STAT 410
Fall 2020

Exam 2

Thursday, November 19, 8:00 – 9:50 pm CDT

There are 4 problems on the exam, with 1, 8, 1, and 2 parts, respectively. The point value of each question is shown in parentheses before the question. The total number of points for the exam is 75.

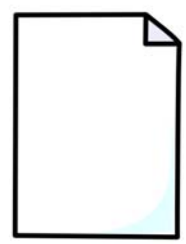
Make sure that you include everything you wish to submit, and that the submission process has completed. You do not need to include the question statements with your work. However, please label you work clearly. Neatness and organization are appreciated. Please put your final answers at the end of your work and mark them clearly.

Be sure to show all your work; your partial credit might depend on it.

No credit will be given without supporting work.

The exam is closed book and closed notes.

You are allowed to use a calculator and one 8½" x 11" sheet (both sides) with notes.



You are allowed to use
https://www.wolframalpha.com/calculators/integral-calculator/
https://www.symbolab.com/solver/definite-integral-calculator
https://www.integral-calculator.com/
https://www.desmos.com/calculator

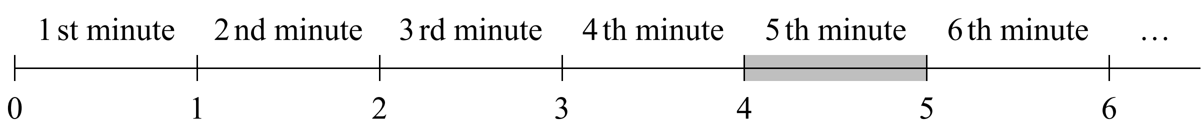
You are allowed to use
R, RStudio, and Microsoft Excel.

1. (6) Suppose that students arrive to a certain on-campus COVID-19 testing location according to a Poisson process with average rate 1 person per 24 seconds. Find the probability that the 10th student would arrive during the 5th minute.

X_t = number of students arriving in t minutes. Poisson(λt)

T_k = the arrival time of the kth student. Gamma, $\alpha = k$.

1 person per 24 seconds $\Rightarrow$ $\lambda = \dfrac{60}{24} = 2.5$ students per minute.



Need $P(4 < T_{10} < 5) = P(T_{10} < 5) - P(T_{10} < 4)$.

```
> pgamma(5,10,2.5)-pgamma(4,10,2.5)
[1] 0.2564986
```

OR

$$P(4 < T_{10} < 5) = \int_4^5 \frac{2.5^{10}}{\Gamma(10)} t^{10-1} e^{-2.5t}\, dt = \int_4^5 \frac{2.5^{10}}{9!} t^{9} e^{-2.5t}\, dt = \ldots$$

OR

If T_α has a Gamma$(\alpha, \theta = 1/\lambda)$ distribution, where α is an integer, then

$$P(T_\alpha \le t) = P(X_t \ge \alpha) \quad \text{and} \quad P(T_\alpha > t) = P(X_t \le \alpha - 1),$$

where X_t has a Poisson(λt) distribution.

$$P(4 < T_{10} < 5) = P(T_{10} > 4) - P(T_{10} > 5) = P(X_4 \le 9) - P(X_5 \le 9)$$

$$= P(\text{Poisson}(10.0) \le 9) - P(\text{Poisson}(12.5) \le 9) = 0.458 - 0.201 = \mathbf{0.257}.$$

```
> ppois(10-1,2.5*4)-ppois(10-1,2.5*5)
[1] 0.2564986
```

OR

If T_α has a Gamma$\left(\alpha, \theta = {}^{1}\!/_{\lambda}\right)$ distribution,

then $\frac{2}{\theta} T_\alpha = 2 \lambda T_\alpha$ has a $\chi^2(2\alpha)$ distribution.

$P(4 < T_{10} < 5) \;=\; P(20 < \chi^2(20) < 25).$

```
> pchisq(2*2.5*5,2*10)-pchisq(2*2.5*4,2*10)
[1] 0.2564986
```

2. Let $\beta > 0$. Consider a probability distribution with probability density function

$$f(x;\beta) = \frac{\beta}{2}\, e^{-\sqrt{\beta x}}, \qquad x > 0, \qquad \text{zero otherwise.}$$

a) (5) Show that $W = \sqrt{X}$ has a Gamma distribution with parameters $\alpha = 2$ and $\theta = \dfrac{1}{\sqrt{\beta}}$.

No credit will be given without proper justification.

You are welcome to use this result to help you with other parts of the problem.

$$W = \sqrt{X} = g(X) \qquad\qquad x = w^2 = g^{-1}(w)$$

$$\frac{dx}{dw} = 2\,w$$

$$f_W(w) = f_X(g^{-1}(w)) \cdot \left|\frac{dx}{dw}\right| = \frac{\beta}{2}\, e^{-\sqrt{\beta}\,w} \cdot 2\,w$$

$$= \beta\, w\, e^{-\sqrt{\beta}\,w} = \frac{\left(\sqrt{\beta}\right)^2}{\Gamma(2)}\, w^{2-1}\, e^{-\sqrt{\beta}\,w}, \qquad w > 0.$$

$\Rightarrow$ $W = \sqrt{X}$ has Gamma$\left(\alpha = 2,\ \theta = \dfrac{1}{\sqrt{\beta}}\right)$ distribution.

b) (5) Find the Fisher information $\mathrm{I}(\beta)$.

$$\ln f(x;\beta) = -\sqrt{\beta}\cdot\sqrt{x} + \ln\beta - \ln 2$$

$$\frac{\partial}{\partial\beta}\ln f(x;\beta) = -\frac{\sqrt{x}}{2\sqrt{\beta}} + \frac{1}{\beta}$$

$$\frac{\partial^2}{\partial\beta^2}\ln f(x;\beta) = \frac{\sqrt{x}}{4\sqrt{\beta^3}} - \frac{1}{\beta^2}$$

$$\mathrm{I}(\beta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\beta^2}\ln f(\mathrm{X};\beta)\right] = -\mathrm{E}\left[\frac{\sqrt{\mathrm{X}}}{4\sqrt{\beta^3}} - \frac{1}{\beta^2}\right] = -\frac{\mathrm{E}(\mathrm{W})}{4\sqrt{\beta^3}} + \frac{1}{\beta^2}$$

$$= -\frac{\alpha\theta}{4\sqrt{\beta^3}} + \frac{1}{\beta^2} = -\frac{2\cdot\frac{1}{\sqrt{\beta}}}{4\sqrt{\beta^3}} + \frac{1}{\beta^2} = \frac{1}{2\beta^2}.$$

OR

$$\mathrm{I}(\beta) = \mathrm{Var}\left[\frac{\partial}{\partial\beta}\ln f(\mathrm{X};\beta)\right] = \mathrm{Var}\left[-\frac{\sqrt{\mathrm{X}}}{2\sqrt{\beta}} + \frac{1}{\beta}\right] = \frac{1}{4\beta}\mathrm{Var}(\mathrm{W})$$

$$= \frac{1}{4\beta}\alpha\theta^2 = \frac{1}{4\beta}\cdot 2\cdot\frac{1}{\beta} = \frac{1}{2\beta^2}.$$

2. (continued)

Let $\beta > 0$ and let $X_1, X_2, \ldots, X_n$ be a random sample from a probability distribution with probability density function

$$f(x;\beta) = \frac{\beta}{2}\, e^{-\sqrt{\beta x}}, \qquad x > 0, \qquad \text{zero otherwise.}$$

Obviously, $\sum_{i=1}^{n} \sqrt{X_i}$ is a sufficient statistic for β.

c) (9) (i) Find the maximum likelihood estimator for β, $\hat{\beta}$.

(ii) Suppose $n = 5$, and $x_1 = 0.09$, $x_2 = 0.36$, $x_3 = 1.96$, $x_4 = 6.25$, $x_5 = 10.24$.

Find the maximum likelihood estimate for β, $\hat{\beta}$.

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\beta}{2}\, e^{-\sqrt{\beta x_i}} \right) = \frac{\beta^n}{2^n}\, e^{-\sqrt{\beta}\, \sum_{i=1}^{n} \sqrt{x_i}}.$$

$$\ln L(\beta) = n \cdot \ln \beta - n \cdot \ln 2 - \sqrt{\beta} \cdot \sum_{i=1}^{n} \sqrt{x_i}.$$

$$(\ln L(\beta))' = \frac{n}{\beta} - \frac{1}{2\sqrt{\beta}} \sum_{i=1}^{n} \sqrt{x_i} = 0.$$

$$\Rightarrow \qquad \hat{\beta} = \left(\frac{2n}{\sum_{i=1}^{n} \sqrt{X_i}} \right)^2 = \frac{4n^2}{\left(\sum_{i=1}^{n} \sqrt{X_i} \right)^2}.$$

$n = 5$, $\quad x_1 = 0.09$, $\quad x_2 = 0.36$, $\quad x_3 = 1.96$, $\quad x_4 = 6.25$, $\quad x_5 = 10.24$.

$\sum_{i=1}^{n} \sqrt{x_i} = 8.$ $\qquad\qquad$ $\hat{\beta} = \left(\frac{2 \cdot 5}{8} \right)^2 = \mathbf{1.5625}$.

d) (7) Is the maximum likelihood estimator $\hat{\beta}$ an unbiased estimator of β? $\quad (n > 1)$

If $\hat{\beta}$ is not an unbiased estimator of β, construct an unbiased estimator of β from $\hat{\beta}$.

$W = \sqrt{X}$ has Gamma$\left(\alpha = 2,\ \theta = \frac{1}{\sqrt{\beta}}\right)$ distribution,

$$\Rightarrow \qquad Y = \sum_{i=1}^{n} \sqrt{X_i} = \sum_{i=1}^{n} W_i \text{ has a Gamma}\left(\alpha = 2n,\ \theta = \frac{1}{\sqrt{\beta}}\right) \text{ distribution.}$$

The maximum likelihood estimator for β is $\hat{\beta} = \dfrac{4n^2}{\left(\sum\limits_{i=1}^{n} \sqrt{X_i}\right)^2} = \dfrac{4n^2}{Y^2}$.

If T_α has a Gamma$(\alpha, \theta = 1/\lambda)$ distribution, then

$$E(T_\alpha^m) = \frac{\theta^m\, \Gamma(\alpha+m)}{\Gamma(\alpha)} = \frac{\Gamma(\alpha+m)}{\lambda^m\, \Gamma(\alpha)}, \qquad m > -\alpha.$$

$$E\left(\frac{1}{Y^2}\right) = E\left(Y^{-2}\right) = \frac{\Gamma(2n-2)}{\left(\sqrt{\beta}\right)^{-2}\, \Gamma(2n)} = \frac{\beta}{(2n-1)(2n-2)}.$$

$$E(\hat{\beta}) = E\left(\frac{4n^2}{Y^2}\right) = 4n^2\, E\left(\frac{1}{Y^2}\right) = 4n^2 \cdot \frac{\beta}{(2n-1)(2n-2)} \neq \beta.$$

$\hat{\beta}$ is NOT an unbiased estimator of β.

$$\hat{\hat{\beta}} = \frac{(2n-1)(2n-2)}{Y^2} = \frac{(2n-1)(2n-2)}{4n^2} \cdot \hat{\beta} = \frac{(2n-1)(2n-2)}{\left(\sum\limits_{i=1}^{n} \sqrt{X_i}\right)^2}$$

is an unbiased estimator of β.

e) (9) Suppose $n = 5$, and $x_1 = 0.09$, $x_2 = 0.36$, $x_3 = 1.96$, $x_4 = 6.25$, $x_5 = 10.24$.

Use $\sum_{i=1}^{n} \sqrt{X_i}$ to construct a 95% confidence interval for β.

$$Y = \sum_{i=1}^{n} \sqrt{X_i} = \sum_{i=1}^{n} W_i \text{ has a Gamma}\left(\alpha = 2n,\ \theta = \frac{1}{\sqrt{\beta}}\right) \text{ distribution.}$$

If T_α has a Gamma$\left(\alpha, \theta = {}^{1}\!/_{\lambda}\right)$ distribution,

then $\frac{2}{\theta} T_\alpha = 2\lambda T_\alpha$ has a $\chi^2(2\alpha)$ distribution.

$$\Rightarrow \quad 2\sqrt{\beta}\, Y = 2\sqrt{\beta} \sum_{i=1}^{n} \sqrt{X_i} \text{ has a } \chi^2(2\alpha = 4n) \text{ distribution.}$$

$$\Rightarrow \quad P\left(\chi^2_{1-\alpha/2}(4n) < 2\sqrt{\beta} \sum_{i=1}^{n} \sqrt{X_i} < \chi^2_{\alpha/2}(4n)\right) = 1 - \alpha.$$

$$\Rightarrow \quad P\left(\frac{\chi^2_{1-\alpha/2}(4n)}{2\sum_{i=1}^{n} \sqrt{X_i}} < \sqrt{\beta} < \frac{\chi^2_{\alpha/2}(4n)}{2\sum_{i=1}^{n} \sqrt{X_i}}\right) = 1 - \alpha.$$

$$\Rightarrow \quad P\left(\left(\frac{\chi^2_{1-\alpha/2}(4n)}{2\sum_{i=1}^{n} \sqrt{X_i}}\right)^2 < \beta < \left(\frac{\chi^2_{\alpha/2}(4n)}{2\sum_{i=1}^{n} \sqrt{X_i}}\right)^2\right) = 1 - \alpha.$$

A $(1-\alpha)\,100\,\%$ confidence interval for β:

$$\left(\left(\frac{\chi^2_{1-\alpha/2}(4n)}{2\sum_{i=1}^{n} \sqrt{X_i}}\right)^2, \left(\frac{\chi^2_{\alpha/2}(4n)}{2\sum_{i=1}^{n} \sqrt{X_i}}\right)^2\right)$$

$$n = 5, \qquad x_1 = 0.09, \quad x_2 = 0.36, \quad x_3 = 1.96, \quad x_4 = 6.25, \quad x_5 = 10.24.$$

$$\sum_{i=1}^{n} \sqrt{x_i} = 8.$$

$$\chi^2_{0.975}(20) = 9.591, \qquad \chi^2_{0.025}(20) = 34.17.$$

$$\left(\left(\frac{9.591}{2\cdot 8}\right)^2, \left(\frac{34.17}{2\cdot 8}\right)^2 \right) \approx \mathbf{(0.3593, 4.5609)}.$$

```
> qchisq(0.025,4*5)
[1] 9.590777
> (qchisq(0.025,4*5)/(2*8))^2
[1] 0.3593086
>
> qchisq(0.975,4*5)
[1] 34.16961
> (qchisq(0.975,4*5)/(2*8))^2
[1] 4.560789
```

f) (4) Show that $\mathrm{E}(\mathrm{X}^k) = \beta^{-k}\,\Gamma(2k+2)$ for $k > -1$.

No credit will be given without proper justification. Do NOT use technology.

You are welcome to use this result to help you with other parts of the problem.

$$\mathrm{E}(\mathrm{X}^k) = \mathrm{E}(\mathrm{W}^{2k}) = \ldots$$

$\mathrm{W} = \sqrt{\mathrm{X}}$ has Gamma$\left(\alpha = 2,\ \theta = \frac{1}{\sqrt{\beta}}\right)$ distribution. $\qquad \mathrm{W} = \mathrm{T}_2$.

If T_α has a Gamma$(\alpha, \theta = 1/\lambda)$ distribution, then

$$\mathrm{E}(\mathrm{T}_\alpha^m) = \frac{\theta^m\,\Gamma(\alpha+m)}{\Gamma(\alpha)} = \frac{\Gamma(\alpha+m)}{\lambda^m\,\Gamma(\alpha)}, \qquad m > -\alpha.$$

$m = 2k$

$$\ldots = \frac{\Gamma(2+2k)}{\left(\sqrt{\beta}\right)^{2k}\,\Gamma(2)} = \beta^{-k}\,\Gamma(2k+2),$$

$$2k > -2 \quad \Rightarrow \quad k > -1.$$

OR

$$\mathrm{E}(\mathrm{X}^k) = \int_0^\infty x^k \cdot \frac{\beta}{2}\, e^{-\sqrt{\beta x}}\, dx \qquad u = \sqrt{\beta x} \qquad du = \frac{\sqrt{\beta}}{2\sqrt{x}}\, dx$$

$$x = \frac{u^2}{\beta} \qquad dx = \frac{2u}{\beta}\, du$$

$$= \int_0^\infty \left(\frac{u^2}{\beta}\right)^k \cdot \frac{\beta}{2}\, e^{-u} \cdot \frac{2u}{\beta}\, du$$

$$= \frac{1}{\beta^k} \int_0^\infty u^{2k+1}\, e^{-u}\, du = \frac{1}{\beta^k} \int_0^\infty u^{(2k+2)-1}\, e^{-u}\, du = \frac{\Gamma(2k+2)}{\beta^k}.$$

g) (7) (i) Obtain a method of moments estimator for β, $\tilde{\beta}$.

(ii) Suppose $n = 5$, and $x_1 = 0.09$, $x_2 = 0.36$, $x_3 = 1.96$, $x_4 = 6.25$, $x_5 = 10.24$.

Find a method of moments estimate of β, $\tilde{\beta}$.

$$E(X) = E(X^1) = \frac{\Gamma(4)}{\beta} = \frac{3!}{\beta} = \frac{6}{\beta}.$$

OR

$$E(X) = E(W^2) = \mathrm{Var}(W) + [E(W)]^2 = \alpha\theta^2 + (\alpha\theta)^2 = \frac{2}{\beta} + \frac{4}{\beta} = \frac{6}{\beta}.$$

$$\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i = \frac{6}{\beta} \qquad \Rightarrow \qquad \tilde{\beta} = \frac{6}{\overline{X}} = \frac{6 \cdot n}{\sum_{i=1}^{n} X_i}.$$

$n = 5$, $x_1 = 0.09$, $x_2 = 0.36$, $x_3 = 1.96$, $x_4 = 6.25$, $x_5 = 10.24$.

$\sum_{i=1}^{n} x_i = 18.9.$ $\qquad \overline{x} = 3.78.$ $\qquad \tilde{\beta} = \frac{6}{3.78} \approx$ **1.5873**.

h) (4) Is your method of moments estimator $\tilde{\beta}$ an unbiased estimator of β?

If $\tilde{\beta}$ is not an unbiased estimator of β, does $\tilde{\beta}$ underestimate or overestimate β (on average)?

$$\tilde{\beta} = \frac{6}{\overline{X}}.$$

Consider $g(x) = \dfrac{6}{x}$.

Then $g(\overline{X}) = \tilde{\beta}$.

$$g'(x) = -\frac{6}{x^2}.$$

$$g''(x) = \frac{12}{x^3} > 0 \qquad \text{for } x > 0.$$

Since $g(x) = \dfrac{6}{x}$, $x > 0$, is strictly convex (that is, it curves up),

and $\overline{X}$ is not a constant random variable, by Jensen's Inequality (Theorem 1.10.5),

$$\mathrm{E}(\tilde{\beta}) = \mathrm{E}(g(\overline{X})) > g(\mathrm{E}(\overline{X})) = g(\mu).$$

$$g(\mu) = g\left(\frac{6}{\beta}\right) = \frac{6}{\left(\frac{6}{\beta}\right)} = \beta.$$

$$\Rightarrow \qquad \mathrm{E}(\tilde{\beta}) > \beta.$$

$\tilde{\beta}$ is NOT an unbiased estimator for β.

On average, $\tilde{\beta}$ **overestimates** β.

3. (5) Let $X_1, X_2, \ldots, X_n$ be a random sample from a probability distribution with mean $\mu = 3$ and unknown variance θ. Consider the following estimator for θ:

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - 3 \right)^2 .$$

Is $\hat{\theta}$ a consistent estimator of θ? *Justify your answer.*

No credit will be given without proper justification.

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - 3 \right)^2 \overset{\text{WLLN}}{\xrightarrow{P}} E[(X-3)^2] \overset{\mu = 3}{=} E[(X-\mu)^2] \overset{\text{definition}}{=} \operatorname{Var}(X) = \theta.$$

$\Rightarrow$ $\hat{\theta}$ is a consistent estimator of θ?

OR

For those who do not remember that

$$\operatorname{Var}(X) = E[(X-\mu)^2]$$

is the definition of the variance of a random variable,

but remember that

$$\operatorname{Var}(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \operatorname{Var}(X) + [E(X)]^2 = \theta + 3^2 = \theta + 9.$$

$$E[(X-3)^2] = E[X^2 - 6X + 9] = E(X^2) - 6E(X) + 9 = \theta + 9 - 6 \cdot 3 + 9 = \theta.$$

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - 3 \right)^2 \overset{\text{WLLN}}{\xrightarrow{P}} E[(X-3)^2] = \theta.$$

OR

$$E(X^2) = \text{Var}(X) + [E(X)]^2 = \theta + 3^2 = \theta + 9.$$

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - 3\right)^2 = \frac{1}{n}\left[\sum_{i=1}^{n} X_i^2 - 6\sum_{i=1}^{n} X_i + 9\right] = \overline{X^2} - 6\overline{X} + 9.$$

By WLLN, $\overline{X^2} \xrightarrow{P} E(X^2) = \theta + 9,$

$\overline{X} \xrightarrow{P} E(X) = 3.$

$$\hat{\theta} = \overline{X^2} - 6\overline{X} + 9 \xrightarrow{P} E(X^2) - 6E(X) + 9 = \theta + 9 - 6 \cdot 3 + 9 = \theta.$$

$\hat{\theta}$ is a consistent estimator of θ means $\hat{\theta} \xrightarrow{P} \theta.$

$E(\hat{\theta}) = \theta$ means $\hat{\theta}$ is an unbiased estimator of $\theta.$

unbiased does NOT imply consistent,

consistent does NOT imply unbiased.

Here, $\hat{\theta}$ is an unbiased estimator of θ:

$$E(\hat{\theta}) = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - 3\right)^2\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[\left(X_i - 3\right)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\theta = \theta.$$

However, this does NOT imply that $\hat{\theta}$ is a consistent estimator of θ.

MSE$(\hat{\theta})$ may not be finite, since we are not assuming $E(X^4) < \infty$.

4. For the students majoring in Philosophical Engineering at Anytown State University, the pre-COVID-19 GPA, X, and the COVID-19 era GPA, Y, jointly follow a bivariate normal distribution with

$$\mu_X = 3.07, \quad \sigma_X = 0.20, \quad \mu_Y = 2.95, \quad \sigma_Y = 0.25, \quad \rho = 0.80.$$

a) (7) What proportion of Philosophical Engineering majors at Anytown State University had their pre-COVID-19 GPA higher than their COVID-19 era GPA? That is, find $P(X > Y)$.

$P(X > Y) = P(X - Y > 0) = ?$

$X - Y$ has Normal distribution,

$E(X - Y) = \mu_X - \mu_Y = 3.07 - 2.95 = 0.12,$

$\operatorname{Var}(X - Y) = \sigma_X^2 - 2\sigma_{XY} + \sigma_Y^2 = \sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2$

$= 0.20^2 - 2 \cdot 0.80 \cdot 0.20 \cdot 0.25 + 0.25^2 = 0.0225$ (standard deviation 0.15).

$P(X - Y > 0) = P(Z > \frac{0 - 0.12}{0.15}) = P(Z > -0.80) = \mathbf{0.7881}.$

b) (7) For a student (a Philosophical Engineering majors at Anytown State University) with COVID-19 era GPA of 2.8, what is the probability that the pre-COVID-19 GPA was above 3.1? That is, find $P(X > 3.1 \mid Y = 2.8)$.

Given $Y = 2.8$, X has Normal distribution

with mean $3.07 + 0.80 \cdot \frac{0.20}{0.25} \cdot (2.8 - 2.95) = 2.974$

and variance $\left(1 - 0.80^2\right) \cdot 0.20^2 = 0.0144$ (standard deviation 0.12).

$P(X > 3.1 \mid Y = 2.8) = P(Z > \frac{3.1 - 2.974}{0.12}) = P(Z > 1.05) = \mathbf{0.1469}.$