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/variable** 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**.

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

---

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

2 Distributions

2.1 Construction of Uniform Distributions

Definition 2.1. For any $n \in \mathbb{N}$, a random variable X is said to have a **(finite discrete) uniform distribution** on the sample space $\Omega = \{1, \cdots, n\}$ if

$$P(X = k) = \frac{1}{n} \ \forall k \in \{1, \cdots, n\} \tag{2.1}$$

Denoted as $X \sim unif\{1, \cdots, n\}$.

Definition 2.2. A random variable U wit $\Omega = [0, 1]$ is said to follow a **(continuous) standard uniform** if

$$P(U \leq u) = u \ \forall u \in [0, 1] \tag{2.2}$$

Denoted as $U \sim unif[0, 1]$.

Construction of Continuous Uniform Let Y be a random variable on $\{0, \cdots, 9\}$. For each sequence of realizations of Y, $(Y_i)_{i \in \mathbb{N}}$, one can construct $U \in [0, 1]$ using the following decimal expansion:

$$U := \sum_{i=1}^{\infty} \frac{Y_i}{10^i} \tag{2.3}$$

Proposition 2.1. It is evident that each finite decimal of Y, $(Y_1, \cdots, Y_n)$, partitions the unitary into 10^n small intervals

$$P(0.Y_1 \cdots Y_n \leq U \leq 0.Y_1 \cdots Y_n + 10^{-n}) = \frac{1}{10^n} \tag{2.4}$$

and

$$P(0 \leq U \leq 0.Y_1 \cdots Y_n) = \frac{Y_1 Y_2 \cdots Y_n}{10^n} = 0.Y_1 \cdots Y_n \tag{2.5}$$

Proposition 2.2. $P(a \leq U \leq b) = b - a$.

Corollary 2.1. $P(U = u) = P(u \leq U \leq u) = u - u = 0$.

Proposition 2.3. Let $U \sim unif[0, 1]$ and $V := 1 - U$, then $V \stackrel{d}{=} U$.

Proof.

$$P(V \leq u) = P(1 - U \leq u) = P(U \geq 1 - u) \tag{2.6}$$
$$= 1 - P(U \leq 1 - u) = 1 - 1 + u = u \tag{2.7}$$

$\blacksquare$

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

Proof. The result should be evident by construction of uniform distribution, (ii) is simply the p-bit expansion of real numbers, which is equivalent to the decimal expansion. ∎

Definition 2.3. Two random processes W_1, W_2 defined on Ω are **identically distributed**, $W_1 \stackrel{d}{=} W_1$ if

$$\mathbb{E}g(W_1) = \mathbb{E}g(W_2) \quad \forall g : \Omega \to \mathbb{R} \tag{2.8}$$

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

$$\varphi(X) \stackrel{d}{=} \varphi(Y) \quad \forall \varphi : \Omega \to \mathcal{X} \tag{2.9}$$

Note that $\varphi(X)$ and $\varepsilon(Y)$ are random variable defined on sample space $\mathcal{X}$.

Proof. Let $h : \mathcal{X} \to \mathbb{R}$, note that $h \circ \varepsilon : \Omega \to \mathbb{R}$. It is evident that $\mathbb{E}h \circ \varphi(X) = \mathbb{E}h \circ \varphi(Y)$. ∎

Theorem 2.3 (Invariance II).

$$W_1 \stackrel{d}{=} W_2 \iff g(W_1) \stackrel{d}{=} g(W_2) \quad \forall g : \Omega \to \mathbb{R} \tag{2.10}$$

Proof. ($\implies$) The sufficient direction is direct by previous theorem.
($\impliedby$) Suppose $g(W_1) \stackrel{d}{=} g(W_2) \quad \forall g : \Omega \to \mathbb{R}$. The proof is immediate by applying identity mapping in the definition of $g(W_1) \stackrel{d}{=} g(W_2)$, which implies $\mathbb{E}g(W_1) = \mathbb{E}g(W_2)$ for any function $g : \Omega \to \mathbb{R}$. ∎

Corollary 2.2. Let $A \stackrel{d}{=} B$, for each $A \subseteq \Omega$, take $g = I_A$. Then,

$$P(X \in A) = \mathbb{E}I_A(X) = \mathbb{E}I_A(Y) = P(Y \in A) \tag{2.11}$$

2.2   Constructing Other Distributions

Definition 2.4. Let $Z = -\ln U$ with $u \sim unif[0,1]$. Then Z is said to follow **exponential distribution** such that

$$P(s \le Z \le t) = e^{-s} - e^{-t} \tag{2.12}$$

The derivation of distribution is immediate from the monotone property of $-\ln$.

Definition 2.5. The random variable Z is said to have a **standard exponential distribution** on $\mathbb{R}_+$, denoted as $Z \sim exp(1)$ if

$$P(Z \le z) = 1 - \exp(-z) \tag{2.13}$$

Definition 2.6. The random variable X is said to have a **scaled exponential distribution** with some $\theta > 0$, denoted as $X \sim exp(\theta)$ if

$$X \stackrel{d}{=} \theta Z \quad Z \sim exp(1) \tag{2.14}$$

Definition 2.7. For any real-valued random variable, X^2, the **distribution function** of X is defined as

$$F_X(x) := P(X \le x) \ \forall x \in \mathbb{R} \tag{2.15}$$

---

[2]A real-valued random variable can be deemed as the composite of an arbitrary random variable W on Ω and a function $g : \Omega \to \mathbb{R}$

Definition 2.8. A real-valued random variable, X, is said to be **absolutely continuous** with respective to length measure if there exists $f : \mathbb{R} \to \mathbb{R}_+$ such that

$$P(s < X \leq t) = \int_s^t f_X(x) \; dx \quad \forall s \leq t \tag{2.16}$$

Where f_X is defined as the **probability density function** of X.
Remark 1: A random variable is absolutely continuous if there exists an integrable density function f.
Remark 2: The density function f_X is not necessarily unique.

Definition 2.9. The **percentile/quantile function**, $x_p = g(p) : [0, 1] \to \mathbb{R}$, is defined so that

$$F(X \leq x_p) = p \tag{2.17}$$

Remark 2.1. In any event, it is (or certainly should be) perfectly clear that each of the three methods outlined above will lead to the same basic result, each with its own probability density function, f, its own distribution function, F, and its own percentile function, g.

2.3 More on Expectation Operators

Remark 2.2. The **expectation** operator

$$\mathbb{E} : \mathcal{R} \to \mathbb{R} \cup \{\pm\infty\} \cup \{\text{DNE}\} \tag{2.18}$$

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

Theorem 2.4. By the algebraic limit theorem, it is evident that for any pair of real-valued random process X, Y and scalar a, the expectation operator is linear:

(i) $\mathbb{E}(X + Y) + \mathbb{E}X + \mathbb{E}Y$;

(ii) $\mathbb{E}(aX) = a\mathbb{E}X$.

Proposition 2.4 (Expectation is Normed). $\mathbb{E}c = c \; \forall c \in \mathbb{R}$.

Proposition 2.5. $X \geq 0 \implies \mathbb{E}X \geq 0$ by order limit theorem.

Theorem 2.5. Let W be a real-valued random variable. Then for any finite mutually disjoint set $(A_1, \cdots, A_n)$,

$$P\left(W \in \sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(W \in A_i) \tag{2.19}$$

That is, the probability measure is linear given mutual disjoining.

Proposition 2.6. Properties of probability measure:

(i) $I_\Omega(W) = 1 \implies P(W \in \Omega) = \mathbb{E}1 = 1$;

(ii) $0 \leq I_A(W) \leq 1 \implies 0 \leq P(W \in A) = \mathbb{E}I_A(W) \leq 1$.

2.4 Expected Value for an Arbitrary Discrete Distribution

Definition 2.10. A **finite scheme** or a **finite discrete distribution** can be written as

$$W \sim \begin{pmatrix} \omega_1 & \cdots & \omega_N \\ p_1 & \cdots & p_N \end{pmatrix} \tag{2.20}$$

with **probability mass function** (pmf)

$$P\left(W = \omega_i\right) = p_i \ s.t. \ \sum_{i=1}^{N} p_i = 1 \tag{2.21}$$

With vector notation

$$W \sim \begin{pmatrix} \omega \\ p \end{pmatrix} \ s.t. \ \omega \subseteq \Omega, \ \langle p, 1 \rangle = 1 \tag{2.22}$$

Proposition 2.7. The expected value of a finite discrete random variable is more or less obvious:

$$Eg(W) = \sum_{i=1}^{N} g\left(\omega_i\right) P\left(W = \omega_i\right) = \sum_{i=1}^{N} g\left(\omega_i\right) p_i \tag{2.23}$$

Proposition 2.8. A finite discrete random variable, $g(W)$, can be explicitly represented as a *finite linear combination of simple indicator functions*:

$$g(W) = \sum_{i=1}^{N} g\left(\omega_i\right) I\left(W = \omega_i\right) = \sum_{i=1}^{N} g\left(\omega_i\right) I_{\{\omega_i\}}(W) \tag{2.24}$$

2.5   Discrete Uniform Distributions

Proposition 2.9. Let $W \sim unif\{1, \cdots, n\}$, then

$$n + 1 - W \overset{d}{=} W \tag{2.25}$$

$$\implies (n + 1 - W)^2 \overset{d}{=} W^2 \tag{2.26}$$

$$\implies (n + 1)^2 - 2(n + 1)W + W^2 \overset{d}{=} W^2 \tag{2.27}$$

$$\implies \mathbb{E}[(n + 1)^2 - 2(n + 1)W + W^2] = \mathbb{E}[W^2] \tag{2.28}$$

$$\implies \mathbb{E}[W] = \frac{n + 1}{2} \tag{2.29}$$

Proposition 2.10.

$$(n + 1 - W)^3 \overset{d}{=} W^3 \tag{2.30}$$

$$\implies 2\mathbb{E}[W^3] = (n + 1)^3 - 3(n + 1)^2\mathbb{E}[W] + 3(n + 1)\mathbb{E}[W^2] \tag{2.31}$$

$$\implies 2\mathbb{E}[W^3] = (n + 1)^3 - 3(n + 1)^2\frac{n + 1}{2} + 3(n + 1)\mathbb{E}[W^2] \tag{2.32}$$

$$\implies 2\mathbb{E}[W^3] = -\frac{(n + 1)^2}{2} + 3(n + 1)\mathbb{E}[W^2] \tag{2.33}$$

$$\implies \mathbb{E}[W^3] = n(\mathbb{E}[W])^2 \tag{2.34}$$

Proposition 2.11. $\mathbb{E}[W^4]$. <mark>TODO</mark>

Definition 2.11. $W \sim unif\{1, \cdots, n\}$, then the *distance between W^2 and $\mathbb{E}[W^2]$* is defined as

$$d(W^2, \mathbb{E}[W^2]) := \sqrt{\mathbb{E}[W^2 - \mathbb{E}[W])^2]^2} = \sqrt{\mathbb{V}[W^2]} = \sigma_{W^2} \tag{2.35}$$

Corollary 2.3 (Corollary of Jensen's Inequality)**.**

$$\mathbb{E}[W^2] \geq (\mathbb{E}[W])^2 \tag{2.36}$$

and equality holds <u>if and only if</u>

$$\mathbb{E}[(W - \mathbb{E}[W])^2] = 0 \tag{2.37}$$

which is equivalent to

$$P(W = \mathbb{E}[W]) = 1 \tag{2.38}$$

Proof.

$$\mathbb{V}[W] = \mathbb{E}[(W - \mathbb{E}[W])^2] \geq 0 \tag{2.39}$$

$\blacksquare$

2.6   Bernoulli Trials

Definition 2.12. A random variable Z is said to be a **Bernoulli trial** if

$$Z \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix} \tag{2.40}$$

with $p \in [0,1]$, denoted as $Z \sim bern(p)$.

Remark 2.3. Let $A \subseteq \Omega$, the **Bernoulli trial** associated with A is defined to be the finite scheme distribution

$$Z = I(A) \sim \begin{pmatrix} 0 & 1 \\ 1 - P(A) & P(A) \end{pmatrix} \tag{2.41}$$

Proposition 2.12 (Invariance)**.** Given the outcome space of Bernoulli trials to be $\{0, 1\}$, $Z^s = Z$ for every $s \in \mathbb{N}$.
Remark: Z^2 and Z are indeed equal, which is a stronger statement than equal in distribution.

Proposition 2.13 (Negation)**.** $Z \sim bern(p) \quad \Leftrightarrow \quad 1 - Z \sim bern(q)$ where $p + q = 1$.

2.7   Standard Uniform Distribution

2.8   Scaled Exponential Distribution

2.9   Gamma Distribution

2.10   Distribution Functions in General

Definition 2.13. A probability measure $P : \mathcal{F} \to [0,1]$ is said to be σ**-additive** if for any countable mutually disjoint events (A_n),

$$P(\sum_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n) \tag{2.42}$$

Definition 2.14. Let (A_n) be a sequence of subsets of Ω, let $A \subseteq \Omega$. Then $(A_n) \to A$ if and only if $(I(A_n))$ point-wise converges to $I(A)$.

Definition 2.15. A probability measure P is **continuous** if for any convergent sequence $(A_n) \to A$ in $\mathcal{F}$, $P(A_n) \to P(A)$.
Remark: the continuity of P is only defined through the sequential definition, there is δ-ε definition for it.

Remark 2.4. The *constructor of indicator function* is a bijection between $\mathcal{P}(\Omega)$ and function space $\{0,1\}^\Omega$.

Proposition 2.14 (Set Theoretic Limits). A sequence of sets (A_n) converges if and only if that $I(A_n)$ converges, which is equivalent to $\lim_{n\to\infty} I(A_n) = \lim \sup I(A_n) = \lim \inf I(A_n)$:

$$\lim_{n\to\infty} I(A_n) = \overset{\infty}{\underset{n=1}{\lim \sup_{k\geq n}}} I(A_k) \tag{2.43}$$

$$= \overset{\infty}{\underset{n=1}{\inf}} \sup_{k\geq n} I(A_k) \tag{2.44}$$

$$= \overset{\infty}{\underset{n=1}{\inf}} I\left(\bigcup_{k\geq n} A_k\right) \tag{2.45}$$

$$= I\left(\overset{\infty}{\underset{n=1}{\bigcap}} \bigcup_{k\geq n} A_k\right) \tag{2.46}$$

Similarly,

$$\lim_{n\to\infty} I(A_n) = \overset{\infty}{\underset{n=1}{\lim \inf_{k\geq n}}} I(A_k) \tag{2.47}$$

$$= \overset{\infty}{\underset{n=1}{\sup}} \inf_{k\geq n} I(A_k) \tag{2.48}$$

$$= I\left(\overset{\infty}{\underset{n=1}{\bigcup}} \bigcap_{k\geq n} A_k\right) \tag{2.49}$$

Therefore,

$$(A_n) \to A \iff A = \overset{\infty}{\underset{n=1}{\bigcup}} \bigcap_{k\geq n} A_k = \overset{\infty}{\underset{n=1}{\bigcap}} \bigcup_{k\geq n} A_k \tag{2.50}$$

Proposition 2.15 (Special Cases)**.**