Notes for Math 149C: Probability and Mathematical Statistics
Miguel Carrión Álvarez

§7.3#7.19. Let $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ be the order statistics of a random sample of size 5 from the uniform distribution having p.d.f.

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ 0 & \text{elsewhere.} \end{cases}$$

Show that $2Y_3$ is an unbiased estimator for $\theta$. Determine the joint p.d.f. of $Y_3$ and the sufficient statistic $Y_5$ for $\theta$. Find the conditional expectation $E(2Y_3 \mid y_3) = \varphi(y_3)$. Compare the variances of $2Y_3$ and $\varphi(Y_3)$.

**Short answer.**

The p.d.f. of $Y_3$ is

$$f_{Y_3}(y_3; \theta) = \frac{5!}{2!2!} \frac{y_3^2(\theta - y_3)^2}{\theta^5}, \quad 0 < y_3 < \theta,$$

which is symmetric under the exchange $y_3 \leftrightarrow (\theta - y_3)$, so the expected value of $Y_3$ must satisfy

$$E(Y_3) = \theta - E(Y_3),$$

and

$$E(2Y_3) = 2E(Y_3) = \theta.$$

The joint p.d.f. of $Y_3$ and $Y_5$ is

$$f_{Y_3, Y_5}(y_3, y_5; \theta) = \frac{5!}{2!} \frac{y_3^2(y_5 - y_3)}{\theta^5}, \quad 0 < y_3 < y_5 < \theta,$$

and the p.d.f. of $Y_5$ is

$$f_{Y_5}(y_5; \theta) = \frac{5!}{4!} \frac{y_5^4}{\theta^4}, \quad 0 < y_5 < \theta,$$

so the conditional p.d.f. of $Y_3$ given $y_5$ is

$$f_{Y_3}(y_3 \mid y_5; \theta) = \frac{f_{Y_3, Y_5}(y_3, y_5; \theta)}{f_{Y_5}(y_5; \theta)} = \frac{4!}{2!} \frac{y_3^2(y_5 - y_3)}{y_5^4}, \quad 0 < y_3 < y_5,$$

which is independent of $\theta$, as it should be since $Y_5$ is a sufficient statistic for $\theta$.

The conditional expectation $E(2Y_3 \mid y_3)$ is

$$\varphi(y_3) = \int_0^{y_3} 2y_3 f_{Y_3}(y_3 \mid y_3) dy_3 = 4! \int_0^{y_3} \frac{y_3^2(y_5 - y_3)}{y_5^4} dy_3 = \frac{4!3!}{5!} y_3 = \frac{6}{5} y_3.$$

We have

$$E(Y_3^2) = \frac{5!}{2!2!} \int_0^\theta \frac{y_3^4(\theta - y_3)^2}{\theta^5} dy_3 = \frac{5!4!}{2!7!} \theta^2 = \frac{2}{7} \theta^2,$$

so

$$\text{Var}(2EY_3) = \frac{8}{7} \theta^2 - \theta^2 = \frac{1}{7} \theta^2.$$

On the other hand, $\varphi(Y_3) = \frac{6}{5} Y_3$,

$$E(y_3) = \frac{5!}{4!} \int_0^\theta \frac{y_3^4}{\theta^4} dy_3 = \frac{5!4!}{4!6!} \theta = \frac{5}{6} \theta$$

and

$$E(y_3^2) = \frac{5!}{4!} \int_0^\theta \frac{y_3^6}{\theta^6} dy_3 = \frac{5!6!}{4!7!} \theta^2 = \frac{5}{7} \theta^2,$$

so

$$\text{Var}(\frac{6}{5} Y_3) = \frac{6^2}{5^2} \left[ \frac{5}{7} - \left( \frac{5}{6} \right)^2 \right] \theta^2 = \frac{1}{35} \theta^2.$$

We see that $\text{Var}(2Y_3) = 5 \text{Var}(\varphi(Y_3))$. 

1
Behind the scenes.

We have a random sample of size 5, \(X_1, X_2, \ldots, X_5\) drawn from the uniform distribution \(U(0, \theta)\). The order statistics \(Y_1 < Y_2 < \ldots < Y_5\) are the result of ordering the \(X_i\) in increasing order.

**p.d.f. of \(Y_3\).** The probability that \(Y_3\) falls between \(y_3\) and \(y_3 + dy_3\) is the probability that two of the \(X_i\) are smaller than \(y_3\), one falls between \(y_3\) and \(y_3 + dy_3\), and two are larger than \(y_3 + dy_3\) (but always smaller than \(\theta\)). This is

\[
f_{Y_3}(y_3; \theta)dy_3 = \frac{5!}{2!2!1!2!}\left(\frac{y_3 - 0}{\theta}\right)^2\left(\frac{\theta - y_3 - dy_3}{\theta}\right)^2 dy_3 = \frac{5!}{2!2!} \frac{y_3^2(\theta - y_3)^2}{\theta^5} dy_3 + O(dy_3^2).
\]

The term \(y_3/\theta\) is the probability that an \(X_i\) falls between 0 and \(y_3\), \(dy_3/\theta\) is the probability that it falls between \(y_3\) and \(y_3 + dy_3\), and \((\theta - y_3)/\theta\) is the probability that it falls between \(y_3\) and \(\theta\). The number \(\frac{5!}{2!2!}\) is the number of ways to distribute 5 items in three groups, such that the first group consists of two objects, the second group of one object, and the third of two objects. The 5 items are the random variables \(X_i\), and the three groups are the intervals \((0, y_3)\), \((y_3, y_3 + dy_3)\) and \((y_3 + dy_3, \theta)\).

The conclusion is

\[
f_{Y_3}(y_3; \theta)dy_3 = \frac{5!}{2!} \frac{y_3^2(\theta - y_3)^2}{\theta^5} dy_3.
\]

**Joint p.d.f. of \(Y_3\) and \(Y_5\).** By the same method of the previous paragraph,

\[
f_{Y_3,Y_5}(y_3, y_5; \theta) = \frac{5!}{2!1!1!1!}\left(\frac{y_3 - 0}{\theta}\right)^2 dy_3 \frac{y_5 - y_3 - dy_3 - dy_5}{\theta} \frac{dy_3 dy_5}{\theta} = \frac{5!}{2!} \frac{y_3^2(y_5 - y_3)}{\theta^5} dy_3 dy_5 + O(dy_3^2),
\]

so

\[
f_{Y_3,Y_5}(y_3, y_5; \theta) = \frac{5!}{2!} \frac{y_3^2(y_5 - y_3)}{\theta^5} dy_3 dy_5.
\]

**Doing Beta integrals.** All the integrals (expectations and variances) in the problem are of the Beta type

\[
\int_0^\theta y^m(\theta - y)^n dy,
\]

which can be calculated by observing that \(y^m(\theta - y)^n\) is proportional to the p.d.f. of the \((m + 1)\)st order statistic of a random sample of size \(m + n + 1\) drawn from a uniform distribution on \((0, \theta)\). That is,

\[
\int_0^\theta \frac{(m + n + 1)!}{m!n!} \frac{y^m(\theta - y)^n}{\theta^{m+n+1}} dy = 1,
\]

where the integrand is the correctly normalized p.d.f. of the order statistic, and it must therefore integrate to 1. We can solve for the original Beta integral to obtain

\[
\int_0^\theta y^m(\theta - y)^n dy = \frac{m!n!}{(m + n + 1)!} \theta^{m+n+1}.
\]