Lecture 9: Independence, conditional independence, conditional distribution

Definition 1.7. Let $(\Omega, \mathcal{F}, P)$ be a probability space.
(i) Let $\mathcal{C}$ be a collection of subsets in $\mathcal{F}$. Events in $\mathcal{C}$ are said to be *independent* if and only if for any positive integer n and distinct events A_1,...,A_n in $\mathcal{C}$,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n).$$

(ii) Collections $\mathcal{C}_i \subset \mathcal{F}$, $i \in \mathcal{I}$ (an index set that can be uncountable), are said to be independent if and only if events in any collection of the form $\{A_i \in \mathcal{C}_i : i \in \mathcal{I}\}$ are independent.
(iii) Random elements X_i, $i \in \mathcal{I}$, are said to be independent if and only if $\sigma(X_i)$, $i \in \mathcal{I}$, are independent.

A useful result for checking the independence of σ-fields.

Lemma 1.3. Let $\mathcal{C}_i$, $i \in \mathcal{I}$, be independent collections of events. Suppose that each $\mathcal{C}_i$ has the property that if $A \in \mathcal{C}_i$ and $B \in \mathcal{C}_i$, then $A \cap B \in \mathcal{C}_i$. Then $\sigma(\mathcal{C}_i)$, $i \in \mathcal{I}$, are independent.

Random variables X_i, $i = 1, ..., k$, are independent according to Definition 1.7 if and only if

$$F_{(X_1,...,X_k)}(x_1, ..., x_k) = F_{X_1}(x_1)\cdots F_{X_k}(x_k), \quad (x_1, ..., x_k) \in \mathcal{R}^k$$

Take $\mathcal{C}_i = \{(a, b] : a \in \mathcal{R}, b \in \mathcal{R}\}$, $i = 1, ..., k$

If X and Y are independent random vectors, then so are $g(X)$ and $h(Y)$ for Borel functions g and h.

Two events A and B are independent if and only if $P(B|A) = P(B)$, which means that A provides no information about the probability of the occurrence of B.

Proposition 1.11. Let X be a random variable with $E|X| < \infty$ and let Y_i be random k_i-vectors, $i = 1, 2$. Suppose that (X, Y_1) and Y_2 are independent. Then

$$E[X|(Y_1, Y_2)] = E(X|Y_1) \text{ a.s.}$$

Proof. First, $E(X|Y_1)$ is Borel on $(\Omega, \sigma(Y_1, Y_2))$, since $\sigma(Y_1) \subset \sigma(Y_1, Y_2)$. Next, we need to show that for any Borel set $B \in \mathcal{B}^{k_1+k_2}$,

$$\int_{(Y_1,Y_2)^{-1}(B)} X dP = \int_{(Y_1,Y_2)^{-1}(B)} E(X|Y_1) dP. \tag{1}$$

If $B = B_1 \times B_2$, where $B_i \in \mathcal{B}^{k_i}$, then $(Y_1, Y_2)^{-1}(B) = Y_1^{-1}(B_1) \cap Y_2^{-1}(B_2)$ and

$$\begin{aligned}
\int_{Y_1^{-1}(B_1)\cap Y_2^{-1}(B_2)} E(X|Y_1)dP &= \int I_{Y_1^{-1}(B_1)} I_{Y_2^{-1}(B_2)} E(X|Y_1)dP \\
&= \int I_{Y_1^{-1}(B_1)} E(X|Y_1)dP \int I_{Y_2^{-1}(B_2)} dP \\
&= \int I_{Y_1^{-1}(B_1)} X dP \int I_{Y_2^{-1}(B_2)} dP \\
&= \int I_{Y_1^{-1}(B_1)} I_{Y_2^{-1}(B_2)} X dP \\
&= \int_{Y_1^{-1}(B_1)\cap Y_2^{-1}(B_2)} X dP,
\end{aligned}$$

where the second and the next to last equalities follow the independence of (X, Y_1) and Y_2, and the third equality follows from the fact that $E(X|Y_1)$ is the conditional expectation of X given Y_1. This shows that (1) holds for $B = B_1 \times B_2$. We can show that the collection $\mathcal{H} = \{B \subset \mathcal{R}^{k_1+k_2} : B \text{ satisfies (1) } \}$ is a σ-field. Since we have already shown that $\mathcal{B}^{k_1} \times \mathcal{B}^{k_2} \subset \mathcal{H}$, $\mathcal{B}^{k_1+k_2} = \sigma(\mathcal{B}^{k_1} \times \mathcal{B}^{k_2}) \subset \mathcal{H}$ and thus the result follows.

The result in Proposition 1.11 still holds if X is replaced by $h(X)$ for any Borel h and, hence,

$$P(A|Y_1, Y_2) = P(A|Y_1) \quad \text{a.s. for any } A \in \sigma(X), \tag{2}$$

if (X, Y_1) and Y_2 are independent.

We say that given Y_1, X and Y_2 are *conditionally independent* if and only if (2) holds.

Proposition 1.11 can be stated as: if Y_2 and (X, Y_1) are independent, then given Y_1, X and Y_2 are conditionally independent.

Conditional distribution

For random vectors X and Y, is $P[X^{-1}(B)|Y = y]$ a probability measure for given y?
The the following theorem shows that there exists a version of conditional probability such that $P[X^{-1}(B)|Y = y]$ is a probability measure for any fixed y.

Theorem 1.7. (i) (Existence of conditional distributions). Let X be a random n-vector on a probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{A}$ be a sub-σ-field of $\mathcal{F}$. Then there exists a function $P(B, \omega)$ on $\mathcal{B}^n \times \Omega$ such that (a) $P(B, \omega) = P[X^{-1}(B)|\mathcal{A}]$ a.s. for any fixed $B \in \mathcal{B}^n$, and (b) $P(\cdot, \omega)$ is a probability measure on $(\mathcal{R}^n, \mathcal{B}^n)$ for any fixed $\omega \in \Omega$.
Let Y be measurable from $(\Omega, \mathcal{F}, P)$ to $(\Lambda, \mathcal{G})$. Then there exists $P_{X|Y}(B|y)$ such that (a) $P_{X|Y}(B|y) = P[X^{-1}(B)|Y = y]$ a.s. P_Y for any fixed $B \in \mathcal{B}^n$, and (b) $P_{X|Y}(\cdot|y)$ is a probability measure on $(\mathcal{R}^n, \mathcal{B}^n)$ for any fixed $y \in \Lambda$.
Furthermore, if $E|g(X, Y)| < \infty$ with a Borel function g, then

$$E[g(X, Y)|Y = y] = E[g(X, y)|Y = y] = \int_{\mathcal{R}^n} g(x, y) dP_{X|Y}(x|y) \quad \text{a.s. } P_Y.$$

(ii) Let $(\Lambda, \mathcal{G}, P_1)$ be a probability space. Suppose that P_2 is a function from $\mathcal{B}^n \times \Lambda$ to $\mathcal{R}$ and satisfies (a) $P_2(\cdot, y)$ is a probability measure on $(\mathcal{R}^n, \mathcal{B}^n)$ for any $y \in \Lambda$, and (b) $P_2(B, \cdot)$ is Borel for any $B \in \mathcal{B}^n$. Then there is a unique probability measure P on $(\mathcal{R}^n \times \Lambda, \sigma(\mathcal{B}^n \times \mathcal{G}))$ such that, for $B \in \mathcal{B}^n$ and $C \in \mathcal{G}$,

$$P(B \times C) = \int_C P_2(B, y) dP_1(y). \tag{3}$$

Furthermore, if $(\Lambda, \mathcal{G}) = (\mathcal{R}^m, \mathcal{B}^m)$, and $X(x, y) = x$ and $Y(x, y) = y$ define the coordinate random vectors, then $P_Y = P_1$, $P_{X|Y}(\cdot|y) = P_2(\cdot, y)$, and the probability measure in (3) is the joint distribution of (X, Y), which has the following joint c.d.f.:

$$F(x, y) = \int_{(-\infty, y]} P_{X|Y}((-\infty, x]|z) dP_Y(z), \quad x \in \mathcal{R}^n, y \in \mathcal{R}^m, \tag{4}$$

where $(-\infty, a]$ denotes $(-\infty, a_1] \times \cdots \times (-\infty, a_k]$ for $a = (a_1, ..., a_k)$.

For a fixed y, $P_{X|Y=y} = P_{X|Y}(\cdot|y)$ is called the conditional distribution of X given $Y = y$.

Two-stage experiment theorem:
If $Y \in \mathcal{R}^m$ is selected in stage 1 of an experiment according to its marginal distribution $P_Y = P_1$, and X is chosen afterward according to a distribution $P_2(\cdot, y)$, then the combined two-stage experiment produces a jointly distributed pair (X, Y) with distribution $P_{(X,Y)}$ given by (3) and $P_{X|Y=y} = P_2(\cdot, y)$.
This provides a way of generating dependent random variables.

Example 1.23. A market survey is conducted to study whether a new product is preferred over the product currently available in the market (old product). The survey is conducted by mail. Questionnaires are sent along with the sample products (both new and old) to N customers randomly selected from a population, where N is a positive integer. Each customer is asked to fill out the questionnaire and return it. Responses from customers are either 1 (new is better than old) or 0 (otherwise). Some customers, however, do not return the questionnaires. Let X be the number of ones in the returned questionnaires. What is the distribution of X?

If every customer returns the questionnaire, then (from elementary probability) X has the binomial distribution $Bi(p, N)$ in Table 1.1 (assuming that the population is large enough so that customers respond independently), where $p \in (0, 1)$ is the overall rate of customers who prefer the new product. Now, let Y be the number of customers who respond. Then Y is random. Suppose that customers respond independently with the same probability $\pi \in (0, 1)$. Then P_Y is the binomial distribution $Bi(\pi, N)$. Given $Y = y$ (an integer between 0 and N), $P_{X|Y=y}$ is the binomial distribution $Bi(p, y)$ if $y \geq 1$ and the point mass at 0 if $y = 0$. Using (4) and the fact that binomial distributions have p.d.f.'s w.r.t. counting measure, we obtain that the joint c.d.f. of (X, Y) is

$$\begin{aligned} F(x, y) &= \sum_{k=0}^{y} P_{X|Y=k}((-\infty, x]) \binom{N}{k} \pi^k (1-\pi)^{N-k} \\ &= \sum_{k=0}^{y} \sum_{j=0}^{\min\{x,k\}} \binom{k}{j} p^j (1-p)^{k-j} \binom{N}{k} \pi^k (1-\pi)^{N-k} \end{aligned}$$

for $x = 0, 1, ..., y$, $y = 0, 1, ..., N$. The marginal c.d.f. $F_X(x) = F(x, \infty) = F(x, N)$. The p.d.f. of X w.r.t. counting measure is

$$\begin{aligned} f_X(x) &= \sum_{k=x}^{N} \binom{k}{x} p^x (1-p)^{k-x} \binom{N}{k} \pi^k (1-\pi)^{N-k} \\ &= \binom{N}{x} (\pi p)^x (1-\pi p)^{N-x} \sum_{k=x}^{N} \binom{N-x}{k-x} \left(\frac{\pi - \pi p}{1-\pi p}\right)^{k-x} \left(\frac{1-\pi}{1-\pi p}\right)^{N-k} \\ &= \binom{N}{x} (\pi p)^x (1-\pi p)^{N-x} \end{aligned}$$

for $x = 0, 1, ..., N$. It turns out that the marginal distribution of X is the binomial distribution $Bi(\pi p, N)$.