NAME $\rightarrow$

ISyE 3044 — Test 3 Solutions — Fall 2012

(Revised 12/5/13)

You have 120 minutes. You are allowed 3 cheat sheets. Good Luck!!

1. (3 pts each) Short-answer questions on less-recent topics — Just write your answer.

   (a) If X and Y have joint p.d.f. $f(x,y) = cxy$, $0 \le x \le 2$, $0 \le y \le 1$, for some appropriate constant c, find $\mathsf{E}[X]$.

   Solution: Since $f(x,y) = a(x)b(y)$ for all x and y, the RV's X and Y are independent. Thus, we can write $f(x) = c_1 x$ for $0 \le x \le 2$.

   To solve for c_1, set $1 = \int_0^2 c_1 x\,dx = 2c_1$, so that $c_1 = 1/2$. Then get $\mathsf{E}[X] = \int_0^2 x f_X(x)\,dx = \frac{1}{2}\int_0^2 x^2\,dx = 4/3$. $\diamondsuit$

   (b) TRUE or FALSE? If X is a continuous random variable that's always positive, we have $\mathsf{E}[\ell\mathrm{n}(X)] = \int_0^\infty \ell\mathrm{n}(x)\mathsf{P}(X > x)\,dx$.

   Solution: FALSE. $\diamondsuit$

   (c) TRUE or FALSE? The Kolmogorov-Smirnov test is a test of independence.

   Solution: FALSE. It's a goodness-of-fit test. $\diamondsuit$

   (d) TRUE or FALSE? In an Arena `PROCESS` block, it is possible to initiate a `DELAY` *without* a `SEIZE` or a `RELEASE`.

   Solution: TRUE. $\diamondsuit$

   (e) TRUE or FALSE? An Arena `ASSIGN` block can be used to change an entity's picture.

   Solution: TRUE. $\diamondsuit$

(f) In Arena, you `SEIZE` a resource. What analogous thing do you do to a conveyor?

Solution: `ACCESS`. $\diamondsuit$

(g) What is the variance of the random variable generated by the Arena function `NORM(3,4)`?

Solution: 16. $\diamondsuit$

(h) Suppose that a Tausworthe generator gave you the series of bits 01111110. If you use all 8 bits, what Unif(0,1) random number would that translate to?

Solution: $\frac{01111110_2}{2^8} = \frac{126}{256} = 0.4922.$ $\diamondsuit$

(i) TRUE or FALSE? Suppose that $U_1, U_2, \ldots$ are truly i.i.d. Unif(0,1) random variables. Then a χ^2 goodness-of-fit test for uniformity with $\alpha = 0.1$ will fail to reject about 10% of the time.

Solution: FALSE. It will incorrectly reject about 10% of the time. $\diamondsuit$

(j) TRUE or FALSE? If Z_1 and Z_2 are i.i.d. Nor(0,1), then the ratio Z_1/Z_2 has both the Cauchy distribution and the t distribution with one degree of freedom.

Solution: TRUE. They're the same. $\diamondsuit$

(k) Phun Phact: If X and Z are i.i.d. standard normals, then the ratio $(X+\delta)/Z$ is said to have the *noncentral t distribution* with one degree of freedom and noncentrality parameter δ. Suppose that $X = -1.0$ and $Z = 0.25$ are two standard normal outcomes. Generate a noncentral t random variable with one degree of freedom and noncentrality parameter 1.

Solution: $(X+\delta)/Z = (-1+1)/0.25 = 0$. I guess I would also accept $(0.25+1)/-1 = -1.25$ $\diamondsuit$

(l) Suppose $X_1, X_2, \ldots, X_{100}$ are i.i.d. Bern(0.1). Further, denote the sample mean by $\bar{X} \equiv \sum_{i=1}^{100} X_i/100$. Use the CLT to find an approximate expression for $\mathsf{P}(0.07 \le \bar{X} \le 0.13)$.

Solution: By the CLT,

$$\bar{X} \approx \text{Nor}(\mathsf{E}[\bar{X}], \mathsf{Var}(\bar{X})) \sim \text{Nor}(\mathsf{E}[X_i], \mathsf{Var}(X_i)/n) \sim \text{Nor}(p, pq/n).$$

Thus, $\bar{X} \approx \text{Nor}(0.1, 0.0009)$. Standardizing yields

$$\begin{aligned} &\mathsf{P}(0.07 \le \bar{X} \le 0.13) \\ &= \mathsf{P}\left(\frac{0.07 - 0.1}{\sqrt{0.0009}} \le \frac{\bar{X} - 0.1}{\sqrt{0.0009}} \le \frac{0.13 - 0.1}{\sqrt{0.0009}}\right) \\ &\approx \mathsf{P}(-1 \le Z \le 1) = 0.683. \quad \diamondsuit \end{aligned}$$

(m) Suppose $X_1, X_2, \ldots$ is a stationary process with mean μ and variance parameter $\sigma^2 \equiv \lim_{n\to\infty} n\mathsf{Var}(\bar{X})$, where the sample mean $\bar{X} \equiv \frac{1}{n}\sum_{i=1}^n X_i$. What colorful stochastic process does $\sum_{i=1}^{\lfloor nt \rfloor}(X_i - \mu)/(\sigma\sqrt{n})$ converge to (as a function of t)? **Hint:** The answer is *not* Redian motion or Yellowian motion.

Solution: Brownian motion. $\diamondsuit$

2. (3 pts each) Short-answer questions on more-recent topics — Just write your answer.

(a) If $X_1, \ldots, X_n$ are i.i.d. $\text{Bin}(6, 0.2)$, what is the expected value of the sample variance S^2?

Solution: $\mathsf{E}[S^2] = \mathsf{Var}(X_i) = npq = 0.96$. $\diamondsuit$

(b) If $X_1, \ldots, X_{10}$ are i.i.d. $\text{Nor}(-3, 10)$, what is the expected value of the maximum likelihood estimator for the variance σ^2?

Solution: $\mathsf{E}[\widehat{\sigma^2}] = \frac{n-1}{n}\mathsf{E}[S^2] = \frac{n-1}{n}\sigma^2 = 9$. $\diamondsuit$

(c) Find the sample variance of -4, 0, 4.

Solution: $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 = 16$. $\diamondsuit$

(d) Suppose we observe the Geom(p) realizations $X_1 = 3$, $X_2 = 9$, and $X_3 = 6$. What is the maximum likelihood estimate of p?

Solution: The likelihood function is

$$L(p) \;=\; \prod_{i=1}^{n} f(x_i) \;=\; \prod_{i=1}^{n} q^{x_i - 1} p \;=\; q^{\sum_{i=1}^{n} x_i - n} p^n.$$

Thus,

$$\begin{aligned} \ell\text{n}(L(p)) &= \Big(\sum_{i=1}^{n} x_i - n\Big)\ell\text{n}(q) + n\ell\text{n}(p). \\ &= \Big(\sum_{i=1}^{n} x_i - n\Big)\ell\text{n}(1-p) + n\ell\text{n}(p). \end{aligned}$$

This implies that

$$\frac{d}{dp}\ell\text{n}(L(p)) \;=\; \frac{-\Big(\sum_{i=1}^{n} x_i - n\Big)}{1-p} + \frac{n}{p}.$$

Setting the derivative to 0 and solving yields $\hat{p} = 1/\bar{x} = 1/6$. $\diamondsuit$

(e) Again suppose we observe the Geom(p) realizations $X_1 = 3$, $X_2 = 9$, and $X_3 = 6$. What's the maximum likelihood estimate of $\ell\text{n}(p)$?

Solution: By invariance,

$$\widehat{\ell\text{n}(p)} \;=\; \ell\text{n}(\hat{p}) \;=\; \ell\text{n}(1/\bar{x}) \;=\; \ell\text{n}(1/6) \;=\; -1.79. \quad \diamondsuit$$

(f) TRUE or FALSE? The mean squared error of an estimator is square of its variance plus the bias.

Solution: FALSE. It's the var + bias^2. $\diamondsuit$

(g) Find the MLE for σ if X has p.d.f.

$$f(x) \;=\; \frac{\sigma}{x\sqrt{\pi}} \exp\Big\{-\sigma^2[\ell\text{n}(x)]^2\Big\}, \quad x \ge 0.$$

Solution: The likelihood function is

$$L(\sigma) \;=\; \prod_{i=1}^{n} f(x_i) \;=\; \frac{\sigma^n}{(\prod_{i=1}^{n} x_i)\pi^{n/2}} \exp\left\{-\sigma^2 \sum_{i=1}^{n} [\ell\text{n}(x_i)]^2\right\}.$$

Thus,

$$\ell\text{n}(L(\sigma)) \;=\; n\ell\text{n}(\sigma) - \ell\text{n}\left(\prod_{i=1}^{n} x_i\right) - \frac{n}{2}\ell\text{n}(\pi) - \sigma^2 \sum_{i=1}^{n} [\ell\text{n}(x_i)]^2.$$

This implies that

$$\frac{d}{d\sigma}\ell\text{n}(L(\sigma)) \;=\; \frac{n}{\sigma} - 2\sigma \sum_{i=1}^{n} [\ell\text{n}(x_i)]^2.$$

Setting the derivative to 0 and solving yields

$$\hat{\sigma} \;=\; \sqrt{\frac{n}{2\sum_{i=1}^{n} \ell\text{n}^2(x_i)]}}. \quad \diamondsuit$$

(h) Suppose we're conducting a χ^2 goodness-of-fit test to determine whether or not 100 i.i.d. observations are from a *Johnson* distribution with unknown parameters α, β, γ, and δ, all of which must be estimated. If we divide the observations into 20 equal-probability intervals, how many degrees of freedom will our test have?

Solution: $\nu = n - s - 1 = 20 - 4 - 1 = 15$. $\diamondsuit$

(i) TRUE or FALSE? Newton's method can help you find the zeros of a continuous function $g(x)$, but you need to know the derivative $g'(x)$.

Solution: TRUE. $\diamondsuit$

(j) Do you use the method of batch means for (a) terminating simulations or (b) steady-state simulations (choose one)?

Solution: steady-state. $\diamondsuit$

(k) Consider the following 6 consecutive observations arising from a simulation:

$$154 \quad 180 \quad 175 \quad 173 \quad 191 \quad 183$$

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 three.

Solution: The batch size is $m = 3$, the number of batches is $b = 2$, and the total number of observations is $n = 6$. The grand sample mean is $\bar{X}_n = 176$. The batch means are

$$\bar{X}_{1,3} = 169.67 \quad \text{and} \quad \bar{X}_{2,3} = 182.33.$$

The batch means variance estimator is

$$\widehat{V}_B \;=\; \frac{m}{b-1}\sum_{i=1}^{b}(\bar{X}_{i,m}-\bar{X}_n)^2 \;=\; \frac{3}{1}\Big[(169.67-176)^2+(182.33-176)^2\Big] \;=\; 240.41.$$

The batch means confidence interval is

$$\begin{aligned}
\mu &\in \bar{X}_n \pm t_{\alpha/2,b-1}\sqrt{\widehat{V}_B/n} \\
&= 176 \pm t_{0.05,1}\sqrt{240.41/6} \\
&= 176 \pm 6.314(6.330) \\
&= 176 \pm 39.97 \;=\; [136,\ 216]. \quad \diamondsuit
\end{aligned}$$

(l) Consider a particular data set of 60000 stationary waiting times obtained from a large queuing system. Suppose your goal is to get a confidence interval for the unknown mean. Would you rather use (a) 30 batches of 2000 observations or (b) 6000 batches of 10 observations each?

Solution: (a). $\diamondsuit$

(m) Suppose $[-2, 2]$ is a 95% nonoverlapping batch means confidence interval for the mean μ based on 10 batches of size 500. Now the boss has decided that she wants a 90% CI based on those same 10 batches of size 500. What is it?

Solution: The confidence interval is of the form

$$[0, 4] \;=\; \bar{X}_n \pm t_{\alpha/2,b-1}\sqrt{\widehat{V}_B/n}.$$

This implies that $\bar{X}_n = 0$ and the half-length is $t_{0.025,9}\sqrt{\widehat{V}_B/n} = 2$. Thus, the new 90% confidence interval is

$$\begin{aligned}\text{new CI} &= \bar{X}_n \pm t_{0.05,9}\sqrt{\widehat{V}_B/n} \\ &= 0 \pm \frac{t_{0.05,9}}{t_{0.025,9}} t_{0.025,9}\sqrt{\widehat{V}_B/n} \\ &= \pm\frac{1.833}{2.262} \times 2 \;=\; \pm 1.621. \quad \diamondsuit\end{aligned}$$

(n) Consider the following observations:

54 80 75 62

If we choose a batch size of 3, calculate all of the overlapping batch means for me.

Solution: $\bar{X}^{\mathrm{o}}_{1,3} = \frac{1}{3}\sum_{i=1}^{3} X_i = 69.67$ and $\bar{X}^{\mathrm{o}}_{2,3} = \frac{1}{3}\sum_{i=2}^{4} X_i = 72.33$. $\diamondsuit$

(o) Suppose that $X_1, X_2, \ldots$ is a stationary stochastic process with covariance function $R_k \equiv \mathsf{Cov}(X_1, X_{1+k})$, for $k = 0, 1, \ldots$. We know from class that the variance of the sample mean can be represented as

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

We also know from class that for a simple AR(1) process, $R_i = \phi^i$, $i = 0, 1, 2, \ldots$. Compute $\mathsf{Var}(\bar{X}_n)$ for an AR(1) process with $n = 3$ and $\phi = 0.9$.

Solution:

$$\begin{aligned}\mathsf{Var}(\bar{X}_n) &= \frac{1}{3}\left[R_0 + 2\sum_{k=1}^{2}\left(1 - \frac{k}{3}\right) R_k\right] \\ &= \frac{1}{3}\left[R_0 + 2\left(1 - \frac{1}{3}\right) R_1 + 2\left(1 - \frac{2}{3}\right) R_2\right] \\ &= \frac{1}{3}\left[\phi^0 + \frac{4}{3}\phi^1 + \frac{2}{3}\phi^2\right] \\ &= \frac{1}{3}\left[1 + \frac{4}{3}(0.9) + \frac{2}{3}(0.81)\right] \;=\; 0.913. \quad \diamondsuit\end{aligned}$$

(p) TRUE or FALSE? Using the notation of the previous question, $\lim_{n\to\infty} n\mathsf{Var}(\bar{X}_n) = R_0 + 2\sum_{i=1}^{\infty} R_i$.

Solution: TRUE. Just let the n's get big (though you have to be a little non-rigorous in this class.) $\diamondsuit$

(q) We know from class that both $X = -\ell\text{n}(U)$ and $Y = -\ell\text{n}(1-U)$ are Exp(1) RV's. What I may not have told you is that X and Y aren't independent. Before I ask my question, fill out the following table.

i	1	2	3	4
U_i	0.05	0.72	0.94	0.36
$X_i = -\ell\text{n}(U_i)$				
$Y_i = -\ell\text{n}(1-U_i)$				

After having filled out the table, I'd like you to calculate the usual estimate for covariance,

$$\widehat{\text{Cov}}(X,Y) = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right].$$

Report only your answer for $\widehat{\text{Cov}}(X,Y)$ on the answer sheet.

Solution: Let's first fill in the table.

i	1	2	3	4
U_i	0.05	0.72	0.94	0.36
$X_i = -\ell\text{n}(U_i)$	2.996	0.329	0.062	1.022
$Y_i = -\ell\text{n}(1-U_i)$	0.051	1.273	2.813	0.446

Now,

$$\begin{aligned}\widehat{\text{Cov}}(X,Y) &= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right] \\ &= \frac{1}{3}[1.202 - 4(1.102)(1.146)] = -1.283. \quad \diamondsuit\end{aligned}$$

By the way, it can also be shown that the estimated correlation between X and Y is

$$\widehat{\rho} = \frac{\widehat{\text{Cov}}(X,Y)}{S_X S_Y} = \frac{-1.283}{(1.326)(1.223)} = -0.791.$$

where S_X and S_Y are the sample standard deviations of the X's and Y's. This is a fairly high negative correlation; and this result can be used as a so-called *antithetic* variate to reduce the variance of an estimator for $\mathsf{E}[X]$.

3. (10 points) We are interested in seeing if the daily order volume of a particular item at a warehouse is geometric distributed. Here are the statistics for a 100-day period. We'll assume that the numbers from day to day are i.i.d.

# of orders	# of days
1	52
2	22
3	13
4	7
5	6

Thus, for example, there were 22 days during when we had exactly 2 orders.

Perform a 90% test to see if the number of orders each day is geometric. I leave it to you to find the appropriate intervals, degrees of freedom, etc. You need to *neatly* show your work and *clearly* state your final answer in plain English.

Solution: We'll do a χ^2 g-o-f test for the Geom(p) distribution. The MLE of p is (by a previous problem on this test) $\hat{p} = 1/\bar{X} = 1/1.93 = 0.518$. Then by invariance, we can calculate

$$\widehat{\mathsf{P}}(X=i) \;=\; (1-\hat{p})^{i-1}\hat{p} \;=\; (0.482)^{i-1}(0.518), \quad \text{for } i = 1, 2, \ldots.$$

Then the expected number of observations for each value of i is

$$E_i \;=\; n\widehat{\mathsf{P}}(X=i) \;=\; 100(0.482)^{i-1}(0.518),$$

and we obtain the following table.

i	$\widehat{\mathsf{P}}(X=i)$	E_i	O_i
1	0.5180	51.80	52
2	0.2497	24.97	22
3	0.1203	12.03	13
4	0.0580	5.80	7
5	0.0280	2.80	6
≥ 6	0.0260	2.60	0
	1	100	100

where we have added the "≥ 6" row to make things add up properly.

In order to assure that all of the E_i's we use are at least 5, we'll need to combine a couple of those latter rows. Here's what we get.

i	$\widehat{\mathsf{P}}(X = i)$	E_i	O_i
1	0.5180	51.80	52
2	0.2497	24.97	22
3	0.1203	12.03	13
4	0.0580	5.80	7
≥ 5	0.0540	5.40	6
	1	100	100

We can now calculate the χ^2 goodness-of-fit statistic,

$$\chi_0^2 \;=\; \sum_i \frac{(O_i - E_i)^2}{E_i} \;=\; 0.747.$$

Meanwhile, let's conduct a formal test at level $\alpha = 0.1$, for which the appropriate quantile is $\chi^2_{\alpha,k-1-s} = \chi^2_{0.1,3} = 6.25$, where $k = 5$ is the number of intervals and $s = 1$ is the number of parameters that we had to estimate. Since $\chi_0^2 < \chi^2_{0.1,3}$, we fail to reject the hypothesis that the number of accidents is geometric. (Actually, since the χ_0^2 statistic was so small, I really didn't need to look up anything in a table.) So the bottom line is that we are willing to assume that the data are geometric. $\diamondsuit$