ISyE 6412A **Theoretical Statistics** **Fall 2023**

HW #1 (due at midnight on Wednesday, Aug 30, ET.)

(There are 2 questions, and please look at both sides. Peer grading will be assigned at 8:00am on Thursday Aug 31)

Problem 1. Suppose that X is a normal random variable with variance 1 and unknown mean θ. It is desired to guess the value of unknown mean θ. Since the experimenter feels the loss is roughly like square error $(d-\theta)^2$ when the true θ is small but is like squared *relative* error $(\theta^{-1}d - 1)^2$ when $|\theta|$ is large, he or she chooses loss function $(\theta - d)^2/(1+\theta^2)$ to reflect this behavior.

(a) Specify S, Ω, D, and L (i.e., the sample space, the set of all possible distribution functions, the decision space, and the loss function).

Answer: $S = \{x : -\infty < x < \infty\},\ \Omega = \{N(\theta, 1) : \theta \in \mathcal{R}\}$ or simply $\Omega = \{\theta : -\infty < \theta < \theta\},\ D = \{d : -\infty < d < \infty\},\ L(\theta, d) = (\theta - d)^2/(1+\theta^2)$.

(b) Determine and plot on the same graph the risk function of the 6 procedures δ_i defined by

$$\delta_1(X) = X; \qquad \delta_2(X) = (1+X)/2; \qquad \delta_3(X) = X/2;$$
$$\delta_4(X) = 2X; \qquad \delta_5(X) = 0; \qquad \delta_6(X) = 1;$$

[You can save time by working (e) first but may find it easier to work (b) first. Your calculation will be made simpler if you first compute the risk function of a general procedure of the form $\delta(X) = a + bX$. A check: $R_{\delta_4}(\theta) = (\theta^2+4)/(1+\theta^2)$.]

Answer: For $\delta(X) = a + bX$, we have

$$\begin{aligned} R_\delta(\theta) &= \mathbf{E}_\theta L(\theta, \delta(X)) = \mathbf{E}_\theta \frac{(a+bX-\theta)^2}{1+\theta^2} \\ &= \frac{1}{1+\theta^2}\mathbf{E}_\theta (a+bX-\theta)^2 = \frac{1}{1+\theta^2}\Big((a + b\mathbf{E}_\theta X - \theta)^2 + Var(a+bX)\Big) \\ &= \frac{(a+(b-1)\theta)^2 + b^2}{1+\theta^2}, \end{aligned}$$

where we use the fact that $\mathbf{E}(Y-\theta)^2 = (\mathbf{E}Y - \theta)^2 + Var(Y)$. In particular,

$$R_{\delta_1}(\theta) = \frac{1}{1+\theta^2} \text{ (as } a=0, b=1), \qquad R_{\delta_2}(\theta) = \frac{(\theta-1)^2+1}{4(1+\theta^2)} \text{ (as } a=b=1/2),$$
$$R_{\delta_3}(\theta) = \frac{1}{4} \text{ (as } a=0, b=1/2), \qquad R_{\delta_4}(\theta) = \frac{\theta^2+4}{1+\theta^2} \text{ (as } a=0, b=2),$$
$$R_{\delta_5}(\theta) = \frac{\theta^2}{1+\theta^2} \text{ (as } a=b=0), \qquad R_{\delta_6}(\theta) = \frac{(\theta-1)^2}{1+\theta^2} \text{ (as } a=1, b=0).$$

The risk functions of these 6 procedures are plotted in Figure 1.

(c) From these calculations, can you assert that any of these six procedures is inadmissible?

Answer: $\delta_4(X) = 2X$ is inadmissible, since δ_1 (or δ_5) is better than δ_4.

(d) On the basis of the risk functions, if one of these 6 procedures must be used, which procedure would you use, and why? (*Note:* Don't consult any references in answering this. Later you will find out the precise meaning of your present intuition.)

Answer: Probably use $\delta_3(X) = X/2$, due to its constant risk $1/4$ at every θ. However, any other reasonable answers will receive full credits (but definitely not $\delta_4(X)$).

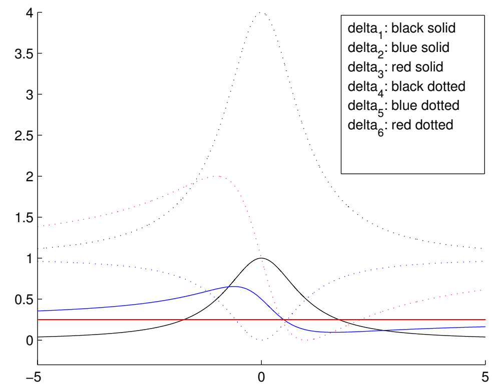

Figure 1: plot of the risk functions of the 6 procedures

(e) Suppose X is replaced by the vector $(X_1, \ldots, X_n)$ of iid normal $N(\theta, 1)$ random variables. The procedures corresponding to $\delta_1, \delta_2, \delta_3, \delta_6$ are

$$\delta_{1,n}(X_1, \ldots, X_n) = \bar{X}_n; \qquad \delta_{2,n}(X_1, \ldots, X_n) = \frac{\bar{X}_n + n^{-1}}{1 + n^{-1}};$$

$$\delta_{3,n}(X_1, \ldots, X_n) = \frac{\sqrt{n}\ \bar{X}_n}{1 + \sqrt{n}}; \qquad \delta_{6,n}(X_1, \ldots, X_n) = 1.$$

Compute the risk functions of these four procedures, and plot graphs of these four risk functions (or, rather, of $nR_{\delta_i,n}$ to make the results comparable to those of part (b)) for n large (e.g., for $n = 10,000$). [Use the fact that $\bar{X}_n$ is $N(\theta, n^{-1})$ distributed. Again, you may find it is easier first to find $(1+\theta^2)^{-1}\mathbf{E}_\theta(a + b\bar{X}_n - \theta)^2$ for general a, b.]

Answer: Note that in this case, we have
$S = \{(x_1, \cdots, x_n) : -\infty < x_i < \infty \text{ for all } i = 1, \cdots, n\}$ or simply $\mathcal{R}^n$,
$\Omega = \{N(\theta, 1) : \theta \in \mathcal{R}\}$ or simply $\Omega = \{\theta : -\infty < \theta < \theta\}$ or $\mathcal{R}$,
$D = \{d : -\infty < d < \infty\}$, or simply $D = \mathcal{R}$ and
$L(\theta, d) = (\theta - d)^2/(1 + \theta^2)$.
Denote $\mathbf{X} = (X_1, \ldots, X_n)$ to simplify the notation. For $\delta_{a,b,n}(\mathbf{X}) = a + b\bar{X}_n$, we have

$$\begin{aligned} R_{\delta_{a,b,n}}(\theta) &= \mathbf{E}_\theta L(\theta, \delta_{a,b,n}(\mathbf{X})) = \mathbf{E}_\theta \frac{(a + b\bar{X}_n - \theta)^2}{1 + \theta^2} \\ &= \frac{1}{1 + \theta^2}\Big((a + b\mathbf{E}_\theta \bar{X}_n - \theta)^2 + Var(a + b\bar{X}_n)\Big) \\ &= \frac{(a + (b-1)\theta)^2 + b^2 \frac{1}{n}}{1 + \theta^2}. \end{aligned}$$

In particular,

$$R_{\delta_{1,n}}(\theta) = \frac{1}{n(1+\theta^2)} \text{ (as } a = 0, b = 1), \qquad R_{\delta_{2,n}}(\theta) = \frac{(\theta - 1)^2 + n}{(1+n)^2(1+\theta^2)} \text{ (as } a = 1 - b = \tfrac{1}{1+n}),$$

$$R_{\delta_{3,n}}(\theta) = \frac{1}{(1+\sqrt{n})^2} \text{ (as } a = 0, b = \tfrac{\sqrt{n}}{1+\sqrt{n}}), \qquad R_{\delta_{6,n}}(\theta) = \frac{(\theta - 1)^2}{1+\theta^2} \text{ (as } a = 1, b = 0).$$

Fig 2 plots the function $nR_{\delta_i,n}$.

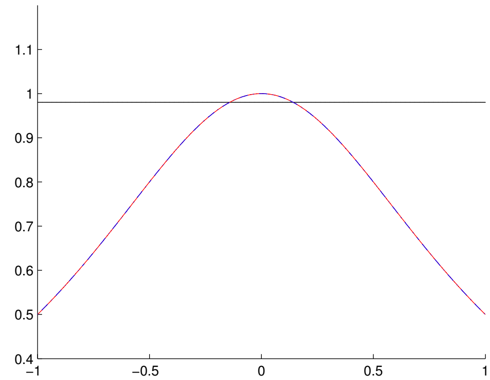

Figure 2: plot of the function $nR_{\delta_i,n}$. Red: $\delta_{1,n}$; Blue: $\delta_{2,n}$; and Black: $\delta_{3,n}$. The plot shows that the curves for $\delta_{1,n}$ and $\delta_{2,n}$ are almost identical.

(f) If n is large, which of the four procedures of part (e) would you use, and why?
(Your answer to this last may differ from the answer to part (d) for the case $n = 1$; does it?)

Answer: From Fig 2, the benefit of constant risk of $\delta_{3,n}$ seems to be not so significant for large n. Thus, one may prefer to use $\delta_{1,n}$ or $\delta_{2,n}$ when n is large.

(g) Suppose the statistician decides to restrict consideration to procedures $\delta_{a,b,n} = a + b\bar{X}_n$ of the form mentioned at the end of (e). He or she is concerned about the behavior of the risk function when $|\theta|$ is large. Show that the risk function approaches 0 as $|\theta| \to \infty$ if and only if $b = 1$. In addition, *among procedures with* $b = 1$, show that the choice $a = 0$ gives uniformly smallest risk function.

[This justification of the procedure $\delta_{1,n} = \bar{X}_n$ under the restriction to procedures of the form $\delta_{a,b,n}$ will seem more sensible to many people than a justification in terms of the "unbiasedness" criterion to be discussed later].

Answer: It is evident from the above computation that as $|\theta|$ goes to ∞, the risk $R_{\delta_{a,b,n}}(\theta)$ converges to the constant value $(b-1)^2$, which is 0 if and only if $b = 1$. Among all procedures $\delta_{a,b,n}$ with $b = 1$, for all real number a, we have

$$R_{\delta_{a,b=1,n}}(\theta) \quad = \quad \frac{a^2 + b^2\frac{1}{n}}{1+\theta^2} \geq \frac{b^2\frac{1}{n}}{1+\theta^2} = R_{\delta_{a=0,b=1,n}}(\theta),$$

Thus the choice $a = 0$ gives uniformly smallest risk function among all procedures $\delta_{a,b,n}$ with $b = 1$.

(h) Show that the procedure $\delta_{6,n}$, defined by $\delta_{6,n}(X_1, \ldots, X_n) \equiv 1$, is admissible for each n. [Hints: how can another procedure δ' satisfy $R_{\delta'}(\theta) \leq R_{\delta_{6,n}}(\theta)$ when $\theta = 1$?]

Answer: We will prove it by contradiction. Assume that $\delta_{6,n}(X_1, \ldots, X_n) \equiv 1$ were inadmissible, and there were a procedure δ' which would be better than $\delta_{6,n}$, or equivalently, the procedure δ' satisfies $R_{\delta'}(\theta) \leq R_{\delta_{6,n}}(\theta)$ for all θ, with inequality holding for at least one θ. Then we have

$$\mathbf{E}_\theta \frac{(\delta'(\mathbf{X}) - \theta)^2}{1+\theta^2} \leq \mathbf{E}_\theta \frac{(\delta_{6,n} - \theta)^2}{1+\theta^2} \text{ for all } -\infty < \theta < \infty.$$

For any given θ, the term $1/(1+\theta^2)$ is constant, and thus this can be written as

$$\frac{1}{1+\theta^2}\mathbf{E}_\theta(\delta'(\mathbf{X})-\theta)^2 \le \frac{1}{1+\theta^2}\mathbf{E}_\theta(\delta_{6,n}-\theta)^2$$

and canceling the term $1+\theta^2$ both sides yield that

$$\mathbf{E}_\theta(\delta'(\mathbf{X})-\theta)^2 \le \mathbf{E}_\theta(\delta_{6,n}-\theta)^2 = (1-\theta)^2 \text{ for all } -\infty < \theta < \infty. \tag{1}$$

Rewriting the definition of $\mathbf{E}_\theta(\delta'(\mathbf{X})-\theta)^2$ in term of the joint density function of $\mathbf{X}$, we have

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Big[(\delta'(x_1,\ldots,x_n)-\theta)^2\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_1-\theta)^2}{2})\Big)\cdots\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_n-\theta)^2}{2})\Big)dx_1\cdots dx_n \le (1-\theta)^2, \tag{2}$$

for all $-\infty < \theta < \infty$, where we use the fact that $\phi(u) = (1/\sqrt{2\pi})\exp(-u^2/2)$ is the pdf of $N(0,1)$.
Now we consider the special case of $\theta = 1$, in which the above inequality becomes

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Big[(\delta'(x_1,\ldots,x_n)-1)^2\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_1-1)^2}{2})\Big)\cdots\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_n-1)^2}{2})\Big)dx_1\cdots dx_n \le 0.$$

Note that the integrand is non-negative and the terms $\exp(-\frac{(x_i-1)^2}{2}) > 0$ for all x_i, the above relation holds if and only if $\delta'(x_1,\ldots,x_n) = 1$ almost everywhere, i.e., $\delta'(\mathbf{X}) = 1$ a.e. Hence, for every θ,

$$\begin{aligned}
R_{\delta'}(\theta) &= \mathbf{E}_\theta\frac{(\delta'(\mathbf{X})-\theta)^2}{1+\theta^2} \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{(\delta'(x_1,\ldots,x_n)-\theta)^2}{1+\theta^2}\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_1-\theta)^2}{2})\Big)\cdots\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_n-\theta)^2}{2})\Big)dx_1\cdots dx_n \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{(1-\theta)^2}{1+\theta^2}\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_1-\theta)^2}{2})\Big)\cdots\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_n-\theta)^2}{2})\Big)dx_1\cdots dx_n \\
&\qquad \text{(since } \delta'(\mathbf{X}) = 1 \text{ a.e.)} \\
&= \frac{(1-\theta)^2}{1+\theta^2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_1-\theta)^2}{2})\Big)\cdots\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_n-\theta)^2}{2})\Big)dx_1\cdots dx_n \\
&= \frac{(1-\theta)^2}{1+\theta^2} \quad \text{(since the intergal is just } \mathbf{E}_\theta(1) = \textstyle\int\cdots\int f_\theta(x_1,\cdots,x_n)dx_1\cdots dx_n = 1) \\
&= R_{\delta_{6,n}}(\theta),
\end{aligned}$$

i.e., $R_{\delta'}(\theta) = R_{\delta_{6,n}}(\theta)$ for all $-infty < \theta < \infty$. This is a contradiction with our assumption that the procedure δ' is better than $\delta_{6,n}$! So $\delta_{6,n}$ is admissible for all n.

Problem 2. Assume that we observe a binomial random variable X with parameter (n,θ), i.e., the probability mass function of X is given by $\mathbf{P}(X=i) = \binom{n}{i}\theta^i(1-\theta)^{n-i}$ for $i = 0,1,\ldots,n$, where $n \ge 1$ is a known integer and $0 \le \theta \le 1$ is unknown. Consider the problem of estimating θ under the so-called "absolute deviation" loss function defined by $L(\theta,d) = |\theta - d|$.

(a) Specify S, Ω, D, and L (i.e., the sample space, the set of all possible distribution functions, the decision space, and the loss function).

(b) When $n = 20$, graph and compare the risk functions of the following three procedures:

$$\delta_1(X) = \frac{X}{n}, \qquad \delta_2(X) = \frac{1}{3}, \qquad \text{and} \qquad \delta_3(X) = 1.$$

Note that the risk functions may not have simple expressions, and it will be OK to use some computer software to plot the risk functions.

(c) Show that for any given integer $n \geq 1$, the procedure $\delta_2(X) = 1/3$ is admissible. [Hints: how can another procedure δ' satisfy $R_{\delta'}(\theta) \leq R_{\delta_2}(\theta)$ when $\theta = 1/3$?]

(d) Show that when $n = 2$, the procedure $\delta_3(X) = 1$ is admissible.

Remarks: Parts (c) and (d) suggest that an admissible estimator may not be appealing. Of course, it is clear that inadmissible estimators are definitely not desirable.

In Part (b), the following R code can be used to plot the risk functions. For more information about the free statistical software R, please see the website <http://www.r-project.org/>.

```
theta <- seq(0,1,0.0001);
R1 <- 0;
for (i in 0:20){
 R1 <- R1+choose(20,i)*(theta^i)*((1-theta)^(20-i))*abs(i/20 - theta);
}
R2 <- abs(1/3  - theta);
R3 <- abs(1  - theta);
plot(theta, R1,"l", ylab="Risk Function", ylim=c(0,1));
lines(theta, R2, col="red");
lines(theta, R3, col="blue")
```

Answer: (a) $S = \{0, 1, 2, \ldots, n\}$, $\Omega = \{\text{Binomial}(n, \theta) : 0 \leq \theta \leq 1\}$ or simply $\Omega = \{\theta : 0 \leq \theta \leq 1\}$ or $\Omega = [0, 1]$, $D = \{d : 0 \leq d \leq 1\} = [0, 1]$, $L(\theta, d) = |\theta - d|$.
(b) The risk functions are

$$\begin{aligned} R_{\delta_1}(\theta) &= \mathbf{E}_\theta L(\theta, \delta_1(X)) = \mathbf{E}_\theta |\frac{X}{n} - \theta| = \sum_{i=0}^{n} |\frac{i}{n} - \theta| \mathbf{P}_\theta(X = i) \\ &= \sum_{i=0}^{n} \binom{n}{i} \theta^i (1-\theta)^{n-i} |\frac{i}{n} - \theta| \end{aligned}$$

and

$$R_{\delta_2}(\theta) = \mathbf{E}_\theta L(\theta, \delta_2(X)) = \mathbf{E}_\theta |\frac{1}{3} - \theta| = |\frac{1}{3} - \theta|.$$

See Fig 3 for the plot.
(c) Assume that δ' is a better procedure than $\delta_2 = 1/3$. Then $0 \leq R_{\delta'}(\theta) \leq R_{\delta_2}(\theta) = |\theta - 1/3|$ for all $0 \leq \theta \leq 1$. Since

$$\begin{aligned} R_{\delta'}(\theta) &= \mathbf{E}_\theta L(\theta, \delta'(X)) = \mathbf{E}_\theta |\delta'(X) - \theta| \\ &= \sum_{i=0}^{n} |\delta'(i) - \theta| \mathbf{P}_\theta(X = i) \\ &= \sum_{i=0}^{n} |\delta'(i) - \theta| \binom{n}{i} (\theta)^i (1-\theta)^{n-i}, \end{aligned}$$

we have that for any $0 \leq \theta \leq 1$,

$$\sum_{i=0}^{n} \left|\delta'(i) - \theta\right| \binom{n}{i} (\theta)^i (1-\theta)^{n-i} \leq |1/3 - \theta|.$$

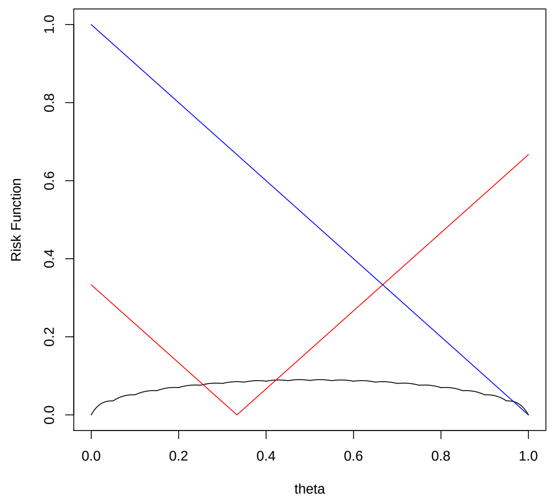

Figure 3: Plot of the risk functions: back curve for $R_{\delta_1}(\theta)$ and red curve for $R_{\delta_2}(\theta)$.

Letting $\theta = 1/3$ both sides yields that

$$\sum_{i=0}^{n} |\delta'(i) - \frac{1}{3}| \binom{n}{i} (\frac{1}{3})^i (\frac{2}{3})^{n-i} \leq 0.$$

Since all terms on the left-hand side are non-negative, we have $\delta'(i) - \frac{1}{3} = 0$ for all $i = 0, 1, \ldots, n$ and this implies that

$$R_{\delta'}(\theta) = |1/3 - \theta| \sum_{i=0}^{n} \binom{n}{i} (\theta)^i (1-\theta)^{n-i} = |1/3 - \theta| = R_{\delta_2}(\theta)$$

for all θ (in fact, we have $\delta'(X) = \frac{1}{3} = \delta_2(X)$ in this case). This is a contradiction with our assumption that the procedure δ' is better than δ_2! So $\delta_2(X) \equiv 1/3$ is admissible.

(d) Assume that $\delta_3(X) \equiv 1$ were inadmissible, and the existed a procedure $\delta^*(X)$ that better than $\delta_3(X) \equiv 1$. Then $0 \leq R_{\delta^*}(\theta) \leq R_{\delta_3}(\theta) = |1 - \theta|$ for all $0 \leq \theta \leq 1$. When $n = 2$, we have that for all $0 \leq \theta \leq 1$, we have

$$\big|\delta^*(0) - \theta\big|(1-\theta)^2 + \big|\delta^*(1) - \theta\big|(2\theta(1-\theta)) + \big|\delta^*(2) - \theta\big|(\theta)^2 \quad \leq \quad |1 - \theta|. \tag{3}$$

Here $\delta^*(0), \delta^*(1)$ and $\delta^*(2)$ are three real-valued constants that does not depend on θ, and we need to investigate these three constants.

It involves three steps. **First,** if we let $\theta = 1$, equation (3) becomes $0 + 0 + |\delta^*(2) - 1| \leq 0$, and thus $\delta^*(2) = 1$. Note that this question is different from problem 1(h) or 2(c), in the sense that $\delta^*(\mathbf{X}) = 1$ a.e. under probability measure $\mathbf{P}_{\theta=1}$ only implies that $\delta^*(X = 2) = 1$, but it does not explicitly tell us the values of $\delta^*(X)$ for $X = 0$ or 1. The main mathematical reason is that a set can have probability 0 under $\mathbf{P}_{\theta=1}$, but does not necessarily have probability 0 under other probability measure $\mathbf{P}_{\theta}$ when $0 \leq \theta < 1$.

Second, to derive the value of $\delta^*(1)$, we plug the relation $\delta^*(2) = 1$ back into (3),

$$\big|\delta^*(0) - \theta\big|(1-\theta)^2 + \big|\delta^*(1) - \theta\big|(2\theta(1-\theta)) \quad \leq \quad |1 - \theta|\big(1 - \theta^2\big) \tag{4}$$

for all $0 \leq \theta \leq 1$. Dividing both-sides by $1 - \theta > 0$, we have

$$\big|\delta^*(0) - \theta\big|(1-\theta) + \big|\delta^*(1) - \theta\big|(2\theta) \leq |1 - \theta|\big(1 + \theta\big) \tag{5}$$

for all $0 \leq \theta < 1$. It is crucial to observe that while (5) is derived only for $0 \leq \theta < 1$, both left-hand and right-hand sides of (5) are continuous functions of θ. Thus, (5) still holds if we let $\theta \to 1$. In other words, the continuous property implies that (5) holds for all $0 \leq \theta \leq 1$. Letting $\theta = 1$ in (5), we now have $\big|\delta^*(1) - 1\big| \leq 0$ and thus $\delta^*(1) = 1$.

The third step is to find the value of $\delta^*(0)$. Plugging the just proved result $\delta^*(1) = 1$ back to (4), we have

$$\big|\delta^*(0) - \theta\big|(1-\theta)^2 \leq |1-\theta|(1-\theta^2-2\theta(1-\theta)) = |1-\theta|(1-\theta)^2$$

for all $0 \leq \theta \leq 1$. When $0 \leq \theta < 1$, divided both sides by $(1-\theta)^2 > 0$, we have

$$\big|\delta^*(0) - \theta\big| \leq |1-\theta|$$

for all $0 \leq \theta < 1$. Again, both sides are continuous function of θ, and letting $\theta \to 1$ implies that this relation hold for all $0 \leq \theta \leq 1$. Letting $\theta = 1$ yields that $|\delta^*(0) - 1| \leq 0$ and thus $\delta^*(0) = 1$.

Hence, we have showed that if $\delta^*(X)$ were a better procedure than $\delta_3(X) \equiv 1$, then we must have $\delta^*(X) \equiv 1$ for $X = 0, 1, 2$. This means that $R_{\delta^*}(\theta) = |1-\theta| = R_{\delta_3}(\theta)$ for all $0 \leq \theta \leq 1$. This is a contradiction with our assumption that the procedure δ^* is better than $\delta_3 \equiv 1$!!! So $\delta_3(X) \equiv 1$ is admissible when $n = 2$.

Remarks: Note that we essentially use (3) three times to show that $\delta^*(x) = 1$ for all $x = 0, 1, 2$. Some students may feel confused why we need to use (3) three times. To illustrate this, you can consider the procedure

$$\delta_d(X) = \begin{cases} 1/2, & \text{if } X = 0; \\ 1/2, & \text{if } X = 1; \\ 1, & \text{if } X = 2. \end{cases}$$

As in (3), the risk function of δ_d is

$$R_{\delta_d}(\theta) = \Big|\frac{1}{2} - \theta\Big|(1-\theta)^2 + \Big|\frac{1}{2} - \theta\Big|(2\theta(1-\theta)) + \Big|1 - \theta\Big|(\theta)^2$$

and clearly $R_{\delta_d}(\theta = 1) = 0$. That is, δ_d satisfies (3) in the special case of $\theta = 1$ but not in other cases of $0 \leq \theta < 1$. The reason is that $\delta_d = 1$ **with probability** 1 under probability $\mathbf{P}_{\theta=1}$, as $\mathbf{P}_{\theta=1}(X = 2) = 1$, but clearly the probability of $\delta_d = 1$ is not 1 under other probability $\mathbf{P}_\theta$ when $0 \leq \theta < 1$.

In general, in probability we only have one specific probability model (or probability measure), but in statistics, we have a sequence of probability models (or probability measures) that depend on the value of θ. Thus it is important in the statistical problem to emphasize which probability model/measure we are talking about, and it will be better to write $\mathbf{P}_{\theta=\theta_0}$ (or $\mathbf{P}_\theta$ or $\mathbf{P}_{F_\theta}$) instead of $\mathbf{P}$. This is a major difference between statistics and probability.