

DETAILS FOR THE SIMPLE EXAMPLES IN THE FIRST LECTURE

1. HOW DO WE GET THE CONCLUSION THAT THE LIKELIHOOD FUNCTION IS IN THE SHAPE OF A $N(\bar{x}, \sigma^2)$ DISTRIBUTION?

The log likelihood we looked at in the lecture is

$$\begin{aligned}
 (1) \quad & -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{j=1}^N \frac{(x_j - \mu)^2}{2\sigma^2} \\
 & = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{\sum_{j=1}^N x_j^2}{2\sigma^2} + \frac{\sum_{j=1}^N 2x_j\mu}{2\sigma^2} - \frac{N\mu^2}{2\sigma^2} \\
 & = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{\sum_{j=1}^N x_j^2}{2\sigma^2} + \frac{N(2\bar{x}\mu - \mu^2)}{2\sigma^2}.
 \end{aligned}$$

Remember that we are holding all the x_j 's fixed at their observed values here, and treating σ^2 as known. And we are going to divide the likelihood by a constant that makes it integrate to one. Since multiplication of the likelihood by a scaling constant adds the log of the constant to the log likelihood, this means that any additive terms in the log likelihood that do not depend on μ can be ignored — they will all wash out as we normalize to make the likelihood integrate to one. So only the last additive term in (1) matters, and we can write the likelihood itself as proportional to

$$(2) \quad e^{\frac{N(2\bar{x}\mu - \mu^2)}{2\sigma^2}}.$$

You may recall from high school algebra an exercise called “completing the square”. We do that here. That is, we note that

$$(3) \quad 2\bar{x}\mu - \mu^2 = -(\mu - \bar{x})^2 + \bar{x}^2,$$

So that the likelihood as a function of μ is proportional to

$$e^{\frac{N(\bar{x} - \mu)^2}{2\sigma^2}},$$

which is in turn proportional to a $N(\bar{x}, \sigma^2/N)$ density.

2. HOW TO COMPUTE THE LIKELIHOOD-BASED PROBABILITY INTERVALS IN SIMPLE EXAMPLE 2

As we noted in class, the pdf of the data is $p^n(1-p)^{N-n}$, where p is the unknown parameter, the probability of a default, N is the number of observations in the sample, and n is the number of defaults.

Though we call this a pdf, many, perhaps most, textbooks distinguish between probability *density* functions and probability *mass* functions, with the latter term reserved for probability distributions over a discrete set of points. Since here we are describing probabilities of finite sequences of zeroes and ones, there are only finitely many possible observations, so you might call this a probability mass function. We will just call it a pdf over a discrete set of points. (Though there are finitely many possible sequences, there are a lot — 2^{100} for $N = 100$.)

This form, $p^n(1-p)^{N-n}$, is proportional, as a function of p , to what is called the Beta($n+1, N-n+1$) distribution. When normalized to integrate to one, the Beta(n, m) pdf is

$$(4) \quad \frac{p^{n-1}(1-p)^{m-1}}{\text{beta}(p, q)},$$

where $\text{beta}(p, q)$ is the beta function. (The beta function is just defined as the integral over p of the numerator of (4).) This is a standard distribution, and R has functions that provide useful operations with most standard distributions. For the $\text{Beta}(n, m)$ distribution R provides

- pbeta** (**q**, **n**, **m**): the probability that a $\text{Beta}(n, m)$ distributed variable z satisfies $z < q$.
- qbeta** (**p**, **n**, **m**): the number $z^*(p)$ (called the “ p ’th quantile”) such that $P[z < z^*(q)] = p$.
- dbeta** (**p**, **n**, **m**): the value of the Beta density function at $z = p$.
- rbeta** (**N**, **n**, **m**): a vector of N i.i.d. random numbers, each with the $\text{Beta}(n, m)$ distribution.

There are corresponding functions for other standard distributions: `pnorm`, `qnorm`, etc for the normal, `pbinom`, `qbinom`, etc. for the binomial, and so on.

So the two-sided, equal-tailed, 95% probability interval for our $p^5(1-p)^{95}$ pdf is easily obtained. The left end of the interval is `qbeta(.025, 6, 96)` and the right end is `qbeta(.975, 6, 96)`.

The shortest, or HPD, 95% interval based on the likelihood is a little more work. As we noted in class, it has the same density height at each end. I actually initially found the interval just by trial and error — pick a p , find `pp <- pbeta(p, 6, 96)`, add .95 to this, find `pup <- qbeta(pp + .95, 6, 96)` (assuming `pp + .95 < 1`), compare `dbeta()` evaluated at p and at pup , keep going until we’ve made the `dbeta` values at each end the same. If we use the R function `uniroot()`, we can avoid the trial and error part. The clever way to do this is to define a small R function:

```
lhend <- function(p) {
  pp <- pbeta(p, 6, 96)
  ppup <- pp + .95
  if (ppup < 1) {
    qup <- qbeta(ppup, 6, 96)
    rv <- dbeta(p, 6, 96) - dbeta(qup, 6, 96)
  } else {
    rv <- 4          # just prevents solutions for non-.95 intvls
    qup <- -1
  }
  attr(rv, "qup") <- qup # so don't need separate
                        # computation of qup for solution
  return(rv)
}
```

After “sourcing” this function (there’s a button for sourcing in the RStudio edit window), the command `uniroot(lhend, c(0, 1))` spits out the lower end of the interval and, as the “qup” attribute, the upper end as well.

The function `uniroot(fcn, interval)` searches for a number in the interval defined by its second argument (the interval (0,1) in our case) that makes the function given as its first argument zero.

Note that this works only because this pdf has a single peak, declines monotonically on each side of the peak, and reaches zero at each end of the interval. When $n = 0$, the pdf is monotonically declining, so the HPD interval always has 0 as left end point.

3. HOW TO COMPUTE THE CONFIDENCE INTERVALS IN SIMPLE EXAMPLE 2

To do this we need to work with the **binomial** distribution, another standard distribution. The binomial distribution with parameters N and p gives the probability, for each possible n , of all sequences of 0’s and 1’s of length N and containing n ones, assuming each element of the sequence is an independent random variable in which the probability of a one is p . So an equal-tail, two-sided test of the null hypothesis that the true p is p^* in our sample with $n = 5$, $N = 100$, rejects p ’s for which **the probability of $n \leq 5$ is less than .025 or the probability of $n \geq 5$ is less than .05. In R notation these conditions are `pbinom(5, 100, p) < .025` or `1 - pbinom(4, 100, p) < .025`. To find the right-hand end of the confidence interval based on these two-sided tests, then, we solve for a value of p that makes `pbinom(5, 100, p) = .025`,**

and for the left-hand end we solve for a p that makes $\text{pbinom}(4, 100, p) = .975$. This can be done via `uniroot(function(p) pbinom(5,100, p) - .025, c(0,1))` and `uniroot(function(p) pbinom(4,100, p) - .975, c(0,1))`. As in the case of the post-sample intervals, these intervals can, for certain values of n and N , turn out to have a left limit of zero or a right limit of one, in which case these `uniroot()` invocations will not find any answer.

[The first version of these notes mistakenly used `pbinom(5, 100, p) > .975` where it should have used `pbinom(4,100, p) > .975.`]