{+}(x, x') = \Phi^{+}(x)\Phi^{+}(x') + \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \phi^{(j-1)}(x)\phi^{(j-1)}(x').$$

Replacing in (17) gives

$$\begin{aligned} \text{Cov}_1(x, x') &= \frac{1}{N^2} \sum_{i \neq i'} \left[\Phi^{+}(x)\Phi^{+}(x') + \sum_{j=1}^{\infty} \frac{\rho_{ii'}^j}{j!} \phi^{(j-1)}(x)\phi^{(j-1)}(x') \right] - \left(1 - \frac{1}{N}\right) \Phi^{+}(x)\Phi^{+}(x') \\ &= \frac{1}{N^2} \sum_{j=1}^{\infty} \frac{1}{j!} \left(\sum_{i \neq i'} \rho_{ii'}^j \right) \phi^{(j-1)}(x)\phi^{(j-1)}(x'), \end{aligned}$$

yielding (2).

Derivation of (3)

The upper bounds are immediate from the definition of α_1 and α_2 . To see the lower bound on α_1 , let $\mathbf{R} = \{\rho_{ii'}\}$ be the correlation matrix of the vector $(Z_1, \dots, Z_N)^T$ and let $\mathbf{1}$ be a column vector of ones. Since $\mathbf{1}^T \mathbf{R} \mathbf{1} \geq 0$ by the positive semidefiniteness of $\mathbf{R}$,

$$\alpha_1 = \frac{1}{N(N-1)} \sum_{i \neq i'} \rho_{ii'} = \frac{1}{N(N-1)} (\mathbf{1}^T \mathbf{R} \mathbf{1} - N) \geq -\frac{1}{N-1}.$$

Figure 1. (a) First three basis functions of the Gram-Charlier expansion (7). (b) Examples of realizations of the random density (7).

Figure 2. Normal empirical null $0.93 * N(0.09, 1.68^2)$ as in Efron's Figure 1 (solid); empirical null $0.73 * f_0$ by truncated Gram-Charlier expansion (13) with $\hat{W}_1 = 0.024$ and $\hat{W}_2 = 0.244$ (dashed).