Definition

Let $X$ be a random variable on a probability space $(\Omega, \mathcal{F}_0, \text{Prob})$ satisfying $E[|X|] < \infty$ and let $\mathcal{F}$ be a $\sigma$-algebra, $\mathcal{F} \subset \mathcal{F}_0$. The conditional expectation of $X$ given $\mathcal{F}$, $E[X|\mathcal{F}]$ is any random variable $Y$ such that

1. $Y \in \mathcal{F}$, that is, is $\mathcal{F}$ measurable

2. For all $A \in \mathcal{F}$, $\int_A XdP = \int_A YdP$. 
Lemma

If $Y$ is a conditional expectation of integrable variable $X$ then $Y$ is integrable.

Proof.

Let $A = \{ Y > 0 \} \in \mathscr{F}$. Then

$$
\int_A YdP = \int_A XdP \leq \int_A |X|dP
$$

$$
\int_{A^c} -YdP = \int_{A^c} -XdP \leq \int_{A^c} |X|dP.
$$

Thus $E[|Y|] \leq E[|X|]$. 

**Lemma**

Let \( X \) be an integrable random variable on probability space \((\Omega, \mathcal{F}_0, \text{Prob})\), with \( \sigma \)-field \( \mathcal{F} \subset \mathcal{F}_0 \), and let \( Y \) and \( Y' \) be two conditional expectations of \( X \) given \( \mathcal{F} \). Then \( Y = Y' \) \( \mathcal{F} \)-a.s.
Proof.

For each set $A \in \mathcal{F}$, $\int_A YdP = \int_A Y'dP$. Given $\epsilon > 0$, let $A = \{Y - Y' \geq \epsilon\}$. One finds

$$0 = \int_A X - XdP = \int_A Y - Y'dP \geq \epsilon \text{Prob}(A).$$
Let $X$ be an integrable random variable on probability space $(\Omega, \mathcal{F}_0, \text{Prob})$, and let $\mathcal{F} \subseteq \mathcal{F}_0$ be a $\sigma$-algebra. Then there exists $Y = \mathbb{E}[X|\mathcal{F}]$. 
Proof.

- By splitting $X$ into its positive and negative parts, we may assume that $X \geq 0$.
- Let $\mu = \text{Prob}$ and let $\nu$ be the measure on $\mathcal{F}$ defined by
  
  $$\nu(A) = \int_A XdP, \quad A \in \mathcal{F}.$$  

- By the definition of the integral, $\nu \ll \mu$.
- Let $Y = \frac{d\nu}{d\mu}$ be the Radon-Nikodym derivative of $\nu$ with respect to $\mu$, which is $\mathcal{F}$-measurable. We have, for $A \in \mathcal{F}$,

  $$\int_A XdP = \nu(A) = \int_A YdP.$$  

Stein’s method of Poisson Approximation

- Stein has given a general method of proving limit theorems via a perturbative method which avoids the use of characteristic functions and handles dependence.

- The following discussion of Poisson Approximation is based on the article ‘Two moments suffice for Poisson approximations: the Chen-Stein method’ by R. Arratia, L. Goldstein, L. Gordon.
Set-up

- Let \( I \) be an arbitrary index set, and for \( \alpha \in I \), let \( X_\alpha \) be a Bernoulli random variable with

\[
p_\alpha = \text{Prob}(X_\alpha = 1) = 1 - \text{Prob}(X_\alpha = 0) > 0.
\]

- Set

\[
W = \sum_{\alpha \in I} X_\alpha, \quad \lambda = \mathbb{E}[W] = \sum_{\alpha \in I} p_\alpha, \quad \lambda \in (0, \infty).
\]
For $\alpha \in I$, let $B_\alpha \subset I$, $\alpha \in B_\alpha$ be a ‘neighborhood of dependence.’

Set

\[
b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta
\]

\[
b_2 = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} p_{\alpha \beta}, \quad p_{\alpha \beta} = E[X_\alpha X_\beta]
\]

\[
b_3 = \sum_{\alpha \in I} s_\alpha.
\]

\[
s_\alpha = E \left[ \left\| E \left[ X_\alpha - p_\alpha \right| \sigma \left( X_\beta : \beta \in I - B_\alpha \right) \right| \right].
\]
Recall the definition of the total variation norm.

**Definition**

If $Z, W$ are two $\mathbb{Z}_{\geq 0}$ valued random variables with distributions (laws) $\mathcal{L}(Z), \mathcal{L}(W)$. The total variation distance between $\mathcal{L}(Z)$ and $\mathcal{L}(W)$ is

\[
\|\mathcal{L}(Z) - \mathcal{L}(W)\|_{TV} = \frac{1}{2} \sup_{\|h\|_{\infty} = 1} |E[h(W)] - E[h(Z)]| = \sup_{A \subset \mathbb{Z}^+} |\text{Prob}(W \in A) - \text{Prob}(Z \in A)|.
\]
The following theorem is due to Chen.

**Theorem**

Let $W$ be the number of occurrences of dependent events, and let $b_1, b_2, b_3$ be as in the set-up. Let $Z$ be a Poisson$(\lambda)$ random variable. Then

$$\| \mathcal{L}(W) - \mathcal{L}(Z) \|_{TV} \leq b_1 + b_2 + b_3.$$
Stein’s operators

Let $\lambda$ be a parameter, let $Z \sim \text{Poisson}(\lambda)$ and define linear operators $S, T$ on functions on $\mathbb{Z}_{\geq 0}$ by

$$Tf(w) = wf(w) - \lambda f(w + 1)$$

$$Sf(w + 1) = -\frac{E[f(Z)1(Z \leq w)]}{\lambda \text{Prob}(Z = w)}, \quad Sf(0) = 0.$$
Stein’s operators

Lemma

$T$ and $S$ are inverse, in the sense that $TSf = f$. 
Stein’s operators

Proof.

We have, for $x \neq 0$,

$$TSf(x) = xSf(x) - \lambda Sf(x + 1)$$

$$= xSf(x) + \frac{E[h(Z\mathbf{1}_{(Z \leq x)})]}{\text{Prob}(Z = x)}$$

$$= -\frac{x E[f(Z) \mathbf{1}_{(Z \leq x - 1)}]}{\lambda \text{Prob}(Z = x - 1)} + \frac{E[f(Z) \mathbf{1}_{(Z \leq x)}]}{\text{Prob}(Z = x)}$$

$$= f(x)$$

For $x = 0$, $xSf(x) = 0$, the result is the same. $\square$
Lemma

Let $\lambda$ be a parameter, and let $Z$ be a $\mathbb{Z}_{\geq 0}$ valued random variable. $Z \sim \text{Poisson}(\lambda)$ if and only if for all bounded $f$,

$$E[Tf(Z)] = 0.$$
Stein’s criterion

**Proof.**

- To check the necessity, write

\[
E[Tf(Z)] = e^{-\lambda} \sum_{n \geq 0} Tf(n) \frac{\lambda^n}{n!}
\]

\[
= e^{-\lambda} \sum_{n \geq 0} (nf(n) - \lambda f(n + 1)) \frac{\lambda^n}{n!}
\]

\[
= e^{-\lambda} \sum_{n \geq 1} (f(n) - f(n)) \frac{\lambda^n}{(n-1)!} = 0.
\]
Stein’s criterion

Proof.

To prove the sufficiency, set \( f(x) = 1_{(x=n)} \) for \( n = 1, 2, \ldots \) to obtain

\[
\text{Prob}(Z = n - 1) = \frac{n}{\lambda} \text{Prob}(Z = n).
\]

The result follows.
Define $\Delta f(n) = f(n + 1) - f(n)$.

**Lemma**

Suppose that $\forall w \geq 0$, $h(w) \in [0, 1]$ and $f = S(h(\cdot) - \mathbb{E}[h(Z)])$. Then

$$\|\Delta f\|_\infty \leq \frac{1 - e^{-\lambda}}{\lambda} \quad \text{and} \quad \|f\|_\infty \leq \min \left(1, \frac{1.4}{\lambda^{1/2}}\right).$$

Furthermore, if $h(w) = 1(w = 0) - e^{-\lambda}$ then $\|f\|_\infty = \frac{1 - e^{-\lambda}}{\lambda}$. 
Bounding the Stein operator

Proof.

Observe

\[
f(m + 1) = \frac{E[h(Z)] \text{Prob}(Z \leq m)}{\lambda \text{Prob}(Z = m)} - \frac{E[h(Z)1(Z \leq m)]}{\lambda \text{Prob}(Z = m)}
\]

\[
= \frac{E[h(Z)1(Z > m)] \text{Prob}(Z \leq m)}{\lambda \text{Prob}(Z = m)}
\]

\[
- \frac{E[h(Z)1(Z \leq m)] \text{Prob}(Z > m)}{\lambda \text{Prob}(Z = m)}.
\]

Hence \(|f(m + 1)| \leq \frac{\text{Prob}(Z \leq m) \text{Prob}(Z > m)}{\lambda \text{Prob}(Z = m)}|.

\]
Proof.

For $m < \lambda$,

$$|f(m + 1)| \leq \frac{\text{Prob}(Z \leq m)}{\lambda \text{Prob}(Z = m)} = \frac{1}{\lambda} \sum_{j=0}^{m} \frac{m!}{\lambda^j (m - j)!}$$

$$\leq \frac{1}{\lambda} \sum_{j=0}^{m} \left( \frac{m}{\lambda} \right)^j \leq (\lambda - m)^{-1}.$$

Hence $|f(m)| \leq 1$ if $m \leq \lambda$. 
Bounding the Stein operator

**Proof.**

- For $m \geq \lambda - 3$

$$|f(m + 1)| \leq \frac{\text{Prob}(Z > m)}{\lambda \text{Prob}(Z = m)} = \sum_{j=0}^{\infty} \frac{\lambda^j m!}{(m + 1 + j)!}$$

$$\leq \frac{1}{m + 1} \left[ 1 + \frac{\lambda}{m + 2} \sum_{j=0}^{\infty} \left( \frac{\lambda}{m + 3} \right)^j \right]$$

$$= \frac{(m + 2)(m + 3) + \lambda}{(m + 1)(m + 2)(m + 3 - \lambda)}.$$ 

This restricts bounding $|f(m)| < 1$ to a finite check, which we’ll ignore.
Bounding the Stein operator

Proof.

Using \( \Pr(Z \leq m) \Pr(Z > m) \leq \frac{1}{4} \) and Stirling’s approximation

\[
|f(m + 1)| \leq \frac{1}{4\lambda \Pr(Z = m)}
\]

\[
\leq \frac{\sqrt{2\pi}}{4\lambda^{\frac{1}{2}}} \left( \frac{m}{\lambda} \right)^{m + \frac{1}{2}} \exp \left( \lambda - m + \frac{1}{12m} \right)
\]

\[
\leq \frac{\sqrt{2\pi}}{4} \lambda^{-\frac{1}{2}} \exp \left( \frac{(m - \lambda)(m - \lambda + \frac{1}{2})}{\lambda} + \frac{1}{12m} \right).
\]

Using this for \(|\lambda - m| \leq \lambda^{\frac{1}{2}}\) and the previous inequalities otherwise obtains the bound \( |f(m + 1)| \leq \frac{c}{\lambda^{2}}.\)
Bounding the Stein operator

Proof.

- Define $f_j$ by taking $h(x) = 1(x = j)$. Hence

$$f_j(m + 1) = \begin{cases} 
\lambda^{j-m-1} \frac{m!}{j!} \text{Prob}(Z > m) & m \geq j \\
-\lambda^{j-m-1} \frac{m!}{j!} \text{Prob}(Z \leq m) & m < j 
\end{cases}.$$ 

- One easily checks that $f_j$ is positive and decreasing in $m \geq j + 1$ and is negative and decreasing in $m \leq j$.

- The only positive value of $f_j(m + 1) - f_j(m)$ is

$$f_j(j + 1) - f_j(j) = \frac{e^{-\lambda}}{\lambda} \left[ \sum_{r=j+1}^{\infty} \frac{\lambda^r}{r!} + \sum_{r=1}^{j} \frac{\lambda^r r}{r! j} \right]$$

$$\leq \frac{e^{-\lambda}}{\lambda} (e^\lambda - 1) = \frac{1 - e^{-\lambda}}{\lambda}.$$
Bounding the Stein operator

Proof.

- Writing the general $f$ as $f = \sum_j h(j) f_j$ proves

  $$f(m + 1) - f(m) \leq f_m(m + 1) - f_m(m) \leq \frac{1 - e^{-\lambda}}{\lambda}.$$ 

- This last calculation contains the claim that $\|f_0\| = \frac{1 - e^{-\lambda}}{\lambda}$ as this is the value at 1.
Proof of Stein’s Poisson approximation theorem.

- Let $h$ be given with $\|h\|_\infty = 1$ and let $Z \sim \text{Poisson}(\lambda)$.
- Let $\overline{h}(\cdot) = h(\cdot) - \mathbb{E}[h(Z)]$, $f = S\overline{h}$ and $Tf = \overline{h}$, so

$$\mathbb{E}[Tf(W)] = \mathbb{E}[h(W) - h(Z)].$$
Proof of Stein’s Poisson approximation

Proof of Stein’s Poisson approximation theorem.

- Let \( V_\alpha = \sum_{\beta \in I - B_\alpha} X_\beta \) and \( W_\alpha = W - X_\alpha \). We have
  \[ X_\alpha f(W) = X_\alpha f(W_\alpha + 1) \]
  \[ f(W_\alpha + 1) - f(W + 1) = X_\alpha [f(W_\alpha + 1) - f(W_\alpha + 2)] \]

- Calculate

\[
E[h(W) - h(Z)] = E[Wf(W) - \lambda f(W + 1)]
= \sum_{\alpha \in I} E[X_\alpha f(W) - p_\alpha f(W + 1)]
= \sum_{\alpha \in I} E[p_\alpha f(W_\alpha + 1) - p_\alpha f(W + 1)]
+ \sum_{\alpha \in I} E[X_\alpha f(W_\alpha + 1) - p_\alpha f(W_\alpha + 1)]
\]
Proof of Stein’s Poisson approximation

Proof of Stein’s Poisson approximation theorem.

- Calculate further

\[
E[h(W) - h(Z)] = \sum_{\alpha \in I} E[p_{\alpha} X_{\alpha} \left( f(W_{\alpha} + 1) - f(W_{\alpha} + 2) \right)] \\
+ \sum_{\alpha \in I} E[(X_{\alpha} - p_{\alpha}) \left( f(W_{\alpha} + 1) - f(V_{\alpha} + 1) \right)] \\
+ \sum_{\alpha \in I} E[(X_{\alpha} - p_{\alpha}) f(V_{\alpha} + 1)].
\]

- The first term may be bounded by \( \|\Delta f\|_{\infty} \sum_{\alpha \in I} p_{\alpha}^2 \).
Proof of Stein’s Poisson approximation theorem.

To bound \( \sum_{\alpha \in I} E \left[ (X_\alpha - p_\alpha) \left[ f(W_\alpha + 1) - f(V_\alpha + 1) \right] \right] \), write \( E \left[ (X_\alpha - p_\alpha) \left[ f(W_\alpha + 1) - f(V_\alpha + 1) \right] \right] \) as a telescoping sum of \( |B_\alpha| - 1 \) terms of the form

\[
E \left[ (X_\alpha - p_\alpha)(f(U + X_\beta) - f(U)) \right] \\
= E \left[ (X_\alpha - p_\alpha)X_\beta(f(U + 1) - f(U)) \right] \\
= E[X_\alpha X_\beta \Delta f(U)] - E[p_\alpha X_\beta \Delta f(U)] \\
\leq \| \Delta f \|_\infty (p_{\alpha\beta} + p_\alpha p_\beta).
\]

Thus the second term is bounded by

\[
\| \Delta f \|_\infty \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} (p_{\alpha\beta} + p_\alpha p_\beta).
\]
Proof of Stein’s Poisson approximation theorem.

- The third term is bounded by

\[
\left| \sum_{\alpha \in I} \mathbb{E}[(X_\alpha - p_\alpha)f(V_\alpha + 1)] \right| 
\leq \| f \|_\infty \sum_{\alpha \in I} \mathbb{E} \left[ \mathbb{E} \left[ |X_\alpha - p_\alpha| \sum_{\beta \in I - B_\alpha} X_\beta \right] \right] = \| f \|_\infty b'_3.
\]

- This completes the proof.
A random graph problem

Example

- On the hypercube \( \{0, 1\}^n \), assume each of the \( n2^{n-1} \) edges is assigned a random direction by tossing a fair coin, and let \( W \) be the number of vertices at which all \( n \) edges point inward.

- Let \( I \) be the set of all \( 2^n \) vertices, and \( X_\alpha \) the indicator that vertex \( \alpha \) has all edges pointing inward. Thus \( p_\alpha = 2^{-n} \). Set \( \lambda = 1 \), \( Z = \text{Poisson}(1) \).

- \( B_\alpha = \{ \beta : |\alpha - \beta| \leq 1 \} \).
A random graph problem

Example

- Calculate

\[ b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta = |I|(n + 1)2^{-2n} = \frac{n + 1}{2^n}. \]

- Calculate

\[ b_2 = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} \mathbb{E}[X_\alpha X_\beta] = 0, \]

since the events \( \{X_\alpha = 1\} \) and \( \{X_\beta = 1\} \) are mutually exclusive.

- \( b_3 = 0 \) since \( X_\alpha \) is independent of \( \sigma(X_\beta : \beta \in I - B_\alpha) \).

- \( \|\mathcal{L}(W) - \mathcal{L}(Z)\|_{TV} \leq (n + 1)2^{-n} \).
Example

- Suppose \( n \) balls (people) are uniformly and independently distributed into \( d \) boxes (days of the year). We seek an estimate for the probability that at least one box contains \( k \) or more balls for \( k = 2, 3, 4, ..., \).

- Let \( I = \{ \alpha \subset \{1, 2, 3, ..., n\} : |\alpha| = k \} \), and let \( X_\alpha \) be the event that each ball in \( \alpha \) goes into the same box.

- Set \( W = \sum_{\alpha \in I} X_\alpha \), \( p_\alpha = \operatorname{Prob}(X_\alpha = 1) = d^{1-k} \), \( \lambda = \binom{n}{k} d^{1-k} \), and \( Z \sim \text{Poisson}(\lambda) \).

- The goal is to approximate \( W \Rightarrow Z \) as \( n \to \infty \). To do so, we assume that \( \lambda \) is held essentially fixed, so that \( d \approx n^{\frac{k}{k-1}} \) as \( n \to \infty \).
The birthday problem

Example

- \( B_\alpha = \{ \beta \in I : \alpha \cap \beta \neq \emptyset \} \). Hence \( X_\alpha \) is independent of \( \sigma(X_\beta : \beta \in B_\alpha) \), so \( b_3 = 0 \).

- One has \( |B_\alpha| = \binom{n}{k} - \binom{n-k}{k} \), so

\[
b_1 = p_\alpha^2 |I||B_\alpha| = \lambda^2 \frac{|B_\alpha|}{|I|} = \lambda^2 \left( 1 - \frac{n-k}{n} \frac{n-k-1}{n-1} \cdots \frac{n-2k+1}{n-k+1} \right) < \lambda^2 \left( 1 - \left( 1 - \frac{k^2}{n-k+1} \right) \right) = \frac{\lambda^2 k^2}{n-k+1}.
\]

- For \( \lambda \) and \( k \) fixed, this tends to 0 with increasing \( n \).
The birthday problem

Example

- For fixed $\alpha$,

$$
\sum_{\beta \in B_\alpha \setminus \{\alpha\}} \mathbb{E}[X_\alpha X_\beta] = \sum_{j=1}^{k-1} \binom{k}{j} \binom{n-k}{k-j} d^{1+j-2k}.
$$

When $\frac{d}{n}$ is large, the dominant term comes from $j = k - 1$, so that

$$
b_2 \lesssim k \binom{n}{k} (n-k) d^{-k} = k \lambda \frac{n-k}{d}.
$$

- Recalling $d \asymp n^{\frac{k}{k-1}}$, $b_2 \to 0$. 
The longest perfect head run

Example

- Let $0 < p < 1$ and $Y_1, Y_2, \ldots$ be an i.i.d. sequence
  $p = \text{Prob}(Y_i = 1) = 1 - \text{Prob}(Y_i = 0)$.
- Let $R_n$ be the length of the longest consecutive run of heads starting within the first $n$ tosses.
- Let $I = \{1, 2, \ldots, n\}$.
- Fix positive integer $t$ and set $X_1 = Y_1 Y_2 \cdots Y_t$, and for $2 \leq \alpha \leq n$,
  \[ X_\alpha = (1 - Y_{\alpha-1}) Y_\alpha Y_{\alpha+1} \cdots Y_{\alpha+t-1}. \]
The longest perfect head run

Example

- Let \( B_\alpha = \{ \beta \in I : |\alpha - \beta| \leq t \} \).
- One has \( b_3 = 0 \) by independence, and \( b_2 = 0 \), since for \( \beta \neq \alpha \), \( \beta \in B_\alpha \), the events \( \{X_\alpha = 1\} \) and \( \{X_\beta = 1\} \) are exclusive.
- We have
  \[
  b_1 < p^{2t} (1 + 2t(1 - p)) + n(2t + 1)p^{2t}(1 - p)^2
  \]
  and
  \[
  \lambda = \lambda(n, t) = E[W] = p^t [(n - 1)(1 - p) + 1].
  \]
- Since \( \{R_n < t\} = \{W = 0\} \), with \( Z \sim \text{Poisson}(\lambda) \)
  \[
  \left| \text{Prob}(R_n < t) - e^{-\lambda(n,t)} \right| \leq \|W - Z\|_{TV} \leq b_1 \min(1, \lambda^{-1}).
  \]
  Keeping \( \lambda \) fixed as \( n \to \infty \), \( b_1 \to 0 \).
Our discussion of Stein’s method of normal approximation is taken from Stein’s 1986 monograph “Approximate computation of expectations.” For the remainder of the lecture $Z$ is a standard normal random variable.
Stein’s operators

- Let $\mathcal{X}$ be the space of all piecewise continuous $h : \mathbb{R} \to \mathbb{R}$ such that, for all $k > 0$
  \[
  \int_{-\infty}^{\infty} |x|^k |h(x)| e^{-\frac{x^2}{2}} \, dx < \infty.
  \]

- Let $\mathcal{F}$ be the space of all continuous and piecewise continuously differentiable $f : \mathbb{R} \to \mathbb{R}$ with $f' \in \mathcal{X}$.

- Define operators $T : \mathcal{F} \to \mathcal{X}$, $Tf(w) = f'(w) - wf(w)$ and $U : \mathcal{X} \to \mathcal{F}$, 
  \[
  Uh(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} [h(x) - \mathbb{E}[h(Z)]] e^{-\frac{x^2}{2}} \, dx.
  \]
Lemma

For all $f \in \mathcal{F}$, $Tf \in \mathcal{H}$. For all $h \in \mathcal{H}$, $Uh \in \mathcal{F}$. Let $Z$ be standard normal. For $h \in \mathcal{H}$, $T \circ Uh(w) = h(w) - \mathbb{E}[h(Z)]$. 
Stein’s operators

Proof.

For \( f \in \mathcal{F} \) and \( k > 0 \),

\[
\int_{0}^{\infty} w^{k+1} |f(w) - f(0)| e^{-\frac{w^2}{2}} \, dw = \int_{0}^{\infty} w^{k+1} \left| \int_{0}^{w} f'(x) \, dx \right| e^{-\frac{w^2}{2}} \, dw
\]

\[
\leq \int_{0}^{\infty} |f'(x)| \int_{x}^{\infty} w^{k+1} e^{-\frac{w^2}{2}} \, dw \, dx
\]

\[
\leq \int_{0}^{\infty} |f'(x)| C(1 + |x|^k) e^{-\frac{x^2}{2}} \, dx < \infty.
\]

Similarly \( \int_{-\infty}^{0} |w|^{k+1} |f(w) - f(0)| e^{-\frac{w^2}{2}} \, dw < \infty \). Hence \( w \mapsto wf(w) \in \mathcal{X}^\prime \), so \( Tf \in \mathcal{X}^\prime \).
Stein’s operators

Proof.

Given \( h \in \mathcal{X} \), \( k \geq 0 \),

\[
\int_0^\infty w^{k+1} |Uh(w)| e^{-\frac{w^2}{2}} \, dw
\]

\[
\leq \int_0^\infty w^{k+1} \int_w^\infty |h(x) - E[h(Z)]| e^{-\frac{x^2}{2}} \, dx \, dw
\]

\[
= \int_0^\infty |h(x) - E[h(Z)]| \frac{x^{k+2}}{k+2} e^{-\frac{x^2}{2}} \, dx < \infty.
\]

Similarly \( \int_{-\infty}^0 |w|^{k+1} |Uh(w)| e^{-\frac{w^2}{2}} \, dw < \infty \), so that \( w \mapsto w Uh(w) \in \mathcal{X} \).
Proof.

Differentiate

\[ Uh(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} [h(x) - E[h(Z)]] e^{-\frac{x^2}{2}} \, dx \]

to obtain \((Uh)'(w) - w(Uh)(w) = h(w) - E[h(Z)]\).
Lemma

In order that the real random variable $W$ has a standard normal distribution, it is necessary and sufficient that, for all continuous and piecewise continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ with $E[|f'(Z)|] < \infty$, $Z$ standard normal, we have

$$E[f'(W)] = E[Wf(W)].$$
Proof of necessity.

Let $W$ have a standard normal distribution. Then

$$E[f'(W)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(w) e^{-\frac{w^2}{2}} \, dw$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(w) \left( \int_{-\infty}^{w} (-z) e^{-\frac{z^2}{2}} \, dz \right) \, dw$$

$$+ \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} f'(w) \left( \int_{w}^{\infty} ze^{-\frac{z^2}{2}} \, dz \right) \, dw$$
Stein’s method of normal approximation

Proof of necessity.

\[
\begin{align*}
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \left( \int_{z}^{0} f'(w) \, dw \right) (-z) e^{-\frac{z^2}{2}} \, dz \\
+ \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \left( \int_{0}^{z} f'(w) \, dw \right) ze^{-\frac{z^2}{2}} \, dz \\
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[ f(z) - f(0) \right] ze^{-\frac{z^2}{2}} \, dz = \mathbb{E}[Wf(W)].
\end{align*}
\]
Proof of sufficiency.

- Given \( w_0 \in \mathbb{R} \), let \( f_{w_0} = U1(w \leq w_0) \).
- Hence
  \[
  E[f'_{w_0}(W) - Wf_{w_0}(W)] = E[1(W \leq w_0) - E[1(Z \leq w_0)]] \\
  = \text{Prob}(W \leq w_0) - \text{Prob}(Z \leq w_0).
  \]

Hence, if this is zero for all \( w_0 \) then \( W \) has a standard normal distribution.
Explicit estimates

The special functions $f_{w_0} = U_1(w \leq w_0)$ are given by

$$f_{w_0}(w) = \begin{cases} \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(w)[1 - \Phi(w_0)] & w \leq w_0 \\ \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(w_0)[1 - \Phi(w)] & w \geq w_0 \end{cases}$$

where $\Phi(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w} e^{-\frac{x^2}{2}} dx$. 
Explicit estimates

Lemma

The functions $f_{w_0}$ satisfies

$$0 < f_{w_0}(w) \leq \frac{\sqrt{2\pi}}{4}, \quad |w f_{w_0}(w)| < 1, \quad |f'_{w_0}(w)| < 1$$

for all real $w_0, w$.

We omit this explicit calculation.
Lemma

For bounded absolutely continuous \( h : \mathbb{R} \to \mathbb{R} \),

\[
\|Uh\|_{\infty} \leq \sqrt{\frac{\pi}{2}} \|h - E[h(Z)]\|_{\infty}
\]
\[
\|Uh'\|_{\infty} \leq 2\|h - E[h(Z)]\|_{\infty}
\]
\[
\|Uh''\|_{\infty} \leq 2\|h'\|_{\infty}.
\]
Proof.

For \( w \leq 0 \),

\[
|Uh(w)| \leq \left[ \sup_{x \leq 0} |h(x) - E[h(Z)]| \right] e^{\frac{w^2}{2}} \int_{-\infty}^{w} e^{-\frac{x^2}{2}} \, dx,
\]

and, for \( w \geq 0 \),

\[
|Uh(w)| \leq \left[ \sup_{x \geq 0} |h(x) - E[h(Z)]| \right] e^{\frac{w^2}{2}} \int_{w}^{\infty} e^{-\frac{x^2}{2}} \, dx.
\]

The first claim follows since the maximum of \( e^{\frac{w^2}{2}} \int_{-\infty}^{w} e^{-\frac{x^2}{2}} \, dx \) in \( w \leq 0 \) is attained at 0.
Proof.

For $w \geq 0$ use

$$(Uh)'(w) = h(w) - E[h(Z)] - we^{\frac{w^2}{2}} \int_w^\infty [h(x) - E[h(Z)]] e^{-\frac{x^2}{2}} \, dx.$$ 

Hence

$$\sup_{w \geq 0} |(Uh)'(w)| \leq \left[ \sup_{w \geq 0} |h - E[h(Z)]| \right] \left[ 1 + \sup_{w \geq 0} we^{\frac{w^2}{2}} \int_w^\infty e^{-\frac{x^2}{2}} \, dx \right]$$

$$\leq 2 \sup_{w \geq 0} |h - E[h(Z)]|.$$ 

The bound for $w \leq 0$ is similar.
Bounds for the Stein operator

Proof.

The bound for $\|(Uh)''\|_\infty$ in terms of $\|h'\|_\infty$ is a more involved computation, which we omit.
Exchangeable pairs

Definition

A pair \((X, X')\) of random variables on a probability space \((\Omega, \mathcal{B}, \text{Prob})\) is called an exchangeable pair if, for all \(B, B'\),

\[
\text{Prob}(X \in B, X' \in B') = \text{Prob}(X \in B', X' \in B).
\]
The following lemma is key.

Lemma

Let $0 < \lambda < 1$ and let $(W, W')$ be an exchangeable pair of real random variables, such that

$$E[W'|W] = (1 - \lambda)W.$$ 

Let $h: \mathbb{R} \rightarrow \mathbb{R}$ be a bounded continuous function with bounded piecewise continuous derivative $h'$.

$$E[h(W)] = E[h(Z)] + E\left[(Uh)'(W) \left[1 - \frac{1}{2\lambda} E[(W' - W)^2|W]\right]\right] + \frac{1}{2\lambda} \int E\left[(W - W') \left(z - \frac{W + W'}{2}\right) [1(z \leq W') - 1(z \leq W)]\right] d(Uh)'(z)$$

Bob Hough

Math 639: Lecture 7

February 16, 2017 55 / 61
Stein’s method for normal approximation

Proof.

From the identity

\[ 0 = E \left[ Wf(W) - \frac{1}{2\lambda} (W' - W)(f(W') - f(W)) \right] \]

\[ = E[Wf(W) - f'(W)] + E \left[ f'(W) - \frac{1}{2\lambda} (W' - W)(f(W') - f(W)) \right] \]

\[ = E[h(Z)] - E[h(W)] + E[f'(W)] - \frac{1}{2\lambda} E[(W' - W)(f(W') - f(W))] \]

obtain

\[ E[h(W)] = E[h(Z)] + E[f'(W)] - \frac{1}{2\lambda} E[(W - W')(f(W) - f(W'))]. \]
Proof.

Rewrite part of the last line as

\[
E \left[ f'(W) - \frac{1}{2\lambda} (W' - W)(f(W') - f(W)) \right] \\
= E \left[ f'(W) \left[ 1 - \frac{1}{2\lambda} E \left[ (W' - W)^2 | W \right] \right] \right] \\
- \frac{1}{2\lambda} E \left[ (W' - W) \left[ f(W') - f(W) - (W' - W)f'(W) \right] \right].
\]
Stein’s method for normal approximation

Proof.
Write

\[ f(W') - f(W) - (W' - W)f'(W) = \int_{W}^{W'} (W' - y)f''(y) dy \]

\[ = \int (W' - y)[1(y \leq W') - 1(y \leq W)]f''(y) dy. \]

Take expectation and use the exchangeability of \( W, W' \) to obtain the claim.
**Theorem**

Let $h$ be a bounded continuous function with bounded piecewise continuous derivative $h'$. Let $W, W'$ as in the previous lemma. Then

$$
|E[h(W)] - E[h(Z)]| \leq \frac{1}{4\lambda} \|h'\|_{\infty} E[|W' - W|^3] \\
+ 2\|h - E[h(Z)]\|_{\infty} \sqrt{E \left[ \left( 1 - \frac{1}{2\lambda} E[(W' - W)^2|W] \right)^2 \right]}.
$$

and for all real $w_0$,

$$
|\text{Prob}(W \leq w_0) - \Phi(w_0)| \leq 2\sqrt{E \left[ \left( 1 - \frac{1}{2\lambda} E[(W' - W)^2|W] \right)^2 \right]} \\
+ (2\pi)^{-\frac{1}{4}} \sqrt{\frac{1}{\lambda} E[|W' - W|^3]}.
$$
Stein’s method for normal approximation

Proof.

\[
E[h(W)] - E[h(Z)] = E \left[ (Uh)'(W) \left( 1 - \frac{1}{2\lambda} E[(W' - W)^2 | W] \right) \right] + \\
\frac{1}{2\lambda} \int E \left[ (W - W') \left( z - \frac{W + W'}{2} \right) \left[ 1(z \leq W') - 1(z \leq W) \right] \right] \\
\times (Uh)''(z) \, dz
\]

So

\[
\left| E[h(W)] - E[h(Z)] \right| \leq \| (Uh)' \|_{\infty} E \left[ 1 - \frac{1}{2\lambda} E[(W - W')^2 | W] \right] \\
+ \| (Uh)'' \|_{\infty} \frac{1}{2\lambda} E \left[ \int_{\min(W, W')}^{\max(W, W')} |W - W'| \left| z - \frac{W + W'}{2} \right| \, dz \right].
\]
Proof.

Recall $\|(Uh)'\|_\infty \leq 2\|h - E[h(Z)]\|_\infty$ and $\|(Uh)''\|_\infty \leq 2\|h'\|_\infty$. Hence

$$|E[h(W)] - E[h(Z)]| \leq$$

$$2\|h - E[h(Z)]\|_\infty \sqrt{E\left[\left(1 - \frac{1}{2\lambda} E[(W - W')^2|W]\right)^2\right]}$$

$$+ 2\|h'\|_\infty \frac{1}{2\lambda} E\left[\frac{|W - W'|^3}{4}\right].$$

This proves the first bound.

To prove the second, bound $1(w \leq w_0)$ from above and below using piece-wise linear functions. We omit the details.