GAUSSIAN PROCESSES

0. Definitions and overview

[1]This course is on Gaussian processes. Our aim is to understand what makes Gaussians special, i.e. why do they appear so widely and what makes them easier to study. We will both meet properties that seem special, but are actually shared by a larger family of models, but also properties that really are special to Gaussians.

Definition 0.1 (Gaussian random variable)**.** *A real-valued random variable X is called Gaussian, if there exist $m \in \mathbb{R}, \sigma > 0$ such that the density of X w.r.t the Lebesgue measure dx is given by*

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

We will denote the law of X by $N(m,\sigma^2)$. The law $N(0,1)$ is called the standard normal or standard Gaussian law.

Equivalently we can characterize the law of a random variable via its characteristic function (why?). It is an exercise to show that if $X \sim N(m,\sigma^2)$, then

$$\mathbb{E}\exp(i\lambda X) = \exp\left(i\lambda m - \frac{\lambda^2\sigma^2}{2}\right).$$

Similarly, there are several ways to define Gaussian random vectors in $\mathbb{R}^n$.

Definition 0.2 (Gaussian random variable)**.** *A $\mathbb{R}^n$-valued random variable $\overline{X} = (X_1, \dots, X_n)$ is called Gaussian, if for any non-zero $\overline{\lambda} = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ the inner product $\langle \overline{\lambda}, \overline{X}\rangle := \sum_{i=1}^n \lambda_i X_i$ is a Gaussian random variable.*

We denote the Gaussian law of $\mathbb{R}^n$ by $N(\overline{M}, \Sigma^2)$, where $\overline{M} = \mathbb{E}\overline{X}$ and $\Sigma^2(i,j) = \mathbb{E}X_iX_j$. If $\overline{M} = \overline{0}$ and $\Sigma^2 = \mathrm{Id}_n$, we call the vector a standard Gaussian on $\mathbb{R}^n$.

The following lemma giving 3 more equivalent ways of defining Gaussian vectors will be on the exercise sheet.

Lemma 0.3 (Equivalent definitions of a Gaussian vector)**.** *Show that $\overline{X}$ is a Gaussian vector if and only if*

1. *There exist a $n \times n$ matrix A of rank n, and a vector $\overline{M}$ such that if $\overline{Y}$ is a vector of n independent Gaussian random variables, then $\overline{X} = A\overline{Y} + \overline{M}$ almost surely.*
2. *There is a symmetric positive definite matrix D and a vector $\overline{M}$ such that that density of $\overline{X}$ w.r.t. the Lebesgue measure $d\overline{x}$ on $\mathbb{R}^n$ is given by $\frac{1}{(2\pi)^{n/2}\det(D)^{1/2}}\exp\left(-\frac{1}{2}(\overline{x}-\overline{M})^T D(\overline{x}-\overline{M})\right)$.*
3. *There is a symmetric positive definite matrix C and a vector $\overline{M}$ such that for all vectors $\overline{\lambda} \in \mathbb{R}^n$, we have that $\mathbb{E}\exp(\langle\overline{\lambda},\overline{X}\rangle) = \exp\left(\langle\overline{\lambda},\overline{M}\rangle + \frac{1}{2}\overline{\lambda}^T C\overline{\lambda}\right)$.*

Remark 0.4. *Sometimes it's useful to also include degenerate Gaussians, meaning the case where $\det(D) = 0$ (i.e. $\sigma = 0$ in the 1D case). In this case the Gaussian vector lives actually in a smaller dimensional subspace, in the 1D case it would be a 0-dimensional subspace, i.e. a single point. We will be pretty free about this, and include this degenerate setting when it makes things easier to state.*

[1]Please let me know if you meet any typos or unclarities!

Finally, a Gaussian process is nothing else than a collection of random variables, now indexed by any set, such that the joint law of any finite number of them is Gaussian.

Definition 0.5 (Gaussian process)**.** *Let I be any index set. Then a collection of random variables $(X_i)_{i\in I}$ defined on the same probability space is called a Gaussian process if for any finite subset $(i_1, \dots, i_n)$ we have that $\overline{X} := (X(i_1), \dots, X(i_n))$ is a Gaussian vector.*

I have left aside a fundamental, quasi-philosophical question - do Gaussians exist? They do exist in the rigorous framework of probability theory, and I let you ponder upon it on the example sheet.

1. Stable laws and CLTs

In this section we look at a property of Gaussians shared by a whole family of probability laws, called stability. We will see the relations between stability and Central limit theorems (CLTs), and try to understand the place of the Gaussians among these stable laws.

Definition 1.1 (Stable law)**.** *A probability law μ is said to be strictly stable, if for any $n \geq 1$ there is some $b_n > 0$ such that if we take i.i.d random variables $X_1, \dots, X_n \sim \mu$, we have that $b_n^{-1}\sum_{i=1}^n X_i \sim \mu$.*

P. Levy introduced stable laws in this form, and later they have been slightly generalized: in the general case, we ask that there are some $b_n > 0$ and $a_n \in \mathbb{R}$ with $a_n + b_n^{-1}\sum_{i=1}^n X_i \sim \mu$. We will stick to the strictly stable setting and when we are lazy we call them stable laws. Often a different characterization of stable laws is used and it's a good exercise to check its equivalence with the definition above:

Lemma 1.2. *μ is a (strictly) stable law if and only if for all $a > 0, b > 0$ there is some $c > 0$ such that if $X_1, X_2 \sim \mu$ are independent and $X \sim \mu$, then $aX_1 + bX_2 \sim cX$.*

Proof. The "if" part is pretty direct: one can just prove it by induction. Indeed, given that for some $b_n > 0$ we have that $b_n^{-1}(X_1 + \cdots + X_n) \sim \mu$, we know that $X_1 + \cdots + X_n + X_{n+1}$ equals in law with $b_nX_1 + X_{n+1}$ to which we can apply the hypothesis.

The "only if" part needs a bit of work and will be sketched in the example sheet, possibly in the non-examinable setting. □

In this course, the first thing to verify is that we are not going to far from Gaussians:

Claim 1.3. *The Gaussian law $N(0, \sigma^2)$ is strictly stable.*

Proof. The easiest way to see this is probably to use the characteristic function: let $X_1, X_2, \dots, X_n \sim N(0, \sigma^2)$ be independent. Then we can directly calculate the characeteristic function of $S_n = \sum_{i=1}^n X_i$. Indeed, using independence and then plugging in the characeteristic function of the Gaussian, we have that

$$\mathbb{E}\exp(i\lambda S_n) = (\mathbb{E}\exp(i\lambda X_1))^n = \exp(-n\lambda^2\sigma^2/2).$$

Thus we recognize that $S_n \sim N(0, n)$ and we see that Gaussian law satisfies the condition of Definition 1.1 with $b_n = \sqrt{n}$. □

But Gaussians are certainly not the only stable laws:

Claim 1.4. *Suppose that X is a random variable with characteristic function $\mathbb{E}\exp(i\lambda X) = \exp(-C|\lambda|^\alpha)$ with $\alpha \in (0, 2]$ and $C > 0$, then X has a stable law.*

Proof. This is on the exercise sheet. As soon as one accepts that this is a characteristic function of a probability measure, the proof is basically the same as for the Gaussian case. □

In fact, the family of stable laws is larger than this, but again for our purposes this gives enough context. Notice that we had a restriction $\alpha \leq 2$, although if you go through the proof we don't seem to use it. Where is the catch?

Exercise 1.1. *Prove that for $\alpha > 2$ and $C > 0$, $\mathbb{E}\exp(i\lambda X) = \exp(-C|\lambda|^\alpha)$ cannot be the characteristic function of a real-valued random variable.*

It comes out that in fact every stable law is related to a central limit theorem:

Proposition 1.5 (Stable laws and CLTs)**.** *A probability law μ is (strictly) stable if and only if we can find i.i.d random variables $X_1 X_2, \dots$ and $c_n > 0$ such that $c_n^{-1}\sum_{i=1}^n X_i$ converges in law to μ.*

Proof. One direction is easy - if μ is a stable law, then we can just take $X_i \sim \mu$ for all $i \geq 1$ and $c_n = b_n$. Then as μ is stable, we have that $b_n^{-1}\sum_{i=1}^n X_i \sim \mu$ and when we take $n \to \infty$, we still obtain μ as a limit.

In the other direction we want to show that if μ is a limit of a renormalized sum $b_n^{-1}\sum_{i=1}^n X_i$ for some i.i.d random variables X_i, then in fact μ is a stable law. Thus for every $m \in \mathbb{N}$, we want to show that if $Y_1, \dots, Y_m \sim \mu$ are independent, then there is some $b_m > 0$ such that $b_m^{-1}\sum_{i=1}^m Y_i \sim \mu$. So fix such m.

By the assumption, there are i.i.d random variables $X_1, X_2, \dots$ and $c_n > 0$ such that $c_n^{-1}\sum_{i=1}^n X_i \to \mu$ in law. Now consider the sequence only along $n = mk$ for $k \geq 1$ and decompose the sum into m pieces: for each $i = 1 \dots m$ set $S_{i,n} = \sum_{j=0}^{k-1} X_{jm+i}$. Then by the assumptions as $k \to \infty$ we have that $c_k^{-1} S_{i,n} \to Y_i$ in law, with $Y_i \sim \mu$. Moreover as all $S_{i,n}$ are independent, $Y_1, \dots, Y_m$ are independent and we have that $c_k^{-1}\sum_{i=1}^m S_{i,n} \to Y_1 + \dots + Y_m$. But also $c_n^{-1}\sum_{i=1}^m S_{i,k} \to \mu$ in law. Thus if we knew that $c_k c_n^{-1}$ converges to a limit b_m^{-1}, then we would have $b_m^{-1}\sum_{i=1}^m Y_i \sim \mu$.

It remains to argue that $a_n := c_k c_n^{-1}$ should converge to a limit as $n \to \infty$. By denoting $W_n := c_k^{-1}\sum_{i=1}^n X_i$, we can deduce this from a general claim:

Claim 1.6. *Let $(a_n)_{n\geq 1}$ be positive real numbers. Suppose that for some random variables W_n we have that W_n converges in law to a non-trivial random variable X and $a_n W_n$ also converges in law to a non-trivial random variable Y. Then also a_n converges.*

Proof. First of all, notice that a_n has to be bounded and bounded away from zero. Indeed, if $a_{n_k} \to 0$ along some subsequence, then the characteristic function of $\mathbb{E}\exp(i\lambda a_{n_k} W_{n_k})$ would converge to 1 by Dominated Convergence Theorem, and thus the limit of $a_n W_n$ would be trivial. Similarly if $a_{n_k} \to \infty$, we would have that $a_n W_n$ does not converge in law.

Thus a_n has convergent subsequences. Consider such subsequence a_{n_k} converging to a_0. Then, again by using pointwise convergence and dominated convergence theorem, we see that $\mathbb{E}\exp(i\lambda a_{n_k} W_{n_k})$ converges to $\mathbb{E}\exp(i\lambda a_0 X)$. As on the other hand $a_{n_k} W_{n_k}$ also converges to some variable Y, we have that $\mathbb{E}\exp(i\lambda a_0 X) = \mathbb{E}\exp(i\lambda Y)$ for all $\lambda \in \mathbb{R}$. But now if there was a different subsequential limit b_0, we would obtain $\mathbb{E}\exp(i\lambda a_0^{-1} Y) = \mathbb{E}\exp(i\lambda b_0^{-1} Y)$ for all $\lambda \in \mathbb{R}$. This can be rewritten as $\mathbb{E}\exp(i\lambda Y) = \mathbb{E}\exp(i\lambda a_0 b_0^{-1} Y)$, which by iterating implies $\mathbb{E}\exp(i\lambda Y) = \mathbb{E}\exp(i\lambda (a_0/b_0)^n)$, which again converges to 1. Thus as Y is non-trivial we obtain a contradiction. Thus all convergent subsequences have a unique limit and hence the sequence a_n converges. □

□

Thus there are many laws and many Central limit theorems, each associated to one of the laws. Our Gaussian law is one of these laws, and it comes out that it's a special one:

Proposition 1.7. *The Gaussian law is the only stable law with finite variance.*

Why should such a statement be true? A first hint comes from the following observation:

Lemma 1.8. *If a strictly stable law μ has finite variance, then $b_n = n^{1/2}$.*

Proof. By the definition of stability, for all $n \in \mathbb{N}$ there is some $b_n > 0$ such that if $X_1, \dots, X_n \sim \mu$ are i.i.d, then $b_n^{-1}(X_1 + \cdots + X_n) \sim \mu$. In particular the variance of two sides has to agree. But all the X_i are independent, and hence the variance of their sum is equal to the sum of their variances. Thus as all $X_i \sim \mu$, we have that

$$\mathrm{Var}(b_n^{-1}(X_1 + \cdots + X_n)) = b_n^{-2} n \mathrm{Var}(\mu).$$

As this has to equal $\mathrm{Var}(\mu)$, we deduce that $b_n = n^{1/2}$. $\square$

Given this it seems reasonable that there would be only one finite variance stable law, indeed one could hope that the set of equalities in distribution $n^{-1/2} \sum_{i=1}^n X_i \sim \mu$ with $X_i \sim \mu$ i.i.d. could have only one solution up to scaling. Whereas this is true, proving this is not quite trivial. It can be done in different ways, but maybe the most natural way is to see it as a consequence of the CLT:

Theorem 1.9 (Central limit theorem)**.** *Let $X_1, X_2, \dots$ be i.i.d. random variables of finite variance σ^2. Then $n^{-1/2} \sum_{i=1}^n (X_i - \mathbb{E}X_i)$ converges in law to $N(0, \sigma^2)$.*

The theorem implies the proposition basically in the same way as we obtained CLTs for stable laws:

Proof of Proposition using CLT. Let μ be a strictly stable law of finite variance. Then we can take $X_1, X_2, \dots \sim \mu$ to be i.i.d. As μ is stable and by the lemma just above $b_n = n^{1/2}$, we have that $n^{-1/2} \sum_{i=1}^n X_i \sim \mu$ for all $n \geq 1$ and in particular these laws converge to μ. But by the CLT these partial sums also converge to the Gaussian law and thus μ has to be the Gaussian law. $\square$