STAT 740, Fall 2017: Homework 2

1. **Properties of estimators**. Redo my simulation example with median (i.e. least absolute deviation) regression, but for samples of size $n = 100$ and also compare to Huber regression using `rlm` in the `MASS` package (Huber is the default). For the `lm` use t-intervals; if `fit` is your fitted normal-errors model, `confint(fit)` gives the t-intervals. Prepare three tables, one for each method and discuss which of the methods is "best" for this type of data. Note: it would be better to use bootstrapped CIs for the median-regression and Huber fits; we will come back to this later.

2. Recalling that Gauss-Hermite quadrature is *exact* for polynomials of degree $2J - 1$, use Gauss-Hermite quadrature to estimate $\pi = 3.1415...$. Hint: what integral does the following give?

```
library(fastGHQuad)
rule=gaussHermiteData(1)
sum(rule$w)
```

3. You will argue that the following R code estimates π

```
n=100000; 4*sum(runif(n)^2+runif(n)^2<1)/n
```

   (a) Try running the code a few times in R. Does it at least provide a couple digits worth of accuracy?

   (b) Write this as an expectation w.r.t. $(u_1, u_2) \sim U\{[0, 1]^2\}$, i.e. (u_1, u_2) are uniform over the unit square. Specifically, find $g(u_1, u_2)$ s.t. $E\{g(u_1, u_2)\} = \pi$.

   (c) What is the distribution of the function $g(u_1, u_2)$? Show that its variance is $\pi(4 - \pi)$.

   (d) What n is needed so that the standard error of the estimator above is 0.0001? That is, so we can trust it to 3 decimal places? Try it in R. Is this a reasonable way to estimate π? Why or why not?

   Estimating π has a long and storied history; you may have brushed up against it in STAT 712–713 as "Buffon's needle problem". Read `https://www.exploratorium.edu/pi/history-of-pi`.

4. For the density $f(x) \propto 3\exp\{-0.5(x+2)^2\} + 7\exp\{-0.5(x-2)^2\}$ (problems 3 and 4, Homework 1),

   (a) Compute the exact normalizing constant, both in closed form using "π" (pencil and paper, one line), and also to 10 decimal places (by evaluating the exact version in R).

   (b) What effective range and J is required to approximate this integral with a Riemann sum to three decimal places?

   (c) Let $\mu_0 = -2$ and $\mu_1 = 2$. Show that marginal distribution of X where $X|Y \sim N(\mu_Y, 1)$ and $Y \sim \text{Bern}(0.7)$ has the density above (one line).

   (d) Let $X \sim f(\cdot)$ according to the above density. Use iterated expectation and part (c) to find $\mu = E(X)$ exactly. Use the law of total variance to find $\sigma^2 = var(X)$.

   (e) The skew of a random variable is defined to be $\gamma_1 = E\{(X-\mu)^3\}/\sigma^3$. Use Gauss-Hermite quadrature (regular, not adaptive) to compute this exactly via

$$E\{(X-\mu)^3\} = \int_{-\infty}^{\infty} (x-\mu)^3 f(x)dx = 0.3\int_{-\infty}^{\infty} (x-\mu)^3 \phi(x|-2,1)dx + 0.7\int_{-\infty}^{\infty} (x-\mu)^3 \phi(x|2,1)dx.$$

   (f) Use the method of composition and (c) to obtain a Monte Carlo sample of size $m = 10000$ $X_1, \ldots, X_{10000}$. Use this sample to compute (i) $E(X)$, (ii) $var(X)$, (iii) $90th$ percentile of X, and (iv) a highest-density 95% probability interval for X.

5. Let the Ache hunting data be $\{(m_i, a_i, t_i)\}_{i=1}^{47}$ where the ith hunter, aged a_i years, killed m_i monkeys out of t_i days hunting. Assume the following logistic regression model with random intercepts

$$m_i|\boldsymbol{\beta}, u_i \overset{ind.}{\sim} \text{bin}\left(t_i, \frac{\exp(\beta_0 + \beta_1 a_i + u_i)}{1+\exp(\beta_0 + \beta_1 a_i + u_i)}\right), \; u_1, \ldots, u_{47}|\sigma \overset{iid}{\sim} N(0, \sigma^2).$$

   Obtain the MLE of $\boldsymbol{\theta} = (\beta_0, \beta_1, \log\sigma)'$ and their standard errors by using adaptive Gauss-Hermite quadrature; you can use my Poisson regression code (which now works, look under "Now by hand...") as a template. Check your results using `glmer` in the `lme4` package. Interpret $\hat{\beta}_1$ in terms of odds ratios. Find a (large sample) 95% CI for β_1 and then exponentiate to get a CI for e^{β_1}. Is the probability of killing a monkey significantly associated with age?

   Note: You can fit this model in `glm` without random effects or `glmer` with random effects using something like

```
glm(cbind(m,t-m)~a,family="binomial")
glmer(cbind(m,t-m)~a+(1|id),family="binomial",nAGQ=100)
```