STA 6934
Problem 3.28 (a,b,c)

Vladimir L. Boginski

We need to evaluate $P(x > a) = \int_a^\infty f(x)dx$, $X \sim f$.

a) We have the following estimators:

$$\delta_1 = \frac{1}{n}\sum_{i=1}^n I(X_i > a), X_i \sim f, iid$$

$$\delta_3 = \frac{1}{n}\sum_{i=1}^n I(X_i > \mu),$$

where $P(X > \mu)$ is known, and $a > \mu$.
The control variate estimator:

$$\delta_2 = \frac{1}{n}\sum_{i=1}^n I(X_i > a) + \beta[\sum_{i=1}^n I(X_i > \mu) - P(X > \mu)].$$

Since $\text{var}(\delta_2) = \text{var}(\delta_1) + \beta^2 \text{var}(\delta_3) + 2\beta \text{cov}(\delta_1, \delta_3)$, and $\text{cov}(\delta_1, \delta_3) = \frac{1}{n} P(X > a)[1 - P(X > \mu)]$, we can find $\text{var}(\delta_3)$:

$$\text{var}(\delta_3) = \frac{1}{n^2}\sum_{i=1}^n \text{var}(I(X_i > \mu)),$$

(because X_i are iid).

$$\text{var}\, I(X_i > \mu) = E[I(X_i > \mu)]^2 - [E(I(X_i > \mu))]^2 =$$

$$= E[I(X_i > \mu)] - [E(I(X_i > \mu))]^2 = \int_\mu^\infty f(x)dx - [\int_\mu^\infty f(x)dx]^2 =$$

$$= P(X > \mu) - [P(X > \mu)]^2 = P(X > \mu)[1 - P(X > \mu)]$$

(We used the fact that $E[I(X_i > \mu)]^2 = E[I(X_i > \mu)]$.)
So,

$$\text{var}(\delta_3) = \frac{1}{n^2} nP(x > \mu)[1 - P(x > \mu)] = \frac{1}{n} P(x > \mu)[1 - P(x > \mu)].$$

b) $\text{var}(\delta_2) = \text{var}(\delta_1) + \beta^2 \text{var}(\delta_3) + 2\beta \text{cov}(\delta_1, \delta_3)$.
If we want δ_2 to improve δ_1 (in the sense of reducing variance) , we need the following condition to be satisfied:

$$\beta^2 \text{var}(\delta_3) + 2\beta \text{cov}(\delta_1, \delta_3) < 0.$$

Since $\beta^2 \operatorname{var}(\delta_3) \geq 0$, and $\operatorname{cov}(\delta_1, \delta_3) \geq 0$ (see part a), we need

$$\beta < 0.$$

Also from the inequality $\beta^2 \operatorname{var}(\delta_3) + 2\beta \operatorname{cov}(\delta_1, \delta_3) < 0$ we get

$$\beta^2 \operatorname{var}(\delta_3) < -2\beta \operatorname{cov}(\delta_1, \delta_3),$$

$$-\beta < 2\frac{\operatorname{cov}(\delta_1, \delta_3)}{\operatorname{var}(\delta_3)}$$

or $|\beta| < 2\dfrac{\operatorname{cov}(\delta_1, \delta_3)}{\operatorname{var}(\delta_3)}$, since $\beta < 0$.

So, the conditions on β are:

$$\beta < 0 \text{ and } |\beta| < 2\frac{\operatorname{cov}(\delta_1, \delta_3)}{\operatorname{var}(\delta_3)}.$$

c)

$f = N(0,1)$, find $P(X > a)$ for $a = 3, 5, 7$.

After generating 1000000 random variables from N(0,1) and applying the estimator δ_1, we get:

a	estimator
3	0.00134
5	0.00000122
7	0