

Theory of Inference: Homework 4

1. Here is a slightly technical question about hypothesis tests. Suppose that

$$Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda) \quad \lambda > 0.$$

The two competing hypotheses are

$$H_0 : \lambda = \lambda_0 \quad \text{versus} \quad H_1 : \lambda = \lambda_1$$

for specified values of λ_0 and λ_1 .

(a) Derive the form of the likelihood ratio, $f_0(\mathbf{y})/f_1(\mathbf{y})$.

Answer. First, we need the PMF for \mathbf{y} , which is

$$f(\mathbf{y}; \lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{y_i}}{y_i!} \propto e^{-n\lambda} \lambda^{\sum_i y_i} = e^{-n\lambda} \lambda^{n\bar{y}}$$

where \bar{y} is the sample mean; we can ignore multiplicative terms that do not depend on λ (since they will cancel in ratios). Then

$$\frac{f_0(\mathbf{y})}{f_1(\mathbf{y})} = e^{-n(\lambda_0 - \lambda_1)} \left(\frac{\lambda_0}{\lambda_1} \right)^{n\bar{y}}.$$

(b) Evaluate this ratio for the values $\lambda_0 = 0.50$, $\lambda_1 = 1.25$ and the dataset

$$\mathbf{y}^{\text{obs}} = \{1, 0, 2, 2, 0, 1, 1, 1, 1, 1, 3\}.$$

Thinking like a Bayesian, what does the dataset say about the two hypotheses?

Answer. Here is the R code I used to generate the dataset and evaluate the likelihood ratio.

```

set.seed(101)
lambda0 <- 0.50
lambda1 <- 1.25
n <- 11
y <- rpois(n, lambda1); print(paste(y, collapse = ", "))

ybar <- mean(y)
logLR <- -n * (lambda0 - lambda1) + (n * ybar) * log(lambda0 / lambda1)
print(exp(logLR)) # 0.02568676

```

This is a likelihood ratio of about 1/40. To a Bayesian who did not have strong prior beliefs about λ , this would indicate strong support for H_1 over H_0 . This is reassuring because you will see from the code that the ‘true’ hypothesis is H_1 .

- (c) (Harder—maybe just think about this one over Easter.) Evaluate the Type 1 and Type 2 errors of the rejection region

$$\mathcal{R} := \left\{ \mathbf{y} : \frac{f_0(\mathbf{y})}{f_1(\mathbf{y})} \leq k \right\}$$

for $k \leftarrow 0.5$. Hint: you will find the Normal approximation to the Poisson very helpful for a pen-and-paper calculation (leave your results expressed in terms of the Normal distribution function), or else you can do the exact calculation in R, using the `ppois` function.

Answer. Taking $\lambda_0 < \lambda_1$, we have

$$\mathbf{y} \in \mathcal{R} \iff n\bar{y} \geq \frac{\log k + n(\lambda_0 - \lambda_1)}{\log(\lambda_0/\lambda_1)} =: c$$

(make sure you check this!). If the Y ’s are IID Poisson with rate λ , then the sum of the Y ’s is also Poisson (prove this using the Moment Generating Function), and its rate is its expectation, $n\lambda$. Using the Normal approximation to the Poisson, we then have

$$n\bar{Y} \sim N(\cdot; n\lambda, n\lambda),$$

approximately, remembering that the variance of a Poisson is the same as the

expectation. Therefore the Type 1 error is approximately

$$\alpha := \Pr_0(\mathbf{Y} \in \mathcal{R}) = \Pr_0(n\bar{Y} \geq c) \approx 1 - \Phi(c; n\lambda_0, n\lambda_0)$$

where Φ is the Normal distribution function, with specified expectation and variance. The Type 2 error is

$$\beta := \Pr_1(\mathbf{Y} \notin \mathcal{R}) \approx \Phi(c; n\lambda_1, n\lambda_1)$$

Here is some R code for the Type 1 and the Type 2 error:

```
k <- 0.5
c <- (log(k) + n * (lambda0 - lambda1)) / log(lambda0 / lambda1)
print(c) # 9.760163

alpha <- pnorm(c, n * lambda0, sqrt(n * lambda0), lower.tail = FALSE)
print(alpha) # 0.03464381

beta <- pnorm(c, n * lambda1, sqrt(n * lambda1))
print(beta) # 0.1409683
```

Now let's also do the exact calculation in R:

```
alpha <- ppois(floor(c), n * lambda0, lower.tail = FALSE)
print(alpha) # 0.05377747
beta <- ppois(c, n * lambda1)
print(beta) # 0.1217738
```

showing that the Normal approximation is accurate to about 2 percentage points, in this case.

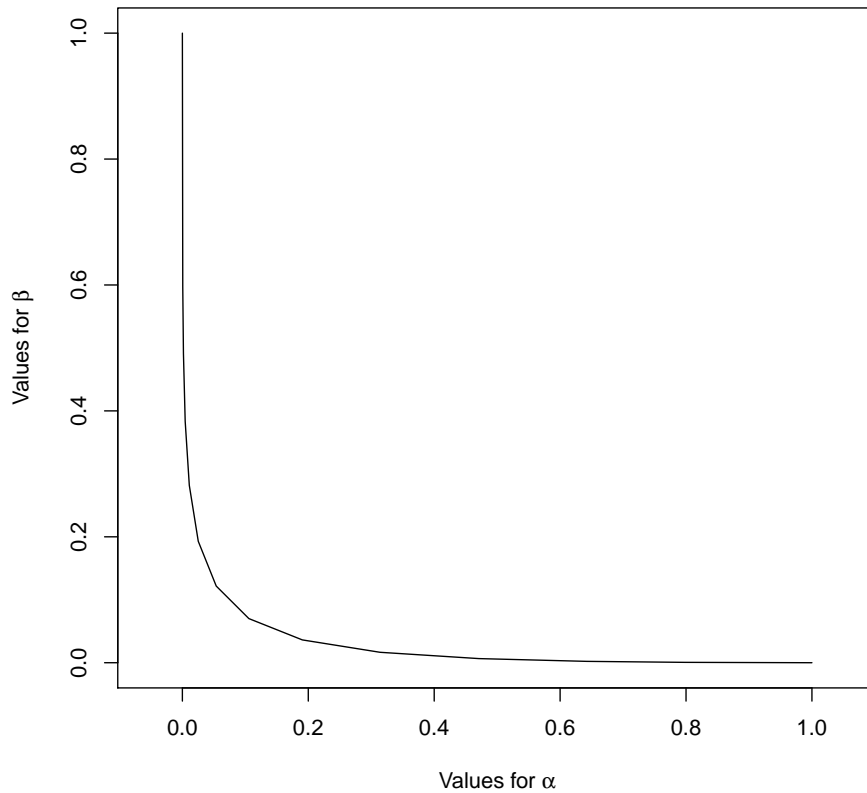
- (d) (Another one to think about.) Produce a plot of the Type 1 error (x -axis) versus the Type 2 error (y -axis) for a range of values for k . This curve is known as the *operating characteristics curve* for this model and this pair of hypotheses. It is obvious that operating characteristics curve goes through $(0, 1)$ and $(1, 0)$, and it is fairly easy to prove that it is always convex (follows from the Neyman-Pearson Lemma).

Answer.

```
kvals <- seq(from = -2, to = 2, length = 101)
kvals <- 10^kvals
opchar <- sapply(kvals, function(k) {
  c <- (log(k) + n * (lambda0 - lambda1)) / log(lambda0 / lambda1)
  alpha <- ppois(floor(c), n * lambda0, lower.tail = FALSE)
  beta <- ppois(c, n * lambda1)
  c(k = k, alpha = alpha, beta = beta)
})
opchar <- t(opchar) # has columns: k, alpha, beta
opchar <- rbind(c(0, 0, 1), opchar, c(Inf, 1, 0)) # always have
these endpoints

plot(opchar[, c("alpha", "beta")], type = "l", col = "black",
     asp = 1, xlim = c(0, 1), ylim = c(0, 1),
     xlab = expression(paste("Values for ", alpha)),
     ylab = expression(paste("Values for ", beta)),
     main = "Operating characteristics curve")
dev.print(pdf, "opchar.pdf")
```

Operating characteristics curve



2. Here is an exam-style revision question about hypothesis tests. Part 2d is unseen and a bit harder. Each part carries five marks.
- (a) Describe the general framework for Hypothesis Testing, and give illustrations of the types of hypotheses that might be tested.

Answer. I'll just sketch the answers, without reference to the notes.

A hypothesis is a statement about the values of the parameters in a parametric model for some observables. In a hypothesis test we assess the evidential support for two competing hypotheses. So if the parametric model is $\mathbf{Y} \sim p(\cdot; \theta)$ for $\theta \in \Omega$, then the two hypotheses can be written, in general, as

$$H_0 : \theta \in \Omega_0 \quad \text{versus} \quad H_1 : \theta \in \Omega_1$$

where $\Omega_0, \Omega_1 \subset \Omega$, and $\Omega_0 \cap \Omega_1 = \emptyset$. If Ω_i is a single point, then H_i is a ‘simple’ hypothesis; otherwise it is a ‘composite’ hypothesis.

One illustration is for two competing models: $\mathbf{Y} \sim p_0(\cdot)$ versus $\mathbf{Y} \sim p_1(\cdot)$, which can be thought of as special cases of a single more general model with $\theta \in \{0, 1\}$. In this case both H_0 and H_1 are simple.

Another common case is where some value θ_0 divides the scalar parameter space into two parts. In this case we have $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, and both H_0 and H_1 are composite. One special case of this is where $\Omega = [\theta_0, \infty)$, in which case H_0 is simple and H_1 is composite.

A third illustration is where a hypothesis concerns only one of the parameters, say $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$, in the case where $\mathbf{Y} \sim N(\cdot; \mu, \sigma^2)$. In this case both H_0 and H_1 are composite, because the value of σ^2 is not specified.

- (b) Explain how a Bayesian might choose between two simple hypotheses.

Answer. Let the simple hypotheses be

$$H_0 : \mathbf{Y} \sim p_0(\cdot) \quad \text{versus} \quad H_1 : \mathbf{Y} \sim p_1(\cdot).$$

According to Bayes’s theorem in odds form,

$$\frac{\Pr(H_0 | \mathbf{Y} \doteq \mathbf{y})}{\Pr(H_1 | \mathbf{Y} \doteq \mathbf{y})} = \frac{p_0(\mathbf{y}) \Pr(H_0)}{p_1(\mathbf{y}) \Pr(H_1)}.$$

The Bayesian will typically favour the hypothesis with the larger posterior probability; i.e. H_0 if the lefthand side (the posterior odds for H_0 versus H_1) is greater than one, and H_1 if it is less than one. The first term on the righthand side is the likelihood ratio. If this is much larger or much smaller than one, then it will typically dominate the second term, the prior odds. So the Bayesian will typically favour H_0 if the likelihood ratio is large, say more than 20, and H_1 if the likelihood ratio is small, say less than $1/20$. In the middle, he will need to include explicit prior probabilities: if he does not have these, then he remains undecided.

- (c) State and prove the Neyman-Pearson Lemma.

Answer. You can look this one up in the lecture notes! My proof was adapted

from section 8.2 of M. DeGroot and M. Schervish, 2002, *Probability and Statistics*, 3rd edn, Addison-Wesley.

- (d) Consider two simple hypotheses, H_0 and H_1 , and the rejection region

$$\mathcal{R} := \left\{ \mathbf{y} : \frac{f_0(\mathbf{y})}{f_1(\mathbf{y})} \leq k \right\}.$$

Let α be the Type 1 error level of \mathcal{R} , and β be the Type 2 error level (these are both functions of k). Show that

$$\frac{\Delta\alpha}{\Delta\beta} \approx -k,$$

where Δ indicates the change that arises from a small change in k .

Answer. Imagine ordering all the values of $\mathbf{y} \in \mathcal{Y}$ in terms of the values of $f_0(\mathbf{y})/f_1(\mathbf{y})$; denote this ordering $\mathbf{y}_{(1)}, \mathbf{y}_{(2)}$ and so on. Fix a value of k and let $\mathbf{y}_{(j)}$ be the value in \mathcal{Y} for which

$$\frac{f_0(\mathbf{y}_{(j-1)})}{f_1(\mathbf{y}_{(j-1)})} < k < \frac{f_0(\mathbf{y}_{(j)})}{f_1(\mathbf{y}_{(j)})},$$

or $f_0(\mathbf{y}_{(j)})/f_1(\mathbf{y}_{(j)}) \approx k$. Now slightly increase k , so that the rejection region now includes $\mathbf{y}_{(j)}$. The increase in Type 1 error is $f_0(\mathbf{y}_{(j)})$, and the increase in the Type 2 error is $-f_1(\mathbf{y}_{(j)})$, i.e. the Type 2 error goes down because the rejection region is bigger. So

$$\frac{\Delta\alpha}{\Delta\beta} = \frac{f_0(\mathbf{y}_{(j)})}{-f_1(\mathbf{y}_{(j)})} \approx -k$$

as stated.

[Comment: this is an important and useful result, but we did not prove it in the lectures, and it is not in the notes. If you look back to the operating characteristics picture above, the gradient of the curve is $\Delta\beta/\Delta\alpha \approx -1/k$, so you can almost read the value of k from the curve, providing that the aspect ratio of the plot is 1, which it is, because I specified `asp = 1` as an argument to the `plot` function.]

(e) Explain briefly how we might examine the hypotheses

$$H_0 : \theta = 0 \quad \text{versus} \quad H_1 : \theta > 0$$

in the case of the model $\mathbf{Y} \sim p(\cdot; \theta)$ with $\theta \geq 0$.

Answer. [Note: I have corrected an ambiguity in the question, by writing $\theta_0 \leftarrow 0$.]

Briefly, the *lower bound* on the likelihood ratio, over all possible prior distributions for θ with support on H_1 , is

$$\frac{p_0(\mathbf{y})}{p_1(\mathbf{y})} \geq \frac{p(\mathbf{y}; \theta_0)}{p(\mathbf{y}; \hat{\theta}_1(\mathbf{y}))}$$

where $\hat{\theta}_1(\mathbf{y}) := \operatorname{argmax}_{t>0} p(\mathbf{y}; t)$. Now if this lower bound is quite large, say above 0.3, then this rules out the possibility of strong evidence against H_0 , and so, if there is a reasonable preference for H_0 over H_1 in the prior (say, because H_0 is the current theory and H_1 is a speculative new theory), then H_0 will be chosen. On the other hand, if this lower bound is very small, say below 0.001, then the likelihood ratio could easily be small, say about 1/20, in which case we would want to choose H_1 . But it could also be large, say 20, in which case we would want to choose H_0 . It depends on the prior for θ under H_1 , which we are reluctant to specify. So in this case we are undecided.

This homework is not due until the first Tue of next Term (21 Apr). In the meantime I will circulate another homework on p -values.

Jonathan Rougier

Mar 2015