Lecture February, 1, 2013

BINOMIAL DISTRIBUTION, $b(p,n)$: NORMAL APPROXIMATION

- Use of tables or direct computation of $\binom{n}{x}p^x(1-p)^{n-x}$ awkward for large n, x.
- Since $X = X_1 + X_2 + \ldots + X_n$ we can use a normal approximation, invoking the central limit theorem.
- Remember:

$$\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n = e^a \qquad \lim_{n\to\infty}\left(1+\frac{a}{n}+o(1/n)\right)^n = e^a$$

$$\varphi_X(u) = 1 + \alpha_1(u) + \alpha_2\frac{1}{2!}u^2 + \alpha_3\frac{1}{3!}u^3 + \ldots$$

$$\varphi_{aX+b}(u) = E[e^{u(aX+b)}] = E[e^{(ua)X}e^{ub}] = e^{ub}\varphi_X(ua)$$

- Now easy to prove that $X \sim b(p.n)$ converges to a $N(np, \sigma^2 = npq)$ as $n \to \infty$, or equivalently $\frac{X-np}{\sqrt{npq}} \to N(0,1)$.

$$\begin{aligned}
Z = \frac{X-np}{\sqrt{npq}} &= \underbrace{\frac{1}{\sqrt{n}}\frac{X_1-p}{pq}}_{Z_1} + \ldots + \underbrace{\frac{1}{\sqrt{n}}\frac{X_n-p}{pq}}_{Z_n} \\
\varphi_Z(u) &= \prod_{i=1}^{n}\varphi_{Z_i}(u) \\
&= \prod_{i=1}^{n}\left(1+\frac{1}{2!}\left(\frac{u}{\sqrt{n}}\right)^2 + o(1/n)\right) \\
\lim_{n\to\infty}\varphi_Z(u) &= \lim_{n\to\infty}\left(1+\frac{1}{2!}\left(\frac{u}{\sqrt{n}}\right)^2 + o(1/n)\right)^n \\
&= e^{\frac{u^2}{2}}
\end{aligned}$$

- Use of approximation: n moderately large, np "away" from zero. Continuity correction.

$$P(a \le Z \le b) \approx \Phi\left(\frac{b+\frac{1}{2}-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-\frac{1}{2}-np}{\sqrt{npq}}\right)$$

  where $\Phi(x)$ is the cumulative distribution function of the $N(0,1)$. Will have an alternative approximation with Poisson distribution. $\frac{X}{n}$

Exercises.

1. Assume you have a regular coin ($P(\text{heads}) = P(\text{tails}) = 0.5$).
   (a) What is the probability that you get 6 or more heads in 10 throws?
   (b) What is the probability that you get 60 or more heads in 100 throws?

(c) What is the probability that you get 600 or more heads in 1000 throws?

2. There are to candidates A and B running for office in an upcoming election. You want to estimate the proportion of people who will vote for A; you interview a sample of 1000 randomly chosen individuals, out of which 550 declare they will vote A.

   (a) What is the probability of getting 550 A-voters or more out of 1000 if, in fact, only 45% of the people are willing to vote for A?

3. You are selling airplane tickets for a plane with 340 seats. You know from experience that 15% of the people who buy a ticket never board the plane, because of last minute problems, late connections, etc.

   (a) If you sell 355 tickets, what is the expected number of people showing up at the boarding gate? What is the probability that you will not have enough room?

   (b) How many tickets can you sell if you want that the probability of not being able to accommodate all passengers be less than 0.01?

4. An insurance company specializes in fire risks. They charge a premium of 500€ per year per house. The probability that in a year a house catches fire, is 0.002, in which case the indemnity the insurance company has to pay is 200.000€. They have insured 10000 houses.

   (a) What is the expected gross profit (excess of premiums over the cost of claims) per year? What is the probability that they incur a loss?

   (b) Assume that the company enters a reinsurance agreement with another similar company (same number of houses insured, same premium, same indemnity in case of fire). They agree to share all premiums and claims 50% each. What is now the expected gross profit and probability of loss for the first company?

Reading. [3] § 7.3 and 7.4, or [4], Chapter 25. Many other books cover these topics. Problem 2 is adapted from [1]. problem 3 from [2].

References

[1] R.B. Ash. *Basic Probability Theory.* Dover Pub., 1970.

[2] A. Garín and F. Tusell. *Problemas de Probabilidad e Inferencia Estadística.* Ed. Tébar-Flores, Madrid, 1991. In the reserved collection, signature AL-519.2(076) an in the general collection, signature 519.2(076.1) TUS.

[3] J. Martín Pliego and L. Ruiz-Maya. *Estadística I : Probabilidad.* Ediciones AC, 2004. In the reserved collection, signature AL-519.2 MAR.

[4] A. Fz. Trocóniz. *Probabilidades. Estadística. Muestreo.* Tebar-Flores, Madrid, 1987.