# ST2334 Finals cheatsheet

by Khoo Jing Hong, Derrick

## Chapter 1: Probability

**MORE EVENT OPERATIONS**

(a) $A \cap A' = \emptyset$ (b) $A \cap \emptyset = \emptyset$

(c) $A \cup A' = S$ (d) $(A')' = A$

(e) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ Distributive law

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

(g) $A \cup B = A \cup (B \cap A')$

(h) $A = (A \cap B) \cup (A \cap B')$

Draw venn diagram for proof

**DE MORGAN'S LAW**

For any $n$ events $A_1, A_2, \ldots, A_n$,

(i) $(A_1 \cup A_2 \cup \ldots \cup A_n)' = A_1' \cap A_2' \cap \ldots \cap A_n'$.

A special case: $(A \cup B)' = A' \cap B'$.

(j) $(A_1 \cap A_2 \cap \ldots \cap A_n)' = A_1' \cup A_2' \cup \ldots \cup A_n'$.

A special case: $(A \cap B)' = A' \cup B'$.

**AXIOMS OF PROBABILITY**

Probability, denoted by $P(\cdot)$, is a **function** on the collection of events of the sample space $S$, satisfying:

Axiom 1. For any event $A$,

$$0 \le P(A) \le 1.$$

Axiom 2. For the sample space,

$$P(S) = 1.$$

Axiom 3. For any two mutually exclusive events $A$ and $B$, that is, $A \cap B = \emptyset$,

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

**PROPOSITION 5**

*For any two events A and B,*

$$P(A) = P(A \cap B) + P(A \cap B').$$

### Conditional probability:

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

### Bayes Theorem:

Used to flip a conditional probability from P(A|B) to P(B|A)

**SPECIAL CASE: BAYES' THEOREM**

Let us consider a special case of Bayes' Theorem when $n = 2$.

$\{A, A'\}$ becomes a partition of $S$, and we have

$$P(A|B) = \frac{P(A)P(B|A)}{\underbrace{P(A)P(B|A) + P(A')P(B|A')}_{= P(B)}}.$$

law of total probability

### Law of total probability:

**SPECIAL CASE: LAW OF TOTAL PROBABILITY**

For any events $A$ and $B$, we have

$$P(B) = P(A)P(B|A) + P(A')P(B|A').$$

$P(A \cap B) + P(A' \cap B)$

## Chapter 2: Random variables

### Discrete Random Variables(DRV)

Properties of PMF:

**PROPERTIES OF THE PROBABILITY MASS FUNCTION**

The probability mass function $f(x)$ of a discrete random variable **must** satisfy:

(1) $f(x_i) \ge 0$ for all $x_i \in R_X$;

(2) $f(x) = 0$ for all $x \notin R_X$;

(3) $\sum_{i=1}^{\infty} f(x_i) = 1$, or $\sum_{x_i \in R_X} f(x_i) = 1$.

For any set $B \subset \mathbb{R}$, we have

$$P(X \in B) = \sum_{x_i \in B \cap R_X} f(x_i).$$

### Continuous Random Variables(CRV)

Properties of PDF:

**DEFINITION 4 (PROBABILITY DENSITY FUNCTION)**

*The **probability density function** of a continuous random variable X, denoted by f(x), is a function that satisfies:*

(1) $f(x) \ge 0$ *for all* $x \in R_X$; *and* $f(x) = 0$ *for* $x \notin R_X$;

(2) $\int_{R_X} f(x)\,dx = 1$;

(3) *For any a and b such that* $a \le b$,

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

Properties of CDF:

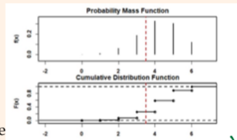
- Non-decreasing in general

Case (discrete random variable)

**CDF: DISCRETE RANDOM VARIABLE**

If $X$ is a discrete random variable, we have

$$F(x) = \sum_{t \in R_X; t \le x} f(t) = \sum_{t \in R_X; t \le x} P(X = t)$$





The cumulative distribution function of a discrete random variable is a step function.

For any two numbers $a < b$, we have does not incl a (open interval)

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

where "$a-$" represents the "largest value in $R_X$ that is smaller than $a$". Mathematically,

$$F(a-) = \lim_{x \uparrow a} F(x).$$

Case (Continuous Random Variable)

**CDF: CONTINUOUS RANDOM VARIABLE**

If $X$ is a continuous random variable,

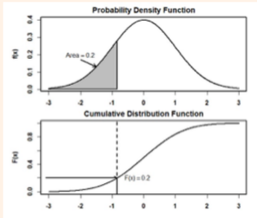
$$F(x) = \int_{-\infty}^{x} f(t)\,dt,$$

and

$$f(x) = \frac{dF(x)}{dx}.$$

Further

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





### Expectation and Variance

DRV: $E(X) = \sum x_i f(x_i)$.

CRV: $\mu_X = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{x \in R_X} x f(x)\,dx$.

Variance: $\sigma_X^2 = V(X) = E(X - \mu_X)^2$. OR

$V(X) = E(X^2) - (E(X))^2$

Properties:

$V(aX + b) = a^2 V(X)$

**Graphing Calculator**

X ~ N(65,25),

P(47.5 < X ≤ 80) -> normalcdf(47.5,80,65,5)

Z ~ N(0,1) (p-value)

P(Z > 0.5) -> normalcdf(0.5, E99, 0, 1)

P(Z < 0.5) -> normalcdf(-E99,0.5,0,1)

Φ(c) = x -> -1 * invNorm(x, 0, 1, RIGHT)

## Chapter 3: Joint Distributions

- Discrete variant

Properties:

**PROPERTIES OF THE DISCRETE JOINT PROBABILITY FUNCTION**

The joint probability mass function has the following properties:

(1) $f_{X,Y}(x,y) \ge 0$ for any $(x,y) \in R_{X,Y}$. (Non-negative)

(2) $f_{X,Y}(x,y) = 0$ for any $(x,y) \notin R_{X,Y}$.

(3) $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(X = x_i, Y = y_j) = 1$.

Equivalently, $\sum\sum_{(x,y) \in R_{X,Y}} f(x,y) = 1$.

(4) Let $A$ be any subset of $R_{X,Y}$, then

$$P((X,Y) \in A) = \sum\sum_{(x,y) \in A} f_{X,Y}(x,y).$$

Defn for joint probability function

- Continuous variant

Properties:

**PROPERTIES OF THE CONTINUOUS JOINT PROBABILITY FUNCTION**

The joint probability density function has the following properties:

(1) $f_{X,Y}(x,y) \ge 0$, for any $(x,y) \in R_{X,Y}$.

(2) $f_{X,Y}(x,y) = 0$, for any $(x,y) \notin R_{X,Y}$.

(3) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$.

Equivalently, $\iint_{(x,y) \in R_{X,Y}} f_{X,Y}(x,y)\,dx\,dy = 1$.

defn for joint probability density function

- Marginal probability distribution

$f_X(x) = \sum_y f_{X,Y}(x,y)$. $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$.

Key idea: Ignoring 1 var and do sum/integr on 1 var

- Conditional distribution

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

**Key ideas:**

1. **X** and Y are independent if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y).$$

2. Covariance:

$$\text{cov}(X,Y) = E[(X-E(X))(Y-E(Y))].$$

$$\text{cov}(X,Y) = E(XY) - E(X)E(Y).$$

If X,Y are independent, then cov(X,Y) = 0. Also,

$$\text{cov}(aX+b, cY+d) = ac \cdot \text{cov}(X,Y).$$
$$\text{cov}(X,Y) = \text{cov}(Y,X);$$
$$\text{cov}(X+b,Y) = \text{cov}(X,Y);$$
$$\text{cov}(aX,Y) = a\,\text{cov}(X,Y).$$

Regarding variance:

$$V(aX+bY) = a^2V(X) + b^2V(Y) + 2ab \cdot \text{cov}(X,Y).$$
$$V(aX) = a^2V(X);$$
$$V(X+Y) = V(X) + V(Y) + 2\text{cov}(X,Y).$$

## Chapter 4: Probability Distributions

DRV:

**Bernoulli Trial**: 2 possible outcomes -> Bernoulli random variable

**Bernoulli Process**: Sequence of independent and identically distributed Bernoulli random vars

**Negative Binom**: Until the **k**th success

**Geometric**: Until the **1**st success

**Poisson**: Fixed period of time or fixed region(eg certain amt of time)

- From 1 day to 2 day -> multiply to lambda
- **Approx to Binom: when n -> inf, p -> 0** and if n >= 20 and p <= 0.05 **OR** n >= 100 and np <= 10, where lambda = np

CRV:

**Exponential**: First success in continuous time

- Memoryless property

**Normal**:

- Standardised: $Z = \frac{X-\mu}{\sigma}.$
  - For any $z$, $\Phi(z) = P(Z \leq z) = P(Z \geq -z) = 1 - \Phi(-z)$;
- **Normal approx to Binom:**
  - np > 5, n(1-p) > 5
  - **when n -> inf, p remain constant**
  - **MUST APPLY continuity correction**

### Continuity Correction

We apply continuity correction when approximating the binomial using the normal.

$$P(X=k) \approx P(k - 1/2 < X < k + 1/2)$$

$$P(a \leq X \leq b) \approx P(a - 1/2 < X < b + 1/2)$$
$$P(a < X \leq b) \approx P(a + 1/2 < X < b + 1/2)$$
$$P(a \leq X < b) \approx P(a - 1/2 < X < b - 1/2)$$
$$P(a < X < b) \approx P(a + 1/2 < X < b - 1/2)$$

$$P(X \leq c) = P(0 \leq X \leq c) \approx P(-1/2 < X < c + 1/2)$$

$$P(X > c) = P(c < X \leq n) \approx P(c + 1/2 < X < n + 1/2)$$

* capture equality watch out for the rules

## Chapter 5: Sampling and sampling dist

1. For random samples of size n taken from population with mean μ and variance σ^2, sample has mean = μ, var = σ^2/n
2. **Central Limit Theorem** - Sampling distribution of sample mean is approximately normal if n is large
3. **Law of Large Numbers** - Sample mean converges to population mean when n increases, because σ^2/n decreases.
4. **Chi-Square distribution** - Used for the square of a random variable(eg. sample variance)

### Properties of $\chi^2$ Distributions

1. If $Y \sim \chi^2(n)$, then $E(Y) = n$ and $V(Y) = 2n$.
2. For large $n$, $\chi^2(n)$ is approximately $N(n, 2n)$.
3. If $Y_1$ and $Y_2$ are independent $\chi^2$ random variables with $m$ and $n$ degrees of freedom respectively, then $Y_1 + Y_2$ is a $\chi^2$ random variable with $m+n$ degrees of freedom.
4. The $\chi^2$ distribution is a family of curves, each determined by the degrees of freedom $n$. All the density functions have a long right tail.

Special case: Sample variance distribution

**Theorem 12**

*If $S^2$ is the variance of a random sample of size $n$ taken from a normal population having the variance $\sigma^2$, then the random variable*

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2}$$

*has a $\chi^2$ distribution with $n-1$ degrees of freedom.*

5. t-Distribution - Used when population is normal, sample size n < 30

If $T \sim t(n)$, then $E(T) = 0$ and $V(T) = n/(n-2)$ for $n > 2$.

$$t = \frac{X-\mu}{S/\sqrt{n}}$$

6. F-distribution - Used when looking at the distribution of sample statistic/sample statistic and sample statistic follows a **chi-square** distribution.

$$F = \frac{U/m}{V/n}$$

## Chapter 6: Estimation

1. Maximum error of estimate - with probability (1 - alpha), error E = $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$
2. Minimum sample size - so that with probability (1 - alpha), error is at most E.

length of confidence interval $= 2\left(z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{E_0}\right)^2.$$

$n \geq 30$ : large

### Different Cases

| | Population | $\sigma$ | $n$ | Statistic | $E$ | $n$ for desired $E_0$ and $\alpha$ |
|---|---|---|---|---|---|---|
| I | Normal | known | any | $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ | $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ | $\left(\frac{z_{\alpha/2} \cdot \sigma}{E_0}\right)^2$ |
| II | any | known | large | $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ | $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ | $\left(\frac{z_{\alpha/2} \cdot \sigma}{E_0}\right)^2$ |
| III | Normal | unknown | small | $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ | $t_{n-1;\alpha/2} \cdot \frac{s}{\sqrt{n}}$ | $\left(\frac{t_{n-1;\alpha/2} \cdot s}{E_0}\right)^2$ |
| IV | any | unknown (estimate sigma) | large | $Z = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ | $z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ | $\left(\frac{z_{\alpha/2} \cdot s}{E_0}\right)^2$ |

3. Equal variance - s1/s2 ~= [1/2, 2]
4. Paired data, D = Difference

$$T = \frac{\overline{D}-\mu_D}{S_D/\sqrt{n}}, \quad \text{where} \quad \overline{D} = \frac{\sum_{i=1}^{n} D_i}{n}, \quad S_D^2 = \frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1}.$$

### Confidence Intervals: Paired Data

For paired data, if $n$ is small ($n < 30$) and the population is normally distributed, a $(1-\alpha)100\%$ confidence interval for $\mu_D$ is

$$\overline{d} \pm t_{n-1;\alpha/2} \cdot \frac{s_D}{\sqrt{n}}.$$

If $n$ is large ($n \geq 30$), a $(1-\alpha)100\%$ confidence interval for $\mu_D$ is

$$\overline{d} \pm z_{\alpha/2} \cdot \frac{s_D}{\sqrt{n}}.$$

## Chapter 7: Hypothesis Test

### How to do a hypothesis test

There are five main steps to hypothesis testing.

Step 1: Set your competing hypotheses: null and alternative.

Step 2: Set the level of significance.

Step 3: Identify the test statistic, its distribution and the rejection criteria.

Step 4: Compute the observed test statistic value, based on your data.

Step 5: Conclusion.

- If $p$-value $< \alpha$, reject $H_0$; else
- If $p$-value $\geq \alpha$, do not reject $H_0$.

| | Do not reject $H_0$ | Reject $H_0$ |
|---|---|---|
| $H_0$ is true | Correct Decision | **Type I error** |
| $H_0$ is false | **Type II error** | Correct Decision |