Chapter 2

Recap Sheet

• Continuous Random Variable

Idea : Instead of assigning probability to single points, we assign it to intervals (areas under f)

Definition . A random variable X is continuous with density f if there exists a non-negative function $f : \mathbb{R} \to \mathbb{R}_+$ such that, for any set $B \subset \mathbb{R}$,

$$P(X \in B) = \int_B f(x)\,dx$$

f is the probability density function of X.

Properties : $\int_{-\infty}^{+\infty} f(x)\,dx = 1$; $P(X = a) = \int_a^a f(x)\,dx = 0$

Core probability query : For any $a < b$,

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

Ex : We are given : $f(x) = \begin{cases} \frac{1}{6}x + k & \text{if } 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$

1. Find k. Since $f(x)$ is a density, we have $\int_{-\infty}^{+\infty} f(x) = 1$. But $f(x)$ is nonzero only on $[0;3]$ so :

$$\int_0^3 \left(\frac{1}{6}x + k\right)dx = 1 \iff \int_0^3 \frac{1}{6}x\,dx + \int_0^3 k\,dx = 1$$

$$\iff \frac{1}{12}\left[x^2\right]_0^3 + k\left[x\right]_0^3 = 1$$

$$\iff \frac{3}{4} + 3k = 1$$

$$\iff k = \frac{\left(1 - \frac{3}{4}\right)}{3} = \frac{1}{12}$$

2. Compute $P(1 \leq X \leq 2)$. By definition,

$$P(1 \leq X \leq 2) = \int_1^2 \left(\frac{1}{6}x + \frac{1}{12}\right)dx = \frac{1}{6}\int_1^2 x\,dx + \frac{1}{12}\int_1^2 dx$$

$$= \frac{1}{12}\left[x^2\right]_1^2 + \frac{1}{12}\left[x\right]_1^2$$

$$= \frac{1}{4} + \frac{1}{12} = \frac{1}{3}$$

• Distribution function

Definition : For a random variable X with density f, the distribution function is

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

Properties :

- Monotone : if $a < b$ then $F_X(a) \leq F_X(b)$
- Limits : $\lim_{x \to -\infty} F_X(x) = 0$; $\lim_{x \to +\infty} F_X(x) = 1$
- Link to pdf : where f is continuous, $F_X'(x) = f(x)$
- Interval probability : for $a < b$,

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

Ex : We use the same pdf as in the previous example. Let us compute $F_X(a) = P(X \leq a)$.

- If $a < 0$: $F_X(a) = 0$
- If $0 \leq a \leq 3$: $F_X(a) = \int_0^a \left(\frac{1}{6}x + \frac{1}{12}\right)dx = \frac{a^2 + a}{12}$
- If $a > 3$. $F_X(a) = 1$

So :

$$F_X(a) = \begin{cases} 0, & a < 0 \\ \frac{a^2+a}{12}, & 0 \leq a \leq 3 \\ 1, & a > 3 \end{cases}$$

- **Function of a CRD (Transformations)**

Idea : Given a continuous X with pdf f_X and a (nice) function h, we define $Y = h(X)$. We want the pdf f_Y.

Method :

- CDF method (always work) : $F_Y(y) = P(Y \leq y) = P(h(X) \leq y)$

$\rightarrow$ Differentiate . $f_Y(y) = \frac{d}{dy} F_Y(y)$

Ex : We have $f_X(x) = 2x \, \mathbb{1}_{[0,1]}(x)$, its CDF is :

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

(A) $Y = 3X + 1$ $(X \in [0,1] \Rightarrow T \in [1,4])$

$$F_Y(y) = P(3X + 1 \leq y) = P\left(X \leq \frac{y-1}{3}\right) = \begin{cases} 0, & y < 1 \\ \left(\frac{y-1}{3}\right)^2, & 1 \leq y \leq 4 \\ 1, & y > 4 \end{cases}$$

Hence $f_Y(y) = \frac{d}{dy}\left(\frac{y-1}{3}\right)^2 = \frac{2(y-1)}{9}$, $1 \leq y \leq 4$

(B) $z = X^2$ $(X \in [0,1] \Rightarrow z \in [0,1])$

$$F_z(z) = P(X^2 \leq z) = P(X \leq \sqrt{z}) = \begin{cases} 0, & z < 0 \\ z, & 0 \leq z \leq 1 \\ 1, & z > 1 \end{cases}$$

Hence $f_Z(z) = \frac{d}{dz}(z) = 1, \quad 0 < z < 1$

(c.) $T = e^X$ $(X \in [0,1] \Rightarrow T \in [1, e])$

$$F_T(t) = P(e^X \leq t) = P(X \leq \ln(t)) = \begin{cases} 0, & t < 1 \\ \ln^2(t), & 1 \leq t \leq e \\ 1, & t > e \end{cases}$$

Hence $f_T(t) = \frac{d}{dt}\left(\ln^2(t)\right) = \frac{2\ln(t)}{t}, \quad 1 \leq t \leq e$

- **Expected Value**

Definition : For a continuous X with pdf f,

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx$$

Transform theorem : For any (integrable) function g,

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x)\,dx$$

Linearity : $E(aX + b) = aE(X) + b$

Ex : Given $f_X(x) = 2x\,\mathbb{1}_{[0,1]}(x)$.

- $E(X) = \int_0^1 x(2x)\,dx = 2\int_0^1 x^2\,dx = \frac{2}{3}[x^3]_0^1 = \frac{2}{3}$
- $E(Y) = E(3X+1) = \int_0^1 (3x+1)(2x)\,dx = \int_0^1 (6x^2 + 2x)\,dx$
  $= 6\int_0^1 x^2\,dx + 2\int_0^1 x\,dx$
  $= 2[x^3]_0^1 + [x^2]_0^1$
  $= 3$

- $E(T) = E(e^X) = \int_0^1 e^x(2x)\,dx \rightarrow$ Integration by parts

$u = 2x \longrightarrow u' = 2$
$v' = e^x \qquad v = e^x$

so $E(T) = [2xe^x]_0^1 - 2\int_0^1 e^x\,dx$
$= 2e - 2[e^x]_0^1$
$= 2e - (2e - 2)$
$= 2$

- **Variance and standard deviation**

Definition : Let X be continuous with pdf f and finite mean $E(X)$. The variance of X is

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

where

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx \quad \text{and} \quad E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\,dx$$

The standard deviation is $\sigma_X = \sqrt{Var(X)}$

Ex : Given $f_X(x) = 2x\,\mathbb{1}_{[0,1]}(x)$, we already computed $E(X) = \frac{2}{3}$

Also $E(X^2) = \int_0^1 x^2(2x)\,dx = 2\int_0^1 x^3\,dx = \frac{1}{2}[x^4]_0^1 = \frac{1}{2}$

Hence $Var(X) = E(X^2) - (E(X))^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$

Properties :

- $Var(X) \geq 0$
- $Var(X+b) = Var(X)$
- $Var(aX) = a^2 Var(X)$
- If X, Y indpdt : $Var(X+Y) = Var(X) + Var(Y)$

• Uniform distribution $U(a,b)$

Definition : A random variable X is Uniform on $[a,b]$ (with $a<b$) if its pdf is

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Distribution function :

$$F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

Mean and variance :

$$E(X) = \frac{a+b}{2} \qquad Var(X) = \frac{(b-a)^2}{12}$$

Interval probability : for $a \leq u < v \leq b$,

$$P(u \leq X \leq v) = \int_u^v \frac{1}{b-a}\,dx = \frac{v-u}{b-a}$$

Ex . Let $X \sim U(2,5)$.

- $P(3 \leq X \leq 4) = \frac{4-3}{5-2} = \frac{1}{3}$
- $P(X < 3.5) = F_X(3.5) = \frac{3.5-2}{5-2} = 0.5$
- $E(X) = \frac{2+5}{2} = 3.5$
- $Var(X) = \frac{(5-2)^2}{12} = \frac{3}{4}$

• Exponential distribution $E(\lambda)$

Definition : A random variable X is Exponential with rate $\lambda > 0$, written $X \sim E(\lambda)$, if

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

CDF : $F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$

Mean and variance : $E(X) = \frac{1}{\lambda}$ $\qquad Var(X) = \frac{1}{\lambda^2}$

When to use it :

- The variable is time-to-event and cannot be negative (e.g. , minutes until next bus arrives)
- The event rate is roughly constant over time
- The past doesn't change the future : after waiting t, the distribution of the remaining wait is the same as at time 0
- Arrivals look like a Poisson process

Ex : Let $X \sim E(\lambda = 0.5$ per hour$)$. Then $E(X) = \frac{1}{\lambda} = 2$ hours, $Var(X) = 4$.

1. $P(X > 3) = e^{-\lambda \cdot 3} = e^{-1.5} \approx 0.2231$
2. $P(1 < X < 4) = P(X > 1) - P(X > 4) = e^{-0.5} - e^{-2} \approx 0.4712$

- **Normal distribution $N(\mu, \sigma^2)$**

Definition : A continuous random variable X is Normal with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ if

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

Mean and variance : $E(X) = \mu \qquad Var(X) = \sigma^2$

When to use it :

- The data/noise is approximately symmetric around a central value
- We are dealing with sums/averages of many small, independant effects

Standardization and the CDF : The Normal CDF has no elementary closed form. We use the standard normal $Z \sim \mathcal{N}(0,1)$ with CDF Φ. The link is :

$$Z = \frac{X-\mu}{\sigma} \Rightarrow P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Ex : Let $X \sim \mathcal{N}(3, 4)$ so $\sigma = 2$. Standardize with $Z = \frac{X-3}{2}$

1. $P(X > 0) = P\left(Z > \frac{0-3}{2} = -1.5\right) = \Phi(1.5) \approx 0.9332$
2. $P(2 < X < 5) = P(-0.5 < Z < 1) = \Phi(1) - \Phi(-0.5)$
   Using symmetry $\Phi(-0.5) = 1 - \Phi(0.5) = 0.3085$
   So $\Phi(1) - \Phi(-0.5) = 0.8413 - 0.3085 = 0.5328$
3. Find x with $P(X \leq x) = 0.975$.
   We know $z_{0.975} \approx 1.96$. Thus

   $x = \mu + \sigma z = 3 + 2(1.96) = 6.92$

CDF : If $Z \sim \mathcal{N}(0,1)$ with pdf $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, its distribution function is :

$$\Phi(x) = P(Z \leq x) = \int_{-\infty}^{x} \varphi(t)\,dt$$

Properties :
- $\lim_{x \to -\infty} \Phi(x) = 0$
- $\lim_{x \to +\infty} \Phi(x) = 1$
- Symmetry : $\Phi(-x) = 1 - \Phi(x)$
  $\Rightarrow \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$

Normal approximation to the Binomial :

If $X \sim \mathrm{Bin}(n,p)$ with n large (and p not too close to 0 or 1), then

$$X \approx \mathcal{N}(\mu = np,\ \sigma^2 = np(1-p))$$

Equivalently, the standardized variable $Z = \frac{X - np}{\sqrt{np(1-p)}}$ is approximately $\mathcal{N}(0,1)$.

- **Other distributions**

Chi - squared distribution . Let $X_1, \dots, X_\nu$ be independant standard normals $X_i \sim \mathcal{N}(0,1)$. The random variable

$$Y = \sum_{i=1}^{\nu} X_i^2$$

follows a chi - squared distribution with ν degrees of freedom, written $Y \sim \chi^2(\nu)$

- ## Statistical Inference

Descriptive stats	Inferential stats
Summarize what you saw (tables, means, plots)	Go beyond the sample to the population by using probability models

Workflow :

1. Choose a probabilistic model for the phenomenon
   (ex : lifetime X of a bulb is exponential $\mathcal{E}(\lambda)$)
2. Collect data (a sample) : $x_1, \dots, x_n$
3. Estimate the model's parameter(s) (e.g. λ)
4. Test claims (e.g., is $\lambda \leq \lambda_0$ so mean life $\geq 1/\lambda_0$?)
5. Predict quantities of interest (e.g., expected failures in 50h)

- ## Sampling and the statistic $\overline{X_n}$

What is "sampling" ?

- Population : the whole set we care about
- Parameter(s) : unknown numbers describing the population (mean μ, variance σ^2 ...)
- Sample : n observations we actually collect $(x_1, \dots, x_n)$
- Random sample : model these as i.i.d random variables $X_1, \dots, X_n$ with the same distribution as the population

Sample mean : $\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$

Properties . $E(\overline{X_n}) = \mu$ $\quad Var(\overline{X_n}) = \frac{\sigma^2}{n} \Rightarrow SE = \frac{\sigma}{\sqrt{n}}$ standard error

- ## Law of Large Numbers

If $X_1, \dots, X_n$ are i.i.d with mean μ and finite variance σ^2, then the sample mean $\overline{X_n}$ converges in probability to μ :

$$\overline{X_n} \xrightarrow{P} \mu \quad \text{as} \quad n \to \infty$$

- ## Central Limit Theorem

For i.i.d $X_1, \dots, X_n$ with mean μ and finite variance σ^2,

$$Z_n = \frac{\sqrt{n}\,(\overline{X_n} - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1)$$

Equivalently, for large n,

$$\overline{X_n} \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) = \mathcal{N}(E(\overline{X_n}), Var(\overline{X_n}))$$

When we use it :

- To approximate probabilities about $\overline{X_n}$:

$$P(a \leq \overline{X_n} \leq b) \approx \Phi\left(\frac{b-\mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{a-\mu}{\sigma/\sqrt{n}}\right)$$

- To build approximate confidence intervals for μ when n is large (and either σ known, or replace by s if n is big)

Rule of thumb : $n \geqslant 30$ is often "large enough"