MATH3733 – Stochastic Financial Modelling

Part 3 – Probability background

“beta– version” 30.11.2008

(Revision and new material)

1 Probability, random variables (revision)

Definition 1 **Probability space** *is a triple* $(\Omega; F; P)$ *where* Ω *is any nonempty set which consists of points called* **outcomes** *or* **elementary events***;* F *is the set of* **all events** *which are subsets of* Ω *and for which* **probability** P *is defined.*

Example 1 *If we toss a coin once, we can describe this with the help of a probability space* $(\Omega; F; P)$ *with* $\Omega = \{0; 1\}$ *which represents two possible outcomes. An* **event** *is* **any subset of** Ω*, – remember that there is an empty set which is a subset of any other set, – and* **probability** *is* **a function** *on all events defined by the formula* $P(A) = \#A/\#\Omega$ *where* $\#A$ *means the number of points in* A*.*

Example 2 *If we roll a die once, we can describe this with the help of a* **probability space** $(\Omega; F; P)$ *with* $\Omega = \{1, 2, 3, 4, 5, 6\}$*, which represents six possible* **outcomes***. In this example* **an event** *is again* **any subset of** Ω *and* **probability** *is* **a function** *on all events defined by the formula* $P(A) = \#A/\#\Omega$*.*

Example 3 *If we throw a point on the interval* $[0; 1]$ *“randomly”, we can describe this with the help of a* **probability space** $(\Omega; F; P)$ *with* $\Omega = [0; 1]$ *which represents a continuum of all possible* **outcomes***. In this example* **an event** *is* **any collection of sub intervals** *of* $[0; 1]$*, and* **probability** *is a* **function** *on the set of all events defined by the formula* $P(A) = |A|/|\Omega|$ *where* $|A|$ *means the full length of* A *and* $|\Omega| = 1$*.*

Probability P must satisfy certain properties or axioms:

1. $P(A \bigcup B) = P(A) + P(B)$ if $A \bigcap B \equiv AB = \emptyset$;
2. $P(\Omega) = 1$.

Definition 2
1. *Events* A *and* B *are called* **independent** *if*
$$P(AB) = P(A)P(B);$$
2. *Events* $A_1, A_2, \ldots, A_n$ *are called* **independent** *if for any subset of* *different* *indices* $i_1, i_2, \ldots, i_k,\ k \geq 1$,
$$P(A_{i_1}, \ldots, A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j});$$

3. *If $P(B) > 0$ then conditional probability $P(A|B)$ is defined by the formula*

$$P(A|B) = P(AB)/P(B) \qquad (\text{if } P(B) = 0, \text{ then } P(A|B) \text{ is } \underline{\text{not defined}}).$$

Exercise 1 *Show the multiplication rule for probabilities:*

$$P(AB) = P(B)P(A|B).$$

Definition 3 1. **Random variable** *X is a* **function on a probability space** *$(\Omega; F; P)$ with values in R such that for any $x \in R$, the set $\{X \leq x\}$ is an* **event**, *that is, $\{X \leq x\} \in F$ and hence the probability $P(X \leq x)$ is defined.*

2. *Then the function*

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

is called a **distribution function** *or* **cumulative distribution function** *of X. Function $p(x)$ is called a* **density** *or* **distributional density** *of X if $F'(x) = p(x)$, or for any x,*

$$F(x) = \int_{-\infty}^{x} p(y)\, dy.$$

Definition 4 1. *Let X, Y be two random variables. The couple (X, Y) is called* **random vector**. *Its* **joint distribution function** *is $F(x; y) = P(X \leq x,\ Y \leq y)$.*

2. *A random vector (X, Y) is called* **continuous** *or* **continuously distributed** *if for any x, y,*

$$\partial^2 F(x, y)/\partial x \partial y = p(x; y),$$

or

$$\int_{-\infty}^{x} \int_{-\infty}^{y} p(u, v)\, du dv = F(x, y).$$

Exercise 2 *Show that marginal densities (also called one-dimensional densities) of both components are expressed as*

$$p_X(x) = \int_{-\infty}^{\infty} p(x, y)\, dy, \quad p_Y(y) = \int_{-\infty}^{\infty} p(x, y)\, dx$$

3. *In continuous case,* **conditional density** *of r.v. X* **given Y** *is function*

$$p_{X|Y}(x|Y) := \left.\frac{p(x, y)}{p_Y(y)}\right|_{y=Y} = \frac{p(x, Y)}{p_Y(Y)}.$$

4. In discrete case, **conditional distribution** *of r.v. X* **given** *Y is function*

$$P(X = x_k|Y)\,|_{Y=y_m} := \frac{P(X = x_k, Y = y_m)}{P(Y = y_m)}, \quad m = 1, \ldots, n.$$

Otherwise it may be expressed by

$$P(X = x_k|Y) := \sum_{m=1}^{n} \frac{P(X = x_k, Y = y_m)}{P(Y = y_m)} 1(Y = y_m).$$

- **Lemma** If a density $f(x)$ does exist and is continuous, then

$$f(x) = dF(x)/dx.$$

  Follows from the fundamental theorem of calculus.

- **Lemma**:
  For independent r.v.'s variances are additive,

$$var(\sum_{j=1}^{n} X_j) = \sum_{j=1}^{n} var(X_j).$$

- **Proof**. To simplify[1], we assume all $|X_j| \le C_0$. If one square out in

$$var(\sum_{j=1}^{n} X_j) = E\left(\sum_{j=1}^{n}(X_j - EX_1)\right)^2,$$

  it suffices to show that

$$(A) \qquad E(X_1 - EX_1)(X_2 - EX_2) = 0,$$

  and similarly for all other pairs. By the definition of independence,

$$(B) \qquad Eg_1(X_1 - EX_1)g_2(X_2 - EX_2) = Eg_1(X_1 - EX_1)Eg_2(X_2 - EX_2)$$

  for any bounded g_1, g_2. Take

$$g_1(x) = g_2(x) = x1(|x| \le 2C_0);$$

  then $g_1(X_1 - EX_1) = X_1 - EX_1$ and likewise $g_2(X_2 - EX_2) = X_2 - EX_2$. Hence, the desired property (A) follows.

[1] A general case here does require some care (believe it or not, but this general case is advanced reading!).

2 Expectations, conditional expectations, martingales

Definition 5 *Expectation and conditional expectation.*

1. **Expectation** revision *or mean value of X is*

$$EX = \int_{-\infty}^{\infty} x p_X(x)\,dx \quad \text{(continuous case)}, \quad EX = \sum_k x_k P(X = x_k) \quad \text{(discrete case)}.$$

2. **Conditional expectation** new **of X given Y** *is*

$$E(X|Y) = \int_{-\infty}^{\infty} x p_{X|Y}(x|Y)\,dx \qquad \text{(continuous case)}$$

(that is, integration with respect to the conditional density), and

$$E(X|Y)|_{Y=y_m} = \sum_k x_k P(X = x_k|Y = y_m), \quad m = 1, \dots, n \quad \text{(discrete case)},$$

or equivalently, $E(X|Y) = \sum_k x_k P(X = x_k|Y = y_m) 1(Y = y_m), \quad m = 1, \dots, n.$

Lemma 1 1. *In either case, for any function f such that expressions make sense,*

$$Ef(X) = \int_{-\infty}^{\infty} f(x) p_X(x)\,dx \text{ (continuous)}, \quad Ef(X) = \sum_k f(x_k) P(X = x_k) \text{ (discrete)}.$$

2. *In either case,*

$$E(E(X|Y)) = EX. \tag{1}$$

Proof of (1) for continuous case.

$$E(E(X|Y)) = E\left(\int_{-\infty}^{\infty} x p_{X|Y}(x|Y)\,dx\right) = \int \left[\int x \frac{p(x,y)}{\int p(x',y)\,dx'}\,dx\right] p_Y(y)\,dy$$

$$= \int \left[\frac{\int x p(x,y)\,dx}{\int p(x',y)\,dx'}\right] p_Y(y)\,dy = \int x \left[\int p(x,y)\,dy\right] dx = \int x p_X(x)\,dx = E(X).$$

Exercise 3 *Consider discrete case.*

Definition 6 *A* **random process** (X_t) *is a set of random variables while t runs over* $(0, 1, 2, \dots)$ *or* $[0, \infty)$.

Definition 7 (new material) *A* **Markov** *(will be recalled on the next lecture) random process (X_t) is called* **a martingale** *if for any $t < t'$, $E|X_t| < \infty$ and*

$$E(X_{t'} - X_t | X_t) = 0, \qquad \text{or equivalently}, \qquad E(X_{t'}|X_t) = X_t.$$

Lemma 2 *If (X_t) is a martingale, then*

$$E(X_t) = const \qquad \text{(i.e. the mean value does not depend on time)}.$$

Proof. By previous lemma, $E(X_t - X_0) = E(E(X_t - X_0|X_0)) = 0$, hence, $E(X_t) = E(X_0)$, $\forall t$.

3 Independent random variables

Definition 8 *We call r.v.'s* $X_1, X_2, \dots$ **independent** *if either of the following holds:*

1. *For any* n *and any* $x_1, x_s, \dots, x_n$,

$$P(X_1 \le x_1, \dots, X_n \le x_n) = \prod_{k=1}^{n} P(X_k \le x_k);$$

2. *for any* n*, if the random vector* $(X_1, \dots, X_n)$ *is continuous, for any* $x_1, x_s, \dots, x_n$,

$$p_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{k=1}^{n} p_{X_k}(x_k);$$

3. *for any* n *and for any bounded functions* $f_1(x), \dots, f_n(x)$,

$$E \prod_{k=1}^{n} f(X_k) = \prod_{k=1}^{n} Ef(X_k).$$

Concerning equivalence see textbooks on probability.

4 Random vectors and conditional distributions

Definition 9 1. *Let* $X_1, X_2, \dots, X_n$ *be* n *random variables. The* n*-tuple* $(X_1, X_2, \dots, X_n)$ *is called* n**-dimensional random vector**. *Its* **joint distribution function** *is* $F(x_1, x_2, \dots, x_n) = P(X_1 \le x_1, \dots, X_n \le x_n)$.

2. *This random vector is called* **continuous** *or* **continuously distributed** *if*

$$\partial^n F(x_1, \dots, x_n)/\partial x_1 \dots \partial x_n = p(x_1, \dots, x_n),$$

or

$$\underbrace{\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n}}_{n} p(u_1, \dots, u_n)\, du_1 \dots du_n = F(x_1, \dots, x_n).$$

Exercise 4 *Show that all marginal densities (also called one-dimensional densities) are equal to* $(n-1)$*-fold integrals,*

$$p_{X_1}(x_1) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-1} p(x_1, x_2, \dots, x_n)\, dx_2 \dots dx_n, \quad \dots \quad \textit{etc.},$$

$$p_{X_n}(x_n) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-1} p(x_1, \dots, x_{n-1}, x_n)\, dx_1 \dots dx_{n-1},$$

and also

$$p_{X_1,\dots X_{n-1}}(x_1, \dots, x_{n-1}) = \int_{-\infty}^{\infty} p(x_1, x_2, \dots, x_n)\, dx_n.$$

5 Conditional distributions for random vectors

Definition 10 *1. In continuous case,* **conditional density** *of r.v. X* **given Y** *is function*

$$p_{X|Y}(x|Y) := \left.\frac{p(x,y)}{p_Y(y)}\right|_{y=Y} = \frac{p(x,Y)}{p_Y(Y)}.$$

2. In discrete case, **conditional distribution** *of r.v. X* **given Y** *is function*

$$P(X = x_k|Y)\,|_{Y=y_m} := \frac{P(X = x_k, Y = y_m)}{P(Y = y_m)}, \quad m = 1, \ldots, n.$$

Otherwise it may be expressed by

$$P(X = x_k|Y) := \sum_{m=1}^{n} \frac{P(X = x_k, Y = y_m)}{P(Y = y_m)} 1(Y = y_m).$$

6 Conditional distributions for random vectors

Definition 11 *1. Let a n-dimensional random vector $(X_1, \ldots X_n)$ have a joint (n-dimensional) density $p(x_1, \ldots, x_n)$. Denote $Y = (X_1, \ldots, X_{n-1})$. We call a conditional density of X_n given Y the function*

$$p_{X_n|Y}(x_n|Y) := \frac{p(X_1, \ldots, X_{n-1}, x_n)}{p_Y(X_1, \ldots, X_{n-1})}.$$

2. Let a n-dimensional random vector $(X_1, \ldots X_n)$ be discrete one with a finite set of possible values of all coordinates. Denote $Y = (X_1, \ldots, X_{n-1})$. Conditional distribution of X_n given $Y = (X_1, \ldots, X_{n-1})$ is function (here y_m is an $(n-1)$-tuple)

$$P(X = x_k|Y)\,|_{Y=y_m} := \frac{P(X = x_k, Y = y_m)}{P(Y = y_m)}, \quad m = 1, \ldots, n,$$

or equivalently,

$$P(X = x_k|Y) := \sum_{m=1}^{n} \frac{P(X = x_k, Y = y_m)}{P(Y = y_m)} 1(Y = y_m).$$

7 Conditional expectations

Definition 12

1. **A conditional expectation of** X **given** $Y = (X_1, \ldots, X_{n-1})$ *is*

$$E(X|Y) = \int_{-\infty}^{\infty} x p_{X|Y}(x|Y)\, dx \qquad \textit{(continuous case)}$$

(that is, integration with respect to the conditional density), and

$$E(X|Y)|_{Y=y_m} = \sum_k x_k P(X = x_k|Y = y_m), \quad m = 1, \ldots, n \qquad \textit{(discrete case)};$$

otherwise, the latter formula can be expressed by

$$E(X|Y) = \sum_{k,m} x_k P(X = x_k|Y = y_m) 1(Y = y_m), \quad m = 1, \ldots, n.$$

8 Information

Information is crucial for the market. Hence, a mathematical financial theory uses it under the name **filtration**; other names: **information**, **sigma-field**, **sigma-algebra**).

Definition 13 *Let market consist of one stock which is modelled by a random process* $X_t,\ t = 0, 1, 2, \ldots$. **A filtration or information to time** t *is* **the set** $\mathcal{F}_t^X$ (*this is a notation*) **of all events which correspond to the values of the process** X **until this time**, $(X_0, X_1, \ldots, X_t)$*; in the other words,* $\mathcal{F}_t^X$ *consists of* **all available information on the market till time** t.

The filtration includes the knowledge of any **event** connected with the values $X_0, X_1, \ldots, X_t$, that is, if we know information till time t, we must know whether any such event was realized or not.

Definition 14

1. **A conditional expectation of** X_n **given information** $\mathcal{F}_{n-1}^X$ *is*

$$E(X_n|\mathcal{F}_{n-1}^X) := E(X_n|X_1, \ldots, X_{n-1}).$$

9 Markov process and martingales

The first definition relates to the simplest case of a process with just three moments of time.

Definition 15 *We call process* $(X_t,\ t = 0, \delta, 2\delta)$ **Markov** *(or* **markovian***) if for any* $0 < t < 2\delta$ *(i.e., just for* $t = \delta$*!), the conditional probability of the event* $\{X_t \le x\}$ **given all the past** *of the process* X *equals to the (again conditional) probability of this event* **given only the last value of the process***, that is, given* $X_{t-\delta}$.

The second definition relates to the case of a process with $t = 0, \delta, 2\delta, \ldots$.

Definition 16 *We call process* $(X_t,\ t = 0, \delta, 2\delta, \ldots)$ **Markov** *if for any available* t, *the conditional probability of the event* $\{X_t \le x\}$ **given all the past** *of the process* X *equals to the (again conditional) probability of this event* **given only the last value of the process**, *that is, given* $X_{t-\delta}$*:*

$$P(X_t \le x|\mathcal{F}^X_{t-1}) = P(X_t \le x|X_{t-1}).$$

Remind that the information $\mathcal{F}^X_{t-1}$ coincides with the vector $(X_0, X_1, \ldots, X_{t-1})$, so that one could rewrite the definition above as

$$P(X_t \le x|X_1, \ldots, X_{t-1}) = P(X_t \le x|X_{t-1}).$$

Definition 17 *Markov process* $(X_t,\ t = 0, \delta, 2\delta, \ldots)$ *is a* **martingale** *if for any* $t < t'$,

$$E|X_t| < \infty \quad \textit{and} \quad E(X_{t'}|X_t) = X_t, \quad \textit{or} \quad E(X_{t'} - X_t|X_t) = 0.$$

Definition 18 *Markov process* $(X_t,\ t = 0, \delta, 2\delta, \ldots)$ *is a* **discounted martingale with the rate** r *if* $\underline{\exp(-rt)X_t \textit{ is a martingale}}$, *i.e. for any* $t < t'$,

$$E|X_t| < \infty \quad \textit{and} \quad \exp(-rt')E(X_{t'}|X_t) = \exp(-rt)X_t.$$

10 More about conditional expectations

- The following formula can be shown by induction, similarly to the proof of lemma 1 from lecture 6.

  $$E(E(X_n|\mathcal{F}^X_{n-1})|\mathcal{F}^X_{n-2}) = E(X_n|\mathcal{F}^X_{n-2}). \tag{2}$$

  More generally, for any $0 < k \le n$, by induction

  $$E(E(X_n|\mathcal{F}^X_{n-1})|\mathcal{F}^X_{n-k}) = E(X_n|\mathcal{F}^X_{n-k}). \tag{3}$$

  They say, while taking conditional expectations, a **smaller** filtration "eats" a greater one. $\mathcal{F}^X_{n-2}$ is smaller than $\mathcal{F}^X_{n-1}$ in the usual sense: at time $n-2$ less information is available than at time $n-1$.

  Exercise 5 *Show it.*

  Solution (in continuous case). To show (2), we denote

$$\int_{-\infty}^{\infty} x_n \frac{p_{X_1,\dots,X_{n-2},X_{n-1},X_n}(X_1,\dots,X_{n-2},X_{n-1},x_n)}{p_{X_1,\dots,X_{n-2},X_{n-1}}(X_1,\dots,X_{n-2},X_{n-1})}\,dx_n =: \phi(X_1,\dots,X_{n-2},X_{n-1}),$$

and write,

$$E(E(X_n|\mathcal{F}^X_{n-1})|\mathcal{F}^X_{n-2}) = E(\int_{-\infty}^{\infty} x_n p_{X_n|X_1,\dots,X_{n-1}}(x_n|X_1,\dots,X_{n-1})\,dx_n \mid \mathcal{F}^X_{n-2})$$

$$= E(\int_{-\infty}^{\infty} x_n \frac{p_{X_1,\dots,X_{n-2},X_{n-1},X_n}(X_1,\dots,X_{n-2},X_{n-1},x_n)}{p_{X_1,\dots,X_{n-2},X_{n-1}}(X_1,\dots,X_{n-2},X_{n-1})}\,dx_n \mid \mathcal{F}^X_{n-2})$$

$$= E(\phi(X_1,\dots,X_{n-2},X_{n-1}) \mid \mathcal{F}^X_{n-2})$$

$$= \int \phi(X_1,\dots,X_{n-2},x_{n-1})\frac{p_{X_1,\dots,X_{n-2},X_{n-1}}(X_1,\dots,X_{n-2},x_{n-1})}{p_{X_1,\dots,X_{n-2}}(X_1,\dots,X_{n-2})}\,dx_{n-1}$$

$$= \int\int x_n \frac{p_{X_1,\dots,X_{n-2},X_{n-1},X_n}(X_1,\dots,X_{n-2},x_{n-1},x_n)}{p_{X_1,\dots,X_{n-2},X_{n-1}}(X_1,\dots,X_{n-2},x_{n-1})}$$

$$\times \frac{p_{X_1,\dots,X_{n-2},X_{n-1}}(X_1,\dots,X_{n-2},x_{n-1})}{p_{X_1,\dots,X_{n-2}}(X_1,\dots,X_{n-2})}\,dx_{n-1}dx_n$$

$$= \int\int x_n \frac{p_{X_1,\dots,X_{n-2},X_{n-1},X_n}(X_1,\dots,X_{n-2},x_{n-1},x_n)}{p_{X_1,\dots,X_{n-2}}(X_1,\dots,X_{n-2})}\,dx_{n-1}dx_n.$$

On the other hand side,

$$E(X_n|\mathcal{F}^X_{n-2}) = \int x_n \frac{p_{X_1,\dots,X_{n-2},X_n}(X_1,\dots,X_{n-2},x_n)}{p_{X_1,\dots,X_{n-2}}(X_1,\dots,X_{n-2})}\,dx_n$$

$$= \int x_n \left(\int \frac{p_{X_1,\dots,X_{n-2},X_{n-1},X_n}(X_1,\dots,X_{n-2},x_{n-1},x_n)}{p_{X_1,\dots,X_{n-2}}(X_1,\dots,X_{n-2})}\,dx_{n-1}\right) dx_n,$$

too, because of the identity,

$$\int p_{X_1,\dots,X_{n-2},X_{n-1},X_n}(X_1,\dots,X_{n-2},x_{n-1},x_n)\,dx_{n-1} = p_{X_1,\dots,X_{n-2},X_n}(X_1,\dots,X_{n-2},x_n).$$

In words, the last identity reads, "if we integrate the joint density of several r.v.'s with respect to one of variables, the result is the density of the remaining r.v.'s".

Hence, (2) follows. The second identity is proved similarly.

- *Useful Exercise:* show the same in discrete case.
  Hint: replace integrals by sums and joint densities by discrete joint densities.

- If X and Y are independent then $E(X|Y) = EX$. **Hint:** $p(x,y) = p_X(x)p_Y(y)$, hence, $p_{X|Y}(x|Y) \equiv p_X(x)$ (continuous case); likewise in discrete case.

11 Weak convergence and characteristic functions

key topic

There are three **equivalent** notions of "weak probabilistic convergence" (again, see textbooks on probability).

Definition 19 *Sequence of random variables* $X_1, X_2, \ldots$ **converges weakly** *to a r.v.* X_0*, notation* $X_n \Longrightarrow X_0$*, if any of the following holds true:*

1. $F_{X_n}(x) \to F_{X_0}(x)$*, for any* x *where* F_{X_0} *is continuous* $(F_{X_n}(x) = P(X_n \le x))$*;*
2. $Ef(X_n) \to Ef(X_0),\ n \to \infty, \forall f$ *continuous and bounded;*
3. $E\exp(i\lambda X_n) \to E\exp(i\lambda X_0),\ n \to \infty, \forall \lambda \in R$.

Definition 20 *An expression* $E\exp(i\lambda X)$ *as function of* $\lambda \in R$ *is called* **characteristic function** *of r.v.* X.

Hence, if we are going to show that (cf. below)

$$\frac{S_n - na}{\sigma\sqrt{n}} \Longrightarrow Y \sim \mathcal{N}(0,1), \tag{4}$$

we may use any of equivalent definitions above. A connection between the second and the third definitions appeals to Fourier analysis, namely, the Weierstrass theorem saying that any continuous function $f(x)$ on $[a,b]$ can be approximated by finite sums of trigonometric polynomials of the form $\sum_k \exp(i\lambda_k x)$. [In turn, this may be shown using simple probabilistic methods, namely, the Law of Large Numbers theorem, see [W. Feller, An Introduction To Probability Theory And Its Applications, vol.1]].

12 Characteristic function for $\mathcal{N}(0,1)$ very important

To use the third definition, we may need to know the characteristic function for $\mathcal{N}(0,1)$, which we now calculate:

1. **Exercise 6** *Let* $Y \sim \mathcal{N}(0,1)$*. Show that for any* $n = 1, 3, 5, \ldots$*,* $EY^n = 0$*, while for any* $n = 2k,\ k = 1, 2, \ldots$*,*

   $$EY^n = (n-1)!! \qquad (k!! = k(k-2)\ldots).$$

   Hint: integrate by parts using $\int (x\exp(-x^2/2))dx = -\exp(-x^2/2)$ (which aims to suggest you what is to be chosen as u and what by v', in the expression EY^n represented via a Gaussian integral and further as $\int uv'$)*, and use induction.*

2. **Compulsory;** *advanced reading* in this calculus would be to explain why $E\sum = \sum E$.

$$Ee^{i\lambda Y} = E\sum_{n=0}^{\infty}\frac{(i\lambda Y)^n}{n!} = \sum_{n=0}^{\infty}E\frac{(i\lambda Y)^n}{n!} = \sum_{k=0}^{\infty}E\frac{(i\lambda Y)^{2k}}{(2k)!} = \sum_{n=0}^{\infty}\frac{(-\lambda^2)^k((2k-1)!!)}{(2k)!}$$

$$= \sum_{n=0}^{\infty}\frac{(-\lambda^2)^k(2k-1)(2k-3)\dots}{(2k)(2k-1)(2k-2)\dots} = \sum_{n=0}^{\infty}\frac{(-\lambda^2)^k}{(2k)(2k-2)\dots} = \sum_{n=0}^{\infty}\frac{(-\lambda^2)^k}{2^k\,k!} = e^{-\lambda^2/2}.$$

3. *Useful Exercise.* Compute all moments and characteristic function of $Z \sim \mathcal{N}(0,\sigma^2)$ (with $\sigma^2 > 0$).
Hint: Consider $Y := Z/\sigma$, show that $Y \sim \mathcal{N}(0,1)$, and use the results of the previous two Exercises for Y.

13 Properties of characteristic functions important

Lemma 3 *Assuming that all the expectations below exist,*

$$(E\exp(i\lambda X))' = E(iX)\exp(i\lambda X)), \qquad E\exp(i\lambda X))'' = -E(X)^2\exp(i\lambda X)),$$
$$E\exp(i\lambda X))''' \;=\; -iE(X)^3\exp(i\lambda X)).$$

Proof. We will show the formulas for finite sums in discrete case:

$$(E\exp(i\lambda X))' = (\sum_k \exp(i\lambda x_k)p_k)' = \sum_k(\exp(i\lambda x_k)p_k)'$$
$$= \sum_k(ix_k)\exp(i\lambda x_k)p_k = E(iX)\exp(i\lambda X),$$

and similarly,

$$(E\exp(i\lambda X))'' = \sum_k(\exp(i\lambda x_k)p_k)'' = \sum_k(ix_k)^2\exp(i\lambda x_k)p_k = -E(X)^2\exp(i\lambda X),$$
$$E\exp(i\lambda X))''' = \sum_k(\exp(i\lambda x_k)p_k)''' = \sum_k(ix_k)^3\exp(i\lambda x_k)p_k = -iE(X)^3\exp(i\lambda X).$$

Lemma 4 *Let $EX = a$, $EX^2 = b$, $E|X|^3 = c < \infty$. Then*

$$E\exp(i\lambda X) = 1 + i\lambda a - \lambda^2 b/2 + o(\lambda^2)\ (\lambda \to 0),\quad \textit{where } o(\lambda^2) = O(c^3|\lambda|^3). \qquad (5)$$

Proof follows from the **Taylor expansion** with two terms and the reminder in the integral or Lagrange form, if we use the previous lemma and assertions

$$(E\exp(i\lambda X))'|_{\lambda=0} = iE(X),\ (E\exp(i\lambda X))''|_{\lambda=0} = i^2E(X)^2,\ (E\exp(i\lambda X))'''|_{\lambda=0} = i^3E(X)^3.$$

14 Central Limit Theorem

(The following theorem can be proved without assumption $E|X|^3 < \infty$, but we would like to simplify the calculus and prepare ourselves to theorem 2 below.)

Theorem 1 *Let $X_1, X_2, \dots$ be independent and identically distributed r.v.'s, such that $EX_k = a$, $var(X_k) = \sigma^2 > 0$, and $c = E|X_k|^3 < \infty$; denote $S_n = \sum_{k=1}^n X_k$. Then*

$$\frac{S_n - na}{\sigma\sqrt{n}} \Longrightarrow Y \sim \mathcal{N}(0,1).$$

Proof. We use characteristic functions:

$$E\exp(\frac{i\lambda(S_n - na)}{\sigma\sqrt{n}}) = \prod_{k=1}^n E\exp(\frac{i\lambda(X_k - a)}{\sigma\sqrt{n}}) \quad \text{(we used independence)}$$

$$= \left(E\exp(\frac{i\lambda(X_1 - a)}{\sigma\sqrt{n}})\right)^n \quad \text{(we used that all } X\text{'s are identically distributed)}$$

$$= \left(1 - \frac{\lambda^2\sigma^2}{2n\sigma^2} + O(\frac{|\lambda|^3 c}{\sigma^3 n^{3/2}})\right)^n \quad \text{(we used lemma 2)}$$

$$\to \exp(-\lambda^2/2), \quad n \to \infty. \quad \text{The latter is the characteristic function of } Y.$$

15 Another CLT very important

There is a lot of versions of CLT. In our course we will need the following one.

Theorem 2 *Let for any n, $X_1(n), X_2(n), \dots$ be independent and identically distributed r.v.'s, – perhaps different for different n's, – such that $EX_k(n) = a(n)$, $var(X_k) = \sigma^2(n) \to \sigma^2 > 0$, and $c_n = E|X_k|^3 \le c \le \infty$; denote $S_n = \sum_{k=1}^n X_k(n)$. Then*

$$\frac{S_n - na(n)}{\sigma(n)\sqrt{n}} \Longrightarrow Y \sim \mathcal{N}(0,1).$$

Proof, in fact, repeats the previous proof: as $n \to \infty$,

$$E\exp(\frac{i\lambda(S_n(n) - na(n))}{\sigma(n)\sqrt{n}}) = \prod_{k=1}^n E\exp(\frac{i\lambda(X_k - a(n))}{\sigma(n)\sqrt{n}}) = \left(E\exp(\frac{i\lambda(X_1 - a(n))}{\sigma(n)\sqrt{n}})\right)^n$$

$$= \left(1 - \frac{\lambda^2\sigma^2(n)}{2n\sigma^2(n)} + O(\frac{|\lambda|^3 c_n}{\sigma(n)^3 n^{3/2}})\right)^n \sim \left(1 - \frac{\lambda^2}{2n} + O(\frac{|\lambda|^3 c}{\sigma^3 n^{3/2}})\right)^n \to \exp(-\lambda^2/2).$$

Remark. There is another way to prove both CLTs, without characteristic functions, see the additional handout on my home page. That handout has been distributed, but is considered as advanced reading. The calculus is more lengthy, however, remember that no use of characteristic functions is required, just Taylor's expansion.

- **The Bienaimé-Chebyshev inequality** **important for any statistician**

$$\boxed{P(|\xi - E\xi| \geq c) \leq c^{-2} var(\xi),} \qquad \forall \ c > 0.$$

  - Proof for discrete case: denote $E\xi = a$;

$$P(|\xi - a| \geq c) = \sum_{i:\,|x_i - a| \geq c} p_i \leq \sum_{i:\,|x_i - a| \geq c} p_i \frac{|x_i - a|^2}{c^2} \leq \sum_{\text{all } i} p_i \frac{|x_i - a|^2}{c^2}.$$

  - Proof for continuous case: denote $E\xi = a$;

$$P(|\xi - a| \geq c) = \int\limits_{|x-a| \geq c} p(x)\, dx \leq \int\limits_{|x-a| \geq c} p(x) \frac{|x - a|^2}{c^2}\, dx \leq \int p(x) \frac{|x - a|^2}{c^2}\, dx.$$

  - **Corollary 1** *If $var(\xi) \approx 0$ then $\xi \approx E\xi$ in the sense that the difference between ξ and a might be greater than any small c only with a very small probability.*

- **Law of Large Numbers** **important for any statistician**

If $X_1, X_2, \ldots$ are independent and identically distributed r.v's (IID) with $EX_1 < \infty$ and $var(X_1) < \infty$, then for any $\epsilon > 0$,

$$P\left(\left| n^{-1} \sum_{j=1}^{n} X_j - EX_1 \right| > \epsilon \right) \to 0, \quad n \to \infty.$$

(Holds true without the assumption on var; the latter simplifies the proof.)

- Proof: by Bienaimé-Chebyshev, as $n \to \infty$,

$$P\left(\left| n^{-1} \sum_{j=1}^{n} X_j - EX_1 \right| > \epsilon \right) \leq \frac{var\left(n^{-1} \sum_{j=1}^{n} (X_j - EX_1) \right)}{\epsilon^2} = \frac{n\ var(X_1)}{n^2 \epsilon^2} \to 0.$$

Usually, in any textbook this Theorem precedes CLT.

16 Further reading on probability background

W. Feller, Introduction to Probability, *vol.1*;
A. N. Shiryaev, Probability.