

# 1 Probability and random variables

- **Probability:**  $S$  sample space (all possible states of the system),  $F \subset \mathcal{P}(S)$  a  $\sigma$ -algebra,  $P : F \rightarrow \mathbb{R}$  a measure, such that  $P(S) = 1$ .
- **Random variable:**  $X : S \rightarrow \mathbb{R}$ , such that preimages of open sets are in  $F$  (i.e. has a well defined probability).
- **Cumulative distribution function** of random variable:  $F_X(t) = P(X \leq t)$ .
- **Probability distribution** of random variable:  $g$  such that  $F_X(t) = \sum_{x \leq t, x \in C} g(x)$ .
- **Probability density function:**  $f$  such that  $F_X(t) = \int_{-\infty}^t f(s)ds$ .
- Two random variables have the **same distribution** if they have the same cdf.

Example: **uniform distribution:**

- $S$  a finite interval  $[a, b]$
- $F$ : Set of Borel sets on  $S$  (sets with a well defined “length”)
- $P$ : Borel measure (“length”) divided by  $b - a$
- $X = id$ .

## 1.1 Expectation of random variables and their functions

- $X$  is a random variable, the **expectation** of  $X$  is  $E[X] = \int_S X dP$ .
- The **variance** of  $X$  is  $E[(X - E[X])^2]$ .
- The  $k$ -th **moment** of  $X$  is  $E[X^k]$ .
- The **moment generating function** of  $X$  is  $E[e^{Xt}]$  (two sided Laplace transform)
- The **characteristic function** of  $X$  is  $E[e^{itX}]$  (Fourier transform)

Since expectation is defined via integration, one can use the properties of integration to prove statements regarding expectation.

Example: **Chebyshev’s theorem:**  $E[X] = 0$ ,  $E[X^2] = 1$ , then  $P(|X| < k) \geq 1 - \frac{1}{k^2}$ .  
Proof:

$$1 = E[X^2] = \int_S X^2 dP \geq k^2 \int_{|X| \geq k} 1 dP = k^2(1 - P(|X| < k))$$

Example: If  $X$  has p.d.f.  $f_X$ , then  $E[g(X)] = \int_{-\infty}^{\infty} g f_X dt$ . We prove it when  $g(X)$  is bounded via Fubini's theorem:

$$\begin{aligned} E[g(X)] &= \int_S g(X) dP \\ &= \int_{g(X) \geq 0} \int_0^{g(X)} 1 dy dP - \int_{g(X) < 0} \int_{g(X)}^0 1 dy dP \\ &= \int_0^{\infty} \int_{g^{-1}([y, \infty))} f_X(t) dt dy - \int_{-\infty}^0 \int_{g^{-1}([-\infty, y])} f_X(t) dt dy \\ &= \int_{-\infty}^{\infty} g f_X dt \end{aligned}$$

There is a multivariate version of this formula, and one can also write down  $E[g(X)]$  when only the c.d.f. of  $X$  is known (via Fubini's theorem or integration by parts).

Can you write down a random variable with neither probability distribution nor p.d.f.?

Can you write down a random variable with no expectation?

## 1.2 Independence and conditional probability for random events

- $A, B \in \mathcal{F}$  are **independent** iff  $P(A \cap B) = P(A)P(B)$ .
- If  $P(B) \neq 0$ ,  $P(A \cap B) = P(B)P(A|B)$ . Here  $P(A|B)$  is the **conditional probability** of  $A$  when  $B$  is known to happen.

## 1.3 Joint distribution, marginal distribution, conditional distribution

### 1.3.1 Joint distribution

- $X$  and  $Y$  are two random variables. The **joint cumulative distribution function** is  $F(s, t) = P(X \leq s, Y \leq t)$ .
- If  $F(s, t) = \sum_{(x, y) \in C, x \leq s, y \leq t} g(s, t)$ , we call  $g$  the **joint probability distribution**.
- If  $F(s, t) = \int_{(-\infty, s] \times (-\infty, t]} f(x, y) dx dy$  we call  $f$  the **joint probability density function**.
- $X$  and  $Y$  are called independent iff the joint c.d.f. is  $F(x, y) = F_X(x)F_Y(y)$ .
- The **covariance** between  $X$  and  $Y$  is  $E[(X - E[X])(Y - E[Y])]$

Example:  $X$  and  $Y$  are two independent random variable with uniform distribution on  $[0, 1]$ . What is the joint distribution function of  $X$  and  $Y$ ? How about  $\max(X, Y)$  and  $\min(X, Y)$ ? What are their covariances?

### 1.3.2 Marginal distribution

Knowing the joint c.d.f. of  $X$  and  $Y$ , the c.d.f. of  $X$  or  $Y$  are called the **marginal cumulative distribution function**, their p.d. or p.d.f. the **marginal p.d. or marginal p.d.f.**

### 1.3.3 Conditional distribution

- If  $A$  is a set such that  $P(Y \in A) > 0$ , then the **conditional cumulative distribution function** of  $X$  is  $F_{X|Y \in A}(t) = P(X \leq t | Y \in A) = P(X \leq t \cap Y \in A) / P(Y \in A)$ . The **conditional p.d.f.**, **conditional p.d.** and **conditional expectation** are defined similarly.
- If  $P(Y \in A) = 0$  there isn't a definition of conditional distribution that works in all cases. For example, if  $X, Y$  has joint p.d.f.  $f_{X,Y}$ , and the marginal p.d.f. of  $Y$ , denoted as  $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx$ , exists and is non zero at  $y_0$ , then the conditional p.d.f. at  $Y = y_0$  is defined as  $f_{X|Y=y_0} = f_{X,Y}(x, y_0) / f_Y(y_0)$ . The conditional c.d.f. is its integral.

Remark: The definition of conditional distribution for the case  $P(Y \in A) = 0$  depends on  $Y$  and not just  $Y^{-1}(A)$ . For example, if  $Z = Ye^X$ ,  $f_{X|Y=0} \neq f_{X|Z=0}$ .

Example:  $X$  is a random variable with uniform distribution on  $[0, 1]$ ,  $P(Y = 1 | X = p) = p$  (i.e.  $P(Y = 1 | X \in A) = \int_A p dF_x(p)$ ),  $P(Y = 0 | X = p) = 1 - p$ . Find the conditional distribution of  $X$  when  $Y = 1$ .

When there are  $N$  random variables,  $N \geq 3$ , the joint/marginal/conditional distributions can be defined analogously.

## 2 Special probability distributions, central limit theorem

### 2.1 Special discrete distributions

- **Bernoulli distribution:**  $f(1) = \theta$ ,  $f(0) = 1 - \theta$ .
- **Binomial distribution** (sum of iid Bernoulli):  $f(x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$ ,  $x = 0, 1, \dots, n$ .

- **Negative Binomial distribution** (waiting time for the  $k$ -th success of iid trials):  $f(x) = \binom{x-1}{k-1} \theta^k (1-\theta)^{x-k}$ ,  $x = k, k+1, \dots$ . When  $k = 1$  it is the **geometric distribution**.
- **Hypergeometric distribution** (randomly pick  $n$  elements at random from  $N$  elements, the number of elements picked from a fixed subset of  $M$  elements)  $f(x) = \binom{M}{x} \binom{N-M}{n-x} \binom{N}{n}^{-1}$ .
- **Poisson distribution** (limit of binomial as  $n \rightarrow \infty$ ,  $n\theta \rightarrow \lambda$ )  $f(x) = \lambda^x e^{-\lambda} / x!$ .
- **Multinomial distribution**  $f(x_1, \dots, x_k) = \binom{n}{x_1, \dots, x_k} \theta_1^{x_1} \dots \theta_k^{x_k}$ ,  $\sum_i x_i = n$ ,  $\theta_i \theta_i = 1$ .
- **Multivariate Hypergeometric distribution**  $f(x_1, \dots, x_k) = \prod_i \binom{M_i}{x_i} \binom{N}{n}^{-1}$ .  $\sum_i x_i = n$ ,  $\sum_i M_i = N$ .

## 2.2 Special continuous distributions

- **Uniform distribution:**  $f(x) = \begin{cases} 1/(b-a) & x \in (a, b) \\ 0 & x \notin (a, b) \end{cases}$ .
- **Normal distribution:**  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ .
- **Multivariate Normal distribution:**  $x \in \mathbb{R}^d$ ,  $\Sigma$  positive definite  $d \times d$  symmetric matrix,  $f(x) = (2\pi)^{-d/2} |\Sigma|^{-1/2} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$ .
- **$\chi^2$  distribution**  $d$ : degrees of freedom. Squared sum of  $d$  normal distributions:  $f(x) = \begin{cases} \frac{1}{2^{d/2} \Gamma(d/2)} x^{\frac{d-2}{2}} e^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases}$ .
- **Exponential distribution**  $f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x > 0 \\ 0 & x \leq 0 \end{cases}$ .
- **Gamma-distribution:**  $f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x > 0 \\ 0 & x \leq 0 \end{cases}$ .
- **Beta distribution:** (conjugate prior of Bernoulli distribution)  $f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & x \in (0, 1) \\ 0 & x \notin (0, 1) \end{cases}$ .

## 2.3 Law of Large Numbers and Central Limit Theorem

### 2.3.1 Convergence

- **Convergence in distribution:** cdf pointwise convergence.
- **Convergence almost surely:**  $P(\lim_i X_i \neq X) = 0$ .

Example:  $X$  uniform on  $[0, 1]$ ,  $Y_i = \begin{cases} 1 & \exists n \in \mathbb{Z} (X + n \in [\sum_{j=1}^i \frac{1}{j}, \sum_{j=1}^{i+1} \frac{1}{j}]) \\ 0 & \text{otherwise} \end{cases}$ .

Then  $Y_i$  converges to 0 in distribution but not almost surely.

### 2.3.2 CLT and weak LLN

**Levy's continuity theorem:** If  $\phi_{X_j} \rightarrow \phi_X$  pointwise, then  $X_j$  converges to  $X$  in distribution.

**Weak Law of Large Numbers**  $X_i$  i.i.d. with expectation  $\mu$ .  $S_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Then  $S_n$  converges to  $\mu$  in distribution.

**(Levy's) Central Limit Theorem**  $X_i$  i.i.d. with expectation  $\mu$  and variance  $\sigma^2 > 0$ .  $Y_n = \sqrt{\frac{1}{n\sigma^2}} \sum_i (X_i - \mu)$ , then  $Y_n$  converges in distribution to standard normal distribution (normal distribution with  $\mu = 0$  and  $\sigma^2 = 1$ ).

Proof of both theorems (assume  $X_i$  bounded): Taylor expansion of the characteristic function.

One can also use the continuity of moment generating function, which is the argument in the textbook.

### 2.3.3 Strong Law of Large Numbers

**Borel-Cantelli Lemma**  $A_i$  events,  $i = 1, 2, \dots$ ,  $\sum_i (A_i) < \infty$ , then  $P(\cap_i (\cup_{j>i} A_j)) = 0$ . (the probability of infinitely many  $A_i$  happening is 0)

Proof:  $P(\cap_i (\cup_{j>i} A_j)) \leq P(\cup_{j>i} A_j) \leq \sum_{j>i} P(A_j)$  which converges to 0 as  $i \rightarrow \infty$ .

**Strong Law of Large Numbers**  $X_i$ ,  $i = 1, 2, \dots$  i.i.d. (independent with identical distribution) and  $E(X_i) = \mu$ , then  $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$  converges a.s. to constant  $\mu$ .

Proof (assume  $X_i$  bounded by  $M$ ): Suppose  $Var(X_i) = m$ .  $\sqrt{\frac{n}{m}}(Y_n - \mu)$  has expectation 0 and variance 1, so  $P(|Y_n - \mu| > C\sqrt{\frac{m}{n}}) < 1/C^2$  by Chebyshev's theorem. Now let  $n_k = k^4$ ,  $C_k = k$ , then  $Y_{n_k} = Y_{k^4}$  converges a.s. to  $\mu$  by Borel-Cantelli.

$Y_n = (\lfloor n^{1/4} \rfloor^4 Y_{\lfloor n^{1/4} \rfloor^4} + X_{\lfloor n^{1/4} \rfloor^4 + 1} + \dots + X_n) / n = Y_{\lfloor n^{1/4} \rfloor^4} + (M + |\mu|) \frac{n - \lfloor n^{1/4} \rfloor^4}{n}$ .  
The first term converges to  $\mu$  as  $n \rightarrow \infty$ , and the second converges to 0.

### 3 Sample statistics

#### 3.1 Some important distributions

- Standard Normal Distribution:  $\mathcal{N}(0, 1)$
- $\chi^2(k)$ : squared sum of  $k$  independent standard normal distribution.
- $t$  distribution:  $Z$  standard normal,  $Y \sim \chi^2(k)$ ,  $Z$  and  $Y$  independent, then  $T = \frac{Z}{\sqrt{Y/k}}$  is said to have  $t$ -distribution with  $k$  degrees of freedom.
- $F$  distribution:  $U$  and  $V$  independent,  $U \sim \chi^2(m)$ ,  $V \sim \chi^2(n)$ , then  $F = \frac{U/m}{V/n}$  is said to have  $F$  distribution with degrees of freedom  $m$  and  $n$ ,

#### 3.2 Sample statistics

$X_1, \dots, X_n$  i.i.d. (independent with identical distributions).

- **Sample mean:**  $\bar{X} = \frac{\sum_i X_i}{n}$ 
  - $E[\bar{X}] = E[X_1]$ ,  $Var(\bar{X}) = \frac{1}{n} Var(X_1)$ .
  - If  $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ ,  $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ .
- **Sample variance:**  $S^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$ .
  - $E[S^2] = Var(X_1)$ .
  - If  $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ , then
    - \*  $\bar{X}$  and  $S$  are independent
    - \*  $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
    - \*  $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$
  - If  $S_1^2$  is the sample variance of  $n_1$  i.i.d.  $\mathcal{N}(\mu, \sigma^2)$  random variables  $Y_i$ ,  $S_2^2$  the sample variance of  $n_2$  i.i.d.  $\mathcal{N}(\mu', \sigma'^2)$  random variables  $Z_j$  independent from  $Y_i$ , then  $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$
- **Order statistics** The  $k$ -th order statistics is the  $k$ -th smallest element in  $\{X_i\}$ , denoted as  $Y_k$ . Then, if  $X_1$  has pdf  $f$ , then  $f_{Y_k}(t) = r \binom{n}{r} (\int_0^t f ds)^{r-1} f(t) (\int_t^\infty f ds)^{n-r}$ .

- 3.3 PDF of  $\chi^2$ -, t- and F- distributions
- 3.4 Drawing without replacement
- 4 Point estimators and their properties
- 5 Method of moments, Maximum likelihood
- 6 Maximum a posteriori
- 7 Hypothesis testing
- 8 Examples of hypothesis testing
- 9 Confidence interval
- 10 Linear Regression
- 11 ANOVA
- 12 Example of non parametric methods