MATH 565 Monte Carlo Methods

Fred J. Hickernell | Test 1 | Wednesday, September 24, 2025

Instructions:

i. *This test has* ***THREE*** *questions. Attempt them all. The maximum number of points is* ***100****.*
ii. *The time allowed is 75 minutes.*
iii. *This test is closed book, but you may use 4 double-sided letter-size sheets of notes.*
iv. *No calculators or other devices are allowed.*
v. *Keep at least* three *significant digits in your intermediate calculations and final answers.*
vi. *Show all your work to justify your answers. Answers without adequate justification will not receive credit.*
vii. *Off-site students may contact the instructor as directed by your syllabus.*

This table provides values of the standard normal cumlative distribution function:

z	-3	-2	-1	0	1	2	3
$\Phi(z)$	0.00135	0.0228	0.1587	0.5000	0.8413	0.9772	0.99865

1. (30 points) Consider a Bernoulli random variable, $Y \sim \text{Bernoulli}(\mu)$, that is evaluated by a time-consuming computer simulation. The sample space is $\{0, 1\}$, with 1 denoting "success". The *unkown* mean, $\mu = \mathbb{E}(Y)$, is the probability of success. Suppose that one has independent and identically distributed (IID) instances of this random variable, namely $Y_0, \ldots, Y_{n-1}$.

   a. (7 points) Show that the sample mean, $\hat{\mu}_n = n^{-1} \sum_{i=0}^{n-1} Y_i$, is an unbiased estimator of μ.

   Answer:

$$\mathbb{E}(\hat{\mu}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=0}^{n-1} Y_i\right) = \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}(Y_i) = \frac{1}{n}\sum_{i=0}^{n-1} \mu = \mu$$

   b. (8 points) Derive the function g for which the *variance* of Y equals $g(\mu)$.

   Answer: Since $Y = Y^2$,

$$\text{var}(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 = \mu - \mu^2 = \mu(1-\mu) =: g(\mu)$$

   c. (15 points) Is $g(\hat{\mu}_n)$ and *unbiased* estimate of the variance of Y? If not, how can one correct it to make it unbiased?

Answer:

$$\mathbb{E}[g(\hat{\mu}_n)] = \mathbb{E}[\hat{\mu}_n - \hat{\mu}_n^2] = \mu - \mathbb{E}(\hat{\mu}_n^2)$$

$$\begin{aligned}
\mathbb{E}(\hat{\mu}_n^2) &= \mathbb{E}\left[\frac{1}{n}\sum_{i=0}^{n-1} Y_i \times \frac{1}{n}\sum_{j=0}^{n-1} Y_j\right] \\
&= \frac{1}{n^2}\sum_{i,j=0}^{n-1} \mathbb{E}(Y_i Y_j) \\
&= \frac{1}{n^2}\left[\sum_{i=0}^{n-1} \mathbb{E}(Y_i^2) + \sum_{\substack{i,j=0\\ i\neq j}}^{n-1} \mathbb{E}(Y_i Y_j)\right] \\
&= \frac{1}{n^2}\left[\sum_{i=0}^{n-1} \mu + \sum_{\substack{i,j=0\\ i\neq j}}^{n-1} \mu^2\right] \\
&= \frac{1}{n^2}\left[n\mu + n(n-1)\mu^2\right] = \frac{\mu + (n-1)\mu^2}{n}
\end{aligned}$$

and so

$$\mathbb{E}[g(\hat{\mu}_n)] = \mu - \frac{\mu + (n-1)\mu^2}{n} = \frac{(n-1)\mu(1-\mu)}{n}$$

Thus, $g(\hat{\mu}_n)$ is biased, but $ng(\hat{\mu}_n)/(n-1)$ is unbiased, provided $n > 1$.
Note that for $n = 1$, $g(\hat{\mu}_n) = Y_1(1 - Y_1) = 0$, so we cannot expect $g(\hat{\mu}_n)$ to be unbiased.

One should recognize that p, the probability of success is the same as μ, the mean. There seemed to be confusion between mean and sample mean.

2. (40 points) The random variable X has the following triangular probability density function (PDF)

$$\varrho(x) = \begin{cases} 0, & -\infty < x \le -2 \\ c(x+2), & -2 < x \le 0, \\ c(2-x), & 0 < x \le 2, \\ 0, & 2 < x < \infty. \end{cases}$$

for some value of c.

a. (10 points) What should the value of c be?

Answer: The cumulative distribution function (CDF) is

$$F(x) = \begin{cases} 0, & -\infty < x \le -2 \\ F(-2) + \int_{-2}^{x} c(t+2)\mathrm{d}t = \frac{c}{2}(x+2)^2, & -2 < x \le 0, \\ F(0) + \int_{0}^{x} c(2-t)\mathrm{d}t = 2c + 2c - \frac{c}{2}(2-x)^2 = 4c - \frac{c}{2}(2-x)^2, & 0 < x \le 2, \\ 1 = F(2) = 4c, & 2 < x < \infty. \end{cases}$$

So $c = 1/4$ *and*

$$F(x) = \begin{cases} 0, & -\infty < x \le -2 \\ \frac{(x+2)^2}{8}, & -2 < x \le 0, \\ 1 - \frac{(2-x)^2}{8}, & 0 < x \le 2, \\ 1, & 2 < x < \infty. \end{cases}$$

b. (15 points) What is the *quantile* function for X?

Answer: We can restrict ourselves to $-2 \le x \le 2$.

$$\begin{aligned} u = F(x) &\iff \begin{cases} u = \frac{(x+2)^2}{8}, & -2 < x \le 0, \\ u = 1 - \frac{(2-x)^2}{8}, & 0 < x \le 2. \end{cases} \\ &\iff \begin{cases} x = -2 + \sqrt{8u}, & -2 < x \le 0, \\ x = 2 - \sqrt{8(1-u)}, & 0 < x \le 2. \end{cases} \\ &\iff \begin{cases} x = -2 + \sqrt{8u}, & 0 < u \le 1/2, \\ x = 2 - \sqrt{8(1-u)}, & 1/2 < x \le 1. \end{cases} \end{aligned}$$

This means that the quantile function is

$$Q(u) = \begin{cases} -2 + \sqrt{8u} = -2 + 2\sqrt{2u}, & 0 < u \le 1/2, \\ 2 - \sqrt{8(1-u)} = 2 - 2\sqrt{2(1-u)}, & 1/2 < x \le 1. \end{cases}$$

c. (15 points) Given IID $\mathcal{U}[0, 1]$ samples of 0.125 and 0.1, what would *one* reasonable sample for X be?

Answer: Applying the quantile function to 0.125 *gives* $X = Q(0.125) = -2 + \sqrt{8 \times 0.125} = -1$.
One could also use acceptance-rejection sampling with the proposal density $Z \sim \mathcal{U}[-2, 2]$ *and note that since the density of* Z *is* $1/4$, *one has* $\varrho_X(x) \le 2\varrho_Z(x)$ *for* $-2 \le x \le 2$, *Then transforming* $U_1 = 0.125$ *to a* $\mathcal{U}[-2, 2]$ *distribution by the transformation* $Z = 4U - 2 = -1.5$, *and using* $U_2 = 0.1$ *as the decision variable, we note that*

$$\frac{\varrho_X(-1.5)}{2\varrho_Z(-1.5)} = \frac{(-1.5+2)/4}{2 \times 1/4} = \frac{1}{4} \ge 0.1 = U_2,$$

so we may accept $X = -1.5$.

This question requires a clear understanding of probability density, cumulative disribution, and quantile. The probability density must be non-negative and integrate to one. It also requires some facility with integrating piecewise linear fundtions. The reason that two uniform samples were given in part c. was to allow you to use acceptance-rejection sampling.

3. (30 points) Given two independent random variables, Y and Z, with the *same* mean, μ, and different variances, one generates two pilot IID samples, $Y_1, \ldots, Y_{100}$ and $Z_1, \ldots, Z_{100}$, and computes their sample measn and variances:

	sample mean, $\widehat{\mu}_{100}$	sample variance, $\widehat{\sigma}^2_{100}$	time required (seconds)
Y	35	16	3
Z	40	36	1

a. (10 points) Approximating the true variance of Y by its sample variance, i.e., $\sigma^2_Y \approx \widehat{\sigma}^2_{Y,100}$, what sample size would be required to make twice the standard deviation of the sample mean for Y no greater than 0.01?

Answer: The variance of the sample mean is $\sigma^2_Y/n \approx \widehat{\sigma}^2_{Y,100}/n = 16/n$. *This means that the sample standard deviation is approximately* $4/\sqrt{n}$. *To make twice the standard deviation of the sample mean be no greater than* 0.01, *we need*

$$\frac{8}{\sqrt{n}} \le 0.01 \iff 800 \le \sqrt{n} \iff n \ge 640\,000$$

b. (10 points) Using the Central Limit Theorem approximation, what confidence interval for the mean would the sample size in part a. give you based on sampling Y, and what would be the confidence level?

Answer: We have a confidence interval of

$$\Pr[\widehat{\mu}_{640\,000} - 0.01 \le \mu \le \widehat{\mu}_{640\,000} + 0.01] \ge 0.9772 - 0.0228 = 0.9544 = 95.44\%.$$

c. (10 points) From the perspective of *time required*, is it better to sample Y or Z to approximate μ to a given tolerance?

Answer: To get a root mean square error of ε *using* Y *samples costs* $n = \sigma^2_Y/\varepsilon^2 = 16/\varepsilon^2$ *samples and* $0.03n = 0.48/\varepsilon^2$ *seconds.*
To get that same root mean square error using Z *samples costs* $n = \sigma^2_Z\varepsilon^2 = 36/\varepsilon^2$ *samples and* $0.01n = 0.36/\varepsilon^2$ *seconds. Thus, it is more cost-effective to sample from* Z.
An argument based on confidence interval widths would lead to the same conclusion.

Some were confused by the difference between the sample standard deviation, and the standard deviation of the sample mean. Note that we expect $\widehat{\mu}_{100}$ to be within $2\sigma/\sqrt{100}$ of μ. This means that 35 ± 0.8 and 40 ± 1.2 should both contain μ, which they cannot. There was either highly unusual behavior, or this problem is flawed.

Exam Scores Summary
Number of Students: 36, Minimum: 2, Maximum: 100, Mean: 47.2, Median: 40
Standard Deviation: 29.1, Quartiles (Q1, Q3): (23.5, 75.5)

10	0
9	23
8	3469
7	44569
6	
5	0122
4	003
3	0235
2	034489
1	44689
0	25