Riemann-Stieltjes integrals

Dragi Anevski
Mathematical Sciences
Lund University

October 28, 2012

1 Introduction

This short note gives an introduction to the Riemann-Stieltjes integral on $\mathbb{R}$ and $\mathbb{R}^n$. Some natural and important applications in probability theory are discussed. The reason for discussing the Riemann-Stieltjes integral instead of the more general Lebesgue and Lebesgue-Stieltjes integrals are that most applications in elementary probability theory are satisfactorily covered by the Riemann-Stieltjes integral. In particular there is no need for invoking the standard machinery of monotone convergence and dominated convergence that hold for the Lebesgue integrals but typically do not for the Riemann integrals.

The reason for introducing Stieltjes integrals is to get a more unified approach to the theory of random variables, in particular for the expectation operator, as opposed to treating discrete and continuous random variables separately. Also it makes it possible to treat mixtures of discrete and continuous random variables: It is for instance **not possible** to show that the expectation of the sum of a discrete and a continuous r.v. is the sum of the expectations, without using Stieltjes integrals. There are also many advantages in inference theory, for instance in the discussion of plug-in estimators.

In Section 2 we introduce the Riemann-Stieltjes integral on $\mathbb{R}$. In Section 3 we discuss some important applications to probability theory. In Section **??** we introduce the Riemann-Stieltjes integral on $\mathbb{R}^n$. Section 3 contains applications to probability theory.

2 The Riemann-Stieltjes integral on $\mathbb{R}$

The Reimann integral corresponds to making no transformation of the x-axis in the Reimann sum

$$\sum_{i=1}^{n} g(\xi_i)(x_{i-1} - x_i).$$

Sometimes one would like to make such a transformation.

Thus let F be a monotone real-valued transformation of $I \subset \mathbb{R}$, so

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

Now assume that h is a real-valued step-function on the interval I, so that we can write

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

for some constants $c_1, \dots, c_n$, and with $I = \cup_{i=1}^n I_i$ a partition, and with I_j intervals. Then we define the Reimann-Stieltjes integral of h as

$$\int h(u)\,dF(u) \quad = \quad \sum_{i=1}^n c_i |F(I_i)|.$$

Note that $|F(I)| = F(\max(I_i)) - F(\min(I_i))$, and that if $I_i = (a_i, b_i)$ then $|F(I_i)| = F(b_i) - F(b_i)$, and that then

$$\int_I h(u)\,dF(u) \quad = \quad \sum_{i=1}^n c_i (F(b_i) - F(a_i)).$$

We can next make the following definition.

Definition 1 *Let F be an increasing function defined on the 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 stepfunctions 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, similarly to the Riemann integral.

Note 1 *The Riemann integral is a special case of the Riemann-Stieltjes integral, when $F(x) = id(x) = x$ is the identity map. Thus in that case, if g is Riemann integrable,*

$$\int g(x)\,dF(x) \quad = \quad \int g(x)\,dx$$

when $F(x) = x$. □

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

Proof. That F is increasing and bounded on $I = [a, b]$ means that

$$-\infty < c := F(a) = \inf_I F \leq \sup_I F = F(b) =: C < \infty.$$

(i): Assume first that I is finite. Thus since g is continuous on the compact interval I it is uniformly continuous so there are ϵ, δ such that

$$|x - y| \leq \delta \quad \Rightarrow \quad |g(y) - g(x)| \leq \epsilon.$$

for ϵ, δ not depending on x, y. Next let $\cup_{i=1}^n I_i = I$ be an arbitrary finite partition of I, with I_i intervals that satisfy $|F(I_i)| \leq \delta$ (this is possible to obtain since F is bounded), and let

$$\begin{aligned} m_i &= \inf_{t \in I_i} g(t), \\ M_i &= \sup_{t \in I_i} g(t). \end{aligned}$$

and note that $M_i - m_i < \epsilon$, by the uniform continuity of g. The step functions

$$\begin{aligned} h_1(t) &= \sum_{i=1}^n m_i 1\{t \in I_i\}, \\ h_2(t) &= \sum_{i=1}^n M_i 1\{t \in I_i\}, \end{aligned}$$

satisfy $h_1 \leq g \leq h_2$. Furthermore

$$\int_I h_1(u)\, dF(u) = \sum_{i=1}^n m_i |F(I_i)| \quad \leq \quad \sum_{i=1}^n M_i |F(I_i)| = \int_I h_2(u)\, dF(u),$$

so that

$$\int_I h_2(u)\, dF(u) - \int_I h_1(u)\, dF(u) \quad \leq \quad \sum_{i=1}^n (M_i - m_i)|F(I_i)| \leq \epsilon \sum_{i=1}^n |F(I_i)| = \epsilon |F(I)|,$$

where the last equality follows since the sets $F(I_i)$ are disjoint (by the monotonicity of F together with the fact that I_i are disjoint). Since for every $\epsilon > 0$ we can get h_1, h_2 step functions so that this holds, we have shown that g is Riemann-Stieltjes integrable.

(ii): Next, assume I not finite. Then since F is increasing it is also piecewise continuous. Therefore for every $\tilde{\epsilon} > 0$ there is finite $\tilde{I} \subset I$ such that

$$\max(\sup_I F - \sup_{\tilde{I}} F, \inf_{\tilde{I}} F - \inf_I F) \quad < \quad \tilde{\epsilon}.$$

Also, since g is absolutely bounded,

$$\sup_{I \setminus \tilde{I}} |g| \leq G$$

so that

$$\begin{aligned} -G \leq g \leq G \text{ on } I \setminus \tilde{I}, \\ \int_{I \setminus \tilde{I}} G\, dF(u) - \int_{I \setminus \tilde{I}} -G\, dF(u) \quad \leq \quad 2G\tilde{\epsilon}. \end{aligned}$$

Thus we can use the construction under (i) on $\tilde{I}$, and concatenate to get the step functions $\tilde{h}_1 = \text{conc}(-G, h_1, -G), \tilde{h}_2 = \text{conc}(G, h_2, G)$ bounding g on all of I and such that

$$\begin{aligned} \int_I \tilde{h}_2\, dF(u) - \int_I \tilde{h}_1\, dF(u) &= \int_{\tilde{I}} h_2(u)\, dF(u) - \int_{\tilde{I}} h_1(u)\, dF(u) \\ &\quad + \int_{I \setminus \tilde{I}} G\, dF(u) - \int_{I \setminus \tilde{I}} -G\, dF(u) \\ &\leq \epsilon |F(\tilde{I})| + 2G\tilde{\epsilon}. \end{aligned}$$

Since $\epsilon, \tilde{\epsilon}$ are arbitrary we have shown that g is Riemann-Stieltjes integrable. □

The Riemann-Stieltjes integral can be obtained as a limit of Riemann-Stieltjes sums. We prove the statements for continuous functions g:

Theorem 2 *Assume g is continuous and F increasing on the interval I. Then*

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

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

Proof. Use a similar construction of h_1, h_2 as in the proof of Theorem 1. Thus we have

$$h_1 \le g \le h_2.$$

and

$$\int_I h_1(u)\,dF(u) \le \sum_{i=1}^n g(\xi_i)(F(x_i) - F(x_{i-1})) \le \int_I h_2(u)\,dFu.$$

Since g is integrable, we can let $\epsilon \downarrow 0$, and make the partition finer and finer as $n \to \infty$, so that the difference between the right hand side and the left hand side (which is smaller than ϵ) goes to zero, which shows the result. □

We note the following two important special cases.

Lemma 1 *Assume F is (an increasing) step function on I, so that*

$$F(t) \quad = \quad \sum_{i=1}^N a_i 1\{t \le t_i\},$$

with $t_0 = \min(I) < t_1 \ldots < t_N = \max(I)$*, and* $a_i \ge 0$[1]*. Then, if g is contiunous,*

$$\int_I g(x)\,dF(x) \quad = \quad \sum_i^N g(t_i) a_i.$$

Proof. When forming the Riemann-Stieltjes sum

$$\sum_{i=1}^n g(\xi_i)(F(x_i) - F(x_{i-1}))$$

[1] The condition $a_i \ge 0$ ensures that F is increasing. An equivalent way to write F is

$$F(x) \quad = \quad \sum_{i=1}^N (a_i - a_{i-1}) 1\{t \in (t_{i-1}, t_i)\}.$$

in Theorem 4, the factor $F(x_i) - F(x_{i-1})$ is

$$F(x_i) - F(x_{i-1}) = \begin{cases} a_k \text{ if } < x_{i-1} < t_k < x_i, \text{ and } x_i - x_{i-1} \text{ small enough,} \\ 0 \text{ if } (x_{i-1}, x_i) \not\ni t_k \text{ for any } k. \end{cases}$$

Thus for large enough n

$$\sum_{i=1}^{n} g(\xi_i)(F(x_i) - F(x_{i-1})) = \sum_{i=1}^{N} g(\xi_i) a_i,$$

Since g is continuous, and $\xi \to t_i$ as $n \to \infty$, we obtain $g(\xi) \to g(t_i)$, so that

$$\begin{aligned} \lim_{n\to\infty} \sum_{i=1}^{n} g(\xi_i)(F(x_i) - F(x_{i-1})) &= \lim_{n\to\infty} \sum_{i=1}^{N} g(\xi_i) a_i \\ &= \sum_{i}^{N} g(t_i) a_i, \end{aligned}$$

which ends the proof. □

Lemma 2 *Assume F if differentiable with $F' = f$ continuous. Then if g is integrable*

$$\int_I g(u)\, dF(u) = \int_I g(u) f(u)\, du.$$

Proof. We derive the result via Riemann-Stiltjes sums: In the sum

$$\sum_{i=1}^{n} g(\xi_i)(F(x_i) - F(x_{i-1}))$$

the factor $(F(x_i) - F(x_{i-1})) = f(\eta_i)(x_i - x_{i-1})$ for some $\eta_i \in (x_{i-1}, x_i)$, by the mean value theorem. Therefore

$$\sum_{i=1}^{n} g(\xi_i)(F(x_i) - F(x_{i-1})) = \sum_{i=1}^{n} g(\xi_i) f(\eta_i)(x_i - x_{i-1}),$$

which we recognize as a Riemann sum for the integral $\int gf\, du$, and the result is proven. □

3 Application to probability theory

In probability theory the basic setup is a probability space $(\Omega, \mathcal{F}, P)$. Here Ω is the outcome space, i.e. the set of all possible outcomes for the random experiment we want to model. The set $\mathcal{F}$ is a collection of subsets of Ω, satisfying the three following conditions, making it into a σ-algebra:

$$\begin{array}{ll} (i) & \emptyset \in \mathcal{F}, \\ (ii) & A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}, \\ (iii) & A_i \in \mathcal{F} \text{ for } i = 1, 2, \ldots \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}. \end{array}$$

The function P is a probability measure defined on $\mathcal{F}$, i.e. a function $P : \mathcal{F} \to [0, 1]$ such that:

$$\begin{array}{ll} (i) & P(\emptyset) = 0, \\ (ii) & P(A^c) = 1 - P(A), \\ (iii) & P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i), \text{ if } A_i \cap A_j = \emptyset \text{ when } i \neq j. \end{array}$$

We suppose in (ii) that $A \in \mathcal{F}$, and in (iii) that $A_i \in \mathcal{F}$ for every i for us to be able to apply P to the resulting sets; the corresponding conditions $(ii), (iii)$ in the definition of a σ-algebra ensures that this is alright.

A r.v. X is a function $\Omega \to \mathbb{R}$, such that

$$\{\omega : X(\omega) \leq x\} \quad \in \quad \mathcal{F}$$

for every $x \in \mathbb{R}$; this condition is called *measurability.* It ensures that one can define the distribution function F of X, which is defined as

$$F(x) \quad = \quad P(\omega : X(\omega) \leq x),$$

since the set $\{(\omega : X(\omega) \leq x\}$ is in $\mathcal{F}$, and therefore we can apply P on it.

Definition 2 *Assume that X is a r.v. with distribution function F. The expectation $E(X)$ of X is defined, if it exists, as*

$$E(X) \quad = \quad \int x \, dF(x).$$

□

Note 2 *The interpretation of the expectation $E(X)$ is clear from the Riemann-Stieltjes approximation*

$$E(X) \quad \approx \quad \sum_{i=1}^{n} \xi_i (F(x_i) - F(x_{i-1}))$$

with $\xi \in (x_{i-1}, x_i]$. Namely one takes a value that X can take, ξ_i, and multiplies it with $(F(x_i) - F(x_{i-1}))$, which if $x_i - x_{i-1}$ is small is close to $P(X = \xi_i)$, and sums over all possible such values ξ_i. □

This can be simplified in the following two (extreme) cases, that can be derived by Lemmas 1 and 2.

(i) F a step function with jumps at the points x_i. Then

$$E(X) \quad = \quad \sum_{k=1}^{\infty} x_i(F(x_i) - F(x_i-))$$

(ii) F is differentiable with derivative $f = F'$. Then

$$E(X) \quad = \quad \int x f(x)\, dx.$$

Exercise 1 *Prove the previous statement, using Lemmas 1 and 2.* □

Let g be a function $g : \mathbb{R} \to \mathbb{R}$ such that $\int g(x)\, dF(x) < \infty$ and such that

$$\{\omega : g(X(\omega)) \leq u\} \quad \in \quad \mathcal{F},$$

for every $u \in \mathbb{R}$.[2] This implies that we can define the distribution function of Y, F_Y by

$$F_Y(y) \quad = \quad P(\omega : Y(\omega) \leq y).$$

Then we can of course define the expectation of Y as

$$E(Y) \quad = \quad \int y\, dF_Y(y),$$

if it exists. The next theorem tells us that it is not necessary to derive the distribution of Y, in order to calculate the expectation $E(Y)$.

Theorem 3 *If X is a r.v. and g as above then the r.v. $Y = g(X)$ has expectation*

$$E(Y) \quad = \quad \int g(x)\, dF(x).$$

Proof. A Riemann-Stieltjes sum for the left hand side is

$$\begin{aligned} \sum_i \eta_i(F_Y(y_i) - F_Y(y_{i-1})) \quad &= \quad \sum_i \eta_i P(Y \in (y_{i-1}, y_i]) \\ &= \quad \sum_i \eta_i P(g(X) \in (y_{i-1}, y_i]) \\ &= \quad \sum_i \eta_i P(X \in g^{-1}\{(y_{i-1}, y_i]\}), \end{aligned}$$

with $\eta_i \in (y_{i-1}, y_i]$. Note that

$$\begin{aligned} \eta_i \in (y_{i-1}, y_i] \quad &\Leftrightarrow \quad \xi_i := g^{-1}(\eta_i) \in g^{-1}\{(y_{i-1}, y_i]\} \\ &\Leftrightarrow \quad g(\xi_i) \in (y_{i-1}, y_i]. \end{aligned}$$

[2]This means that $g(X)$ is measurable.

Therefore the above is equal to

$$\sum_i g(\xi_i) P(X \in g^{-1}\{(y_{i-1}, y_i]\}),$$

with $\xi_i \in g^{-1}\{(y_{i-1}, y_i]\}$.

Note furthermore that if the intervals $(y_{i-1}, y_i]$ form a partition (so are disjoint and have as union the whole interval), then the intervals $(x_{i-1}, x_i] = g^{-1}\{(y_{i-1}, y_i]\}$ also form a partition. Therefore the above can be written as

$$\sum_i g(\xi_i) P(X \in (x_{i-1}, x_i]),$$

with $\xi_i \in (x_{i-1}, x_i]$, which is a Riemann-Stieltjes sum for the right hand side, and we are done. □

The statement of the theorem, can be written, in the two special cases of F a step function and F differentiable with derivative $f = F'$

$$E(g(X)) = \begin{cases} \int g(x) f(x)\, dx, \\ \sum g(x_i)(F(x_i) - F(x_{i-1})). \end{cases}$$

The special case of $g(X) = 1\{X \in A\}$ for A a set in $\mathbb{R}$ is of particular interest. This is a Bernoulli random variable, it's distribution function is a step function with jumps at 0 and 1, so that

$$\begin{aligned} E(1\{X \in A\}) &= 0(F(0) - F(0-)) + 1(F(1) - F(1-)) \\ &= F(1) - F(1-) \\ &= P(g(X) = 1) \\ &= P(X \in A). \end{aligned}$$

Note also that the left hand side of this is

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

and thus we get the very useful (and important!) formula

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

Note that in particular $1 = P(X \in \mathbb{R}) = \int_{\mathbb{R}} dF(x)$.

Lemma 3 *Let X be a r.v. with distribution function F. Then*

$$E(aX + b) \quad = \quad aE(X) + b.$$

Proof. We have that

$$\begin{aligned} E(aX + b) &= \int (ax + b)\, dF(x) \\ &= a \int x\, dF(x) + b \int dF(x) \\ &= aE(X) + b. \end{aligned}$$

□

We next define the variance of a random varible.

Definition 3 *The variance of a r.v. X with distribution function is defined (if it exists) as*

$$\mathrm{Var}(X) \quad = \quad E((X - \mu)^2),$$

with $\mu = E(X)$. □

We immediately get the following easy (and very important) formula for calculating the variance:

$$\begin{aligned} \mathrm{Var}(X) &= E((X - \mu)^2) = E(X^2 - 2X\mu + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2 = E(X^2) - (E(X))^2. \end{aligned}$$

We state this as a simple Lemma.

Lemma 4 *If $\mathrm{Var}(X)$ exists then*

$$\mathrm{Var}(X) \quad = \quad E(X^2) - (E(X))^2.$$

□

Note 3 *The expectation $E(X)$ of a random variable is a measure of where the distribution is concentrated. Compare for instance the two distributions $Un(0,1)$ and $Un(1,2)$ A random variable X_1 which is distributed according to $Un(0,1)$ has expecion $E(X) = \int x 1\{0 \leq x \leq 1\} dx = 1/2$. A r.v. X_1 distributed according to $Un(1,2)$ has expectation $E(X) = \int x 1\{1 \leq 2\} dx = 3/2$. Check that the two variables X_1, X_2have the same variance!*

The variance $\mathrm{Var}(X)$ of a random variable is a measure of how spread out the distribution of X is. The two random variables $Y_1 \in Un(-1,1)$ and $Y_2 \in Un(-2,2)$ have the same expectation $E(Y_1) = E(Y_2) = 0$. Check that $\mathrm{Var}(Y_1) < \mathrm{Var}(Y_2)$. Draw a graph of the distribution functions. Does this make sense? □

Example 1 *We calculate the expectation and variance of $X \in Bern(p), Y \in Exp(\theta)$.*
Since

$$\begin{aligned} E(X) &= \int x dF_X(x) = \sum_{x_i} x_i f(x_i) = 0f(0) + 1f(1) = 0(1-p) + 1p = p \\ E(X^2) &= \int x^2 dF_x = \sum_{x_i} x_i^2 f(x_i) = 0^2 f(0) + 1^2 f(1) = 0(1-p) + 1p = p. \end{aligned}$$

we have

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p).$$

Also

$$\begin{aligned} E(Y) &= \int y dF_Y(y) = \int y f(y) = \int_0^\infty y \frac{1}{\theta} e^{-y/\theta}\, dy = \ldots = \theta, \\ E(Y^2) &= \int y^2 dF_Y(y) = \int_0^\infty y^2 \frac{1}{\theta} e^{-y/\theta} = \ldots = 2\theta^2, \end{aligned}$$

so that

$$\mathrm{Var}(Y) = E(Y^2) - (E(Y))^2 = 2\theta^2 - \theta^2 = \theta^2.$$

□

Lemma 5 *Let X be a r.v. with distribution function F. Then*

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X).$$

Proof. Note first that $E(aX + b) = aE(X) + b$. Therefore

$$\begin{aligned} \mathrm{Var}(aX + b) &= E((aX + b - aE(X) - b)^2) \\ &= E((aX - aE(X))^2) \\ &= E(a^2(X - \mu)^2) \\ &= a^2 E((X - \mu)^2). \end{aligned}$$

□

Note that since the variance has an another dimension than X, namely the dimension of $\mathrm{Var}(X)$ is $\dim(X)^2$, we introduce the standard deviation $D(X)$ as

$$D(X) = \sqrt{\mathrm{Var}(X)},$$

which exists if the variance exists.

Example 2 *If $X \in Exp(\theta)$ then $D(X) = \theta$.* □

4 The Riemann-Stieltjes integral on $\mathbb{R}^n$

In analogy with integration over $\mathbb{R}$, assume $F : \mathbb{R}^n \to \mathbb{R}$ is a function that is increasing in each variable (when the others are kept fixed). Let h be a step function that is piecewise constant (equal to c_i say) on the rectangles $I_i = [a_1^i, b_1^i] \times \ldots \times [a_n^i, b_n^i]$ in $\mathbb{R}^n$, and $I = \cup_i^n I_i$ is a partition. Then we can define the integral

$$\int_I h(x)\, dF(x) \quad = \quad \sum_{i=1}^{n} c_i |F(I_i)|,$$

with $|F(I_i)|$ the area of the transformed rectangle $F(I_i)$. To see how we should define the area of the rectangle let us remind ourselves that the area of the rectangle I_i is

$$|I_i| \quad = (b_1^i - a_1^i) \cdot \ldots \cdot (b_n^i - a_n^i).$$

We would like to write the transformed area, in analogy with this and *in increments of F*.

Recall that in one dimension the area of a interval $I = [a, b]$ is

$$|I| \quad = \quad b - a,$$

and the area of the transformed interval is

$$|F(I)| \quad = \quad F(b) - F(a).$$

In two dimensions the area of the rectangle $I_i = [a_1^i, b_1^i] \times [a_2^i, b_2^i]$ is

$$\begin{aligned} |I_i| \quad &= \quad (b_1^i - a_1^i)(b_2^i - a_2^i) \\ &= \quad b_1^i b_2^i - a_1^i b_2^i - b_1^i a_2^i + a_1^i a_2^i, \end{aligned}$$

and we define the area of the transformed rectangle $F(I_i)$ as

$$|F(I_i)| \quad = \quad F(b_1^i, b_2^i) - F(a_1^i, b_2^i) - F(b_1^i, a_2^i) + F(a_1^i, a_2^i)$$

Note that in using F we are making a generalization of the ordinary area measure, for which the area of the interval $I = [0, b_1] \cdots [0, b_n] \subset \mathbb{R}^n$ is $|I| = b_1 \cdots b_n$, to the area measure $|F(I)| = F(b_1, \ldots, b_n)$.

For general dimension, let us use the notation

$$\Delta_j F(I_i) \quad = \quad F(x_1^i, \ldots, x_{j-1}, b_j^i, x_{j+1}, x_n) - F(x_1, \ldots, x_{j-1}, a_j^i, x_{j+1}, x_n).$$

(Note that the argument is fixed at all positions except at position j where we have the increment). Then we define the area by

$$|F(I_i)| \quad = \quad \Delta_1 \ldots \Delta_n F(I_i)$$

Definition 4 *Let F be an increasing function defined on the interval $I \subset \mathbb{R}^n$, 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 stepfunctions 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, the Riemann-Stieltjes integral of g is defined as

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

□

We start by noting that this is indeed an generalization of the ordinary Riemann integral in several dimensions.

Note 4 *The Riemann integral on $\mathbb{R}^n$ is a special case of the Riemann-Stieltjes integral, with $F(I) = |I|$, for every interval I, the ordinary area length. In that case, if g is integrable, we write*

$$\int g(x)\,dF(x) \quad = \quad \int g(x)\,dx.$$

To see this more explicit, let us look at the case of dimension $n = 2$. Then for $I_i = [a_1^i, b_1^i] \times [a_2^i, b_2^i]$ the area of $F(I)$ becomes

$$\begin{aligned} \Delta_1\Delta_2 F(I_i) &= |(b_1^i, b_2^i)| - |(a_1^i, b_2^i)| - |(b_1^i, a_2^i)| + |(a_1^i, a_2^i)| \\ &= (b_1^i - a_1^i)(b_2^i - a_2^i) \\ &= |I_i|. \end{aligned}$$

Therefore the Riemann-Stieltjes integral of step functions g as above becomes in this case

$$\begin{aligned} \int gdF &= \sum_{i=1}^{n} c_i |I_i| \\ &= \int gdx. \end{aligned}$$

Thus to check that a function g is Riemann-Stieltjes integrable over $\mathbb{R}^n$ is the same as checking that it is Riemann integrable over $\mathbb{R}^n$. □

Similarly to for one-dimensional case, the Riemann-Stieltjes integral can be obtained as a limit of a Riemann-Stieltjes sum.

Theorem 4 *Assume that g is integrable and F is bounded. Then*

$$\int_I gdF \quad = \quad \lim_{\max_i |I_i| \to 0} \sum_{i=1}^{n} g(\xi_i)|F(I_i)|,$$

where $\cup_{i=1}^n I_i = I$ is a partition and with $\xi \in I_i$.

Proof. The proof is similar to the proof of Theorem 2. □

We will mainly use Theorem 4 as a convenient device to prove nice results. The following two special cases are of particular importance.

Lemma 6 *Assume that* $F : \mathbb{R}^n \to \mathbb{R}$ *is such that* $\frac{\partial^n}{\partial_1 \ldots \partial_n} F =: f$ *exists and is continuous, and assume that* g *is R-S integrable. Then*

$$\int g dF \quad = \quad \int g f dx.$$

Proof. We can approximate the area in the Riemann-Stieltjes sum, with the use of e.g. an intermediate value theorem, to obtain

$$\begin{aligned} \sum_{i=1}^{n} g(\xi_i)|F(I_i)| &= \sum_{i=1}^{n} g(\xi_i) \frac{\partial^n}{\partial_1 \ldots \partial_n} F(\eta_i)|I_i| \\ &= \sum_{i=1}^{n} g(\xi_i) f(\eta_i)|I_i| \end{aligned}$$

with $\eta_i \in I_i$. We see that this is (approximately) an ordinary Riemann-sum for the integral $\int g f dx$ and this together with the fact that f is continuous, implies that the sum converges to the integral $\int g f dx$. □

Lemma 7 *Assume* F *is a step function, with jumps on the grid* $\mathcal{G} \subset I$*, of size*

$$\Delta_1 \ldots \Delta_n F(\{\gamma\}) \quad = \quad f_\gamma.$$

Then, if g *is continuous,*

$$\int g dF \quad = \quad \sum_{\gamma \in \mathcal{G}} g(\gamma) f_\gamma.$$

Proof. If $I = \cup_i^m I_i$ is a partition fine enough, each interval I_k will contain at most one point $\gamma \in \mathcal{G}$. Let J be the indices for those I_k's that contain one grid point. Then $\Delta_1 \ldots \Delta_n F(I_j) = f_\gamma$ for some $g \in \mathcal{G}$, if $j \in J$. For all other $j \notin J$, $\Delta_1 \ldots \Delta_n F(I_j) = 0$. Thus, if the partition is fine enough, the Riemann-Stieltjes sum is

$$\sum_{i=1}^{m} g(\xi_i) \Delta_1 \ldots \Delta_n F(I_j) \quad = \quad \sum_{\gamma \in \mathcal{G}} g(\xi_\gamma) f_\gamma,$$

which converges to the right hand side in the statement of the Lemma, if g is continuous. □

[3]

5 Applications to probability theory

Let $X_1, \ldots, X_n$ be a random vector with distribution function $F = F_{X_1,\ldots,X_n}$. Assume $g : \mathbb{R}^n \to \mathbb{R}$ is a function (nice enough) such $Y = g(X_1, \ldots, X_n)$ is a random variable. Then Y has a distribution function F_y and we can of course define (if it exists) it's expectation $E(Y)$. The next result however is often handy.

[3]Skriv ner fÃűr resultatet fÃűr blandningar av diskreta och kontinuerliga s.v.

Theorem 5 *If $X = (X_1, \ldots, X_n)$ is a random vector with distribution function F, and $g : \mathbb{R}^n \to \mathbb{R}$ a function, then the r.v. $Y = g(X_1, \ldots, X_n)$ has expectation*

$$E(Y) \quad = \quad \int g(x_1, \ldots, x_n) dF(x_1 \ldots, x_n).$$

Proof. The proof is similar to the one-dimensional case. Thus a Riemann-Stieltjes sum for the left hand side is

$$\begin{aligned} \sum_i \eta_i (F_Y(y_i) - F_Y(y_{i-1})) &= \sum_i \eta_i P(Y \in (y_{i-1}, y_i]) \\ &= \sum_i \eta_i P(g(X) \in (y_{i-1}, y_i]) \\ &= \sum_i \eta_i P(X \in g^{-1}\{(y_{i-1}, y_i]\}), \end{aligned}$$

with $\eta_i \in (y_{i-1}, y_i]$. Since

$$\begin{aligned} \eta_i \in (y_{i-1}, y_i] \quad &\Leftrightarrow \quad \xi_i := g^{-1}(\eta_i) \in g^{-1}\{(y_{i-1}, y_i]\} \\ &\Leftrightarrow \quad g(\xi_i) \in (y_{i-1}, y_i]. \end{aligned}$$

Therefore the above is equal to

$$\sum_i g(\xi_i) P(X \in g^{-1}\{(y_{i-1}, y_i]\}),$$

with $\xi_i \in g^{-1}\{(y_{i-1}, y_i]\}$.

If the intervals $(y_{i-1}, y_i]$ form a partition of an interval in $\mathbb{R}$ (so are disjoint and have as union the whole interval), then the intervals $I_i = g^{-1}\{(y_{i-1}, y_i]\}$ form a partition in $\mathbb{R}^n$. Therefore the above sum can be written as

$$\sum_i g(\xi_i) P(X \in I_i),$$

with $\xi_i \in I_i$, which is a Riemann-Stieltjes sum for the right hand side, and the theorem is proved. □

In particular if g is an indicator function X beeing in a subset $A \subset \mathbb{R}^n$ we get as a consequence that

$$P(X \in A) \quad = \quad E(1\{X \in A\}) = \int 1\{x \in A\} dF(x) = \int_A dF.$$

Further consequences are summarized in the following Lemma.

Lemma 8 *If X_1, X_2 are random variables, and a_1, a_2 are real numbers then*

$$\begin{aligned} &(i) \qquad E(a_1 X_1 + a_2 X_2) = a_1 E(X_1) + a_2 E(X_2). \\ &(ii) \qquad \mathrm{Var}(a_1 X_1 + a_2 X_2) = a_1^2 \mathrm{Var}(X_1) + a_2^2 \mathrm{Var}(X_2) + 2 a_1 a_2 \mathrm{Cov}(X_1, X_2). \end{aligned}$$

In particular if X_1, X_2 are independent then

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2).$$

Proof. To show (i),

$$\begin{aligned} E(a_1X_1 + a_2X_2) &= \int (a_1x_! + a_2x_2)dF(x_1, x_2) \\ &= \int a_1x_1dF(x_1, x_2) + \int a_2x_2dF(x_1, x_2) \\ &= E(a_1X_1) + E(a_2X_2) \\ &= a_1E(X_1) + a_2E(X_2). \end{aligned}$$

For (ii), let $\mu_1 = E(X_1), \mu_2 = E(X_2)$. Then

$$\begin{aligned} \text{Var}(a_1X_1 + a_2X_2) &= E[(a_1X_1 + a_2X_2 - a_1\mu_1 - a_2\mu_2)^2] \\ &= E[(a_1X_1 - a_1\mu_1)^2 + (a_2X_2 - a_2\mu_2)^2 \\ &\quad +2(a_1X_1 - a_1\mu_1)(a_2X_2 - a_2\mu_2)] \\ &= E[(a_1X_1 - a_1\mu_1)^2] + E[(a_2X_2 - a_2\mu_2)^2] \\ &\quad +2E[(a_1X_1 - a_1\mu_1)(a_2X_2 - a_2\mu_2)] \\ &= a_1^2\text{Var}(X_1) + a_2^2\text{Var}(X_2) + 2a_1a_2\text{Cov}(X_1, X_2), \end{aligned}$$

where the third equality follows from the linearity of the Riemann-Stieltjes integral. □

Note 5 *Since we have define expectation using the Riemann-Stieltjes integral, and thus obtaining a general formula that holds for continuous and discrete r.v.'s as well as for mixtures, we can define the expectation of a sum $X + Y$ where X is continuous and Y is discrete.*

The expectation of $X + Y$ as above is **not possible** *to define if we define (only) the expectation of discrete and continuous r.v. cases separately, as sums and ordinary Riemann integrals respectively. This is a flaw with most introductory texts on Probability Theory, and is one of the main reasons that we define and use the Riemann-Stieltjes integral in this text.*
□

5.1 Integration on $\mathbb{R}$

Let I be a finite or infinite interval in $\mathbb{R}$ and assume that $h : \mathbb{R} \to \mathbb{R}$ is a step function, i.e. a function that can be written

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

where $\cup_{i=1}^{n} A_i = I$ is a partition of I, so $A_i \cap A_j = \emptyset$ if $i \neq j$, and $\{c_i\}_{i=1}^{n}$ are finite constants. Then we can define the integral of h over the interval I by

$$\int_I h(u)\, du \quad = \quad \sum_{i=1}^{n} c_i |A_i|$$

with $|A|$ the length of A.

Recall the elementary definition of integrability of a function $g : \mathbb{R} \to \mathbb{R}$ over a finite interval I.

Definition 5 *Assume that for every $\epsilon > 0$ there are step functions h_1, h_2 such that $h_1 \leq g \leq h_2$, and such that*

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

Then we say that g is (Riemann) integrable. □

If g is integrable, we next define $\int_I g\, du$.

Definition 6 *Assume g is Riemann integrable. Then we define*

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

□

To see that the definition makes sense, note that if h^* is a fixed but arbitrary stepfunction such that $h^* \geq g$ then

$$\int_I h^*\, du \geq \int_I g\, du \geq \int_I h\, du$$

for every step function $h \leq g$. This means that the set

$$C \quad := \quad \{\int_I h\, du : h \text{ step function}, h \leq g\}$$

is a set of real numbers that is bounded from above, by $\int_I h^*\, du$, and therefore it has a least upper bound, i.e. $\sup C$ exists, and this is what we define as $\int_I g\, du$.

It is easy to see (reasoning as above), that if g is integrable then also

$$\int_I g\, du \quad = \quad \inf\{\int h\, du : h \text{ step function}, h \geq g\}.$$

Exercise 2 *Prove this!*

What functions are Riemann integrable?

Theorem 6 *Assume that I is a finite interval, and that g is a continuous function. Then g is Riemann integrable.*

Proof. Since g is continuous on I, it is uniformly continuous, i.e. for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$|x - y| \le \delta \quad \Rightarrow \quad |g(y) - g(x)| \le \epsilon,$$

and the same δ, ϵ can be used for every x, y. With this choice of ϵ, δ, now let $\cup_{i=1}^n I_i = I$ be an arbitrary partition of I, with I_i intervals of length $|I_i| \le \delta$, and define

$$\begin{aligned} m_i &= \inf_{t \in I_i} g(t), \\ M_i &= \sup_{t \in I_i} g(t). \end{aligned}$$

Then we have $M_i - m_i < \epsilon$. Define the step functions

$$\begin{aligned} h_1(t) &= \sum_{i=1}^n m_i 1\{t \in I_i\}, \\ h_2(t) &= \sum_{i=1}^n M_i 1\{t \in I_i\}, \end{aligned}$$

and note that $h_1 \le g \le h_2$. Then

$$\int_I h_1(u)\, du = \sum_{i=1}^n m_i \delta \quad \le \quad \sum_{i=1}^n M_i \delta = \int_I h_2(u)\, du,$$

and

$$\int_I h_2(u)\, du - \int_I h_1(u)\, du \quad = \quad \delta \sum_{i=1}^n (M_i - m_i) \le \delta \epsilon n \le \epsilon |I|.$$

Since for every $\epsilon > 0$ we can get h_1, h_2 step functions so that this holds, we thus have shown that g is integrable. □

Corollary 1 *Assume that g is piecewise continuous on I. Then g is Riemann integrable.*

Proof. Let $I = \cup_{i=1}^k J_k$ be a partition of I such that g is continuous on each J_k. Then g is integrable on each J_k and we can define

$$\int_I g(x)\, dx \quad = \quad \sum_{i=1}^k \int_{J_k} g(x)\, dx.$$

(To explicitly exhibit upper and lower step functions h_1, h_2 that satisfy the condition in the definition of integrability on all of I choose

$$\begin{aligned} h_1 &= \text{conc}(h_1^{(1)}, \ldots, h_1^{(k)}), \\ h_2 &= \text{conc}(h_2^{(1)}, \ldots, h_2^{(k)}), \end{aligned}$$

with conc denoting the concatenation of functions.) □

If g is Riemann-integrable one can obtain the integral $\int_I g\,dx$ as a limit of Riemann sums. We prove the result for continuous functions:

Theorem 7 *Assume g is continuous. Then*[4]

$$\int_I g(x)\,dx \quad = \lim_{\max_{1\le i\le n}|x_i-x_{i-1}|\to 0} \sum_{i=1}^n g(\xi_i)(x_i - x_{i-1}),$$

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

Proof. Since g is continuous is it Riemann integrable. Let $x_1 < \ldots < x_n$ be a partition of I. Form the (upper and lower) step functions h_1, h_2, (via the bounds M_i, m_i,) using $I_i = [x_{i-1}, x_i)$ as in the proof of the previous theorem. Note that $m_i \le g(\xi_i) \le M_i$ for arbitrary $\xi_i \in [x_{i-1}, x_i)$ and every i such that

$$h_1 \le g \le h_2.$$

Therefore

$$\int_I h_1(u)\,du \le \sum_{i=1}^n g(\xi_i)(x_i - x_{i-1}) \le \int_I h_2(u)\,du.$$

Since g is integrable, letting $\epsilon \downarrow 0$, i.e. making the partition finer and finer as $n \to \infty$, the difference between the right hand side and the left hand side (which is smaller than ϵ) goes to zero which shows the result. □

5.2 Integration on $\mathbb{R}^n$

What has been covered for $\mathbb{R}$ can be readily extended to $\mathbb{R}^n$

Indeed, if I is a finite or infinite interval in $\mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}$ is a step function,

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

with $\cup_{i=1}^n A_i = I$ a partition of I and c_i constants, then we can define the integral of h over the interval I by

$$\int_I h(u)\,du \quad = \quad \sum_{i=1}^n c_i |A_i|$$

with $|A|$ the Euclidian length of A.

As for the univariate we can define integrability of $g : \mathbb{R}^n \to \mathbb{R}$.

[4]The limit is a limit as $n \to \infty$ if we with the extra condition that the maximum grid length $\max|x_i - x_{i-1}|$ goes to zero.

Definition 7 *Assume that for every $\epsilon > 0$ there are step functions h_1, h_2 such that $h_1 \leq g \leq h_2$, and such that*

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

Then we say that g is (Riemann) integrable. □

If g is integrable, we next define $\int_I g\,du$.

Definition 8 *Assume g is Riemann integrable. Then we define*

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

□

As for the univariate case it is easy to see that the definition is sensible. The next theorem has a proof that is analogous to the univariate case.

Theorem 8 *Assume that I is a finite interval, and that g is a (piecewise) continuous function. Then g is Riemann integrable.*