A SIMPLE UNIFYING FORMULA FOR TAYLOR'S THEOREM AND CAUCHY'S MEAN VALUE THEOREM

Jochen Einbeck
University of Munich, Department of Statistics,
Akademiestr. 1, D-80799 München, Germany
e-mail: einbeck@stat.uni-muenchen.de

Abstract: We introduce a formula which generalizes Taylor's theorem from powers of linear terms $z - x$ to functional terms $\phi(z) - \phi(x)$, leading to a formula which reduces in a special case to Cauchy's generalized mean value theorem. In other words, regarding Cauchy's mean value theorem as an extension of the simple mean value theorem, we provide the analogous extension of Taylor's theorem. The filling of this gap is easy and requires only mathematics on an undergraduate level, so that the mentioned analogy might be a useful tool for illustration at schools and universities.

AMS Subj. Classification: 41A58

KeyWords: Taylor's formula, Generalized mean value theorem, Widder's theorem, Nonparametric smoothing.

One of the most widely applied mathematical theorems is that of Taylor [1], which allows to approximate a function m in a neighborhood of x by a linear combination of polynomials:

$$m(z) = \sum_{j=0}^{p} \frac{m^{(j)}(x)}{j!}(z-x)^j + r_{p+1}(z).$$

Taylor did not specify the remainder term, the first representation of which is due to Lagrange [2]. We provide the following simple extension of Taylor's theorem in Lagrange respresentation:

Theorem 1. *Let $m, \phi : [v,w] \longrightarrow \mathbb{R}$ be p times continuously differentiable and $p+1$ times differentiable in (v,w), ϕ invertible in $[v,w]$ and let $x \in [v,w]$. Then for each $z \in [v,w]$ with $z \neq x$ there exists a point $\zeta \in (x,z)$ resp. (z,x) such that*

$$m(z) = \sum_{j=0}^{p} \frac{\psi_{(j)}(x)}{j!}(\phi(z)-\phi(x))^j + s_{p+1}(z) \tag{1}$$

with

$$\psi_{(j+1)}(\cdot) = \frac{\psi'_{(j)}(\cdot)}{\phi'(\cdot)}, \psi_{(0)}(\cdot) = m(\cdot), \tag{2}$$

holds, whereby the remainder term can be written as

$$s_{p+1}(z) = \frac{\psi_{(p+1)}(\zeta)}{(p+1)!}(\phi(z)-\phi(x))^{p+1}.$$

Proof. *Define*

$$\begin{aligned} g(y) &= m(z) - m(y) - \psi_{(1)}(y)(\phi(z)-\phi(y)) - \ldots - \\ &- \frac{\psi_{(p)}(y)}{p!}(\phi(z)-\phi(y))^p - \frac{M}{(p+1)!}(\phi(z)-\phi(y))^{p+1}, \end{aligned} \tag{3}$$

where $M \in \mathbb{R}$ is chosen so that $g(x) = 0$ is fulfilled. Using $g(x) = g(z) = 0$ Rolle's theorem yields that there exists a $\zeta \in (x,z)$ resp. (z,x) with $g'(\zeta) = 0$. Since

$$g'(y) = -\frac{\psi'_{(p)}(y)}{p!}(\phi(z)-\phi(y))^p - \frac{M}{p!}(\phi(z)-\phi(y))^p(-\phi'(y))$$

it follows that $0 = -\psi'_{(p)}(\zeta) + M\phi'(\zeta)$ and thus $M = \psi_{(p+1)}(\zeta)$. Setting $y = x$ in (3) yields Theorem 1. □

Remark 1. Let ϕ, m as above and $g(\cdot) = (m \circ \phi^{-1})(\cdot)$. Applying Taylor's theorem on g at point $\phi(x)$ and comparing the result with (1) yields

$$\psi_{(k)}(\cdot) = (m \circ \phi^{-1})^{(k)}(\phi(\cdot)). \tag{4}$$

As a consequence, the property of uniqueness of coefficients carries over from Taylor's theorem to Theorem 1.

Remark 2. There exist a variety of formulations for the remainder term in Taylor's theorem, one of them being the extact integral representation (Cauchy [3]). For an overview of remainder forms we refer to Blumenthal [4]. Using observation (4), or directly via partial integration, the remainder term in (1) in integral form can be written as

$$s_{p+1}(z) = \frac{1}{p!} \int_x^z (\phi(z) - \phi(t))^p \psi_{(p+1)}(t) \phi'(t)\, dt\,, \tag{5}$$

however requiring m and ϕ to be $p+1$ times *continuously* differentiable.

Remark 3. We observe that Theorem 1 reduces for $p = 0$ to the well-known generalized mean value theorem (Cauchy [3]), however with the small constraint that one of the two involved functions has to be invertible in the considered interval. Using the remainder form (5), it reduces for $p = 0$ to the fundamental theorem of integral calculus.

Regarding Table 1, we notice that Theorem 1 fits quite naturally in the existing framework. Considering its simplicity and the age of the theorems surrounding it, it seems to be surprising that it, to our knowledge, never has been mentioned before. Research about generalizations of Taylor's theorem started late, what gave Widder [5] reason to state that *"in view of the great importance of Taylor's series in analysis, it may be regarded as extremely surprising that so few attempts at generalization have been made"*. Meanwhile exist a large variety of extensions, whereby only that of Widder [5], cited in the following remark, is related to our one:

Remark 4. Recall that for basis functions $\phi_0(z), \ldots, \phi_p(z) \in C^p$ (i.e. having p continuous derivatives) Wronski's determinant

$$\begin{tikzcd}[column sep=huge, row sep=huge]
\boxed{\text{Theorem 1}} \arrow[r, "\phi = id"] \arrow[d, "p=0"] & \boxed{\begin{array}{c}\text{Taylor's Theorem}\\ \text{(Taylor 1715,}\\ \text{Lagrange 1797)}\end{array}} \arrow[d, "p=0"] \\
\boxed{\begin{array}{c}\text{Generalized Mean}\\ \text{Value Theorem}\\ \text{(Cauchy 1821)}\end{array}} \arrow[r, "\phi = id"] & \boxed{\begin{array}{c}\text{Mean Value Theorem}\\ \text{(Lagrange 1797)}\end{array}}
\end{tikzcd}$$

Table 1: Relation between Taylor's and other theorems.

is defined by $W_p(z) = \begin{vmatrix} \phi_0(z) & \phi_1(z) & \cdots & \phi_p(z) \\ \phi_0'(z) & \phi_1'(z) & \cdots & \phi_p'(z) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0^{(p)}(z) & \phi_1^{(p)}(z) & \cdots & \phi_p^{(p)}(z) \end{vmatrix}$. Widder's generalization of Taylor's formula is as follows.

If the functions $m(z), \phi_0(z), \phi_1(z), \ldots, \phi_p(z)$ are of class C^p in a neighborhood of $z = x$, and if $W_p(x) \neq 0$, then there exists a unique function of approximation

$$a_p(z) = \sum_{j=0}^{p} c_j \phi_j(z) =$$

$$= \quad -\left(\frac{1}{W_p(x)}\right) \begin{vmatrix} 0 & \phi_0(z) & \phi_1(z) & \cdots & \phi_p(z) \\ m(x) & \phi_0(x) & \phi_1(x) & \cdots & \phi_p(x) \\ m'(x) & \phi_0'(x) & \phi_1'(x) & \cdots & \phi_p'(x) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m^{(p)}(x) & \phi_0^{(p)}(x) & \phi_1^{(p)}(x) & \ldots & \phi_p^{(p)}(x) \end{vmatrix} \qquad (6)$$

of order p for $z = x$.

Thereby $a_p(z) = \sum_{j=0}^{p} c_j \phi_j(z)$ is a function of approximation for function $m(z)$ of order p at point $z = x$ if the functions $\phi_j(z)$ are of class C^p in a neighborhood of $z = x$, and if $a_p(z)$ has contact

$$\boxed{\begin{matrix}\text{Widder's}\\ \text{Theorem}\end{matrix}} \xrightarrow[=\phi^j(z)]{\phi_j(z)} \boxed{\text{Theorem 1}} \xrightarrow[=z]{\phi(z)} \boxed{\begin{matrix}\text{Taylor's}\\ \text{Theorem}\end{matrix}}$$

Table 2: From Widder's to Taylor's formula.

of order p at least with $m(z)$ at $z = x$, i.e. if

$$a_p^{(k)}(x) = g^{(k)}(x), \quad k = 0, \ldots, p.$$

In the further course of his paper, Widder also provides expressions for the remainder terms. Obviously, Widder's generalization covers Theorem 1 widely. However, the fact that it theoretically *covers* Theorem 1, does not necessarily mean that the latter one might be directly *derived* from it. Indeed, when setting $\phi_j(z) = \phi^j(z), j = 0, \ldots, p$, with $\phi \in C^p$ invertible, one sees easily the equivalence of Widder's expansion and Theorem 1 in the case $p = 1$. However, for higher degrees p, one notices rapidly that the complexity of the Wronskian as well as the determinant in (6) makes Widder's formula hard to apply. The problem is thereby that with each iterative differentation (from one line of the Wronskian to the deeper one) the number of terms rises exponentially, leading to intractable expressions already after some few steps. We encourage the reader to verify this. Meanwhile, Theorem 1 might serve as a kind of bridge between Taylor's and Widder's formula, as illustrated in Table 2.

Remark 5. We remark finally that there exists an application of Theorem 1 in nonparametric statistics. Assume that $(X_i, Y_i), i = 1, \ldots, n$ is a sample of independent and identically distributed random variables, sampled from a population (X, Y). X and Y are modelled by $Y = m(X) + \epsilon$, where $m(\cdot)$ is a smooth function and ϵ some random noise with $E(\epsilon) = 0$ which is independent from X. Then, one might calculate a local approximation

$\hat{m}(x)$ of $m(x)$ by minimizing

$$\sum_{i=1}^{n}\left\{Y_i - \sum_{j=0}^{p}\gamma_j(x)(\phi(X_i)-\phi(x))^j\right\}^2 K\left(\frac{X_i - x}{h}\right)$$

in terms of $\gamma_j(x), j = 0,\ldots,p$, where $K(\cdot)$ is a kernel function, usually a symmetric probability density function, and ϕ serves as a basis function which might be selected data-adaptively to reduce the bias of the fit. Calculation of the asymptotic bias of this approximation requires the presented theorem. See details, in the multivariate setting, in Einbeck [6].

Acknowledgements

The author is grateful to Daniel Rost and Hubert Kalf, Department of Mathematics, University of Munich, for fruitful discussions.

References

[1] Taylor, B. (1715). *Methodus incrementum directa et inversa.* London.

[2] Lagrange, J. L. (1797). *Théorie des fonctions analytique.* Paris: Imprimerie de la Republique.

[3] Cauchy, A.-L. (1821). *Cours d'analyse de l'Ècole polytechnique.* Paris: L'Imprimerie Royale.

[4] Blumenthal, L. M. (1926). Concerning the remainder term in Taylor's formula. *American Mathematican Monthly* **33**, 424-426.

[5] Widder, D. V. (1928). A generalization of Taylor's series. *Transactions of the American Mathematical Society* **31**, 43-52.

[6] Einbeck, J. (2003). Multivariate local fitting with general basis functions. *Computational Statistics* **18**, 185-203.