1 Simple Random Variables

Poisson distribution. Let $\lambda \in [0, \infty)$ and let $X : \Omega \to \mathbb{N}_0$ be such that

$$P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad n \in \mathbb{N}_0$$

Then $X$ is called the Poisson random variable with parameter $\lambda$, or $X \sim \text{Poi}(\lambda)$. $E[X] = \text{Var}[X] = \lambda$.

Poisson distribution is usually used to model the probability of a given number of events occurring in a fixed interval of time, for example, the number of buses passing certain bus stop per hour.

Exponential distribution. Let $\theta > 0$ and let $X$ be a nonnegative random variable such that

$$P(X \leq x) = \int_0^x \theta e^{-\theta t} dt \quad \text{for} \quad x > 0$$

Then $X$ is an exponential random variable with parameter $\theta$, or $X \sim \text{Exp}(\theta)$. $E[X] = 1/\theta, \text{Var}[X] = 1/\theta^2$.

Exponential distribution is the distribution that describes the time between events in a Poisson process. In the bus example, the waiting time between two buses is exponential distributed.

Gamma distribution. Let $\theta, r > 0$ and let $\Gamma_{\theta, r}$ be the distribution on $[0, \infty)$ with density

$$\theta^r \Gamma(r) x^{r-1} e^{-\theta x}.$$ 

Then $\Gamma_{\theta, r}$ is called the Gamma distribution with scaled parameter $\theta$ and shape parameter $r$. $E[\Gamma_{\theta, r}] = r/\theta, \text{Var}[\Gamma_{\theta, r}] = r/\theta^2$. In particular, if $\{X_i\}_{i=1}^n$ are i.i.d random variables of $\text{Exp}(\theta)$, then $\sum_{i=1}^n X_i \sim \Gamma_{\theta, n}$.

2 Borel-Cantelli Lemma

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Consider a sequence of subsets $\{A_n\}$ of $\Omega$. We define

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega \in \Omega : \#\{n \in \mathbb{N} : \omega \in A_n\} = \infty\}$$

and

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \{\omega \in \Omega : \#\{n \in \mathbb{N} : \omega \notin A_n\} < \infty\}$$

In other words, the limit superior is the event where infinitely many of the $A_n$ occur. On the other hand, the limit inferior is the event where eventually all of the $A_n$ occur. It is clear that

$$\limsup 1_{A_n} = 1_{\limsup A_n}, \quad \liminf 1_{A_n} = 1_{\liminf A_n}$$

where $1_A(x)$ is the indicator function of set $A$. It is more convenient to write $\limsup A_n = \{\omega : \omega \in A_n \text{ i.o.}\}$, where i.o. means infinitely often.

Theorem 2.1. Borel-Cantelli lemma 1 (BC1) If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n \text{ i.o.}) = 0$.

Proof. By definition and the summability of $P(A_n)$,

$$P(A_n \text{ i.o.}) = P(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m) = \lim_{n \to \infty} P(\bigcup_{m=n}^{\infty} A_m) \leq \lim_{n \to \infty} \sum_{m=n}^{\infty} P(A_m) = 0$$
(BC1) provides a sufficient condition to verify $\mathbb{P}(A_n \text{ i.o.}) = 0$, which is very useful to prove some almost sure results in probability. Recall that a family of random variables $\{X_n\}$ converge to $X$ in probability means, for any $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0 \text{ as } n \to \infty.$$ 

$X_n$ converges to $X$ almost surely (a.s.) is to say

$$\mathbb{P}(\omega : |X_n(\omega) - X(\omega)| \to 0) = 1.$$ 

**Proposition 2.1.** $X_n \to X$ a.s. if and only if for any $\varepsilon > 0, \mathbb{P}(|X_n - X| > \varepsilon \text{ i.o.}) = 0$.

**Proof.** By definition,

$$X_n \to X \text{ a.s. } \iff \mathbb{P}(\omega : |X_n(\omega) - X(\omega)| \to 0) = 1$$ 

$$\iff \mathbb{P}(\forall \varepsilon > 0, \exists N, \text{s.t.}|X_n - X| \leq \varepsilon \text{ for } n > N) = 1$$ 

$$\iff \mathbb{P}(\cap_{\varepsilon_k} \cup_{n=1}^\infty \cap_{m=n}^\infty \{ |X_m - X| \leq \varepsilon_k \}) = 1$$ 

$$\iff \forall \varepsilon > 0, \mathbb{P}(\cap_{n=1}^\infty \cup_{m=n}^\infty \{ |X_m - X| > \varepsilon \}) = 0$$ 

$$\iff \forall \varepsilon > 0, \mathbb{P}(|X_n - X| > \varepsilon \text{ i.o.}) = 0$$

\[\square\]

The following result gives an application of (BC1).

**Proposition 2.2.** $X_n \to X$ in probability if and only if for every subsequence $X_{n(m)}$ such that there is a further subsequence $X_{n(m_k)}$ that converges almost surely to $X$. 

**Proof.** Let $\varepsilon_k$ be a sequence of positive numbers converging to zero. For each $k$, there is an $n(m_k) > n(m_{k-1})$ so that $\mathbb{P}(|X_{n(m_k)} - X| > \varepsilon_k) \leq 2^{-k}$. Noticing that

$$\sum_{k=1}^\infty \mathbb{P}(|X_{n(m_k)} - X| > \varepsilon_k) < \infty,$$

(BC1) implies $\mathbb{P}(|X_{n(m_k)} - X| > \varepsilon_k \text{ i.o.}) = 0$, i.e., $X_{n(m_k)} \to X$ a.s. On the other hand, if for every subsequence $X_{n(m)}$ there is a further subsequence $X_{n(m_k)} \to X$ a.s., then for any $\delta > 0$, any subsequence of the sequence $\{y_n\}$ with $y_n := \mathbb{P}(|X_n - X| > \delta)$ has a convergent further subsequence (with limit zero), thus $y_n \to 0$. 

\[\square\]

**Theorem 2.2.** Borel-Cantelli lemma 2 (BC2). Assume that $A_n$ are independent, then $\sum_{n=1}^\infty \mathbb{P}(A_n) = \infty$ implies $\mathbb{P}(A_n, \text{ i.o.}) = 1$.

**Proof.** We only need to show $\mathbb{P}(\{A_n, \text{ i.o.}\}^c) = 0$. Indeed, by the independence of $A_n$

$$\mathbb{P}(\{A_n, \text{ i.o.}\}^c) = \mathbb{P}(\cup_{n=1}^\infty \cap_{m=n}^\infty A_m^c) = \lim_{n \to \infty} \mathbb{P}(\cap_{m=n}^\infty A_m^c) = \lim_{n \to \infty} \prod_{m=n}^\infty (1 - \mathbb{P}(A_m))$$

$$\leq \lim_{n \to \infty} \prod_{m=n}^\infty \exp(-\mathbb{P}(A_m)) = \lim_{n \to \infty} \exp\left(-\sum_{m=n}^\infty \mathbb{P}(A_m)\right) = 0$$

where we have used the fact that $1 - x \leq e^{-x}$ when $x > 0$ for the inequality. 

\[\square\]

**Theorem 2.3.** Strong law of large numbers (SLLN). Let $\{X_i\}$ be an i.i.d sequence with $\mathbb{E}|X_i| < \infty$. Assume that $\mathbb{E}|X_i| = \mu$ and $S_n = \sum_{i=1}^n X_i$. Then $S_n/n \to \mu$ a.s. as $n \to \infty$. 

2
As an application of (BC2), the following result shows that $\mathbb{E}|X_t| < \infty$ is necessary for SLLN.

**Proposition 2.3.** If $\{X_n\}$ is a family of i.i.d random variables with $\mathbb{E}|X_1| = \infty$, then $\mathbb{P}(|X_n| \geq n, \text{ i.o.}) = 1$. So if $S_n = \sum_{i=1}^{n} X_i$ then $\mathbb{P}(|\lim_{n \to \infty} S_n/n \text{ exits } \in (-\infty, \infty)) = 0$.

**Proof.** Note that

$$\mathbb{E}|X_1| = \int_0^{\infty} \mathbb{P}(|X_1| > x)dx \leq \sum_{n=0}^{\infty} \mathbb{P}(|X_1| > n).$$

If $\mathbb{E}|X_1| = \infty$ and are i.i.d, by (BC2) $\mathbb{P}(|X_n| \geq n, \text{ i.o.}) = 1$. Moreover, observe that

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}$$

and on $C := \{\omega : \lim_{n \to \infty} S_n/n \text{ exists } \in (-\infty, \infty)\}, S_n/(n(n+1)) \to 0$. Thus on $C \cap \{\omega : |X_n| \geq n \text{ i.o. }\}$, we have

$$\left| \frac{S_n}{n} - \frac{S_{n+1}}{n+1} \right| > \frac{2}{3} \text{ i.o.}$$

contradicting the fact that $\omega \in C$. Thus

$$\{\omega : |X_n| \geq n \text{ i.o.}\} \cap C = \emptyset$$

and since $\mathbb{P}(|X_n| \geq n, \text{ i.o.}) = 1$, we have $\mathbb{P}(C) = 0$. \qed

## 3 Kolmogorov’s Extension Theorem

In this section, we discuss the problem of constructing stochastic processes which are infinite collections $\{X_t : t \in T\}$ of random variables. In general, this construction starts from the finite distribution of the process. Then Kolmogorov’s Extension Theorem guarantees the existence of a consistent measure on some infinite dimensional space.

Let $(E, \mathcal{E})$ be a measurable space and let $T$ be an index set. We denote by $E^T$ the set of all functions $f$ from $T$ to $E$. For every $t \in T$, define $\pi_t : E^T \to E$ to be the evaluation (projection) map

$$\pi_t(f) := f(t)$$

Then one can define a $\sigma$-algebra $\mathcal{E}^T$ on $E^T$ which is the smallest $\sigma$-algebra such that each $\pi_t$ is $(\mathcal{E}^T / \mathcal{E})$-measurable.

**Definition 3.1. (Stochastic process)** Let $T$ be an index set, $(E, \mathcal{E})$ a measurable space and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. A stochastic process is defined as a collection $\{X_t : t \in T\}$ of $(E, \mathcal{E})$-valued random variables carried by the triple $(\Omega, \mathcal{F}, \mathbb{P})$.

Let $X$ be a stochastic process, then for $\omega \in \Omega$, the map $X(\omega) : t \in T \mapsto X_t(\omega) \in E$ is called the sample function of $X$. Here is another view of $X$, that is the map $X : \omega \in \Omega \mapsto X(\omega) \in E^T$ or an $(E^T, \mathcal{E}^T)$-valued random variable.

**Definition 3.2. (Law of the stochastic process)** The law of the stochastic process is defined as the probability measure $\mu : \mathbb{P} \circ X^{-1}$ on $(E^T, \mathcal{E}^T)$.

For any non-empty subset $S$ of $T$, we define $\pi_S X : \Omega \mapsto E^S$ via

$$(\pi_S X)(\omega) := \pi_S(X(\omega)) = X(\omega)|_{S}$$

and

$$\mu_S := \mathbb{P} \circ (\pi_S X)^{-1}$$

on $(E^S, \mathcal{E}^S)$. 

---

3
Definition 3.3. (Finite dimensional distribution) Let \( \text{Fin}(T) \) be the set of non-empty finite set of \( T \). The probability measures \( \{\mu_S : S \in \text{Fin}(T)\} \) are called the finite-dimensional distribution of \( X \).

Definition 3.4. (Compatibility condition) Let \( U \subset V \in \text{Fin}(T) \) and let \( \pi_U^V \) denote the restriction map from \( E^V \) to \( E^U \), then we say that the finite-dimensional distribution \( \{\mu_S\} \) has the Compatibility condition if

\[
\mu_U = \mu_V \circ (\pi_U^V)^{-1}
\]

Theorem 3.1. (Kolmogorov’s Extension Theorem) Let \( E \) be a compact metrisable space, and let \( \mathcal{E} = \mathcal{B}(E) \). Let \( T \) be a set. Suppose that for each \( S \) in \( \text{Fin}(T) \), there exits a probability measure \( \mu_S \) on \( (E^S, \mathcal{E}^S) \) and that the measures \( \{\mu_S : S \in \text{Fin}(T)\} \) are compatible, i.e. \( \mu_U = \mu_V \circ (\pi_U^V)^{-1} \) holds whenever \( U, V \in \text{Fin}(T) \) for \( U \subset V \). Then there exists a unique measure \( \mu \) on \( (E^T, \mathcal{E}^T) \) such that

\[
\mu_S = \mu \circ \pi_S^{-1} \quad \text{on} \quad (E^S, \mathcal{E}^S)
\]

Remark 3.1. The compact metrisable condition on \( (E, \mathcal{E}) \) in Theorem 3.1 can be weaken to be a topological space endowed with the Borel \( \sigma \)-algebra.

Example (Gaussian processes). Let \( C \) be a symmetric positive quadratic form on \( \mathbb{R}^N \). Then there exists a unique Gaussian processes with index set \( \mathbb{N} \), state space \( \mathbb{R} \), such that for all finite \( J \in \mathbb{N} \), the finite dimensional distributions are \( |J| \)-dimensional Gaussian vectors with covariance \( C^J \). Here \( C^J \) is the finite-dimensional sub-matrices of \( C \).

References
