Jensen's for Special Polynomials

Jensen's inequality says that for any convex function f, $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$ (You don't need to know this for CS 70). In this problem, we will prove that Jensen's inequality holds for a subclass of convex functions called "special polynomials" (this is a made up name). We define a special polynomial as any function that can be written as

$$f(x) = a_n x^{(2^{n-1})} + a_{n-1} x^{(2^{n-2})} + \ldots + a_1 x$$

for some $n \in \mathbb{N}$ and $\forall 1 \leq i \leq n$, $a_i \geq 0$

(a) Prove that $\mathbb{E}[X^2] \geq E[X]^2$ (*Hint: use the definition of variance*)

Solution: Recall from the definition of Variance that $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, implying that

$$\mathbb{E}[X^2] = Var(X) + \mathbb{E}[X]^2 \tag{1}$$

Now let $Y = (X - \mu)^2$. Then, $Var(X) = E(Y)$. Since Y is a non-negative random variable (a square of a real-valued function can never be negative), we must have that $E(Y) = Var(X) \geq 0$, since the weighted average of non-negative values must be non-negative. Thus,

$$\mathbb{E}[X^2] \geq \mathbb{E}[X]^2 \tag{2}$$

(b) Use part (a) to prove that $\mathbb{E}[X^{(2^k)}] \geq E[X]^{(2^k)}$ for some $k \in \mathbb{N}$

Solution: From part (a), we have proven that the statements holds when $k = 1$. We use induction to prove it holds for arbitrary k.

Suppose the statement holds for $k - 1$. Then,

$$\begin{aligned}
\mathbb{E}[X^{(2^k)}] &= \mathbb{E}[(X^{(2^{k-1})})^2] && (3)\\
&\geq \mathbb{E}[X^{(2^{k-1})}]^2 && (4)\\
&\geq (\mathbb{E}[X]^{2^{k-1}})^2 && (5)\\
&= \mathbb{E}[X]^{(2^k)} && (6)
\end{aligned}$$

A lot happened here. Let's break it down line by line.

On line 3, we rearranged exponents to isolate $k - 1$.

Then, from line 3 to line 4, we applied the inequality in part (a).

From line 4 to line 5, we applied the inductive hypothesis, and finally we rearranged exponents from line 5 to line 6. This completes the proof!

(c) Use part (b) and properties of expectation to prove that $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$

Solution: Most of the heavy lifting is done. We now apply the function definition from the problem statement and use linearity of expectation to isolate each term. Then, we use the inequality from part (b) on each term

$$\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[a_n X^{(2^{n-1})} + a_{n-1} x^{(2^{n-2})} + \ldots + a_1 x] && (7)\\
&= \mathbb{E}[a_n X^{(2^{n-1})}] + \mathbb{E}[a_{n-1} X^{(2^{n-2})}] + \ldots + \mathbb{E}[a_1 X] && (8)\\
&= a_n \mathbb{E}[X^{(2^{n-1})}] + a_{n-1} \mathbb{E}[X^{(2^{n-2})}] + \ldots + a_1 \mathbb{E}[X] && (9)\\
&\geq a_n \mathbb{E}[X]^{(2^{n-1})} + a_{n-1} \mathbb{E}[X]^{(2^{n-2})} + \ldots + a_1 \mathbb{E}[X] && (10)\\
&= f(\mathbb{E}[X]) && (11)
\end{aligned}$$

Boom!