Definition: NB(r, p) *distribution*

In a sequence of independent Bernoulli trials with success probability p,
if Y is the number of failures before the rth success,
then Y has the negative binomial distribution with parameters r and p,
i.e. $Y \sim \text{NB}(r, p)$.

Let $q = 1 - p$.
$P(Y = k) = P(\text{some string of } k \text{ failures and } r - 1 \text{ successes})P(\text{success})$
$= \binom{k+r-1}{k} p^{r-1} q^k p$
$= \binom{k+r-1}{k} p^r q^k.$

If $W \sim \text{Geo}(p)$ (counting failures before the first success), then
$E[W] = \frac{q}{p}$,
$V[W] = \frac{q}{p^2}$.
We can consider $Y \sim \text{NB}(r, p)$ as the sum of r iid $\text{Geo}(p)$ RVs, so
$Y = W_1 + \cdots W_r$,
$E[Y] = rE[W_1] = \frac{rq}{p}$,
$V[Y] = rV[W_1] = \frac{rq}{p^2}$.

Definition: NB2(μ, r) *distribution, i.e. mean parameterization of NB distribution*

In the original NB(r, p) distribution,
$P(Y = k) = \binom{k+r-1}{k} p^r q^k,\ k = 0,1,2 \ldots$.
Reparameterize by letting $\mu = E[Y] = \frac{rq}{p}$, so that $p = \frac{r}{r+\mu}$ and $q = \frac{\mu}{r+\mu}$, and we have
$P(Y = k) = \binom{k+r-1}{k} \left(\frac{r}{r+\mu}\right)^r \left(\frac{\mu}{r+\mu}\right)^k.$
We say that $Y \sim \text{NB2}(\mu, r)$.
In this case, $V[Y] = \frac{rq}{p^2} = \frac{\mu}{p} = \frac{\mu(r+\mu)}{r} = \mu + \frac{\mu^2}{r}$.

Extension of NB2 distribution to $r \in \mathbb{R}^+$

Using the gamma function, we can allow for positive real values of r:
$P(Y = k) = \frac{\Gamma(r+k)}{k!\Gamma(r)} \left(\frac{r}{r+\mu}\right)^r \left(\frac{\mu}{r+\mu}\right)^k,\ k = 0,1,2 \ldots$.
It may require proving that the expectation and variance are still the same for noninteger r.

Theorem: limit of $NB2(\mu, r)$ is $\text{Pois}(\mu)$

As $r \to \infty$, the $\text{NB2}(\mu, r)$ distribution approaches the $\text{Pois}(\mu)$ distribution.
Equivalently, the $\text{NB}(r, p)$ distribution approaches the Pois $\left(\frac{rq}{p}\right)$ distribution.

Proof

For $k = 0,1,2, \ldots$, $P(Y = k) = \frac{\Gamma(r+k)}{k!\Gamma(r)}\left(\frac{r}{r+\mu}\right)^r\left(\frac{\mu}{r+\mu}\right)^k$

$= \frac{\mu^k}{k!}\frac{\Gamma(r+k)}{\Gamma(r)(r+\mu)^k}\left(1+\frac{\mu}{r}\right)^{-r}$ by algebraic rearrangement.

Since $\frac{\Gamma(r+k)}{\Gamma(r)(r+\mu)^k} = \frac{(r+k-1)(r+k-2)\ldots(r)\Gamma(r)}{(r+\mu)\ldots(r+\mu)\Gamma(r)}$

$= \left(\frac{r+k-1}{r+\mu}\right)\left(\frac{r+k-2}{r+\mu}\right)\cdots\left(\frac{r}{r+\mu}\right)$

$= \left(\frac{r}{r+\mu}+\frac{k-1}{r+\mu}\right)\left(\frac{r}{r+\mu}+\frac{k-1}{r+\mu}\right)\cdots\left(\frac{r}{r+\mu}\right).$

we have $P(Y = k) = \frac{\mu^k}{k!}\left[\left(\frac{r}{r+\mu}+\frac{k-1}{r+\mu}\right)\left(\frac{r}{r+\mu}+\frac{k-1}{r+\mu}\right)\cdots\left(\frac{r}{r+\mu}\right)\right]\left(1+\frac{\mu}{r}\right)^{-r}$.

Taking limits as $r \to \infty$,

$$P(Y = k) = \frac{\mu^k}{k!}[(1)(1)\ldots(1)]e^{-\mu} = \frac{e^{-\mu}\mu^k}{k!},$$

which is the $\text{Poisson}(\mu)$ PMF. ■

Definition: zero-inflated $NB2(\mu, r)$ distribution

In a zero-inflated negative binomial model, we hypothesize that Y is generated:
* by a zero process with probability ζ, and
* by a negative binomial process with probability $1 - \zeta$.

If $f(\cdot)$ is the $NB2(\mu, r)$ PMF, then the $ZINB2(\mu, r, \zeta)$ PMF is given by

$$P(Y = k) = \begin{cases} \zeta + (1-\zeta)f(0), & k = 0 \\ (1-\zeta)f(k), & k > 0 \end{cases}.$$