Change of Measure
Article for EoAS

Tomasz Rolski
Mathematical Institute, Wrocław University
pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland

August 5, 2003

Change of measure is a technique used for deriving saddlepoint approximations, bounds or optimal methods of simulation. It is used in risk theory.

Theoretical backround; Radon-Nikodym theorem

Let Ω be a basic space and $\mathcal{F}$ a σ-field of subsets. We consider two probability measures $\mathbf{P}$ and $\tilde{\mathbf{P}}$ on $(\Omega, \mathcal{F})$. Expectations with respect to $\mathbf{P}$ and $\tilde{\mathbf{P}}$ are denoted by $\mathbf{E}$ and $\tilde{\mathbf{E}}$ respectively. We say that $\tilde{\mathbf{P}}$ is *absolute continuous* with respect to $\mathbf{P}$ if for all $A \in \mathcal{F}$ such that $\mathbf{P}(A) = 0$ we have $\tilde{\mathbf{P}}(A) = 0$. We write then $\tilde{\mathbf{P}} \ll \mathbf{P}$. If $\tilde{\mathbf{P}} \ll \mathbf{P}$ and $\mathbf{P} \ll \tilde{\mathbf{P}}$, then we say that $\tilde{\mathbf{P}}$ is *equivalent* to $\mathbf{P}$ and write $\tilde{\mathbf{P}} \equiv \mathbf{P}$. A useful here version of *Radon-Nikodym theorem* states that there exists a nonnegative random variable Z on $(\Omega, \mathcal{F})$ such that $\mathbf{E}Z = 1$ and

$$\mathbf{E}[Z; A] = \tilde{\mathbf{P}}(A) \tag{0.1}$$

for all $A \in \mathcal{F}$, where we denote $\mathbf{E}[Z; A] = \int_A Z\,dP$. Random variable Z is usually referred to as a *Radon-Nikodym derivative*, written $\mathrm{d}\tilde{\mathbf{P}}/\mathrm{d}\mathbf{P}$. The relation (0.1) is basic for the change of measure technique and can be stated equivalently as follows. For each random variable X such that $\tilde{\mathbf{E}}|X| < \infty$

$$\mathbf{E}[ZX] = \tilde{\mathbf{E}}[X]. \tag{0.2}$$

Conversely, if moreover $\mathbf{P} \equiv \tilde{\mathbf{P}}$, then

$$\tilde{\mathbf{E}}[Z^{-1}; A] = \mathbf{P}(A).$$

We now survey important cases and applications.

Importance sampling

In the theory of simulation one of the basic variance reduction technique is *importance sampling.* This technique is in fact a change of measure and use the basic identity (0.2). Let e.g. $\Omega = \mathbb{R}$ (although derivation can be easily adapted to $\Omega = \mathbb{R}^k$), $\mathcal{F} = \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ denotes the σ-field of Borel subsets, $X(\omega) = \omega$, and probability measures $\mathbf{P}$ and $\tilde{\mathbf{P}}$ have densities $g(x)$ and $\tilde{g}(x)$ respectively. We want to compute by simulation

$$\tilde{\mathbf{E}}\phi(X) = \int_{-\infty}^{\infty} \phi(x)\tilde{g}(x)\,dx.$$

An *at hoc* estimator is $(1/n)\sum_{j=1}^{n}\phi(X_j)$, where $X_1,\ldots,X_n$ are independent replications of X with density $\tilde{g}(x)$, but frequently this estimator is not good. Assuming that the set $\{x : g(x) = 0\}$ has $\mathbf{P}$-probability zero, we can write

$$\begin{aligned} \tilde{\mathbf{E}}\phi(X) &= \int_{-\infty}^{\infty} \phi(x)\tilde{g}(x)\,dx \\ &= \int_{-\infty}^{\infty} \phi(x)\frac{\tilde{g}(x)}{g(x)}g(x)\,dx \\ &= \mathbf{E}\phi(X)\frac{\tilde{g}(X)}{g(X)}. \end{aligned}$$

A suitable choice of g allows to improve properties of the importance sampling estimator $(1/n)\sum_{j=1}^{n}\frac{\phi(X)\tilde{g}(X_j)}{g(X_j)}$, where now $X_1,\ldots,X_n$ are independent replications of X_1 considered on the probability space $(\Omega,\mathcal{F},\mathbf{P})$. Notice that $\tilde{g}(x)/g(x)$ is called *likelihood ratio.* See more in [1, 4, 7].

Likelihood ration sequence

Consider now a sequence of random variables $X_1, X_2, \ldots$ on $(\Omega,\mathcal{F})$ under two different probability measures $\mathbf{P}$ and $\tilde{\mathbf{P}}$ and let $\mathcal{F}_n$ denote the σ-field generated by $X_1,\ldots,X_n$. That is $\{\mathcal{F}_n\}$ is a filtration; see FILTRATIONS*. Let $\mathbf{P}_{|n}$ and $\tilde{\mathbf{P}}_{|n}$ be the restrictions of $\mathbf{P}$ and $\tilde{\mathbf{P}}$ to $\mathcal{F}_n$ and assume that $\tilde{\mathbf{P}}_{|n} \ll \mathbf{P}_{|n}$ for all $n = 1,2,\ldots.$ Then $Z_n = \mathrm{d}\tilde{\mathbf{P}}_{|n}/\mathrm{d}\mathbf{P}_{|n}$, $n = 1,2,\ldots$ is an $\mathcal{F}_n$–martingale on $(\Omega,\mathcal{F},\mathbf{P})$ because for $n \le m$ and $A \in \mathcal{F}_n \subset \mathcal{F}_m$, we have by (0.1)

$$\mathbf{E}[Z_n; A] = \tilde{\mathbf{P}}_{|n}(A) = \tilde{\mathbf{P}}_{|m}(A) = \mathbf{E}[Z_m; A];$$

see MARTINGALES*. Note that $\tilde{\mathbf{P}} \ll \mathbf{P}$ if $Z_n \to Z$ a.s. and $\{Z_n\}$ is uniformly integrable, which holds very rarely and typically this is not true in most of applications.

The following converse construction is important for various applications. Suppose that $X_1, X_2, \ldots$ is a sequence of random variables on $(\Omega, \mathcal{F}, \mathbf{P})$ and $\{\mathcal{F}_n\}$ as before. Let Z_n be a positive $\mathcal{F}_n$-martingale such that $\mathbf{E}Z_1 = 1$. Under these assumptions, the relation $d\tilde{\mathbf{P}}_{|n} = Z_n \, d\mathbf{P}_{|n}$, which means that

$$\tilde{\mathbf{P}}_{|n}(A) = \mathbf{E}[Z_n; A], \qquad A \in \mathcal{F}_n, \qquad n = 1, 2, \ldots,$$

defines a consistent sequence of probability measures. We assume that $\mathcal{F} = \bigvee_{j=1}^{\infty} \mathcal{F}_n$ and that there exists the unique probability measure $\tilde{\mathbf{P}}$ on $\mathcal{F}$ such that $\tilde{\mathbf{P}}(A) = \tilde{\mathbf{P}}_{|n}(A)$ for all $A \in \mathcal{F}_n$. A standard version of Kolmogorov consistency theorem suffices if $\Omega = \mathbb{R} \times \mathbb{R} \times \ldots$. For other spaces one needs a more general version of the Kolmogorov theorem; see e.g. [9], Theorem 4.2 on page 143. Suppose furthermore that for all n we have $Z_n > 0$ $\tilde{\mathbf{P}}$-a.s. Then

$$\mathbf{P}_{|n}(A) = \tilde{\mathbf{E}}[Z_n^{-1}; A], \qquad A \in \mathcal{F}_n.$$

Moreover for each $\tilde{\mathbf{P}}$-finite stopping time τ the above can be extended to the basic identity:

$$\mathbf{P}(A) = \tilde{\mathbf{E}}[Z_\tau^{-1}; A], \qquad A \in \mathcal{F}_\tau. \tag{0.3}$$

As a special case, assume that $X_1, X_2, \ldots$ are independent and identically distributed random variables; under $\mathbf{P}$ or $\tilde{\mathbf{P}}$ random variable X_n has distribution F or $\tilde{F}$ respectively. Then $\tilde{\mathbf{P}}_{|n} \ll \mathbf{P}_{|n}$ if and only if $\tilde{F} \ll F$. If furthermore F and $\tilde{F}$ have densities $\tilde{g}(x)$ and $g(x)$ respectively, then

$$Z_n = \frac{\mathrm{d}\tilde{\mathbf{P}}_{|n}}{\mathrm{d}\mathbf{P}_{|n}} = \frac{\tilde{g}(X_1)}{g(X_1)} \frac{\tilde{g}(X_2)}{g(X_2)} \cdots \frac{\tilde{g}(X_n)}{g(X_n)}$$

which is called *likelihood ratio sequence*. See for details [2, 4].

Associated sequences

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with a common distribution F and let $\hat{m}(s) = \mathbf{E}\exp(sX)$ be the moment generating function of F. We assume that $\hat{m}(s)$ is finite in an interval $(\theta_l < 0 < \theta_r)$. For all $\theta_l < \theta < \theta_r$ we define the so called *associated distribution*

$$\tilde{F}_\theta(B) = \frac{\int_B e^{\theta x} \, F(dx)}{\hat{m}(\theta)}.$$

We have the following formulas for the moment generating function, mean and variance of $\tilde{F}_\theta$

$$
\begin{aligned}
\hat{m}_\theta(s) &= \tilde{\mathbf{E}}_\theta e^{sX} = \frac{\hat{m}(s+\theta)}{\hat{m}(\theta)}, \\
\mu_\theta &= \tilde{\mathbf{E}}_\theta X_1 = \frac{\hat{m}'(\theta)}{\hat{m}(\theta)}, \\
\sigma_\theta^2 &= \tilde{\mathbf{E}}_\theta X_1 = \frac{\hat{m}''(\theta)\hat{m}(\theta) - (\hat{m}'(\theta))^2}{(\hat{m}(\theta))^2}.
\end{aligned}
$$

Clearly $F \equiv \tilde{F}_0$ and

$$
\mathrm{d}\tilde{\mathbf{P}}_{|n}/\mathrm{d}\mathbf{P}_{|n} = L_{n;\theta} = \frac{e^{\theta(X_1+\ldots+X_n)}}{\hat{m}^n(\theta)}, \qquad n = 1, 2, \ldots.
$$

The basic identity can be expressed now as follows: if τ is a finite $\tilde{\mathbf{P}}_\theta$-a.s. stopping time and $A \in \mathcal{F}_\tau$, then

$$
\begin{aligned}
\mathbf{P}(A) &= \tilde{\mathbf{E}}_\theta[\frac{-e^{\theta(X_1+\ldots+X_\tau)}}{\hat{m}^\tau(\theta)}; A] \\
&= \tilde{\mathbf{E}}_\theta[e^{-\theta S_\tau + \tau\kappa^\tau(\theta)}; A], \qquad (0.4)
\end{aligned}
$$

where $S_n = X_1 + \ldots + X_n$ and $\kappa(\theta) = \log \hat{m}(\theta)$. More details on associated distributions can be found in [2, 11].

An example; proof of Chernov bound and efficient simulation

Under assumption from the last paragraph on associated sequences we derive now an inequality for $\mathbf{P}(A(n))$, where

$$
A(n) = \{S_n > (\mathbf{E}X_1 + \epsilon)n\}
$$

and $S_n = X_1 + \ldots + X_n$. The relevant choice of θ is fulfilling

$$
\mathbf{E}_\theta X_1 = \frac{\hat{m}'(\theta)}{\hat{m}'(\theta)} = \mathbf{E}X_1 + \epsilon.
$$

Then $I = \theta(\mathbf{E}X_1 + \epsilon) - \kappa(\theta) > 0$. Using the basic identity (0.4) with $\tau = n$ we have

$$
\begin{aligned}
\mathbf{P}(S_n > (\mathbf{E}X_1 + \epsilon)n) &= \tilde{\mathbf{E}}_\theta[\frac{e^{-\theta S_n}}{\hat{m}^n(\theta)}; S_n > (\mathbf{E}X_1 + \epsilon)n] \\
&= e^{-nI}\tilde{\mathbf{E}}_\theta[e^{-\theta(S_n - n(\mathbf{E}X_1+\epsilon))}; S_n > (\mathbf{E}X_1 + \epsilon)n] \\
&\leq e^{-nI}.
\end{aligned}
$$

According to the theory of simulation, the sequence of events $A(n)$ is rare because $\mathbf{P}(A(n) \to 0$. It turns out that the above change of measure is in some sense optimal for simulation of $\mathbf{P}(A(n))$. One has to sample $X_1, \ldots, X_n$ according to $\tilde{\mathbf{P}}_\theta$ and use the estimator $(1/n)\sum_{j=1}^n Y_j$, where $Y_j = L_{n;\theta}\mathbb{I}(S_n > n(\mathbf{E}X_1 + \epsilon))$. Details can be found in [1].

Continuous time stochastic processes; a general scheme

All considered processes are assumed to be cádlág, that is right continuous and with left limits realizations, and they are defined on $(\Omega, \mathcal{F}, \mathbf{P})$ equipped with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, where $\mathcal{F} = \bigvee_{t\geq 0}\mathcal{F}_t$ is the smallest σ-field generated by all subsets from $\mathcal{F}_t$ $(t \geq 0)$. For a probability measure $\mathbf{P}$ we denote the restriction to $\mathcal{F}_t$ by $\mathbf{P}_{|t}$. Suppose that $\{Z(t)\}$ is a positive martingale such that $\mathbf{E}Z(t) = 1$ and define

$$\tilde{\mathbf{P}}_{|t}(A) = \mathbf{E}[Z(t); A], \qquad A \in \mathcal{F}_{|t}$$

for all $t \geq 0$. Then $\tilde{\mathbf{P}}_{|t}$ is a probability measure and $\{\tilde{\mathbf{P}}_{|t},\ t \geq 0\}$ form a consistent family. We assume that the standard set-up is fulfilled, that is

(i) the filtered probability space $(\Omega, \mathbb{F}, \mathbf{P})$ has right-continuous filtration $\{\mathcal{F}_t\}$,

(ii) there exists the unique probability measure $\tilde{\mathbf{P}}$ such that $\tilde{\mathbf{P}}(A) = \tilde{\mathbf{P}}_{|t}(A)$ for $A \in \mathcal{F}_t$.

A useful model where the standard set-up holds is $\Omega = D[0, \infty)$, $X(t, \omega) = \omega(t)$, with filtration $\mathcal{F}_t = \sigma\{X(s),\ 0 \leq s \leq t\}$ $(t \geq 0)$, and $\mathcal{F} = \bigvee_{t\geq 0}\mathcal{F}_t$. Assuming moreover $Z(t) > 0$ $\tilde{\mathbf{P}}$-a.s. we also have

$$\mathbf{P}_{|t}(A) = \tilde{\mathbf{E}}[Z^{-1}(t); A], \qquad A \in \mathcal{F}_t$$

and furthermore, for a stopping time τ which is finite $\tilde{\mathbf{P}}$ a.s.

$$\mathbf{P}(A) = \tilde{\mathbf{E}}[Z^{-1}(\tau); A], \qquad A \in \mathcal{F}_\tau. \tag{0.5}$$

Assume from now on that $\{X(t)\}$ is a Markov process. For such the class of processes one defines a generator $\boldsymbol{A}$, which can be thought as an operator from a class of function to a class of function (on the state space). Define for a strictly positive function f

$$E^f(t) = \frac{f(X(t))}{f(X(0))}\exp\left(-\int_0^t \frac{(\boldsymbol{A}f)(X(s))}{f(X(s))}\,ds\right), \qquad t \geq 0\,. \tag{0.6}$$

If for some h the process $\{E^h(t)\}$ is a mean one martingale, then we can make the change of measure with $Z(t) = E^h(t)$. Note that $\mathbf{E}E^h(t) = 1$. Then from Kunita–Watanabe theorem we know, the process $\{X(t)\}$ on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \tilde{\mathbf{P}})$ is Markovian. It can be shown, with some care, that a generator $\tilde{\boldsymbol{A}}$ of this new stochastic process fulfills

$$\tilde{\boldsymbol{A}}f = h^{-1} \cdot [\boldsymbol{A}(f \cdot h) - f \cdot \boldsymbol{A}h]. \tag{0.7}$$

Details can be found in [8]

Continuous time Markov chains (CTMC) The simplest case we can analyze completely is when $\{X(t)\}$ is a CTMC with a finite state space. In this case the generator $\boldsymbol{A}$ is $\boldsymbol{Q} = (q_{ij})_{i,j=1,\ldots,\ell}$ the intensity matrix of the process, which acts on the space of all finite column vectors. Thus functions f, h are column vectors. and h is positive. We change the measure by E^h. Then $\{X(t)\}$ is a CTMC with new intensity matrix $\tilde{\boldsymbol{Q}} = (\tilde{q}_{ij})$ given by

$$\tilde{q}_{ij} = \begin{cases} q_{ij}\frac{h_j}{h_i} & i, \neq j \\ -\sum_{k=1}^{\ell} q_{ij}\frac{h_j}{h_i}, & i = j \end{cases} .$$

Poisson process Suppose that under $\mathbf{P}$ ($\tilde{\mathbf{P}}$) the stochastic process $\{X(t)\}$ is the Poisson process with intensity $\lambda > 0$ (with $\tilde{\lambda} > 0$). Then $\mathbf{P}_{|t} \equiv \tilde{\mathbf{P}}_{|t}$ for all $t > 0$ and

$$Z(t) = \frac{\mathrm{d}\tilde{\mathbf{P}}_{|t}}{\mathrm{d}\mathbf{P}_{|t}} = \left(\frac{\tilde{\lambda}}{\lambda}\right)^{X(t)-X(0)} e^{-(\tilde{\lambda}-\lambda)t}, \qquad t \geq 0.$$

Note that we can rewrite the above in form (0.6):

$$Z(t) = \frac{h(X(t)}{h(X(0))} e^{-\int_0^t \frac{(\boldsymbol{A}h)(X(t))}{h(X(t))}\, dt}$$

where $h(x) = (\tilde{\lambda}/\lambda)^x$ and formally for a function f we define $(\boldsymbol{A}f)(x) = \lambda(f(x+1) - f(x))$.

Claim surplus process in the classical Poisson model We consider claims arriving according to a Poisson process $\{N(t)\}$ with rate λ and independent and identically distributed claims $U_1, U_2, \ldots$ with a common distribution B. Let $\beta > 0$ be the premium rate. We assume that the Poisson process and claim sequence are independent. The claim surplus process in such the model is

$$S(t) = \sum_{j=1}^{N(t)} U_j - \beta t, \qquad t \geq 0.$$

A classical problem of risk theory is to compute the ruin probability $\mathbf{P}(\tau(u) < \infty)$, where the ruin time $\tau(u) = \min\{t : \ S(t) > u\}$. Notice that $\mathbf{P}(\tau(u) < \infty) < 1$ if $(\lambda \mathbf{E} U_1)/\beta < 1$. We show how the technique of change of measure can be used to analyze this problem. We need some preliminary results which are of interest on its own and are useful in other applications too.

The stochastic process

$$Z_\theta(t) = e^{\theta S(t) - t\kappa(\theta)}$$

is a mean-one martingale for θ such that $\mathbf{E}e^{\theta S(t)} < \infty$, where $\kappa(\theta) = \lambda(\hat{m}(\theta) - 1) - \beta\theta$ and $\hat{m}(\theta) = \mathbf{E}\exp(\theta U)$ is the moment generating function of U. Assume moreover that there exists $\gamma > 0$ such that $\kappa(\gamma) = 0$. For this is necessary that $(\lambda \mathbf{E} U_1)/\beta < 1$. Then the process $\{Z(t)\}$, where $Z(t) = \exp(\gamma S(t))$ is a positive mean one martingale. Moreover $Z(t) = E^h(t)$ for $h(x) = \exp(\gamma x)$. We may define the new measure $\tilde{\mathbf{P}}$ by $\{Z(t)\}$. It can be shown that $\{S(t)\}$ under $\tilde{\mathbf{P}}$ is a claim surplus process in the classical Poisson model, but with the claim arrival rate $\lambda \hat{m}(\gamma)$, the claim size distribution $\tilde{B}_\theta(dx) = \exp(\gamma x)\, B(dx)/\hat{m}(\gamma)$ and premium rate β. This means that the process $\{S(t)\}$ is a Levy process and to find its drift we compute, using Wald identity, that

$$\tilde{\mathbf{E}} S(1) = \int_0^\infty x \exp(\gamma x)\, B(dx) - \beta = \kappa^{'}(\gamma) > 0.$$

The strict positivity follows from the analysis of function $\kappa(s)$. Hence $S(t) \to \infty$ $\tilde{\mathbf{P}}$-a.s. and therefore $\tau(u)$ is finite $\tilde{\mathbf{P}}$-a.s. Now the ruin pobability can be written using (0.5) and recalling $\tilde{\mathbf{P}}(\tau(u) < \infty) = 1$, we have

$$\mathbf{P}(\tau(u) < \infty) = \tilde{\mathbf{E}}[e^{-\gamma S(\tau(u))}; \tau < \infty] = e^{-\gamma u} \tilde{\mathbf{E}}[e^{-\gamma(S(\tau(u)) - u)}].$$

Since $S(\tau(u)) - u \geq$, we obtain immediately a Lundberg inequality that

$$\mathbf{P}(\tau(u) < \infty) \leq e^{-\gamma u}, \qquad u \geq 0.$$

Cramer-Lundberg estimations follows from the observation that, under $\tilde{\mathbf{P}}$, the process $\{\xi(t)\}$ defined by $\xi(t) = S(\tau(t)) - t$ is the residual lifetime process in a renewal process and therefore

$$\tilde{\mathbf{E}} e^{-\xi(t)} \to \tilde{\mathbf{E}} e^{-\xi(\infty)}.$$

Further reading recommendation

A related material on the change of measure or Girsanov type theorems can be found for example in books Revuz and Yor (1991) and Kűchler and Sorensen (1997), Jacod and Shirayev (1987).

References

[1] Asmussen, S. (1998) *Stochastic simulation with a few towards stochastic processes.* Lecture Notes MaphySto no 2, Aarhus.

[2] Asmussen, S. (2000) *Ruin probabilities.* World Scientific, Singapore.

[3] Jacod, J. & Shirayev, A.N. (1987) *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin, Heidelberg.

[4] Glynn, P.W. & Iglehart, D.L. (1989) Importance sampling for stochastic simulation. *Management Sci.* **35**, 1367–1392.

[5] Kűchler, U. & Sorensen, M. (1997) *Exponential Families of Stochastic processes.* Springer-Verlag, New York.

[6] Jacobsen, M. (1982) *Statistical Analysis of Counting Processes.* Springer, New York.

[7] Madras, N. (2002) *Lectures on Monte Carlo Methods.* Fields Institute Monographs, AMS, Providence.

[8] Palmowski, Z. & Rolski, T. (2002) A technique for exponential change of measure for Markov processes. *Bernoulli* **8**, 767–785.

[9] Parthasarathy, K.R. (1967) *Probability measures on metric spaces*, Academic Press, New York.

[10] Revuz, D. & Yor, M. (1991) *Continuous Martingales and Brownian Motion*, Springer–Verlag, Berlin.

[11] Rolski, T., Schmidt, V., Schmidli, H. & Teugels, J. (1999) *Stochastic Processes for Insurance and Finance.* Wiley, Chichester.