Statistical Methods and Data Analysis I
Lecture 9: Binomial Distribution. Central Limit Theorem.

Oleg Goldshmidt
oleg.goldshmidt@post.idc.ac.il

Arison School of Business
Interdisciplinary Center (IDC)
Herzliya, Israel

March 31, 2019

Binomial Distribution

Definition: Bernoulli trial

An experiment with only 2 *disjoint* outcomes: *success* or *failure* — *Bernoulli random variable*.

- Assume that the observations are *independent and identically distributed* (IID).
  - Coin toss corresponds to an *IID Bernoulli variable* with $p = 0.5$.
- *What is the probability to get k successes in n experiments if the probability of success is p?*
- The probability of a *single scenario* leading to k successes is

$$P_1 = p^k(1-p)^{n-k}$$

- We also need to figure out how many such scenarios are possible... *Patience!*

Definition: binomial distribution

The *binomial distribution* describes the probability of having exactly k successes in n independent Bernoulli trials with probability of success p.

The Choose Function

- The number of ways to *choose* a scenario with k successes out of n is given by the so-called *choose function*:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

(where the *factorial* of n is $n! = 1 \times 2 \times 3 \times \cdots \times (n-1) \times n$).

- Example:

$$\binom{9}{4} = \frac{9!}{4!(9-4)!} = \frac{9 \times 8 \times 7 \times 6 \times 5!}{4 \times 3 \times 2 \times 1 \times 5!} = 3 \times 7 \times 6 = 126$$

- Luckily, we can use R:

`choose(9, 4)` $\rightarrow$ `126`

Binomial and Bernoulli Distributions

Probability of k successes in n IID Bernoulli trials

The probability of getting k successes out of n in independent Bernoulli trials with probability of success p is given by

$$P = \binom{n}{k} p^k (1-p)^{n-k}$$

Bernoulli distribution

The Bernoulli distribution is a special case of binomial distribution with number of trials $n = 1$.

- Denote the value of success as 1, the value of failure as 0.
- The *expectation of success* is given by

$$E[x] = \mu = p \times 1 + (1-p) \times 0 = p$$

- The *variance* is given by

$$Var[x] = (1-p)^2 p + (0-p)^2 (1-p) = [(1-p)p + p^2](1-p) = p(1-p)$$

Binomial Distribution: Mean, Variance, Standard Deviation

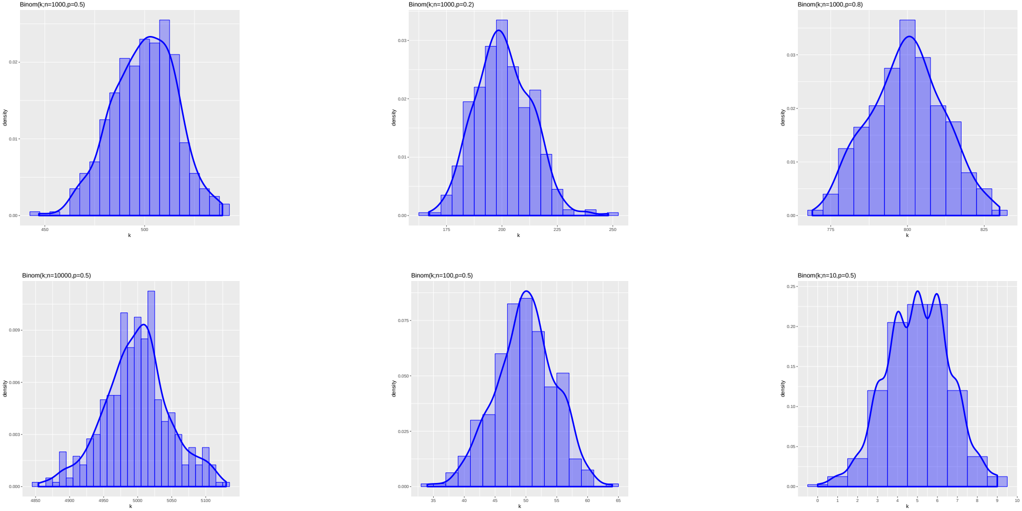
- A 2012 Gallup survey showed that roughly 26.2% of Americans are obese. *How many obese people do you expect to find in a random sample of 100 Americans?*
- Any single American is either obese ("success") or not ("failure"). The probability of success is $p = 0.262$. The probability to find k obese Americans in a sample of n is given by the *binomial distribution*.
- *The binomial random variable is simply a sum of n IID Bernoulli variables* $x_1, x_2, \ldots, x_n$
- The expected value (mean) is simple:

$$\mu = np = 100 \times 0.262 = 26.2$$

- We cannot possibly find 26.2 obese Americans in a sample of 100. There will necessarily be a deviation... *How much do we expect the number of obese Americans in a random sample of 100 to vary?*
- The variance and standard deviation of a binomial random variable are

$$Var\left[x\right] = Var\left[x_1\right] + Var\left[x_2\right] + \cdots + Var\left[x_n\right] = np(1-p), \quad \sigma = \sqrt{np(1-p)} \approx 4.4$$

Binomial Distribution: Illustration



Example: Running Advertising Campaigns

Scenario: online advertising

Our company's marketing department used to run online advertising for the company's products using the Google platform. Now Marketing want to expand into Facebook, and they are wondering whether it will be more efficient.
Advertising costs money, and we are not sure it will be efficient. So we run a small trial campaign... Our success criterion is the click-through rate.

Example: Running Advertising Campaigns

Scenario: online advertising

Our company's marketing department used to run online advertising for the company's products using the Google platform. Now Marketing want to expand into Facebook, and they are wondering whether it will be more efficient.
Advertising costs money, and we are not sure it will be efficient. So we run a small trial campaign... Our success criterion is the click-through rate.

Initial trial of 1000 impressions resulted in 116 clicks (11.6% click-through rate), which is great.

Example: Running Advertising Campaigns

Scenario: online advertising

Our company's marketing department used to run online advertising for the company's products using the Google platform. Now Marketing want to expand into Facebook, and they are wondering whether it will be more efficient.
Advertising costs money, and we are not sure it will be efficient. So we run a small trial campaign... Our success criterion is the click-through rate.

Initial trial of 1000 impressions resulted in 116 clicks (11.6% click-through rate), which is great.

Marketing don't believe the result: all the previous experience on Google shows that the success rate was closer to 3.2%, out of 10000 impressions.

Example: Running Advertising Campaigns

Scenario: online advertising

Our company's marketing department used to run online advertising for the company's products using the Google platform. Now Marketing want to expand into Facebook, and they are wondering whether it will be more efficient.
Advertising costs money, and we are not sure it will be efficient. So we run a small trial campaign... Our success criterion is the click-through rate.

Initial trial of 1000 impressions resulted in 116 clicks (11.6% click-through rate), which is great.

Marketing don't believe the result: all the previous experience on Google shows that the success rate was closer to 3.2%, out of 10000 impressions.

It makes sense to use the historical experience as an estimate of the prior probability...

Bernoulli trials: recap

A *Bernoulli random process* is a set of experiments whose *binary* outcomes are *independent* and *identically distributed*:

- *Binary*: there are only 2 disjoint outcomes: success (click-through) or failure (no click-through)
- *Independent*: experiments don't affect each other
- *Identical*: probability of success (click-through) p is the same for each experiment (impression). Probability of failure is $1 - p$.

Click-Through Model: Bernoulli Trials

Bernoulli trials: recap

A *Bernoulli random process* is a set of experiments whose *binary* outcomes are *independent* and *identically distributed*:

- *Binary*: there are only 2 disjoint outcomes: success (click-through) or failure (no click-through)
- *Independent*: experiments don't affect each other
- *Identical*: probability of success (click-through) p is the same for each experiment (impression). Probability of failure is $1-p$.

Probability of m successes out of n trials given probability of success p:

$$P(m \text{ of } n|p) = p^m(1-p)^{n-m}$$

IDC HERZLIYA

Bayesian Inference in Advertising

If the real click-through rate is θ, what is the probability of having m click-throughs out of n impressions?

$$P(m \text{ of } n|\theta) \propto \theta^m (1-\theta)^{n-m} = \left(\theta^{m/n}(1-\theta)^{1-m/n}\right)^n = \theta^{0.116n}(1-\theta)^{0.884n}$$

since $m/n = 0.116$:

Bayesian Inference in Advertising

If the real click-through rate is θ, what is the probability of having m click-throughs out of n impressions?

$$P(m \text{ of } n|\theta) \propto \theta^m(1-\theta)^{n-m} = \left(\theta^{m/n}(1-\theta)^{1-m/n}\right)^n = \theta^{0.116n}(1-\theta)^{0.884n}$$

since $m/n = 0.116$:

- Hypothesis: the new Facebook campaign is similar to the earlier Google campaigns.
- Assume prior probability distribution of the click-through rate θ (from N impressions):

$$P(\theta) \propto \theta^{0.032N}(1-\theta)^{0.968N}$$

- Posterior probability distribution:

$$P(\theta|m \text{ of } n) = P(m \text{ of } n|\theta) \times P(\theta) \propto \theta^{0.116n}(1-\theta)^{0.884n}\theta^{0.032N}(1-\theta)^{0.968N}$$

IDC HERZLIYA

Bayesian Inference in Advertising (Cont.)

Computing posterior estimate of the mean click rate:

$$
\begin{aligned}
P(\theta|m\ of\ n) &\propto \theta^{0.116n}(1-\theta)^{0.884n}\theta^{0.032N}(1-\theta)^{0.968N} \\
&= \theta^{0.032[1+2.625n/(N+n)](N+n)}(1-\theta)^{0.968[1-0.087n/(N+n)](N+n)} \\
&= \theta^{0.04(N+n)}(1-\theta)^{0.96(N+n)} \quad \Leftarrow \quad (N=10000,\ \ n=1000)
\end{aligned}
$$

i.e., the posterior estimate of the mean click rate is 4%.

Bayesian Inference in Advertising (Cont.)

Computing posterior estimate of the mean click rate:

$$
\begin{aligned}
P(\theta|m\ of\ n) &\propto \theta^{0.116n}(1-\theta)^{0.884n}\theta^{0.032N}(1-\theta)^{0.968N} \\
&= \theta^{0.032[1+2.625n/(N+n)](N+n)}(1-\theta)^{0.968[1-0.087n/(N+n)](N+n)} \\
&= \theta^{0.04(N+n)}(1-\theta)^{0.96(N+n)} \quad \Leftarrow \quad (N=10000,\ \ n=1000)
\end{aligned}
$$

i.e., the posterior estimate of the mean click rate is 4%.
We are encouraged and run another 5000 impressions resulting in 9.5% click-through rate:

$$
\begin{aligned}
P(\theta|m\ of\ n) &\propto \theta^{0.04[1+1.375n/(N+n)](N+n)}(1-\theta)^{0.96[1-0.057n/(N+n)](N+n)} \\
&= \theta^{0.057(N+n)}(1-\theta)^{0.943(N+n)} \quad \Leftarrow \quad (N=11000,\ \ n=5000)
\end{aligned}
$$

Standard Error

Definition

Assume we estimate the mean of a random sample of size n. What if we analyze any such random samples? The estimated means will be distributed, hopefully around the *real population mean*. and the distribution will have a standard deviation σ.
This standard deviation of the distribution of the estimates of the mean is called the *standard error of the mean* (SEM).

- Assume $x_1, x_2, \ldots, x_n$ are IID (independent, identically distributed) random variables — independent observations drawn from a population with mean μ and standard deviation σ.
- The mean of the observations is a *linear combination* of individual IID Bernoulli variables:

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Standard Error (Cont.)

- The *variance* of the mean is computed as the variance of a *linear combination* of individual IID Bernoulli variables:

$$Var\left[m\right] = Var\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

- The *standard error of the mean* (SEM) is

$$SEM = SD\left[m\right] = \sqrt{Var\left[m\right]} = \frac{\sigma}{\sqrt{n}}$$

- The more observations we make (the larger the sample) the more precise our estimate of the mean is.
  - However, the dependency on the number of observations is weak ($1/\sqrt{n}$):
    to improve precision by a factor of 2 we need to make 4 times more observations!

Central Limit Theorem

Central Limit Theorem (CLT)

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n, i.e., a sequence of IID random variables drawn from a population with mean μ and *finite* variance σ^2.

By the law of large numbers, the *sample average* s_n converges to μ as $n \to \infty$: $\lim_{n\to\infty} s_n = \mu$.

The *distribution of sample average* s_n converges to $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ as $n \to \infty$, *regardless of the distribution of the individual* x_i.

- We will not attempt to prove the theorem at this time — it requires advanced mathematics.
- There are multiple variants and generalizations:
  - The assumption of identical distributions may be relaxed, under certain conditions.
  - The assumption of independence may be relaxed, under certain conditions.
- The CLT is valid only *asymptotically*, i.e., for *very large n*.
- For finite n the approximation is *good near the peak* of the normal distribution, but *not for the tails*.