Notes on Probability and Computing

Xu Zhean

January 15, 2022

Contents

1 Events and Probability

A *probability space* is a *measure space* $(\Omega, \mathcal{F}, \mathbb{P})$ consisting of:

- the *sample space* Ω — a set of outcomes called *sample*;
- the *σ-algebra* $\mathcal{F}$ — a family of subsets of Ω, called *events*, such that $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complements (i.e. $\forall A \in \mathcal{F}$, $\Omega \backslash A \in \mathcal{F}$) and countable unions (i.e. $\forall A_i \in \mathcal{F}$, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$);
- the *probability function* $\mathbb{P} : \mathcal{F} \to [0,1]$ such that $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}$ is *σ-additive* (i.e. $\mathbb{P}\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$).

The motivation behind this complicated definition is that some sets are *non-measurable*, thus mathematicians developed the theory of *measure*. For instance, *Borel set* on real line forms a σ-algebra which is *generated by* open intervals. *Stieltjes measures* is a *Borel measure* and builds the measure-theoretic foundation of *continuous probability distribution*.

Lemma 1.1 (Inclusion-exclusion principle) *Let $E_1, \cdots, E_n$ be any n events. Then*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{\ell=1}^{n} (-1)^{\ell+1} \sum_{i_1 < i_2 < \cdots < i_\ell} \mathbb{P}\left(\bigcap_{r=1}^{\ell} E_{i_r}\right).$$

Events $E_1, E_2, \cdots, E_n$ are *mutually independent* (simply called *independent* when $k = 2$) if and only if, for any subset $I \subseteq \{1, 2, \cdots, k\}$, $\mathbb{P}\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} \mathbb{P}(E_i)$. Note that events $X, Y, Z, \cdots$ are unnecessarily mutually independent when they are pairwise independent.

The *conditional probability* that event E occurs given that event F occurs is $\mathbb{P}(E \mid F) = \mathbb{P}(E \cap F) / \mathbb{P}(F)$ $(\mathbb{P}(F) > 0)$.

Theorem 1.2 (Law of total probability) *Let events $\bigsqcup_{i=1}^{n} E_i = \Omega$. Then we have $\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \mid E_i) \cdot \mathbb{P}(E_i)$.*

Theorem 1.3 (Bayes's law) *Let events $E_1, E_2, \cdots, E_n$ satisfy $\bigsqcup_{i=1}^{n} E_i = \Omega$. Then we have*

$$\mathbb{P}(E_k \mid B) = \frac{\mathbb{P}(E_k \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid E_k) \cdot \mathbb{P}(E_k)}{\sum_{i=1}^{n} \mathbb{P}(B \mid E_i) \cdot \mathbb{P}(E_i)}.$$

In the *Bayesian approach* one starts with a *prior* model, giving some initial value to the model parameters. This model is then modified, by incorporating new observations, to obtain a *posterior* model that captures the new information.

Exercise 1.6 Using mathematical induction, we have $p_{i,j} = \frac{i-1}{i+j-1} \cdot p_{i-1,j} + \frac{j-1}{i+j-1} \cdot p_{i,j-1} = \frac{i+j-2}{i+j-1} \cdot \frac{1}{i+j-2} = \frac{1}{i+j-1}$.

Exercise 1.7.b Let $F_{b_1 b_2 \cdots b_n}$ be the intersection of events E_i $(b_i = 1)$ or $\Omega \backslash E_i$ $(b_i = 0)$, and P_k be the sum of $\mathbb{P}(F_b)$ where b consists of k one and $n-k$ zero. Then for every $k \geq 1$, we have $\sum_{i=1}^{l} (-1)^{i+1} \binom{k}{i} = 1 + (-1)^{l+1} \binom{k-1}{l} \geq 1$. Multiply both sides by P_k and sum them up. We eventually reach the desired inequality.

Exercise 1.11.b $p_3 = p_1 \cdot (1-p_2) + (1-p_1) \cdot p_2 \Rightarrow q_3 = 1 - 2p_3 = (1-2p_1)(1-2p_2) = q_1 q_2$. Is there any underlying motivation?

Exercise 1.24 (Karger's algorithm) Let K be the minimum r-way cut-set. Considering all r-way cut-sets consisting of $r-1$ single vertex, the total size is $m \cdot \binom{n-2}{r-1}$ with an upper bound $(m - |K|) \cdot \binom{n}{r-1}$. It follows that

$$m \cdot \binom{n-2}{r-1} \leq (m - |K|) \cdot \binom{n}{r-1} \quad \Rightarrow \quad 1 - \frac{|K|}{m} \geq \binom{n-2}{r-1} \binom{n}{r-1}^{-1} = \frac{(n-r+1)(n-r)}{n(n-1)}.$$

The probability that K survives all the $n-r$ iterations is at least

$$\prod_{i=0}^{n-r-1} \frac{(n-i+1-r)(n-i-r)}{(n-i)(n-i-1)} = r \cdot \binom{n}{r-1}^{-1} \binom{n-1}{r-1}^{-1}$$

and its reciprocal is the maximum possible number of minimum cardinality of r-way cut-sets.

2 Discrete Random Variables and Expectation

A *(real-valued) random variable* X on a sample space Ω is a *measurable function* $X : \Omega \to \mathbb{R}$, and a *discrete random variable* is one which may take on only a countable number of distinct values. "$X = a$" represents the set $\{s \in \Omega \mid X(s) = a\}$, and we denote the probability of that event by $\mathbb{P}(X = a) = \sum_{s \in \Omega : X(s) = a} \mathbb{P}(s)$.

Random variables $X_1, X_2, \cdots, X_n$ are *mutually independent* (simply called *independent* when $k = 2$) if and only if, for any subset $I \subseteq \{1, 2, \cdots, k\}$ and any values x_i $(i \in I)$, $\mathbb{P}(\bigcap_{i \in I} (X_i = x_i)) = \prod_{i \in I} \mathbb{P}(X_i = x_i)$.

The *expectation* of a discrete random variable X, denoted by $\mathbb{E}[X]$, is given by $\mathbb{E}[X] = \sum_i i \cdot \mathbb{P}(X = i)$. Note that the infinite series needs to be *absolutely convergent* (i.e. rearrangements do not change the value of the sum).

Theorem 2.1 (Linearity of expectation) *For discrete random variables $X_1, X_2, \cdots, X_n$ with finite expectations and any contants $c_1, c_2, \cdots, c_n$, we have $\mathbb{E}[\sum_{i=1}^n c_i X_i] = \sum_{i=1}^n c_i \mathbb{E}[X_i]$.*

proof. Observe that we only need to prove the following two cases:

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_i \sum_j (i+j) \cdot \mathbb{P}((X=i) \cap (Y=j)) \\
&= \sum_i i \sum_j \mathbb{P}((X=i) \cap (Y=j)) + \sum_j j \sum_i \mathbb{P}((X=i) \cap (Y=j)) = \mathbb{E}[X] + \mathbb{E}[Y], \\
\mathbb{E}[cX] &= \sum_i i \cdot \mathbb{P}(cX = j) = c \cdot \sum_j (j/c) \cdot \mathbb{P}(X = j/c) = c \cdot \sum_k k \cdot \mathbb{P}(X = k) = c \cdot \mathbb{E}[X].
\end{aligned}$$

When there are countably infinite variables, the situation becomes more subtle. We will discuss it later. □

Theorem 2.2 (Jensen's inequality) *If f is a convex function, then $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$.*

proof. Assume that f has a Taylor expansion. Let $\mu = \mathbb{E}[X]$. By Taylor's theorem, there is a value c such that

$$f(x) = f(\mu) + f'(\mu)(x-\mu) + \frac{f''(c)(x-\mu)^2}{2} \geq f(\mu) + f'(\mu)(x-\mu)$$

Taking expectations of both sides

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(\mu) + f'(\mu)(X-\mu)] = \mathbb{E}[f(\mu)] + f'(\mu)(\mathbb{E}[X] - \mu) = f(\mu) = f(\mathbb{E}[X])$$

An alternative proof will be presented in Exercise 2.10. □

Define *conditional expectation* $\mathbb{E}[Y \mid Z = z] = \sum_y y \cdot \mathbb{P}(Y = y \mid Z = z)$ and $\mathbb{E}[Y \mid Z]$ as a random variable $f(Z)$ that takes on the value $\mathbb{E}[Y \mid Z = z]$ when $Z = z$.

Theorem 2.3 (Law of total expectation) *For any random variables X and Y,*

$$\mathbb{E}[X] = \sum_y \mathbb{P}(Y = y) \cdot \mathbb{E}[X \mid Y = y] = \mathbb{E}[\mathbb{E}[X \mid Y]].$$

A *Bernoulli* random variable X takes 1 with probability p and 0 with probability $1-p$. A *binomial* random variable X with parameters n and p, denoted by $B(n,p)$, is defined by *probability distribution* $\mathbb{P}(X = k) = \binom{n}{k} \cdot p^k (1-p)^{n-k}$, $n = 0, 2, \cdots, n$. Its expectation is np.

A *geometric* random variable X with parameter p is defined by probability distribution $\mathbb{P}(X = n) = (1-p)^{n-1} p$, $n = 1, 2, \cdots$. Its expectation is $1/p$. Geometric random variables are *memoryless*, that is, one ignores past failures as distribution does not change. Formally, we have the following statement.

Lemma 2.4 (Memorylessness) *Let X be a geometric random variable with parameter p. Then, for $n > 0$,*

$$\mathbb{P}(X = n + k \mid X > k) = \mathbb{P}(X = n).$$

Lemma 2.5 *Let X be a discrete random variable that takes on only nonnegative integer values. Then,*

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) = \sum_{1 \leq i \leq k} \mathbb{P}(X = k) = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i)$$