Local false discovery rates

Patrick Breheny

January 26

Introduction

- We concluded the previous lecture with a look at how false discovery rates can be viewed as either a frequentist methodology or an empirical Bayes estimate
- From a Bayesian standpoint, however, the false discovery rate is somewhat strange, in the sense that it involves conditioning on a rejection region $z_j \in \mathcal{Z}$
- A more natural thing to do, as least from a Bayesian perspective, is to condition on the actual value of z; in other words, to estimate

$$\text{fdr}(z_0) = \mathbb{P}(H_0 | Z = z_0);$$

the *local false discovery rate* for H_{0j} is therefore $\text{fdr}(z_j)$

FDR applies to the group, not a specific test

- One reason that the FDR is somewhat unsatisfying is that, by conditioning on $z_j \in \mathcal{Z}$, we calculate a probability/rate applying generally to all hypotheses in that region
- This, however, ignores the fact that some z-values are much more extreme than others, or to put it another way, that not all hypotheses are equally likely to be contributing the false discoveries
- In the leukemia data for example, at an FDR of 1%, we can claim 734 discoveries; among them, $|z_j|$ ranges from 3.3 to 9.5
- FDR tells us to expect ≈ 7 false discoveries; those false discoveries are presumably much more likely to be coming from the tests with $z \approx 3$ than $z \approx 9$

The tale of the dishonest statistician

- To see why this might be a problem, let's take this line of reasoning to an extreme end: suppose we test $h = 1,000$ hypotheses, and the smallest p-value we get is $p = 0.001$
- If we want to control the FDR at 10%, this is well above the BH cutoff to reject the first gene (here, 0.0001)
- Suppose that the statistician, disappointed by the fact that we cannot reject any hypotheses, decides to add 10 additional tests for which they know in advance that the null hypothesis is false

The tale of the dishonest statistician (cont'd)

- As expected, the results for those 10 tests are highly significant
- Now, they go back to control the FDR for these 1,010 tests; the p-value cutoff for the 11th test is now $p = 0.0011$, so now we *can* reject the hypothesis that we couldn't on the previous slide
- This approach allows the statistician to publish a list of 11 "discoveries", of which 10 were known in advance, but hey, there's one interesting new discovery that we have "significant" statistical evidence for

Exchangeability

- This obviously flawed approach illustrates that false discovery rates come with a key assumption of exchangeability: if we're going to make significance statements about a *group* of tests, those tests should be as homogeneous as possible
- It isn't incorrect to say that the false discovery rate for those 11 discoveries is under 10%, but it's certainly misleading – it's pretty obvious which result is likely to be the false discovery
- This example is (hopefully) unrealistic, but the question of which hypotheses can be combined to form a relevant group arises quite often: for example, should we be combining the left and right tails?

Bayes rule again

- Following the same reasoning as at the end of the previous lecture, we can use Bayes rule to obtain an expression for the local false discovery rate:

$$\text{fdr}(z) = \frac{\pi_0 f_0(z)}{f(z)},$$

  where $f(z) = \pi_0 f_0(z) + \pi_1 f_1(z)$ is the marginal density of z-values and $f_0(z)$ is the null density

- Note: Many authors (including me) use Fdr to refer to the false discovery rate and fdr to refer to the local FDR, reflecting the F/f convention for denoting distribution and density functions, respectively

Remarks

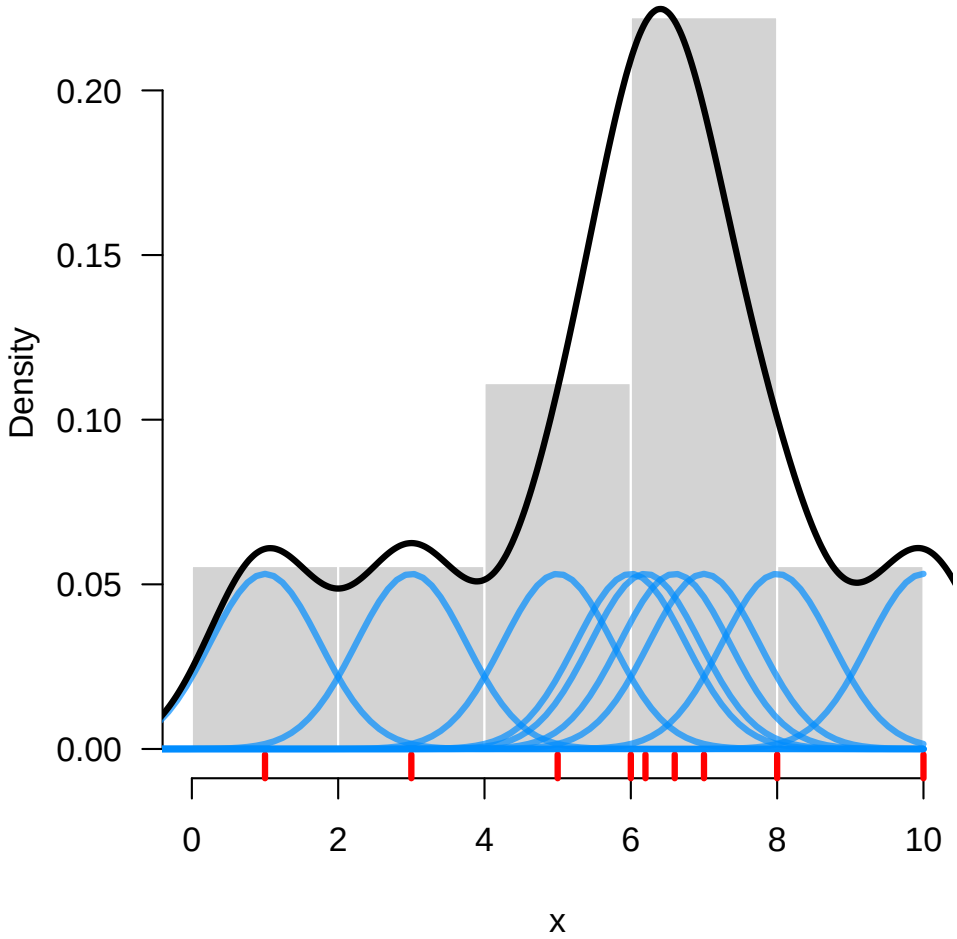
- Local FDRs offer a number of advantages over tail-area FDRs; for example, from a Bayesian perspective, conditioning on z is correct, not $z \in \mathcal{Z}$; in fact, the quantity $f_1(z)/f_0(z)$ is known as the *Bayes factor* for quantifying the level of empirical support for hypothesis 1 over hypothesis 0
- However, local FDR has faced two main challenges in terms of gaining widespread acceptance relative to tail-area FDR:
  - No interpretation as a frequentist error rate control procedure is available
  - Estimating a density (f) is far less straightforward than estimating a distribution (F), meaning that there are many variants of local FDR, unlike tail area FDR
- This may be changing (I've started to see local FDRs in prominent journals more often), but time will tell

Three ingredients

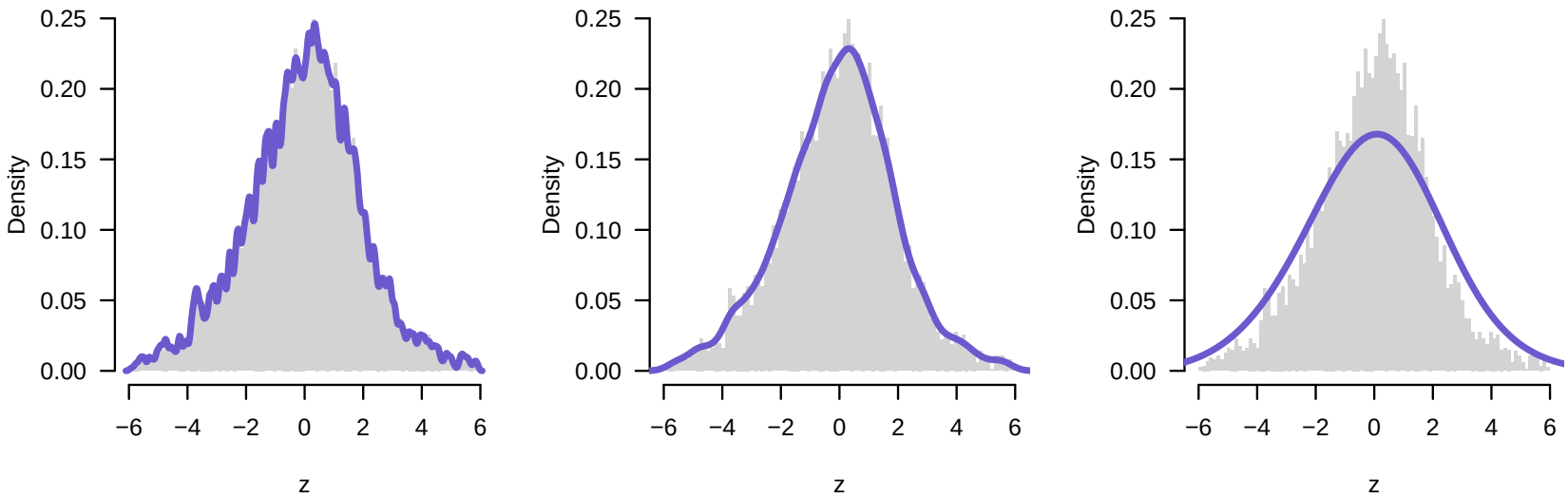
- The local false discovery rate has three components:
  - π_0
  - f
  - f_0
- Each of these can potentially be varied, producing different estimates of fdr
- We will start by looking at a simple approach for estimating these quantities, then discuss more sophisticated/complex approaches and alternatives

Density estimation using Gaussian kernels

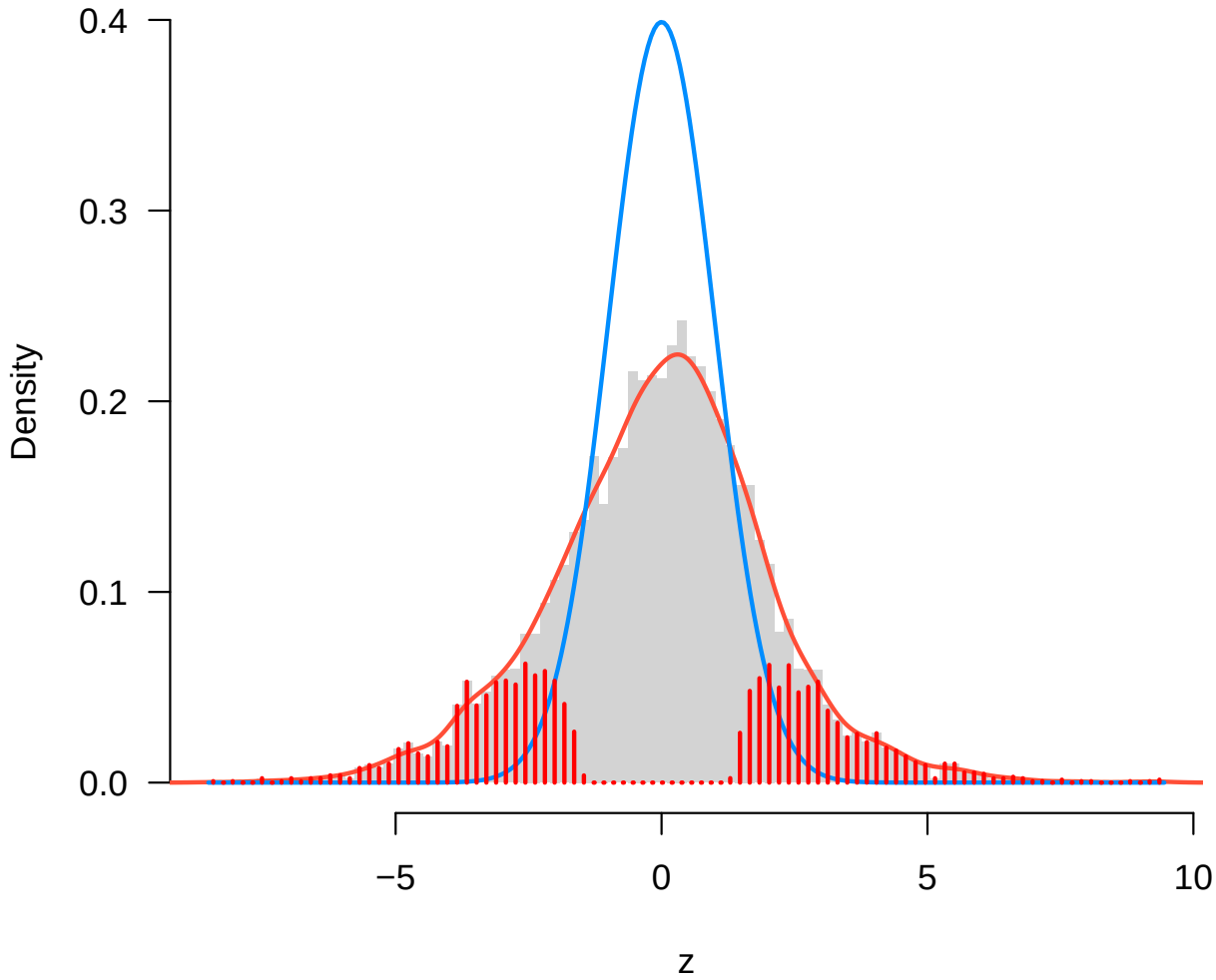
One common approach is *kernel density estimation*:



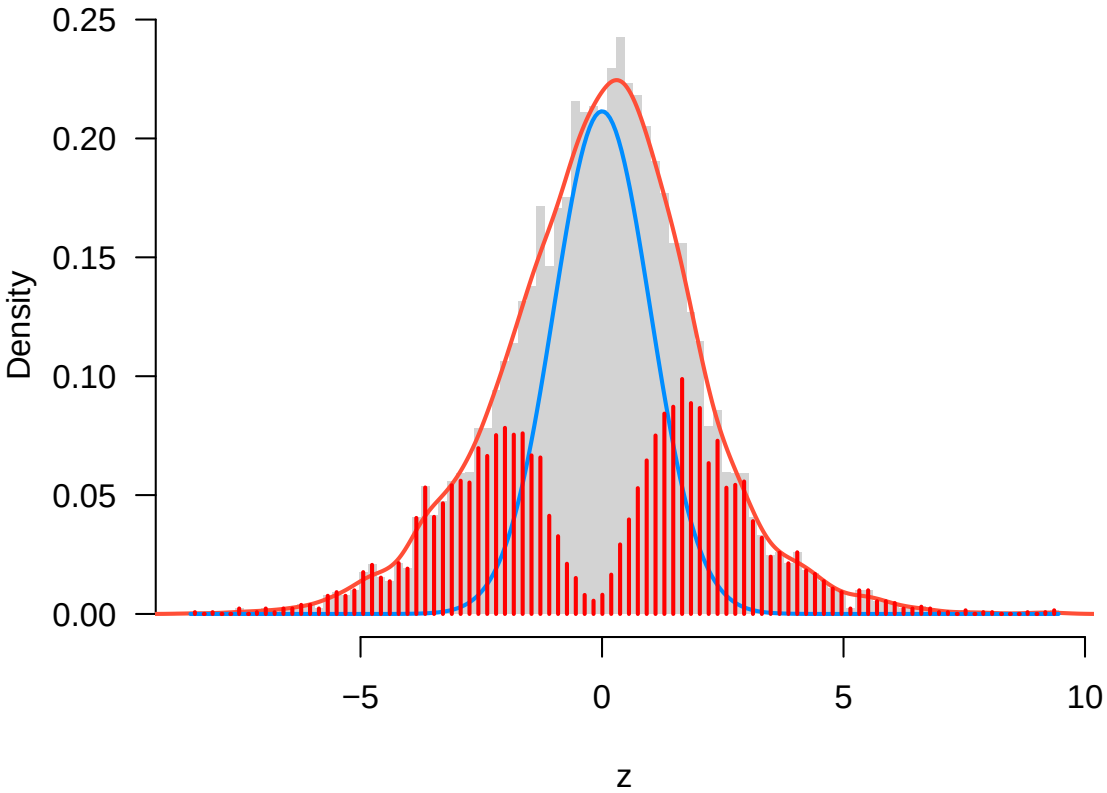
Choice of bandwidth



Local fdr for leukemia data: Illustration

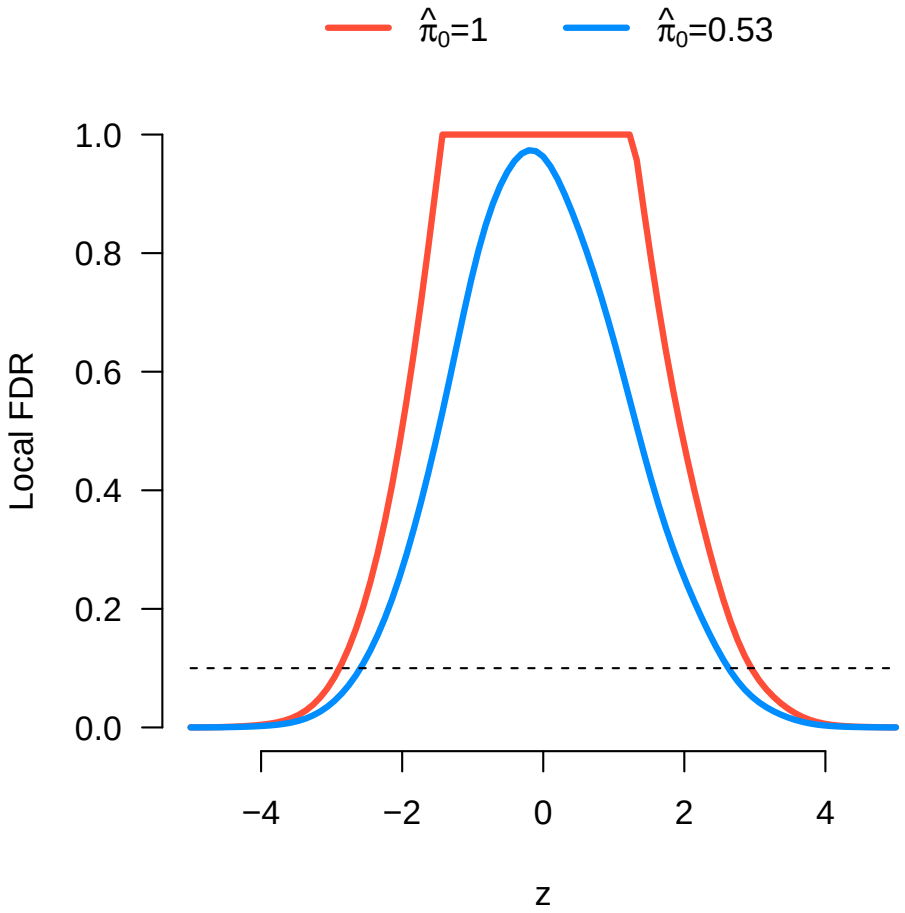


Local fdr for leukemia data: $\hat{\pi}_0 = 0.53$



Using $\hat{\pi}_0 = 0.53$, our estimate from the previous lecture, we seem to obtain more realistic estimations of the null and alternative distributions

z vs local FDR



For a 10% local FDR cutoff:

- Using $\hat{\pi}_0 = 1$, critical value of $z = 2.95$; 986 significant results
- Using $\hat{\pi}_0 = 0.53$, critical value of $z = 2.63$; 1,266 significant results

Estimating a null distribution?

- Lastly, one could consider estimating f_0 as well
- This is admittedly a somewhat weird idea – using the data to estimate the null – however, it has been proposed in the literature and studied by many authors
- The basic idea is to assume that $Z \sim \mathrm{N}(\delta_0, \sigma_0^2)$ and use the "central" part of the data to estimate δ_0 and σ_0
- It is certainly possible, for a variety of reasons, for the theoretical null $N(0, 1)$ not to hold; whether we can fix these problems by estimating a null is not always clear
- It's an interesting idea, but I'm not going to say much more about it in this lecture

Cutoff comparison

- It is worth spending a few slides on a deeper examination of Fdr versus fdr in terms of results and interpretation
- Using $\pi_0 = 1$, and a 10% cutoff,
  - Fdr: Critical $z = 2.27$; 1,635 significant findings
  - fdr: Critical $z = 2.95$; 986 significant findings
- For any given percentage cutoff, local FDR is considerably more conservative than tail-area FDR about declaring a result significant – a 10% Fdr means something quite different from 10% fdr

Conditional expectation relationship

- Further insight into the relationship between FDR and local FDR is given by this result:

$$\mathbb{E}\{\text{fdr}(z)|z \in \mathcal{Z}\} = \text{Fdr}(\mathcal{Z})$$

- Roughly, then, we should expect the average local FDR among the significant features to equal the FDR:
  - Left tail: Average fdr for features with $\text{Fdr} < 0.1$ is 0.102
  - Right tail: Average fdr for features with $\text{Fdr} < 0.1$ is 0.097
- This relationship does not exactly work out for two-sided tests unless we specifically estimate a combined tail density $f(|z|)$

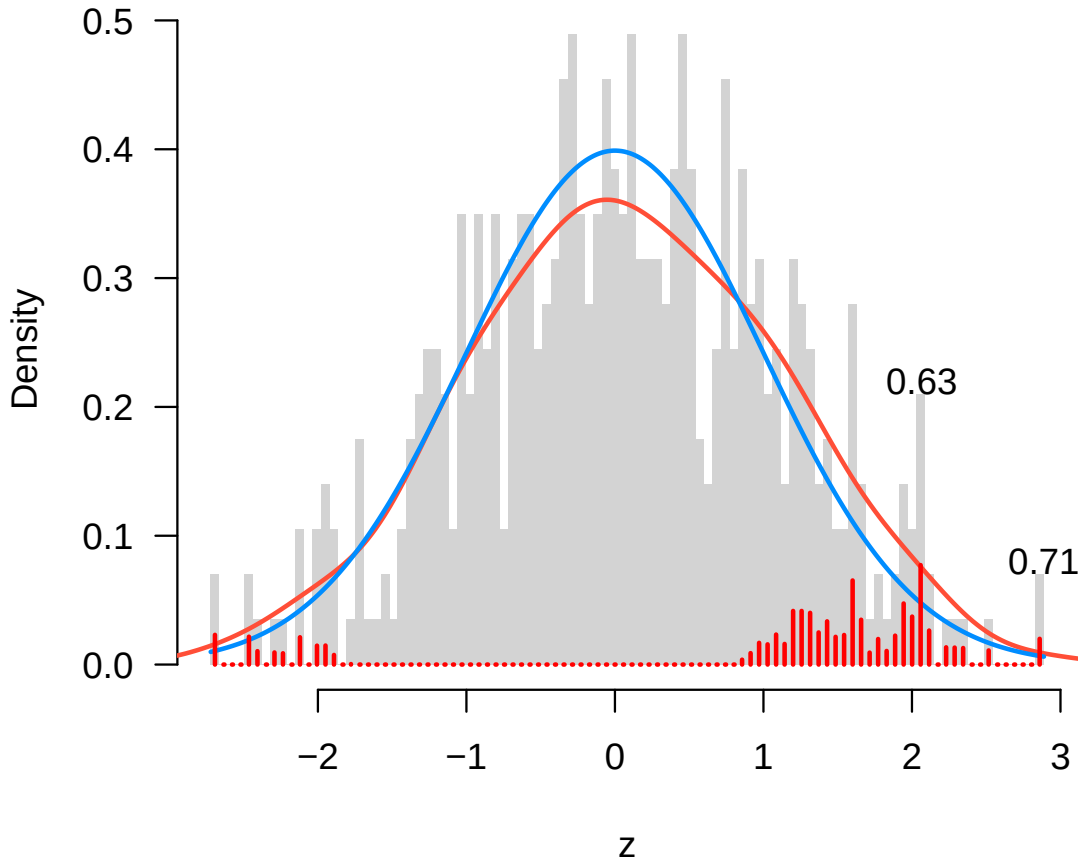
More sophisticated approaches

There are a number of R packages for calculating local FDRs, all of which take different approaches to density estimation (and potentially π_0 and f_0):

- **locfdr**: Uses Poisson regression with splines to model histogram counts
- **fdrtool**: Uses a special form of density estimation that enforces a monotonicity constraint (avoiding density "bumps")
- **ashr**: Uses mixture models (one for the null, many for the alternatives)

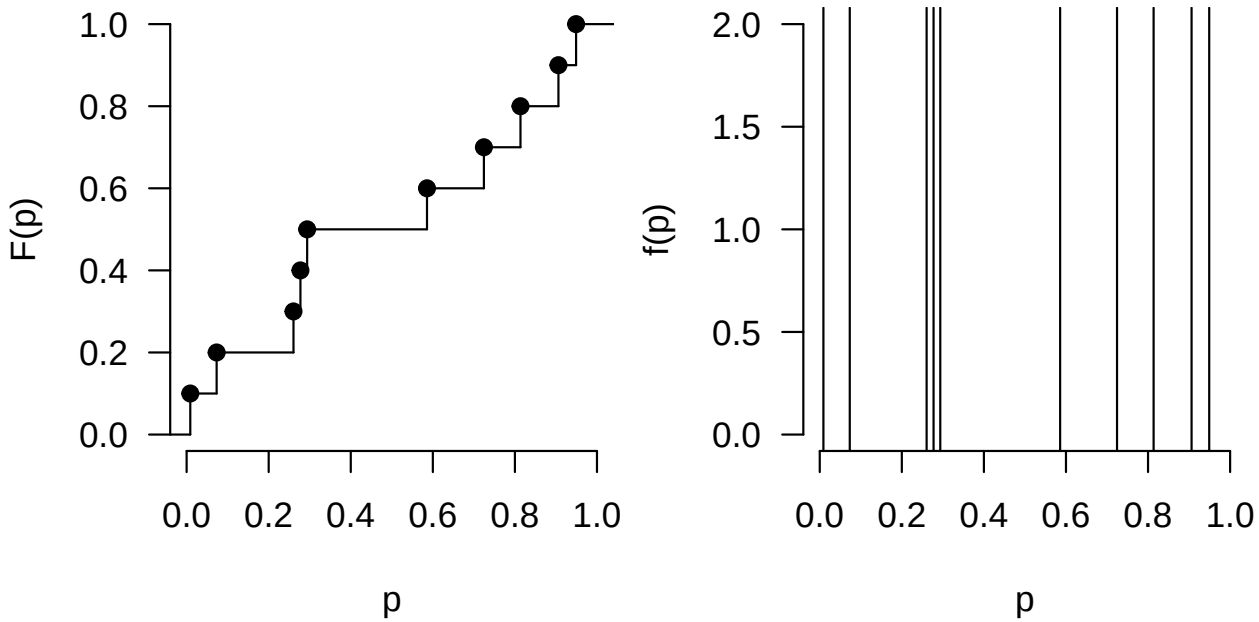
The problem with density "bumps"

One potential issue with kernel density estimation is that the fdr is not necessarily a monotone function of z or p:



CDF and density

- The main idea of `fdrtool` is to estimate density via the CDF
- Among other things, this has the advantage that Fdr and fdr are based on the same estimate and always logically consistent
- However, we can't just use the CDF directly:

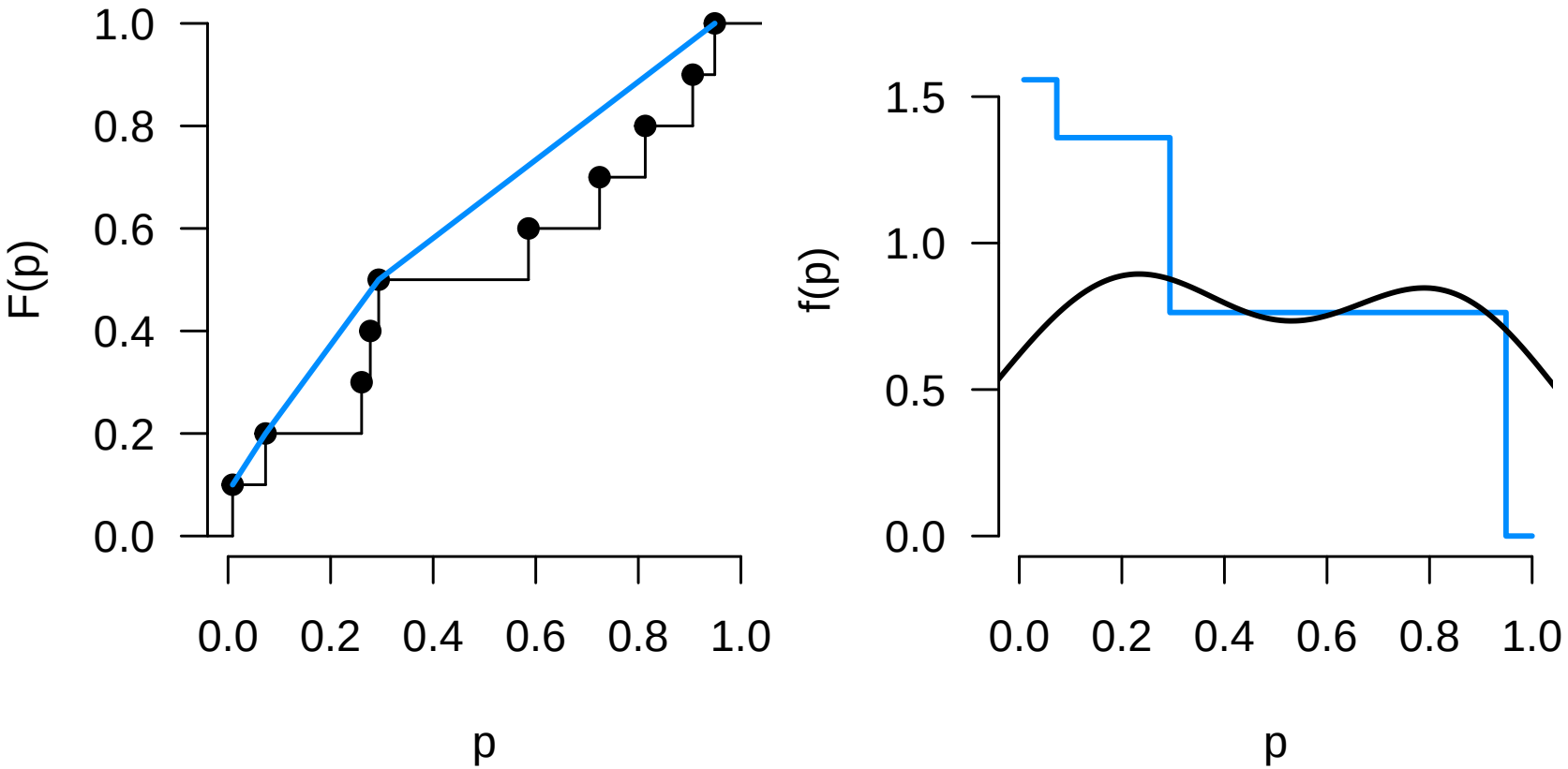


Grenander estimator

- Furthermore, we want our CDF estimator to produce a monotone density: $f(p)$ should be increasing as $p \to 0$
- A classical method for accomplishing this was proposed by Ulf Grenander in 1956
- Recognizing that a monotone decreasing density corresponds to a concave distribution function (likewise, increasing density means convex distribution), Grenander proposed estimating the CDF using the *least concave majorant* of the empirical CDF

Grenander estimator: Illustrated

Doing so produces a piecewise constant density:



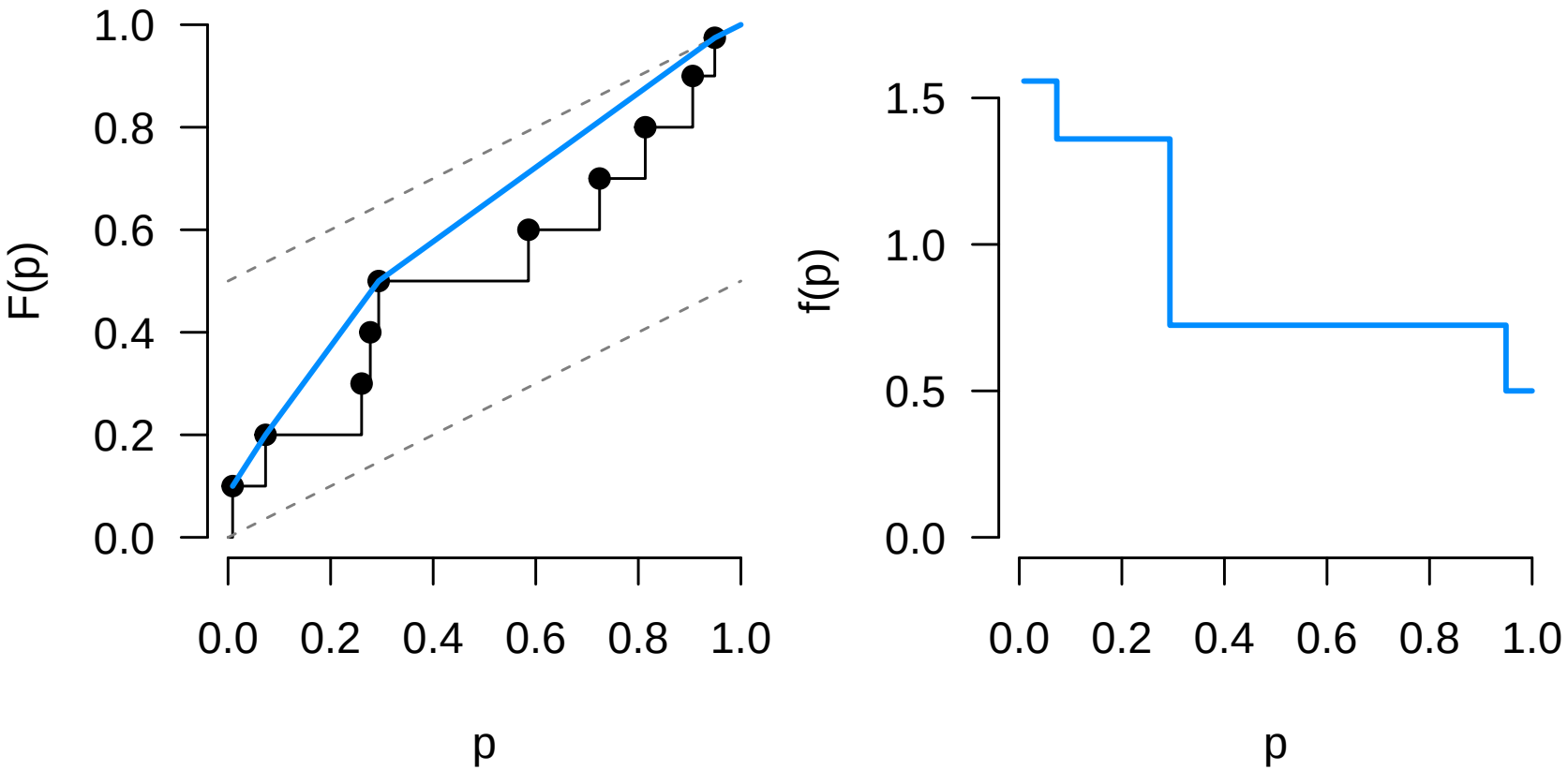
Accommodating π_0

- This achieves the desired result – local fdr will now be a monotone function of the p-value
- To accommodate a mixture of null and alternative densities, we must subtract $\pi_0 \text{Unif}(0, 1)$ from our density estimate
- This potentially introduces problems, since we could end up with a negative density
- To avoid this, note that the mixture model $F = \pi_0 F_0 + (1 - \pi_0) F_1$ introduces two constraints:

$$F(p) \geq \pi_0 p$$
$$F(p) \leq 1 - \pi_0(1 - p)$$

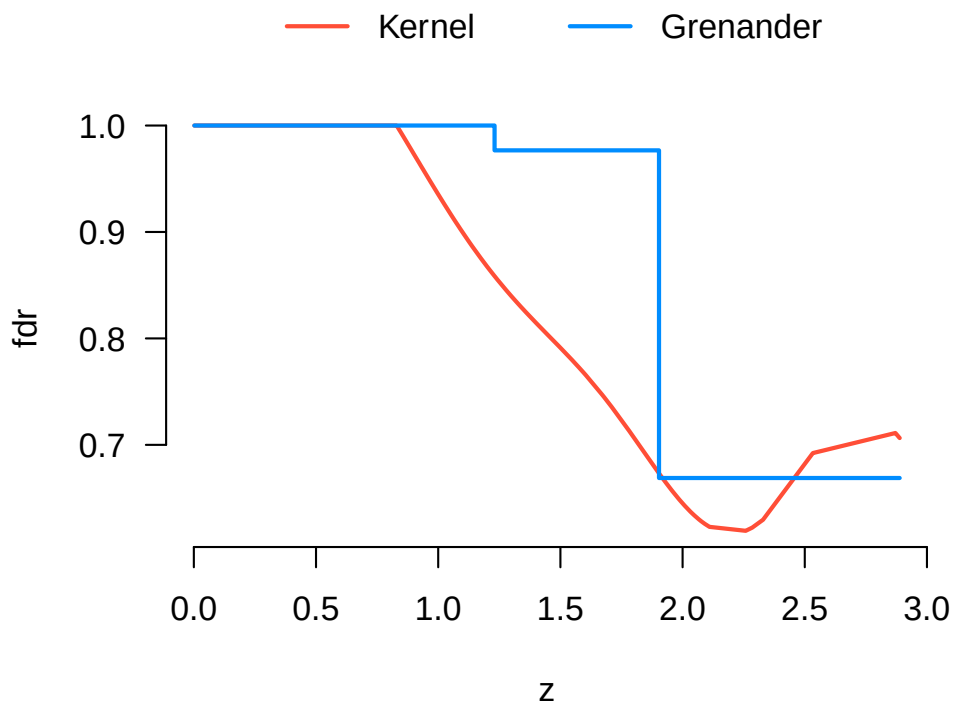
Modified Grenander estimator: Illustrated

With $\pi_0 = 0.5$:



Revisiting the "bumpy" data from earlier

Thus, we obtain a robust, unified framework for estimating both Fdr and fdr, and ensure both estimates agree with each other and are monotone functions of the original test statistic:



Usage

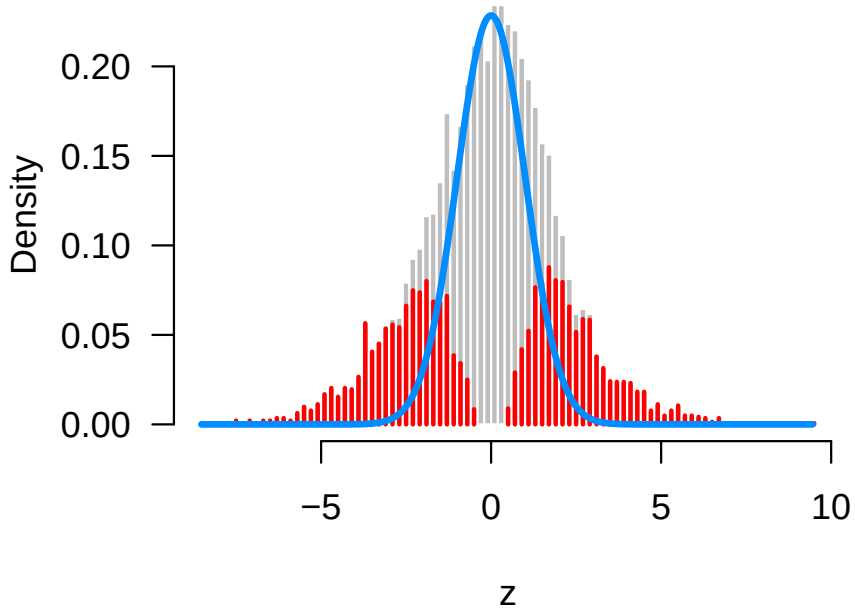
- Usage of the **fdrtool** package is fairly straightforward:

```
res <- fdrtool(p, statistic='pvalue')
```

  this returns a list containing (among other things) `res$qval` and `res$lfdr`
- By default, the function also produces some diagnostic plots; there are options to turn this off and to choose from various approaches for estimating π_0
- One very important thing to be aware of is that if you supply z statistics, `fdrtool` will estimate the null distribution as well as the alternative (this may or may not be what you want)

Application to leukemia data

- Applying this method to the leukemia data, we have 1,202 genes with $\text{fdr} < 0.1$ (similar to the kernel approach, which found 1,266 genes at this threshold)



- Note: If we model the null, 40 genes with $\text{fdr} < 0.1$