Appendix A: Justification of the Bootstrap

A classic example of bootstrap failure occurs when one considers the maximum of a series of observations and one may question whether something similar is occurring in this context. Let $Z_1, Z_2, \ldots, Z_n$ be independently and identically distributed uniform random variables on the interval given by $[0, \theta]$ with θ unknown and suppose confidence intervals for θ are sought. Let $T = max\{Z_1, \ldots, Z_n\}$. Then it may be shown that

$$Q = n(\theta - T)/\theta \rightarrow \text{ standard exponential distribution.} \tag{A.1}$$

Now consider an obvious bootstrap procedure to elicit information regarding θ: let $t =$ observed T, T^* denote the maximum value observed in a bootstrap sample of $Z_1, Z_2, \ldots, Z_n$ and $Q^* = n(t - T^*)/t$. For the bootstrap procedure to work in any meaningful way it should be the case that Q^* also converge to a standard exponential distribution. However it is easy to see that if $Z_{(n)}$ is in the bootstrap sample we have $Q^* = 0$. Further, bootstrap sampling with replacement implies

$$Prob(Q^* = 0) = Prob(Z_{(n)} \text{ in bootstrap sample }) = 1 - \left(\frac{n-1}{n}\right)^n \rightarrow .632 \tag{A.2}$$

Consequently the limiting distribution of Q^* contains a point mass at 0 with probability .632 so clearly the limiting distribution of Q^* cannot be exponential.

This problem does not arise in the context presented here (investigating the distribution of τ_{r_G}) as the maximum is taken over G variables, the number of which is fixed. In the failing example given above the maximum is taken over n observations – an index that increases asymptotically.

To give a more formal justification of why the bootstrap is appropriate for this overestimation problem we demonstrate that

$$\tau_{r_G} = \frac{\sqrt{n}\left(\bar{d}_{r_G} - \mu_{r_G}\right)}{\sqrt{2}\, s_{r_G}} \text{ and } \tau^*_{r^*_G} = \frac{\sqrt{n}\left(\bar{d}^*_{r^*_G} - \bar{d}_{r^*_G}\right)}{\sqrt{2}\, s_{r^*_G}} \tag{A.3}$$

have the same asymptotic distribution, i.e. the bootstrap procedure works in at least an asymptotic sense. Let $r^0_1, r^0_2, \ldots, r^0_G$ order the true effect sizes,

$$\frac{\mu_{r^0_1}}{\sigma_{r^0_1}} \leq \frac{\mu_{r^0_2}}{\sigma_{r^0_2}} \leq \cdots \leq \frac{\mu_{r^0_G}}{\sigma_{r^0_G}} \tag{A.4}$$

To proceed further we will focus upon τ_{r_G} in the simple and common situation that there exists a single variable/gene with maximal true effect size greater than all other effect sizes, i.e.

$$\frac{\mu_{r^0_G}}{\sigma_{r^0_G}} > \frac{\mu_j}{\sigma_j} \text{ for all } j \neq r^0_G. \tag{A.5}$$

Recall that

$$t_j = \tau_j + \frac{\sqrt{n}\, \mu_j}{\sqrt{2}\, s_j} \text{ where } \tau_j = \frac{\sqrt{n}\left(\bar{d}_j - \mu_j\right)}{\sqrt{2} s_j} \tag{A.6}$$

so for $j \neq r^0_G$ we have

$$Pr\left[t_{r^0_G} > t_j\right] = Pr\left[\tau_{r^0_G} + \frac{\sqrt{n}\, \mu_{r^0_G}}{\sqrt{2}\, s_{r^0_G}} > \tau_j + \frac{\sqrt{n}\, \mu_j}{\sqrt{2}\, s_j}\right]. \tag{A.7}$$

From (A.5) and the realization that the τ terms are governed by a $t-$distribution it is clear the term $\frac{\sqrt{n}\, \mu_{r^0_G}}{\sqrt{2}\, s_{r^0_G}}$ will dominate the inequality above so that with probability 1 $t_{r^0_G}$ will exceed t_j as $n \to \infty$. As this holds for all $j \neq r^0_G$ this implies $r_G \to r^0_G$ with probability 1. From similar reasoning one can deduce that this behavior also occurs in the bootstrap sample so that $r^*_G \to r_G \to r^0_G$ with probability

1 (assuming the number of bootstrap replications increases to ∞ with n). So, because $r_G^* \to r_G^0$ this implies

$$\frac{\sqrt{n}\left(\bar{d}^*_{r_G^*} - \bar{d}_{r_G^*}\right)}{\sqrt{2}s^*_{r_G^*}} - \frac{\sqrt{n}\left(\bar{d}^*_{r_G^0} - \bar{d}_{r_G^0}\right)}{\sqrt{2}s^*_{r_G^0}} \to 0 \text{ with probability 1, or} \tag{A.8}$$

$$\tau^*_{r_G^*} \to \tau^*_{r_G^0} = \frac{\sqrt{n}\left(\bar{d}^*_{r_G^0} - \bar{d}_{r_G^0}\right)}{\sqrt{2}s^*_{r_G^0}}. \tag{A.9}$$

For a fixed index, e.g. r_G^0, it is well known that under general conditions the bootstrap has appropriate asymptotic behavior (Bickel and Freedman, 1981), i.e.

$$\tau^*_{r_G^0} = \frac{\sqrt{n}\left(\bar{d}^*_{r_G^0} - \bar{d}_{r_G^0}\right)}{\sqrt{2}s^*_{r_G^0}} \text{ and } \tau_{r_G^0} = \frac{\sqrt{n}\left(\bar{d}_{r_G^0} - \mu_{r_G^0}\right)}{\sqrt{2}s_{r_G^0}} \tag{A.10}$$

both converge weakly to a Gaussian distribution. From (A.9) and (A.10) this means

$$\tau^*_{r_G^*} = \frac{\sqrt{n}\left(\bar{d}^*_{r_G^*} - \bar{d}_{r_G^*}\right)}{\sqrt{2}s^*_{r_G^*}} \text{ converges in distribution to } N(0,1). \tag{A.11}$$

Similarly, because $r_G \to r_G^0$ with probability 1 this means

$$\tau_{r_G} \text{ converges with probability 1 to } \tau_{r_G^0} = \frac{\sqrt{n}\left(\bar{d}_{r_G^0} - \mu_{r_G^0}\right)}{\sqrt{2}s_{r_G^0}} \tag{A.12}$$

where the right hand term has a $t-$distribution and hence also converges to a $N(0,1)$ distribution. Consequently it has been demonstrated that both $\tau^*_{r_G^*}$ and τ_{r_G} have the same limiting distribution and thus the use of the bootstrap is justified in this asymptotic sense.

Appendix B: Calculation of the Bias

Here an effort is made to sketch the degree of bias that may be expected and link this magnitude to some factors such as sample size, distribution of true effect sizes, and the number of tests. Simplifying assumptions will be employed as necessary. Here attention will be focused upon τ_{r_G} though analogous results hold for τ_{r_1}. One may write

$$E\left[\tau_{r_G}\right] = \sum_{j=1}^{G} E\left[\tau_{r_G} \left| r_G = j\right.\right] P\left[r_G = j\right]. \tag{B.1}$$

$$\text{Then } E\left[\tau_{r_G} \left| r_G = j\right.\right] P\left[r_G = j\right] = \left(\int \tau f_{\tau_j | r_G = j}\left(\tau\right) d\tau\right) P\left[r_G = j\right] \tag{B.2}$$

$$= \frac{\int \tau f_{\tau_j}\left(\tau, r_G = j\right) d\tau}{P\left[r_G = j\right]} P\left[r_G = j\right] \tag{B.3}$$

$$= \int \tau f_{\tau_j}\left(\tau, r_G = j\right) d\tau \tag{B.4}$$

where $f_{\tau_j | r_G = j}$ is the conditional distribution of τ_j given $r_G = j$ and $f_{\tau_j}\left(\tau, r_G = j\right)$ describes the joint distribution of τ_j and the event $r_G = j$.

$$\text{Now } r_G = j \text{ if and only if } t_j > \max_{k \neq j} t_k \tag{B.5}$$

$$\text{if and only if } \tau_{r_j} > \max_{k \neq j}\left(\tau_k + \frac{\sqrt{n}}{\sqrt{2}}\left(\frac{\mu_k}{s_k} - \frac{\mu_j}{s_j}\right)\right). \tag{B.6}$$

To simplify we will approximate the s_j and s_k terms by σ_j and σ_k. Then we obtain

$$E\left[\tau_{r_G}\right] = \sum_{j=1}^{G} \int \tau f_{\tau_j}\left(\tau, r_G = j\right) d\tau \tag{B.7}$$

$$\approx \sum_{j=1}^{G} E\left[\int_{M_{-j}}^{\infty} \tau f_{\tau}\left(\tau\right) d\tau\right] \tag{B.8}$$

$$\text{where } M_{-j} = \max_{k \neq j}\left(\tau_k + \frac{\sqrt{n}}{\sqrt{2}}\left(\frac{\mu_k}{\sigma_k} - \frac{\mu_j}{\sigma_j}\right)\right) \tag{B.9}$$

f_τ denotes a $t-$distribution with $2n-2$ degrees of freedom and the expectation in (B.8) is necessary because M_{-j} contains random elements τ_k. To simplify further we will approximate f_τ by a standard Gaussian distribution and assume the G variables are independent. Then we may rewrite terms as

$$E\left[\tau_{r_G}\right] \approx \frac{1}{\sqrt{2\pi}} \sum_{j=1}^{G} E\left[e^{\frac{-M_{-j}^2}{2}}\right] \tag{B.10}$$

From (B.10) one sees that bias is inversely related to the absolute value of the M_{-j} terms. Some consequences of this derivation are as follows.

Consider the effect of increasing the sample size holding all else constant. It is worthwhile to examine M_{-j} for the case when $j = r_G^0$ and $j \neq r_G^0$ separately where we assume only one variable (with index r_G^0) has the most positive effect size, i.e. there are no ties. Then

$$\begin{aligned} \lim_{n\to\infty} M_{-j} &= \lim_{n} \max_{k\neq j} \left(\tau_k + \frac{\sqrt{n}}{\sqrt{2}}\left(\frac{\mu_k}{\sigma_k} - \frac{\mu_j}{\sigma_j}\right)\right) && \text{(B.11)} \\ &= -\infty \text{ if } j = r_G^0 && \text{(B.12)} \\ &= \infty \text{ if } j \neq r_G^0. && \text{(B.13)} \end{aligned}$$

In either case we have that $M_{-j}^2 \to \infty$ so from (B.10) one sees $E\left[\tau_{r_G}\right] \approx 0$.

The case for expanding the differences among effect sizes is similar – at least for the simplified example below. For a given pattern of effect sizes among the G variables (again with no ties for the most extreme effect size), consider a new pattern of effect sizes given by multiplying each original effect by a constant $c > 0$.

Then if r_G^0 designates the most positive effect size

$$\lim_{c\to\infty} M_{-j} = \lim_{c} \max_{k\neq j} \left(\tau_k + c\frac{\sqrt{n}}{\sqrt{2}}\left(\frac{\mu_k}{\sigma_k} - \frac{\mu_j}{\sigma_j}\right)\right) \tag{B.14}$$

$$= -\infty \text{ if } j = r_G^0 \tag{B.15}$$

$$= \infty \text{ if } j \neq r_G^0. \tag{B.16}$$

Consequently the same conclusion of no bias follows. If one reverses the limiting action of c so that $c \to 0$ from above then

$$\lim_{c\downarrow 0} M_{-j} = \max_{k\neq j} \tau_k \tag{B.17}$$

where the τ_k are identically and independently distributed $t-$statistics and the bias is then positive. This conclusion is applicable situation to the situation of no variables showing differential expression.

The case for increasing G, the number of variables is less clear cut as it depends upon the combination of effect sizes. As an example, suppose originally, all true effect sizes are equal (either zero or not) – then there will be non-trivial overestimation. If one additional variable is added that has a much larger effect size then as demonstrated in Table 3 this may reduce or eliminate the bias. Then if an additional variable is added with the same larger effect size some degree of overestimation will then by reintroduced. Empirically it seems that adding variables with effect sizes at or near the size of the largest preexisting effect sizes exaggerates the bias effects for μ_{r_G}. In terms of figuring the change of M_{-j} terms as above there is more ambiguity as some terms M_{-j}^2 terms will likely increase, others decrease, and some new terms will be introduced.

References

Bickel, P.J. and Freedman, D.A.(1981). Some asymptotic theory for the bootstrap. *Annals of Statistics* **9**, 1196-1217.