1. DELTA METHOD

Local linear approximations: The idea of using a first-order (linear) Taylor expansion of a function g, in the neighborhood of that constant limit is a very useful technique known as the *delta method*. This module introduces this method, named after the Δ in the following expansion

$$g(x + \Delta x) \approx g(x) + \Delta x g'(x).$$

Asymptotic distributions of transformed sequences: In the central limit theorem, we consider a sequence $X_1, X_2, \ldots$ of independent and identically distributed (i.i.d.) random variables with finite variance σ^2. In this case, the central limit theorem states that

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \xrightarrow{D} Z \text{ as } n \to \infty,$$

where $\mu = \mathbf{E}[X_1]$ and Z is a standard normal random variable. In this module, we wish to consider the asymptotic distribution of some function of $\bar{X}_n$.

In the simplest case, the answer depends on results already known: Consider a *linear* function $g(t) = at + b$ for some known constants a and b. Since $\mathbf{E}[\bar{X}_n] = \mu$, clearly $\mathbf{E}[g(\bar{X}_n)] = a\mu + b = g(\mu)$ (linearity of the expectation). Therefore, it is rensonable to ask whether $\sqrt{n}\left[g\left(\bar{X}_n\right) - g(\mu)\right]$ tends to some distribution as $n \to \infty$. But, the linearity of $g(t)$ allows one to write

$$\sqrt{n}\left[g(\bar{X}_n) - g(\mu)\right] = a\sqrt{n}\left(\bar{X}_n - \mu\right).$$

We conclude by Slutsky's theorem that

$$\frac{\sqrt{n}\left[g(\bar{X}_n) - g(\mu)\right]}{\sigma} \xrightarrow{D} aZ \text{ as } n \to \infty.$$

Of course, the distribution on the right hand side above is $N\left(0, a^2\right)$.

None of the preceding developments is especially deep; one might even say that it is obvious that a linear transformation of the random variable $\bar{X}_n$ alters its asymptotic distribution by a constant multiple. Yet, what if the function $g(t)$ is non-linear? It is in this non-linear case that a strong understanding of the argument above, as simple as it may be, pays real dividends. Since $\bar{X}_n$ is consistent for μ (say), then we know that, roughly speaking, $\bar{X}_n$ will be very close to μ for large n. Therefore, the only meaningful aspect of the behavior of $g(t)$ will be its behavior in a small neighborhood of μ. In a small neighborhood of μ, $g(\mu)$ may be considered to be roughly a linear function if we use a first-order Taylor expansion. In particular, we may approximate

$$g(t) \approx g(\mu) + g'(\mu)(t - \mu)$$

for t in a small neighborhood of μ. We see that $g'(\mu)$ is the multiple of t, and the logic of the linear case above suggests

$$\frac{\sqrt{n}\left\{g(\bar{X}_n) - g(\mu)\right\}}{\sigma} \xrightarrow{D} g'(\mu)Z \text{ as } n \to \infty,$$

where $Z \sim N(0, 1)$.

Indeed, this expression is a special case of the powerful theorem known as the delta method, which we now state.

Theorem 1. *(Delta method) If $g'(a) \neq 0$ and $n^b (X_n - a) \xrightarrow{D} X$ for $b > 0$, then*

$$n^b \left\{g(X_n) - g(a)\right\} \xrightarrow{D} g'(a)X \text{ as } n \to \infty.$$

The often used *special case* of this theorem (in which the limiting random variable is normally distributed) states the following:

Theorem 2. *If $g'(\mu) \neq 0$ and $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} Z$ as $n \to \infty$, then*

$$\frac{\sqrt{n}\left\{g(\bar{X}_n) - g(\mu)\right\}}{\sigma} \xrightarrow{D} g'(\mu)Z \sim N(0, [g'(\mu)]^2) \text{ as } n \to \infty.$$

2. Poisson limit for rare events

Let $X_k \sim \text{Ber}(p)$ be independent random variables. The central limit theorem says that if p is fixed and n is large, the distribution of $(S_n - np)/\sqrt{np(1-p)}$ is close to the $N(0,1)$ distribution.

Now, we consider a slightly different situation. Let $X_1, \dots, X_n$ have $\text{Ber}(p_n)$ distribution, where $p_n = \frac{\lambda}{n}$ and $\lambda > 0$ is fixed. Then, we shall show that the distribution of $X_1 + \dots + X_n$ is close to that of $\text{Pois}(\lambda)$. Note that the distribution of X_1 changes with n and hence it would be more correct to write $X_{n,1}, \dots, X_{n,n}$.

Theorem 3. *Let $\lambda > 0$ be fixed and let $X_{n,1}, \dots, X_{n,n}$ be i.i.d. Ber(λ/n). Let $S_n = X_{n,1} + \dots + X_{n,n}$. Then, for every $k \geq 0$*

$$\mathbf{P}\{S_n = k\} \to e^{-\lambda} \frac{\lambda^k}{k!} \text{ as } n \to \infty.$$

Proof. Fix k and observe that

$$\mathbf{P}\{S_n = k\} = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Note that $\frac{n(n-1)\cdots(n-k+1)}{n^k} \to 1$ as $n \to \infty$ (since k is fixed). Also, $(1 - \frac{\lambda}{n})^{n-k} \to e^{-\lambda}$ (if not clear, note that $(1 - \frac{\lambda}{n})^n \to e^{-\lambda}$ and $(1 - \frac{\lambda}{n})^{-k} \to 1$). Hence, the right hand side above converges to $e^{-\lambda} \frac{\lambda^k}{k!}$ which is what we wanted to show. ■

What is the meaning of this? Bernoulli random variables may be thought of as indicators of events, i.e., think of $X_{n,1}$ as $\mathbf{1}_{A_1}$, etc. The theorem considers n events which are independent and each of them is "rare" (since the probability of it occurring is λ/n which becomes small as n increases). The number of events increases, but the chance of each events decreases in such a way that the expected number of events that occur stays constant. Then, the total number of events that actually occur has an approximately Poisson distribution.

Example 4. (A physical example). A large amount of custard is made in the hostel mess to serve 100 students. The cook adds 300 raisins and mixes the custard so that on an average they get 3 raisins per student. But, the number of raisins that a given student gets is random and the above theorem says that it has approximately Pois(3) distribution. How so? Let X_k be the indicator of the event that the kth raisin ends up in your cup. Since there are 100 cups, the chance of this happening is $1/100$. The number of raisins in your cup is precisely $X_1 + X_2 + \dots + X_{300}$. Apply the theorem (take $n = 100$ and $\lambda = 3$).

Exercise 5. Place r balls into m bins at random. If $m = 1000$ and $r = 500$, then the number of balls in the first bin has approximately Pois$(1/2)$ distribution. Work out how this comes from the theorem.