IMPORTANCE SAMPLING

Importance Sampling Background: let $\mathbf{x} = (x_1, \ldots, x_n)$,

$$\begin{aligned} \theta &= E[h(\mathbf{X})] = \int h(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &\approx \frac{1}{N} \sum_{i=1}^{N} h(\mathbf{X}_i) = \bar{\Theta}, \end{aligned}$$

if $\mathbf{X}_i \sim F(\mathbf{X})$, and $F(\mathbf{x})$ is cdf for $f(\mathbf{x})$. For many problems, $F(\mathbf{x})$ is difficult to sample from and/or $Var(h)$ is large.

- If a related, easily sampled pdf $g(\mathbf{x})$ is available, could use

$$\begin{aligned} \theta = E_g[\frac{h(\mathbf{X}) f(\mathbf{X}}{g(\mathbf{X})}] &= \int \frac{h(\mathbf{x}) f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} \\ &\approx \frac{1}{N} \sum_{i=1}^{N} \frac{h(\mathbf{X}_i) f(\mathbf{X}_i)}{g(\mathbf{X}_i)}, \end{aligned}$$

  with $\mathbf{X}_i \sim G(\mathbf{X})$, for associated cdf $G(\mathbf{X})$.

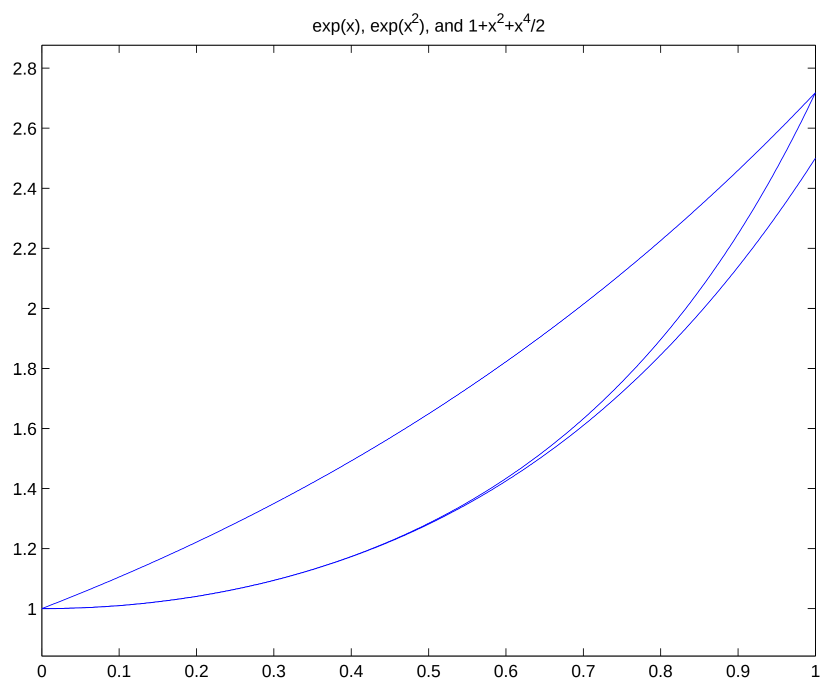
- **Importance sampling**: if $Var(\frac{h(\mathbf{x}) f(\mathbf{x})}{g(\mathbf{x})})$ is small, $g(\mathbf{x})$ samples are concentrated where $h(\mathbf{x}) f(\mathbf{x})$ is "important":

IMPORTANCE SAMPLING CONT.

- Importance Sampling Example: $\theta = \int_0^1 e^{x^2} dx$.
  Try $g(x) = e^x$; so $\theta = \int_0^1 e^{x^2-x} e^x dx$.
  To find $G(x)$, note $\int_0^1 e^x dx = e - 1$, so
  $$G(x) = \frac{1}{e-1}\int_0^x e^t dt = (e^x - 1)/(e-1);$$
  then using $X_i = \ln(1 + (e-1)U_i)$,
  $$\theta = (e-1)\int_0^1 e^{x^2-x}\frac{e^x}{e-1}dx \approx \frac{(e-1)}{N}\sum_{i=1}^{N} e^{X_i^2 - X_i},$$

```
N = 10000; U = rand(1,N); Y = exp(U.^2);
disp( [mean(Y) 2*std(T)/sqrt(N)]) % simple MC
       1.4672     0.009463
e = exp(1); X = log(1+(e-1)*U);
T = (e-1)*exp(X.*(X-1));
disp( [mean(T) 2*std(T)/sqrt(N)]) % importance
       1.4628    0.0022348
```

Error reduction by $\approx 1/4$.



IMPORTANCE SAMPLING CONT.

Alternative: $g(x) = 1 + x^2$, $G(x)$?

$$\theta = \frac{4}{3}\int_0^1 \frac{e^{x^2}}{1+x^2}\frac{3(1+x^2)}{4}dx \approx \frac{4}{3N}\sum_{i=1}^N \frac{e^{X_i^2}}{1+X_i^2},$$

with $X_i \sim \frac{3}{4}X + \frac{1}{4}X^3$.

```
N = 10000; U = rand(1,N); I = rand(1,N)<3/4;
X = I.*U + (1-I).*U.^(1/3);
T = 4*exp(X.^2)./(3*(1+X.^2));
disp( [mean(T) 2*std(T)/sqrt(N)]) % importance
        1.4627     0.0028178
```

Better $g(x) = 1 + x^2 + x^4/2$, $G(x) = \frac{30}{43}x + \frac{10}{43}x^3 + \frac{3}{43}x^5$;

$$\begin{aligned}\theta &= \frac{43}{30}\int_0^1 \frac{e^{x^2}}{1+x^2+x^4/2}\frac{30(1+x^2+x^4/2)}{43}dx \\ &\approx \frac{43}{30N}\sum_{i=1}^N \frac{e^{X_i^2}}{1+X_i^2+X_i^4/2},\end{aligned}$$

with $X_i \sim \frac{30}{43}x + \frac{10}{43}x^3 + \frac{3}{43}x^5$.

```
N = 10000; U = rand(1,N); V = rand(1,N);
I = V<30/43; J = V>40/43;
X = I.*U + (1-I).*(1-J).*U.^(1/3)+ J.*U.^(1/5);
T = 43*exp(X.^2)./(30*(1+X.^2+X.^4/2));
disp( [mean(T) 2*std(T)/sqrt(N)]) % importance
        1.4623   0.00072228
```

IMPORTANCE SAMPLING CONT.

- Higher dimensional problems: often
  $$f(\mathbf{x}) \approx g(\mathbf{x}) = g_1(x_1)g_2(x_2)\cdots g_n(x_n),$$
  so samples are from a sequence of 1-d samples.

  2-d example: $\theta = \int_0^1 \int_0^1 e^{(x_1+x_2)^2} d\mathbf{x}$; if $g(\mathbf{x}) = e^{x_1}e^{x_2}$;
  $\theta = \int_0^1 \int_0^1 e^{((x_1+x_2)^2 - x_1 - x_2} e^{x_1+x_2} d\mathbf{x}$.

  After scaling, with $X_{ij} = \ln(1 + (e-1)U_{ij})$,
  $$\begin{aligned} \theta &= (e-1)^2 \int_0^1 \int_0^1 e^{((x_1+x_2)^2 - x_1 - x_2)} \frac{e^{x_1+x_2}}{(e-1)^2} d\mathbf{x} \\ &\approx \frac{(e-1)^2}{N} \sum_{i=1}^{N} e^{(X_{1i}+X_{2i})^2 - X_{1i} - X_{2i}}. \end{aligned}$$

```
N = 10000; U = rand(2,N); T = exp(sum(U).^2);
disp( [mean(T) 2*std(T)/sqrt(N)]) % simple MC
      4.9204      0.12261
e = exp(1); X = log(1+(e-1)*U);
T = (e-1)^2*exp(sum(X).^2-sum(X));
disp( [mean(T) 2*std(T)/sqrt(N)])
      4.8863     0.065169
```

  Better $g(\mathbf{x}) = e^{2x_1}e^{2x_2}$, with $g(1,1) = f(1,1)$? Then
  $G_i(x) = \frac{e^{2x}-1}{e^2-1}$, $X_{ij} = \ln(1 + (e^2-1)U_{ij})/2$, and
  $$\theta = \frac{(e^2-1)^2}{4} \int_0^1 \int_0^1 e^{((x_1+x_2)^2 - 2(x_1+x_2))} \frac{4e^{2(x_1+x_2)}}{(e^2-1)^2} d\mathbf{x},$$

```
e = exp(1); X = log(1+(e^2-1)*U)/2;
T = (e^2-1)^2*exp(sum(X).^2-2*sum(X))/4;
disp( [mean(T) 2*std(T)/sqrt(N)])
      4.9008    0.0082436
```

  Better $g(\mathbf{x}) = 1 + (x_1 + x_2)^2$?

IMPORTANCE SAMPLING CONT.

Tilted Densities $g(x)$: given pdf $f(x)$

let $M(t) = \int e^{tx} f(x)dx$ (the moment generating function).

The **tilted density** for $f(x)$ is $f_t(x) = \frac{e^{tx}f(x)}{M(t)}$.

- Examples
  - Exponential densities: if $f(x) = \lambda e^{-\lambda x},\ x \in [0, \infty)$,
$$f_t(x) = (\lambda - t)e^{-(\lambda - t)x}, \quad t < \lambda.$$
  - Bernoullli pmf's: $f(x) = p^x(1-p)^{1-x}$, $x = 0,\ 1$.
    $M(t) = E_f[e^{tx}] = e^t p + (1-p)$, so
    $f_t(x) = \frac{e^{tx}p^x(1-p)^{1-x}}{e^t p + (1-p)} = \left(\frac{e^t p}{e^t p + (1-p)}\right)^x \left(\frac{1-p}{e^t p + (1-p)}\right)^{1-x}$,
    a Bernoulli RV with $p_t = \frac{e^t p}{e^t p + (1-p)}$.
    So $f/f_t = \frac{e^t p + (1-p)}{e^{tx}} = e^{-tx}(e^t p + (1-p))$
    Generalization: if $f(x)$ is a $Binomial(n, p)$ pmf, $f_t(x)$ is $Binomial(n, e^t p + 1 - p)$, with $M(t) = (e^t p + 1 - p)^n$.
  - Normal densities: if $f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}},\ x \in (-\infty, \infty)$,
$$e^{tx} f(x) = \frac{e^{xt}e^{-x^2/2}}{\sqrt{2\pi}} = \frac{e^{-(x-t)^2/2}e^{-t^2/2}}{\sqrt{2\pi}}$$
    so $f_t(x) = \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}}$, $Normal(t, 1)$, with $M(t) = e^{-t^2/2}$.
    Generalization: if $f(x)$ is a $Normal(\mu, \sigma^2)$ pdf, then $f_t(x)$ is a $Normal(\mu + \sigma^2 t, \sigma^2)$ pdf.
- Choosing t: pick t with small $Var(\frac{h(\mathbf{x})f(\mathbf{x})}{f_t(\mathbf{x})})$.
  Text heuristic for exponentials and Bernoullis:
  if $h = I\{\sum X_i > a\}$, choose $t = t^*$ with $E_{t^*}[\sum X_i] \approx a$.

IMPORTANCE SAMPLING CONT.

- Examples:

1. Bernoulli RV Examples: if $X_i's$ are independent Bernoulli(p_i) RV's and $\theta = I\{\sum_{i=1}^n X_i > a\} = I\{S > a\}$.

$$\hat{\theta} = I\{S > a\}e^{-tS}\prod_{i=1}^{n}(e^t p_i + (1-p_i)),$$

and

$$E_t[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n}\frac{e^t p_i}{e^t p_i + (1-p_i)}.$$

Example with $n = 20$, $p_i = .4$, $a = 16$; choose t so that
$E_t[S] = 20\frac{.4e^t}{.4e^t+.6} = 16$, with solution $e^{t^*} = 6$;
then $p_t = .8$, $e^{t^*}p + (1-p) = 3$, and estimator is
$\hat{\theta} = I\{\sum X_i > a\}6^{-S}3^{20} = 3^{20-S}I\{\sum X_i > a\}/2^S$.

Matlab

```
N = 100000; p = .4; n = 20;
I = sum( rand(n,N) < p ) > 16; % Simple MC
disp([mean(I) 2*std(I)/sqrt(N)])
   6e-05   4.8989e-05
S = sum( rand(n,N) < .8 ); % importance
I = 3.^(20-S).*( S > 16 )./2.^S;
disp([mean(I) 2*std(I)/sqrt(N)])
   4.7575e-05   5.1608e-07
N = 10000000; p = .4; n = 20;
I = sum( rand(n,N) < p ) > 16; % Simple MC
disp([mean(I) 2*std(I)/sqrt(N)])
   4.82e-05   4.3908e-06
```

IMPORTANCE SAMPLING CONT.

2. Exponential RV Example:
   if $X_i \sim Exp(\frac{1}{i+2})$, $i = 1, \ldots, 4$, $S(\mathbf{X}) = \sum_{i=1}^{4} X_i$, find

$$\theta = \int_0^\infty \cdots \int_0^\infty h(\mathbf{x}) \frac{e^{-\sum_{i=1}^{4} \frac{x_i}{i+2}}}{3 \cdot 4 \cdot 5 \cdot 6} \, d\mathbf{x},$$

with $h(\mathbf{x}) = S(\mathbf{x}) I\{S(\mathbf{x}) > 62\}$.

Raw simulation uses $X_{ij} \sim Exp(\frac{1}{i+2})$, to estimate

$$\theta \approx \frac{1}{N} \sum_{j=1}^{N} h(\mathbf{X}_j).$$

Matlab

```
N = 100000; U = rand(4,N);
X = -diag([3:6])*log(1-U);
S = sum(X); h = S.*( S > 62 );
disp( [mean(h) 2*std(h)/sqrt(N)])
     0.066974     0.013647
```

Note: to find $E[S|S > 62]$, divide by $E[I(S > 62)]$

```
E = h/mean((S>62));
disp([mean(E) 2*std(E)/sqrt(N)])
       66.948        15.237
```

IMPORTANCE SAMPLING CONT.

For tilted density, use common tilt parameter t,
so that $X_i \sim Exp(1/(i+2)-t)$,

$$\theta \;=\; \prod_{i=1}^{4} \frac{i+2}{1-(i+2)t} \int_{[0,\infty)^4} \frac{h(\mathbf{x})e^{-tS(\mathbf{x})}}{\prod_{i=1}^{4}(i+2)} \frac{e^{-\sum_{i=1}^{4} x_i(\frac{1}{i+2}-t)}}{\prod_{i=1}^{4} \frac{i+2}{1-(i+2)t)}}\, d\mathbf{x};$$

$$\approx\; \frac{C}{N}\sum_{j=1}^{N} h(\mathbf{X}_j)e^{-tS(\mathbf{X}_j)}, \text{ with } C = \prod_{i=1}^{4} \frac{1}{1-(i+2)t}.$$

Text estimates "good" $t = .14$, by approximately solving

$$\sum_{i=1}^{4} E_t[X_i] = \frac{3}{1-3t} + \frac{4}{1-4t} + \frac{5}{1-5t} + \frac{6}{1-6t} = 62.$$

But "guess and check" with Matlab finds "better" $t \approx .136$.
Matlab tests:

```
t = .14; Cd = 1./(1-[3:6]*t); C = prod(Cd);
St = -([3:6].*Cd)*log(1-U);
ht = C*St.*( St > 62 ).*exp(-t*St);
disp([mean(ht) 2*std(ht)/sqrt(N)])
     0.063281    0.0010059
t = .136; Cd = 1./(1-[3:6]*t); C = prod(Cd);
St = -([3:6].*Cd)*log(1-U);
ht = C*St.*( St > 62 ).*exp(-t*St);
disp([mean(ht) 2*std(ht)/sqrt(N)])
     0.06201    0.00099896
E = ht/mean(C*(St > 62).*exp(-t*St));
disp([mean(E) 2*std(E)/sqrt(N)])% Expected Value
      68.215        1.0931
```

Note smaller standard errors compared to raw sampling.

IMPORTANCE SAMPLING CONT.

- Tilting for Normal Densities: if $f(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2}$, tilted density $f_t(x) = \frac{f(x)e^{xt}}{M(t)}$ is a shifted normal.
  For multidimensional problems, t could be a vector $\mathbf{t}$.
  Choice of $\mathbf{t}$? Try to make $Var(\frac{h(\mathbf{x})f(\mathbf{x})}{f(\mathbf{x}-\mathbf{t})})$ small:
  a) choose point $\mathbf{t}$ where $h(\mathbf{x})f(\mathbf{x})$ is maximum (mode), or
  b) choose $\mathbf{t} = E[\mathbf{x}h(\mathbf{x})]/E[h(\mathbf{x})]$ (mean).
  Asian Option example: this has $S_m = S_{m-1}e^{(r-\frac{\sigma^2}{2})\delta+\sigma\sqrt{\delta}Z_m}$,
  with $\delta = T/M$, $Z_m \sim Normal(0,1)$ and expected profit

$$
\begin{aligned}
\theta &= E[e^{-rT}\max(\frac{1}{M}\sum_{i=1}^{M} S_i(\mathbf{Z}) - K, 0)] \\
&= e^{-rT}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\max(\frac{1}{M}\sum_{i=1}^{M} S_i(\mathbf{z}) - K, 0)\frac{e^{-\sum_{i=1}^{M} z_i^2/2}}{(\sqrt{2\pi})^m}d\mathbf{z} \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(\mathbf{z})\frac{e^{-\sum_{i=1}^{M} z_i^2/2}}{(\sqrt{2\pi})^m}d\mathbf{z},
\end{aligned}
$$

  with $h(\mathbf{z}) = e^{-rT}\max(\frac{1}{M}\sum_{i=1}^{M} S_i(\mathbf{Z}) - K, 0)$.
  For method a), find $\mathbf{t}$ to maximize $h(\mathbf{z})e^{-\sum_{i=1}^{M} z_i^2/2}$.
  For method b), $\mathbf{t}$ can be estimated from data.
  Given $\mathbf{t}$, use

$$
\begin{aligned}
\hat{\theta} &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(\mathbf{z})\frac{e^{-\sum_{i=1}^{M} z_i^2/2}}{e^{-\sum_{i=1}^{M}(z_i-t_i)^2/2}}\frac{e^{-\sum_{i=1}^{M}(z_i-t_i)^2/2}}{(\sqrt{2\pi})^m}d\mathbf{z} \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(\mathbf{y}+\mathbf{t})\frac{e^{-\sum_{i=1}^{M}(y_i+t_i)^2/2}}{e^{-\sum_{i=1}^{M} y_i^2/2}}\frac{e^{-\sum_{i=1}^{M} y_i^2/2}}{(\sqrt{2\pi})^m}d\mathbf{y}.
\end{aligned}
$$

IMPORTANCE SAMPLING CONT.

Example with $M = 16$,

$S_0 = K = 50,\ T = 1,\ r = .05,\ \sigma = .1.$

Matlab test using method b):

```
M = 16; S0 = 50; K = 50; T = 1; dlt = T/M;
r = 0.05; s = 0.1; rd = ( r - s^2/2 )*dlt;
N = 10000; z = randn(M,N); % Simple MC
S = S0*exp(cumsum(rd + s*sqrt(dlt)*z));
h = exp(-r*T)*max( mean(S)-K, 0 );
disp([mean(h) var(h) 2*std(h)/sqrt(N)])
      1.9465       4.825      0.043932
t = z*h'/sum(h); % Tilt Vector
y = z; z = y + t*ones(1,N);
S = S0*exp(cumsum(rd + s*sqrt(dlt)*z));
h = exp(-r*T)*max( mean(S)-K, 0 );
ht = h.*exp(sum(y.*y-z.*z)/2); % Importance
disp([mean(ht) var(ht) 2*std(ht)/sqrt(N)])
      1.9136      0.66366     0.016293
```

Notice variance reduction from tilted sampling.

Using method a) with Matlab "fminsearch" to find t:

```
Sf = @(z)S0*exp(cumsum(rd+s*sqrt(dlt)*z));
hf = @(z)exp(-z'*z/2)*max(mean(Sf(z))-K,0);
t = fminsearch(@(z)-hf(z), ones(M,1) ); % Tilt t
y = z; z = y + t*ones(1,N);
S = S0*exp(cumsum(rd + s*sqrt(dlt)*z));
h = exp(-r*T)*max( mean(S)-K, 0 );
ht = h.*exp(sum(y.*y-z.*z)/2); % Importance
disp([mean(ht) var(ht) 2*std(ht)/sqrt(N)])
       1.8868      0.35124     0.011853
```