1. GAUSSIAN RANDOM VARIABLES

Standard normal: A standard normal or Gaussian random variable is one with density $\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Its distribution function is $\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt$ and its tail distribution function is denoted $\overline{\Phi}(x) := 1 - \Phi(x)$. If X_i are i.i.d. standard normals, then $X = (X_1, \ldots, X_n)$ is called a standard normal vector in $\mathbb{R}^n$. It has density $\prod_{i=1}^{n} \varphi(x_i) = (2\pi)^{-n/2} \exp\{-|\mathbf{x}|^2/2\}$ and the distribution is denoted by γ_n, so that for every Borel set A in $\mathbb{R}^n$ we have $\gamma_n(A) = (2\pi)^{-n/2} \int_A \exp\{-|\mathbf{x}|^2/2\} d\mathbf{x}$.

Exercise 1. [Rotation invariance] If $P_{n\times n}$ is an orthogonal matrix, then $\gamma_n P^{-1} = \gamma_n$ or equivalently, $PX \stackrel{d}{=} X$. Conversely, if a random vector with independent co-ordinates has a distribution invariant under orthogonal transformations, then it has the same distribution as cX for some (non-random) scalar c.

Multivariate normal: If $Y_{m\times 1} = \mu_{m\times 1} + B_{m\times n} X_{n\times 1}$ where $X_1, \ldots, X_n$ are i.i.d. standard normal, then we say that $Y \sim N_m(\mu, \Sigma)$ with $\Sigma = BB^t$. Implicit in this notation is the fact that the distribution of Y depends only on Σ and not on the way in which Y is expressed as a linear combination of standard normals (this follows from Exercise 1). It is a simple exercise that $\mu_i = \mathbf{E}[X_i]$ and $\sigma_{i,j} = \mathrm{Cov}(X_i, X_j)$. Since matrices of the form BB^t are precisely positive semi-definite matrices (defined as those $\Sigma_{m\times m}$ for which $\mathbf{v}^t \Sigma \mathbf{v} \geq 0$ for all $\mathbf{v} \in \mathbb{R}^m$), it is clear that covariance matrices of normal random vectors are precisely p.s.d. matrices. Clearly, if $Y \sim N_m(\mu, \Sigma)$ and $Z_{p\times 1} = C_{p\times m} Y + \theta_{p\times 1}$, then $Z \sim N_p(\theta + C\mu, C\Sigma C^t)$. Thus, affine linear transformations of normal random vectors are again normal.

Exercise 2. The random vector Y has density if and only if Σ is non-singular, and in that case the density is

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}\mathbf{y}^t \Sigma^{-1} \mathbf{y}\right\}.$$

If Σ is singular, then X takes values in a lower dimensional subspace in $\mathbb{R}^n$ and hence does not have density.

Exercise 3. Irrespective of whether Σ is non-singular or not, the characteristic function of Y is given by

$$\mathbf{E}\left[e^{i\langle \lambda, Y\rangle}\right] = e^{-\frac{1}{2}\lambda^t \Sigma \lambda}, \text{ for } \lambda \in \mathbb{R}^m.$$

In particular, if $X \sim N(0, \sigma^2)$, then its characteristic function is $\mathbf{E}[e^{i\lambda X}] = e^{-\frac{1}{2}\sigma^2\lambda^2}$ for $\lambda \in \mathbb{R}$.

Exercise 4. If $U_{k\times 1}$ and $V_{(m-k)\times 1}$ are such that $Y^t = (U^t, V^t)$, and we write $\mu = (\mu_1, \mu_2)$ and $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ are partitioned accordingly, then

(1) $U \sim N_k(\mu_1, \Sigma_{11})$.

(2) $U \big|_V \sim N_k(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1/2} V,\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ (assume that Σ_{22} is invertible).

Moments: All questions about a centered Gaussian random vector must be answerable in terms of the covariance matrix. In some cases, there are explicit answers.

Exercise 5. Prove the *Wick formula* (also called *Feynman diagram formula*) for moments of centered Gaussians.

(1) Let $X \sim N_n(0, \Sigma)$. Then, $\mathbf{E}[X_1 \ldots X_n] = \sum_{M \in \mathcal{M}_n} \prod_{\{i,j\} \in M} \sigma_{i,j}$, where $\mathcal{M}_n$ is the collection of all matchings of the set $[n]$ (thus $\mathcal{M}_n$ is empty if n is odd) and the product is over all matched pairs. For example, $\mathbf{E}[X_1 X_2 X_3 X_4] = \sigma_{12}\sigma_{34} + \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23}$.

(2) If $\xi \sim N(0,1)$, then $\mathbf{E}[\xi^{2n}] = (2n-1)(2n-3)\dots(3)(1)$.

Cumulants: Let X be a real-valued random variable with $\mathbf{E}[e^{tX}] < \infty$ for t in a neighbourhood of 0. Then, we can write the power series expansions

$$\mathbf{E}[e^{i\lambda X}] = \sum_{k=0}^{\infty} m_n(X)\frac{\lambda^n}{n!}, \qquad \log \mathbf{E}[e^{i\lambda X}] = \sum_{k=1}^{\infty} \kappa_n[X]\frac{\lambda^n}{n!}.$$

Here $m_n[X] = \mathbf{E}[X^n]$ are the moments while $\kappa_n[X]$ is a linear combination of the first n moments ($\kappa_1 = m_1$, $\kappa_2 = m_2 - m_1^2$, etc). Then κ_n is called the nth cumulant of X. If X and Y are independent, then it is clear that $\kappa_n[X+Y] = \kappa_n[X] + \kappa_n[Y]$.

Exercise 6. (optional). Prove the following relationship between moments and cumulants. The sums below are over partitions Π of the set $[n]$ and $\Pi_1, \dots, \Pi_{\ell_\Pi}$ denote the blocks of Π.

$$m_n[X] = \sum_{\Pi}\prod_i \kappa_{|\Pi_i|}[X], \qquad \kappa_n[X] = \sum_{\Pi}(-1)^{\ell_\Pi - 1}\prod_i m_{|\Pi_i|}[X].$$

Thus $\kappa_1 = m_1$, $\kappa_2 = m_2 - m_1^2$,

Exercise 7. If $\xi \sim N(0,1)$, then $\kappa_1 = 0$, $\kappa_2 = 1$ and $\kappa_n = 0$ for all $n \geq 3$.

The converse of this result is also true and often useful in proving that a random variable is normal. For instance, the theorem below implies that to show that a sequence of random variables converges to normal, it suffices to show that cumulants $\kappa_m[X_n] \to 0$ for all $m \geq m_0$ for some m_0.

Result 8 (Marcinkiewicz)**.** If X is a random variable with finite moments of all orders and $\kappa_n[X] = 0$ for all $n \geq n_0$ for some n_0, then X is Gaussian.

Convergence and Gaussians:

Exercise 9. The family of distributions $N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $0 \leq \sigma^2 < \infty$, is closed under convergence in distribution (for this statement to be valid we include $N(\mu, 0)$ which means δ_μ). Indeed, $N(\mu_n, \sigma_n^2) \xrightarrow{d} N(\mu, \sigma^2)$ if and only if $\mu_n \to \mu$ and $\sigma_n^2 \to \sigma^2$.

A vector space of Gaussian random variables: Let $Y \sim N_m(0, \Sigma)$ be a random vector in some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then, for every vector $\mathbf{v} \in \mathbb{R}^m$, define the random variable $Y_\mathbf{v} := \mathbf{v}^t Y$. Then, for any $\mathbf{v}_1, \dots, \mathbf{v}_j$, the random variables $Y_{\mathbf{v}_1}, \dots, Y_{\mathbf{v}_j}$ are jointly normal. The joint distribution of $\{Y_\mathbf{v}\}$ is fully specified by noting that $Y_\mathbf{v}$ have zero mean and $\mathbf{E}[Y_\mathbf{v} Y_\mathbf{u}] = \mathbf{v}^t \Sigma \mathbf{u}$.

We may interpret this as follows. If Σ is p.d. (p.s.d. and non-singular), then $(\mathbf{v}, \mathbf{u})_\Sigma := \mathbf{v}^t \Sigma \mathbf{u}$ defines an inner product on $\mathbb{R}^m$. On the other hand, the set $L_0^2(\Omega, \mathcal{F}, \mathbf{P})$ of real-valued random variables on Ω with zero mean and finite variance, is also an inner product space under the inner product $\langle U, V \rangle := \mathbf{E}[UV]$. The observation in the previous paragraph is that $\mathbf{v} \to Y_\mathbf{v}$ is an isomorphism of $(\mathbb{R}^m, (\cdot,\cdot)_\Sigma)$ into $L_0^2(\Omega, \mathcal{F}, \mathbf{P})$.

In other words, given any finite dimensional inner-product space $(V, \langle \cdot, \cdot \rangle)$, we can find a collection of Gaussian random variables on some probability space, such that this collection is isomorphic to the given inner-product space. Later we shall see the same for Hilbert spaces[1].

[1]This may seem fairly pointless, but here is one thought-provoking question. Given a vector space of Gaussian random variables, we can multiply any two of them and thus get a larger vector space spanned by the given normal random variables and all pair-wise products of them. What does this new vector space correspond to in terms of the original $(V, \langle \cdot, \cdot \rangle)$?

2. THE GAUSSIAN DENSITY

Recall the standard Gaussian density $\varphi(x)$. The corresponding cumulative distribution function is denoted by Φ and the tail is denoted by $\overline{\Phi}(x) := \int_x^\infty \varphi(t)dt$. The following estimate will be used very often.

Exercise 10. For all $x > 0$, we have $\frac{1}{\sqrt{2\pi}} \frac{x}{1+x^2} e^{-\frac{1}{2}x^2} \le \overline{\Phi}(x) \le \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{1}{2}x^2}$ In particular[2], $\overline{\Phi}(x) \sim x^{-1}\varphi(x)$ as $x \to \infty$. Most often the following simpler bound, valid for $x \ge 1$, suffices.

$$\frac{1}{10x} e^{-\frac{1}{2}x^2} \le \overline{\Phi}(x) \le e^{-\frac{1}{2}x^2}.$$

For $t > 0$, let $p_t(x) := \frac{1}{\sqrt{t}}\varphi(x/\sqrt{t})$ be the $N(0,t)$ density. We interpret $p_0(x)dx$ as the degenerate measure at 0. These densities have the following interesting properties.

Exercise 11. Show that $p_t \star p_s = p_{t+s}$, i.e., $\int_{\mathbb{R}} p_t(x-y)p_s(y)dy = p_{t+s}(x)$.

Exercise 12. Show that $p_t(x)$ satisfies the heat equation: $\frac{\partial}{\partial t} p_t(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2} p_t(x)$ for all $t > 0$ and $x \in \mathbb{R}$.

Remark 13. Put together, these facts say that $p_t(x)$ is the *fundamental solution* to the heat equation. This just means that the heat equation $\frac{\partial}{\partial t} u(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2} u(t,x)$ with the initial condition $u(0,x) = f(x)$ can be solved simply as $u(t,x) = (f \star p_t)(x) := \int_{\mathbb{R}} f(y)p_t(x-y)dy$. This works for reasonable f (say $f \in L^1(\mathbb{R})$).

We shall have many occasions to use the following "integration by parts" formula.

Exercise 14. Let $X \sim N_n(0,\Sigma)$ and let $F : \mathbb{R}^n \to \mathbb{R}$. Under suitable conditions on F (state sufficient conditions), show that $\mathbf{E}[X_i F(X)] = \sum_{j=1}^n \sigma_{ij}\mathbf{E}[\partial_j F(X)]$. As a corollary, deduce the Wick formula of Exercise 5.

Stein's equation: Here we may revert to $t = 1$, thus $p_1 = \varphi$. Then, $\varphi'(x) = -x\varphi(x)$. Hence, for any $f \in C_b^1(\mathbb{R})$, we integrate by parts to get $\int f'(x)\varphi(x)dx = -\int f(x)\varphi'(x)dx = \int f(x)x\varphi(x)dx$. If $X \sim N(0,1)$, then we may write this as

(1) $$\mathbf{E}[(Tf)(X)] = 0 \quad \text{for all } f \in C_b^1(\mathbb{R}), \text{ where } (Tf)(x) = f'(x) - xf(x).$$

The converse is also true. Suppose (1) holds for all $f \in C_b^1(\mathbb{R})$. Apply it to $f(x) = e^{i\lambda x}$ for any fixed $\lambda \in \mathbb{R}$ to get $\mathbf{E}[Xe^{i\lambda X}] = i\lambda\mathbf{E}[e^{i\lambda X}]$. Thus, if $\psi(\lambda) := \mathbf{E}[e^{i\lambda X}]$ is the characteristic function of X, then $\psi'(\lambda) = -\lambda\psi(\lambda)$ which has only one solution, $e^{-\lambda^2/2}$. Hence X must have standard normal distribution.

Digression - central limit theorem: One reason for the importance of normal distribution is of course the central limit theorem. The basic central limit theorem is for $W_n := (X_1 + \ldots + X_n)/\sqrt{n}$ where X_i are i.i.d. with zero mean and unit variance. Here is a sketch of how central limit theorem can be proved using Stein's method. Let $f \in C_b^1(\mathbb{R})$ and observe that $\mathbf{E}[W_n f(W_n)] = \sqrt{n}\mathbf{E}[X_1 f(W_n)]$. Next, write

$$f\left(\frac{X_1 + \ldots + X_n}{\sqrt{n}}\right) \approx f\left(\frac{X_2 + \ldots + X_n}{\sqrt{n}}\right) + \frac{X_1}{\sqrt{n}} f'\left(\frac{X_2 + \ldots + X_n}{\sqrt{n}}\right)$$

where we do not make precise the meaning of the approximation. Let $\hat{W}_n = \frac{X_2 + \ldots + X_n}{\sqrt{n}}$. Then,

$$\mathbf{E}[W_n f(W_n)] \approx \sqrt{n}\mathbf{E}[X_1]\mathbf{E}[f(\hat{W}_n)] + \mathbf{E}[X_1^2]\mathbf{E}[f'(\hat{W}_n)] = \mathbf{E}[f'(\hat{W}_n)].$$

Since $\hat{W}_n \approx W_n$, this shows that $\mathbf{E}[Tf(W_n)] \approx 0$. We conclude that $W_n \approx N(0,1)$.

There are missing pieces here, most important being the last statement - that if a random variable satisfies Stein's equation approximately, then it must be approximately normal. When included, one does get a proof of the standard CLT.

[2] The notation $f(x) \sim g(x)$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.