Logistic Regression -IV

Prof. Dan A. Simovici

UMB

The Bernoulli Distribution

A random variable X has a *Bernoulli distribution* if

$$X : \begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix}$$

We write this as $X \sim \text{Bernoulli}(p)$.

Example

If we flip a coin that has the probability p of coming up heads and $1 - p$ of coming up tails the random variable that describes this experiment has a Bernoulli distribution.

Binomial Distribution

A random variable X has a *Binomial distribution* if

$$X : \begin{pmatrix} 0 & 1 & \cdots & k & \cdots n \\ (1-p)^n & np(1-p)^{n-1} & \cdots & \binom{n}{k} p^k (1-p)^{n-k} & \cdots & p^n \end{pmatrix}$$

We write this as $X \sim \text{Binomial}(n, p)$.

If $X_1, \ldots, X_n$ are independent random variables such that $X_i \sim \text{Bernoulli}(p)$, then $X_1 + \cdots + X_n \sim \text{Binomial}(n, p)$.

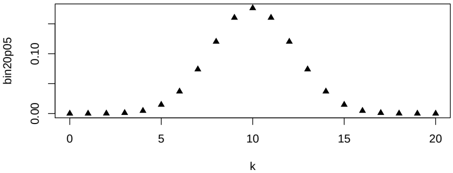
Theorem

If $X \sim Binomial(n, p)$, then $E[X] = np$ and $var(X) = np(1 - p)$.

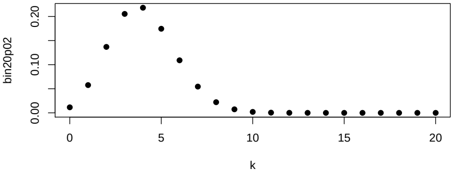
Binomial distributions can be drawn with this code:

```
k <- seq(from=0,by=1,to=20)
bin20p05 <- choose(20,k)*0.5^k*0.5^(20-k)
bin20p02 <- choose(20,k)*0.2^k*0.8^(20-k)
mylayout <- c(1,2)
layout(mylayout)
plot(k,bin20p05,pch=17,sub="Binomial dist. for p =0.5")
plot(k,bin20p02,pch=19,sub="Binomial dist. for p =0.2")
```

The previous slide code results in the graphs shown below:



Binomial dist. for p =0.5



Binomial dist. for p =0.2

Note that if X is a Bernoulli random variable with parameter p, $X \sim \text{Bernoulli}(p)$, we have:

$$P(X = k) = p^k (1 - p)^{1-k}$$

for $k \in \{0, 1\}$.

As we saw in the previous group of slides, linear regression is not suitable for predicting a probability p because it may lead to values outside the interval $[0, 1]$.
So, we replace p with the odds ratio

$$\text{odds(p)} = \frac{p}{1-p}$$

and to the *logit* function

$$\ell(p) = \ln \frac{p}{1-p}$$

for $p \in (0, 1)$. Note that $\lim_{p \to 0+} \ell(p) = -\infty$ and $\lim_{p \to 1-} \ell(p) = \infty$.
If η is a value of the logit function, $\eta = \text{logit}(p)$, then $p = \frac{e^p}{1+e^p} = L(p)$.

The likelihood function

The *likelihood function* is a basic concept in statistical inference. Suppose that we have a statistical model of an experiment involving a binomially distributed variable with a parameter p and we record the results of n experiments $x_1, \ldots, x_n$. These results are assummed to be statistically independent, so their probability is

$$P(x_1, \ldots, x_n|p) = f(x_1|p)f(x_2|p)\cdots f(x_n|p)$$

The notation "$|p$" means that the value of the parameter is supposed to be p.

Starting now from a sequence $x_1, \ldots, x_n$ we seek to determine p such that the probability $P(x_1, \ldots, x_n|p)$ is maximized. To reflect this new approach we consider the likelihood function L defined as

$$L(p|x_1, \ldots, x_n) = P(x_1, \ldots, x_n|p) = f(x_1|p)f(x_2|p)\cdots f(x_n|p)$$

and we seek p^* that maximizes $L(p|x_1, \ldots, x_n)$, that is

$$p^*(x_1, \ldots, x_n) = \mathrm{argmax}_p L(p|x_1, \ldots, x_n).$$

This is the *maximum likelihood estimate* of p.

Since $p^*(x_1, \ldots, x_n) = \text{argmax}_p L(p|x_1, \ldots, x_n)$, it follows that we also have

$$p^*(x_1, \ldots, x_n) = \text{argmax}_p \alpha L(p|x_1, \ldots, x_n)$$

for any positive α. Thus, the value p^* does not change if we multiply the likelihood function by a positive constant.

Example

The maximum likelihood for Bernoulli trials:

$$L(p|x_1,\ldots,x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i},$$

so

$$\ln L(p|x_1,\ldots,x_n) = \sum_{i=1}^{n}(x_i \ln p + (1-x_i)\ln(1-p)).$$

By differentiating $\ln L(p, x_1, \ldots, x_n)$ with respect to p and setting $\frac{\partial L(p, x_1, \ldots, x_n)}{\partial p} = 0$ we have

$$\frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\sum_{i=1}^{n}(1 - x_i) = 0,$$

hence

$$p = \frac{\sum_{i=1}^{n} x_i}{n}$$

achieves the maximum of the log likelihood. This is equivalent to

$$\text{logit}(p) = \frac{p}{1-p} = \frac{\sum_{i=1}^{n} x_i}{n - \sum_{i=1}^{n} x_i}.$$

Maximum Likelihood for the Binomial Distribution

Example

The likelihood for a binomial distribution is:

$$L(p|x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{n!}{x_i!(n-x_i)!} p^{x_i} (1-p)^{n-x_i}.$$

With the exception of the factor $\frac{n!}{x_i!(n-x_i)!}$ the likelihood is the same as the likelihood for n independent Bernoulli trials; note that the factor $\frac{n!}{x_i!(n-x_i)!}$ does not depend on p and does not affect the maximum likelohood estimate.

As we saw, we seek p such that $\text{logit}(p) = \mathbf{r}'\mathbf{x}$.
Maximum Likelihood (ML) Principle: choose as an estimate the parameter value $p*$ which would maximise the probability of what we have already observed, or the likelihood, or the logarithm of the likelihood (all equivalent).
Since $p^* = \text{argmax} L(p, \mathbf{x}) = \text{argmax} \log L(p, \mathbf{x})$, it follows that any constant multiple of the likelihood produces the same result.