Carson Slater
Dr. Harvill
STA 5351

Homework 8

1) $X_1, \ldots, X_n$ is iid from $f_X(x|\theta)$ such that

$$f_X(x|\theta) = \tfrac{1}{2}(1+\theta x), \quad -1 < x < 1, \quad -1 < \theta < 1.$$

Finding a consistent estimator for θ, we have if we consider the method of moments estimator for θ, $\hat{\theta}$.

So then setting

$$\mu = E[X] = \int_{-1}^{1} \left(\frac{1}{2} + \frac{\theta x}{2}\right) dx = \bar{X}$$

$$\Rightarrow \mu = \left(\frac{1}{2}x + \frac{\theta x^2}{4}\right)\Bigg|_{-1}^{1} = \frac{\theta}{3} = \bar{X}$$

$$\Rightarrow \hat{\theta} = 3\bar{X}.$$

Using the MOM estimator, we need to show

$$\lim_{n\to\infty} \text{Bias}[\hat{\theta}] = 0 \quad \text{and} \quad \lim_{n\to\infty} \text{Var}[\hat{\theta}] = 0$$

So then

$$\lim_{n\to\infty} \text{Bias}[\hat{\theta}] = \lim_{n\to\infty} E[\hat{\theta}_n] - \theta = \lim_{n\to\infty} 3E[\bar{X}_n] - \theta = \lim_{n\to\infty} 3\frac{\theta}{3} - \theta = 0$$

$$\lim_{n\to\infty} \text{Var}[\hat{\theta}] = \lim_{n\to\infty} 9\,\text{Var}[\bar{X}] = \lim_{n\to\infty} 9\frac{\sigma^2}{n} = 0 \quad (1)$$

We need to show $\sigma^2 < \infty$ for (1) to hold.

$$\sigma^2 = E\left[(X-\mu)^2\right] = \int_{-1}^{1} (x-\mu)^2 \tfrac{1}{2}(1+\theta x)\,dx = \tfrac{1}{2}\int_{-1}^{1} (x-\mu)^2 dx + \tfrac{1}{2}\int_{-1}^{1} (x-\mu)^2 \theta x\,dx$$

$$= \frac{1}{2}\left(\int_{-1}^{1} x^2 dx - \mu \int_{-1}^{1} 2x\,dx + \mu^2 x + \theta \int_{-1}^{1} x^3 dx - \theta\mu \int_{-1}^{1} x^2 dx + \mu^2\theta \int_{-1}^{1} x\,dx \right)$$

$$= \frac{3-\theta^2}{9} < \infty .$$

So $\hat{\theta}$ is a consistent estimator for θ.

2) Given $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$.

a) To find the UMVUE of $P(X=0)$, we have that $P(X=0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda}$. So $\tau(\lambda) = e^{-\lambda}$. Define the estimator for $e^{-\lambda}$ $W(X_1) = \begin{cases} 0 & \text{if } X_1 \neq 0 \\ 1 & \text{if } X_1 = 0 \end{cases}$. The estimator is unbiased, as $E[W(X_1)] = \sum_{x=0}^{\infty} W(X_1) P(X=x) = P(X=0) = e^{-\lambda}$.

So then also, we need to find a sufficient statistic for λ. By the factorization theorem, we have that

$f(\underline{x}|\lambda) = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!} I_{\{0,1,\ldots\}}(x_1) = g(\underline{x}|\theta) h(\underline{x})$, where $g(\underline{x}|\theta) = \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}$

and $h(\underline{x}) = \frac{I_{\{0,1,\ldots\}}(x_1)}{\prod_{i=1}^n x_i!}$. So $T(\underline{x}) = \sum_{i=1}^n X_i$ is sufficient for λ. Then assume there exists a function $g(T)$ such that $E_\lambda[g(T)] = 0 \;\; \forall \; \lambda > 0$. Then

$$E_{n\lambda}[g(T)] = \sum_{T=1}^{\infty} g(T) \frac{(n\lambda)^T e^{-n\lambda}}{T!}$$

$$= e^{-n\lambda} \sum_{T=0}^{\infty} \frac{g(T)(n\lambda)^T}{T!},$$

which is only 0 iff $g(T) = 0 \; \forall \; T$. So then if $E_{n\lambda}[g(T)] = 0$, then $P_{n\lambda}(g(T)=0) = 1 \;\; \forall \; n\lambda$, and $\sum_{i=1}^n X_i$ is a complete sufficient statistic.

By the Rao-Blackwell Theorem, the UMVUE is

$$\phi(W) = E[W \mid T = t]$$
$$= P(X_1 = 0 \mid T = t)$$
$$= \frac{P(X_1 = 0, \sum_{i=1}^{n} X_i = t)}{P(T = t)}$$
$$= \frac{P(X_1 = 0) P(\sum_{i=2}^{n} X_i = t)}{P(T = t)}$$
$$= \frac{e^{-\lambda}\left[((n-1)\lambda)^t e^{-(n-1)\lambda} / t!\right]}{n\lambda^t e^{-n\lambda} / t!}$$
$$= \left(\frac{n-1}{n}\right)^t$$
$$= \left(\frac{n-1}{n}\right)^{n\bar{x}}.$$

To find the MLE of $P(X=0) = e^{-\lambda}$. We can find the MLE of λ, $\hat{\lambda}$ and apply the invariance property of the MLE.

So then $\hat{\lambda}$ is

$$L(\lambda \mid \underline{x}) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} I_{\{0,1,\ldots\}}(x) \Rightarrow \mathcal{L}(\lambda \mid \underline{x}) = \sum_{i=1}^{n} x_i \log(\lambda) - n\lambda - \log\left(\prod_{i=1}^{n} x_i\right).$$

$$\frac{\partial}{\partial \lambda} \mathcal{L}(\lambda \mid \underline{x}) = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n \overset{\text{set}}{=} 0$$

$$\Rightarrow \hat{\lambda} = \bar{x}.$$

$$\frac{\partial^2}{\partial^2 \lambda} \mathcal{L}(\lambda \mid \bar{x}) = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0 \;\; \forall \lambda, \Rightarrow \hat{\lambda} \text{ is a}$$

maximum.

So the MLE of $P(X=0)$ is $e^{-\bar{x}}$ by the invariance property of the MLE.

b) To find the UMVUE statistic for $P(X=1)$, we define $\tau(\theta) = \frac{\lambda^1 e^{-\lambda}}{1!} = \lambda e^{-\lambda}$. We define a statstic $W(X_1) = \begin{cases} 0 & \text{if } X_1 \neq 1 \\ 1 & \text{if } X_1 = 1 \end{cases}$, which is unbiased, as $E[W] = \sum_{i=1}^{n} W \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = (1)P(X=1) = \lambda e^{-\lambda}$.

So $\sum_{i=1}^{n} X_i$ is a complete sufficient statistic for λ. By the Rao-Blackwell Theorem, we have that

$$
\begin{aligned}
\phi(W) &= E[W \mid T=t] \\
&= \frac{P(X_1 = 1, \sum_{i=1}^{n} X_i = t)}{P(T=t)} \\
&= \frac{P(X_1=1) P(\sum_{i=2}^{n} X_i = t-1)}{P(T=t)} \\
&= \frac{\lambda e^{-\lambda} \left((n-1)\lambda\right)^{t-1} e^{-n\lambda} / (t-1)!}{(n\lambda)^t e^{-n\lambda} / t!} \\
&= \frac{t!}{(t-1)!} \frac{1}{n} \left(\frac{n-1}{n}\right)^{t-1} \\
&= \frac{t}{n} \left(\frac{n-1}{n}\right)^{nt-1} \\
&= \bar{x} \left(\frac{n-1}{n}\right)^{n\bar{x}-1}
\end{aligned}
$$

is the UMVUE for $P(X=1)$.

Since the MLE for λ is $\hat{\lambda} = \bar{x}$, the MLE of $P(X=1)$ is $\bar{x} e^{-\bar{x}}$ by the invariance property of the MLE.

From Kendull, Stwart + Ord.

c) Finding the ARE for $\phi(T) = \left(\frac{n-1}{n}\right)^{n\bar{x}}$ w.r.t. $\hat{\lambda} = e^{-\bar{x}}$, we need to find the distribution of $\phi(T)$, which can be approximated by the delta method

$$\sqrt{n}\left(\phi(T) - \tau(\lambda)\right) \to N(0, \sigma^2)$$

where σ^2 is appoximated by

$$\sigma^2 = \mathrm{Var}\left(\{\phi(T)^2\}\right) \approx \frac{\phi'(\lambda)^2}{I(\lambda)}.$$

where

$$\phi'(\lambda) = n\left(\frac{n-1}{n}\right)^{n\lambda} \log\left(\frac{n-1}{n}\right)$$

and

$$I(\lambda) = -nE_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log(f_x(x\mid\lambda))\right] = -n\left(-\frac{\lambda}{\lambda^2}\right) = \frac{n}{\lambda}$$

So $\sqrt{n}\left(\phi(T) - \tau(\lambda)\right) \xrightarrow{D} N\left(0, \left\{\lambda\left(\frac{n-1}{n}\right)^{n\lambda}\log\left(\frac{n-1}{n}\right)\right\}^2\right)$

The variance of the MLE, $\hat{\lambda} = e^{-\bar{x}}$, is the CRLB because of the asymptotic efficiency of MLE's.
So the variance of $\hat{\lambda}$ is $\left\{\frac{de^{-\lambda}}{d\lambda}\right\}^2 = e^{-2\lambda}$

$$\sqrt{n}\left(\hat{\lambda} - \tau(\lambda)\right) \xrightarrow{D} N\left(0, \frac{\lambda e^{-2\lambda}}{n}\right).$$

The ARE of $\phi(T)$ w.r.t. $\hat{\lambda}$ is $\left\{\frac{e^{-\lambda}}{\left(\frac{n-1}{n}\right)^{n\lambda}\log\left(\frac{n-1}{n}\right)^n}\right\}^2$.

Finding the ARE for $\phi(T) = \bar{X}\left(\frac{n-1}{n}\right)^{n\bar{X}-1}$ w.r.t. $\hat{\lambda} = \bar{x}e^{-\bar{x}}$

we need to find the distribution of $\phi(T)$, which can be approximated by the delta method

$$\sqrt{n}\left(\phi(T) - \tau(\lambda)\right) \to N(0, \sigma^2)$$

where σ^2 is appoximated by

$$\sigma^2 = Var\left(\{\phi(T)^2\}\right) \approx \frac{\phi'(\lambda)^2}{I(\lambda)}.$$

where

$$\phi'(\lambda) = \left(\frac{n-1}{n}\right)^{n\lambda}\left(\lambda n \log\left(\frac{n-1}{n}\right) + 1\right)$$

and again,

$$I(\lambda) = -nE_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log(f_x(x\mid\lambda))\right] = -n\left(-\frac{\lambda}{\lambda^2}\right) = \frac{n}{\lambda}$$

So $\sqrt{n}\left(\phi(T) - \tau(\lambda)\right) \xrightarrow{D} N\left(0, \left[\frac{\lambda}{n}\left(\frac{n-1}{n}\right)^{n\lambda}\left(\lambda n \log\left(\frac{n-1}{n}\right)+1\right)\right]^2\right)$

The variance of the MLE, $\hat{\lambda} = \bar{x}e^{-x}$, is the CRLB because of the asymptotic efficiency of MLE's.

So the variance of $\hat{\lambda}$ is $\left\{\frac{de^{-\lambda}}{d\lambda}\right\}^2 = e^{-2\lambda}$

$$\sqrt{n}\left(\hat{\lambda} - \tau(\lambda)\right) \xrightarrow{D} N\left(0, \frac{\lambda(\lambda-1)^2 e^{-2\lambda}}{n}\right)$$

The ARE of $\phi(T)$ w.r.t. $\hat{\lambda}$ is $\left\{\frac{(\lambda-1)^2 e^{-2\lambda}}{\left(\frac{n}{n-1}\right)\left(\frac{n-1}{n}\right)^{n\lambda}\left(1+\log\left(\frac{n-1}{n}\right)^n\right)}\right\}^2.$

Both ARE $\to 1$ as $n \to \infty$, as they contain $\left\{\frac{(n-1)}{n}\right\}^n \to \frac{1}{e}$ as $n\to\infty$

d) From the data, $\bar{x} = \frac{104}{15} \approx 6.93$

Estimating $P(X=0)$

MLE: $P(X=0) \approx e^{-6.93} \approx 0.000974$

UMVUE: $P(X=0) \approx \left(\frac{15-1}{15}\right)^{104} \approx 0.000765$

I would choose the UMVUE since $n=15$.

Estimating $P(X=1)$

MLE: $P(X=0) \approx 6.93e^{-6.93} \approx 0.006758$

UMVUE: $P(X=0) \approx 6.93\left(\frac{15-1}{15}\right)^{104-1} \approx 0.005685$

Again, I would choose the UMVUE because of the smaller sample size, as it is more precise. We also know the UMVUE is unbiased in both cases.

3) We denote M_n as the sample median, whereas the $\bar{X}_n$ is the sample mean. We are given that if $f(x|\theta)$ is the pmf or pdf of the R.V. X, and it is symmetric about μ, then.

$$\sqrt{n}\left(M_n - \mu\right) \xrightarrow{D} N\left(0, \frac{1}{4nf^2(\mu)}\right)$$

a) Assume σ^2 exists for $f_x(x|\theta)$. By the CLT

$$\sqrt{n}\left(\bar{X}-\mu\right) \xrightarrow{D} N\left(0, \sigma^2/n\right).$$

A general expression for $ARE(M_n, \bar{X}_n)$ is

$$ARE\left(M_n, \bar{X}_n\right) = \frac{\sigma^2/n}{1/4nf^2(\mu)} = 4\sigma^2 f^2(\mu)$$

b) i) $ARE(M_n, \bar{X}_n) = 4\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu-\mu}{\sigma}\right)^2}\right)^2 = \frac{2}{\pi} e^{-\left(\frac{\mu-\mu}{\sigma}\right)^2} = \frac{2}{\pi} < 1$

So the mean is more asymptotically efficient.

ii) $ARE(M_n, \bar{X}_n) = 4 \frac{\pi^2\beta^2}{3}\left(\frac{1}{\beta^2}\left(\frac{e^{-(\mu-\mu)/\beta}}{(1+e^{-(\mu-\mu)/\beta})^2}\right)\right) = 4\frac{\pi^2\beta^2}{3}\left(\frac{1}{4\beta}\right)^2 = \frac{\pi}{12} < 1$

So the mean is more asymptotically efficient.

iii) $ARE(M_n, \bar{X}_n) = 4\left(2\sigma^2\right)\left(\frac{1}{2\sigma} e^{-|\mu-\mu|/\sigma}\right)^2 = 4\left(2\sigma^2\right)\left(\frac{1}{2\sigma}\right)^2 = 2 > 1.$

So the median is more asymptotically efficient.

iv) For the t_ν distribution $E[X] = \mu = 0$.

So then generally,

$$ARE\left(M_n, \bar{X}\right) = 4 \frac{\nu}{\nu-2}\left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\right)^2 = \frac{4}{\pi(\nu-2)}\left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)^2$$

and given

$\nu = 3$, $ARE(M_n, \bar{X}) = \frac{4}{\pi}\left(\frac{\Gamma(2)}{\Gamma(1.5)}\right)^2 > 1$

So the median is more asymptotically efficient.

$\nu = 5 \quad ARE(M_n, \bar{X}) = \left(\frac{4}{3\pi}\right)\left(\frac{\Gamma(3)}{\Gamma(2.5)}\right)^2 < 1$

So the mean is more asymptotically efficient.

$\nu = 10 \quad ARE(M_n, \bar{X}) = \left(\frac{1}{2\pi}\right)\left(\frac{\Gamma(5.5)}{\Gamma(5)}\right)^2 < 1$

So the mean is more asymptotically efficient.

$\nu = 25 \quad ARE(M_n, \bar{X}) = \left(\frac{4}{23\pi}\right)\left(\frac{\Gamma(13)}{\Gamma(12.5)}\right)^2 < 1$

So the mean is more asymptotically efficient.

$\nu = 50 \quad ARE(M_n, \bar{X}) = \left(\frac{1}{12\pi}\right)\left(\frac{\Gamma(25.5)}{\Gamma(25)}\right)^2 < 1$

So the mean is more asymptotically efficient.

$\nu = \infty$, as $\nu \to \infty$, $t_\nu \xrightarrow{D} Z$, so then for $\nu = \infty$, we would get $ARE(M_n, \bar{X}) = \frac{2}{\pi}$

c) Suppose $X_1, \ldots, X_n$ are iid with pdf/pmf $f_X(x|\theta)$. We define $E[X] = \mu$ and $Var[X] = \sigma^2$, which implies that from (a), $ARE[M_n, \bar{X}_n] = 4\sigma^2 f^2(\mu)$. Let W be the transformation $W_i = \frac{X_i - \mu}{\sigma}$. Then $f_Y(y) = \sigma f_X(\sigma y + \mu)$ with $ARE(M_n, \bar{Y}_n) = 4(1) f_Y^2(0) = 4\sigma^2 f_X^2(\mu)$. So ARE is unaffected by changes in scale.

4) a) To find a test statistic for $H_0: p_1 = p_2$ vs. $H_1: p_1 \neq p_2$, we know each draw from a binomial population is the sum of bernoulli random variables where $\hat{p}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\hat{p}_2 = \frac{1}{m}\sum_{i=1}^{m} Y_i$. This means that

$$E[\hat{p}_1] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{np_1}{n} = p_1, \quad E[\hat{p}_2] = \frac{1}{m}E\left[\sum_{i=1}^{m} Y_i\right] = \frac{mp_2}{m} = p_2$$

and $Var[\hat{p}_1] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right] = \frac{np_1(1-p_1)}{n^2} = \frac{p_1(1-p_1)}{n}$,

$Var[\hat{p}_2] = \frac{1}{m^2} Var\left[\sum_{i=1}^{m} Y_i\right] = \frac{mp_2(1-p_2)}{m^2} = \frac{p_2(1-p_2)}{m}$. Then by the CLT, we have that for

$$\frac{\hat{p}_1 - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} \xrightarrow{D} Z_1 \sim N(0,1) \quad \text{and} \quad \frac{\hat{p}_2 - p_2}{\sqrt{\frac{p_2(1-p_2)}{m}}} \xrightarrow{D} Z_2 \sim N(0,1)$$

Define $S_1 = \sum_{i=1}^{n} X_i$ and $S_2 = \sum_{i=1}^{m} Y_i$, so that $\hat{p} = \frac{S_1 + S_2}{n+m}$. Under H_0, $p_1 = p_2 = p$. So then the estimator for $Var[p] = \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}$

$$= \frac{mp_1 + np_2 - (mp_1^2 + np_2^2)}{nm} = p(1-p)\left(\frac{1}{n} + \frac{1}{m}\right).$$

Then, by Slutsky's Theorem, we have that

$$\frac{(\hat{p}_1 - \hat{p}_2) - \underbrace{(p_1 - p_2)}_{=0 \text{ under } H_0}}{\sqrt{(\hat{p}(1-\hat{p}))\left(\frac{1}{n} + \frac{1}{m}\right)}} \xrightarrow{D} Z$$

since $$\frac{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}{p(1-p)\left(\frac{1}{n}+\frac{1}{m}\right)} \xrightarrow{P} 1.$$

So then,

$$Y_{n,m} = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)} \xrightarrow{D} Z^2 \sim \chi^2_1.$$

b) So let $W_{n,m} = \sum \frac{(\text{observed} - \text{expected})^2}{\text{observed}}$, then we have that $S_1 = np_1$, $S_2 = mp_2$, $F_1 = n(1-p_1)$, and $F_2 = m(1-p_2)$.

Notice how algebraically, $Y_{n,m} = \frac{\left(\frac{nm}{n+m}\right)(\hat{p}_1 - \hat{p}_2)}{\hat{p}(1-\hat{p})}$.

So then we would write W as

$$W_{nm} = \frac{\left(S_1 - \frac{nS}{n+m}\right)^2}{\frac{nS}{n+m}} + \frac{\left(S_2 - \frac{mS}{n+m}\right)^2}{\frac{mS}{n+m}} + \frac{\left(F_1 - \frac{nF}{n+m}\right)^2}{\frac{nF}{n+m}} + \frac{\left(F_2 - \frac{mF}{n+m}\right)^2}{\frac{mF}{n+m}}$$

$$= \frac{(n\hat{p}_1 - n\hat{p})^2}{n\hat{p}} + \frac{(m\hat{p}_2 - m\hat{p})^2}{m\hat{p}} + \frac{(n(1-\hat{p}_1) + n(1-\hat{p}))^2}{n(1-\hat{p})} + \frac{(m(1-\hat{p}_2) + m(1-\hat{p}_2))^2}{m(1-\hat{p})}$$

$$= \frac{n(\hat{p}_1 - \hat{p})^2 + m(\hat{p}_2 - \hat{p})^2}{\hat{p}} + \frac{n(\hat{p}_1 - \hat{p})^2 + m(\hat{p}_2 - \hat{p})}{(1-\hat{p})}$$

$$= \frac{n(\hat{p}_1 - \hat{p})^2}{\hat{p}(1-\hat{p})} + \frac{m(p_2 - \hat{p})}{\hat{p}(1-\hat{p})} = \frac{n\left(\hat{p}_1 - \frac{n_1\hat{p}_1 + m\hat{p}_2}{n+m}\right)^2}{(1-\hat{p})} + \frac{m\left(\hat{p}_2 - \frac{n_1\hat{p}_1 + m\hat{p}_2}{n+m}\right)^2}{(1-\hat{p})}$$

$$= \frac{\left(\frac{n}{(n+m)^2}\right)(m(\hat{p}_1 - \hat{p}_2))^2}{\hat{p}(1-\hat{p})} + \frac{\left(\frac{m}{(n+m)^2}\right)(n(\hat{p}_1 - \hat{p}_2))^2}{\hat{p}(1-\hat{p})} = \frac{\left(\frac{nm}{n+m}\right)(\hat{p}_1 - \hat{p}_2)^2}{\hat{p}(1-\hat{p})} = Y_{n,m}.$$

which implies $W_{n,m} \xrightarrow{D} \chi^2_1$. QED

c) Since $\hat{p}_1$ is the MLE for p_1 and $\hat{p}_2$ is the MLE of p_2, by the invariance property, as well as asymptotic efficiency of the MLE, we know $\hat{p}_1(1-\hat{p}_1) \xrightarrow{P} p(1-p)$ and $\hat{p}_2(1-\hat{p}_2) \xrightarrow{P} p(1-p)$, as $p_1 = p_2 = p$ under H_0. So then also applying Slutsky's Theorem, we then get

$$Z_{n,m} = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}} \xrightarrow{D} Z \sim N(0,1)$$

So then by properties of standard normal R.V.'s, we have that $(Z_{n,m})^2 \sim \chi^2_1$

Also, since $\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{n} \neq \hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)$, it is trivial that $(Z_{n,m})^2 \neq Y_{n,m}$.

d) We know that given H_0, the MLE of $Var(\hat{p}_1 - \hat{p}_2)$ is

$$Var(\hat{p}_1 - \hat{p}_2) = p(1-p)\left(\frac{1}{n}+\frac{1}{m}\right).$$

by properties of variance, if H_0 is false, then

$$Var(p_1 - p_2) = \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}$$

So for a hypothesis test, would be a better estimator, as it is the more consistent estimator under H_0. Hence since the variance estimator used in

$Y_{n,m}$ is equal to the variance estimator in $Z_{n,m}$ under H_0, $Z_{n,m}$ would be a better statistic for a confidence interval.

e) So $\hat{p}_1 = 0.85$ and $\hat{p}_2 = 0.54$, $\hat{p} = 0.71$.

Then $Y_{n,m} = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\hat{p}(1-\hat{p})(\frac{1}{n}+\frac{1}{m})} = 8.49516$, p-Value $= P(Y_{n,m} > \chi^2_1) = 0.003561$

and $Z_{n,m} = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}} = 3.02962$, p-value $= P(|z| > 3.029) = 0.002449$

$\Big\}$ computed with R

For all reasonable values of α, both tests would reject H_0 in favor of H_1.

5) Given $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, considering the hypotheses

$$H_0: p = p_0 \quad \text{vs.} \quad H_1: p \neq p_0,$$

we derive a likelihood ratio test statistic. We know from prior homeworks that the MLE of p is $\hat{p} = \overline{X}$

$$\lambda(\underline{x}) = \frac{L(p_0 | \underline{x})}{L(\hat{p} | \underline{x})} = \frac{p_0^{\sum_{i=1}^n x_i}(1-p_0)^{n-\sum_{i=1}^n x_i}}{\hat{p}^{\sum_{i=1}^n x_i}(1-\hat{p})^{n-\sum_{i=1}^n x_i}} = \left(\frac{p_0}{\hat{p}}\right)^{n\hat{p}}\left(\frac{1-p_0}{1-\hat{p}}\right)^{n(1-\hat{p})}$$

a) So then,

$$-2\log(\lambda(\hat{p})) = -2n\left(\hat{p}(\log p_0 - \log\hat{p}) + (1-\hat{p})(\log(1-p_0) - \log(1-\hat{p}))\right)$$

$$= -2n\left(\hat{p}\log p_0 - \hat{p}\log\hat{p} + \log(1-p_0) - \log(1-\hat{p}) - \hat{p}\log(1-p_0) + \hat{p}\log(1-\hat{p})\right)$$

$$= -2n\left(\hat{p}\log\left(\frac{p_0}{\hat{p}}\right) + \hat{p}\log\left(\frac{1-\hat{p}}{1-p_0}\right) + \log\left(\frac{1-p_0}{1-\hat{p}}\right)\right) \quad \square$$

b) See End of Homework.

6) Given $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

a) Suppose σ^2 is known, μ unknown. Consider the null hypothesis

$$H_0: \mu = \mu_0.$$

Given that $\bar{X}$ is an estimator for μ, we have that the sampling distribution for $\bar{X} \sim N(\mu, \sigma^2/n)$. So then if σ^2 is known, then let $Z_n = \sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)$. Since the normal distribution is a location-scale family, then $Z_n \sim Z$. So a Wald statistic for testing H_0 is

$$Z_n = \sqrt{n}\left(\frac{\bar{X}-\mu_0}{\sigma}\right).$$

b) Given μ is known and σ^2 is unknown. We want to test $H_0: \sigma_0 = \sigma$. We want a Wald statistic, based on a standardized estimator that is asymptotically normal, as well as consistent. We can use the MLE for σ^2, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ as a consistent estimator for σ^2 because of asymptotic efficiency for MLE's, we have

$$\sqrt{n}\left(\hat{\sigma}^2 - \sigma^2\right) \xrightarrow{D} N\left(0, \frac{1}{I(\sigma^2)}\right).$$

where $I(\sigma^2)$ is Fisher's information number.

So $\frac{1}{I(\sigma^2)} = \left(-n E\left[\left(\frac{\partial^2}{\partial(\sigma^2)^2} \log f(\underline{x}|\sigma^2)\right)\right]\right)^{-1} = \left(-n E\left[\frac{\partial^2}{\partial(\sigma^2)^2}\left(-\frac{1}{2}\log(\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}\right)\right]\right)^{-1}$

$= \left(-n E\left[-\frac{1}{2(\sigma^2)^2} + \frac{(x-\mu)^2}{2(\sigma^2)^3}\right]\right)^{-1} = \left(-\frac{n}{2(\sigma^2)^2} + \frac{nE[(x-\mu)^2]}{2(\sigma^2)^3}\right)^{-1} = \left(-\frac{n}{2(\sigma^2)^2} + \frac{n(\sigma^2)^2}{(\sigma^2)^3}\right)^{-1}$

$= \frac{2(\sigma^2)^2}{n}$. So the asymptotic distribution.

Because the MLE of σ^2, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ is consistent, we can apply the invariance property of the MLE to yield $\sqrt{\hat{\sigma}^2} = \hat{\sigma}$ as a consistent estimator for σ. Since the parent population is normally distributed, and is a full exponential family, we can apply the Delta method for $\sqrt{\sigma^2}$. If $g(x) = \sqrt{x}$, then

$$\sqrt{n}\left(\sqrt{\hat{\sigma}^2} - \sqrt{\sigma^2}\right) \xrightarrow{D} N\left(0, \frac{2(\sigma^2)^2}{n}\left(\frac{1}{2\sqrt{\sigma^2}}\right)^2\right)$$

$$\implies \sqrt{n}\left(\sqrt{\hat{\sigma}^2} - \sqrt{\sigma^2}\right) \xrightarrow{D} N\left(0, \frac{\sigma^2}{n}\right).$$

We can estimate this variance with $\hat{\sigma}^2$, the MLE of σ^2, which is consistent. Hence

$$\frac{\hat{\sigma}^2}{2} \xrightarrow{P} \frac{\sigma^2}{2} \implies \frac{\hat{\sigma}^2}{\sigma^2} = 1$$

Via Slutsky's Theorem,

$$Z_n = \frac{\hat{\sigma} - \sigma_0}{\sqrt{\frac{\sigma^2}{2n}}} = \sqrt{n}\left(\frac{\hat{\sigma} - \sigma}{\hat{\sigma}/\sqrt{2}}\right) \xrightarrow{D} Z.$$

So the Wald statistic is $\quad Z_n = \dfrac{\hat{\sigma} - \sigma_0}{\hat{\sigma}/2}$.

7) Given $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

Suppose σ^2 is known, μ unknown. Consider the null hypothesis

$$H_0 : \mu = \mu_0 .$$

a) If μ is unknown and σ^2 is known, the score function is over μ and is

$$S(\mu) = \frac{\partial}{\partial \mu}\left(-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 \right)$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$$

$$= \frac{n(\bar{x} - \mu)}{\sigma^2} .$$

To find the information number, or variance of $S(\mu)$, we appeal to the fact that the parent population is a full exponential family. So then

$$I(\sigma^2) = -E\left[\left\{\frac{\partial^2}{\partial \mu^2}\mathcal{L}(\mu \mid \underline{x})\right\}\right]$$

$$= -E\left[\frac{\partial}{\partial \mu} \frac{n(\bar{x} - \mu)}{\sigma^2}\right]$$

$$= \frac{n}{\sigma^2}$$

So the score test statistic is

$$Z_s = \frac{\frac{n(\bar{x} - \mu_0)}{\sigma^2}}{\sqrt{\frac{n}{\sigma^2}}} = \sqrt{n}\left(\frac{\bar{x} - \mu_0}{\sigma}\right). \quad \text{QED}$$

b) Suppose σ^2 is unknown and μ is known. We need to find a Score statistic to test $H_0: \sigma = \sigma_0$.

Here the Score function is

$$S(\sigma) = \frac{\partial}{\partial \sigma} L(\sigma \mid \underline{x})$$

$$= \frac{\Sigma (x_i - \mu)^2}{\sigma^3} - \frac{n}{\sigma}$$

So the variance of the score function is the information number.

$$Var(S(\sigma)) = -E\left[\left\{\frac{\partial^2}{\partial \sigma^2} \log L(\sigma \mid \underline{X})\right\}\right]$$

$$= -E\left[\frac{\partial}{\partial \sigma}\left\{\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^3} - \frac{n}{\sigma}\right\}\right]$$

$$= E\left[\frac{n}{\sigma^2} - \frac{3\sum_{i=1}^{n}(x_i - \mu)}{\sigma^4}\right]$$

$$= \frac{n}{\sigma^2} - \frac{3\sigma^2}{\sigma^4} = \frac{n-3}{\sigma^2}$$

So the Score statistic is

$$Z_s = \frac{S(\sigma_0)}{\sqrt{I(\sigma_0)}} = \frac{\left(\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma_0^3} - \frac{n}{\sigma_0}\right)}{\sqrt{\frac{n-3}{\sigma_0}}} = \sqrt{n-3}\left(\frac{\hat{\sigma}^2 - \sigma_0^2}{\sigma_0^2}\right).$$

8)

a) We are given $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$. To find the limiting distribution of $\sqrt{n}\left(\frac{\bar{X}-\lambda}{\sqrt{\lambda}}\right)$, we first recognize that $\bar{X}$ is the MLE of λ. Additionally, $E[\bar{X}] = \lambda$, and $Var[\bar{X}] = \frac{\lambda}{n}$. The Poisson family is a full exponential family, it satisfies all regularity conditions, so the we can write,

$$\sqrt{n}(\bar{x}-\lambda) \xrightarrow{D} N\left(0, \frac{\lambda}{n}\right).$$

So then we recognize that $\sqrt{n}\left(\frac{\bar{X}-\lambda}{\sqrt{\lambda}}\right) \sim Z$.

b) We can show

$$\frac{\frac{\partial}{\partial\lambda}\mathcal{L}(\lambda|\underline{x})}{\sqrt{-E_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\mathcal{L}(\lambda|x)\right]}} = \frac{\frac{\partial}{\partial\lambda}\left(-n\lambda + \left(\sum_{i=1}^n x_i\right)\log\lambda - \sum_{i=1}^n \log(x_i!)\right)}{\sqrt{-E_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\left(-n\lambda + \left(\sum_{i=1}^n x_i\right)\log\lambda - \sum_{i=1}^n \log(x_i!)\right)\right]}}$$

$$= \frac{\frac{n(\bar{x}-\lambda)}{\lambda}}{\sqrt{-E\left[-\frac{\sum_{i=1}^n x_i}{\lambda^2}\right]}} = \frac{\frac{n(\bar{x}-\lambda)}{\lambda}}{\sqrt{\frac{n}{\lambda}}} = \sqrt{n}\left(\frac{\bar{x}-\lambda}{\sqrt{\lambda}}\right)$$

So this shows that $\sqrt{n}\left(\frac{\bar{x}-\lambda}{\sqrt{\lambda}}\right)$ is the optimal result.

c) As $n \to \infty$, we have that

$$\frac{\frac{\partial}{\partial \lambda} \ell(\lambda \mid \underline{x})}{\sqrt{-E_{\lambda}\left[\frac{\partial^2}{\partial \lambda^2} \ell(\lambda \mid x)\right]}} \xrightarrow{D} Z \sim N(0,1)$$

So via a Score interval, the $1-\alpha$ approximate CI is $\left\{\lambda : \left|\sqrt{n}\left(\frac{\bar{X}-\lambda}{\lambda}\right)\right| \leq z_{\alpha/2}\right\}$. So this implies that

$$\bar{X} - \frac{z_{\alpha/2}^2 - z_{\alpha/2}\sqrt{4\bar{X} + z_{\alpha/2}^2}}{2} \leq \lambda \leq \bar{X} + \frac{z_{\alpha/2}^2 - z_{\alpha/2}\sqrt{4\bar{X} + z_{\alpha/2}^2}}{2}$$

by solving via the quadratic formula. $\square$

Theory 2 Homework 8

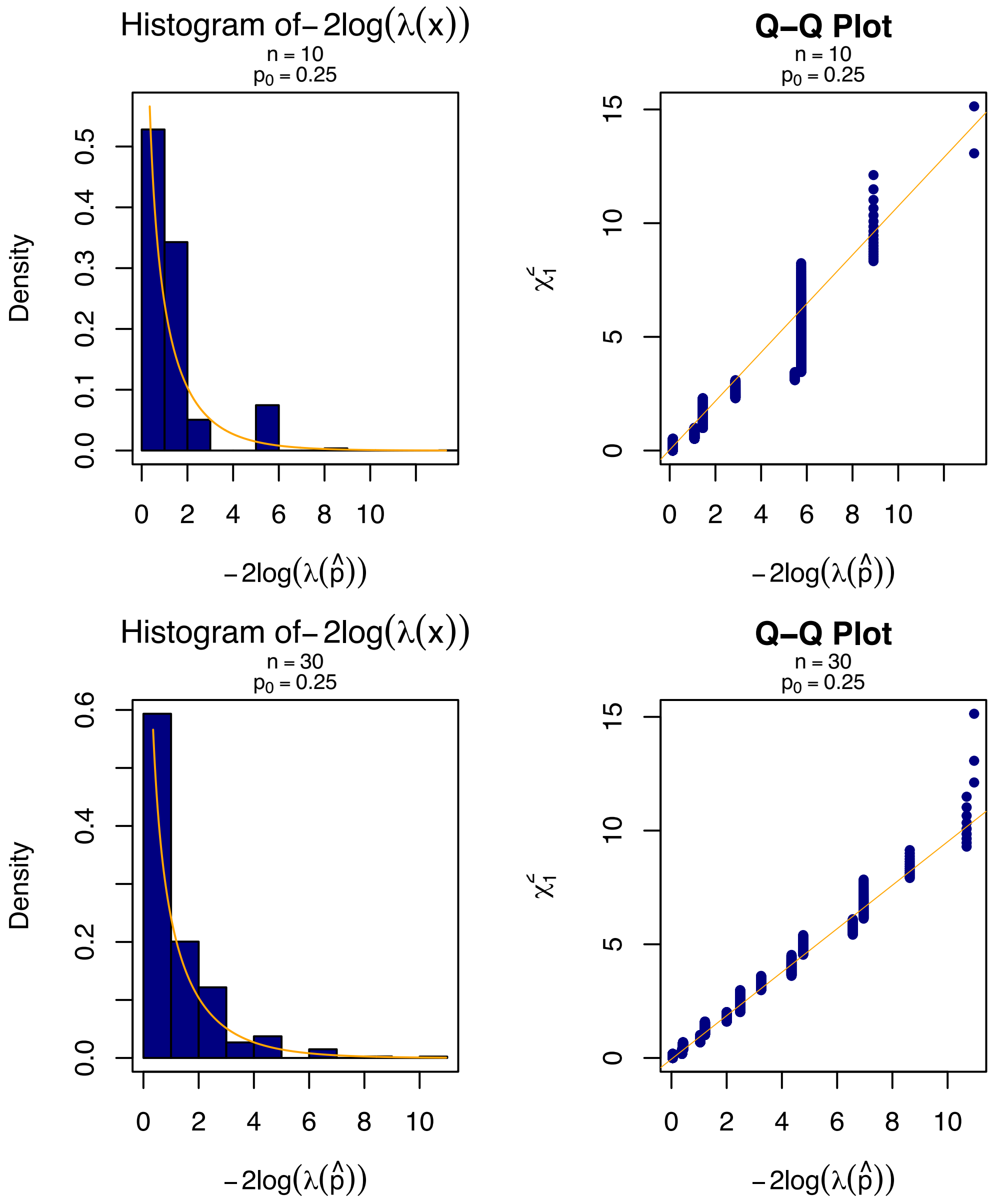
Question 5 (b)

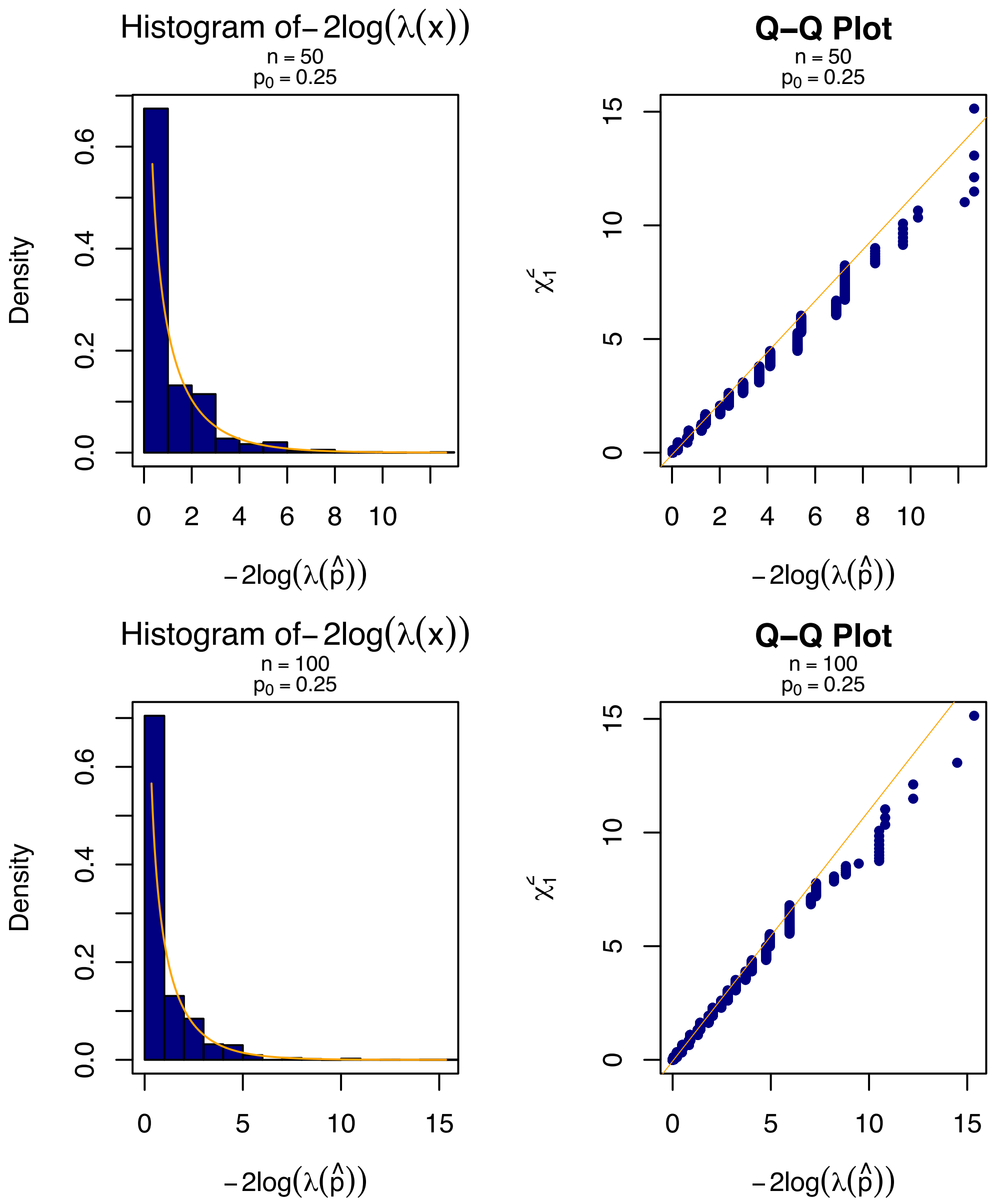
```r
asymp.bin.lrt <- function(n = 10, p0 = 0.5, reps = 5000, seed = 0) {
  percs <- c(0.25, 0.50, 0.75, 0.95, 0.99)
  if(seed > 0) set.seed(seed)
  X     <- matrix(rbinom(n = n*reps, size = 1, prob = p0),
                  nrow = reps, ncol = n)
  phat  <- rowMeans(X)
  rat   <- (p0/phat)^(n*phat)
  ratom <- ((1 - p0)/(1 - phat))^(n*(1 - phat))
  LRT   <- -2*log(rat*ratom)
  min.x <- min(LRT)
  max.x <- max(LRT)
  x     <- seq(from = 0.35, to = max.x, length.out = 501)
  y     <- dchisq(x, df = 1)
  max.y <- max(hist(LRT, plot = FALSE)$density, y)
  par(mfrow = c(1, 2), pch = 20, cex = 0.85)
  q.x <- ((1:(n+1)) - 0.5)/n
  hist(LRT, prob = T,
       xlim = c(min.x, max.x), ylim = c(0, max.y),
       xlab = expression(-2*log(lambda(hat(p)))),
       ylab = "Density",
       main = expression(paste("Histogram of", -2*log(lambda(x)))),
       col = "navy")
  mtext(bquote(n == .(n)), line = 0.75, cex = 0.7)
  mtext(bquote(p[0] == .(p0)), line = 0, cex = 0.7)
  lines(x, y, col = "orange")
  box()
  q.x <- ppoints(reps)
  qqplot(LRT, qchisq(q.x, df = 1),
         xlab = expression(-2*log(lambda(hat(p)))),
```

```
        ylab = expression(chi[1]^2), main = "Q-Q Plot",
        col = "navy")
  qqline(LRT, distribution = function(q.x) qchisq(q.x, df = 1),
        col = "orange", lwd = 0.5)
  mtext(bquote(n == .(n)), line = 0.75, cex = 0.7)
  mtext(bquote(p[0] == .(p0)), line = 0, cex = 0.7)
  emp.percs <- quantile(LRT, probs = percs, names = TRUE)
  return(list(Empirical.quantiles = emp.percs,
              Chi.square.quantiles = qchisq(percs, 1)))
}
```

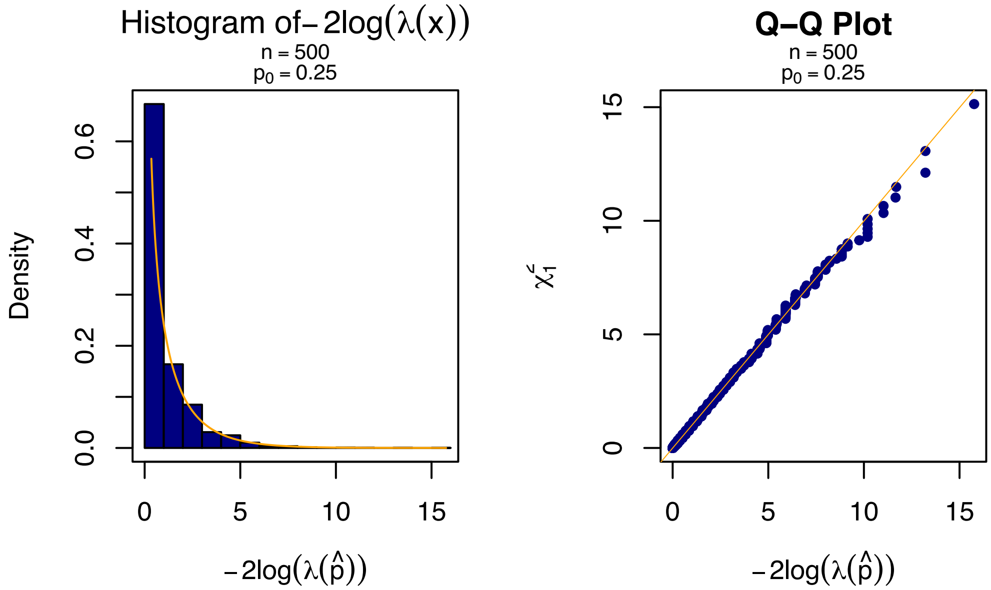
For $p_0 = 0.25$

```r
seed <- 12345
ns <- c(10, 30, 50, 100, 500)
lapply(ns, \(n) {
  data.frame(asymp.bin.lrt(n = n, p0 = 0.25, seed = seed),
             "n" = n)
  }) |> knitr::kable(digits = 3)
```





Histogram of − 2log(λ(x))
n = 50
p_0 = 0.25
Density
−2log(λ(p̂))
Q–Q Plot
n = 50
p_0 = 0.25
χ^2_1
−2log(λ(p̂))
Histogram of − 2log(λ(x))
n = 100
p_0 = 0.25
Density
−2log(λ(p̂))
Q–Q Plot
n = 100
p_0 = 0.25
χ^2_1
−2log(λ(p̂))


Histogram of −2log(λ(x))
n = 500
p0 = 0.25
Density
−2log(λ(p̂))
Q–Q Plot
n = 500
p0 = 0.25
χ²1
−2log(λ(p̂))

	Empirical.quantiles	Chi.square.quantiles	n
25%	0.140	0.102	10
50%	0.140	0.455	10
75%	1.449	1.323	10
95%	5.754	3.841	10
99%	5.754	6.635	10

	Empirical.quantiles	Chi.square.quantiles	n
25%	0.045	0.102	30
50%	0.420	0.455	30
75%	1.213	1.323	30
95%	4.348	3.841	30
99%	6.960	6.635	30

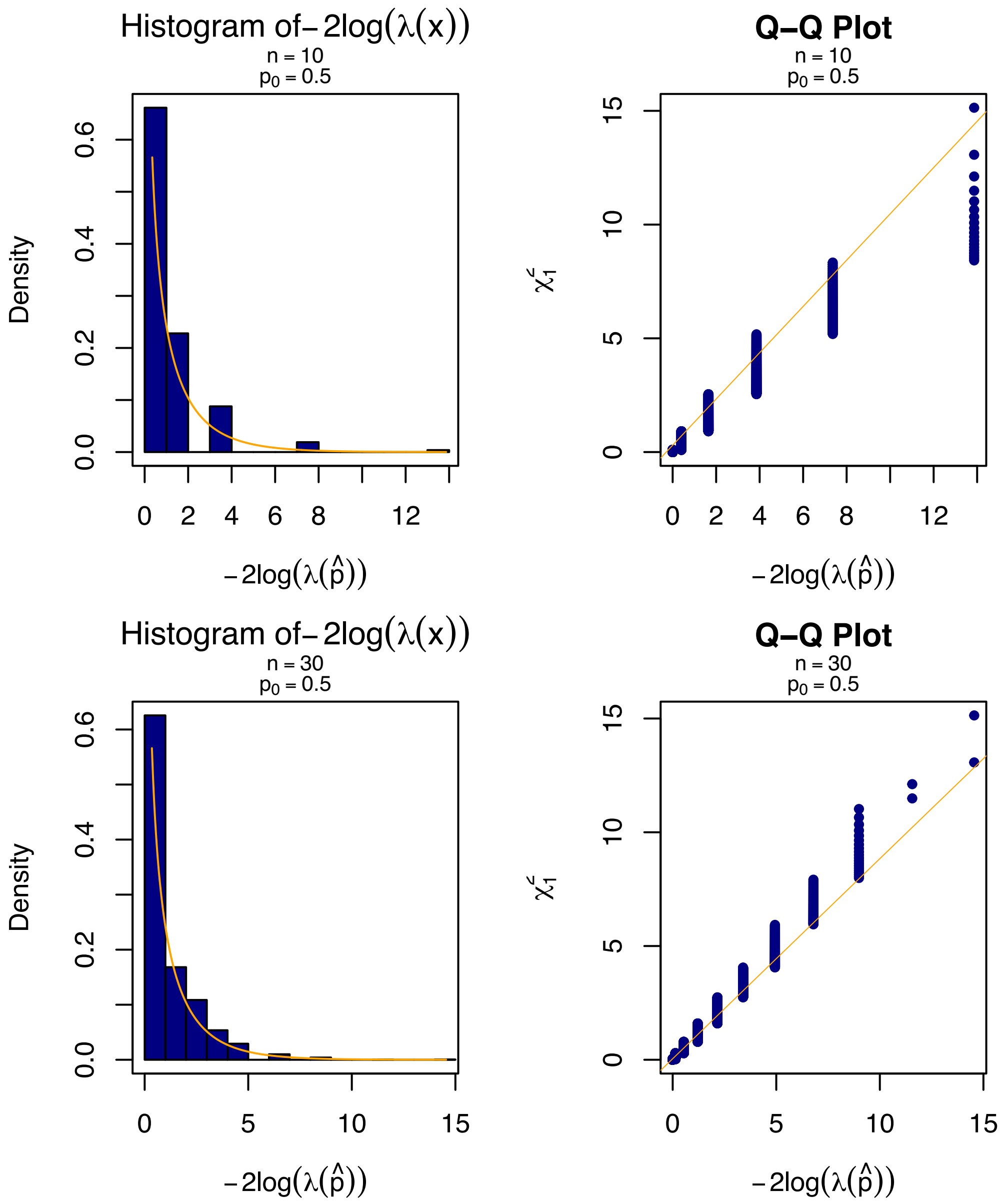
	Empirical.quantiles	Chi.square.quantiles	n
25%	0.027	0.102	50
50%	0.640	0.455	50
75%	1.404	1.323	50
95%	4.109	3.841	50
99%	6.881	6.635	50

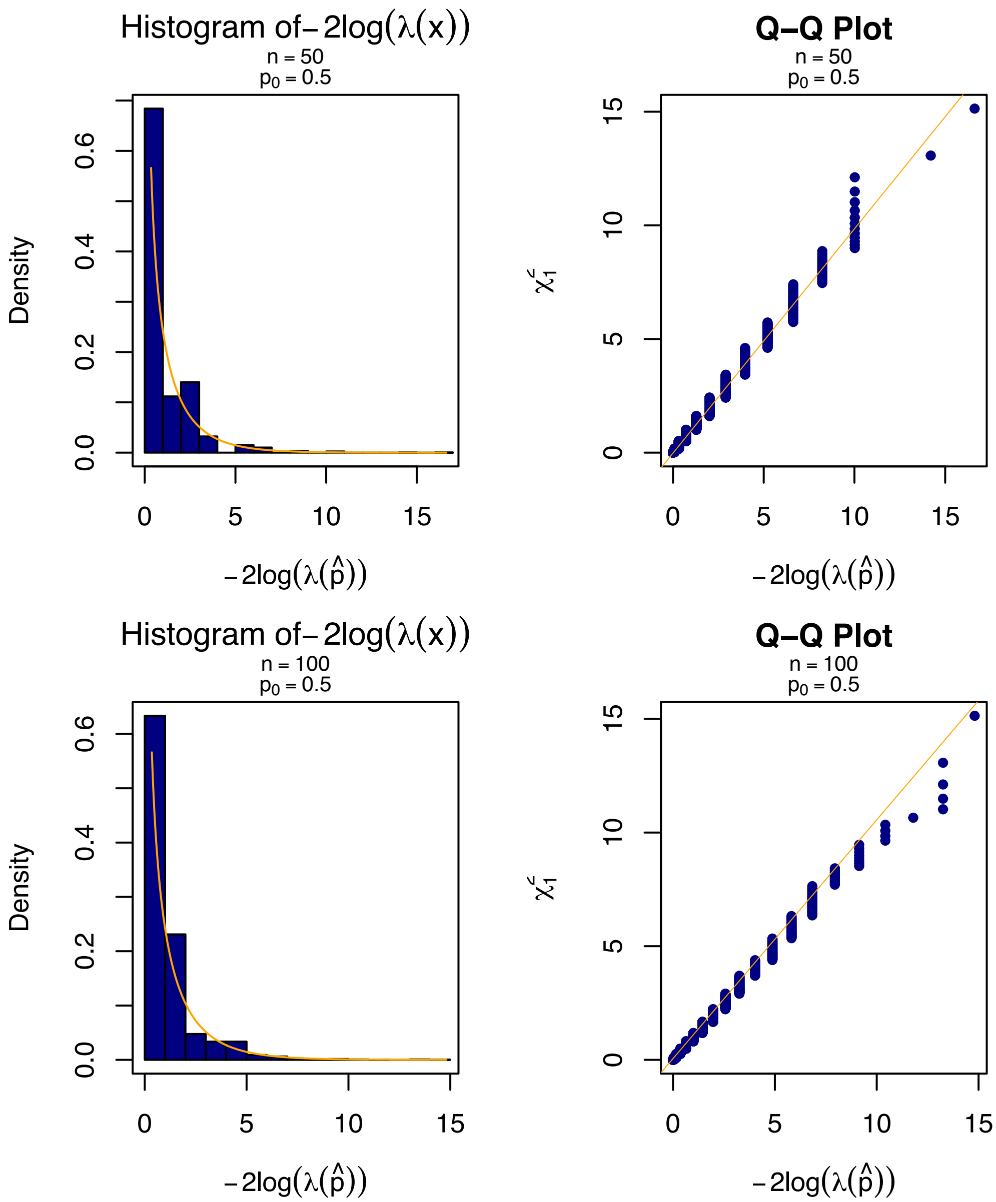
	Empirical.quantiles	Chi.square.quantiles	n
25%	0.054	0.102	100
50%	0.468	0.455	100
75%	1.400	1.323	100
95%	3.712	3.841	100
99%	5.953	6.635	100

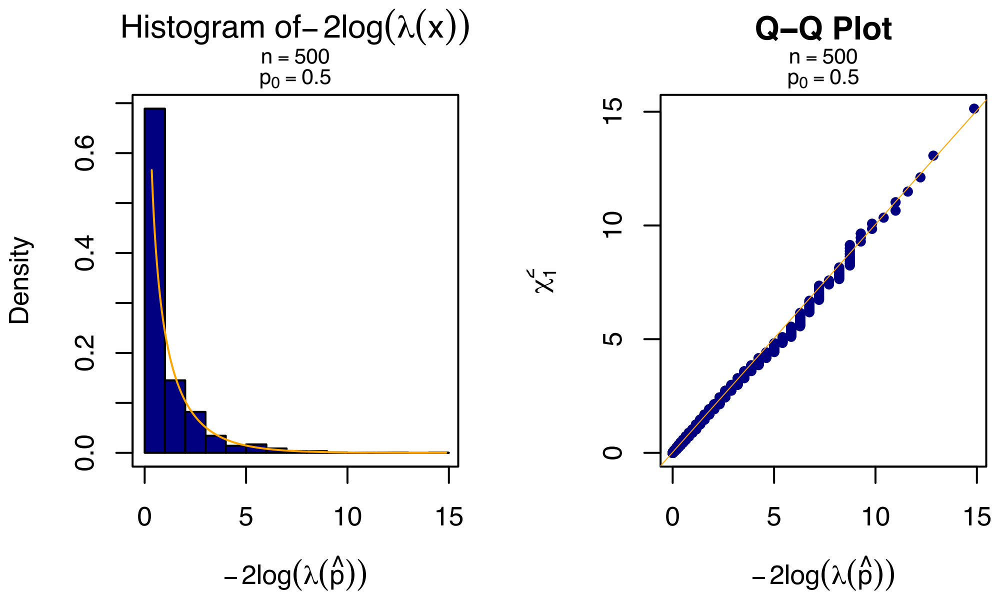
	Empirical.quantiles	Chi.square.quantiles	n
25%	0.097	0.102	500
50%	0.516	0.455	500
75%	1.317	1.323	500
95%	3.993	3.841	500
99%	6.439	6.635	500

For $p_0 = 0.5$

```r
lapply(ns, \(n) {
  data.frame(asymp.bin.lrt(n = n, p0 = 0.5, seed = seed),
             "n" = n)
  }) |> knitr::kable(digits = 3)
```




Histogram of $-2\log(\lambda(x))$
$n = 50$
$p_0 = 0.5$
Density
$-2\log(\lambda(\hat{p}))$
Q–Q Plot
$n = 50$
$p_0 = 0.5$
χ^2_1
$-2\log(\lambda(\hat{p}))$
Histogram of $-2\log(\lambda(x))$
$n = 100$
$p_0 = 0.5$
Density
$-2\log(\lambda(\hat{p}))$
Q–Q Plot
$n = 100$
$p_0 = 0.5$
χ^2_1
$-2\log(\lambda(\hat{p}))$


Histogram of $-2\log(\lambda(x))$
n = 500
$p_0 = 0.5$
Density
0.0
0.2
0.4
0.6
0
5
10
15
$-2\log(\lambda(\hat{p}))$
Q–Q Plot
n = 500
$p_0 = 0.5$
χ^2_1
0
5
10
15
0
5
10
15
$-2\log(\lambda(\hat{p}))$

	Empirical.quantiles	Chi.square.quantiles	n
25%	0.403	0.102	10
50%	0.403	0.455	10
75%	1.646	1.323	10
95%	3.855	3.841	10
99%	7.361	6.635	10

	Empirical.quantiles	Chi.square.quantiles	n
25%	0.133	0.102	30
50%	0.535	0.455	30
75%	1.208	1.323	30
95%	3.398	3.841	30
99%	6.794	6.635	30

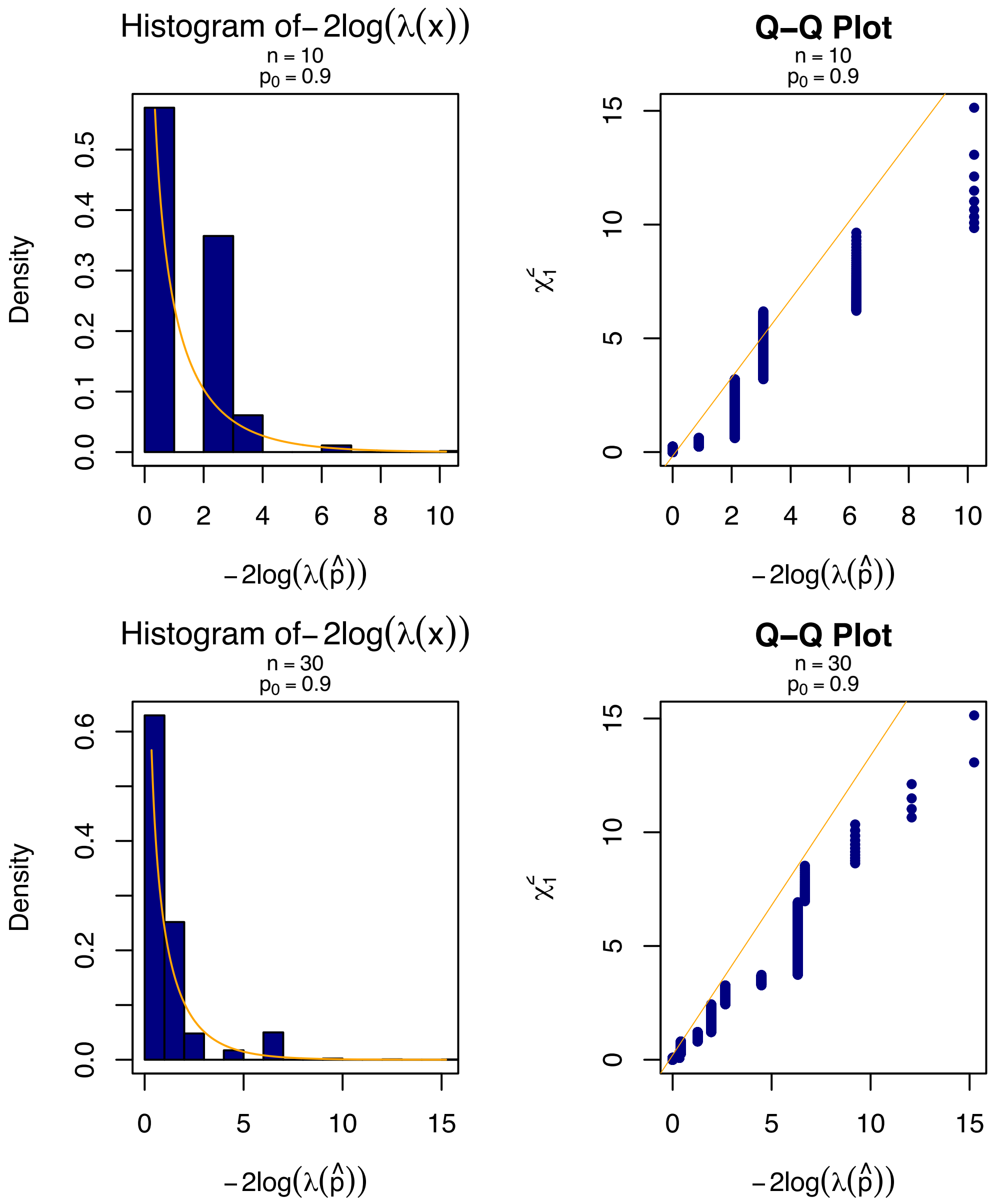
	Empirical.quantiles	Chi.square.quantiles	n
25%	0.080	0.102	50
50%	0.320	0.455	50
75%	1.286	1.323	50
95%	3.973	3.841	50
99%	6.628	6.635	50

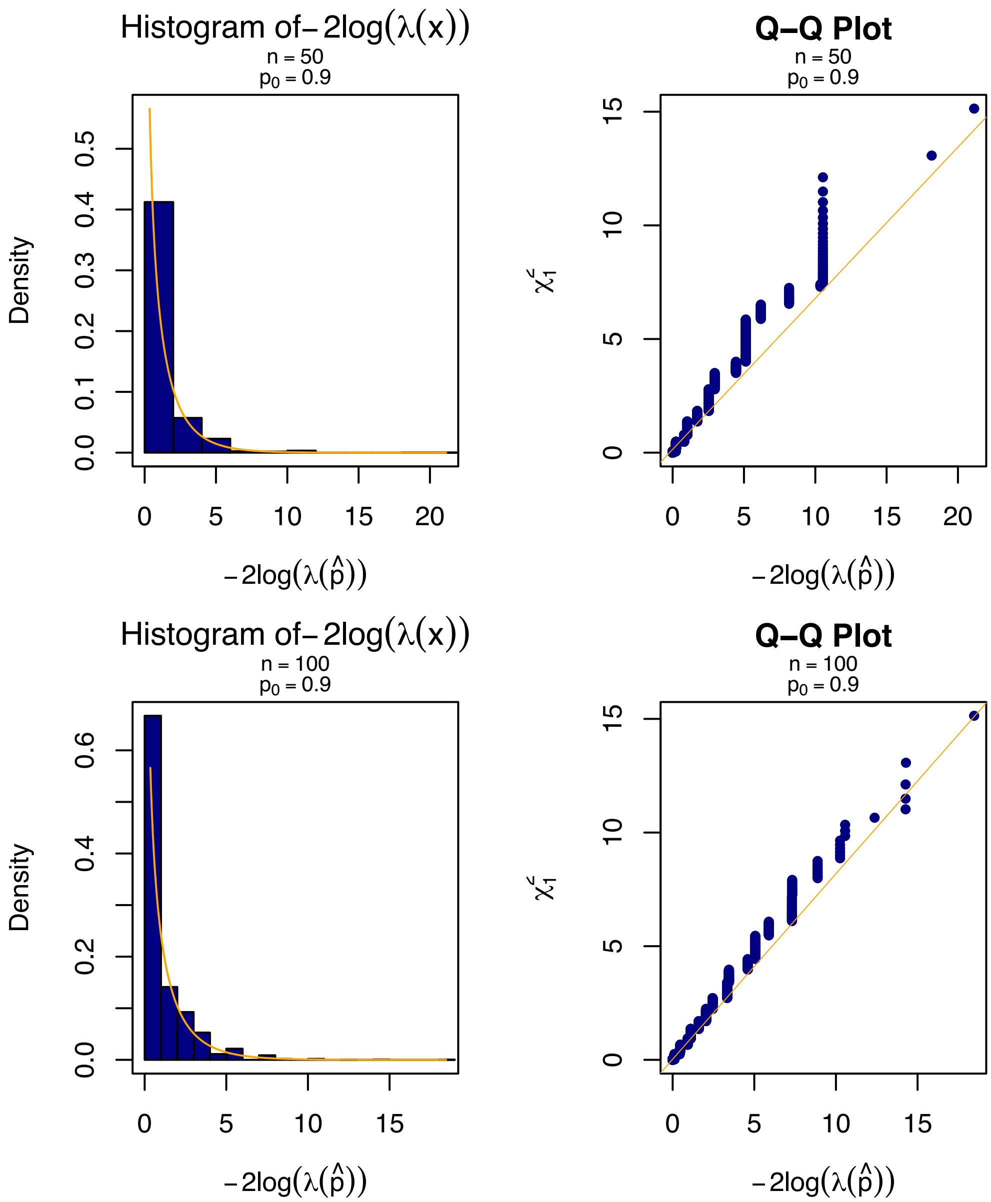
	Empirical.quantiles	Chi.square.quantiles	n
25%	0.160	0.102	100
50%	0.360	0.455	100
75%	1.443	1.323	100
95%	4.027	3.841	100
99%	6.838	6.635	100

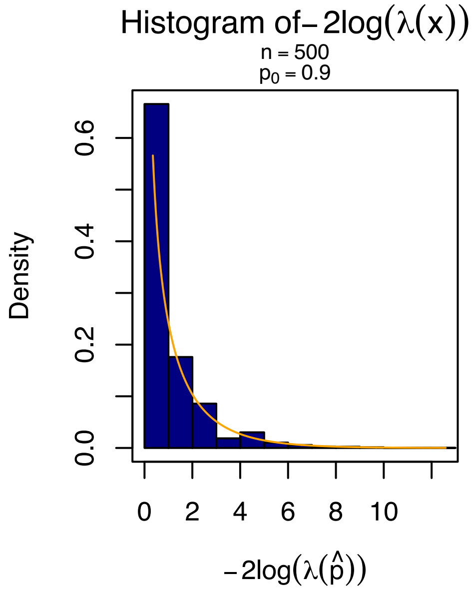
	Empirical.quantiles	Chi.square.quantiles	n
25%	0.128	0.102	500
50%	0.512	0.455	500
75%	1.353	1.323	500
95%	3.877	3.841	500
99%	6.743	6.635	500

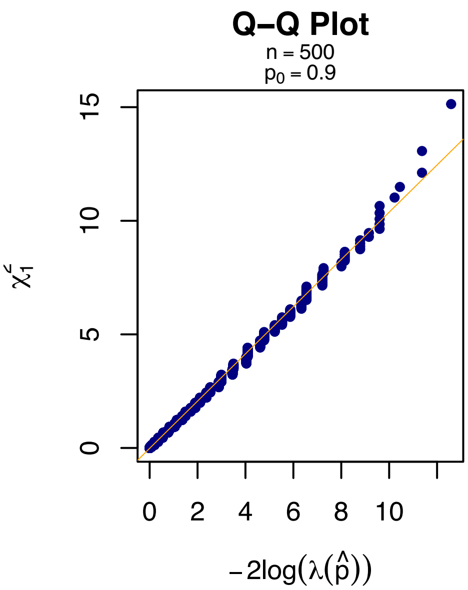
For $p_0 = 0.9$

```
lapply(ns, \(n) {
  data.frame(asymp.bin.lrt(n = n, p0 = 0.9, seed = seed),
             "n" = n)
  }) |> knitr::kable(digits = 3)
```




Histogram of −2log(λ(x))
n = 50
p0 = 0.9
Density
−2log(λ(p̂))
Q–Q Plot
n = 50
p0 = 0.9
χ²1
−2log(λ(p̂))
Histogram of −2log(λ(x))
n = 100
p0 = 0.9
Density
−2log(λ(p̂))
Q–Q Plot
n = 100
p0 = 0.9
χ²1
−2log(λ(p̂))


Histogram of −2log(λ(x))
n = 500
$p_0 = 0.9$
Density
0.0
0.2
0.4
0.6
0
2
4
6
8
10
$-2\log(\lambda(\hat{p}))$


Q–Q Plot
n = 500
$p_0 = 0.9$
χ_1^2
0
5
10
15
0
2
4
6
8
10
$-2\log(\lambda(\hat{p}))$

	Empirical.quantiles	Chi.square.quantiles	n
25%	0.000	0.102	10
50%	0.888	0.455	10
75%	2.107	1.323	10
95%	3.073	3.841	10
99%	6.225	6.635	10

	Empirical.quantiles	Chi.square.quantiles	n
25%	0.339	0.102	30
50%	0.415	0.455	30
75%	1.947	1.323	30
95%	6.322	3.841	30
99%	6.322	6.635	30

	Empirical.quantiles	Chi.square.quantiles	n
25%	0.210	0.102	50
50%	0.237	0.455	50
75%	1.023	1.323	50
95%	4.440	3.841	50
99%	8.161	6.635	50

	Empirical.quantiles	Chi.square.quantiles	n
25%	0.108	0.102	100
50%	0.474	0.455	100
75%	1.105	1.323	100
95%	3.449	3.841	100
99%	7.307	6.635	100

	Empirical.quantiles	Chi.square.quantiles	n
25%	0.090	0.102	500
50%	0.540	0.455	500
75%	1.360	1.323	500
95%	4.038	3.841	500
99%	6.548	6.635	500