Exam III

STAT 430
FALL 2017

Instructions

- The exam is a take home exam. It is due December 15th on blackboard by 5:00 pm.
- If you have any questions about, or need clarification on the meaning of an item on this exam, please ask your instructor during my office hours or email.
- No other form of external help is permitted attempting to receive help or provide help to others will be considered cheating.
- **Do not cheat on this exam.** Academic integrity demands an honest and fair testing environment. Cheating will not be tolerated and will result in an immediate score of 0 on the exam and an incident report will be submitted to the dean's office.

Name: ______________________________

Student ID: ______________________________

1. (5 points) Suppose that X is a Poisson random variable with rate λ. Find $E(1/(X+1))$.
2. (5 points) Suppose that X is a uniform on the interval $[0,1]$. With $Y = \sqrt{X}$, find $E(Y)$ and $Var(Y)$.
3. Suppose that $Z_1, Z_2, \ldots$ are independent standard normal random variables.
   (a) (5 points) Find the moment generating function of $X_n = \sum_{i=1}^{n} \frac{1}{3^i} Z_i$.
   (b) (5 points) Using the moment generating function, find the mean and variance of X_n.
   (c) (5 points) As $n \to \infty$, what happens to the distribution of X_n?
4. (10 points) In R, use the Monte Carlo method of integration to estimate the value of $\int_0^1 sin(2\pi x)dx$ with $n = 100$, $n = 1000$ and $n = 10000$. Compare the estimated integral to the exact value.
5. (5 points) Let $\{X_i\}$ be a sequence of independent random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma_i^2$ (i.e., the variances of the random variables are different). Show that if $\sum_{i=1}^{n} {}_i^2/n^2 \to 0$ then $\bar{X}$ converges to μ in probability.
6. Consider an sample of n independent random variables with density function

$$f(y|\theta) = \frac{\theta}{2} e^{-|x|/\theta}, -\infty < x < \infty$$

   (a) (5 points) Find the MME estimator of θ.
   (b) (5 points) Find the MLE estimator of θ.
7. Suppose that ϵ_i are independent normal random variables with mean 0 and (unknown) variance σ^2 for $i = 1, 2, ..., n$. For known values of x_i, consider the statistical model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$$

   (a) (10 points) Find the maximum likelihood estimates $\hat{\beta}_0$ $\hat{\beta}_1$ $\hat{\beta}_2$ and $\hat{\sigma^2}$ in terms of the observable values of x_i and y_i.
   (b) (5 points) Suppose that an expiriment is performed and the following observations are collected: Using this data, provided the fitted version of the model from part a),

x	1	2	3	4	5	6	7	8	9	10	11	12
y	5.63	3.41	-0.92	-8.96	-20.75	-38.23	-60.54	-85.79	-118.26	-147.34	-182.94	-226.32
	5.88	3.03	-0.16	-8.98	-22.54	-41.10	-60.98	-85.47	-117.15	-148.61	-185.78	-225.89

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2$$

   (c) (5 points) In R, create a plot with the fitted values vs the observed values (i.e., the fitted values $\hat{y}_i$ on the x-axis). What does this plot indicate about the overall quality of the fitted model?
   (d) (5 points) Using the fitted model from part b), find the values of the residuals

$$e_i = y_i - \hat{y}_i$$

   and provide a histogram of these values. What does the shape of the residuals histogram indicate about the assumptions we make when fitting the model we used in part b)?
8. Suppose that ϵ_i are independent normal random variables with mean 0 and (unknown) variance σ^2 for $i = 1, 2, ..., n$. For known values of x_i, consider two statistical models:

Model A

$$y_i = \beta_0 e^{\beta_1 x_i} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2) \text{ (independent)}$$

Model B

$$log(y_i) = \alpha_0 + \alpha_1 x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2) \text{ (independent)}$$

(a) (10 points) Using maximum likelihood estimators, provided the estimated parameters of Model A in terms of x_i and y_i.

(b) (10 points) Using maximum likelihood estimators, provided the estimated parameters of Model B in terms of x_i and y_i.

(c) (5 points) Are these models equivalent? That is, can we model an exponential relationship between y_i and x_i by modeling a linear relationship between $log(y_i)$ and x_i? Explain.