Final Review

1. Set operations: Let E, F be two events in S (Draw the Venn Diagram).

   (a) Union $E \cup F = \{x : x \in E \text{ or } x \in F\}$.

   (b) Intersection $EF = E \cap F = \{x : x \in E \text{ and } x \in F\}$.

   (c) Complement $E^c = \{x : x \notin E\}$.

   (d) Difference $E - F = E \cap F^c = \{x : x \in E \text{ and } x \notin F\}$.

   Let $E_1, E_2, \cdots$, be a sequence of events in S. We define

$$\bigcup_{n=1}^{\infty} E_n = \{x : \exists n \geq 1, x \in E_n\}$$

   and

$$\bigcap_{n=1}^{\infty} E_n = \{x : \forall n \geq 1, x \in E_n\}.$$

   Theorem 1 (DeMorgan's Law) *For each positive integer $n \geq 1$,*

$$\Big(\bigcup_{i=1}^{n} E_i\Big)^c = \bigcap_{i=1}^{n} E_i^c$$

$$\Big(\bigcap_{i=1}^{n} E_i\Big)^c = \bigcup_{i=1}^{n} E_i^c$$

2. Some propositions for the probability

   (a) $P(E) = 1 - P(E^c)$.

   (b) If $E \subset F$, then $P(E) \leq P(F)$.

   (c) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

   (d) Inclusion–exclusion identity:

$$\begin{aligned} P\Big(\bigcup_{i=1}^{n} E_i\Big) &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) \\ &= \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \cdots \\ &\quad + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) + \cdots \\ &\quad + (-1)^{n+1} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}). \end{aligned}$$

3. For two events E, F such that $P(E) > 0$, the conditional probability that F occurs given that E has occurred is denoted by

$$P(F|E) = \frac{P(F \cap E)}{P(E)}.$$

4. **Theorem 2 (The multiplication rule)** *For events E, F, we have*

$$P(E \cap F) = P(E) \cdot P(F|E).$$

More generally,

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) \times P(E_2|E_1) \times P(E_3|E_2 \cap E_1) \cdots P(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}).$$

5. For two events E, F, we say E and F are independent if

$$P(E \cap F) = P(E) \cdot P(F).$$

Otherwise, they are dependent. More generally, for a sequence of events $E_1, E_2, \cdots$, they are independent if and only if for any $i_1 < i_2 < \cdots < i_k$ with $k \geq 1$,

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = P(E_{i_1}) \times P(E_{i_2}) \times \cdots P(E_{i_k}).$$

6. **Total Probability Formula**: Let $A_1, \cdots, A_n$ be mutually exclusive so that $S = A_1 \cup \cdots \cup A_n$. Then for any event B,

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

Bayes' Theorem: Let $A_1, \cdots, A_n$ be mutually exclusive so that $S = A_1 \cup \cdots \cup A_n$. Then for any event B,

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}.$$

7. For a discrete random variable X, we define the probability mass function (abbreviated as pmf) $p(m)$ of X by

$$p(m) = P(X = m).$$

8. A nonnegative function $f : (-\infty, \infty) \to [0, \infty)$ is called a probability density function (pdf) if

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Then a random variable X is called a continuous random variable if there is a probability density function f such that for all $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

9. For any function g, we have

$$E[g(X)] = \int g(x) f_X(x) dx \quad \text{or} \quad E[g(X)] = \sum_{k \in \mathbb{Z}} g(k) P(X = k).$$

10. Properties of expectations: Let X, Y be two random variables.

    (a) If $X \geq 0$, then $EX \geq 0$;
    (b) If $c \in \mathbb{R}$, then $E[cX] = cEX$;
    (c) $E[X + Y] = EX + EY$.

11. For a random variable X, the cumulative distribution function (cdf) of X is defined by

$$F(b) = P(X \leq b), \quad \forall b \in \mathbb{R}.$$

If X is continuous, then

$$F(b) = \int_{-\infty}^{b} f(x) dx.$$

If X is discrete, then

$$F(b) = P(X \leq b) = \sum_{m=-\infty}^{[b]} P(X = m),$$

where $[b]$ is the largest integer less than or equal to b. It's an nondecreasing function having jumps at each m with $p(m) > 0$.

12. Steps for finding the probability density function of a random variable:

Step 1. Find the cdf of the random variable;
Step 2. Differentiate to find the density;
Step 3. Specify in what region the result holds.

13. The joint pdf $f(x, y)$ satisfies that

$$1 = \int \int f(x, y) dx dy.$$

14. The marginal density functions of X and Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Two continuous random variables X and Y are independent if and only if

$$f(x, y) = f_X(x) f_Y(y).$$

15. For any jointly pdf $f(x,y)$, we have for any function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$,

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dx dy$$

16. The Bivariate Normal Distribution. We say that the random variables X, Y have a bivariate normal distribution if, for constants μ_x, μ_y, $\sigma_x > 0$, $\sigma_y > 0$, $-1 < \rho < 1$, their joint density function is given, for all $-\infty < x, y < \infty$, by

$$\begin{aligned} f(x,y) =& \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \\ & \exp\Big\{ -\frac{1}{2(1-\rho^2)} \Big[\Big(\frac{x-\mu_x}{\sigma_x}\Big)^2 + \Big(\frac{y-\mu_y}{\sigma_y}\Big)^2 - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \Big] \Big\}. \end{aligned}$$

We say $(X,Y) \sim \mathcal{N}(\mu_x, \mu_y; \sigma_x^2, \sigma_y^2; \rho)$.

17. Consider two jointly continuous random variables (X,Y) with joint density function f_{XY}. Let another pair of random variables (U,V) be defined as functions of (X,Y): $U = g(X,Y)$, $V = h(X,Y)$. The goal is to find the joint density function f_{UV} of (U,V) through a multivariate change of variable.

Let K be a region of the xy plane so that $f_{XY}(x,y) = 0$ outside K. This implies $P((X,Y) \in K) = 1$. Let $G(x,y) = (g(x,y), h(x,y))$ be a bijective function that maps K onto a region L on the plane. Then $P((U,V) \in L) = 1$. Solve $u = g(x,y)$ and $v = h(x,y)$ to get

$$x = q(u,v) \text{ and } y = r(u,v).$$

Let

$$J(u,v) = \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} = \begin{vmatrix} \frac{\partial q(u,v)}{\partial u} & \frac{\partial q(u,v)}{\partial v} \\ \frac{\partial r(u,v)}{\partial u} & \frac{\partial r(u,v)}{\partial v} \end{vmatrix}.$$

Theorem 3 *The joint density function of $(U,V) = (g(X,Y), h(X,Y))$ is given by*

$$f_{UV}(u,v) = f_{XY}(q(u,v), r(u,v)) |J(u,v)|$$

for $(u,v) \in L$.

18. If a discrete random variable X has a pmf $p(m)$, then the variance of X is

$$Var(X) = E((X-\mu)^2) = \sum_{-\infty}^{\infty} (m-\mu)^2 p(m),$$

where $\mu = EX$ is the expectation of X.

Similarly, if a continuous random variable X has a pdf $f(x)$, then the variance of X is

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

19. Properties of variance: Let $E(X) = \mu$ and $Var(X) < \infty$. Let $X_1, X_2, \cdots, X_n$ be a sequence of random variables with $Var(X_i) < \infty$ for all i.

    (a)
$$Var\Big(a + bX\Big) = b^2 Var(X).$$

    (b)
$$Var(X) = E((X-\mu)^2) < E((X-c)^2), \quad \forall c \neq \mu.$$

    (c) If $Var(X) = 0$, then $X = \mu$ a.s.

    (d)
$$Var(\sum_{j=1}^{n} X_j) = \sum_{i=1}^{n}\sum_{j=1}^{n}[E(X_iX_j) - \mu_i\mu_j].$$

    (e) If $X_1, \cdots, X_n$ are independent, then
$$Var(\sum_{j=1}^{n} X_j) = \sum_{j=1}^{n} Var(X_j).$$

20. The covariance between X and Y, denoted by $Cov(X, Y)$, is defined by
$$\sigma_{XY} = Cov(X, Y) = E[(X - EX)(Y - EY)].$$

21. Define
$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$
to be the correlation of X and Y.

22. Properties of covariance:

    (a) $Cov(X, Y) = E(XY) - (EX)(EY)$.

    (b) $Cov(X, X) = Var(X)$.

    (c) $Cov(aX, Y) = aCov(X, Y)$.

    (d) $Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$. In particular,
$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n}\sum_{i=1}^{n} Cov(X_i, Y_j) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{1\le i<j\le n} Cov(X_i, X_j).$$

23. If $\rho = 0$ or $Cov(X, Y) = 0$, we say X and Y are uncorrelated. If X and Y are independent, then X and Y are uncorrelated

24. Markov's inequality: For any random variable X, we have

$$P(|X| > t) \leq \frac{E|X|}{t}, \quad \forall t > 0.$$

Proof by using $1_{|X|>t} \leq \frac{|X|}{t}$.

25. Chebyshev's inequality: For any random variable X with mean μ and variance σ^2, we have

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}, \quad \forall t > 0.$$

26. Four convergence modes:

(1) a.s. convergence:

$$P(\lim_{n\to\infty} X_n = X) = 1.$$

(2) converge in probability: For any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

(3) converge in distribution: For any $x \in \mathbb{R}$ such that x is a continuity point of F_X,

$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x).$$

(4) L^p convergence for $p > 0$:

$$\lim_{n\to\infty} E(|X_n - X|^p) = 0.$$

27. a.s. convergence implies convergence in probability: by definition.

28. L^p convergence implies convergence in probability: Markov's inequality.

29. Convergence in probability implies convergence in distribution. First proof: Use equivalent definition. Second proof, use cdf.

30. If X_n converges in distribution to a constant c, then X_n converges in probability to c. Use equivalent definition for convergence in probability to c from HW 13.4.

31. Convergence in distribution does not imply convergence in probability:

Counterexample: $X_n = Z$ and $X = -Z$, then X_n converges in distribution to X but X_n does not converge a.s nor in probability.

32. Convergence in probability does not imply a.s. convergence:

Counterexample: $X_n(\omega) = 1(\omega \in B_n)$ where B_n is the n-th set of $A_{jk} = [(j-1)/k, j/k]$ with $1 \leq j \leq k$, then X_n converges in probability to X but X_n does not converge a.s.

33. L^p convergence does not imply a.s. convergence:

    Counterexample: same as convergence in probability does not imply a.s. convergence.

34. a.s. convergence does not imply L^p convergence:

    Counterexample: Let $X_n(\omega) = n$ for $0 < \omega < 1/n$.

35. Weak Law of Large Numbers: Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having finite mean μ and variance σ^2. Then

$$\lim_{n\to\infty} P\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - \mu\Big| > \varepsilon\Big) = 0.$$

    That is, $\frac{X_1+\cdots+X_n}{n}$ converges in probability to μ as $n \to \infty$. Proof is by Chebyshev's inequality.

36. Strong Law of Large Numbers: Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having finite mean μ. Then

$$P\Big(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\Big) = 1.$$

    That is, $\frac{X_1+\cdots+X_n}{n}$ converges a.s. to μ as $n \to \infty$.

37. Central Limit Theorem (CLT): Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having finite mean μ and variance σ^2. Then

$$\lim_{n\to\infty} P\Big(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \le x\Big) = P(Z \le x), \quad \forall x \in \mathbb{R},$$

    where $Z \sim \mathcal{N}(0,1)$. That is, $\frac{X_1+\cdots+X_n-n\mu}{\sqrt{n\sigma^2}}$ converges in distribution to standard normal as $n \to \infty$.