Carson Slater
5/3/2024

Dr. Harvill
Theory 2

Final Exam

1) Given $X_1, \dots, X_n \overset{iid}{\sim} \text{Geometric}(p)$. The pmf is

$$f(x|p) = p(1-p)^{x-1}, \; x = 1, 2, \dots, \; 0 < p < 1.$$

Also, $E[X] = \frac{1}{p}$, $Var[X] = \frac{1-p}{p^2}$

a) To find the method of moments estimator (MOM) for p, $\tilde{p}$, we can set the first population moment ($E[X]$) equal to the first sample moment (M_1). So then

$$\frac{1}{p} = E[X] = M_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\Rightarrow \tilde{p} = \frac{n}{T}, \text{ where } T = \sum_{i=1}^{n} X_i. \; \square$$

alternatively written as $\tilde{p} = \frac{1}{\bar{X}}$.

b) To find the maximum likelihood estimator (MLE) for p, we derive the likelihood function, $L(p|\underline{x})$.

$$L(p|\underline{x}) = \prod_{i=1}^{n} f(x_i|p) = \prod_{i=1}^{n} p(1-p)^{x_i - 1}.$$

$$= p^n (1-p)^{\sum_{i=1}^{n} x_i - n}.$$

We can maximize $\log(L(p|\underline{x})) = \ell(p|\underline{x})$ to find $\hat{p}$.

$$\frac{\partial}{\partial p}\ell(p|\underline{x}) = \frac{\partial}{\partial p}\left(n\log p + \sum_{i=1}^{n} x_i \log(1-p) - n\log(1-p)\right)$$

$$\Rightarrow 0 \overset{set}{=} \frac{n}{p} + \frac{\sum_{i=1}^{n} x_i - n}{1-p} \Rightarrow 0 = \frac{1}{p} + \frac{\bar{x} - 1}{p}$$

$$\Rightarrow \frac{1-\bar{x}}{1-p} = \frac{1}{p} \Rightarrow 1-\bar{x} = \left(\frac{1-p}{p}\right)\frac{\frac{1}{p}}{\frac{1}{p}}$$

"well chosen 1"

$$\Rightarrow 1-\bar{x} = \frac{1}{p} - 1 \Rightarrow \hat{p} = \frac{1}{\bar{x}}. \quad \square$$

Concavity

$$\frac{\partial^2}{\partial p^2}\ell(p|\underline{x}) = -\frac{n}{p^2} + \frac{\sum_{i=1}^n x_i - n}{(1-p)^2}$$

$$= \frac{n - 2np - np^2 + p^2\sum_{i=1}^n x_i - np^2}{p^2(1-p)^2}$$

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

c) We have that $Y = P(X=3) = p(1-p)^2$. By the invariance property of the MLE, we know that the MLE of Y is $\hat{Y} = \hat{p}(1-\hat{p})^2$. So then the MLE of

$$Y = \frac{1}{\bar{x}}\left(1 - \frac{1}{\bar{x}}\right)^2. \quad \square$$

d) Given properties of full exponential families of the form,

$$f(x|\theta) = h(x)c(\theta)e^{\omega(\theta)t(x)}$$

we know that $\sum_{i=1}^n t(x_i)$ is a complete sufficient statistic. So we can show that a Geo(p) distribution is an exponential family. For $X \sim \text{Geo}(p)$,

multiplied by "well chosen 1"

$$f(x|p) = I_{\{1,2,\dots\}}(x)\frac{p}{1-p}\exp\{(x)\log(1-p)\}$$

$$= h(x)c(p)\exp\{t(x)\omega(x)\}$$

where $h(x) = I_{\{1,2,\dots\}}(x)$, $c(p) = \frac{p}{1-p}$, $t(x) = x$, and $\omega(p) = \log(1-p)$.

So the Geo(p) is an exponential family.

Hence, $\sum_{i=1}^{n} t(x_i) = \sum_{i=1}^{n} x_i$ is a complete, sufficient statistic for p. $\square$

e) A statistic $T(\underline{x}) = \sum_{i=1}^{n} x_i$ is minimally sufficient for p if for every two sample points $\underline{x}$ and $\underline{y}$, $\frac{f(\underline{x}|p)}{f(\underline{y}|p)}$ is constant as a function of p if and only if $T(\underline{x}) = T(\underline{y})$. So

$$\frac{f(\underline{x}|p)}{f(\underline{y}|p)} = \frac{p^n(1-p)^{T(\underline{x})-n}}{p^n(1-p)^{T(\underline{y})-n}} = \frac{(1-p)^{T(\underline{x})}}{(1-p)^{T(\underline{y})}} = (1-p)^{T(\underline{x})-T(\underline{y})}.$$

is constant as a function of p iff $T(\underline{x}) = T(\underline{y})$.

$\Longrightarrow$ $T(\underline{x}) = \sum_{i=1}^{n} x_i$ is a minimally sufficient, complete statistic.

f) Given sufficient statistic $T(\underline{x})$ for p and unbiased estimator $W(\underline{x})$ for p, the Rao-Blackwell Theorem states that estimator $\phi(T) = E_p[W|T]$ is the UMVUE of p. So consider $W(\underline{x}) = \frac{1}{\bar{x}}$. And then $E[W] = E\left[\frac{1}{\bar{x}}\right] = E\left[\frac{n}{\sum_{i=1}^{n} x_i}\right] = \frac{n}{\sum_{i=1}^{n} E[x_i]} = \frac{n}{\sum_{i=1}^{n} \frac{1}{p}} = \frac{np}{n} = p.$

So we have that $W(\underline{x})$ is unbiased and $T(\underline{x}) = \sum_{i=1}^{n} x_i$ is sufficient for p. By the Rao-Blackwell Theorem, we have that

$$\phi(T) = E_p\left[\frac{1}{X} \,\middle|\, \sum_{i=1}^{n} X_i = t\right]$$

$$= E_p\left[\frac{1}{X} \,\middle|\, \bar{X} = \frac{t}{n}\right] \longrightarrow \text{Any function of a sufficient statistic is sufficient.}$$

$$= \sum_{x=1}^{\infty} \frac{n}{t} p(1-p)^{x-1} = \frac{n}{t} \underbrace{\sum_{x=1}^{\infty} p(1-p)^{x-1}}_{\substack{\text{pmf of geometric dist over} \\ \text{its support} \\ = 1}}$$

$$= \frac{n}{t},$$

where $\sum_{i=1}^{n} X_i = t$ is the UMVUE of p.

2) Let $X_1, \ldots, X_n \sim \text{Gamma}(\eta, \theta)$, $\eta > 0$ and $\theta > 0$, where θ is the scale. Given $T(\underline{x}) = \sum_{i=1}^{n} x_i$ is sufficient for θ and the MLE of θ, $\hat{\theta} = \frac{\bar{X}}{\eta}$. We consider testing

$$H_0: \theta \le \theta_0 \quad \text{and} \quad H_a: \theta > \theta_0.$$

a) The likelihood ratio test first involves finding the likelihood ratio test statistic, $\lambda(\underline{x})$. We have that $\lambda(\underline{x}) = \dfrac{L(\theta_0 | \underline{x})}{L(\hat{\theta} | \underline{x})}$, where $L(\theta | \underline{x}) = \prod_{i=1}^{n} \frac{1}{\Gamma(\eta)\theta^{\eta}} x_i^{\eta - 1} e^{-x_i/\theta}$

$$\Rightarrow \frac{L(\theta_0|\underline{x})}{L(\hat{\theta}|\underline{x})} = \frac{\prod_{i=1}^{n} \frac{1}{\Gamma(\eta)\theta_0^{\eta}} x_i^{\eta-1} e^{-x_i/\theta_0}}{\prod_{i=1}^{n} \frac{1}{\Gamma(\eta)\hat{\theta}^{\eta}} x_i^{\eta-1} e^{-x_i/\hat{\theta}}} = \frac{\frac{1}{\theta_0^{\eta n}} \cancel{\left(\prod_{i=1}^{n} x_i^{\eta-1}\right)} e^{-\frac{1}{\theta_0}\sum_{i=1}^{n} x_i}}{\frac{1}{\hat{\theta}^{\eta n}} \cancel{\left(\prod_{i=1}^{n} x_i^{\eta-1}\right)} e^{-\frac{1}{\hat{\theta}}\sum_{i=1}^{n} x_i}} = \left(\frac{\hat{\theta}}{\theta_0}\right)^{\eta n} \exp\left\{\sum_{i=1}^{n} x_i \left(\frac{1}{\hat{\theta}} - \frac{1}{\theta_0}\right)\right\}$$

$$= \left(\frac{\hat{\theta}}{\theta_0}\right)^{\eta n} \exp\left\{\sum_{i=1}^{n} x_i \left(\frac{1}{\hat{\theta}} - \frac{1}{\theta_0}\right)\right\} = \left(\frac{\hat{\theta}}{\theta_0}\right)^{\eta n} \exp\left\{\eta n \left(1 - \frac{\hat{\theta}}{\theta_0}\right)\right\}$$

$\eta n \hat{\theta} = \sum_{i=1}^{n} x_i$

So $$\lambda(\underline{x}) = \begin{cases} \left(\frac{\hat{\theta}}{\theta_0}\right)^{\eta n} \exp\left\{\eta n \left(1 - \frac{\hat{\theta}}{\theta_0}\right)\right\} & \text{if } \theta_0 < \hat{\theta} \\ 1 & \text{if } \theta_0 \ge \hat{\theta}. \end{cases}$$

The LRT is the test that rejects H_0 for some value $0 \le c \le 1$ such that $\{\underline{x} : \lambda(\underline{x}) \le c\}$.

For this to be a size α test we need to choose constant c such that c satisfies

$P(\lambda(\underline{x}) \le c) = \alpha$. $\square$

b)

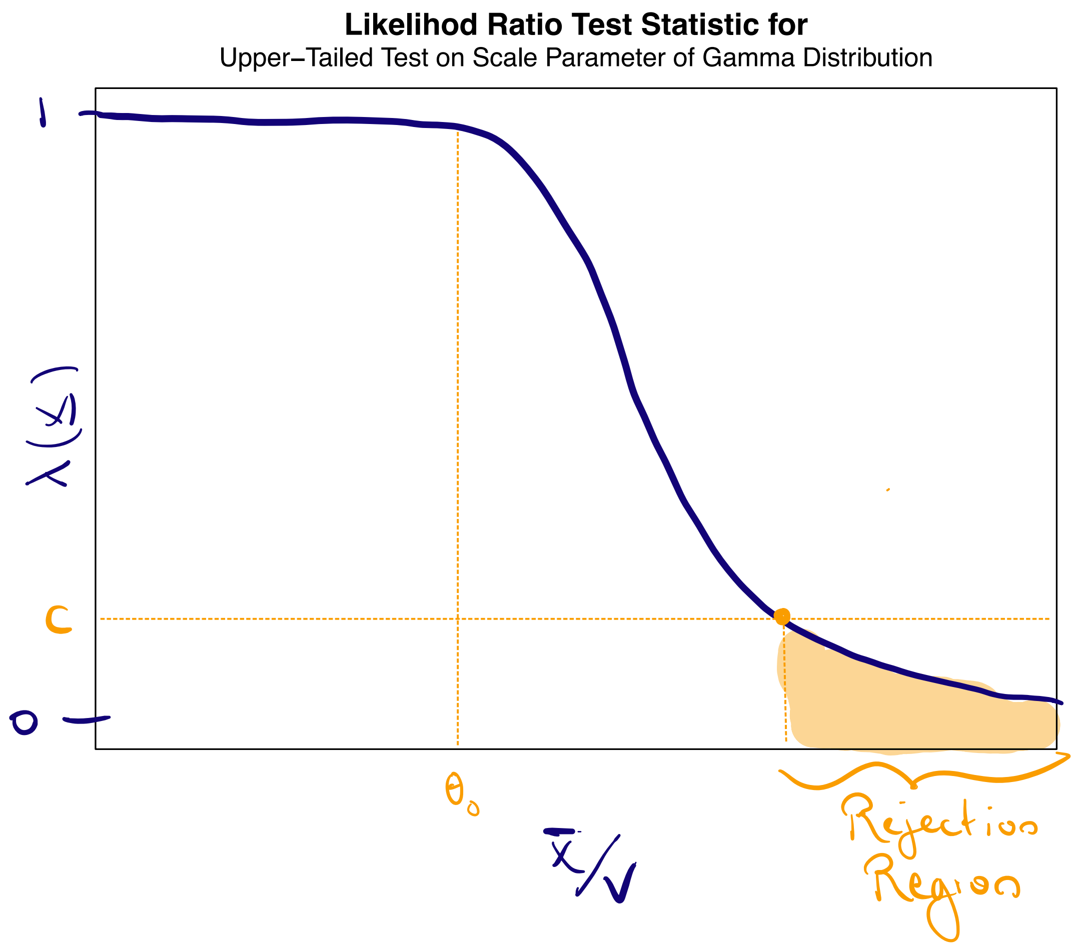

Figure 1: Axes for graph of likelihood ratio test statistic for an upper-tailed test of the gamma scale parameter.

c) The Karlin Rubin Theorem states that given a sufficient statistic $T(\underset{\sim}{x})$ whose sampling distribution has monotone likelihood ratio, the UMP level α test is the that rejects H_0 iff $T(\underset{\sim}{x}) > t_0$ such that $P(T(\underset{\sim}{x}) \leq t_0) = \alpha$. We are given that $T_1(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i$ is a sufficient statistic for θ, and since every function of a sufficient statistic is also sufficient, we have that $T_2(\underset{\sim}{x}) = \frac{\sum_{i=1}^{n} x_i}{n\eta} = \frac{\bar{x}}{\eta} \sim \text{Gamma}\left(n\eta, \frac{\theta}{\eta n}\right)$ is also sufficient, and its sampling distribution is an exponential family and has monotone likelihood ratio. So by the Karlin-Rubin theorem, the test that rejects H_0 is $\frac{\bar{x}}{\eta} > K\theta_0$, where K is chosen to satisfy $P(\frac{\bar{x}}{\eta} > K\theta_0) = \alpha$. $\square$

d) A valid p-value $p(\underset{\sim}{x})$ is such that for every $\theta \leq \theta_0$ and every $\alpha \in [0,1]$, that $P\left(p(\underset{\sim}{x}) \leq \alpha\right) \leq \alpha$. So for this test, a valid p-value is $P(\frac{\bar{x}}{\eta} \geq \theta_0 K)$, where K is the α upper-tail quantile of the gamma$\left(\eta n, \frac{\theta_0}{\eta n}\right)$ under H_0. If $\theta > \theta_0$, then $P\left(\frac{\bar{x}}{\eta\theta} > K\right) \leq P\left(\frac{\bar{x}}{\eta\theta_0} > K\right)$. Which follows the behavior we are looking for.

e) The power function is the probability of rejecting the null hypothesis. Recall $\frac{\bar{X}}{\theta\eta} \sim \text{Gamma}(n\eta, \frac{1}{\eta n})$

$$\beta(\theta) = P(\underline{X} \in R) = P\left(\frac{\bar{X}}{\eta} > \theta_0 k\right)$$

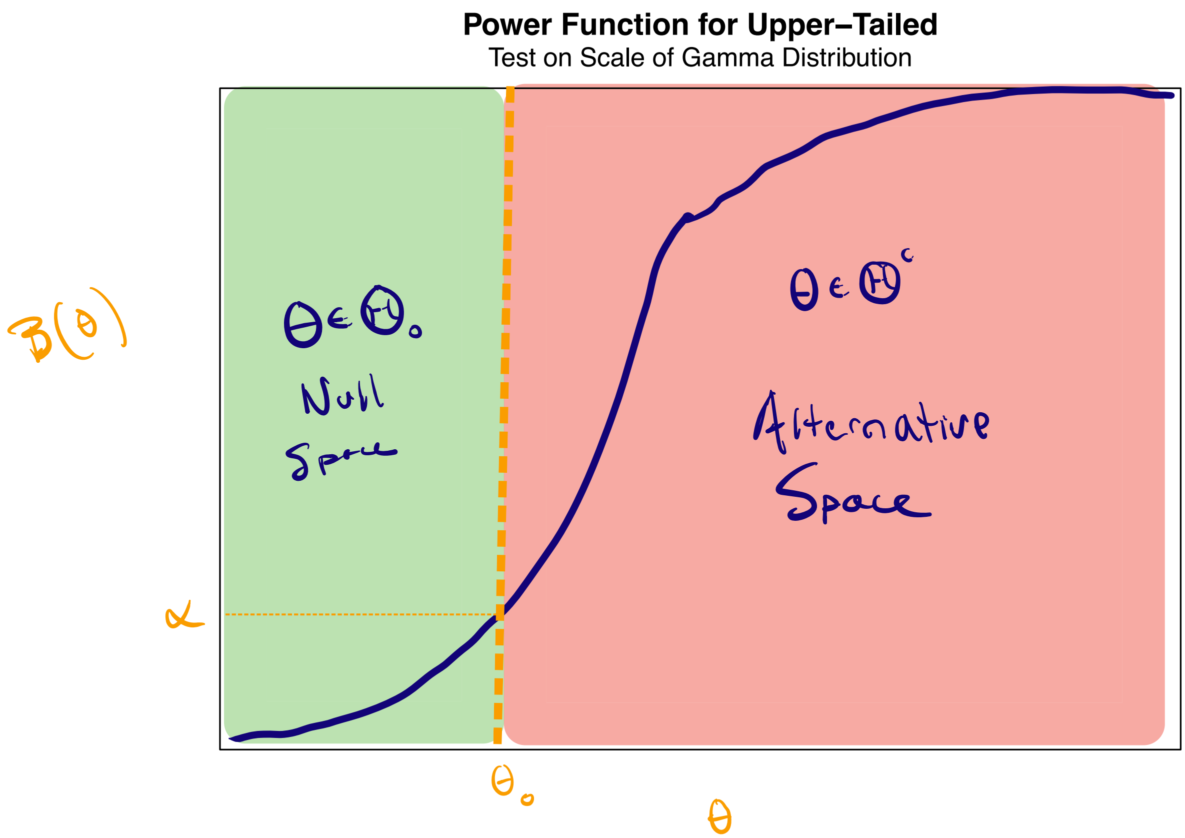

Figure 2: Axes for graph of the power function of the UMP upper-tailed test of the gamma scale parameter.

f) The $1-\alpha$ uniformly most accurate (UMA) confidence interval can be found by inverting the acceptance region of the UMP level α test. Here we know the acceptance region

$$A(\theta_0) = \left\{ \underline{x} : \frac{\bar{x}}{\eta} \leq \theta_0 K \right\}.$$

Inverting this into $1-\alpha$ a confidence set, we can write that

$$C(\underline{x}) = \left\{ \theta : \theta \geq \frac{\bar{x}}{\eta K} \right\}$$

where K is the α upper-tail quantile for the $\text{gamma}\left(\eta n, \frac{1}{\eta n}\right)$ distribution. So $C(\underline{x}) = \left[\frac{\bar{x}}{\eta K}, \infty\right]$.

g) Consider, for forming a Wald test, let statistic $T_n = \frac{\bar{X}_n}{\eta}$. Recall $\frac{\bar{X}}{\eta}$ is the MLE for θ, and since a gamma population is an exponential family, it satisfies regularity conditions, and is asymptotically consistent, and efficient. So

$$\sqrt{n}\left(\frac{\bar{X}_n}{\eta} - \theta\right) \xrightarrow{D} N(0, \sigma^2)$$

where σ^2 is the delta method variance of $\frac{\bar{X}_n}{\eta}$.

So

$$\sigma^2 = \lim_{n\to\infty} n \operatorname{Var}\left(\frac{\bar{X}}{\eta}\right) = \lim_{n\to\infty} \frac{n}{\eta^2} \operatorname{Var}(\bar{x}) = \frac{\eta\theta^2}{\eta^2} = \frac{\theta^2}{\eta}.$$

where η is known (given at beginning).

Since $\frac{\theta}{\sqrt{\eta}}$ is unknown, we must show that

$\frac{\theta^2}{\eta} \Big/ \frac{\hat{\theta}^2}{n} = \frac{\theta^2}{\hat{\theta}^2} \xrightarrow{P} 1$. We can say this because of the consistency of the MLE of θ, $\hat{\theta}$. Also by the invariance property of the MLE, we then have that $\frac{\theta}{\sqrt{\eta}} \Big/ \frac{\hat{\theta}}{\sqrt{n}} = \frac{\theta}{\hat{\theta}} \xrightarrow{P} 1$. Now, by Slutsky's Theorem

$$\sqrt{n}\left(\frac{\frac{\bar{X}}{\eta} - \theta}{\sqrt{\hat{\theta}^2/n}}\right) \xrightarrow{D} N(0,1).$$

So, under H_0, a Wald Test statistic is $\sqrt{n}\left(\frac{\frac{\bar{X}}{\eta} - \theta}{\sqrt{\hat{\theta}^2/n}}\right)$, and the α level Wald test rejects H_0 if and only if $\sqrt{n}\left(\frac{\frac{\bar{X}}{\eta} - \theta}{\sqrt{\hat{\theta}^2/n}}\right) > z_\alpha$, where z_α is the upper tail α quantile of the standard normal disbribution.

h) The acceptance region of the Wald test is

$$A(\bar{x}) = \left\{ \underline{x} : \sqrt{n}\left(\frac{\frac{\bar{X}}{\eta} - \theta}{\sqrt{\hat{\theta}^2/n}}\right) < z_\alpha \right\}.$$

Inverting this yields the $1-\alpha$ confidence region

$$C(\theta) = \left\{ \theta : \theta > \frac{\bar{X}}{\eta} - z_\alpha \frac{\bar{X}}{\eta^{3/2}\sqrt{n}} \right\} \quad \square$$

As $\frac{\hat{\theta}}{\sqrt{\eta n}} = \frac{\bar{X}}{\eta} \cdot \frac{1}{\sqrt{\eta n}}$

3) Given $Y_1, \ldots, Y_n \overset{iid}{\sim}$ Gamma(η, ω), where shape $\eta > 0$ and rate $\omega > 0$. We impose a gamma(α, β) prior on ω.

a) To find $\pi(\omega)$, we consider that

$$\pi(\omega \mid \underline{y}) = \frac{f(\underline{y} \mid \omega)\pi(\omega)}{m(\underline{y})}.$$

We know the joint pdf of $\underline{y} \mid \omega$ is

$$f(\underline{y} \mid \omega) = \left(\frac{\omega^{\eta}}{\Gamma(\eta)}\right)^{n} \prod_{i=1}^{n} \left(y_i^{\eta-1}\right) e^{-\sum_{i=1}^{n} y_i \omega}$$

So then

$$\pi(\omega \mid \underline{y}) \propto \left(\frac{\omega^{\eta}}{\Gamma(\eta)}\right)^{n} \left(\prod_{i=1}^{n} y_i\right)^{\eta-1} e^{-\sum_{i=1}^{n} y_i \omega} \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)} \omega^{\alpha-1} e^{-\beta\omega}\right)$$

$$= \frac{\beta^{\alpha}\left(\prod_{i=1}^{n} y_i\right)^{\eta-1}}{\Gamma(\eta)^{n}\Gamma(\alpha)} \omega^{\eta n+\alpha-1} e^{-(\beta+\sum_{i=1}^{n} y_i)\omega}$$

$$\propto \frac{(\beta+\sum_{i=1}^{n} y_i)^{\eta n+\alpha-1}}{\Gamma(\eta n+\alpha)} \omega^{\eta n+\alpha-1} e^{-(\beta+\sum_{i=1}^{n} y_i)\omega} \quad \square$$

So $\omega \mid \underline{y} \sim$ gamma$(\alpha+\eta n, \beta+\sum_{i=1}^{n} y_i)$.

b) Under Bayes rule for absolute error loss, the Bayes estimator for ω is the median of the posterior distribution, which is the value ξ which satisfies

$$\int_0^{\xi} \frac{(\beta+\sum_{i=1}^{n} y_i)^{\eta n+\alpha-1}}{\Gamma(\eta n+\alpha)} \omega^{\eta n+\alpha-1} e^{-(\beta+\sum_{i=1}^{n} y_i)\omega} \, d\omega = \frac{1}{2}. \quad \square$$

c) Under Bayes rule for squared error loss the Bayes estimator for ω is the mean of the posterior distribution, which is $(\alpha+\eta n)(\beta+\sum_{i=1}^{\hat{n}} y_i)$ $\square$

d) The Bayes test of $H_0: \omega \leq \omega_0$ versus $H_1: \omega > \omega_0$ is the the test that accepts H_0 if and only if

$$P(\omega \leq \omega_0 \mid \underline{y}) \geq \frac{1}{2}.$$

Since the posterior is not symmetric about its mean the test will accept if $\omega_0 \geq \xi$, the population median.

(In this problem, $0 \leq \gamma \leq 1$ is what I used for what α normally is)

e) We know $\omega \mid \underline{y} \sim \text{gamma}(\alpha+\eta n, \beta+\sum_{i=1}^{\hat{n}} y_i)$. Consider the pivot $\frac{2\omega}{\beta+\sum_{i=1}^{\hat{n}} y_i} \sim \chi^2_{\alpha+\eta n}$. From here the shortest $1-\gamma$ credible set is

$$\chi^2_{\alpha+\eta n, \frac{1-\gamma}{2}} \leq \frac{2\omega}{\beta+\sum_{i=1}^{\hat{n}} y_i} \leq \chi^2_{\alpha+\eta n, \frac{\gamma}{2}},$$

which can be rewritten as

$$\frac{2\chi^2_{\alpha+\eta n, \frac{1-\gamma}{2}}}{\beta+\sum_{i=1}^{\hat{n}} y_i} \leq \omega \leq \frac{2\chi^2_{\alpha+\eta n, \frac{\gamma}{2}}}{\beta+\sum_{i=1}^{\hat{n}} y_i} \quad \square$$

where $\chi^2_{\alpha+\eta n, \_}$ are respective quantiles from the $\chi^2_{\alpha+\eta n}$ distribution.

4) Given $X_1, X_2, \dots$ be iid with $\mu = E[X_i]$ and $Var[X_i] = \sigma^2$ where $\sigma^2 < \infty$. Consider

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Chebychev's inequality states that

$$P(g(S_n) \geq \varepsilon) \leq \frac{E[g(S_n)]}{\varepsilon}.$$

By Chebychev's Inequality, we can write that $g(S_n) = (S_n^2 - \sigma^2)$, so

$$P(|S_n^2 - \sigma^2| \leq \varepsilon) = P((S_n^2 - \sigma^2)^2 \leq \varepsilon^2) \leq \frac{E[(S_n^2 - \sigma^2)^2]}{\varepsilon^2} = \frac{Var(S_n^2)}{\varepsilon^2}$$

taking the limit as $n \to \infty$, we get

$$\lim_{n\to\infty} P(|S_n - \sigma^2| < \varepsilon) \leq \lim_{n\to\infty} \frac{Var(S_n^2)}{\varepsilon^2}. \quad (1)$$

If $\lim_{n\to\infty} Var(S_n^2) = 0$, then (1) is the definition of S_n^2 converging in probability to σ^2, which makes it a consistent estimator. So to be consistent, $\lim_{n\to\infty} Var(S_n^2) \to 0$.

5) Given $W_1, W_2, \ldots$ be iid Bernoulli(p) draws where $p = P(W_i = 1)$, $0 < p < 1$ for each $i = 1, 2, \ldots$ We know the maximum likelihood estimator for p, $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} W_i$.

a) We have that

$$\lim_{n\to\infty} Var(\hat{p}_n) = \lim_{n\to\infty} \frac{1}{n^2} Var\left(\sum_{i=1}^{n} W_i\right) = \lim_{n\to\infty} \frac{p(1-p)}{n} = 0.$$

Choose $k_n = n$. Now the limiting variance of $\hat{p}_n$ is

$$\lim_{n\to\infty} n\, Var(\hat{p}_n) = \frac{1}{n} Var\left(\sum_{i=1}^{n} W_i\right) = p(1-p). \quad \square$$

b) i) Consider the odds $\tau(p) = \frac{p}{1-p}$. From the invariance property of the MLE, we know that the MLE for the odds, $\hat{\tau} = \tau(\hat{p})$, written as

$$\tau(\hat{p}) = \frac{\hat{p}_n}{1-\hat{p}_n} = \frac{\sum_{i=1}^{n} W_i}{n - \sum_{i=1}^{n} W_i}.$$

ii) For estimator $\hat{p}$, with function $\tau(\hat{p})$, we have that, by the first order delta method,

$$\sqrt{n}\left(\tau(\hat{p}_n) - \tau(p)\right) \xrightarrow{D} N\left(0, \sigma^2(\tau'(p))^2\right)$$

where $\sigma^2 = p(1-p)$. Here $\tau'(p) = \frac{1}{(1-p)^2}$, then $(\tau'(p))^2 = \frac{1}{(1-p)^4}$. So then

$$\sqrt{n}\left(\tau(\hat{p}_n) - \tau(p)\right) \xrightarrow{D} N\left(0, \frac{1}{(1-p)^3}\right)$$

"approximately distributed as"

which implies asymptotically, $\frac{\hat{p}_n}{1-\hat{p}_n} \approx N\left(\frac{p}{1-p}, \frac{1}{n(1-p)^3}\right)$.

c) Consider testing $H_0: p = p_0$ vs. $H_1: p_0 \neq 0$. We are given the likelihood ratio test statistic

$$\lambda(\hat{p}_n) = \left(\frac{p_0}{\hat{p}_n}\right)^{n\hat{p}_n} \left(\frac{1-p_0}{1-\hat{p}_n}\right)^{n(1-\hat{p}_n)}.$$

Given from test.

i) The likelihood ratio test is the test that rejects H_0 when $\lambda(\hat{p}_n) < c \implies -2\log(\lambda(\hat{p}_n)) \geq -2\log c$.

Consider that as $n \to \infty$, $-2\log(\lambda(\hat{p}_n)) \xrightarrow{D} \chi^2_1$.

So choose c such that $-2\log c$ is the $\chi^2_{1,\alpha}$ upper α level quantile (area α falls to right of quantile.

ii) The finite sample is always preffered since the large sample is an approximation. If the finite sample test is computable, it is the better.

iii) To find the score test, we consider the score function $S(p) = \frac{\partial}{\partial p}\mathcal{L}(p|\underline{w})$, where $\mathcal{L}(p|\underline{x})$ is the log-likelihood function for the Bernoulli distribution.

So

$$S(p) = \frac{\partial}{\partial p}\log\left(\prod_{i=1}^{n} p^{w_i}(1-p)^{1-w_i}\right) = \frac{\partial}{\partial p}\left(\sum_{i=1}^{n} w_i \log(p) + \left(n - \sum_{i=1}^{n} w_i\right)\log(1-p)\right)$$

$$= \frac{\sum_{i=1}^{n} w_i}{p} + \frac{n - \sum_{i=1}^{n} w_i}{1-p} = \frac{n\hat{p}_n}{p} + \frac{n(1-\hat{p})}{1-p}.$$

Also the Variance of the Score test is the Fischer information number since the Bernoulli(p) family can be written as an exponential family.

$$f(w|p) = p^{w}(1-p)^{1-w} I_{(0,1)}(w) = h(w)c(p)\exp\{T(w)\omega_1(p) + \omega_2(p)\}$$

where $h(w) = I_{(0,1)}(w)$, $c(p) = 1$, $T(w) = x$, $\omega_1(p) = \log\left(\frac{p}{1-p}\right)$, $\omega_2(p) = \log(1-p)$.

So then $I_n = -E\left[\frac{\partial^2}{\partial p^2}\ell(p|\underline{w})\right]$

$$= -E\left[\frac{\partial}{\partial p}\left(\frac{\sum_{i=1}^{n} w_i}{p} + \frac{n - \sum_{i=1}^{n} w_i}{1-p}\right)\right]$$

$$= -E\left[-\frac{\sum_{i=1}^{n} w_i}{p^2} + \frac{n - \sum_{i=1}^{n} w_i}{(1-p)^2}\right]$$

$$= -\left(-\frac{\sum_{i=1}^{n} E[w_i]}{p^2} + \frac{n - \sum_{i=1}^{n} E[w_i]}{(1-p)^2}\right)$$

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

$$= \frac{n}{p} - \frac{n(1-p)}{(1-p)^2} = \frac{n}{p(1-p)}$$

The score statistic for H_0 is $Z = \frac{S(p_0)}{\sqrt{I_n(p_0)}}$

So

$$Z_s = \sqrt{n}\left(\frac{\frac{n\hat{p}_n}{p} + \frac{n(1-\hat{p})}{1-p}}{\sqrt{p(1-p)}}\right).$$

If $|Z_s| > z_{\alpha/2}$, then the decision is to reject H_0. That is the α level Score test.

c) Given variables Y_{ij} are observed according to the model $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, we have that there are $k+1$ parameters $(\mu, \tau_1, \ldots, \tau_k)$ but only k means $(E[Y_{1j}], \ldots, E[Y_{kj}])$. This is an overparameterized model. Without loss of generality, suppose $k = 2$, where the parameters of the model are μ, τ_1, τ_2. Then the set of parameters $(\mu = 1, \tau_1 = 2, \tau_2 = -1)$ yield the distribution $E[Y_{ij}] = \begin{cases} 3 & \text{if } i = 1 \\ 0 & \text{if } i = 2 \end{cases}$. Consider the parameter vector $(\mu = -2, \tau_1 = 5, \tau_2 = 2)$. Then the distribution of $E[Y_{ij}] = \begin{cases} 3 & \text{if } i = 1 \\ 0 & \text{if } i = 2 \end{cases}$ also, which implies the treatment means model is not identifiable. Different parameters yield the same distributions.

7) Considering the cell means ANOVA model

$$Y_{ij} = \theta_i + \varepsilon_{ij},\ i=1,2,\dots k,\ j=1,2,\dots,n_i.$$

where $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, where $\sigma^2 < \infty$.

a) The classic ANOVA hypothesis is:

$H_0: \theta_1 = \theta_2 = \dots = \theta_k$

versus

H_1: At least one of the cell means is different from all the other cell means.

b) Consider $\bar{Y}_{i.}$. Under the assumptions of ANOVA, then $Y_{ij} \overset{iid}{\sim} N(\theta_i, \sigma^2)$, assuming homoscedasticity. So then $\bar{Y}_{i.}$ is a linear combination of n_i normal random variables, and so is normal as well. Again, consider $\bar{Y}_{..} = \sum_{i=1}^{k} \bar{Y}_{i.}$ is also implies $\bar{Y}_{..}$ is a linear combination of normally disbributed $\sum_{i=1}^{k} n_i$ random variables. So $(\bar{Y}_{i.} - \bar{Y}_{..})$, $i=1,2,\dots,k$ is furthermore also a linear combination of $\left(\sum_{l=1}^{k} n_l\right) - n_i$ normal random variables and is also normally distributed. Consider then, the sampling distribution of $\bar{Y}_{i.} \sim N(\theta, \frac{\sigma^2}{n_i})$ under H_0 and the sampling distribution of $Y_{ij} \sim N(\theta, \sigma^2)$. let $\bar{Y}_{..}$ be an unbiased estimator for θ, ($\theta = \theta_1, \ldots, \theta_k$ under H_0). Then by the central limit theorem, $\left(\frac{Y_{ij} - \bar{Y}_{..}}{\sigma}\right) \sim N(0,1)$ and $\left(\frac{\bar{Y}_{i.} - \bar{Y}_{..}}{\sigma/\sqrt{n}}\right) \sim N(0,1)$. So consider:

$$\frac{MSB}{MSW} \cdot \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2}} = \frac{SSB/\sigma^2(k-1)}{SSW/\sigma^2(n-k)} = \frac{\sum_{j=1}^{k} \frac{n_i}{\sigma^2}(\bar{Y}_{i.} - \bar{Y}_{..})^2 / k-1}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} \frac{1}{\sigma^2}(Y_{ij} - \bar{Y}_{..}) / n-k}$$

$$= \frac{\sum_{i=1}^{k} \left(\frac{\bar{Y}_{i.} - \bar{Y}_{..}}{\sigma/\sqrt{n}}\right)^2 / k-1}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\frac{Y_{ij} - \bar{Y}_{..}}{\sigma^2}\right)^2 / n-k},$$

where by Cochran's Theorem:

$$\sum_{i=1}^{k} \left(\frac{\bar{Y}_{i.} - \bar{Y}_{..}}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2_{k-1}, \quad \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\frac{Y_{ij} - \bar{Y}_{..}}{\sigma^2}\right)^2 \sim \chi^2_{n-k}.$$

and since we have the ratio of two χ^2 random variables divided by their degrees of freedom, we have an $F_{k-1, n-k}$ distributed random variable, where

$$\frac{MSB}{MSW} \sim F_{k-1, n-k}. \quad \square$$