A Handbook to
Conquer *Casella and Berger* Book in Ten Days

Oliver Y. Chén

Last update: June 25, 2016

Introduction

(Casella and Berger, 2002) is arguably the finest classic statistics textbook for advanced undergraduate and first-year graduate studies in statistics. Nonetheless, to thoroughly comprehend the entire book within a compact time frame while having a list of other books to read, research projects to conduct, and deadlines approaching during college or graduate school is almost impossible. Therefore, we write this handbook to help our readers to graspe the key components of the *Casella and Berger* book via an enhancement training in ten days. The handbook is a by-product from the author's preparation for an examination. In fact, without any prior supporting guidebook like this one, the author went through the entire book, and did every problem set in exercises sections, and found this approch rather time-consuming (and, alas, unwise!) Thanks to the author's torment, we are able to selectively screen the most important concepts, theories, examples, and problems sets in the book. We hope this handbook could help our readers prepare for an upper-leverl college or graduate-level statistics class and corresponding homeworks in a time-efficient manner; and to effectively attack examinations, such as qualificaiton exams, midterm and finals exams. The content, admittedly, is subjective.

Guideline

If you are comfortable with real analysis, measure theory, regression analysis, and experimental design, please choose the advanced schedule; otherwise choose the standard schedule. Once you have chosen a schedule, and have arranged time for the enhancement study, please try your best to stick to corresponding scheule. My personal experience shows that if one does not stick to the schedule, the ten-day period may extend to a month; and by the end of the third month, the task may still be incomplete. We advice you spend at least eight hours a day during this period. If you cannot finish your daily schedule within eight hours, you can go over to ten, and so on (but do not be too exhausted, because this is a ten-day effort). If by the end of the day you still cannot finish the corresponding schedule, we advice you to move on to the next-day schedule the second day. The purpose of this handbook is to grasp the key components of the *Casella and Berger* book via an enhancement training in ten days; it does not contradict with our belief that statistics learning is a long-term effort; and the best way to understand the philosophy and beauty of statistics is to read articles critically, practice problem-solving repetitively, apply knowledge in real-world scenarios, and think out of the box (e.g. derive new theory and methods). To seek a deeper and broader reach of statistics knowledge, you can read the book before (but more time-efficiently and effectively, after) the ten-day period; other books that you may find helpful in furthering your study in statistics are listed below[1].

For advanced readers, we still recommend you to at least quickly practice the recommended problem sets in these chapters listed hereinafter, just to verify if you indeed *know* these contents, or just *know of* them.

Preffixes (*)-(*****) indicate importance from low to high. For contents and theorems with more than three asterisks, we recommend our readers to know them by heart; for problem sets with more than three asterisks, we recommend our readers to do them as if you were taking a

[1] `https://www.stat.berkeley.edu/mediawiki/index.php/Recommended_Books`

Standard Schedule		Advanced Schedule	
Days	Chapters	Days	Chapters
1	1-3	1	6
2	4-5	2	6
3	6	3	7
4	6	4	7
5	7	5	8
6	7	6	8
7	8	7	9
8	8	8	10
9	10	9	10
10	9, 11-12	10	1-5, 11-12

Table 1: Schedule

test.

1 Chapter 1: Probability Theory

2 Chapter 2: Expectation

2.1 Self-check

1. What is a *parameter*? See Example 2.1.1 (p.48);
2. What is a *support (set)*? See Example 2.1.1 (p.50);
3. What are *one-to-one* and *onto*? See Example 2.1.1 (p.50);
4. What is the *kernal* of a function? See Example 2.3.8 (p.63);
5. $X \wedge Y = \min(X, Y)$ and $X \vee Y = \max(X, Y)$; hence $X + Y = X \wedge Y + X \vee Y$. See Exercise 2.15 (p.78);

2.2 Theorems

1. (*) Theorem 2.1.3 (p.51);

2. (****) Theorem 2.1.10 , and the *general inverse function* in equation (2.1.13). This is very helpful in generating a desired distribution from i.i.d. uniforms;

3. (**) *Moment generating function (mgf)*. Definition 2.3.6 (*); and Theorem 2.3.7 (p.62) ;

4. (*) Convergence of mgfs to a mgf implies convergence of cdfs. Theorem 2.3.12 (p.66);

5. (*) *Laplace* transformation. The mgf $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$ is the Laplace transformation of $f_X(x)$. Theorem 2.3.12 (p.66);

6. (*) *Poisson* approximation of *Binomial*. Example 2.3.13 (p.66);

7. (**) Lemma 2.3.14. If $\lim_{n\to\infty} a_n = a$, then $\lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$;

8. (**) *Leibnitz's Rule*. If $f(x,\theta), a(\theta)$ and $b(\theta)$ are differentiable w.r.t. θ, then

   $$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx = f(b(\theta),\theta)\frac{d}{d\theta}b(\theta) - f(a(\theta),\theta)\frac{d}{d\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta}f(x,\theta)dx.$$

   See Theorem 2.4.1 (p.69)

9. (****) Lebesgue's *Dominated Convergence Theorem*. See Theorem 2.4.2 (p.69);

10. (***) *Lipschitz Continuous*. It imposes smoothness on a function by bounding its first derivative by a function with finite integral. It leads to interchangeability of intergration and differentiaiton. See Theorem 2.4.3 (p.70), Corollary 2.4.4, and Examples 2.4.5 -2.4.6;

2.3 Recommended Exercises

1. (*) Ex 2.1 (c). Hint: Theorem 2.1.5 (p.51);

2. (**) Ex 2.3. Hint: For a continuous r.v., start working with $F_Y(y) = \mathbb{P}(Y \leq y)$; whereas for a discrete r.v., start working with $\mathbb{P}(Y = y)$;

3. (*) Ex 2.6 (a);

3 Chapter 3: Families of Distributions

4 Chapter 4: Multiple Random Variables

4.1 Self-check

1. What is an *n-dimensional random vector*? It is a function $f : \mathbb{S} \mapsto \mathbb{R}^n$, where $\mathbb{S}$ is the sample space. For example, consider $n = 2$. Take $(2, 3) \in \mathbb{S}$. An f can be $(2, 3) \xrightarrow{f} (X, Y)$, where $X = 2 + 3$ an $Y = |2 - 3|$. See Definition 4.1.1 (p. 139);

2. What is the *joint probability mass function (pmf)* - for discrete r.v.'s? It is a function $f : \mathbb{R}^n \mapsto \mathbb{R}^1$, where $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$. The subscript $X_1, \ldots, X_n$ emphasizes that f is the joint pmf for vector $(X_1, \ldots, X_n)$, instead of some other vector. See Definition 4.1.3 (p. 140);

3. What is the *marginal probability mass function (pmf)*? See Definition 4.1.6 (p. 143);

4. What is the *joint* and *marginal probability density function (pdf)* - for continuous r.v.'s? See Definition 4.1.10 (p. 144);

5. *Conditional expectation*. $\mathbb{E}(Y|X = x) = \int_{y \in \mathcal{Y}} y f(y|x) dy$ Example 4.2.4 (p. 151);

6. *Conditional variance* $Var(Y|X = x) = \mathbb{E}(Y^2|x) - \mathbb{E}^2(Y|x)$ Example 4.2.4 (p. 151);

7. The family of conditional probability distributions: $Y|X$ or $Y|X \sim \mathcal{P}(X, \theta)$ v.s. *conditional* pdf pf Y given $X = x$, where x (yes, small x) acts as a local parameter. Comments on p. 151;

8. *Borel Paradox*. It deals with conditional probability conditioning on measure zero sets. See 4.9.3 (p. 204);

9. *Bivariate transformation*. Discrete case (p.157); continuous case (p. 158), the *Jacobian transformation* (p. 158), multivariate case (Example 4.6.13, p. 185). Read carefully the definition of sets $\mathcal{A}$ and $\mathcal{B}$ on p.157;

10. *Hierarchical models*. See an example of Binomial-Poisson hierarchy in Example 4.4.1 (p.163);

11. *Mixture distribution*. A r.v. X has a mixture distribution if the distribution of X depends on a quantity that also has a distribution. Definition 4.4.4 (p.165);

12. *Covariance* and *correlation*. Definition 4.5.1-2; Theorem 4.5.3 (p.169-170);

13. *Bivariate normal*. Definition 4.5.10 (p.175);

14. *Convex* and *concave* functions. $g(x)$ is convex if $g(\lambda x+(1-\lambda)y) \leq \lambda g(x)+(1-\lambda)g(y)$, $\forall x, y$ and $0 < \lambda < 1$. $g(x)$ is convace if $-g(x)$ is convex. Definition 4.7.6 (p. 189)

4.2 Theorems

1. (*) Checking independence by using cross-product. See Lemma 4.2.7 (p.153);

2. (*) $X \sim$ Poisson (θ) and $Y \sim$ Poisson (λ), X and Y are independent, then $X + Y \sim$ Poisson $(\theta + \lambda)$;

3. (***) *Conditional Expectation (iterative expectation)*. $\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|Y))$: carefully think about the subscript. Rigorously, it should be written as: $\mathbb{E}_X X = \mathbb{E}_Y(\mathbb{E}_{X|Y}(X|Y))$ because $\mathbb{E}(X|Y)$ is a r.v. (random in Y), $\mathbb{E}(X|Y = y) = \int x f_{X|Y}(x|Y = y)dx$ is a con-

stant, and $\mathbb{E}_Y \mathbb{E}(X|Y=y) = \int \left\{ \int x f_{X|Y}(x|y) dx \right\} f_Y(y) dy$. See Theorem 4.4.3 (p.164). A more rigorous definition is given w.r.t. a sigma-field, see Section 5.1 in (Durrett, 2010);

4. (***) *Conditional Variance (iterative variance)*. $VarX = \mathbb{E}(Var(X|Y)) + Var(\mathbb{E}(X|Y))$. See Theorem 4.4.7 (p.167);

5. (*) $Var(aX+bY) = a^2 VarX + b^2 VarY + 2abCov(X,Y)$; See Theorem 4.5.6 (p.171);

6. (*) Multinomial distribution and multinomial theory. See Definition 4.6.2 and Theorem 4.6.4 (p.181). Note that the marginal distribution X_n of a multinomial distribution $(X_1, \dots, X_n)$ with $(p_1, \dots, p_n)$ and $\sum_{i=1}^n x_i = m$ is binomial (m, p_n); and what is the distribution of $X_i|X_j = x_j$ and what is $Cov(X_i, X_j)$? See Ex 4.39 (p.198);

7. (*****) Inequalities. Section 4.7 (p.186).

   - *Hölder's Inequality*: if $\frac{1}{p} + \frac{1}{q} = 1$, then $|\mathbb{E}XY| \leq \mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{1/p} (\mathbb{E}|X|^q)^{1/q}$;

   - *Cauchy-Schwarz Inequality* (Sepcial case of Hölder's Inequality when $p = q = 2$): $|\mathbb{E}XY| \leq \mathbb{E}|XY| \leq (\mathbb{E}|X|^2)^{1/2} (\mathbb{E}|X|^2)^{1/2}$;

   - *Covariance Inequality* (Special case of Cauchy-Schwarz Inequality): $(Cov(X,Y))^2 \leq \sigma_X^2 \sigma_Y^2$. eXAMPLE 4.7.4 (P.188);

   - *Liapounov's Inequality* (Sepcial case of Hölder's Inequality): $\{\mathbb{E}|X|^r\}^{1/r} \leq \{\mathbb{E}|X|^s\}^{1/s}$, for $1 < r < s < \infty$;

   - *Minkowski's Inequality*: For $1 \leq p > \infty$, $[\mathbb{E}|X+Y|^p]^{1/p} \leq [\mathbb{E}|X|^p]^{1/p} + [\mathbb{E}|Y|^p]^{1/p}$;

   - *Jensen's Inequality*: For $g(x)$ convex, $\mathbb{E}g(X) \geq g(\mathbb{E}X)$. Do: show Harmonic mean $\leq$ Geometric mean $\leq$ Arithmetic mean (Example 4.7.8, p.191).

4.3 Recommended Exercises

1. (****) Ex 4.39 (p.198). Hint: Apply iterative expectation and Theorem 4.6.4 (p.182);

5 Chapter 5: Random Sample

5.1 Self-check

1. What does *iid random variable* really mean: *independent and identically distributed random variables with pdf or pmf* $f(x)$. Definition 5.1.1 (P. 207);

2. What is a *statistic* and what is a *sampling distribution*? Definition 5.2.1 (P. 211). Note a statistic cannot be a function of a parameter;

3. *Student's t-distribution*. Definition 5.3.4 (P. 223);

4. *Snedecor's F-distribution*. Definition 5.3.6 (P. 224);

5. *Order statistics*. Section 5.4 (p.227);

6. How to generate random sample from a given distribution, using WLLN? See Examples 5.6.1 and 5.6.2 (p. 246-247);

7. How to generate random sample from a *uniform distribution*? See Examples 5.6.3 (p.247) for continuous, and Examples 5.6.4 (p.247) for discrete case;

8. How to generate random sample using *Accept-reject Algorithm*? See Theorem 5.6.8 (p.253). Comments: essentionally, if we want to generate Y that follows $f_Y(y)$. (1), independently generate $U \sim Uniform[0,1]$, and V from $f_V(v)$ (could be another uniform!) - which we know how to generate; (2) plug V into $\mathbf{1}/M f_Y(V)/f_V(V)$, where $M = \sup_y f_Y(y)/f_V(y)$, the expected number of trials needed - we do not know how to generate from f_Y but we know the expression of f_Y and can evaluate $f_Y(V = v)$; (3) if $U < \mathbf{1}/M f_Y(V)/f_V(V)$, we accept V as a desired candidate from f_Y, and let $Y = V$; and (4) repeat;

9. How to generate random sample using *Metroplis Algorithm*? See p.254. Also read MCMC, Miscellanea 5.8.5 (p.269).

5.2 Theorems

1. (***) *Exponential family*. See Theorem 5.2.11 (p.217);

2. (*) Lemma 5.3.3 (p. 220);

3. (*) Relationship between t- and F-distributions. Theorem 5.3.8 (p.225);

4. (****) Convergence in *probability* (Definition 5.5.1, p.232), *almost-sure* convergence (Definition 5.5.6, Examples 5.5.7-8 p.234), convergence in *distribution* (Defintion 5.5.10, p. 235), and a *subsequnce* of a sequence that converges in *probability* converges *almost surely* (Comment, p.235);

5. (****) Application of convergence in *probability*: *Weak Law of Large Numbers (WLLN* (Theorem 5.5.2, p. 232), *continous mapping* (Theorem 5.5.4, p. 233);

6. (****) Application of *almost sure* convergence: *Strong Law of Large Numbers (SLLN* (Theorem 5.5.9, p. 235);

7. (****) *Central Limit Thorem (CLT)*: weak version (assume existence of mgf): Theorem 5.5.14 (p.236); strong version (only assume finite variance): Theorem 5.1.15 (p.238);

8. (***) Proving tools. *Slutsky's Theorem* (Theorem 5.5.17, p.239), *Taylor expansion* (Definition 5.5.20 and Theorem 5.5.21, p.241), and *Delta Method* (Theorem 5.5.24, p.243, second order, Theorem 5.5.26, p.244, Multivariate, Theorem 5.5.28, p.245);

9. (**) Application of *Taylor* expansion: approximate of general mean and variance. (5.5.8) and (5.5.9), p.241-242.

6 Chapter 6: Data Reduction

6.1 Self-check

1. What are three *principles* of data reducton? Sufficiency, Likelihood, and Equivariance;

2. (***) A *Sufficient statistic* for a parameter θ is a statistic, $T(\mathbf{X})$, such that, $X|T(\mathbf{X})$ does not depend on θ. See definition 6.2.1 (p. 274); Finding Sufficient statistics: *Factorization Theorem*: Theorem 6.2.6 (p.276); for *exponential family* Theorem 6.2.10 (p.276); *minimal sufficient statistics*: Theorem 6.2.11 (p.280) and Theorem 6.2.13 (p.281)

3. (*) An *ancillary statistic*, $S(\mathbf{X})$, is a statistic whose distribution does not depend on the parameter θ;

4. (**) *Complete statistics*. See Definition 6.2.21 (p.285);

5. (*) A *complete* and *minimal sufficient statistic* is independent of every *ancillary statistic*? See Basu's Theorem (p.287);

6. (*) Any *complete statistic* is also a *minimal sufficient statistic*, provided a *minimal sufficient statistic* exisits? See Theorem 6.2.28 (p.289);

7. (**) What is the difference between $f(\mathbf{x}|\theta)$ (a pdf or pmf) and $L(\theta|\mathbf{x})$ (the likelihood function)? see Definition 6.3.1 (p.290);

8. (**) *Likelihood principle* about θ. What is the difference between "plausible" and "probable"? See p.291;

9. (*) *Fiducial inference*. See p.291;

10. (*) *Evidence* and *evidence function* $Ev(E, \mathbf{x})$. Suppose we have an experiment $E = (\mathbf{X}, \theta, \{f(\mathbf{x}|\theta\})$. Knowing how the experiment is performed, we will observe $\mathbf{X} = \mathbf{x}$ and

wish to draw conclusion or *inference* about θ. This *inference* we denote by $Ev(E, \mathbf{x})$, the *evidence about* θ *arising from* E and $\mathbf{x}$. E.g.: $Ev(E, \mathbf{x}) = (\bar{x}, \sigma/\sqrt{n})$, where $\bar{x}$ depends on $\mathbf{x}$ and $\sigma/\sqrt{n}$ depends on the knowledge of E See Example 6.3.4;

11. (*) *Formal sufficient principle*, *conditionality principle*, *formal likelihood principle*, *Birnbaum's Theorem*. See Section 6.3.2. p.292;

12. (*) The *equivariance principle*. See Section 6.4 p. 296.

7 Chapter 7: Point Estimation

7.1 Self-check

1. Why do we need conduct *point estimation*? Conducting point estimation is to find a good estimator of the point θ (a *point estimate* is any function $W(\mathbf{X})$ of a sample), which yields knowledge of the entire population, because we sample from a population described by a pdf or pmf $f(x|\theta)$; moreover, θ may have a meaning physical intepretation (e.g. population mean). see p.311;

2. What is the difference between an *estimator* ($W(\mathbf{X})$) and an *estimate* ($W(\mathbf{X} = \mathbf{x})$)? See p.312;

3. How to find a point estimate of θ, and $\tau(\theta)$, a function of θ? How to evaluate the estimate of θ?

4. What are the four common methods for finding estimators? *Method of Eoments* (MOM, a.k.a., *moment matching*), *Maximal Likelihood Estimation (MLE)*, *Bayes estimation*, *Expectation-Maximazation (EM)* algorithm; See Section 7.2 (p.312);

5. What are the four common methods for evaluating estimators? They are, via evaluating the *Mean Squared Errors (MSE)*, via choosing an unbiased estimator with uniformly

smallest variance (what is a *UMVUE*, see Definition 7.3.7, p.334, how to find a UMVUE - via Cramér-Rao bound, See Theorem 7.3.9), via finding a uniformly better unbiased estimator (using both sufficiency and unbiasedness, how? via *Rao-Blackwellization*, see Theorem 7.3.17, p.342), and via *loss function optimality*;

6. What is the *induced likelihood function*? See (7.2.5) on p.320;

7. *Conjugate family*. Definition 7.2.15 (p.325);

8. What is *information number* or *Fisher information*?

9. Cramér-Rao is not sharp; namely, the bound may be *strictly smaller than* any unbiased estimator (p.340). See also Corollary 7.3.15 regarding attainment;

7.2 Theorems

1. (**) *Invariate property of MLE*. Theorem 7.2.10 (p.320);

2. (****) (Cramér-Rao Inequality, Information Inequality). Let $\mathbf{X}_i$ be a *sample* with pdf $f(\mathbf{x}|\theta)$, and let $W(\mathbf{X})$ be any estimator satisfying *exchange of deferentiation and intergration* and with *finite variance*, then

$$\text{Var}_\theta(W(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta W(\mathbf{X})\right)^2}{I(\theta)} = \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta W(\mathbf{X})\right)^2}{\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right)^2} \overset{\text{Lemma 7.3.11}}{=} \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta W(\mathbf{X})\right)^2}{-\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{X}|\theta)\right)^2}.$$

See Theorem 7.3.9 (p.335), and Corollary 7.3.10 for iid case (p.337).

3. Any unbiased estimator W for $\tau(\theta)$ attaining the Cramér-Rao lower bound is a best unbiased estimator of $\tau(\theta)$. See Example 7.3.12;

4. (**) *Rao-Blackwell Theorem*: to obtain an updated unbiased estimator that has smaller variance than the original unbiased estimator, by taking the conditional expectation of the *sufficient*

statistic. Specifically, if W is unbised for $\tau(\theta)$, and T is a sufficient statistic for $\underline{\theta}$, than $\mathbb{E}(W|T)$ has smaller variance than original W. See Theorem 7.3.17 (p.342);

5. Relationship between *Completeness* and *best unbiasedness*. See Theorem 7.3.23 and Example 7.3.24 (p.347);

6. *Loss function*, *risk function*, and *Bayes risk*: consider an *estimator* $\delta(x)$ of θ, the *risk function* is defined (see p.349) as

$$R(\theta,\delta) = \underbrace{\mathbb{E}_\theta \qquad \overbrace{L(\theta,\delta(\mathbf{X}))}^{\text{pre-specified distance between } \theta \text{ and } \delta(\mathbf{X})}}_{\text{average pre-specified distance between } \theta \text{ and } \delta(\mathbf{X})},$$

and the *Bayes risk* is defined

$$\int_\Theta R(\theta,\delta)\pi(\theta)d\theta,$$

and the *Bayes rule w.r.t. a prior* π is defined (see p.352) as

$$\hat{\delta}^\pi = \arg\min_\delta \int_\Theta R(\theta,\delta)\pi(\theta)d\theta.$$

7.3 Recommended Exercises

1. (**) *Moment matching* Example 7.2.1, p.314. The key is to understand the essense of MOM:

$$m_1 := \boxed{\frac{\sum_i X_i}{n}} \overset{\text{matching}}{=} \mu_1 := \mathbb{E}X \overset{\text{e.g. Normal case}}{=} \boxed{\mu};$$

$$m_2 := \boxed{\frac{\sum_i X_i^2}{n}} \overset{\text{matching}}{=} \mu_2 := \mathbb{E}X \overset{\text{e.g. Normal case}}{=} \boxed{\mu^2+\sigma^2};$$

2. Comparison between $\hat{\sigma}^{2MLE} = \frac{\sum_i (X_i-\bar{X})^2}{n}$ and $S^2 = \frac{\sum_i (X_i-\bar{X})^2}{n-1}$. Show on averge, $\hat{\sigma}^{2MLE}$ will be closer to σ^2, but biased. Calculate $\mathbb{E}\hat{\sigma}^{2MLE}$, $\mathbb{E}S^2$, $Var\left(\hat{\sigma}^{2MLE}\right)$, $VarS^2$, $\mathbb{E}(\hat{\sigma}^{2MLE}-\sigma^2)^2$, and $\mathbb{E}(S^2-\sigma^2)^2$. See Examples 7.3.3 and 7.3.4 (p.331);

8 Chapter 8: Hypothesis Testing

8.1 Self-check

1. What is a *hypothesis*? A hypothesis is a statement about a population parameter, based upon a *sample* of the population. See p.373;

2. A hypothesis test is specified in terms of a *test statistic* $W(\mathbf{X})$, a function of the *sample*. See p.374;

3. What are the three common methods of finding test procedures? *Likelihood Ratio Test (LRT)*; *Bayesian Test*, and *Union-Intersection and Union-Intersection Test*;

4. *Likelihood Ratio Test (LRT)*: $\lambda(\mathbf{x}) = \frac{\sup_{\theta\in\Theta_0} L(\theta|\mathbf{x})}{\sup_{\theta\in\Theta} L(\theta|\mathbf{x})} = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$, where $\hat{\theta}_0$ is the MLE over Θ_0 and $\hat{\theta}$ is the globle MLE over Θ; then the *reject region* is $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$. $\lambda(\mathbf{x})$ depends on $\mathbf{x}$ only through $T(\mathbf{x})$, the sufficient statistic;

5. *Bayesian Test*: Classical statistican consider θ to be a fixed number; and hence a hypothesis is either *true* or *false*. For example, if $\theta \in \Theta_0, \mathbb{P}(H_0 \text{is ture}|\mathbf{x}) = 1$. For *Bayesians*, accept H_0 if $\mathbb{P}(\theta \in \Theta_0|\mathbf{X}) \geq \mathbb{P}(\theta \in \Theta_0^c|\mathbf{X})$, the test statistic (a function of the sample) is $\mathbb{P}(\theta \in \Theta_0|\mathbf{X})$, and the rejection region is $\{\mathbf{x} : \mathbb{P}(\theta \in \Theta_0|\mathbf{x}) > \frac{1}{2}\}$. To guard against falsely rejecting H_0, a Bayesian rejects H_0 only if $\mathbb{P}(\theta \in \Theta_0|\mathbf{X}) > c$, where c takes, say, 0.99;

6. *Union-Intersection and Union-Intersection Test*. See Section 8.2.3 (p.380-381);

7. How to evaluate tests? Type I and II errors, and power function;

8. *Type I and II errors*: Define R as the rejection region. Then the probability of a Type I error is $\mathbb{P}(\mathbb{X} \in \mathbb{R}|\theta)$, for $\theta \in \Theta_0$; and the probability of a Type II error is $\mathbb{P}(\mathbb{X} \in \mathbb{R}|\theta) = 1 - \mathbb{P}(\mathbb{X} \in \mathbb{R}|\theta)$, for $\theta \in \Theta_0^c$;

9. *Power function*: $\beta(\theta) = \mathbb{P}(\mathbf{X} \in R|\theta)$. It is a function of θ; there is NO indication whatsoever whether $\theta \in \Theta_0$ or Θ_0^c;

10. *Size*, and *size* α *test*: a size α test is one such that $\sup_{\theta\in\Theta_0} \beta(\theta) = \alpha$, for $0 \leq \alpha \leq 1$; if $\sup_{\theta\in\Theta_0} \beta(\theta) \leq \alpha$, it is a level α test. See Definitions 8.3.5-8.3.6 (p.385);

8.2 Theorems

1. Uniformly most powerful (UMP) test. Definition 8.3.11 (p.388);

2. Neyman-Pearson Lemma. Theorem 8.3.12 (p.388), Corollary 8.3.13 (p.389);

3. Karlin-Rubin Theorem 8.3.17 (p.391);

8.3 Recommended Exercises

1. 8.1;

2. 8.3;

3. 8.5 (a);

4. 8.6 (c);

5. 8.8;

References

Casella, G. and R. L. Berger (2002). *Statistical inference*, Volume 2. Duxbury Pacific Grove, CA.

Durrett, R. (2010). *Probability: theory and examples*. Cambridge university press.