ST402
Principles and Methods of Statistical Practice

Weeks 1 to 5: Distribution Theory
Michaelmas Term
2006-2007

Teacher: Martin Knott B607

Course Aims and Objectives

This part of the course aims to convey a thorough understanding of probability and distribution theory. A range of methods, related to distributions of practical importance, are taught. The course provides a strong theoretical foundation for MSc Statistics students.

The following list gives you an idea of the sort of things you will be able to do by the end of the first five weeks – it does not come close to covering everything. By the end of the first part of the course you should:

- be able to work out probabilities associated with simple (and not so simple) experiments,
- know the distinction between a random variable and an instance of a random variable,
- for any given distribution, be able to select suitable methods and use them to work out moments,
- be familiar with a large number of distributions,
- understand relationships between variables, conditioning, independence and correlation,
- feel at ease with joint distributions and conditional distributions,
- be able to put together all the theory and techniques you have learnt to solve practical problems.

Introductory Material

Structure of the Course

Each student should usually attend **three hours per week** in the Michaelmas term. This is made up of:

1. two hour lecture/problem class Tuesday 15:00-17:00 in S300,
2. one problem class Friday 09:00-10:00 in S221,

The course in assessed by a two hour written exam in the summer term.

Books

There are a large number of books that cover at least part of the material in the course. Finding a useful book is partly a question of personal taste. I suggest you look at what is available in the library and find a text that covers the material in a way that you find appealing and intelligible. Reproduced below is the reading list along with some additional texts that may be worth looking at.

Main texts

- Casella, G. and R. L. Berger. *Statistical inference.* [QA276 C33]
- Larson, H. J. *Introduction to probability theory and statistical inference.* [QA273.A5 L33]
- Meyer, P. L. *Introductory probability and statistical applications.* [QA273.A5 M61]

Books covering similar material to main texts

- Hogg, R. V. and A. T. Craig. *Introduction to mathematical statistics.* [QA276.A2 H71]
- Freund, J. E. *Mathematical statistics* [QA276 F88]
- Hogg, R. V. and E. A. Tanis. *Probability and statistical inference.* [QA273 H71]
- Mood, A. M., F. A. Graybill and D. C. Boes. *Introduction to the theory of statistics.* [QA276.A2 M81]
- Bartoszyński, R. and M. Niewiadomska-Bugaj. *Probability and Statistical Inference.* [QA273 B29]

Worth a look

- Cox, D. R. and D. V. Hinkley. *Theoretical statistics.* [QA276.A2 C87]
  (Not great to learn from but a good reference source.)
- Stuart, A. and J. K. Ord. *Kendall's Advanced Theory of Statistics* **1**, *Distribution theory.* [QA276.A2 K31]
  (A bit arcane but covers just about everything.)
- Grimmett, G. R. and D. R. Stirzaker. *Probability and random processes.* [QA273 G86]
  (Very succinct and very precise. One for those who like it mathematical.)
- Johnson, N. L. and S. Kotz (some volumes also with N. Balakrishnan) *Discrete Univariate Distributions*, *Continuous Univariate Distributions*, etc. [QA273.6 J61]
  (The place to go if you have a particular question about a distribution.)
- Larsen R. J. and M. L. Marx. *An Introduction to Mathematical Statistics and its Applications.* [QA276 L33]
  (Good introduction to probability. Lots of examples)

Pre-requisites

The background of students on this course is very varied. The course does not assume much knowledge of probability or statistics but we will move very quickly through the material. Below is a, far from exhaustive, list of mathematical tool that we will be using.

- **Sets**: union, intersection, complement.
- **Series**: arithmetic, geometric, Taylor.
- **Integration**: standard integrals, integration by parts.
- **Differentiation**: standard differentials, product rule, function of a function rule.

A Guide to Content

The following is a guide to the content of the course rather than a definitive syllabus. Throughout the course examples, with varying degrees of realism, will be used to illustrate the theory. The material that goes into the exam will be determined by what is covered in the lectures.

1. **Events and their Probabilities**: Interpretations of probability / Sample space: experiments, outcomes / Events / Probability Space: probability measure, permutations and combinations / Conditional Probability: partitions, law of total probability, Bayes' theorem, independence.

2. **Random Variables and their Distributions**: Random Variables: cumulative distribution function / Discrete Random Variables: probability mass function / Continuous Random Variables: probability density function, Lebesgue-Stieltjes integral / Support and indicators / Mean / Variance / Expectation Operator: Markov, and Chebyshev inequalities / Moments: generating functions / Application: functions of random variables and simulation.

3. **Multivariate Distributions**: Joint Distributions: bivariate distributions, marginal distributions, expectations of functions of random variables, multivariate generalisation / Conditional Distribution: conditional mass function, conditional density function, conditional expectation, conditional variance / Dependence: independence, covariance, correlation / Application: sums of independent random variables, random sums / Application: multivariate normal density, fundamentals of linear prediction / Joint Moments: joint generating function, conditional generating function / Application: laws of large numbers and the central limit theorem / Transformations of random variables.

We will encounter many different distributions during the course. These include:

- **Discrete Distributions**: degenerate, Bernoulli, binomial, geometric, negative binomial, Poisson, hypergeometric, uniform.

- **Continuous Distributions**: uniform, normal, exponential, chi-square, t, F, gamma, beta.

Course Handouts

The printed notes given out in lectures are intended as a **reference**. They provide key equations and course structure but are not intended as lecture notes or as a replacement for copying stuff down from the board. I have made some corrections to these notes but **there will still be errors!** If you spot something that looks strange or wrong, please let me know. The notes, questions and some solutions are available in the public folders under `Statistics/ST402`, and in the course handouts boxes in the Statistics department (6th floor of Columbia House). If you are missing any handouts, please collect them from one of these three places.

1 Events and their Probabilities

1.1 Sample Space

1. **Experiment**: a procedure, that we usually think of as being repeatable any number of times, which has a well-defined set of possible outcomes.

2. **Sample outcome**: a potential eventuality of the experiment. The notation ω is used for an outcome.

3. **Sample space**: the set of all possible outcomes. The notation Ω is used for the sample space of an experiment. An outcome ω is a member of the sample space Ω, that is, $\omega \in \Omega$.

4. **A simple experiment**: A fair six-sided die is thrown twice. The outcomes are pairs of numbers between 1 and 6. For example, $(3, 5)$ denotes a 3 on the first throw and 5 on the second. The sample space is given by $\Omega = \{(i, j) \ : \ i = 1, \ldots, 6, \ j = 1, \ldots, 6\}$. In this example the sample space is finite so can be written out in full:

$$\left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}.$$

1.2 Events

For any experiment, the events form a collection of subsets of Ω which we denote $\mathcal{F}$. The collection $\mathcal{F}$ has the following properties;

i $\emptyset \in \mathcal{F}$,

ii if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$,

iii if $A_1, A_2, \ldots, \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

[Aside: Any collection of subsets with these properties is known as a sigma field or a sigma algebra. It is also, less helpfully, sometimes called a Borel field,] We associate the pair $(\Omega, \mathcal{F})$ with our experiment; the sample space consisting of all of the possible outcomes, and a collection of events which are just subsets of the outcomes.

1.3 Probability Space

1. **Probability measure**: P on $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \longrightarrow [0, 1]$ satisfying;

   i $P(A) \geq 0$,

   ii $P(\Omega) = 1$,

   iii if $A_1, A_2, \ldots,$ is a collection of mutually exclusive members of $\mathcal{F}$ then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

   We can associate a probability space $(\Omega, \mathcal{F}, P)$ with any experiment.

2. **Properties of probability measures**

   (a) $P(A^c) = 1 - P(A)$.

   (b) If $A \subseteq B$ then $P(B) = P(A) + P(B \backslash A) \geq P(A)$.

   (c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

   (d) More generally if $A_1, \ldots, A_n$ are events then

   $$P(\bigcup_{i=1}^{n} A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n+1} P(A_1 \cap \ldots \cap A_n).$$

   (e) Boole inequality: $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$.

3. **Probability from counting outcomes**

   (a) Multiplication rule for counting ordered sequences: Operation A_i can be performed in n_i different ways for $i = 1, \ldots, k$. The ordered sequence (operation A_1, operation A_2, ..., operation A_k) can be performed in $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ ways. We write this product as $\prod_{i=1}^{k} n_i$.

   (b) Permutations and combinations:

      i. Permutations of k out of n distinct objects: The number of permutations of length k that can be formed from n distinct elements (not allowing repetition) is denoted ${}_nP_k$, where

      $${}_nP_k = \frac{n!}{(n-k)!}.$$

      ii. Permutations of n objects that are not all distinct: The numbers of ways of arranging n objects where n_1 are of type 1, n_2 are of type 2, ..., n_r are of type r is

      $$\frac{n!}{n_1! n_2! \ldots n_r!},$$

      where $\sum_{i=1}^{r} n_i = n$.

      iii. Combinations of k out of n distinct objects: The number of combinations of length k that can be formed from n distinct elements is denoted ${}_nC_k$, where

      $${}_nC_k = \frac{n!}{k!(n-k)!}.$$

1.4 Conditional Probability

Let A and B be events with $P(B) > 0$. The conditional probability of A given B is the probability that A will occur give that B has occurred;

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

1. **Multiplication rule for intersections**: Let $A_1, \ldots, A_n$ be a set of events defined on Ω,

   $$P(\bigcap_{i=1}^{n} A_i) = \prod_{j=1}^{n} P(A_j | \bigcap_{i=0}^{j-1} A_i),$$

   where we define $A_0 = \Omega$.

2. **Partitions**: A set of events $B_1, \ldots, B_n$ defined on Ω is a partition if
   (a) Exhaustive: $\bigcup_{i=1}^n B_i = \Omega$,
   (b) Mutually exclusive: $B_i \cap B_j = \emptyset$ for $i \neq j$,
   (c) Non-zero probability: $P(B_i) > 0$ for $i = 1, \ldots, n$.

3. **Law of total probability**: If $B_1, \ldots, B_n$ is a partition of Ω and A is any other event defined on Ω, then
$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

4. **Bayes' theorem**: If $B_1, \ldots, B_n$ is a partition of Ω and A is any other event defined on Ω, then
$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)}.$$

5. **Independence**
   (a) Two events A and B are independent ($A \perp B$) iff $P(A \cap B) = P(A)P(B)$.
   (b) Events $A_1, \ldots, A_n$ (mutually) independent if every subset is a subset of independent events and $P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2)\ldots P(A_n)$ (note – this definition is recursive).

6. **Some properties of independent events**
   (a) If $P(A) > 0$ then $P(B|A) = P(B) \iff A \perp B$.
   If $P(B) > 0$ then $P(A|B) = P(A) \iff A \perp B$.
   (b) If $A \perp B$ then $A^c \perp B^c$, $A^c \perp B$ and $A \perp B^c$.

Elementary Set Theory

1. **Notation**

	Set terminology	Probability terminology
Ω	Collection of objects	Sample space
ω	Member of Ω	Outcome
A	Subset of Ω	Event - an outcome in A occurs
A^c	Complement	Event - an outcome not in A occurs
$A \cap B$	Intersection	Event - an outcome in A and B occurs
$A \cup B$	Union	Event - an outcome in A and/or B occurs
$A \backslash B$	Difference	Event - an outcome in A but not in B occurs
$A \subseteq B$	Inclusion	if outcome in A also in B
$\emptyset$	Empty set	Impossible event
Ω	Whole space	Certain event

2. **Properties of Intersections and Unions**

Commutative	$A \cap B = B \cap A$	$A \cup B = B \cup A$
Associative	$A \cap (B \cap C) = (A \cap B) \cap C$	$A \cup (B \cup C) = (A \cup B) \cup C$
Distributive	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
With whole space	$A \cap \Omega = A$	$A \cup \Omega = \Omega$
With empty set	$A \cap \emptyset = \emptyset$	$A \cup \emptyset = A$

3. **Complement of** A: is $A^c = \Omega \backslash A$, that is, $\omega \in A^c \Longleftrightarrow \omega \notin A$.

   (a) $(A^c)^c = A$.

   (b) $A \cap A^c = \emptyset$.

   (c) $A \cup A^c = \Omega$.

   (d) $(A \cup B)^c = A^c \cap B^c$.

4. **Collection of subsets of** Ω: say $\{A_1, \ldots, A_n\}$, where n is not necessarily finite.

   (a) Mutually exclusive: if $A_i \cap A_j = \emptyset$ for any $i \neq j$, so $A_1, \ldots, A_n$ are disjoint sets.

   (b) Exhaustive: if $\bigcup_{i=1}^n A_i = \Omega$.

   (c) De Morgan's theorem (a generalization of 3d above): $(\bigcup_{i=1}^n A_i)^c = \bigcap_{i=1}^n A_i^c$.

2 Random Variables and their Distributions

2.1 Random Variables

A random variable is a function $X : \Omega \longrightarrow \mathbb{R}$ with the property that, if $A_x = \{\omega \in \Omega : X(\omega) \leq x\}$ then $A_x \in \mathcal{F}$ for all $x \in \mathbb{R}$. Thus, A_x is an event, for every real-valued x.

1. **(Cumulative) distribution function**: of a random variable X is the function $F : \mathbb{R} \longrightarrow [0, 1]$ given by $F(x) = P(X \leq x)$.
2. **Properties of distribution functions**
   (a) Limits: $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.
   (b) Non-decreasing: if $x < y$ then $F(x) \leq F(y)$.
   (c) Right continuous: $\lim_{h \downarrow 0} F(x + h) = F(x)$.
3. **Probabilities from distribution functions**
   (a) $P(X > x) = 1 - F(x)$.
   (b) $P(x < X \leq y) = F(y) - F(x)$.
   (c) $P(X < x) = \lim_{h \downarrow 0} F(x - h) = F(x-)$.
   (d) $P(X = x) = F(x) - F(x-)$.

2.2 Discrete Random Variables

A random variable X is *discrete* if it only takes values in some countable subset $\{x_1, x_2, \ldots\}$ of $\mathbb{R}$.

1. **Probability mass function**: of a discrete random variable X is the function $f : \mathbb{R} \longrightarrow [0, 1]$ given by $f(x) = P(X = x)$.
2. **Properties of mass functions**
   (a) $f(x) = F(x) - F(x-)$.
   (b) $\sum_{i:x_i \leq x} f(x_i) = F(x)$.
   (c) $\sum_i f(x_i) = 1$.
   (d) $f(x) = 0$ if $x \notin \{x_1, x_2, \ldots\}$.

2.3 Continuous Random Variables

A random variable X is *continuous* if its distribution function can be expressed as

$$F(x) = \int_{-\infty}^{x} f(u)du \qquad \text{for } x \in \mathbb{R},$$

for some integrable function $f : \mathbb{R} \longrightarrow [0, \infty)$. The function f is called the **(probability) density function** of X.

1. **Properties of continuous random variables**

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

   (b) $\int_{-\infty}^{\infty} f(x)dx = 1$.

   (c) $\int_a^b f(x)dx = P(a < X \leq b)$.

2. **Unified notation**: known as Lebesgue-Stieltjes integral. For a random variable X;

$$P(a < X \leq b) = \int_a^b dF(x) = \begin{cases} \sum_{i: a < x_i \leq b} f(x_i), & \text{if } X \text{ discrete,} \\ \int_a^b f(x)dx, & \text{if } X \text{ continuous.} \end{cases}$$

2.4 Support and indicator functions

1. **Support**: The support of a non-negative real-valued function f is the subset of $\mathbb{R}$ on which f takes non-zero values, that is, $\{x \in \mathbb{R} : f(x) > 0\}$.

2. **Indicator function**: The indicator function for a set $S \subset \mathbb{R}$ is $I_S(x)$ where

$$I_S(x) = \begin{cases} 1, & x \in S \\ 0, & \text{otherwise.} \end{cases}$$

   The set S often takes the form of an interval such as $[0, 1]$ or $[0, \infty)$.

2.5 Mean

If X is a random variable with $\int_{-\infty}^{\infty} |x| dF(x) < \infty$, the mean of X is defined as

$$\mu = E(X) = \int_{-\infty}^{\infty} x dF(x) = \begin{cases} \sum_i x_i f(x_i), & \text{if } X \text{ discrete,} \\ \int_{-\infty}^{\infty} x f(x)dx, & \text{if } X \text{ continuous.} \end{cases}$$

2.6 Variance

If X is a random variable, the variance of X is defined as

$$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 dF(x) = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i), & \text{if } X \text{ discrete,} \\ \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx, & \text{if } X \text{ continuous.} \end{cases}$$

The standard deviation is defined as $\sigma = \sqrt{\text{Var}(X)}$.

2.7 Expectation operator

For any function $g : \mathbb{R} \longrightarrow \mathbb{R}$, with $\int_{-\infty}^{\infty} |g(x)| dF(x) < \infty$, the *expectation*, or *expected value* of $g(X)$ is defined as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) dF(x) = \begin{cases} \sum_i g(x_i) f(x_i), & \text{if } X \text{ discrete,} \\ \int_{-\infty}^{\infty} g(x) f(x)dx, & \text{if } X \text{ continuous.} \end{cases}$$

1. **Properties of expectation**: for constant a and random variables X and Y with defined expected values.
   (a) $E(a) = a$.
   (b) $E(aX) = aE(X)$.
   (c) $E(X + Y) = E(X) + E(Y)$.
   (d) If $X \geq Y$ then $E(X) \geq E(Y)$.
2. **Inequalities involving expectation**
   (a) Markov inequality: Let Y be a random variable with $Y \geq 0$ and $E(Y)$ defined, then $P(Y \geq a) \leq E(Y)/a$ for any constant $a > 0$.
   (b) Chebyshev inequality: Let X be a random variable with E(X) defined, then $P(|X - E(X)| \geq a) \leq \sigma^2/a^2$ for any constant $a > 0$.
   (c) Jensen inequality: If X is a random variable with $E(X)$ defined and g is a convex function with $E(g(X))$ defined, then $E(g(X)) \geq g(E(X))$.

2.8 Moments

If r is a positive integer then the r^{th} *moment*, m_r, of X is

$$m_r = E(X^r).$$

The r^{th} *central moment*, μ_r is

$$\mu_r = E((X - m_1)^r).$$

1. **Properties of moments**
   (a) $m_1 = E(X) = \mu$ (mean); $\mu_1 = 0$.
   (b) $\mu_2 = E(X^2) - E(X)^2 = \mathrm{Var}(X) = \sigma^2$ (variance).
   (c) Coefficient of skewness: $\gamma_1 = E[(X - \mu)^3]/\sigma^3 = \mu_3/\mu_2^{\frac{3}{2}}$.
   (d) Coefficient of kurtosis: $\gamma_2 = (E[(X - \mu)^4]/\sigma^4) - 3 = (\mu_4/\mu_2^2) - 3$.
2. **Moment generating function**: of a random variable X is a function $M : \mathbb{R} \longrightarrow [0, \infty)$ given by

$$M(t) = E(e^{tX}),$$

   where $M(t) < \infty$ for $|t| < h$ and some $h > 0$.
   (a) Evaluation via integration:

$$M(t) = \int_{-\infty}^{\infty} e^{tx} dF(x) = \begin{cases} \sum_i e^{tx_i} f(x_i), & \text{if } X \text{ discrete,} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx, & \text{if } X \text{ continuous.} \end{cases}$$

   (b) Taylor expansion:

$$M(t) = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \ldots + \frac{t^r}{r!}E(X^r) + \ldots = \sum_{j=0}^{\infty} \frac{E(X^j)}{j!} t^j.$$

(c) Derivatives at zero:

$$M^{(r)}(0) = \frac{d^r}{dt^r} M(t)\Big|_{t=0} = E(X^r) = m_r.$$

(d) Uniqueness: If $M_X(t) = M_Y(t)$ for all $|t| < h$ and some $h > 0$, then $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$.

3. **Cumulant generating function**: of a random variable X with moment generating function $M(t)$, is defined as

$$K(t) = \log M(t),$$

The r^{th} *cumulant*, κ_r, is the coefficient of $t^r/r!$ in the expansion of the cumulant generating function $K(t)$;

$$\kappa_r = K^{(r)}(0) = \frac{d^r}{dt^r} K(t)\Big|_{t=0}.$$

(a) $\kappa_1 = m_1 = \mu$ (mean, first moment).

(b) $\kappa_2 = \mu_2 = \sigma^2$ (variance, second central moment).

(c) $\kappa_3 = \mu_3$ (third central moment).

(d) $\kappa_4 + 3\kappa_2^2 = \mu_4$ (forth central moment).

4. **Probability generating function**: of a discrete random variable X, taking non-negative integer values, is defined as

$$G(t) = E(t^X) = \sum_{i=0}^{\infty} t^i P(X = i).$$

(a) $P(X = 0) = G(0)$.

(b) $E(X) = G'(1)$.

(c) $E(X(X-1)\dots(X-k+1)) = G^{(k)}(1)$.

5. **Characteristic function**: of a random variable X is the function $\phi : \mathbb{R} \longrightarrow \mathbb{C}$ defined by

$$\phi(t) = E(e^{itX})$$

where $i = \sqrt{-1}$. Note that $E(X^k) = i^{-k}\phi^{(k)}(0)$.

2.9 Functions of random variables

If X is a random variable with distribution function F_X and g is a real-valued function, then $Y = g(X)$ is a random variable.

- If g is strictly increasing $F_Y(y) = F_X(g^{-1}(y))$.
- If g is strictly decreasing $F_Y(y) = 1 - F_X(g^{-1}(y))$.

Discrete Distributions

1. **Degenerate**: all probability concentrated at a single point.
   - $f(x) = 1 \quad$ for $x = a$.
   - $M(t) = e^{at}, \quad K(t) = at$.
   - $\mu = a, \quad \sigma^2 = 0$.

2. **Bernoulli**: trials with two possible outcomes, here labeled $X = 0$ and $X = 1$. Equivalent to a binomial distribution with $n = 1$.
   - $f(x) = p^x(1-p)^{1-x} \quad$ for $x = 0,\ 1$.
   - $M(t) = 1 - p + pe^t, \quad K(t) = \log(1 - p + pe^t)$.
   - $\mu = p, \quad \sigma^2 = p(1-p)$.

3. **Binomial**: number of successes in n independent Bernoulli trials with probability of success p. Notation $\mathrm{Bin}(n, p)$.
   - $f(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \quad$ for $x = 0,\ 1, \ldots,\ n$.
   - $M(t) = (1 - p + pe^t)^n, \quad K(t) = n\log(1 - p + pe^t)$.
   - $\mu = np, \quad \sigma^2 = np(1-p)$.

4. **Geometric**: number of trials before first success in repeated independent Bernoullis. Equivalent to a negative binomial with $r = 1$.
   - $f(x) = (1-p)^{x-1}p \quad$ for $x = 1,\ 2, \ldots$.
   - $M(t) = \frac{pe^t}{1-(1-p)e^t}, \quad K(t) = -\log\{(1 - \frac{1}{p}) + \frac{1}{p}e^{-t}\} \quad$ for $|t| < -\log(1-p)$.
   - $\mu = \frac{1}{p}, \quad \sigma^2 = \frac{1-p}{p^2}$.

5. **Negative Binomial**: number of trials before r^{th} success in repeated independent Bernoullis.
   - $f(x) = \binom{x-1}{r-1}p^r(1-p)^{x-r} \quad$ for $x = r,\ r+1, \ldots$.
   - $M(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^r, \quad K(t) = -r\log\{(1 - \frac{1}{p}) + \frac{1}{p}e^{-t}\} \quad$ for $|t| < -\log(1-p)$.
   - $\mu = \frac{r}{p}, \quad \sigma^2 = \frac{r(1-p)}{p^2}$.

6. **Poisson**: often used for the number of events which occur in an interval of time. Notation $\mathrm{Pois}(\lambda)$.
   - $f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad$ for $x = 0,\ 1, \ldots$.
   - $M(t) = e^{\lambda(e^t-1)}, \quad K(t) = \lambda(e^t - 1)$.
   - $\mu = \lambda, \quad \sigma^2 = \lambda$.

7. **Hypergeometric**: number of type 1 chosen when selected n without replacement from urn containing a total of N made up of N_1 of type 1 and N_2 of type 2.
   - $f(x) = \binom{N_1}{x}\binom{N_2}{n-x}/\binom{N}{n} \quad$ for $x \in \{0, \ldots, n\} \cap \{n - N_2, \ldots, N_1\}$.
   - $\mu = n\frac{N_1}{N}, \quad \sigma^2 = n\frac{N_1}{N}\frac{N_2}{N}\frac{N-n}{N-1}$.

8. **Uniform**: choosing from m with equal probability.
   - $f(x) = \frac{1}{m} \quad$ for $x = 1,\ 2, \ldots, m$.
   - $\mu = \frac{m+1}{2}, \quad \sigma^2 = \frac{m^2-1}{12}$.

Continuous Distributions

1. **Uniform**: random number chosen from a given closed interval $[a,\ b]$. Notation $U(a,b)$.
   - $f(x) = \frac{1}{b-a}$ for $a \le x \le b$.
   - $M(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}$ for $t \neq 0$ and $M(0) = 1$.
   - $\mu = \frac{a+b}{2}$, $\sigma^2 = \frac{(b-a)^2}{12}$.
2. **Normal**: characterized by first two moments. Notation $N(\mu, \sigma^2)$.
   If $X \sim N(\mu, \sigma^2)$ then $(X-\mu)/\sigma = Z \sim N(0,1)$, a standard normal distribution.
   A sequence of r independent normals is written $X_j \sim iN(\mu, \sigma^2)$ for $j = 1, \ldots, r$.
   - $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$ for $-\infty < x < \infty$.
   - $M(t) = e^{\mu t + \sigma^2 t^2/2}$.
   - $E(X) = \mu$, $\text{Var}(X) = \sigma^2$.
3. **Exponential**: waiting time between events when observing a Poisson process. Notation $\text{Exp}(\theta)$ (somewhat ambiguous). Equivalent to a gamma distribution with $\alpha = 1$.
   - $f(x) = \theta e^{-\theta x}$ for $0 \le x < \infty$.
   - $M(t) = \frac{\theta}{\theta - t}$ for $t < \theta$.
   - $\mu = 1/\theta$, $\sigma^2 = 1/\theta^2$.
4. **Chi-square**: If $Z_j \sim iN(0,1)$ then $X = \sum_{j=1}^{r} Z_j$ has a chi-square distribution on r degrees of freedom. Notation χ^2_r or $\chi^2(r)$. Equivalent to a gamma distribution with $\alpha = r/2$ and $\theta = 1/2$.
   - $f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2}$ for $0 \le x < \infty$.
   - $M(t) = \frac{1}{(1-2t)^{r/2}}$ for $t < 1/2$.
   - $\mu = r$, $\sigma^2 = 2r$.
5. **Gamma**: characterized by parameters $\alpha > 0$ and θ. The gamma function is defined by
   $$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy, \quad \text{for } t > 0.$$
   Properties of the gamma function $\Gamma(t) = (t-1)\Gamma(t-1)$ for $t > 1$ and $\Gamma(n) = (n-1)!$ for positive integer n. Notation for the gamma distribution; $\text{Gamma}(\alpha, \theta)$ or $G(\alpha, \theta)$.
   - $f(x) = \frac{1}{\Gamma(\alpha)} \theta^\alpha x^{\alpha-1} e^{-\theta x}$ for $0 \le x < \infty$.
   - $M(t) = \frac{1}{(1-t/\theta)^\alpha}$ for $t < \theta$.
   - $\mu = \alpha/\theta$, $\sigma^2 = \alpha/\theta^2$.
6. **Beta**: characterized by parameters $p > 0$ and $q > 0$. The beta function is defined by
   $$B(p,q) = \int_0^1 y^{p-1}(1-y)^{q-1} dy = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$
   Notation $\text{Beta}(p,q)$.
   - $f(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}$ for $0 \le x \le 1$.
   - $\mu = \frac{p}{p+q}$, $\sigma^2 = \frac{pq}{(p+q+1)(p+q)^2}$.

3 Multivariate Distributions

3.1 Joint Distributions

1. **Joint distribution function**: If $X_1, \dots, X_n$ are random variables, the joint (cumulative) distribution function is

$$F(x_1, x_2, \dots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n).$$

   We sometimes use the notation $F_{X_1,\dots,X_n}(x_1, \dots, x_n)$ when we need to distinguish the joint distribution function of $X_1, \dots, X_n$ from other distribution functions.

2. **Bivariate distribution function**: For two random variable X and Y the joint distribution functions is

$$F(x, y) = P(X \le x, Y \le y).$$

   We sometimes use the notation $F_{X,Y}(x, y)$ for the joint distribution function of X and Y.

   (a) $F(-\infty, y) = \lim_{x\to-\infty} F(x, y) = 0$,
   $F(x, -\infty) = \lim_{y\to-\infty} F(x, y) = 0$,
   $F(\infty, \infty) = \lim_{x\to\infty,y\to\infty} F(x, y) = 1$.

   (b) $P(x_1 < X \le x_2, y_1 < Y \le y_2) = F(x_2, y_2) - F(x_1, y_2) - [F(x_2, y_1) - F(x_1, y_1)]$.

   (c) Right continuous in x: $\lim_{h\downarrow 0} F(x + h, y) = F(x, y)$,
   Right continuous in y: $\lim_{h\downarrow 0} F(x, y + h) = F(x, y)$.

3. **Marginal distribution functions**: If $F_{X,Y}$ is the joint distribution function of X and Y then the marginal distribution functions are given by

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y) = F_{X,Y}(x, \infty),$$
$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x, y) = F_{X,Y}(\infty, y).$$

4. **Discrete bivariate distributions**: for discrete random variables X and Y.

   (a) Joint mass function: $f(x, y) = f_{X,Y}(x, y) = P(X = x, Y = y)$.

   (b) Probabilities: $P(x_1 < X \le x_2, y_1 < Y \le y_2) = \sum_{x_1<x\le x_2} \sum_{y_1<y\le y_2} f(x, y)$.

   (c) Marginal mass functions: $f_X(x) = \sum_y f_{X,Y}(x, y)$,
   $f_Y(y) = \sum_x f_{X,Y}(x, y)$.

5. **Continuous bivariate distributions**: for jointly continuous random variables X and Y.

   (a) Joint density function: is a positive real-valued function f such that

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v)dudv \quad \text{for each } x, y \in \mathbb{R}.$$

   (b) Probabilities: $P(x_1 < X \le x_2, y_1 < Y \le y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x, y)dxdy$.

   (c) Marginal density functions: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$,
   $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$,

6. **Expectation of a function of two variables**: If g is a well-behaved, real-valued function of two variables ($g : \mathbb{R}^2 \longrightarrow \mathbb{R}$) and X and Y are random variable with joint mass/density function $f_{X,Y}$ then

$$E[g(X,Y)] = \begin{cases} \sum_y \sum_x g(x,y) f_{X,Y}(x,y), & \text{discrete case,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy, & \text{continuous case.} \end{cases}$$

7. **Multivariate generalisations**: for random variables $X_1, \ldots, X_n$ and for $j = 1, \ldots, n$.

   (a) Marginal distribution: $F_{X_j}(x_j) = F_{X_1,\ldots,X_n}(\infty, \ldots, \infty, x_j, \infty, \ldots, \infty)$

   (b) Marginal mass/density

$$f_{X_j}(x_j) = \begin{cases} \sum_{x_1} \ldots \sum_{x_{j-1}} \sum_{x_{j+1}} \ldots \sum_{x_n} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n), & \text{discrete case,} \\ \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_n, & \text{continuous case.} \end{cases}$$

   (c) Expectation of a function of n variables:

$$E[g(X_1, \ldots, X_n)] = \begin{cases} \sum_{x_1} \ldots \sum_{x_n} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n), & \text{discrete,} \\ \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n, & \text{continuous.} \end{cases}$$

3.2 Conditional Distributions

1. **Discrete conditional distributions**: X and Y discrete random variables with $P(X = x) > 0$.

   (a) Conditional distribution function of Y given $X = x$: $F_{Y|X}(y|x) = P(Y \le y | X = x)$.

   (b) Conditional mass function of Y given $X = x$: $f_{Y|X}(y|x) = P(Y = y|X = x) = f_{X,Y}(x,y)/f_X(x)$.

   (c) Relationship between distribution and mass functions: $F_{Y|X}(y|x) = \sum_{y_i \le y} f_{Y|X}(y_i|x)$.

2. **Continuous conditional distributions**: X and Y jointly continuous random variables with $f_X(x) > 0$.

   (a) Conditional distribution function of Y given $X = x$: $F_{Y|X}(y|x) = \int_{-\infty}^{y} (f_{X,Y}(x,v)/f_X(x))dv$.

   (b) Conditional density function of Y given $X = x$: $f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x)$.

3. **Conditional, joint and marginal densities**: given $f_X(x) > 0$ and (for (d)) $f_Y(y) > 0$.

   (a) Conditional mass/density

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{f_{X,Y}(x,y)}{\sum_y f_{X,Y}(x,y)}, & \text{discrete case,} \\ \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy}, & \text{continuous case.} \end{cases}$$

   (b) Joint mass/density: $f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x)$.

   (c) Marginal mass/density

$$f_Y(y) = \begin{cases} \sum_x f_{Y|X}(y|x) f_X(x), & \text{discrete case,} \\ \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)dx, & \text{continuous case.} \end{cases}$$

   (d) Reverse conditioning

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x)}{f_Y(y)} f_{Y|X}(y|x).$$

4. **Conditional expectation**: X and Y random variables.

   (a) Condition expectation of Y given X: define

   $$\psi(x) = E(Y|X=x) = \begin{cases} \sum_y y f_{Y|X}(y|x), & \text{discrete case,} \\ \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy, & \text{continuous case.} \end{cases}$$

   The conditional expectation of Y given X is $E(Y|X) = \psi(X)$ (a random variable).

   (b) Condition expectation of $g(Y)$ given X: If g is a well-behaved, real-valued function, define

   $$h(x) = E(g(Y)|X=x) = \begin{cases} \sum_y g(y) f_{Y|X}(y|x), & \text{discrete case,} \\ \int_{-\infty}^{\infty} g(y) f_{Y|X}(y|x)dy, & \text{continuous case.} \end{cases}$$

   The conditional expectation of $g(Y)$ given X is $E(g(Y)|X) = h(X)$ (a random variable).

   (c) Law of iterated expectations:

   $$E[\psi(X)] = E[E(Y|X)] = E(Y).$$

   Useful consequence,

   $$E(Y) = \begin{cases} \sum_x E(Y|X=x) f_X(x), & \text{discrete case,} \\ \int_{-\infty}^{\infty} E(Y|X=x) f_X(x)dx, & \text{continuous case.} \end{cases}$$

5. **Conditional variance**: for random variables X and Y, define

   $$\omega(x) = \text{Var}(Y|X=x) = \begin{cases} \sum_y [y - E(Y|X=x)]^2 f_{Y|X}(y|x), & \text{discrete case,} \\ \int_{-\infty}^{\infty} [y - E(Y|X=x)]^2 f_{Y|X}(y|x)dy, & \text{continuous case.} \end{cases}$$

   The conditional variance of Y given X is $\text{Var}(Y|X) = \omega(X) = E(Y^2|X) - [E(Y|X)]^2$. The conditional variance is a random variable and, using the law of iterated expectations, we can show that $E[\text{Var}(Y|X)] = \text{Var}(Y) - \text{Var}[E(Y|X)]$.

3.3 Dependence

1. **Independence of random variables** The random variables $X_1, X_2, \ldots, X_n$ are (mutually) independent if and only if the events $\{X_1 \le x_1\}, \{X_2 \le x_2\}, \ldots, \{X_n \le x_n\}$ are independent for all choices of $x_1, x_2, \ldots, x_n$.

   If $X_1, X_2, \ldots, X_n$ independent, then for all $x_1, x_2, \ldots, x_n$:

   (a) $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n)$,

   (b) $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n)$,

   (c) $E(X_1X_2\ldots X_n) = E(X_1)E(X_2)\ldots E(X_n)$,

   (d) $g_1(X_1), g_2(X_2), \ldots, g_n(X_n)$ are independent for real-valued functions $g_1, g_2, \ldots, g_n$.

2. **Covariance function**: for random variables X and Y,

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

   (a) Symmetry: $\text{Cov}(X,Y) = \text{Cov}(Y,X)$.

   (b) With constant multipliers: $\text{Cov}(aX, bY) = ab\text{Cov}(X,Y)$.

(c) Bilinearity: $\mathrm{Cov}(X_1+X_2, Y_1+Y_2) = \mathrm{Cov}(X_1, Y_1)+\mathrm{Cov}(X_1, Y_2)+\mathrm{Cov}(X_2, Y_1)+\mathrm{Cov}(X_2, Y_2)$.

(d) Variance: $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$,
$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$,
$\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X, Y)$.

(e) If X and Y are independent, $\mathrm{Cov}(X, Y) = 0$.

3. **Correlation coefficient**: for random variables X and Y,

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Correlation is scaled covariance, $|\mathrm{Corr}(X, Y)| \leq 1$. Independent implies uncorrelated but the reverse implication does not hold.

3.4 Sums of Independent Random Variables

1. **Bivariate case**: for independent random variables X and Y, let $Z = X + Y$ be the sum.

(a) Density of sum: is given by

$$f_Z(z) = \begin{cases} \sum_x f_X(x) f_Y(z-x), & \text{discrete case,} \\ \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx, & \text{continuous case.} \end{cases}$$

This type of function is known as a *convolution*, so f_Z is the convolution of f_X and f_Y denoted $f_Z = f_X * f_Y$. Note that the operation is commutative so $f_X * f_Y = f_Y * f_X$.

(b) Moment generating function of sum: $M_Z(t) = M_X(t)M_Y(t)$.

(c) Cumulant generating function of sum: $K_Z(t) = K_X(t) + K_Y(t)$.

2. **Multivariate identically distributed case**: for independent identically distributed random variables $X_1, \ldots, X_n$ with common density function f, let $S = \sum_{j=1}^{n} X_j$.

(a) Density of sum: $f_S(x) = (f * f * \ldots * f)(x)$, f convolved with itself n times.

(b) Moment generating function of sum: $M_S(t) = [M_X(t)]^n$.

(c) Cumulant generating function of sum: $K_S(t) = nK_X(t)$.

3. **Examples**

(a) Normal: $X_j \sim iN(\mu, \sigma^2) \Rightarrow S \sim N(n\mu, n\sigma^2)$.

(b) Poisson: $X \sim \mathrm{Pois}(\lambda_1),\ Y \sim \mathrm{Pois}(\lambda_2) \Rightarrow Z \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$,
$X_j \sim i\mathrm{Pois}(\lambda) \Rightarrow S \sim \mathrm{Pois}(n\lambda)$.

(c) Gamma: $X \sim \mathrm{Gamma}(r_1, \theta),\ Y \sim \mathrm{Gamma}(r_2, \theta) \Rightarrow Z \sim \mathrm{Gamma}(r_1 + r_2, \theta)$,
$X_j \sim i\mathrm{Exp}(\lambda) \Rightarrow S \sim \mathrm{Gamma}(n, \lambda)$.

3.5 Joint Moments

If X and Y are random variables with joint mass/density function $f_{X,Y}$ then the $(r, s)^{\text{th}}$ joint moment is

$$m_{r,s} = E(X^r Y^s) = \begin{cases} \sum_y \sum_x x^r y^s f_{X,Y}(x, y), & \text{discrete case,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s f_{X,Y}(x, y)dxdy, & \text{continuous case.} \end{cases}$$

The $(r,s)^{\text{th}}$ joint central moment is

$$\mu_{r,s} = E[(X-E(X))^r(Y-E(Y))^s] = \begin{cases} \sum_y \sum_x [(x-\mu_X)^r(y-\mu_Y)^s] f_{X,Y}(x,y), & \text{discrete case,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [(x-\mu_X)^r(y-\mu_Y)^s] f_{X,Y}(x,y)dxdy, & \text{continuous case.} \end{cases}$$

1. **Properties of joint moments**

   (a) r^{th} moment for X: $m_{r,0} = E(X^r)$.

   (b) r^{th} central moment for X: $\mu_{r,0} = E[(X-\mu_X)^r]$.

   (c) Covariance: $\mu_{1,1} = E[(X-E(X))(Y-E(Y))] = \text{Cov}(X,Y)$.

2. **Joint moment generating function**

$$M_{X,Y}(t,u) = E(e^{tX+uY}) = \begin{cases} \sum_y \sum_x e^{tx+uy} f_{X,Y}(x,y), & \text{discrete case,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{tx+uy} f_{X,Y}(x,y)dxdy, & \text{continuous case.} \end{cases}$$

   (a) Coefficient of $t^r u^s$:

$$\frac{1}{r!s!}E(X^rY^s) = \frac{1}{r!s!}m_{r,s}$$

   .

   (b) Moment generating function for marginal: $M_X(t) = E(e^{tX}) = M_{X,Y}(t,0)$,
   $M_Y(t) = E(e^{uY}) = M_{X,Y}(0,u)$.

   (c) Derivatives at zero:

$$M_{X,Y}^{(r,s)}(0,0) = \frac{d^{r+s}}{dt^r du^s} M_{X,Y}(t,u)\bigg|_{t=0,u=0} = E(X^rY^s) = m_{r,s}.$$

   (d) If X and Y independent: $M_{X,Y}(t,u) = M_X(t)M_Y(u)$.

3. **Conditional moment generating function**: define

$$\nu(u,x) = M_{Y|X}(u|x) = E(e^{uY}|X=x) = \begin{cases} \sum_y e^{uy} f_{Y|X}(y|x), & \text{discrete case,} \\ \int_{-\infty}^{\infty} e^{uy} f_{Y|X}(y|x)dy, & \text{continuous case.} \end{cases}$$

   The conditional moment generating function of Y given X is $M_{Y|X}(u|X) = \nu(u,X) = E(e^{uY}|X)$. This is a conditional expectation so it is a random variable. We can calculate joint moment generating function and moment generating function for marginal Y from the conditional moment generating function,

$$\begin{aligned} M_{X,Y}(t,u) &= E(e^{tX+uY}) = E[e^{tX}M_{Y|X}(u|X)], \\ M_Y(u) &= M_{X,Y}(0,u) = E[M_{Y|X}(u|X)]. \end{aligned}$$

4. **Joint cumulants**: let $K_{X,Y}(t,u) = \log M_{X,Y}(t,u)$, then we define the $(r,s)^{\text{th}}$ joint cumulant $\kappa_{r,s}$ as the coefficient of $(t^r u^s)/(r!s!)$ in the expansion of $K_{X,Y}$. Thus, $\kappa_{1,1} = \text{Cov}(X,Y)$.

5. **Multivariate generalisation**: for random variables $X_1, \ldots, X_n$ with joint mass/density function $f_{X_1,\ldots,X_n}$.

   (a) Joint moments:

$$\begin{aligned} m_{r_1,\ldots,r_n} &= E(X_1^{r_1}\ldots X_n^{r_n}) \\ &= \begin{cases} \sum_{x_1}\cdots\sum_{x_n} x_1^{r_1}\ldots x_n^{r_n} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n), & \text{discrete case,} \\ \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{r_1}\ldots x_n^{r_n} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)dx_1\ldots dx_n, & \text{continuous case.} \end{cases} \end{aligned}$$

(b) Joint central moments: $\mu_{r_1,\ldots,r_n} = E[(X_1 - E(X_1))^{r_1} \ldots (X_n - E(X_n))^{r_n}]$

(c) Joint moment generating function: $M_{X_1,\ldots,X_n}(t_1,\ldots,t_n) = E(e^{t_1X_1+\ldots+t_nX_n})$. The coefficient of $t_1^{r_1} \ldots t_n^{r_n}$ in the expansion of $M_{X_1,\ldots,X_n}$ is $E(X_1^{r_1} \ldots X_n^{r_n})/(r_1! \ldots r_n!)$.

(d) Joint cumulant generating function: $K_{X_1,\ldots,X_n}(t_1,\ldots,t_n) = \log(M_{X_1,\ldots,X_n}(t_1,\ldots,t_n))$. The $(r_1,\ldots,r_n)^{\text{th}}$ joint cumulant is defined as the coefficient of $(t_1^{r_1} \ldots t_n^{r_n})/(r_1! \ldots r_n!)$ in the expansion of $K_{X_1,\ldots,X_n}$.

(e) Independence: if $X_1,\ldots,X_n$ are independent then $M_{X_1,\ldots,X_n}(t_1,\ldots,t_n) = M_{X_1}(t_1)\ldots M_{X_n}(t_n)$.

3.6 Compounding and Random Sums

1. **From conditional to marginal**: for random variables X and Y suppose that we know the conditional distribution of Y given X and the marginal distribution of X.

   (a) Joint density: $f_{X,Y}(x,y) = f_{Y|X}(y|x)f(x)$.

   (b) Marginal density for Y:

   $$f_Y(y) = \begin{cases} \sum_x f_{Y|X}(y|x)f(x), & \text{discrete case,} \\ \int_{-\infty}^{\infty} f_{Y|X}(y|x)f(x)dx, & \text{discrete case.} \end{cases}$$

   (c) Moment generating function for Y:

   $$M_Y(t) = E[M_{Y|X}(t|X)] \begin{cases} \sum_x M_{Y|X}(t|x)f(x), & \text{discrete case,} \\ \int_{-\infty}^{\infty} M_{Y|X}(t|x)f(x)dx, & \text{discrete case.} \end{cases}$$

2. **Random sums**: suppose that $X_1, X_2, \ldots$ is a sequence of independent identically distributed random variables and $S = X_1 + \ldots + X_N$ where N is a random variable.

   (a) Conditional expectation: $E(S|N) = NE(X)$,
   variance: $\text{Var}(S|N) = N\text{Var}(X)$,
   moment generating function: $M_{S|N}(t|N) = [M_X(t)]^N$.

   (b) Marginal expectation: $E(S) = E(N)E(X)$,
   variance: $\text{Var}(S) = E(N)\text{Var}(X) + \text{Var}(N)[E(X)]^2$,
   moment generating function: $M_S(t) = M_N(\log M_X(t))$,
   cumulant generating function: $K_S(t) = K_N(K_X(t))$.

3. **Examples**: let $Y|X = x$ denote the random variable with density $f_{Y|X}(y|x)$.

   (a) Binomial and Poisson: $Y|N = n \sim \text{Bin}(n,p)$ and $N \sim \text{Pois}(\lambda) \Rightarrow Y \sim \text{Pois}(\lambda p)$,
   (equivalently $Y = \sum_{j=1}^N X_j$ where $X_j \sim \text{Bernoulli}(p)$).

   (b) Binomial and beta: $Y|\Theta = \theta \sim \text{Bin}(m,\theta)$ and $\Theta \sim \text{Beta}(\alpha,\beta)$
   $\Rightarrow f_Y(y) = \binom{m}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} / \frac{\Gamma(\alpha+\beta+m)}{\Gamma(\beta+m-y)\Gamma(\alpha+y)}$.

   (c) Poisson and gamma: $Y|\Lambda = \lambda \sim \text{Pois}(\lambda)$ and $\Lambda \sim \text{Gamma}(r,\alpha)$
   $\Rightarrow f_Y(y) = \frac{\Gamma(r+y)}{\Gamma(r)y!}(\frac{\alpha}{\alpha+1})^r(\frac{1}{\alpha+1})^y$.
   Y has a negative binomial distribution
   (in which r is not required to be an integer).

   (d) Poisson mixing: $Y|\Lambda = \lambda \sim \text{Pois}(\lambda)$ and Λ distributed on $[0,\infty)$
   $\Rightarrow E(Y) = E(\Lambda),\ \text{Var}(Y) = E(\Lambda) + \text{Var}(\Lambda)$.

(e) Poisson limit random sum: $S = \sum_{j=1}^{N} X_j$ where $N \sim \text{Pois}(\lambda)$

$\Rightarrow$ $E(S) = \lambda E(X)$, $\text{Var}(S) = \lambda E(X^2)$,
$M_S(t) = M_N(\log M_X(t)) = e^{\lambda(M_X(t)-1)}$,
$K_S(t) = \lambda(M_X(t) - 1)$, $\kappa_r = \lambda E(X^r)$.

(f) Random sum of exponentials: $S = \sum_{j=1}^{N} X_j$ where $X_j \sim \text{Exp}(\theta)$ and $N \sim \text{Geometric}(p)$

$\Rightarrow$ $S \sim \text{Exp}(\theta p)$.

3.7 Multivariate Normal Distribution

1. **Bivariate case**

(a) Independent normals: if U and V are $iN(0,1)$ then

$$f_{U,V}(u,v) = \frac{1}{2\pi} e^{-(u^2+v^2)/2},$$
$$M_{U,V}(s,t) = e^{(s^2+t^2)/2}.$$

(b) Standard bivariate normal: if $X = U$ and $Y = \rho U + \sqrt{1-\rho^2}V$ then $Y \sim N(0,1)$ and $\text{Corr}(X,Y) = \rho$. The joint distribution is a standard bivariate normal;

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(x^2-2\rho xy+y^2)/(2(1-\rho^2))}.$$

(c) General bivariate normal: if $X^* = \mu_X + \sigma_X X$ and $Y^* = \mu_Y + \sigma_Y Y$ then $X^* \sim N(\mu_X, \sigma_X^2)$ and $Y^* \sim N(\mu_Y, \sigma_Y^2)$ with $\text{Corr}(X^*, Y^*) = \rho$ and

$$f_{X^*,Y^*}(x,y) = \frac{1}{\sigma_X \sigma_Y} f_{X,Y}(\frac{x-\mu_X}{\sigma_X}, \frac{y-\mu_Y}{\sigma_Y}).$$

(d) Conditional distribution: of Y^* given X^* is

$$Y^*|X^* = x \ \sim \ N(\mu_y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X),\ \sigma_Y^2(1-\rho^2)).$$

2. **Multivariate case**

(a) Multivariate normal density: let $X_1, \ldots, X_n$ be random variables and define a random vector $\mathbf{X} = (X_1, \ldots, X_n)'$, where $'$ denotes transpose, so $\mathbf{X}$ is an $n \times 1$ vector. If $X_1, \ldots, X_n$ are jointly normal then $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the mean $\boldsymbol{\mu} = E(\mathbf{X})$ is an $n \times 1$ vector and the covariance matrix $\boldsymbol{\Sigma} = \text{Var}(\mathbf{X})$ is an $n \times n$ matrix whose $(i,j)^{\text{th}}$ entry is $\text{Cov}(X_i, X_j)$. The joint density functions is

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2} e^{-(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})/2}$$

(b) Conditional expectation for multivariate normal: suppose that $\mathbf{X} = (X_1, \ldots, X_n)'$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)'$, for some integers n and m, and $X \sim N(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ and $Y \sim N(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$. If, in addition, $\text{Cov}(\mathbf{X}, \mathbf{Y}) = \boldsymbol{\Sigma}_{XY} = \boldsymbol{\Sigma}_{YX}'$, then

$$E(\mathbf{Y}|\mathbf{X}) = \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_X^{-1}(\mathbf{X} - \boldsymbol{\mu}_X),$$
$$\text{Var}(\mathbf{Y}|\mathbf{X}) = \boldsymbol{\Sigma}_Y - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_X^{-1}\boldsymbol{\Sigma}_{XY}.$$

3.8 Laws of Large Numbers

1. **Types of convergence** Let $Y, Y_1, Y_2, \ldots,$ be a collection of random variables.
   (a) Convergence in distribution: $Y_n \xrightarrow{D} Y$ if $P(Y_n \leq y) \longrightarrow P(Y \leq y)$ as $n \longrightarrow \infty$.
   (b) Convergence in probability: $Y_n \xrightarrow{P} Y$ if $P(|Y_n - Y| > \varepsilon) \longrightarrow 0$ as $n \longrightarrow \infty$ for all $\varepsilon > 0$.
   (c) Convergence almost surely: $Y_n \xrightarrow{a.s.} Y$ if $A = \{\omega \in \Omega : Y_n(\omega) \longrightarrow Y(\omega)$ as $n \longrightarrow \infty\} \Rightarrow P(A) = 1$.

2. **Convergence of sample mean** Let $X_1, X_2, \ldots,$ be a sequence of independent identically distributed random variables with distribution function F_X, finite mean μ and, for part (d), finite non-zero variance σ^2. We definite $S_n = \sum_{j=1}^{n} X_j$ as the partial sum of n independent identically distributed random variables and $\frac{1}{n}S_n$ is the sample mean (a random variable).
   (a) The law of large numbers: $\frac{1}{n}S_n \xrightarrow{D} \mu$ as $n \longrightarrow \infty$.
   (b) Weak law of large numbers: $\frac{1}{n}S_n \xrightarrow{P} \mu$ as $n \longrightarrow \infty$.
   (c) Strong law of large numbers: $\frac{1}{n}S_n \xrightarrow{a.s.} \mu$ as $n \longrightarrow \infty$.
   (d) Central limit theorem:
   $$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{D} N(0,1) \quad \text{as} \quad n \longrightarrow \infty.$$

   Note: convergence to μ is convergence to the degenerate random variable with all of the mass concentrated at the single value μ.

3.9 Transformations of Random Variables

1. **Bivariate transformations**: Suppose that continuous random variables X_1 and X_2 are transformed into new random variables Y_1 and Y_2 by a one-to-one function $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ so $(Y_1, Y_2)' = T(X_1, X_2)'$. The joint density of Y_1 and Y_n is given by
$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} f_{X_1,X_2}(x_1(y_1,y_2), x_2(y_1,y_2))|J(y_1,y_2)|, & \text{if } (y_1,y_2) \text{ is in the range of T,} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$
   where $(x_1(y_1,y_2), x_2(y_1,y_2))' = T^{-1}(y_1,y_2)'$ and
$$J(y_1,y_2) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} - \frac{\partial x_2}{\partial y_1}\frac{\partial x_1}{\partial y_2}. \tag{2}$$
   This is the Jacobian of the inverse transformation.

2. **Ratios of random variables**: To illustrate the bivariate transformation, consider the following question. What is the probability density function of the ratio X_1/X_2. Let
$$Y_1 = X_1/X_2 \quad \text{and} \quad Y_2 = X_2.$$
   Simple rearrangement gives us $X_1 = Y_1Y_2$ and $X_2 = Y_2$. Thus in term of coordinates the inverse transformation is
$$x_1(y_1,y_2) = y_1y_2 \quad \text{and} \quad x_2(y_1,y_2) = y_2.$$
   Taking partial derivatives yields
$$\frac{\partial x_1}{\partial y_1} = y_2, \quad \frac{\partial x_1}{\partial y_2} = y_1, \quad \frac{\partial x_2}{\partial y_1} = 0, \quad \frac{\partial x_2}{\partial y_2} = 1.$$

From equation (2) above, we see that the Jacobian of the inverse transformation is $|J(y_1, y_2)| = |y_2|$. Substitution in equation (1) gives us the joint density,

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(y_1 y_2, y_2)|y_2|.$$

Integrating over y_2 yields the marginal for the ratio Y_1,

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(y_1 y_2, y_2)|y_2| dy_2.$$

3. **Multivariate generalisation** Consider the n-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_n)'$ and a one-to-one transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ to give another random vector $\mathbf{Y} = T\mathbf{X}$. The joint density function for $\mathbf{Y}$ is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} f_{\mathbf{X}}(\mathbf{x}(\mathbf{y}))|J(\mathbf{y})|, & \text{if } \mathbf{y} \text{ is in the range of T,} \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathbf{x}(\mathbf{y}) = T^{-1}\mathbf{y}$ and

$$J(\mathbf{y}) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial x_1}{\partial y_n} & \cdots & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix},$$

is the Jacobian of the inverse transformation.

