Transformations of Random Variables

September, 2009

We begin with a random variable X and we want to start looking at the random variable $Y = g(X) = g \circ X$ where the function

$$g : \mathbb{R} \to \mathbb{R}.$$

The **inverse image** of a set A,

$$g^{-1}(A) = \{x \in \mathbb{R}; g(x) \in A\}.$$

In other words,

$$x \in g^{-1}(A) \text{ if and only if } g(x) \in A.$$

For example, if $g(x) = x^3$, then $g^{-1}([1, 8]) = [1, 2]$

For the singleton set $A = \{y\}$, we sometimes write $g^{-1}(\{y\}) = g^{-1}(y)$. For $y = 0$ and $g(x) = \sin x$, $g^{-1}(0) = \{k\pi; k \in \mathbb{Z}\}$.

If g is a one-to-one function, then the inverse image of a singleton set is itself a singleton set. In this case, the inverse image naturally defines an inverse function. For $g(x) = x^3$, this inverse function is the cube root. For $g(x) = \sin x$ or $g(x) = x^2$ we must limit the domain to obtain an inverse function.

Exercise 1. *The inverse image has the following properties:*

- $g^{-1}(\mathbb{R}) = \mathbb{R}$
- *For any set* A*,* $g^{-1}(A^c) = g^{-1}(A)^c$
- *For any collection of sets* $\{A_\lambda; \lambda \in \Lambda\}$*,*

$$g^{-1}\left(\bigcup_\lambda A_\lambda\right) = \bigcup_\lambda g^{-1}(A).$$

As a consequence the mapping

$$A \mapsto P\{g(X) \in A\} = P\{X \in g^{-1}(A)\}$$

satisfies the axioms of a probability. The associated probability $\mu_{g(X)}$ is called the **distribution** of $g(X)$.

1 Discrete Random Variables

For X a discrete random variable with probabiliity mass function f_X, then the probability mass function f_Y for $Y = g(X)$ is easy to write.

$$f_Y(y) = \sum_{x \in g^{-1}(y)} f_X(x).$$

Example 2. *Let X be a uniform random variable on $\{1, 2, \ldots n\}$, i. e., $f_X(x) = 1/n$ for each x in the state space. Then $Y = X + a$ is a uniform random variable on $\{a + 1, 2, \ldots a + n\}$*

Example 3. *Let X be a uniform random variable on $\{-n, -n + 1, \ldots, n - 1, n\}$. Then $Y = |X|$ has mass function*

$$f_Y(y) = \begin{cases} \frac{1}{2n+1} & \text{if } x = 0, \\ \frac{2}{2n+1} & \text{if } x \neq 0. \end{cases}$$

2 Continuous Random Variable

The easiest case for transformations of continuous random variables is the case of g one-to-one. We first consider the case of g increasing on the range of the random variable X. In this case, g^{-1} is also an increasing function.

To compute the cumulative distribution of $Y = g(X)$ in terms of the cumulative distribution of X, note that

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \leq g^{-1}(y)\} = F_X(g^{-1}(y)).$$

Now use the chain rule to compute the density of Y

$$f_Y(y) = F_Y'(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y).$$

For g decreasing on the range of X,

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \geq g^{-1}(y)\} = 1 - F_X(g^{-1}(y)),$$

and the density

$$f_Y(y) = F_Y'(y) = -\frac{d}{dy} F_X(g^{-1}(y)) = -f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y).$$

For g decreasing, we also have g^{-1} decreasing and consequently the density of Y is indeed positive,

We can combine these two cases to obtain

$$\boxed{f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.}$$

Example 4. *Let U be a uniform random variable on $[0, 1]$ and let $g(u) = 1 - u$. Then $g^{-1}(v) = 1 - v$, and $V = 1 - U$ has density*

$$f_V(v) = f_U(1 - v)|-1| = 1$$

on the interval $[0, 1]$ and 0 otherwise.

Example 5. *Let X be a random variable that has a uniform density on $[0,1]$. Its density*

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{if } x > 1. \end{cases}$$

Let $g(x) = x^p$, $p \neq 0$. Then, the range of g is $[0,1]$ and $g^{-1}(y) = y^{1/p}$. If $p > 0$, then g is increasing and

$$\frac{d}{dy} g^{-1}(y) = \begin{cases} 0 & \text{if } y < 0, \\ \frac{1}{p} y^{1/p-1} & \text{if } 0 \le y \le 1, \\ 0 & \text{if } y > 1. \end{cases}$$

This density is unbounded near zero whenever $p > 1$.

If $p < 0$, then g is decreasing. Its range is $[1, \infty)$, and

$$\frac{d}{dy} g^{-1}(y) = \begin{cases} 0 & \text{if } y < 1, \\ -\frac{1}{p} y^{1/p-1} & \text{if } y \ge 1, \end{cases}$$

In this case, Y is a Pareto distribution with $\alpha = 1$ and $\beta = -1/p$. We can obtain a Pareto distribution with arbitrary α and β by taking

$$g(x) = \left(\frac{x}{\alpha}\right)^{1/\beta}.$$

If the transform g is not one-to-one then special care is necessary to find the density of $Y = g(X)$. For example if we take $g(x) = x^2$, then $g^{-1}(y) = \sqrt{y}$.

$$F_y(y) = P\{Y \le y\} = P\{X^2 \le y\} = P\{-\sqrt{y} \le X \le \sqrt{y}\} = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Thus,

$$\begin{aligned} f_Y(y) &= f_X(\sqrt{y}) \frac{d}{dy}(\sqrt{y}) - f_X(-\sqrt{y}) \frac{d}{dy}(-\sqrt{y}) \\ &= \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) \end{aligned}$$

If the density f_X is symmetric about the origin, then

$$f_y(y) = \frac{1}{\sqrt{y}} f_X(\sqrt{y}).$$

Example 6. *A random variable Z is called a* **standard normal** *if its density is*

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}).$$

A calculus exercise yields

$$\phi'(z) = -\frac{1}{\sqrt{2\pi}} z \exp(-\frac{z^2}{2}) = -z\phi(z), \qquad \phi''(z) = \frac{1}{\sqrt{2\pi}} (z^2 - 1) \exp(-\frac{z^2}{2}) = (z^2 - 1)\phi(z).$$

Thus, ϕ has a global maximum at $z = 0$, it is concave down if $|z| < 1$ and concave up for $|z| > 1$. This show that the graph of ϕ has a bell shape.

*$Y = Z^2$ is called a χ^2 (***chi-square***) random variable with one degree of freedom. Its density is*

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} \exp(-\frac{y}{2}).$$

3 The Probability Transform

Let X a continuous random variable whose distribution function F_X is strictly increasing on the possible values of X. Then F_X has an inverse function.

Let $U = F_X(X)$, then for $u \in [0,1]$,

$$P\{U \le u\} = P\{F_X(X) \le u\} = P\{U \le F_X^{-1}(u)\} = F_X(F_X^{-1}(u)) = u.$$

In other words, U is a uniform random variable on $[0,1]$. Most random number generators simulate independent copies of this random variable. Consequently, we can simulate independent random variables having distribution function F_X by simulating U, a uniform random variable on $[0,1]$, and then taking

$$X = F_X^{-1}(U).$$

Example 7. *Let X be uniform on the interval $[a,b]$, then*

$$F_X(x) = \begin{cases} 0 & \text{if } x < a, \\ \frac{x-a}{b-a} & \text{if } a \le x \le b, \\ 1 & \text{if } x > b. \end{cases}$$

Then

$$u = \frac{x-a}{b-a}, \quad (b-a)u + a = x = F_X^{-1}(u).$$

Example 8. *Let T be an exponential random variable. Thus,*

$$F_T(t) = \begin{cases} 0 & \text{if } t < 0, \\ 1 - \exp(-t/\beta) & \text{if } t \ge 0. \end{cases}$$

Then,

$$u = 1 - \exp(-t/\beta), \quad \exp(-t/\beta) = 1 - u, \quad t = -\frac{1}{\beta}\log(1-u).$$

Recall that if U is a uniform random variable on $[0,1]$, then so is $V = 1-U$. Thus if V is a uniform random variable on $[0,1]$, then

$$T = -\frac{1}{\beta}\log V$$

is a random variable with distribution function F_T.

Example 9. *Because*

$$\int_\alpha^x \frac{\beta\alpha^\beta}{t^{\beta+1}}\,dt = -\alpha^\beta t^{-\beta}\Big|_\alpha^x = 1 - \left(\frac{\alpha}{x}\right)^\beta.$$

A Pareto random variable X has distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x < \alpha, \\ 1 - \left(\frac{\alpha}{x}\right)^\beta & \text{if } x \ge \alpha. \end{cases}$$

Now,

$$u = 1 - \left(\frac{\alpha}{x}\right)^\beta \quad 1 - u = \left(\frac{\alpha}{x}\right)^\beta, \quad x = \frac{\alpha}{(1-u)^{1/\beta}}.$$

As before if $V = 1 - U$ is a uniform random variable on $[0, 1]$, then

$$X = \frac{\alpha}{V^{1/\beta}}$$

is a Pareto random variable with distribution function F_X.

$$f_Y(y) = f_X(x)\cdot\frac{dx}{dy} \qquad Y = g(X) \text{ (strictly monotone)}$$

written all x in terms of y

$$\frac{dx}{dy} = \left(\frac{dy}{dx}\right)^{-1} \quad \text{(chain rule)}$$

proof: $F_Y(y) = P(Y \le y) = P(g(X) \le y)$

$= P(X \le g^{-1}(y))$

$= F_X(g^{-1}(y)) = F_X(x)$

$\Rightarrow f_Y(y) = f_X(x)\frac{dx}{dy}$

for CDF, equivalently to do nothing

(remember $f_Y(y)\cdot dy = f_X(x)\cdot dx$)

Example log normal

$Y = e^Z$, $Z \sim N(0,1)$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\lg y)^2} \cdot \frac{dz}{dy}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\lg y)^2} \frac{1}{y}$$

multi-dimensional version

$\vec{Y} = g(\vec{X}) \quad g: R^n \to R^m$

joint PDF of $\vec{Y}$ is

$$f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}(\vec{x})\left|\frac{d\vec{x}}{d\vec{y}}\right| \qquad \text{Jacobian}$$

matrix of all possible partial derivatives

$$\left(\left|\frac{d\vec{x}}{d\vec{y}}\right| = \left|\frac{d\vec{y}}{d\vec{x}}\right|^{-1}\right)$$

Convolution (sums) Let $T = X+Y$, independent

discrete

$$P(T=t) = \sum_x P(X=x)P(Y=t-x)$$

Continuous

$$P(T=t) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(t-x) \cdot dx$$

$$\begin{aligned} F_T(t) &= P(T \leq t) \\ &= \int P(X+Y \leq t \mid X=t) f_X(x) \cdot dx \\ &= \int P(Y \leq t-X \mid X=t) f_X(x) \cdot dx \\ &= \int_{-\infty}^{\infty} F_Y(t-x) \cdot f_X(x) \cdot dx \end{aligned}$$

(see below interchange derivative and integral)

Idea: ① prove existence of object with desired property A using prob

Show $P(A) > 0$ for a random object

② Suppose each object has a number associated "score"

Show there is an object with "good" score.

there is an object with score is at least $E(X)$

Ex. 100 people, 5 commitees of 20, each person is on 3 commitees

problem. there exists 2 commitees with overlap ≥ 3

find average overlap of 2 commitees

$$\begin{aligned} E(\text{overlap}) &= 100 \cdot E(\text{person } i \text{ in both}) \\ &= 100 \times P \\ &= 100 \times \frac{\binom{3}{2}}{\binom{5}{2}} = \underbrace{\frac{20}{7}}_{\Downarrow} \end{aligned}$$

(one must round up to 3)

Supplementary Notes 3

Interchange of Differentiation and Integration

The theme of this course is about various limiting processes. We have learnt the limits of sequences of numbers and functions, continuity of functions, limits of difference quotients (derivatives), and even integrals are limits of Riemann sums. As often encountered in applications, exchangeability of limiting processes is an important topic. For example, we learnt

$$\frac{d}{dx}\int_a^x f = \int_a^x \frac{df}{dx}, \quad f(a) = 0,$$

whenever $\frac{df}{dx}$ is integrable; also

$$\lim_{n\to\infty} \frac{d}{dx} f_n(x) = \frac{d}{dx} \lim_{n\to\infty} f_n(x),$$

if $\{f_n\}$ and $\{f_n'\}$ converge uniformly.

Here we consider the following situation. Let $f(x, y)$ be a function defined in $[a, b] \times [c, d]$ and

$$\phi(y) = \int_a^b f(x, y)dx.$$

It is natural to ask if continuity and differentiability are preserved under integration.

Theorem 1. *Let $f(x, y)$ be continuous in $[a, b] \times [c, d]$. Then ϕ defined above is a continuous function on $[c, d]$.*

Proof. Since f is continuous in $[a, b]\times[c, d]$, it is bounded and uniformly continuous. In other words, for any $\varepsilon > 0$, $\exists\delta$ such that

$$\begin{aligned} |\phi(y) - \phi(y')| &\le \int_a^b |f(x, y) - f(x, y')|dx \\ &< \varepsilon(b - a) \qquad \forall y, |y - y'| < \delta, \end{aligned}$$

which shows that ϕ is uniformly continuous on $[c, d]$. $\square$

Theorem 2. *Let f and $\dfrac{\partial f}{\partial y}$ be continuous in $[a,b]\times[c,d]$. Then ϕ is differentiable and*

$$\frac{d}{dy}\phi(y)=\int_a^b \frac{\partial f}{\partial y}(x,y)dx$$

holds.

Proof. Fix $y\in(c,d)$, $y+h\in(c,d)$ for small $h\in\mathbb{R}$,

$$\begin{aligned}\frac{\phi(y+h)-\phi(y)}{h}&=\frac{1}{h}\int_a^b (f(x,y+h)-f(x,y))dx\\&=\int_a^b \frac{\partial f}{\partial y}(x,z)dx\end{aligned}$$

where z is a point between y and $y+h$ which depends on x . In any case,

$$\Big|\frac{\phi(y+h)-\phi(y)}{h}-\int_a^b \frac{\partial f}{\partial y}(x,y)dx\Big|\le\int_a^b \Big|\frac{\partial f}{\partial y}(x,z)-\frac{\partial f}{\partial y}(x,y)\Big|dx.$$

Since $\dfrac{\partial f}{\partial y}$ is uniformly continuous on $[a,b]\times[c,d]$, for $\varepsilon>0$, $\exists\delta$ such that

$$\Big|\frac{\partial f}{\partial y}(x,y')-\frac{\partial f}{\partial y}(x,y)\Big|<\varepsilon,\qquad \forall|y'-y|<\delta \text{ and } \forall x.$$

Taking $h\le\delta$, we get

$$\Big|\frac{\phi(y+h)-\phi(y)}{h}-\int_a^b \frac{\partial f}{\partial y}(x,y)dx\Big|<\varepsilon,$$

whence the condition follows.

When $y=c$ or d, the same proof works with some trivial changes. □

In many applications, the rectangle is replaced by an unbounded region. When this happens, we need to consider improper integrals. As a typical case, let's assume f is defined in $[a,\infty)\times[c,d]$ and set

$$\phi(y)=\int_a^\infty f(x,y)dx.$$

The function $\phi(y)$ makes sense if the improper integral $\int_a^\infty f(x)dx$ is well-defined for each y. Recall that this means

$$\lim_{b\to\infty}\int_a^b f(x,y)dx$$

exists. We introduce the following definition: The improper integral

$$\int_a^\infty f(x,y)dx$$

is uniformly convergent if $\forall\varepsilon$, $\exists b_0>0$ such that

$$\Big|\int_b^{b'} f(x,y)dx\Big|<\varepsilon, \qquad \forall b',b\ge b_0.$$

Notice that in particular, this implies that $\int_a^\infty f(x,y)dx$ exists for every y.

Uniform convergence of an improper integral may be studied parallel to the uniform convergence of infinite series. In fact, if we let

$$\phi_n(y)=\int_a^n f(x,y)dx,$$

it is not hard to see that the improper integral converges uniformly iff the infinite series $\sum_{n=n_0}^\infty \phi_n(y)$ converges uniformly when $f(x,y)\ge 0$. When f changes sign, the equivalence does not always hold. Nevertheless, techniques in establishing uniform convergence can be borrowed and applied to the present situation. As a sample, we have the following version of M-test, whose proof is omitted.

Theorem 3. *Suppose that $|f(x,y)|\le h(x)$ and h has an improper integral on $[a,\infty)$. Then $\int_a^\infty f(x,y)dx$ converges uniformly and absolutely.*

Theorem 4. *Let f be continuous in $[a,\infty)\times[c,d]$. Then ϕ is continuous in $[c,d]$ if the improper integral $\int_a^\infty f(x,y)dx$ converges uniformly.*

Proof. By Theorem 1, the function

$$\phi_n(y)=\int_a^n f(x,y)dx$$

is continuous on $[c, d]$ for every n. By assumption, $\forall \varepsilon > 0$, $\exists b_0$ such that

$$|\phi_n(y) - \phi_m(y)| = \Big| \int_n^m f(x, y) dx \Big| < \varepsilon, \qquad \forall n, m \geq b_0.$$

Hence $\{\phi_n\}$ is a Cauchy sequence in sup-norm. Since any Cauchy sequence in sup-norm converges, ϕ_n converges uniformly to some continuous function ψ. As ϕ_n converges pointwisely to ϕ, ϕ and ψ coincide, so ϕ is continuous. $\square$

Theorem 5. *Let f and $\dfrac{\partial f}{\partial y}$ be continuous in $[a, \infty) \times [c, d]$. Suppose that the improper integrals $\displaystyle\int_a^\infty f$ and $\displaystyle\int_a^\infty \frac{\partial f}{\partial y}$ are uniformly convergent. Then ϕ is differentiable, and*

$$\frac{d\phi}{dy}(x) = \int_a^\infty \frac{\partial f}{\partial y}(x, y) dy$$

holds.

Proof. Applying the mean-value theorem to $\phi_n - \phi_m$,

$$\phi_n(y) - \phi_m(y) - (\phi_n(y_0) - \phi_m(y_0)) = (y - y_0)(\phi_n'(z) - \phi_m'(z))$$

for some z between y and y_0. According to Theorem 2 and the uniform convergence of $\displaystyle\int_a^\infty \frac{\partial f}{\partial y}$,

$$|\phi_n'(z) - \phi_m'(z)| = \Big| \int_m^n \frac{\partial f}{\partial y}(x, y) dy \Big| \to 0$$

as $n, m \to \infty$. This shows that $\forall \varepsilon > 0$, $\exists b_0$ such that

$$\Big| \frac{\phi_n(y) - \phi_n(y_0)}{y - y_0} - \frac{\phi_m(y) - \phi(y_0)}{y - y_0} \Big| < \varepsilon, \qquad n, m \geq b_0.$$

Letting $m \to \infty$,

$$\Big| \frac{\phi_n(y) - \phi_n(y_0)}{y - y_0} - \frac{\phi(y) - \phi(y_0)}{y - y_0} \Big| \leq \varepsilon, \qquad \forall n \geq b_0.$$

By triangle inequality,

$$\begin{aligned} &\Big| \frac{\phi(y) - \phi(y_0)}{y - y_0} - \int_a^\infty \frac{\partial f}{\partial y}(x, y) dx \Big| \\ &\leq \Big| \frac{\phi(y) - \phi(y_0)}{y - y_0} - \frac{\phi_n(y) - \phi_n(y_0)}{y - y_0} \Big| + \Big| \frac{\phi_n(y) - \phi_n(y_0)}{y - y_0} - \int_a^n \frac{\partial f}{\partial y}(x, y) dx \Big| \\ &+ \Big| \int_a^n \frac{\partial f}{\partial y}(x, y) dx - \int_a^\infty \frac{\partial f}{\partial y}(x, y) dx \Big|. \end{aligned}$$

Fix a large $n \geq b_0$ such that

$$\Big| \int_n^\infty \frac{\partial f}{\partial y}(x,y)dx \Big| < \varepsilon$$

and, by Theorem 2, we can also find $\delta > 0$ such that

$$\Big| \frac{\phi_n(y) - \phi_n(y_0)}{y - y_0} - \int_a^n \frac{\partial f}{\partial y}(x,y)dx \Big| < \varepsilon \qquad |y - y_0| < \delta.$$

Putting things together, we conclude

$$\Big| \frac{\phi(y) - \phi(y_0)}{y - y_0} - \int_a^\infty \frac{\partial f}{\partial y}(x,y)dx \Big| \leq \varepsilon + \varepsilon + \varepsilon < 4\varepsilon.$$

$\square$

One may appreciate these results when considering its relevance in partial differential equations. Consider the Laplace equation

$$u_{xx} + u_{yy} = 0$$

on the disk $D = \{(x,y) : x^2 + y^2 < 1\}$. Expressed in polar coordinates, the equation becomes

$$u_{rr} + \frac{u_r}{r} + \frac{u_{\theta\theta}}{r^2} = 0, \qquad (r,\theta) \in [0,1) \times [0, 2\pi].$$

To solve this equation means to find a function $u = u(r,\theta)$ which satisfies this equation, and, moreover, u is periodic in θ for $r \in [0,1)$. This is because when returning to the rectangular coordinates, u is continuous in D.

We observe that the Laplace equation is rotationally invariant. More precisely, for any solution $u(r,\theta)$, the function $v(r,\theta) = u(r, \theta + \theta_0)$ is a solution for each θ_0. From linearity it follows that $\sum_{j=1}^n c_j u(r, \theta + \theta_j)$ is again a solution. In limit form, the function

$$\tilde{u}(r,\theta) = \int_0^{2\pi} g(\alpha) u(r, \theta + \alpha) d\alpha$$

should also be a solution for any continuous g. Indeed, define $f(r,\theta,\alpha) = g(\alpha)u(r,\theta + \alpha)$. The functions f, $\dfrac{\partial f}{\partial \theta}$, $\dfrac{\partial^2 f}{\partial \theta^2}$, $\dfrac{\partial f}{\partial r}$, $\dfrac{\partial^2 f}{\partial r^2}$, are continuous in $[0,d] \times [0, 2\pi]$, $d < 1$. It follows from Theorem 2 that $\tilde{u}$ is also harmonic. Noting that g is arbitrary, in this way we have found many many harmonic functions from a single one. In fact, taking the special harmonic function to be

$$u(r,\theta) = \frac{1}{1 - r\cos\theta + r^2},$$

one can show that every harmonic function in D which is continuous in $\{(x,y) : x^2 + y^2 \le 1\}$ arises in this way.

We shall prove a more sophisticated criterion for uniform convergence. Indeed, recall that the comparison test is only effective in proving absolute convergence of infinite series. We need Abel's and Dirichlet's criteria to handle the convergence of alternating series. Here the situation is similar. We shall establish a version of Abel's criterion. The following lemma, which is usually called the second mean value theorem, is an integral analog of the Abel's lemma.

Theorem 6. *Let f be integrable and g be non-negative, decreasing and continuous on $[a,b]$. Then there exists $c \in [a,b]$ such that*

$$\int_a^b fg = g(c)\int_a^b f.$$

Proof. Divide $[a,b]$ equally by the partition $a = x_0 < x_1 < \cdots < x_n = b$. We have

$$\int_a^b fg = \sum_{j=1}^n \int_{x_{j-1}}^{x_j} fg = \sum_{j=1}^n g(x_{j-1})\int_{x_{j-1}}^{x_j} f + \sum_{j=1}^n \int_{x_{j-1}}^{x_j} (g(x) - g(x_j))f(x)dx.$$

As g is continuous and f is bounded on $[a,b]$, the second term on RHS of this equation tends to 0 as $n \to \infty$. Writing $F(x) = \displaystyle\int_a^x f$, the second term

$$\begin{aligned}
\sum_{j=1}^n g(x_{j-1})\int_{x_{j-1}}^{x_j} f &= \sum_{j=1}^n g(x_{j-1})(F(x_j) - F(x_{j-1})) \\
&= \sum_{j=2}^{n+1} g(x_{j-2})F(x_{j-1}) - \sum_{j=1}^n g(x_{j-1})F(x_{j-1})) \\
&= g(x_{n-1})F(b) + \sum_{j=2}^n (g(x_{j-2}) - g(x_{j-1}))F(x_{j-1})) - g(a)F(a).
\end{aligned}$$

As g is decreasing,

$$\begin{aligned} g(a)m &= [g(x_{n-1}) + \sum_{j=2}^{n}(g(x_{j-2}) - g(x_{j-1}))]m \le \sum_{j=1}^{n} g(x_{j-1}) \int_{x_{j-1}}^{x_j} f \\ &\le [g(x_{n-1}) + \sum_{j=2}^{n}(g(x_{j-2}) - g(x_{j-1}))]M = g(a)M \end{aligned}$$

for $M = \sup F$ and $m = \inf F$. By mean-value theorem then there exists $\xi_n \in [a, b]$ such that

$$g(a) \int_a^{x_{in}} f = \sum_{j=1}^{n} g(x_{j-1}) \int_{x_{j-1}}^{x_j} f.$$

Taking $n \to \infty$, by passing to a convergent subsequence of $\{\xi_n\}$ we get

$$g(c) \int_a^b f = \int_a^b gf.$$

□

Now, we can prove the criterion of Abel.

Theorem 7. *Suppose that $\int_a^\infty f(x, y)dx$ converges uniformly for $y \in [c, d]$, and $g(x, y)$ is decreasing for each fixed y and is bounded. Then $\displaystyle\int_a^\infty f(x, y)g(x, y)dx$ converges uniformly on $[c, d]$.*

Proof. By Theorem 6, there exists $\xi \in [b, b']$ such that

$$\begin{aligned} \Big| \int_b^{b'} f(x, y)g(x, y)dx \Big| &= \Big| g(\xi, y) \int_b^{b'} f(x, y)dx \Big| \\ &< (\sup |g|)\varepsilon, \end{aligned}$$

for large b and b', the conclusion follows. □

The following application is of technical nature.

Let's evaluate the Dirichlet integral

$$I = \int_0^\infty \frac{\sin x}{x} dx.$$

The trick is to consider the integral

$$\varphi(y) = \int_0^\infty e^{-yx} \frac{\sin x}{x} dx, \qquad y \ge 0.$$

Letting

$$f(x,y) = \begin{cases} 1, & x = 0 \\ e^{-yx}\frac{sin x}{x}, & x \neq 0 \end{cases},$$

We showed that $\displaystyle\int_0^\infty \frac{\sin x}{x}dx$ converges, so it is uniformly convergent in y trivially. By Abel's criterion, $\displaystyle\int_0^\infty e^{-yx}\frac{\sin x}{x}dx$ converges uniformly. The y-derivative of f is given by

$$\frac{\partial f}{\partial y} = -e^{-yx}\sin x.$$

For each $y \geq \delta$, $\displaystyle\int_\delta^\infty \frac{\partial f}{\partial y}(x,y)dx$ is clearly uniformly convergent. By Theorem 2 , we conclude that

$$\begin{aligned}\varphi'(y) &= \int_0^\infty \frac{d}{dy}(e^{-yx}\frac{\sin x}{x})dx \\ &= -\int_0^\infty e^{-yx}\sin x dx \\ &= -\frac{1}{1+y^2},\end{aligned}$$

which holds for $y \geq \delta > 0$. By integration we get

$$\varphi(y) = -\tan^{-1} y + C.$$

As

$$|\varphi(y)| = \Big|\int_0^\infty e^{-yx}\frac{\sin x}{x}\Big| \leq \int_0^\infty e^{-yx} = \frac{1}{y} \to 0$$

as $y \to \infty$, $C = \tan^{-1}\infty = \dfrac{\pi}{2}$. So

$$I = \lim_{y\to 0^+}\varphi(y) = \frac{\pi}{2}.$$