Continuous Random Variables - Class 17

October 24, 2013 Debdeep Pati

1 Normal approximation to the Binomial distribution

If n is large, a binomial random variable with parameters n and p will have approximately the same distribution of a normal random variable with the same mean and variance.

Theorem 1 *If S_n denotes the number of successes that occur when n independent trials, each resulting in success with probability p are performed, then, for any $a < b$,*

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \to \Phi(b) - \Phi(a).$$

Because we are using a continuous r.v. to approximate the integer-valued binomial r.v., continuity correction is usually necessary, e.g.,

1. $P(X = i) \to P(X \in (i - 0.5, i + 0.5))$
2. $P(X \geq i) \to P(X \geq i - 0.5)$
3. $P(X > i) = P(X \geq i + 1) \to P(X \geq i + 1 - 0.5)$
4. $P(X \leq i) = P(X \leq i + 0.5)$.

2 Examples of normal approximation

1. Let X be the number of times that a fair coin, flipped 40 times, lands heads. Find the probability that $X = 20$. Use the normal approximation and then compare it to the exact solution.

   $$\begin{aligned} P(X = 20) \approx P(19.5 < X < 20.5) &= P((19.5 - 20)/\sqrt{10} < X < (20.5 - 20)/\sqrt{10}) \\ &= \Phi(0.16) - \Phi(-0.16) \approx 0.1272 \end{aligned}$$

   and $P(X = 20) = \binom{40}{20}(0.5)^{40} = 0.1254$.

2. The ideal size of a first-year class at a particular college is 150 students. The college, knowing from past experience that on the average only 30 percent of those accepted for admission will actually attend, uses a policy of approving the applications of 450 students. Compute the probability that more than 150 first-year students attend this college.

   $$P(X \geq 150.5) \approx 1 - \Phi(1.59) \approx 0.0559$$

3. To determine the effectiveness of a certain diet in reducing the amount of cholesterol in the bloodstream, 100 people are put on the diet. After they have been on the diet for a sufficient length of time, their cholesterol count will be taken. The nutritionist running this experiment has decided to endorse the diet if at least 65 percent of the people have a lower cholesterol count after going on the diet. What is the probability that the nutritionist endorses the new diet if, in fact, it has no effect on the cholesterol level?

4. An examination is often regarded as being good if the test scores of those taking the examination can be approximated by a normal density function. The instructor often uses the test scores to estimate the normal parameters μ, and σ^2 and then assign the letter grade A to those whose test score is greater than $\mu + \sigma$, B to those whose score is between μ and $\mu + \sigma$, C to those whose score is between $\mu - \sigma$ and μ, D to those whose score is between $\mu - 2\sigma$ and $\mu - \sigma$, and F to those getting a score below $\mu - 2\sigma$. This is sometimes referred to as grading on the curve.

3 Exponential distribution

1. $X \sim \exp(\lambda)$ if

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, x < 0 \end{cases}$$

   Find $F(x), E(X), V(X)$

2. Examples in the real-world: the amount of time (starting from now) till an earthquake occurs, the length of a phone call in minutes

3. $X \sim Exp(\lambda)$ is memoryless in terms of

$$P(X > s + t \mid X > t) = P(X > s), s, t > 0$$

   (Similar to Geometric r.v. which has $P(X = n + k \mid X > n) = P(X = k)$)

4. An example: The probability that the instrument survives for at least s + t hours given that it has survived t hours, is the same as the initial probability that it survives for at least s hours.

5. Suppose the number of miles that a car can run before its battery wears out follows $Exp(1/10)$. If a person desires to take a 5-mile trip, what is the prob. that he completes the trip without having to replace the battery? What can be said when the distribution is not exponential? (By the memoryless property, the remaining

lifetime of the battery, denoted by X, follows $Exp(1/10)$, and thus $P(X > 5) = 1 - F(5) = e^{-5/10} = e^{-0.5} = 0.6$ However, if the lifetime distribution F is not exponential, the relevant prob is $P(lifetime > t+5 \mid lifetime > t)$, where t is the # of miles the battery had been in use. We need additional information of t to calculate the probability