10 Change of Probability Measure

10.1 Change of Measure for Random Variables

Let $(\Omega, \mathcal{F})$ be a measurable space on which two probability measures P and Q are defined. We make the following definitions:

- Q is absolutely continuous with respect to P ($Q << P$) if $P(A) = 0 \Rightarrow Q(A) = 0$ for $A \in \mathcal{F}$.
- P and Q are equivalent ($P \sim Q$) if $P << Q$ and $Q << P$.
- P and Q are singular ($P \perp Q$) if $P(A) = 0$ and $Q(A) = 1$ for some $A \in \mathcal{F}$.

If $Q << P$, then by the Radon-Nikodym theorem there exists a (unique except for the values on a null-event) random variable $\Lambda \geq 0$ with $E(\Lambda) = 1$ such that

$$Q(A) = E_P(\Lambda I_A) = \int_A \Lambda \, dP \quad \text{for } A \in \mathcal{F}.$$

For obvious reasons one uses the notation dQ/dP for Λ and it is called the derivative and also the density of Q with respect to P.

Conversely, given a probability space $(\Omega, \mathcal{F}, P)$ and a random variable $\Lambda \geq 0$ with $E(\Lambda) = 1$, one may define a new probability Q on $(\Omega, \mathcal{F})$ by $Q(A) = E(\Lambda I_A)$ for $A \in \mathcal{F}$. In that case we have $Q << P$ with $dQ/dP = \Lambda$.

There is an abstract version of Bayes' theorem from elementary probability theory called the general Bayes' formula: If $Q << P$ with $dQ/dP = \Lambda$ and $\mathcal{G} \subseteq \mathcal{F}$ is a σ-field, then we have

$$E_Q(X|\mathcal{G}) = \frac{E_P(X\Lambda|\mathcal{G})}{E_P(\Lambda|\mathcal{G})}$$

for random variables X with $E_Q(|X|) < \infty$.

Proof For $A \in \mathcal{G}$ we have

$$E_Q\left(1_A \frac{E_P(X\Lambda|\mathcal{G})}{E_P(\Lambda|\mathcal{G})}\right) = E_P\left(1_A \Lambda \frac{E_P(X\Lambda|\mathcal{G})}{E_P(\Lambda|\mathcal{G})}\right) = E_P\left(E_P\left(1_A \Lambda \frac{E_P(X\Lambda|\mathcal{G})}{E_P(\Lambda|\mathcal{G})}\Big|\mathcal{G}\right)\right)$$

$$= E_P\left(1_A \frac{E_P(X\Lambda|\mathcal{G})}{E_P(\Lambda|\mathcal{G})} E_P(\Lambda|\mathcal{G})\right) = E_P(E_P(1_A X\Lambda|\mathcal{G})) = E_P(1_A \Lambda X)$$

$$= E_Q(1_A X)$$

Let X be a continuously distributed (in the elementary sense) random variable with PDF $f: \mathbb{R} \to \mathbb{R}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. For any other PDF $g: \mathbb{R} \to \mathbb{R}$ such that $g^{-1}(\{0\}) \supseteq f^{-1}(\{0\})$ we may define a new probability measure Q on $(\Omega, \mathcal{F})$

$E_Q(e^{j\omega X}) = E\left(\frac{e^{j\omega X} g(X)}{f(X)}\right) = \int_{-\infty}^{+\infty} e^{j\omega x} g(x)\,dx = (\mathcal{F} g)(\omega)$

by

$$Q(A) = E\left(\frac{g(X)I_A}{f(X)}\right) \quad \text{for } A \in \mathcal{F}.$$

It is easy to see that X will have PDF g when viewed a random variable on $(\Omega, \mathcal{F}, Q)$. Note the definitions of a random variable as a measurable function $X : \Omega \to \mathbb{R}$ doesn't involve any probability measure. The probability distribution of X will depend on what probability measure we select to use. If we change that measure the probability distribution of (the one and same random variable) X will change in general.

Let X be a normal $N(\mu_1, \sigma_1^2)$-distributed random variable defined on a probability space $(\Omega, \mathcal{F}, P)$. Taking

$$\Lambda = \frac{\sigma_1}{\sigma_2} \exp\left(\frac{(X-\mu_1)^2}{2\sigma_1^2} - \frac{(X-\mu_2)^2}{2\sigma_2^2}\right)$$

it follows from what we did in the previous paragraph that X is normal $N(\mu_2, \sigma_2^2)$-distributed on the probability space $(\Omega, \mathcal{F}, Q)$ with probability measure $Q(A) = E_P(\Lambda I_A)$ for $A \in \mathcal{F}$ so that $Q << P$ with $dQ/dP = \Lambda(T)$.

Example (Importance sampling). One cannot estimate $P(N(0,1) > 10^6)$ by calcutating $\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > 10^6\}}$ for $\{X_i\}_{i=1}^{n}$ independent $N(0,1)$. But taking

$$\Lambda = \exp\left(\frac{X^2}{2} - \frac{(X-10^6)^2}{2}\right) = e^{10^6 X - \frac{1}{2}10^{12}} \quad \text{and} \quad Q(A) = E_P(1_A \Lambda)$$

we have

$$E_P(1_{\{X>10^6\}}) = E_Q(1_{\{X>10^6\}}\Lambda^{-1}) \approx \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > 10^6\}} e^{-10^6 X_i + \frac{1}{2}10^{12}}$$

for $X_i \sim N(10^6, 1)$ which is same as

$$\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > 0\}} e^{-10^6 X_i - \frac{1}{2}10^{12}} \quad \text{for } X_i \; N(0,1).$$

10.2 Change of Measure for Processes

Let $\{\Lambda(t)\}_{t\in[0,T]}$ be a positive martingale on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, P)$ such that $E(\Lambda(T)) = 1$. Define a new probability Q on $(\Omega, \mathcal{F})$ by $Q(A) = E(\Lambda(T)I_A)$ for $A \in \mathcal{F}$. By application of the general Bayes' formula it follows that

$$E_Q(X|\mathcal{F}_t) = \frac{E_P(X\Lambda(T)|\mathcal{F}_t)}{E_P(\Lambda(T)|\mathcal{F}_t)} = \frac{E_P(X\Lambda(T)|\mathcal{F}_t)}{\Lambda(t)} = E_P\left(\frac{\Lambda(T)}{\Lambda(t)} X \middle| \mathcal{F}_t\right), \; t \in [0,T],$$

whenever X is a random variable with $E_Q(|X|) < \infty$. If in addition X is $\mathcal{F}_t$-measurable for a $t \in [0,T]$ it further holds that, by using above formula with t

replaced by s and then "towering" inside with $\mathcal{F}_t$

$$E_Q(X|\mathcal{F}_s) = E_P\left(\frac{\Lambda(t)}{\Lambda(s)}X \,\middle|\, \mathcal{F}_s\right) \quad \text{for } s\in[0,t].$$

Now, an adapted process $\{M(t)\}_{t\in[0,T]}$ is a Q-martingale if and only if $\{\Lambda(t)M(t)\}_{t\in[0,T]}$ is a P-martingale. In particular $\{1/\Lambda(t)\}_{t\in[0,T]}$ is a Q-martingale.

Proof $\boxed{\Leftarrow}$ $E_Q(M(t)|\mathcal{F}_s) = E_P\left(\frac{\Lambda(t)}{\Lambda(s)}M(t)\,\middle|\,\mathcal{F}_s\right) = \Lambda(s)M(s)/\Lambda(s) = M(s)$

$\boxed{\Rightarrow}$ $M(s) = E_Q(M(t)|\mathcal{F}_s) = E_P\left(\frac{\Lambda(t)}{\Lambda(s)}M(t)\,\middle|\,\mathcal{F}_s\right) = \frac{1}{\Lambda(s)}E_P(\Lambda(t)M(t)|\mathcal{F}_s)$. #

As we have seen in the previous paragraph (and unsurprisingly one must say) conditional expectations change with change of probability measure. However, quadratic variation and covariation do not. More specifically, if a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ satisfy $X_n \to_P X$ and if $Q << P$ then it follows (from so called absolute continuity of the Lebesgue integral) that $X_n \to_Q X$. And so if a quadratic variation or covariation is well-defined in the sense of convergence in probability on the probability space $(\Omega, \mathcal{F}, P)$, then it is well-defined on the probability space $(\Omega, \mathcal{F}, Q)$ as well and with the same value.

Girsanov's theorem for change of drift in diffusions: Let $\{X(t)\}_{t\in[0,T]}$ satisfy the (BM) SDE

$$dX(t) = \mu_1(X(t),t)\,dt + \sigma(X(t),t)\,dB(t) \quad \text{for } t\in[0,T],$$

where σ is assumed to be strictly positive for simplicity. Select a new drift coefficient $\mu_2 : \mathbb{R}\times[0,T]\to\mathbb{R}$ and set

$$H(t) = \frac{\mu_2(X(t),t) - \mu_1(X(t),t)}{\sigma(X(t),t)} \quad \text{for } t\in[0,T].$$

Suppose that $\{\mathcal{E}(\int H\,dB)(t)\}_{t\in[0,T]}$ is a martingale and define a new probability measure Q by

$$\frac{dQ}{dP} = \mathcal{E}(\textstyle\int H\,dB)(T) = \exp\left(\int_0^T H\,dB - \frac{1}{2}\int_0^T H(t)^2\,dt\right).$$

Then the process

$$W(t) = B(t) - \int_0^t H(s)\,ds \quad \text{for } t\in[0,T]$$

is a BM on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t^B\}_{t\in[0,T]}, Q)$ and X satisfies the SDE

$$dX(t) = \mu_2(X(t),t)\,dt + \sigma(X(t),t)\,dW(t) \quad \text{for } t\in[0,T]$$

(on that probability space). A direct proof that W is BM is by Lévy's characterisation of BM: As $W(t)$ obviously has quadrati variation t it is enough to check that $W(t)$ is a Q-martingale which in turn is the case iff

$\mathcal{E}(\int H\,dB)(t)\left(B(t)-\int_0^t H(s)\,ds\right)$ is a P-martingale.
However, taking the differential of that process we obtain

$$\mathcal{E}(\int H\,dB)(t)\,H(t)\,dB(t)\left(B(t)-\int_0^t H(s)\,ds\right)$$

$$+\ \mathcal{E}(\int H\,dB)(t)\,dB(t) - \mathcal{E}(\int H\,dB)(t)\,H(t)\,dt$$

$$+\ \mathcal{E}(\int H\,dB)(t)\,H(t)\,dB(t)\,dB(t)$$ where the last two terms cancel each other.

Further $X(t)$ satisfies the SDE

$$\mu_2(X(t),t)\,dt + \sigma(X(t),t)\,dW(t)$$

$$= \mu_2(X(t),t)\,dt + \sigma(X(t),t)\left(dB(t) - \frac{\mu_2(X(t),t)-\mu_1(X(t),t)}{\sigma(X(t),t)}\,dt\right)$$

$$= \mu_1(X(t),t)\,dt + \sigma(X(t),t)\,dB(t) = dX(t) \qquad \#$$

<u>Example</u> Taking $\sigma(x,t)=1$, $\mu_1(x,t)=-\alpha x$ and $\mu_2(x,t)=0$ we have $dX(t)=-\alpha X(t)\,dt + dB(t)$ and $dX(t)=dW(t)$

so the Ornstein-Uhlebeck process $X(t)$ on $(\Omega,\mathcal{F}_t,P)$ is BM on $(\Omega,\mathcal{F}_t,Q)$ for Q given by

$$\frac{dQ}{dP} = \mathcal{E}(\int H\,dB)(T) \quad \text{with } H(t)=\alpha X(t) \text{ so}$$

$$dQ/dP = \exp\left(\int_0^T \alpha X(t)\,dB(t) - \frac{1}{2}\int_0^T \alpha^2 X(t)^2\,dt\right).$$

10.4 Likelihood Functions

Now we will see how Girsanov's theorem for change of drift in diffusions can be used to make statistical inference (hypotheses testing and estimation) of the drift coefficient.

Again let $\{X(t)\}_{t\in[0,T]}$ satisfy the (BM) SDE

$$dX(t) = \mu_1(X(t),t)\,dt + \sigma(X(t),t)\,dB(t) \quad \text{for } t\in[0,T],$$

where σ is stritcly positive. Select another drift $\mu_2 : \mathbb{R} \times [0,T] \to \mathbb{R}$ and set

$$H(t) = \frac{\mu_2(X(t),t) - \mu_1(X(t),t)}{\sigma(X(t),t)} \quad \text{for } t \in [0,T].$$

Suppose that $\{\mathcal{E}(\int H\,dB)(t)\}_{t\in[0,T]}$ is a martingale and define a new probability measure Q by the so called likelihood ratio

$$\frac{dQ}{dP} = \exp\bigg(\int_0^T H\,dB - \frac{1}{2}\int_0^T H(t)^2\,dt\bigg) \qquad \text{use } dB(t) = \frac{dX(t) - \mu_1(X(t),t)dt}{\sigma(X(t),t)}$$

$$= \exp\bigg(\int_0^T \frac{\mu_2(X(t),t) - \mu_1(X(t),t)}{\sigma(X(t),t)^2}\,dX(t) - \frac{1}{2}\int_0^T \frac{\mu_2(X(t),t)^2 - \mu_1(X(t),t)^2}{\sigma(X(t),t)^2}\,dt\bigg).$$

Recall that

$$W(t) = B(t) - \int_0^t H(s)\,ds \quad \text{for } t \in [0,T]$$

is a Q-BM and that X satisfies the SDE

$$dX(t) = \mu_2(X(t),t)\,dt + \sigma(X(t),t)\,dW(t) \quad \text{for } t \in [0,T].$$

With the framework of the previous paragraph, assume that we have made an observation of $\{X(t)\}_{t\in[0,T]}$ and want to make statistical inference about the drift coefficient. In a hypotheses test of

$$H_0 : dX(t) = \mu_1(X(t),t)\,dt + \sigma(X(t),t)\,dB(t) \quad (= \text{the drift coefficient is } \mu_1)$$

against

$$H_1 : dX(t) = \mu_2(X(t),t)\,dt + \sigma(X(t),t)\,dW(t) \quad (= \text{the drift coefficient is } \mu_2)$$

we simply reject H_0 if dQ/dP is sufficiently much larger than 1. Similarly we can make estimation of the drift by taking $\mu_1 = 0$ and maximizing dQ/dP with respect to μ_2 where the $\mu_2 = \mu$ that gives the maximum is the estimated drift. (Of course, typically this requires a parametric choice of μ_2 to be carried out in practice.) These statistical procedures can be given a solid theoretical framework with appropriate central limit theorems, convergence rates and optimality properties etc. However the theory is rather complicated.

The diffusion coefficient $\sigma : \mathbb{R} \times [0,T] \to \mathbb{R}$ cannot be accessed by likelihood ratios. Here one can instead employ quadratic variation and simply make use of the fact that

$$d[X](t) = \sigma(X(t),t)^2\,dt \quad \text{for } t \in [0,T].$$

One fits a suitable σ to this equation with the observation $\{X(t)\}_{t\in[0,T]}$ inserted, for example, using least squares. (Again, typically this requires a parametric model for σ.)

Example. We have observed an Ornstein-Uhlenbeck process $\{X(t)\}_{t\in[0,T]}$ with SDE $dX(t) = -\alpha X(t)\,dt + \sigma\,dB(t)$ with unknown parameters $\alpha \in \mathbb{R}$ and $\sigma > 0$ and want to estimate these parameters.

$\boxed{\sigma}$ As we have $[X](t) = \sigma^2 t$ we estimate

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{n}\left(X(t_i) - X(t_{i-1})\right)^2, \quad 0 = t_0 < t_1 < \dots < t_n = T.$$

$\boxed{\alpha}$ Set $\mu_1(x,t) = 0$, $\mu_2(x,t) = -\alpha x$ and $\sigma(x,t) = \sigma$ in what we just did to obtain

$$\frac{dQ}{dP} = \exp\left(-\int_0^T \frac{\alpha X(t)}{\sigma^2}\,dX(t) - \frac{1}{2}\int_0^T \frac{\alpha^2 X(t)^2}{\sigma^2}\,dt\right)$$

for $dX(t) = \sigma\,d\hat{B}(t)$ and $dX(t) = -\alpha X(t)\,dt + \sigma\,dW(t)$

other BM's than OU $B(t)$

Maximize wrt. α by differentiating to get

$$\hat{\alpha} = -\int_0^T X(t)\,dX(t) \Big/ \int_0^T X(t)^2\,dt. \qquad \#$$

These calculations are done in applied lecture also.