University of California, Santa Cruz
Department of Applied Mathematics and Statistics
Baskin School of Engineering
Classical and Bayesian Inference - AMS 132

Homework 3

Instructions: You have until Wednesday, February 21, to complete the assignment. It has to be returned during 10 first minutes of class (4:55 pm to 5:05 pm) or between 1:00 pm and 3:00 pm in office BE 357B.

1. Suppose that a random sample is to be taken from the Bernoulli distribution with unknown parameter p. Suppose also that it is believed that the value of p is in the neighborhood of 0.2. How large a random sample must be taken in order that $P(|\overline{X}_n - p| \leq 0.1) \geq 0.75$ when $p = 0.2$?
   To do this, first find the distribution of $n\overline{X}_n$. Later, try values of n between 8 and 11. For computing the probabilities you can use: functions `dbinom(x, size=n, prob=p)` or `pbinom(x, size=n, prob=p)` from `R` that computes $P(X = x)$, or $P(X \leq x)$, respectively, when $X \sim Binomial(n, p)$; a calculator; or the table in the back of the book, page 854. **[15pts]**

   Since $X_i \overset{i.i.d.}{\sim} Bernoulli(p)$, it follows that $Y = n\overline{X}_n = \sum_{i=1}^{n} X_i \sim Binomial(n, p)$, where $p = 0.2$. Now,

$$P(|\overline{X}_n - 0.2| \leq 0.1) = P(-0.1 \leq \overline{X}_n - 0.2 \leq 0.1) = P(-0.1 + 0.2 \leq \overline{X}_n \leq 0.1 + 0.2),$$

$$= P(0.1n \leq Y \leq 0.3n).$$

   For $n = 8$, $P(|\overline{X}_n - 0.2| \leq 0.1) = P(0.8 \leq Y \leq 2.4) = P(Y = 1) + P(Y = 2) = 0.6291$, where $Y \sim Binomial(n = 8, p = 0.2)$. For $n = 9$, $P(|\overline{X}_n - 0.2| \leq 0.1) = P(0.9 \leq Y \leq 2.7) = P(Y = 1) + P(Y = 2) = 0.6039$, where $Y \sim Binomial(n = 9, p = 0.2)$. For $n = 10$, $P(|\overline{X}_n - 0.2| \leq 0.1) = P(1 \leq Y \leq 3) = P(Y = 1) + P(Y = 2) + P(Y = 3) = 0.7717$, where $Y \sim Binomial(n = 10, p = 0.2)$. For $n = 11$, $P(|\overline{X}_n - 0.2| \leq 0.1) = P(1.1 \leq Y \leq 3.3) = P(Y = 2) + P(Y = 3) = 0.5167$, where $Y \sim Binomial(n = 11, p = 0.2)$. So, in order to have $P(|\overline{X}_n - p| \leq 0.1) \geq 0.75$ when $p = 0.2$, we need a sample of size $n = 10$. The `R` code for computing the above probabilities is

```
sum(dbinom(c(1,2), size=8, prob=0.2))
[1] 0.6291456
> sum(dbinom(c(1,2), size=9, prob=0.2))
[1] 0.6039798
```

```
> sum(dbinom(c(1,2,3), size=10, prob=0.2))
[1] 0.7717519
> sum(dbinom(c(2,3), size=11, prob=0.2))
[1] 0.5167383
```

2. Considering that $X_1, \ldots, X_n$ form a random sample from the Bernoulli distribution with parameter p, use the central limit theorem in Sec. 6.3 to find approximately the size of a random sample that must be taken in order that$P(\mid \overline{X}_n - p \mid \leq 0.1) \geq 0.95$ when $p = 0.2$.
   Recall that the central limit theorem states that if $X_1, \ldots, X_n$ are independent and identically distributed random variable with mean μ and variance σ^2, then $\overline{X}_n \sim N(\mu, \sigma^2/n)$, for large enough n.

   Since $X_i \overset{iid}{\sim} Bernoulli(p)$, $p = 0.2$, it follows that $\mu = E(X_i) = p = 0.2$ and $\sigma^2 = Var(X_i) = p(1-p) = 0.2 * 0.8 = 0.16$. By the central limit theorem, $\overline{X}_n \sim N(0.2, 0.16/n)$, therefore $Z = \frac{\overline{X}_n - 0.2}{\sqrt{0.16/n}} \sim N(0,1)$. Now, note that

$$P(\mid \overline{X}_n - p \mid \leq 0.1) = P(-0.1 \leq \overline{X}_n - 0.2 \leq 0.1) = P(-\frac{0.1}{\sqrt{0.16/n}} \leq Z \leq \frac{0.1}{\sqrt{0.16/n}})$$
$$= 2P(Z \leq \frac{0.1}{\sqrt{0.16/n}}) - 1.$$

   Notice that $P(\mid \overline{X}_n - p \mid \leq 0.1) \geq 0.95$ is equivalent to $2P(Z \leq \frac{0.1}{\sqrt{0.16/n}}) - 1 \geq 0.95$. so we need to find the sample size such that $P(Z \leq \frac{0.1}{\sqrt{0.16/n}}) \geq 1.95/2 = 0.975$. This is, $\frac{0.1}{\sqrt{0.16/n}} \geq 1.9599$. Therefore, $n \geq \left(\frac{1.9599\sqrt{0.16}}{0.1}\right)^2 = 61.49$, this is, the sample size has to be equal or larger than 62.

   , this is, we need the sample size such that the 0.975 quantile of a standard normal distributed random variable. The value of n that

3. Suppose that $X_1, \ldots, X_n$ form a random sample from the normal distribution with mean μ and variance σ^2. Assuming that the sample size n is 16, determine the following probability

$$P\left(\frac{1}{2}\sigma^2 \leq \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \leq 2\sigma^2\right).$$

   For computing the probabilities you can use: functions `dchisq(x, df=m)` or `pchisq(x, df=m)` from `R` that computes $f(x)$, or $P(X \leq x)$, respectively, when $X \sim \chi^2_{(m)}$; or the table in the back of the book, page 858. **[15pts]**

   Since $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, it follows that $\left(\frac{X_i - \mu}{\sigma}\right) \sim N(0,1)$, then $\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_{(1)}$. Now, since the random variables $\frac{X_i - \mu}{\sigma}$ are independent, $Y = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_{(n)}$. So, for $n = 16$,

$$P\left(\frac{1}{2}\sigma^2 \leq \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \leq 2\sigma^2\right) = P(\frac{n}{2} \leq Y \leq 2n)$$

$$= P(8 \leq Y \leq 32) = P(Y \leq 32) - P(Y \leq 8) = 0.9388,$$

where $Y \sim \chi^2_{(16)}$. The `R` code for computing the above probability is

```
pchisq(32, df=16)-pchisq(8, df=16)
[1] 0.9388666
```

4. Suppose that $X_1, \ldots, X_n$ form a random sample from the normal distribution with mean μ and variance σ^2, and let $\widehat{\sigma_0}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$. Determine the smallest values of n for which the following relations are satisfied:

   (a) $P\left(\frac{\widehat{\sigma_0}^2}{\sigma^2} \leq 1.9\right) \geq 0.95$.

   Since $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, it follows that $Y = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_{(n)}$. Now

   $$P\left(\frac{\widehat{\sigma_0}^2}{\sigma^2} \leq 1.9\right) = P(\frac{Y}{n} \leq 1.9) = P(Y \leq 1.9n).$$

   In order to $P(Y \leq 1.9n) \geq 0.95$, wee neet to find n such that the 0.95 quantile of a $\chi^2_{(n)}$ is $1.9n$. After trying different values of n, it follows that $n = 9$ is such that $P(Y \leq 1.9 * 9) = 0.9528 \geq 0.95$, where $Y \sim \chi^2_{(9)}$.

   (b) $P\left(\mid \widehat{\sigma_0}^2 - \sigma^2 \mid \leq \frac{1}{2}\sigma^2\right) \geq 0.8$.

   Since $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, it follows that $Y = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_{(n)}$. Now

   $$P\left(\mid \widehat{\sigma_0}^2 - \sigma^2 \mid \leq \frac{1}{2}\sigma^2\right) = P(-1/2 \leq \sigma_0^2/\sigma^2 - 1 \leq 1/2) = P(1/2 \leq \sigma_0^2/\sigma^2 \leq 3/2)$$
   $$= P(n/2 \leq Y \leq 3n/2),$$

   where $Y \sim \chi^2_{(n)}$. After trying different values for n, it folows that $n = 12$ is such that $P(n/2 \leq Y \leq 3n/2) = 0.8003 \geq 0.8$, where $Y \sim \chi^2_{(12)}$.

5. Suppose that the random variables X_1, X_2, and X_3 are i.i.d., and that each has the standard normal distribution. Also, suppose that

   $$Y_1 = 0.8X_1 + 0.6X_2,$$
   $$Y_2 = \sqrt{2}(0.3X_1 - 0.4X_2 - 0.5X_3),$$
   $$Y_3 = \sqrt{2}(0.3X_1 - 0.4X_2 + 0.5X_3).$$

   Find the distribution of Y_1, Y_2, and Y_3.

   Since Y_1, Y_2, and Y_3 are linear combinations of standard normally and independently distributed random variables, it follows that all have a normal distribution. Now

   $$E(Y_1) = 0.8E(X_1) + 0.6E(X_2) = 0,$$

$$Var(Y_1) = 0.8^2 Var(X_1) + 0.6^2 Var(X_2) = 0.8^2 + 0.6^2 = 1,$$

$$E(Y_2) = \sqrt{2}(0.3E(X_1) - 0.4E(X_2) - 0.5E(X_3)) = 0,$$
$$Var(Y_2) = 2(0.3^2 Var(X_1) + 0.4^2 Var(X_2) + 0.5^2 Var(X_3)) = 2(0.3^2 + 0.4^2 + 0.5^2) = 1,$$

$$E(Y_3) = \sqrt{2}(0.3E(X_1) - 0.4E(X_2) + 0.5E(X_3)) = 0,$$
$$Var(Y_3) = 2(0.3^2 Var(X_1) + 0.4^2 Var(X_2) + 0.5^2 Var(X_3)) = 2(0.3^2 + 0.4^2 + 0.5^2) = 1,$$

Therefore, Y_1, Y_2, and Y_3 have standard normal distributions.

6. Suppose that a point (X, Y) is to be chosen at random in the xy-plane, where X and Y are independent random variables and each has the standard normal distribution. If a circle is drawn in the xy-plane with its center at the origin, what is the radius of the smallest circle that can be chosen in order for there to be probability 0.99 that the point (X, Y) will lie inside the circle?

   We need to find the smallest radius r, such that $P(X^2 + Y^2 \leq r^2) = 0.99$. Since X and Y are independent random variables with standard normal distribution, it follows that $W = X^2 + Y^2$ has a χ^2 distribution with 2 degrees of freedom. So, r^2 is the 0.99 quantile of W. Therefore, $r^2 = 9.2103$, which implies that the smallest radius is 3.0348.

7. Suppose that six random variables $X_1, \ldots, X_6$ form a random sample from the standard normal distribution, and let
$$Y = (X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2.$$
   Determine a value of c such that the random variable cY will have a χ^2 distribution. **[15pts]**

   Since $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ it follows that $X_1 + X_2 + X_3 \sim N(0, 3)$ and $X_4 + X_5 + X_6 \sim N(0, 3)$. Then $\frac{X_1+X_2+X_3}{\sqrt{3}} \sim N(0, 1)$ and $\frac{X_4+X_5+X_6}{\sqrt{3}} \sim N(0, 1)$. This implies that, $\frac{(X_1+X_2+X_3)^2}{3} \sim \chi^2_{(1)}$ and $\frac{(X_4+X_5+X_6)^2}{3} \sim \chi^2_{(1)}$. Since these are independent random variables, it follows that $\frac{(X_1+X_2+X_3)^2}{3} + \frac{(X_4+X_5+X_6)^2}{3} \sim \chi^2_{(2)}$. This is, $\frac{1}{3}Y \sim \chi^2_{(2)}$. Therefore, if $c = 1/3$, the random variable cY will have a χ^2 distribution with 2 degrees of freedom.