Chapter 1

Recap Sheet

• Probability Basics

Sample space (Ω) : set of all possible outcomes of a random experiment

Ex : Throw a die $\rightarrow \Omega = \{1,2,3,4,5,6\}$

Event : subset of Ω

Ex : $A = \{2,4,6\}$ = "getting an even number"

Probability function : function assigning a number between 0 and 1 to each event, such that :

- $P(\Omega) = 1$
- $P(A \cup B) = P(A) + P(B)$ for disjoint events
  can't happen at the same time

Ex : $P(A) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$

Union probability : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional probability : $P(A|B) = \frac{P(A \cap B)}{P(B)}$, if $P(B) > 0$

↳ prob of A knowing that B has already occured

Law of total probability : if (B_i) is a partition of Ω (disjoint cases covering the whole space), then :

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$$

Ex : A school has 70% of students in CS and 30% in electronics.
- 90% of students in CS do sports.
- 40% of students in electronics do sports.

If we choose student randomly, what is the probability that he does sports ?

$$P(S) = P(S|CS)P(CS) + P(S|E)P(E) = 0.90 \times 0.70 + 0.40 \times 0.30 = 0.26$$

Baye's theorem : allows us to reverse conditional probabilities

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}$$

Ex : We do a medical test.
- The disease touches 1% of the population.
- The test is positive 99% of the time when sick.
- The test is positive 5% of the time when healthy (false pos.)

If the patient's test is positive, what is the probability of him really being sick ?

$$P(S|P) = \frac{P(P|S)P(S)}{P(P|S)P(S) + P(P|\bar{S})P(\bar{S})} = \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.05 \times 0.99} \approx 0.166$$

Independant events : $P(A \cap B) = P(A)P(B)$
↳ learning that B happened doesn't change the chance of A

• Discrete random variables

Random variable : rule that assigns a number to each outcome :
$X : \Omega \to \mathbb{R}$.

Discrete : if X only takes finitely (contably) many values

- Support (possible values) : $Supp(X) = \{x : P(X = x) > 0\}$
- PMF (probability mass function) : $p_X(x) = P(X = x)$
  ↳ for discrete X : $p_X(x) \geq 0$; $\sum_x p_X(x) = 1$
- CDF (cumulative) : $F_X(a) = P(X \leq a) = \sum_{x \leq a} p_X(x)$

Ex : We flip a coin 3 times. Let X = "number of heads".
Then $Supp(X) = \{0, 1, 2, 3\}$. Count outcomes :

$P(X = 0) = \frac{1}{8}$ (TTT) $\quad P(X = 1) = \frac{3}{8}$ (HTT, THT, TTH)

$P(X = 2) = \frac{3}{8}$ (HHT, HTH, THH) $\quad P(X = 3) = \frac{1}{8}$ (HHH)

The CDF at this support points :

$F_X(0) = \frac{1}{8} \quad F_X(1) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2} \quad F_X(2) = \frac{1}{2} + \frac{3}{8} = \frac{7}{8} \quad F_X(3) = 1$

Expected value (mean) of a **discrete random variable** X with possible values $x_1, x_2 \ldots$ and probabilities $p(x_i) = P(X = x_i)$:

$$E(X) = \sum_i x_i \, p(x_i)$$

Properties :
- $E(X + a) = E(X) + a$
- $E(aX) = a\,E(X)$
- $E(X + Y) = E(X) + E(Y)$

Ex : Using the previous example values :

$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5$

Expectation of a function $g(X)$ for a RV X with expected values x_i and pmf $p(x_i) = P(X = x_i)$:

$$E[g(X)] = \sum_i g(x_i)\, p(x_i)$$

Ex : Using the previous example values.
We'll take $g(x) = x^2$. Then :

$E(X) = 0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8} = 3$

Variance : $Var(X) = E(X^2) - (E(X))^2$ **Standard deviation** : $\sigma = \sqrt{Var(X)}$
↳ same units as X

Properties : $Var(X + c) = Var(X)$; $Var(aX) = a^2\, Var(X)$
↳ constant

Interpretation : variance measures spread around the mean
↳ small variance = values cluster near mean

Bienaymé - Chebyshev inequality : gives a universal bound on how often X strays (dévier) far from its mean
↳ useful when we don't know the exact distribution

For any random variable X with mean $\mu = E(X)$ and variance $\sigma^2 = Var(X)$,

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \text{ for any } \varepsilon > 0$$

Ex : We have the following data for an exam scores : $\mu = 70$ and $\sigma = 10$.
What can we guarantee about scores within 20% of the mean?

→ We choose $\varepsilon = 20$: $P(|X-70| < 20) \geq 1 - \frac{100}{20^2} = 1 - \frac{100}{400} = 0.75$

So no matter the shape of the score distribution, at least 75 % of students score between 50 and 90.

Two discrete random variables : the joint probability gives it for pairs : $P_{X,Y}(x,y) = P(X=x, Y=y)$; $\sum_x \sum_y P_{X,Y}(x,y) = 1$

Idea : a grid whose cell (x,y) is the probability that both happen together.

Marginals : $P_X(x) = \sum_y P_{X,Y}(x,y)$; $P_Y(y) = \sum_x P_{X,Y}(x,y)$

Ex : Urn with 2 red, 2 white, 1 black balls. We draw 2 without replacement.
X = # of red drawn $\in \{0,1,2\}$
Y = # of white drawn $\in \{0,1,2\}$

We have the following joint table :

X \ Y	0	1	2	$P_X(x)$
0	0	1/5	1/10	3/10
1	1/5	2/5	0	3/5
2	1/10	0	0	1/10
$P_Y(y)$	3/10	3/5	1/10	1

Conditionals : from the joint table, we pick the row (if conditioning on X) or the column (if conditioning on Y) and divide each cell by its marginal sum

Ex : $P(X=1 \mid Y=0) = \frac{1/5}{3/10} = \frac{2}{3}$

$P(Y=0 \mid X=1) = \frac{1/5}{3/5} = \frac{1}{3}$

Independance . $P(X=x_i, Y=y_i) = P(X=x_i)P(Y=y_i)$

Ex : Here we have $P(X=1, Y=1) = 2/5$ but $P(X=1)P(Y=1) = \frac{9}{25} \neq \frac{2}{5}$ so the random variables are not independant.

Covariance : for discrete X, Y with joint pmf $p_{i,j} = P(X = x_i, Y = y_j)$

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

Interpretation :
- $\mathrm{Cov}(X,Y) > 0$: when X is above its mean, Y tends to be above its mean too (co-movement)
- $\mathrm{Cov}(X,Y) < 0$: when X is above its mean, Y tends to be below its mean (anti-movement)

Properties :
- $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
- $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$
- If X, Y independant $\Rightarrow \mathrm{Cov}(X,Y) = 0$

Correlation : $\rho(X,Y) = \dfrac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$

Always $-1 \leq \rho \leq 1$

Interpretation : $|\rho|$ close to 1 = strong linear relationship.
$|\rho|$ close to 0 = weak / no linear relationship.
$\rho > 0$ = Y tends to increase when X increases
$\rho < 0$ = opposite tendency

Ex : $E(X) = 0 \times \frac{3}{10} + 1 \times \frac{3}{5} + 2 \times \frac{1}{10} = \frac{4}{5}$, $E(Y) = \frac{4}{5}$

$E(X^2) = 0^2 \times \frac{3}{10} + 1^2 \times \frac{3}{5} + 2^2 \times \frac{1}{10} = 1 \Rightarrow \mathrm{Var}(X) = 1 - \left(\frac{4}{5}\right)^2 = \frac{9}{25}$

Same for Y : $\mathrm{Var}(Y) = \frac{9}{25}$

$E(XY) = \sum_{x,y} xy\,P(x,y)$

$= (1.1) \times \frac{2}{5} + (1.0) \times \frac{1}{5} + (0.1) \times \frac{1}{5} + (2.0) \times \frac{1}{10} + (0.2) \times \frac{1}{10} = \frac{2}{5}$

$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{2}{5} - \frac{4}{5} \cdot \frac{4}{5} = \frac{-6}{25} = -0.24$

Given only 2 draws, "more red" leaves fewer slots for white, so they tend to move in opposite directions.

$\rho = \dfrac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \dfrac{-6/25}{\sqrt{\frac{9}{25}} \times \sqrt{\frac{9}{25}}} = -\frac{2}{3} \approx -0{,}66$

There is a moderate negative linear relationship.

• Usual distributions

Discrete uniform distribution : if every value is equally likely

$$X \sim U(n) \; : \; P(X = x_i) = \frac{1}{n} \text{ for } i = 1, \dots, n$$

If the values of x_i are their rank (exactly $\{1, 2, \dots, n\}$) then :

$$E(X) = \frac{n+1}{2} \qquad \mathrm{Var}(X) = \frac{n^2 - 1}{12}$$

Ex : We throw a fair die, the support is $S = \{1,2,3,4,5,6\}$, each with prob $\frac{1}{6}$. So $X \sim U(6)$ and :

$E(X) = \frac{6+1}{2} = \frac{7}{2} = 3.5$ $\quad$ $Var(X) = \frac{6^2-1}{12} = \frac{35}{12} \approx 2.92$

Bernoulli distribution : if we have an indicator random variable (event A happens or not) like

$$X = 1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \quad (\text{success}) \\ 0 & \text{if } \omega \notin A \quad (\text{failure}) \end{cases}$$

Then X only takes values $\{0,1\}$, with

$$P(X=1) = P(A) = p \,, \quad P(X=0) = 1 - P(A) = q \,, \quad p+q = 1$$

We say $X \sim B(p)$.

Mean : $E(X) = 1 \cdot P(X=1) + 0 \cdot P(X=0) = p$, also $E(X^2) = E(X) = p$

Variance : $Var(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p) = pq$

Standard deviation : $\sigma_X = \sqrt{pq}$

Ex : Let $A = \{\text{die} = 6\}$. Then $p = P(A) = \frac{1}{6}$ and $q = \frac{5}{6}$.

We have $E(X) = \frac{1}{6} \approx 0.167$ and $Var(X) = \frac{1}{6} \times \frac{5}{6} = \frac{5}{36} \approx 0.139$

Binomial distribution : we repeat the same Bernoulli trial n times.

- each trial has 2 outcomes : success (1) with prob p and failure (0) with prob $q = 1-p$
- trials are independant and identically distributed (same p each time)

We define X = number of successes in the n trials

Then X follows a Binomial distribution : $X \sim B(n,p)$ with support $\{0,1,\dots,n\}$. We have :

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

Mean : $E(X) = np$ $\quad$ Variance : $Var(X) = np(1-p)$

Std var : $\sigma = \sqrt{np(1-p)}$

Ex : 5 fair coin flips. Let $n=5$, $p = \frac{1}{2}$. Then $X \sim B(5, \frac{1}{2})$.

$P(X=k) = \binom{5}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{5-k} = \binom{5}{k} \cdot \left(\frac{1}{2}\right)^{k+5-k} = \binom{5}{k} \cdot 2^{-5}$

$E(X) = np = \frac{5}{2}$ $\quad$ $Var(X) = \frac{5}{2} \times \frac{1}{2} = \frac{5}{4}$

Poisson distribution : models the count of events occuring in a fixed interval of time / space when :
- events happen independantly of each other
- the average rate λ is constant in the interval
- events are "rare" relative to the size of the interval

A random variable X has a Poisson distribution with parameter $\lambda > 0$ if :

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

We write $X \sim \mathcal{P}(\lambda)$

Mean and variance : $E(X) = Var(X) = \lambda$

Ex : Let X = number of laptops sold in a day, with average $\lambda = 5$. Assume $X \sim \mathcal{P}(\lambda)$.

1. Exactly 5 sold : $P(X = 5) = \frac{e^{-5} 5^5}{5!} \approx 0.176$
2. At least 2 sold : $P(X \geq 2) = 1 - P(0) - P(1)$
$= 1 - e^{-5} - e^{-5} \cdot 5$
≈ 0.960

Also $E(X) = Var(X) = 5$ so $\sigma = \sqrt{5} \approx 2.236$

Binomial → Poisson approximation : when n is large and p is small with $\lambda = np$ fixed, we can approximate :

$$\text{If } X \sim \mathcal{B}(n, p), \text{ then } P(X = k) \approx \frac{e^{-\lambda} \lambda^k}{k!}$$

Geometric distribution : we repeat independant Bernoulli (p) trials until the first success occurs.

There are 2 standard ways to define the random variable :

1. X = number of trials up to and including the first success
Support : $\{1, 2, 3, \dots\}$. We write $X \sim G_1(p)$

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \dots$$

Idea : $k - 1$ failures, then a success

2. Y = number of failures before the first success
Support $\{0, 1, 2, \dots\}$. We write $Y \sim G_0(p)$

$$P(Y = y) = (1 - p)^y p, \quad y = 0, 1, 2, \dots$$

Relation : $X = Y + 1$

For $X \sim G_1(p)$:

$$E(X) = \frac{1}{p} \qquad Var(X) = \frac{1 - p}{p^2} \qquad \sigma_X = \frac{\sqrt{1 - p}}{p}$$

Negative Binomial distribution : we repeat independantly Bernoulli (p) trials until we have obtained r successes.

X = numbers of trials needed to see the r-th success
Support : $\{r, r+1, r+2, \dots\}$. We write $X \sim NB(r,p)$

$$P(X=k) = \binom{k-1}{r-1} p^r q^{k-r}$$

Mean and variance : $E(X) = \frac{r}{p}$ $\quad Var(X) = \frac{r(1-p)}{p^2}$