NAME →

ISyE 6644 — Spring 2018 — Practice Test #3

Hi One Last Time, Gentle Class!

Our Test 3 will cover *every freaking thing in the course*! However, I'll place more weight on the new stuff since Test 2.

Practice Test 3 below is actually an augmentation of the Test 3 originally given in my "live" 6644 class during the Fall 2013 semester. Your Test 3 won't be as long. You'll get to have 3 cheat sheets (6 sides total) and plenty of time. ☺

Also, all of our test questions will be multiple choice, in spite of what you see below.

Dave

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

   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

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

   Thus,

$$\mathsf{E}[X] \;=\; \int_0^1 12(x^3 - x^4)\,dx \;=\; 0.6. \quad \Box$$

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

   Solution:

$$\begin{aligned}
\mathsf{Var}(XY) &= \mathsf{E}[(XY)^2] - (\mathsf{E}[XY])^2 \\
&= \mathsf{E}[X^2]\mathsf{E}[Y^2] - (\mathsf{E}[X]\mathsf{E}[Y])^2 \quad X, Y \text{ indep} \\
&= (\mathsf{E}[X^2])^2 - (\mathsf{E}[X])^4 \quad X, Y \text{ i.i.d.} \\
&= (\mathsf{Var}(X) + (\mathsf{E}[X])^2)^2 - (\mathsf{E}[X])^4 \quad X, Y \text{ i.i.d.} \\
&= (3+9)^2 - 3^4 \;=\; 63. \quad \Box
\end{aligned}$$

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. $\Box$

4. Draw an Arena `DECIDE` module.

   Solution: It's a sideways diamond. $\Box$

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

   Solution: `Advanced Transfer`. $\Box$

6. In Arena, what module can you use to merge multiple entities into one?

   Solution: `Batch`. $\Box$

7. Arena Program (this will take a little work): Explicitly describe and show how to implement in Arena the following inventory system. Customers arrive according to a Poisson process at the rate of 10/day. Each customer demands a discrete uniform number of items from 1 to 8 (like an 8-sided die toss). When inventory goes below 50, we order 200 items. We allow for backlogs and lead-times. I want appropriate English verbiage and Arena block diagrams with adequate explanation of the model. Also discuss what stats you'd collect and how you'd do it.

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.) $\Box$

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 $\text{Erlang}_3(1/2)$. $\Box$

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

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

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

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

    Solution:

    $$-\ell\text{n}(\sqrt{U}) - \ell\text{n}(\sqrt{V}) \;=\; -\frac{1}{2}\ell\text{n}(U) - \frac{1}{2}\ell\text{n}(V) \;\sim\; \text{Erlang}_2(2). \quad \Box$$

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 λ for the asymptotics to work and the efficiency to kick in.) $\square$

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.) $\square$

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 ϵ_i's are i.i.d. Exp(λ), and $0 < \alpha < 1$. Find $\mathsf{Cov}(X_0, X_1)$.

    Solution:

$$\begin{aligned} \mathsf{Cov}(X_0, X_1) &= \mathsf{Cov}\Big(X_0,\, \alpha X_0\Big)\alpha + \mathsf{Cov}\Big(X_0,\, \alpha X_0 + \epsilon_1\Big)(1-\alpha) \\ &= \alpha^2 \mathsf{Var}(X_0) + \alpha(1-\alpha)\mathsf{Var}(X_0) + 0 \\ &= \alpha/\lambda^2. \quad \square \end{aligned}$$

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

    Solution: $\mathsf{E}[S^2] = \sigma^2 = 81.$ $\square$

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 $\mathsf{E}[X_i^2]$?

    Solution: By invariance, the MLE is

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

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

    Solution: Bisection search quickly reveals that x is about 0.4265. $\Box$

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

    Solution: FALSE. ($\mathsf{E}[S^2] = \sigma^2 \not\Rightarrow \mathsf{E}[S] = \sigma$.) $\Box$

19. Suppose we're conducting a χ^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: $10 - 4 - 1 = 5$. $\Box$

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.

# of visits	# of days
0	6
1	30
2	61
3	62
4	34
5	7

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

We'll perform a 95% χ^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 $E_i = n\,\Pr(X = i) = 200\binom{5}{i}(0.5)^5$. So we have the following table.

i	O_i	E_i
0	6	6.25
1	30	31.25
2	61	62.5
3	62	62.5
4	34	31.25
5	7	6.25

Since all of the E_i's are ≥ 5, we can use all of the cells. Thus, the answer is 6 intervals. $\square$

(b) What is your statistic value?

Solution: $\chi_0^2 = \sum_{i=0}^{5}(O_i - E_i)^2/E_i = 0.432.$ $\square$

(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 $\chi^2_{\alpha,k-1} = \chi^2_{0.05,5} = 11.07$, 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.

i	1	2	3	4
$R_{(i)}$	0.18	0.33	0.61	0.92
$\frac{i}{n} - R_{(i)}$	0.07	0.17	0.14	0.07
$R_{(i)} - \frac{i-1}{n}$	0.18	0.08	0.11	0.18

This indicates that $D_n^+ = \max_i[\frac{i}{n} - R_{(i)}] = 0.17$ and $D_n^- = \max_i[R_{(i)} - \frac{i-1}{n}] = 0.18$, 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 $R_k = \mathsf{Cov}(X_1, X_{1+k}) = 3 - k$ for $k = 0, 1, 2, 3$, and $R_k = 0$ for $k \geq 4$. Find $\mathsf{Var}(\bar{X}_4)$.

    Solution: From class notes, we have

$$\begin{aligned}
\mathsf{Var}(\bar{X}_n) &= \frac{1}{n}\left[R_0 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right) R_k\right] \\
&= \frac{1}{4}\left[R_0 + 2(\frac{3}{4})R_1 + 2(\frac{2}{4})R_2 + +2(\frac{1}{4})R_3\right] \\
&= \frac{1}{4}\left[3 + 2(\frac{3}{4})2 + 2(\frac{2}{4})1 + +2(\frac{1}{4})0\right] \\
&= 1.75. \quad \Box
\end{aligned}$$

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 40 10 50

    Use the method of independent replications to calculate a two-sided 90% confidence interval for the mean μ.

    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

$$\begin{aligned}
\mu &\in \bar{Z}_b \pm t_{\alpha/2,b-1}\sqrt{S_Z^2/b} \\
&= 32.5 \pm t_{0.05,3}\sqrt{291.67/4} \\
&= 32.5 \pm 2.353(8.539) \\
&= 32.5 \pm 20.09 \;=\; [12.4,\ 52.6]. \quad \Box
\end{aligned}$$

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

    Solution: Steady-State. $\Box$

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. $\Box$

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 μ. In particular, use two batches of size 5.

Solution: We use $b = 2$ batches here 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

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

Then the desired CI is

$$\begin{aligned}
\mu &\in \bar{X}_n \pm t_{\alpha/2,b-1}\sqrt{\widehat{V}_B/n} \\
&= 128.8 \pm t_{0.05,1}\sqrt{435.6/10} \\
&= 128.8 \pm 6.314(6.6) \\
&= 128.8 \pm 41.67 \;=\; [87.1,\ 170.5]. \quad \Box
\end{aligned}$$

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

Solution: Let's get a 99% CI for μ first. To begin with, the original 90% CI for μ 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 μ will have half-length

$$t_{0.005,\,b-1}\sqrt{\frac{S_Z^2}{b}} \;=\; \frac{t_{0.005,\,b-1}}{t_{0.05,\,b-1}}t_{0.05,\,b-1}\sqrt{\frac{S_Z^2}{b}} \;=\; \frac{t_{0.005,\,9}}{t_{0.05,\,9}}(0.5) \;=\; \frac{3.250}{1.833}(0.5) \;=\; 0.887.$$

Now we can get the 99% CI for μ:

$$\mu \in \bar{Z}_b \pm t_{0.005,\,b-1}\sqrt{\frac{S_Z^2}{b}} \;=\; 0.5 \pm 0.887.$$

This immediately implies that the 99% CI for 2μ is

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

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.) $\Box$

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. $\Box$

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

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

$$\Pr(A > 1/\sqrt{12}) \;=\; \Pr(\mathsf{Nor}(0,1) > 1) \;=\; 0.1587. \quad \Box$$

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.

replication	system 1	system 2
1	10	25
2	20	10
3	5	40
4	30	30

Solution: This is a two-sample CI problem assuming unknown and unequal variances. We have $\bar{X} = 16.25$, $\bar{Y} = 26.25$, $S_X^2 = 122.917$, and $S_Y^2 = 156.25$. The estimated d.f. is

$$\nu \;\equiv\; \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{(S_X^2/n)^2}{n+1} + \frac{(S_Y^2/m)^2}{m+1}} - 2 \;=\; 7.86 \;=\; 7.$$

Then the appropriate CI is

$$\begin{aligned}
\mu_X - \mu_Y \;&\in\; \bar{X} - \bar{Y} \pm t_{\alpha/2,\nu}\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}} \\
&= -10 \pm t_{0.025,7}\sqrt{\frac{122.9}{4} + \frac{156.3}{4}} \\
&= -10 \pm 2.365(8.355) \\
&= -10 \pm 19.76 \;=\; [-29.76,\, 9.76]. \quad \Box
\end{aligned}$$

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.

replication	system 1	system 2
1	10	25
2	20	30
3	5	10
4	30	40

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

replication	X_i	Y_i	D_i
1	10	25	-15
2	20	30	-10
3	5	10	-5
4	30	40	-10

Now,

$$\begin{aligned}
\mu_X - \mu_Y &\in \bar{D} \pm t_{\alpha/2,n-1}\sqrt{\frac{S_D^2}{n}} \\
&= -10 \pm t_{0.025,3}\sqrt{\frac{16.67}{4}} \\
&= -10 \pm 3.182(2.041) \\
&= -10 \pm 6.50 \;=\; [-16.50, -3.5]. \quad \square
\end{aligned}$$

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: $I = e - 1 = 1.718$. $\square$

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

   Solution: $\hat{I}_n = \frac{1}{n}\sum_{i=1}^n e^{U_i} = 1.787$. $\square$

   (c) Now use antithetics to estimate I.

   Solution: $\tilde{I}_n = \frac{1}{n}\sum_{i=1}^n e^{1-U_i} = 1.656$. $\square$

   (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. $\square$

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). $\square$

36. There may also be a couple additional easy ranking and selection questions from Module 10.