Statistics
Semester 4

Contents

Chapter 1

Revision of Probability

I'm simply gonna list rules.

$$\begin{aligned}
\mathbb{E}(X) &= \mu = \sum_{i\in\Omega} X_i \Pr(X_i) \\
\mathbb{E}(g(X)) &= \sum_{i\in\Omega} g(X_i)\Pr(X_i) \\
\mathbb{E}(aX+b) &= a\mathbb{E}(X)+b \\
\mathbb{E}(X+Y) &= \mathbb{E}(X)+\mathbb{E}(Y) \quad \text{if both variables are independent}
\end{aligned}$$

$$\begin{aligned}
\operatorname{Var}(X) &= \sigma^2 = \mathbb{E}(X^2)-\mu^2 \\
\operatorname{Var}(aX+bY) &= a^2\operatorname{Var}(X)+b^2\operatorname{Var}(Y)+2ab\operatorname{cov}(X,Y)
\end{aligned}$$

where

$$\operatorname{cov}(X,Y) = \mathbb{E}(XY)-\mathbb{E}(X)\cdot\mathbb{E}(Y).$$

1.1 Discrete Distributions

1. **Uniform discrete law**

$$\begin{aligned}
&X(\Omega) = \{1,2,3,\dots,n\} \\
&\Pr(X=k) = \frac{1}{n} \quad \forall k = 1,2,3,\dots,n \\
&\begin{cases} \mathbb{E}(X) = \frac{n+1}{2} \\ \operatorname{Var}(X) = \frac{n^2-1}{12} \end{cases}
\end{aligned}$$

2. **Bernoulli law of parameters** p $(0 < p < 1)$

$$\begin{aligned}
&X \sim \mathrm{B}(p) \\
&X(\Omega) = \{0,1\} \\
&\Pr(X=1) = p \quad \Pr(X=0) = 1-p \\
&\begin{cases} \mathbb{E}(X) = p \\ \operatorname{Var}(X) = p(1-p) \end{cases}
\end{aligned}$$

3. **Binomial law of parameters n and p**

$$X \sim \mathrm{Bin}(n,p)$$
$$X(\Omega) = \{1,2,\ldots,n\}$$
$$\Pr(X=1) = C_n^k p^k q^{n-k} \quad \forall k \in \{0,1,2,\ldots,n\}$$
$$\begin{cases} \mathbb{E}(X) = np \\ \mathrm{Var}\,(\mathrm{X}) = np(1-p) \end{cases}$$

4. **Hypergeometric law**

$$X \sim \mathcal{H}(N,n,p)$$
$$X(\Omega) = [\max\{0, n-N+M\}, \min\{M,n\}]$$
$$\Pr(X=k) = \frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n} \quad \forall k \in X(\Omega)$$
$$\begin{cases} \mathbb{E}(X) = np \\ \mathrm{Var}\,(X) = np(1-p)\left(\frac{N-n}{N-1}\right) \end{cases}$$

5. **Geometric law**

$$\mathrm{X} \sim \mathrm{G}(p)$$
$$X(\Omega) = \mathbb{N}^*$$
$$\Pr(\mathrm{X}=k) = p(1-p)^{k-1} \quad \forall k \in \mathbb{N}^*$$
$$\begin{cases} \mathbb{E}(\mathrm{X}) = \frac{1}{p} \\ \mathrm{Var}\,(X) = \frac{1-p}{p^2} \end{cases}$$

6. **Poisson's law of parameter λ ($\lambda \in \mathbb{R}_+^*$)**

$$X \sim \mathcal{P}(\lambda)$$
$$X(\Omega) = \mathbb{N}$$
$$\Pr(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}$$
$$\begin{cases} \mathbb{E}(\mathrm{X}) = \lambda \\ \mathrm{Var}\,(\mathrm{X}) = \lambda \end{cases}$$

1.2 Continuous Distributions

1. **Uniform law**

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{else} \end{cases}$$
$$\begin{cases} \mathbb{E}(x) = \frac{a+b}{2} \\ \mathrm{Var}\,(x) = \frac{(b-a)^2}{12} \end{cases}$$

2. **Continuous law**

$$x \sim \xi(\lambda)$$
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$$
$$\begin{cases} \mathbb{E}(x) = \frac{1}{\lambda} \\ \mathrm{Var}\,(x) = \frac{1}{\lambda^2} \end{cases}$$

3. **Normal law**

$$
\begin{aligned}
& x \sim \mathcal{N}(\mu, \sigma) \\
& f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
& \begin{cases} \mathbb{E}(x) = \mu \\ \operatorname{Var}(x) = \sigma^2 \end{cases}
\end{aligned}
$$

For $\mathcal{N}(0,1)$

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x.$$

$$\pi(z) = \Phi(z) - 0.5 = \int_{0}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x.$$

1.3 Convergence

Theorem 1.3.1 Chebyshev's inequality

Let X be a random variable of expectation $\mathbb{E}(X)$ and variance $\operatorname{Var}(X)$. Then $\forall \varepsilon$

$$\Pr(|X - \mathbb{E}(X)| \geqslant \varepsilon) \leqslant \frac{\operatorname{Var}(X)}{\varepsilon^2}.$$

it can also be stated as

$$\Pr(|X - \mathbb{E}(X)| < \varepsilon) \geqslant 1 - \frac{\operatorname{Var}(X)}{\varepsilon^2}.$$

We say a sequence of random variables X_n converges to a $(X_n) \xrightarrow{\Pr} a$ if $\forall \varepsilon$

$$\lim_{n \to +\infty} \Pr(|X_n - a| > \varepsilon) = 0.$$

or

$$\lim_{n \to +\infty} \Pr(|X_n - a| \leqslant \varepsilon) = 1.$$

Theorem 1.3.2 Weak law of large numbers

Consider a random variable (X_n) of mean μ and variance σ^2. Consider the random variable $\tilde{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$. It can be shown that $\tilde{X}_n$ converges to μ meaning $\forall \varepsilon$

$$\lim_{n \to +\infty} \Pr\left(|\tilde{X}_n - \mu| > \varepsilon\right) = 0.$$

1.4 Approximations

Theorem 1.4.1 Binomial by a Poisson

$$\operatorname{Bin}(n, p) \sim \mathcal{P}(np) \quad \text{if } \begin{cases} n \geqslant 30 \\ p \leqslant 0.1 \\ np < 15 \end{cases}.$$

Theorem 1.4.2 Hypergeometric by a Binomial

$$\mathcal{H}(N, n, p) \sim \mathrm{Bin}(n, p) \quad \text{if } n \leqslant 0.05N.$$

Theorem 1.4.3 De Moivre–Laplace theorem

$$\mathrm{Bin}(n, p) \sim \mathcal{N}\left(np, \sqrt{np(1-p)}\right) \quad \text{if } \begin{cases} n \geqslant 30 \\ np \geqslant 5 \\ n(1-p) \geqslant 5 \end{cases}.$$

In this case the event $X = k$ can be replaced by $k - 0.5 < X < l + 0.5$

Theorem 1.4.4 Central limit theorem

Let (X_n) be a sequence of independent random variables following the same law of expectation μ and of standard deviation σ. Let $S_n = \sum_{i=1}^{n} X_i$ and $S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. It can be shown that S_n^* converges in law to $\mathcal{N}(0, 1)$.

$$\mathbb{E}(S_n) = n\mu$$
$$\mathrm{Var}\,(S_n) = n\sigma^2$$

1.5 Further laws

Theorem 1.5.1 Chi square law

Let $X_1, X_2, \ldots, X_n$ be n independent random variables following the standard normal law $\mathcal{N}(0, 1)$. Let $Y = {X_1}^2 + {X_2}^2 + \cdots + {X_n}^2$. We say that Y follows a chi-square law with n degrees of freedom. $Y \sim {\chi_n}^2$.

$$\mathbb{E}(Y) = n$$
$$\mathrm{Var}\,(Y) = 2n$$

It can be shown that the density function of Y is

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}.$$

where Γ is the gamma function

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} \mathrm{d}t \quad \forall x > 0.$$

Theorem 1.5.2 Student law(t-distribution)

Let X, Z be two independent random random variables such that $X \sim \mathcal{N}(0, 1)$ and $Z \sim {\chi_n}^2$. Hence the random variable

$$T = \frac{X}{\sqrt{\frac{Z}{n}}}.$$

is said to be following a student law. $T \sim \mathcal{T}_n$

$$f(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$