Stat 582
Emerson, Win 12
Homework #2 Key
January 18, 2012

Written problems to be handed in Wednesday, January 25, 2012.

1. Let $X_1, X_2, \ldots$ be independent random variables having mean μ, variance σ^2, and finite fourth central moment μ_4. Further let $\gamma_2 = \mu_4/\sigma^4 - 3$ be the coefficient of excess kurtosis and $\phi = \gamma_2/2 + 1$, which can be viewed as an alternative coefficient of excess kurtosis. Based on the first n observations, define the first sample moment $\overline{X}_n$ and second central sample moment $\hat{\sigma}_n^2$ as

$$\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \hat{\sigma}_n^2 \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

and define sample variance $s_n^2 \equiv n\hat{\sigma}_n^2/(n-1)$.

a. Show that both $\hat{\sigma}_n^2$ and s_n^2 are asymptotically unbiased estimators of σ^2.

Ans: For the second central sample moment we have

$$\begin{aligned} E[\hat{\sigma}_n^2] &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i^2 - n(\overline{X}_n)^2\right] \\ &= \frac{1}{n}E\left[\sum_{i=1}^{n} X_i^2\right] - E\left[(\overline{X}_n)^2\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i^2\right] - E\left[(\overline{X}_n)^2\right] \\ &= E\left[X_i^2\right] - E\left[(\overline{X}_n)^2\right] = \left(\sigma^2 + \mu^2\right) - \left(\frac{\sigma^2}{n} + \mu^2\right) \\ &= \frac{n-1}{n}\sigma^2 \to \sigma^2 \quad \text{as } n \to \infty, \end{aligned}$$

and $bias(\hat{\sigma}_n^2) = E[\hat{\sigma}_n^2] - \sigma^2 = -\sigma^2/n$ converges to 0 as $n \to \infty$, so it is asymptotically unbiased. For the sample variance

$$E[s_n^2] = E\left[\frac{n}{n-1}\hat{\sigma}_n^2\right] = \frac{n}{n-1}\left[\frac{n-1}{n}\sigma^2\right] = \sigma^2,$$

so s_n^2 is unbiased, and thus asymptotically unbiased.

b. Derive expressions for the variance of $\hat{\sigma}_n^2$ and s_n^2. Express the variances both in terms of γ_2 and ϕ.

Ans: Because the variance (and sample variance) is unaffected by a location shift, we can for this part of the problem assume $\mu = 0$ without loss of generality. Working first with the second central sample moment, we note

$$\begin{aligned} Var(\hat{\sigma}_n^2) = E(\hat{\sigma}_n^4) - E^2(\hat{\sigma}_n^2) &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right]^2 - \frac{(n-1)^2}{n^2}\sigma^4 \\ &= E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \overline{X}_n^2\right]^2 - \frac{(n-1)^2}{n^2}\sigma^4 \\ &= E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_i^2\ X_j^2 - \frac{2}{n}\sum_{i=1}^{n} X_i^2\overline{X}_n^2 + \overline{X}_n^4\right] - \frac{(n-1)^2}{n^2}\sigma^4. \\ &= \frac{1}{n^2}E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} X_i^2\ X_j^2\right] - \frac{2}{n}E\left[\sum_{i=1}^{n} X_i^2\overline{X}_n^2\right] + E\left[\overline{X}_n^4\right] - \frac{(n-1)^2}{n^2}\sigma^4. \end{aligned}$$

We thus consider the first term, first finding

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} X_i^2\ X_j^2\right] &= \sum_{i=j} E[X_i^4] + \sum_{i\neq j} E[X_i^2]E[X_j^2] \\
&= n\mu_4 + n(n-1)\sigma^4,
\end{aligned}
$$

where we have used the independence of the random variables to find $E[X_i^2 X_j^2] = E[X_i^2]E[X_j^2]$ for $i \neq j$, and the fact that $E[X_i^2] = \sigma^2$ under the WLOG assumption that $\mu = 0$, and the identical distribution of the X_i's. For the second term we find

$$
\begin{aligned}
E\left[\sum_{i=1}^{n} X_i^2 \overline{X}_n^2\right] &= E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} X_i^2 X_j X_k\right] \\
&= \frac{1}{n^2}\Big[\sum_{i=j=k} E[X_i^4] + \sum_{i\neq j=k} E[X_i^2]E[X_j^2] + \sum_{i=j\neq k} E[X_i^3]E[X_k] + \\
&\qquad \sum_{i=k\neq j} E[X_i^3]E[X_j] + \sum_{i\neq j\neq k\neq i} E[X_i]^2 E[X_j]E[X_k]\Big] \\
&= \frac{1}{n^2}\left[n\mu_4 + n(n-1)\sigma^4 + 0 + 0 + 0\right] = \frac{1}{n}\mu_4 + \frac{n-1}{n}\sigma^4,
\end{aligned}
$$

where again we have used the fact that the random variables are i.i.d and the WLOG assumption that $E[X_i] = 0$. Finally, for the third term we find

$$
\begin{aligned}
E\left[\overline{X}_n^4\right] &= \frac{1}{n^4} E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n} X_i X_j X_k X_\ell\right] \\
&= \frac{1}{n^4}\Big[\sum_{i=j=k=\ell} E[X_i^4] + \sum_{i=j\neq k=\ell} E[X_i^2]E[X_k^2] + \sum_{i=k\neq j=\ell} E[X_i^2]E[X_k^2] + \\
&\qquad \sum_{i=\ell\neq j=k} E[X_i^2]E[X_k^2] + 0\Big] \\
&= \frac{1}{n^4}\left[n\mu_4 + 3n(n-1)\sigma^4\right] = \frac{1}{n^3}\mu_4 + \frac{3(n-1)}{n^3}\sigma^4
\end{aligned}
$$

Putting this all together, we find

$$
\begin{aligned}
Var(\hat{\sigma}_n^2) &= \left(\frac{1}{n} - \frac{2}{n^2} + \frac{1}{n^3}\right)\mu_4 + \left(\frac{n-1}{n} - \frac{2(n-1)}{n^2} + \frac{3(n-1)}{n^3} - \frac{(n-1)^2}{n^2}\right)\sigma^4 \\
&= \frac{(n-1)^2}{n^3}\mu_4 - \frac{(n-1)(n-3)}{n^3}\sigma^4
\end{aligned}
$$

We can then find the variance of s_n^2 as

$$
Var(s_n^2) = Var\left(\frac{n}{n-1}\hat{\sigma}_n^2\right) = \frac{n^2}{(n-1)^2}\left[\frac{(n-1)^2}{n^3}\mu_4 - \frac{(n-1)(n-3)}{n^3}\sigma^4\right] = \frac{1}{n}\mu_4 - \frac{n-3}{n(n-1)}\sigma^4.
$$

In terms of the coefficient of excess kurtosis γ_2, we use $\mu_4 = (\gamma_2 + 3)\sigma^4$ to find

$$
\begin{aligned}
Var(\hat{\sigma}_n^2) &= \frac{n-1}{n^3}\left[(n-1)\gamma_2 + 2n\right]\sigma^4 \\
Var(s_n^2) &= \frac{1}{n(n-1)}\left[(n-1)\gamma_2 + 2n\right]\sigma^4
\end{aligned}
$$

and in terms of the alternative coefficient of excess kurtosis $\phi = \gamma_2/2 + 1$,

$$Var(\hat{\sigma}_n^2) = \frac{2(n-1)}{n^3}\left[(n-1)\phi + 1\right]\sigma^4$$
$$Var(s_n^2) = \frac{2}{n(n-1)}\left[(n-1)\phi + 1\right]\sigma^4$$

c. Which estimator has lower variance? Which estimator has lower mean squared error?

Ans: $Var(\hat{\sigma}_n^2)/Var(s_n^2) = (n-1)^2/n^2 < 1$, thus $\hat{\sigma}_n^2$ has lower variance. Using the fact that mean squared error is the sum of the variance and the square of the bias, we find

$$\begin{aligned} MSE(s_n^2) - MSE(\hat{\sigma}_n^2) &= Var(s_n^2) - Var(\hat{\sigma}_n^2) - bias^2(\hat{\sigma}_n^2) \\ &= \frac{2}{n(n-1)}\left[(n-1)\phi + 1\right]\sigma^4 - \frac{2(n-1)}{n^3}\left[(n-1)\phi + 1\right]\sigma^4 - \frac{1}{n^2}\sigma^4 \\ &= \left[(4n^2 - 6n + 2)\phi - (n^2 - 5n + 2)\right]\frac{\sigma^4}{n^3(n-1)}. \end{aligned}$$

Now $4n^2 - 6n + 2 > 0$ for $n > 1$ (and with $n = 1$, $\hat{\sigma}_n^2 \equiv 0$ and s_n^2 is undefined, so it seems silly to worry about that case). Hence we can find that the difference is positive (so a larger MSE for the sample variance s_n^2 than for $\hat{\sigma}_n^2$) for distributions having

$$\phi > \frac{(n^2 - 5n + 2)}{(4n^2 - 6n + 2)}.$$

This threshold is increasing in n for $n > 1$. The quadratic term $(4n^2 - 6n + 2)$ in the denominator is positive for $n > 1$, and the quadratic term in the numerator is positive for $n \geq 5$. Hence for $2 \leq n \leq 4$ the sample variance has larger MSE for all distributions having a finite variance, at $n = 5$ the sample variance has larger MSE for all distributions having $\phi > 1/36$, and at $n = 10$, the sample variance has larger MSE for all distributions having $\phi > 0.152$. As $n \to \infty$, the threshold increases to an asymptotic limit dictating that the sample variance has larger MSE for all distributions having $\phi > 0.25$.

d. Without appealing to the weak law of large numbers, show that both $\hat{\sigma}_n^2$ and s_n^2 are consistent estimators of σ^2.

Ans: Both $\hat{\sigma}_n^2$ and s_n^2 are asymptotically unbiased and have variances that asymptotically approach 0. Hence by problem 3 of homework # 1, we know that $\hat{\sigma}_n^2 \to_p \sigma^2$ and $s_n^2 \to_p \sigma^2$, and thus both estimators are consistent for σ^2.

2. In the setting of problem 1 with identically distributed random variables, the Levy Central Limit Theorem provides that

$$Z_n \equiv \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \to_d \mathcal{N}(0, 1).$$

In practice, however, we rarely know σ and thus typically compute

$$T_n \equiv \sqrt{n}\frac{\overline{X}_n - \mu}{s_n}.$$

a. Show that $T_n \to_d Z$ for some random variable Z. What is the distribution of Z?

Ans: By part d of problem 1, $s_n^2 \to_p \sigma^2$. Noting that $g(u) = \sigma/\sqrt{u}$ is a continuous function for $u > 0$, the continuous mapping theorem (Mann-Wald) tells us that $\sigma/s_n = g(\sigma_n^2) \to_p g(\sigma^2) = 1$. Hence, we can use Slutsky's theorem to show

$$T_n = \frac{\sigma}{s_n} Z_n \to_d 1 \cdot Z = Z \;\sim\; \mathcal{N}(0, \sigma^2).$$

b. Show that inference based on using a t distribution for T_n is asymptotically correct for arbitrary distributions in this setting.

Ans: In the one sample t test, we typically base inference on the assumption that $T_n \sim t_{n-1}$, t distribution with $n-1$ degrees of freedom. We thus calculate $Pr(T_n \le c) = Pr(t_{n-1} \le c)$. Hence, we want to show that $\forall c \; Pr(t_{n-1} \le c) \to \Phi(c)$, where Φ is the cumulative distribution function for a standard normal.

Now the t_n distribution is defined as the distribution of $Z/\sqrt{X/n}$, where $Z \sim \mathcal{N}(0,1)$ is independent of $X \sim \chi^2_n$, a chi squared distribution with n degrees of freedom. And the chi squared distribution having n degrees of freedom is defined as the distribution of $X = \sum_{i=1}^n W_i^2$, for $W_1, \ldots, W_n$ independent, identically distributed random variables with $W_i \sim \mathcal{N}(0,1)$.

So, letting Z and $W_1, W_2, \ldots$ be a sequence of totally independent standard normal random variables, and defining

$$Z_n = Z \qquad \text{and} \qquad X_n = \sum_{i=1}^n W_i^2 \qquad \text{and} \qquad T_n = \frac{Z_n}{\sqrt{X_n/n}},$$

we have $X_n \sim \chi^2_n$ and $T_n \sim t_{n-1}$. But the weak law of large numbers tells us that

$$\frac{X_n}{n} = \frac{1}{n}\sum_{i=1}^n W_i^2 \to_p E[W_i^2] = 1,$$

and because $g(u) = 1/\sqrt{u}$ is a continuous function for $u > 0$, the continuous mapping theorem (Mann-Wald) provides that $1/\sqrt{X_n/n} \to_p 1$. And because $Z_n \to_d Z \sim \mathcal{N}(0,1)$ (trivially), Slutsky's theorem tells us

$$T_n = \frac{1}{\sqrt{X_n/n}} Z_n \to_d 1 \cdot Z = Z \sim \mathcal{N}(0,1),$$

and the distribution functions for t_n random variables converge to the distribution function for a standard normal. (And as $n \to \infty$, so does $n - p \to \infty$ for any fixed, finite p, so this result holds for standard inference in regression problems quite generally, once we establish a central limit theorem for ordinary least squares estimates.)

c. Suppose we want to know that T_n is arbitrarily close to Z_n. Derive an expression that would allow you to put bounds on the probability that Z_n/T_n might be far from 1 for an arbitrary distribution.

Ans: Noting that $Z_n/T_n = s_n/\sigma$, for an arbitrary small $\epsilon > 0$ we find using Chebyshev's inequality (eventually) that

$$\begin{aligned}
Pr\left(\left|\frac{s_n}{\sigma} - 1\right| \ge \epsilon\right) &= 1 - Pr\left(-\epsilon < \frac{s_n}{\sigma} - 1 < \epsilon\right) \\
&= 1 - Pr\left(1 - \epsilon < \frac{s_n}{\sigma} < 1 + \epsilon\right) \\
&= 1 - Pr\left(\epsilon^2 - 2\epsilon < \frac{s_n^2 - \sigma^2}{\sigma^2} < \epsilon^2 + 2\epsilon\right) \\
&\le 1 - Pr\left(\left|\frac{s_n^2 - \sigma^2}{\sigma^2}\right| < \epsilon^2 + 2\epsilon\right) \\
&= Pr\left(\left|\frac{s_n^2 - \sigma^2}{\sigma^2}\right| \ge \epsilon^2 + 2\epsilon\right) \\
&\le \frac{1}{(\epsilon^2 + 2\epsilon)^2} Var\left(\frac{s_n^2}{\sigma^2}\right) \qquad \text{by Chebyshev's inequality} \\
&= \frac{2\left[(n-1)\phi + 1\right]}{n(n-1)(\epsilon^2 + 2\epsilon)^2}.
\end{aligned}$$

d. Find an expression for the smallest sample size n such that $Pr(\,|1 - Z_n/T_n| \geq \delta) \leq \epsilon$ for arbitrary $\delta > 0$ and arbitrary $\epsilon > 0$ as a function of γ_2. Also provide your expression as a function of ϕ.

Ans: Using the results of part c, we want to find n such that

$$\epsilon = \frac{2\,[\,(n-1)\phi + 1\,]}{n(n-1)(\delta^2 + 2\delta)^2}.$$

Letting $\kappa = \epsilon(\delta^2 + 2\delta)^2$, we thus find the solution to

$$n^2 - \left(\frac{\kappa + 2\phi}{\kappa}\right) n + \frac{2(\phi - 1)}{\kappa} = 0,$$

yielding

$$n = \frac{1}{2} + \frac{\phi}{\kappa} \pm \sqrt{\left(\frac{\phi}{\kappa} - \frac{1}{2}\right)^2 + \frac{2}{\kappa}}.$$

Noting that $\phi = \gamma_2/2 + 1$ this becomes

$$n = \frac{1}{2} + \frac{\gamma_2 + 2}{2\kappa} \pm \sqrt{\left(\frac{1}{2} - \frac{\gamma_2 + 2}{2\kappa}\right)^2 + \frac{2}{\kappa}}.$$

e. For what distribution does the expression in part d achieve its minimum?

Ans: Note that the range of ϕ is $[0, \infty)$, because $\mu_4 = E(X_i - \mu)^4 \geq [E(X_i - \mu)^2]^2 = \sigma^4$ owing to the convexity of the function $g(u) = u^2$ along with Jensen's inequality. Within that range, the expression in part d is minimized by $\phi = 0$.

f. Suppose that the X_i's are known to be normally distributed. How does the bound you found in part c compare with the probability computed using the exact distribution?

Ans: If we have $X_i \sim \mathcal{N}(\mu, \sigma^2)$, then $\phi = 1$ and $(n-1)s_n^2/\sigma^2 \sim \chi^2_{n-1}$. So the expression for part d becomes

$$n = \frac{1}{2} + \frac{1}{\kappa} \pm \sqrt{\left(\frac{1}{\kappa} - \frac{1}{2}\right)^2 + \frac{2}{\kappa}}.$$

Using the exact distribution, we want to find n such that

$$P\left(\,(n-1)(1-\delta) < \chi^2_{n-1} < (n-1)(1+\delta)\,\right) = 1 - \epsilon.$$

This yields a much smaller sample size. For instance, choosing $\delta = .1$ and $\epsilon = .1$, using the following R code

```
delta <- 0.1
epsilon <- 0.1
k <- epsilon * (delta^ 2 + 2 * delta)^ 2
phi <- 1
n <- 2:1000
e <- pchisq((n-1)*(1+delta)^ 2,n-1) - pchisq((n-1)*(1-delta)^ 2,n-1))
cbind(exact=sum(e<1-epsilon) + 1,
    Chebyshev=1/2 + phi/k + sqrt((phi/k - 1/2)^ 2 + 2/k))
```

we get a sample size of 136 using the exact distribution and 455 using Chebyshev.

g. What do the above results suggest about the sample sizes for which approximate statistical inference based on the asymptotic distribution of T_n will be accurate?

Ans: The results of part d do not tell us very much, because the Chebyshev bound is not very tight in general. Formulas based on an asymptotic distribution for s_n^2 (such as found in problem 3) will be far more informative. But even then, we will find that the larger the coefficient of excess kurtosis, the large sample size will be required.

3. Again, consider the setting of problem 1 with identically distributed random variables.

a. Show that

$$W_n \equiv \sqrt{n}\left(s_n^2 - \sigma^2\right) \rightarrow_d W$$

for some random variable W, and specify the distribution of W.

Ans: We first note that $X_i \sim_{iid} (\mu, \sigma^2, \mu_3, \mu_4)$ implies $Y_i = (X_i - \mu)^2$ are independent and identically distributed with $E(Y_i) = \sigma^2$ and $Var(Y_i) = \mu_4 - \sigma^4$. Hence, by the Levy central limit theorem we know

$$\sqrt{n}(\overline{Y}_n - \sigma^2) \rightarrow_d \mathcal{N}(0, \mu_4 - \sigma^4).$$

However,

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 + \frac{1}{n}\sum_{i=1}^{n}(\overline{X}_n - \mu)^2 = \hat{\sigma}_n^2 + (\overline{X}_n - \mu)^2,$$

so substituting this in the above asymptotic result we have

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) + \sqrt{n}(\overline{X}_n - \mu)^2 \rightarrow_d \mathcal{N}(0, \mu_4 - \sigma^4).$$

But by the Levy CLT we have $\sqrt{n}(\overline{X}_n - \mu) \rightarrow_d \mathcal{N}(0, 1)$ and by the WLLN we have $(\overline{X}_n - \mu) \rightarrow_p 0$, so by Slutsky's, we have $\sqrt{n}(\overline{X}_n - \mu)^2 \rightarrow_p 0$, and we thus obtain

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \rightarrow_d \mathcal{N}(0, \mu_4 - \sigma^4).$$

Furthermore, because $\hat{\sigma}_n^2 = (n-1)s_n^2/n$, we have

$$\sqrt{n}(s_n^2 - \sigma^2) - \frac{1}{\sqrt{n}}\hat{\sigma}_n^2 \rightarrow_d \mathcal{N}(0, \mu_4 - \sigma^4),$$

and because $\hat{\sigma}_n^2 \rightarrow_p \sigma^2$ and $1/\sqrt{n} \rightarrow 0$, then $\frac{1}{\sqrt{n}}\hat{\sigma}_n^2 \rightarrow_p 0$, and we obtain

$$W_n = \sqrt{n}(s_n^2 - \sigma^2) \rightarrow_d \mathcal{N}(0, \mu_4 - \sigma^4) =_d \mathcal{N}(0, 2\phi\sigma^4).$$

Note that for $X_i \sim \mathcal{B}(1, 0.5)$, $\mu = 0.5$, $\sigma^2 = 0.25$, $\mu_4 = 0.25$, $\phi = 0$, and

$$\sqrt{n}(s_n^2 - \sigma^2) \rightarrow_p 0.$$

b. Suppose again that the X_i's are normally distributed. What is the exact distribution of W_n? Compare the accuracy of using an approximate distribution based on W to the exact distribution for W_n for a range of quantiles and a range of sample sizes.

Ans: Based on the asymptotics, we would use approximate distribution

$$s_n^2 \dot{\sim} \mathcal{N}\left(\sigma^2, \frac{2\phi\sigma^4}{n}\right).$$

For $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $\phi = 1$ and

$$(n-1)\frac{s_n^2}{\sigma^2} \sim \chi^2_{n-1}.$$

The following table presents selected quantiles for selected samples sizes based on the exact ("Ex") and asymptotic ("As") distributions when the variance $\sigma^2 = 1$. Note that the exact distribution is skewed in small samples, while the asymptotic distribution is symmetric. Clear from the table is the need for quite large sample sizes in order for the asymptotic distribution to be a good approximation in the tails.

Table 1
Comparison of Exact and Asymptotic Quantiles of Sample Variance for Normal Data

			Quantiles						
	p	n=5	n=10	n=25	n=50	n=75	n=100	n=200	n=500
Ex	0.001	0.023	0.128	0.337	0.489	0.568	0.618	0.719	0.816
As	0.001	-0.954	-0.382	0.126	0.382	0.495	0.563	0.691	0.805
Ex	0.005	0.052	0.193	0.412	0.556	0.627	0.672	0.761	0.844
As	0.005	-0.629	-0.152	0.271	0.485	0.579	0.636	0.742	0.837
Ex	0.010	0.074	0.232	0.452	0.591	0.658	0.699	0.782	0.859
As	0.010	-0.471	-0.040	0.342	0.535	0.620	0.671	0.767	0.853
Ex	0.025	0.121	0.300	0.517	0.644	0.704	0.741	0.813	0.880
As	0.025	-0.240	0.123	0.446	0.608	0.680	0.723	0.804	0.876
Ex	0.050	0.178	0.369	0.577	0.692	0.746	0.778	0.841	0.898
As	0.050	-0.040	0.264	0.535	0.671	0.731	0.767	0.836	0.896
Ex	0.100	0.266	0.463	0.652	0.751	0.796	0.823	0.874	0.920
As	0.100	0.189	0.427	0.638	0.744	0.791	0.819	0.872	0.919
Ex	0.250	0.481	0.655	0.793	0.857	0.885	0.901	0.931	0.957
As	0.250	0.573	0.698	0.809	0.865	0.890	0.905	0.933	0.957
Ex	0.500	0.839	0.927	0.972	0.986	0.991	0.993	0.997	0.999
As	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ex	0.750	1.346	1.265	1.177	1.128	1.105	1.092	1.066	1.042
As	0.750	1.427	1.302	1.191	1.135	1.110	1.095	1.067	1.043
Ex	0.900	1.945	1.632	1.383	1.266	1.216	1.186	1.130	1.082
As	0.900	1.811	1.573	1.362	1.256	1.209	1.181	1.128	1.081
Ex	0.950	2.372	1.880	1.517	1.354	1.285	1.245	1.170	1.106
As	0.950	2.040	1.736	1.465	1.329	1.269	1.233	1.164	1.104
Ex	0.975	2.786	2.114	1.640	1.433	1.347	1.297	1.206	1.128
As	0.975	2.240	1.877	1.554	1.392	1.320	1.277	1.196	1.124
Ex	0.990	3.319	2.407	1.791	1.529	1.422	1.360	1.248	1.153
As	0.990	2.471	2.040	1.658	1.465	1.380	1.329	1.233	1.147
Ex	0.995	3.715	2.621	1.898	1.597	1.474	1.404	1.277	1.171
As	0.995	2.629	2.152	1.729	1.515	1.421	1.364	1.258	1.163
Ex	0.999	4.617	3.097	2.132	1.742	1.586	1.497	1.339	1.207
As	0.999	2.954	2.382	1.874	1.618	1.505	1.437	1.309	1.195

c. Now consider non-normal X_i's. Compare the accuracy of using an approximate distribution based on W to using the exact distribution for W_n when the data are normally distributed. Consider several distributions covering a range of departures from the normal distribution.

Ans: The key issue here is that using the chi squared distribution is in large samples presuming that $\phi = 1$. It is thus of interest to consider a few distributions that differ in the value of ϕ. In all cases, I standardized so the variance $\sigma^2 = 1$.

- Exponential with mean $\mu = 1$, so $\sigma^2 = 1$ and $\phi = 4$.

- Lognormal with $meanlog = 0.496044$, $sdlog^2 = 0.2530205$, so $\sigma^2 = 1$ and $\phi = 4$.
- A Bernoulli with $p = 0.9082483$ and then multiplied by $\sqrt{12}$, so $\sigma^2 = 1$ and $\phi = 4$.
- Gamma with $scale = 1/\sqrt{3}$ and $shape = 3$, so $\sigma^2 = 1$ and $\phi = 2$.
- A Bernoulli with $p = 0.8535534$ and then multiplied by $\sqrt{8}$, so $\sigma^2 = 1$ and $\phi = 2$.
- A uniform $\mathcal{U}(0, \sqrt{12})$, so $\sigma^2 = 1$ and $\phi = 0.4$
- A Bernoulli with $p = 0.7041241$ and then multiplied by $\sqrt{4.8}$, so $\sigma^2 = 1$ and $\phi = 0.4$.
- A Bernoulli with $p = 0.5$ and then multiplied by $\sqrt{4}$, so $\sigma^2 = 1$ and $\phi = 0$.

The following table contains the quantiles estimated assuming a normal distribution for the data (so using the χ^2_{n-1} distribution and labeled "N"), the quantiles estimated using the true value of ϕ and the asymptotic distribution (labeled "As"), and the quantiles estimated using 10,000 simulated data sets (labeled by the distribution used to simulate the data and "Si"). As might be expected, the quantiles using the assumption of normality are incorrect for all of these cases in which $\phi \neq 1$.

Apparent again is the relatively large sample size that is needed to obtain a good approximation in the tails of the distribution, as well as the tendency for worse approximations for higher values of ϕ. (Note that when $\phi = 0$, the asymptotic distribution is degenerate.) Note also that the small sample properties of the approximation do appear to depend on other aspects of the distribution beyond σ^2 and μ_4, with the lognormal and exponential cases seeming to be less well approximated by the asymptotic distribution than the corresponding Bernoulli, and the uniform case seeming to be better approximated by the asymptotic distribution than the corresponding Bernoulli.

Table 2
Comparison of Exact and Asymptotic Quantiles of Sample Variance for Non-normal Data

			Quantiles					
		p	n=10	n=25	n=50	n=100	n=200	n=500
Normal (χ^2_{n-1})	N	0.025	0.300	0.517	0.644	0.741	0.813	0.880
Asymptotic ($\phi = 4$)	As	0.025	-0.753	-0.109	0.216	0.446	0.608	0.752
Exponential ($\phi = 4$)	Si	0.025	0.134	0.303	0.439	0.562	0.667	0.778
Lognormal ($\phi = 4$)	Si	0.025	0.181	0.346	0.476	0.592	0.690	0.788
Bernoulli ($\phi = 4$)	Si	0.025	0.000	0.000	0.240	0.465	0.627	0.741
Normal (χ^2_{n-1})	N	0.500	0.927	0.972	0.986	0.993	0.997	0.999
Asymptotic ($\phi = 4$)	As	0.500	1.000	1.000	1.000	1.000	1.000	1.000
Exponential ($\phi = 4$)	Si	0.500	0.755	0.876	0.929	0.962	0.974	0.993
Lognormal ($\phi = 4$)	Si	0.500	0.753	0.871	0.919	0.949	0.971	0.989
Bernoulli ($\phi = 4$)	Si	0.500	1.200	0.920	0.901	0.993	0.988	1.004
Normal (χ^2_{n-1})	N	0.975	2.114	1.640	1.433	1.297	1.206	1.128
Asymptotic ($\phi = 4$)	As	0.975	2.753	2.109	1.784	1.554	1.392	1.248
Exponential ($\phi = 4$)	Si	0.975	3.543	2.457	1.983	1.666	1.439	1.274
Lognormal ($\phi = 4$)	Si	0.975	3.217	2.420	1.985	1.653	1.448	1.272
Bernoulli ($\phi = 4$)	Si	0.975	2.800	2.000	1.807	1.545	1.408	1.251

Table 2 (cont.)
Comparison of Exact and Asymptotic Quantiles of Sample Variance for Non-normal Data

		p	Quantiles n=10	n=25	n=50	n=100	n=200	n=500
Normal (χ^2_{n-1})	N	0.025	0.300	0.517	0.644	0.741	0.813	0.880
Asymptotic ($\phi = 2$)	As	0.025	-0.240	0.216	0.446	0.608	0.723	0.825
Gamma ($\phi = 2$)	Si	0.025	0.235	0.426	0.561	0.670	0.751	0.837
Bernoulli ($\phi = 2$)	Si	0.025	0.000	0.320	0.460	0.595	0.724	0.822
Normal (χ^2_{n-1})	N	0.500	0.927	0.972	0.986	0.993	0.997	0.999
Asymptotic ($\phi = 2$)	As	0.500	1.000	1.000	1.000	1.000	1.000	1.000
Gamma ($\phi = 2$)	Si	0.500	0.848	0.929	0.965	0.980	0.989	0.996
Bernoulli ($\phi = 2$)	Si	0.500	0.800	1.120	0.983	0.973	0.997	0.999
Normal (χ^2_{n-1})	N	0.975	2.114	1.640	1.433	1.297	1.206	1.128
Asymptotic ($\phi = 2$)	As	0.975	2.240	1.784	1.554	1.392	1.277	1.175
Gamma ($\phi = 2$)	Si	0.975	2.715	1.979	1.652	1.453	1.305	1.189
Bernoulli ($\phi = 2$)	Si	0.975	2.133	1.680	1.489	1.387	1.286	1.173
Normal (χ^2_{n-1})	N	0.025	0.300	0.517	0.644	0.741	0.813	0.880
Asymptotic ($\phi = 0.4$)	As	0.025	0.446	0.649	0.752	0.825	0.876	0.922
Uniform ($\phi = 0.4$)	Si	0.025	0.408	0.639	0.747	0.824	0.878	0.923
Bernoulli ($\phi = 0.4$)	Si	0.025	0.000	0.537	0.735	0.818	0.882	0.936
Normal (χ^2_{n-1})	N	0.500	0.927	0.972	0.986	0.993	0.997	0.999
Asymptotic ($\phi = 0.4$)	As	0.500	1.000	1.000	1.000	1.000	1.000	1.000
Uniform ($\phi = 0.4$)	Si	0.500	0.988	0.995	0.997	1.000	0.999	1.000
Bernoulli ($\phi = 0.4$)	Si	0.500	1.139	1.025	1.046	1.015	1.020	1.019
Normal (χ^2_{n-1})	N	0.975	2.114	1.640	1.433	1.297	1.206	1.128
Asymptotic ($\phi = 0.4$)	As	0.975	1.554	1.351	1.248	1.175	1.124	1.078
Uniform ($\phi = 0.4$)	Si	0.975	1.659	1.380	1.257	1.181	1.126	1.079
Bernoulli ($\phi = 0.4$)	Si	0.975	1.356	1.269	1.213	1.173	1.130	1.088
Normal (χ^2_{n-1})	N	0.025	0.300	0.517	0.644	0.741	0.813	0.880
Asymptotic ($\phi = 0$)	As	0.025	1.000	1.000	1.000	1.000	1.000	1.000
Bernoulli ($\phi = 0$)	Si	0.025	0.711	0.840	0.916	0.961	0.979	0.992
Normal (χ^2_{n-1})	N	0.500	0.927	0.972	0.986	0.993	0.997	0.999
Asymptotic ($\phi = 0$)	As	0.500	1.000	1.000	1.000	1.000	1.000	1.000
Bernoulli ($\phi = 0$)	Si	0.500	1.067	1.027	1.014	1.006	1.003	1.001
Normal (χ^2_{n-1})	N	0.975	2.114	1.640	1.433	1.297	1.206	1.128
Asymptotic ($\phi = 0$)	As	0.975	1.000	1.000	1.000	1.000	1.000	1.000
Bernoulli ($\phi = 0$)	Si	0.975	1.111	1.040	1.020	1.010	1.005	1.002

4. Suppose we now have two groups of interest. Let $X_1, X_2, \ldots$ be independent random variables having mean μ, variance $\sigma^2 < \infty$ and $Y_1, Y_2, \ldots$ be independent random variables having mean ν, variance $\tau^2 < \infty$. Based on the first n observations of the X_i's and the first m observations of the Y_i's define sample means $\overline{X}_n$ and $\overline{Y}_m$ and sample variances s^2_{Xn} and s^2_{Ym}. Further define pooled sample variance

$$s^2_{Pmn} = \frac{(n-1)s^2_{Xn} + (m-1)s^2_{Ym}}{m+n-2}.$$

We consider the following variations of t tests comparing μ to ν:

$$T_{mn}^{(e)} = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu - \nu)}{s_{Pmn}\sqrt{\frac{1}{n} + \frac{1}{m}}}$$
$$T_{mn}^{(u)} = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu - \nu)}{\sqrt{\frac{s_{Xn}^2}{n} + \frac{s_{Ym}^2}{m}}}$$
$$T_{mn}^{(x)} = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu - \nu)}{s_{Xn}\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

a. Derive asymptotic distributions for each of the above statistics as $n \to \infty$ and $m/n \to \lambda$, a positive constant. Provide sufficient conditions that each of those limiting distributions would have variance 1. Under what conditions would each statistic out perform the others to detect a difference between μ and ν?

Ans: Note that from problems 2 and 3 above, we know that as $k \to \infty$

$$\sqrt{k}(\overline{X}_k - \mu) \to_d \mathcal{N}(0, \sigma^2)$$
$$\sqrt{k}(\overline{Y}_k - \nu) \to_d \mathcal{N}(0, \tau^2)$$
$$s_{Xk}^2 \to_p \sigma^2$$
$$s_{Yk}^2 \to_p \tau^2$$

Now consider a sequence $n_1, n_2, \ldots$ such that $n_{i+1} \geq n_i$ for $i \geq 1$ and $n_k \to \infty$ as $k \to \infty$. Similarly consider $m_1, m_2, \ldots$ such that $m_{i+1} \geq m_i$ for $i \geq 1$ and $m_k \to \infty$ as $k \to \infty$. Furthermore, suppose $m_k/n_k \to \lambda$. Then we have

$$W_k \equiv \sqrt{n_k}(\overline{X}_{n_k} - \mu) \to_d \mathcal{N}(0, \sigma^2) \equiv W$$
$$V_k \equiv \sqrt{m_k}(\overline{Y}_{m_k} - \nu) \to_d \mathcal{N}(0, \tau^2) \equiv V$$
$$s_{Xn_k}^2 \to_p \sigma^2$$
$$s_{Ym_k}^2 \to_p \tau^2,$$

as $k \to \infty$. Furthermore, by the independence of the X_i's and Y_i's, we know that the joint distribution of (W_k, V_k) satisfies

$$P(W_k \leq w, V_k \leq v) = P(W_k \leq w)P(V_k \leq v),$$

and because the normal distribution function is continuous everywhere and we have $W_k \to_d W$ and $V_k \to_d V$,

$$P(W_k \leq w)P(V_k \leq v) \to P(W \leq w)P(V \leq v),$$

which is the cdf for a bivariate normal so

$$\begin{pmatrix} W_n \\ V_n \end{pmatrix} \to_d \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \tau^2 \end{pmatrix}\right).$$

Now, let

$$\vec{a}_k = \begin{pmatrix} \sqrt{\frac{m_k}{n_k+m_k}} \\ -\sqrt{\frac{n_k}{n_k+m_k}} \end{pmatrix} \to \begin{pmatrix} \sqrt{\frac{\lambda}{1=\lambda}} \\ -\sqrt{\frac{1}{1+\lambda}} \end{pmatrix} \equiv \vec{a}.$$

Then

$$\vec{a}_k' \begin{pmatrix} W_n \\ V_n \end{pmatrix} = \sqrt{\frac{m_k n_k}{n_k + m_k}}\left((\overline{X}_{n_k} - \overline{Y}_{m_k}) - (\mu - \nu)\right) \to_d \mathcal{N}\left(0, \frac{\lambda}{1+\lambda}\sigma^2 + \frac{1}{1+\lambda}\tau^2\right),$$

by Slutsky's theorem. This gives the asymptotic distribution of the numerator of the T statistics (after inflation by a factor of $\sqrt{mn/(m+n)}$.

Now we consider the various denominators (where we also multiply by a factor of $\sqrt{mn/(m+n)}$ so it will cancel with our previous result).

$$\sqrt{\frac{m_k n_k}{n_k+m_k}} s_{Pmn}\sqrt{\frac{1}{n_k}+\frac{1}{m_k}} = \sqrt{\frac{(n_k-1)s^2_{Xn_k}+(m_k-1)s^2_{Ym_k}}{m_k+n_k-2}} \rightarrow_p \sqrt{\frac{1}{1+\lambda}\sigma^2+\frac{\lambda}{1+\lambda}\tau^2},$$

using Slutsky's and Mann-Wald. Thus by Slutsky's

$$T_k^{(e)} \rightarrow_d \mathcal{N}\left(0, \frac{\frac{\lambda}{1+\lambda}\sigma^2+\frac{1}{1+\lambda}\tau^2}{\frac{1}{1+\lambda}\sigma^2+\frac{\lambda}{1+\lambda}\tau^2} = \frac{\lambda\sigma^2+\tau^2}{\sigma^2+\lambda\tau^2}\right),$$

which is the standard normal distribution if $\sigma^2 = \tau^2$ OR $\lambda = 1$.

For the denominator of $T_n^{(u)}$

$$\sqrt{\frac{m_k n_k}{n_k+m_k}}\sqrt{\frac{s^2_{Xn_k}}{n_k}+\frac{s^2_{Ym}}{m}} \rightarrow_p \sqrt{\frac{\lambda}{1+\lambda}\sigma^2+\frac{1}{1+\lambda}\tau^2},$$

so using Slutsky's we find

$$T^{(u)} \rightarrow_d \mathcal{N}(0,1)$$

in all cases.

For the denominator of $T_n^{(x)}$

$$\sqrt{\frac{m_k n_k}{n_k+m_k}} s_{Xn_k}\sqrt{\frac{1}{n_k}+\frac{1}{m_k}} \rightarrow_p \sigma,$$

so using Slutsky's we find

$$T^{(x)} \rightarrow_d \mathcal{N}\left(0, \frac{\lambda}{1+\lambda}+\frac{1}{1+\lambda}\frac{\tau^2}{\sigma^2}\right),$$

which is the standard normal distribution if $\sigma^2 = \tau^2$.

Note that any of the three statistics are valid to test the strong null hypothesis $H_0 : X \sim Y$, or the slightly weaker null hypothesis $H_0 : \mu = \nu; \sigma^2 = \tau^2$. Only the T test that allows for unequal variances can test the weakest null hypothesis of $H_0 : \mu = \nu$. Of course, the efficiency of the tests will depend upon how the distributions might differ under the alternative. In implementing the tests, we generally use the t distribution, and we tend to use a t with higher degrees of freedom when we use $T^{(e)}$, thereby gaining a bit more statistical power to test the stronger null hypotheses. If we posit a situation in which $\tau^2 > \sigma^2$ under the alternative (but not the null), then $T^{(x)}$ may be the most powerful of the three tests, but again in its implementation the degrees of freedom used in the inference can make a difference.

b. Derive asymptotic distributions for each of the above statistics as $\min(n, m) \rightarrow \infty$.

Ans: Here we appeal to the Lindeberg-Feller central limit theorem. Because we anticipate root-n consistency, the goal is to find independent random variables that have mean 0 and variance proportional to the sample size. From the results we obtained in part a, we anticipate that we want the sum S_k to look like $\sqrt{n_k m_m/(n_k+m_k)}\left[(\overline{X}_{n_k} - \overline{Y}_{m_k}) - (\mu-\nu)\right]$, so we do what it takes to get that.

Hence, for $k = 1, 2, \ldots$, and sequence n_k and m_k defined as above except without the assumption about m_k/n_k, let

$$W_{k:i} = \begin{cases} \sqrt{\frac{m_k}{n_k(m_k+n_k)}}(X_i-\mu) & \text{for } 1 \leq i \leq n_k \\ \sqrt{\frac{n_k}{m_k(m_k+n_k)}}(\nu - Y_{i-n_k}) & \text{for } n_k+1 \leq i \leq n_k+m_k \end{cases}$$

Then $W_{k:i} \sim \left(0, \frac{m_k n_k}{m_k+n_k}\left((\sigma^2/n_k^2)1_{[i\le n_k]} + (\tau^2/m_k^2)1_{[i>n_k]}\right)\right)$ independently. Thus we have (as desired)

$$S_k = \sum_{i=1}^{n_k+m_k} W_{k:i} = \sqrt{\frac{m_k n_k}{m_k+n_k}}\left[(\overline{X}_{n_k} - \overline{Y}_{m_k}) - (\mu - \nu)\right]$$
$$V_k = \sum_{i=1}^{n_k+m_k} Var(W_{k:i}) = \left(\frac{m_k n_k}{m_k+n_k}\right)\left(\frac{\sigma^2}{n_k} + \frac{\tau^2}{m_k}\right).$$

Now we have

$$Z_k^{(u)} \equiv \frac{S_k}{\sqrt{V_k}} \to_d \mathcal{N}(0,1)$$

if and only if the Lindeberg condition holds:

$$\forall \epsilon > 0 \qquad \lim_{k\to\infty} \frac{1}{V_k}\sum_{i=1}^{n_k+m_k} E\left[|W_i|^2 1_{[|W_i|\ge\epsilon\sqrt{V_k}]}\right] = 0.$$

Now for $1 \le i \le n_k$,

$$E\left[|W_i|^2 1_{[|W_i|\ge\epsilon\sqrt{V_k}]}\right] = \frac{m_k}{n_k(m_k+n_k)} E\left[(X_1-\mu)^2 1_{[(X_1-\mu)^2\ge\epsilon^2 V_k n_k(m_k+n_k)/m_k]}\right],$$

and

$$\begin{aligned} M_X &\equiv \epsilon^2 V_k n_k(m_k+n_k)/m_k \\ &= \epsilon^2\left(\frac{m_k n_k}{m_k+n_k}\right)\left(\frac{\sigma^2}{n_k}+\frac{\tau^2}{m_k}\right)\left(\frac{n_k(m_k+n_k)}{m_k}\right) \\ &= \epsilon^2\left(n_k\sigma^2 + \frac{n_k^2\tau^2}{m_k}\right) \\ &\to \infty \end{aligned}$$

as $k\to\infty$ (because $n_k\to\infty$). And because $\sigma^2 = E(X_1-\mu)^2 < \infty$, then as $M_X > \infty$

$$A_X \equiv E\left[(X_1-\mu)^2 1_{[(X_1-\mu)^2\ge M_X]}\right] \to 0.$$

Similar derivations for $n_k + 1 \le i \le n_k + m_k$ yield

$$E\left[|W_i|^2 1_{[|W_i|\ge\epsilon\sqrt{V_k}]}\right] = \frac{n_k}{m_k(m_k+n_k)} E\left[(Y_1-\nu)^2 1_{[(Y_1-\nu)^2\ge\epsilon^2 V_k m_k(m_k+n_k)/n_k]}\right],$$

and

$$\begin{aligned} M_Y &= \epsilon^2 V_k m_k(m_k+n_k)/n_k \\ &= \epsilon^2\left(\frac{m_k n_k}{m_k+n_k}\right)\left(\frac{\sigma^2}{n_k}+\frac{\tau^2}{m_k}\right)\left(\frac{m_k(m_k+n_k)}{n_k}\right) \\ &= \epsilon^2\left(\frac{m_k^2\sigma^2}{n_k} + m_k\tau^2\right) \\ &\to \infty \end{aligned}$$

as $k\to\infty$ (because $m_k\to\infty$). And because $\tau^2 = E(Y_1-\nu)^2 < \infty$, then as $M_Y > \infty$

$$A_Y \equiv E\left[(Y_1-\nu)^2 1_{[(Y_1-\nu)^2\ge M_Y]}\right] \to 0.$$

Now owing to the X_i's being i.i.d. and the Y_i's being i.i.d, and because then as $\min(n_k, m_k)\to\infty$, $V_k\to\infty$, $A_X\to 0$, $A_Y\to 0$ and

$$\begin{aligned} \frac{1}{V_k}\sum_{i=1}^{n_k+m_k} E\left[|W_i|^2 1_{[|W_i|\ge\epsilon\sqrt{V_k}]}\right] &= \frac{1}{V_k}\left[\frac{m_k}{m_k+n_k}A_X + \frac{n_k}{m_k+n_k}A_Y\right] \\ &\to 0 \end{aligned}$$

thus satisfying the Liindeberg condition.

Now, having shown that $Z_n^{(u)} \to_d \mathcal{N}(0,1)$, the same arguments used in part a for the consistency of the various estiimates of the σ^2 and τ^2 give the asymptotic distribution for $T_n^{(e)}$, $T_n^{(u)}$, and $T_n^{(x)}$.