Understanding parameter differences between analyses employing nested data subsets

Steven Gratton\(^1,2\)\(^\star\) and Anthony Challinor\(^1,2,3\)\(^\dagger\)

\(^1\)Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
\(^2\)Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge CB3 0HA, UK
\(^3\)DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

18 November 2019

ABSTRACT
We provide an analytical argument for understanding the likely nature of parameter shifts between those coming from an analysis of a dataset and from a subset of that dataset, assuming differences are down to noise and any intrinsic variance alone. This gives us a measure against which we can interpret changes seen in parameters and make judgements about the coherency of the data and the suitability of a model in describing those data.

Key words: methods: analytical – methods: statistical

1 INTRODUCTION
One would typically expect the posterior distributions of the parameters of a model to change as the datasets used to constrain them are changed. A part of this must be attributable to “scatter”, i.e., noise and any intrinsic variance assumed by the model. However, it could also be an indication of a problem, either in the data (e.g., a systematic error in one of the datasets or an unaccounted-for relative calibration between two datasets) or in the model (i.e., the model is incomplete and unable to well-describe all of the data).

In this paper we provide an analytical argument for understanding the likely nature of parameter differences in the ideal, scatter-only, case. This gives us a diagnostic measure which we can use to interpret changes seen between parameters inferred from using subsets of the data. Thus we can make judgements about the internal coherency of the data and the appropriateness of a model for describing those data.

The method used here can also be used to derive simply other useful results for understanding differences between models and data. For example, Wilks (1938) derived the distribution that one should expect for the change in \(\chi^2\) when allowing additional model parameters to vary if in fact they are unnecessary. In a cosmological context, this is very useful in knowing how strongly to take \(\chi^2\) improvements as potential detections of systematics or new physics. In an appendix we show how the approach presented here can be used to rederive quickly this result.

Some of the techniques described in detail in this paper were used in Planck Collaboration et al. (2016a, 2017, 2020a). For similarly-motivated work see Raveri & Hu (2019) and Lemos et al. (2020). The former paper discusses a number of “concordance/discordance estimators” within and between cosmological data sets, including some based on parameter shifts as here but assuming the data is Gaussian-distributed. The latter paper develops a “suspiciousness statistic” for “correlated” data sets but does not specifically investigate nested data subsets.

2 DERIVATION OF MAIN RESULT
Imagine one has a collection of data, denoted by the vector \(X\). One has a parametric model in mind to describe these data, depending on a collection of parameters denoted by a vector \(P\). Let us write the probability for a realisation of the data to be within \(dX\) of \(X\) under the model as

\[ p(X) dX = e^{-S} dX, \]

where the “action” \(S\) is a function of \(X\) and the parameters \(P\), and \(dX\) is the appropriate multi-dimensional measure on the data. Thus we can make judgements about the internal coherency of the data and the appropriateness of a model for describing those data.

The method used here can also be used to derive simply other useful results for understanding differences between models and data. For example, Wilks (1938) derived the distribution that one should expect for the change in \(\chi^2\) when allowing additional model parameters to vary if in fact they are unnecessary. In a cosmological context, this is very useful in knowing how strongly to take \(\chi^2\) improvements as potential detections of systematics or new physics. In an appendix we show how the approach presented here can be used to rederive quickly this result.

Some of the techniques described in detail in this paper were used in Planck Collaboration et al. (2016a, 2017, 2020a). For similarly-motivated work see Raveri & Hu (2019) and Lemos et al. (2020). The former paper discusses a number of “concordance/discordance estimators” within and between cosmological data sets, including some based on parameter shifts as here but assuming the data is Gaussian-distributed. The latter paper develops a “suspiciousness statistic” for “correlated” data sets but does not specifically investigate nested data subsets.

\[ S = S_0 + S^T \delta P + \frac{1}{2} \delta P^T \hat{S} \delta P + \cdots \]

where the “action” \(S\) is a function of \(X\) and the parameters \(P\), and \(dX\) is the appropriate multi-dimensional measure on the data. (This form is motivated by analogy with the path integral approach to Euclidean quantum field theory, where \(S\) plays the part of the Euclidean action.)

Let us now assume that the model can indeed describe the data and that the true parameter values are \(P_0\). Let us expand \(S\) to second order in the parameters about \(P_0\):

\[ S = S_0 + S^T \delta P + \frac{1}{2} \delta P^T \hat{S} \delta P + \cdots \]

\[ \approx S_0 + S^T \delta P + \frac{1}{2} \delta P^T \hat{S} \delta P + \frac{1}{2} \left( \delta P + \hat{S}^{-1} \delta P + \frac{1}{2} \left( \delta P + \hat{S}^{-1} \delta P + \cdots \right) \right) + \text{const.} \]

© 2019 The Authors
Here $S'$ denotes the vector of derivatives of $S$ with respect to the parameters, superscript $T$ denotes transpose, and $S''$ denotes the matrix of second derivatives. In the second line we have made the (typically-good) approximation of replacing $S''$ with its average $\overline{S''}$, where here and onwards an overline denotes an ensemble average of the indicated object with respect to data realizations from the assumed model with parameters $P_0$. Equation (2) motivates the “maximum-likelihood” estimator for the parameters, and we now briefly recap some of its properties in preparation for what will follow. The fluctuation in parameters around $P_0$ for any given realisation of the data is given by
\[
\delta P = -\overline{S''}^{-1} S',
\]
where $S'$ is evaluated for the realization in question. This can be seen to be unbiased as follows. From Eq. (3) we have
\[
\begin{align*}
\delta P &= -\overline{S''}^{-1} S' \\
&= - \overline{\int dX S_i e^{-S}} \\
&= - \left( \overline{\int dX e^{-S}} \right)_i \\
&= -1_i = 0,
\end{align*}
\]
where $S_i = \delta S_i/\delta P_i$. In the last line we have used the fact that the probability distribution for $X$ is normalized to unity. We see that $\delta P$ for the correct model averages to zero over realizations of that model, and so $\overline{\delta P} = 0$.

To obtain the covariance of the parameters we average the outer product of Eq. (3) over the ensemble. By judicious integrations by parts we can relate the average of this product of first derivatives to the average of a second derivative:
\[
\begin{align*}
\overline{S_i S_j} &= \overline{\int dX S_i S_j e^{-S}} \\
&= - \overline{\int dX S_i (e^{-S})_j} \\
&= - \left( \overline{\int dX e^{-S}} \right)_j + \overline{\int dX S_{ij} e^{-S}} \\
&= 1_{ij} + \overline{S_{ij}},
\end{align*}
\]
Hence
\[
\overline{\delta P \delta P^T} = \overline{S''}^{-1} \overline{S''} S''^{-1} = \overline{S''}^{-1}.
\]
and corresponding to Eq. (6) we have
\[
\begin{align*}
\delta P_1 \delta P_1^T &= \overline{S''}^{-1}. \\
\end{align*}
\]

Now we are in a position to investigate the distribution of parameter differences, $\delta P_1 - \delta P$, over the full ensemble. With each term averaging to zero, the parameter differences also average to zero. For the covariance, we have
\[
\begin{align*}
(\delta P_1 - \delta P_1) (\delta P_1 - \delta P)^T &= \overline{S''}^{-1} S_1' S_1^T S''^{-1} \\
&= \overline{S''}^{-1} - \overline{S''}^{-1} S_1' S_1^T S''^{-1} \\
&= \overline{S''}^{-1} - \overline{S''}^{-1} S_1' S_1^T S''^{-1} + \overline{S''}^{-1} S_1' S_1^T S''^{-1} - \overline{S''}^{-1},
\end{align*}
\]
and we see the average of the “mixed” quantity $S_1' S'' S''^{-1} S_1$. As with Eq. (5), using integrations by parts we can manipulate this into the form of the average of a second derivative:
\[
\begin{align*}
\overline{S_{1,i} S_{1,j}} &= \overline{\int dX S_{1,i} S_{1,j} e^{-S}} \\
&= - \overline{\int dX S_{1,i} (e^{-S})_{1,j}} \\
&= - \left( \overline{\int dX e^{-S}} \right)_{1,j} + \overline{\int dX S_{1,ij} e^{-S}} \\
&= - \left( \overline{\int dX S_{1,i} e^{-S}} \right)_{1,j} + \overline{\int dX S_{1,ij} e^{-S}} \\
&= 1_{ij} + \overline{S_{1,ij}},
\end{align*}
\]
where we have used the fact that by its definition $S_1$ must be independent of $X_2$. Hence we find
\[
\overline{\delta P_1 \delta P_1^T} = \overline{S''}^{-1},
\]
the same as for $\overline{\delta P \delta P^T}$ itself. Substituting into Eq. (10) gives us the elegant final result
\[
(\delta P_1 - \delta P) (\delta P_1 - \delta P)^T = \overline{S''}^{-1} - \overline{S''}^{-1},
\]
i.e., the covariance of the parameter differences between the partial and full analyses is the difference of their respective covariances.

3 Interpreting Differences in Multiple Parameters

For multiple parameters, one can form a “$\chi^2$” for the differences in the parameters between the two analyses. This allows one to treat all parameters fairly, neither focussing on one outlier in particular nor neglecting degeneracies when judging how unlikely multiple shifts are. If we write the parameter shifts as $\Delta = \delta P_1 - \delta P$, in the Gaussian approxima-
tion the probability \( p(\Delta) d\Delta \) of being within \( d\Delta \) of \( \Delta \) is

\[
p(\Delta) d\Delta = \frac{d\Delta}{2\pi \left( S_0 - S_0^{-1} \right)^{1/2}} \exp \left[ -\frac{1}{2} \lambda \left( \frac{S_0 - S_0^{-1}}{S_0 - S_0^{-1}} \right)^{-1} \right].
\]

(14)

If the effective \( \chi^2 \) (i.e., minus twice the argument of the exponential) was to be large compared to the number of fitted model parameters then one might begin to worry about the fidelity of some aspect of the data taken as a whole or indeed about the applicability of the model to all of the data. The former would point to systematic effects, the latter to new physics.

4 COSMIC MICROWAVE BACKGROUND EXAMPLE ANALYSIS

Let us apply the formalism to an example drawn from cosmic microwave background (CMB) analysis in cosmology. Relevant introduction, motivation and definitions of the model parameters may be found in the Planck “cosmological parameters” series of papers (Planck Collaboration et al. 2014, 2016b, 2020b).

Here one takes maps of the microwave sky and compares them to predictions from a parameterized model, which we take here to be a standard six-parameter \( \Lambda \text{CDM} \) cosmology. The model does not predict the actual pattern of fluctuations of the CMB, only their statistical properties. The primordial fluctuations in the early Universe are assumed to be Gaussian-distributed, giving an intrinsic “cosmic variance” to observations. The primordial 3D power spectrum is taken to change smoothly with scale and so is fully described by an amplitude and “tilt” parameter. In addition to temperature, “T”, or intensity fluctuations, the linear polarization of the light also varies across the sky. In the simplest \( \Lambda \text{CDM} \) models that we shall consider here, this polarization can be described with the help of an additional scalar field, “E”. From the T and E fields, there are three angular power multipoles and forming their auto- and cross-power spectra. CAMB model of the Planck 2015 analysis (Planck Collaboration et al. 2016b) using CAHB (Lewis et al. 2000). Next, we generate a realization of the TT, TE and EE spectra from this model by first drawing Gaussian realizations of the T and E multipoles and forming their auto- and cross-power spectra. Finally, we perform Markov-chain Monte-Carlo (MCMC) analyses on both the full set of spectra and the partial set and find best-fit models in both cases using CosmoMC (Lewis & Bridle 2002). Parameter means, standard deviations and best-fits are listed in Table 1 and 2D marginalised posterior distributions are illustrated in Fig. 1.

CosmoMC provides estimates of the covariance matrices for the posterior distributions derived from the MCMC chains. We use these on the right-hand side in Eq. (13) to obtain an estimate of the covariance of the parameter differences that we should expect to see. We are now able to compare predictions from this covariance to what is found in this test. To start with, from the diagonal of the matrix we derive standard deviations for the shifts in parameters between the full and partial analyses. We list in Table 2 these standard deviations, along with the measured shifts in both the best-fits and the means in terms of these standard deviations; the shifts are of the right magnitude to agree with the prediction.

Considering more than one parameter at a time, one might perform a singular-value-decomposition of Eq. (13) to identify the “most likely” shifts one should expect to see. One can also use the entire covariance of Eq. (13), as discussed in Sec. 3, to compute an overall “goodness-of-fit" for the shifts in all of the parameters. We obtain a \( \chi^2 \) of 9.84 for the shifts in the means (11.3 for shifts in best-fits) for our six degrees of freedom, a value greater than which would be expected about 13% (8%) of the time under the distribution in Eq. (14).

In computing the standard deviations of parameter shifts shown in Table 2 and associated \( \chi^2 \) values, we have used the covariances estimated from the MCMC chains, in a similar manner to what one would need to do in a real problem. In our simulation here, however, we know what the underlying model is and so can calculate \( S_{\mu\nu} \) analytically in terms of derivatives of the fiducial spectra with respect to the model parameters (these derivatives being evaluated numerically). Using these matrices, the standard deviations of the shifts change very little, but the \( \chi^2 \) of the shift in the means changes to 8.93 and the \( \chi^2 \) of the shift in the best-fits changes to 9.93.

We go on to generate 100 further realizations of the power spectra, and find their best-fitting parameter values under the full and partial treatments. We compare the shifts...
Figure 1. Two-dimensional marginalised posterior distributions for a full (blue; smaller contours) and a partial (red; larger contours) analysis of a simulated CMB dataset.

Table 1. Cosmological parameter constraints from a full and partial analysis of a simulated CMB data set (the posterior distributions are shown in Fig. 1). These results serve as our starting point for comparing parameter shifts to expectation.

| Parameter | Full analysis | Partial analysis |
|-----------|---------------|-----------------|
|           | best fit | mean | std. dev. | best fit | mean | std. dev. |
| $\Omega_b h^2$ | 0.02221 | 0.02222 | 0.00012 | 0.02219 | 0.02224 | 0.00032 |
| $\Omega_c h^2$ | 0.12013 | 0.12011 | 0.00083 | 0.11844 | 0.11804 | 0.00267 |
| $100 \theta_{MC}$ | 1.04070 | 1.04071 | 0.00025 | 1.03886 | 1.03905 | 0.00118 |
| $\tau$ | 0.07255 | 0.07265 | 0.00183 | 0.07337 | 0.07365 | 0.00254 |
| $\ln(10^{10} A_s)$ | 3.07673 | 3.07689 | 0.00399 | 3.07529 | 3.07498 | 0.00522 |
| $n_s$ | 0.96030 | 0.96040 | 0.00335 | 0.96384 | 0.96522 | 0.00937 |

between the analyses to predictions from Eq. (13) in Fig. 2. We plot a histogram of the $\chi^2$s of the shifts, using the analytic covariances, in Fig. 3 (performing a classic binned Pearson’s chi-squared test comparing the histogram to the curve, we find a value of 9.5 for 10 effective degrees of freedom, indicating good agreement).

In order to obtain the good agreement with expectation shown, it was necessary to take into account the effect of “numerical noise” in the minimization procedure. The result returned by the CosmoMC minimizer routine is expected to be good to within a 0.05-sigma tolerance of the true minimum for each parameter. Assuming such errors are independent...
Understanding parameter differences

Table 2. Comparison of the shifts in parameters between the two analyses reported in Table 1 to the predictions of Eq. (13). The first column shows the prediction of the expected standard deviations of the shifts in best-fitting parameters. The second and third columns show the shifts found, in units of these standard deviations, in the best-fits and posterior means respectively. For the simulation considered, the shifts are seen to be broadly in line with the predictions.

| Parameter | std. dev. of shift | Δ(best fit) in shift std. dev.'s | Δ(mean) in shift std. dev.'s |
|-----------|--------------------|---------------------------------|-----------------------------|
| $\Omega_b h^2$ | 0.00030 | -0.10 | -0.06 |
| $\Omega_c h^2$ | 0.00254 | -0.67 | -0.82 |
| $100\theta_{MC}$ | 0.00115 | -1.59 | -1.53 |
| $\tau$ | 0.00175 | +0.47 | +0.57 |
| $\ln(10^{10} A_s)$ | 0.00337 | -0.43 | -0.56 |
| $n_s$ | 0.00875 | +0.40 | +0.55 |

Figure 3. Normalized histogram showing the effective $\chi^2$ from Eq. (14) evaluated for the difference of best-fit parameters between the partial and full analyses of 100 simulated CMB datasets, using analytic covariances computed around the fiducial model (including terms accounting for tolerances in the minimization procedure), compared to a $\chi^2$ distribution for six degrees of freedom. (A classic binned Pearson’s chi-squared test comparing the histogram to the curve confirms the visual impression that the two are in reasonable agreement with each other.)

5 COMMENTS AND EXTENSIONS

Our result Eq. (13) should have a wide applicability, enabling one to compare analyses with differing combinations of datasets, e.g., to illuminate tensions between late-time measurements of the expansion rate of the Universe (see,
e.g., Riess et al. 2018) and inferences from the ΛCDM framework with Planck (see, e.g., the discussion in Planck Collaboration et al. 2020b), as well as comparing subsets of data taken with the same experiment.

We can understand generic features of parameter shifts from the form of Eq. 13, by considering how partial cancellations between the two covariance matrices may or may not occur as the constraining power between the two analyses changes. For example, if large parameter degeneracies that exist using the partial data set are broken using the full data set, then one would expect parameter differences to lie along those parameter degeneracies also. This is because the large degenerate eigenmode in the partial-analysis covariance matrix would not be significantly reduced by any corresponding mode in the full-analysis covariance matrix.

Note that in order to obtain the simple result of Eq. (13), we had to look at differences between one data combination and another “nested” within it. No such simple result exists in general for parameter differences between, say, two non-nested datasets X₁ and X₂. If the data sets happen to be independent, then our method does yield

\[\langle \delta P_1 - \delta P_2 \rangle = \frac{1}{\sigma^2} \mathbf{C}_1^{-1} \mathbf{C}_2 \mathbf{C}_1^{-1} \mathbf{C}_2^{-1} \mathbf{C}_1^{-1} \mathbf{C}_2^{-1},\]

with uncertainties now adding in quadrature as expected.

By applying Eq. (13) to multiple nested subsets, one can build up a “grand” covariance matrix for expectations of the parameter differences between all the analyses. Let parameters \(\delta P_i\) come from an analysis involving a subset of the data that yielded parameters \(P_i\) itself from an analysis involving a subset of the data that yielded parameters \(P\). By considering appropriate conditional distributions associated with this covariance matrix, certain properties of the parameters obtained may be understood. For example, using Eqs. (6) and (12) we can compute the joint covariance of the differences \(P_1 - P\) and \(P_2 - P\) to compare their behaviour to expectation.

Our formalism can be used to investigate how parameter estimates should “track” towards the truth as more data is added. Imagine one happened to know what the true parameters to our data, and then fit an extended model with an additional \(n_b\) parameters. In the case where the first model is the correct one, i.e., the additional parameters are not required, by how much should we expect the extended model to improve the goodness of fit by chance? For a cosmological example, one might allow the amplitude of the 3D power spectrum of primordial gravitational waves to vary in the extended model when they are actually negligible. Wilks’ Theorem (Wilks 1938) addresses expectations for the improvement in the log-likelihood for the wider analysis over the more specific one. The theorem shows the improvement to be

\[\chi^2 - \chi^2_{\text{min}} \sim \nu_{\text{df}},\]

where \(\nu_{\text{df}}\) is the number of degrees of freedom equal to the number \(n_b\) of additional parameters.

Finally, one can generalise the argument of Sec. 2 to include Gaussian priors on the parameters. However, the result is not as easily expressed in terms of the covariance matrices involved as it was in Eq. (13).

6 CONCLUSIONS

In this note we have presented arguments aimed to aid the understanding of relations between inferences using full and partial amounts of data and, in the appendix, between inferences using standard and extended models.

Our work provides some insight into the sorts of posterior variations one might expect when comparing related Bayesian parameter estimations. A “rule of thumb” for a parameter that is well-constrained by the data (so that any effect of priors may be neglected) is as follows: when more data is added, a significant change in the width of the posterior distribution will be accompanied by a difference in the most likely parameter value that can be a significant fraction of the larger width. If the posteriors have similar widths then there should be little shift in the peak position.

ACKNOWLEDGEMENTS

We thank George Efstathiou, Antony Lewis and the Planck Parameters and Likelihood teams for helpful comments and discussions over the development of this work. We thank the anonymous referee of this paper for useful comments on the original manuscript.

Our MCMC analyses were performed using the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (http://www.csd3.cam.ac.uk/), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council, and DiRAC funding from the Science and Technology Facilities Council (https://dirac.ac.uk).

SG and AC acknowledge support from the UK Science and Technology Facilities Council (grant numbers ST/N000927/1 and ST/S000623/1).

APPENDIX A: WILKS’ THEOREM

Imagine we fit a model with \(n_p\) parameters to our data, and then fit an extended model with an additional \(n_b\) parameters. In the case where the first model is the correct one, i.e., the additional parameters are not required, by how much should we expect the extended model to improve the goodness of fit by chance? For a cosmological example, one might allow the amplitude of the 3D power spectrum of primordial gravitational waves to vary in the extended model when they are actually negligible. Wilks’ Theorem (Wilks 1938) addresses expectations for the improvement in the log-likelihood for the wider analysis over the more specific one. The theorem shows the improvement to be \(\chi^2\)-distributed with a number of degrees of freedom equal to the number \(n_b\) of additional parameters.

Equations (2) and (3) can be used to rederive this result. Here we need to keep track of the constant to evaluate the
Understanding parameter differences

Let us introduce an \( n \times n_a \) projection matrix \( \mathbf{M} \) with \( n = n_a + n_b \), corresponding to the total number of parameters varied in the wider analysis.

\[
\mathbf{M} = \begin{pmatrix} \mathbf{I}_{n_a} & \mathbf{0}_{n \times n_b} \end{pmatrix},
\]

where \( \mathbf{I}_{n_a} \) denotes the \( n_a \times n_a \) identity matrix and \( \mathbf{0}_{n \times n_b} \) denotes an \( n \times n_b \) matrix of zeros. This allows us to express \( S_{ab} - S_0 \) for the usual model, \( S_{ab} \) being its best-fit action, as

\[
S_{ab} - S_0 = -\frac{1}{2} (\delta P_{a}^T \mathbf{S}_e \delta P_{a})
= -\frac{1}{2} S_a^T \mathbf{S}_e^{-1} S_a
= -\frac{1}{2} S_a^T \mathbf{M} \mathbf{S}_e^{-1} \mathbf{M}^T S_a,
\]

i.e., in terms of \( S' \) rather than just \( S_a' \). Subtracting this from the same quantity evaluated for the extended model yields

\[
S_{bf} - S_{ab} = -\frac{1}{2} S'^T (\mathbf{S}_e^{-1} - \mathbf{M} \mathbf{S}_e^{-1} \mathbf{M}^T) S'
= -\frac{1}{2} \delta P^T (\mathbf{S}_e^{-1} - \mathbf{M} \mathbf{S}_e^{-1} \mathbf{M}^T \mathbf{S}_e^{-1}) \delta P,
\]

an expression now involving only \( \delta P \), the shift from the underlying model to the extended best-fit model, and not also \( \delta P_a \). Using Eq. (A4) the bracketed term becomes

\[
\begin{pmatrix} 0 \\ \mathbf{0}_{n_b \times n} \end{pmatrix} \left( \mathbf{S}_e^{-1} - \mathbf{M} \mathbf{S}_e^{-1} \mathbf{M}^T \mathbf{S}_e^{-1} \right) \begin{pmatrix} 0 \\ \mathbf{0}_{n \times n_b} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix} \left( \mathbf{S}_e^{-1} \right) \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix} \left( \mathbf{S}_e^{-1} \right) \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix},
\]

with \( \mathbf{S}_e^{-1} \) being the lower-right block of the inverse of \( \mathbf{S}_e^{-1} \). Hence

\[
S_{bf} - S_{ab} = -\frac{1}{2} \delta P_{b}^T (\mathbf{S}_e^{-1} \mathbf{S}_e^{-1}) ^{-1} \delta P_{b},
\]

which we note only depends on the additional parameters \( \delta P_b \). So to understand how the action is distributed, we need to know how the \( \delta P_b \) are distributed.

From Eq. (6), we see that the \( \delta P \) have covariance \( \mathbf{S}_e^{-1} \), and hence the \( \delta P_b \) have covariance \( \mathbf{S}_e^{-1} \).

**REFERENCES**

Efstathiou G., Gratton S., 2019, arXiv e-prints, p. arXiv:1910.00483

Lemos P., Köhlinger F., Handley W., Joachimi B., Whiteway L., Lahav O., 2020, MNRAS, 496, 4647

Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511

Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473

Planck Collaboration et al., 2014, A&A, 571, A16

Planck Collaboration et al., 2016a, A&A, 594, A11

Planck Collaboration et al., 2016b, A&A, 594, A13

Planck Collaboration et al., 2017, A&A, 607, A95

Planck Collaboration et al., 2020a, A&A, 641, A5

Planck Collaboration et al., 2020b, A&A, 641, A6

Raveri M., Hu W., 2019, Phys. Rev. D, 99, 043506

Riess A. G., et al., 2018, ApJ, 855, 136

Wilks S. S., 1938, Ann. Math. Statist., 9, 60

This paper has been typeset from a \textsc{TeX}/\textsc{l\textbackslash{}a\textbackslash{}tex} file prepared by the author.