Statistics 200 Homework 3 Solutions

Nicolas Annoni

January 27, 2009

1 5.1

Using Chebyshev's Inequality, we have that

$$\begin{aligned} P(|\bar{X} - \mu > \epsilon) &\leq \frac{\text{var}(\bar{X})}{\epsilon^2} \\ &= \frac{\text{var}(\sum_i \frac{X_i}{n})}{\epsilon^2} \\ &= \frac{\frac{1}{n^2}\sum_i \text{var}X_i}{\epsilon^2} \\ &= \frac{\frac{1}{n^2}\sum_i \sigma_i^2}{\epsilon^2} \\ &\to 0, \end{aligned}$$

by assumption.

2 5.16

First note that

$$EX_i = \int_0^1 2x^2 dx = 2/3$$

and

$$\text{var}X_i = \int_0^1 2x^3 dx - (2/3)^2 = 1/2 - 4/9 = 1/18.$$

Thus, we have that

$$\begin{aligned} P(S \leq 10) &= P\left(\frac{(S - 20(2/3))}{\sqrt{20/18}} \leq \frac{(10 - 20(2/3))}{\sqrt{20/18}}\right) \\ &\approx P\left(Z \leq -\frac{\sqrt{18/20} \cdot 10}{3}\right) \\ &\approx 0.0008. \end{aligned}$$

3 5.21

a) We have that

$$E\hat{I}(f) = \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{f(X_i)}{g(X_i)}\right) = \frac{1}{n}\sum_{i=1}^{n}\int_a^b \frac{f(x)}{g(x)}g(x)dx = I(f).$$

b) First calculate $E(\hat{I}(f))^2$. We have that:

$$\begin{aligned}
E(\hat{I}(f))^2 &= \frac{1}{n^2}E\left(\sum_{i=1}^{n}\frac{f(X_i)}{g(X_i)}\right)^2 \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}E\frac{f^2(X_i)}{g^2(X_i)} + \sum_{i\neq j}E\frac{f^(X_i)f(X_j)}{g(X_i)g(X_j)}\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}E\frac{f^2(X_i)}{g^2(X_i)} + \sum_{i\neq j}E\frac{f^(X_i)}{g(X_i)}E\frac{f^(X_j)}{g(X_j)}\right) \\
&= \frac{1}{n^2}\left(n\int_a^b \frac{f^2(x)}{g(x)}dx + n(n-1)I^2(f)\right).
\end{aligned}$$

So that

$$\begin{aligned}
\mathrm{var}(\hat{I}(f)) &= \frac{1}{n}\int_a^b \frac{f^2(x)}{g(x)}dx + \frac{n-1}{n}I^2(f) - I^2(f) \\
&= \frac{1}{n}\left(\int_a^b \frac{f^2(x)}{g(x)}dx - I^2(f)\right).
\end{aligned}$$

For the finite case, let $g(x) = 1/(b-a)$; i.e. we generate X_i from the uniform distribution over $[a,b]$. Then, provided f and f^2 are integrable on $[a,b]$, the variance will be finite for each n. For the infinite case, let $a = 0, b = 1$ with $f(x) = 1$ and $g(x) = 2x$. Then, the integral will be infinite.

c) Then, we have that $g(x) = 1$ so that $\hat{I}(f) = \sum_{i=1}^{n} f(X_i)$, with the $X_i \sim U(0,1)$, which is the Monte Carlo estimate in the example. Finally, one possibility is to consider

$$g(x) = \frac{|f(x)|}{\int_a^b |f(x)|dx}.$$

Comparing the variance when $g(x) = 1$ to the variance with g above reduces to showing that

$$\int_a^b \frac{f^2(x)}{g(x)}dx \leq \int_a^b f^2(x)dx.$$

Then, the variance will be smaller with our new $g(x)$. So, observe that

$$\begin{aligned}\int_a^b \frac{f^2(x)}{g(x)}dx &= \int_a^b \frac{f^2(x)}{|f(x)|}\left(\int_a^b |f(y)|dy\right)dx \\ &= \left(\int_a^b |f(x)|dx\right)^2 \\ &\leq \int_a^b f^2(x)dx,\end{aligned}$$

where the last line is an application of Jensen's Inequality.

4 5.24

Let X_i be a Bernoulli R.V. that takes on a value of 1 if a point lands inside region A, and 0 otherwise. Then, we estimate the area of A with

$$\hat{A} = \frac{\sum_{i=1}^n X_i}{n}.$$

Noting that $EX_i = 0.2$ and $\mathrm{SD}X_i = \sqrt{0.2 \cdot 0.8} = 0.4$, we have that

$$P\left(-0.01 < \frac{\sum_{i=1}^n X_i}{n} - 0.2 < 0.01\right) = P\left(-\frac{\sqrt{n}}{40} < \frac{\sum_{i=1}^n X_i - 0.2n}{0.4\sqrt{n}} < \frac{\sqrt{n}}{40}\right),$$

when multiply through by $\sqrt{n}/0.4$. Treating

$$G_n = \frac{\sum_{i=1}^n X_i - 0.2n}{0.4\sqrt{n}}$$

as a standard normal random variable, we set the 99.5% quantile of that distribution equal to $\frac{\sqrt{n}}{40}$, which implies that

$$n = 1600Q_{99.5}^2 = 1600(2.57)^2 = 10568.$$

Thus, we need 10568 points to keep the error of the estimate under 0.01 with 99% probability.

5 7.1

The population mean and variance are given by:

$$\begin{aligned}EX &= \frac{1+2+2+4+8}{5} = 3.4 \\ \mathrm{var}X &= \frac{2.4^2 + 1.4^2 + 1.4^2 + 0.6^2 + 4.6^2}{5} = 6.24\end{aligned}$$

The possible samples (multiplicities included) are: $\{(1,2),(1,2),(1,4),(1,8),(2,2),$ $(2,4),(2,8),(2,4),(2,8),(4,8)\}$, each having the probability of $1/10$ of being drawn at random. Thus, the sampling distribution of the mean is given by:

$$\begin{aligned}
P(\bar{X}=1.5) &= 0.2 \\
P(\bar{X}=2) &= 0.1 \\
P(\bar{X}=2.5) &= 0.1 \\
P(\bar{X}=3) &= 0.2 \\
P(\bar{X}=4.5) &= 0.1 \\
P(\bar{X}=5) &= 0.2 \\
P(\bar{X}=6) &= 0.1
\end{aligned}$$

The mean of the sampling distribution is 3.4 and the variance is 2.34. Because the mean of the population and the mean of the sampling distribution match, these results are consistent with Theorem A. Because

$$\frac{6.24}{2}\left(\frac{3}{4}\right) = 2.34,$$

the results are consistent with Theorem B.

6

Suppose that $Y \Rightarrow y$. Then, we have that

$$P(Y_n \leq c) \to \begin{cases} 1 & c \geq y \\ 0 & c < y \end{cases} \qquad (*),$$

except at $c = y$, since this is not a continuity point. Observe that for any $\epsilon > 0$:

$$\begin{aligned}
P(|Y_n - y| > \epsilon) &= P(Y_n - y > \epsilon) + P(Y_n - y < -\epsilon) \\
&= P(Y_n > y + \epsilon) + P(Y_n < y - \epsilon) \\
&\leq 1 - P(Y_n \leq y + \epsilon) + P(Y_n \leq y - \epsilon) \\
&\to 1 - 1 + 0 \\
&= 0,
\end{aligned}$$

with the convergence of the probabilities given by $(*)$. Thus, Y converges in probability to y.

7

Squaring all terms, we get that

$$\begin{aligned}
& n(p_n - p)^2 \leq z^2 p(1-p) \\
\Rightarrow\ & n(p_n^2 - 2p_n p + p^2) \leq z^2 p - z^2 p^2 \\
\Rightarrow\ & (n + z^2)p^2 - (2p_n n + z^2)p + p_n^2 n \leq 0,
\end{aligned}$$

where we set $z^2 = z_{1-\alpha/2}^2 = (-z_{1-\alpha/2})^2$. Using the quadratic formula and noting that $(n+z^2)$ is positive, we have that:

$$\begin{aligned}
& p \in \frac{(2p_n n + z^2) \mp \sqrt{(2p_n n + z^2)^2 - 4(n+z^2)p_n^2 n}}{2(n+z^2)} \\
\Rightarrow \quad & p \in \frac{(2p_n n + z^2) \mp \sqrt{4p_n n z^2 + z^4 - 4z^2 p_n^2 n}}{2(n+z^2)} \\
\Rightarrow \quad & p \in \frac{(2p_n n + z^2) \mp z\sqrt{4p_n n(1-p_n) + z^2}}{2(n+z^2)} \\
\Rightarrow \quad & p \in \frac{p_n + \frac{z^2}{2n} \mp z\sqrt{\frac{p_n(1-p_n)}{n} + \frac{z^2}{4n^2}}}{1+\frac{z^2}{n}},
\end{aligned}$$

which is known as the Wilson Score Interval. Using the given parameters, we have that $z = 1.96$ and $p_n = 0.6$. Then,

$$\begin{aligned}
\hat{I}_n &= 0.6 \mp 1.96 \cdot \frac{1}{10}\sqrt{0.24} = [0.504, 0.696] \\
\tilde{I}_n &= \frac{0.6 + \frac{1.96^2}{200} \mp 1.96\sqrt{0.0024 + \frac{1.96^2}{40000}}}{1 + \frac{1.96^2}{100}} \\
&= [0.502, 0.691].
\end{aligned}$$