Question 1

(1a)(i) $\log f(y,\omega) = y\theta(\omega) - k(\theta) + g(y)$

$$\frac{\partial \log f(y,\omega)}{\partial \omega} = \frac{\partial \theta}{\partial \omega}\left\{y - \frac{\partial k}{\partial \theta}\right\}$$

$$\frac{\partial^2 \log f(y,\omega)}{\partial \omega^2} = \frac{\partial^2 \theta}{\partial \omega^2}\left\{y - \frac{\partial k}{\partial \theta}\right\} - \left(\frac{\partial \theta}{\partial \omega}\right)^2 \frac{\partial^2 k}{\partial \theta^2}$$

The Fisher information is

$$I(\omega) = E\left[\left(\frac{\partial \theta}{\partial \omega}\right)^2 \frac{\partial^2 k}{\partial \theta^2} - \frac{\partial^2 \theta}{\partial \omega^2}\left\{y - \frac{\partial k}{\partial \theta}\right\}\right] = \left(\frac{\partial \theta}{\partial \omega}\right)^2 \frac{\partial^2 k}{\partial \theta^2}$$

(ii) $$P\{Y = k \mid Y \neq j\} = \frac{\pi(1-\pi)^{k-1}}{1 - \pi(1-\pi)^{j-1}} \quad k \neq 0 \text{ and } k \neq j$$

(iii) The log-likelihood of $\{u_i\}_{i=1}^{n}$ is

$$\mathcal{L}_n(\pi) = \sum_{i=1}^{n}\left\{\log \pi + (u_i - 1)\log(1-\pi)\right\} - n\log\left[1 - \pi(1-\pi)^{j-1}\right]$$

let

$$= \sum_{i=1}^{n} u_i \log(1-\pi) + n\left[\log\frac{\pi}{1-\pi} - \log\left[1-\pi(1-\pi)^{j-1}\right]\right]$$

let $\log(1-\pi) = \theta$ hence $1-\pi = e^{\theta}$

and $\pi = 1 - e^{\theta}$ $\quad \theta \in (-\infty, 0]$

Thus the above likelihood can be written as

$$\mathcal{L}_n(\theta) = \theta \sum_{i=1}^{n} u_i + n \underbrace{\left[\log\left(\frac{1-e^{\theta}}{e^{\theta}}\right) - \log\left(1 - e^{\theta(j-1)}\left(1-e^{\theta}\right)\right)\right]}_{-k(\theta)}$$

$$g(y) = 0$$

<u>Note</u> The function $\pi = 1 - e^{\theta}$ is a bijection.

(iv) minimal sufficient statistic is $\sum_i u_i$

(v) Since $\pi(\theta) = 1 - e^{\theta}$ is a bijection. Thus

The Hessian of $\mathcal{L}_n(\theta)$ is $-K''(\theta)$ which is negative.

The Hessian of $\mathcal{L}_n(\pi)$ must also be negative. (since a bijection cannot change the sign). Hence $\mathcal{L}_n(\pi)$ is a concave function, which is easy to maximise using any numerical scheme (we don't need to worry about initial values).

(b) $V = \delta Y + (1-\delta)\theta$

(i) $P(V=k) = P(V=k \mid \delta=1)P(\delta=1) + P(V=k \mid \delta=0)P(\delta=0)$

$$= p\,\pi(1-\pi)^{k-1} + (1-p)\,I(\theta = k).$$

indicator variable

(ii) The log-likelihood is

$$\mathcal{L}(\pi, p, \theta) = \sum_{i=1}^{n} \log\left\{ p\,\pi(1-\pi)^{V_i - 1} + (1-p)\,I(\theta = V_i)\right\}.$$

(iii) Suppose θ is known, then we try to obtain good initial values.

From part (a) we obtained the likelihood of $Y_i \mid Y_i \neq \theta$;

Define $\{u_i \neq \theta\}_{i=1}^{m}$ all Y_i where $Y_i \neq \theta$.

Then construct likelihood

$$\mathcal{L}_n(\pi; u \neq \theta) = \sum_{i=1}^{m} \left\{ u_i \mid \theta) \log(1-\pi) \right\} + m\left[\log\frac{\pi}{1-\pi} - \log\left[1 - \pi(1-\pi)^{\theta-1}\right]\right\}$$

let $\hat{\pi}_\theta = \arg\max_{\pi} \mathcal{L}_n(\pi; \text{not } \theta)$

[recall from (a)] this can easily be maximised].

This gives an initivalue for $\hat{\pi}_\theta$ (when θ is known).

$$P(Y=\theta) = p\pi(1-\pi)^{(\theta-1)} + (1-p)$$

Estimate

$$\hat{q}_\theta = \frac{\text{No } Y_i = \theta}{n} \quad \text{hence solve for}$$

$$\hat{p}_\theta = \underset{p}{\text{solution}} \left\{ p\hat{\pi}_\theta (1-\hat{\pi}_\theta)^{\theta-1} + (1-p) = \hat{q}_\theta \right\}.$$

This gives initial values for the maximisation.

Now use Newton-Raphson

$$\underbrace{\hat{\pi}(\theta), \hat{p}(\theta)}_{\text{maximum likelihood estimator when } \theta \text{ is known}} = \arg\max_{p,\pi} \left\{ \sum_{i=1}^{n} \log\left[p\pi(1-\pi)^{k-1} + (1-p) I(\theta=k) \right] \right\}.$$

maximum likelihood estimator when θ is known.

(iv) In part (iii) we evaluated 1 - 5

the mle for θ known. ~~Now we~~ Now we go back

to this likelihood and maximise over θ.

$$\hat{\theta} = \underset{\theta \in \mathbb{Z}^+}{\arg\max} \left\{ \sum_{i=1}^{n} \log \left[\hat{p}(\theta)\, \hat{\pi}(\theta)\, (1-\pi(\theta))^{k-1} + (1-\hat{p}(\theta))\, I(\theta = k) \right\} \right\}.$$

Thus the mle is

$$\left[\hat{\theta}, \hat{p}(\hat{\theta}), \hat{\pi}(\hat{\theta}) \right].$$

(*) Note that ~~θ~~ since $\theta \in \mathbb{Z}^+$, then θ cannot be asymptotically normal.

Question 2

(2-1)

~~Suppose $\{X_i\}$ are iid ~~[illegible]~~ normal random variables~~

and we test

$$H_0 : \mu^2 = \sigma^2 \quad \text{vs} \quad H_A : \mu^2 \neq \sigma^2. \qquad (1)$$

We use the log-likelihood ratio statistic.

In order to derive the distribution under the null, we note that the likelihood can be reparameterised as

$$\mathcal{L}(\mu, \sigma^2) = -n \log \sigma^2 - \frac{1}{2\sigma^2} \sum_i (X_i - \mu)^2$$

$$\tilde{\mathcal{L}}(\mu, \gamma) = -n \log [\mu + \gamma]^2 - \frac{1}{2[\mu + \gamma]^2} \sum (X_i - \mu)^2.$$

using the reparameterisation the hypotheses stated in (1) is equivalent to

$$H_0 : \gamma = 0 \qquad H_A : \gamma \neq 0$$

Thus.

[illegible]

$$T_n = 2\left\{ \max_{\mu,\sigma^2} \mathcal{L}_n(\mu,\sigma^2) - \max_{\mu} \mathcal{L}_n(\mu,\mu^2)\right\} \quad (2\text{-}2)$$

$$= \underbrace{2\left\{ \max_{\mu,\gamma} \tilde{\mathcal{L}}_n(\mu,\gamma) - \max_{\mu} \tilde{\mathcal{L}}_n(\mu,0)\right\}}.$$

Now using results for the (nested) Profile likelihood under the null $\xrightarrow{D} \chi^2_{2-1} = \chi^2_1$.

However since the two quantities are the same we do not have to use the second quantity we can maximize over the 'usual' likelihood. I.e.

$$T_n = 2\left[\mathcal{L}_n(\hat{\mu},\hat{\sigma}^2) - \mathcal{L}_n(\tilde{\mu},\tilde{\mu}^2)\right]$$

$$\xrightarrow{D} \chi^2_1 \quad \text{(under the null hypothesis)}.$$

Question 3

1) $\overline{X}^2 = \left\{ \frac{\sigma \sqrt{n}}{\sqrt{n}} \underbrace{\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \right)}_{\to N(0,1)} + \mu \right\}^2$

$$= \underbrace{\frac{\sigma^2}{n} \left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \right)^2}_{O_p\left(\frac{1}{n}\right)} + \underbrace{2 \frac{\sigma}{\sqrt{n}} \mu \left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \right)}_{O_p\left(\frac{1}{\sqrt{n}}\right)} + \mu^2 \qquad (1)$$

Hence $\sqrt{n}\left[\overline{X}^2 - \mu^2\right] = 2\sigma\mu \overbrace{\left[\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \right]}^{N(0,1)} + O_p\left(\frac{1}{\sqrt{n}}\right)$

$$\to N(0, 4\sigma^2 \mu^2).$$

Note that in order for the second term in (1) to dominate the first we require $\mu \neq 0$.

If $\mu = 0$, then $\overline{X}$ will be a weighted χ^2.

(ii) $\cancel{\text{If}}$ $var(\bar{X}^2) = 4\sigma^2\mu^2/n + O\left(\frac{1}{n^{3/2}}\right)$ [by using (1)] 3- ② previous page.

$$var[s^2] = \frac{1}{(n-1)^2}\sum_{i=1}^{n} var[(X_i - \bar{X})^2]$$

& calculate out.

$$var[(X_i-\bar{X})^2] = 2[var(X_i - \bar{X})]^2$$

$$= cum_4[X_i - \bar{X}]$$

$$+ cum[(X_i-\bar{X}),(X_i-\bar{X}),(X_i-\bar{X}),(X_i-\bar{X})]$$

(the other terms we ignore since the ~~variance~~ mean is zero).

$$var(X_i - \bar{X}) = var(X_i) - 2cov(X_i,\bar{X}) + var(\bar{X})$$

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

$$cum_4[X_i - \bar{X}] = \underbrace{cum_4[X_i]}_{\kappa_4} + \underbrace{cum[X_i, X_i, X_i, \bar{X}]}_{=\frac{1}{n}\kappa_4} + \dots$$

$$+ \dots$$

$\cancel{+\kappa_4 + \frac{1}{n}\kappa_4}$

$$+ cum[\bar{X},\bar{X},\bar{X},\bar{X}].$$

Calculating $cum(\bar{X},\bar{X},\bar{X},\bar{X})$ is a bit tricky. Calculating it out should give you $\frac{\kappa_4}{n^3}$. However you can also guess that it is "small", since $\bar{X}$ is asympt. normal

Thus

$$\operatorname{cum}_4[X_i - \bar{X}] = \underbrace{\kappa_4}_{\text{main term.}} + O\left(\frac{1}{n}\right)$$

This gives

$$\operatorname{var}[(X_i - \bar{X})^2] = 2\left(\operatorname{var}[X_i - \bar{X}]\right)^2 + \operatorname{cum}_4[X_i - \bar{X}]$$

$$= 2\sigma^2 + \kappa_4 + O\left(\frac{1}{n}\right).$$

$$\operatorname{cov}[\bar{X}, s^2] = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{cov}[X_i, (X_i - \bar{X})^2]$$

Again consider

(since mean is zero)

$$\operatorname{cov}[X_i, (X_i - \bar{X})^2] = \cancel{2\{E\,\operatorname{cov}(X_i, (X_i - \bar{X}))\,E(X_i - \bar{X})}$$

$$\operatorname{cum}(X_i, X_i - \bar{X}, X_i - \bar{X})$$

$$= \underbrace{\operatorname{cum}(X_i, X_i, X_i)}_{\kappa_3} - \underbrace{\operatorname{cum}(X_i, X_i, \bar{X})}_{\frac{\kappa_3}{n}} + \dots - \underbrace{\operatorname{cum}(X_i, \bar{X}, \bar{X})}_{\frac{\kappa_3}{n^2}}.$$

Altogether this gives

$$\mathrm{var}\left[\sqrt{n}\,\bar{X}^2\right] = 4\sigma^2\mu^2 + o(1)$$

$$\mathrm{var}\left[\sqrt{n}\,S^2\right] = 2\sigma^2 + k_4 + o(1)$$

$$\mathrm{cov}\left[\sqrt{n}\,\bar{X}^2, \sqrt{n}\,S^2\right] = k_3 + o(1)$$

Th

$$\sqrt{n}\begin{bmatrix}\bar{X}^2-\mu^2\\ S^2-\sigma^2\end{bmatrix} \xrightarrow{D} N\left(0, \begin{pmatrix}4\sigma^2\mu^2 & k_3\\ k_3 & 2\sigma^2+k_4\end{pmatrix}\right).$$

(iii) Consider the statistic

$$T_n = \sqrt{n}\left(\frac{\bar{X}^2}{S^2}-1\right) = \sqrt{n}\left(\frac{\bar{X}^2-\cancel{\sigma^2}}{S^2}\right) \quad \text{(make Taylor exp of } S^{2}\text{)}$$

$$= \sqrt{n}\left\{\frac{(\bar{X}^2-\mu^2) - (S^2-\mu^2)}{\mu^2}\right\}\left\{1+\underbrace{\frac{\mu^2(S^2-\mu^2)}{\alpha S^2+(1-\alpha)\mu^2}}_{O_p(1/\sqrt{n})}\right\}$$

By part (ii)

$$= \frac{\sqrt{n}\{(\bar{X}^2-\mu^2)-(S^2-\mu^2)\}}{\mu^2} + O_p\left(\frac{1}{\sqrt{n}}\right)$$

Now we calculate the "limiting variance" using the 3- results from part (ii). 6

$$\mathrm{Var}\left\{\frac{\sqrt{n}}{\mu^2}\left[(\bar{X}^2-\mu^2)-(s^2-\mu^2)\right]\right\}$$

$$= \frac{1}{\mu^4}\left\{\Sigma_{11}-2\Sigma_{12}+\Sigma_{22}\right\} \quad \text{(General)}$$

$$= \frac{1}{\mu^4}\left\{4\sigma^2\mu^2-2\kappa_3+2\sigma^2+\kappa_4\right\}. \quad \text{(using results)}.$$

Thus if $\mu \neq 0$ and under the null

$$\sqrt{n}\left(\frac{\bar{X}^2}{s^2}-1\right) \xrightarrow{D} N\left(0, \frac{1}{\mu^4}\left[4\sigma^2\mu^2-2\kappa_3+2\sigma^2+\kappa_4\right]\right).$$

(iv) Note that $\frac{\bar{Y}^2}{s^2}$ are positive random variables hence their finite same distribution will be right skewed. This will mean that the normal approximation will not be particular good for small samples.

[If we measured the skewness by using the third order cumulant it would be of order $O(1/\sqrt{n})$].

If $\{X_i\}$ are iid normal random variables then $\bar{X}$ and s^2 are independent and $\frac{\bar{Y}^2}{s^2}$ follows an F-distribution.

Power calculation for test statistic

$$T_n = \sqrt{n}\left\{\frac{\bar{X}}{S^2} - 1\right\} = \sqrt{n}\left\{\frac{\bar{X}}{S^2} - \frac{\mu^2}{(\mu+\theta)^2}\right\}$$

$$+ \underbrace{\sqrt{n}\left\{\frac{\mu^2}{(\mu+\theta)^2} - 1\right\}}.$$

Let

$$\sqrt{n}\left\{\frac{(\mu+\theta)^2 - 2\mu\theta + \theta^2}{(\mu+\theta)^2} - 1\right\}$$

$$= \sqrt{n}\left\{\frac{\mu\theta[-2+\theta]}{(\mu+\theta)^2}\right\}$$

Then if θ is kept constant as $\sqrt{n} \to \infty$ the power of test $\to 100\%$.

To assess power set $\theta = \phi/\sqrt{n}$

Then $$\sqrt{n}\left\{\frac{\mu^2}{(\mu+\theta)^2} - 1\right\} \simeq \frac{-2\mu\phi}{(\mu + \phi/\sqrt{n})^2} \simeq \frac{-2\phi}{\mu}$$

Let $$\tilde{\sigma}^2 = \frac{1}{\mu^4}\left[4\sigma^2\mu^2 - 2k_3 + 2\sigma^2 + k_4\right].$$

Then under the null

$$\frac{T_n}{\tilde{\sigma}} \longrightarrow N(0, 1)$$

Suppose we do the test at the 5% level.

Then under the alternative of $\theta = \phi/\sqrt{n}$ $(\phi > 0)$ (3-7)

$$P\left\{\left|\frac{T_n}{\tilde{\sigma}}\right| > z_{1-\alpha/2}\right\}$$

$$> P\left\{\frac{T_n}{\tilde{\sigma}} < \cancel{0} - z_{1-\alpha/2}\right\} = P\left\{\underbrace{\frac{\sqrt{n}}{\tilde{\sigma}}\left[\frac{\bar{X}}{S^2} - \frac{\mu^2}{(\mu+\phi/\sqrt{n})}\right]}_{\simeq Z} - \frac{2\phi}{\mu\tilde{\sigma}} \cancel{<} z_{1-\alpha/2}\right\}$$

$$+ \cancel{\sqrt{n}}$$

$$= P\left\{Z \cancel{<} - z_{1-\alpha/2} + \frac{2\phi}{\mu\tilde{\sigma}}\right\}$$

$$= \Phi\left[-z_{1-\alpha/2} + \frac{2\phi}{\mu\tilde{\sigma}}\right] \rightarrow \text{power function}$$

which grows as $\phi \to \infty$.

Power Transforms

How to make $\frac{\bar{X}^2}{s^2}$ more normal?

•

One method is to do a power transform of the Type:

$$0 < \lambda < 1$$

$$\frac{\left(\frac{\bar{X}^2}{s^2}\right)^{\lambda} - 1}{\lambda} \quad \text{if} \quad \lambda \neq 0 \qquad \boxed{\text{Box-Cox transform}}$$

$$\log\left(\bar{X}^2/s^2\right) \quad \text{if} \quad \lambda = 0$$

However one needs to calculate the mean and variance of this (under the null); This can be done using the delta meth.

• Another method is to do a Bootstrap; in the hope of capturing the skewness.

One can prove that the power transform tends to reduce the skewness in the distribution.