Solutions: Assignment #1 STA355H1S

1. (a) $Z \sim \mathcal{N}(0, \sigma^2)$. Then

$$\begin{aligned} P(|Z| \le x) &= P(-x \le Z \le x) \\ &= \Phi(x/\sigma) - \Phi(-x/\sigma) \\ &= \Phi(x/\sigma) - [1 - \Phi(x/\sigma)] \ \ \text{(symmetry of the } \mathcal{N}(0,\sigma^2) \text{ distribution)} \\ &= 2\Phi(x/\sigma) - 1 \end{aligned}$$

Likewise, $G^{-1}(\tau)$ satisfies $2\Phi(G^{-1}(\tau)/\sigma) - 1 = \tau$; thus

$$\begin{aligned} \Phi(G^{-1}(\tau)/\sigma) &= \frac{\tau+1}{2} \\ \text{and so } \ G^{-1}(\tau) &= \sigma\Phi^{-1}\left(\frac{\tau+1}{2}\right). \end{aligned}$$

(b) From part (a), it follows that the density of $|Z|$ is

$$g(x) = G'(x) = \frac{2}{\sigma}\phi(x/\sigma) \ \ \text{for } x \ge 0$$

where ϕ is the $\mathcal{N}(0,1)$ density. Now using the results from lecture, it follows that if $k/n \approx \tau$ then

$$\sqrt{n}\left[W_{(k)} - \sigma\Phi^{-1}\left(\frac{\tau+1}{2}\right)\right] \xrightarrow{d} \mathcal{N}\left(0, \frac{\tau(1-\tau)}{g^2(G^{-1}(\tau))}\right)$$

and

$$g^2(G^{-1}(\tau)) = \frac{4}{\sigma^2}\phi^2\left(\Phi^{-1}\left(\frac{\tau+1}{2}\right)\right).$$

Thus

$$\begin{aligned} \sqrt{n}(\widehat{\sigma}_k - \sigma) &= \left[\Phi^{-1}\left(\frac{\tau+1}{2}\right)\right]^{-1}\sqrt{n}\left[W_{(k)} - \sigma\Phi^{-1}\left(\frac{\tau+1}{2}\right)\right] \\ &\xrightarrow{d} \mathcal{N}(0, \gamma^2(\tau)) \end{aligned}$$

where

$$\gamma^2(\tau) = \frac{\sigma^2}{4}\left\{\tau(1-\tau)\left[\Phi^{-1}\left(\frac{\tau+1}{2}\right)\phi\left(\Phi^{-1}\left(\frac{\tau+1}{2}\right)\right)\right]^{-2}\right\}.$$

We can now use the following R code to find the value of τ minimizing the asymptotic variance above; note that we can ignore the $\sigma^2/4$ term.

```
> tau <- c(1:999)/1000
> quants <- qnorm((tau+1)/2)
> avar <- tau*(1-tau)/(quants*dnorm(quants))^2
> plot(tau,log(avar),type="l",lwd=4)
> tau[avar==min(avar)]
[1] 0.862
```

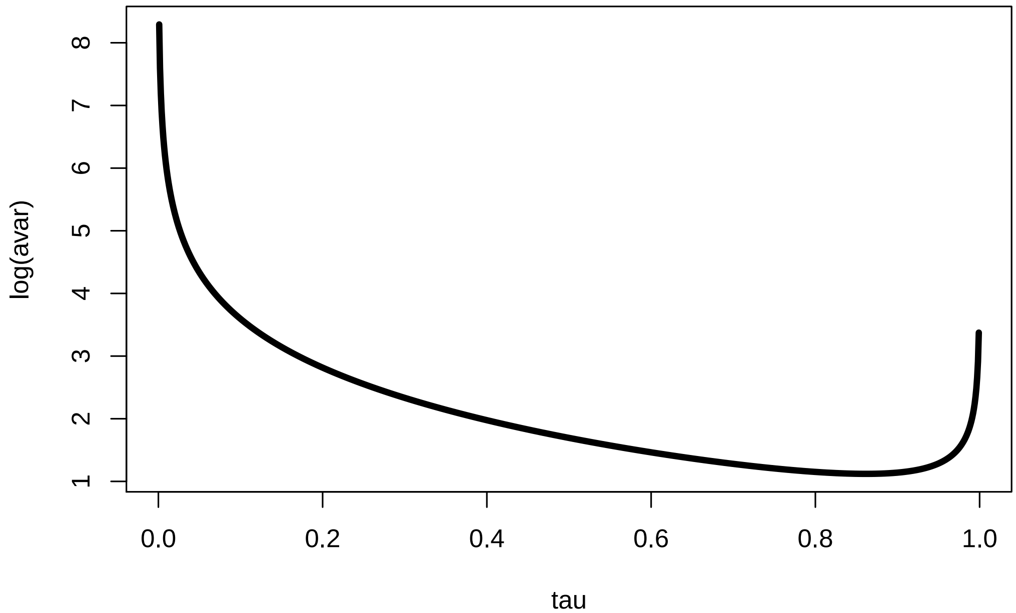

Figure 1: Graph of $\ln(4\gamma^2(\tau)/\sigma^2)$ versus τ.

The graph of $\ln(4\gamma^2(\tau)/\sigma^2)$ versus τ is shown in Figure 1.

(c) For simplicity, we will set $\sigma^2 = 1$; the general proof requires only a minor modification. First of all, suppose that $U_1 \sim \mathcal{N}(\mu, 1)$ and $U_2 \sim \mathcal{N}(-\mu, 1)$. Then

$$
\begin{aligned}
P(|U_1| \le x) &= P(-x \le U_1 \le x) \\
&= P(-x - \mu \le U_1 - \mu \le x - \mu) \\
&= \Phi(x - \mu) - \Phi(-x - \mu) \\
&= \Phi(x - \mu) + \Phi(x + \mu) - 1 \\
&= \Phi(x - |\mu|) + \Phi(x + |\mu|) - 1 \\
&= P(|U_2| \le x)
\end{aligned}
$$

Thus if $U \sim \mathcal{N}(\mu, 1)$ then the distribution of $|U|$ depends on $|\mu|$.

Now we can take $\mu > 0$ and show that if $U \sim \mathcal{N}(\mu, 1)$ then $P(|U| \le x)$ decreases as μ increases for each $x > 0$. To do this, it suffices to show that the (partial) derivative of $P(|U| \le x)$ with respect to μ is negative:

$$
\begin{aligned}
\frac{d}{d\mu} P(|U| \le x) &= \phi(x + \mu) - \phi(x - \mu) \\
&= \frac{1}{\sqrt{2\pi}} \left[\exp\left(-\frac{1}{2}(x + \mu)^2 \right) - \exp\left(-\frac{1}{2}(x - \mu)^2 \right) \right] \\
&< 0
\end{aligned}
$$

since

$$(x+\mu)^2 - (x-\mu)^2 = 4\mu x > 0$$

for $\mu, x > 0$.

(d) The assumption here is that all but a small subset of the data come from a $\mathcal{N}(0, \sigma^2)$ distribution. There are several approaches to identifying the observations coming from $\mathcal{N}(\mu \neq 0, \sigma^2)$ distributions, the main idea (which follows from part (c)) being that these observations are likely to be the largest in absolute value. Defining $w_i = |x_i|$ and ordering $\{w_i\}$ so that $w_{(1)} \leq w_{(2)} \leq \cdots \leq w_{(n)}$, we can successively remove the largest values until the half-normal plot of the remaining data looks like a half-normal plot of independent $\mathcal{N}(0, \sigma^2)$ observations.

The R code below illustrates this approach, sequentially deleting the largest values of w_i until the half-normal plot "looks" OK:

```
> x <- scan("data.txt")
> ord <- order(abs(x))
> for (k in 0:999) {
+     devAskNewPage(ask = T)
+     w <- ord[1:(1000-k)]
+     halfnormal(x[w],tau=0.862)
+     title(main=paste(k, ifelse(k==1,"observation removed",
+               "observations removed")))
+     }
```

Figure 2 shows half-normal plots for the original data as well as the plots when 7, 12, and 15 observations are removed. The half-normal plot with 7 observations removed looks quite good while the plots for 12 and 15 observations suggest that perhaps too many observations have been removed. (In these plots, I used the optimal $\tau = 0.862$ to estimate σ but you can use other values of τ.) The true number of observations with non-zero means in these data is 10.

2. (a) This is easier than it looks! Note that

$$\frac{1}{h(F^{-1}(\tau))} = \frac{1 - F(F^{-1}(\tau))}{f(F^{-1}(\tau))} = \frac{1-\tau}{f(F^{-1}(\tau))}.$$

Thus

$$\int_0^1 \frac{1}{h(F^{-1}(\tau))}\, d\tau = \int_0^1 \frac{1-\tau}{f(F^{-1}(\tau))}\, d\tau$$

and making the change of variables $x = F^{-1}(\tau)$ so that $\tau = F(x)$ and $d\tau = f(x)\, dx$, we get

$$\int_0^1 \frac{1-\tau}{f(F^{-1}(\tau))}\, d\tau = \int_0^\infty (1 - F(x))\, dx = E(X).$$

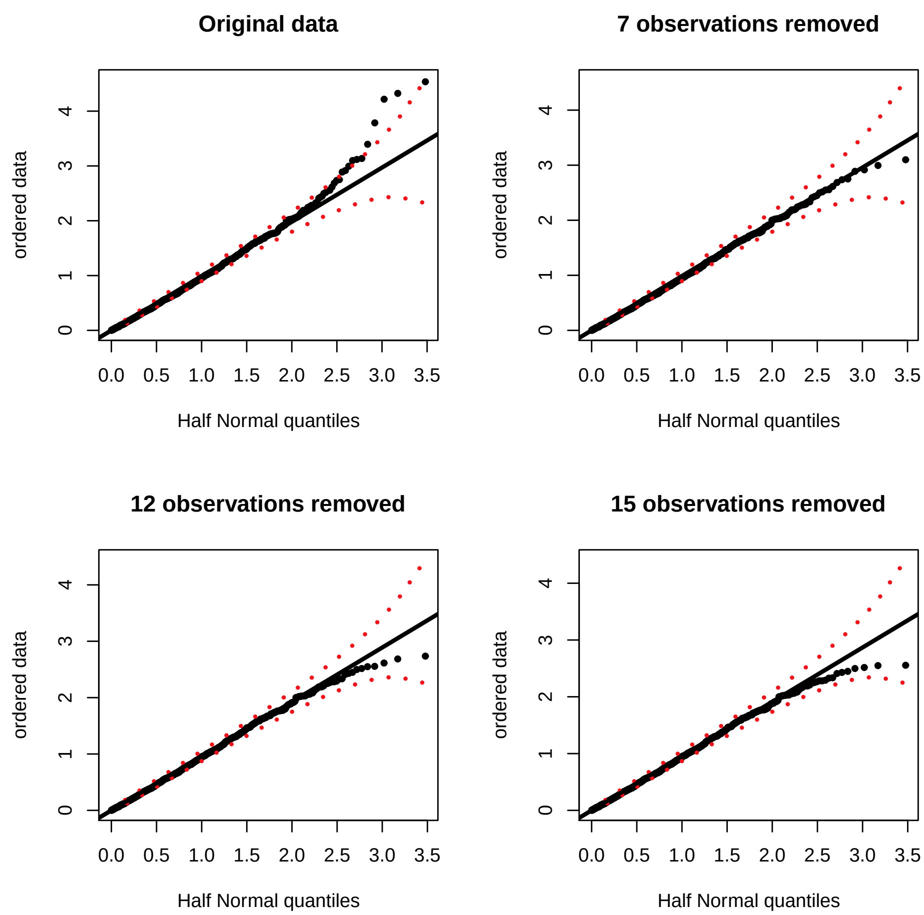

Figure 2: Half-normal plots for original data and for data with 7, 12, and 15 observations removed

(b) Since $k/n \approx (k-1)/n \approx \tau$, we have

$$(n-k+1)D_k = n\left(1 - \frac{k-1}{n}\right) D_k \approx (1-\tau)nD_k.$$

Thus

$$(n-k+1)D_k \xrightarrow{d} (1-\tau)Z$$

where Z has an Exponential distribution with mean $1/f(F^{-1}(\tau))$ and so $(1-\tau)Z$ has an Exponential distribution with mean $(1-\tau)/f(F^{-1}(\tau)) = 1/h(F^{-1}(\tau))$ from part (a).

(c) We use the following R code to produce the TTT plot (shown below):

```
> kevlar <- scan("kevlar.txt")
Read 76 items
> kevlar <- sort(kevlar)
> n <- length(kevlar)
> d <- c(n:1)*c(kevlar[1],diff(kevlar))
```

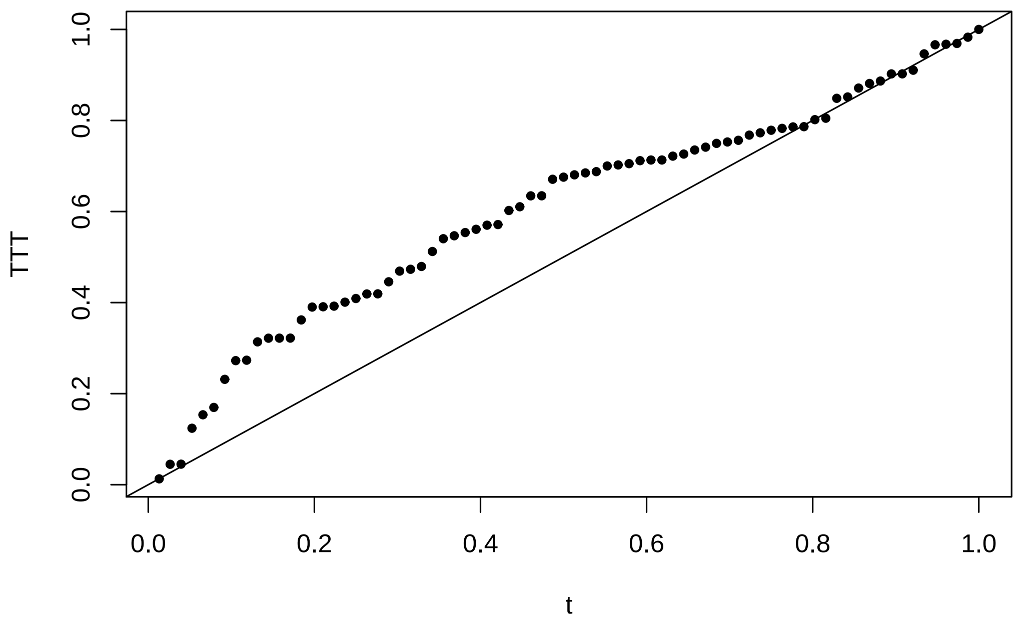

Figure 3: TTT plot for the kevlar data

```
> plot(c(1:n)/n, cumsum(d)/sum(kevlar),xlab="t", ylab="TTT",pch=20)
> abline(0,1) # add 45 degree line to plot
```

The TTT plot is shown in Figure 3; since the points lie above the 45^o line, this suggests that the hazard function is increasing.

Supplemental problems:

3. (a) First of all, note that λX has a Gamma distribution with shape parameter α and scale parameter 1; since skewness and kurtosis are scale-invariant, it follows that $\text{skew}(\lambda X) = \text{skew}(X)$ and $\text{kurt}(\lambda X) = \text{kurt}(X)$. Thus it suffices to compute the skewness and kurtosis for X with a Gamma distribution with shape parameter α and scale parameter 1.

To compute the skewness and kurtosis, we need to evaluate $E(X^k)$ for $k = 1, 2, 3, 4$:

$$E(X^k) = \frac{1}{\Gamma(\alpha)} \int_0^\infty x^{k+\alpha-1} \exp(-x)\, dx = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)}.$$

Using the identity, $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, we get $E(X) = \alpha$, $E(X^2) = \alpha(\alpha + 1)$, $E(X^3) = \alpha(\alpha + 1)(\alpha + 2)$, $E(X^4) = \alpha(\alpha + 1)(\alpha + 2)(\alpha + 3)$. Thus $\text{Var}(X) = \alpha$ and so

$$\text{skew}(X) = \frac{1}{\alpha^{3/2}} E[(X - \alpha)^3]$$

$$
\begin{aligned}
&= \frac{1}{\alpha^{3/2}}\left[E(X^3) - 3\alpha E(X^2) + 3\alpha^2 E(X) - \alpha^3\right] \\
&= \frac{2\alpha}{\alpha^{3/2}} \\
&= \frac{2}{\alpha^{1/2}}
\end{aligned}
$$

$$
\begin{aligned}
\text{kurt}(X) &= \frac{1}{\alpha^2} E[(X-\alpha)^4] \\
&= \frac{1}{\alpha^2}\left[E(X^4) - 4\alpha E(X^3) + 6\alpha^2 E(X^2) - 4\alpha^3 E(X) + \alpha^4\right] \\
&= \frac{3\alpha^2 + 6\alpha}{\alpha^2} \\
&= 3 + \frac{6}{\alpha}.
\end{aligned}
$$

Note that as $\alpha \to \infty$, $\text{skew}(X) \to 0$ and $\text{kurt}(X) \to 3$, which are the skewness and kurtosis of a Normal distribution. This is not surprising: For example, if $X_1, \cdots, X_n$ are independent Gamma random variables with shape parameter 1 then $X_1 + \cdots + X_n$ has a Gamma distribution with shape parameter n and we know (by the CLT) that the distribution of $X_1 + \cdots + X_n$ will also be approximately Normal.

(b) This is much easier than it looks. First of all, we can assume that $E(X_i) = 0$ since we are taking expected values of products of random variables with mean 0. Also $\text{Var}(S_n) = \sigma_1^2 + \cdots + \sigma_n^2$. Then

$$
\begin{aligned}
E(S_n^3) &= E\left[\left(\sum_{i=1}^n X_i\right)^3\right] \\
&= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n E(X_i X_j X_k) \\
&= \sum_{i=1}^n E(X_i^3)
\end{aligned}
$$

where the latter equality holds since $E(X_i X_j X_k) = E(X_i)E(X_j)E(X_k) = 0$ if i, j, k are distinct and $E(X_i^2 X_j) = E(X_i^2)E(X_j) = 0$ if $i \neq j$. Thus

$$
\begin{aligned}
\text{skew}(S_n) &= \frac{E(S_n^3)}{(\sigma_1^2 + \cdots + \sigma_n^2)^{3/2}} \\
&= \frac{1}{(\sigma_1^2 + \cdots + \sigma_n^2)^{3/2}} \sum_{i=1}^n E(X_i^3) \\
&= \frac{1}{(\sigma_1^2 + \cdots + \sigma_n^2)^{3/2}} \sum_{i=1}^n \sigma_i^3 \text{skew}(X_i)
\end{aligned}
$$

4. (a) By the Delta Method

$$
\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, [g'(\mu)]^2 \sigma^2(\mu))
$$

and so $[g'(\mu)]^2\sigma^2(\mu) = 1$, which means that

$$g'(\mu) = \pm\frac{1}{\sigma(\mu)}.$$

(b) For the Poisson distribution $\sigma^2(\mu) = \mu$ and so

$$g'(\mu) = \pm\frac{1}{\sqrt{\mu}}.$$

We can take $g(\mu) = 2\sqrt{\mu}$ (or more generally $g(\mu) = \pm 2\sqrt{\mu} + \text{constant}$).

For the exponential distribution $\sigma^2(\mu) = \mu^2$ and so

$$g'(\mu) = \pm\frac{1}{\mu}.$$

We can take $g(\mu) = \ln(\mu)$ (or more generally $g(\mu) = \pm\ln(\mu) + \text{constant}$).

For the Bernoulli distribution, $\sigma^2(\mu) = \mu(1-\mu)$ and so

$$g'(\mu) = \pm\frac{1}{\sqrt{\mu(1-\mu)}}$$

Solving this differential equation is not easy - however, it turns out that $g(\mu) = 2\sin^{-1}(\sqrt{\mu})$ is a solution.

5. The R code given below gives the calculations for the normal and Cauchy distributions. Note that this code can be condensed somewhat; in fact, we can compute the expected proportion of outliers in a single line. The calculations for the Laplace are analytical.

(a) Normal distribution

```
> IQR <- qnorm(3/4)-qnorm(1/4)
> IQR
[1] 1.348980
> prop <- 2*(1-pnorm(qnorm(3/4)+1.5*IQR))
> prop
[1] 0.006976603
```

(b) Laplace distribution: For $x > 0$, we have $P(X > x) = \exp(-x)/2$ and so the upper quartile $F^{-1}(3/4)$ satisfies $\exp(-F^{-1}(3/4))/2 = 0.25$ — thus $F^{-1}(3/4) = \ln(2)$, the lower quartile $F^{-1}(1/4) = -\ln(2)$, and the interquartile range is $2\ln(2)$. Thus the expected outlier proportion is

$$\exp(-(\ln(2) + 1.5 \times 2\ln(2))) = \exp(-4\ln(2)) = 2^{-4} = 0.0625.$$

(c) Cauchy distribution:

```
> IQR <- qcauchy(3/4)-qcauchy(1/4)
> IQR
[1] 2
> prop <- 2*(1-pcauchy(qcauchy(3/4)+1.5*IQR))
> prop
[1] 0.1559583
```

(d) Generally speaking, heavier tailed distributions tend to produce more outliers (at least, using the boxplot definition). The Laplace distribution actually produces more outliers than one might expect given its light tails. For example, the logistic distribution (which has similar tails to the exponential) produces an expected proportion of outliers of 0.024. The reason for this is the fact that the Laplace distribution is more "peaked" around 0, which makes its interquartile range somewhat smaller than for other densities with similar tails.

6. (a) S_n^2 is the sample variance and should converge in probability to σ^2. To see this, note that

$$\begin{aligned} S_n^2 &= \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \\ &= \frac{n}{n-1}\left\{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2\right\} \\ &\xrightarrow{p} E(X_i^2) - \mu^2 = \sigma^2 \end{aligned}$$

using the WLLN and the fact that $n/(n-1) \to 1$.

(b) By the CLT, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} Z \sim \mathcal{N}(0, \sigma^2)$ and from (a), $S_n \xrightarrow{p} \sigma$. Thus using Slutsky's Theorem, $\sqrt{n}(\bar{X}_n - \mu)/S_n \xrightarrow{d} Z/\sigma \sim \mathcal{N}(0,1)$.

(c) Applying the Delta Method with $g(x) = \exp(x)$, we have $\sqrt{n}(\exp(\bar{X}_n) - \exp(\mu)) \xrightarrow{d} \exp(\mu)Z$ where $Z \sim \mathcal{N}(0,\sigma^2)$ and by Slutsky's Theorem, it follows that $\sqrt{n}(\exp(\bar{X}_n) - \exp(\mu))/S_n \xrightarrow{d} \exp(\mu)Z/\sigma \sim \mathcal{N}(0, \exp(2\mu))$.

(d) This is a little trickier although our intuition tells us that the limit (in probability) should be $E(|X_i - \mu|)$. To see this, we use the triangle inequality to get upper and lower bounds for each $|X_i - \bar{X}_n|$:

$$\begin{aligned} |X_i - \bar{X}_n| &= |X_i - \mu + \mu - \bar{X}_n| \\ &\leq |X_i - \mu| + |\bar{X}_n - \mu| \\ \text{and} \quad |X_i - \mu| &= |X_i - \bar{X}_n + \bar{X}_n - \mu| \\ &\leq |X_i - \bar{X}_n| + |\bar{X}_n - \mu|. \end{aligned}$$

Thus

$$|X_i - \mu| - |\bar{X}_n - \mu| \leq |X_i - \bar{X}_n| \leq |X_i - \mu| + |\bar{X}_n - \mu|$$

and so

$$\frac{1}{n}\sum_{i=1}^{n} |X_i - \mu| - |\bar{X}_n - \mu| \leq \frac{1}{n}\sum_{i=1}^{n} |X_i - \bar{X}_n| \leq \frac{1}{n}\sum_{i=1}^{n} |X_i - \mu| + |\bar{X}_n - \mu|.$$

By the WLLN,

$$\frac{1}{n}\sum_{i=1}^{n} |X_i - \mu| \xrightarrow{p} E(|X_i - \mu|) \quad \text{and} \quad |\bar{X}_n - \mu| \xrightarrow{p} 0$$

and so

$$\frac{1}{n}\sum_{i=1}^{n} |X_i - \bar{X}_n| \xrightarrow{p} E(|X_i - \mu|).$$

7. (a) Using the Taylor series expansion, we have

$$\begin{aligned} g(X_n) - g(\theta) &= g'(\theta)(X_n - \theta) + \frac{1}{2}g''(\theta)(X_n - \theta)^2 + r_n \\ &= \frac{1}{2}g''(\theta)(X_n - \theta)^2 + r_n \end{aligned}$$

Multiplying by a_n^2 gives

$$\begin{aligned} a_n^2(g(X_n) - g(\theta)) &= \frac{1}{2}g''(\theta)a_n^2(X_n - \theta)^2 + a_n^2 r_n \\ &= \frac{1}{2}g''(\theta)\left\{a_n(X_n - \theta)\right\}^2 + a_n^2(X_n - \theta)^2\frac{r_n}{(X_n - \theta)^2} \end{aligned}$$

Now $a_n^2(X_n - \theta)^2 \xrightarrow{d} Z^2$ and $r_n/(X_n - \theta)^2 \xrightarrow{p} 0$; thus by Slutsky's Theorem,

$$a_n^2(X_n - \theta)^2\frac{r_n}{(X_n - \theta)^2} \xrightarrow{p} 0$$

and so (again by Slutsky's Theorem),

$$a_n^2(g(X_n) - g(\theta)) \xrightarrow{d} \frac{1}{2}g''(\theta)Z^2.$$

Note: Z need not be Normal and so the form of the limiting distribution given above is as good as we can do! Of course, if $Z \sim \mathcal{N}(0, \sigma^2)$ then the distribution of Z^2 is σ^2 times a chi-square distribution with 1 degree of freedom.

(b) Following the procedure from part (a), we have

$$\begin{aligned} g(X_n) - g(\theta) &= g'(\theta)(X_n - \theta) + \frac{1}{2}g''(\theta)(X_n - \theta)^2 + \cdots \\ &\quad + \frac{1}{(k-1)!}g^{(k-1)}(\theta)(X_n - \theta)^{k-1} + \frac{1}{k!}g^{(k)}(\theta)(X_n - \theta)^k + r_n \\ &= \frac{1}{k!}g^{(k)}(\theta)(X_n - \theta)^k + r_n \end{aligned}$$

and multiplying by a_n^k, we get

$$\begin{aligned} a_n^k(g(X_n) - g(\theta)) &= \frac{1}{k!}g^{(k)}(\theta)a_n^k(X_n - \theta)^k + a_n^k r_n \\ &= \frac{1}{k!}g^{(k)}(\theta)\left\{a_n(X_n - \theta)\right\}^k + a_n^k(X_n - \theta)^k \frac{r_n}{(X_n - \theta)^k} \\ &\xrightarrow{d} \frac{1}{k!}g^{(k)}(\theta)Z^k \end{aligned}$$

applying Slutsky's Theorem as in part (a).

Note: In parts (a) and (b), we've assumed that the remainder term r_n satisfies, respectively, $r_n/(X_n - \theta)^2 \xrightarrow{p} 0$ and $r_n/(X_n - \theta)^k \xrightarrow{p} 0$. A rigorous proof of this is a little bit more subtle – for example, in part (a), the remainder term r_n satisfies $r_n/(x_n - \theta)^2$ if $\{x_n\}$ is a sequence of numbers converging to θ. The idea behind a rigorous proof lies in using the fact that X_n is close to θ (in which case $r_n/(X_n - \theta)^2$ or $r_n/(X_n - \theta)^k$ is close to 0 for sufficiently large n) with probability close to 1 and so the probability that $r_n/(X_n - \theta)^2$ (or $r_n/(X_n - \theta)^k$) is not close to 0 is small.