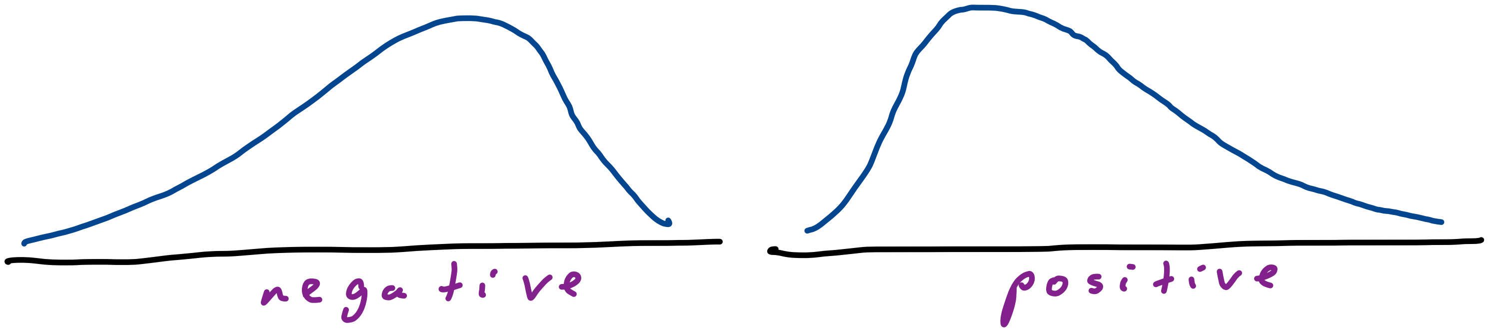
A new life of Pearson's skewness

Yevgeniy Kovchegov
Oregon State University

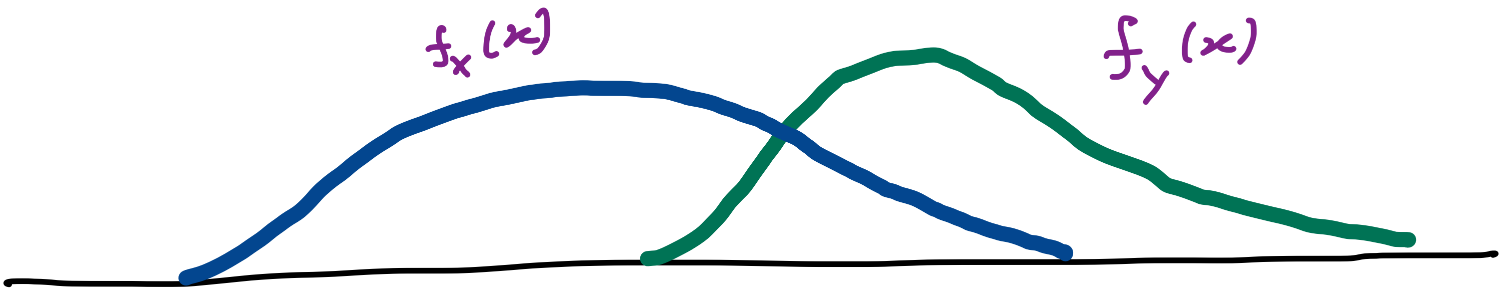
Skewness and Stochastic Dominance.

In this talk, we will connect the following two concepts.

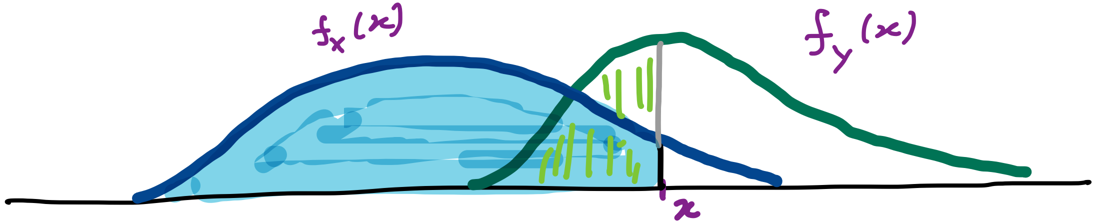
- **Skewness**



- **Stochastic Dominance**



Stochastic Dominance.



Consider random variables X and Y. If the cumulative distribution functions F_X and F_Y satisfy

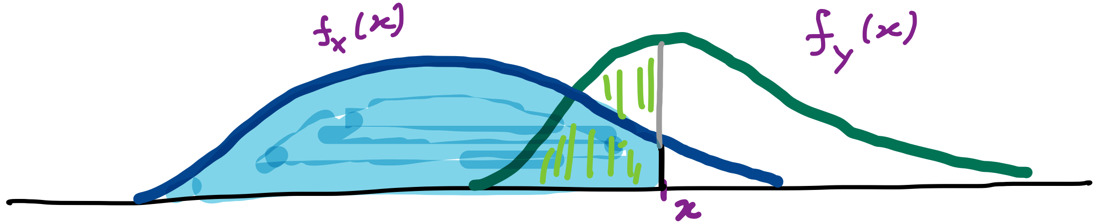
$$F_X(x) \geq F_Y(x) \qquad \forall x \in \mathbb{R},$$

then Y exhibits **stochastic dominance** over X.

Furthermore, if $F_X \not\equiv F_Y$, then, Y exhibits **strict stochastic dominance** over X.

Lemma. If Y exhibits stochastic dominance over X, then, for any increasing function $h : \mathbb{R} \to \mathbb{R}$ we have $E[h(Y)] \geq E[h(X)]$. Moreover, if Y exhibits strict stochastic dominance over X, and if $h(x)$ is strictly increasing, then $E[h(Y)] > E[h(X)]$.

Stochastic Dominance.



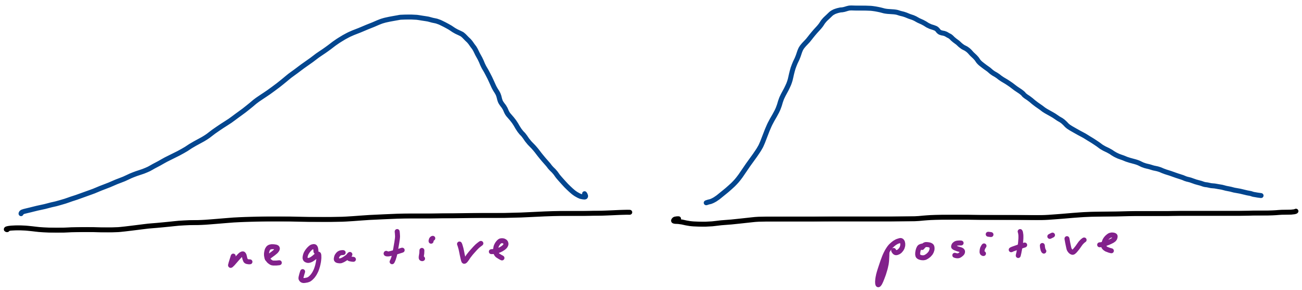
Lemma. If Y exhibits stochastic dominance over X, then, for any increasing function $h : \mathbb{R} \to \mathbb{R}$ we have $E[h(Y)] \geq E[h(X)]$. Moreover, if Y exhibits strict stochastic dominance over X, and if $h(x)$ is strictly increasing, then $E[h(Y)] > E[h(X)]$.

The above lemma is usually proved via a coupling argument. For continuous random variables, sometimes can use integration by parts:

$$\int_a^b h(x) f_Y(x) dx = h(b) - \int_a^b h'(x) F_Y(x) dx \geq h(b) - \int_a^b h'(x) F_X(x) dx = \int_a^b h(x) f_X(x) dx$$

Skewness.

Let $\mu = E[X]$ and $\sigma = \sqrt{Var(X)} = \sqrt{E[(X-\mu)^2]}$.



Pearson's moment coefficient of skewness

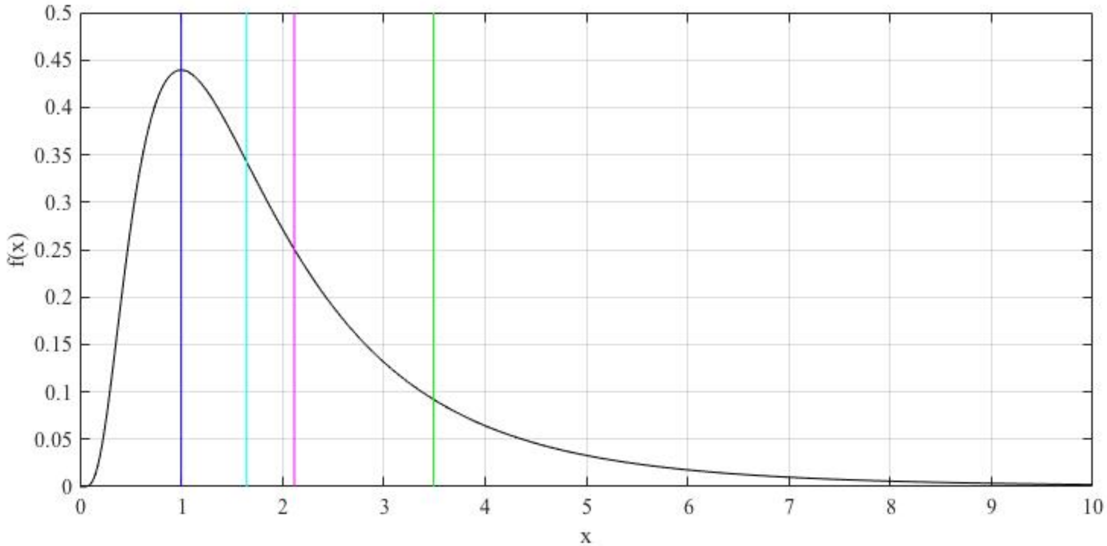
$$\gamma = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

The sign of γ determines positive/negative skewness.

Positive skewness $\Rightarrow$ **mean-median-mode inequality:**

mode $<$ median $<$ mean.

Skewness.

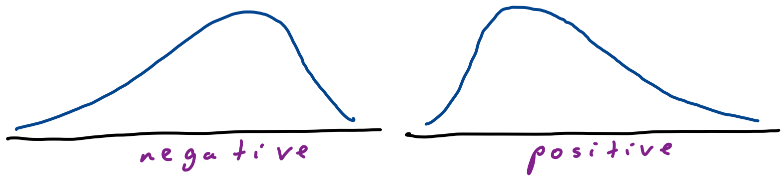


Positive skewness $\Rightarrow$ **mean-median-mode inequality:**

mode $<$ median $<$ mean.

Negative skewness $\Rightarrow$ mean $<$ median $<$ mode.

Skewness.



Pearson's moment coefficient of skewness

$$\gamma = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

Positive skewness $\Rightarrow$ **mean-median-mode inequality:**

mode $<$ median $<$ mean.

Pearson's first skewness coefficient (mode skewness):

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

Pearson's second skewness coefficient (median skewness):

$$3 \times \frac{\text{mean} - \text{median}}{\text{standard deviation}}.$$

Fréchet p-mean.

For $p \in [1, \infty)$ and a random variable X with the finite $(p-1)$-st moment, $E[|X|^{p-1}] < \infty$, the p-**mean** ν_p is the unique solution of

$$E\left[(X - \nu_p)_+^{p-1}\right] = E\left[(\nu_p - X)_+^{p-1}\right],$$

where $x_+ = \max\{0, x\}$.

The above defined p-**mean** is extending the notion of the **Fréchet** p-**mean**:

$$\nu_p = \text{argmin}_{a \in \mathbb{R}} E\left[|X - a|^p\right]$$

– required finiteness p-th moment, $E[|X|^p] < \infty$.

Notice that ν_1 is the **median**:

$$P(X > \nu_1) = P(X < \nu_1).$$

Fréchet p-mean.

For $p \in [1, \infty)$ and a random variable X with the finite $(p-1)$-st moment, $E[|X|^{p-1}] < \infty$, the p-**mean** ν_p is the unique solution of

$$E\left[(X - \nu_p)_+^{p-1}\right] = E\left[(\nu_p - X)_+^{p-1}\right],$$

where $x_+ = \max\{0, x\}$.

Notice that ν_1 is the **median**:

$$P(X > \nu_1) = P(X < \nu_1).$$

Similarly, ν_2 is the **mean**:

$$E\left[(X - \nu_2)_+\right] = E\left[(\nu_2 - X)_+\right] \quad \Leftrightarrow \quad E[X] = \nu_2.$$

Fréchet p-mean.

$$E\big[(X-\nu_p)_+^{p-1}\big] = E\big[(\nu_p - X)_+^{p-1}\big]$$

Notice that ν_1 is the **median**:

$$P(X > \nu_1) = P(X < \nu_1).$$

Similarly, ν_2 is the **mean**:

$$E\big[(X-\nu_2)_+\big] = E\big[(\nu_2 - X)_+\big] \quad \Leftrightarrow \quad E[X] = \nu_2.$$

Proposition. Consider a random variable X with $E[|X|^3] < \infty$. The Pearson's moment coefficient of skewness $\gamma > 0$ if and only if $\nu_4 > \nu_2$.

Proof.

$$\gamma = \left(\frac{\nu_4 - \nu_2}{\sigma}\right)^3 + 3\left(\frac{\nu_4 - \nu_2}{\sigma}\right)$$

Skewness.

Pearson's moment coefficient of skewness

$$\gamma = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] > 0 \quad \Leftrightarrow \quad \nu_4 > \nu_2$$

In the **unimodal case** let ν_0 denote the **mode**.

Positive skewness $\Rightarrow \; \nu_0 \; < \; \nu_1 \; < \; \nu_2$.

Pearson's first skewness coefficient (mode skewness):

$$\frac{\nu_2 - \nu_0}{\sigma}$$

Pearson's second skewness coefficient (median skewness):

$$\frac{3(\nu_2 - \nu_1)}{\sigma}.$$

True positive skewness.

Let

$$\mathcal{D} = \left\{p \geq 1 \,:\, E[|X|^{p-1}] < \infty\right\}$$

and, in the unimodal case, $\mathcal{D}_0 = \mathcal{D} \cup \{0\}$.

Definition. A random variable X is truly positively skewed if and only if ν_p is an increasing function of $p \in \mathcal{D}$. Analogously, for true negative skewness.

The above defined *true positive skewness* insures

$$\nu_1 < \nu_2 < \nu_4.$$

Definition. A unimodal distribution is truly mode positively skewed if and only if ν_p is an increasing function of of $p \in \mathcal{D}_0$. Analogously, for true negative skewness.

Exponential distribution.

Exponential random variable X: $f(x) = \lambda e^{-\lambda x}\mathbf{1}_{(0,\infty)}(x)$.
W.l.o.g. let $\lambda = 1$. Equation

$$\int_0^{\nu_p} (\nu_p - x)^{p-1} e^{-x}\, dx = \int_{\nu_p}^{\infty} (x - \nu_p)^{p-1} e^{-x}\, dx.$$

simplifies to

$$\int_0^{\nu_p} x^{p-1} e^{x}\, dx = \Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x}\, dx.$$

Differentiating $\frac{d}{dp}$ yields

$$\nu_p^{p-1} e^{\nu_p} \frac{d\nu_p}{dp} + \int_0^{\nu_p} x^{p-1} e^{x} \log x\, dx = \int_0^{\infty} x^{p-1}\, e^{-x} \log x\, dx.$$

Exponential distribution.

$$\int_0^{\nu_p} x^{p-1} e^{x}\,dx = \Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x}\,dx.$$

Since $\frac{1}{\Gamma(p)} x^{p-1}\, e^{-x} \mathbf{1}_{(0,\infty)}(x)$ stochastically dominates $\frac{1}{\Gamma(p)} x^{p-1}\, e^{x} \mathbf{1}_{(0,\nu_p)}(x)$,

$$\nu_p^{p-1} e^{\nu_p} \frac{d\nu_p}{dp} = \int_0^{\infty} x^{p-1}\, e^{-x} \log x\, dx - \int_0^{\nu_p} x^{p-1} e^{x} \log x\, dx \; > 0$$

by

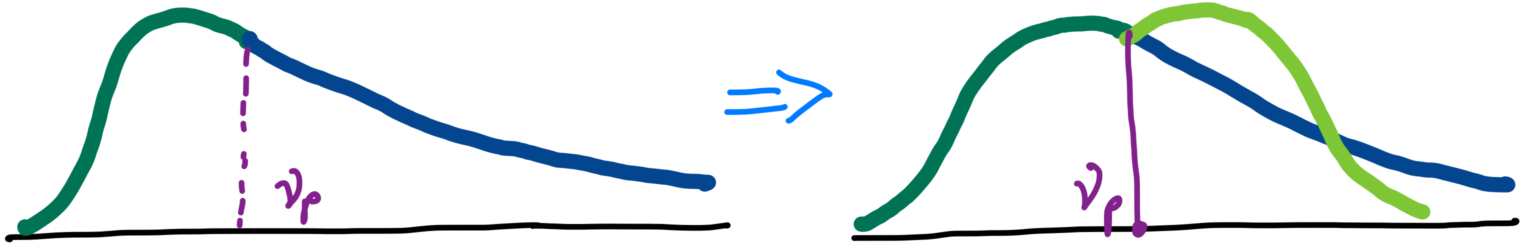
Lemma. If Y exhibits stochastic dominance over X, then, for any increasing function $h : \mathbb{R} \to \mathbb{R}$ we have $E[h(Y)] \geq E[h(X)]$. Moreover, if Y exhibits strict stochastic dominance over X, and if $h(x)$ is strictly increasing, then $E[h(Y)] > E[h(X)]$.

Skewness and stochastic dominance.

Let X be a continuous random variable with density $f(x)$ and $supp(f) = (L, R)$. Then

$$H_p := \int_0^{\nu_p - L} x^{p-1} f(\nu_p - x)\, dx = \int_0^{R-\nu_p} x^{p-1} f(\nu_p + x)\, dx.$$

Positive skewness: the left tail is "spreading short" and the right tail is "spreading longer".



Interpretation: $\frac{1}{H_p} x^{p-1} f(\nu_p + x) \mathbf{1}_{(0,R-\nu_p)}(x)$ to exhibit strict stochastic dominance over $\frac{1}{H_p} x^{p-1} f(\nu_p - x) \mathbf{1}_{(0,\nu_p - L)}(x)$.

Skewness and stochastic dominance.

Positive skewness: the left tail is "spreading short" and the right tail is "spreading longer".

Interpretation: $\frac{1}{H_p} x^{p-1} f(\nu_p + x) \mathbf{1}_{(0,R-\nu_p)}(x)$ to exhibit strict stochastic dominance over $\frac{1}{H_p} x^{p-1} f(\nu_p - x) \mathbf{1}_{(0,\nu_p - L)}(x)$.

Theorem. If $\frac{1}{H_p} x^{p-1} f(\nu_p + x) \mathbf{1}_{(0,R-\nu_p)}(x)$ exhibits strict stochastic dominance over $\frac{1}{H_p} x^{p-1} f(\nu_p - x) \mathbf{1}_{(0,\nu_p - L)}(x)$, then function ν_p is increasing at p.

Consequently, if the above stochastic dominance holds for all p in the interior of $\mathcal{D}$, the distribution is truly positively skewed

True positive skewness: examples.

• **Gamma random variable**: $f(x) = \frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}e^{-\lambda x}$ with parameters $\alpha > 0$ and $\lambda > 0$ is truly (mode) positively skewed.

• **Beta random variable**: $f(x) = \frac{1}{\mathcal{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ with parameters $\beta > \alpha > 1$ (and mode $\nu_0 = \frac{\alpha-1}{\alpha+\beta-2} < \frac{1}{2}$) is truly mode positively skewed.

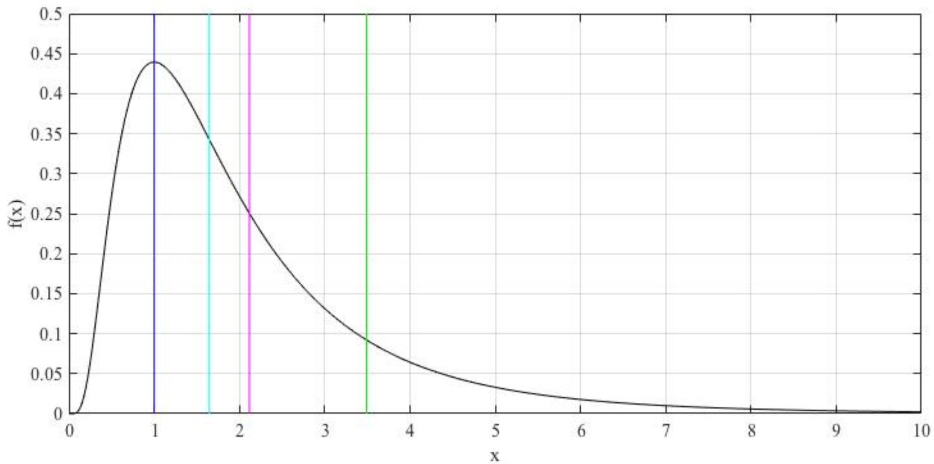
• **Pareto random variable**: $f(x) = \frac{\alpha}{x^{\alpha+1}}$ with parameter $\alpha > 0$ is truly mode positively skewed.

Notice that for $\alpha \in (0,1)$, the quantities ν_1, ν_2, and γ do not exist.

- **Log-normal random variable:**

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}$$

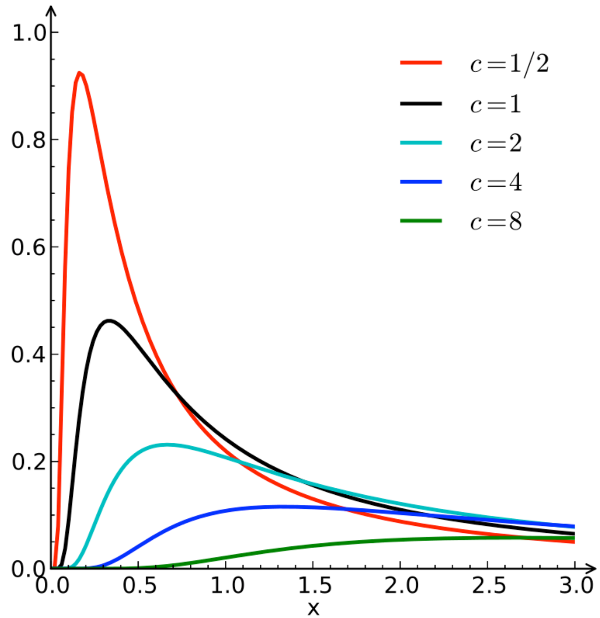
over $(L, R) = (0, \infty)$, with parameters μ and σ^2.



Theorem.

$$\nu_p = \exp\left\{\mu + \frac{p-1}{2}\sigma^2\right\}$$

Lévy distribution.



For scale parameter $c > 0$,

$$f(x) = \sqrt{\frac{c}{2\pi}}\frac{1}{x^{3/2}}e^{-c/(2x)}, \ x > 0$$

Mean: ∞

Skewness: undefined

REU'21 team
Alex Negrón, Clarice Pertel, and Christopher Wang

have shown that

- **Lévy distribution** is truly positively skewed:

$$\nu_p \uparrow \quad \text{for } p \in \mathcal{D} = [1, 3/2).$$