FINAL EXAMINATION

Statistics 305

Term 1, 2006-2007

Thursday, December 14, 2006

Time: 8:30am – 11:00am

Student Name (**Please print in caps**): SOLUTION

Student Number: ______________________

Notes:

- This exam has 7 problems on the 10 following pages, plus 3 pages of statistical tables. Check to ensure that you have a complete paper.
- The amount each part of each question is worth is shown in [ ] on the left-hand side of the page.
- Where appropriate, record your answers in the blanks provided on the right-hand side of the page.
- Your solutions must be justified; show all the work and state all the reason(s) leading to your answer for each question in the space provided immediately under the question.
- Clear and complete solutions are essential; little partial credit will be given.
- This is a closed book exam.
- A single two-sided 8.5 x 11 page of notes is allowed.
- Calculators are allowed (but not for symbolic differentiation or integration).
- No devices (including calculators) that can store text or send/receive messages are allowed.

Problem	Total Available	Score
1.	9	
2.	20	
3.	23	
4.	16	
5.	13	
6.	8	
7.	11	
Total	100	

1. Suppose X is a random variable with moment generating function $M_X(t)$ given by

$$M_X(t) = 0.2\exp(\theta t) + 0.6\exp(2\theta t) + 0.2\exp(3\theta t) \quad \text{for } -\infty < t < \infty.$$

[3] a) Find the expected value of X. 2θ

$$\frac{d}{dt} M_X(t) = 0.2\theta \exp(\theta t) + 0.6(2\theta)\exp(2\theta t) + 0.2(3\theta)\exp(3\theta t)$$

$$\Big|_{t=0} = 0.2\theta + 1.2\theta + 0.6\theta = 2\theta$$

$$\Rightarrow \underline{\underline{E(X) = 2\theta}}$$

[3] b) Find the standard deviation of X. $\sqrt{0.4}\,\theta$

$$\frac{d^2}{dt^2} M_X(t) = 0.2\theta^2 \exp(\theta t) + 0.6(2\theta)^2 \exp(2\theta t) + 0.2(3\theta)^2 \exp(3\theta t)$$

$$\Big|_{t=0} = 0.2\theta^2 + 0.6(4\theta^2) + 0.2(9\theta^2) = 4.4\theta^2$$

$$\Rightarrow E(X^2) = 4.4\theta^2$$

$$\Rightarrow Var(X) = 4.4\theta^2 - (2\theta)^2 = \underline{\underline{0.4\theta^2}}$$

[3] c) What is the distribution of X (Be sure to explain clearly.)? ________

Suppose W has the distribution:

$$W = \begin{cases} \theta & \text{with probability } 0.2 \\ 2\theta & \quad '' \quad 0.6 \\ 3\theta & \quad '' \quad 0.2 \end{cases}$$

Then $M_W(t) = M_X(t)$

$$\quad \equiv E(e^{tW})$$

By the uniqueness theorem for moment-generating functions, the distribution of X is the same as that of W.

2. Suppose $X_1, X_2, \ldots, X_n$ is a simple random sample from a Poisson distribution with parameter λ. Let Y be the number of the X_i's equal to 0 and consider the estimator:

$$\tilde{\lambda} = -\log(Y/n).$$

[2] a) What is the distribution of Y ? ______________

$$Y \sim B(n, P(X=0))$$

$$\Rightarrow Y \sim B(n, e^{-\lambda})$$

[8] b) Find a second-order approximation to the expected value of the estimator $\tilde{\lambda}$. Is the estimator $\tilde{\lambda}$ asymptotically unbiased? ______________

$$\tilde{\lambda} = -\log X, \text{ where } X = \frac{Y}{n} \Rightarrow E(X) = e^{-\lambda}$$

$$Var(X) = e^{-\lambda}(1-e^{-\lambda})/n$$

So $\tilde{\lambda} = f(X)$, where $f(x) = -\log x$

$$\Rightarrow f'(x) = -1/x$$

$$\Rightarrow f''(x) = 1/x^2$$

By Delta Method:

$$\Rightarrow E(\tilde{\lambda}) \cong f(E(X)) + \frac{1}{2} f''(E(X)) \cdot Var(X)$$

$$= -\log[e^{-\lambda}] + \frac{1}{2} \frac{1}{(e^{-\lambda})^2} \cdot \frac{e^{-\lambda}(1-e^{-\lambda})}{n}$$

$$\Rightarrow E(\tilde{\lambda}) \cong \underline{\underline{\lambda + \frac{1}{2}\left(\frac{1-e^{-\lambda}}{e^{-\lambda}}\right)\frac{1}{n}}}$$

$$\Rightarrow Bias(\tilde{\lambda}) \equiv E(\tilde{\lambda}) - \lambda = \frac{1}{2}\left(\frac{1-e^{-\lambda}}{e^{-\lambda}}\right)\frac{1}{n} \longrightarrow \underline{\underline{0}} \text{ as } n \to \infty$$

So $\tilde{\lambda}$ <u>is</u> asymptotically unbiased

[3] c) Find the asymptotic variance of the estimator $\tilde{\lambda}$. __________

By Delta Method: $\mathrm{Var}(\tilde{\lambda}) \cong [f'(E(X))]^2 \cdot \mathrm{Var}(X)$

$$= \left[-\frac{1}{e^{-\lambda}}\right]^2 \cdot \frac{e^{-\lambda}(1-e^{-\lambda})}{n}$$

$$\Rightarrow \mathrm{Var}(\tilde{\lambda}) \cong \left(\frac{1-e^{-\lambda}}{e^{-\lambda}}\right) \cdot \frac{1}{n}$$

[2] d) Find the leading term of the asymptotic MSE of the estimator $\tilde{\lambda}$. __________

$$\mathrm{MSE} = \mathrm{Var} + \mathrm{Bias}^2$$

$$\Rightarrow \mathrm{MSE}(\tilde{\lambda}) \cong \left(\frac{1-e^{-\lambda}}{e^{-\lambda}}\right) \cdot \frac{1}{n} + \left[\frac{1}{2}\left(\frac{1-e^{-\lambda}}{e^{-\lambda}}\right)\frac{1}{n}\right]^2$$

$$\Rightarrow \text{leading term} = \left(\frac{1-e^{-\lambda}}{e^{-\lambda}}\right)\frac{1}{n} \equiv \text{A(symptotic) MSE}_{\text{say}}$$

[5] e) Find the expression for the asymptotic efficiency of the estimator $\tilde{\lambda}$ relative to the maximum likelihood estimator (MLE) $\hat{\lambda}_{ML} = \bar{X}$. What is the limiting value of the asymptotic relative efficiency as $\lambda \to 0$? As $\lambda \to \infty$? __________

$\hat{\lambda}_{ML} = \bar{X}$ has $E = \lambda$ and $\mathrm{Var} = \lambda/n$

$\Rightarrow \mathrm{MSE}(\hat{\lambda}_{ML}) = \lambda/n$ (exact, not just asymptotic)

$$\mathrm{ARE}(\tilde{\lambda}, \hat{\lambda}_{ML}) = \frac{\mathrm{MSE}(\hat{\lambda}_{ML})}{\mathrm{AMSE}(\tilde{\lambda})} = \frac{\lambda/n}{\left(\frac{1-e^{-\lambda}}{e^{-\lambda}}\right)\frac{1}{n}} = \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}} = \frac{\lambda}{e^{\lambda}-1}$$

As $\lambda \downarrow 0$, $e^{\lambda} - 1 \simeq \lambda + \lambda^2/2 + \lambda^3/6 + \cdots$

$$\Rightarrow \frac{\lambda}{e^{\lambda}-1} \approx \frac{1}{1+\lambda/2+\cdots} \cong 1 - \frac{\lambda}{2} \to \underline{1}$$

As $\lambda \uparrow \infty$, $\mathrm{ARE} \approx \lambda e^{-\lambda} \longrightarrow \underline{0}$

3. Suppose $X_1, X_2, \ldots, X_n$ is a simple random sample from an exponential distribution with density given by:

$$f_\theta(x) = (1/\theta)\exp(-x/\theta) \quad \text{for } 0 \le x < \infty .$$

[5] a) Find $\hat{\theta}_{ML}$, the maximum likelihood estimator (MLE) of θ. $\overline{X}$

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \exp\left(-\frac{x_i}{\theta}\right) = \frac{1}{\theta^n} \exp\left(-\frac{n\bar{x}}{\theta}\right)$$

$$\Rightarrow \ell(\theta) = -n \log \theta - \frac{n\bar{x}}{\theta}$$

$$\Rightarrow \ell'(\theta) = -\frac{n}{\theta} + \frac{n\bar{x}}{\theta^2} \Rightarrow \ell'(\theta) > 0 \iff \frac{n\bar{x}}{\theta^2} > \frac{n}{\theta}$$

$$\iff \theta < \bar{x}$$

So $\bar{x}$ is max!

$$\Rightarrow \hat{\theta}_{ML} = \overline{X}$$

[3] b) What is the exact sampling distribution of the MLE $\hat{\theta}_{ML}$?

$$n\hat{\theta}_{ML} = \sum_{i=1}^{n} X_i$$

But $X_i \sim \mathcal{G}(1, \frac{1}{\theta})$ and they are independent

$$\Rightarrow n\hat{\theta}_{ML} \sim \mathcal{G}(n, \tfrac{1}{\theta})$$

Even better:

All are equivalent statements

$$\hat{\theta}_{ML} \sim \mathcal{G}\left(n, \frac{n}{\theta}\right)$$

OR $$\hat{\theta}_{ML} \overset{d}{=} \frac{1}{n} \cdot \mathcal{G}\left(n, \frac{1}{\theta}\right)$$

OR $$\frac{\hat{\theta}_{ML}}{\theta} \sim \mathcal{G}(n, n)$$

[2] c) What is the (exact) bias of the MLE $\hat{\theta}_{ML}$? 0

$$E\,\hat{\theta}_{ML} = \frac{1}{n}\left(\frac{n}{1/\theta}\right) = \theta$$

$E(W)$, where $W \sim \mathcal{G}(n, \frac{1}{\theta})$

$$\Rightarrow \text{Bias} \equiv E(\hat{\theta}_{ML}) - \theta = 0$$

$\Rightarrow \hat{\theta}_{ML}$ is unbiased!

$\dfrac{\theta^2}{n}$

[2] d) What is the (exact) variance of the MLE $\hat{\theta}_{ML}$?

$$Var(\hat{\theta}_{ML}) = \frac{1}{n^2} Var(\overline{W})$$
$$= \frac{1}{n^2} \frac{n}{\left(\frac{1}{\theta}\right)^2} = \frac{\theta^2}{n}$$

No

[5] e) Is there any other unbiased estimator of θ with a smaller variance?
Note: Be sure to explain clearly.

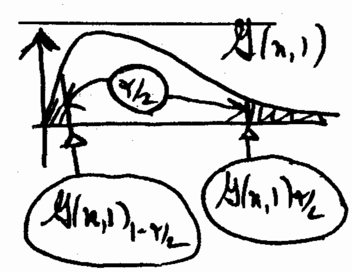
Need to evaluate the Cramer-Rao lower bound:

$$n I(\theta) = E\left[-\ell''(\theta)\right]$$
$$= E\left[-\frac{n}{\theta^2} + 2\frac{n\overline{X}}{\theta^3}\right] = -\frac{n}{\theta^2} + \frac{2}{\theta^3} n E(\overline{X})$$
$$= -\frac{n}{\theta^2} + \frac{2n}{\theta^3}\theta$$
$$= -\frac{n}{\theta^2} + \frac{2n}{\theta^2} = \frac{n}{\theta^2}$$

Note: The general result that $A.Var(\hat{\theta}_{ML}) = \frac{1}{nI(\theta)}$ is not good enough here as we are talking about exact (fixed n) results.

$\Rightarrow CRLB = \theta^2/n$. No unbiased estimator can have a smaller variance than CRLB $\Rightarrow$ NO!

[6] f) Find an exact $1-\alpha$ confidence interval for θ based on the MLE $\hat{\theta}_{ML}$.



$$n\hat{\theta}_{ML} \sim G\left(n, \frac{1}{\theta}\right) \Rightarrow \frac{n\hat{\theta}_{ML}}{\theta} \sim G(n,1)$$

$$\Rightarrow 1-\alpha = Pr\left\{ G(n,1)_{1-\alpha/2} \le \frac{n\hat{\theta}_{ML}}{\theta} \le G(n,1)_{\alpha/2} \right\}$$
$$= Pr\left\{ \frac{1}{G(n,1)_{1-\alpha/2}} \geq \frac{\theta}{n\hat{\theta}_{ML}} \geq \frac{1}{G(n,1)_{\alpha/2}} \right\}$$
$$= Pr\left\{ \frac{n\hat{\theta}_{ML}}{G(n,1)_{\alpha/2}} \le \theta \le \frac{n\hat{\theta}_{ML}}{G(n,1)_{1-\alpha/2}} \right\}$$

$$\Rightarrow 1-\alpha \text{ CI for } \theta \text{ is given by } \left(\frac{n\hat{\theta}_{ML}}{G(n,1)_{\alpha/2}}, \frac{n\hat{\theta}_{ML}}{G(n,1)_{1-\alpha/2}} \right)$$

Note: If $W \sim G(n,1) \Rightarrow 2W \sim G(n, \frac{1}{2}) \equiv G(\frac{2n}{2}, \frac{1}{2}) \equiv \chi^2_{2n}$. So, can use cut-offs from χ^2_{2n}

4. According to genetic theory, flies of three types resulting from certain cross-breedings should occur with probabilities θ^2, $2\theta(1-\theta)$ and $(1-\theta)^2$ respectively, where $0<\theta<1$. Let X_1, X_2 and X_3 denote the frequencies of these three types of flies in an experiment with a total of n independent such cross-breedings $(X_1+X_2+X_3=n)$.

[2] a) Write down the likelihood function, $L(\theta)$, for the data in this experiment in terms x_1, x_2 and x_3, the observed values of X_1, X_2 and X_3.

$$L(\theta) = \frac{n!}{x_1!\,x_2!\,x_3!}\left[\theta^2\right]^{x_1}\left[2\theta(1-\theta)\right]^{x_2}\left[(1-\theta)^2\right]^{x_3}$$

$$= \frac{n!}{x_1!\,x_2!\,x_3!}\cdot 2^{x_2}\cdot \theta^{2x_1+x_2}\cdot(1-\theta)^{x_2+2x_3}$$

[4] b) Express $\hat{\theta}_{ML}$, the MLE of θ, in terms of X_1, X_2 and X_3.

$$\ell(\theta) = c + (2x_1+x_2)\log\theta + (x_2+2x_3)\log(1-\theta)$$

$$\ell'(\theta) = \frac{(2x_1+x_2)}{\theta} - \frac{(x_2+2x_3)}{(1-\theta)}$$

$$\Rightarrow \ell'(\theta)>0 \Leftrightarrow \frac{2x_1+x_2}{\theta} > \frac{x_2+2x_3}{1-\theta}$$

$$\Leftrightarrow (2x_1+x_2)(1-\theta) > (x_2+2x_3)\theta$$

$$\Leftrightarrow \theta(2x_1+2x_2+2x_3) < 2x_1+x_2$$

$$\Leftrightarrow \theta < \frac{2x_1+x_2}{2n} \quad \Leftarrow \text{So, max!}$$

$$\Rightarrow \hat{\theta}_{ML} = \frac{2X_1+X_2}{2n}$$

[7] c) Find the asymptotic variance of $\hat{\theta}_{ML}$.

Evaluate the Fisher information for θ in the sample:

$$\ell''(\theta) = -\frac{2x_1+x_2}{\theta^2} - \frac{x_2+2x_3}{(1-\theta)^2}$$

$$E[-\ell''(\theta)] = \frac{1}{\theta^2}E(2X_1+X_2) + \frac{1}{(1-\theta)^2}E(X_2+2X_3)$$

$$X_i \sim B(n,\cdot) \Rightarrow \quad = \frac{1}{\theta^2}\left\{2n\theta^2 + n\,2\theta(1-\theta)\right\} + \frac{1}{(1-\theta)^2}\left\{n\,2\theta(1-\theta) + 2n(1-\theta)^2\right\}$$

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

$$= \frac{2n}{\theta} + \frac{2n}{1-\theta} = \frac{2n}{\theta(1-\theta)}$$

$$\Rightarrow A.Var(\hat{\theta}_{ML}) = \underline{\underline{\frac{\theta(1-\theta)}{2n}}}$$

[3] d) Find the form of an approximate 99% confidence interval for θ based on $\hat{\theta}_{ML}$.

$$\hat{\theta}_{ML} \stackrel{\sim}{\approx} N\left(\theta, \frac{\theta(1-\theta)}{2n}\right)$$

$$\Rightarrow \hat{\theta}_{ML} \pm z_{\alpha/2} \cdot \sqrt{\frac{\theta(1-\theta)}{2n}}$$ is an approximate $1-\alpha$ CI for θ

But, need to estimate θ here

$$\Rightarrow \underline{\underline{\hat{\theta}_{ML} \pm 2.58\sqrt{\frac{\hat{\theta}_{ML}(1-\hat{\theta}_{ML})}{2n}}}}$$

5. Suppose $X_1, X_2, \ldots, X_9$ is a random sample of size $n = 9$ from a $N(\mu, 1)$ population. You have decided to test H_0: $\mu = 0$ versus H_1: $\mu \neq 0$ by rejecting H_0 if either $\bar{x} < -0.5$ or $\bar{x} > 1.0$, where $\bar{x}$ is the observed value of the sample mean $\bar{X}$.

[4] a) What is the probability of Type I error of this test? 0.0681

Type I error ≡ reject H_0 when it is true

$\Rightarrow P(\text{Type I error}) = P_{\mu=0}(\bar{X} < -0.5 \text{ or } \bar{X} > 1.0)$

$= 1 - P_{\mu=0}(-0.5 < \bar{X} < 1.0)$ Under H_0, $\bar{X} \sim N(0, \frac{1}{\sqrt{9}})$

$= 1 - P(-0.5\sqrt{9} < \sqrt{9}\bar{X} < 1.0\sqrt{9})$

$= 1 - P(-1.5 < Z < 3.0)$

$= 1 - \{0.9987 - 0.0668\} = \underline{0.0681}$

[4] b) What is its probability of Type II error when the true value of μ is 1.2? 0.2743

Type II error = accept H_0 when it is false

$\Rightarrow P_{\mu=1.2}(\text{Type II error}) = P_{\mu=1.2}(-0.5 < \bar{X} < 1.0)$ When $\mu=1.2$, $\bar{X} \sim N(1.2, \frac{1}{\sqrt{9}})$

$= P((-0.5-1.2)\sqrt{9} < \sqrt{9}(\bar{X}-1.2) < (1.0-1.2)\sqrt{9})$

$= P(-5.1 < Z < -0.6)$

$= P(Z < -0.6) - P(Z < -5.1)$

$= 0.2743 - 0.0000 = \underline{0.2743}$

[5] c) What is the smallest sample size that would result in a power of at least 0.90 when the true value of μ is 1.2? 41

$\text{Power}_{\mu=1.2} = P_{\mu=1.2}(\text{reject } H_0) = 1 - P_{\mu=1.2}(\text{accept } H_0)$

When $n=9$, this is $1 - 0.2743 = 0.7257$ from b)

$\Rightarrow$ Require $n > 9$, for sure

$\text{Power}_{\mu=1.2} = 1 - P(-1.7\sqrt{n} < Z < -0.2\sqrt{n})$ as in b)

$= 1 - \{P(Z < -0.2\sqrt{n}) - P(Z < -1.7\sqrt{n})\}$

But $n > 9 \Rightarrow$ last piece $= 0$

$\Rightarrow 1 - P(Z < -0.2\sqrt{n}) \geq 0.9$ is required

$\Leftrightarrow P(Z \geq 0.2\sqrt{n}) \leq 0.1 \Leftrightarrow 0.2\sqrt{n} \geq 1.28 \Leftrightarrow \sqrt{n} \geq 6.4$

$\Leftrightarrow \underline{n \geq 40.96}$

6. Suppose $X_1, X_2, \ldots, X_n$ is a simple random sample from a double exponential distribution with density given by:

$$f_\theta(x) = (\theta/2)\exp(-\theta\,|\,x\,|) \qquad \text{for } -\infty < x < \infty .$$

Consider testing the simple null hypothesis H_0: $\theta = \theta_0$ against the simple alternative hypothesis H_1: $\theta = \theta_1$, where θ_0 and $\theta_1 > \theta_0$ are specified values.

[6] a) Derive the form of the most powerful test of H_0 versus H_1.
Note: You do not need to determine the critical (cut-off) value for the test.

$\sum_{i=1}^{n} |X_i| < c$

Neymann-Pearson Lemma says MP test for simple vs simple is:

Reject Ho $\Leftrightarrow$ $\frac{L_0}{L_1}$ is too small

$$\Leftrightarrow \quad \frac{\left(\frac{\theta_0}{2}\right)^n \exp\left\{-\theta_0 \Sigma |x_i|\right\}}{\left(\frac{\theta_1}{2}\right)^n \exp\left\{-\theta_1 \Sigma |x_i|\right\}} = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left\{(\theta_1-\theta_0)\,\Sigma|X_i|\right\}$$

is too small

$\Leftrightarrow$ $(\theta_1-\theta_0)\ \Sigma\,|X_i|$ is too small

But $\theta_1 > \theta_0$, so $\Leftrightarrow$ $\Sigma\,|x_i|$ is too small

[2] b) Is this test uniformly most powerful for testing H_0: $\theta = \theta_0$ against the composite alternative hypothesis H_1: $\theta > \theta_0$ (Be sure to explain clearly.)?

Yes

No matter what value of θ_1 we consider (provided $\theta_1 > \theta_0$), we obtain exactly the same (form of) test as above as the MP test.
But Ho doesn't change $\Rightarrow$ it is actually exactly the same test (the specific value of θ_1 does not enter into the determination of the appropriate critical value — only Ho and that $\theta_1 > \theta_0$)

$\Rightarrow$ this test is UMP for $H_0: \theta = \theta_0$ vs $H_1: \theta > \theta_0$.

7. For two factors, starchy/sugary and green/white base leaf, the counts in the following table were observed for a total of 3839 progeny of self-fertilized heterozygotes (Fisher 1958). According to genetic theory, the cell probabilities are as in the table, where θ $(0 < \theta < 1)$ is a parameter related to the linkage of the two factors.

Type	Count	Probability
Starchy green	1997	$(2+\theta)/4$
Starchy white	906	$(1-\theta)/4$
Sugary green	904	$(1-\theta)/4$
Sugary white	32	$\theta/4$

Note: Use of $X^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$ leads to $X^2 = 5.8716$ ⇒ p-value is a bit more than 0.10 (R yields $p = 0.118$) ⇒ X^2 yields same conclusion (but I asked you to use the GLRT ⇒ G^2.

[11] Use the generalized likelihood ratio test (GLRT) to assess whether the data provide convincing evidence against the null hypothesis H_0: $\theta = 0.05$. Note: You do not need to derive the form of the GLRT, but be sure to explain clearly all the steps in the work leading to your conclusion.

When H_0 is true, the expected counts are:

Useful check on calculations

$$3839 \times \tfrac{1}{4}(2+0.05) = 1967.4875$$
$$3839 \times \tfrac{1}{4}(1-0.05) = 911.7625$$
$$3839 \times \tfrac{1}{4}(1-0.05) = 911.7625$$
$$3839 \times \tfrac{1}{4}(0.05) = 47.9875$$

Total = 3,839.0000, as it must be!

Then GLRT is based on $G^2 = 2 \sum_{i=1}^{4} O_i \log\left(\frac{O_i}{E_i}\right)$

$$= 2\{29.7327 - 5.7443 - 7.7294 - 12.9665\}$$
$$= 2\{3.2926\} = 6.5852$$

Now 4 cells and no parameters estimated under H_0 (H_0 is a simple hypothesis here) ⇒ under H_0, $G^2 \approx \chi^2_{(3)}$

⇒ p-value is between 0.10 ($\chi^2_3(0.10) = 6.25$) and 0.05 ($\chi^2_3(0.05) = 7.81$).

Linear interpolation yields $p \approx 0.089$ (R yields $p = 0.086$)

⇒ Roughly a 1 in 10 chance of seeing a value as extreme or more extreme than the one we saw ($\chi^2_{(3)} = 6.5852$) ⇒ the data do NOT provide convincing evidence against H_0