9. Hypothesis Testing

9.1. Statistical hypotheses and their tests

- A statistical hypothesis is a conjecture about the distribution of a random vector.
- A hypothesis can be simple (if the conjecture contains a single distribution) or composite (if the conjecture is composed of several simple hypotheses).
- A test of a hypothesis is a rule that allows to reject or not the hypothesis depending on each foreseeable result of a random experiment.

- The hypothesis we want to test is called the "null hypothesis" H_0.
- A null hypothesis is tested against an alternative hypothesis H_1.
- Let $\theta \in \Theta \subset \mathbb{R}^K$ be an unknown parameter vector characterizing the distribution of the random vector $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$.
- **Examples:**
- Both the null and the alternative hypotheses are simple:

$$H_0 : \theta = \theta_0 \ , \quad H_1 : \theta = \theta_1.$$

- The null hypothesis is simple but the alternative is composite:

$$H_0 : \theta = \theta_0 \ , \quad H_1 : \theta \neq \theta_0.$$

- Both the null and the alternative hypotheses are composite:

$$H_0 : \theta \in B_0 \ , \quad H_1 : \theta \in B_1 \left(\equiv B_0^c\right).$$

- Tests are usually based on the value X of a sample $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ or on the value t of a statistic $\tilde{t} = g\left(\tilde{X}\right)$. This statistic is called the **test statistic**.
- For a test based on a statistic $\tilde{t} = g\left(\tilde{X}\right)$, we construct a function $\widehat{\phi}$ called the test function (or simply "the test") as follows:

$$\widehat{\phi}(\tilde{t}) = \begin{cases} 1 & \text{if } \tilde{t} \in C_{\widehat{\phi}} \\ \\ 0 & \text{if } \tilde{t} \notin C_{\widehat{\phi}} \end{cases}$$

and we reject H_0 if $\widehat{\phi}(t) = 1$ $\left(\text{or if } t \in C_{\widehat{\phi}}\right)$ and we do not reject H_0 if $\widehat{\phi}(t) = 0$ $\left(\text{or if } t \notin C_{\widehat{\phi}}\right)$.

- Note that the previous test function can we written as a function of the sample:

$$\phi(\tilde{X}) = \begin{cases} 1 & \text{if } \tilde{X} \in C_\phi \\ \\ 0 & \text{if } \tilde{X} \notin C_\phi \end{cases}$$

  and we reject H_0 if $\phi(X) = 1$ $\left(\text{or if } X \in C_\phi\right)$ and we do not reject H_0 if $\phi(X) = 0$ $\left(\text{or if } X \notin C_\phi\right)$.

- Obviously, $\widehat{\phi}(\tilde{t}) = \widehat{\phi}(g(\tilde{X})) = \phi(\tilde{X})$, i.e., $\phi = \widehat{\phi}(g)$ and, since $\tilde{t} = g(\tilde{X})$, then $C_\phi = g^{-1}\left(C_{\widehat{\phi}}\right)$.

- The set C_ϕ (or $C_{\widehat{\phi}}$) where H_0 is rejected is called the critical region, whereas its complement C_ϕ^c (or $C_{\widehat{\phi}}^c$) is the acceptance region.

- Note that the function ϕ (or $\widehat{\phi}$) is the indicator function of the critical region C_ϕ (or $C_{\widehat{\phi}}$), $\phi = \mathbb{I}_{C_\phi}$ and $\widehat{\phi} = \mathbb{I}_{C_{\widehat{\phi}}}$.

- Assume that the parameter value characterizing the distribution of the random vector $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ is θ. The distribution of $\tilde{X}$ when the parameter is θ is denoted by $P_{\tilde{X}}(\cdot; \theta)$.

- Let $\mathrm{E}_\theta(\cdot)$ be the expectation when $\tilde{X} \sim P_{\tilde{X}}(\cdot; \theta) : \mathcal{B} \longrightarrow [0, 1]$. Let $P(\cdot; \theta) : \mathcal{F} \longrightarrow [0, 1]$ be a probability that induces the distribution $P_{\tilde{X}}(\cdot; \theta)$ for $\tilde{X}$.

- The rejection of the null hypothesis $H_0 : \theta \in B_0$ when it is true is called a **type I error (**or **false positive)**. The probability of committing a type I error under the test function ϕ when the value of the distribution parameter is $\theta \in B_0$ is denoted by $\alpha_\phi(\theta)$.

- The non-rejection (or acceptance) of the null hypothesis $H_0 : \theta \in B_0$ when it is false is called a **type II error (**or **false negative).** The probability of committing a type II error under the test function ϕ when the value of the distribution parameter is $\theta \in B_1$ is denoted by $\beta_\phi(\theta)$.

- Observe that, if $\theta \in B_0$, then

$$\alpha_\phi(\theta) = P\left\{\tilde{X} \in C_\phi; \theta\right\} = P_{\tilde{X}}\left(C_\phi; \theta\right) = \int_{C_\phi} dP_{\tilde{X}}(X; \theta)$$

$$= \int_{\mathbb{R}^n} \mathbb{I}_{C_\phi}(X)\, dP_{\tilde{X}}(X; \theta) = \mathrm{E}_\theta(\phi(\tilde{X}))$$

and, if $\theta \notin B_0$ (or $\theta \in B_1$), then

$$\beta_\phi(\theta) = P\left\{\tilde{X} \notin C_\phi; \theta\right\} = P_{\tilde{X}}\left(C_\phi^c; \theta\right) = 1 - P_{\tilde{X}}\left(C_\phi; \theta\right)$$

$$= 1 - \int_{C_\phi} dP_{\tilde{X}}(X; \theta) = 1 - \int_{\mathbb{R}^n} \mathbb{I}_{C_\phi}(X)\, dP_{\tilde{X}}(X; \theta) = 1 - \mathrm{E}_\theta(\phi(\tilde{X})).$$

- Note that $\alpha_\phi : B_0 \longrightarrow [0, 1]$, whereas $\beta_\phi : B_1 \longrightarrow [0, 1]$.

- If the null hypothesis is simple, $H_0 : \theta = \theta_0$, then $\alpha_\phi \equiv \alpha_\phi(\theta_0)$ is called the **size** of the critical region or "the level of significance" of the test ϕ.

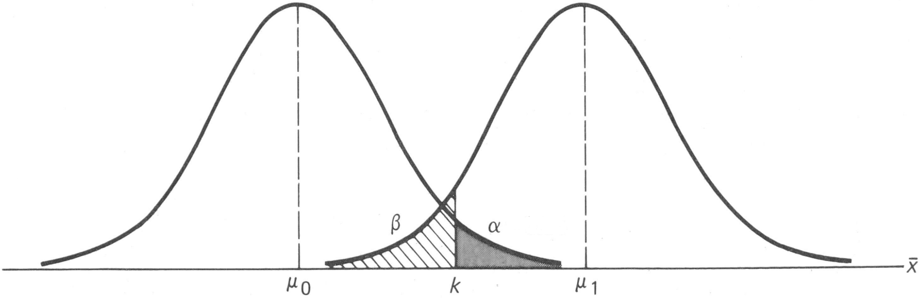
- When the null hypothesis is composite, $H_0 : \theta \in B_0$, the **size** of the critical region or the level of significance of the test ϕ is the supremum of the possible probabilities of a type I error, $\sup_{\theta \in B_0} \alpha_\phi(\theta)$.

- If $\theta \in B_1$, then $1 - \beta_\phi(\theta) = P_{\tilde{X}}\left(C_\phi; \theta\right) = P\left\{\tilde{X} \in C_\phi; \theta\right\}$ is the probability of rejecting the null hypothesis H_0 when it is false. Note that $1 - \beta_\phi(\theta) = \mathrm{E}_\theta(\phi(\tilde{X}))$ when $\theta \in B_1$.

- $1 - \beta_\phi(\theta)$ is called the **power** of the test ϕ when the distribution parameter is $\theta \in B_1$.

- The ideal test will be that for which $\alpha_\phi(\theta) = 0$ when $\theta \in B_0$ and $\beta_\phi(\theta) = 0$ when $\theta \in B_1$ or, equivalently, $\alpha_\phi(\theta) = 0$ when $\theta \in B_0$ and $1 - \beta_\phi(\theta) = 1$ when $\theta \in B_1$.

- However, α_ϕ and β_ϕ are related.



Critical region for testing $\mu = \mu_0$ against $\mu = \mu_1$.

If we move the critical value k, we will modify $\alpha = \alpha_\phi(\mu_0)$ and $\beta = \beta_\phi(\mu_1)$ in opposite directions.

9.2. The power function of a test

- Let $\theta \in \Theta \subset \mathbb{R}^K$ be a parameter vector characterizing the distribution of the random vector $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$.

- **Definition.** The power function $\pi_\phi : \Theta \longrightarrow [0, 1]$ of a test ϕ of a statistical hypothesis $H_0 : \theta \in B_0$ against an alternative hypothesis $H_1 : \theta \in B_1 = B_0^c$ is given by

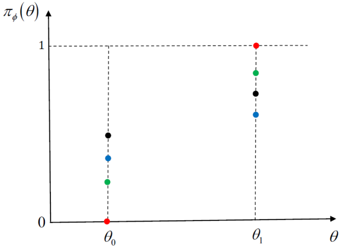
$$\pi_\phi(\theta) = \begin{cases} \alpha_\phi(\theta) & \text{for values of } \theta \text{ under } H_0 \ \ (\theta \in B_0) \\ \\ 1 - \beta_\phi(\theta) & \text{for values of } \theta \text{ under } H_1 \ \ (\theta \in B_1). \end{cases}$$

or

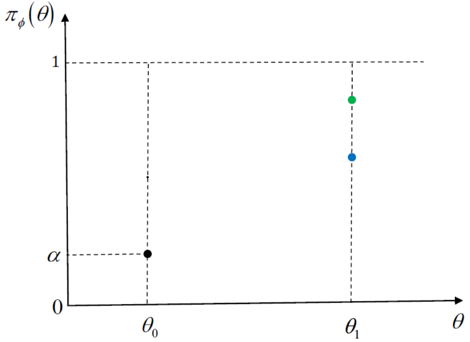
$$\pi_\phi(\theta) = P_{\tilde{X}}\left(C_\phi; \theta\right) = P\left\{\tilde{X} \in C_\phi; \theta\right\}$$
$$= \mathrm{E}_\theta(\phi(\tilde{X})) \equiv \mathrm{E}_\theta(\phi), \quad \text{for all } \theta \in \Theta.$$

- The power function gives us the probability of rejecting the null hypothesis for all possible values of the unknown parameter θ.

$$H_0 : \theta = \theta_0 \, , \; H_1 : \theta = \theta_1$$

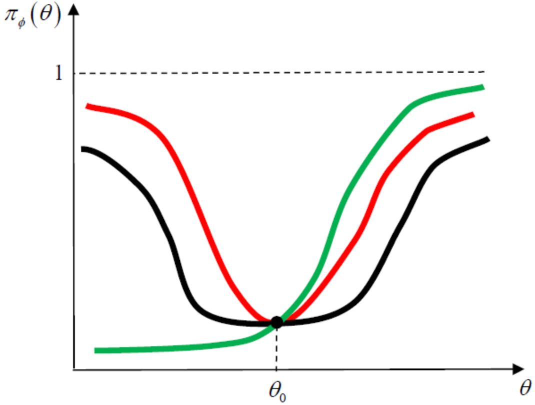


- The test with red power function is the ideal test. The test with green power function is better than the tests with either blue or black power functions since the test with green power function rejects the null hypothesis with a lower probability when it is true (which is desirable) and with a higher probability when it is false (which is also desirable).
- The tests with blue and black power function cannot be ranked since the test with blue power function rejects with a lower probability the null hypothesis when it is true (which is good) but it rejects also the null hypothesis with a lower probability when it is false (which is bad).



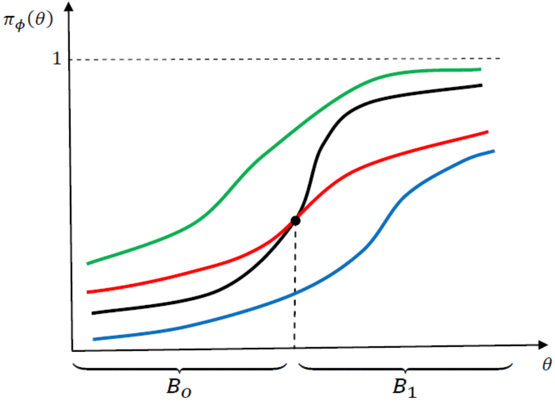
If we fix the probability $\alpha_\phi(\theta_0) \equiv \alpha_\phi$ of committing a type I error (or level of significance or size of the critical region) at the level α for all tests under consideration and then we look for the highest possible power $\pi_\phi(\theta_1) \equiv \pi_\phi = 1 - \beta_\phi(\theta_1) \equiv 1 - \beta_\phi$ (i.e., for the lowest probability $\beta_\phi(\theta_1) \equiv \beta_\phi$ of committing a type II error), then the test with the green power function is better than the test with the blue power function.

$$H_0 : \theta = \theta_0 \,,\ H_1 : \theta \neq \theta_0$$



The test with the red power function is better than that with the black power function. The test with the green power function cannot be ranked with the other two tests.

$$H_0 : \theta \in B_0\,,\; H_1 : \theta \in B_1 = B_0^c$$



The test with the black power function is better than that with the red power function. The tests with the green and blue power functions cannot be ranked with the other tests.

- A related function is the operating characteristic function of the test,

$$OC_{\phi}(\theta) = P\left\{\tilde{X} \notin C_{\phi}; \theta\right\} = P\left\{\tilde{X} \in C_{\phi}^{c}; \theta\right\} = P_{\tilde{X}}\left(C_{\phi}^{c}; \theta\right)$$

$$1 - P_{\tilde{X}}\left(C_{\phi}; \theta\right) = 1 - \pi_{\phi}(\theta) = 1 - \mathrm{E}_{\theta}(\phi\left(\tilde{X}\right))$$

or

$$OC_{\phi}(\theta) = \begin{cases} 1 - \alpha_{\phi}(\theta) & \text{for values of } \theta \text{ under } H_0 \;\; (\theta \in B_0) \\ \beta_{\phi}(\theta) & \text{for values of } \theta \text{ under } H_1 \;\; (\theta \in B_1). \end{cases}$$

- The operating characteristic function gives us the probability of not rejecting the null hypothesis for all possible values of the unknown parameter θ.

9.3. Uniformly most powerful tests

- **Definition:** A test function $\phi(\tilde{X})$ is a Uniformly Most Powerful (UMP) test if for any other test function $\phi'(\tilde{X})$ such that

$$\sup_{\theta \in B_0} \alpha_{\phi'}(\theta) \equiv \alpha' \leq \alpha \equiv \sup_{\theta \in B_0} \alpha_{\phi}(\theta),$$

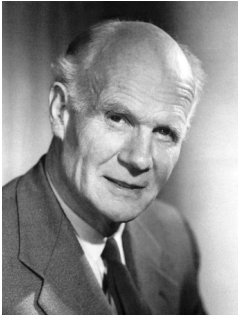
we have that for all $\theta \in B_1$,

$$\pi_{\phi'}(\theta) \equiv 1 - \beta_{\phi'}(\theta) \leq 1 - \beta_{\phi}(\theta) \equiv \pi_{\phi}(\theta).$$

- That is, a UMP test with size (or level of significance) α is a test that has the greatest power among all possible tests with size no larger than α.

9.4. The Neyman-Pearson lemma

Jerzy Neyman (1894 - 1981)



Egon Pearson (1895 - 1980)

- Assume that both $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$ are simple hypotheses and that the test ϕ is based on the value X of the sample $\tilde{X}$. Then, the probability of committing a type I error is

$$\alpha_\phi (\theta_0) = P \left\{ \tilde{X} \in C_\phi ; \theta_0 \right\} \equiv P \left\{ \tilde{X} \in C_\phi ; H_0 \right\} = P_{\tilde{X}} \left(C_\phi ; H_0 \right),$$

the probability of committing a type II error is

$$\beta_\phi (\theta_1) = P \left\{ \tilde{X} \notin C_\phi ; \theta_1 \right\} \equiv P \left\{ \tilde{X} \notin C_\phi ; H_1 \right\} = P_{\tilde{X}} \left(C_\phi^c ; H_1 \right),$$

and the power of the test is

$$\pi_\phi (\theta_1) = P \left\{ \tilde{X} \in C_\phi ; \theta_1 \right\} \equiv P \left\{ \tilde{X} \in C_\phi ; H_1 \right\} = P_{\tilde{X}} \left(C_\phi ; H_1 \right).$$

- We usually fix the probability $\alpha_\phi (\theta_0) \equiv \alpha_\phi$ of committing a type I error (or level of significance or size of the critical region) at the (low) level α (the most commonly used levels are 0.05, 0.01 or 0.001) and then we look for the highest possible power $\pi_\phi (\theta_1) \equiv \pi_\phi = 1 - \beta_\phi (\theta_1) \equiv 1 - \beta_\phi$ (i.e., for the lowest probability $\beta_\phi (\theta_1) \equiv \beta_\phi$ of committing a type II error).

- Let $h_0(X)$ be the density (probability function) of the random vector $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ under the null hypothesis, whereas $h_1(X)$ is the density (probability function) of the random vector $\tilde{X}$ under the alternative hypothesis.
- If $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ constitutes a random sample, then

$$h_j(X) = h_j(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f_j(x_i), \quad j = 0, 1,$$

  where f_0 (f_1) is the pdf or pmf of the population under the null (alternative) hypothesis.
- The value of the ratio

$$\frac{h_0(X)}{h_1(X)}$$

  should be small if H_0 is false and large if H_0 is true. Therefore, we should expect that the critical region of the test contains the sample values X for which the previous ratio is small.
- *Note:* we assume values of X for which $h_0(X) / h_1(X)$ is well-defined in $\overline{\mathbb{R}}_+$, i.e., when it is not equal to $0/0$, which means that we must have $h_0(X) > 0$ or $h_1(X) > 0$ or both.

- Note that, for a given sample value X, $h_0(X) = L_0$, where L_0 is the value of the likelihood function under the null hypothesis, and $h_1(X) = L_1$, where L_1 is the value of the likelihood function under the alternative hypothesis.
- **Neyman-Pearson Lemma.** In testing the simple null hypothesis $H_0 : \theta = \theta_0$ against the simple alternative hypothesis $H_1 : \theta = \theta_1$, the test

$$\phi_k\left(\tilde{X}\right) = \begin{cases} 1 & \text{if } \tilde{X} \in C_{\phi_k} \\ 0 & \text{if } \tilde{X} \notin C_{\phi_k}, \end{cases}$$

where

$$C_{\phi_k} = C_k \equiv \left\{ \frac{h_0(X)}{h_1(X)} \leq k \right\},$$

with $k > 0$, is a UMP test with level of significance α_{ϕ_k}, i.e., $\phi_k\left(\tilde{X}\right)$ is the most powerful test with significance level no larger than α_{ϕ_k}, where

$$\alpha_{\phi_k} = P_{\tilde{X}}\left(C_{\phi_k}; \theta_0\right) \equiv P_{\tilde{X}}\left(C_{\phi_k}; H_0\right).$$

- The lemma says that the power of any other test ϕ with $\alpha_\phi \leq \alpha_{\phi_k}$ does not exceed the power of ϕ_k. That is,

$$\alpha_{\phi_k} \geq \alpha_\phi \implies 1 - \beta_{\phi_k} \geq 1 - \beta_\phi.$$

- In particular, if we fix the level of significance at the level α,

$$\alpha = \alpha_\phi = \alpha_{\phi_k} \implies 1 - \beta_{\phi_k} \geq 1 - \beta_\phi.$$

- **Proof.** We want to prove that

$$\left(1 - \beta_{\phi_k}\right) - \left(1 - \beta_\phi\right) = \mathrm{E}_1(\phi_k) - \mathrm{E}_1(\phi) = \mathrm{E}_1\left(\phi_k - \phi\right) \geq 0$$

  if ϕ is any other test such that

$$\alpha_{\phi_k} - \alpha_\phi = \mathrm{E}_0\left(\phi_k\right) - \mathrm{E}_0\left(\phi\right) = \mathrm{E}_0\left(\phi_k - \phi\right) \geq 0,$$

  where $\mathrm{E}_0\left(\cdot\right) \equiv \mathrm{E}_{\theta_0}\left(\cdot\right)$ and $\mathrm{E}_1\left(\cdot\right) \equiv \mathrm{E}_{\theta_1}\left(\cdot\right)$.

Assume first that $X \in C_k$. Then,

$$\phi_k - \phi = 1 - \phi \geq 0 \quad \text{and} \quad h_1 \geq \frac{h_0}{k} \geq 0$$

so that

$$(\phi_k - \phi)\, h_1 \geq \frac{1}{k}(\phi_k - \phi)\, h_0 \geq 0.$$

Assume now that $X \notin C_k$. Then,

$$\phi_k - \phi = 0 - \phi \leq 0 \quad \text{and} \quad 0 \leq h_1 < \frac{h_0}{k}$$

so that

$$0 \geq (\phi_k - \phi)\, h_1 \geq \frac{1}{k}(\phi_k - \phi)\, h_0.$$

Therefore,

$$\left(1 - \beta_{\phi_k}\right) - \left(1 - \beta_{\phi}\right) = \mathrm{E}_1(\phi_k - \phi) \geq \frac{1}{k}\underbrace{\mathrm{E}_0(\phi_k - \phi)}_{\left(\alpha_{\phi_k} - \alpha_{\phi}\right) \geq 0} \geq 0. \qquad Q.E.D.$$

9.5. The Monotone Likelihood Ratio property

- The Karlin-Rubin theorem is an extension of the Neyman-Pearson lemma for composite hypothesis. Before stating that theorem, we need to introduce a new property for pdfs/pmfs.
- **Definition:** The family of density (or probability) functions $f(x;\theta)$, with $\theta \in \Theta \subset \mathbb{R}$, for univariate distributions $(x \in \mathbb{R})$ has a monotone likelihood ratio (MLR) if, for every $\theta_2 > \theta_1$,

$$\frac{f(x;\theta_2)}{f(x;\theta_1)}$$

  is a non-decreasing function of x.
- *Note:* we will only consider again values of x for which $f(x;\theta_2)/f(x;\theta_1)$ is not equal to $0/0$, which means that we must have $f(x;\theta_2) > 0$ or $f(x;\theta_1) > 0$ or both.
- Thus, if $f(x;\theta)$ has a MLR, then for any pair $\theta_2 > \theta_1$, the greater is x the more likely that it is drawn from the distribution with the parameter value θ_2 rather than θ_1.

- **Application of the MLR property:**
- Assume that a worker exerts either a high or a low effort $\tilde{e}$ in her job. Thus, $\tilde{e}$ can take only two values e_H and e_L with $e_H > e_L$. The prior probabilities of these two values are $f_{\tilde{e}}(e_H)$ and $f_{\tilde{e}}(e_L)$, respectively. Let us assume that the amount of output $\tilde{y}$ produced by this worker arises from a stochastic technology that depends on the amount of effort exerted by the worker so that $g_{\tilde{y}|\tilde{e}}(y|e)$ is the conditional probability that the amount of output is y when the effort is e.
- Using Bayes' theorem, we can compute the conditional probability that $\tilde{e} = e_H$ given $\tilde{y} = y$,

$$f_{\tilde{e}|\tilde{y}}(e_H|y) = \frac{f_{\tilde{e}}(e_H) \cdot g_{\tilde{y}|\tilde{e}}(y|e_H)}{f_{\tilde{e}}(e_H) \cdot g_{\tilde{y}|\tilde{e}}(y|e_H) + f_{\tilde{e}}(e_L) \cdot g_{\tilde{y}|\tilde{e}}(y|e_L)}$$

$$= \frac{f_{\tilde{e}}(e_H)}{f_{\tilde{e}}(e_H) + f_{\tilde{e}}(e_L) \cdot \dfrac{1}{\dfrac{g_{\tilde{y}|\tilde{e}}(y|e_H)}{g_{\tilde{y}|\tilde{e}}(y|e_L)}}}.$$

- If $g_{\tilde{y}|\tilde{e}}(y|e)$ has a strict MLR, i.e.,

$$\frac{g_{\tilde{y}|\tilde{e}}(y|e_H)}{g_{\tilde{y}|\tilde{e}}(y|e_L)}$$

  is strictly increasing in y, then $f_{\tilde{e}|\tilde{y}}(e_H|y)$ is strictly increasing in y. Note that here we view the amount of effort as the parameter for the distribution of output.

- This means that the higher is the (observable) amount of output, the higher is the probability that the worker has exerted a high amount of effort (which is not observable).

- The MLR property plays a crucial role in contract theory, economics of uncertainty, and mechanism design.

- **Proposition 1.** If the family of density (or probability) functions $f(x;\theta)$, with $\theta \in \Theta \subset \mathbb{R}$ and $x \in \mathbb{R}$, has a MLR then, for every $\theta_2 > \theta_1$, the following holds:

  (a) First-order stochastic dominance:

$$F(x;\theta_2) \leq F(x;\theta_1) \text{ for all } x \in \mathbb{R}. \quad (1)$$

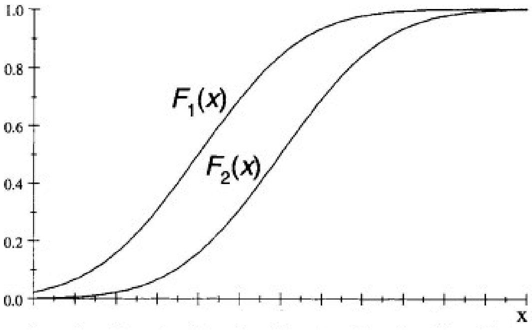
  (b)

$$P\{\tilde{x} > x;\theta_2\} \geq P\{\tilde{x} > x;\theta_1\} \text{ for all } x \in \mathbb{R}. \quad (2)$$

  (c) Non-increasing hazard rate:

$$\frac{f(x;\theta_1)}{1-F(x;\theta_1)} \geq \frac{f(x;\theta_2)}{1-F(x;\theta_2)} \text{ for all } x \in \mathbb{R}. \quad (3)$$

- **Proof:** See the handout.
- When (1) in the previous proposition holds, we say that the distribution with parameter θ_2 first-order "stochastically dominates" the distribution with parameter θ_1. Stochastic dominance is a key concept in the theory of decision under uncertainty.



- The distribution characterized by the distribution function F_2 first-order stochastically dominates the distribution characterized by the distribution function F_1.
- Note that the ordering induced on the set of distributions by the first-order stochastic dominance relation is very partial as any two arbitrary distributions cannot be typically ranked according to this criterion, i.e., we usually have $F_1(x) > F_2(x)$ for some values of x and $F_1(x) < F_2(x)$ for some other values of x.

- The following proposition tells us that, when choosing ex-ante between two distributions of consumption, which can be ranked according to the first-order stochastic dominance relation, all the expected utility maximizing individuals with non-decreasing utility (i.e., individuals who weakly prefer more to less) will unanimously weakly prefer the distribution of consumption that first-order stochastically dominates the other.
- **Proposition 2 (Rothschild and Stiglitz).** Consider two random variables $\tilde{x}$ and $\tilde{y}$ having distributions whose supports are subsets of the bounded interval $[a, b]$ and their associated distribution functions are $F_{\tilde{x}}$ and $F_{\tilde{y}}$, respectively. Then, $F_{\tilde{x}}(z) \leq F_{\tilde{y}}(z)$ for all $z \in \mathbb{R}$ if and only if

$$\mathrm{E}\left[u(\tilde{x})\right] \equiv \int_{\mathbb{R}} u(z) dF_{\tilde{x}}(z) \geq \mathrm{E}\left[u(\tilde{y})\right] \equiv \int_{\mathbb{R}} u(z) dF_{\tilde{y}}(z),$$

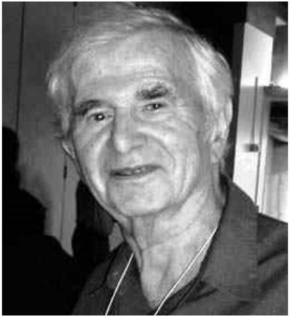
for all Bernoulli utility functions u that are non-decreasing and continuous on $[a, b]$.
- **Proof:** See the handout.

- When (3) in Proposition 1 holds, we say that the "hazard rate" is non-increasing in θ.
- Imagine the following situation in survival/reliability analysis: the parameter θ measures the health/quality of an individual/component and $f(x;\theta)$ is the (unconditional) probability/density of dying/failure at age x. The hazard rate

$$\eta(x;\theta) = \frac{f(x;\theta)}{1 - F(x;\theta)}$$

  is the conditional probability/density of dying/failing given that the individual/component is at age x. Obviously, this requires to have survived until the age x, which is given by the "survival function" $1 - F(x;\theta)$.
- Under MLR, the higher is the health/quality θ of an individual/component the lower is the hazard rate, i.e., $\eta(x;\theta)$ is non-increasing in θ.
- The MLR property implies first-order stochastic dominance and decreasing hazard rate. The converse is not true.

9.6. The Karlin-Rubin theorem



Samuel Karlin (1924 - 2007)

Herman Rubin (1926 - 2018)

- **Karlin-Rubin theorem.** Assume that (i) $\tilde{t} = t\left(\tilde{X}\right)$ is a test statistic that is a real-valued (or univariate) sufficient estimator for the parameter $\theta \in \Theta \subset \mathbb{R}$ characterizing the distribution of the random vector $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ and (ii) the family $g(t;\theta)$ of density (or probability) functions for $\tilde{t}$ has a MLR. Then, in testing the null hypothesis $H_0 : \theta \leq \theta_0$ against the alternative hypothesis $H_1 : \theta > \theta_0$, the test

$$\widehat{\phi}\left(\tilde{t}\right) = \begin{cases} 1 & \text{if } \tilde{t} > t_0 \\ 0 & \text{if } \tilde{t} \leq t_0, \end{cases}$$

is a UMP test with size $\alpha = P\left\{\tilde{t} > t_0; \theta_0\right\} = \mathrm{E}_{\theta_0}\left(\widehat{\phi}\left(\tilde{t}\right)\right)$.

- **Proof.** See the handout.
- In all the tests we consider in this chapter, the critical (acceptance) region can be defined by either a weak or a strict inequality since, if the test statistic $\tilde{t}$ is absolutely continuous then $P\left\{\tilde{t} = t_0; \theta_0\right\} = 0$, while if $\tilde{t}$ is discrete we can choose t_0 so that $P\left\{\tilde{t} = t_0; \theta_0\right\} = 0$.

- Note that, in the proof of the Karlin-Rubin theorem, we invert the ratio of pdfs/pmfs of the Neyman-Pearson lemma and, therefore, we had to change de direction of the inequality in the test.

- Analogously, under the condition of the Karlin-Rubin theorem, the test that rejects $H_0 : \theta \geq \theta_0$ in favor of $H_1 : \theta < \theta_0$ if and only if $\tilde{t} < t_0$ is a UMP test with significance level $\alpha = P\left\{\tilde{t} < t_0; \theta_0\right\}$.

- The Karlin-Rubin theorem also covers the particular case of testing $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta > \theta_0$ and, thus, the case of testing $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta < \theta_0$. Obviously, the particular case of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ is also covered by the Karlin-Rubin theorem.

- Finally, if we he have a non-increasing likelihood ratio rather than non-decreasing, then we only have to change the direction of the inequalities in the previous tests.

- Obviously, if we use a single observation x from a population distribution characterized by the parameter θ, then $\tilde{x}$ is (trivially) a sufficient estimator for θ. Thus, if the family of density (or probability) functions $f(x;\theta)$, with $\theta \in \Theta \subset \mathbb{R}$, has a MLR then, in testing the null hypothesis $H_0 : \theta \leq \theta_0$ against the alternative hypothesis $H_1 : \theta > \theta_0$, the test

$$\phi(\tilde{x}) = \begin{cases} 1 & \text{if } \tilde{x} > x_0 \\ 0 & \text{if } \tilde{x} \leq x_0, \end{cases}$$

  is a UMP test with size $\alpha = P\{\tilde{x} > x_0; \theta_0\} = \mathrm{E}_{\theta_0}(\phi(\tilde{x}))$.
- For most of the two-sided tests (i.e., $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$) no UMP test exists.
- This is so because the test that is UMP for $\theta < \theta_0$ is not the same as the test that is UMP for $\theta > \theta_0$. A UMP test must be most powerful across every value in H_1.
- A UMP test does not exist in general when we are testing for the value of a parameter vector, $\theta \in \Theta \subset \mathbb{R}^K$.

- If we use the sample mean (or the sum) of a random sample as a (sufficient) estimator $\tilde{t}$ for the parameter θ of the exponential distribution or for the parameter μ of the normal distribution (with σ known), then the pdf of $\tilde{t}$ has a MLR.
- If we use the number (or the percentage) of successes in independent Bernoulli experiments as a (sufficient) estimator $\tilde{t}$ for the parameter θ of the binomial distribution (with n known) or for the parameter λ of the Poisson distribution, then the pmf of $\tilde{t}$ has a MLR.
- The family of exponential pdfs or pmfs, which are the ones with the functional form

$$g(t;\theta) = c(\theta)h(t)e^{w(\theta)t},$$

has a MLR if $w(\theta)$ is non-decreasing.
- Obviously, if $\theta_2 > \theta_1$ and $w(\theta)$ is non-decreasing, then

$$\frac{g(t;\theta_2)}{g(t;\theta_1)} = \frac{c(\theta_2)h(t)e^{w(\theta_2)t}}{c(\theta_1)h(t)e^{w(\theta_1)t}} = \frac{c(\theta_2)}{c(\theta_1)}e^{[w(\theta_2)-w(\theta_1)]t}$$

is non-decreasing in t.

9.7. Likelihood ratio tests

- Let $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be a random vector (or sample) with density (probability function) $h(X;\theta)$ where $\theta \in \Theta \subset \mathbb{R}^K$. Consider the following null and alternative hypotheses: $H_0 : \theta \in B_0$ and $H_1 : \theta \in B_1 = B_0^c$ (i.e., $B_0 \cup B_1 = \Theta$).

- If $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ is a random sample, then

$$h(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta),$$

  where f is the density (or probability function) of the common population of the i.i.d. random variables $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n$.

- Let $L(\theta; X) \equiv h(X;\theta)$ be the likelihood function, which is equal to $\prod_{i=1}^{n} f(x_i;\theta)$ if $\tilde{X}$ is a random sample.

- The value

$$\lambda = \frac{\sup\limits_{\theta \in B_0} L(\theta; X)}{\sup\limits_{\theta \in \Theta} L(\theta; X)} \in [0, 1],$$

which is the value of the likelihood ratio statistic

$$\tilde{\lambda} = \frac{\sup\limits_{\theta \in B_0} L(\theta; \tilde{X})}{\sup\limits_{\theta \in \Theta} L(\theta; \tilde{X})},$$

should be small if H_0 is false and large if H_0 is true.

- Sometimes, we can find the value of the parameter θ that maximizes the likelihood under the null hypothesis,

$$\widehat{\theta} = \arg\max_{\theta \in B_0} L(\theta; X).$$

- Similarly, we could find the maximum likelihood estimate,

$$\widehat{\theta}_{ML} = \arg\max_{\theta \in \Theta} L(\theta; X).$$

- Each of the previous maximization problems has solution if B_0 (resp. Θ) is compact and the likelihood function $L(\cdot; X)$ is continuous on B_0 (resp. Θ).

- Then, the value of the likelihood ratio test statistic becomes

$$\lambda = \frac{L\left(\widehat{\theta}; X\right)}{L\left(\widehat{\theta}_{ML}; X\right)} \in [0, 1] .$$

- Recall also that, sometimes, by modifying the density function on a set with zero Lebesgue measure, we can make

$$\max_{\theta \in B_0} L(\theta; X) = \sup_{\theta \in B_0} L(\theta; X) \qquad \text{and} \qquad \max_{\theta \in \Theta} L(\theta; X) = \sup_{\theta \in \Theta} L(\theta; X) ,$$

where each of the previous suprema exists if the likelihood function $L(\cdot; X)$ is bounded and B_0 (resp. Θ) is compact.

- *Note:* If both the null and the alternative hypotheses were simple, $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$, then the value of the likelihood ratio test statistic becomes

$$\lambda = \frac{L(\theta_0; X)}{\max\{L(\theta_0; X), L(\theta_1; X)\}} = \frac{L_0}{\max\{L_0, L_1\}}$$

$$= \frac{1}{\max\left\{1, \dfrac{1}{L_0 / L_1}\right\}} = \frac{1}{\max\left\{1, \dfrac{1}{h_0(X) / h_1(X)}\right\}} \in [0, 1],$$

which yields a test equivalent to the one appearing in the Neyman-Pearson lemma.

- Thus, the likelihood ratio test applies the logic of the Neyman-Pearson lemma to test hypotheses that are not necessarily simple and for which the lemma does not necessarily apply!

- **Definition.** The test

$$\widehat{\phi}\left(\tilde{\lambda}\right) = \begin{cases} 1 & \text{if } \tilde{\lambda} \leq k \\ \\ 0 & \text{if } \tilde{\lambda} > k, \end{cases}$$

where $k \in (0, 1)$, is a likelihood ratio test of the null hypothesis $H_0 : \theta \in B_0$ against the alternative $H_1 : \theta \in B_1$.

- If we fix the level of significance equal to α, we should find the value k such that

$$\sup_{\theta \in B_0} P\left\{\tilde{\lambda} \in [0, k]\, ; \theta\right\} = \sup_{\theta \in B_0} P_{\tilde{\lambda}}\left([0, k]\, ; \theta\right) = \sup_{\theta \in B_0} \int_{[0,k]} dP_{\tilde{\lambda}}(\lambda; \theta) = \alpha.$$

- The problem is that we need to find the distribution $P_{\tilde{\lambda}}(\cdot;\theta)$ of the test statistic $\tilde{\lambda}$. This is a difficult task!

- Assume now that $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ is a random sample. Under the null hypothesis,

$$-2\ln\tilde{\lambda} \longrightarrow \chi^2_M \text{ as } n \to \infty. \qquad \text{(Wilks' Theorem)}$$

  where M is the difference between the dimensions of Θ and B_0.

- In this case, for n large, we should reject (not reject) the null hypothesis if

$$-2\ln\lambda \geq (<)\chi^2_{\alpha,M}.$$

9.8. The Holy Trinity: likelihood ratio, Wald, and score tests.

- Let $L(\theta; X)$ be the likelihood function with $\theta \in \Theta \subset \mathbb{R}$, where X is the value of the sample $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$, and let $\widehat{\boldsymbol{\theta}}_{ML}$ be the maximum likelihood estimator for θ,

$$\widehat{\boldsymbol{\theta}}_{ML} = \underset{\theta \in \Theta}{\arg\max}\ L\left(\theta; \tilde{X}\right) = \underset{\theta \in \Theta}{\arg\max}\left[\ln L\left(\theta; \tilde{X}\right)\right].$$

- Let us just consider the simple null hypothesis $H_0 : \theta = \theta_0$ against the composite alternative hypothesis $H_1 : \theta \neq \theta_0$.

- **The likelihood ratio (LR) test statistic** is

$$\tilde{\lambda} = \frac{L\left(\theta_0; \tilde{X}\right)}{\sup\limits_{\theta \in \Theta} L\left(\theta; \tilde{X}\right)} = \frac{L\left(\theta_0; \tilde{X}\right)}{\max\limits_{\theta \in \Theta} L\left(\theta; \tilde{X}\right)} = \frac{L\left(\theta_0; \tilde{X}\right)}{L\left(\widehat{\theta}_{ML}; \tilde{X}\right)},$$

  and the null hypothesis is rejected when the value λ of this test statistic is low.

- Note that the LR test requires to evaluate the likelihood function both at the null hypothesis and at its unconstrained maximum.

- **Wald test statistic.** It is based on the (squared) standardized difference between the hypothesized value θ_0 and the unconstrained maximizing argument $\widehat{\boldsymbol{\theta}}_{ML}$ of the likelihood function:

$$\widetilde{W} = \frac{\left(\widehat{\boldsymbol{\theta}}_{ML} - \theta_0\right)^2}{\mathrm{Var}\left(\widehat{\boldsymbol{\theta}}_{ML}\right)} = \left(\frac{\widehat{\boldsymbol{\theta}}_{ML} - \theta_0}{\sigma_{\widehat{\boldsymbol{\theta}}_{ML}}}\right)^2,$$

  and the null hypothesis is rejected when the value W of this test statistic is high, i.e., when θ_0 is relatively far from $\widehat{\boldsymbol{\theta}}_{ML}$.

- Note that the Wald test only requires to find the unconstrained maximizing argument $\widehat{\boldsymbol{\theta}}_{ML}$ of the likelihood function.

Abraham Wald (1902 - 1950)

- **Score (or Rao or Lagrange multiplier) test statistic.** It is based on the square of the slope (relative to its curvature) of the log-likelihood function evaluated at the hypothesized value θ_0,

$$\widetilde{S} = \left. \frac{\left(\frac{\partial \ln L\left(\theta; \tilde{X}\right)}{\partial \theta} \right)^2}{-\mathrm{E}\left(\frac{\partial^2 \ln L\left(\theta; \tilde{X}\right)}{\partial \theta^2} \right)} \right|_{\theta=\theta_0},$$

  and the null hypothesis is rejected when the value S of this test statistic is high, i.e., when the slope of $\ln L\left(\theta; \tilde{X}\right)$ evaluated at θ_0 is far from zero so that $\ln L\left(\theta; \tilde{X}\right)$ evaluated at θ_0 is relatively far from its unconstrained maximum.

- Note that the denominator of $\widetilde{S}$, $-\mathrm{E}\left(\frac{\partial^2 \ln L\left(\theta; \tilde{X}\right)}{\partial \theta^2} \right)$, is the Fisher information $I\left(\theta\right)$ since $L\left(\theta; \tilde{X}\right) = h\left(\tilde{X}; \theta\right)$, where $h\left(X; \theta\right)$ is the density (probability function) of the random vector $\tilde{X}$. This denominator is a measure of the curvature of the function $\ln h\left(\cdot; \tilde{X}\right)$.

- The Fisher information measures also the amount of information that the random vector $\tilde{X}$ contains about the unknown parameter θ.

- If $h\left(\tilde{X};\cdot\right) = L\left(\cdot;\tilde{X}\right)$ (and, thus, its logarithm) is sharply peaked with respect to changes in θ, it is easy to infer the "correct" value of θ from observing the realization of $\tilde{X}$, i.e., the data provides a lot of information about the parameter θ. If $h\left(X;\cdot\right)$ is flat and spread-out, then it would more difficult to infer the actual "true" value of θ from observing the realization of $\tilde{X}$.

- The derivative of $\ln h\left(\tilde{X};\theta\right) = \ln L\left(\theta;\tilde{X}\right)$ with respect to θ is called the score, $\dfrac{\partial \ln h\left(\tilde{X};\theta\right)}{\partial\theta} = \dfrac{\partial \ln L\left(\theta;\tilde{X}\right)}{\partial\theta}$.

- In fact, the Fisher information is equal to the variance of the score,

$$I\left(\theta\right) = \mathrm{Var}\left(\frac{\partial \ln h\left(\tilde{X};\theta\right)}{\partial\theta}\right) = \mathrm{Var}\left(\frac{\partial \ln L\left(\theta;\tilde{X}\right)}{\partial\theta}\right).$$

- This is so because, under the assumptions needed for the proof of the Cramér-Rao lower bound, we know that

$$\mathrm{E}\left[\frac{\partial \ln h(\tilde{X};\theta)}{\partial \theta}\right] = 0,$$

so that

$$\mathrm{Var}\left(\frac{\partial \ln h(\tilde{X};\theta)}{\partial \theta}\right) = \mathrm{E}\left[\left(\frac{\partial \ln h(\tilde{X};\theta)}{\partial \theta}\right)^2\right],$$

and in the the handout about the Cramér Rao lower bound we proved that

$$\mathrm{E}\left[\left(\frac{\partial \ln h(\tilde{X};\theta)}{\partial \theta}\right)^2\right] = -\mathrm{E}\left[\frac{\partial^2 \ln h(\tilde{X};\theta)}{\partial \theta^2}\right].$$

- Therefore, the test statistic $\widetilde{S}$ is the square of the standardized score.

- Note also that the score test only requires to evaluate the derivatives of the likelihood function at the hypothesized value θ_0.
- The score test is also called Lagrange multiplier test. To see why, let us consider the following (trivial) constrained maximization problem:

$$\max_{\theta \in \Theta} \ln L\left(\theta; \tilde{X}\right)$$

subject to

$$\theta = \theta_0.$$

- The corresponding Lagrangean is

$$\mathrm{L}(\theta, \ell) = \ln L\left(\theta; \tilde{X}\right) + \ell\left(\theta_0 - \theta\right),$$

where ℓ is the Lagrange multiplier.

- The F.O.C. with respect to θ is

$$\frac{\partial \mathrm{L}(\theta, \ell)}{\partial \theta} = 0 \iff \frac{\partial \ln L\left(\theta; \tilde{X}\right)}{\partial \theta} = \ell.$$

- Therefore, the score test can be understood as a test of the magnitude of the Lagrange multiplier associated with the constraint $\theta = \theta_0$ at the constrained maximum.

- Note that, if the constraint were non-binding at the constrained maximum, the Lagrange multiplier ℓ should not differ from zero. Conversely, if the constraint $\theta = \theta_0$ is very binding then ℓ should be very different from zero.

- Note that, in this section, we assume that the set of parameter values under the null hypothesis has zero dimension $(\theta = \theta_0)$ and and the parameter space $\Theta \subset \mathbb{R}$ has dimension equal to one.

- If $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ is a random sample, then $-2 \ln \tilde{\lambda} = -2 \left[\ln L \left(\theta_0; \tilde{X} \right) - \ln L \left(\theta_{ML}; \tilde{X} \right) \right]$, $\widetilde{W}$, and $\widetilde{S}$ are all asymptotically distributed as a χ^2_M under the null hypothesis, where $M = 1$.

- Thus, for large random samples, the null hypothesis is rejected with a level of significance α when

$$-2 \ln \lambda \geq \chi^2_{\alpha,1} \ ,$$

$$W \geq \chi^2_{\alpha,1} \ ,$$

or

$$S \geq \chi^2_{\alpha,1} \ .$$

- The LR, Wald and score tests constitute the so-called "Holy Trinity" of tests.

9.9. Acceptance intervals

- Assume that the parameter value characterizing the distribution of the random vector (or sample) $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ is $\theta \in \Theta \subset \mathbb{R}$.
- Consider a real-valued test statistic $\tilde{t} = g\left(\tilde{X}\right)$ whose distribution is $P_{\tilde{t}}\left(\cdot;\theta\right)$.

- Consider now the following test, based on the test statistic $\tilde{t} = g\left(\tilde{X}\right)$, of the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta \neq \theta_0$:

$$\widehat{\phi}(\tilde{t}) = \begin{cases} 1 & \text{if } \tilde{t} \in C_{\widehat{\phi}} \\ & \\ 0 & \text{if } \tilde{t} \notin C_{\widehat{\phi}} \end{cases}$$

  and fix the level of significance $\alpha_{\widehat{\phi}}\ (= \alpha_{\widehat{\phi}}\ (\theta_0))$ equal to α (recall that the most commonly used levels are 0.05, 0.01 or 0.001).

- Then, if $\theta = \theta_0$, the critical region $C_{\widehat{\phi}}$ of size α must satisfy the following:

$$P\left\{g\left(\tilde{X}\right) \in C_{\widehat{\phi}}; \theta_0\right\} = P\left\{\tilde{t} \in C_{\widehat{\phi}}; \theta_0\right\} = P_{\tilde{t}}\left(C_{\widehat{\phi}}; \theta_0\right) = \alpha.$$

  or

$$P\left\{g\left(\tilde{X}\right) \notin C_{\widehat{\phi}}; \theta_0\right\} = P\left\{\tilde{t} \notin C_{\widehat{\phi}}; \theta_0\right\} = P_{\tilde{t}}\left(C_{\widehat{\phi}}^c; \theta_0\right) = 1 - \alpha.$$

- If we want that the acceptance region (which is the complement of the critical region) be an interval (a, b) then, under the null hypothesis, we must have

$$P\left\{a < g\left(\tilde{X}\right) < b; \theta_0\right\} = P_{\tilde{t}}\left\{(a, b); \theta_0\right\} = 1 - \alpha.$$

- Moreover, for two-tailed tests we make

$$P\left\{g\left(\tilde{X}\right) \leq a; \theta_0\right\} = P_{\tilde{t}}\left\{(-\infty, a]; \theta_0\right\} = \frac{\alpha}{2}$$

and

$$P\left\{g\left(\tilde{X}\right) \geq b; \theta_0\right\} = P_{\tilde{t}}\left\{[b, \infty); \theta_0\right\} = \frac{\alpha}{2}.$$

- Therefore, we reject the null hypothesis if

$$t = g(X) \notin (a, b)$$

and we do not reject the null hypothesis if

$$t = g(X) \in (a, b) \ .$$

- In applications, we use a test statistic $\tilde{t} = g\left(\tilde{X}, \theta\right)$ whose distribution $P_{\tilde{t}}(\cdot;\theta)$ under the null hypothesis is known and it is independent of the parameter θ. The random variable $\tilde{t}$ is thus a pivotal variable (see the next examples).

- **Examples (Two-tailed tests):**
- Let $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be a random sample from a normal population with mean μ and variance σ^2.
- **(1)** If σ^2 is known, the random variable $\tilde{z} = \dfrac{\bar{\mathbf{x}} - \mu_0}{\sigma / \sqrt{n}} \sim \mathrm{N}(0, 1)$ is a test statistic for the null hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$.
- If the level of significance is α, then the acceptance interval is $(-z_{\alpha/2}, z_{\alpha/2})$. We reject the null hypothesis if $|z| \geq z_{\alpha/2}$.

- **(2)** The random variable $\tilde{t} = \dfrac{\bar{\mathbf{x}} - \mu_0}{\mathbf{s} \big/ \sqrt{n}} \sim \mathrm{t}_{n-1}$ is a test statistic for the null hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$.

- If the level of significance is α, then the acceptance interval is $(-t_{\alpha/2,n-1}, t_{\alpha/2,n-1})$. We reject the null hypothesis if $|t| \geq t_{\alpha/2,n-1}$.

- **(3)** The random variable $\widetilde{\chi}^2 = \dfrac{(n-1)\mathbf{s}^2}{\sigma_0^2} \sim \chi^2_{n-1}$ is a test statistic for the null hypothesis $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$.

- If the level of significance is α, then the acceptance interval is $\left(\chi^2_{1-\frac{\alpha}{2},n-1}, \chi^2_{\alpha/2,n-1}\right)$. We reject the null hypothesis if $\chi^2 \notin \left(\chi^2_{1-\frac{\alpha}{2},n-1}, \chi^2_{\alpha/2,n-1}\right)$.

- Let $\mathbf{s_1^2}$ and $\mathbf{s_2^2}$ be the variances of independent random samples of size n_1 and n_2 from two normal populations with variances σ_1^2 and σ_2^2, respectively.

- **(4)** The random variable $\widetilde{F} = \dfrac{\mathbf{s_1^2}}{\mathbf{s_2^2}} \sim \mathrm{F}_{n_1-1,n_2-1}$ is a test statistic for the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$.

- If the level of significance is α, then the acceptance interval is

$$\left(F_{1-\frac{\alpha}{2},n_1-1,n_2-1}, F_{\alpha/2,n_1-1,n_2-1}\right) = \left(\frac{1}{F_{\alpha/2,n_2-1,n_1-1}}, F_{\alpha/2,n_1-1,n_2-1}\right).$$

We reject the null hypothesis if

$$F \notin \left(\frac{1}{F_{\alpha/2,n_2-1,n_1-1}}, F_{\alpha/2,n_1-1,n_2-1}\right).$$

- **Examples (One-tailed tests):**
- Let $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be a random sample from a normal population with mean μ and variance σ^2.
- **(1)** If σ^2 is known, the random variable $\dfrac{\bar{\mathbf{x}} - \mu_0}{\sigma / \sqrt{n}} \sim \mathrm{N}(0, 1)$ is a test statistic for the null hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu > (<) \mu_0$.
- If the level of significance is α, then the acceptance interval is $(-\infty, z_\alpha)$ (or $(-z_\alpha, \infty)$) and the critical interval is $[z_\alpha, \infty)$ (or $(-\infty, -z_\alpha]$).

- **(2)** The random variable $\dfrac{\bar{\mathbf{x}}-\mu_0}{\mathbf{s}/\sqrt{n}} \sim \mathrm{t}_{n-1}$ is a test statistic for the null hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu > (<) \mu_0$.

- If the level of significance is α, then the acceptance interval is $(-\infty, t_{\alpha,n-1})$ (or $(-t_{\alpha,n-1}, \infty)$) and the critical interval is $[t_{\alpha,n-1}, \infty)$ (or $(-\infty, -t_{\alpha,n-1}]$).

- **(3)** The random variable $\dfrac{(n-1)\mathbf{s^2}}{\sigma_0^2} \sim \chi^2_{n-1}$ is a test statistic for the null hypothesis $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 > (<) \sigma_0^2$.

- If the level of significance is α, then the acceptance interval is $\left(0, \chi^2_{\alpha,n-1}\right)$ (or $\left(\chi^2_{1-\alpha,n-1}, \infty\right)$) and the critical intervals is $\left[\chi^2_{\alpha,n-1}, \infty\right)$ (or $\left(0, \chi^2_{1-\alpha,n-1}\right]$).

- Let $\mathbf{s_1^2}$ and $\mathbf{s_2^2}$ be the variances of independent random samples of size n_1 and n_2 from two normal populations with variances σ_1^2 and σ_2^2, respectively.
- **(4)** The random variable $\frac{\mathbf{s_1^2}}{\mathbf{s_2^2}} \sim \mathrm{F}_{n_1-1,n_2-1}$ is a test statistic for the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 > (<) \sigma_2^2$.
- Let $H_1 : \sigma_1^2 > \sigma_2^2$. If the level of significance is α, then the acceptance interval is $\left(0, F_{\alpha,n_1-1,n_2-1}\right)$ and the critical interval is $\left[F_{\alpha,n_1-1,n_2-1}, \infty\right)$.

- Let $H_1 : \sigma_1^2 < \sigma_2^2$. If the level of significance is α, then the acceptance interval is

$$\left(F_{1-\alpha, n_1-1, n_2-1}, \infty\right) = \left(\frac{1}{F_{\alpha, n_2-1, n_1-1}}, \infty\right).$$

and the critical interval is

$$\left(0, F_{1-\alpha, n_1-1, n_2-1}\right] = \left(0, \frac{1}{F_{\alpha, n_2-1, n_1-1}}\right].$$

9.10. The p-value

- Assume again that the parameter value characterizing the distribution of the random vector (or sample) $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ is $\theta \in \Theta \subset \mathbb{R}$.
- Let $\tilde{t} = g\left(\tilde{X}\right)$ be a test statistic whose distribution under the null hypothesis is $P_{\tilde{t}}\left(\cdot; \theta_0\right)$.
- Consider the test of the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta \neq \theta_0$.
- The *p*-value is the probability under the null hypothesis of obtaining a value for the test statistic at least as extreme as that given by the sample value, $t = g(X)$.
- In other words, is the minimal size of the critical region for which we would reject the null hypothesis if the value of the statistic $\tilde{t}$ were t.

- If $t = g(X)$ is such that $P\{\tilde{t} \leq t; \theta_0\} \leq 1/2$, then

$$p\text{-value} = 2P\{\tilde{t} \leq t; \theta_0\} = 2P_{\tilde{t}}\{(-\infty, t]; \theta_0\},$$

whereas, if $P\{\tilde{t} \leq t; \theta_0\} > 1/2$, then

$$p\text{-value} = 2P\{\tilde{t} \geq t; \theta_0\} = 2P_{\tilde{t}}\{[t, \infty); \theta_0\}.$$

- Equivalently,

$$p\text{-value} = 2 \cdot \min\{P\{\tilde{t} \leq t; \theta_0\}, P\{\tilde{t} \geq t; \theta_0\}\}.$$

- If we fix the level of significance at α, then we reject H_0 if p-value $\leq \alpha$ and we do not reject H_0 if p-value $> \alpha$.

- In the one-tailed test of the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta > (<) \theta_0$, the p-value is either

$$p\text{-value} = P\{\tilde{t} \geq t; \theta_0\} \quad \text{or} \quad p\text{-value} = P\{\tilde{t} \leq t; \theta_0\}.$$

- We use stars to flag levels of significance for three of the most commonly used levels.

- If a p-value is less than 0.05, it is flagged with one star $(*)$. If a p-value is less than 0.01, it is flagged with two stars $(**)$. If a p-value is less than 0.001, it is flagged with three stars $(***)$.

- Note that the p-value gives us directly the strength of the evidence against the null hypothesis. This is an advantage of the p-value over checking whether the value of the test statistic falls in the critical region.

- This is a very interesting article about the use, misuse or abuse of p-values in Economics:

  Imbens, Guido W. (2021). Statistical significance, p-values, and the reporting of uncertainty. *Journal of Economic Perspectives* 95, 157-174.

9.11. Tests concerning means, differences between means, variances, proportions, and differences among several proportions

- See the handout.

9.12. Contingency tables - 9.13. Goodness of fit

- See the handout.

Monotone Likelihood Ratio Property

Proposition 1. If the family of density (or probability) functions $f(x;\theta)$, with $\theta \in \Theta \subset \mathbb{R}$, has a MLR then, for every $\theta_2 > \theta_1$, the following holds:

(a) First-order stochastic dominance:

$$F(x;\theta_2) \leq F(x;\theta_1) \text{ for all } x \in \mathbb{R}. \quad (1)$$

(b)

$$P\left\{\tilde{x} > x;\theta_2\right\} \geq P\left\{\tilde{x} > x;\theta_1\right\} \text{ for all } x \in \mathbb{R}. \quad (2)$$

(c) Non-increasing hazard rate:

$$\frac{f(x;\theta_1)}{1-F(x;\theta_1)} \geq \frac{f(x;\theta_2)}{1-F(x;\theta_2)} \text{ for all } x \in \mathbb{R}. \quad (3)$$

Proof. Let us prove it for the case where $f(x;\theta)$ is a pdf. The case where $f(x;\theta)$ is a pmf has an analogous proof by using sums instead of integrals.

Since $f(x;\theta)$ has a MLR, if $\theta_2 > \theta_1$ then

$$\frac{f(x_1;\theta_2)}{f(x_1;\theta_1)} \leq \frac{f(x_2;\theta_2)}{f(x_2;\theta_1)}, \text{ for } x_2 \geq x_1,$$

or

$$f(x_1;\theta_2)\cdot f(x_2;\theta_1) \leq f(x_1;\theta_1)\cdot f(x_2;\theta_2), \text{ for } x_2 \geq x_1,$$

or

$$f(x_1;\theta_2)\cdot f(x;\theta_1) \leq f(x_1;\theta_1)\cdot f(x;\theta_2) \text{ for } x \geq x_1 \quad (4)$$

or

$$f(x;\theta_2)\cdot f(x_2;\theta_1) \leq f(x;\theta_1)\cdot f(x_2;\theta_2) \text{ for } x_2 \geq x. \quad (5)$$

Let us fix x and integrate both sides of the inequality (4) with respect to x_1 on $(-\infty, x]$,

$$\int_{(-\infty,x]} f(x_1;\theta_2)\cdot f(x;\theta_1)dx_1 \leq \int_{(-\infty,x]} f(x_1;\theta_1)\cdot f(x;\theta_2)dx_1$$

$$\Longrightarrow F(x;\theta_2)\cdot f(x;\theta_1) \leq F(x;\theta_1)\cdot f(x;\theta_2)$$

or

$$\frac{f(x;\theta_2)}{f(x;\theta_1)} \geq \frac{F(x;\theta_2)}{F(x;\theta_1)}. \quad (6)$$

Similarly, let us fix x and integrate both sides of the inequality (5) with respect to x_2 on (x,∞),

$$\int_{(x,\infty)} f(x;\theta_2)\cdot f(x_2;\theta_1)dx_2 \leq \int_{(x,\infty)} f(x;\theta_1)\cdot f(x_2;\theta_2)dx_2$$

$$\Longrightarrow f(x;\theta_2)\cdot[1-F(x;\theta_1)] \le f(x;\theta_1)\cdot[1-F(x;\theta_2)]$$

or

$$\frac{1-F(x;\theta_2)}{1-F(x;\theta_1)} \ge \frac{f(x;\theta_2)}{f(x;\theta_1)}. \tag{7}$$

Combining (6) and (7), we get

$$\frac{1-F(x;\theta_2)}{1-F(x;\theta_1)} \ge \frac{F(x;\theta_2)}{F(x;\theta_1)},$$

which, after rearranging, simplifies to (1)

Expression (1) implies that

$$P\{\tilde{x} > x;\theta_2\} = 1-F(x;\theta_2) \ge 1-F(x;\theta_1) = P\{\tilde{x} > x;\theta_1\}, \text{ for all } x \in \mathbb{R},$$

which proves (2).

Finally, rearranging (7), we get (3). *Q.E.D.*

First-Order Stochastic Dominance

Proposition 2 (Rothschild and Stiglitz). Consider two random variables $\tilde{x}$ and $\tilde{y}$ having distributions whose supports are subsets of the bounded interval $[a,b]$ and their associated distribution functions are $F_{\tilde{x}}$ and $F_{\tilde{y}}$, respectively. Then, $F_{\tilde{x}}(z) \leq F_{\tilde{y}}(z)$ for all $z \in \mathbb{R}$ if and only if

$$\mathrm{E}\left[u(\tilde{x})\right] \equiv \int_{\mathbb{R}} u(z)dF_{\tilde{x}}(z) \geq \mathrm{E}\left[u(\tilde{y})\right] \equiv \int_{\mathbb{R}} u(z)dF_{\tilde{y}}(z),$$

for all Bernoulli utility functions u that are non-decreasing and continuous on $[a,b]$.

Proof.
(a) $F_{\tilde{x}}(z) \leq F_{\tilde{y}}(z)$ for all $z \in \mathbb{R} \Longrightarrow \mathrm{E}\left[u(\tilde{x})\right] \geq \mathrm{E}\left[u(\tilde{y})\right]$ for all u non-decreasing and continuous on $[a,b]$.

From the integration by parts formulae in Section 4.4 of Chapter 4, and since the function u can be viewed as a continuous distribution function on $[a,b]$, we have

$$\mathrm{E}\left[u(\tilde{x})\right] = \int_{[a,b]} u\left(z\right) dF_{\tilde{x}}(z) = u(a)\left[F_{\tilde{x}}(a) - F_{\tilde{x}}(a^{-})\right] + \int_{(a,b]} u\left(z\right) dF_{\tilde{x}}(z)$$

$$= F_{\tilde{x}}(a)u(a) - F_{\tilde{x}}(a^{-})u(a) + \overbrace{F_{\tilde{x}}(b)u(b) - \underbrace{F(a^{+})}_{\substack{F(a) \\ \text{by right-continuity}}} u(a) - \int_{(a,b]} F_{\tilde{x}}(z)du(z)}^{\text{integrating by parts}}$$

$$= \underbrace{F_{\tilde{x}}(b)}_{1}u(b) - \underbrace{F_{\tilde{x}}(a^{-})}_{P\{\tilde{x}<a;\theta\}=0} u(a) - \int_{(a,b]} F_{\tilde{x}}(z)du(z) = u(b) - \int_{(a,b]} F_{\tilde{x}}(z)du(z)$$

$$= u(b) - \int_{[a,b]} F_{\tilde{x}}(z)du(z),$$

where the last equality follows from the continuity of the "distribution function" u on $[a,b]$.

Similarly,

$$\mathrm{E}\left[u(\tilde{y})\right] = \int_{[a,b]} u\left(z\right) dF_{\tilde{y}}(z) = u(b) - \int_{[a,b]} F_{\tilde{y}}(z)du(z).$$

From the previous two expressions,

$$\mathrm{E}\left[u(\tilde{x})\right] - \mathrm{E}\left[u(\tilde{y})\right] = -\overbrace{\int_{[a,b]} \underbrace{\left[F_{\tilde{x}}(z) - F_{\tilde{y}}(z)\right]}_{\substack{\leq 0 \\ \text{Stochastic dominance}}} du\left(x\right)}^{\leq 0} \geq 0. \tag{8}$$

(b) $\mathrm{E}[u(\tilde{x})] \geq \mathrm{E}[u(\tilde{y})]$ for all u non-decreasing and continuous on $[a, b] \Longrightarrow F_{\tilde{x}}(z) \leq F_{\tilde{y}}(z)$ for all $z \in \mathbb{R}$.

By contradiction. Assume that there exists a $z' \in [a, b)$ such that $F_{\tilde{x}}(z') > F_{\tilde{y}}(z')$. The right-continuity of the distribution functions implies that there exists a $c \in (z', b)$ such that $F_{\tilde{x}}(z) > F_{\tilde{y}}(z)$ for all $z \in [z', c) \subset [a, b]$.

Consider the following Bernoulli utility function u:

$$u(z) = \begin{cases} \bar{u}_0 & \text{for } z < z' \\ u_0 + \left(\dfrac{\bar{u}_1 - \bar{u}_0}{c - z'}\right)(z - z') & \text{for } z \in [z', c) \\ \bar{u}_1 & \text{for } z > c, \end{cases}$$

where the two constants $\bar{u}_0$ and $\bar{u}_1$ satisfy $\bar{u}_1 > \bar{u}_0$. Note that this Bernoulli utility function u is non-decreasing and continuous for all $z \in \mathbb{R}$, and differentiable a.e. (it is non-differentiable only at z' and c).

Then, from (8),

$$\mathrm{E}[u(\tilde{x})] - \mathrm{E}[u(\tilde{y})] = -\int_{[a,b]} [F_{\tilde{x}}(z) - F_{\tilde{y}}(z)]\, du(z)$$

$$= -\int_{[z',c)} \underbrace{[F_{\tilde{x}}(z) - F_{\tilde{y}}(z)]}_{>0} \overbrace{\underbrace{\left(\frac{\bar{u}_1 - \bar{u}_0}{c - z'}\right)}_{>0} dz}^{du(z)} < 0$$

since $du(z) = u'(z)dz = 0$ for $z \notin [z', c)$. Thus, we arrive at a contradiction. *Q.E.D.*

The Karlin-Rubin Theorem

Theorem. Assume that (i) $\tilde{t} = t\left(\tilde{X}\right)$ is a test statistic that is a real-valued (or univariate) sufficient estimator for the parameter $\theta \in \Theta \subset \mathbb{R}$ characterizing the distribution of the random vector $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ and (ii) the family $g(t;\theta)$ of density (or probability) functions for $\tilde{t}$ has a MLR. Then, in testing the null hypothesis $H_0 : \theta \leq \theta_0$ against the alternative hypothesis $H_1 : \theta > \theta_0$, the test

$$\widehat{\phi}\left(\tilde{t}\right) = \begin{cases} 1 & \text{if } \tilde{t} > t_0 \\ \\ 0 & \text{if } \tilde{t} \leq t_0, \end{cases}$$

is a UMP test with size $\alpha = P\left\{\tilde{t} > t_0; \theta_0\right\} = \mathrm{E}_{\theta_0}\left(\widehat{\phi}\left(\tilde{t}\right)\right)$.

Proof. Fix a parameter value $\theta' \in \Theta$ such that $\theta' > \theta_0$ and consider testing $H_0' : \theta = \theta_0$ versus $H_1' : \theta = \theta'$ by using the test $\widehat{\phi}$. The MLR property implies that

$$t > t_0 \text{ if and only if } \frac{g(t;\theta')}{g(t;\theta_0)} > k',$$

with

$$k' = \inf_{t \in \mathcal{T}} \frac{g(t;\theta')}{g(t;\theta_0)},$$

where $\mathcal{T}$ is the set of real numbers t such that $t > t_0$ and $g(t;\theta')/g(t;\theta_0)$ is well defined in $\overline{\mathbb{R}}$ (i.e., $g(t;\theta') > 0$ or $g(t;\theta_0) > 0$ or both).

Since $\tilde{t}$ is a sufficient estimator for θ then, according to the factorization theorem and its proof in Chapter 8, the pdf/pmf of the random vector $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ can be written as

$$f\left(X;\theta\right) = h\left(X\right) \cdot g(t;\theta),$$

where the function $g(t;\theta)$ is the pdf/pmf of the estimator $\tilde{t}$. Thus,

$$\frac{g(t;\theta')}{g(t;\theta_0)} = \frac{h\left(X\right) \cdot g(t;\theta')}{h\left(X\right) \cdot g(t;\theta_0)} = \frac{f\left(X;\theta'\right)}{f\left(X;\theta_0\right)}.$$

Therefore, under the test $\widehat{\phi}$ we reject H_0' if and only if

$$\frac{f\left(X;\theta'\right)}{f\left(X;\theta_0\right)} > k' \iff \frac{f\left(X;\theta_0\right)}{f\left(X;\theta'\right)} < \frac{1}{k'} \equiv k.$$

Thus, according to the Neyman-Pearson lemma, the test $\phi(X) = \widehat{\phi}\left(t\left(\tilde{X}\right)\right)$ is a most powerful test with significance level no larger than $\alpha = P\left\{\tilde{t} > t_0; \theta_0\right\}$ for testing $H_0' : \theta = \theta_0$ versus $H_1' : \theta = \theta'$. Since θ' was arbitrary, the test based on the test function $\widehat{\phi}\left(\tilde{t}\right)$ is simultaneously UMP with significance level α for every $\theta' > \theta_0$. Therefore, the test $\widehat{\phi}$ is UMP with significance level α for testing $H_0' : \theta = \theta_0$ against $H_1 : \theta > \theta_0$.

Let

$$\pi_{\widehat{\phi}}\left(\theta\right) = P\left\{\tilde{t} > t_0; \theta\right\} = \mathrm{E}_\theta\left(\widehat{\phi}\left(\tilde{t}\right)\right)$$

be the power function of the test $\widehat{\phi}$ for testing the null hypothesis $H_0 : \theta \leq \theta_0$ against the alternative hypothesis $H_1 : \theta > \theta_0$. From the MLR assumption and equation (2) in Proposition 1, the power function $\pi_{\widehat{\phi}}\left(\theta\right)$ is non-decreasing since, for any $\theta_2 > \theta_1$,

$$\pi_{\widehat{\phi}}\left(\theta_2\right) = P\left\{\tilde{t} > t_0; \theta_2\right\} \geq P\left\{\tilde{t} > t_0; \theta_1\right\} = \pi_{\widehat{\phi}}\left(\theta_1\right).$$

Therefore,

$$\sup_{\theta \leq \theta_0} \pi_{\widehat{\phi}}\left(\theta\right) = \pi_{\widehat{\phi}}\left(\theta_0\right) = \alpha,$$

which means that the significance level of the test $\widehat{\phi}$ when the null hypothesis is $H_0 : \theta \leq \theta_0$ is the same as when the null hypothesis is $H_0' : \theta = \theta_0$.

Hence, the test based on the test function $\widehat{\phi}\left(\tilde{t}\right)$ it is a UMP test with significance level $\alpha = P\left\{\tilde{t} > t_0; \theta_0\right\}$ for testing the composite null hypothesis $H_0 : \theta \leq \theta_0$ against the composite alternative hypothesis $H_1 : \theta > \theta_0$. *Q.E.D.*

Hypothesis Testing

The test procedure for simple null hypotheses consists of the following four steps:

(1) State the null hypothesis H_0 and an appropriate alternative hypothesis H_1.

(2) Using the sampling distribution of an appropriate test statistic, determine a critical region of size α, where α is specified.

(3) Compute the value of the test statistic from sample data.

(4) Decide whether to reject the null hypothesis or whether to accept it.

Tests concerning means $(H_0 : \mu = \mu_0)$

Use the value of the test statistic

$$\frac{\overline{\boldsymbol{x}} - \mu_0}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1),$$

where $\overline{\boldsymbol{x}}$ is the mean of the random sample from a normal population.

If $n \geq 30$, we can replace σ by s. Moreover, for large n we can use the central limit theorem, even if the population is not normal.

If $n < 30$, σ^2 is unknown, and the random sample comes from a normal population, use the statistic

$$\frac{\overline{\boldsymbol{x}} - \mu_0}{\boldsymbol{s}/\sqrt{n}} \sim \mathrm{t}_{n-1}.$$

Notes (arising from the Karlin-Rubin theorem): (a) The previous procedure can be used to test the simple null hypothesis $\mu = \mu_0$ against the alternative hypothesis $\mu \neq \mu_0$ (composite alternative), $\mu = \mu_1 > \mu_0$ (simple alternative), $\mu = \mu_1 < \mu_0$ (simple alternative), $\mu > \mu_0$ (composite alternative), or $\mu < \mu_0$ (composite alternative).

(b) The same procedure can also be used to test the composite null hypothesis $\mu \geq \mu_0$ (or $\mu \leq \mu_0$) against the composite alternative hypothesis $\mu < \mu_0$ (or $\mu > \mu_0$). In this case α is the the maximum probability of committing a type I error for any value of μ assumed under the null hypothesis.

(c) The same logic applies to the next tests concerning differences between means, variances, proportions, and differences between proportions.

Tests concerning differences between means $(H_0 : \mu_1 - \mu_2 = \delta)$

Use the test statistic

$$\frac{\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2 - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \overset{a}{\sim} \mathrm{N}(0, 1)$$

where $\overline{\boldsymbol{x}}_1$ and $\overline{\boldsymbol{x}}_2$ are the means of two independent random samples.

If $n_1 \geq 30$ and $n_2 \geq 30$, we can replace σ_1^2 and σ_2^2 by s_1^2 and s_2^2, respectively.

If $n_1 < 30$ or $n_2 < 30$, σ_1^2 and σ_2^2 are equal but unknown, and the two random samples are independent and come from a normal population, use the test statistic

$$\frac{\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2 - \delta}{\boldsymbol{s}_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim \mathrm{t}_{n_1+n_2-2}, \quad \text{where} \quad \boldsymbol{s}_p^{\boldsymbol{2}} = \frac{(n_1-1)\boldsymbol{s}_1^{\boldsymbol{2}} + (n_2-1)\boldsymbol{s}_2^{\boldsymbol{2}}}{n_1+n_2-2}.$$

Tests concerning variances

Assume that you have random samples from normal populations.

a) $H_0 : \sigma^2 = \sigma_0$. Use the test statistic

$$\frac{(n-1)\boldsymbol{s}^{\boldsymbol{2}}}{\sigma_0^2} \sim \chi^2_{n-1}.$$

b) $H_0 : \sigma_1^2 = \sigma_2^2$. Use the test statistic

$$\frac{\boldsymbol{s}_1^{\boldsymbol{2}}}{\boldsymbol{s}_2^{\boldsymbol{2}}} \sim \mathrm{F}_{n_1-1,n_2-1}.$$

Tests concerning proportions $(H_0 : \theta = \theta_0)$

Let $\tilde{x} \sim \mathrm{B}(n;\theta)$. Then, we use the test statistic

$$\frac{\tilde{x} - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \overset{a}{\sim} \mathrm{N}(0,1).$$

Tests concerning differences among several proportions

Suppose that $x_1, x_2, \ldots, x_k$ are observed values of a set of random variables $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_k$ having binomial distributions with the respective parameters n_1 and θ_1, n_2 and θ_2, …, n_k and θ_k. If the n's are sufficiently large we use the test statistic

$$\frac{\tilde{x}_i - n_i\theta_i}{\sqrt{n_i\theta_i(1-\theta_i)}} \overset{a}{\sim} \mathrm{N}(0,1) \quad \text{for } i = 1,\ 2,\ \ldots,\ k.$$

Therefore,

$$\sum_{i=1}^{k} \frac{(\tilde{x}_i - n_i\theta_i)^2}{n_i\theta_i(1-\theta_i)} \overset{a}{\sim} \chi^2_k.$$

To test the null hypothesis $H_0 : \theta_1 = \theta_2 = \ldots = \theta_k = \theta_0$, against the alternative that at least one of the θ's does not equal θ_0, we can use the test statistic

$$\sum_{i=1}^{k} \frac{(\tilde{x}_i - n_i\theta_0)^2}{n_i\theta_0(1-\theta_0)} \overset{a}{\sim} \chi^2_k.$$

When θ_0 is not specified, that is, when we are interested only in the null hypothesis $\theta_1 = \theta_2 = \ldots = \theta_k$, we use the pooled estimator

$$\boldsymbol{\theta} = \frac{\tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_k}{n_1 + n_2 + \ldots + n_k}.$$

Then

$$\sum_{i=1}^{k} \frac{(\tilde{x}_i - n_i\boldsymbol{\theta})^2}{n_i\boldsymbol{\theta}(1-\boldsymbol{\theta})} \overset{a}{\sim} \chi^2_{k-1}. \tag{1}$$

The loss of one degree of freedom is due to the fact that we use an estimate for the unknown parameter θ (see the discussion in the next section).

An alternative form of the chi-square statistic for this kind of test is the following:

	successes	failures
Sample 1	x_1	$n_1 - x_1$
Sample 2	x_2	$n_2 - x_2$
...	...	...
Sample k	x_k	$n_k - x_k$

Let us refer to its entries as the observed cell frequencies, n_{ij}, $i = 1, 2, \ldots, k$ and $j = 1, 2$. Under the null hypothesis $\theta_1 = \theta_2 = \ldots = \theta_k = \theta_0$ the expected cell frequencies for the first column are $e_{i1} = n_i\theta_0$ for $i = 1, 2, \ldots, k$ and those for the second column are $e_{i2} = n_i(1-\theta_0)$. When θ_0 is unknown, we substitute for it, as before, the value θ of the pooled estimator $\boldsymbol{\theta}$ and estimate the cell frequencies as

$$e_{i1} = n_i\theta \quad \text{and} \quad e_{i2} = n_i(1-\theta), \quad \text{for } i = 1, 2, \ldots, k.$$

The value of the test statistic (1) can be written as

$$\sum_{i=1}^{k}\sum_{j=1}^{2} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}. \quad \leftarrow \text{Check it!!} \tag{2}$$

Contingency tables (Generalization of the previous section)

A $r \times c$ contingency table has c columns representing the different categories $A_1, A_2, \ldots, A_c$ of one variable, r rows representing the different categories $B_1, B_2, \ldots, A_r$ of another variable, and n_{ij} is the cell frequency for $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, c$. In the previous section we had $r = k$ and $c = 2$.

Schematically,

	A_1	A_2	...	A_c	
B_1	n_{11}	n_{12}	...	n_{1c}	$n_{1\cdot}$
B_2	n_{21}	n_{22}	...	n_{2c}	$n_{2\cdot}$
...	...	...	...	...	...
B_r	n_{r1}	n_{r2}	...	n_{rc}	$n_{r\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot c}$	n

,

where the row and the column totals $n_{i\cdot} = \sum_{j=1}^{c} n_{ij}$ and $n_{\cdot j} = \sum_{i=1}^{r} n_{ij}$ are called the marginal frequencies and n, the sum of all cell frequencies, is called the grand total.

The null hypothesis that we want to test is that the two variables are independent. If θ_{ij} is the probability that an item will fall into the cell (i, j), $\theta_{i\cdot}$

is the probability that an item will fall into the i^{th} row and $\theta_{\cdot j}$ is the probability that an item will fall into the j^{th} column. The null hypothesis we want to test is that

$$\theta_{ij} = (\theta_{i\cdot})(\theta_{\cdot j}) \quad \text{for } i = 1, 2, \ldots, r \text{ and } j = 1,\, 2, \ldots, c.$$

To test the null hypothesis against the alternative, that $\theta_{ij} \neq (\theta_{i\cdot})(\theta_{\cdot j})$ for at least a pair of values of i and j, we estimate the probabilities $\theta_{i\cdot}$ and $\theta_{\cdot j}$ as

$$\hat{\theta}_{i\cdot} = \frac{n_{i\cdot}}{n} \quad \text{and} \quad \hat{\theta}_{\cdot j} = \frac{n_{\cdot j}}{n},$$

and, hence, the expected cell frequencies as

$$e_{ij} = \left(\hat{\theta}_{i\cdot}\right)\left(\hat{\theta}_{\cdot j}\right) n = \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n} \cdot n = \frac{(n_{i\cdot})(n_{\cdot j})}{n}.$$

Then, we base our decision on

$$\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

and reject the null hypothesis if the value we obtain exceeds $\chi^2_{\alpha,(r-1)(c-1)}$. The degrees of freedom come from the fact that, if we know the frequencies of the $(r-1)(c-1)$ cells corresponding to the first $(r-1)$ rows and to the first $(c-1)$ columns, then we know the frequencies of the remaining cells since the total frequencies of rows and columns are fixed (think about it!).

Notes: (a) Since the previous test statistic has only approximately a chi-square distribution, we should use this test only when none of the e_{ij}'s is less than five. This sometimes requires that we combine some of the cells with a corresponding loss of degrees of freedom.

(b) *A general method to find the number of degrees of freedom:* Whenever expected cell frequencies in chi-square formulas are estimated on the basis of sample data, the number of degrees of freedom is $s-t-1$, where s is the number of terms in the sum and t is the number of independent parameters replaced by estimates. The -1 in the previous formula is due to the additional restriction that the sum of all the cell frequencies must be equal to the fixed grand total n.

(c) When testing for differences among k proportions (previous section), we had $s = 2k$ and $t = k$, since we had to estimate the k parameters $\theta_1, \theta_2, \ldots, \theta_k$ and the number of degrees of freedom was $2k - k - 1 = k - 1$.

(d) When testing for independence with an $r \times c$ contingency table, we have $s = r \cdot c$ and $t = r + c - 2$ since the $r + c$ parameters $\theta_{i\cdot}$ and $\theta_{\cdot j}$ are subject to the two restrictions that their respective sum must equal one. Hence, $s - t - 1 = rc - (r + c - 2) - 1 = (r-1)(c-1)$.

Goodness of fit (for discrete random variables)

This test applies to the situation in which we want to determine whenever a set of data may be looked as a random sample from a population having a given distribution. Let n_i be the observed absolute frequencies and e_i the expected absolute frequencies. To test the null hypothesis that a set of observed data comes from a population having a specified distribution against the alternative, that the population has some other distribution, we compute the value of the Pearson's statistic,

$$\sum_{i=1}^{m} \frac{(n_i - e_i)^2}{e_i},$$

and reject H_0 at the level of significance α, if the computed value exceeds $\chi^2_{\alpha,m-t-1}$, where m is the number of terms in the summation and t is the number of independent parameters estimated on the basis of sample data (see previous discussion).

Notes: (a) The set of values of the random variable is partitioned in adjacent classes so that each expected absolute frequency is at least five.

(b) The statistic considered in this section has an asymptotic distribution that is chi-square regardless of the distribution being tested.

Goodness of fit (for all types of random variables)

Consider a random sample of size n and let us order the values of the random sample as $x_1 < x_2 < \ldots < x_n$. The "sample distribution function" is constructed as follows:

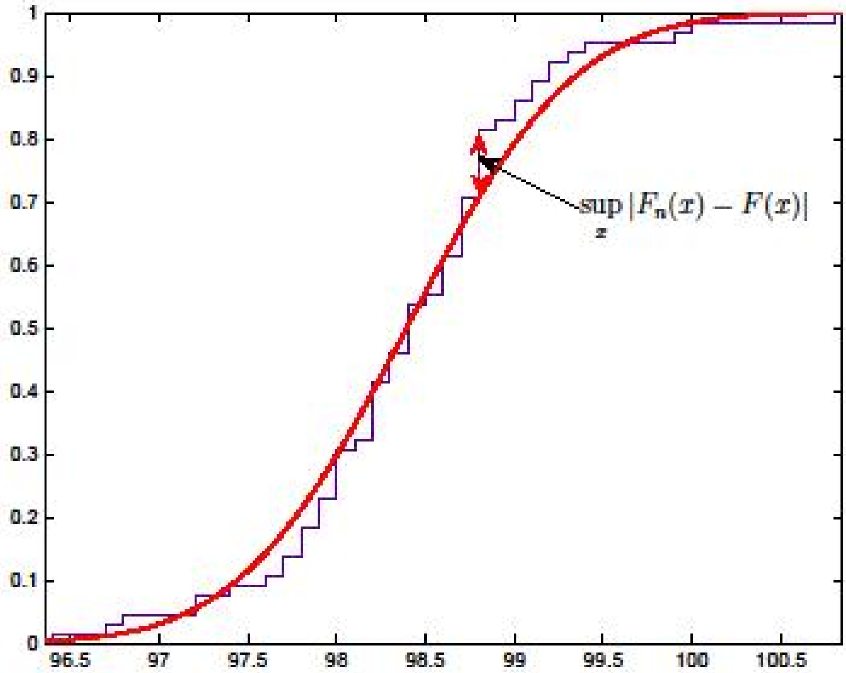
$$F_n(x) = \frac{j}{n} \quad \text{for} \quad x_j \leq x < x_{j+1}, \; j = 0, 1, \ldots, n,$$

with $x_0 = -\infty$ and $x_{n+1} = \infty$.

We compute the value of the Kolmogorov-Smirnov test statistic as follows:

$$D_n = \sup_x |F_n(x) - F(x)|,$$

where $F(x)$ is the distribution function of the population according to the null hypothesis H_0. If H_0 is true, D_n should be small.



Sample and true distribution functions.

The distribution of D_n is independent of F, but not of n. The following table gives the acceptance limits for such a test:

SAMPLE SIZE (n)	LEVEL OF SIGNIFICANCE $\propto$ FOR $D_n = \sup_x \lvert F_n(x) - F(x) \rvert$				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.210	.220	.240	.270	.320
30	.190	.200	.220	.240	.290
35	.180	.190	.210	.230	.270
OVER 35	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

The most serious limitation of this test is that the distribution must be fully specified. That is, if the parameters of $F(x)$ are estimated from the data, the critical region of the Kolmogorov-Smirnov test is no longer valid. It typically must be determined by simulation.

Comparison between two distributions

Let $F_{n_1}(x)$ be the sample distribution function of a random sample of size n_1 and $G_{n_2}(x)$ be the sample distribution function of a random sample of size n_2

(see previous discussion). The two random samples are independent.

Consider the following value of the two sample Kolmogorov-Smirnov test statistic:

$$D_{n_1,n_2} = \sup_x |F_{n_1}(x) - G_{n_2}(x)|.$$

The null hypothesis is that the two random samples come from identical populations. If H_0 is true, D should be small. The following table gives the acceptance limits for such a test:

Table gives critical *D*-values for $\alpha = 0.05$ (upper value) and $\alpha = 0.01$ (lower value) for various sample sizes. * means you cannot reject H_0 regardless of observed *D*.

$n_2 \backslash n_1$	3	4	5	6	7	8	9	10	11	12
1	* *	* *	* *	* *	* *	* *	* *	* *	* *	* *
2	* *	* *	* *	* *	* *	16/16 *	18/18 *	20/20 *	22/22 *	24/24 *
3	* *	* *	15/15 *	18/18 *	21/21 *	21/24 24/24	24/27 27/27	27/30 30/30	30/33 33/33	30/36 36/36
4		16/16 *	20/20 *	20/24 24/24	24/28 28/28	28/32 32/32	28/36 32/36	30/40 36/40	33/44 40/44	36/48 44/48
5			* *	24/30 30/30	30/35 35/35	30/40 35/40	35/45 40/45	40/50 45/50	39/55 45/55	43/60 50/60
6				30/36 36/36	30/42 36/42	34/48 40/48	39/54 45/54	40/60 48/60	43/66 54/66	48/72 60/72
7					42/49 42/49	40/56 48/56	42/63 49/63	46/70 53/70	48/77 59/77	53/84 60/84
8						48/64 56/64	46/72 55/72	48/80 60/80	53/88 64/88	60/96 68/96
9							54/81 63/81	53/90 70/90	59/99 70/99	63/108 75/108
10								70/100 80/100	60/110 77/110	66/120 80/120
11									77/121 88/121	72/132 86/132
12										96/144 84/144

For larger sample sizes, the approximate critical value D_α is given by the equation

$$D_\alpha = c(\alpha)\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

where the coefficient is given by the table below.

α	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

Examples: (1) At $\alpha = 0.05$ and samples sizes 5 and 8, $D_\alpha = 30/40 = 0.75$.

(2) At $\alpha = 0.01$ and samples sizes 15 and 28, $D_\alpha = 1.63\sqrt{\frac{15+28}{15\cdot 28}} = 0.522$.

Exercises. Probability and Statistics. IDEA.
9. Hypothesis Testing

1. Suppose we want to test the null hypothesis that the mean μ of a normal population with the variance $\sigma^2 = 1$ is μ_0 against the alternative hypothesis that it is μ_1, where $\mu_1 > \mu_0$. Let $\bar{x}$ be the value of the mean of a random sample of size n. Find the value of k (as a function of n) such that $\bar{x} > k$ provides a critical region of size $\alpha = 0.05$. Also determine the minimum sample size n needed for testing the null hypothesis $\mu = 10$ against the alternative hypothesis $\mu = 11$ by this procedure so that the probability β of committing a type II error is smaller than 0.05, $\beta \leq 0.05$.

2. A random sample of size n from a normal population with known variance σ^2 is to be used to test the null hypothesis $\mu = \mu_0$ against the alternative hypothesis $\mu = \mu_1$, where $\mu_1 > \mu_0$. Use the Neyman-Pearson lemma to find the uniformly most powerful test of size α.

3. Consider a coin having the probability θ of landing on heads and the following composite hypotheses:

$$H_0 \; : \; 0 \leq \theta \leq 0.5 \qquad \text{(null hypothesis)}$$

$$H_1 \; : \; 0.5 < \theta \leq 1 \qquad \text{(alternative hypothesis)}$$

   (a) Find the power functions of the following two tests:

   - *Test 1*: observe three independent flips of the coin and reject H_0 if three heads come up;

   - *Test 2*: observe three independent flips of the coin and reject H_0 if either two or three heads come up.

   (b) Discuss the properties of the previous two tests in terms of the associated probabilities of committing type I and type II errors.

4. Let $\tilde{x}_i$ have a Bernoulli distribution with unknown parameter θ, and let

$$\tilde{y} = \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 + \tilde{x}_4 + \tilde{x}_5$$

   denote the number of successes in five independent trials. To test $\theta = 0.5$ against $\theta \neq 0.5$ we make a test with critical region $\tilde{y} = 0, 1, 4$ or 5. Plot the power function and find the significance level of the test.

5. (a) Find the critical region of the likelihood ratio test with a level of significance α for testing the null hypothesis

$$H_0 : \; \mu = \mu_0$$

   against the composite alternative

$$H_1 : \; \mu \neq \mu_0$$

on the basis of a random sample of size n from a normal population with the known variance σ^2.

(b) Do the same with both the Wald test and the score test.

6. Suppose it is known from experience that the standard deviation of the weight of 8-ounce packages of cookies made by a certain bakery is 0.16 ounce. Let us assume that the weight of those packages of cookies is normally distributed. To check whether its production is under control on a given day, namely, to check whether the true average weight of the packages is 8 ounces, they select a random sample of 25 packages and find that their mean weight is $\bar{x} = 8.112$ ounces. Since the bakery stands to lose money when $\mu > 8$ and the customer loses out when $\mu < 8$, test the null hypothesis $\mu = 8$ against the alternative $\mu \neq 8$ using $\alpha = 0.01$. Find the corresponding p-value.

7. Suppose that 100 tires of a certain brand lasted on the average 21431 miles with standard deviation of 1295 miles. Using $\alpha = 0.05$, test the null hypothesis $\mu = 22000$ miles against the alternative hypothesis $\mu < 22000$ miles. Find the corresponding p-value.

8. Suppose that the specifications for a certain kind of ribbon call for a mean breaking strength of 185 pounds, and that five pieces randomly selected from different rolls have a mean breaking strength of 183.1 pounds with a sample standard deviation of 8.2 pounds. Assuming that we can look upon the data as a random sample from a normal population, test the null hypothesis $\mu = 185$ against the alternative $\mu < 185$ at $\alpha = 0.05$. Find the corresponding p-value.

9. Suppose that the nicotine content of two brands of cigarettes are being measured. If in an experiment fifty cigarettes of the first brand had an average nicotine content of $\bar{x}_1 = 2.61$ milligrams with a sample standard deviation of $s_1 = 0.12$ milligram, while forty cigarettes of the second brand had an average nicotine content of $\bar{x}_2 = 2.38$ milligrams with a sample standard deviation of $s_2 = 0.14$ milligram, test the null hypothesis $\mu_1 - \mu_2 = 0.2$ against the alternative $\mu_1 - \mu_2 \neq 0.2$, using $\alpha = 0.05$.

10. In the comparison of two kind of paint, a consumer testing service finds that four one-gallon cans of one brand cover on average 512 square feet with a standard deviation of 31 square feet, while four one-gallon cans of another brand cover on the average 492 square feet with a sample standard deviation of 26 square feet. Test the null hypothesis $\mu_1 - \mu_2 = 0$ against the alternative $\mu_1 - \mu_2 \neq 0$ at the significance level $\alpha = 0.05$. Assume that the two populations are normal and have equal variances.

11. Suppose that the thickness of a part used in a semiconductor is its critical dimension and that measurements of the thickness of a random sample of 18 such parts have the variance $s^2 = 0.68$, where the measurements are in the thousandths of an inch. The process is considered to be under control if the variation of the thickness is given by a variance not greater that 0.36. Assuming

that the measurements constitute a random sample from a normal population, test the null hypothesis $\sigma^2 = 0.36$ against the alternative hypothesis $\sigma^2 > 0.36$ at $\alpha = 0.05$. Find the corresponding p-value.

12. In comparing the variability of the tensile strength of two kind of structural steel, an experiment yielded the following results: $n_1 = 13$, $s_1^2 = 19.2$, $n_2 = 16$ and $s_2^2 = 3.5$, where the units of measurement are 1000 pounds per square inch. Assuming that the measurements constitute independent random samples from two normal populations, test the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative hypothesis $\sigma_1^2 \neq \sigma_2^2$ at the $\alpha = 0.02$ level of significance. Find the corresponding p-value.

13. An oil company claims that at most 20 percent of all automobile market owners buy brand A gasoline. Test this claim at $\alpha = 0.01$, if a random check indicates that 58 of 200 automobile owners buy brand A gasoline.

14. Determine, on the basis of the sample data shown in the following table, whether the true proportion of shoppers favoring detergent A over detergent B is the same in all three cities:

	Number favoring A	Number favoring B	
Los Angeles	232	168	$n_1 = 400$
San Diego	260	240	$n_2 = 500$
Fresno	197	203	$n_3 = 400$

Use the significance level $\alpha = 0.05$.

15. For the data shown in the table below, test for independence between a person's ability in mathematics and his or her interest in statistics. Use the 0.01 significance level.

		Ability in Math		
		Low	Average	High
Interest in Statistics	Low	63	42	15
	Average	58	61	31
	High	14	47	29

16. For the data in the table below, test whether the number of errors the compositor makes in setting a galley of type is a random variable having a Poisson distribution. Use the significance level $\alpha = 0.05$.

Number of errors per galley	Observed frequencies n_i
0	18
1	53
2	103
3	107
4	82
5	46
6	18
7	10
8	2
9	1
	440

17. Test at the 5% significance level the hypothesis that a distribution is normal with mean 32 and variance 3.24, using the following observed values of a random sample of size 10: 31.0, 31.4, 33.3, 33.4, 33.5, 33.7, 34.4, 34.9, 36.2, and 37.0.

18. The random variable $\tilde{x}$ has a lognormal density function,

$$f(x;\mu,\sigma) = \frac{1}{\left(\sqrt{2\pi}\right)\sigma \cdot x} e^{-\frac{1}{2}\left(\frac{(\ln x)-\mu}{\sigma}\right)^2}, \text{ for } 0 < x < \infty.$$

Note that the random variable $\ln \tilde{x}$ is normal with the mean μ and the variance σ^2.

Consider the null and alternative hypotheses:

$$H_0: \quad \mu = 80; \ \sigma = 20$$

$$H_1: \quad \mu = 110; \ \sigma = 20$$

Given a random sample of size n from a population having the density $f(x;\mu,\sigma)$, use the Neyman-Pearson Lemma to devise a statistical test to discriminate between H_0 and H_1 with the level of significance $\alpha = 0.05$. That is, apply the Lemma to find the appropriate test statistic and the precise critical region with size $\alpha = 0.05$ yielding the most powerful test. The critical region you must construct should be a function of the sample size n only.

19. On the basis of **a single** observation, we want to test the simple null hypothesis that the probability function of $\tilde{x}$ is

x	1	2	3	4	5	6	7
$f(x)$	1/12	1/12	1/12	1/4	1/6	1/6	1/6

against the composite alternative that the probability function is

x	1	2	3	4	5	6	7
$g(x)$	$a/3$	$b/3$	$c/3$	$2/3$	0	0	0

where $a+b+c=1$.

(a) Using the likelihood ratio technique, find the critical region with size $\alpha = 0.25$ for which the null hypothesis is rejected. Which is the corresponding probability of a type II error?

(b) Consider the critical region for which the null hypothesis is rejected only when $x = 4$. Which is the size of such a critical region? Which is the corresponding probability of a type II error?

(c) Compare the test obtained in (a) with the one in (b). Which one do you prefer? Does the comparison contradict the Neyman-Pearson lemma?

20. R.D. Clarke divided the south of London into 576 squares of 0.1 square miles each and counted the number of flying bombs hitting those squares during the WWII. He obtained the following results:

Number of bombs per square	0	1	2	3	4	5	6 or more
Number of squares (frequency)	229	211	93	35	7	1	0

Test the null hypothesis that the bombs were hitting randomly those squares against the alternative that the bombs were directed towards a specific target. Use a significance level of 0.01.

21. Let $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$ be a random variable whose range is $\tilde{x}(\Omega) = \{0, 1, 2\}$. We observe the value $\{x_1, x_2\}$ of a random sample $\{\tilde{x}_1, \tilde{x}_2\}$ of size 2 from the population $\tilde{x}$.

We want to test the null hypothesis that the probability function of the random variable $\tilde{x}$ is given by the following uniform probability function:

$$f(x) = \frac{1}{3}, \quad \text{for } x = 0, 1, 2; \qquad (*)$$

against the alternative that the probability function of $\tilde{x}$ is

$$g(x) = \begin{cases} 0 & \text{for } x = 0 \\ \frac{2}{3} & \text{for } x = 1 \\ \frac{1}{3} & \text{for } x = 2 \end{cases}$$

(a) Devise the most powerful test with a level of significance $\alpha = 1/9$. What is the power of this test?

(b) Consider a test where the sample mean $\bar{\boldsymbol{x}} = \dfrac{\tilde{x}_1 + \tilde{x}_2}{2}$ is the test statistic. Find the value of k such that $\bar{x} \geq k$ provides a critical region of size $\alpha = 1/9$, where $\bar{x}$ is the value of the sample mean $\bar{\boldsymbol{x}}$. What is the power of this test?

Suppose now that we want to test the null hypothesis that the probability function of the discrete random variable $\tilde{x}$ is the uniform probability function given in $(*)$ against the alternative COMPOSITE hypothesis that the probability function of $\tilde{x}$ IS NOT the one given in $(*)$.

(c) Using again the value $\{x_1, x_2\}$ of a random sample $\{\tilde{x}_1, \tilde{x}_2\}$ of size 2 from the population $\tilde{x}$, devise a likelihood ratio test with a level of significance $\alpha = 1/3$. You should provide the test statistic and the critical region for this test.

22. Prove that the two test statistics given in (1) and (2) in the handout "Hypothesis Testing" are equivalent.

23. Assume that you have obtained x "successes" in three identical and independent trials. Let θ be the probability of success in each trial. Devise a likelihood ratio test with a level of significance $\alpha = 1/4$ to test the null hypothesis that $\theta = 1/2$ against the alternative hypothesis that $\theta \neq 1/2$. You should provide the test statistic and the critical region for this test.

    Hint: Note that $\lim\limits_{x \to 0} x^x = 1$ or, equivalently, $\lim\limits_{x \to 0} (x \cdot \ln x) = 0$, which follows from L'Hôpital's rule.

24. Let $\{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n\}$ be a random sample of size n from a normal population with unknown mean μ and unknown variance σ^2. Consider the following statistic:

$$\hat{\boldsymbol{s}}^2 = \frac{\sum_{i=1}^{n} (\tilde{x}_i - \overline{\boldsymbol{x}})^2}{n},$$

    where $\overline{\boldsymbol{x}}$ is the sample mean.

    Assume that the sample size is $n = 20$. Use a level of significance $\alpha = 0.05$ to construct an acceptance interval (a, b) for the statistic $\hat{\boldsymbol{s}}^2$ to test the null hypothesis $H_0 : \sigma^2 = 5$ against the alternative $H_1 : \sigma^2 \neq 5$. The interval should be such that you do not reject H_0 if $a < \hat{s}^2 < b$ and you reject it if $\hat{s}^2 \notin (a, b)$. Moreover, $P\left\{\hat{\boldsymbol{s}}^2 \leq a\right\} = P\left\{\hat{\boldsymbol{s}}^2 \geq b\right\} = \alpha/2$ under the null hypothesis.

25. Let $\{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n\}$ be a random sample of size n from a normal population with unknown mean μ and known variance σ^2. Consider testing $H_0 : \mu \geq \mu_0$ against $H_1 : \mu < \mu_0$. Use the Karlin-Rubin theorem to construct a uniformly most powerful (UMP) test with significance level α based on the test statistic $\bar{\boldsymbol{x}}$ (the sample mean).

26. You know that $\tilde{x}$ is a normally distributed random variable with a standard deviation of 200 and that its mean is either 700 or 800. You wish to make a judgement about the value of the mean of $\tilde{x}$. You may choose among the following two pieces of information (or statistics) for your judgement:

(a) The mean $\overline{\boldsymbol{x}}$ of a random sample of size 25 from the population of $\tilde{x}$.

(b) The proportion $\hat{\boldsymbol{\theta}}$ of observations with a value smaller than 800 in a random sample of size 49 from the population of $\tilde{x}$.

We impose that the probability of the mistake of saying $\mu = 700$ if in fact $\mu = 800$ must be the same as the probability of the mistake of saying $\mu = 800$ if in fact $\mu = 700$.

Which of the previous two statistics do you prefer for judging μ?

Note: For the statistic in (b) use the corresponding approximation by the normal distribution (as dictated by the central limit theorem) since such an approximation is excellent in this case.

27. Let $\{\tilde{x}_i\}_{i=1}^n$ be a random sample of size $n = 70$ from a population $\tilde{x}$. Assume that the population distribution is Poisson with the parameter $\lambda > 0$. Consider the null hypothesis $H_0 : \lambda = 5$ and the alternative hypothesis $H_1 : \lambda = 6$.

    (a) Use the Neyman-Pearson Lemma to test H_0 against H_1 with the level of significance $\alpha = 0.05$. That is, apply the Lemma to find the appropriate test statistic and the precise critical region with size $\alpha = 0.05$ yielding the most powerful test. *Hint:* Use the Central Limit Theorem to find a good approximation for the standardization of the sample mean.

    (b) What is the power of the test you have devised in part (a)?

28. Check that the two formulae to compute the expected utility shown in the handout "First-Order Stochastic Dominance",

$$\mathrm{E}\left[u(\tilde{x})\right] = \int_{[a,b]} u\left(x\right) dF_{\tilde{x}}(x) = u(b) - \int_{[a,b]} F_{\tilde{x}}(x) du(x),$$

    yield exactly the same value for the following example. The random variable $\tilde{x}$, which represents consumption, is discrete and its support is a subset of the closed interval $[0, 1]$. In particular, $\tilde{x}$ only takes on three values: 0 with probability 1/2, 1/2 with probability 1/4, and 1 with probability 1/4. The Bernoulli utility function is $u(x) = -x^2 + 2x + 1$, which is continuous and increasing for $x \leq 1$.

29. Consider a normal population with the known variance $\sigma^2 = 3$ and a random sample $\{\tilde{x}_1, ..., \tilde{x}_n\}$ with size $n = 12$.

    (a) We want to test the null hypothesis that the mean μ of this population is equal to 7 ($H_0 : \mu = 7$) against the alternative simple hypothesis that it is 9 ($H_1 : \mu = 9$). Let $\bar{x}$ be the value of the mean of the random sample. Find the value of k such that $\bar{x} > k$ provides a critical region of size $\alpha = 0.05$. Compute the power of this test.

    (b) Construct a likelihood ratio test with a level of significance $\alpha = 0.05$ for testing the null hypothesis $H_0 : \mu = 7$ against the composite alternative

$H_1 : \mu \neq 7$. You should provide the appropriate test statistic and the precise critical region. *Hint:* for this part, you should prove that

$$\sum_{i=1}^{12} \left[(x_i - 7)^2 - (x_i - \bar{x})^2 \right] = 12(\bar{x} - 7)^2.$$

(c) Plot the power function of the test found in part (b).