Binomial Distribution

Let p be the probability of success in a single trial and define q as $q = 1 - p$. The probability of k successes in N trials is $p(k|N) = \binom{N}{k} p^k q^{N-k}$.

1 Mean and Variance

We use the trick that $< x + y >=< x > + < y >$ and $\mathrm{Var(x + y) = Var(x) + Var(y)}$ where x and y are independent random variables. We take the N trials as N independent random variables x_i with $x_i = 0$ or 1 for failure or success, so that $< k >= N \cdot < x_i >$ and $\sigma^2 = \mathrm{Var(k) = N \cdot Var(x_i)}$. Then

$$< x_i >= 0 \cdot q + 1 \cdot p = p \qquad \rightarrow \qquad \boxed{\mu =< k >= Np}$$

(not a surprise...). For the variance:

$$< x_i^2 >= 0^2 \cdot q + 1^2 \cdot p = p$$

$$\mathrm{Var(x_i) =< x_i^2 > - < x_i >^2 = p - p^2 = p(1 - p) = pq} \qquad \rightarrow \qquad \boxed{\sigma^2 = Npq}$$

2 Gaussian limit

For large Np and large Nq the binomial distribution can be approximated as a gaussian in the "neighborhood" of the mean, i.e., for k not to different from Np. We use Stirling formula $X! \approx \sqrt{2\pi X} X^X e^{-X}$ to approximate $\binom{k}{N}$:

$$p(k|N) \approx \frac{N^N e^{-N} \sqrt{2\pi N}}{k^k e^{-k} \sqrt{2\pi k} \;\; (N-k)^{N-k} e^{N-k} \sqrt{2\pi(N-k)}} \;\; p^k q^{N-k}$$

$$p(k|N) \approx \left(\frac{Np}{k}\right)^k \left(\frac{Nq}{N-k}\right)^{N-k} \sqrt{\frac{N}{2\pi k(N-k)}} \tag{1}$$

Define a small $\delta = k - Np = Nq - (N - k)$. Then

$$\log\left(\frac{Np}{k}\right) = -\log\left(1 + \frac{\delta}{Np}\right) \qquad \text{and} \qquad \log\left(\frac{\mathrm{Nq}}{\mathrm{N - k}}\right) = -\log\left(1 - \frac{\delta}{\mathrm{Nq}}\right)$$

Using the expansion $\log(1+x) \approx x - x^2/2$, after a bit of algebra we get

$$\log\left[\left(\frac{Np}{k}\right)^k \left(\frac{Nq}{N-k}\right)^{N-k}\right] \approx -\frac{\delta^2}{2Npq} \;=\; -\frac{(k-Np)^2}{2Npq} \;=\; -\frac{(k-\mu)^2}{2\sigma^2}$$

where $\mu = Np$. Taking the exponent of the log and plugging it back into equation 1:

$$p(k|N) \approx \sqrt{\frac{N}{2\pi k(N-k)}} \quad e^{-\frac{(k-\mu)^2}{2\sigma^2}} \tag{2}$$

But $k(N-k) = (\delta + Np)(Nq - \delta) \approx N^2pq = N\sigma^2$. Plugging this into equation 2 gives

$$\boxed{p(k|N) \approx \frac{1}{\sqrt{2\pi}\sigma} \quad e^{-\frac{(k-\mu)^2}{2\sigma^2}}} \tag{3}$$

3 Bayesian Estimates of p

For k successes in N trials the posterior pdf for p is

$$\pi(p)dp \;\propto\; p^k(1-p)^{N-k}\Pi(p)dp \;\propto\; b(p; k, N)\Pi(p)dp$$

where I used the symbols π and Π for the posterior and prior, respectively; $b(x; a, b)$ is the beta distribution defined for $0 \leq x \leq 1$:

$$b(x; a, b) \;=\; \frac{1}{B(a,b)}\, x^{a-1}(1-x)^{b-1} \;=\; \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}(1-x)^{b-1}$$

where $B(a, b)$ is the Beta function. It is often convenient to pick a beta prior $\Pi(p) = b(p; a, b)$ which then yields a posterior[1]

$$\boxed{\pi(p)dp = b(p; k+a, N-k+b)\ dp} \tag{4}$$

Note $(a, b) = (1, 1)$ is a flat prior and $(a, b) = (0.5, 0.5)$ is the Jeffreys prior. The beta prior for the binomial process is a "conjugate prior", i.e., a prior that yields a posterior of the same functional form. Then based on the posterior one can construct credible intervals.

[1]See for example https://tinyurl.com/3n9b222k

3.1 Bayesian Gaussian Approximation

Given the gaussian limit derived in Section 2, it is not surprising that in the same limiting case the posterior for p will be a gaussian. The factor in the exponent of equation 3 is $-(k-Np)^2/(2Npq)$. Writing it in terms of the observed quantity $p_0 \equiv k/N$ and approximating $p \approx p_0$ and $q = (1-p) \approx (1-p_0) \equiv q_0$, the posterior with a flat prior becomes

$$\boxed{\pi(p)dp \propto e^{-\frac{(p-p_0)^2}{2\sigma_0^2}} dp}$$

with $\sigma_0^2 = p_0 q_0 / N$.

We can get to the same result starting from the beta distribution of equation 4. The mean and variance of $b(x;\alpha,\beta)$ are $\mu = \alpha/(\alpha+\beta)$ and $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$, respectively. (This can be shown easily from the integrals $\int_0^1 x b(x;\alpha,\beta)dx \propto \int_0^1 b(x;\alpha+1,\beta)dx$ and $\int_0^1 x^2 b(x;\alpha,\beta)dx \propto \int_0^1 b(x;\alpha+2,\beta)dx$ and using the normalization of the beta distribution in terms of the Beta function). With a flat prior $a = b = 1$, and for $k >> 1$ and $N-k >> 1$, the posterior of equation 4 has mean $\frac{(k+1)}{(k+1+N-k+1)} \approx \frac{k}{N} = p_0$ and variance $\frac{(k+1)(N-k+1)}{(k+1+N-k+1)^2(k+1+N-k+1+1)} \approx \frac{k(N-k)}{N^3} = \frac{p_0 q_0}{N} = \sigma_0^2$. The fact that the functional form of this posterior is also gaussian can be shown going through a very similar procedure as was done in Section 2. The algebra is a little different because the normalization factors are not the same, i.e, $\binom{N}{k} = 1/((N+1)B(k+1,N-k+1))$, see lemma A.1 of https://projecteuclid.org/euclid.ejs/1472829397.

4 Frequentist Intervals, Clopper-Pearson

For k successes in N trials the two-sided central frequentist interval $[p_L, p_H]$ at a CL of α is given by solving

$$\sum_{i=0}^{k} \binom{i}{N} p_L^i (1-p_L)^{N-i} = \frac{\alpha}{2}$$

$$\sum_{i=k}^{N} \binom{i}{N} p_H^i (1-p_H)^{N-i} = \frac{\alpha}{2}$$

The interval can also be expressed in terms of the incomplete beta function $B(x;a,b) = \int_0^x t^{a-1}(1-t)^{b-1}dt$ as

$$B(\frac{\alpha}{2}; k, N-k+1) \quad < \quad p \quad < \quad B(1-\frac{\alpha}{2}; k+1, N-k)$$

There are software packages to calculate these intervals. Note that in a strict frequentist sense the Clopper-Pearson intervals constructed this way are conservative, i.e., they overcover. This has to do with the quantized nature of the observation (k is an integer).