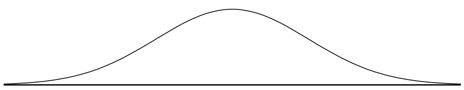
1 Basic notions: finite dimension

1a Gaussian measures on $\mathbb{R}$, or normal distributions

1a1 Definition. The *standard one-dimensional Gaussian measure* γ^1, known also as the standard normal distribution $\mathrm{N}(0,1)$, is defined by

$$\gamma^1(A) = \frac{1}{\sqrt{2\pi}} \int_A \mathrm{e}^{-x^2/2} \, \mathrm{d}x$$

for all measurable $A \subset \mathbb{R}$.



The image of $\mathrm{N}(0,1)$ under a linear map $x \mapsto a + \sigma x$ (where $a \in \mathbb{R}$ and $\sigma \in [0,\infty)$ are parameters) is called the normal distribution $\mathrm{N}(a,\sigma^2)$. In other words,

$$\mathrm{N}(a,\sigma^2)(A) = \gamma^1\big(\{x : a + \sigma x \in A\}\big) \, ;$$

thus, $\mathrm{N}(a,0)$ is a single atom at a, and

$$\mathrm{N}(a,\sigma^2)(A) = \frac{1}{\sqrt{2\pi}\sigma} \int_A \exp\Big(- \frac{(x-a)^2}{2\sigma^2} \Big) \, \mathrm{d}x$$

if $\sigma > 0$.

Like the famous number π, the normal distribution appears here and there, again and again. Some simple examples follow, just for your information (they will not be used).

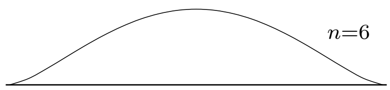
1a2 Example. Let λ_n denote the uniform distribution (in other words, the normalized surface measure) on the sphere $S^{n-1}(\sqrt{n}) = \{x \in \mathbb{R}^n : |x| = \sqrt{n}\}$

and denote by μ_n the corresponding distribution of the first coordinate; that is, $\mu_n(A) = \lambda_n(A \times \mathbb{R}^{n-1})$. Then $\mu_n \to \gamma^1$ as $n \to \infty$ in the sense that

$$\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\gamma^1 \quad \text{as } n \to \infty \tag{1a3}$$

for every bounded continuous $f : \mathbb{R} \to \mathbb{R}$. Moreover, μ_n has a density,

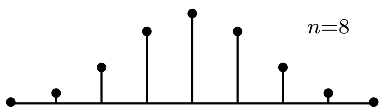
$$\mu_n(A) = \mathrm{const}_n \cdot \int_{A \cap (-\sqrt{n}, \sqrt{n})} \Big(1 - \frac{x^2}{n} \Big)^{\frac{n-3}{2}} \, \mathrm{d}x \,,$$



and the density converges to the normal density. For details see Exercise 2.1.40 in a book by D. Stroock,[1] who cites F. Mehler (1866) and notes that "in terms of statistical mechanics, this result can be interpreted as a derivation of the Maxwell distribution of velocities for a gas of free particles (...)". See also [2], Exercise 2.12 in Sect. 2.3.

1a4 Example. Let λ_n denote the uniform distribution (that is, the normalized counting measure) on the finite set $\{-1, +1\}^n$, and denote by μ_n the corresponding distribution of the (linear) function $(x_1 + \dots + x_n)/\sqrt{n}$;

$$\mu_n\big(\{k/\sqrt{n}\}\big) = 2^{-n} \frac{n!}{\big(\frac{n-k}{2}\big)!\big(\frac{n+k}{2}\big)!} \quad \text{for } k = -n, -n+2, \dots, n \,.$$



Then $\mu_n \to \gamma^1$ (in the sense of (1a3)), which is the De Moivre (1733) - Laplace (1770s) theorem, the simplest special case of Central Limit Theorem.

1a5 Example. Let λ_n denote the uniform distribution (that is, the normalized Lebesgue measure) on the cube $[-1, +1]^{2n+1}$, and μ_n the corresponding distribution of the (nonlinear) function $(x_1, \dots, x_{2n+1}) \mapsto \sqrt{2n}\, x_{(n+1)}$; here $(x_{(1)}, \dots, x_{(2n+1)})$ is the increasing rearrangement of $(x_1, \dots, x_{2n+1})$. It appears (see [2], Example 2.2.6) that $\mu_n \to \gamma^1$. In fact, μ_n has the density $\mathrm{const}_n \cdot \big(1 - \frac{x^2}{2n}\big)^n$ for $|x| < \sqrt{2n}$.

[1] Daniel W. Stroock, "Probability theory, an analytic view", Cambridge 1993.

1a6 Example. Let $\lambda_{n,p}$ denote the product measure

$$\lambda_{n,p}\bigl(\{(x_1,\dots,x_n)\}\bigr) = p^{x_1+\dots+x_n}(1-p)^{n-x_1-\dots-x_n} = p^k(1-p)^{n-k}$$

on $\{0,1\}^n$; define λ_n on $[0,1]\times\{0,1\}^n$ by

$$\lambda_n\bigl(A\times\{x\}\bigr) = \int_A \lambda_{n,p}\bigl(\{(x_1,\dots,x_n)\}\bigr)\,\mathrm{d}p = \int_A p^k(1-p)^k\,\mathrm{d}p\,.$$

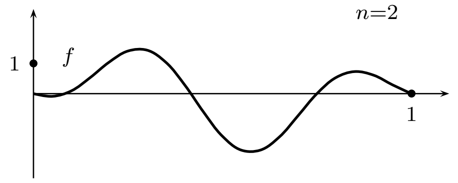
(It means tossing n times an unfair coin with parameter p chosen at random, uniformly on $[0,1]$.) The conditional distribution of p given $x\in\{0,1\}^n$ is

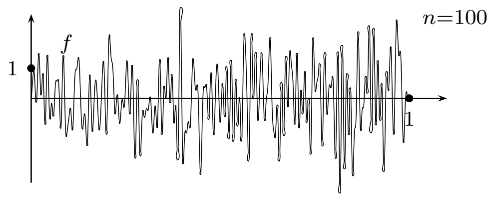
$$A \mapsto \frac{\lambda_n\bigl(A\times\{x\}\bigr)}{\lambda_n\bigl([0,1]\times\{x\}\bigr)} = \mathrm{const}_n\cdot\int_A p^k(1-p)^{n-k}\,\mathrm{d}p\,.$$

Denote by μ_{2n} the conditional distribution of $2\sqrt{2n}(p-0.5)$ given $x\in\{0,1\}^{2n}$ such that $x_1+\dots+x_{2n}=n$. It appears that $\mu_n\to\gamma^1$, which is the simplest case of asymptotic normality in Bayesian (and non-Bayesian) statistics. In fact, μ_n has the density $\mathrm{const}_n\cdot\bigl(1-\frac{x^2}{2n}\bigr)^n$ for $|x|<\sqrt{2n}$.

1a7 Example. Consider 2^{2n} trigonometric polynomials of the form

$$f(x) = \frac{1}{\sqrt{n}}\bigl(\pm\cos(2\pi\omega)\pm\sin(2\pi\omega)\pm\dots\pm\cos(2\pi n\omega)\pm\sin(2\pi n\omega)\bigr)\,;$$





each f has its distribution μ_f,

$$\mu_f(A) = \operatorname{mes} f^{-1}(A) = \int_0^1 \mathbf{1}_A(f(\omega))\,\mathrm{d}\omega\,.$$

(By 'mes' I denote Lebesgue measure.) For most (but not all) of these f, μ_f is close to γ^1 (provided that n is large).

The so-called central limit problem for convex bodies, not even formulated here, is deeper.[1]

Among all probability measures μ on $\mathbb{R}$ such that $\int x\,\mu(\mathrm{d}x)=0$ and $\int x^2\,\mu(\mathrm{d}x)=1$, γ^1 minimizes the Poincare constant

$$\frac12\sup_f\frac{\iint|f(x)-f(y)|^2\,\mu(\mathrm{d}x)\mu(\mathrm{d}y)}{\int|f'(x)|^2\,\mu(\mathrm{d}x)}$$

(see [1], 1.10.2 and 1.6.4) and maximizes the entropy (see [1], 1.10.23).

[1] M. Antilla, K. Ball, I. Perissinaki, "The central limit problem for convex bodies", Trans. Amer. Math. Soc. **355** (2003), 4723–4735.

1b Gaussian measures on $\mathbb{R}^n$, or multinormal distributions

1b1 Definition. The *standard n-dimensional Gaussian measure* γ^n, known also as the standard multinormal distribution, is defined by

$$\gamma^n(A) = (2\pi)^{-n/2} \int_A \mathrm{e}^{-|x|^2/2} \, \mathrm{d}x$$

for all measurable $A \subset \mathbb{R}^n$.

Here $x = (x_1, \dots, x_n)$, $|x| = \sqrt{x_1^2 + \dots + x_n^2}$ and $\mathrm{d}x = \mathrm{d}x_1 \dots \mathrm{d}x_n$. Note that $\gamma^n = \gamma^1 \times \dots \times \gamma^1$, that is,

$$\int \cdots \int f_1(x_1) \dots f_n(x_n) \, \gamma^n(\mathrm{d}x_1 \dots \mathrm{d}x_n) = \\ \Big(\int f_1(x) \gamma^1(\mathrm{d}x) \Big) \dots \Big(\int f_n(x) \gamma^1(\mathrm{d}x) \Big)$$

for measurable $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$ such that the latter integrals converge. In other words, γ^n makes $x_1, \dots, x_n$ independent, each distributed $\mathrm{N}(0,1)$.

The image γ of γ^n under a linear map $L : \mathbb{R}^n \to \mathbb{R}^n$ is, by definition, a centered (or zero-mean) multinormal distribution on $\mathbb{R}^n$. If $\dim L(\mathbb{R}^n) = 0$ then γ is a single atom at 0. If $1 \le \dim L(\mathbb{R}^n) \le n-1$ then γ is singular. If $\dim L(\mathbb{R}^n) = n$, that is, L is invertible, then γ has a density of the form $x \mapsto \mathrm{const} \cdot \exp\big(-Q(x)\big)$, where $Q : \mathbb{R}^n \to [0, \infty)$ is a quadratic form. In the latter case we say that γ is nondegenerate.

An arbitrary (not just centered) multinormal distribution is, by definition, the image of a centered multinormal distribution under a shift $x \mapsto x + a$.

The two-dimensional case is of special interest.

First, the density of γ^2 is easy to integrate in polar coordinates,

$$\iint \frac{1}{2\pi} \mathrm{e}^{-(x_1^2+x_2^2)/2} \, \mathrm{d}x_1 \mathrm{d}x_2 = \frac{1}{2\pi} \iint \mathrm{e}^{-r^2/2} r \, \mathrm{d}r \mathrm{d}\varphi = 1 \, ,$$

which verifies not only the constant $\frac{1}{2\pi}$ for $n=2$ but also $\frac{1}{\sqrt{2\pi}}$ for $n=1$, thus, $(2\pi)^{-n/2}$ for any n.

Second, γ^2 is invariant under rotations $(x_1, x_2) \mapsto (x_1 \cos\alpha - x_2 \sin\alpha, \, x_1 \sin\alpha + x_2 \cos\alpha)$. Therefore the distribution of $x_1 \cos\alpha - x_2 \sin\alpha$ does not depend on α, it is $\mathrm{N}(0,1)$ for any α; we get

$$a x_1 + b x_2 \sim \mathrm{N}(0, a^2 + b^2) \, ;$$
$$\mathrm{N}(0, a^2) * \mathrm{N}(0, b^2) = \mathrm{N}(0, a^2 + b^2) \, .$$

Also, $x_1 \cos\alpha - x_2 \sin\alpha$ and $x_1 \sin\alpha + x_2 \cos\alpha$ are independent. More generally, for $a, b \in \mathbb{R}^2$,

$$\langle x, a\rangle \text{ and } \langle x, b\rangle \text{ are independent whenever } \langle a, b\rangle = 0\,.$$

That is, $\langle a, b\rangle = 0$ implies

$$\iint f(\langle x,a\rangle) g(\langle x,b\rangle)\,\gamma^2(\mathrm{d}x) = \\ \Big(\iint f(\langle x,a\rangle)\,\gamma^2(\mathrm{d}x) \Big) \Big(\iint g(\langle x,b\rangle)\,\gamma^2(\mathrm{d}x) \Big) = \\ \Big(\int f(|a|x)\gamma^1(\mathrm{d}x) \Big) \Big(\int g(|b|x)\gamma^1(\mathrm{d}x) \Big)$$

for measurable $f, g : \mathbb{R} \to \mathbb{R}$ such that the latter integrals converge. No other distribution has such properties (see [1], Sect. 1.9).

The same holds in $\mathbb{R}^n$. Namely, γ^n makes $\langle x, a\rangle \sim \mathrm{N}(0, |a|^2)$ and $\langle x, a_1\rangle, \dots, \langle x, a_m\rangle$ independent whenever $a_1, \dots, a_m$ are orthogonal.

Here are n-dimensional counterparts of Examples 1a2–1a7.

1b2 Example. Generalizing 1a2, let λ_N be the uniform distribution on the sphere $S^{N-1}(\sqrt{N})$ and μ_N the corresponding distribution of the first three coordinates. Then $\mu_N \to \gamma^3$ (Mehler, Maxwell-Boltzmann). The same holds for all n (not just 3), and is often (unjustly) called Poincaré's lemma (or Poincaré's limit).

1b3 Example. Generalizing 1a4, let λ_N denote the uniform distribution on the finite set $\{-e_1, e_1, -e_2, e_2, -e_3, e_3\}^N$, where (e_1, e_2, e_3) is the standard basis of $\mathbb{R}^3$. Denote by μ_N the corresponding distribution of $\sqrt{3/N}(x_1 + \dots + x_N)$. Then $\mu_N \to \gamma^3$. The same holds for all n (not just 3), see [2], Chap. 2, Example 9.1.

1b4 Example. In order to generalize 1a5 we need a median of a 3-dimensional sample $(x_1, \dots, x_N)$. We may define it as the minimizer of the (strictly convex) function $x \mapsto |x - x_1| + \dots + |x - x_N|$. The asymptotic normality holds for all n (not just 3).

1b5 Example. Generalizing 1a6 we replace the unfair coin with an experiment having 3 outcomes whose probabilities p_1, p_2, p_3 are parameters chosen at random uniformly on the simplex $p_1 + p_2 + p_3 = 1$, $p_1 \ge 0$, $p_2 \ge 0$, $p_3 \ge 0$. The conditional distribution of $\big(\sqrt{N}(p_1 - \frac13), \sqrt{N}(p_2 - \frac13), \sqrt{N}(p_3 - \frac13)\big)$ converges to a degenerate multinormal distribution. The same holds for all n (not just 3).

1b6 Example. Similarly to 1a7, most of the 2^{6N} triples (f_1, f_2, f_3) of trigonometric polynomials lead to distributions close to γ^3. The same holds for all n (not just 3).

Among all probability measures μ on $\mathbb{R}^n$ such that $\int x\,\mu(\mathrm{d}x) = 0$ and $\int x_k^2\,\mu(\mathrm{d}x) = 1$ for $k = 1, \dots, n$, γ^n minimizes the Poincare constant (see [1], 1.10.2 and 1.6.4) and maximizes the entropy.

1c Gaussian measures on finite-dimensional linear spaces

1c1 Lemma. Let E be an m-dimensional linear space and $V_1 : \mathbb{R}^n \to E$ a linear operator onto (that is, $V_1(\mathbb{R}^n) = E$). Then there exists an *invertible* linear operator $V_2 : \mathbb{R}^m \to E$ such that $V_1(\gamma^n) = V_2(\gamma^m)$ (written also as $\gamma^n \circ V_1^{-1} = \gamma^m \circ V_2^{-1}$), that is,

$$\gamma^n(V_1^{-1}(A)) = \gamma^m(V_2^{-1}(A)) \quad \text{for all measurable } A \subset E\,.$$

Proof (sketch). We choose an orthonormal basis $(e_1, \dots, e_n)$ of $\mathbb{R}^n$ such that $e_{m+1}, \dots, e_n$ span the kernel $\{x \in \mathbb{R}^n : V_1(x) = 0\}$, then $V_1(e_1), \dots, V_1(e_m)$ are a basis of E. By rotation invariance of γ^n we may assume that $(e_1, \dots, e_n)$ is the standard basis of $\mathbb{R}^n$. We have

$$\begin{aligned}\gamma^n(V_1^{-1}(A)) = \gamma^n\{(x_1, \dots, x_n) : V_1(x_1e_1 + \dots + x_me_m) \in A\} = \\ \gamma^m\{(x_1, \dots, x_m) : V_1(x_1e_1 + \dots + x_me_m) \in A\} = \gamma^m(V_2^{-1}(A))\end{aligned}$$

where $V_2(x_1, \dots, x_m) = x_1V_1(e_1) + \dots + x_mV_1(e_m)$. $\square$

1c2 Definition. A probability measure γ on a finite-dimensional linear space E is a *centered Gaussian measure,* if for some $n \in \{0, 1, 2, \dots\}$ there exists a one-to-one linear operator $V : \mathbb{R}^n \to E$ such that $V(\gamma^n) = \gamma$.

Usually we deal only with *centered* Gaussian measures, and omit the word 'centered'. When needed, we can say 'not just centered' or 'shifted'.

1c3 Exercise. If $V : \mathbb{R}^n \to E$ is a linear operator (not just one-to-one) then $V(\gamma^n)$ is a Gaussian measure. (Centered, of course. . .)

Prove it.

1c4 Exercise. If E_1, E_2 are finite-dimensional linear spaces, $V : E_1 \to E_2$ a linear operator and γ a Gaussian measure on E_1, then $V(\gamma)$ is a Gaussian measure on E_2.

Prove it.

1c5 Exercise. The number n in Def. 1c2 is uniquely determined by γ.
Prove it.

This number is, by definition, the dimension of γ. If $\dim\gamma = \dim E$, we say that γ is nondegenerate.

1c6 Exercise. Define the support of γ (it should be a linear subspace whose dimension is equal to the dimension of γ).

We define the *ellipsoid of concentration* of γ as the set of all $x \in E$ such that (see [1, p. 5], [4, p. 98])

$$|f(x)|^2 \le \int f^2 \, \mathrm{d}\gamma \quad \text{for all linear } f : E \to \mathbb{R}\,. \tag{1c7}$$

1c8 Exercise. The ellipsoid of concentration of γ^n is the unit ball of $\mathbb{R}^n$.
Prove it.
(See also [4], Exercise 2 to Sect. 9.)

1c9 Exercise. $\int f^2 \, \mathrm{d}\gamma = \sup f(x)$, where x runs over the ellipsoid of concentration of γ.
Prove it.

1c10 Exercise. If E_1, E_2 are finite-dimensional linear spaces, $V : E_1 \to E_2$ a linear operator and γ a Gaussian measure on E_1, then V maps the ellipsoid of concentration of γ *onto* the ellipsoid of concentration of $V(\gamma)$.
Prove it.

The ellipsoid of concentration of a nondegenerate Gaussian measure γ on E is the unit ball of a norm $|\cdot|_\gamma$ on E,

$$|x|_\gamma = \sup \Big\{ |f(x)| : \textstyle\int f^2 \, \mathrm{d}\gamma \le 1 \Big\} \,.$$

The pair $(E, |\cdot|_\gamma)$ is a Euclidean space, and γ has the density $\mathrm{const} \cdot \mathrm{e}^{-|x|_\gamma^2/2}$. For a degenerate γ the same holds on its support.

1c11 Exercise. Let $E = E_1 \oplus E_2$ (that is, $E_1, E_2 \subset E$ are linear subspaces, $E_1 \cap E_2 = \{0\}$ and $E_1 + E_2 = E$), and E_1, E_2 are orthogonal in $(E, |\cdot|_\gamma)$ (that is, $|x+y|_\gamma^2 = |x|_\gamma^2 + |y|_\gamma^2$ for $x \in E_1$, $y \in E_2$). Then there exist Gaussian measures γ_1 on E_1 and γ_2 on E_2 such that

$$\int f \, \mathrm{d}\gamma = \iint f(x+y) \, \gamma_1(\mathrm{d}x) \gamma_2(\mathrm{d}y)$$

for every bounded measurable $f : E \to \mathbb{R}$.
Prove it.
Hint: recall the proof of 1c1.

We may write $\gamma = \gamma_1 \times \gamma_2$ or $\gamma = \gamma_1 * \gamma_2$. Note that γ_1, γ_2 are uniquely determined by γ (just take $f(x+y) = g(x)$ or $h(y)$). These γ_1, γ_2 are projections (marginals) of γ. Ellipsoids of concentration of γ_1, γ_2 are both sections and projections of the ellipsoid of concentration of γ.

Let γ be a nondegenerate Gaussian measure on E and $V : E \to E_1$ a linear operator onto. Then $E = \tilde{E}_1 \oplus E_2$ where $E_2 = \{x : V(x) = 0\}$ is the kernel and $\tilde{E}_1 = E \ominus E_2$ its orthogonal (w.r.t. $|\cdot|_\gamma$) complement. The restriction $V|_{\tilde{E}_1}$ is an isometry $\tilde{E}_1 \to E_1$, provided that E_1 is equipped with $|\cdot|_{V(\gamma)}$. Denoting the inverse isometry by $\tilde{V} : E_1 \to \tilde{E}_1$ we have

$$\int f \, \mathrm{d}\gamma = \iint_{\tilde{E}_1 \times E_2} f(x+y) \, \gamma_1(\mathrm{d}x)\gamma_2(\mathrm{d}y) = \iint_{E_1 \times E_2} f(\tilde{V}(x)+y) \, V(\gamma)(\mathrm{d}x)\gamma_2(\mathrm{d}y) = \int_{E_1} \biggl(\int_E f(\tilde{V}(x)+y) \, \gamma_2(\mathrm{d}y) \biggr) V(\gamma)(\mathrm{d}x) \, ,$$

which means that the conditional distribution γ_x of $z \in E$ given $V(z) = x \in E_1$ is γ_2 shifted by $\tilde{V}(x)$. We see that all conditional measures are shifts of a single Gaussian measure, and the shift vector depends linearly on the condition. This is known as the *normal correlation theorem;* see also [3], Sect. 9.3 and [1], 1.2.8 and 3.10.

1c12 Exercise. Consider a random trigonometric polynomial

$$X(t) = \zeta_1 \cos t + \eta_1 \sin t + \frac{1}{2}\zeta_2 \cos 2t + \frac{1}{2}\eta_2 \sin 2t \, ,$$

where $\zeta_1, \eta_1, \zeta_2, \eta_2$ are independent $\mathrm{N}(0,1)$ random variables. Describe the conditional distribution of X given $X(0)$.

References

[1] V.I. Bogachev, *Gaussian measures,* AMS 1998.

[2] R. Durrett, *Probability: theory and examples* (second edition), 1996.

[3] S. Janson, *Gaussian Hilbert spaces,* Cambridge 1997.

[4] M.A. Lifshits, *Gaussian random functions,* Kluwer 1995.

Index