PART I. Basic probability.

1. Permutations and combinations.

(a) (Permutations.) The number of ways of picking r objects out of n objects, keeping track of order, is

$$n \cdot (n-1) \cdot (n-2) \cdots (n-r+1) = \frac{n!}{(n-r)!}.$$

This number is sometimes called "n pick r," written ${}_nP_r$.

(b) (Combinations.) The number of ways of choosing r objects out of n objects, without keeping track of order, is

$$\frac{n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)}{r!} = \frac{n!}{r!(n-r)!}.$$

This number is sometimes called "n choose r," written ${}_nC_r$ or $\binom{n}{r}$.

2. Probability axioms.

(a) $P(A) \geq 0$ for any event A.

(b) $P(S) = 1$, where S is the sample space.

(c) If the events $A_1, A_2, A_3, A_4, \ldots$ are mutually exclusive (no two of them can happen together), then

$$P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + P(A_4) + \cdots.$$

(The list $A_1, A_2, A_3, A_4, \ldots$ could be finite or infinite.)

(d) If all outcomes in a sample space S are equally likely, and $n(A)$ denotes the number of outcomes in the event A, then

$$P(A) = \frac{n(A)}{n(S)}$$

(assuming the sample space is a finite set).

(e) If all outcomes in a sample space S are equally likely, and $n(A)$ denotes the number of outcomes in the event A, then

$$P(A) = \frac{n(A)}{n(S)}$$

(assuming the sample space is a finite set).

(f) If events $A_1, A_2, A_3, A_4, \ldots$ are independent (they don't affect each other), then

$$P(A_1 A_2 A_3 A_4 \cdots) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot P(A_4) \cdots.$$

(The list $A_1, A_2, A_3, A_4, \ldots$ could be finite or infinite.)

3. Probability rules.

(a) For any event A,

$$P(A) = 1 - P(A^c),$$

where A^c denotes the complement of A (meaning all outcomes in the sample space except those in A).

(b) For any events A and B (not necessarily mutually exclusive), we have

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

(c) For any events A, B, and C (not necessarily mutually exclusive), we have

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

PART II. Conditional probability.

1. Formulas for $P(A|B)$.

(a) Suppose all events in your sample space are equally likely. Then for any events A and B, we have

$$P(A|B) = \frac{n(AB)}{n(B)}.$$

(b) Given any events A and B, we have

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

(whether or not events are equally likely).

2. Formulas for $P(AB)$.

(a) Given any events A and B, we have

$$P(AB) = P(B) \cdot P(A|B).$$

(b) (Generalization.) Given any events A, B, and C, we have

$$P(ABC) = P(C) \cdot P(B|C) \cdot P(A|BC).$$

(c) (Further generalization.) Given any finite or infinite list of events, we have

$$P(\text{all events happen}) = P(\text{first one happens}) \cdot P(\text{second happens given that first does}) \cdot P(\text{third does given that first two do}) \cdot P(\text{fourth does given that first three do}) \cdots .$$

3. Conditional probability and mutually exclusive events.

(a) For any events A and B,

$$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c)$$

(again, B^c denotes the complement of B).

(b) (Generalization.) Suppose $B_1, B_2, B_3, \ldots, B_n$ are mutually exclusive events, and the event A can only happen if one of the events B_i happens. Then

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) + \cdots + P(B_n)P(A|B_n).$$

PART III. Random variables, expected value, variance, standard deviation.

1. Definition of random variable.

A random variable X is a function defined on the sample space S of an experiment. That is, a random variable X is a way of assigning a number to each possible outcome of an experiment.

2. Random variable probabilities.

If X is a random variable, then $P(X = x)$ is the probability of X taking the value x.

3. Probability mass function.

"Compute the probability mass function of X" just means "Compute $P(X = x)$ for every possible value x of X."

4. Expected value.

If X is a random variable, then the expected value $E(X)$ is defined by

$$E(X) = \sum_x x \cdot P(X = x),$$

where the sum is over all possible values x of X.

5. Sum rule for expected values.

If $X_1, X_2, X_3, \dots X_n$ are random variables, then

$$E(X_1 + X_2 + X_3 \cdots + X_n) = E(X_1) + E(X_2) + E(X_3) + \cdots + E(X_n).$$

6. Variance and standard deviation.

If X is a random variable, then the variance $\text{var}(X)$ is defined by

$$\text{var}(X) = E((X - \mu)^2 = \sum_x (x - \mu)^2 \cdot P(X = x),$$

where $\mu = E(X)$. Also, the standard deviation $\text{std}(X)$ is defined by $\text{std}(X) = \sqrt{\text{var}(X)}$.

7. Sum rule for variance.

If the random variables $X_1, X_2, X_3, \dots X_n$ are independent, then

$$\text{var}(X_1 + X_2 + X_3 \cdots + X_n) = \text{var}(X_1) + \text{var}(X_2) + \text{var}(X_3) + \cdots + \text{var}(X_n).$$

PART IV. Statistical inference.

1. Relative frequency density (RFD).

$$RFD = \frac{\text{frequency (number of data points in given bin)}}{(\text{bin width } B) \times (\text{data set size } n)}$$

2. Sample mean and standard deviation.

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}, \quad s = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n - 1}}.$$

3. Sample mean and standard deviation for grouped data.

$$\overline{x} = \frac{f_1 y_1 + f_2 y_2 + \cdots + f_m y_m}{n}, \quad s = \sqrt{\frac{f_1(y_1 - \overline{x})^2 + f_2(y_2 - \overline{x})^2 + \cdots + f_m(y_m - \overline{x})^2}{n - 1}},$$

where $y_1, y_2, \dots, y_m$ are the distinct values that the data assumes, and each value y_k occurs f_k times.

4. Standard normal probabilities. If X is $N(0,1)$, then

$$P(-L < x < L) = p,$$

where the "critical value" L corresponds to the probability p as follows:

- $L = 1$: $p = 0.683 = 68.3\%$
- $L = 2$: $p = 0.955 = 95.5\%$
- $L = 3$: $p = 0.997 = 99.7\%$
- $L = 1.96$: $p = 0.950 = 95\%$
- $L = 2.33$: $p = 0.980 = 98\%$
- $L = 2.576$: $p = 0.990 = 99\%$

5. NISNID (normal is standard normal in disguise) fact.

If X is $N(\mu, \sigma)$, then $Z = \dfrac{X - \mu}{\sigma}$ is $N(0,1)$.

6. Confidence intervals.

$$\left(\overline{x} - L\frac{s}{\sqrt{n}}, \overline{x} + L\frac{s}{\sqrt{n}}\right).$$

Here $L = 1.96$ for a 95% confidence interval; $L = 2.33$ for a 98% confidence interval; $L = 2.576$ for a 99% confidence interval.

7. The "test statistic" or "z-score" for hypothesis testing of a population mean.

$$z = \frac{\overline{x} - \mu_0}{(s/\sqrt{n})}.$$

If $|z|$ is larger than or equal to 1.96/2.33/2.576, then reject H_0 and accept H_A at the 95%/98%/ 99% level respectively.

PART V. The Binomial distribution.

1. Mean, variance and standard deviation of (a single trial of) a binomial experiment.

Suppose X is a binomial random variable, meaning $X = 1$ if a certain event happens and $X = 0$ if not. Suppose the probability of that event happening (that is, the probability of a "success") is p. Then

$$E(X) = p, \qquad \operatorname{var} X = p(1-p), \qquad \operatorname{std} X = \sqrt{p(1-p)}.$$

2. Probability mass function, mean, variance and standard deviation of a repeated binomial experiment. Suppose a binomial experiment – that is, one with only two possible outcomes, a "success" and a "failure" – has $P(\text{success}) = p$. Suppose this experiment is repeated n times. (Assume all n trials are independent of each other.) Let X denote the number of successes in the n trials. Then

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad (0 \le k \le n),$$

$$E(X) = np, \qquad \operatorname{var} X = np(1-p), \qquad \operatorname{std} X = \sqrt{np(1-p)}.$$

3. The normal approximation to the binomial distribution.

Let X denote the number of successes in n trials of a binomial experiment, with $P(\text{success}) = p$. Assume $np \ge 10$ and $n(1-p) \ge 10$. Then for any numbers j and k between 0 and n inclusive:

(i) $P(X \leq j) \approx P(Y < j + 0.5)$.

(ii) $P(X \geq k) \approx P(Y > k - 0.5)$.

(iii) $P(k \leq X \leq j) \approx P(k - 0.5 < Y < j + 0.5)$.

PART VI. The Poisson distribution.

1. **Probability mass function.** Suppose a certain event happens, on average, λ times in each interval of a given extent. Let X denote the actual number of times it happens in such an interval. Then, for $k = 0, 1, 2, \ldots$,

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

and we say that X has a Poisson distribution with parameter λ.

2. **Mean, variance, standard deviation.** If X has a Poisson distribution with parameter λ, then

$$E(X) = \text{var}(X) = \lambda, \qquad \text{std}(X) = \lambda.$$