Part IA — Probability
Theorems

Based on lectures by R. Weber
Notes taken by Dexter Chua

Lent 2015

Basic concepts
Classical probability, equally likely outcomes. Combinatorial analysis, permutations and combinations. Stirling's formula (asymptotics for $\log n!$ proved). [3]

Axiomatic approach
Axioms (countable case). Probability spaces. Inclusion-exclusion formula. Continuity and subadditivity of probability measures. Independence. Binomial, Poisson and geometric distributions. Relation between Poisson and binomial distributions. Conditional probability, Bayes's formula. Examples, including Simpson's paradox. [5]

Discrete random variables
Expectation. Functions of a random variable, indicator function, variance, standard deviation. Covariance, independence of random variables. Generating functions: sums of independent random variables, random sum formula, moments.

Conditional expectation. Random walks: gambler's ruin, recurrence relations. Difference equations and their solution. Mean time to absorption. Branching processes: generating functions and extinction probability. Combinatorial applications of generating functions. [7]

Continuous random variables
Distributions and density functions. Expectations; expectation of a function of a random variable. Uniform, normal and exponential random variables. Memoryless property of exponential distribution. Joint distributions: transformation of random variables (including Jacobians), examples. Simulation: generating continuous random variables, independent normal random variables. Geometrical probability: Bertrand's paradox, Buffon's needle. Correlation coefficient, bivariate normal random variables. [6]

Inequalities and limits
Markov's inequality, Chebyshev's inequality. Weak law of large numbers. Convexity: Jensen's inequality for general random variables, AM/GM inequality.

Moment generating functions and statement (no proof) of continuity theorem. Statement of central limit theorem and sketch of proof. Examples, including sampling. [3]

Contents

0 Introduction **3**

1 Classical probability **4**
1.1 Classical probability 4
1.2 Counting 4
1.3 Stirling's formula 4

2 Axioms of probability **5**
2.1 Axioms and definitions 5
2.2 Inequalities and formulae 5
2.3 Independence 5
2.4 Important discrete distributions 6
2.5 Conditional probability 6

3 Discrete random variables **7**
3.1 Discrete random variables 7
3.2 Inequalities 8
3.3 Weak law of large numbers 9
3.4 Multiple random variables 9
3.5 Probability generating functions 10

4 Interesting problems **11**
4.1 Branching processes 11
4.2 Random walk and gambler's ruin 11

5 Continuous random variables **12**
5.1 Continuous random variables 12
5.2 Stochastic ordering and inspection paradox 12
5.3 Jointly distributed random variables 12
5.4 Geometric probability 12
5.5 The normal distribution 12
5.6 Transformation of random variables 12
5.7 Moment generating functions 13

6 More distributions **14**
6.1 Cauchy distribution 14
6.2 Gamma distribution 14
6.3 Beta distribution* 14
6.4 More on the normal distribution 14
6.5 Multivariate normal 14

7 Central limit theorem **15**

8 Summary of distributions **16**
8.1 Discrete distributions 16
8.2 Continuous distributions 16

0 Introduction

1 Classical probability

1.1 Classical probability

1.2 Counting

Theorem (Fundamental rule of counting). Suppose we have to make r multiple choices in sequence. There are m_1 possibilities for the first choice, m_2 possibilities for the second etc. Then the total number of choices is $m_1 \times m_2 \times \cdots m_r$.

1.3 Stirling's formula

Proposition. $\log n! \sim n \log n$

Theorem (Stirling's formula). As $n \to \infty$,

$$\log\left(\frac{n!e^n}{n^{n+\frac{1}{2}}}\right) = \log\sqrt{2\pi} + O\left(\frac{1}{n}\right)$$

Corollary.

$$n! \sim \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}$$

Proposition (non-examinable). We use the $1/12n$ term from the proof above to get a better approximation:

$$\sqrt{2\pi}n^{n+1/2}e^{-n+\frac{1}{12n+1}} \leq n! \leq \sqrt{2\pi}n^{n+1/2}e^{-n+\frac{1}{12n}}.$$

2 Axioms of probability

2.1 Axioms and definitions

Theorem.

(i) $\mathbb{P}(\emptyset) = 0$

(ii) $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

(iii) $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

(iv) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Theorem. If $A_1, A_2, \cdots$ is increasing or decreasing, then

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right).$$

2.2 Inequalities and formulae

Theorem (Boole's inequality)**.** For any $A_1, A_2, \cdots$,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Theorem (Inclusion-exclusion formula)**.**

$$\mathbb{P}\left(\bigcup_i^n A_i\right) = \sum_1^n \mathbb{P}(A_i) - \sum_{i_1<i_2} \mathbb{P}(A_{i_1} \cap A_{j_2}) + \sum_{i_1<i_2<i_3} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \cdots$$
$$+ (-1)^{n-1}\mathbb{P}(A_1 \cap \cdots \cap A_n).$$

Theorem (Bonferroni's inequalities)**.** For any events $A_1, A_2, \cdots, A_n$ and $1 \leq r \leq n$, if r is odd, then

$$\mathbb{P}\left(\bigcup_1^n A_i\right) \leq \sum_{i_1} \mathbb{P}(A_{i_1}) - \sum_{i_1<i_2} \mathbb{P}(A_{i_1}A_{i_2}) + \sum_{i_1<i_2<i_3} \mathbb{P}(A_{i_1}A_{i_2}A_{i_3}) + \cdots$$
$$+ \sum_{i_1<i_2<\cdots<i_r} \mathbb{P}(A_{i_1}A_{i_2}A_{i_3}\cdots A_{i_r}).$$

If r is even, then

$$\mathbb{P}\left(\bigcup_1^n A_i\right) \geq \sum_{i_1} \mathbb{P}(A_{i_1}) - \sum_{i_1<i_2} \mathbb{P}(A_{i_1}A_{i_2}) + \sum_{i_1<i_2<i_3} \mathbb{P}(A_{i_1}A_{i_2}A_{i_3}) + \cdots$$
$$- \sum_{i_1<i_2<\cdots<i_r} \mathbb{P}(A_{i_1}A_{i_2}A_{i_3}\cdots A_{i_r}).$$

2.3 Independence

Proposition. If A and B are independent, then A and B^C are independent.

2.4 Important discrete distributions

Theorem (Poisson approximation to binomial)**.** Suppose $n \to \infty$ and $p \to 0$ such that $np = \lambda$. Then

$$q_k = \binom{n}{k} p^k (1-p)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

2.5 Conditional probability

Theorem.

(i) $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$.

(ii) $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \mid B \cap C)\mathbb{P}(B \mid C)\mathbb{P}(C)$.

(iii) $\mathbb{P}(A \mid B \cap C) = \frac{\mathbb{P}(A \cap B \mid C)}{\mathbb{P}(B \mid C)}$.

(iv) The function $\mathbb{P}(\,\cdot \mid B)$ restricted to subsets of B is a probability function (or measure).

Proposition. If B_i is a partition of the sample space, and A is any event, then

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

Theorem (Bayes' formula)**.** Suppose B_i is a partition of the sample space, and A and B_i all have non-zero probability. Then for any B_i,

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}{\sum_j \mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}.$$

Note that the denominator is simply $\mathbb{P}(A)$ written in a fancy way.

3 Discrete random variables

3.1 Discrete random variables

Theorem.

(i) If $X \geq 0$, then $\mathbb{E}[X] \geq 0$.

(ii) If $X \geq 0$ and $\mathbb{E}[X] = 0$, then $\mathbb{P}(X = 0) = 1$.

(iii) If a and b are constants, then $\mathbb{E}[a + bX] = a + b\mathbb{E}[X]$.

(iv) If X and Y are random variables, then $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. This is true even if X and Y are not independent.

(v) $\mathbb{E}[X]$ is a constant that minimizes $\mathbb{E}[(X - c)^2]$ over c.

Theorem. For any random variables $X_1, X_2, \cdots X_n$, for which the following expectations exist,

$$\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i].$$

Theorem.

(i) $\operatorname{var} X \geq 0$. If $\operatorname{var} X = 0$, then $\mathbb{P}(X = \mathbb{E}[X]) = 1$.

(ii) $\operatorname{var}(a + bX) = b^2 \operatorname{var}(X)$. This can be proved by expanding the definition and using the linearity of the expected value.

(iii) $\operatorname{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, also proven by expanding the definition.

Proposition.

- $\mathbb{E}[I[A]] = \sum_\omega p(\omega) I[A](\omega) = \mathbb{P}(A)$.
- $I[A^C] = 1 - I[A]$.
- $I[A \cap B] = I[A]I[B]$.
- $I[A \cup B] = I[A] + I[B] - I[A]I[B]$.
- $I[A]^2 = I[A]$.

Theorem (Inclusion-exclusion formula)**.**

$$\mathbb{P}\left(\bigcup_i^n A_i\right) = \sum_1^n \mathbb{P}(A_i) - \sum_{i_1 < i_2} \mathbb{P}(A_{i_1} \cap A_{j_2}) + \sum_{i_1 < i_2 < i_3} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \cdots$$
$$+ (-1)^{n-1}\mathbb{P}(A_1 \cap \cdots \cap A_n).$$

Theorem. If $X_1, \cdots, X_n$ are independent random variables, and $f_1, \cdots, f_n$ are functions $\mathbb{R} \to \mathbb{R}$, then $f_1(X_1), \cdots, f_n(X_n)$ are independent random variables.

Theorem. If $X_1, \cdots, X_n$ are independent random variables and all the following expectations exists, then

$$\mathbb{E}\left[\prod X_i\right] = \prod \mathbb{E}[X_i].$$

Corollary. Let $X_1, \cdots X_n$ be independent random variables, and $f_1, f_2, \cdots f_n$ are functions $\mathbb{R} \to \mathbb{R}$. Then

$$\mathbb{E}\left[\prod f_i(x_i)\right] = \prod \mathbb{E}[f_i(x_i)].$$

Theorem. If $X_1, X_2, \cdots X_n$ are independent random variables, then

$$\operatorname{var}\left(\sum X_i\right) = \sum \operatorname{var}(X_i).$$

Corollary. Let $X_1, X_2, \cdots X_n$ be independent identically distributed random variables (iid rvs). Then

$$\operatorname{var}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\operatorname{var}(X_1).$$

3.2 Inequalities

Proposition. If f is differentiable and $f''(x) \geq 0$ for all $x \in (a, b)$, then it is convex. It is strictly convex if $f''(x) > 0$.

Theorem (Jensen's inequality)**.** If $f : (a, b) \to \mathbb{R}$ is convex, then

$$\sum_{i=1}^{n} p_i f(x_i) \geq f\left(\sum_{i=1}^{n} p_i x_i\right)$$

for all $p_1, p_2, \cdots, p_n$ such that $p_i \geq 0$ and $\sum p_i = 1$, and $x_i \in (a, b)$.

This says that $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$ (where $\mathbb{P}(X = x_i) = p_i$).

If f is strictly convex, then equalities hold only if all x_i are equal, i.e. X takes only one possible value.

Corollary (AM-GM inequality)**.** Given $x_1, \cdots, x_n$ positive reals, then

$$\left(\prod x_i\right)^{1/n} \leq \frac{1}{n}\sum x_i.$$

Theorem (Cauchy-Schwarz inequality)**.** For any two random variables X, Y,

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

Theorem (Markov inequality)**.** If X is a random variable with $\mathbb{E}|X| < \infty$ and $\varepsilon > 0$, then

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}|X|}{\varepsilon}.$$

Theorem (Chebyshev inequality)**.** If X is a random variable with $\mathbb{E}[X^2] < \infty$ and $\varepsilon > 0$, then

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}[X^2]}{\varepsilon^2}.$$

3.3 Weak law of large numbers

Theorem (Weak law of large numbers)**.** Let $X_1, X_2, \cdots$ be iid random variables, with mean μ and $\operatorname{var} \sigma^2$.

Let $S_n = \sum_{i=1}^n X_i$.

Then for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \to 0$$

as $n \to \infty$.

We say, $\frac{S_n}{n}$ tends to μ (in probability), or

$$\frac{S_n}{n} \to_p \mu.$$

Theorem (Strong law of large numbers)**.**

$$\mathbb{P}\left(\frac{S_n}{n} \to \mu \text{ as } n \to \infty\right) = 1.$$

We say

$$\frac{S_n}{n} \to_{\text{as}} \mu,$$

where "as" means "almost surely".

3.4 Multiple random variables

Proposition.

(i) $\operatorname{cov}(X, c) = 0$ for constant c.

(ii) $\operatorname{cov}(X + c, Y) = \operatorname{cov}(X, Y)$.

(iii) $\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X)$.

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

(v) $\operatorname{cov}(X, X) = \operatorname{var}(X)$.

(vi) $\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X, Y)$.

(vii) If X, Y are independent, $\operatorname{cov}(X, Y) = 0$.

Proposition. $|\operatorname{corr}(X, Y)| \leq 1$.

Theorem. If X and Y are independent, then

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X]$$

Theorem (Tower property of conditional expectation)**.**

$$\mathbb{E}_Y[\mathbb{E}_X[X \mid Y]] = \mathbb{E}_X[X],$$

where the subscripts indicate what variable the expectation is taken over.

3.5 Probability generating functions

Theorem. The distribution of X is uniquely determined by its probability generating function.

Theorem (Abel's lemma)**.**

$$\mathbb{E}[X] = \lim_{z \to 1} p'(z).$$

If $p'(z)$ is continuous, then simply $\mathbb{E}[X] = p'(1)$.

Theorem.

$$\mathbb{E}[X(X-1)] = \lim_{z \to 1} p''(z).$$

Theorem. Suppose $X_1, X_2, \cdots, X_n$ are independent random variables with pgfs $p_1, p_2, \cdots, p_n$. Then the pgf of $X_1 + X_2 + \cdots + X_n$ is $p_1(z)p_2(z)\cdots p_n(z)$.

4 Interesting problems

4.1 Branching processes

Theorem.

$$F_{n+1}(z) = F_n(F(z)) = F(F(F(\cdots F(z)\cdots)))) = F(F_n(z)).$$

Theorem. Suppose

$$\mathbb{E}[X_1] = \sum kp_k = \mu$$

and

$$\operatorname{var}(X_1) = \mathbb{E}[(X-\mu)^2] = \sum (k-\mu)^2 p_k < \infty.$$

Then

$$\mathbb{E}[X_n] = \mu^n, \quad \operatorname{var} X_n = \sigma^2\mu^{n-1}(1 + \mu + \mu^2 + \cdots + \mu^{n-1}).$$

Theorem. The probability of extinction q is the smallest root to the equation $q = F(q)$. Write $\mu = \mathbb{E}[X_1]$. Then if $\mu \leq 1$, then $q = 1$; if $\mu > 1$, then $q < 1$.

4.2 Random walk and gambler's ruin

5 Continuous random variables

5.1 Continuous random variables

Proposition. The exponential random variable is *memoryless*, i.e.

$$\mathbb{P}(X \geq x + z \mid X \geq x) = \mathbb{P}(X \geq z).$$

This means that, say if X measures the lifetime of a light bulb, knowing it has already lasted for 3 hours does not give any information about how much longer it will last.

Theorem. If X is a continuous random variable, then

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq x)\;\mathrm{d}x - \int_0^\infty \mathbb{P}(X \leq -x)\;\mathrm{d}x.$$

5.2 Stochastic ordering and inspection paradox

5.3 Jointly distributed random variables

Theorem. If X and Y are jointly continuous random variables, then they are individually continuous random variables.

Proposition. For independent continuous random variables X_i,

(i) $\mathbb{E}[\prod X_i] = \prod \mathbb{E}[X_i]$

(ii) $\mathrm{var}(\sum X_i) = \sum \mathrm{var}(X_i)$

5.4 Geometric probability

5.5 The normal distribution

Proposition.

$$\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x - \mu)^2}\;\mathrm{d}x = 1.$$

Proposition. $\mathbb{E}[X] = \mu$.

Proposition. $\mathrm{var}(X) = \sigma^2$.

5.6 Transformation of random variables

Theorem. If X is a continuous random variable with a pdf $f(x)$, and $h(x)$ is a continuous, strictly increasing function with $h^{-1}(x)$ differentiable, then $Y = h(X)$ is a random variable with pdf

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

Theorem. Let $U \sim U[0, 1]$. For any strictly increasing distribution function F, the random variable $X = F^{-1}U$ has distribution function F.

Proposition. $(Y_1, \cdots, Y_n)$ has density

$$g(y_1, \cdots, y_n) = f(s_1(y_1, \cdots, y_n), \cdots s_n(y_1, \cdots, y_n))|J|$$

if $(y_1, \cdots, y_n) \in S$, 0 otherwise.

5.7 Moment generating functions

Theorem. The mgf determines the distribution of X provided $m(\theta)$ is finite for all θ in some interval containing the origin.

Theorem. The rth moment X is the coefficient of $\frac{\theta^r}{r!}$ in the power series expansion of $m(\theta)$, and is

$$\mathbb{E}[X^r] = \frac{\mathrm{d}^n}{\mathrm{d}\theta^n} m(\theta)\bigg|_{\theta=0} = m^{(n)}(0).$$

Theorem. If X and Y are independent random variables with moment generating functions $m_X(\theta), m_Y(\theta)$, then $X+Y$ has mgf $m_{X+Y}(\theta) = m_X(\theta)m_Y(\theta)$.

6 More distributions

6.1 Cauchy distribution

Proposition. The mean of the Cauchy distribution is undefined, while $\mathbb{E}[X^2] = \infty$.

6.2 Gamma distribution

6.3 Beta distribution*

6.4 More on the normal distribution

Proposition. The moment generating function of $N(\mu, \sigma^2)$ is

$$\mathbb{E}[e^{\theta X}] = \exp\left(\theta\mu + \frac{1}{2}\theta^2\sigma^2\right).$$

Theorem. Suppose X, Y are independent random variables with $X \sim N(\mu_1, \sigma_1^2)$, and $Y \sim (\mu_2, \sigma_2^2)$. Then

(i) $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

(ii) $aX \sim N(a\mu_1, a^2\sigma_1^2)$.

6.5 Multivariate normal

7 Central limit theorem

Theorem (Central limit theorem)**.** Let $X_1, X_2, \cdots$ be iid random variables with $\mathbb{E}[X_i] = \mu$, $\mathrm{var}(X_i) = \sigma^2 < \infty$. Define

$$S_n = X_1 + \cdots + X_n.$$

Then for all finite intervals (a, b),

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\;\mathrm{d}t.$$

Note that the final term is the pdf of a standard normal. We say

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to_D N(0, 1).$$

Theorem (Continuity theorem)**.** If the random variables $X_1, X_2, \cdots$ have mgf's $m_1(\theta), m_2(\theta), \cdots$ and $m_n(\theta) \to m(\theta)$ as $n \to \infty$ for all θ, then $X_n \to_D$ the random variable with mgf $m(\theta)$.

8 Summary of distributions

8.1 Discrete distributions

8.2 Continuous distributions