Motivation

- Cannot access \( p(\omega \mid y^o, A) \) analytically.
- Instead, simulate \( \omega^{(1)} \sim p(\omega \mid y^o, A) \).
- Approximate posterior moments \( E[h(\omega) \mid y^o, A] \).
- Bayes actions \( \hat{a} = \arg \min_{a \in A} E[L(a, \omega) \mid y^o, A] \).
Motivation

- Cannot access $p(\omega | y^o, A)$ analytically
- Instead, simulate $\omega^{(1)} \sim p\left( \frac{\omega}{y^o, A} \right)$
- Approximate posterior moments $E[h(\omega) | y^o, A]$
- Bayes actions $\hat{a} = \arg \min_{a \in A} E[L(a, \omega) | y^o, A]$
Notation

- The methods we discuss today can be used for a probability distribution of interest, $I$
- To this end:
  - $\theta^{(m)} \sim p(\theta \mid I)$
  - $\omega^{(m)} \sim p(\omega \mid \theta^{(m)}, I)$
Notation

- The methods we discuss today can be used for a probability distribution of interest, $I$
- To this end:
  - $\theta^{(m)} \sim p(\theta \mid I)$
  - $\omega^{(m)} \sim p(\omega \mid \theta^{(m)}, I)$
The setup

- Suppose the following is possible:
  - $\theta^{(m)} \overset{iid}{\sim} p(\theta | I)$
  - $\omega^{(m)} \overset{iid}{\sim} p(\omega | \theta^{(m)}, I)$

- Our setup (Theorem 4.1.1)
  - $\left\{ \theta^{(m)}, \omega^{(m)} \right\}$ is i.i.d.; $\theta^{(m)} \in \Theta$, $\omega^{(m)} \in \Omega$, $\theta^{(m)} \sim p(\theta | I)$
  - $h : \Omega \rightarrow \mathbb{R}^1$
The setup

- Suppose the following is possible:
  - $\theta^{(m)} \sim_{iid} p(\theta | l)$
  - $\omega^{(m)} \sim_{iid} p(\omega | \theta^{(m)}, l)$
- Our setup (Theorem 4.1.1)
  - $\{\theta^{(m)}, \omega^{(m)}\}$ is i.i.d.; $\theta^{(m)} \in \Theta$, $\omega^{(m)} \in \Omega$, $\theta^{(m)} \sim p(\theta | l)$
- $h : \Omega \rightarrow \mathbb{R}^1$
Some conditions

- First moment condition: $E[h(\omega) \mid l] = \bar{h}$;
- Second moment condition: $\text{var}[h(\omega) \mid l] = \sigma^2$;
- Probability mass condition: For given $p \in (0, 1)$, there is a unique $h_p$ such that the statements
  
  $$P[h(\omega) \leq h_p \mid l] \geq p \text{ and } P[h(\omega) \geq h_p \mid l] \geq 1 - p$$

  are both true;
- Positive p.d.f. condition: For the unique $h_p$ corresponding to $p$, $p [h(\omega) = h_p \mid l] > 0$. 
Some conditions

- First moment condition: $E[h(\omega) \mid l] = \bar{h}$;
- Second moment condition: $\text{var}[h(\omega) \mid l] = \sigma^2$;
- Probability mass condition: For given $p \in (0, 1)$, there is a unique $h_p$ such that the statements
  \[ P[h(\omega) \leq h_p \mid l] \geq p \text{ and } P[h(\omega) \geq h_p \mid l] \geq 1 - p \]
  are both true;
- Positive p.d.f. condition: For the unique $h_p$ corresponding to $p$, \( p[h(\omega) = h_p \mid l] > 0 \).
First result (a)

Given the setup,

\[ \omega^{(m)} \overset{i.i.d.}{\sim} p(\omega \mid I) \]

whether or not any of conditions 1-4 are true.

Second result (b)

Given the first moment condition,

\[ \overline{h}^{(M)} = M^{-1} \sum_{m=1}^{M} h(\omega^{(m)}) \overset{a.s.}{\longrightarrow} \overline{h}. \]
First result (a)

Given the setup,

$$\omega^{(m)} \overset{i.i.d.}{\sim} p(\omega \mid I)$$

whether or not any of conditions 1-4 are true.

Second result (b)

Given the first moment condition,

$$\overline{h}^{(M)} = M^{-1} \sum_{m=1}^{M} h(\omega^{(m)}) \overset{a.s.}{\rightarrow} \overline{h}.$$
Third result (c)

Given the first and second moment conditions,

\[ M^{1/2} \left( \bar{h}^{(M)} - \bar{h} \right) \xrightarrow{d} N(0, \sigma^2) \]

and

\[ \hat{\sigma}^2(M) = M^{-1} \sum_{m=1}^{M} \left[ h \left( \omega^{(m)} \right) - \bar{h}^{(M)} \right]^2 \xrightarrow{a.s.} \sigma^2 \]
Third result (c)

Given the first and second moment conditions,

\[ M^{1/2} \left( \overline{h}^{(M)} - \overline{h} \right) \xrightarrow{d} N(0, \sigma^2) \]

and

\[ \hat{\sigma}^2(M) = M^{-1} \sum_{m=1}^{M} \left[ h \left( \omega^{(m)} \right) - \overline{h}^{(M)} \right]^2 \xrightarrow{a.s.} \sigma^2 \]
Notation for quantiles

Let $\hat{h}_p^{(M)}$ be any real number such that

$$M^{-1} \sum_{m=1}^{M} I\left[ (-\infty, \hat{h}_p^{(M)}) \right] \left[ h\left( \omega^{(m)} \right) \right] \geq p$$

and

$$M^{-1} \sum_{m=1}^{M} I\left[ \hat{h}_p^{(M)}, \infty \right] \left[ h\left( \omega^{(m)} \right) \right] \geq 1 - p.$$
Notation for quantiles

Let $\hat{h}_p^{(M)}$ be any real number such that

$$M^{-1} \sum_{m=1}^{M} I\left(-\infty, \hat{h}_p^{(M)}\right) h\left(\omega^{(m)}\right) \geq p$$

and

$$M^{-1} \sum_{m=1}^{M} I\left[\hat{h}_p^{(M)}, \infty\right) h\left(\omega^{(m)}\right) \geq 1 - p.$$
Fourth result (d)

- Given the probability mass condition,

\[ \hat{h}_p^{(M)} \xrightarrow{a.s.} h_p \]

Fifth result (e)

- Given the probability mass and positive p.d.f. conditions,

\[ M^{1/2} \left[ \hat{h}_p^{(M)} - h_p \right] \xrightarrow{d} N \left\{ 0, p (1 - p) / p \left[ h(\omega) = h_p \mid I \right]^2 \right\} . \]
Fourth result (d)

- Given the probability mass condition,

\[ \hat{h}_p^{(M)} \xrightarrow{a.s.} h_p \]

Fifth result (e)

- Given the probability mass and positive p.d.f. conditions,

\[ M^{1/2} \left[ \hat{h}_p^{(M)} - h_p \right] \xrightarrow{d} \mathcal{N} \left\{ 0, \frac{p(1-p)}{p[h(\omega) = h_p | I]^2} \right\} \]
Assessment of approximation error

- Recall the third result (c):

\[ M^{1/2} \left( \bar{h}^{(M)} - \bar{h} \right) \xrightarrow{d} N(0, \sigma^2) \]

and

\[ \hat{\sigma}^2(M) = M^{-1} \sum_{m=1}^{M} \left[ h\left( \omega^{(m)} \right) - \bar{h}^{(M)} \right]^2 \xrightarrow{a.s.} \sigma^2 \]

- The numerical standard error of \( \bar{h}^{(M)} \) is

\[ \left( \hat{\sigma}^2(M) / M \right)^{1/2} \]
Assessment of approximation error

- Recall the third result (c):
  \[ M^{1/2} \left( \bar{h}^{(M)} - \bar{h} \right) \xrightarrow{d} N(0, \sigma^2) \]

  and
  \[ \hat{\sigma}^2(M) = M^{-1} \sum_{m=1}^{M} \left[ h \left( \omega^{(m)} \right) - \bar{h}^{(M)} \right]^2 \xrightarrow{a.s.} \sigma^2 \]

- The numerical standard error of \( \bar{h}^{(M)} \) is
  \[ \left( \hat{\sigma}^2(M) / M \right)^{1/2} \]
Approximation of Bayes actions by direct sampling: The setup

- $\theta^{(m)} \sim p(\theta | l)$, $\omega^{(m)} | (\theta^{(m)}, l) \sim p(\omega | \theta^{(m)}, l)$.
- $L(a, \omega) \geq 0$ is a loss function defined on $\Omega \times A$, where $A$ is an open subset of $\mathbb{R}^q$.
- The risk function

$$R(a) = \int_{\Omega} \int_{\Theta} L(a, \omega) p(\theta | l) p(\omega | \theta, l) \, d\theta d\omega$$

has a strict global minimum at $\hat{a} \in A \subseteq \mathbb{R}^m$. 
Approximation of Bayes actions by direct sampling: The setup

- \( \theta^{(m)} \sim p(\theta | I) \), \( \omega^{(m)} | (\theta^{(m)}, I) \sim p(\omega | \theta^{(m)}, I) \).
- \( L(a, \omega) \geq 0 \) is a loss function defined on \( \Omega \times A \), where \( A \) is an open subset of \( \mathbb{R}^q \).
- The risk function

\[
R(a) = \int_{\Omega} \int_{\Theta} L(a, \omega) \ p(\theta | I) \ p(\omega | \theta, I) \ d\theta d\omega
\]

has a strict global minimum at \( \hat{a} \in A \subseteq \mathbb{R}^m \).
First result

Under weak regularity conditions (Theorem 4.1.2, conditions (1 - 2))

\[
\lim_{M \to \infty} P \left[ \inf_{\mathbf{a} \in A_M} (\mathbf{a} - \hat{\mathbf{a}})' (\mathbf{a} - \hat{\mathbf{a}}) > \varepsilon \mid I \right] = 0.
\]
First result

Under weak regularity conditions (Theorem 4.1.2, conditions (1 - 2))

\[
\lim_{M \to \infty} P \left[ \inf_{a \in A_M} (a - \hat{a})' (a - \hat{a}) > \varepsilon \mid I \right] = 0.
\]
Second result

This requires Conditions 1-6, which include

\[ B = \text{var} \left[ \frac{\partial L(a, \omega)}{\partial a} \bigg|_{a=\hat{a}} \right] \]

exists and is finite;

\[ H = \mathbb{E} \left[ \frac{\partial^2 L(a, \omega)}{\partial a \partial a'} \bigg|_{a=\hat{a}} \right] \]

exists and is finite and nonsingular.
Second result

This requires Conditions 1-6, which include

\[ B = \text{var} \left[ \frac{\partial L(a, \omega)}{\partial a} \bigg|_{a=\hat{a}} \right] \]

exists and is finite;

\[ H = \mathbb{E} \left[ \frac{\partial^2 L(a, \omega)}{\partial a \partial a'} \bigg|_{a=\hat{a}} \right] \]

exists and is finite and nonsingular.
Second result (continued)

Then

\[ M^{1/2} (\hat{a}_M - \hat{a}) \xrightarrow{d} N (0, H^{-1} BH^{-1}) , \]

\[ M^{-1} \sum_{m=1}^{M} \partial L \left( a, \omega^{(m)} \right) / \partial a | a=\hat{a}_M \]

\[ \cdot \partial L \left( a, \omega^{(m)} \right) / \partial a' | a=\hat{a}_M \xrightarrow{p} B , \]

\[ M^{-1} \sum_{m=1}^{M} \partial^2 L \left( a, \omega^{(m)} \right) / \partial a \partial a' | a=\hat{a}_M \xrightarrow{p} H . \]
Second result (continued)

Then

- $M^{1/2} (\hat{a}_M - \hat{a}) \xrightarrow{d} N \left( 0, H^{-1}BH^{-1} \right)$,
- $M^{-1} \sum_{m=1}^{M} \frac{\partial L \left( a, \omega^{(m)} \right)}{\partial a} \bigg|_{a=\hat{a}_M}$
  \cdot \frac{\partial L \left( a, \omega^{(m)} \right)}{\partial a'} \bigg|_{a=\hat{a}_M} \xrightarrow{p} B,$
- $M^{-1} \sum_{m=1}^{M} \frac{\partial^2 L \left( a, \omega^{(m)} \right)}{\partial a \partial a'} \bigg|_{a=\hat{a}_M} \xrightarrow{p} H.$
Importance sampling: Motivation

Suppose that we cannot derive a method for drawing $\theta^{(m)} \sim p(\theta | l)$ but we can simulate $\theta^{(m)} \sim p(\theta | S)$ where $p(\theta | S)$ is similar to $p(\theta | l)$. 
Importance sampling: Motivation

Suppose that we cannot derive a method for drawing $\theta^{(m)} \sim p(\theta \mid I)$ but we can simulate $\theta^{(m)} \sim p(\theta \mid S)$ where $p(\theta \mid S)$ is similar to $p(\theta \mid I)$. 
Main results: The setup

\[ u\{\theta^{(m)}, \omega^{(m)}\} \text{ is independent and identically distributed,} \]
\[ \theta^{(m)} \sim p(\theta | S) \]
\[ \omega^{(m)} \sim p(\omega | \theta^{(m)}, l). \]

Define the weighting function

\[ w(\theta) = p(\theta | l) / p(\theta | S) \]

As before, \( h : \Omega \to \mathbb{R}^1 \)
Main results: The setup

\[ u\{\theta^{(m)}, \omega^{(m)}\} \text{ is independent and identically distributed,} \]
\[ \theta^{(m)} \sim p(\theta | S) \]
\[ \omega^{(m)} \sim p(\omega | \theta^{(m)}, I). \]

Define the weighting function

\[ w(\theta) = \frac{p(\theta | I)}{p(\theta | S)} \]

As before, \( h : \Omega \to \mathbb{R}^1 \)
Some conditions

- First moment condition: $E[h(\omega) \mid l] = \bar{h}$ exists
- Second moment condition: $\text{var}[h(\omega) \mid l] = \sigma^2$ exists
- Essential condition: The support of $p(\theta \mid S)$ includes $\Theta$.
- Bounded weight condition: $w(\theta) = \frac{p(\theta \mid l)}{p(\theta \mid S)}$ is bounded above on $\Theta$. 
Some conditions

- First moment condition: $E[h(\omega) \mid I] = \bar{h}$ exists
- Second moment condition: $\text{var}[h(\omega) \mid I] = \sigma^2$ exists
- Essential condition: The support of $p(\theta \mid S)$ includes $\Theta$.
- Bounded weight condition: $w(\theta) = p(\theta \mid I) / p(\theta \mid S)$ is bounded above on $\Theta$. 
More notation

Kernels

\[ k(\theta \mid I) = c_I \cdot p(\theta \mid I) \]
\[ k(\theta \mid S) = c_S \cdot p(\theta \mid S) \]
More notation

Kernels
\[ k(\theta | I) = c_I \cdot p(\theta | I) \]
\[ k(\theta | S) = c_S \cdot p(\theta | S) \]
First result

Given the first moment condition and the essential condition,

\[ h^{(M)} = \frac{\sum_{m=1}^{M} w(\theta^{(m)}) h(\omega^{(m)})}{\sum_{m=1}^{M} w(\theta^{(m)})} \overset{a.s.}{\to} h \]

The proof is straightforward.
First result

Given the first moment condition and the essential condition,

\[
\bar{h}^{(M)} = \frac{\sum_{m=1}^{M} w(\theta^{(m)}) h(\omega^{(m)})}{\sum_{m=1}^{M} w(\theta^{(m)})} \quad \text{a.s.} \quad \bar{h}
\]

The proof is straightforward.
Second result

Given all four conditions,

\[ M^{1/2} \left( \bar{h}^{(M)} - \bar{h} \right) \xrightarrow{d} N(0, \tau^2) \]

and

\[ \hat{\tau}^2(M) = \frac{M \sum_{m=1}^{M} \left[ h \left( \omega^{(m)} \right) - \bar{h}^{(M)} \right]^2 \nu \left( \theta^{(m)} \right)^2}{\left[ \sum_{m=1}^{M} \nu \left( \theta^{(m)} \right) \right]^2} \xrightarrow{a.s.} \tau^2. \]

The essential part of the proof is the “delta method.”
Second result

Given all four conditions,

\[ M^{1/2} \left( \bar{h}^{(M)} - \bar{h} \right) \overset{d}{\rightarrow} N (0, \tau^2) \]

and

\[ \hat{\tau}^2(M) = \frac{M \sum_{m=1}^{M} \left[ h \left( \omega^{(m)} \right) - \bar{h}^{(M)} \right]^2 w \left( \theta^{(m)} \right)^2}{\left[ \sum_{m=1}^{M} w \left( \theta^{(m)} \right) \right]^2} \overset{a.s.}{\rightarrow} \tau^2. \]

The essential part of the proof is the “delta method.”
Bayes actions

These results can be extended to provide simulation approximations of the Bayes action

\[ \hat{a} = \arg \min E \left[ l(\omega) \mid y^o, A \right] \]

and the Bayes risk

\[ R(a) = \int_{\Omega} \int_{\Theta} L(a, \omega) p(\theta \mid l) p(\omega \mid \theta, l) d\theta d\omega. \]

(Theorem 4.2.3)
Bayes actions

These results can be extended to provide simulation approximations of the Bayes action

\[ \hat{a} = \arg \min E \left[ l(\omega) \mid y^o, A \right] \]

and the Bayes risk

\[ R(a) = \int_\Omega \int_\Theta L(a, \omega) p(\theta \mid l) p(\omega \mid \theta, l) d\theta d\omega. \]

(Theorem 4.2.3)
A particularly useful property of importance sampling

Recall the marginal likelihood

\[ p(y^o \mid A) = \int_{\Theta_A} p(\theta_A \mid A) \, p(y^o \mid \theta_A, A) \, d\theta_A \]

\[ = \int_{\Theta_A} \frac{p(\theta_A \mid A) \, p(y^o \mid \theta_A, A)}{p(\theta_A \mid S)} \, p(\theta_A \mid S) \, d\theta_A \]

\[ = E \left[ \frac{p(\theta_A \mid A) \, p(y^o \mid \theta_A, A)}{p(\theta_A \mid S)} \right] \]

if \( \theta_A \sim p(\theta_A \mid S) \).
A particularly useful property of importance sampling

Recall the marginal likelihood

\[
 p (y^o \mid A) = \int_{\Theta_A} p (\theta_A \mid A) p (y^o \mid \theta_A, A) \, d\theta_A
\]

\[
= \int_{\Theta_A} \frac{p (\theta_A \mid A) p (y^o \mid \theta_A, A)}{p (\theta_A \mid S)} p (\theta_A \mid S) \, d\theta_A
\]

\[
= E \left[ \frac{p (\theta_A \mid A) p (y^o \mid \theta_A, A)}{p (\theta_A \mid S)} \right]
\]

if \( \theta_A \sim p (\theta_A \mid S) \).
Useful property (continued)

\[ p(y^o \mid A) = E \left[ \frac{p(\theta_A \mid A) p(y^o \mid \theta_A, A)}{p(\theta_A \mid S)} \right] \]

If \( \theta_A^{(m)} \overset{iid}{\sim} p(\theta_A \mid S) \) then

\[ E \left[ \frac{1}{M} \sum_{m=1}^{M} \frac{p(\theta_A^{(m)} \mid A) p(y^o \mid \theta_A^{(m)}, A)}{p(\theta_A^{m} \mid S)} \right] = E \left[ \frac{1}{M} \sum_{m=1}^{M} w(\theta_A^{(m)}) \right] . \]
**Useful property** (continued)

\[
p(y^o \mid A) = E \left[ \frac{p(\theta_A \mid A) p(y^o \mid \theta_A, A)}{p(\theta_A \mid S)} \right]
\]

If \( \theta_A^{(m)} \overset{iid}{\sim} p(\theta_A \mid S) \) then

\[
E \left[ \frac{1}{M} \sum_{m=1}^{M} \frac{p(\theta_A^{(m)} \mid A) p(y^o \mid \theta_A^{(m)}, A)}{p(\theta_A^{m} \mid S)} \right]
\]

\[
= E \left[ \frac{1}{M} \sum_{m=1}^{M} w(\theta_A^{(m)}) \right].
\]