# 1 Discrete distributions

## 1.1 The binomial distribution

Consider a bowl containing $r$ red balls and $N - r$ white balls, where $0 < r < N$. Draw randomly $n$ balls from this bowl *with replacement,* i.e., shake the bowl thoroughly, draw blindfolded a ball, take the blindfold off, observe the color of the ball you have drawn, *put the ball back* in the bowl (and the blindfold on!), and repeat this procedure $n$ times.

The number of ways you can draw an *ordered* sequence of $k$ red balls and $n - k$ white balls in this way is: $r^k (N-r)^{n-k}$, and the number of ways you can draw an ordered sequence of $n$ balls (of any color) is $N^n$. Thus, the probability that you draw a sequence of $k$ red balls and $n - k$ white balls *in a particular order* is: $r^k (N-r)^{n-k} / N^n = (p)^k (1-p)^{n-k}$, where $p = r/N$. But the number of *ordered* sequences of $k$ red balls and $n - k$ white balls is:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

Therefore, if $Y$ is the number of red balls you have drawn, then

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \ \ k = 0, 1, ..., n.$$

This distribution is called the Binomial $(n, p)$ distribution.

The expectation of $Y$ is:

$$E[Y] = n.p$$

## 1.2 The negative binomial distribution

Consider a sequence of independent repetitions of a random experiment with constant probability $p$ of success. Let the random variable $Y$ be the total number of failures in this sequence before the $m$-th success, where $m \geq 1$. Thus, $Y + m$ is equal to the number of trials necessary to produce exactly $m$ successes. The probability $P(Y = k)$, $k = 0, 1, 2, ....$, is the product of the probability of obtaining exactly $m - 1$ successes in the first $k + m - 1$ trials, which is equal to the (Binomial) probability

$$\binom{k+m-1}{m-1} p^{m-1} (1-p)^{k+m-1-(m-1)},$$

and the probability $p$ of a success on the $(k+m)$-th trial:

$$P(Y=k)=\binom{k+m-1}{m-1}p^m(1-p)^k,\ k=0,1,2,....$$

This distribution is called the Negative Binomial $(m,p)$ distribution.

The expectation of $Y$ is:

$$E[Y]=m\left(p^{-1}-1\right).$$

### 1.3 The Poisson distribution

Let $Y_n$ be Binomial $(n,p_n)$ distributed:

$$P(Y_n=k)=\binom{n}{k}p_n^k(1-p_n)^{n-k},\ k=0,1,...,n,$$

and suppose that for $n=1,2,....$, $p_n\downarrow 0$ as $n\to\infty$, such that for $n>c$, $np_n=c$, where $c>0$ is a constant. Then for fixed $k=0,1,2,.....$, $\lim_{n\to\infty}P(Y_n=k)=P(Y=k)$, where $Y$ is a random variable with probability function

$$P(Y=k)=\exp(-c)\frac{c^k}{k!}.$$

This distribution is called the Poisson $(c)$ distribution. Since it is the limit of a Binomial $(n,p)$ distribution with $p=c/n$ for $n>c$, the Poisson distribution is often used to model the distribution of *rare* events.

The expectation of $Y$ is:

$$E[Y]=c.$$

## 2 Count data models

These three distributions are often used to model count data. Let $Y$ be a dependent variable which is a count of something, and let $X$ be a vector of explanatory variables, including 1 for the constant term.

## 2.1 Conditional binomial

If $Y$ has a finite largest value $n$, say, so that $n$ is the smallest natural number such that $P[Y \in \{0,1,2,...,n\}] = 1$, then the conditional distribution of $Y$ may be modelled as a conditional Binomial distribution:

$$P(Y = k|X) = \binom{n}{k} p(X)^k (1 - p(X))^{n-k}, \ k = 0, 1, ..., n,$$

where

$$p(X) = F(\beta' X) \tag{1}$$

with $F$ a distribution function and $\beta$ is a parameter vector. Then the conditional expectation of $Y$ given $X$ is

$$E[Y|X] = n.p(X) = n.F(\beta' X).$$

Note that if component $\beta_i$ of $\beta$ is positive, then the corresponding component $X_i$ of $X$ has a positive effect on $E[Y|X]$:

$$\frac{\partial E[Y|X]}{\partial X_i} = n.f(\beta' X)\beta_i > 0,$$

where $f$ is the density corresponding to $F$.

If $Y$ does not have a finite upper bound, then either the negative binomial distribution or the Poisson distribution may be used to model $P[Y = k|X]$.

## 2.2 Conditional negative binomial

In the negative binomial case the model is

$$P(Y = k|X) = \binom{k + m - 1}{m - 1} p(X)^m (1 - p(X))^k, \ k = 0, 1, 2, ....$$

where

$$p(X) = F(-\beta' X), \tag{2}$$

with $F$ a distribution function and $\beta$ is a parameter vector. The reason for the minus sign is that then

$$E[Y|X] = m.\left(p(X)^{-1} - 1\right) = m.\left(F(-\beta' X)^{-1} - 1\right)$$

is increasing in $\beta' X$, so that the effect of a component $X_i$ of $X$ on $E[Y|X]$ is positive if component $\beta_i$ of $\beta$ is positive:

$$\frac{\partial E[Y|X]}{\partial X_i} = m.\left(f(-\beta' X)F(-\beta' X)^{-2}\right)\beta_i > 0.$$

## 2.3 Conditional Poisson

In the Poisson case the model for $P[Y = k|X]$ is:

$$P(Y = k|X) = \exp(-c(X))\frac{c(X)^k}{k!},$$

where

$$c(X) = \exp(\beta' X).$$

Again, if component $\beta_i$ of $\beta$ is positive, then the corresponding component $X_i$ of $X$ has a positive effect on $E[Y|X]$:

$$\frac{\partial E[Y|X]}{\partial X_i} = \exp(\beta' X)\beta_i > 0.$$

# 3 Ordered probability models

If the discrete dependent variable $Y$ represents an ordering of attributes, so that a larger $Y$ means more or better, but **not** a count of something, and $Y$ has a finite largest value $n$, say, so that $n$ is the smallest natural number such that $P[Y \in \{0, 1, 2, ..., n\}] = 1$, then $P[Y = k|X]$ may be modelled as

$$\begin{aligned}
&P[Y = 0|X] = F(-\beta' X) \qquad (3)\\
&P[Y = 1|X] = F(-\beta' X + \mu_1) - F(-\beta' X)\\
&P[Y = 2|X] = F(-\beta' X + \mu_2) - F(-\beta' X + \mu_1)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&P[Y = n-1|X] = F(-\beta' X + \mu_{n-1}) - F(-\beta' X + \mu_{n-2})\\
&P[Y = n|X] = 1 - P[Y = 0|X] - \ldots\ldots - P[Y = n-1|X],
\end{aligned}$$

where

$$0 < \mu_1 < \mu_2 < \ldots < \mu_{n-1},$$

and $F$ is a distribution function. The ordering of the parameters $\mu_j$ can be enforced easily by reparametrizing the $\mu_j$'s as

$$\begin{aligned}
&\mu_1 = \exp(\gamma_1)\\
&\mu_2 = \exp(\gamma_1) + \exp(\gamma_2)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\mu_{n-1} = \exp(\gamma_1) + \exp(\gamma_2) + \ldots + \exp(\gamma_{n-1})
\end{aligned}$$

The interpretation of the coefficients in $\beta$ is explained in the guided tour on discrete dependent variables models.

# 4 Qualitative response models

If $Y$ takes only two values, $Y = 0$ and $Y = 1$, then a the conditional distribution of $Y$ given $X$ may be modelled as:

$$P[Y = 1|X] = F(\beta'X), \tag{4}$$

where $F$ is a distribution function. If component $\beta_i$ of $\beta$ is positive, then the corresponding component $X_i$ of $X$ has a positive effect on $P[Y = 1|X]$:

$$\frac{\partial P[Y = 1|X]}{\partial X_i} = F'(\beta'X)\beta_i > 0.$$

Moreover, if $Y$ has a finite largest value $n$, say, so that $n$ is the smallest natural number such that $P[Y \in \{0, 1, 2, ..., n\}] = 1$, and $Y$ represents different attributes rather than a count or an ordering, the multinomial logit model may be an appropriate model:

$$\begin{aligned} P[Y &= 0|X] = \frac{1}{1 + \exp(\beta_1'X) + \ldots \exp(\beta_n'X)} \\ P[Y &= k|X] = \frac{\exp(\beta_k'X)}{1 + \exp(\beta_1'X) + \ldots \exp(\beta_n'X)},\ k = 1, 2, .., n. \end{aligned}$$

# 5 The choice of the distribution function F

In EasyReg International you have two options for the distribution function $F$ in (1), (2), (3) and (4), the **Logit** specification

$$F(u) = \frac{1}{1 + \exp(-u)}$$

and the **Probit** specification

$$F(u) = \int_{-\infty}^{u} \frac{\exp(-z^2/2)}{\sqrt{2\pi}} dz.$$