RMSC5102 Simulation Techniques in Risk Management and Finance

Tutorial Notes

Spring, 2021

LEUNG Man Fung, Heman

Contents

I) Probability and statistics

Discrete random variables

Random variables: numeric quantities that take different values with specified probabilities

Discrete random variable: a r.v. that takes value from a discrete set of numbers

Probability mass function: a pmf assigns a probability to each possible value x of the discrete random variable X, denoted by $f(x) = \mathbb{P}(X = x)$

$\sum_{i=1}^{n} f(x_i) = 1$ (total probability rule)

Cumulative distribution function: a cdf gives the probability that X is less than or equal to the value x, denoted by $F(x) = \mathbb{P}(X \leq x)$

Expected value: $\mu = \mathbb{E}(X) = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i)$ (the idea is "probability weighted average")

Variance: $\sigma^2 = \mathrm{Var}(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \mathbb{P}(X = x_i)$ (the idea is "probability weighted distance from mean")

Alternatively, $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$

Translation/rescale: $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$, $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$

Linearity of expectation: $\mathbb{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbb{E}(X_i)$

Law of the unconscious statistician: $\mathbb{E}[g(X)] = \sum_{i=1}^{n} g(x_i) \mathbb{P}(X = x_i)$

Moment generating function: $M_X(t) = \mathbb{E}(e^{tX}) = \sum_{i=1}^{n} e^{tx_i} \mathbb{P}(X = x_i)$

$\mathbb{E}(X^n) = \frac{d^n M_X(t)}{dt^n}|_{t=0}$

Binomial distribution

Factorial: $n! = n \times (n-1) \times \ldots \times 1$, note that $0! = 1$

Permutation (order is important): $\mathrm{P}_k^n = \frac{n!}{(n-k)!}$

Combination (order is not important): $\mathrm{C}_k^n = \frac{n!}{k!(n-k)!}$, also denoted as $\binom{n}{k}$

Binomial distribution: probability distribution on the number of successes X in n independent experiments, each experiment has a probability of success p; denoted by $X \sim \mathrm{B}(n, p)$

If $X_1, \ldots X_n \sim \mathrm{B}(1, p)$, then $\sum_{i=1}^{n} X_i \sim \mathrm{B}(n, p)$ (sum of i.i.d. Bernoulli r.v.s is a binomial)

Pmf: $\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, 2, \dots, n$

Mgf: $M_X(t) = (1 - p + pe^t)^n$

Mean: $\mathbb{E}(X) = np$

Variance: $\mathrm{Var}(X) = np(1-p)$

Poisson distribution

Poisson distribution: probability distribution on the number of occurrence X (usually of a rare event) over a period of time or space with rate λ; denoted by $X \sim \mathrm{Po}(\lambda)$. Useful in modelling jump

Pmf: $\mathbb{P}(X = x) = \frac{e^{-\lambda}\mu^x}{x!}$ for $x = 0, 1, 2, \dots$

Mgf: $M_X(t) = \exp[\lambda(e^t - 1)]$

Mean: $\mathbb{E}(X) = \lambda$

Variance: $\mathrm{Var}(X) = \lambda$

Geometric distribution

Geometric distribution: probability distribution on the number of Bernoulli trials X needed to get 1 success, each trial has a probability of success p; denoted by $X \sim \mathrm{Geo}(p)$

Pmf: $\mathbb{P}(X = x) = (1-p)^{x-1}p$ for $x = 1, 2, \dots$

Mgf: $M_X(t) = \frac{pe^t}{1-(1-p)e^t}$ for $t < -\ln(1-p)$

Mean: $\mathbb{E}(X) = \frac{1}{p}$

Variance: $\mathrm{Var}(X) = \frac{1-p}{p^2}$

Continuous random variables

Continuous random variable: a r.v. that takes value over an interval of numbers

Probability density function: a pdf specifies the probability of the random variable falling within a particular range of values, denoted by $f(x)$

$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$, which is the area under the curve from a to b

$\mathbb{P}(X = a) = \int_a^a f(x)dx = 0$ for all a

$\int_{-\infty}^{\infty} f(x)dx = 1$ (total probability rule)

Cumulative distribution function: a cdf gives the probability that X is less than or equal to the value x, denoted by $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)dt$

$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$ (by the fundamental theorem of calculus)

Expected value: $\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$ (the idea is "probability weighted average")

Variance: $\sigma^2 = \mathrm{Var}(X) = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$ (the idea is "probability weighted distance from mean")

Alternatively, $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$

Translation/rescale: $\mathbb{E}(aX+b) = a\mathbb{E}(X) + b$, $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$

Linearity of expectation: $\mathbb{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbb{E}(X_i)$

Law of the unconscious statistician: $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$

Moment generating function: $M_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}f(x)dx$

Uniform distribution

Uniform distribution: if X follows uniform distribution on the interval $[a, b]$, then it has the same probability density at any point in the interval and we denote it by $X \sim \mathrm{U}(a, b)$. Basic R.V. in inverse transform

Pdf: $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$, otherwise 0

Cdf: $F(x) = \int_a^x \frac{1}{b-a}dt = \left[\frac{t}{b-a}\right]_a^x = \frac{x-a}{b-a}$ for $a \leq x \leq b$

Mgf: $M_X(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}$ for $t \neq 0$

Mean: $\mathbb{E}(X) = \frac{a+b}{2}$

Variance: $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$

Normal distribution

Normal distribution: if X follows normal distribution with mean μ and variance σ^2, then $X \sim \mathrm{N}(\mu, \sigma^2)$. Often used to represent continuous random variable with unknown distributions

Pdf: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ for $-\infty < x < \infty$

Mgf: $M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$

Standard normal distribution: $Z \sim \mathrm{N}(0,1)$

Cdf of standard normal: denoted as $\Phi(z) = \mathbb{P}(Z \le z)$

$\mathbb{P}(a \le Z \le b) = \mathbb{P}(Z \le b) - \mathbb{P}(Z \le a) = \Phi(b) - \Phi(a)$

$\Phi(-z) = 1 - \Phi(z)$ by symmetric property

Percentile of standard normal: $\Phi(1.645) = 0.95, \Phi(1.96) = 0.975$

Standardization: if $X \sim \mathrm{N}(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$

$$\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Exponential distribution

Exponential distribution: if X follows exponential distribution with rate λ, we denote it by $X \sim \mathrm{Exp}(\lambda)$. Continuous analogue of the geometric distribution

Pdf: $f(x) = \lambda e^{-\lambda x}$ for $x \ge 0$, otherwise 0

Cdf: $F(x) = 1 - e^{-\lambda x}$ for $x \ge 0$

Mgf: $M_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$

Mean: $\mathbb{E}(X) = \frac{1}{\lambda}$

Variance: $\mathrm{Var}(X) = \frac{1}{\lambda^2}$

Memoryless property: $\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t)$ for $s, t \ge 0$

Exponential distribution is the only continuous distribution that has this property

Useful representation: if $Y \sim \mathrm{Exp}(1)$, then $\frac{Y}{\lambda} \sim \mathrm{Exp}(\lambda)$

Some remarks

Covariance: $\mathrm{Cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)$

Variance of sum: $\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)$

Tower rule of expectation: $\mathbb{E}(X)=\mathbb{E}[\mathbb{E}(X|Y)]$

Law of total variance (EVE): $\mathrm{Var}(X)=\mathbb{E}[Var(X|Y)]+\mathrm{Var}[\mathbb{E}(X|Y)]$

Sum of poisson: if $X\sim\mathrm{Po}(\lambda_1), Y\sim\mathrm{Po}(\lambda_2)$ independently, then $X+Y\sim\mathrm{Po}(\lambda_1+\lambda_2)$

Sum of normal: if $X\sim\mathrm{N}(\mu_1,\sigma_1^2), Y\sim\mathrm{N}(\mu_2,\sigma_2^2)$ independently, then

$$X+Y\sim\mathrm{N}(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$$

Square of standard normal: if $X\sim\mathrm{N}(\mu,\sigma^2)$, the $Z^2=\left[\frac{X-\mu}{\sigma}\right]^2\sim\chi_1^2$

Sum of chi square: if $X\sim\chi_n^2, Y\sim\chi_m^2$, then $X+Y\sim\chi_{n+m}^2$

II) Financial derivative

Forward

Payoff: $\mathrm{S_t} - K$ (long), $K - S_t$ (short)

Pricing: $f = S - Ke^{-r(T-t)}, F = Se^{r(T-t)}$

With known cash income: $f = S - I - Ke^{-r(T-t)}, F = (S - I)e^{r(T-t)}, I = \mathrm{PV(income)}$

With known dividend yield: $f = Se^{-q(T-t)} - Ke^{-r(T-t)}, F = Se^{(r-q)(T-t)}$

Minimum variance hedge ratio: $h^* = \rho \times \frac{\sigma_S}{\sigma_F} \Rightarrow N_F^* = h^* \times N_A$ (since $h = \frac{N_F}{N_S}$)

Option

Upper bounds: $C_E \leq C_A \leq S, P_E \leq Ke^{-r(T-t)}, P_A \leq K$

Lower bounds: $\max\left(S - Ke^{-r(T-t)}, 0\right) \leq C_E \leq C_A, \max\left(Ke^{-r(T-t)} - S, 0\right) \leq P_E \leq P_A$

Put-call parity: $C_E - P_E = S - I - Ke^{-r(T-t)}$ (idea is call – put = forward)

Put call inequality: $S - K \leq C_A - P_A \leq S - Ke^{-r(T-t)}$

European-American relationship: $P_A > P_E, C_A = C_E$ (for non-dividend-paying)

Binomial tree

Risk neutral probability: $q = \frac{e^{r\delta t} - d}{u - d}, u = e^{\sigma\sqrt{\delta t}}, d = u^{-1} = e^{-\sigma\sqrt{\delta t}}$

Pricing: $f = e^{-r\delta t}[qf_u + (1-q)f_d]$

Backward induction: start from payoff as terminal prices (American: take max between payoff and f)

Black–Scholes–Merton model

Black-Scholes equation: $\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} = rV$

Black-Scholes formula: $C(S_t, t) = \Phi(d_1)S_t - \Phi(d_2)Ke^{-r(T-t)}$,

$P(S_t, t) = Ke^{-r(T-t)} - \mathrm{S_t} + C(S_t, t) = \Phi(-d_2)Ke^{-r(T-t)} - \Phi(-d_1)S_t$

where $d_1 = \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right], d_2 = d_1 - \sigma\sqrt{T-t}$

Implied volatility: the value of volatility when back-solving an option pricing model (such as BS) with current market price

III) Stochastic calculus

Brownian motion

Wiener process: W_t is called a Wiener process if the following holds

Stationary increment: $W_t - W_s \sim \mathrm{N}(0, t-s)$

Independent increment: $W_{t_4} - W_{t_3} \perp W_{t_2} - W_{t_1}$

Starts at zero: $\mathbb{P}\left(W_{t_0} = 0\right) = 1$

Properties: $\mathrm{Cov}(W_s, W_t) = \min(s,t)$, $[dW_t]^2 = dt$ (quadratic variation), nowhere differentiable

Itô's process: X_t is an Itô's process if it is solution to the following stochastic differential equation

$$\begin{cases} dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t \\ X_0 = a \end{cases}$$

Where $\mu(t, X_t)$ is known as the drift function and $\sigma(t, X_t)$ is known as the volatility function. **You may think $dX_t \approx X_{t+\delta t} - X_t$ and $dt \approx \delta t$ (useful in simulation)**

Stochastic integral

Definition: $\int_0^T f(s, W_s)dW_s = \lim_{\delta(\pi)\to 0} \sum_{j=0}^{N(\pi)-1} f(t_j, W_{t_j})(W_{t_{j+1}} - W_{t_j})$

Itô's lemma: $df(t, X_t) = \left[\frac{\partial f}{\partial t} + \mu(t, X_t)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2(t, X_t)\frac{\partial^2 f}{\partial x^2}\right]dt + \sigma(t, X_t)\frac{\partial f}{\partial x}dW_t$

Geometric Brownian motion: $dS_t = rS_t dt + \sigma S_t dW_t \Rightarrow S_t = S_0 e^{\left(r-\frac{1}{2}\sigma^2\right)t+\sigma W_t}$

Consequently, $S_{t+\delta t} = S_t e^{\left(r-\frac{1}{2}\sigma^2\right)\delta t+\sigma\sqrt{\delta t}Z}$ where $Z \sim \mathrm{N}(0,1)$

Finding stochastic integral: "guess" the function such that it will contain the integrand in its SDE. Use Itô's lemma to find SDE of the guess and then integrate both sides

Solving SDE: "guess" a solution and use Itô's lemma to verify that the solution satisfies the SDE (the following table is borrowed from Prof. Yau Chun Yip's notes on Stochastic Calculus)

Name	SDE	Solution (X_t)
Ornstein-Uhlenbeck(OU) process	$dX_t = -\alpha X_t dt + \sigma dW_t$	$ce^{-\alpha t} + \sigma e^{-\alpha t}\int_0^t e^{\alpha s} dW_s$
Mean reverting OU	$dX_t = (m - \alpha X_t)dt + \sigma dW_t$	$\frac{m}{\alpha} + \left(c - \frac{m}{\alpha}\right) e^{-\alpha t} + \sigma \int_0^t e^{\alpha(s-t)} dW_s$
Geometric Brownian motion	$dX_t = aX_t dt + bX_t dW_t$	$ce^{(a-b^2/2)t+bW_t}$
Brownian bridge	$dX_t = \frac{b-X_t}{1-t} dt + dW_t$	$a(1-t) + bt + (1-t)\int_0^t \frac{dW_s}{1-s}$
	$dX_t = \left(\sqrt{1+X_t^2} + \frac{1}{2}X_t\right) dt + \sqrt{1+X_t^2} dW_t$	$\sinh(c+t+W_t)$
	$dX_t = X_t^3 dt + X_t^2 dW_t$	$\frac{1}{c-W_t}$
	$dX_t = -\frac{1}{2}X_t dt + \sqrt{1-X_t^2} dW_t$	$\sin(c+W_t)$
	$dX_t = -\frac{1}{1+t}X_t dt + \frac{1}{1+t} dW_t$	$(c+W_t)/(1+t)$
	$dX_t = r dt + \alpha X_t dW_t$	$ce^{\alpha W_t - \frac{1}{2}\alpha^2 t} + r\int_0^t e^{\alpha(W_t-W_s)-\frac{1}{2}\alpha^2(t-s)} ds$

Integrating factor: add e^{rt} to both sides of a SDE (target: cancel some terms)

Martingale property: $\mathbb{E}\left[\int_0^T f(t,W_t)dW_t \middle| \mathcal{F}_s\right] = \int_0^s f(t,W_t)dW_t$

In particular, $\mathbb{E}\left(\int_0^T f(t,W_t)dW_t\right) = 0$

Itô isometry: $\mathbb{E}\left[\left(\int_0^T f(t,W_t)dW_t\right)^2\right] = \int_0^T \mathbb{E}[f(t,W_t)^2]dt$

Similarly, $\mathbb{E}\left[\left(\int_0^T f(t,W_t)dW_t\right)\left(\int_0^T g(t,W_t)dW_t\right)\right] = \int_0^T \mathbb{E}[f(t,W_t)g(t,W_t)]dt$

Product rule: $d(X_tY_t) = X_t dY_t + Y_t dX_t + d[\sigma(t,X_t)W_t, \bar{\sigma}(t,Y_t)\bar{W}_t]$

IV) Simulation methods

Theoretical support

Sample mean: $\bar{X}_n = \sum_{i=1}^{n} X_i$

Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

Law of large numbers (WLLN): Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean θ and variance σ^2, then $\bar{X}_n \approx \theta$ as $n \to \infty$

Central limit theorem: Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean θ and finite variance σ^2, then $\bar{X}_n \approx \mathrm{N}\left(\theta, \frac{\sigma^2}{n}\right)$ as $n \to \infty$

Modulo operation: find the remainder of a division; denoted by mod. Commonly used in generating pseudorandom number

Probability integral transform: if X is a continuous random variable with cdf F_X, then $Y = F_X(X) \sim \mathrm{Unif}(0,1)$. In other words, if we can find the cdf inverse F_X^{-1}, then $F_X^{-1}(U)$ and X have the same distributions

Standard Monte Carlo

Idea: take average of independent replications/scenarios of the reality/future

Algorithm:

1) Generate random variable X_i
2) Calculate $h_i = h(X_i)$, where h is the target function
3) Repeat 1 and 2 for n times
4) $\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n} h_j$ (remember to do discounting if necessary)

Margin of error: terminate a simulation when $\frac{s^2}{\sqrt{n}} \leq d$, where s^2 is the sample variance and d is the maximum tolerable error

Inverse transform

Idea: if we know $X \sim F_X$ (i.e., the cdf), we can generate X out of uniform random numbers

Algorithm (discrete):

1) Generate $U \sim \mathrm{Unif}(0,1)$
2) $X = x_j$ if $\sum_{i=0}^{j-1} p_i \le U < \sum_{i=0}^{j} p_i$

Algorithm (continuous):

1) Generate $U \sim \mathrm{Unif}(0,1)$
2) $X = F_X^{-1}(U)$ assuming the inverse exists

Rejection sampling

Idea: if we can simulate $Y \sim G_Y$ easily, we can use the proportional distribution (likelihood ratio) as a basis to simulate X with pdf $f(x)$

Algorithm:

1) Find $\mathrm{c} = \max_y \frac{f(y)}{g(y)}$
2) Generate Y_i from a density g, e.g., $U_1 \sim \mathrm{Unif}(0,1) \Rightarrow Y_i = G^{-1}(U_1)$
3) Generate $U_2 \sim \mathrm{Unif}(0,1)$
4) If $U_2 \le \frac{1}{c} \cdot \frac{f(Y_i)}{g(Y_i)}$, set $X_i = Y_i$. Otherwise return to 2

Number of iterations needed: $N \sim \mathrm{Geo}\left(\frac{1}{c}\right) \Rightarrow \mathbb{E}(N) = \mathrm{c}$

V) Variance reduction

Antithetic variables

Idea: if we are able to generate negatively correlated underlying random variables, the estimator can have lower variance as compared with independent samples. This requires the target function $h(x)$ to be monotone

Algorithm:

1) Generate $U \sim \text{Unif}(0,1)$
2) Set $X_i = F^{-1}(U), Y_i = F^{-1}(1-U)$ (note: want X, Y same distribution but negative correlation)
3) Repeat 1 and 2 for $i = 1, \dots, n$
4) $\hat{\theta} = \frac{1}{2n}\sum_{i=1}^{n}[h(X_i) + h(Y_i)]$

Useful corollary: if $h(x)$ is monotone, then $\text{Cov}\big(h(U), h(1-U)\big) \leq 0$ where $U \sim \text{Unif}(0,1)$

Stratified sampling

Idea: if we have information about grouping in the population, then we may use conditional mean (mean of subgroup) as the sample from the population

Algorithm:

1) Generate $V_{i,j} = \frac{1}{B}(U_{i,j} + i - 1)$ where $U_{i,j} \sim \text{Unif}(0,1)$ for $i = 1, \dots, B; j = 1, \dots, N_B$
2) Set $X_{i,j} = F^{-1}(V_{i,j})$
3) $\hat{\theta} = \frac{1}{B \times N_B}\sum_{j=1}^{N_B}\big[h\big(X_{1,j}\big) + h\big(X_{2,j}\big) + \cdots + h\big(X_{B,j}\big)\big]$ (average over subsamples and bins, remember to adjust for conditional probability)

Control variate

Idea: if we combine the estimate of our target unknown quantity with estimates of some known quantities, we can exploit the known information

Algorithm:

1) Find μ_Y for Y with a known distribution (or estimate μ_Y via pilot simulation)
2) Generate X_i, Y_i for $i = 1, \dots, n$
3) Compute $\bar{X}, \bar{Y}, \hat{\sigma}_{XY}, \hat{\sigma}_Y^2$

4) $\hat{\theta} = \bar{X} - \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_Y^2}(\bar{Y} - \mu_Y)$

Pilot simulation: we can run a simulation with a small sample size (e.g., $m = 100$) and compute $\hat{\sigma}_{XY}, \hat{\sigma}_Y^2$ and $\mu_Y = \bar{Y}_m$ based on this pilot sample. Then we can use their values when we compute $\hat{\theta} = \bar{X}_n - \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_Y^2}(\bar{Y}_n - \mu_Y)$ for our target n samples

Properties of effective control: evaluable from simulation data, known mean and high correlation with the simulation variable. Possible candidates are underlying random variable (e.g. uniform when we use inverse transform) and martingale transform (will not be tested)

Importance sampling

Idea: if certain values of the simulation variable have more impact on the parameter of interest (e.g. probability of a rare event), we can try to "emphasize" those values by sampling them more frequently and reduce variance. This can be done by changing the probability measure using the likelihood ratio (technically it is called Radon–Nikodym derivative) as weight

Algorithm:

1) Find the likelihood ratio $\frac{f(x)}{g(x)}$ where $f(x)$ is the original target pdf
2) Generate $X_i \sim G$ for $i = 1, \dots, n$
3) $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \frac{h(X_i)f(X_i)}{g(X_i)}$

Maximum principle: choose g such that both $g(x)$ and $h(x)f(x)$ take maximum values at the same $x = x^*$

Exponential tilting

Tilted density: $f_t(x) = \frac{e^{tx}f(x)}{M_X(t)}$ where $M_X(t)$ is the mgf of X. Useful for rare event simulation such as choosing g in rejection sampling or importance sampling

Choice of t for importance sampling: choose t such that the upper bound of $\frac{h(x)f(x)}{f_t(x)} = \frac{h(x)M_X(t)}{e^{tx}}$ is minimized. In particular, we first find $x = x_t^*$ (subscript t because it may depend on t) such that $\frac{h(x)f(x)}{f_t(x)} \leq \frac{h(x_t^*)f(x_t^*)}{f_t(x_t^*)}$ for all x in the support. Then we minimize $\frac{h(x_t^*)f(x_t^*)}{f_t(x_t^*)}$ with respect to t