Normal distribution

Continuous distribution

$$p(x) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$

$$\mathrm{Prob}(a<x<b) = \int_a^b dx\, p(x)$$

One says $X \sim N(\mu, \sigma^2)$

For moments of continuous distributions, one often uses characteristic functions, which is Generating function with $z \to e^{ik}$

$$F(k) = \langle e^{ikx} \rangle$$

Note that $F(k) = \langle e^{ikx} \rangle = \int_{-\infty}^{\infty} dx\, p(x)\, e^{ikx}$

is really the Fourier transformation of the probability distribution.

$$F(0) = 1$$

$$\ln F(k) = (ik)\langle x \rangle + \frac{(ik)^2}{2} \mathrm{Var}(x) + \cdots$$

Also if $w = x + y$ $\quad F_w(k) = F_x(k) F_y(k)$

when x, y are independent random variables.

for normal distribution

$$F(k)=\int_{-\infty}^{\infty} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\, e^{ikx}\, dx$$

change variable to $x-\mu=y$

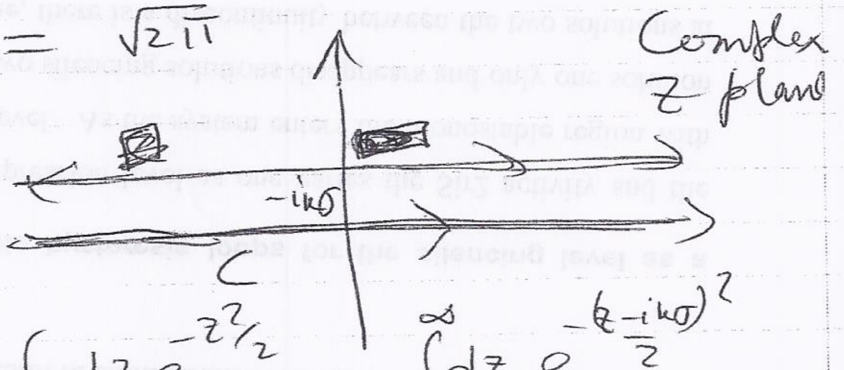
$$F(k)=e^{ik\mu}\int_{-\infty}^{\infty}\frac{e^{-\frac{y^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\, e^{iky}\, dy = e^{ik\mu+\frac{(ik\sigma^2)^2}{2\sigma^2}} \int_{-\infty}^{\infty} dy\, \frac{e^{-\frac{(y-ik\sigma^2)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

Now change variable once more to $z=\frac{y}{\sigma}$

$$F(k)=e^{ik\mu-\frac{k^2\sigma^2}{2}}\int_{-\infty}^{\infty} dz\, \frac{e^{-\frac{(z-ik\sigma)^2}{2}}}{\sqrt{2\pi}}$$

$$\int_{-\infty}^{\infty} dz\, e^{-z^2/2}=\sqrt{2\pi}$$

turns out that



$$1=\int_{-\infty}^{\infty} dz\,\frac{e^{-z^2/2}}{\sqrt{2\pi}}=\int_C dz\,\frac{e^{-z^2/2}}{\sqrt{2\pi}}=\int_{-\infty}^{\infty} dz\,\frac{e^{-\frac{(z-ik\sigma)^2}{2}}}{\sqrt{2\pi}}$$

So $F(k)=e^{ik\mu-\frac{k^2\sigma^2}{2}}$

Mean $\to \mu$ Variance $\to \sigma^2$

You can also get other moments, for example

$$\langle (x-\mu)^4 \rangle = \frac{d^4}{dk^4} e^{-k^2\sigma^2/2}\Big|_{k=0} = \frac{d^4}{dk^4}\frac{1}{2}\left(\frac{k^2\sigma^2}{2}\right)^2$$

$$= 4!\,\frac{1}{8}\,\sigma^4 = 3\sigma^4$$

Turns out that Binomial Distribution tends to normal for $N \to \infty$ in the following sense.

$$z = \frac{n - Np}{\sqrt{Npq}}$$ is distributed like $N(0,1)$.

We will do the proofs via two routes, one directly using Stirling approximation, and the other, using characteristic function.

Stirling's Formula

$$N! \simeq \sqrt{2\pi N}\, N^N e^{-N}, \text{ for large } N.$$

Start with Binomial Distr$^{\underline{n}}$

$$\boxed{P(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n}}$$

We know that this distribution is peaked around $n \simeq Np$ with a width order $\sqrt{Npq}$. So we expect log of the probability distribution change by order 1 over that scale.

Strategy: Take log and Taylor expand around the maximum in n, while use

$$\ln P(n) = \ln N! - \ln n! - \ln(N-n)! + n \ln p + (N-n) \ln q$$

$$\simeq \ln N! - \ln\left(\sqrt{2\pi n}\, n^n e^{-n}\right) - \ln\left(\sqrt{2\pi(N-n)}\,(N-n)^{N-n} e^{-(N-n)}\right) + n \ln p + (N-n) \ln q$$

independent of n

$$= \text{Const} - \left(n+\tfrac{1}{2}\right)\ln n + \cancel{n}$$

$$- \left(N-n+\tfrac{1}{2}\right)\ln(N-n) - N + \cancel{n}$$

$$+ n\ln p + (N-n)\ln q$$

$$\frac{\partial}{\partial n}\ln P(n) = \frac{n+\frac{1}{2}}{n} - \ln n$$

$$+ \frac{N-n+\frac{1}{2}}{N-n} + \ln(N-n) + \ln p - \ln q$$

$$= \underbrace{\frac{1}{2n} + \frac{1}{2(N-n)}}_{\text{ignore}} + \ln\frac{N-n}{n} - \ln(q/p)$$

The derivative is zero when $n = \bar{n}$

$$\boxed{\frac{N-\bar{n}}{\bar{n}} = \frac{q}{p}} \Longrightarrow \boxed{\bar{n} = Np}$$

$$\frac{\partial^2}{\partial n^2}\ln P(n) \approx -\frac{1}{n} - \frac{1}{N-n} + o\left(\frac{1}{N^2}\right)$$

$$= -\frac{N}{n(N-n)} + o\left(\frac{1}{N^2}\right)$$

At $n = \bar{n}$ $\quad \frac{\partial^2}{\partial n^2}\ln P(n) = -\frac{N}{\bar{n}(N-\bar{n})} = -\frac{1}{Npq}$

So the ~~[illegible]~~ Taylor expansion around $n=\bar{n}$ (which is the maximum, given that the 1st derivative is zero, and that the 2nd derivative is negative)

$$\ln P(n) = \ln P(\bar{n}) - \frac{1}{2Npq}(n-\bar{n})^2 + O\left(\frac{1}{N^2}\right)$$

$$= \text{const} - \frac{1}{2}\frac{(n-Np)^2}{Npq} + \dots$$

$$P(n) \propto e^{\text{const} - \frac{1}{2}\frac{(n-Np)^2}{Npq}}$$

~~Alter~~ Alternative proof sketch:

$$n = \sum_{i=0}^{N} s_i \qquad s_i = 0, 1$$

failure, success in the ith trial

$$F(k) = \langle e^{ikn} \rangle = \sum_{i=0}^{N} \langle e^{iks_i} \rangle = (e^{ik}p + q)^N$$

Look at the modulus

$$|e^{ik}p+q|^2 = p^2+q^2+2pq\cos k$$

$$= (p+q)^2 - 2pq(1-\cos k) = 1-4pq\sin^2\frac{k}{2}$$

$$|F(k)|^2 = \left(1-4pq\sin^2\frac{k}{2}\right)^N \approx e^{-Npqk^2}$$

The phase part $\frac{F(k)}{F(k)^*} = \left(\frac{e^{ik}p+q}{e^{-ik}p+q}\right)^N = \left(\frac{1+p(e^{ik}-1)}{1+p(e^{-ik}-1)}\right)^N$

$$= (1+2ikp+\cdots)^N \approx e^{2ikNp}$$

$$F(k) \simeq e^{iNpk - \frac{Npqk^2}{2}}$$

Compare with $F(k) = e^{ik\mu - \frac{\sigma^2k^2}{2}}$ for $N(\mu, \sigma^2)$.

In general, if a distr$\underline{\underline{n}}$ has a well defined mean μ, X_i are indep^t draw

$\bar{X}_N = \frac{X_1+\cdots+X_N}{N}$ converges in probability to μ. [Law of Large Number]

If the variance σ^2 is well defined, the distribution of $Z_N = (\bar{X}_N - \mu)/\sqrt{N\sigma^2}$ tends to $N(0,1)$

[Central Limit Theorem]