1. Suppose $B_i \overset{\text{iid}}{\sim} \text{Ber}(p_i)$, and let $B = \sum_{i=1}^n B_i$. Unfortunately, B does **not** follow a binomial distribution. We might try to approximate B with $C \sim \text{Binom}(n, \overline{p})$, where $\overline{p} = \frac{1}{n}\sum_{i=1}^n p_i$.

   (a) Show that $M_B(t) \leq M_C(t)$ for all $t \in \mathbb{R}$. **Hint:** Use the AM-GM inequality

   $$\left(\prod_{i=1}^n a_i\right)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{i=1}^n a_i$$

   for non-negative $a_1, \cdots, a_n$.

   (b) Show that $\mathbb{E}B = \mathbb{E}C$ and $\text{var}(B) \leq \text{var}(C)$.

Solution.

(a) We have

$$\begin{aligned} M_B(t) &= \prod_{i=1}^n (1 - p_i + p_i e^t) \\ &\leq \left(\frac{1}{n}\sum_{i=1}^n (1 - p_i + p_i e^t)\right)^n \\ &= (1 - \overline{p} + \overline{p}e^t)^n \\ &= M_C(t) \end{aligned}$$

(b) *Way 1.* We may show it directly:

$$\mathbb{E}B = \sum_{i=1}^n \mathbb{E}B_i = \sum_{i=1}^n p_i = n\overline{p} = \mathbb{E}C$$

Showing $\text{var}(B) \leq \text{var}(C)$ is equivalent to showing

$$\begin{aligned} &\sum_{i=1}^n p_i(1 - p_i) \leq n\overline{p}(1 - \overline{p}) \\ \iff &\sum_{i=1}^n p_i^2 \geq n\overline{p}^2 \\ \iff &\sum_{i=1}^n (p_i - \overline{p})^2 \geq 0 \end{aligned}$$

with the last inequality true trivially, and hence just work your way backwards.

Way 2. Using part (a),

$$1 + t\mathbb{E}B + \frac{t^2}{2}\mathbb{E}B^2 \leq 1 + t\mathbb{E}C + \frac{t^2}{2}\mathbb{E}C^2 + \mathcal{O}(t^3) \qquad (\star)$$

where $\mathcal{O}(t^3)$ is bounded by t^3 for sufficiently small t. Subtracting 1 and dividing by t,

$$\mathbb{E}B + \frac{t}{2}\mathbb{E}B^2 \leq \mathbb{E}C + \frac{t}{2}\mathbb{E}C^2 + \mathcal{O}(t^2) \qquad \text{for } t > 0$$
$$\mathbb{E}B + \frac{t}{2}\mathbb{E}B^2 \geq \mathbb{E}C + \frac{t}{2}\mathbb{E}C^2 + \mathcal{O}(t^2) \qquad \text{for } t < 0$$

Hence the right and left hand limits as $t \to 0$ produces $\mathbb{E}B \leq \mathbb{E}C$ and $\mathbb{E}B \geq \mathbb{E}C$, respectively, together giving $\mathbb{E}B = \mathbb{E}C$. Then, subtracting $1 + t\mathbb{E}B$ and diving by $t^2/2$ on both sides of $(\star)$,

$$\mathbb{E}B^2 \leq \mathbb{E}C^2 + \mathcal{O}(t) \quad \text{for all } t$$

Hence $t \to 0$ gives us $\mathbb{E}B^2 \leq \mathbb{E}C^2$.

□

2. Suppose $Y|X \sim \text{Ber}(\Phi(\alpha + \beta X))$, where $\Phi(t) = \int_{-\infty}^{t} \phi(x)dx$ is the standard normal CDF. Instead of observing X, I give you Z, which is a mis-measured version of X. Suppose $X|Z \sim N(\gamma + \delta Z, \tau^2)$.

   (a) First, prove the integral

$$\int_{-\infty}^{\infty} \Phi(a + bx)\phi(x)dx = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right)$$

   (b) Assuming $Y|(X, Z) = Y|X$, show that $Y|Z \sim \text{Ber}(\Phi(\widetilde{\alpha} + \widetilde{\beta}Z))$, where

$$\widetilde{\alpha} = \frac{\alpha + \beta\gamma}{\sqrt{1+\beta^2\tau^2}}, \qquad \widetilde{\beta} = \frac{\beta\delta}{\sqrt{1+\beta^2\tau^2}}$$

   In most measurement error contexts, $0 < \delta < 1$ and $\tau^2 > 0$, hence $|\widetilde{\beta}| < |\beta|$, so the mis-measured covariate provides a bias towards the null for the effect of X on Y.

   Solution.

   (a) The LHS can be reformulated as $\mathbb{P}(Y \leq a + bX)$, where $X, Y \overset{\text{iid}}{\sim} N(0, 1)$. Hence

$$\begin{aligned}\mathbb{P}(Y \leq a + bX) &= \mathbb{P}(Y - bX \leq a) = \mathbb{P}(N(0, 1 + b^2) \leq a) \\ &= \mathbb{P}\left(N(0,1) \leq \frac{a}{\sqrt{1+b^2}}\right) = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right)\end{aligned}$$

   (b) We have

$$\begin{aligned}\mathbb{P}(Y = 1|Z) &= \mathbb{E}[Y|Z] = \mathbb{E}[\mathbb{E}[Y|X, Z]|Z] \\ &= \mathbb{E}[\mathbb{E}[Y|X]|Z] = \mathbb{E}[\Phi(\alpha + \beta X)|Z] \\ &= \int_{-\infty}^{\infty} \Phi(\alpha + \beta x) f_{X|Z}(x|z)dx \\ &= \int_{-\infty}^{\infty} \Phi(\alpha + \beta(\gamma + \delta Z) + \beta\tau v)\phi(v)dv \quad \text{Transform } v = \frac{x - (\gamma + \delta Z)}{\tau}\end{aligned}$$

   Let $a = \alpha + \beta\gamma + \beta\delta Z$ and $b = \beta\tau$, and using the result from part (a),

$$\begin{aligned}\mathbb{P}(Y = 1|Z) &= \Phi\left(\frac{\alpha + \beta\gamma + \beta\delta Z}{\sqrt{1+\beta^2\tau^2}}\right) \\ &= \Phi(\widetilde{\alpha} + \widetilde{\beta}Z)\end{aligned}$$

   as desired.

□

3. Suppose $X_1, \cdots, X_n$ are independent RV's, and let $Z = h(X_1, \cdots, X_n)$. Let $\mathbf{X}_{-i} = (X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n)$ and $\mathbf{X}_{i:j} = (X_i, X_{i+1}, \cdots, X_{j-1}, X_j)$. Define

$$\mathbb{E}_i Z = \mathbb{E}[Z|\mathbf{X}_{1:i}] = \int_{\mathbb{R}^{n-i}} h(X_1, \cdots, X_i, x_{i+1}, \cdots, x_n) dF_{X_{i+1}}(x_{i+1}) \cdots dF_{X_n}(x_n)$$
$$\mathbb{E}^{(i)} Z = \mathbb{E}[Z|\mathbf{X}_{-i}] = \int_{\mathbb{R}} h(X_1, \cdots, X_{i-1}, x_i, X_{i+1}, \cdots, X_n) dF_{X_i}(x_i)$$

(a) Show that $\mathbb{E}_i \mathbb{E}^{(i)} Z = \mathbb{E}_{i-1} Z$.

(b) Define $\Delta_i = \mathbb{E}_i Z - \mathbb{E}_{i-1} Z$. Show that $\mathbb{E}_i \Delta_j = 0$ for $j > i$. Along with the representation $Z - \mathbb{E}Z = \sum_{i=1}^n \Delta_i$, conclude that

$$\text{var}(Z) = \sum_{i=1}^n \mathbb{E}\Delta_i^2$$

(c) Using part (a), prove that

$$\Delta_i^2 \leq \mathbb{E}_i (Z - \mathbb{E}^{(i)} Z)^2$$

(d) Use (b) and (c) to conclude that

$$\text{var}(Z) \leq \sum_{i=1}^n \mathbb{E}(Z - \mathbb{E}^{(i)} Z)^2$$

This is known as the *Efron-Stein inequality*, which provides a bound on the variance of a function of independent RV's.

Solution.

(a) Writing longhand, $\mathbb{E}_i \mathbb{E}^{(i)} Z = \mathbb{E}[\mathbb{E}[Z|\mathbf{X}_{-i}]|\mathbf{X}_{1:i}]$. The inner expectation marginalizes out X_i, while the outer expectation marginalizes out $X_{i+1}, \cdots, X_n$. Together, they marginalize out $X_i, \cdots, X_n$. This is exactly $\mathbb{E}_{i-1} Z$. More formally,

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Z|\mathbf{X}_{-i}]|\mathbf{X}_{1:i}] &= \int_{\mathbb{R}^{n-i}} \left(\int_{\mathbb{R}} h(X_1, \cdots, X_{i-1}, x_i, x_{i+1}, \cdots, x_n) dF_{X_i}(x_i) \right) dF_{\mathbf{X}_{(i+1):n}}(\mathbf{x}_{(i+1):n}) \\
&= \int_{n-i-1} h(X_1, \cdots, X_{i-1}, x_i, x_{i+1}, \cdots, x_n) dF_{X_i}(x_i) \cdots dF_{X_n}(x_n) \\
&= \mathbb{E}_{i-1} Z
\end{aligned}$$

(b) For $j = i$, it is clear that $\mathbb{E}_i \mathbb{E}_j Z = \mathbb{E}_i Z$. For $j > i$, note that $\mathbb{E}_i \mathbb{E}_j Z = \mathbb{E}_i Z$ as well, for the outer expectation marginalizes out the $X_{i+1}, \cdots, X_j$ that wasn't marginalized out by E_j in the first place. Hence, for $j > i$ (or $j \geq i+1$), we have

$$\mathbb{E}_i \Delta_j = \mathbb{E}_i \mathbb{E}_j Z - \mathbb{E}_i \mathbb{E}_{j-1} Z = \mathbb{E}_i Z - \mathbb{E}_i Z = 0$$

Hence,

$$\begin{aligned} \mathrm{var}(Z) = \mathbb{E}(Z - \mathbb{E}Z)^2 &= \mathbb{E}\left(\sum_{i=1}^{n} \Delta_i\right)^2 \\ &= \sum_{i=1}^{n} \mathbb{E}\Delta_i^2 + 2\sum_{j>i} \mathbb{E}\Delta_i \Delta_j \end{aligned}$$

But note that $\mathbb{E}\Delta_i\Delta_j = \mathbb{E}[\mathbb{E}_i \Delta_i \Delta_j] = \mathbb{E}[\Delta_i \mathbb{E}_i \Delta_j] = \mathbb{E}[\Delta_i \cdot 0] = 0$, so all the cross-terms go away. Hence, the result follows.

(c) By part (a), we have the representation $\Delta_i = \mathbb{E}_i[Z - \mathbb{E}^{(i)} Z]$ then

$$0 \leq \mathrm{var}(Z - \mathbb{E}^{(i)} Z | \mathbf{X}_{1:i}) = \mathbb{E}_i (Z - \mathbb{E}^{(i)} Z)^2 - \Delta_i^2$$

hence the result follows.

(d) We have

$$\begin{aligned} \mathrm{var}(Z) &= \sum_{i=1}^{n} \mathbb{E}\Delta_i^2 \\ &\leq \sum_{i=1}^{n} \mathbb{E}[\mathbb{E}_i (Z - \mathbb{E}^{(i)} Z)^2] \\ &= \sum_{i=1}^{n} \mathbb{E}(Z - \mathbb{E}^{(i)} Z)^2 \end{aligned}$$

□