Math 639: Lecture 7

Stein's method

Bob Hough

February 16, 2017

Conditional expectation

Definition

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

1. $Y \in \mathscr{F}$, that is, is $\mathscr{F}$ measurable
2. For all $A \in \mathscr{F}$, $\int_A X dP = \int_A Y dP$.

Conditional expectation

Lemma

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

Proof.

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

$$\int_A Y dP = \int_A X dp \leqslant \int_A |X| dP$$
$$\int_{A^c} -Y dP = \int_{A^c} -X dP \leqslant \int_{A^c} |X| dP.$$

Thus $\mathrm{E}[|Y|] \leqslant \mathrm{E}[|X|]$. $\square$

Conditional expectation

Lemma

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

Conditional expectation

Proof.

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

$$0 = \int_A X - X dP = \int_A Y - Y' dP \geqslant \epsilon \operatorname{Prob}(A).$$

$\square$

Conditional expectation

Lemma

Let X be an integrable random variable on probability space $(\Omega, \mathscr{F}_0, \mathrm{Prob})$, and let $\mathscr{F} \subset \mathscr{F}_0$ be a σ-algebra. Then there exists $Y = \mathrm{E}[X|\mathscr{F}]$.

Proof.

- By splitting X into its positive and negative parts, we may assume that $X \geqslant 0$.
- Let $\mu = \text{Prob}$ and let ν be the measure on $\mathscr{F}$ defined by

$$\nu(A) = \int_A X dP, \qquad A \in \mathscr{F}.$$

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

$$\int_A X dP = \nu(A) = \int_A Y dP.$$

$\square$

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_α be a Bernoulli random variable with

$$p_\alpha = \mathrm{Prob}(X_\alpha = 1) = 1 - \mathrm{Prob}(X_\alpha = 0) > 0.$$

- Set

$$W = \sum_{\alpha \in I} X_\alpha, \qquad \lambda = \mathrm{E}[W] = \sum_{\alpha \in I} p_\alpha, \qquad \lambda \in (0, \infty).$$

Set-up

- For $\alpha \in I$, let $B_\alpha \subset I$, $\alpha \in B_\alpha$ be a 'neighborhood of dependence.'
- Set

$$\begin{aligned} b_1 &= \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta \\ b_2 &= \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} p_{\alpha\beta}, \qquad p_{\alpha\beta} = \mathrm{E}[X_\alpha X_\beta] \\ b_3 &= \sum_{\alpha \in I} s_\alpha. \end{aligned}$$

$$s_\alpha = \mathrm{E}\left[\left|\mathrm{E}\left[X_\alpha - p_\alpha \middle| \sigma\left(X_\beta : \beta \in I - B_\alpha\right)\right]\right|\right].$$

Set-up

Recall the definition of the total variation norm.

Definition

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

$$\begin{aligned}\|\mathscr{L}(Z) - \mathscr{L}(W)\|_{\mathrm{TV}} &= \frac{1}{2} \sup_{\|h\|_\infty = 1} |\mathrm{E}[h(W)] - \mathrm{E}[h(Z)]| \\ &= \sup_{A \subset \mathbb{Z}^+} \left| \mathrm{Prob}(W \in A) - \mathrm{Prob}(Z \in A) \right|.\end{aligned}$$

Stein's method of Poisson approximation

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 $\mathrm{Poisson}(\lambda)$ random variable. Then

$$\|\mathscr{L}(W) - \mathscr{L}(Z)\|_{\mathrm{TV}} \leqslant b_1 + b_2 + b_3.$$

Stein's operators

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

$$Tf(w) = wf(w) - \lambda f(w+1)$$
$$Sf(w+1) = -\frac{\mathrm{E}\left[f(Z)\mathbf{1}_{(Z \leqslant w)}\right]}{\lambda \,\mathrm{Prob}(Z = w)}, \qquad Sf(0) = 0.$$

Stein's operators

Lemma

T and S are inverse, in the sense that $TSf = f$.

Stein's operators

Proof.

We have, for $x \neq 0$,

$$\begin{aligned} TSf(x) &= xSf(x) - \lambda Sf(x+1) \\ &= xSf(x) + \frac{\mathrm{E}[h(Z\mathbf{1}_{(Z\leqslant x)})]}{\mathrm{Prob}(Z=x)} \\ &= -\frac{x\,\mathrm{E}[f(Z)\mathbf{1}_{(Z\leqslant x-1)}]}{\lambda\,\mathrm{Prob}(Z=x-1)} + \frac{\mathrm{E}[f(Z)\mathbf{1}_{(Z\leqslant x)}]}{\mathrm{Prob}(Z=x)} \\ &= f(x) \end{aligned}$$

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

Stein's criterion

Lemma

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

$$\mathrm{E}[Tf(Z)] = 0.$$

Stein's criterion

Proof.

- To check the necessity, write

$$\begin{aligned} E[Tf(Z)] &= e^{-\lambda} \sum_{n \geqslant 0} Tf(n) \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n \geqslant 0} (nf(n) - \lambda f(n+1)) \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n \geqslant 1} (f(n) - f(n)) \frac{\lambda^n}{(n-1)!} = 0. \end{aligned}$$

□

Proof.

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

$$\mathrm{Prob}(Z = n-1) = \frac{n}{\lambda}\mathrm{Prob}(Z = n).$$

The result follows. $\square$

Bounding the Stein operator

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

Lemma

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

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

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

Bounding the Stein operator

Proof.

- Observe

$$f(m+1) = \frac{\mathrm{E}[h(Z)]\,\mathrm{Prob}(Z \leqslant m)}{\lambda\,\mathrm{Prob}(Z=m)} - \frac{\mathrm{E}[h(Z)\mathbf{1}(Z \leqslant m)]}{\lambda\,\mathrm{Prob}(Z=m)}$$
$$= \frac{\mathrm{E}[h(Z)\mathbf{1}(Z > m)]\,\mathrm{Prob}(Z \leqslant m)}{\lambda\,\mathrm{Prob}(Z=m)}$$
$$- \frac{\mathrm{E}[h(Z)\mathbf{1}(Z \leqslant m)]\,\mathrm{Prob}(Z > m)}{\lambda\,\mathrm{Prob}(Z=m)}.$$

Hence $|f(m+1)| \leqslant \frac{\mathrm{Prob}(Z\leqslant m)\,\mathrm{Prob}(Z>m)}{\lambda\,\mathrm{Prob}(Z=m)}$.

□

Bounding the Stein operator

Proof.

- For $m < \lambda$,

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

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

Hence $|f(m)| \leqslant 1$ if $m \leqslant \lambda$.

$\square$

Bounding the Stein operator

Proof.

- For $m \geqslant \lambda - 3$

$$|f(m+1)| \leqslant \frac{\mathrm{Prob}(Z > m)}{\lambda\,\mathrm{Prob}(Z = m)} = \sum_{j=0}^{\infty} \frac{\lambda^j m!}{(m+1+j)!}$$
$$\leqslant \frac{1}{m+1}\left[1 + \frac{\lambda}{m+2}\sum_{j=0}^{\infty}\left(\frac{\lambda}{m+3}\right)^j\right]$$
$$= \frac{(m+2)(m+3)+\lambda}{(m+1)(m+2)(m+3-\lambda)}.$$

This restricts bounding $|f(m)| < 1$ to a finite check, which we'll ignore. □

Bounding the Stein operator

Proof.

- Using $\mathrm{Prob}(Z \leqslant m)\,\mathrm{Prob}(Z > m) \leqslant \frac{1}{4}$ and Stirling's approximation

$$\begin{aligned} |f(m+1)| &\leqslant \frac{1}{4\lambda\,\mathrm{Prob}(Z=m)} \\ &\leqslant \frac{\sqrt{2\pi}}{4\lambda^{\frac{1}{2}}}\left(\frac{m}{\lambda}\right)^{m+\frac{1}{2}} \exp\left(\lambda - m + \frac{1}{12m}\right) \\ &\leqslant \frac{\sqrt{2\pi}}{4}\lambda^{-\frac{1}{2}} \exp\left(\frac{(m-\lambda)(m-\lambda+\frac{1}{2})}{\lambda} + \frac{1}{12m}\right). \end{aligned}$$

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

□

Bounding the Stein operator

Proof.

- Define f_j by taking $h(x) = \mathbf{1}(x = j)$. Hence

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

- One easily checks that f_j is positive and decreasing in $m \geqslant j+1$ and is negative and decreasing in $m \leqslant j$.
- The only positive value of $f_j(m+1) - f_j(m)$ is

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

□

Bounding the Stein operator

Proof.

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

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

- This last calculation contains the claim that $\|f_0\| = \frac{1-e^{-\lambda}}{\lambda}$ as this is the value at 1.

$\square$

Proof of Stein's Poisson approximation

Proof of Stein's Poisson approximation theorem.

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

$$\mathrm{E}[Tf(W)] = \mathrm{E}[h(W) - h(Z)].$$

$\square$

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 $X_\alpha f(W) = X_\alpha f(W_\alpha + 1)$ and $f(W_\alpha + 1) - f(W + 1) = X_\alpha [f(W_\alpha + 1) - f(W_\alpha + 2)]$.
- Calculate

$$\begin{aligned} \mathrm{E}\left[h(W) - h(Z)\right] &= \mathrm{E}\left[Wf(W) - \lambda f(W+1)\right] \\ &= \sum_{\alpha \in I} \mathrm{E}\left[X_\alpha f(W) - p_\alpha f(W+1)\right] \\ &= \sum_{\alpha \in I} \mathrm{E}\left[p_\alpha f(W_\alpha + 1) - p_\alpha f(W+1)\right] \\ &\quad + \sum_{\alpha \in I} \mathrm{E}\left[X_\alpha f(W_\alpha + 1) - p_\alpha f(W_\alpha + 1)\right] \end{aligned}$$

$\square$

Proof of Stein's Poisson approximation

Proof of Stein's Poisson approximation theorem.

- Calculate further

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

- The first term may be bounded by $\|\Delta f\|_\infty \sum_{\alpha \in I} p_\alpha^2$.

$\square$

Proof of Stein's Poisson approximation

Proof of Stein's Poisson approximation theorem.

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

$$\begin{aligned}
&\mathrm{E}\left[(X_\alpha - p_\alpha)(f(U + X_\beta) - f(U))\right]\\
&= \mathrm{E}\left[(X_\alpha - p_\alpha)X_\beta(f(U+1) - f(U))\right]\\
&= \mathrm{E}[X_\alpha X_\beta \Delta f(U)] - \mathrm{E}[p_\alpha X_\beta \Delta f(U)]\\
&\leqslant \|\Delta f\|_\infty (p_{\alpha\beta} + p_\alpha p_\beta).
\end{aligned}$$

- 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

Proof of Stein's Poisson approximation theorem.

- The third term is bounded by

$$\left|\sum_{\alpha \in I} \mathrm{E}[(X_\alpha - p_\alpha) f(V_\alpha + 1)]\right|$$

$$\leqslant \|f\|_\infty \sum_{\alpha \in I} \mathrm{E}\left[\left|\mathrm{E}\left[X_\alpha - p_\alpha \middle| \sum_{\beta \in I - B_\alpha} X_\beta\right]\right|\right] = \|f\|_\infty b_3'.$$

- 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_α the indicator that vertex α has all edges pointing inward. Thus $p_\alpha = 2^{-n}$. Set $\lambda = 1$, $Z = \text{Poisson}(1)$.
- $B_\alpha = \{\beta : |\alpha - \beta| \leqslant 1\}$.

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

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

  since the events $\{X_\alpha = 1\}$ and $\{X_\beta = 1\}$ are mutually exclusive.
- $b_3 = 0$ since X_α is independent of $\sigma(X_\beta : \beta \in I - B_\alpha)$.
- $\|\mathscr{L}(W) - \mathscr{L}(Z)\|_{\mathrm{TV}} \leqslant (n+1)2^{-n}$.

The birthday problem

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_α be the event that each ball in α goes into the same box.
- Set $W = \sum_{\alpha \in I} X_\alpha$, $p_\alpha = \mathrm{Prob}(X_\alpha = 1) = d^{1-k}$, $\lambda = \binom{n}{k} d^{1-k}$ and $Z \sim \mathrm{Poisson}(\lambda)$.
- The goal is to approximate $W \Rightarrow Z$ as $n \to \infty$. To do so, we assume that λ is held essentially fixed, so that $d \asymp n^{\frac{k}{k-1}}$ as $n \to \infty$.

The birthday problem

Example

- $B_\alpha = \{\beta \in I : \alpha \cap \beta \neq \varnothing\}$. Hence X_α 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

$$\begin{aligned} b_1 &= p_\alpha^2 |I||B_\alpha| \\ &= \lambda^2 \frac{|B_\alpha|}{|I|} \\ &= \lambda^2 \left(1 - \frac{n-k}{n}\frac{n-k-1}{n-1}\cdots\frac{n-2k+1}{n-k+1}\right) \\ &< \lambda^2 \left(1 - \left(1 - \frac{k^2}{n-k+1}\right)\right) = \frac{\lambda^2 k^2}{n-k+1}. \end{aligned}$$

- For λ and k fixed, this tends to 0 with increasing n.

The birthday problem

Example

- For fixed α,

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

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 $2 \leqslant \alpha \leqslant n$,

$$X_\alpha = (1 - Y_{\alpha-1}) Y_\alpha Y_{\alpha+1} \cdots Y_{\alpha+t-1}.$$

The longest perfect head run

Example

- Let $B_\alpha = \{\beta \in I : |\alpha - \beta| \leqslant t\}$.
- 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

$$b_1 < p^{2t}\left(1 + 2t(1-p)\right) + n(2t+1)p^{2t}(1-p)^2$$

and

$$\lambda = \lambda(n,t) = \mathrm{E}[W] = p^t\left[(n-1)(1-p) + 1\right].$$

- Since $\{R_n < t\} = \{W = 0\}$, with $Z \sim \mathrm{Poisson}(\lambda)$

$$\left|\mathrm{Prob}(R_n < t) - e^{-\lambda(n,t)}\right| \leqslant \|W - Z\|_{\mathrm{TV}} \leqslant b_1 \min(1, \lambda^{-1}).$$

Keeping λ fixed as $n \to \infty$, $b_1 \to 0$.

Stein's method of normal approximation

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 $\mathscr{X}$ 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 $\mathscr{F}$ be the space of all continuous and piecewise continuously differentiable $f : \mathbb{R} \to \mathbb{R}$ with $f' \in \mathscr{X}$.
- Define operators $T : \mathscr{F} \to \mathscr{X}$, $Tf(w) = f'(w) - wf(w)$ and $U : \mathscr{X} \to \mathscr{F}$,
$$Uh(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} [h(x) - \mathrm{E}[h(Z)]] e^{-\frac{x^2}{2}} dx.$$

Stein's operators

Lemma

For all $f \in \mathscr{F}$, $Tf \in \mathscr{X}$. For all $h \in \mathscr{X}$, $Uh \in \mathscr{F}$. Let Z be standard normal. For $h \in \mathscr{X}$, $T \circ Uh(w) = h(w) - \mathrm{E}[h(Z)]$.

Stein's operators

Proof.

- For $f \in \mathscr{F}$ and $k > 0$,

$$\int_0^\infty w^{k+1}|f(w) - f(0)|e^{-\frac{w^2}{2}}dw = \int_0^\infty w^{k+1}\left|\int_0^w f'(x)dx\right| e^{-\frac{w^2}{2}}dw$$
$$\leqslant \int_0^\infty |f'(x)| \int_x^\infty w^{k+1}e^{-\frac{w^2}{2}}dwdx$$
$$\leqslant \int_0^\infty |f'(x)|C(1 + |x|^k)e^{-\frac{x^2}{2}}dx < \infty.$$

Similarly $\int_{-\infty}^0 |w|^{k+1}|f(w) - f(0)|e^{-\frac{w^2}{2}}dw < \infty$. Hence $w \mapsto wf(w) \in \mathscr{X}$, so $Tf \in \mathscr{X}$.

$\square$

Stein's operators

Proof.

- Given $h \in \mathscr{X}$, $k \geqslant 0$,

$$\begin{aligned} &\int_0^\infty w^{k+1}|Uh(w)|e^{-\frac{w^2}{2}}dw \\ &\leqslant \int_0^\infty w^{k+1}\int_w^\infty |h(x) - \mathrm{E}[h(Z)]|e^{-\frac{x^2}{2}}dxdw \\ &= \int_0^\infty |h(x) - \mathrm{E}[h(Z)]|\frac{x^{k+2}}{k+2}e^{-\frac{x^2}{2}}dx < \infty. \end{aligned}$$

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

□

Stein's operators

Proof.

- Differentiate

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

to obtain $(Uh)'(w) - w(Uh)(w) = h(w) - \mathrm{E}[h(Z)]$.

$\square$

Stein's method of normal approximation

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 $\mathrm{E}[|f'(Z)|] < \infty$, Z standard normal, we have

$$\mathrm{E}[f'(W)] = \mathrm{E}[Wf(W)].$$

Stein's method of normal approximation

Proof of necessity.

Let W have a standard normal distribution. Then

$$\begin{aligned}
\mathrm{E}[f'(W)] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(w) e^{-\frac{w^2}{2}} dw \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(w) \left(\int_{-\infty}^{w} (-z) e^{-\frac{z^2}{2}} dz \right) dw \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} f'(w) \left(\int_{w}^{\infty} z e^{-\frac{z^2}{2}} dz \right) dw
\end{aligned}$$

$\square$

Proof of necessity.

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \left(\int_{z}^{0} f'(w)dw \right) (-z) e^{-\frac{z^2}{2}} dz \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \left(\int_{0}^{z} f'(w)dw \right) z e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [f(z) - f(0)] z e^{-\frac{z^2}{2}} dz = \mathrm{E}[Wf(W)].
\end{aligned}$$

$\square$

Stein's method of normal approximation

Proof of sufficiency.

- Given $w_0 \in \mathbb{R}$, let $f_{w_0} = U\mathbf{1}(w \leqslant w_0)$.
- Hence

$$\begin{aligned} \mathrm{E}[f'_{w_0}(W) - Wf_{w_0}(W)] &= \mathrm{E}[\mathbf{1}(W \leqslant w_0) - \mathrm{E}[\mathbf{1}(Z \leqslant w_0)]] \\ &= \mathrm{Prob}(W \leqslant w_0) - \mathrm{Prob}(Z \leqslant w_0). \end{aligned}$$

Hence, if this is zero for all w_0 then W has a standard normal distribution.

$\square$

Explicit estimates

The special functions $f_{w_0} = U\mathbf{1}(w \leqslant w_0)$ are given by

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

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

$$0 < f_{w_0}(w) \leqslant \frac{\sqrt{2\pi}}{4}, \ |wf_{w_0}(w)| < 1, \ \left|f'_{w_0}(w)\right| < 1$$

for all real w_0, w.

We omit this explicit calculation.

Bounds for the Stein operator

Lemma

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

$$\|Uh\|_\infty \leqslant \sqrt{\frac{\pi}{2}}\|h - \mathrm{E}[h(Z)]\|_\infty$$

$$\|Uh'\|_\infty \leqslant 2\|h - \mathrm{E}[h(Z)]\|_\infty$$

$$\|Uh''\|_\infty \leqslant 2\|h'\|_\infty.$$

Bounds for the Stein operator

Proof.

For $w \leqslant 0$,

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

and, for $w \geqslant 0$,

$$|Uh(w)| \leqslant \left[\sup_{x \geqslant 0} |h(x) - \mathrm{E}[h(Z)]|\right] e^{\frac{w^2}{2}} \int_{w}^{\infty} e^{-\frac{x^2}{2}} dx.$$

The first claim follows since the maximum of $e^{\frac{w^2}{2}} \int_{-\infty}^{w} e^{-\frac{x^2}{2}} dx$ in $w \leqslant 0$ is attained at 0. $\square$

Bounds for the Stein operator

Proof.

For $w \geqslant 0$ use

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

Hence

$$\begin{aligned} \sup_{w \geqslant 0} |(Uh)'(w)| &\leqslant [\sup |h - \mathrm{E}[h(Z)]|] \left[1 + \sup_{w \geqslant 0} we^{\frac{w^2}{2}} \int_w^\infty e^{-\frac{x^2}{2}} dx\right] \\ &\leqslant 2 \sup |h - \mathrm{E}[h(Z)]|. \end{aligned}$$

The bound for $w \leqslant 0$ 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. $\square$

Exchangeable pairs

Definition

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

$$\mathrm{Prob}(X \in B, X' \in B') = \mathrm{Prob}(X \in B', X' \in B).$$

The following lemma is key.

Lemma

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

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

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

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

Stein's method for normal approximation

Proof.

From the identity

$$
\begin{aligned}
0 &= \mathrm{E}\left[Wf(W) - \frac{1}{2\lambda}(W' - W)(f(W') - f(W))\right] \\
&= \mathrm{E}[Wf(W) - f'(W)] + \mathrm{E}\left[f'(W) - \frac{1}{2\lambda}(W' - W)(f(W') - f(W))\right] \\
&= \mathrm{E}[h(Z)] - \mathrm{E}[h(W)] + E[f'(W)] - \frac{1}{2\lambda}\mathrm{E}[(W' - W)(f(W') - f(W))]
\end{aligned}
$$

obtain

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

□

Stein's method for normal approximation

Proof.

Rewrite part of the last line as

$$
\begin{aligned}
&\mathrm{E}\left[f'(W) - \frac{1}{2\lambda}(W' - W)(f(W') - f(W))\right] \\
&= \mathrm{E}\left[f'(W)\left[1 - \frac{1}{2\lambda}\mathrm{E}\left[(W' - W)^2 | W\right]\right]\right] \\
&\quad - \frac{1}{2\lambda}\mathrm{E}\left[(W' - W)\left[f(W') - f(W) - (W' - W)f'(W)\right]\right].
\end{aligned}
$$

□

Stein's method for normal approximation

Proof.

Write

$$f(W') - f(W) - (W' - W)f'(W) = \int_W^{W'} (W' - y)f''(y)dy$$

$$= \int (W' - y)[\mathbf{1}(y \leqslant W') - \mathbf{1}(y \leqslant W)]f''(y)dy.$$

Take expectation and use the exchangeability of W, W' to obtain the claim. $\square$

Stein's method for normal approximation

Theorem

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

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

and for all real w_0,

$$\begin{aligned}|\operatorname{Prob}(W \leqslant w_0) - \Phi(w_0)| \leqslant &2\sqrt{\operatorname{E}\left[\left(1 - \frac{1}{2\lambda}\operatorname{E}\left[(W'-W)^2|W\right]\right)^2\right]} \\ &+ (2\pi)^{-\frac{1}{4}}\sqrt{\frac{1}{\lambda}\operatorname{E}\left[|W'-W|^3\right]}.\end{aligned}$$

Stein's method for normal approximation

Proof.

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

so

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

$\square$

Stein's method for normal approximation

Proof.

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

$$
\begin{aligned}
&|\,\mathrm{E}[h(W)] - \mathrm{E}[h(Z)]| \leqslant \\
&2\|h - \mathrm{E}[h(Z)]\|_\infty \sqrt{\mathrm{E}\left[\left(1 - \frac{1}{2\lambda}\mathrm{E}\left[(W - W')^2|W\right]\right)^2\right]} \\
&+ 2\|h'\|_\infty \frac{1}{2\lambda}\mathrm{E}\left[\frac{|W - W'|^3}{4}\right].
\end{aligned}
$$

This proves the first bound.
To prove the second, bound $\mathbf{1}(w \leqslant w_0)$ from above and below using piece-wise linear functions. We omit the details. □