STA347: Probability

Tianyu Du

October 26, 2019

Contents

1 Preliminaries

Definition 1.1. A **process**[1] W is a mechanism generating **outcomes** w from a sample space Ω. Any realized trail of process W can be denoted as a potentially infinite sequence in Ω:

$$W : w_1, w_2, \cdots, w_n, \cdots \tag{1.1}$$

Definition 1.2. A **random variable** (extended process), $X := g(W)$, can be constructed from a process W and a real-valued function $g : \Omega \to \mathbb{R}$.

Definition 1.3. Given a random variable $X = g(W)$, the **sample mean** (i.e. empirical expectation) of the first n trials from a sequence of realizations, $g(w_1), \cdots, g(w_n), \cdots$, is defined to be

$$\hat{\mathbb{E}}_n g(W) := \frac{\sum_{i=1}^n g(w_i)}{n} \tag{1.2}$$

Definition 1.4. A process W is said to be a **random process** if it satisfies the *empirical law of large numbers*, in that, $\forall g \in \mathbb{R}^\Omega$:

(i) *stability*: $(\hat{\mathbb{E}}_n g(W))_{n\in\mathbb{N}}$ converges;

(ii) *Invariance*: $\forall\ (w_n)_{n\in\mathbb{N}} \subseteq \Omega$, the limits of $(\hat{\mathbb{E}}_n g(W))_{n\in\mathbb{N}}$ are the same.

Definition 1.5. Let W be a random process and $g \in \mathbb{R}^\Omega$, the **expected value** of $g(W)$ is defined as

$$\mathbb{E}g(W) := \lim_{n\to\infty} \hat{\mathbb{E}}_n g(W) \tag{1.3}$$

the limit is well-defined given ELLN.

Definition 1.6. Let W be a random process. For every $A \subseteq \Omega$, take $g := I_A \in \mathbb{R}^\Omega$, the **empirical relative frequencies** (i.e. empirical probability) is defined as

$$\hat{P}(W \in A) := \hat{\mathbb{E}}_n I_A(W) \tag{1.4}$$

Given ELLN, the limit is well-defined, then the **probability** is defined to be the limit:

$$P(W \in A) := \lim_{n\to\infty} \hat{P}(W \in A) \tag{1.5}$$

Remark 1.1. The notation of expected values and probabilities on W is well-defined only when W satisfies the empirical law of large numbers, that is, W is a random process.
Given W defined on Ω satisfies ELLN, the behaviour of W can be fully characterized by its **probability distribution**.

$$W \sim P_W \text{ on } \Omega \tag{1.6}$$

2 Distributions

Definition 2.1. A **standard uniform** is defined to be $\mathcal{U} \sim unif[0,1]$ if and only if

$$P(\mathcal{U} \leq u) = u\ \forall u \in [0,1] \tag{2.1}$$

[1] This is just a process, not necessarily a random process.

Definition 2.2. $Z \sim unif\{0, \cdots, p-1\}$ if and only if

$$P(Z = i) = P(Z = j) \quad \forall i, j \in \{0, \cdots, p-1\} \tag{2.2}$$

Theorem 2.1. If $U = \sum_{n=1}^{\infty} Z_i p^{-i}$, then the following are equivalent:

(i) $U \sim unif[0, 1]$;

(ii) $Z_i \overset{i.i.d.}{\sim} Z \overset{d}{=} unif\{0, \cdots, p-1\}$.

Definition 2.3. Two random processes X, Y on a common sample space $\mathcal{X}$ are **identically distributed**, $X \overset{d}{=} Y$ if and only if

$$\mathbb{E}[g(X)] = \mathbb{E}[g(Y)] \quad \forall g : \mathcal{X} \to \mathbb{R} \tag{2.3}$$

Proposition 2.1. Specifically, for $A \overset{d}{=} B$, take $g = I_A$ where $A \subset \mathcal{X}$. It is evident that for every such subset, the probability **probability** as

$$\mathbb{P}[X \in A] = \mathbb{E}[I_A(X)] = \mathbb{E}[I_A(Y)] = \mathbb{P}[Y \in A] \tag{2.4}$$

Theorem 2.2 (Invariance)**.** If $X \overset{d}{=} Y$, then

$$\varphi(X) \overset{d}{=} \varphi(Y) \quad \forall \varphi : \mathcal{X} \to \mathcal{Y} \tag{2.5}$$

Proof.

$$\mathbb{E}[h \circ \varphi(X)] = \mathbb{E}[h \circ \varphi(Y)] \quad \forall h : \mathcal{Y} \to \mathbb{R} \tag{2.6}$$

■

Definition 2.4. The **expectation** operator

$$\mathbb{E} : \mathcal{R} \to \mathbb{R} \cup \{\pm\infty\} \cup \{\text{DNE}\} \tag{2.7}$$

where $\mathcal{R}$ is the space of *real-valued* random processes.

Proposition 2.2. Let $W \sim unif\{1, \cdots, n\}$, then

$$n + 1 - W \overset{d}{=} W \tag{2.8}$$
$$\implies (n + 1 - W)^2 \overset{d}{=} W^2 \tag{2.9}$$
$$\implies (n + 1)^2 - 2(n + 1)W + W^2 \overset{d}{=} W^2 \tag{2.10}$$
$$\implies \mathbb{E}[(n + 1)^2 - 2(n + 1)W + W^2] = \mathbb{E}[W^2] \tag{2.11}$$
$$\implies \mathbb{E}[W] = \frac{n + 1}{2} \tag{2.12}$$

Proposition 2.3.

$$(n+1-W)^3 \stackrel{d}{=} W^3 \tag{2.13}$$
$$\implies 2\mathbb{E}[W^3] = (n+1)^3 - 3(n+1)^2\mathbb{E}[W] + 3(n+1)\mathbb{E}[W^2] \tag{2.14}$$
$$\implies 2\mathbb{E}[W^3] = (n+1)^3 - 3(n+1)^2\frac{n+1}{2} + 3(n+1)\mathbb{E}[W^2] \tag{2.15}$$
$$\implies 2\mathbb{E}[W^3] = -\frac{(n+1)^2}{2} + 3(n+1)\mathbb{E}[W^2] \tag{2.16}$$
$$\implies \mathbb{E}[W^3] = n(\mathbb{E}[W])^2 \tag{2.17}$$

Proposition 2.4. $\mathbb{E}[W^4]$. TODO

Definition 2.5. $W \sim unif\{1, \cdots, n\}$, then the *distance between* W^2 and $\mathbb{E}[W^2]$ is defined as

$$d(W^2, \mathbb{E}[W^2]) := \sqrt{\mathbb{E}[W^2 - \mathbb{E}[W])^2]^2} = \sqrt{\mathbb{V}[W^2]} = \sigma_{W^2} \tag{2.18}$$

Corollary 2.1 (Corollary of Jensen's Inequality)**.**

$$\mathbb{E}[W^2] \geq (\mathbb{E}[W])^2 \tag{2.19}$$

and equality holds if and only if

$$\mathbb{E}[(W - \mathbb{E}[W])^2] = 0 \tag{2.20}$$

which is equivalent to

$$P(W = \mathbb{E}[W]) = 1 \tag{2.21}$$

Proof.

$$\mathbb{V}[W] = \mathbb{E}[(W - \mathbb{E}[W])^2] \geq 0 \tag{2.22}$$

■

Lemma 2.1. $u = \sum_{i=1}^{\infty} z_i p^{-i}$, and let $z = (z_i : i \in \mathbb{N}) \in \dot{p}^{\infty}$, then

$$z_1 = b_1 \tag{2.23}$$