VARIANCE REDUCTION TECHNIQUES

Simulation model: a process on $(\Omega, \mathcal{F}, \mathbb{P})$ with its natural filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$. The estimator of the desired quantity θ is in general a functional of a stochastic process.

In the simplest case of iid simulation runs, one is interested in $\theta = \mathbb{E}(X)$, and simulations are assumed to produce iid variables $\{X_1, X_2, ..\}$ with (perhaps) same distribution as X, or approximate distribution.

Let $X = h(w)$ for a given representation, $w \in \Omega$, and let $F(dx)$ be the probability distribution of X. We can write θ in two ways:

$$\theta = \mathbb{E}(X) = \int_\Omega h(w)\,\mathbb{P}(dw) = \int_S x F(dx).$$

REMARK: Because we are interested only in θ, <u>any</u> combination of $(h, \mathbb{P})$ that has the same value can be used. In particular, $\mathbb{E}(X+Y) = \theta \;\; \forall Y$ such that $\mathbb{E}(Y) = 0$.

$$\beta = \int_\Omega h(w)\,\mathbb{P}(dw) - \theta \text{ is the bias.}$$

$$MSE = \int_\Omega (h(w) - \theta)^2\,\mathbb{P}(dw) \quad (\text{if } \beta = 0 \Rightarrow MSE = Var)$$

To increase the efficiency we seek to REDUCE the variance:

- change $h(\cdot)$ (representation)
- change $\mathbb{P}(\cdot)$ (measure)

Classification of methods

- Correlation methods
  - antithetic random numbers,
  - control variables
- Partition methods
  - Latin hypercube and quasi-MC method
  - Stratification
- Changes of probability measure
  - Importance sampling
  - Conditional Monte Carlo

Proposition: ~~If~~ ~~//~~ ~~$\hat{\theta}_n$~~ Let $L(x)=x^2$. If $Var(\hat{\theta}_n) \le K < \infty$ for all n, then it follows from the Dominated Convergence Theorem that $\lim_{c\to\infty} R(c) = 0$.

Definition: If there ~~is a~~ are finite real numbers $r>0$ ~~such~~ and $\xi>0$ such that

$$\lim_{c\to\infty} c^r R(c) = \frac{1}{\xi}$$

then r is called the asymptotic efficiency rate and ξ is called the asymptotic efficiency.

Proposition: Let $\{\hat{\theta}(n), C(n)\}$ represent our simulation output and let $L(\cdot)$ be a loss function with $L'(0)=0$, $L''(0)>0$, and assume that $\exists\, \gamma, \sigma^2$:

$$n^{\gamma}(\hat{\theta}(n) - \theta) \Rightarrow N(0,\sigma^2).$$

Also, assume that $\exists\, \beta>0$: $n^{-\beta} C(n) \to \lambda$, where λ is the cost rate per replication (CPU time) in the long run.

Then: $\quad r = \frac{2\gamma}{\beta}, \quad \xi = \frac{2}{L''(0)\, \lambda^{2\gamma} \sigma^2}$

Example: Simulations where the usual CLT is valid, $\gamma = \frac{1}{2}$ because usually

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i) \text{ ~~has~~ is a sample avg.}$$

and

$$\sqrt{n}(\hat{\theta}_n - \theta)^2 \Rightarrow N(0,\sigma^2).$$

In this case, if simulations require the same amount of time λ then $\beta = 1$ and we have, for the MSE function $L(x) = x^2$ that:

$$r = 1, \qquad \xi = \frac{1}{\lambda\sigma^2}.$$

Remark: Suppose that there are two estimators $Y_1^{(n)}$ and $Y_2^{(n)}$ of a parameter θ, via different simulations.

The "error" terms are given by:

$$(Y_1(n) - \theta)^2 \approx n^{-r_1} N(0,\sigma_1^2)$$

$$(Y_2(n) - \theta)^2 \approx n^{-r_2} N(0,\sigma_2^2)$$

For same error, which one requires more CPU time?

- rate that is larger is preferable for long sims.
- if $r_1 = r_2 \Rightarrow$ smaller variance is prefrable

SIMULATION EFFICIENCY AND VARIANCE REDUCTION

(See SimSpiders)

- Confidence intervals for estimators $\hat{\theta}$ of a parameter θ, assuming $\theta = \mathbb{E}(\phi(X_t ; t \leq T))$ for X_t a stochastic process on $(\Omega, \mathfrak{F}, \mathbb{P})$.
- ~~Love~~ Approximate confidence intervals are based on approximations to the distribution of the estimator $\hat{\theta}$.
- Central Limit Theorems CLT provide useful approximations,

use information when possible instead of approximations

example: Bernoulli $(p) \Rightarrow Var(x) = p(1-p)$ so a useful estimate of variance is $\hat{p}(1-\hat{p})$ instead of usual sample variance.

Definition: The sample variance is

$$\hat{\sigma}_n = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

~~Prop~~

CLT's theorems help to ~~as~~ estimate approximate limit distributions of the ERROR terms in the form: $n^{\beta}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$

Precision against Speed

X_t a s.p. on $(\Omega, \{\mathfrak{F}_t\}, \mathbb{P})$

Suppose that we wish to estimate

$$\theta = \mathbb{E}(\phi(X_t ; t \leq T)) \qquad (T \text{ can be } +\infty)$$

and we use n replications of a simulation (or steps in a long-run simulation), to obtain $\hat{\theta}_n$.

We assume that:

$$\hat{\theta}_n \xrightarrow[n\to\infty]{} \mathbb{E}(\phi(X_t ; t \leq T)) \quad \text{w.p.1}$$

$$\lim_{n\to\infty} \hat{\theta}_n = \theta \quad \text{w.p.1}$$

and in some cases we will also have that

$$\lim_{n\to\infty} \mathbb{E}\hat{\theta}_n = \theta.$$

$C(n)$: ~~...~~ CPU time or ~~"cost"~~ "cummulative cost" of the n replications.

$L(x)$: loss function $L: \mathbb{R} \to \mathbb{R}^+$ a convex function with $L(0)=0$. Normally we use $L(x) = x^2$

Definition: Given a ~~...~~ simulation budget c, ~~the~~ $T(c) = \min(n \geq 0 : C(n) \geq c)$ is the simulation time, and

$R(c) = \mathbb{E}(L(\hat{\theta}(T(c)) - \theta))$ is the expected loss.

Antithetic Variables

Let X_1, X_2 have the same distribution with $\mathbb{E}(X_i) = \theta$.

Then $\mathbb{E}\left(\frac{X_1+X_2}{2}\right) = \theta$ and the mean can be used as an estimator. Then:

$$\mathrm{Var}\left(\frac{X_1+X_2}{2}\right) = \frac{1}{4}\left(\sigma^2+\sigma^2+2\underbrace{\sigma_{X_1X_2}}_{\mathrm{Cov}(X_1,X_2)}\right) = \frac{1}{2}\left(\sigma^2+\mathrm{Cov}(X_1,X_2)\right)$$

$\Rightarrow$ seek negative correlation, so this estimator has smaller variance than the naive estimator.

Theorem: Let $(X_1, \dots X_n)$ be independent rv's on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $f, g : \mathbb{R}^n \to \mathbb{R}^+$ be non-negative functions and monotonically increasing in each component. Then:

$$\mathbb{E}\left(f(X_1,\dots X_n)\, g(X_1,\dots X_n)\right) \geq \mathbb{E}\left(f(X_1,\dots X_n)\right)\cdot\mathbb{E}\left(g(X_1,\dots X_n)\right).$$

Corollary: Let $h : [0,1]^n \to \mathbb{R}$ be monotone in each ~~component~~ argument. Then if $(U_1, \dots U_n) \sim$ iid $U(0,1)$:

$$\mathrm{Cov}\left\{h(U_1,\dots,U_n),\, h(1-U_1,\dots,1-U_n)\right\} \leq 0$$

Idea of proof: $n=2$, assume wlog $h(x_1,x_2)$ $(\uparrow,\downarrow)$ & let now $f(u_1,u_2) = h(U_1, 1-U_2)$ and $g(U_1,U_2) - h(1-U_1,U_2)$, then applying the Theorem $\Rightarrow$

$$\mathbb{E}\left(h(U_1,1-U_2)\, h(1-U_1,U_2)\right) \leq \mathbb{E}\left(h(U_1,1-U_2)\right)\mathbb{E}\left(h(U_1,U_2)\right)$$

$$\mathbb{E}\left(h(u,v)\, h(1-u,1-v)\right) - \mathbb{E}\, h(u,v)\cdot\mathbb{E}\left(h(1-u,1-v)\right) \leq 0.$$

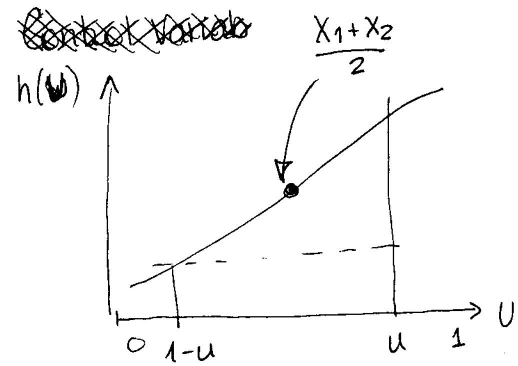
Examples: acetats (use the option pricing). [Ross, CH8].

Result: Observe that because any distribution function is monotonically non-decreasing, then

$$h(U) = F^{-1}(U) = X(U)$$

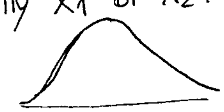
will always be monotone in U. This is an argument for the use of inverse method.

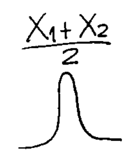
[extension: common random variables for comparison].



Idea: by averaging "extreme variabilities" will compensate.

only X_1 or X_2:





extreme case if h is linear: $h(U) = \alpha u \Rightarrow$

$$\frac{\alpha(u)+\alpha(1-u)}{2} = \frac{\alpha}{2} = \mathbb{E}h(U) \text{ and } \mathrm{var}=0.$$

Examples in Option Pricing:

(1). Evaluation of Asian options. $\{S(t)\,;\,t\le T\}$ GBM: $\{B(t);t\le T\}$ is the non-risky asset $B(t)=e^{rt}$, r: interest rate.

European option payoff: $(S(T)-K)_+$, K: strike price.

Asian option payoff: (Bermuda)

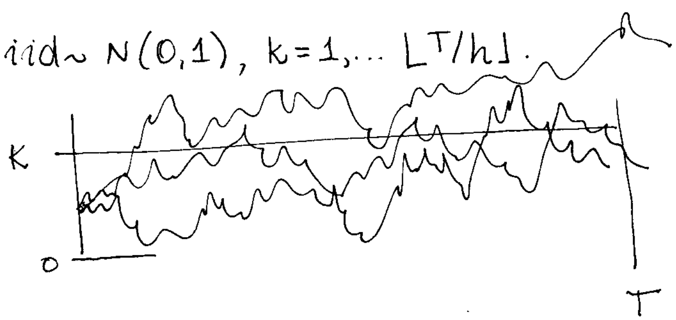
$$C(T)=e^{-rT}\Big(\frac{1}{R}\sum_{k=1}^{R}S(t_k)-K\Big)_+$$

(t_k are the "dividend" times, or an approximation $t_k=hk$.)

The option value is $\mathbb{E}_Q\,C(T)$ where Q is the risk-neutral measure, under which:

$$S(t_{k+1}) = S(t_k)\exp\Big\{\Big(r+\frac{\sigma^2}{2}\Big)h+\sigma\sqrt{h}\,Z_{k+1}\Big\}$$

$\{Z_k\}$ iid $\sim N(0,1)$, $k=1,\dots\lfloor T/h\rfloor$.



The function $h(z)=e^z$ is monotonically increasing, and if $Z\sim N(0,1)$, then $-Z\sim N(0,1)$ also. Negative correlation is guaranteed: $\{Z_1^{(i)},\dots Z_R^{(k)}\}$ iid $\sim N(0,1)$ for nominal sim, and use $\tilde Z_m^{(i)}\equiv -Z_m^{(i)}$ for antithetic sim:

$$C_1^{(i)}=e^{-rT}\Big(\frac{1}{R}\sum_{k=1}^{R}S^{(i)}(t_k)-K\Big)_+ \quad (\text{use } Z)$$

$$C_2^{(i)}=e^{-rT}\Big(\frac{1}{R}\sum_{k=1}^{R}\tilde S^{(i)}(t_k)-K\Big)_+ \quad (\text{use } \tilde Z=-Z)$$

	$\sigma=0.2$			$\sigma=0.4$			$\sigma=0.6$		
$K/S_0=$	.9	1	1.1	0.9	1.0	1.1	0.9	1.0	1.1
Ind	0.053	0.344	0.566	0.308	0.694	1.017	0.632	1.052	1.443
~~AA~~	0.050	0.231	0.068	0.297	0.506	0.381	0.583	0.817	0.759

(2) Use of control variable:

Geometric mean: $G(T)=\Big\{\prod_{k=1}^{R}S(t_k)\Big\}^{1/R}$

Result: $G(T)\sim$ lognormal with mean and variance

$$m_G=\ln S_0+\Big(r+\frac{\sigma^2}{2}\Big)\Big(h\frac{(R+1)}{2}\Big)$$

$$\sigma_G^2=\frac{\sigma^2 h(R+1)(2R+1)}{6R}$$

Use as a control variable $y=e^{-rT}(G(T)-K)_+$:

for each replication $(Z_1^{(i)},\dots Z_R^{(i)})$ calculate both $X_i=C_1^{(i)}(T)$ and y_i, and also evaluate $\mathrm{Cov}(X,Y)$.

It turns out that even using $c^*=1$ works wonderfully well

$$X_i^c=X_i+(\mathbb{E}Y-Y_i)$$

Control:

0.003	0.004	0.006	0.014	0.017	0.021	0.032	0.038	0.047

up to 100 variance reduction!

Control Variables

$$\theta = \mathbb{E}X = \int_{\Omega} h(\omega)\,\mathbb{P}(d\omega)$$

Let $Y = g(\omega)$ be a rv on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}(Y) = \mu$ is known. Then $\mathbb{E}(X^c) = \theta$; for any $c \in \mathbb{R}$, where:

$$X^c(\omega) = h(\omega) + c\,(g(\omega) - \mu).$$

Goal: find $c \in \mathbb{R}$ that minimizes the variance:

$$\mathrm{Var}(X^c) = \sigma_X^2 + c^2 \sigma_Y^2 + 2c\,\mathrm{Cov}(X,Y) \text{ and}$$

$$\boxed{c^* = \frac{\mathrm{Cov}(X,Y)}{\sigma_Y^2}}$$ always minimizes $\mathrm{Var}(X^c)$ regardless of the correlation

and $$\mathrm{Var}(X^{c^*}) = \sigma_X^2\left(1 - \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y}\right) \text{ for } c^*$$

Discuss idea of "correction": if Y is positively correlated and we observe $Y >> \mathbb{E}(Y)$ we also "correct" the observation X to bring it "closer" to θ.

Problem: c^* is not known, even if we know σ_Y^2, because σ_{XY} is not known.

Solutions: - if sign(σ_{XY}) is known $\Rightarrow$ attempt reduction with correct sign, even if not optimal, or:

- estimate optimal c^* (pilot or concurrent)

$$\hat{\sigma}_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \hat{X}_N)(Y_i - \hat{Y}_N)$$

$$\hat{\sigma}_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \hat{Y}_N)^2$$

$$\boxed{\hat{c}^* = -\frac{\hat{\sigma}_{XY}}{\hat{\sigma}_Y^2}}$$

Examples from Ross and option-pricing.

Remark: $\hat{c}^*$ is correlated with $\{(X_i, Y_i)\}$ in concurrent estimation $\Rightarrow \mathbb{E}(X^{\hat{c}^*}) \neq \theta$ (bias), so one can use "jack-knife" method (acetats).

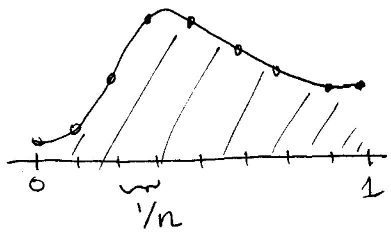
Endogenous variables: variables generated during the nominal simulation (little or no extra computing effort).

Exogenous variables: similar processes that are "adapted" or synchronised, example: use a (known) M/M/1 process to simulate in parallel with an (unknown) G/GI/1 process and consider the M/M/1 for the control. Requires good SYNCHRONIZATION (of random number seeds).

Weighted Means: If we have 2 different unbiased estimators then $Y = X_1 - X_2$ can be used: $X^c = X_1 + c(X_1 - X_2)$.

Latin Hypercube

$$\mathbb{E}(h(U)) = \int_0^1 h(u)\,du$$



$$\approx \frac{1}{n}\sum_{i=1}^{n} h(1/n)$$

approximation

Notice that MC method is $\frac{1}{n}\sum_{i=1}^{n} h(X_i)$, $\{X_i\}$ iid $\sim U(0,1)$.

To reduce variance, use instead of $X_i \sim U(0,1)$,

$V_i \sim U\left[\frac{i-1}{n}, \frac{i}{n}\right]$ so exactly one point per interval.

But $V_i \overset{d}{\neq} U(0,1)$.

Let π be a permutation of the indices $\{1,\dots n\}$, chosen with uniform distribution:

$$\mathbb{P}(\pi(i) = j) = \frac{1}{n} \quad \forall j = 1,\dots n$$

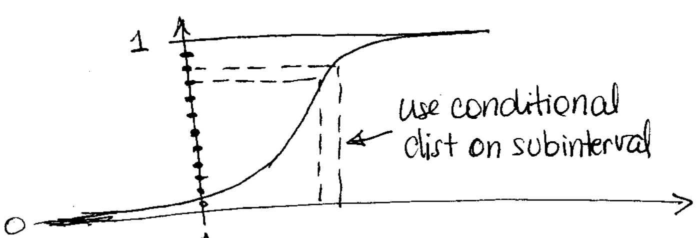
and let $$U_i = V_{\pi(i)} \sim U\left[\frac{\pi(i)-1}{n}, \frac{\pi(i)}{n}\right],$$

then:

(a). $\mathbb{P}(U \le x) = x$ (marginal is $U(0,1)$)

(b). $\exists!\, k : U_k \in \left[\frac{j-1}{n}, \frac{j}{n}\right]$ for each $j \in \{1,..n\}$

(c). $\forall\, x_1, x_2 \ \ \mathbb{P}(U_1 \le x_1, U_2 \le x_2) \le \mathbb{P}(U_1 \le x_1)\mathbb{P}(U_2 \le x_2)$



n points equally spaced

Inverse method: general dist

$X_i = F^{-1}(U_i)$

[see examples option pricing] [QMC methods]

Stratification

IDEA: Partition the sample space into events ("strata"), where the variability is known to be smaller. Then use conditioning

Remember Evve? ....

$$Var(X) = \underbrace{\mathbb{E}(Var(X|Y))}_{\text{stratification}} + \underbrace{Var(\mathbb{E}(X|Y))}_{\text{conditional MC}}$$

Let $Y \in \{1,\dots,k\}$ label the different strata, so that $\{\{\omega : Y(\omega) = i\}\ i = 1,..k\}$ is a partition of Ω, and consider the estimation of:

$$\theta = \mathbb{E}(X) = \sum_{i=1}^{k} \mathbb{E}(X|Y=i)\,p_i$$

where $p_i = \mathbb{P}(Y=i)$ is assumed to be <u>known</u>.

The idea is to fix each "scenario" $Y=i$ and perform a simulation experiment for EACH stratum at a time. Because $Var(X|Y_i) \le Var(X)$, we can expect a reduction in variance.

Let n_i = # replications @ $Y=i$, and

under iid model

$$\hat{\theta}_i = \frac{1}{n_i}\sum_{k=1}^{n_i} X_k\big|_{Y=i}, \quad Var(\hat{\theta}_i) = \frac{1}{n_i} Var(X|Y=i)$$

where $X|_{Y=i}$ is a random variable restricted (conditional) to $\{Y=i\}$. Then use

$$\hat{\theta}_n = \sum_{i=1}^{k} \hat{\theta}_i \cdot p_i$$ as estimator.

Because $Var(\hat{\theta}_n) = \sum_{i=1}^{K} p_i^2 \frac{Var(X|Y=i)}{n_i}$, we can find the optimal values of n as an allocation problem:

$$\underset{(n_1,..n_K)}{\text{Min}} \quad Var(\hat{\theta}_n)$$

$$\text{s.t.} \quad n_1+..+n_K = n$$

when $\sigma_i^2 \equiv Var(X|Y=i)$ are known, the solution is:

$$n_i^* = n \frac{p_i \sigma_i}{\sum_i p_i \sigma_i}.$$

Normally σ_i^2 are not known. They can be estimated, but then the optimal allocation must be done sequentially (or use a pilot simulation).

Example: good & bad days for a queueing system (hospital, bank, etc). $\lambda = 12$ or $\lambda = 4$. We know $p(Y=1,2) = 1/2$.
Simulate the system using the two scenarios:

$$X|_{y=i} \sim \text{Poisson}(\lambda_s(1-e^{-10})) \Rightarrow$$

$$Var(X|Y=1) \simeq 12, \; Var(X|Y=2) \approx 4,$$

$$\theta \simeq 8 \text{ and } Var(X) \approx 24.$$

$n_1^* = 0.634n \Rightarrow$ do 63.4% of simulations for $\lambda = 12$ and the rest for $\lambda = 4$.

Importance Sampling

Example: Let X have density $f(x) = 2x$, $0 < x < 1$ and $\theta = \mathbb{E}(h(X)) = \int_0^1 x^4 f(x)\,dx = \frac{1}{3}$.

Instead of generating $X \sim F$, consider a change of measure (Radon-Nikodym derivatives) as follows:

NEW DENSITY $g(x) = 6x^5$, $0 < x < 1$

$$L(x) = \frac{f(x)}{g(x)},$$

so that:

$$\mathbb{E}_g[h(X)L(X)] = \int_0^1 x^4 L(x) g(x)\,dx = \theta$$

and we only need to weigh the observations $h(x)$ by their "Likelihood Ratio": Here, the estimator is:

$$h(X)\frac{f(X)}{g(X)} = \frac{X^4(2X)}{6X^5} \equiv \frac{1}{3}$$

and it has zero variance $Var(h(X)L(X)) = 0$!!!

~/~

Assume that $\mathbb{P}, \mathbb{Q}$ are measures on $(\Omega, \mathcal{F})$ with $\mathbb{P} << \mathbb{Q}$, and call $L(\mathbb{P},\mathbb{Q},\omega) = \left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)(\omega)$ the R-N derivative, then for any integrable function h:

$$\theta = \mathbb{E}_{\mathbb{Q}}(Lh) = \int_\Omega h(\omega) L(\mathbb{P},\mathbb{Q},\omega)\,d\mathbb{Q}(\omega) = \mathbb{E}_{\mathbb{P}}(h).$$

If we use an estimator (IS) Lh under $\mathbb{Q}$, when is this going to have reduced variance?

Proposition: If $L(\mathbb{P},\mathbb{Q},\omega) \le 1$ $\forall \omega$ such that $h(\omega) \neq 0$, then $\mathrm{Var}_{\mathbb{Q}}(Lh) \le \mathrm{Var}_{\mathbb{P}}(h)$.

Recall that $\mathbb{E}_{\mathbb{Q}}(L) = 1$, so the above proposition means that a good change of measure must be tailored to the performance function that we wish to estimate. Interpretation: "region of importance" $\{\omega : h(\omega) \neq 0\} \subset \Omega$.

In general:

$$\mathrm{Var}_{\mathbb{Q}}(hL) = \mathbb{E}_{\mathbb{Q}}(h^2L^2) - \theta^2$$

$$= \int_\Omega h^2(\omega)\frac{d\mathbb{P}(\omega)}{d\mathbb{Q}} \cdot d\mathbb{P}(\omega) - \theta^2,$$

and the "optimal" change of measure is to use the conditional probability $\mathbb{P}$ on the "importance" set $\{\omega : h(\omega) \neq 0\}$. Clearly, this is usually not known when θ is not known, so it cannot help for simulations. However, it provides insight.

Let the measure $\mathbb{Q}$ be defined by:

$$\forall A \in \mathcal{F} \qquad \mathbb{Q}(A) = \frac{1}{\theta^*}\int_A |h(\omega)|\,d\mathbb{P}(\omega)$$

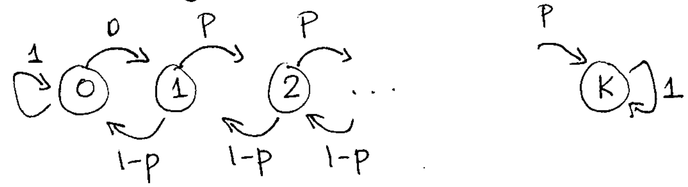
$$\theta^* = \int |h(\omega)|\,d\mathbb{P}(\omega),$$

then $L(\mathbb{P},\mathbb{Q},\omega) = \dfrac{\theta^*}{|h(\omega)|}$ and:

$$\mathrm{Var}_{\mathbb{Q}}[hL] = \int |h(\omega)|\,\theta^*\,d\mathbb{P}(\omega) = (\theta^*)^2 - \theta^2.$$

If h is non-negative, $\mathrm{Var}_{\mathbb{Q}}[hL] = 0$, but of course $\theta^* = \theta$ ... (not known!)

Example: gambler's ruin problem. (acetats)



Absorbing boundaries @ 0 or K

(Random walk, birth & death processes, etc).

$$\tau = \inf(n : X_n = 0 \text{ or } X_n = K)$$

We wish to estimate $\theta = \mathbb{E}(h(\omega))$, with

$$h(\omega) = \mathbb{1}(X_\tau(\omega) = K),$$

so that $h(\omega) = 0$ when $X_\tau = 0$. Here,

$$P_{i,i+1} = p \qquad 0 < i \le K-1$$

$$P_{i,i-1} = q = 1-p \quad 0 < i \le K-1$$

Replace the measure by considering a different transition matrix: "swapping" the parameters:

$$\tilde{P}_{i,i+1} = q \quad ; 0 < i < K-1$$
$$= 1 - \tilde{P}_{i,i-1}$$

with same absorbing states. We assume that $p << q$, so that very long simulations would be required to observe a significant number of winning events. Under $\mathbb{Q}$, however, winning is very likely.

Consider any trajectory ω : $X_\tau(\omega) = K \Rightarrow$ here, there are always K steps more to the right than to the left:

$$\tau \begin{cases} \frac{\tau-K}{2} \text{ left}: (q \text{ or } p) \\ \frac{\tau+K}{2}\text{-right}: (p \text{ or } q) \end{cases}$$

$$L(\mathbb{P},\mathbb{Q},\omega) = \left(\frac{q}{p}\right)^{\frac{z-K}{2}} \left(\frac{1-q}{1-p}\right)^{\frac{z+K}{2}-1} = \left(\frac{p}{1-p}\right)^{K-1}$$

if we choose $q=1-p$ for the change of measure

Because $h(\omega)=0$ when $X_z \neq K$, then

$$Var_{\mathbb{Q}}(hL) = \left(\frac{p}{1-p}\right)^{K-1} \mathbb{E}_p(h^2) - \theta^2 \leq Var_{\mathbb{P}}(h)$$

provided that $p<1/2$. It is more advantageous when K is large and $p \approx 0$.

Rare Event Estimation.