Mean-preserving Contraction and Stochastic Dominance: An Information Design Perspective

Zhicheng Du*

April 14, 2025

In information design, when both the sender and the risk-neutral receiver only care about the mean of the posterior distribution induced by a signal, the sender's feasible space of information strategies can be precisely characterized by the notion of mean-preserving contraction (henceforth MPC). While this point is well-known in the literature, it is rarely discussed in detail. Beginners may wonder about these two questions:

1. What is the connection between the notion of mean-preserving contraction and the notion of second-order stochastic dominance?
2. Why does the notion of mean-preserving contraction fully characterize the sender's feasible strategy space when the sender and the risk-neutral receiver only care about the posterior mean?

This note is written to answer these two questions.

1 Answer to the first question

Introduction to Mean-preserving Contraction. The notion of MPC provides a partial ordering of random variables based on their cumulative density functions (CDFs).

Definition 1 (MPC)**.** *Distribution $G \in \Delta([0,1])$ is an MPC of distribution $F \in \Delta([0,1])$ if*

(i) $\int_0^t G(x)\,\mathrm{d}x \le \int_0^t F(x)\,\mathrm{d}x, \quad \forall t \in [0,1]$*; and*

(ii) $\int_0^1 G(x)\,\mathrm{d}x = \int_0^1 F(x)\,\mathrm{d}x$.

The opposite definition of MPC is mean-preserving spread (henceforth, MPS). If distribution G is an MPC of distribution F, then distribution F is an MPS of distribution G, and vice versa. In information design, the distribution of posterior means forms an MPC of the prior when the sender and the risk-neutral receiver only care about the posterior mean.

Intuitively, constructing an MPC based on one distribution is essentially a process of making the distribution more concentrated around its mean while keeping the mean unchanged,

*The author thank Zihe Wang for helpful comments and suggestions.

thereby reducing uncertainty and dispersion of the distribution.

From the perspective of probability density, an MPC of some prior F is generated from the prior F through a series of mean-preserving contractions – each a process that moves the probability density from both sides to the middle while keeping the mean unchanged.

Introduction to Stochastic Dominance. In decision theory and economics, stochastic dominance is a fundamental concept used to compare the risk-return profiles of uncertain outcomes. Similar to MPC, it also provides a partial ordering of random variables based on their CDFs. In this note, we only care about the first-order stochastic dominance (henceforth FOSD) and the second-order stochastic dominance (henceforth SOSD).

Here is the original definition of FOSD, which is from the perspective of risk-return profile. We call the definition below as the *utility condition* of FOSD.

Definition 2 (FOSD). *Distribution $G \in \Delta([0,1])$ is FOSD to distribution $F \in \Delta([0,1])$, that is $G \succ_1 F$, if for all (weakly) increasing function $u : [0,1] \to \mathbb{R}$,*[1] *it holds*

$$\mathbb{E}_{x\sim G}[u(x)] = \int_0^1 u(x)\ \mathrm{d}G(x) \geq \int_0^1 u(x)\ \mathrm{d}F(x) = \mathbb{E}_{x\sim F}[u(x)]\ .$$

Below Lemma 1 means that the *utility condition* and the *CDF condition* of FOSD are equivalent. In other words, there are two kinds of equivalent definitions of FOSD: the *utility condition* and the *CDF condition*.

Lemma 1. *$\int_0^1 u(x)\ \mathrm{d}G(x) \geq \int_0^1 u(x)\ \mathrm{d}F(x)$ holds for all (weakly) increasing function $u : [0,1] \Leftrightarrow \mathbb{R} \Leftrightarrow F(x) \leq G(x)$ for all $x \in [0,1]$.*

Proof of Lemma 1. First, we prove that $F(x) \leq G(x)$ for all $x \in [0,1] \Rightarrow \int_0^1 u(x)\ \mathrm{d}G(x) \geq \int_0^1 u(x)\ \mathrm{d}F(x)$ holds for all (weakly) increasing function $u : [0,1] \to \mathbb{R}$. Since $F(x) \leq G(x)$ for all $x \in [0,1]$, we have $\int_0^1 (G(x) - F(x))\ \mathrm{d}u(x) \geq 0$ for all (weakly) increasing function u. Through integration by parts, we achieve the target result.

Second, we prove that $\int_0^1 u(x)\ \mathrm{d}G(x) \geq \int_0^1 u(x)\ \mathrm{d}F(x)$ holds for all (weakly) increasing function $u : [0,1] \to \mathbb{R} \Rightarrow F(x) \leq G(x)$ for all $x \in [0,1]$. We prove it by contradiction. Suppose that there exists $x_0 \in (0,1)$ such that $F(x_0) > G(x_0)$. We construct a weakly increasing function $u : [0,1] \to \mathbb{R}$ as follows

$$u(x) = \begin{cases} 1, & \text{if } x \leq x_0\ , \\ 0, & \text{otherwise}\ . \end{cases}$$

By the fact that $\int_0^1 u(x)\ \mathrm{d}G(x) \geq \int_0^1 u(x)\ \mathrm{d}F(x)$, we have

$$\int_0^1 u(x)\ \mathrm{d}G(x) - \int_0^1 u(x)\ \mathrm{d}F(x) = \int_0^{x_0} u(x)\ \mathrm{d}G(x) - \int_0^{x_0} u(x)\ \mathrm{d}F(x) = G(x_0) - F(x_0) \geq 0\ ,$$

which forms a contradiction with the assumption.

Combining these two directions above, we have finished the proof. $\square$

[1] Function u satisfies that $u'(x) \geq 0$ for all $x \in [0,1]$.

Here is the original definition of SOSD, which is from the perspective of risk-return profile. We call the definition below as the *utility condition* of SOSD.

Definition 3 (SOSD)**.** *Distribution $G \in \Delta([0,1])$ is SOSD to distribution $F \in \Delta([0,1])$, that is $G \succ_2 F$, if for all (weakly) increasing and concave function $u : [0,1] \to \mathbb{R}$,*[2] *it holds*

$$\mathbb{E}_{x\sim G}[u(x)] = \int_0^1 u(x)\ \mathrm{d}G(x) \geq \int_0^1 u(x)\ \mathrm{d}F(x) = \mathbb{E}_{x\sim F}[u(x)]\ .$$

Below Lemma 2 means that the *utility condition* and the *CDF condition* of SOSD are equivalent. In other words, there are two kinds of equivalent definitions of SOSD: the *utility condition* and the *CDF condition*.

Lemma 2. *$\int_0^1 u(x)\ \mathrm{d}G(x) \geq \int_0^1 u(x)\ \mathrm{d}F(x)$ holds for all (weakly) increasing and concave function $u : [0,1] \to \mathbb{R} \Leftrightarrow \int_0^t G(x)\ \mathrm{d}x \leq \int_0^t F(x)\ \mathrm{d}x$ for all $t \in [0,1]$.*

Proof of Lemma 2. First, we prove that $\int_0^t G(x)\ \mathrm{d}x \leq \int_0^t F(x)\ \mathrm{d}x$ for all $t \in [0,1] \Rightarrow \int_0^1 u(x)\ \mathrm{d}G(x) \geq \int_0^1 u(x)\ \mathrm{d}F(x)$ holds for all (weakly) increasing and concave function $u : [0,1] \to \mathbb{R}$. Through integration by parts, we have

$$\begin{aligned}
&\int_0^1 u(x)\ \mathrm{d}G(x) - \int_0^1 u(x)\ \mathrm{d}F(x) \\
&= \int_0^1 u'(x)[F(x) - G(x)]\ \mathrm{d}x = \int_0^1 u'(x)\ \mathrm{d}\int_0^x [F(t) - G(t)]\ \mathrm{d}t \\
&= \left[u'(x)\int_0^x [F(t) - G(t)]\ \mathrm{d}t\right]_0^1 - \int_0^1 u''(x)\left(\int_0^x [F(t) - G(t)]\ \mathrm{d}t\right)\ \mathrm{d}x \\
&= u'(1)\left[\int_0^1 F(t)\ \mathrm{d}t - \int_0^1 G(t)\ \mathrm{d}t\right] - \int_0^1 u''(x)\left(\int_0^x [F(t) - G(t)]\ \mathrm{d}t\right)\ \mathrm{d}x\ .
\end{aligned}$$

By the fact that $\int_0^t G(x)\ \mathrm{d}x \leq \int_0^t F(x)\ \mathrm{d}x$ for all $t \in [0,1]$, we have that $\int_0^1 u(x)\ \mathrm{d}G(x) - \int_0^1 u(x)\ \mathrm{d}F(x) \geq 0$.

Second, we prove that $\int_0^1 u(x)\ \mathrm{d}G(x) \geq \int_0^1 u(x)\ \mathrm{d}F(x)$ holds for all (weakly) increasing and concave function $u : [0,1] \to \mathbb{R} \Rightarrow \int_0^t G(x)\ \mathrm{d}x \leq \int_0^t F(x)\ \mathrm{d}x$ for all $t \in [0,1]$. For any fixed $t \in [0,1]$, we can construct the function $u_t(x) = \min(x,t)$. Replacing the function u with

[2] Function u satisfies that $u'(x) \geq 0$ and $u''(x) \geq 0$ for all $x \in [0,1]$.

the function u_t and using integration by parts, we have

$$\begin{aligned}
\int_0^1 u_t(x)\ \mathrm{d}G(x) &\geq \int_0^1 u_t(x)\ \mathrm{d}F(x) \\
\int_0^t x\ \mathrm{d}G(x) + t\int_t^1\ \mathrm{d}G(x) &\geq \int_0^t x\ \mathrm{d}F(x) + t\int_0^1\ \mathrm{d}F(x)\ , \\
\int_0^t x\ \mathrm{d}G(x) + t(1-G(t)) &\geq \int_0^t x\ \mathrm{d}F(x) + t(1-F(t))\ , \\
\left(tG(t) - \int_0^t G(x)dx\right) + t(1-G(t)) &\geq \left(tF(t) - \int_0^t F(x)dx\right) + t(1-F(t))\ , \\
\int_0^t G(x)\ \mathrm{d}x &\leq \int_0^t F(x)\ \mathrm{d}x\ .
\end{aligned}$$

Combining these two directions above, we have finished the proof. □

The Relationship between MPC and SOSD. We observe that the definitions of MPC and SOSD, though seemingly similar, are not equivalent. Indeed, MPC is a stronger version of SOSD since the former also requires two distributions to share a common mean. Given a distribution F, an MPS of distribution F is also SOSD to distribution F, while the converse does not hold. Given that distribution G is SOSD to distribution F, if these two distributions further share a common mean, then distribution G also forms an MPC of distribution F. The relationship can be exactly summarized in the following lemma.

Lemma 3. *Given two distributions over $[0,1]$, F and G, $G \in \mathsf{MPC}(F)$ if and only if $G \succ_2 F$ and $\mathbb{E}_{x\sim F}[x] = \mathbb{E}_{x\sim G}[x]$.*

2 Answer to the second question

Given the sender's prior F, we use $\mathsf{MPC}(F)$ to represent the space of all mean-preserving contractions induced from the prior F. According to Rothschild and Stiglitz (1970); Blackwell and Girshick (1979); Gentzkow and Kamenica (2016), we have the following lemma.

Theorem 4. *There exists an information structure that induces the distribution G over posterior means from prior distribution F if and only if $G \in \mathsf{MPC}(F)$.*

To prove this theorem, we have to introduce the intermediary notion below.

Definition 4 (Martingale Coupling)**.** *Given two distributions over $[0,1]$, F and G, a two-dimensional distribution H over $[0,1]\times[0,1]$ is a Martingale Coupling (Sub-Martingale Coupling) for F and G, if it holds: (i) F and G are both H's marginal distributions; and (ii) $\mathrm{E}_{x\sim H_{x|y}}[x] = y$ ($\mathrm{E}_{x\sim H_{x|y}}[x] \leq y$) for each $y \in [0,1]$.*[3]

Lemma 5 (Strassen's Theorem) brings an equivalent condition to the existence of a Sub-Martingale Coupling for two distributions, that is, for any Sub-Martingale coupling $H \in \Delta([0,1]\times[0,1])$, the relationship between its two marginals is SOSD.

[3]Here $H_{x|y}$ denotes the conditional distribution of x given y under joint distribution H.

Lemma 5 (Strassen's Theorem)**.** *Given two distributions over $[0,1]$, F and G, $G \succ_2 F$ if and only if there exists a Sub-Martingale Coupling $H \in \Delta([0,1] \times [0,1])$ for F and G.*

Proof of Lemma 5. The existence of a Sub-Martingale coupling H for distributions F and G is equivalent to the fact that there is a feasible solution to the following infinite-dimensional linear system (we omit the objective of the linear system since it is useless). We use f and g to denote the probability density functions (PDFs) of F and G.

$$
\begin{aligned}
\text{subject to} \quad & -\int_0^1 \mathrm{d}H(x,y) = -f(x), && \forall x \in [0,1] \\
& \int_0^1 \mathrm{d}H(x,y) = g(y), && \forall y \in [0,1] \\
& \int_0^1 (x-y)\ \mathrm{d}H(x,y) \le 0, && \forall y \in [0,1]
\end{aligned}
$$

By introducing dual variables for these constraints and then applying Farkas' lemma,[4] the above fact is further equivalent to the fact that there exists no such pair of dual variables (λ, μ, γ) that is a solution to the following linear system.

$$
\begin{aligned}
\text{subject to} \quad & \lambda(x) \le \mu(y) + \gamma(y)(x-y), && \forall (x,y) \in [0,1]^2 && (1) \\
& \gamma(y) \ge 0, && \forall y \in [0,1] && (2) \\
& \int_0^1 \mu(y) g(y)\ \mathrm{d}y - \int_0^1 \lambda(x) f(x)\ \mathrm{d}x < 0 && && (3)
\end{aligned}
$$

There exists a feasible solution to the linear system above if and only if there exist functions u and s such that

$$
\begin{aligned}
& u(x) \le u(y) + s(y)(x-y), && \forall (x,y) \in [0,1]^2 && (4) \\
& s(y) \ge 0, && \forall y \in [0,1] && (5) \\
& \int_0^1 u(y) g(y)\ \mathrm{d}y - \int_0^1 u(x) f(x)\ \mathrm{d}x < 0 && && (6)
\end{aligned}
$$

Here is the proof: First, we prove the sufficiency ($\Leftarrow$). Let $\lambda(x) = u(x)$ for each $x \in [0,1]$, $\mu(y) = u(y)$ for each $y \in [0,1]$, and $\gamma(y) = s(y)$ for each $y \in [0,1]$, then the sufficiency directly holds. Second, we prove the necessity ($\Rightarrow$). Let $u(x) = \min_{y \in [0,1]}\{\mu(y) + s(y)(x-y)\}$ for each $x \in [0,1]$. Function u is increasing and concave, so let function s denote the supergradient of function u, then Conditions (4) and (5) hold. By the fact $\lambda(x) \le u(x)$ for each $x \in [0,1]$ and $u(y) \le \mu(y)$ for each $y \in [0,1]$, Condition (6) holds.

With this intermediate result, now we can finish our proof. Conditions (4) and (5) imply that, function u is an increasing concave function. Condition (6) states that $\mathrm{E}_{x \sim G}[u(x)] <$

[4] Farkas' lemma states that exactly one of the following systems has a solution: (i) $Ax = b$ with $x \ge 0$; and (ii) $A^\top y \ge 0$ with $b^\top y < 0$, where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, and $y \in \mathbb{R}^m$.

$\mathrm{E}_{x\sim F}[u(x)]$. Thus there is no feasible solution to the dual linear system is equivalent to the fact that, for all increasing concave function u, $\mathrm{E}_{x\sim G}[u(x)] \geq \mathrm{E}_{x\sim F}[u(x)]$, which is exactly the definition of distribution G being SOSD to distribution F. $\square$

With Lemma 5, it suffices to prove that, for two distributions F and G, there exists an information structure that induces distribution G over posterior mean from prior distribution F, if and only if, $\mathbb{E}_{x\sim F}[x] = \mathbb{E}_{x\sim G}[x]$ and there exists a Sub-Martingale Coupling for these two distributions. Note that, if distributions F and G share a common mean, then any Sub-Martingale Coupling for these two distributions is also a Martingale Coupling.

Proof of Theorem 4. First, we prove the proof of necessity ($\Rightarrow$). Suppose that information structure $(\pi(\cdot|\cdot), S)$ induces the distribution of posterior mean G from the prior F. Based on the Bayes Rule, we directly have that

$$\mathbb{E}_{x\sim F}[x] = \mathbb{E}_{x\sim G}[x] = \mathbb{E}_{v\sim F}[\mathbb{E}_{s\sim\pi(\cdot|v)}[\mathbb{E}[x|s]]] \ .$$

We construct a two-dimensional distribution H as follows

$$H(x,y) \triangleq F(x) \times G(y), \ \forall x,y \in [0,1] \ .$$

It is obvious that distributions F and G are two marginals of the two-dimensional distribution H. Based on the fact that distribution G is the distribution over posterior mean from prior F, we have that $\mathrm{E}_{x\sim H_{x|y}}[x] = y$ for each $y \in [0,1]$. Thus, the two-dimensional distribution H indeed forms a Martingale Coupling of distributions F and G.

Second, we prove the proof of sufficiency ($\Leftarrow$). We can construct such an information structure: let $S \triangleq [0,1]$ and $\pi(s = y|x) = H(y|x) = \frac{H(x,y)}{F(x)}$ for any $x \in [0,1]$ and $y \in [0,1]$. Since the two-dimensional distribution H is a Martingale Coupling of distributions F and G, it is easy to check that the distribution of posterior means – induced from the prior F and the constructed information structure after updating using Bayes Rule is exactly the distribution G. $\square$

References

David A Blackwell and Meyer A Girshick. *Theory of games and statistical decisions*. Courier Corporation, 1979.

Matthew Gentzkow and Emir Kamenica. A rothschild-stiglitz approach to bayesian persuasion. *American Economic Review*, 106(5):597–601, 2016.

Michael Rothschild and Joseph Stiglitz. Increasing risk: I. a definition. *Journal of Economic Theory*, 2(3):225–243, 1970.