Since Bernoulli trials are independent from each other, we can consider that

P(having k trials until r successes) = P(having k-1trials until r-1 success) · P(having 1 success trial).

Considering that P(having k-1trials until r-1 success) meets binomial distribution, we can calculate that

$$\mathrm{P(having\ k-1trials\ until\ r-1\ success)}\ = \binom{\mathrm{k-1}}{\mathrm{r-1}} p^{r-1}(1-p)^{k-r}.$$

Therefore,

$$P(\text{having k trials until r successes})\ = \binom{\mathrm{k-1}}{\mathrm{r-1}} p^{r-1}(1-p)^{k-r}\cdot p = \binom{\mathrm{k-1}}{\mathrm{r-1}} p^{r}(1-p)^{k-r}.$$

Considering the negative binomial distribution can be seen as a combination of geometric distribution repeated for r times, and assuming that in the ith Bernoulli trial, X_i times of trials are needed to achieve the first success, we know that

$$X\ =\ \sum_{i=1}^{r} X_i.$$

Thus,

$$E(X) = E\left(\sum_{i=1}^{r} X_i\right) = \sum_{i=1}^{r} E(X_i),$$

and

$$Var(X) = Var\left(\sum_{i=1}^{r} X_i\right) =\ \sum_{i=1}^{r} E(X_i)\ .$$

From the conclusions about the geometric distribution that, knowing that

$$E(X_i)\ =\ \frac{1}{p},$$

and

$$Var(X_i)\ =\ \frac{1-p}{p^2}.$$

Hence, the expectation and standard variance of the negative binomial distribution are:

$$E(X)\ =\ \sum_{i=1}^{r} E(Xi) = r\cdot\frac{1}{p} = \frac{r}{p},$$

and

$$Var(X)\ =\ \sum_{i=1}^{r} E(X_i)\ =\ r\ \cdot\ \frac{1-p}{p^2} = \frac{r(1-p)}{p^2}.$$