

Introduction to Probability and Statistics
for Engineers and Scientists
Sheldon M Ross
Notes and Exercises

Anirudh Krishnan

August 24, 2021

Contents

1	Parameter Estimation	2
2	Hypothesis Testing	13

Chapter 1

Parameter Estimation

“Please explain in detail your reasons for choosing this prior distribution”

Consider a distribution F_θ from which a random sample $\{X_i\}$ is drawn. It is an extremely common problem to estimate the vector of parameters θ of the underlying distribution using only the information from the sample.

Estimates of these parameters can either be provided as a *point* or an *interval*, alongside a *confidence* value associated with these estimates.

An important class of such problems is *Bayesian estimation*, which involves estimating an unknown parameter only partial information about it is known.

Maximum Likelihood Estimator : An estimator of an unknown parameter θ is any sample-derived statistic used to estimate its value. The observed value of an estimator is called the *estimate*.

As an example consider n random samples drawn from an underlying exponential distribution whose mean is the unknown parameter θ . The joint PDF of the RVs $\{X_i\}$, which would then be used to estimate θ is given by

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n) \\ &= \frac{1}{\theta} e^{-x_1/\theta} \frac{1}{\theta} e^{-x_2/\theta} \dots \frac{1}{\theta} e^{-x_n/\theta} \quad x_i > 0 \\ &= \frac{1}{\theta^n} \exp \left\{ \sum_{i=1}^n \frac{-x_i}{\theta} \right\} \end{aligned} \tag{1.1}$$

Likelihood function : Define $f(x_1, x_2, \dots, x_n | \theta)$ as the joint PDF or PMF of the n random samples from an underlying distribution with unknown parameter θ . This term is the likelihood that the values $\{x_i\}$ will actually be observed when θ is the true value of the parameter.

The *maximum likelihood estimate* $\hat{\theta}$ is that value of the parameter θ which maximizes the likelihood function. A useful computational convenience is that log likelihood also has its maximum at the same point as likelihood itself.

The algorithm usually involves identifying the likelihood function after incorporating the unknown parameter θ , taking logarithm wherever convenient, and then differentiating with respect to θ in order to maximize the likelihood.

A sample of size n from an underlying Bernoulli distribution whose unknown mean p is to be estimated. The above procedure yields the maximum likelihood estimator of p to be the sample mean $\sum X_i/n$.

The above problem when the underlying distribution is Poisson with rate λ yields the MLE as $\hat{\lambda} = \sum X_i/n$.

However, for a uniform distribution on $[0, \theta]$, the parameter is estimated by $\hat{\theta} = \max(\{X_i\})$ and not by the sample mean.

The more relevant problem is with an underlying normal distribution with unknown μ and σ , to be estimated using a sample of size n .

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X} \quad \text{and} \quad \hat{\sigma} = \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{1/2} \quad (1.2)$$

$$(1.3)$$

Note that $\hat{\sigma}$ differs from the sample standard deviation S , in the denominator being n instead of $(n - 1)$.

Failure and survival rate : Defining failure rate λ_i and survival rate $s_i = 1 - \lambda_i$ for an underlying discrete distribution taking values $i \geq 1$,

$$\lambda_i = P\{X = i \mid X > i - 1\} = \frac{P\{X = i\}}{P\{X > i - 1\}} \quad (1.4)$$

$$s_i = 1 - \lambda_i = \frac{P\{X > i\}}{P\{X > i - 1\}} \quad (1.5)$$

For illustration, let X denote a human's age at death. Now, failure rate is a measure of the fraction of all people surviving at least $i - 1$ years who will die at age i .

Using the relations,

$$\begin{aligned} P\{X > i\} &= s_1 s_2 \dots s_i \\ P\{X = n\} &= \lambda_n P\{X > n - 1\} = s_1 s_2 \dots s_{n-1} (1 - s_n) \end{aligned} \quad (1.6)$$

Using the above PMF for X requires estimating the survival rates \hat{s}_i . This is a simple task of looking at what fraction of the sample who reached age i one year ago are alive today.

Interval estimates : It is often useful to provide an interval for the estimate of a parameter along with an associated confidence value instead of just a point value. This involves looking at the probability distribution of the estimator itself.

As an example, consider the MLE of the mean of a normal RV from which a sample of size n is drawn. The MLE \bar{X} is known to have a normal distribution with mean μ and variance σ^2/n .

Rearranging the sample mean to form a standard normal RV Z , and using its CDF to find a 95% confidence interval gives,

$$\begin{aligned} \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} &\sim Z \\ P\left\{|\bar{X} - \mu| \leq \frac{1.96 \sigma}{\sqrt{n}}\right\} &= 0.95 \\ \text{with 95 \% confidence, } \mu &\in \left[\bar{x} - \frac{1.96 \sigma}{\sqrt{n}}, \bar{x} + \frac{1.96 \sigma}{\sqrt{n}}\right] \end{aligned} \quad (1.7)$$

The above relation states that if the sample mean $\bar{X} = \bar{x}$, then the population mean μ lies within this interval with 95% confidence.

In addition to the *two-sided* confidence interval for μ above, the *one-sided upper* and *one-sided lower* 95% confidence intervals for μ are,

$$\mu \in \left[\bar{x} - \frac{1.645 \sigma}{\sqrt{n}}, \infty \right) \quad \text{and} \quad \mu \in \left(-\infty, \bar{x} + \frac{1.645 \sigma}{\sqrt{n}} \right] \quad (1.8)$$

The factors of 1.96 and 1.645 used above were obtained using the inverse CDF of the normal distribution for $p = 0.975$ and $p = 0.95$ respectively. A different confidence interval would involve recalculating the inverse CDF and replacing the pre-factors used above.

Confidence interval for a normal mean with unknown variance : Consider a normal RV with unknown μ and σ , from which a sample of size n is drawn. Since the above procedure can no longer be used to construct confidence intervals, the sample variance S is now of interest.

Using the fact that $\sqrt{n}(\bar{X} - \mu)/S \sim t_{n-1}$, and the symmetry of the t-distribution, for some observed value of sample mean and variance \bar{x}, s^2

$$1 - \alpha = P \left\{ |\bar{X} - \mu| \leq \frac{t_{\alpha/2, n-1} S}{\sqrt{n}} \right\}$$

$$\mu \in \left[\bar{x} - \frac{(t_{\alpha/2, n-1}) s}{\sqrt{n}}, \bar{x} + \frac{(t_{\alpha/2, n-1}) s}{\sqrt{n}} \right] \quad (1.9)$$

This becomes the $100(1 - \alpha)\%$ confidence interval for μ . As $n \rightarrow \infty$, the t-distribution reduces to the standard normal and the new confidence interval expression reduces to the old one. It is common, but not necessary that the confidence interval when σ is unknown is larger than when it is known, by virtue of the t-distribution being wider than the standard normal distribution, and the general trend of $S^2 > \sigma^2$.

The above methods have assumed an underlying normal distribution. However, for a large enough sample size, they may be used even when the underlying PDF is not a normal

RV because of the Central Limit theorem.

Monte-Carlo simulation : The process of randomly sampling from a uniform distribution to estimate difficult integrals using the power of the central limit theorem acting on a large sample consisting of i.i.d. RVs.

For example, consider a set of n uniform RVs $\{U_i\}$ in the domain $[a, b]$

$$\theta = \int_a^b g(x) \, dx = \mathbb{E}[g(x)] \quad (1.10)$$

A new set of RVs $\{X_i\} = g(\{U_i\})$, which are random samples drawn from the PDF $g(x)$, would now have sample mean and variance \bar{x} and s^2 . A confidence interval for $\mathbb{E}[g(x)]$ is now constructed using the t-distribution method described above.

Prediction Interval : The problem of predicting a confidence interval for the next sample X_{n+1} drawn from a normal distribution with parameters μ, σ given a set of n samples already available, is similar to the above procedure.

Let S_n^2 and \bar{X}_n be the sample variance and mean using the first n samples. Since \bar{X}_n is a normal RV with mean μ and variance σ^2/n , whereas X_{n+1} is also a normal RV independent of \bar{X}_n ,

$$\begin{aligned} \frac{X_{n+1} - \bar{X}_n}{\sigma \sqrt{1 + 1/n}} &\sim Z \\ \frac{X_{n+1} - \bar{X}_n}{S_n \sqrt{1 + 1/n}} &\sim t_{n-1} \end{aligned} \quad (1.11)$$

Once again, this leads to the confidence intervals using the t-distribution defined above.

Confidence interval for a normal variance with unknown mean : Using the fact that $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$, the $100(1-\alpha)\%$ confidence interval for a sample of size n with sample variance $S^2 = s^2$ is calculated as,

$$\sigma^2 \in \left[\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \right] \quad (1.12)$$

Note the use of $\chi_{\alpha/2, n-1}^2$ as well as $\chi_{1-\alpha/2, n-1}^2$ arising from the fact that the χ_n^2 distribution is not symmetric about its mean.

Difference in means of two normal populations : The problem is to estimate $\mu_1 - \mu_2$, where two normal RVs X, Y with means μ_1, μ_2 and variances σ_1^2, σ_2^2 respectively are used to generate two sets of samples $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_m\}$.

Additionally, the sample sets $\{X_i\}, \{Y_j\}$ are independent.

The point estimator of $\mu_1 - \mu_2$ is simply $\bar{X} - \bar{Y}$, using the previous section. For an interval estimate, the distribution of $\bar{X} - \bar{Y}$ is needed.

$$\begin{aligned} \bar{X} - \bar{Y} &\sim \mathcal{N} \left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right) \\ Z &\sim \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \end{aligned} \quad (1.13)$$

When the variances σ_1^2, σ_2^2 are known, the $100(1 - \alpha)\%$ confidence interval is,

$$\mu_1 - \mu_2 \in \left[\bar{x} - \bar{y} - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \bar{x} - \bar{y} + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right] \quad (1.14)$$

Next, consider the special case where σ_1, σ_2 are unknown but happen to be equal. Now, using the sample variances S_1^2, S_2^2 , and the additive property of chi-squared distributions, the *pooled estimator* S_P^2 of this common variance is

$$S_P^2 = \frac{(n-1) S_1^2 + (m-1) S_2^2}{(n+m-2)}$$

Rearranging the sample statistics into a t-distribution after replacing σ with S_P ,

$$\begin{aligned} Z &\sim \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2(1/n + 1/m)}} \\ \chi_{n+m-2}^2 &\sim (n + m - 2) \frac{S_P^2}{\sigma^2} \\ t_{n+m-2} &\sim \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_P^2(1/n + 1/m)}} \end{aligned} \quad (1.15)$$

Now, the confidence interval is defined using this t-distribution as

$$\mu_1 - \mu_2 \in \left[\bar{x} - \bar{y} - t_{\alpha/2, n+m-2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{x} - \bar{y} + t_{\alpha/2, n+m-2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right] \quad (1.16)$$

Mean of a Bernoulli RV - confidence interval : Consider a sample of size n drawn from a Bernoulli RV with parameter p . The point estimate for p is simply $\hat{p} = \sum X_i/n$.

For an interval estimate, assume $X = \sum X_i$ has the observed value x , and define the confidence interval for p as,

$$\begin{aligned} Z &\sim \frac{X - np}{\sqrt{np(1-p)}} \\ p &\in \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] \end{aligned} \quad (1.17)$$

For a given interval of width b in the above relation, the number of samples required n is given by

$$b = 2z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{and} \quad n = (2z_{\alpha/2})^2 \frac{\hat{p}(1-\hat{p})}{b^2} \quad (1.18)$$

To guarantee a confidence interval of size b or smaller, the minimum number of samples needed is $n^* = (Z_{\alpha/2}/b)^2$

Mean of exponential distribution - confidence interval : Consider a set of samples $\{X_1, \dots, X_n\}$ drawn from an exponential RV with unknown mean θ . The point estimate for θ is just the sample mean \bar{X} .

For an interval estimate, consider the following RV transformations,

$$\begin{aligned} \sum X_i &\sim \Gamma(n, 1/\theta) \\ \frac{2}{\theta} \sum X_i &\sim \Gamma(n, 1/2) \sim \chi_{2n}^2 \end{aligned} \quad (1.19)$$

Using the above chi-squared distribution, the confidence interval for θ is given by,

$$\theta \in \left[\frac{2 \sum X_i}{\chi_{\alpha/2, 2n}^2}, \frac{2 \sum X_i}{\chi_{1-\alpha/2, 2n}^2} \right] \quad (1.20)$$

Evaluating point estimators : The question of how good a proposed point estimator $d(\mathbf{X})$ for a parameter θ drawn from a sample $\mathbf{X} = \{X_1, \dots, X_n\}$ is often answered by looking at the mean squared error $r(d, \theta)$

$$r(d, \theta) = \mathbb{E}[(d(X) - \theta)^2] \quad (1.21)$$

Generally, there is no universal value of d that minimizes the error above for all values of θ .

Bias of an estimator : For the definitions used above, the bias $b_\theta(d)$ is expressed as,

$$b_\theta(d) = \mathbb{E}[d(\mathbf{X})] - \theta \quad (1.22)$$

If an estimator has 0 bias for all values of θ , then it is called an *unbiased* estimator of θ . For example, a weighted sum of the random samples from an unknown distribution with mean θ is always an unbiased estimator.

$$\begin{aligned}\mathbb{E}\left[\sum \lambda_i X_i\right] &= \sum \mathbb{E}[X_i] \lambda_i \\ &= \theta \sum \lambda_i = \theta \\ \text{if } \sum \lambda_i &= 1\end{aligned}\tag{1.23}$$

The mean squared error of an unbiased estimator is equal to its variance.

$$\begin{aligned}r(d, \theta) &= \mathbb{E}[(d(\mathbf{X}) - \theta)^2] = \mathbb{E}[(d(\mathbf{X}) - \mathbb{E}[d(\mathbf{X})])^2] \\ &= \text{Var}(d(\mathbf{X}))\end{aligned}\tag{1.24}$$

For a given parameter θ , a linear interpolant between two independent unbiased estimators d_1, d_2 is also unbiased. If σ_1^2, σ_2^2 are the respective known variances of these estimators, then this new estimator has the form

$$\begin{aligned}d &= \lambda d_1 + (1 - \lambda) d_2 \\ r(d, \theta) &= \text{Var}(d) = \lambda^2 \sigma_1^2 + (1 - \lambda)^2 \sigma_2^2 \\ \hat{\lambda} &= \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad \text{to minimize} \quad r(d, \theta)\end{aligned}\tag{1.25}$$

as a generalization of the above, consider a composite estimator d made up of many independent unbiased estimators d_i each with the same mean θ and separate variances σ_i^2 . The best construction of d would then be,

$$d = \frac{\sum d_i / \sigma_i^2}{\sum 1 / \sigma_i^2} \quad (1.26)$$

$$r(d, \theta) = \text{Var}(d) = \frac{1}{\sum (1 / \sigma_i^2)} \quad (1.27)$$

For a general estimator with some nonzero bias, its mean-squared error is given by

$$r(d, \theta) = \text{Var}(d) + b_\theta^2(d) \quad (1.28)$$

Bayes estimator : Consider the special case when some information about a parameter θ is available even before the observations $\{X_i\}$ are obtained. This is called the *prior distribution* of θ .

However, once the set of n random variables $\{X_i = x_i\}$ have been observed, it is possible to update the prior PDF $p(\theta)$ using this new information to yield the *posterior distribution* $f(\theta \mid x_1, \dots, x_n)$.

The update procedure uses the *likelihood* $f(x \mid \theta)$ and Bayes' theorem to calculate the posterior distribution.

$$\begin{aligned} f(\theta \mid x_1, \dots, x_n) &= \frac{f(\theta, x_1, \dots, x_n)}{f(x_1, \dots, x_n)} \\ &= \frac{f(x_1, \dots, x_n \mid \theta) p(\theta)}{\int f(x_1, \dots, x_n \mid \theta) p(\theta) d\theta} \end{aligned} \quad (1.29)$$

The Bayes estimator is simply the mean of the posterior distribution $\mathbb{E}[\theta \mid X_1, \dots, X_n]$, as it minimizes the expected square error in estimating θ . If the set of RVs take on the values $\{X_i = x_i\}$,

$$\mathbb{E}[\theta \mid X_1, \dots, X_n] = \int \theta f(\theta \mid x_1, \dots, x_n) d\theta \quad (1.30)$$

As an example, consider a set of independent normal RVs $\{X_i\}$ each with unknown mean θ and known variance σ_0^2 . Now, suppose the prior distribution of θ is itself normal with known parameters (μ, σ^2) . The Bayes estimator of θ happens to be a normal RV with

$$\mathbb{E}[\theta \mid X_1, \dots, X_n] = \left(\frac{n\sigma^2}{n\sigma^2 + \sigma_0^2} \right) \bar{X} + \left(\frac{\sigma_0^2}{n\sigma^2 + \sigma_0^2} \right) \mu \quad (1.31)$$

$$\text{Var}(\theta \mid X_1, \dots, X_n) = \frac{\sigma^2 \sigma_0^2}{n\sigma^2 + \sigma_0^2} \quad (1.32)$$

Choosing a normal RV for the prior distribution is justifiable if

- The parameter θ is likely to be close to a value μ .
- θ is just as likely to be a certain distance away from μ whether larger or smaller. This justifies a symmetric distribution about μ .
- There exists some number σ such that

$$P \{ |\theta - \mu| \leq z_{\alpha/2} \} = 1 - \alpha$$

for some conventional choice of α such as 0.05.

For an interval estimate of θ , say in the interval $[a, b]$, with probability $1 - \alpha$ use

$$\int_a^b f(\theta \mid x_1, \dots, x_n) \, d\theta = 1 - \alpha \quad (1.33)$$

Chapter 2

Hypothesis Testing

“Ch8 quote here”

In addition to the problem of estimating the unknown parameters of an underlying distribution from which samples are drawn, it is also useful to test some assertions about the sample itself.

Hypothesis : A statement about the parameters of the underlying distribution from which a sample has been drawn. Direct information about these parameters is most often unavailable. A hypothesis is *accepted* if a random sample from the population is consistent with the assertion made.

Statistical tests to *accept* a hypothesis do not conclude that it is true. They merely conclude that the sample used to test it is consistent with the hypothesis. In other words, a hypothesis, if true, might have reasonably led to the sample having the observed values.

Significance levels : Consider a distribution F_θ where θ is an unknown parameter. In order to test some hypothesis about θ , it is useful to define a *null hypothesis* H_0 . A null hypothesis is *simple* or *composite*, depending on whether or not it completely specifies the probability distribution when true.

Critical region : When a sample $\{X_i\}$ of size n is drawn to test a null hypothesis, the condition for rejecting it is for the sample to lie in some region of n -dimensional space $C \in \mathbb{R}^n$, called the critical region.

Errors in testing a hypothesis : The two possible types of errors when testing a hypothesis are the false rejection (*Type I*) and the false acceptance (*Type II*). Depending on the real-world application, one of these types of errors might be considered a much bigger issue than the other.

Level of significance : Define a scalar α , such that whenever H_0 is true, the probability

of being rejected is never greater than α . This is called the significance level of the test. Conventional values of $\alpha = 0.1, 0.05, 0.005$ are common in science literature.

$$P \{ \text{Type I error} \} \leq \alpha \quad (2.1)$$

The general procedure to develop a null hypothesis is as follows

- Determine a point estimator $d(\mathbf{X})$ of the parameter θ .
- Find the probability distribution of $d(\mathbf{X})$ provided H_0 is true.
- Determine the critical region corresponding to the required level of significance α .
- For a null hypothesis which asserts that the parameter θ lies in some region w , $H_0 : \theta \in w$ is rejected if $d(\mathbf{X})$ lies in the critical region C or equivalently is 'far away' from w .

Tests - Mean of a normal RV : A normal RV is extremely common in many real-world applications which makes the problem of testing hypotheses about its mean value very important, with strategies differing based on knowledge of its variance.

Known Variance : Using the usual sample notation, consider the null hypothesis $H_0 : \mu = \mu_0$ proposed against the alternative hypothesis $H_1 : \mu \neq \mu_0$. Here μ_0 is some assumed constant.

Rearranging the sample mean \bar{X} , which is the point estimator of μ , to yield a standard normal distribution gives the critical region and thus the hypothesis rejection condition.

$$P_{\mu_0} \left\{ \left| \bar{X} - \mu_0 \right| > c \right\} = \alpha \quad (2.2)$$

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim Z$$

$$P \{ Z > z_{\alpha/2} \} = \alpha/2$$

$$\text{reject } H_0 \text{ if } \frac{\sqrt{n}}{\sigma} \left| \bar{X} - \mu_0 \right| > z_{\alpha/2} \quad (2.3)$$

$$\text{accept } H_0 \text{ if } \frac{\sqrt{n}}{\sigma} \left| \bar{X} - \mu_0 \right| \leq z_{\alpha/2} \quad (2.4)$$

The above notation P_{μ_0} denotes the probability given that $\mu = \mu_0$.

p-value : The probability that a standard normal in absolute value $|Z|$ will exceed the quantity above on the LHS, is called the p-value of a test.

$$P \left\{ |Z| > \frac{\sqrt{n}}{\sigma} |\bar{X} - \mu_0| \right\} = p \quad (2.5)$$

$$\text{reject } H_0 \text{ if } p \leq \alpha \quad (2.6)$$

$$\text{accept } H_0 \text{ if } p > \alpha \quad (2.7)$$

Power function of a test : In order to measure the probability of a type-II error, define the *Operating Characteristic* (OC) curve $\beta(\mu)$ as,

$$\beta(\mu) = P_{\mu} \{ \text{acceptance of } H_0 \} \quad (2.8)$$

$$\begin{aligned} &= P_{\mu} \left\{ Z \in \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} \pm z_{\alpha/2} \right\} \\ &= \Phi \left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2} \right) - \Phi \left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\alpha/2} \right) \end{aligned} \quad (2.9)$$

The probability of rejection when μ is the true value is $1 - \beta(\mu)$. This is called the power-function of the test.

Using the power-function, the probability of hypothesis acceptance when the true mean is μ_1 is expressed as $\beta(\mu_1) \approx \beta$. The sample size n needed to ensure this is calculated as

$$\begin{aligned} \beta &= \Phi \left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2} \right) - \Phi \left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \right) \\ &\approx \Phi \left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2} \right) = P \{ Z > z_{\beta} \} \end{aligned}$$

The second term above is considered so small as to be close to zero and thus ignored.

$$\Phi(-z_\beta) = \Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2}\right)$$

$$n \approx \frac{\sigma^2 (z_{\alpha/2} + z_\beta)^2}{(\mu_1 - \mu_0)^2} \quad (2.10)$$

One-sided tests : Using the same reasoning as the two-sided tests above, the two possible kinds of one-sided testing problem are defined as

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0 \quad (2.11)$$

$$H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0 \quad (2.12)$$

$$\text{reject } H_0 \text{ if } \frac{\sqrt{n}}{\sigma} (\bar{X} - \mu_0) > z_\alpha \quad (2.13)$$

$$\text{accept } H_0 \text{ if } \frac{\sqrt{n}}{\sigma} (\bar{X} - \mu_0) \leq z_\alpha \quad (2.14)$$

This leads to the corresponding power function and OC curve being a decreasing function of μ . In the above one-sided tests, a reversal of the inequalities reverses the sign of z_α

$$\beta(\mu) = \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_\alpha\right) \quad (2.15)$$

$$\beta(\mu_0) = 1 - \alpha \quad (2.16)$$

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0 \quad (2.17)$$

$$H_0 : \mu \geq \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0 \quad (2.18)$$

$$\text{reject } H_0 \text{ if } \frac{\sqrt{n}}{\sigma} (\bar{X} - \mu_0) < -z_\alpha \quad (2.19)$$

$$\text{accept } H_0 \text{ if } \frac{\sqrt{n}}{\sigma} (\bar{X} - \mu_0) \geq -z_\alpha \quad (2.20)$$