Math 361 X1, Spring 2003

Normal Approximation versus Poisson Approximation

The following table gives probabilities for getting probability of having (exactly) k successes in 100 trials with success probability p for each trial, for different choices of k and p: $(p,k) = (0.5,1),(0.5,40),(0.02,1),(0.02,3),(0.02,40)$. The data was computed using the formulas

$$B(p,k) = \binom{100}{k} p^k (1-p)^{100-k}, \quad N(p,k) = \frac{e^{-\frac{(k-100p)^2}{200p(1-p)}}}{\sqrt{100p(1-p)\cdot 2\pi}}, \quad P(p,k) = e^{-100p}\frac{(100p)^k}{k!}$$

for the binomial (exact) probability , normal approximation, and Poisson approximation. (The second and third formulas are obtained by using the parameters $\mu = np = 100p$ and $\sigma = \sqrt{np(1-p)} = \sqrt{100p(1-p)}$ in the formula for the normal and Poisson approximations for P(k successes in 100 trials).

The numerical values are given by the following table. Those values of the normal and Poisson approximation that are close to the actual values are in boldface.

Parameters	Binomial $B(p,k)$	Normal $N(p,k)$	Poisson $P(p,k)$
$p=0.5, k=1$	$7.88861\cdot 10^{-29}$	$1.1146\cdot 10^{-22}$	$9.64375\cdot 10^{-21}$
$p=0.5, k=3$	$1.27559\cdot 10^{-25}$	$5.18573\cdot 10^{-21}$	$4.01823\cdot\cdot 10^{-18}$
$p=0.5, k=40$	0.0108439	**0.0107982**	0.0214995
$p=0.1, k=1$	0.000295127	0.00147728	0.000045399
$p=0.1, k=3$	0.0058916	0.00874063	0.0075665
$p=0.1, k=40$	$2.47021\cdot 10^{-15}$	$2.56487\cdot 10^{-23}$	$5.56429\cdot 10^{-13}$
$p=0.02, k=1$	0.270652	0.220797	**0.270671**
$p=0.02, k=3$	0.182276	0.220797	**0.180447**
$p=0.02, k=40$	$4.49726\cdot 10^{-41}$	$2.98453\cdot 10^{-161}$	$1.82375\cdot 10^{-37}$

The table confirms that normal and Poisson approximation should only be used when the appropriate conditions are satisfied; otherwise the results obtained by these approximations may be off by several orders of magnitude. For the normal approximation p should not be too small (nor should the complementary probability $1-p$ be too small) and k should be relatively close to $\mu = np$, the "center" the distribution; the latter condition is only satisfied in the case $k = 40$ and $p = 0.5$ and this is indeed the only case where the normal approximation is accurate (to within 0.5 %). For all other cases, the normal distribution is way off and should not be used. The Poisson approximation is extremely good (within 0.01% of the actual value) when $p = 0.02$ (which is roughly the reciprocal of $n = 100$ and $k = 1$ or $k = 3$; it gets worse when $p = 0.1$, and it is completely off in the case $p = 0.5$.