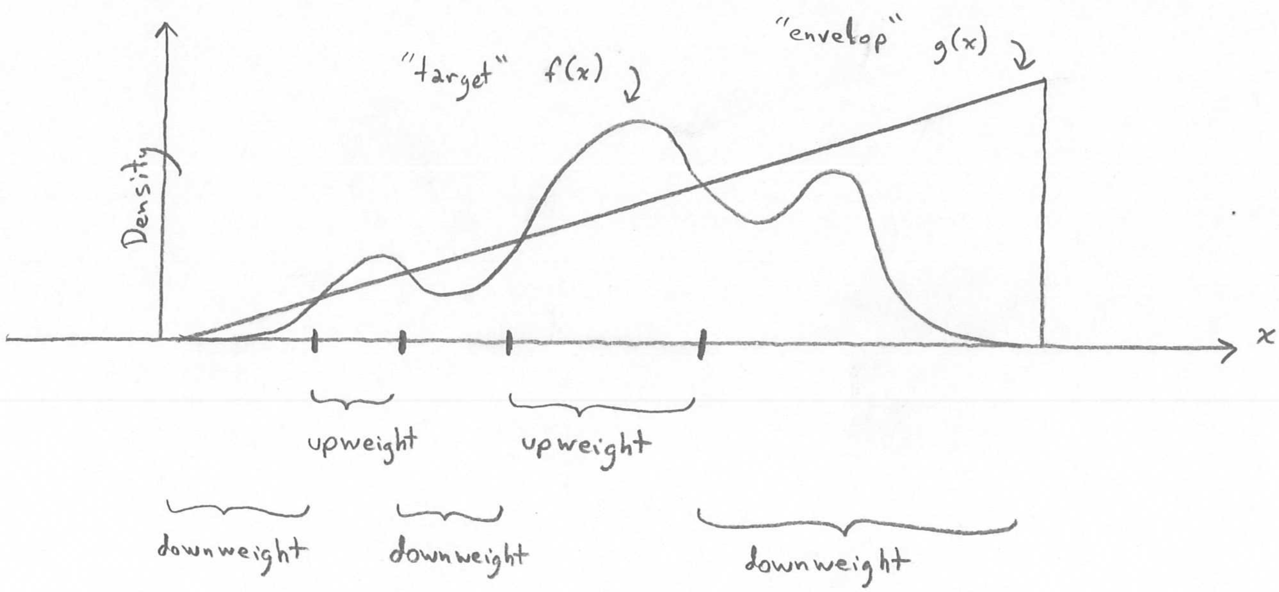
Importance Sampling

- Recall that in rejection sampling, we either keep a draw from the envelop distribution (thereby accepting it as a draw from the target) or we reject it (thereby completely disregarding it).
- In effect, rejection sampling puts weight of 1 on some draws and 0 on other draws.
- Why be so rigid? Why not permit weights other than 0 or 1? That's the idea behind importance sampling.



- Importance sampling accepts all the draws from the envelop distribution and gives them the weight $w(x) = f(x)/g(x)$, where $f(x)$ is the target and $g(x)$ is the envelop distribution.

- Theory:

$$\underbrace{E_f(h(X))}_{\text{"The expected value of } h(X) \text{ with respect to the density } f(x)\text{"}} = \int_{\text{support of } f} h(x)\, f(x)\, dx = \int h(x) \frac{f(x)}{g(x)}\, g(x)\, dx$$

$$= \underbrace{E_g\left(h(X)\, w(X)\right)}_{\text{"The expected value of } h(X)\,w(X) \text{ with respect to the density } g(x)\text{"}}$$

- Thus, we can estimate $E_f(h(X))$ through Monte Carlo integration:

  - Suppose $X_1, X_2, \ldots, X_B$ are draws from the distribution having density $g(x)$.
  - Then, $E_f(h(X)) = E_g(h(X)\,w(X)) \approx \underbrace{\frac{1}{B}\sum_{i=1}^{B} h(x_i)\,w(x_i)}_{\hat{\theta}}$
  - Assess Monte Carlo error!

    95% C.I. for $E_f(h(X)) = \theta$.

    $$\hat{\theta} \pm 1.96 \frac{s.e.}{\sqrt{B}},$$

    where $s.e. = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}\left(h(x_i)\,w(x_i) - \hat{\theta}\right)^2}$

  - Note that Monte Carlo error can be very large if the weights $w(x_i)$ have high variance!

- What if the target density $f(x)$ can only be evaluated up to a normalizing constant? That is, $f(x) = c f^*(x)$ and we can evaluate $f^*(x)$ but not $f(x)$ because we don't know c such that $\int f(x)\,dx = \int c f^*(x)\,dx = 1$.

- Note that $E_g(w(X)) = E_g\left(\frac{f(X)}{g(X)}\right)$

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

$$= \int f(x)\,dx$$

$$= 1$$

- By the Law of Large Numbers

$$\frac{1}{B}\sum_{i=1}^{B} w(x_i) \xrightarrow[p]{} E_g(w(X)) = 1$$

- We cannot compute $w(x) = \frac{f(x)}{g(x)} = \frac{c f^*(x)}{g(x)} = c\,w^*(x)$

but we can compute $w^*(x) = f^*(x)/g(x)$

- So, $\frac{1}{c}\left[\frac{1}{B}\sum_{i=1}^{B} c w^*(x_i)\right] \xrightarrow[p]{} \frac{1}{c}$

$$\underbrace{\frac{1}{B}\sum_{i=1}^{B} w^*(x_i)}_{\text{We can compute this for finite } B.} \xrightarrow[p]{} \frac{1}{c}$$

- Therefore $\overbrace{B \Big/ \sum_{i=1}^{B} w^*(x_i)} \xrightarrow[p]{} c$

- Putting it all together

$$E_f(h(x)) = E_g(h(x)\, w(x))$$

$$= E_g(h(x)\, c\, w^*(x))$$

$$= c\, E_g(h(x)\, w^*(x))$$

$$\approx \frac{B}{\sum_{i=1}^{B} w^*(x_i)} \frac{1}{B} \sum_{i=1}^{B} h(x_i)\, w^*(x_i)$$

$$= \frac{1}{B} \sum_{i=1}^{B} h(x_i) \underbrace{\left[\frac{B\, w^*(x_i)}{\sum_{j=1}^{B} w^*(x_j)} \right]}_{\hat{w}(x_i) \text{ are standardized weights.}}$$

$$= \frac{1}{B} \sum_{i=1}^{B} h(x_i)\, \hat{w}(x_i)$$