STAT 512 hw 3

1. Let $X_1, \dots, X_{25} \overset{\text{ind}}{\sim} \text{Normal}(\mu = 3, \sigma^2 = 4)$.

   (a) Give the mgf of X_1.

   $M_{X_1}(t) = e^{3t + 4t^2/2}$.

   (b) Give the mgf of $\bar{X}_{25} = (1/25)(X_1 + \cdots + X_{25})$.

   $M_{X_1}(t) = e^{3t + (4/25)t^2/2}$.

   (c) Give $P(X_1 < 2)$.

   We have

   $$P(X_1 < 2) = P((X_1 - 3)/2 < (2-3)/2) = P(Z < -1/2), \quad Z \sim \text{Normal}(0,1),$$

   and $P(Z < -1/2) = \texttt{pnorm(-1/2)} = 0.3085375$.

   (d) Give $P(\bar{X}_{25} < 2)$.

   We have

   $$P(\bar{X}_{25} < 2) = P(5(\bar{X}_{25} - 3)/2 < 5(2-3)/2) = P(Z < -5/2), \quad Z \sim \text{Normal}(0,1),$$

   and $P(Z < -5/2) = \texttt{pnorm(-2.5)} = 0.006209665$.

   (e) Give $P(|X_1 - 3| > 1)$.

   We have

   $$\begin{aligned} P(|X_1 - 3| > 1) &= 1 - P(-1 < X_1 - 3 < 1) \\ &= 1 - P(-1/2 < (X_1 - 3)/2 < 1/2) \\ &= 1 - P(-1/2 < Z < 1/2), \quad Z \sim \text{Normal}(0,1), \\ &= 2(1 - P(Z < 1/2)) \\ &= \texttt{2*(1 - pnorm(1/2))} \\ &= 0.6170751. \end{aligned}$$

   (f) Give $P(|\bar{X}_{25} - 3| > 1)$.

We have

$$\begin{aligned}
P(|\bar{X}_{25} - 3| > 1) &= 1 - P(-1 < \bar{X}_{25} - 3 < 1) \\
&= 1 - P(-5/2 < 5(X_1 - 3)/2 < 5/2) \\
&= 1 - P(-5/2 < Z < 5/2), \quad Z \sim \text{Normal}(0,1), \\
&= 2(1 - P(Z < 5/2)) \\
&= \texttt{2*(1 - pnorm(5/2))} \\
&= 0.01241933.
\end{aligned}$$

(g) Identify the distribution of $5(\bar{X}_{25} - 3)/2$.

This has the $\text{Normal}(0,1)$ distribution.

(h) Give $P([5(\bar{X}_{25} - 3)/2]^2 > 3.841459)$. *Hint: Use the result from Question 2.*

Since $Z^2 \sim \chi_1^2$ if $Z \sim \text{Normal}(0,1)$, the answer is

$$P([5(\bar{X}_{25} - 3)/2]^2 > 3.841459) = \texttt{1 - pchisq(3.841459,1)} = 0.05.$$

(i) Now let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Normal}(\mu = 3, \sigma^2 = 4)$ for $n \geq 1$ and set $\bar{X}_n = (1/n)(X_1 + \cdots + X_n)$.

i. Consider the probability $P(|\bar{X}_n - 3| > \varepsilon)$, for some small $\varepsilon > 0$. Is this an increasing or a decreasing function of n?

We have

$$P(|\bar{X}_n - 3| > \varepsilon) = P(\sqrt{n}|\bar{X}_n - 3|/2 > \sqrt{n}\varepsilon/2) = P(|Z| > \sqrt{n}\varepsilon/2), \quad Z \sim \text{Normal}(0,1)$$

which is a decreasing function of n.

ii. What does your answer say about the quality of $\bar{X}_n$ as an estimator of the mean?

As n increases, the probability that $\bar{X}_n$ is within any fixed distance of the mean increases. This means that $\bar{X}_n$ can be made very close to the mean, with high probability, if n (which we may call the sample size) is made large enough. This seems like a good quality for an estimator to have! Maybe this property should have a name…

2. Let $Z \sim \text{Normal}(0,1)$. Show that Z^2 has the χ_1^2 distribution (chi-squared with degrees of freedom 1), i.e. show that the pdf of $Y = Z^2$ is

$$f_Y(y) = \frac{1}{\Gamma(1/2)2^{1/2}} y^{\frac{1}{2}-1} e^{-y/2} \mathbf{1}(y > 0).$$

Hint: $\Gamma(1/2) = \sqrt{\pi}$.

We have $Z \sim \phi(z) = (2\pi)^{-1/2}e^{-z^2/2}$ for $z \in \mathbb{R}$. Let $\Phi(z) = \int_{-\infty}^{z} \phi(t)dt$ denote the cdf of Z. The random variable $Y = Z^2$ has support $(0, \infty)$ and cdf given by

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \end{aligned}$$

for all $y \in \mathbb{R}$. For $y \in (0, \infty)$, the pdf of Y is given by

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) \\ &= \frac{d}{dy}[\Phi(\sqrt{y}) - \Phi(-\sqrt{y})] \\ &= \phi(\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right) - \phi(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right) \\ &= \left(\frac{1}{2\sqrt{y}}\right)[\phi(\sqrt{y}) + \phi(-\sqrt{y})] \\ &= \frac{1}{2\sqrt{y}}\phi(\sqrt{y}) \qquad \{\text{since } \phi(\sqrt{y}) = \phi(-\sqrt{y})\} \\ &= \frac{1}{\Gamma(1/2)2^{1/2}}y^{1/2-1}e^{y}/2. \end{aligned}$$

3. Let $W \sim \text{Chi-squared}(\nu)$, so that the pdf of W is given by

$$f_W(w) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}}w^{\nu/2-1}e^{-w/2}\mathbf{1}(w > 0),$$

where $\nu > 0$ is the degrees of freedom.

(a) Give α and β such the $\text{Gamma}(\alpha, \beta)$ and $\text{Chi-squared}(\nu)$ distributions are the same.

The pdf of the $\Gamma(\nu/2, 2)$ is that of the $\text{Chi-squared}(\nu)$ distribution.

(b) Give $\mathbb{E}W$ in terms of the degrees of freedom parameter ν.

The expected value of a $\text{Gamma}(\alpha, \beta)$ random variable is $\alpha\beta$, so $\mathbb{E}W = (\nu/2)2 = \nu$.

(c) Give $\text{Var}\, W$ in terms of the degrees of freedom parameter ν.

The variance of a $\text{Gamma}(\alpha, \beta)$ random variable is $\alpha\beta^2$, so $\text{Var}\, W = (\nu/2)2^2 = 2\nu$.

4. Let $Y_1, \ldots, Y_n \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$ distribution, $p \in (0, 1)$. Let $\hat{p}_n = \bar{Y}_n = n^{-1}(Y_1 + \cdots + Y_n)$.

   (a) Show that

$$\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2 = \frac{n}{n-1}\hat{p}_n(1 - \hat{p}_n).$$

Write

$$\begin{aligned}\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2 &= \frac{1}{n-1}\left[\sum_{i=1}^{n} Y_i^2 - n\bar{Y}_n^2\right] \\ &= \frac{1}{n-1}\left[n\hat{p}_n - n\hat{p}_n^2\right] \\ &= \frac{n}{n-1}\hat{p}_n(1 - \hat{p}_n).\end{aligned}$$

The trick is to note that $Y_i^2 = Y_i$ since $Y_i \in \{0, 1\}$ for all $i = 1, \ldots, n$.

   (b) Find $\mathbb{E}\left[n(n-1)^{-1}\hat{p}_n(1 - \hat{p}_n)\right]$.

Since this is the same as $\mathbb{E}S_n^2$, where $S_n^2 = (n-1)^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2$, the expectation is $p(1-p)$, since this is the variance of the $\text{Bernoulli}(p)$ distribution.

5. Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Poisson}(\lambda)$ distribution. Find the expected value of $\hat{\lambda} = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. *Hint:* $\mathbb{E}\hat{\lambda} \neq \lambda$.

Since variance of the $\text{Poisson}(\lambda)$ distribution is λ, $\mathbb{E}S_n^2 = \lambda$, where $S_n^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. We can write $\hat{\lambda}$ as $\hat{\lambda} = (n-1)n^{-1}S_n^2$. Therefore

$$\mathbb{E}\hat{\lambda} = (n-1)n^{-1}\lambda.$$

6. Let $X_1, \ldots, X_n$ be a random sample from the $\text{Weibull}(k, \lambda)$ distribution, which has pdf

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, & x \geq 0 \\ 0, & x < 0, \end{cases}$$

for $k, \lambda > 0$.

(a) Find the cdf F_X of the Weibull(k, λ) distribution. *Hint: Set up the integral* $F_X(x) = \int_0^x f_X(t)dt$ *and do the change of variable* $u = (t/\lambda)^k$.

For $x > 0$ we have

$$\begin{aligned}
F_X(x) &= \int_0^x \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} e^{-(t/\lambda)^k} dt \\
&= \int_0^{(x/\lambda)^k} e^{-u} dt \qquad \text{by } u = (x/\lambda)^k \\
&= 1 - e^{-(x/\lambda)^k}.
\end{aligned}$$

(b) Find the pdf of $X_{(1)}$.

According to the formula in the lecture notes

$$\begin{aligned}
f_{X_{(1)}}(x) &= n[1 - F_X(x)]^{n-1} f_X(x) \\
&= ne^{-(n-1)(x/\lambda)^k} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \\
&= \frac{k}{\lambda n^{-1/k}}\left(\frac{x}{\lambda n^{-1/k}}\right)^{k-1} \exp\left[-\left(\frac{x}{\lambda n^{-1/k}}\right)^k\right]
\end{aligned}$$

for $x > 0$.

(c) Show that $X_{(1)}$ has the Weibull$(k, \lambda n^{-1/k})$ distribution.

We note that the pdf of $X_{(1)}$ can be written such that it is recognized as the pdf of the Weibull$(k, \lambda n^{-1/k})$ distribution.

7. Let $U_1, \ldots, U_n$ be a random sample from the Uniform$(0, \theta)$ distribution.

(a) Find the joint density of order statistics $U_{(1)}$ and $U_{(n)}$.

The pdf of the Uniform$(0, \theta)$ distribution is $f_U(u) = \theta^{-1}\mathbf{1}(0 < u < \theta)$ and the cdf is

$$F_U(u) = \begin{cases} 0, & u < 0 \\ u/\theta, & 0 \le u \le \theta \\ 1, & u > \theta \end{cases}$$

From the lecture notes, we have

$$\begin{aligned}
f_{U_{(1)},U_{(n)}}(u_1, u_n) &= \frac{n(n-1)}{\theta^2}\left[\frac{1}{\theta}u_n - \frac{1}{\theta}u_1\right]^{n-2} \mathbf{1}(0 < u_1 < u_n < \theta) \\
&= \frac{n(n-1)}{\theta^n}(u_n - u_1)^{n-2}\mathbf{1}(0 < u_1 < u_n < \theta).
\end{aligned}$$

(b) Find the joint density of the random variables $R = U_{(1)}/U_{(n)}$ and $M = U_{(n)}$.

We have

$$\begin{aligned} r &= u_1/u_n =: g_1(u_1, u_n) \\ m &= u_n =: g_2(u_1, u_n) \end{aligned} \iff \begin{aligned} u_1 &= rm =: g_1^{-1}(r, m) \\ u_n &= m =: g_2^{-1}(r, m) \end{aligned}$$

with Jacobian

$$J(r, m) = \begin{vmatrix} \frac{d}{dr} rm & \frac{d}{dm} rm \\ \frac{d}{dr} m & \frac{d}{dm} m \end{vmatrix} = \begin{vmatrix} m & r \\ 0 & 1 \end{vmatrix} = m.$$

Note that the support of (R, M) is the set of values $\{r, m : 0 < r < 1, 0 < m < \theta\}$. So the joint pdf of R and M is given by

$$\begin{aligned} f_{R,M}(r, m) &= \frac{n(n-1)}{\theta^n}(m - rm)^{n-2}|m|\mathbf{1}(0 < r < 1, 0 < m < \theta) \\ &= \frac{n(n-1)}{\theta} m^{n-1}(1 - r)^{n-2}\mathbf{1}(0 < r < 1, 0 < m < \theta). \end{aligned}$$

(c) State whether R and M are independent.

We can write down the joint pdf of R and M as

$$f_{R,M}(r, m) = \frac{n(n-1)}{\theta^n} m^{n-1}\mathbf{1}(0 < m < 1) \cdot (1 - r)^{n-2}\mathbf{1}(0 < r < 1),$$

which is the product of a function of just r and a function of just m, so R and M are independent.

(d) Give the marginal pdf of R and identify the distribution.

For $r \in (0, 1)$, integrating the joint density of R and M over $m \in (0, \theta)$ gives

$$f_R(r) = \int_0^\theta \frac{n(n-1)}{\theta^n} m^{n-1}(1 - r)^{n-2} dm = (n-1)(1 - r)^{n-2} \frac{m^n}{\theta^n}\Big|_0^\theta = (n-1)(1 - r)^{n-2}.$$

We can write this as

$$f_R(r) = \frac{\Gamma(n - 1 + 1)}{\Gamma(n-1)\Gamma(1)} r^{1-1}(1 - r)^{(n-1)-1}\mathbf{1}(0 < r < 1),$$

which we recognize as the pdf of the $\text{Beta}(1, n-1)$ distribution.

8. Let $X_1, \dots, X_n$ be a random sample from the $\text{Uniform}(0, 1)$ distribution, where n is an odd number. Show that

(a) The expected value of the median is $1/2$.

From the lecture, $X_{(k)} \sim \text{Beta}(k, n-k+1)$ so that

$$X_{((n+1)/2)} \sim \text{Beta}((n+1)/2, n-(n+1)/2+1),$$

giving

$$\mathbb{E}X_{((n+1)/2)} = \frac{(n+1)/2}{(n+1)/2+n-(n+1)/2+1} = 1/2.$$

(b) The variance of the median is $\frac{1}{4}\frac{1}{n+2}$.

$$\begin{aligned}\text{Var}\, X_{((n+1)/2)} &= \frac{(n+1)/2 \times (n-(n+1)/2+1)}{((n+1)/2+n-(n+1)/2+1)^2((n+1)/2+n-(n+1)/2+1+1)}\\ &= \frac{1}{4}\frac{1}{n+2}\end{aligned}$$

9. Use R to run the following simulation. Choose a sample size $n \leq 20$ and draw $1{,}000$ samples of size n from the $\text{Uniform}(0,1)$ distribution. In so doing:

(a) Choose a value of k, $1 < k < n$, and from each of the $1{,}000$ samples, save the kth order statistic. Make a histogram of the $1{,}000$ values of the kth order statistic and overlay the pdf of the sampling distribution of the kth order statistic (you must figure out and input the shape parameters of the beta distribution). Use the following code to get started:

```
S <- 1000 # number of random samples to generate
Yk <- numeric(S) # create empty vector in which to store values

for(s in 1:S) # run a loop of length S
{
    Y <- runif(n)
    Yk[s] <- sort(Y)[k] # get kth order statistic
}

hist(Yk,freq=FALSE,xlim=c(0,1))
y.seq <- seq(0,1,length=100)
lines(dbeta(y.seq,shape1=???,shape2=???)~y.seq,col="blue",lwd=2)
```

(b) Do the same thing for the nth order statistic and the 1st order statistic.

Turn in all three histograms with densities overlaid and all your code.

```
n <- 10
S <- 1000
k <- 3
Ymin <- Ymax <- Yk <- numeric(S)

for(s in 1:S)
{

        Y <- runif(n)
        Ymin[s] <- min(Y)
        Ymax[s] <- max(Y)
        Yk[s] <- sort(Y)[k]

}

par(mfrow=c(1,3))

hist(Yk,freq=FALSE,xlim=c(0,1))
y.seq <- seq(0,1,length=100)
lines(dbeta(y.seq,shape1=k,shape2=n-k+1)~y.seq,col="blue",lwd=2)

hist(Ymin,freq=FALSE,xlim=c(0,1))
y.seq <- seq(0,1,length=100)
lines(dbeta(y.seq,shape1=1,shape2=n)~y.seq,col="blue",lwd=2)

hist(Ymax,freq=FALSE,xlim=c(0,1))
y.seq <- seq(0,1,length=100)
lines(dbeta(y.seq,shape1=n,shape2=1)~y.seq,col="blue",lwd=2)
```