Stats 426 Lecture 10

Thomas Cohn

2/22/21

Central Limit Theorem

Thm: Let $X_1, X_2, \ldots$ be a sequence of independent random variables, with mean 0 and variance σ^2, common cdf F, and common mgf M defined in a neighborhood of 0. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) = \phi(x),\ -\infty < x < \infty$$

Proof: We know $E(S_n) = n\mu$, $\text{Var}(S_n) = n\sigma^2$. Suppose that $E(X) = \mu$. Let $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$. Then $Z_n \xrightarrow{D} \text{N}(0,1)$. That is, for large n, the distribution of S_n is approximately $\text{N}(n\mu = 0, n\sigma^2)$. So $\bar{X}_n$ is approximately $\text{N}(\mu, \sigma^2/n)$. $\square$

Ex: Suppose $Y_n \sim \text{Bin}(n, p)$, where $p \in (0,1)$. Then

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} \text{N}(0,1)$$

Or Y_n is approximately $\text{N}(np, np(1-p))$. Our rule of thumb here is that he normal approximation of the CLT is valid when $np \geq 5$ and $n(1-p) \geq 5$. The closer p is to 0 or 1, the larger the sample size must be.

Ex: Suppose $Y_n \sim \text{Poisson}(\lambda_n)$, where $\lambda_n \to \infty$. Then

$$\frac{Y_n - \lambda_n}{\sqrt{\lambda_n}} \xrightarrow{D} \text{N}(0,1)$$

Or Y_n is approximately $\text{N}(\lambda_n, \lambda_n)$.

Ex: A coin is tossed 100 times, and 60 are heads. Is it fair?

Well, let X be the number of heads in 100 tosses. $E(X) = np = 50$, $\text{Var}(X) = np(1-p) = 25$ if the coin is fair. Assume the coin is fair. Then

$$P(X \geq 60) = P\left(\frac{X - 50}{5} \geq \frac{60 - 50}{5}\right) \approx 1 - \phi(2) \approx 0.0228$$

So it's probably not a fair coin!

Distributions Derived from the Normal Distribution

Defn: The chi-square distribution with n degrees of freedom is defined as follows: $Z \sim \text{N}(0,1)$, $U = Z^2 \sim \chi_1^2$. If $U_1, \ldots, U_n$ are independent chi-square distributions with 1 degree of freedom, then their sum is a chi-square distribution with n degrees of freedom, i.e., $Y = U_1 + \cdots + U_n \sim \chi_n^2$.

Defn: The t-distribution with n degrees of freedom is defined as follows: $Z \sim \text{N}(0,1)$, $U \sim \chi_n^2$, $Z \perp U$, then $T = \frac{Z}{\sqrt{U/n}}$ has a t_n-distribution.

Defn: The F-distribution with m and n degrees of freedom is defined as follows: $U \sim \chi_n^2$, $V \sim \chi_m^2$, $U \perp V$. $F = \frac{V/m}{U/n}$ has a $F_{m,n}$ distribution: $F \sim F_{m,n}$.

Chi-Square Distribution

- $Y \sim \chi^2_n \equiv Y \sim \text{Gamma}(\lambda = 1/2, \alpha = n/2)$
- $f_Y(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{(n/2)-1} e^{-(1/2)y}$
- $E(Y) = \frac{\alpha}{\lambda} = \frac{n/2}{1/2} = n$
- $\text{Var}(Y) = \frac{\alpha}{\lambda^2} = \frac{n/2}{1/4} = 2n$
- The mgf for $\text{Gamma}(\lambda, \alpha)$ is $(1 - t/\lambda)^{-\alpha}$ for $t < \lambda$, so the mgf for χ^2_n is

$$(1-2t)^{-n/2} = \underbrace{(1-2t)^{-1/2} \cdots (1-2t)^{-1/2}}_{n}$$

Exer: Suppose $Z_1, \ldots, Z_N \sim \text{N}(0,1)$ are i.i.d. Then $Z_1 + \cdots + Z_n \sim \chi^2_n$.

Prove by induction: For $n = 1$, use transformation or the cdf. Note that $U_1 \sim \text{Gamma}(\lambda, \alpha_1)$ and $U_2 \sim \text{Gamma}(\lambda, \alpha_2)$, with $U_1 \perp U_2$, implies that $U_1 + U_2 \sim \text{Gamma}(\lambda, \alpha_1 + \alpha_2)$.

What about the CLT? Well, if $U_i = Z_i^2$, where $Z_i \sim \text{N}(0,1)$ i.i.d., then $U_i \sim \chi^2_1$, or $\text{Gamma}(1/2, 1/2)$.
Let $S_n = U_1 + \cdots + U_n$. Then CLT implies that S_n is approximately normal, but we need to find its mean and variance.

$$E(S_n) = nE(U_1) = n \cdot 1 = n \qquad \text{Var}(S_n) = n\,\text{Var}(U_1) = n \cdot 2 = 2n$$

Thus, S_n is approximately $\text{N}(n, 2n)$.