1. If $X$ and $Y$ have joint p.d.f. $f(x,y) = c(1-x)y$, $0 \leq y \leq x \leq 1$, for some appropriate constant $c$, find $E[X]$.

**Solution:** First of all,

$$1 = \int_0^1 \int_0^x c(1-x)y \, dy \, dx = \frac{c}{2} \int_0^1 (x^2-x^3) \, dx = \frac{c}{24},$$

so that $c = 24$. Then

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y) \, dy = \int_0^x 24(1-x)y \, dy = 12(x^2-x^3), \quad 0 < x < 1.$$

Thus,

$$E[X] = \int_0^1 12(x^3-x^4) \, dx = 0.6. \quad \square$$

2. If $X$ and $Y$ are i.i.d. Pois(3) RV’s, find $\text{Var}(XY)$.

**Solution:**

$$\text{Var}(XY) = \text{E}[(XY)^2] - (\text{E}[XY])^2 = \text{E}[X^2\text{E}[Y^2]] - (\text{E}[X]\text{E}[Y])^2 \quad X,Y \text{ indep}$$

$$= (\text{E}[X^2])^2 - (\text{E}[X])^4 \quad X, Y \text{ i.i.d.}$$

$$= (\text{Var}(X) + (\text{E}[X])^2)^2 - (\text{E}[X])^4 \quad X, Y \text{ i.i.d.}$$

$$= (3 + 9)^2 - 3^4 = 63. \quad \square$$
3. TRUE or FALSE? The runs “up and down” test is most often used to test for goodness-of-fit of observations.

**Solution:** FALSE. It’s a test for independence. □

4. Draw an Arena **DECIDE** block.

**Solution:** It’s a sideways diamond. □

5. In what Arena template would you find the **Sequence** spreadsheet?

**Solution:** Advanced Transfer. □

6. In Arena, you can **LEAVE** a station. What is the analogous block to use to get into a station?

**Solution:** **ENTER**. □

7. Which method is *not* a viable way for an entity to **LEAVE** a resource? (i) Route; (ii) Connect; (iii) Move; (iv) Transport; (v) Convey. (There may be more than one correct answer.)

**Solution:** (iii) Move. □

8. TRUE or FALSE? Suppose that $U_1, U_2, \ldots$ are truly i.i.d. U(0,1) random variables. Then a 95% chi-square goodness-of-fit test for uniformity will *incorrectly* reject uniformity of the observations about 5% of the time.

**Solution:** TRUE. (That’s what Type I error is.) □

9. TRUE or FALSE? Suppose $Z_1, \ldots, Z_6$ are i.i.d. standard normal random variables obtained by the Box–Muller method. Then $\sum_{i=1}^{6} Z_i^2 \sim \text{Erlang}_3(1/2)$.
Solution: TRUE. The sum of 6 i.i.d. $Z_i^2$ random variables is a $\chi^2(6)$, which is itself an Erlang$_3(1/2)$.  

10. Suppose $X_1, X_2, \ldots, X_{100}$ are i.i.d. with $\Pr(X = 0) = \Pr(X = 2) = 0.5$. Define the sample mean $\bar{X} \equiv \sum_{i=1}^{100} X_i / 100$. Use the Central Limit Theorem to find an approximate expression for $\Pr(\bar{X} < 1.1)$.

Solution: Note that $E[X] = \sum_x x \Pr(X = x) = 1$ and $E[X^2] = \sum_x x^2 \Pr(X = x) = 2$, so that $\Var(X) = E[X^2] - (E[X])^2 = 1$. These results makes sense since $X \sim 2\text{Bern}(0.5)$.

Anyway, the CLT now implies that $\bar{X} \approx \text{Nor}(\mu, \sigma^2 / n) = \text{Nor}(1, 0.01)$, and so

$$\Pr(\bar{X} < 1.1) = \Pr\left(\frac{\bar{X} - 1}{\sqrt{0.01}} < \frac{1.1 - 1}{\sqrt{0.01}}\right) \approx \Pr(\text{Nor}(0, 1) < 1) = 0.8413.$$ 

11. If $U, V$ are i.i.d. $U(0,1)$, what’s the distribution of $-\ln(\sqrt{U}) - \ln(\sqrt{V})$?

Solution:

$$-\ln(\sqrt{U}) - \ln(\sqrt{V}) = \frac{1}{2} \ln(U) - \frac{1}{2} \ln(V) \sim \text{Erlang}_2(2).$$ 

12. TRUE or FALSE? If we want to generate $X \sim \text{Pois}(4.5)$, then it’s better to use a normal approximation than acceptance-rejection.

Solution: FALSE. (The normal approximation requires larger $\lambda$ for the asymptotics to work and the efficiency to kick in.)

13. TRUE or FALSE? The acceptance-rejection’s majorizing function $t(x)$ is usually a p.d.f.

Solution: FALSE. (It integrates to something $> 1$.)
14. Consider the stationary first-order exponential autoregressive process (EAR(1)),

\[ X_i = \begin{cases} \alpha X_{i-1}, & \text{w.p. } \alpha \\ \alpha X_{i-1} + \epsilon_i, & \text{w.p. } 1 - \alpha, \end{cases} \]

where \( X_0 \) and the \( \epsilon_i \)'s are i.i.d. \( \text{Exp}(\lambda) \), and \( 0 < \alpha < 1 \). Find \( \text{Cov}(X_0, X_1) \).

**Solution:**

\[
\text{Cov}(X_0, X_1) = \text{Cov}(X_0, \alpha X_0) \alpha + \text{Cov}(X_0, \alpha X_0 + \epsilon_1) (1 - \alpha) \\
= \alpha^2 \text{Var}(X_0) + \alpha (1 - \alpha) \text{Var}(X_0) + 0 \\
= \alpha / \lambda^2. \quad \square
\]

15. If \( X_1, \ldots, X_n \) are i.i.d. \( \text{Exp}(1/9) \), what is the expected value of the sample variance \( S^2 \)?

**Solution:** \( \mathbb{E}[S^2] = \sigma^2 = 81. \quad \square \)

16. If \( X_1, X_2, X_3 \) are i.i.d. normal, with \( X_1 = 3 \), \( X_2 = 2 \), and \( X_3 = 7 \), what is the MLE for \( \mathbb{E}[X_i^2] \)?

**Solution:** By invariance, the MLE is

\[
\hat{\mathbb{E}}[X_i^2] = \hat{\mu}^2 + \hat{\sigma}^2 = \bar{X}^2 + \frac{n-1}{n} S^2 = 4^2 + \frac{14}{3} = 20.67. \quad \square
\]

17. Find \( x \) such that \( e^{2x} = 1/x \). (Get within two decimals.)

**Solution:** Bisection search quickly reveals that \( x \) is about 0.4265. \quad \square

18. TRUE or FALSE? The square root of the sample variance is unbiased for the standard deviation.

**Solution:** FALSE. \( \mathbb{E}[S^2] = \sigma^2 \Rightarrow \mathbb{E}[S] = \sigma \). \quad \square
19. Suppose we’re conducting a $\chi^2$ goodness-of-fit test to determine whether or not 200 i.i.d. observations are from a Johnson distribution, which has 4 parameters that must be estimated. If we divide the observations into 10 equal-probability intervals, how many degrees of freedom will our test have?

**Solution:** $10 - 4 - 1 = 5$. □

20. We are interested in seeing if the number of emergency department visits occurring each day at the Georgia Tech clinic is Binomial(5,0.5). Below are the results for a 200-day period. We’ll assume that the numbers from day to day are i.i.d.

<table>
<thead>
<tr>
<th># of visits</th>
<th># of days</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

Thus, for example, there were 30 days during when the ED had exactly 1 visit.

We’ll perform a 95% $\chi^2$ goodness-of-fit test to see if the number of accidents each day is Binomial(5,0.5).

(a) How many intervals will you use for your test?

**Solution:** Let $X$ denote the number of visits on a particular day. Under the null hypothesis, the expected number of occurrences of $i$ visits is $E_i = n \Pr(X = i) = 200 \binom{5}{i} (0.5)^i$. So we have the following table.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$O_i$</th>
<th>$E_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
<td>6.25</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>31.25</td>
</tr>
<tr>
<td>2</td>
<td>61</td>
<td>62.5</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>62.5</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>31.25</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>6.25</td>
</tr>
</tbody>
</table>

Since all of the $E_i$’s are $\geq 5$, we can use all of the cells. Thus, the answer is 6 intervals. □
(b) What is your statistic value?

Solution: \( \chi^2_0 = \sum_{i=0}^{5} (O_i - E_i)^2 / E_i = 0.432. \) \( \square \)

(c) What is your conclusion? Binomial(5,0.5) or not?

Solution: Since the previous answer is so small, we don’t really need to look up the quantile, but I’ll do it anyway. The critical value is \( \chi^2_{\alpha,4} = \chi^2_{0.05,4} = 9.49 \), indicating a fail to reject. (So, yes, we’ll assume it’s Binomial.) \( \square \)

21. Consider the following PRN’s: 0.18, 0.92, 0.61, 0.33. If we use the Kolmorogov-Smirnov goodness-of-fit test to see if these numbers are U(0,1), what is the value of the test statistic?

Solution: Let’s make the usual table with the ordered PRN’s.

<table>
<thead>
<tr>
<th>( i )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R(i) )</td>
<td>0.18</td>
<td>0.33</td>
<td>0.61</td>
<td>0.92</td>
</tr>
<tr>
<td>( \frac{i}{n} - R(i) )</td>
<td>0.07</td>
<td>0.17</td>
<td>0.14</td>
<td>0.07</td>
</tr>
<tr>
<td>( R(i) - \frac{i-1}{n} )</td>
<td>0.18</td>
<td>0.08</td>
<td>0.11</td>
<td>0.18</td>
</tr>
</tbody>
</table>

This indicates that \( D^+_n = \max_i \left[ \frac{i}{n} - R(i) \right] = 0.17 \) and \( D^-_n = \max_i \left[ R(i) - \frac{i-1}{n} \right] = 0.18 \), so that \( D_n = \max(D^+, D^-) = 0.18 \). \( \square \)

22. Do the PRN’s in Question 21 pass the Kolmogorov-Smirnov goodness-of-fit test for uniformity at level \( \alpha = 0.05 \)?

Solution: From the one-sided table, we have \( D_{\alpha,n} = D_{0.05,4} = 0.565 \). Since \( D_n < D_{\alpha,n} \), we fail to reject uniformity. \( \square \)

23. Consider a stationary stochastic process \( X_1, X_2, \ldots \), with covariance function \( R_k = \text{Cov}(X_1, X_{1+k}) = 3 - k \) for \( k = 0, 1, 2, 3 \), and \( R_k = 0 \) for \( k \geq 4 \). Find \( \text{Var}(\bar{X}_4). \)
Solution: From class notes, we have

\[
\text{Var}(\bar{X}_n) = \frac{1}{n} \left[ R_0 + 2 \sum_{k=1}^{n-1} \left( 1 - \frac{k}{n} \right) R_k \right]
\]

\[
= \frac{1}{4} \left[ R_0 + 2\left(\frac{3}{4}\right)R_1 + 2\left(\frac{2}{4}\right)R_2 + +2\left(\frac{1}{4}\right)R_3 \right]
\]

\[
= \frac{1}{4} \left[ 3 + 2\left(\frac{3}{4}\right)2 + 2\left(\frac{2}{4}\right)1 + +2\left(\frac{1}{4}\right)0 \right]
\]

\[
= 1.75. \quad \square
\]

24. Consider the following (approximately normal) average waiting times from 4 independent replications of a complicated queueing network. Suppose that each output is based on the average of 500 waiting times:

\[
30 \quad 40 \quad 10 \quad 50
\]

Use the method of independent replications to calculate a two-sided 90% confidence interval for the mean \( \mu \).

Solution: We use \( b = 4 \) reps here. \( \bar{Z}_b = \sum_{i=1}^{b} Z_i/b = 32.5 \) and \( S_Z^2 = \frac{1}{b-1} \sum_{i=1}^{b} (Z_i - \bar{Z}_b)^2 = 291.67 \). Then the desired CI is

\[
\mu \in \bar{Z}_b \pm t_{\alpha/2,b-1} \sqrt{S_Z^2/b}
\]

\[
= 32.5 \pm t_{0.05,3} \sqrt{291.67/4}
\]

\[
= 32.5 \pm 2.353(8.539)
\]

\[
= 32.5 \pm 20.09 = [12.4, 52.6]. \quad \square
\]

25. Which is the method of batch means more appropriate for: terminating or steady-state simulations?

Solution: Steady-State. \quad \square

26. Which is usually a better way to deal with initialization bias in steady-state simulation analysis: (i) make an extremely long run to overwhelm the bias, or (ii)
perform truncation?

**Solution:** Truncate. □

27. Consider the following 10 snowfall totals in Buffoonalo, NY over consecutive years:

```
130 94 125 112 150 123 141 133 128 152
```

Use the method of batch means to calculate a two-sided 90% confidence interval for the mean $\mu$. In particular, use two batches of size 5.

**Solution:** We use $b = 2$ batches here. $\bar{X}_n = 128.8$, the batch means are $\bar{X}_{1,5} = 122.2$ and $\bar{X}_{2,5} = 135.4$, and the batch means estimator for the variance parameter is

$$\hat{V}_B = \frac{m}{b-1} \sum_{i=1}^{b} (\bar{X}_{i,m} - \bar{X}_n)^2 = 5 \sum_{i=1}^{2} (\bar{X}_{i,5} - \bar{X}_{10})^2 = 435.6.$$

Then the desired CI is

$$\mu \in \bar{X}_n \pm t_{\alpha/2, b-1} \sqrt{\hat{V}_B/n} = 128.8 \pm t_{0.05, 9} \sqrt{435.6/10} = 128.8 \pm 6.314(6.6) = 128.8 \pm 41.67 = [87.1, 170.5].$$

28. Suppose $[0, 1]$ is a 90% confidence interval for the mean $\mu$ based on 10 independent replications of size 1000. Now the boss has decided that she wants a 99% CI for $2\mu$ based on those same 10 replications of size 1000. What is it?

**Solution:** Let’s get a 99% CI for $\mu$ first. To begin with, the original 90% CI for $\mu$ is

$$\mu \in \bar{Z}_b \pm t_{0.05, b-1} \sqrt{\frac{S_{Z}^2}{b}} = 0.5 \pm 0.5.$$

This implies that the 99% CI for $\mu$ will have half-length

$$t_{0.005, b-1} \sqrt{\frac{S_{Z}^2}{b}} \times t_{0.005, b-1} \frac{S_{Z}^2}{b} = \frac{t_{0.005, 9}(0.5)}{t_{0.05, 9}} = \frac{3.250}{1.833}(0.5) = 0.887.$$
Now we can get the 99% CI for \( \mu \):

\[
\mu \in \bar{Z}_b \pm t_{0.005,b-1}\sqrt{\frac{S^2}{b}} = 0.5 \pm 0.887.
\]

This immediately implies that the 99% CI for \( 2\mu \) is

\[
2\mu \in 1 \pm 1.77 = [-0.77, 2.77]. \quad \square
\]

29. Suppose I use the method of overlapping batch means with sample size \( n = 10000 \) and batch size \( m = 500 \). Approximately how many degrees of freedom will the resulting variance estimator have?

**Solution:** Denote \( b = n/m = 20 \). You get approximately \( \frac{3}{2}(b - 1) = 28.5 \) d.f. (Will also accept \( 3b/2 = 30 \) or anything reasonably close.) \( \square \)

30. If \( \mathcal{W}(t) \) is a standard Brownian motion process and \( a < b \), find \( \Pr(\mathcal{W}(a) < \mathcal{W}(b)) \).

**Solution:** \( \mathcal{W}(a) - \mathcal{W}(b) \) is normally distributed with mean 0. This immediately yields a probability of 1/2. \( \square \)

31. Suppose that \( A = \int_0^1 B(t) \, dt \) is the area under a Brownian bridge process. Find \( \Pr(A > 1/\sqrt{12}) \).

**Solution:** From class notes, we have \( A \sim \text{Nor}(0, \frac{1}{12}) \). Then

\[
\Pr(A > 1/\sqrt{12}) = \Pr(\text{Nor}(0, 1) > 1) = 0.1587. \quad \square
\]

32. We are studying the waiting times arising from two queueing systems. Suppose we make 4 independent replications of both systems, where the systems are simulated independently of each other. Assuming that the average waiting time results from each replication are approximately normal, find a two-sided 95% CI for the difference in the means of the two systems.

<table>
<thead>
<tr>
<th>replication</th>
<th>system 1</th>
<th>system 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>
Solution: This is a two-sample CI problem assuming unknown and unequal variances. We have $\bar{X} = 16.25$, $\bar{Y} = 26.25$, $S^2_X = 122.917$, and $S^2_Y = 156.25$. The estimated d.f. is

$$
\nu \equiv \frac{\left( \frac{S^2_X}{n} + \frac{S^2_Y}{m} \right)^2}{\frac{(S^2_X/n)^2}{n+1} + \frac{(S^2_Y/m)^2}{m+1}} - 2 = 7.86 = 7.
$$

Then the appropriate CI is

$$
\mu_X - \mu_Y \in \bar{X} - \bar{Y} \pm t_{\alpha/2,\nu} \sqrt{\frac{S^2_X}{n} + \frac{S^2_Y}{m}}
= -10 \pm t_{0.025,7} \sqrt{\frac{122.9}{n} + \frac{156.3}{m}}
= -10 \pm 2.365(8.355)
= -10 \pm 19.76 = [-29.76, 9.76], \quad \Box
$$

33. This is sort of the same as Question 32, except we have now used common random numbers to induce positive correlation between the results of the two systems. Again find a two-sided 95% CI for the difference in the means of the two systems.

<table>
<thead>
<tr>
<th>replication</th>
<th>system 1</th>
<th>system 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>40</td>
</tr>
</tbody>
</table>

Solution: This is a paired-$t$ CI problem assuming unknown variance of the differences.

<table>
<thead>
<tr>
<th>replication</th>
<th>$X_i$</th>
<th>$Y_i$</th>
<th>$D_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>25</td>
<td>-15</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>30</td>
<td>-10</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>10</td>
<td>-5</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>40</td>
<td>-10</td>
</tr>
</tbody>
</table>
Now,
\[
\mu_X - \mu_Y \in \bar{D} \pm t_{\alpha/2,n-1} \sqrt{\frac{S_D^2}{n}}
\]
\[
= -10 \pm t_{0.025,3} \sqrt{\frac{16.67}{4}}
\]
\[
= -10 \pm 3.18(2.041)
\]
\[
= -10 \pm 6.50 = [-16.50, -3.5]. \quad \Box
\]

34. Let’s use the basic Monte Carlo technique from class to integrate \( I = \int_0^1 e^x \, dx \).

(a) First of all, what is the exact value of \( I \)?

**Solution:** \( I = e - 1 = 1.718. \quad \Box \)

(b) Use the PRN’s 0.95, 0.63, 0.15, and 0.42 to estimate \( I \).

**Solution:** \( \hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} e^{U_i} = 1.787. \quad \Box \)

(c) Now use antithetics to estimate \( I \).

**Solution:** \( \tilde{I}_n = \frac{1}{n} \sum_{i=1}^{n} e^{1-U_i} = 1.656. \quad \Box \)

(d) Combine your last two answers.

**Solution:** \( \bar{I}_n = \frac{1}{2}(\hat{I}_n + \tilde{I}_n) = 1.721, \) which is a great answer. \( \Box \)

35. Suppose that I’m interested in selecting the most popular television show during a particular time period. What kind of selection problem is this — (a) normal, (b) multinomial, or (c) Bernoulli?

**Solution:** (b). \( \Box \)