Statistics 581, Solutions, Problem Set 3

Wellner; 10/18/00

1. Suppose that $Y \sim \chi^2_n(\delta)$. Compute $E(Y)$ and $Var(Y)$.
Hint: use the formulas $E(Y) = E\{E(Y|K)\}$ and

$$Var(Y) = Var(E(Y|K)) + E\{Var(Y|K)\}\,.$$

Solution: Now $(Y|K) \sim \chi^2_{2K+n}$ where $K \sim$ Poisson$(\delta/2)$, so

$$E(Y) = E\{E(Y|K)\} = E\{2K+n\} = n + 2(\delta/2) = n + \delta\,.$$

Furthermore,

$$\begin{aligned} Var(Y) &= E\{Var(Y|K)\} + Var\{E(Y|K)\} \\ &= E\{2(2K+n)\} + Var\{2K+n\} \\ &= 4(\delta/2) + 2n + 4(\delta/2) \\ &= 2n + 4\delta\,. \end{aligned}$$

2. Lehmann and Casella, TPE, problem 8.25, page 77. [Note problem 8.24 on the same page.]
Solution: (8.21).
8.21(a) If $R_n = o_p(1/k_n)$, then $R_n = O_p(1/k_n)$: Since $k_n R_n \rightarrow_p 0$ as $n \rightarrow \infty$, we have $P(|k_n R_n| > \epsilon) \rightarrow 0$ for every $\epsilon > 0$, and this implies $k_n R_n = O_p(1)$ or $R_n = O_p(1/k_n)$.
8.21(b) This one is a bit tricky to translate from the deterministic setting to the language of random variables. One possible translation is:
$R_n = O_p(1)$ if and only if for every $\epsilon > 0$ there exists an $M = M(\epsilon)$ such that $\sup_n P(|R_n| > M) \le \epsilon$. Proof: Clearly if for every $\epsilon > 0$ there is an M such that $\sup_n P(|R_n| > M) \le \epsilon$, then $R_n = O_p(1)$. It remains to prove the reverse implication. Fix $\epsilon > 0$. By the hypothesis that $R_n = O_p(1)$, there exist $M = M_\epsilon$ and $n_0 = n_0(\epsilon)$ such that $P(|R_n| > M) < \epsilon$ for $n > n_0$. But since $P(|R_n| < \infty) = 1$ for $n = 1, \ldots, n_0$, it follows that we can choose an M' so that

$$\max_{1 \le n \le n_0} P(|R_n| > M') < \epsilon\,.$$

But then it follows that

$$\sup_n P(|R_n| > \max\{M, M'\}) < \epsilon\,.$$

8.21(c) $R_n = o_p(1)$ if and only if $R_n \rightarrow_p 0$: This follows immediately from the definition.
8.21(d) If $R_n = O_p(1/k_n)$ and k_n'/k_n converges to a finite limit, then $R_n = O_p(1/k_n')$:

Since $R_n = O_p(1/k_n)$ means $k_n R_n = O_p(1)$, it follows that $k'_n R_n = (k'_n/k_n) k_n R_n = O(1) O_p(1) = O_p(1)$; i.e. $R_n = O_p(1/k'_n)$.
(8.22) (a) If R_n and R'_n are both $O_p(1/k_n)$, then so is $R_n + R'_n$: By the hypothesis for every $\epsilon > 0$ there are numbers M and M' and integers n_0 and n'_0 such that

$$P(|k_n R_n| > M) < \epsilon \qquad \text{for} \ \ n > n_0$$

and

$$P(|k_n R'_n| > M') < \epsilon \qquad \text{for} \ \ n > n'_0 \, .$$

But then, for $n > \max\{n_0, n'_0\}$,

$$\begin{aligned} P(|k_n(R_n + R'_n)| > M + M') &\le P(|k_n R'_n| > M') + P(|k_n R'_n| > M') \\ &\le \epsilon + \epsilon = 2\epsilon \, , \end{aligned}$$

which implies that $R_n + R'_n = O(1/k_n)$.
8.22(b) If R_n and R'_n are both $o_p(1/k_n)$, then so is $R_n + R'_n$: By the hypothesis $k_n R_n \to_p 0$ and $k_n R'_n \to_p 0$. Hence $k_n(R_n + R'_n) = k_n R_n + k_n R'_n \to_p 0 + 0 = 0$. Thus $R_n + R'_n = o(1/k_n)$.
8.23(a) If $k'_n/k_n \to \infty$, $R_n = O_p(1/k_n)$, and $R'_n = O_p(1/k'_n)$, then $R_n + R'_n = O_p(1/k_n)$: By hypothesis, $k_n R_n = O_p(1)$ and $k'_n R'_n = O_p(1)$. Hence

$$\begin{aligned} k_n(R_n + R'_n) &= k_n R_n + \frac{k_n}{k'_n} k'_n R'_n \\ &= O_p(1) + o(1) O_p(1) \\ &= O_p(1) + o_p(1) = O_p(1) \, . \end{aligned}$$

Hence $R_n + R'_n = O_p(1/k_n)$.
8.23(b) If $k'_n/k_n \to \infty$, $R_n = o_p(1/k_n)$, and $R'_n = o_p(1/k'_n)$, then $R_n + R'_n = o_p(1/k_n)$: By hypothesis, $k_n R_n = o_p(1)$ and $k'_n R'_n = o_p(1)$. Hence

$$\begin{aligned} k_n(R_n + R'_n) &= k_n R_n + \frac{k_n}{k'_n} k'_n R'_n \\ &= o_p(1) + o(1) o_p(1) \\ &= o_p(1) + o_p(1) = o_p(1) \, . \end{aligned}$$

Hence $R_n + R'_n = o_p(1/k_n)$.

3. Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with mean μ, variance σ^2, and $E|X|^4 < \infty$.
(a) Show that the sample variance $S_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2/(n-1)$ satisfies

$$\sqrt{n}(S_n^2 - \sigma^2)/\sqrt{2}\sigma^2 \to_d N(0, 1 + \gamma_2/2).$$

where $\mu_4 \equiv E(X - \mu)^4$ and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$ is called the *excess of kurtosis*.
(b) Suppose that you want to test $H : \sigma \le \sigma_0^2$ versus $K : \sigma^2 > \sigma_0^2$ for σ_0 a fixed number, and you base your test on normal theory, but in fact the X's are *not normal* with $\gamma_2 \ne 0$. What effect does this have on the level (or size or actual type one error) of the normal theory test?
Solution: (a) Since

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

where $Y_i \equiv (X_i - \mu)^2$ are i.i.d. $(E(Y_1), Var(Y_1)) = (\sigma^2, \mu_4 - \sigma^4)$, it follows from the classical CLT that

$$\begin{aligned}\sqrt{n}(\overline{S}_n^2 - \sigma^2) &= \sqrt{n}(\overline{Y}_n - \sigma^2) - \sqrt{n}(\overline{X}_n - \mu)(\overline{X}_n - \mu) \\ &\to_d N(0, \mu_4 - \sigma^4) - N(0, \sigma^2) \cdot 0 \\ &= N(0, \mu_4 - \sigma^4).\end{aligned}$$

Hence

$$\begin{aligned}\sqrt{n}(\overline{S}_n^2 - \sigma^2)/(\sqrt{2}\sigma^2) &\to_d N(0, (\mu_4/\sigma^4 - 1)/2) \\ &= N(0, (\mu_4/\sigma^4 - 3 + 2)/2) = N(0, 1 + \gamma_2/2).\end{aligned}$$

If instead of $\overline{S}_n^2$ we consider the more usual $S_n^2 = (n/(n-1))\overline{S}_n^2$, it is easily seen that

$$\sqrt{n}(S_n^2 - \overline{S}_n^2) = \sqrt{n}(\frac{n}{n-1} - 1)\overline{S}_n^2 = o(1)O_p(1) = o_p(1).$$

Thus we also have

$$\sqrt{n}(S_n^2 - \sigma^2)/(\sqrt{2}\sigma^2) \to_d N(0, 1 + \gamma_2/2).$$

(b) When the X_i's are normal, $\gamma_2 = 0$ and $(n-1)S_n^2/\sigma_0^2 \sim \chi_{n-1}^2$ when $\sigma = \sigma_0$ is true. Hence the size of the normal theory test when normal theory is true is

$$\begin{aligned}\alpha &= P_{\sigma_0}((n-1)S_n^2/\sigma_0^2 \geq \chi_{n-1,\alpha}^2) \\ &= P_{\sigma_0}(\sqrt{n}(S_n^2/\sigma_0^2 - 1)/\sqrt{2} \geq \sqrt{\frac{n}{2}}(\frac{\chi_{n-1,\alpha}^2}{n-1} - 1)).\end{aligned}$$

Since $\sqrt{n}(S_n^2/\sigma_0^2 - 1)/\sqrt{2} \to_d N(0,1)$ under normality, this forces

$$\sqrt{\frac{n}{2}}(\frac{\chi_{n-1,\alpha}^2}{n-1} - 1)) \to z_\alpha.$$

Thus when the X_i's *are not normal* we have

$$\begin{aligned}P_{\sigma_0}((n-1)S_n^2/\sigma_0^2 \geq \chi_{n-1,\alpha}^2) &= P_{\sigma_0}(\sqrt{n}(S_n^2/\sigma_0^2 - 1)/\sqrt{2} \geq \sqrt{\frac{n}{2}}(\frac{\chi_{n-1,\alpha}^2}{n-1} - 1)) \\ &\to P(N(0, 1 + \gamma_2/2) \geq z_\alpha) \\ &= P(Z \geq \frac{z_\alpha}{\sqrt{1 + \gamma_2/2}}) = 1 - \Phi(\frac{z_\alpha}{\sqrt{1 + \gamma_2/2}}).\end{aligned}$$

When $\gamma_2 = -2$, the asymptotic size is 0; when $\gamma_2 = 0$, the asymptotic size is α; when $\gamma_2 = \infty$, the asymptotic size is $1/2$. For $\gamma_2 \in [-2, 0)$ the asymptotic size is $< \alpha$ while for $\gamma_2 \in (0, \infty)$ the asymptotic size is $> \alpha$.

4. Suppose that $X_1, \ldots, X_n$ are independent $N(0,1)$ random variables, and let $Y_i = X_i^2$, for $i = 1, \ldots, n$. Thus $\sum_1^n Y_i \sim \chi_n^2$.
(a) Show that $\sqrt{n}(\overline{Y}_n - 1) \to_d N(0,$ "*something*"), and find "something".
(b) Show that for each $r > 0$, $\sqrt{n}(\overline{Y}_n^r - 1) \to_d N(0, V^2(r))$ and find $V^2(r)$ as a

function of r.
(c) Show that

$$\frac{\sqrt{n}(\overline{Y}_n^{1/3} - (1 - 2/(9n)))}{\sqrt{2/9}} \to_d N(0,1)\,.$$

Does this agree with your result in (b)?
(d) Make normal probablility plots to compare the approximations in (a) and (c). [The transformation in (c) is called the "Wilson-Hilferty" transformation of a χ^2 random variable.]

Solution: (a) Since the Y_i's are i.i.d. with $E(Y_i) = 1$ and $Var(Y_i) = E(X_i^4) - E(X_i^2)^2 = 3 - 1 = 2$, it follows from the CLT that

$$\sqrt{n}(\overline{Y}_n - 1) \to_d Z \sim N(0,2)\,.$$

(b) For $g(x) = x^r$ we have $g'(x) = rx^{r-1}$. Hence by the g'-theorem

$$\begin{aligned} \sqrt{n}(\overline{Y}_n^r - 1) &= \sqrt{n}(g(\overline{Y}_n) - g(1)) \\ &\to_d g'(1)Z = rN(0,2) = N(0, 2r^2)\,. \end{aligned}$$

Thus $V^2(r) = 2r^2$.
(c) When $r = 1/3$, we find from (b) that

$$\sqrt{n}(\overline{Y}_n^{1/3} - 1) \to_d (1/3)Z \sim N(0, 2/9)\,.$$

Hence it follows that

$$\begin{aligned} &\sqrt{n}(\overline{Y}_n^{1/3} - (1 - 2/(9n)) \\ &= \sqrt{n}(\overline{Y}_n^{1/3} - 1) + (2/9\sqrt{n}) \\ &\to_d N(0, 2/9) + 0 = N(0, 2/9)\,. \end{aligned}$$

Hence

$$\frac{\sqrt{n}(\overline{Y}_n^{1/3} - (1 - 2/(9n)))}{\sqrt{2/9}} \to_d N(0,1)$$

in complete agreement with (b). (The added term $(2/9n)$ gives a higher order approximation to the mean.)
(d) See the plots at the end of this solution set.

5. Suppose that $X_1, X_2, \ldots$ are i.i.d. positive random variables, and define $\overline{X}_n \equiv n^{-1}\sum_{i=1}^n X_i$, $H_n \equiv 1/(n^{-1}\sum_{i=1}^n (1/X_i))$, and $G_n \equiv \{\prod_{i=1}^n X_i\}^{1/n}$ to be the *arithmetic, harmonic, and geometric* means respectively. We know that $\overline{X}_n \to_{a.s.} E(X_1) = \mu$ if and only if $E|X_1| < \infty$.
(a) Use the SLLN together with appropriate additional hypotheses to show that $H_n \to_{a.s.} 1/\{E(1/X_1)\} \equiv h$, and $G_n \to_{a.s.} \exp(E\{\log X_1\}) \equiv g$.
(b) Use the multivariate CLT and the delta method to find the joint limiting distribution of $\sqrt{n}(\overline{X}_n - \mu, H_n - h, G_n - g)$. You will need to impose or assume additional moment conditions to be able to prove this. Specify these additional assumptions carefully.

Solution: (a) If $0 < E(1/X_1) < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n}(1/X_i) \rightarrow_{a.s.} E(1/X_1) > 0.$$

If $E|\log(X_1)| < \infty$, then

$$\log G_n = \frac{1}{n}\sum_{i=1}^{n}\log(X_i) \rightarrow_{a.s.} E\log X_1.$$

Thus by the continuous mapping theorem if both $E(1/X_1) < \infty$ and $E|\log X_1| < \infty$, it follows that

$$(H_n, G_n) \rightarrow_{a.s.} (1/E(1/X_1), \exp(E\log X_1)) \equiv (h, g).$$

(c) By the multivariate CLT, if $EX_1^2 < \infty$, $E(1/X_1)^2 < \infty$, and $E(\log X_1)^2 < \infty$, then

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ \overline{X^{-1}}_n - E(1/X_1) \\ \overline{\log X}_n - E\log X_1 \end{pmatrix} \rightarrow_d \underline{Z} \sim N_3(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} Var(X_1) & Cov(X_1, 1/X_1) & Cov(X_1, \log(X_1)) \\ Cov(X_1, 1/X_1) & Var(1/X_1) & Cov(1/X_1, \log X_1) \\ Cov(X_1, \log(X_1)) & Cov(1/X_1, \log X_1) & Var(\log(X_1)) \end{pmatrix}.$$

Hence by the delta method with $g(x, y, z) = (x, 1/y, \exp(z))$ so that $\nabla g(x, y, z) = \text{diag}(1, -y^{-2}, \exp(z))$ and $\nabla g(\mu, E(1/X_1), E(\log X_1)) = \text{diag}(1, -h^2, g)$, it follows that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ H_n - h \\ G_n - g \end{pmatrix} \rightarrow_d \nabla g \cdot \underline{Z} \sim N_3(0, \nabla g \Sigma \nabla g^T).$$

Plots for Problem 4, part (d)

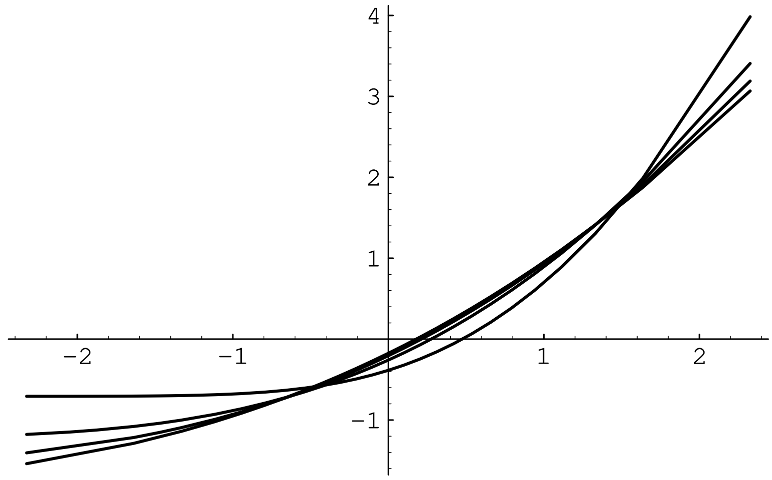

Figure 1: Basic CLT, $n = 3, 5, 7, 9$.

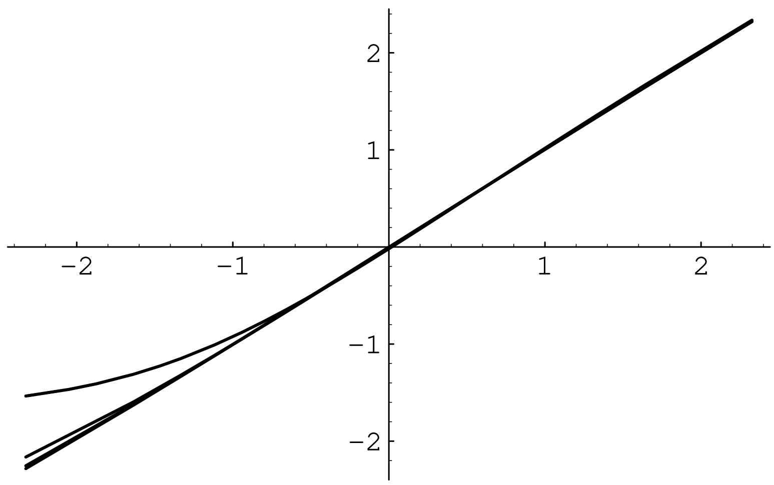

Figure 2: Wilson-Hilferty transform of chi-square, $n = 1, 3, 5, 7$.

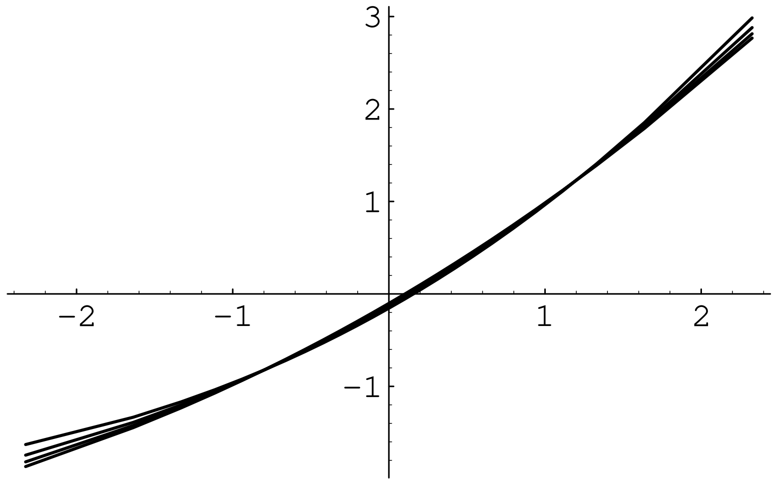

Figure 3: Basic CLT, $n = 9, 13, 17, 21$.

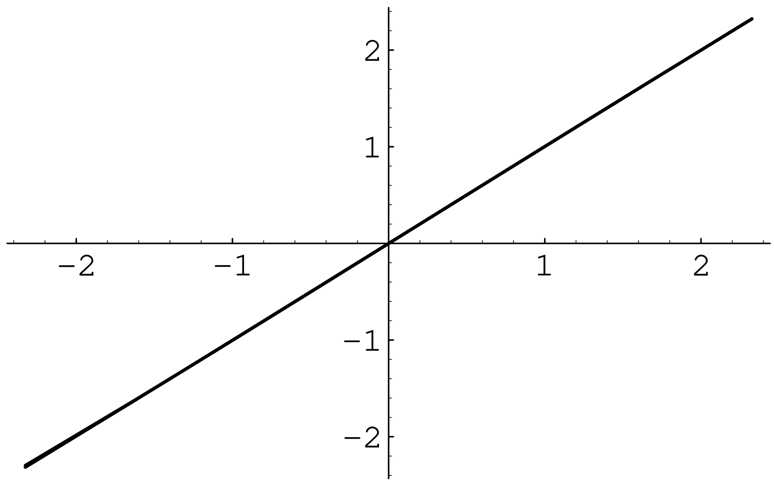

Figure 4: Wilson-Hilferty transform of chi-square,$n = 9, 13, 17, 21$.

Mathematica Code for Figures 1 and 2:

```
<<Statistics`ContinuousDistributions`
dist[j_] := ChiSquareDistribution[j]
gdist := NormalDistribution[0,1]
Qn[u_,j_] := Quantile[dist[j], u]
Tn[u_,j_] := Sqrt[j/2]*(Qn[u,j]/j - 1)
Sn[u_,j_] := Sqrt[9*j/2]*((Qn[u,j]/j)^(1/3) - (1 - 2/(9*j)))
QN[u_] := Quantile[gdist, u]
ParametricPlot[{{QN[u],Tn[u,1]}, {QN[u],Tn[u,3]},
            {QN[u],Tn[u,5]}, {QN[u],Tn[u,7]}}, {u,0.01,.99}]
ParametricPlot[{{QN[u],Sn[u,1]}, {QN[u],Sn[u,3]},
            {QN[u],Sn[u,5]}, {QN[u],Sn[u,7]}}, {u,0.01,.99}]
```