27. GIRSANOV'S THEOREM

We will now move to another technique that can be used to remove, or modify, the drift term in an SDE but that has other uses as well. The technique, whose ultimate formulation bears Girsanov's name, is based on the idea of "exponential change of measure" which is a standard method to adjust the mean in large deviation theory.

27.1 Exponential change of measure.

To illustrate how exponential change of measure works in general, we note:

Lemma 27.1 (Exponential tilt) *Let $X_1,\dots,X_n$ be independent and $\varphi_i(\lambda) := E\mathrm{e}^{\lambda X_i} < \infty$ for all $\lambda \in \mathbb{R}$ and all $i = 1,\dots,n$. For $A \in \sigma(X_1,\dots,X_n)$ and $\lambda = (\lambda_1,\dots,\lambda_n) \in \mathbb{R}^n$, set*

$$P_\lambda(A) := E\biggl(1_A \exp\Bigl\{\sum_{i=1}^n \bigl[\lambda_i X_i - \log \varphi_i(\lambda_i)\bigr]\Bigr\}\biggr) \tag{27.1}$$

Then $(X_1,\dots,X_n)$ remain to be independent under P_λ with expectation

$$E_\lambda(X_i) = \frac{\varphi_i'(\lambda_i)}{\varphi_i(\lambda_i)} \tag{27.2}$$

for all $i = 1,\dots,n$.

Proof. The product structure of P_λ is verified directly from (27.1). For (27.2) the product structure shows $E_\lambda(X_i) = \varphi_i(\lambda_i)^{-1}E(X_i\mathrm{e}^{\lambda X_i}) = \varphi_i'(\lambda_i)/\varphi_i(\lambda_i)$. ☐

The point of the above lemma is that, since $\lambda \mapsto \varphi_i(\lambda)$ is strictly convex (under the assumption of having all exponential moments), its logarithmic derivative is strictly increasing. It follows that, by tuning λ_i appropriately, we can adjust the expectation of each X_i to whatever value in the interior of the convex hull of the original support of X_i.

The latter observation is exactly what makes the exponential change of measure technique indispensable in the study of *large deviations*: For $X_1,\dots,X_n$ i.i.d. with $\varphi(\lambda) := E\mathrm{e}^{\lambda X}$ finite for all $\lambda \in \mathbb{R}$, given any a in the interior of the convex hull of the support of X_1, we can find λ so that $\varphi'(\lambda)/\varphi(\lambda) = a$. Then for any $\epsilon > 0$, the Intermediate Value Theorem guarantees

$$\begin{aligned}
&P\biggl(\frac{1}{n}\sum_{i=1}^n X_i \in [a-\epsilon, a+\epsilon]\biggr) \\
&\qquad = \mathrm{e}^{[a+O(\epsilon)]n} E\biggl(\exp\Bigl\{a\sum_{i=1}^n X_i\Bigr\}\,;\, \frac{1}{n}\sum_{i=1}^n X_i \in [a-\epsilon, a+\epsilon]\biggr) \\
&\qquad = \bigl[\varphi(\lambda)\mathrm{e}^{a+O(\epsilon)}\bigr]^n P_\lambda\biggl(\frac{1}{n}\sum_{i=1}^n X_i \in [a-\epsilon, a+\epsilon]\biggr)
\end{aligned} \tag{27.3}$$

where $O(\epsilon)$ is a quantity in $[-\epsilon,\epsilon]$. Thanks to the choice of λ, the Weak Law of Large Numbers then tells us that the probability on the right tends to 1 as $n \to \infty$ and so the exponential "cost" of having $\frac{1}{n}\sum_{i=1}^n X_i \approx a$ is to the leading order $[\mathrm{e}^{\lambda a}\varphi(\lambda)]^n$. (A convexity argument ensures that $\lambda a + \log \varphi(\lambda) \leqslant 0$ with the inequality strict whenever $a \neq EX_1$.) The above is, more or less, the proof of the celebrated *Cramér theorem*.

The title of Lemma 27.1 brings up another name — "tilting" — for the above technique as this is what the exponential factor — sometimes referred to as the "tilt" — does to the initial distribution. For general underlying random variables, a drawback of the "tilting" technique is that is that varying λ_i is inevitably accompanied by changes to the whole distribution of the X_i's. As our next lemma shows, this is not the case when Gaussian random variables are of concern as these are determined by only two parameters — the mean and covariance structure.

Lemma 27.2 (Tilted Gaussian law) *Let $X = (X_1, \dots, X_n)$ be a multivariate normal with mean zero and covariance $C = \{\operatorname{Cov}(X_i, X_j)\}_{i,j=1}^n$. For $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ let*

$$P_\lambda(A) := E\bigl(1_A \mathrm{e}^{\lambda\cdot X - \frac12 \lambda\cdot C\lambda}\bigr) \tag{27.4}$$

Then

$$X - C\lambda \text{ under } P_\lambda \overset{\text{law}}{=} X \text{ under } P \tag{27.5}$$

Here "$\cdot$" denotes the Euclidean inner product in $\mathbb{R}^n$.

Proof. Let $t \in \mathbb{R}^n$. Then

$$\begin{aligned} E_\lambda\bigl(\mathrm{e}^{t\cdot(X-C\lambda)}\bigr) &= E\bigl(\mathrm{e}^{t\cdot(X-C\lambda)+\lambda\cdot X-\frac12\lambda\cdot C\lambda}\bigr) \\ &= E\bigl(\mathrm{e}^{(\lambda+t)\cdot X-\frac12(\lambda+t)\cdot C(\lambda+t)}\bigr)\mathrm{e}^{\frac12 t\cdot Ct} = \mathrm{e}^{\frac12 t\cdot Ct} \end{aligned} \tag{27.6}$$

Since the latter is the Laplace transform of $\mathcal{N}(0, C)$, the claim follows using the Curtiss Theorem and the Cramér-Wold device. □

The mean-zero restriction is made for convenience of expression; if the mean equals μ, then replace X by $X - \mu$ above. While the underlying Gaussian nature allowed us to treat a fully general case in one step, its special case of independent Gaussians could have also been dealt with inductively, by integrating out one variable at the time. This leads to another version of "exponential change of measure" which, this time, is very close to the one we are ultimately aiming for.

Lemma 27.3 (Discrete-time Girsanov Theorem) *Let $X_1, \dots, X_n$ be independent with $X_i = \mathcal{N}(0, \sigma_i^2)$ for all $i = 1, \dots, n$. For each $k = 1, \dots, n$, let $\lambda_k \colon \mathbb{R}^{k-1} \to \mathbb{R}$ be a Borel-measurable function. (In particular, λ_1 is a constant.) Then*

$$P_\lambda(A) := E\left(1_A \exp\left\{\sum_{k=1}^n \Bigl[\lambda_k(X_1, \dots, X_{k-1})X_k - \tfrac12 \lambda_k(X_1, \dots, X_{k-1})^2 \sigma_k^2\Bigr]\right\}\right) \tag{27.7}$$

is a probability measure and

$$\left\{X_k - \sum_{j=1}^k \lambda_j(X_1, \dots, X_{j-1})\sigma_j^2\right\}_{k=1}^n \text{ under } P_\lambda \overset{\text{law}}{=} X \text{ under } P \tag{27.8}$$

Proof. For $k = 1, \dots, n$, let

$$M_k := \prod_{j=1}^{k} \mathrm{e}^{\lambda_j(X_1,\dots,X_{j-1})X_j - \frac{1}{2}\lambda_j(X_1,\dots,X_{j-1})^2\sigma_j^2} \tag{27.9}$$

and set $\mathcal{F}_k := \sigma(X_1, \dots, X_k)$. We claim that $\{M_k\}_{k=1}^n$ is a martingale for filtration $\{\mathcal{F}_k\}_{k=1}^n$ under P. To see this note that

$$E\big(M_{k+1}\big|\mathcal{F}_k\big) = M_k E\big(\mathrm{e}^{\lambda(X_1,\dots,X_k)X_{k+1} - \frac{1}{2}\lambda(X_1,\dots,X_k)^2\sigma_{k+1}^2} \,\big|\, \mathcal{F}_k\big) \tag{27.10}$$

where the expectation on the right equals one by the fact that we can regard $\lambda(X_1, \dots, X_k)$ as a constant under the conditional expectation. It follows that $E(M_k) = 1$ for all $k = 1, \dots, n$ and since

$$\forall A \in \mathcal{F}_n\colon \quad P_\lambda(A) = E\big(1_A M_n\big), \tag{27.11}$$

we also get that P_λ is a probability measure.

In order to prove (27.8), abbreviate

$$\widetilde{X}_k := X_k - \sum_{j=1}^{k} \lambda_j(X_1, \dots, X_{j-1})\sigma_j^2 \tag{27.12}$$

and, given any $t = (t_1, \dots, t_n) \in \mathbb{R}^n$, let

$$N_k := \prod_{j=1}^{k} \mathrm{e}^{t_j \cdot \widetilde{X}_j - \frac{1}{2}\sigma_j^2 t_j^2} \tag{27.13}$$

A calculation shows that

$$N_k M_k = \prod_{j=1}^{k} \mathrm{e}^{[t_j + \lambda_k(X_1,\dots,X_{j-1})]X_j - \frac{1}{2}[t_j + \lambda_j(X_1,\dots,X_{j-1})]^2\sigma_j^2} \tag{27.14}$$

and so, by the same argument as above, also $\{N_k M_k\}_{k=1}^n$ is a martingale under P. It follows that

$$1 = E(N_0 M_0) = E(N_n M_n) = E_\lambda(N_n) \tag{27.15}$$

Using the explicit form of N_n, this can be written as

$$E_\lambda \mathrm{e}^{t \cdot \widetilde{X}} = \mathrm{e}^{\frac{1}{2} t \cdot \sigma^2 t} = E\mathrm{e}^{t \cdot X} \tag{27.16}$$

The Curtiss and Cramér-Wold theorems now imply the claim. □

27.2 Statement of Girsanov's theorem and its proof.

The ultimate result of this lecture is a continuous-time version of above lemma. The main difference is that here we have to assume that the tilted measure is a probability, rather than derive it as part of the proof.

Theorem 27.4 (Girsanov 1960) *Assume a Brownian motion B and a Brownian filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$ are given. For $Y \in \mathcal{V}^{\text{loc}}$ and $t \geqslant 0$ set*

$$M_t := \exp\biggl\{\int_0^t Y_s \mathrm{d}B_s - \frac{1}{2}\int_0^t Y_s^2 \mathrm{d}s\biggr\} \tag{27.17}$$

and let

$$\forall A \in \mathcal{F}_t\colon \quad \widetilde{P}(A) := E\bigl(1_A M_t\bigr) \tag{27.18}$$

If $EM_t = 1$, then $\widetilde{P}$ is a probability measure and

$$\biggl\{B_s - \int_0^s Y_u \mathrm{d}u\colon s \in [0,t]\biggr\} \text{ under } \widetilde{P} \tag{27.19}$$

is a standard Brownian motion.

We start by a lemma that reveals why the assumption $EM_t = 1$ is important:

Lemma 27.5 *For M as in Theorem 27.4, if $EM_t = 1$ then $\{M_{s\wedge t}\colon s \geqslant 0\}$ is a martingale.*

Proof. First note that, by the Itô formula,

$$\mathrm{d}M_t = M_t\bigl(Y_t \mathrm{d}B_t - \tfrac{1}{2}Y_t^2 \mathrm{d}t\bigr) + \frac{1}{2}M_t Y_t^2 \mathrm{d}t = M_t Y_t \mathrm{d}B_t \tag{27.20}$$

and so $\{M_s\colon s \geqslant 0\}$ is a local martingale. Letting

$$\tau_n := \inf\biggl\{u \geqslant 0\colon \int_0^u Y_s \mathrm{d}B_s \geqslant n\biggr\} \tag{27.21}$$

the process $\{M_{s\wedge\tau_n}\colon s \geqslant 0\}$ is bounded uniformly by e^n and is thus a martingale with

$$\forall s \geqslant 0\colon \quad EM_{t\wedge s\wedge\tau_n} = EM_0 = 1 \tag{27.22}$$

Since $\tau_n < \infty$ a.s., the Optional Sampling Theorem then shows

$$\forall s \geqslant 0\colon \quad E\bigl(M_{t\wedge\tau_n} \,\big|\, \mathcal{F}_s\bigr) = M_{t\wedge s\wedge\tau_n} \quad \text{a.s.} \tag{27.23}$$

As $\tau_n \to \infty$ P-a.s. as $n \to \infty$ by the fact that $Y \in \mathcal{V}^{\text{loc}}$, we have $M_{u\wedge\tau_n} \to M_u$ pointwise a.s. Fatou's lemma applied in (27.22) shows

$$\forall s \geqslant 0\colon \quad E(M_{t\wedge s}) \leqslant 1 \tag{27.24}$$

while its conditional version used in (27.23) gives

$$\forall s \geqslant 0\colon \quad E(M_t|\mathcal{F}_s) \leqslant M_{t\wedge s} \quad \text{a.s.} \tag{27.25}$$

But the expectation of the left hand side equals $EM_t = 1$ which in conjunction with (27.24) forces $EM_{t\wedge s} = 1$ for all $s \geqslant 0$. This is only possible if equality holds in (27.25) a.s. Hence, $\{M_{s\wedge t}\colon s \geqslant 0\}$ is a martingale, as claimed. □

We are now ready to give:

Proof of Theorem 27.4. In order to prove (27.19), we will invoke a continuous-time version of the argument (27.13–27.16) but, since integrability is no longer automatic, we will rely on characteristic functions instead of Laplace transforms. Abbreviate

$$\widetilde{B}_s := B_s - \int_0^s Y_u \mathrm{d}u \tag{27.26}$$

Given any $Z \in \mathcal{V}_0$, set

$$N_s := \exp\Big\{\mathrm{i}\int_0^s Z_s \mathrm{d}\widetilde{B}_s + \frac12\int_0^s Z_s^2 \mathrm{d}s\Big\} \tag{27.27}$$

Note that $\widetilde{B}$ is used in the first integral. This shows that

$$M_sN_s = \exp\Big\{\int_0^s (Y_s+\mathrm{i}Z_s)\mathrm{d}\widetilde{B}_s - \frac12\int_0^s (Y_s+\mathrm{i}Z_s)^2\mathrm{d}s\Big\} \tag{27.28}$$

The Itô formula then gives

$$\mathrm{d}M_sN_s = M_sN_s(Y_s+\mathrm{i}Z_s)\mathrm{d}B_s \tag{27.29}$$

and so $\{M_sN_s\colon s \geqslant 0\}$ is a local martingale.

Let τ_n be as in (27.21). Since $Z \in \mathcal{V}_0$ gives $|N_s| \leqslant \exp\{\frac12 t\sup_{u\geqslant 0}\|Z_u\|_\infty\}$, the process $\{N_{s\wedge t}\colon s \geqslant 0\}$ is bounded. Then (27.23) gives that $\{M_{s\wedge t\wedge\tau_n}N_{s\wedge t\wedge\tau_n}\colon s\geqslant 0\}$ is uniformly integrable and so is a martingale. This means that, for all $s \geqslant 0$ and $n \geqslant 1$,

$$E\big(M_{t\wedge\tau_n}N_{t\wedge\tau_n}\,\big|\,\mathcal{F}_s\big) = M_{s\wedge t\wedge\tau_n}N_{s\wedge t\wedge\tau_n} \quad \text{a.s.} \tag{27.30}$$

The proof of Lemma 27.5 shows that $M_{t\wedge\tau_n} \to M_t$ both P-a.s. and in $L^1(P)$ and so we also get that $M_{t\wedge\tau_n}N_{t\wedge\tau_n} \to M_tN_t$ in $L^1(P)$. Using this in (27.30) yields

$$\forall s \geqslant 0\colon \quad E(M_tN_t\,|\,\mathcal{F}_s) = M_{s\wedge t}N_{s\wedge t} \quad \text{a.s.} \tag{27.31}$$

Applying this to $s = 0$ shows

$$1 = E(M_0N_0) = E(M_tN_t) = \widetilde{E}(N_t) \tag{27.32}$$

For the choice

$$Z_s := \sum_{j=1}^n \lambda_j 1_{(t_{j-1},t_j]}(s) \tag{27.33}$$

where $0 = t_0 < t_1 < \dots < t_n = t$ and $\lambda_1,\dots,\lambda_n \in \mathbb{R}$, this becomes

$$\widetilde{E}\bigg(\exp\Big\{\mathrm{i}\sum_{j=1}^n \lambda_j(\widetilde{B}_{t_j} - \widetilde{B}_{t_{j-1}})\Big\}\bigg) = \exp\Big\{-\frac12\sum_{j=1}^n \lambda_j^2(t_j - t_{j-1})\Big\} \tag{27.34}$$

Using the Cramér-Wold device we get that, under $\widetilde{P}$, the process $\widetilde{B}$ has the same finite-dimensional distributions as B under P. Since $\widetilde{B}$ is continuous, it is a standard Brownian motion as claimed. □

The above proof is somewhat subtle because it works under the minimal possible conditions. Indeed, a slight upgrade of Lemma 27.3 gives the statement for all Y simple without the proviso $EM_t = 1$ (which comes automatically in this case) so another strategy to address the general case could rely on approximation of Y by simple processes. The problem here is that, while the $L^2([0,t])$-convergence $\int_0^t [Y_s^{(n)} - Y_s]^2 \mathrm{d}s \to 0$ in probability implies that the associated exponential martingales $M_t^{(n)}$ converge to M_t in probability (see Lemma 12.9), we need $M_t^{(n)} \to M_t$ in L^1 which entails uniform integrability for which we have no direct argument.

27.3 Why do we care?.

A natural question is what makes Girsanov's theorem useful. There are a number of reasons but one perhaps most convincing comes in:

Corollary 27.6 *Let $x \in \mathbb{R}$ and let X a standard Brownian motion started from $X_0 = x$ under the probability measure P^x. Given $t > 0$, suppose that $a\colon [0,t] \times \mathbb{R} \to \mathbb{R}$ is Borel measurable such that the integrals in*

$$M_t := \exp\left\{\int_0^t a(s, X_s)\mathrm{d}X_s - \frac{1}{2}\int_0^t a(s, X_s)^2 \mathrm{d}s\right\} \tag{27.35}$$

are well defined and $E^x M_t = 1$ holds. Then, under the probability measure $\widetilde{P}^x$ on $(\Omega, \mathcal{F}_t)$ defined by $\widetilde{P}^x(A) := E^x(1_A M_t)$, the process $\{B_s\colon s \in [0,t]\}$ defined by

$$B_s := X_s - \int_0^s a(s, X_s)\mathrm{d}s \tag{27.36}$$

is a standard Brownian motion with $\widetilde{P}^x(B_0 = x) = 1$. In particular, the SDE

$$\mathrm{d}X_u = a(u, X_u)\mathrm{d}u + \mathrm{d}B_u \tag{27.37}$$

admits a weak solution up to time t with $X_0 = x$.

Proof. That B is a standard Brownian motion under $\widetilde{P}^x$ follows from Girsanov's Theorem and the observation that the starting point does not matter for the proof. Writing (27.36) as

$$X_s = \int_0^s a(u, X_u)\mathrm{d}u + B_s \tag{27.38}$$

then gives X the meaning of the solution to the SDE (27.37). □

Recall form Lemma 26.1 that we can generally solve the SDE (27.37) path-by-path by converting it to an ODE. However, this requires an existence theorem (e.g., Peano's theorem) which generally needs that $x \mapsto a(t,x)$ is at least continuous. The upshot of the solution via Girsanov's theorem is that no regularity of $t, x \mapsto a(t,x)$ beyond joint measurability (and existence of the integrals in M_t along with $E^x M_t = 1$) is required to get a weak solution for a.e. path of Brownian motion.

Further reading: Karatzas-Shreve, Section 3.5