Intro Prob Lecture Notes

William Sun

May 1, 2017

Central Limit Theorem

- Trivial Application: Suppose $X_1, X_2, \cdots \sim$ i.i.d. Normal(μ, σ^2)
  - Unlike Poisson, these random variables are continuous
  - Define $S_n = \sum\limits_{i=1}^{n} X_i$
  - Mean, variance of S_n?
    - Please check: $E(S_n) = n\mu, Var(S_n) = n\sigma^2$
  -

    ### What is the mgf of S_n?

    $$\begin{aligned} M_{S_n}(t) &= E(e^{tS_n}) \\ &= E(e^{t\sum\limits_{i=1}^{n} X_i}) \text{ (remember independent)} \\ &= M_{X_1}(t) \cdot M_{X_2}(t) \ldots M_{X_n}(t) \text{ (remember i.i.d.!)} \\ &= \left(M_{X_1}(t)\right)^n \\ &= \left(e^{\mu t + \frac{\sigma^2 t^2}{2}}\right)^n \\ &= e^{n\mu t + \frac{n\sigma^2 t^2}{2}} \end{aligned}$$

    - $S_n \sim$ Normal$(n\mu, n\sigma^2) \to \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim$ Normal(0, 1)
  - $\frac{S_n - mean(S_n)}{stddev(S_n)}$ approaches the normal
- De Moivre's Theorem

– If $X \sim$ binomial(n, p) then

$$Y_n := \frac{X - np}{\sqrt{npq}}$$

has a distribution that is converging to the standard normal

* Calculus fact:

$$\lim_{n\to\infty} (1 + \frac{x}{n})^n = e^x$$

· Furthermore, this is a robust result.

$$\lim_{n\to\infty} (1 + \frac{x}{n} + \text{ terms with n in denominator with powers greater than 1})^n = e^x$$

· Ex: for some constant a,

$$\lim_{n\to\infty} (1 + \frac{x}{n} + \frac{a}{n^{1.2}})^n = e^x$$

– Proof:

–

$$\begin{aligned}
M_{Y_n}(t) &= E(e^{t(\frac{X-np}{\sqrt{npq}})}) \\
&= e^{-\frac{ptn}{\sqrt{npq}}} E(e^{\frac{t}{\sqrt{npq}}X}) \\
&= \text{ Note: the second term is mgf of binomial } M_X(\frac{t}{\sqrt{npq}}) \\
&= e^{-\frac{ptn}{\sqrt{npq}}} (q + p^{e^{\frac{t}{\sqrt{npq}}}})^n \\
&= \left(e^{-\frac{pt}{\sqrt{npq}}} (q + pe^{\frac{t}{\sqrt{npq}}}\right)^n \\
&= \left(qe^{-\frac{pt}{\sqrt{npq}}} + pe^{\frac{qt}{\sqrt{npq}}}\right)^n \\
&\text{Taylor Expansion} \\
&= \left(q(1 - \frac{pt}{\sqrt{npq}} + \frac{p^2t^2}{2!npq} + \ldots) + p(1 + \frac{qt}{\sqrt{npq}} + \frac{q^2t^2}{2!npq} + \ldots)\right)^n \\
&= (1 + \frac{qp^2t^2}{2npq} + \frac{pq^2t^2}{2npq} + \ldots)^n \\
&= (1 + \frac{t^2/2}{n} + \ldots)^n \\
&= \lim_{n\to\infty} (1 + \frac{t^2/2}{n} + \ldots)^n = e^{\frac{t^2}{2}}
\end{aligned}$$

* So Y_n is converging in distribution to a standard normal

A Central Limit Theorem

- If $X_1, X_2, X_3, \ldots$ i.i.d. random variables, and each has mean μ and variance σ^2, then

$$F_{Y_n}(y) = P\Big(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq y\Big) \to \Phi(y) \text{ as } n \to \infty$$

- Idea of proof: Assuming the identical distributions have an MGF (only the first two moments is sufficient)

-

$$\begin{aligned}
M_{Y_n}(t) &= E(e^{t\frac{S_n - n\mu}{\sigma\sqrt{n}}}) \\
&= e^{-\frac{\mu t n}{\sigma\sqrt{n}}} E(e^{\frac{t}{\sigma\sqrt{n}} \cdot S_n}) \\
&= e^{\frac{-\mu t n}{\sigma\sqrt{n}}} \Big(M_X(\frac{t}{\sigma\sqrt{n}})\Big)^n \\
&= e^{\frac{-\mu t n}{\sigma\sqrt{n}}} \Big(1 + \mu \cdot \frac{t}{\sigma\sqrt{n}} + \frac{\sigma^2 + \mu^2)t^2}{2\sigma^2 n} + \ldots\Big)^n \\
&= e^{-\frac{\mu t}{\sigma\sqrt{n}}} \Big(1 + \frac{\mu t}{\sigma\sqrt{n}} + \frac{\sigma^2 + \mu^2)t^2}{2\sigma^2 n} + \ldots\Big)^n \\
&= \Big((1 - \frac{\mu t}{\sigma\sqrt{n}} + \frac{\mu^2 t^2}{2\sigma^2 n} + \ldots)(1 + \frac{\mu t}{\sigma\sqrt{n}} + \frac{(\sigma^2 + \mu^2)t^2}{2\sigma^2 n} + \ldots)\Big)^n \\
&= \Big(1 + \frac{\mu t}{\sigma\sqrt{n}} + \frac{(\sigma^2 + \mu^2)t^2}{2\sigma^2 n} - \frac{\mu^2 t^2}{\sigma^2 n} + \frac{\mu^2 t^2}{2\sigma^2 n} + \ldots\Big)^n \\
&= (q + \frac{t^2}{2n} + \ldots)^n \to e^{\frac{t^2}{2}}
\end{aligned}$$

- How large should n be in order for the Central Limit Theorem approximation to be "good"?