Introduction to Probability Theory for Graduate Economics

Brent Hickman

January 27, 2015

3 Jointly Distributed RVs

Up to this point, we have only spoken of RVs as being single-dimensional objects. We will now turn to the case of vector-valued random variables, which are functions mapping a sample space into $\mathbb{R}^n$. Thus, when we talk about the distribution or density of $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, we refer to n-dimensional real-valued functions. Alternatively, one could think of the individual elements X_i as distinct univariate random variables jointly sharing some connection through the distribution of $\mathbf{X}$. For this reason, we call the distribution and density of $\mathbf{X}$ the "joint" distribution and the "joint" density of the X_is.

Example 1 The following is an example of a bivariate, discrete probability distribution for a RV $\mathbf{X} = (X_1, X_2)$, where $X_i \in \{0, 1, 2, 3\}$:

		x_2				
		0	1	2	3	
	0	0.000	0.048	0.090	0.064	0.210
x_1	1	0.048	0.180	0.190	0.012	0.430
	2	0.096	0.192	0.000	0.000	0.288
	3	0.064	0.000	0.008	0.000	0.072
		0.216	0.420	0.288	0.076	1.000

Within the 4×4 matrix above, the ij^{th} entry represents $\Pr[X_1 = i \cap X_2 = j]$. It's important to remember that these are *joint* probabilities, not conditional probabilities. Notice that the event $X_1 = 0$ can be broken up into four mutually exclusive and exhaustive events of the form $(X_1 = 0 \cap X_2 = i), \quad i = 0, 1, 2, 3$. Thus, by the law of total probability we can sum over the probabilities of these events to get the **marginal probability** $\Pr[X_1 = 0]$. The numbers in the margins represent the unconditional probabilities on X_1 and X_2 and they are referred to as the **marginal distributions**. Conditional probabilities can be computed by dividing joint probabilities by marginal probabilities. For example,

$$\Pr[X_1 = 1 | X_2 = 1] = \frac{\Pr[X_1 = 1 \cap X_2 = 1]}{\Pr[X_2 = 1]} = \frac{0.18}{0.42} \approx 0.43. \blacksquare$$

It is generally easiest to talk about probabilities in terms of a **joint CDF**.

Definition 1 *For an n-dimensional RV* $\mathbf{X}$*, the joint cumulative distribution function of its n RVs is the function defined by*

$$F(x_1, \ldots, x_n) = \Pr[X_1 \leq x_1, \ldots, X_n \leq x_n]$$

.

Theorem 1 *For an n-dimensional RV* $\mathbf{X}$*, the joint cumulative distribution function* $F(x_1, \ldots, x_n)$ *satisfies the following properties:*

1. $\lim_{x_i \to -\infty} F(x_1, \ldots, x_n) = 0 \;\; \forall i$
2. $\lim_{\mathbf{X} \to (\infty_1, \ldots, \infty_n)} F(x_1, \ldots, x_n) = 1$
3. $\lim_{h \to 0^+} F(x_1 + h, \ldots, x_n) = \ldots = \lim_{h \to 0^+} F(x_1, \ldots, x_n + h) = F(x_1, \ldots, x_n)$ *(coordinatewise right-continuity)*

The above three properties are necessary in order for F to assign valid probabilities to all events, *but they are not sufficient by themselves*. Recall that a single dimensional CDF must also be non-decreasing. Similarly, in $\mathbb{R}^n$ there is another monotonicity property which also must be satisfied, but it is more complicated now. The following theorem characterizes this analogous monotonicity condition for the bivariate case:

Theorem 2 *A function* $F(x_1, x_2)$ *is a bivariate CDF if and only if it satisfies the properties in Theorem 1, along with the following property:*

$$F(b, d) - F(b, c) - F(a, d) + F(a, c) \geq 0 \;\text{ for all }\; a < b, \; c < d.$$

SKETCH PROOF: The reasoning behind the theorem can be easily seen by representing $\mathbb{R}^2$ as a Venn diagram and by depicting the event $A = (a, b] \times (c, d]$ as a rectangle which does not include its southwestern boundary. Graphically representing the probability $F(b,d) - F(b,c) - F(a,d) + F(a,c)$ is a simple matter of subtracting the regions $x_1 \leq b, x_2 \leq c$ and $x_1 \leq a, x_2 \leq d$ from the region $x_1 \leq b, x_2 \leq d$ and then adding back in the region $x_1 \leq a, x_2 \leq c$. As it turns out, $\Pr[a < X_1 \leq b, c < X_2 \leq d] = F(b,d) - F(b,c) - F(a,d) + F(a,c)$, which must be non-negative. ■

In order to see what's going on in the previous theorem, recall that for a one-dimensional CDF, monotonicity means that $F(x) = Pr[X \leq x] \leq F(x') = Pr[X \leq x']$ whenever $x < x'$. For higher dimensions, the intuition behind the monotonicity condition is the same:

Theorem 3 *A function $F : \mathbb{R}^n \to [0,1]$ is a joint CDF if and only if it satisfies the properties in Theorem 1, and in addition,*

$$F(x_1, \ldots, x_n) \leq F(x'_1, \ldots, x'_n)$$

whenever $x_i \leq x'_i$, $i = 1, \ldots, n$ and $x_i < x'_i$ for some $i \in \{1, \ldots, n\}$.

Corollary 1 *If $F : \mathbb{R}^n \to [0,1]$ is a joint CDF then for any two sets $A, A' \subseteq \mathbb{R}^n$, if $A \subseteq A'$ then the probability assigned to event A' under F will be no less than the probability assigned to event A.*

As in the univariate case, we say that an n-dimensional RV $\mathbf{X} = (X_1, \ldots, X_n)$ is a **multivariate continuous RV** if its joint CDF is absolutely continuous with respect to Lebesgue measure. This is equivalent to requiring that none of the individual X_is have mass-points. For continuous RVs, we can define a **joint PDF**.

Definition 2 *The joint PDF $f : support(\mathbf{X}) \to \mathbb{R}_+$ of an n-dimensional continuous RV is defined by the following:*

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \ldots, t_n) dt_1 \cdots dt_n, \quad \forall \mathrm{x} \in \mathbb{R}^n.$$

Joint PDFs must satisfy similar properties as univariate PDFs.

Theorem 4 *A function $f(x_1, \ldots, x_n)$ is a joint PDF of an n-dimensional RV $\mathbf{X}$ if and only if*

1. $f(x_1, \ldots, x_n) \geq 0, \quad \forall \mathrm{x} \in support(\mathbf{X}) \subseteq \mathbb{R}^n$ *and*
2. $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_1 \cdots dx_n = 1.$

If the joint PDF of $\mathbf{X} \in \mathbb{R}^n$ exists, then the monotonicity property mentioned in Theorems 2 and 3 has another intuitive representation:

Corollary 2 *Suppose that an n-dimensional RV $\mathbf{X} = (X_1, \ldots, X_n)$ is absolutely continuous. Then a function $F : \mathbb{R}^n \to [0,1]$ is a joint CDF of $\mathbf{X}$ if and only if it satisfies the properties in Theorem 1, as well as the* *following:*

$$\frac{d^n F(x_1,\ldots,x_n)}{dx_1\cdots dx_n} \geq 0,\ \forall (x_1,\ldots,x_n)\in support(\mathbf{X}).$$

In other words, the joint PDF of $\mathbf{X}$ *is non-negative everywhere on its support.*

As with the discrete example above, knowledge of the joint distribution allows one to derive the marginal distributions and densities of the component RVs. The only difference with continuous RVs is that now we integrate over a continuum rather than summing over a discrete set.

Definition 3 *If* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ *is an n-dimensional RV with joint CDF* $F(x_1,\ldots,x_n)$*, then the marginal CDF of* X_j *is*

$$F_j(x_j) = \lim_{x_i\to\infty,\, i\neq j} F(x_1,\ldots,x_n).$$

If $\mathbf{X}$ *is continuous then the marginal PDF of* X_j *is derived by integrating out all other components:*

$$f_j(x_j) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\ldots,x_j,\ldots,x_n)dx_1\cdots dx_{j-1}dx_{j+1}\cdots dx_n.$$

3.1 Independent RVs

Recall from Chapter 1 that if A and B are two independent events in some sample space S, then it will be true that$P(A\cap B) = P(A)P(B)$, $P(A|B) = P(A)$ and $P(B|A) = P(B)$. Random variables can also be thought of as independent if joint probabilities share a similar relationship with marginal probabilities:

Definition 4 *Random variables* $X_1, X_2, \ldots, X_n$ *are said to be independent if for every* $a_i < b_i$*,*

$$\Pr[a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n] = \prod_{i=1}^{n} \Pr[a_i \leq X_i \leq b_i].$$

This statement can be reformulated in terms of joint distributions and densities fairly straightforwardly:

Theorem 5 *The following statements are equivalent for RVs* $X_1, X_2, \ldots, X_n$ *having joint CDF* F *and joint PDF* f*:*

1. *The* X_i*s are independent*

2. $F(x_1,\ldots,x_n) = F_1(x_1)\cdots F_n(x_n)$

3. $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$

4. *The following two statements are both true:*

   A. *The support set* $\{(x_1, \dots, x_n) | f(x_1, \dots, x_n) > 0\}$ *is a Cartesian product and*

   B. *The joint pdf can be factored into the product of univariate functions of* $x_1, \dots, x_n$, $f(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n)$.

Exercise 1 Prove that if X and Y are independent RVs, then

$$\mathrm{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y)dydx = \mathrm{E}[X]\mathrm{E}[Y]. \blacksquare$$

3.2 Conditional Distributions

Recall from Chapter 1 the conditional probability rule that for two events A and B, we have $P(A|B) = \frac{P(A \cap B)}{P(B)}$. Once again, a similar concept holds for jointly-distributed RVs in terms of joint and marginal probabilities.

Definition 5 *For any two RVs* X_1 *and* X_2 *having joint PDF* $f(x_1, x_2)$ *and marginal PDFs* f_1 *and* f_2*, the conditional PDF of* X_2 *given that* $X_1 = x_1$ *is*

$$f(x_2|x_1) = \begin{cases} \frac{f(x_1,x_2)}{f_1(x_1)} & \textit{for } x_1 \textit{ s.t. } f_1(x_1) > 0, \\ 0 & \textit{otherwise.} \end{cases}$$

Exercise 2 Prove that if X_1 and X_2 are independent RVs, then $f(x_2|x_1) = f_2(x_2)$. $\blacksquare$

Note that for jointly-distributed X_1 and X_2, a given value x_1 defines a potentially distinct distribution $f(x_2|x_1)$ over outcomes of X_2. This concept is easily generalizable to higher dimensions, where one could talk about conditional distributions like $f(x_1, \dots, x_k | x_{k+1}, \dots, x_n) = \frac{f(x_1,\dots,x_n)}{f_{k+1}(x_{k+1}) \cdots f_n(x_n)}$. It is also fairly straightforward to define conditional CDFs in similar fashion.

Now that we have defined the notion of a conditional distribution, we can perform familiar operations with it like computing **conditional expectations** and **conditional variance**.

Definition 6 *If* X *and* Y *are two jointly distributed RVs, then the conditional expectation of* Y*, given that* $X = x$ *is*

$$\mathrm{E}(Y|x) = \int_{-\infty}^{\infty} yf(y|x)dy$$

Definition 7 *If X and Y are two jointly distributed RVs, then the conditional variance of Y, given that $X = x$ is*

$$\text{Var}(Y|x) = \text{E}\{[Y - \text{E}(Y|x)]^2|x\}.$$

Theorem 6 *If X and Y are two jointly distributed RVs, then*

$$\text{E}(Y) = \text{E}_X[\text{E}_Y(Y|X)].$$

The above result is sometimes referred to as the **law of iterated expectations**, and in economics it has important applications to dynamic stochastic decision processes. One notable example is asset pricing.

Example 2 The **Lucas tree model** is a basic framework for thinking about asset pricing. In the model, there is a set of risk-neutral consumers who each own shares in a productive tree which produces random outputs each period. Assume that states of the world are H (high output) and L (low output), and that outputs in consecutive periods are related by some joint distribution. One common assumption is that outputs form a **Markov chain**, which will be covered in chapter 5. The objective of the model is to study how shares of the productive tree are priced. It can be shown that the price of a share of the tree in period t in state $i = \{H, L\}$ is

$$p_{it} = \beta \text{E}_t[p_{t+1} + y_{t+1}],$$

where the E_t operator denotes an expectation conditional on information available in period t, β is the discount factor, p_{t+1} is the prevailing price of the asset next period (given tomorrow's state) and y_{t+1} is the output next period (given tomorrow's state).

However, the problem is that the above expression implicitly involves many nested expectations involving information which is not available in period t. For example, we could expand it for 3 periods into the future to get

$$p_{it} = \beta \text{E}_t \left[\beta \text{E}_{t+1} \left[\beta \text{E}_{t+2} \left[\beta \text{E}_{t+3}[p_{t+4} + y_{t+4}] + y_{t+3}\right] + y_{t+2}\right] + y_{t+1}\right].$$

The law of iterated expectations implies that

$$\text{E}_{t+k}[p_{t+k+2} + y_{t+k+2}] = \text{E}_{t+k}\left[\text{E}_{t+k+1}[p_{t+k+2} + y_{t+k+2}]\right],$$

which allows one to write down a closed-form expression $p_{it} = \text{E}_t \left[\sum_{i=1}^{\infty} \beta^i y_{t+i}\right]$. ■

Theorem 7 *If X and Y are two jointly distributed RVs, then*

$$\text{Var}(Y) = \text{E}_X[\text{Var}_Y(Y|X)] + \text{Var}_X[\text{E}_Y(Y|X)].$$

The two previous theorems say some interesting things about the relation between conditional moments and unconditional moments. The second is particularly useful, as it indicates that, on average (over X), the conditional variance will be smaller than the unconditional variance. This highlights the usefulness of conditioning in decision problems: agents can often use information not only to refine their expectations, but also to eliminate some of the noise from their choice process.

Exercise 3 Show that if X and Y are independent RVs, then $\text{E}(Y|X) = \text{E}(Y)$ and $\text{E}(X|Y) = \text{E}(X)$. ■

In Chapter 2 we learned that many univariate distributions can be expressed in terms of moment-generating functions. As it turns out, many joint distributions can also be characterized by **joint moment-generating functions (joint MGFs)**.

Definition 8 *The joint MGF of* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$*, if it exists, is given by*

$$M_{\mathbf{X}}(t_1, \ldots, t_n) = \text{E}[e^{t_1 x_1 + \ldots + t_n x_n}].$$

The properties of joint MGFs are similar to univariate MGFs. They can be used to compute mixed moments of the form $\text{E}[X_i^r X_j^s]$ by differentiating $M_{\mathbf{X}}(t_1, \ldots, t_n)$ r times with respect to t_i and s times with respect to t_j, and then setting all $t_k = 0$. Thus, the joint MGF also uniquely determines the joint distribution by characterizing all of its moments.

Another use for joint MGFs is in computing univariate **marginal MGFs**: $M_i(t_i) = M_{\mathbf{X}}(0, \ldots, 0, t_i, 0, \ldots, 0)$. We can also use the joint MGF to determine whether a set of RVs is independent:

Theorem 8 *If* $M_{X,Y}(t_1, t_2)$ *exists, then the RVs X and Y are independent if and only if*

$$M_{X,Y}(t_1, t_2) = M_X(t_1) M_Y(t_2).$$

Example 3 ***Conditional Expectation:*** *Winner-Pay Auctions with Symmetric, Independent Private Values (IPV)*

Consider a game in which there are N bidders competing for the right to purchase some object and let $M \equiv (N-1)$ denote the number of opponents each bidder faces. Under the symmetric IPV assumption, each bidder has a privately known value $v \in [0, \overline{v}]$, his subjective reservation price for the object, which is an iid draw from a distribution $F(v)$. Assume F is atomless and differentiable. Assume also that only the winner actually has to pay anything to the auctioneer, so that the payoff to the winner is his private value minus the price, and losers get a payoff of zero. A Bayesian-Nash equilibrium of this model is a function β mapping private values v into bids b, such that each player's expected payoff is maximized by bidding $b = \beta(v)$, given that his opponents behave similarly. It can be shown *a priori* that under the above assumptions β exists, it is strictly increasing and differentiable in v, and $\beta(0) = 0$.

a. Show that in a symmetric equilibrium of the first-price auction (where the winner pays his own bid), bidders will choose a bid equal to the conditional expectation of the highest opponent private value, given that it is less than their own. In other words, show that

$$\beta(v) = \mathrm{E}[\max(\mathbf{V}) | \max(\mathbf{V}) \leq v],$$

where $\mathbf{V}$ is a random $(N-1)$-vector of private values. Is this equilibrium bid less than or greater than ones private value?

b. Using a probability argument, show that $\beta(v)$ converges pointwise to $\overline{\beta}(v) = v$ (the 45°-line) as $N \to \infty$. (*HINT*: What happens to the above conditional expectation as $N \to \infty$? Show that for $N < N'$, the conditional distribution of $[\max(\mathbf{V}) | \max(\mathbf{V}) \leq v, N']$ stochastically dominates that of $[\max(\mathbf{V}) | \max(\mathbf{V}) \leq v, N]$.)

c. In a second-price auction (where the winner pays the highest competing bid), there is a dominant strategy equilibrium in which everyone simply bids $b = v$. Show that the expected revenue to the seller under this equilibrium is the same as in the symmetric equilibrium of a first-price auction. This is known as the **Revenue Equivalence Theorem**. (*HINT*: Show that when someone wins, the expected payment is also the conditional expectation of the highest opponent private value, given that it is less than ones own.)

a.) Since the equilibrium is monotonic (*i.e.*, people with higher private values bid more), one only wins when ones private value exceeds the highest order statistic V_M of all opponent private values. Recall from Exercise 10 in Chapter 1 that $\Pr[V_M \leq v]$ is simply

$$G_M(v) = F(v)^M.$$

Note also that G_M is a valid CDF, so we could find the density of V_M by $g_M(v) = G'_M(v)$.

When someone with a private value of v wins the auction, their payoff is simply $(v - b)$. Consider such a bidder who is choosing a bid of b so as to maximize his expected payoff. In equilibrium, his decision problem is

$$\max_b G_M\left(\beta^{-1}(b)\right)(v-b),$$

with FOC

$$\frac{g_M\left(\beta^{-1}(b)\right)}{\beta'\left(\beta^{-1}(b)\right)}(v-b) - g_M\left(\beta^{-1}(b)\right) = 0.$$

In equilibrium the bidder chooses $\beta(v) = b$, so we can substitute and rearrange to get

$$\beta'(v)\,G_M(v) + g_M(v)\,\beta(v) = v g_M(v).$$

Note that the LHS of the above expression is the same as $d\beta(v)G_M(v)/dv$, so by using the boundary condition $\beta(0) = 0$, we can integrate (or anti-differentiate) both sides of the FOC to get

$$\begin{aligned} \beta(v)G_M(v) &= \int_0^v u g_M(u)\,du \\ \Rightarrow\ \beta(v) &= \int_0^v u\frac{g_M(u)}{G_M(v)}du = \mathrm{E}[V_M|V_M \le v], \end{aligned} \tag{1}$$

where the final equality comes from the fact that $\frac{g_M(u)}{G_M(v)}$ is the conditional density of V_M, given that $V_M \le v$. This expectation is strictly less than v, as the above density places mass on values less than v.

b.) Fix some $v \in [0, \overline{v}]$ and define a RV $U = V_M$, but only on the restricted sample space $V_M \in [0, v]$. The distribution of U is the anti-derivative of the conditional density in equation (1) above,

$$F_U(u; v, N) = \frac{G_M(u)}{G_M(v)} = \frac{F(u)^{N-1}}{F(v)^{N-1}}.$$

I claim that for $N', N'' \in \mathbb{N}$, if $N' < N''$, then $F_U(u; v, N'')$ FOSD $F_U(u; v, N')$. If we let $k = N'' - N'$, the truth of this claim can be seen by noting that

$$\frac{F(u)^{N''-1}}{F(v)^{N''-1}} = \left(\frac{F(u)}{F(v)}\right)^k \frac{F(u)^{N'-1}}{F(v)^{N'-1}} \le \frac{F(u)^{N'-1}}{F(v)^{N'-1}},$$

where the inequality follows from the fact that $F(u)/F(v) \le 1$ for any $u \in [0, v]$. This tells us that as more bidders enter the auction, the conditional expectation in (1) will rise. Moreover, since $F(u)/F(v) < 1$ for any $u < v$, it follows that

$$\lim_{N\to\infty} \frac{F(u)^{N-1}}{F(v)^{N-1}} = \lim_{N\to\infty}\left(\frac{F(u)}{F(v)}\right)^{N-1} = 0$$

for any $u < v$. In other words, not only is more mass being shifted to higher realizations of U as N gets large, but in the limit, *all* of the mass is shifted to the upper bound, v. Therefore, we have

$\lim_{N\to\infty}\beta(v) = \lim_{N\to\infty} \mathrm{E}[V_M|V_M \leq v] = v$ and we are done.

c.) Since the symmetric equilibrium of the second-price auction has everyone bidding $b = v$, it is still true that in equilibrium a bidder only wins if he has the highest private value. Moreover, the price he pays is once again the RV U described in the answer to part b., so his expected payment is $\mathrm{E}[V_M|V_M \leq v]$. Since every bidder in each auction has the same probability of winning and the same expected payment conditional on winning, it follows that the first-price and second-price auctions with symmetric IPV are **revenue equivalent** in expectation. ■

3.3 Covariance, Correlation and Affiliation

The joint distribution among RVs provides a comprehensive characterization of the relationship between them, but this information expressed in terms of a high-dimensional function can often be difficult to interpret. Another measure of the relation between two jointly distributed RVs is **covariance**. Covariance is a more superficial concept of stochastic relatedness, but at the same time, its simplicity makes it easy to understand and interpret.

Definition 9 *The covariance of two jointly distributed RVs X and Y is given by*

$$\mathrm{COV}(X,Y) = \sigma_{XY} \equiv \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy,$$

where μ_X and μ_Y are, respectively, the unconditional expectations of X and Y.

Note from the definition above that $\mathrm{COV}(X,X) = \mathrm{E}[(X-\mu_X)^2] = \mathrm{VAR}(X)$.

Exercise 4 Show that if X and Y are joint RVs, then

$$\mathrm{COV}(X,Y) = \mathrm{E}(XY) - E(X)\mathrm{E}(Y)$$

and

$$\mathrm{COV}(X,Y) = 0$$

when X and Y are independent. ■

Although covariance is a useful measure of the effect of X on the conditional distribution of Y, one problem is that it is expressed in the units of X and Y, which may not be desirable if comparisons are to be made with covariances of other RVs, say W and Z. A unit-free measure of covariance can be found by computing the **correlation coefficient**.

Definition 10 *If X and Y are joint random variables with unconditional variances of σ_X^2 and σ_Y^2 and covariance σ_{XY}, then the correlation coefficient of X and Y is*

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Theorem 9 *If ρ_{XY} is the correlation coefficient of RVs X and Y, then*

1. $-1 \leq \rho_{XY} \leq 1$,
2. *X and Y independent* $\Rightarrow \rho_{XY} = 0$ *and*
3. $\rho_{XY} \in \{-1, 1\} \quad \Leftrightarrow \quad Y = aX + b$ *for some* $a \neq 0$ *and* b.

When $\rho_{XY} = 0$, then X and Y are said to be **uncorrelated**; otherwise, they are **correlated**. Note that X and Y being uncorrelated is a necessary condition for independence, but it is NOT sufficient. For example, if $X \sim N(\mu, \sigma)$, then X and $Y = X^2$ are clearly dependent RVs, although it turns out that they are uncorrelated. This simple example highlights the fact that correlation is in some sense a measure of linear relations among RVs.

In microeconomic models, there is a strong form of positive correlation for joint RVs which occasionally comes up. It is called **stochastic affiliation**.

Definition 11 *RVs $X_1, \ldots, X_n$ are said to be affiliated if their joint PDF $f(\mathbf{x})$ is log-supermodular, meaning that for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}_n$, we have*

$$f\left(\mathbf{x} \vee \mathbf{x}'\right) f\left(\mathbf{x} \wedge \mathbf{x}'\right) \geq f(\mathbf{x}) f(\mathbf{x}'),$$

where

$$\mathbf{x} \vee \mathbf{x}' = (\max\{x_1, x_1'\}, \ldots, \max\{x_n, x_n'\})$$

and

$$\mathbf{x} \wedge \mathbf{x}' = (\min\{x_1, x_1'\}, \ldots, \min\{x_n, x_n'\}).$$ [1]

When f is twice continuously differentiable, then equivalently $X_1, \ldots, X_n$ are affiliated iff for all $i \neq j$,

$$\frac{\partial^2}{\partial x_i \partial x_j} \ln f \geq 0.$$

[1] A multivariate real-valued function f is said to be **supermodular** if for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}_n$, we have

$$f\left(\mathbf{x} \vee \mathbf{x}'\right) + f\left(\mathbf{x} \wedge \mathbf{x}'\right) \geq f(\mathbf{x}) + f(\mathbf{x}').$$

f is **log-supermodular** if $\log(f)$ is supermodular.

Intuitively, stochastic affiliation implies that conditional on observing a high (low) outcome of X, the probability of observing a high (low) outcome of Y is increased. Consider two joint RVs X and Y. Affiliation tells us a lot about the conditional distribution $f(Y|x)$. If $x \leq x'$, $\;y \leq y'$, then affiliation implies that

$$f(x',y)f(x,y') \leq f(x',y')f(x,y),$$

which implies

$$\frac{f(x,y')}{f(x,y)} \leq \frac{f(x',y')}{f(x',y)},$$

which can be rewritten as

$$\frac{f(y'|x)f_X(x)}{f(y|x)f_X(x)} = \frac{f(y'|x)}{f(y|x)} \leq \frac{f(y'|x')}{f(y|x')} = \frac{f(y'|x')f_X(x')}{f(y|x')f_X(x')}.$$

This last inequality tells us that the likelihood ratio

$$\frac{f(\cdot|x)}{f(\cdot|x)}$$

is increasing in x, or in other words, it displays the **monotone likelihood ratio property** or MLRP.

Recall from Chapter 2 that the MLRP implies that for all $x \leq x'$, $F(y|x')$ stochastically dominates $F(y|x)$ according to the likelihood ratio order. Recall also that in Chapter 2 we showed that monotone likelihood ratio dominance implies hazard rate dominance, reverse hazard rate dominance and first-order stochastic dominance. This is the sense in which high (low) realizations of X imply a higher probability of high (low) realizations of Y. This is also why affiliation is considered to be a stronger form of correlation, because it gives information on the overall conditional distribution of $Y|x$, rather than just the conditional expectation.

Exercise 5 BIVARIATE NORMAL DISTRIBUTION: Consider $Z = (X,Y)$ distributed as a bivariate normal RV with joint pdf

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right]\right\},$$

where $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$, and σ_{12} is the covariance. Do the following:

a. Graphically depict the level curves of the joint distribution, showing the regression line, and draw some conditional distributions of X for different values of Y.

b. Using the result from Exercise 4, assume that X and Y are independent and demonstrate the above theorems on independence, including the ones related to joint vs. marginal distributions, conditional expectations, the expectation of the product of X and Y, and joint vs. marginal MGFs. For joint normal RVs, the joint MGF is

$$M_{XY}(t_1, t_2) = \exp\left(t_1\mu_1 + t_2\mu_2 + t_1^2\frac{\sigma_1^2}{2} + t_2^2\frac{\sigma_2^2}{2} + t_1 t_2 \sigma_{12}\right)$$

and for univariate normal RVs, the MGF is

$$M_X(t) = \exp\left(t\mu + t^2\frac{\sigma^2}{2}\right).$$

c. Determine whether there is any configuration of the parameters $\mu_1, \mu_2, \sigma_1\sigma_2,$ and σ_{12} for which X and Y are affiliated. ■

References

[1] BAIN, LEE J. and MAX ENGELHARDT (1992): *Introduction to Probability Theory and Mathematical Statistics*, Duxbury.

[2] KRISHNA, VIJAY (2002): *Auction Theory*, Academic Press.

[3] "Multivariate normal distribution." *Wikipedia, The Free Encyclopedia*. 17 Nov 2008, 00:14 UTC. 20 Nov 2008 http://en.wikipedia.org/w/index.php?title=Multivariate_normal_distribution&oldid=252257083.

[4] WILLIAMSON, STEVE (2006): *Notes on Macroeconomic Theory; Chapter 6: Consumption and Asset Pricing*, Typescript, Dept. of Economics, Washington University in St. Louis.