Princeton University

Spring 2025 MAT425: Measure Theory

HW10 Sample Solutions

Apr 28 2025

Serban Cicortas (Q6, Q7, Q9, Q10) and Chayim Lowen (Q1-Q5)

May 5, 2025

Problem 1.

Find probability measures μ, ν on $\mathbb{R}^2$ such that

$$\mu(B_1 \times B_2) = \nu(B_2 \times B_1) \qquad (B_1, B_2 \in \mathfrak{B}(\mathbb{R}))$$

fails.

Solution. Let $\mu = \nu$ be the measure defined by $d\mu = d\nu = \chi_{(0,1)\times(1,2)}\, d\lambda$ where λ is the Lebesgue measure. That is,

$$\mu(S) = \nu(S) = \lambda(S \cap ((0,1) \times (1,2)))$$

for all $S \in \mathfrak{B}(\mathbb{R}^2)$. Let $B_1 := (0,1)$ and $B_2 := (1,2)$. Then

$$\mu(B_1 \times B_2) = 1 \neq 0 = \nu(B_2 \times B_1) \qquad \square$$

Problem 2.

Find probability measures μ on $\mathbb{R}^2$ and ν on $\mathbb{R}$ such that

$$\mu(B \times \mathbb{R}) = \nu(B) \quad (B \in \mathfrak{B}(\mathbb{R}))$$

fails.

Solution. Define μ by $d\mu = \chi_{(0,1)^2}\, d\lambda$, and ν by $d\nu = \chi_{(1,2)}\, d\lambda$. That is,

$$\mu(S) := \lambda(S \cap (0,1)^2) \quad \text{and} \quad \nu(T) := \lambda((1,2) \cap T)$$

Let $B := [0,1]$. Then

$$\mu(B \times \mathbb{R}) = 1 \neq 0 = \nu(B)$$

□

Problem 3.

Let $(S_N)_{N\in\mathbb{N}}$ be a simple random walk on $\mathbb{R}$ defined by IID increments $(X_n)_{n\in\mathbb{N}}$ with a priori probability measure $\mu_0 : \mathfrak{B}(\mathbb{R}) \to [0,1]$. Show that these give rise to a system of probability measures satisfying the Kolmogorov consistency conditions.

Solution. We show that the "joint probability distributions" of the increments satisfy the consistency conditions. To see this, let $B_1, \dots, B_m$ be Borel subsets of $\mathbb{R}$ and let $j_1, \dots, j_m$ be pairwise distinct positive integers. Then

$$\mathbb{P}_{X_{j_1},\dots,X_{j_m}}(B_1 \times \cdots \times B_m) \equiv \left(\prod_{k=1}^{m} \mu_0\right)(B_1 \times \cdots \times B_m) \equiv \mu_0(B_1)\mu_0(B_2)\cdots\mu_0(B_n)$$

Note that the expression on the left is independent of the choice of $j_1, \dots, j_m$ and of the order in which the B_i's appear (since multiplication of real numbers is commutative and associative). In particular, this means it remains invariant when we permute the X_{j_i}'s and the B_i. The second consistency condition follows from this.

For the third consistency condition, let $j_1, \dots, j_m, i_1, \dots, i_l$ be pairwise distinct positive integers and let $B_1, \dots, B_m$ be Borel subsets of $\mathbb{R}$. Then

$$\begin{aligned}
\mathbb{P}_{X_{j_1},\dots,X_{j_m},X_{i_1},\dots,X_{i_l}}(B_1 \times \cdots \times B_m \times \mathbb{R}^l) &\equiv \left(\prod_{k=1}^{m} \mu_0\right)(B_1 \times \cdots \times B_m \times \mathbb{R}^l) \\
&\equiv \mu_0(B_1)\mu_0(B_2)\cdots\mu_0(B_n)\mu_0(\mathbb{R})^l \\
&= \mu_0(B_1)\mu_0(B_2)\cdots\mu_0(B_n) \\
&\equiv \mathbb{P}_{X_{j_1},\dots,X_{j_m}}(B_1 \times \cdots \times B_n)
\end{aligned}$$

where we have used the fact that μ_0 is a probability measure to conclude that $\mu_0(\mathbb{R})^l = 1$. □

Problem 4.

Let $m : [0,T] \to \mathbb{R}$ be a given function and C a self-adjoint positive linear operator on $L^2([0,T] \to \mathbb{R})$ induced by a kernel $(t,s) \mapsto C(t,s)$. We associate to m and C the *Gaussian stochastic process* $(B_t)_{t\in[0,T]}$ specified by its finite marginals for $0 \le t_1 < \cdots < t_n \le T$ via

$$\frac{d\mathbb{P}_{B_{t_1},\ldots,B_{t_n}}}{d\lambda}(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det K}} \exp\left(-\frac{1}{2}\left\langle (x-\mu), K^{-1}(x-\mu)\right\rangle\right) \quad (x \in \mathbb{R}^n)$$

where K is the $n \times n$ matrix given by $K_{ij} := C(t_i, t_j)$ and μ is the vector given by $\mu_i := m(t_i)$.

(a) Show that this system satisfies the Kolmogorov consistency conditions.

Solution. We have only defined the joint marginals for strictly increasing sequences of t_j's. The defnition in the other cases is determined in the *unique* way that makes the second Kolmogorov condition hold true.

The first Kolmogorov condition holds because it follows from our calculation on HW9Q10(a) that

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\left\langle (x-\mu), K^{-1}(x-\mu)\right\rangle\right) d\lambda(x) = \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\left\langle x, K^{-1}x\right\rangle\right) d\lambda(x) = \frac{(2\pi)^{n/2}}{(\det K^{-1})^{1/2}} = (2\pi)^{n/2}\sqrt{\det K}$$

and thus the functions above integrate to 1.

To see that the third Kolmogorov condition holds, let $t_1, \ldots, t_n, s_1, \ldots, s_m$ be pairwise distinct real numbers in $[0, T]$ (not necessarily in increasing order).

The distribution $\mathbb{P}_{B_{t_1},\ldots,B_{t_n},B_{s_1},\ldots,B_{s_m}}$ is a multivariate Gaussian distribution. It is possible but messy to show directly that the marginal of this distribution with respect to the first n variables gives $\mathbb{P}_{B_{t_1},\ldots,B_{t_n}}$. Instead we appeal to properties of Gaussian distributions.

We appeal to the results in Chapter 1.2 of Bogachev's *Gaussian Measures*. There it is shown that a probability measure has the multivariate Gaussian form that our distributions do if and only if its 1-dimensional marginals are ordinary Gaussians. This implies that marginals of Gaussians are Gaussian. Two multivariate Gaussians with the same means and covariances are equal. (This also follows from Bogachev's results.)

More precisely, Bogachev shows[1] that (in our notation) the vector μ gives the means of the B-variables and the matrix K gives all their covariances. More precisely, $\mu(j) = m(t_j)$ is the mean of B_{t_j} and $C(t_j, t_k)$ is the covariance of B_{t_j} and B_{t_k}.[2]

Since covariances are preserved upon taking marginals, we see that the means and covariances of $\mathbb{P}_{B_{t_1},\ldots,B_{t_n}}$ agree with those of the marginal of $\mathbb{P}_{B_{t_1},\ldots,B_{t_n},B_{s_1},\ldots,B_{s_m}}$ with respect to the first n variables. It follows that $\mathbb{P}_{B_{t_1},\ldots,B_{t_n}}$ is the appropriate marginal of $\mathbb{P}_{B_{t_1},\ldots,B_{t_n},B_{s_1},\ldots,B_{s_m}}$. □

(b) Show that a Gaussian process is characterized by its first two moments by relating them to m and C.

[1] This also follows from our work in HW9Q10.

[2] One cannot yet talk of *the* random variable B_{t_j} since we do not yet know that a global probability space exists. However, this concept does make sense with respect to any of our given distributions.

Solution. The function

$$\frac{d\,\mathbb{P}_{B_t}}{d\lambda}(x) = \frac{1}{\sqrt{2\pi C(t)}} \exp\left(-\frac{1}{2}C(t,t)^{-1}(x-m(t))^2\right)$$

is visibly symmetric around the point $m(t)$. It follows that $\mathbb{E}[B_t] = m(t)$.

Our calculation in HW9Q10(c) (in the special case when v_1 and v_2 are standard basis vectors) gives

$$\mathbb{E}[\overline{B_t}B_u] = C(t,u)$$

Together, these facts show that m and C can be recovered from the first two moments of the Gaussian process. □

Problem 5.

5. What is the operator

$$C : L^2([0,1] \to \mathbb{R}) \to L^2([0,1] \to \mathbb{R})$$

so that

$$C(t,s) = \min(\{\,t,s\,\})\,?$$

Conclude a more "appealing" expression for

$$\frac{\mathrm{d}\mathbb{P}_{\left(B_{t_1},\cdots,B_{t_n}\right)}}{\mathrm{d}\lambda}(x)$$

when $(B_t)_t$ is Brownian motion then.

Solution. We claim that that operator is the inverse of the 1D Laplacian on $L^2([0,1] \to \mathbb{R})$ with certain boundary conditions. Let us verify this. Consider then the operator K so that

$$(Kf)(s) := \int_{t=0}^1 C(s,t)f(t)\mathrm{d}t = \int_{t=0}^1 \min(\{\,s,t\,\})f(t)\mathrm{d}t = \int_{t=0}^s tf(t)\mathrm{d}t + s\int_{t=s}^1 f(t)\mathrm{d}t\,.$$

We calculate ∂_s^2 of this to obtain

$$(Kf)''(s) = \partial_s\left(sf(s) + \int_{t=s}^1 f(t)\mathrm{d}t - sf(s)\right) = -f(s)\,.$$

Moreover, $(Kf)(0) = 0$ whereas $(Kf)'(1) = 0$ too. Hence we identify

$$K = \left(-\partial^2\right)^{-1}$$

on $L^2([0,1])$. Note that while the first boundary condition $(Kf)(0) = 0$ naturally corresponds to the conditioning $B_0 = 0$, $(Kf)'(1) = 0$ makes the operator K self-adjoint and corresponds to the fact that B_1

is unconstrained (this is vague handwaving but that part of the argument is not necessary to answer the question).

In light of Q4, for any $0 \leq t_1 < \cdots < t_n \leq 1$, let κ be the $n \times n$ matrix comprised via $\kappa_{ij} := C(t_i, t_j)$. Then κ^{-1} is clearly a finite mesh approximation (with mesh size n) of the Laplacian, i.e., $\langle x, \kappa^{-1} x\rangle = \sum_{i=1}^{n} \frac{1}{t_{i+1}-t_i}(x_{i+1}-x_i)^2$. Hence

$$\frac{\mathrm{d}\mathbb{P}_{(B_{t_1},\cdots,B_{t_n})}}{\mathrm{d}\lambda}(x) \approx \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det\left(\left(-\partial^{-2}\right)_n\right)}} \exp\left(-\frac{1}{2}\left\langle x, \left(-\partial^2\right)_n x\right\rangle_{\mathbb{R}^n}\right) \qquad (x \in \mathbb{R}^n)\ .$$

□

Problem 6.

Solution. a). The rate function is Cramer's function, given as the Legendre transform of the cumulant generating function. For an introduction to the problem and sketch of the proof of the first condition for LDP see Example 7.47 in the lecture notes. For the full proof see Klenke - Probability Theory attached below. One should first read the proof of Theorem 23.3 (pages 508-509), then Example 23.10 (pages 512-513).

b). Consider the case when each X_n is a symmetric coin flip for ± 1. The cumulant generating function is $\kappa(\theta) = \log \mathbb{E}[e^{\theta X}] = \log\left(\frac{1}{2}(e^{\theta} + e^{-\theta})\right) = \log\cosh\theta$. According to part a), the rate function is the Legendre transform of $\log\cosh\theta$. To compute this, we denote by $\theta^*(x)$ the value that realizes the supremum in the definition of the Legendre transform, $I(x) = x\theta^*(x) - \log\cosh\theta^*(x)$. As a result, we get that:

$$x = \big(\log\cosh\big)'\theta^*(x) = \tanh\theta^*(x)$$

We can then compute the rate function:

$$I(x) = x\tanh^{-1}(x) - \log\cosh\tanh^{-1}(x) = \frac{(1+x)\log(1+x) + (1-x)\log(1-x)}{2}$$

c). By Cramer's theorem proved above we have that:

$$\lim_{N\to\infty} \frac{1}{N}\log\mathbb{P}[A_N \geq a] = -I(a)$$

As a result, we obtain that $\mathbb{P}[A_N \geq a]$ is asymptotically $e^{-NI(a)}$.

d). By Varadhan's lemma, we know that:

$$\lim_{N\to\infty} \frac{1}{N}\log\mathbb{E}_{A_N}[\exp Nf(A_N)] = \sup_{\omega\in\mathbb{R}}\big(f(\omega) - I(\omega)\big)$$

As a result, we obtain that $\mathbb{E}_{A_N}[\exp Nf(A_N)]$ is asymptotically $\exp\Big(N\sup_{\omega\in\mathbb{R}}\big(f(\omega) - I(\omega)\big)\Big)$. □

Problem 7.

Solution. See Theorem 7.54 in the lecture notes for the proof of the scaling law. □

Problem 8.

Solution. Recall the Feynman-Kac formula: if $V : \mathbb{R} \to \mathbb{R}$ is a potential then we can rewrite a solution of the Schroedinger-heat equation

$$\partial_t \psi = -(-\Delta + V(X))\psi$$

for the unknown $\psi : [0,\infty) \to L^2(\mathbb{R} \to \mathbb{C})$ with initial condition $\psi(0) = \psi_0 \in L^2(\mathbb{R} \to \mathbb{C})$ using Brownian motion

$$(\exp(-t(-\Delta + V(X)))\psi_0)(x) = \mathbb{E}\left[\psi_0(x + B_t)\exp(-\int_{s=0}^{t} V(x + B_s)\mathrm{d}s)\right]$$

where $(B_t)_t$ is standard Brownian motion which is conditioned so that $B_0 = 0$ almost-surely. Thus the question is asking us to evaluate the asymptotics of the heat-kernel

$$I_\lambda(x,y) := (\exp(-\frac{T}{\lambda}(-\Delta + \lambda^2 V(X))))(x,y)$$

for large λ. Let us define pinned Brownian motion

$$B_t^{\text{pinned},(x@0,y@T)} := x + (y-x)\frac{t}{T} + B_t - \frac{t}{T}B_T$$

where $(B_t)_t$ is standard Brownian motion. Then clearly we have $B_0^{\text{pinned},(x@0,y@T)} = x$, $B_T^{\text{pinned},(x@0,y@T)} = y$, hence the notation. W.r.t. that process, the heat kernel becomes, using the Feynman-Kac formula and conditional expectation

$$\begin{aligned}
\exp(-\frac{T}{\lambda}(-\Delta + \lambda^2 V(X))(x,y) &= \mathbb{E}\left[\delta(x + B_{T/\lambda} - y)\exp(-\lambda^2\int_{t=0}^{T/\lambda} V(x+B_t)\mathrm{d}t)\right] \\
&= \mathbb{E}\left[\exp(-\lambda^2\int_{t=0}^{T/\lambda} V(x+B_t)\mathrm{d}t) \mid B_{T/\lambda} = y - x\right]\mathbb{P}\left[B_{T/\lambda} = y-x\right] \\
&= \mathbb{E}\left[\exp(-\lambda^2\int_{t=0}^{T/\lambda} V(x+B_t^{\text{pinned},(x@0,y@T/\lambda)})\mathrm{d}t)\right]\mathbb{P}\left[B_{T/\lambda} = y-x\right]
\end{aligned}$$

which is basically what originally appeared in the question (with the pinned process fully written out).

Let us make a change of variables in the integral within the exponent. We find it more convent to actually use the first line of the above, so we have

$$\begin{aligned}
\lambda^2\int_{t=0}^{T/\lambda} V(x+B_t)\mathrm{d}t &= \lambda\int_{s=0}^{T} V(x + B_{s/\lambda})\mathrm{d}s \\
&\overset{\mathrm{d}}{=} \lambda\int_{s=0}^{T} V(x + \frac{1}{\sqrt{\lambda}}B_s)\mathrm{d}s
\end{aligned}$$

where we have used the scaling property of Brownian motion. We are being told in the question that the family of measures $\left\{\mathbb{P}_{B_{\sqrt{\varepsilon}t}}\right\}_{\varepsilon>0}$ obeys an LDP, so following Varadhan's lemma we find

$$\lim_{\lambda\to\infty} -\frac{1}{\lambda}\log\left(\exp(-\frac{T}{\lambda}(-\Delta + \lambda^2 V(X))(x,y)\right) = -\inf_\gamma\left(I(\gamma) + \int_0^T V(\gamma(s))\mathrm{d}s\right)$$

where I is the rate function given in the question (the kinetic energy) and in principle the infimum is over all continuous functions $[0,T] \to \mathbb{R}$ which obey $\gamma(0) = x, \gamma(T) = y$, but actually by the definition of I we can take the infimum over the smaller subsets of those functions which are absolutely continuous and whose first derivative is in L^2. □

Problem 9.

Solution. a). Using the reflection principle we have:

$$\mathbb{P}[\tau_a \leq t] = \mathbb{P}\left[\sup_{s\in[0,t]} B_s \geq a\right] = 2\mathbb{P}[B_t \geq a] = 2\int_{a/\sqrt{t}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

and we also obtain that $\mathbb{P}[\tau_a < \infty] = 1$.

b). By the strong Markov property, we have that both processes $\{B(t+\tau_a) - B(\tau_a) : \ t \geq 0\}$ and $\{-B(t+\tau_a) + B(\tau_a) : \ t \geq 0\}$ are Brownian motions and agree in distribution. As a result, we also get that $B_t \overset{d}{=} \tilde{B}_t$.

c). This was already computed in part a). □

Problem 10.

Solution. We compute that:

$$u(t,x) = \mathbb{E}[f(x+B_t)] = \int_{\mathbb{R}} f(x+y) \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} dy = \int_{\mathbb{R}} f(y) \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} dy$$

This is simply the convolution of f with the heat kernel, so we get that u solves the heat equation with initial data given by f. See for example Section 3.2 in Chapter 5 of Stein, Shakarchi - Fourier Analysis. □

http://dx.doi.org/10.1090/surv/062/01

CHAPTER 1

Finite Dimensional Gaussian Distributions

The connection of literature purposes with the purely scientific ones, the desire to occupy imagination and at the same time to enrich life with ideas and knowledge create considerable difficulties in composing the different parts of the book and hamper the unity of exposition.

Alexander von Humboldt. Views of Nature

1.1. Gaussian measures on the real line

We start our discussion of Gaussian measures by recalling the identity

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\Bigl(-\frac{(t-a)^2}{2\sigma^2}\Bigr)\,dt=1, \tag{1.1.1}$$

known for all real a and $\sigma > 0$. A standard way of verifying this equality is to evaluate the double integral $\int\int \exp(-x^2-y^2)\,dx\,dy$ in polar coordinates and apply Fubini's theorem.

1.1.1. Definition. *A Borel probability measure γ on $\mathbb{R}^1$ is called Gaussian if it is either the Dirac measure δ_a at a point a or has density*

$$p(\,\cdot\,,a,\sigma^2)\colon\ t\mapsto\frac{1}{\sigma\sqrt{2\pi}}\exp\Bigl(-\frac{(t-a)^2}{2\sigma^2}\Bigr)$$

with respect to Lebesgue measure. In the latter case the measure γ is called nondegenerate.

The parameters a and σ^2 are called the *mean* and the *variance* of γ, respectively. The quantity σ is called the mean-square deviation. For any Dirac measure (i.e., a probability measure concentrated at a point) we put $\sigma=0$. The mean and the variance of a Gaussian measure γ admit the following representation:

$$a=\int_{-\infty}^{+\infty}t\,\gamma(dt),\quad \sigma^2=\int_{-\infty}^{+\infty}(t-a)^2\,\gamma(dt).$$

The measure with density $p(\,\cdot\,,0,1)$ is called *standard.* A mean zero Gaussian measure is called *centered* or *symmetric.*

1.1.2. Definition. *A Gaussian random variable is a random variable with Gaussian distribution.*

A Gaussian random variable with a centered distribution is called centered or symmetric. Clearly, an arbitrary Gaussian random variable with mean a and

Licensed to Princeton Univ. Prepared on Sun May 4 18:22:35 EDT 2025for download from IP 128.112.200.107.

variance σ^2 can be represented as $\sigma\xi + a$, where ξ is a random variable with the standard Gaussian distribution. Gaussian distributions are often called *normal.*

Using equality (1.1.1) it is easy to find the Fourier transform (the characteristic functional) of the Gaussian measure γ with parameters (a, σ^2). We have

$$\widetilde{\gamma}(y) := \int_{\mathbb{R}^1} \exp(iyx)\,\gamma(dx) = \exp\Big(iay - \frac{1}{2}\sigma^2 y^2\Big).$$

The normal (standard Gaussian) distribution function Φ is defined by the relation

$$\Phi(t) = \int_{-\infty}^{t} p(s, 0, 1)\,ds.$$

The inverse function Φ^{-1} is defined on $(0,1)$. It is convenient to employ the following convention: $\Phi^{-1}(0) = -\infty$ and $\Phi^{-1}(1) = +\infty$.

The rate of decreasing of $1 - \Phi$ at infinity is estimated as follows.

1.1.3. Lemma. *For any $t > 0$, one has*

$$\frac{1}{\sqrt{2\pi}}\Big(\frac{1}{t} - \frac{1}{t^3}\Big)e^{-t^2/2} \le 1 - \Phi(t) \le \frac{1}{\sqrt{2\pi}}\frac{1}{t}e^{-t^2/2}. \qquad (1.1.2)$$

PROOF. By virtue of the integration by parts formula, we get

$$\int_t^\infty s e^{-s^2/2}\frac{1}{s}\,ds = \frac{e^{-t^2/2}}{t} - \int_t^\infty \frac{1}{s^2}e^{-s^2/2}\,ds \le \frac{e^{-t^2/2}}{t}.$$

The lower bound is proved in a similar manner. □

The next classical result has a lot of applications in the theory of Gaussian measures.

1.1.4. Theorem. *Let ξ_n be a sequence of independent centered Gaussian random variables with variances σ_n^2. Then the following conditions are equivalent:*

(i) *The series* $\sum_{n=1}^{\infty} \xi_n$ *converges almost everywhere;*

(ii) *The series in* (i) *converges in probability;*

(iii) *The series in* (i) *converges in L^2;*

(iv) $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$.

PROOF. According to the dominated convergence theorem and the independence condition, either of conditions (i) — (iii) implies convergence of the product

$$\prod_{n=1}^{\infty}\int \exp(i\xi_n)\,dP = \prod_{n=1}^{\infty}\exp(-\sigma_n^2/2),$$

which yields condition (iv). Conversely, condition (iv) implies condition (iii), since by virtue of independence and symmetry of the random variables in question, we have

$$\int (\xi_k + \ldots + \xi_{k+m})^2\,dP = \int \sum_{j=0}^{m} \xi_{k+j}^2\,dP.$$

Hence (ii) is satisfied as well. Thus, the only non trivial implication in our claim is (iv)⇒(i). In order to prove it, note that the conditional expectation of the

Licensed to Princeton Univ. Prepared on Sun May 4 18:22:35 EDT 2025for download from IP 128.112.200.107.

square-integrable random variable $\xi = \sum_{n=1}^{\infty} \xi_n$ with respect to the σ-field $\mathcal{A}_m$, generated by $\xi_1, \ldots, \xi_m$, coincides with $\sum_{n=1}^{m} \xi_n$, since the random variables ξ_n are independent and have zero means. Therefore, the sequence of the partial sums of the series defining ξ is a martingale with respect to $\{\mathcal{A}_n\}$. Hence Theorem A.3.5 in Appendix applies and yields the almost sure convergence. □

One more classical result related to one dimensional Gaussian distributions is the central limit theorem. We only give its special case that will be used below.

1.1.5. Theorem. *Let $\{\xi_n\}$ be a sequence of independent random variables with one and the same distribution such that $\mathbb{E}\xi_1 = 0$ and $\sigma^2 = \mathbb{E}\xi_1^2 < \infty$. Put $S_n = \xi_1 + \ldots + \xi_n$. Then, for any x, we get*

$$P\left\{\frac{S_n}{\sigma\sqrt{n}} \leq x\right\} \to \Phi(x) \qquad \text{as } n \to \infty.$$

In addition, the distributions of $\dfrac{S_n}{\sigma\sqrt{n}}$ converge weakly to the standard Gaussian measure.

PROOF. Since $\mathbb{E}\xi_1^2 < \infty$, the function $\varphi(t) = \mathbb{E}e^{it\xi_1}$ is twice differentiable. We have

$$\varphi(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2), \quad t \to 0.$$

The characteristic functionals φ_n of $\dfrac{S_n}{\sigma\sqrt{n}}$ are given by $\varphi\Big(\dfrac{t}{\sigma\sqrt{n}}\Big)^n$. Hence, for any fixed t, we get

$$\lim_{n\to\infty} \varphi_n(t) = \lim_{n\to\infty}\left[1 - \frac{\sigma^2 t^2}{2\sigma^2 n} + o\Big(\frac{1}{n}\Big)\right]^n = e^{-t^2/2}.$$

Both claims follow from this equality (see [**697**, Chapter III]). □

1.2. Multivariate Gaussian distributions

1.2.1. Definition. *A Borel probability measure γ on $\mathbb{R}^n$ is called Gaussian if for every linear functional f on $\mathbb{R}^n$, the induced measure $\gamma \circ f^{-1}$ is Gaussian.*

We shall use the standard identification of the space of linear functions on $\mathbb{R}^n$ with $\mathbb{R}^n$. The inner product in $\mathbb{R}^n$ is denoted by $(\,\cdot\,,\,\cdot\,)$ or by $\langle\,\cdot\,,\,\cdot\,\rangle$.

Recall that the Fourier transform $\widetilde{\mu}$ of a finite Borel measure μ on $\mathbb{R}^n$ is defined by the formula

$$\widetilde{\mu}\colon\ \mathbb{R}^n \to \mathbb{C}^1, \quad \widetilde{\mu}(y) = \int_{\mathbb{R}^n} \exp\big[i(y,x)\big]\,\mu(dx).$$

Recall that measures on $\mathbb{R}^n$ are uniquely determined by their Fourier transforms.

1.2.2. Proposition. *A measure γ on $\mathbb{R}^n$ is Gaussian if and only if its Fourier transform has the form*

$$\widetilde{\gamma}(y) = \exp\Big(i(y,a) - \frac{1}{2}(Ky,y)\Big), \tag{1.2.1}$$

Licensed to Princeton Univ. Prepared on Sun May 4 18:22:35 EDT 2025for download from IP 128.112.200.107.

where a is a vector in $\mathbb{R}^n$ and K is a nonnegative matrix. The measure γ has a density if and only if the matrix K is nondegenerate. In this case, the density of the measure γ is given by

$$x \mapsto \frac{1}{\sqrt{(2\pi)^n \det K}} \exp\left\{-\frac{1}{2}\left(K^{-1}(x-a), x-a\right)\right\}.$$

PROOF. Let f be a linear function on $\mathbb{R}^n$. Using the change of variables formula (see formula (A.3.1) in Appendix), one evaluates the Fourier transform of the measure $\nu = \gamma \circ f^{-1}$ as follows:

$$\widetilde{\nu}(t) = \int\limits_{\mathbb{R}^1} \exp(its)\,\nu(ds) = \int\limits_{\mathbb{R}^n} \exp\bigl(itf(x)\bigr)\,\gamma(dx).$$

Let us denote the vector representing the functional f by the same symbol. From (1.2.1) we get

$$\widetilde{\nu}(t) = \exp\left(it(a,f) - \frac{1}{2}t^2(Kf,f)\right),$$

which means that the measure ν is Gaussian. Conversely, suppose that all such measures are Gaussian. Denote their means and variances by $a(f)$ and $\sigma(f)^2$, respectively. Then the following equalities hold true:

$$a(f) = \int\limits_{\mathbb{R}^1} t\,\gamma\circ f^{-1}(dt) = \int\limits_{\mathbb{R}^n} f(x)\,\gamma(dx),$$

$$\sigma(f)^2 = \int\limits_{\mathbb{R}^1} \Bigl(t - a(f)\Bigr)^2 \gamma\circ f^{-1}(dt) = \int\limits_{\mathbb{R}^n} \Bigl((f,x) - a(f)\Bigr)^2 \gamma(dx).$$

Hence the function $f \mapsto a(f)$ is linear, and the function $f \mapsto \sigma(f)^2$ is a nonnegative quadratic form. Therefore, there exist a vector a and a nonnegative symmetric operator K such that $a(f) = (f,a)$ and $\sigma(f)^2 = (Kf,f)$. This yields (1.2.1). The assertion about densities reduces to the one dimensional case, since we can use the coordinates corresponding to the eigenvectors of the matrix K. □

1.2.3. Corollary. *Let γ be a Gaussian measure on $\mathbb{R}^n$ with the Fourier transform* (1.2.1). *Then*

$$a = \int\limits_{\mathbb{R}^n} x\,\gamma(dx), \tag{1.2.2}$$

$$(Ku,v) = \int\limits_{\mathbb{R}^n} (u, x-a)(v, x-a)\,\gamma(dx), \quad \forall\, u,\, v \in \mathbb{R}^n. \tag{1.2.3}$$

The vector a given by equality (1.2.2) is called the *mean* of the Gaussian measure γ, and the operator K defined by means of (1.2.3) is called its *covariance operator*.

Clearly, Gaussian measures on $\mathbb{R}^n$ can be described as the images of the *standard Gaussian measure* on $\mathbb{R}^n$ (i.e., the product of n copies of the standard Gaussian measure on $\mathbb{R}^1$) under affine mappings $x \mapsto \sqrt{K}x + a$.

On the linear subspace $\sqrt{K}(\mathbb{R}^n)$ we define the inner product

$$(u,v)_\gamma := \left(\sqrt{K}^{-1}u, \sqrt{K}^{-1}v\right).$$

Licensed to Princeton Univ. Prepared on Sun May 4 18:22:35 EDT 2025for download from IP 128.112.200.107.

Theorem 23.3 (Cramér (1938)). *Let $X_1, X_2, \ldots$ be i.i.d. real random variables with finite* ***logarithmic moment generating function***

$$\Lambda(t) := \log \mathbf{E}\big[e^{tX_1}\big] < \infty \quad \text{for all } t \in \mathbb{R}. \tag{23.9}$$

Let

$$\Lambda^*(x) := \sup_{t\in\mathbb{R}} \big(tx - \Lambda(t)\big) \quad \text{for } x \in \mathbb{R},$$

the Legendre transform of Λ. Then, for every $x > \mathbf{E}[X_1]$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}\big[S_n \geq xn\big] = -I(x) := -\Lambda^*(x). \tag{23.10}$$

Proof. By passing to $X_i - x$ if necessary, we may assume $\mathbf{E}[X_i] < 0$ and $x = 0$. (In fact, if $\tilde{X}_i := X_i - x$, and $\tilde{\Lambda}$ and $\tilde{\Lambda}^*$ are defined as Λ and Λ^* above but for $\tilde{X}_i$ instead of X_i, then $\tilde{\Lambda}(t) = \Lambda(t) - t \cdot x$ and thus $\tilde{\Lambda}^*(0) = \sup_{t\in\mathbb{R}}(-\tilde{\Lambda}(t)) = \Lambda^*(x)$.) Define $\varphi(t) := e^{\Lambda(t)}$ and

$$\varrho \;:=\; e^{-\Lambda^*(0)} \;=\; \inf_{t\in\mathbb{R}} \varphi(t).$$

By (23.9) and the differentiation lemma (Theorem 6.28), φ is differentiable infinitely often and the first two derivatives are

$$\varphi'(t) = \mathbf{E}\big[X_1\, e^{tX_1}\big] \quad \text{and} \quad \varphi''(t) = \mathbf{E}\big[X_1^2\, e^{tX_1}\big].$$

Hence φ is strictly convex and $\varphi'(0) = \mathbf{E}[X_1] < 0$.

First consider the case $\mathbf{P}[X_1 \leq 0] = 1$. Then $\varphi'(t) < 0$ for every $t \in \mathbb{R}$ and $\varrho = \lim_{t\to\infty} \varphi(t) = \mathbf{P}[X_1 = 0]$. Therefore,

$$\mathbf{P}[S_n \geq 0] \;=\; \mathbf{P}[X_1 = \ldots = X_n = 0] \;=\; \varrho^n$$

and thus the claim follows.

Now let $\mathbf{P}[X_1 < 0] > 0$ and $\mathbf{P}[X_1 > 0] > 0$. Then $\lim_{t\to\infty} \varphi(t) = \infty = \lim_{t\to-\infty} \varphi(t)$. As φ is strictly convex, there is a unique $\tau \in \mathbb{R}$ at which φ assumes its minimum; hence

$$\varphi(\tau) = \varrho \quad \text{and} \quad \varphi'(\tau) = 0.$$

Since $\varphi'(0) < 0$, we have $\tau > 0$. Using Markov's inequality (Theorem 5.11), we estimate

$$\mathbf{P}[S_n \geq 0] \;=\; \mathbf{P}\big[e^{\tau S_n} \geq 1\big] \;\leq\; \mathbf{E}\big[e^{\tau S_n}\big] \;=\; \varphi(\tau)^n \;=\; \varrho^n.$$

Thus we get the upper bound

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}[S_n \geq 0] \;\leq\; \log \varrho \;=\; -\Lambda^*(0).$$

The remaining part of the proof is dedicated to verifying the reverse inequality:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}[S_n \geq 0] \;\geq\; \log \varrho. \tag{23.11}$$

We use the method of an exponential size-biasing of the distribution $\mu := \mathbf{P}_{X_1}$ of X_1, which turns the atypical values that are of interest here into typical values. That is, we define the **Cramér transform** $\hat{\mu} \in \mathcal{M}_1(\mathbb{R})$ of μ by

$$\hat{\mu}(dx) = \varrho^{-1} e^{\tau x} \mu(dx) \quad \text{ for } x \in \mathbb{R}.$$

Let $\hat{X}_1, \hat{X}_2, \ldots$ be independent and identically distributed with $\mathbf{P}_{\hat{X}_i} = \hat{\mu}$. Then

$$\hat{\varphi}(t) := \mathbf{E}\big[e^{t\hat{X}_1}\big] = \frac{1}{\varrho} \int_{\mathbb{R}} e^{tx} e^{\tau x} \, \mu(dx) = \frac{1}{\varrho}\, \varphi(t+\tau).$$

Hence

$$\mathbf{E}\big[\hat{X}_1\big] = \hat{\varphi}'(0) = \frac{1}{\varrho}\, \varphi'(\tau) = 0,$$
$$\mathbf{Var}\big[\hat{X}_1\big] = \hat{\varphi}''(0) = \frac{1}{\varrho}\, \varphi''(\tau) \in (0,\infty).$$

Defining $\hat{S}_n = \hat{X}_1 + \ldots + \hat{X}_n$, we get

$$\begin{aligned}
\mathbf{P}[S_n \geq 0] \;&=\; \int_{\{x_1+\ldots+x_n \geq 0\}} \mu(dx_1)\cdots\mu(dx_n) \\
&=\; \int_{\{x_1+\ldots+x_n \geq 0\}} \big(\varrho\, e^{-\tau x_1}\big)\hat{\mu}(dx_1)\cdots\big(\varrho\, e^{-\tau x_n}\big)\hat{\mu}(dx_n) \\
&=\; \varrho^n\, \mathbf{E}\left[e^{-\tau \hat{S}_n}\, \mathbb{1}_{\{\hat{S}_n \geq 0\}}\right].
\end{aligned}$$

Thus, in order to show (23.11), it is enough to show

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{E}\left[e^{-\tau \hat{S}_n}\, \mathbb{1}_{\{\hat{S}_n \geq 0\}}\right] \;\geq\; 0. \tag{23.12}$$

However, by the central limit theorem (Theorem 15.37), for every $c > 0$,

$$\begin{aligned}
\frac{1}{n} \log \mathbf{E}\left[e^{-\tau \hat{S}_n}\, \mathbb{1}_{\{\hat{S}_n \geq 0\}}\right] &\geq \frac{1}{n} \log \mathbf{E}\left[e^{-\tau \hat{S}_n}\, \mathbb{1}_{\{0 \leq \hat{S}_n \leq c\sqrt{n}\}}\right] \\
&\geq \frac{1}{n} \log \left(e^{-\tau c \sqrt{n}}\, \mathbf{P}\left[\frac{\hat{S}_n}{\sqrt{n}} \in [0,c]\right]\right) \\
&\overset{n\to\infty}{\longrightarrow} \;\lim_{n\to\infty} \frac{-\tau c \sqrt{n}}{n} + \lim_{n\to\infty} \frac{1}{n} \log \left(\mathcal{N}_{0,\mathbf{Var}[\hat{X}_1]}([0,c])\right) \\
&= 0.
\end{aligned}$$

□

Proof. Assume that $(\mu_\varepsilon)_{\varepsilon>0}$ satisfies an LDP with rate functions I and J. Then, for every $x \in E$ and $\delta > 0$,

$$\begin{aligned} I(x) &\geq \inf I(B_\delta(x)) \\ &\geq -\liminf_{\varepsilon\to 0} \varepsilon \log\big(\mu_\varepsilon(B_\delta(x))\big) \\ &\geq -\limsup_{\varepsilon\to 0} \varepsilon \log\big(\mu_\varepsilon\big(\overline{B_\delta(x)}\big)\big) \\ &\geq \inf J\big(\overline{B_\delta(x)}\big) \stackrel{\delta\to 0}{\longrightarrow} J(x). \end{aligned}$$

Hence $I(x) \geq J(x)$. Similarly, we get $J(x) \geq I(x)$. □

Lemma 23.9. *Let $N \in \mathbb{N}$ and let a_ε^i, $i = 1, \ldots, N$, $\varepsilon > 0$, be nonnegative numbers. Then*

$$\limsup_{\varepsilon\to 0} \varepsilon \log \sum_{i=1}^N a_\varepsilon^i = \max_{i=1,\ldots,N} \limsup_{\varepsilon\to 0} \varepsilon \log(a_\varepsilon^i).$$

Proof. The sum and maximum differ at most by a factor N:

$$\max_{i=1,\ldots,N} \varepsilon \log(a_\varepsilon^i) \leq \varepsilon \log \sum_{i=1}^N a_\varepsilon^i \leq \varepsilon \log(N) + \max_{i=1,\ldots,N} \varepsilon \log(a_\varepsilon^i).$$

The maximum and limit (superior) can be interchanged and hence

$$\begin{aligned} \max_{i=1,\ldots,N} \limsup_{\varepsilon\to 0} \varepsilon \log(a_\varepsilon^i) &= \limsup_{\varepsilon\to 0} \varepsilon \log\Big(\max_{i=1,\ldots,N} a_\varepsilon^i\Big) \\ &\leq \limsup_{\varepsilon\to 0} \varepsilon \log\Big(\sum_{i=1}^N a_\varepsilon^i\Big) \\ &\leq \limsup_{\varepsilon\to 0} \varepsilon \log(N) + \max_{i=1,\ldots,N} \limsup_{\varepsilon\to 0} \varepsilon \log(a_\varepsilon^i) \\ &= \max_{i=1,\ldots,N} \limsup_{\varepsilon\to 0} \varepsilon \log(a_\varepsilon^i). \end{aligned}$$

□

Example 23.10. Let $X_1, X_2, \ldots$ be i.i.d. real random variables that satisfy the condition of Cramér's theorem (Theorem 23.3); i.e., $\Lambda(t) = \log(\mathbf{E}[e^{tX_1}]) < \infty$ for every $t \in \mathbb{R}$. Furthermore, let $S_n = X_1 + \ldots + X_n$ for every n. We will show that Cramér's theorem implies that $P_n := \mathbf{P}_{S_n/n}$ satisfies an LDP with rate n and with good rate function $I(x) = \Lambda^*(x) := \sup_{t\in\mathbb{R}}(tx - \Lambda(t))$. Without loss of generality, we can assume that $\mathbf{E}[X_1] = 0$. The function I is everywhere finite, continuous, strictly convex and has its unique minimum at $I(0) = 0$. Cramér's theorem says that $\lim_{n\to\infty} \frac{1}{n} \log(P_n([x, \infty))) = -I(x)$ for $x > 0$ and (by symmetry) $\lim_{n\to\infty} \frac{1}{n} \log(P_n((-\infty, x])) = -I(x)$ for $x < 0$. Clearly, for $x > 0$,

$$\begin{aligned} -I(x) &\geq \lim_{n\to\infty} \frac{1}{n} \log P_n((x, \infty)) \\ &\geq \sup_{\varepsilon>0} \lim_{n\to\infty} \frac{1}{n} \log P_n([x+\varepsilon, \infty)) = -\inf_{\varepsilon>0} I(x+\varepsilon) = -I(x). \end{aligned}$$

Similarly, $\lim_{n\to\infty} \frac{1}{n} \log P_n((-\infty, x)) = -I(x)$ for $x < 0$.

The main work has been done by showing that the family $(P_n)_{n\in\mathbb{N}}$ satisfies conditions (LDP 1) and (LDP 2) at least for unbounded intervals. It remains to show by some standard arguments (LDP 1) and (LDP 2) for *arbitrary* open and closed sets, respectively.

First assume that $C \subset \mathbb{R}$ is closed. Define $x_+ := \inf\big(C \cap [0,\infty)\big)$ as well as $x_- := \sup\big(C \cap (-\infty, 0]\big)$. By monotonicity of I, on $(-\infty, 0]$ and $[0,\infty)$, we get $\inf I(C) = I(x_-) \wedge I(x_+)$ (with the convention $I(-\infty) = I(\infty) = \infty$). If $x_- = 0$ or $x_+ = 0$, then $\inf(I(C)) = 0$, and (LDP 2) holds trivially. Now let $x_- < 0 < x_+$. Using Lemma 23.9, we get

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} & \log P_n(C) \\
&\le \limsup_{n\to\infty} \frac{1}{n} \log\big(P_n\big((-\infty, x_-]\big) + P_n\big([x_+,\infty)\big)\big) \\
&= \max\Big\{\limsup_{n\to\infty} \frac{1}{n} \log P_n\big((-\infty, x_-]\big), \limsup_{n\to\infty} \frac{1}{n} \log P_n\big([x_+,\infty)\big)\Big\} \\
&= \max\big\{-I(x_-), -I(x_+)\big\} = -\inf I(C).
\end{aligned}$$

This shows (LDP 2).

Now let $U \subset \mathbb{R}$ be open. Let $x \in U \cap (0,\infty)$ (if such an x exists). Then there exists an $\varepsilon > 0$ with $(x-\varepsilon, x+\varepsilon) \subset U \cap (0,\infty)$. Now

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log P_n\big((x-\varepsilon,\infty)\big) &= -I(x-\varepsilon) > -I(x+\varepsilon) \\
&= \lim_{n\to\infty} \frac{1}{n} \log P_n\big([x+\varepsilon,\infty)\big).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log P_n(U) &\ge \lim_{n\to\infty} \frac{1}{n} \log P_n((x-\varepsilon, x+\varepsilon)) \\
&= \lim_{n\to\infty} \frac{1}{n} \log\big(P_n\big((x-\varepsilon,\infty)\big) - P_n\big([x+\varepsilon,\infty)\big)\big) \\
&= \lim_{n\to\infty} \frac{1}{n} \log\big(P_n\big((x-\varepsilon,\infty)\big)\big) = -I(x-\varepsilon) \ \ge\ -I(x).
\end{aligned}$$

Similarly, this also holds for $x \in U \cap (-\infty, 0)$; hence

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(U) \ \ge\ \inf\ I(U \setminus \{0\}) \ =\ \inf\ I(U).$$

Note that in the last step, we used the fact that U is open and that I is continuous. This shows the lower bound (LDP 1). $\diamond$

In fact, the condition $\Lambda(t) < \infty$ for all $t \in \mathbb{R}$ can be dropped. Since $\Lambda(0) = 0$, we have $\Lambda^*(x) \ge 0$ for every $x \in \mathbb{R}$. The map Λ^* is a convex rate function but is, in