SMSTC (2007/08)

Probability

www.smstc.ac.uk

Contents

SMSTC (2007/08)
Probability
Lecture 5: Important special distributions

[a]

www.smstc.ac.uk

Contents

5.1 Binomial, Poisson, and multinomial

Consider the coin tossing experiment, i.e. consider a sequence $\xi_1, \xi_2, \dots$ of i.i.d. random variables with

$$\mathbf{P}(\xi_1 = 1) = p, \quad \mathbf{P}(\xi_1 = 0) = 1 - p.$$

The law of

$$S_n = \xi_1 + \cdots + \xi_n$$

is called Binomial with parameters n and p. From this we have

$$\mathbf{E}S_n = np, \quad \operatorname{var} S_n = n \operatorname{var} \xi_1 = np(1-p),$$

and, since $\xi_{n+1} + \xi_{n+m}$ is Binomial with parameters m and p, and independent of S_n, we have that the sum of two independent Binomial random variables with the same p is again Binomial. We also have

$$\mathbf{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

and, from the Binomial theorem,

$$\mathbf{E}e^{\theta S_n} = (pe^\theta + 1 - p)^n.$$

EXERCISE 1. Compute the moment generating function of $n^{-1/2}(S_n - \mathbf{E}S_n)$ and show that, as $n \to \infty$, it converges to the a function of the form $e^{-c\theta^2}$. (This is moment generating function of a Gaussian random variable).

EXERCISE 2. Use Chernoff's inequality to estimate $\mathbf{P}(S_n > n(p + x))$ for $x > 0$.

By letting p vary with n and taking limits we obtain a different fundamental law, the Poisson law. Specifically,

Lemma 5.1. *If $p_n = \frac{\lambda}{n} + o(1/n)$, as $n \to \infty$, then*

$$\mathbf{P}(S_n = k) \to \frac{\lambda^k}{k!} e^{-\lambda},$$

for all $k = 0, 1, 2, \ldots$.

Proof Use Stirling's formula. □

This is the Poisson law with parameter λ. The Poisson law is fundamental when, roughly speaking, we deal with independent rare events.

Let X be a Poisson random variable. Then We have

$$\mathbf{E}e^{\theta X} = \sum_{k=0}^{\infty} \frac{(\lambda e^{\theta})^k}{k!} e^{-\lambda} = e^{\lambda(e^{\theta}-1)}.$$

Differentiating a couple of times, we find

$$\mathbf{E}X = \lambda, \quad \operatorname{var} X = \lambda.$$

Furthermore,

Lemma 5.2. *If $X_1, X_2, \ldots$ are independent Poisson random variables with parameters $\lambda_1, \lambda_2, \ldots$ such that $\sum \lambda_k < \infty$ then $\sum_k X_k$ is Poisson with parameter $\sum_k \lambda_k$.*

Proof We prove this for a finite number n of random variables (which is enough by the way we construct a probability on an infinite product). The moment generating function of $X_1 + \ldots + X_n$ is

$$\prod_{k=1}^{n} e^{\lambda_k (e^{\theta}-1)} = e^{(e^{\theta}-1) \sum_{k=1}^{n} \lambda_k}$$

and this is the moment generating function of a Poisson law with parameter $\sum_k \lambda_k$. □

Next we consider conditional probabilities:

Lemma 5.3. *Suppose that X_1, X_2 are independent Poisson with parameters λ_1, λ_2. Then, conditional on $X_1 + X_2 = n$, we have that X_1 is Binomial with parameters n, $\lambda_1/(\lambda_1 + \lambda_2)$.*

Proof Elementary conditioning: For $0 \le k \le n$,

$$\begin{aligned}
\mathbf{P}(X_1 = k | X_1 + X_2 = n) &= \frac{\mathbf{P}(X_1 = k, X_2 = n-k)}{\mathbf{P}(X_1 + X_2 = n)} \\
&= \frac{\frac{\lambda_1^k}{k!} e^{-\lambda_1} \; \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2}}{\frac{(\lambda_1+\lambda_2)^n}{n!} e^{-\lambda_1-\lambda_2}} \\
&= \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}.
\end{aligned}$$

□

Generalising this, we have:

Lemma 5.4. *Suppose that $X_1, X_2, \dots, X_d$ are independent Poisson with parameters $\lambda_1, \lambda_2, \dots, \lambda_d$, respectively. Then, conditionally on $\sum_k X_k = n$, the random vector $(X_1, \dots, X_d)$ has law given by*

$$\mathbf{P}(X_1 = n_1, \dots, X_d = n_d \mid \sum_k X_k = n) = \binom{n}{n_1, \dots, n_d} \left(\frac{\lambda_1}{\lambda}\right)^{n_1} \cdots \left(\frac{\lambda_d}{\lambda}\right)^{n_d},$$

where $(n_1, \dots, n_d)$ are nonnegative integers with sum equal to n, and $\lambda = \sum_k \lambda_k$, and where

$$\binom{n}{n_1, \dots, n_d} = \frac{n!}{n_1! \cdots n_d!}$$

EXERCISE 3. Show this.

The symbol $\binom{n}{n_1, \dots, n_d}$ is the multinomial coefficient since it appears in the algebraic identity known as multinomial theorem:

$$(x_1 + \cdots + x_d)^n = \sum \binom{n}{n_1, \dots, n_d} x_1^{n_1} \cdots x_d^{n_d}.$$

The sum is taken over all nonnegative integers $(n_1, \dots, n_d)$, with sum equal to n.
The random variable $(X_1, \dots, X_d)$ with values in $\mathbb{Z}_+^d$ is said to have a multinomial distribution with parameters $d, n, p_1, \dots, p_d$ (where $p_1 + \cdots + p_d = 1$, so one of them is superfluous) if

$$\mathbf{P}(X_1 = n_1, \dots, X_d = n_d) = \binom{n}{n_1, \dots, n_d} p_1^{n_1} \cdots p_d^{n_d},$$

where $(n_1, \dots, n_d)$ are nonnegative integers with sum equal to n,
Of course, a multinomial distribution with parameters $2, n, p, 1-p$ is a binomial distribution with parameters n, p.

5.2 Thinning

Suppose an urn contains n balls. There are d colours available. Let the colours be denoted by $c_1, c_2, \dots, c_d$. To each ball assign colour c_i with probability p_i, independently from ball to ball. Let S_n^i be the number of balls that have colour c_i, $1 \le i \le d$. It is easy to see that $(S_n^1, \dots, S_n^d)$ has a multinomial law:

$$\mathbf{P}(S_n^1 = k_1, \dots, S_n^d = k_d) = \binom{n}{k_1, \dots, k_d} p_1^{k_1} \cdots p_d^{k_d},$$

where $(k_1, \dots, k_d)$ are nonnegative integers summing up to n. Clearly,

$$S_n^1 + \cdots S_n^d = n,$$

so the random variable $S_n^1, \dots, S_n^d$ cannot be independent. The question is:

> Suppose that the number of balls is itself a random variable, independent of everything else. Is there a way to choose the law of this random variable so that the above variables are independent?

To put the problem in precise mathematical terms, let $\xi_1, \xi_2, \dots$ be i.i.d. random colours, i.e. random variables taking values in $\{c_1, \dots, c_d\}$ such that

$$\mathbf{P}(\xi_1 = c_i) = p_i, \quad 1 \le i \le d.$$

Let

$$S_n^i := \sum_{k=1}^{n} \mathbf{1}(\xi_k = c_i).$$

(In physical terms, S_n^i denotes precisely what we talked about earlier using a more flowery language.) Now, independent of the sequence $\xi_1, \xi_2, \ldots$, let N be a random variable with values in $\mathbb{Z}_+$. The problem is to find its law so that

$$S_N^1, \ldots, S_N^d \text{ are independent random variables.} \qquad (\star)$$

It turns out that this is a characterising property of the Poisson law. We will contend ourselves by proving one direction:

Lemma 5.5. *If N is Poisson then $(\star)$ holds. Moreover, if N has expectation λ, then S_N^i is also Poisson with expectation λp_i.*

EXERCISE 4. Prove the last lemma.

5.3 Geometric

A random variable X with values in $\mathbb{N}$ is geometric if it has the memoryless property:

> For each $k \in \mathbb{N}$, the conditional distribution of $X - k$ given $\{X > k\}$ is the same as the distribution of X:
>
> $$\mathbf{P}(X - k = n | X > k) = \mathbf{P}(X = n).$$

Think of X as a random time, e.g. the day on which a certain volcano will erupt. The property above says that if by day k the volcano has not erupted then the remaining time $X - k$ has the same law as X, no matter how large k is.

Lemma 5.6. *If $q = \mathbf{P}(X > 1)$ then*

$$\mathbf{P}(X > k) = q^k, \quad k \in \mathbb{N}$$

and

$$\mathbf{P}(X = k) = pq^{k-1},$$

where $p = 1 - q$.

Proof By the property of X,

$$\mathbf{P}(X > k + n | X > k) = \mathbf{P}(X > n),$$

for all k, n, which means that

$$\mathbf{P}(X > k + 1) = \mathbf{P}(X > k)\mathbf{P}(X > 1).$$

Iterating this, we find

$$\mathbf{P}(X > k) = \mathbf{P}(X > 1)^k, \quad k = 0, 1, \ldots.$$

□

People refer to X as geometric with parameter p. The terminology is not standard because other people refer to $X - 1$ (which also has the memoryless property but takes values in $\mathbb{Z}_+$) as geometric with parameter q. A matter of taste, really.

It is easy to see that

Lemma 5.7. *If X is geometric in $\mathbb{N}$ with $\mathbf{P}(X=1)=p$ then*

$$\mathbf{E}X = 1/p, \quad \operatorname{var} X = (1-p)/p^2, \quad \mathbf{E}e^{\theta X} = \frac{pe^\theta}{1-(1-p)e^\theta}.$$

EXERCISE 5. Do all that.

A concrete way to get a geometric random variable is by considering $\xi_1, \xi_2, \ldots$ to be i.i.d. with

$$\mathbf{P}(\xi_1 = 1) = 1 - \mathbf{P}(\xi_1 = 0) = p$$

and by letting

$$X = \inf\{k \geq 1 : \ \xi_k = 1\}.$$

We have $\mathbf{P}(X < \infty) = 1$, so

$$X = \min\{k \geq 1 : \ \xi_k = 1\}$$

and

$$\mathbf{P}(X > k) = \mathbf{P}(\xi_1 = \cdots = \xi_k = 0) = (1-p)^k,$$

as required.
We have that

Lemma 5.8. *If $X_1, X_2, \ldots, X_d$ are independent and geometric then $X = \min(X_1, \ldots, X_d)$ is geometric.*

Proof

$$\begin{aligned}\mathbf{P}(X > k) &= \mathbf{P}(X_1 > k, \ldots, X_d > k) \\ &= \mathbf{P}(X_1 > k) \cdots \mathbf{P}(X_d > k) \\ &= q_1^k \cdots q_d^k = (q_1 \cdots q_d)^k.\end{aligned}$$

□

EXERCISE 6. Let X, Y be independent and geometric. Show that

$$\mathbf{P}(X - Y > n | X > Y) = \mathbf{P}(X > n),$$

for all n, and interpret the formula.

5.4 Uniform

We have already seen, in detail, how to construct a uniform random variable, from first principles. Recall that U is uniform in the interval $[0,1]$ if $\mathbf{P}(U \leq x) = x$, $0 \leq x \leq 1$. More generally,

> X is uniform in $[a,b]$ if, for all intervals I, the probability $\mathbf{P}(X \in I)$ is proportional to the length of I.

Of course, if X is uniform in $[a,b]$, then $cX + d$ is uniform in the interval with endpoints $ca + d$ and $cb + d$.
Recall that if F is a distribution function and U is uniform in $[0,1]$ then $F^{-1}(U)$ is a random variable with distribution function F.

Lemma 5.9. *Let $p_1, \dots, p_d$ be positive numbers adding up to 1. Split the interval $[0,1]$ into consecutive intervals $I_1, \dots, I_d$ of lengths $p_1, \dots, p_d$, respectively. Let $U_1, \dots, U_n$ be i.i.d. uniform in $[0,1]$. Let*

$$S_n^i = \sum_{k=1}^n \mathbf{1}(U_k \in I_i), \quad 1 \le i \le d.$$

Then $(S_n^1, \dots, S_n^d)$ has a multinomial law. In particular, S_n^i is Binomial with parameters n, p_i.

EXERCISE 7. Show this last lemma.

EXERCISE 8. Let $U_1, \dots, U_d$ be i.i.d. uniform in $[0,1]$. Compute the probability $\mathbf{P}(U_1 < U_2 < \cdots < U_d)$.

EXERCISE 9. Consider a stick of length 1 and break it into 3 pieces, by choosing the two break points at random. Find the probability that the 3 smaller sticks can be joined to form a triangle.

EXERCISE 10. Pick a random variable U_1 uniform in $[0,1]$. Let U_2 be the midpoint of the interval $[0, U_1]$ or of $[U_1, 1]$, with equal probability. Continue in the same manner and define U_3 to be the midpoint of $[0, U_2]$ or of $[U_2, 1]$, with equal probability. Show that the $x \mapsto \lim_{n\to\infty} \mathbf{P}(U_n \le x)$ is continuous but not absolutely so.

5.5 Exponential

A random variable T with values in $\mathbb{R}_+$ is exponential if it has the memoryless property:

For all $t, s > 0$,

$$\mathbf{P}(T - t > s | T > t) = \mathbf{P}(T > s).$$

Lemma 5.10. *If T is exponential then there is $\lambda > 0$ such that*

$$\mathbf{P}(T > t) = e^{-\lambda t}, \quad t \ge 0.$$

Proof Implicit in the definition is that $\mathbf{P}(T > t) > 0$ for all t. Hence $\alpha := \mathbf{P}(T > 1) \in (0,1)$. We have

$$\mathbf{P}(T > t + s) = \mathbf{P}(T > t)\mathbf{P}(T > s)$$

for all t, s. Using this and induction, we have that, for all $n \in \mathbb{N}$,

$$\mathbf{P}(T > nt) = \mathbf{P}(T > t)^n.$$

This gives that, for all $m \in \mathbb{N}$,

$$\mathbf{P}(T > 1 = m(1/m)) = \mathbf{P}(T > 1/m)^m$$

and so $\mathbf{P}(T > 1/m) = \mathbf{P}(T > 1)^{1/m}$. Letting $t = 1/m$ in the pre-last display, we have

$$\mathbf{P}(T > n/m) = \alpha^{n/m}.$$

Now, for $t > 0$, let $q_1 > q_2 > \dots$ be rational numbers with $\inf\{q_1, q_2, \dots\} = t$. Then

$$\mathbf{P}(T > t) = \mathbf{P}(\cup_k \{T > q_k\}) = \sup_k \mathbf{P}(T > q_k) = \sup_k \alpha^{q_k} = \alpha^{\inf_k q_k} = \alpha^t.$$

Since $\alpha < 1$, we have $\lambda := -\log \alpha > 0$. □

We say that T is exponential with parameter (rate) λ.

It is easy to see that

Lemma 5.11. *If T is exponential with rate λ then T has density*

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

Also,

$$\mathbf{E}e^{\theta T} = \frac{\lambda}{\lambda - \theta},$$

defined for all $\theta < \lambda$, and

$$\mathbf{E}T = 1/\lambda, \quad \operatorname{var} T = 1/\lambda^2.$$

Proof Note that

$$\int_0^t f(s) ds = 1 - e^{-\lambda t},$$

showing that f is a density for T. The rest are trivial. □

Lemma 5.12. *Let $T_1, T_2, \ldots, T_d$ be independent exponential random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_d$, respectively. Then* $\min(T_1, \ldots, T_d)$ *is exponential with parameter $\lambda_1 + \cdots + \lambda_d$.*

Proof

$$\mathbf{P}(\min(T_1, \ldots, T_d) > t) = \mathbf{P}(T_1 > t) \cdots \mathbf{P}(T_d > t).$$

□

Whereas an exponential is the natural analogue of a geometric, in that they are both memoryless, the former also enjoys the important scaling property:

If T is exponential with rate λ then, for any $c > 0$, cT is exponential with rate λ/c,

and this is obvious.

EXERCISE 11. Let $T_1, T_2, \ldots, T_d$ be independent exponential random variables all with rate 1. Show that

$$\text{law of } \max(T_1, \ldots, T_d) = \text{ law of } \left(T_1 + \frac{T_2}{2} + \cdots + \frac{T_d}{d}\right).$$

Another relation between geometric and exponential is the following: Let p be very small. Let X be geometric with parameter p. Consider a scaling of X by p, i.e. the random variable pX which takes values $p, 2p, 3p, \ldots$. Then the law of pX converges to an exponential law with rate 1:

Lemma 5.13 (Rényi)**.** *For X geometric with parameter p,*

$$\lim_{p \to 0} \mathbf{P}(pX > t) = e^{-t}, \quad t > 0.$$

Proof $\mathbf{P}(pX > t) = \mathbf{P}(X > t/p)$, and, since X is an integer,

$$\mathbf{P}(X > t/p) = \mathbf{P}(X \geq \lceil t/p \rceil) = (1-p)^{\lceil t/p \rceil + 1}.$$

Since $\lim_{n \to \infty}(1 + n^{-1})^n = e$, we have that the last expression converges to e^{-t} as $p \to 0$. □

5.6 Gaussian (normal) variables

5.6.1 The Gaussian law

The motivation of the Gaussian probability comes from the central limit theorem (which, long time ago, was known as the "law of errors"). This was stated, without proof, in 2.3.2. We work heuristically in order to motivate the definitions. Let $S_n = \xi_1 + \cdots + \xi_n$ be the sum of n independent indicator (also known as Bernoulli) random variables ξ_i with $\mathbf{P}(\xi_i = 1) = p$ for all i. Let $\widehat{S}_n = S_n - \mathbf{E}S_n$. Then, as $n \to \infty$, the distribution function of $\widehat{S}_n/\sqrt{n}$ converges to an absolutely continuous distribution function with a famous formula. Let X be a random variable with such a distribution function. Note that

$$\frac{\widehat{S}_{2n}}{\sqrt{n}} = \frac{1}{\sqrt{2}}\frac{\widehat{S}_n}{\sqrt{n}} + \frac{1}{\sqrt{2}}\frac{\widehat{S}'_n}{\sqrt{n}},$$

where $\widehat{S}'_n = \xi_{n+1} + \cdots + \xi_{2n}$ is a random variable with the same law as $\widehat{S}_n$. So the distribution of $\frac{\widehat{S}'_n}{\sqrt{n}}$ will also converge to the distribution of X. Moreover, $\widehat{S}'_n$, $\widehat{S}_n$ are independent. Therefore, if X_1, X_2 are independent random variables with the same law as X, then we expect that

$$X = \frac{X_1 + X_2}{\sqrt{2}}.$$

Even if we do not know what this famous distribution is, it should be such that this "addition law" holds. From this, we can discover its density. One way to do that is by imposing the extra assumption (which is not necessary) that the generating function of X exists for all θ: $M(\theta) = \mathbf{E}e^{\theta X}$. Then

$$M(\theta) = M(\theta/\sqrt{2})^2.$$

Letting $\theta = \sqrt{\eta}$ and taking logarithms, we have

$$\log M(\sqrt{\eta}) = 2\log M(\sqrt{\eta/2}).$$

So, if we temporarily let $m(\eta) = \log M(\sqrt{(\eta)})$ we have

$$m(\eta) = 2m(\eta/2).$$

So $m(0) = 0$ and with some work, we can actually find that the only continuous function satisfying the latter is linear: $m(\eta) = c\eta$. (This should be geometrically obvious.) Hence

$$M(\theta) = e^{c\theta^2}.$$

We know that the moments of X are given by the derivatives of M at 0. We find

$$M'(0) = 0, \quad M''(0) = 2c.$$

So $\mathbf{E}X = 0$ (as it should), while $\mathbf{E}X^2 = 2c$. So $c > 0$. Since $\mathbf{E}X = 0$, the second moment is the variance and we customarily denote it by σ^2. We arrive at

$$M(\theta) = e^{\frac{1}{2}\sigma^2\theta^2}.$$

We know that there is only one probability distribution with a given moment generating function. Instead of trying to figure out which one it is, let us reveal the result and then just verify that it is correct. We claim that the probability distribution corresponding to the last M is absolutely continuous with density

$$f(x) = Ce^{-x^2/2\sigma^2}.$$

Here, C is such that $\int_{\mathbb{R}} f(x)dx = 1$. This is the famous NORMAL OR GAUSSIAN DENSITY with mean 0 and variance σ^2. It is called standard if $\sigma^2 = 1$.

EXERCISE 12. Show that $\int_{-\infty}^{\infty} e^{\theta x} f(x)dx = e^{\frac{1}{2}\sigma^2\theta^2}$. (Hint: Complete the square and use the definition of C.)

Now, to find C is a very important thing. It is based on the following:

Lemma 5.14.

$$\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}.$$

(Liouville said that a Mathematician is someone for whom this integral is obvious.)

EXERCISE 13. Use Fubini's theorem to write $\left(\int_{\mathbb{R}} e^{-x^2} dx\right)^2 = \int_{\mathbb{R}^2} e^{-x^2-y^2} dxdy$ and do the latter integral using polar coördinates.

Using this we find that

$$C = \frac{1}{\sqrt{2\pi\sigma^2}}.$$

Therefore, the standard normal density is

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

We write $\mathcal{N}(0,1)$ for the law of a random variable X with standard normal density. We write $\mathcal{N}(\mu,\sigma^2)$ for the law of $\sigma X + \mu$.

EXERCISE 14. Show that a density for $\mathcal{N}(\mu,\sigma^2)$ is

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

EXERCISE 15. Show that if $X_i, i = 1,\ldots,d$ are independent and X_i having law $\mathcal{N}(\mu_i,\sigma_i^2)$ then $\sum_{i=1}^d X_i$ has law $\mathcal{N}(\sum_i \mu_i, \sum_i \sigma_i^2)$. Therefore linear combinations of independent normal variables are normal.

5.6.2 The multidimensional Gaussian random vector

We now pass on to defining a Gaussian (or normal) random variable in $\mathbb{R}^d$.

> We say that $(X_1,\ldots,X_d)$ is Gaussian in $\mathbb{R}^d$ if, for all $a_1,\ldots,a_d \in \mathbb{R}$, the random variable $a_1X_1 + \cdots + a_dX_d$ is normal.

The next lemma shows what the moment generating function of a normal vector is:

Lemma 5.15. *If $(X_1,\ldots,X_d)$ is Gaussian vector with*

$$\mu_j = \mathbf{E}X_j, \quad r_{jk} = \operatorname{cov}(X_j,X_k),$$

then

$$\mathbf{E}\exp\sum_{j=1}^d \theta_j X_j = \exp\left\{\sum_{j=1}^d \mu_j\theta_j + \frac{1}{2}\sum_{j=1}^d\sum_{k=1}^d r_{jk}\theta_j\theta_k\right\}.$$

Proof By definition $\sum_{j=1}^d \theta_j X_j$ should be normal, i.e. have a law $\mathcal{N}(\mu,\sigma^2)$, for some μ,σ^2. We have

$$\mu = \mathbf{E}\sum_{j=1}^d \theta_j X_j = \sum_{j=1}^d \mu_j\theta_j, \quad \sigma^2 = \operatorname{cov}\sum_{j=1}^d \theta_j X_j = \sum_{j=1}^d\sum_{k=1}^d \theta_j\theta_k \operatorname{cov}(X_j,X_k).$$

□

Since the moment generating function of a Gaussian vector is the exponential of a quadratic form, there is no better way to express it other than using Linear Algebra. To this end, we think of the elements x of $\mathbb{R}^d$ as column vectors. We use x' to denote transposition, i.e. the corresponding row when x is a column. And, of course, $(x')' = x$. Consider the mean (column) vector

$$\mu = (\mu_1, \ldots, \mu_d)'$$

and the symmetric covariance matrix

$$R = [r_{jk}].$$

Write also X for the column with entries $X_1, \ldots, X_d$. We then have

$$\mathbf{E}e^{\theta' X} = \exp\left\{\theta'\mu + \frac{1}{2}\theta' R\theta\right\}.$$

This uniquely defines the law of $(X_1, \ldots, X_d)$. This law is denoted by $\mathcal{N}(\mu, R)$, where μ is the mean vector and R the covariance matrix.

Lemma 5.16. *A Gaussian vector $(X_1, \ldots, X_d)$ is absolutely continuous if and only if R is invertible. In this case, its density is given by*

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det(R)}} \exp\left(-\frac{1}{2}(x-\mu)' R^{-1}(x-\mu)\right).$$

Proof Assume $\mu = 0$, to ease notation. Assume R is invertible. It is easily seen that

$$R = \mathbf{E}(XX')$$

where XX' is a $d \times d$ matrix and $\mathbf{E}(XX')$ is the matrix formed by taking the expectations of the entries of XX'. This R has two important properties:
(i) it is symmetric (obviously);
(ii) it is positive semi-definite, i.e. the quadratic form

$$u'Ru = \sum_k \sum_\ell R_{k\ell} u_k u_\ell \geq 0$$

for all values of the variables, positive or negative. The reason for the latter is that $u'Ru$ is the expectation of a non-negative quantity:

$$u'Ru = u'\mathbf{E}XX'u = \mathbf{E}(X'u)'(X'u) = \mathbf{E}(X'u)^2 \geq 0.$$

Standard linear algebra shows that R has exactly d (counting multiplicities) non-negative eigenvalues. (In fact, they are strictly positive, due to the invertibility of R which is tantamount to $\det(R) \neq 0$). Furthermore, the eigenvectors can be chosen to be orthonormal. Letting U be the matrix whose columns are these d orthonormal eigenvectors and Λ the diagonal matrix with the eigenvalues in its diagonal, we have

$$RU = U\Lambda,$$

from the very definition of the eigenvectors. Now

$$U' = U^{-1},$$

hence

$$R = U\Lambda U' = U\Lambda^{1/2}\Lambda^{1/2}U' = PP',$$

with

$$P = U\Lambda^{1/2}.$$

The matrix P is non-singular and is called the square root of R. We now define new random variables Z by

$$X = PZ\,.$$

The thing to observe is that the covariance matrix of Z is

$$\text{cov}(Z) = \mathbf{E}ZZ' = \mathbf{E}P^{-1}XX'P'^{-1} = P^{-1}R'P'^{-1} = P^{-1}PP'P'^{-1} = I,$$

I being the identity matrix. Thus

$$\mathbf{E}e^{\theta' Z} = e^{\theta'\theta/2} = \prod_{j=1}^{d} e^{\theta_j^2/2} = \prod_{j=1}^{d} \mathbf{E}e^{\theta_j Z_j},$$

implying that the components of Z are independent standard normal. Hence the density g of Z is product:

$$g(z) = \prod_j \frac{1}{\sqrt{2\pi}} e^{-z_j^2/2}.$$

Now $X = PZ$, hence its density f is computed easily by

$$f(x) = g(P^{-1}x)/|\det(P)|,$$

which yields the desired formula. □

If R has determinant zero then it is possible to "reduce the dimension" of the random vector X so that the density exists. In fact,

Lemma 5.17. *The support of X is the range of its covariance matrix R.*

Proof Suppose that R has rank r. Then it has $d-r$ eigenvalues at zero, so that the matrix Λ consists of a $d-r$ size block of zeros and the remaining non-zero eigenvalues. Hence now $R = PP'$, where P is a $d \times r$ matrix with rank r. We try again to find Z so that

$$X = PZ,$$

where Z is an r-dimensional random vector with density. If we manage to do this we will finish, since the range of P is the range of R. Observing that $P'P$ is an $r \times r$ non-singular matrix, we pre-multiply by it to get $P'PZ = P'X$, hence $Z = (P'P)^{-1}P'X$. So if we define Z this way, we see that

$$\mathbf{E}ZZ' = (P'P)^{-1}P'RP(P'P)^{-1} = (P'P)^{-1}P'PP'P(P'P)^{-1} = I,$$

i.e. Z is a collection of r independent standard normal variables. The formula for Z is the formula that solves a full-rank over-estimated linear system. We usually write $Z = P^+X$ and call P^+ the pseudo-inverse of P. It remains to show that every linear function of X is a linear function of Z. Let

$$F = \lambda' X a \quad , G = \lambda' PZ$$

with $Z = P^+X$. Consider

$$F - G = \lambda'(X - PZ).$$

Then

$$\mathbf{E}(F-G)^2 = \mathbf{E}\lambda'(X - PZ)(X' - Z'P')\lambda = \mathbf{E}\lambda'(XX' - XZ'P' - PZX' + PZZ'P')\lambda.$$

But

$$\mathbf{E}ZX' = (P'P)^{-1}P'\mathbf{E}XX' = (P'P)^{-1}P'PP' = P'.$$

So

$$\mathbf{E}(F-G)^2 = \lambda'(PP' - PP' - PP' + PP')\lambda = 0.$$

□

Terminology: If X has law $\mathcal{N}(0,R)$ with R having rank r then $\sum_j X_j^2$ is called χ^2 with r degrees of freedom.

5.6.3 Conditional Gaussian law

If it appears that we've done a lot of Linear Algebra, then this is because it is so: Dealing with Gaussian random variables (and processes!) is mostly dealing with Linear Algebra (or Linear Analysis!).

Without proof, we mention the following:

Lemma 5.18. *Let $(X;Y_1,\dots,Y_d)$ be a Gaussian random variable in $\mathbb{R}^{1+d}$. Then the conditional law $\mathbf{P}(X \in \cdot|Y_1,\dots,Y_d)$ is normal with mean $\mathbf{E}(X|Y_1,\dots,Y_d)$ and deterministic covariance matrix. Moreover, $\mathbf{E}(X|Y_1,\dots,Y_d)$ is a linear function of $Y_1,\dots,Y_d)$ and is characterised by the fact that $X - E(X|Y_1,\dots,Y_d)$ is independent of $(Y_1,\dots,Y_d)$.*

EXERCISE 16. Let X, Y be independent standard Gaussian variables. Based on the last part of the lemma above, compute $\mathbf{E}(X+2Y|3X-4Y)$.