On Selecting a Transformation: With Applications

L. Brown, Tony Cai, and Anirban DasGupta

2014

1 Introduction

2 Binomial Case

Let $X \sim Bin(n, p)$; it is well known that the transformation $\arcsin \sqrt{\frac{X}{n}}$ is a variance stabilizing transformation (VST) in this case, in the sense $\sqrt{n}(\arcsin \sqrt{\frac{X}{n}} - \arcsin \sqrt{p}) \Rightarrow N(0, 1/4)$, with the asymptotic variance independent of the parameter $p$. Anscombe (1952) pointed out that the transformation $g(X) = \sqrt{n + 1/2(\arcsin \sqrt{\frac{X+3/8}{n+3/4}} - \arcsin \sqrt{p})}$ is a better VST, in the sense an asymptotic expansion for its variance works out to $\text{var}[g(X)] = 1/4 + O(n^{-2})$, although in general, and in particular, for the more traditional VST $\arcsin \sqrt{\frac{X}{n}}$, an asymptotic expansion for the variance would be of the form $1/4 + c_1(p)/n + O(n^{-2})$, i.e., the first nontrivial term depends on $p$. In this sense, Anscombe’s $g(X)$ is at the same time a normalizing and a better variance stabilizing transformation. The following calculations are done in this context. We give a third order weak expansion
for a general VST, which would allow us to simultaneously answer many questions, Anscombe’s observation being a corollary.

2.1 The Weak Expansion

Towards this end, let $a, b \geq 0, c \in \mathcal{R}$ and let $g(X) = g_{a,b,c}(X) = \sqrt{n + c(\arcsin\sqrt{\frac{X+a}{n+b}} - \arcsin\sqrt{p})}$. Let $Z_n$ be the standardized Binomial variable $Z_n = \frac{X-np}{\sqrt{npq}}, q = 1-p$. We derive a third order weak expansion for $g(X)$ of the form

$$g(X) = Z_n/2 + c_1(a, b, p, Z_n)/\sqrt{n} + c_2(a, b, c, p, Z_n)/n + O_p(n^{-3/2})$$

This allows us to derive asymptotic expansions to certain moments of $g(X)$ to desired number of terms, which one can use to answer questions such as what is the recommended transformation to achieve asymptotic normality as well as asymptotic variance stabilization at the same time in some optimal way.

The idea is to use the trigonometric identity

$$\arcsin(x) - \arcsin(y) = \arcsin\left(\sqrt{1 - y^2x} - \sqrt{1 - x^2y}\right)$$

, with $x = \sqrt{\frac{X+a}{n+b}}$ and $y = p$. This will require first a weak expansion for $\frac{X+a}{n+b}$, and then its square root. Finally, the normalizing term $\sqrt{n + c}$ is brought in.

**Step 1.** By definition of $Z_n$,

$$\frac{X+a}{n+b} = \frac{np + \sqrt{npq}Z_n + a}{n+b}$$
\[(p + \sqrt{pq}Z_n/\sqrt{n} + a/n)(1 - b/n + O(n^{-2}) = p + \sqrt{pq}Z_n/\sqrt{n} + (a-bp)/n - b\sqrt{pq}Z_n/n^{3/2} + O_p(n^{-2}) = p(1 + \sqrt{q/pZ_n/\sqrt{n}} + (a/p-b)/n - b\sqrt{q/p}/n^{3/2}) + O_p(n^{-2})\]

\[
\Rightarrow \sqrt{\frac{X + a}{n + b}} = \sqrt{p}[1 + \sqrt{q/pZ_n/(2\sqrt{n}) + (a/p-b)/2n} - b\sqrt{pq}Z_n/2n^{3/2}]
\]

\[
= \sqrt{p}[1 + \sqrt{q/pZ_n/(2\sqrt{n}) + (4a - 4bp - qZ_n^2)/8np} + (\sqrt{q/pZ_n[qZ_n^2 - 4a - 4bp]}/16pn^{3/2})] + O_p(n^{-2}), \text{ on algebra.}
\]

**Step 2** Using \(b - a\) for \(a\), \(b\) for \(b\), \(q\) for \(p\), and \(-Z_n\) for \(Z_n\), similarly,

\[
\sqrt{1 - \frac{X + a}{n + b}} = \sqrt{q}[1 - \sqrt{p/qZ_n/(2\sqrt{n}) + (4bq - 4(b-a) - pZ_n^2)/(8nq)}
\]

\[
+ (\sqrt{p/qZ_n[pZ_n^2 - 4(b - a) - 4bq]}/16/(16qn^{3/2})] + O_p(n^{-2})
\]

**Step 3** Combining these two separate weak expansions, on careful algebra, \(\sqrt{q}\sqrt{\frac{X+a}{n+b}} - \sqrt{p}\sqrt{1 - \frac{X+a}{n+b}} = Z_n/(2\sqrt{n}) + (4a - 4bp + (p-q)Z_n^2)/(8n\sqrt{pq}) + [(4a(p-q) - 4bp)Z_n + (1 - 3pq)Z_n^3]/(16pqn^{3/2}) + O_p(n^{-2}).\)

**Step 4** Now we are going to use the aforementioned trigonometric identity. Indeed, it gives

\[
\arcsin(\sqrt{\frac{X+a}{n+b}} - \arcsin(\sqrt{p}) = \arcsin(\sqrt{q}\sqrt{\frac{X+a}{n+b}} - \sqrt{p}\sqrt{1 - \frac{X+a}{n+b}}).
\]

Furthermore, on using the series representation \(\arcsin x = x + x^3/6 + O(x^5)\) as \(x \to 0\), on several lines of algebra, \(\arcsin(\sqrt{\frac{X+a}{n+b}} -\)
arcsin(\sqrt{p}) = Z_n/(2\sqrt{n}) + (4a - 4bp + (p - q)Z_n^2)/(8\sqrt{npq}) + [(12a(p - q) - 12bp)Z_n + (3 - 8pq)Z_n^3]/(48pnq^{3/2}) + O_p(n^{-2}).

Finally, by multiplying by \sqrt{n + c} = \sqrt{n(1 + c/n)} and recombining coefficients, after additional careful algebra, we arrive at the three term weak expansion

\sqrt{n + c}[arcsin(\sqrt{\frac{X+a}{n+b}}) - arcsin(\sqrt{p})] = Z_n/2 + (4a - 4bp + (p - q)Z_n^2)/(8npq) + [(12a(p - q) - 12bp + 12cpq)Z_n + (3 - 8pq)Z_n^3]/(48npq) + O_p(n^{-2}).

The \(O_p(n^{-2})\) term actually remains \(O(n^{-2})\) on taking an expectation. This may be formally verified by using von Bahr(1965). This allows us to use the weak expansion and formulae for successive moments of a Binomial to derive asymptotic expansions for the moments of \(\sqrt{n + c}[arcsin(\sqrt{\frac{X+a}{n+b}}) - arcsin(\sqrt{p})]\). In particular, we can answer questions such as which choices of \(a, b\), kill the \(1/\sqrt{n}\) term in the bias and which choices of \(a, b, c\) kill the \(1/n\) term in the variance. We can also find asymptotic expansions for the skewness and kurtosis of \(\sqrt{n + c}[arcsin(\sqrt{\frac{X+a}{n+b}}) - arcsin(\sqrt{p})]\), in order to assess close approximation to the \(N(0,1/4)\) distribution.

### 2.2 Expansions for Bias and Variance

In order to derive expansions for the mean and variance of the generalized arcsine transform as above, we will need the follow-

ing binomial moment formulae:

\[ E(Z_n) = 0; E(Z_n^2) = 1; E(Z_n^3) = (1-2p)/\sqrt{npq}; E(Z_n^4) = (1+3(n-2)pq)/(npq) \]

\[ = 3 + (1 - 6pq)/(npq) \]

From the weak expansion and the moment expressions above,

\[ E[\sqrt{n} + c[arcsin(\sqrt{n} + a/n + b) - arcsin(\sqrt{p})]] = (4a - 4bp + (p - q))/(8\sqrt{npq}) + O(n^{-3/2}). \]

Naturally, \( c \) is absent in the two term expansion for the bias. Immediately from this expansion, if we choose \( a = 1/4 \) and \( b = 1/2 \), then the coefficient of the \( 1/\sqrt{n} \) term becomes zero and then we have, for any \( c \),

\[ E[\sqrt{n} + c[arcsin(\sqrt{X + a/n + b} - n + 1/2) - arcsin(\sqrt{p})]] = O(n^{-3/2}) \]

That is, by choosing \( a = 1/4 \) and \( b = 1/2 \), one can attain second order mean-matching.

The expansion for the variance takes a little more calculation than this.

Again, from the weak expansion and the moment expressions,

\[ \text{var}[\sqrt{n} + c[arcsin(\sqrt{X + a/n + b}) - arcsin(\sqrt{p})]] = \text{var}[Z_n/2] + (p - q)^2 \text{var}[Z_n^2]/(64npq) + 2Cov[Z_n/2, (12a(p-q) - 12bp + 12cpq)Z_n]/(48npq) + 2Cov[Z_n/2, (3 - 8pq)Z_n^3]/(48npq) + O(n^{-2}) = 1/4 + [c/4 + 3 - 8a + 4(4a - 2b - 1)p + O(n^{-2})] \]
If we choose $c = 1/2$, $a = 3/4$ and $b = 3/8$, then, interestingly, on a little algebra, the coefficient of the $1/n$ term becomes zero, and one attains second order variance-stabilization, i.e.,

$$\text{var}\left[\sqrt{n + 1/2}\left[\arcsin\left(\frac{X + 3/8}{n + 3/4}\right) - \arcsin\sqrt{p}\right]\right] = 1/4 + O(n^{-2}),$$

which is what Anscombe (1952) had pointed out.

We thus notice that for second order mean-matching and for second-order variance matching, we need incongruous choices of $a, b$. For mean-matching, we need $(a, b) = (1/4, 1/2)$, while for variance matching we need $(a, b) = (3/8, 3/4)$. This is actually true in greater generality, as we shall see later.

Because of this impossibility to simultaneously achieve second order mean-matching and second order variance-matching, one is naturally led to the question which one is more important, and also if an intermediary choice between the points $(a, b) = (1/4, 1/2), (a, b) = (3/8, 3/4)$ is in some sense the recommended choice. The answer depends on what one wants to do with the transformation. For example, if the transformation is to be used for writing confidence intervals, then the choice will depend on coverage and length of the ultimate confidence interval. These questions can be asked and worked out in much greater generality, as we shall see.
2.3 Expansions for Skewness and Kurtosis

The corresponding expansions for the skewness and kurtosis of 
\( Y_n = \sqrt{n + c}\arcsin(\sqrt{\frac{X+a}{n+b}}) - \arcsin(\sqrt{p}) \) are less informative, in the sense the first nontrivial terms in their asymptotic expansions do not involve the free constants \( a, b, c \). The next term does; however, the ultimate effect of killing the second nontrivial term by judicious choice of \( a, b, c \) is likely to be limited. For completeness, and some insight too, we outline the one term expansions for skewness and kurtosis below.

First, note that in this section we may take \( c = 0 \), due to the scale invariance property of the coefficients of skewness and kurtosis. Writing \( \beta = 12a(p - q) - 12bp \), and \( \gamma = (3 - 8pq) \), we have,

\[
Y_n - E(Y_n) = Z_n/2 + (p - q)(Z_n^2 - 1)/(8\sqrt{npq}) + (\beta Z_n + \gamma Z_n^3)/(48npq) + O_p(n^{-3/2})
\]

Therefore,

\[
E(Y_n - E(Y_n))^3 = 1/8E(Z_n^3) + 3(p - q)/(8\sqrt{npq})E(Z_n^4 - Z_n^2) + O(n^{-3/2})
\]

\[
= (1 - 2p)/(8\sqrt{npq}) + 3(p - q)/(16\sqrt{npq}) + O(n^{-3/2})
\]

\[
= (2p - 1)/(16\sqrt{npq}) + O(n^{-3/2}).
\]

On the other hand, we already know that \( E(Y_n - E(Y_n))^2 = 1/4 + O(1/n) \). Therefore, the coefficient of skewness

\[
\kappa_3(n, a, b, c) = E(Y_n - E(Y_n))^3/(E(Y_n - E(Y_n))^2)^{3/2} = (2p - 1)/(2\sqrt{npq}) + O(n^{-3/2}).
\]
The effect of $a, b, c$ is absent in the $1/\sqrt{n}$ term. By similar and a bit more tedious algebra, the corresponding kurtosis expansion is obtained as the following; we omit the steps.

$$\kappa_4(n, a, b, c) = E(Y_n - E(Y_n))^4/(E(Y_n - E(Y_n))^2)^2 = 3+(1-2pq)/(npq)+O(n^{-2})$$

Again, the effect of $a, b, c$ is absent in the $1/n$ term.

3 Poisson Case

Let $X \sim Poi(\lambda)$; it is well known that in this case, $\sqrt{X}$ is a variance-stabilizing transformation, in the sense $\sqrt{n}(\sqrt{X} - \sqrt{\lambda}) \xrightarrow{d} N(0, 1/4)$, and so the asymptotic variance is free of the parameter $\lambda$. Once again, from Anscombe(1952), $\sqrt{n}(\sqrt{X + 3/8} - \sqrt{\lambda})$ is a better VST, in the sense an asymptotic expansion for its variance is of the form $\text{var}[\sqrt{n}(\sqrt{X + 3/8} - \sqrt{\lambda})] = 1/4+O(n^{-2})$, while in general, the expansion for the variance would be of the form $1/4 + c_1(\lambda)/n + O(n^{-2})$. Thus, $\sqrt{X + 3/8}$ has the second order variance-matching property. Analogous to the previous section for the binomial case, we give here the third order weak expansion for a general square-root transformation in the Poisson case, which we can use to simultaneously answer many questions, Anscombe’s observation once again being a corollary.
3.1 The Weak Expansion

To keep the expansions notationally similar looking, we assume that we have \( n \) iid observations \( X_1, \ldots, X_n \) from \( \text{Poi}(\lambda) \), and let \( X = \sum X_i \). Let \( Z_n = (X - n\lambda)/\sqrt{n\lambda} \) be the standardized sum.

**Step 1.** By definition of \( Z_n \),

\[
\frac{X + a}{n + b} = \frac{n\lambda + Z_n\sqrt{n\lambda} + a}{n + b} = \lambda(1 + Z_n/\sqrt{n\lambda} + a/(n\lambda))(1 - b/n + O(n^{-2})
\]

\[
\Rightarrow \sqrt{\frac{X + a}{n + b}} = \sqrt{\lambda}(1 - b/(2n) + O(n^{-2}))\left[1 + Z_n/(2\sqrt{n\lambda}) + (4a - Z_n^2)/(8n\lambda) + (Z_n^3 - 4aZ_n)/(16n\lambda)\right] + O_p(n^{-2}).
\]

**Step 2.** From here, by combining terms,

\[
\sqrt{\frac{X + a}{n + b}} = \sqrt{\lambda}\left[1 + Z_n/(2\sqrt{n\lambda}) + (4a - Z_n^2 - 4b\lambda)/(8n\lambda) + ((Z_n^3 - (4a + 4b\lambda)Z_n)/(16n\lambda))\right] + O_p(n^{-2}).
\]

**Step 3.** Bringing in the normalizing factor \( \sqrt{n + c} \), and combining terms again, after some more careful algebra,

\[
\sqrt{n + c}\left[\sqrt{\frac{X + a}{n + b}} - \sqrt{\lambda}\right] = Z_n/2 + (4a - 4b\lambda - Z_n^2)/(8\sqrt{n\lambda}) + (Z_n^3 + [4(c - b)\lambda - 4a]Z_n)/(16n\lambda) + O_p(n^{-2})
\]

As in the Binomial case, this three term weak expansion can be used to answer many questions simultaneously. In particular, bias, variance, skewness and kurtosis expansions will follow.
3.2 Expansions for Bias and Variance

Once again we will need the following Poisson moment formulas; see DasGupta(1998) for formulas involving the Stirling numbers of the second kind for Poisson moments of arbitrary order. The formulas we need are:

\[ E(Z_n) = 0; E(Z_n^2) = 1; E(Z_n^3) = 1/\sqrt{n\lambda}; E(Z_n^4) = 3+1/(n\lambda) \]

From the weak expansion and these moment formulae,

\[
E[\sqrt{n + c(\sqrt{X + a/n + b} - \sqrt{\lambda})}]
\]

\[ = (4a - 1 - 4b\lambda)/(8\sqrt{n\lambda}) + O(n^{-3/2}) \]

Interestingly, if we choose \( a = 1/4 \) and \( b = 0 \), then the coefficient of the \( 1/\sqrt{n\lambda} \) term becomes zero, resulting in the second order mean-matching property

\[
E[\sqrt{n + c(\sqrt{X + 1/4/n} - \sqrt{\lambda})}] = O(n^{-3/2}).
\]

Again, the variance expansion takes a little more calculation. Since the steps are essentially the same as in the binomial case, we omit the calculation, and simply state it:

\[
\text{var}[\sqrt{n + c(\sqrt{X + a/n + b} - \sqrt{\lambda})}]
\]

\[ = 1/4 + (3 - 8a + 8(c - b)\lambda)/(32n\lambda) + O(n^{-2}). \]
We see that if we choose $a = 3/8$ and $b = c$ (but otherwise arbitrary), the coefficient of the $1/n$ term becomes zero, and so one attains the second order variance stabilization

$$\text{var}[(\sqrt{n} + b)(\sqrt{X + 3/8/n + b} - \sqrt{\lambda})] = 1/4 + O(n^{-2}),$$

for any $b$ (we will probably want to take $b > 0$, but if the true $\lambda > 0$, asymptotically it does not matter what the sign of $b$ is).

Once again, akin to the binomial case, we see that the choices of $a, b, c$ are incongruous for second order mean matching and second order variance matching. Thus, the choice question resurfaces.

### 3.3 Expansions for Skewness and Kurtosis

The qualitative nature of the expansion for skewness is the same as in the binomial case. In the first nontrivial term in the asymptotic expansion for skewness, the free coefficients $a, b, c$ do not appear, although they have an effect on the next term in the expansion. Due to essential similarity of the steps in the derivation to the binomial case, here we simply state the expansion.

$$\kappa_3(n, a, b, c) = -1/(2\sqrt{n\lambda}) + O(n^{-3/2})$$

In contrast, for the kurtosis, something different seems to be going on. The free coefficients $a, b$ do appear in the one term
expansion for kurtosis and it is possible to completely annihilate it by choosing \( a = \frac{3}{8} \) and \( b = 0 \). We outline the derivation due to the special nature of this, in contrast to the binomial case.

Again, we may take \( c = 0 \) due to the scale invariance nature of the definition of kurtosis. We will also need the additional Poisson moment formulas:

\[
E(X^5) = (n\lambda + 10n^2\lambda^2), \quad E(X^6) = (n\lambda + 25n^2\lambda^2 + 15n^3\lambda^3)
\]

Then,

\[
Y_n = \sqrt{n}(\sqrt{(X + a)/(n + b)} - \sqrt{\lambda})
\]

\[
= Z_n/2 + ((4(a - b\lambda) - Z_n^2)/(8\sqrt{n\lambda}) + (Z_n^3 - 4(a + b\lambda)Z_n)/(16n\lambda) + O_p(n^{-3/2})
\]

This gives, on using the binomial expansion of \((u + v + w)^4\),

\[
E(Y_n - E(Y_n))^4 = 1/16E(Z_n^4) + 4 \times 1/8E(Z_n^3(1 - Z_n^2))/(8\sqrt{n\lambda}) + 4 \times 1/8E(Z_n^3(Z_n^3 - 4(a + b\lambda)Z_n))/(16n\lambda) + 6 \times 1/4E(Z_n^2(1 - 2Z_n^2 + Z_n^4))/(64n\lambda) + O(n^{-2})
\]

\[
= 3/16 + (9 - 24(a + b\lambda))/(64n\lambda) + O(n^{-2}),
\]

on algebra. Notice that if we choose \( a = \frac{3}{8} \) and \( b = 0 \), this becomes \( E(Y_n - E(Y_n))^4 = 3/16 + O(n^{-2}) \).

We already know that \( E(Y_n - E(Y_n))^2 = 1/4 + O(n^{-2}) \) for the choices \( a = \frac{3}{8} \) and \( b = 0 \). Therefore,

\[
\kappa_4(n, a, b, c) = E(Y_n - E(Y_n))^4/(E(Y_n - E(Y_n))^2)^2 = 3 + O(n^{-2}),
\]

for the special choices \( a = \frac{3}{8} \) and \( b = 0 \). It is interesting that the same choices \( a = \frac{3}{8} \) and \( b = 0 \) annihilate the first nontrivial
terms in the variance as well as the kurtosis. The Poisson case seems to be a special case, for some reason.

4 General Regular Problems

The basic idea of starting with a consistent estimate \( \hat{\theta} \), and transform it as \( g(\hat{\theta} + a/n) \) and then finally renormalize by \( \sqrt{n + c} \) can evidently be pursued in much greater generality than the binomial and the Poisson cases. The formal expansions can all be carried out. We can even keep the transform \( g \) to be general, subject to sufficient smoothness. This generalization may be fruitful to obtain certain insights; for example, transforms other than a VST may well have more desirable properties in some other ways. Also, we can keep the estimate \( \hat{\theta} \) very general, subject to some assumptions. After the formal expansions are derived, we can look at the entire picture comprehensively and explore what \( g, \hat{\theta}, a, b, c \) are optimal in a specific problem in some well formulated sense. Without the generalization, we cannot obtain such insights. Below, we provide the formal expansions in general. We call them formal because the \( O_p(n^{-r}) \) terms have to integrate to \( O(n^{-r}) \) for whatever \( r \) we need. Whether they do will usually depend on what is \( \hat{\theta} \), and the exact probability model. It is not something that can be guaranteed universally.
4.1 The General Weak Expansion

Let then \( \hat{\theta} = \hat{\theta}(X_1, \cdots, X_n) \) be a sequence of estimates such that \( \sqrt{n}(\hat{\theta} - \theta) \overset{D}{\to} N(0, \sigma(\theta)), \sigma(\theta) > 0 \). We will take the parameter space \( \Theta \subset \mathcal{R} \) to be open. We suppose that \( \hat{\theta} \) admits the moment expansions:

\[
E(\hat{\theta} - \theta) = b(\theta)/n + O(n^{-2}); \ Var(\hat{\theta}) = \sigma^2(\theta)/n \times (1+c(\theta)/n) + O(n^{-3})
\]

\[
E(\hat{\theta} - \theta)^3 = d_{31}(\theta)/n^2 + d_{32}(\theta)/n^3 + O(n^{-4}); \ E(\hat{\theta} - \theta)^4 = d_{41}(\theta)/n^2 + d_{42}(\theta)/n^3 + O(n^{-4})
\]

\[
E(\hat{\theta} - \theta)^5 = d_{51}(\theta)/n^3 + d_{52}(\theta)/n^4 + O(n^{-5}); \ E(\hat{\theta} - \theta)^6 = d_{61}(\theta)/n^3 + d_{62}(\theta)/n^4 + O(n^{-5})
\]

Let \( g(.) : \mathcal{R} \to \mathcal{R} \) be three times continuously differentiable, and let \( a, b, c \) be constants. Let \( T = T_n = g(\hat{\theta} + a/n) \). The task is to derive a three term expansion in probability for \( Y_n = \sqrt{n + c(T_n - g(\theta))} \), from which expansions for its bias, variance, skewness and kurtosis will follow.

Towards this, define \( Z_n = \frac{\sqrt{n}(\hat{\theta} - \theta)}{\sigma(\theta)} \). Then,

\[
T_n = g(\theta) + [\sigma(\theta)Z_n/\sqrt{n} + (a-b\theta)/n - b\sigma(\theta)Z_n/n^{3/2}]g'(\theta) + 1/2[\sigma(\theta)Z_n/\sqrt{n} + (a-b\theta)/n - b\sigma(\theta)Z_n/n^{3/2}]^2g''(\theta) + \cdots + R_n
\]

\[
= g(\theta) + g'(\theta)\sigma(\theta)Z_n/\sqrt{n} + [(a-b\theta)g'(\theta)]/2! + [((a-b\theta)g''(\theta) - bg'(\theta))\sigma(\theta)Z_n + \sigma(\theta)^3g''(\theta)Z_n^3/6]/n^{3/2} + O_p(n^{-5})
\]
4.2 Expansions for Bias and Variance

We present below formal expansions for the mean and variance of $Y_n$. It is implicitly understood that the relevant $O_p(n^{-r})$ terms have been assumed to integrate to $O(n^{-r})$.

From the weak expansion and the assumptions made on the moments of $\hat{\theta}$,

$$E[Y_n] = \frac{g'(\theta)b(\theta)}{\sqrt{n}} + \left[ (a-b\theta)g'(\theta) + \sigma^2(\theta)g''(\theta)/2 \right]/\sqrt{n} + O(n^{-3/2})$$

The first order mean matching property holds iff

$$g'(\theta)\{a - b\theta + b(\theta)\} + \sigma^2(\theta)g''(\theta)/2 = 0 \forall \theta$$

That is, given an estimate sequence $\hat{\theta}$, any triplet $(a, b, g)$ which satisfies the above second order linear differential equation results in the first order mean matching property. It is clear that for a given $b(\theta)$ (which would come from the choice of $\hat{\theta}$), there is essentially a unique family of transforms $g(.)$ and constants $a, b$ which leads to first order mean matching.

Straightforward solution of the linear differential equation yields the following :

$$g(\theta) = c \int e^{-2\int \frac{a-b\theta+b(\theta)}{\sigma^2(\theta)} d\theta} d\theta + d,$$

for arbitrary constants $c, d$, where both $\int$ signs are interpreted as primitives. That is, if we choose an estimate sequence $\hat{\theta}$, and
constants \(a, b\), then subject to existence of the required integrals, there is a two-parameter family of transformations that lead to the first order mean matching property. *However*, if we start with a specific \(g\), e.g., the choice \(g(\theta) = \int \frac{1}{\sigma(\theta)} d\theta\) (namely, a VST), then there could be \((a, b)\) which lead to the identity \(g'(\theta)\{a - b\theta + b(\theta)\} + \sigma^2(\theta)g''(\theta)/2 = 0 \forall \theta\)

\[
\Leftrightarrow \frac{1}{\sigma(\theta)}\{a - b\theta + b(\theta)\} - \frac{\sigma'(\theta)}{2} = 0 \forall \theta.
\]

The existence of a mean matching VST is explored in more detail a bit later.

We give a few examples as illustrations.

**Example; Poisson** Suppose \(X \sim Poi(\lambda)\), and consider transformations of the form \(g(X) = X^\alpha, 0 < \alpha \leq 1\). Suppose the basic estimate sequence is \(\hat{\lambda} = \bar{X}\). Then, in the notation of the calculation above, \(\sigma(\lambda) = \sqrt{\lambda}, b(\lambda) = 0,\) and

\[
g'(\lambda)[a - b\lambda + b(\lambda)] + g''(\lambda)\sigma^2(\lambda)/2 = \alpha[a\lambda^{\alpha-1} - b\lambda^\alpha + \frac{\alpha - 1}{2}\lambda^{\alpha-1}],
\]

which is identically zero if \(a = \frac{1-\alpha}{2}\), and \(b = 0\). Thus, with \(n\) iid observations from \(Poi(\lambda)\), for any \(0 < \alpha \leq 1\), \((\bar{X} + \frac{1-\alpha}{2})^\alpha\) is a *mean matching transformation*. In particular, if \(\alpha = \frac{1}{2}\), we get Anscombe’s observation as a corollary. Note that the choice \(\alpha = \frac{1}{2}\) results simultaneously in variance stabilization and mean matching, but *does not* result in symmetrizing optimally. The choice \(\alpha = \frac{2}{3}\) results simultaneously in symmetrization and
mean matching. The desires to achieve variance stabilization, symmetrization, and mean matching are not mutually compatible. There is a conflict and one has to examine which ones are more important for the ultimate goal, perhaps good confidence intervals. This is discussed in greater detail in the subsequent section on symmetrizing transformations.

**Example; Binomial** Let \( X \sim Bin(n, p) \). Let \( F(\alpha, x) \) denote the CDF of a symmetric Beta random variable on \([0, 1]\) with parameter \( \alpha \). Then, we claim that transformations of the form \( g(x) = F(\alpha, x) \) with \( \alpha < 1 \) result in the required mean matching property, if we then choose \( b = 2a = 1 - \alpha > 0 \) and choose the basic estimate to be \( \frac{X}{n} \).

Indeed, if \( g(x) = F(\alpha, x) \), then \( g'(x) = c(x(1 - x))^{\alpha - 1} \) for a suitable constant \( c \), and \( g''(x) = c(\alpha - 1)(x(1 - x))^{\alpha - 2}(1 - 2x) \). Furthermore, in the notation of the general calculation above, the bias function \( b(.) = 0 \).

Then,

\[
g'(p)[a - bp + b(p)] + g''(p)\sigma^2(p)/2 = c\{a(1 - 2p)(p(1 - p))^{\alpha - 1} \\
+ p(1 - p)(\alpha - 1)(p(1 - p))^{\alpha - 2}(1 - 2p)/2\} = 0
\]

for all \( p \).

In particular, if \( \alpha = \frac{1}{2} \), then we have \( a = 1/4 \) and \( b = 1/2 \), with \( g(x) = \arcsin[\sqrt{x(1 - x)}] \) (other than a constant multiplier), which is the observation due to Anscombe. We see that
there is a much broader class that will result in the mean matching property.

Again, it is the case that while it is possible to achieve variance stabilization and mean matching at the same time, and also symmetrization and mean matching at the same time, it is not possible to achieve all three at the same time. One has to examine what is more important. More about this is said in the section on symmetrizing transformations.

**Example; General Location Parameters** An interesting thing occurs in location parameter problems, which we point out in this example. Let \( X \sim f(x - \theta) \), and assume \( f(.) \) is symmetric about zero. In that case, the Pitman estimate \( \hat{\theta} \) based on \( n \) iid observations is already exactly unbiased. Thus, in the notation of our general calculation, \( b(\theta) = 0 \), and \( \sigma(\theta) = \text{constant} \), which we assume to be 1. We point out that the standard normal CDF \( \Phi(.) \) has a special property \textit{vis-a-vis} mean matching.

Recall now our previous general observation that given \( a, b \), subject to the existence of the relevant integrals, transformations of the form \( g(\theta) = \int e^{-2 \int \frac{a-b\theta+b(\theta)}{\sigma^2(\theta)} d\theta} d\theta \) result in the mean matching property. It follows that if we choose \( b \) to be \(< 0\), \( a \) to be arbitrary, then the relevant integrals exist and \( g(\theta) = \Phi(\sqrt{-2b[\theta - \frac{a}{b}]} \) leads to a mean matching transformation. Precisely, take \( \hat{\theta} \) to be the Pitman estimate, and then consider
the penultimate transformation \( \Phi[(\hat{\theta} + a/n - a/b)\sqrt{-2b}] \) is mean matching for any \( a \) and any \( b < 0 \). The special case \( a = b = 0 \) can be handled directly, which results in the identity transformation, i.e., it simply says that the Pitman estimate is mean matching, which of course we already knew.

Coming next to an expansion for the variance of \( Y_n \), first we write an expansion for \( \text{var}[T_n] \). From the weak expansion and the moment assumptions on \( \hat{\theta} \),

\[
\text{var}[T_n] = \frac{(g'(\theta))^2 \sigma^2(\theta)}{n} \left[ 1 + c(\theta)/n \right] + \frac{(g''(\theta))^2 \sigma^4(\theta)}{4} \times \frac{2}{n^2} \\
+ \frac{g'(\theta)g''(\theta)\sigma^3(\theta)}{\sqrt{n}} \left( \frac{d(\theta) - b(\theta)/\sigma(\theta)}{\sqrt{n}} \right) \times \frac{1}{n^{3/2}} \\
+ \left\{ 2g'(\theta)\sigma(\theta)[(a-b\theta)g''(\theta) - bg'(\theta)]\sigma(\theta) + g'(\theta)g^{(3)}(\theta)\sigma^4(\theta) \right\} \times \frac{1}{n^2} + O(n^{-3})
\]

19
\[ \frac{(g'(\theta))^2 \sigma^2(\theta)}{n} + \frac{1}{n^2} \times \{(g'(\theta))^2 \sigma^2(\theta)c(\theta) + (g''(\theta))^2 \sigma^4(\theta)/2 \\
+ g'(\theta)g''(\theta)\sigma^3(\theta)[d(\theta) - b(\theta)/\sigma(\theta)] + 2g'(\theta)\sigma^2(\theta) \times \\
[(a - b\theta)g''(\theta) - bg'(\theta)] + g'(\theta)g^{(3)}(\theta)\sigma^4(\theta)\} + O(n^{-3}) \]

Multiplying now by the normalizing \(\sqrt{n+c}\) factor in the definition of \(Y_n\), after careful algebra and recombination of terms, the expansion for \(\text{var}[Y_n]\) finally is

\[ \text{var}[Y_n] = (g'(\theta))^2 \sigma^2(\theta) + \frac{1}{n} \times \{c(\theta)(g'(\theta))^2 \sigma^2(\theta) \\
+ c(g'(\theta))^2 \sigma^2(\theta) + g'(\theta)g''(\theta)[d(\theta) - b(\theta)/\sigma(\theta)]\sigma^3(\theta) \\
+(g''(\theta))^2 \sigma^4(\theta)/2 + g'(\theta)g^{(3)}(\theta)\sigma^4(\theta) + 2g'(\theta)\sigma^2(\theta)[-bg'(\theta) + g''(\theta)(a-b\theta)]\} + O(n^{-2}) \\
= (g'(\theta))^2 \sigma^2(\theta) + \\
\frac{1}{n} \times \{(g'(\theta))^2 \sigma^2(\theta)[c-2b+c(\theta)] + g'(\theta)g''(\theta)[d(\theta)\sigma^3(\theta) - b(\theta)/\sigma^2(\theta)] + 2(a-b\theta)\sigma^2(\theta) \\
+ g''^2(\theta)\sigma^4(\theta) + g'(\theta)g^{(3)}(\theta)\sigma^4(\theta)\} + O(n^{-2}) \]

In particular, if \(g(.)\) is already a VST, then the leading term is a constant, and one wants to know for which choices, if any, of the free constants \(a, b, c\), the coefficient of \(\frac{1}{n}\) completely disappears. On more algebra using the final expression just above, we see that this second order variance stabilization occurs iff

\[ c - 2b + c(\theta) - d(\theta)\sigma'(\theta) + \frac{\sigma'(\theta)}{\sigma(\theta)} \]
\[ \times [b(\theta) + 2b\theta - 2a] + \frac{5}{2}\sigma''(\theta) - \sigma(\theta)\sigma''(\theta) = 0 \forall \theta \]

Again, the existence of a second order VST, or colloquially, a variance matching VST is explored in more detail below.

### 4.3 Existence of a Mean or Variance Matching VST

We know from Anscombe(1952) that in the Binomial and the Poisson case, there exist special VSTs that have the previously described mean matching property. That is the coefficient of the \(\frac{1}{\sqrt{n}}\) term in the expansion for the bias vanishes. It seems interesting to ask in what generality would this be possible. We have a complete answer below in the entire one parameter Exponential family. Similar analysis can be carried out for other families, for example a curved one parameter Exponential family.

Assume then that \(X\) has a density wrt some \(\sigma\)–finite measure of the form \(e^{\theta x - \psi(\theta)}h(x)\), \(\theta \in \Theta\), that we conveniently parametrize in terms of the mean itself; this is possible as in the full rank Exponential family, the mean is a strictly increasing function of the natural parameter in its interior. Thus, we assume that

\[ X \sim e^{x(\psi')^{-1}(\mu) - \psi((\psi')^{-1}(\mu))}h(x); \]

writing \(k(\mu) = (\psi')^{-1}(\mu)\), the density then is

\[ X \sim e^{xk(\mu) - \psi(k(\mu))}h(x) \]
We consider the problem of estimating $\mu$ itself, for which the most natural basic estimate is $\bar{X}$. Since $\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \psi''(k(\mu)))$, for a smooth $g(.)$, $\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{D} N(0, (g'(\mu))^2 \psi''(k(\mu)))$. Thus, the basic VST is given by

$$g'(\mu) = \frac{1}{\sqrt{\psi''(k(\mu))}}$$

and

$$g(\mu) = \int \frac{1}{\sqrt{\psi''(k(\mu))}} d\mu$$

For the following calculation, we are going to need the following derivative formulas which are relatively straightforward, and we omit the proofs.

$$k'(\mu) = \frac{1}{\psi''(k(\mu))}, k''(\mu) = \frac{-\psi^{(3)}(k(\mu))}{[\psi''(k(\mu))]^3}$$

Recall now our previous general characterization for a given transform to be mean matching:

$$g'(\mu)(a - b\mu) + \frac{g''(\mu)\sigma^2(\mu)}{2} = 0 \forall \mu$$

We are going to use this with $g(\mu) = \int \frac{1}{\sqrt{\psi''(k(\mu))}} d\mu$, and $\sigma^2(\mu) = \psi''(k(\mu))$.

On doing so, and on using the derivative identities quoted just above, on simplification one gets that a mean matching VST exists in a one parameter Exponential family density iff there exist $a, b$ such that

$$\psi^{(3)}(k(\mu))k'(\mu) + 4b\mu = 4a \forall \mu$$
\[4b\mu - k'(\mu)k''(\mu)(\psi''(k(\mu)))^3 = 4a\forall \mu\]
\[
\Leftrightarrow \frac{\psi^{(3)}(\theta)}{4\psi''(\theta)} + b\psi'(\theta) = a\forall \theta \in \Theta^0
\]
The last line is perhaps the most algebraically convenient version of the characterization of mean matching. The corresponding necessary and sufficient condition for the existence of a variance matching VST in the Exponential family is the following; we will not present all the steps in the derivation.

Towards this end, recall the general characterization given earlier for the existence of a variance matching VST:

\[c - 2b + c(\theta) - d(\theta)\sigma'(\theta) + \frac{\sigma'(\theta)}{\sigma(\theta)}\]
\[
\times [b(\theta) + 2b\theta - 2a] + \frac{5}{2}\sigma^2(\theta) - \sigma(\theta)\sigma''(\theta) = 0\forall \theta,
\]
which we now specialize to the Exponential family situation at hand. We need the following moment formulas towards this specialization:

\[E(X) = \psi'(k(\mu)); E(X^2) = \psi'^2(k(\mu)) + \psi''(k(\mu));\]
\[E(X^3) = (\psi'(k(\mu)))^3 + 3\psi'(k(\mu))\psi''(k(\mu)) + \psi^{(3)}(k(\mu))\]
These formulas are all available in, e.g., Brown(1986). Furthermore, for the estimate sequence \(\hat{\mu} = \bar{X}\), we have the additional
facts $\sigma(\mu) = \sqrt{\psi''(k(\mu))}; b(\mu) = c(\mu) = 0$. Also, recall the notation $d(\mu) = \frac{d_{31}(\mu)}{\sigma^3(\theta)}$.

For notational simplicity, in the following lines we will write just $k$ for $k(\mu)$ and $k'$ for $k'(\mu)$, etc. Likewise the derivatives $\psi^{(j)}(k(\mu))$ will be denoted as just $\psi^{(j)}$. The following formulas for the case of an Exponential family are also needed in substituting for $\sigma, \sigma', \sigma'', d_{31}$ in our general characterization above for the existence of a variance matching VST:

\[
k' = \frac{1}{\psi''}; k'' = -\psi^{(3)}/(\psi'')^3; \sigma = \sqrt{\psi''}; \sigma' = \psi^{(3)}/(2(\psi'')^{3/2}); \sigma'' = \psi^{(4)}/(2(\psi'')^{5/2}) - (\psi^{(3)})^2/(2(\psi'')^{7/2}) - 1/4 \times (\psi^{(3)})^2/(2(\psi'')^{7/2}).
\]

These are straightforward. Now substituting all of these expressions into our general characterization just above, after algebra, the final result works out to the following:

A variance matching VST exists in a one parameter Exponential family density iff there exist constants $a, b, c$ such that

\[
8(c-2b)(\psi''(\theta))^3 + 8b\psi'(\theta)\psi''(\theta)\psi^{(3)}(\theta) - [8a\psi''(\theta)\psi^{(3)}(\theta) + 4\psi''(\theta)\psi^{(4)}(\theta) - 7(\psi^{(3)}(\theta))^2] \forall \theta \in \Theta_0.
\]

Remark We can see that a necessary and sufficient condition for the existence of a variance matching VST in the general Exponential family is complicated. It says that $\psi(\theta)$ must satisfy a complex fourth order (and highly nonlinear) differential equation everywhere in the interior of the natural parameter space.
It is surprising that there are any common densities in the Exponential family at all for which such a complex nonlinear differential equation can hold. The next examples say that there are indeed such densities in the Exponential family.

**Example: Poisson**

### 4.4 Demonstrable Gains in Using Mean or Variance Matching VST

The question naturally arises if adjusting the VST to have the first order mean matching or perhaps the second order variance stabilization property is but a pedagogic exercise, or there is actually something concrete to gain in problems that matter. To answer this, one has to decide what is a VST or an adjustment of it would be used for. The most obvious uses are to form tests and confidence intervals. By the asymptotic nature of all VSTs, the actual error probabilities and the actual coverage probabilities are obviously not equal to the nominal values. Demonstrable gains would be established if we really do see type one error rates closer to the nominal value and larger powers at specific \( n \), and most interestingly, for specific small \( n \). The gains can only be established by extensive case by case computation. Note that no simulations are required below. We can compute exactly.

**Example: Fisher’s z** Fisher’s \( z \) is an old classic in everyday statistics. The sampling distribution of \( r \), the sample correla-
tion coefficient is certainly asymptotically normal under four finite (joint) moments, and so, in particular, for the bivariate normal case. But it has long been known that the convergence to normality is extremely slow. Fisher’s $z$ is the usual VST in the problem; namely, in the bivariate normal case, with $z = \frac{1}{2} \log\left[\frac{1+r}{1-r}\right]$, $\xi = \frac{1}{2} \log\left[\frac{1+\rho}{1-\rho}\right]$, $\sqrt{n}(z - \xi) \overset{L}{\Rightarrow} N(0, 1)$. The quality of the normal approximation improves significantly by using $z$ instead of $r$. This has been commented on in works as old as Hotelling(1953). Note that $z$ is not itself mean matching in the sense we are discussing. More precisely, the $\frac{1}{n}$ term in the expectation of $z - \xi$ does not vanish.

From usual Taylor expansion arguments, on using the moment expansions for $r$ given by $E(r) = \rho - \frac{\rho(1-\rho^2)}{2n} + O(n^{-2})$ (see, e.g., Hotelling(1953); note that Fisher’s expansion is incorrect), one gets $E(z) = \xi + \frac{\rho}{2n} + O(n^{-2})$, if the true $\rho$ is bounded away from $\pm 1$. Thus, a first order bias correction is obtained by considering $z - \frac{r}{2n}$. This adjustment of $z$ is different from a MSE adjustment that Hotelling advocated, and Hotelling’s adjustment is not as simple as this one. One would be curious why we are trying to obtain a mean matched version in this manner, as opposed to our general adjustment which would have been of the form $\frac{z + \frac{\rho}{2n}}{1 + \frac{\rho}{n}}$. The reason is, on calculation, which
we do not show here, it turns out that
\[
E[z + \frac{a}{n} + \frac{b}{n}] = \xi + \frac{1}{n} \times \frac{2a + (1 - 2b)\rho - \rho^3}{2(1 - \rho^2)},
\]
and therefore no choice of \(a, b\) can make the \(\frac{1}{n}\) term zero for all \(\rho\). So, let us look at the adjustment \(z - \frac{r}{2n}\). Suppose we want to compare its exact type I error probability with the test based on simply \(z\). We will consider the one sided alternative here, and merely for notational simplicity, we will just discuss \(H_0 : \rho = 0\) vs. \(H_1 : \rho > 0\). As a function of \(r, z\) is strictly monotone, and so is \(z - \frac{r}{2n}\) with probability 1 for all large \(n\). Thus the critical region of each test based on the asymptotic normal distribution is of the form \(r > r(\alpha, n)\), \(\alpha\) being the nominal level. \(r(\alpha, n)\) is trivial to compute. The exact type I error probability is then \(P_{\rho=0}(r > r(\alpha, n))\). This is easy to compute in the null case (and even for nonzero \(\rho\), by using well known rapidly converging asymptotic expansions, as in Hotelling(1953), or even by using Tables, as in the classic Tables of F.N.David(1938)). We give plots of the exact type I error probability of the tests based on simply \(z\) and the mean matched \(z\) for sample sizes 10 to 30, with nominal levels as 5% and 1%. It is obvious that although Fisher’z has a bias of only about .01 in the type I error rate, the mean matched version is visibly better. We gain by using the mean matched version of Fisher’s \(z\).

Interestingly, there is a corresponding variance correction as
well, and as we will see, it *shrinks* Fisher’z towards \( r \) itself. One may wonder why the adjusted \( z \) would shrink towards \( r \), because \( z \) and \( r \) do *not* estimate the same parameter. After all, \( z \) estimates \( \xi \) and \( r \) estimates \( \rho \). It seems to us that the shrinkage occurs towards \( r \) because \( \xi \approx \rho \) for \( \rho \approx 0 \). Here is this variance corrected adjustment of \( z \).

The starting point is to observe the fact that \( \text{var}(r) = \frac{(1 - \rho^2)^2}{n} + \frac{11\rho^2(1-\rho^2)^2}{2n^2} + O(n^{-3}) \) (this is well known; see, e.g., Hotelling(1953)). Therefore, an obvious idea is to define the *second order VST*

\[
\begin{align*}
z^* &= \int_0^r \frac{1}{(1 - \rho^2)\sqrt{1 + \frac{11\rho^2}{2n}}} d\rho \\
&= \frac{1}{2\sqrt{1 + \frac{11}{2n}}} \log\left[\frac{(1 + r)(1 + \frac{11r}{2n} + \sqrt{1 + \frac{11}{2n}\sqrt{1 + \frac{11}{2n}r^2}})}{(1 - r)(1 - \frac{11r}{2n} + \sqrt{1 + \frac{11}{2n}\sqrt{1 + \frac{11}{2n}r^2}})}\right],
\end{align*}
\]
on actually doing the exact integration. It can be done.

Obviously, this is far too complicated to take seriously. So the obvious idea is to expand this (in probability) in powers of \( n^{-1} \). This results, on further careful calculation, to

\[
z^* = z + \frac{11}{4n}(r - z) + O_p(n^{-2}).
\]

We will now use the RHS of the latest line itself to denote \( z^* \). We have thus arrived at a second order VST given by the more informative form

\[
z^* = r + (1 - \frac{11}{4n})(z - r),
\]
which we recognize to be a shrinkage estimate, a shrinkage of Fisher’s original $z$ towards $r$. Hotelling himself attempted such variance corrections, but his adjustments lacked the interesting shrinkage interpretation because he used variance of $z$ rather than variance of $r$ to find the adjustments. But the spirit of the delta theorem clearly suggests that one should try to stabilize variance by starting with variance of $r$ itself. If we do that, the shrinkage form emerges.

We provide below plots for the exact type 1 error of the one sided test for $H_0 : \rho = 0$ based on $z^*$ also for the 5% and the 1% case. In the 5% case, the mean matched version of $z$ is a shade better in terms of coming close to the nominal level, but in the 1% case, they are virtually the same. In any case, the mean matched version and this new $z^*$ both improve on Fisher’s original $z$. We repeat that much more exact computing would be called for to make firm statements that they always do. But we expect that to be in fact true. We should gain by using either the mean matched version or the new $z^*$.

**Example; Poisson** We provide below a small collection of plots that compare the type one error rates and the power of one sided Poisson tests based on the traditional VS T and its mean and variance matching versions, respectively. The computations are exact after some algebra; no simulation is required at all. In these plots, we notice that the mean matching version
does have a type one error closer to the nominal value in an overall sense. The variance matching version appears to be a very small amount even closer to the nominal value, on close scrutiny. But the difference between the mean and the variance matching versions is not significant in these plots. We also see in the power plots that again the mean matching version offers quite a bit of gain in power when the alternative is not too far from the null, and apparently, never an actual decrease in power. Essentially the same power gains were observed by using the variance matching VST; there was not much to distinguish.

We cannot have a basis to conclude that this is always the case without massive exact computations. But in the computations that we did, we see meaningful gains, especially considering the small sample sizes that we tried.
True Type I Error of 5% Test using Mean matched z

n

0.06

0.05

0.04

0.03

0.02

0.01

15

20

25

30

n
True Type I Error of the 5% One sided Correlation Test using Shrinkage $z$
True Type I Error of the 1% One sided Correlation Test using Shrinkage z
True Type I Error of 1% Test using Plain z

- x-axis: n
- y-axis: 0.01 to 0.07
True Type I Error of 5% Poisson Test using VST; n = 10
True Type I Error of 5% Poisson Test using Mean Matching VST; n = 10
True Type I Error of 5% Poisson Test using Variance Matching VST; n = 10
Power Difference of 5% Poisson Tests using Mean Matching VST and VST; \( n = 10 \), Null = 0.2
Power Difference of 5% Poisson Tests using Variance Matching VST and VST; n = 10, null = .2
True Type I Error of 1% Poisson Test using VST; n = 10
True Type I Error of 1% Poisson Test using Mean Matching VST; $n = 10$
True Type I Error of 1% Poisson Test using Variance Matching VST; n = 10
Power Difference of 1% Poisson Tests using Mean Matching VST and VST; $n = 10$, Null = 1.8
Example; $N(\theta, \theta)$ The normal distribution with an equal
mean and variance is an interesting member of the Exponential family, and is a continuous analog of the Poisson. \( \sum X_i^2 \) is the minimal sufficient statistic, and on a straightforward calculation, the MLE of \( \theta \) is \( \hat{\theta} = \frac{\sqrt{1 + 4c_2} - 1}{2} \), where \( c_2 = \frac{1}{n} \sum X_i^2 \). Since the Fisher information function is \( I(\theta) = \frac{2\theta + 1}{2\theta^2} \), and the problem is very smooth, it follows that \( \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{L} N(0, \frac{\theta^2}{2\theta+1}) \). It follows that the VST is given by the function

\[
\xi = g_v(\theta) = \int \frac{\sqrt{2\theta + 1}}{\theta} = 2[\sqrt{2\theta + 1} - \text{arccoth}(\sqrt{2\theta + 1})]
\]

\[
= 2[\sqrt{2\theta + 1} - \text{arctanh}((2\theta + 1)^{-\frac{1}{2}})],
\]

on actually doing the primitive calculation and on using the obvious trigonometric identity. Plugging in the MLE \( \hat{\theta} \) for \( \theta \), the VST is

\[
g_v(\hat{\theta}) = (1 + 4c_2)^{1/4} - \text{arctanh}((1 + 4c_2)^{-1/4}),
\]

and \( \sqrt{n}(g_v(\hat{\theta}) - \xi) \xrightarrow{L} N(0, 1/2) \).

We next derive a mean matched version in the sense we have been discussing. To derive the mean matched VST, we will need the second term in the asymptotic expansion of the mean of \( g_v(\hat{\theta}) \). On using the usual Taylor series argument, and on using the derivative formula

\[
h''(x) = \frac{1 - 3\sqrt{1 + 4x}}{(1 + 4x)^{5/4}(\sqrt{1 + 4x} - 1)^2},
\]

47
it turns out on some algebra, which we do not show, that

\[ E[g_v(\hat{\theta})] = \xi - \frac{1 + 3\theta}{2(2\theta + 1)^{3/2}n}, \]

and so, the mean matched version is

\[ g_{vm}(\hat{\theta}) = g_v(\hat{\theta}) + \frac{1 + 3\hat{\theta}}{2(2\hat{\theta} + 1)^{3/2}n}. \]

We now want to examine the type I error probabilities of one sided tests using the plain VST \( g_v \) and the mean matched version \( g_{vm} \). The exact calculation of these two error probabilities is quite a bit more complicated than it was in the previous examples. This is because the distribution of \( c_2 \) is a (scaled) noncentral chisquare. The critical region of the test based on \( g_v \) works out to \( c_2 > c(\theta_0, \alpha, n) \), where \( \theta_0 \) is the null value, and \( c(\theta_0, \alpha, n) \) is the unique root of the equation

\[
(1 + 4c_2)^{1/4} - \text{arctanh}((1 + 4c_2)^{-1/4}) - \xi(\theta_0) = z_\alpha/\sqrt{2n}.
\]

Thus, the type I error of this test is the probability

\[ P(\sum X_i^2 > nc(\theta_0, \alpha, n)), \]

which is a tail probability in the noncentral chisquare distribution with \( n \) degrees of freedom and noncentrality parameter \( n\theta_0 \). Precisely, the type I error is

\[ P(\text{NC}\chi^2(n, n\theta_0) > \frac{nc(\theta_0, \alpha, n)}{\theta_0}), \]

48
which can be computed easily by using the density formula
\[ e^{-n\theta_0/2} \frac{1}{2(n\theta_0)^{n/4-1/2}} e^{-u/2} u^{n/4-1/2} I_{n/2-1}(\sqrt{n\theta_0 u}), \]
for the above noncentral chi-square distribution, with \( I_\nu \) denoting the Bessel \( I \) function of index \( \nu \). Notice that all calculations are exact; no simulations were performed in the reports below.

The type I error of the test based on the mean matched VST is formally exactly the same, except we must implicitly assume that \( n \) is sufficiently large so that as a function of \( c_2 \), \( g_{vm} \) is monotone increasing in \( c_2 \). A table of the type I error of each test is given below for the 5% and the 1% case for selected values of \( n \) and \( \theta_0 = 1 \), the null value. It really is surprising how deficient the true type I error of the test based on the ordinary VST is, and it is also suprising how much relief the mean matched version provides. We have yet another example where the mean matched version obviously does better.

<table>
<thead>
<tr>
<th>( n )</th>
<th>VST(5%)</th>
<th>Mean Matched(5%)</th>
<th>VST(1%)</th>
<th>Mean Matched(1%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>.028</td>
<td>.043</td>
<td>.003</td>
<td>.005</td>
</tr>
<tr>
<td>30</td>
<td>.036</td>
<td>.045</td>
<td>.005</td>
<td>.007</td>
</tr>
<tr>
<td>100</td>
<td>.042</td>
<td>.047</td>
<td>.007</td>
<td>.008</td>
</tr>
<tr>
<td>300</td>
<td>.045</td>
<td>.048</td>
<td>.008</td>
<td>.009</td>
</tr>
<tr>
<td>500</td>
<td>.046</td>
<td>.049</td>
<td>.009</td>
<td>.009</td>
</tr>
<tr>
<td>1000</td>
<td>.047</td>
<td>.049</td>
<td>.009</td>
<td>.009</td>
</tr>
</tbody>
</table>
4.5 Expansions for Skewness

As in the special Binomial and the Poisson case in the previous sections, the asymptotic expansion for skewness does not involve the free constants $a, b, c$ in the first nontrivial term. The first nontrivial term involves only the functions $b(\theta), d(\theta)$ and the transform $g(.)$ and its derivatives. However, at this stage $g(.)$ is arbitrary. No VST assumptions are needed right now. The leading term in the skewness expansion is presented below. Note again that the free constant $c$ can be assumed to be zero without loss of generality for the skewness expansion. We do not assume $a, b$ are zero; they will automatically vanish when the leading term is calculated below.

Using the same notation as above, 

$$ Y_n - E(Y_n) = g'(\theta)\sigma(\theta)Z_n + \frac{1}{2\sqrt{n}}\left[\sigma^2(\theta)g''(\theta)(Z_n^2 - 1) - 2b(\theta)g'(\theta)\right] + O_p(n^{-1}) $$

Therefore,

$$ E[Y_n - E(Y_n)]^3 = (g'(\theta)\sigma(\theta))^3 E[Z_n^3] + 3(g'(\theta)\sigma(\theta))^2 \frac{1}{2\sqrt{n}} \times $$

$$ [\sigma^2(\theta)g''(\theta)(E(Z_n^4 - Z_n^2)) - 2b(\theta)g'(\theta)E(Z_n^2)] + O(n^{-3/2}) $$

Using now the moment assumptions we made on the basic estimate sequence $\hat{\theta}$, on a little bit of algebra,

$$ E[Y_n - E(Y_n)]^3 = (g'(\theta)\sigma(\theta))^3 \left[d(\theta) - 3 \frac{b(\theta)}{g'(\theta)\sigma(\theta)} + 3\sigma(\theta) \frac{g''(\theta)}{g'(\theta)}\right], $$
assuming that \( g'(\theta) \) is never zero. Use also the already derived variance expansion

\[
\text{var}[Y_n] = (g'(\theta)\sigma(\theta))^2 + O(n^{-1}).
\]

Putting the two together, the coefficient of skewness of \( Y_n \) admits the expansion

\[
\kappa_3(n, a, b, c) = \frac{1}{\sqrt{n}} \times \text{sgn}(g'(\theta)\sigma(\theta))[d(\theta) - 3 \frac{b(\theta)}{g'(\theta)\sigma(\theta)} + 3\sigma(\theta)\frac{g''(\theta)}{g'(\theta)}] + O(n^{-3/2}).
\]

From this expansion, a transformation \( g \) is first order symmetrizing if and only if

\[
d(\theta) - 3 \frac{b(\theta)}{g'(\theta)\sigma(\theta)} + 3\sigma(\theta)\frac{g''(\theta)}{g'(\theta)} = 0 \forall \theta
\]

\[
\Leftrightarrow g''(\theta) + \frac{1}{3\sigma(\theta)}d(\theta)g'(\theta) = \frac{b(\theta)}{\sigma^2(\theta)} \forall \theta
\]

4.6 Existence and Analytic Characterization of Symmetrizing Transformations

From the above result, first order symmetrizing transformations are characterized by the (monotone) solutions of a second order non-homogeneous linear differential equation. Towards this, it is first instructive to write all solutions of this differential equation. Standard theory asks for a particular solution of the corresponding homogeneous equation, followed by an use of Abel’s transform resulting in all solutions of the nonhomogeneous equation.
We carry out this agenda below. It would be useful to exactly state the form of all solutions of a second order nonhomogeneous equation.

**Theorem (Abel)** Consider the differential equation \( y'' + f(x)y' + h(x)y = R(x) \). Let \( \phi \) be a particular solution of the corresponding homogeneous equation \( y'' + f(x)y' + h(x)y = 0 \). Let also \( z_0 \) be any solution of the first order (in \( z' \)) equation \( \phi z'' + (2\phi' + f\phi)z' = R \). Then a general solution of the original equation \( y'' + f(x)y' + h(x)y = R(x) \) is of the form \( z_0(x)\phi(x) \).

A straightforward application of Abel’s theorem results in the following characterization of all solutions of our equation

\[
g''(\theta) + \frac{1}{3\sigma(\theta)}d(\theta)g'(\theta) = \frac{b(\theta)}{\sigma^2(\theta)} : \\
g(\theta) = z(\theta)\phi(\theta)
\]

where

\[
\phi(\theta) = \int e^{-\frac{1}{3} \int \frac{d(\theta)}{\sigma(\theta)} d\theta} d\theta,
\]

and

\[
z'(\theta) = \frac{\phi'(\theta)}{\phi^2(\theta)}[C + \int \frac{\phi^2(\theta)}{\phi'(\theta)} \frac{b(\theta)}{\sigma^2(\theta)} d\theta].
\]

Once again, it is useful to completely pin down the form of the first order symmetrizing transformation in the general Exponential family. It turns out that there is always essentially a unique one. By comparing the symmetrizing transformation with the
variance stabilizing transformation, analytically and graphically, we can gain useful insight into the exact extent of incompatibility between transforming for achieving the best symmetry and transforming for achieving variance stabilization. Efron (1982), and Bickel and Doksum (1981) have commented on the importance of understanding this tension.

Using the same notation as before, \( X \sim e^{xk(\mu) - \psi(k(\mu))h(x)} \).

To characterize the symmetrizing transformation, from our general representation above, we need the functions \( b(\mu), d(\mu) \) and \( \sigma(\mu) \). Of these, by choice of the problem and the basic estimate sequence, \( b(\mu) = 0 \), and we recall that \( \sigma(\mu) = \sqrt[3]{\psi''(k(\mu))}, d(\mu) = \frac{\psi'''(k(\mu))}{[\psi''(k(\mu))]^{3/2}} \). Then, on calculations, the following are obtained :

\[
\int \frac{d(\mu)}{\sigma(\mu)} d\mu = \log \psi''(k(\mu)); \phi(\mu) = \int [\psi''(\theta)]^{2/3} d\theta |_{\theta=k(\mu)}
\]

The \( \phi \) function is the same one as in our general representation for symmetrizing transforms from just above. Finally, on calculations again,

\[
z'(\mu) = \frac{[\psi''(k(\mu))]^{-1/3}}{(\int [\psi''(\theta)]^{2/3} d\theta |_{\theta=k(\mu)})^2}
\]

On integrating this last expression, one gets

\[
z(\mu) = C - \frac{1}{\int [\psi''(\theta)]^{2/3} d\theta |_{\theta=k(\mu)}},
\]

for an arbitrary constant \( C \). Finally, then, from our general representation, a symmetrizing transformation in a member of
the general one parameter Exponential family is of the following form:

\[ g(\mu) = C_1 \int [\psi''(\theta)]^{2/3} d\theta |_{\theta=k(\mu)} + C_2 \]

In other words, there is essentially just one symmetrizing transformation, namely, \( g(\mu) = \int [\psi''(\theta)]^{2/3} d\theta |_{\theta=k(\mu)} \).

Now that we have an analytic representation of the symmetrizing transformation and we (generally) always have an analytic formula for the VST, we are in a position to analyze the conflict between the two. Here are some examples. For notational unambiguity, we will denote the VST by \( g_v \) and the symmetrizing transformation by \( g_s \).

**Example; Poisson** In this case, \( \psi(\theta) = e^\theta \) and \( k(\mu) = \log \mu \). Therefore, the only symmetrizing transform of \( \bar{X} \) is \( g_s(\bar{X}) = \bar{X}^{2/3} \). In contrast, the VST is \( g_v(\bar{X}) = \bar{X}^{1/2} \). Away from zero, the transforms diverge from each other. Better symmetrization is achieved by making the transform closer to the linear function \( \bar{X} \). The VST seems to have a little too much curvature near zero and a bit too little away from zero. See the plot for a visual comparison.

**Example; Binomial** In this case, \( \psi(\theta) = \log(1 + e^\theta) \) and \( k(\mu) = \log(\frac{\mu}{1-\mu}) \); note that \( \mu \) is just the traditional parameter \( p \). On performing the requisite primitive calculation as asked by our general Exponential family result above, the only sym-
metrizing transform of \( \bar{X} = \frac{x}{n} \) is \( g_s(\bar{X}) = F(\frac{2}{3}, \frac{2}{3}, \bar{X}) \), where \( F(\alpha, \beta, p) \) denotes the CDF of a \( \text{Beta}(\alpha, \beta) \) density. In contrast, the VST, namely the arcsine transform, is the CDF of a \( \text{Beta}(\frac{1}{2}, \frac{1}{2}) \) density. Very interestingly, in this case, the conflict is not so great as in the Poisson case. Not only do the transforms (obviously) not diverge from each other, graphically, they look almost like translations of each other; see the plot. The most noticeable difference is near zero, where, one again, the VST has a significantly larger curvature, the phenomenon we earlier evidenced in the Poisson case.

**Example; Gamma.** Let \( X \sim e^{-x\theta}x^{\alpha-1}\theta^{\alpha}/\Gamma(\alpha) \); we take \( \theta \) to be the parameter and \( \alpha \) as known, although the other problem also has some interest, e.g., in reliability problems.

In this case, \( \psi(\theta) = \alpha \log \theta \) and \( k(\mu) = \frac{\alpha}{\mu} \); note that \( \mu \) is just \( \frac{\alpha}{\theta} \). Again, on performing the necessary primitive calculation, the only symmetrizing transformation is \( g_s(\bar{X}) = \bar{X}^{1/3} \), which is the well known Wilson-Hilferty transform. On the other hand, the VST is \( g_v(\bar{X}) = \log \bar{X} \). We again notice from the plot that the VST has a sharper curvature near \( x = 1 \), the lower tail.

**Example; Bivariate Normal Correlation.** Here is an example where the VST is a time honored transform, and it would be interesting to know which transforms symmetrize. There are two distinctions from the three previous examples. There is no longer a unique symmetrizing transform. By choosing the num-

55
ber $C$ in our general formula for a symmetrizing transform, we can obtain two linearly independent solutions of the requisite differential equation. As $C \to \infty$, the transform converges to the trivial constant transformation, which of course is always symmetrizing. The nontrivial symmetrizing transforms cannot be written in *closed form* now. The function $\phi$ of the general formula above can be found analytically, because asymptotic expansions for moments of $r$ are known. Indeed, on a little calculation using the expansions for the moments of $r$, one gets $\phi(\rho) = (1 + \rho)\log(1 + \rho) - (1 - \rho)\log(1 - \rho) - 2\rho$. But the next integral for $z'$ in our general formula cannot be done in closed form. It is, however, straightforward to write expansions in powers of $r$ for the subsequent quantities that cannot be written in terms of elementary functions. For instance, 

$$\frac{\phi^2(\rho)}{\phi'(\rho)} \frac{b(\rho)}{\sigma^2(\rho)} = \frac{5\rho^5}{72} + O(\rho^7); \frac{\phi'(\rho)}{\sigma^2(\rho)} = -\frac{36}{5\rho^4} + O(\rho^{-2});$$

and, if we choose $C = 0$ in the formula for the general symmetrizing transformation, then $g(\rho) = \frac{5}{432} \rho^6 + O(\rho^8)$. Of course, the constant term is irrelevant, and what we have is that while the VST, namely Fisher’s $z$ is $O(r)$ near zero, the symmetrizing transformation is of the order of $r^6$ near the origin. Note that we can also compute $z', z$ and hence $z\phi$ by numerical integrations. It is also very fast.

We have given plots of the symmetrizing transform and Fisher’s $z$ below, using $C = 0$, and deliberately arranging things so that the functions coincide at both ends. However, in this case, it is
the symmetrizing transform that has a greater curvature near zero.

Variance Stabilizing and Symmetrizing Transformations in Binomial; Upper = VST
Variance Stabilizing and Symmetrizing Transformations in Poisson; Lower = VST
Variance Stabilizing and Symmetrizing Transformations in Gamma; Upper at High End = VST
Symmetrizing Transformation and Fisher’s z in Normal case Near zero
Symmetrizing Transformation and Fisher’s z in Normal Case; Lower at High End = z
4.7 Demonstrable Gains in Using the Symmetrizing Transformation?

The intuitive appeal of the symmetrizing transformation stems from the fear that, especially in small samples, the ordinary estimate $\hat{\theta}$ or a suitable VST, even when mean or variance matched, may be vulnerable to considerable skewness. If there is considerable skewness, the normal approximation would be inadequate, perhaps even risky, and the advantage achieved in variance stabilization may get substantially neutralized. Unfortunately, we do not know of a single interesting example where the dual goals of variance stabilization and symmetrization can be simultaneously achieved. This conflict, and the subsequent natural issue of which one is better to use has been raised by prominent researchers. See, in particular, Fisher (1921), Hotelling (1953), discussions by Kendall and Anscombe of Hotelling’s paper, specifically, Bickel and Doksum (1981), and Efron (1982, The Wald Lecture). It would clearly be useful to have some general investigation of that question, especially because we have now a complete analytic characterization of what the symmetrizing transformation would be. For now, we present a couple of illustrative examples, a discrete one (the Poisson) and a continuous one (the Gamma).

Example; Gamma. Consider the gamma density $e^{-\theta x}x^{\beta-1}/\Gamma(\beta)$.

The basic symmetrizing transformation in the Gamma case is the Wilson-Hilferty cube root transform $\bar{X}^{\frac{3}{2}}$, while the basic
VST is \( \log X \). By using simple Taylor expansions (see, e.g., Bickel and Doksum(2001)), they can be made mean matching by using, respectively,

\[
g_s(\bar{X}) = \bar{X}^{\frac{1}{3}} + \frac{\beta^{2/3}}{9n} \bar{X}^{\frac{1}{3}} = (1 + \frac{\beta^{2/3}}{9n}) \bar{X}^{\frac{1}{3}},
\]

and,

\[
g_v(\bar{X}) = \log \bar{X} + \frac{1}{2n\beta}.
\]

Therefore, for the one sided alternative \( \mu > \mu_0 \), where \( \mu = \frac{1}{\theta} \), on some algebra which we omit, the test based on \( g_s \) rejects if

\[
\theta_0 \sum X_i > n\left(\frac{9n}{9n+1}\right)^3(1 + z_{n/(3 \sqrt{n})})^3,
\]

while the test based on \( g_v \) rejects if

\[
\theta_0 \sum X_i > n\beta e^{z_{n/(3 \sqrt{n})-1}/(2\beta)}.
\]

clearly, the type I error probabilities of both of these tests are actually independent of \( \theta_0 = \frac{1}{\mu_0} \), and are incomplete Gamma functions, and hence are very easily computable. The next plots show the exact type I error probabilities of each test at the nominal 5% and the 1% level for small \( n \), namely, \( n \) between 2 and 15, in the Exponential case, i.e., with \( \beta = 1 \). One notices that the test based on the mean matched symmetrizing transform gives error probabilities remarkably close to the nominal values even at such small \( n \), while the test based on the mean matched VST is not bad, but less accurate, in comparison. There is some evidence that for small \( n \), and skewed populations, the symmetrizing
transformation deserves a closer look than it has been popularly awarded in the literature.

The exact mean and the variance of the VST, the mean matched VST, the Wilson-Hilferty, and the mean matched Wilson-Hilferty transforms are presented below for the Exponential case for some selected values of $n$. The mean and the variance are reported in that order.

<table>
<thead>
<tr>
<th>$n$</th>
<th>VST</th>
<th>Mean matched</th>
<th>Symm.</th>
<th>Mean matched</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>(-.23,1.11)</td>
<td>(-.007,1.11)</td>
<td>(.003,.96)</td>
<td>(.003,1.00)</td>
</tr>
<tr>
<td>10</td>
<td>(-.16,1.05)</td>
<td>(-.003,1.05)</td>
<td>(.001,98)</td>
<td>(.001,1.000)</td>
</tr>
<tr>
<td>30</td>
<td>(-.09,1.02)</td>
<td>(-.0005,1.02)</td>
<td>(.0002,.99)</td>
<td>(.0002,1.0000)</td>
</tr>
</tbody>
</table>

A few things strike out. The bias of the plain VST is quite shocking. The mean matched Wilson-Hilferty transform kills the bias nearly completely and also makes the variance virtually indistinguishable from 1. The mean matched VST is much better. The mean matched symmetrizing transform is a slightly better performer.

As another comparison of the mean matched VST and mean matched symmetrizing transform, we have provided below Q-Q plots of 100 simulated values of each in the Exponential(1) case. In the first plot, $n = 10$, and in the second plot $n = 20$. The Q-Q plots for the two transforms look almost identical under each $n$. The mean matched versions of each are comparable.
Q-Q Plots of 100 Simulated Values of Mean matched VST & Symm. Trans; Exp(1); VST = Top
Q-Q Plots of 100 Simulated Values of Mean matched VST & Symm. Trans; Exp(1); VST = Top
Exact Type I Error of the 5% Exponential One sided Test using the Symmetrizing Transformation
Exact Type I Error of the 1% Exponential One sided Test using the Symmetrizing Transformation
Example; Poisson In this case, we compare the mean match-
ing VST \( g_v(\bar{X}) = \sqrt{\bar{X} + \frac{1}{4n}} \) with the mean matching symmetrizing transformation \( g_s(\bar{X}) = \bar{X}^{\frac{3}{2}} + \frac{1}{9n\bar{X}^{\frac{3}{2}}} \). That \( g_v \) is mean matching is already known to us; a few lines of algebra will show that \( g_s \) is mean matching as well. The proof uses Taylor expansions and we do not produce it here.

On some more algebra, which we also omit, the test based on \( g_v \) rejects if
\[
\sum X_i \geq \left\lfloor n\left(\sqrt{\lambda_0} + \frac{z_\alpha}{(2\sqrt{n})}\right)^2 - \frac{1}{4}\right\rfloor + 1,
\]
where \( \left\lfloor . \right\rfloor \) above denotes integer part. Therefore, the type I error probability of the test based on \( g_v \) is the Poisson probability
\[
P(\text{Poi}(\lambda_0) \geq \left\lfloor n\left(\sqrt{\lambda_0} + \frac{z_\alpha}{(2\sqrt{n})}\right)^2 - \frac{1}{4}\right\rfloor + 1).
\]
And, the test based on \( g_s \) rejects if
\[
\bar{X}^{\frac{2}{3}} + \frac{1}{9n\bar{X}^{\frac{2}{3}}} \geq \lambda_0^{2/3} + \frac{(2z_\alpha\lambda_0^{\frac{1}{6}})}{(3\sqrt{n})}
\]
\[
\iff y^3 - ay + \frac{1}{9n} \geq 0,
\]
where \( y \) is used to denote \( \bar{X}^{\frac{1}{3}} \), and \( a = \lambda_0^{2/3} + \frac{(2z_\alpha\lambda_0^{\frac{1}{6}})}{(3\sqrt{n})} \).
For all sufficiently large \( n \), with probability 1, \( y^3 - ay + \frac{1}{9n} \geq 0 \iff y \geq y_0 \), where \( y_0 \) is the only positive root of the said cubic function. Of course, the root \( y_0 \) depends on each of \( \lambda_0, n, \) and \( \alpha \).
Therefore, the type I error of the test based on \( g_s \) is the Poisson probability
\[
P(\text{Poi}(n\lambda_0) \geq 1 + \left\lfloor ny_0^3\right\rfloor),
\]
where, again, \( \left\lfloor . \right\rfloor \) denotes integer part.
Since Poisson CDFs are very easily calculated, we can plot the type I error of each test for selected values of $\lambda_0, n, \alpha$. 
True Type I Error of the 5% One Sided Poisson test using Mean Matching Symm. Transform; $n = 10$
True Type I Error of the 1% One Sided Poisson test using Mean Matching Symm. Transform; \( n = 10 \)
True Type I Error of the 5% One Sided Poisson test using Mean Matching Symm. Transform; n = 20
In this case, we see that for the smaller sample size 10, as well
as the larger sample size 20, the type I error appears to have a
greater bias than the test based on the mean matched VST did,
as we previously saw in the plots for the mean matched VST. In
contrast to the preceding Exponential example, in the Poisson
case, the $\frac{2}{3}$-rd transform seems less suitable than the square root
transform. More convincing evidence of this is given below.

We provide below an independent avenue for comparing the
desirability of the four choices we are emphasizing; the plain
VST, the mean matched VST, the plain symmetrizing trans-
form, and its mean matched version. The most natural question
ask is which of the four, when studentized (for the two VSTs,
studentization is the same as standardization), gives most nearly
a $N(0,1)$ variable? There are evidently many ways to formul-
late most nearly $N(0,1)$. Below we have simply looked at the
mean, the variance, the skewness, and the kurtosis, for some se-
lected values of the parameter $\lambda$. The mean, variance, skewness
and kurtosis are reported for each procedure in that order. The
calculations are exact; no simulation was done.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>VST</th>
<th>Mean matched</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>(-.13,1.14,-.49,4.07)</td>
<td>(-.004,1.04,-.28,3.30)</td>
</tr>
<tr>
<td>20</td>
<td>(-.06,1.02,-.12,3.07)</td>
<td>(0.1,0.12,3.06)</td>
</tr>
</tbody>
</table>
Curiously, the symmetrizing transformation does not symmetrize that well. In fact, the mean matched VST has uniformly smaller skewness than the symmetrizing as well as the mean matched symmetrizing transform. The mean matched VST also nearly kills the bias of the VST, which, incidentally, is quite serious. At the higher value of \( \lambda \), the mean matched VST even has a kurtosis value closest to 3 than any of its competitors. It does look as though in the Poisson case, the mean matched VST is a winner.

4.8 VST and Symmetrizing Transforms in Microarray

4.9 Construction of Tolerance Intervals from VST and Symmetrizing Transforms

Tolerance intervals are of wide use in engineering, process quality control, and reliability studies. The limits of a tolerance interval enclose a region which will probably contain a large proportion of the population. Thus, given \( X_1, X_2, \ldots, X_n \) from a distribution \( F \), a tolerance interval \( (a(X_1, X_2, \ldots, X_n), b(X_1, X_2, \ldots, X_n)) \) is such that \( P(F(b) - F(a) \geq 1 - \alpha) \geq p \). Thus, tolerance intervals come with two layers of
confidence levels. We will call $p$ the credibility and $1 - \alpha$ the coverage.

Tolerance intervals can be found from exact calculations, which can be tedious or can be found, approximately, by using a suitable pivot. We can also construct tolerance intervals by using a suitable VST, or a symmetrizing transform, when they can be found. A natural question to ask is which transform produces the most credible tolerance interval at a given $\alpha$ and a given length. We consider the Poisson and the Gamma cases here, as they have been two of our continuing exploratory cases in this article, and also because Poisson and Gamma distributions are two of the mainstays in process quality control and reliability and survival analysis problems.

**Example; Poisson** Five candidate tolerance intervals constructed from a normal approximation of a Poisson distribution itself, and a VST, a mean matched VST, a variance matched VST, and a symmetrizing transform are as follows; we use $k$ for $z_{1/2}^\alpha$.

\[
\bar{X} \pm k\sqrt{\bar{X}} \text{(standard)} \\
(\sqrt{\bar{X}} \pm k/2)^2 \text{ (VST)} \\
(\sqrt{\bar{X}} \pm k/2)^2 - \frac{1}{4} \text{ (Mean matched VST)} \\
(\sqrt{\bar{X}} \pm k/2)^2 - \frac{3}{8} \text{ (Variance matched VST)}
\]
\((\bar{X}^{\bar{x}} \pm \frac{2}{3} k \bar{X}^{\frac{1}{\delta}})^{\frac{1}{2}}\) (Symmetrizing Transform)

The exact credibility of any of these tolerance intervals is the Poisson expectation \(E_{X_1,X_2,\ldots,X_n|\lambda}[I_F(b|\lambda)-F(a|\lambda)\geq 1-\alpha]\), and therefore, each of the five credibilities can be computed exactly, by a numerical summation of a Poisson weighted infinite series. The following Tables provide some values for selected \((n, \lambda)\) in order to analyze what is going on. The nominal coverage is 95\% in each case.

<table>
<thead>
<tr>
<th>(k)</th>
<th>MLE</th>
<th>VST</th>
<th>Mean matched</th>
<th>Variance matched</th>
<th>Symmetrizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>.21</td>
<td>0</td>
<td>.25</td>
<td>.30</td>
<td>0</td>
</tr>
<tr>
<td>1.75</td>
<td>.30</td>
<td>0</td>
<td>.66</td>
<td>.54</td>
<td>0</td>
</tr>
<tr>
<td>2.0</td>
<td>.54</td>
<td>0</td>
<td>.87</td>
<td>.78</td>
<td>0</td>
</tr>
<tr>
<td>2.1</td>
<td>.54</td>
<td>0</td>
<td>.93</td>
<td>.87</td>
<td>0</td>
</tr>
<tr>
<td>2.2</td>
<td>.54</td>
<td>0</td>
<td>.97</td>
<td>.93</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: \(n = 10, \lambda = 1\)

<table>
<thead>
<tr>
<th>(k)</th>
<th>MLE</th>
<th>VST</th>
<th>Mean matched</th>
<th>Variance matched</th>
<th>Symmetrizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>.11</td>
<td>0</td>
<td>.27</td>
<td>.21</td>
<td>0</td>
</tr>
<tr>
<td>1.75</td>
<td>.28</td>
<td>0</td>
<td>.70</td>
<td>.62</td>
<td>0</td>
</tr>
<tr>
<td>2.0</td>
<td>.53</td>
<td>0</td>
<td>.96</td>
<td>.90</td>
<td>0</td>
</tr>
<tr>
<td>2.1</td>
<td>.53</td>
<td>0</td>
<td>.98</td>
<td>.96</td>
<td>0</td>
</tr>
<tr>
<td>2.2</td>
<td>.62</td>
<td>0</td>
<td>.99</td>
<td>.99</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: \(n = 20, \lambda = 1\)
<table>
<thead>
<tr>
<th>$k$</th>
<th>MLE</th>
<th>VST</th>
<th>Mean matched</th>
<th>Variance matched</th>
<th>Symmetrizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2.0</td>
<td>.52</td>
<td>.27</td>
<td>.38</td>
<td>.43</td>
<td>.34</td>
</tr>
<tr>
<td>2.1</td>
<td>.66</td>
<td>.64</td>
<td>.75</td>
<td>.76</td>
<td>.72</td>
</tr>
<tr>
<td>2.2</td>
<td>.77</td>
<td>.82</td>
<td>.87</td>
<td>.86</td>
<td>.85</td>
</tr>
</tbody>
</table>

Note: $n = 10; \lambda = 10$

<table>
<thead>
<tr>
<th>$k$</th>
<th>MLE</th>
<th>VST</th>
<th>Mean matched</th>
<th>Variance matched</th>
<th>Symmetrizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2.0</td>
<td>.59</td>
<td>.37</td>
<td>.48</td>
<td>.53</td>
<td>.43</td>
</tr>
<tr>
<td>2.1</td>
<td>.75</td>
<td>.79</td>
<td>.90</td>
<td>.90</td>
<td>.87</td>
</tr>
<tr>
<td>2.2</td>
<td>.85</td>
<td>.93</td>
<td>.96</td>
<td>.96</td>
<td>.95</td>
</tr>
</tbody>
</table>

Note: $n = 20; \lambda = 10$

**Discussion** Except for one or two stray cases, the tolerance intervals based on the ordinary MLE and the plain VST show poor and even very poor performance. The mean matched VST is usually a little better than the variance matched VST; but both do well. The interval based on the symmetrizing transform is worse than the mean or variance matched VST, except when $\lambda, k, n$ are large. In that case, the symmetrizing transform performs at about the same level as the mean and variance matched VSTs. Note, however, although the first four tolerance intervals have the same length, the one based on the symmetrizing trans-
form has approximately the same length *only when* \( \lambda \to \infty \). So the comparison is not really fair unless \( \lambda \) is large. Still, a reasonable conclusion would seem to be that the tolerance intervals based on the ordinary MLE or the plain VST should not be used at all, and the two adjusted VSTs are the safest, unless \( \lambda, k, n \) are all large, in which case the symmetrizing transform is also safe and does well. Again, we see that in a new problem, mean or variance matching of the usual VST leads to a lot of gain.

**Example; Gamma** Again, five candidate tolerance intervals in this case are as follows:

\[
(1 \pm \frac{k}{\sqrt{\beta}} \bar{X}) \text{ (standard)}
\]

\[
Exp[\pm \frac{k}{\sqrt{\beta}}] \bar{X} \text{ (VST)}
\]

\[
Exp[\pm \frac{k}{\sqrt{\beta}} - \frac{1}{2\beta}] \bar{X} \text{ (Mean matched VST)}
\]

\[
Exp[\pm \frac{k}{\sqrt{\beta}}] (\bar{X})^{\frac{1}{\sqrt{1 - \frac{1}{2\beta}}}} \text{ (Variance matched VST)}
\]

\[
(1 \pm \frac{k}{3\sqrt{\beta}})^3 \bar{X} \text{ (Symmetrizing)}
\]

Once again, the exact credibility of each of these tolerance intervals can be computed exactly, by the numerical integration of an incomplete Gamma function multiplied by a Gamma density function. The numbers below required no simulations. The computing was quite fast, and the entire set of Tables took less than one minute. Again, we take the nominal coverage to be 95%.
## Discussion

Except when the credibility of all the intervals is unacceptably small, i.e., except for the smallest values of $k$, the tolerance interval based on the symmetrizing transformation is clearly the best. Moreover, the interval based on the variance
matched VST is clearly the worst. In comparison, the mean matched VST does a lot better; it improves significantly on the plain VST. However, it does not match up to the interval based on the symmetrizing Wilson-Hilferty transform. Consistent with what we had previously seen for the confidence interval problem, once again, in the Gamma case we see the symmetrizing transform as the winner.

For additional information, we have also provided below plots of the credibilities of the five tolerance intervals for $k$ between 2 and 2.5, in the order usual, symmetrizing, plain vst, mean matched vst, and variance matched vst. The plots show the same conclusion that we reached from the Tables.
Credibility of T1_1, T1_5, T1_2, T1_3, T1_4 in Gamma case; n = 10, beta = 5, theta = 1
4.10 VST and Symmetrizing Transforms for Grouped Data

4.11 VST and Symmetrizing Transforms in Selection Bias Models

4.12 Time Series Applications

Another important family for practical applications of these ideas is stationary time series models. Since we are constrained, generally, by one parameter models to talk about variance stabilization and symmetrization, the three most obvious examples are the AR(1), MA(1), and a long memory process with a scalar Hurst parameter. These three examples are investigated below.

**Example; AR(1)** Consider stationary autoregression of order 1, given by the model \((X_t - \mu) = \phi(X_{t-1} - \mu) + Z_t\), where \(Z_t \sim N(0, \sigma^2)\), the \(Z_y\) being independent. We consider estimation of the autoregression parameter \(\phi\) and assume \(\mu\) to be zero. Let \(\hat{\rho}_i\) denote the lag \(i\) autocorrelation. Then for each fixed \(i\), \((\hat{\rho}_i)^\frac{1}{2}\) can be used to consistently estimate \(\phi\). Precisely, \(\sqrt{n}[\hat{\rho}_i - \phi^i] \overset{d}{\Rightarrow} N(0, w_{ii}(\phi))\), where \(w_{ii} = \frac{(1-\phi^2_i)(1+\phi^2_i)}{1-\phi^2_i}\), from which the asymptotic normal distribution of \((\hat{\rho}_i)^\frac{1}{2}\) is deduced by applying the Delta theorem.

The MLE \(\hat{\phi}\) corresponds, essentially, to \(i = 1\). Indeed, the MLE \(\sqrt{n}[\hat{\phi} - \phi] \overset{d}{\Rightarrow} N(0, 1 - \phi^2)\). Therefore, the plain VST

86
is given by the transform $g(\phi) = \int \frac{1}{\sqrt{1-\phi^2}} = \arcsin(\phi)$. This is popularly called the Jenkins transformation. The mean matched version requires an asymptotic expansion for the bias of $\hat{\rho}_1$. The variance matched version requires an asymptotic expansion for the variance of the mle. Fortunately, these are both available; see White(1961). Indeed,

$$E[\hat{\phi}] = \phi + \frac{1 - \phi^2}{n} + O(n^{-2}); \text{var}[\hat{\phi}] = \frac{1 - \phi^2}{n} + \frac{9\phi^2 - 1}{n^2} + O(n^{-3}).$$

We note here that neither a mean matched or a variance matched VST can be found in the Anscombe fashion. For instance, from our general theory on existence of Anscombe style mean matched VSTs, the necessary and sufficient condition now is the existence of constants $a, b$ such that $a + (1 - b)\phi + 1 - \phi^2 \equiv 0$. Clearly, no such $a, b$ exist. However, a mean matched VST can be found in the usual, more direct way. From the asymptotic expansion for the bias of $\hat{\phi}$ stated just above, a mean matched version is:

$$g_m(\hat{\phi}) = \arcsin(\hat{\phi}) + \frac{(\hat{\phi} - \frac{1}{4})^2 + \frac{15}{16}}{n\sqrt{1 - \hat{\phi}^2}}.$$

The variance matched version is found from the general shrinkage idea we described previously. The two term expansion for $n\text{var}[\hat{\phi}]$ is $1 - \phi^2 + \frac{9\phi^2 - 1}{n}$, and so our variance matched VST is:

$$g_{vm}(\hat{\phi}) = \int \frac{1}{\sqrt{1 - \phi^2} + \frac{9\phi^2 - 1}{n}}d\phi|_{\phi=\hat{\phi}}.$$
This can be integrated in closed form, giving the formula

$$g_{vm}(\hat{\phi}) = \arcsin\left[\sqrt{\frac{1 - 9/n}{1 - 1/n}} \hat{\phi}\right].$$

Notice that $g_{vm}$ has a much simpler form than the mean matched version $g_m$. Also, note that while the plain VST and the variance matched version are odd functions of $\hat{\phi}$, the mean matched version, annoyingly, is not. Plots of each of these three functions are given below. The plain and the variance matched versions are similar, the plain VST having a greater curvature. The mean matched VST is very different, and is actually unbounded as $\phi \rightarrow \pm 1$. 
Plain and Variance matched VST in AR(1) Case; Upper at High End = Plain VST
Example; MA(1) Consider next a moving average of order
1, given by the model \( X_t = Z_t + \theta Z_{t-1} \), where \( Z_t \sim WN(0, \sigma^2) \), and we have taken the stationary mean to be zero. We consider estimation of \( \theta \). Again, we use sample autocorrelations of lag 1, although any lag \( i \) can be used, as in the case of autoregression. Then, \( \sqrt{n} \left[ \hat{\rho}_1 - \frac{\theta}{1+\theta^2} \right] \overset{L}{\rightarrow} N(0, \frac{1+4\theta^2+\theta^4}{(1+\theta^2)^2}) \). Therefore, the plain VST is given by the transform \( g(\theta) = \int \frac{1+\theta^2}{\sqrt{1+4\theta^2+\theta^4}}. \)

Let \( F(\alpha, q), E(\alpha, q) \) denote respectively Elliptic integrals of the first and the second kind respectively. Then, by using formulae 3.152.1 and 3.153.1 in pp 276-277 in Gradshteyn and Ryzhik(2000),

\[
g(\theta) = \frac{1}{a} F(\arctan\left[\frac{\theta}{b}\right], \frac{c}{a}) + \theta \sqrt{\frac{a^2 + \theta^2}{b^2 + \theta^2}} - aE(\arctan\left[\frac{\theta}{b}\right], \frac{c}{a}),
\]

where \( a = \sqrt{2 + \sqrt{3}}, b = \sqrt{2 - \sqrt{3}}, c = 12^{\frac{1}{3}} \). This is a very unusual example of a variance stabilizing transformation. The mean matched version requires an asymptotic expansion for the bias of \( \hat{\rho}_1 \) in the MA(1) model.

**Example; ARIMA(0,d,0)** The stationary ARIMA(0,d,0) process is one of the most common processes that exhibit long range dependence when \( 0 < d < \frac{1}{2} \), and has been extensively used in many disciplines. Estimation of parameters involves atypical phenomena compared to short range dependence time series models. For example, the sample mean is no longer \( \sqrt{n} \) consistent, and all M estimates are asymptotically indistinguishable. Beran(1994) is a standard reference for classical methods.
under long range dependence. Bayesian theory is available in Koop et.al(1997), Jensen(2004), and DasGupta and Zang(2005). We consider estimation of the long memory parameter $d$.

4.13 General One Sample U - Statistics