Hypothesis testing

• What is hypothesis testing?

Question 7.1 (What is a hypothesis testing question?)

1. **observed data.** $x_1, \ldots, x_n$
2. **statistical modeling.** Regard $x_1, \ldots, x_n$ as a realization of random variables $X_1, \ldots, X_n$, and assign $X_1, \ldots, X_n$ a joint distribution:
   a joint cdf $F(\cdot|\Theta)$, or a joint pdf $f(\cdot|\Theta)$, or a joint pmf $p(\cdot|\Theta)$,
   where $\Theta = (\theta_1, \ldots, \theta_k) \in \Omega$, and $\theta_i's$ are **fixed constants**, but their values are **unknown**.
3. **point estimation.** What is the value of Θ? Find a function of $X_1, \ldots, X_n$, $\widehat{\Theta}$, to estimate Θ or a function of Θ.
4. **hypothesis testing.** Separate Ω into two disjoint sets Ω_0 and Ω_A, i.e.,
   $$\Omega_0 \cap \Omega_A = \emptyset \quad \text{and} \quad \Omega_0 \cup \Omega_A = \Omega.$$
   Use the data $X_1, \ldots, X_n$ to answer the question: which of the two hypotheses,
   $$H_0 : \Theta \in \Omega_0 \quad \text{versus} \quad H_A : \Theta \in \Omega_A$$
   is more favorable.

Question 7.2

Can we obtain an estimate of Θ and accept H_0 if $\hat{\Theta} \in \Omega_0$ and reject H_0 if $\hat{\Theta} \in \Omega_A$?

Hint. Consider the case: $X \sim \text{Binomial}(n, p)$,

$$H_0 : p = 0.5 \quad \text{v.s.} \quad H_1 : p \neq 0.5,$$

What if $n = 10^5$ and $\hat{p} = 0.50001$?

Definition 7.1 (null and alternative hypotheses, simple and composite hypotheses, TBp.331,332,334)

- H_0 is called **null hypothesis**.
- H_A (sometimes denoted as H_1) is called **alternative hypothesis**.
- An hypothesis is said to be a **simple hypothesis** if that hypothesis uniquely specifies the distribution.
- Any hypothesis that is not a simple hypothesis is called a **composite hypothesis**.

Question 7.3 (asymmetry between H_0 and H_A)

Is there a difference between the roles of H_0 and H_A? Can we arbitrary exchange the two hypotheses?

Example 7.1 (some null and alternative hypotheses, TBp.329, 334)

1. Two coins experiment
   - Data and problem
     - Suppose that I have two coins
     - Coin 0 has probability of heads equal to 0.5, and coin 1 has probability of heads equal to 0.7.
     - I choose one of the coins, toss it 10 times, and tell you the number of heads X
     - On the basis of observed X, your task is to decide which coin it was.
   - Statistical modeling.
     - $X \sim \text{Binomial}(10, p)$
     - $\Omega = \{\text{Binomial}(10, 0.5), \text{Binomial}(10, 0.7)\}$
   - Problem formulation
     - H_0: coin 0 $\Rightarrow \Omega_0 = \{\text{Binomial}(10, 0.5)\}$
     - H_A: coin 1 $\Rightarrow \Omega_A = \{\text{Binomial}(10, 0.7)\}$
     - Both hypotheses are simple.

2. Testing for ESP
   - Data and problem
     - A subject is asked to identify, without looking, the suits of 20 cards drawn randomly with replacement from a 52 card deck.
     - Let T be the number of correct identifications.
     - We would like to know whether the person is purely guessing or has extrasensory ability.
   - statistical modeling
     - $T \sim \text{Binomial}(20, p)$.
     - Note that $p = 0.25$ means that the subject is mearly guessing and has no extrasensory ability.
     - $\Omega = \{p : p \in [0.25, 1]\}$ or $\Omega = \{p : p \in [0, 1]\}$
   - Problem formulation
     - $\Omega = [0.25, 1]$
       $$H_0 : p = 0.25 \quad \text{v.s.} \quad H_A : p > 0.25$$
       * Then, $\Omega_0 = \{0.25\}$ and $\Omega_A = (0.25, 1]$.
       * The H_0 is simple and H_A is composite. Furhermore, H_A is called a **one-sided hypothesis**.

– $\Omega = [0,1]$

$$H_0 : p = 0.25 \quad \text{v.s.} \quad H_A : p \neq 0.25$$

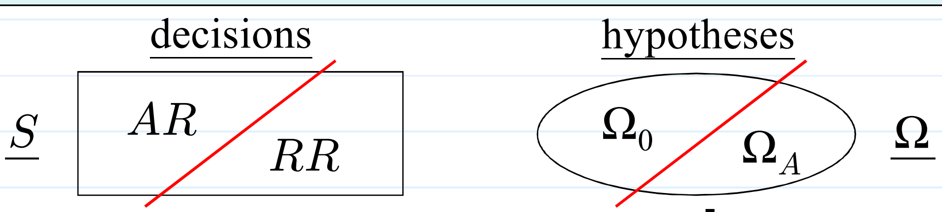
* Then, $\Omega_0 = \{0.25\}$ and $\Omega_A = [0, 0.25) \cup (0.25, 1]$.
* The H_0 is simple and H_A is composite. Furhermore, H_A is called a **two-sided hypothesis**.

3. *goodness-of-fit test* (for Poisson distribution)
   - Data and problem
     – Observe an i.i.d. data $X_1, \ldots, X_n$.
     – Suppose that we only know that X_i's are discrete data.
     – We would like to know whether the observations came from a Poisson distribution.
   - Statistical modeling:
     – $X_1, \ldots, X_n$ are i.i.d. from a discrete distribution with cdf F.
     – $\Omega = \{F : F \text{ is a discrete cdf}\}$
   - Problem formulation
     – H_0 : Data came from some Poisson $\Rightarrow \Omega_0 = \{F : F \text{ is a } P(\lambda) \text{ cdf}\}$
     – H_A : Data not from Poisson $\Rightarrow \Omega_A = \Omega \setminus \Omega_0$.
     – Both hypotheses are composite.

• Neyman-Pearson paradigm --- concept and procedure of hypothesis testing

Definition 7.2 (test, rejection region, acceptance region, test statistic, TBp. 331)

- A **test** is a rule defined on sample space of data (denoted as S) that specifies:
  1. for which sample values $\mathbf{x} \in S$, the decision is made to accept H_0
  2. for which sample values $\mathbf{x} \in S$, reject H_0 in favor of H_A
- The subset of the sample space for which H_0 will be rejected is called the **rejection region**, denoted as RR.
- The complement of the rejection region is called the **acceptance region**, denoted as $AR = S \setminus RR = RR^c$.



Note. Two types of errors may be incurred in applying this paradigm

		decisions	
		accept H_0	reject H_0
hypo-theses	H_0 is true	Bingo!	Type I error
	H_A is true	Type II error	Bingo!

Definition 7.3 (type I error, significance level, type II error, power, TBp. 331)

Let Θ_0 denote the true value of parameters.

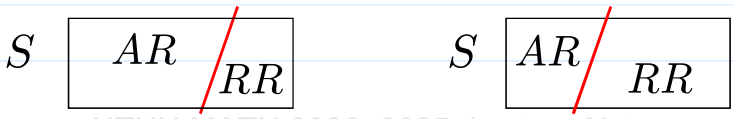
- **Type I error**: reject H_0 when H_0 is ture, i.e., $\mathbf{x} \in RR$ when $\Theta_0 \in \Omega_0$.
  - α_{Θ_0} = probability of Type I error $= P(RR|\Theta_0)$, where $\Theta_0 \in \Omega_0$.
  - **Significance level** α: the maximum (or the least upper bound) of the Type I error probability, i.e., $\alpha = \max_{\Theta \in \Omega_0} \alpha_\Theta$ or $\alpha = \sup_{\Theta \in \Omega_0} \alpha_\Theta$.
- **Type II error**: accept H_0 when H_A is ture, i.e., $\mathbf{x} \in AR$ when $\Theta_0 \in \Omega_A$
  - β_{Θ_0} = probability of Type II error $= P(AR|\Theta_0)$, where $\Theta_0 \in \Omega_A$
  - For $\Theta_0 \in \Omega_A$, $\textbf{power}_{\Theta_0} = 1 - \beta_{\Theta_0} = P(RR|\Theta_0)$ = probability of rejecting H_0 when H_A is ture

Question 7.4 (dilemma between type I and type II errors)

An ideal test would have $\alpha_\Theta = 0$, for $\Theta \in \Omega_0$ and $\beta_\Theta = 0$, for $\Theta \in \Omega_A$; but this can be achieved only in trivial cases. (Example?).

In general, when α decreases, β increases, and vice versa.

smaller α, larger β



larger α, smaller β

A solution to the dilemma

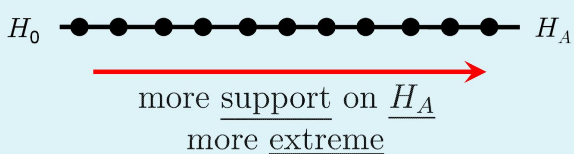
The Neyman-Pearson approach imposes an asymmetry between H_0 and H_A:

1. the significance level α is fixed in advance, usually at a rather small number (e.g. 0.1, 0.05, and 0.01 are commonly used), and
2. then try to construct a test yielding a small value of β_Θ for $\Theta \in \Omega_A$.

Example 7.2 (cont. Ex. 7.1 item 1 (LNp.3), TBp.330-331)

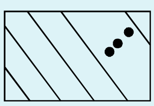
- Testing the value of the parameter p of a Bernoulli distribution with 10 trials. Let X be the number of successes, then $X \sim B(10, p)$.

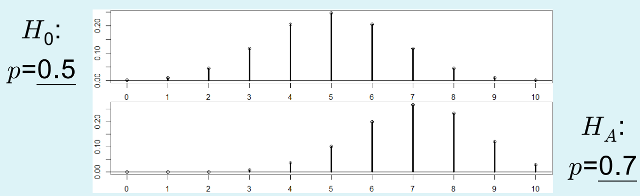
$$H_0 : p = 0.5 \quad \text{vs.} \quad H_A : p = 0.7$$



- *Different observations* of X give *different support* on H_A

$$X = T(X_1, \ldots, X_{10})$$





- In this case, larger observations of X give more support on H_A.
- Rejection region should consist of large values of X.
- The observations that give more support on H_A are called "*more extreme*" observations

- If the rejection region is $\{7, 8, 9, 10\}$, then

$$\alpha = P(X \in \{7, 8, 9, 10\} \,|\, p = 0.5) = 1 - P(X \leq 6 \,|\, p = 0.5) = 0.18.$$

- *Set the significance level as* 0.18.
  - The probability of Type II error (i.e., β) is

$$P(X \notin \{7, 8, 9, 10\} \,|\, p = 0.7) = P(X \leq 6 \,|\, p = 0.7) = 0.35,$$

  - and the power (i.e., $1 - \beta$) is

$$P(X \in \{7, 8, 9, 10\} \,|\, p = 0.7) = 1 - P(X \leq 6 \,|\, p = 0.7) = 0.65.$$

- Suppose that we are interested in testing

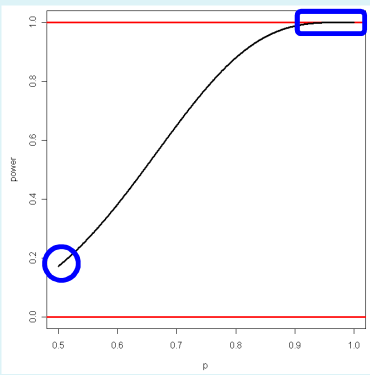
$$H_0 : p = 0.5 \quad \text{vs.} \quad H_A : p > 0.5$$

and the rejection region is still set to be $\{7,\ 8,\ 9,\ 10\}$:

- Note that β_p is a function of p:

$$\beta_p = P(X \notin \{7, 8, 9, 10\} \,|\, p\,).$$

- power function $\equiv 1 - \beta_p$.
- $1 - \beta_p \to 1$ as $p \to 1$, and $1 - \beta_p \to \alpha$ as $p \to 0.5$.
- $1 - \beta_p \to 0$ as $p \to 0$.



Question 7.5

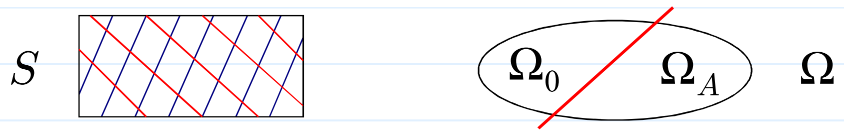
1. Which hypothesis get more *protection*? What kind of protection?
2. Can the protection make P(Type I error) always smaller than P(Type II error), i.e., $\alpha < \beta_\Theta$ for any $\Theta \in \Omega_A$?

Definition 7.4 (test statistic, null distribution, critical value, TBp. 331 & 334)

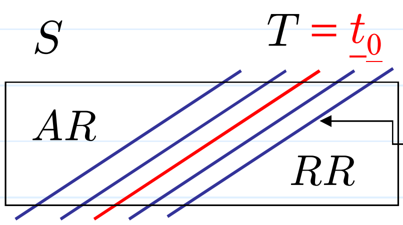
- The **test statistic** is a function of data, denoted as
$$T(X_1, \ldots, X_n),$$
upon which the statistical decision will be based, i.e., the rejection region and acceptance region are defined based on the values of T.
- The distribution of a test statistic *under the null hypothesis* is called **null distribution**.
- If the rejection region is of the form $\{T > t_0\}$ or $\{T < t_0\}$, the number t_0 is called **critical value**, which separate the rejection region and acceptance region.

$T = T(X_1, \ldots, X_n)$ hypotheses

Q1: How to find a good test statistic?



Q2: How to find the critical value?



based on a function of data, T
$T \Rightarrow$ test statistic

Example 7.3 (cont. Ex. 7.2)

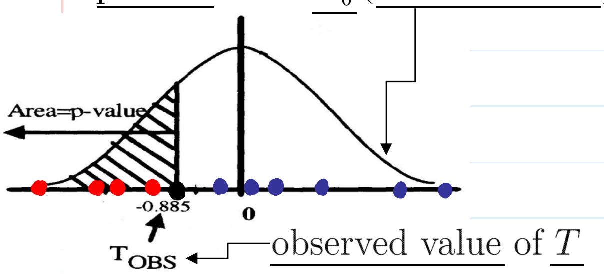
In Ex. 7.2, LNp. 8, the original data $X_1, \ldots, X_{10}$ are i.i.d. from Bernoulli $B(p)$, the test statistics is $X = X_1 + \cdots + X_{10}$, and the null distribution is Binomial(10, 0.5). The critical value can be set to be 6.5.

Question 7.6 (difficulty with level-α tests)

Different persons have different criteria of significance levels, and could come up with opposite conclusions even based on the same observed data.

Example. test statistic: T, and rejection region is of the type: $T < t$,
$\Rightarrow$ smaller values of T give more support on H_A, i.e., cast more doubt on H_0
$\Rightarrow$ on the other hand, larger values of T cast more doubt on H_A
$\Rightarrow$ smaller values of T are more "extreme" (**Note.** More "extreme" values cast more doubts on H_0.)

pdf of T under H_0 (null distribution)



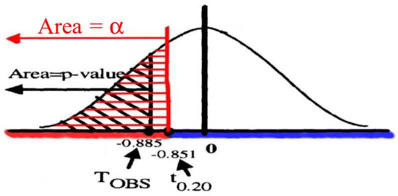
p-value = the probability that a value of T that is as "extreme" or more "extreme" than the observed value of T would occur by chance if the null hypothesis was true.

Note. If p-value is very small (i.e., T_{obs} is a very extreme value), then either H_0 holds and an extremely rare event has occurred, or H_0 is false.

---

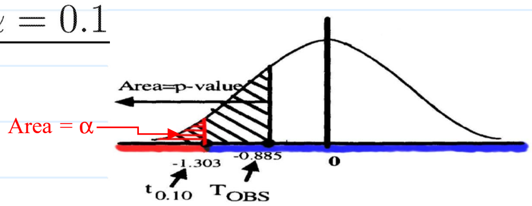
relationship between p-value and $\alpha = P(\text{Type I error})$

$\alpha = 0.2$



$T_{obs} < t_{0.2} \Leftrightarrow p\text{-value} < 0.2$

$\alpha = 0.1$



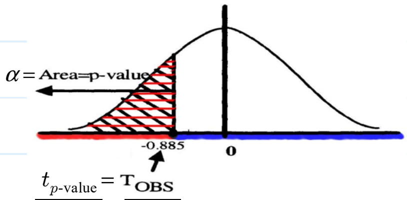
$T_{obs} > t_{0.1} \Leftrightarrow p\text{-value} > 0.1$

Decision rule: $T_{obs} \in$ rejection region $\Leftrightarrow$ p-value $\leq \alpha$

Note 1. The rejection region depends on the value of α, while the p-value depends on the observed value of T

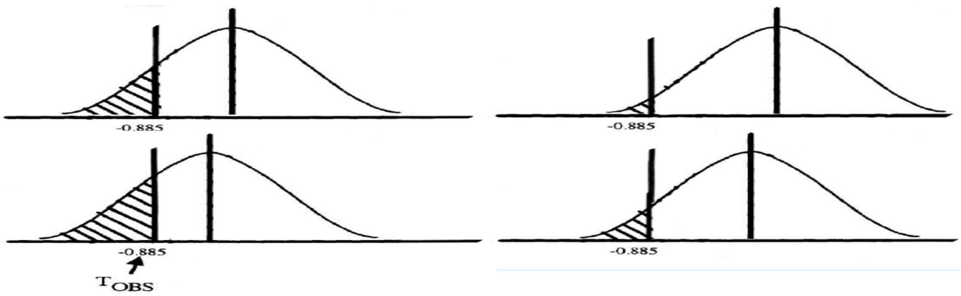
Note 2. It makes more sense to report a p-value than merely report whether the null hypothesis was rejected or not.

Question : Is p-value a statistic?



p-value = the smallest value of α for which the null hypothesis will be rejected.

Question: What if H_0 is composite, i.e., Ω_0 contains more than one distribution?



Definition 7.5 (p-value)

Report the testing result in terms of the "observed significance level":

$$p\text{-value} = \sup_{\Theta \in \Omega_0} P(\text{reject } H_0 \text{ on observing } \mathbf{X} = \mathbf{x} \text{ or more conclusive} | \Theta).$$

Example 7.4 (cont. Ex. 7.2, LNp.8)

If we observed $X = k$, then the p-value is

$$P(X \geq k \,|\, p = 0.5).$$

$\boxed{\alpha = 0.01}$

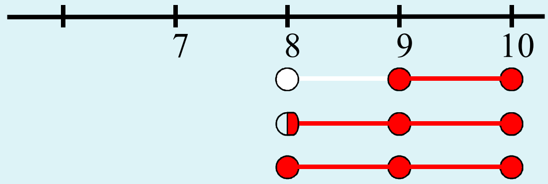
– If we observe 9 successes then the p-value is 0.0107. $\boxed{\alpha = 0.05}$

– If we observe 8 successes then the p-value is 0.0547. $\boxed{\alpha = 0.10}$

$P(X = 10 \,|\, p = 0.5) = 0.0010$

$P(X \geq 9 \,|\, p = 0.5) = 0.0107$

$P(X \geq 8 \,|\, p = 0.5) = 0.0547$



$\alpha = 0.0107$

$\alpha = 0.03$

$\alpha = 0.0547$

Question 7.7 (how to determine H_0 and H_A)

Note that there is an asymmetry between H_0 and H_A. How to determined which of the two hypotheses should be placed in H_0?

Some guidelines. (TBp. 335-336)

1. It is conventional to choose the simpler of two hypotheses as the null hypothesis, e.g. Poisson goodness of fit.
2. The consequences of incorrectly rejecting one hypothesis may be graver than those of incorrectly rejecting the other. In this case, the former should be chosen as the null hypothesis. e.g. screening new drugs.
3. The null hypothesis is often a simple explanation that must be discredited in order to demonstrate the presence of some physical phenomenon or effect. e.g. ESP : $H_0 : p = 0.25$ must be conclusively disproved in order to convince a skeptic that there is any ESP effect.

❖ **Reading**: textbook, 9.1, 9.2, 9.2.1, 9.2.2; **Further reading**: Hogg et al., 5.5

• Neyman Pearson Lemma --- most powerful tests

Question 7.8

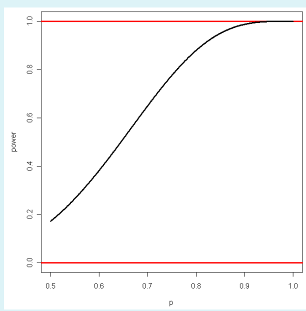
Among all tests with significance level α, what is the "best" test?

Definition 7.6 (uniformly most powerful test, TBp. 336)

For testing

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

a test is called **uniformly most powerful (UMP)** test with significance level α, if $\forall\, \Theta \in \Omega_A$ the power of the test is larger than (i.e., β_Θ is smaller) or equal to the power of any other tests with significance level α.



• most powerful test for simple vs. simple hypotheses

Definition 7.7 (likelihood ratio, TBp.329)

- let $\mathbf{X} = (X_1, \ldots, X_n)$ have a joint pmf/pdf f, where $f \in \Omega = \{f_0, f_A\}$.
- null and alternative hypotheses: $\quad H_0 : f_0 \quad \text{vs.} \quad H_A : f_A$

- likelihood ratio:
$$\frac{f_0(\mathbf{x})}{f_A(\mathbf{x})}$$

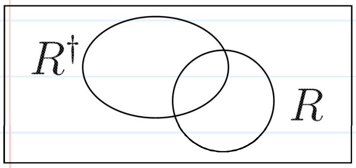
Question: We should reject H_0 when $f_0(\mathbf{x})/f_A(\mathbf{x})$ is large or small? Why?

Theorem 7.1 (Neyman-Pearson lemma, TBp. 332)

Suppose that the likelihood ratio test that rejects H_0 when
$$f_0(\mathbf{x})/f_A(\mathbf{x}) \le c$$
has significance level α. Then any other test which has significance level $\alpha^* \le \alpha$ has power less than or equal to the power of the likelihood ratio test.

Proof. (*for continuous case*) Let $R^\dagger$ be the rejection region of the likelihood ratio test, and R be the rejection region of any other test with significance level $\alpha^* \le \alpha$. Then,
$$\int_{R^\dagger} f_0(x)\,dx = \alpha \ge \alpha^* = \int_R f_0(x)\,dx,$$
and hence the difference of the powers is
$$\begin{aligned}\int_{R^\dagger} f_A(x)\,dx - \int_R f_A(x)\,dx &= \int_{R^\dagger\setminus R} f_A(x)\,dx - \int_{R\setminus R^\dagger} f_A(x)\,dx\\ &\ge \int_{R^\dagger\setminus R} \tfrac{1}{c} f_0(x)\,dx - \int_{R\setminus R^\dagger} \tfrac{1}{c} f_0(x)\,dx\\ &= \tfrac{1}{c}\int_{R^\dagger} f_0(x)\,dx - \tfrac{1}{c}\int_R f_0(x)\,dx \ge 0.\end{aligned}$$

S



Example 7.7 (cont. Ex. 7.2 (LNp.8), TBp. 329, 330)

If $X \sim \text{Binomial}(10, p)$ and consider testing
$$H_0 : p = 0.5 \quad \text{vs.} \quad H_A : p = 0.7.$$
The likelihood ratios are

x	0	1	2	3	4	5	6	7	8	9	10
$f_0(x)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001
$f_A(x)$	.0000	.0001	.001	.009	.037	.103	.200	.267	.234	.121	.028
$f_0(x)/f_A(x)$	165.4	70.9	30.4	13.0	5.58	2.39	1.03	.44	.19	.08	.03

and the likelihood ratio test rejects H_0 for small values of $f_0(x)/f_A(x)$ which corresponding to large values of X.

Example 7.8 (TBp. 333)

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$ with σ^2 known.
$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_A : \mu = \mu_A,$$
where μ_0, μ_A are given constants.
- Suppose the significance level is prescribed as α.
- The likelihood ratio is
$$\frac{f_0(\mathbf{x})}{f_A(\mathbf{x})} = \frac{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu_0)^2\right]}{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu_A)^2\right]} \propto \exp\left[2n\,\overline{X}\,(\mu_0-\mu_A)+n\mu_A^2-n\mu_0^2\right].$$

- Suppose $\mu_0 - \mu_A < 0$ ($\mu_A > \mu_0$). Likelihood ratio is small $\Leftrightarrow$ $\overline{X}$ is large.
- Thus, the most powerful test rejects H_0 for
$$\overline{X} > x_0 \quad \text{for some } x_0.$$
- The x_0 is chosen so that the test has the desired level α
$$\alpha = P(\overline{X} > x_0 \,|\, H_0) = P(\overline{X} > x_0 \,|\, \mu = \mu_0) = P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{x_0 - \mu_0}{\sigma/\sqrt{n}}\right).$$
- Hence, solve
$$\frac{x_0 - \mu_0}{\sigma/\sqrt{n}} = z(\alpha)$$
for x_0, where $z(\alpha)$ is the $1-\alpha$ quantile of the $N(0,1)$ distribution.
- Question: What if $\mu_0 > \mu_A$ (i.e., $\mu_A < \mu_0$)?

Question 7.9

Why not just reject H_0 when $f_0(\mathbf{x})/f_A(\mathbf{x}) \leq 1$?

- UMP examples for testing certain composite hypotheses

Example 7.9 (cont. Ex. 7.8 (LNp.18), TBp.336)

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$, σ^2 known.
$$H_0^{(2)}: \mu = \mu_0 \quad \text{vs.} \quad H_A^{(2)}: \mu > \mu_0$$
- In Ex. 7.8, for any particular simple alternative H_A: $\mu = \mu_A$ ($> \mu_0$), the UMP test rejects H_0 for
$$\overline{X} > x_0,$$
where x_0 **does not depend on the value of** μ_A**.**
- Thus, this test is also UMP for $H_A^{(2)}$: $\mu > \mu_0$.

Example 7.10 (cont. Ex. 7.9, TBp.336)

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$, σ^2 known. Consider testing
$$H_0^{(3)}: \mu \leq \mu_0 \quad \text{vs.} \quad H_A^{(3)}: \mu > \mu_0.$$
- **Claim**: The test $\phi^*(\mathbf{x})$ which rejects H_0 when
$$\overline{X} > x_0,$$
where x_0 is determined by:
$$\alpha = P(\overline{X} > x_0 \,|\, \mu = \mu_0) \quad = \quad P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{x_0 - \mu_0}{\sigma/\sqrt{n}} \,\middle|\, \mu = \mu_0\right)$$
is an level-α UMP test for $H_0^{(3)}$ vs. $H_A^{(3)}$.

- This follows from
  1. ϕ^* is a level-α test because for $\mu \le \mu_0$,

$$\begin{aligned} P\left(\overline{X} > x_0 \,\middle|\, \mu\right) &= P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \ge \frac{x_0-\mu}{\sigma/\sqrt{n}} \,\middle|\, \mu\right) \\ &= 1-\Phi\left(\frac{x_0-\mu}{\sigma/\sqrt{n}}\right) \le 1-\Phi\left(\frac{x_0-\mu_0}{\sigma/\sqrt{n}}\right) = \alpha. \end{aligned}$$

  2. ϕ^* is UMP because if there is another level-α test $\phi(\mathbf{x})$ for testing

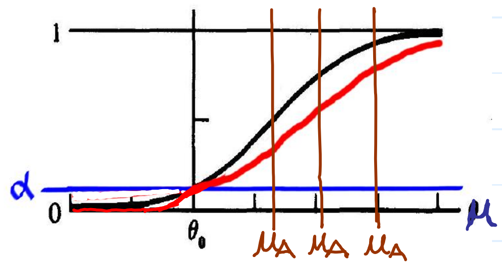
$$H_0^{(3)} : \mu \le \mu_0 \text{ v.s. } H_A^{(3)} : \mu > \mu_0,$$

  then ϕ is also a level-α test for

$$H_0^{(2)} : \mu = \mu_0 \text{ v.s. } H_A^{(2)} : \mu > \mu_0,$$

  and ϕ is always less powerful than ϕ^*.

Plot of $P(RR|\theta)$ $= E_\theta(\phi)$ against θ



$$H_0^{(1)} : \mu = \mu_0 \text{ vs. } H_A^{(1)} : \mu = \mu_A$$

$$H_0^{(2)} : \mu = \mu_0 \text{ vs. } H_A^{(2)} : \mu > \mu_0$$

$$H_0^{(3)} : \mu \le \mu_0 \text{ vs. } H_A^{(3)} : \mu > \mu_0$$

Definition 7.7 (test function, nonrandomized test, randomized test)

- For a (nonrandomized) test, its **test function** $\phi(\mathbf{x})$ is
  - a function defined on sample space S and
  - taking values in $[0,1]$ such that

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \text{rejection region} \\ 0, & \text{if } \mathbf{x} \in \text{acceptance region} \end{cases}$$

- A *randomized* test function $\phi(\mathbf{x})$:
  - It has a form of:

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \text{rejection region} \\ \gamma, & \text{if } \mathbf{x} \in \text{boundry of rejection and acceptance regions} \\ 0, & \text{if } \mathbf{x} \in \text{acceptance region} \end{cases}$$

    where $0 < \gamma < 1$.
  - When $\phi(\mathbf{x}) = \gamma$, a coin, whose probability of falling head is γ, is tossed and H_0 is rejected or accepted when head or tail appears, respectively.

Theorem 7.2 (significance level, power, and test function)

$$E_\Theta\left[\phi(\mathbf{X})\right] = \begin{cases} \alpha_\Theta, & \text{for } \Theta \in \Omega_0 \\ 1-\beta_\Theta, & \text{for } \Theta \in \Omega_A \end{cases}$$

• UMP tests for one-parameter exponential family

Recall. A one-parameter exponential family has the joint pmf/pdf of the form

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

where the set A is independent of θ.

Theorem 7.3 (UMP test for one-parameter exponential family, one-sided hypothesis)

- Suppose $\mathbf{X} = (X_1, \cdots, X_n)$ have a joint pdf/pmf $f(\mathbf{x}|\theta)$ which has the form of the one-parameter exponential family.
- Suppose that $C(\theta)$ is a strictly monotone increasing function of θ.
- Then the level-α UMP test for testing
$$H_0 : \theta \le \theta_0 \quad (\text{or } H_0 : \theta = \theta_0) \quad \text{vs.} \quad H_A : \theta > \theta_0$$
is given by
$$\phi(\mathbf{X}) = \begin{cases} 1, & \text{if } T(\mathbf{X}) > c, \\ \gamma, & \text{if } T(\mathbf{X}) = c, \\ 0, & \text{if } T(\mathbf{X}) < c. \end{cases}$$
- The c and γ are determined by
$$E_{\theta_0}(\phi) = P(T(\mathbf{X}) > c \,|\, \theta = \theta_0) + \gamma\, P(T(\mathbf{X}) = c \,|\, \theta = \theta_0) = \alpha.$$

Proof. The proof follows the same argument as in Examples 7.8 to 7.10.

Question 7.10

What if we are interested in $H_0 : \theta \ge \theta_0$ vs. $H_A : \theta < \theta_0$?

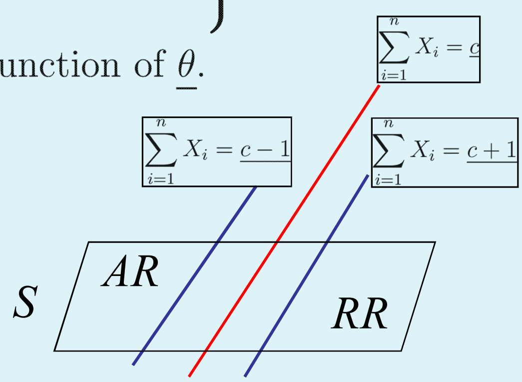
$$\begin{aligned} \theta^* &= -\theta \\ C^*(\theta^*) &= -C(-\theta^*) \\ T^*(\mathbf{x}) &= -T(\mathbf{x}) \end{aligned}$$

$H_0 : \theta^* \le -\theta_0$ vs. $H_A : \theta^* > -\theta_0$ | $C^*(\theta^*)\uparrow$, as $\theta^*\uparrow$

$$\phi(\mathbf{X}) = \begin{cases} 1, & \text{if } T^*(\mathbf{X}) > c^* \Leftrightarrow T(\mathbf{X}) < -c^* \\ \gamma, & \text{if } T^*(\mathbf{X}) = c^* \Leftrightarrow T(\mathbf{X}) = -c^* \\ 0, & \text{if } T^*(\mathbf{X}) < c^* \Leftrightarrow T(\mathbf{X}) > -c^* \end{cases}$$

Example 7.11 (UMP for i.i.d. Bernoulli)

- Let $\mathbf{X} = (X_1, \cdots, X_n)$ be i.i.d. from $B(1, \theta)$, $\theta \in (0, 1)$. Then, the joint pmf is
$$f(\mathbf{x}|\theta) = \exp\left\{\log\left(\frac{\theta}{1-\theta}\right)\sum_{i=1}^n x_i + n\log(1-\theta)\right\} I_{\{0,1\}^n}(\mathbf{x}).$$
- Here, $C(\theta) = \log\left(\frac{\theta}{1-\theta}\right)$ is a strictly increasing function of θ.
- The level-α UMP test for
$$H_0 : \theta \le \theta_0 \quad \text{vs.} \quad H_A : \theta > \theta_0$$
is given by
$$\phi(\mathbf{X}) = \begin{cases} 1, & \text{if } \sum_{i=1}^n X_i > c, \\ \gamma, & \text{if } \sum_{i=1}^n X_i = c, \\ 0, & \text{if } \sum_{i=1}^n X_i < c. \end{cases}$$



- The c and γ are determined by

$$E_{\theta_0}(\phi) = P\left(\sum_{i=1}^n X_i > c \,\middle|\, \theta=\theta_0\right) + \gamma P\left(\sum_{i=1}^n X_i = c \,\middle|\, \theta=\theta_0\right) = \alpha,$$

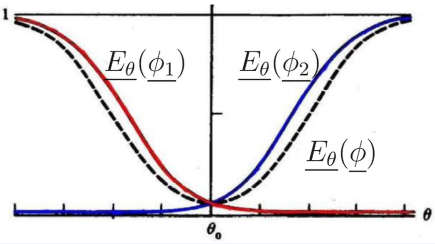
where $\sum_{i=1}^n X_i \sim \text{Binomial}(n, \theta_0)$.

❖ **Reading**: textbook, 9.1, 9.2, 9.2.3; **Further reading**: Roussas, 13.1, 13.2, 13.3

• UMPU tests (for one-parameter exponential family)

Question 7.11

Does a UMP test exist for two-sided alternative hypothesis?



- - - - ϕ: a "good" test for $H_0: \theta=\theta_0$ vs. $H_A: \theta\neq\theta_0$

Question: How should the power function of ϕ look like?

—— ϕ_1: UMP for $H_0: \theta=\theta_0$ (or $\theta\geq\theta_0$) vs. $H_A: \theta<\theta_0$

—— ϕ_2: UMP for $H_0: \theta=\theta_0$ (or $\theta\leq\theta_0$) vs. $H_A: \theta>\theta_0$

Definition 7.8 (unbiased test)

A level-α test function ϕ for $H_0: \Theta\in\Omega_0$ versus $H_A: \Theta\in\Omega_A$ is said to be **unbiased** if $E_\Theta(\phi)$ satisfies

$$E_\Theta(\phi)\leq\alpha \text{ if } \Theta\in\Omega_0 \quad\text{and}\quad E_\Theta(\phi)\geq\alpha \text{ if } \Theta\in\Omega_A.$$

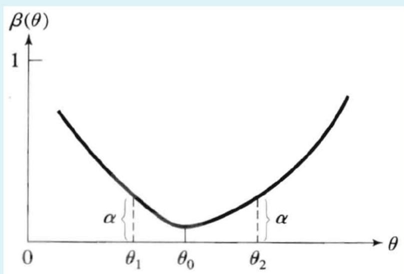
Theorem 7.4 (UMPU tests for one-parameter exponential family, two-sided hypothesis)

- Suppose $\mathbf{X}=(X_1,\cdots,X_n)$ have a joint pdf/pmf $f(\mathbf{x}|\theta)$ which has the form of the one-parameter exponential family.
- Suppose that $C(\theta)$ is a strictly monotone increasing function of θ.
- Null and alternative hypotheses:

$$H_0: \theta_1\leq\theta\leq\theta_2 \quad\text{vs.}\quad H_A: \theta<\theta_1 \text{ or } \theta>\theta_2$$

- The level-α uniformly most powerful unbiased test is given by

$$\phi(\mathbf{X}) = \begin{cases} 1, & \text{if } T(\mathbf{X}) < c_1 \text{ or } > c_2 \\ \gamma_i, & \text{if } T(\mathbf{X}) = c_i, \quad i=1,2 \\ 0, & \text{if } c_1 < T(\mathbf{X}) < c_2 \end{cases} \qquad (1)$$



- The c_1, c_2 and γ_1, γ_2 are determined by

$$E_{\theta_1}[\phi(\mathbf{X})] = E_{\theta_2}[\phi(\mathbf{X})] = \alpha.$$

- **Note**. When $\theta_1 = \theta_2$,
  - Null and alternative hypotheses:
  $$H_0 : \theta = \theta_1 \text{ against } H_A : \theta \neq \theta_1$$
  - The form of the UMP unbiased test is the same as (1) in LNp.26.
  - The constants c_1, c_2 and γ_1, γ_2 are now determined by
  $$E_{\theta_1}[\phi(\mathbf{X})] = \alpha,$$
  and
  $$E_{\theta_1}[\phi(\mathbf{X})T(\mathbf{X})] = E_{\theta_1}[\phi(\mathbf{X})]\, E_{\theta_1}[T(\mathbf{X})] = \alpha\, E_{\theta_1}[T(\mathbf{X})].$$

Example 7.12 (UMPU test for normal mean)

- Let $\mathbf{X} = (X_1, \cdots, X_n)$ be i.i.d. from $N(\mu, \sigma^2)$, where σ is known.
- The joint pdf is
$$f(\mathbf{x}|\mu) = \exp\left\{\frac{n\mu}{\sigma^2}\,\bar{x} - \frac{2n\mu^2}{\sigma^2} - \frac{\sum_{i=1}^n x_i^2}{2\sigma^2} - n\log(\sqrt{2\pi})\right\}.$$
  which belongs to one-parameter exponential family.
- In this case, $C(\mu) = n\mu/\sigma^2$, which is an increasing function of μ.
- Null and alternative hypotheses:
$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_A : \mu \neq \mu_0$$
  where μ_0 is a given value.

- The level-α UMP unbiased test is given by
$$\phi(\mathbf{X}) = \begin{cases} 1, & \text{if } \overline{X} < c_1 \text{ or } > c_2, \\ 0, & \text{if } c_1 < \overline{X} < c_2. \end{cases}$$
- The c_1, c_2 are determined by
$$E_{\mu_0}[\phi(\mathbf{X})] = \alpha, \quad E_{\mu_0}[\phi(\mathbf{X})\,\overline{X}] = \alpha\, E_{\mu_0}(\overline{X}).$$
- Further derivation:
  - The $\phi(\mathbf{X})$ can be expressed as
  $$\phi(\mathbf{X}) = \begin{cases} 1, & \text{if } \frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma} < c_1' \text{ or } > c_2', \\ 0, & \text{if } c_1' < \frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma} < c_2'. \end{cases}$$
  where $c_i' = \sqrt{n}(c_i - \mu_0)/\sigma$.
  - Because the pdf of $N(0,1)$ is symmetry about 0, $c_1' = -c_2'$.
  - Therefore, we have
  $$\phi(\mathbf{X}) = \begin{cases} 1, & \text{if } \left[(\overline{X} - \mu_0)/(\sigma/\sqrt{n})\right]^2 > c, \\ 0, & \text{otherwise.} \end{cases}$$
  - The constant c is determined by $P(\chi_1^2 > c) = \alpha$ since under H_0,
  $$\left[(\overline{X} - \mu_0)/(\sigma/\sqrt{n})\right]^2 \sim \chi_1^2.$$

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.
- The joint pdf is $f(\mathbf{x}|\sigma) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - n\log\sqrt{2\pi}\right\}$.

  which belongs to one-parameter exponential family.
- Here, $C(\sigma) = -1/(2\sigma^2)$ is an increasing function of σ.
- Null and alternative hypotheses:

$$H_0 : \sigma = \sigma_0 \quad \text{vs.} \quad H_A : \sigma \neq \sigma_0$$

- The level-α UMP unbiased test is given by

$$\phi(\mathbf{X}) = \begin{cases} 1, & \text{if } T \equiv \sum_{i=1}^{n}(X_i-\mu)^2/\sigma_0^2 < c_1 \text{ or } > c_2, \\ 0, & \text{otherwise.} \end{cases}$$

- The c_1, c_2 are determined by

$$E_{\sigma_0}(1-\phi) = \int_{c_1}^{c_2} f_n(y)\,dy = 1-\alpha,$$

  and

$$E_{\sigma_0}[(1-\phi)\,T] = \int_{c_1}^{c_2} y\,f_n(y)\,dy = E_{\sigma_0}(1-\phi)\,E_{\sigma_0}(T) = n\,(1-\alpha),$$

  where $f_n(y)$ is the pdf of the χ^2_n 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

$$\int_{c_1}^{c_2} f_{n+2}(y)\,dy = 1-\alpha.$$

  - 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.
    - ∗ This follows from the fact that χ^2_n distribution tends to Normal distribution for large n by CLT.

Definition 7.9 (monotone likelihood ratio)

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

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

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

Theorem 7.5

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

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

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

• (Generalized) likelihood ratio tests

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)

- Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ have a joint pdf/pmf $f(\mathbf{x}|\theta)$, where $\theta \in \Omega$.
- Consider testing the hypotheses

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

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

$$\Lambda^*(\mathbf{x}) = \frac{\max_{\theta \in \Omega_0} \mathcal{L}(\theta, \mathbf{x})}{\max_{\theta \in \Omega_A} \mathcal{L}(\theta, \mathbf{x})},$$

where $\mathcal{L}(\theta, \mathbf{x})$ is the likelihood function.
- Small values of Λ^* tend to discredit H_0.
- For technical reasons, it is preferable to use

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

- 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.
- Consider the hypotheses $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$.

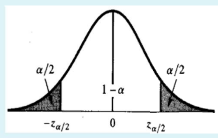
  Then $\Omega_0 = \{\mu_0\}$, $\Omega_A = \{\mu : \mu \neq \mu_0\}$, $\Omega = \{-\infty < \mu < \infty\}$.
- The LR statistic is

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

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

$$-2\log(\Lambda) = \frac{(\overline{x} - \mu_0)^2}{\sigma^2/n}.$$

- Under H_0, $\overline{X} \sim N(\mu_0, \sigma^2/n)$ and $-2\log\Lambda \sim \chi_1^2$.
- Thus, the LR test rejects when



$$\frac{(\overline{X} - \mu_0)^2}{\sigma^2/n} > \chi_1^2(\alpha) \quad \text{or equivalently} \quad \frac{|\overline{X} - \mu_0|}{\sigma/\sqrt{n}} > z(\alpha/2).$$

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

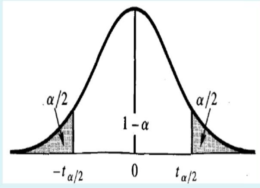
- Let $X_1, \ldots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$, where μ and σ are unknown.
- Consider the hypotheses $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$.

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

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

  where $\hat{\sigma}_0^2 = \sum_{i=1}^n (X_i - \mu_0)^2/n$ and $\hat{\sigma}_1^2 = \sum_{i=1}^n (X_i - \overline{X})^2/n$.

- Now,
$$\Lambda = \left[1 + (\overline{X} - \mu_0)^2/\hat{\sigma}_1^2\right]^{-n/2}.$$
- Thus the LR test rejects H_0 for small values of Λ, i.e., large values of
$$\left(\frac{\overline{X} - \mu_0}{\hat{\sigma}_1}\right)^2.$$
- Under H_0, $\sqrt{n-1}\,(\overline{X} - \mu_0)/\hat{\sigma}_1 \sim t_{n-1}$.
- Thus, the level-α LR test rejects H_0 when
$$\left|\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2/(n-1)}}\right| > t_{n-1}(\alpha/2),$$



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

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.
Q: *Similar result exist for the null distribution of GLR statistic*?

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
$$-2\log(\Lambda)$$
tends to a χ^2_m distribution, with degree of freedom
$$m = \dim(\Omega) - \dim(\Omega_0),$$
as the sample size tends to infinity.

(**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, $\dim(\Omega) = 1$, $\dim(\Omega_0) = 0$, $m = 1 - 0 = 1$.)

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

❖ **Reading**: textbook, 9.4; **Further reading**: Roussas, 13.7

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

Example 7.16 (GLR test for multinomial distribution)

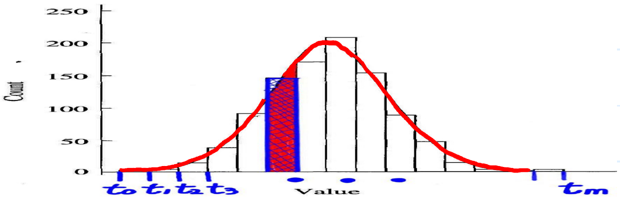
- Data: $(X_1, \ldots, X_m)$.
  Consider an experiment which may result in m possible different outcomes. In n independent repetitions of the experiment, let X_i be the number of trials which result in the ith outcome.

- Statistical Modeling.
  - $(X_1, \ldots, X_m) \sim$ 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
  $$\Omega = \{\mathbf{p} = (p_1, \ldots, p_m) : p_i \geq 0, i = 1, 2, \ldots, m, \ \sum_{i=1}^m p_i = 1\}$$
- Problem.
  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.
- Problem formulation.
  We can formulate this as a test with null and alternative hypotheses:
  $$H_0 : \mathbf{p} \in \Omega_0 = \{(p_{10}, \ldots, p_{m0})\} \quad \text{vs.} \quad H_A : \mathbf{p} \in \Omega \backslash \Omega_0$$
- Likelihood Ratio Test.
  - $$\Lambda = \frac{\frac{n!}{x_1! \cdots x_m!} p_{10}^{x_1} \cdots p_{m0}^{x_m}}{\max_{\mathbf{p} \in \Omega} \left[\frac{n!}{x_1! \cdots x_m!} p_1^{x_1} \cdots p_m^{x_m} \right]} = \frac{\frac{n!}{x_1! \cdots x_m!} p_{10}^{x_1} \cdots p_{m0}^{x_m}}{\frac{n!}{x_1! \cdots x_m!} \hat{p}_1^{x_1} \cdots \hat{p}_m^{x_m}},$$
  where $\hat{p}_i = x_i/n$ is the MLE in Ω.

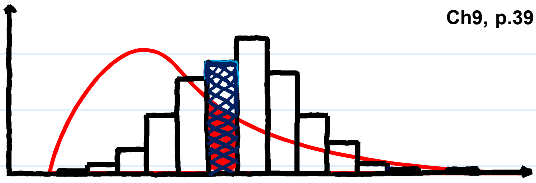
Therefore,
$$\Lambda = \prod_{i=1}^m \left(\frac{np_{i0}}{x_i} \right)^{x_i} \quad \text{and} \quad -2 \log \Lambda = 2 \sum_{i=1}^m x_i \log \left(\frac{x_i}{np_{i0}} \right).$$
  - H_0 is rejected if
  $$-2 \log \Lambda(X_1, \ldots, X_m) \geq c.$$
  - The constant c can be determined by the fact that under H_0,
  $$-2 \log \Lambda \xrightarrow{D} \chi^2_{m-1}$$
  (i.e., $-2 \log \Lambda$ is asymptotically χ^2_{m-1} distributed when $n \to \infty$) because
  $$\dim(\Omega) - \dim(\Omega_0) = (m-1) - 0 = m - 1.$$

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

- In statistical modeling, we often see the statement:
  $X_1, \ldots, X_n$ are i.i.d. from a distribution with cdf $F(\cdot|\Theta)$.
- Suppose that the independent and identical assumptions are true, can we examine (using data) whether the distribution assupmtion (i.e., $F(\cdot|\Theta)$) is reasonable?



Ω_0 is true



Ω_0 is false

- Observe data $X_1, \ldots, X_n$.
- Let O_i, $i = 1, \ldots, m$, be the number of $X_1, \ldots, X_n$ that fall in $(t_{i-1}, t_i]$.
- Then,

$$O_1, \ldots, O_m \sim \text{Multinomial}(n, p_1, \ldots, p_m),$$

where p_i is the "red area" in the graph when Ω_0 is true.

Example 7.17 (GLR tests for goodness-of-fit, TBp.341-342)

- Ω: the vector of cell probabilities $\mathbf{p} = (p_1, \ldots, p_m)$ that are free except for the constraints that $p_i \geq 0$, $\sum_{i=1}^m p_i = 1$.
  - Note that $\dim(\Omega) = m - 1$.
- H_0: $\Omega_0 \subset \Omega$,

$$\Omega_0 = \{\mathbf{p}(\Theta) : \mathbf{p}(\Theta) = (p_1(\Theta), \ldots, p_m(\Theta))\}$$

is the vector of cell probabilities from $F(\cdot|\Theta)$, where Θ is a k-dimensional unknown parameter, i.e., $p_i(\Theta) = F(t_i|\Theta) - F(t_{i-1}|\Theta)$.
  - Note that $\dim(\Omega_0) = k$.

- Goodness-of-fit tests judge plausibility of the models in H_0 relative to H_A: $\Omega \backslash \Omega_0$ using data $O_1, \ldots, O_m$.
- Likelihood ratio test
  - The LR statistic is

$$\Lambda = \frac{\max_{\mathbf{p} \in \Omega_0} \left[\frac{n!}{o_1! o_2! \cdots o_m!} \, p_1(\Theta)^{o_1} \cdots p_m(\Theta)^{o_m} \right]}{\max_{\mathbf{p} \in \Omega} \left(\frac{n!}{o_1! o_2! \cdots o_m!} \, p_1^{o_1} \cdots p_m^{o_m} \right)} = \prod_{i=1}^m \left(\frac{p_i(\hat{\Theta})}{\hat{p}_i} \right)^{o_i}$$

  where
    - $\hat{\Theta}$ is the MLE of Θ, and
    - $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_m$ are the unrestricted MLE, i.e.,

$$\hat{p}_i = o_i / n \, , \; i = 1, 2, \ldots, m.$$

  - Then, since $O_i = n\hat{p}_i$,

$$-2 \log \Lambda = 2 \sum_{i=1}^m n\hat{p}_i \log \left(\frac{n\hat{p}_i}{np_i(\hat{\Theta})} \right) = 2 \sum_{i=1}^m O_i \log \left(\frac{O_i}{E_i} \right),$$

  where
    - $O_i = n\hat{p}_i$ is the *observed* counts, and
    - $E_i = np_i(\hat{\Theta})$ is the *expected* counts.
  - H_0 is rejected if $-2 \log \Lambda \geq c$.

– To decide the constant c, since

$$\dim(\Omega) - \dim(\Omega_0) = m - k - 1,$$

under H_0, the large sample (i.e., n is large) distribution of $-2\log\Lambda$ is χ^2_{m-k-1}, i.e., $c = \chi^2_{m-k-1}(\alpha)$.

Question 7.15

Compare
- the Ω_0 and Ω in item 3, LNp.5, and
- the Ω_0 and Ω in Ex.7.17, LNp.39.

Are they different? What and Why different?

Example 7.18 (Pearson's Chi-square test, TBp.342-343)

- For the same H_0 vs. H_A and data $O_1, \ldots, O_m$ in Ex.7.17, LNp.39, *Pearson's chi-square test* for goodness of fit:
  – The test statistic is:
  $$X^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}.$$
  – H_0 is rejected if $X^2 > c$.
  – Under H_0, $X^2 \xrightarrow{D} \chi^2_{m-k-1}$.

- Under H_0, the X^2 statistic is asymptotically equivalent (note: $n \to \infty$) to the LR test statistic given in LNp.40 because:

$$\begin{aligned}
-2\log\Lambda &= 2\,n \sum_{i=1}^{m} \hat{p}_i \log\left(\frac{\hat{p}_i}{p_i(\hat{\theta})}\right) \\
&= 2n\sum_{i=1}^{m}\left[\hat{p}_i - p_i(\hat{\theta})\right] + n\sum_{i=1}^{m}\frac{1}{p_i(\hat{\theta})}\left[\hat{p}_i - p_i(\hat{\theta})\right]^2 + \cdots \\
&= 0 + \sum_{i=1}^{m}\frac{\left(n\,\hat{p}_i - n\,p_i(\hat{\theta})\right)^2}{n\,p_i(\hat{\theta})} + \cdots \\
&\approx 0 + X^2
\end{aligned}$$

  – by Taylor expansion of $x\log(x/x_0)$ about x_0
  $$f(x) = x\log(x/x_0) = 0 + (x - x_0) + \frac{1}{2}\frac{1}{x_0}(x - x_0)^2 + \cdots$$
  – and, the fact that when H_0 is true and n is large, $\hat{p}_i \approx p_i(\hat{\theta})$.
- **Note:** Pearson's chi-square test is more commonly used than the LR test, since it is easier to calculate.

Remarks.

1. There is a distinction between $O_1, \ldots, O_m$ and $X_1, \ldots, X_n$ (especially for continuous case)
   - $X_1, \ldots, X_n \Rightarrow X_{(1)}, \ldots, X_{(n)} \Rightarrow O_1, \ldots, O_m$
   - $O_1, \ldots, O_m \Rightarrow X_{(1)}, \ldots, X_{(n)} \Rightarrow X_1, \ldots, X_n$
2. The MLE of θ based on $O_1, \ldots, O_m$ can be different from the MLE of θ based on $X_1, \ldots, X_n$.
3. Different choices of $(t_{i-1}, t_i]$, $i = 1, \ldots, m$, can cause different results. (**Note.** The choice should not depend heavily on observed data.)
4. It is recommended that O_i, $E_i \geq 5$.

Example 7.19 (Hardy-Weinberg Equilibrium, TBp.343-344, or Ex.6.15, LN, Ch8, p.24)

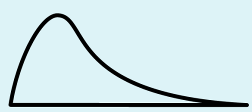
- $n = 1029$, the cell probabilities are $(1-\theta)^2$, $2\theta(1-\theta)$, θ^2 under the Hardy-Weinberg Equilibrium model and the MLE of θ is $\hat{\theta} = 0.4247$.

	Blood Type		
	M	MN	N
O_i	342	500	187
E_i	340.6	502.8	185.6

- Consider the test:

  H_0: $(p_1(\theta), p_2(\theta), p_3(\theta))$ are specified by the Hardy-Weihberg model

  H_A: (p_1, p_2, p_3) do not have that specified form
- Pearson's chi-square test:
  1. Pearson's chi-square test statistic is

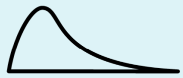
     $$X^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} \xrightarrow{D} \chi_1^2 \quad \text{under } H_0.$$
  2. Set $\alpha = 0.05$. Thus, reject H_0 if the value of X^2 statistic exceeds 3.84, the 95%-quantile of the χ_1^2 distribution.
  3. Since
     $$X^2 = \frac{(342 - 340.6)^2}{340.6} + \frac{(500 - 502.8)^2}{502.8} + \frac{(187 - 185.6)^2}{185.6} \doteq 0.0319,$$
     H_0 is not rejected.

2. Why $\alpha = 0.05$? Will conclusion be different if we choose other α?
   The p-value is more useful:

$$p\text{-value} = P_{H_0}(X^2 > 0.0319) = P(\chi_1^2 \geq 0.0319) = 0.86.$$

   If the null model were correct, deviations this large or larger would occur 86% of the time. Thus, the data give us no reason to doubt the null model.

- Likelihood ratio statistic is



$$-2\log\Lambda = 2\sum_{i=1}^{3} O_i \log\left(\frac{O_i}{E_i}\right) = 0.032 \quad (\approx 0.0319 = X^2).$$

The two tests leads to the same conclusion.

Note: $\Lambda = \exp(0.032/(-2)) = 0.98 \approx 1$ $(0 \leq \Lambda \leq 1)$. Hardy-Weinberg model is almost as likely as the most general possible model.

Example 7.20 (Bacterial Clumps, TBp. 344-345)

- In testing milk for bacterial contamination, 0.01mL of milk is spread over an area of 1cm^2. 400 counts of bacterial clumps:

number per 1cm^2	0	1	2	$\cdots$	9	10	19
Frequency	56	104	80	$\cdots$	3	2	1

- H_0: The data are from Poisson $P(\lambda)$

- MLE for the λ of Poisson model (H_0) is

$$\hat{\lambda} = \frac{0 \times 56 + 1 \times 104 + \cdots + 19 \times 1}{400} = 2.44,$$

giving the expected frequencies E_i in the following table.

number per 1cm^2	0	1	2	$\cdots$	6	≥ 7
O_i (Obs. freq.)	56	104	80	$\cdots$	9	20
E_i (Exp. freq.)	34.9	85.1	103.8	$\cdots$	10.2	5.0
Component of χ^2	12.8	4.2	5.5	$\cdots$	0.14	45.0

- The chi-square statistic is $X^2 = 75.6 > 18.55 = \chi_6^2(0.005)$. So, p-value < 0.005 and the goodness of fit test rejects the Poisson model (H_0).
- 1. bacteria held by surface tension on lower surface of the drop may adhere to the glass slide on contact.
  2. film not of uniform thickness.

Example 7.21 (Fisher's reexamination of Mendel's data, TBp. 345-346)

- Mendel crossed 556 smooth, yellow male peas with wrinkled, green female peas.

Type	Observed Count	Expected Count	Probability
Smooth yellow	315	312.75	9/16
Smooth green	108	104.25	3/16
Wrinkled yellow	102	104.25	3/16
Wrinkled green	31	34.75	1/16

- For the data,
  $-2\log\Lambda = 2\sum_{i=1}^{4} O_i \log(O_i/E_i) = 0.618$,
  $\Lambda = \exp(-0.618/2) = 0.73$, the p-value is slightly less than 0.9. (Peason's statistic is $X^2 = 0.604$.)
- Fisher pooled the results of all of Mendel's experiments:
  - Two independent experiments give chi-square statistic T_1, T_2 with p and r degrees of freedom under H_0.
  - Under H_0, $T_1 + T_2 \sim \chi^2_{p+r}$.
  - Adding all the chi-square statistics for all the independent experiments gives p-value=0.99996!
- The best explanation is perhaps that Mendel continued experimenting until the results looked good. The statistical analysis here assume n is fixed before data are collected.

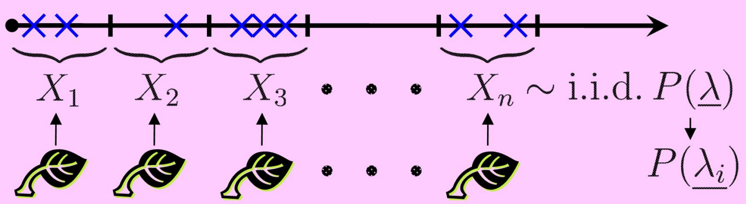
- Dorfman (1978) studied the goodness of fit of the intelligence scores of fathers and sons to a normal distribution (H_0) using Pearson's chi-square test. The p-values were greater than $1 - 10^{-7}$ and $1 - 10^{-6}$, respectively.

❖ **Reading**: textbook, 9.5

- Application of GLR test II --- Poisson dispersion test

Question 7.16

- Recall the insect counts example (Ex. 6.31, LN, Ch8, p.68), the Poisson model did not fit well.



- Poisson model assumptions:
  1. The rate is constant.
  2. Counts in one interval are independent of counts in disjoint intervals.
- For counts of insects on leaves, some assumptions may be violated, e.g.,
  - Leaves are of different sizes and occur at various locations on different plants. Hence, 1. may fail.
  - If the insects hatched from eggs that were deposited in groups, there may be clustering of the insects. Then, 2. may fail.
- How to examine whether the rate is a constant for the insect data?

Example 7.22 (GLR test for Poisson dispersion, TBp. 347-348)

- statistical modeling for solving the question:
  - $X_i \sim P(\lambda_i), i = 1, \ldots, n.$
  $$\Omega = \{(\lambda_1, \ldots, \lambda_n) : \lambda_i > 0, i = 1, \ldots, n\} \qquad \Rightarrow \dim(\Omega) = n.$$
  - Null hypothesis H_0: the counts are Poisson with common parameter λ.
  $$\Omega_0 = \{(\lambda_1, \ldots, \lambda_n) : \lambda_1 = \cdots = \lambda_n \equiv \lambda\} \qquad \Rightarrow \dim(\Omega_0) = 1.$$
  - Alternative hypothesis H_A: the counts are Poisson with different rates $\lambda_1, \lambda_2, \ldots, \lambda_n$, i.e., $\Omega_A = \Omega \backslash \Omega_0$.
- Under Ω_0, MLE of λ is $\hat{\lambda} = \overline{X}$.
- Under Ω, MLE of λ_i's are X_i's, denoted by $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_n$.
- Thus the likelihood ratio is

$$\Lambda = \frac{\prod_{i=1}^n \hat{\lambda}^{x_i} e^{-\hat{\lambda}}/x_i!}{\prod_{i=1}^n \tilde{\lambda}_i^{x_i} e^{-\tilde{\lambda}_i}/x_i!} = \prod_{i=1}^n \left(\frac{\overline{x}}{x_i}\right)^{x_i} e^{x_i - \overline{x}}$$

$$-2\log\Lambda = 2\sum_{i=1}^n \left[x_i \log\left(\frac{x_i}{\overline{x}}\right) + (x_i - \overline{x})\right] = 2\sum_{i=1}^n x_i \log\left(\frac{x_i}{\overline{x}}\right)$$

$$\approx \frac{1}{\overline{x}} \sum_{i=1}^n (x_i - \overline{x})^2 = \frac{n\,\hat{\sigma}^2}{\overline{x}}$$

  - $\hat{\sigma}^2/\overline{x}$: measure of clustering
  - Null Poisson model: variance=mean.
  - Alternative Poisson model: variance>mean.
  - (asymptotic) null distribution: χ^2_{n-1}
- Poisson dispersion test (for goodness-of-fit)
  - has high power against alternative that are overdispersed relative to Poisson,
  - often used when there is not enough data to be accumulated into several cells.

Example 7.23 (Asbestos Fibers, Poisson dispersion test, TBp. 348)

- For the data in Ex. 6.4, LN, Ch8, p.9,

$n\hat{\sigma}^2/\overline{x} = \frac{1}{\overline{x}} \sum_{i=1}^n (x_i - \overline{x})^2 = 26.56$; $-2\log\Lambda = 2\sum_{i=1}^n x_i \log\left(\frac{x_i}{\overline{x}}\right) = 27.11$

- $\dim(\Omega) - \dim(\Omega_0) = 23 - 1 = 22$
- p-value for 27.11 is 0.21. So there is not enough evidence against the null hypothesis.
- **Note.** Sample size 23 is small and the test may have low power.

Example 7.24 (Bacterial Clumps, Poisson dispersion test, TBp. 348-349)

- For the data in Ex.7.20, LNp.45,

$$\bar{x} = 2.44, \qquad \hat{\sigma}^2 = 4.59 \qquad \Rightarrow \qquad \frac{n\hat{\sigma}^2}{\bar{x}} = \frac{400 \times 4.59}{2.44} = 752.7$$

- The p-value is:

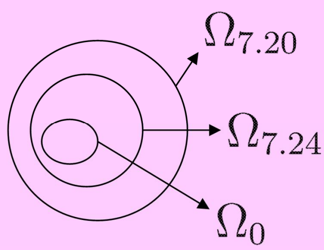
$$p\text{-value} = P\left(\frac{n\hat{\sigma}^2}{\bar{X}} \geq 752.7 \,\middle|\, H_0\right) = P\left(\frac{\frac{n\hat{\sigma}^2}{\bar{X}} - 399}{\sqrt{2 \times 399}} \geq \frac{752.7 - 399}{\sqrt{2 \times 399}}\right)$$

$$\approx 1 - \Phi(12.5) \approx 0 \qquad \text{(normal approximation to } \chi^2_{m=399}\text{)}$$

- Thus, there is almost no doubt that the Poisson distribution fails to fit the data.

Question 7.17

- Compare Ex. 7.20 and Ex. 7.24. They test the same null hypothesis H_0. Why are the test statistics in the two examples different?
  - In Ex. 7.20,
    $\Omega_{7.20} = \{X_i \text{ can be any discrete r.v.'s}\}$
  - In Ex. 7.24,
    $\Omega_{7.24} = \{X_i \sim P(\lambda_i), i = 1, \ldots, n.\}$

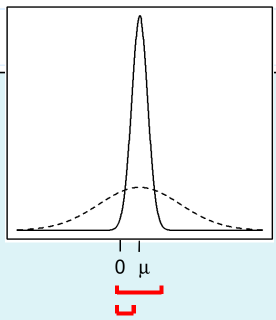


- Is it appropriate to use the test statistic in Ex. 7.20 to test the H_0 and H_A in Ex. 7.24? How is the opposite?
- For same data set, which of the tests in the two example would be expected to have smaller p-value? Why?

Note. If one has a specific alternative hypothesis in mind, better power can be obtained by developing a test against that alternative rather than against a more general alternative.

❖ **Reading**: textbook, 9.6

- Some concerns about hypothesis testing



- Question: Suppose modeling is correct. For

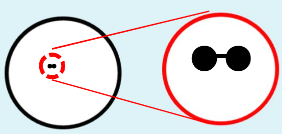
$$H_0\colon \theta = \theta_0 \quad \text{vs.} \quad H_A\colon \theta \neq \theta_0$$

when H_0 is not rejected, does it mean we accept $\theta = \theta_0$?

e.g.: $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$, σ known,
$\mu \approx 0$, but not zero, reject H_0 if

$$\left|\frac{\overline{Y} - 0}{\sigma/\sqrt{n}}\right| \geq c \Leftrightarrow |\overline{Y}| > c\,\frac{\sigma}{\sqrt{n}} = c\sqrt{Var(\overline{Y})}.$$

Consider the two cases:
(i) n=10, and (ii) n=10000.



Note: this is why we prefer **not** to say

"accept H_0",

but rather

"sample size is not large enough to reject H_0",

or

"fail to reject H_0"

- When sample size n is large enough, it is very possible that almost every tests are significant (i.e., H_0 rejected).

  example: average temperature yesterday: $\mu_1 = 28°C$,
  average temperature today: $\mu_2 = 28.001°C$,
  $\mu_1 - \mu_2 = 0.001°C$.

  Test $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_A : \mu_1 - \mu_2 \neq 0$. The null $\mu_1 = \mu_2$ will be easily rejected if n is very large.

- Statistical significance may be inconsistent with practical/physical significance. (example?)
  - Q: what causes the inconsistency? To claim significance,

    $$\text{statistical standard} \quad \xleftrightarrow{\text{different}} \quad \text{physical standard}$$

  - for datasets with large n, easy to get *statistically significant* results for θ_i's (e.g., every $H_0 : \theta_i = 0$ rejected). But, the magnitudes of some θ_i's may be small (i.e., $\hat{\theta}_i \approx 0$) $\Rightarrow$ not physically important.

- duality of confidence intervals and hypothesis tests

Example 7.25 (Normal mean, TBp. 337)

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

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_A : \mu \neq \mu_0$$

- Test statistic and rejection region: rejects H_0 for

$$|\overline{X} - \mu_0| > x_0 .$$

- The x_0 is determined by

$$P\left(|\overline{X} - \mu_0| > x_0 \middle| H_0\right) = \alpha \quad \Rightarrow \quad x_0 = \frac{\sigma}{\sqrt{n}} z(\alpha/2).$$

- Thus, the test accepts $H_0 : \mu = \mu_0$ when

$$|\overline{X} - \mu_0| < \frac{\sigma}{\sqrt{n}} z(\alpha/2).$$

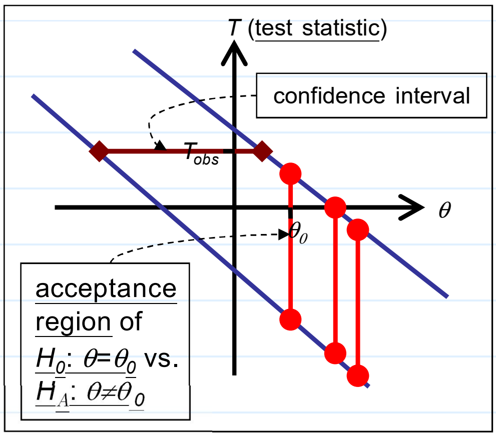
$$\Leftrightarrow \quad P\left(AR \middle| H_0\right) = 1 - \alpha.$$

$$\Leftrightarrow \quad P\left(\overline{X} - \frac{\sigma}{\sqrt{n}} z(\alpha/2) < \mu_0 < \overline{X} + \frac{\sigma}{\sqrt{n}} z(\alpha/2) \middle| \mu = \mu_0\right) = 1 - \alpha.$$

- On the other hand, a $(1-\alpha)\times 100\%$ confidence interval for μ is

$$\left[\overline{X}-\frac{\sigma}{\sqrt{n}}z(\alpha/2)\,,\ \ \overline{X}+\frac{\sigma}{\sqrt{n}}z(\alpha/2)\right]$$

- Thus, μ_0 lies in the $(1-\alpha)\times 100\%$ confidence interval if and only if the level-α test accepts $H_0:\mu=\mu_0$.



In general,

θ : parameter of a family of distributions

Ω : the set of all possible values of θ (parameter space)

Theorem 7.8 (TBp. 338)

Suppose for every θ_0 in Ω, there is a level-α test of the hypothesis

$$H_0:\theta=\theta_0 \quad \text{vs.} \quad H_A:\theta\neq\theta_0$$

Denote the acceptance region of the test by $AR(\theta_0)$. Then the set

$$C(X)=\left\{\theta \,\middle|\, X\in AR(\theta)\right\}$$

is a $(1-\alpha)\times 100\%$ confidence region for θ.

Proof. First note that

$$P\left(X\in AR(\theta_0)\,\middle|\,\theta=\theta_0\right)=1-\alpha.$$

This implies that

$$P\left(\theta_0\in C(X)\,\middle|\,\theta=\theta_0\right)=P\left(X\in AR(\theta_0)\,\middle|\,\theta=\theta_0\right)=1-\alpha.$$

Theorem 7.9 (TBp. 338)

Suppose $C(X)$ is a $(1-\alpha)\times 100\%$ confidence region for θ. Then an acceptance region for a level-α test of $H_0:\theta=\theta_0$ is

$$AR(\theta_0)=\left\{X\,\middle|\,\theta_0\in C(X)\right\}.$$

Proof. $P\left(X\in AR(\theta_0)\,\middle|\,\theta=\theta_0\right)=P\left(\theta_0\in C(X)\,\middle|\,\theta=\theta_0\right)=1-\alpha.$

Note.

- Theorem 7.8 says that a $(1-\alpha)\times 100\%$ confidence region for θ consists of all the values of θ_0 for which $H_0:\theta=\theta_0$ will not be rejected at level α.
- Theorem 7.9 says that the nlll hypothesis $H_0:\theta=\theta_0$ is accepted if θ_0 lies in the confidence region.

Example 7.26 (Normal variance)

- Let $X_1, \cdots, X_n$ be an i.i.d. sample from $N(\mu, \sigma^2)$, where μ is known and σ is unknown.
- The acceptance region of the level-α UMP unbiased test for
$$H_0 : \sigma = \sigma_0 \quad \text{vs.} \quad H_A : \sigma \neq \sigma_0$$
is approximately
$$\chi_n^2(1-\alpha/2) \leq \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2} \leq \chi_n^2(\alpha/2)$$
- It is equivalent to
$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2(\alpha/2)} \leq \sigma_0^2 \leq \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2(1-\alpha/2)}.$$
- Therefore, a $(1-\alpha) \times 100\%$ confidence interval for σ^2 is
$$\left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2(\alpha/2)}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2(1-\alpha/2)}\right].$$

- a good test $\Leftrightarrow$ a good confidence interval (region)
- UMPU test $\Leftrightarrow$ UMA (Uniformally Most Accurate) confidence interval

❖ **Reading**: textbook, 9.3

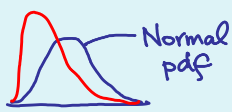
- Test for normality

Question 7.18

"$X_1, \ldots, X_n$ are i.i.d. from $N(\mu, \sigma^2)$" is a very popular assumption in statistical analysis. How can we test the goodness of fit to the normal distribution?

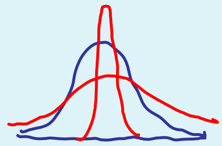
- **Pearson's chi-square (or GLR) test for goodness of fit.** In order for the limiting distribution to be chi-square, the parameters must be estimated from the grouped data. (Chernoff and Lehmann, 1954).
- Departures from normality often takes the form of skewness or asymmetry.
- A goodness-of-fit test can be based on the **coefficient of skewness** for the sample:

$$b_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^3}{s^3}$$



  Reject the null hypothesis that $X_1, \cdots, X_n$ is a sample from a Normal distribution for large values of $|b_1|$.
- Symmetric distributions can depart from normality by being heavy-tailed or light-tailed or too peaked or too flat in the center. Thesis forms of departures may be detected by **coefficient of kurtosis** for the sample:

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^4}{s^4}.$$



- Linearity of the normal probability plot can be measured by the correlation coefficient r of the normal quantiles and the ordered observations. The test rejects for small values of r.
- Sampling distributions of b_1, b_2, and r under normal model can be approximated by simulation.

❖ **Reading**: textbook, 9.9

- Some graphical methods for accessing goodness of fit

***Recall*: the 3rd step in the procedure of statistics --- Data Analysis**

- numerical methods
  - ➢ *estimation* (point estimation & interval estimation)
  - ➢ *hypothesis testing*
  - ▪ accurate conclusion
  - ▪ but only based on one (or several) numbers
- graphical methods
  - ▪ offer intuitive perception and various information
  - ▪ (could bring) divergent conclusions

- Hanging rootograms

Definition 7.11 (hanging histogram, rootograms, chi-gram TBp. 349-352)

Plot differences between the histogram (i.e., observed counts, O_i) and expected counts E_i:

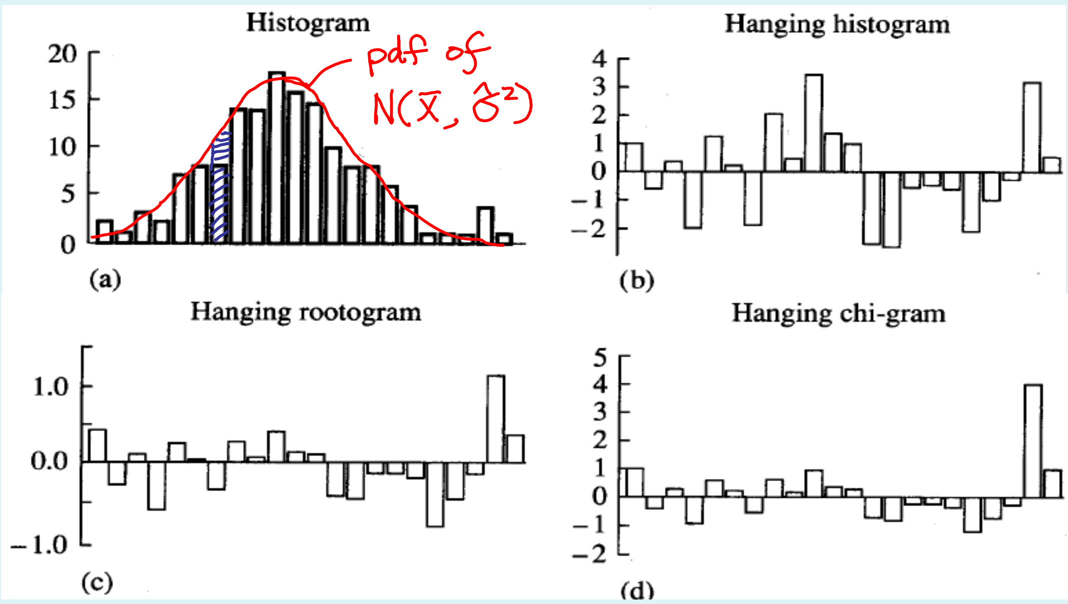


Figure (a): histogram — roughly bell shaped

To examine carefully, if the jth interval is $[x_{j-1}, x_j]$, the probability that an observation falls in it is $\hat{p}_j = \Phi\left((x_j - \bar{x})/\hat{\sigma}\right) - \Phi\left((x_{j-1} - \bar{x})\hat{\sigma}\right)$ under the normal model (μ, σ estimated by $\bar{x}$,$\hat{\sigma}$).

Figure (b): hanging histogram of the differences:

n_j (observed count) $-$ $\hat{n}_j$ (expected count) , where $\hat{n}_j = n\hat{p}_j$

$$\begin{aligned} Var(n_j - \hat{n}_j) &\doteq Var(n_j) \qquad \text{(ignore variability from } \hat{n}_j) \\ &= np_j(1-p_j) = n\left[\frac{1}{4} - \left(p_j - \frac{1}{2}\right)^2\right] \\ &\approx np_j \qquad \text{(when } p_j \text{ is small)} \end{aligned}$$

$\Rightarrow$ the cell variances are unequal

Theorem 7.10 (variance stabilizing transformation, TBp.351)

Let X be a random variable with mean μ and variance $\sigma^2 = \sigma^2(\mu)$. Let $Y = f(X)$, then

$$Var(Y) \approx Var(X)[f'(\mu)]^2 = \sigma^2(\mu) \cdot f'(\mu)^2$$

If f is chosen so that $\sigma^2(\mu)f'(\mu)^2$ is constant, the variance of Y will not depend on μ. Such an f is called a variance stabilizing transformation.

Figure (c): hanging rootogram showing $\sqrt{n_j} - \sqrt{\hat{n}_j}$.

Applying the variance stabilizing transformation to this case: under H_0 (i.e., the model is correct),

$$E(n_j) = np_j = \mu$$

$$Var(n_j) \approx np_j = \sigma^2(\mu)$$

Then $\mu[f'(\mu)]^2$ is constant if $f(x) = \sqrt{x}$ and

$$E(\sqrt{n_j}) \approx \sqrt{np_j}$$

$$Var(\sqrt{n_j}) \approx 1/4\,.$$

Note that in Figure (c), the deviations in the center were down weighted and deviations in the tails were emphasized.

Figure (d): hanging chi-gram plot, a plot of the components of Pearsons chi-square statistic:
$$\frac{n_j - \hat{n}_j}{\sqrt{\hat{n}_j}}$$
since, neglecting the variability in $\hat{n}_j$, $Var(n_j - \hat{n}_j) \approx np_j = \hat{n}_j$, so

$$Var\left(\frac{n_j - \hat{n}_j}{\sqrt{\hat{n}_j}}\right) \approx 1.$$

Question: What more information can you obtain from the plots than a goodness-of-fit test?

❖ **Reading**: textbook, 9.7

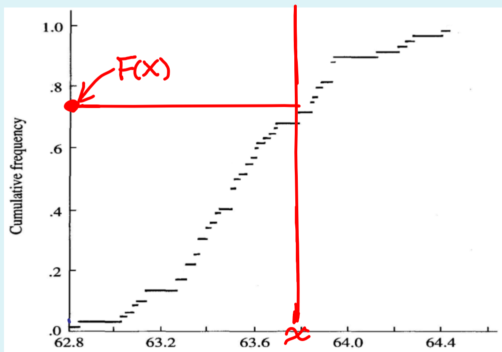
- Probability Plots

Question 7.19

- histogram , pdf/pmf ? , cdf

Definition 7.12 (empirical cumulative distribution function, TBp. 378)

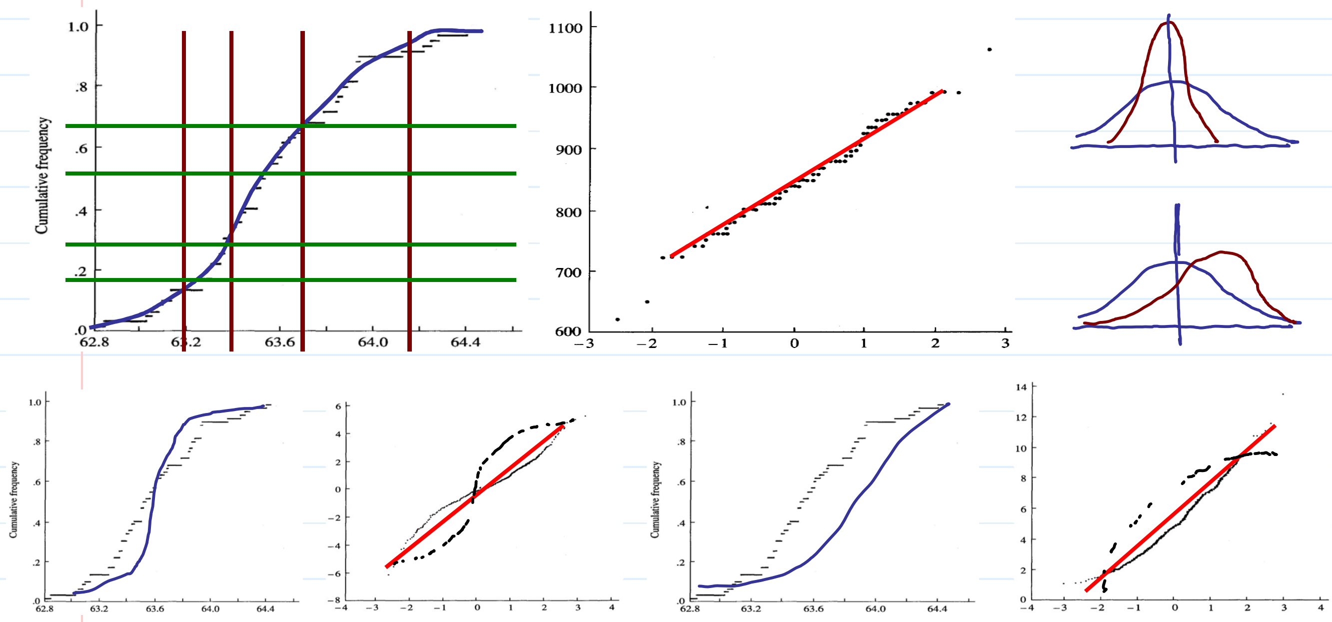
- $x_1, x_2, \ldots, x_n$: a batch of numbers $\Rightarrow$ a realization of random variables $X_1, \ldots, X_n$ that are i.i.d. with cdf F.
- **empirical cumulative distribution function (ecdf)** is
$$F_n(x) = \frac{1}{n}(\#x_i \le x) \quad \Rightarrow \quad \text{an estimate of } F(x) = P(X \le x)$$
- Let $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ be the ordered numbers of $x_1, \ldots, x_n$, then
$$F_n(x) = \begin{cases} 0, & \text{if } x < x_{(1)} \\ \frac{k}{n}, & \text{if } x_{(k)} \le x < x_{(k+1)}, \text{ where } k = 1, 2, \ldots, n-1 \\ 1, & \text{if } x \ge x_{(n)} \end{cases}$$
or
$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(x_i)$$
where I is the indicator function:
$$I_{(-\infty,x]}(X_i) = \begin{cases} 1 & \text{if } X_i \le x \\ 0 & \text{if } X_i > x \end{cases}$$



- $I_{(-\infty,x]}(X_i)$, $i = 1, 2, \ldots, n$ are independent Bernoulli random variables. Thus, $nF_n(x)$ is a $B(n, F(x))$ random variable.
$$E[F_n(x)] = F(x); \qquad Var[F_n(x)] = \frac{1}{n}F(x)[1 - F(x)]$$
$F_n(x)$ unbiased, has maximum variance at the median of F.

Question 7.20

How to examine by graph whether $F_n(x)$ is similar to $F(x)$?



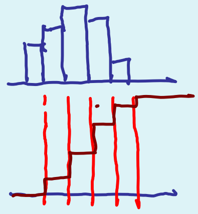
Definition 7.13 (probability plot, TBp. 352-354)

- Suppose that it is hypothesized that X follows a certain distribution F. Given an i.i.d. sample $X_1, \ldots, X_n \sim F$, plot
$$F(X_{(k)}) \quad \text{versus} \quad \frac{k}{n+1}$$
or equivalently
$$X_{(k)} \quad \text{versus} \quad F^{-1}\left(\frac{k}{n+1}\right).$$
- Note that we are plotting the ordered observations (which may be viewed as the observed or empirical quantiles) versus the quantiles of the hypothetical distribution.
- **a rationale of probability plot**: If X is a continuous random variable with a strictly increasing c.d.f. F. Then $Y = F(X)$ has a Uniform(0,1) distribution
$$P(Y \le y) = P(F(X) \le y) = P(X \le F^{-1}(y)) = F\big(F^{-1}(y)\big) = y,$$
for $0 < y < 1$, and
$$E(Y_{(k)}) = E(F(X_{(k)})) = \frac{k}{n+1}.$$

- **probability plot for location-scale family.** Suppose F is of the form
$$F_{\mu,\sigma}(x) = G\left(\frac{x-\mu}{\sigma}\right),$$
where μ is the location parameter and σ is the scale parameter. Plot
$$\frac{X_{(k)} - \mu}{\sigma} \qquad \text{against} \qquad G^{-1}\left(\frac{k}{n+1}\right)$$
or plot
$$X_{(k)} \qquad \text{against} \qquad G^{-1}\left(\frac{k}{n+1}\right).$$
If the model were correct, the result would be approximately a straight line:
$$X_{(k)} \approx \sigma G^{-1}\left(\frac{k}{n+1}\right) + \mu.$$
A slight modification uses $E(X_{(k)})$ instead of $G^{-1}\left(\frac{k}{n+1}\right)$. But it can be argued that
$$E(X_{(k)}) \approx F_{\mu,\sigma}^{-1}\left(\frac{k}{n+1}\right) \approx \sigma G^{-1}\left(\frac{k}{n+1}\right) + \mu$$
so this modification yields very similar results to the original procedure.

- **probability plot for grouped data** Suppose the grouping gives $t_0, \cdots, t_m$ for the bin boundaries of the histogram. In the interval $[t_{i-1}, t_i)$ there are O_i counts, $i = 1, \cdots, m$. Denote the cumulative frequencies by $N_j = \sum_{i=1}^{j} O_i$. We plot

$$t_j \quad \text{versus} \quad G^{-1}\left(\frac{N_j}{n+1}\right), \ j = 1, \cdots, m.$$



- **Caution**
  - Probability plots are by nature monotone increasing. Some experience is necessary in gauging "straightness"
  - Simulations are very useful in sharpening one's judgment.

❖ **Reading**: textbook, 9.8, 10.2.1