Probability Theory

Sets

Definitions

Keyword	Definition
Sample space S	The set of all possible outcomes.
Event A	An event A is a subset of S.
$A \cap B$	A and B occured.
$A \cup B$	A or B occured.
$\overline{A}$	A did not occur.

Properties of set operations.

Commutative

$$A \cap B = B \cap A$$
$$A \cup B = B \cup A$$

Associative

$$(A \cap B) \cap C = A \cap (B \cap C)$$
$$(A \cup B) \cup C = A \cup (B \cup C)$$

Definition
A and B are **mutually exclusive** iff $A \cap B = \emptyset$

The following identities are true.

$$A \cap \overline{A} = \emptyset$$
$$A \cup \overline{A} = S$$
$$S - B = \overline{B}$$
$$A - B = A \cap \overline{B}$$
$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

Probability

The probability $P(A)$ of A happening is defined as

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

Where N_A is the number of events in N experiments.

Properties of probabilities.
- $0 \leq P(A) \leq 1$
- $P\left(A \cap \overline{A}\right) = 0$
- $P\left(A \cup \overline{A}\right) = 1$
- $P\left(\overline{A}\right) = 1 - P(A)$

The union of two events $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If A and B are **mutually exclusive**, then $P(A \cup B) = P(A) + P(B)$
- Extending for three events:

$P(A \cup B \cup C)$
$= P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$

This can be proved using the rule for two events on $P(A \cup (B \cup C))$

Definition
Conditional probability: $P(B|A)$ is the probability of B occuring given A.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Because $P(A \cap B) = P(A)P(B|A)$, we have

$$\begin{aligned} P(A)P(A|B)P(A|B \cap C) &= P(A \cap B)P(A|B \cap C) \\ &= P(A \cap B \cap C) \end{aligned}$$

Bayes Theorem
It is obvious that

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(A)P(A|B)}{P(B)} \end{aligned}$$

Provided $P(B) \neq 0$

It is also true that

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P\left(\overline{A}\right)P\left(B|\overline{A}\right)}$$

Combinatorics

Permutations and Combinations

- The number of permutations for choosing r elements from n elements is

$$^nP_r = \frac{n!}{(n-r)!}$$

- The number of combinations (order doesn't matter) is therefore

$$^nC_r = \frac{n!}{r!(n-r)!}$$

^nC_r is called the **binomial coefficients**.

$$(p+q)^n = \sum_{r=0}^{n} {^nC_r} q^r p^{n-r}$$

From **Pascal's triangle** we have $^nC_r = {^{n-1}C_r} + {^{n-1}C_{r-1}}$, which we can also prove

$$\begin{aligned}\frac{(n-1)!}{r!(n-1-r)!}+\frac{(n-1)!}{(r-1)!(n-r)!}&=\frac{(n-r)(n-1)!+r(n-1)!}{r!(n-r)!}\\&=\frac{n!}{r!(n-r)!}\end{aligned}$$

Arrangements

Suppose there are r identical objects R, t of T, s of S, then the number of distinguishable arrangment is

$$\frac{n!}{r!s!t!}$$

Where $n = r + s + t$, and $r!$ is the number of (non-distinguishable) arrangments for R, etc.

Balls in Boxes

How to put p balls into q boxes.

We can write the situation out in $\square\square \mid \square \mid \mid \square \mid \cdots$, there are p balls and $q - 1$ walls.
- Number of permutations of all objects is $(p + q - 1)!$
- Number of distinguishable arrangment is

$$\frac{(p+q-1)!}{p!(q-1)!}$$

Discrete Probability Distributions

Definitions

Keyword	Definition
Random variable	A variable whose value is determined by the outcome of an experiment.
Discrete variable	The variable only takes discrete values.
Probability function	$f(x)$ is the probability that X takes the value x.

If there are only n values that $f(x)$ can take, is true that

$$\sum_{i=0}^{n-1} f(x_i) = 1$$

The **cumulative probability function** is

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

$$P(a < x \le b) = F(b) - F(a)$$

Mean and Variance

Definition

Mean is also called the expected value, denoted $E(X)$ or $<X>$

$$E(X) = \sum_{i=0}^{n-1} x_i f(x_i)$$

Properties of Mean

$$E(aX) = aE(X)$$
$$E(X+Y) = E(X) + E(Y) \text{ where } X \text{ and } Y \text{ are independent}$$
$$E(g(X)) = \sum_{i=0}^{n-1} g(x_i) f(x_i)$$

Definition
Variance measures how the results spread around the mean.

$$\sigma^2 = E((X - E(x))^2) = \overline{\left(X - \overline{X}\right)^2}$$

The **standard deviation** $\sigma = \sqrt{\sigma^2}$

An alternative way to calculate the variance is

$$\begin{aligned}\sigma^2 &= E((X-\mu)^2) \\ &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2\end{aligned}$$

Discrete Probability Distributions

The Binomial Distribution

Note
The handout use a different way of calculating expected value and variance.

Expecation of value of X

$$\begin{aligned}E(X) &= \sum_{k=0}^{m} k \binom{m}{k} p^k (1-p)^{m-k} \\ &= \sum_{k=0}^{m} q \times \binom{m}{k} \left(\frac{d}{dq} q^k\right) (1-p)^{m-k} \text{ where } p = q \\ &= q \times \frac{d}{dq} \sum_{k=0}^{m} \binom{m}{k} q^k (1-p)^{m-k} \\ &= q \times \frac{d}{dq} (q+1-p)^m \text{ by binomial theorem} \\ &= mq(q+1-p)^m \\ &= mp\end{aligned}$$

Variance of X

$$\begin{aligned}E(X^2) &= \sum_{k=0}^{m} k^2 \binom{m}{k} q^k (1-p)^{n-k} \\ &= \sum_{k=0}^{m} (k(k-1)+k) \binom{m}{k} q^k (1-p)^{n-k} \\ &= mp + \sum_{k=0}^{m} k(k-1) \binom{m}{k} q^k (1-p)^{n-k} \text{ by previous result} \\ &= mp + \sum_{k=0}^{m} k(k-1) \binom{m}{k} q^k (1-p)^{n-k} \\ &= mp + q^2 \times \frac{d^2}{dq^2} \sum_{k=0}^{m} \binom{m}{k} q^k (1-p)^{n-k} \\ &= mp + q^2 \times \frac{d^2}{dq^2} (q+1-p)^m \\ &= mp + p^2 \times m(m-1) \\ E(X^2) - E(X)^2 &= mp(1-p)\end{aligned}$$

The Poisson Distribution

For binomial distribution where $p \ll 1$ and $np = \lambda$, as $n \to \infty$, we have another well defined distribution

$$\begin{aligned}\mu &= np = \lambda \\ \sigma^2 &= np(1-p) \\ &= \lambda\end{aligned}$$

When n is large and p is small, the poisson distribution is a good approximation for the binomial distribution.

> **Note**
> Missing notes because the camera in the recording pointing towards the sheet of paper is off.

$$\begin{aligned}P(X=r) &= \frac{\lambda^r e^{-\lambda}}{r!} \\ \sum_{r=0}^{\infty} P(X=r) &= e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} \\ &= 1\end{aligned}$$

The poissoin distribution is useful for events that happens in time:
- Probability of something happening at a time is very small.
- **Over a unit time**, the probability is $\lambda^r e^{-\lambda}/r!$

Continuous Probability Distribution

The normal distribution is a **continuous probability distribution** given by the **probability density function**, which gives the *probabilty per unit length.*

$$P(x \le X \le x + dx) = f(x)dx$$

The probability for any individual $P(X = x)$ is zero. For a fixed range $\alpha \le X \le \beta$

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x)dx$$

$f(x)$ obeys properties:
- $0 \leq f(x) < \infty$ for all x
- $$\int_{-\infty}^{\infty} f(x)dx = 1$$

Definition
The **cumulative probability function** is defined as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

Which has property $F'(x) = f(x)$.

And also these equations are true.

$$P(\alpha \leq X \leq \beta) = F(\beta) - F(\alpha)$$

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$\sigma^2 = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

Uniform Probability Distribution

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{when } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{when } x < \alpha \\ \frac{x-\alpha}{\beta-\alpha} & \text{when } \alpha \leq x \leq \beta \\ 1 & \text{when } x > \beta \end{cases}$$

$$\mu = \frac{\alpha + \beta}{2}$$

$$\sigma^2 = \frac{(\beta - \alpha)^2}{12}$$

This makes sense because σ^2 should only depend on the difference between α and β, not their individual values.

The Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2\right)$$

$$F(x) = \frac{1}{2} + \frac{1}{2} \exp\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)$$

You are required to know how to prove the result for $F(x)$ and the mean, variance of the normal distribution from definition, not included here.

Central Limit Theorem

If x are samples taken from a distribution, then the arithmetic mean over n samples is normally distributed as n becomes large.

$$\overline{x} = \frac{1}{n} \sum_{i=0}^{n-1} x_i$$