2.1 Independence
2.1 Independence

Fix a probability space \((\Omega, \mathcal{F}, P)\).

2 events \(A\) and \(B\) are said to be independent if \(P(A \cap B) = P(A)P(B)\).

2 random variables \(X\) and \(Y\) are said to be independent if for all \(C, D \in \mathcal{R}\),
\[
P(X \in C, Y \in D) = P(X \in C)P(Y \in D).
\]

2 sub-\(\sigma\)-fields \(\mathcal{F}_1\) and \(\mathcal{F}_2\) of \(\mathcal{F}\) are said to be independent if for all \(A \in \mathcal{F}_1\) and \(B \in \mathcal{F}_2\), \(A\) and \(B\) are independent.

Obviously, 2 events \(A\) and \(B\) are independent iff \(1_A\) and \(1_B\) are independent.

The random variables \(X\) and \(Y\) are independent iff \(\sigma(X)\) and \(\sigma(Y)\) are independent.
Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

2 events $A$ and $B$ are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

2 random variables $X$ and $Y$ are said to be independent if for all $C, D \in \mathcal{R}$,

$$\mathbb{P}(X \in C, Y \in D) = \mathbb{P}(X \in C)\mathbb{P}(Y \in D).$$

2 sub-$\sigma$-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ of $\mathcal{F}$ are said to be independent if for all $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$, $A$ and $B$ are independent.

Obviously, 2 events $A$ and $B$ are independent iff $1_A$ and $1_B$ are independent.

The random variables $X$ and $Y$ are independent iff $\sigma(X)$ and $\sigma(Y)$ are independent.
2.1 Independence

\( n \) sub-\( \sigma \)-fields \( \mathcal{F}_i \) of \( \mathcal{F} \) are said to be independent if whenever \( A_i \in \mathcal{F}_i \), \( i = 1, \ldots, n \), we have

\[
P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i).
\]

\( n \) random variables \( X_1, \ldots, X_n \) are said to be independent if \( \sigma(X_1), \ldots, \sigma(X_n) \) are independent.

\( n \) events \( A_1, \ldots, A_n \) are said to be independent if \( 1_{A_1}, \ldots, 1_{A_n} \) are independent.

Compare with the definition of independence in undergraduate probability!
Collections of sets $\mathcal{A}_1, \ldots, \mathcal{A}_n \subset \mathcal{F}$ are said to be independent if whenever $A_i \in \mathcal{A}_i$ and $I \subset \{1, \ldots, n\}$, $\mathbb{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i)$. If each $\mathcal{A}_i$ consists of a single set $A_i$, that is, $\mathcal{A}_i = \{A_i\}$, this definition reduces to the definition of independence of $n$ events we learned in undergraduate probability.

If $\Omega \in \mathcal{A}_i$ for all $i = 1, \ldots, n$, then $\mathcal{A}_1, \ldots, \mathcal{A}_n \subset \mathcal{F}$ are independent iff whenever $A_i \in \mathcal{A}_i$

$$\mathbb{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i).$$
2.1 Independence

Collections of sets $A_1, \ldots, A_n \subset \mathcal{F}$ are said to be independent if whenever $A_i \in A_i$ and $I \subset \{1, \ldots, n\}$, $\mathbb{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i)$. If each $A_i$ consists of a single set $A_i$, that is, $A_i = \{A_i\}$, this definition reduces to the definition of independence of $n$ events we learned in undergraduate probability.

If $\Omega \in A_i$ for all $i = 1, \ldots, n$, then $A_1, \ldots, A_n \subset \mathcal{F}$ are independent iff whenever $A_i \in A_i$

$$\mathbb{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i).$$
Theorem 2.1.7

Suppose $\mathcal{A}_1, \ldots, \mathcal{A}_n \subset \mathcal{F}$ are independent and each $\mathcal{A}_i$ is a $\pi$-system containing $\Omega$, then $\sigma(\mathcal{A}_1), \ldots, \sigma(\mathcal{A}_n)$ are independent.

Proof

Let $A_2, \ldots, A_n$ be sets with $A_i \in \mathcal{A}_i$, and let $F = A_2 \cap \cdots A_n$ and

$$\mathcal{L} = \{A \in \mathcal{F} : P(A \cap F) = P(A)P(F)\}.$$ 

Obviously $\Omega \in \mathcal{L}$. If $A, B \in \mathcal{L}$ with $A \subset B$, then

$$P((B \setminus A) \cap F) = P((B \cap F) \setminus (A \cap F)) = P(B \cap F) - P(A \cap F)$$

$$= P(B)P(F) - P(A)P(F) = P(B \setminus A)P(F),$$

and so $B \setminus A \in \mathcal{L}$. Let $B_k \in \mathcal{L}$ and $B_k \uparrow B$, then

$$P(B \cap F) = \lim_k P(B_k \cap F) = \lim_k P(B_k)P(F) = P(B)P(F),$$

and so $B \in \mathcal{L}$. Thus $\mathcal{L}$ is a $\lambda$-system.
**Theorem 2.1.7**

Suppose $A_1, \ldots, A_n \subset \mathcal{F}$ are independent and each $A_i$ is a $\pi$-system containing $\Omega$, then $\sigma(A_1), \ldots, \sigma(A_n)$ are independent.

**Proof**

Let $A_2, \ldots, A_n$ be sets with $A_i \in A_i$, and let $F = A_2 \cap \cdots \cap A_n$ and

$$
\mathcal{L} = \{ A \in \mathcal{F} : P(A \cap F) = P(A)P(F) \}.
$$

Obviously $\Omega \in \mathcal{L}$. If $A, B \in \mathcal{L}$ with $A \subset B$, then

$$
P((B \setminus A) \cap F) = P((B \cap F) \setminus (A \cap F) = P(B \cap F) - P(A \cap F)
$$

$$
= P(B)P(F) - P(A)P(F) = P(B \setminus A)P(F),
$$

and so $B \setminus A \in \mathcal{L}$. Let $B_k \in \mathcal{L}$ and $B_k \uparrow B$, then

$$
P(B \cap F) = \lim_k P(B_k \cap F) = \lim_k P(B_k)P(F) = P(B)P(F),
$$

and so $B \in \mathcal{L}$. Thus $\mathcal{L}$ is a $\lambda$-system.
Proof (cont)

By the $\pi$-$\lambda$-theorem, $\mathcal{L} \supset \sigma(A_1)$. Thus if $A_1 \in \sigma(A_1)$ and $A_i \in \mathcal{A}_i$, $i = 2, \ldots, n$, then

$$
\mathbb{P}(\cap_{i=1}^n A_i) = \mathbb{P}(A_1)\mathbb{P}(\cap_{i=2}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i).
$$

Hence $\sigma(A_1), A_2, \ldots, A_n$ are independent.

Applying the above to $A_2, \ldots, A_n, \sigma(A_1)$ gives $\sigma(A_2), A_3, \ldots, A_n, \sigma(A_1)$ are independent. After $n$ iterations, we reach the desired conclusion.
Theorem 2.1.8

Suppose $X_1, \ldots, X_n$ are random variables. If for all $x_1, \ldots, x_n \in (-\infty, \infty]$, 

$$
\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x_i),
$$

then $X_1, \ldots, X_n$ are independent.

Proof

Let $\mathcal{A}_i = \{X_i \in (-\infty, x_i] : x_i \in (-\infty, \infty]\}$. Then $\mathcal{A}_i$ is a $\pi$-system. We know that $\sigma(\mathcal{A}_i) = \sigma(X_i)$. Applying Theorem 2.1.7, we get the desired conclusion.
Theorem 2.1.8

Suppose $X_1, \ldots, X_n$ are random variables. If for all $x_1, \ldots, x_n \in (-\infty, \infty]$, 

$$
P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} P(X_i \leq x_i),$$

then $X_1, \ldots, X_n$ are independent.

Proof

Let $\mathcal{A}_i = \{X_i \in (-\infty, x_i] : x_i \in (-\infty, \infty]\}$. Then $\mathcal{A}_i$ is a $\pi$-system. We know that $\sigma(\mathcal{A}_i) = \sigma(X_i)$. Applying Theorem 2.1.7, we get the desired conclusion.
Theorem 2.1.9

Suppose the sub-$\sigma$-fields $\mathcal{F}_{i,j}$, $i = 1, \ldots, n$, $j = 1, \ldots, m(i)$, are independent and let $G_i = \sigma(\bigcup_{j=1}^{m(i)} \mathcal{F}_{i,j})$. Then $G_1, \ldots, G_n$ are independent.

Proof

For $i = 1, \ldots, n$, let $A_i = \{\bigcap_{j=1}^{m(i)} A_{i,j} : A_{i,j} \in \mathcal{F}_{i,j}\}$. Each $A_i$ is a $\pi$-system containing $\bigcup_{j=1}^{m(i)} \mathcal{F}_{i,j}$, and $A_1, \ldots, A_n$ are independent. Applying Theorem 2.1.7, we get that $G_1, \ldots, G_n$ are independent.
Theorem 2.1.9

Suppose the sub-σ-fields \( \mathcal{F}_{i,j}, i = 1, \ldots, n, j = 1, \ldots, m(i), \) are independent and let \( G_i = \sigma(\bigcup_{j=1}^{m(i)} \mathcal{F}_{i,j}) \). Then \( G_1, \ldots, G_n \) are independent.

Proof

For \( i = 1, \ldots, n \), let \( A_i = \{ \bigcap_{j=1}^{m(i)} A_{i,j} : A_{i,j} \in \mathcal{F}_{i,j} \} \). Each \( A_i \) is a \( \pi \)-system containing \( \bigcup_{j=1}^{m(i)} \mathcal{F}_{i,j} \), and \( A_1, \ldots, A_n \) are independent. Applying Theorem 2.1.7, we get that \( G_1, \ldots, G_n \) are independent.
Theorem 2.1.10

If $X_{i,j}$, $i = 1, \ldots, n$, $j = 1, \ldots, m(i)$, are independent random variables, and $f_i : \mathbb{R}^{m(i)} \mapsto \mathbb{R}$, $i = 1, \ldots, n$, are measurable, then $f_i(X_{i,1}, \ldots, X_{i,m(i)})$, $i = 1, \ldots, n$, are independent.

Proof

Let $\mathcal{F}_{i,j} = \sigma(X_{i,j})$ and $\mathcal{G}_i = \sigma(\cup_{j=1}^{m(i)} \mathcal{F}_{i,j})$. The desired result follows from Theorem 2.1.9.
Theorem 2.1.10

If \( X_{i,j}, i = 1, \ldots, n, j = 1, \ldots, m(i), \) are independent random variables, and \( f_i: \mathbb{R}^{m(i)} \to \mathbb{R}, i = 1, \ldots, n, \) are measurable, then \( f_i(X_{i,1}, \ldots, X_{i,m(i)}), i = 1, \ldots, n, \) are independent.

Proof

Let \( \mathcal{F}_{i,j} = \sigma(X_{i,j}) \) and \( \mathcal{G}_i = \sigma(\bigcup_{j=1}^{m(i)} \mathcal{F}_{i,j}), \) The desired result follows from Theorem 2.1.9.
2.1 Independence

**Theorem 2.1.11**

Suppose $X_1, \ldots, X_n$ are independent random variables and $X_i$ has distribution $\mu_i$, then $(X_1, \ldots, X_n)$ has distribution $\mu_1 \times \cdots \times \mu_n$.

**Proof**

Note that, for $A_1, \ldots, A_n \in \mathcal{R}$,

$$
P((X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n) = P(X_1 \in A_1, \ldots, X_n \in A_n)
$$

$$= \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i) = \prod_{i=1}^{n} \mu_i(A_i) = (\mu_1 \times \cdots \times \mu_n)(A_1 \times \cdots \times A_n).$$

Thus the distribution of $(X_1, \ldots, X_n)$ and $\mu_1 \times \cdots \times \mu_n$ agree on $
\{A_1 \times \cdots \times A_n : A_i \in \mathcal{R}\}$, a $\pi$-system that generates $\mathcal{R}^n$. So the $\pi$-$\lambda$ theorem implies they must agree.
Theorem 2.1.11
Suppose $X_1, \ldots, X_n$ are independent random variables and $X_i$ has
distribution $\mu_i$, then $(X_1, \ldots, X_n)$ has distribution $\mu_1 \times \cdots \times \mu_n$.

Proof
Note that, for $A_1, \ldots, A_n \in \mathcal{R}$,

$$
P((X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n) = P(X_1 \in A_1, \ldots, X_n \in A_n)
$$

$$
= \prod_{i=1}^{n} P(X_i \in A_i) = \prod_{i=1}^{n} \mu_i(A_i) = (\mu_1 \times \cdots \times \mu_n)(A_1 \times \cdots \times A_n).
$$

Thus the distribution of $(X_1, \ldots, X_n)$ and $\mu_1 \times \cdots \times \mu_n$ agree on
$
\{A_1 \times \cdots \times A_n : A_i \in \mathcal{R}\}$, a $\pi$-system that generates $\mathcal{R}^n$. So the $\pi$-$\lambda$
theorem implies they must agree.
2.1 Independence

**Theorem 2.1.12**

Suppose $X$ and $Y$ are independent and have distributions $\mu$ and $\nu$ respectively. If $h : \mathbb{R}^2 \mapsto \mathbb{R}$ is a measurable function with $h \geq 0$ or $\mathbb{E}|h(X, Y)| < \infty$, then

$$
\mathbb{E}h(X, Y) = \int \int h(x, y) \mu(dx) \nu(dy).
$$

In particular, if $h(x, y) = f(x)g(y)$ where $f, g : \mathbb{R} \mapsto \mathbb{R}$ are measurable functions with $f, g \geq 0$ or $\mathbb{E}|f(X)| < \infty$ and $\mathbb{E}|g(Y)| < \infty$, then

$$
\mathbb{E}(f(X)g(Y)) = \mathbb{E}f(X) \cdot \mathbb{E}g(Y).
$$

**Proof**

Using Theorem 1.6.9 and then Fubini’s theorem, we have

$$
\mathbb{E}h(X, Y) = \int \int h(x, y)(\mu \times \nu)(dx dy) = \int \int h(x, y)\mu(dx)\nu(dy).
$$
Theorem 2.1.12

Suppose $X$ and $Y$ are independent and have distributions $\mu$ and $\nu$ respectively. If $h : \mathbb{R}^2 \mapsto \mathbb{R}$ is a measurable function with $h \geq 0$ or $\mathbb{E}|h(X, Y)| < \infty$, then

$$\mathbb{E} h(X, Y) = \int \int h(x, y) \mu(dx) \nu(dy).$$

In particular, if $h(x, y) = f(x)g(y)$ where $f, g : \mathbb{R} \mapsto \mathbb{R}$ are measurable functions with $f, g \geq 0$ or $\mathbb{E}|f(X)| < \infty$ and $\mathbb{E}|g(Y)| < \infty$, then

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}f(X) \cdot \mathbb{E}g(Y).$$

Proof

Using Theorem 1.6.9 and then Fubini’s theorem, we have

$$\mathbb{E} h(X, Y) = \int_{\mathbb{R}^2} h(x, y) (\mu \times \nu)(dx dy) = \int \int h(x, y) \mu(dx) \nu(dy).$$
Proof (cont)

To prove the second result, we start with the result when $f, g \geq 0$. In this case, using the first result, the fact that $g(y)$ does not depend on $x$ and then Theorem 1.6.9 twice we get

$$
\mathbb{E}(f(X)g(Y)) = \int \int f(x)g(y)\mu(dx)\nu(dy) = \int g(y) \int f(x)\mu(dx)\nu(dy)
$$

$$
= \int \mathbb{E}f(X)g(y)\nu(dy) = \mathbb{E}f(X) \cdot \mathbb{E}g(Y).
$$

Applying the result for non-negative $f$ and $g$ to $|f|$ and $|g|$, we get

$$
\mathbb{E}|f(X)g(Y)| = \mathbb{E}|f(X)| \cdot \mathbb{E}|g(Y)| < \infty. \text{ We can then use } f = f^+ - f^- \text{ and } g = g^+ - g^- \text{ to get the desired result.}
$$
**Theorem 2.1.13**

Suppose the random variables $X_1, \ldots, X_n$ are independent. If either (a) $X_i \geq 0$ for all $i = 1, \ldots, n$ or (b) $E|X_i| < \infty$ for all $i = 1, \ldots, n$, then

$$E \left( \prod_{i=1}^{n} X_i \right) = \prod_{i=1}^{n} E X_i.$$  

i.e., the expectation on the left exists and has the value given on the right.

**Proof**

$X = X_1$ and $Y = X_2 \cdots X_n$ are independent by Theorem 2.1.10 so taking $f(x) = |x|$ and $g(y) = |y|$, we get

$$E|X_1 X_2 \cdots X_n| = E|X_1| \cdot E|X_2 \cdots X_n|.$$  

By induction, we get that if $1 \leq m \leq n$, 

Theorem 2.1.13

Suppose the random variables $X_1, \ldots, X_n$ are independent. If either (a) $X_i \geq 0$ for all $i = 1, \ldots, n$ or (b) $\mathbb{E}|X_i| < \infty$ for all $i = 1, \ldots, n$, then

$$\mathbb{E} \left( \prod_{i=1}^{n} X_i \right) = \prod_{i=1}^{n} \mathbb{E}X_i.$$

i.e., the expectation on the left exists and has the value given on the right.

Proof

$X = X_1$ and $Y = X_2 \cdots X_n$ are independent by Theorem 2.1.10 so taking $f(x) = |x|$ and $g(y) = |y|$, we get

$$\mathbb{E}|X_1 X_2 \cdots X_n| = \mathbb{E}|X_1| \cdot \mathbb{E}|X_2 \cdots X_n|.$$

By induction, we get that if $1 \leq m \leq n$,
Proof (cont)

\[ \mathbb{E}|X_m \cdots X_n| = \prod_{i=m}^{n} \mathbb{E}|X_i|. \]

If \( X_i \geq 0 \) for all \( i = 1, \ldots, n \), the desired result follows from the special case \( m = 1 \). To prove the result in general note that the special case \( m = 2 \) implies \( \mathbb{E}|Y| = \mathbb{E}|X_2 \cdots X_n| < \infty \), so using Theorem 2.1.12 with \( f(x) = x \) and \( g(y) = y \), we get \( \mathbb{E}(X_1 \cdots X_n) = \mathbb{E}X_1 \cdot \mathbb{E}(X_2 \cdots X_n) \) and the desired result follows by induction.
Theorem 2.1.15

If $X$ and $Y$ are independent, $F(x) = \mathbb{P}(X \leq x)$ and $G(y) = \mathbb{P}(Y \leq y)$, then

$$\mathbb{P}(X + Y \leq z) = \int F(z - y) dG(y).$$

Proof

Let $h(x, y) = 1_{\{x+y\leq z\}}$. Let $\mu$ and $\nu$ be the distributions of $X$ and $Y$ respectively. For fixed $y$,

$$\int h(x, y) \mu(dx) = \int 1_{(-\infty, z-y]}(x) \mu(dx) = F(z - y),$$

thus

$$\mathbb{P}(X + Y \leq z) = \int \int h(x, y) \mu(dx) \nu(dy) = \int F(z - y) \nu(dy) = \int F(z - y) dG(y).$$
### Theorem 2.1.15

If $X$ and $Y$ are independent, $F(x) = \mathbb{P}(X \leq x)$ and $G(y) = \mathbb{P}(Y \leq y)$, then

$$
\mathbb{P}(X + Y \leq z) = \int F(z - y) dG(y).
$$

### Proof

Let $h(x, y) = 1_{\{x+y\leq z\}}$. Let $\mu$ and $\nu$ be the distributions of $X$ and $Y$ respectively. For fixed $y$,

$$
\int h(x, y) \mu(dx) = \int 1_{(-\infty, z-y]}(x) \mu(dx) = F(z - y),
$$

thus

$$
\mathbb{P}(X + Y \leq z) = \int \int h(x, y) \mu(dx) \nu(dy) = \int F(z - y) \nu(dy) = \int F(z - y) dG(y).
$$
Theorem 2.1.16

Suppose that $X$ with density $f$ and $Y$ with distribution function $G$ are independent. Then $X + Y$ has density

$$h(x) = \int f(x - y)dG(y).$$

When $Y$ has density $g$, the last formula can be written as

$$h(x) = \int f(x - y)g(y)dy.$$

Proof

By the previous theorem,

$$\mathbb{P}(X + Y \leq z) = \int F(z - y)dG(y) = \int \int_{-\infty}^{z} f(x - y)dx dG(y)$$

$$= \int_{-\infty}^{z} \int f(x - y)dG(y)dx.$$
Theorem 2.1.16
Suppose that $X$ with density $f$ and $Y$ with distribution function $G$ are independent. Then $X + Y$ has density

$$h(x) = \int f(x - y) dG(y).$$

When $Y$ has density $g$, the last formula can be written as

$$h(x) = \int f(x - y) g(y) dy.$$

Proof
By the previous theorem,

$$P(X + Y \leq z) = \int F(z - y) dG(y) = \int \int_{-\infty}^{z} f(x - y) dx dG(y)$$

$$= \int_{-\infty}^{z} \int f(x - y) dG(y) dx.$$
Theorem 2.1.18
Suppose $X$ and $Y$ are independent, $X$ is a gamma random variable with parameters $(\alpha_1, \lambda)$ and $Y$ is a gamma random variable with parameters $(\alpha_2, \lambda)$. Then $X + Y$ is a gamma random variable with parameters $(\alpha_1 + \alpha_2, \lambda)$.

Theorem 2.1.20
Suppose $X$ and $Y$ are independent, $X$ is a normal random variable with parameters $(\mu_1, \sigma_1^2)$ and $Y$ is a normal random variable with parameters $(\mu_2, \sigma_2^2)$. Then $X + Y$ is a normal random variable with parameters $(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Similar results for binomial, Poisson and negative binomial random variables.
Theorem 2.1.18
Suppose $X$ and $Y$ are independent, $X$ is a gamma random variable with parameters $(\alpha_1, \lambda)$ and $Y$ is a gamma random variable with parameters $(\alpha_2, \lambda)$. Then $X + Y$ is a gamma random variable with parameters $(\alpha_1 + \alpha_2, \lambda)$.

Theorem 2.1.20
Suppose $X$ and $Y$ are independent, $X$ is a normal random variable with parameters $(\mu_1, \sigma_1^2)$ and $Y$ is a normal random variable with parameters $(\mu_2, \sigma_2^2)$. Then $X + Y$ is a normal random variable with parameters $(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Similar results for binomial, Poisson and negative binomial random variables.
Theorem 2.1.18
Suppose $X$ and $Y$ are independent, $X$ is a gamma random variable with parameters $(\alpha_1, \lambda)$ and $Y$ is a gamma random variable with parameters $(\alpha_2, \lambda)$. Then $X + Y$ is a gamma random variable with parameters $(\alpha_1 + \alpha_2, \lambda)$.

Theorem 2.1.20
Suppose $X$ and $Y$ are independent, $X$ is a normal random variable with parameters $(\mu_1, \sigma_1^2)$ and $Y$ is a normal random variable with parameters $(\mu_2, \sigma_2^2)$. Then $X + Y$ is a normal random variable with parameters $(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Similar results for binomial, Poisson and negative binomial random variables.