Testing Simple* Hypotheses

* means "Distribution of data is completely specified, with no parameters to estimate".

X data

$f(x|\theta)$ model

$H_0 : \theta = \theta_0$, or dist. of X is $f(x|\theta_0)$

$H_1 : \theta = \theta_1$, or dist of X is $f(x|\theta_1)$

Neyman-Pearson Lemma (1933): Best test to use is Likelihood ratio (LR) test.

Reject H_0 if $\dfrac{f(x|\theta_1)}{f(x|\theta_0)} > K$.

$\alpha = P(\text{Rej. } H_0 \mid H_0 \text{ true})$ "Type 1" "False Positive"

$\beta = P(\text{Acc. } H_0 \mid H_1 \text{ true})$ "Type 2" "False Negative"

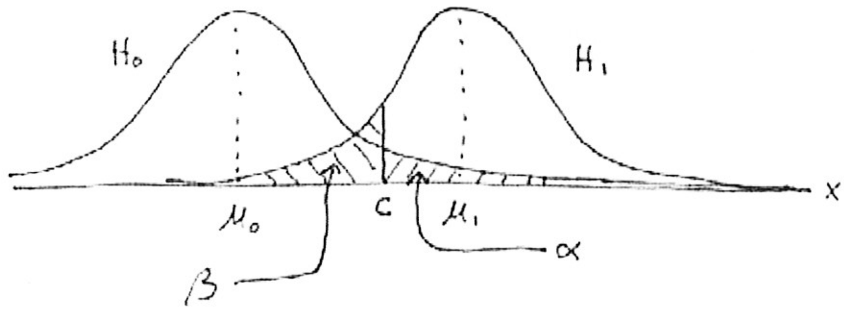
$\pi = 1 - \beta$ = power of the test.

The LR $\dfrac{f(x|\theta_1)}{f(x|\theta_0)}$ orders x values

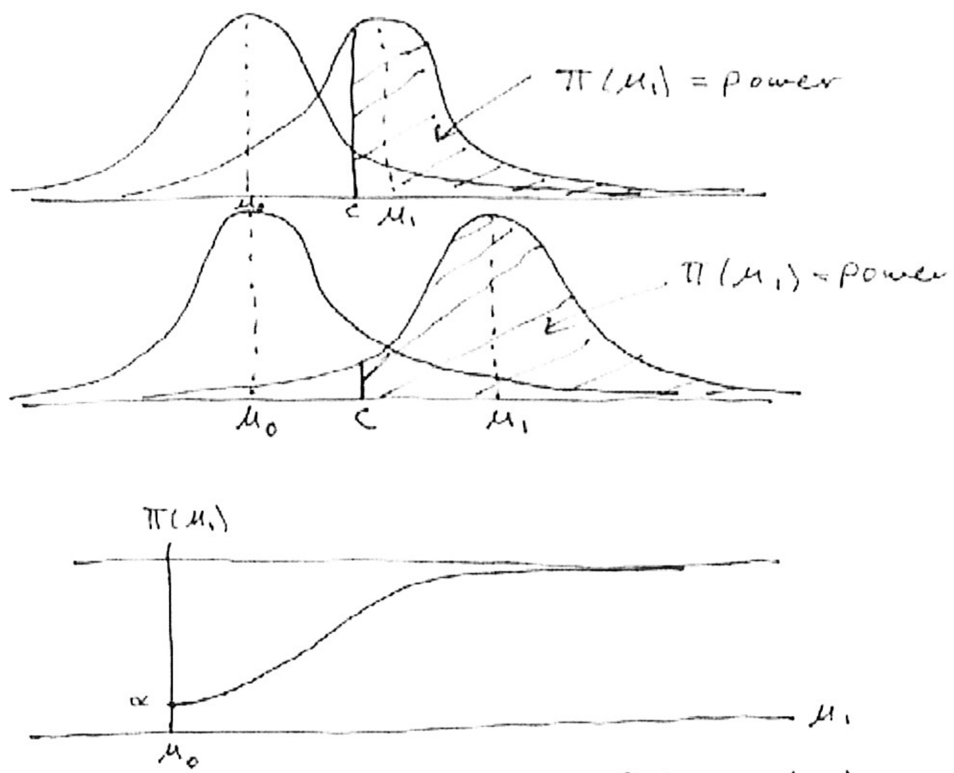
high LR is stronger evidence for H_1
low LR is " " " H_0

K draws the line

Figure 6.3



(a) Testing a simple hypothesis vs. a simple alternative.



(b) The power function $\pi(\mu_1) = \Pr(\text{Reject} \mid \mu_1)$, as a function of the alternative μ_1

[From Stigler, Chap 6]

The Neyman - Pearson Lemma

Given α, no test with the same or lower α has a lower β than the likelihood ratio with the given α.

Proof

The LR test rejects if $X = x$ for any x satisfying

$$f(x|\theta_1) > K f(x|\theta_0).$$

Define an "indicator function"

$$I_{NP} = \begin{cases} 1 & \text{iff } f(x|\theta_1) > K f(x|\theta_0) \\ 0 & \text{otherwise} \end{cases}$$

I_{NP} is a Bernoulli random variable.

$$E[I_{NP}] = 0 \cdot Pr(I_{NP} = 0) + 1 \cdot Pr(I_{NP} = 1) = Pr(I_{NP} = 1)$$

Let α_{NP} be the probability of a type I error for the NP test. Then

$$\alpha_{NP} = E[I_{NP}(X)|\theta_0] \qquad \left(\begin{array}{c}\text{reject } H_0 \\ H_0 \text{ true}\end{array}\right)$$

$$1 - \beta_{NP} = E[I_{NP}(X)|\theta_1] \qquad \left(\begin{array}{c}\text{reject } H_0 \\ H_1 \text{ true} \\ \text{"power"}\end{array}\right)$$

Let T be <u>any</u> <u>other</u> <u>test</u> with $\alpha_T \le \alpha_{NP}$, and let $I_T = \begin{cases} 1 & \text{iff } T \text{ rejects } H_0 \text{ data } X = x \\ 0 & \text{otherwise} \end{cases}$

Then $\alpha_T = E[I_T(X) | \theta_0]$ ("reject θ_0 H_0 true")

$1 - \beta_T = E[I_T(x) | \theta_1]$ ("accept θ_1 H_1 true")

Claim: for all x,

$$I_{NP}(x)\left[f(x|\theta_1) - K f(x|\theta_0)\right]$$
$$\geq I_T(x)\left[f(x|\theta_1) - K f(x|\theta_0)\right]$$

why? The part in [ ] is the same on both sides. If $I_{NP}(x) = 1$, then $[\ ] \geq 0$, and since $I_{NP}(x) = 1 \geq I_T(x)$, the inequality is true. If $I_{NP}(x) = 0$, $[\ ] \leq 0$ and the inequality holds because $I_T(x) \geq 0 = I_{NP}(x)$.

Multiply out the inequality to get

$$I_{NP}(x) f(x|\theta_1) - K I_{NP}(x) f(x|\theta_0)$$
$$\geq I_T(x) f(x|\theta_1) - K I_T(x) f(x|\theta_0)$$

Now, sum or itegrate or multiply integrate over x to give expectations:

$$E[I_{NP}(x)|\theta_1] - K\,E[I_{NP}(X)|\theta_0] \geq$$
$$E[I_T(x)|\theta_1] - K\,E[I_T(X)|\theta_0]$$

Now let's change notation back to α's and β's:

$$1 - \beta_{NP} - K\alpha_{NP} \geq 1 - \beta_T - K\alpha_T$$
$$1 - \beta_{NP} \geq 1 - \beta_T + K(\alpha_{NP} - \alpha_T)$$

but: $\alpha_{NP} - \alpha_T \geq 0$, $K \geq 0$, so

$$1 - \beta_{NP} \geq 1 - \beta_T$$
$$\beta_T \geq \beta_{NP}$$

$\therefore$

So: the LR test is the <u>most</u> <u>powerful</u> for a particular θ_1, given α.

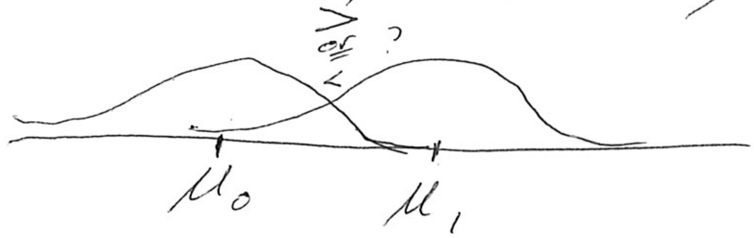
What about composite tests?

Restricted solution <u>may</u> solve more general problem

<u>Ex</u> $X_1, \ldots, X_n$ $N(\mu, \sigma_0^2)$, σ_0^2 known

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu = \mu_1 > \mu_0$$



Test: Reject H_0 if $\bar{X} > C = \mu_0 + Z_\alpha \frac{\sigma_0}{\sqrt{n}}$

<u>Note</u>

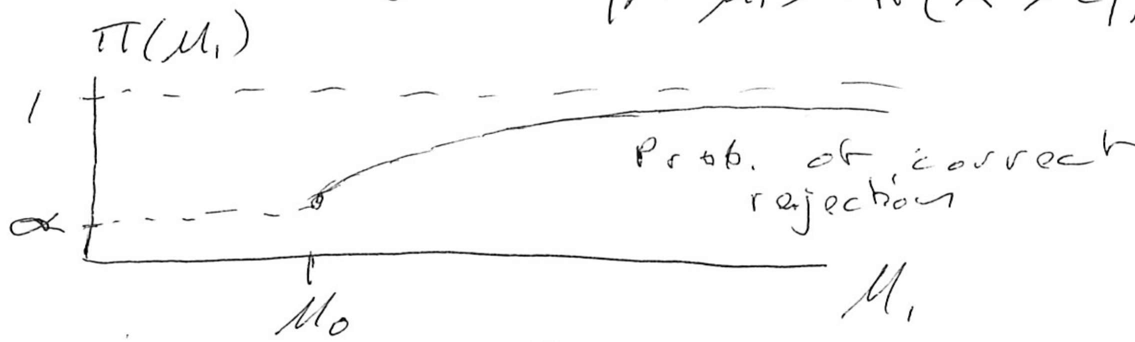
Same test for <u>any</u> $\mu_1 > \mu_0$

But the <u>power</u> depends on μ_1

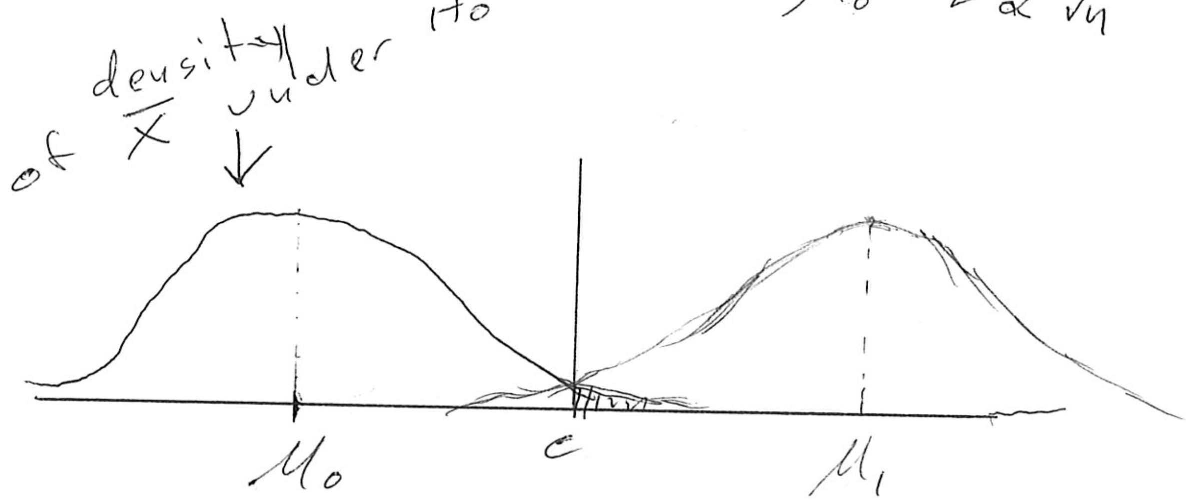
The test is <u>uniformly</u> <u>most</u> <u>powerful</u>

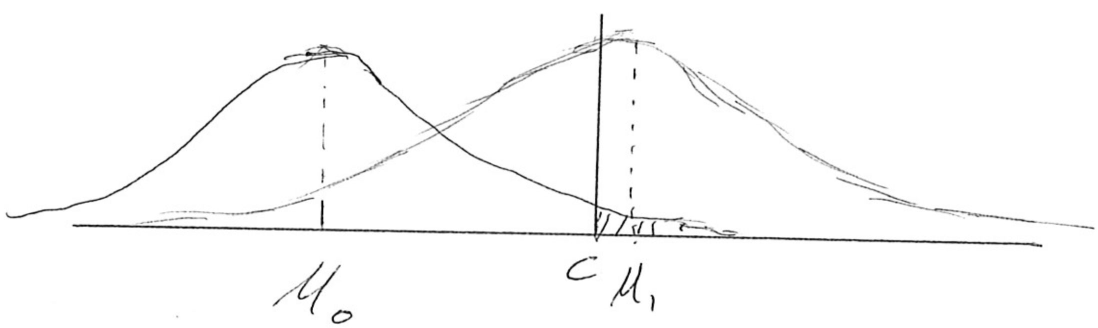
Describe peformance with <u>power</u> <u>function</u>

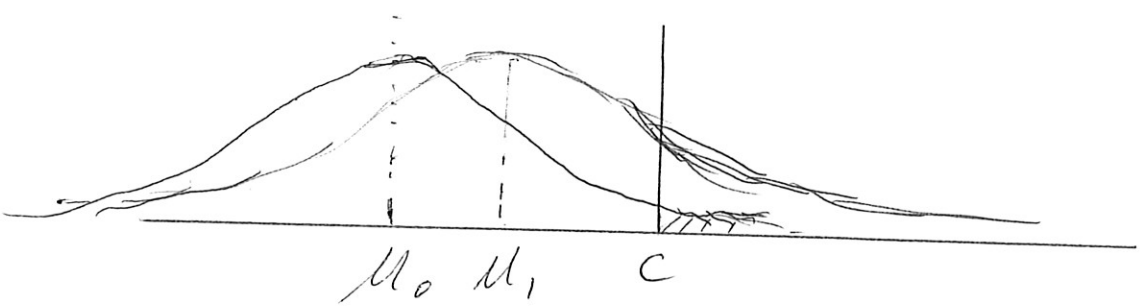
$$\pi(\mu_1) = \Pr(\text{Reject } H_0 \mid \mu = \mu_1) = \Pr(\bar{X} > C \mid \mu = \mu_1)$$



$$C = \mu_0 + Z_\alpha \frac{\sigma_0}{\sqrt{n}}$$







Example: X binomial (n, θ)

$H_0: \theta = \theta_0$ ($= \frac{1}{2}$? "fair coin")

$H_1: \theta = \theta_1 > \theta_0$

Likelihood Ratio: $\dfrac{p(X|\theta_1)}{p(X|\theta_0)} = \underbrace{\left(\dfrac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)}_{>1}^{X} \left(\dfrac{1-\theta_1}{1-\theta_0}\right)^{n}$

because: $\dfrac{\theta_1}{\theta_0} > 1$, $\dfrac{1-\theta_0}{1-\theta_1} > 1$

large when X large

Test: Reject H_0 if $X > c$,

want $P(X > c \,|\, \theta = \theta_0) = \alpha$

(not possible exactly for all α)

Ex: $n = 5$ $\theta_0 = \frac{1}{2}$

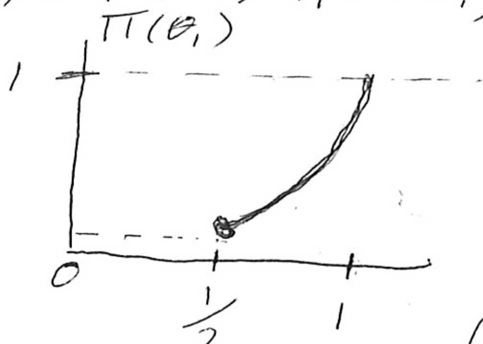
X	0	1	2	3	4	5
$p(X\|\theta_0)$	.03	.16	.31	.31	.16	.03

$\alpha = .03 \longrightarrow$ Reject H_0 if $X > 4$ (≥ 5)

$\alpha = .19 \longrightarrow$ Reject H_0 if $X > 3$

Power function? For $\alpha = .03$, $c = 4$

$\pi(\theta_1) = P(X > 4 | \theta = \theta_1) = P(X = 5 | \theta = \theta_1) = \theta_1^5$



UMP

But: In General, when testing composite* hypotheses [*ie more than one distribution in H_0 and/or H_1]

there is no UMP test.

Ex: $X_1, \ldots, X_n$ $\mathcal{N}(\mu, \sigma_0^2)$ ⟶ Known

$H_0: \mu = \mu_0$

$H_1: \mu = \mu_1 \neq \mu_0$ [composite]

Possible Tests:

① Reject if $\bar{X} > c$ ————

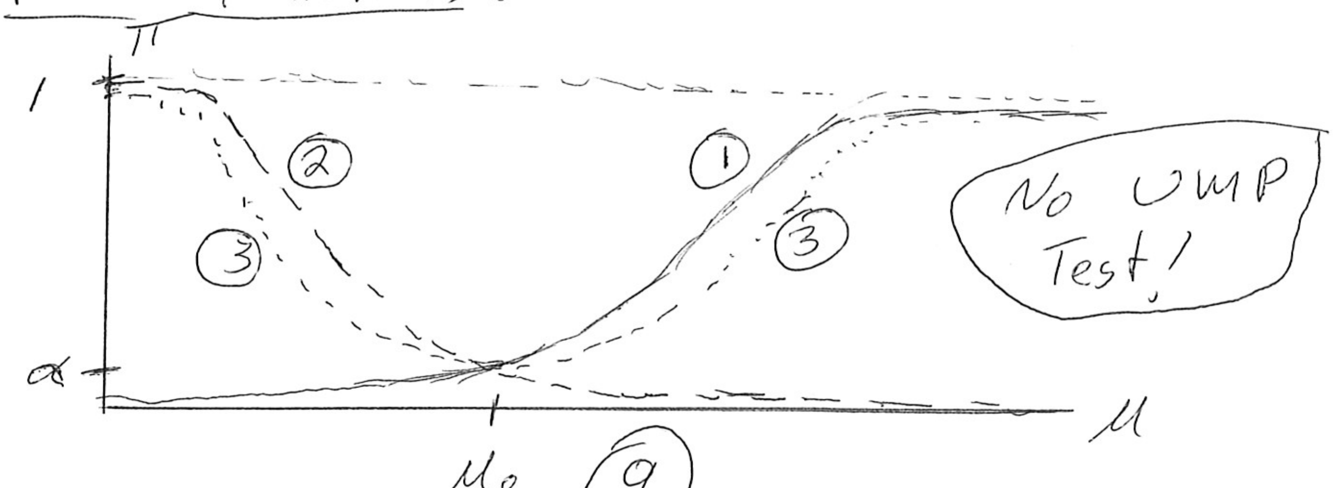
(Best vs $\mu_1 > \mu_0$

② Reject if $\bar{X} < c'$ - - - - -

(Best vs $\mu_1 < \mu_0$)

③ Reject if $|\bar{X} - \mu_0| > c''$ ....

Power Functions:



Likelihood Ratio Tests – General Case

Test H_0: Group of θ_0's

H_1: Group of θ_1's

Idea: Compare "champ" of H_0 to "champ" of H_1.

Could compare $\max\limits_{\theta_0\text{'s}} L(\theta)$ to $\max\limits_{\theta_1\text{'s}} L(\theta)$

Instead, compare $\max\limits_{\theta_0\text{'s}} L(\theta)$ to $\underbrace{\max\limits_{\text{all } \theta} L(\theta)}_{\text{(max at } \underline{\underline{\text{MLE}}}\text{!)}}$

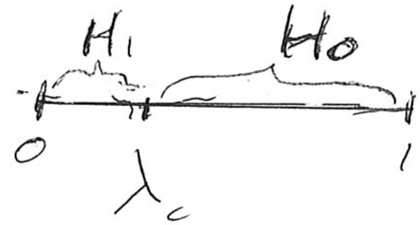
Let $\lambda = \dfrac{\max\limits_{\theta_0\text{'s}} L(\hat{\theta}_0)}{\max\limits_{\text{all } \theta\text{'s}} L(\hat{\theta}_1)}$

If H_0 clearly best, $\lambda \doteq 1$

If H_1 clearly best, $\lambda \ll 1$

$\boxed{\text{So: Reject } H_0 \text{ if } \lambda < \lambda_c, \text{ with } P(\lambda < \lambda_c \mid H_0) \leq \alpha}$

"<u>Likelihood Ratio Test</u>"



Examples of Likelihood Ratio Tests

1) Neyman - Pearson Tests are a special case
2) Student's t-tests, tests of 1 or 2 means with <u>unknown</u> variances.
3) ANOVA - "Analysis of Variance"
4) Regression tests
5) Variance comparisons
6) Chi-Square tests

2, 3, 4, and 5 : 245!

Chi-Square Tests
Contingency Tables
Tests of Fit

But First:

<u>Multinomial</u> Distributions

Multinomial: Generalization of Binomial

n independent trials

For each trial:

(mutually exclusive) Outcomes $A_1, A_2, \ldots, A_K$

Probabilities $\theta_1, \theta_2, \ldots, \theta_K$

$$(\theta_1 + \theta_2 + \cdots + \theta_K = 1)$$

Counts $X_1, X_2, \ldots, X_K$

$$(X_1 + X_2 + \cdots + X_K = n)$$

Ex: $K=2$ $A_1 = $ "H", $A_2 = $ "T"

$X_1 = X$, $X_2 = n - X$, X Binomial

Ex: $K = 11$ $A_i = $ "pair of dice total $i+1$"

$(i = 1, 2, \ldots, 11)$

Roll pair of dice n times

$X_1 = \#2\text{'s}, \ldots, X_{11} = \#12\text{'s}, \sum X_i = n$

Ex: $K = 38$. Roulette wheel, n spins

$X_1 = \#1\text{'s}, \ldots, X_{36} = \#36\text{'s}, X_{37} = \#0\text{'s}, X_{38} = \#00\text{'s}$

n trials, K outcomes each

X_i = count of # A_i's

Note: For any A_i (say A_3) we can regroup: A_3 vs. all others. Then $X_3 = \# A_3$'s can be seen to have a binomial marginal distribution:

$$P(X_3 = j) = \binom{n}{j} \theta_3^j (1-\theta_3)^{n-j}$$

$$E(X_3) = n\theta_3, \quad Var(X_3) = n\theta_3(1-\theta_3)$$

Same for any single X_i

Multinomial Distribution:

$(X_1, X_2, \ldots, X_K)$ are dependent, multivariate

$$p(x_1, x_2, \ldots, x_n | \theta's) = P_r(X_1 = x_1, \ldots, X_n = x_n | \theta's)$$

$$= \frac{n!}{x_1!\, x_2! \ldots x_n!} \theta_1^{x_1} \theta_2^{x_2} \cdots \theta_K^{x_K}$$

if $x_1 + x_2 + \cdots + x_K = n$

$= 0$ otherwise

Estimation

$$L(\theta_1, \dots, \theta_k) = \frac{n!}{x_1!\, x_2! \dots x_n!} \theta_1^{x_1} \cdots \theta_k^{x_k}$$

$$\theta_1 + \cdots + \theta_k = 1$$

MLE's $\hat{\theta}_i = \frac{x_i}{n}$ – sample fractions

[can show this from $\frac{\partial}{\partial \theta_i} L(\vec{\theta})$

Testing: $H_0 : \theta_1 = a_1, \dots, \theta_k = a_k$

H_1 : "otherwise"

L. R. Test:

$$\lambda = \frac{\max_{H_0\ \theta's} L(\theta_1, \dots, \theta_n)}{\max_{\text{all } \theta's} L(\theta_1, \dots, \theta_n)}$$

Reject if $\lambda < \lambda_c$. Let $E(x_i | H_0) = m_i = n a_i$

Then $\lambda = \frac{L(a_1, \dots, a_k)}{L(\hat{\theta}_1, \dots, \hat{\theta}_n)} = \left(\frac{a_1}{\hat{\theta}_1}\right)^{x_1} \cdots \left(\frac{a_k}{\hat{\theta}_k}\right)^{x_k}$

$$= \left(\frac{m_1}{x_1}\right)^{x_1} \cdots \left(\frac{m_k}{x_k}\right)^{x_k}$$

λ small $\rightarrow$ $-\log \lambda$ large, reject if

$-\log \lambda > K$

Multinomial LR test, cont:

$$\lambda = \left(\frac{m_1}{X_1}\right)^{X_1} \cdots \left(\frac{m_k}{X_k}\right)^{X_k}$$

Let's invoke Taylor's Thm:

$$f(x-m) = f(m) + f'(m)(x-m) + \frac{1}{2} f''(m)(x-m)^2 + \text{Rem}$$

Let $f(x) = x \log\left(\frac{x}{m}\right)$, so $f(m) = 0$, $f'(m) = 1$, $f''(m) = \frac{1}{m}$

Then $f(x) = (x-m) + \frac{1}{2}\frac{(x-m)^2}{m} + \cdots$

$$-\log\lambda = -\sum x_i \log\left(\frac{m_i}{x_i}\right)$$

$$= \sum x_i \log\left(\frac{x_i}{m_i}\right)$$

$$= \underbrace{\sum (x_i - m_i)} + \frac{1}{2}\sum \frac{(x_i - m_i)^2}{m_i} + \text{Rem}$$

$\sum (x_i - m_i)$
$= \sum x_i - \sum m_i$
$= n - n = 0$

$$= 0 + \frac{1}{2}\sum \frac{(x_i - m_i)^2}{m_i} + \text{Rem}$$

$$= \frac{1}{2}\chi^2 + \text{Rem}$$

$$\boxed{-2\log\lambda \stackrel{\sim}{=} \chi^2}$$ (if Rem small)

So... "reject if $\chi^2 > c$" is almost the L. R. test...

more next time.