Approximation for the Order Statistic Error

by

Kazuhiro Iwasawa*
Jonathan Goodman†

December 2, 1998

1 Introduction

In this paper, an approximation formula for any order statistics will be given. It is based on Taylor expanding the Distribution function near the confidence interval. The remarkable feature of this formula is that it gives much simpler formula compared to the one given by David and Johnson.

2 Preliminary Results

Theorem 1 (Approximation Theorem for the Order Statistic Error) *Suppose we have i-th order statistic given n points. Assume that i and n are large anough. Then, the order statistic error for any distribution can be approximated by the following formula:* $Error \approx \sqrt{\frac{(1-F)F}{(F')^2(n-1)}}$ *where* $F = \frac{(i-1)}{(n-1)}$.

$(Proof)$

Suppose we are given density function f(x) and its distribution function F(x), then the distribution funxtion at x for the i-th order statistic out of n saples is given by:

$$
\begin{aligned}
F_{i:n}(x) &= Pr(X_{i:n} \leq x) \\
&= Pr(\text{at least i of X's are less than x}) \\
&= \sum_{r=i}^{n} \binom{n}{r} (F(x))^r (1-F(x))^{n-r}.
\end{aligned}
$$

*Princeton University, Department of Mathematics
†New York University, Department of Mathematics

Therefore, differentiating the above formula with respect to x to obtain its density function as:

$$
\begin{aligned}
f_{i:n}(x) &= \frac{\partial F_{i:n}(x)}{\partial x} \\
&= \sum_{r=i}^{n} \left[\frac{n!}{(r-1)!(n-r)!}(F(x))^{r-1}(1-F(x))^{n-r} - \frac{n!}{r!(n-r-1)!}(F(x))^{r}(1-F(x))^{n-r-1} \right] f(x) \\
&= \frac{n!}{(i-1)!(n-i)!}(F(x))^{i-1}(1-F(x))^{n-i} f(x).
\end{aligned}
$$

There is only one term for the density function due to the fact that the above summation is a telescoping sum resulting in calcellation of all terms except the very first term.

Next, we would like to express the above function in terms of something we are familiear with.

If we are interested in the local behavior of this density function near the confidence interval say, μ. Then we can Taylor expand this density function around the μ so that we have: $x = \mu + \Delta x$, if the function F(x) and f(x) are smooth enough (Let's for the sake of simplicity, assume that they are C^{∞}). Then we can put the Taylor expansion for F(x) into the above density function to yield:

$$
\begin{aligned}
f_{i:n}(x) &= \frac{n!}{(i-1)!(n-i)!}(F(x))^{i-1}(1-F(x))^{n-i} f(x) \\
&= \frac{n!}{(i-1)!(n-i)!}(\sum_{k=0}^{\infty} \frac{F^k(\mu)}{k!}(\Delta x)^k)^{i-1}(1-\sum_{k=0}^{\infty} \frac{F^k(\mu)}{k!}(\Delta x)^k)^{n-i} \sum_{k=1}^{\infty} \frac{F^k(\mu)}{(k-1)!}(\Delta x)^{k-1}.
\end{aligned}
$$

If Δx is very samll, then one can truncate the above series beyond $(\Delta x)^3$ to get the following approximation for the density function:

$$
f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!}(F(x))^{i-1}(1-F(x))^{n-i} f(x)
$$

$$
\approx \frac{n!}{(i-1)!(n-i)!}(F+F^{'}\Delta x+\frac{1}{2}F^{''}(\Delta x)^2)^{i-1}(1-F-F^{'}(\Delta x)-\frac{1}{2}F^{''}(\Delta x)^2)^{n-i}(F^{'}+F^{''}(\Delta x))
$$

$$
= \frac{n!}{(i-1)!(n-i)!}F^{i-1}(1-F)^{n-i}(1+\frac{F^{'}}{F}(\Delta x)+\frac{1}{2}\frac{F^{''}}{F}(\Delta x)^2)^{i-1}(1-\frac{F^{'}}{(1-F)}(\Delta x)-\frac{1}{2}\frac{F^{''}}{(1-F)}(\Delta x)^2)^{n-i}
$$

$$
\cdot(F^{'}+F^{''}(\Delta x)).
$$

$$\approx \frac{n!}{(i-1)!(n-i)!} F^{i-1}(1-F)^{n-i}(F^{'}+F^{''}(\Delta x))$$

$$\cdot e^{\left[(\frac{F^{'}}{F}(\Delta x)+\frac{1}{2}\frac{F^{''}}{F}(\Delta x)^2-\frac{1}{2}(\frac{F^{'}}{F})^2(\Delta x)^2)(i-1)+(-\frac{F^{'}}{(1-F)}(\Delta x)-\frac{1}{2}\frac{F^{''}}{(1-F)}(\Delta x)^2-\frac{1}{2}(\frac{F^{'}}{(1-F)})^2(\Delta x)^2)(n-i)\right]}.[1]$$

On other hand, if we collect the coefficient for Δx, we can quickly discover that the coefficient for Δx dissaperars if and only if $F = \frac{i-1}{n-1}$, which is precisely what the distribution F satisfies near the small neighborhood of μ.

Therefore, we can simplify the density function further to abtain:

$$f_{i:n}(x) \approx \frac{n!}{(i-1)!(n-i)!} F^{i-1}(1-F)^{n-i}(F^{'}+F^{''}(\Delta x))$$

$$\cdot e^{\left[(\frac{1}{2}\frac{F^{''}}{F}(\Delta x)^2-\frac{1}{2}(\frac{F^{'}}{F})^2(\Delta x)^2)(i-1)+(-\frac{1}{2}\frac{F^{''}}{(1-F)}(\Delta x)^2-\frac{1}{2}(\frac{F^{'}}{(1-F)})^2(\Delta x)^2)(n-i)\right]}.$$

$$= \frac{n!}{(i-1)!(n-i)!} F^{i-1}(1-F)^{n-i}(F^{'}+F^{''}(\Delta x))$$

$$\cdot e^{\left[(\frac{1}{2}\frac{F^{''}}{F}(\Delta x)^2-\frac{1}{2}(\frac{F^{'}}{F})^2(\Delta x)^2)(n-1)F+(-\frac{1}{2}\frac{F^{''}}{(1-F)}(\Delta x)^2-\frac{1}{2}(\frac{F^{'}}{(1-F)})^2(\Delta x)^2)(n-1)F\right]}.[2]$$

$$= \frac{n!}{(i-1)!(n-i)!} F^{i-1}(1-F)^{n-i}(F^{'}+F^{''}(\Delta x))e^{-\frac{1}{2}\frac{(F^{'})^2(n-1)}{(1-F)F}(\Delta x)^2}.$$

$$\approx \frac{n!}{(i-1)!(n-i)!} F^{i-1}(1-F)^{n-i}(F^{'})e^{-\frac{1}{2}\frac{(F^{'})^2(n-1)}{(1-F)F}(\Delta x)^2}.$$

Here, we made further simplification by ignoring $F^{''}(\Delta x)$ term so that all factors outside the exponential becomes a constant. This step is justified since we assume that i and n are large enough, and therefore Δx is negligible above. Therefore reducing the formula to approximate Gaussian with variance of $\frac{(1-F)F}{(F^{'})^2(n-1)}$. Consequently, one can estimate the error for the i-th order statistic given n points as $\sqrt{\frac{(1-F)F}{(F^{'})^2(n-1)}}$.

[1] We have used the following approximation: $1+\epsilon \approx e^{(\epsilon+(\epsilon)^2)}$

[2] We have used the following relation implied by $F = \frac{(i-1)}{(n-1)}$: $(n-i) = (n-1)(1-F)$

$QED.$

One word of caution here. This approximation is good only for large enough n, and large enough i. Otherwise, Δx becomes relatively large compared to μ and the Taylor approxiamtion above no longer becomes accurate. Also, $F^{'}$ can be approximated numerically by finite difference scheme.

References

[BC] Balakrishnan, N., and Cohen, A. Clifford., (1991). *Order Statistics and Inference, Estimation Methods.* Academic Press, Inc.

[D] David, H.A. (1981). *Order Statistics.* John Wiley and Sons, Inc.

[DJ] David, F.N., and Johnson, N.L. (1956). Statistical Treatment of censored data. *Biometrika 41.* 228-40.