556: MATHEMATICAL STATISTICS I

INEQUALTIES

1. CONCENTRATION INEQUALITIES

LEMMA (CHEBYCHEV'S LEMMA)
If X is a random variable, then for non-negative function h, and $c > 0$,

$$P[h(X) \geq c] \leq \frac{E_{f_X}[h(X)]}{c}$$

Proof. (continuous case) : Suppose that X has density function f_X which is positive for $x \in \mathbb{X}$. Let $\mathcal{A} = \{x \in \mathbb{X} : h(x) \geq c\} \subseteq X$. Then, as $h(x) \geq c$ on A,

$$\begin{aligned} E_{f_X}[h(X)] = \int h(x) f_X(x)\,dx &= \int_{\mathcal{A}} h(x) f_X(x)\,dx + \int_{\mathcal{A}'} h(x) f_X(x)\,dx \\ &\geq \int_{\mathcal{A}} h(x) f_X(x)\,dx \\ &\geq \int_{\mathcal{A}} c f_X(x)\,dx \\ &= cP[X \in \mathcal{A}] = cP[h(X) \geq c] \end{aligned}$$

and the result follows. ∎

- **SPECIAL CASE I - THE MARKOV INEQUALITY**
  If $h(x) = |x|^r$ for $r > 0$, so
  $$P[|X|^r \geq c] \leq \frac{E_{f_X}[|X|^r]}{c}.$$
  Alternately stated (by Casella and Berger) as follows: If $P[Y \geq 0] = 1$ and $P[Y = 0] < 1$, then for any $r > 0$
  $$P[Y \geq r] \leq \frac{E_{f_X}[Y]}{r}$$
  with equality if and only if
  $$P[Y = r] = p = 1 - P[Y = 0]$$
  for some $0 < p \leq 1$.

- **SPECIAL CASE II - THE CHEBYCHEV INEQUALITY**
  Suppose that X is a random variable with expectation μ and variance σ^2. Then $h(x) = (x - \mu)^2$ and $c = k^2\sigma^2$, for $k > 0$,
  $$P\left[(X - \mu)^2 \geq k^2\sigma^2\right] \leq 1/k^2$$
  or equivalently
  $$P[|X - \mu| \geq k\sigma] \leq 1/k^2.$$
  Setting $\epsilon = k\sigma$ gives
  $$P[|X - \mu| \geq \epsilon] \leq \sigma^2/\epsilon^2$$
  or equivalently
  $$P[|X - \mu| < \epsilon] \geq 1 - \sigma^2/\epsilon^2.$$

CHERNOFF BOUNDS

THEOREM
Suppose that $X_1, \ldots, X_n$ are independent binary trials (known as "Poisson trials") such that

$$P[X_i = x] = \begin{cases} 1 - p_i & x = 0 \\ p_i & x = 1 \end{cases}$$

and zero otherwise. Let $X = (X_1 + \cdots + X_n)$, so that

$$E_{f_X}[X] = \sum_{i=1}^{n} p_i = \mu$$

say. Then for $d > 0$

$$P[X \geq (1+d)\mu] \leq \exp\left\{ \frac{e^d}{(1+d)^{(1+d)}} \right\}^{\mu}$$

and for $0 \leq d \leq 1$

$$P[X \geq (1+d)\mu] \leq \exp\left\{ -\mu d^2/3 \right\}$$

Proof. Let $a > 0$. Then

$$\begin{aligned} P[X \geq (1+d)\mu] &= P[\exp\{aX\} \geq \exp\{a(1+d)\mu\}] \\ &\leq E_{f_X}[\exp\{aX\}] \exp\{-a(1+d)\mu\} \end{aligned}$$

using the previous Chebychev Lemma with $h(x) = e^{ax}$ and $c = e^{a(1+d)\mu}$. But

$$E_{f_X}[\exp\{aX\}] = \prod_{i=1}^{n} E_{f_{X_i}}[\exp\{aX_i\}] = \prod_{i=1}^{n} [p_i e^a + (1 - p_i)] = \prod_{i=1}^{n} [1 + p_i(e^a - 1)]$$

Now for $y > 0$, $1 + y < e^y$, so setting $y = p_i(e^a - 1)$, we conclude that

$$E_{f_X}[\exp\{aX\}] < \prod_{i=1}^{n} \exp\{p_i(e^a - 1)\} = \exp\left\{ \sum_{i=1}^{n} p_i(e^a - 1) \right\} = \exp\left\{ \mu(e^a - 1) \right\}$$

Hence

$$P[X \geq (1+d)\mu] \leq \exp\left\{ \mu(e^a - 1) \right\} \exp\{-a(1+d)\mu\}$$

and setting $a = \log(1+d)$ yields

$$P[X \geq (1+d)\mu] \leq \left\{ \frac{e^{\mu d}}{(1+d)^{\mu(1+d)}} \right\} = \left\{ \frac{e^d}{(1+d)^{(1+d)}} \right\}^{\mu}$$

Now, for $0 \leq d \leq 1$, the right hand side is bounded above by $\exp\{-\mu d^2/3\}$. To see this, consider (after taking logs),

$$f(d) = d - (1+d)\log(1+d) + d^2/3.$$

We need to show that $f(d)$ is bounded above by zero for $0 \leq d \leq 1$. Now, clearly $f(0) = 0$, and taking derivatives twice we have

$$\begin{aligned} f^{(1)}(d) &= -\log(1+d) + 2d/3 \\ f^{(2)}(d) &= -\frac{1}{(1+d)} + 2/3 \end{aligned}$$

so $f^{(1)}(0) = 0$ and $f^{(1)}(d)$ is **negative** for all $0 \leq d \leq 1$. Thus $f(d)$ must be negative for all d in this range. ∎

NOTE : In fact, for any integer $k \geq 2$, the bound for $0 \leq d \leq 1$

$$P[X \geq (1+d)\mu] \leq \exp\left\{-\mu d^k/3\right\}$$

holds, but the bound is **tighter** if k is **smaller**. The bound does **not** hold if $k = 1$. To see this, consider again

$$f_k(d) = d - (1+d)\log(1+d) + d^k/3.$$

and

$$\begin{aligned} f_k^{(1)}(d) &= -\log(1+d) + kd^{k-1}/3 \\ f_k^{(2)}(d) &= -\frac{1}{(1+d)} + k(k-1)d^{k-2}/3 \end{aligned}$$

Now $f_k(0) = 0$ and $f_k(1) = 1 - 2\log 2 + 1/3 < 0$, and as there is only one solution of

$$\log(1+x) = kx^{k-1}/3$$

on $0 < x < 1$, there is precisely **one** turning point of $f(d)$ on this interval. Thus $f_k(d)$ never becomes positive on $(0,1)$.

See also the graph below of the function $f_k(d)$ for $k = 1, 2, 3, 4, 5$.

LEMMA (A CHERNOFF BOUND USING MGFS)
If X is a random variable, with mgf $M_X(t)$ defined on a neighbourhood $(-h, h)$ of zero. Then

$$P[X \geq a] \leq e^{-at} M_X(t) \qquad \text{for } 0 < t < h$$

Proof. Using the Chebychev Lemma with $h(x) = e^{tx}$ and $c = e^{at}$, for $t > 0$,

$$P[X \geq a] = P[tX \geq at] = P[\exp\{tX\} \geq \exp\{at\}] \leq \frac{E_{f_X}[e^{tX}]}{e^{at}} = \frac{M_X(t)}{e^{at}}$$

provided $t < h$ also. Using similar methods,

$$P[X \leq a] \leq e^{-at} M_X(t) \qquad \text{for } -h < t < 0$$

∎

THEOREM Tail bounds for the Normal density
If $Z \sim N(0,1)$, then for $t > 0$

$$P[|Z| \geq t] \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}$$

Proof.

$$P[Z \geq t] = \left(\frac{1}{2\pi}\right)^{1/2} \int_t^\infty e^{-x^2/2}\, dx \leq \left(\frac{1}{2\pi}\right)^{1/2} \int_t^\infty \frac{x}{t} e^{-x^2/2}\, dx = \left(\frac{1}{2\pi}\right)^{1/2} \frac{e^{-t^2/2}}{t}$$

and by symmetry $P[|Z| \geq t] = 2P[Z \geq t]$.

Note: Using similar methods

$$P[|Z| \geq t] \geq \sqrt{\frac{2}{\pi}} \frac{te^{-t^2/2}}{1+t^2}$$

yielding a lower bound on this probability. ∎

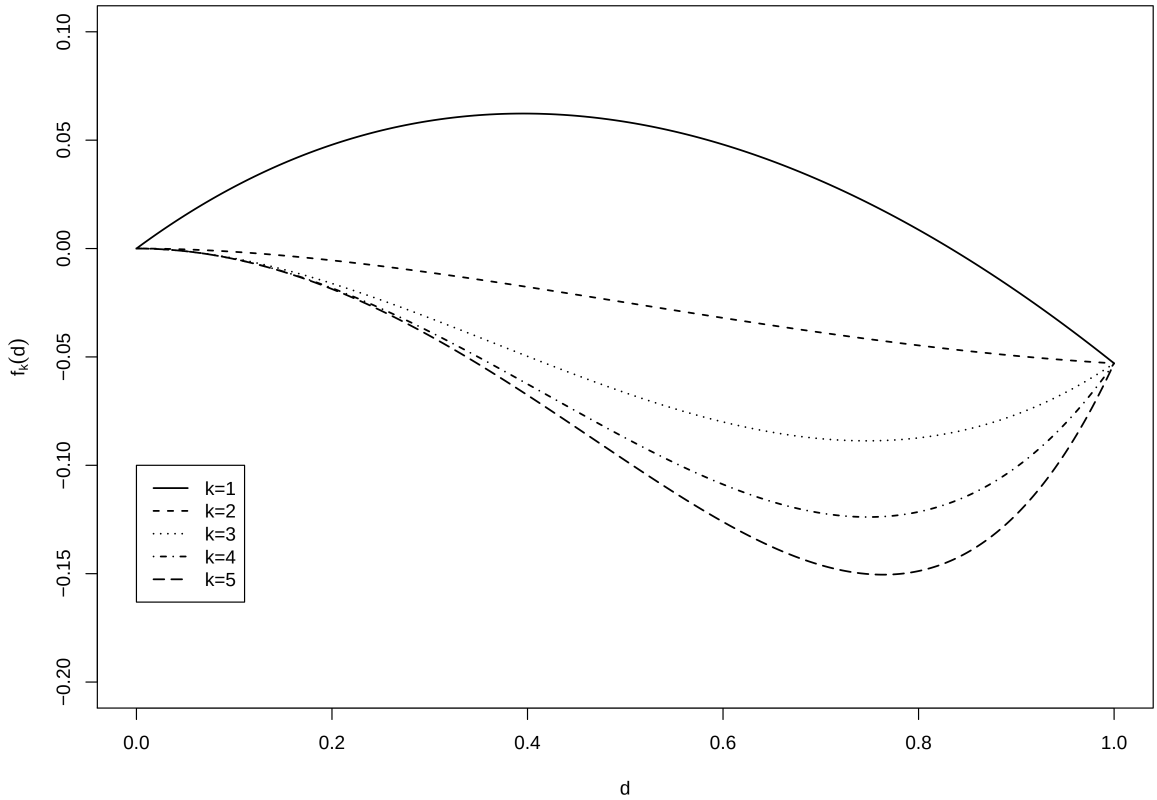

Figure 1: The function $f_k(d) = d - (1+d)\log(1+d) + d^k/3$ for $k = 1, 2, 3, 4, 5$. The function is negative on $0 < d < 1$ for each $k \geq 2$.

2. INEQUALITIES FOR MULTIPLE RANDOM VARIABLES

LEMMA
Let $a, b > 0$ and $p, q > 1$ satisfy

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{1}$$

Then

$$\frac{1}{p}\, a^p + \frac{1}{q}\, b^q \geq ab$$

with equality if and only if $a^p = b^q$.

Proof. Fix $b > 0$. Let

$$g(a;b) = \frac{1}{p}\, a^p + \frac{1}{q}\, b^q - ab.$$

We require that $g(a;b) \geq 0$ for all a. Differentiating wrt a for fixed b yields

$$g^{(1)}(a;b) = a^{p-1} - b$$

so that $g(a;b)$ is minimized (the second derivative is strictly positive at all a) when $a^{p-1} = b$, and at this value of a, the function takes the value

$$\frac{1}{p}\, a^p + \frac{1}{q}\, (a^{p-1})^q - a(a^{p-1}) = \frac{1}{p}\, a^p + \frac{1}{q}\, a^p - a^p = 0$$

as, by equation (1), $1/p + 1/q = 1 \implies (p-1)q = p$. As the second derivative is strictly positive at all a, the minimum is attained at the **unique** value of a where $a^{p-1} = b$, where, raising both sides to power q yields $a^p = b^q$. ∎

THEOREM (HÖLDER'S INEQUALITY)
Suppose that X and Y are two random variables, and $p, q > 1$ satisfy 1. Then

$$|E_{f_{X,Y}}[XY]| \leq E_{f_{X,Y}}[|XY|] \leq \{E_{f_X}[|X|^p]\}^{1/p} \{E_{f_Y}[|Y|^q]\}^{1/q}$$

Proof. (continuous case) For the first inequality,

$$E_{f_{X,Y}}[|XY|] = \iint |xy| f_{X,Y}(x,y)\ dxdy \geq \iint xy f_{X,Y}(x,y)\ dxdy = E_{f_{X,Y}}[XY]$$

and

$$E_{f_{X,Y}}[XY] = \iint xy f_{X,Y}(x,y)\ dxdy \geq \iint -|xy| f_{X,Y}(x,y)\ dxdy = -E_{f_{X,Y}}[|XY|]$$

so

$$-E_{f_{X,Y}}[|XY|] \leq E_{f_{X,Y}}[XY] \leq E_{f_{X,Y}}[|XY|] \qquad \therefore \qquad |E_{f_{X,Y}}[XY]| \leq E_{f_{X,Y}}[|XY|].$$

For the second inequality, set

$$a = \frac{|X|}{\{E_{f_X}[|X|^p]\}^{1/p}} \qquad b = \frac{|Y|}{\{E_{f_Y}[|Y|^q]\}^{1/q}}.$$

Then from the previous lemma

$$\frac{1}{p}\frac{|X|^p}{E_{f_X}[|X|^p]} + \frac{1}{q}\frac{|Y|^q}{E_{f_Y}[|Y|^q]} \geq \frac{|XY|}{\{E_{f_X}[|X|^p]\}^{1/p}\{E_{f_Y}[|Y|^q]\}^{1/q}}$$

and taking expectations yields, on the left hand side,

$$\frac{1}{p}\frac{E_{f_X}[|X|^p]}{E_{f_X}[|X|^p]} + \frac{1}{q}\frac{E_{f_Y}[|Y|^q]}{E_{f_Y}[|Y|^q]} = \frac{1}{p} + \frac{1}{q} = 1$$

and on the right hand side

$$\frac{E_{f_{X,Y}}[|XY|]}{\{E_{f_X}[|X|^p]\}^{1/p}\{E_{f_Y}[|Y|^q]\}^{1/q}}$$

and the result follows. ∎

THEOREM (CAUCHY-SCHWARZ INEQUALITY)
Suppose that X and Y are two random variables.

$$|E_{f_{X,Y}}[XY]| \leq E_{f_{X,Y}}[|XY|] \leq \left\{E_{f_X}[|X|^2]\right\}^{1/2} \left\{E_{f_Y}[|Y|^2]\right\}^{1/2}$$

Proof. Set $p = q = 2$ in the Hölder Inequality. ∎

Corollaries:

(a) Let μ_X and μ_Y denote the expectations of X and Y respectively. Then, by the Cauchy-Schwarz inequality

$$|E_{f_{X,Y}}[(X-\mu_X)(Y-\mu_Y)]| \leq \left\{E_{f_X}[(X-\mu_X)^2]\right\}^{1/2}\left\{E_{f_Y}[(Y-\mu_Y)^2]\right\}^{1/2}$$

so that

$$E_{f_{X,Y}}[(X-\mu_X)(Y-\mu_Y)] \leq E_{f_X}[(X-\mu_X)^2]E_{f_Y}[(Y-\mu_Y)^2]$$

and hence

$$\left\{Cov_{f_{X,Y}}[X,Y]\right\}^2 \leq Var_{f_X}[X]\,Var_{f_Y}[Y].$$

(b) **Lyapunov's Inequality**: Define $Y=1$ with probability one. Then, for $1<p<\infty$

$$E_{f_X}[|X|] \leq \{E_{f_X}[|X|^p]\}^{1/p}.$$

Let $1<r<p$. Then

$$E_{f_X}[|X|^r] \leq \{E_{f_X}[|X|^{pr}]\}^{1/p}$$

and letting $s=pr>r$ yields

$$E_{f_X}[|X|^r] \leq \{E_{f_X}[|X|^s]\}^{r/s}$$

so that

$$\{E_{f_X}[|X|^r]\}^{1/r} \leq \{E_{f_X}[|X|^s]\}^{1/s}$$

for $1<r<s<\infty$.

THEOREM (MINKOWSKI'S INEQUALITY)
Suppose that X and Y are two random variables, and $1 \leq p < \infty$. Then

$$\left\{E_{f_{X,Y}}[|X+Y|^p]\right\}^{1/p} \leq \{E_{f_X}[|X|^p]\}^{1/p} + \{E_{f_Y}[|Y|^p]\}^{1/p}$$

Proof. Write

$$\begin{aligned} E_{f_{X,Y}}[|X+Y|^p] &= E_{f_{X,Y}}[|X+Y||X+Y|^{p-1}] \\ &\leq E_{f_{X,Y}}[|X||X+Y|^{p-1}] + E_{f_{X,Y}}[|Y||X+Y|^{p-1}] \end{aligned}$$

by the triangle inequality $x+y| \leq |x|+|y|$. Using Hölder's Inequality on the terms on the right hand side, for q selected to satisfy $1/p+1/q=1$,

$$E_{f_{X,Y}}[|X+Y|^p] \leq \{E_{f_X}[|X|^p]\}^{1/p}\left\{E_{f_{X,Y}}[|X+Y|^{q(p-1)}]\right\}^{1/q} + \{E_{f_Y}[|Y|^p]\}^{1/p}\left\{E_{f_{X,Y}}[|X+Y|^{q(p-1)}]\right\}^{1/q}$$

and dividing through by $\left\{E_{f_{X,Y}}[|X+Y|^{q(p-1)}]\right\}^{1/q}$ yields

$$\frac{E_{f_{X,Y}}[|X+Y|^p]}{\left\{E_{f_{X,Y}}[|X+Y|^{q(p-1)}]\right\}^{1/q}} \leq \{E_{f_X}[|X|^p]\}^{1/p} + \{E_{f_Y}[|Y|^p]\}^{1/p}$$

and the result follows as $q(p-1)=p$, and $1-1/q=1/p$. ∎

3. JENSEN'S INEQUALITY

Jensen's Inequality gives a lower bound on expectations of convex functions. Recall that a function $g(x)$ is **convex** if, for $0 < \lambda < 1$, $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$ for all x and y. Alternatively, function $g(x)$ is **convex** if

$$\frac{d^2}{dt^2}\{g(t)\}_{t=x} = g^{(2)}(x) \geq 0.$$

Conversely, $g(x)$ is concave if $-g(x)$ is convex.

THEOREM (JENSEN'S INEQUALITY)
Suppose that X is a random variable with expectation μ, and function g is convex. Then

$$E_{f_X}[g(X)] \geq g(E_{f_X}[X])$$

with equality if and only if, for every line $a + bx$ that is a tangent to g at μ

$$P[g(X) = a + bX] = 1.$$

that is, $g(x)$ is linear.

Proof. Let $l(x) = a + bx$ be the equation of the tangent at $x = \mu$. Then, for each x, $g(x) \geq a + bx$ as in the figure below.

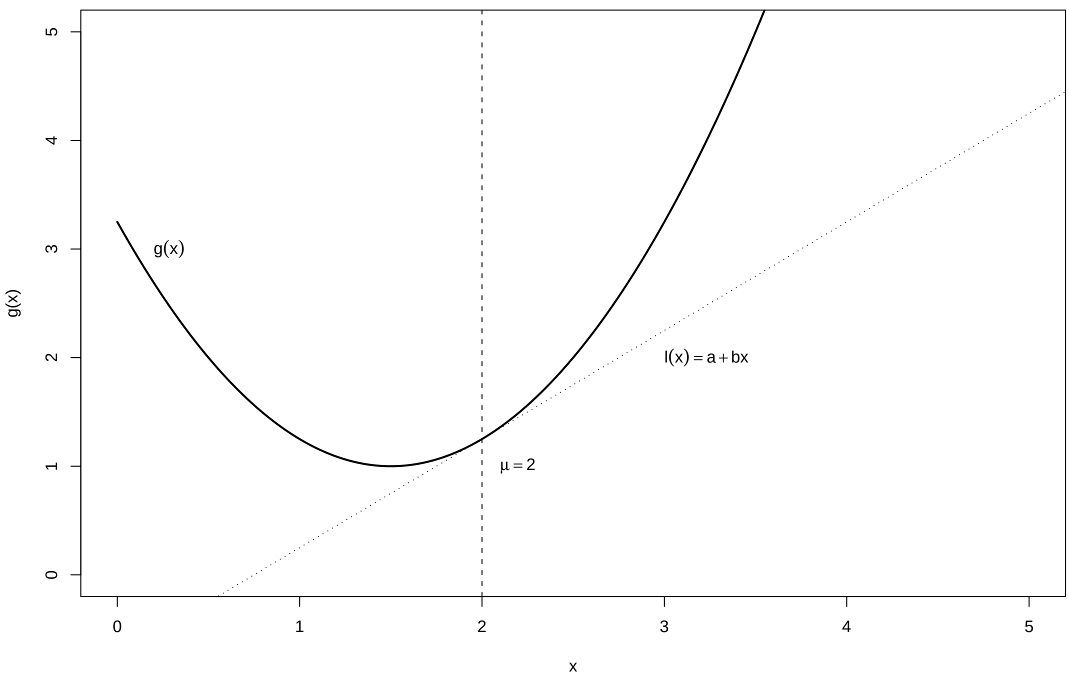

Figure 2: The function $g(x)$ and its tangent at $x = \mu$.

Thus

$$E_{f_X}[g(X)] \geq E_{f_X}[a + bX] = a + bE_{f_X}[X] = l(\mu) = g(\mu) = g(E_{f_X}[X])$$

as required. Also, **if** $g(x)$ is linear, then equality follows by properties of expectations. Suppose that

$$E_{f_X}[g(X)] = g(E_{f_X}[X]) = g(\mu)$$

but $g(x)$ is convex, but not linear. Let $l(x) = a + bx$ be the tangent to g at μ. Then by convexity

$$g(x) - l(x) > 0 \qquad \therefore \qquad \int (g(x) - l(x)) f_X(x)\, dx = \int g(x) f_X(x)\, dx - \int l(x) f_X(x)\, dx > 0$$

and hence

$$E_{f_X}[g(X)] > E_{f_X}[l(X)].$$

But $l(x)$ is linear, so $E_{f_X}[l(X)] = a + bE_{f_X}[X] = g(\mu)$, yielding the contradiction

$$E_{f_X}[g(X)] > g(E_{f_X}[X]).$$

and the result follows. ∎

Corollary and examples:

- If $g(x)$ is **concave**, then
$$E_{f_X}[g(X)] \leq g(E_{f_X}[X])$$
- $g(x) = x^2$ is **convex**, thus
$$E_{f_X}\left[X^2\right] \geq \{E_{f_X}[X]\}^2$$
- $g(x) = \log x$ is **concave**, thus
$$E_{f_X}[\log X] \leq \log\{E_{f_X}[X]\}$$

LEMMA
Suppose that X is a random variable, with finite expectation μ. Let g be a non-decreasing function. Then

$$E_{f_X}[g(X)(X-\mu)] \geq 0$$

Proof. Using the indicator random variable $I_A(X)$,

$$\begin{aligned}
E_{f_X}[g(X)(X-\mu)] &= E_{f_X}[g(X)(X-\mu)I_{(-\infty,0)}(X-\mu)] + E_{f_X}[g(X)(X-\mu)I_{[0,\infty)}(X-\mu)] \\
&= \int_{-\infty}^{\mu} g(x)(x-\mu)f_X(x)dx + \int_{\mu}^{\infty} g(x)(x-\mu)f_X(x)dx \\
&\geq \int_{-\infty}^{\mu} g(\mu)(x-\mu)f_X(x)dx + \int_{\mu}^{\infty} g(\mu)(x-\mu)f_X(x)dx \\
&= E_{f_X}[g(\mu)(X-\mu)I_{(-\infty,0)}(X-\mu)] + E_{f_X}[g(\mu)(X-\mu)I_{[0,\infty)}(X-\mu)] \\
&= E_{f_X}[g(\mu)(X-\mu)] = 0
\end{aligned}$$

∎