Of course there are many functions $X$ which are not Riemann integrable, but are RVs w.r.t. Lebesgue measure, and thus have well defined expected values.

**Example:** Let $X(\omega) = \begin{cases} 0 & \text{if } \omega \in \mathbb{Q} \text{ (rational)} \\ 1 & \text{if } \omega \notin \mathbb{Q} \text{ (irrational)} \end{cases}$

i.e. $X = \mathbb{1}_{\mathbb{Q}^c}$ where $\mathbb{Q}^c = \text{irrational numbers}$

Then $X$ is not Riemann-integrable but we still have $E[X] = \int X \, dP = 1$ well-defined.

We don’t need to restrict attention to intervals over $[0,1]$. Let $X : \mathbb{R} \to [0, \infty)$ be a Borel-measurable function.

Then we can define $\int X \, d\lambda$ by

$$\int X \, d\lambda = \int_{-\infty}^{\infty} X(t) \lambda(dt) = \sum_{n \in \mathbb{Z}} \int_{nt}^{nt+1} X(t) P(dt) \leq E[Y_n]$$
Where \( Y_n(t) = X(n+t) \)

and the expectation is w.r.t. Lebesgue measure on \([0,1]\).

In other words:
- We integrate a non-negative function over \(\mathbb{R}\) by adding up its integral over each interval \([n, n+1)\).
- We represent \(\lambda\) (Lebesgue meas. on \(\mathbb{R}\)) as a sum of countably many copies of Lebesgue meas. on unit intervals.

---

Understanding the Integral - Expectation Connection

\(X\) is a RV on a probability space \((\Omega, \mathcal{A}, P)\)

\[
E[X] = \sum_{\omega \in \Omega} X(\omega) \, dP(\omega) = \int_{\Omega} X \, dP
\]

Think of \(P\) as the CDF of \(X\)
Theorem: Let $X$ be a RV on prob. space $(\Omega, \mathcal{F}, P)$. Then for each Borel-measurable function $g$ on $\mathbb{R}$, we have

$$E[g(X)] = \int g(x) \, d\mu_X(x) \quad \text{M&amp;W notation}$$

where $\mu_X$ is the "probability distribution" of $X$.

Recall: $\mu_X$ is defined by

$$\mu_X(B) = P(X \in B) \quad \text{for } B \in \mathcal{B} \text{ the Borel } \sigma\text{-algebra}$$

e.g. $X$ is an absolutely continuous RV if

$\exists$ non-neg. Borel-meas. function $f_X$ (PDF of $X$)

s.t. $\mu_X(B) = \int_B f_X(d\lambda) \quad \forall B \in \mathcal{B}$.

Alternate def:

$$E[g(X)] = \int g(x) \, dF(x)$$

where $F$ is CDF of $X$.

Given a distribution function $F$, there is a unique probability function $P_F$ s.t.

$$P_F((a,b]) = F(b) - F(a)$$

$$= \mu_X((a,b])$$

from notation above.
Inequalities and Convergence

(ref: §5 Rosenthal)

**Markov's Inequality**: If $X$ is a non-negative random variable, then for all $\alpha > 0$,

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}.$$

**PF**: Define a new RV $Z$ by

$$Z(w) = \begin{cases} \alpha, & X(w) \geq \alpha \\ 0, & X(w) < \alpha \end{cases}$$

Then $Z \leq X$ so it follows that $E[Z] \leq E[X]$ (order-preserving property). Compute

$$E[Z] = \alpha \cdot P(X \geq \alpha) + 0 \cdot P(X < \alpha) = \alpha P(X \geq \alpha)$$

$$\Rightarrow P(X \geq \alpha) = \frac{E[X]}{\alpha}.$$
Chebychev’s Inequality: Let $Y$ be an arbitrary RV with finite mean $\mu_Y$. Then for all $\alpha > 0$,

$$P(|Y - \mu_Y| \geq \alpha) \leq \frac{\text{Var}(Y)}{\alpha^2}$$

applies more generally.

**PF:** Set $X = (Y - \mu_Y)^2$. Then $X$ is a non-neg. RV.

By Markov’s Inequality,

$$P(|Y - \mu_Y| \geq \alpha) = P(X \geq \alpha^2) \leq \frac{E[X]}{\alpha^2} = \frac{\text{Var}(Y)}{\alpha^2}.$$

We will use these 2 inequalities extensively, including to prove the laws of large numbers (see below).

Two other sometimes useful inequalities:

**Cauchy–Schwarz Inequality:**

Let $X$ and $Y$ be RVs with $E[X^2] < \infty$ and $E[Y^2] < \infty$.

Then

$$E[|XY|] \leq \sqrt{E[X^2]} \sqrt{E[Y^2]}.$$

**PF:** Let $Z = \frac{|X|}{\sqrt{E[X^2]}}$ and $W = \frac{|Y|}{\sqrt{E[Y^2]}}$, so that

Since
\[
E[Z^2] = E\left( \frac{|X|^2}{E[X^2]} \right) = \frac{E[X^2]}{E[X^2]} = 1
\]

Then,
\[
0 \leq E[(Z-W)^2] = E[Z^2 + W^2 - 2ZW]
\]
\[
\]
\[
= 1 + 1 - 2E[ZW]
\]
\[
\Rightarrow 2E[ZW] \leq 2
\]
\[
\Rightarrow E[ZW] \leq 1
\]

Thus,
\[
E[ZW] = E\left[ \frac{|X|}{\sqrt{E[X^2]}} \cdot \frac{|Y|}{\sqrt{E[Y^2]}} \right] = \frac{E[|XY|]}{\sqrt{E[X^2]E[Y^2]}} \leq 1
\]
\[
\Rightarrow E[|XY|] \leq \sqrt{E[X^2]} \cdot E[Y^2]
\]

Jensen's Inequality:

Let \( X \) be a RV with finite mean and let \( \phi: \mathbb{R} \rightarrow \mathbb{R} \) be a convex function, i.e. a function s.t.
\[
\lambda \phi(x_1) + (1-\lambda) \phi(x_2) \geq \phi(\lambda x_1 + (1-\lambda)x_2) \quad \text{for} \quad x_1, x_2, \lambda \in \mathbb{R}
\]

Then \( E[\phi(X)] \geq \phi(E[X]) \), and \( 0 \leq \lambda \leq 1 \).
Pf: Since $\phi$ is convex, we can find a linear function $g(x) = ax + b$ which lies entirely below the graph of $\phi$ but which touches it at the point $x = E[X]$, i.e. s.t. $g(x) \leq \phi(x) \quad \forall x \in \mathbb{R}$ and $g(E[X]) = \phi(E[X])$.

Then $E[\phi(X)] \geq E[g(X)]$ \quad order-preserving of $E(\cdot)$

\[= E[aX + b]\]

\[= aE[X] + b\]

\[= g(E[X]) = \phi(E[X]).\]

Convex function: if line segment b/t any 2 points on $\phi$ lies above or on the graph of $\phi$. 
Convergence of Random Variables (ref: R §5.2)

If $Z$ and $Z_1, Z_2, \ldots$ are random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, what does it mean to say that $\{Z_n\}$ converges to $Z$ as $n \to \infty$?

Different notions of convergence:

- pointwise
- almost surely (with probability 1)
- in probability

Q. Which ones are strong, weak?

**def:** Pointwise convergence: $\lim_{n \to \infty} Z_n(\omega) = Z(\omega) \ \forall \omega.$

**def:** Convergence in distribution: $\lim_{n \to \infty} F_{X_n}(\omega) = F(\omega) \ \forall \omega$

Pointwise convergence of CDFs
def: convergence almost surely (a.s.): \( Z_n \xrightarrow{a.s.} Z \)

\[ P(\lim_{n \to \infty} Z_n = Z) = 1 \quad \text{(or} \quad P(\lim_{n \to \infty} |X_n - X| \leq \varepsilon) = 1) \quad \forall \varepsilon > 0 \]

i.e. \( P(\exists w \in \Omega : \lim_{n \to \infty} Z_n(w) = Z(w)) = 1 \)

**Lemma**: Let \( Z \) and \( Z_1, Z_2, \ldots \) be RVs. Suppose for each \( \varepsilon > 0 \), we have \( P(|Z_n - Z| \geq \varepsilon \text{ i.o.}) = 0 \).

Then \( P(Z_n \to Z) = 1 \), i.e. \( Z_n \xrightarrow{a.s.} Z \).

Combine this Lemma with Borel-Cantelli Lemmas:

**Corollary**: Let \( Z \) and \( Z_1, Z_2, \ldots \) be RVs. Suppose for each \( \varepsilon > 0 \), we have \( \sum_{n} P(|Z_n - Z| \geq \varepsilon) < \infty \).

Then \( P(Z_n \to Z) = 1 \), i.e. \( Z_n \xrightarrow{a.s.} Z \).

def: Convergence in Probability \( X_n \xrightarrow{P} X \)

If \( \forall \varepsilon > 0 \), \( \lim_{n \to \infty} P(|Z_n - Z| \geq \varepsilon) = 0 \).

**Note**: Convergence a.s. \( \Rightarrow \) convergence in probability
but converse NOT true.