This solution is provided by Xiao Luo.

1. To define the Riemann integral of a bounded function \( f \) on a finite interval \([a, b]\) (Here, we use Darboux integral with notations adapted from Wikipedia), we need first to define a partition \( P \) of \([a, b]\) which is a finite sequence of values \( x_i \) such that \( a \leq x_0 \leq x_1 \leq \cdots \leq x_n \leq b \). Then we have the corresponding upper Darboux sum \( U_{f,P} \) and lower Darboux sum \( u_{f,P} \) defined as follows:

\[
U_{f,P} = \sum_{i=1}^{n} M_i(x_i - x_{i-1}), \\
u_{f,P} = \sum_{i=1}^{n} m_i(x_i - x_{i-1}),
\]

where \( M_i = \sup_{x \in [x_{i-1}, x_i]} f(x) \), \( m_i = \inf_{x \in [x_{i-1}, x_i]} f(x) \). By the Darboux sums we can define the Darboux integrals

\[
U_f = \inf \{ U_{f,P} : P \text{ is a partition of } [a, b] \}, \\
u_f = \sup \{ u_{f,P} : P \text{ is a partition of } [a, b] \}.
\]

If \( U_f = u_f \), then the Riemann integral \( \int_a^b f(x) \, dx \) is defined as \( U_f = u_f \).

As for the definition of a Lebesgue integral of a measurable function \( f \) on some measurable set \( A \) under a measure \( \mu \), we shall start with an indicator function \( f = \mathbb{1}_B(x) \) where \( B \) is a measurable set. Then, the integral of \( f \) over \( A \) is defined as \( \int_A f(x) \, d\mu(x) = \mu(A \cap B) \), then by linear expansion, we can naturally define the Lebesgue integral of a simple function over \( A \) where a simple function is a linear sum of disjoint indicator functions. Now for any positive measurable function, we can define the Lebesgue integral of a measurable function over \( A \) as

\[
\int_A f(x) \, d\mu(x) = \sup \{ \int_A s \, d\mu(x) : s \leq f \text{ is a simple function} \}.
\]
Finally for any measurable function $f$, it can be split as the difference of two positive measurable function $f = f^+ - f^-$, so we can define the Lebesgue integral of $f$ over $A$ as

$$\int_A f(x) \, d\mu(x) = \int_A f^+(x) \, d\mu(x) - \int_A f^-(x) \, d\mu(x).$$

One condition that Riemann integrable implies Lebesgue integrable is that the domain of integration is a bounded interval and the integrand is also bounded over the domain of integration, or in other words, the Riemman integral is a proper Riemann integral.

2. Let

$$f_n(x) = \int_{-n}^{n} \frac{1}{\sqrt{2}} e^{-\frac{(y-x)^2}{2}} \, dy,$$

then there is no analytic expression for $f_n(x)$, however

$$\lim_{n \to \infty} f_n(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2}} e^{-\frac{(y-x)^2}{2}} \, dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2}} e^{-\frac{y^2}{2}} \, dy = \sqrt{\pi}.$$ 

3. We have the inequality $v(A) = \int_A f(x) \, d\mu(x) \leq \int_A |f(x)| \, d\mu(x)$. So to prove that $v(A) = 0$, we just need to show $\int_A |f(x)| \, d\mu(x) = 0$. For any $n > 0$, we have

$$\int_A |f(x)| \cdot 1_{\{|f(x)| \leq n\}} \, d\mu(x) \leq \int_A n \, d\mu(x) = n\mu(A) = 0.$$

Then by Fatou's Lemma,

$$\int_A |f(x)| \, d\mu(x) \leq \liminf_n \int_A |f(x)| 1_{\{|f(x)| \leq n\}} \, d\mu(x) = 0,$$

so $\int_A |f(x)| \, d\mu(x) = 0$ and hence $v(A) = 0$.

4. Assume that $\alpha_i > 0, i = 1, \cdots, n$ and $\sum_{i=1}^{n} \alpha_i = 1$. We shall prove by induction. For $k = 2$, the desired inequality holds by definition.
Suppose it also holds for \( k - 1 \), then
\[
g\left(\sum_{i=1}^{k} \alpha_i x_i\right) = g\left((1 - \alpha_n) \sum_{i=1}^{k-1} \frac{\alpha_i}{1 - \alpha_n} x_i + \alpha_n x_n\right)
\]
\[
\leq (1 - \alpha_n) g\left(\sum_{i=1}^{k-1} \frac{\alpha_i}{1 - \alpha_n} x_i\right) + \alpha_n g(x_n)
\]
\[
\leq (1 - \alpha_n) \sum_{i=1}^{k-1} \frac{\alpha_i}{1 - \alpha_n} g(x_i) + \alpha_n g(x_n)
\]
\[
\leq \sum_{i=1}^{k-1} \alpha_i g(x_i) + \alpha_n g(x_n)
\]
\[
= \sum_{i=1}^{k} \alpha_i g(x_i),
\]

hence the desired inequality holds.

5. Assume that \( \rho(x, \theta) \geq 0 \) almost surely. Fix \( \theta_0 \). By the definition of \( \liminf_{\theta \to \theta_0} E\rho(x, \theta) \), there exists a sequence of points \( \theta_n \to \theta_0 \) such that
\[
\lim_{n \to \infty} E\rho(x, \theta_n) = \liminf_{\theta \to \theta_0} E\rho(x, \theta).
\]

By Fatou’s Lemma and the lower semi-continuity of \( \rho(x, \theta) \) in \( \theta \), we know
\[
\liminf_{n \to \infty} E\rho(x, \theta_n) \geq \int \liminf_{n \to \infty} \rho(x, \theta_n) f(\zeta) \, d\zeta
\]
\[
= \int \rho(x, \theta_0) f(\zeta) \, d\zeta
\]
\[
= E\rho(x, \theta_0),
\]

hence
\[
E\rho(x, \theta_0) \leq \liminf_{n \to \infty} E\rho(x, \theta_n) = \liminf_{\theta \to \theta_0} E\rho(x, \theta).
\]

6. We need first to show that \( \inf_C h(\zeta) \) is finite. Suppose by contradiction that \( \inf_C h(\zeta) = -\infty \), then there exists a sequence of points \( \{x_n\} \in C \) such that \( \lim_n h(x_n) = -\infty \). Because \( C \) is compact, every infinite sequence in \( C \) has an accumulation point in \( C \). So suppose without loss of generality that \( \lim_n x_n = x' \in C \). Then by the lower semi-continuity
of $h$, we have $h(x') \leq \inf_n h(x_n) = -\infty$ which is impossible. Hence we have shown that $\inf_C h(x) = c > -\infty$. By similar argument as above, there exists a sequence of points $\{y_n\} \in C$ with limit $x_0 \in C$ such that $\lim_n h(x_n) = c$. Then by the lower semi-continuity of $h$, $h(x_0) \leq \lim_n h(x_n) = c$. Apparently $h(x_0) \geq c$, therefore $h(x_0) = c = \inf_C h(x)$.

7. Because $\frac{1}{n} \sum_{i=1}^{n} |X_i - \theta| = \frac{1}{n} \sum_{i=1}^{n} |X_i^\ast - \theta|$ where $X_i^\ast$ is the order statistic, we can assume without loss of generality that $X_1 \leq X_2 \cdots \leq X_n$. Let $f_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} |X_i - \theta|$, then

$$f(x) = \begin{cases} 
\sum_{i=1}^{n} (\theta - X_i) & \text{if } \theta \geq X_n \\
\sum_{i=k+1}^{n} (X_i - \theta) + \sum_{i=1}^{k} (\theta - X_i) & \text{if } X_k \leq \theta < X_{k+1} \\
\sum_{i=1}^{n} (X_i - \theta) & \text{if } \theta \leq X_1,
\end{cases}$$

i.e. $f$ is a linear function in each of the following intervals $(-\infty, X_1)$, $(X_1, X_2), \cdots, (X_{n-1}, X_n), (X_n, \infty)$, thus the minimum of $f$ can only be possibly obtained at the end points $\{X_1, \cdots, X_n\}$ of those intervals. Then by comparing the values of $f$ at two different end points, we have

$$f(X_{k+1}) - f(X_k) = (\sum_{i=k+2}^{n} (X_i - X_{k+1}) + \sum_{i=1}^{k+1} (X_{k+1} - X_i))$$

$$- (\sum_{i=k+1}^{n} (X_i - X_k) + \sum_{i=1}^{k} (X_k - X_i))$$

$$= \sum_{i=k+1}^{n} (X_k - X_{k+1}) + \sum_{i=1}^{k} (X_{k+1} - X_k)$$

$$= (2k - n)(X_{k+1} - X_k).$$

Hence, $f(X_{k+1}) \geq f(X_k)$ if $k \leq \frac{n}{2}$ and $f(X_{k+1}) \leq f(X_k)$ if $k > \frac{n}{2}$. Therefore, for $n$ being odd, $f(X_1) \leq \cdots \leq f(X_{\frac{n+1}{2}}) \leq f(X_{\frac{n+3}{2}})$ and $f(X_{\frac{n+1}{2}}) \geq f(X_{\frac{n+3}{2}}) \geq \cdots \geq f(X_n)$, so $f(x)$ is minimized at $X_{\frac{n+1}{2}}$, the median value of $\{X_i\}$. For the case $n$ being even, similar argument shows that $f(x)$ is minimized at either $X_{\frac{n+1}{2}}$ or $X_{\frac{n}{2}}$. In conclusion, we have shown that $f(x)$ is minimized at $\text{median}(X_1, \cdots, X_n)$, a median value of $\{X_i\}$.

VERIFYING HUBER’S ASSUMPTIONS of consistent estimators. Note we need to assume that $\mathbb{E}|X_1|$ is finite.
(1). From the previous arguments, we know that median\((X_1, \ldots, X_n)\) minimizes the function \(\frac{1}{n} \sum_{i=1}^{n} \rho(\theta, X_i)\), hence assumption (1) of Huber’s theorem holds.

(2). The estimator median\((X_1, \ldots, X_n)\) is measurable because it equals the middle one of the order statistics \(\{X_{(1)}, \ldots, X_{(n)}\}\) that are all measurable functions.

(3). The distance function \(\rho\) is continuous and is thus lower semi-continuous. Moreover, \(E\rho(\theta, X_1) \leq E|X_1| + \theta < \infty\) as long as \(E|X_1| < \infty\). So assumption (3) is met.

(4). For assumption (4), let \(C = [\theta_* - c_1, \theta_* + c_2]\) such that there exists a \(\delta\) such that \(0 < \delta < \frac{1}{2}\) and \(F(\theta_* - c_1) = \frac{1}{2} - \delta, F(\theta_* + c_2) = \frac{1}{2} + \delta\) where \(F\) is the cumulative distribution function for \(X_1\). Then \(C\) is a compact set and we shall show that median\((X_1, \ldots, X_n) = \theta_*\) will eventually stays in \(C\) with probability 1. To achieve this, we want to show that \(\lim_{n \to \infty} P(\text{median}(X_1, \ldots, X_n) \notin C) = 0\). Now let \(f\) be the density function of \(X\) (Discrete case can be similarly handled.), and suppose for the moment that \(n = 2m + 1\) is odd, then we know median\((X_1, \ldots, X_n) = X_{(m+1)}\) with density function \(\frac{(2m+1)!}{m!m!} f(x) F(x)^m (1 - F(x))^m\), so we have

\[
P(\text{median}(X_1, \ldots, X_n) \notin C) = \int_{x \notin [\theta_* - c_1, \theta_* + c_2]} \frac{(2m+1)!}{m!m!} f(x) F(x)^m (1 - F(x))^m \, dx
\]

\[
= \int_{y \in [0,1], y \notin [\frac{1}{2} - \delta, \frac{1}{2} + \delta]} \frac{(2m+1)!}{m!m!} y^m (1 - y)^m \, dy.
\]

For \(y \notin [\frac{1}{2} - \delta, \frac{1}{2} + \delta]\) and \(y \in [0,1]\), the supremum of \(y(1-y)\) is \(\frac{1}{4} - \delta^2\). And by Stirling’s formula that \(n! \sim \sqrt{2\pi n^{n+\frac{1}{2}} e^{-n}}\), we can get \(\frac{(2m+1)!}{m!m!} \sim \frac{m}{\sqrt{\pi}} 4^m\), hence

\[
P(\text{median}(X_1, \ldots, X_n) \notin C) = \int_{y \in [0,1], y \notin [\frac{1}{2} - \delta, \frac{1}{2} + \delta]} \frac{(2m+1)!}{m!m!} y^m (1 - y)^m \, dy
\]

\[
\leq \int_{y \in [0,1], y \notin [\frac{1}{2} - \delta, \frac{1}{2} + \delta]} \frac{m}{\sqrt{\pi}} 4^m (\frac{1}{4} - \delta^2)^m \, dy
\]

\[
= \int_{y \in [0,1], y \notin [\frac{1}{2} - \delta, \frac{1}{2} + \delta]} \frac{m}{\sqrt{\pi}} (1 - 4\delta^2)^m \, dy
\]

\[
= (1 - 2\delta)^m \frac{m}{\sqrt{\pi}} (1 - 4\delta^2)^m.
\]

So by letting \(n \to \infty\) hence \(m \to \infty\), we have \((1 - 2\delta)^m \frac{m}{\sqrt{\pi}} (1 - 4\delta^2)^m \to 0\) and hence \(P(\text{median}(X_1, \ldots, X_n) \notin C)\) goes to 0 as
well. Similarly we can also show that when $n$ is even and $n$ goes to infinity, $P(\text{median}(X_1, \cdots, X_n) \notin C)$ goes to 0. So we have proved that with probability 1, and the estimator $\text{median}(X_1, \cdots, X_n)$ will finally stay in $C$.

(5). For this last assumption, we need to show $\gamma(\theta, X)$ is minimized at $\theta = \theta_*$. By the definition of $\theta_*$, we have $P(X \leq \theta_*) \geq \frac{1}{2}$ and $P(X \geq \theta_*) \geq \frac{1}{2}$, hence we have

$$P(X \leq \theta_*) - P(X > \theta_*) \geq 0,$$

$$P(X \geq \theta_*) - P(X < \theta_*) \geq 0.$$

For any $\theta < \theta_*$, we have

$$E|X - \theta| - E|X - \theta_*|$$

$$= \int_{x \geq \theta_*} [(x - \theta) - (x - \theta_*)] dP(x) + \int_{x \leq \theta} [(\theta - x) - (\theta_* - x)] dP(x)$$

$$+ \int_{\theta < x < \theta_*} [(x - \theta) - (\theta_* - x)] dP(x)$$

$$= (\theta - \theta_*)[P(X \leq \theta) - P(X \geq \theta_*)] + \int_{\theta < x < \theta_*} (2x - \theta - \theta_*) dP(x)$$

$$= (\theta - \theta_*)[P(X < \theta_*) - P(X \geq \theta_*)] + 2 \int_{\theta < x < \theta_*} (x - \theta) dP(x)$$

$$\geq 2 \int_{\theta < x < \theta_*} (x - \theta) dP(x) > 0.$$

Similarly for $\theta > \theta_*$, we have

$$E|X - \theta| - E|X - \theta_*|$$

$$= \int_{x \geq \theta} [(x - \theta) - (x - \theta_*)] dP(x) + \int_{x \leq \theta_*} [(\theta - x) - (\theta_* - x)] dP(x)$$

$$+ \int_{\theta_* < x < \theta} [(\theta - x) - (x - \theta_*)] dP(x)$$

$$= (\theta - \theta_*)[P(X \leq \theta_*) - P(X \geq \theta)] + \int_{\theta_* < x < \theta} (\theta + \theta_* - 2x) dP(x)$$

$$= (\theta - \theta_*)[P(X \leq \theta_*) - P(X > \theta_*)] + 2 \int_{\theta_* < x < \theta} (\theta - x) dP(x)$$

$$\geq 2 \int_{\theta_* < x < \theta} (\theta - x) dP(x) > 0.$$

So we have shown that $\theta_*$ is the unique point to minimize $\gamma(\theta)$. And we have shown that the assumptions of Huber’s consistence theorem are met.

The conclusion from Huber’s consistence theorem is that the estimator $\text{median}(X_1, \cdots, X_n)$ is a consistent estimator for $\theta_*$. 