Math 563 - Fall 15 - Homework 4

5. (Resnick) Let X be a random variable. A number m is called a median of X if $P(X \geq m) \geq 1/2$ and $P(X \leq m) \geq 1/2$. You should convince yourself that there is always at least one median, but it need not be unique. Recall that the mean of X is $\mu = E[X]$ if X is integrable.

(b) Let m be a median of X. Prove that the minimum of $g(a) = E[|X - a|]$ occurs at m.

Solution: Define

$$I_- = \{t : P(X \leq t) \geq \frac{1}{2}\}, \quad I_+ = \{t : P(X \geq t) \geq \frac{1}{2}\}$$

and let $m_- = \inf I_-$ and $m_+ = \sup I_+$. Note that $P(X \leq t)$ is increasing in t and $P(X \geq t)$ is decreasing in t. This and the continuity of P implies $I_- = [m_-, \infty)$ and $I_+ = (-\infty, m_+]$. Every t belongs to at least one of I_- or I_+ since $P(X \leq t) + P(X \geq t) \geq 1$. So $m_- \leq m_+$. The medians are all the values in the interval $[m_-, m_+]$. (This interval could just be a single point.)

We claim that

$$E|X - c| = \int_c^\infty P(X \geq t)\, dt + \int_{-\infty}^c P(X \leq t)\, dt \tag{1}$$

We demonstrate this as follows. By the transformation theorem

$$\begin{aligned} E|X - c| &= \int_{-\infty}^\infty |x - c|\, d\mu_X \\ &= \int_{-\infty}^\infty [1_{x>c} \int_c^x dt + 1_{x<c} \int_x^c dt]\, d\mu_X \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty [1_{c \leq t \leq x} + 1_{x \leq t \leq c}] dt\, d\mu_X \end{aligned}$$

Using Fubini's theorem

$$\begin{aligned} &= \int_{-\infty}^\infty \int_{-\infty}^\infty [1_{c \leq t \leq x} + 1_{x \leq t \leq c}] d\mu_X\, dt \\ &= \int_{-\infty}^\infty [1_{c \leq t} P(X \geq t) + 1_{t \leq c} P(X \leq t)]\, d\mu_X \\ &= \int_c^\infty P(X \geq t)\, dt + \int_{-\infty}^c P(X \leq t)\, dt \end{aligned}$$

Now let m be a median. Using (1) we have for $c > m$

$$E|X - c| - E|X - m| = \int_m^c [P(X \leq t) - P(X \geq t)]dt$$

and if $c < m$ we have

$$E|X - c| - E|X - m| = \int_c^m [P(X \geq t) - P(X \leq t)]dt$$

We will complete the proof by showing that $P(X \leq t) - P(X \geq t)$ is 0 on (m_-, m_+), is positive on (m_+, ∞), and is negative on $(-\infty, m_-)$.

If $t > m_+$ then $P(X \geq t) < \frac{1}{2}$ and $P(X \leq t) \geq \frac{1}{2}$.
So $P(X \leq t) - P(X \geq t) > 0$.

If $t < m_-$ then $P(X \geq t) \geq \frac{1}{2}$ and $P(X \leq t) < \frac{1}{2}$.
So $P(X \leq t) - P(X \geq t) < 0$.

If $m_- < m_+$ then the events $X \leq m_-$ and $X \geq m_+$ are disjoint. So the sum of their probabilities is at most 1. Since $P(X \leq m_-) \geq \frac{1}{2}$ and $P(X \geq m_+) \geq \frac{1}{2}$, it must be that $P(X \leq m_-) = P(X \geq m_+) = \frac{1}{2}$ and $P(m_- < X < m_+) = 0$. So for $m_- < t < m_+$ we have
$P(X \leq t) - P(X \geq t) = 0$. Note that this quantity may not be zero at m_- and m_+.