Complementary Material to A Concise Introduction to Mathematical Statistics

Dragi Anevski

December 9, 2019

1 Preface

The material in this text is intended to serve as clarification to some parts of the material in my book "A Concise Introduction to Mathematical Statistics", first published by Studentlitteratur in 2017. These notes are the result of my experience in teaching from my book at the first course in mathematical statistics at the Faculty of Science at Lund University.

The main take from the feedback I have received from the students, and with a for me pleasant surprise, is that the mathematical exposition in the book can be presented (even more) strictly. In fact, at several occasions during the teaching I have been posed quite technical questions on concepts, and after the course I have also received comments that a certain topic should be further explained, and that certain concepts have been understood by the students to be loosely explained. This is on the one hand somewhat surprising to me, and in fact a common feedback from colleagues is that the book is too theoretical or too difficult for a first course. On the other hand, this has not been my belief, nor my experience from teaching on the material over a number of years, and I am therefore both glad and relieved that my students not only accept the level of mathematical stringency in the book, but in fact want to see more of mathematical stringency.

One main concept in the book that I expand on here is the Riemann-Stieltjes integral, in giving some refinements of the Riemann-Stieltjes integral. In fact I have in the book, on purpose, been slightly opaque in order to not cloud the exposition and the main results by "un-necessary technicalities". In particular I have been on purpose vague about integration of functions that are at most piecewise continuous.

In conclusion I am happy to be able to add these comments on extensions and clarifications, to the benefit for the interested students, and also as extra help for the teacher that is using my book. In fact I have noticed that some parts in my book have been able to trip up also colleagues; for this I apologise and I hope these further notes may help to alleviate some obstacles in the understanding and clarification of the material.

2 Complements to the Riemann-Stieltjes integral on $\mathbb{R}$

2.1 Integration of piecewice continuous functions

Recall the definition of the Riemann-Stieltjes integral.

Let $I \subseteq \mathbb{R}$ be a finite interval $I = [a, b]$, and F a monotone (increasing) real-valued transformation of $I \subseteq \mathbb{R}$, so

$$F : I \mapsto F(I).$$

Assume that h is a real-valued step-function on the interval I,

$$h(t) \quad = \quad \sum_{i=1}^{n} c_i 1\{t \in I_i\},$$

with some constants $c_1, \ldots, c_n$, and where $I = \cup_{i=1}^{n} I_i$ is a partition of I, with I_j intervals. Define the Riemann-Stieltjes integral of h as

$$\begin{aligned} \int h(u)\, dF(u) \quad &= \quad \sum_{i=1}^{n} c_i |F(I_i)| \qquad (1) \\ &= \quad \sum_{i=1}^{n} c_i (F(b_i) - F(a_i)). \end{aligned}$$

We next define the Riemann-Stieltjes integral.

Definition 1. *Let F be an increasing function defined on the finite interval I, and let g be a function defined on I. Then g is called Riemann-Stieltjes integrable, w.r.t. F, if for every $\epsilon > 0$, there are step functions h_1, h_2 such that $h_1 \leq g \leq h_2$ and*

$$\int_I h_2(u)\, dF(u) - \int_I h_1(u)\, dF(u) \quad < \quad \epsilon.$$

If g is Riemann-Stieltjes integrable, we define the Riemann-Stieltjes integral of g as

$$\int_I g(u)\, dF(u) \quad = \quad \sup\{\int_I h(u)\, dF(u) : h \leq g, h \text{ step function.}\}$$

Recall also the following theorem, for which proof is given in the book.

Theorem 1. *If g is continuous, and F is increasing and bounded on the finite interval I, then g is Riemann-Stieltjes integrable on I.*

Recall also that the Riemann-Stieltjes integral can be obtained as a limit of Riemann-Stieltjes sums, and that we have proved the following statement: If g is continuous and F increasing and bounded on the interval I, then

$$\int_I g(x)\,dFx \quad = \quad \lim_{\max_{1\leq i\leq n}|x_i - x_{i-1}|\to 0} \sum_{i=1}^{n} g(\xi_i)(F(x_i) - F(x_{i-1})), \qquad (2)$$

where $\min I = x_0 < x_1 < \ldots < x_n = \max I$ are partitions of I, and ξ_i are arbitrary points in $[x_{i-1}, x_i)$.

I would like to make the following note about integration over open intervals, that will be of use in the sequel: We have treated an interval I as a closed interval, and in fact the above theorem says that if g is continuous and F increasing and bounded on the compact (i.e. finite/bounded and closed) interval then g is Riemann-Steiltjes integrable. In the sequel we will need to integrate also over bounded and *open* intervals. The definition of the integral will then be analogous to Definition 1, with the difference that the partition of the interval $(a, b) = I = \cup_{i=1} I_i$, in the definition of the integral of step functions (1), will have that the two sets I_1 and I_n are open on the left and on the right, respectively, i.e. $I_1 = (a, \tilde{a}]$ and $I_n = [\tilde{b}, b)$ for some $a < \tilde{a} < \tilde{n} < b$. We can still define $\int_i gdF$ analogously to as in Definition 1. Theorem 1 is however then not true; we then need to demand that g is continuous and *bounded* on the bounded and open interval for $\int_I gdF$ to exist. The proof is a simple modification of the proof of Theorem 1, as given in the book[1]. This extension of Theorem 1 will be of interest in Theorem 2 below.

Now, one may note that we actually use the Riemann-Stieltjes integral on some functions g that *have discontinuities.* For instance, if F is the distribution function for a random variable X, and we want to calculate e.g. $P(X \in [2, 3))$, then we claim that

$$\begin{aligned} P(X \in [2,3)) &= \int_{X\in[2,3)} dF(x) \\ &= \int g(x)dF(x), \end{aligned}$$

with g the function $g(x) = 1\{2 \leq x < 3\}$. Obviously g is not continuous. We may have even more seemingly worrysome examples, for instance we claim that, for the same random variable X above,

$$\begin{aligned} P(X = 2) &= P(X \in \{2\}) \\ &= \int_{\{2\}} dF(x) \\ &= \int g(x)dF(x), \end{aligned}$$

[1] This result is analogous to results for the ordinary Rimann integral of improper integrals. For instance the integral $\int_{[0,1]} x^{-1}dx$ does not exist due to the fact that x^{-1} is unbounded at 0, and neither does the integral $\int_{(0,1)} x^{-1}dx$. For this reason we need to demand that the function g is bounded on the interval in question.

for the "function" $g(x) = 1\{x = 2\}$, which strictly speaking is the Dirac measure, or generalised function, $\delta_2(x)$. Now, again g is *not* continuous and it is not even a proper function, it is merely a distribution, in the Schwarz generalised function sense.

Furthermore, the function F may be discontinuous, and the set of discontinuity points D_F of F and the set of discontinuity points D_g of g may overlap, so that one may have $D_F \cap D_g \neq \emptyset$. For example, if F is the distribution function of a Bernoulli random variable, with parameter p, then one may ask how to define e.g. the two expressions

$$\int_{(1/2,1]} dF(x) \text{ and } \int_{\{1\}} dF(x).$$

Clearly we would like the first integral to denote $P(X \in (1/2, 1])$ which should be equal to $P(X = 1) = p$, and we would like the second integral to denote $P(X = 1) = p$. Recall that have made the claim that for any "nice" set A, we can obtain $P(X \in A)$ as

$$\begin{aligned} P(X \in A) &= \int_A dF(X) \\ &= \int 1_A(x) dF(x), \end{aligned}$$

and of course the sets in the above examples, $A = (1/2, 1]$ and $A = \{1\}$, should definitely be allowed to be labeled as nice. The function 1_A, however, is not continuous and is possibly not even a proper function.

The solution to the above problem is straightforward. The problem seems to be, as we have already indicated, that the discontinuity points of g and of F may overlap. We solve the problem in a fashion that is appropriate for the problems that we encounter in probability theory. Recall that then F is a distribution function, and therefore right continuous. If x_0 is a discontinuity point of F, so if $x_0 \in D_F$, then F is right continuous (and not left continuous) at x_0. If x_0 is also a discontinuity point of g, so $x_0 \in D_g$, then in the definition of the Riemann-Stieltjes integral, the point x_0 is a problem, since in any partition $I = \cup_{i=1}^n I_i$, if $I_k \ni x_0$, it is not clear how to chose c_k in the term

$$c_k |F(I_k|,$$

in (10), or, equivalently, how to chose ξ_k in the term $g(\xi_k)(F(x_k) - F(x_{k-1}))$ in the Riemann-Stieltjes approximating sum (2); choosing c_k, or ξ_k, to the left or to the right of the discontinuity point x_0 will give different values for the expressions.

To extend, or modify, the definition of the Riemann-Stieltjes integral so that it will be applicable to possibly discontinuous functions, we first note that of course we will not be able to make an extension that will manage all functions, only those that are nice enough for the applications we have in mind. Therefore let us first decide that we want to extend the definition to all piecewise continuous and bounded functions g, and let C_{pc} denote those functions. Then if

D_g is the set of discontinuity points of g, this set may be infinite but must be countable, and furthermore between each subsequent two discontinuity points there must be a positive distance. Therefore we can make a partition

$$(D_g)^c = \cup_{i=1}^m J_i \tag{3}$$

of the set $(D_g)^c$ of continuity points of g, where each open interval J_i has boundary points in D_g, and with $m \leq \infty$ (so that $m = \infty$ is allowed).

Now let F be increasing and $g \in C_{pc}$ be arbitrary but fixed. Then if $D_F \cap D_g = \emptyset$, we will see that then the Riemann-Stieltjes definition above is actually valid, since then we can make the partition (3) and define

$$\int_{(D_g)^c} g(x)dF(x) \quad = \quad \sum_{i=1}^m \int_{J_i} g(x)dF(x).$$

Note here that the intervals J_i are open so that the extension of Theorem 1, in which one needs that g is bounded in addition to continuous, in in force. In the above sum, g is continuous and bounded on each J_i, and therefore every term is well-defined. On the remaining points, i.e. for any $x_0 \in D_g$, F is continuous, and therefore any approximating upper and lower step function $h_1 \leq g \leq h_2$ will have a difference in the contribution to the total integral that is

$$(h_2(\xi) - h_1(\xi))|F(I)|, \tag{4}$$

where ξ is some point in an open interval I containg x_0. The term (4) can be made arbitrarily small, smaller than ϵ say, since g is bounded and by letting I decrease towards the one-point set $\{x_0\}$ then since F is continuous at x_0, we will have $|F(I)| \downarrow 0$. Since there are countably many points in D_g, we may enumerate them, arbitrarily, as $D_g = \{x_1, x_2, \ldots\}$ and for each discontinuity point x_i in D_g, we may choose an interval I_i that contains x_i and that is small enough and some $\xi_i \in I_i$, and step functions $h_{1,i}$ and $h_{2,i}$, such that

$$(h_{2,i}(\xi_i) - h_{1,i}(\xi_i))|F(I_i)| \quad < \quad \epsilon/2^i. \tag{5}$$

Then the total contribution to the difference between the upper and lower integral over all discontinuity points, with the for arguments sake assumption that there are infinitely many points in D_g so that $m = \infty$, is

$$\begin{aligned} \sum_{i=1}^{\infty}(h_{2,i}(\xi_i) - h_{1,i}(\xi_i))|F(I_i)| \quad &< \quad \sum_{i=1}^{\infty} \epsilon/2^i \\ &= \quad 2\epsilon, \end{aligned}$$

and thus the total contribution to the difference between the upper and lower integral over all discontinuity points can be made arbitrarily small. Therefore, formally,

$$\int_{D_g} g(x)dF(x) \quad = \quad 0.$$

Since we can write, formally,

$$\int g(x)dF(x) \quad = \quad \int_{D_g} g(x)dF(x) + \int_{(D_g)^c} g(x)dF(x),$$

we may state the conclusion of the reasonings as a theorem.

Theorem 2. *Suppose that g is bounded and piecewise continuous with a countable number of discontinuity points D_g and such that $D_F \cap D_g = \emptyset$. Then g is Riemann-Stieltjes integrable and*

$$\begin{aligned} \int g(x)dF(x) &= \int_{(D_g)^c} g(x)dF(x) \\ &= \sum_{i=1}^{m} \int_{J_i} g(x)dF(x), \end{aligned}$$

with $(D_g)^c = \cup_{i=1}^m J_i$ the partition defined in (3).

By applying Riemann-Stieltjes sum approximations to each term in the above sum we obtain a generalisation of the result about Riemann-Stieltjes sum approximations to piecewise continuous functions.

Theorem 3. *Assume g is bounded and piecewise continuous with D_g the set of discontinuity points of g, F increasing and bounded on the interval I, and $D_g \cap D_F = \emptyset$. Then*

$$\int_I g(x)\, dFx \quad = \quad \lim_{\max_{1\le i\le n}|J_i|\to 0} \sum_{i=1}^{n} g(\xi_i)|F(J_i)|,$$

where

$$\cup_{i=1}^n I_i \quad = \quad (D_g)^c$$

are partitions of $(D_g)^c$, with $I_i \subseteq J_k$ for each i and some k, where J_k are the partition sets in (3), and ξ_i are arbitrary points in I_i.

We are left with the remaining problem, namely how to define the integral for function $g \in C_{pc}$ such that $D_g \cap D_F \neq \emptyset$. This is easy, and merely convention. In fact suppose that $x_0 \in D_F \cap D_g$. Then, for any $\delta > 0$ that is small enough, the open interval $(x_0 - \delta, x_0)$ will be included in one of the open sets J_k in (3). Therefore g is continuous in the open interval $(x_0 - \delta, x_0)$, if $\delta > 0$ is small enough, and discontinuous at x_0. Then for such a δ, if g is *left* continuous at x_0, we define, by convention,

$$\int_{(x_0-\delta,x_0]} g(x)dF(x) \quad = \quad \lim_{\epsilon\downarrow 0} \int_{(x_0-\delta,x_0+\epsilon)} \tilde{g}(x)dF(x), \qquad (6)$$

where $\tilde{g}(x)$ is the constant extrapolation of g to the *right* of x_0, namely,

$$\tilde{g}(x) = g(x)1\{x \le x_0\} + g(x_0)1\{x > x_0\}.$$

Example 1. *Suppose that $g(x) = 1\{x \leq 0\}$, and F is the distribution function of a Bernoulli distributed random variable with parameter p. Then F has jump discontinuities at 0 and 1, and g is left continuous at the (sole) point $0 \in D_F \cap D_g$. The constant extrapolation $\tilde{g}$ is the identity function*

$$\tilde{g}(x) \equiv 1.$$

The above definition then gives us, for any finite $C < 0$,

$$\begin{aligned}
\int_{(C,0]} g(x)dF(x) &= \lim_{\epsilon \downarrow 0} \int_{(C,\epsilon)} \tilde{g}(x)dF(x) \\
&= \lim_{\epsilon \downarrow 0} F(\epsilon) - F(C) \\
&= 1 - p.
\end{aligned}$$

Using probabilistic notation, the right hand side of the above is $P(X = 0)$ while the left hand side is $\int_{(C,0]} dF(x)$. Thus we have obtained, with the above definition,

$$\begin{aligned}
P(X = 0) &= \int_{(C,0]} dF(x) \\
&= P(X \in (C, 0]).
\end{aligned}$$

The example illustrates why we have chosen to make the definition in (6). It is in order to be able to state that

$$P(X \in A) = \int_A dF(x),$$

for all reasonable sets.

Thus if x_0 is a point in $D_F \cap D_g$ and g is left-continuous at x_0, then we define the contribution of the integral over $(x_0 - \delta, x_0]$, for $\delta > 0$ small, to the total integral by (6). If instead g is *right* continuous at x_0 then we define, by convention,

$$\int_{(x_0-\delta,x_0]} g(x)dF(x) = \int_{(x_0-\delta,x_0)} g(x)dF(x). \tag{7}$$

Example 2. *(continued) If F is the distribution function of $X \in Be(p)$, then, with the above definition,*

$$\begin{aligned}
P(X \in (C, 0)) &= \int_{(C,0)} dF(x) \\
&= 0.
\end{aligned}$$

We need one more final convention in order to be able to integrate all the functions that arise in the applications we have in mind. We would also like to be able to state the relation $P(X = c) = \int_{\{c\}} dF(x)$. In fact the definition is basically a simplified version of (6). The one-point set $\{c\}$ can be written as

$$\begin{aligned} \{c\} &:= \cap_{n\geq 1}(c - 1/n, c] \\ &= \lim_{n\uparrow\infty}(c - 1/n, c], \end{aligned}$$

and

$$(c - 1/n, c] \quad \downarrow \quad \{c\}.$$

Since

$$\int_{(c-1/n,c]} dF(x) \quad = \quad F(c) - F(c - 1/n)$$

it makes sense to define $\int_{\{c\}} dF(x)$ by

$$\begin{aligned} \int_{\{c\}} dF(x) &= F(x) - \lim_{n\to\infty} F(x - 1/n) \\ &= F(c) - \lim_{x\uparrow c} F(x). \end{aligned} \tag{8}$$

From the definition (8) we see that if F is continuous at c, then

$$\int_{\{c\}} dF(x) \quad = \quad 0,$$

and if F is discontinuous at c, with a jump discontinuity of size a, then

$$\int_{\{c\}} dF(x) \quad = \quad a.$$

These two relations imply that if F is a distribution function for the random varible X, then with the definition (8), we have established the relation

$$\int_{\{c\}} dF(x) \quad = \quad P(X = c).$$

Recall that we are dealing with random variables that are discrete, continuous or of mixed kind. With the above definitions and conventions we are able to state the relation, with the use of the Riemann-Stieltjes integral,

$$\int_A dF(x) \quad = \quad P(X \in A), \tag{9}$$

for all types of random variables X that we will encounter, and all sets A that are countable unions of one point sets $\{x_0\}$, and open or half-open intervals, $(a, b), [a, b)$ or $(a, b]$, for $a < b$.

2.2 Integration over unbounded sets

We have in the book stated that we can define the integral over unbounded sets

$$\int_{\mathbb{R}} g(x)dF(x)$$

as a limit of integrals over compacts sets, with the limit taken as the size of the compacts sets goes to infinity. This is in analogy with the construction for the usual Riemann integral. However, if we note that the functions F that we integrate with respect to are distribution functions and are therefore bounded, we may note that there is no need to take limits. In fact the unbounded real line $\mathbb{R}$ is transformed, by the application of F, to the unit interval $[0, 1]$, i.e. $F(\mathbb{R}) \subseteq [0, 1]$. Therefore, we can let $I \subseteq \mathbb{R}$ be possibly unbounded, e.g. $I = \mathbb{R}$, and let F be an arbitry distribution function, and note that now

$$F : I \mapsto F(I) \subseteq [0, 1].$$

Next, we assume that h is a real-valued step-function that is bounded on the interval I, so that

$$h(t) \quad = \quad \sum_{i=1}^{n} c_i 1\{t \in I_i\},$$

with some finite constants $c_1, \ldots, c_n$, and where $I = \cup_{i=1}^{n} I_i$ is a partition of I, with I_j intervals, some of wich are possibly infinite. Then we can define the Riemann-Stieltjes integral of h with respect to the distribution function F as

$$\begin{aligned} \int h(u)\, dF(u) \quad &= \quad \sum_{i=1}^{n} c_i |F(I_i)| \qquad (10) \\ &= \quad \sum_{i=1}^{n} c_i (F(b_i) - F(a_i)). \end{aligned}$$

Note that by the assumptions, the integral will be finite. Having defined the integral of step functions, the definition of the Riemann-Stieltjes integral is now straightforward. We state it and the subsequent theorems for slightly more general F.

Definition 2. *Let F be an increasing and bounded function defined on the, possibly infinite, interval I, and let g be a function defined on I. Then g is called Riemann-Stieltjes integrable, with respect to F, if for every $\epsilon > 0$, there are step functions h_1, h_2 such that $h_1 \leq g \leq h_2$ and*

$$\int_I h_2(u)\, dF(u) - \int_I h_1(u)\, dF(u) \quad < \quad \epsilon.$$

If g is Riemann-Stieltjes integrable, we define the Riemann-Stieltjes integral of g as

$$\int_I g(u)\, dF(u) \quad = \quad \sup\{\int_I h(u)\, dF(u) : h \leq g, h \text{ step function.}\}$$

One can check that the integral is well defined. Furthermore, the claims that functions g that are continuous are integrable, that one can approximate the integral by Riemann-Stieltjes sums, and that one can make the extensions as in the previous subsection are straightforward. We state the results for $I = \mathbb{R}$ directly.

Theorem 4. *Suppose that F is an increasing and bounded function and that g is bounded and piecewise continuous on $\mathbb{R}$, and such that $D_g \cap D_F = \emptyset$. Then g is Riemann-Stieltjes integrable with respect to F, and*

$$\int_{\mathbb{R}} g(x)\, dF(x) \quad = \quad \sum_{i=1}^{m} \int_{J_i} g(x) dF(x),$$

with $(D_g)^c = \cup_{i=1}^m J_i$ the partition defined in (3).

Theorem 5. *Assume g is bounded and piecewise continuous with D_g the set of discontinuity points of g, F increasing and bounded on $\mathbb{R}$, and $D_g \cap D_F = \emptyset$. Then*

$$\int_{\mathbb{R}} g(x)\, dFx \quad = \quad \lim_{\max_{1\le i\le n} |I_i| \to 0} \sum_{i=1}^{n} g(\xi_i)|F(I_i)|,$$

where

$$\cup_{i=1}^n I_i \quad = \quad (D_g)^c$$

are partitions of $(D_g)^c$, with $I_i \subseteq J_k$ for each i and some k, where J_k are the partition sets in (3), and ξ_i are arbitrary points in I_i.

The extensions for functions $g \in C_{pc}$ such that $D_g \cap D_F \neq \emptyset$ is identical to the construction for integrals over finite sets, and so is the construction of integrals over one-point sets. Thus we again obtain the relation (9), with A possibly a set of total length infinity.

3 Complements to testing

Suppose that we aim to construct a test (T, C), for a given significance level α based on i.i.d. observations $x_1, \ldots, x_n$ of the random variable X, with the fixed and unknown distribution F_θ, or equivalently, with the distribution function F_θ for the fixed and unknown parameter $\theta \in \Theta$, of the hypotheses

$$\begin{aligned} H_0 : &\quad \theta \in \Theta_0, \\ H_1 : &\quad \theta \in \Theta_1. \end{aligned}$$

Here $\Theta = \Theta_0 \cup \Theta_1$ is a partition of the parameter space, and Θ_0 can be a one point set, such as $\Theta_0 = \{\theta_0\}$, in which case we call the null hypothesis a *simple* hypothesis, or it can consist of several points, such as e.g. an interval $\Theta_0 = (-\infty, \theta_0]$, in which case we call the null hypothesis *composite*.

We have claimed that, in the two cases, we find a critical region C as a set such that

$$\begin{aligned} \alpha &= P(T \in C | \theta = \theta_0), \\ \alpha &= \sup_{\theta \in \Theta_0} P(T \in C | \theta), \end{aligned}$$

respectively. In the two cases, we have interpreted $P(T \in C|\theta = \theta_0)$ as the probability that T is in C given that the true value of θ is θ_0 and $P(T \in C|\theta)$ as the probability that T is in C given that θ is the true value, respectively. In both cases we are stating a probability of something *given that θ takes a value.* However, θ is not a random variable, it is a constant, so how can we condition on it being something? In fact, we can not, the notation is quite common and is really an abuse of the conditional probability notation.

The correct way to express the above becomes clear when we recall that a probability of the type above, as $P(T \in C)$, is expressed as an integral with respect to a distribution function. To clarify this, and in order to avoid cumbersome notation with multiple integrals, suppose that we have one observation x of X. Then the statistic $T = T(x)$ is a function on $\mathbb{R}$, and if we treat it as a random variable $T = T(X)$, we can write, for any (nice) set C

$$\begin{aligned} P(T(X) \in C) &= E(1\{T(X) \in C\}) \\ &= \int 1\{T(x) \in C\} dF(x). \end{aligned} \tag{11}$$

Here we have merely used the formula $E(g(X)) = \int g(x)dF(x)$, for the function $g(x) = 1\{T(x) \in C\}$. But we see that in the last integral in (11) the distribution function $dF(x)$ is really $dF_\theta(x)$, i.e. that it should be written

$$\int 1\{T(x) \in C\} dF_\theta(x),$$

where θ is the parameter that determines the distribution function for the random variable X. That means that the value of the integral depends on θ. This of course also means that the probability $P(T(X) \in C)$ depends on θ. We can therefore write

$$P_\theta(T(X) \in C) = P_\theta(T \in C)$$

meaning the probability that T is in C, when θ is the parameter that determines the distribution of x, or *when θ is the true value.* With this notation we can now give a correct formalisation of the above significance level expressions. They are

$$\begin{aligned} \alpha &= P_{\theta_0}(T \in C), \\ \alpha &= \sup_{\theta \in \Theta_0} P_\theta(T \in C), \end{aligned}$$

respectively.

Similarly the power function for a test (T, C) can be written as

$$\pi(\theta) = P_\theta(T \in C).$$

4 Some results on stochastic convergence

In this section we state and prove two results that are useful for the inference theory applications that are treated in the book. Both results are useful for deriving limit distributions. They are Slutsky's theorem and the continuous mapping theorem. We will state and prove only the simplest versions of the two theorems, they are usually stated in greater generality, see e.g. [1] for more general versions.

Theorem 6. *(Slutsky) Suppose that $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ are two sequences of random variables, that $\{X_n\}$ are independent of $\{Y_n\}$, X is a random variable and $c \neq 0$ is a constant, and that*

$$Y_n \stackrel{d}{\to} Y \qquad X_n \stackrel{P}{\to} c.$$

Then

$$X_n Y_n \quad \stackrel{d}{\to} \quad Yc$$

Proof. Let x be an arbitrary continuity point of the distribution function F_{cY}. Note that this implies that x/c is a continuity point of F_Y. We want to show that $F_{X_nY_n}(x) \to F_{cY}(x)$ as $n \to \infty$.

Let $\delta > 0$ and $\epsilon > 0$ be given. We have that

$$\begin{aligned} F_{X_nY_n}(x) &= P(X_nY_n \leq x) \\ &= P(X_nY_n \leq x, |X_n - c| > \epsilon) + P(X_nY_n \leq x, |X_n - c| \leq \epsilon) \\ &\leq P(|X_n - c| > \epsilon) + P(X_nY_n \leq x, |X_n - c| \leq \epsilon) \\ &\leq \delta + P(X_nY_n \leq x, |X_n - c| \leq \epsilon), \qquad (12) \end{aligned}$$

where the last inequality holds if $n \geq N$ for some $N = N(\epsilon, \delta)$, since $X_n \stackrel{P}{\to} c$. On the other hand, from the second equality above, we have

$$F_{X_nY_n}(x) \quad \geq \quad P(X_nY_n \leq x, |X_n - c| \leq \epsilon). \qquad (13)$$

Note that

$$\begin{aligned} A_n(\epsilon) &:= \{X_nY_n \leq x\} \cap \{|X_n - c| \leq \epsilon\} \\ &= \{X_nY_n \leq x\} \cap \{c - \epsilon \leq X_n \leq c + \epsilon\}. \end{aligned}$$

Assume that $c - \epsilon > 0$ (which follows from $c > 0$ if ϵ is small enough, and which we can assume without loss of generality; the proof for the case $c < 0$ is left as an exercise.) Then

$$\{Y_n \leq \frac{x}{c+\epsilon}\} \cap C_n(\epsilon) \subseteq A_n(\epsilon) \subseteq \{Y_n \leq \frac{x}{c-\epsilon}\} \cap C_n(\epsilon), \qquad (14)$$

with $C_n(\epsilon) = \{c - \epsilon \leq X_n \leq c + \epsilon\}$. Furthermore

$$\begin{aligned} P(Y_n \leq \frac{x}{c+\epsilon}) &= P(\{Y_n \leq \frac{x}{c+\epsilon}\} \cap C_n(\epsilon)) + P(\{Y_n \leq \frac{x}{c+\epsilon}\} \cap C_n(\epsilon)^c) \\ &\leq P(A_n(\epsilon)) + P(C_n(\epsilon)^c) \\ &\leq P(A_n(\epsilon)) + \delta, \qquad (15) \end{aligned}$$

if $n \geq N$.

Thus (13), (15) imply

$$\begin{aligned} F_{X_n Y_n}(x) &\geq P(A_n(\epsilon)) \\ &= P(Y_n \leq \frac{x}{c+\epsilon}) - \delta. \end{aligned}$$

Since x/c is a continuity point of F_Y, x/c lies in an open set on which F_Y is continuous, and thus if ϵ is small enough also $x/(c+\epsilon)$ is a continuity point of F_Y. This implies

$$\lim_{n\to\infty} F_{X_n Y_n}(x) \geq F_Y(\frac{x}{c+\epsilon}) - \delta. \tag{16}$$

On the other hand, (12), (14) imply

$$\begin{aligned} F_{X_n Y_n}(x) &\leq \delta + P(\{Y_n \leq \frac{x}{c-\epsilon}\} \cap C_n(\epsilon)) \\ &\leq \delta + P(Y_n \leq \frac{x}{c-\epsilon}) \end{aligned}$$

Reasoning similarly to as above we see that if ϵ is small enough then $x/(c-\epsilon)$ is a continuity point for F_Y, which implies that

$$\lim_{n\to\infty} F_{X_n Y_n}(x) \leq \delta + F_Y(\frac{x}{c-\epsilon}). \tag{17}$$

Since $\delta > 0$ and $\epsilon > 0$ are arbitrary and F_Y is continuous at x/c, (16), (17) imply that

$$\lim_{n\to\infty} F_{X_n Y_n}(x) = F_Y(\frac{x}{c}). \tag{18}$$

Finally, noting that $F_Y(x/c) = F_{cY}(x)$, we have proved the theorem. □

The next results states that (certain forms of) stochastic convergence is invariant under continuous transformations.

Theorem 7. *(Continuous mapping theorem) Suppose that $\{X_n\}_{n\geq 1}$ is a sequence of random variables such that either of*

$$\begin{aligned} (i) &\quad X_n \overset{a.s.}{\to} X, \\ (ii) &\quad X_n \overset{P}{\to} c, \end{aligned}$$

hold, for some random variable X and constant c. Suppose that $g : \mathbb{R} \mapsto \mathbb{R}$ is a continuous function. Then

$$\begin{aligned} (i) &\quad g(X_n) \overset{a.s.}{\to} g(X), \\ (ii) &\quad g(X_n) \overset{P}{\to} g(c) \end{aligned}$$

hold, respectively.

Proof. We note first, as proved in the book, that if X_n are random variables, on the probability space $(\Omega, \mathcal{F}, P)$, and g is continuous then $g(X_n)$ are random variables.

(i): This follows from the characterisation of almost sure convergence: There exists a set $B \in \mathcal{F}$ with $P(B) = 1$, such that for every $\omega \in B$

$$X_n(\omega) \quad \to \quad X(\omega),$$

which means that for every $\delta > 0$ there exists a finite $N = N(\omega, \delta)$ such that if $n \geq N$ then $|X_n(\omega) - X(\omega)| < \delta$. Now since g is continuous at $X(\omega)$, we have that for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon, \omega) > 0$ such that if $|X_n(\omega) - X(\omega)| < \delta$ then $|g(X_n(\omega)) - g(X(\omega))| < \epsilon$.

Thus we have shown that for arbitrary fixed $\omega \in B$, for every $\epsilon > 0$ there is a $\delta = \delta(\epsilon, \omega) > 0$ and then a finite $N = N(\epsilon, \delta, \omega)$ such that if $n \geq N$ then $|X_n(\omega) - X(\omega)| < \epsilon$. This proves

$$g(X_n) \quad \stackrel{a.s.}{\to} \quad g(X).$$

(ii): Let $\epsilon > 0$ be arbitrary but fixed. Then since g is continuous at c, there is a $\delta > 0$ such that if $|X_n - c| < \delta$ then $|g(X_n) - g(c)| < \epsilon$, i.e.

$$\{|X_n - c| < \delta\} \quad \subseteq \quad \{|g(X_n) - g(c)| < \epsilon\}. \tag{19}$$

This implies that

$$\begin{aligned} P(|g(X_n) - g(c)| < \epsilon) &= P(|g(X_n) - g(c)| < \epsilon, |X_n - c| < \delta) \\ &\quad + P(|g(X_n) - g(c)| < \epsilon, |X_n - c| \geq \delta) \\ &= P(|X_n - c| < \delta) \\ &\quad + P(|g(X_n) - g(c)| < \epsilon, |X_n - c| \geq \delta) \\ &\geq P(|X_n - c| < \delta). \end{aligned}$$

where the second equality follows by (19). The right hand side converges to one by the assumption for (ii), which proves (ii). □

Note that only part (ii) of the previous theorem is used in the book. It is then typically used together with Slutsky's theorem, and it is e.g. used to conclude that if

$$s^2 \quad \stackrel{P}{\to} \quad \sigma^2$$

then for instance

$$\begin{aligned} s^2/\sigma^2 &\quad \stackrel{P}{\to} \quad 1, \\ \sqrt{s^2} &\quad \stackrel{P}{\to} \quad \sqrt{\sigma^2}. \end{aligned}$$

We may finish by noting that the continuous mapping theorem also holds for convergence in distribution, and is in fact most often applied on that form. Thus it states that if $X_n \xrightarrow{d} X$ and g is continuous (real-valued) then $g(X_n) \xrightarrow{d} g(X)$. This is not a difficult result to prove but is still best left to more advanced texts, since we do not use that version of the continuous mapping theorem anywhere in this book.

References

[1] A.W. van der Vaart. *Asymptotic Statistic*. Cambridge University Press, 2000.