Chapter 3

Properties of Information

Many economic settings involve an unknown type or quality θ and a signal X about θ, where both θ and X are ordered (i.e., there are "better" qualities θ and "higher" signal realizations X). In these settings, we might think that higher realizations of X are good news about θ—for example, that higher test scores suggest higher ability or that better reviews for a product suggest higher quality. These positive inferences are not in general justified, requiring assumptions on the joint distribution of (θ, X).

Section 3.1 presents three useful definitions of positive dependence between random variables, which are applied to our motivating problem (inference about θ from observation of a signal X) in Section 3.2. Section 3.3 presents an example of the kind of counterintuitive result that can obtain when these properties are not imposed on the informational environment.

3.1 Definitions

3.1.1 Monotone Likelihood Ratio Property

Consider two random variables Z and $\widetilde{Z}$ with distributions F and $\widetilde{F}$ that admit densities f and $\tilde{f}$. To simplify exposition, all densities mentioned in this chapter are assumed to be everywhere positive.

Definition 3.1. *The distribution F* likelihood-ratio dominates *the distribution $\widetilde{F}$ if*[1]

$$\frac{f(z)}{f(z')} \geq \frac{\widetilde{f}(z)}{\widetilde{f}(z')} \qquad \textit{for all } z > z'$$

Intuitively, moving up in the likelihood-ratio dominance order renders higher realizations of z more likely relative to lower realizations.

This definition is often specialized to conditional densities in the following way. Suppose θ and X are real-valued random vectors defined on the same

[1] The assumption that densities are everywhere strictly positive allows us to define the monotone likelihood ratio property in terms of likelihood ratios. More generally, we can consider a distribution F to likelihood-ratio dominate another distribution $\widetilde{F}$ if $f(z)\tilde{f}(z') \geq f(z')\tilde{f}(z)$ for all $z > z'$.

probability space with densities f_θ and f_X and conditional densities $f_{\theta|X}$ and $f_{X|\theta}$.

Definition 3.2. *The family of conditional densities* $\{f_{X|\theta}(\cdot \mid \theta)\}_{\theta\in\Theta}$ *have the* monotone likelihood ratio property *(MLRP) if for every* $x > x'$ *and* $\theta > \theta'$,

$$\frac{f_{X|\theta}(x \mid \theta)}{f_{X|\theta}(x' \mid \theta)} \geq \frac{f_{X|\theta}(x \mid \theta')}{f_{X|\theta}(x' \mid \theta')}. \tag{3.1}$$

If the inequality above holds strictly at every $x > x'$*, then we say that* $\{f_{X|\theta}(\cdot \mid \theta)\}$ *have the* strict *monotone likelihood ratio property.*

Remark 3.1. If $\{f_{X|\theta}(\cdot \mid \theta)\}$ satisfy MLRP, then $\{f_{\theta|X}(\cdot \mid X)\}$ also satisfy MLRP. To see this, observe that by Bayes' rule, (3.1) can be rewritten

$$\frac{f_{\theta|X}(\theta \mid x)f_X(x)}{f_\theta(\theta)}\frac{f_\theta(\theta)}{f_{\theta|X}(\theta \mid x')f_X(x')} \geq \frac{f_{\theta|X}(\theta' \mid x)f_X(x)}{f_\theta(\theta')}\frac{f_\theta(\theta')}{f_{\theta|X}(\theta' \mid x')f(x')}$$

which simplifies to the condition that $\{f_{\theta|X}(\cdot \mid X)\}$ have the monotone likelihood ratio property.

In the special case of an additive signal $X = \theta + \varepsilon$, where ε is independent of θ and has density f_ε,

$$\frac{f_{\theta|X}(\theta \mid x)}{f_{\theta|X}(\theta' \mid x)} = \frac{f_\varepsilon(x-\theta)}{f_\varepsilon(x-\theta')}$$

so the MLRP condition in (3.1) becomes

$$\frac{f_\varepsilon(x-\theta)}{f_\varepsilon(x'-\theta)} \geq \frac{f_\varepsilon(x-\theta')}{f_\varepsilon(x'-\theta')} \qquad \text{for every } x > x' \text{ and } \theta > \theta',$$

i.e., for every $\theta > \theta'$, the function $\frac{f_\varepsilon(x-\theta)}{f_\varepsilon(x-\theta')}$ is nondecreasing in x. It turns out that this is precisely the condition that f_ε is log concave.

Definition 3.3. *A function* f *that maps a convex set into the positive reals is* log-concave *if the function* $\ln f$ *is concave.*

Proposition 5 (Saumard and Wellner (2014)). *A density function* f *on* $\mathbb{R}$ *is log-concave if and only if for every* $\theta > \theta'$*, the ratio* $\frac{f(x-\theta)}{f(x-\theta')}$ *is a non-decreasing function of* x.

Thus, in any model where (1) $X = \theta + \varepsilon$, (2) the noise term ε is independent of θ, and (3) ε has a log-concave density, we can be guaranteed that $\{f_{\theta|X}(\cdot \mid x)\}$ has the monotone likelihood ratio property (no matter the distribution of θ).

Many distributions have log-concave densities—for example, normal distributions, exponential distributions, the uniform distribution over any convex set, the logistic distribution, and the extreme value distribution. But others do not—for example, the Pareto distribution and Cauchy distribution. See Saumard and Wellner (2014) or Bagnoli and Bergstrom (2005) for other examples and properties of log-concave distributions.

3.1.2 Affiliation

Let $Z_1, \ldots, Z_n$ be real-valued random variables taking values in $\mathbb{R}^n$ and admitting joint density f, which again we'll assume to be everywhere strictly positive. For any $z, z' \in \mathbb{R}^n$, let $z \wedge z'$ ("z meet z') denote the component-wise minimum of z and z', and $z \vee z'$ ("z join z') denote the component-wise maximum, i.e.,

$$z \vee z' = (\max(z_1, z_1'), \ldots, \max(z_n, z_n'))$$
$$z \wedge z' = (\min(z_1, z_1'), \ldots, \min(z_n, z_n'))$$

DEFINITION 3.4. *The variables* $Z_1, \ldots, Z_n$ *are* affiliated *if*

$$f(z \vee z')f(z \wedge z') \geq f(z)f(z') \tag{3.2}$$

for all $z, z' \in \mathbb{R}^n$.

This condition loosely says that larger realizations of any one variable make larger realizations of the other variables more likely. Figure 3.1 depicts this relationship for two binary variables.

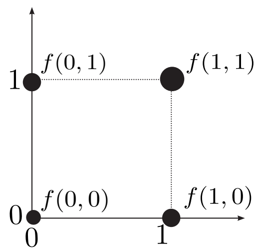

Figure 3.1: Two binary variables with joint density f are affiliated if $f(1,1)f(0,0) \geq f(1,0)f(0,1)$.

REMARK 3.2. If $Z_1, \ldots, Z_n$ are mutually independent, then they are affiliated.

Besides Definition 3.4, there are several equivalent ways to characterize affiliation. The first follows by taking logs of both sides of (3.2).

Proposition 6. $Z_1, Z_2, \ldots, Z_n$ *are affiliated if and only if* f *is log-supermodular, i.e.*

$$\log f(z \vee z') + \log f(z \wedge z') \geq \log f(z) + \log f(z')$$

for all z, z'.

Proposition 7. *Suppose the joint density* f *is twice-differentiable. Then* $Z_1, Z_2, \ldots, Z_n$ *are affiliated if and only if* $\frac{\partial^2 \log f}{\partial z_i z_j} \geq 0$.

We show the only if direction below, leaving the if direction for an exercise.
Proof. Without loss let $i = 1$ and $j = 2$. Choose any $z_1, z_1', z_2, z_2' \in \mathbb{R}$ where $z_1 > z_1'$ and $z_2 > z_2'$. Suppose $Z_1, Z_2, \ldots, Z_n$ are affiliated. Then by definition

$$\log f(z_1, z_2, z_{-12}) - \log f(z_1', z_2, z_{-12}) \geq \log f(z_1, z_2', z_{-12}) - \log f(z_1', z_2', z_{-12})$$

where z_{-12} is shorthand for $(z_3, \ldots, z_n)$. Rewrite z_1 as $z_1' + \varepsilon$ and divide both sides by ε. Taking the limit as $\varepsilon \to 0$, we have

$$\lim_{\varepsilon \to 0} \left(\frac{\log f(z_1' + \varepsilon, z_2, z_{-12}) - \log f(z_1', z_2, z_{-12})}{\varepsilon} \right) \geq \lim_{\varepsilon \to 0} \left(\frac{\log f(z_1' + \varepsilon, z_2', z_{-12}) - \log f(z_1', z_2', z_{-12})}{\varepsilon} \right)$$

so $\frac{\partial \log f}{\partial z_1}$ is increasing in z_2, as desired. ■

EXERCISE 3.1 (G). *Prove the 'if' direction of Proposition 7: If the joint density f is twice-differentiable and satisfies $\frac{\partial^2 \log f}{\partial z_i z_j} \geq 0$, then $Z_1, Z_2, \ldots, Z_n$ are affiliated.*

The next characterization simplifies (3.2) to a pairwise condition. Specifically, for any (Z_i, Z_j) and any realization of the remaining variables Z_{-ij}, higher realizations of Z_i must imply higher realizations of Z_j.

Proposition 8. *$Z_1, \ldots, Z_n$ are affiliated if and only if*

$$f(z_i, z_j, z_{-ij}) f(z_i', z_j', z_{-ij}) \geq f(z_i', z_j, z_{-ij}) f(z_i, z_j', z_{-ij}) \tag{3.3}$$

for every pair of distinct indices i, j, and every $z_i > z_i'$, $z_j > z_j'$, and $z_{-ij} \in \mathbb{R}^{n-2}$.

EXERCISE 3.2 (G). *Prove Proposition 8.*

This pairwise characterization immediately implies the following characterization, which says that $(Z_1, \ldots, Z_n)$ are affiliated if and only if for every pair of variables i, j, and every realization of z_{-ij}, the family of conditional densities $\{f(\cdot \mid z_j, z_{-ij})\}_{z_j \in \mathbb{R}}$ has the monotone-likelihood ratio property.

Proposition 9. *$Z_1, \ldots, Z_n$ are affiliated if and only if*

$$f(z_i \mid z_j, z_{-ij}) f(z_i' \mid z_j', z_{-ij}) \geq f(z_i \mid z_j', z_{-ij}) f(z_i' \mid z_j, z_{-ij}) \tag{3.4}$$

for every pair of distinct indices i, j, and every $z_i > z_i'$, $z_j > z_j'$, and $z_{-ij} \in \mathbb{R}^{n-2}$.

Proof. The displays in (3.3) and (3.4) are equivalent to one another by Bayes' rule, so Proposition 8 implies Proposition 9. ■

Operations that preserve affiliation include:

Proposition 10 (Monotone Functions)**.** *Suppose $Z_1, \ldots, Z_n$ are affiliated, and the functions $g_i : \mathbb{R} \to \mathbb{R}$, $1 \leq i \leq n$, are either all nondecreasing or all nonincreasing. Then the variables $g_1(Z_1), \ldots, g_n(Z_n)$ are affiliated.*

Proposition 11 (Subsets)**.** *Suppose $Z_1, \ldots, Z_n$ are affiliated and let $A \subseteq \{1, \ldots, n\}$ be any subset of these variables. Then the variables $(Z_i)_{i \in A}$ are affiliated.*

Proposition 12 (Order Statistics)**.** *For each $1 \leq i \leq n$, let $z^{(i)}$ denote the i-th largest realization among $(z_1, \ldots, z_n)$. Then the variables $(Z^{(1)}, \ldots, Z^{(n)})$ are affiliated.*

EXERCISE 3.3 (G). *Show that affiliation is not preserved under arbitrary linear combinations of affiliated variables by constructing an example of random variables Z_1, Z_2, Z_3 where (Z_1, Z_2, Z_3) are affiliated but $(Z_1 + Z_2, Z_3)$ are not.*

3.1.3 First-Order Stochastic Dominance

Again consider two real-valued random variables, a parameter θ and a signal X, defined on the same probability space with joint distribution F. In many applications we may expect a higher signal realization to lead to a higher inference about the unknown parameter. We now formalize 'higher inference' as a first-order stochastic dominance shift in the posterior belief.

DEFINITION 3.5. *A distribution F* first-order stochastically dominates $\widetilde{F}$, *which we denote by $F \geq_{FOSD} \widetilde{F}$, if*

$$\int u(\theta) dF(\theta) \geq \int u(\theta) d\widetilde{F}(\theta)$$

for every nondecreasing function $u : \mathbb{R} \to \mathbb{R}$. Equivalently, $F(\theta) \leq \tilde{F}(\theta)$ at every $\theta \in \Theta$.

If u is interpreted as a utility function over money, then a monetary gamble distributed according to F is preferred over one distributed according to $\tilde{F}$ by every agent that prefers more money over less, regardless of the specific shape of the agent's utility function. We can use this definition to compare conditional beliefs about θ.

DEFINITION 3.6. *Say that F has the* FOSD property *if $F_{\theta|X}(\cdot \mid X = x) \geq_{FOSD} F_{\theta|X}(\cdot \mid X = x')$ for all $x > x'$.*

Milgrom (1981) proposed a closely related property, which is imposed on conditional distributions $F_{X|\theta}$ rather than joint distributions F. (This is analogous to considering a signal $\sigma : \Theta \to \Delta(S)$ without fixing a prior on Θ.)

DEFINITION 3.7. *Say that a signal realization x is* more favorable than *signal realization x' if for every prior distribution $F_\theta \in \Delta(\Theta)$, the posterior distribution $F_{\theta|X}(\cdot \mid x)$ first-order stochastically dominates the posterior distribution $F_{\theta|X}(\cdot \mid x')$.*

That is, x is more favorable than x' if observing the realization x leads to a FOSD-higher posterior belief about θ (compared to observing x'). If x is more favorable than x' for all $x > x'$, then we have a stronger version of the FOSD property (given in (3.6)) that holds not only for the specific joint distribution F, but for all joint distributions F that are generated by $F_{X|\theta}$ and some choice of prior F_θ.

EXAMPLE 3.1. Recall that in the normal-updating setting with $\theta \sim \mathcal{N}(\mu, \sigma_\theta^2)$, $X = \theta + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, and $\theta \perp\!\!\!\perp \varepsilon$, the agent's posterior belief about θ conditional on X is

$$\mathcal{N}\left(\frac{\sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2} X + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\theta^2} \mu, \frac{\sigma_\varepsilon^2 \sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2}\right).$$

This distribution is increasing (in the FOSD order) in the realization of X for all parameters μ and σ_θ^2. So x is more favorable than x' for every pair $x > x'$.

3.2 How They are Related

Let θ and X be real-valued random vectors defined on the same probability space. We'll use F to denote their joint distribution, and assume throughout that the densities f_θ and f_X and conditional densities $f_{\theta|X}$ and $f_{X|\theta}$ exist. In this setting, our main properties from above are:

A: (X, θ) are affiliated.

MLRP: $\{f_{X|\theta}(\cdot \mid \theta)\}$ satisfies MLRP.

FOSD: For all $x > x'$, $F(\cdot \mid X = x) \geq_{FOSD} F(\cdot \mid X = x')$

MF: For all $x > x'$, x is more favorable than x'

These properties are related in the following way:

$$\textbf{(A)} \iff \textbf{(MLRP)} \iff \textbf{(MF)} \implies \textbf{(FOSD)}$$

where the one-directional implication from (MF) to (FOSD) is strict. See de Castro (2009) for an example of a distribution satisfying (FOSD) but not (MLRP).

Remark 3.3. (MLRP) is equivalent to (MF) but strictly stronger than (FOSD). Thus if a joint distribution F satisfies (MLRP) then it must satisfy (FOSD), but F can satisfy (FOSD) and fail (MLRP). On the other hand, a conditional distribution $F_{X|\theta}$ that satisfies (FOSD) for every completion to a joint distribution F (i.e., for every choice of prior F_θ) must also satisfy (MLRP). So "FOSD for every prior" is equivalent to MLRP, while "FOSD for some prior" is weaker.

We've already established the equivalence between (A) and (MLRP) in Proposition 9. Since (FOSD) is necessary for (MF), clearly (MF) implies (FOSD). The following result proves equivalence of (MLRP) and (MF).

Proposition 13 (Milgrom (1981))**.** *x is more favorable than x' if and only if for every $\theta > \theta'$,*

$$\frac{f_{X|\theta}(x \mid \theta)}{f_{X|\theta}(x' \mid \theta)} \geq \frac{f_{X|\theta'}(x \mid \theta')}{f_{X|\theta'}(x' \mid \theta')} \tag{3.5}$$

Proof. We will first show that if (3.5) is satisfied at every $\theta > \theta'$, then x must be more favorable than x'. Fix any prior F_θ and parameter $\theta^* \in \Theta$. If $F_\theta(\theta^*) \in \{0, 1\}$ then the conclusion is trivially reached. So suppose $F_\theta(\theta^*) \in (0, 1)$.

For any $\theta \leq \theta^*$ and $\tilde{\theta} > \theta^*$, (3.5) implies

$$\frac{f(x \mid \tilde{\theta})}{f(x \mid \theta)} \geq \frac{f(x' \mid \tilde{\theta})}{f(x' \mid \theta)}$$

where we omit subscripts on the densities here and elsewhere in the proof to ease notation. Integrating over all $\tilde{\theta}$ such that $\tilde{\theta} > \theta^*$ (with respect to the prior distribution F_θ), we obtain

$$\frac{\int_{\tilde{\theta} > \theta^*} f(x \mid \tilde{\theta}) dF_\theta(\tilde{\theta})}{f(x \mid \theta)} \geq \frac{\int_{\tilde{\theta} > \theta^*} f(x' \mid \tilde{\theta}) dF_\theta(\tilde{\theta})}{f(x' \mid \theta)}$$

or equivalently

$$\frac{f(x \mid \theta)}{\int_{\tilde{\theta}>\theta^*} f(x \mid \tilde{\theta})dF_\theta(\tilde{\theta})} \leq \frac{f(x' \mid \theta)}{\int_{\tilde{\theta}>\theta^*} f(x' \mid \tilde{\theta})dF_\theta(\tilde{\theta})}.$$

Integrating over all θ such that $\theta \leq \theta^*$, we obtain

$$\frac{\int_{\theta\leq\theta^*} f(x \mid \theta)dF_\theta(\theta)}{\int_{\tilde{\theta}>\theta^*} f(x \mid \tilde{\theta})dF_\theta(\tilde{\theta})} \leq \frac{\int_{\theta\leq\theta^*} f(x' \mid \theta)dF_\theta(\theta)}{\int_{\tilde{\theta}>\theta^*} f(x' \mid \tilde{\theta})dF_\theta(\tilde{\theta})}.$$

Recall that $f(x \mid \theta)f(\theta) = f(\theta \mid x)f(x)$, so the above display implies

$$\frac{\int_{\theta\leq\theta^*} f(\theta \mid x)d\theta}{\int_{\tilde{\theta}>\theta^*} f(\tilde{\theta} \mid x)d\tilde{\theta}} \leq \frac{\int_{\theta\leq\theta^*} f(\theta \mid x')d\theta}{\int_{\tilde{\theta}>\theta^*} f(\tilde{\theta} \mid x')d\tilde{\theta}}$$

or more simply

$$\frac{F(\theta^* \mid x)}{1 - F(\theta^* \mid x)} \leq \frac{F(\theta^* \mid x')}{1 - F(\theta^* \mid x')}$$

Since $\frac{y}{1-y}$ is a strictly increasing function in y, we have $F(\theta^* \mid x) \leq F(\theta^* \mid x')$ as desired.

In the other direction, we will show that if x is more favorable than x', then (3.5) holds everywere. Consider any two parameter values $\theta > \theta'$, and let F_θ be a prior distribution supported on these two points with equal probability on each.

Since by assumption x is more favorable than x', then $F(\theta' \mid x) \leq F(\theta' \mid x')$, implying

$$\frac{F(\theta' \mid x)}{1 - F(\theta' \mid x)} \leq \frac{F(\theta' \mid x')}{1 - F(\theta' \mid x')}$$

or equivalently

$$\frac{f(\theta' \mid x')}{f(\theta \mid x')} \geq \frac{f(\theta' \mid x)}{f(\theta \mid x)}.$$

Applying Bayes' rule again, we can rewrite the above as $\frac{f(x|\theta)}{f(x'|\theta)} \geq \frac{f(x|\theta')}{f(x'|\theta')}$, which is the desired conclusion. ■

Remark 3.4. Milgrom (1981)'s result is not precisely the proposition above, but instead the equivalence between strict MLRP (as defined in Definition 3.2) and a definition of "more favorable" that replaces FOSD with strict FOSD.

Specifically, say that F strictly first-order stochastically dominates $\widetilde{F}$ if $F(\theta) \leq \widetilde{F}(\theta)$ everywhere with strict inequality at some θ. (Equivalently, $\int u(\theta)dF(\theta) > \int u(\theta)d\widetilde{F}(\theta)$ for every strictly increasing function $u : \mathbb{R} \to \mathbb{R}$.) Say that x is strictly more favorable than x' if for every prior distribution F_θ, the posterior distribution $F_{\theta|X}(\cdot \mid x)$ strictly first-order stochastically dominates $F_{\theta|X}(\cdot \mid x')$. Then, by substituting strict inequalities in place of weak inequalities in the proof above where appropriate, we can conclude that $\{f_{X|\theta}(\cdot \mid \theta)\}$ satisfies strict MLRP if and only if x is strictly more favorable than x'.[2]

[2] Indeed, the same proof demonstrates a stronger (if slightly more cumbersome to state) result: If and only if $\{f_{X|\theta}(\cdot \mid \theta)\}$ satisfies strict MLRP, then $F_{\theta|X}(\theta \mid x) < F_{\theta|X}(\theta \mid x')$ at every θ such that $0 < F(\theta) < 1$.

We conclude by briefly summarizing other notions of positive dependence and placing the above properties relative to these.

Positive covariance (C): $Cov(X, \theta) \geq 0$

Positive quadrant dependence (QD): $Cov(g(X), h(\theta)) \geq 0$ for all non-decreasing functions g and h

Association (As): $Cov(g(X, \theta), h(X, \theta)) \geq 0$ for all non-decreasing functions g and h

Left-Tail Decreasing (LT): For all x, $F_{X|\theta}(X \leq x \mid \theta \leq t)$ is non-increasing in t, and for all t, $F_{\theta|X}(\theta \leq t \mid X \leq x)$ is non-increasing in x.

Inverse Hazard Rate Decreasing (IHR): For all x, $F_{X|\theta}(x \mid t)/f_{X|\theta}(x \mid t)$ is non-increasing in t, and for all t, $F_{\theta|X}(t \mid x)/f_{\theta|X}(t \mid x)$ is non-increasing in x.

These properties are extensively studied in, for example, Lehmann (1966), Esary, Proschan and Walkup (1967), de Castro (2009), and Chapter 3 of Balakrishna and Lai (2009). The following chain of implications is summarized in de Castro (2009):

Theorem 3.1. (A) $\iff$ (MLRP) $\implies$ (IHR) $\implies$ (FOSD) $\implies$ (LT) $\implies$ (As) $\implies$ (QD) $\implies$ (C)

Thus the standard properties of affiliation and MLRP are in fact strong, implying all of the other properties but not in general implied by them. These properties are equivalent to one another in the special case in which the two variables are jointly normal.

EXERCISE 3.4 (G). *Suppose $(X_1, \ldots, X_n)$ are jointly normal and exchangeable, where $\sigma^2 = Var(X_i)$ for each i, and $\rho = Cov(X_i, X_j)$ for each pair of indices i, j. Prove that these variables are affiliated if and only if $\rho \geq 0$.*

HINT. Use the fact given in Exercise 2.10.

3.3 When These Conditions Fail

An example from Lagziel and Lehrer (2019) demonstrates the kind of counterintuitive result that can hold in settings where (A) and (MLRP) fail.

An editor chooses which papers to publish. Papers have unknown quality graded on a 9-point scale (A+, A, A-, B+, B, B-, C+,C,C-), whose prior distribution is given in Figure 3.2.

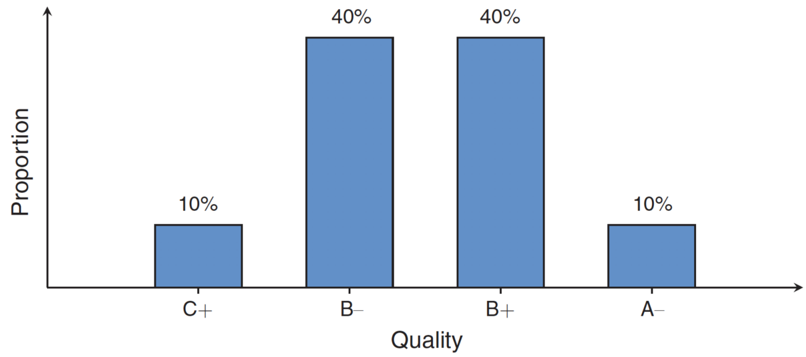

Figure 3.2: Distribution of Papers' Quality

The editor learns about quality via a noisy refereeing process, which generates an unbiased signal X about the paper. The realization of X is equal to the true quality with probability 0.8, and otherwise exactly two levels higher or lower than the true quality (each with probability 0.1). The distribution of X is reported in Figure 3.3:

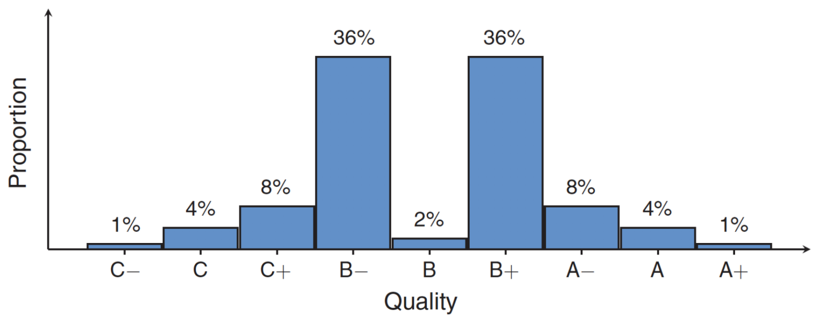

Figure 3.3: Distribution of Referee Signal

The editor chooses a threshold and accepts all papers whose expected quality (given the referee's report) exceeds this threshold. Intuitively, we may expect that the editor faces a tradeoff between publishing more papers versus publishing higher quality papers, where a higher threshold corresponds to publishing fewer but higher quality papers.

But observe that if the editor chooses to publish only papers with an expected quality that (weakly) exceeds A (i.e., the top-rated 5% of papers), then the expected value of the published work is close to $B+$. If the editor lowers the bar to $A-$ (i.e., the top-rated 13%), then the expected value of the published work *increases* to $A-$. Not only are more papers published, but their expected quality is higher.

In this example, we have $\mathbb{E}(\theta \mid X = x) < \mathbb{E}(\theta \mid X = x')$ even while $x > x'$, so clearly the posterior belief at x' does not first-order stochastically dominate the posterior belief at x. Chambers and Healy (2011) demonstrate an even stronger reversal by constructing signals such that the posterior belief at the lower signal realization first-order stochastically dominates the posterior belief at the higher signal realization. Notably, their result relies on natural-seeming signals that satisfy various reasonable properties.

Theorem 3.2. *For every non-degenerate, bounded θ there exists a signal structure X and two signal realizations $x' > x$ such that $f(\theta \mid X = x')$ is strictly first-order stochastically dominated by $f(\theta \mid X = x)$. Furthermore, X can be chosen to have the following properties: i) X is an additive signal structure, and ii) $e := X - \theta$ is mean-zero, symmetric, quasiconcave, and has bounded support.*

See Heinsalu (2020) for a strengthening of Lagziel and Lehrer (2019)'s example using this result, in which lowering the threshold not only increases the expected quality, but results in a quality distribution for published papers that first-order stochastically dominates the one that would obtain at the higher threshold.

3.4 Additional Exercises

Exercise 3.5 (G). *Let $Z_1, \dots, Z_n$ be affiliated and let $h : \mathbb{R}^n \to \mathbb{R}$ be any function that is nondecreasing in each of its coordinates. Prove that the function*

$$\mathbb{E}(h(Z_1, \dots, Z_n) \mid Z_1 = z_1)$$

is nondecreasing in z_1.

Exercise 3.6 (G). *Let X be any real-valued random variable and let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be bounded nondecreasing functions. Prove that $Cov(f(X), g(X)) \geq 0$. (Do not apply the FKG inequality.)*

Hint. There are at least two short proofs, one that uses Fubini's theorem and the fact that $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ for any independent random variables X and Y, and another which relies entirely on elementary (if not obvious) arguments.

Exercise 3.7 (G). *Ann and Bob share the same prior p over an unknown real-valued state θ, and observe a common realization of the signal X, but disagree about the distribution of S. Ann believes that $X = \theta + \varepsilon$, where $\theta \perp\!\!\!\perp \varepsilon$ and ε is a real-valued noise term with density f_ε. Bob believes that $X = \theta + \varepsilon + \Delta$ for some $\Delta > 0$. That is, Ann perceives Bob as adding Δ to the realization of the signal, while Bob perceives Ann as subtracting Δ from the realization of the signal.*

Let f_A denote the joint density of (θ, X) according to Ann's model and f_B denote the joint density according to Bob's model, with $\mathbb{E}^A$ and $\mathbb{E}^B$ denoting their respective expectation operators. Impose the monotone likelihood-ratio property on $\{f_A(\cdot \mid \omega)\}$, that is,

$$\frac{f_A(x' \mid \theta')}{f_A(x \mid \theta')} \geq \frac{f_A(x' \mid \theta)}{f_A(x \mid \theta)} \quad \forall x' > x, \theta' > \theta$$

(a) Prove that $\{f_B(\cdot \mid \theta)\}$ also satisfies MLRP.

(b) Prove that $\mathbb{E}^A[\mathbb{E}^B[\theta \mid X]]$ is decreasing in Δ, and interpret this result.

(c) *Suppose that Ann and Bob now additionally observe a common vector of iid signals* $(Y_1, Y_2, \ldots, Y_N)$ *where each* $Y_i = \theta + \delta_i$ *with* $\theta \perp\!\!\!\perp \delta_i$ *and* δ_i *are iid across signals. Prove that*

$$\mathbb{E}^A[\mathbb{E}^B[\theta \mid X, Y_1, \ldots, Y_N]] \leq \mathbb{E}^A[\mathbb{E}^B[\theta \mid X, Y_1, \ldots, Y_N, Y_{N+1}]]$$

for every $N \geq 1$*. Again, interpret the result.*