MATH4240 Tutorial 1 Notes

This is a review note on the basics of probability. See also the Summary Note of Chapter 0.

1 Basic Concepts

Probability is a measurement of uncertainty.

A *probability space* consists of

- a *sample space* Ω which consists of all possible outcomes
- an *event space* $\mathcal{F} \subseteq 2^{\Omega}$[1] which encodes the information available as a σ-algebra[2]
- a *probability* measure $P : \mathcal{F} \to [0, 1]$ which measures the uncertainty of a given event

Example 1. Suppose I have a coin, which when tossed gives either **H** (head) or **T** (tail). I now toss the coin 2 times. Let us denoted the ultimate outcome as ω. Then

- the sample space is $\Omega = \{ \mathbf{HH}, \mathbf{HT}, \mathbf{TH}, \mathbf{TT} \}$
- before any coin toss, a natural event space is $\mathcal{F}_0 = \{ \emptyset, \Omega \}$
- after the first coin toss, a natural event space is $\mathcal{F}_1 = \{ \emptyset, \Omega, \{ \mathbf{HH}, \mathbf{HT} \}, \{ \mathbf{TH}, \mathbf{TT} \} \}$
- after both coin tosses, a natural event space is $\mathcal{F}_2 = 2^{\Omega}$

The basic properties of a probability measure are

- $0 \leq P(E) \leq 1$, with $P(\emptyset) = 0$, $P(\Omega) = 1$
- if $E_1, E_2, \ldots$ are disjoint, then $P(\bigcup E_n) = \sum P(E_n)$
- if $E_1 \subseteq E_2 \subseteq \ldots$, then $P(E_n) \to P(\bigcup E_n)$
- if $E_1 \supseteq E_2 \supseteq \ldots$, then $P(E_n) \to P(\bigcap E_n)$

A *random variable* is a (measurable) function $X : \Omega \to A \subseteq \mathbb{R}$ that quantifies the outcome. Commonly probability is measured according to the value of X via events of the form $\{ \omega \mid X(\omega) \in E \}$[3], with shorthand notation $P(X \in E) = P(\{ \omega \mid X(\omega) \in E \})$.

2 Conditional, Independence

On two events A, B, assuming $P(B) \neq 0$, then the *conditional probability* of A given B is $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Two events A, B are *independent* if $P(A \cap B) = P(A) P(B)$. If $P(B) \neq 0$, this is equivalent to $P(A|B) = P(A)$.

Two random variables X, Y are *independent* ($X \perp Y$) if for all (Borel) sets $A, B \subseteq \mathbb{R}$, $\{ X \in A \}$ and $\{ Y \in B \}$ are independent.

Theorem 2.1 (total probability)**.** *If $E_1, \ldots$ are disjoint, then $P(A) = \sum P(A \cap E_n)$.*
In particular, if $P(E_n) \neq 0$ for each n, then $P(A) = \sum P(A|E_n) P(E_n)$

Theorem 2.2 (Bayes)**.** $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. *Usually $P(B)$ can be computed with total probability.*

[1]Here 2^{Ω} is the power set of Ω.

[2]A collection of sets $\mathcal{F} \subseteq 2^{\Omega}$ is a σ-algebra if $\emptyset \in \mathcal{F}$, $E \in \mathcal{F}$ implies $\Omega \setminus E \in \mathcal{F}$, and $E_1, E_2, \ldots \in \mathcal{F}$ implies $\bigcup E_n \in \mathcal{F}$. Precisely speaking, it is a collections of sets that we can "assign" a probability to.

[3]That is, we measure the probability on X by measuring the probability on the corresponding ω.

3 Distribution

For a random variable X, the *cumulative distribution function (cdf)* is $F_X(t) = P(X \leq t)$, and

- If the range of X is countable, then we call it a *discrete* random variable.
  For simplicity, we will assume that its range is $\mathbb{N}$ (or $\mathbb{Z}$).
  Its *probability mass function (pmf)* is $p_X(n) = P(X = n)$.
- If there exists a function $f_X : \mathbb{R} \to \mathbb{R}$ such that $P(X \in A) = \int_A f_X$, then we call X a *continuous* random variable, and f_X its *probability density function (pdf)*.
  Note that pdf *may not be unique* (e.g. differ at a single point).
- If X is neither discrete nor continuous, we call it *mixed*.

We will call both pmf and pdf *density*.

We can also consider the *joint distribution* of two random variables X, Y:

- if X, Y are discrete, the joint pmf is $p_{X,Y}(x, y) = P(X = x,\ Y = y)$
- if X, Y are continuous, the joint pdf $f_{X,Y}(x, y)$ is a function that satisfies $P(X \in A,\ y \in B) = \int_{A \times B} f_{X,Y}$ for all A, B

The basic properties are

- $0 \leq F_X \leq 1$ is non-decreasing right-continuous with left limit everywhere
- for pmf, $0 \leq p_X \leq 1$, $F_X(n) = \sum_{i \leq n} p_X(i)$, and $\sum p_X(i) = 1$
- for pdf, $f_X \geq 0$, with $a \leq b$, $P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X$, and $\int_{-\infty}^{\infty} f_X = 1$
- if X, Y are independent, then $p_{X,Y} = p_X p_Y$ if discrete and $f_{X,Y} = f_X f_Y$ if continuous

We can also consider the conditional distribution: assuming X is continuous with pdf f_X,

- on event A with nonzero probability, $F_{X|A}(t) = P(\ X \leq t \mid A\ )$
- on continuous random variable Y, $f_{X|Y}(x \mid y) = f_{X,Y}(x, y)/f_Y(y)$ assuming $f_Y(y) > 0$

Example 2. The common examples of probability distributions are:

distribution	notation*	density
discrete uniform	$\mathrm{Unif}\{a, b\}$	$p_X(i) = \frac{1}{b-a+1}$ for $i \in \{\ a, \ldots, b\ \}$
continuous uniform	$\mathrm{Unif}(a, b)$	$f_X(x) = \frac{1}{b-a}\chi_{[a,b]}$
Bernoulli	$\mathrm{Bern}(p)$	$p_X(1) = p$, $p_X(0) = 1 - p$
binomial	$\mathrm{Bin}(n, p)$	$p_X(r) = \binom{n}{r} p^r (1-p)^{n-r}$ for $r \in \{\ 0, \ldots, n\ \}$
geometric	$\mathrm{Geom}(p)$	$p_X(n) = (1-p)^{n-1} p$ for $n \geq 1$
Poisson	$\mathrm{Poi}(\lambda)$	$p_X(n) = e^{-\lambda} \frac{\lambda^n}{n!}$ for $n \geq 0$
exponential	$\mathrm{Exp}(\lambda)$	$f_X(x) = \lambda e^{-\lambda x} \chi_{x \geq 0}$
normal	$N(\mu, \sigma^2)$	$f_X(x) = (2\pi\sigma^2)^{-1/2} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$

* Notation seems to vary per reference.

with $\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$ being the indicator function of set A, and (as a shorthand notation) $\chi_{x \geq 0} = \chi_{[0,\infty)}$.

4 Moment

The *expectation / expected value / mean* of a random variable X is $E(X) = \sum np_X(n)$ if discrete, $E(X) = \int xf_X(x)$ if continuous (if exists).

The *variance* of a random variable X with a finite mean $\mu = E(X)$ is $\mathrm{Var}(X) = E\left((X-\mu)^2\right)$ (if exists).

Conditional expectation and variance are defined similarly, only that e.g. the conditional expectation with respect to a random variable $E(X|Y) : \mathbb{R} \to \mathbb{R}$ is now a random variable $(E(X|Y))(y) = E(X|Y=y)$:

- if Y is discrete, $E(X|Y=y) = \sum xp_{X|Y}(x \mid y)$
- if Y is continuous, $E(X|Y=y) = \int xf_{X|Y}(x \mid y)$

The basic properties are

- E is linear: $E(aX) = aE(X)$ and $E(X+Y) = E(X) + E(Y)$, *whether X, Y are independent or not*
- $\mathrm{Var}(X) = E\left(X^2\right) - E(X)^2$ and $\mathrm{Var}(cX) = c^2 \mathrm{Var}(X)$
- if X, Y are independent, then $E(XY) = E(X)E(Y)$ and (thus) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.
  In general, we have $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(E(XY) - E(X)E(Y))$.
- (total expectation) if $A_1, \ldots$ are disjoint, then $E(X) = \sum E(X|A_n) P(A_n)$.
  In particular, (tower property) $E(X) = E(E(X|Y))$ and $P(X \in A) = E(P(\ X \in A \mid Y\ ))$
- (total variance) $\mathrm{Var}(X) = E(\mathrm{Var}(X|Y)) + \mathrm{Var}(E(X|Y))$

As an exercise, you should compute the mean and the variance of the distributions listed in the table above.

Example 3.

- if $X \sim \mathrm{Poi}(\lambda)$, then $E(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$
- if $X \sim \mathrm{Exp}(\lambda)$, then $E(X) = 1/\lambda$ and $\mathrm{Var}(X) = 1/\lambda^2$

Example 4. For a randomly shuffled deck of 52 poker cards, where 4 of them are aces, the expected number of cards placed before the first ace can be computed as follows: number the non-ace cards as $1, \ldots, 48$, and for each i let X_i be the random variable $X_i = \begin{cases} 1 & \text{if card } i \text{ is placed before the first ace} \\ 0 & \text{otherwise} \end{cases}$. Then the expected number of cards placed before the first ace is $E\left(\sum X_i\right) = \sum E(X_i) = 48/5$ as you can easily argue that $E(X_i) = 1/5$ for each card.

5 Computations

5.1 Sum of Random Variables

If X, Y are independent continuous random variables, then the cdf of their sum $X+Y$ can be computed as

$$\begin{aligned} F_{X+Y}(z) &= P(X+Y \le z) \\ &= \int P(X+Y \le z | X=x) f_X(x) \\ &= \int P(Y \le z-x) f_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_Y(y) f_X(x) \,\mathrm{d}y \,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_Y(y-x) f_X(x) \,\mathrm{d}y \,\mathrm{d}x \end{aligned}$$

and so

$$f_{X+Y}(z) = \frac{\mathrm{d}}{\mathrm{d}z} F_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z-x) f_X(x) \,\mathrm{d}x = (f_X * f_Y)(z)$$

which means that the pdf f_{X+Y} of $X+Y$ is the *convolution* of f_X and f_Y. A similar conclusion holds for discrete random variables.

Example 5. If $X \sim \mathrm{Poi}(\lambda_X)$, $Y \sim \mathrm{Poi}(\lambda_Y)$ are independent, then $X+Y \sim \mathrm{Poi}(\lambda_X + \lambda_Y)$

5.2 Order Statistics

If $X_1, \dots, X_n$ are independent random variables, then the cdf of their maximum $Y = \max(X_1, \dots, X_n)$ can be computed as

$$F_Y(y) = P\left(\max(X_1, \dots, X_n) \le y\right) = \prod P\left(X_i \le y\right) = \prod F_{X_i}(y)$$

Same approach can also give the cdf of the k-largest value.

Example 6. If $X_1, \dots, X_n$ are independent random variables and $X_i \sim \mathrm{Exp}(\lambda_i)$, then for $Y = \min(X_1, \dots, X_n)$,

$$F_Y(y) = 1 - P\left(Y > y\right) = 1 - \prod P\left(X_i > y\right) = 1 - \prod(1 - F_{X_i}(y)) = 1 - \prod e^{-\lambda_i y}\chi_{y \ge 0} = 1 - e^{-(\sum \lambda_i)y}\chi_{y \ge 0}$$

This implies $\min(X_1, \dots, X_n) = Y \sim \mathrm{Exp}(\sum \lambda_i)$.

Furthermore, on $Z = \min(X_2, \dots, X_n) \sim \mathrm{Exp}(\lambda)$ with $\lambda = \sum_{i \ge 2} \lambda_i$,

$$P\left(\min(X_1, \dots, X_n) = X_1\right) = P\left(X_1 \le Z\right) = \int P\left(X_1 \le z\right) f_Z(z) = \int_0^\infty (1 - e^{-\lambda_1 z})\lambda e^{-\lambda z} = \frac{\lambda_1}{\lambda_1 + \lambda} = \frac{\lambda_1}{\sum \lambda_i}$$