Section 5.3. Central Limit Theorem

Note. In this section, we state the Central Limit Theorem and give a partial proof. We consider several examples to illustrate the use of the Central Limit Theorem.

Note. In Corollary 3.4.1 we saw that if $X_1, X_2, \ldots, X_n$ are random samples from a population with distribution $N(\mu, \sigma^2)$, then $\overline{X}_n = \sum_{i=1}^n X_i/n$ has a $N(\mu, \sigma^2/n)$ distribution. Then $\sqrt{n}\,\overline{X}_n/\sigma$ has distribution

$$N((\sqrt{n}/\sigma)\mu, (\sqrt{n}/\sigma)^2(\sigma^2/n)) = N(\sqrt{n}\mu/\sigma, 1)$$

by Theorem 3.4.2, and so $\dfrac{\sqrt{n}\,\overline{X}_n}{\sigma} - \dfrac{\sqrt{n}\mu}{\sigma} = \dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$ has a $N(0,1)$ distribution. The Central Limit Theorem states that if $X_1, X_2, \ldots, X_n$ are random samples from a population with mean μ and variance σ^2 and *any* distribution, then $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ has a $N(0,1)$ distribution. The Central Limit Theorem then allows us to compute confidence intervals, p-values, and so forth. Hence, it is of central (!) importance in statistics.

Note. Next, we state the Central Limit Theorem. We present a "partial proof" by assuming the existence of a moment generating function for the distribution of the population from which the samples are taken. Recall that not every distribution has a moment generating function; Example 1.9.5 gives an example of a discrete random variable which does not have a moment generating function and Example 1.9.6 gives an example of a continuous random variable which does not have a moment generating function (in both examples, divergence of $E(e^{tX})$ is the problem).

Theorem 5.3.1. Central Limit Theorem.

Let $X_1, X_2, \ldots, X_n$ denote the observations of a random sample from a distribution that has mean μ and positive variance σ^2. Then the random variable $Y_n = (\sum_{i=1}^n X_i - n\mu)/(\sqrt{n}\sigma) = \sqrt{n}(\overline{X}_n - \mu)/\sigma$ converges in distribution to a random variable that has a normal distribution with mean 0 and variance 1.

Note. We discussed the existence of characteristic functions in some detail in Section 1.9. Some Special Expectations. In Robert B. Ash's (with contributions from Catherine Doleans-Dade) *Probability and Measure Theory*, 2nd Edition (Academic Press, 2000), a proof of the following version of the Central Limit Theorem is given. The proof is based on characteristic functions as defined in Ash (the definition is stated in our Section 1.9 notes).

Theorem 5.3.A. A k-Dimensional Central Limit Theorem of Ash.

Let $X_1, X_2, \ldots$ be independent and identically distributed ("iid") k-dimensional random vectors with finite mean μ and covariance Σ. If $S_n = \sum_{j=1}^n X_j$ then random variable $\dfrac{S_n - n\mu}{\sqrt{n}}$ converges weakly to Y, where Y has a Gaussian [normal] distribution with mean 0 and covariance Σ.

Note 5.3.A. We can also state the Central Limit Theorem as $\sqrt{n}(\overline{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$ where $\overline{X} = \sum_{i=1}^n X_i/n$. So when n is "large" (and fixed) then random variable $\overline{X}$ as a distribution which is approximately a normal distribution with mean μ and variance σ^2/n (which is the statement of the Central Limit Theorem as given in Introduction to Probability and Statistics [MATH 1530]; see my online notes for this class on Chapter 11. Sampling Distributions). We now consider applications of the Central Limit Theorem.

Example 5.3.1. Large Sample Inference for μ.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and positive variance σ^2, where μ and σ^2 are unknown. Let $\overline{X}$ and S be the sample mean and sample standard deviation, respectively. By Theorem 5.1.B, S is a consistent estimator of σ; that is, $S \xrightarrow{D} \sigma$. By Theorem 5.1.3 we have $S/\sigma \xrightarrow{P} a$, and by Theorem 5.1.4 we have $\sigma/S \xrightarrow{P} 1$. Now $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1)$. So by Slutsky's Theorem (Theorem 5.3.1; take $B_n = \sigma/S$ and $X_n = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ in Sultsky's Theorem) we have

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \xrightarrow{D} (1)N(0,1) = N(0,1).$$

Example 5.3.3. Normal Approximation to the Binomial Distribution.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a binomial distribution $b(1,p)$. As seen in our Section 3.1. The Binomial and Related Distributions, $\mu = p$, $\sigma^2 = p(1-p)$, and the moment generating function is (by Note 3.1.A) $M(t) = (1-p+pe^t)^n$ for $t \in \mathbb{R}$. If $Y_n = X_1 + X_2 + \cdots + X_n$ then by Theorem 3.1.1 Y_n had binomial distribution $b(n,p)$ which has mean np and variance $np(1-p)$. Now

$$\frac{Y_n - np}{\sqrt{np(1-p)}} = \frac{n\overline{X}_n - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\overline{X}_n - p)}{\sqrt{p(1-p)}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}.$$

So by the Central Limit Theorem (Theorem 5.3.1),

$$\frac{Y_n - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\overline{X}_n - \mu}{\sigma} \xrightarrow{D} N(0,1).$$

So we can approximate a binomial distribution with a normal distribution (Y_n itself approximately has a distribution of $N(np, np(1-p))$). This is done in Introduction to Probability and Statistics (MATH 1530) in Chapter 13. Binomial Distributions where the "rule of thumb" is to require $np \geq 10$ and $n(1-p) \geq 10$ for a good approximation. Hogg, McKean, and Craig observe that for n as small as 10 with $p = 1/2$ (so $np = 5$), the approximation can still be rather good as is shown in Figure 5.3.1 where $b(n,p) = b(10, 1/2)$ as compared to $N(np, np(1-p)) = N(5, 5/2)$.

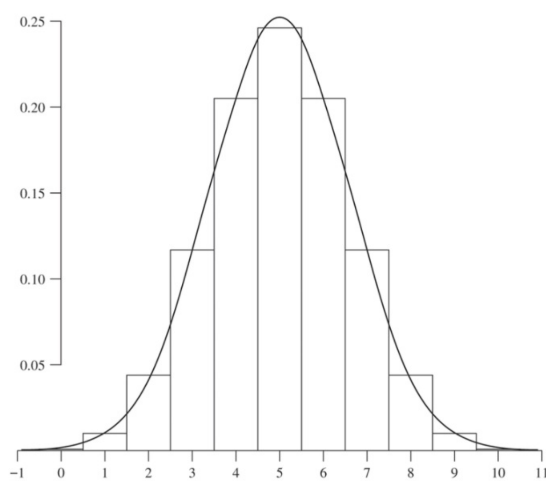

Figure 5.3.1. The binomial distribution $b(10, 1/2)$ and the normal distribution $N(5, 5/2)$.

Example 5.3.4. Consider the binomial distribution $b(n,p) = b(100, 1/2)$. Notice $np = n(1-p) = 50$. We approximate $P(Y = 48, 49, 50, 51, 52)$ using the approximation of Example 5.3.3. Since Y is a discrete integer-valued random variable then the events $\{Y = 48, 49, 50, 51, 52\}$ and $\{47.5 < Y < 52.5\}$ are equivalent events. Now

$$P(47.5 < Y < 52.5) = P\left(\frac{47.5 - 50}{5} < \frac{Y - 50}{5} < \frac{52.5 - 50}{5}\right)$$
$$= P\left(-0.5 < \frac{Y - 50}{5} < 0/5\right).$$

Since $(Y - 50)/5$ has distribution $N(0, 1)$, then the probability is $P(-0.5 < Z < 0.5) \approx 0.3829$. Notice Figure 5.3.1 again. The binomial distribution involves summing areas of rectangles of width 1 (and heights which reflect probabilities) whereas the normal approximation involves an area under a continuous function. The difference in these areas is called the *continuous correction*.

Example 5.3.5. Large Sample Inference for Proportions.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with p as the probability of success. Let $\hat{p}$ be the sample proportion of success. Then $\hat{p} = \overline{X} = \sum_{i=1}^n X_i/n$. Hence, by Exercise 5.3.13, $\frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})/n}} \xrightarrow{D} N(0, 1)$. This idea will be used in Chapter 4 to get confidence intervals for parameter p.

Example 5.3.6. Large Sample Inference for χ^2-Tests.

As in Example 5.3.3, let random variable Y_n have binomial distribution $b(n, p)$. Again (as in Example 5.3.3) $\frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})/n}} \xrightarrow{D} N(0, 1)$. By Theorem 3.4.1, the square of a normally distributed random variable has distribution $\chi^2(1)$, and by Theorem 5.2.4 if $(X_n) \xrightarrow{D} X$ and g is continuous on the support of X then $g(X_n) \xrightarrow{D} g(X)$. So with $g(x) = x^2$ we have $\left(\frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})/n}}\right)^2 \xrightarrow{D} \chi^2(1)$.