Mathematical thoughts on statistics

Daniel Miller

May 19, 2017

1 General definitions

Let $\mathbf{R}$ be the set of real numbers, $R = R(\mathbf{R})$ the space of Radon measures on $\mathbf{R}$. Then R has a "positive subspace" R^+ consisting of all non-negative Radon measures, and if we put $R^- = -R^+$, then $R = R^+ + R^-$. Any $\nu \in R$ has a unique decomposition $\nu = \nu^+ + \nu^-$. There is a natural norm on the vector space R given by $|\nu| = \nu^+(1) - \nu^-(1)$.

Let $P = R^{+,|\cdot|=1}$ be the space of probability measure on $\mathbf{R}$. We can also interpret P as the positive subset of $R^{\cdot(1)=1}$, so its tangent space is $R^{\cdot(1)=0}$, the space of Radon measures ν with $\nu(1) = 0$.

2 Statistics and test statistics

Definition 1. A *statistic* is a function $T\colon P \to \mathbf{R}$.

In practice, we are never actually given the true probability distribution underlying data, so the best we can do is estimate it.

Definition 2. An *estimate* is a family $T_\bullet = (T_n)_{n\geqslant 1}$, where $T_n\colon \mathbf{R}^n \to \mathbf{R}$.

Given n points $x_1, \ldots, x_n$ drawn from $\nu \in P$, we estimate $T(\nu)$ by $T_n(x_1, \ldots, x_n)$. The distribution of the $T_n(x_1, \ldots, x_n)$ is $T_{n*}\nu^{\times n}$. We also call $T_n(x_1, \ldots, x_n)$ the *test statistic.*

Definition 3. The estimate $T_\bullet$ is *consistent* if $T_{n*}\nu^{\times n}$ converges to $\delta_{T(\nu)}$ in probability for any ν.

Write t for the identity function $\mathbf{R} \to \mathbf{R}$. Our first example of a statistic is the *mean*, $E\colon P \to \mathbf{R}$, given by $E(\nu) = \nu(t)$. Another example is the *variance*, given by $V(\nu) = \nu((t - E(\nu))^2) = \nu(t^2) - \nu(t)^2$.

Definition 4. The estimate $T_\bullet$ is *unbiased* if $E(T_{n*}\nu^{\times n}) = T(\nu)$ for all $\nu \in P$ and for all $n \geqslant 1$.

Ideally, there will be some kind of theoretical "best unbiased estimate."

We will also generally assume our estimates are *permutation invariant*, that is $T_n(x_1, \ldots, x_n) = T_n(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for all permutations σ.

Try $\nu = \delta_x$; then $T_{n*}\delta_x^{\times n} = \delta_{T_n(x,\ldots,x)}$. Unbiasedness tells us that at the very least, $T_n(x, \ldots, x) = T(\delta_x)$. This is our first example of the "plug-in estimate."

Now try $\nu = \frac{1}{2}(\delta_{x_1} + \delta_{x_2})$. Then

$$T_{n*}\nu^{\times n} = 2^{-n} \sum_{\sigma:[n]\to[2]} \delta_{T_n(x_{\sigma(1)},\ldots,x_{\sigma(n)})},$$

which has expected value (mean)

$$2^{-n} \sum_{\sigma:[n]\to[2]} T_n(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = 2^{-n} \sum_{k=0}^{n} \binom{n}{k} T_n(\overbrace{x_1, \ldots, x_1}^{k}, \overbrace{x_2, \ldots, x_2}^{n-k})$$

which we want to equal $\frac{1}{2}(x_1 + x_2)$.

What can we say: one should have $T_1(x) = T(\delta_x)$.

3 Linear statistics

A statistic T is *linear* if it extends to a linear map $T\colon R \to \mathbf{R}$. Equivalently, whenever $a_1, \ldots, a_n \geqslant 0$ and $\sum a_i = 1$ and $\nu_i \in P$, then $T(\sum a_i\nu_i) = \sum a_i T(\nu_i)$.

Any reasonable linear statistic will be of the form $T(\nu) = \nu(f)$ for some f a function on $\mathbf{R}$.

Let's see if there is a "best unbiased estimate" for linear statistics of this form.