Journal of Neuroscience Methods 145 (2005) 133–144

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

Applying resampling methods to neurophysiological data

Eran Stark [a, *], Moshe Abeles [a, b, c]

[a] *Department of Physiology, Hadassah Medical School, The Hebrew University of Jerusalem, Jerusalem 91120, Israel*
[b] *The Interdisciplinary Center for Neural Computation, The Hebrew University of Jerusalem, Jerusalem 91904, Israel*
[c] *Gonda Brain Research Center, Bar-Ilan University, Ramat-Gan 52900, Israel*

Received 8 September 2004; received in revised form 12 December 2004; accepted 14 December 2004

Abstract

Standard statistical techniques do not always provide answers to complex physiological questions because often there are no parametric or non-parametric distributions on which significance can be estimated. Resampling methods provide a battery of tests that can be used in such circumstances. In the past few years these methods have been explored theoretically and are now employed frequently. In this paper we describe a unified framework for the use of such methods in the context of neurophysiological data analysis. We construct specific tests for placing confidence limits on estimates of mutual information and on parameters of circular data, and we present procedures for testing hypotheses on circular and on partitioned data. These tests are explained in detail and illustrated with real data from experiments with behaving monkeys.

Keywords: Circular statistics; Confidence limits; Information theory; Monkey recordings; Non-parametric statistics; Prehension; Premotor cortex; Spatial organization

1. Introduction

Resampling methods have conventionally been used as a means of tackling problems which are too complicated to be solved analytically ('Monte Carlo' techniques, Metropolis and Ulam, 1949). Over the past 30 years, the theoretical foundations for these methods have been expanded and substantiated (Efron, 1979). Among others, the jackknife, bootstrap, and permutation tests have been used extensively in various fields and applications such as, for example, in astronomy (Barrow et al., 1984), medical imaging (Holmes et al., 1996), and econometrics (Mills and Zandvakili, 1997). These methods are particularly suitable for hypothesis testing and for determining the accuracy of non-parametric and/or complex statistics for which closed-form formulae, if they exist, depend on extensive assumptions.

Neurophysiological data analysis, a field that utilizes methods from multiple areas of research (information theory, signal processing, statistics, and so on), often encounters situations in which a physiological question: (1) cannot be answered in a parametric framework for which closed-form formulae for accuracy exist; (2) may need to be examined by standard, existing tools, but the results exhibit a bias that influences inference; and/or (3) can only be assessed by specially tailored algorithms or procedures that, in turn, require objective validation.

For these reasons, as well as due to the availability of fast computers, resampling methods have gained increased popularity in neuroscience (Georgopoulos et al., 1988; Optican and Richmond, 1987, are early examples). These methods are flexible, easy to implement, applicable in non-parametric settings, and require a minimal set of assumptions. In this paper we hope to contribute to this area by providing a comprehensive framework for the application of resampling methods to data obtained from neurophysiological experiments, with an emphasis on circular data. We develop specific tests in this

* Corresponding author. Tel.: +972 2 675 8381; fax: +972 2 643 9736.
E-mail address: eranstark@md.huji.ac.il (E. Stark).

0165-0270/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2004.12.005

framework, assess their properties, and provide examples of their application.

The organization of the paper is as follows. In Section 2 we describe the statistical framework, review essential topics from resampling methods, and point out a few subtleties of application. Section 3 describes the experimental procedures used to obtain neural data. In Section 4, the core of this paper, we describe several problems of a statistical nature that do not have a closed-form solution and present procedures based on resampling to solve them. We conclude with a brief discussion.

2. A review of resampling methods

We begin with a brief formulation of the statistical framework that constitutes the basis of our approach. Readers already familiar with the formalities of resampling methods (for instance, Davidson and Hinkley, 1997; Efron and Tibshirani, 1993) may wish to skim this section.

2.1. Data-generation model

Denote some data sample by $\vec{X}$, a vector of n observations (each observation may be a scalar or some complex data structure). The sample is assumed to be generated by an unknown stationary probability distribution function (pdf) F, through independent and identically distributed (iid) sampling, $F \xrightarrow{\text{iid}} \vec{X} = \{x_1, x_2, \ldots, x_n\}$. The empirical distribution $\hat{F}$ is defined as the output of a discrete function that assigns equal ($1/n$) probabilities to all observations, $\text{prob}(A) = \#\{x_i \in A\}/n$. The syntax $\#\{x_i \in A\}$ means "the number of observations that belong to A". It can be shown (Efron and Tibshirani, 1993) that $\hat{F}$ is a sufficient statistic for F, the unknown distribution.

We are interested in estimating a parameter θ of the distribution F (such as its mean or variance) obtained by a known transformation T of F. We denote this by $\theta = T(F)$. The 'plug-in estimate' (also called the 'empirical maximum likelihood' estimate) of this parameter is defined by $\hat{\theta} = T(\hat{F})$. Denote the equivalent statistic by $\hat{\theta} = t(\vec{X})$.

For example, if $T(F)$ is the first moment $\mu = E_F[X]$, then $t(\vec{X})$ is the sample's mean, $\mu_{\hat{F}} = E_{\hat{F}}[X] = \bar{\vec{X}}$. If $\vec{X}$ is the only source of information on F, then $\hat{\theta} = T(\hat{F})$ is a consistent estimator of θ, that is, $\lim_{n\to\infty} \hat{\theta} = \theta$ (by the weak law of large numbers).

The *scatter* (or precision) of an estimator may be measured by its standard error; for example, for the first moment, plug-in application of the closed-form formula for the standard error of the mean gives $\widehat{se}(\vec{X}) = \sigma_{\hat{F}}/\sqrt{n} = (1/n)\sqrt{\sum_{i=1}^{n} (x_i - \bar{\vec{X}})^2}$. Disappointingly, closed-form formulae for standard errors (or for a related measure – confidence intervals) exist only for the mean and a few other well-known statistics.

A measure of estimation accuracy is the *bias*, defined as the difference between the expected value of an estimator and the estimated parameter itself, $\text{bias}_F \equiv E_F[t(\vec{X})] - T(F)$. In the described setup, where F is sampled once ($\vec{X}$), we cannot estimate a bias directly from the data. Moreover, closed-form formulae for estimation of (or correction for) bias exist for only a few statistics (for example, the sample variance).

2.2. Resampling approaches to parameter estimation

Several of the problems referred to in Section 1 can be circumvented in the general case of parameter estimation by iid sampling from the data-generating probability mechanism $F \to \vec{X}$. Resampling (with repetitions allowed, i.e. with replacement) the empirical distribution $\hat{F}$ results in a surrogate dataset, $\hat{F} \to \vec{X}^{\text{bs}} = \{x_1^{\text{bs}}, x_2^{\text{bs}}, \ldots, x_n^{\text{bs}}\}$. Note that x_i^{bs} is not necessarily the i'th element of $\vec{X}$ and that x_i^{bs} may be the same as x_j^{bs}, $i \neq j$, because resampling is carried out with replacement. Calculating the statistic $\hat{\theta} = t(\vec{X})$ of this dataset yields another estimate of the same parameter, $\hat{\theta}^{\text{bs}} = t(\vec{X}^{\text{bs}})$. By repeating this process many times a *distribution* of the statistic of interest, also called the *bootstrapped distribution*, is generated.

The bootstrapped distribution of a statistic may be used to measure the scatter and/or bias of an estimator. From this distribution, estimates of the standard error of a statistic, bias, and (non-parametric) confidence limits may be obtained. The procedure is quite simple:

1. Given a sample $\vec{X}$ of size n, compute a statistic of interest, $\hat{\theta} = t(\vec{X})$.
2. Resample $\vec{X}$ (with replacement) N^{bs} times, arriving at $\vec{X}_b^{\text{bs}} = \{x_1^{\text{bs}}, x_2^{\text{bs}}, \ldots, x_n^{\text{bs}}\}$, $b = 1 \ldots N^{\text{bs}}$, surrogate datasets.
3. For each surrogate set, compute the statistic $\hat{\theta}_b^{\text{bs}} = t(\vec{X}_b^{\text{bs}})$, $b = 1 \ldots N^{\text{bs}}$.
4. Compute the mean of the bootstrapped distribution $\bar{\hat{\theta}}^{\text{bs}} = \sum \hat{\theta}_b^{\text{bs}}/N^{\text{bs}}$.
5. Evaluate the accuracy of estimation by computing, from *the same* bootstrapped distribution:
   - 5.1 the standard error of $\hat{\theta}$, $\widehat{se} = \sqrt{\sum_{b=1}^{N^{\text{bs}}} (\hat{\theta}_b^{\text{bs}} - \bar{\hat{\theta}}^{\text{bs}})^2 / (N^{\text{bs}} - 1)}$ (the S.D. of $\hat{\theta}^{\text{bs}}$);
   - 5.2 the bias, $\bar{\hat{\theta}}^{\text{bs}} - \hat{\theta}$;
   - 5.3 and the confidence limits (see below).

From the tails of the bootstrapped distribution, confidence limits for $\hat{\theta}$ may be established. Denote by $1 - \alpha$ the desired confidence interval for the estimate of the statistic and by $\hat{\theta}_{\text{sorted}}^{\text{bs}}$ the sorted (in ascending order) values of the statistics. The lower confidence limit is the $N^{\text{bs}}\alpha/2$ value of $\hat{\theta}_{\text{sorted}}^{\text{bs}}$ and the upper confidence limit is the $N^{\text{bs}}\alpha/2$ value from the end: denote these limits by $\{\hat{\theta}_{\text{lo}}, \hat{\theta}_{\text{hi}}\}$. Confidence limits computed in this manner are robust to translations, transforma-

tions, and, contrary to some parametric confidence limits, obey restrictions on the values that the parameter can take (a range-preserving property). However, they are sensitive to bias, which must be assessed separately. Methods for 'automatic' consideration of bias during estimation of confidence intervals have been developed (Efron and Tibshirani, 1993); we will not elaborate upon them here.

The above procedure is cluttered by two potential sources of noise: *sampling noise* that is due to a small sample size n or to non-iid sampling, and *resampling noise*, resulting from a small number of resampling repetitions, N^{bs}. Given a fixed sample size, the first source imposes a lower bound on the estimation error, and cannot be reduced by repeated resampling. As for the second source of noise, for a sample size n, there are $2n-1$ choose n distinct surrogate combinations of sample *indices*, or datasets $\left(N^{\mathrm{distinct}} = \binom{2n-1}{n}\right)$; increasing the number of unique resampling repetitions beyond N^{distinct} does not improve estimation accuracy. Fig. 1 shows, on a logarithmic scale, N^{distinct} as a function of sample size n; for instance, when $n=5$, there are only 126 distinct samples. If the data permit, it is advisable to use tens-to-hundreds of repetitions in order to estimate standard errors or bias; thousands of repetitions are required for confidence limits (Efron and Tibshirani, 1993). Thus, when sample size is small, the reliability of confidence limits (estimating the tails of the distribution) obtained by bootstrapping is limited, and a larger sample is then required to estimate trustworthy confidence limits.

An estimate of bias can also be obtained by an alternative method, the *jackknife* ('leave-one-out'). Briefly, jackknifed samples are obtained by generating reduced datasets, $\vec{X}_j^{\mathrm{jk}} = \{x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n\}, \quad j = 1 \ldots n$. For each reduced set, the statistic of interest is computed by $\hat{\theta}_j^{\mathrm{jk}} = t(\vec{X}_j^{\mathrm{jk}})$; bias is evaluated by $(n-1)((1/n)\sum \hat{\theta}_j^{\mathrm{jk}} - \hat{\theta})$. This estimate can be shown to be a quadratic approximation to the bootstrap estimate of bias when N^{bs} is very large. Compared with a practical number of bootstrap repetitions, the jackknife bias may be a more *accurate* estimate. However, jackknife estimates are less *precise* (have higher variance) and behave erratically for some non-smooth statistics (for example, quantiles which the bootstrap handles easily), so in what follows we will not use them.

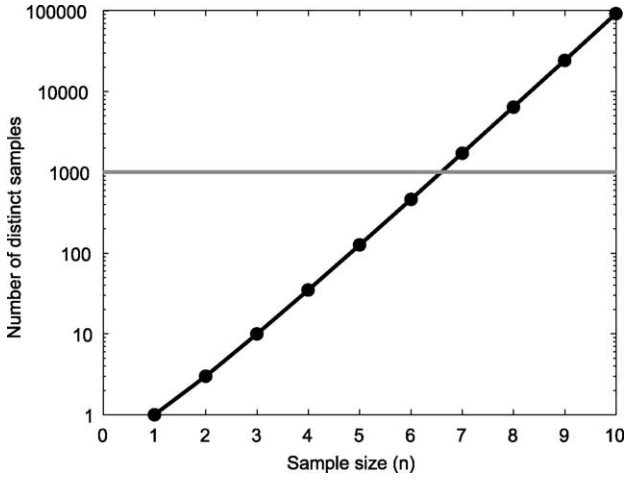

Fig. 1. Number of distinct bootstrapping samples as a function of sample size. Ordinate is in a logarithmic (base 10) scale. Horizontal line is at 1000 repetitions, the minimal number advisable for estimating confidence limits of a statistic.

2.3. Hypothesis testing by resampling methods

Resampling techniques have applications other than parameter estimation. The statistic of interest does not have to be an estimator of some parameter; it can be any transformation on (function of) the sample $\vec{X}$, $t(\vec{X})$. The bootstrapping procedure (steps 1–3) described above may then be applied to the test statistic, arriving at a bootstrapped distribution of this statistic, $t(\vec{X}_b^{\mathrm{bs}})$. In order to test some null hypothesis (for example, H_0: $t(\vec{X}) = \Theta$), the distribution, resampled *under the null hypothesis*, is compared to the statistic (or value) of interest; in our example, the probability to obtain $t(\vec{X})$ equal to or larger than Θ under H_0 is given by the tail of the bootstrapped distribution, $\mathrm{prob}(H_0) = \#\{t(\vec{X}_b^{\mathrm{bs}}) \geq \Theta\}/N^{\mathrm{bs}}$. The process is entirely equivalent to the computation of a $(1-2$ prob) confidence interval for the statistic of interest; its merit lies in its usefulness in more complex scenarios.

Consider an extension of the data-generation model where, *independently* of the n_1 observations of $F_1 \rightarrow \vec{X}_1$, there are n_2 observations of $F_2 \rightarrow \vec{X}_2$ (a two-sample scenario). If a test statistic is defined as a function of both samples, $t(\vec{X}_1, \vec{X}_2)$, composite hypotheses about the (unknown) distributions F_1 and F_2 may be tested. For example, H_0: $F_1 = F_2$ may be tested by defining an appropriate statistic (for instance, the difference between the means) and comparing the observed value to the bootstrapped distribution. The bootstrapping is done *under* H_0: in this example, each surrogate dataset will contain $n_1 + n_2$ elements of the concatenated set, $\{\vec{X}_1, \vec{X}_2\}$. Complicated situations can be handled using the same framework (see Section 4).

In the special case where the null hypothesis is equality of distributions, a more powerful method is available: a *permutation test*. Under H_0: $F_1 = F_2$ it does not matter which distribution is the source of an observation $x_i \in \{\vec{X}_1, \vec{X}_2\}$. A statistic is computed, grouping of individual observations to either sample is randomly permuted (sampled *without* replacement) N^{perm} times, and the observed statistic is compared to the distribution of permuted values as was previously explained. Note that since sampling is without replacement, there are only $\binom{n_1+n_2}{n_1}$ distinct permuted datasets. Rather than estimating the pdf generating the data as the bootstrap test of hypotheses does, a permutation test exploits the special symmetry ($F_1 = F_2$) that H_0 imposes on the distributions, allowing for the mixing of $\vec{X}_1$ and $\vec{X}_2$. The latter method is

preferable in the specific case of testing equality of distributions; other, more complex, null hypotheses may be tested using the bootstrap method.

Regardless of the resampling technique (bootstrapping or permutations), an appropriate statistic should be defined in order for a test of hypotheses to be reliable. The number of repetitions used should be large – on the order of thousands – as in bootstrap estimation of confidence limits.

3. Data source

In the following sections we describe tests constructed within the above framework and we apply them to neurophysiological data obtained from monkey experiments. In this section we describe the procedure used to acquire the data. Recordings were from the dorsal and the ventral premotor (PMd, PMv) cortices of two *Macaca fascicularis* monkeys performing a prehension task. Targets were 3D objects requiring different types of grasp (precision grip, power grip, finger opposition, and so on) and were presented in six directions, equally spaced relative to a central touch pad. In order to initiate a trial, a monkey had to press and hold the touch pad: a target object was then presented for a short period of time, followed by a delay (1000–1500 ms) during which objects were not visible. A 'Go' signal prompted the monkey to reach for, grasp, and hold the target object.

Neural data were acquired by inserting during each recording session up to 16 independently movable (EPS 1.31, Alpha-Omega Engineering, Nazareth, Israel) glass-plated tungsten micro-electrodes (impedance 0.2–2 MΩ at 1 kHz) through the dura mater. Electrodes were arranged in two independently adjustable guide tubes, such that up to eight electrodes were inserted into each area (PMd, PMv). The signals from these electrodes were amplified (10K), bandpass filtered (5–6000 Hz), sampled at 25 kHz, and stored on disk (Alpha-Map 5.4, Alpha-Omega Eng.). Behavioral events were sampled (6 kHz) and digitized. The signal from each electrode was subject to manual offline spike-sorting (Alpha-Sort 4.0, Alpha-Omega Eng.) resulting in a set of well-isolated units. All animal handling procedures were in accordance with the *NIH Guide for the Care and Use of Laboratory Animals* (1996), complied with Israeli law, and were approved by the Ethics Committee of the Hebrew University.

4. Results

Within the general framework described in Section 2, one may devise procedures for estimating parameters and for testing specific statistical hypotheses without parametric assumptions. In what follows we describe several novel procedures for handling cases in which the analytical formulae are either based on assumptions that we cannot realistically make or do not exist altogether. For each problem, the resampling solution is compared to a parametric or to a non-parametric method if one is available, and the information gained by the new application is emphasized.

4.1. Mutual information estimates from finite data

Mutual information (I) is a measure of the reduction in uncertainty of a random variable that occurs when the value of another random variable is known. This measure can be used for objective quantification of the dependence between neural activity and external events (stimuli) or behaviors, and (following Cover and Thomas, 1991) is defined by

$$I(R;S) \equiv H(S) - H(S/R) = H(R) - H(R/S) = \sum_{r\in R, s\in S} p(r,s)\log\frac{p(r,s)}{p(r)p(s)}, \quad (1)$$

where H is entropy; R the neural activity (quantized in some manner); S the stimulus set (or array of behaviors); and both are discrete random variables. This definition indicates that I is a symmetric measure. It has further advantage as it can indicate statistical independence between R and S because then $I(R;S)$ is, by definition, zero.

In order to measure mutual information from experimental data, an estimate of the joint pdf $\hat{F} = \hat{p}(r,s)$ is required. From this estimate, the estimates of the marginal probabilities $\hat{p}(r)$ and $\hat{p}(s)$ can be obtained by summation. Therefore, the estimated quantity $\hat{I}$ depends on the (discrete) sampling of the stimulus–response pairs (r,s) from the unknown distribution F, and is subject to error. Bias, in particular, hinders the estimation process, since the *empirical* joint pdf, $\hat{F}$, estimated from a finite dataset, is bound to differ from the product of the marginal distributions even when the two variables are statistically independent. Several procedures have been suggested to overcome these difficulties, such as kernel methods (Optican and Richmond, 1987), shuffling of stimulus–response pairs (Chee-Orts and Optican, 1993), asymptotic series expansions (Panzeri and Treves, 1996), and asymptotic theory (Strong et al., 1998), yielding improved estimates of $\hat{I}$. Regardless of the exact procedure, we would like to measure the scatter of the estimator by setting confidence limits on the estimate; to the best of our knowledge, there are no closed-form formulae for doing so.

4.1.1. Confidence limits of mutual information by bootstrapping

Formally, we have a set of n pairs of observations $F \to \vec{X} = \{(r_1,s_1),\ldots,(r_n,s_n)\}$, and we want to estimate the parameter I from the empirical joint pdf $\hat{F}$ and arrive at the estimator $\hat{I}$. Both scatter and bias may be estimated from the data using a bootstrap procedure. In the case of a single measurement, we resample values (with replacement); in this case, we resample the stimulus–response pairs (r,s) (with replacement) and recompute the statistic $\hat{I}_b^{\mathrm{bs}}$ (for each resampling repetition $b = 1\ldots N^{\mathrm{bs}}$). The bias of $\hat{I}$ is estimated by $\mathrm{bias}^{\mathrm{bs}} = (\sum \hat{I}_b^{\mathrm{bs}}/N^{\mathrm{bs}}) - \hat{I}$, and the confidence limits of the

(assumed to be unbiased) $\hat{I}$ by $\{\hat{I}_{\mathrm{lo}}, \hat{I}_{\mathrm{hi}}\}$ (Section 2.2). Alternatively, the bias may be estimated by one of the abovementioned procedures (but see below). If the bias is not negligible, it can be corrected for by subtraction, arriving at 'de-biased' estimates,

$$\hat{I}^{\text{de-biased}} = \hat{I} - \text{bias}, \tag{2.1}$$

$$\{\hat{I}_{\mathrm{lo}}^{\text{de-biased}}, \hat{I}_{\mathrm{hi}}^{\text{de-biased}}\} = \{\hat{I}_{\mathrm{lo}}, \hat{I}_{\mathrm{hi}}\} - 2\,\text{bias}. \tag{2.2}$$

The factor 2 in Eq. (2.2) is necessary because the bootstrapped distribution is centered around a biased estimate; equivalently, Eq. (2.1) could be written as $I^{\text{de-biased}} = (\sum \hat{I}_b^{\text{bs}}/N^{\text{bs}}) - 2\,\text{bias}$.

4.1.2. Application: an example

Fig. 2 illustrates an application of this procedure to neural data obtained from prehension experiments. As an example, we estimate the mutual information between spike counts of a PMv unit and the planned direction of movement. In Fig. 2A, raster plots of the activity of a PMv unit are illustrated; the unit was recorded during {93, 93, 91, 89, 93, 94} trials in different behavioral conditions (six movement directions). During the 400 ms just prior to the 'Go' signal, this unit exhibited modulation of spiking with respect to behavior; spikes were counted during this period for each trial, and the joint pdf $\hat{F} = \hat{p}(r, s)$ was estimated by direct quantization of the response space (Fig. 2B). The raw mutual information, $\hat{I}$, was estimated from $\hat{F}$ using Eq. (1) and corrected by plugging into Eq. (2.1) $\text{bias}^{\text{analytic}} = (\sum_s \tilde{R}_s - \tilde{R} - S + 1)/(2n \log 2)$ (Panzeri and Treves, 1996). In the latter formula, summation

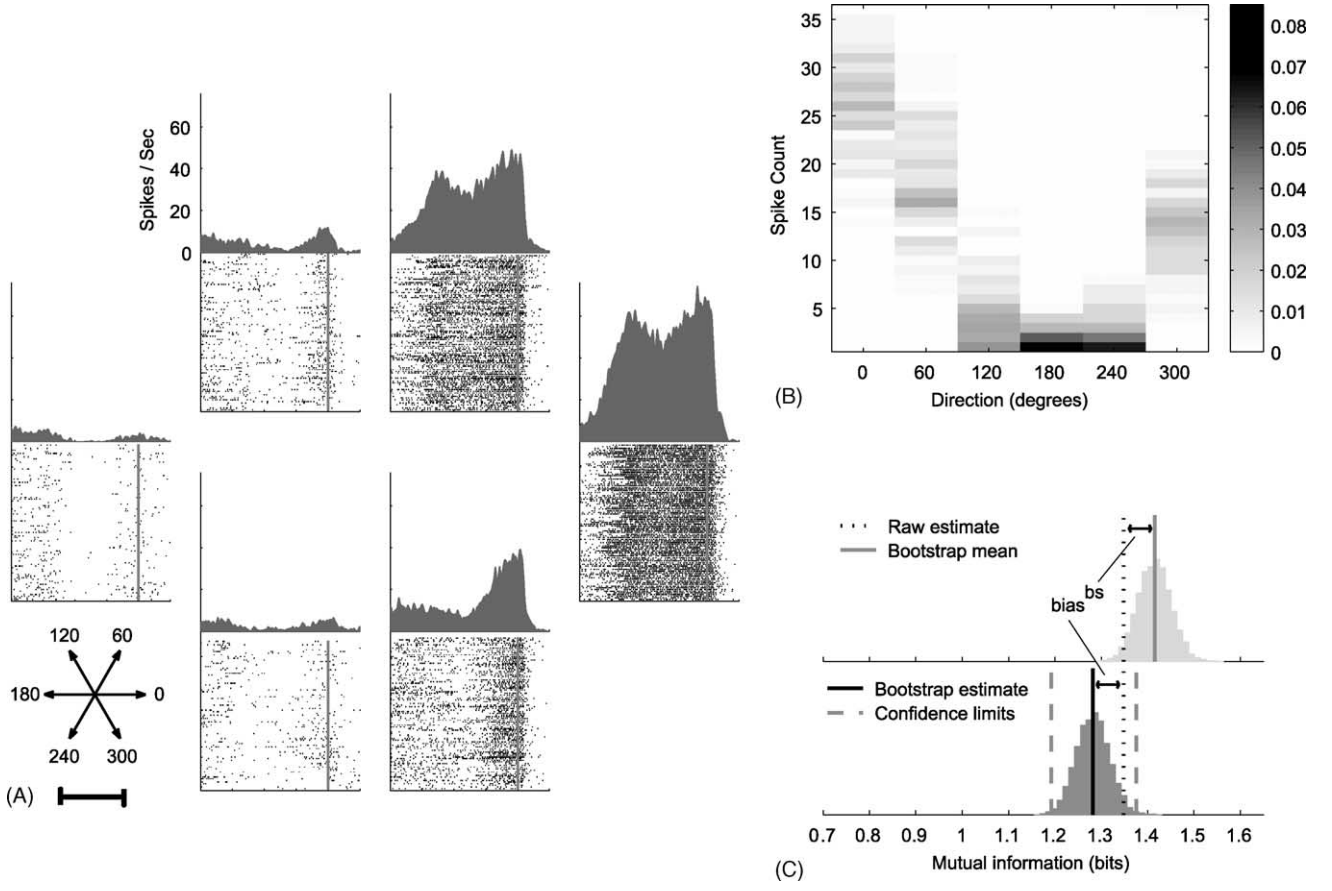

Fig. 2. Confidence limits for mutual information of a PMv unit prior to movement. (A) A PMv unit's spiking activity during trials requiring movement in six directions shown in different panels (see Section 4.1.2). The bottom part of each panel shows a standard raster diagram, aligned on the 'Go' signal (vertical line); the upper part of each panel shows the corresponding peri-event time histogram (obtained by smoothing rasters with a Gaussian window with $\sigma = 15$ ms). Scale bar is 1 s. (B) Empirical joint probability distribution $\hat{F} = \hat{p}(r, s)$ (Section 4.1) of spike counts in the 400 ms just prior to the 'Go' signal (shown in the six panels of (A) as vertical lines) and movement directions. Gray levels give the joint probabilities of getting a certain amount of spikes (ordinate) at a given movement direction (abscissa, corresponding to the directions shown by the compass at the lower left-hand side of (A)). Measurements were discrete (every 60° and 1 count). For display purposes the height (gray level) is spread so as to fill the intervals between the discrete values. (C) Confidence limits of bias-corrected mutual information. Vertical dotted black line shows the raw estimate $\hat{I}$. Histogram in light gray (top) shows the bootstrapped distribution ($\hat{I}^{\text{bs}}$) obtained by bootstrapping the (r, s) pairs; the continuous dark gray line is its mean. The bootstrap estimate of bias is the distance between $\hat{I}$ and the distribution mean, Eq. (2.1); this estimate closely matches the analytic correction described in the text (Section 4.1.2). The bootstrapped distribution was shifted by twice the bias to the left for presentation and is shown by the dark gray histogram (bottom); the dark gray dashed lines indicate the translated 99% confidence limits, obtained by Eq. (2.2).

is over non-zero response bins $\tilde{R}_s$ in each behavior (s) out of the S behaviors and $\tilde{R}$ is the number of non-zero response bins in $\hat{p}(r)$ (this correction is intended to compensate for limited sample size n). The raw value was $\hat{I} = 1.35$ bits and the corrected value was $\hat{I}^{\text{de-biased}} = 1.29$ bits.

Next, the procedure explained in Section 4.1.1 was carried out. The distribution of the bootstrapped statistic $\hat{I}^{\text{bs}}$ is shown at the top of Fig. 2C in light gray: bootstrapped values were, on average, higher than the raw estimate $\hat{I}$, suggesting that some bias indeed shifted the estimate upwards. The bias estimated by bootstrapping was 0.067 bits—very close to the first-order analytically estimated bias of 0.06. After computing and correcting the 99% confidence limits as in Eq. (2.2), final estimates were obtained (dark gray dashed lines in the bottom of Fig. 2C), suggesting that the unit transmitted between 1.19 and 1.38 bits of information (out of the possible $\log_2 6 \cong 2.58$) about the upcoming movement direction.

4.2. Estimating parameters for circular distributions

When measuring responses to moving gratings or neural activity related to reaching movements, the measurements are made in relation to the direction of motion, which is a *circular* independent variable. In the preceding section, the possible circular nature of the data was ignored for the sake of generality; this omission can be justified on the grounds that the stimuli were treated as discrete. However, due to the common use of circular data and their physiological meaning, it may be preferable to explore certain circular properties of the data. Before describing specific tests designed to test hypotheses on such data, we introduce several basic terms from circular statistics.

Let us extend the data-generation model (Section 2.1) to a circular one-sample scenario. Each of the n observations in $\vec{X}$ is now a vector, described in polar coordinates by $\vec{x}_i = (f_i, \phi_i)$ of (amplitude, direction). Note that in this model sampling is not limited to the circumference of the unit circle, so both amplitude and direction are sampled from a continuum. The sample's resultant (i.e. the vectorial sum of all $\vec{x}_i$'s), normalized to a length between 0 and 1, may be represented in Cartesian coordinates by

$$\vec{R} = \{R_x, R_y\}, \quad R_x = \frac{\sum f_i \cos\phi_i}{\sum f_i}, \quad R_y = \frac{\sum f_i \sin\phi_i}{\sum f_i} \tag{3}$$

(all summations are over $i = 1 \ldots n$). The direction (argument) of $\vec{R}$, $\angle\vec{R} = \tan^{-1}(R_y/R_x)$, resolved to the proper quadrant, is the first moment of the circular sample, or preferred direction (PD); its amplitude, $|\vec{R}| = \sqrt{R_x^2 + R_y^2}$, gives the complement of the second moment, the circular variance ($S = 1 - |\vec{R}|$). Note that if sampling is limited to the unit circle itself, $\vec{R}$ simplifies to the expression $\{R_x = (1/n)\sum \cos\theta_i,\ R_y = (1/n)\sum \sin\theta_i\}$ (Mardia, 1972). The normalized amplitude of the resultant $\left|\vec{R}\right|$ is invariant to rotations and can be regarded as a measure of concentration (and $S = 1 - |\vec{R}|$ as a measure of width): for a homogenously distributed sample, $|\vec{R}|$ is close to zero, while for samples clustered around their mean it approaches the value of one.

4.2.1. Non-parametric confidence limits of circular parameters

We want to estimate the PD of a circular sample as well as its $1-\alpha$ confidence interval. The suitable statistic is the direction of the sample's resultant, $\angle\vec{R}$, as defined above. Closed-form formulae for its confidence limits exist only under certain assumptions: Mardia (1972) gives such formulae for the specific case where the observations are from the unit circle and originate in a circular counterpart of the Gaussian distribution; that is, the von Mises distribution (which is unimodal, symmetric, and completely defined by two parameters). If such assumptions cannot be made, the bootstrapping method described below can be used for estimating confidence limits.

Assume that $\vec{X}$ is obtained by sampling the circle in M equally spaced discrete directions $\vec{\phi} = \{\phi_1, \phi_2, \ldots, \phi_M\}$. Otherwise, discretize the observation directions. In each direction we have a set of n_m observations, $m = 1 \ldots M$, totaling $n = \sum_{m=1}^{M} n_m$ such that $\vec{X} = \{(f_1^1, \phi_1), \ldots, (f_1^{n_1}, \phi_1), \ldots, (f_M^1, \phi_M), \ldots, (f_M^{n_M}, \phi_M)\}$. In order to estimate $\vec{R}$, compute the mean amplitude of the observations in each direction, $\bar{f}_m = (1/n_m)\sum_{i=1}^{n_m} f_m^i$, and use Eq. (3). Note that this process is not necessarily equivalent to direct application of Eq. (3) to $\vec{X}$: only when all n_m's are identical do the two estimates converge. Next, rearrange the sample $\vec{X}$ such that each set of M observations from different directions forms a new observation:

$$\hat{\vec{x}}_j = \{(f_1^j, \phi_1), \ldots, (f_m^j, \phi_m), \ldots, (f_M^j, \phi_M)\},$$
$$m = 1 \ldots M, \quad j = 1 \ldots \max(n_m). \tag{4}$$

If n_m's have different sizes due to unbalanced finite sampling, the appropriate elements of $\hat{\vec{x}}_j$ should be left empty. Neither $\vec{R}$ nor the statistic $\angle\vec{R}$ change following this rearrangement. We are now in a position to directly apply a bootstrapping procedure (Section 2.2) to the set of rearranged observations and arrive at confidence limits for the PD. A similar procedure, using the test statistic $|\vec{R}|$, will yield confidence limits for the resultant's amplitude, or sample width. The same bootstrapped distribution of $\vec{R}$ can be used for the computation of confidence limits of both $\angle\vec{R}$ and $|\vec{R}|$.

4.2.2. Application: confidence limits for PD and concentration of directional tuning curves

The activity of a single PMv unit was recorded during $n = 245$ trials; we want to estimate the PD and the concentration of the directional tuning curve (TC) of this unit as well as confidence limits for these parameters. Trials included presentation of target objects in six fixed directions, $\vec{\phi} = \{(0, 1/6, \ldots, 5/6)2\pi\}$. During movement in each trial,

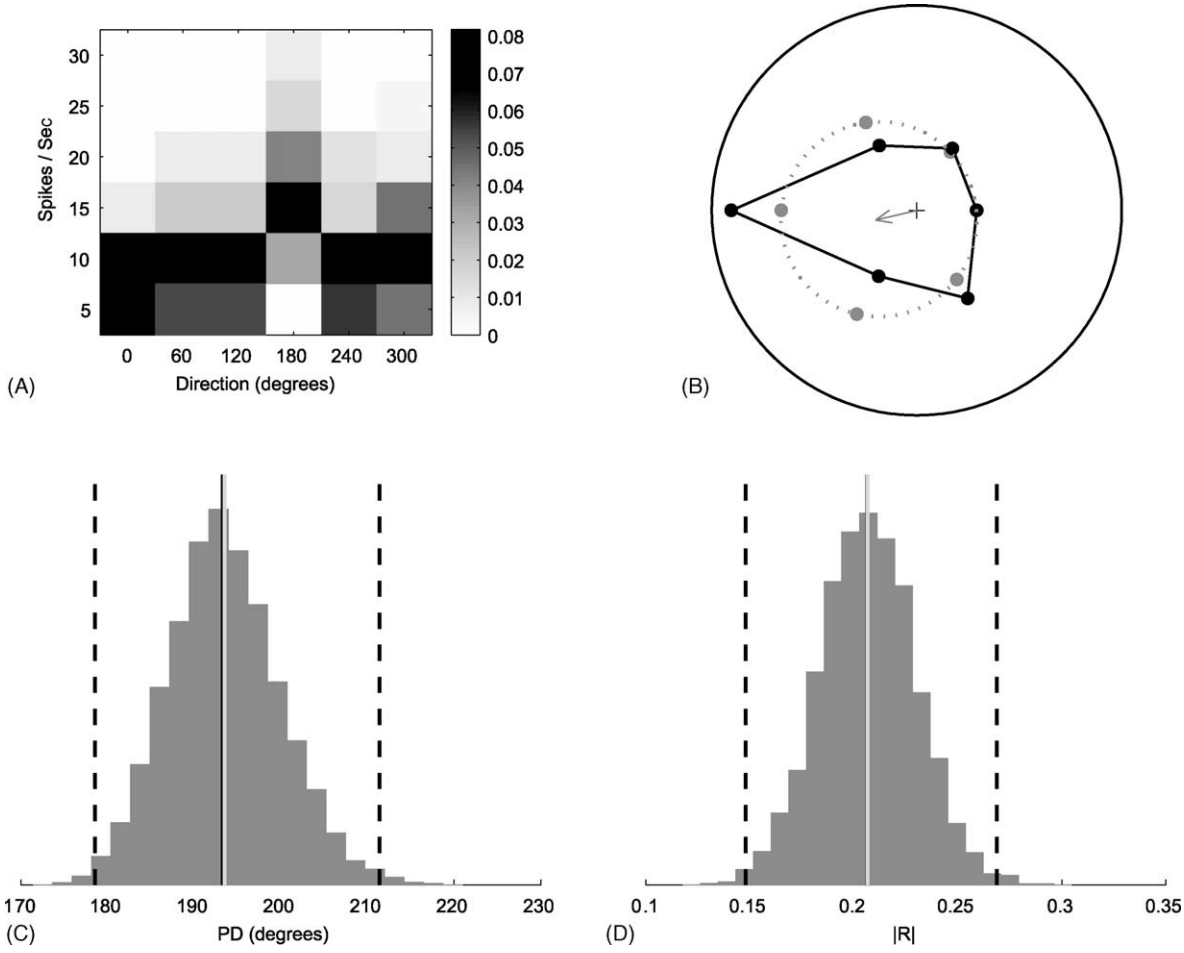

Fig. 3. Confidence limits of PD and concentration of a PMv unit's directional tuning curve during movement. (A) Empirical joint distribution of a PMv unit's firing rate during movement and movement direction. Measurements on the ordinate were quantized into bins of five spikes per second for display purposes. Other conventions are the same as in Fig. 2B. (B) Directional tuning curve of the same unit shown in A, obtained by averaging spike counts in each direction. Radial scale corresponds to 15 spikes/sec; the resultant's length is shown as a fraction of this radius. In a dotted line, the maximum likelihood estimate of a von Mises distribution is shown. Clearly, the experimental data diverge from the von Mises distribution. (C) The tuning curve's PD ($\angle\vec{R}$, vertical continuous black line) and its 99% confidence limits (dashed lines), estimated by $N^{bs} = 5000$ resampling repetitions. The mean of the distribution is also shown (vertical continuous gray line); the estimate is essentially unbiased. (D) Estimation of tuning concentration $|\vec{R}|$, along with confidence limits. Conventions are the same as in (C).

the firing rate during a window of 400 ms around movement initiation was noted as f_i, so in this example we have $x_i = (f_i, \phi_m)$, $i = 1...n$, $n_m = \{42, 40, 41, 40, 40, 42\}$, $\phi_m \in \vec{\phi}$. Fig. 3A illustrates the joint distribution of f and ϕ. From this distribution, the TC of the unit was point-estimated by averaging the various f_i's in each of the M directions, $(\bar{f}_m, \phi_m)$ (Fig. 3B); the PD and concentration of this TC were computed, as were their non-parametric 99% confidence limits (Fig. 3C and D). As the figure suggests, this unit had sharp directional tuning, with a PD of 193° (with 99% confidence limits at (179°, 211°)) and a resultant length of 0.21 (range (0.15, 0.27)). As can be appreciated from Fig. 3B, the tuning curve did not fit a von Mises distribution well (likelihood-ratio test, $p < 0.01$). Nevertheless, we estimated parametric confidence limits for the PD and the resultant's length based on a von Mises distribution (Mardia, 1972), arriving at wider (and biased) estimates for both parameters (PD: (161°, 225°); $|\vec{R}|$: (0.06, 0.27)). Obviously, lack of fit between the data and the von Mises pdf resulted in erroneous estimates of the confidence limits both in terms of range and bias: application of our procedure to data that did fit a von Mises distribution yielded confidence limits similar to those obtained directly under the assumption that the distribution is von Mises.

4.3. Testing circular distributions for equality

Assume that neural responses to certain circular stimuli under two conditions (say attentive and non-attentive) are measured, and that we want to know whether the responses are affected by these conditions. Specifically, it is of interest to know whether *any* aspect of the responses (that is, their distributions) differs across the conditions. Statistically speaking, this is a *circular two-sample* scenario, where there are n_1 observations of $F_1 \rightarrow \vec{X}_1$ and, independently, n_2 observations of $F_2 \rightarrow \vec{X}_2$. Assume that the two circular distributions are sampled in the same M discrete direc-

tions. We want to test H_0: $F_1 = F_2$. For linear, continuous data, a standard two-sample non-parametric test, such as the Kolmogorov–Smirnov goodness-of-fit test, can be used. For two samples evaluated *on* the unit circle itself, a Uniform-Scores (Wheeler's) test (Mardia, 1972) is adequate. However, our case does not fall within these categories since the vectors can be anywhere and not just on the circumference of the unit circle; we will therefore use a circular permutation test.

4.3.1. Circular permutation test

Define the test statistic $t(\vec{X}_1, \vec{X}_2)$ as the absolute difference between the resultants of the two samples

$$\Delta R = ||\vec{R}_1 - \vec{R}_2||. \quad (5)$$

Under H_0, this statistic should be close to zero; if the two samples have equal concentrations but the PDs are displaced by π radians, ΔR could approach a value of two. Note that ΔR does not depend solely on the first moments of the distributions: if the PDs are similar but one sample is dispersed and the other is concentrated, ΔR should be close to one. Applying a permutation test in this setup is now straightforward: we randomly assign values from each discrete direction to either sample (taking advantage of the identical sampling frequency M of the two distributions) N^{perm} times, and recompute the statistic, arriving at a permuted distribution, ΔR^{perm}. The probability of obtaining a statistic as large as or larger than ΔR under H_0 is then approximated by $\text{prob}(H_0) = \#\{\Delta R^{\text{perm}} \geq \Delta R\}/N^{\text{perm}}$.

Since sampling is discrete (or when discretization is plausible), a conventional test of goodness-of-fit (such as the χ^2 or the likelihood-ratio test) can be applied. However, the permutation test described above is more sensitive to revealing deviation from equality. Moreover, a slight variation in the test statistic enables us to test the same H_0 with different alternatives. Assume that we wish to test the null hypothesis H_0: $F_1 = F_2$ against the alternative H_1: $\angle\vec{R}_1 \neq \angle\vec{R}_2$; that is, that the samples have different PDs regardless of their widths. An appropriate test statistic could be $\Delta R^{(1)} = \pi - |\pi - |\angle\vec{R}_1 - \angle\vec{R}_2||$, the absolute difference between two preferred directions. If the alternative H_1 is non-equal sample concentrations, $\Delta R^{(2)} = ||\vec{R}_1| - |\vec{R}_2||$; that is, the absolute difference between the amplitudes of the resultants, could be utilized.

4.3.2. Application: comparison of tuning curves

Trials included presentation of and movement towards different objects in the same six directions. A physiological question of interest is whether a certain unit changed its directional tuning in response to different objects; in order to address this question for one specific PMd unit, directional tuning curves during trials involving power grips ($\vec{n} = \{22, 20, 19, 20, 19, 18\}$) and precision grips ($\vec{n} = \{20, 22, 20, 20, 23, 21\}$) were estimated (Fig. 4A). During preparation for movement, this unit had significant directional tuning to *both* objects (determined by showing statistical significance for both a Kruskal–Wallis test and a re-sampling test (Crammond and Kalaska, 1996) for each object separately). As the figure suggests, the circular distributions appear to differ from one another. Applying a non-parametric (likelihood-ratio) test of goodness-of-fit to these data indicated that the two discrete distributions indeed differed ($p < 0.01$); the permutation test described above yielded a similar result (Fig. 4B). Importantly, the permutation test indicated that the difference was mainly due to a change of *tuning width* between objects rather than due to a change of PDs (Fig. 4C and D).

4.4. Testing hypotheses on partitioned data: spatial organization

For each unit recorded in the experiment, a preferred object (PO) was computed as the object that elicited the maximal response (Mann–Whitney U-test, $p < 0.01$). During each recording session, up to eight electrodes were inserted through a common guide tube (inter-electrode distance $\sim$300 μm), so multiple POs could be obtained from electrodes in close proximity. Fig. 5A shows, in different rows, POs from different sessions (recording sites). Usually, several units were recorded by a single electrode and sometimes several had POs. These are indicated by heavy line boxes in Fig. 5A.

A question of physiological importance is whether POs recorded in the same site are as similar as expected from the entire sample. A negative answer would indicate some kind of spatial arrangement of units with similar preferred objects, since units recorded by the same electrode (electrode set, $S_{\text{electrodes}}$) are anatomically closer than units recorded by different electrodes at the same recording site (S_{sites}), which are, in turn, closer than units recorded in different sites (S_{PMd}). Examination of the data of Fig. 5A indicates that overall, both object types were preferred to a similar degree (28 and 21 preferences for precision and power grips, respectively). Closer inspection of the different rows (sites) suggests, however, that multiple units recorded in the same site often had the same PO.

How can this tendency be quantified? A simple and direct method would be to compute all pair-wise PO 'differences' within sites and compare them to pair-wise differences in the entire sample. Define a pair-wise difference as 0 if the two objects are identical and as 1 if they are different. For the data of Fig. 5A, the mean pair-wise PO-difference within sites was 0.32 and across all observations it was 0.5. The two populations differed significantly (Mann–Whitney U-test, $p < 0.01$), indicating that nearby neurons tended to 'prefer' similar objects.

The direct method exemplified above has, however, three drawbacks: (1) it considers only pair-wise differences (or similarities), neglecting *higher-order similarities*; (2) it does not account for possible *intra-electrode effects*; (3) an *approximation* of the probability calculation was used. While this latter technicality can be easily amended by using an exact binomial test, the issue of higher-order similarities cannot.

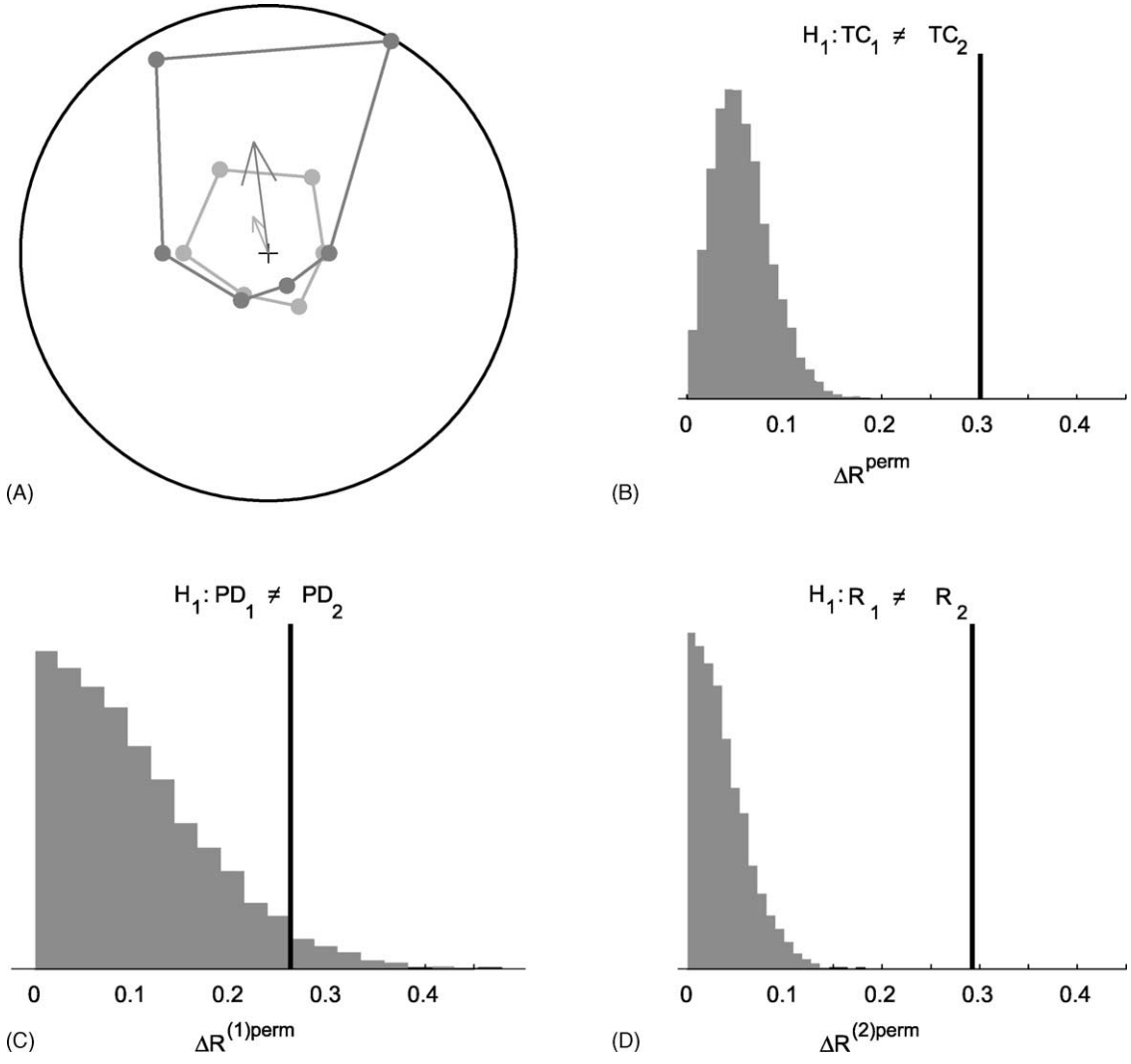

Fig. 4. Differences between tuning curves of a PMd unit during preparation for movement. (A) Tuning curve for a power grip object (light gray), superimposed on a tuning curve for a precision grip object (dark gray). Radius corresponds to 25 spikes/s; other conventions are the same as in Fig. 3B. (B) Histogram of the permuted statistics ΔR^{perm} values, testing the null hypothesis that the circular distributions are the same. Vertical line indicates the value of the observed ΔR. The null hypothesis is strongly rejected. $N^{\text{perm}} = 5000$. (C) Values of the permuted statistic $\Delta R^{(1)\text{perm}}$, testing whether the two distributions have identical PDs. The observed $\Delta R^{(1)}$ is larger than only 94% of the permuted statistics' values, hence the null hypothesis cannot be rejected at the 5% level. (D) Histogram of the permuted statistic $\Delta R^{(2)\text{perm}}$. The observed value (vertical line) was larger than all bootstrapped values, indicating that the difference between the distributions is likely to be due to different tuning widths.

This can be illustrated with a simple example. Assume that four Bernoulli trials with a success probability of 1/2 are performed (analogous to four POs within the same site), and that in all trials failures were observed. If pair-wise probabilities are computed, then six pairs, each consisting of two failures with a probability of 1/4, should be observed; the probability of observing such a pair can be estimated from the sample by the geometric mean of all observations, which in this case is 1/4. However, the a-priori probability of four failures in four trials is 1/16. The discrepancy is due to the fact that by limiting quantification to pairs, the higher-order structure was missed altogether.

The second problem which arises from ignoring intra-electrode similarities is more severe. Intra-site similarities (rows of Fig. 5A) are obviously affected by intra-electrode similarities (heavy boxes within rows): in six out of nine cases, intra-electrode pairs were identical. In addition, some intra-site similarities were entirely due to intra-electrode similarities (see, for example, seventh row from the bottom). The direct method used above does not take these similarities into account at all.

Both the first and second problems could be solved by using a bootstrap test of hypotheses. The null hypothesis is that intra-site similarities are the same as similarities throughout the entire sample. For each 'element' s (for example, a site, i.e., a row of Fig. 5A) a statistic $t(s)$ is computed; in our example, it can simply be the number of 'ones' in the site, denoted by m. Then, a probability is estimated under some assumption A^{global}, for example, of random allocation from the entire sample. Denote the global probability of the occurrence of a

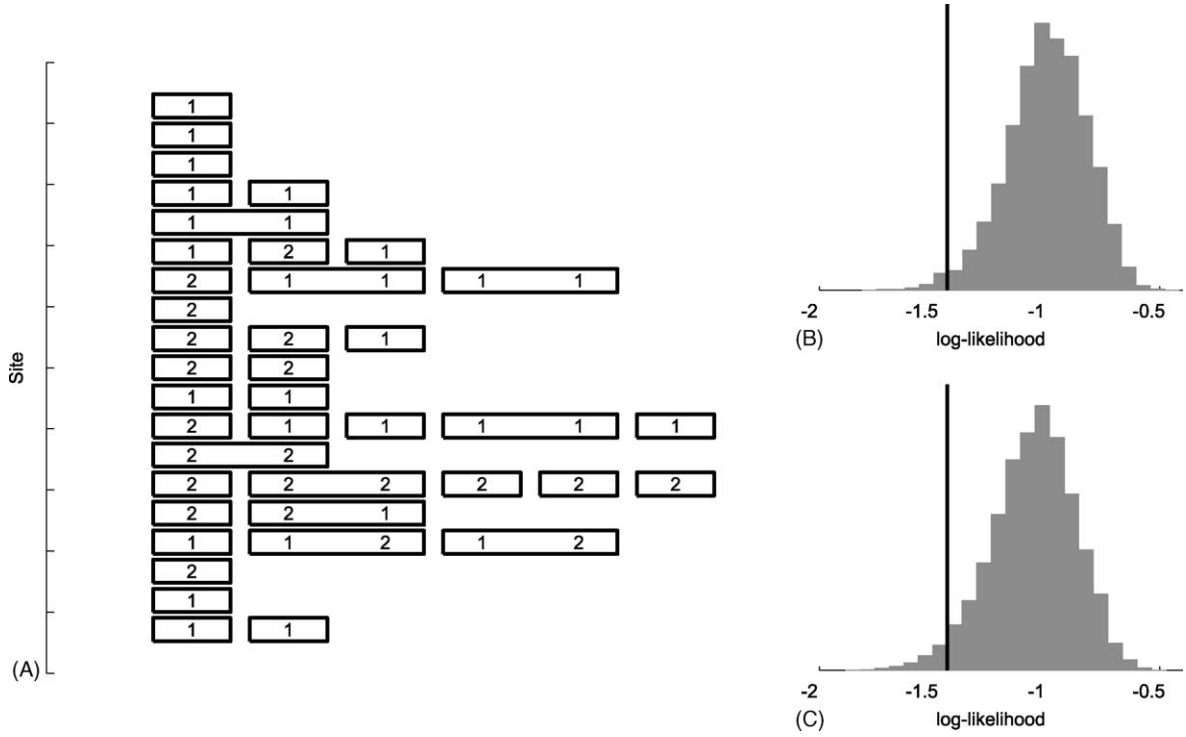

Fig. 5. Partitioned binary observations: spatial arrangement of preferred objects. (A) Example of a dataset (see also Section 4.4). Observations are preferred objects (POs) of PMd units from one monkey, determined by applying a Mann–Whitney *U*-test to spike counts during the 400 ms just before the 'Go' signal. '1' indicates preference for a precision grip, while '2' indicates preference for a power grip; all observations together constitute S_{PMd}. Different rows correspond to different recording sites (S_{sites}), hence observations within the same row are from units anatomically closer than between rows. Heavy line boxes within each row enclose numbers corresponding to objects preferred by units recorded from the same electrode ($S_{\mathrm{electrodes}}$). (B) Testing equality of intra-site and population (sample) similarities (sets S_{sites} and S_{PMd}). Vertical line shows the observed value of the log-likelihood, $L(S_{\mathrm{PMd}})$. Similarities were unlikely to be equal ($p < 0.01$), since almost all bootstrapped values were larger than the observed $L(S_{\mathrm{sites}})$ (black line). (C) Comparison of intra-site similarities to the similarities in the entire population, taking into account intra-electrode influences. Resampling was done separately for electrodes with two POs (heavy line boxes) and separately for electrodes with a single PO (see Section 4.4). Comparison of the sample statistic, $L(S_{\mathrm{sites}})$ with its bootstrapped distribution suggested that intra-site similarities are neither explained by intra-electrode similarities nor are likely to arise by chance ($p < 0.05$); however, intra-electrode similarities clearly affected this result (compare with (B)).

'one' (a precision grip object) by p (here, 0.57), the number of such objects in a given site (row) by m, and the total number of observations in that site by M. The probability to observe m or fewer 'ones' (or $M-m$ or fewer 'twos', whichever is smaller) is then given by

$$p(t(s)|A^{\mathrm{global}}) = \min\left\{\sum_{k=0}^{m}\binom{M}{k}p^k(1-p)^{M-k}, \sum_{k=0}^{M-m}\binom{M}{k}(1-p)^k p^{M-k}\right\}. \quad (6)$$

Since the sample includes a number of sites, the average log-likelihood of the set is computed by

$$L(S) = \frac{1}{|S|}\sum_{s\in S}\log(p(t(s)|A^{\mathrm{global}})), \quad (7)$$

where, in our case, S is the set of sites S_{sites} and $|S|$ the total number of sites (19). The statistic $L(S)$ reflects the probability of obtaining the observed intra-site similarities of *all orders*, normalized by the number of sites in the sample; in the case of our example, it is equal to $L(S_{\mathrm{sites}}) = -1.44$.

Next, this statistic is compared to its expected value. Since sites of different size were used (tackling the first problem described above by taking higher order similarities into consideration), there is no analytically-computable expected value. However, by random assignment (with replacement) of POs from the entire sample to sites and repetitive calculations of $L(S^{\mathrm{bs}}_{\mathrm{sites}})$, a bootstrapped distribution of $L(S_{\mathrm{sites}})$ can be generated, and the probability of accepting the null hypothesis can be estimated from its tail (by $\mathrm{prob}(H_0) = \#\{(L(S^{\mathrm{bs}}_{\mathrm{sites},b}) \le L(S_{\mathrm{sites}})\}/N^{\mathrm{bs}}$). Fig. 5B shows the results of applying this procedure to our dataset: the observed $L(S_{\mathrm{sites}})$ was indeed surprising, being smaller than 99% of its bootstrapped values, suggesting that within-site similarities could not be attributed to chance alone.

To address the second problem above (the confounding effect of intra-electrode similarities on intra-site similarities), resampling should be performed in a manner that *conserves subset structure*, the composition of units recorded by each electrode. If a certain row of S_{sites} contains d electrodes (for

example, POs of units recorded by three electrodes in the same site, Fig. 5A, fourth row from the bottom), resampled 'sites' should also contain *d* electrodes, each of the same size and structure (POs of three electrodes, two with two POs and one with one PO, but not necessarily all from the same site (row)). Note that the value of $L(S_{\text{sites}})$ does not change—only the null hypothesis does, as it now considers intra-electrode effects. This procedure yielded *smaller* bootstrapped values (Fig. 5C), indicating that some intra-site similarities could indeed be attributed to intra-electrode similarities. Nevertheless, $L(S_{\text{sites}})$ was smaller than over 95% of these bootstrapped values, suggesting that some anatomical organization beyond similarities within the same electrode does exist among the recording sites.

Two points should be stressed. First, the assumption A^{global} used to estimate the probabilities is not verified during the procedure; in fact, it does not have to be exactly correct: the only requirement is for it to be applied to *all* (the original and the resampled) sets during the bootstrapping procedure. Second, assignment of observations from set to set should *conserve subset structure*, as exemplified above.

The above procedure may be generalized to any dataset obeying similar rules of partitioning (in each set, each observation appears once, and each set is a union of the elements of another, hierarchically lower set). A variant of this procedure, the *Circular Variance Test*, was applied by Ben-Shaul et al. (2003) to PDs of units recorded in the Macaque motor cortex. In this paper we have described an application to nominal variables (preference for discrete object types). Clearly, other applications are possible as well.

5. Discussion

We have presented applications of resampling methods to four types of physiological issues. The first procedure was applied to estimates of mutual information from discrete neural data, and allowed placing confidence limits on such estimates, a task that does not have an analytical equivalent. The second procedure involved confidence limits as well, this time on circular parameters that often describe data obtained from neurophysiological experiments. Here, the goal could be achieved by making parametric assumptions on the data, assumptions that, as shown, can yield incorrect results. These two procedures used the bootstrap method of resampling. The third procedure was also related to circular data, and involved a comparison of distributions; for this application, use of a permutation test was appropriate, due to the symmetric null hypothesis. While a non-parametric test of goodness-of-fit could be used with lower sensitivity, it could not reveal sources of differences. We concluded with a procedure designed to test specific hypotheses on spatially ordered data, which can be applied to any dataset that obeys certain rules of partitioning. Due to the complicated data structure and possible interactions between 'sets', we used a bootstrapping-based test with non-trivial resampling rules.

Although these applications are by no means exhaustive, they do represent an array of problems that can be addressed in the framework of resampling methods. The procedures constructed here were applied to scalar and complex (polar) variables. Resampling methods may be easily applied to vector statistics, for example, to a cross-correlation function or to the spike-triggered average. In order to construct confidence limits for such statistics computed from a finite dataset, data could be resampled, the vector statistic could be recomputed, and the confidence limits could be estimated from the sorted values of *each element* of the vectors.

When can resampling methods be used to answer a statistical question? It turns out that under most circumstances and for most statistics, such methods are readily applicable. Exceptions to straightforward application include dependent sampling (deviation from iid sampling such as a time series), and non-smooth statistics (for instance, the sample's maximal value). Such cases may, however, be 'worked around' by particular resampling techniques or by smoothing data prior to resampling (Chee-Orts and Optican, 1993; Davidson and Hinkley, 1997).

A question that is particularly relevant to the neurophysiologist is therefore when *should* resampling methods be used. One of the main advantages of resampling methods, as demonstrated by the procedures presented in this paper, is that they are tuned to the data at hand: the bootstrap estimates the probability mechanism underlying data generation (Section 4.1) and permutation tests exploit the symmetry of the null hypothesis (Section 4.3). These properties imply that no parametric assumptions must be made during the process. Nevertheless, this is not all goodness. As is well known from conventional statistics, parametric tests, when their assumptions are met, are preferred over non-parametric tests due to their higher sensitivity (power, the probability to reject the null hypothesis when the alternative is correct) as well as due to their inferential value.

We thus propose the following scheme (consult Table 1). If data and statistic conform nicely to some parametric distribution, a *parametric test* is preferable. If this is not the case and a non-resampling, *non-parametric* test is available, it can be used, saving computer time and additional programming. If such a test does not exist or additional verification of test assumptions or performance is needed, a *resampling* (non-parametric) test should be utilized.

If only few assumptions can be made about the data but an appropriate non-resampling test does not exist, parametric resampling is possible. One would need to estimate parameters of the distribution assumed to have generated the data, and resample from *that* distribution: such resampling was originally termed 'Monte Carlo' simulations and is often used to evaluate properties of statistical tests. Alternatively, resampling may be performed directly under certain assumptions, tuning the test to fit the *type* of data in hand, rather than to a single, specific, dataset.

In summary, we presented a unified framework for the use of resampling methods by the analyst of neurophysiological

Table 1
Properties of different families of statistical tests

Test	Parametric	Non-parametric	Resampling
Applicability	Specific cases	Wide	Almost all cases
Assumptions	Fixed	Fixed, but usually weak	Flexible
Computation time	Short	Usually short	Could be long
Inference	Parametric	Non-parametric	Both non-parametric and parametric
Usage	Straightforward	Straightforward	May require additional planning or programming

data. We have shown that resampling-based procedures can be easily applied to a host of different types of problems yielding meaningful results, results that often cannot be obtained using conventional methods.

Routines for implementing the procedures described in this paper were written in C and in MATLAB and are available from ES upon request.

Acknowledgements

We thank Moshe Nakar for help with the construction of the experimental setup, Varda Sharkansky for technical help, and Itay Asher and Rotem Drori for help in carrying out the experiments. This research was supported in part by a Center of Excellence grant (8006/00) administered by the Israel Science Foundation, the Horowitz Foundation, RICH center, GIF, and DIP.

References

Barrow JD, Bhavsar SP, Sonoda DH. A bootstrap resampling analysis of galaxy clustering. Mon Not Roy Astron Soc 1984;210:19P–23P.

Ben-Shaul Y, Stark E, Asher I, Drori M, Nadasdy Z, Abeles M. Dynamical organization of directional tuning in the primate premotor and primary motor cortex. J Neurophysiol 2003;89:1136–42.

Chee-Orts MN, Optican LM. Cluster method for analysis of transmitted information in multivariate neuronal data. Biol Cybern 1993;69:29–35.

Cover TM, Thomas JA. Elements of information theory. New York: Wiley; 1991.

Crammond DF, Kalaska JF. Differential relation of discharge in primary motor and premotor cortex to movement posture during reaching movements. Exp Brain Res 1996;108:45–61.

Davidson AC, Hinkley DV. Bootstrap methods and their application. Cambridge: Cambridge University Press; 1997.

Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979;7:1–26.

Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman and Hall; 1993.

Georgopoulos AP, Kettner RE, Schwartz AB. Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of arm movement by a neural population. J Neurosci 1988;8:2928–37.

Holmes AP, Blair RC, Watson JDG, Ford I. Nonparametric analysis of statistic images from functional mapping experiments. J Cereb Blood Flow Metab 1996;16:7–22.

Mardia KV. Statistics of directional data. London: Academic Press; 1972.

Metropolis N, Ulam S. The Monte Carlo method. J Am Stat Assoc 1949;44:335–41.

Mills JA, Zandvakili S. Statistical inference via bootstrapping for measures of inequality. J Appl Econ 1997;12:133–50.

Optican LM, Richmond BH. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. J Neurophysiol 1987;57:162–78.

Panzeri S, Treves A. Analytical estimates of limited sampling biases in different information measures. Netw-Comput Neural Syst 1996;7:87–107.

Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W. Entropy and information in neural spike trains. Phys Rev Lett 1998;80:197–200.