

# Probability Memorandum

## 1 Events and probability

- **Definition:** A **probability space** is a triple  $(\Omega, F, P)$ , where:

1.  $\Omega$  is the **sample space**;
2.  $F$  is a collection of subsets of  $\Omega$ , called **events**, satisfying axioms **F**<sub>1</sub> – **F**<sub>3</sub> below;
3.  $P$  is a **probability measure**, which is a function  $P : F \rightarrow [0, 1]$ , satisfying axioms **P**<sub>1</sub> – **P**<sub>3</sub> below;

- **The axioms of probability:**

- $F$  is a collection of subsets of  $\Omega$ , with:

**F**<sub>1</sub>:  $\emptyset \in F$ .

**F**<sub>2</sub>: If  $A \in F$ , then also  $A^c \in F$ .

**F**<sub>3</sub>: If  $\{A_i, i \in I\}$  is a finite or countably infinite collection of members of  $F$ , then  $\cup_{i \in I} A_i \in F$

- $P$  is a function from  $F$  to  $\mathbb{R}$ , with:

**P**<sub>1</sub>: For all  $A \in F$ ,  $P(A) \geq 0$ .

**P**<sub>2</sub>:  $P(\Omega) = 1$ .

**P**<sub>3</sub>: If  $\{A_i, i \in I\}$  is a finite or countably infinite collection of members of  $F$ , and  $A_i \cap A_j = \emptyset$  for  $i \neq j$ , then  $P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$ .

- **Theorem:** Let  $(\Omega, F, P)$  be a probability space, then:

1.  $P(A^c) = 1 - P(A)$ .
2. If  $A \subseteq B$ , then  $P(A) \leq P(B)$ .

- **Definition:** Let  $(\Omega, F, P)$  be a probability space. If  $A, B \in F$  and  $P(B) > 0$  then the **conditional probability** of  $A$  given  $B$  is

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

. (If  $P(B) = 0$ , then  $P(A|B)$  is not defined).

- **Theorem:** Let  $(\Omega, F, P)$  be a probability space and let  $B \in F$  satisfy  $P(B) > 0$ . Define a new function  $Q : F \rightarrow \mathbb{R}$  by  $Q(A) = P(A|B)$ . Then  $(\Omega, F, Q)$  is also a probability space.

- **Definition:**

1. Events  $A$  and  $B$  are independent if  $P(A \cap B) = P(A)P(B)$ .  
More generally
2. More generally, a family of events  $A = \{A_i, i \in I\}$  is independent if

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)$$

for all finite subsets  $J$  of  $I$ .

3. A family  $A$  of events is pairwise independent if  $P(A_i \cap A_j) = P(A_i)P(A_j)$ , whenever  $i \neq j$ .

- **Theorem:** Suppose that  $A$  and  $B$  are independent. Then:

1.  $A$  and  $B^c$  are independent.
2.  $A^c$  and  $B^c$  are independent.

- **Definition:** A family of events  $\{B_1, B_2, \dots\}$  is a **partition** of  $\Omega$  if:

1.  $\Omega = \bigcup_{i \geq 1} B_i$ ;
2.  $B_i \cap B_j = \emptyset$  whenever  $i \neq j$ .

- **The law of total probability:** Suppose  $\{B_1, B_2, \dots\}$  is a partition of  $\Omega$  by sets from  $F$  such that  $P(B_i) > 0$  for all  $i \geq 1$ . Then, for any  $A \in F$ ,

$$P(A) = \sum_{i \geq 1} P(A|B_i)P(B_i)$$

- **Bayes' Theorem:** Suppose  $\{B_1, B_2, \dots\}$  is a partition of  $\Omega$  by sets from  $F$  such that  $P(B_i) > 0$  for all  $i \geq 1$ . Then, for any  $A \in F$  such that  $P(A) > 0$ ,

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i \geq 1} P(A|B_i)P(B_i)}$$

## 2 Discrete random variables

- **Definition:** A **discrete random variable**  $X$  on a probability space  $(\Omega, F, P)$  is a function  $X : \Omega \rightarrow \mathbb{R}$  such that:

1.  $\{\omega \in \Omega : X(\omega) = x\} \in F$  for each  $x \in \mathbb{R}$ ;
2.  $ImX := \{X(\omega) : \omega \in \Omega\}$  is a finite or countably infinite subset of  $\mathbb{R}$ .

- **Definition:** The **probability mass function** (p.m.f.) of  $X$  is the function  $p_X : \mathbb{R} \rightarrow [0, 1]$  defined by

$$p_X(x) = P(X = x)$$

- **Definition:** The **expectation** (or **mean** or **expected value**) of  $X$  is

$$E[X] = \sum_{x \in ImX} xP(X = x)$$

provided that  $\sum_{x \in ImX} |x|P(X = x) < \infty$ . If  $\sum_{x \in ImX} |x|P(X = x)$  is infinite, we say that the expectation does not exist.

- **Theorem:** If  $h : \mathbb{R} \rightarrow \mathbb{R}$ , then

$$E[h(X)] = \sum_{x \in ImX} h(x)P(X = x)$$

provided that  $\sum_{x \in ImX} h(x)P(X = x) < \infty$ .

- **Theorem:** Let  $X$  be a discrete random variable such that  $E[X]$  exists.

1. If  $X$  is non-negative, then  $E[X] \geq 0$ .
2. If  $a, b \in \mathbb{R}$  then  $E[aX + b] = aE[X] + b$ .

- **Definition** For a discrete random variable  $X$ , the **variance** of  $X$  is defined by

$$var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

provided that this quantity exists.

- **Theorem:** Let  $X$  be a discrete random variable and  $a, b \in \mathbb{R}$ , then

$$\text{var}(aX + b) = a^2 \text{var}(X)$$

- **Definition:** Suppose that  $B$  is an event such that  $P(B) > 0$ . Then the **conditional distribution** of  $X$  given  $B$  is

$$P(X = x|B) = \frac{P(\{X = x\} \cap B)}{P(B)}$$

for  $x \in \mathbb{R}$ . The **conditional expectation** of  $X$  given  $B$  is

$$E[X|B] = \sum_x xP(X = x|B)$$

whenever the sum converges absolutely. We write  $p_{X|B}(x) = P(X = x|B)$ .

- **Partition theorem for expectations:** Suppose  $\{B_1, B_2, \dots\}$  is a partition of  $\Omega$  by sets from  $\mathcal{F}$  such that  $P(B_i) > 0$  for all  $i \geq 1$  then

$$E[X] = \sum_{i \geq 1} E[X|B_i]P(B_i)$$

- **Definition:** Given two random variable  $X$  and  $Y$ , their **joint distribution** is

$$p_{X,Y}(x, y) = P(\{X = x\} \cap \{Y = y\})$$

. The **marginal distribution** of  $X$  is

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

- **Theorem:** Suppose  $X$  and  $Y$  are discrete random variables and  $a, b \in \mathbb{R}$ . Then

$$E[aX + bY] = aE[X] + bE[Y]$$

provided that both  $E[X]$  and  $E[Y]$  exist.

- **Theorem:** If  $X$  and  $Y$  are independent random variables whose expectations exist, then

$$E[XY] = E[X]E[Y]$$

- **Definition:** The **covariance** of two random variables  $X$  and  $Y$  is defined by

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

. Also,

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$$

- **Theorem:**

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{i < j} \text{cov}(X_i, X_j)$$

### 3 Difference equations and random walks

- **Definition:** A  $k$ th order **linear recurrence relation** (or **difference equation**) has the form

$$\sum_{j=0}^k a_j u_{n+j} = f(n)$$

with  $a_0 \neq 0$  and  $a_k \neq 0$ , where  $a_0 \dots a_k$  are constants independent of  $n$ . A **solution** to such a difference equation is a sequence  $(u_n)_{n \geq 0}$  satisfying the above equation for all  $n \geq 0$ .

- **Theorem:** The general solution  $(u_n)_{n \geq 0}$  (if the boundary conditions are not specified) can be written as  $u_n = v_n + w_n$ , where  $(v_n)_{n \geq 0}$  is a particular solution to the equation and  $(w_n)_{n \geq 0}$  solves the homogeneous equation

$$\sum_{j=0}^k a_j w_{n+j} = 0$$

- **Theorem:** Consider a random walk on the integer  $\mathbb{Z}$ , started from some  $n > 0$ , which at each step increases by 1 with probability  $p$ , and decreases by 1 with probability  $q = 1 - p$ . Then the probability  $u_n$  that the walk ever hits 0 is given by

$$u_n = \begin{cases} \left(\frac{q}{p}\right)^n & \text{if } p > q, \\ 1 & \text{if } p \leq q. \end{cases}$$

### 4 Probability generating functions

- **Definition:** Let  $X$  be a non-negative integer-valued random variable. Let

$$S = \left\{ s \in \mathbb{R} : \sum_{k=0}^{\infty} |s|^k P(X = k) < \infty \right\}$$

. Then the **probability generating function** (p.g.f.) of  $X$  is  $G_X : S \subset \mathbb{R}$  defined by

$$G_X(s) = E[s^X] = \sum_{k=0}^{\infty} s^k P(X = k)$$

- **Theorem:** The distribution of  $X$  is uniquely determined by its probability generating function,  $G_X$ .

- **Theorem:** If  $X$  and  $Y$  are independent random variables, then

$$G_{X+Y}(s) = G_X(s)G_Y(s)$$

- **Theorem:** Suppose that  $X_1, X_2, \dots, X_n$  are independent  $Ber(p)$  random variables and let  $Y = X_1 + \dots + X_n$ . Then  $Y \sim Bin(n, p)$ .

- **Theorem:** Suppose that  $X_1, X_2, \dots, X_n$  are independent random variables such that  $X_i \sim Po(\lambda_i)$  and let  $Y = X_1 + \dots + X_n$ . Then  $Y \sim Po(\sum_{i=1}^n \lambda_i)$ .

- **Theorem:** Let  $X$  be a random variable with p.f.g.  $G_X$ . Then

1.  $E[X] = G'_X(1)$
2.  $var(x) = G''_X(1) + G'_X(1) - (G'_X(1))^2$
3. In general,  $E[X(X-1)\dots(X-k+1)] = \frac{d^k G_X}{ds^k}(1)$

- **Theorem:** Let  $X_1, X_2, \dots$  be i.i.d. non-negative integer-valued random variables with p.g.f.  $G_X(s)$ . Let  $N$  be another non-negative integer-valued random variable with p.g.f.  $G_N(s)$ , independent of  $X_1, X_2, \dots$ . Then the p.g.f. of  $\sum_{i=1}^N X_i$  is  $G_N(G_X(s))$ .
- **Branching process:** Suppose we have a population. Each individual in the population survives just one unit of time and, just before dying, gives birth to a random number of children in the next generation. This number of children has probability mass function  $p(i)$ , called the **offspring distribution**. Let  $X_n$  be the size of the population in generation  $n$ , so that  $X_0 = 1$ . Let  $C_i^{(n)}$  be the number of children of the  $i$ th individual in generation  $n \geq 0$ , so that we can write

$$X_{n+1} = C_1^{(n)} + C_2^{(n)} + \dots + C_{X_n}^{(n)}$$

- **Theorem:** For  $n \geq 0$ ,

$$G_{n+1}(s) = G(G_n(s)) = \underbrace{G(G(\dots G(s) \dots))}_{n+1 \text{ times}} = G(G_n(s))$$

Hence, suppose that the mean number of children of a single individual is  $\mu = \sum_{i=1}^{\infty} ip(i)$ , then

$$E[X_n] = \mu^n$$

- **Theorem:** The extinction probability  $q$  is the smallest non-negative solution of

$$x = G(x)$$

- **Theorem:** Assume  $p(1) \neq 1$ . Then  $q = 1$  if  $\mu \leq 1$  and  $q < 1$  if  $\mu > 1$ .

## 5 Continuous random variables

- **Definition:** A **random variable**  $X$  defined on a probability space  $(\Omega, F, P)$  is a function  $X : \Omega \rightarrow \mathbb{R}$  such that  $\{\omega : X(\omega) \leq x\} \in F$  for each  $x \in \mathbb{R}$ .
- **Definition:** The **cumulative distribution function** (c.d.f.) of a random variable  $X$  is the function  $F_X : \mathbb{R} \rightarrow [0, 1]$  defined by

$$F_X(x) = P(X \leq x)$$

- **Theorem:**

1.  $F_X$  is non-decreasing.
2.  $P(a < X \leq b) = F_X(b) - F_X(a)$  for  $a < b$ .
3.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$ .
4.  $\lim_{x \rightarrow \infty} F_X(x) = 1$ .

- **Definition:** A **continuous random variable**  $X$  is a random variable whose c.d.f. satisfies

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u) du$$

, where  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  is a function such that

1.  $f_X(u) \geq 0$  for all  $u \in \mathbb{R}$ .

2.  $\int_{-\infty}^{\infty} f_X(u)du = 1.$

$f_X$  is called the **probability density function** (or **density**) of  $X$ .

- **Theorem:** From the Fundamental Theorem of Calculus, we have

$$\frac{dF_X(x)}{dx} = f_X(x)$$

, at any point  $x$  where  $f_X(x)$  is continuous.

- **Theorem:** If  $X$  is a continuous random variable with p.d.f.  $f_X(x)$  then

$$P(X = x) = 0 \text{ for all } x \in \mathbb{R}$$

and

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

- **Definition:** Let  $X$  be a continuous random variable with p.d.f.  $f_X(x)$ . The expectation of  $X$  is defined to be

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

whenever  $\int_{-\infty}^{\infty} x f_X(x)dx < \infty$ . Otherwise, we say that the mean is undefined.

- **Theorem:** Let  $X$  be a continuous random variable and  $h$  a function  $h : \mathbb{R} \rightarrow \mathbb{R}$ . Then

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x)dx$$

.

- **Theorem:** Suppose that  $X$  is a continuous random variable with density  $f_X$  and that  $h : \mathbb{R} \rightarrow \mathbb{R}$  is a differentiable function which is strictly increasing (i.e.  $\frac{dh}{dx}(x) > 0$  for all  $x$ ). Then  $Y = h(X)$  is a continuous random variable with p.d.f.

$$f_Y(y) = f_X(h^{-1}(y)) \frac{d}{dy} h^{-1}(y)$$

where  $h^{-1}$  is the inverse function of  $h$ .

- **Definition:** Let  $X$  and  $Y$  be random variables such that

$$F_{X,Y}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{X,Y}(u, v)dudv$$

for some function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  such that

1.  $f_{X,Y}(u, v) \geq 0$  for all  $u, v \in \mathbb{R}$
2.  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(u, v)dudv = 1$

Then  $X$  and  $Y$  are **jointly continuous** and  $f_{X,Y}$  is their **joint density function**.

- **Theorem:** For a pair of jointly continuous random variables  $X$  and  $Y$ , we have

$$P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f_{X,Y}(u, v)dudv$$

for  $a < b$  and  $c < d$ .

- **Theorem:** Suppose  $X$  and  $Y$  are jointly continuous with joint density  $f_{X,Y}$ . Then  $X$  is a continuous random variable with density

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

.

- **Definition:** Jointly continuous random variables  $X$  and  $Y$  with joint density  $f_{X,Y}$  are independent if

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

for all  $x, y \in \mathbb{R}$ .

## 6 Random samples and the weak law of large numbers

- **Definition:** Let  $X_1, X_2, \dots, X_n$  denote i.i.d. random variables. Then these random variables are said to constitute a **random sample** of size  $n$  from the distribution.

- **Definition:** The **sample mean** is defined to be

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

- **Theorem:** Suppose that  $X_1, X_2, \dots, X_n$  form a random sample from a distribution with mean  $\mu$  and variance  $\sigma^2$ . Then the expectation and variance of the sample mean are

$$E[\bar{X}_n] = \mu \text{ and } \text{var}(\bar{X}_n) = \frac{1}{n} \sigma^2$$

- **Weak law of large numbers:** Suppose that  $X_1, X_2, \dots$  are i.i.d. random variables with mean  $\mu$ . Then for any fixed  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \epsilon \right) = 0$$

- **Markov's inequality:** Suppose that  $Y$  is a non-negative random variable whose expectation exists. Then

$$P(Y \geq t) \leq \frac{E[Y]}{t}$$

for all  $t > 0$ .

- **Chebyshev's inequality:** Suppose that  $Z$  is a random variable with a finite variance. Then, for any  $t > 0$ ,

$$P(|Z - E[Z]| \geq t) \leq \frac{\text{var}(Z)}{t^2}$$

## 7 Common discrete distributions

Distribution	Probability mass function	Mean	Variance	Generating function
<b>Uniform</b> $U\{1, 2, \dots, n\},$ $n \in \mathbb{N}$	$P(X = k) = \frac{1}{n}, 1 \leq k \leq n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$G_X(s) = \frac{s-s^{n+1}}{n(1-s)}$
<b>Bernoulli</b> $Ber(p), p \in [0, 1]$	$P(X = 1) = p, P(X = 0) = 1 - p$	$p$	$p(1 - p)$	$G_X(s) = 1 - p + ps$
<b>Binomial</b> $Bin(n, p),$ $n \in \mathbb{N}, p \in [0, 1]$	$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, k = 0, 1, \dots, n$	$np$	$np(1 - p)$	$G_X(s) = (1 - p + ps)^n$
<b>Poisson</b> $Po(\lambda), \lambda \geq 0$	$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$	$\lambda$	$\lambda$	$G_X(s) = e^{\lambda(s-1)}$
<b>Geometric</b> $Geom(p)$ $p \in [0, 1]$	$P(X = k) = (1 - p)^{k-1} p, k = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$G_X(s) = \frac{ps}{1-(1-p)s}$
<b>Alternative geometric</b> $p \in [0, 1]$	$P(X = k) = (1 - p)^k p, k = 0, 1, \dots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$G_X(s) = \frac{p}{1-(1-p)s}$
<b>Negative binomial</b> $NegBin(k, p)$ $k \in \mathbb{N}, p \in [0, 1]$	$P(X = k) = \binom{n-1}{k-1} (1 - p)^{n-k} p^k, n = k, k + 1, \dots$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$	$G_X(s) = \left( \frac{ps}{1-(1-p)s} \right)^k$

## 8 Common continuous distributions

Distribution	Probability density function	Cumulative distribution function	Mean	Variance
<b>Uniform</b> $U[a, b], a < b$	$f_X(x) = \frac{1}{b-a}, a \leq x \leq b$	$F_X(x) = \frac{x-a}{b-a}, a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
<b>Exponential</b> $Exp(\lambda), \lambda > 0$	$f_X(x) = \lambda e^{-\lambda x}, x \geq 0$	$F_X(x) = 1 - e^{-\lambda x}, x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
<b>Gamma</b> $Gamma(\alpha, \lambda)$ $\alpha > 0, \lambda \geq 0$	$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x \in \mathbb{R}$		$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
<b>Normal</b> $N(\sigma^2)$ $\mu \in \mathbb{R}, \sigma^2 > 0$	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$	$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$	$\mu$	$\sigma^2$
<b>Beta</b> $Beta(\alpha, \beta)$	$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, x \in [0, 1]$	$\frac{\alpha}{\alpha+\beta}$		$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta-1)}$