Statistics 581, Problem Set 3 Solutions

Wellner; 10/19/2005

1. A. Ferguson, ACILST, page 24, problem 4. One strategy for evaluating the integral

$$I = \int_1^\infty \frac{1}{x} \sin(2\pi x) dx = .152645...$$

by Monte Carlo approximation is as follows. Write the integral, by a change of variable $y = 1/x$, as

$$I = \int_0^1 \frac{1}{y} \sin\left(\frac{2\pi}{y}\right) dy$$

and approximate I by

$$\hat{I}_{1n} = \frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i} \sin\left(\frac{2\pi}{Y_i}\right)$$

where $Y_1, \ldots, Y_n$ is a sample from the uniform distribution on $[0, 1]$. How well does this approximation work? Does $\hat{I}_{1n}$ converge to I almost surely?

B. An alternative way to rewrite the integral is to first integrate by parts: thus

$$\begin{aligned} I &= \int_1^\infty x^{-1}\sin(2\pi x)dx = \frac{-\cos(2\pi x)}{x}\Big|_1^\infty - \int_1^\infty \frac{1}{2\pi} x^{-2}\cos(2\pi x)dx \\ &= \frac{1}{2\pi}\left(1 - \int_1^\infty x^{-2}\cos(2\pi x)dx\right) \\ &= \frac{1}{2\pi}\left(1 - \int_0^1 \cos(2\pi/y)dy\right). \end{aligned}$$

Propose an alternative Monte Carlo approximation $\hat{I}_{n2}$ to the integral I based on this re-expression of the integral. Does $\hat{I}_{2n}$ converge to I almost surely?

C. Does $\sqrt{n}(\hat{I}_{2n} - I) \rightarrow_d$ "something"? If so, find "something".

Solution: A. By the chance of variable $x = 1/y$ it follows that

$$I = \int_1^\infty \frac{1}{x}\sin(2\pi x)dx = \int_0^1 \frac{1}{y}\sin(2\pi/y)dy\,.$$

Thus a natural estimator is

$$\hat{I}_{1n} = \frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i} \sin\left(\frac{2\pi}{Y_i}\right)$$

where $Y_1, \ldots, Y_n$ is a sample from the uniform distribution on $[0, 1]$. Unfortunately, however,

$$E|\frac{1}{Y_1}\sin(2\pi/Y_1)| \;=\; E\left\{\frac{1}{Y_1}|\sin(2\pi/Y_1)|\right\} = \int_0^1 \frac{1}{y}|\sin(2\pi/y)|dy$$

$$
\begin{aligned}
&= \int_1^\infty \frac{1}{x}|\sin(2\pi x)|dx = \sum_{k=3}^\infty \int_{(k-1)/2}^{k/2} \frac{1}{x}|\sin(2\pi x)|dx \\
&\geq \sum_{k=3}^\infty \frac{1}{k/2}\int_{(k-1)/2}^{k/2} |\sin(2\pi x)|dx \\
&= \sum_{k=3}^\infty \frac{2}{k}\frac{1}{\pi} = \frac{2}{\pi}\sum_{k=3}^\infty \frac{1}{k} = \infty\,.
\end{aligned}
$$

Hence the strong law of large numbers fails, and it is not at all clear that $\hat{I}_n$ will be a reasonable estimator of I. Here are plots of the functions being integrated on $[1,\infty)$ and on $(0,1]$:

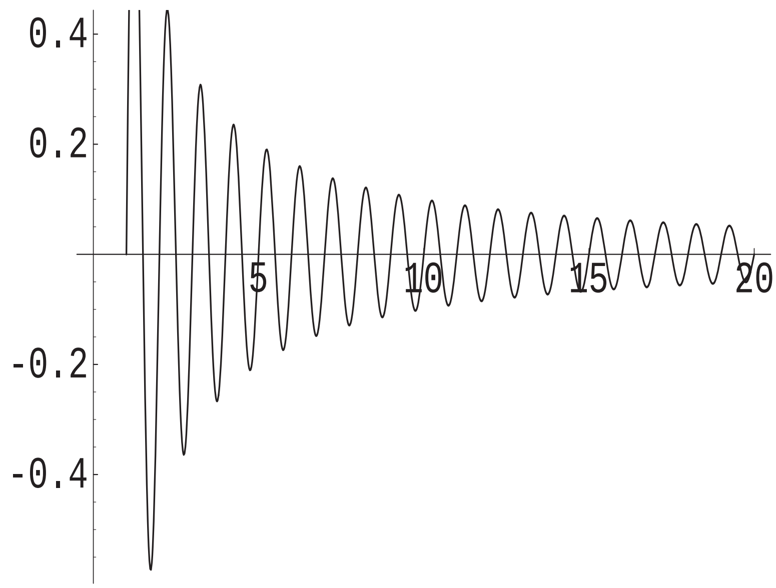

Figure 1: Plot of $x^{-1}\sin(2\pi x)$ on $[1,\infty)$

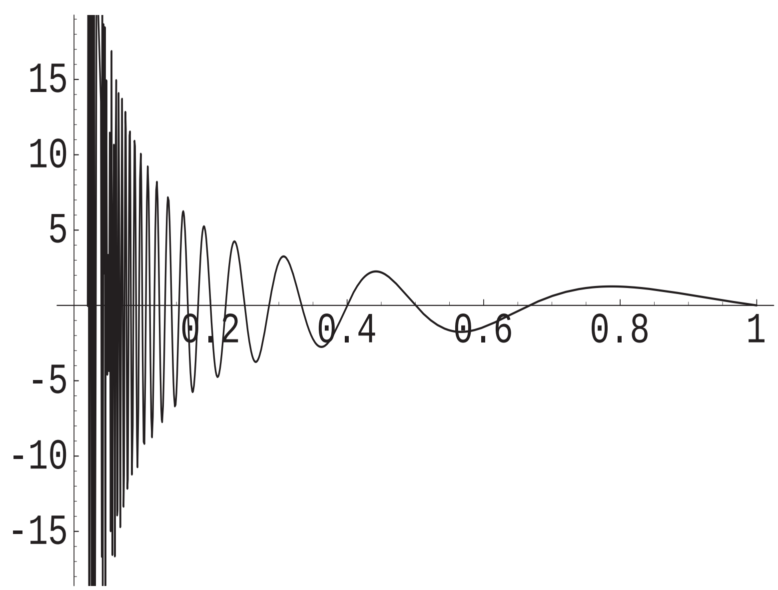

Figure 2: Plot of $y^{-1}\sin(2\pi/y)$ on $(0,1]$

B. With the integral I rewritten as

$$I = \frac{1}{2\pi}\left(1 - \int_0^1 \cos(2\pi/y)dy\right),$$

an alternative Monte Carlo approximation $\hat{I}_{2n}$ would be

$$\hat{I}_{2n} = \frac{1}{2\pi}\left(1 - \frac{1}{n}\sum_{i=1}^{n}\cos(2\pi/Y_i)\right)$$

where $Y_1, \ldots, Y_n$ are i.i.d. Uniform$(0,1)$. In this case we have

$$E|\cos(2\pi/Y_1)| \le 1 < \infty$$

since $|\cos(x)| \le 1$ for all $x > 0$. Thus it follows from the strong law of large numbers and the Mann-Wald theorem that $\hat{I}_{2n} \to_{a.s.} I$.

C. Now

$$\sqrt{n}(\hat{I}_{2n} - I) = -\sqrt{n}(\overline{C}_n - E(C_1))$$

where $C_i \equiv \cos(2\pi/Y_i)/(2\pi)$, $i = 1, \ldots, n$ are i.i.d. with $E(C_1^2) < \infty$ (since $|\cos(x)| \le 1$ for all $x > 0$ as in B). Therefore the (ordinary or Lindeberg) CLT yields

$$\sqrt{n}(\hat{I}_{2n} - I) \to_d N(0, Var(C_1))$$

where

$$Var(C_1) = \frac{1}{(2\pi)^2}\left\{\int_0^1 [\cos(2\pi/y)]^2 dy - \left(\int_0^1 \cos(2\pi/y)dy\right)^2\right\} = 0.01274....$$

2. Ferguson, ACILST, page 34, problem 1 (modified slightly).
Suppose that $X_1, \ldots, X_n$ is a sample from the Poisson distribution with parameter $\lambda > 0$: $P(X_1 = k) = \exp(-\lambda)\lambda^k/k!$, $k = 0, 1, \ldots$. Let $Z_n = (1/n)\sum_{i=1}^n 1_{[X_i=0]}$.
(a) What is the joint asymptotic distribution of

$$\sqrt{n}((\overline{X}_n, Z_n)' - (\lambda, e^{-\lambda})')?$$

(b) Let $p_0(\lambda) \equiv P_\lambda(X_1 = 0)$. What is the asymptotic distribution of $\hat{p}_0 \equiv p_0(\hat{\lambda}_n)$ where $\hat{\lambda}_n = \overline{X}_n$?
(c) What is the joint asymptotic distribution of $(Z_n, \hat{p}_0)$ (after centering and rescaling)?

(d) Compute the ratio of the asymptotic variances of the two estimators Z_n and $\hat{p}_0$ of $p_0(\lambda)$. Which estimator would you prefer if the Poisson model (assumption) holds? Which estimator would you prefer if the Poisson model (assumption) fails?

Solution: (a). Let $W_i \equiv (X_i, Y_i) \equiv (X_i, 1_{[X_i=1]})$. Then the W_i's are i.i.d. with mean $E(W_1) = (\lambda, e^{-\lambda})'$ and covariance matrix

$$\Sigma = \begin{pmatrix} \lambda & -\lambda e^{-\lambda} \\ -\lambda e^{-\lambda} & e^{-\lambda}(1 - e^{-\lambda}) \end{pmatrix}. \qquad (1)$$

Hence the multivariate CLT implies that

$$\sqrt{n}(\overline{W} - E(W_1)) = \sqrt{n}((\overline{X}_n, Z_n)' - (\lambda, e^{-\lambda})) \to_d T \sim N_2(0, \Sigma) \tag{2}$$

where Σ is given in (1).

(b). Now $\hat{p}_0 = g(\overline{X}_n)$ where $g(v) = e^{-v}$. Hence $g'(v) = -e^{-v}$, $g'(\lambda) = -e^{-\lambda}$, and $\sqrt{n}(\overline{X}_n - \lambda) \to_d N(0, \lambda)$ by the CLT (or the first component of the convergence in distribution in part (a)). Hence it follows from the delta-method that

$$\sqrt{n}(\hat{p}_0 - p_0(\lambda)) = \sqrt{n}(g(\overline{X}_n) - g(\lambda)) \to_d g'(\lambda)N(0, \lambda) = N(0, \lambda e^{-2\lambda}) .$$

(c). At this point it is a bit easier to study $(\hat{p}_0, Z_n) = g(\overline{X}_n, Z_n)$ where $g(u, v) \equiv (e^{-u}, v)$. Then in view of (2) and

$$\nabla g(\lambda, e^{-\lambda}) = \begin{pmatrix} -e^{-\lambda} & 0 \\ 0 & 1 \end{pmatrix} ,$$

it follows from the delta-method that

$$\sqrt{n}((\hat{p}_0, Z_n)' - e^{-\lambda}(1, 1)') \to_d \nabla g(\lambda, e^{-\lambda})T \sim N_2(0, \nabla g \Sigma (\nabla g)')$$

where

$$\nabla g \Sigma (\nabla g)' = \begin{pmatrix} \lambda e^{-2\lambda} & \lambda e^{-2\lambda} \\ \lambda e^{-2\lambda} & e^{-\lambda}(1 - e^{-\lambda}) \end{pmatrix} .$$

This is a situation in which we have two estimators of $P_\lambda(X_1 = 0) = p_0(\lambda)$, namely the MLE $\hat{p}_1 = p_0(\hat{\lambda})$ and the empirical (or "plug-in" estimator $Z_n = \#\{i \le n : X_i = 0\}/n$. Note that the ratio of the asymptotic variance of $\hat{p}_0$ to the asymptotic variance of Z_n is

$$ARE(\hat{p}_0, Z_n) \equiv \frac{\lambda e^{-2\lambda}}{e^{-\lambda}(1 - e^{-\lambda})} = \frac{\lambda e^{-\lambda}}{(1 - e^{-\lambda})} < 1$$

for all $\lambda > 0$. See the figure below

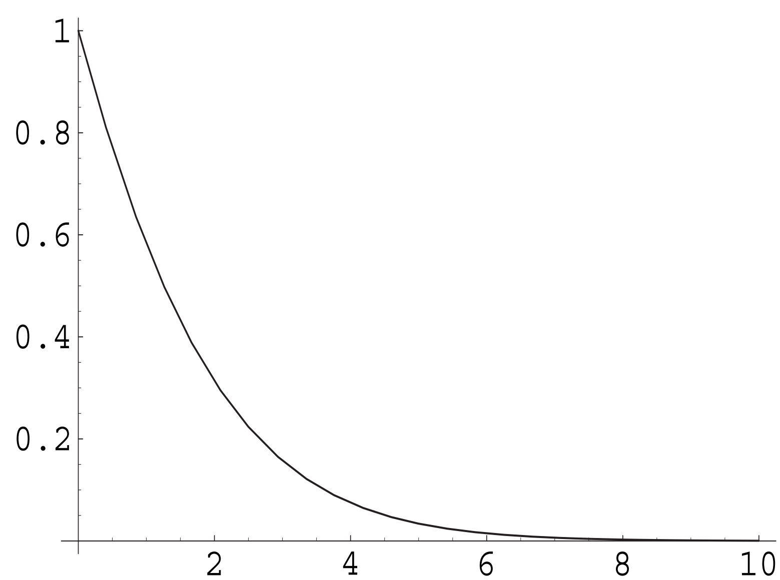

Figure 3: ARE of MLE relative to Plug-In

3. Suppose that $X_1, X_2, \ldots$ are i.i.d. (μ, σ^2) with $\mu_4 < \infty$. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ be the sample mean and sample variance respectively.
(a) Show that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{pmatrix} \to_d \underline{Z} \sim N_2(0, \Sigma)$$

where

$$\begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{pmatrix}.$$

(b) Suppose $\sigma > 0$. Use (a) to find the limiting distribution of the sample *signal-noise ratio* $D_n \equiv \overline{X}_n / S_n$; i.e. show that $\sqrt{n}(D_n - d) \to_d N(0, V^2)$ with $d \equiv \mu/\sigma$ and find V^2.
Solution: (a) Since $S_n^2 = n^{-1}\sum_{i=1}^n (X_i - \mu)^2 + o_p(1/\sqrt{n})$, we have

$$\begin{aligned}
\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{pmatrix} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} X_i - \mu \\ (X_i - \mu)^2 - \sigma^2 \end{pmatrix} + o_p(1) \\
&\to_d \underline{Z} \sim N_2(0, \Sigma)
\end{aligned}$$

by the multivariate CLT where Σ is as given above.
(b) The function $g(u, v) = u/\sqrt{v}$ is differentiable at points (u, v) with $v \neq 0$, and the derivative is $\nabla g(u, v) = (1/\sqrt{v}, u(-1/2)v^{-3/2})$ so that $\nabla g(\mu, \sigma^2) = (1/\sigma, (-1/2)\mu\sigma^{-3}) = (1/\sigma)(1, -(1/2)\mu/\sigma^2))$. Hence it follows from the delta method (g' theorem) that

$$\begin{aligned}
\sqrt{n}(D_n - d) &= \sqrt{n}(g(\overline{X}_n, S_n^2) - g(\mu, \sigma^2)) \\
&\to_d \nabla g \cdot \underline{Z} \sim N(0, \nabla g^T \Sigma \nabla g)
\end{aligned}$$

and it is easy to calculate that

$$\begin{aligned}
\nabla g^T \Sigma \nabla g &= \frac{1}{\sigma^4}\left\{\sigma^4 - \mu\mu_3 + \frac{1}{4}d^2(\mu_4 - \sigma^4)\right\} \\
&= 1 - d\gamma_1 + \frac{1}{4}d^2(2 + \gamma_2)
\end{aligned}$$

where $\gamma_1 \equiv \mu_3/\sigma^3$ and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$. Note that when the X_i's are normal (so $\gamma_1 = \gamma_2 = 0$), this reduces to $1 + d^2/2$. Thus under normality we have

$$\sqrt{n}(g(D_n) - g(d)) \to_d N(0, 1)$$

if $g(x) \equiv \sqrt{2}\text{arcsinh}(x/\sqrt{2})$.

4. Suppose that X is a random variable with finite fourth moment; $E|X|^4 < \infty$. Then $\mu_4 = E(X - \mu)^4$ is the fourth central moment of X. The ratio $\mu_4/\sigma^4 \equiv \kappa$ is the *kurtosis* of X (or of the distribution function F of X), and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$ is called the *excess of kurtosis*; note that for any $N(\mu, \sigma^2)$ random variable, $\gamma_2 = 0$. Investigate the value of γ_2 for various classical distributions (t_r, uniform, bernoulli, Poission(λ), ...). How big can γ_2 be? How small can γ_2 be?

Solution: Note that $\mu_4^{1/4} = \{E(X-\mu)^4\}^{1/4} \ge \{E(X-\mu)^2\}^{1/2} = \sigma$ by Liapunov's inequality. Thus $\mu_4/\sigma^4 \ge 1$ always, or $\gamma_2 \equiv \mu_4/\sigma^4 - 3 \ge -2$ with equality if $X = \pm 1$ with probability 1/2 each: then $\mu = 0$, $\sigma^2 = 1$, $\mu_4 = 1$, and $\gamma_2 = -2$.
For $X \sim N(0,1)$, $\gamma_2 = 0$ since $EX^4 = 3$.
For $X \sim t_r$, $r > 4$, $\gamma_2 = 6/(r-4) \nearrow \infty$ as $r \searrow 4$; $\gamma_2 \searrow 0$ as $r \nearrow \infty$.
For $X \sim \text{Gamma}(\alpha, \beta)$, $\gamma_2 = 6/\alpha \nearrow \infty$ as $\alpha \searrow 0$.
For $X \sim \text{Poisson}(\lambda)$, $\gamma_2 = 1/\lambda \nearrow \infty$ as $\lambda \searrow 0$.
For $X \sim \text{Bernoulli}(p)$, $\gamma_2 = (1-p)^2/p + p^2/(1-p) - 3$ which $= -2$ when $p = 1/2$, and $\nearrow \infty$ when $p \to 0, 1$.

5. 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 $\sqrt{n}(\overline{Y}_n - 1) \to_d N(0, \text{"}something\text{"})$, 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) Here are some plots showing the effect of taking the 1/3 power.

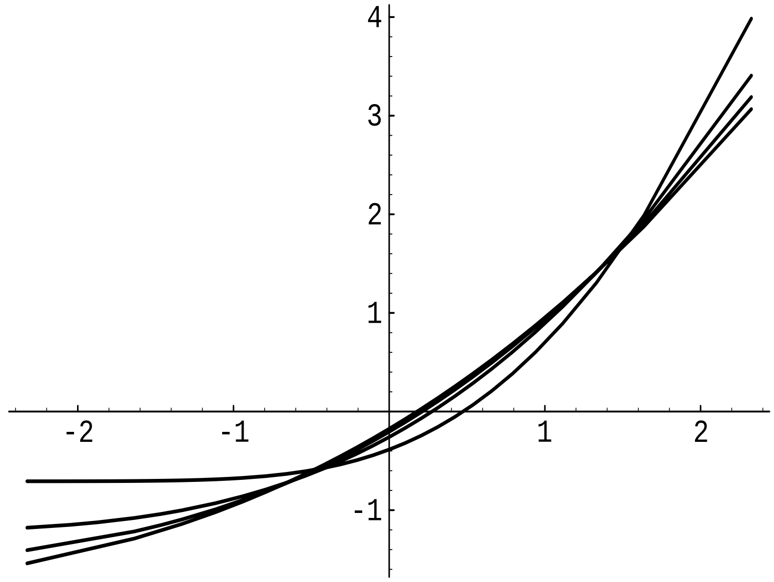

Figure 4: Q-Q plot for $\sqrt{n/2}(\overline{Y}_n - 1)$, $n = 1, 3, 5, 7$

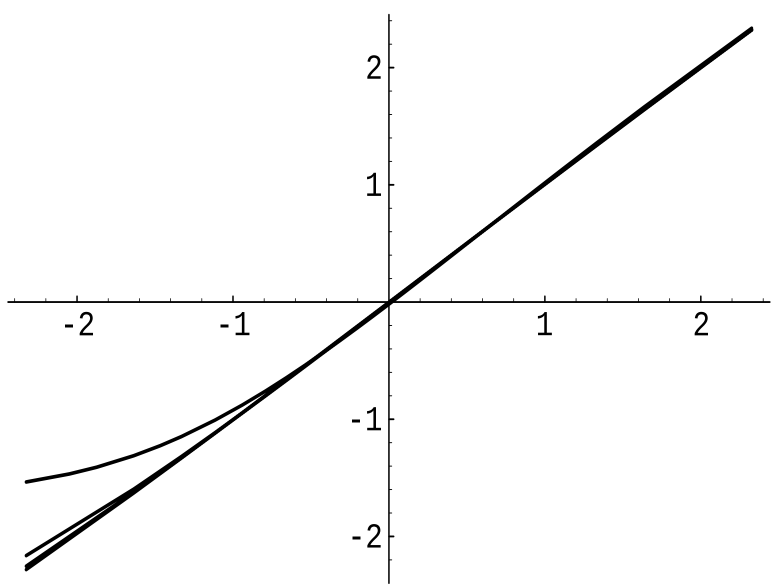

Figure 5: Q-Q plot for $\sqrt{9n/2}((\overline{Y}_n)^{1/3} - (1 - 2/(9n)))$, $n = 1, 3, 5, 7$

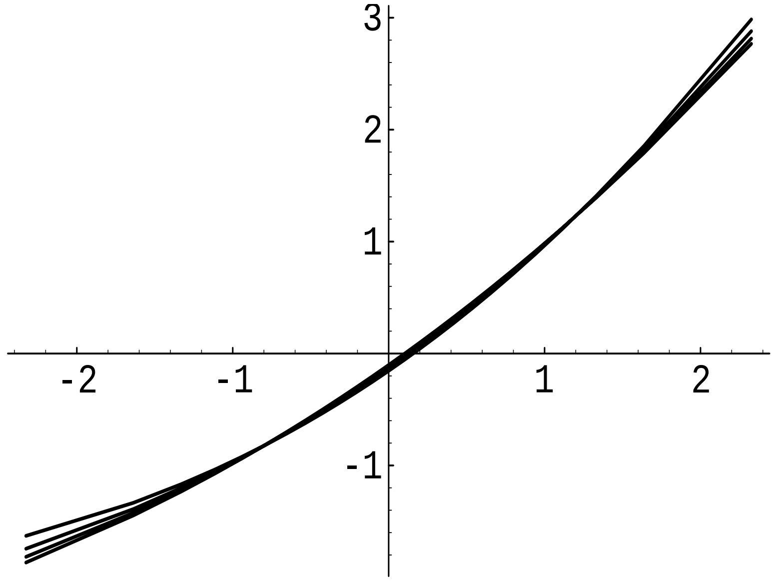

Figure 6: Q-Q plot for $\sqrt{n/2}(\overline{Y}_n - 1)$, $n = 9, 13, 17, 21$

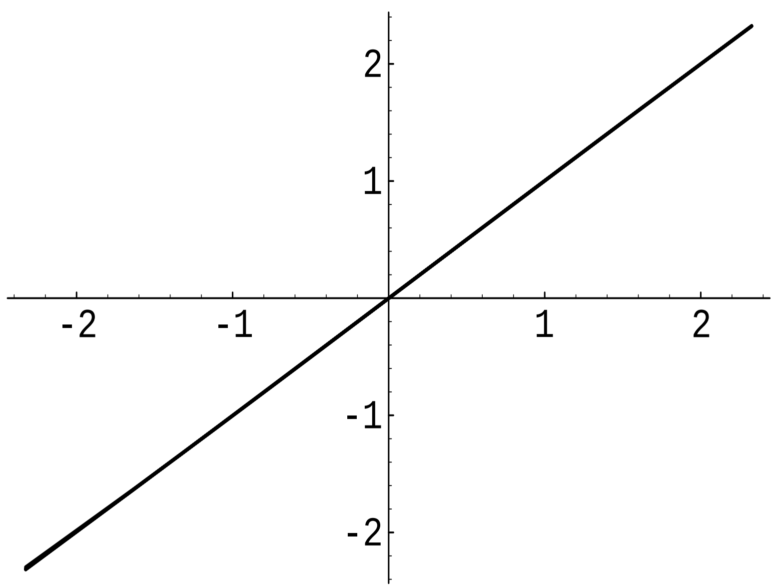

Figure 7: Q-Q plot for $\sqrt{9n/2}((\overline{Y}_n)^{1/3} - (1 - 2/9n))$, $n = 9, 13, 17, 21$

6.

7. Ferguson, ACILST, problem 1, page 65, modified.

   In a multinomial experiment with sample size $n = 100$ and 3 cells with null hypothesis $H_0 : \underline{p}_0 = (1/3, 1/3, 1/3)$, what is the approximate power at the alternative $\underline{p} = (.25, .5, .25)$ when the level of significance is $\alpha = .05$? $\alpha = .01$? How large a sample size is need to achieve power 0.8 at this alternative when $\alpha = .05$? $\alpha = .01$?

(a) In a multinomial experiment with sample size 100 and 3 cells with null hypothesis $H_0 : p_1 = 1/4, p_2 = 1/2, p_3 = 1/4$, what is the approximate power at the alternative $p_1 = 0.2, p_2 = 0.6, p_3 = 0.2$ when the level of significance is $\alpha = 0.05$? $\alpha = 0.01$? (b) How large a sample size is needed to achieve power 0.9 at this alternative when $\alpha = 0.05$? $\alpha = 0.01$?

Solution: Now

$$n^{1/2}(\underline{p}-\underline{p}_0) = 10((1/4, 1/2, 1/4)-(1/3, 1/3, 1/3)) = 10(-1/12, 1/6, -1/12) = (-5/6, 5/3, -5/6)\,,$$

so the non-centrality parameter is

$$\delta = \frac{(5/6)^2}{1/3} + \frac{(5/3)^2}{1/3} + \frac{(5/6)^2}{1/3} = 3 \cdot 25 \left\{ \frac{1}{36} + \frac{1}{9} + \frac{1}{36} \right\} = \frac{25}{2} = 12.5\,.$$

Thus the approximate power via $\chi_2^2(\delta)$ is

$$P(\chi_2^2(12.5) \geq \chi_{2,.05}) = P(\chi_2^2(12.5) \geq 5.991) = .896\,, \qquad \text{when} \quad \alpha = .05\,,$$

and

$$P(\chi_2^2(12.5) \geq \chi_{2,.01}) = P(\chi_2^2(12.5) \geq 9.210) = .744 \qquad \text{when} \quad \alpha = .01\,,$$

(b) Now we want to find n so that

$$P(\chi_2^2(\delta_n) \geq 5.991) = .90$$

where

$$\delta_n = n \left\{ \frac{(1/12)^2}{1/3} + \frac{(1/6)^2}{1/3} + \frac{(1/12)62}{1/3} \right) = n/8\,.$$

In this case we find that $\delta_n = n/8 = 12.6539$, so that $n = 8 \cdot 12.6539 \approx 102$. When $\alpha = .01$ we find that $\delta_n = n/8 = 17.4267$ so that $n = 8 \cdot 17.4267/(.04) \approx 140$.

The alternative approximation to power that we derived in class is

$$\begin{aligned} P_p(Q_n \geq \chi^2_{k-1,\alpha}) &= P_p\left(\sqrt{n}(n^{-1}Q_n - q) \geq \sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - q)\right) \\ &\doteq P\left(N(0, d^T Ad) \geq \sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - q)\right) \\ &= 1 - \Phi(\sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - q)/\sqrt{d^T Ad}) \end{aligned}$$

where $d \equiv 2\text{diag}(1/p_0)(p - p_0)$, $A = \text{diag}(p) - pp^T$, and $q = \sum_{j=1}^k (p_j - p_{j0})^2/p_{j0}$. In the present case I calculate $q = 1/8$, $d = (-1/2, 1, -1/2)^T = (-1, 2, -1)/2$, and

$$A = \text{diag}(p) - pp^T = \frac{1}{16} \begin{pmatrix} 3 & -2 & -1 \\ -2 & 4 & -2 \\ -1 & -2 & 3 \end{pmatrix}$$

so that $d^T Ad = (3/4)^2$. Thus the approximation becomes

$$P_p(Q_n \geq \chi^2_{2,\alpha}) \doteq 1 - \Phi(\sqrt{n}(n^{-1}\chi^2_{2,\alpha} - 1/8)/(3/4)).$$

When I calculate I get

$$\begin{aligned} P_p(Q_n \geq \chi^2_{2,.05}) &\doteq 1 - \Phi(\sqrt{n}(n^{-1}\chi^2_{2,.05} - 1/8)/(3/4)) = 0.807249 \\ P_p(Q_n \geq \chi^2_{2,.05}) &\doteq 1 - \Phi(\sqrt{n}(n^{-1}\chi^2_{2,.01} - 1/8)/(3/4)) = 0.669532, \end{aligned}$$

which are both somewhat lower than suggested by the non-central chi-square approximation. A Monte-Carlo study not shown here shows that the non-central chi-square approximation is quite accurate in this case. I suspect that the fixed alternative limit theorem and resulting normal approximation to power will do better for more extreme alternatives with a larger number of cells, but I have not carried out a thorough study.