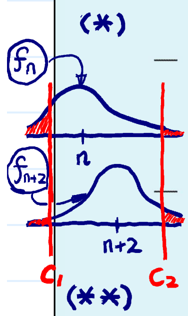
Example 7.13 (UMPU test for normal variance)

- Let $\mathbf{X} = (X_1, \cdots, X_n)$ be i.i.d. from $N(\mu, \sigma^2)$, where μ is known and σ is unknown. ← σ: parameter >0
- The joint pdf is $f(\mathbf{x}|\sigma) = \exp\left\{ \underbrace{-\frac{1}{2\sigma^2}}_{C(\sigma)} \underbrace{\sum_{i=1}^{n}(x_i - \mu)^2}_{T(\mathbf{x})\text{: a statistic}} - n\log\sqrt{2\pi}\sigma \right\}$,
  $\because \mu$ is a known constant
  which belongs to one-parameter exponential family.
- Here, $C(\sigma) = -1/(2\sigma^2)$ is an increasing function of σ.
- Null and alternative hypotheses: two-sided
  $$H_0: \sigma = \sigma_0 \quad \text{vs.} \quad H_A: \sigma \neq \sigma_0$$
  $\frac{(X_i-\mu)^2}{\sigma_0^2} \overset{\text{i.i.d.}}{\sim} \chi_1^2$ under H_0
  $T \sim \chi_n^2$ under H_0 ← null distribution
- The level-α UMP unbiased test is given by

  $T \propto$ sample variance ← reasonable?

  $$\phi(\mathbf{X}) = \begin{cases} 1, & \text{if } T \equiv \sum_{i=1}^{n}(X_i-\mu)^2/\sigma_0^2 < c_1 \text{ or } > c_2, \\ 0, & \text{otherwise.} \end{cases}$$
  T: test statistic
- The c_1, c_2 are determined by ← using null distribution
  $$(*) — E_{\sigma_0}(1-\phi) = \int_{c_1}^{c_2} f_n(y)\,dy = 1-\alpha, \Leftrightarrow E_{\sigma_0}(\phi(\mathbf{X})) = \alpha.$$
  $E_{\sigma_0}(1-\phi) = P_{\sigma_0}(c_1<T<c_2)$
  and
  $$E_{\sigma_0}[\phi(\mathbf{X})T(\mathbf{X})] = E_{\sigma_0}[\phi(\mathbf{X})]\cdot E_{\sigma_0}[T(\mathbf{X})] \Leftrightarrow$$
  $$E(T)-E(\phi T) = E_{\sigma_0}[(1-\phi)T] = \int_{c_1}^{c_2} y\, f_n(y)\,dy = E_{\sigma_0}(1-\phi)\, E_{\sigma_0}(T) = n\,(1-\alpha),$$
  $E_{\sigma_0}(1-\phi)\,E_{\sigma_0}(T) = E(T) - E(\phi)E(T)$
  where $f_n(y)$ is the pdf of the χ_n^2 distribution.



- Further derivation:
  - It is convenient to use the identity $y\, f_n(y) = n\, f_{n+2}(y)$. (Exercise.)
  - Rewrite the second condition as
    $$\int_{c_1}^{c_2} f_{n+2}(y)\,dy = 1-\alpha. \quad — (**)$$
  - Unless n is very small, or σ_0 is very close to 0 or ∞, the equal-tails test given by
    $$\int_0^{c_1} f_n(y)\,dy = \int_{c_2}^{\infty} f_n(y)\,dy = \alpha/2$$
    is a good approximation to the UMPU test.

  $f_n \approx f_{n+2}$ and symmetric about n ← ⊛ This follows from the fact that χ_n^2 distribution tends to Normal distribution for large n by CLT.
  $Z_1, \ldots, Z_n \overset{\text{i.i.d.}}{\sim} \chi_1^2$, $\sum_{i=1}^{n} Z_i \sim \chi_n^2$

FYI → More general condition (than exponential family) for finding $\begin{matrix}\text{UMP}\\ \text{UMPU}\end{matrix}$ tests.

Definition 7.9 (monotone likelihood ratio)

△ in LNp.18~19 & △ in LNp.23 (cf.)

We say that the likelihood function $\mathcal{L}(\theta, \mathbf{x})$ has **monotone likelihood ratio (MLR)** in the statistic $T(\mathbf{x})$, if for any $\theta_1 < \theta_2$, the ratio

used to determine what observations "more extreme"

$$\frac{\mathcal{L}(\theta_1, \mathbf{x})}{\mathcal{L}(\theta_2, \mathbf{x})}$$

irrelevant to θ

Recall Neymann-Pearson Lemma
$$\frac{f_0(\mathbf{x})}{f_A(\mathbf{x})} = \frac{\mathcal{L}(\theta_0, \mathbf{x})}{\mathcal{L}(\theta_A, \mathbf{x})}\downarrow \Leftrightarrow T(\mathbf{x})\uparrow \Leftrightarrow RR: T > c$$

is a descreasing function of $T(\mathbf{x})$.

Theorem 7.5

LNp.23 (UMP) LNp.26 (UMPU)

Theorems 7.3 and 7.4 still hold when we replace the one-parameter exponential family by a family of pdfs/pmfs that has MLR property in a statistic T.

Theorem 7.6

Let $\mathbf{X} = (X_1, \dots, X_n)$ have a joint pdf/pmf from the one-parameter exponential family:

$$\frac{\mathcal{L}(\theta_1, \underset{\sim}{x})}{\mathcal{L}(\theta_2, \underset{\sim}{x})} \propto \exp\{[C(\theta_1) - C(\theta_2)] \cdot T(\underset{\sim}{x})\}$$

It has the MLR property.

$$f(\mathbf{x}|\theta) = \exp\{C(\theta)\, T(\mathbf{x}) + d(\theta) + S(\mathbf{x})\} I_A(\mathbf{x}).$$

Suppose $C(\theta)$ is increasing. The likelihood function $\mathcal{L}(\theta, \mathbf{x})$ has MLR property in $T(\mathbf{x})$. Suppose $C(\theta)$ is decreasing. The $\mathcal{L}$ has MLR property in $-T(\mathbf{x})$.

❖ **Further reading**: Roussas, 13.4, 13.5

composite H_0 or H_A

- (Generalized) likelihood ratio tests

Recall. Estimation
1. UMVUE may not exist → MLE (good asymptotic property)
2. Likelihood.

Identify what observations "more extreme"

Simple vs. Simple: MP → UMP → UMPU; MP ↕ LR (Thm 7.1 (LNp.17)); LR → GLR; may not exist

Question 7.12

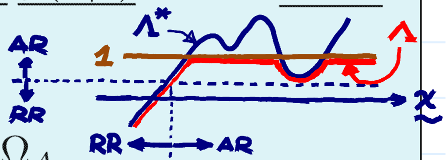
UMP/UMPU tests may not exist, and even though exist, may be difficult to derive. Any other procedure for finding a good/reasonable test?

Definition 7.10 (generalized likelihood ratio test, TBp. 339)

parameter space

- Suppose $\mathbf{X} = (X_1, \dots, X_n)$ have a joint pdf/pmf $f(\mathbf{x}|\theta)$, where $\theta \in \Omega$.
- Consider testing the hypotheses (not necessarily i.i.d.)

$$H_0 : \theta \in \Omega_0 \quad \text{vs.} \quad H_A : \theta \in \Omega_A$$

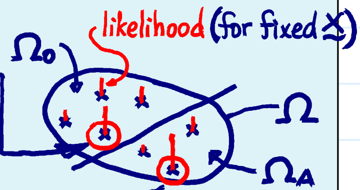


where $\Omega_0, \Omega_A \subset \Omega$, $\Omega_0 \cap \Omega_A = \phi$, and $\Omega = \Omega_0 \cup \Omega_A$.

- The **(generalized) likelihood ratio** (**GLR** or **LR**) is given by

$$\underset{\text{function of data only}}{\Lambda^*(\mathbf{x})} = \frac{\max_{\theta \in \Omega_0} \mathcal{L}(\theta, \mathbf{x})}{\max_{\theta \in \Omega_A} \mathcal{L}(\theta, \mathbf{x})},$$

sup; substitute MLE in Ω_0; substitute MLE in Ω_A; cf. Neymann-Pearson Lemma



where $\mathcal{L}(\theta, \mathbf{x})$ is the likelihood function.

- Small values of Λ^* tend to discredit H_0. ← determine "more extreme"
- For technical reasons, it is preferable to use

$$\Lambda(\mathbf{x}) = \frac{\max_{\theta \in \Omega_0} \mathcal{L}(\theta, \mathbf{x})}{\max_{\theta \in \Omega} \mathcal{L}(\theta, \mathbf{x})} \quad (\text{note: } \Lambda = \min(\Lambda^*, 1)).$$

$0 < \Lambda \leq 1$; a statistic; sup; substitute MLE in Ω_0; substitute MLE in Ω

$$\max_{\theta \in \Omega} \mathcal{L}(\theta) = \begin{cases} \max_{\theta \in \Omega_0} \mathcal{L}, & \text{if } \arg\max \Omega \in \Omega_0 \ (\leftrightarrow \Lambda^* > 1) \\ \max_{\theta \in \Omega_A} \mathcal{L}, & \text{if } \arg\max \Omega \in \Omega_A \ (\leftrightarrow \Lambda^* < 1) \end{cases}$$

problem 7.9 (LNp.19)

- Small values of Λ tend to discredit H_0.
- The rejection region of a GLR test consists of observations of $\mathbf{X}$ that correspond to small values of Λ.

Example 7.14 (GLR tests for normal mean with known variance, two-sided, TBp.339-340)

- Suppose that $X_1, \ldots, X_n$ are i.i.d. from $N(\mu, \sigma^2)$, where σ^2 is known. μ. parameter
- Consider the hypotheses simple $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$.

  Then $\Omega_0 = \{\mu_0\}$ (dim=0), $\Omega_A = \{\mu : \mu \neq \mu_0\}$, $\Omega = \{-\infty < \mu < \infty\}$ (dim=1).
- The LR statistic is

$$\begin{aligned}
\Lambda &= \frac{(\sqrt{2\pi}\sigma)^{-n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_0)^2\right]}{\max_{-\infty<\mu<\infty}\left[(\sqrt{2\pi}\sigma)^{-n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right]\right]} \\
&= \frac{\cancel{(\sqrt{2\pi}\sigma)^{-n}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_0)^2\right]}{\cancel{(\sqrt{2\pi}\sigma)^{-n}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \overline{x})^2\right]} \\
&= \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \overline{x})^2\right]\right) \\
&= \exp\left[-\frac{n}{2\sigma^2}(\overline{x} - \mu_0)^2\right]
\end{aligned}$$

substitute the MLE of $\mu = \overline{X}$

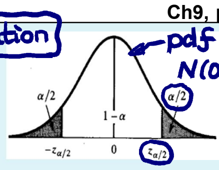
$-\overline{X}+\overline{X}$ → $\sum(x_i-\overline{x})^2 + n(\overline{x}-\mu_0)^2$

- Thus the LR test rejects H_0 for small values of Λ, i.e., large values of

RR: $\left(\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right)^2 > c$ a decreasing function of Λ → $-2\log(\Lambda) = \dfrac{(\overline{x}-\mu_0)^2}{\sigma^2/n}$. cf the test statistic of UMPU test in LNp.28

$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu_0, \sigma^2)$ test statistic

- Under H_0, $\overline{X} \sim N(\mu_0, \sigma^2/n)$ and $-2\log\Lambda \sim \chi^2_1$. null distribution



- Thus, the LR test rejects when

different test stat. but, same RR

$$\frac{(\overline{X} - \mu_0)^2}{\sigma^2/n} > \chi^2_1(\alpha) \quad \text{or equivalently} \quad \frac{|\overline{X} - \mu_0|}{\sigma/\sqrt{n}} > z(\alpha/2).$$

$Z \sim N(0,1) \Rightarrow Z^2 \sim \chi^2_1$

test statistic — Is this UMPU test? Check Ex.7.12(LNp.28) — test statistic

Example 7.15 (GLR tests for normal mean with unknown variance, two-sided)

- Let $X_1, \ldots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$, where μ and σ are unknown.

  parameters (2-parameter exponential family)
- Consider the hypotheses not simple $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$.

  Then $\Omega_0 = \{(\mu_0, \sigma) : \sigma > 0\}$ (dim=1), $\Omega_A = \{(\mu, \sigma) : \mu \neq \mu_0,\ \sigma > 0\}$,

  $\Omega = \{(\mu, \sigma) : -\infty < \mu < \infty,\ \sigma > 0\}$ (dim=2).
- The LR statistic is

MLE of σ in Ω_0

$$\begin{aligned}
\Lambda &= \frac{\max_{0<\sigma<\infty} (\sqrt{2\pi}\sigma)^{-n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_0)^2\right]}{\max_{-\infty<\mu<\infty,\, 0<\sigma<\infty}\left[(\sqrt{2\pi}\sigma)^{-n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right]\right]} \\
&= \frac{(\cancel{\sqrt{2}\pi}\hat{\sigma}_0)^{-n}\, \cancel{\exp(-n/2)}}{(\cancel{\sqrt{2}\pi}\hat{\sigma}_1)^{-n}\, \cancel{\exp(-n/2)}} = \left(\hat{\sigma}_0^2/\hat{\sigma}_1^2\right)^{-n/2},
\end{aligned}$$

$\hat{\sigma}_0^2 = \hat{\sigma}_1^2 + \frac{n(\overline{X}-\mu_0)^2}{n}$ ← $-\overline{X}+\overline{X}$

MLE of σ in Ω — MLE of μ in $\Omega = \overline{X}$

where $\hat{\sigma}_0^2 = \sum_{i=1}^n (X_i - \mu_0)^2/n$ and $\hat{\sigma}_1^2 = \sum_{i=1}^n (X_i - \overline{X})^2/n$. LN.CH8, p.20~21

- Now, determine what observations "more extreme" $\Lambda = \left[1 + (\overline{X} - \mu_0)^2/\hat{\sigma}_1^2\right]^{-n/2}$. $\Lambda\downarrow \Leftrightarrow \frac{(\bar{x}-\mu_0)^2}{\hat\sigma_1^2}\uparrow$
- Thus the LR test rejects H_0 for small values of Λ, i.e., large values of

Why use this as the test statistic? to get a common null distribution

$$\left(\frac{\overline{X} - \mu_0}{\hat{\sigma}_1}\right)^2 \Leftrightarrow \left|\frac{\bar{x}-\mu_0}{\hat\sigma_1}\right|$$

(reasonable?) cf. test statistic in LNp.28

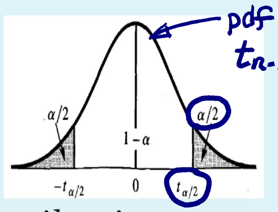
- Under H_0, $\sqrt{n-1}\,(\overline{X} - \mu_0)/\hat{\sigma}_1 \sim t_{n-1}$. null distribution (LN, CH-6, p.80)

test statistic

- Thus, the level-α LR test rejects H_0 when

$$\left|\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2/(n-1)}}\right| > t_{n-1}(\alpha/2),$$

S_n^2 in LN, CH-6, p.80



where $t_{n-1}(\alpha/2)$ is the $(1 - \alpha/2)$-quantile of the t_{n-1} distribution.

Q: Why is it useful? To calculate estimated standard error, relative efficiency, confidence interval, ...

Question 7.13

In point estimation, the asymptotic distribution of MLE tends to Normal when sample size is large enough. The property is useful when the exact distribution of MLE is difficult to find. to calculate p-value or determine critical value

Q: *Similar result exist for the null distribution of GLR statistic?* Why need it?

Theorem 7.7 (large sample theory for null distribution of GLR statistics, TBp.341)

Under smoothness conditions on the density function, the null distribution of

$\boxed{0 \le -2\log\Lambda < \infty}$ $\quad -2\log(\Lambda)$ ← test statistic

tends to a χ^2_m distribution, with degree of freedom

$$m = \dim(\Omega) - \dim(\Omega_0),$$

as the sample size tends to infinity. sample size large enough

Note. $-2\log\Lambda\uparrow \Leftrightarrow \Lambda\downarrow$
RR: $\Lambda < c$
$\Leftrightarrow -2\log\Lambda > c' = \chi^2_m(\alpha)$
pdf of χ^2_m, m, α

(**Note.** $\dim(\Omega)$ and $\dim(\Omega_0)$ are the dimensions of free parameters under Ω and Ω_0, respectively. For example, in Ex. 7.14, LNp.33, $\dim(\Omega) = 1$, $\dim(\Omega_0) = 0$, $m = 1 - 0 = 1$.)

beyond the scope of the course

Proof. The proof is based on a second-order Taylor expansion of $-2\log\Lambda$.

❖ **Reading**: textbook, 9.4; **Further reading**: Roussas, 13.7 e.g. $n\,I_1(\theta_0)\cdot(\hat\theta_{MLE}-\theta_0)^2$ ← LN, CH8, p.39

- Application of GLR test I --- tests for multinomial distribution, goodness-of-fit tests

Example 7.16 (GLR test for multinomial distribution)

- Data: $(X_1, \ldots, X_m)$.
  Consider an experiment which may result in m possible different outcomes. In n independent repetitions of the experiment, let X_i be the number of trials which result in the ith outcome.

sample size

- Statistical Modeling. p_i: probability that the ith outcome occurs
  - $(X_1, \ldots, X_m) \sim$ multinomial$(n, p_1, \ldots, p_m)$, where $\sum_{i=1}^m p_i = 1$, and the pmf is:
  $$P(X_1 = x_1, \ldots, X_m = x_m) = \frac{n!}{x_1! \cdots x_m!} {p_1}^{x_1} \cdots {p_m}^{x_m},$$
  where $\sum_{j=1}^m x_j = n$.
  - The parameter space is
  $$\dim = m-1 \rightarrow \Omega = \{\mathbf{p} = (p_1, \ldots, p_m) : p_i \geq 0, i = 1, 2, \ldots, m, \sum_{i=1}^m p_i = 1\}$$
- Problem. For example, fair die, $p_{10} = p_{20} = \cdots = p_{60} = 1/6$

  We suspect that the $\mathbf{p}$ has certain specified values $\mathbf{p}_0 = (p_{10}, \ldots, p_{m0})$, where $p_{10}, \ldots, p_{m0}$ are given values. $\boxed{\text{known}}$

  $\boxed{\mathbf{p} = \mathbf{p}_0 \text{ or } \mathbf{p} \neq \mathbf{p}_0}$
- Problem formulation.

  We can formulate this as a test with null and alternative hypotheses:
  $$\boxed{\text{simple}} \rightarrow H_0 : \mathbf{p} \in \underbrace{\Omega_0}_{\dim=0} = \{(p_{10}, \ldots, p_{m0})\} \quad \text{vs.} \quad H_A : \mathbf{p} \in \Omega \backslash \Omega_0$$

  $\boxed{\text{Q: Why arrange } \Omega_0 \& \Omega_A \text{ in the way?}}$

  $\boxed{\text{guidelines in LN p.15}}$
- Likelihood Ratio Test. $\boxed{\text{LN, CH8, p.23-24}}$
  - $$\Lambda = \frac{\frac{n!}{x_1! \cdots x_m!} {p_{10}}^{x_1} \cdots {p_{m0}}^{x_m}}{\max_{\mathbf{p} \in \Omega} \left[\frac{n!}{x_1! \cdots x_m!} {p_1}^{x_1} \cdots {p_m}^{x_m} \right]} = \frac{\cancel{\frac{n!}{x_1! \cdots x_m!}} p_{10}^{x_1} \cdots p_{m0}^{x_m}}{\cancel{\frac{n!}{x_1! \cdots x_m!}} \hat{p}_1^{x_1} \cdots \hat{p}_m^{x_m}},$$
  where $\hat{p}_i = x_i/n$ is the MLE in Ω.

Therefore, $X_i \sim B(n, p_i) \Rightarrow Var(X_i) = np_i(1-p_i)$ — large when $p = 1/2$; small when $p \approx 0$ or 1
$$\Lambda = \prod_{i=1}^m \left(\frac{np_{i0}}{x_i} \right)^{x_i} \quad \text{and} \quad -2 \log \Lambda = 2 \sum_{i=1}^m x_i \log \left(\frac{\overset{\text{observed counts}}{x_i}}{\underset{\text{expected counts}}{np_{i0}}} \right).$$

$\boxed{\text{reasonable?}}$ $\boxed{\text{Q: Why multipled by } x_i\text{?}}$

$\boxed{\text{Consider, e.g., } n=1000 \quad ① x_i = 2, np_{i0} = 1 \quad ② x_i = 200, np_{i0} = 100}$

- H_0 is rejected if
$$-2 \log \Lambda(X_1, \ldots, X_m) \geq c.$$
- The constant c can be determined by the fact that under H_0,
$$\text{test statistic} \rightarrow -2 \log \Lambda \xrightarrow{D} \chi^2_{m-1} \leftarrow \text{asymptotic null distribution}$$
(i.e., $-2 \log \Lambda$ is asymptotically χ^2_{m-1} distributed when $n \to \infty$) because
$$\dim(\Omega) - \dim(\Omega_0) = (m-1) - 0 = m - 1.$$
$\boxed{\text{Note. Not } m \to \infty \text{ (LN, CH8, p.86)}}$

$\boxed{5/6}$

Question 7.14 (examine distribution assumption in statistical modeling $\Rightarrow$ goodness of fit)

- In statistical modeling, we often see the statement:
$$X_1, \ldots, X_n \text{ are i.i.d. from a distribution with cdf } F(\cdot|\Theta).$$
$\boxed{\text{determine useful \& useless information. Recall: sufficient statistics}}$
- Suppose that the independent and identical assumptions are true, can we examine (using data) whether the distribution assumption (i.e., $F(\cdot|\Theta)$) is reasonable?

original statistical model $\Omega = \{F(\cdot|\Theta)\} \xrightarrow{\text{enlarge}} \Omega = \overbrace{\{F(\cdot|\Theta)\}}^{\Omega_0} \cup \overbrace{\{\text{other distributions}\}}^{\Omega_A}$